No. 20-1191

UNITED STATES COURT OF APPEALS
FOR THE FOURTH CIRCUIT

WIKIMEDIA FOUNDATION,

Plaintiff–Appellant,

v.

NATIONAL SECURITY AGENCY, *et al.*,

Defendants–Appellees.

On Appeal from the United States District Court
for the District of Maryland at Baltimore

JOINT APPENDIX—VOLUME 6 OF 7 (JA3407–JA3878)

H. Thomas Byron III
Joseph Busa
Michael Shih
U.S. DEPARTMENT OF JUSTICE
950 Pennsylvania Ave. NW
Washington, DC 20530
Phone: (202) 616-5367
Fax: (202) 307-2551
h.thomas.byron@usdoj.gov

Counsel for Defendants–Appellees

Patrick Toomey
Ashley Gorski
Charles Hogle
AMERICAN CIVIL LIBERTIES
UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
ptoomey@aclu.org

Counsel for Plaintiff–Appellant
(Additional counsel on next page)

Alex Abdo
Jameel Jaffer
KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY
475 Riverside Drive, Suite 302
New York, NY 10115
Phone: (646) 745-8500
alex.abdo@knightcolumbia.org

Deborah A. Jeon
David R. Rocah
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF MARYLAND
3600 Clipper Mill Rd., #350
Baltimore, MD 21211
Phone: (410) 889-8555
Fax: (410) 366-7838
rocah@aclu-md.org

Benjamin H. Kleine
COOLEY LLP
101 California Street, 5th Floor
San Francisco, CA 94111
Phone: (415) 693-2000
Fax: (415) 693-2222
bkleine@cooley.com

***Wikimedia Foundation v. National Security Agency, et al.*,**
No. 20-1191 (4th Cir.)

JOINT APPENDIX
Table of Contents

VOLUME 1

U.S. District Court for the District of Maryland, Docket Sheet,
Case No. 1:15-cv-00662 ..... JA0001

Plaintiff Wikimedia Foundation's Amended Complaint
(June 22, 2015), ECF No. 72 ..... JA0036

Exhibits to Wikimedia Foundation's Motion to Compel

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation
(Mar. 26, 2018), ECF No. 125-3 ..... JA0096

Exhibit 1: Chart Identifying Discovery Requests at Issue on
Wikimedia Foundation's Motion to Compel,
ECF No. 125-4..... JA0101

Exhibit 2: Wikimedia Foundation's Requests for Admission
and attachments (Nov. 7, 2017), ECF No. 125-5 ..... JA0118

Exhibits to Defendants' Opposition to Wikimedia Foundation's Motion to Compel

Declaration of Daniel R. Coats, Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-2 ..... JA0170

Declaration of Lauren L. Bernick, Senior Associate Civil Liberties
Protection Officer in the Office of Civil Liberties, Privacy, and
Transparency at the Office of the Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-3 ..... JA0190

Notice of Filing Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141 ..... JA0199

Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141-1 ..... JA0201

Exhibits to Wikimedia Foundation's Reply in Support of Its Motion to Compel

Declaration of Ashley Gorski, Counsel for Wikimedia Foundation (May 18, 2018), ECF No. 143-1 ..... JA0270

Exhibit 1: Chart Identifying Deposition Questions at Issue on Wikimedia Foundation's Motion to Compel, ECF No. 143-2 ..... JA0272

Exhibit 2: Transcript of Deposition of NSA's Designated Witness, Rebecca J. Richards, Pursuant to Fed. R. Civ. P. 30(b)(6) (Apr. 16, 2018), ECF No. 143-3 ..... JA0286

Opinion & Order Denying Wikimedia Foundation's Motion to Compel

Memorandum Opinion (Aug. 20, 2018), ECF No. 150 ..... JA0689

Order Denying Plaintiff's Motion to Compel Discovery Responses & Deposition Testimony (Aug. 20, 2018), ECF No. 151 ..... JA0716

Exhibits to Defendants' Motion for Summary Judgment

Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Nov. 13, 2018), ECF No. 164-4 ..... JA0719

Declaration of James Gilligan, Counsel for Defendants (Nov. 13, 2018), ECF No. 164-5 ..... JA0818

Exhibit 3: Wikimedia Foundation's Amended and Supplemental Responses and Objections to NSA's First Set of Interrogatories (Mar. 23, 2018), ECF No. 164-6 ..... JA0821

Exhibit 4: Wikimedia Foundation's Amended Responses and Objections to ODNI's Interrogatory No. 19 (Apr. 6, 2018), including Technical Statistics Chart, ECF No. 164-7 ..... JA0861

Exhibit 5: Wikimedia Foundation's Responses and Objections to NSA's First Set of Interrogatories (Jan. 11, 2018), ECF No. 164-8..... JA0876

VOLUME 2

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment

Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Dec. 18, 2018), ECF No. 168-2 ..... JA0920

Appendices A through Z to Declaration of Scott Bradner (Dec. 18, 2018), ECF Nos. 168-3 to 168-4 ..... JA1067

VOLUME 3

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Appendices AA through FF to Declaration of Scott Bradner (Dec. 18, 2020), ECF No. 168-5 ..... JA1791

Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Dec. 18, 2018), ECF No. 168-6 ..... JA2151

Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-7 ..... JA2218

Declaration of James Alexander, Former Manager for Trust and Safety and Former Legal and Community Advocacy Manager at Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-8 ..... JA2244

Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Dec. 18, 2018), ECF No. 168-9 ..... JA2253

Declaration of Emily Temple-Wood (Dec. 18, 2018), ECF No. 168-10 ..... JA2268

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-11 ..... JA2278

Exhibit 8: Wikimedia-hosted email list discussing NSA slide with Wikimedia logo, from July to August 2013, ECF No. 168-12 ..... JA2283

Exhibit 9: Wikimedia "Talk page" discussing its non-public information policy, from September to December 2013, ECF No. 168-13 ..... JA2305

Exhibit 10: "OTRS" ticket showing Wikimedia user requesting Tor permissions in September 2013, ECF No. 168-14 ..... JA2349

VOLUME 4

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 11: Wikimedia webpage showing Wikimedia user requesting Tor permissions in September 2017, ECF No. 168-15 ..... JA2353

Exhibit 12: Wikimedia document compiling German-user-

community appeal concerning privacy in 2013,
ECF No. 168-16 ..... JA2357

Exhibit 13: Wikimedia "Talk page" discussing NSA
surveillance from June to December 2013,
ECF No. 168-17 ..... JA2363

Exhibit 14: Wikimedia Technical Statistics Chart & Supporting
Exhibits A-G, ECF No. 168-18 ..... JA2396

Exhibit 15: Privacy & Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* (July 2014), ECF No. 168-19 ..... JA2434

Exhibit 16: FISC Memorandum Opinion, [*Redacted*], 2011 WL
10945618 (Oct. 3, 2011), ECF No. 168-20 ..... JA2631

Exhibit 17: Office of the Director of National Intelligence, *DNI Declassifies Intelligence Community Documents Regarding Collection Under Section 702 of FISA* (Aug. 21, 2013),
ECF No. 168-21 ..... JA2717

Exhibit 18: Defendant NSA's Objections and Responses to
Plaintiff's First Set of Interrogatories (Dec. 22, 2017),
ECF No. 168-22 ..... JA2721

Exhibit 19: FISC Submission, *Clarification of National Security Agency's Upstream Collection Pursuant to Section 702 of FISA*
(May 2, 2011), ECF No. 168-23 ..... JA2743

Exhibit 20: Office of the Director of National Intelligence,
Statistical Transparency Report Regarding Use of National Security Authorities, Calendar Year 2017 (Apr. 2018),
ECF No. 168-24 ..... JA2748

Exhibit 21: FISC Memorandum Opinion & Order
(Apr. 26, 2017), ECF No. 168-25 ..... JA2790

VOLUME 5

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 22: FISC Submission, *Government's Response to the Court's Briefing Order of May 9, 2011* (June 1, 2011), ECF No. 168-26 ..... JA2890

Exhibit 23: *Big Brother Watch & Others v. United Kingdom*, App. Nos. 58170/13, 62322/14, 24960/15, Eur. Ct. H.R. (2018), ECF No. 168-27 ..... JA2932

Exhibit 24: NSA Director of Civil Liberties & Privacy Office, *NSA's Implementation of FISA Section 702* (Apr. 16, 2014), ECF No. 168-28 ..... JA3145

Exhibit 25: *Legal Issues Relating to the Testing, Use, and Deployment of an Intrusion-Detection System (EINSTEIN 2.0)*, 33 Op. O.L.C. 1 (Jan. 9, 2009), ECF No. 168-29 ..... JA3157

Exhibit 26: Minimization Procedures Used by the NSA in Connection with Acquisitions of Foreign Intelligence Information Pursuant to Section 702 of FISA (July 2014), ECF No. 168-30 ..... JA3193

Exhibit 27: Glenn Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet,"* Guardian, July 31, 2013, ECF No. 168-31 ..... JA3209

Exhibit 28: NSA slide, excerpted from Exhibit 27 (Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet"*), ECF No. 168-32 ..... JA3220

Exhibit 29: Morgan Marquis-Boire, et al., *XKEYSCORE: NSA's Google for the World's Private Communications*, Intercept, July 1, 2015, ECF No. 168-33 ..... JA3222

Exhibit 30: NSA slide deck, *XKEYSCORE for Counter-CNE*, published in The Intercept on July 1, 2015, ECF No. 168-34 ... JA3237

Exhibit 31: Wikimedia, *Founding Principles* (accessed Mar. 14, 2018), ECF No. 168-35 ..... JA3259

Exhibit 32: Yana Welinder, *Opposing Mass Surveillance on the Internet*, Wikimedia Blog (May 9, 2014), ECF No. 168-36 ..... JA3262

Exhibit 33: Wikimedia Public Policy, *Privacy* (accessed Mar. 14, 2018), ECF No. 168-37 ..... JA3266

Exhibit 34: Wikipedia, *Sock Puppetry* (accessed Mar. 14, 2018), ECF No. 168-38 ..... JA3273

Exhibit 35: Wikimedia, *Privacy Policy* (accessed Feb. 14, 2018), ECF No. 168-39 ..... JA3286

Exhibit 36: Ryan Lane, *The Future of HTTPS on Wikimedia Projects*, Wikimedia Blog (Aug. 1, 2013), ECF No. 168-40 ..... JA3311

Exhibit 37: Yana Welinder, et al., *Securing Access to Wikimedia Sites with HTTPS*, Wikimedia Blog (June 12, 2015), ECF No. 168-41 ..... JA3317

Exhibit 38: Wikimedia email describing Tech/Ops goals and the importance of HTTPS (May 23, 2014), ECF No. 168-42 .... JA3325

Exhibit 39: Wikimedia document discussing IPsec implementation, including July 8, 2013 statement from a Wikimedia engineer, ECF No. 168-43 ..... JA3328

Exhibit 40: Wikimedia job posting for Traffic Security Engineer (accessed Feb. 8, 2018), ECF No. 168-44 ..... JA3364

Exhibit 41: Michelle Paulson, *A Proposal for Wikimedia's New Privacy Policy and Data Retention Guidelines*, Wikimedia Blog (Feb. 14, 2014), ECF No. 168-45 ..... JA3367

Exhibit 42: Wikimedia's Supplemental Exhibit C in response

to NSA Interrogatory No. 8 (volume of HTTP border-crossing communications by country), ECF No. 168-46 ..... JA3375

Exhibit 43: Wikimedia's Supplemental Exhibit D in response to NSA Interrogatory No. 8 (volume of HTTPS border-crossing communications by country), ECF No. 168-47 ..... JA3388

Exhibit 44: Wikimedia analytics document showing monthly unique visitors to Wikimedia by region, from December 2007 to May 2015, ECF No. 168-48 ..... JA3400

Exhibit 45: Press Release, NSA, *NSA Stops Certain Section 702 "Upstream" Activities*, Apr. 28, 2017, ECF No. 168-49..... JA3404

VOLUME 6

Exhibits to Defendants' Reply in Support of Their Motion for Summary Judgment

Second Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Feb. 15, 2019), ECF No. 178-2 ..... JA3407

Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Feb. 15, 2019), ECF No. 178-3 ..... JA3452

Second Declaration of James Gilligan, Counsel for Defendants (Feb. 15, 2019), ECF No. 178-4 ..... JA3725

Exhibit 9: Wikimedia Foundation's Responses and Objections to DOJ's First Set of Interrogatories (Jan. 11, 2018), ECF No. 178-5..... JA3728

Exhibit 10: Relevant Portions of the Deposition of James Alexander, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-6..... JA3761

Exhibit 11: Relevant Portions of the Deposition of Michelle

Paulson, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-7 ..... JA3777

Exhibit 12: Wikimedia Foundation, *Securing access to Wikimedia sites with HTTPS*, June 12, 2015 (WIKI0007108-7114), ECF No. 178-8 ..... JA3791

Exhibit 13: Wikipedia: Village pump (technical)/Archive 138 (WIKI0006872-6938), ECF No. 178-9 ..... JA3800

Exhibit 14: Jimmy Wales and Lila Tretikov, "Stop Spying on Wikimedia Users," N.Y. Times, Mar. 10, 2015, ECF No. 178-10 ..... JA3869

Exhibit 15: Wikimedia Foundation, *Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance*, Mar. 10, 2015, ECF No. 178-11 ..... JA3873

VOLUME 7

Exhibits to Wikimedia Foundation's Sur-reply in Opposition to Defendants' Motion for Summary Judgment

Second Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Mar. 8, 2019), ECF No. 181-1 ..... JA3879

Second Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Mar. 8, 2019), ECF No. 181-2 ..... JA3940

Second Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Mar. 8, 2019), ECF No. 181-3 ..... JA4006

Second Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Mar. 8, 2019), ECF No. 181-4 ..... JA4012

Second Declaration of Emily Temple-Wood (Mar. 8, 2019),
ECF No. 181-5 ..... JA4015

Exhibits to Defendants' Sur-reply in Support of Their Motion for Summary Judgment

Third Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University
(Mar. 22, 2019), ECF No. 182-2 ..... JA4019

Second Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Mar. 22, 2019), ECF No. 182-3 ..... JA4048

Opinion & Order Granting Defendants' Motion for Summary Judgment

Memorandum Opinion (Dec. 16, 2019), ECF No. 188 ..... JA4073

Order Granting Defendants' Motion for Summary Judgment
(Dec. 16, 2019), ECF No. 189 ..... JA4123

Wikimedia Foundation's Notice of Appeal

Notice of Appeal (Feb. 14, 2020), ECF No. 191 ..... JA4124

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 6

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	)	No. 1:15-cv-0662 (TSE)

(SECOND) DECLARATION OF DR. HENNING SCHULZRINNE

Dr. Henning Schulzrinne, for his (second) declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

INTRODUCTION AND SUMMARY

1. I am the Julian Clarence Levi Professor of Computer Science at Columbia University in New York, New York. I previously submitted a declaration in this case, dated November 12, 2018. I submit this second declaration at the request of the United States Department of Justice to address the conclusions reached by Mr. Scott Bradner in his December 18, 2018, declaration, including Mr. Bradner's assessment of conclusions reached in my earlier declaration. My background and qualifications in the fields of computer science, electrical engineering, and digital communications technology; the sources of information I considered in arriving at the conclusions stated in this case (apart from those cited herein); and my compensation for my services in this matter, are all stated in my prior declaration.

2. For the reasons I detail herein, it remains my conclusion that the hypothesis advanced in this case by plaintiff Wikimedia Foundation ("Wikimedia"), that the National Security Agency ("NSA"), in the course of conducting Upstream collection, must as a matter of technological necessity be intercepting, copying, and reviewing at least some of Wikimedia's electronic communications that traverse the Internet, is incorrect. Based on what is publicly

known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications. Nothing stated in Mr. Bradner's declaration, or in Wikimedia's summary judgment opposition brief ("Wikimedia Brief") alters that conclusion.

3. Mr. Bradner addresses a question quite different from the issue that I was asked to address: the likelihood that the NSA has copied and scanned for targeted selectors at least some of Wikimedia's international text-based Internet communications in the course of conducting Upstream collection. Declaration of Scott Bradner (Plaintiff's Exhibit 1) ("Bradner Decl.") ¶¶ 1(c), 6(e). As both Mr. Bradner and Wikimedia remark, my earlier declaration did not address that issue. Bradner Decl. ¶ 7(a); Wikimedia Brief at 12. Nor do I address it here. That is in part because the Department of Justice has not asked that I opine on that question, which implicates matters that remain classified. I also do not address that question because to arrive at an answer would require that I, like Mr. Bradner, engage in speculation about the NSA's surveillance priorities, practices, and capabilities, and the number and nature of its Upstream surveillance targets, matters about which neither I nor, apparently, Mr. Bradner, has any specialized knowledge or information. I have explained what it is technologically possible for the NSA to do, if it wishes, in conducting Upstream surveillance, and explained how it could avoid interaction (whether by design, or effect) with Wikimedia's communications. But I do not attempt to reach conclusions about what it actually does based on assumptions that lack a basis either in Internet technology or engineering, or pertinent information about the Upstream program.

4. What I do address in this declaration, at the Justice Department's request, are the bases on which Mr. Bradner concludes that it is "virtually certain" that the NSA, in conducting Upstream surveillance, has copied and reviewed at least some of Wikimedia's communications. Bradner Decl. ¶ 6(e). Specifically, I discuss the bases of his conclusions (i) that the NSA, in conducting Upstream surveillance, "most likely" copies, reassembles, and scans for selectors all

communications packets traversing an international Internet link that is monitored by the NSA (if any); (ii) that it is "implausible" that the NSA uses the traffic-mirroring techniques (white- and blacklisting) described in my first declaration; and (iii) that even if the NSA uses one or more of the techniques I described, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. As I explain below, Mr. Bradner's conclusions rest principally on assumptions he makes about the NSA's practices and priorities, its resources and capabilities, and its Upstream surveillance targets, and are not based on facts and information concerning Internet technology and engineering. (As was the case with my first declaration, in reaching the conclusions stated herein I have not considered, nor have I been provided with, any classified or other non-public information concerning Upstream surveillance.)

5. It is not the case, as Wikimedia states, that my analysis "ignores key features" of and "critical" disclosures about Upstream surveillance, such as the supposed comprehensiveness of its goals, collection of wholly domestic "about" communications, and the trade-offs of various possible collection system configurations. *See* Wikimedia Brief at 1, 21-22. Rather, these matters were not pertinent to the question of technical feasibility that I was asked to address. And as I explain below, they also provide no support, and certainly none based in Internet technology and engineering, for concluding that the NSA "almost certainly" (Bradner Decl. ¶ 6(a)) copies and scans all communications traversing any circuit it monitors, including Wikimedia's.

6. Finally, at the Justice Department's request, I discuss reasons why, in today's digital communications environment, (i) an organization operating any but the smallest websites would not want to compromise the interests of security, and user privacy, by failing to implement HTTPS encryption "by default," and (ii) an organization that transmits confidential information over the Internet from one business location to another would not want to risk compromising the security and privacy of that information by failing to encrypt it through, for example, the Internet Protocol Security (IPsec) suite.

SUMMARY OF CONCLUSIONS REACHED IN MY FIRST DECLARATION

7. In my prior declaration, I explained that, contrary to Wikimedia's central hypothesis, the NSA could be conducting Upstream surveillance, at least as it is envisioned by Wikimedia, in a number of technically feasible, readily implemented ways that would not require the NSA to electronically copy, electronically scan for selectors (the process Mr. Bradner refers to, less precisely, as "reviewing"), or otherwise interact with *all* communications traversing any given Internet backbone link where Upstream surveillance might theoretically be conducted. I also concluded, more specifically, that if the NSA (hypothetically) were using these techniques to conduct Upstream surveillance, then it could do so without accessing or interacting with Wikimedia's communications. Declaration of Dr. Henning Schulzrinne (Government Exhibit 1) (hereinafter, "First Decl.") ¶¶ 1, 15, 53. These techniques, types of traffic-mirroring known as whitelisting and blacklisting, involve programming a router or switch at a monitored link, using access control lists to selectively mirror (that is, copy) only those communications that are deemed most likely to include communications of interest, without copying or otherwise handling those that are not.

8. For example, if a collecting entity has information indicating that communications of interest are associated with particular IP addresses (or blocks of IP addresses), then at each monitored link an assisting telecommunications service provider could configure its router or switch with a whitelist of the specified IP addresses, which would allow only those communications packets containing source or destination IP addresses on the whitelist to be copied and passed through an interface with collection equipment to be scanned, and (where targeted selectors are detected) retained in the collector's databases. First Decl. ¶¶ 65-66.

9. Conversely, if the collecting entity determines that communications traffic to and from certain IP addresses, or blocks of IP addresses, are of little or no interest for its purposes, then it may request the telecommunications service provider at a given monitored link to configure its router or switch with a blacklist of these IP addresses that would prevent any

packets to or from those addresses from being copied and passed through the interface with the collector's equipment to be scanned. First Decl. ¶¶ 67-68.

10. As I also explained, distinct types of communications, such as e-mail, or web (HTTP and HTTPS) communications, can be whitelisted or blacklisted in the same fashion, based on their assigned port numbers. First Decl. ¶¶ 70-71.

11. Mr. Bradner does not maintain that it is necessary, as a technical matter, to copy all communications traversing a given Internet backbone link that a collecting entity may be monitoring in order to obtain access just to some of them. He repeatedly acknowledges, in fact, that it is possible to copy only a subset of the traffic crossing that link, that is, the communications packets meeting specified criteria such as source or destination IP addresses, or port or protocol numbers. Bradner Decl. ¶¶ 272(b), 280-81, 299, 325, 366.

12. As I also explained, using the traffic-mirroring techniques described in my first declaration, it would be technically feasible for the NSA to conduct Upstream surveillance, at least as it is envisioned in Wikimedia's First Amended Complaint (what I referred to in my first declaration as "Upstream-type" surveillance, *see* First Decl. ¶¶ 15, 77, 88), without copying, scanning, or otherwise interacting with any of the three categories of communications that Wikimedia believes are subject to NSA Upstream collection processes. First Decl. ¶ 77. The first category includes HTTP and (principally) HTTPS requests from individual Internet users to the servers housing Wikimedia's websites, and the responses thereto. First Decl. ¶ 78. These communications could be blacklisted by configuring the router or switch at any monitored link to prevent any communications with port numbers 80 and 443, respectively, from being copied and forwarded to NSA scanning equipment. First Decl. ¶ 79. Alternatively, Wikimedia's HTTP and HTTPS communications could be excluded from those made available to the NSA at a monitored link through (i) blacklisting communications to or from Wikimedia IP addresses, or (ii) using whitelists of IP addresses (that do not include Wikimedia's IP addresses) to exclude Wikimedia HTTP and HTTPS communications (and other communications of no intelligence interest) except

those exchanged with users (if any) who had been assigned targeted (whitelisted) IP addresses. First Decl. ¶¶ 80-81.[1]

13. Similarly, NSA access to Wikimedia's second category of communications, encrypted log communications, could be blocked by blacklisting communications containing the protocol number (50) of the IPSec protocol used by Wikimedia to encrypt its log communications, or, as in the case of Wikimedia's Category 1 communications, whitelisting or blacklisting by IP address. First Decl. ¶¶ 83-84. The third category, various types of online communications engaged in by Wikimedia's staff, could also be blocked, in the same manner as Wikimedia HTTP and HTTPS communications, through white- or blacklisting by IP address. First Decl. ¶ 87.

MR. BRADNER'S ASSERTION THAT THE NSA "MOST LIKELY"CONDUCTS UPSTREAM SURVEILLANCE USING HIS COPY-ALL-THEN-SCAN APPROACH IS WITHOUT A BASIS IN INTERNET TECHNOLOGY AND ENGINEERING

14. Although Mr. Bradner acknowledges the technical feasibility of the white- and blacklisting techniques I describe in my earlier declaration, he states that in his view it is "most likely" that the NSA does not use these filtering techniques, and instead copies and (after reassembling the packets) scans for selectors all communications, including Wikimedia's, that traverse any international Internet backbone link (if any) where the NSA conducts Upstream surveillance. Bradner Decl. ¶¶ 282-89, 366-67.

15. For the most part, Mr. Bradner's conclusions regarding the "likely" manner in which the NSA actually conducts Upstream collection are based on assumptions about the NSA's policies and practices in conducting covert surveillance, its operational priorities, and resources, and the number of its Upstream surveillance targets. As I stated above, I have not been asked to offer views on the "likely" manner in which the NSA actually conducts Upstream collection, because of its classified nature, and I do not venture an opinion on that subject for the very reason that an informed opinion would require knowledge and information about the classified

[1] I further explained that whitelisting and blacklisting by IP address in this fashion would also block NSA access to the SMTP (e-mail) communications, of unspecified number and geographic distribution, that Wikimedia also includes in Category 1 of its communications. First Decl. ¶ 82.

operational details of the Upstream program that I (and by all appearances, Mr. Bradner) lack. To a lesser extent, Mr. Bradner attempts to rest his conclusions on what he assumes would be the preferred equipment configuration of an assisting telecommunications service provider, but ultimately this assumption rests on a misreading of my earlier declaration, and further speculation regarding the number of the NSA's Upstream targets.

16. Mr. Bradner considers it likely that an assisting provider at any monitored link furnishes the NSA with copies of all communications packets traversing that link, after which (in Mr. Bradner's opinion) the NSA reassembles the packets and then reviews (electronically scans) the reconstructed communications for targeted selectors in order to identify those that will be retained in NSA databases. Bradner Decl. ¶¶ 273-79, 282-89. Mr. Bradner refers to this process as a "copy-then-filter" configuration." Bradner Decl. ¶ 272(a). For clarity of distinction, I will refer to this approach herein as "copy-all-then-scan." Mr. Bradner considers this approach more likely than use of the traffic-mirroring techniques described in my first declaration, which he refers to as "in-line filtering," Bradner Decl. ¶¶ 272(b), 280-89, and which I will refer to herein (again for clarity of distinction) as "filter-then-copy-and-scan."

17. Mr. Bradner gives four reasons for considering his copy-all-then-scan approach more likely than the filter-then-copy-and-scan techniques described in my first declaration. The first two concern the supposed operational preferences of the NSA, and the latter two the supposed preferences of an assisting provider. The reasons Mr. Bradner gives are:

a. that whitelisting or blacklisting (by IP address or port or protocol number) before copying communications and scanning them for targeted selectors would require the NSA to share sensitive information about its targets and/or filtering criteria with the assisting provider's personnel. Bradner Decl. ¶¶ 283, 285-87.

b. that whitelisting or blacklisting to reduce the volume of communications that must be reassembled and scanned for selectors would be of little offsetting benefit to the NSA given the real-time processing capacity of modern packet-inspection devices. Bradner Decl. ¶ 288.

c. that his suggested copy-all-then-scan configuration would not require the placement of "an NSA-operated device into the heart of [the provider's] network," which would risk an adverse impact on the provider's network "in the event of an equipment failure or misconfiguration." Bradner Decl. ¶ 284.

d. that configuring a router or switch to filter the communications made available to the NSA could create a risk of "overloading" the router and impairing the provider's ability to support its customers' traffic. Bradner Decl. ¶¶ 288, 366(c).

I address these purported justifications in turn.

18. Sharing Sensitive Information: The extent to which the NSA is willing (or finds it necessary) to share classified information with an assisting provider in order to conduct Upstream surveillance (or any other kind of collection activity) is a matter about which neither I nor (so far as his declaration reveals) Mr. Bradner has any specialized knowledge or information. I do not consider uninformed assumptions about the NSA's willingness to share such information with an assisting provider to be a basis on which to reach conclusions, from the perspective of Internet technology and engineering, about the manner in which the NSA conducts Upstream surveillance. I observe, however, that according to the PCLOB Section 702 Report cited by Mr. Bradner and Wikimedia, Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act* (July 2, 2014), https://www.pclob.gov/library/702-Report.pdf, the NSA already shares sensitive information about its surveillance targets with assisting provider(s), specifically, the selectors (such as their e-mail addresses and telephone numbers) used to identify their communications for acquisition. PCLOB Section 702 Report at 36. This would appear to call Mr. Bradner's premise into question.

19. Operational Benefit of Filtering Traffic Before Scanning: Similarly, whether the NSA places greater importance on the potential intelligence value of scanning every communication that crosses a given Internet backbone link, or the operational efficiencies and cost-savings that would flow from first filtering out communications of low interest, is a matter about which neither I nor, evidently, Mr. Bradner, has knowledge. That said, the practical benefits to be gained from first filtering out low-interest communications cannot be dismissed on the grounds suggested by Mr. Bradner. Mr. Bradner simply asserts, without supporting data or explanation, that there is little to be gained by reducing the volume of traffic that must be copied and scanned for selectors because (in his estimation) the most powerful, commercially

available packet-inspection devices are capable of scanning all traffic in real time as it traverses a given monitored link. *See* Bradner Decl. ¶ 288 ("Modern deep packet inspection devices, individually or operating in parallel, can process or review Internet communications at the same rate that those communications traverse high-bandwidth Internet links."). (He also implicitly assumes that these high-end commercial packet-inspection devices, or devices of similar capacity, are those actually employed by the NSA.).

20. Readily available data concerning traffic volume at the links that Wikimedia claims are monitored by the NSA, and the processing capacity of the packet-inspection devices that Mr. Bradner refers to, contradict his assumptions. The data transfer rates (i.e., the traffic flow) at international Internet links of the kind that Wikimedia presumes are monitored by the NSA are likely to surpass the processing capacity of even the best-resourced entities. For example, the AEConnect link, a transatlantic cable running between Shirley, New York and Killala, Ireland, put into service in 2016, has an aggregate capacity of approximately 40 terabits (over 40 trillion bits) of information per second[2]. The newer MAREA link, a suboceanic cable connecting Virginia Beach, Virginia with Bilbao, Spain, has a capacity of 160 terabits (over 160 trillion bits) of information per second.[3] According to the telecommunications market research firm TeleGeography, inter-regional Internet capacity has increased to 98 terabits per second, while total international capacity (a large fraction of which is known to originate or terminate in the United States) reached 295 terabits per second in 2017[4]. In contrast, the largest commercial deep-packet-inspection (DPI) devices, such as the NIKSUN Supreme Eagle, typically have Ethernet interfaces with speeds of no more than 100 gigabits (100 billion bits) per second,[5] that is, 1/10 terabit per second, as those are the fastest standardized Ethernet interfaces that are commercially available.

[2] https://www.submarinenetworks.com/systems/trans-atlantic/aeconnect
[3] https://en.wikipedia.org/wiki/MAREA
[4] https://blog.telegeography.com/295-tbps-internet-traffic-and-capacity-in-2017
[5] https://www.niksun.com/c/1/ds/NIKSUN_datasheet_Supreme_eagle.pdf

21. Not all transoceanic fiberoptic cables, especially those more recently put into service, operate at full capacity. But even at half capacity (20 terabits per second for AEConnect, 80 terabits per second for MAREA), to copy and scan all traffic crossing the AEConnect link, as Mr. Bradner suggests, would require 200 packet-inspection devices, and at the MAREA link 800 such devices, with racks upon racks full of monitoring gear. In addition, a collection system such as Mr. Bradner envisions would require installation of an opto-electronic device, such as a router, to convert the single high-speed stream of traffic flowing over the link into hundreds of lower-speed streams feeding into the individual packet-inspection devices. Acquiring, deploying, operating, and maintaining one or more collection systems of this kind would be enormously costly, and present formidable technical and logistical challenges. Thus, there are potentially many reasons to have the assisting carrier's router perform the preliminary filtering, removing most of the traffic that is likely not of interest before communications are copied, reassembled, and scanned. (For example, removing encrypted web traffic alone would likely reduce the volume to be copied and scanned by about half.)

22. Mr. Bradner also makes the related observation that "[i]f filtering traffic for performance reasons were desirable, the NSA would get much more result from filtering YouTube than from filtering Wikimedia." Bradner Decl. ¶ 366(c). That observation ignores the fact that the volume of streaming video traffic at international links is far lower than at domestic links, since most video streaming services, such as Netflix, do not transmit programming across international boundaries. (Most viewers access streaming video services, such as Netflix, YouTube, or Hulu, from content distribution networks located within their home countries, or at least their own continents.) Therefore, the benefit of filtering out only video traffic crossing an international Internet link (where Wikimedia supposes that Upstream surveillance is conducted) would likely be modest. Moreover, Mr. Bradner's observation is a non-sequitur; the two options, filtering video traffic, and filtering Wikimedia, are not mutually exclusive. If the NSA wished to avoid scanning large volumes of communications traffic of (hypothetically) no intelligence

interest, then there is no technological reason why it could not filter both video traffic *and* traffic from high-volume websites such as Amazon.com and Wikipedia.

23. Placement of NSA Equipment in a Provider's Network: I turn now to the reasons given by Mr. Bradner for his view that a provider would prefer his copy-all-then-scan configuration to a filter-then-copy-and-scan approach. The central premise of this conclusion is that the filter-then-copy-and-scan approach I describe in my first declaration involves placement of "an NSA-operated device into the heart of [a provider's] network." Bradner Decl. ¶ 284. That assertion is simply mistaken. I did not propose in my earlier declaration a filtering configuration involving the use of "NSA-operated devices." The whitelisting and blacklisting techniques I described all involve the use of existing provider equipment, that is, the provider's own network switch or router, programmed with access control lists according to NSA criteria, in order to perform the very kind of traffic-mirroring function for which routers and switches are designed and utilized in the ordinary course of a provider's business. First Decl. ¶¶ 65-71.

24. Moreover, the techniques I described would pose little risk to the operational integrity of the provider's network, i.e., the unimpeded flow of communications traffic to and from the provider's customers. Adding monitoring capability, with the constraints noted, to an existing carrier router does not affect carrier operations, requires minimal physical additions to carrier facilities (such as running an extra fiber patch cable) and operationally speaking is imperceptible to the carrier. It only requires periodic configuration (programming) changes in the white- and blacklists that are stored electronically in the router's memory, which can be accomplished (either by NSA or provider personnel) through remote rather than onsite access, not unlike the way a document stored on a server can be updated and edited by a user who connects to the server from another location.

25. Risk of "Overloading" the Provider's Router: When Mr. Bradner states that whitelisting or blacklisting "could affect the performance of the [provider's] router and create a risk of overloading the router," Bradner Decl. ¶ 288, he is again making assumptions, this time about the extent to which filtering communications made available to the NSA would place

demands on a router's processing capacity. That would depend in large part on the number of IP addresses (or blocks of addresses) that the NSA designates for white- or blacklisting at a given monitored link, which in turn would be related, to a significant degree, to the number of the NSA's Upstream targets (and to the number of targets whose communications the NSA is seeking to capture at that particular link). These are also matters about which Mr. Bradner has no knowledge or information, at least so far as his declaration reveals.

26. There are, in addition, a number of other practical considerations, overlooked by Mr. Bradner, that weigh against choosing Mr. Bradner's copy-all-then-scan approach over a filter-then-copy-and-scan approach. There are two possible equipment configurations for implementing Mr. Bradner's suggested approach, (a) the use of an optical splitter to create a copy of the entire communications stream flowing across a monitored circuit, *see* First Decl. ¶ 55, or, alternatively, (b) configuring the provider's router to mirror all incoming or outgoing packets to the collection infrastructure. Both pose risks of adverse impacts on the network to which a provider might object.

27. When an optical splitter is connected to an optical fiber, all communications traffic carried on that fiber is directed into the splitter, where the stream is duplicated by dividing the optical power of the stream between two (or more) outputs, allowing the "original" communications stream to continue to its intended destination, albeit at reduced optical power, while the duplicate stream(s) may be diverted for other purposes, including scanning for communications of interest. Thus, while passive optical splitters are relatively simple components, adding a splitter to facilitate Upstream collection would introduce another potential failure point to a provider's network, and at best introduce a degree of optical power loss. Generally, to ensure high reliability and easy maintenance, optical network architectures try to minimize both the number of connectors and optical loss.

28. As an alternative to optical splitting, a router could theoretically be configured to mirror all incoming or outgoing packets to the monitoring infrastructure. However, no commercial router I am aware of is designed to mirror all incoming or outgoing traffic at once.

Rather, carrier-scale routers may only be able to mirror traffic to one or two interfaces by design. To mirror all traffic (if at all possible) would require adding interfaces capable of supporting the total input or output capacity of the router, potentially exceeding its design capacity and doubling its cost. (Interface cards constitute the largest single cost component of a carrier-grade router. Each router model has a maximum interface capacity that constrains the number and speed of interfaces.)

29. I do not mean by these observations to suggest how the NSA or the assisting provider at a given link theoretically monitored by the NSA would choose to configure an Upstream collection system. I am saying, however, that the reasons given by Mr. Bradner for suggesting that the NSA and the provider would prefer his copy-all-then-scan configuration over a filter-then-copy-and-scan approach (i) make assumptions about the NSA's surveillance practices, priorities, and resources, (ii) are based on a misunderstanding of the configuration I proposed in my earlier declaration, (iii) rely on assumptions, rather than information, concerning the number of the NSA's Upstream targets, and (iv) overlook important considerations that a provider concerned about network integrity would take into account. I do not consider them a reliable technological basis on which to conclude that it is "most likely" the NSA is using a copy-all-then-scan an approach, as asserted by Mr. Bradner.

MR. BRADNER'S VIEW THAT USE OF A FILTER-THEN-COPY-AND-SCAN APPROACH IS "IMPLAUSIBLE" ALSO LACKS A BASIS IN INTERNET TECHNOLOGY AND ENGINEERING

30. Conversely, Mr. Bradner gives a number of reasons why he concludes that the filter-then-copy-and-scan approach described in my first declaration is "implausible." Bradner Decl. ¶¶ 366-67. A number of these reasons are the same as those he gives in support of his conclusion that the NSA "most likely" uses his copy-all-then-scan approach. *See* Bradner Decl. ¶ 366(a) (providing sensitive information to the assisting provider); ¶ 366(c) (no need to reduce the processing load on packet-inspection devices). I have already addressed the conjectural nature of these points and why, as a technological matter (that is, from the perspective of Internet technology and engineering) they do not support Mr. Bradner's conclusions. Below I

address the remaining bases of Mr. Bradner's "plausibility" conclusion, which may be divided into three categories: (a) those concerning port and protocol blocking (Bradner Decl. ¶ 366(b), (e)-(h)); (b) those concerning whitelisting (Bradner Decl. ¶ 366(d)); and (c) those concerning blacklisting Wikimedia IP addresses (Bradner Decl. ¶ 367(a)).

Blocking Port or Protocol Numbers (HTTP and HTTPS communications)

31. Mr. Bradner first considers it implausible that the NSA would blacklist particular types of communications by port or protocol number, in particular HTTPS communications (encrypted web communications), for a variety of reasons. Bradner Decl. ¶ 366(b), (e)-(h). As I now explain, each ground given by Mr. Bradner for this conclusion is based on assumptions about the NSA's surveillance priorities, or is unexplained altogether, and provides no technological basis for concluding that the NSA would not (much less could not) utilize this filtering technique.

32. Principally, Mr. Bradner remarks that blocking particular types of communications by port or protocol number would leave "blind spot[s] in the NSA's Upstream surveillance that "[s]ophisticated targets" could "easily probe" to discover and exploit to avoid collection of their communications. Bradner Decl. ¶ 366(b), (e). Mr. Bradner does not explain what targets could "probe," or how, to discover these so-called blind spots. The traffic-mirroring techniques I describe in my earlier declaration are completely invisible to users' end systems (that is, their communicating devices) and to other equipment on the network not directly engaged in handling the monitored traffic. For example, traffic mirroring does not increase the delay or reduce the data flows being mirrored, or change the content or headers of the packets being transmitted in any way. Even if the undefined "probing" Mr. Bradner alludes to were possible, he does not explain what level of technical sophistication would be required, or on what basis he assumes that the NSA's targets possess that level of sophistication.

33. Because port numbers are, in Mr. Bradner's words, "only advisory," Bradner Decl. ¶¶ 109, 366(e), he suggests that if potential NSA targets somehow found out that the NSA was (hypothetically) blacklisting ports 80 and/or 443 (in order to block HTTP and/or HTTPS communications from its collection devices), then these potential targets could assign port 80 or

443 to all their communications (whether in fact they are HTTP or HTTPS communications or not) and thereby avoid detection. Bradner Decl. ¶ 366(b), (e). While port numbers cannot be dismissed as merely "advisory,"[6] it is technically feasible, with certain pre-arrangements, for two or more users to communicate in the manner that Mr. Bradner describes. At bottom, however, whether the creation of "blind spots" is a matter of such genuine intelligence concern as to motivate the NSA to examine all HTTP and HTTPS communications (even if it were not otherwise persuaded of the value in doing so), depends on facts and information concerning its mission priorities and resources known only to the NSA (or at least not presented in Mr. Bradner's declaration).[7]

34. Apart from the "blind spot" issue, Mr. Bradner also remarks that blacklisting HTTP and HTTPS communications (ports 80 and 443) "would leave a very large hole in the NSA's collection ability," including web-based e-mail, webchat, and web-based editors, and that there are many "obvious" reasons, in his view, for the NSA to acquire HTTPS communications. Bradner Decl. ¶ 366(f), (g); *see also* Bradner Decl. ¶¶ 326, 359. In so contending, Mr. Bradner is apparently making assumptions about the value that the NSA places on particular types of communications rather than offering a technological reason why the NSA could not or would not block access to such communications.

[6] Port numbers are assigned to specific applications, such as e-mail or web browsing, by the global Internet Assigned Numbers Authority (IANA), *see* First Decl. ¶¶ 16, 36, and represent established international conventions—default settings, in a manner of speaking—to facilitate automated, "user-friendly" communication between devices connected to the Internet. While use of an assigned port to run an application is neither legally nor even technically mandated, using non-standard ports, as I have explained, requires that users on both ends of an exchange must agree in advance to communicate in this atypical fashion, so that appropriate adjustments can be made to their communications before they are transmitted, to ensure that they are routed on receipt to the agreed-upon, non-standard port.

[7] Mr. Bradner observes that if the NSA blacklisted only HTTPS communications, but not HTTP communications, then it could still obtain access to unblocked HTTP communications to and from Wikimedia servers. Bradner Decl. ¶ 366(h). While true in theory, I have not posited a scenario in which the NSA blocks only HTTPS, but not HTTP communications. I have observed only that it is technically feasible to block both, First Decl. ¶¶ 80-81, a point with which Mr. Bradner does not disagree.

35. Moreover, I observe that if the NSA were interested in communications to and from particular websites, such as webmail sites, or chatroom sites, then it would not be necessary for the NSA to obtain access to all HTTP or HTTPS communications traversing a monitored link in order to do so. It is technically feasible, using a combination of blacklisting and whitelisting, to provide the NSA with access only to communications with websites of particular interest. Specifically, at a monitored link the provider's router or switch could be configured with a blacklist that would block NSA access to all communications with port numbers 80 or 443 (i.e., all HTTP and HTTPS communications), except those HTTP and HTTPS communications to or from the IP addresses included on a whitelist containing the addresses of the sites of interest to the NSA (including, hypothetically, specific webmail and chatroom sites). In this fashion, blacklisting HTTP and HTTPS communications (including Wikimedia's) would not necessarily, at least as a technological matter, carve out so large a "hole" in the NSA's Upstream collection as Mr. Bradner assumes.

36. The additional reasons given by Mr. Bradner for concluding that the NSA is "likely" acquiring HTTPS communications include (a) that the NSA is authorized to collect encrypted communications under the "minimization procedures" that govern its surveillance activities under Section 702 of the Foreign Intelligence Surveillance Act ("FISA"), Bradner Decl. ¶¶ 325, 366(g); (b) that the NSA has acknowledged collecting "web activity" under Section 702 of FISA, Bradner Decl. ¶¶ 314-15, 366(f) & n.126; (c) that the NSA "may, currently or in the future, be able to decrypt important encrypted messages," Bradner Decl. ¶ 326(a), (b); and (d) that the NSA could learn "useful information," such as the IP addresses of potential targets and the websites they visit, from the unencrypted addressing information of encrypted HTTPS communications, Bradner Decl. ¶¶ 326(c), 366(g). None of these reasons is a technological basis on which to assess the likelihood that the NSA does or does not acquire HTTPS communications, and, even on their own terms, they do not support Mr. Bradner's conclusion.

a. Authority under Section 702 to collect encrypted communications: There are a variety of encrypted Internet applications and communications, using a variety of

encryption techniques, apart from HTTPS, and it is my understanding that the NSA conducts at least two forms of collection under authority of FISA Section 702 – PRISM and Upstream. Therefore, to say that the NSA is authorized generally to collect encrypted communications under Section 702, Bradner Decl. ¶¶ 325, 366(g), is not to say specifically that it acquires HTTPS communications using Upstream collection.

b. Collection of "web activity": The reference to "web activity" in the NSA court filing cited by Mr. Bradner, Bradner Decl. ¶¶ 314-15, 366(f), App'x C at 30, appears in a passage comparing the total number of so-called multi-communication transactions (MCTs) acquired through Upstream to "the *total* take of Section 702 upstream collection of web activity" (emphasis mine). This juxtaposition, together with the reference to "total" take, suggests that the term "web" activity may have been intended here to refer to Internet activity as a whole, in light of the fact that MCTs are not necessarily limited to HTTP and HTTPS communications but could include e-mail (SMTP) communications. Moreover, it is common in colloquial usage to use the term "web" when referring to the Internet at large.[8] Even assuming that "web activity" as used in the cited NSA filing refers more precisely to HTTP and HTTPS communications, there are various forms of web activity, such as webmail, and chatrooms, that do not include communications with what we ordinarily think of as websites such as Wikimedia's, and that could be obtained using the combined blacklisting/whitelisting technique I discuss in paragraph 35, above.

c. Possibility of decryption: While the NSA "may" be able, now or in the future, to decipher encrypted communications, Bradner Decl. ¶ 326(a), (b), equally so, it may not, and without information about the NSA's true decryption capabilities, I do not see how the abstract possibility makes it "likely" as a technological matter that the NSA, in fact, collects encrypted communications at all, much less HTTPS communications specifically.

[8] *See* https://en.wikipedia.org/wiki/Internet#World_Wide_Web (observing that "[m]any people use, erroneously, the terms Internet and World Wide Web, or just the Web, interchangeably . . .").

d. Potential intelligence value of addressing information: Mr. Bradner's suggestion that the NSA might find intelligence value in the addressing information of HTTPS communications that would merit their collection even though their contents are encrypted, Bradner Decl. ¶¶ 326(c), 366(g), is simply another assumption on his part about the NSA's surveillance priorities that provides no technological support for his conclusion.

37. However, even taking as given Mr. Bradner's assumption that the NSA, for the reasons above, may be motivated to collect certain HTTPS communications, there are technical means by which it could obtain access to HTTPS communications of interest without copying and scanning all HTTPS communications at a monitored link, including Wikimedia's. As I explained in paragraph 35, above, it is technically feasible, to blacklist ports 80 and 443 (i.e., HTTP and HTTPS communications), while simultaneously whitelisting the IP addresses of websites, webmail services, and/or chatrooms of interest. In this fashion the NSA could obtain access to communications to and from websites of interest while excluding all others, including, hypothetically, Wikimedia's. A configuration of this kind would be entirely consistent, moreover, with the references to acquiring encrypted communications in the NSA's minimization procedures, and to collection of "web activity" in the NSA court filing cited by Mr. Bradner.

38. In short, Mr. Bradner gives no technological reason, as opposed to conjecture about the NSA's practices, priorities, and capabilities, for dismissing port blocking as "implausible."

Blacklisting Wikimedia's IP Addresses

39. Next, Mr. Bradner finds it "basically inconceivable" that the NSA would blacklist Wikimedia's IP addresses, for two reasons. Bradner Decl. ¶ 367(a). Neither is based in Internet technology and engineering.

40. Mr. Bradner states that it is "totally unbelievable" that the NSA would have undertaken the "incredibly resource-intensive task" of sifting through millions of websites to decide which to monitor and which not. Bradner Decl. ¶ 367(a). This is another non-sequitur, as

I did not suggest anything of the kind. What I discussed in my first declaration was the practical possibility of blacklisting certain high-volume but perhaps low-interest websites, such as, hypothetically, Amazon.com, Wikimedia's websites, and perhaps others, First Decl. ¶ 81, in order to reduce unwanted volumes of communications and enhance the efficiency of the collection process. This would be a trivial task.

41. Identifying and removing high-volume, commercial websites and video services is made much easier since page views follow a Zipf distribution, that is, a website's volume of page views declines rapidly as its popularity ranking decreases. For example, according to the Parse.ly blog post,[9] the top roughly 1,000 websites account for 97% of global page views. Thus, while the identity of popular websites shifts over time, these can be tracked relatively easily by widely-known services like Alexa.com, https://www.alexa.com/topsites.[10] According to Alexa, as of early 2019 Wikipedia.org is ranked as the fifth most popular website globally (sixth in the United States), so even a manual process would likely include it on any filtering blacklist (assuming the NSA deemed it of low interest). If desired, the list of these popular sites could be obtained periodically and mechanically, converted to IP addresses by domain name lookups programmatically, and then be used to modify the filter list used in routers.[11] When Mr. Bradner disparages the idea of reviewing millions of websites to decide one by one which sites to monitor and which to ignore, he is taking issue with a proposal that I have not made.

42. Second, Mr. Bradner states that he finds blacklisting Wikimedia IP addresses to be implausible because doing so "would deliberately limit the possible collection of information on the use of Wikimedia resources by NSA targets, a potentially valuable source of information

[9] https://blog.parse.ly/post/10/zipfs-law-of-the-internet-explaining-online-behavior/

[10] Alexa.com tracks the rankings of the popular websites; while also operated by Amazon, it is not related to Amazon's smart-speaker service with the same name.

[11] As Mr. Bradner notes, the NSA, at least in some circumstances, uses IP address filtering to "eliminate potential domestic transactions" from those scanned. PCLOB Section 702 Report at 37; Bradner Decl. ¶ 290. As he also notes, the number of domestic U.S. IP addresses exceeds 60,000. Bradner Decl. ¶ 229. If the NSA can maintain a list of over 60,000 address blocks, "no easy task," Bradner Decl. ¶ 298, then it can also add a few dozen or a hundred IP addresses of organizations that contribute high volumes of traffic, but are unlikely to be of intelligence interest, such as (hypothetically) Wikimedia.

about the online research and reading of its targets." Bradner Decl. ¶ 367(a). Once again, Mr. Bradner is engaging in speculation, without an evident basis in knowledge or information, about the online reading habits of the NSA's targets and the intelligence value that the NSA would ascribe to knowing whether they read Wikimedia websites. That is not a technological basis on which to judge the plausibility of blacklisting Wikimedia websites.

Whitelisting IP Addresses of Interest

43. As I have explained, First Decl. ¶ 65, whitelisting is a filtering technique whereby an assisting telecommunications carrier at an Internet backbone link hypothetically monitored by the NSA could provide the NSA only with copies of communications packets whose source or destination IP addresses appear on a list of particular IP addresses, or blocks of IP addresses, that the NSA has determined are associated with communications (or targets) of interest. Using this technique, the NSA would not obtain access to any of Wikimedia's communications unless users communicating with Wikimedia's websites, or perhaps with its staff, were assigned IP addresses on the targeted whitelist. First Decl. ¶¶ 81, 84, 87.

44. Mr. Bradner suggests no reason why whitelisting, if employed, would not block NSA access to Wikimedia's communications (unless, hypothetically, Wikimedia were receiving or responding to communications from a whitelisted IP address). Rather, he gives a single reason why he finds it implausible that the NSA would employ whitelists in the course of Upstream surveillance. Bradner Decl. ¶ 366(d). Separately, Wikimedia, in its legal brief, suggests that whitelisting would be inconsistent with certain features of Upstream collection. Wikimedia Brief at 22. I address both points in turn.

45. Mr. Bradner opines that whitelists would be "useless" for Upstream collection purposes because it is not "remotely possible" for the NSA to know in advance the IP addresses of all its targets. Bradner Decl. ¶ 366(d). He premises this conclusion, however, on various assumptions about the number, nature, and mobility of the NSA's Upstream surveillance targets, for which he offers no supporting information.

46. First, Mr. Bradner appears to assume that the NSA has over 129,000 Upstream targets, based on a Government report disclosing that in 2017 the NSA had over 129,000 Section 702 targets. As discussed above, however, and as Mr. Bradner also observes, the NSA operates two collection programs under Section 702 of FISA, Upstream and PRISM. *See* Bradner Decl. ¶ 334 & n.108. So far as I am aware, the Government has not disclosed any information concerning the number of Section 702 targets that are targets, specifically, of Upstream collection, as opposed to those that may be targets exclusively of PRISM collection. Mr. Bradner presents no basis on which to assume that the NSA's Upstream targets are too numerous to make whitelisting technically practical.

47. Second, Mr. Bradner simply assumes both (i) that the NSA's Upstream targets include individuals (as opposed to other more stationary types of entities), and (ii) that their movements result in changes to the IP addresses associated with their communications. Bradner Decl. ¶¶ 334, 366(d). Even if the NSA's Upstream targets include individuals who move from place to place, IP addresses, as Mr. Bradner acknowledges, are often associated with specific geographic areas to a "reasonable degree of certainty." Bradner Decl. ¶ 296; *see also* First Decl. ¶¶ 32-34. Therefore, even if an individual NSA target moved to a certain degree from place to place within a given geographic area, in principle the NSA could reliably obtain access to that target's communications by targeting not just a single IP address, but a set of IP addresses, as I posited in my first declaration, First Decl. ¶ 67, associated with geographic areas where the target is believed to be located.

48. In the final analysis, the NSA's capabilities to ascertain and track the IP addresses of communications with its Upstream targets are unknown, and a blanket assumption that it lacks the ability to do so is not a technological basis on which to dismiss whitelisting as "useless."

49. Wikimedia separately argues in its legal brief that the idea of whitelisting "ignores" the NSA's collection of "about" communications that are neither to nor from its targets, and the fact that NSA selectors (such as e-mail addresses) do not appear in the packet headers where

network-layer addressing information, such as IP addresses, is located. Wikimedia Brief at 21-22. Notably, this is not a point echoed by Mr. Bradner, and it is not well taken.

50. Wikimedia appears to be conflating two separate steps in the filter-then-copy-and-scan collection process I have described. First is the filtering of communications within the provider's router or switch as they traverse a monitored link. In the whitelisting context, this step involves a comparison of packets' source and destination IP addresses (contained in the packet header) to a whitelist of specified IP addresses, so that only packets containing source or destination IP addresses on the whitelist are copied and made available for scanning. The second step occurs after the whitelisted packets are copied and passed through the router interface to be scanned by the collection system. This is the point at which the packets' "payload," including, for example, source and destination e-mail addresses as well as their message content, is scanned for targeted selectors, to identify those communications that will actually be retained.

51. This two-step process is entirely compatible with the acquisition of "about" communications as described in public sources, and with scanning for selectors other than IP addresses. If Wikimedia means to suggest otherwise, then it is mistaken. If the NSA, through whitelisting, were to obtain access to communications to and from IP addresses of interest, it could then scan them for the presence of targeted selectors of any (authorized) kind, be they e-mail addresses, telephone numbers, or (hypothetically) other communications identifiers. And if the NSA obtained access to communications to or from a specified (whitelisted) set of IP addresses it has associated with a target (or a target's geographic location), and scanned them for the target's e-mail address, then it could acquire not only communications to or from the target's e-mail address, but also "about" communications between parties other than the target, also originating from or destined for one of the whitelisted IP addresses, that contain the target's e-mail address in their contents.

52. In sum, neither Mr. Bradner, nor Wikimedia, has identified any grounds based in Internet technology and engineering that would render whitelisting "useless," or otherwise impractical or "implausible" for purposes of NSA Upstream collection. They have provided no

reason to retreat from my conclusion that at any Internet backbone link that the NSA might hypothetically be monitoring, the NSA, through whitelisting, could block access to any Wikimedia communications, except those (if any) in which users communicating with Wikimedia, had been assigned a targeted IP address. First Decl. ¶¶ 81, 84, 87.

MISCELLANEOUS POINTS RAISED BY MR. BRADNER AND WIKIMEDIA

53. In the foregoing sections I have addressed (i) the grounds on which Mr. Bradner concludes it is "more likely" that the NSA employs his copy-all-then-scan approach to Upstream collection, rather than a filter-then-copy-and-scan approach, and (ii) the grounds on which Mr. Bradner (and, separately, Wikimedia) argue that a filter-then-copy-and-scan approach using whitelisting and/or blacklisting techniques is "implausible." In this section I turn to several more general observations made by Mr. Bradner, and Wikimedia.

54. Asymmetric routing of Internet communications: Mr. Bradner remarks that my analysis "does not discuss the asymmetric routing of communications on the Internet." Bradner Decl. ¶ 359. To the contrary, as Mr. Bradner and I both have observed, communications exchanged between two users on the Internet (such as e-mail), or communications between a user and a website, do not necessarily follow the same path back and forth, even in real time. Bradner Decl. ¶¶ 104(b), 199, 309; First Decl. ¶ 89. I believe that Mr. Bradner overstates the impact of asymmetric routing between two end points at international gateways. (The number of international gateways between any two countries is relatively small.) But the point is that asymmetric routing has nothing to do with the feasibility or utility of whitelisting and/or blacklisting as (hypothetical) Upstream collection techniques that could block NSA access to Wikimedia communications.

55. As Mr. Bradner himself observes, the upshot of asymmetric routing, when it occurs, is that the response to a communication may cross a different international link than the one crossed by the original communication, and that the NSA, if interested in acquiring both the original communication and the response, would have to monitor both links. Bradner Decl. ¶ 309. This has no impact, however, on the scope of the communications the NSA must monitor

at each link, regardless of the number of links actually monitored. For example, if the source IP address of the original communication were included on a whitelist used at the first link, then, if the response returned over a different link, it could still be captured at that second link, using the same whitelist, since the source IP address of the original communication would now appear as the destination IP address of the response. Thus, it would not ordinarily be necessary for the NSA to copy and scan all communications, at either link, in order to capture both the original communication and the response.

56. Acquisition of wholly domestic "about" communications: Mr. Bradner suggests that the NSA must not be using IP address filtering to eliminate wholly domestic communications before copying and scanning, at least at some monitored links. Bradner Decl. ¶ 293. He bases this conclusion on a statement, in an October 2011 opinion by the Foreign Intelligence Surveillance Court ("FISC"), that the NSA has acknowledged that it "will acquire a wholly domestic 'about' communication" if it is "routed through an international internet link being monitored by the NSA." Bradner Decl. ¶¶ 292-94 & n.83, App'x P. Whether or not that is so, it would be technologically inaccurate to conclude on this basis (as Wikimedia appears to do, Wikimedia Brief at 21-22), that the acquisition of some wholly domestic communications, even at a so-called "international Internet link," is inconsistent with the use of the whitelisting and blacklisting techniques I have described.

57. As Mr. Bradner observes, the "routing of wholly domestic communications over international circuits does occasionally happen." Bradner Decl. ¶ 292 & n.82. There are several scenarios in which this could occur. For example, suppose that a person located in the United States, who uses a foreign-based virtual private network (VPN) service, sends an e-mail to another person, also located in the United States. When a user communicates via a VPN, all of the user's communications are encrypted and first routed through the VPN server before being directed to their ultimate destination. As a result, on that first leg each communications packet

is assigned the VPN server's address as its destination IP address.[12] In the example above, the user's e-mail will cross an international link on its way to the foreign-based VPN server, and again on its way from the VPN server to the U.S. recipient. Similarly, if persons located in the United States use foreign-based servers (perhaps belonging to an overseas corporation, or university) to send and receive e-mail from other persons in the United States, those communications will cross international links as they are routed to and from the foreign servers.

58. NSA acquisition of such communications, assuming it occurs as the FISC described, would be entirely consistent with white- or blacklisting by IP address. In the above examples, on the first leg from the sender to the foreign VPN or e-mail server, the destination IP address of the communications will be the foreign IP address of the server; and on the second leg to the recipient the source IP address of the communications will also be the foreign IP address of the server. Therefore, if the NSA were whitelisting communications to and from certain IP addresses, communications of the kind described above could still be copied and scanned by the NSA (and acquired if they contain targeted selectors), if the IP addresses of the foreign VPN and e-mail servers were included on the whitelist. Equally so, if the NSA were blacklisting communications to and from certain IP addresses, wholly domestic communications of the kinds described above could still be copied and scanned by the NSA (and acquired if they contain targeted selectors), if the IP addresses of the foreign VPN and e-mail servers were not on the NSA's blacklist.

59. U.K. Section 8(4) collection: Mr. Bradner states that his conclusion that the NSA likely follows his copy-all-then-scan approach is "reinforced" by public filings of the U.K. Government in the European Court of Human Rights (ECHR) concerning its "Section 8(4)" collection program conducted by the U.K.'s Government Communications Headquarters (GCHQ). Bradner Decl. ¶¶ 368-69. I see little support, if any, for Mr. Bradner's conclusions in these non-

[12] Once the packets reach the VPN server, the VPN layer of encryption is removed, and the packets are forwarded to their intended destination, but for security are assigned the VPN server's address as their source IP address. The process is reversed for any response to the user's initial communication.

technical documents, which contain information, not discussed in his declaration, that tends to refute, not support, his views.

60. The mere fact that the U.K. Government conducts Section 8(4) surveillance using one possible configuration does not mean that the NSA conducts Upstream surveillance the same way, as opposed to other possible approaches that could be followed. Beyond that observation, the description of Section 8(4) collection that Mr. Bradner relies on is contained in publicly filed legal briefs and a court opinion, not internal technical or operational manuals or schematics detailing the design and operation of Section 8(4) collection systems. *See* Bradner Decl. App'x DD, EE. These sources provide only the roughest outline of the Section 8(4) collection process, and it is difficult, therefore, to draw detailed technical conclusions about the Section 8(4) process or how it compares to Upstream collection based on these sources.

61. To the extent that documents at such a high level of generality can be relied upon for the purpose of drawing conclusions about how Section 8(4) collection (or Upstream) operates, I note that they actually describe a collection approach quite comparable (at least at a general level) to the type of IP address and port and protocol number filtering described in my earlier declaration. In a passage not cited in Mr. Bradner's declaration, the U.K. Government's brief before the ECHR describes Section 8(4) collection as follows:

> First stage: collection
>
> GCHQ selects which bearers [circuits] to access based on an assessment of the likely intelligence value of the communications they are carrying. . . .
>
> Second stage: filtering
>
> GCHQ's processing systems operate on the bearers which it has chosen to access. A degree of filtering is then applied to the traffic on these bearers, designed to select communications of potential intelligence value. As a result of this filtering stage, the processing systems automatically discard a significant proportion of the communications on the targeted bearers.
>
> Third stage: selection for examination
>
> The remaining communications are then subjected to the application of queries, both simple and complex, to draw out communications of intelligence value. Examples of a simple query are searches against a "strong selector" such as a telephone number or email address. Complex queries combine a number of criteria, which may include weaker selectors but which in combination aim to reduce the odds of a false positive. Communications that do not match the chosen

> criteria are automatically discarded. The retained communications are available to analysts for possible examination.

Bradner Decl., App'x EE at 4-5.

62. To summarize, according to the above description of Section 8(4) collection in the U.K. Government's brief, before communications are "querie[d]" for the presence of "selectors" (at the third stage), the Section 8(4) processing systems apply "filtering" (at the second stage) to winnow communications deemed to lack intelligence value and to pass on to stage three only those communications considered to be "of potential intelligence value." The exact type of filtering performed at the second stage is not disclosed, but the general description of the Section 8(4) process contained in these publicly available documents is consistent with application of the filter-then-copy-and-scan techniques I have described.

63. Mr. Bradner focuses attention on the U.K. Government's explanation that in order to conduct these filtering and querying processes it is necessary as a "practical" matter, for "technical reasons . . . to intercept the entire contents of a [circuit]." Bradner Decl. ¶ 368 (citing App'x EE ¶¶ 7-8, at 2-3). These "technical reasons" are not described, so there is no way to know whether they would constrain the NSA's ability to configure its collection systems.

64. Assuming as does Mr. Bradner that the "intercept[ion]" referred to in the U.K. brief involves duplication of the entire communications stream before communications are filtered, it does not necessarily follow that the U.K. Government or the NSA must be given access to copies of all communications traveling across a monitored link. If a provider does not prefer to use its network router or switch to perform the IP address or port or protocol number filtering, then it would also be technically feasible for a provider to use an optical splitter, as both Mr. Bradner and I have discussed, to duplicate the communications stream and divert the copied stream for off-line processing while the communications in the "original" stream continue toward their intended destinations. *See* First Decl. ¶ 55; Bradner Decl. ¶¶ 275-76. The provider then could apply IP address or port or protocol number filtering (whether whitelisting or blacklisting)

to the copied stream using a router or standard firewall "appliance"[13] to make available to GCHQ (or the NSA) only those of the copied communications meeting the filter criteria, while automatically destroying the rest. Either configuration would be consistent with what is said in the U.K. Government's brief concerning the "interception" of all communications on a circuit, and neither would involve, much less require, passing all the communications on a monitored circuit to the GCHQ's or the NSA's possession and control.

65. In short, the documents cited by Mr. Bradner offer no basis for concluding (as opposed to speculating) that in conducting Upstream collection the NSA copies and scans all communications, or even that it acquires copies (which it then scans) of all communications, that cross a monitored Internet backbone link.

66. The EINSTEIN 2.0 System: Wikimedia also states in its brief that Mr. Bradner's conclusions are "corroborated" by the U.S. Government's cyber-defense system known as EINSTEIN 2.0. Wikimedia Brief at 20. In contrast, while Mr. Bradner refers to the EINSTEIN 2.0 system as an example of a deep-packet-inspection system, he does not cite EINSTEIN 2.0 as corroboration for his conclusions. Bradner Decl. ¶ 259.

67. EINSTEIN 2.0 is not, as Wikimedia describes it, a "surveillance program" like Upstream collection. Rather, as described in the Department of Justice Office of Legal Counsel ("OLC") memorandum relied on by Wikimedia, Plaintiff's Exhibit 25, EINSTEIN 2.0 is a cyber-intrusion detection system meant to protect the unclassified information technology systems of civilian U.S. Government agencies against malware and other network-based attacks. It is not meant for intelligence gathering, except possibly tracing the progression of detected cyber-attacks.

68. To perform its function, EINSTEIN 2.0 scans incoming Internet traffic as it reaches the access points connecting these Federal Government systems to the Internet. Plaintiff's Exhibit 25 at 3. As described in the OLC memorandum, "EINSTEIN 2.0 sensors [do] not scan actual

[13] The Barracuda CloudGen firewall is one commercial example, with throughput of up to 46 Gb/s.

Federal Systems Internet Traffic for malicious computer code as that traffic is in transmission, but instead will scan a temporary copy of that traffic created solely for the purpose of scanning by the sensors," while "[t]he 'original' Federal Systems Internet Traffic will continue to its destination without being scanned." Plaintiff's Exhibit 25 at 4. There are at least two reasons why conclusions about the Upstream collection process cannot be drawn from this statement about the configuration of the EINSTEIN 2.0 system.

69. First, because cyber attacks can use any protocol, originate from any external Internet host, and can target any destination system, to be effective an intrusion-detection system must inspect all incoming traffic. For the reasons I have discussed above, and in my first declaration, the NSA could reliably obtain access to its targets' communications crossing a monitored link without copying and scanning all communications that cross that link. Second, it is unlikely that the volume of incoming Internet traffic at any given civilian Government agency exceeds 10 gigabits per second.[14] In comparison, the potential volume of traffic at the AEConnect link, discussed above, 40 terabits per second, is over 4,000 times greater. The relatively small volume of traffic that EINSTEIN 2.0 can be expected to support is well within the capability envelope of a single commercial intrusion detection system or DPI system, meaning that there are likely no processing or capacity constraints in the EINSTEIN 2.0 system that would necessitate filtering out communications to reduce the volume that needs to be scanned for malicious code.

70. Because the purpose and required processing capacity of EINSTEIN 2.0 differ fundamentally from those of Upstream collection, the architecture and operation of EINSTEIN 2.0 are unlikely to provide insight into the operational practices of the Upstream program.

71. Comprehensiveness: Finally, I address Mr. Bradner's and Wikimedia's attempts to draw support from what Mr. Bradner describes as "the NSA's stated desire to be comprehensive

[14] The GSA Enterprise Infrastructure Solutions (EIS) guide supports this conclusion, as the highest available "[d]edicated burstable Internet bandwidth" is 10 Gb/s, as CLIN 22006. (https://eis-public-pricer.nhc.noblis.org/ajax.php/resources/download?type=csv&file=clins) Similarly, the MTIPS (Managed Trusted Internet Protocol Services) item that incorporates EINSTEIN functionality tops out at 10 Gb/s (same file, CLIN MT00060). See also https://www.gsa.gov/technology/technology-products-services/it-security/trusted-internet-connections-tics.

in its [Upstream] collection." Bradner Decl. ¶¶ 228, 333, 359 (citing PCLOB Section 702 Report at 10, 123, 143); Wikimedia Brief at 21. Mr. Bradner infers that "if the NSA's goal is to comprehensively obtain its targets' communications, then it must comprehensively copy, reassemble and review all transactions that could conceivably be to or from a target that transit the circuits being monitored." Bradner Decl. ¶ 335. There are numerous reasons why this conclusion does not follow.

72. The "repeated[ ]" statements by the "[G]overnment" that Mr. Bradner refers to, Bradner Decl. ¶ 333, are actually a single statement by the PCLOB that appears twice in the PCLOB's Section 702 Report. *See, e.g.,* Bradner Decl. ¶ 333 & n.106 (citing PCLOB Section 702 Report at 10, 123). In the cited statement, the PCLOB characterizes "the NSA's acquisition of 'about' communications" as "an inevitable byproduct of the government's efforts to comprehensively acquire communications that are sent to or from its targets." PCLOB Section 702 Report at 10; *see id.* at 123 (same).

73. We cannot simply indulge an assumption that the NSA "comprehensively" acquires the communications of its targets based on the slim reed relied on by Mr. Bradner. I agree with Mr. Bradner that it would be "unsurprising" to discover that the NSA, in a perfect world, would prefer to obtain access to all of its foreign-intelligence targets' communications. It does not follow, however, that the NSA is in fact doing so. It is one thing to state these goals, and quite another to design, construct, deploy, maintain, and pay for the collection systems required, in the numbers and with the capacity needed, to attain such ambitious goals. We cannot assume on the basis of a stated goal alone that the NSA has achieved that desired result without assuming away the technical, logistical, and financial hurdles, the resource constraints and trade-offs, and the competing mission priorities, that would stand in the way. Even if the technical and logistical hurdles could be overcome, we must recognize the possibility that at some point the cost of doing so may, in the NSA's view, outweigh the marginal benefit of potentially discovering still further communications of its targets in some as-yet unexplored stream of communications on the Internet. In short, we must recognize that in the field of large-

scale digital communications engineering as in other practical endeavors, idealized goals must be tempered by gritty reality.

74. Even taking for granted, for a moment, Mr. Bradner's assumption that the NSA has achieved the goal of comprehensively acquiring its targets' online communications, it still would not follow that the NSA "must comprehensively copy, reassemble and review all [communications] . . . transit[ing] the circuits being monitored." Bradner Decl. ¶ 335. I explained in my first declaration how the NSA, using traffic-mirroring techniques such as white- and blacklisting, could reliably obtain access to its Upstream targets' communications without copying and scanning all of the communications traversing a monitored link. And, as I explain herein, Mr. Bradner identifies no technical reason to question that conclusion, instead relying on speculation about the nature and habits of the NSA's targets, the NSA's intelligence priorities, its resources, and capabilities, to support his opinions to the contrary.[15] Moreover, it must be acknowledged that the term comprehensive is a qualitative one, susceptible of a number of meanings other than "exhaustive." Without detailed information, at the least, concerning the types and quantities of communications accessed by the NSA, it is not possible to reverse-engineer detailed conclusions about the methodologies employed and equipment configurations deployed by the NSA from so limited and technically unenlightening a starting point as the supposed comprehensiveness of the agency's objectives.

75. To reiterate, I do not opine on the likelihood that the NSA, in the course of conducting Upstream surveillance, actually may use the traffic-mirroring techniques I have discussed. But it is my opinion that, at bottom, the reasons given by Mr. Bradner for deeming

[15] Mr. Bradner also relies on the PCLOB's characterization of Upstream's objective to conclude that the NSA "is very likely to be monitoring a large number of international circuits, given that it would need to monitor most, if not all, such circuits to accomplish its stated (and unsurprising) goal of reliably and comprehensively collecting the communications of its targets." Bradner Decl. ¶ 353. As I have stated, one cannot simply take for granted that the PCLOB's 2014 description of the NSA's supposed goals reflects the reality of what the NSA has actually accomplished. Be that as it may, I have also already explained that the feasibility of whitelisting and blacklisting does not depend on the number of sites on the Internet that the NSA actually monitors, whether the number is one or many. First Decl. ¶¶ 90-91. Mr. Bradner does not contest this point.

that possibility "implausible," and his alternative approach "most likely," are without basis in Internet technology and engineering.

MR. BRADNER'S CERTAINTY THAT THE NSA HAS COPIED AND SCANNED AT LEAST SOME WIKIMEDIA COMMUNICATIONS, EVEN IF THE NSA EMPLOYS TRAFFIC-MIRRORING TECHNIQUES, IS WITHOUT TECHNICAL OR EMPIRICAL BASIS.

76. Although Mr. Bradner acknowledges that the various types of traffic-mirroring techniques I discuss are technically feasible, he nevertheless maintains that, even if the NSA were using these techniques to filter out Wikimedia communications from those made available to it at a monitored link (whether by design, or effect), it is still "virtually certain" that the NSA, in the course of Upstream collection, "has copied, reassembled, and reviewed at least some of Wikimedia's communications." Bradner Decl. ¶ 370. He does not explain, however, why he believes this to be the case. Although he raises a number of reasons why he believes the NSA likely would not employ these techniques (which I addressed in the preceding sections), he raises no technical objections to the *efficacy* of white- or blacklisting port or protocol numbers, or whitelisting by IP address, as means of blocking access to Wikimedia communications.

77. Rather, Mr. Bradner asserts only that it is "technologically incorrect that blocking [i.e., blacklisting] Wikimedia IP addresses would block all Wikimedia traffic." Bradner Decl. ¶ 367(b). So far as I can discern from Mr. Bradner's declaration, this is the sole "technical[ ] inaccura[cy]," Bradner Decl. ¶ 7(b), that Mr. Bradner ascribes to my analysis. Specifically, Mr. Bradner posits three scenarios in which Wikimedia's communications "would still be copied, reassembled and reviewed by the NSA," even if the NSA blacklisted communications containing Wikimedia IP addresses. Bradner Decl. ¶ 367(b). Each of these scenarios, while theoretically possible, could come to pass only in the uncertain event that particular conditions are met, as I discuss in turn.

78. Scenario One: MCT Containing a Wikimedia Communication: Mr. Bradner observes that if a Wikimedia communication were contained within a so-called "multi-communication transaction" (MCT) that itself was neither to nor from Wikimedia, then blacklisting communications with source or destination IP addresses assigned to Wikimedia

would not prevent the MCT, including the embedded Wikimedia communication, from being copied and scanned at a monitored link. Bradner Decl. ¶ 367(b)(1). (This would be so, because the IP addresses included in the individual packet headers would be the source and destination IP addresses of the enclosing MCT, rather than the IP addresses of the embedded Wikimedia communication.) Mr. Bradner gives, as a hypothetical example of an MCT, the group of e-mail messages that are transmitted together as a single communication from an e-mail service to a subscribing user's Inbox, when the user logs in to check his or her e-mail. Bradner Decl. ¶¶ 132, 317.[16] This situation would result in copying and scanning communications to or from Wikimedia, however, only if several conditions were met:

a. First, Wikimedia maintains that its communications are copied and scanned as they traverse international links on the Internet backbone. For an MCT comprised of e-mails downloaded from a server to the Inbox on an individual user's cellphone or computer to cross an international Internet link, either a user in a foreign location must be downloading e-mails from a server located inside the United States, or a user located in the United States must be downloading e-mails from a foreign server.

b. Second, for the MCT containing these e-mails to include Wikimedia communications, the user must also be one who communicates by e-mail with Wikimedia.

c. Third, for the MCT to be copied and scanned by the NSA, it would have to traverse an international link that is monitored by the NSA.

d. Fourth, the MCT in which the Wikimedia communication is embedded must itself be a communication that has not been blacklisted.

79. Wikimedia claims to communicate with persons in almost every country on Earth, and that it is "virtually certain," therefore, that its communications traverse every circuit on every international cable carrying Internet traffic to and from the United States, *see* Bradner Decl. ¶ 6(d). But the scenario posited by Mr. Bradner is limited to those persons *meeting the first two criteria above.* Neither Wikimedia nor Mr. Bradner cites evidence concerning the number or geographic locations of persons meeting those criteria, if any, who communicate with Wikimedia, and I am aware of none. There is no basis, therefore, to conclude that MCTs enclosing Wikimedia communications almost certainly cross every international Internet link to and from the United

[16] The NSA has not publicly acknowledged what kind of MCTs are acquired during the Upstream collection process.

States, or, critically, that they cross one or more links that happen to be monitored (assuming any) by the NSA (the third criterion). It is a matter of conjecture, therefore, whether MCTs containing Wikimedia communications are copied and scanned by the NSA while crossing international Internet links.

80. The scenario is rendered even more conjectural when the possibility is considered that an MCT in which a Wikimedia communication is embedded might itself be a blacklisted communication (thus failing to meet the fourth criterion). U.S. consumer-based webmail services, such as a person located outside the United States might use, most commonly interface with their subscribers using encrypted HTTPS communications. (Gmail, the most popular webmail service provider, has been 100-percent encrypted since mid-2014.)[17] In addition, most businesses and other organizations strongly discourage or prevent use of unencrypted protocols to send or receive e-mail for business purposes. Thus, for example, employees of U.S.-based companies traveling or stationed abroad who rely on an e-mail server located at a U.S. corporate headquarters most likely will send and retrieve e-mail, respectively, through an encrypted SMTP submission transport connection, using TCP port 587, and an encrypted IMAP connection, using TCP port 993. Overall, the likelihood that a person located outside the United States would use unencrypted e-mail protocols to send or receive e-mail is exceedingly small. As I observed in my earlier declaration, if the NSA lacked the ability to decipher a particular kind (or kinds) of encrypted communications, then it could avoid copying and scanning them, if it wished, by blacklisting their associated port or protocol numbers. First Decl. ¶ 79. That would include encrypted MCTs that might hypothetically contain Wikimedia communications.

81. Scenario Two: E-mail to Wikimedia from Abroad Using a U.S.-Based Service: Mr. Bradner next posits that the NSA could copy and scan Wikimedia communications, even if it blacklisted Wikimedia's IP addresses, in a "case where a person located outside the U.S. is using an email service located inside the U.S. to send email to [and receive email from] Wikimedia." Bradner Decl. ¶ 367(b)(2). (This would be possible because while in transmission from the user

[17] https://transparencyreport.google.com/https/overview.

to the e-mail server, and vice versa, the communication would not include Wikimedia's IP address in the packet headers.) In practical terms, this scenario is simply a variant of the first, and could occur only if the following conditions, quite similar to those in the MCT scenario above, are met:

a. First, for the e-mail in question to cross an international Internet link without a Wikimedia IP address in the packet header, the user must be someone located outside the United States who is using an e-mail service (more precisely, an e-mail server) that is located inside the United States.

b. Second, the user must be sending e-mail to and/or receiving e-mail from Wikimedia.

c. Third, for the e-mail to be copied and scanned by the NSA, it would have to traverse an international link that is monitored by the NSA.

d. Fourth, the e-mail must itself be a communication that has not been blacklisted.

82. As in Mr. Bradner's MCT scenario, there is no basis on which to conclude that e-mail between Wikimedia and users who meet the first two criteria above (if any) almost certainly cross every international Internet link to and from the United States, or that they cross one or more links that happen to be monitored (assuming any) by the NSA (the third criterion). It is also a matter of conjecture, therefore, whether e-mail sent to Wikimedia, even from persons outside the United States who are using U.S.-based e-mail services, are copied and scanned by the NSA while crossing international Internet links. And, as in the MCT scenario, the matter becomes even more uncertain when one considers that any international e-mail communications of the kind posited by Mr. Bradner are quite likely encrypted with secure e-mail and/or HTTPS transmission protocols, raising the possibility that they are blacklisted (thus not meeting the fourth criterion), and so excluded from copying and scanning.

83. Scenario Three: Accessing Wikimedia Websites from Abroad Using a U.S.-Based VPN Service: Finally, Mr. Bradner posits that the NSA could copy and scan communications to and from a Wikimedia website, even if it blacklisted Wikimedia's IP addresses, if a user outside the United States accessed the Wikimedia site using a VPN (virtual private network) service located inside the United States. Bradner Decl. ¶ 367(b)(3). This is possible because, as discussed in paragraph 57, above, when a user communicates via a VPN, all of the user's communications

(including HTTP and HTTPS communications) are first routed through the VPN server, and assigned the server's address as the communications' destination IP address, rather than (as in this scenario) the target website. In reality, however, the scenario envisioned by Mr. Bradner could occur only if the following conditions are met:

a. First, for the HTTP and HTTPS communications in question to cross an international Internet link, the user must be someone located outside the United States.

b. Second, the user must be someone who has decided to use, and perhaps to pay a service fee for, the VPN service. (The user may also be using the communication facilities of an organization, such as an employer, that has decided to use a VPN for its business communications.)

c. Third, although the user is located outside the United States, the user (or the organization whose facilities he or she is using) must have chosen a VPN service based inside, not outside, the United States.

d. Fourth, to be copied and scanned by the NSA, the user's communications with Wikimedia's websites would have to traverse an international link that is monitored by the NSA.

e. Fifth, communications to the VPN server must not themselves be blacklisted.

84. Again, there is no basis on which to conclude that communications with Wikimedia websites from users who meet the first three criteria above (if any) almost certainly cross every international Internet link to and from the United States, or that they cross one or more links that happen to be monitored (assuming any) by the NSA (the fourth criterion). It is also a matter of conjecture, therefore, whether HTTP or HTTPS communications sent to Wikimedia websites, even those from persons outside the United States who are using U.S.-based VPN services, are copied and scanned by the NSA while crossing international Internet links. And it becomes even less certain considering that any such communications would be encrypted by the VPN service, raising the possibility, if the NSA lacked the ability to decipher them, that they are blacklisted (perhaps using the VPN's IP address), and thus, not meeting the fifth criterion, are excluded from copying and scanning.

85. In sum, it is far short of the certainty claimed by Mr. Bradner that any of the scenarios described by him would come to pass at a particular international Internet link that happened to be monitored by the NSA (if any), such that the NSA would copy and scan

communications of Wikimedia's even if it had blacklisted Wikimedia IP addresses. Each of these scenarios requires a confluence of multiple events before it could come to pass, the likelihood of which individually, and certainly collectively, is at best conjectural. And so, while I acknowledge that in theory blacklisting might not eliminate all possibility that the NSA, in conducting Upstream surveillance, obtains copies of and scans Wikimedia communications for targeted selectors, blacklisting would render that possibility a matter of speculation.

86. Moreover, it bears emphasis that even if there remained a possibility of copying and scanning Wikimedia communications despite blacklisting Wikimedia IP addresses, that would not be the case if the NSA were employing a whitelisting technique. In a whitelisting scenario, no communications to or from Wikimedia would be copied and made available for scanning by the NSA, unless a communication to or from Wikimedia itself had a source or destination IP address, respectively, on the target whitelist. Nor does Mr. Bradner suggest otherwise. Therefore, the hypothetical possibility that Wikimedia communications could be copied and reviewed in the limited and uncertain sets of circumstances suggested by Mr. Bradner, notwithstanding blacklisting, does not alter my conclusion that the NSA, through whitelisting, could conduct Upstream-type surveillance as envisioned by Wikimedia, without copying and reviewing or otherwise interacting with Wikimedia communications.

87. For these reasons, I find Mr. Bradner's assertion that it is "virtually certain" the NSA has copied and scanned at least some of Wikimedia's communications, even if the NSA employs one or more of the whitelisting and blacklisting techniques I have described, Bradner Decl. ¶ 370, to be without a basis in Internet technology and engineering that rises above the level of conjecture.

PRACTICAL REASONS FOR MAINTAINING "HTTPS BY DEFAULT" AND IPSec ENCRYPTION IN THE CURRENT COMMUNICATIONS ENVIRONMENT

88. Finally, I discuss (a) the significant reasons why, in the current digital communications environment, any organization that operates a major website would be powerfully motivated to protect communications to and from its site using the HTTPS protocol,

in particular HTTPS by default, and (b) the reasons why an organization that transmits sensitive proprietary or personal information on the Internet would be equally motivated to encrypt those communications using a security protocol such as IPsec or ssh.

89. As I explained in my earlier declaration, communications on the Internet may be encrypted to protect the privacy and integrity of the information they contain. The HTTPS protocol (technically, a combination of the Transport Layer Security (TLS) protocol and the HTTP protocol), is the most common encryption mechanism on the Internet, used to ensure that a user's browser connects to the correct web server (rather than an imposter site), and that the information sent between the user and the website can be read only by the user's web browser and the host web server, but not third parties. First Decl. ¶ 42.

90. Although selective encryption of web communications for online transactions deemed sensitive or confidential (such as online banking) began as long ago as 1994, in 2014 the Internet Engineering Task Force (IETF) (*see* First Decl. ¶ 26) described a range of motivations for using encryption pervasively in today's environment, including: (i) surveillance by nation-state actors; (ii) legal but "privacy-unfriendly" exploitation of user information by commercial enterprises; and (iii) various forms of cybercrime.[18] Online surveillance by nation-state actors is a global phenomenon, not limited to activities conducted by the NSA or even the U.S. Government.[19] Large Internet service providers have indicated their interest in aggregating and monetizing data about the web-browsing patterns of their subscribers by selling data to online advertisers.[20] For example, in March 2016 the Federal Communications Commission imposed a $1.35 million fine on Verizon Wireless for its use of a technology that allowed marketers to track customers' web browsing, without their knowledge, so they could be provided more targeted

[18] IETF RFC 7258 (May 2014), https://tools.ietf.org/html/rfc7258

[19] *See* https://rsf.org/en/news/special-report-internet-surveillance-focusing-5-governments-and-5-companies-enemies-internet; https://en.wikipedia.org/wiki/List_of_government_mass_surveillance_projects.

[20] *See* https://arstechnica.com/tech-policy/2017/03/ad-industry-lobbyists-celebrate-impending-death-of-online-privacy-rules/

online advertising.[21] Cybercrime, such as "hacking" users' credit card numbers or login credentials for purposes of fraud, or theft, is widespread on the Internet.[22] Commenting on this environment in an October 2015 address, Wikipedia founder Jimmy Wales reportedly stated, "There's really no excuse to have any major web property that's not secure[d]" using HTTPS.[23]

91. Although for nearly two decades encryption was used primarily to facilitate online banking transactions and credit card purchases, over the last five years HTTPS encryption of web communications has increasingly become the norm, and a security "best practice." Many websites now automatically redirect visitors to the secure (HTTPS) versions of their sites even if a user's browser first contacted the unencrypted (HTTP) version. By 2017, half of all traffic was encrypted, according to Mozilla, developer of the Firefox browser[24], and over 70 of the top 100 most popular websites worldwide had enabled HTTPS encryption, up from 37 in 2016, according to reports by Google.[25] Google reports that by January 2019 the number had risen to 96 of the 100 top sites (which account for approximately 25 percent of global web traffic), and that as of February 2019 over 90 percent of the webpages viewed using Google Chrome (the world's most popular web browser) were loaded over HTTPS connections.[26] CloudFlare, one of the largest content distribution networks (used to facilitate worldwide distribution of client websites' content to end users) now offers redirection to HTTPS encryption as a standard feature of its services.[27]

92. Any organization that continues to operate a website in the current digital environment without at least offering, if not requiring, a secure HTTPS encryption places at risk

---

[21] *See* https://www.cnet.com/news/verizon-racks-up-1-35-m-bill-for-violating-consumer-privacy/.

[22] *See* https://www.pcworld.com/article/205051/Norton_Study_Says_Cybercrime_is_Rampant.html.

[23] https://motherboard.vice.com/en_us/article/ezvj8k/jimmy-wales-theres-really-no-excuse-not-to-use-encryption

[24] *See* https://www.wired.com/2017/01/half-web-now-encrypted-makes-everyone-safer/

[25] *See* https://www.zdnet.com/article/google-this-surge-in-chrome-https-traffic-shows-how-much-safer-you-now-are-online/.

[26] *See* https://transparencyreport.google.com/https/overview?hl=en.

[27] *See* https://www.cloudflare.com/lp/ssl-for-saas-providers/.

the online privacy and security of its users, and perhaps its own proprietary information and commercial interests. Online vendors whose operation depends on the willingness of users to visit their sites and share sensitive personal information (like credit card numbers) to make purchases would discourage visitors and lose business if they did not offer a secure connection. Non-commercial sites that deal with sensitive subject matters, or that simply value online privacy as a matter of principle, also face increasing pressure to implement HTTPS encryption to protect the confidentiality of their users' communications.

93. The transition is being spearheaded through efforts by Google, and other online organizations, to promote the adoption of HTTPS encryption across the entire web. Since July 2018, Google's Chrome web browser has been labeling all websites that do not use the HTTPS protocol as "Not Secure," a warning that may be interpreted by visitors to mean that the site has been hacked or compromised, or may even be malicious. Mozilla's Firefox web browser is reportedly following suit.[28] Given the increasing demands today for greater Internet security, labeling a website as "not secure" could bring reputational damage to the site (and perhaps its parent organization), and would be contrary to the interests of any organization seeking to maximize visits to its site, whether its content is considered sensitive or not.

94. At this time two methods of HTTPS encryption are available. Opt-in HTTPS gives a user the option, when he or she visits a website, of communicating over an HTTP connection or choosing a secure HTTPS connection. In contrast, HTTPS by default involves the automatic redirection of a user to the HTTPS version of the website, even if the user's browser first seeks to make an unsecure HTTP connection. HTTPS by default also involves use of the HSTS (HTTP Strict Transfer Security) protocol to reconfigure the user's browser to connect automatically via HTTPS transmission whenever the user again visits the same site, and to refuse connection to the unencrypted site. Both versions require roughly the same initial and ongoing investment of resources to develop, operate, and maintain the required technical infrastructure, including the

[28] *See* https://www.deepdotweb.com/2018/01/05/mozilla-label-http-sites-not-secure-future-versions-firefox/

retention of qualified engineering personnel. In some circumstances, opt-in HTTPS encryption may be marginally less costly for an organization to implement, depending on the fraction of unencrypted HTTP traffic visiting the website, because it may require somewhat less server capacity than supporting all site visitors exclusively via HTTPS.

95. All other things being equal, the level of encryption protection provided is the same whether a web communication is encrypted with HTTPS optionally or by default. The difference between the two methods is that HTTPS by default adds a level of protection against so-called SSL stripping attacks. In an SSL stripping attack, if a user's browser makes an unsecure HTTP request to make a secure HTTPS connection to a website, an attacker can intercept that initial, unsecure request and use it to "strip" the encryption that otherwise would have protected the ensuing exchange of communications between the user and the website. The attacker thus obtains access to the contents of the user's communications with the site, including the information the user views or downloads from the site, and any sensitive personal or financial information the user shares with the site. Implementing HTTPS by default mitigates the threat of such attacks through the HSTS reconfiguration of a user's browser to encrypt future connection requests after the user's first visit to the site, or if the site is listed on the HSTS preload list incorporated into the user's browser. Largely for this reason, the adoption of default rather than opt-in HTTPS, like the adoption of HTTPS encryption generally, is increasingly becoming the norm, particularly by sites where users must enter login credentials or provide sensitive personal information in order to conduct financial transactions. For example, nine out of 11 U.S. banking sites and all 13 U.S. Government sites listed at the HTTPSWatch website support HSTS, and all but one redirect from HTTP to HTTPS automatically[29]. This confirms the prediction of an August 2017 study[30]: "Top websites will be almost entirely HTTPS within a year and a half. Half have

[29] See https://httpswatch.com/us; visited Feb. 9, 2019.

[30] Adrienne P. Felt, et al., "Measuring HTTPS Adoption on the Web," 26th USENIX Security Symposium, August 2017, https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-felt.pdf.

moved, more are preparing to move, and the rem[a]inder will feel pressured to meet the changing industry standard."

96. So far as I am aware, SSL stripping attacks have never been publicly identified as a surveillance technique employed by the NSA in connection with its Upstream collection program, whether in any official U.S. Government disclosures. Such an attack would require a man-in-the-middle attack or DNS redirection, not just passive intercept.

97. For the same wide variety of reasons identified by the IETF, going well beyond what has been publicly reported about NSA Upstream surveillance, encryption increasingly has become the default across the Internet as a whole, not simply on the World Wide Web. IPsec, the Internet Protocol Security suite, is a set of network security protocols commonly used by commercial enterprises and other large organizations to encrypt sensitive data that they transmit from one business site to another, such as between large data centers. Wikimedia has implemented IPsec both for the transmission of its web server logs between its foreign servers in the Netherlands and its servers located in the United States, *see* First Decl.¶ 83; Gov't Exh. 4 (Technical Statistics Chart), and apparently for the transmission of cache-to-cache data between its U.S. data centers, Plaintiff's Exhibit 3 (Paulson Decl.) ¶ 53; Plaintiff's Exhibit 39 at WIKI00006566. The transmission of a website's server logs in encrypted form is an accepted best practice, not only to protect proprietary information about the operation of the site, but to protect user information that may be considered personal. Indeed, the European Union's General Data Protection Regulation, which became effective in May 2018, classifies information contained in web server logs (principally IP addresses) as personal data that must be handled in a manner that ensures appropriate security, such as by encrypting them.[31] Finally, other internal communications that traverse the Internet, such as cache-to-cache communications, are protected by encryption primarily to ensure integrity, e.g., to prevent malicious actors from modifying web pages in transit. While this threat is less likely than others, ISPs in China, Russia and Pakistan, among other countries, have temporarily diverted traffic through their country. If

[31] https://www.ctrl.blog/entry/gdpr-web-server-logs

a third party could modify web pages in transit, they could install trackers or malware, for example.

98. Even if the NSA were not conducting Upstream surveillance, there would remain numerous reasons, discussed above, why an organization would be highly motivated to encrypt transmissions of its web server logs and cache-to-cache data. The online transmission of such sensitive proprietary or private information using unencrypted protocols not only poses known risks of interception, modification and theft by unauthorized third parties (including foreign government actors), it has also become an indicator that an organization lacks proper cyber hygiene.

CONCLUSION

99. For the reasons I discuss above and in my first declaration, it remains my opinion that, based on what is publicly known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications.

100. While I offer no opinion on the likelihood that the NSA does or does not, in fact, employ these techniques, I have examined the bases of Mr. Bradner's opinions (i) that the NSA, in conducting Upstream surveillance, "most likely" copies, reassembles, and scans for selectors all communications packets traversing an international Internet link that is monitored by the NSA (if any); (ii) that it is "implausible" that the NSA uses the traffic-mirroring techniques (white- and blacklisting) described in my first declaration; and (iii) that even if the NSA uses one or more of the techniques I described, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. I conclude that these opinions lack a non-speculative foundation in Internet technology and engineering.

101. It is also my opinion that even if the NSA were not conducting Upstream surveillance, in the current digital communications environment there would remain numerous reasons for an organization that operates one or more major websites to implement HTTPS-by-

Total Pages:(3523 of 4208)

default on its websites, and for an organization that transmits large volumes of proprietary or other sensitive data across the Internet to encrypt those transmissions using IPsec or another such encrypted transmission protocol.

I declare of penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Executed in New York, New York on February 15, 2019.

HENNING G. SCHULZRINNE

Pg: 57 of 484
Filed: 07/01/2020
Doc: 18-6
USCA4 Appeal: 20-1191

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 7

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	)	No. 1:15-cv-0662 (TSE)

DECLARATION OF DR. ALAN J. SALZBERG

Dr. Alan Salzberg, for his declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

I. Introduction

1. I am the Principal (and owner) of Salt Hill Statistical Consulting. My work includes statistical sampling, analysis, and review for government and industry. I was asked by the U.S. Department of Justice to review the Declaration of Jonathon Penney filed on December 18, 2018 in the above-captioned case. ("Penney Declaration"). In particular, I was asked to assess and provide my conclusions concerning the validity of both the statistical conclusions reached in the Penney Declaration and the underlying methodology.

2. The Penney Declaration presents an empirical data analysis of Wikipedia page-view data and concludes that "public awareness of NSA surveillance programs, including Upstream surveillance, which became widespread during the June 2013 Snowden disclosures, is highly likely to have had a large-scale chilling effect on Wikipedia users."[1] My review analyzes the data, methodology, and conclusions presented in the Penney Declaration.[2]

3. This declaration proceeds as follows. In the next section, I summarize my opinions. In Section III, I review my qualifications. In Section IV, I detail the reasons for my opinions. And in Section V I set forth my conclusions. Appendix I contains my programming code from which I produced the analyses contained in this report. Appendix II lists the documents and data I considered as part of this report. Appendix III contains my resume, publications for the last 10 years, and testimony history for the last four years. Appendix IV contains a graph showing page views by article for each of the 48 articles the Penney Declaration theorizes were influenced by a chilling effect. Appendix V contains the same 48 articles but for an extended time period that continues through November 2018. Appendix VI contains a graph showing page views by article for each of the 89 articles described in the Penney Declaration as comparative articles (which purportedly were not affected by the June 2013 disclosures). Appendix VII contains the aggregate graphs for each of the five comparison datasets.

II. Summary of Opinions

4. In summary, I find that:

   A. The methodology used in the Penney Declaration—which purportedly shows an upward trend in page views of certain articles posted on Wikipedia through May 2013, followed by an abrupt drop and downward trend in views of those articles beginning in June 2013—is deeply flawed, inappropriate, and likely biased.

---

[1] Penney Declaration, paragraph 10.

[2] The Penney Declaration, in paragraphs 12 through 21, describes research on chilling effects theory. The Penny Declaration's stated conclusions in Paragraph 11 do not rely on that overview section, and I was not provided, nor does the Penney Declaration present, any data on this research. Therefore, I did not review or consider those paragraphs further. Furthermore, it does not appear that any of that research was specific to Upstream.

B. The Penney Model simply assumes that a single change occurred in June 2013, rather than letting the data identify the timing and number of changes in trends that occurred. Even though there is no consistent trend in the data, the design of the Penney Model will create the appearance that the data contain just one inflection point. And, because of its design—even though changes in trend occurred *before* these June 2013 disclosures—the Penney Model will find that the disclosures caused them.

C. Contrary to the hypothesis presented in the Penney Declaration, analysis of page views for the 48 individual articles in the privacy-sensitive group do not show a rising trend followed by an immediate and sustained drop in June 2013.

D. With the one exception of removing the article on Hamas, the Penney Declaration does no analysis or adjustment for factors (such as world events) affecting these individual article page views. Instead, the Penney Declaration inappropriately aggregates the vastly different page view data for individual articles, with the result that these individual differences in page views are masked.

E. Even at that aggregate level, I find that the hypothesized peak in page views of "privacy-sensitive" articles in May 2013 does not exist, and the hypothesized upward and then downward trends in views of privacy-sensitive articles before and after June 2013, respectively, do not exist.

F. Extended data through 2018 regarding page views of the privacy-sensitive articles do not indicate a long-term decline in page views from pre-June 2013 levels.

G. A proper control dataset would exhibit similar page view behavior prior to June 2013. The comparison datasets used in the Penney Declaration do not and are thus inappropriate controls.

H. The Penney Declaration analysis ends in July 2014. No data are presented that shed any light on whether page views at the time the Amended Complaint was filed in 2015 (or thereafter) were affected by Upstream. In other words, even if the purported effect and trends were a correct conclusion for the data examined (and they are not), the Penney Declaration analysis does not and cannot show that the effect continued years after the study ended.

I. Even if a chilling effect occurred in June 2013, there are no data analyzed in the Penney Declaration that show any effect was due specifically to "public awareness of" the specific NSA surveillance program challenged here (known as Upstream surveillance) rather than possible inaccuracies, if any, about the program reported in the press, disclosures about other NSA programs, disclosures about other surveillance programs (e.g., surveillance by Britain), or other, unrelated events of June 2013.

I describe the analyses that led to these findings in Section IV.

III. Qualifications

5. I am the Principal of Salt Hill Statistical Consulting. My work includes statistical sampling, analysis, and review for government and industry. Many of my consulting projects and research papers relate to the detection and measurement of bias. On several occasions, I have written expert statistical reports or testified as a statistical expert, both in court and in depositions. My current and recent work includes:

    - Statistical analysis and modeling regarding the valuation of residential mortgages. Assisted in developing complex models to evaluate portfolios of loans affected in the housing crash of 2008.
    - On behalf of several state public service commissions, directed data analysis and statistical design in a series of systems tests of Bell South, Verizon, SBC-Ameritech, and Qwest. Testified before several state public service commissions, including New York, Virginia, Florida, Michigan, and Colorado. Co-inventor of U.S. Patent related to this work.
    - For a major pharmaceutical company, analyzed company and external marketing data to determine reliability and potential biases in using external data sources. Analyzed physician-specific data for a period of 36 months concerning product marketing to approximately 1 million prescription drug subscribers.
    - Statistical sampling and analysis, including regression modeling and survival analysis, on behalf of the U.S. Department of Labor.
    - Statistical review of the sampling and estimation methodology used to audit Medicaid providers in New York State. Work was performed on behalf of the New York State Office of Medicaid Inspector General.

6. I received a Ph.D. in Statistics from the University of Pennsylvania, where I also received a B.S. in Economics. I have taught courses in statistics and quantitative methods at the University of Pennsylvania and American University and have published statistics papers in peer-reviewed journals. I am also the co-inventor on a U.S. Patent (#6,636,585) for a statistical process design to test the systems of telecommunications companies. A copy of my résumé is attached as Appendix I to this Report, which also includes all publications within the last ten years and a list of testimony within the last four years. My company is being compensated at a rate of $560 per hour for my work in this matter.

IV. Details of Findings

A. Background and Data

7. The analysis presented in the Penney Declaration uses eight datasets to analyze a hypothesized "chilling effect" on Wikipedia users due to "public awareness of NSA

surveillance programs, including Upstream surveillance."[3] The first three datasets (which I will call the "Terror" datasets) contain monthly page-view information for 48 so-called "privacy-sensitive" Wikipedia articles that Dr. Penney selected because they contain terms included in a 2011 U.S. Department of Homeland Security list of "terrorism related keywords."[4] These three overlapping datasets contain page views for Wikipedia articles from January 2012 through August 2014 ("study period").[5] The first dataset contains the monthly page views, by article, for each of the 48 articles, by month, for the study period. I will call this dataset "Terror 48."[6] The second dataset contains monthly page views for 47 articles, which are comprised of all of the original 48 articles except for the article on "Hamas." I will call this dataset "Terror 48 without Hamas." The third dataset, which I will call "High Privacy 31," contains page-view data for 31 of the 48 articles deemed most "privacy-concerning" by the Penney Declaration.[7]

8. The Penney Declaration also considers five comparison datasets. According to the Penney Declaration, these datasets include two datasets of total global article views (which I call "Global 1" and "Global 2");[8] 25 domestic-security related articles ("Security 25"); 34 infrastructure articles ("Infrastructure 34"); and 26 popular ("Popular 26") articles.[9]

9. I supplemented the data in the Penney Declaration using publicly available data from Wikimedia to capture information on page views for each of the Terror 48 articles for the time period from July 2015 through November 2018. Therefore, for some of my analyses, I use data from January 2012 through November 2018, except for the period from September 2014 through June 2015, which was not in the original study period and for which data are also not currently available.[10]

10. The Penney Declaration posits a statistical model (which I will call the "Penney Model") and uses the datasets to estimate the parameters of that model and draw the conclusions described in paragraphs 10, 11, and 58 of the Penney Declaration. The Penney Model posits a straight-line trend in page views for each month from January 2012 through May 2013; an immediate change in June 2013; and a second straight-line trend for each month

---

[3] Penney Declaration, paragraph 10.
[4] Penney Declaration, paragraph 31.
[5] Penney Declaration, paragraph 34.
[6] In the Terror 48 dataset provided as support for the Penney Declaration, the articles "Recruitment" and "Fundamentalism" have exactly the same number of page views in 30 of the 32 months, and therefore I concluded that Penney made a copy/paste error with respect to this data. The inclusion of this error in the analyses makes little difference for the first 32 months, but in comparing page views for the more recent time period where I supplemented the data, I could not determine whether the data for the original 32 months should have been associated with Recruitment or Fundamentalism and therefore I exclude both where noted.
[7] Penney Declaration, paragraph 48. According to the Penney Declaration, the so-called high privacy articles were determined using a survey conducted via an online survey tool named Mechanical Turk, which I did not evaluate for its accuracy or validity.
[8] Penney Declaration, paragraph 49. The Penney Declaration did not include analyses for the Global 2 dataset but since that dataset was provided to me as part of the data that was considered in the Penney Declaration, I include it in my analyses. The Global 2 apparently includes mobile data whereas the Global 1 dataset does not.
[9] Penney Declaration paragraphs 52-56 describe the Popular, Infrastructure, and Security articles.
[10] If available that data could have been used to provide further insight into trends, but its unavailability is irrelevant to my conclusions.

from June 2013 until August 2014. The hypothesis for the articles in the Terror datasets[11] is that there is a steady increase through May 2013, followed by an immediate decline in June 2013, followed by a steady decline thereafter. Furthermore, the hypothesis for the sets of comparator articles is that they experience neither an immediate decline nor a change in monthly trends in June 2013.[12]

B. A Simple Review of Article Page Views Indicates That A Decline in Page Views Does Not Begin in June 2013

11. Before reviewing the specific analysis found in the Penney Declaration, I review the page views for the individual 48 terror-related articles (the Terror 48) that the Penney Declaration claims were subject to a chilling effect in June 2013.[13] I find that the page views per article controvert the Penney Declaration conclusion (based on aggregation of the page view data) that there is a rise until May 2013 followed by "statistically significant and substantial drop in view counts immediately following June 2013."[14]

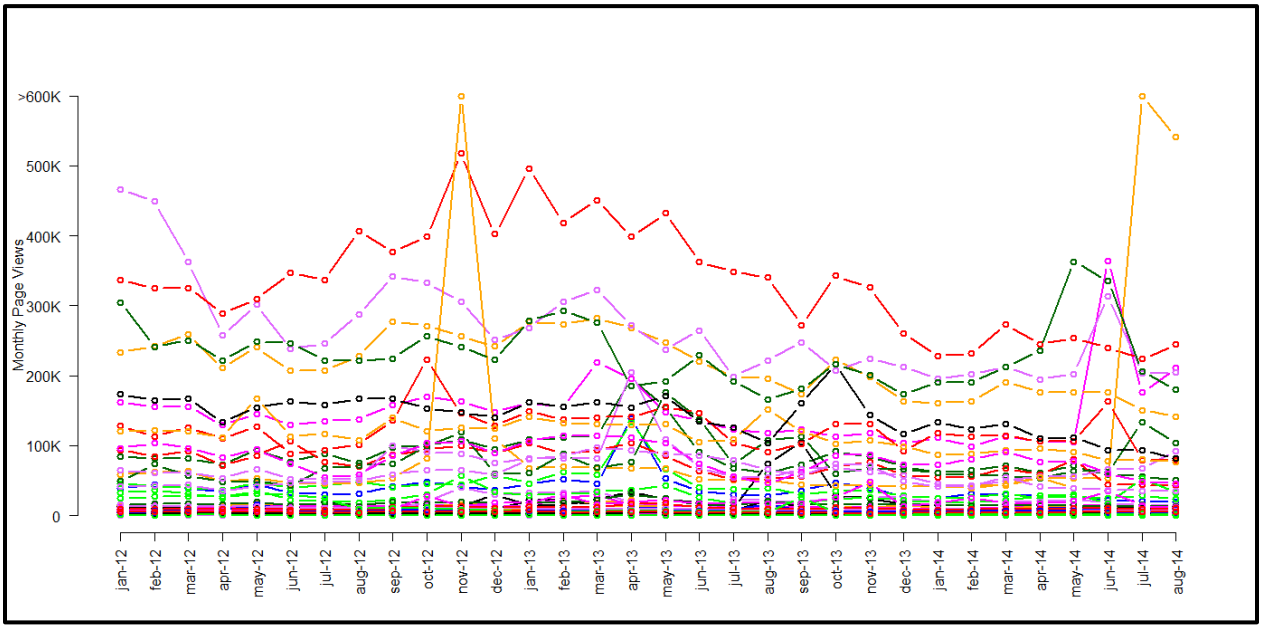
12. My review of the page views for the individual articles shows that almost none of the Terror 48 articles experiences its peak in May 2013 (the hypothesis of the Penney Declaration). For the Terror 48 articles, 17 had already reached their peak number of page views in 2012 and 18 more reached their peak at some point between January and April of 2013. In other words, 35 out of 48 (73%) reached their peak prior to the hypothesized peak of May 2013, and thus the occurrences of June 2013 could not have possibly caused any of these drops in page views. Eleven more of the articles (23%) reached their peak after the disclosures, meaning there was no immediate and sustained drop in June 2013, again controverting the hypothesis in the Penney Declaration. Just two out of 48 (4%) reached their peak in the hypothesized month of May 2013. Even these two articles, though they reached their highest level in May 2013, do not appear to follow the pattern of a steady rise until May 2013 and then a sustained drop afterwards.

13. While many (but not all) of the Terror 48 articles experienced higher numbers of page views in 2012 and early 2013 when compared to late 2013 and early 2014, the decline did not begin in June 2013. Furthermore, the page views did not consistently rise or fall for any sustained period for most articles. To visually demonstrate this fact, I plotted the page views for each of the Terror 48 articles on a single graph. As shown in Figure 1, there is no immediate decline in June 2013, no consistent upward trend through May 2013, and no consistent downward trend that begins in June 2013.

---

[11] The analysis covers all 48 articles but the conclusions made in the Penney Declaration apply only to 47 (the Terror 48 minus Hamas set of articles) and 31 (the High Privacy 31) of those articles.
[12] See Penney Declaration, paragraph 11.
[13] Technically, the Penney Declaration only makes conclusions regarding the Terror 48 articles without Hamas and the High Privacy 31 articles (see paragraph 58 of the Penney Declaration) but I review all 48 articles here for completeness.
[14] Penney Declaration, paragraph 11. The "trend reversal" referred to in Penney Declaration Paragraph 11 is alluding to a purported rise prior to June 2013 and a drop afterward.

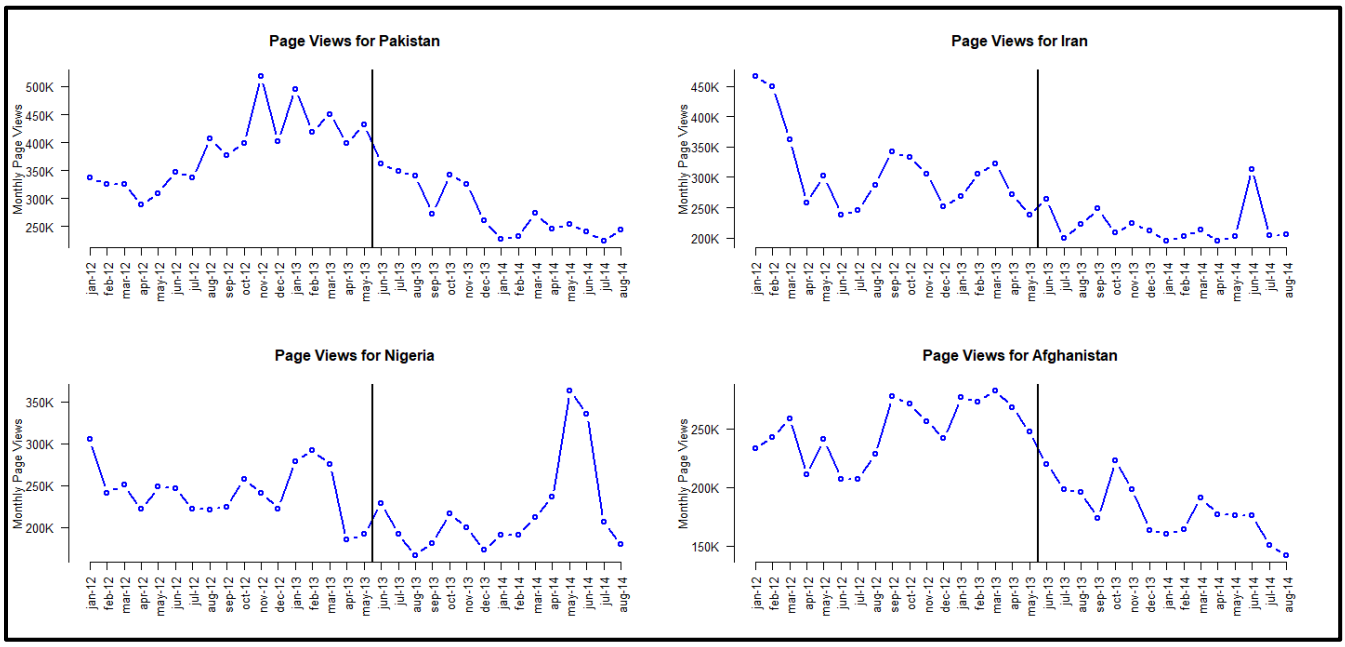
Figure 1: Individual Page Views for Each of the Articles Within the Terror 48, Which The Penney Declaration Hypothesized Show an Immediate Decline Beginning in June 2013



14. In short, the Penney Declaration's conclusions are controverted by a simple disaggregated review of the data for each article. The rest of my report carefully reviews the data and the Penney Declaration to explain the reasons for the incorrect conclusions.

15. While Figure 1 is helpful in showing that there is no overall or consistent downward trend starting in June 2013, reviewing the page view data for individual articles allows one to see that none of the articles follows the hypothesis set forth in the Penney Declaration. (I have included page view data for each of the articles in the Terror 48 set in Appendix IV.) For example, Figure 2 below shows the page views for the four articles with the most page views of the Terror 48. As can be seen in these individual graphs, there does appear to be a general decline in page views. However, that decline did not begin with the June 2013 disclosures. Page views for the Pakistan article peaked in 2012, and followed with an erratic decline. Page views for the Iran article saw their peak in January 2012, and erratically declined thereafter. Page views for the Nigeria article were more erratic, with no clear increase or decline. Page views for the Afghanistan article were erratically increasing or remaining about the same until early 2013 when they began to erratically decline.

Figure 2: Individual Articles show no Association of June 2013 with a Decline in Page Views



16. These four graphs, above, are indicative of the pages views of all 48 articles in that not one of the 48 articles appears to follow the Penney Declaration hypothesis of a steady increase through May 2013 followed by an immediate drop and steady decline beginning in June 2013. In addition, a review of the entire set of individual graphs by article, which I have provided in Appendix IV, reveals that there are vast differences in monthly page views over time in each article.[15] Given those vast differences, it is not statistically appropriate to combine them for the purposes of analysis, as Dr. Penney did in his analysis.

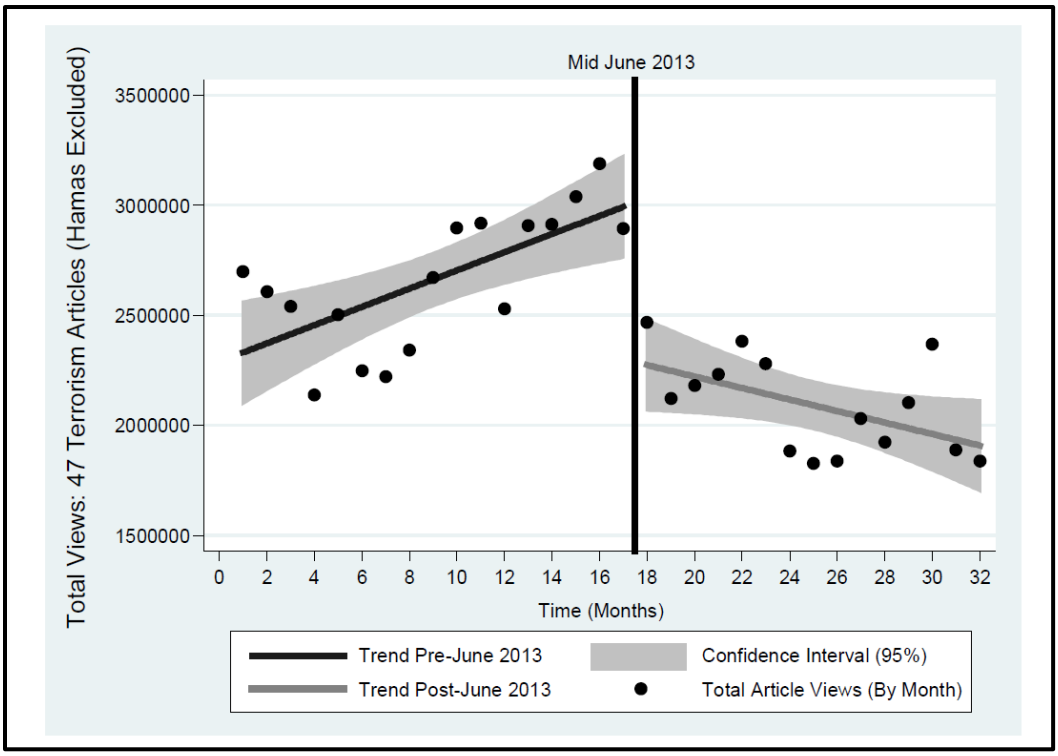
17. As I explain in Section F below, ignoring these differences biases the model and renders it invalid. The simple reason is that such aggregation masks the individual differences in page views. Although aggregation can be appropriate in instances where most of the data tell a consistent and similar story and the aggregation merely eliminates outliers (which would, in that instance, be considered "noise"), where the data are vastly different (as here) aggregation skews the data and tells a misleading story. While I review the aggregate data analyzed in the Penney Declaration in the next section, my review does not imply agreement with the methodology of aggregating the data here.

[15] Note that I scaled each of the 48 graphs according to its page views in order to clearly show the trends. In the aggregate analysis performed in the Penney Declaration, the articles with the most page views are also treated as highly influential because the aggregation of the graphs is influenced according to page view.

C. The Aggregate Data Analyzed in the Penney Declaration Do Not Indicate Either a Peak in May 2013 or a Long Term Decline Beginning in June 2013

18. I begin my analysis of the aggregated data with an analysis of the Penney Declaration's Figure 2, which shows the Terror 48 without Hamas data set (totaling 47 articles) that were analyzed. A careful view of the Penney Declaration's Figure 2 (reproduced below as my Figure 3) indicates that the peak in monthly page views does not occur in May 2013 and there is no immediate drop or trend reversal in June 2013. In other words, even the aggregated figure presented in the Penney Declaration fails to show the hypothesized trend reversal and drop in June 2013.

Figure 3: Penny Declaration Figure 2 Reveals Some of the Flaws of the Penney Declaration Analysis



19. The suggestive trend lines in the Penney Declaration's figure give the impression of a steady increase followed by a decrease, but the points, representing individual months, reveal otherwise. Careful attention to Figure 2 in the Penney Declaration reveals that the page views went up and down several times over the course of the 32 months shown and did not have a single peak in May 2013 (month 17 in the Penney Declaration figure reproduced above).

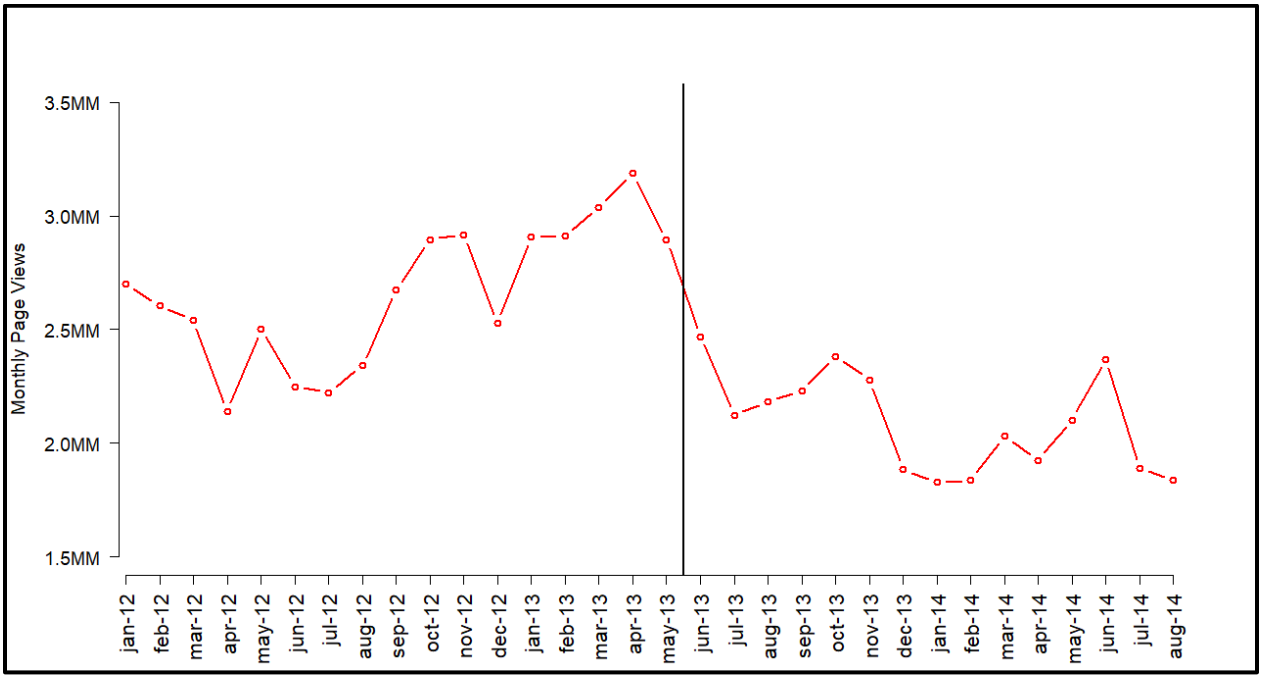
20. Furthermore, only 16 of the 32 months (50%) show page view totals within the model's 95% confidence interval. A properly constructed 95% confidence interval should contain about 95% of the data points. In this instance, the failure to capture a remarkable 50% of

the data points within the 95% confidence interval may be due to an incorrect model, improper construction of the interval, or both.

21. Using the same data points that the Penney Declaration analyzes, I re-drew the Penney Declaration Figure 2 (see Figure 4 below), adding proper labeling of dates and removing suggestive trend lines. In contrast to the solid upward line drawn on the Penney Declaration figure, my plotting of the same points in Figure 4 shows that there are a number of both declines and increases. There is a notable trough in the Summer of 2012, for example, and the number of page views appears to be generally declining through July 2012. Importantly, the highest number of page views occurred in April 2013 and not the hypothesized May 2013.

22. Beyond June 2013, when the Penney Declaration hypothesizes a steady decline, the number of page views go up and down, rising three months in a row from August through October 2013, and again rising three out of four months from March through June 2014.

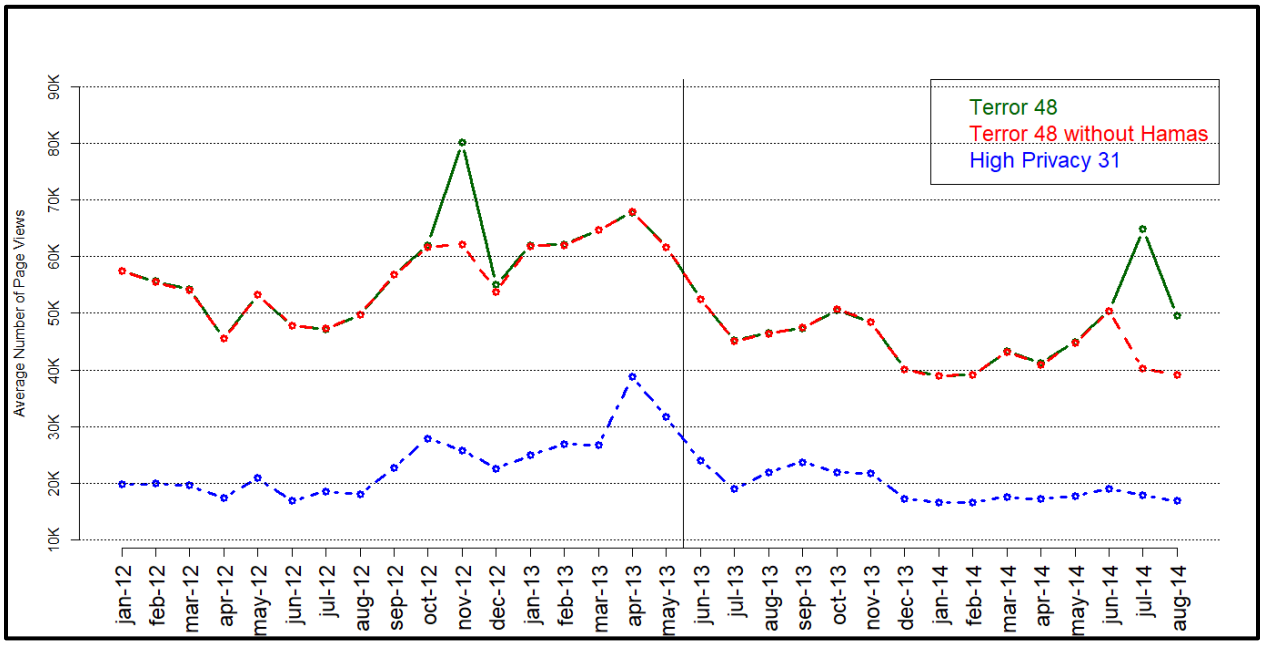
Figure 4: Terror 48 Without Hamas Dataset Without the Penney Declaration "Trend" Lines



23. Figure 5 below adds the other two datasets analyzed (Terror 48 and High Privacy 31) to the Terror 48 Without Hamas dataset graphed above, and I used the average page views per article rather than the sum.[16] Once again, Figure 5 indicates that the peak is in April 2013 (and prior to April for the Terror 48 dataset) and that there is no sudden drop in June 2013.

[16] The red line in Figure 4, which shows the total page views for the Terror 48 without Hamas data, has exactly the same pattern as the red line in Figure 5, which shows the average page views for the same data set. The left axis in Figure 5 is just divided by 47 in order to display the average instead of the total.

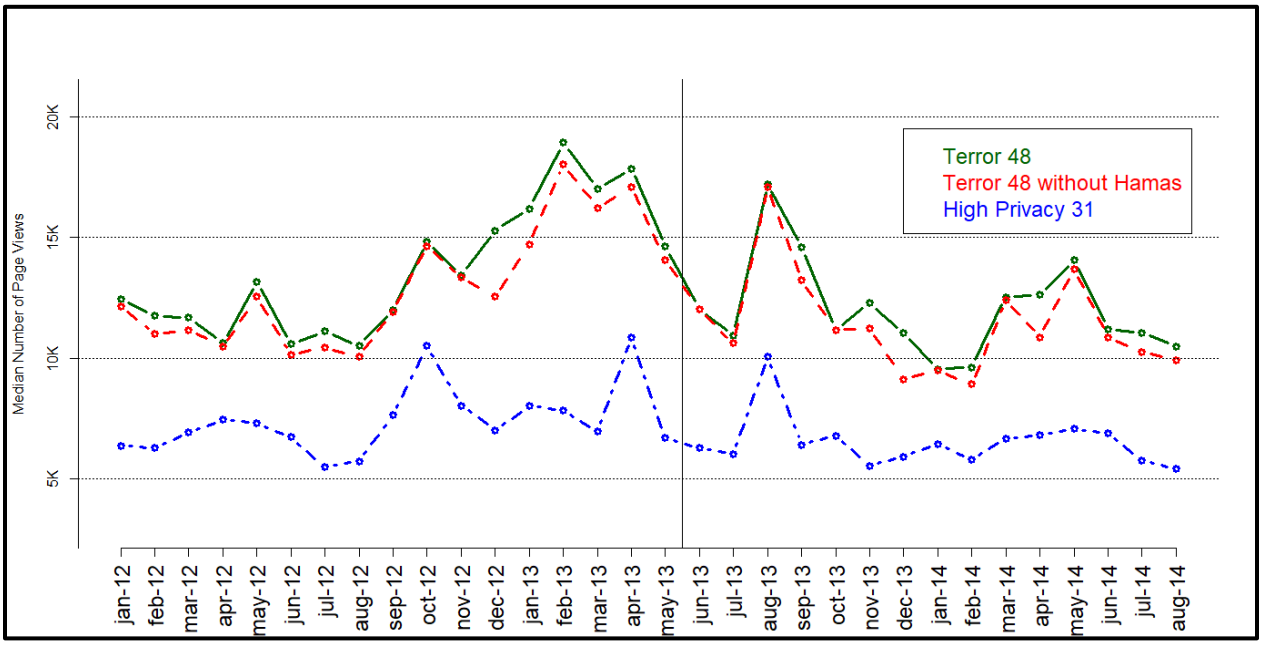
Figure 5: Average Page Views Show a Peak in April 2013 or Before



24. Because the average number of monthly page views can be affected by a single article with a very high number of page views in a particular month, I also show the median number of page views by month in Figure 6, below. The median number of page views for any given month is the middle number of page views when the number of views by article is sorted from the lowest number of views to the highest number of views. Therefore, the median shows the number of page views for the "typical" article in the group for a particular month, and therefore is not sensitive to a few articles with very high (or very low) page views for a month. As shown in Figure 6, the peak in median page views occurs prior to the hypothesized peak of May 2013. These data indicate that a rise in page views began in the Summer or Fall of 2012 and peaked in the Winter or Spring of 2013.

25. Figure 6 indicates that while page views generally rose for some time beginning in late 2012, no dramatic peak or fall occurred. Instead, there was a slow and unsteady rise and decline. The page views appear to level off to about early 2012 levels by the Summer of 2014, when the Penney Declaration data end.

Figure 6: Median Page Views Show a Peak in April 2013 or Before



26. In summary, based on the individual article data and the aggregated data, the Penney Declaration hypothesis of an increase through May 2013 followed by an immediate and continuing drop afterwards has no support.

D. Extended Data on Page Views Does Not Indicate an Immediate or Long Term Decline Beginning in June 2013

27. The individual and aggregate article data are very different but they are consistent in that they both show that there was no abrupt and sustained decline in monthly page views beginning in June 2013. The figures and analyses above, like the Penney Declaration, only use page view data through August of 2014. As I explained, I also supplemented that data with publicly available page view data from Wikimedia, by article, for the period July 2015 through November 2018.[17]

28. While I obtained data for each of the original 48 articles, there are inconsistencies or errors associated with five of those articles. Specifically, there were five articles in which the keywords changed, i.e., that the article was under a prior keyword but now a search for that keyword redirects to a different article (e.g., the "terror" article became "fear").[18]

---

[17] A link to this data ("Hamas" page is shown as an example in this link) is https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2015-07&end=2018-11&pages=Hamas. The data are taken from en.wikipedia.org, with a selection of monthly data on all platforms with an "Agent" of "user."

[18] The five articles in which key words changed are: 1) "weapons grade" is now "weapons grade nuclear material"; 2) "Euskadi ta Askatasuna" is now "ETA (separatist group)"; 3) "pirates" is now "piracy"; 4) "Islamist" is now "Islamism"; and 5) "terror" is now "fear". The article "title" and "keyword" were synonymous prior to the changes (i.e., when a user entered the keyword into Wikipedia's search tool, they were directed to an article of the same name). After the changes, entering the keyword into the search tool directs you to the new article. When I gathered the page view information the keyword terror redirected to an article titled fear, for example. I note that now, on

In addition, I noticed that the data for two other articles containing the keywords recruitment and fundamentalism were exactly the same in the dataset provided along with the Penney Declaration in all but two months. This apparent error in the Penney Declaration data affects comparisons of those keywords with their correctly downloaded page views from 2015 through 2018. Because of the inconsistencies and errors for these seven articles' data, I include these in some analyses and exclude them in others. Their inclusion or exclusion does not change my conclusions.

29. In summary, I created a dataset for all 48 articles from January 2012 through November 2018, excluding September 2014 through June 2015 because Wikimedia does not make the data for those months available. Since there are five articles with differing key words and the two articles with potential data errors, I exclude those seven of the 48 articles from sets (b), (c), and (d), identified below. In short, when presenting the data for the entire 2012-2018 period, I use four datasets analogous to the terror datasets used in the Penney Declaration to examine page views for the 2012 to 2014 period, but which take into account the exclusion of data from the seven articles with anomalies:

    a. Page views for the 48 terror-related articles, which as noted above I call the "Terror 48;"

    b. Page views for the Terror 48 without the seven articles that have inconsistencies in data or naming, which I call "Terror 41;"

    c. Page views for Terror 41 without the Hamas article, which I call "Terror 41 without Hamas";

    d. Page views for the 26 articles that were included in the 31 "high privacy" in the Penney Declaration and that were also part of the Terror 41 articles. I call these articles "High Privacy 26."[19]

30. The four datasets all show that there was no immediate or long term decline in monthly pages views that began in June 2013. I provide graphs for each of the Terror 48 articles over the extended period in Appendix V, and my earlier conclusion is the same: there is no immediate or long-term drop in any of the individual articles' monthly page views beginning in June 2013.

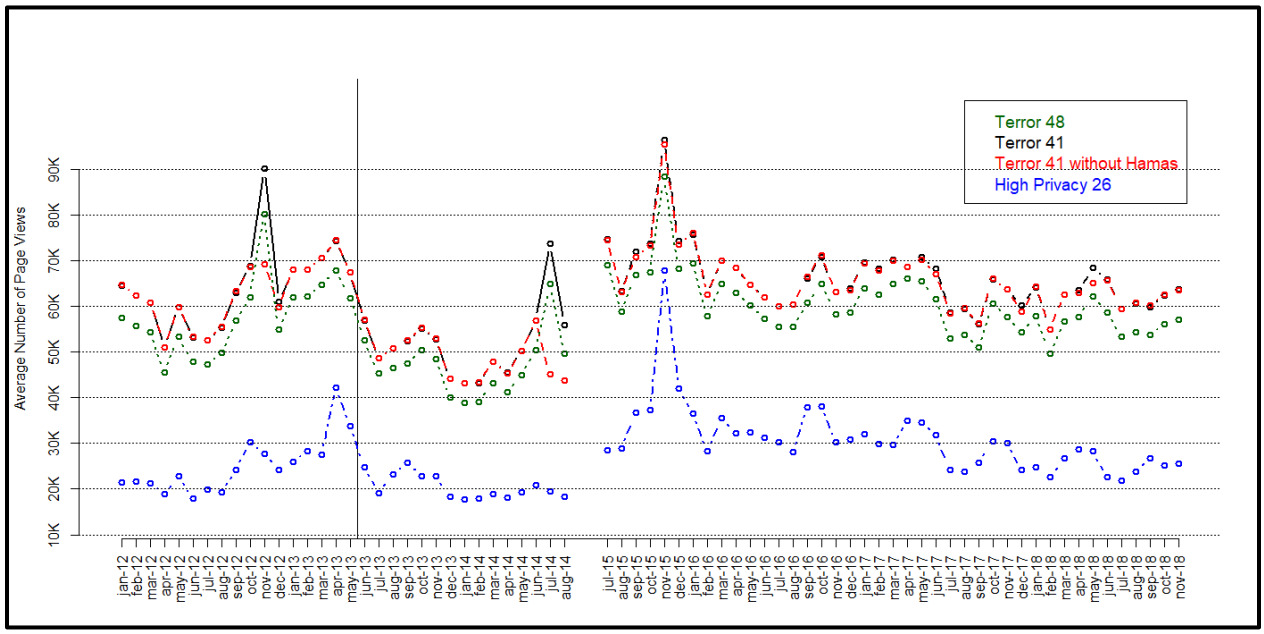
31. I also show the aggregate data over the extended period. Figure 7 below shows the average monthly number of page views for the terror datasets. The later data show many months with average page views in the range of 60,000 to 70,000, about the level of the peak months prior to June 2013. In other words, to the extent that page views did decline in late 2013 and early 2014, that decline appeared to reverse in 2015.[20]

February 14, 2019, terror no longer redirects to fear but instead again goes to a Wikipedia article called "Terror." The other four keywords still redirect as described above (as of February 14, 2019).

[19] The High Privacy 26 contains views for the 31 High Privacy articles after removing the five articles (among the seven articles) that had data issues, *see* above n.18, and were among the 31 High Privacy articles. Those five are Islamist, Recruitment, Weapons Grade, Euskadi ta Askatasuna, and terror.

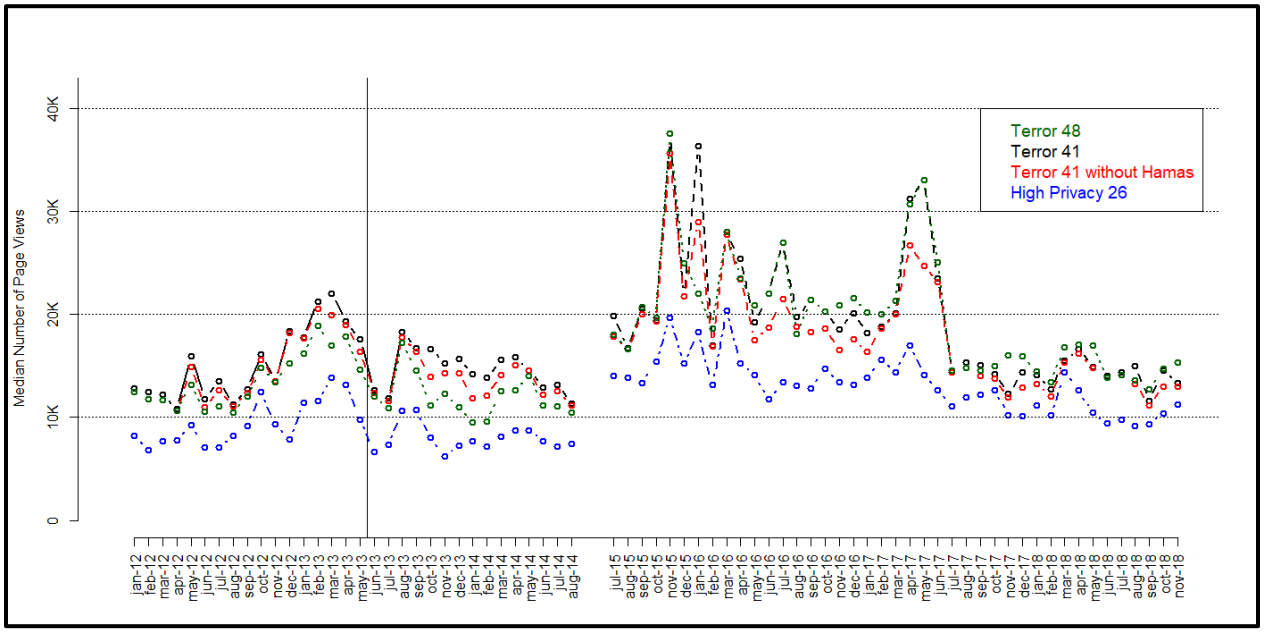
[20] As I will explain further below, the behavior of the aggregate data need not be indicative of the behavior of the individual article data. For example, the aggregate averages have a peak near the November 2015 Paris terror attacks, but that does not mean that all or most of the individual articles peaked around that time.

Figure 7: Average Page Views for Extended Period (Through November 2018) Fail to Support the Theories in the Penney Declaration



32. The average number of monthly page views is heavily influenced by the articles with the largest number of views and can be skewed by a single article with heavy readership in a single month. For that reason, I also calculated the median page views by month for the data through November 2018. As shown in Figure 8, median page views in 2015 and beyond often surpassed June 2013 views, a fact that undermines the theory that page views declined and remained low after June 2013.

Figure 8: Median Page Views for Extended Period (Through November 2018) Undermine the Theories in the Penney Declaration



E. The Comparison Datasets used in the Penney Declaration are not Comparable and So Do Not Corroborate Its Conclusions

33. The Penney Declaration bases its conclusions in part on the fact that following May 2013 the page views in the five comparison datasets did not decrease in a similar manner as the page views in the terror datasets.[21] Even assuming the issues with the extended terror-related datasets discussed above did not exist, the conclusion regarding the comparison datasets is flawed because the Penney Declaration does not demonstrate that the comparison datasets were truly comparable.

34. In particular, the Penney Declaration does not demonstrate that the comparison datasets would have had increases and decreases similar to those of the terror datasets *but for* the June 2013 disclosures. There is no analysis in the Penney Declaration that shows that the trends in page views were similar before June 2013 nor does the Penney Declaration explore whether other factors may have changed the trend of the comparison groups in ways that would not have changed the trend of the terror articles.

35. This issue means there is potential bias in any comparisons due to what is called selection by history. In simple terms, this means that if the comparison groups are not similar to the terror datasets to begin with prior to June 2013 (and thus not changing in a similar

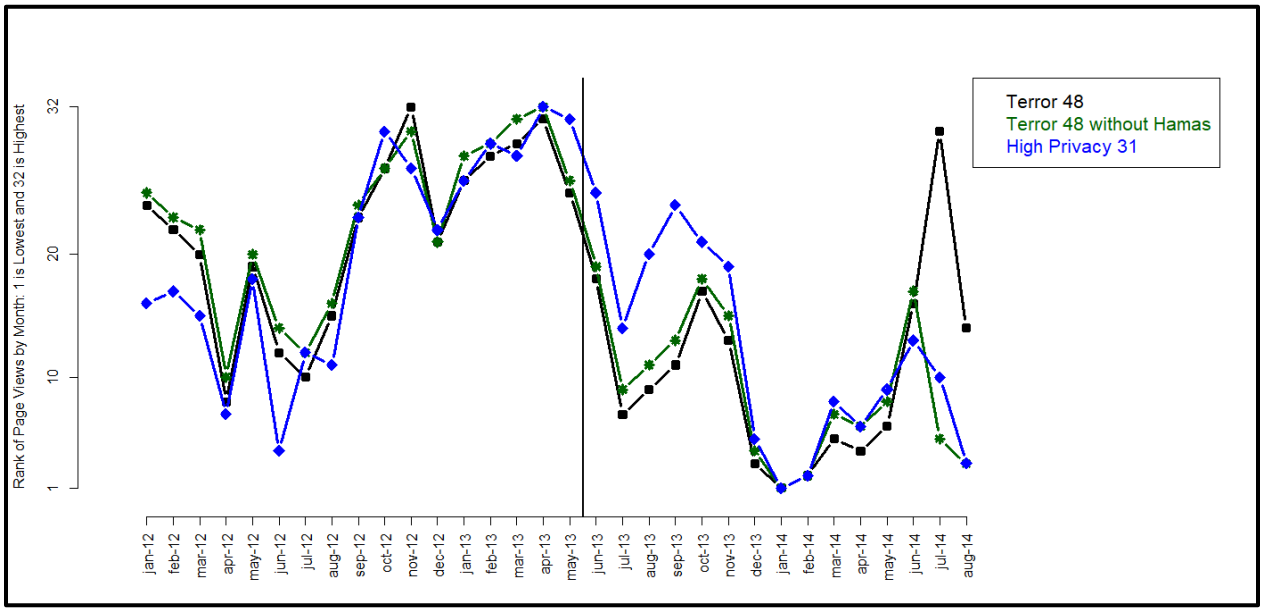
[21] These five datasets consist of "three comparator article groups" cited in paragraph 53 of the Penney Declaration as well as the two global view datasets of Wikipedia home page views used in the Penney Declaration. See my description of these datasets, above, in paragraph 8.

way over time), the estimated effects derived using such comparison groups could be wrong.[22]

36. A simple way to explore whether the terror and comparison datasets are changing in a similar manner prior to the June 2013 disclosures is to review their monthly page views. The magnitude of page views for the five comparison datasets is far different than it is for the terror datasets. Therefore, for each dataset, I ranked the page views by month for each of the 32 months from January 2012 through August 2014. This means that for each dataset, the month with the lowest number of views will have a rank of one, the one with the second lowest will have a rank of two, and so forth, up to the rank of 32, which will be assigned to the month with the highest number of page views.

37. Figure 9 below plots these rankings using the method described in paragraph 37, above, for the following datasets: Terror 48, Terror 48 without Hamas, and High Privacy 31.[23] They are very similar, which is not surprising since two of the three datasets comprise subsets of the articles in the Terror 48 dataset. As shown in the chart, the highest month appears to be either November 2012 or April 2013.

Figure 9: Ranked Page Views for Terror Articles



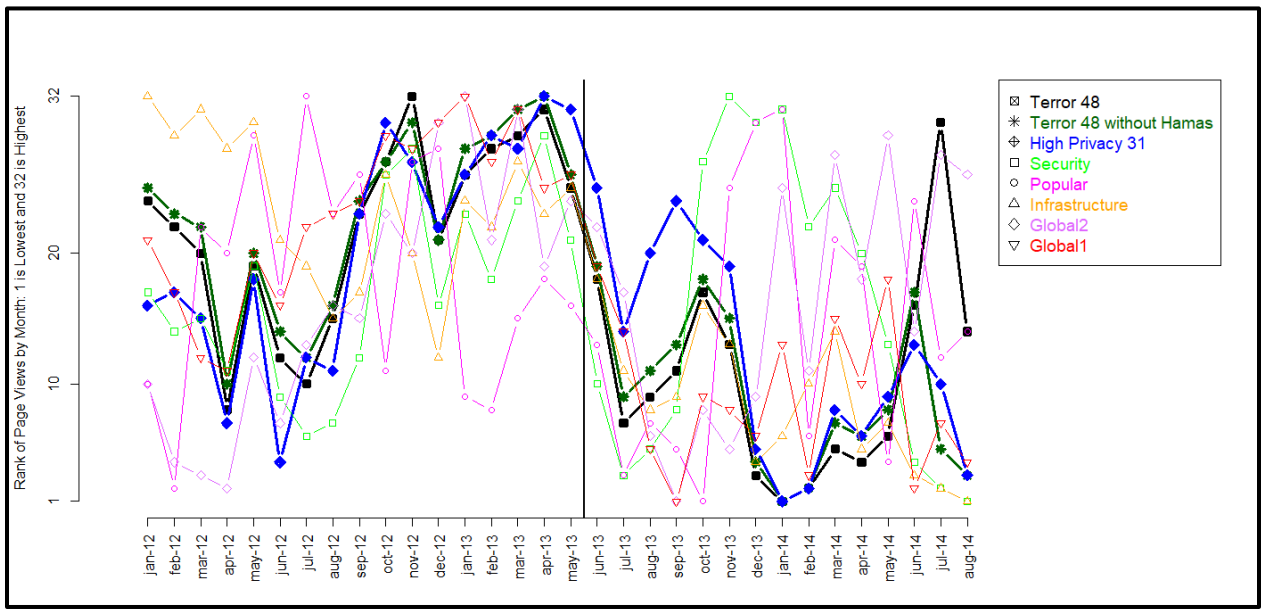
38. Figure 10 below shows the ranked page views for the same three terror datasets along with the five comparison datasets. In order for the comparison between the three terror datasets on one hand and the five comparison datasets on the other hand to be appropriate in determining whether the June 2013 disclosures had a singular effect on the Terror datasets, the trends in page views of the comparison articles would need to be similar prior to June 2013. In other words, a proper control group would roughly follow the

[22] See, for example, "Campbell, Donald, and Stanley, Julian C., Experimental and quasi-experimental Designs for Research, 1963, Houghton-Mifflin, p. 55-57. This issue is also discussed in Salzberg, Alan J., "Removable Selection Bias in Quasi-Experiments," The American Statistician, 1999, pp. 103-107.
[23] See paragraph 7 for detailed descriptions of these datasets.

trend of the Terror articles datasets prior to June 2013, when there is not yet any hypothesized effect. This would mean that a comparison of the data after June 2013 could potentially be used to estimate an effect.

39. Instead, the pre-June 2013 trends of the terror and comparison datasets are not at all alike. Figure 10 shows erratic behavior in the page views for the so-called five comparison datasets prior to June 2013 and that erratic behavior does not mimic the (also) erratic movements in the terror datasets. Therefore, the comparisons made in the Penney Declaration are not appropriate.

Figure 10: Ranked Views of Terror and Comparison Show Very Different Trends Even Prior to June 2013.



40. The comparison in Figure 10, which appears to show that the so-called "comparator" groups are not, in fact, comparable prior to June 2013 is confirmed by the Penney Declaration analysis. The Penney Declaration analysis is summarized after paragraph 53 (in Figure 3 of the Penney Declaration), which I have reproduced below as Figure 11.

Figure 11: Snapshot of Penney Declaration Figure 3

Wikipedia Article Group	Monthly trend pre-June 2013	Change in view count in June 2013	Change in monthly trend after June 2013	Model Fit
47 Terrorism Articles	**41,420.51**** *p*=0.00	**−693,616.9**** *p*=0.00	**−67,513.1**** *p*=0.00	**Yes** *F*=0.00
25 Security Articles	11,135.0 *p*=0.187	−24,638.34 *p*=0.84	−20,465.87 *p*=0.12	No *F*=0.45
34 Infrastructure Articles	**−11,079**** *p*=0.00	−12,721.0 *p*=0.77	2,431.84 *p*=0.61	**Yes** *F*=0.00
26 Popular Articles	−48,458 *p*=0.798	−1,716,643 *p*=0.53	177,324.7 *p*=0.551	No *F*=0.79

Statistically significant findings in bold (*p<0.05, **p<0.01).

41. The first row of Figure 11 shows the results of the Penney Model for the 47 Terrorism articles. The first column shows a statistically significant upward trend prior to June 2013 for that group. The next row shows the results for the first of the three comparator groups that the Penney Declaration analyzed, the 25 Security articles, and shows no statistically significant trend prior to June 2013. This means that there was no possible reversal that could have occurred around June 2013, making the comparison group of Security articles inappropriate and conclusions based on its use incorrect. The second comparator group, the 34 Infrastructure articles, shown in the third row, shows a statistically significant decline prior to June 2013, indicating that the trend for this comparator group was the opposite of the Terrorism articles and, once again, inappropriate as a comparator group. The final group, of 26 Popular articles, shows no statistically significant trend prior to June 2013, and thus this final group is also inappropriate to use as a comparator group.

42. In summary, none of the three datasets of comparator articles that the Penney Declaration analyzes is an appropriate comparator because none of them exhibits the trend prior to June 2013 that the Penney Declaration posits is indicated by the aggregated data of the Terrorism articles.

43. The Penney Declaration also considers two other datasets, one of global Wikipedia homepage views and one of the same data without mobile data.[24] Both of these datasets show an increase through June 2013 followed by a decline after June 2013.[25] In other words, the Penney Model finds an effect at June 2013 for these two comparison datasets even though his theory is that the page views for these two comparison datasets should not have been affected by the June 2013 disclosures. The Penney Declaration attempts to explain away or minimize this effect by explaining that the effect is smaller for global

[24] These are the datasets identified as Global 1 and Global 2 in paragraph 8, above.
[25] Both show an upward trend prior to June 2013. One shows both the immediate and trend change to be statistically significant and one shows only the immediate change to be statistically significant.

views.[26] However, like the three other comparator datasets, the trend prior to June 2013 is also different for these comparator datasets, and thus there is no reason to expect the trend or immediate change would be the same after June 2013. In other words, these datasets are also poor and inappropriate controls.

44. Furthermore, like the page views for the terror-related articles, the page views for the comparison articles vary substantially from one another, not simply in overall number of views but importantly in their trends over time. Graphs of page views for each article used in the comparison datasets, which I provide in Appendix VII, clearly show that among the control articles trends in page views are vastly different. In other words, to the extent that some of the controls might be appropriate, they would need to be used individually (and not in aggregate) and individual factors affecting page views would need to be accounted for, as I explain below.

45. As with the terror-related articles, and as I will explain in detail in the following section, the Penney Model is a flawed and oversimplified model that does not account for any individual differences in page views, and instead assumes the only differences and changes are due to the June 2013 disclosures.

46. In summary, the five comparator datasets used in the Penney Declaration do not support the Penney Declaration conclusions. The three datasets of article page views all have different trends prior to the June 2013 disclosures, making them inappropriate for comparison. The two Wikipedia homepage datasets have a statistically significant trend upward prior to June 2013, but the peak occurs prior to May 2013 and does not correspond to the trend in the terror article views prior to June 2013. This fact means these articles are also not appropriate controls.

F. The Penney Model Estimates are Deeply Flawed, Inappropriate and Likely Biased

47. As explained above, there is no indication of either an abrupt drop in monthly page views of the terror-related articles or an abrupt reversal in an upward trend in views of such articles beginning in June 2013. However, two of the Penney Model estimates are statistically significant, and this statistical significance forms the basis for the Penney Declaration's conclusions.[27] How is it, then, that a simple examination of the data shows no abrupt change or reversal, but two of the Penney Model estimates show a statistically significant change and reversal? The reason is that a deeply flawed model gives deeply flawed results. Because the Penney Model divides the data around an assumed inflection point, it forces the assumption that all changes in page views, beyond a simple trend line, that occurred after that point are caused by the June 2013 disclosures. This flawed assumption drives the spurious statistical significance and other incorrect results. I explain the flaws of the Penney Model in detail below.

48. **The first flaw** in the Penney Model is that the model aggregates the data, and this aggregation masks the differences in the changes in views over time by article. The

[26] As with the terror datasets, the decline actually begins *before* the hypothesized month of June 2013.
[27] Penney Declaration, paragraph 11.

Penney Declaration did not explore whether the claimed reversal in trend existed for each article, and did not explore whether it occurred at the same time, if it occurred at all. Review of the simple graphs of each of the Terror 48 articles, which I provide in Appendix IV (I show four of them in Figure 2), clearly indicates that the trend of page views and their changes over time are not the same for each article. This means that aggregating the data for a single model is inappropriate.

49. As explained earlier, only 2 of the 48 articles' page views peak as hypothesized (in May 2013). Thirty-five of 48 (73%) reach their page view peak earlier than May. In other words the steady march upward followed by an abrupt drop in June 2013 and a steady march downward is a fiction created partly by aggregation of the data.

50. This aggregation is performed without any analysis of the individual datasets to determine whether such aggregation is appropriate. The page views for the 48 articles is an example of what is called "panel data" (in this case the 32 months of page views for each article consists of a panel). Because each of the panels may be different over time, and the panels may be related to one another, a statistical analysis that lumps them together can produce spurious results, as it does in this case.[28] A proper analysis could have used the data for the 48 articles and accounted for the potential effects of specific news events and other influences on each article's page views. There are standard methods for analyzing this kind of panel data but the Penney Model ignores them. [29] Furthermore, as explained in the next paragraphs, even ignoring the differences in the articles and aggregating the data, there is still no indication that the peak is in the hypothesized month of May 2013.

51. **The second flaw** is that the Penney Model assumes a single peak in May 2013 rather than letting the data reveal where, if anywhere, a peak in the data exists.[30] In other words, the Penney Model does not allow for a test of the timing of the change in page views but instead simply assumes that the one and only trend change occurred in June 2013. As a result, the regression model will detect an effect in June 2013 if the period prior to June 2013 generally had increasing page views and the period after generally had declining views, regardless of when the change actually began. That is, even if the change in trend and the decline began *before* the June 2013 disclosures (as it did for 73% of the subject articles, see paragraph 12, above), the Penney Model will find that the disclosures caused them.

52. This model deficiency explains why, despite the aggregate data hitting a peak in April 2013 and not the hypothesized May 2013, the Penney Model indicates the peak was in May 2013 (and the trend reversed starting in June 2013). If I alter the Penney Model to check for an April peak (and a reversal of trend in May instead of June), the altered model "proves" the April peak and trend reversal in May.[31] Thus, for example, the

---

[28] Certain events may cause a change to multiple articles. For example, the rise in views for both "Jihad" and "ammonium nitrate" occurred at the time of the Boston bombings, as I detail below.

[29] For example, see Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, 2012, South-Western Cengage Learning, p. 459-474.

[30] The model also does not allow for there to be multiple peaks in the data.

[31] This is also true when checking for trend reversal in April 2013. The output from these alternative models is contained in the appendix. I do not consider the Penney Model or any of these models appropriate, because they do

alternate (and opposing) theory that the Boston Marathon bombings (which occurred in April 2013) caused the trend reversal beginning in May is also "proven" using the Penney Model.

53. A simple method of checking for the timing of a reversal is possible using what is called a polynomial model. Such a method is common for determining whether and when a trend changes direction (from increasing to decreasing and vice-versa). For reasons outlined below, this simple model, like the Penney Model, is far from adequate and does little to account for the changes in page views.[32] I simply use it to demonstrate that had the Penney Declaration estimated the timing of the reversal in trend in aggregate page views in even this simple fashion, it would not have found that it occurred beginning in June 2013.

54. A polynomial model estimates that views of the Terror 48 article peaked in September 2012; that views of the Terror 48 without Hamas article peaked in November 2012; and that views of the Terror 31 articles peaked in March of 2013. In other words, contrary to the Penney Declaration theory, a model that is forced to select a single peak does not estimate that peak to be the month hypothesized by the Penney Model.

55. **The third flaw** is that the Penney Model is oversimplified, leaving out virtually all factors that could affect page views of terror-related articles from the model. The only factors in the model are a simple trend over time and a single hypothesized cause for the change in June 2013. This means that to the extent that page views change due to factors other than the June 2013 disclosures, those unidentified factors and their concomitant effects on page views will be inappropriately incorporated into the estimates of trend reversal. For example, the Penney Model fails to account for seasonality or major news events that may have affected page views.[33]

56. Such an over-simplified model suffers from what is called "omitted variable bias" and means that the conclusions may be wrong because estimates from the model are biased.[34] This problem means the true effect of the June 2013 disclosures may be non-existent or in the opposite direction of the effect as estimated by the flawed model.[35]

---

not account for seasonality or any other factors (as I explain later). However, the fact that a statistically significant trend reversal can also be found in April and May indicates that the hypothesis that such a change occurred specifically in June 2013 is in no way proven by the Penney Model, even if one assumes that a model with a single change in trend is correct.

[32] For example, it only allows for one change in trend and it does not allow for any effects due to things like world events relevant to individual articles (except for those related to the Hamas article) or seasonality, *see* paragraphs 56-61, below.

[33] Although the Penney Declaration correctly states (in paragraph 26) that the time period is long enough that one could control for seasonality (e.g., lower page views in the summer than at other times of the year), it is barely so, and in any case the Penney Model does not actually attempt to account for any seasonality. This means that the differing number of summer and winter months in the pre-June 2013 and post-June 2013 analysis will affect the results, for example. For some of the regressions, the Penney Model controls what is called "first-order serial autocorrelation," but this correction does not address seasonality.

[34] See, for example, Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, South-Western Cengage Learning, p. 88-91.

[35] For an example of this, see Gujarati, Damodar N., Basic Econometrics, 3rd Edition, McGraw-Hill, 1995, p. 204-207.

57. To demonstrate that there are changes that are not accounted for in the model, I determined if page views dropped during the summer months. In order to check this, I used data from all 48 articles. Therefore, I had a total 1,536 data points, consisting of 32 months, from January 2012 to August 2014, for each article multiplied by 48 articles. The results of my analysis indicate a large and statistically significant reduction in page views in the summer months.[36]

58. Because six of the 15 months considered in Penney's Model are summer months in the period after May 2013 (June 2013 through August 2014), but only three of 17 months are summer months in the period considered before June 2013 (January 2012 through May 2013), a failure to account for the reduction of page views in the summer months means the estimate of an immediate drop and reversal in trend will be overstated in a model like the Penney Model that does not take season into account. As I stated above, the seasonality effect is just one example of a factor that is not accounted for in the Penney Model and is not meant to be exhaustive of the many potential model omissions.

59. The Penney Declaration tacitly acknowledges the fact that it mostly ignores factors affecting page views by excluding the Hamas article from some of its analysis. The reason given for excluding Hamas is that conflicts with Israel occurred in two of the months at-issue and greatly changed page views.[37] While this logically makes sense, the model made no adjustments for any of the other world events occurring during the period of study. The exclusion of the Hamas articles manipulates the data in a way that is favorable to the hypothesis in the Penney Declaration without apparently considering items that may not be favorable.

60. For example, the Boston Marathon bombing occurred two months before the Snowden disclosures, and there was a substantial increase in page views for certain articles. Page views for "Jihad" more than doubled between April and May 2013, from below 100,000 views to above 200,000 views, and page views for Ammonium nitrate (the chemical compound reportedly used in the bomb) had similarly dramatic changes. These dramatic changes corresponding to the Boston bombings were short-term, and, within a month or two, the number of views dropped. Because the Boston bombings occurred prior to June 2013 and are otherwise not accounted for, the increase in page views around April 2013 is improperly incorporated into the estimated "chilling effect" of the June 2013 disclosures by the Penney Model.

61. **The fourth flaw** in the Penney Model is that the 48 terror articles were chosen by Dr. Penney based on their use of terms contained on a 2011 Department of Homeland Security list of terrorism-related terms, and the Model did not take into account that a natural rise or decline in user interest in the topics covered by those articles may occur over time. This could mean that some articles and topics have become less important

[36] Results are in the attached programming log. In order to allow the articles to be comparable despite having different page views, I ranked each article's monthly page views from 1 (lowest) to 32 (highest) prior to performing my analysis. Note that these results do not take into account other factors and therefore the decline in the summer months may be due to particular news events that did or did not occur during those months, for example.
[37] See paragraph 42 of the Penney Declaration.

over time, which could account for a decrease in the number of page views. Also, public interest could shift to newer topics or articles regarding terrorism.

62. I note that while the top few articles in terms of page views were articles about countries, none of the articles in the Terror 48 dataset was about Syria, whose civil war has had an increased news profile over the years. Page views on the article for Syria have averaged nearly 300,000 per month since July 2015, a higher number of views than 47 of the 48 articles explored in the study.[38]

63. Articles about Al Qaeda were included but articles about the Islamic State (including ISIS and ISIL) were not included among the terrorism-related articles considered in the Penney Model. Page views for ISIL (Islamic State of Iraq and the Levant) have averaged more than 600,000 per month since July 2015, higher than any of the 48 articles explored in the Penney Declaration.[39] In short, topics identified in a 2011 list of terrorism related keywords do not necessarily correspond to highly viewed terrorism-related articles during the period of the study or thereafter, and a decline of any static list of articles over time may be expected as "hot" topics change over time.

64. A dramatic demonstration of this issue is the article "Deaths in 2012," which is one of the popular articles used as a control in the Penney Declaration.[40] The page views for this article hovers around 2 million from January through December of 2012 and then quickly drop to nearly zero (for a graph of page views of this article, see Appendix VI). While not necessarily behaving as dramatically as page views for this article, many of the 2011 terrorism-related keywords undoubtably became stale over time, and, subsequently, page views dropped. Such declines have nothing to do with the June 2013 disclosures but are deemed an effect of the June 2013 disclosures by the Penney Model.

65. **The fifth flaw** in the Penney Model relates to the data examined. The data examined only include the 32 months through August of 2014. There is no analysis of any data beyond that date. Therefore, the Penney Model results do not and cannot imply that an effect of the June 2013 disclosures persists today, or did so even in 2015. As I explain above, my own analysis of more recent data shows that page views of the Terror 48 articles are not substantially different than they were prior to June 2013. In addition, changes in the focus of terrorism would mean that some of the articles are less relevant and other articles, not examined at all, are more relevant to the question of whether the Upstream program has a continued chilling effect. This is left unexamined in the Penney Declaration.

66. **The sixth flaw** in the Penney Model is that it fails to isolate the particular effect of public "awareness" about the NSA Upstream program challenged in this suit from the potential effects of, e.g., a) Snowden disclosures about other NSA surveillance activities; b) possible inaccuracies, if any, reported about the Upstream program in the press; c) the Snowden disclosures about British intelligence activities; and d) other events of June

---

[38] Page views found at https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2015-07&end=2018-11&pages=Syria.
[39] See https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2015-07&end=2018-11&pages=Islamic_State_of_Iraq_and_the_Levant .
[40] Penney Declaration, Table 16.

2013. In other words, even if we accept the claim that a chilling effect occurred in June 2013 (and there is no evidence of such an effect), there are no data or statistical analysis offered that indicate such an effect was due to awareness of the specific NSA program at issue here rather than other related or unrelated events of June 2013.

V. Conclusions

67. The Penney Declaration hypothesizes that a chilling effect from the Snowden disclosures caused page views of certain terrorism-related[41] Wikipedia articles to decline beginning in June 2013 and concludes that the Penney Model results regarding page views of these articles are evidence of the decline.

68. My analysis of those articles shows that the Penney Declaration conclusion is wrong. The mistaken conclusion can be observed by performing a simple analysis of the articles' page views and observing that a decline in page views, when it occurred, generally occurred before the disclosures and almost never occurred beginning in the hypothesized month of June 2013. This fact is seen in both the individual and aggregate data.

69. Comparison datasets that are used as controls in the Penney Declaration display different trends prior to 2013, and therefore are inappropriate as control data. Furthermore, as with the terrorism-related articles, the Penney Model inappropriately aggregates articles that have different trends in these comparison datasets.

70. Even assuming that page views of terrorism-related articles fell, as hypothesized, in the data analyzed, the Penney Declaration analyzes data only through August of 2014. Additional data I analyzed, which run through November 2018, indicate that any declines, which in any case began before June 2013, were relatively short-lived.

71. At the root of the mistaken conclusion in the Penney Declaration is a deeply flawed model that aggregates the data and ignores every possible reason for changes in page views except the June 2013 disclosures that concerned Upstream. This means that all changes in page views are presumed to be part of the effects of the disclosures by the Penney Model, no matter what the underlying reason for the page view changes.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Executed in New York, New York, on February 14, 2019.

Alan J. Salzberg

[41] Penney Declaration, paragraph 31.

APPENDIX I: Programming Code

The following is a Stata (Version 14) program and log, used to analyze the data.

This is the program:

```
clear
capture log close
log using readandreplicate_20190115.log, replace
use Penney_regression_data.dta

* note that for July 2015 and beyond:
* terror - now fear
* weapons grade is - now weapons grade nuclear material but didnt exist until
June 2017 even as weapons gade nuclear maerials
* Euskadi ta Askatasuna - now ETA (separatist group)
* pirates is - now piracy
* islamist is - now islamism
* recruitment and fundmanetalism have same data in all but 2 of first 32
months--a clear error

*
rename date viewsdate
rename time monthindex
gen date1=date(viewsdate,"MDY")
format date1 %d
gen month1=month(date1)
gen year1=year(date1)
*
* rename for shorter names
rename terrorarticles48 art_Terror_48
rename terrorarticles47 art_Terror_47
rename globalmilnonmobileraw art_Global1
rename terror31higherprivacy  art_Terror_31
rename securityarticles25comparator  art_Security
rename popular‌articlescomparator  art_Popular
rename infrastructurecomparatorfinal art_Infrastructure
rename globalviewsmilcombined  art_Global2
*
* now index by pct change from median
* and replicate original regressions
foreach var1 of varlist art_*  {
* egen rk_`var1' = rank(`var1')
display "========="
display "`var1'"
display "==========="
regress `var1' monthindex intervention postslope
}
* table 8 replication
regress art_Terror_31 monthindex intervention postslope art_Global1
```

```
* table 9 replication
regress art_Terror_47 monthindex intervention postslope art_Global1
* control regs
regress art_Global2 monthindex intervention postslope

* show that may and april also stat signif
gen interventionmay=intervention
replace interventionmay=1 if monthindex==17
gen postslopemay=postslope
replace postslopemay=postslope+1 if interventionmay==1
gen interventionapril=interventionmay
replace interventionapril=1 if monthindex==16
gen postslopeapril=postslopemay
replace postslopeapril=postslopeapril + 1 if interventionapril==1
list monthindex postslope postslopeapril postslopemay intervention
interventionapril interventionmay
*
* estimate turning point (estimated peak of data)
gen idx2=monthindex^2
regress art_Terror_48 monthindex idx2
predict tmp48
egen max48=max(tmp48)
list viewsdate monthindex if tmp48==max48

regress art_Terror_47 monthindex idx2
predict tmp47
egen max47=max(tmp47)
list viewsdate monthindex if tmp47==max47

regress art_Terror_31 monthindex idx2
predict tmp31
egen max31=max(tmp31)
list viewsdate monthindex if tmp31==max31

drop tmp31 tmp47 tmp48 max31 max47 max48

*
regress art_Terror_31 monthindex intervention postslope
regress art_Terror_31 monthindex interventionmay postslopemay
regress art_Terror_31 monthindex interventionapril postslopeapril

regress art_Terror_47 monthindex intervention postslope
regress art_Terror_47 monthindex interventionmay postslopemay
regress art_Terror_47 monthindex interventionapril postslopeapril

regress art_Terror_47 monthindex intervention postslope
regress art_Terror_47 monthindex interventionmay postslopemay
regress art_Terror_47 monthindex interventionapril postslopeapril
```

```
reshape long art_, i( monthindex date1 month1 year1 intervention postslope)
j(artnmshort) string
rename art_ pageviews
format pageviews %12.0f
egen rankviews=rank(pageviews), by(artnmshort)
 gen yearmonth1=year*100+month1
* most groups peaked in earlier period (not unique to terror articles) and no
group peaked in May 2013 (just before claimed intervention)
list year1 month1 artnmshort if rankviews==32
* trough
list year1 month1 artnmshort if rankviews==1

*
* write out to csv file in order to produce graphs
outsheet using articlesaggregate.csv, comma replace

*****************************
* replicate control regressions
*****************************
clear
use security25
regress sum_view monthindex postslope intervention
outsheet using security25.csv, comma replace

use  infrastructure34
regress sum_view monthindex postslope intervention
outsheet using infrastructure34.csv, comma replace

use popular26
regress sum_view monthindex postslope intervention
outsheet using popular26.csv, comma replace

clear

*******************************
* now use with individual 48
*******************************
clear
use artterror48_origplusrecentdates.dta
gen date1=date(dateorig,"MDY")
gen month1=month(date1)
gen year1=year(date1)
sort date1
gen monthindex=_n
* account for skipped 11 months
replace monthindex = monthindex + 10 if year>=2015
gen intervention=1
replace intervention=0 if date1<date("06/01/2013","MDY")
gen postslope = (monthindex-17)*intervention
```

```
egen totview=rowtotal(art_t*)

* check first regression again
regress totview monthindex postslope intervention if year<=2014
gen totviewminushamas=totview - art_t22
gen totviewminusdup=totview - art_t47
regress totviewminushamas monthindex postslope intervention if year1<=2014
*
regress totviewminusdup monthindex postslope intervention if year1<=2014

*
* now drop totals and reshape
drop totv*
* obvious error in articles on Recruitment and fundamentalism (all numbers
but last couple are the same)
count if art_t46==art_t47

reshape long art_t, i( monthindex date1 month1 year1 intervention postslope)
j(artnum)
*
rename art_t pageviews

* pull in article names
sort artnum
merge m:1 artnum using articlenames48
assert _merge==3
drop _merge
* normalize names for better display and read/write
replace artnames=subinstr(artnames,"(","_",.)
replace artnames=subinstr(artnames,")","_",.)
replace artnames=subinstr(artnames," ","_",.)
replace artnames=subinstr(artnames,"+","_",.)
replace artnames=subinstr(artnames,"-","_",.)
replace artnames=subinstr(artnames,"__","_",.)
replace artnames=subinstr(artnames,"__","_",.)
replace artnames=subinstr(artnames,"__","_",.)

* pull in indicator of whether article was high privacy
sort artnum
merge m:1 artnum using highprivacy31
gen highprivind=_merge==3
assert _merge!=2
drop _merge
*
* indicate 7 articles with issues between early and late period
gen lateissueind=0
replace lateissueind=1 if artname=="terror"
replace lateissueind=1 if artname=="Weapons_grade"
replace lateissueind=1 if artname=="_Euskadi_ta_Askatasuna"
```

```
replace lateissueind=1 if artname=="Pirates"
replace lateissueind=1 if artname=="Islamist"
replace lateissueind=1 if artname=="Recruitment"
replace lateissueind=1 if artname=="Fundamentalism"



* check that high privacy desig is ok by checking reg of sum
egen totview31=sum(pageviews), by(monthindex highprivind)
replace totview31=. if highprivind==0
bysort monthindex highprivind: gen tmpindx=_n
regress totview31 monthindex postslope intervention if tmpindx==1 &
year1<=2014
drop tmpindx
*
* get ranks of first 17, first 32 and all
gen pageviewall=pageviews
gen pageviews17=pageviews
replace pageviews=. if year>2014
replace pageviews17=. if monthindex>=18
egen rankviewsearly=rank(pageviews), by(artnum)
egen maxrankearly=max(rankviewsearly), by(artnum)
egen rankviews17=rank(pageviews17), by(artnum)
egen maxrank17=max(rankviews17), by(artnum)
egen rankviewsall=rank(pageviewall), by(artnum)
egen maxrankall=max(rankviewsall), by(artnum)

sum maxr*
sum rankv*
sort artnum date1

*
gen yearmonth=year1*100 + month1
* summermonths lower in general --inidcation of seasonality
* use rank so all data can be considered on a like to like basis
 table month1, c(mean rankviewsearly median rankviewsearly mean rankviewsall
median rankviewsall n rankviewsall) row format(%6.2f)
 table month1, c(mean rankviewsearly median rankviewsearly mean rankviewsall
median rankviewsall n rankviewsall) row format(%6.2f)
regress rankviewsall i.month1 if lateissueind==0
regress rankviewsall i.month1 if monthindex<=32

 * where is maximum?
 tab yearmonth highpriv  if rankviewsearly==maxrankearly
 tab yearmonth highpriv  if rankviewsall==maxrankall

* output to csv for graphics and other analysis
gen dateformat=date1
format dateformat %d
```

```
outsheet using orig48long.csv, comma replace
*
log close
```

This is the program log:

```
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\readandreplicate_20190115.log
  log type:  text
 opened on:  15 Jan 2019, 18:07:38

. use Penney_regression_data.dta

.
. * note that for July 2015 and beyond:
. * terror - now fear
. * weapons grade is - now weapons grade nuclear material but didnt exist
until June 2017 even as weapons gade nuclear maer
> ials
. * Euskadi ta Askatasuna - now ETA (separatist group)
. * pirates is - now piracy
. * islamist is - now islamism
. * recruitment and fundmanetalism have same data in all but 2 of first 32
months--a clear error
.
. *
. rename date viewsdate

. rename time monthindex

. gen date1=date(viewsdate,"MDY")

. format date1 %d

. gen month1=month(date1)

. gen year1=year(date1)

. *
. * rename for shorter names
. rename terrorarticles48 art_Terror_48

. rename terrorarticles47 art_Terror_47

. rename globalmilnonmobileraw art_Global1

. rename terror31higherprivacy  art_Terror_31
```

```
. rename securityarticles25comparator  art_Security

. rename popular articlescomparator  art_Popular

. rename infrastructurecomparatorfinal art_Infrastructure

. rename globalviewsmilcombined  art_Global2

. *
. * now index by pct change from median
. * and replicate original regressions
. foreach var1 of varlist art_*  {
  2. * egen rk_`var1' = rank(`var1')
. display "========="
  3. display "`var1'"
  4. display "==========="
  5. regress `var1' monthindex intervention postslope
  6. }
=========
art_Terror_48
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
9.16
       Model |  3.1498e+12         3  1.0499e+12   Prob > F        =
0.0002
    Residual |  3.2091e+12        28  1.1461e+11   R-squared       =
0.4953
-------------+----------------------------------   Adj R-squared   =
0.4413
       Total |  6.3590e+12        31  2.0513e+11   Root MSE        =
3.4e+05

------------------------------------------------------------------------------
-
art_Terro~48 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   47038.28   16760.41     2.81   0.009     12706.13
81370.43
intervention |  -995085.2   241987.6    -4.11   0.000     -1490774   -
499396.1
   postslope |  -35517.69   26272.41    -1.35   0.187    -89334.29
18298.91
```

```
       _cons |    2352364   171743.1    13.70   0.000      2000564
2704164
------------------------------------------------------------------------------
-
=========
art_Terror_47
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
24.85
       Model |  3.4887e+12         3  1.1629e+12   Prob > F        =
0.0000
    Residual |  1.3105e+12        28  4.6805e+10   R-squared       =
0.7269
-------------+----------------------------------   Adj R-squared   =
0.6977
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
2.2e+05

------------------------------------------------------------------------------
-
art_Terro~47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   41420.51   10710.65     3.87   0.001     19480.73
63360.29
intervention |  -693616.9   154640.9    -4.49   0.000     -1010384   -
376849.4
   postslope |   -67513.1   16789.25    -4.02   0.000    -101904.3   -
33121.89
       _cons |    2289153   109751.5    20.86   0.000      2064337
2513968
------------------------------------------------------------------------------
-
=========
art_Global2
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
10.06
       Model |   6663270.2         3  2221090.07   Prob > F        =
0.0001
    Residual |   6180561.8        28   220734.35   R-squared       =
0.5188
```

```
-------------+----------------------------------   Adj R-squared    =
0.4672
       Total |    12843832        31  414317.161   Root MSE         =
469.82

------------------------------------------------------------------------------
-
 art_Global2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   114.3824   23.25974     4.92   0.000     66.73693
162.0278
intervention |  -1535.819   335.8252    -4.57   0.000    -2223.726   -
847.9123
   postslope |  -46.97164   36.46029    -1.29   0.208    -121.6572
27.71387
       _cons |     8313.5   238.3414    34.88   0.000      7825.28
8801.72
------------------------------------------------------------------------------
-
=========
art_Terror_31
===========

      Source |       SS           df       MS      Number of obs    =
32
-------------+----------------------------------   F(3, 28)         =
20.87
       Model |  5.1404e+11         3  1.7135e+11   Prob > F         =
0.0000
    Residual |  2.2989e+11        28  8.2102e+09   R-squared        =
0.6910
-------------+----------------------------------   Adj R-squared    =
0.6579
       Total |  7.4392e+11        31  2.3998e+10   Root MSE         =
90610

------------------------------------------------------------------------------
-
art_Terro~31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   28484.13   4485.873     6.35   0.000     19295.24
37673.02
intervention |  -253556.5   64767.24    -3.91   0.001    -386226.2   -
120886.9
```

```
   postslope |  -41554.21    7031.73    -5.91   0.000    -55958.05   -
27150.36
       _cons |   471146.3   45966.52    10.25   0.000     376988.2
565304.5
------------------------------------------------------------------------------
-
=========
art_Security
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
0.91
       Model |  7.5795e+10         3  2.5265e+10   Prob > F        =
0.4470
    Residual |  7.7441e+11        28  2.7657e+10   R-squared       =
0.0891
-------------+----------------------------------   Adj R-squared   =   -
0.0084
       Total |  8.5020e+11        31  2.7426e+10   Root MSE        =
1.7e+05

------------------------------------------------------------------------------
-
art_Security |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   11135.07   8233.343     1.35   0.187     -5730.17
28000.31
intervention |  -24638.34   118873.4    -0.21   0.837    -268139.4
218862.7
   postslope |  -20465.87   12905.99    -1.59   0.124     -46902.6
5970.859
       _cons |   708187.4   84366.66     8.39   0.000     535370.2
881004.7
------------------------------------------------------------------------------
-
=========
art_Popular
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
0.34
       Model |  1.4789e+13         3  4.9297e+12   Prob > F        =
0.7938
```

```
    Residual |  4.0134e+14        28  1.4334e+13   R-squared       =
0.0355
-------------+----------------------------------   Adj R-squared   =   -
0.0678
       Total |  4.1613e+14        31  1.3424e+13   Root MSE        =
3.8e+06

------------------------------------------------------------------------------
-
 art_Popular |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |  -48458.14   187433.7    -0.26   0.798    -432398.7
335482.5
intervention |   -1716643    2706177    -0.63   0.531     -7259994
3826709
   postslope |   177324.7   293807.6     0.60   0.551    -424512.8
779162.2
       _cons |   2.58e+07    1920624    13.41   0.000     2.18e+07
2.97e+07
------------------------------------------------------------------------------
-
=========
art_Infrastructure
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
27.12
       Model |  3.0280e+11         3  1.0093e+11   Prob > F        =
0.0000
    Residual |  1.0421e+11        28  3.7218e+09   R-squared       =
0.7440
-------------+----------------------------------   Adj R-squared   =
0.7165
       Total |  4.0701e+11        31  1.3129e+10   Root MSE        =
61007

------------------------------------------------------------------------------
-
art_Infras~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |  -11079.82   3020.285    -3.67   0.001    -17266.59   -
4893.042
```

```
intervention |  -12721.07   43607.01    -0.29   0.773      -102046
76603.85
   postslope |   2431.841   4734.381     0.51   0.612    -7266.098
12129.78
       _cons |   771772.3   30948.71    24.94   0.000     708376.8
835167.9
------------------------------------------------------------------------------
-
=========
art_Global1
===========

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
20.64
       Model |  10062791.9         3  3354263.97   Prob > F        =
0.0000
    Residual |  4549258.31        28  162473.511   R-squared       =
0.6887
-------------+----------------------------------   Adj R-squared   =
0.6553
       Total |  14612050.2        31  471356.459   Root MSE        =
403.08

------------------------------------------------------------------------------
-
 art_Global1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   70.57598   19.95544     3.54   0.001     29.69912
111.4528
intervention |  -1397.969   288.1175    -4.85   0.000    -1988.151   -
807.7867
   postslope |  -90.97598    31.2807    -2.91   0.007    -155.0516   -
26.90038
       _cons |    7385.11   204.4824    36.12   0.000     6966.247
7803.973
------------------------------------------------------------------------------
-

. * table 8 replication
. regress art_Terror_31 monthindex intervention postslope art_Global1

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(4, 27)        =
16.30
```

```
       Model |  5.2604e+11         4  1.3151e+11   Prob > F        =
0.0000
    Residual |  2.1789e+11        27  8.0700e+09   R-squared       =
0.7071
-------------+----------------------------------   Adj R-squared   =
0.6637
       Total |  7.4392e+11        31  2.3998e+10   Root MSE        =
89833

------------------------------------------------------------------------------
-
art_Terro~31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   32108.35   5349.312     6.00   0.000     21132.46
43084.23
intervention |    -325345   87120.19    -3.73   0.001    -504100.9   -
146589.1
   postslope |  -46226.01   7955.041    -5.81   0.000     -62548.4   -
29903.61
 art_Global1 |  -51.35198   42.11781    -1.22   0.233    -137.7706
35.06662
       _cons |   850386.4   314365.4     2.71   0.012     205361.8
1495411
------------------------------------------------------------------------------
-

. * table 9 replication
. regress art_Terror_47 monthindex intervention postslope art_Global1

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(4, 27)        =
18.49
       Model |  3.5157e+12         4  8.7893e+11   Prob > F        =
0.0000
    Residual |  1.2835e+12        27  4.7538e+10   R-squared       =
0.7326
-------------+----------------------------------   Adj R-squared   =
0.6929
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
2.2e+05

------------------------------------------------------------------------------
-
art_Terro~47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
```

```
-------------+----------------------------------------------------------------
-
  monthindex |   35983.25   12983.28     2.77   0.010     9343.768
62622.74
intervention |  -585915.8   211448.8    -2.77   0.010     -1019773   -
152058.7
   postslope |   -60504.2   19307.63    -3.13   0.004    -100120.2   -
20888.23
 art_Global1 |   77.04117   102.2238     0.75   0.458    -132.7048
286.7872
       _cons |    1720195   762994.1     2.25   0.032     154660.4
3285730
------------------------------------------------------------------------------
-

. * control regs
. regress art_Global2 monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
10.06
       Model |   6663270.2         3  2221090.07   Prob > F        =
0.0001
    Residual |   6180561.8        28   220734.35   R-squared       =
0.5188
-------------+----------------------------------   Adj R-squared   =
0.4672
       Total |    12843832        31  414317.161   Root MSE        =
469.82

------------------------------------------------------------------------------
-
 art_Global2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   114.3824   23.25974     4.92   0.000     66.73693
162.0278
intervention |  -1535.819   335.8252    -4.57   0.000    -2223.726   -
847.9123
   postslope |  -46.97164   36.46029    -1.29   0.208    -121.6572
27.71387
       _cons |     8313.5   238.3414    34.88   0.000      7825.28
8801.72
------------------------------------------------------------------------------
-

.
```

```
. * show that may and april also stat signif
. gen interventionmay=intervention

. replace interventionmay=1 if monthindex==17
(1 real change made)

. gen postslopemay=postslope

. replace postslopemay=postslope+1 if interventionmay==1
(16 real changes made)

. gen interventionapril=interventionmay

. replace interventionapril=1 if monthindex==16
(1 real change made)

. gen postslopeapril=postslopemay

. replace postslopeapril=postslopeapril + 1 if interventionapril==1
(17 real changes made)

. list monthindex postslope postslopeapril postslopemay intervention
interventionapril interventionmay

     +-------------------------------------------------------------------------
-----+
     | monthi~x   postsl~e   postsl~l   postsl~y   interv~n   interv~l
interv~y |
     |-------------------------------------------------------------------------
-----|
  1. |        1          0          0          0          0          0
0 |
  2. |        2          0          0          0          0          0
0 |
  3. |        3          0          0          0          0          0
0 |
  4. |        4          0          0          0          0          0
0 |
  5. |        5          0          0          0          0          0
0 |
     |-------------------------------------------------------------------------
-----|
  6. |        6          0          0          0          0          0
0 |
  7. |        7          0          0          0          0          0
0 |
  8. |        8          0          0          0          0          0
0 |
```

9.	9	0	0	0	0	0
10.	10	0	0	0	0	0
11.	11	0	0	0	0	0
12.	12	0	0	0	0	0
13.	13	0	0	0	0	0
14.	14	0	0	0	0	0
15.	15	0	0	0	0	0
16.	16	0	1	0	0	1
17.	17	0	2	1	0	1
18.	18	1	3	2	1	1
19.	19	2	4	3	1	1
20.	20	3	5	4	1	1
21.	21	4	6	5	1	1
22.	22	5	7	6	1	1
23.	23	6	8	7	1	1
24.	24	7	9	8	1	1
25.	25	8	10	9	1	1
26.	26	9	11	10	1	1
27.	27	10	12	11	1	1
28.	28	11	13	12	1	1

```
 29. |       29         12         14         13          1          1
1 |
 30. |       30         13         15         14          1          1
1 |
     |-------------------------------------------------------------------
-----|
 31. |       31         14         16         15          1          1
1 |
 32. |       32         15         17         16          1          1
1 |
     +-------------------------------------------------------------------
-----+

. *
. * estimate turning point (estimated peak of data)
. gen idx2=monthindex^2

. regress art_Terror_48 monthindex idx2

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(2, 29)        =
2.60
       Model |  9.6611e+11         2  4.8306e+11   Prob > F        =
0.0917
    Residual |  5.3928e+12        29  1.8596e+11   R-squared       =
0.1519
-------------+----------------------------------   Adj R-squared   =
0.0934
       Total |  6.3590e+12        31  2.0513e+11   Root MSE        =
4.3e+05

------------------------------------------------------------------------------
-
art_Terro~48 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   20575.12   34056.48     0.60   0.550     -49078.2
90228.43
       idx2 |  -1120.311   1001.228    -1.12   0.272    -3168.052
927.4307
       _cons |    2589880   243771.8    10.62   0.000      2091311
3088449
------------------------------------------------------------------------------
-

. predict tmp48
(option xb assumed; fitted values)
```

```
. egen max48=max(tmp48)

. list viewsdate monthindex if tmp48==max48

     +-----------------------+
     |  viewsdate   monthi~x |
     |-----------------------|
  9. | 09/01/2012          9 |
     +-----------------------+

.
. regress art_Terror_47 monthindex idx2

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(2, 29)        =
12.52
       Model |  2.2234e+12         2  1.1117e+12   Prob > F        =
0.0001
    Residual |  2.5758e+12        29  8.8822e+10   R-squared       =
0.4633
-------------+----------------------------------   Adj R-squared   =
0.4263
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
3.0e+05

------------------------------------------------------------------------------
-
art_Terro~47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   43574.63      23537     1.85   0.074     -4563.94
91713.19
        idx2 |  -2022.568   691.9654    -2.92   0.007    -3437.796   -
607.3393
       _cons |    2398370   168474.8    14.24   0.000      2053801
2742940
------------------------------------------------------------------------------
-

. predict tmp47
(option xb assumed; fitted values)

. egen max47=max(tmp47)

. list viewsdate monthindex if tmp47==max47
```

```
     +-----------------------+
     |  viewsdate   monthi~x |
     |-----------------------|
 11. | 11/01/2012         11 |
     +-----------------------+

.
. regress art_Terror_31 monthindex idx2

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(2, 29)        =
9.35
       Model |  2.9173e+11         2  1.4586e+11   Prob > F        =
0.0007
    Residual |  4.5220e+11        29  1.5593e+10   R-squared       =
0.3921
-------------+----------------------------------   Adj R-squared   =
0.3502
       Total |  7.4392e+11        31  2.3998e+10   Root MSE        =
1.2e+05

------------------------------------------------------------------------------
-
art_Terro~31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   36223.88   9861.789     3.67   0.001     16054.26
56393.51
        idx2 |  -1193.715   289.9272    -4.12   0.000    -1786.683   -
600.7469
       _cons |   495510.5    70589.4     7.02   0.000       351139
639882
------------------------------------------------------------------------------
-

. predict tmp31
(option xb assumed; fitted values)

. egen max31=max(tmp31)

. list viewsdate monthindex if tmp31==max31

     +-----------------------+
     |  viewsdate   monthi~x |
     |-----------------------|
 15. | 03/01/2013         15 |
     +-----------------------+
```

```
.
. drop tmp31 tmp47 tmp48 max31 max47 max48

.
. *
. regress art_Terror_31 monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
20.87
       Model |  5.1404e+11         3  1.7135e+11   Prob > F        =
0.0000
    Residual |  2.2989e+11        28  8.2102e+09   R-squared       =
0.6910
-------------+----------------------------------   Adj R-squared   =
0.6579
       Total |  7.4392e+11        31  2.3998e+10   Root MSE        =
90610

------------------------------------------------------------------------------
-
art_Terro~31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   28484.13   4485.873     6.35   0.000     19295.24
37673.02
intervention |  -253556.5   64767.24    -3.91   0.001    -386226.2   -
120886.9
   postslope |  -41554.21    7031.73    -5.91   0.000    -55958.05   -
27150.36
       _cons |   471146.3   45966.52    10.25   0.000     376988.2
565304.5
------------------------------------------------------------------------------
-

. regress art_Terror_31 monthindex interventionmay postslopemay

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
14.66
       Model |  4.5452e+11         3  1.5151e+11   Prob > F        =
0.0000
    Residual |  2.8941e+11        28  1.0336e+10   R-squared       =
0.6110
```

```
-------------+----------------------------------   Adj R-squared    =
0.5693
       Total |  7.4392e+11        31  2.3998e+10   Root MSE         =
1.0e+05

------------------------------------------------------------------------------
  art_Terror_31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
----------------+-------------------------------------------------------------
     monthindex |   27831.07   5513.605     5.05   0.000     16536.96
39125.18
interventionmay |    -135552   72099.74    -1.88   0.071    -283241.6
12137.67
   postslopemay |  -47070.54   7797.415    -6.04   0.000    -63042.82   -
31098.26
          _cons |   475064.7   53314.02     8.91   0.000     365855.8
584273.5
------------------------------------------------------------------------------

. regress art_Terror_31 monthindex interventionapril postslopeapril

      Source |       SS           df       MS      Number of obs    =
32
-------------+----------------------------------   F(3, 28)         =
12.16
       Model |  4.2092e+11         3  1.4031e+11   Prob > F         =
0.0000
    Residual |  3.2300e+11        28  1.1536e+10   R-squared        =
0.5658
-------------+----------------------------------   Adj R-squared    =
0.5193
       Total |  7.4392e+11        31  2.3998e+10   Root MSE         =
1.1e+05

------------------------------------------------------------------------------
    art_Terror_31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
------------------+-----------------------------------------------------------
       monthindex |   19718.72   6418.652     3.07   0.005     6570.704
32866.73
interventionapril |   85936.01   75872.03     1.13   0.267    -69480.79
241352.8
   postslopeapril |  -47183.37   8335.046    -5.66   0.000    -64256.94    -
30109.8
```

```
            _cons |   521034.7   58359.17     8.93   0.000     401491.3
640578
------------------------------------------------------------------------------
------

.
. regress art_Terror_47 monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
24.85
       Model |  3.4887e+12         3  1.1629e+12   Prob > F        =
0.0000
    Residual |  1.3105e+12        28  4.6805e+10   R-squared       =
0.7269
-------------+----------------------------------   Adj R-squared   =
0.6977
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
2.2e+05

------------------------------------------------------------------------------
-
art_Terro~47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   41420.51   10710.65     3.87   0.001     19480.73
63360.29
intervention |  -693616.9   154640.9    -4.49   0.000     -1010384   -
376849.4
   postslope |   -67513.1   16789.25    -4.02   0.000    -101904.3   -
33121.89
       _cons |    2289153   109751.5    20.86   0.000      2064337
2513968
------------------------------------------------------------------------------
-

. regress art_Terror_47 monthindex interventionmay postslopemay

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
19.19
       Model |  3.2291e+12         3  1.0764e+12   Prob > F        =
0.0000
    Residual |  1.5701e+12        28  5.6077e+10   R-squared       =
0.6728
```

```
-------------+----------------------------------   Adj R-squared    =
0.6378
       Total |  4.7992e+12        31  1.5481e+11   Root MSE         =
2.4e+05

------------------------------------------------------------------------------
  art_Terror_47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
----------------+-------------------------------------------------------------
     monthindex |   43914.21   12842.55     3.42   0.002     17607.45
70220.98
interventionmay |  -502573.7   167938.1    -2.99   0.006    -846579.3   -
158568.1
   postslopemay |  -83106.85   18162.11    -4.58   0.000    -120310.2   -
45903.46
          _cons |    2274190   124181.5    18.31   0.000      2019816
2528565
------------------------------------------------------------------------------

. regress art_Terror_47 monthindex interventionapril postslopeapril

      Source |       SS           df       MS      Number of obs    =
32
-------------+----------------------------------   F(3, 28)         =
14.09
       Model |  2.8871e+12         3  9.6236e+11   Prob > F         =
0.0000
    Residual |  1.9122e+12        28  6.8291e+10   R-squared        =
0.6016
-------------+----------------------------------   Adj R-squared    =
0.5589
       Total |  4.7992e+12        31  1.5481e+11   Root MSE         =
2.6e+05

------------------------------------------------------------------------------
    art_Terror_47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
------------------+-----------------------------------------------------------
       monthindex |   37869.78   15617.23     2.42   0.022     5879.338
69860.22
interventionapril |  -195021.8   184604.3    -1.06   0.300    -573166.5
183122.9
   postslopeapril |  -91064.94   20280.01    -4.49   0.000    -132606.7   -
49523.23
```

```
            _cons |    2308442   141993.7    16.26   0.000      2017581
2599303
------------------------------------------------------------------------------
------

.
. regress art_Terror_47 monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
24.85
       Model |  3.4887e+12         3  1.1629e+12   Prob > F        =
0.0000
    Residual |  1.3105e+12        28  4.6805e+10   R-squared       =
0.7269
-------------+----------------------------------   Adj R-squared   =
0.6977
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
2.2e+05

------------------------------------------------------------------------------
-
art_Terro~47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   41420.51   10710.65     3.87   0.001     19480.73
63360.29
intervention |  -693616.9   154640.9    -4.49   0.000     -1010384   -
376849.4
   postslope |   -67513.1   16789.25    -4.02   0.000    -101904.3   -
33121.89
       _cons |    2289153   109751.5    20.86   0.000      2064337
2513968
------------------------------------------------------------------------------
-

. regress art_Terror_47 monthindex interventionmay postslopemay

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
19.19
       Model |  3.2291e+12         3  1.0764e+12   Prob > F        =
0.0000
    Residual |  1.5701e+12        28  5.6077e+10   R-squared       =
0.6728
```

```
-------------+----------------------------------   Adj R-squared    =
0.6378
       Total |  4.7992e+12        31  1.5481e+11   Root MSE         =
2.4e+05

------------------------------------------------------------------------------
  art_Terror_47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
----------------+-------------------------------------------------------------
     monthindex |   43914.21   12842.55     3.42   0.002     17607.45
70220.98
interventionmay |  -502573.7   167938.1    -2.99   0.006    -846579.3   -
158568.1
   postslopemay |  -83106.85   18162.11    -4.58   0.000    -120310.2   -
45903.46
          _cons |    2274190   124181.5    18.31   0.000      2019816
2528565
------------------------------------------------------------------------------

. regress art_Terror_47 monthindex interventionapril postslopeapril

      Source |       SS           df       MS      Number of obs    =
32
-------------+----------------------------------   F(3, 28)         =
14.09
       Model |  2.8871e+12         3  9.6236e+11   Prob > F         =
0.0000
    Residual |  1.9122e+12        28  6.8291e+10   R-squared        =
0.6016
-------------+----------------------------------   Adj R-squared    =
0.5589
       Total |  4.7992e+12        31  1.5481e+11   Root MSE         =
2.6e+05

------------------------------------------------------------------------------
    art_Terror_47 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
------------------+-----------------------------------------------------------
       monthindex |   37869.78   15617.23     2.42   0.022     5879.338
69860.22
interventionapril |  -195021.8   184604.3    -1.06   0.300    -573166.5
183122.9
   postslopeapril |  -91064.94   20280.01    -4.49   0.000    -132606.7   -
49523.23
```

```
            _cons |    2308442   141993.7    16.26   0.000      2017581
2599303
------------------------------------------------------------------------------
------

.
. reshape long art_, i( monthindex date1 month1 year1 intervention postslope)
j(artnmshort) string
(note: j = Global1 Global2 Infrastructure Popular Security Terror_31
Terror_47 Terror_48)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       32   ->     256
Number of variables                  20   ->      14
j variable (8 values)                     ->   artnmshort
xij variables:
art_Global1 art_Global2 ... art_Terror_48 ->   art_
-----------------------------------------------------------------------------

. rename art_ pageviews

. format pageviews %12.0f

. egen rankviews=rank(pageviews), by(artnmshort)

.  gen yearmonth1=year*100+month1

. * most groups peaked in earlier period (not unique to terror articles) and
no group peaked in May 2013 (just before claim
> ed intervention)
. list year1 month1 artnmshort if rankviews==32

     +---------------------------------+
     | year1   month1       artnmshort |
     |---------------------------------|
  3. |  2012        1   Infrastructure |
 52. |  2012        7          Popular |
 88. |  2012       11        Terror_48 |
 97. |  2013        1          Global1 |
 98. |  2013        1          Global2 |
     |---------------------------------|
126. |  2013        4        Terror_31 |
127. |  2013        4        Terror_47 |
181. |  2013       11         Security |
     +---------------------------------+

. * trough
. list year1 month1 artnmshort if rankviews==1
```

```
     +---------------------------------+
     | year1   month1       artnmshort |
     |---------------------------------|
161. |  2013        9          Global1 |
162. |  2013        9          Global2 |
172. |  2013       10          Popular |
198. |  2014        1        Terror_31 |
199. |  2014        1        Terror_47 |
     |---------------------------------|
200. |  2014        1        Terror_48 |
251. |  2014        8   Infrastructure |
253. |  2014        8         Security |
     +---------------------------------+

. 
. * 
. * write out to csv file in order to produce graphs
. outsheet using articlesaggregate.csv, comma replace

. 
. ****************************
. * replicate control regressions
. ****************************
. clear

. use security25

. regress sum_view monthindex postslope intervention

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
0.91
       Model |  7.5795e+10         3  2.5265e+10   Prob > F        =
0.4470
    Residual |  7.7441e+11        28  2.7657e+10   R-squared       =
0.0891
-------------+----------------------------------   Adj R-squared   =   -
0.0084
       Total |  8.5020e+11        31  2.7426e+10   Root MSE        =
1.7e+05

------------------------------------------------------------------------------
-
    sum_view |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
```

```
  monthindex |   11135.07   8233.343     1.35   0.187     -5730.17
28000.31
   postslope |  -20465.87   12905.99    -1.59   0.124     -46902.6
5970.859
intervention |  -24638.34   118873.4    -0.21   0.837    -268139.4
218862.7
       _cons |   708187.4   84366.66     8.39   0.000     535370.2
881004.7
------------------------------------------------------------------------------
-

. outsheet using security25.csv, comma replace

.
. use  infrastructure34

. regress sum_view monthindex postslope intervention

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
27.12
       Model |  3.0280e+11         3  1.0093e+11   Prob > F        =
0.0000
    Residual |  1.0421e+11        28  3.7218e+09   R-squared       =
0.7440
-------------+----------------------------------   Adj R-squared   =
0.7165
       Total |  4.0701e+11        31  1.3129e+10   Root MSE        =
61007

------------------------------------------------------------------------------
-
    sum_view |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |  -11079.82   3020.285    -3.67   0.001    -17266.59   -
4893.042
   postslope |   2431.841   4734.381     0.51   0.612    -7266.098
12129.78
intervention |  -12721.07   43607.01    -0.29   0.773      -102046
76603.85
       _cons |   771772.3   30948.71    24.94   0.000     708376.8
835167.9
------------------------------------------------------------------------------
-

. outsheet using infrastructure34.csv, comma replace
```

```
.
. use popular26

. regress sum_view monthindex postslope intervention

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
0.34
       Model |  1.4789e+13         3  4.9297e+12   Prob > F        =
0.7938
    Residual |  4.0134e+14        28  1.4334e+13   R-squared       =
0.0355
-------------+----------------------------------   Adj R-squared   =   -
0.0678
       Total |  4.1613e+14        31  1.3424e+13   Root MSE        =
3.8e+06

------------------------------------------------------------------------------
-
    sum_view |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |  -48458.14   187433.7    -0.26   0.798    -432398.7
335482.5
   postslope |   177324.7   293807.6     0.60   0.551    -424512.8
779162.2
intervention |   -1716643    2706177    -0.63   0.531     -7259994
3826709
       _cons |   2.58e+07    1920624    13.41   0.000     2.18e+07
2.97e+07
------------------------------------------------------------------------------
-

. outsheet using popular26.csv, comma replace

.
. clear

.
. ******************************
. * now use with individual 48
. ******************************
. clear

. use artterror48_origplusrecentdates.dta
```

```
. gen date1=date(dateorig,"MDY")

. gen month1=month(date1)

. gen year1=year(date1)

. sort date1

. gen monthindex=_n

. * account for skipped 11 months
. replace monthindex = monthindex + 10 if year>=2015
(41 real changes made)

. gen intervention=1

. replace intervention=0 if date1<date("06/01/2013","MDY")
(17 real changes made)

. gen postslope = (monthindex-17)*intervention

. egen totview=rowtotal(art_t*)

.
. * check first regression again
. regress totview monthindex postslope intervention if year<=2014

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
9.16
       Model |  3.1498e+12         3  1.0499e+12   Prob > F        =
0.0002
    Residual |  3.2091e+12        28  1.1461e+11   R-squared       =
0.4953
-------------+----------------------------------   Adj R-squared   =
0.4413
       Total |  6.3590e+12        31  2.0513e+11   Root MSE        =
3.4e+05

------------------------------------------------------------------------------
-
     totview |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   47038.28   16760.41     2.81   0.009     12706.13
81370.43
```

```
   postslope |  -35517.69   26272.41    -1.35   0.187    -89334.29
18298.91
intervention |  -995085.2   241987.6    -4.11   0.000     -1490774   -
499396.1
       _cons |    2352364   171743.1    13.70   0.000      2000564
2704164
------------------------------------------------------------------------------
-

. gen totviewminushamas=totview - art_t22

. gen totviewminusdup=totview - art_t47

. regress totviewminushamas monthindex postslope intervention if year1<=2014

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
24.85
       Model |  3.4887e+12         3  1.1629e+12   Prob > F        =
0.0000
    Residual |  1.3105e+12        28  4.6805e+10   R-squared       =
0.7269
-------------+----------------------------------   Adj R-squared   =
0.6977
       Total |  4.7992e+12        31  1.5481e+11   Root MSE        =
2.2e+05

------------------------------------------------------------------------------
-
totviewmin~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   41420.51   10710.65     3.87   0.001     19480.73
63360.29
   postslope |   -67513.1   16789.25    -4.02   0.000    -101904.3   -
33121.89
intervention |  -693616.9   154640.9    -4.49   0.000     -1010384   -
376849.4
       _cons |    2289153   109751.5    20.86   0.000      2064337
2513968
------------------------------------------------------------------------------
-

. *
. regress totviewminusdup monthindex postslope intervention if year1<=2014
```

```
      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
8.83
       Model |  2.9756e+12         3  9.9188e+11   Prob > F        =
0.0003
    Residual |  3.1438e+12        28  1.1228e+11   R-squared       =
0.4863
-------------+----------------------------------   Adj R-squared   =
0.4312
       Total |  6.1195e+12        31  1.9740e+11   Root MSE        =
3.4e+05

------------------------------------------------------------------------------
-
totviewmin~p |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   43262.98   16589.01     2.61   0.014     9281.937
77244.02
   postslope |  -28278.84   26003.73    -1.09   0.286    -81545.06
24987.39
intervention |  -985297.4   239512.8    -4.11   0.000     -1475917   -
494677.6
       _cons |    2325107   169986.8    13.68   0.000      1976905
2673309
------------------------------------------------------------------------------
-

.
. *
. * now drop totals and reshape
. drop totv*

. * obvious error in articles on Recruitment and fundamentalism (all numbers
but last couple are the same)
. count if art_t46==art_t47
  30

.
. reshape long art_t, i( monthindex date1 month1 year1 intervention
postslope) j(artnum)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 4
> 1 42 43 44 45 46 47 48)

Data                               wide   ->   long
-----------------------------------------------------------------------------
```

```
Number of obs.                         73   ->    3504
Number of variables                    55   ->       9
j variable (48 values)                      ->   artnum
xij variables:
              art_t1 art_t2 ... art_t48     ->   art_t
-----------------------------------------------------------------------------

. *
. rename art_t pageviews

.
. * pull in article names
. sort artnum

. merge m:1 artnum using articlenames48
(note: variable artnum was byte, now float to accommodate using data's
values)

    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                             3,504  (_merge==3)
    -----------------------------------------

. assert _merge==3

. drop _merge

. * normalize names for better display and read/write
. replace artnames=subinstr(artnames,"(","_",.)
(146 real changes made)

. replace artnames=subinstr(artnames,")","_",.)
(73 real changes made)

. replace artnames=subinstr(artnames," ","_",.)
(1,679 real changes made)

. replace artnames=subinstr(artnames,"+","_",.)
(73 real changes made)

. replace artnames=subinstr(artnames,"-","_",.)
(146 real changes made)

. replace artnames=subinstr(artnames,"__","_",.)
(73 real changes made)

. replace artnames=subinstr(artnames,"__","_",.)
(73 real changes made)
```

```
. replace artnames=subinstr(artnames,"__","_",.)
(0 real changes made)

.
. * pull in indicator of whether article was high privacy
. sort artnum

. merge m:1 artnum using highprivacy31

    Result                           # of obs.
    -----------------------------------------
    not matched                         1,241
        from master                     1,241  (_merge==1)
        from using                          0  (_merge==2)

    matched                             2,263  (_merge==3)
    -----------------------------------------

. gen highprivind=_merge==3

. assert _merge!=2

. drop _merge

. *
. * indicate 7 articles with issues between early and late period
. gen lateissueind=0

. replace lateissueind=1 if artname=="terror"
(73 real changes made)

. replace lateissueind=1 if artname=="Weapons_grade"
(73 real changes made)

. replace lateissueind=1 if artname=="_Euskadi_ta_Askatasuna"
(73 real changes made)

. replace lateissueind=1 if artname=="Pirates"
(73 real changes made)

. replace lateissueind=1 if artname=="Islamist"
(73 real changes made)

. replace lateissueind=1 if artname=="Recruitment"
(73 real changes made)

. replace lateissueind=1 if artname=="Fundamentalism"
(73 real changes made)
```

```
.
.
.
. * check that high privacy desig is ok by checking reg of sum
. egen totview31=sum(pageviews), by(monthindex highprivind)

. replace totview31=. if highprivind==0
(1,241 real changes made, 1,241 to missing)

. bysort monthindex highprivind: gen tmpindx=_n

. regress totview31 monthindex postslope intervention if tmpindx==1 &
year1<=2014

      Source |       SS           df       MS      Number of obs   =
32
-------------+----------------------------------   F(3, 28)        =
20.87
       Model |  5.1404e+11         3  1.7135e+11   Prob > F        =
0.0000
    Residual |  2.2989e+11        28  8.2102e+09   R-squared       =
0.6910
-------------+----------------------------------   Adj R-squared   =
0.6579
       Total |  7.4392e+11        31  2.3998e+10   Root MSE        =
90610

------------------------------------------------------------------------------
-
   totview31 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
  monthindex |   28484.13   4485.873     6.35   0.000     19295.24
37673.02
   postslope |  -41554.21    7031.73    -5.91   0.000    -55958.05   -
27150.36
intervention |  -253556.5   64767.24    -3.91   0.001    -386226.2   -
120886.9
       _cons |   471146.3   45966.52    10.25   0.000     376988.2
565304.5
------------------------------------------------------------------------------
-

. drop tmpindx

. *
. * get ranks of first 17, first 32 and all
```

```
. gen pageviewall=pageviews
(26 missing values generated)

. gen pageviews17=pageviews
(26 missing values generated)

. replace pageviews=. if year>2014
(1,942 real changes made, 1,942 to missing)

. replace pageviews17=. if monthindex>=18
(2,662 real changes made, 2,662 to missing)

. egen rankviewsearly=rank(pageviews), by(artnum)
(1968 missing values generated)

. egen maxrankearly=max(rankviewsearly), by(artnum)

. egen rankviews17=rank(pageviews17), by(artnum)
(2688 missing values generated)

. egen maxrank17=max(rankviews17), by(artnum)

. egen rankviewsall=rank(pageviewall), by(artnum)
(26 missing values generated)

. egen maxrankall=max(rankviewsall), by(artnum)

.
. sum maxr*

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
maxrankearly |      3,504          32           0         32         32
   maxrank17 |      3,504          17           0         17         17
  maxrankall |      3,504    72.45833    3.304247         50         73

. sum rankv*

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
rankviewse~y |      1,536        16.5    9.235782          1         32
 rankviews17 |        816           9    4.901734          1         17
rankviewsall |      3,478    36.80449    21.02188          1         73

. sort artnum date1

.
. *
. gen yearmonth=year1*100 + month1
```

```
. * summermonths lower in general --inidcation of seasonality
. * use rank so all data can be considered on a like to like basis
.  table month1, c(mean rankviewsearly median rankviewsearly mean
rankviewsall median rankviewsall n rankviewsall) row form
> at(%6.2f)

------------------------------------------------------------------------------
---
   month1 | mean(rank~y)  med(rankv~y)  mean(rank~l)  med(rankv~l)
N(rankvie~l)
----------+-------------------------------------------------------------------
---
        1 |        17.38         17.50         39.98         42.00
286
        2 |        16.72         17.00         35.92         36.00
286
        3 |        19.43         20.00         43.27         46.50
286
        4 |        17.34         17.00         39.69         39.50
286
        5 |        19.58         21.00         42.18         45.00
286
        6 |        14.20         14.00         32.54         31.00
287
        7 |        12.55         11.00         29.64         27.00
333
        8 |        11.77          9.00         28.67         27.00
333
        9 |        17.11         17.50         34.46         33.00
285
       10 |        20.39         22.00         40.94         42.00
286
       11 |        18.85         20.00         40.54         41.50
286
       12 |        14.18         14.00         36.27         39.00
238
          |
    Total |        16.50         16.50         36.80         37.00
3,478
------------------------------------------------------------------------------
---

.  table month1, c(mean rankviewsearly median rankviewsearly mean
rankviewsall median rankviewsall n rankviewsall) row form
> at(%6.2f)

------------------------------------------------------------------------------
---
```

```
   month1 | mean(rank~y)  med(rankv~y)  mean(rank~l)  med(rankv~l)
N(rankvie~l)
----------+-----------------------------------------------------------------
---
        1 |        17.38         17.50         39.98         42.00
286
        2 |        16.72         17.00         35.92         36.00
286
        3 |        19.43         20.00         43.27         46.50
286
        4 |        17.34         17.00         39.69         39.50
286
        5 |        19.58         21.00         42.18         45.00
286
        6 |        14.20         14.00         32.54         31.00
287
        7 |        12.55         11.00         29.64         27.00
333
        8 |        11.77          9.00         28.67         27.00
333
        9 |        17.11         17.50         34.46         33.00
285
       10 |        20.39         22.00         40.94         42.00
286
       11 |        18.85         20.00         40.54         41.50
286
       12 |        14.18         14.00         36.27         39.00
238
          |
    Total |        16.50         16.50         36.80         37.00
3,478
----------------------------------------------------------------------------
---

. regress rankviewsall i.month1 if lateissueind==0

      Source |       SS           df       MS      Number of obs   =
2,993
-------------+----------------------------------   F(11, 2981)     =
16.57
       Model |  76589.9048        11  6962.71861   Prob > F        =
0.0000
    Residual |   1252281.6     2,981  420.087754   R-squared       =
0.0576
-------------+----------------------------------   Adj R-squared   =
0.0542
       Total |   1328871.5     2,992  444.141544   Root MSE        =
20.496
```

```
------------------------------------------------------------------------------
-
rankviewsall |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
      month1 |
          2  |  -4.176829   1.848066    -2.26   0.024    -7.800443   -
.5532154
          3  |    3.03252   1.848066     1.64   0.101    -.5910936
6.656134
          4  |  -.5020325   1.848066    -0.27   0.786    -4.125646
3.121581
          5  |   2.004065   1.848066     1.08   0.278    -1.619549
5.627679
          6  |  -7.971545   1.848066    -4.31   0.000    -11.59516   -
4.347931
          7  |  -11.40418   1.780841    -6.40   0.000    -14.89598   -
7.912379
          8  |  -11.82578   1.780841    -6.64   0.000    -15.31759   -
8.333982
          9  |  -6.107724   1.848066    -3.30   0.001    -9.731337    -
2.48411
         10  |    .851626   1.848066     0.46   0.645    -2.771988
4.47524
         11  |   .6300813   1.848066     0.34   0.733    -2.993533
4.253695
         12  |  -3.890244   1.938268    -2.01   0.045    -7.690722   -
.0897656
             |
       _cons |       40.5    1.30678    30.99   0.000     37.93772
43.06228
------------------------------------------------------------------------------
-

. regress rankviewsall i.month1 if monthindex<=32

      Source |       SS           df       MS      Number of obs   =
1,536
-------------+----------------------------------   F(11, 1524)     =
7.50
       Model |    40176.52        11  3652.41091   Prob > F        =
0.0000
    Residual |  741743.313     1,524  486.708211   R-squared       =
0.0514
-------------+----------------------------------   Adj R-squared   =
0.0445
       Total |  781919.833     1,535  509.394028   Root MSE        =
22.061
```

```
------------------------------------------------------------------------------
-
rankviewsall |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
-
      month1 |
           2  |  -.8854167   2.599969    -0.34   0.733    -5.985312
4.214478
           3  |   3.173611   2.599969     1.22   0.222    -1.926284
8.273506
           4  |  -.4930556   2.599969    -0.19   0.850     -5.59295
4.606839
           5  |   3.975694   2.599969     1.53   0.126      -1.1242
9.075589
           6  |  -6.152778   2.599969    -2.37   0.018    -11.25267   -
1.052883
           7  |  -9.854167   2.599969    -3.79   0.000    -14.95406   -
4.754272
           8  |  -10.05208   2.599969    -3.87   0.000    -15.15198   -
4.952188
           9  |   -.984375   2.906853    -0.34   0.735    -6.686231
4.717481
          10  |   5.869792   2.906853     2.02   0.044     .1679358
11.57165
          11  |   3.151042   2.906853     1.08   0.279    -2.550814
8.852898
          12  |  -5.104167   2.906853    -1.76   0.079    -10.80602
.5976892
              |
       _cons  |   36.38542   1.838455    19.79   0.000     32.77925
39.99159
------------------------------------------------------------------------------
-


.
.  * where is maximum?
.  tab yearmonth highpriv  if rankviewsearly==maxrankearly

           |      highprivind
 yearmonth |         0          1 |     Total
-----------+----------------------+----------
    201201 |         2          0 |         2
    201202 |         0          2 |         2
    201203 |         0          1 |         1
    201205 |         1          1 |         2
    201206 |         1          0 |         1
    201208 |         1          0 |         1
```

201209	0	1	1
201210	1	3	4
201211	2	1	3
201301	0	4	4
201302	1	0	1
201303	2	3	5
201304	0	8	8
201305	1	1	2
201307	1	0	1
201308	0	1	1
201309	0	1	1
201310	1	0	1
201311	0	1	1
201403	0	1	1
201405	1	1	2
201406	1	0	1
201407	1	1	2
Total	17	31	48

```
. tab yearmonth highpriv  if rankviewsall==maxrankall
```

yearmonth	highprivind 0	1	Total
201202	0	1	1
201203	0	1	1
201210	1	2	3
201211	2	0	2
201301	0	1	1
201303	2	1	3
201304	0	6	6
201307	1	0	1
201309	0	1	1
201310	1	0	1
201311	0	1	1
201406	1	0	1
201407	1	0	1
201507	1	0	1
201511	1	6	7
201512	0	1	1
201601	0	1	1
201603	0	2	2
201604	0	1	1
201610	0	2	2
201703	1	0	1
201704	0	3	3
201705	1	0	1

```
    201707 |          1          0 |          1
    201805 |          0          1 |          1
    201806 |          1          0 |          1
    201810 |          1          0 |          1
    201811 |          1          0 |          1
-----------+----------------------+----------
     Total |         17         31 |         48


.
. * output to csv for graphics and other analysis
. gen dateformat=date1

. format dateformat %d

.
. outsheet using orig48long.csv, comma replace

. *
. log close
      name:  <unnamed>
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\readandreplicate_20190115.log
  log type:  text
 closed on:  15 Jan 2019, 18:07:40
```

The following is a R code, used to produce the graphs:

```
# libraries need to be commented in once per session
#library(dplyr)
# library(plyr)
#individual article data
# start with empty dataset
rm(list = ls())
art48incl2018<-
read.csv("D:\\clients_2018\\DOJ_Wiki_NSA\\programsdata\\orig48long.csv",sep="
,",header=T)
# article data as used in regressions (aggregated by group)
artagg<-
read.csv("D:\\clients_2018\\DOJ_Wiki_NSA\\programsdata\\articlesaggregate.csv
",sep=",",header=T)
# comparison datasets
compinfra34<-
read.csv("D:\\clients_2018\\DOJ_Wiki_NSA\\programsdata\\infrastructure34.csv"
,sep=",",header=T)
compsec25<-
read.csv("D:\\clients_2018\\DOJ_Wiki_NSA\\programsdata\\security25.csv",sep="
,",header=T)
```

```
comppop26<-
read.csv("D:\\clients_2018\\DOJ_Wiki_NSA\\programsdata\\popular26.csv",sep=",
",header=T)

# get labels for dates
artagg$dateabbr<-paste0(substr(as.character(artagg$date1),3,5),"-
",substr(as.character(artagg$date1),8,9))
art48incl2018$dateabbr<-
paste0(substr(as.character(art48incl2018$dateformat),3,5),"-
",substr(as.character(art48incl2018$dateformat),8,9))
if
(sum(unique(art48incl2018$monthindex)==sort(unique(art48incl2018$monthindex))
)<73) stop("Dates out of Order")
labellong<-unique(art48incl2018$dateabbr)
labelshort<-labellong[1:32]
# end date label

# create data without NAs and without data that has issues between 2014 and
later data
artincl2018noNA<-art48incl2018[!is.na(art48incl2018$rankviewsall),]
# just time through 2014
art48<-art48incl2018[art48incl2018$monthindex<=32,]
art48$artnames<-as.character(art48$artnames)
########################
# get summary stats
########################
sum2018noissue<-
ddply(artincl2018noNA[artincl2018noNA$lateissueind==0,],.(monthindex,interven
tion,postslope),summarise, mean1=mean(rankviewsall),
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))
sum2018_47noissue<-ddply(artincl2018noNA[artincl2018noNA$lateissueind==0 &
artincl2018noNA$artnames!="Hamas",],.(monthindex,intervention,postslope),summ
arise, mean1=mean(rankviewsall),
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))
sum2018_31noissue<-ddply(artincl2018noNA[artincl2018noNA$lateissueind==0 &
artincl2018noNA$highprivind==1,],.(monthindex,intervention,postslope),summari
se, mean1=mean(rankviewsall),
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))

sum2018all<-
ddply(artincl2018noNA,.(monthindex,intervention,postslope),summarise,
mean1=mean(rankviewsall),
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))
sum2014_48<-ddply(art48,.(monthindex,intervention,postslope),summarise,
mean1=mean(rankviewsall),
```

```
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))
sum2014_47<-
ddply(art48[art48$artnames!="Hamas",],.(monthindex,intervention,postslope),su
mmarise,mean1=mean(rankviewsall),
median1=median(rankviewsall),meanviews=mean(pageviewall),medviews=median(page
viewall))

sum2014_31<-
ddply(art48[art48$highprivind==1,],.(monthindex,intervention,postslope),summa
rise,mean1=mean(rankviewsall),median1=median(rankviewsall),meanviews=mean(pag
eviewall),medviews=median(pageviewall))


####################################
# show aggregate views and ranking by month
####################################
numagg<-length(unique(artagg$artnmshort))
artnms<-sort(unique(artagg$artnmshort),decreasing=T)
artnmslong<-as.character(artnms)
artnmslong[artnms=="Terror_48"]<-"Terror 48"
artnmslong[artnms=="Terror_47"]<-"Terror 48 without Hamas"
artnmslong[artnms=="Terror_31"]<-"High Privacy 31"

cols1<-
c("black","darkgreen","blue","green","magenta","orange","mediumorchid1","red"
)
lwd1<-c(rep(3,3),rep(1,5))
pch1<-c(7:9,0:2,5:6)
# aggregate  rank terror
tmpplot<-artagg[artagg$artnms==artnms[1],]
plot(tmpplot$monthindex,tmpplot$rankviews,type="b",pch=pch1[1],col=cols1[1],l
wd=lwd1[1],xlim=c(0,40),ylim=c(0,33),axes=F,ylab="Rank of Page Views by
Month: 1 is Lowest and 32 is Highest",xlab="")
for (i in 2:3) {
tmpplot<-artagg[artagg$artnms==artnms[i],]
lines(tmpplot$monthindex,tmpplot$rankviews,col=cols1[i],type="b",lwd=lwd1[i],
pch=pch1[i])
}
axis(1,1:32,label=unique(artagg$dateabbr),cex.axis=1,las=2)
axis(2,at=c(1,10,20,32))
legend("topright",legend=artnmslong[1:3],text.col=c(cols1[1:3]),cex=1.3)
abline(v=17.5,lwd=2)
# save with just terror articles
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","aggregate_ran
k.png"),type="png")
# add controls
tmpplot<-artagg[artagg$artnms==artnms[1],]
```

```
plot(tmpplot$monthindex,tmpplot$rankviews,type="b",pch=pch1[1],col=cols1[1],l
wd=lwd1[1],xlim=c(0,40),ylim=c(1,32),axes=F,ylab="Rank of Page Views by
Month: 1 is Lowest and 32 is Highest",xlab="")
axis(1,1:32,label=unique(artagg$dateabbr),cex.axis=1,las=2)
axis(2,at=c(1,10,20,32))
for (i in 2:8) {
tmpplot<-artagg[artagg$artnms==artnms[i],]
lines(tmpplot$monthindex,tmpplot$rankviews,col=cols1[i],type="b",lwd=lwd1[i],
pch=pch1[i])
abline(v=17.5,lwd=2)

}
legend("topright",legend=artnmslong,text.col=cols1,pch=pch1,cex=1.2)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","aggregate_ran
kwithcontrols.png"),type="png")

## now plot each of the 8 separately

for (i in 4:8) {
tmpplot<-artagg[artagg$artnms==artnms[i],]
plot(tmpplot$monthindex,tmpplot$rankviews,type="b",pch=pch1[1],col=cols1[1],x
lim=c(0,32),ylim=c(0,33),axes=F,ylab="Rank of Page Views by Month: 1 is
Lowest and 32 is Highest",xlab="",lwd=2,main=paste("Rank of Views by Month
for Control:",artnms[i]))
axis(1,1:32,label=unique(artagg$dateabbr),cex.axis=1,las=2)
axis(2,at=c(1,10,20,32))
abline(v=17.5)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/",
"aggregate_comp_", artnms[i],".png"),type="png")

}
##################################################
# now show total views (as in Figure 1 of Penney)
##################################################

tmpplot<-artagg[artagg$artnms==artnms[1],]
plot(tmpplot$monthindex,tmpplot$pageviews,type="b",pch=pch1[1],col=cols1[1],l
wd=lwd1[1],xlim=c(0,32),ylim=c(0,4200000),axes=F,ylab="Page Views in
Millions",xlab="")

for (i in 2:3) {
tmpplot<-artagg[artagg$artnms==artnms[i],]
lines(tmpplot$monthindex,tmpplot$pageviews,col=cols1[i],type="b",lwd=lwd1[i],
pch=pch1[i])
}
axis(1,1:32,label=unique(artagg$dateabbr),cex.axis=1,las=2)
axis(2,at=c(0,1,2,3,4,5)*1000000,label=paste((0:5),"MM"),las=2)
legend("topright",legend=artnmslong[1:3],text.col=c(cols1[1:3]))
abline(v=17.5,lwd=2)
```

```
# save with just terror articles
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","aggregate32_s
um.png"),type="png")

#########################################################################
# End aggregate graphs with controls
#########################################################################

###################################
# now look at terror data aggregates
###################################
# look at mean and median views
# first median
plot(sum2018noissue$monthindex[1:32],sum2018noissue$medviews[1:32],type="b",a
xes=F,ylim=c(0,max(sum2018noissue$medviews*1.1)),xlab="",ylab="Median Number
of Page Views",lwd=2,xlim=c(0,75))

axis(1,at=c(1:32,35:75),label=labellong,las=2)
axis(2,at=c(0,10000,20000,30000,40000,50000),label=c("0","10K","20K","30K","4
0K","50K"))
abline(v=17.5)
title(main=" ")

lines(sum2018noissue$monthindex[1:32],sum2018_47noissue$medviews[1:32],col="r
ed",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[1:32],sum2018all$medviews[1:32],col="darkgree
n",type="b",lty=3,lwd=2)
lines(sum2018noissue$monthindex[1:32],sum2018_31noissue$medviews[1:32],col="b
lue",type="b",lty=4,lwd=2)

lines(sum2018noissue$monthindex[33:73]-
8,sum2018noissue$medviews[33:73],col="black",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018_47noissue$medviews[33:73],col="red",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018all$medviews[33:73],col="darkgreen",type="b",lty=3,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018_31noissue$medviews[33:73],col="blue",type="b",lty=4,lwd=2)

abline(h=(1:9)*10000,lty=3)

legend(61,40000,legend=c("Terror 48","Terror 41","Terror 41 without
Hamas","High Privacy
26"),text.col=c("darkgreen","black","red","blue"),cex=1.2)

savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","terror48plusm
edviews.png"),type="png")

# now mean views
```

```
plot(sum2018noissue$monthindex[1:32],sum2018noissue$meanviews[1:32],type="b",
axes=F,ylim=c(min(sum2018noissue$meanviews*.3),max(sum2018noissue$meanviews*1
.1)),xlab="",ylab="Average Number of Page Views",lwd=2,xlim=c(1,75))
axis(1,at=c(1:32,35:75),label=labellong,las=2)
axis(2,at=c(1:9)*10000,label=paste0(c(1:9)*10,"K"))
abline(v=17.5)
title(main=" ")


lines(sum2018noissue$monthindex[1:32],sum2018_47noissue$meanviews[1:32],col="
red",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[1:32],sum2018all$meanviews[1:32],col="darkgre
en",type="b",lty=3,lwd=2)
lines(sum2018noissue$monthindex[1:32],sum2018_31noissue$meanviews[1:32],col="
blue",type="b",lty=4,lwd=2)

lines(sum2018noissue$monthindex[33:73]-
8,sum2018noissue$meanviews[33:73],col="black",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018_47noissue$meanviews[33:73],col="red",type="b",lty=2,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018all$meanviews[33:73],col="darkgreen",type="b",lty=3,lwd=2)
lines(sum2018noissue$monthindex[33:73]-
8,sum2018_31noissue$meanviews[33:73],col="blue",type="b",lty=4,lwd=2)
abline(h=(1:9)*10000,lty=3)

legend(60,105000,legend=c("Terror 48","Terror 41","Terror 41 without
Hamas","High Privacy
26"),text.col=c("darkgreen","black","red","blue"),cex=1.2)

savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","terror48plusa
vgviews.png"),type="png")

##################################
# Just 32 months until aug 2014
##################################

plot(sum2014_48$monthindex,sum2014_48$meanviews,type="b",axes=F,ylim=c(min(su
m2014_48$meanviews*.3),max(sum2014_48$meanviews*1.1)),xlab="",ylab="Average
Number of Page Views",lwd=3,xlim=c(1,32),col="darkgreen")
axis(1,at=c(1:32),label=labelshort,las=2,cex.axis=1.5)
axis(2,at=c(1:9)*10000,label=paste0(c(1:9)*10,"K"))
abline(v=17.5)
#title(main=" ")


lines(sum2014_47$monthindex,sum2014_47$meanviews,col="red",type="b",lty=2,lwd
=3)
```

```
lines(sum2014_31$monthindex,sum2014_31$meanviews,col="blue",type="b",lty=4,lw
d=3)

abline(h=(1:9)*10000,lty=3)

legend("topright",legend=c("Terror 48","Terror 48 without Hamas","High
Privacy 31"),text.col=c("darkgreen","red","blue"),cex=1.5)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","terror48_2014
averageviews.png"),type="png")
#

plot(sum2014_48$monthindex,sum2014_48$medviews,type="b",axes=F,ylim=c(min(sum
2014_48$medviews*.3),max(sum2014_48$medviews*1.1)),xlab="",ylab="Median
Number of Page Views",lwd=3,xlim=c(1,32),col="darkgreen")
axis(1,at=c(1:32),label=labelshort,las=2,cex.axis=1.5)
axis(2,at=c(0:6)*5000,label=paste0(c(0:6)*5,"K"))
abline(v=17.5)
#title(main=" ")


lines(sum2014_47$monthindex,sum2014_47$medviews,col="red",type="b",lty=2,lwd=
3)
lines(sum2014_31$monthindex,sum2014_31$medviews,col="blue",type="b",lty=4,lwd
=3)

abline(h=(1:6)*5000,lty=3)

legend(24,19500,legend=c("Terror 48","Terror 48 without Hamas","High Privacy
31"),text.col=c("darkgreen","red","blue"),cex=1.5)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/","terror48_2014
medianviews.png"),type="png")

#####################################
# End mean and median 48 plots
#####################################

#####################################
# now do all 48 articles individually
#####################################
for (i in 1:48) {
tmpplot<-art48incl2018[art48incl2018$artnum==i,]
tmpname<-unique(tmpplot$artname)
plot(tmpplot$monthindex,tmpplot$pageviewall,main=paste("Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=tmpplot$monthindex,label=tmpplot$dateabbr,las=2)
axis(2,at=1000*pretty(tmpplot$pageviewall/1000),label=paste0(pretty(tmpplot$p
ageviewall/1000),"K"),las=2)
```

```
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","indivg
rph_",tmpname,".png"),type="png")
# just first 32
tmpplot<-art48incl2018[art48incl2018$artnum==i &
art48incl2018$monthindex<=32,]
tmpname<-unique(tmpplot$artname)
plot(tmpplot$monthindex,tmpplot$pageviewall,main=paste("Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=tmpplot$monthindex,label=tmpplot$dateabbr,las=2)
axis(2,at=1000*pretty(tmpplot$pageviewall/1000),label=paste0(pretty(tmpplot$p
ageviewall/1000),"K"),las=2)
abline(v=17.5,lwd=2)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","indiv3
2grph_",tmpname,".png"),type="png")
}

# infrastructure plots
infranames<-names(compinfra34)
for (i in 1:34) {
tmpploty<-compinfra34[,i+4]
tmpname<-infranames[i+4]
plot(1:32,tmpploty,main=paste("Infrastructure: Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=1:32,label=labelshort,las=2)
axis(2,at=1000*pretty(tmpploty/1000),label=paste0(pretty(tmpploty/1000),"K"),
las=2)
abline(v=17.5,lwd=2)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","infra3
4_",tmpname,".png"),type="png")
}
# security plots

securitynames<-names(compsec25)
for (i in 1:25) {
tmpploty<-compsec25[,i+4]
tmpname<-securitynames[i+4]
plot(1:32,tmpploty,main=paste("Security: Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=1:32,label=labelshort,las=2)
axis(2,at=1000*pretty(tmpploty/1000),label=paste0(pretty(tmpploty/1000),"K"),
las=2)
abline(v=17.5,lwd=2)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","sec25_
",tmpname,".png"),type="png")
}
```

```
# popular plots

popnames<-names(comppop26)
for (i in 1:26) {
tmpploty<-comppop26[,i+4]
tmpname<-popnames[i+4]
plot(1:32,tmpploty,main=paste("Popular: Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views in Millions")
axis(1,at=1:32,label=labelshort,las=2)
axis(2,at=1000000*pretty(tmpploty/1000000),label=paste0(pretty(tmpploty/10000
00),"MM"),las=2)
abline(v=17.5,lwd=2)
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","pop26_
",tmpname,".png"),type="png")
}

# multiple per page first 32 months
# just first 32
par(mfrow=c(4,3))
for (i in 1:48) {
tmpplot<-art48incl2018[art48incl2018$artnum==i &
art48incl2018$monthindex<=32,]
tmpname<-unique(tmpplot$artname)
plot(tmpplot$monthindex,tmpplot$pageviewall,main=paste("Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=tmpplot$monthindex,label=tmpplot$dateabbr,las=2)
axis(2,at=1000*pretty(tmpplot$pageviewall/1000),label=paste0(pretty(tmpplot$p
ageviewall/1000),"K"),las=2)
abline(v=17.5,lwd=2)
if (trunc(i/12)==i/12) {
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","mfrow4
3_32grph_",i,".png"),type="png")
}
}
# show top four in terms of page views
par(mfrow=c(2,2))
top4<-c("Pakistan","Iran","Nigeria","Afghanistan")
for (i in 1:4) {
tmpplot<-art48incl2018[art48incl2018$artname==top4[i] &
art48incl2018$monthindex<=32,]
tmpname<-unique(tmpplot$artname)
plot(tmpplot$monthindex,tmpplot$pageviewall,main=paste("Page Views
for",tmpname),col="blue",type="b",lwd=2,axes=F,xlab="",ylab="Monthly Page
Views")
axis(1,at=tmpplot$monthindex,label=tmpplot$dateabbr,las=2)
axis(2,at=1000*pretty(tmpplot$pageviewall/1000),label=paste0(pretty(tmpplot$p
ageviewall/1000),"K"),las=2)
```

```
abline(v=17.5,lwd=2)
}
savePlot(paste0("D:/clients_2018/DOJ_Wiki_NSA/programsdata/R/graphs/","top4_3
2grph_",i,".png"),type="png")


par(mfrow=c(1,1))

#library(dplyr)
# indiv
tmpcol=rep(c("black","darkgreen","blue","green","magenta","orange","mediumorc
hid1","red"),8)
tmpplot<-art48incl2018[art48incl2018$artnum==1 &
art48incl2018$monthindex<=32,]

plot(tmpplot$monthindex,tmpplot$pageviewall,main=paste("
"),col=tmpcol[1],type="b",axes=F,xlab="",ylab="Monthly Page
Views",ylim=c(0,600000),lwd=2)
axis(1,at=tmpplot$monthindex,label=tmpplot$dateabbr,las=2)
axis(2,at=c(0:6)*100000,label=c("0",paste0(1:5,"00K"),">600K"),las=2)

for (i in 2:48) {
tmpplot<-art48incl2018[art48incl2018$artnum==i &
art48incl2018$monthindex<=32,]
tmpplot$pageviewall[tmpplot$pageviewall>600000]<-600000
tmpname<-unique(tmpplot$artname)
lines(tmpplot$monthindex,tmpplot$pageviewall,type="b",col=tmpcol[i],lwd=2)
}
savePlot("all48inonegraph.png",type="png")

#
# Figure 2
plot(sum2014_47$monthindex,sum2014_47$meanviews*47,main=paste("
"),type="b",axes=F,xlab="",ylab="Monthly Page
Views",ylim=c(1500000,3500000),lwd=2,col="red",cex.lab=1.2)
axis(1,at=sum2014_47$monthindex,label=tmpplot$dateabbr,las=2,cex.axis=1.3)
axis(2,at=1000000*c(1.5,2.0,2.5,3.0,3.5),label=c("1.5MM","2.0MM","2.5MM","3.0
MM","3.5MM"),las=2,pos=c(.8,1500000),cex.axis=1.2)
abline(v=17.5,lwd=2)
savePlot("Penneyfig2.png",type="png")
```

APPENDIX II: Documents Considered

1. *Dkt 186-6_Declaration of Jonathon Penney.pdf* ("Penney Declaration")
2. *English Homepage Views (Raw - Non-Mobile).xlsx* – Provided to me as data underlying the Penney Declaration analysis.
3. *Final 25 Article Security Comparator Data Set.xlsx* - Provided to me as data underlying the Penney Declaration analysis.
4. *Higher Privacy Rated Terrorism Articles (above 2) (31 Articles Set).xlsx* - Provided to me as data underlying the Penney Declaration analysis.
5. *IndependentPrivacyRatingResults-Full-Survey.pdf* – Provided to me as data underlying the Penney Declaration analysis.
6. *Infrastructure Security Comparator (34 Articles).xlsx* – Provided to me as data underlying the Penney Declaration analysis.
7. *Popular-Wikipedia-Pages-Comparator (26 Articles).xlsx* – Provided to me as data underlying the Penney Declaration analysis.
8. *Wikipedia Case Study - Key Variables.xlsx* – Provided to me as data underlying the Penney Declaration analysis.
9. *Wikipedia-Case-Study-Article-Traffic-June 2015-Full-48.xlsx* – Provided to me as data underlying the Penney Declaration analysis.
10. *Wikipedia-Case-Study-Article-Traffic-June 2015-Full-48_format_plus2018.xslx* – 48 Articles page views for months through 2018, which I compiled using the website referenced in my Declaration. I call these articles the Terror 48 in the body of my declaration.
11. *ISIS variations pageviews-20150701-20181130* – Article page views for ISIS, which I compiled using the website referenced in my Declaration.
12. Additional documents provided for consideration by the Department of Justice (but which I did not refer to in writing my Declaration).
    1. WIKI0001545.pdf
    2. WIKI0002024.pdf
    3. WIKI0002025.xlsx
    4. WIKI0002263.pdf
    5. WIKI0002274.pdf
    6. WIKI0002607.xlsx
    7. WIKI0002608.xlsx
    8. WIKI0004893.pdf
    9. WIKI0005137.pdf
    10. WIKI0005154.pdf
    11. WIKI0005174.pdf
    12. WIKI0005194.pdf
    13. WIKI0005229.pdf
    14. WIKI0005251.pdf
    15. WIKI0005266.pdf
    16. WIKI0005285.pdf
    17. WIKI0005300.pdf
    18. WIKI0005322.pdf

19. WIKI0005336.pdf
20. WIKI0005360.pdf
21. WIKI0005379.pdf
22. WIKI0005399.pdf
23. WIKI0005420.pdf
24. WIKI0005439.pdf
25. WIKI0005466.pdf
26. WIKI0005487.pdf
27. WIKI0005500.pdf
28. WIKI0005514.pdf
29. WIKI0005528.pdf
30. WIKI0005544.pdf
31. WIKI0005577.pdf
32. WIKI0005693.pdf
33. WIKI0005832.pdf
34. WIKI0005978.pdf
35. WIKI0006146.xlsx
36. WIKI0006147.xlsx
37. WIKI0006148.xlsx
38. WIKI0006149.xlsx
39. WIKI0006282.csv
40. WIKI0006283.pdf
41. WIKI0006295.xlsx
42. WIKI0006296.pdf
43. WIKI0006367.xlsx
44. WIKI0006368.csv
45. WIKI0006369.pdf
46. WIKI0007358.pdf
47. WIKI0007616.xlsx
48. WIKI0008237.pdf
49. WIKI0008262.pdf
50. WIKI0008271.xlsx
51. WIKI0008312.csv
52. WIKI0008313.csv
53. WIKI0009301.csv
54. WIKI0009302.xlsx

APPENDIX III: Resume and Testimony History

Resume of Alan J. Salzberg

EXPERIENCE

Salt Hill Statistical Consulting, Founder and Principal, 2000-present
Founder and Principal of a statistical consulting company (formerly Quantitative Analysis). The firm is skilled at presenting complex ideas to non-experts, including providing expert testimony in court settings. Capabilities include development and implementation of statistical techniques as well as critical review and audit of existing statistical estimates, samples, and models. The company's clients are law firms, government, and private corporations and have included: United States Department of Labor; Pfizer; Barnes & Thornburg; Honeywell; K&L Gates; City of New York.

Summit Consulting, Teaming Partner, 2009-present
Consult on multiple engagements with economic consulting firm on large-scale government projects. Served as a Director at the firm in 2014.

Analysis & Inference, Inc., CEO, 1991-1995 and 2008-2013
Led a statistical consulting company that provides consulting services to corporations, law firms, and government.

KPMG LLP, Practice Leader, Quantitative Analysis Group – New York, 1996-2000
Established and led the New York office of KPMG's Quantitative Analysis Group.
Morgan Stanley, Associate, 1988-1990, 1995-1996
Performed statistical modeling and software design.

EDUCATION

Ph.D., Statistics, Wharton School, University of Pennsylvania, 1995
M.A., Statistics, Wharton School, University of Pennsylvania, 1992
B.S., Economics (concentration in Economics and Finance), *cum laude*, Wharton School, University of Pennsylvania, 1988

ENGAGEMENTS

- Served as a statistical consultant on behalf of the United States government and other entities in the development of dynamic models for residential property valuation in order to determine whether certain residential mortgage-backed securities (RMBS) were fairly valued. Made use of statistical and econometric techniques including regression modeling, statistical sampling, bootstrapping, and bias adjustment.

- Using social security and insurance company data, developed two probability-based models in order to match unclaimed assets with the individual owners of those assets. The models

were successfully implemented at our client, a financial services company, and used to assist state agencies in locating unclaimed assets.

- Served as a statistical expert on behalf of a nuclear power plant owner in a construction delay dispute. Analyzed a statistical sample and model from a population of more than 100,000 comments on design documents. Authored three expert reports and testified before the International Chamber of Commerce's arbitration court in London.

- Served as a statistical sampling expert on behalf of an arbitration panel in a dispute regarding payments on several thousand healthcare claims. Analyzed data from samples of those claims and made recommendations to the arbitration panel regarding proper interpretation and extrapolation of the sample.

- On behalf of the New York State Office of Medicaid Inspector General, reviewed the sampling and estimation methodology used to audit Medicaid providers in New York State. Reviewed and critiqued specific methodologies in ongoing matters, and provided recommendations for improving the statistical audit process.

- On behalf of a Fortune 100 company, evaluated models that estimated the potential liability in more than 10,000 asbestos settlements. In addition, reviewed the likely bias and other issues with a model that predicted the "propensity to sue" for future claims. Wrote two expert reports concerning findings and testified as a statistical expert regarding those findings.

- In a series of matters on behalf of the law department for a major city, created and analyzed a massive real estate database, modeled market and sales values, and wrote expert reports to determine potential biases of alternative methods of valuing commercial real estate. Determined the validity of assumptions about lease lengths, turnover rates, and other issues affecting rents and property values. Testified as a statistical expert in one of these matters.

- On behalf of the United States Department of Labor, acted as the principal investigator on a study of industry compliance with certain labor laws. Developed and pulled a statistical sample for evaluation. Performed survival analysis to better understand how long certain industry investigations would last and the likely outcomes of such investigations.

- For major pharmaceutical company, analyzed company and external marketing data to determine reliability and potential biases in using external data sources. Analyzed physician-specific data for a period of 36 months concerning product marketing to approximately 1 million prescription drug subscribers.

- In complex litigation matter involving an undersea oil field, analyzed data from several years of inspections and repairs to determine likelihood of a catastrophic failure that would result in a major oil spill. Used survival analysis to determine the likelihood of such an event for different inspection and repair cycles.

- On behalf of several state public service commissions, directed data analysis and statistical design in a series of tests of Bell South, Verizon, SBC-Ameritech, and Qwest. Beginning in

1998, developed software and procedures for calculating performance metrics and evaluating the competitive environment. Testified before several state public service commissions, including New York, Virginia, Florida, Michigan, and Colorado.

- Modeled television audience ratings to determine the Public Broadcasting System's share of cable royalty distributions. Used statistical methods to determine a reliable estimate of PBS's cable royalty share. The estimate resulted in a multi-million dollar decision in favor of the Public Broadcasting System by the Cable Royalty Tribunal.

- Lead statistician in the design and implementation of a sample of all personal property and equipment on behalf of the United States Internal Revenue Service. The population of interest involved more than one million items contained in over 1,000 buildings. The sample design, implementation, and resulting estimates and projections were subject to intense scrutiny by the United States General Accounting Office.

- For the United States Department of Justice, designed and implemented a sample to estimate the number of immigrants improperly granted citizenship. The sample was designed to provide precision of plus or minus less than 1%, for a population of more than 1 million immigrants. The work was the focus of intense congressional scrutiny and received substantial review in the media.

- On behalf of Fortune 100 company, created statistical models to determine the probabilities and likely severities of accidents for different employee and accident types. This project resulted in recommended annual savings of $3 million.

- On behalf of the Arava Institute of Environmental Studies, advised on design and sampling methodology for a broad-based survey of environmental education in middle and high schools. More than 7,000 students were surveyed in a sample that was stratified by size of town, income level, and other socio-economic variables. Performed weighted statistical analysis to project survey results to the population. Presented results before Israeli Congressional committee in July 2007.

- For the United States Customs Service (Department of Homeland Security), assisted with sampling of financial statement information. Designed and wrote sampling plans, helped implement the plans, and created spreadsheet calculator to analyze results. In an earlier engagement, evaluated the credibility of statistical sampling and analysis used to track and categorize imports, for the Office of Inspector General. Suggested improved methods of sampling and implementation.

- Provided expert testimony in statistics more than two dozen trials, hearings, and depositions over the last 20 years, including multiple times in United States Federal Court.

RESEARCH

"What are the Chances?" blog, 2007 to present. Excerpts have been included in newspapers and textbooks, including Lundsford, Andrea L. and Ruszkiewicz, John, *Everything's an Argument, 6th Edition,* 2012. The blog is publicly available at https://salthillstatistics.com/blog.

"Resolving a Multi-Million Dollar Contract Dispute with a Latin Square," *American Statistician*, with William B. Fairley, Steven M. Crunk, Peter J. Kempthorne, Julie Novak, and Bee Leng Lee, 2017.

"Law and Statistics of Combining Categories: Wal-Mart and Employment Discrimination Cases", with Albert J. Lee, *Proceedings of the 2010 Joint Statistical Meetings of the American Statistical Association*, 2010.

"Evaluating the Environmental Literacy of Israeli Elementary and High School Students," with Maya Negev, Gonen Sagy, and Alon Tal, *Journal of Environmental Education*, Winter 2008.

"Trends in Environmental Education in Israel," with Gonen Sagy, Maya Negev, Yaakov Garb, and Alon Tal, *Studies in Natural Resources and Environment*, Vol. 6, 2008. [In Hebrew]

"Results from a Representative Sample in the Israeli Educational System," with Gonen Sagy, Maya Negev, Yaakov Garb, and Alon Tal, *Studies in Natural Resources and Environment*, Vol. 6, 2008. [In Hebrew]

"Comment on Local model uncertainty and incomplete-data bias by Copas and Li," with Paul R. Rosenbaum, *Journal of the Royal Statistical Society, Series B*, 2005.

"Determining Air Exchange Rates in Schools Using Carbon Dioxide Monitoring", with D. Salzberg and C. Fiegley, presented at the *American Industrial Hygiene Conference and Expo*, 2004.

"The Modified Z versus the Permutation Test in Third Party Telecommunications Testing", *Proceedings of the 2001 Joint Statistical Meetings of the American Statistical Association*.

"Removable Selection Bias in Quasi-experiments," *The American Statistician*, May 1999.

"Skewed oligomers and origins of replication," with S. Salzberg, A. Kervalage, and J. Tomb, *Gene*, Volume 217, Issue 1-2 (1998), pp. 57-67.

"Selection Bias in Quasi-experiments," (Doctoral Thesis), 1995.

Editorial Contributor (referee for scholarly papers)**,** *American Statistician*.

Patent *(#6,636,585)* One of five inventors on a patent for statistical process design related to information systems testing.

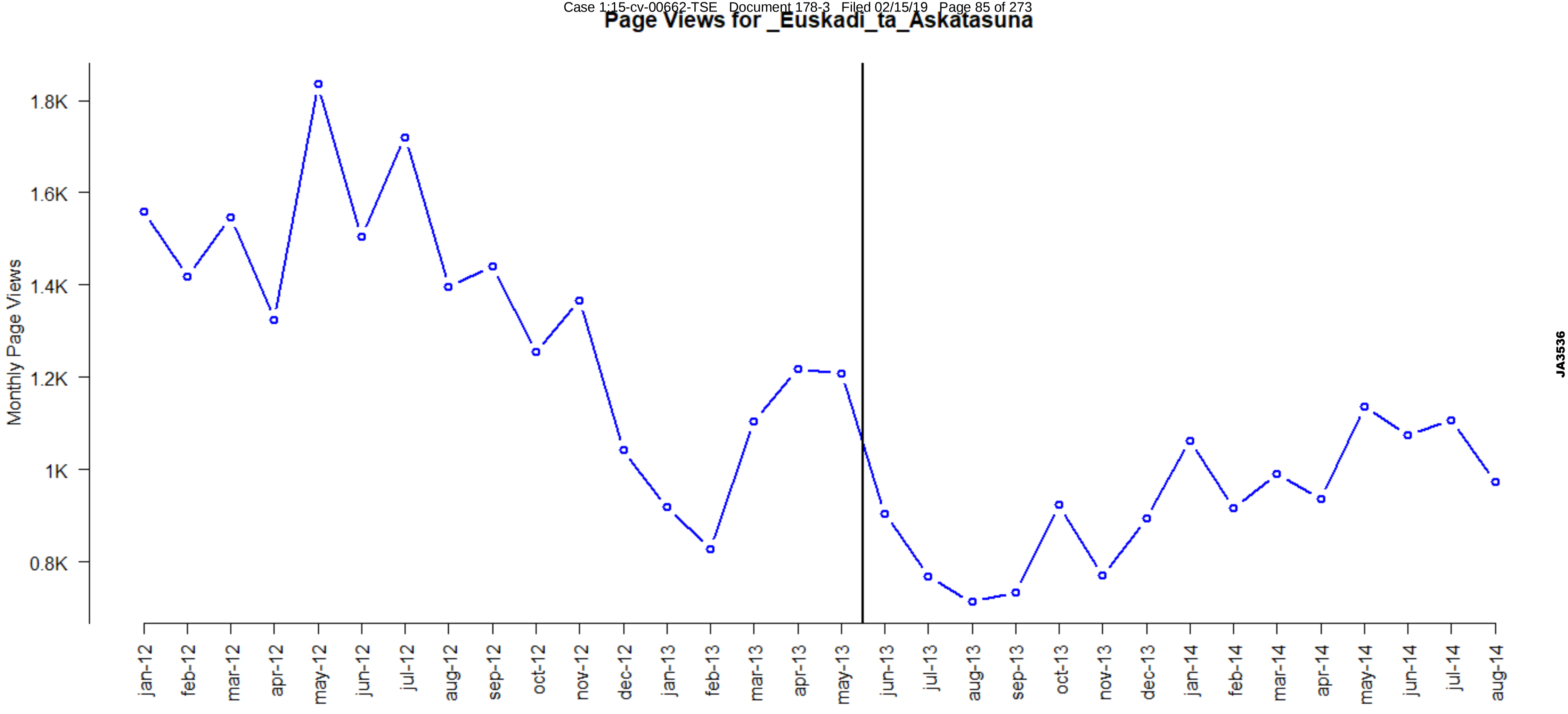
PERSONAL

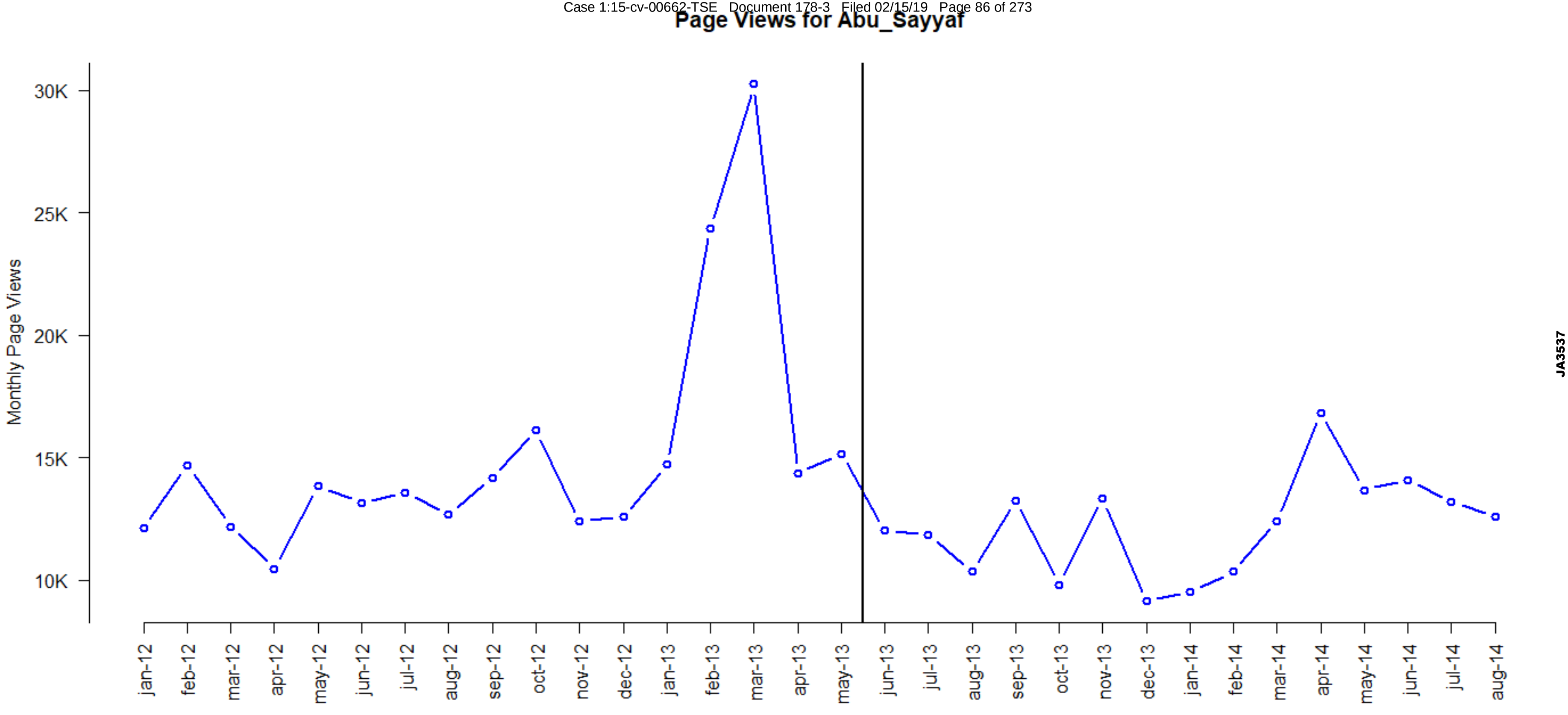
Married, with two daughters and a son.
Languages: English (native), Hebrew (conversational).
Member, Park Slope Food Coop.
Member, 39 Plaza Housing Corp (residential coop). Board member, 2012-2015.
Enjoy ultimate Frisbee, basketball, biking, hiking, running, tennis, chess, and bridge.

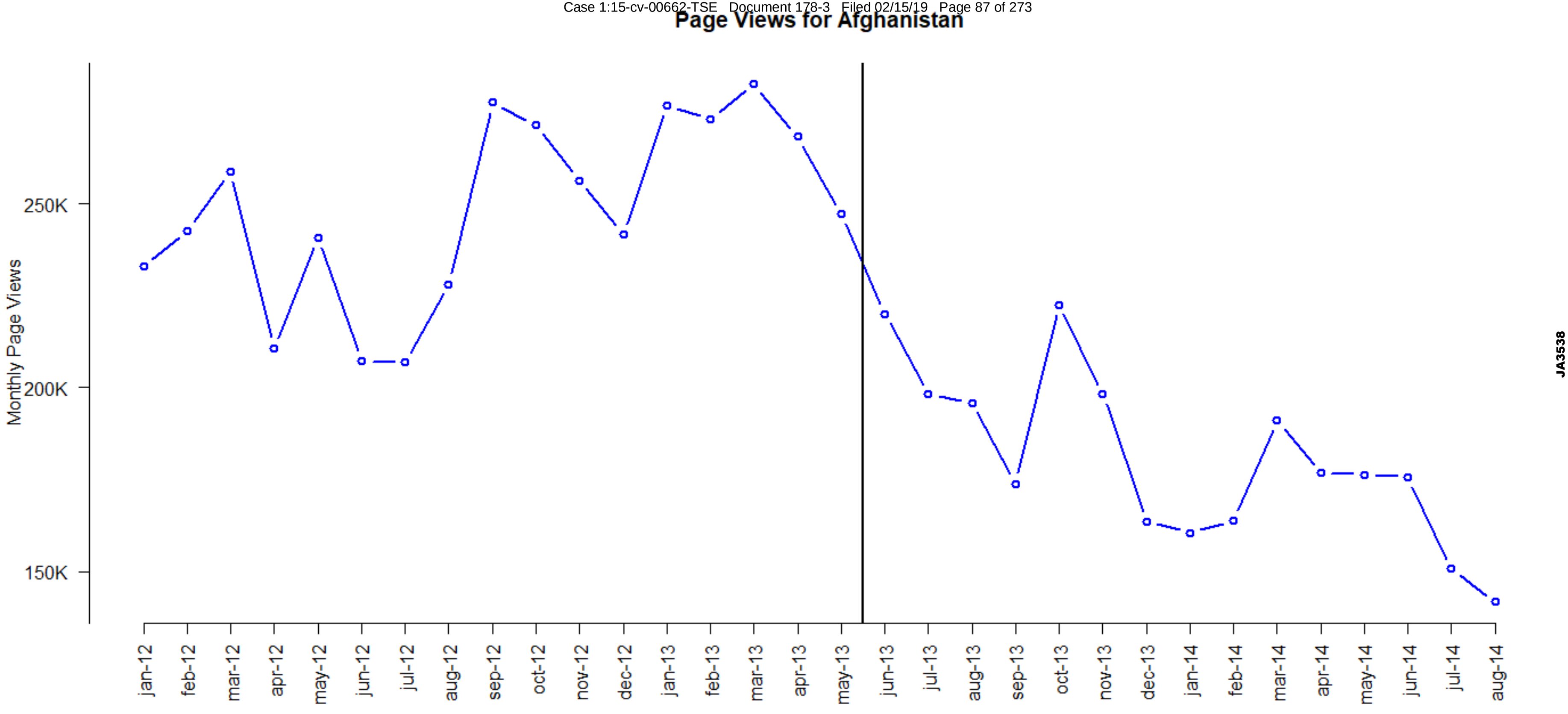
FOUR YEAR TESTIMONY HISTORY

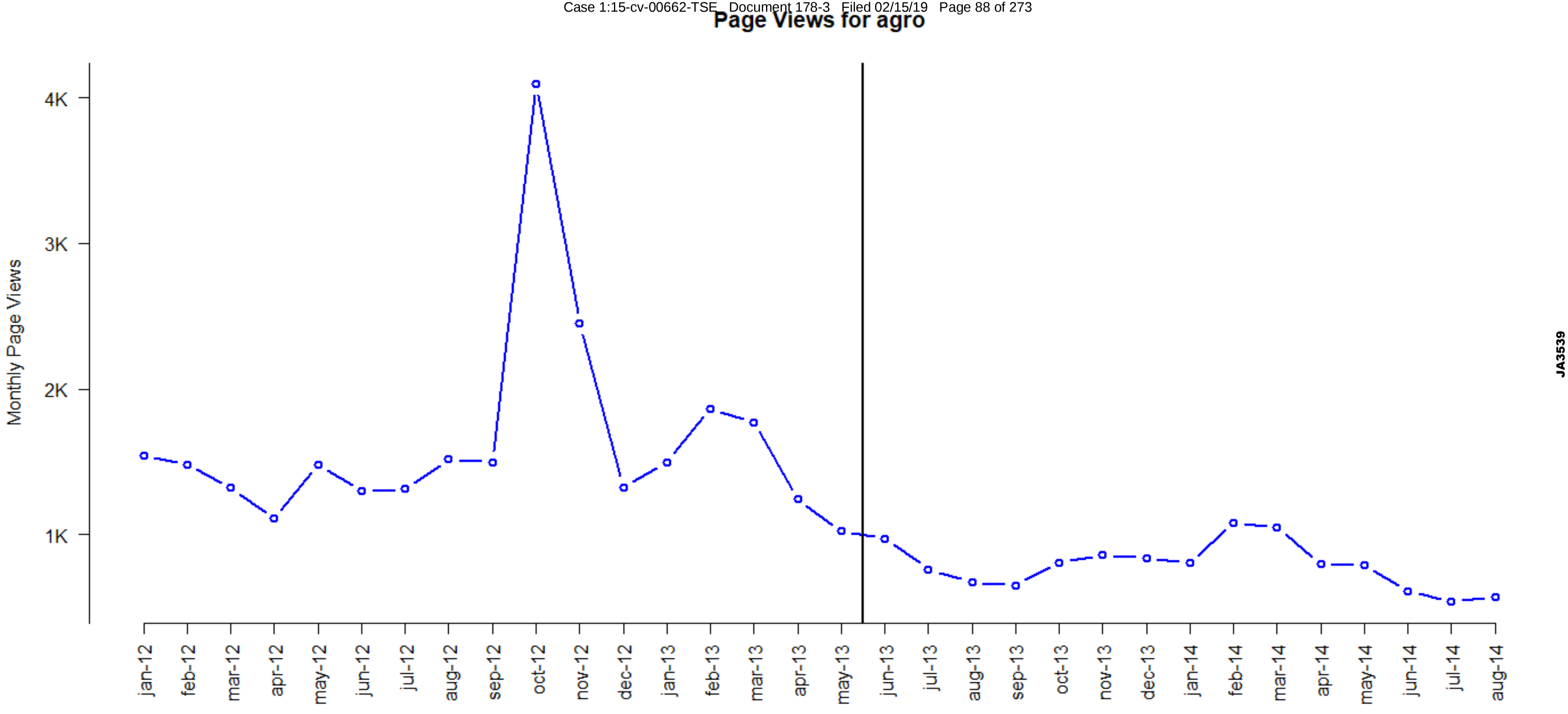
1. [Federal court] Bayer Healthcare LLC, v. Baxalta, et al, 2019.
2. [Federal court] Steward, et al, v. State of Texas, 2018.
3. [deposition] Center for Independence of the Disabled, et al, v. Metropolitan Transit Authority, et al, 2018.
4. [deposition] Bayer Healthcare, LLC, v. Baxalta Inc., et al, 2018.
5. [deposition] New Image Global, Inc. v. U.S., 2017.
6. [Federal court] Steward, et al, v. State of Texas, 2017.
7. [deposition] Home Equity Mortgage Trust, et al., v. DLJ Mortgage Capital, et al., 2017.
8. [court] Regents of the University of California v. County of Sacramento, 2016.
9. [international arbitration] Areva NP GmbH, Areva NP S.A.S. and Siemens Aktiengesellschaft v. Teollisuuden Voima Oyj, 2016.
10. [Federal court] Kerner v. City & County of Denver, 2015.
11. [deposition] Regents of the University of California v. County of Sacramento, 2015.

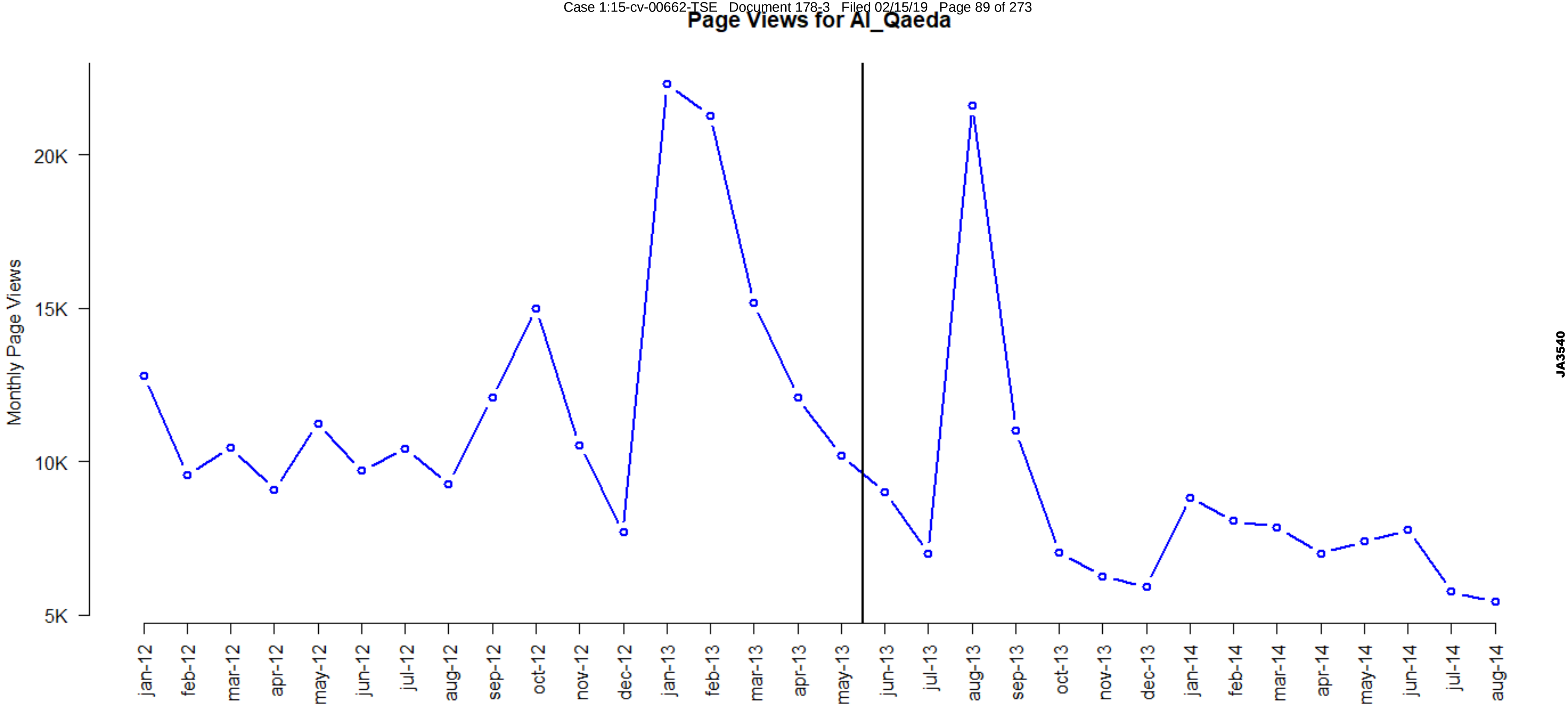
APPENDIX IV: Page Views for 48 Terror Articles, Original Time Period

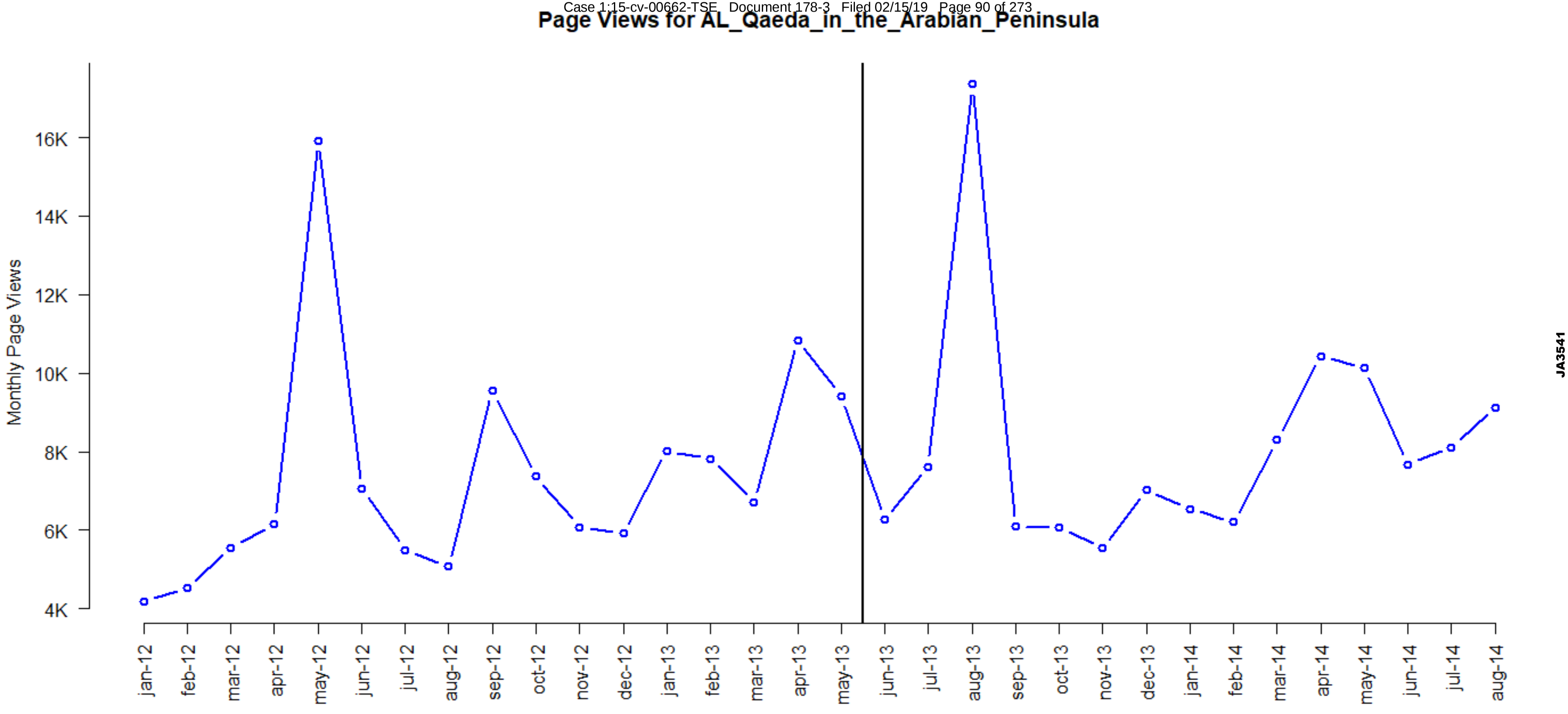

Page Views for _Euskadi_ta_Askatasuna
Monthly Page Views
1.8K
1.6K
1.4K
1.2K
1K
0.8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

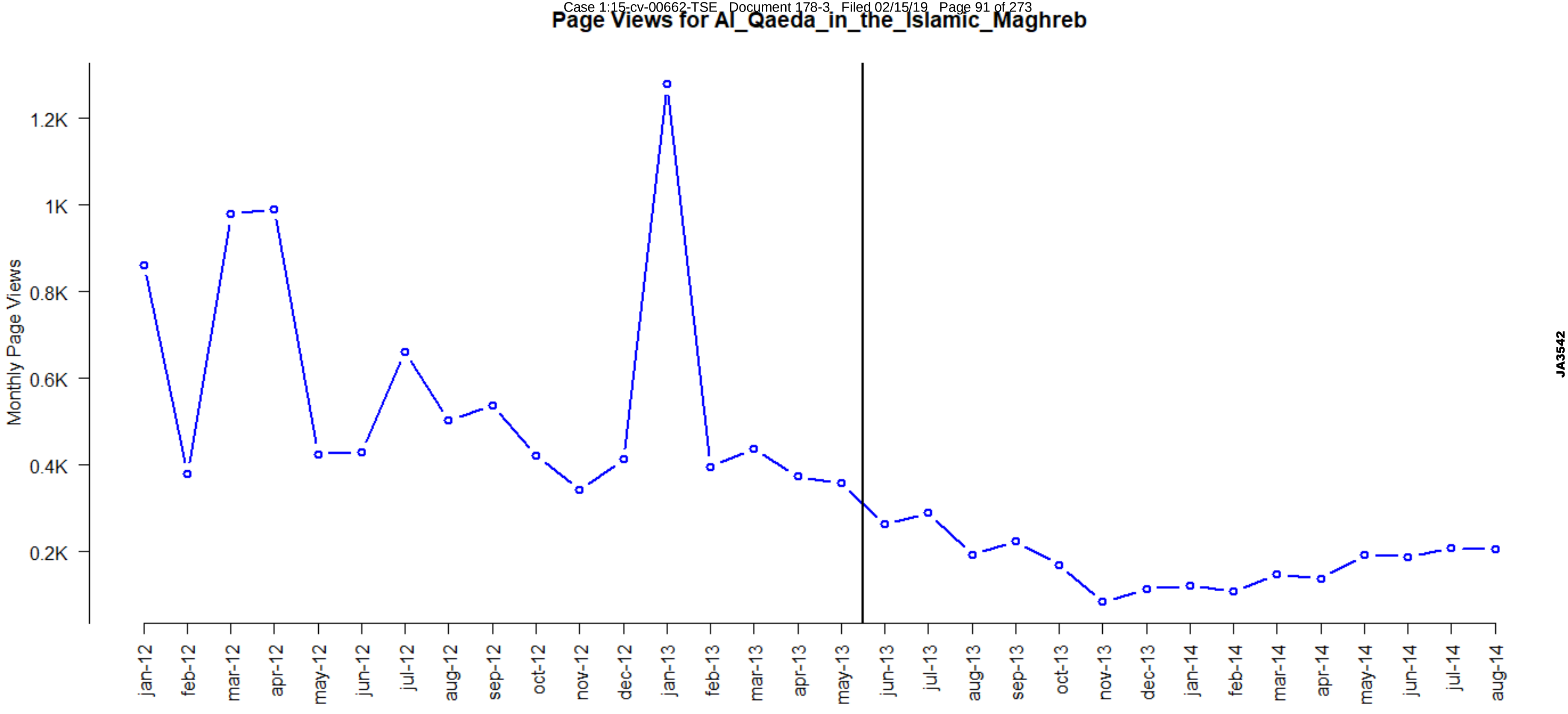

Page Views for Abu_Sayyaf
Monthly Page Views
30K
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

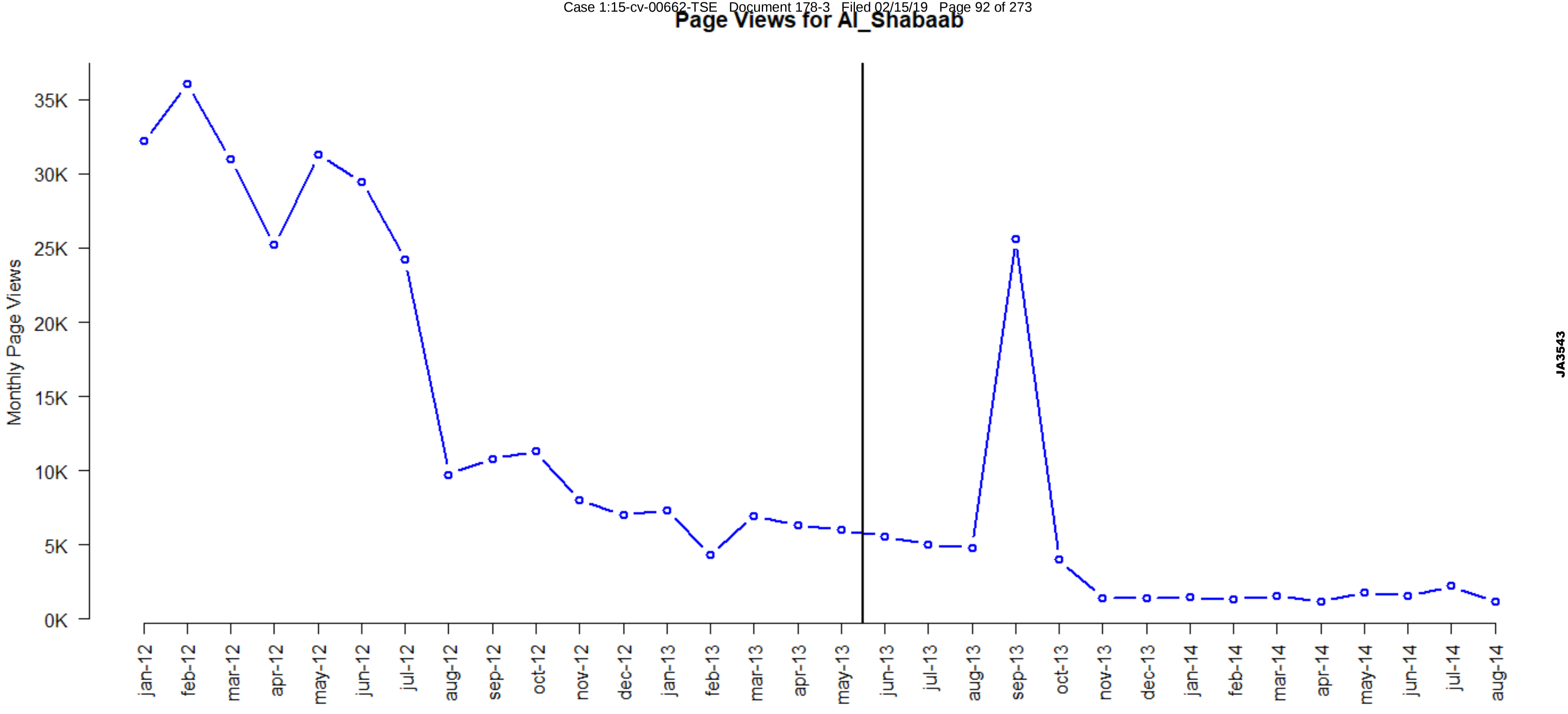

Page Views for Afghanistan
Monthly Page Views
250K
200K
150K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

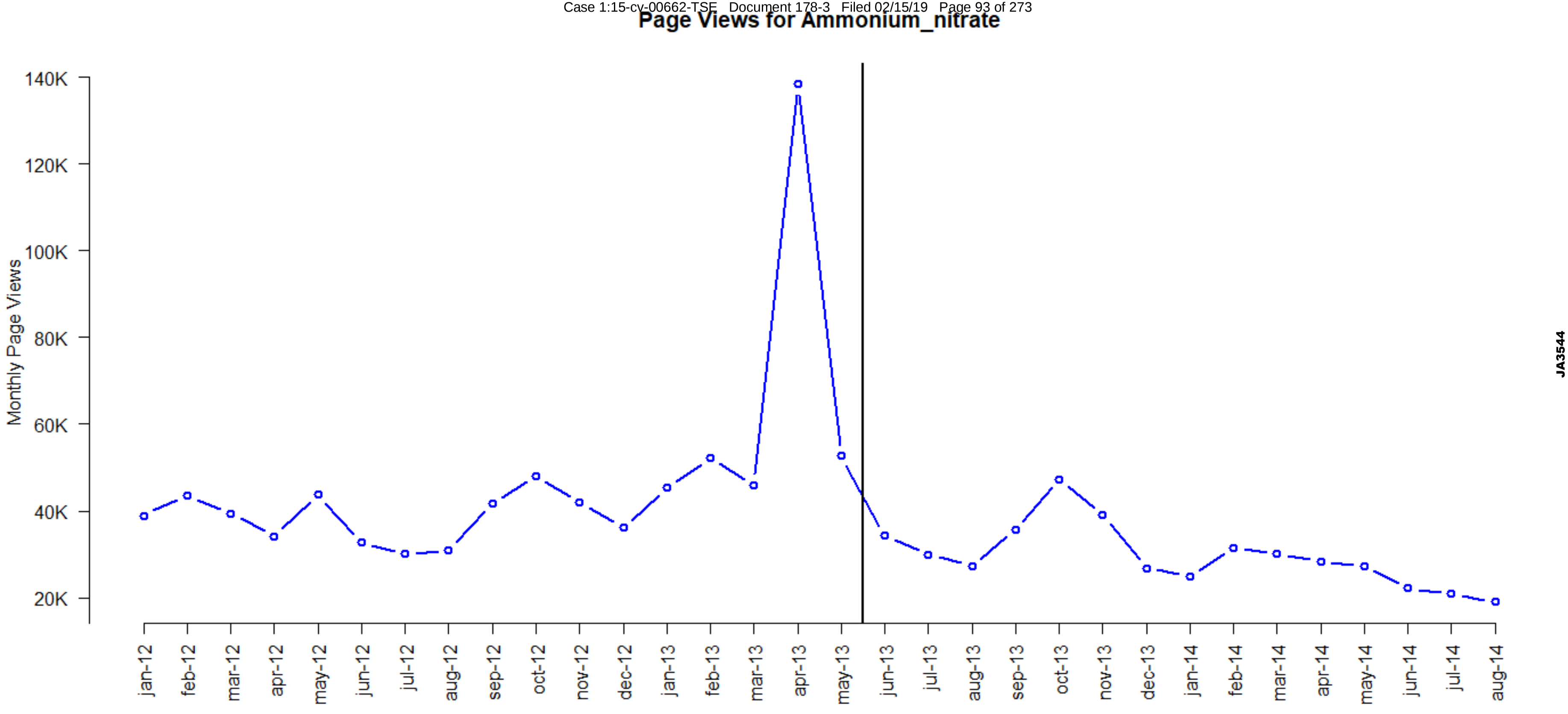

Page Views for agro
Monthly Page Views
4K
3K
2K
1K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

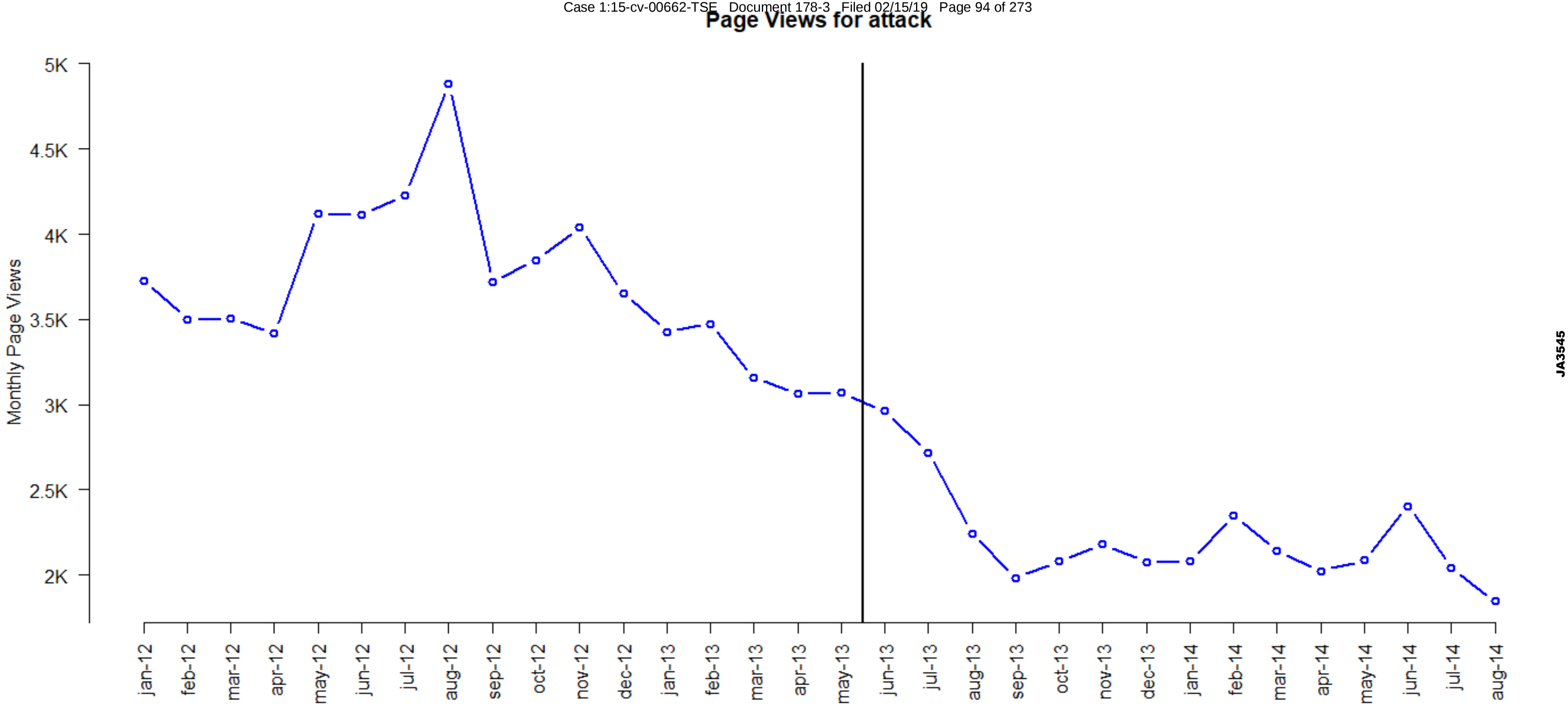

Page Views for Al_Qaeda
Monthly Page Views
20K
15K
10K
5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

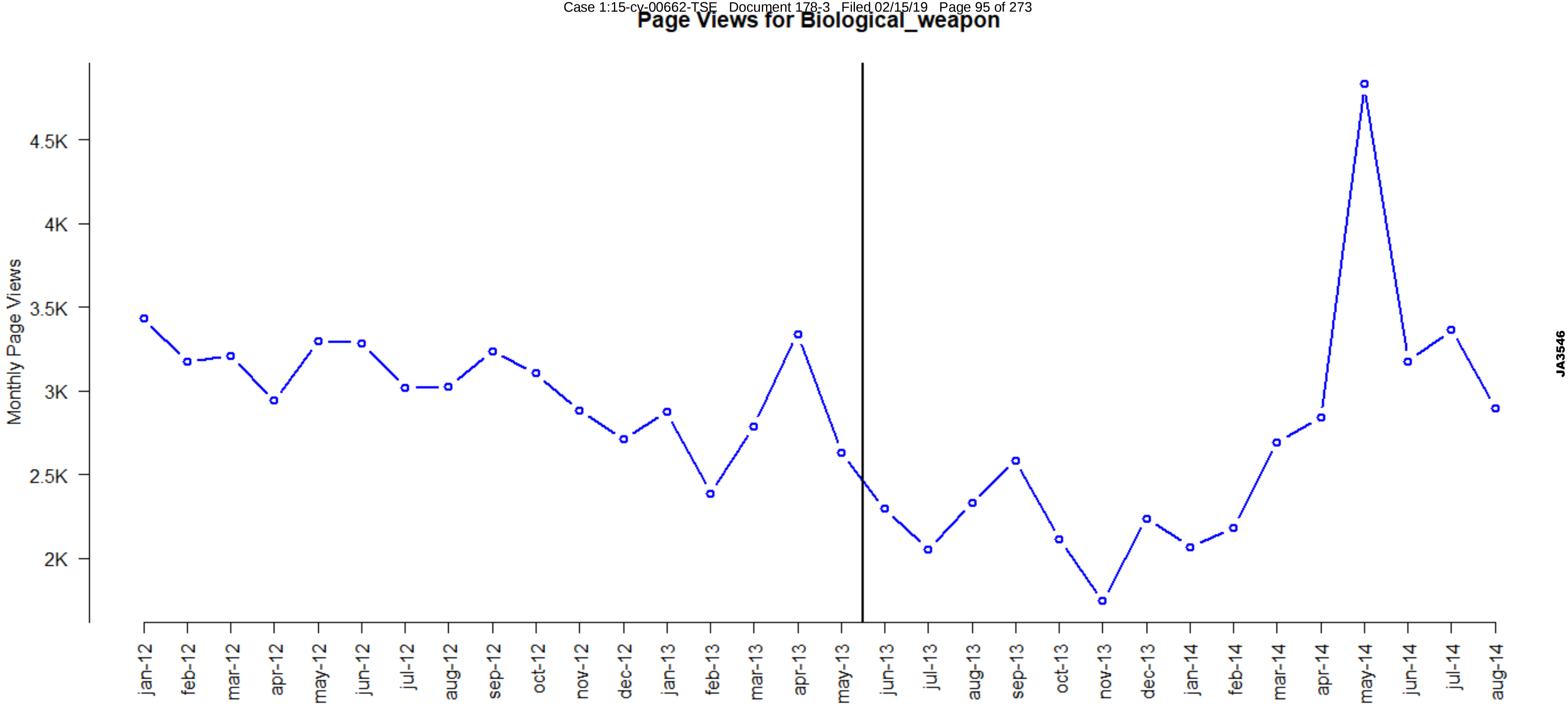

Page Views for AL_Qaeda_in_the_Arabian_Peninsula
Monthly Page Views
16K
14K
12K
10K
8K
6K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

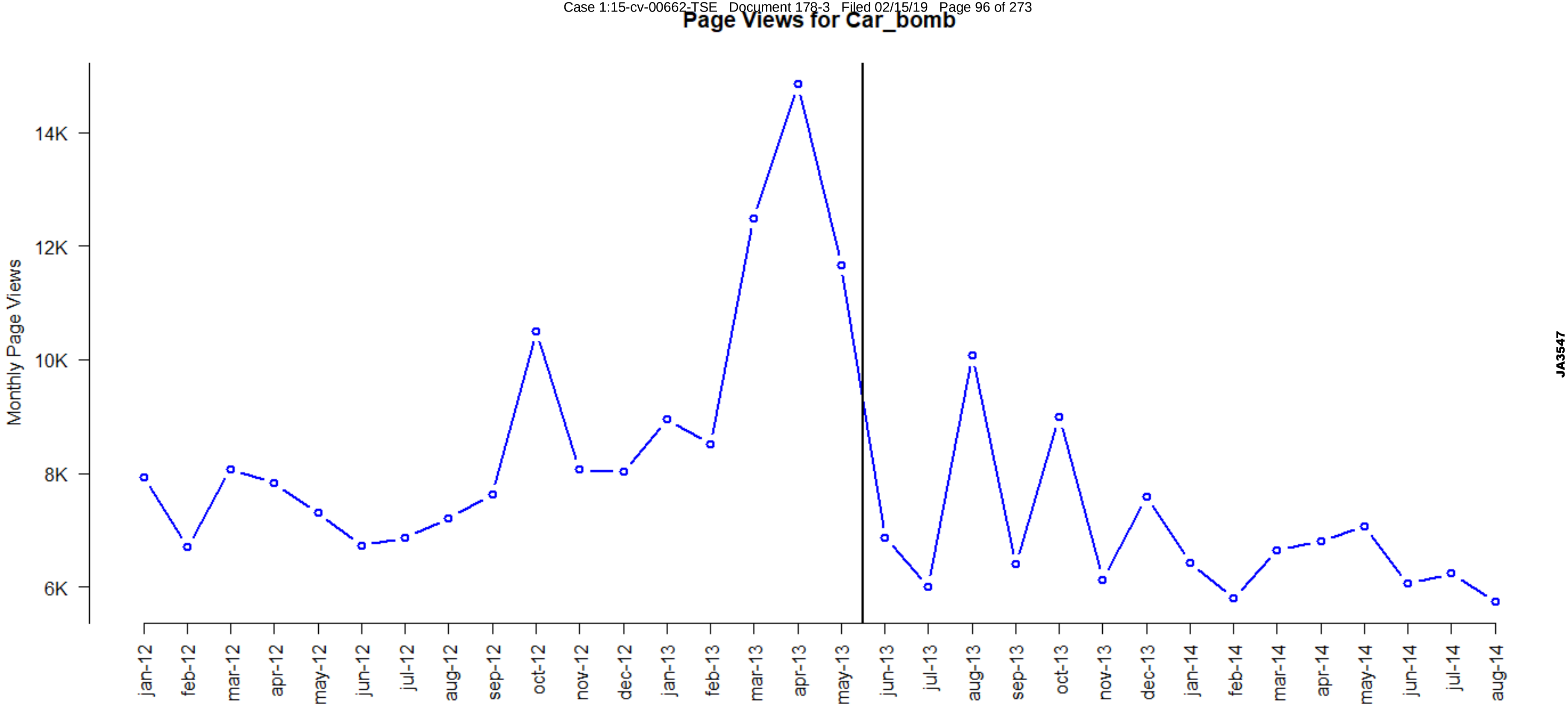

Page Views for Al_Qaeda_in_the_Islamic_Maghreb
Monthly Page Views
1.2K
1K
0.8K
0.6K
0.4K
0.2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

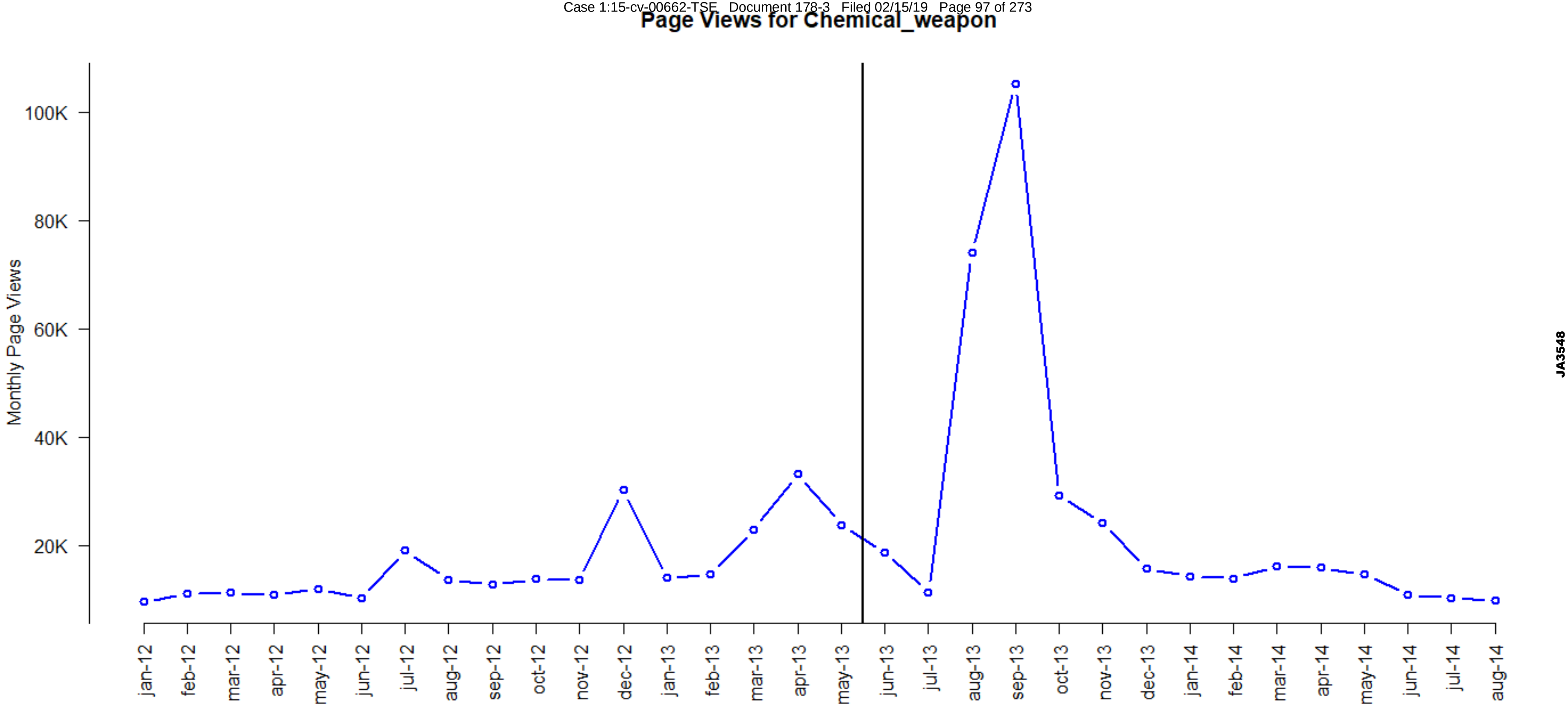

Page Views for Al_Shabaab
Monthly Page Views
35K
30K
25K
20K
15K
10K
5K
0K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

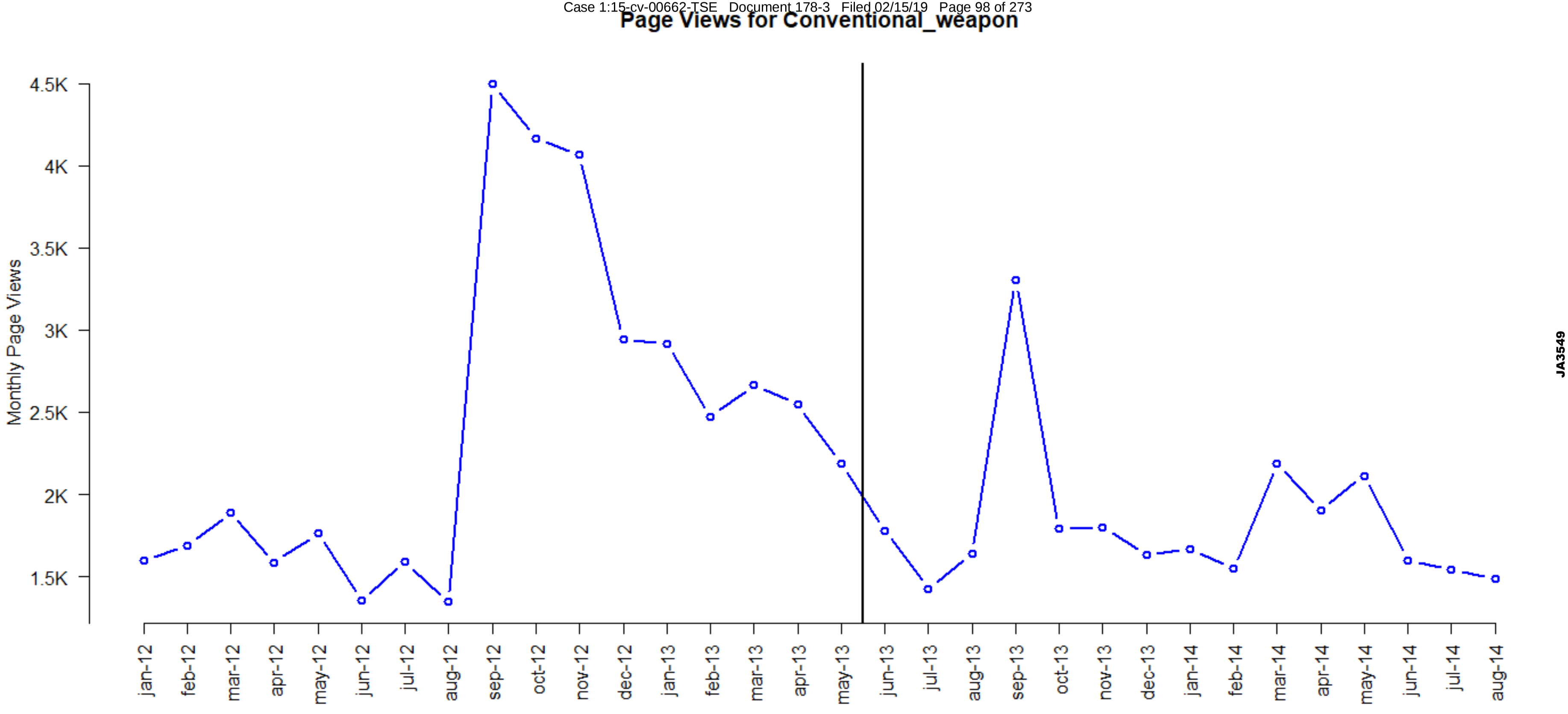

Page Views for Ammonium_nitrate
Monthly Page Views
140K
120K
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

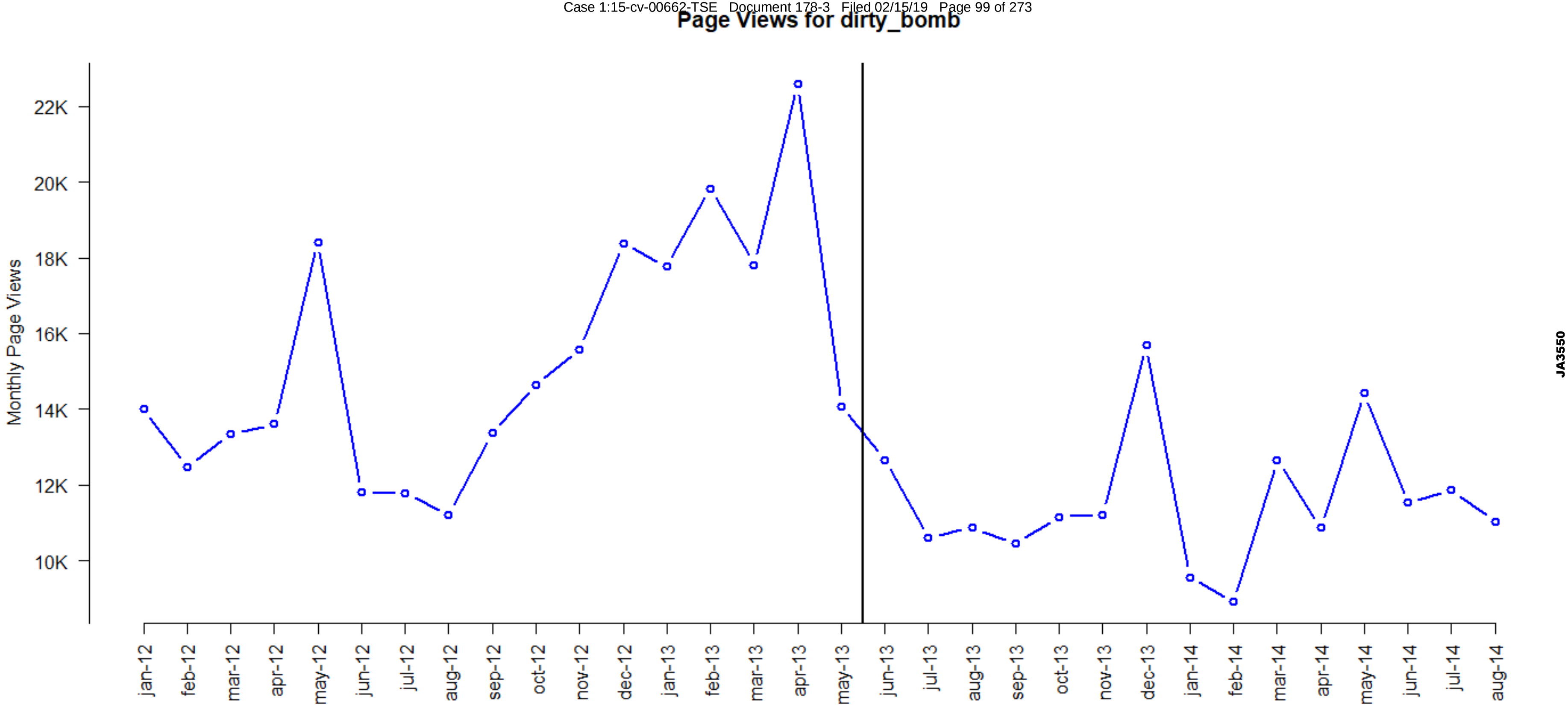

Page Views for attack
Monthly Page Views
5K
4.5K
4K
3.5K
3K
2.5K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

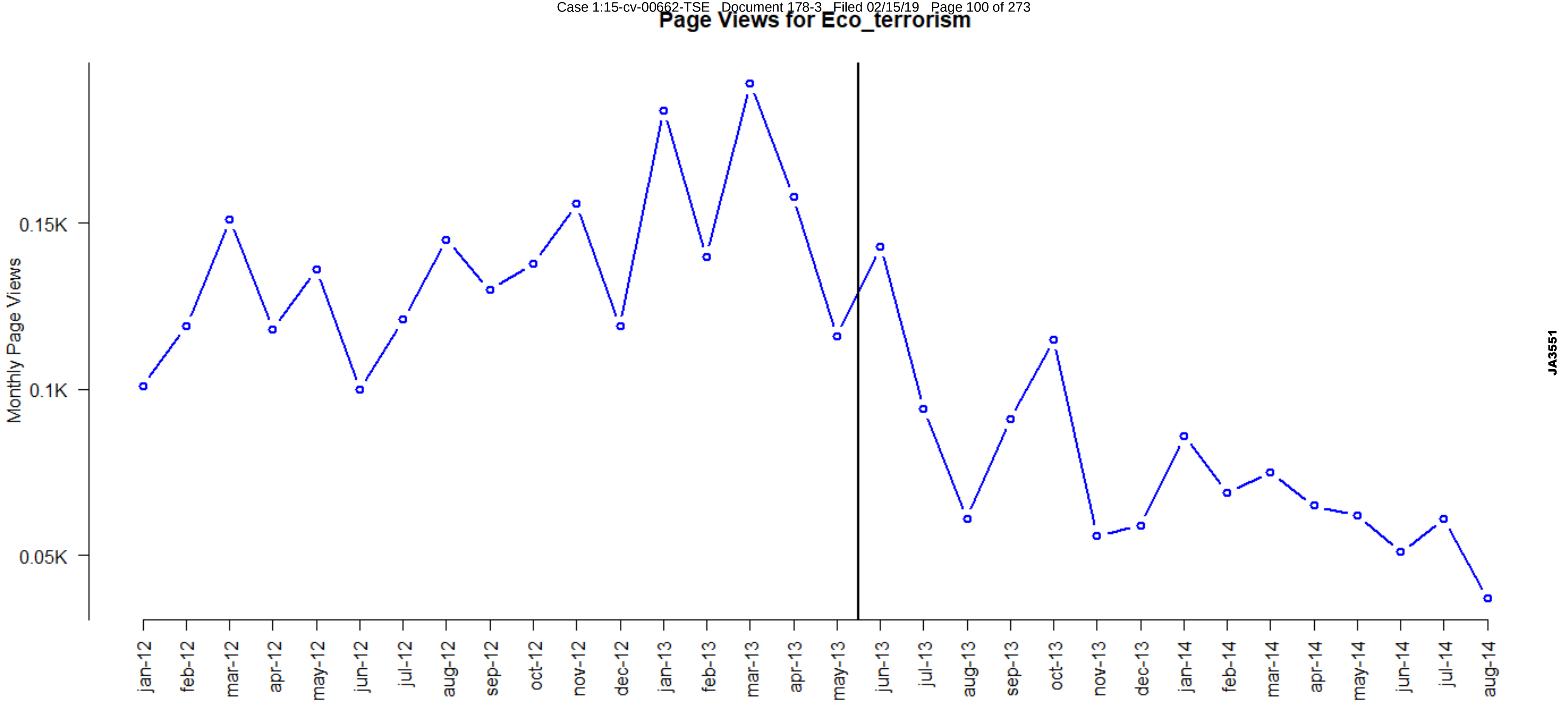

Page Views for Biological_weapon
Monthly Page Views
4.5K
4K
3.5K
3K
2.5K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

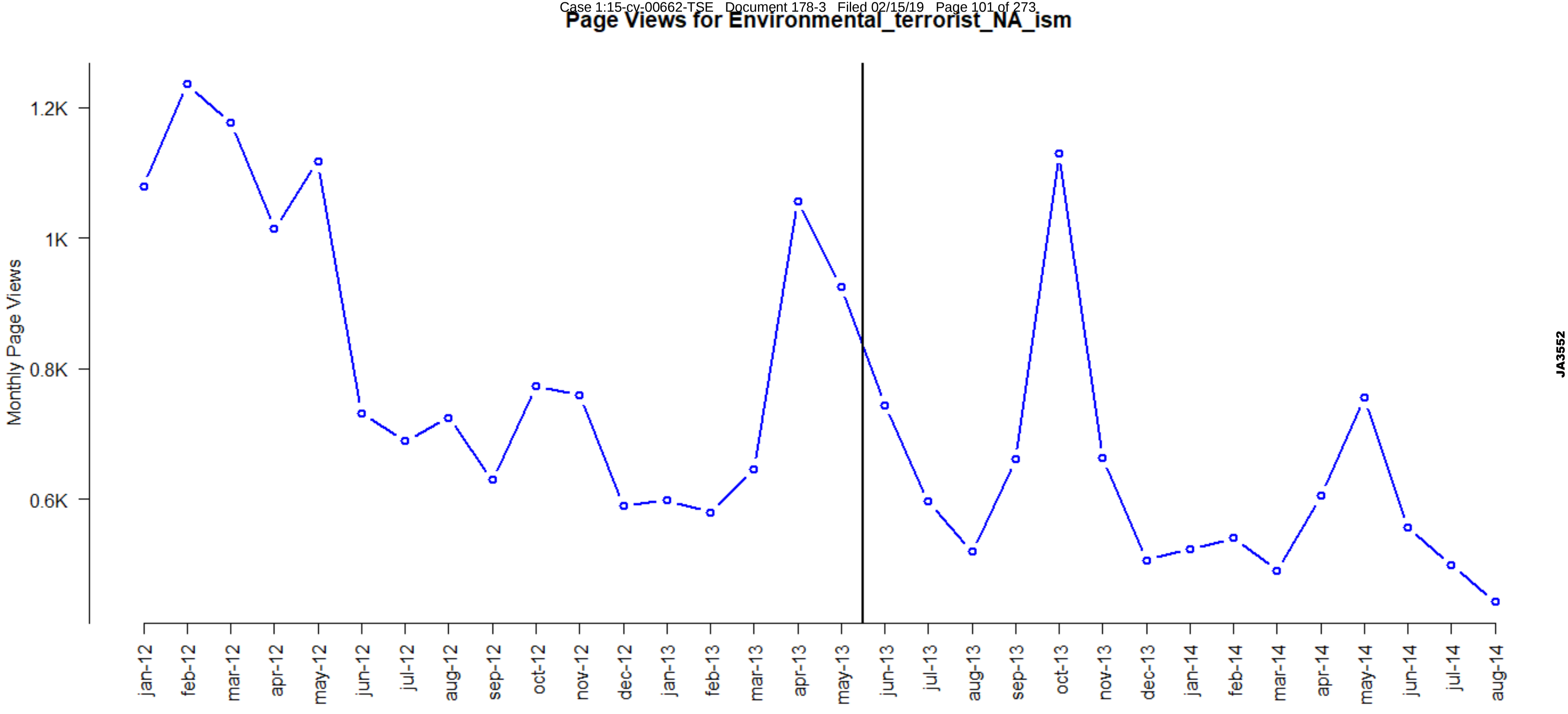

Page Views for Car_bomb
Monthly Page Views
14K
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

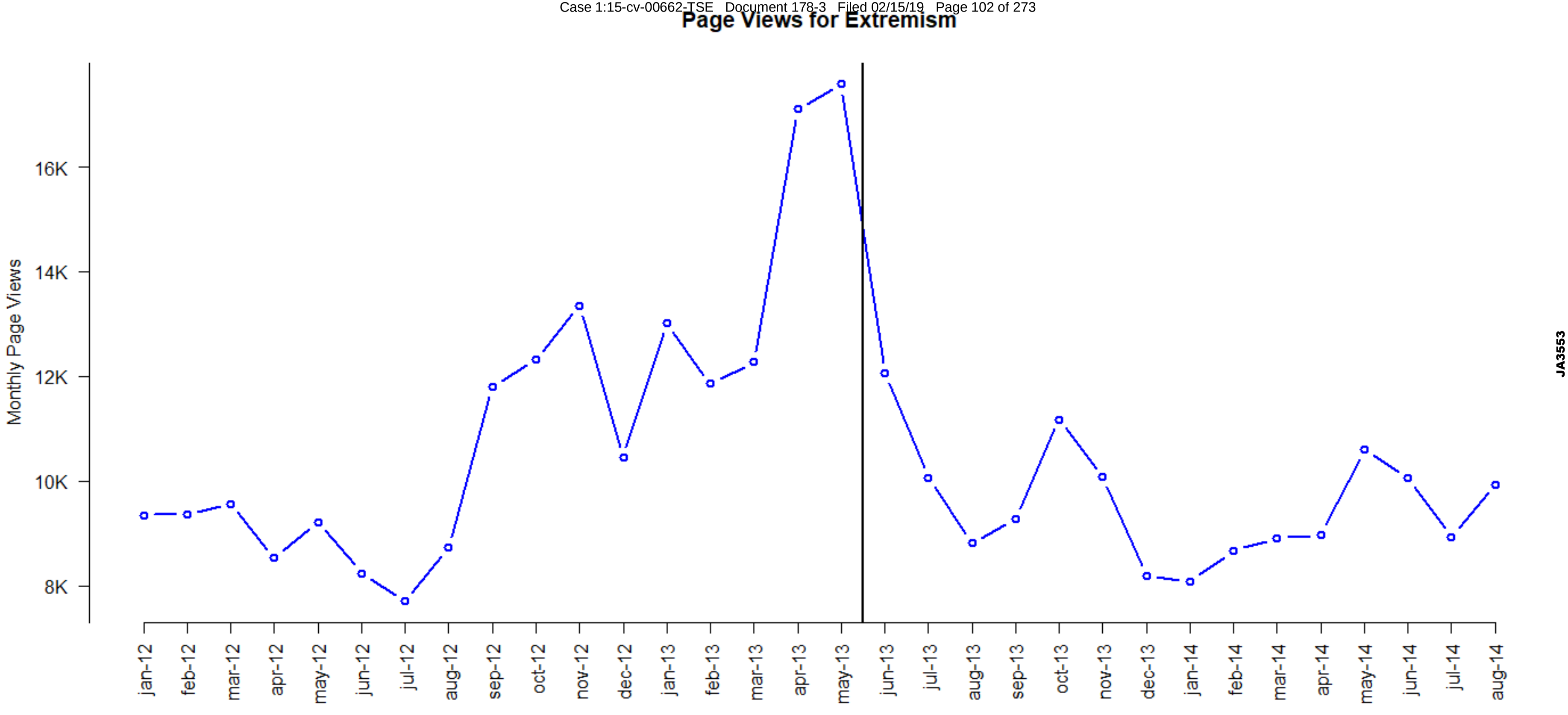

Page Views for Chemical_weapon
Monthly Page Views
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

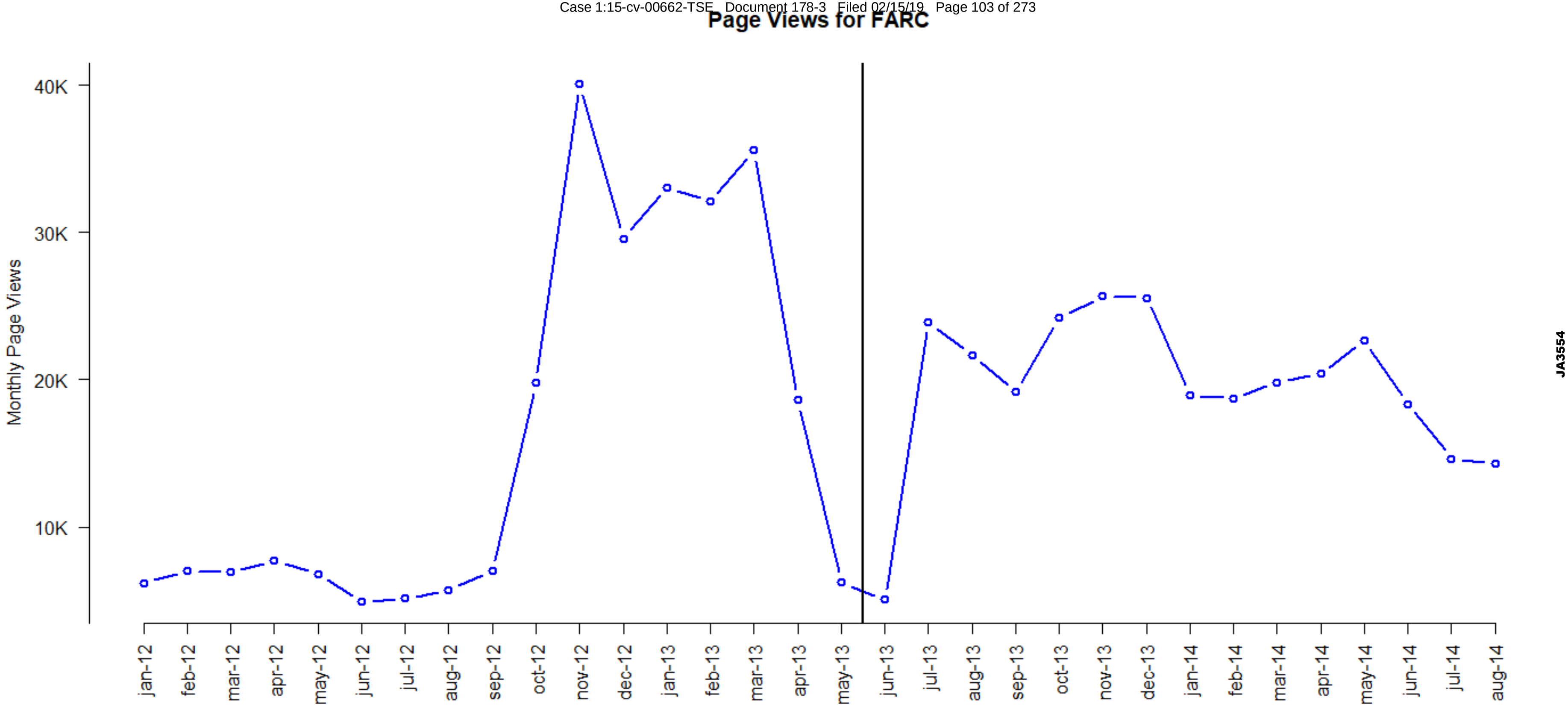

Page Views for Conventional_weapon
Monthly Page Views
4.5K
4K
3.5K
3K
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

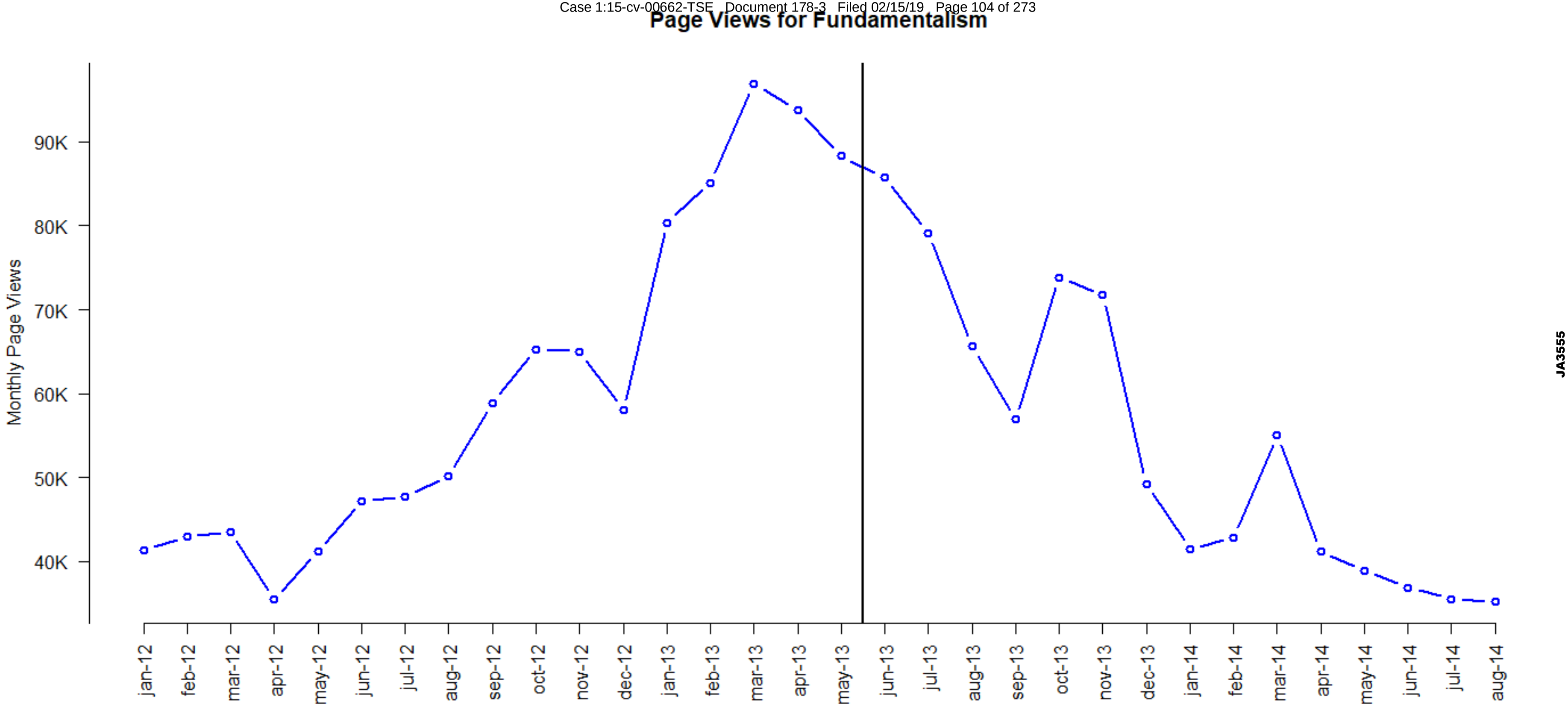

Page Views for dirty_bomb
Monthly Page Views
22K
20K
18K
16K
14K
12K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

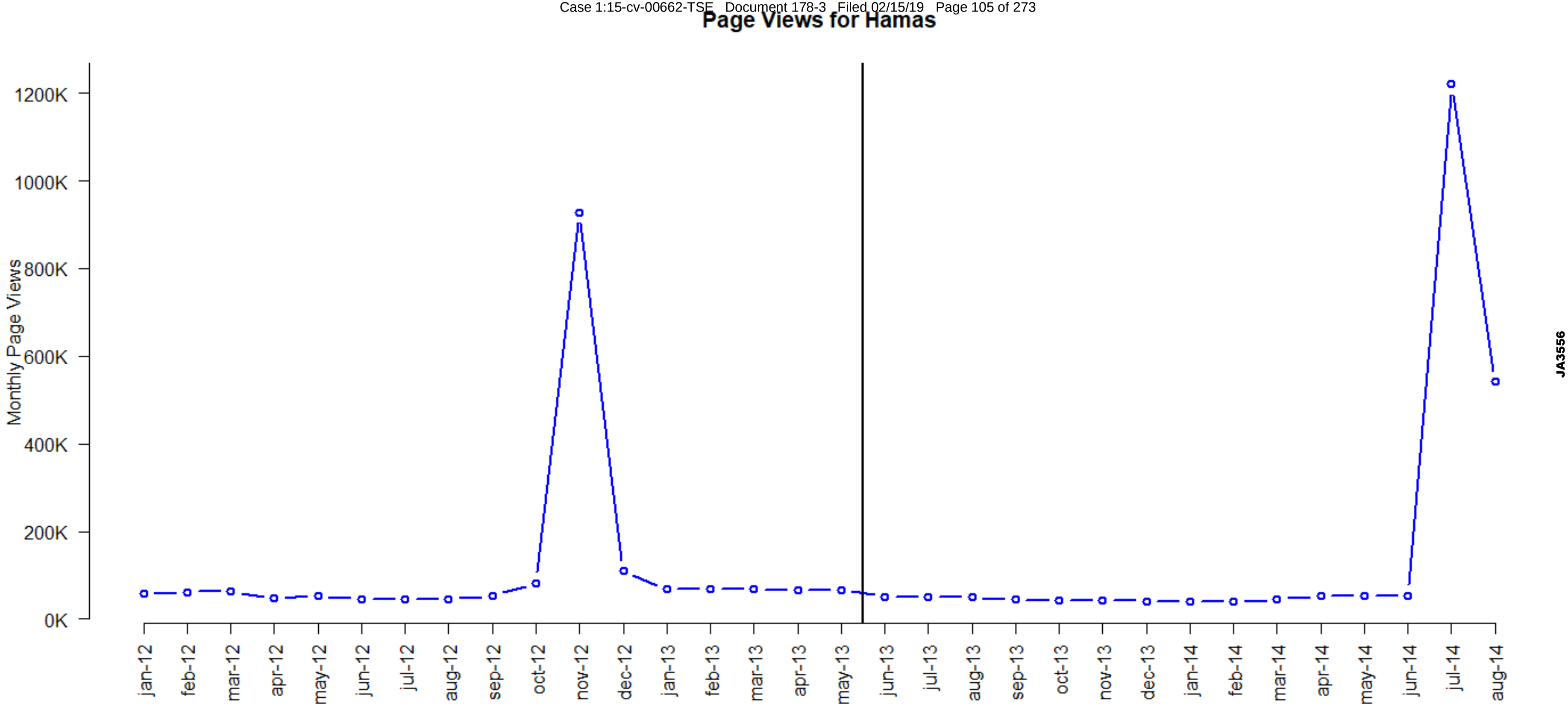

Page Views for Eco_terrorism
Monthly Page Views
0.15K
0.1K
0.05K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

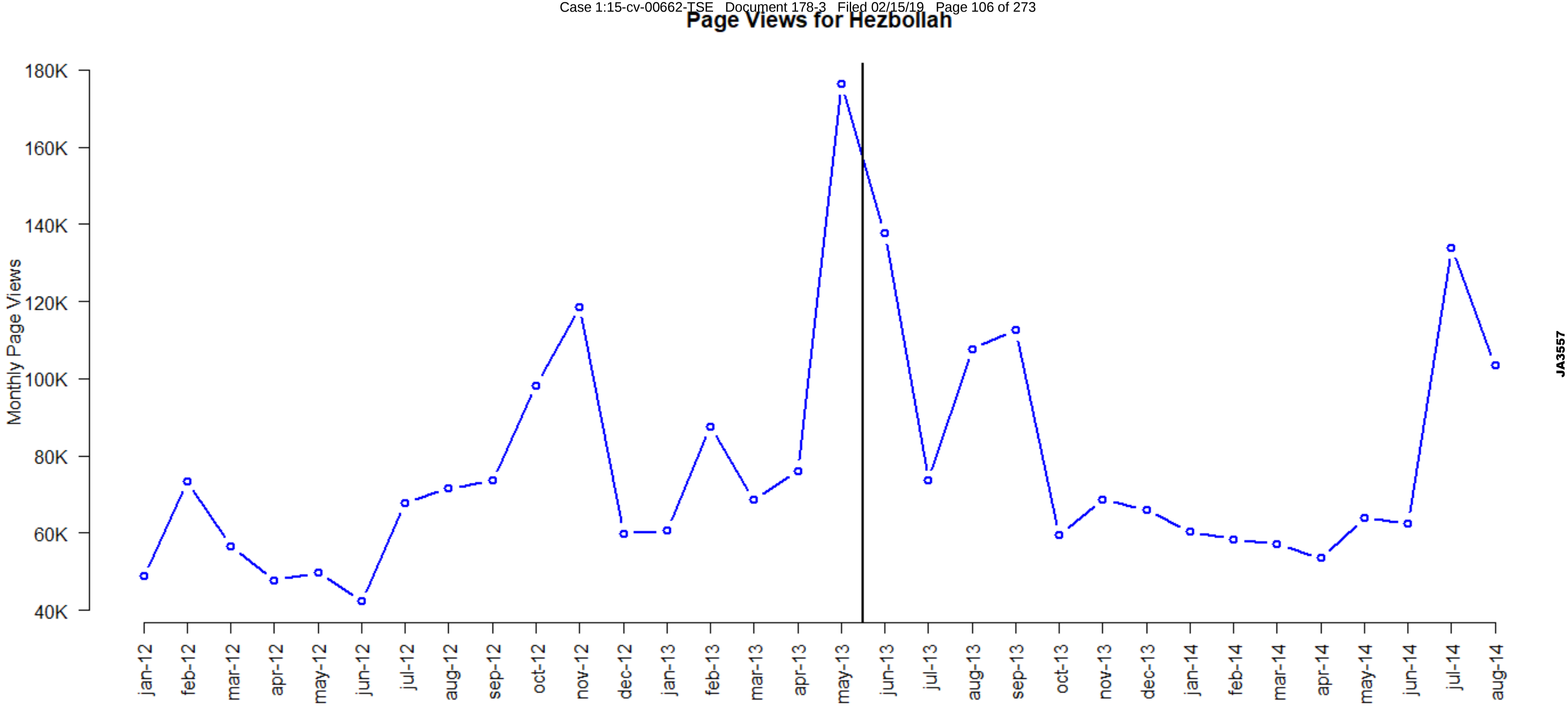

Page Views for Environmental_terrorist_NA_ism
Monthly Page Views
1.2K
1K
0.8K
0.6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

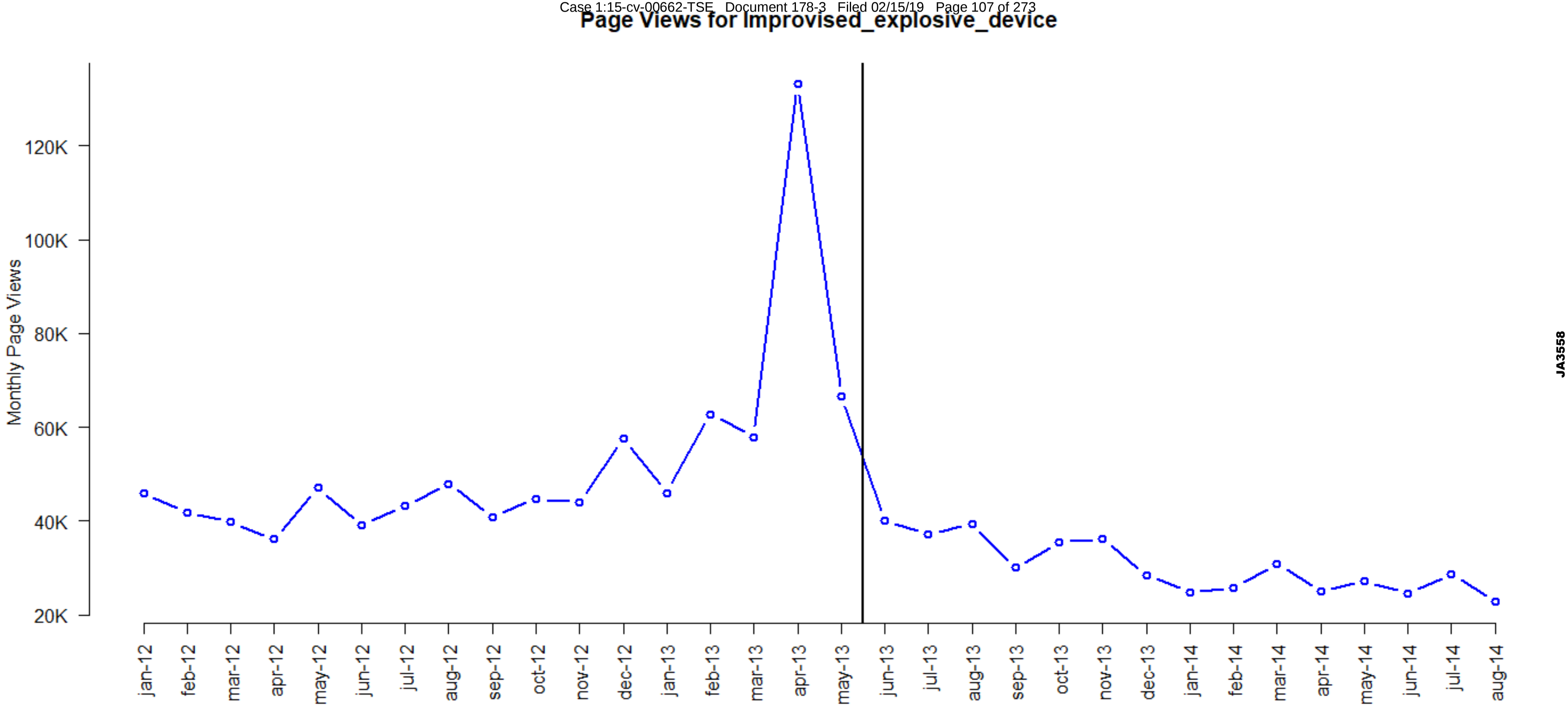

Page Views for Extremism
Monthly Page Views
16K
14K
12K
10K
8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

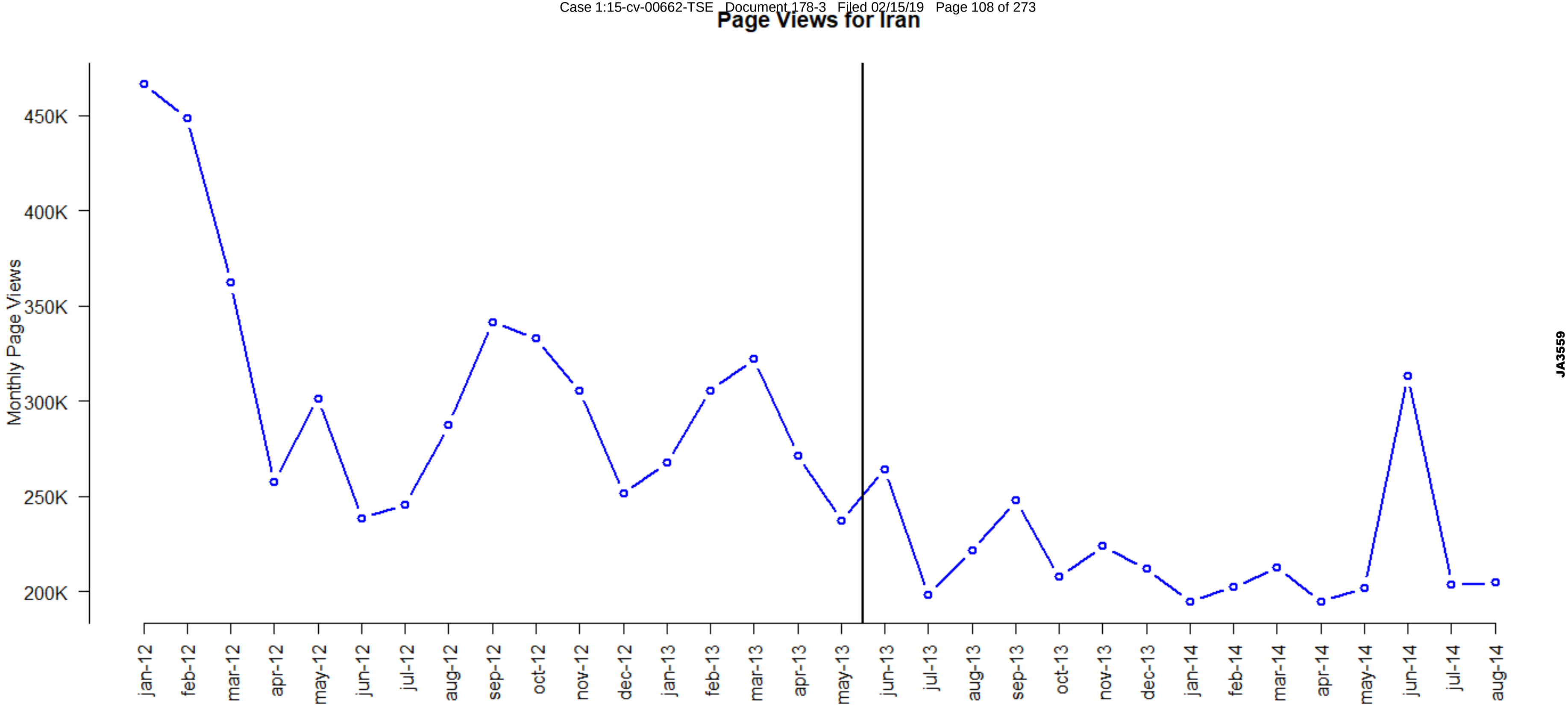

Page Views for FARC
Monthly Page Views
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

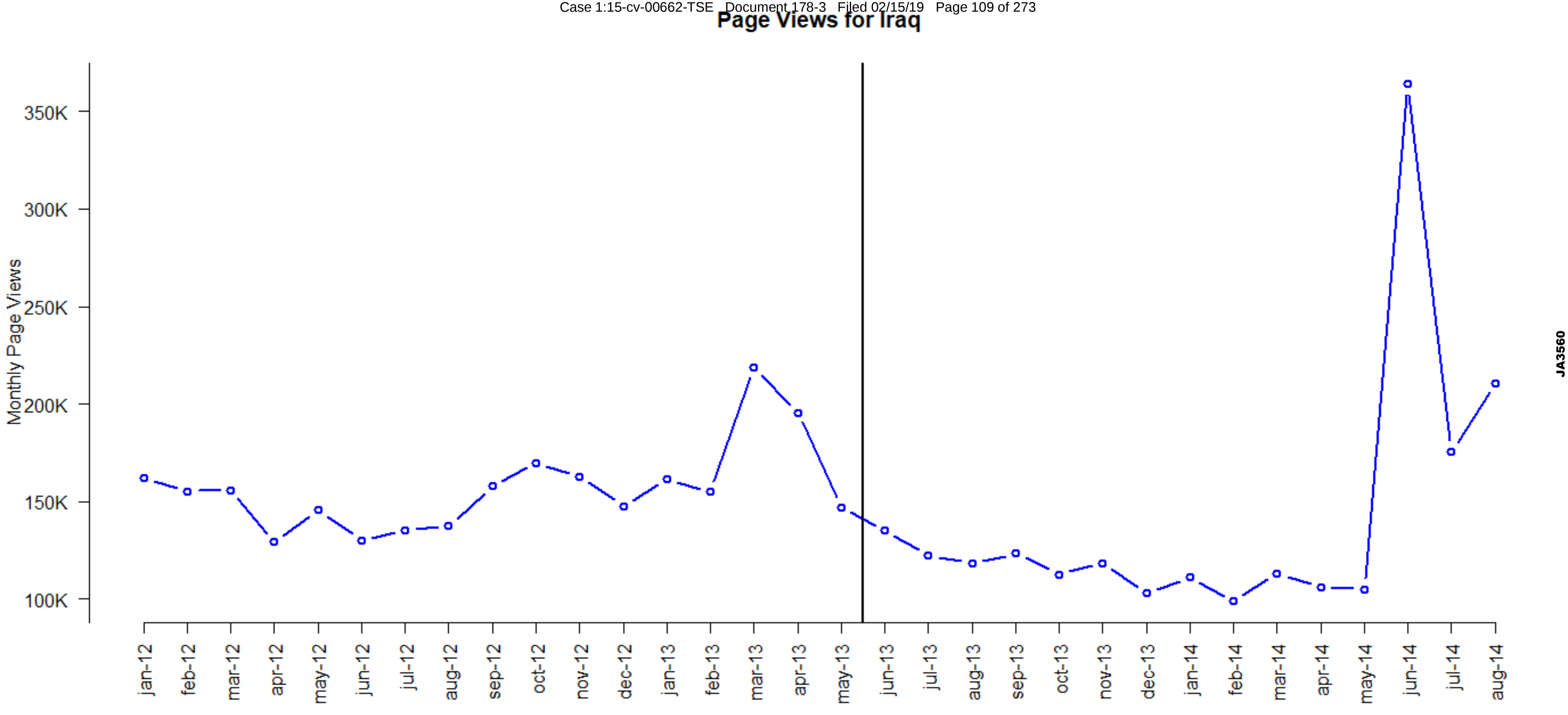

Page Views for Fundamentalism
Monthly Page Views
90K
80K
70K
60K
50K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

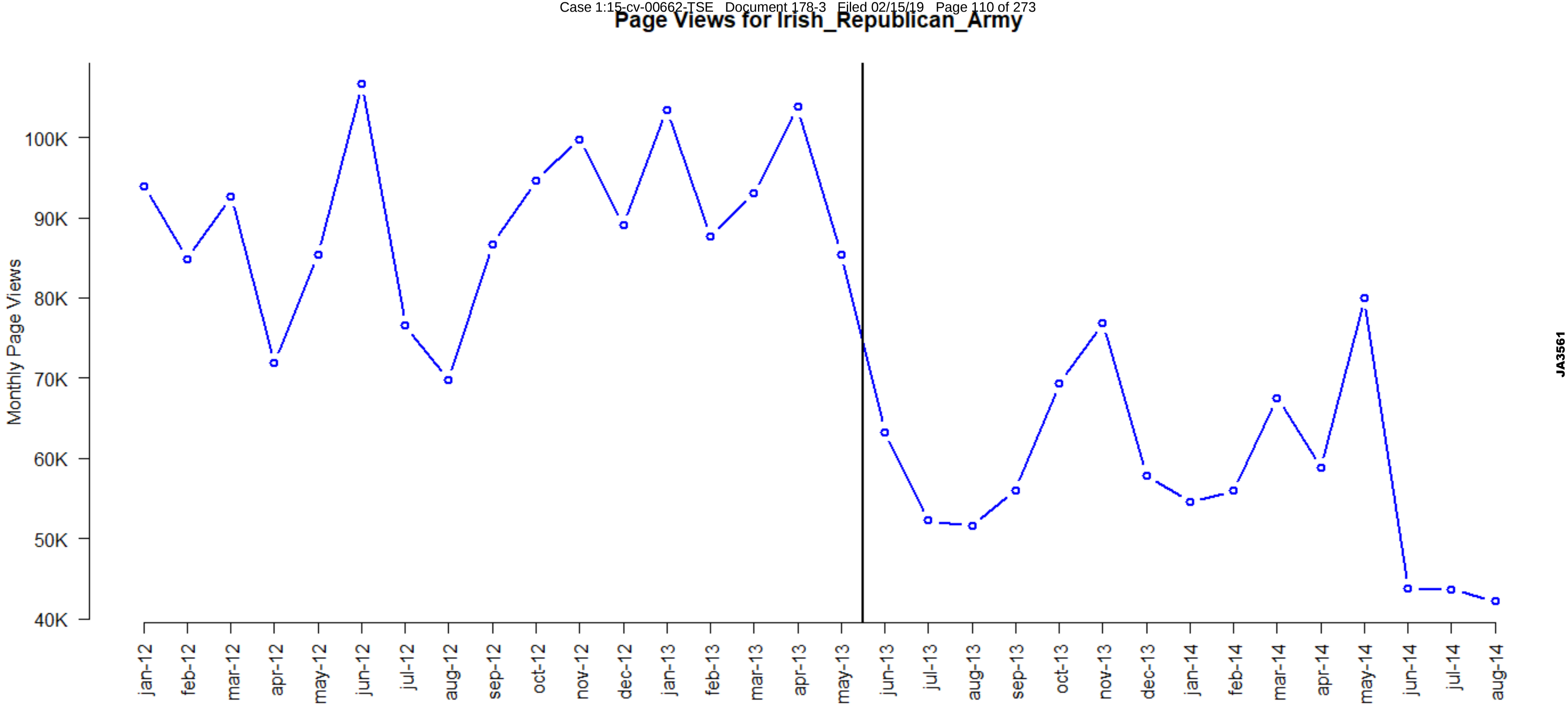

Page Views for Hamas
Monthly Page Views
1200K
1000K
800K
600K
400K
200K
0K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

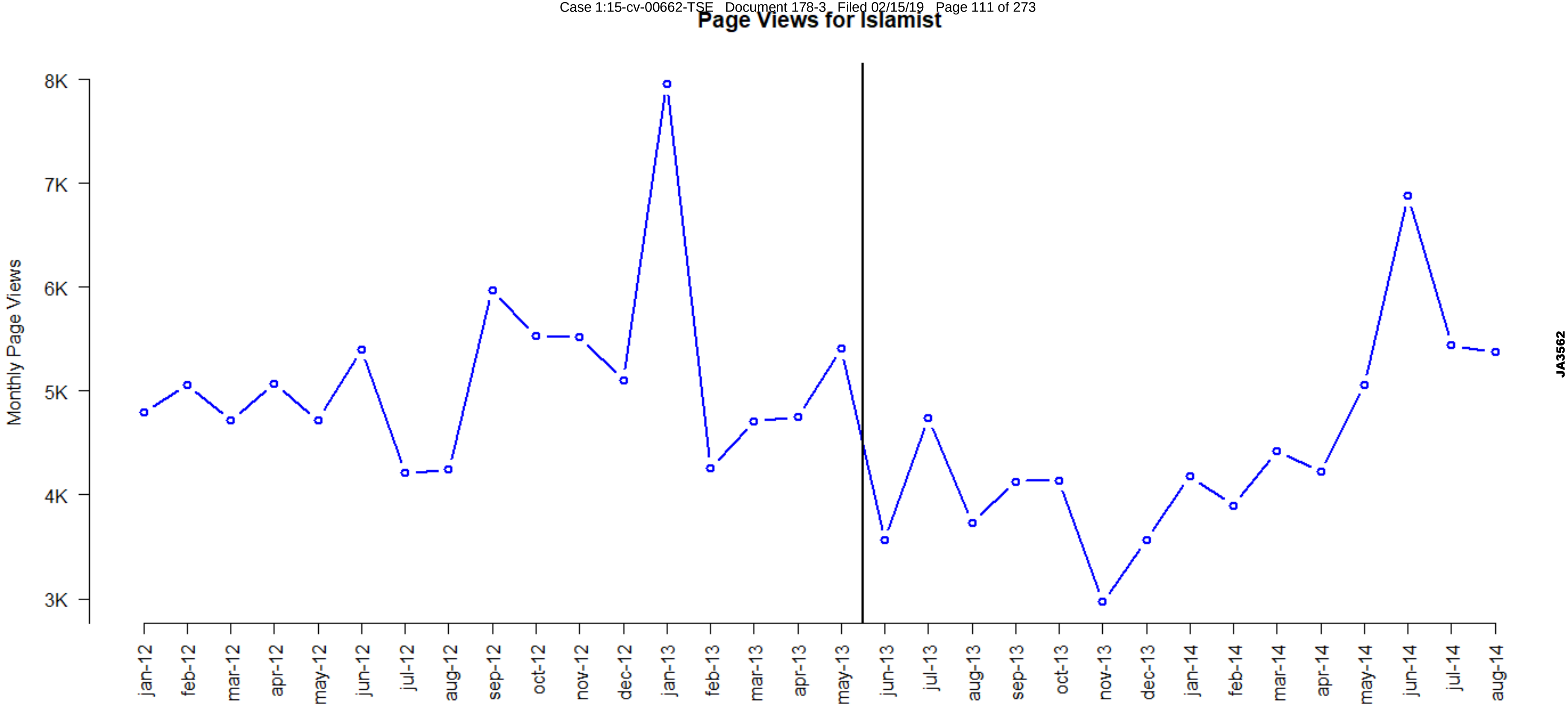

Page Views for Hezbollah
Monthly Page Views
180K
160K
140K
120K
100K
80K
60K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

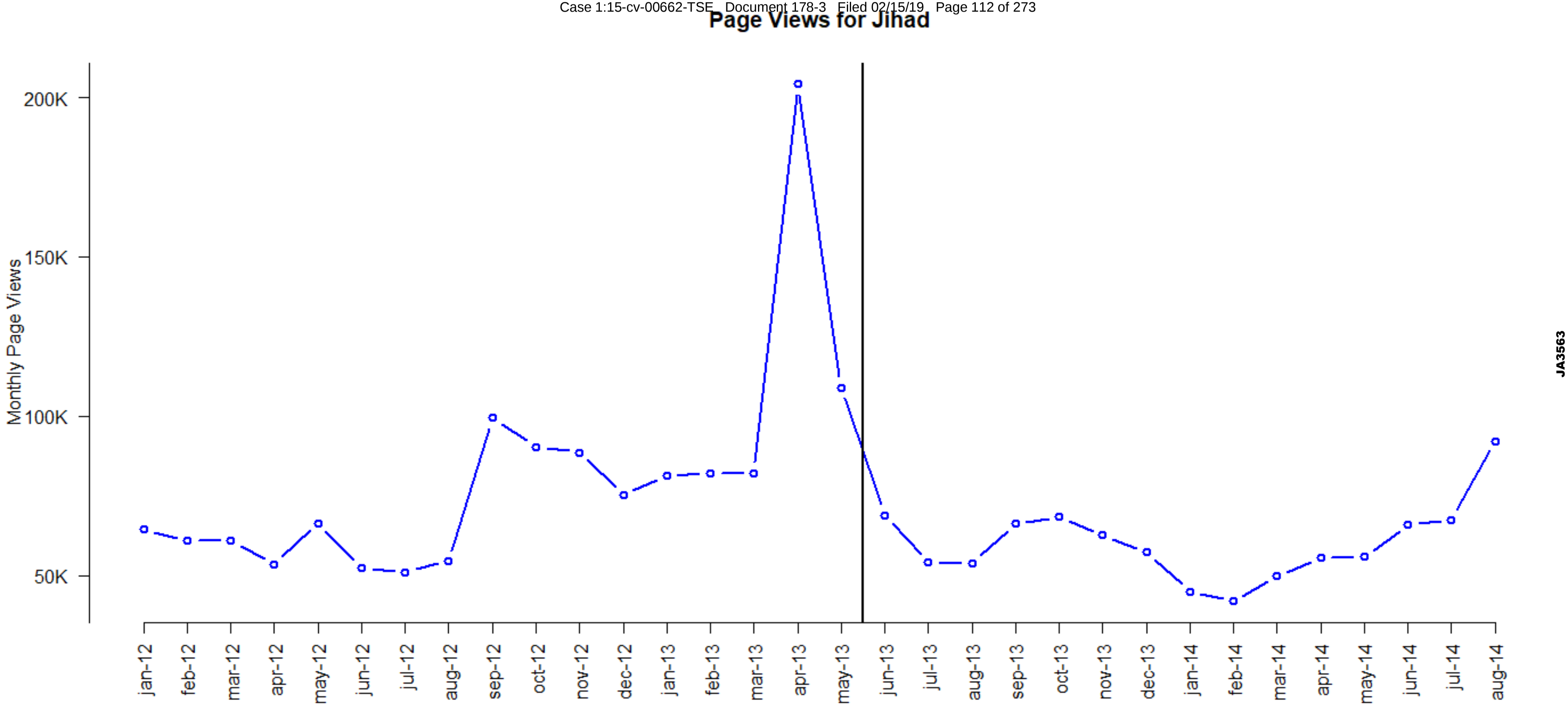

Page Views for Improvised_explosive_device
Monthly Page Views
120K
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

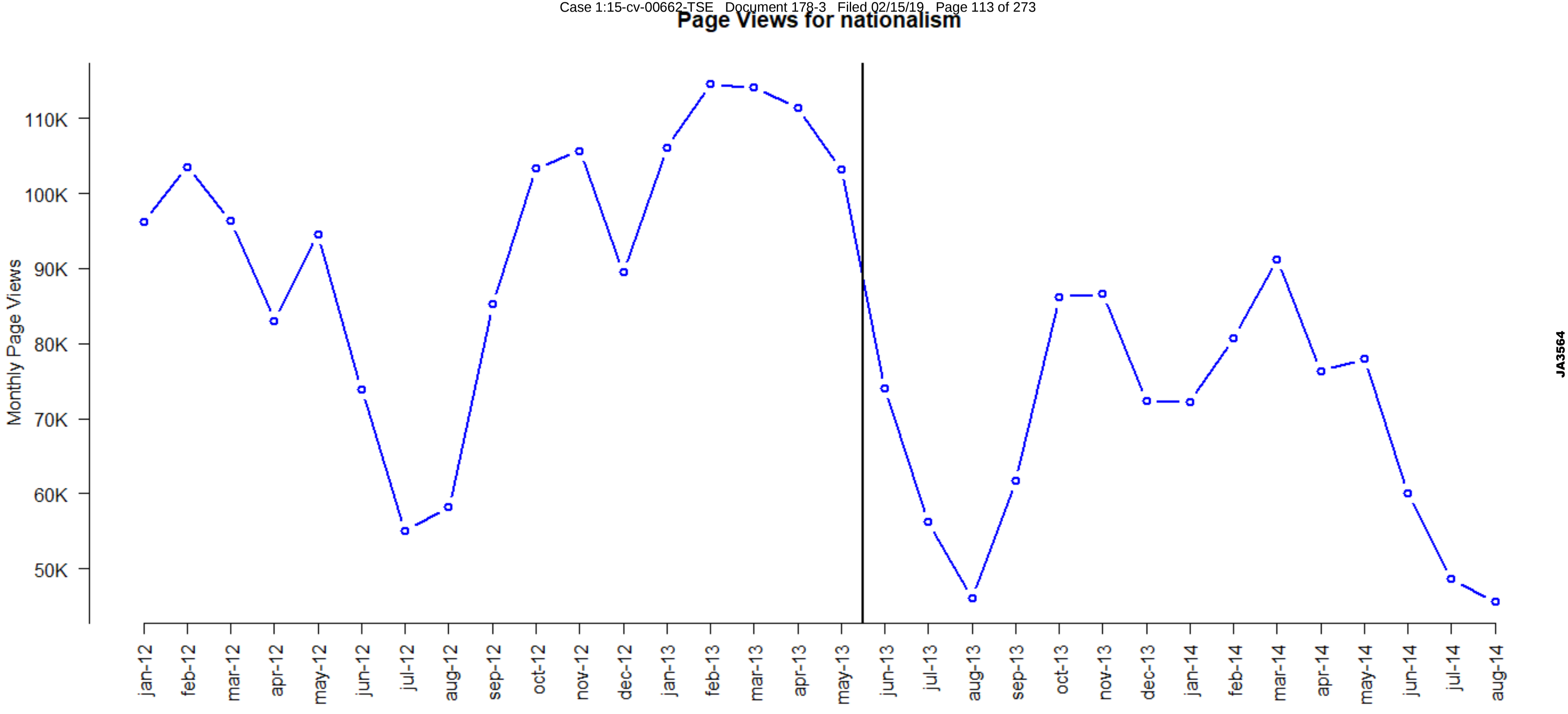

Page Views for Iran
Monthly Page Views
450K
400K
350K
300K
250K
200K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

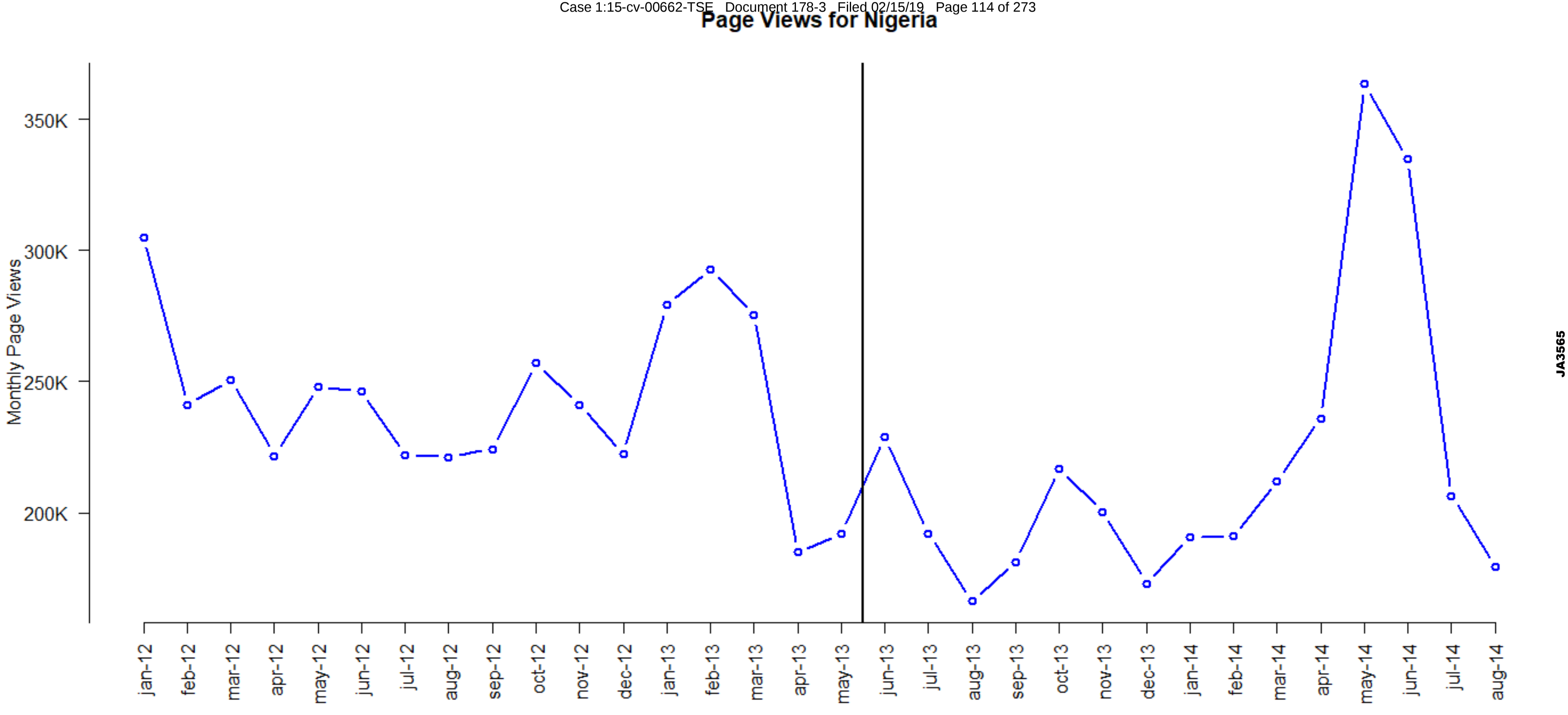

Page Views for Iraq
Monthly Page Views
350K
300K
250K
200K
150K
100K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

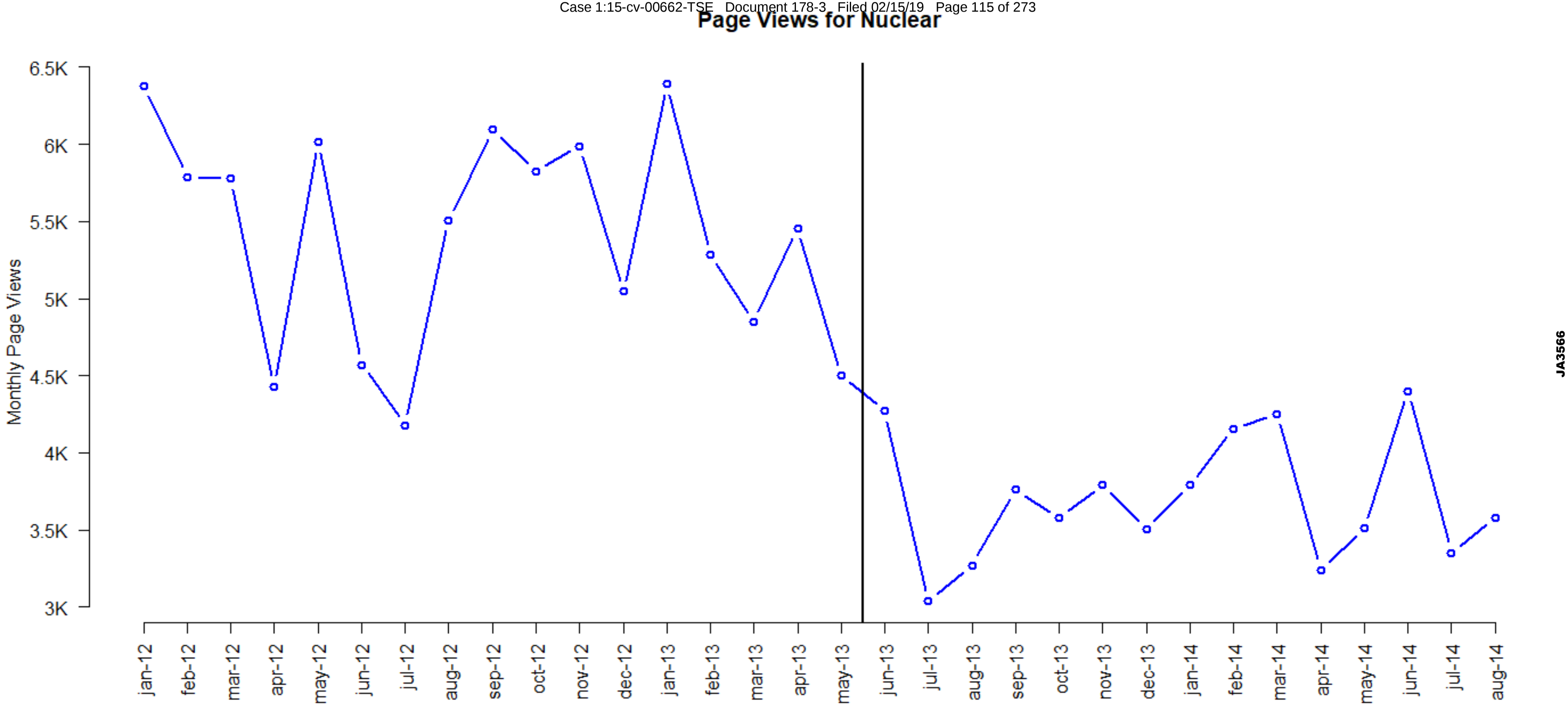

Page Views for Irish_Republican_Army
Monthly Page Views
100K
90K
80K
70K
60K
50K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

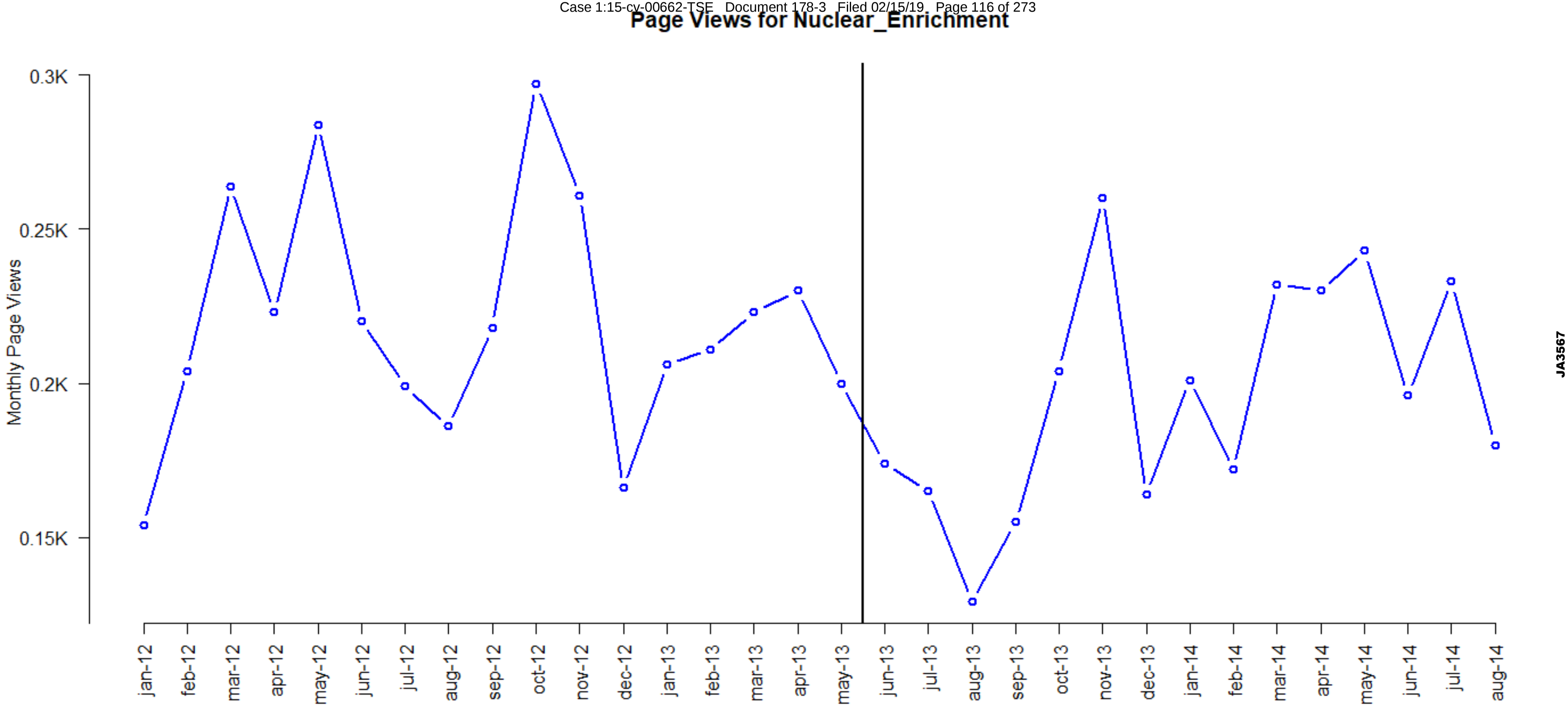

Page Views for Islamist
Monthly Page Views
8K
7K
6K
5K
4K
3K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

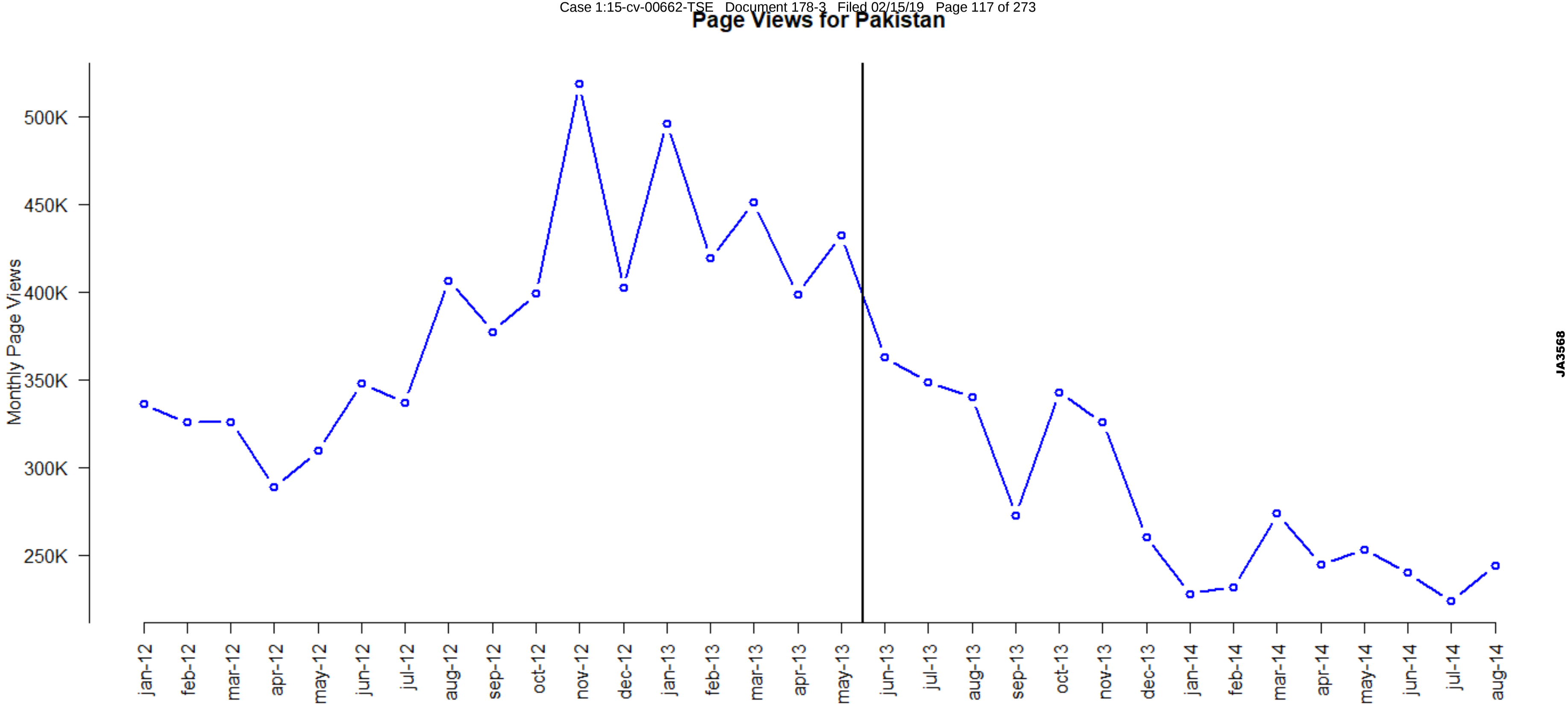

Page Views for Jihad
Monthly Page Views
200K
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

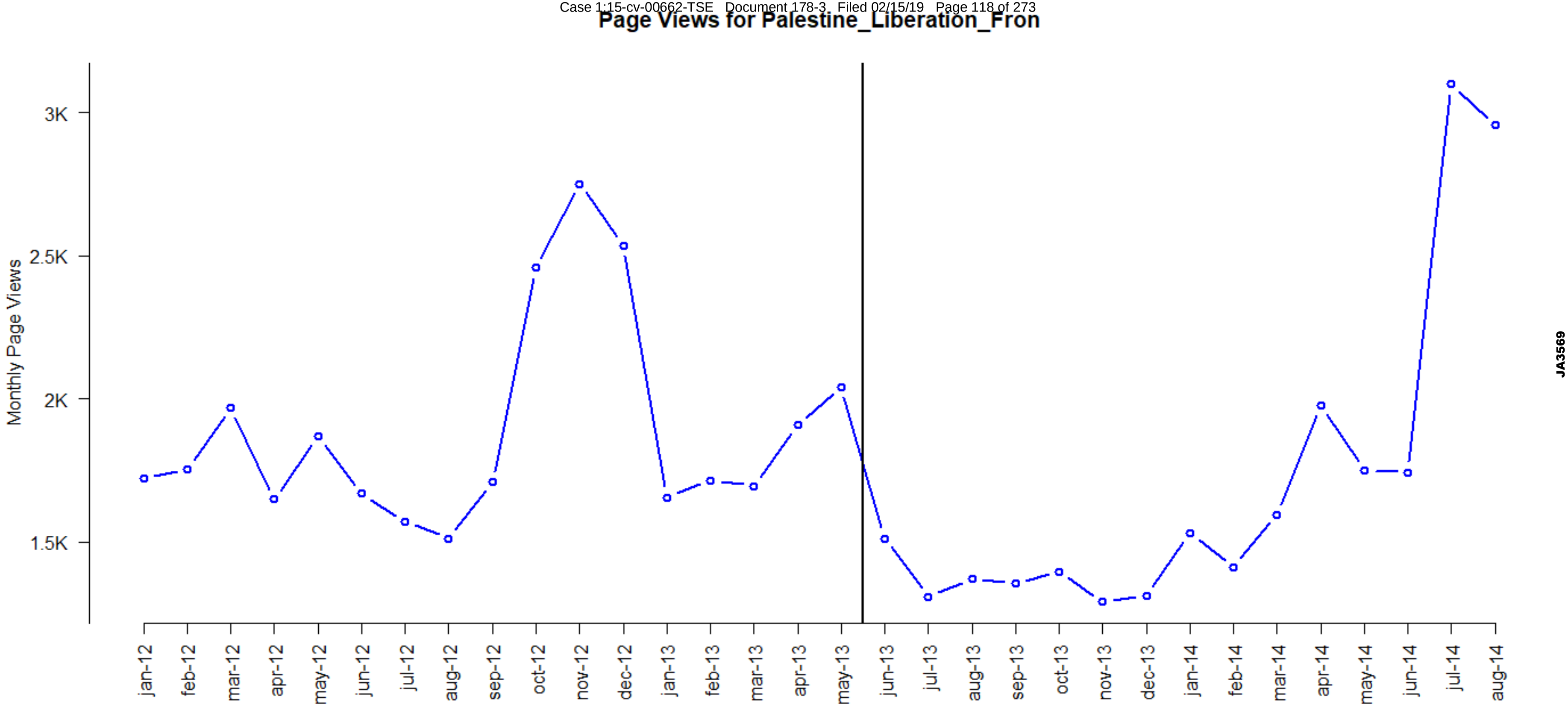

Page Views for nationalism
Monthly Page Views
110K
100K
90K
80K
70K
60K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

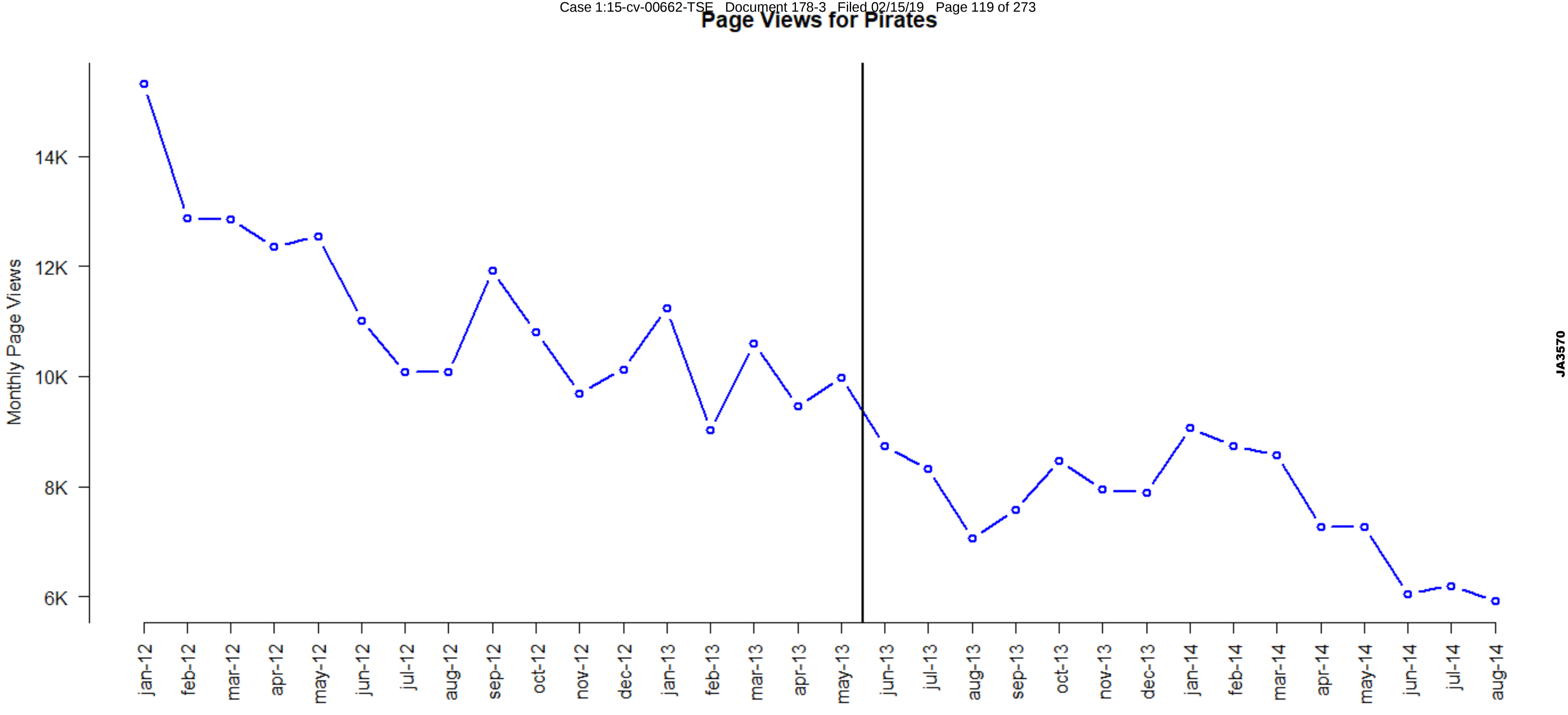

Page Views for Nigeria
Monthly Page Views
350K
300K
250K
200K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

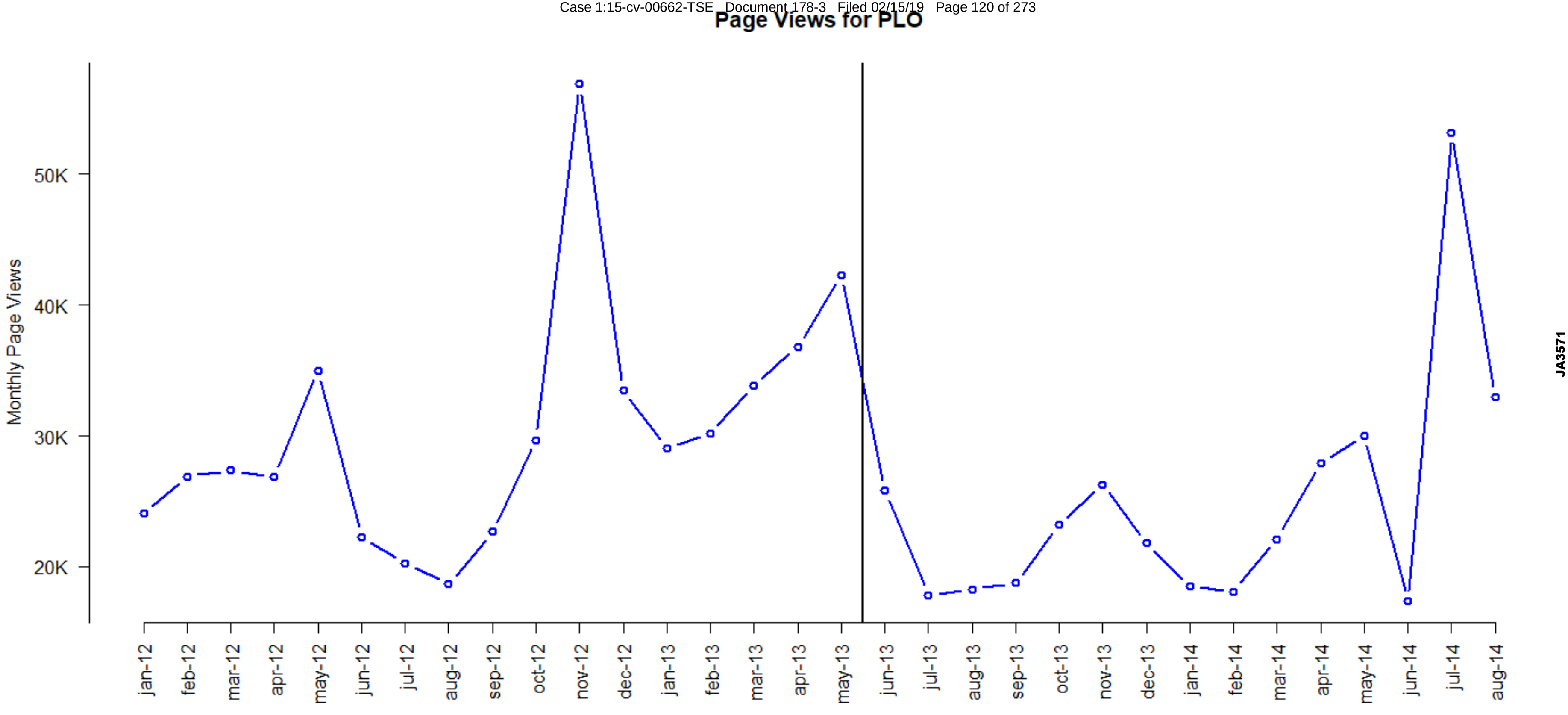

Page Views for Nuclear
Monthly Page Views
6.5K
6K
5.5K
5K
4.5K
4K
3.5K
3K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

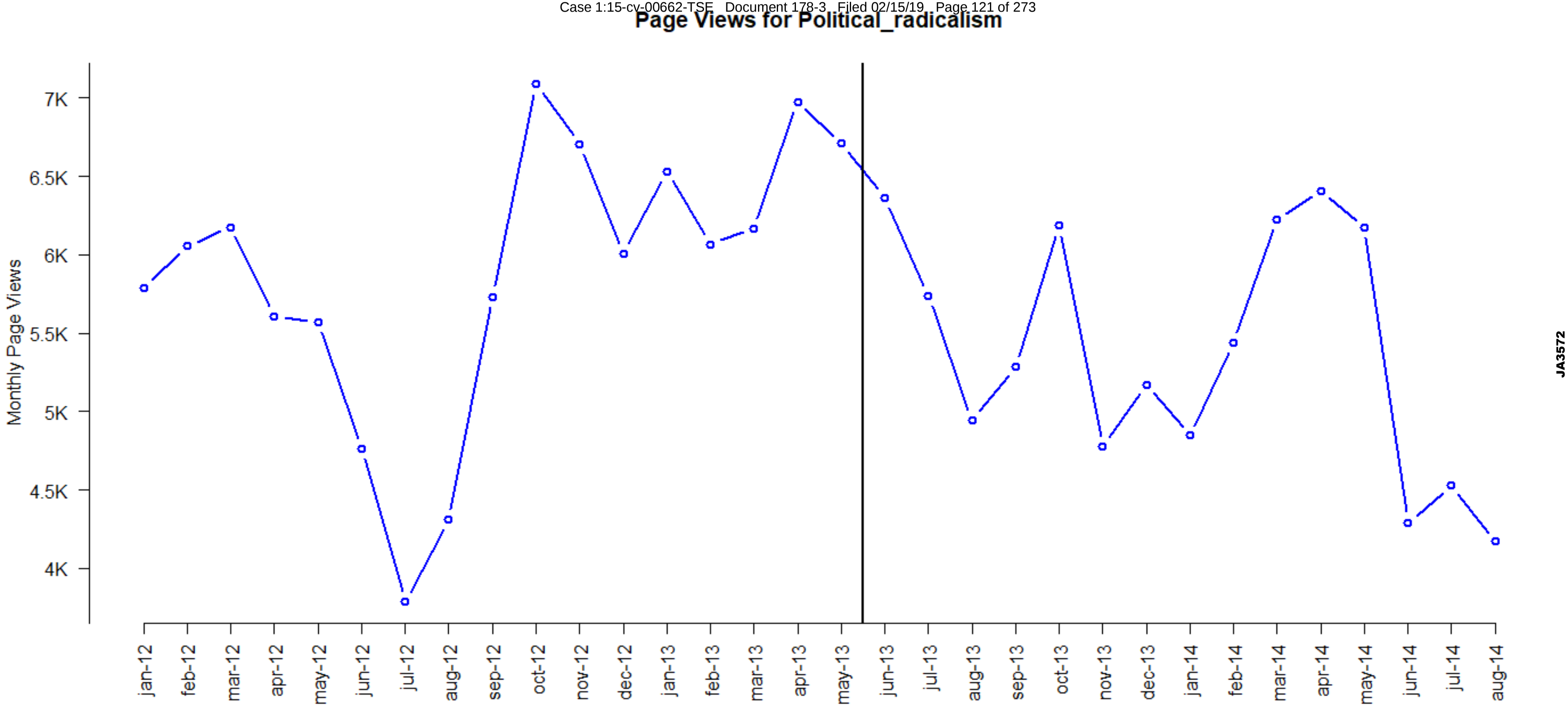

Page Views for Nuclear_Enrichment
Monthly Page Views
0.3K
0.25K
0.2K
0.15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

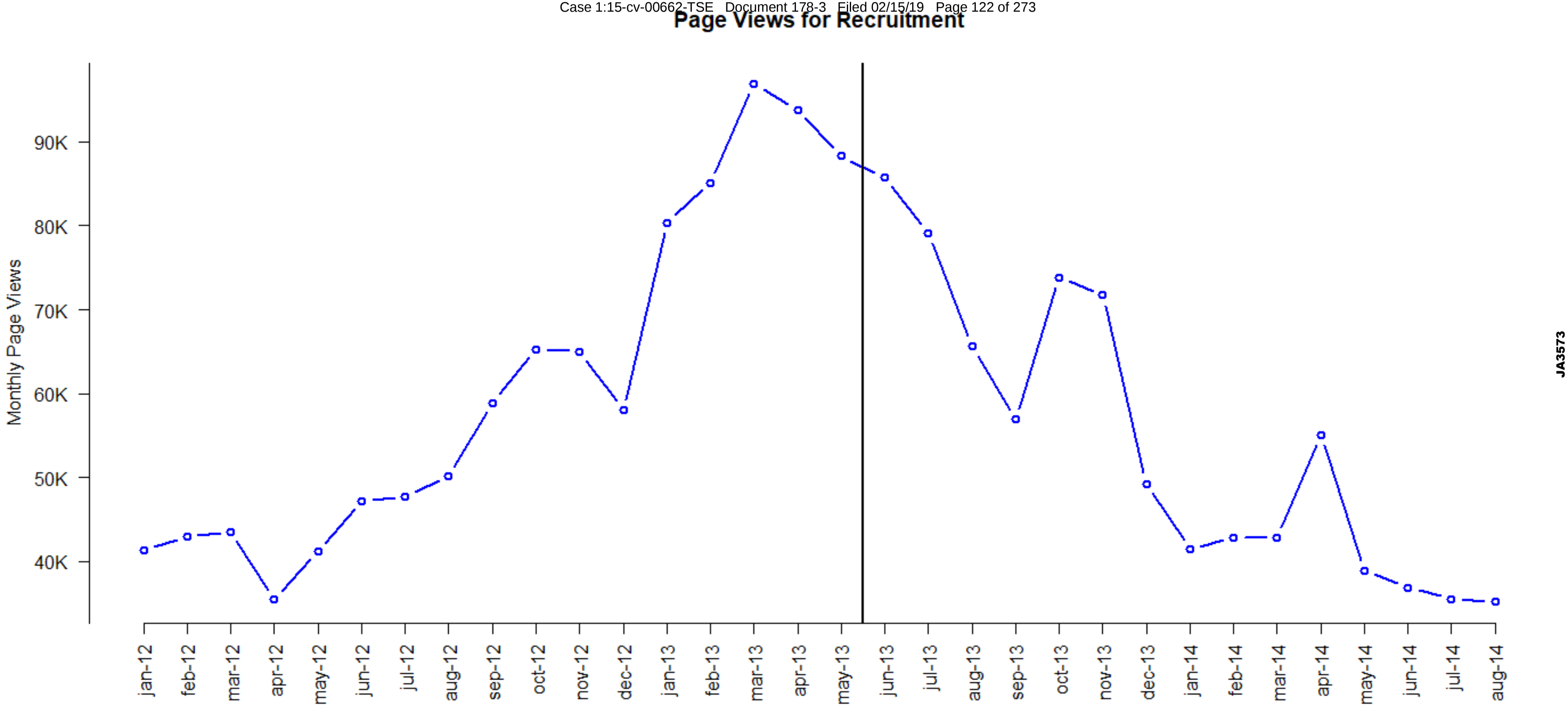

Page Views for Pakistan
Monthly Page Views
500K
450K
400K
350K
300K
250K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

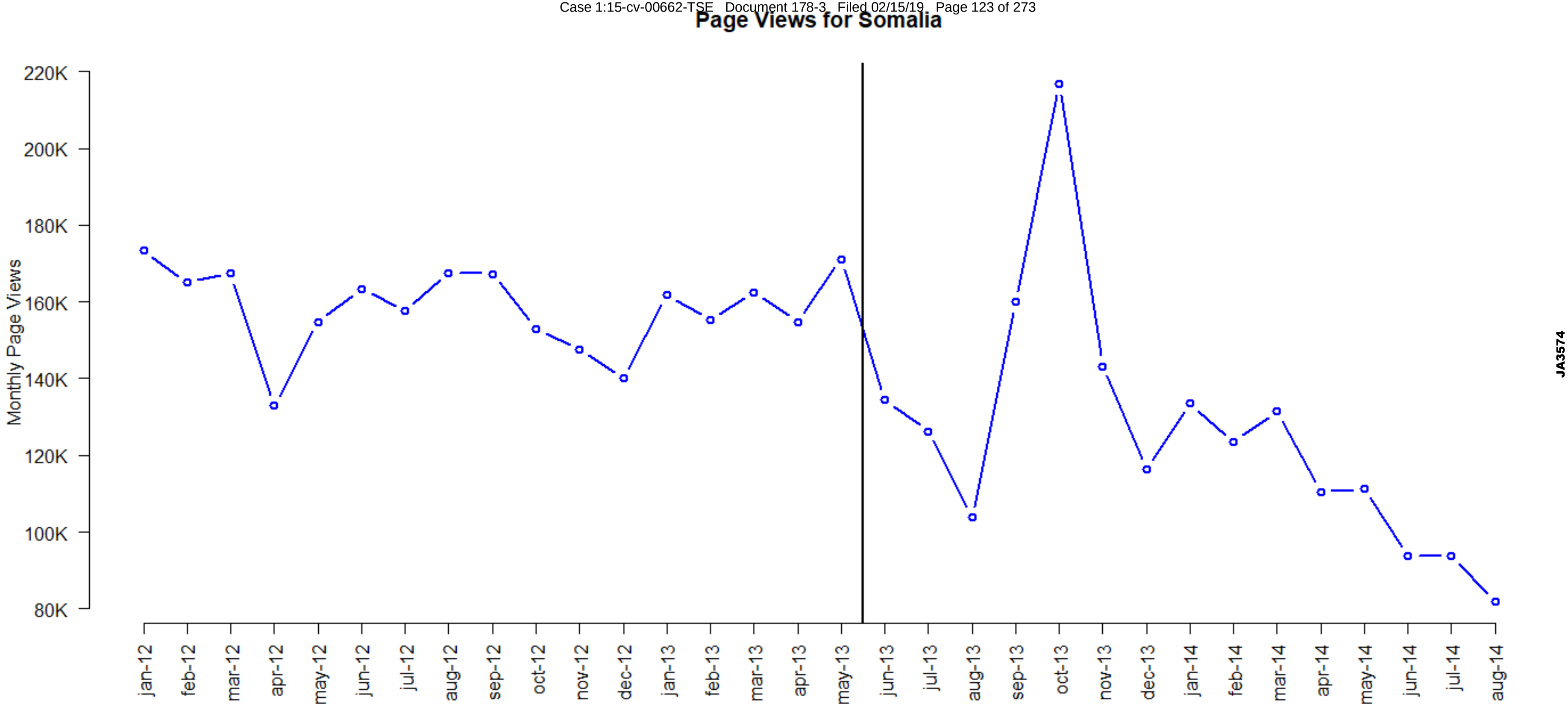

Page Views for Palestine_Liberation_Fron
Monthly Page Views
3K
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

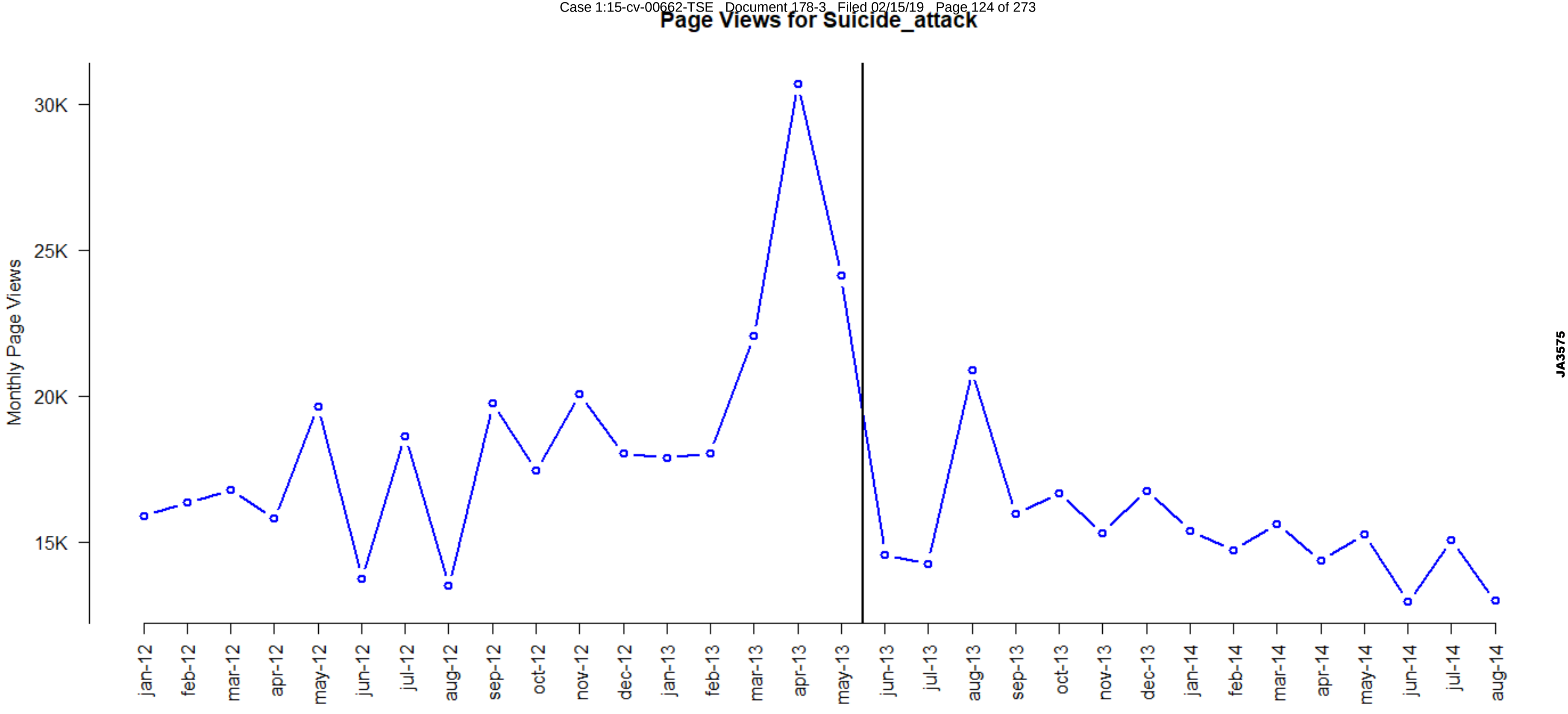

Page Views for Pirates
Monthly Page Views
14K
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

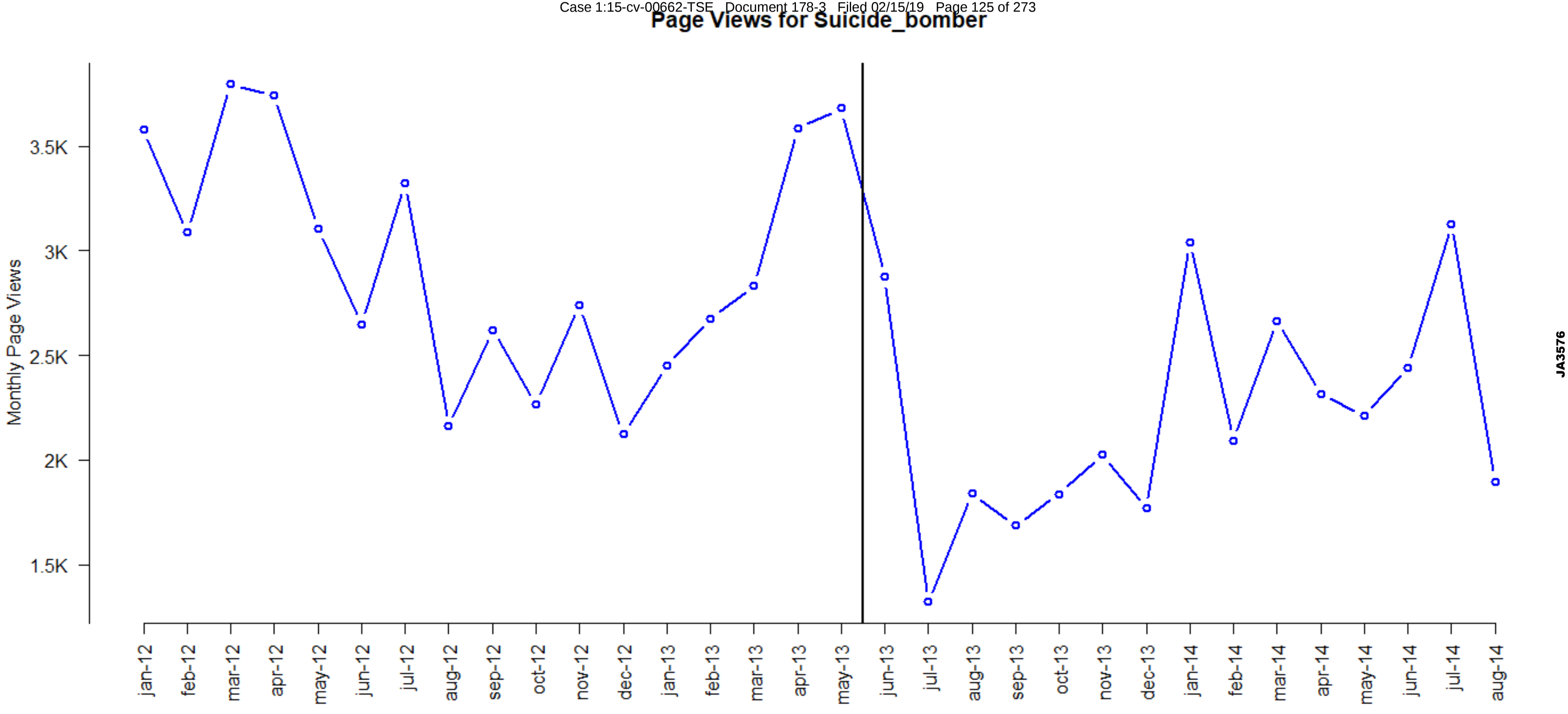

Page Views for PLO
Monthly Page Views
50K
40K
30K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

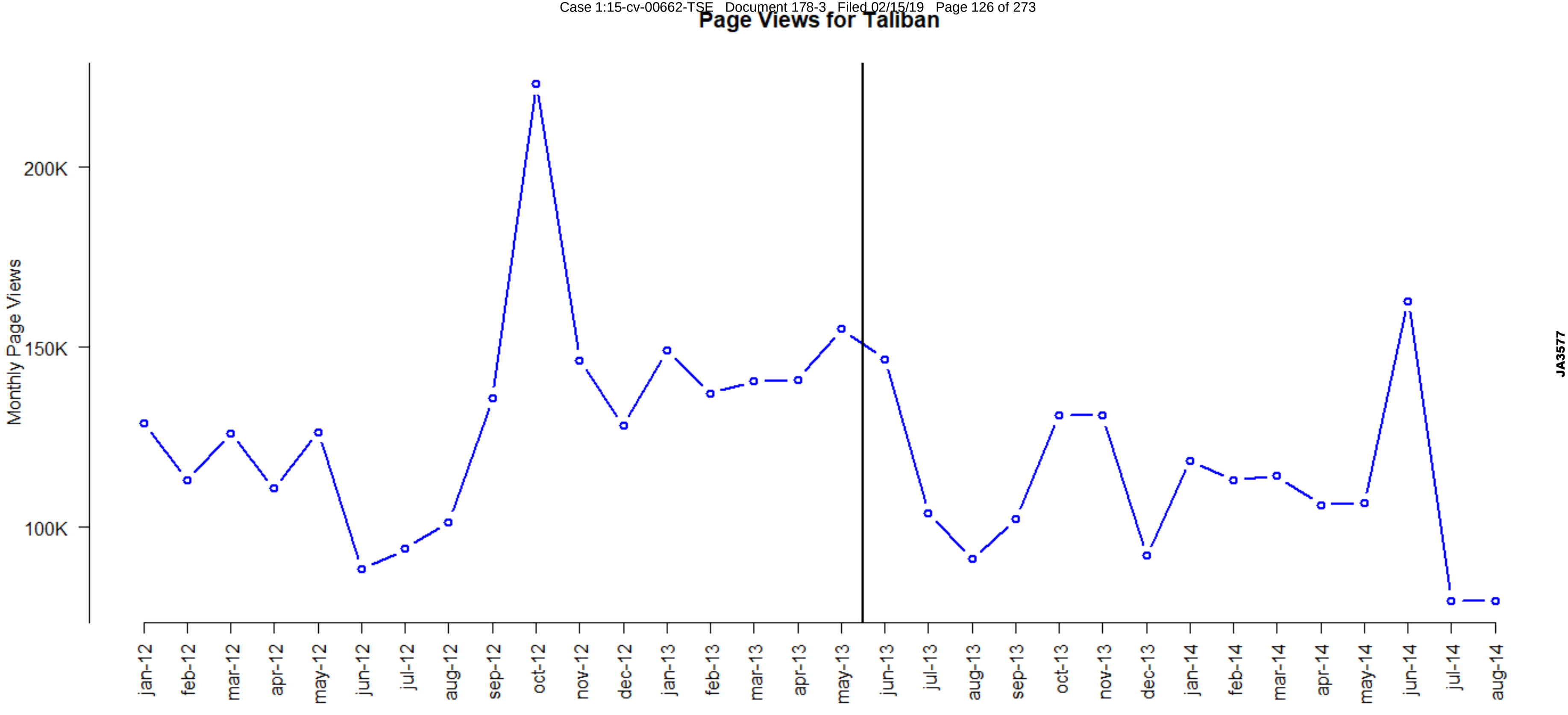

Page Views for Political_radicalism
Monthly Page Views
7K
6.5K
6K
5.5K
5K
4.5K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

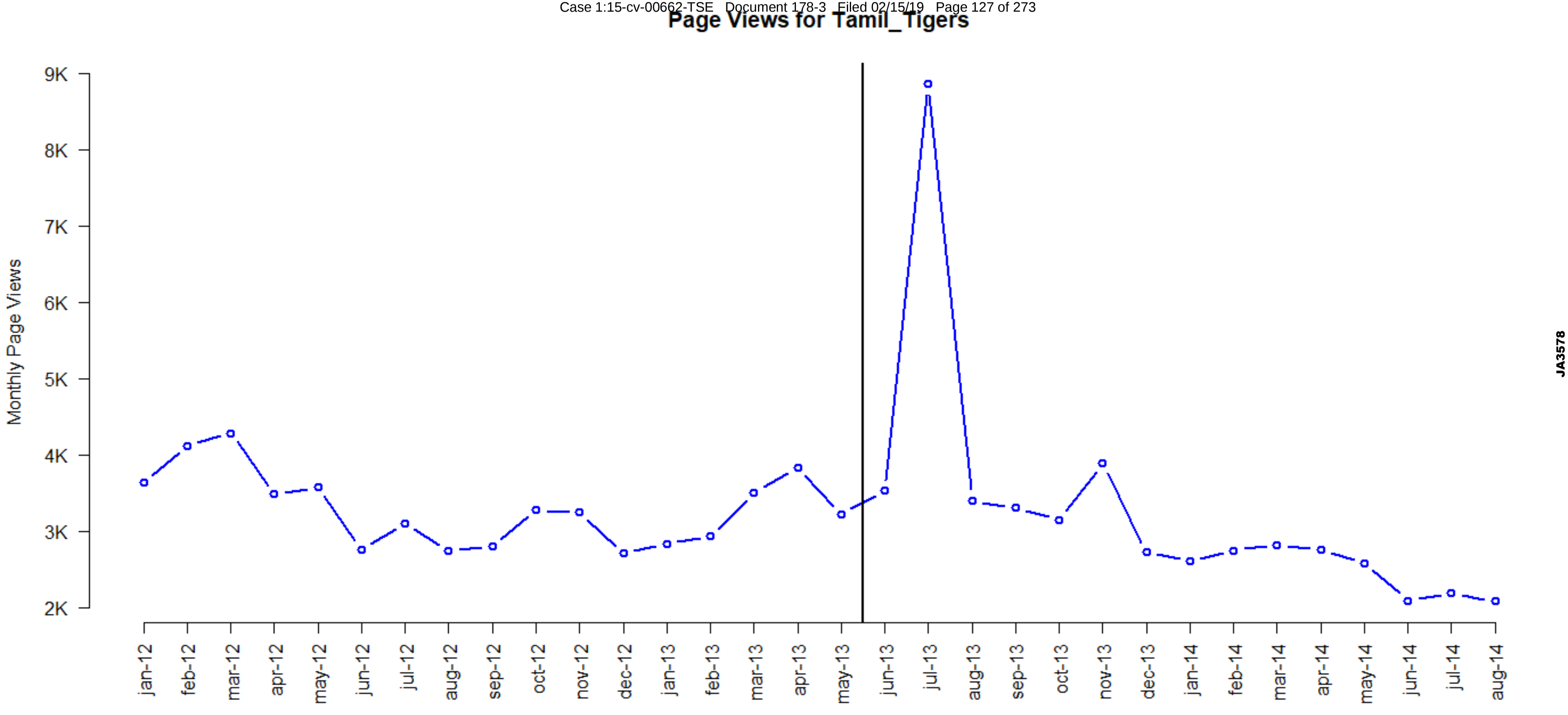

Page Views for Recruitment
Monthly Page Views
90K
80K
70K
60K
50K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

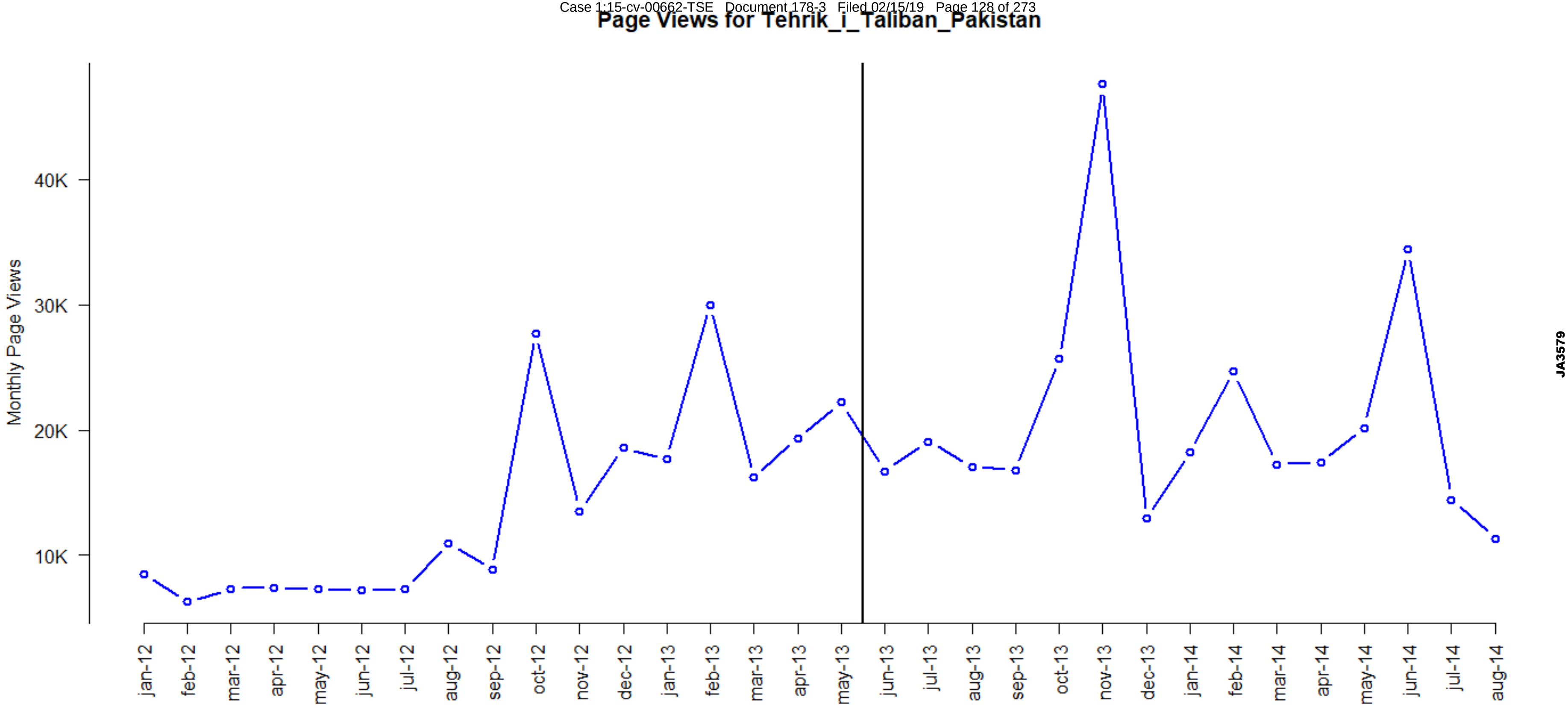

Page Views for Somalia
Monthly Page Views
220K
200K
180K
160K
140K
120K
100K
80K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

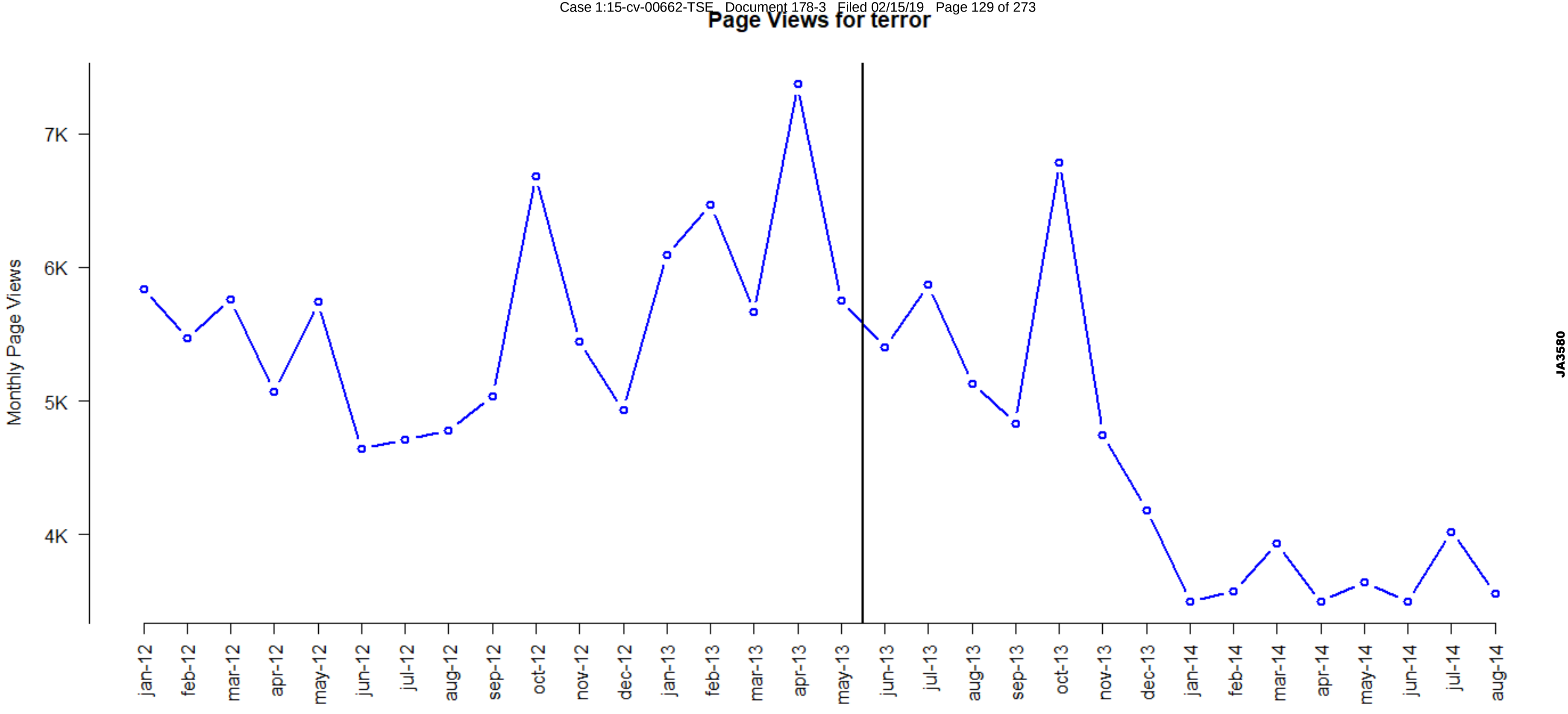

Page Views for Suicide_attack
Monthly Page Views
30K
25K
20K
15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

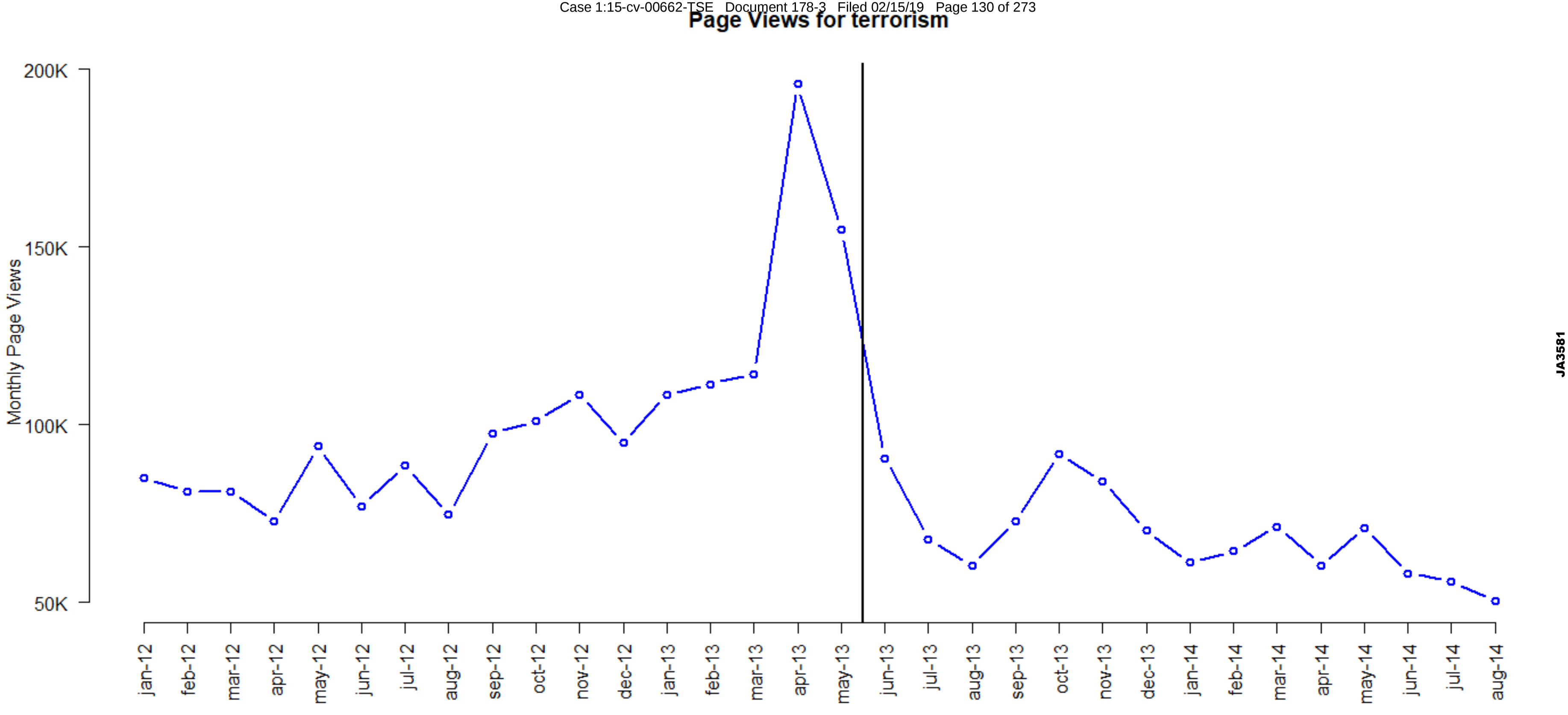

Page Views for Suicide_bomber
Monthly Page Views
3.5K
3K
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

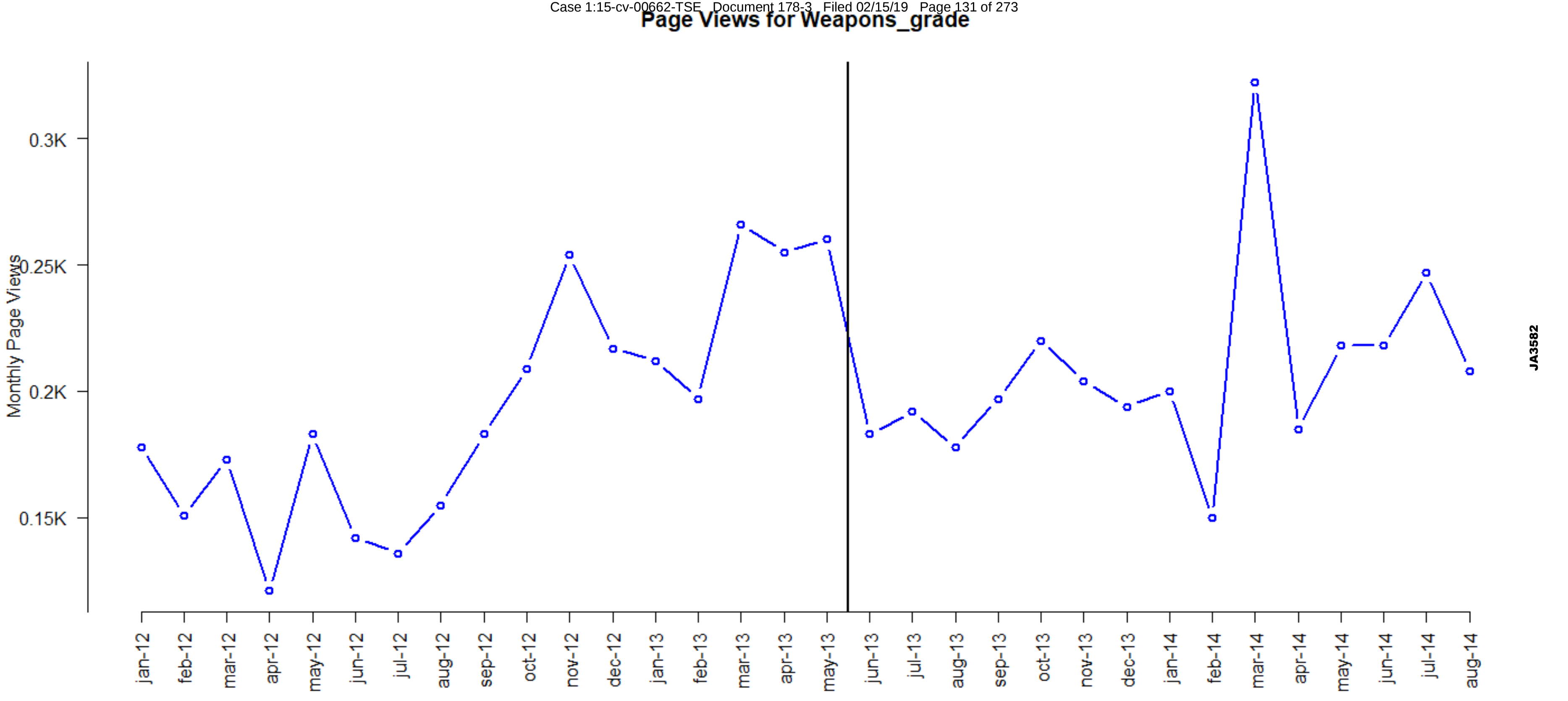

Page Views for Taliban
Monthly Page Views
200K
150K
100K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

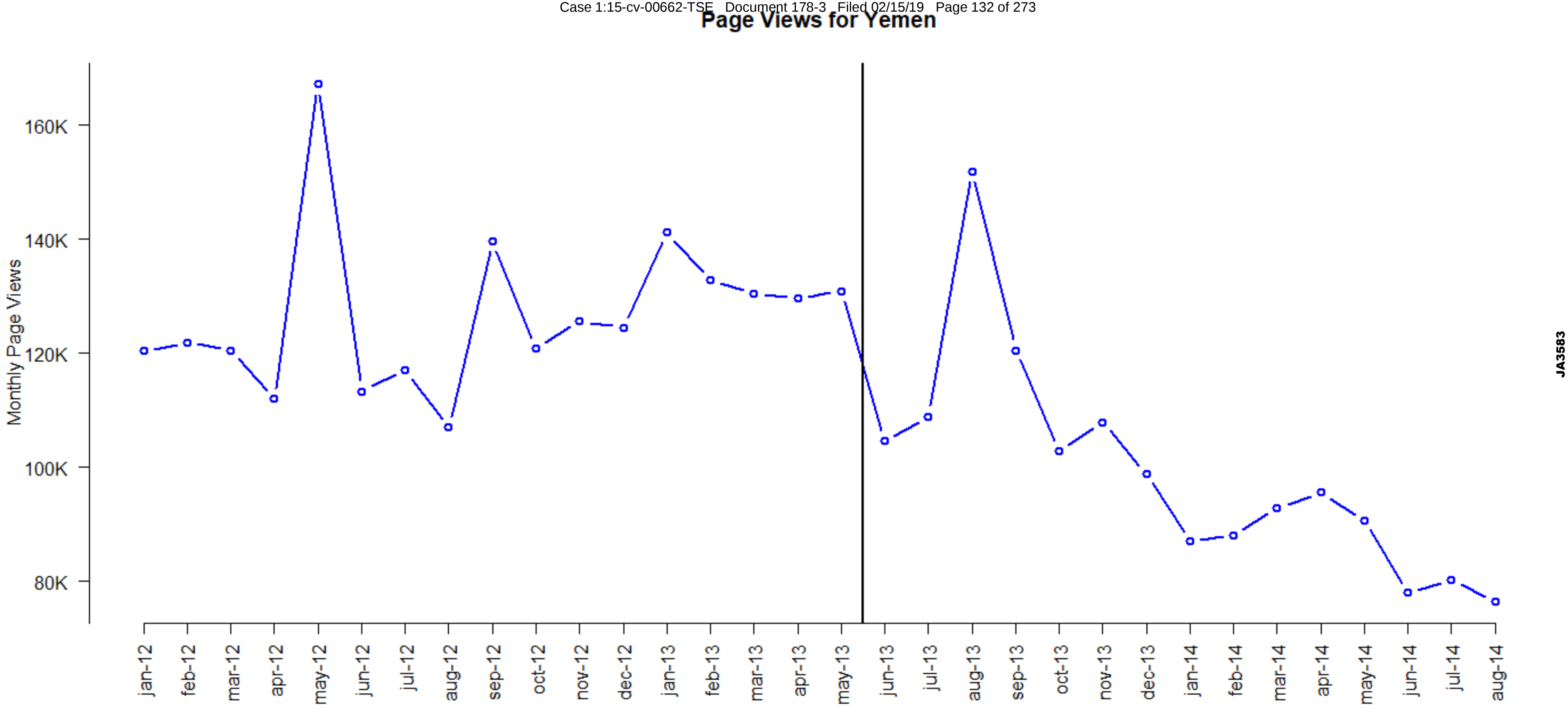

Page Views for Tamil_Tigers
Monthly Page Views
9K
8K
7K
6K
5K
4K
3K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Page Views for Tehrik_i_Taliban_Pakistan
Monthly Page Views
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

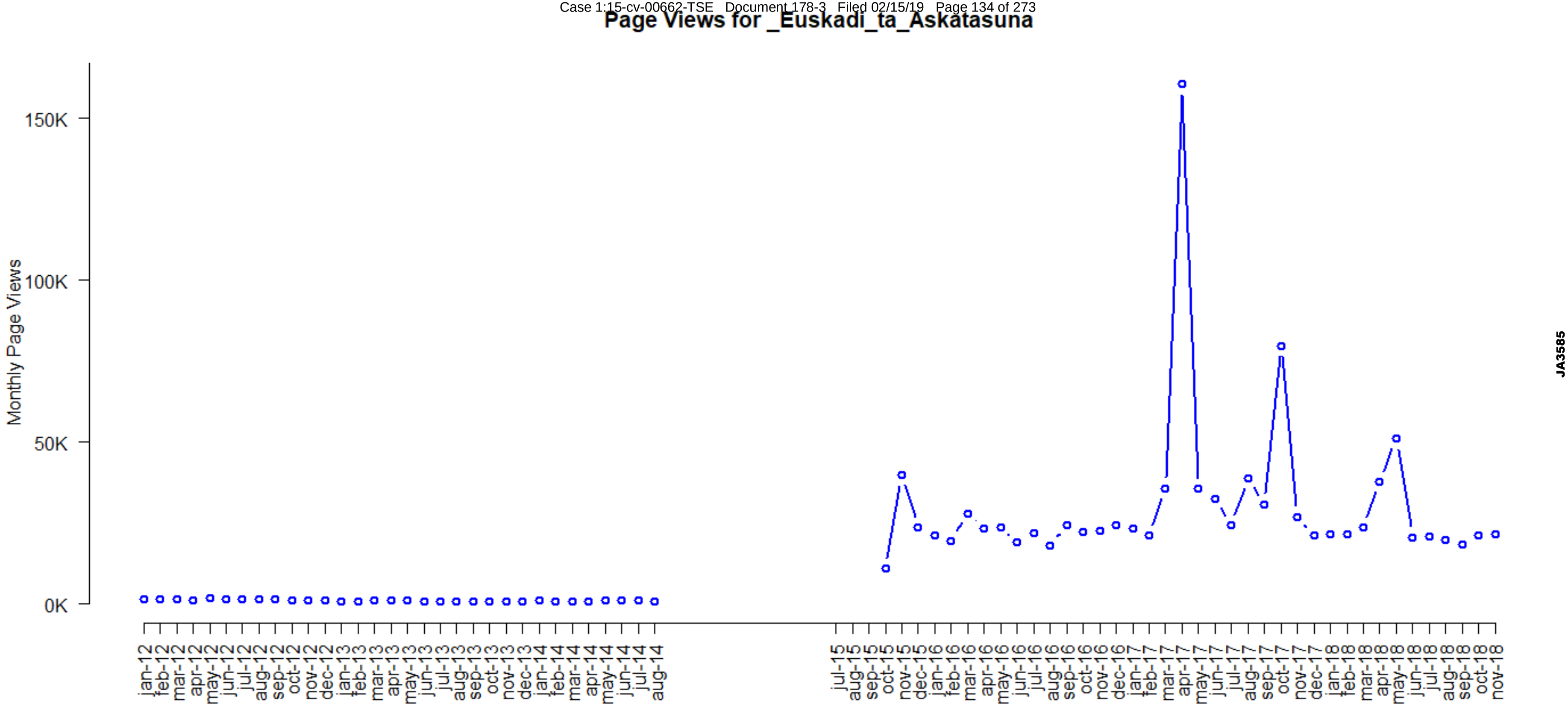

Page Views for terror
Monthly Page Views
7K
6K
5K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

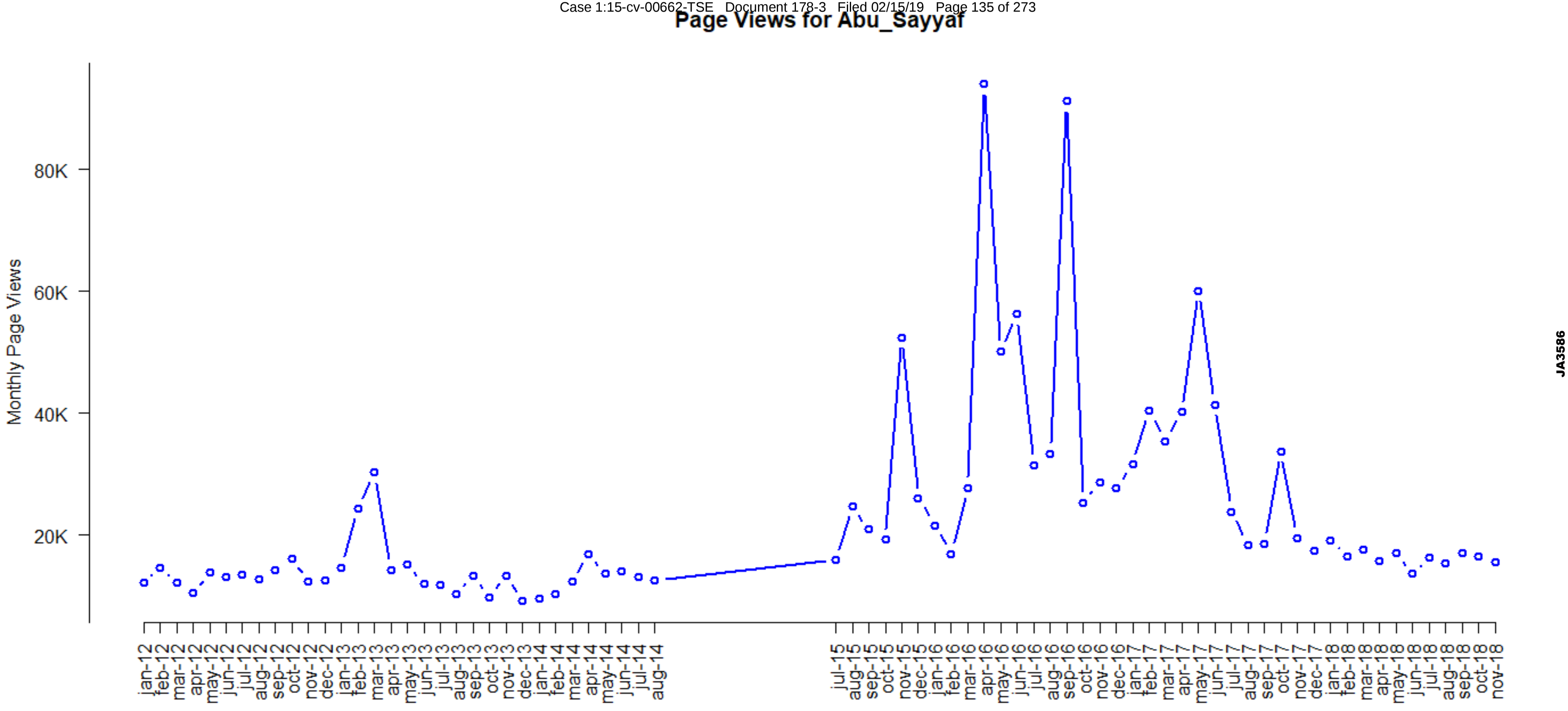

Page Views for terrorism
Monthly Page Views
200K
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Page Views for Weapons_grade
Monthly Page Views
0.3K
0.25K
0.2K
0.15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

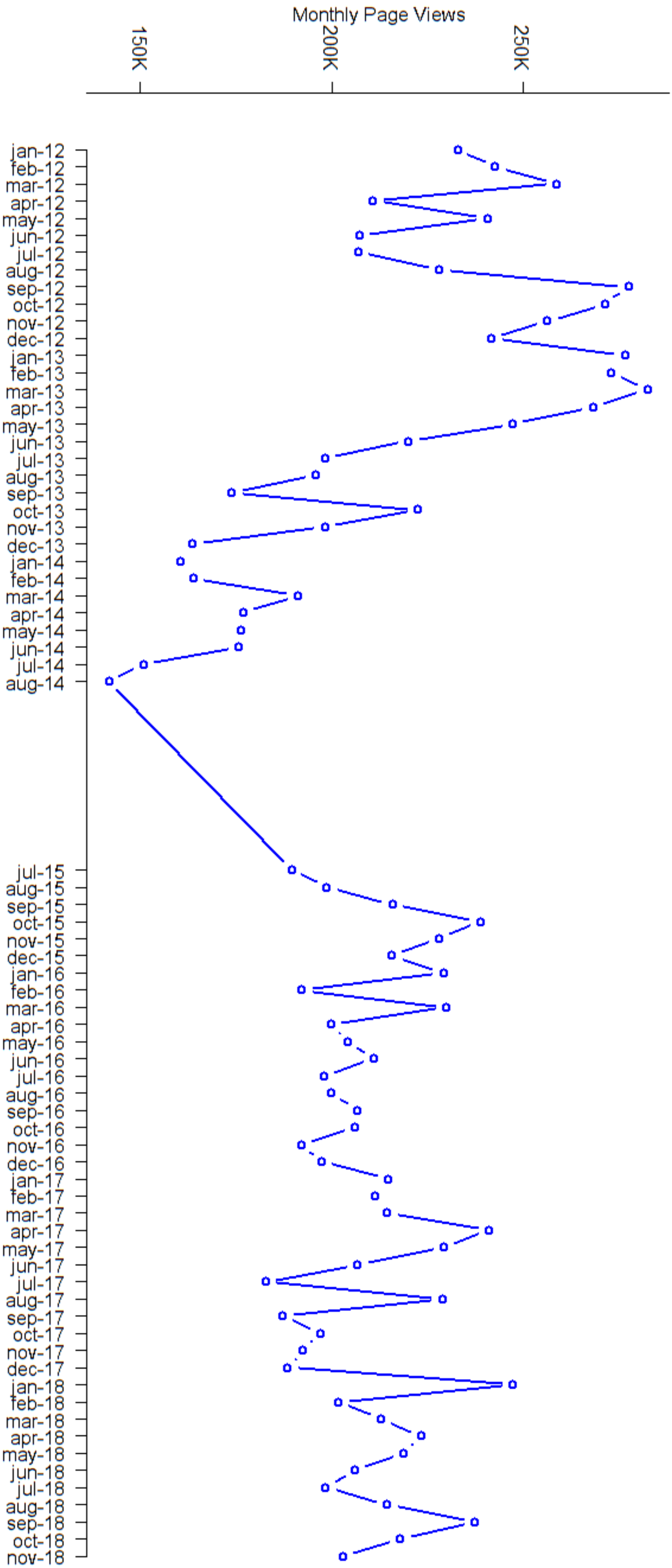

Page Views for Yemen
Monthly Page Views
160K
140K
120K
100K
80K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

APPENDIX V: Page Views for 48 Terror Articles, Extended Time Period

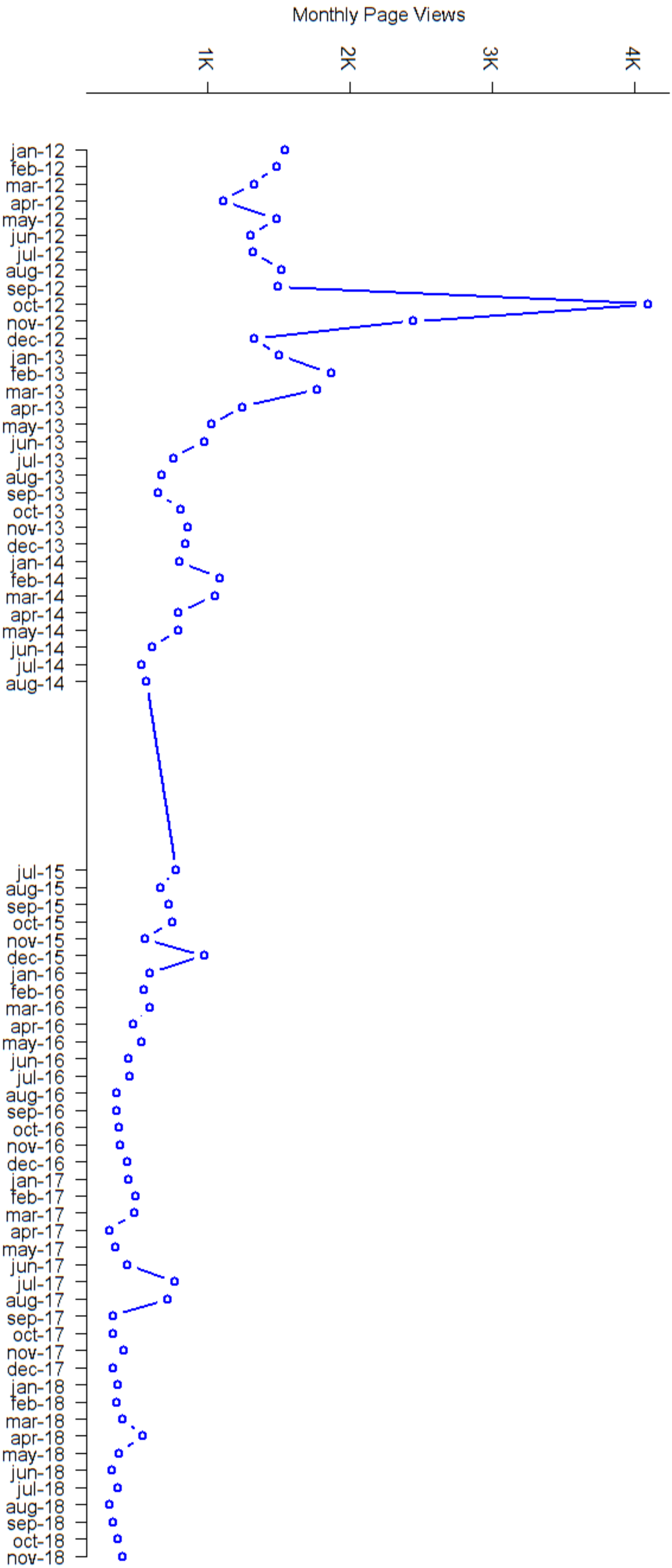

Page Views for _Euskadi_ta_Askatasuna
Monthly Page Views
150K
100K
50K
0K
jan-12 feb-12 mar-12 apr-12 may-12 jun-12 jul-12 aug-12 sep-12 oct-12 nov-12 dec-12 jan-13 feb-13 mar-13 apr-13 may-13 jun-13 jul-13 aug-13 sep-13 oct-13 nov-13 dec-13 jan-14 feb-14 mar-14 apr-14 may-14 jun-14 jul-14 aug-14
jul-15 aug-15 sep-15 oct-15 nov-15 dec-15 jan-16 feb-16 mar-16 apr-16 may-16 jun-16 jul-16 aug-16 sep-16 oct-16 nov-16 dec-16 jan-17 feb-17 mar-17 apr-17 may-17 jun-17 jul-17 aug-17 sep-17 oct-17 nov-17 dec-17 jan-18 feb-18 mar-18 apr-18 may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18

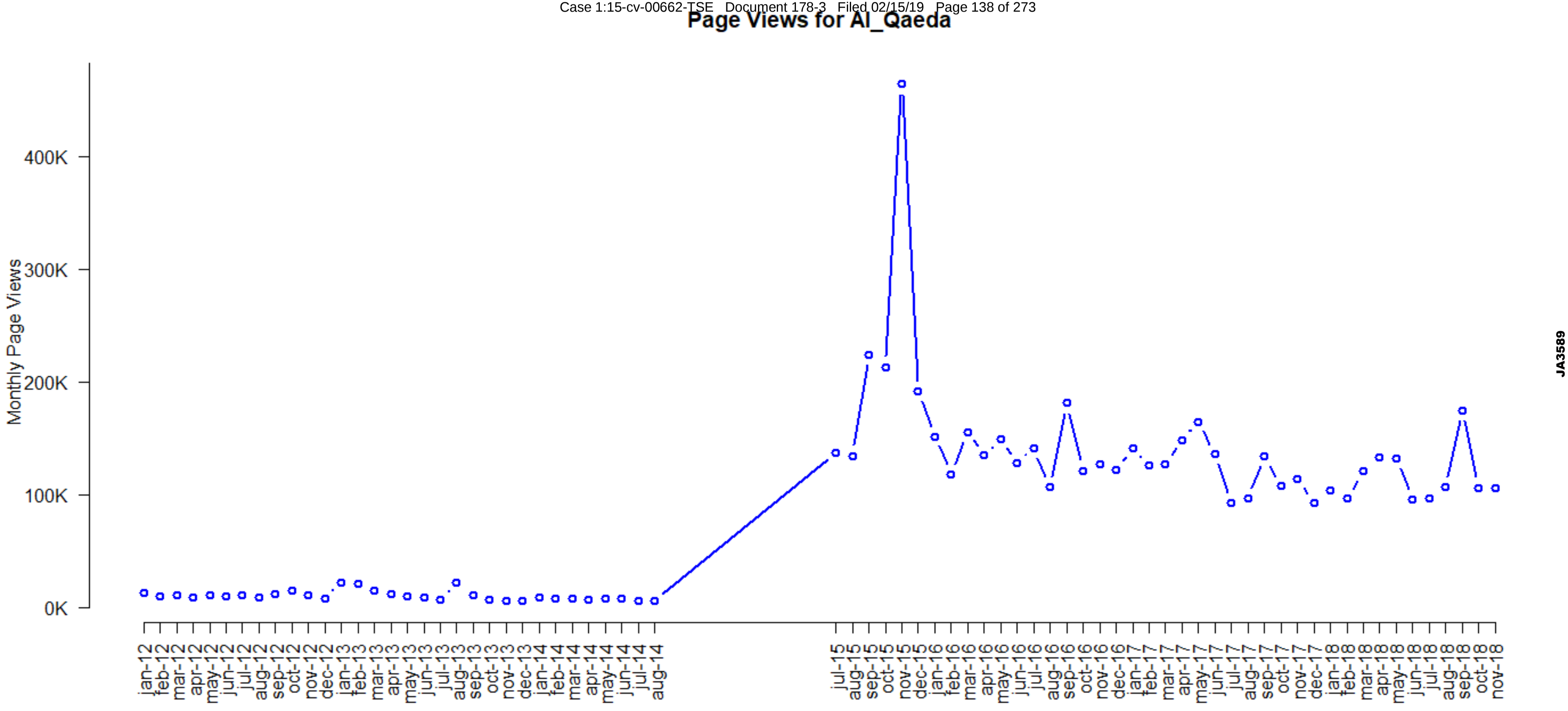

Page Views for Abu_Sayyaf
Monthly Page Views
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

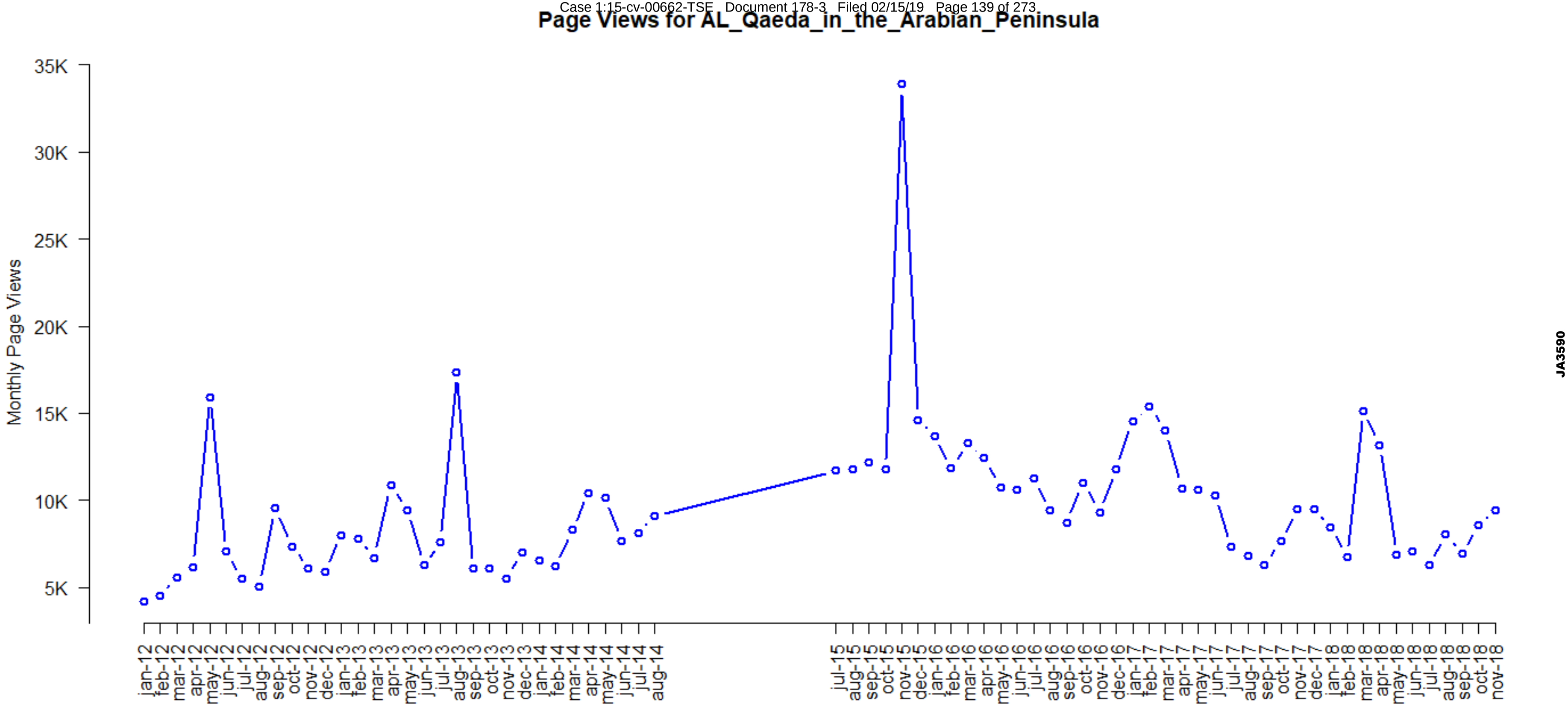
Page Views for Afghanistan

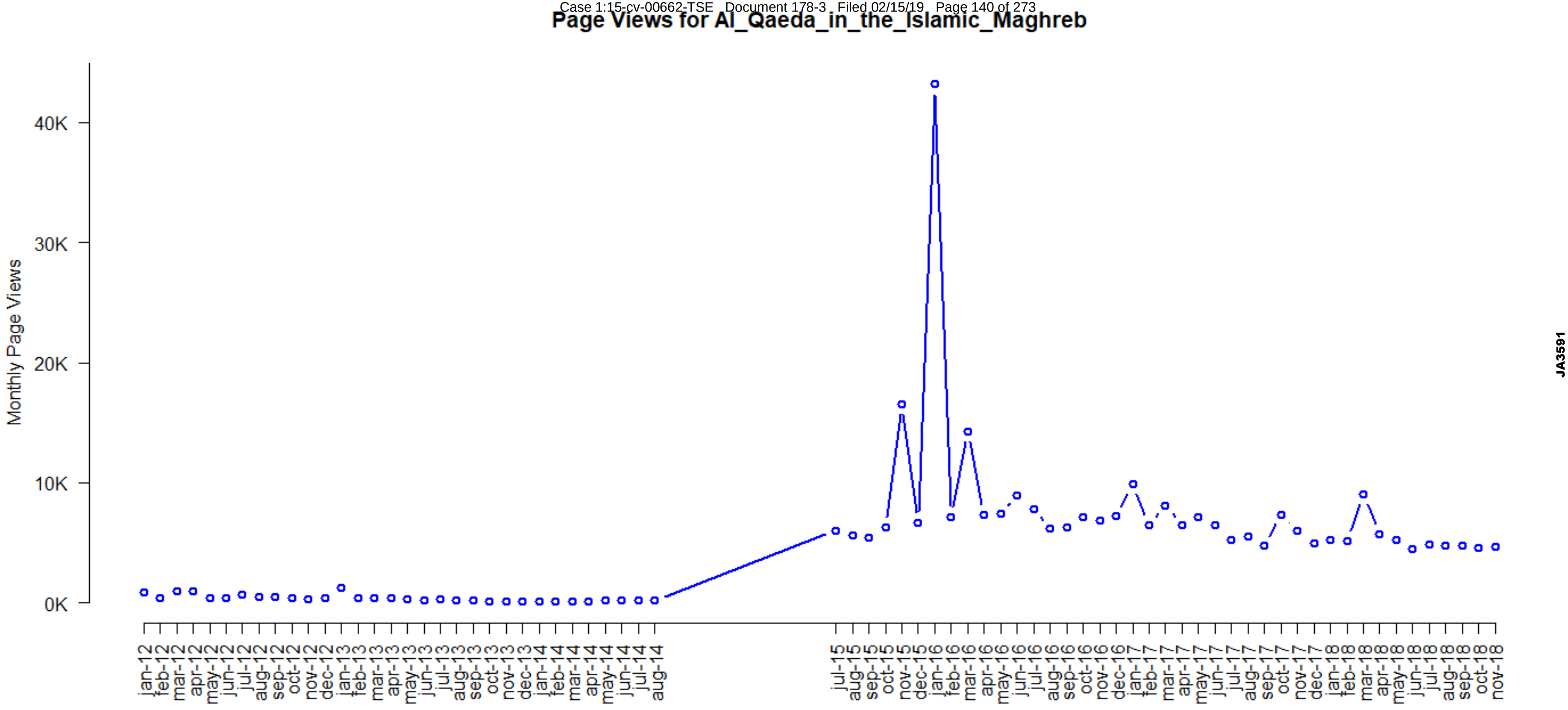


Page Views for agro

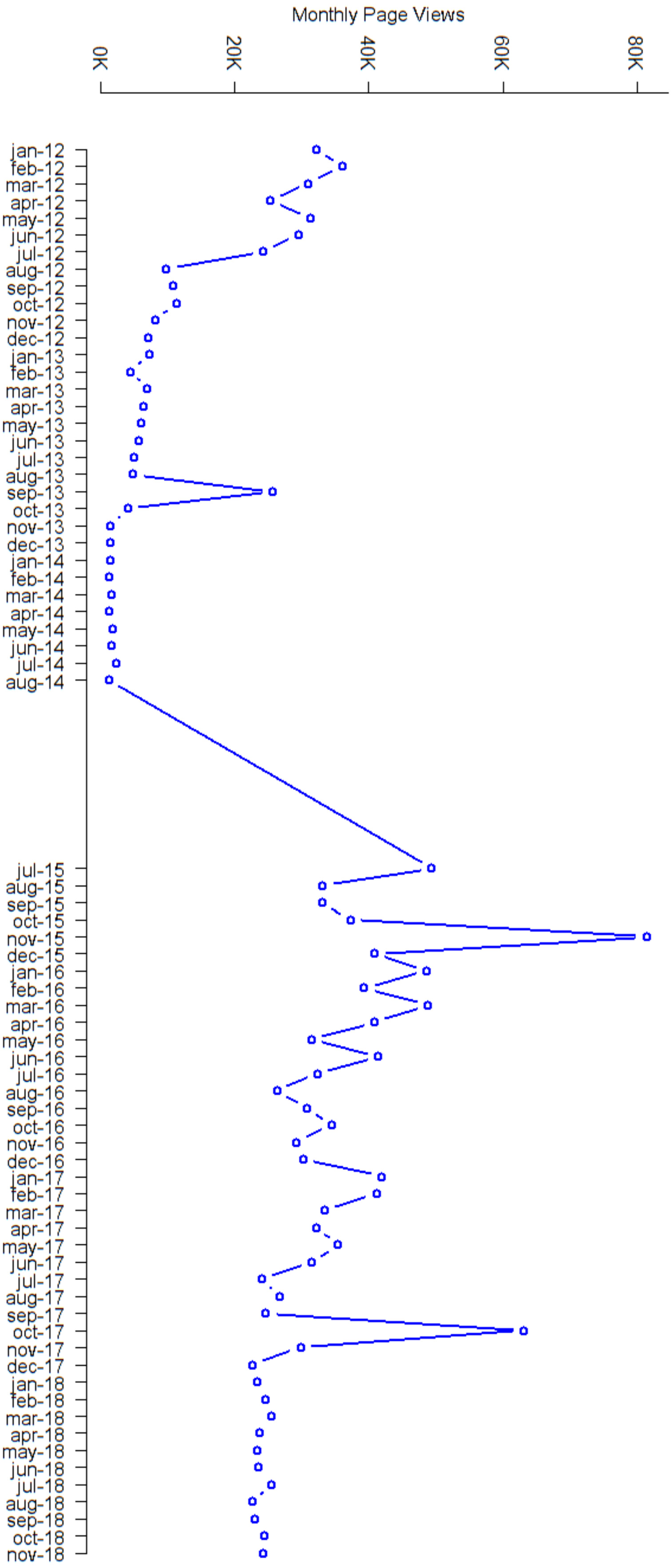



Page Views for Al_Qaeda
Monthly Page Views
400K
300K
200K
100K
0K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

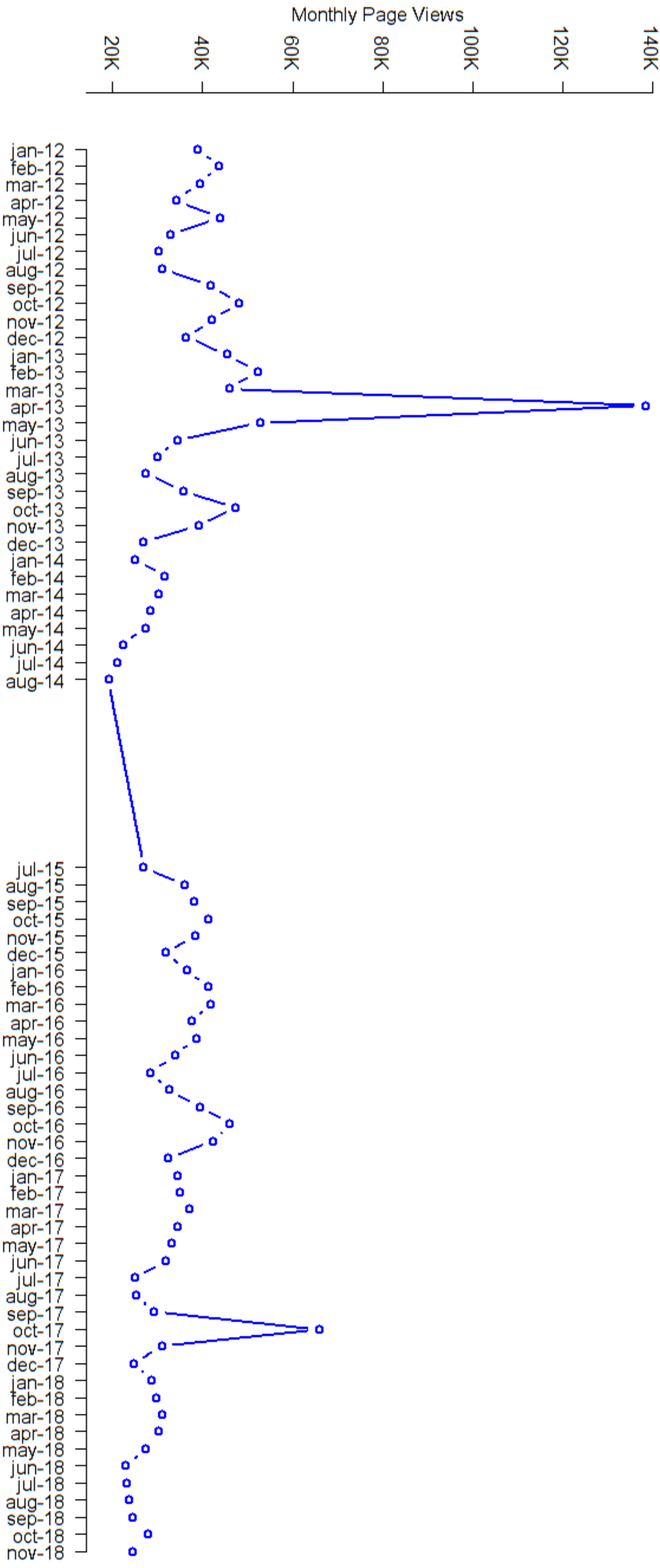

Page Views for AL_Qaeda_in_the_Arabian_Peninsula
Monthly Page Views
35K
30K
25K
20K
15K
10K
5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18


Page Views for Al_Qaeda_in_the_Islamic_Maghreb
Monthly Page Views
0K
10K
20K
30K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

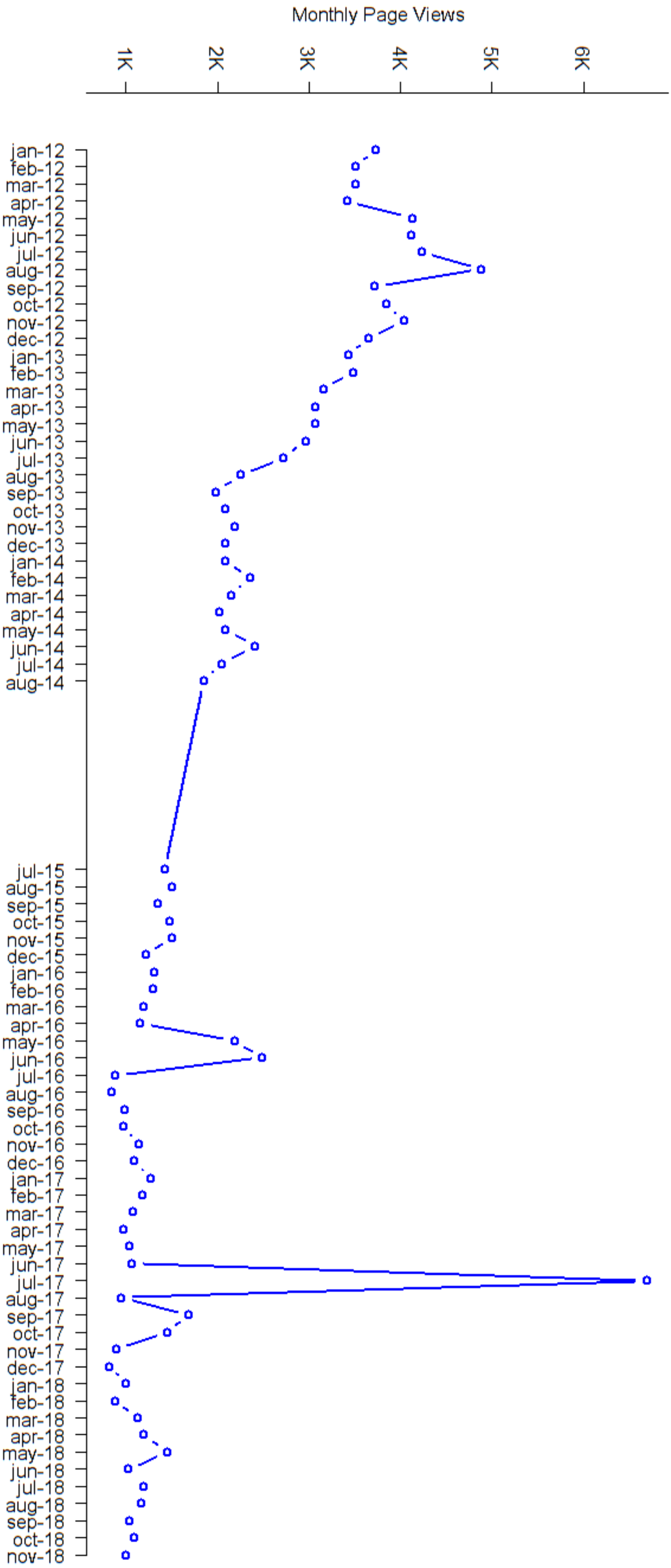
Page Views for Al_Shabaab



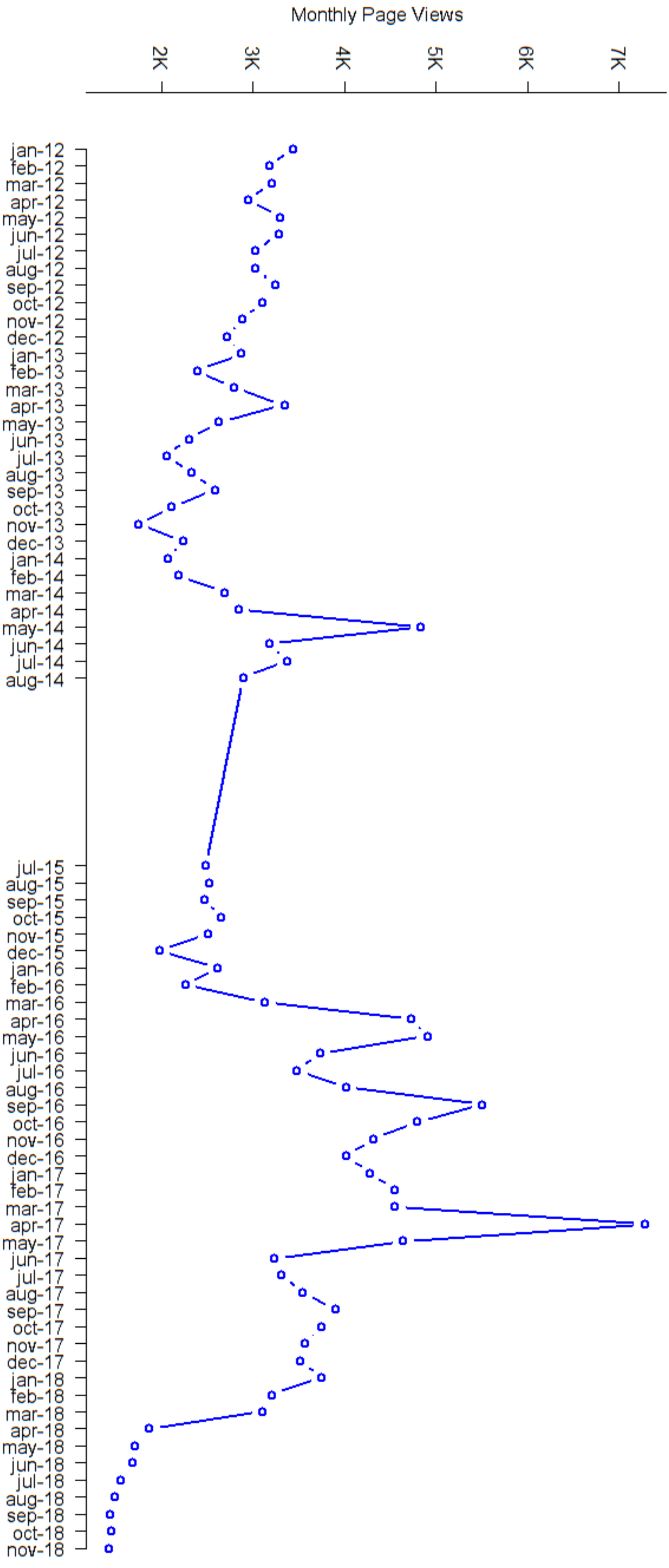
Page Views for Ammonium_nitrate



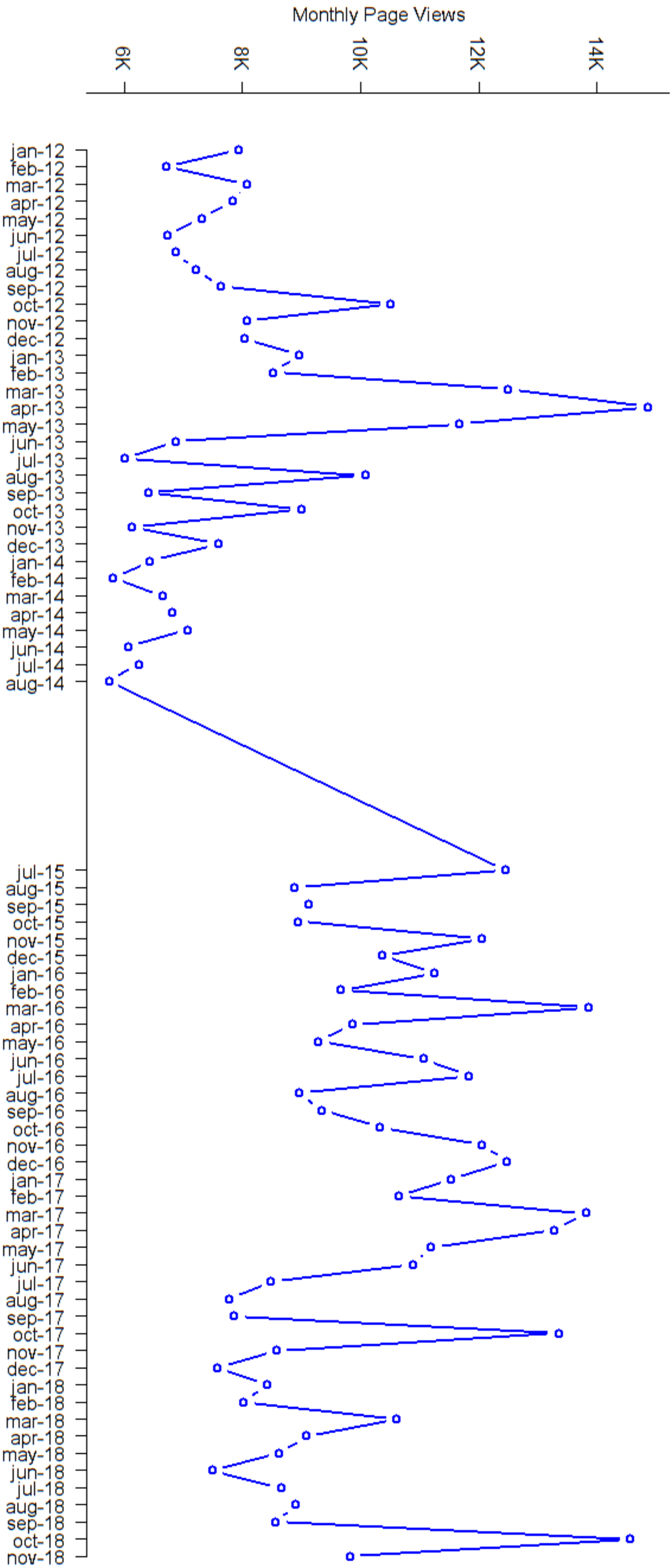
Page Views for attack



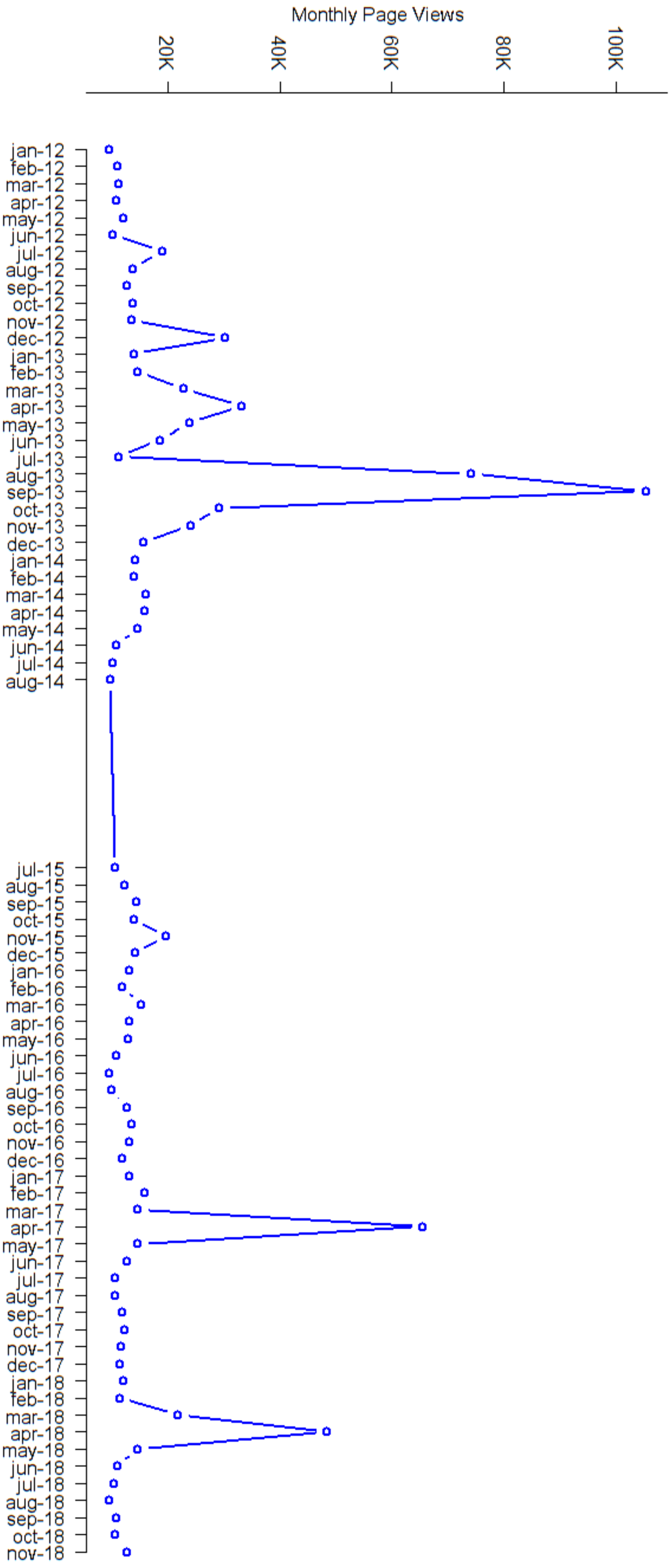
Page Views for Biological_weapon

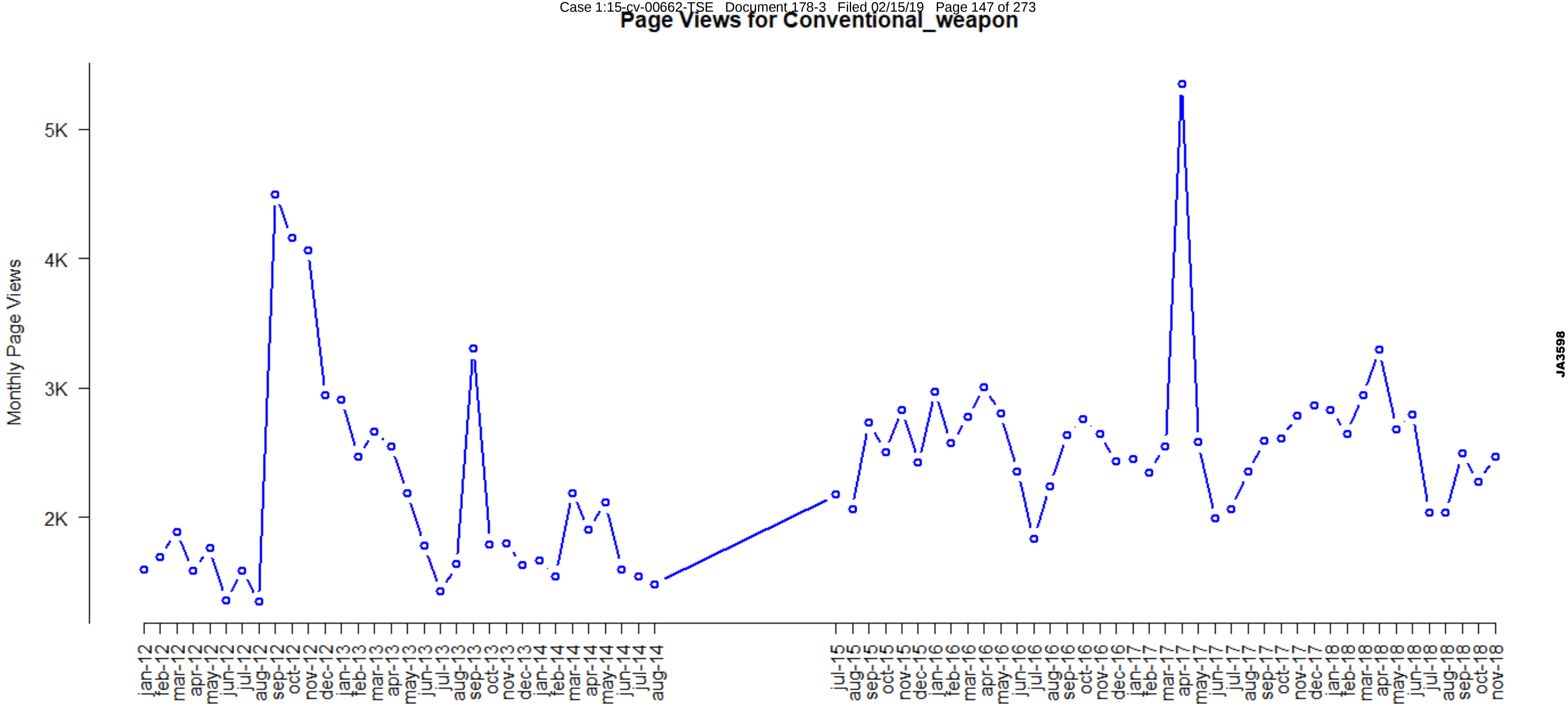


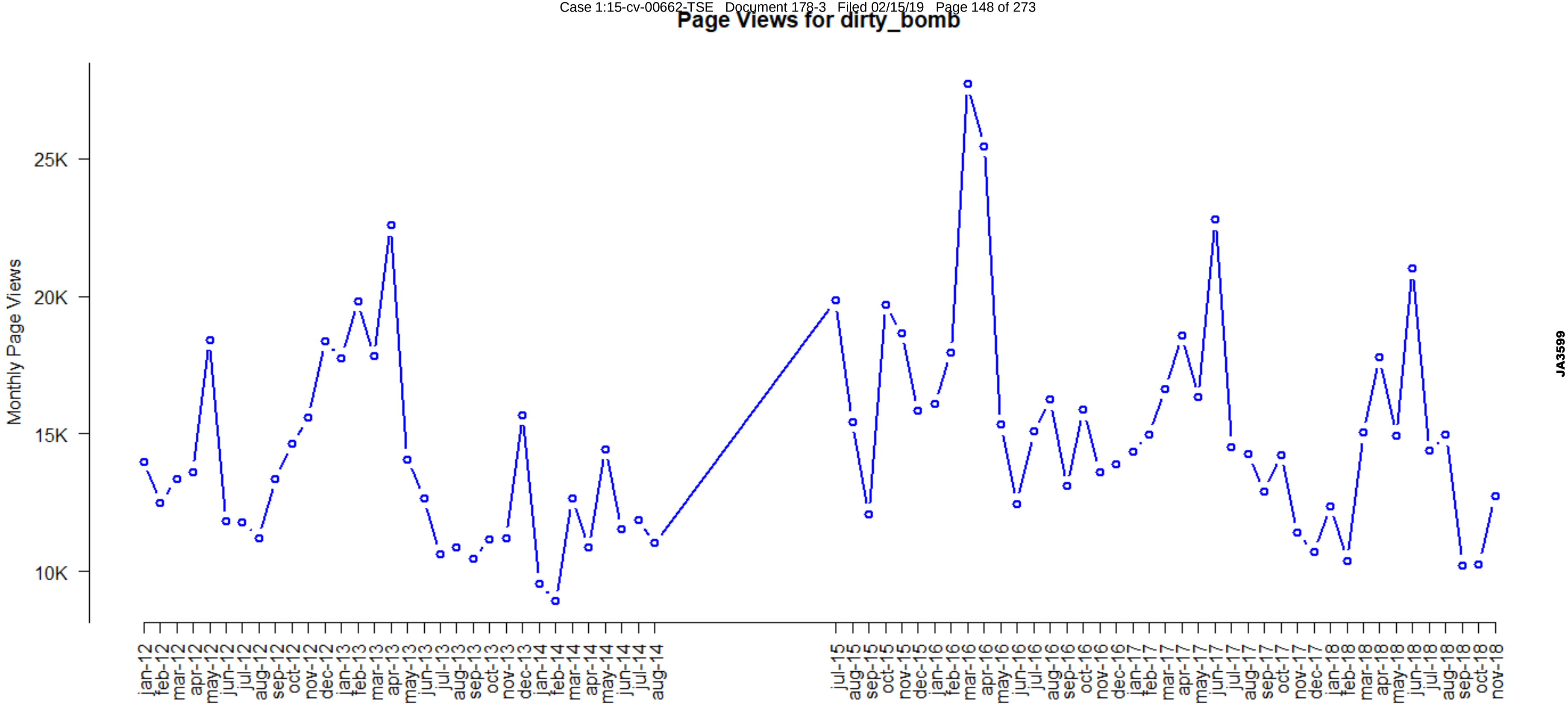
Page Views for Car_bomb

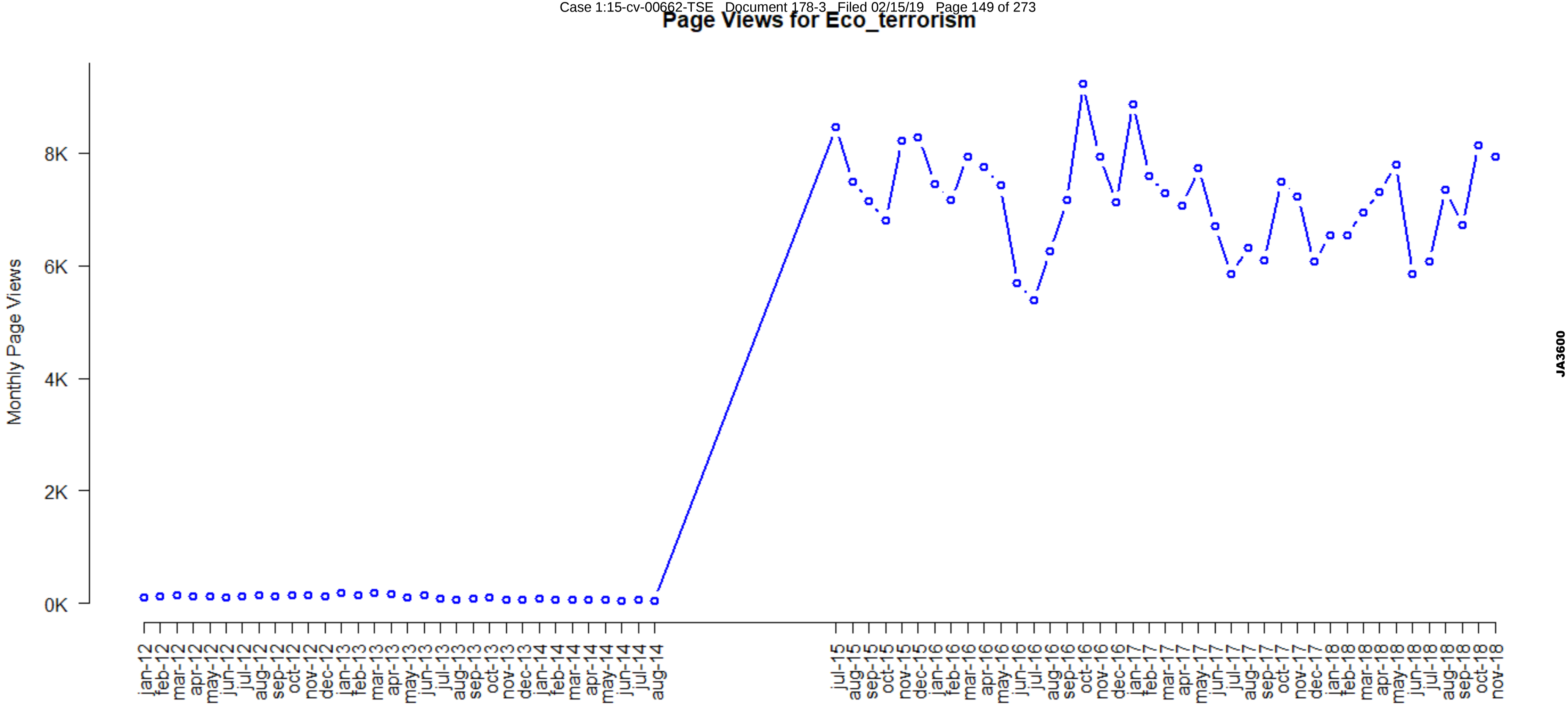


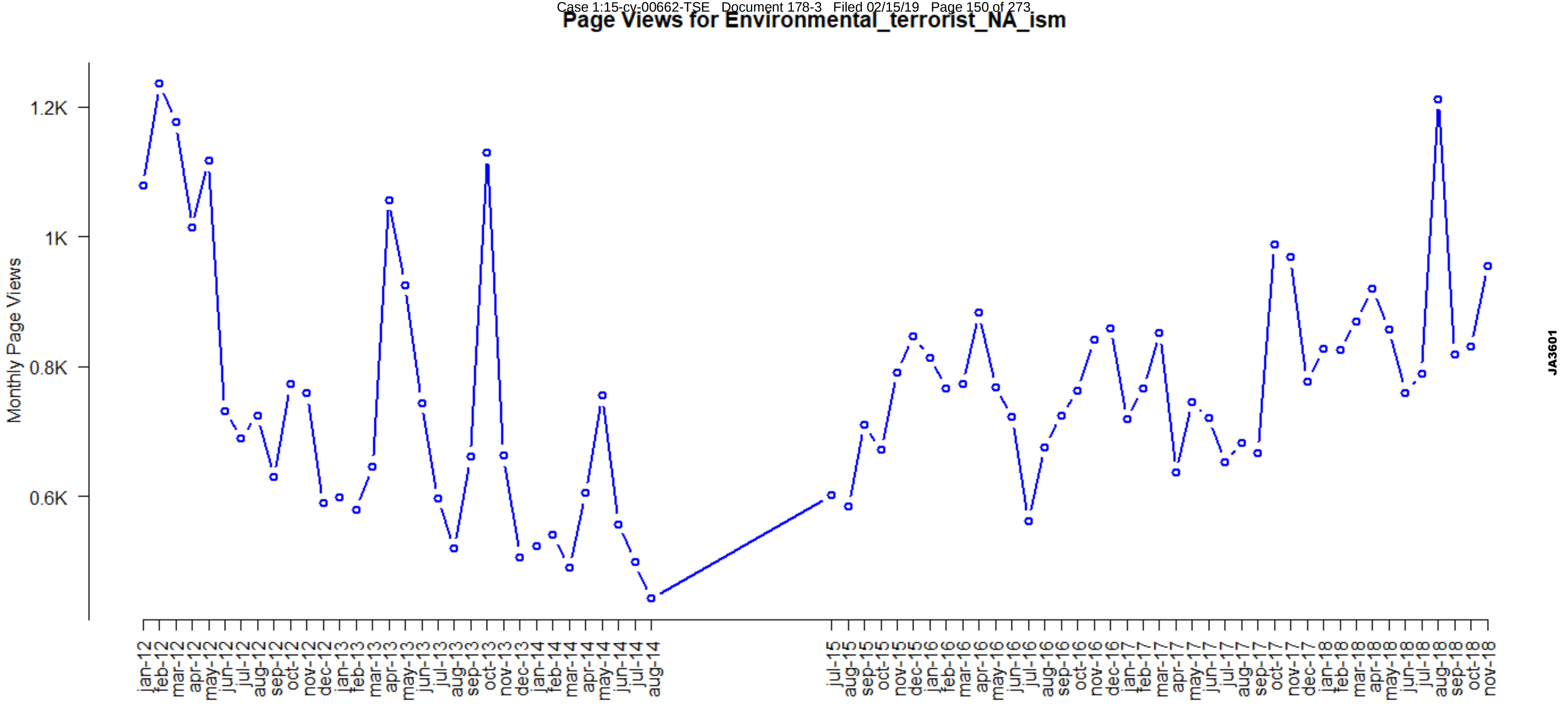
Page Views for Chemical_weapon

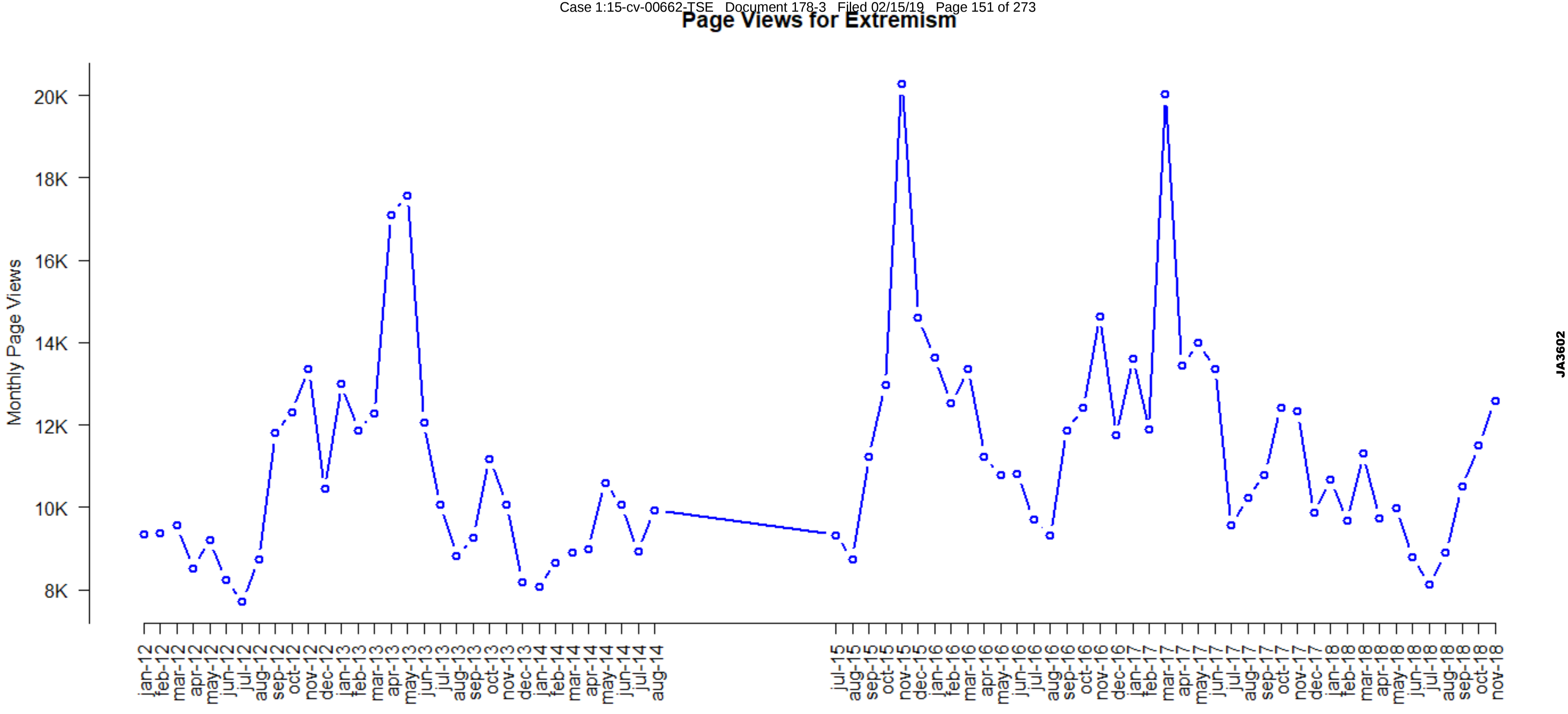


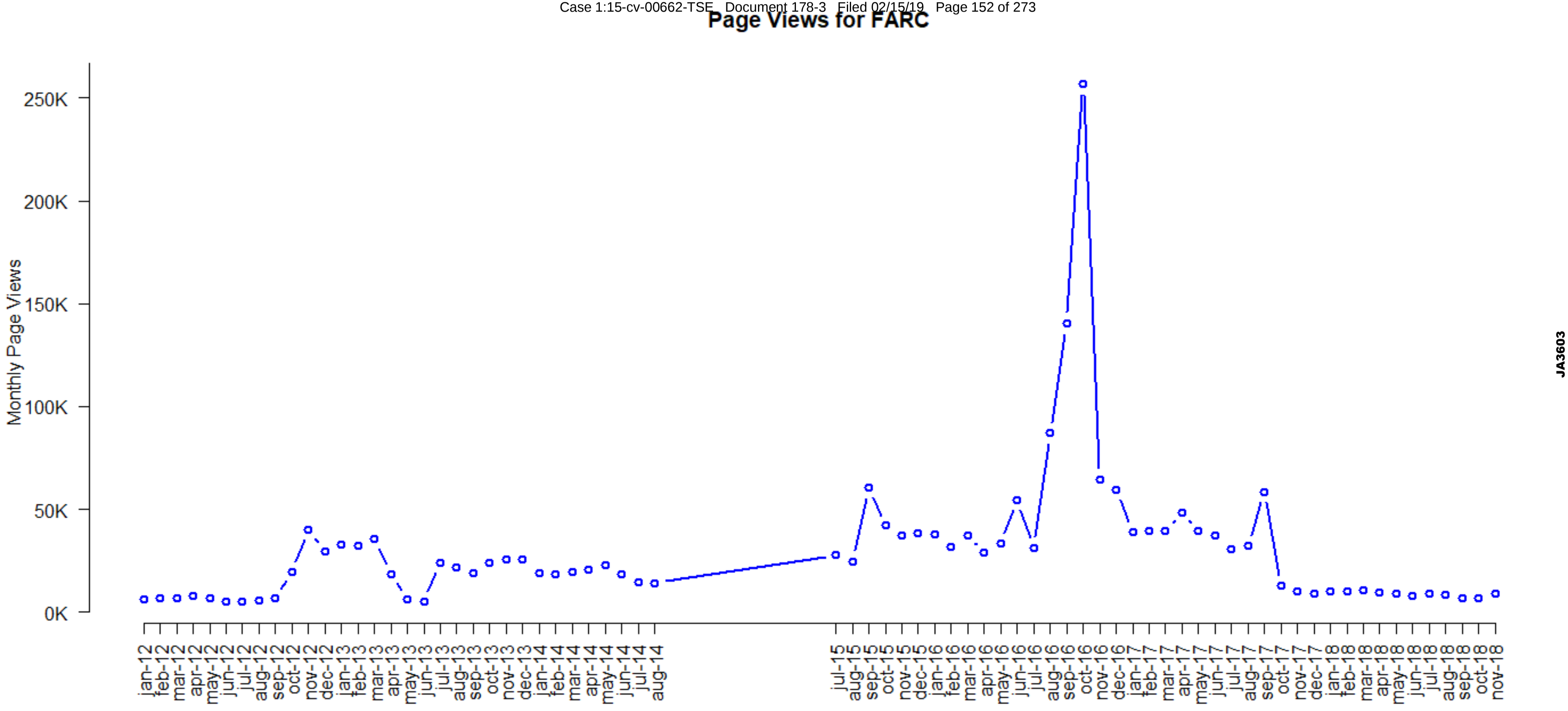

Page Views for Conventional_weapon
Monthly Page Views
5K
4K
3K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

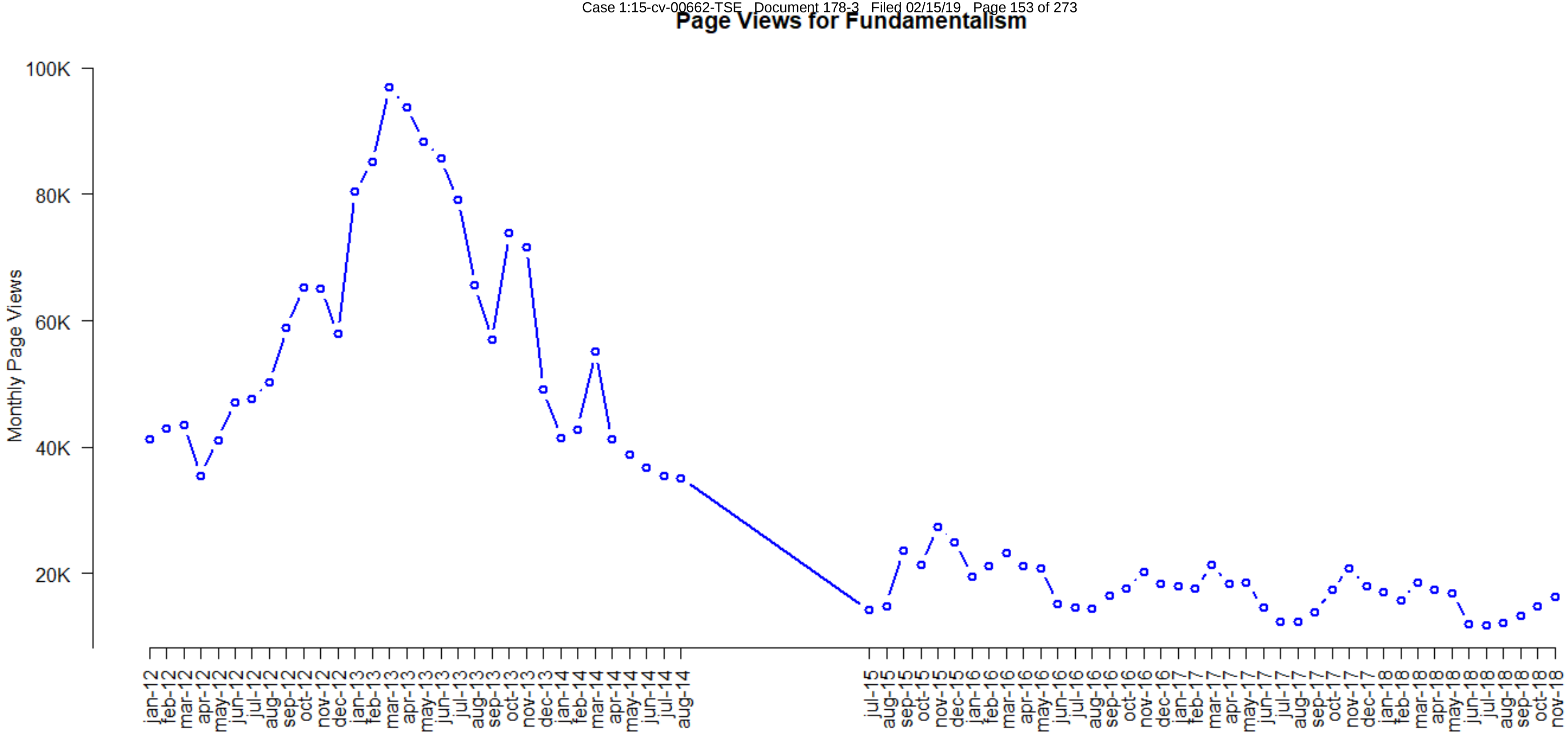

Page Views for dirty_bomb
Monthly Page Views
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18


Page Views for Eco_terrorism
Monthly Page Views
0K
2K
4K
6K
8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

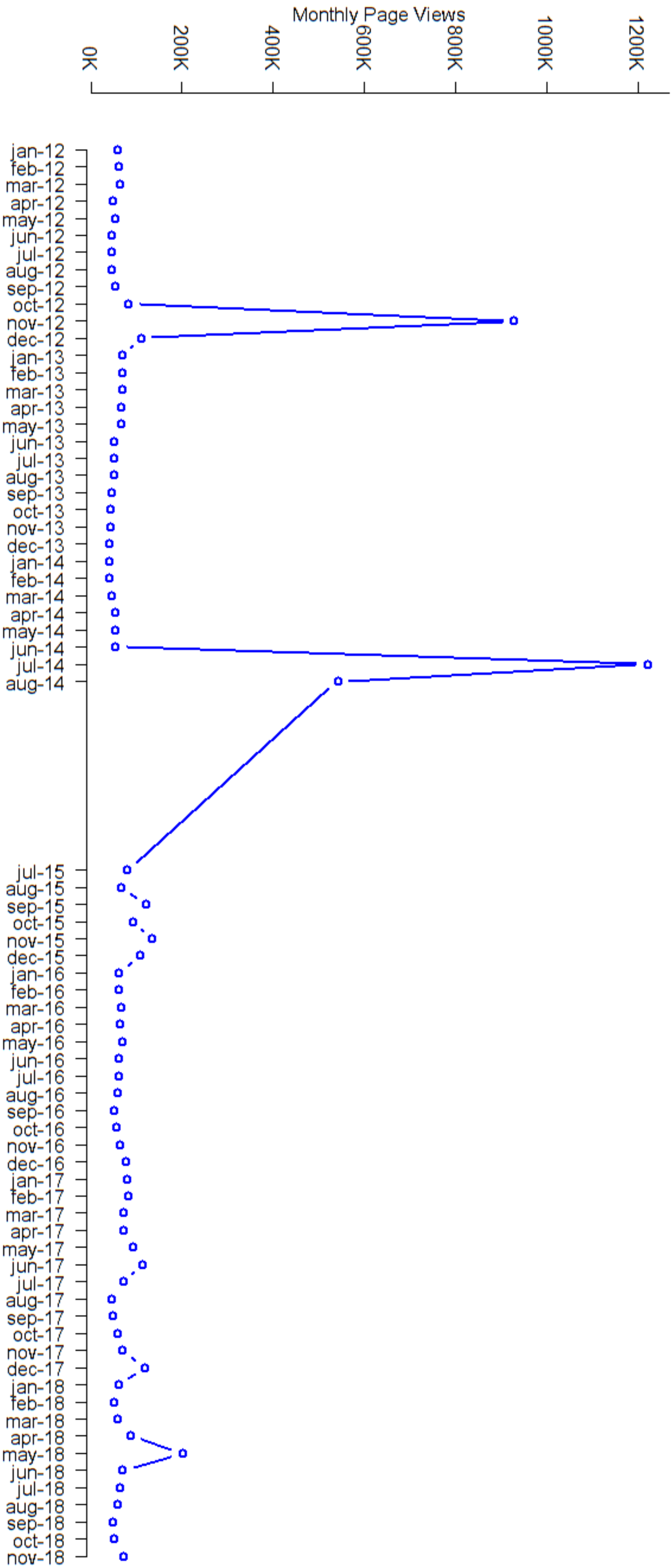

Page Views for Environmental_terrorist_NA_ism
Monthly Page Views
1.2K
1K
0.8K
0.6K
jan-12 feb-12 mar-12 apr-12 may-12 jun-12 jul-12 aug-12 sep-12 oct-12 nov-12 dec-12 jan-13 feb-13 mar-13 apr-13 may-13 jun-13 jul-13 aug-13 sep-13 oct-13 nov-13 dec-13 jan-14 feb-14 mar-14 apr-14 may-14 jun-14 jul-14 aug-14
jul-15 aug-15 sep-15 oct-15 nov-15 dec-15 jan-16 feb-16 mar-16 apr-16 may-16 jun-16 jul-16 aug-16 sep-16 oct-16 nov-16 dec-16 jan-17 feb-17 mar-17 apr-17 may-17 jun-17 jul-17 aug-17 sep-17 oct-17 nov-17 dec-17 jan-18 feb-18 mar-18 apr-18 may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18

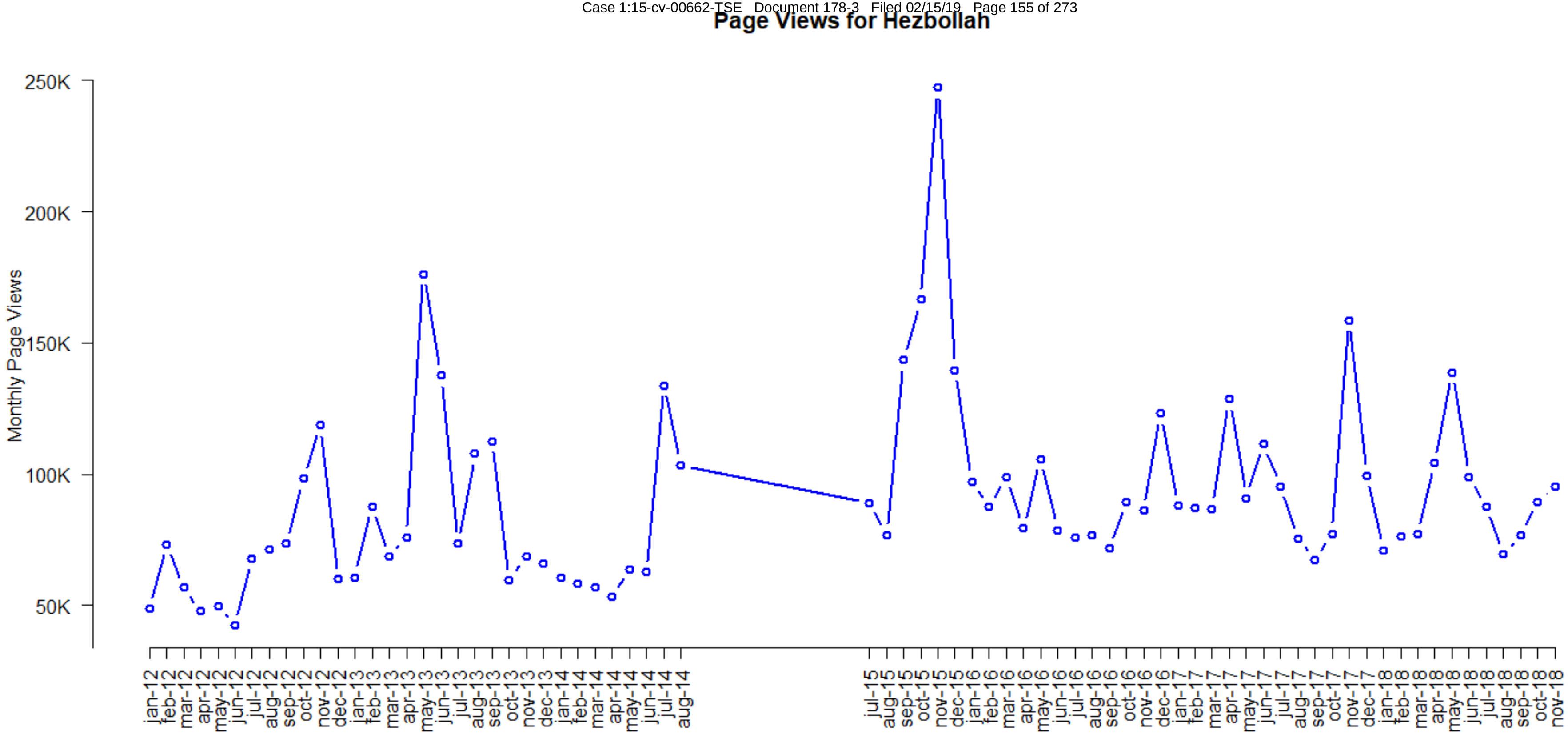

Page Views for Extremism
Monthly Page Views
20K
18K
16K
14K
12K
10K
8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

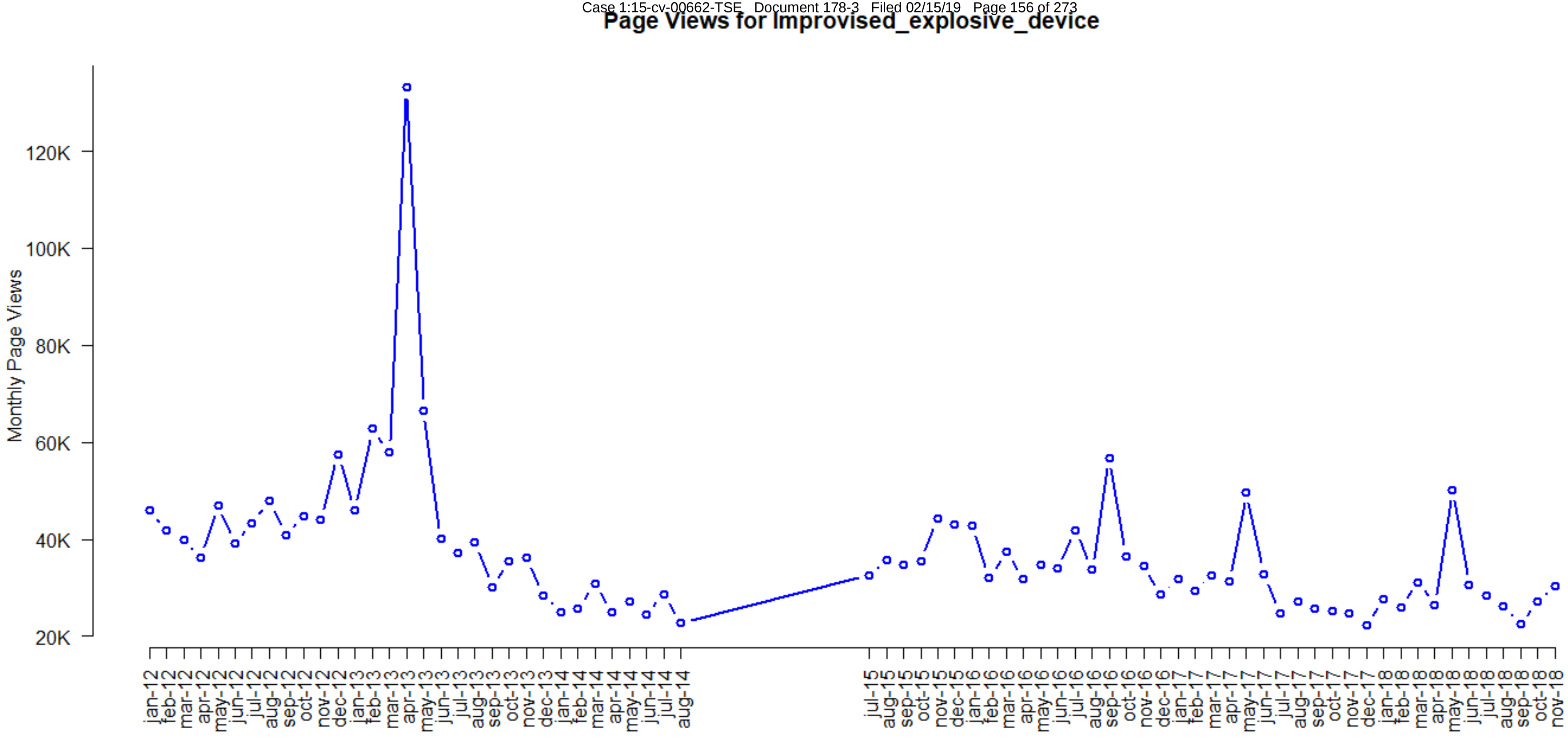

Page Views for FARC
Monthly Page Views
250K
200K
150K
100K
50K
0K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18


Page Views for Fundamentalism
Monthly Page Views
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

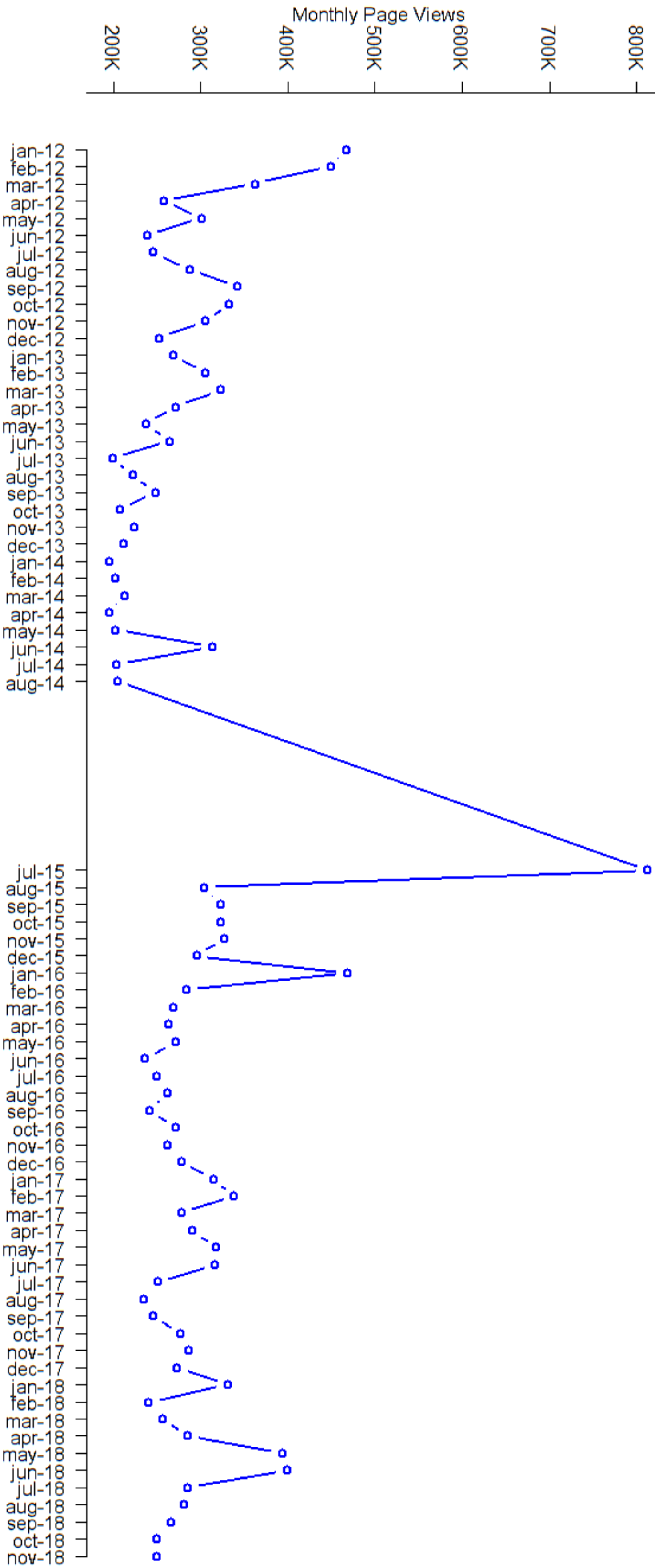
Page Views for Hamas



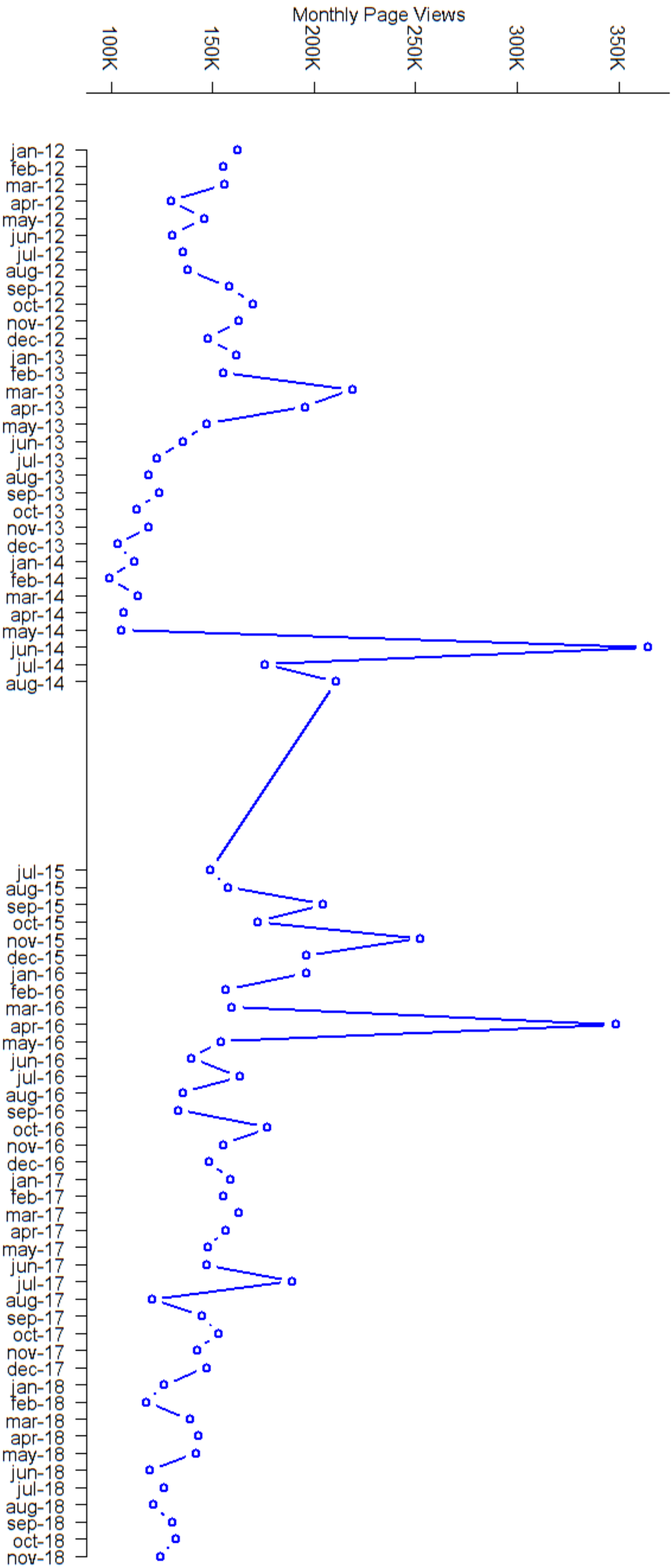

Page Views for Hezbollah
Monthly Page Views
250K
200K
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

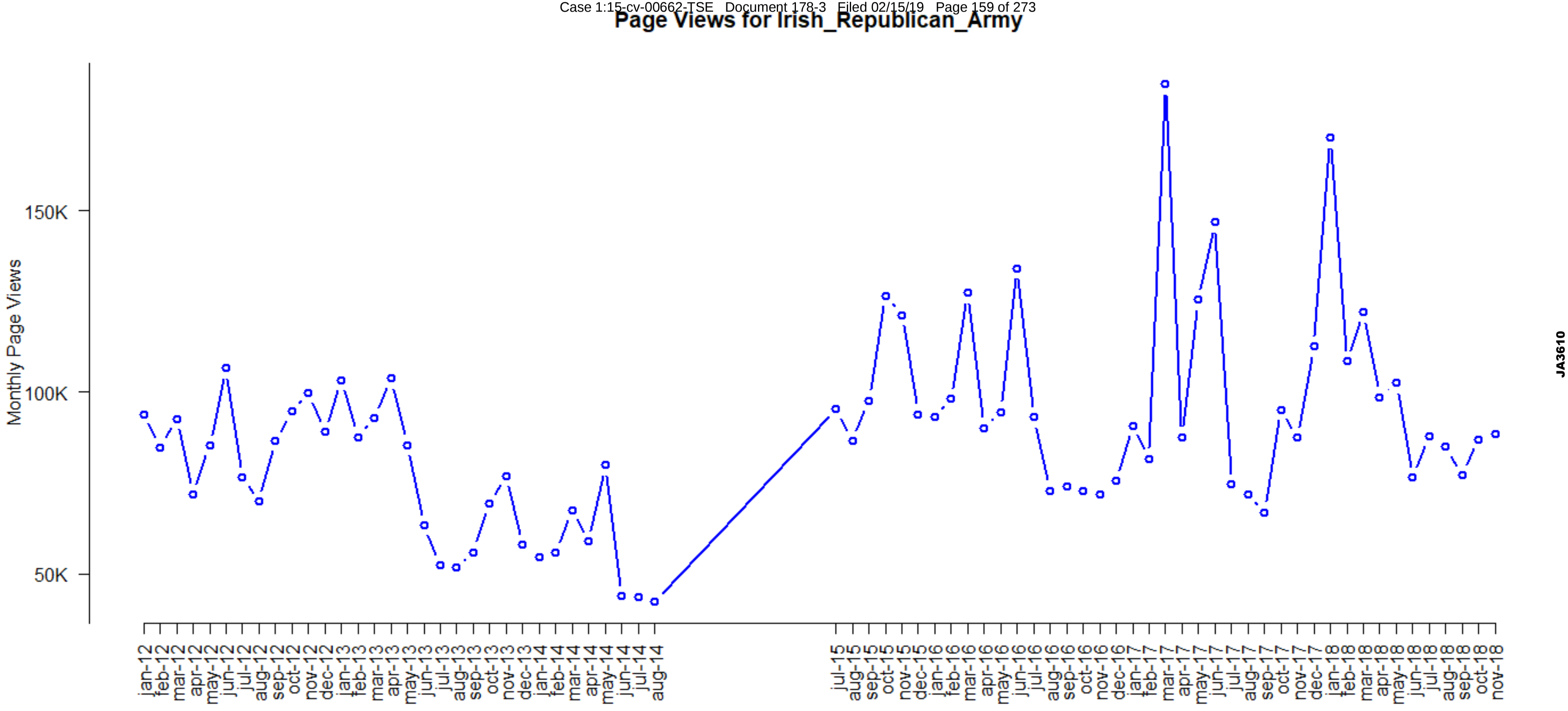

Page Views for Improvised_explosive_device
Monthly Page Views
120K
100K
80K
60K
40K
20K
jan-12 feb-12 mar-12 apr-12 may-12 jun-12 jul-12 aug-12 sep-12 oct-12 nov-12 dec-12 jan-13 feb-13 mar-13 apr-13 may-13 jun-13 jul-13 aug-13 sep-13 oct-13 nov-13 dec-13 jan-14 feb-14 mar-14 apr-14 may-14 jun-14 jul-14 aug-14
jul-15 aug-15 sep-15 oct-15 nov-15 dec-15 jan-16 feb-16 mar-16 apr-16 may-16 jun-16 jul-16 aug-16 sep-16 oct-16 nov-16 dec-16 jan-17 feb-17 mar-17 apr-17 may-17 jun-17 jul-17 aug-17 sep-17 oct-17 nov-17 dec-17 jan-18 feb-18 mar-18 apr-18 may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18

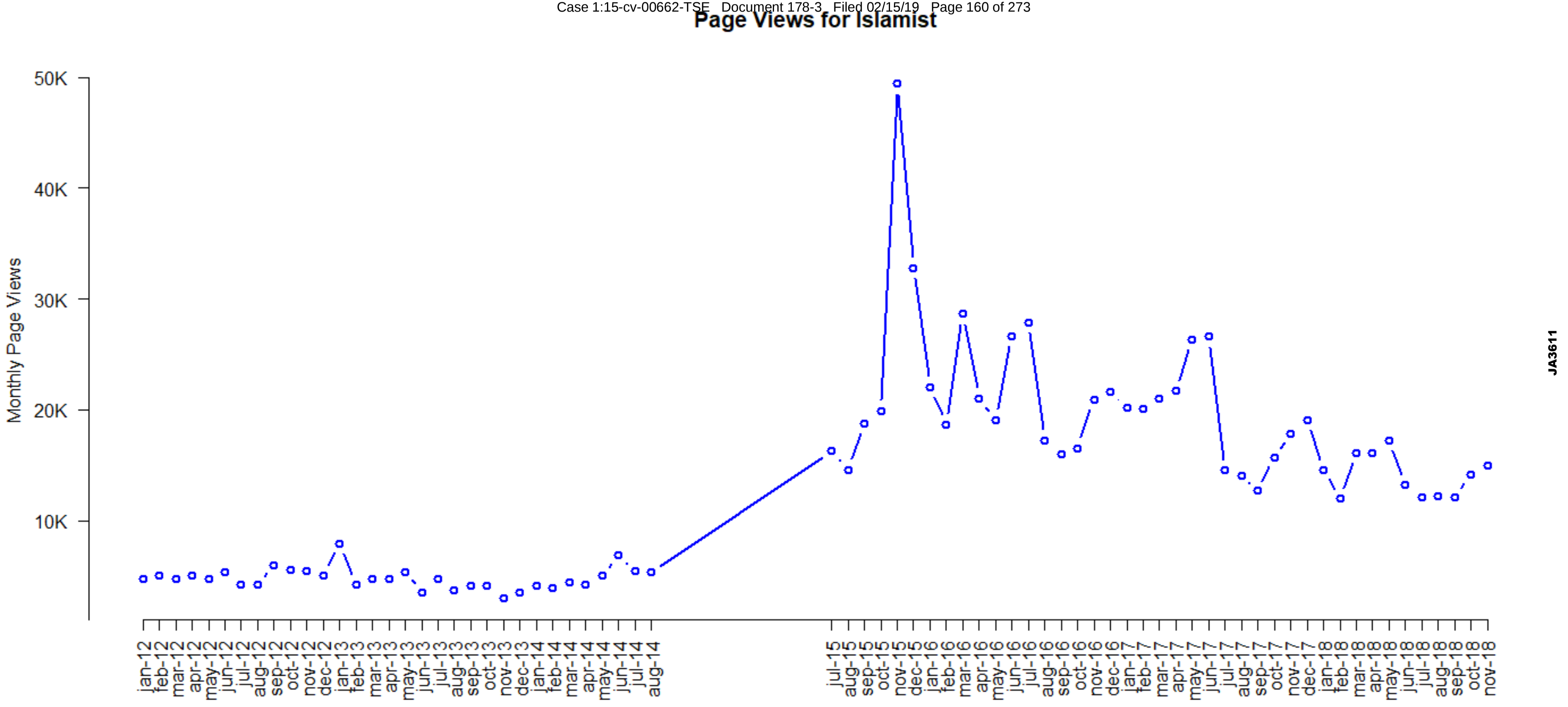
Page Views for Iran

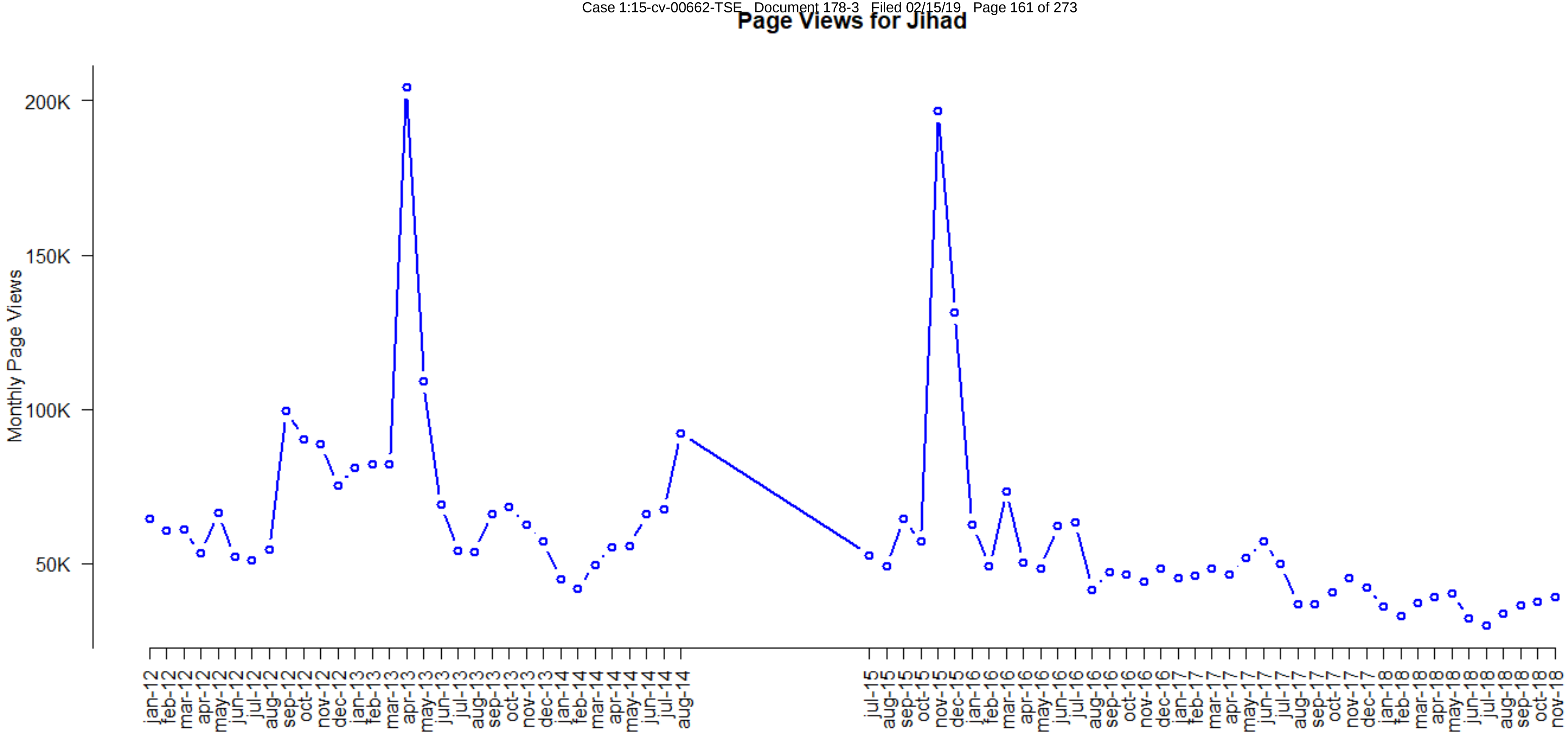


Page Views for Iraq

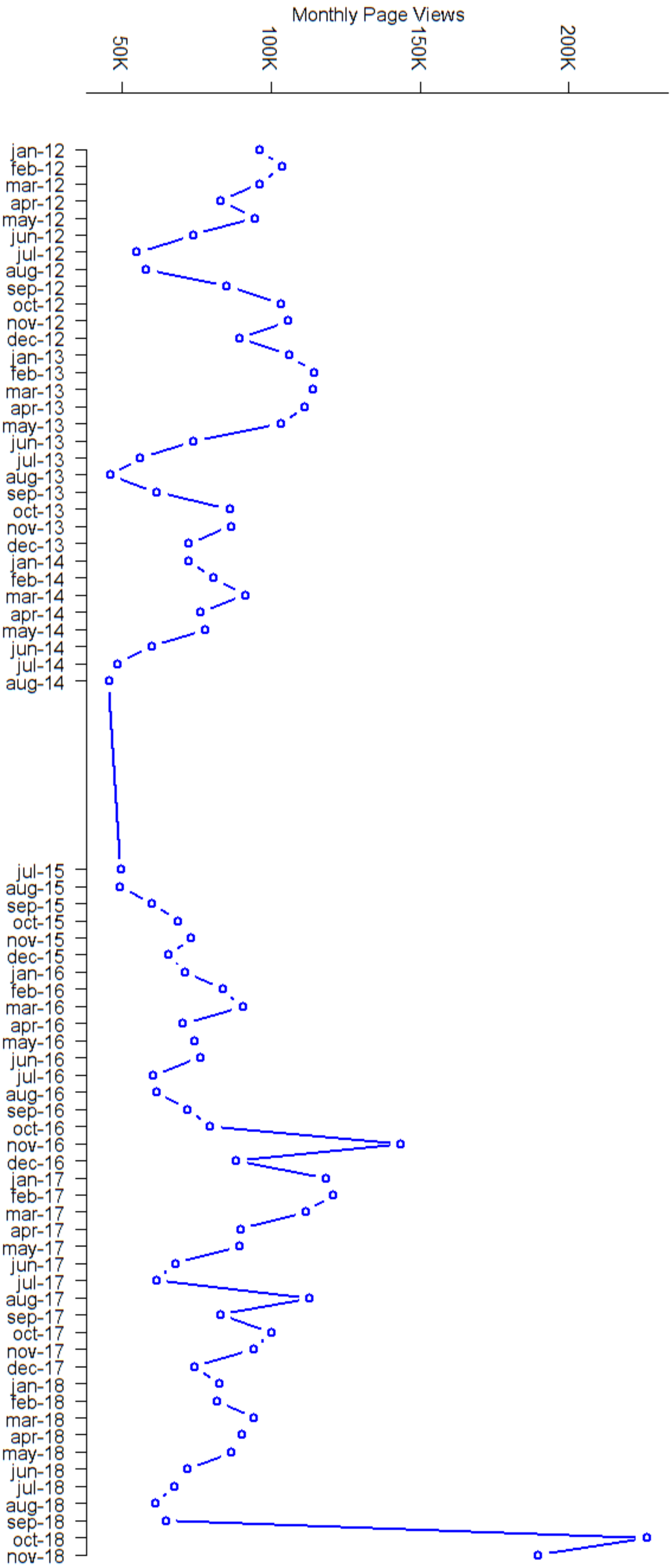


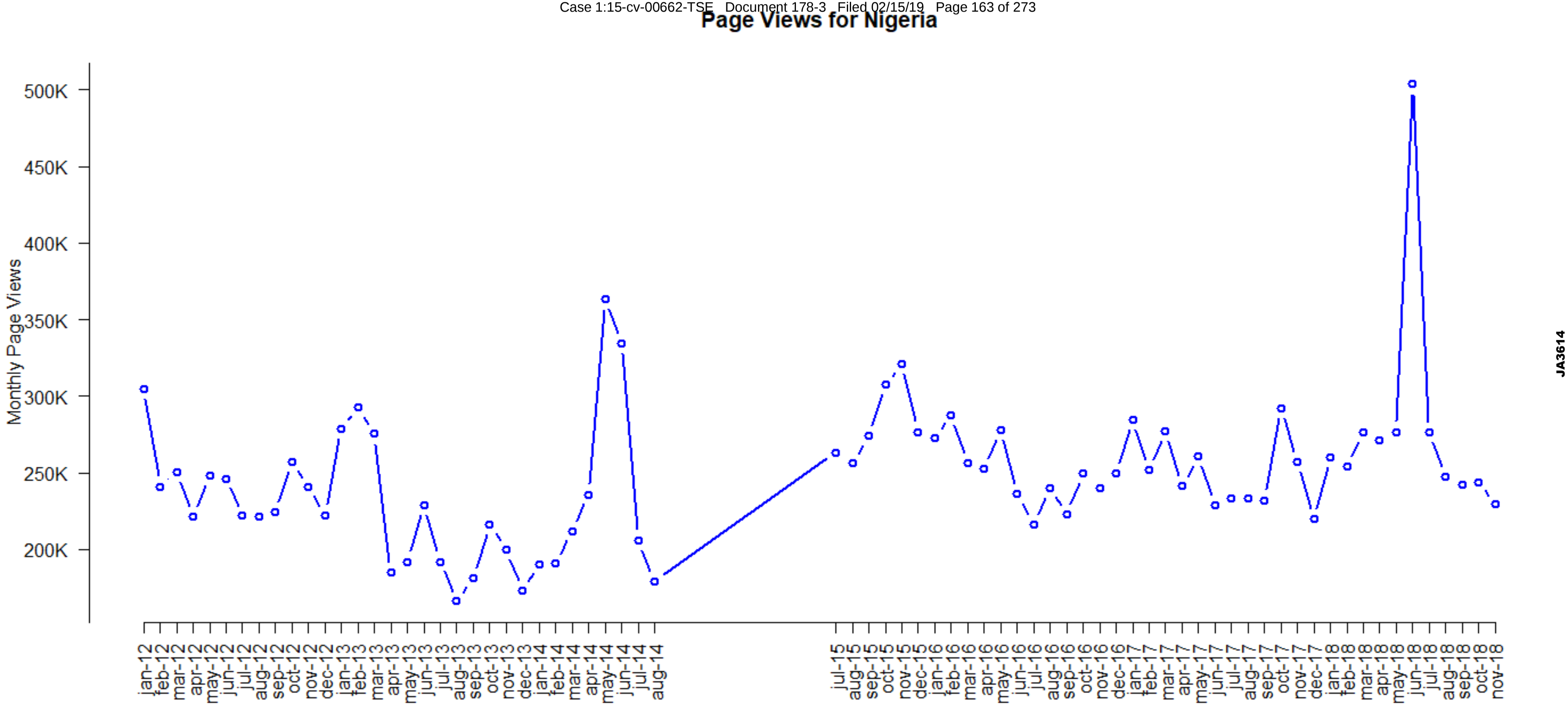

Page Views for Irish_Republican_Army
Monthly Page Views
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

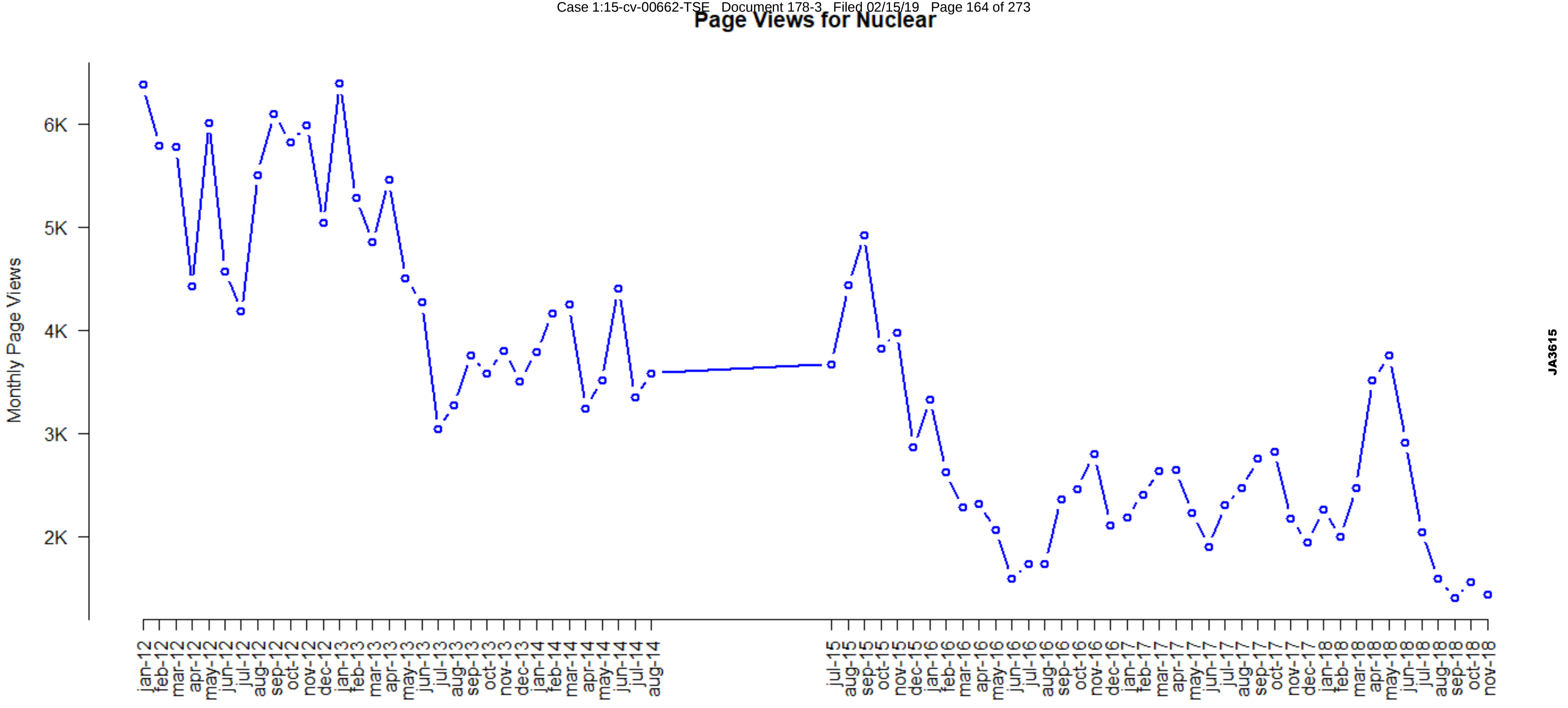

Page Views for Islamist
Monthly Page Views
50K
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

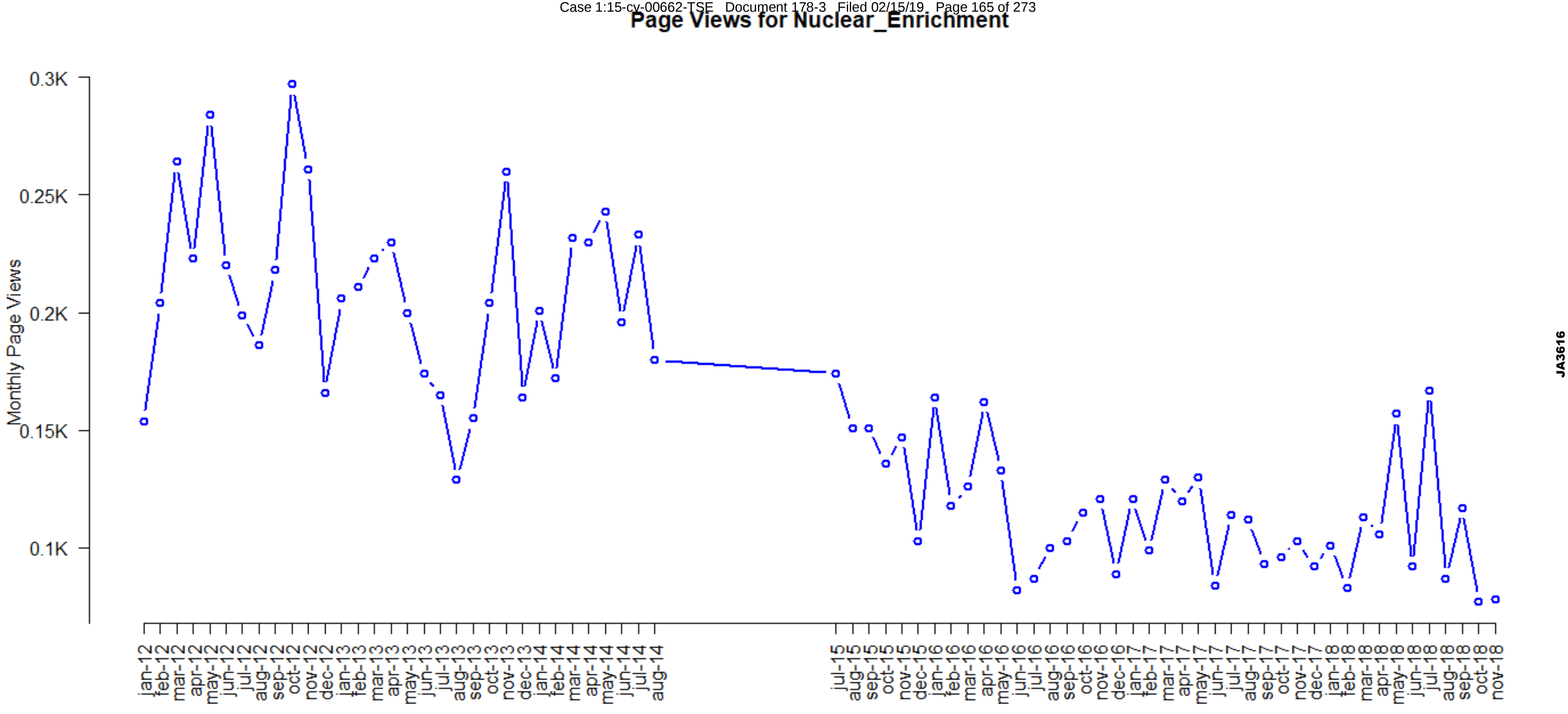

Page Views for Jihad
Monthly Page Views
200K
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

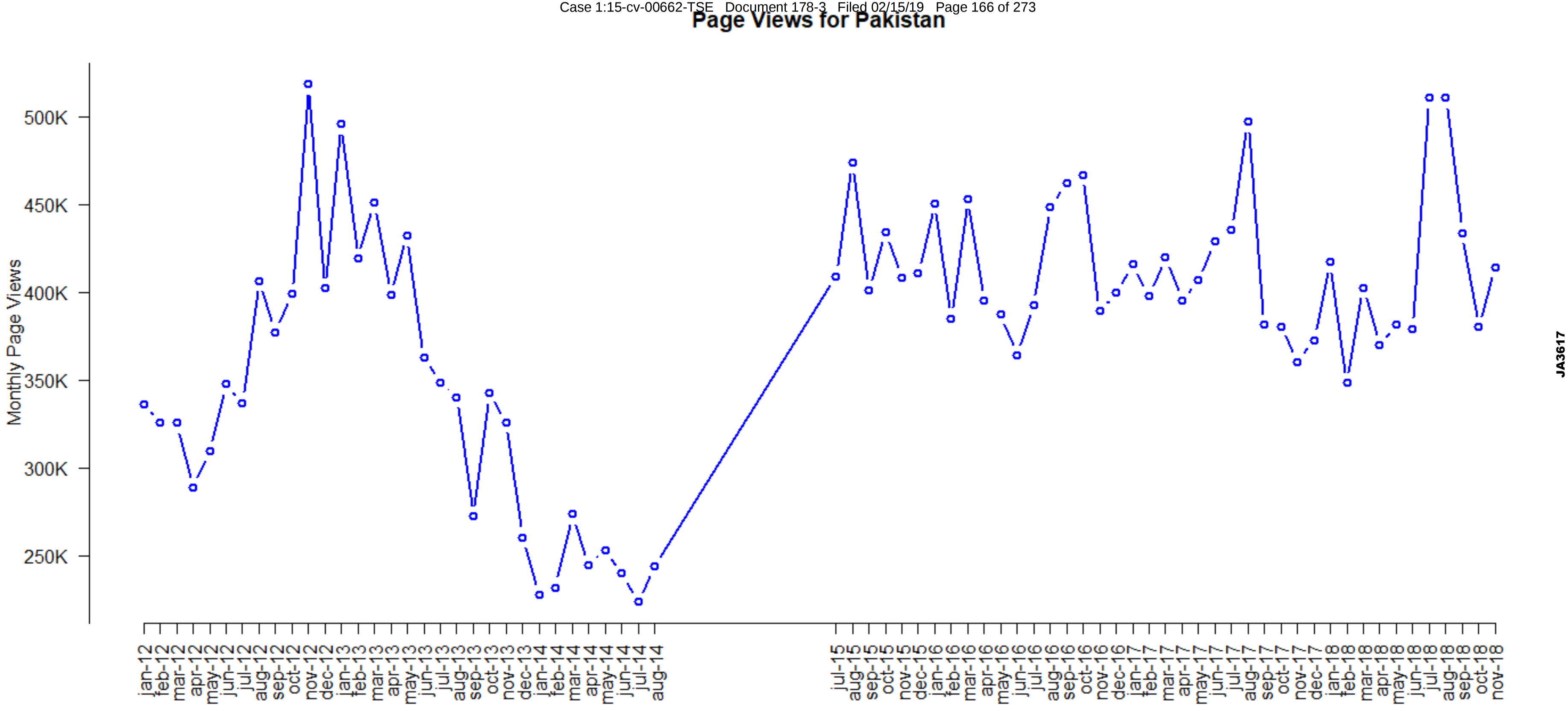
Page Views for nationalism



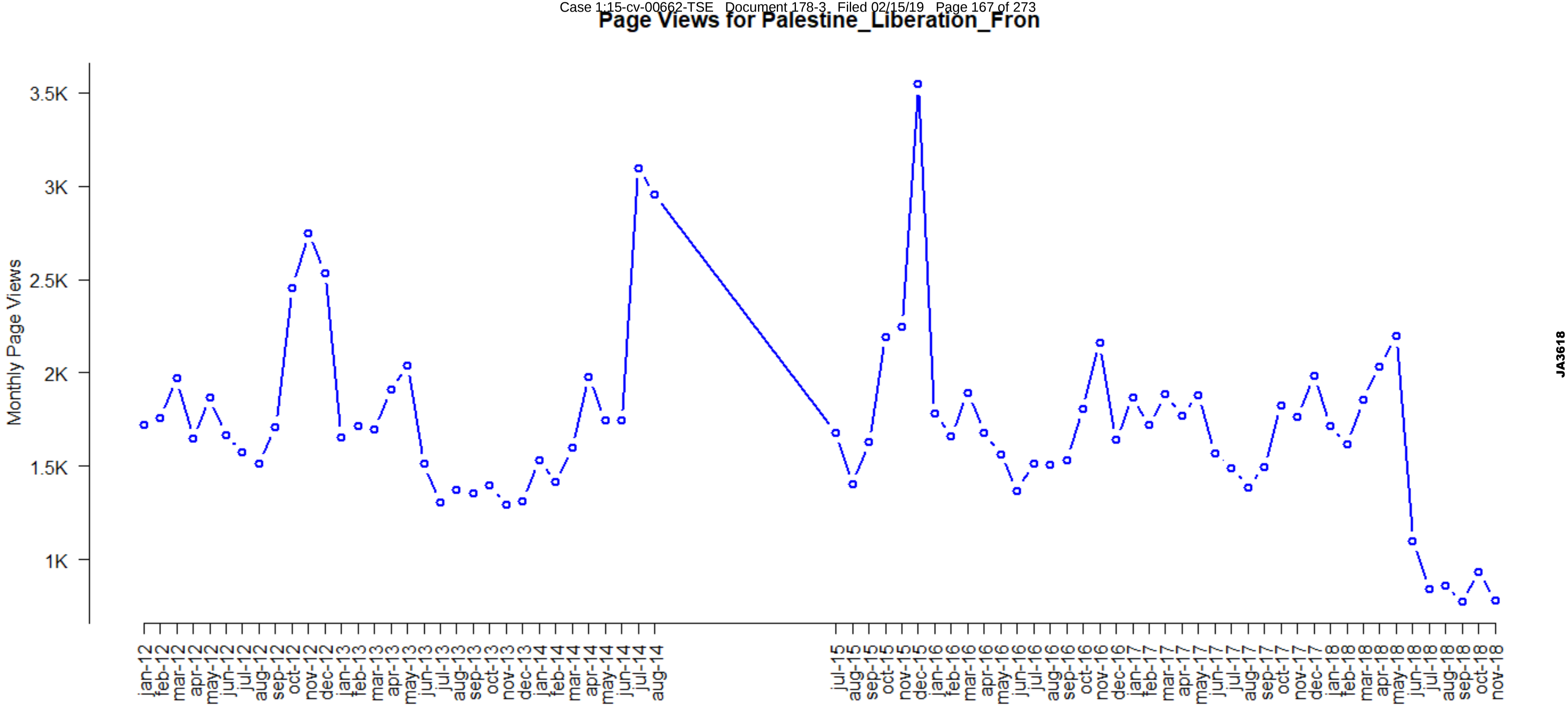

Page Views for Nigeria
Monthly Page Views
500K
450K
400K
350K
300K
250K
200K
jan-12 feb-12 mar-12 apr-12 may-12 jun-12 jul-12 aug-12 sep-12 oct-12 nov-12 dec-12 jan-13 feb-13 mar-13 apr-13 may-13 jun-13 jul-13 aug-13 sep-13 oct-13 nov-13 dec-13 jan-14 feb-14 mar-14 apr-14 may-14 jun-14 jul-14 aug-14
jul-15 aug-15 sep-15 oct-15 nov-15 dec-15 jan-16 feb-16 mar-16 apr-16 may-16 jun-16 jul-16 aug-16 sep-16 oct-16 nov-16 dec-16 jan-17 feb-17 mar-17 apr-17 may-17 jun-17 jul-17 aug-17 sep-17 oct-17 nov-17 dec-17 jan-18 feb-18 mar-18 apr-18 may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18


Page Views for Nuclear
Monthly Page Views
6K
5K
4K
3K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

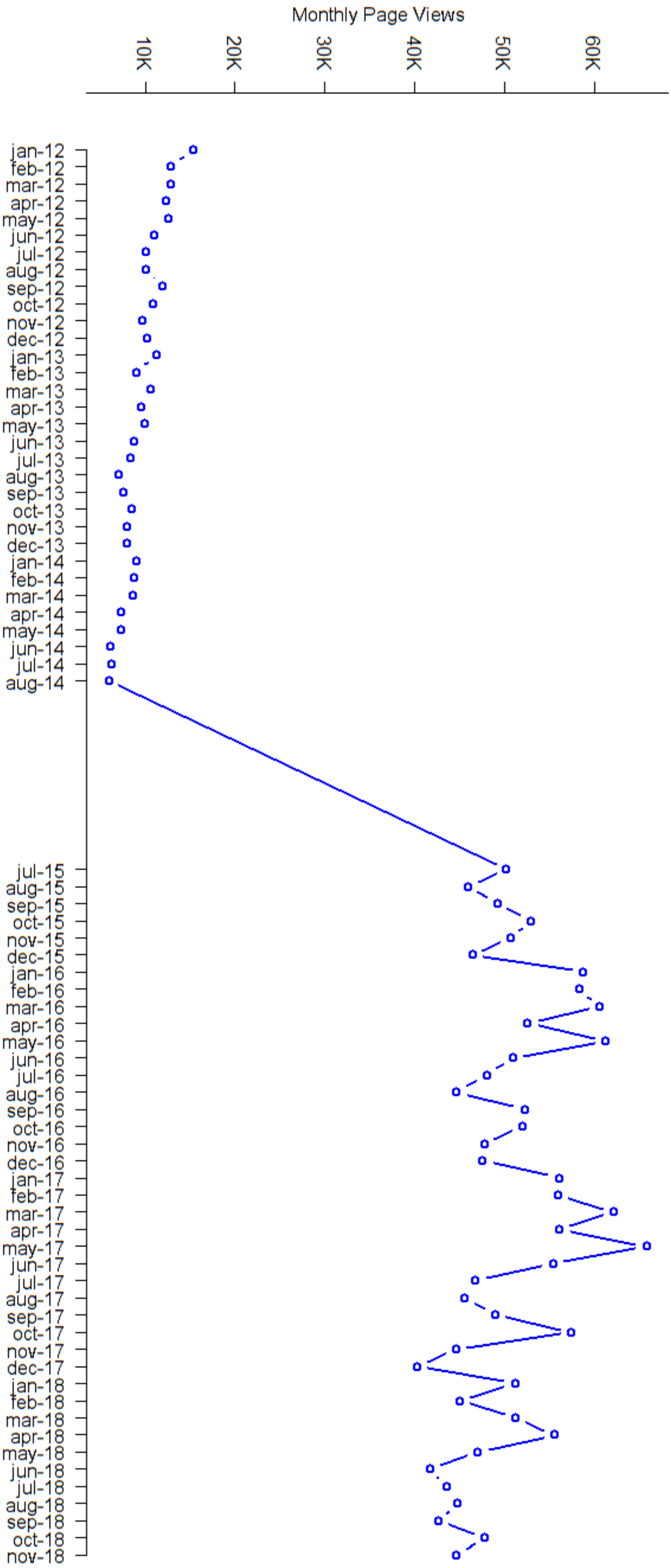

Page Views for Nuclear_Enrichment
Monthly Page Views
0.3K
0.25K
0.2K
0.15K
0.1K
jan-12 feb-12 mar-12 apr-12 may-12 jun-12 jul-12 aug-12 sep-12 oct-12 nov-12 dec-12 jan-13 feb-13 mar-13 apr-13 may-13 jun-13 jul-13 aug-13 sep-13 oct-13 nov-13 dec-13 jan-14 feb-14 mar-14 apr-14 may-14 jun-14 jul-14 aug-14
jul-15 aug-15 sep-15 oct-15 nov-15 dec-15 jan-16 feb-16 mar-16 apr-16 may-16 jun-16 jul-16 aug-16 sep-16 oct-16 nov-16 dec-16 jan-17 feb-17 mar-17 apr-17 may-17 jun-17 jul-17 aug-17 sep-17 oct-17 nov-17 dec-17 jan-18 feb-18 mar-18 apr-18 may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18

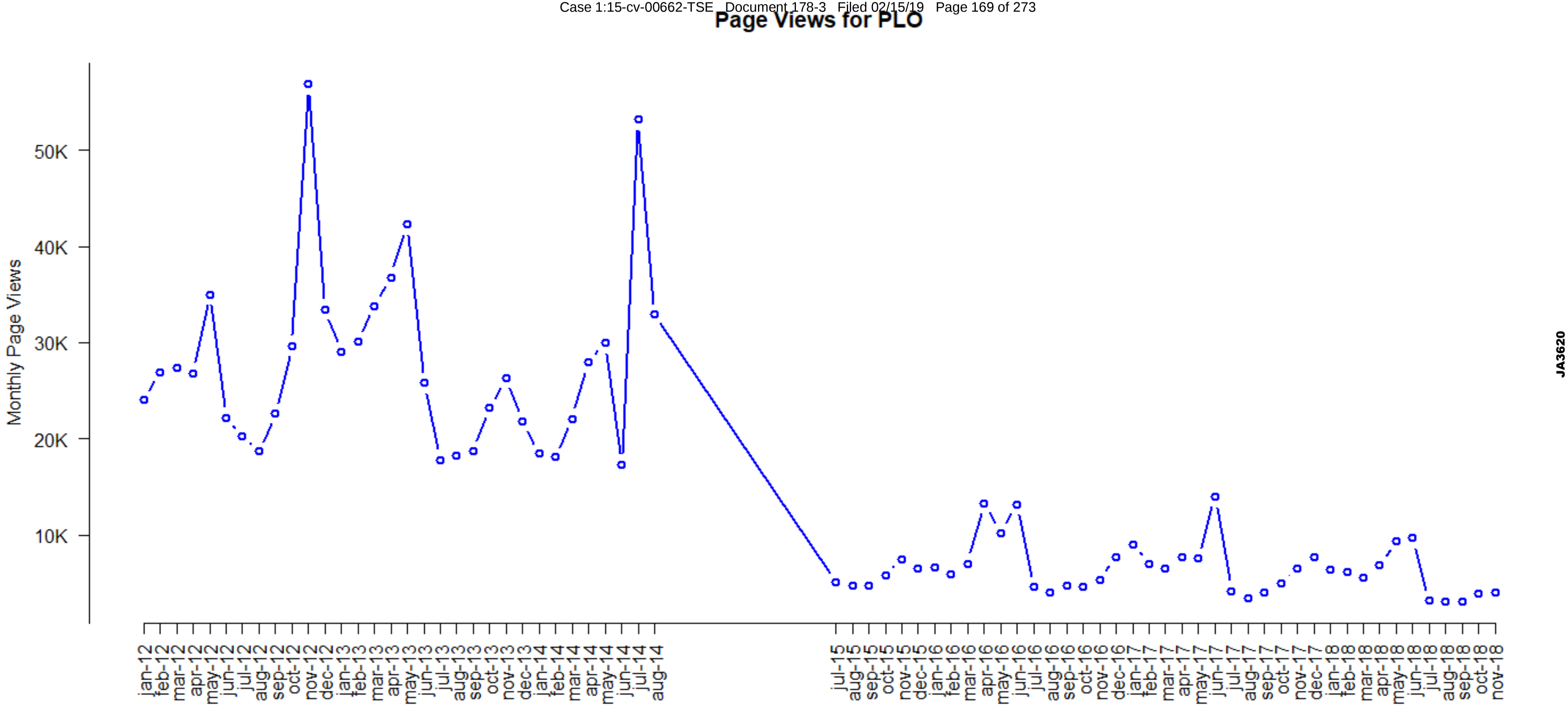

Page Views for Pakistan
Monthly Page Views
500K
450K
400K
350K
300K
250K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

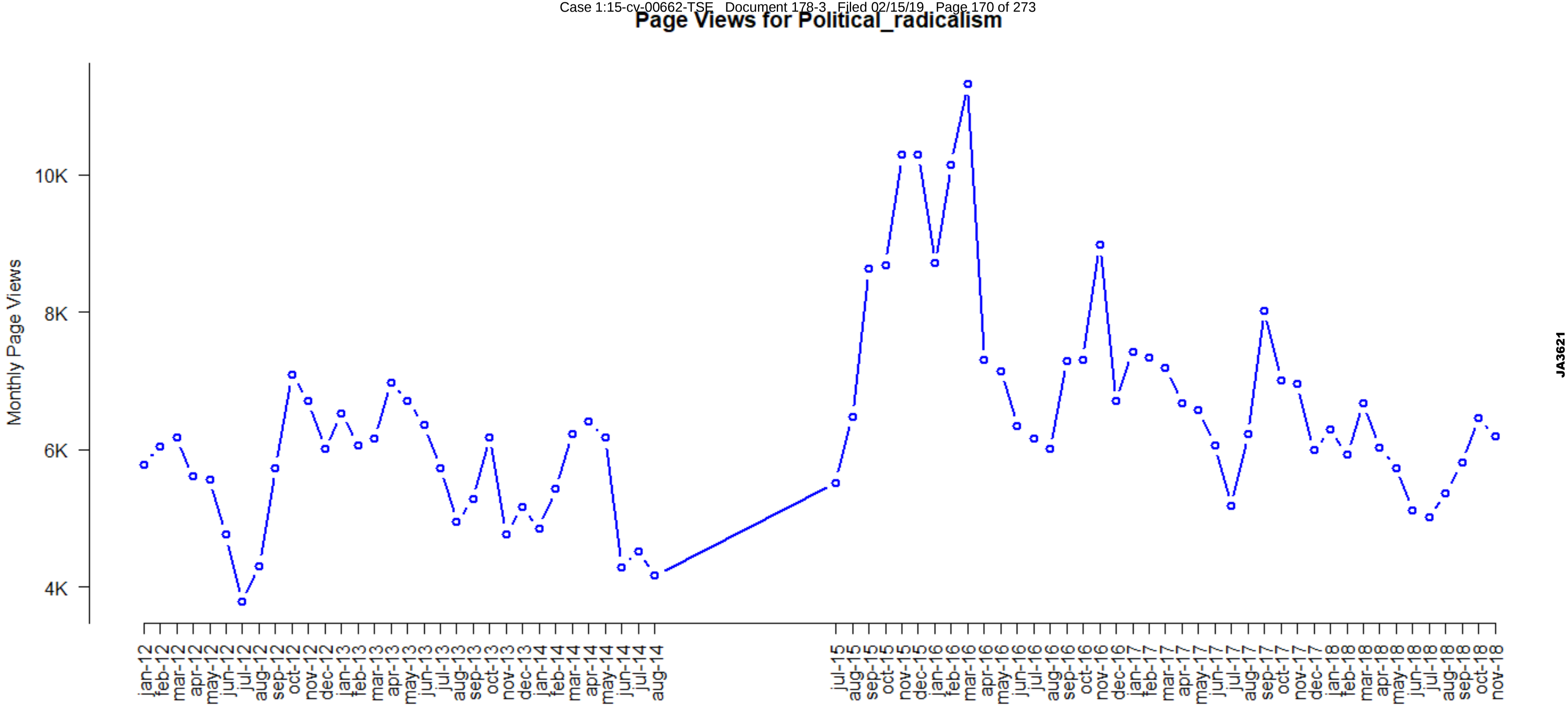
JA3617


Page Views for Palestine_Liberation_Fron
Monthly Page Views
3.5K
3K
2.5K
2K
1.5K
1K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

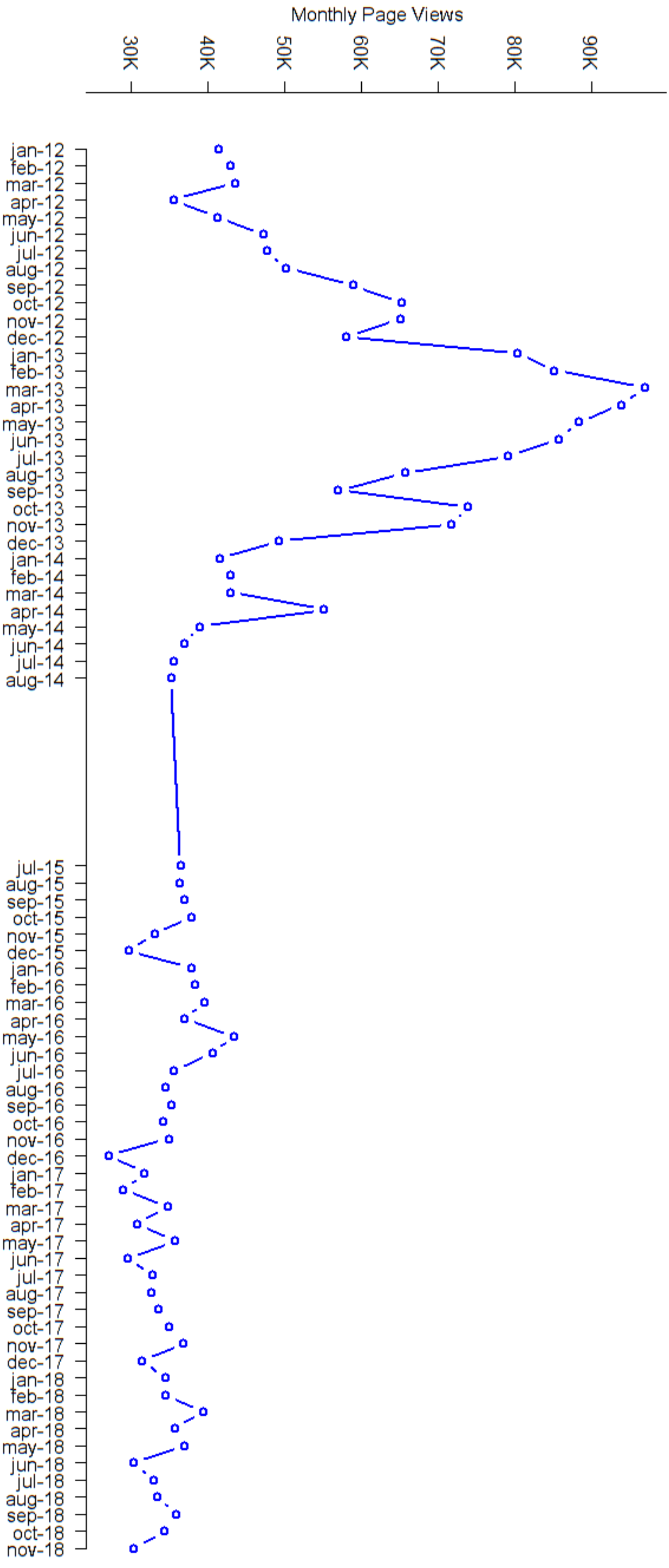
Page Views for Pirates

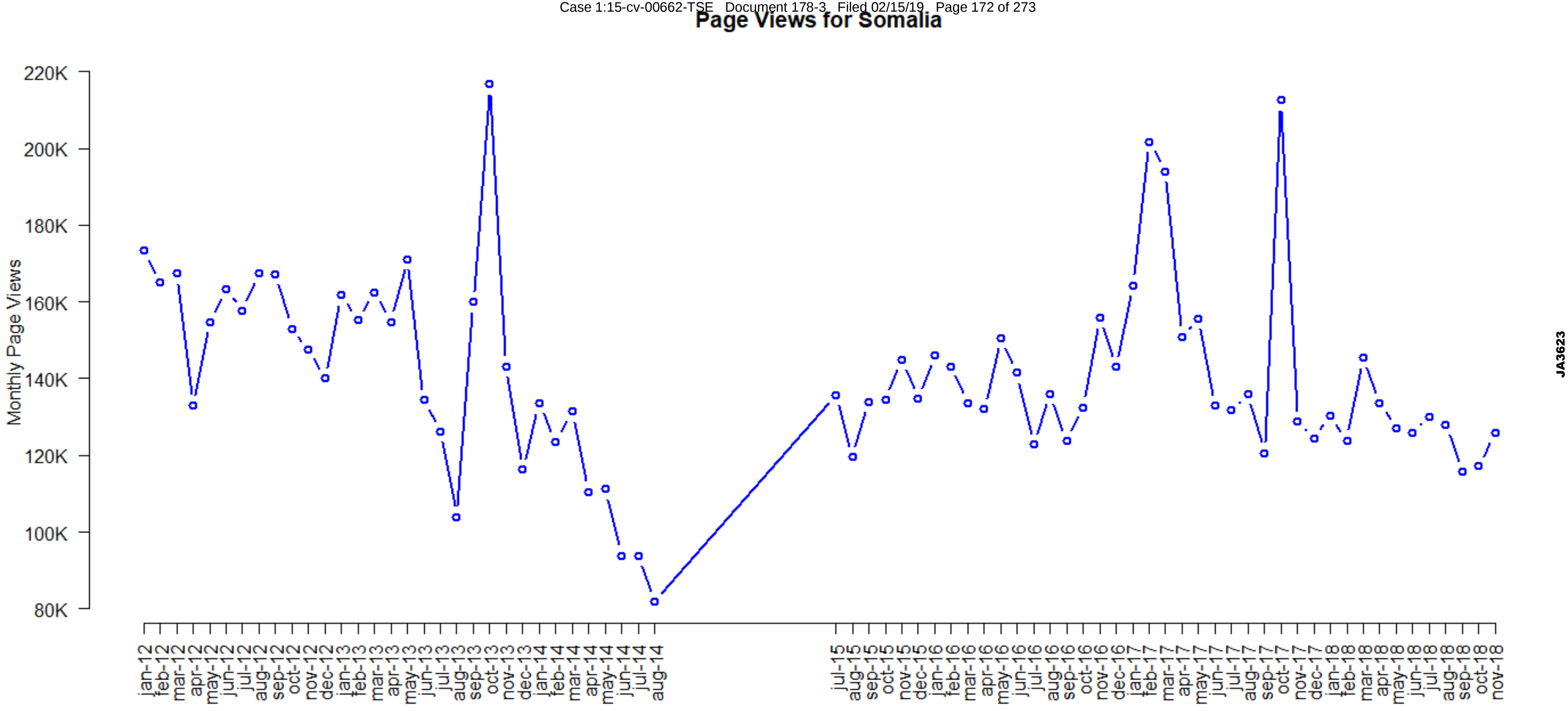


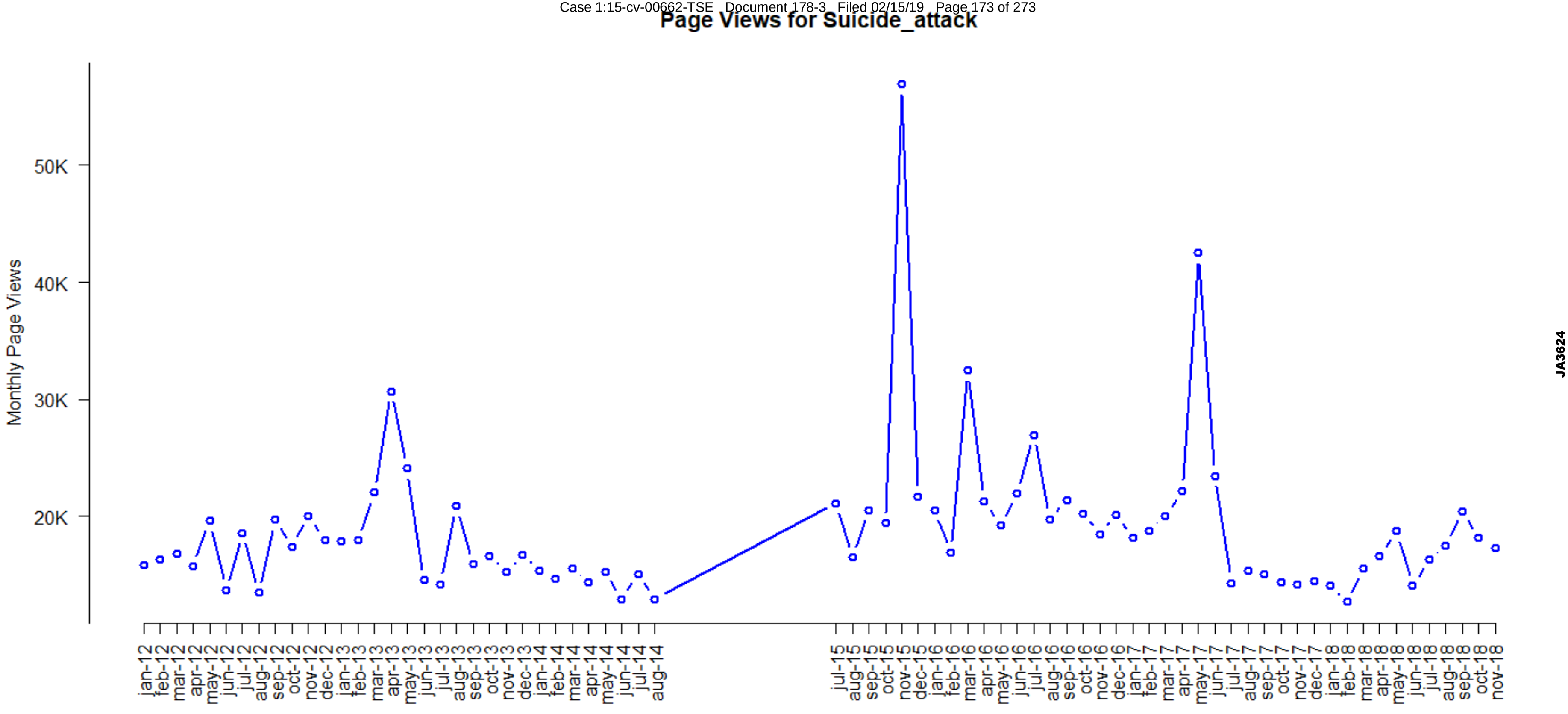

Page Views for PLO
Monthly Page Views
50K
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

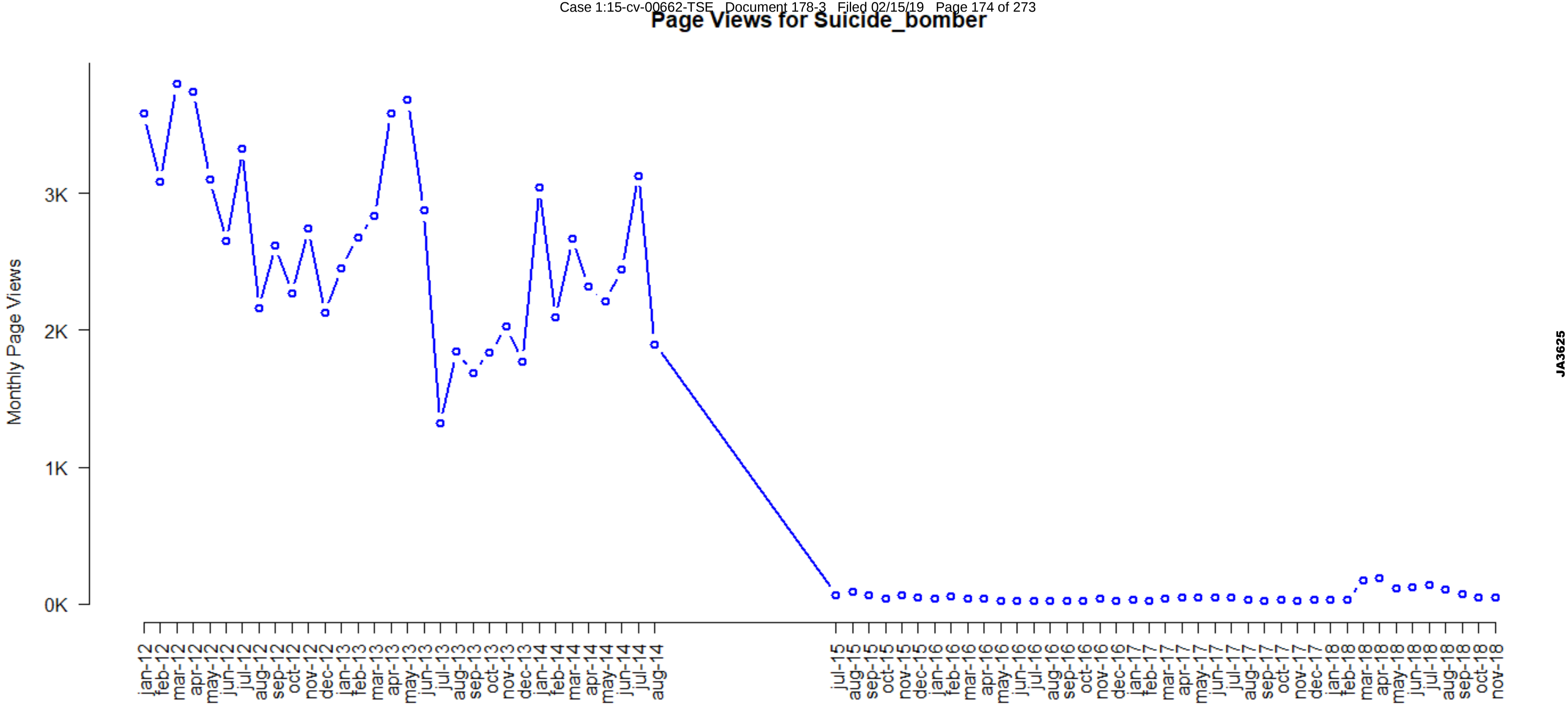

Page Views for Political_radicalism
Monthly Page Views
10K
8K
6K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

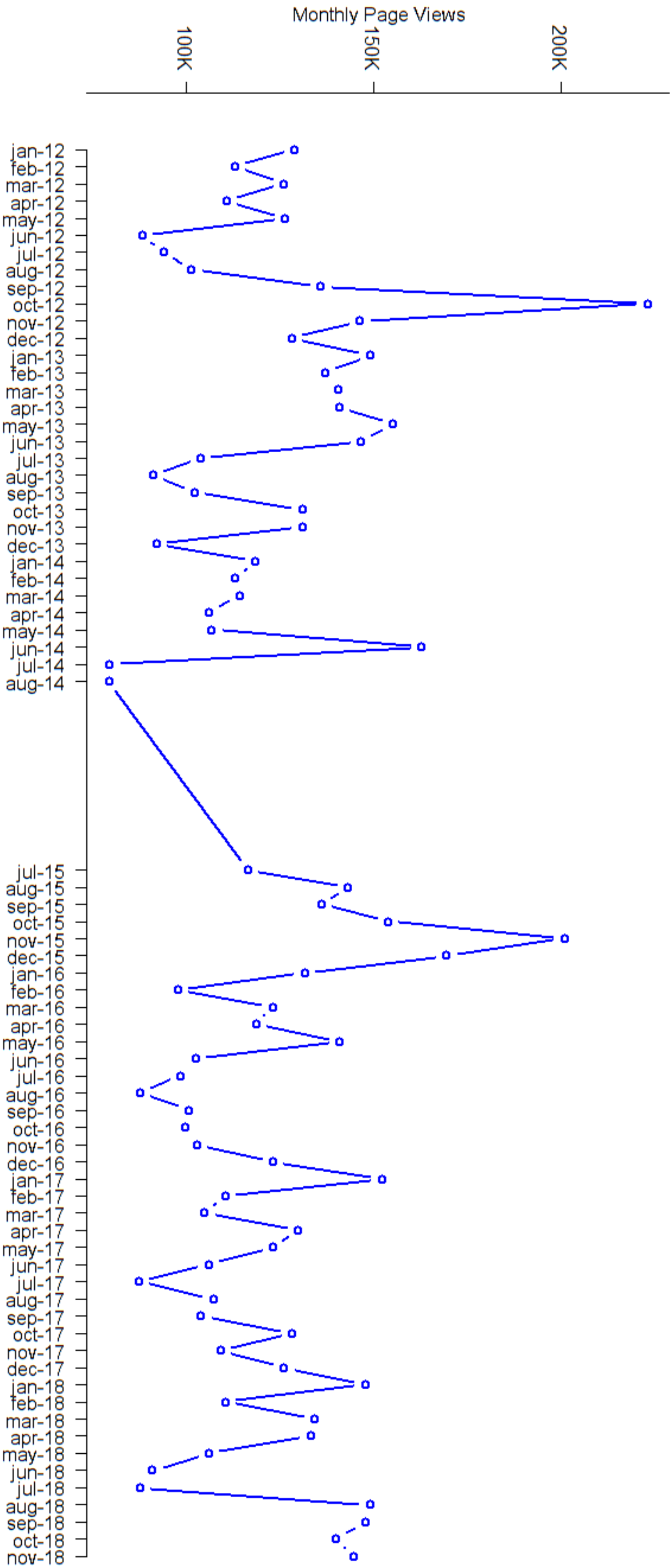
Page Views for Recruitment



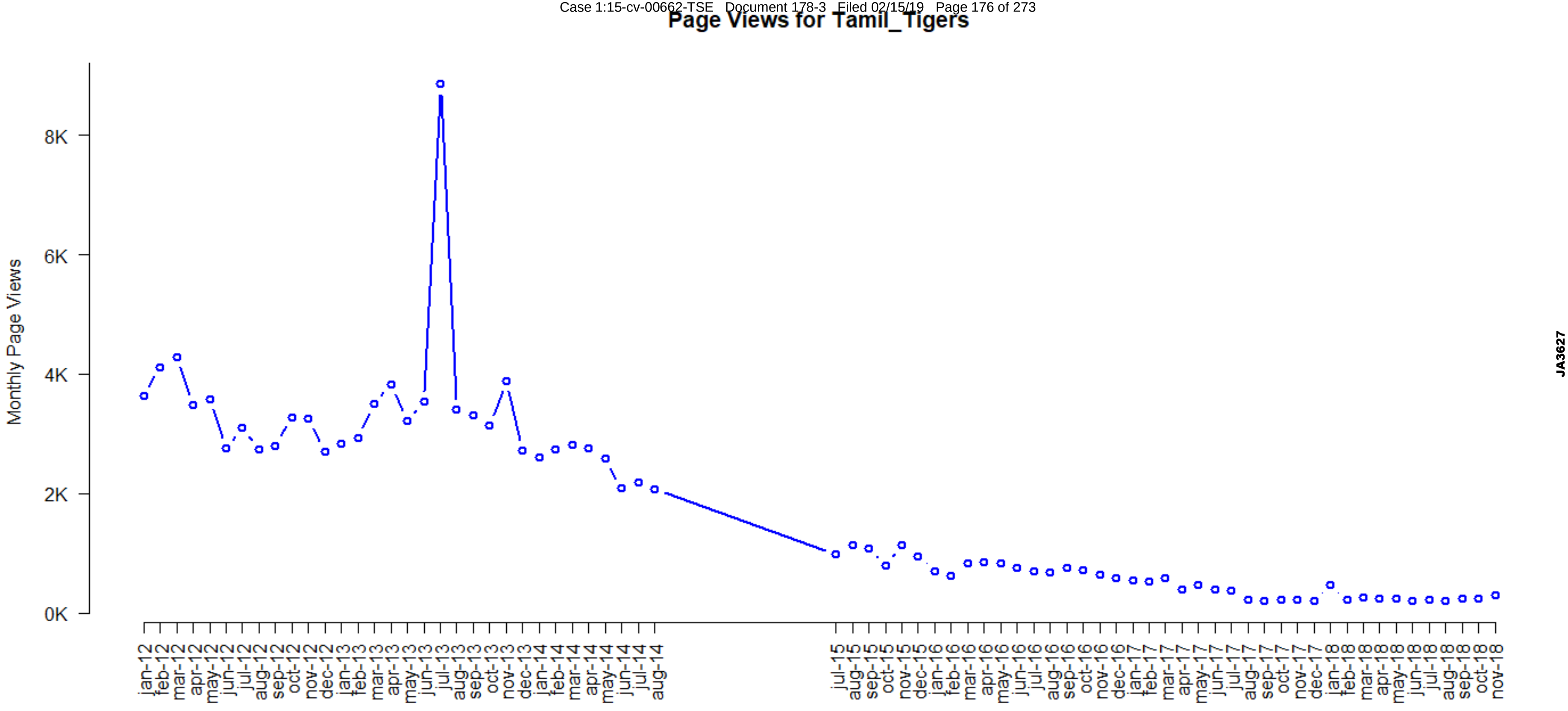

Page Views for Somalia
Monthly Page Views
220K
200K
180K
160K
140K
120K
100K
80K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

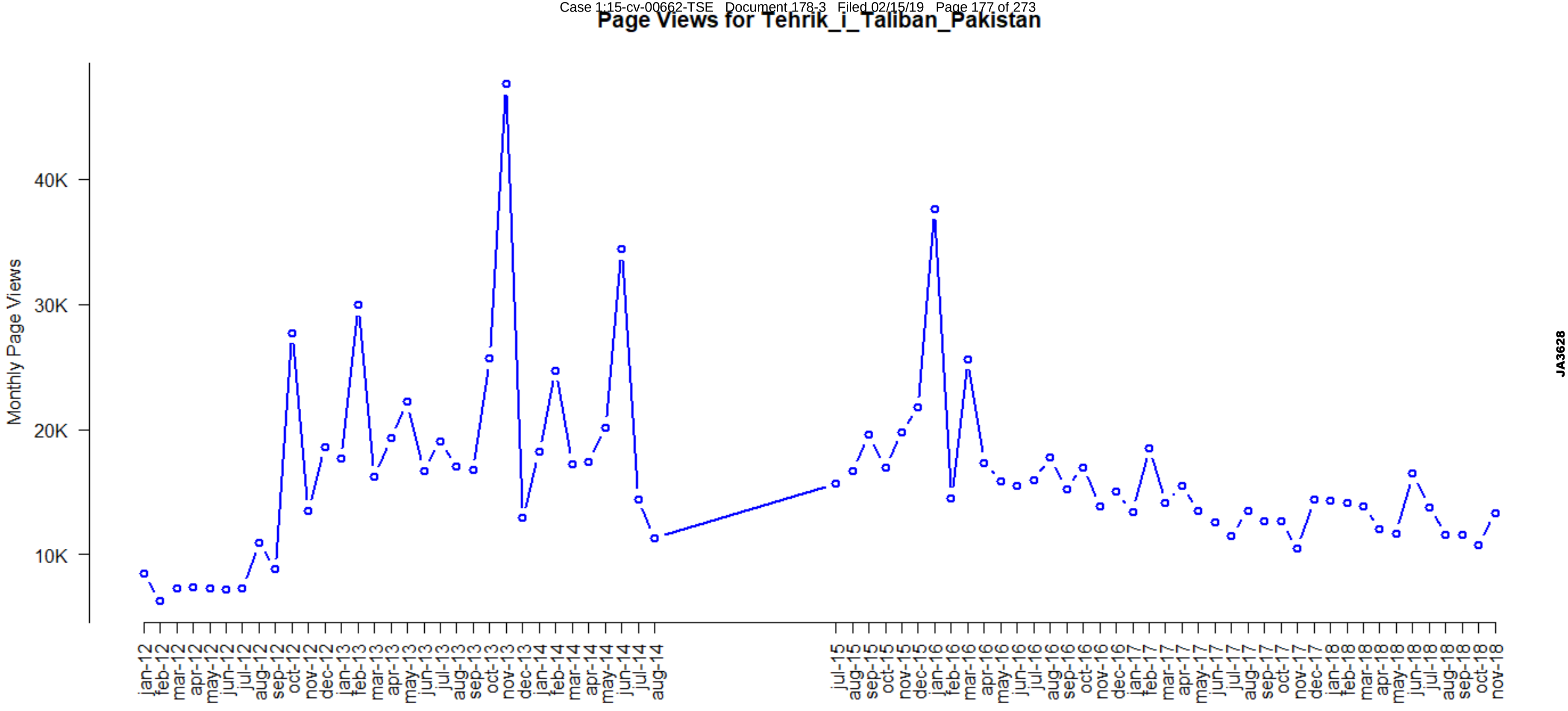

Page Views for Suicide_attack
Monthly Page Views
50K
40K
30K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

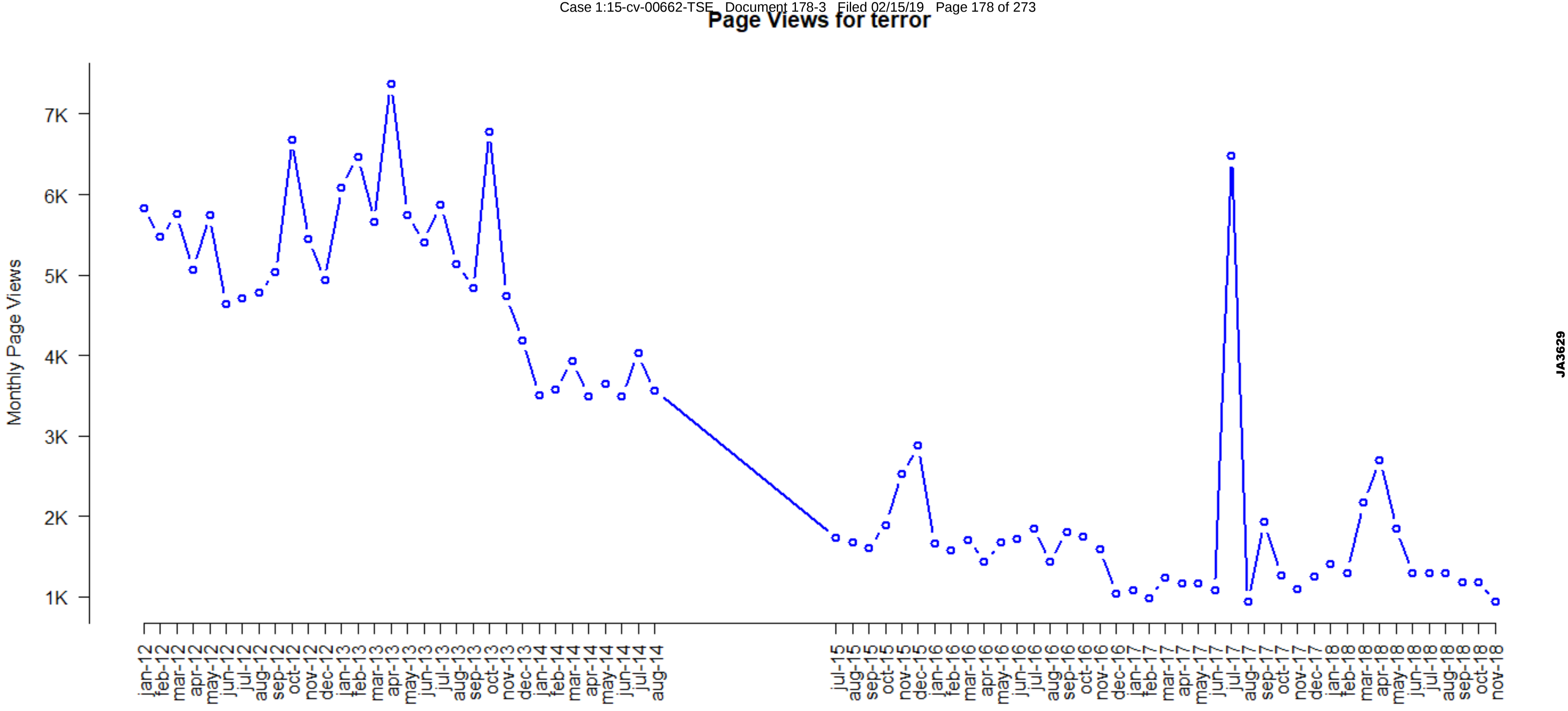

Page Views for Suicide_bomber
Monthly Page Views
0K
1K
2K
3K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18



Page Views for Taliban

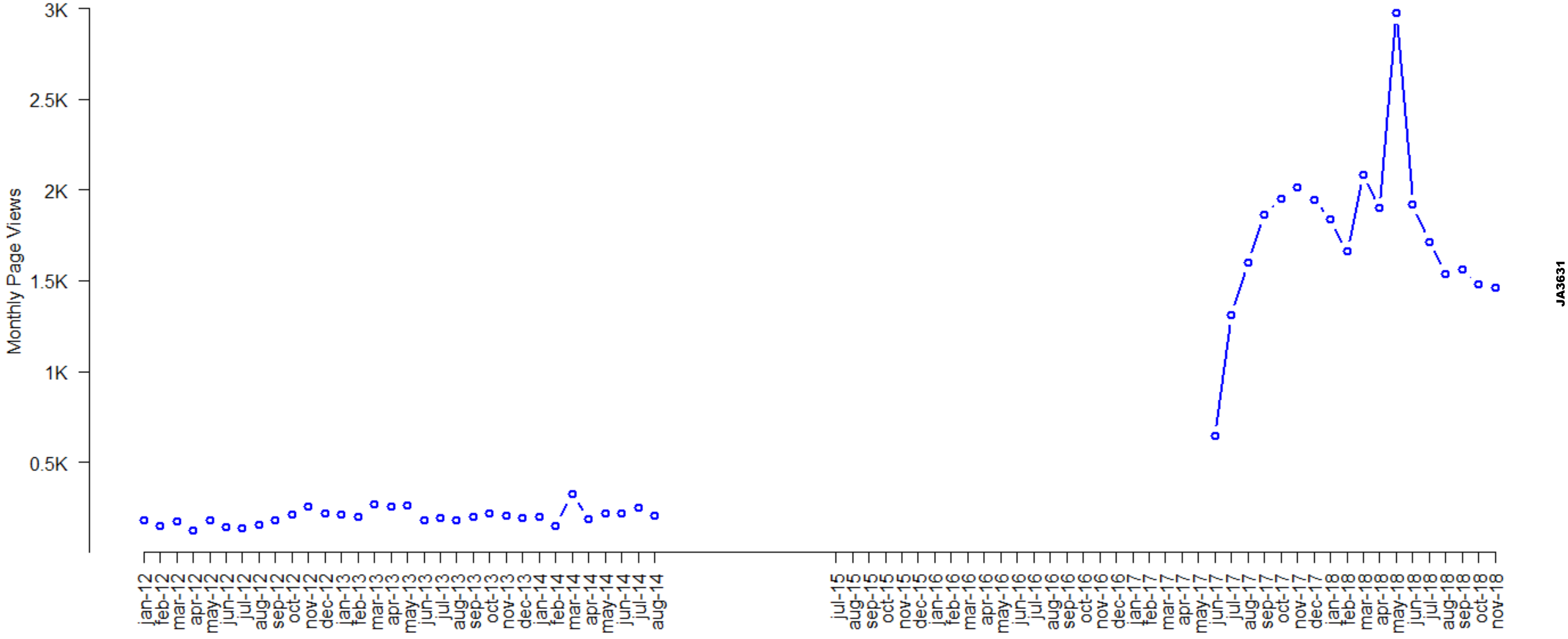

Page Views for Tamil_Tigers
Monthly Page Views
0K
2K
4K
6K
8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18


Page Views for Tehrik_i_Taliban_Pakistan
Monthly Page Views
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

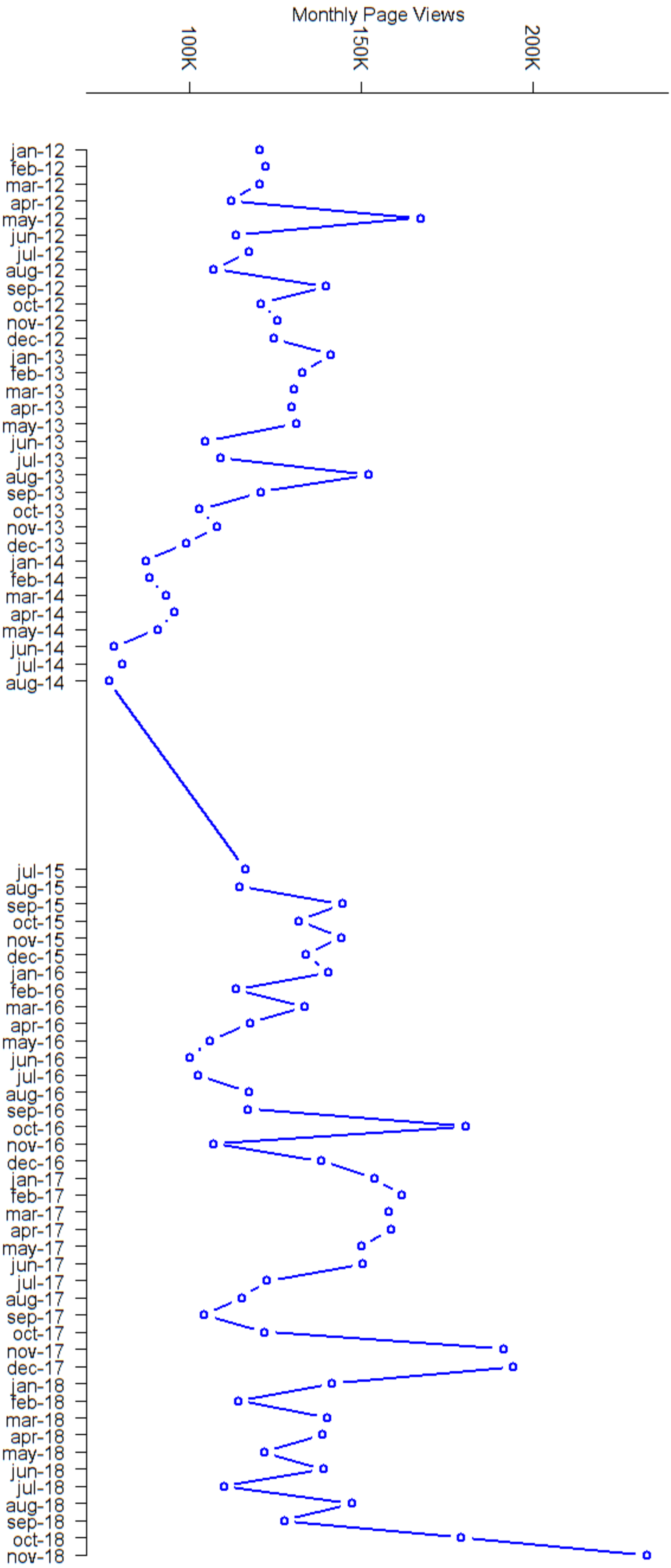

Page Views for terror
Monthly Page Views
7K
6K
5K
4K
3K
2K
1K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14
jul-15
aug-15
sep-15
oct-15
nov-15
dec-15
jan-16
feb-16
mar-16
apr-16
may-16
jun-16
jul-16
aug-16
sep-16
oct-16
nov-16
dec-16
jan-17
feb-17
mar-17
apr-17
may-17
jun-17
jul-17
aug-17
sep-17
oct-17
nov-17
dec-17
jan-18
feb-18
mar-18
apr-18
may-18
jun-18
jul-18
aug-18
sep-18
oct-18
nov-18

Page Views for terrorism

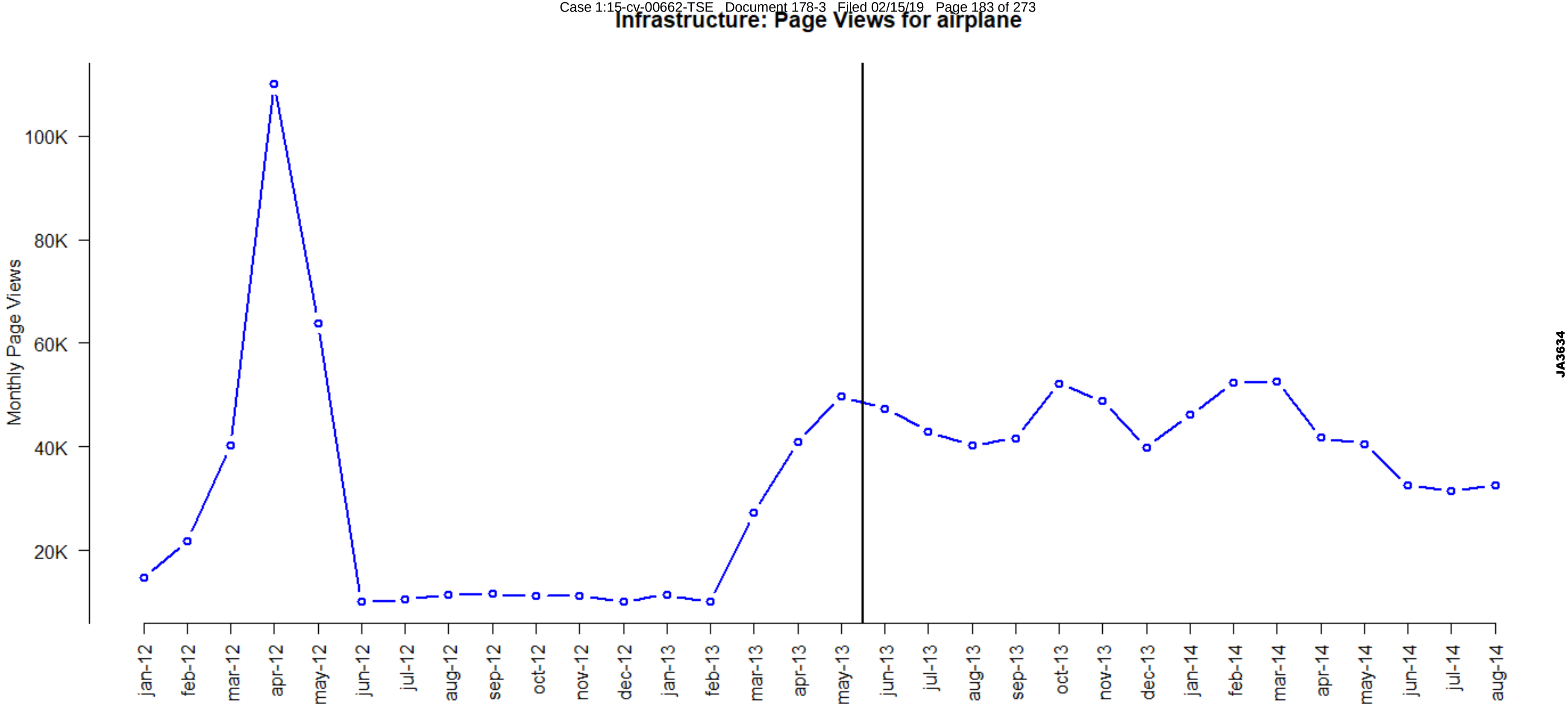
Page Views for Weapons_grade

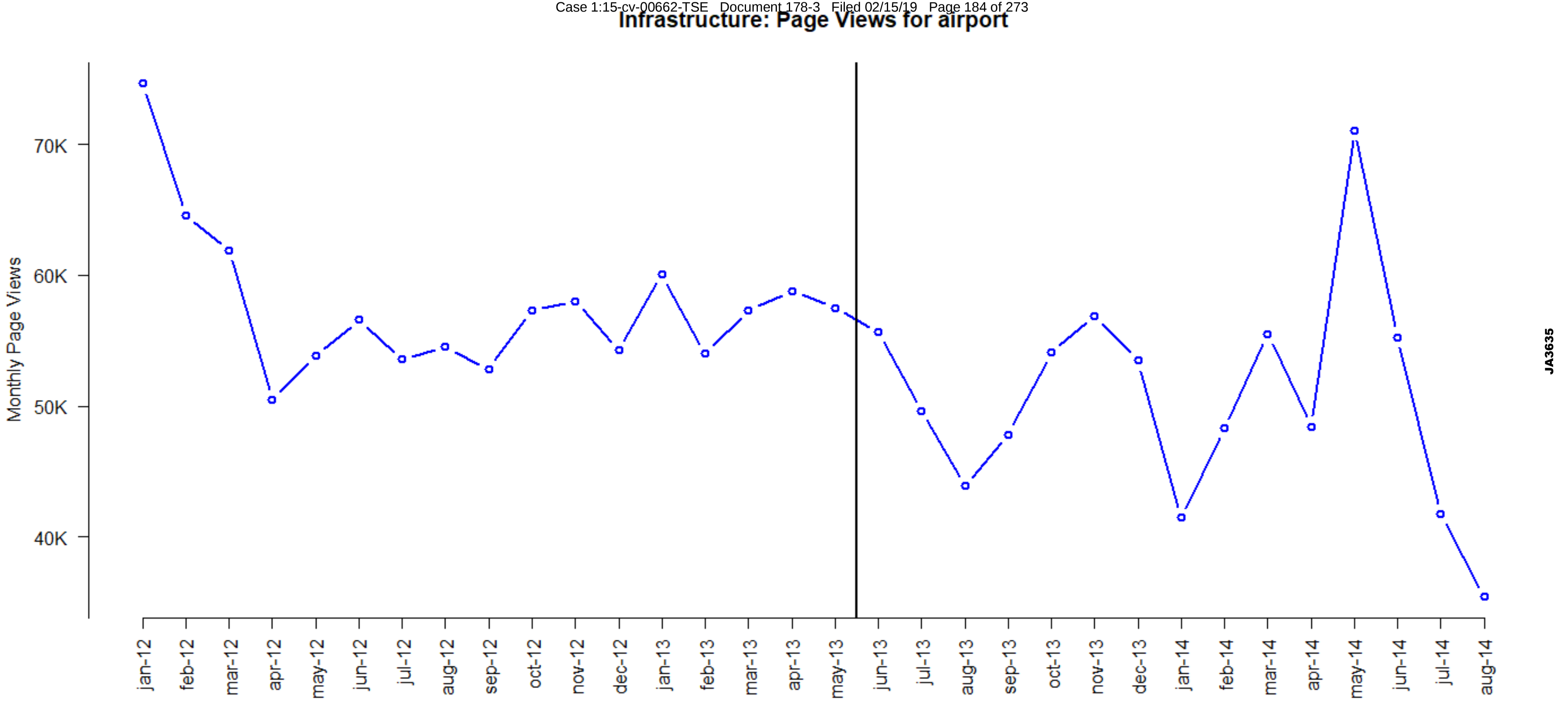


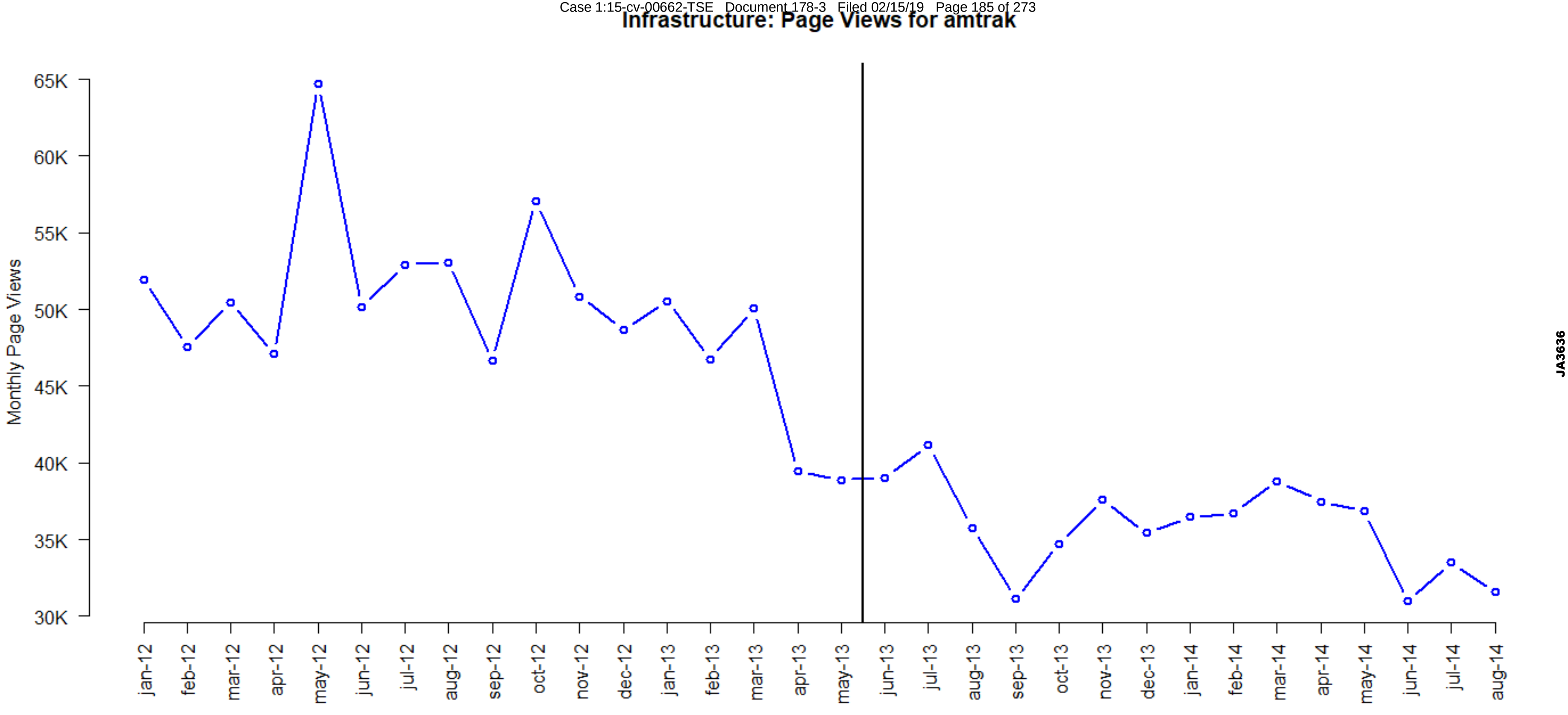
Page Views for Yemen

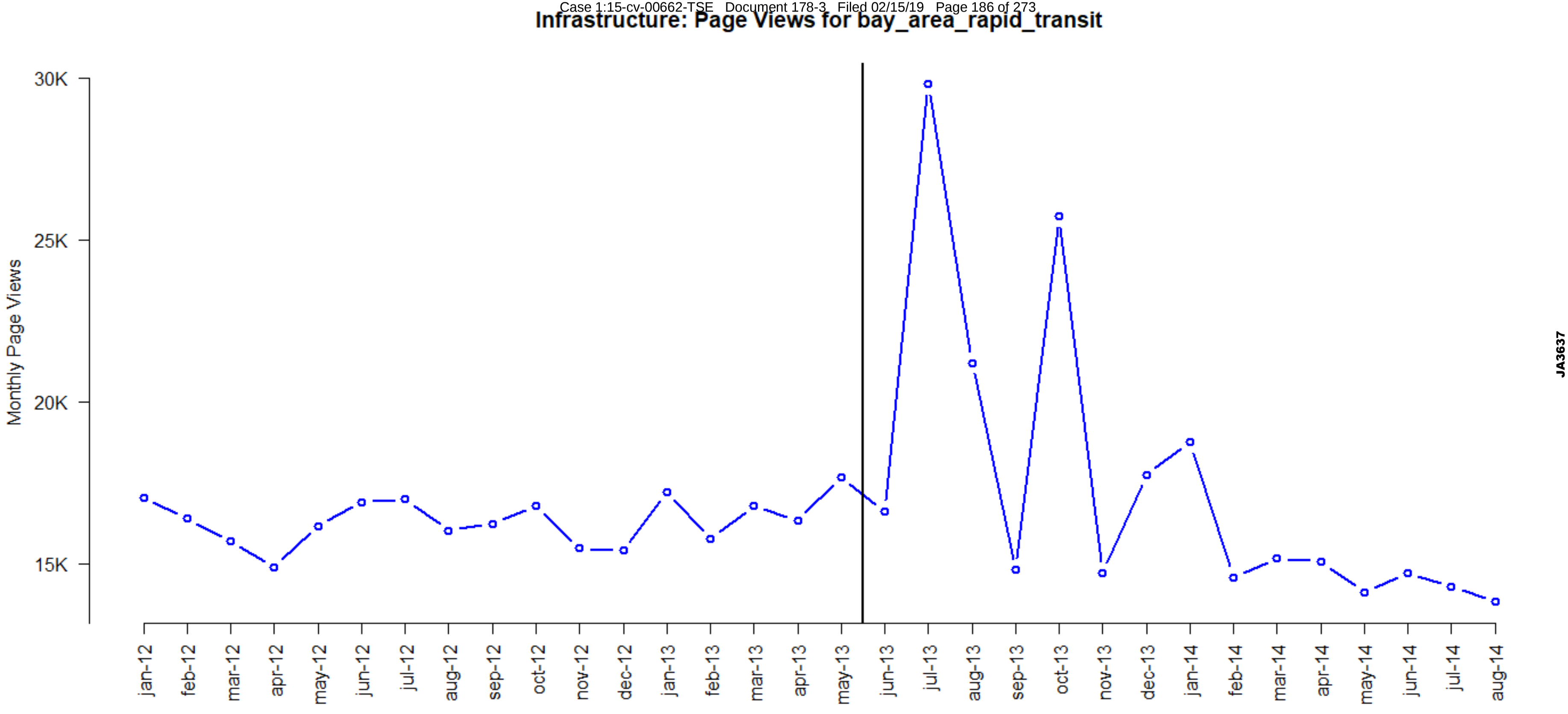


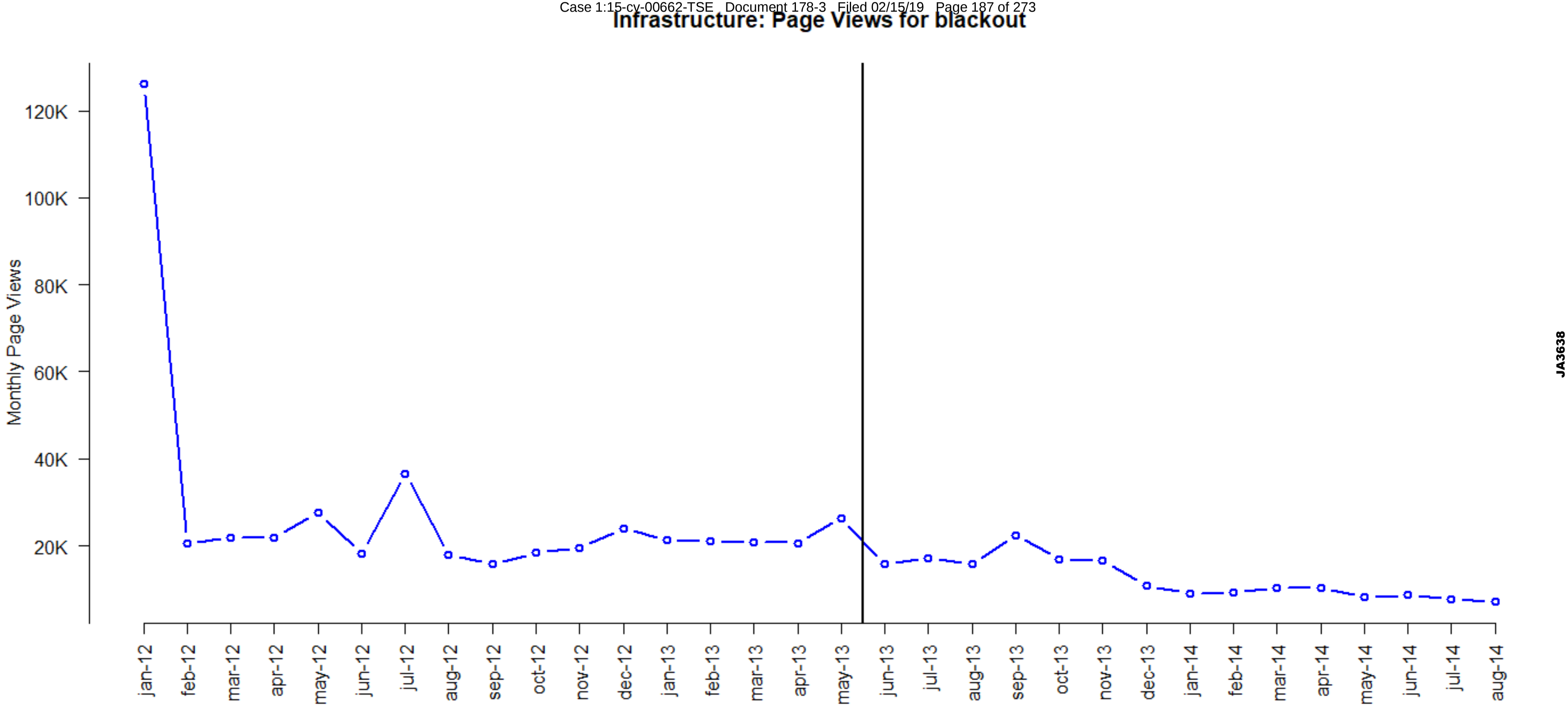
APPENDIX VI: Page Views for 85 Comparative Articles

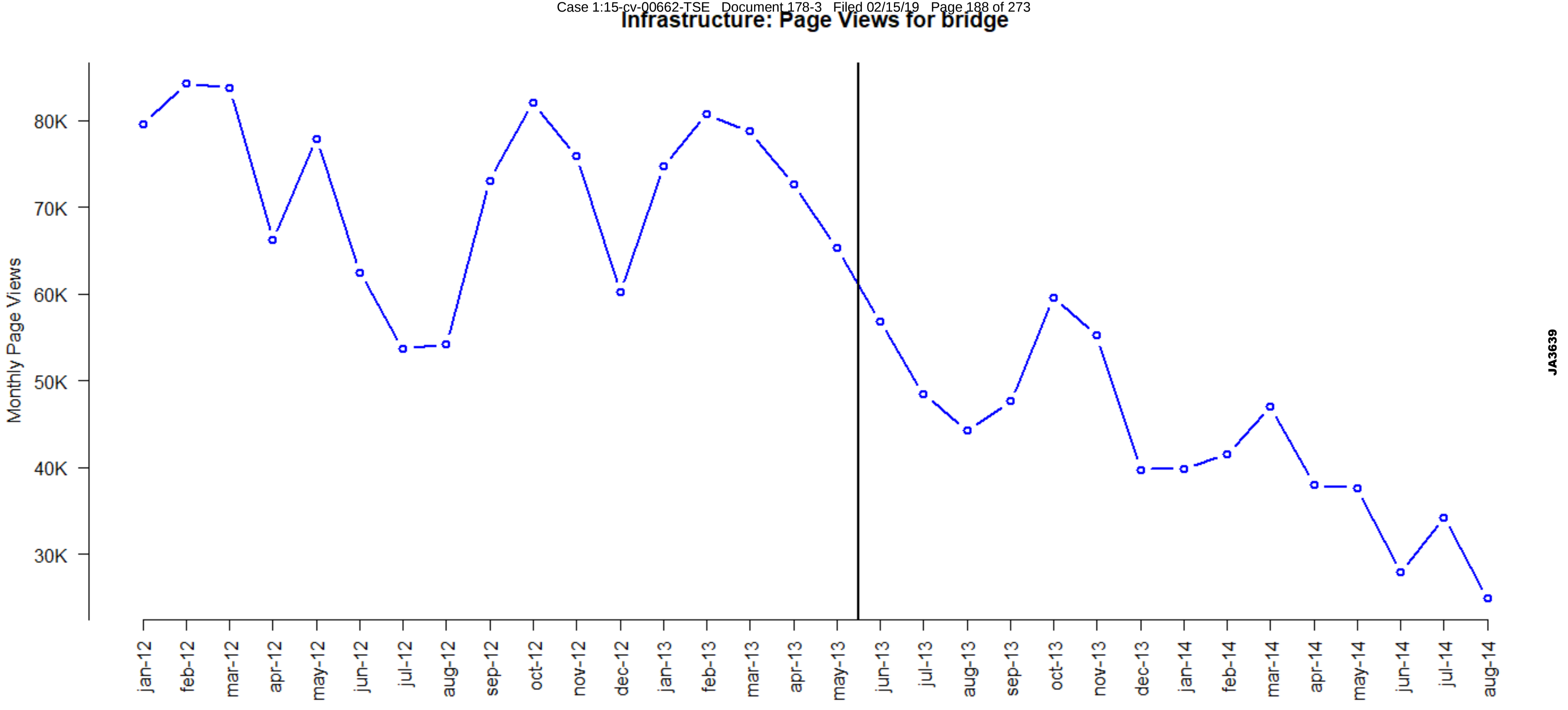

Infrastructure: Page Views for airplane
Monthly Page Views
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

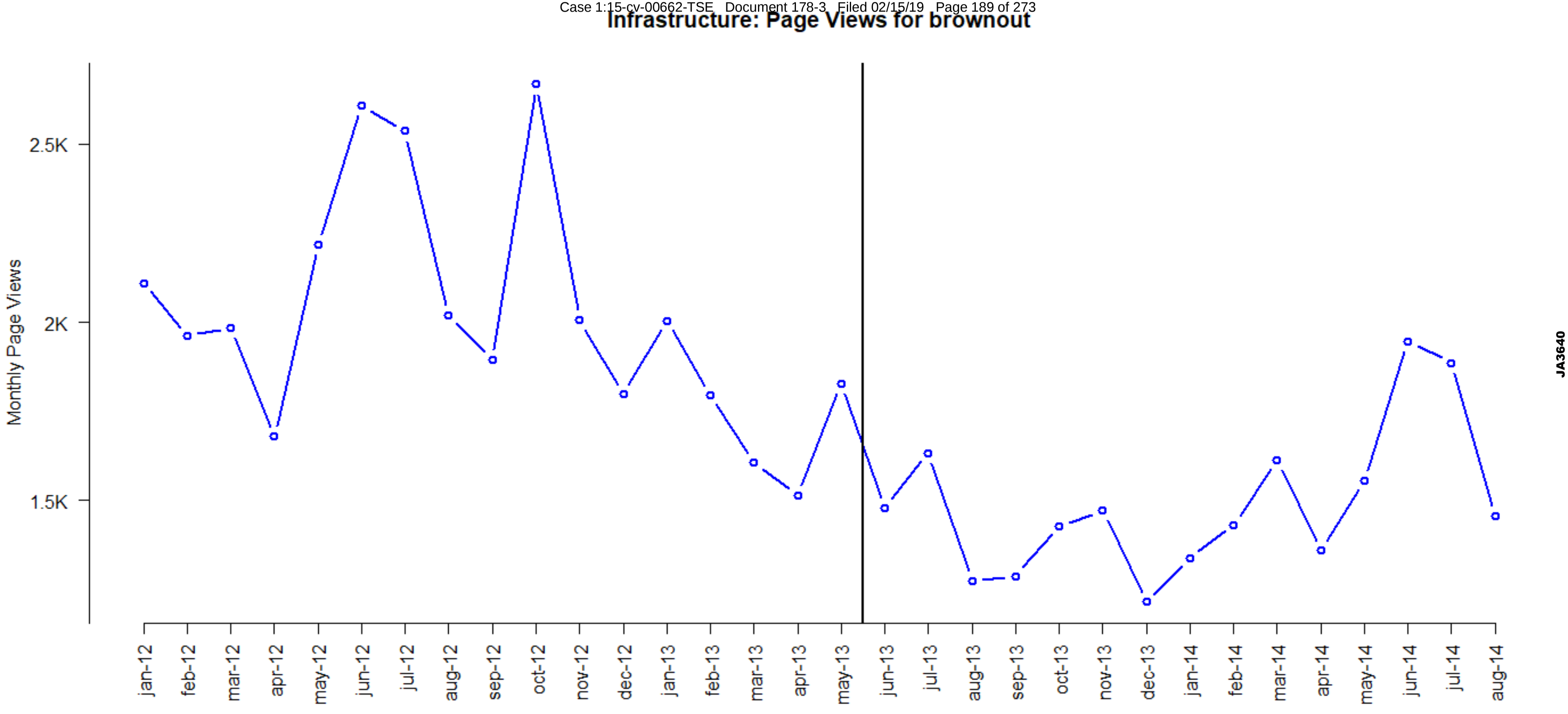

Infrastructure: Page Views for airport
Monthly Page Views
70K
60K
50K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

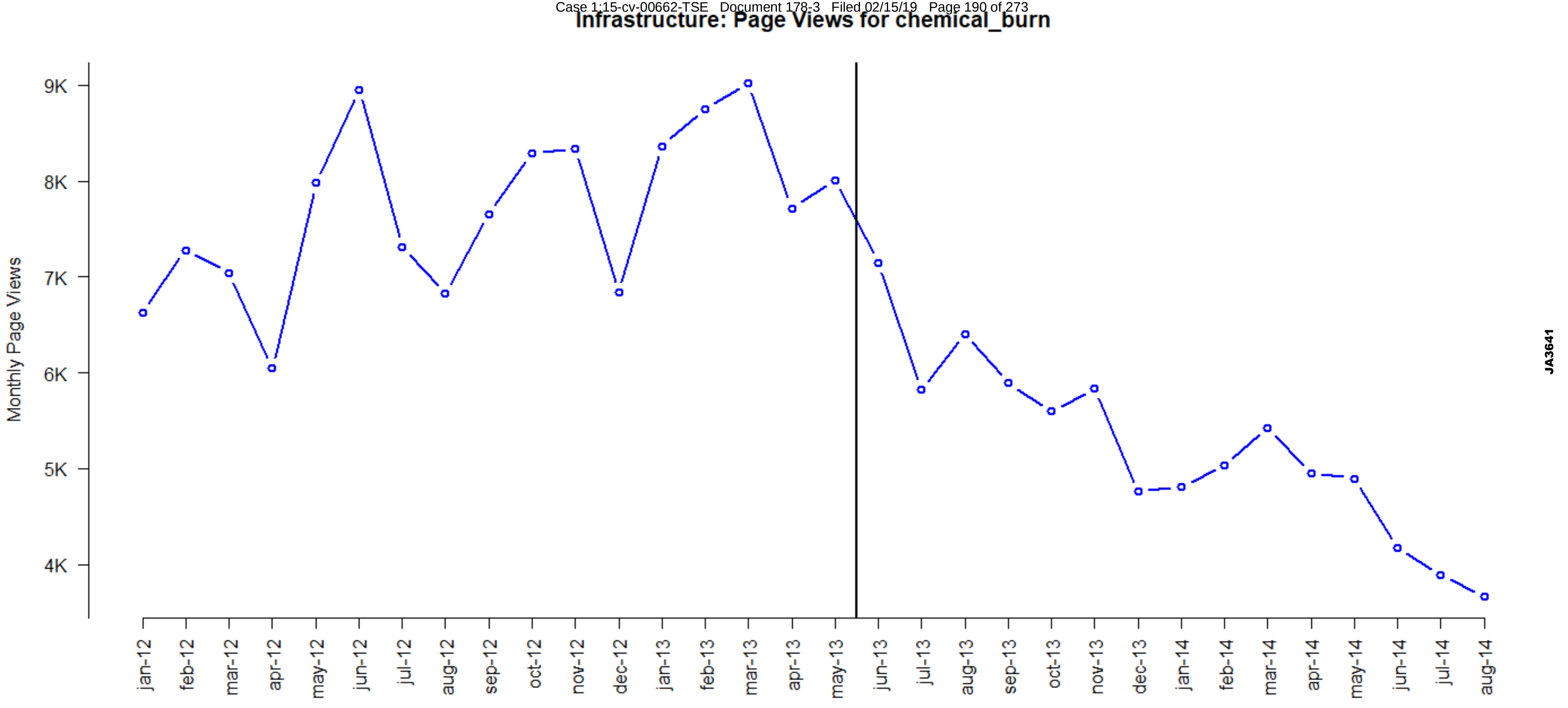

Infrastructure: Page Views for amtrak
Monthly Page Views
65K
60K
55K
50K
45K
40K
35K
30K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

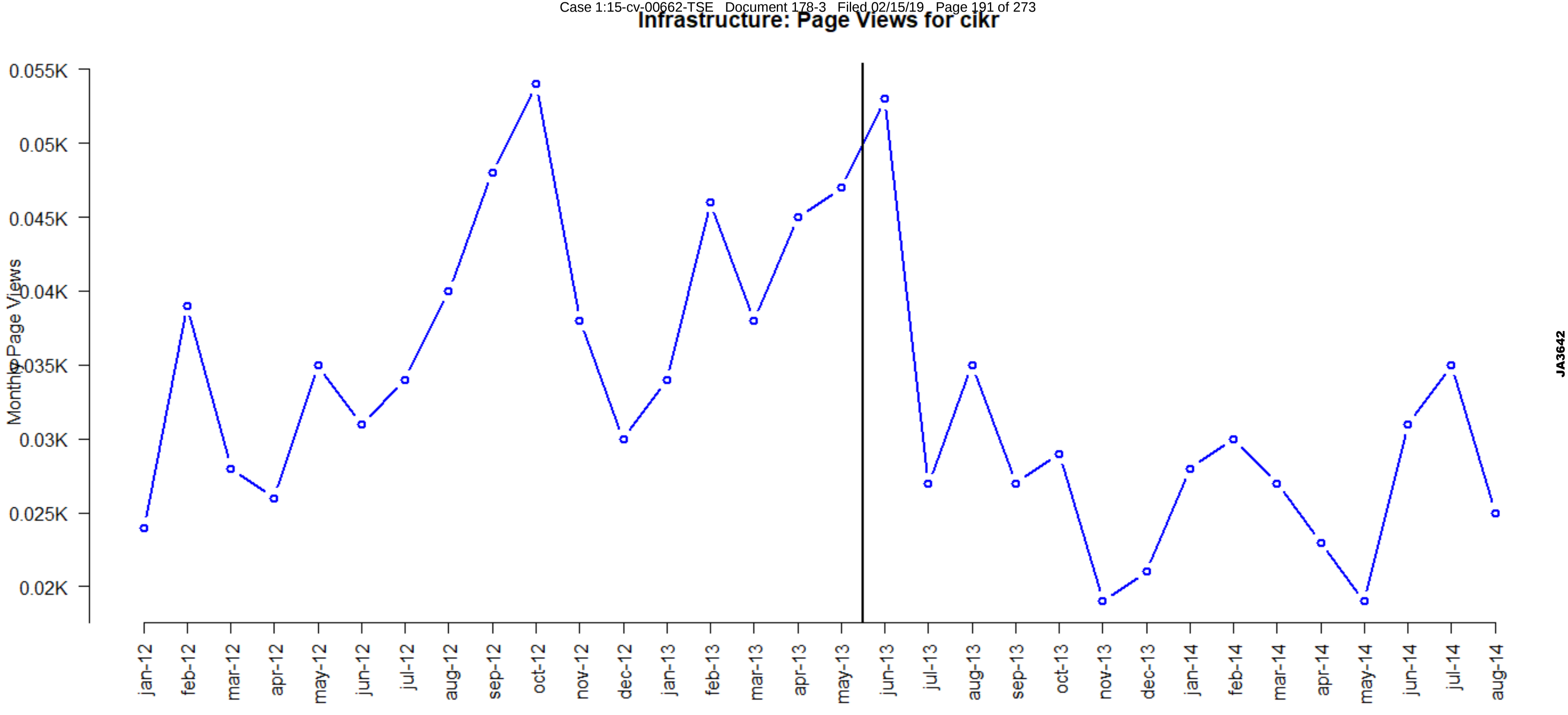

Infrastructure: Page Views for bay_area_rapid_transit
Monthly Page Views
30K
25K
20K
15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

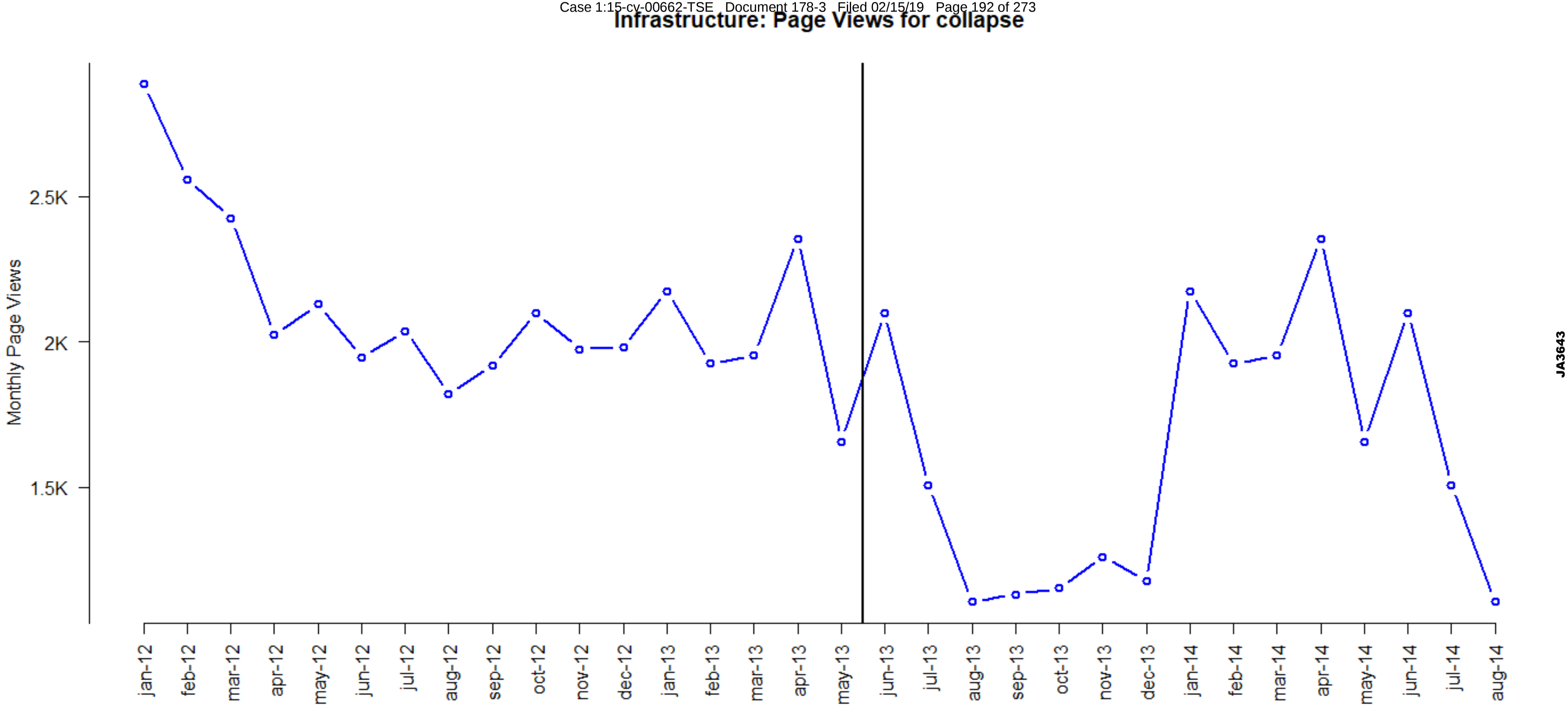

Infrastructure: Page Views for blackout
Monthly Page Views
120K
100K
80K
60K
40K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

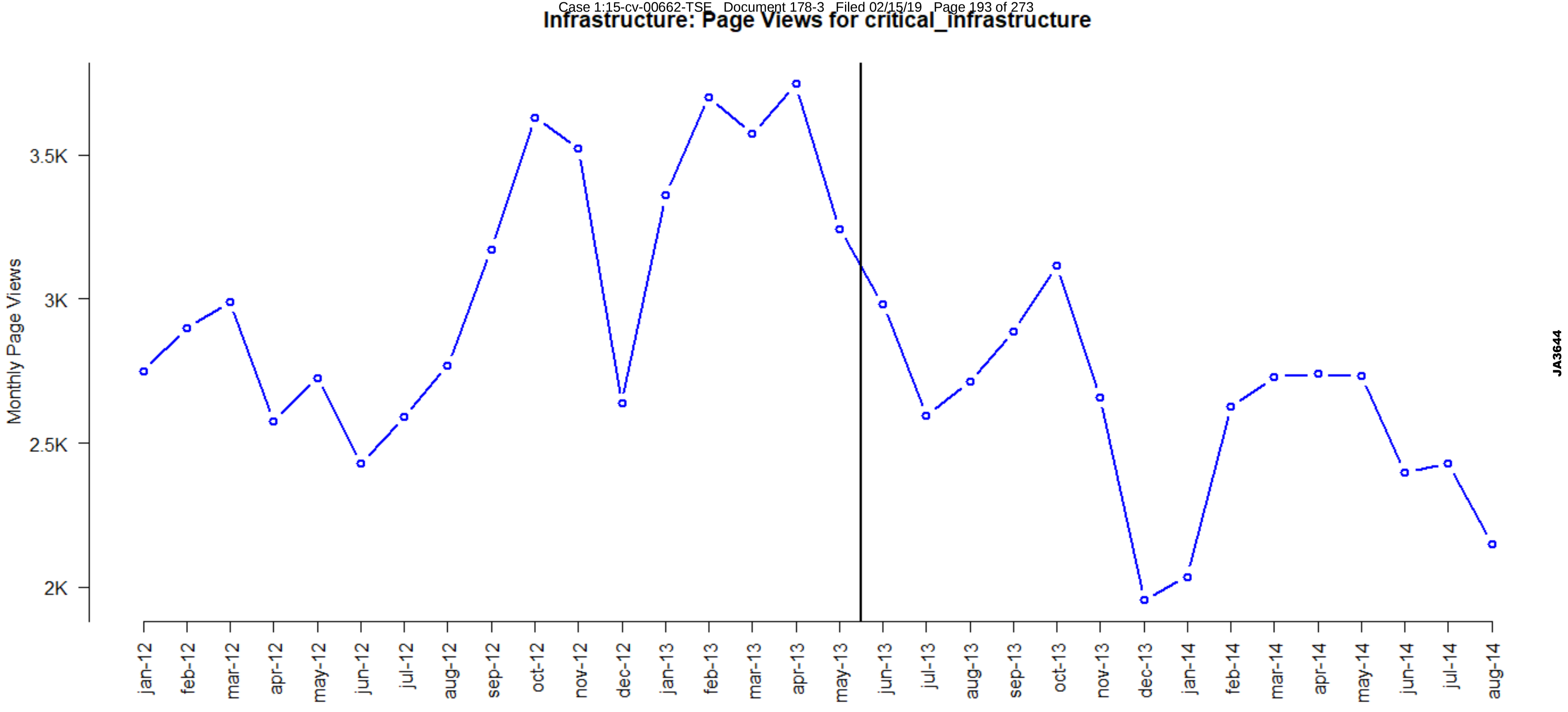

Infrastructure: Page Views for bridge
Monthly Page Views
80K
70K
60K
50K
40K
30K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

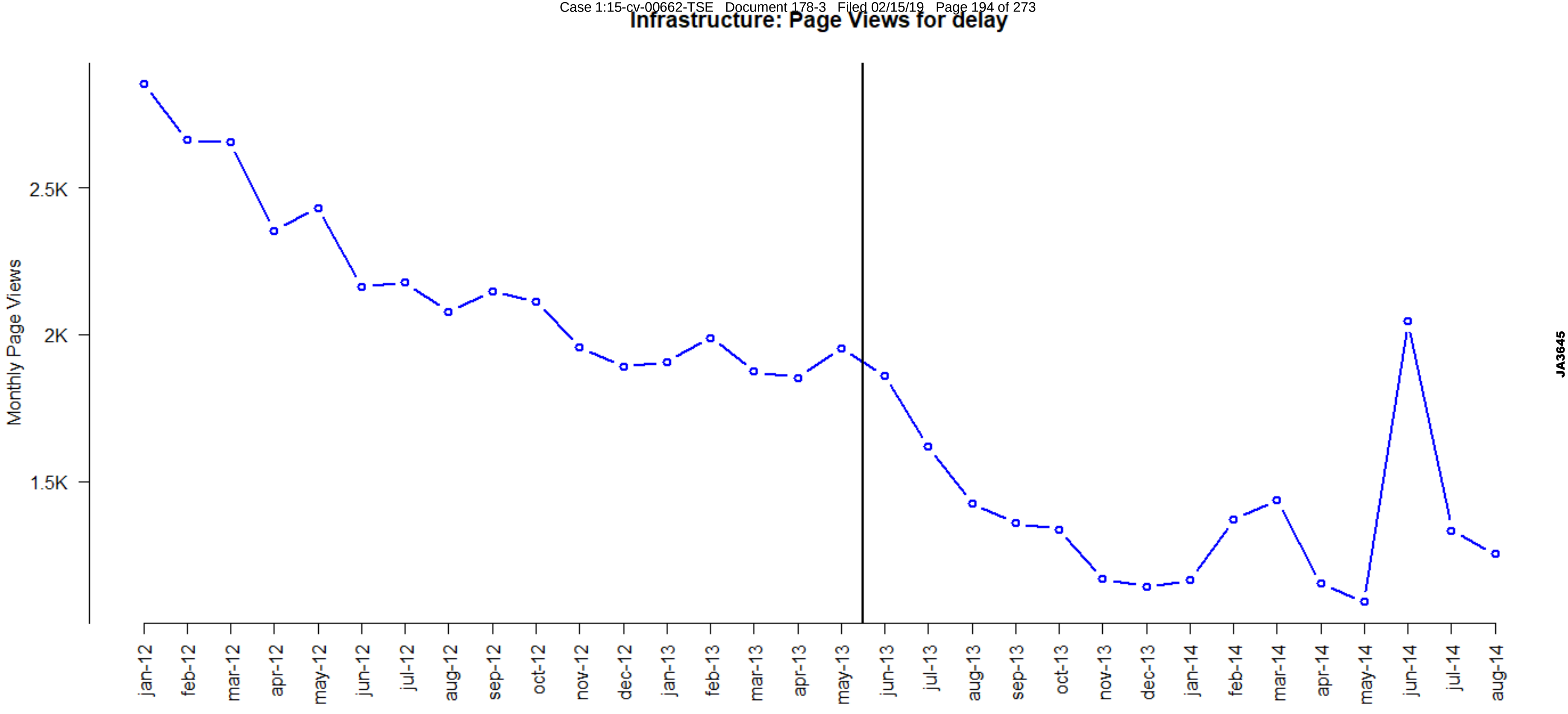

Infrastructure: Page Views for brownout
Monthly Page Views
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

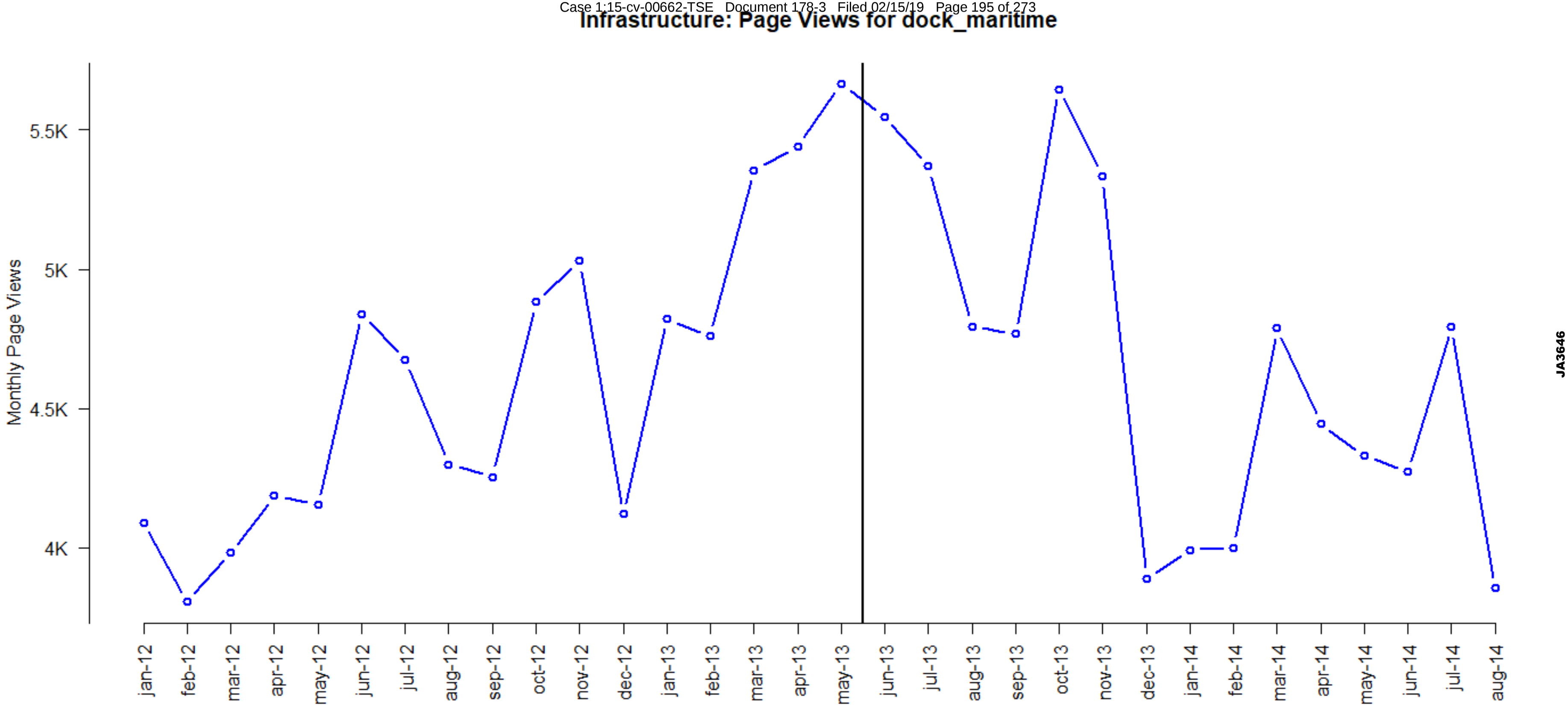

Infrastructure: Page Views for chemical_burn
Monthly Page Views
9K
8K
7K
6K
5K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

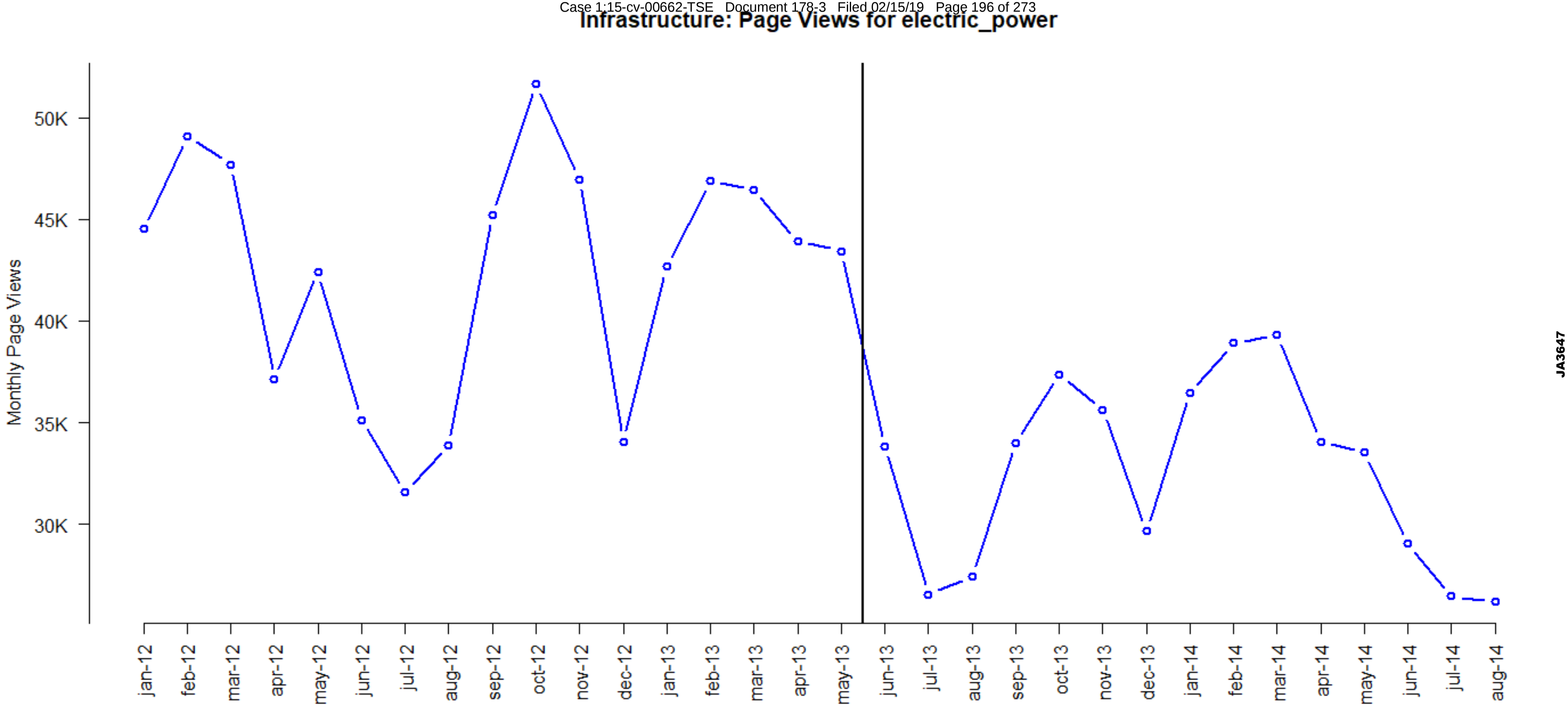

Infrastructure: Page Views for cikr
Monthly Page Views
0.055K
0.05K
0.045K
0.04K
0.035K
0.03K
0.025K
0.02K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

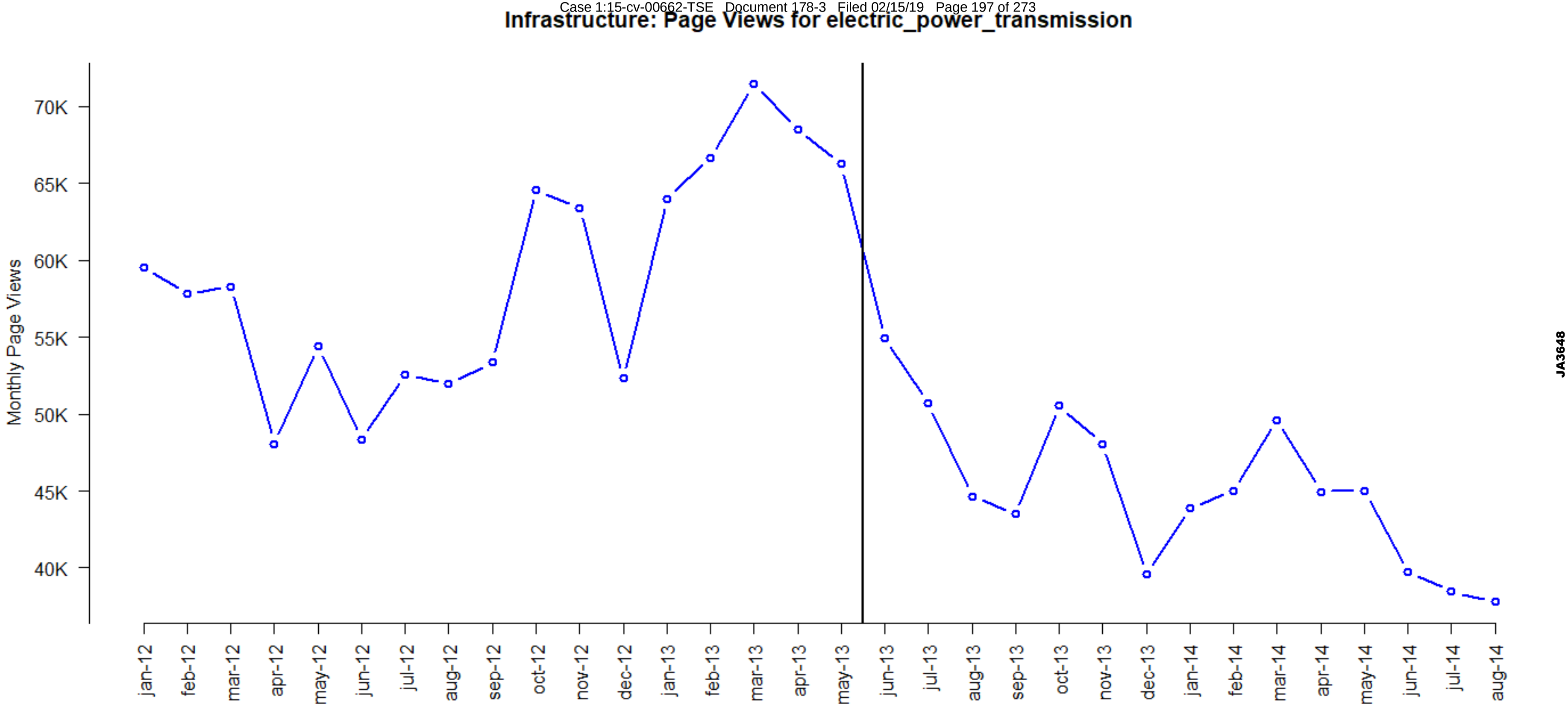

Infrastructure: Page Views for collapse
Monthly Page Views
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

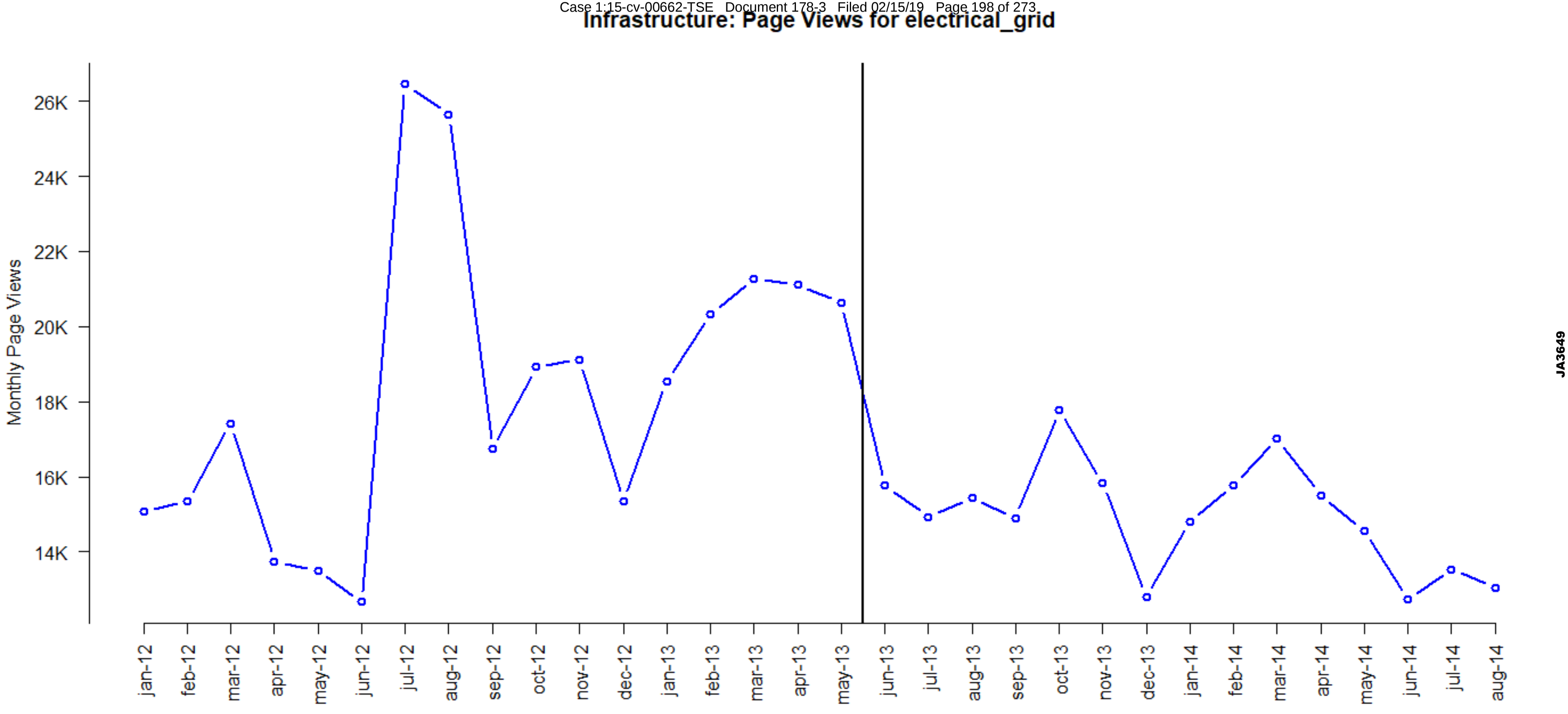

Infrastructure: Page Views for critical_infrastructure
Monthly Page Views
3.5K
3K
2.5K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

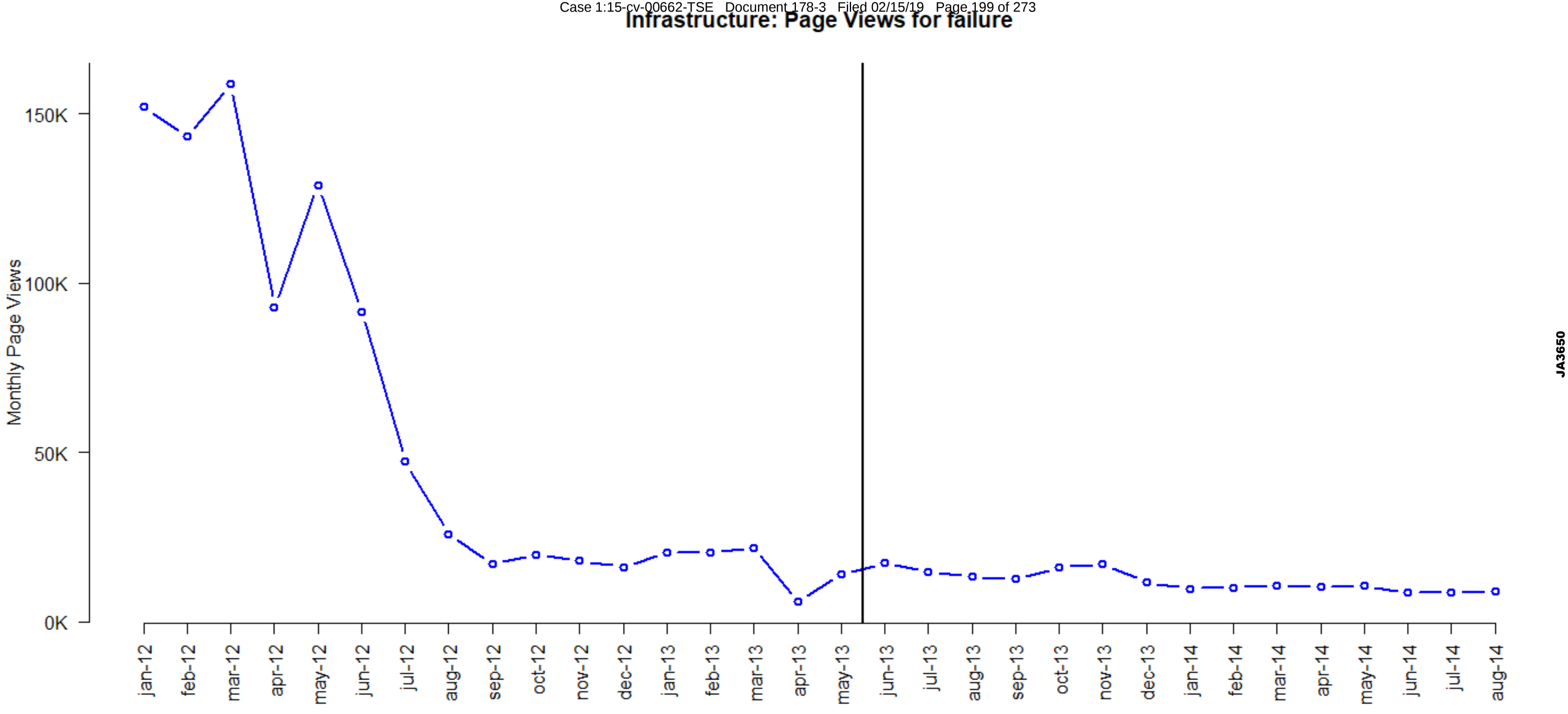

Infrastructure: Page Views for delay
Monthly Page Views
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

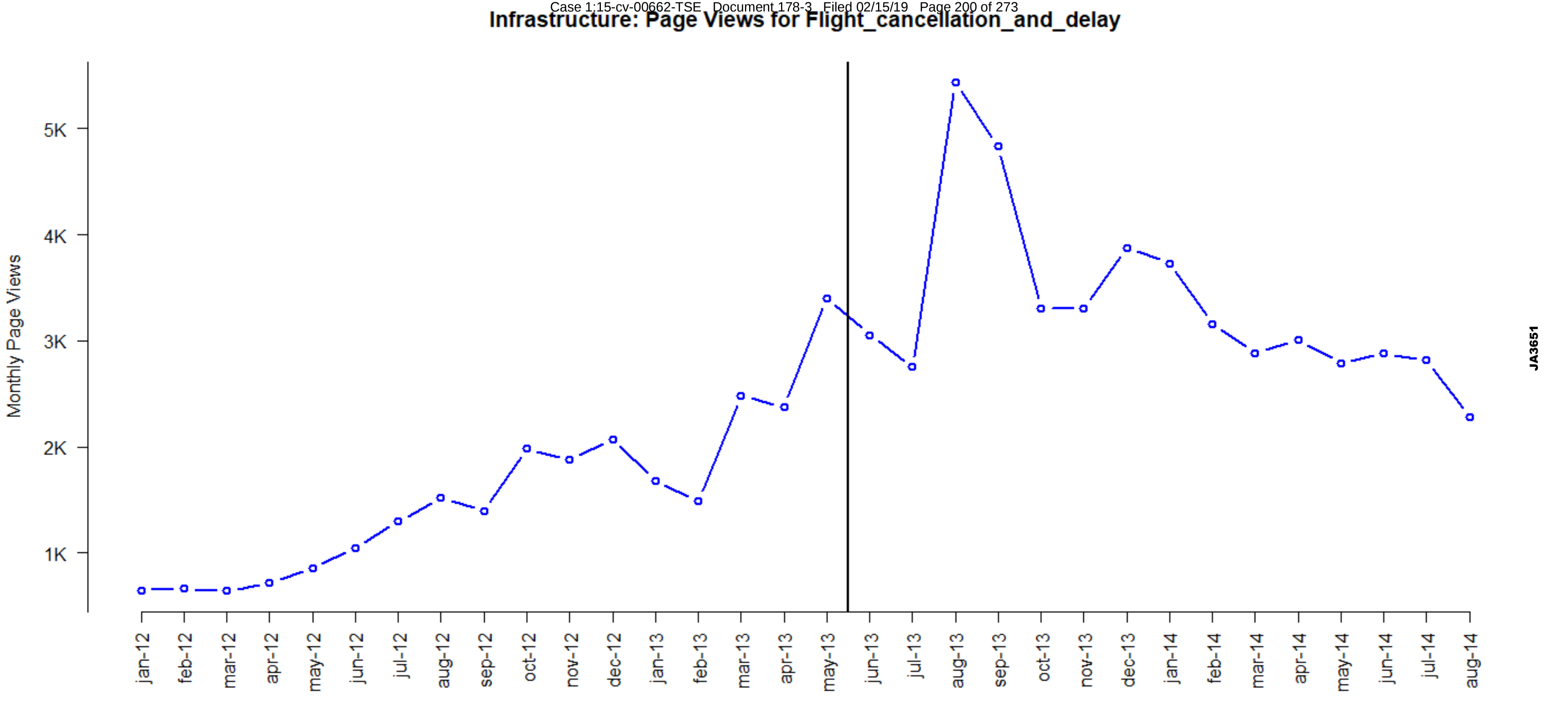

Infrastructure: Page Views for dock_maritime
Monthly Page Views
5.5K
5K
4.5K
4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

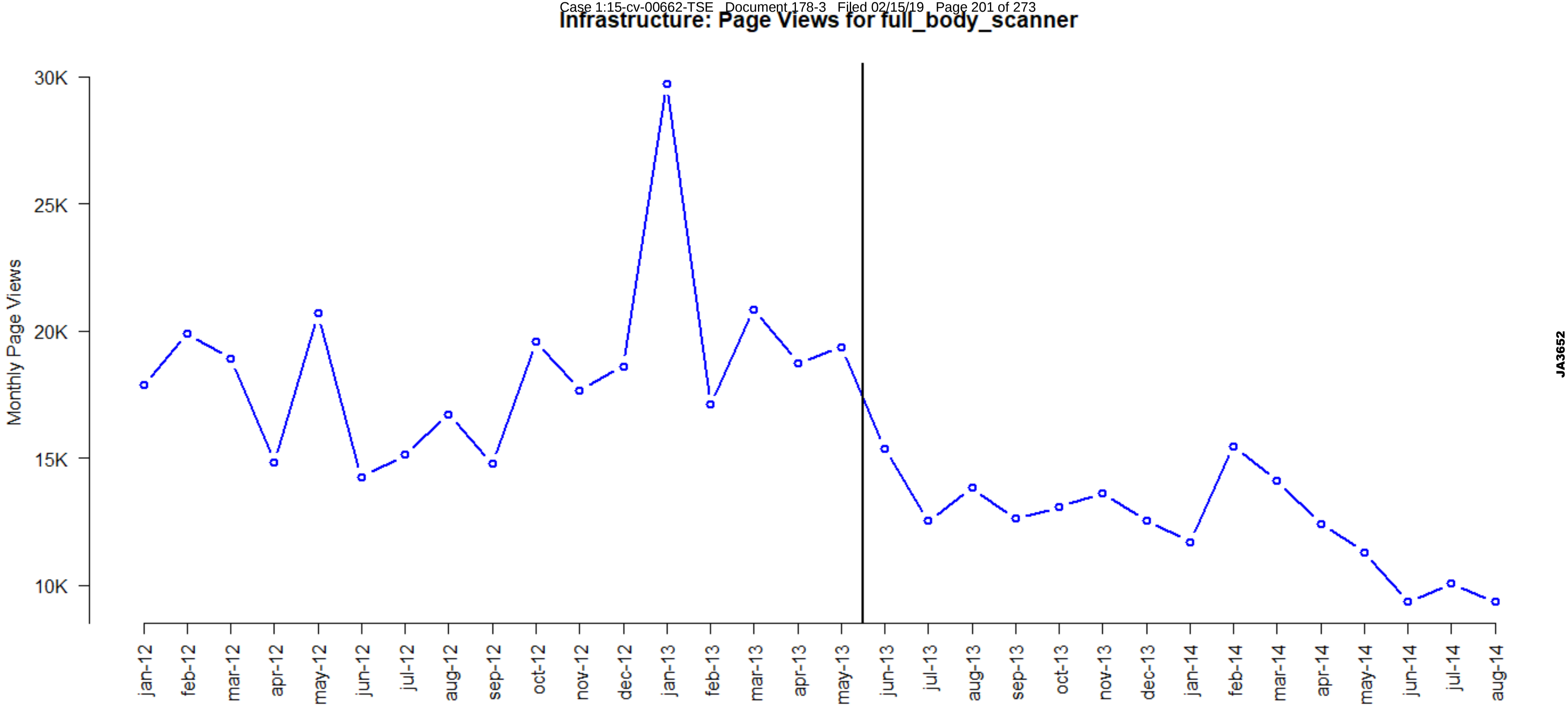

Infrastructure: Page Views for electric_power
Monthly Page Views
50K
45K
40K
35K
30K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

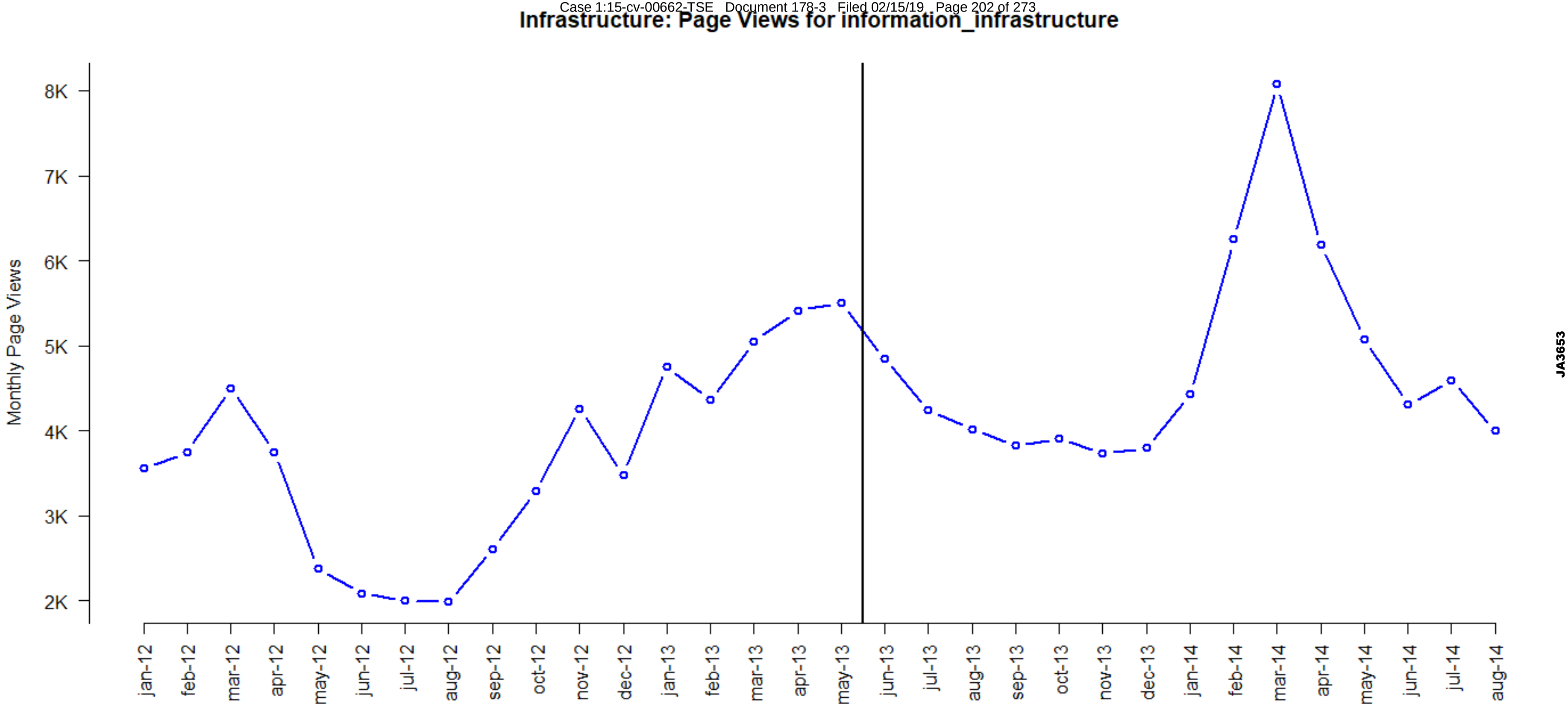

Infrastructure: Page Views for electric_power_transmission
Monthly Page Views
70K
65K
60K
55K
50K
45K
40K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

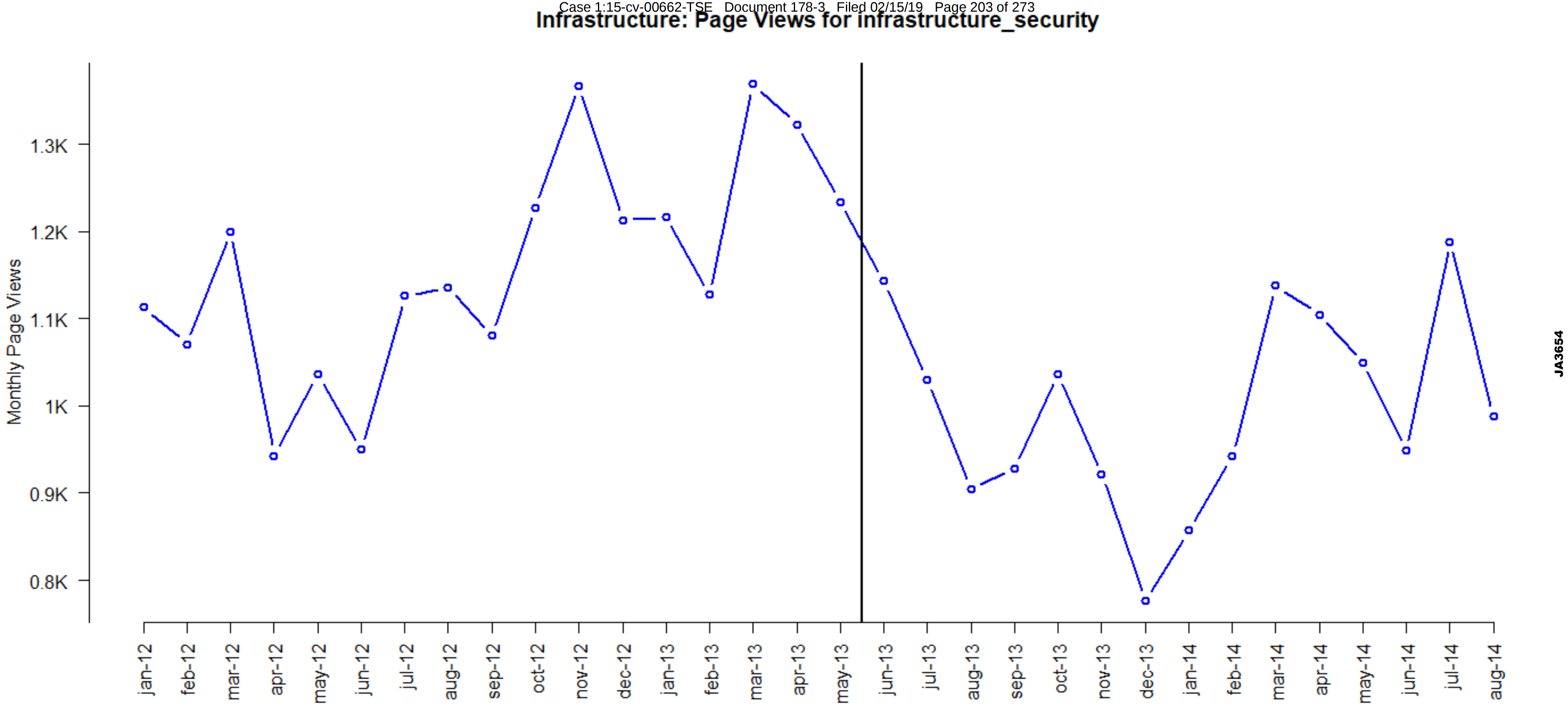

Infrastructure: Page Views for electrical_grid
Monthly Page Views
26K
24K
22K
20K
18K
16K
14K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

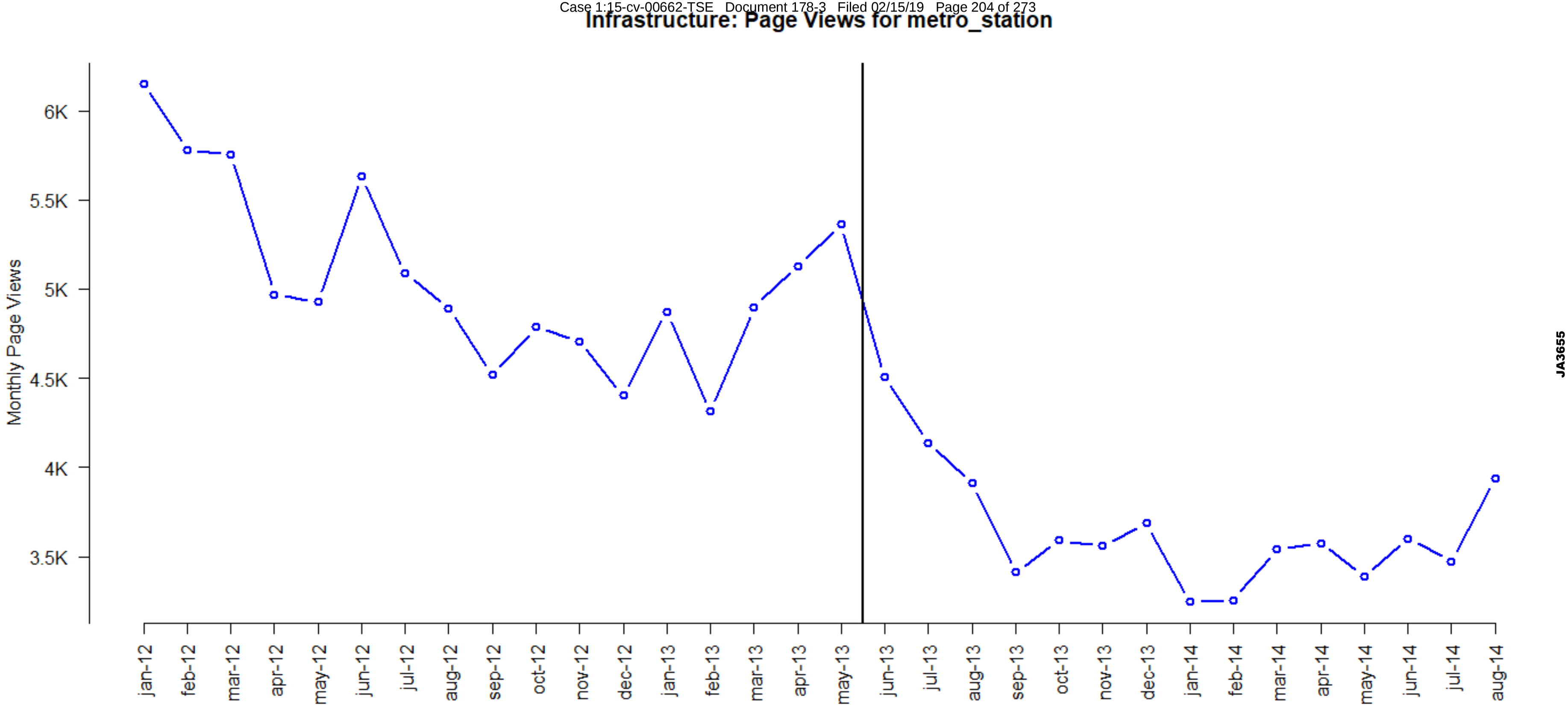

Infrastructure: Page Views for failure
Monthly Page Views
150K
100K
50K
0K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

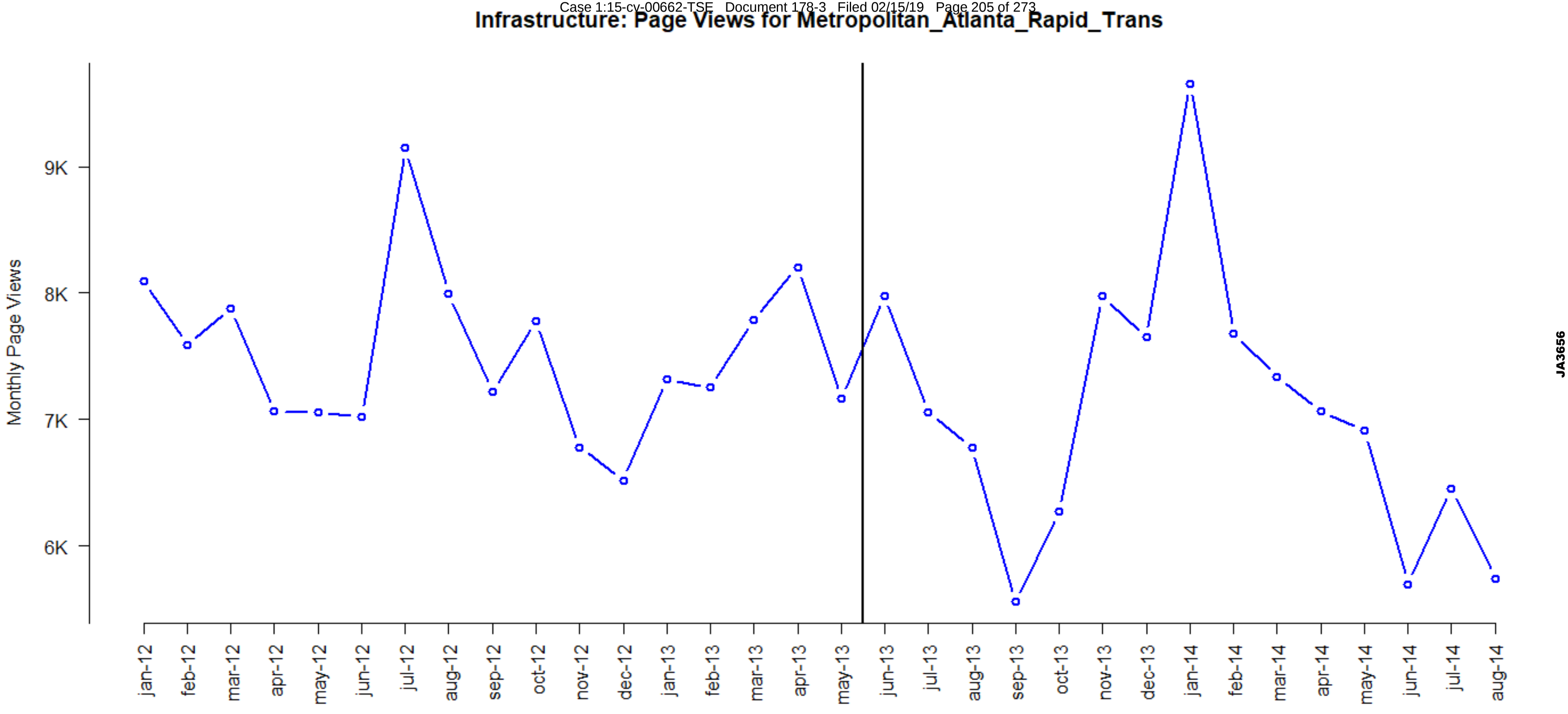

Infrastructure: Page Views for Flight_cancellation_and_delay
Monthly Page Views
5K
4K
3K
2K
1K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

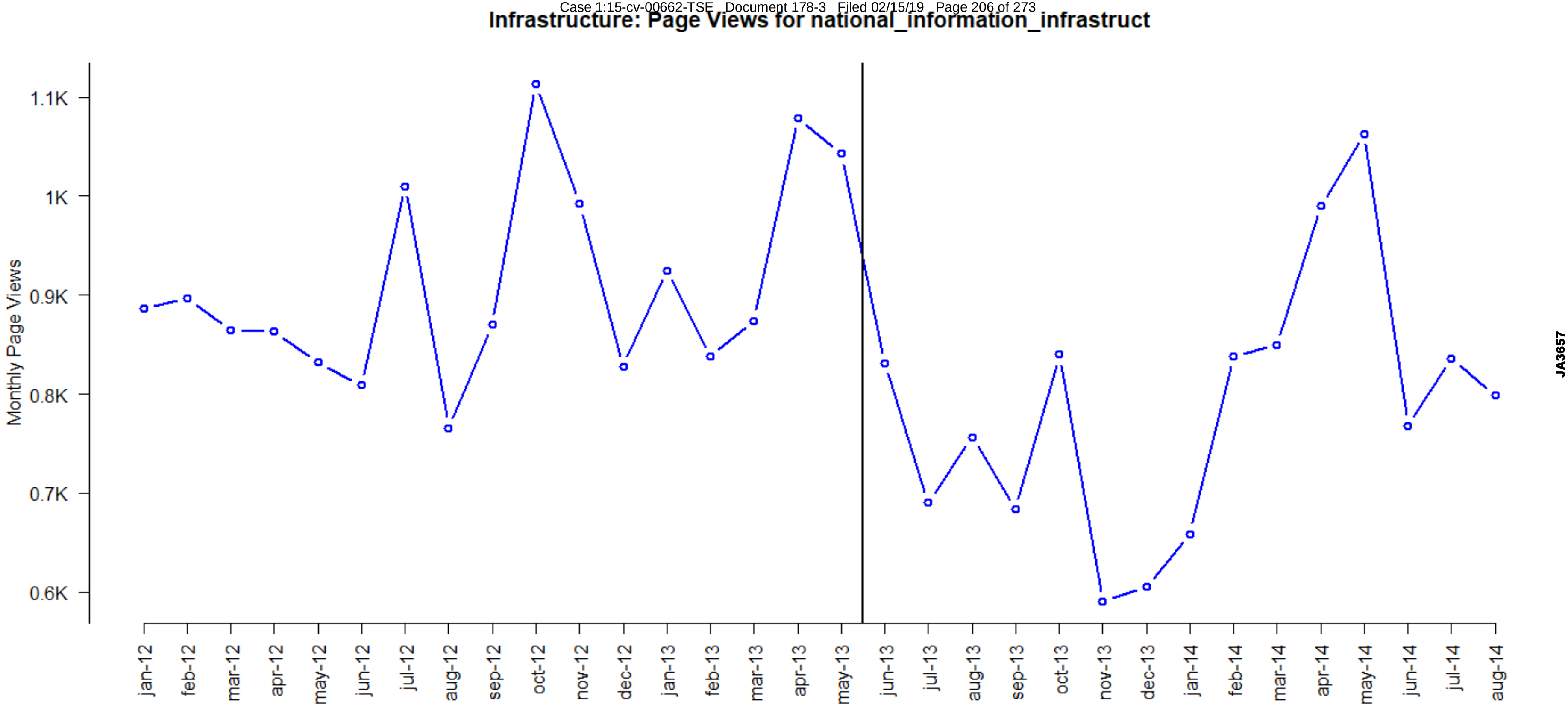

Infrastructure: Page Views for full_body_scanner
Monthly Page Views
30K
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

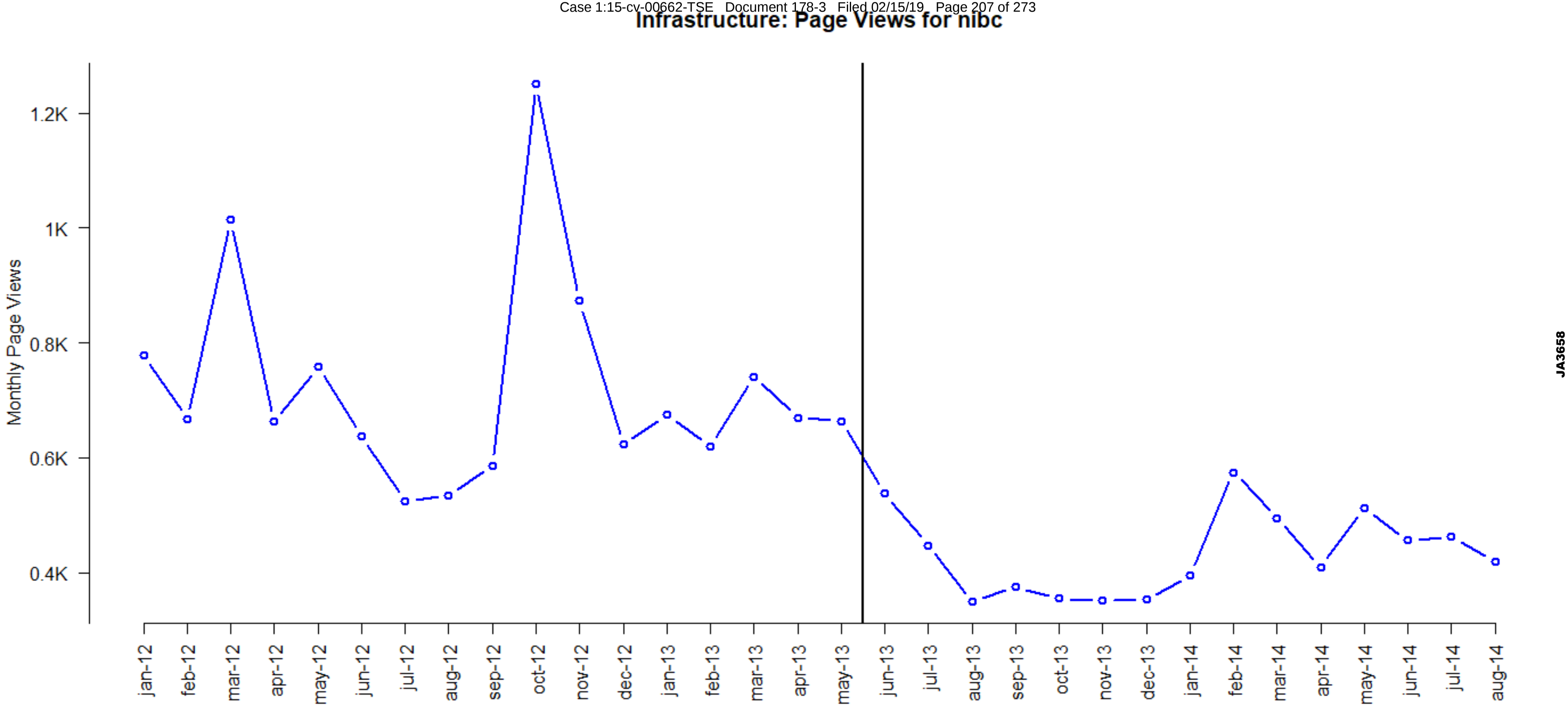

Infrastructure: Page Views for information_infrastructure
Monthly Page Views
8K
7K
6K
5K
4K
3K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

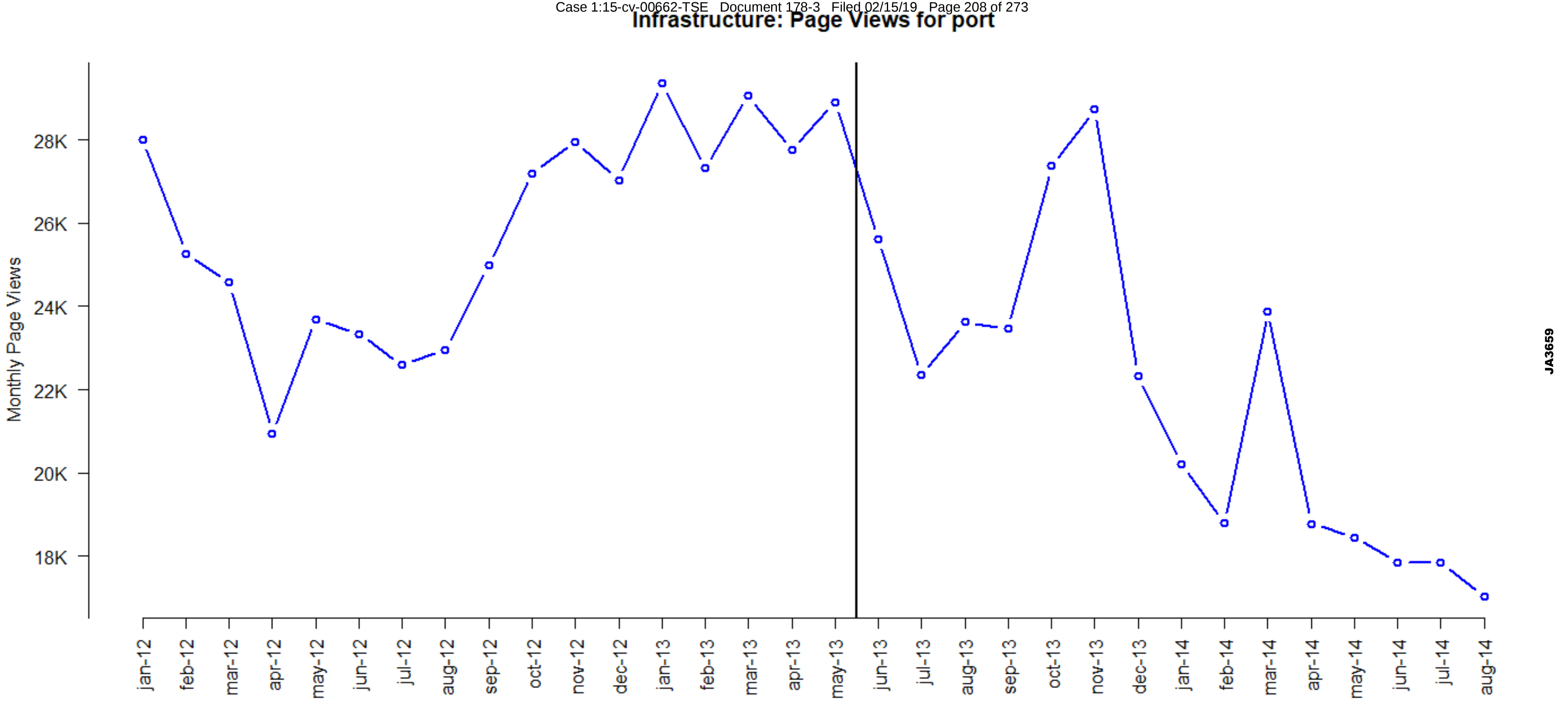

Infrastructure: Page Views for infrastructure_security
Monthly Page Views
1.3K
1.2K
1.1K
1K
0.9K
0.8K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

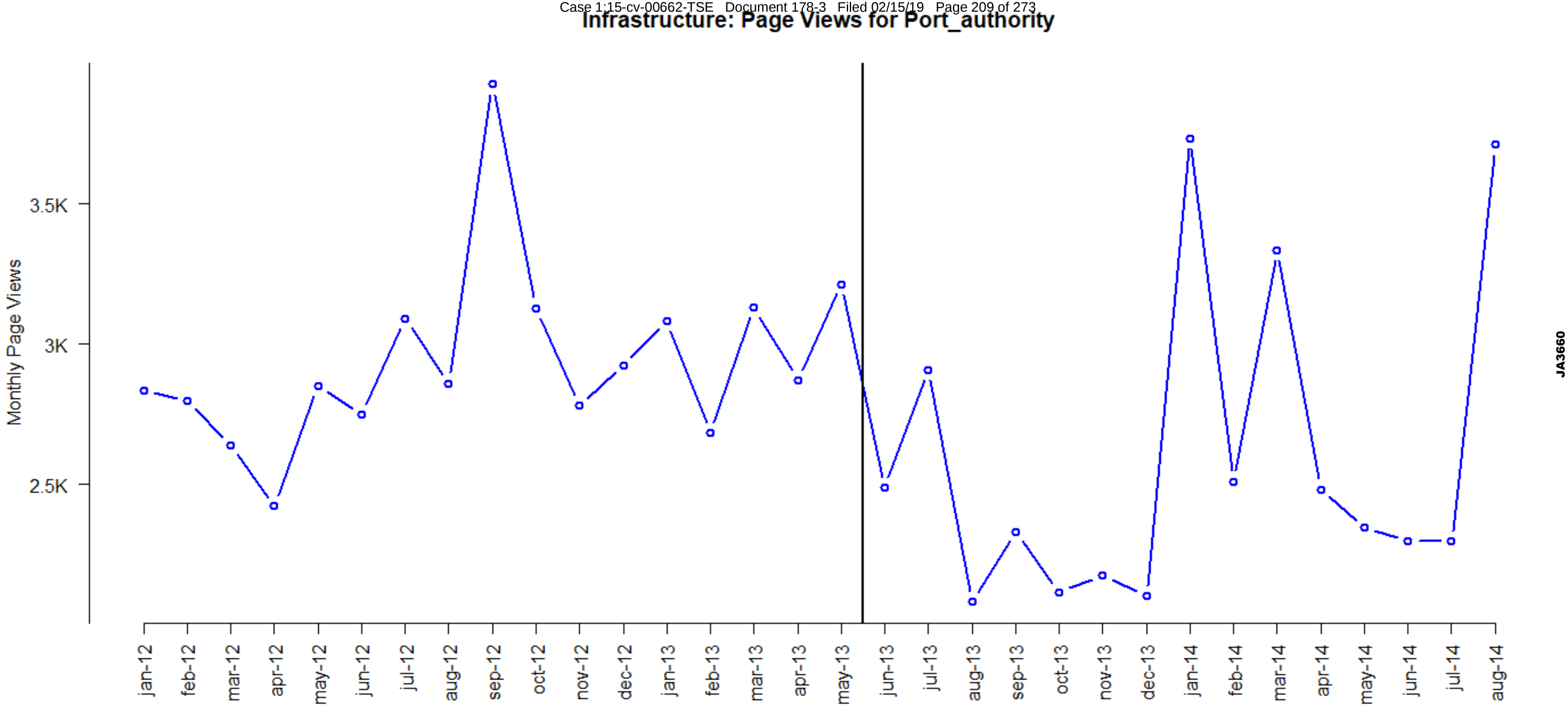

Infrastructure: Page Views for metro_station
Monthly Page Views
6K
5.5K
5K
4.5K
4K
3.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

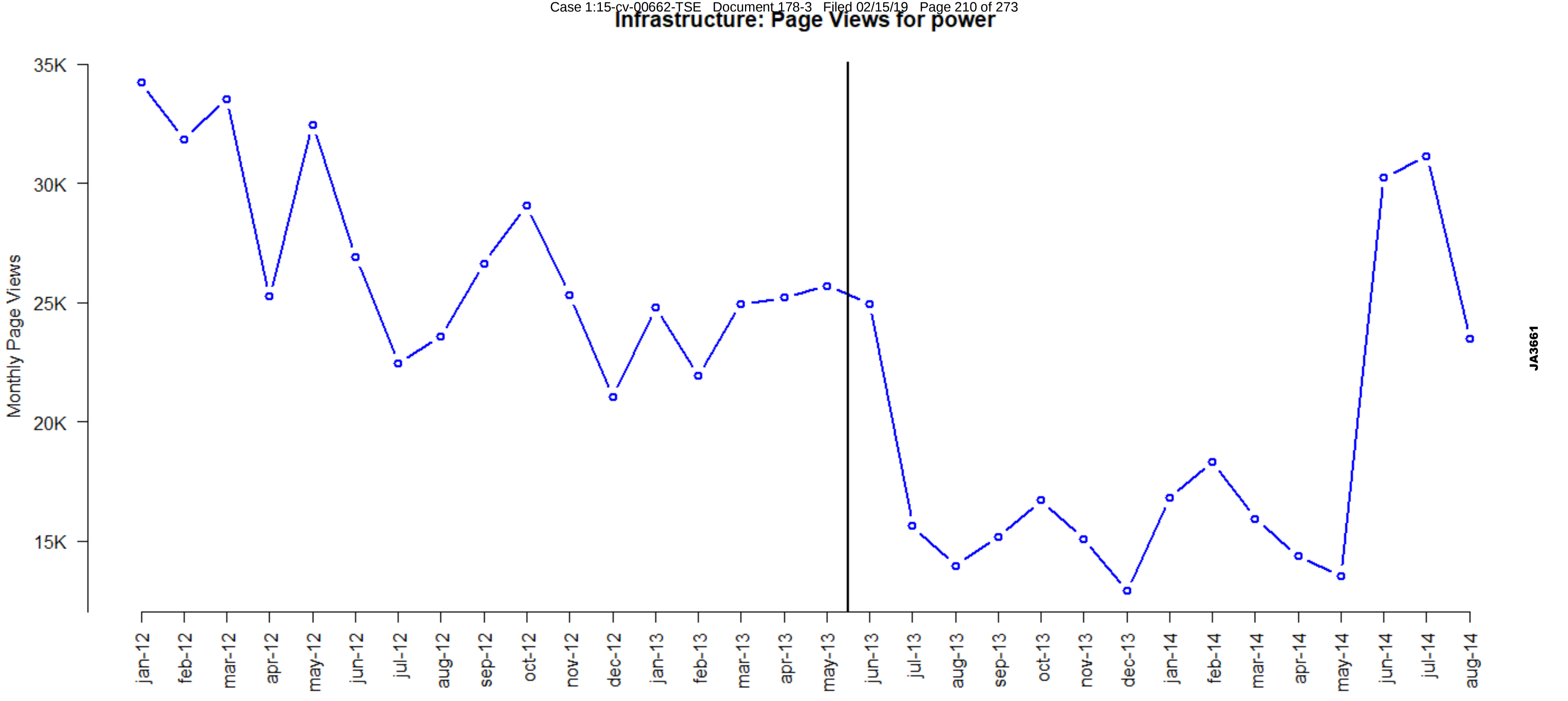

Infrastructure: Page Views for Metropolitan_Atlanta_Rapid_Trans
Monthly Page Views
9K
8K
7K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

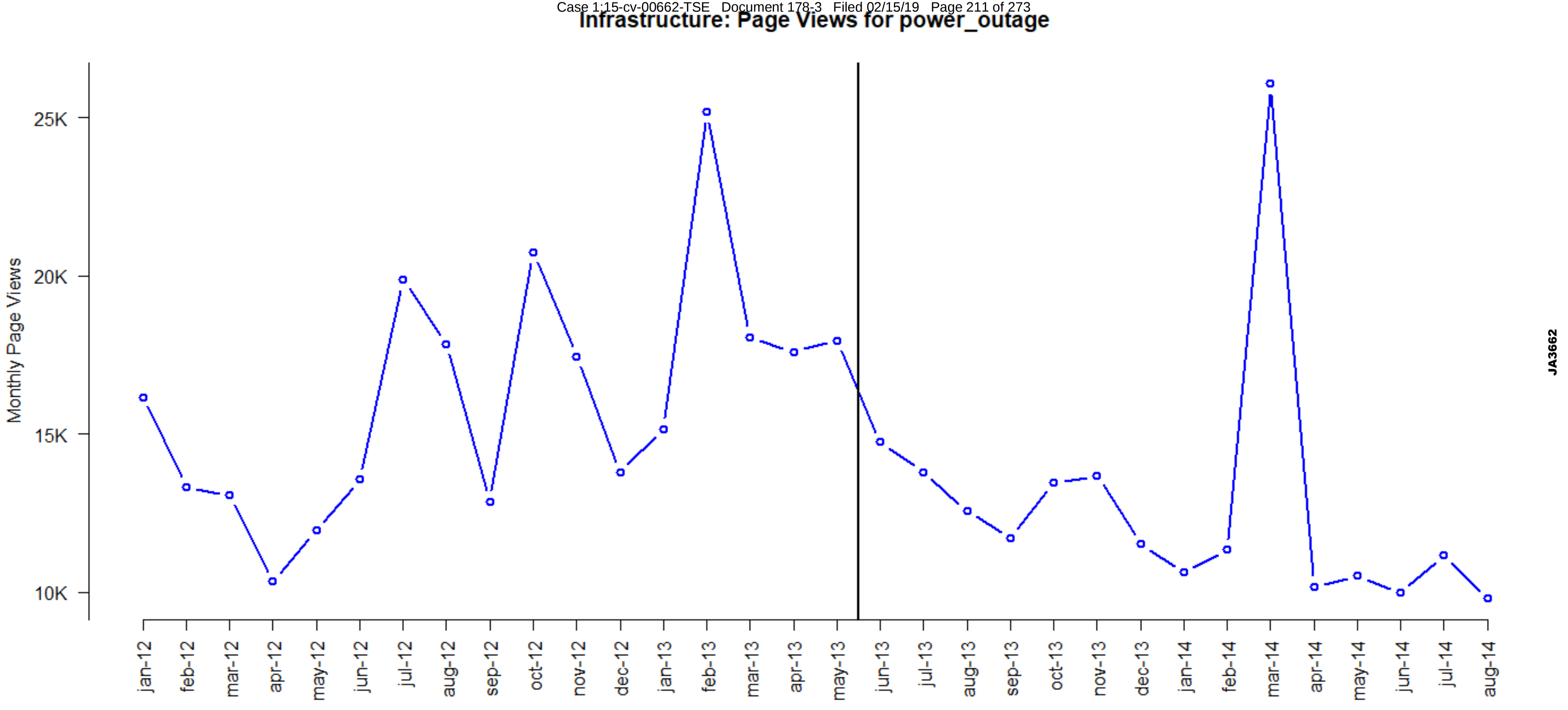

Infrastructure: Page Views for national_information_infrastruct
Monthly Page Views
1.1K
1K
0.9K
0.8K
0.7K
0.6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

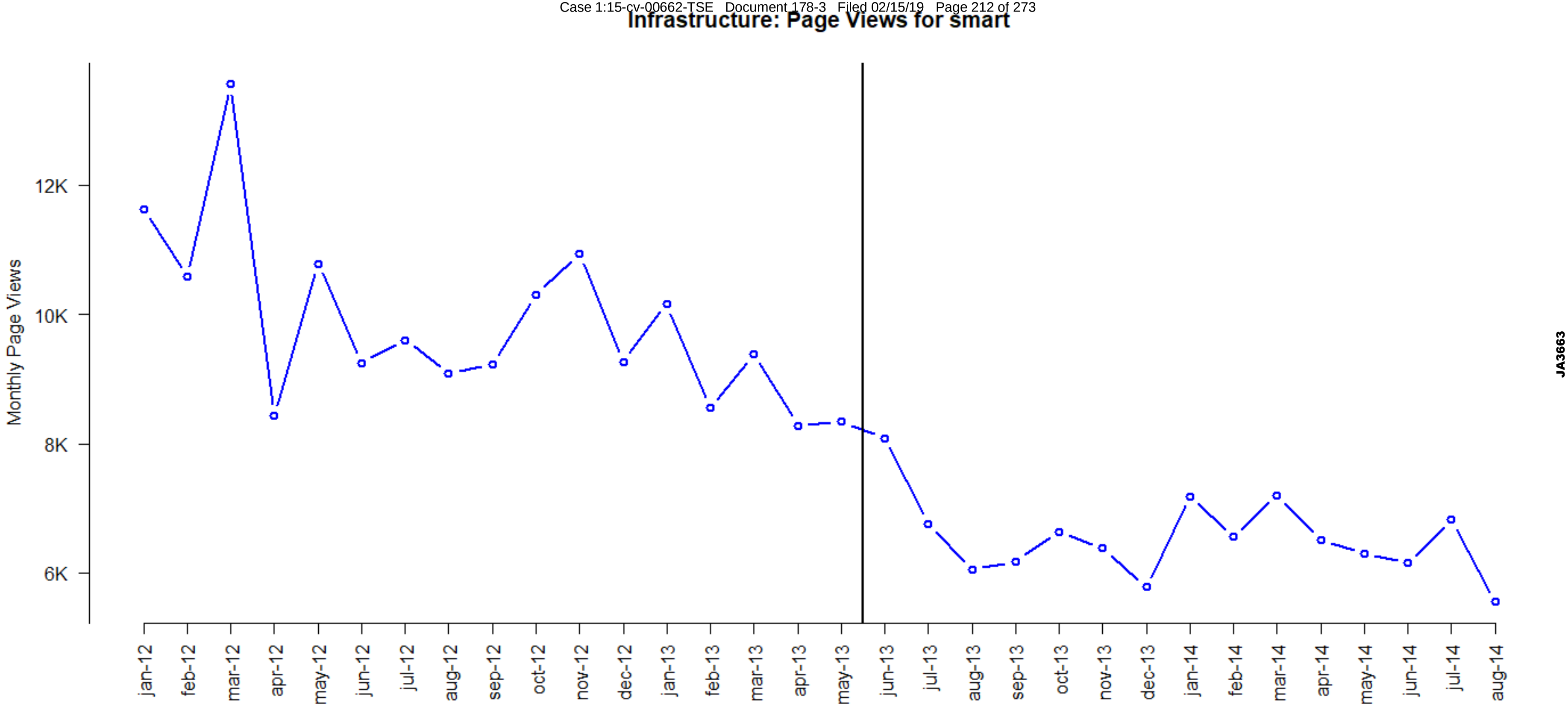

Infrastructure: Page Views for nibc
Monthly Page Views
1.2K
1K
0.8K
0.6K
0.4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

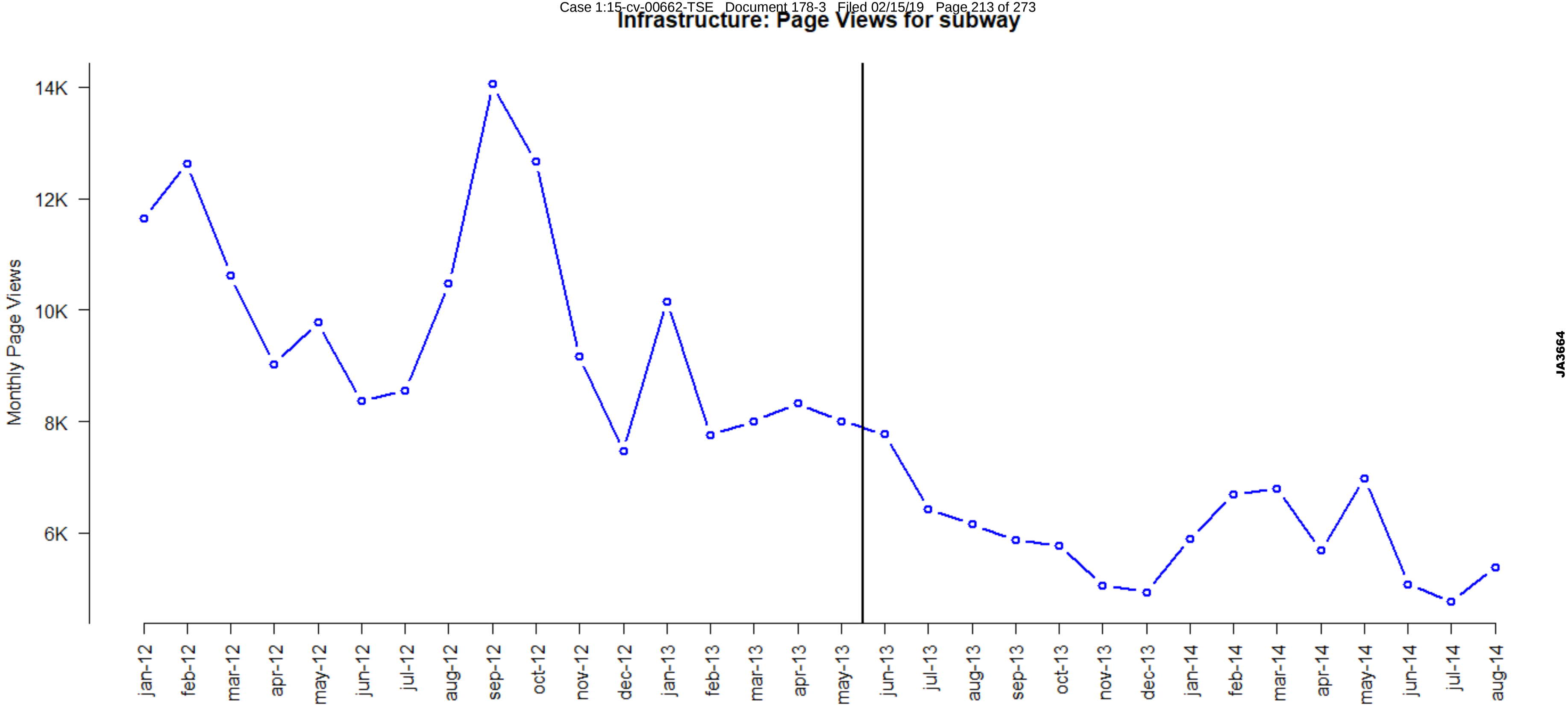

Infrastructure: Page Views for port
Monthly Page Views
28K
26K
24K
22K
20K
18K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

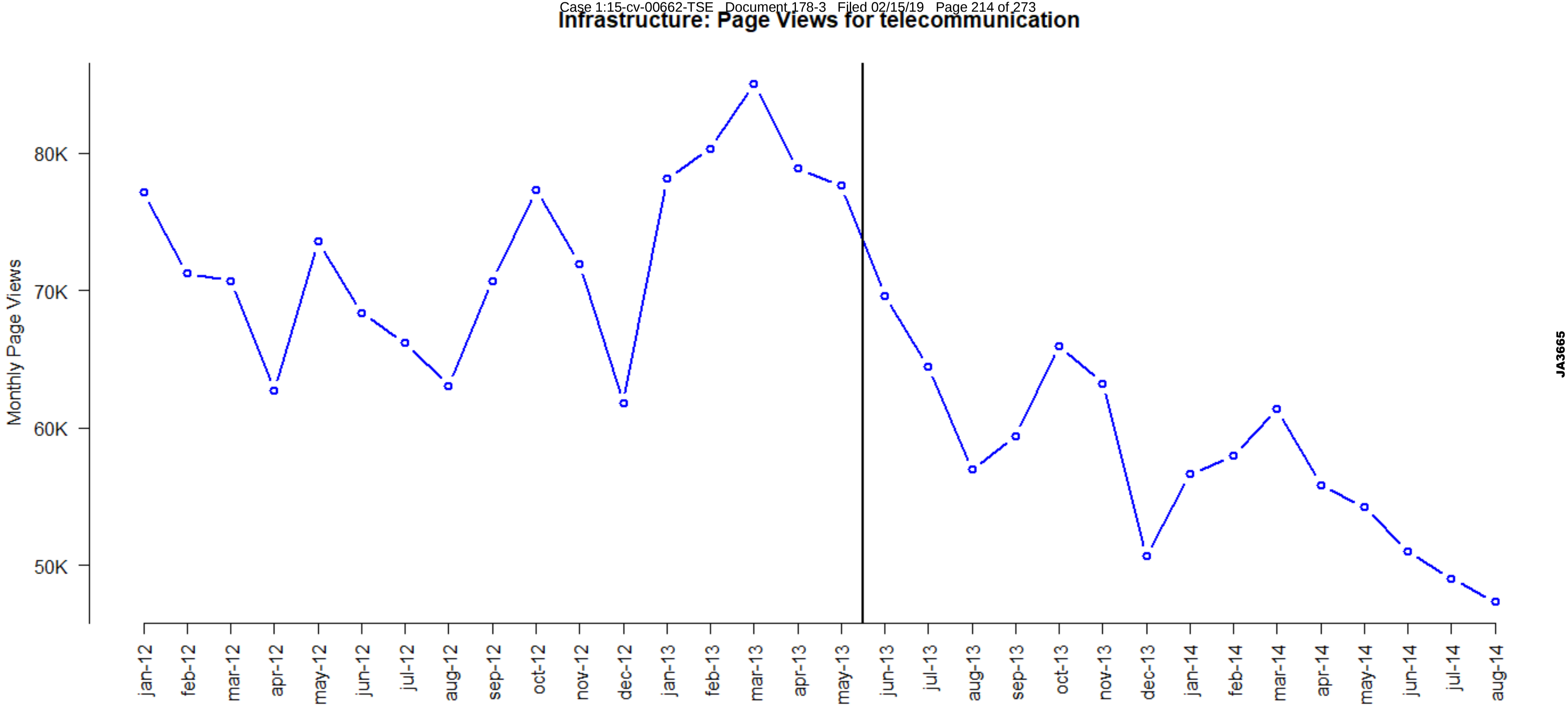

Infrastructure: Page Views for Port_authority
Monthly Page Views
3.5K
3K
2.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

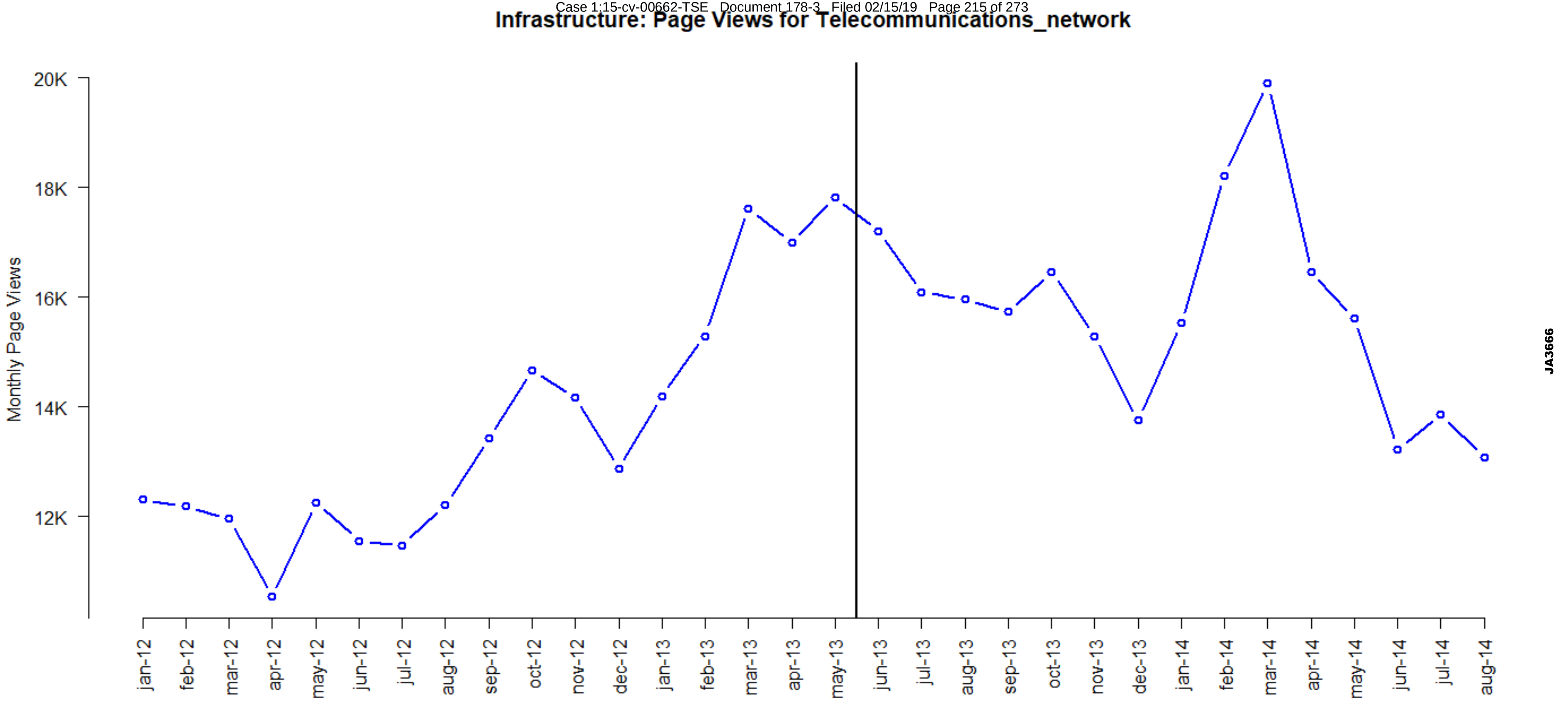

Infrastructure: Page Views for power
Monthly Page Views
35K
30K
25K
20K
15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

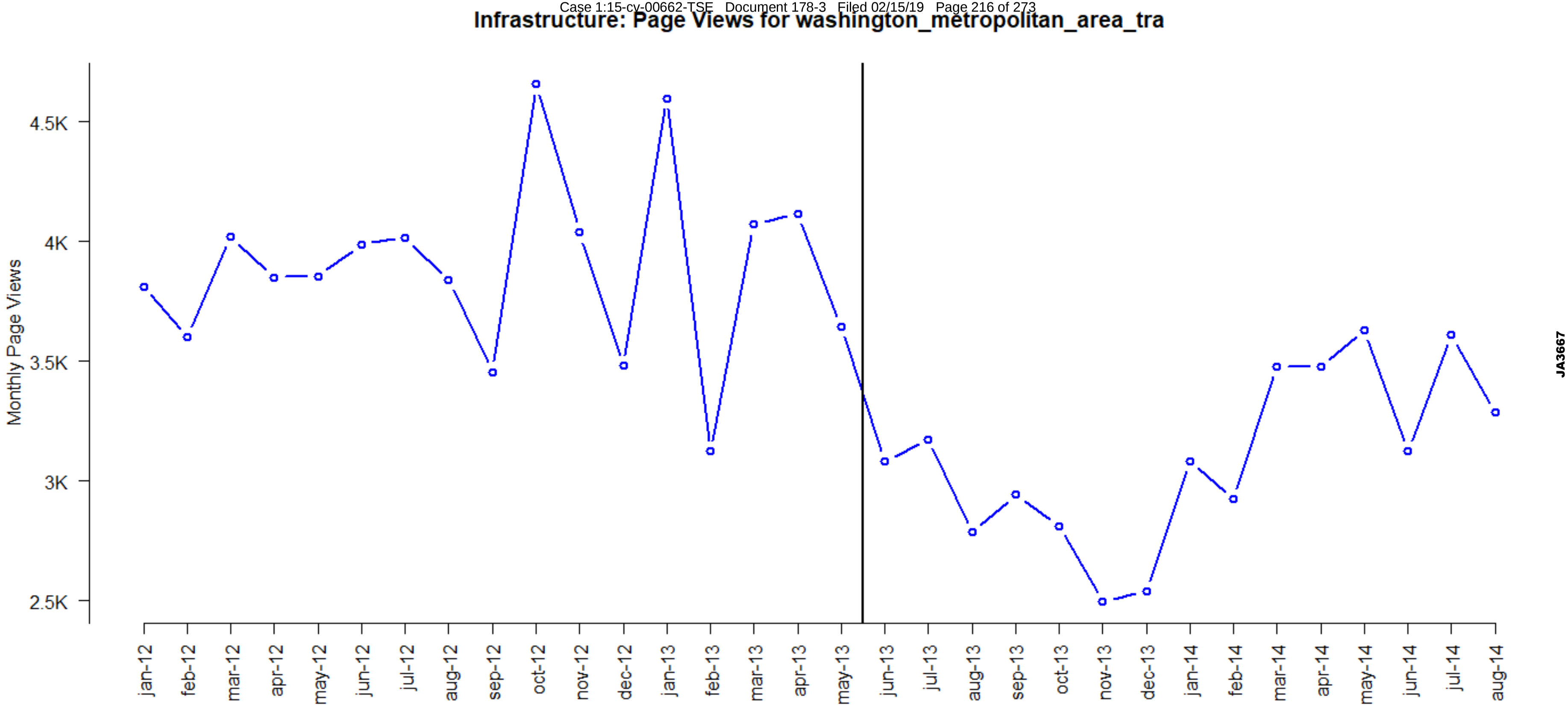

Infrastructure: Page Views for power_outage
Monthly Page Views
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

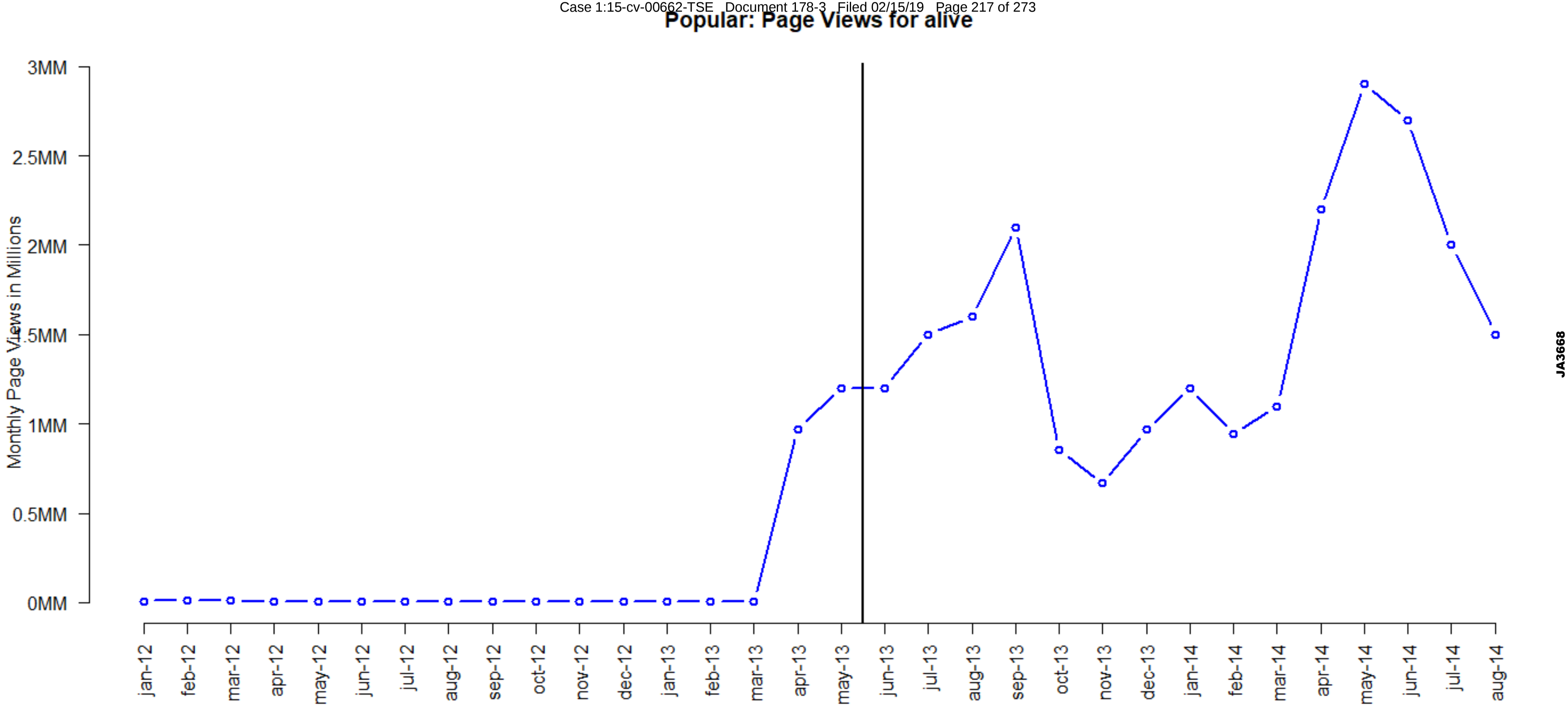

Infrastructure: Page Views for smart
Monthly Page Views
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

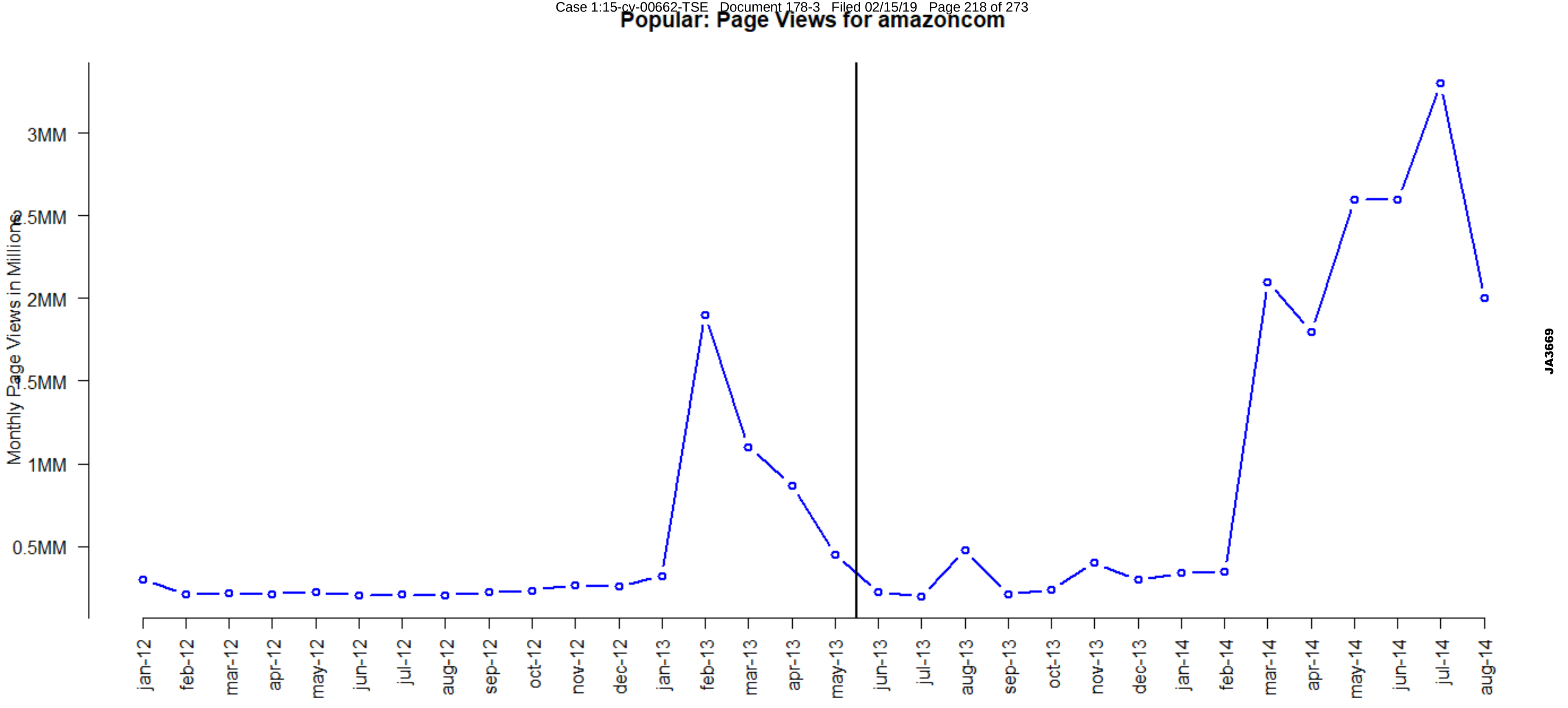

Infrastructure: Page Views for subway
Monthly Page Views
14K
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

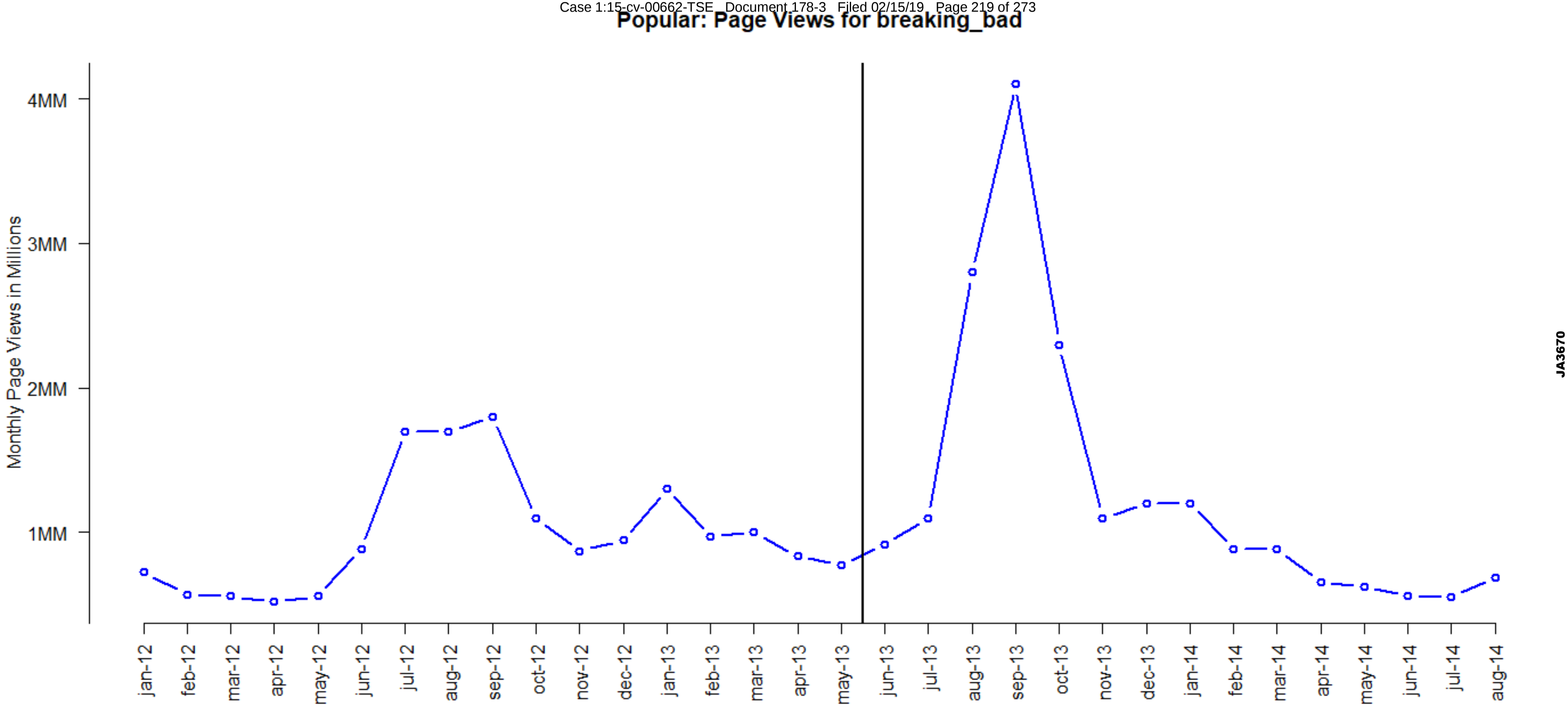

Infrastructure: Page Views for telecommunication
Monthly Page Views
80K
70K
60K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

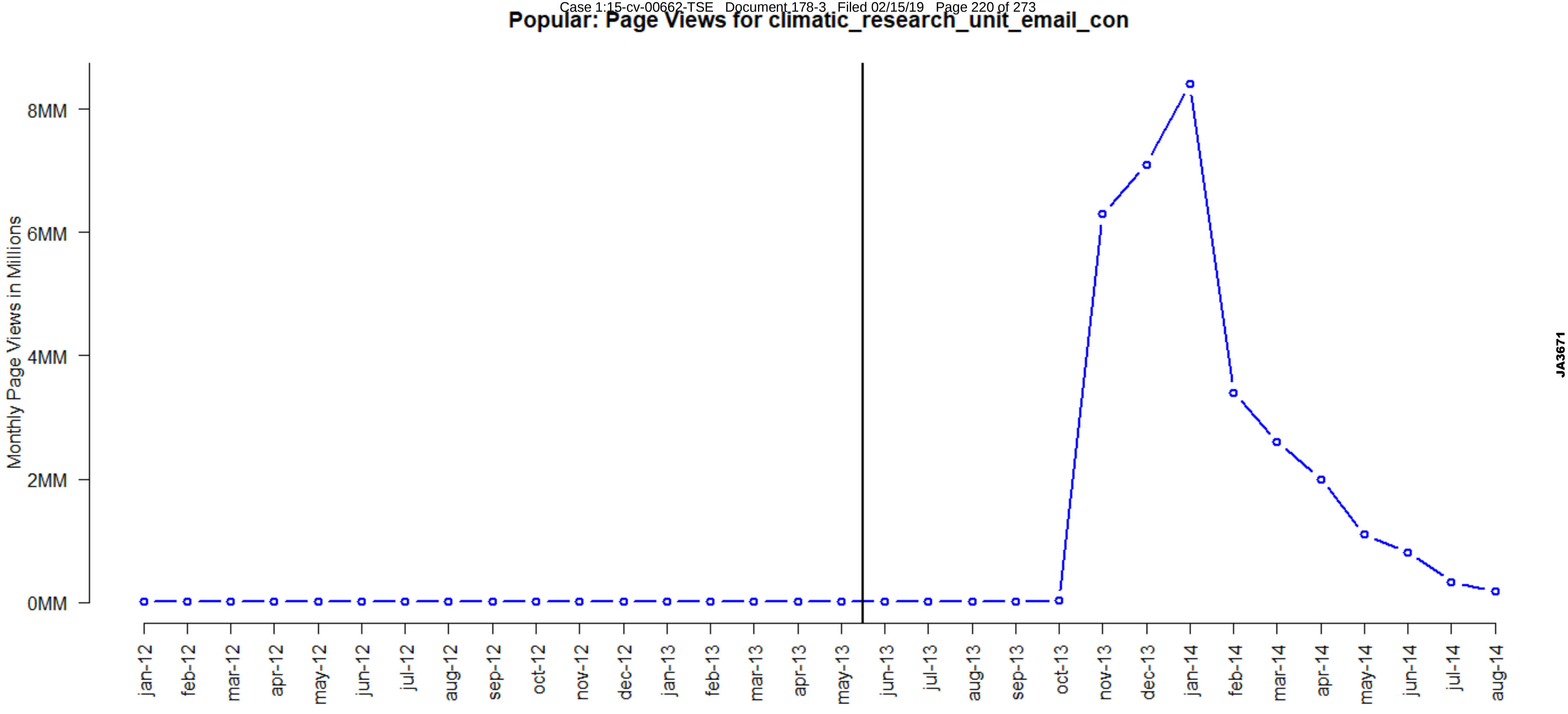

Infrastructure: Page Views for Telecommunications_network
Monthly Page Views
20K
18K
16K
14K
12K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Infrastructure: Page Views for washington_metropolitan_area_tra
Monthly Page Views
4.5K
4K
3.5K
3K
2.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

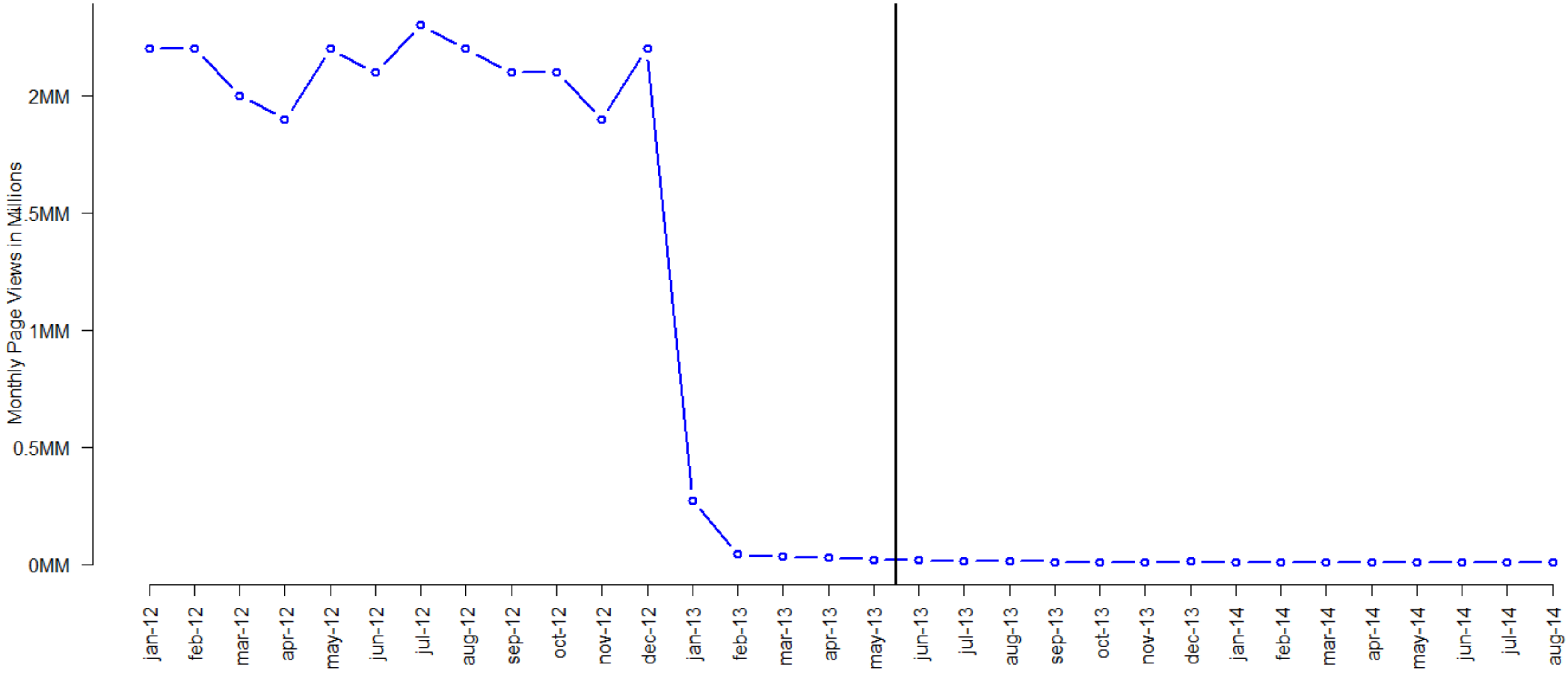

Popular: Page Views for alive
Monthly Page Views in Millions
3MM
2.5MM
2MM
1.5MM
1MM
0.5MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

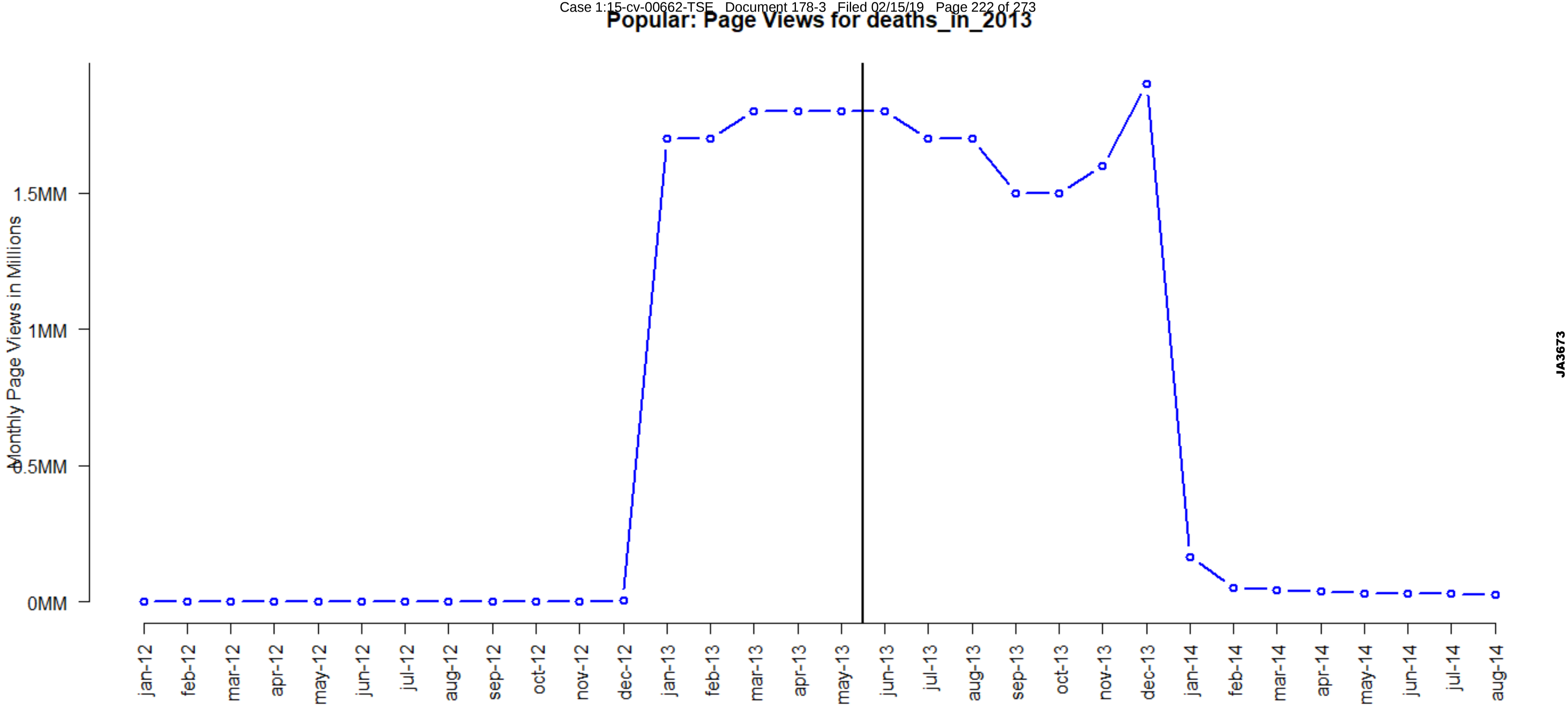

Popular: Page Views for amazoncom
Monthly Page Views in Millions
3MM
2.5MM
2MM
1.5MM
1MM
0.5MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

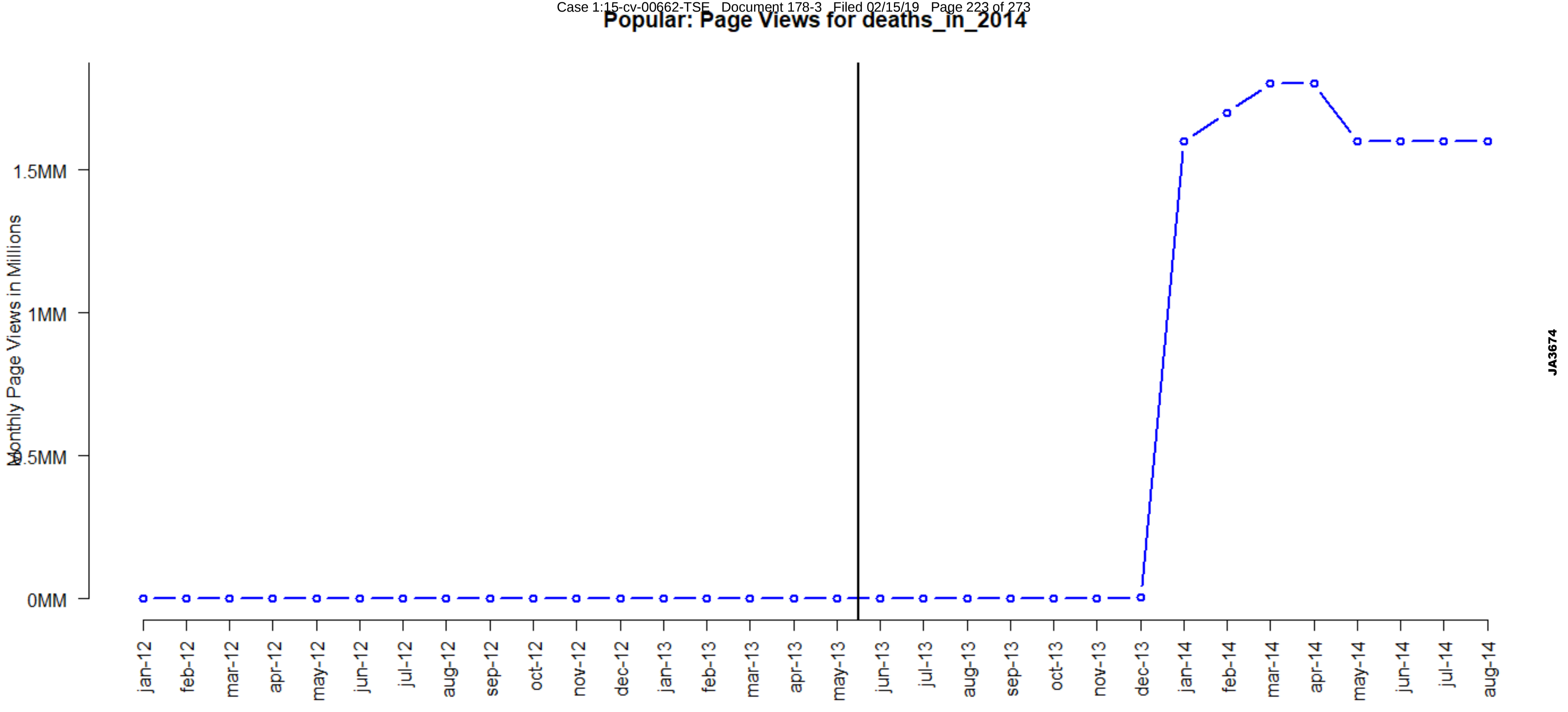

Popular: Page Views for breaking_bad
Monthly Page Views in Millions
4MM
3MM
2MM
1MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

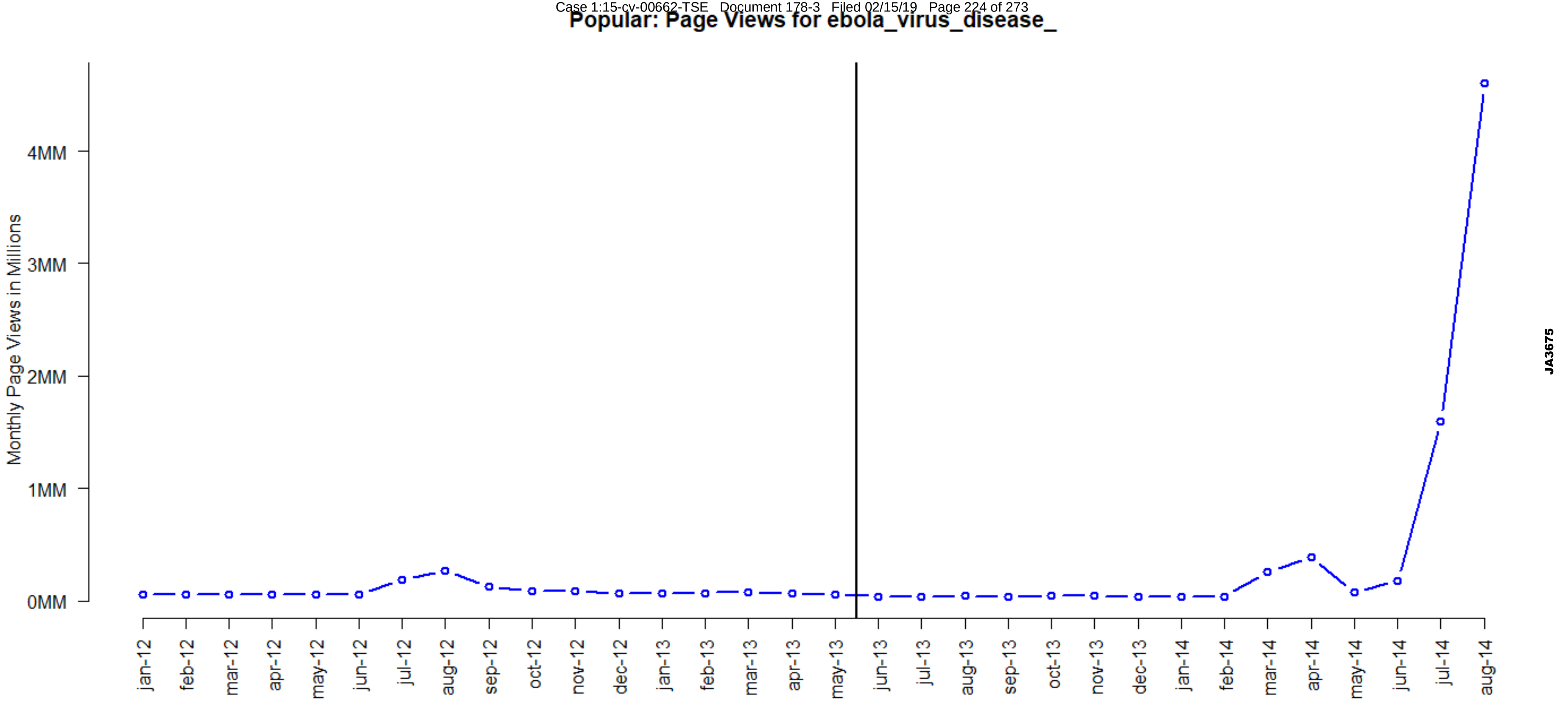

Popular: Page Views for climatic_research_unit_email_con
Monthly Page Views in Millions
0MM
2MM
4MM
6MM
8MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

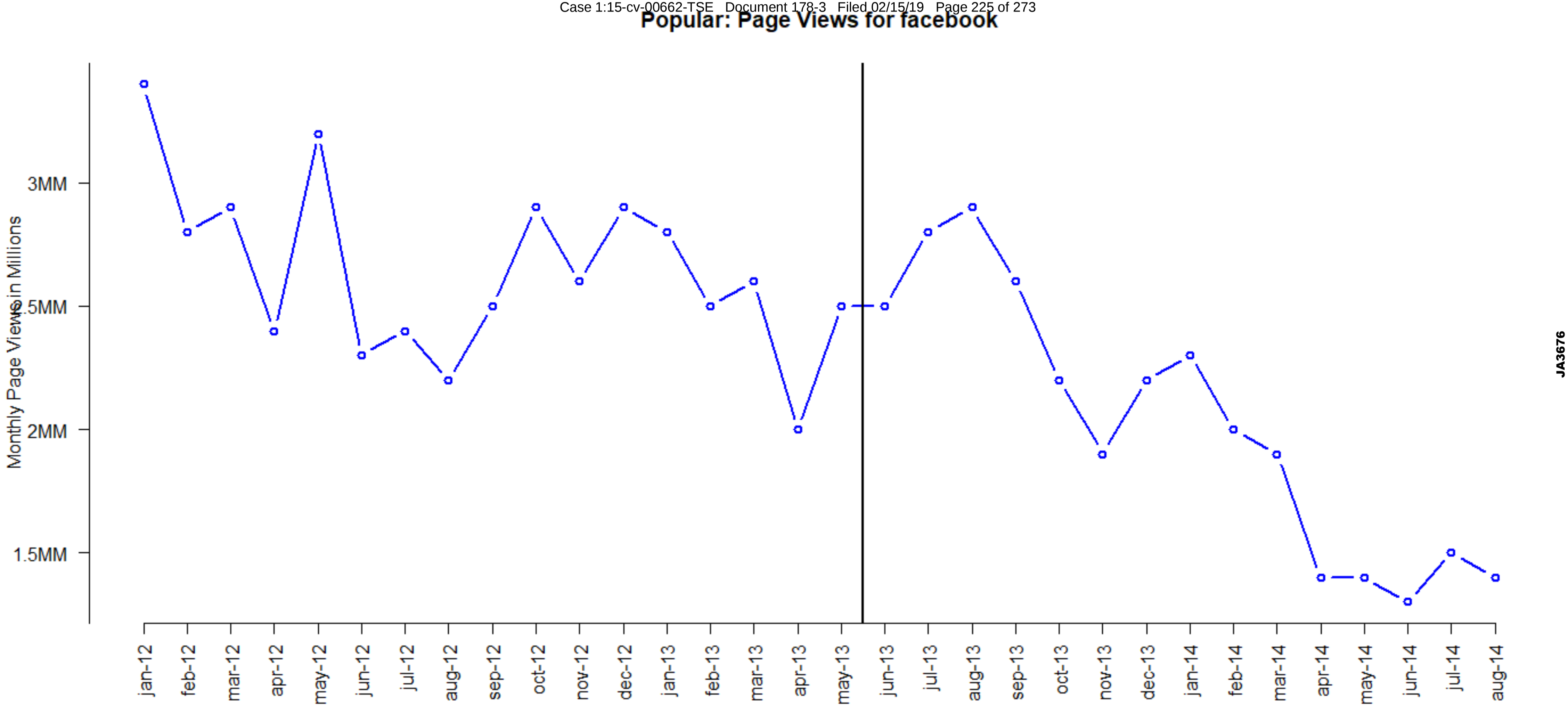
Popular: Page Views for deaths_in_2012

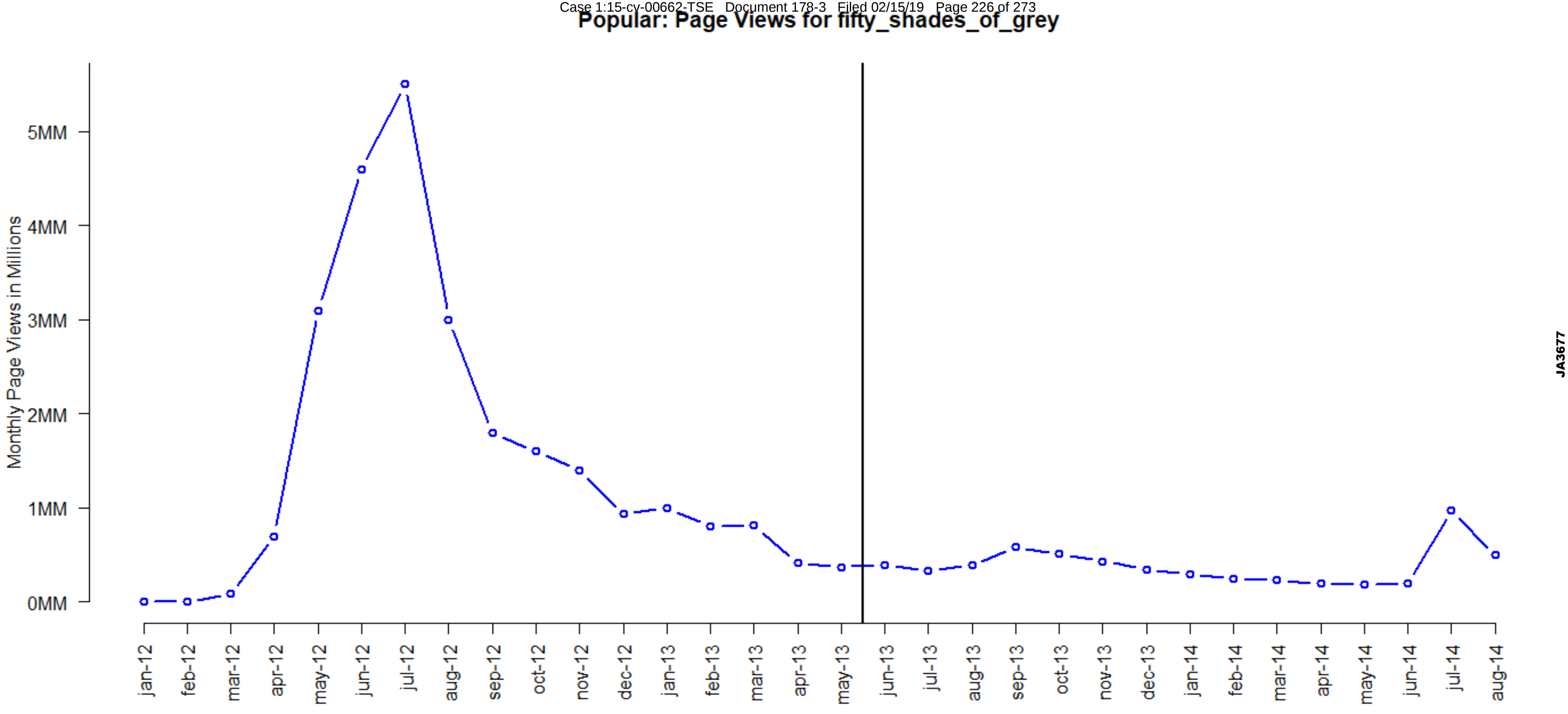


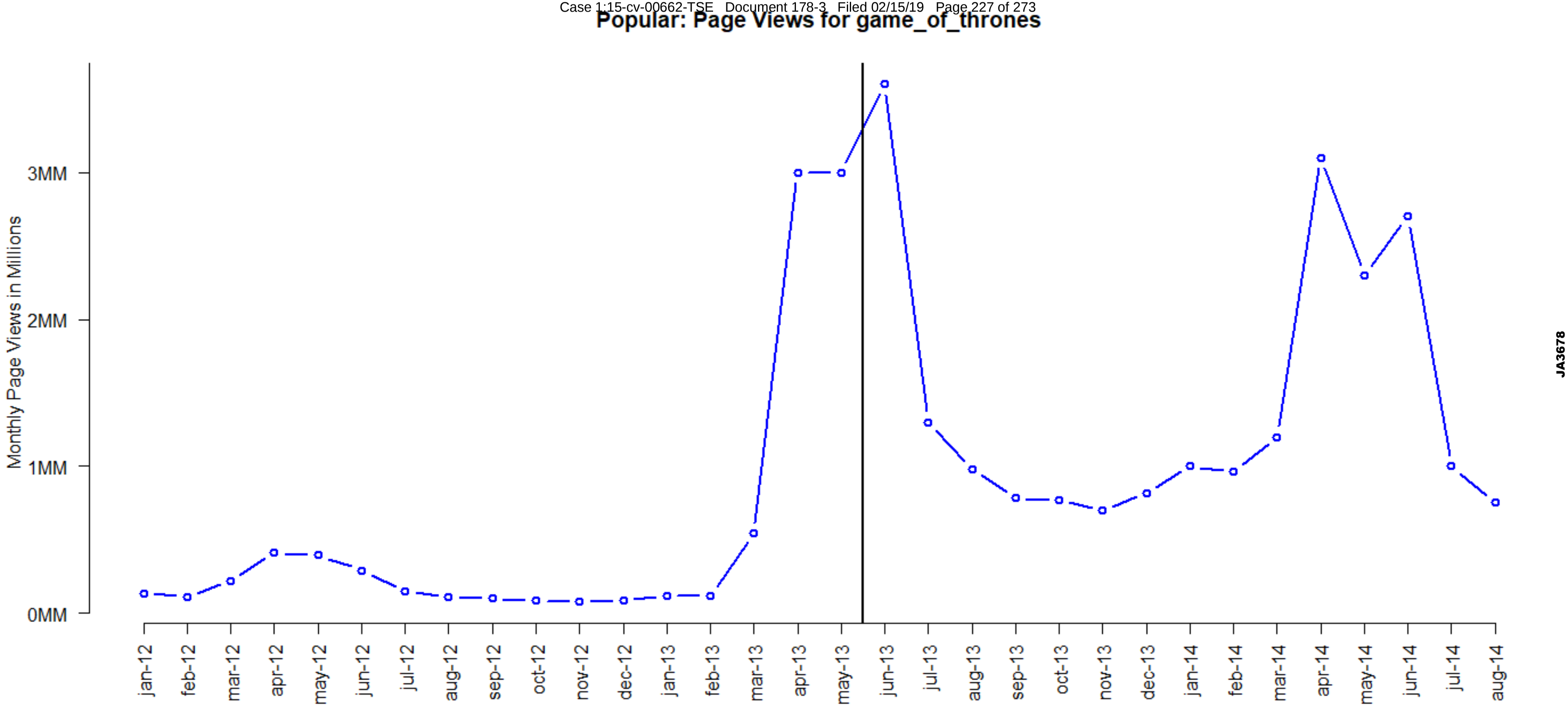

Popular: Page Views for deaths_in_2013
Monthly Page Views in Millions
1.5MM
1MM
0.5MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

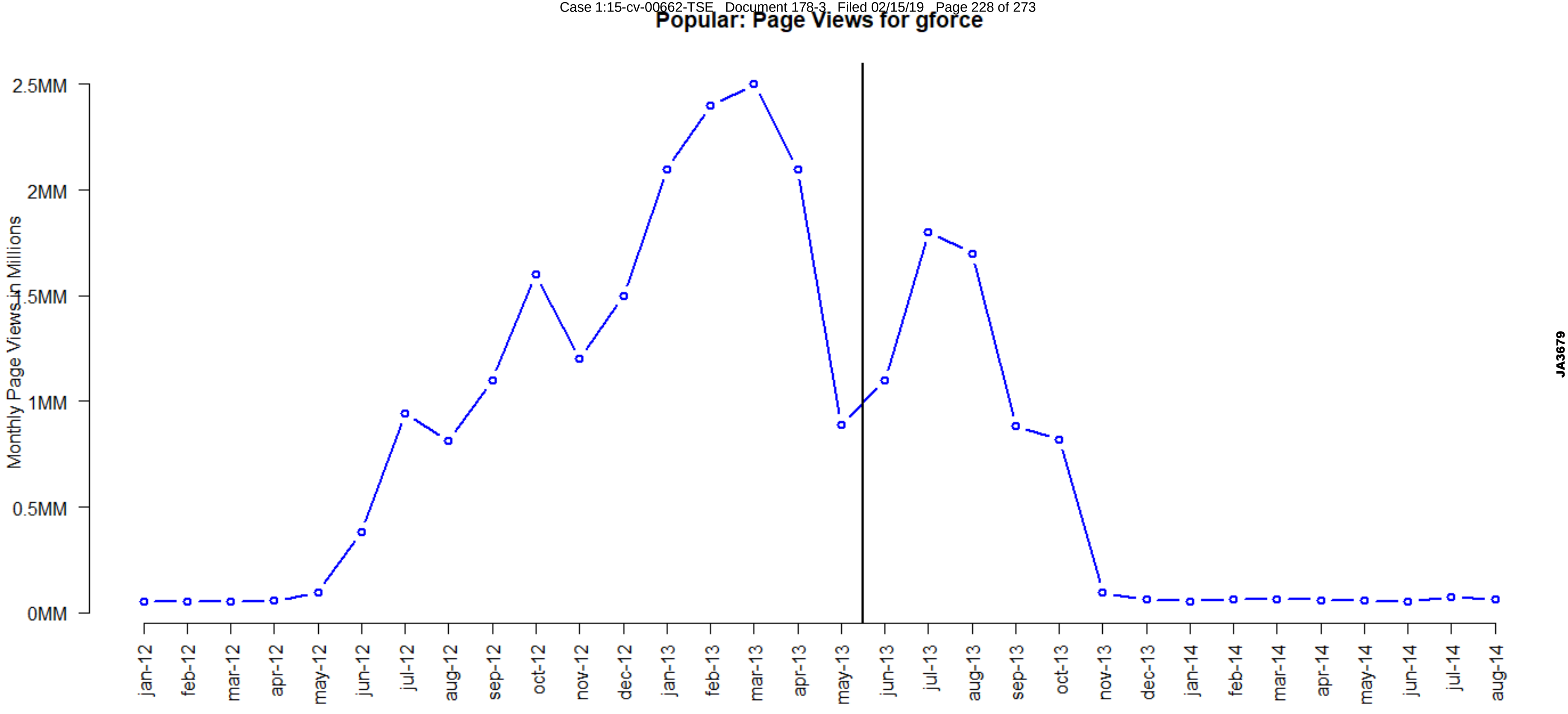

Popular: Page Views for deaths_in_2014
Monthly Page Views in Millions
1.5MM
1MM
0.5MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

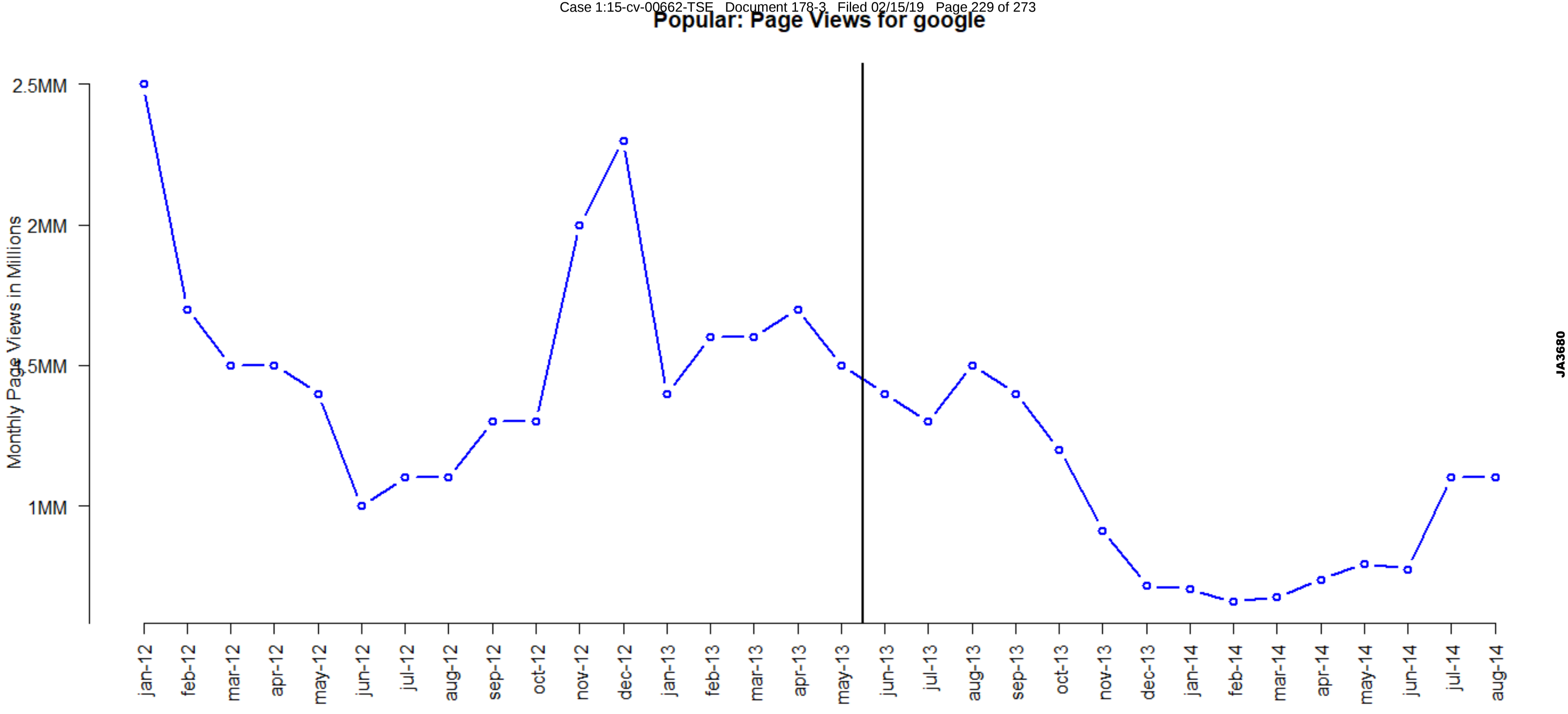

Popular: Page Views for ebola_virus_disease_
Monthly Page Views in Millions
0MM
1MM
2MM
3MM
4MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

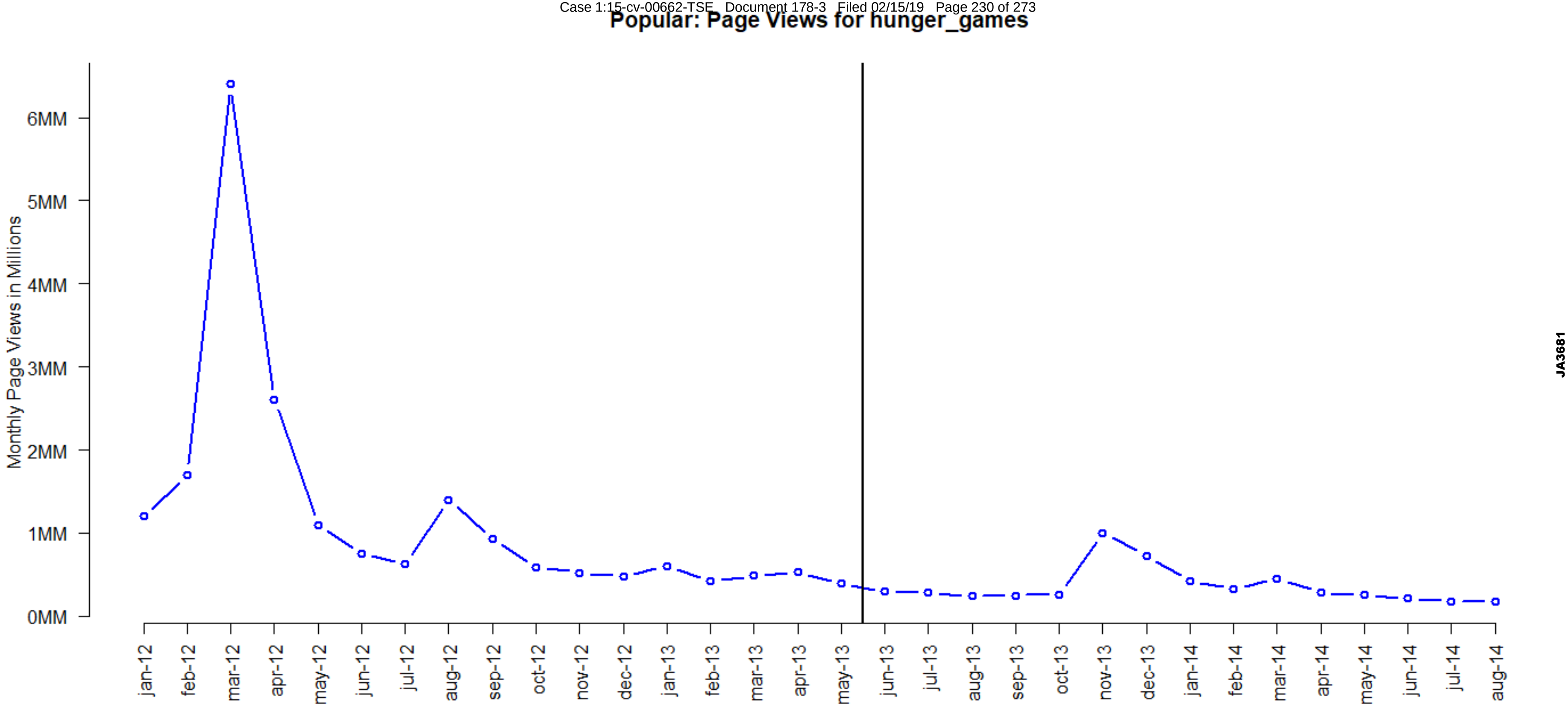

Popular: Page Views for facebook
Monthly Page Views in Millions
3MM
2.5MM
2MM
1.5MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

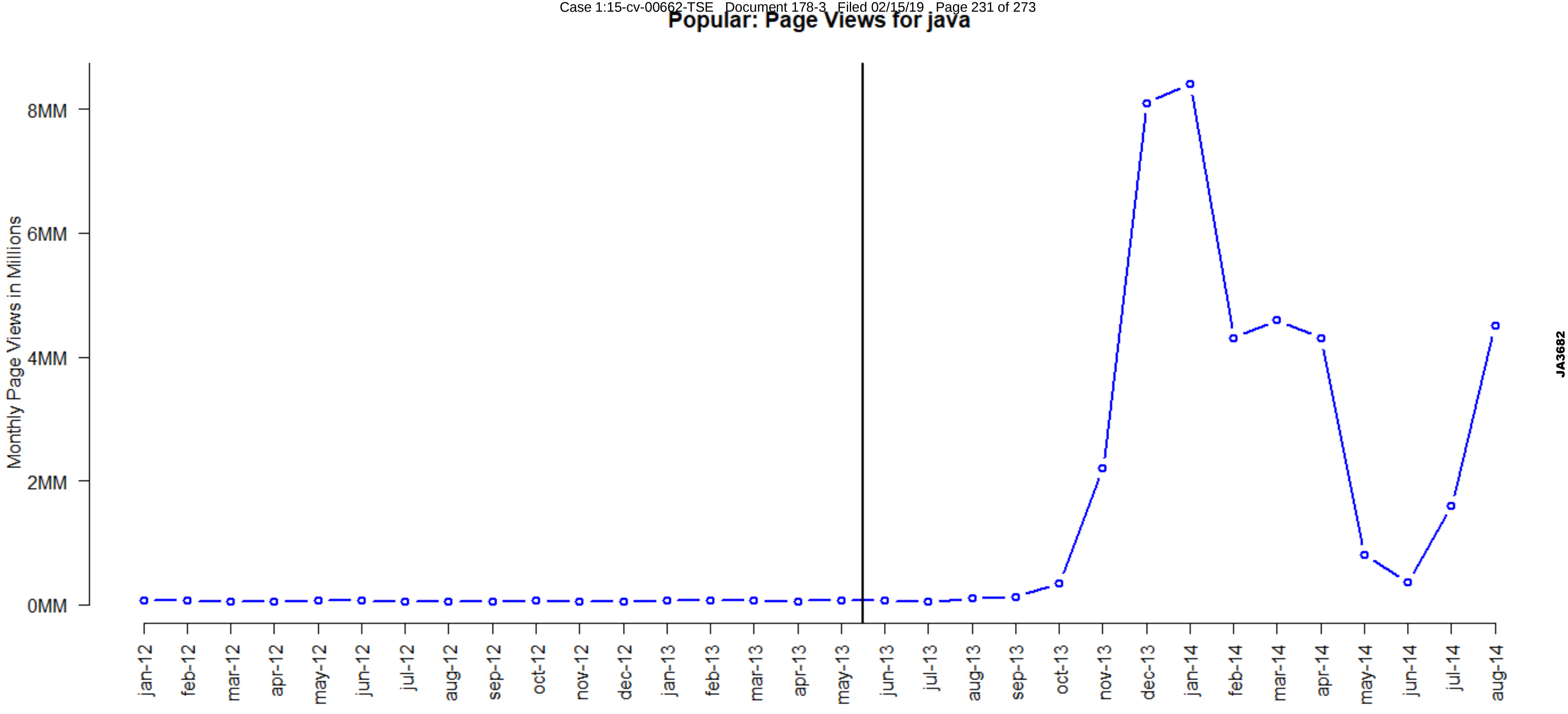

Popular: Page Views for fifty_shades_of_grey
Monthly Page Views in Millions
0MM
1MM
2MM
3MM
4MM
5MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Popular: Page Views for game_of_thrones
Monthly Page Views in Millions
0MM
1MM
2MM
3MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

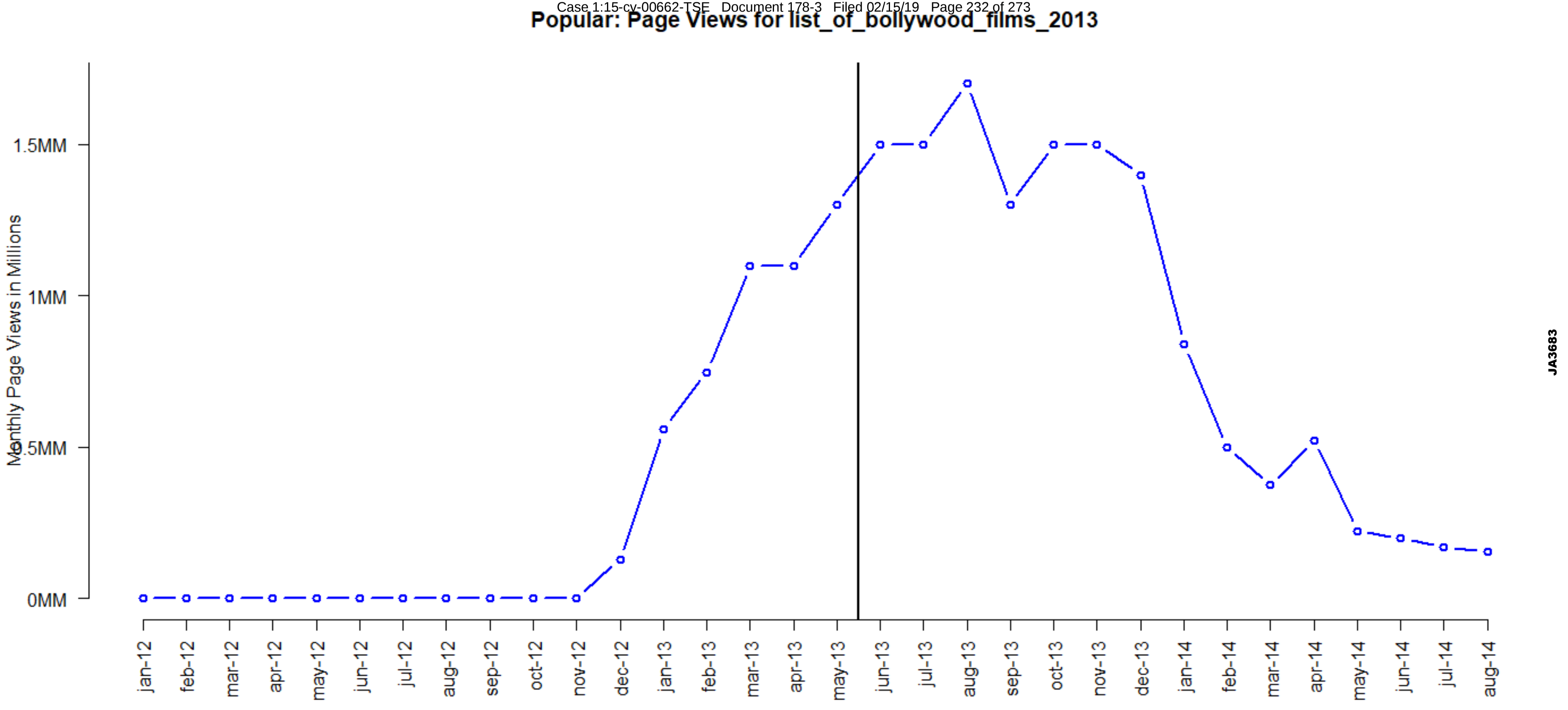

Popular: Page Views for gforce
Monthly Page Views in Millions
2.5MM
2MM
1.5MM
1MM
0.5MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

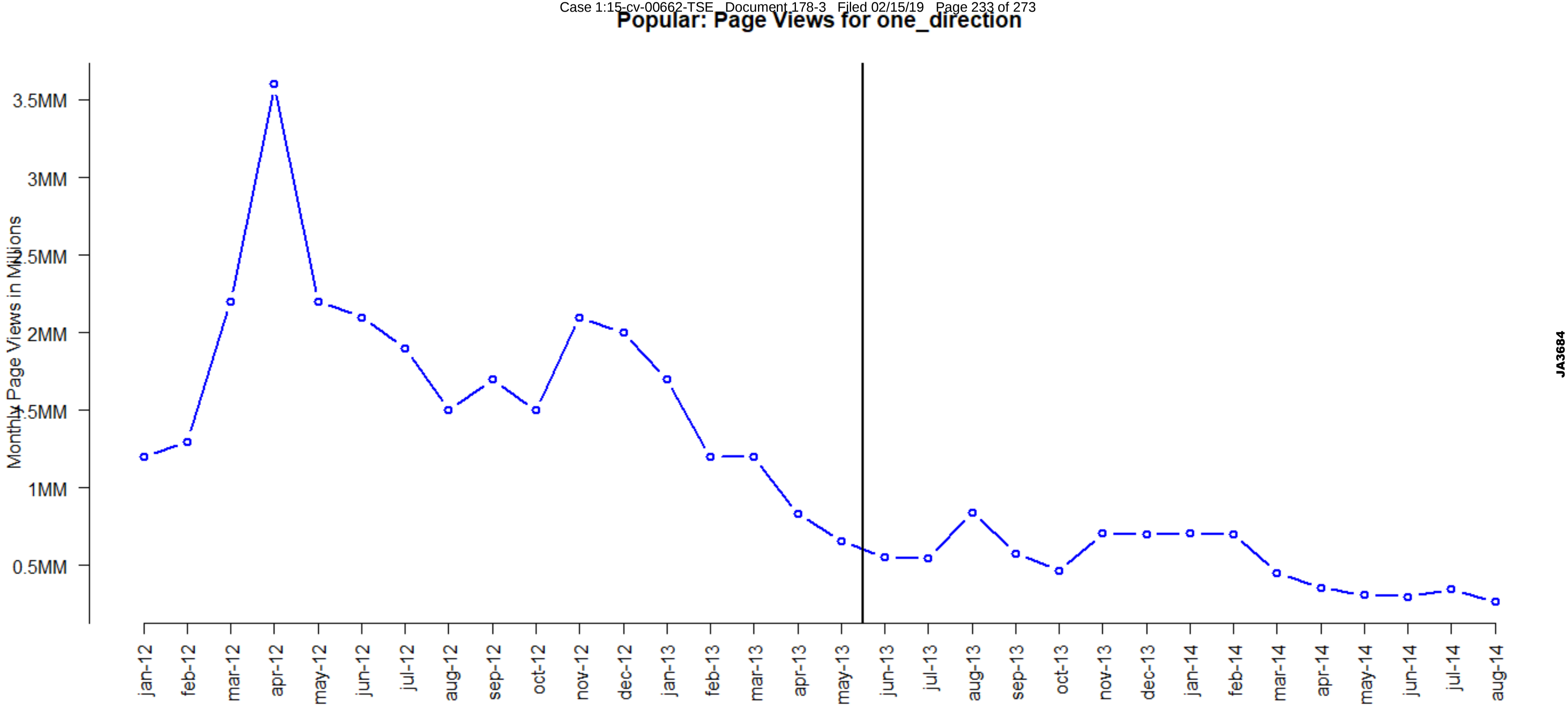

Popular: Page Views for google
Monthly Page Views in Millions
2.5MM
2MM
1.5MM
1MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

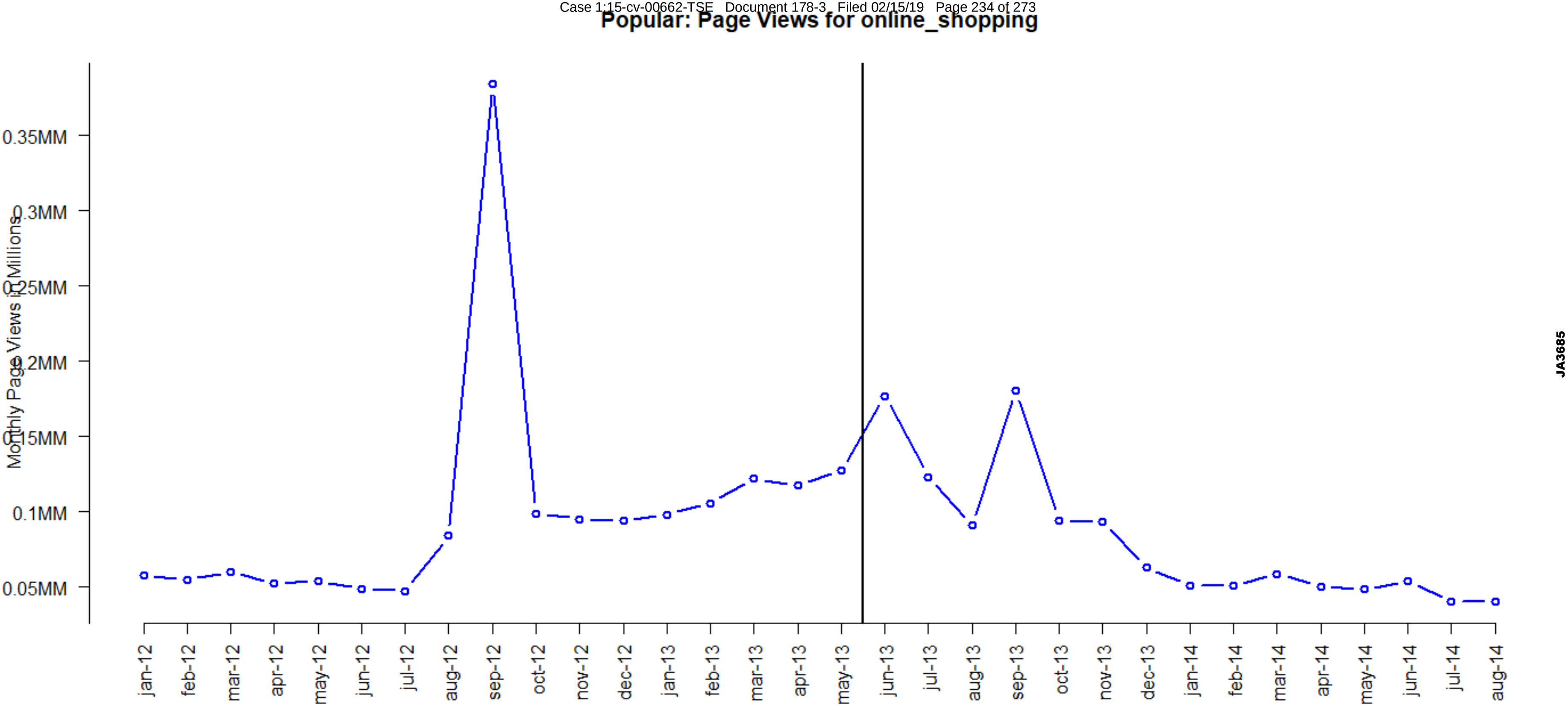

Popular: Page Views for hunger_games
Monthly Page Views in Millions
0MM
1MM
2MM
3MM
4MM
5MM
6MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

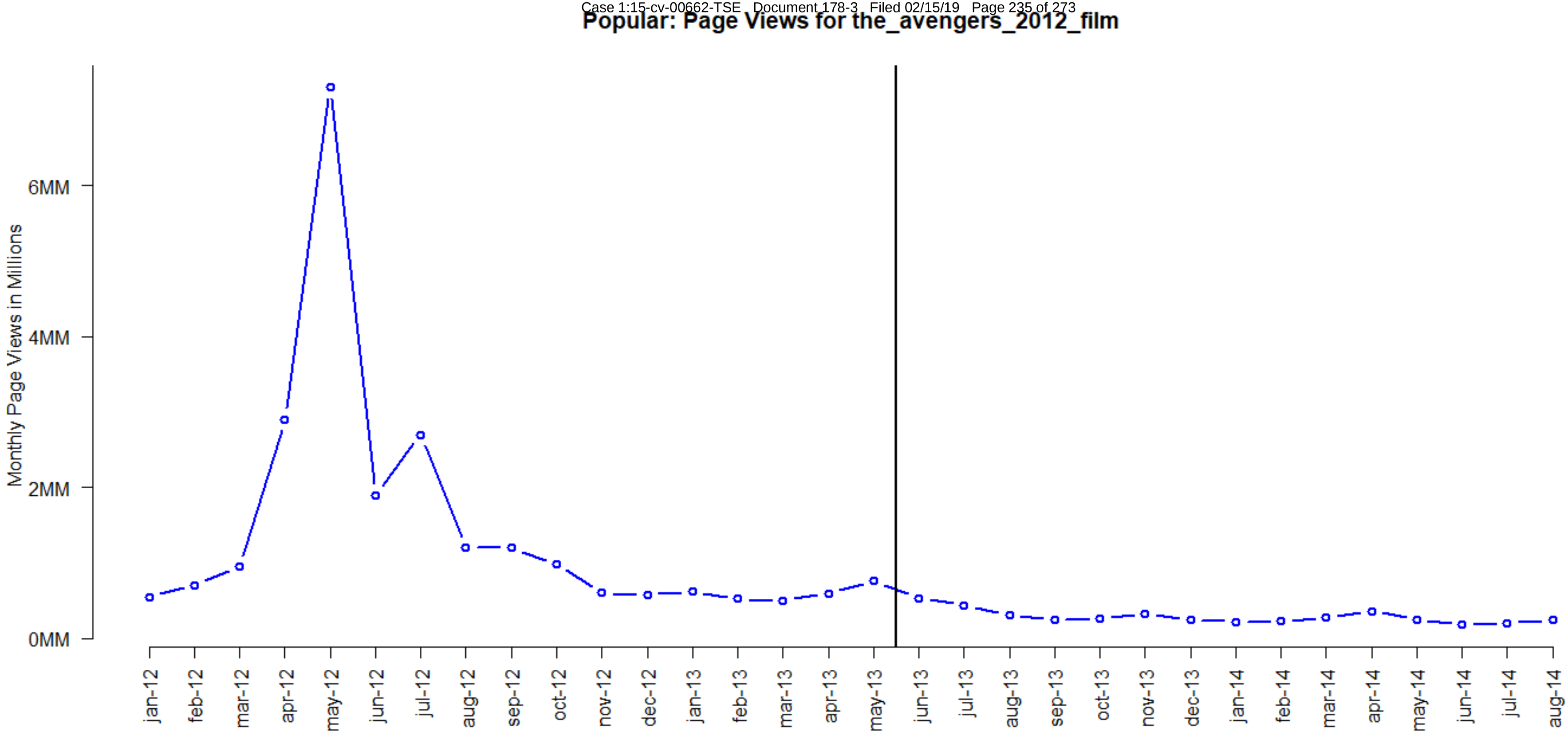

Popular: Page Views for java
Monthly Page Views in Millions
8MM
6MM
4MM
2MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

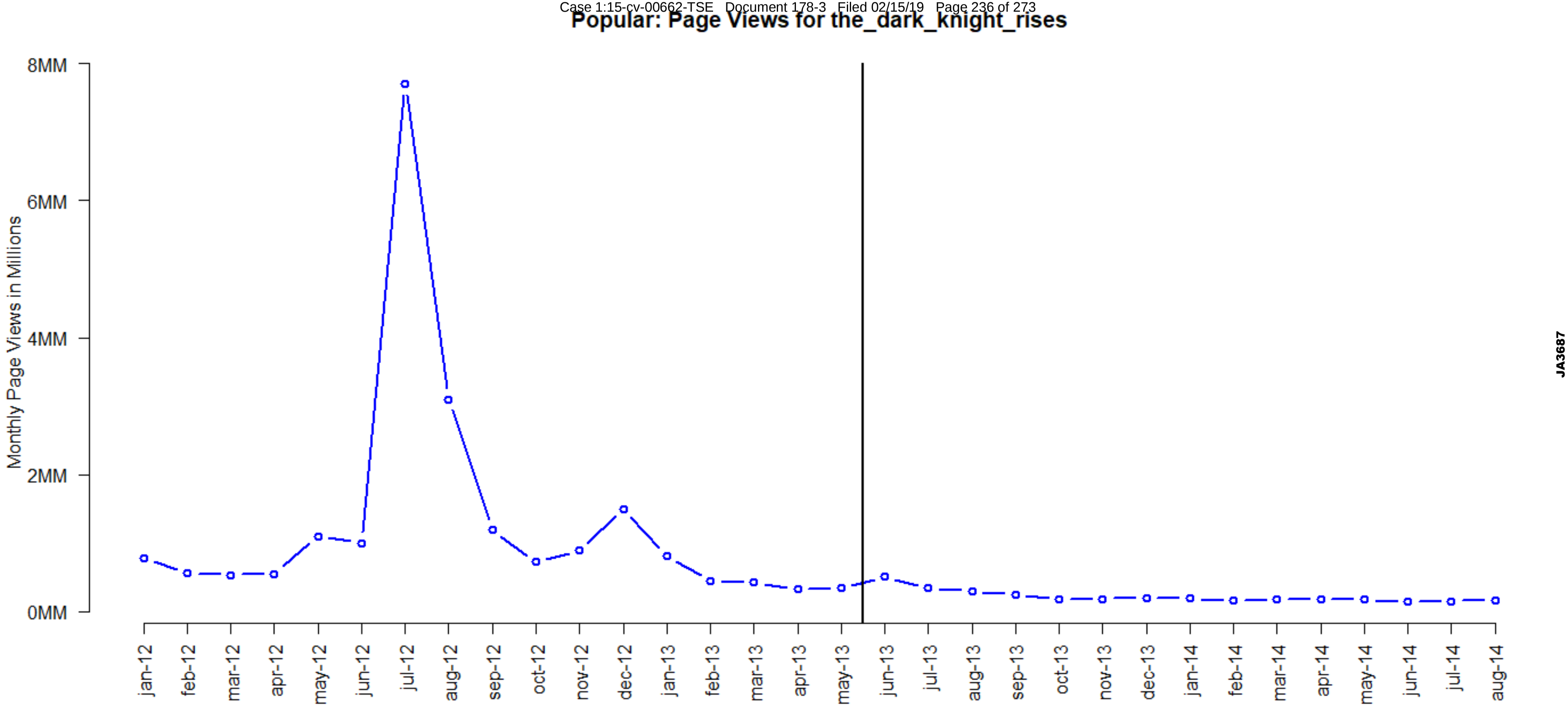
JA3682

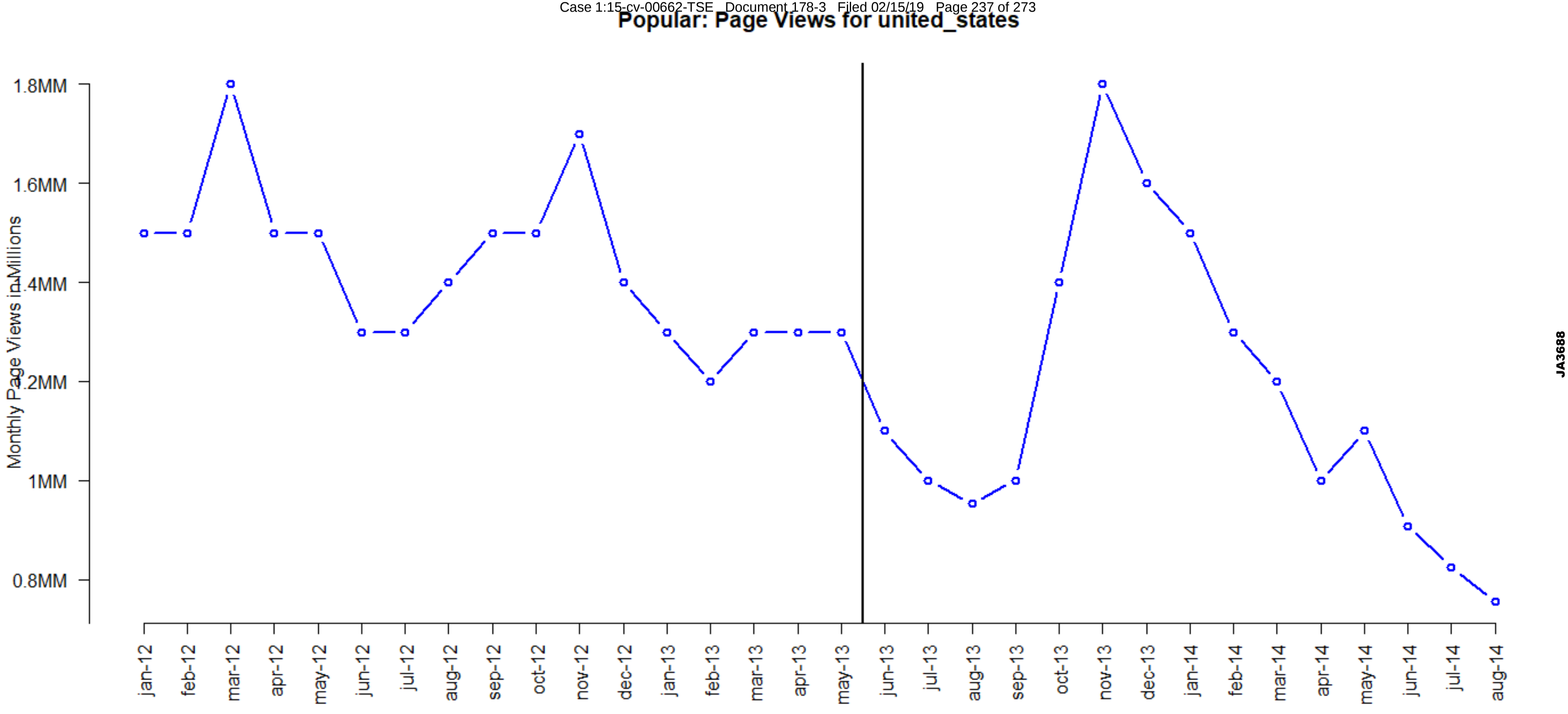

Popular: Page Views for list_of_bollywood_films_2013
Monthly Page Views in Millions
1.5MM
1MM
0.5MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

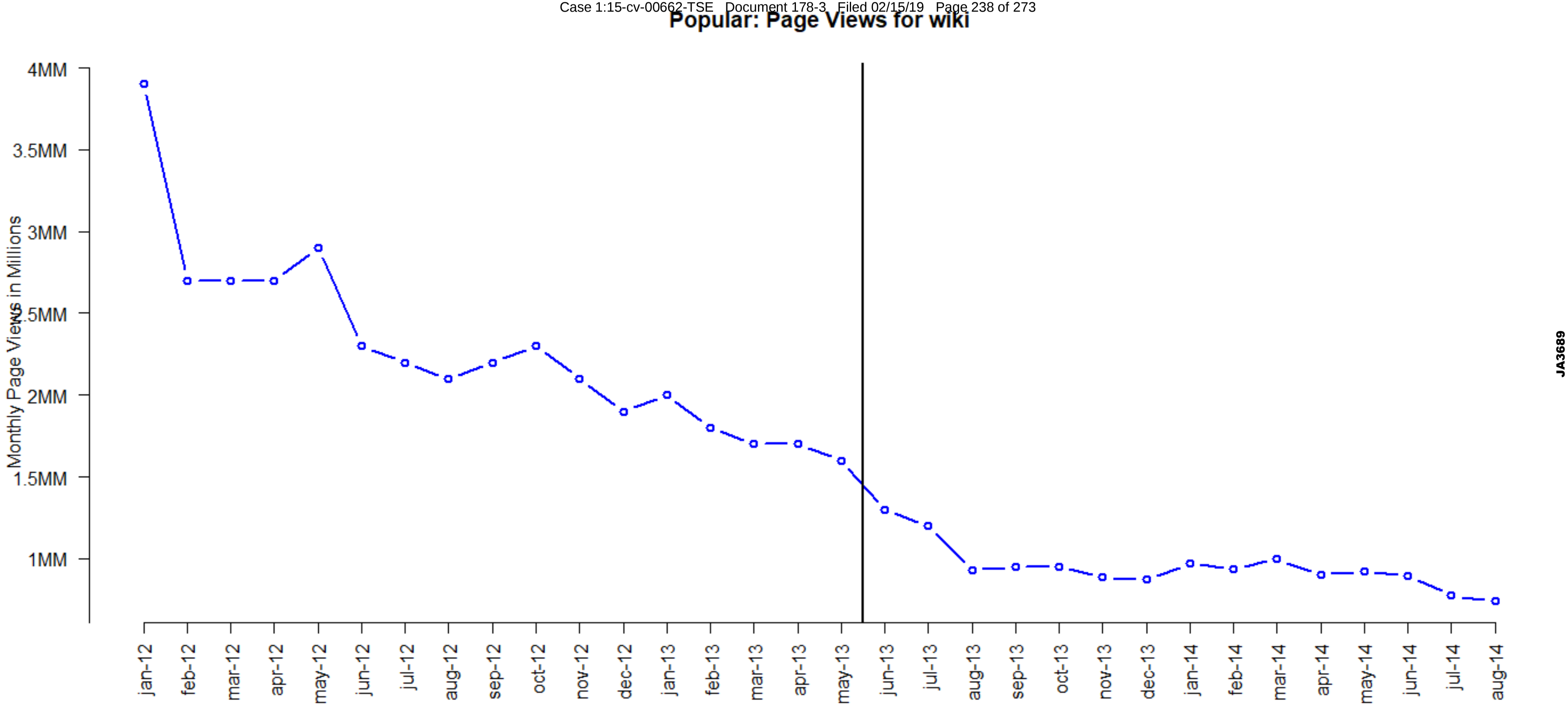

Popular: Page Views for one_direction
Monthly Page Views in Millions
3.5MM
3MM
2.5MM
2MM
1.5MM
1MM
0.5MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

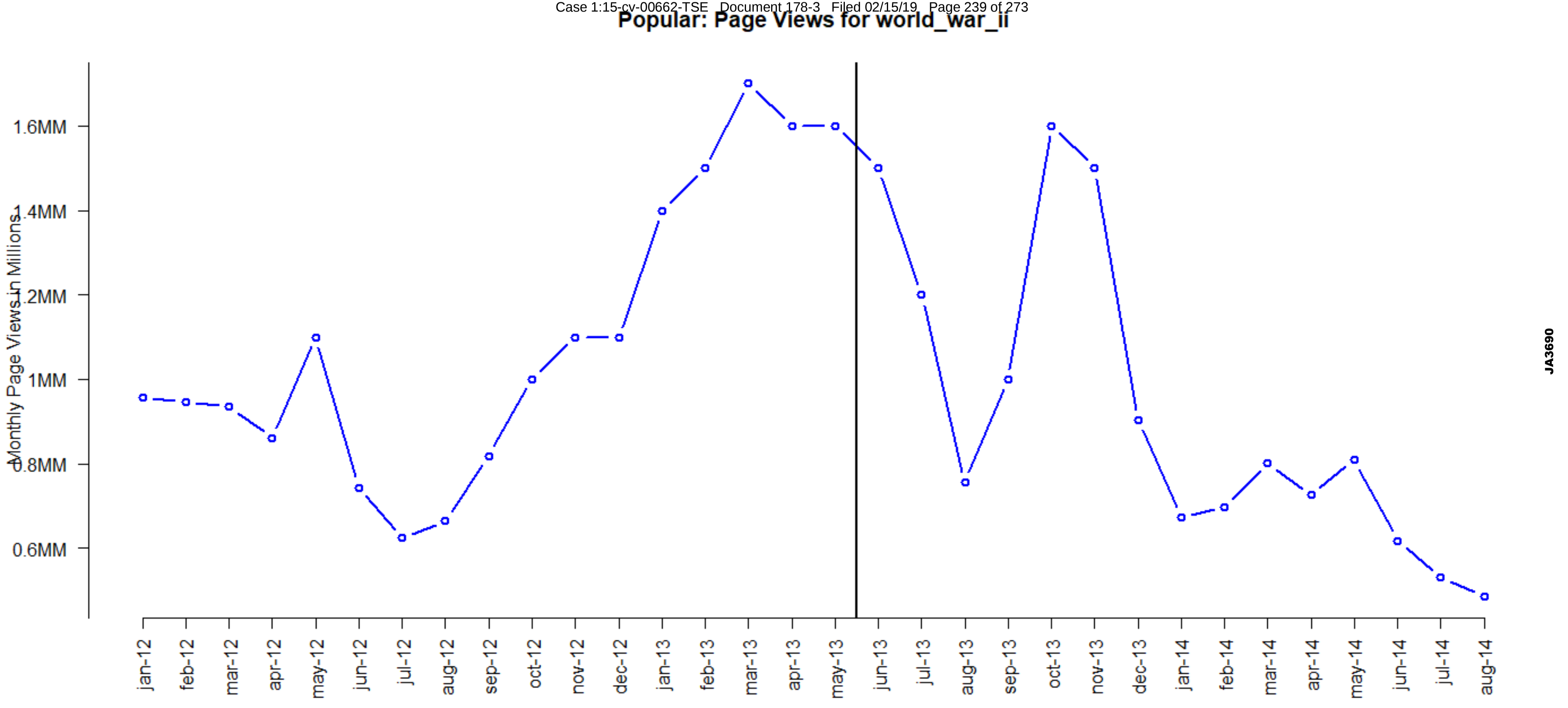

Popular: Page Views for online_shopping
Monthly Page Views in Millions
0.35MM
0.3MM
0.25MM
0.2MM
0.15MM
0.1MM
0.05MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

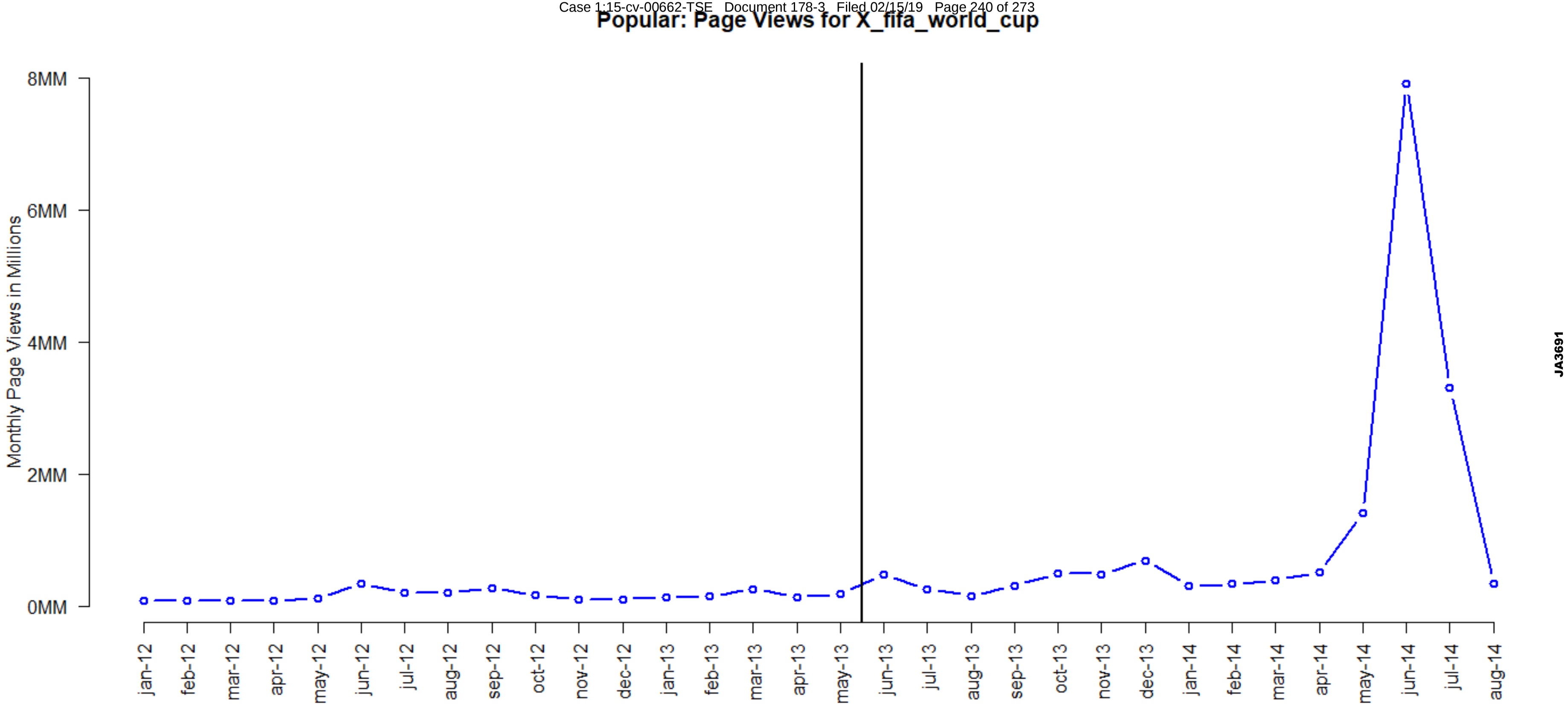

Popular: Page Views for the_avengers_2012_film
Monthly Page Views in Millions
0MM
2MM
4MM
6MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

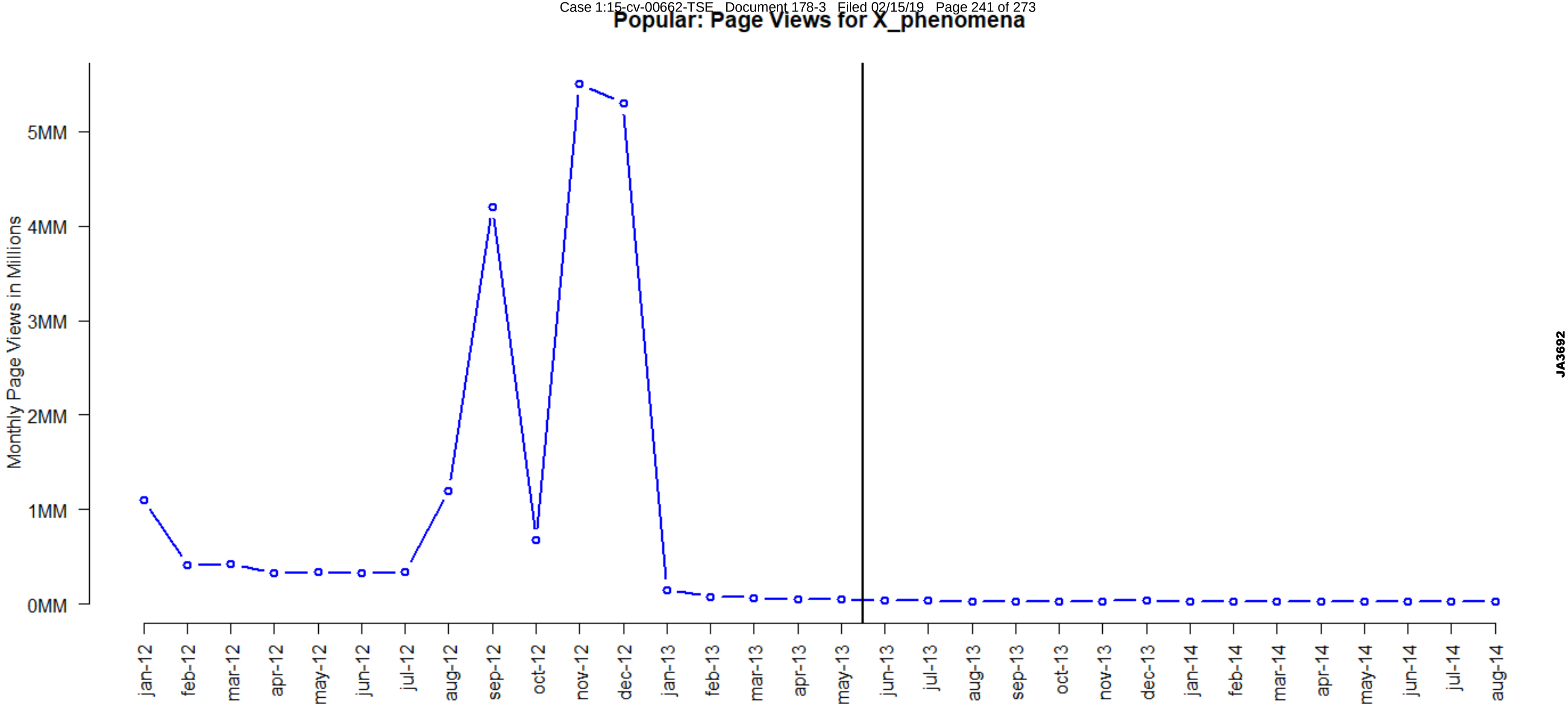

Popular: Page Views for the_dark_knight_rises
Monthly Page Views in Millions
8MM
6MM
4MM
2MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

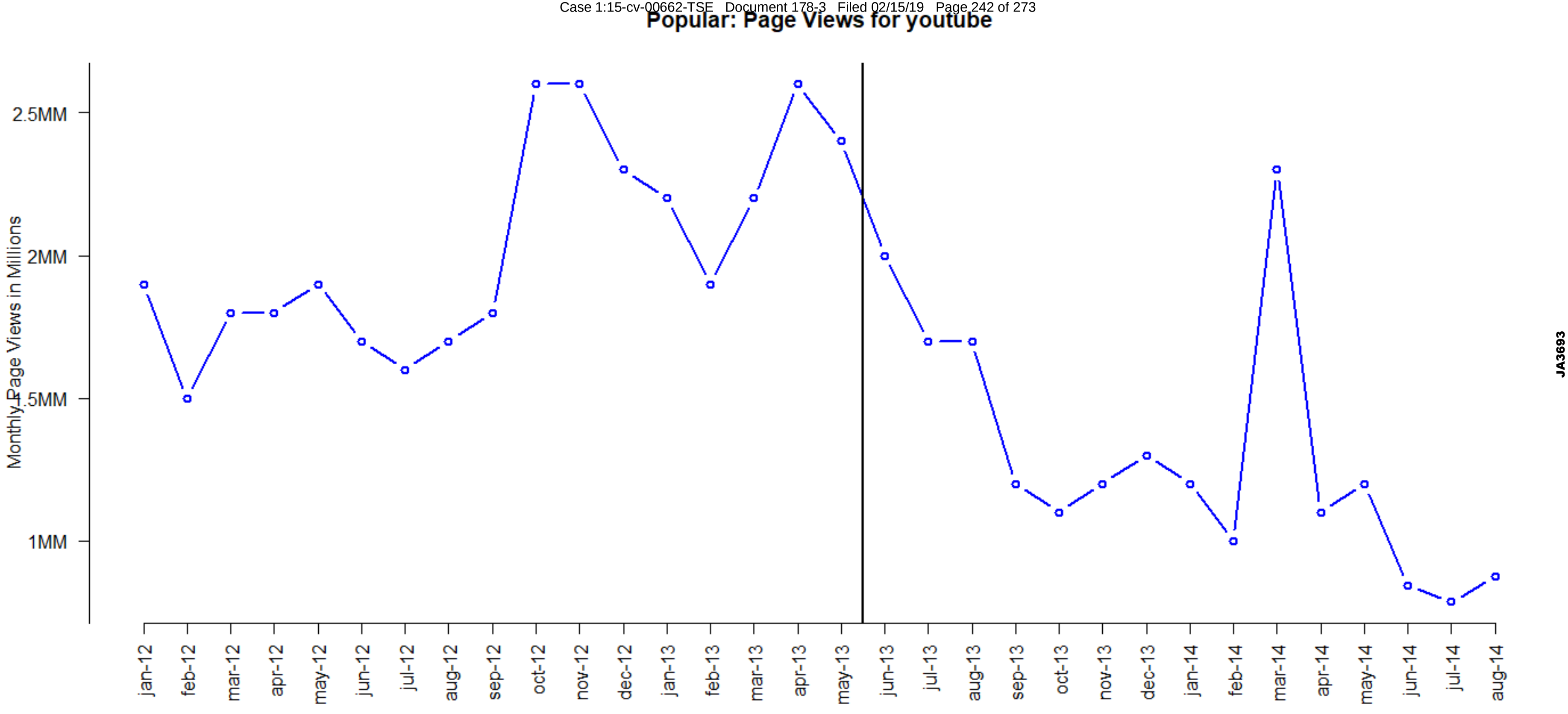

Popular: Page Views for united_states
Monthly Page Views in Millions
1.8MM
1.6MM
1.4MM
1.2MM
1MM
0.8MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

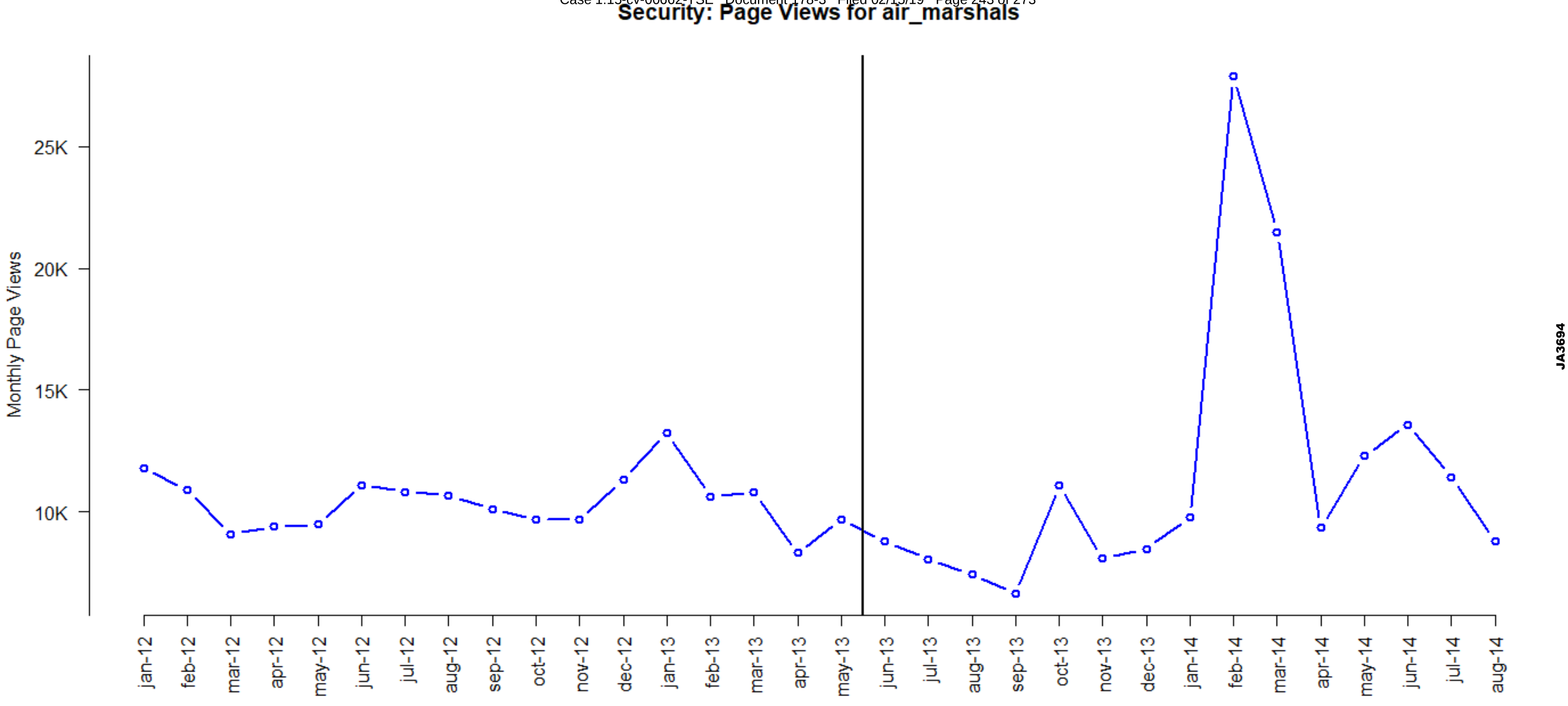

Popular: Page Views for wiki
Monthly Page Views in Millions
4MM
3.5MM
3MM
2.5MM
2MM
1.5MM
1MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

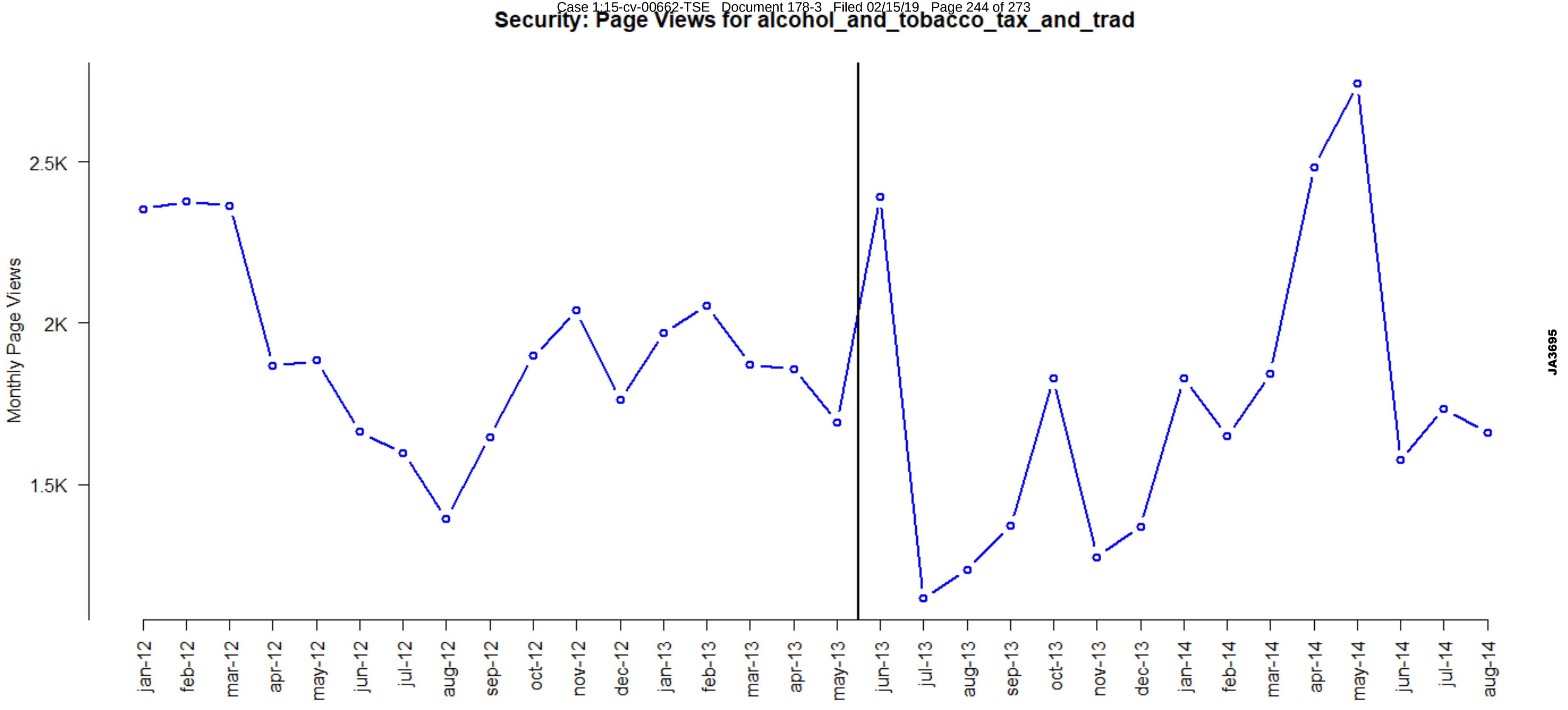

Popular: Page Views for world_war_ii
Monthly Page Views in Millions
1.6MM
1.4MM
1.2MM
1MM
0.8MM
0.6MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

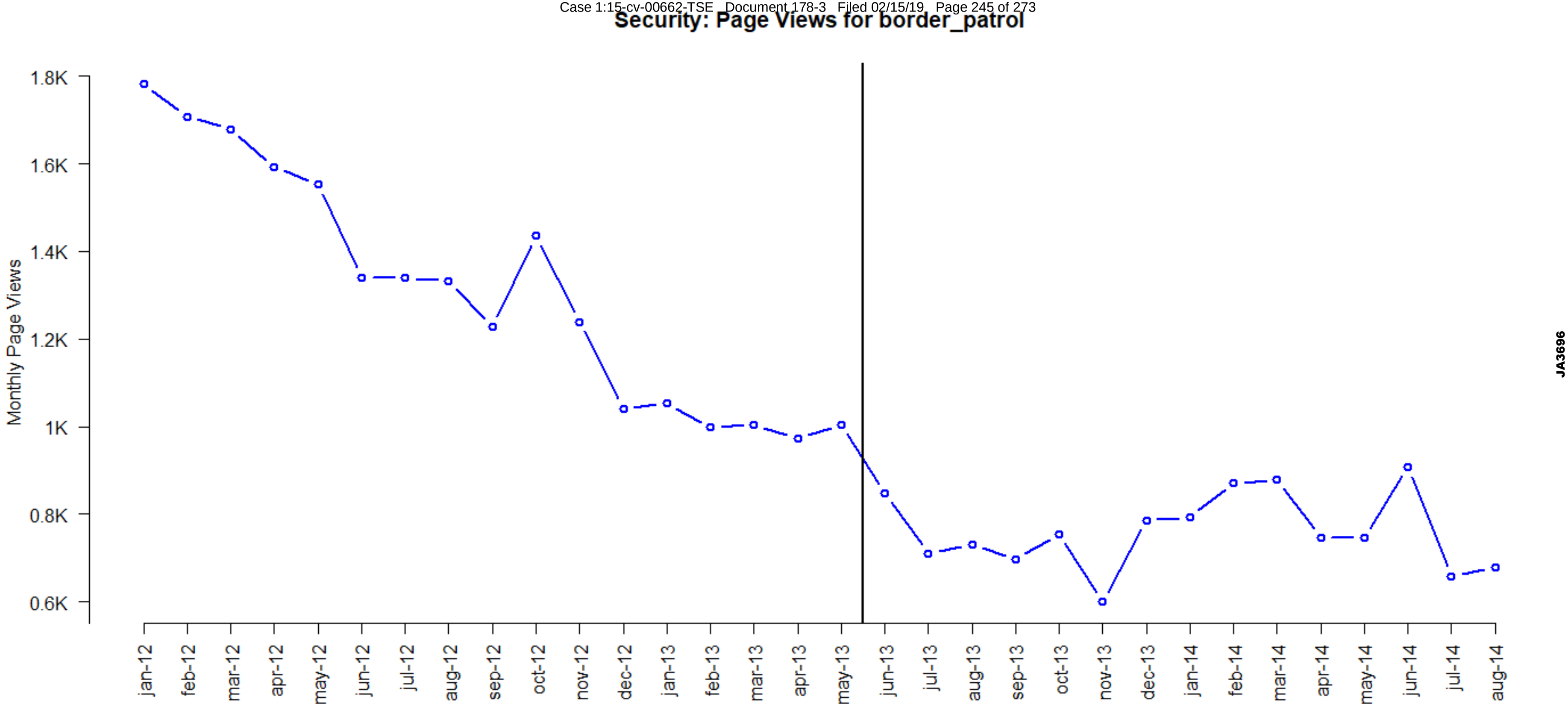

Popular: Page Views for X_fifa_world_cup
Monthly Page Views in Millions
8MM
6MM
4MM
2MM
0MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

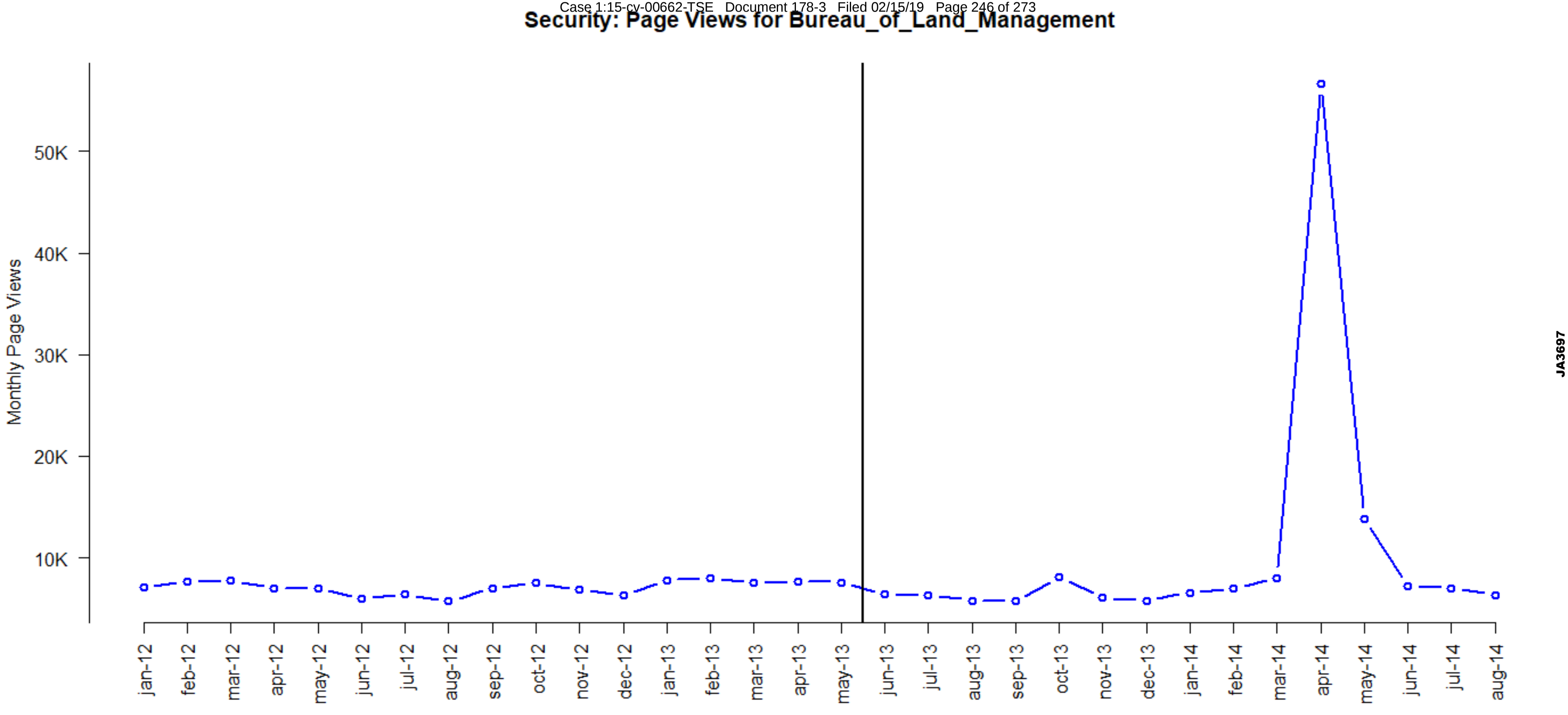

Popular: Page Views for X_phenomena
Monthly Page Views in Millions
0MM
1MM
2MM
3MM
4MM
5MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

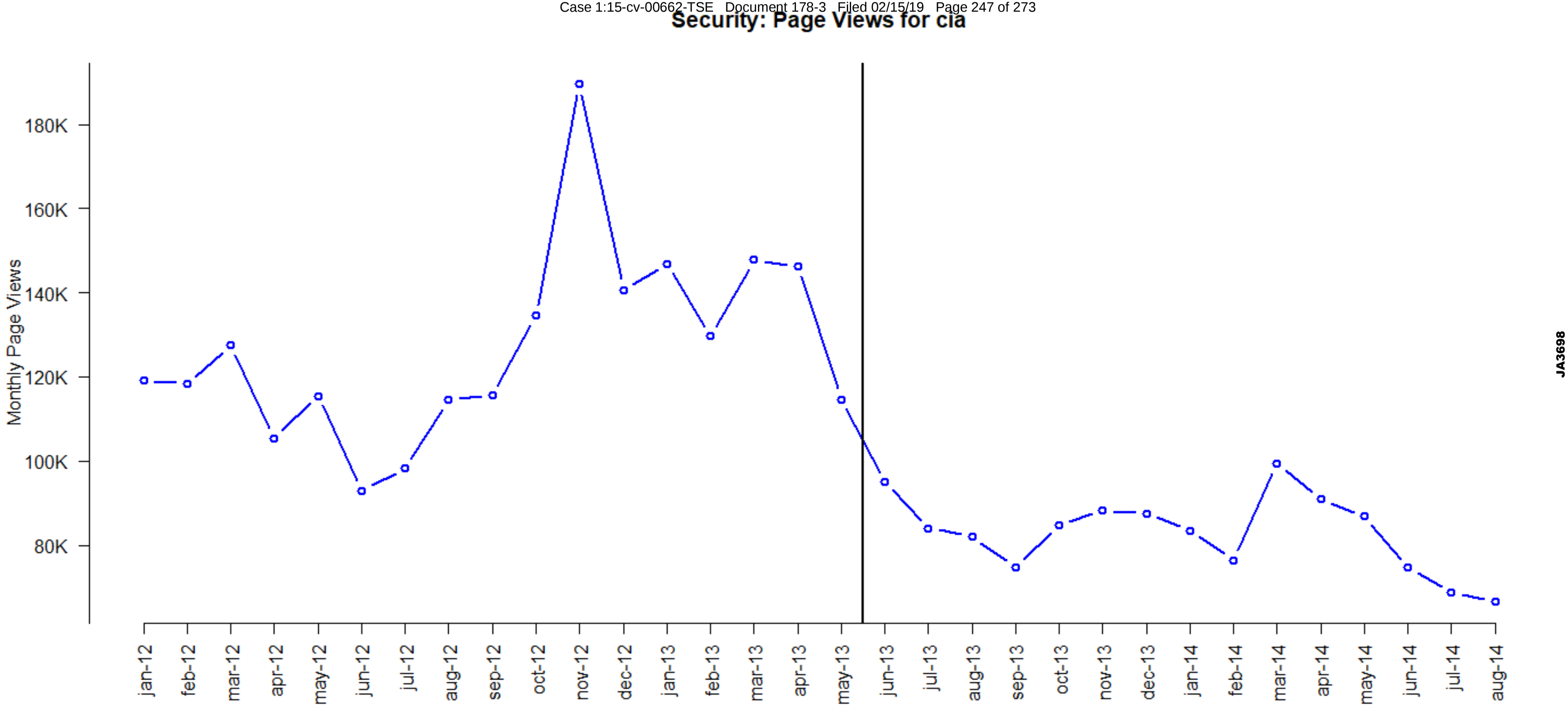

Popular: Page Views for youtube
Monthly Page Views in Millions
2.5MM
2MM
1.5MM
1MM
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

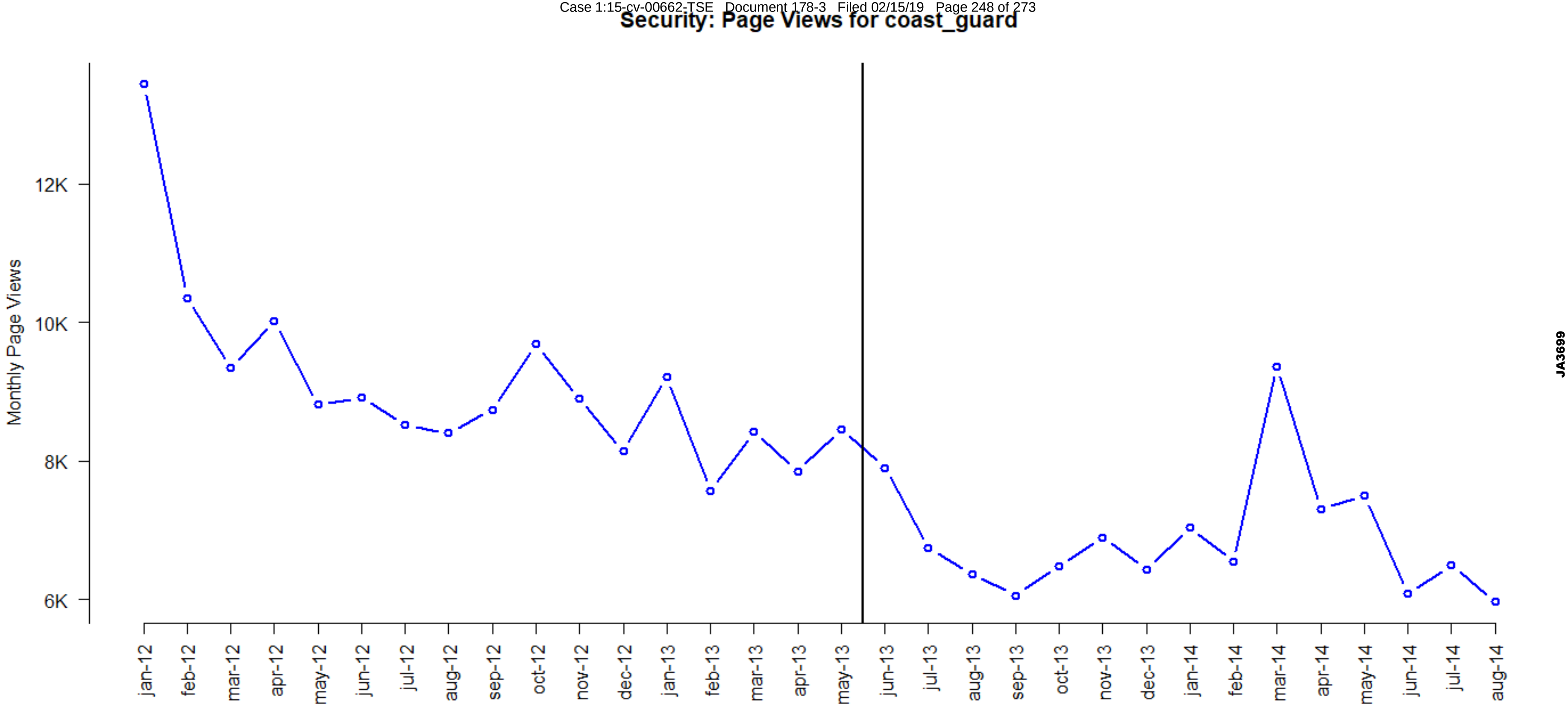

Security: Page Views for air_marshals
Monthly Page Views
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

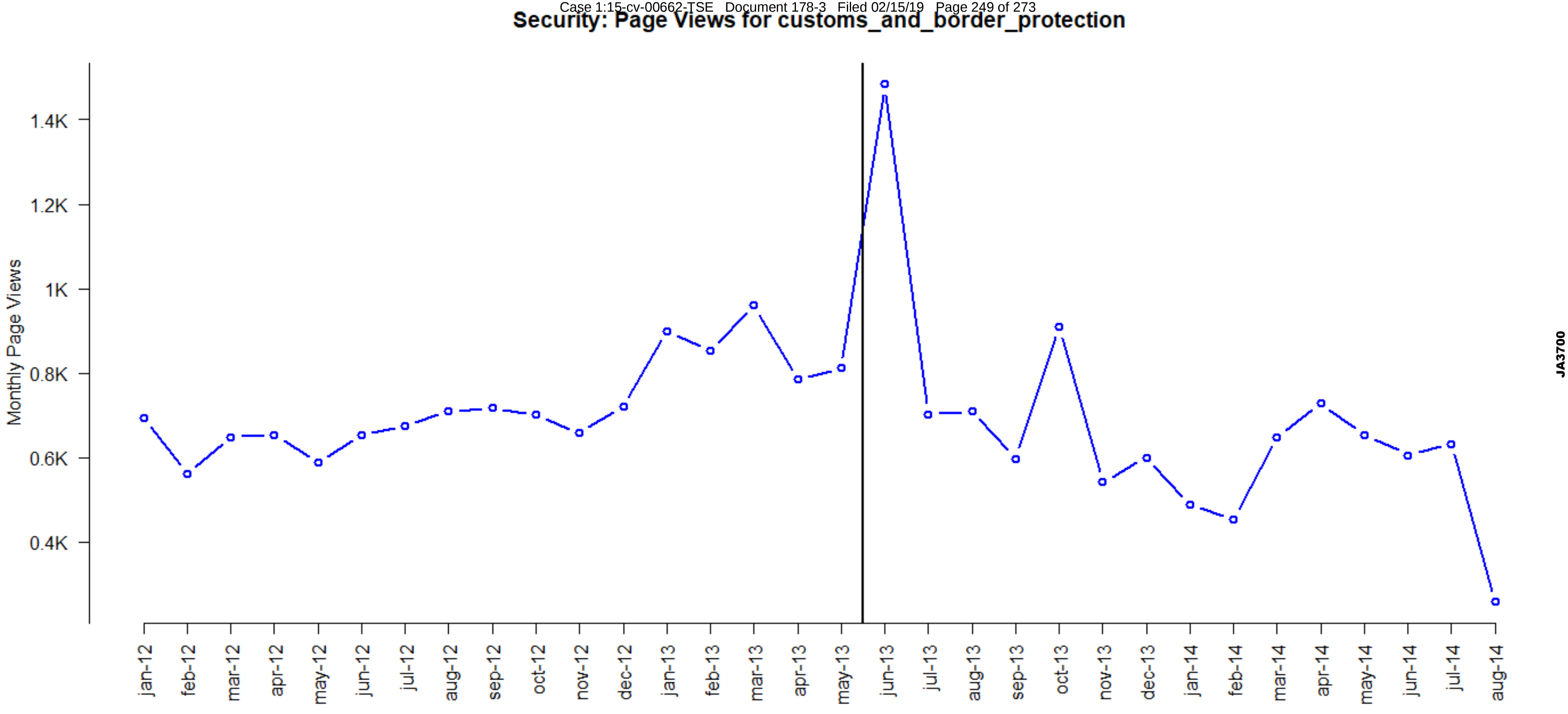

Security: Page Views for alcohol_and_tobacco_tax_and_trad
Monthly Page Views
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

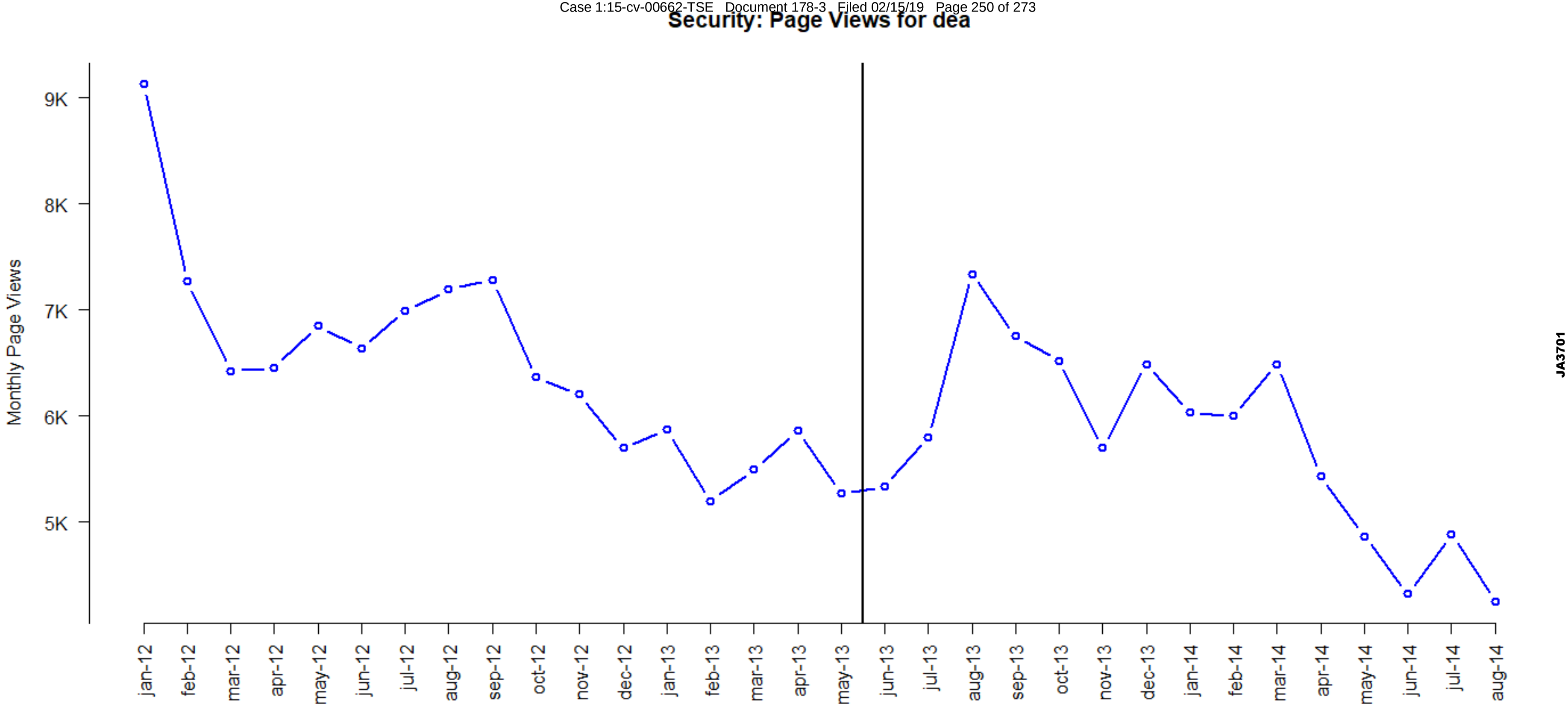

Security: Page Views for border_patrol
Monthly Page Views
1.8K
1.6K
1.4K
1.2K
1K
0.8K
0.6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

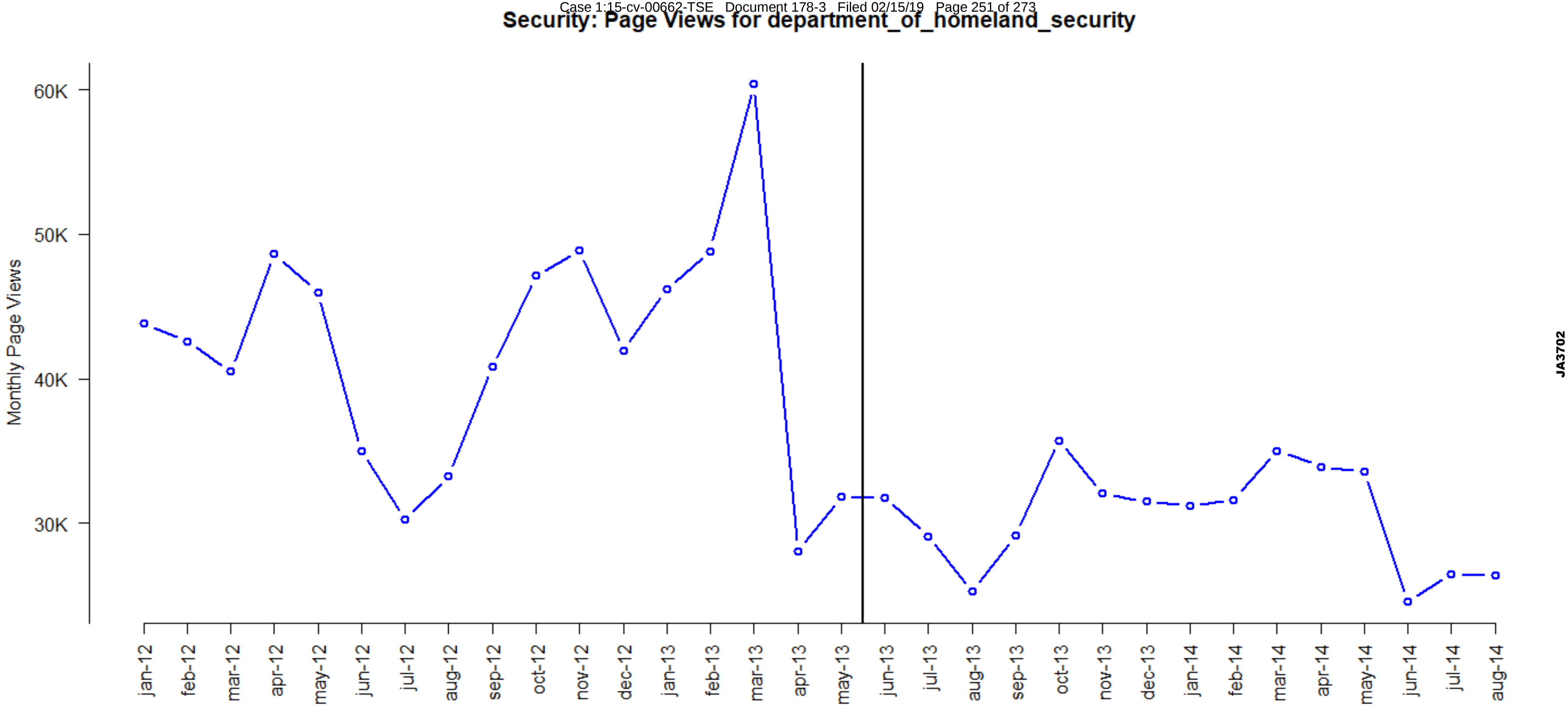

Security: Page Views for Bureau_of_Land_Management
Monthly Page Views
50K
40K
30K
20K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

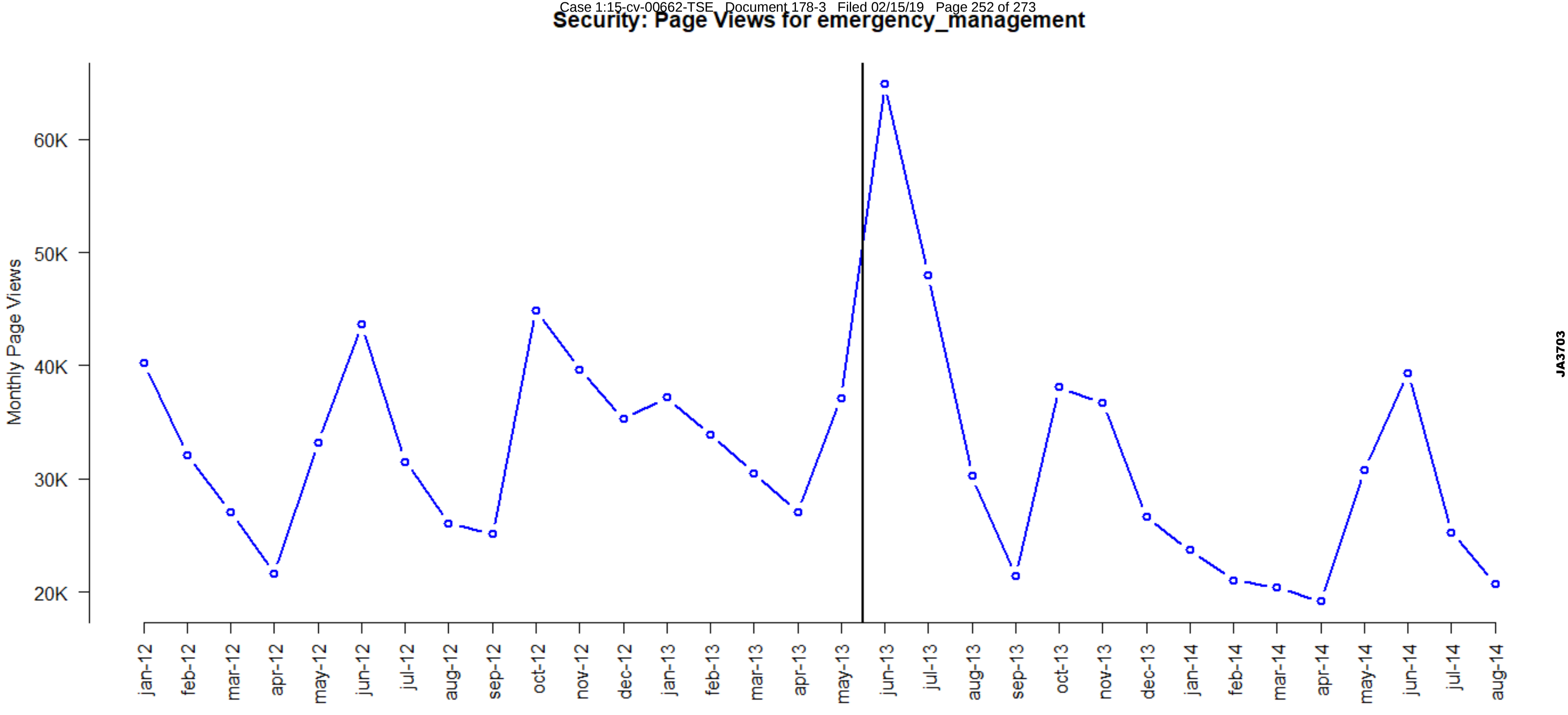

Security: Page Views for cia
Monthly Page Views
180K
160K
140K
120K
100K
80K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

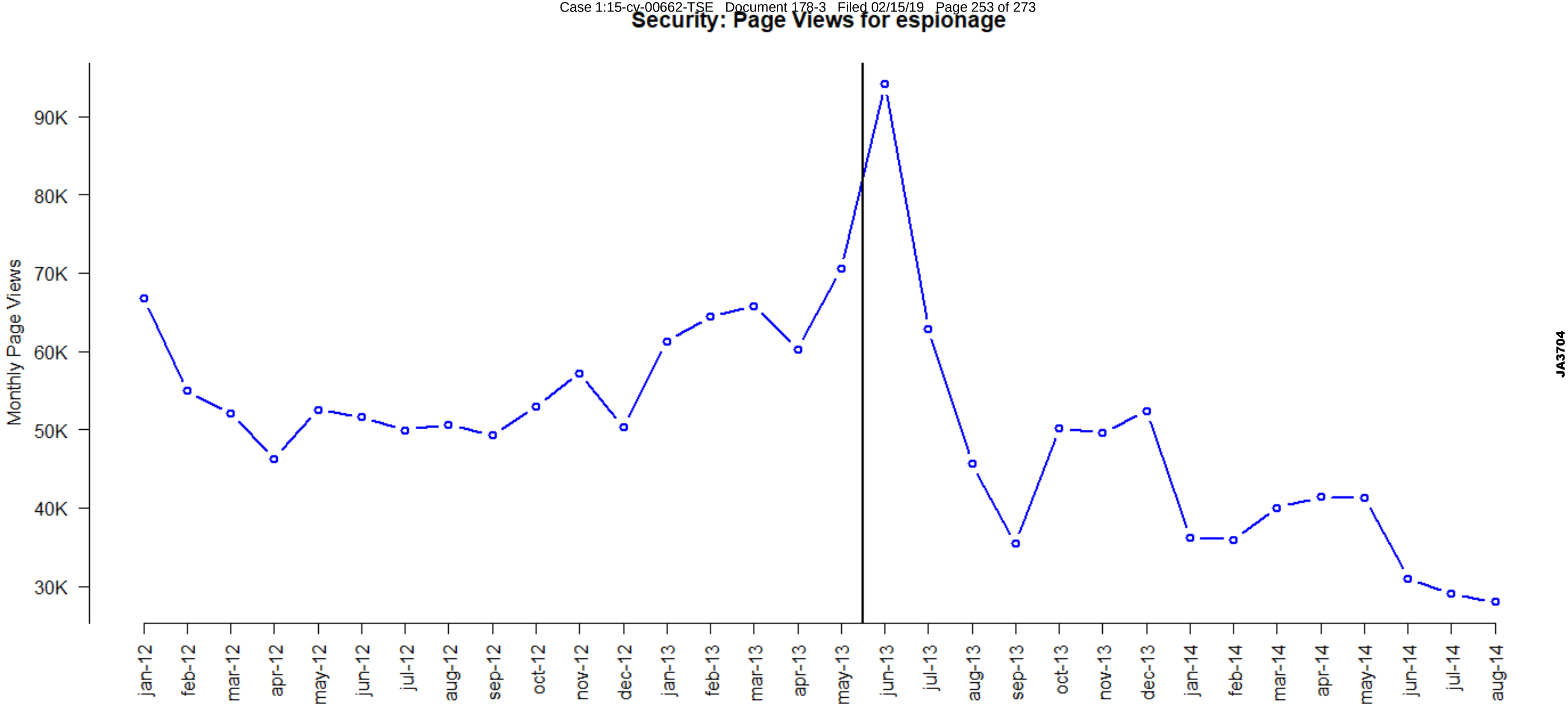

Security: Page Views for coast_guard
Monthly Page Views
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

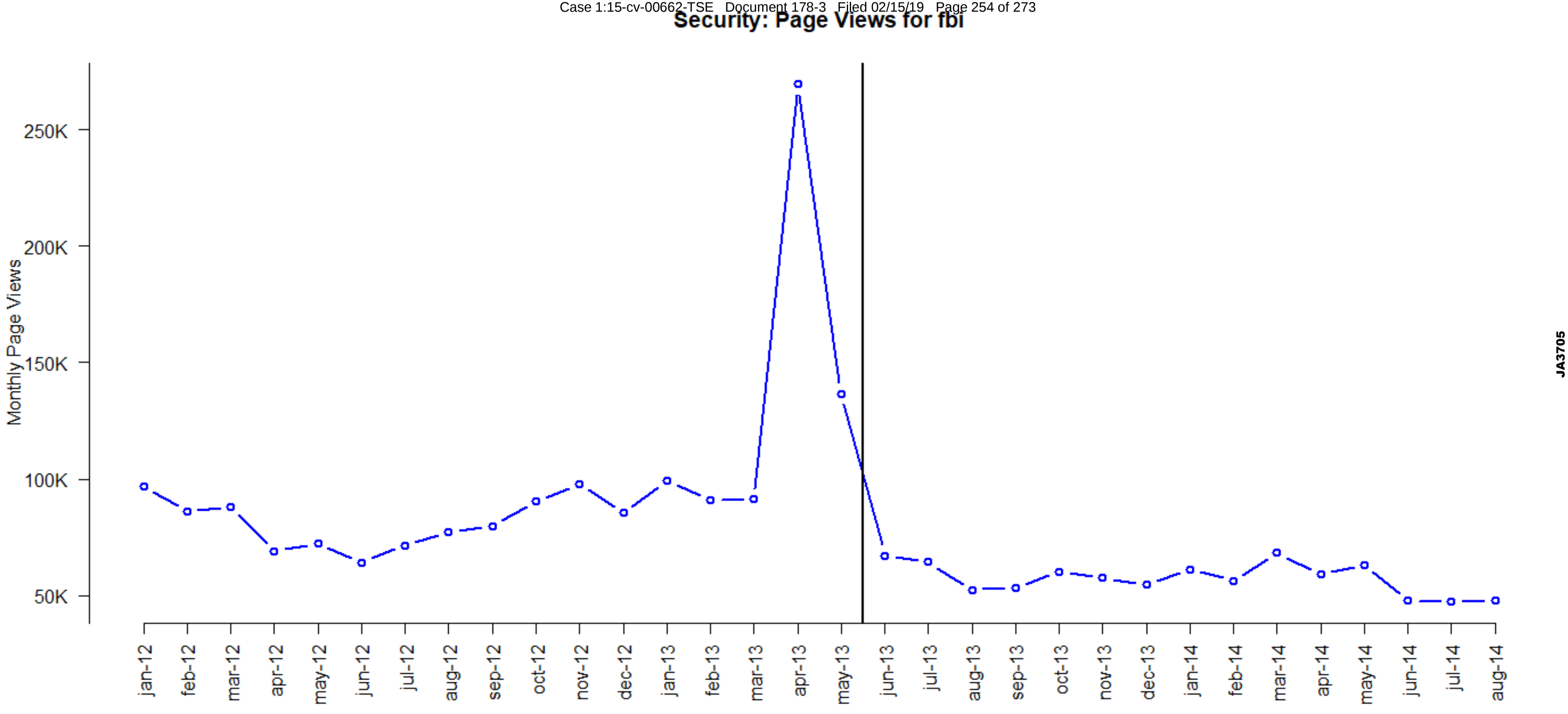

Security: Page Views for customs_and_border_protection
Monthly Page Views
1.4K
1.2K
1K
0.8K
0.6K
0.4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

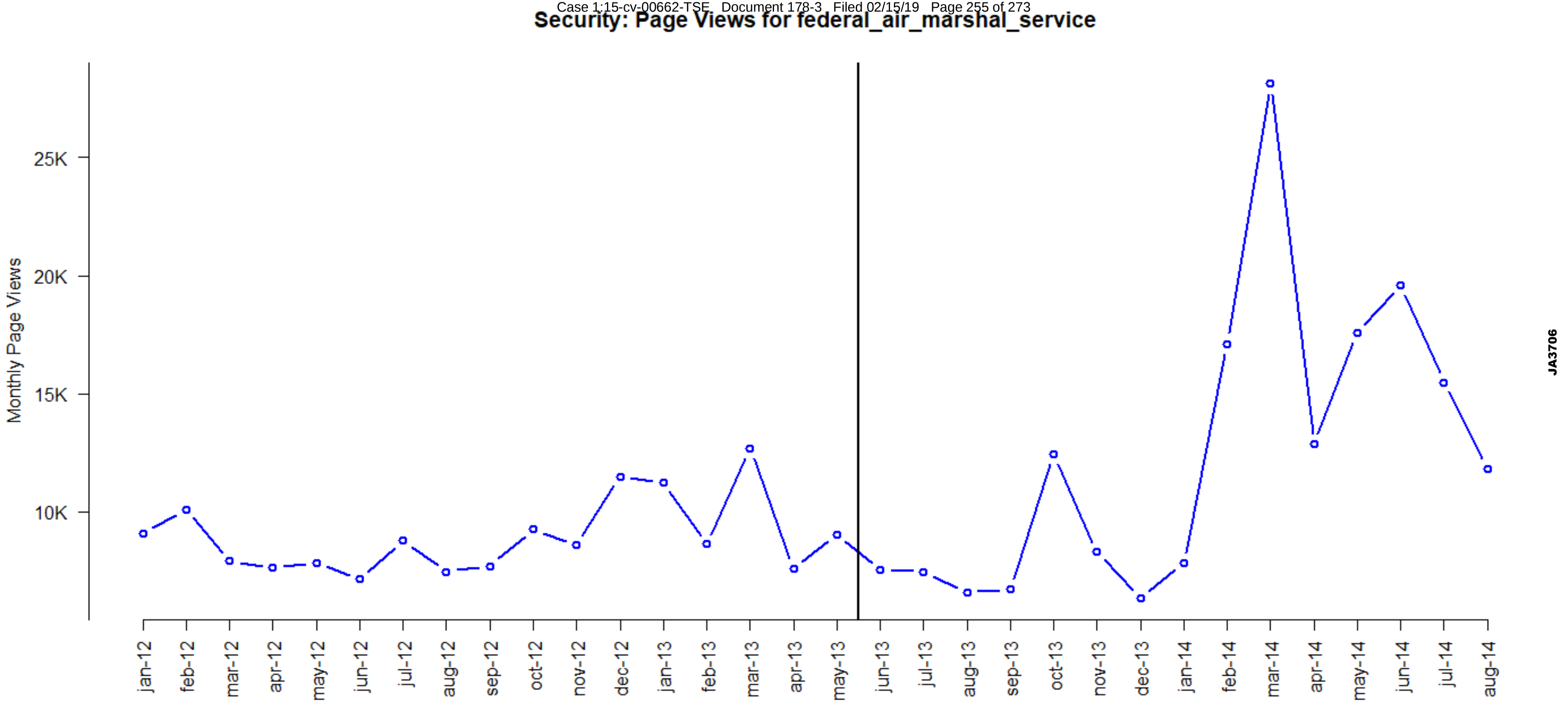

Security: Page Views for dea
Monthly Page Views
9K
8K
7K
6K
5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

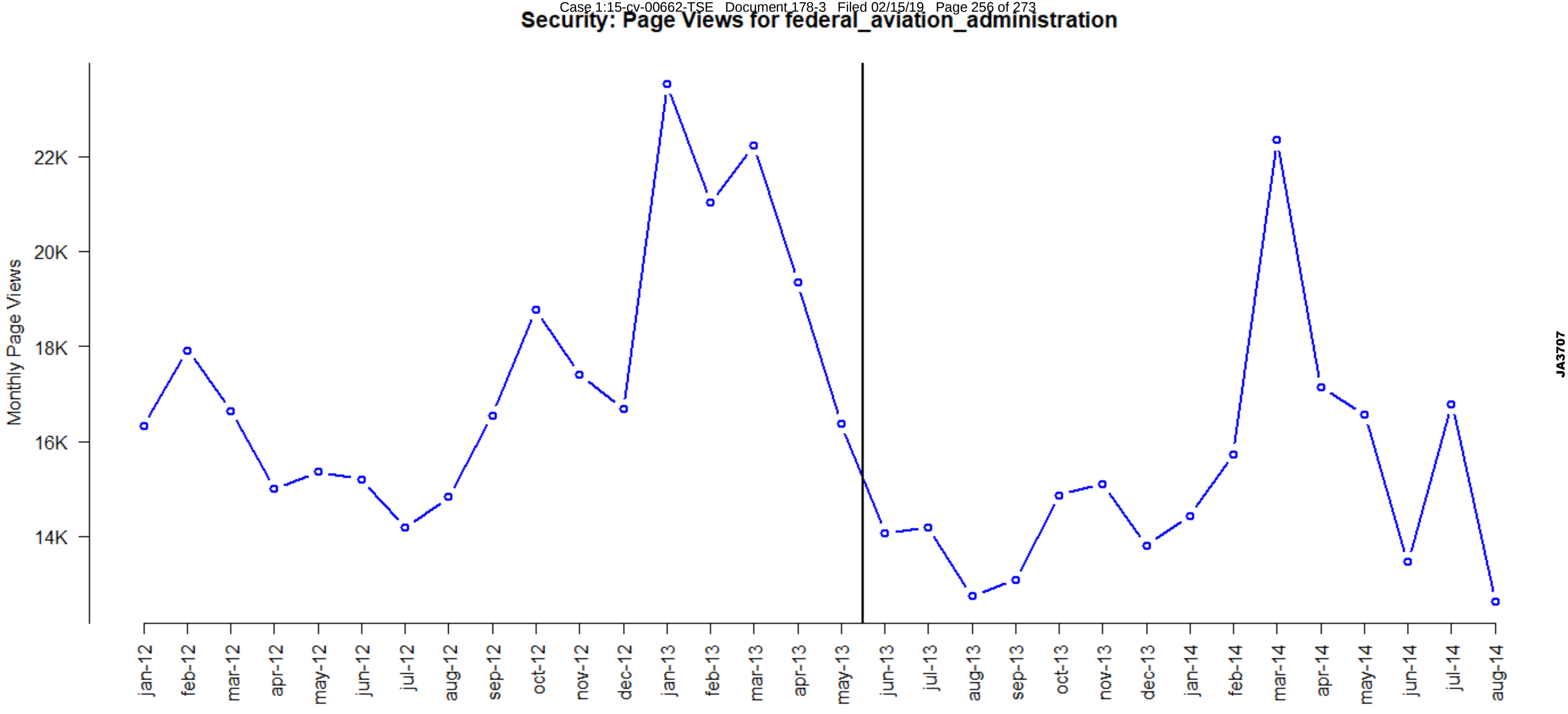

Security: Page Views for department_of_homeland_security
Monthly Page Views
60K
50K
40K
30K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

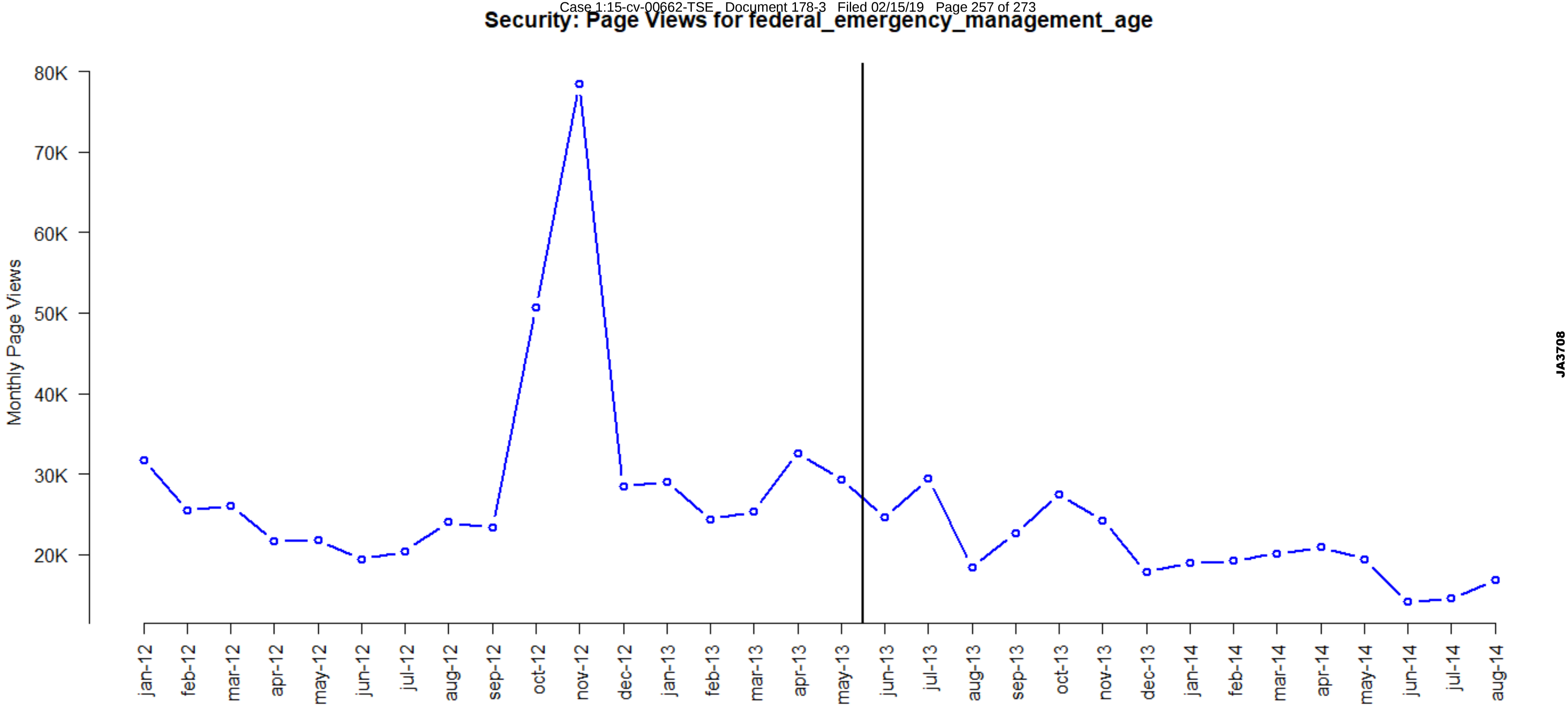

Security: Page Views for emergency_management
Monthly Page Views
60K
50K
40K
30K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

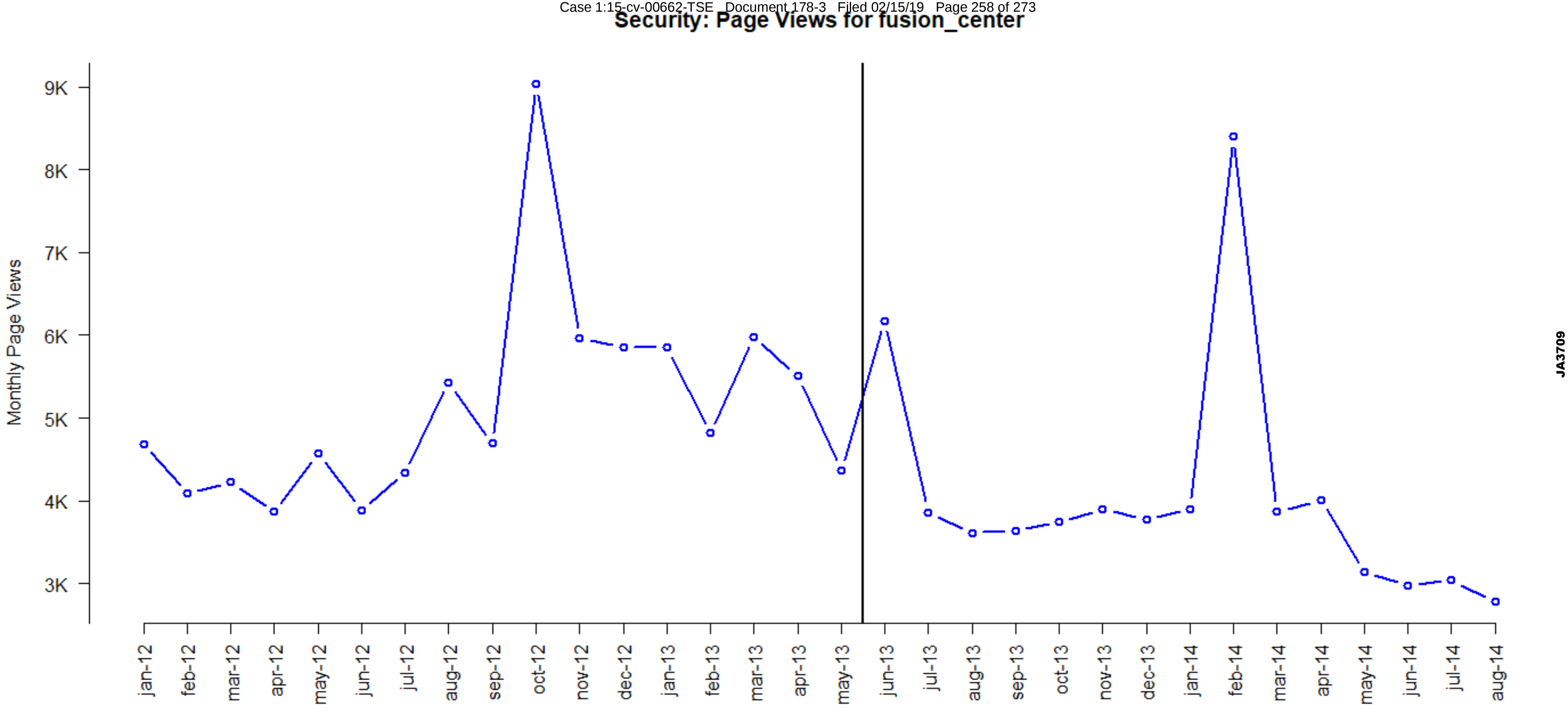

Security: Page Views for espionage
Monthly Page Views
90K
80K
70K
60K
50K
40K
30K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

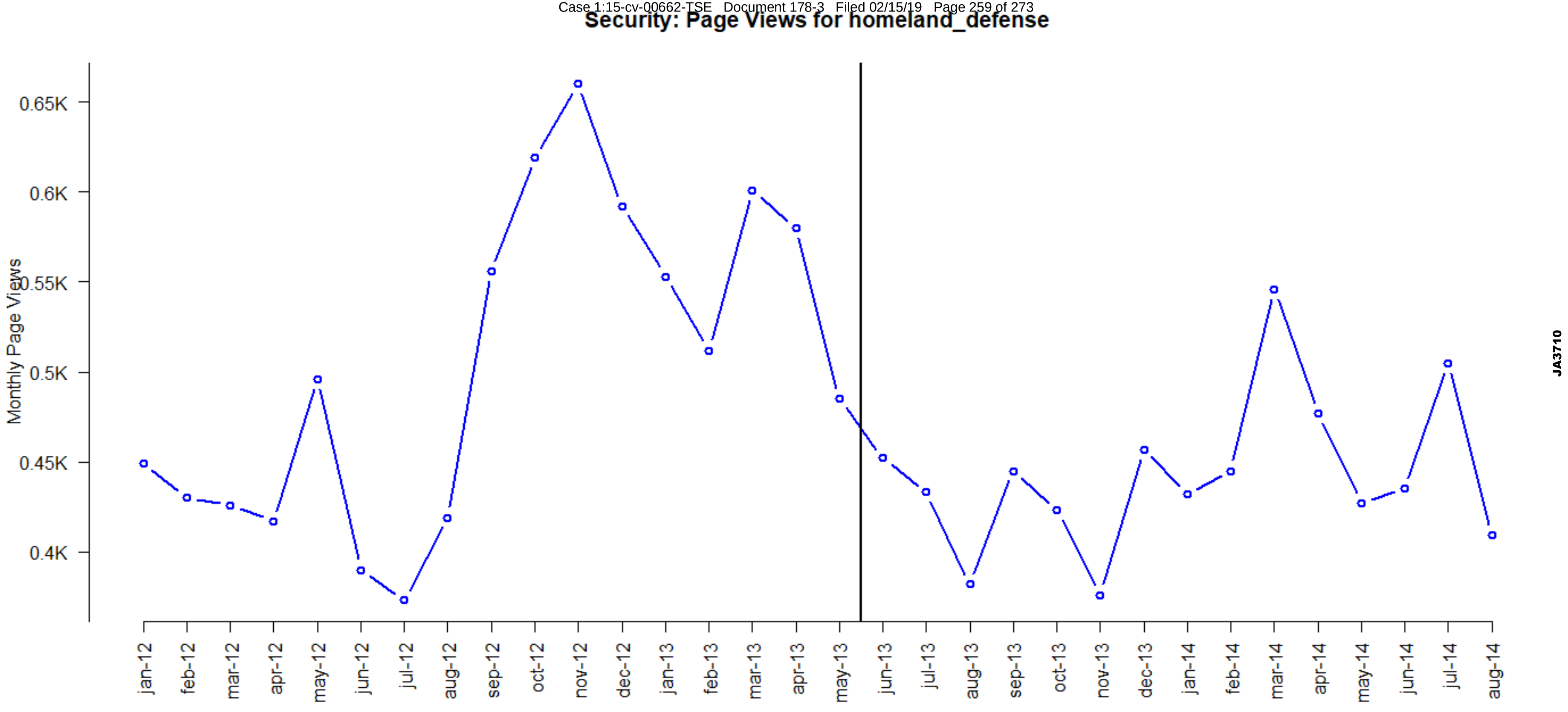

Security: Page Views for fbi
Monthly Page Views
250K
200K
150K
100K
50K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

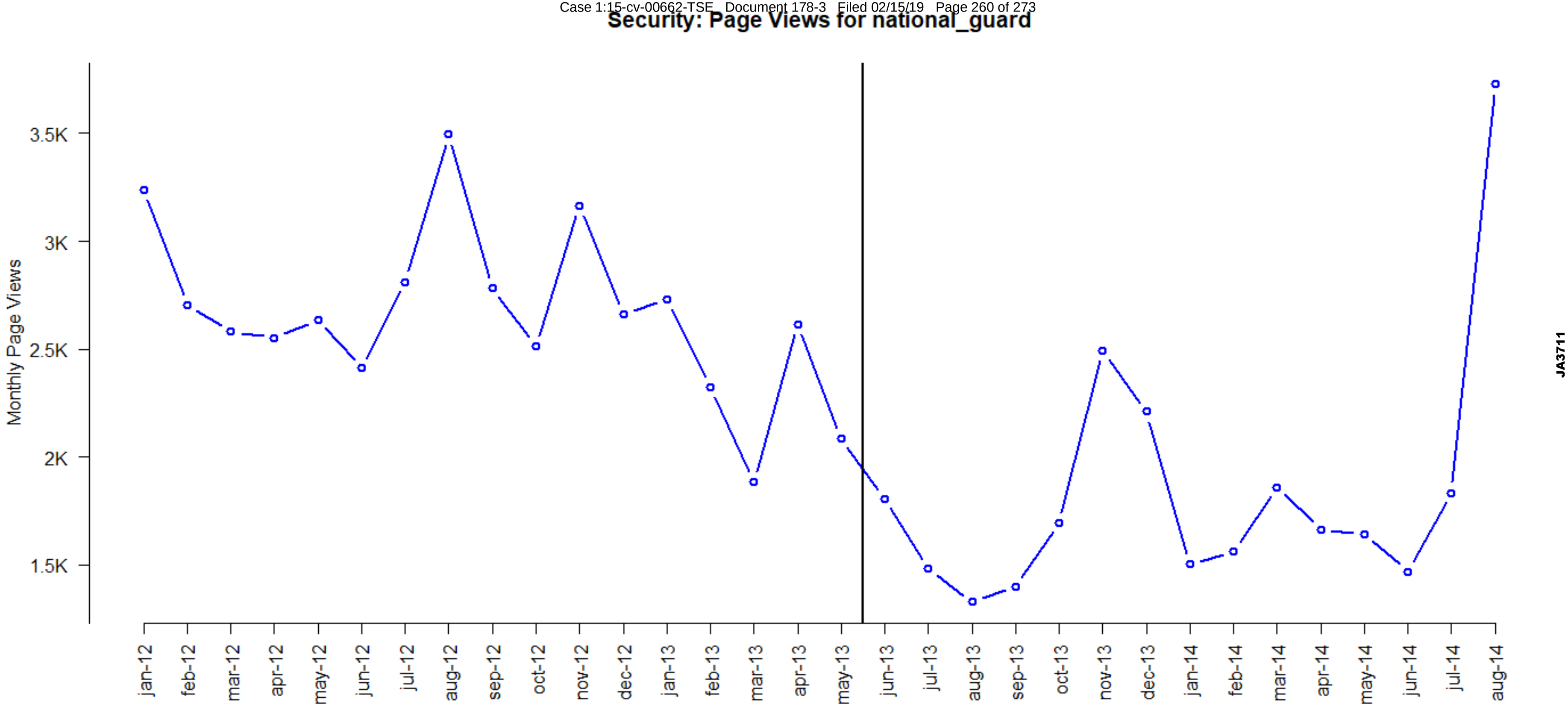

Security: Page Views for federal_air_marshal_service
Monthly Page Views
25K
20K
15K
10K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

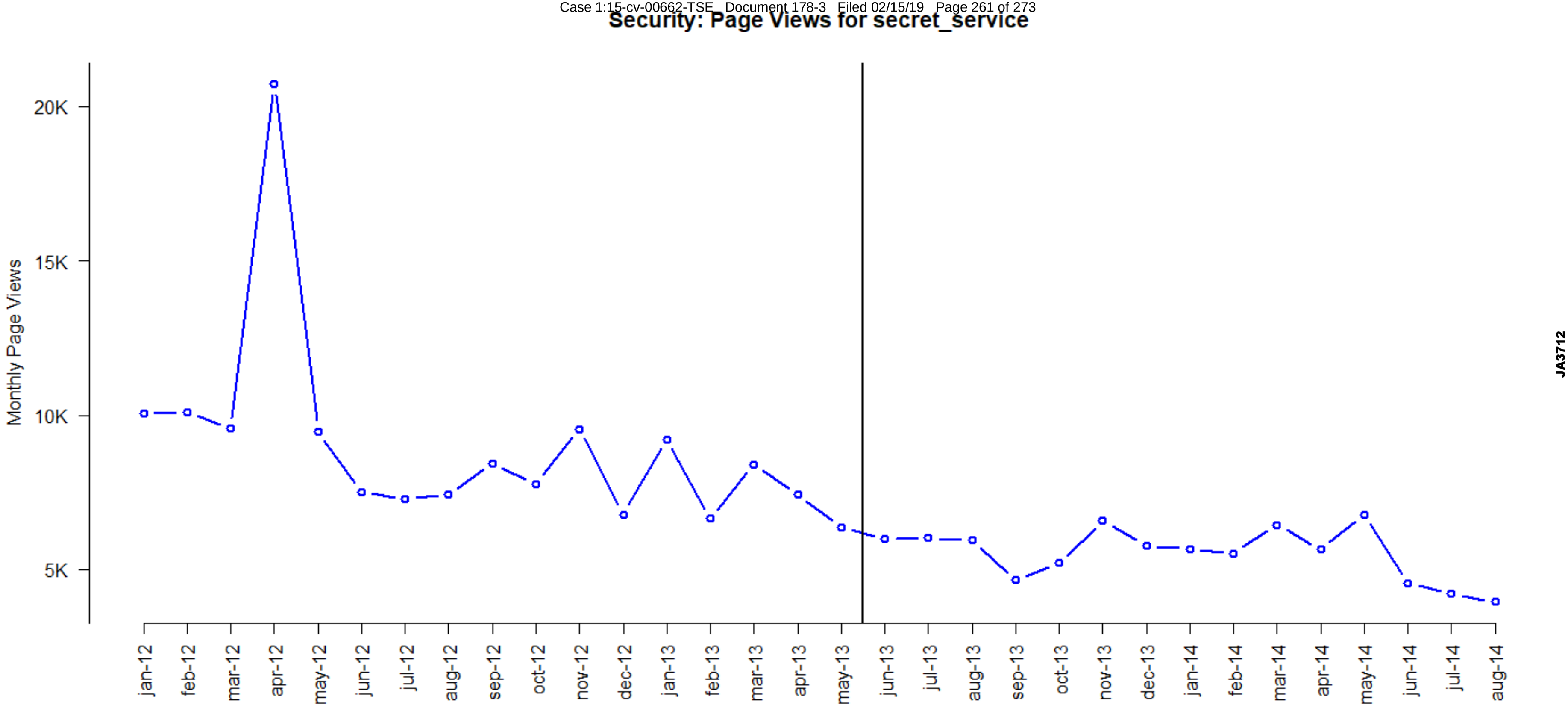

Security: Page Views for federal_aviation_administration
Monthly Page Views
22K
20K
18K
16K
14K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

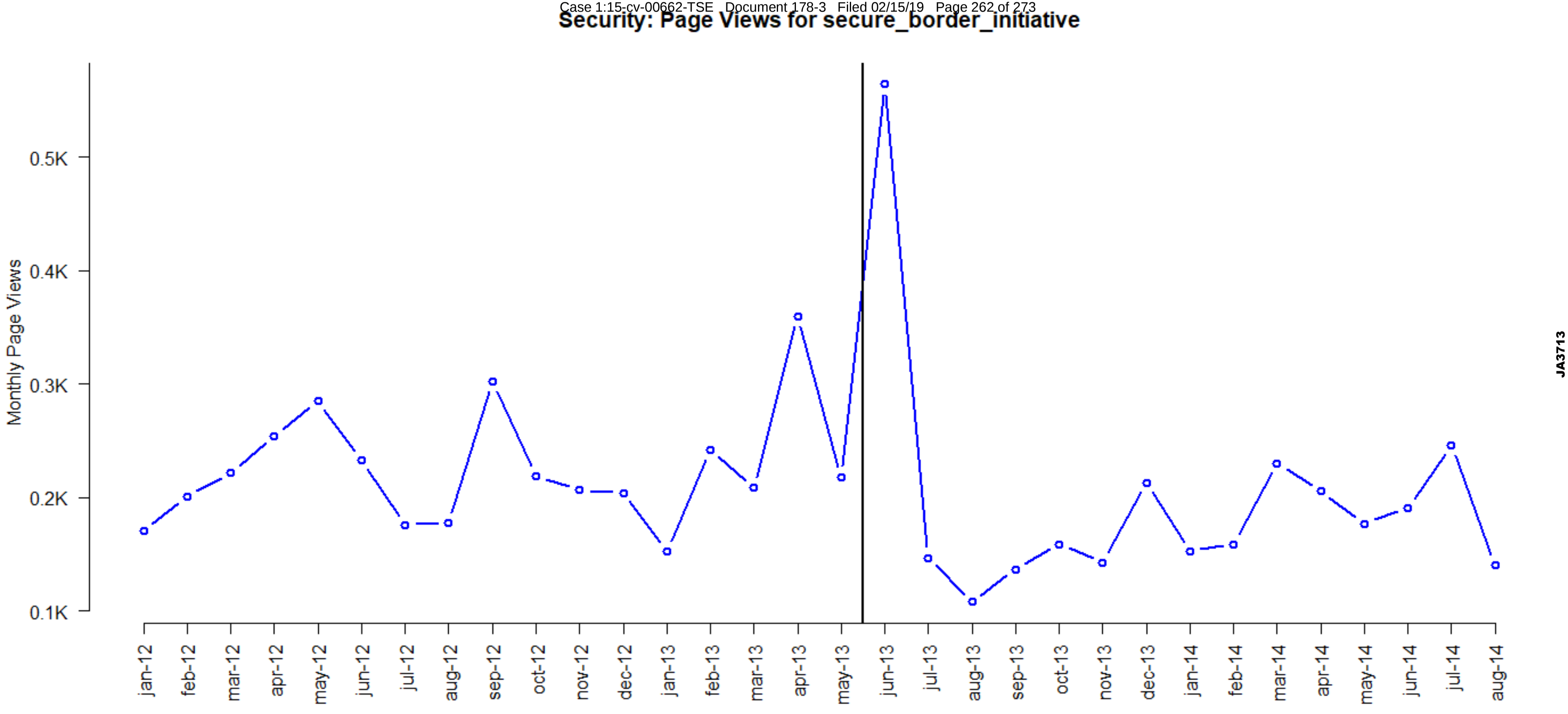

Security: Page Views for federal_emergency_management_age
Monthly Page Views
80K
70K
60K
50K
40K
30K
20K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

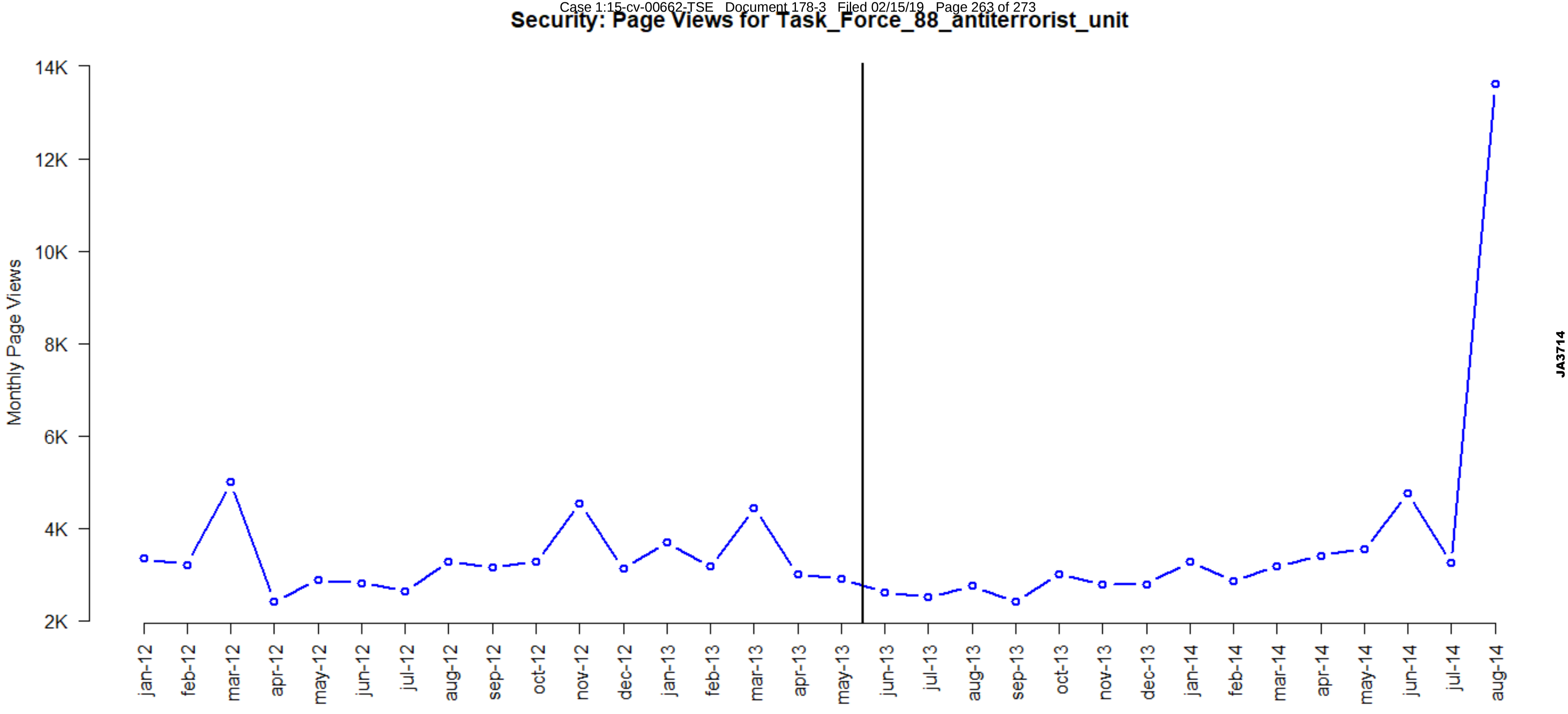

Security: Page Views for fusion_center
Monthly Page Views
9K
8K
7K
6K
5K
4K
3K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

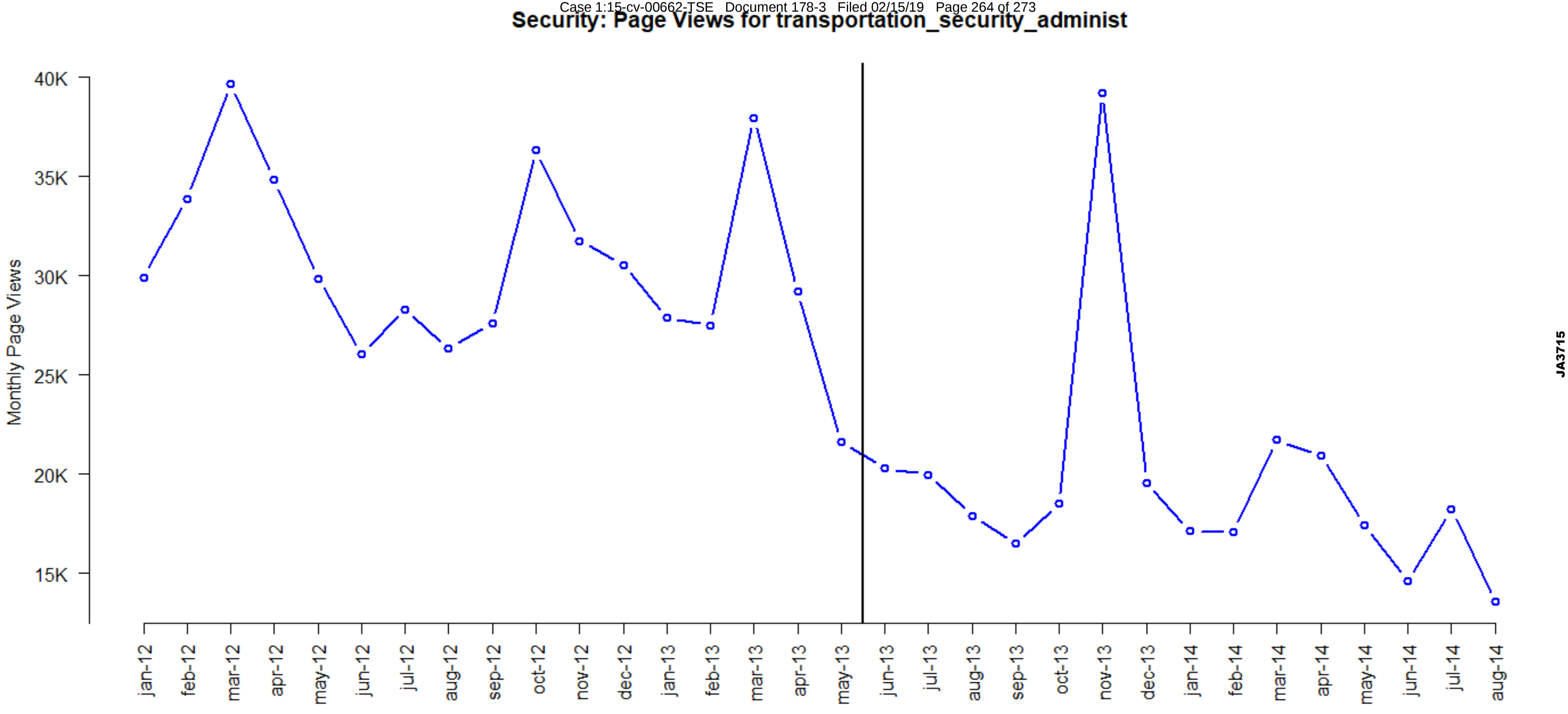

Security: Page Views for homeland_defense
Monthly Page Views
0.65K
0.6K
0.55K
0.5K
0.45K
0.4K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

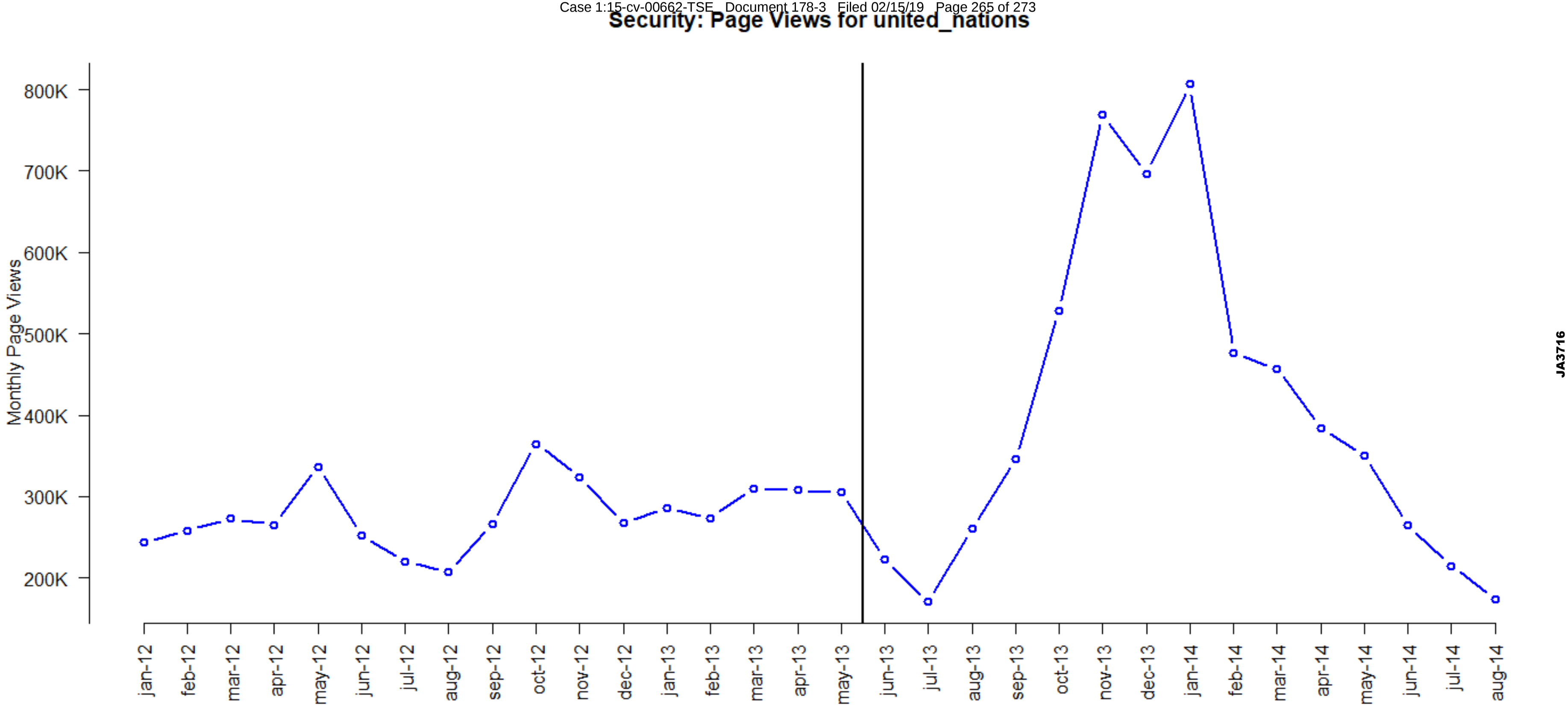

Security: Page Views for national_guard
Monthly Page Views
3.5K
3K
2.5K
2K
1.5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

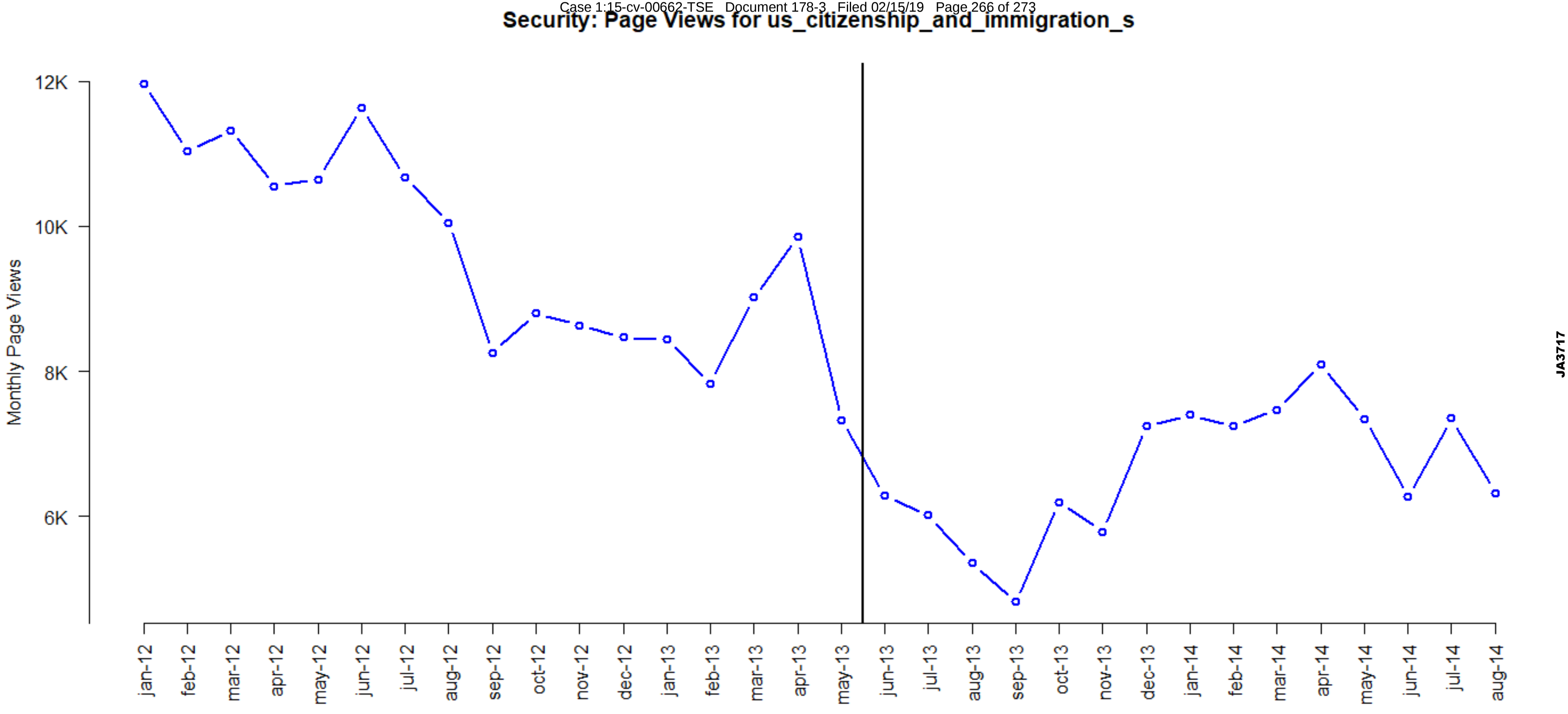

Security: Page Views for secret_service
Monthly Page Views
20K
15K
10K
5K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

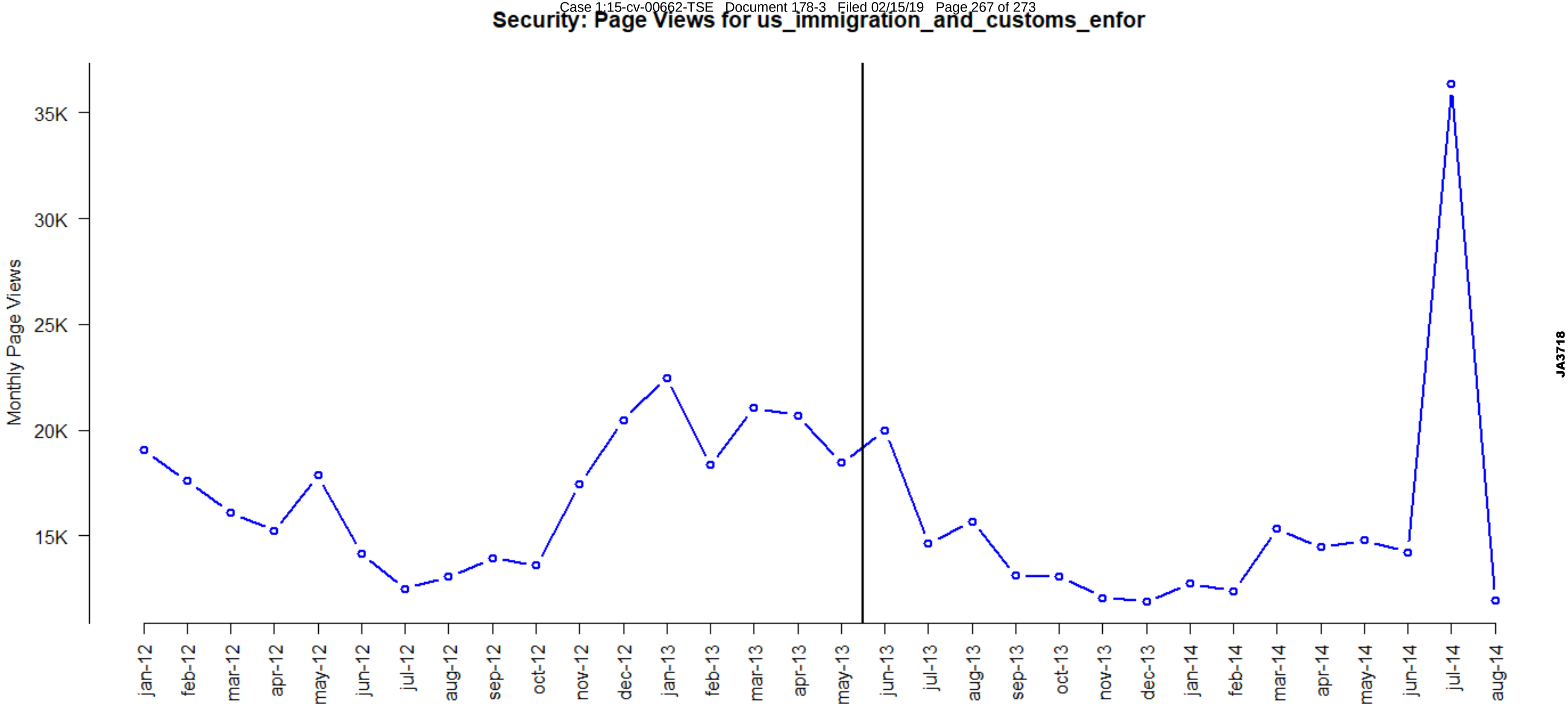

Security: Page Views for secure_border_initiative
Monthly Page Views
0.5K
0.4K
0.3K
0.2K
0.1K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Security: Page Views for Task_Force_88_antiterrorist_unit
Monthly Page Views
14K
12K
10K
8K
6K
4K
2K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

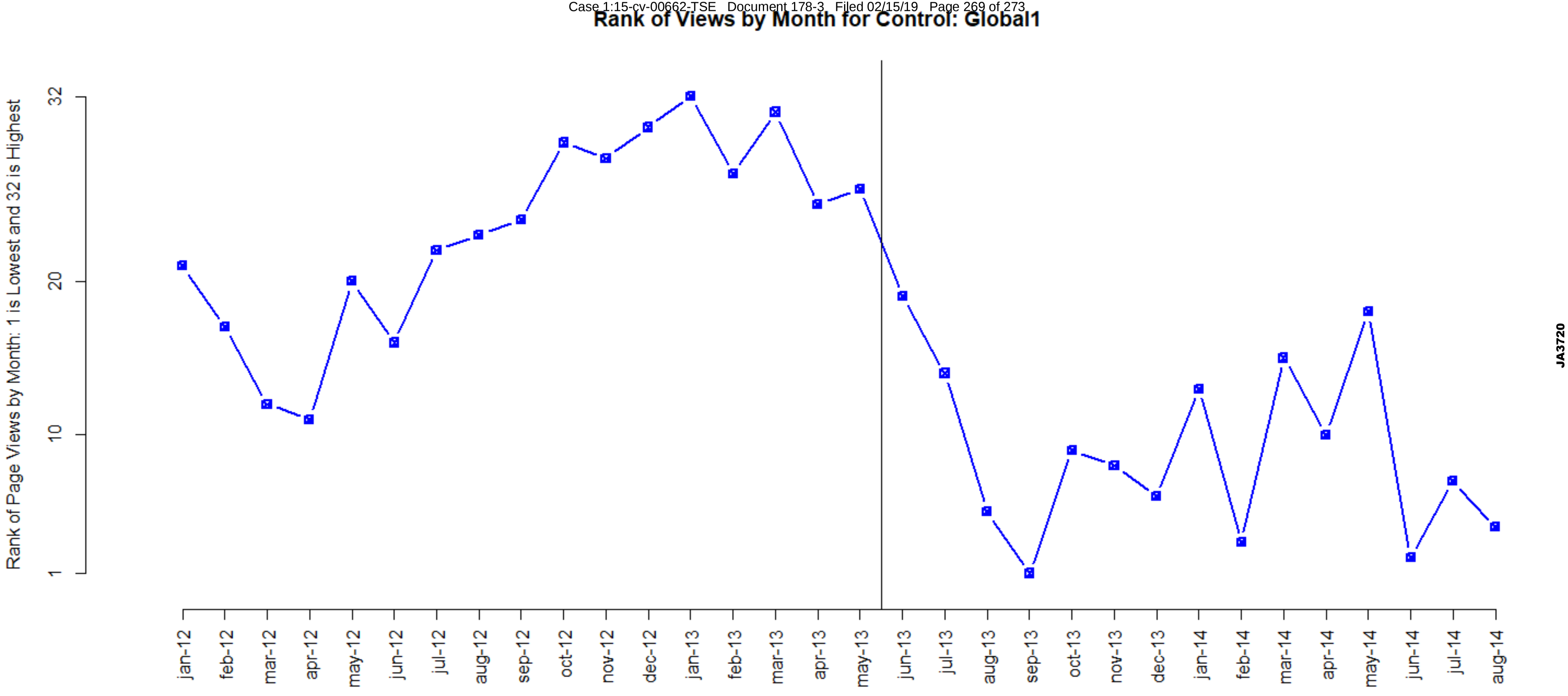

Security: Page Views for transportation_security_administ
Monthly Page Views
40K
35K
30K
25K
20K
15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

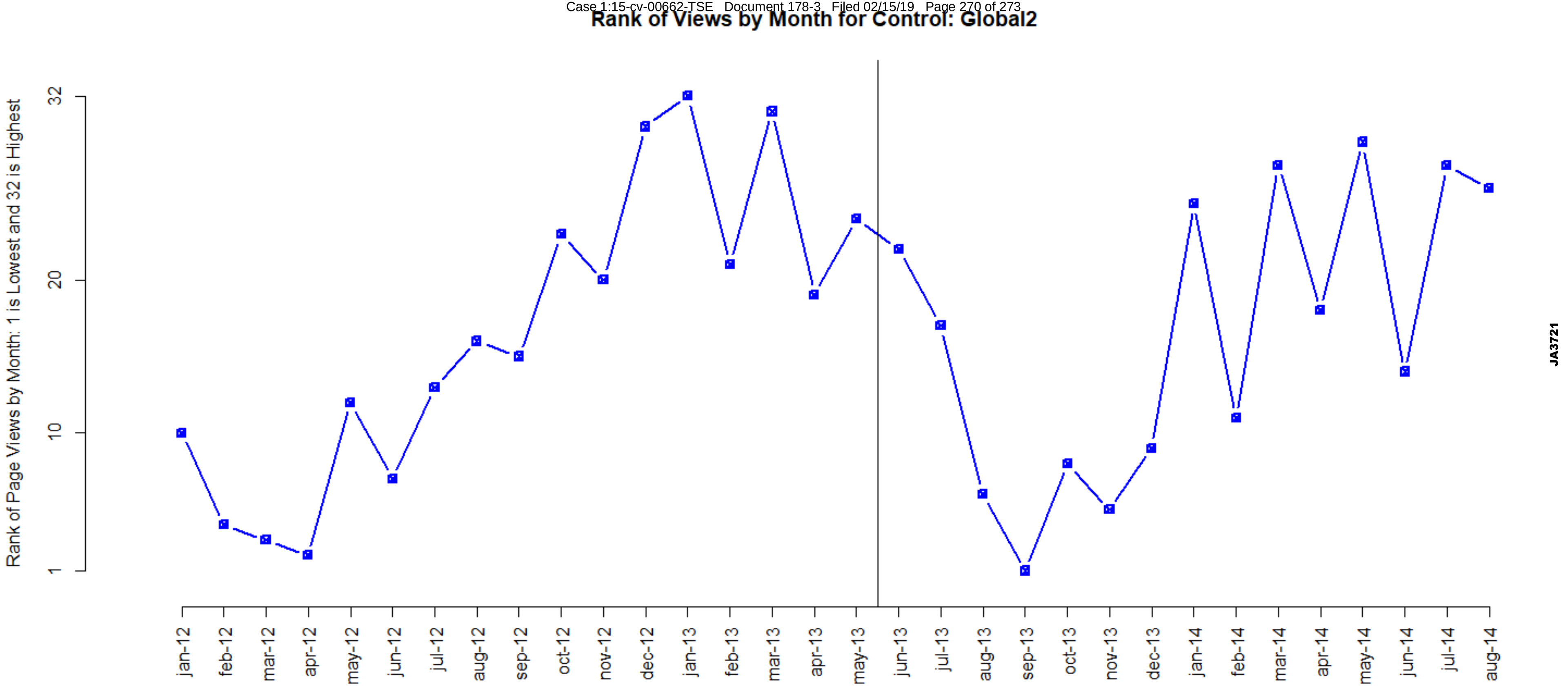

Security: Page Views for united_nations
Monthly Page Views
800K
700K
600K
500K
400K
300K
200K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

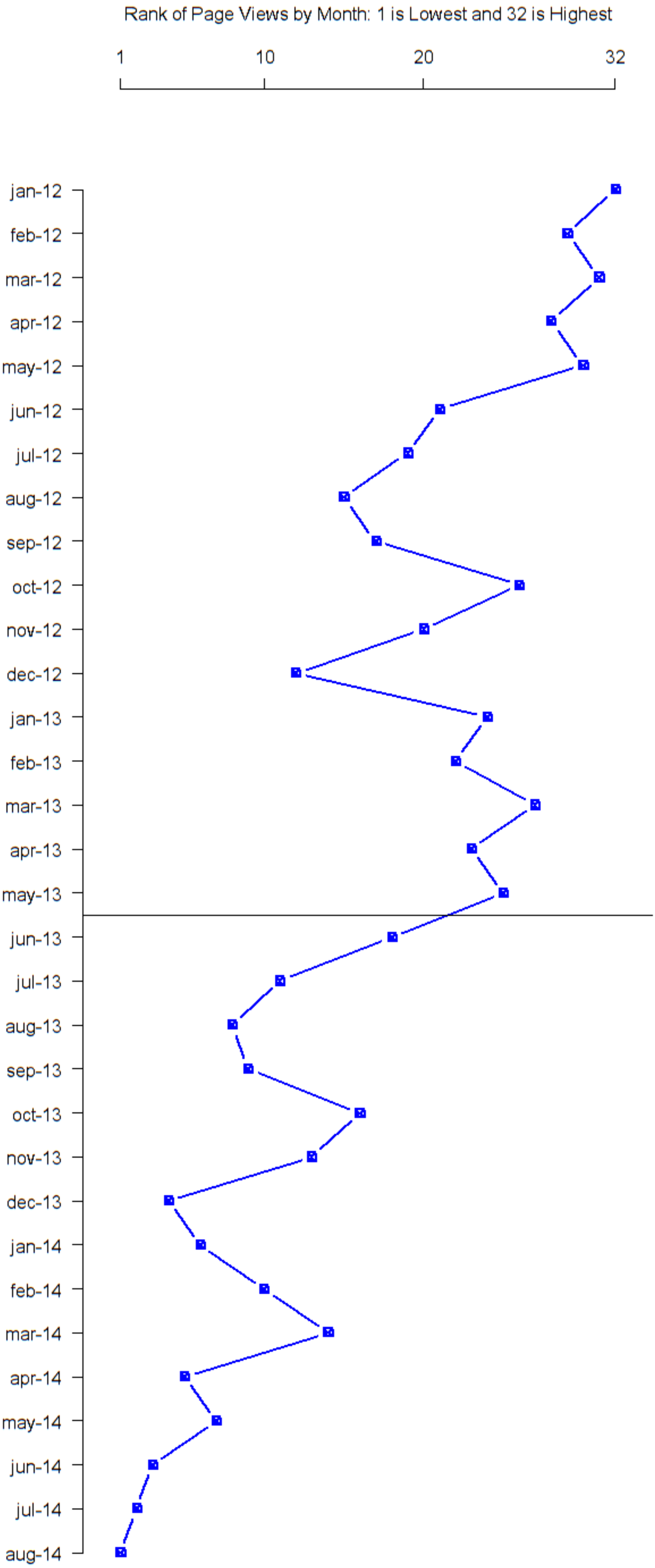

Security: Page Views for us_citizenship_and_immigration_s
Monthly Page Views
12K
10K
8K
6K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Security: Page Views for us_immigration_and_customs_enfor
Monthly Page Views
35K
30K
25K
20K
15K
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14

APPENDIX VII: Page Views for Five Aggregate Comparison Datasets

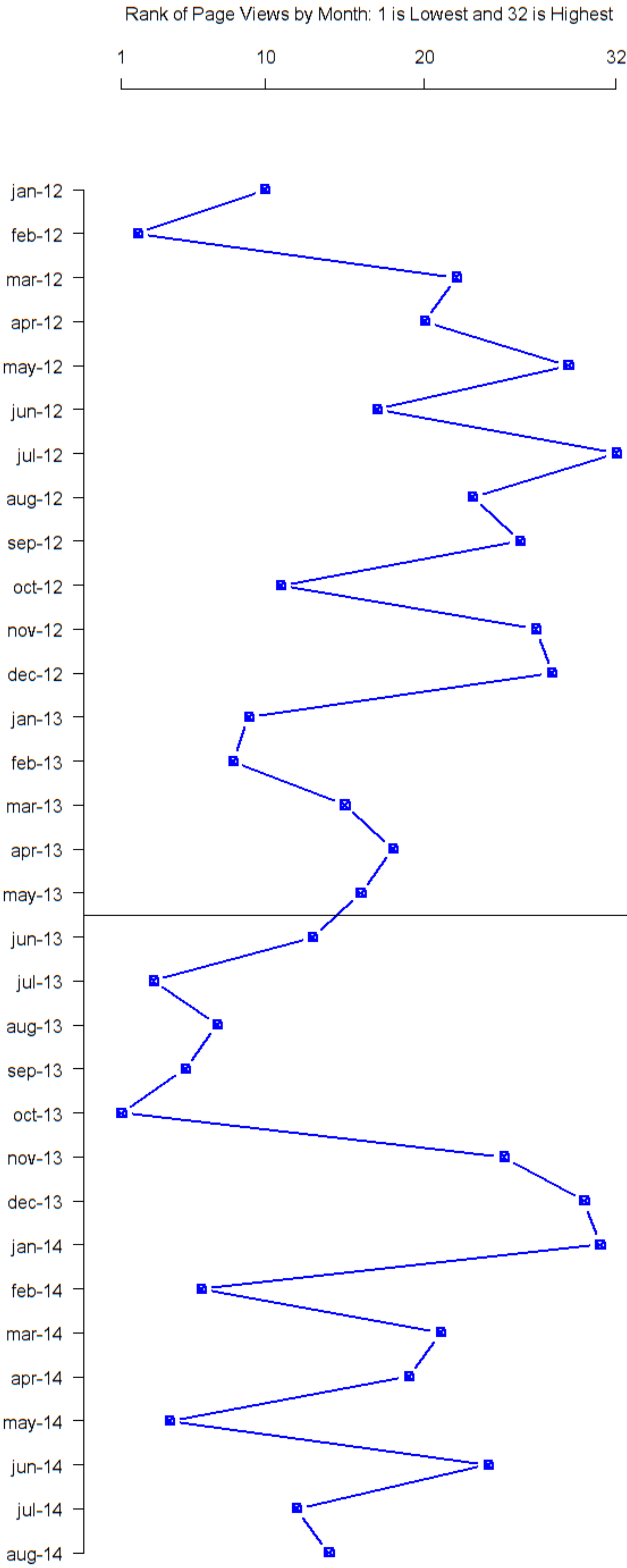

Rank of Views by Month for Control: Global1
Rank of Page Views by Month: 1 is Lowest and 32 is Highest
1
10
20
32
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14


Rank of Views by Month for Control: Global2
Rank of Page Views by Month: 1 is Lowest and 32 is Highest
1
10
20
32
jan-12
feb-12
mar-12
apr-12
may-12
jun-12
jul-12
aug-12
sep-12
oct-12
nov-12
dec-12
jan-13
feb-13
mar-13
apr-13
may-13
jun-13
jul-13
aug-13
sep-13
oct-13
nov-13
dec-13
jan-14
feb-14
mar-14
apr-14
may-14
jun-14
jul-14
aug-14



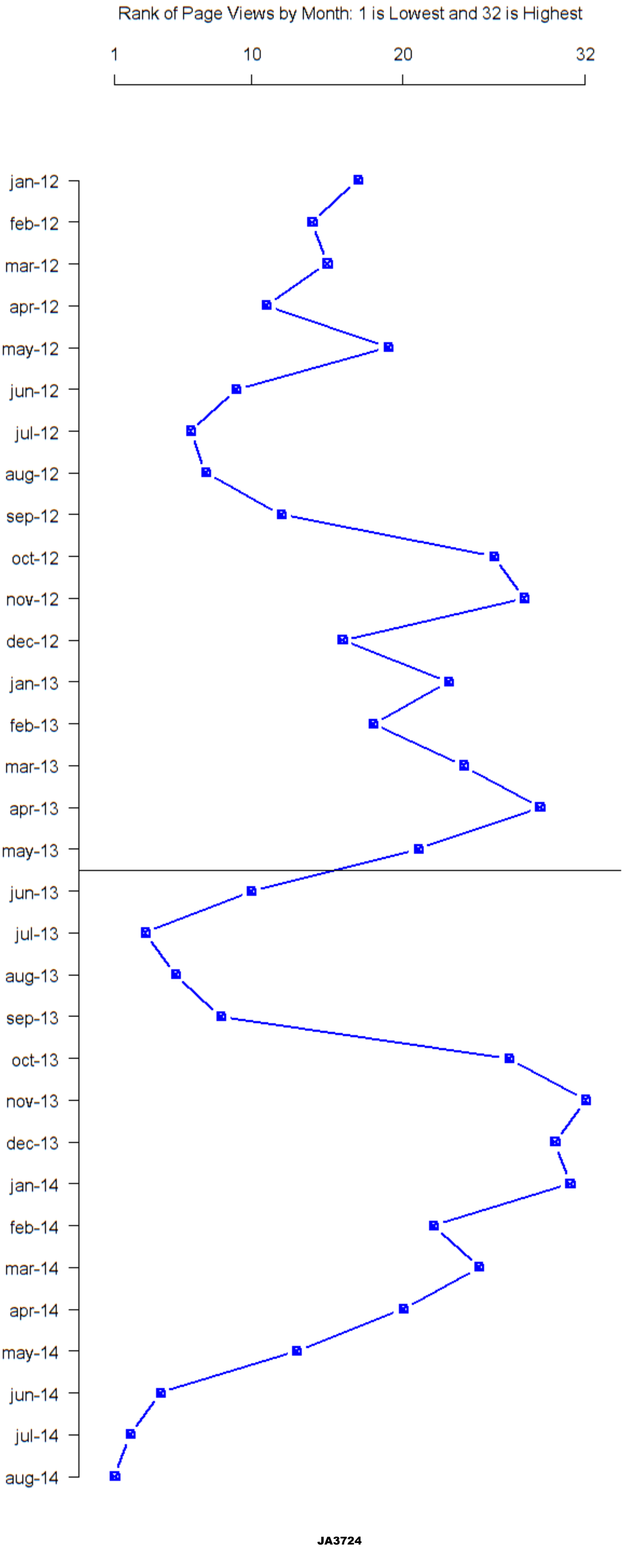
Rank of Views by Month for Control: Infrastructure

Case 1:15-cv-00662-TSE Document 178-3 Filed 02/15/19 Page 271 of 273
USCA4 Appeal: 20-1191 Doc: 18-6 Filed: 07/01/2020 Pg: 328 of 484
Total Pages:(3794 of 4208)



Rank of Views by Month for Control: Popular

USCA4 Appeal: 20-1191 Doc: 18-6 Filed: 07/01/2020 Pg: 329 of 484
Case 1:15-cv-00662-TSE Document 178-3 Filed 02/15/19 Page 272 of 273
Total Pages:(3795 of 4208)



USCA4 Appeal: 20-1191 Doc: 18-6 Filed: 07/01/2020 Pg: 330 of 484
Case 1:15-cv-00662-TSE Document 178-3 Filed 02/15/19 Page 273 of 273 PageID# 4208
Rank of Views by Month for Control: Security
Total Pages:(3796 of 4208)

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 8

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	No. 1:15-cv-0662 (TSE)

SECOND DECLARATION OF JAMES J. GILLIGAN

Pursuant to 28 U.S.C. § 1746, I, James J. Gilligan, hereby declare:

1. I am Special Litigation Counsel in the Civil Division, Federal Programs Branch, of the United States Department of Justice. I serve as lead counsel for the Defendants in the above-captioned case. The statements made herein are based on my personal knowledge, and on information made available to me in the course of my duties and responsibilities as Government counsel in this case.

2. I submit this declaration in support of the Defendants' concurrently filed reply in support of their motion for summary judgment.

3. Filed herewith as Defendants' Exhibits 9-15 are true and correct copies of the following documents:[1]

Exhibit No.	Exhibit Name
9	Wikimedia Foundation, Inc.'s Responses and Objections to DOJ's First Set of Interrogatories, dated January 11, 2018

[1] Defendants' Exhibits 1-5 were filed together with the Brief in Support of Defendants' Motion for Summary Judgment, ECF No. 161 (public versions available at ECF No. 166). Defendants' Exhibits 6-7 are separately filed along with this declaration.

10	Relevant Portions of the Deposition of James Alexander, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6)
11	Relevant Portions of the Deposition of Michelle Paulson, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6)
12	Wikimedia Foundation, *Securing access to Wikimedia sites with HTTPS*, June 12, 2015 (WIKI0007108-7114)
13	Wikipedia: Village pump (technical)/Archive 138 (WIKI0006872-6938)
14	Jimmy Wales and Lila Tretikov, "Stop Spying on Wikimedia Users" (N.Y. Times, March 10, 2015), available at https://www.nytimes.com/2015/03/10/opinion/stop-spying-on-wikipedia-users.html
15	Wikimedia Foundation, *Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance*, March 10, 2015, available at https://blog.wikimedia.org/2015/03/10/wikimedia-v-nsa/

4. I obtained both Exhibits 14 and 15 by downloading them from the websites listed above.

I declare under penalty of perjury that the foregoing is true and correct. Executed in Washington, D.C., this 15th day of February, 2019.

/s/ James J. Gilligan
JAMES J. GILLIGAN

Counsel for Defendants

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 9

IN THE UNITED STATES DISTRICT COURT FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, INC. Plaintiff, v. NATIONAL SECURITY AGENCY, et al., Defendants.	Civil Action No. 1:15-cv-00662-TSE Hon. T.S. Ellis, III

WIKIMEDIA FOUNDATION INC.'S RESPONSES AND OBJECTIONS TO UNITED STATES DEPARTMENT OF JUSTICE'S FIRST SET OF INTERROGATORIES

PROPOUNDING PARTY: **UNITED STATES DEPARTMENT OF JUSTICE**

RESPONDING PARTY: **WIKIMEDIA FOUNDATION, INC.**

SET NUMBER: **ONE**

Pursuant to Federal Rule of Civil Procedure 33, Plaintiff Wikimedia Foundation, Inc. ("Plaintiff" or "Wikimedia") responds as follows to Defendant Department of Justice's ("Defendant" or "DOJ") (collectively with Plaintiff, the "Parties") First Set of Interrogatories (the "Interrogatories"):

I. GENERAL RESPONSES.

1. Plaintiff's response to Defendant's Interrogatories is made to the best of Plaintiff's present knowledge, information, and belief. Discovery in this action is ongoing, and Plaintiff's responses may be substantially altered by further investigation, including further review of Plaintiff's own documents, as well as the review of documents produced by Defendant, which Plaintiff has just begun to receive. Said response is at all times subject to such additional or

different information that discovery or further investigation may disclose and, while based on the present state of Plaintiff's recollection, is subject to such refreshing of recollection, and such additional knowledge of facts, as may result from Plaintiff's further discovery or investigation.

2. Plaintiff reserves the right to make any use of, or to introduce at any hearing and at trial, information and/or documents responsive to Defendant's Interrogatories but discovered subsequent to the date of this response, including, but not limited to, any such information or documents obtained in discovery herein.

3. To the extent that Plaintiff responds to Defendant's Interrogatories by stating that Plaintiff will provide information and/or documents that Plaintiff deems to embody material that is private, business confidential, proprietary, trade secret, or otherwise protected from disclosure pursuant to Federal Rule of Civil Procedure 26(c)(7), Federal Rule of Evidence 501, or other applicable law, Plaintiff will do so only pursuant to the Parties' Stipulated Protective Order (ECF No. 120).

4. Plaintiff reserves all objections or other questions as to the competency, relevance, materiality, privilege, or admissibility as evidence in any subsequent proceeding in or trial of this or any other action for any purpose whatsoever of Plaintiff's responses herein and any document or thing identified or provided in response to Defendant's Interrogatories.

5. Plaintiff's responses will be subject to and limited by any agreements the Parties reach concerning the scope of discovery.

6. Plaintiff reserves the right to object on any ground at any time to such other or supplemental interrogatories as Defendant may at any time propound involving or relating to the subject matter of these Interrogatories.

II. GENERAL OBJECTIONS.

Plaintiff makes the following general objections, whether or not separately set forth in response to each Interrogatory, to each instruction, definition, and Interrogatory made in Defendant's Interrogatories:

1. Plaintiff objects to the Interrogatories in their entirety insofar as any such instruction, definition, or Interrogatory seeks information or production of documents protected by the attorney-client privilege or the work product doctrine. Fed. R. Civ. Proc. 26(b)(1). Such information or documents shall not be provided in response to Defendant's Interrogatories and any inadvertent disclosure or production thereof shall not be deemed a waiver of any privilege with respect to such information or documents or of any work product immunity which may attach thereto. Fed. R. Civ. Proc. 26(b)(5)(B).

2. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory seeks identification of documents, witnesses, or information that Defendant has withheld from Plaintiff. Fed. R. Civ. Proc. 26(b)(1), (2).

3. Plaintiff objects to the Interrogatories in their entirety to the extent any such Interrogatory requires Plaintiff to identify potentially thousands of pages of documents, not all of which have been or can be located and reviewed by counsel within the time period allowed for this response or within a reasonable time. Accordingly, said Interrogatories would subject Plaintiff to unreasonable and undue annoyance, oppression, burden and expense.

4. Plaintiff objects to any Interrogatories that exceed the scope of jurisdictional discovery as defined by Defendants, *see* ECF No. 116 at 4, and ordered by the Court.

5. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory seeks information that is available through or from public

sources or records, or that are otherwise equally available to Defendant, on the ground that such instructions, definitions, and/or Interrogatories unreasonably subject Plaintiff to undue annoyance, oppression, burden, and expense. Fed. R. Civ. Proc. 26(b)(1), (2).

6. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory purport to impose obligations that are greater or more burdensome than or contradict those imposed by the applicable Federal and local rules. *See* Fed. R. Civ. Proc. 26, 33.

7. Plaintiff objects to the Interrogatories in their entirety as the Interrogatories contain more than the "25 written interrogatories, including all discrete subparts," permitted by the Federal Rules of Civil Procedure, Rule 33(a)(1), and Defendant has not sought leave to serve additional interrogatories.

8. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory seeks documents or information no longer in existence or not currently in Plaintiff's possession, custody, or control, or to the extent they refer to persons, entities, or events not known to Plaintiff or controlled by Plaintiff, on the grounds that such definitions or Interrogatories are overly broad, seek to require more of Plaintiff than any obligation imposed by law, would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense, and would seek to impose upon Plaintiff an obligation to investigate, discover, or produce information or materials from third parties or otherwise that are accessible to Defendant or readily obtainable from public or other sources. Fed. R. Civ. Proc. 26(b)(1), (2).

9. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory seeks information or production of documents protected from disclosure by any right to privacy or any other applicable privilege or protection, including

the right to confidentiality or privacy of third parties, any right of confidentiality provided for by Plaintiff's contracts or agreements with such third parties, or by Plaintiff's obligations under applicable law or contract to protect such confidential information. Plaintiff reserves the right to withhold any responsive information or documents governed by a third-party confidentiality agreement until such time as the appropriate notice can be given or the appropriate permissions can be obtained. Plaintiff also objects generally to all instructions, definitions, or Interrogatories to the extent they seek disclosure of trade secrets and other confidential research or analyses, development, or commercial information of Plaintiff or any third party.

10. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory is overbroad and unduly burdensome, particularly to the extent they seek "all," "each," "every," or "any" documents, witnesses, "factors," or facts relating to various subject matters. Fed. R. Civ. Proc. 26(b)(1), (2). To the extent Plaintiff responds to such Interrogatories, Plaintiff will use reasonable diligence to identify responsive documents, witnesses or facts in its possession, custody, or control, based on its present knowledge, information, and belief.

11. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory seeks expert discovery prematurely.

12. Plaintiff objects to any contention Interrogatories in their entirety as premature. Plaintiff will provide its response prior to the close of fact discovery.

13. Plaintiff objects to the Interrogatories in their entirety to the extent any such instruction, definition, or Interrogatory purports to require Plaintiff to restore and/or search data sources that are not reasonably accessible on the grounds that such definitions and Interrogatories would subject Plaintiff to undue burden and expense. Fed. R. Civ. Proc. 26(b)(1), (2).

III. DEFINITIONAL OBJECTIONS.

1. Plaintiff objects to definition number one (1) to the extent it defines "Plaintiff" and "Wikimedia" to include Plaintiff's "parent, subsidiary, and affiliated organizations, and all persons acting on their behalf, including officials, agents, employees, attorneys, and consultants." Said definition is overly broad, seeks irrelevant information not calculated to lead to the discovery of admissible evidence, seeks information outside Plaintiff's possession, custody, or control, and would subject Plaintiff to unreasonable and undue annoyance, oppression, burden and expense. Said definition is also vague and ambiguous in that it cannot be determined what is meant by the terms "affiliated organizations" and "all persons acting on their behalf." Plaintiff shall construe "Plaintiff" and "Wikimedia" to mean Wikimedia, and its present officers, directors, agents, and employees.

2. Plaintiff objects to definition number four (4) and to each Interrogatory that purports to require Plaintiff to "state the basis of," "stating the basis of," "state on what basis," or otherwise "state with particularity" or "identify" "all" facts, documents, or persons whose testimony support or dispute any given factual assertion, on the ground that any response thereto would require subjective judgment on the part of Plaintiff and its attorneys, and would further require disclosure of a conclusion or opinion of counsel in violation of the attorney work product doctrine and/or attorney-client privilege. Plaintiff further objects that this definition and all requests to identify documents in the Interrogatories are premature at this early stage of the litigation, would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense, and would impose an obligation to provide information greater than that required by the Federal Rules of Civil Procedure.

3. Plaintiff objects to definition number five (5) as unduly burdensome in that it

purports to require Plaintiff to "identify" each "natural person" by providing information including "her most current home and business addresses, telephone numbers, and e-mail addresses, the name of her current employer, and her title."

4. Plaintiff objects to definition number six (6) as unduly burdensome in that it purports to require Plaintiff to "identify" an "entity that is not a natural person" by providing information including "its telephone number and e-mail address, and the full names, business addresses, telephone numbers, and e-mail addresses of both its chief executive officer and an agent designated by it to receive service of process."

5. Plaintiff objects to definition number seven (7) as unduly burdensome in that it purports to require Plaintiff to "identify" documents by providing "(a) the nature of the document (*i.e*, letter, memorandum, spreadsheet, database, etc.); (b) its date; (c) its author(s) (including title(s) or position(s)); (d) its recipient(s) (including title(s) or position(s)); (e) its number of pages or size; and (f) its subject matter," or by providing information in accordance with Defendant's "Specifications for Production of ESI and Digitized ('Scanned') Images attached to Defendant National Security Agency's First Set of Requests for Production." Plaintiff further objects that this definition and all requests to identify documents in the Interrogatories are premature at this early stage of the litigation, would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense, and would impose an obligation to provide information greater than that required by the Federal Rules of Civil Procedure.

IV. INSTRUCTIONAL OBJECTIONS

1. Plaintiff objects to instruction number one (1) to the extent it purports to request "knowledge or information" from Wikimedia's "parent, subsidiary, or affiliated organizations, and their officials, agents, employees, attorneys, consultants, and any other person acting on their

behalf." Said request is overly broad, seeks irrelevant information not calculated to lead to the discovery of admissible evidence, seeks information outside Plaintiff's possession, custody, or control, and would subject Plaintiff to unreasonable and undue annoyance, oppression, burden and expense. Moreover, said request is vague and ambiguous in that it cannot be determined what is meant by the term "affiliated organizations" and "any other person acting on their behalf." Where an Interrogatory requests knowledge or information of Plaintiff, Plaintiff shall construe such request to mean knowledge or information from Wikimedia, and its present officers, directors, agents, and employees.

2. Plaintiff objects to instruction number three (3) as unduly burdensome and imposing an obligation to provide information greater than that required by the Federal Rules of Civil Procedure to the extent it purports to require Plaintiff to "identify each person known by Plaintiff to have such knowledge, and in each instance where Plaintiff avers insufficient knowledge or information as a grounds for not providing information or for providing only a portion of the information requested, set forth a description of the efforts made to locate information needed to answer the interrogatory."

3. Plaintiff objects to instruction number four (4) to the extent it seeks to require it to identify anything other than the specific claim of privilege or work product being made and the basis for such claim, and to the extent it seeks to require any information not specified in Discovery Guideline 10, on the grounds that the additional information sought by Defendant would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense, and constitutes information protected from discovery by privilege and as work product. Plaintiff is willing to discuss acceptable reciprocal obligations for disclosure of information withheld on the basis of attorney-client privilege or attorney work-product.

4. Plaintiff objects to instruction number five (5) to the extent it defines "the time period for which each interrogatory seeks a response" as "the period from July 10, 2008 (the date of enactment of the FISA Amendments Act of 2008, Pub. L. 110-261, 121 Stat. 522) until the date of Plaintiff's response." This definition is overly broad, seeks irrelevant information not calculated to lead to the discovery of admissible evidence, and would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense. Where appropriate, Plaintiff has defined the specific time period encompassed by specific responses.

5. Plaintiff objects to instruction number six (6) that the Interrogatories are continuing, to the extent said instruction seeks unilaterally to impose an obligation to provide supplemental information greater than that required by Federal Rule of Civil Procedure 26(e) and would subject Plaintiff to unreasonable and undue annoyance, oppression, burden, and expense. Plaintiff will comply with the requirements of the Federal Rules of Civil Procedure and is willing to discuss mutually acceptable reciprocal obligations for continuing discovery.

V. SPECIFIC OBJECTIONS AND RESPONSES TO INTERROGATORIES.

Without waiving or limiting in any manner any of the foregoing General Objections, Definitional Objections, or Instructional Objections, but rather incorporating them into each of the following responses to the extent applicable, Plaintiff responds to the specific Interrogatories in Defendant's Interrogatories as follows:

ALLEGATIONS REGARDING NSA INTERCEPTION OF WIKIMEDIA'S INTERNATIONAL, TEXT-BASED, INTERNET COMMUNICATIONS

INTERROGATORY NO. 1:

Please state whether Plaintiff continues to contend, for purposes of establishing jurisdiction, that "even if one assumes that a 0.00000001% chance … of the [National Security

Agency ("NSA")] copying and reviewing any particular communication, the odds of the government copying and reviewing at least one of Plaintiff['s] communications in a one-year period would be greater than 99.99999999999%." *See* Amended Complaint ¶ 58.

RESPONSE TO INTERROGATORY NO. 1:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is a contention Interrogatory that is premature at this stage in the litigation. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Subject to and without waiving any of these General or Specific Objections, Plaintiff responds as follows.

No.

INTERROGATORY NO. 2:

Unless Plaintiff's response to Interrogatory No. 1, above, is an unequivocal "no," then please state the basis, including all assumptions, of the allegation in paragraph 58 of the Amended Complaint, that, "if one assumes a 0.00000001% chance … of the NSA copying and reviewing any particular communication, the odds of the government copying and reviewing at least one Wikimedia communication in a one-year period would be greater than 99.99999999999%."

RESPONSE TO INTERROGATORY NO. 2:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is premature at this stage of the litigation because it is a contention Interrogatory and because it seeks information that may be the subject of expert reports and expert testimony. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Because Plaintiff's Response to Interrogatory No. 1 is "no," Plaintiff has not provided a response to this Interrogatory.

INTERROGATORY NO. 3:

Unless Plaintiff's response to Interrogatory No. 1, above, is an unequivocal "no," please identify any conclusions that Plaintiff disputes in paragraph 11 of the Declaration of Dr. Alan Salzberg, ECF No. 77-2, including but not limited to the conclusions that "no statistical foundation is provided" for the alleged .00000001% chance of the NSA copying and reviewing at least one Wikimedia communication in a one-year period contained in paragraph 58 of the Amended Complaint, and that "[i]f that assumption is incorrect, the calculation changes as a direct result," stating the basis on which Plaintiff disputes each conclusion.

RESPONSE TO INTERROGATORY NO. 3:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is premature at this stage of the litigation because it is a contention Interrogatory and because it seeks information that may be the subject of expert reports and expert testimony. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Because Plaintiff's Response to Interrogatory No. 1 is "no," Plaintiff has not provided a response to this Interrogatory.

INTERROGATORY NO. 4:

Unless Plaintiff's response to Interrogatory No. 1, above, is an unequivocal "no," please identify any conclusions in paragraph 14 of the Declaration of Dr. Alan Salzberg, ECF No. 77-2, that Plaintiff disputes, stating the basis on which Plaintiff disputes that conclusion.

RESPONSE TO INTERROGATORY NO. 4:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is premature at this stage of the litigation because it is a contention Interrogatory and because it seeks information that may be the subject of expert reports and expert testimony. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Plaintiff further objects that this Interrogatory is overbroad and unduly burdensome.

Because Plaintiff's Response to Interrogatory No. 1 is "no," Plaintiff has not provided a response to this Interrogatory.

INTERROGATORY NO. 5:

Unless Plaintiff's response to Interrogatory No. 1, above, is an unequivocal "no," please state whether and on what basis on which Plaintiff disputes the conclusion, in paragraph 15 of the Declaration of Dr. Alan Salzberg, ECF No. 77-2, that "[a]ny clustering of the copying and reviewing of communications, whether by country or some other criteria, would mean that some groups would have different chances of being copied than some other groups and that the fact that a particular communication in one group is reviewed or copied means other communications in that group are more likely to be copied."

RESPONSE TO INTERROGATORY NO. 5:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is premature at this stage of the litigation because it is a contention Interrogatory and because it seeks information that may be the subject of expert reports and expert testimony. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Because Plaintiff's Response to Interrogatory No. 1 is "no," Plaintiff has not provided a response to this Interrogatory.

INTERROGATORY NO. 6:

Unless Plaintiff's response to Interrogatory No. 1, above, is an unequivocal "no," please state whether and on what basis on which Plaintiff disputes the conclusion, in paragraph 19 of the Declaration of Dr. Alan Salzberg, ECF No. 77-2, that "Plaintiff's assertions about how the [Upstream] process works – through the copying of 'certain high-capacity cables, switches, and routers' ([Am.] Compl. ¶ 49) – would mean, if accurate, that the process is, in statistical terms, haphazard" rather than random.

RESPONSE TO INTERROGATORY NO. 6:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is premature at this stage of the litigation because it is a contention Interrogatory and because it seeks information that may be the subject of expert reports and expert testimony. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

Because Plaintiff's Response to Interrogatory No. 1 is "no," Plaintiff has not provided a response to this Interrogatory.

ALLEGATIONS REGARDING INJURY TO WIKIMEDIA

INTERROGATORY NO. 7:

For each category of Wikimedia international, text-based, Internet communications identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, please identify each specific type of information contained in that category of communication in which Plaintiff contends it has

a possessory interest.

RESPONSE TO INTERROGATORY NO. 7:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, see ECF No. 116 at 4, and as ordered by the Court. Plaintiff additionally objects that this Interrogatory is compound in that it contains multiple subparts.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 8:

For each category of Wikimedia international, text-based, Internet communications identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, please identify by layer within an Internet communication packet each type of information contained in that category of communication in which Plaintiff contends it has a privacy or proprietary interest.

RESPONSE TO INTERROGATORY NO. 8:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, see ECF No. 116 at 4, and as ordered by the Court. Plaintiff additionally objects that this Interrogatory is compound in that it contains multiple subparts.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 9:

For each category of Wikimedia international, text-based, Internet communications

identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, please identify by layer within an Internet communication packet each type of information contained in that category of communication in which Plaintiff contends it has an expressive interest.

RESPONSE TO INTERROGATORY NO. 9:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, *see* ECF No. 116 at 4, and as ordered by the Court. Plaintiff additionally objects that this Interrogatory is compound in that it contains multiple subparts.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 10:

Please state the basis of Plaintiff's allegation, in paragraph 99 of the Amended Complaint, that "Wikimedia's communications also reveal private information about its operations, including details about its technical infrastructure, its data flows, and its member community writ large."

RESPONSE TO INTERROGATORY NO. 10:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, *see* ECF No. 116 at 4, and as ordered by the Court.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 11:

If Plaintiff contends that in order to conduct Upstream surveillance it is necessary to review

the application layer of a packet or otherwise review the contents of a communication where Upstream surveillance does not involve "about" collection, then please state the basis of that contention.

RESPONSE TO INTERROGATORY NO. 11:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is a contention Interrogatory that is premature at this stage in the litigation. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery. Plaintiff additionally objects that this Interrogatory seeks information that is not reasonably calculated to lead to the discovery of admissible evidence.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 12:

If Plaintiff disputes that with "rare exception" it "do[es] not contribute, monitor, or delete content" on its Project websites, that it "merely host[s] this content," "maintaining the infrastructure and organizational framework that allows [its] users to build the [Project websites] by contributing and editing [the] content themselves"; or that it "do[es] not take an editorial role" but "simply provide[s] access to the content that . . . users have contributed and edited," as stated at http://wikimediafoundation.org/wiki/Terms_of_Use (last visited on November 15, 2017) (copy attached), then please state the basis on which Plaintiff contends those statements are inaccurate.

RESPONSE TO INTERROGATORY NO. 12:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, see ECF No. 116 at 4, and as ordered by the Court.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 13:

Please state which of the categories of Wikimedia international, text-based, Internet communications identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, involve communications in which Wikimedia obtains information about the actual identities of the other parties to those communications, apart from the IP addresses associated with the communications to and from those parties.

RESPONSE TO INTERROGATORY NO. 13:

In addition to the General Objections above which are incorporated herein, Plaintiff further objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, *see* ECF No. 116 at 4, and as ordered by the Court.

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 14:

Please state which of the categories of Wikimedia international, text-based, Internet communications identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, involves communications in which Wikimedia, rather than the other parties to the communications, selects the information sent to them.

RESPONSE TO INTERROGATORY NO. 14:

In addition to the General Objections above which are incorporated herein, Plaintiff further

objects that this Interrogatory seeks information that exceeds the scope of jurisdictional discovery as defined by Defendants, *see* ECF No. 116 at 4, and as ordered by the Court. Plaintiff additionally objects that this Interrogatory seeks information that is not reasonably calculated to lead to the discovery of admissible evidence. Plaintiff further objects that this Interrogatory is vague, ambiguous and unintelligible as to its use of the terms "selects" and "them."

On the basis of these General and Specific Objections, Plaintiff will not provide a response to this Interrogatory.

INTERROGATORY NO. 15:

For each category of Wikimedia international, text-based, Internet communications identified in response to NSA Interrogatory No. 3 that Plaintiff contends is intercepted, copied, and reviewed by the NSA in the course of Upstream surveillance, please state the basis of Plaintiff's allegations, in paragraphs 75 and 109 of the Amended Complaint, that "in part" as a result of Upstream surveillance Wikimedia has undertaken "burdensome" and "costly measures" and "diverted time and monetary resources … from other important organizational work" in order to protect the confidentiality of communications in that category.

RESPONSE TO INTERROGATORY NO. 15:

In addition to the General Objections above which are incorporated herein, Plaintiff additionally objects that this Interrogatory is improperly compound in that it contains multiple subparts.

Subject to and without waiving any of these General or Specific Objections, Plaintiff responds as follows.

(1) Technical measures.

Due in part to Upstream surveillance, Wikimedia transitioned from HTTP to HTTPS as the

default protocol for all Wikimedia project webpages. In order to effectively execute its transition to HTTPS-by-default for all Project pages, Wikimedia has devoted four years of full-time employee work allocated across different members of Wikimedia's staff. This transition to HTTPS-by-default has also created additional burdens on specific Wikimedia projects or initiatives. For example, the HTTPS transition necessitated approximately six months of full-time employee work to: (1) coordinate with Wikimedia's partners regarding the manner in which the transition would affect the "Wikipedia Zero" project; and (2) provide related technical support.

Due in part to Upstream surveillance, Wikimedia implemented Internet Protocol Security ("IPsec"). In order to effectively execute IPsec implementation and maintenance, Wikimedia allocated approximately six months of full-time employee work.

The transition to HTTPS-by-default and IPsec implementation required a capital expenditure on technical infrastructure:

(i) Wikimedia spent approximately €241,148.46 on Cache/TLS-termination servers located in Amsterdam, Netherlands.

(ii) Wikimedia spent approximately $40,384.56 on Cache/TLS-termination servers located in Virginia, U.S.A.

Documentation of the aforementioned expenditures will be produced to Defendants.

Finally, Wikimedia has also hired a full-time Traffic Security Engineer who will be responsible for implementing and maintaining technical efforts to protect its users' reading and editing habits from mass surveillance—including, specifically, from the NSA's Upstream surveillance.

(2) Policy measures.

Wikimedia held internal discussions and community consultations specifically related to

NSA surveillance. Due in part to Upstream surveillance, Wikimedia expedited the negotiation, drafting, and approval of a new Privacy Policy, which included a new Access to Nonpublic Information Policy. Wikimedia staff responded to community concerns over surveillance—including concerns about the NSA's Upstream surveillance—when drafting these policy changes. As part of these efforts, Wikimedia's legal team conducted extensive negotiations with other departments within Wikimedia and consulted with outside counsel.

Due in part to Upstream surveillance, Wikimedia created new Data Retention Guidelines and a new Request for User Information Procedure and Guidelines. Specifically, in drafting each of these guidelines, Wikimedia staff was acting on concerns over NSA surveillance.

These collective processes to overhaul Wikimedia's privacy policies and create related procedures and guidelines required approximately (and at least) 14 months of one full-time employee's work.

(3) Staff practices.

Due in part to Upstream surveillance, Wikimedia staff increasingly relied on telephone communications and encrypted messaging systems, including when interacting with community members who have expressed concerns over privacy/confidentiality issues, including NSA surveillance.

Wikimedia is a non-profit organization with limited staff and financial resources. The aforementioned resources that Wikimedia devoted to protect the confidentiality of its communications were made at the expense of other organization initiatives and activities that Wikimedia could have undertaken to further advance its mission.

INTERROGATORY NO. 16:

For each of the "burdensome" and "costly measures" and "diver[sions] of time and

monetary resources" identified in response to Interrogatory No. 15, above, please identify and describe with particularity every factor other than Upstream surveillance to which that measure or diversion of time and resources is attributable, including but not limited to any other NSA surveillance, any other U.S. government surveillance, any foreign government surveillance, any surveillance by non-government entities, and protection against computer viruses or other computer-crime activities.

RESPONSE TO INTERROGATORY NO. 16:

In addition to the General Objections above which are incorporated herein, Plaintiff additionally objects that this Interrogatory is improperly compound in that it contains multiple subparts. Plaintiff additionally objects that this Interrogatory is overbroad and unduly burdensome to the extent that it requests that Plaintiff identify "with particularity every factor other than Upstream surveillance" that has attributed to Wikimedia's measures to protect the confidentiality of its communications.

(1) Technical measures.

Wikimedia transitioned to HTTPS-by-default primarily due to concerns over Upstream surveillance. Other factors that influenced Wikimedia's transition to HTTPS included Wikimedia's desire to protect against: (i) surveillance practices of foreign state actors; (ii) practices of commercial actors; and (iii) individual computer hackers. A significant majority of Wikimedia's ongoing work to maintain HTTPS standards and practices is due solely to the threat of NSA surveillance.

Wikimedia implemented IPsec primarily due to concerns over Upstream surveillance. Other factors that influenced Wikimedia's decision to implement IPsec included Wikimedia's desire to protect against: (i) other NSA surveillance practices; (ii) other U.S. government

surveillance practices; and (iii) surveillance practices of foreign state actors.

Wikimedia hired a Traffic Security Engineer partially due to concerns over Upstream surveillance. Other factors that influenced this staffing decision included Wikimedia's desire to protect against: (i) other NSA surveillance practices; (ii) other U.S. government surveillance practices; (iii) surveillance practices of foreign state actors; (iv) practices of commercial actors; and (v) individual computer hackers.

(2) Policy Measures.

Wikimedia enacted a new Privacy Policy partially due to concerns over Upstream surveillance. Other factors that influenced Wikimedia's policy update included Wikimedia's concerns regarding: (i) other NSA surveillance practices; (ii) other U.S. government surveillance practices; (iii) surveillance practices of foreign state actors; (iv) responding to civil subpoenas; (v) responding to government subpoenas; (vi) practices of commercial actors; (vii) protecting against individual computer hackers; and (viii) keeping policies up-to-date and transparent.

Wikimedia created new Data Retention Guidelines partially due to concerns over Upstream surveillance. Other factors that influenced Wikimedia's creation of these guidelines included Wikimedia's concerns regarding: (i) other NSA surveillance practices; (ii) other U.S. government surveillance practices; (iii) surveillance practices of foreign state actors; (iv) responding to civil subpoenas; (v) responding to government subpoenas; (vi) practices of commercial actors; (vii) protecting against individual computer hackers; and (viii) keeping policies up-to-date and transparent.

Wikimedia created new Request for User Information Procedure and Guidelines partially due to concerns over Upstream surveillance. Other factors that influenced Wikimedia's creation of the guidelines included Wikimedia's concerns regarding: (i) other NSA surveillance practices;

(ii) other U.S. government surveillance practices; (iii) surveillance practices of foreign state actors; (iv) responding to civil subpoenas; (v) responding to government subpoenas; (vi) practices of commercial actors; and (vii) protecting against individual computer hackers; and (viii) keeping policies up-to-date and transparent.

(3) Staff practices.

Due in part to Upstream surveillance, Wikimedia staff increasingly relied on telephone communications and encrypted messaging systems, including when interacting with community members. Other factors that influenced these practices included: (i) other NSA surveillance practices; (ii) other U.S. government surveillance practices; and (iii) surveillance practices of foreign state actors.

INTERROGATORY NO. 17:

For each of the "burdensome" and "costly measures" and "diver[sions] of time and monetary resources" identified in response to Interrogatory No. 15, above, please state the basis of Plaintiff's contention that the measure or diversion is attributable to Upstream surveillance rather than factors identified in Plaintiff's response to Interrogatory No. 16, above.

RESPONSE TO INTERROGATORY NO. 17:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is overbroad, unduly burdensome, and a contention Interrogatory that is premature at this stage in the litigation. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery.

(1) **Wikimedia's transition to HTTPS-by-default**. Revelations about Upstream surveillance in summer 2013 were a substantial factor in Wikimedia's decision to transition to HTTPS-by-default. Wikimedia initially had significant reservations regarding how the transition

would affect users in large restricted corporate networks or users in countries such as China and Iran, for whom Wikimedia project webpages may or would become inaccessible if they were transitioned to HTTPS. Had it not been for revelations about the NSA's Upstream surveillance, it is likely that Wikimedia would not have transitioned all of its Project webpages to HTTPS-by-default, and instead would have relied on a less burdensome approach through which users could "opt-in" to using HTTPS. Revelations related to Upstream surveillance also contributed to Wikimedia's execution of the transition process on an accelerated basis.

(2) **Wikimedia's IPsec implementation.** Revelations about Upstream surveillance in summer 2013 also prompted and was the decisive factor in Wikimedia's decision to implement IPsec. Wikimedia had considered implementing IPsec before the revelations, but only acted once it learned the extent of the NSA's surveillance practices as disclosed in June 2013. Knowledge that the NSA's Upstream surveillance involved tapping the Internet backbone made IPsec implementation necessary to protect the confidentiality and security of Wikimedia's communications. Revelations related to Upstream surveillance also contributed to Wikimedia's execution of the transition process on an expedited basis.

(3) **Wikimedia's hiring of a Traffic Security Engineer.** Wikimedia's primary motivation in hiring a Traffic Security engineer is to maintain ongoing efforts to protect the confidentiality and security of its Internet communications in response to NSA surveillance practices, including the Upstream surveillance. If it were not for Wikimedia's extensive ongoing efforts to combat the threat of NSA surveillance, Wikimedia would not have expended the additional resources to hire a new employee for this position.

(4) **Policy Measures.** Major substantive work on Wikimedia's Privacy Policy, Data Retention Guidelines, and Request for User Information Procedure and Guidelines occurred

subsequent to revelations about Upstream surveillance in summer 2013. Wikimedia held community consultations related to NSA surveillance in connection with drafting the new policies and guidelines, and community member concerns about surveillance—including specifically the Upstream surveillance—were taken into account by Wikimedia staff members when crafting the policies. With NSA surveillance concerns in mind, Wikimedia's new Privacy Policy was expressly designed to minimize the collection and retention of user information.

(5) **Wikimedia's increased reliance on telephone communications and encrypted messaging systems when interacting with community members.** Revelations about Upstream surveillance in summer 2013 led to a reluctance on the part of international community members to interact with U.S.-based Wikimedia staff. Fears over NSA surveillance of international text-based Internet communications meant that Wikimedia was required to increasingly rely on telephone and in person communications and encrypted messaging systems when interacting with community members.

Wikimedia had been aware of surveillance threats from other state actors, but the sophistication and extent of NSA practices created a heightened emphasis on communications security. Because of Wikimedia's international scope, Upstream surveillance presented the most direct threat to the confidentiality and security of its communications, both with members of the community and within the organization internally.

INTERROGATORY NO. 18:

Please state the basis of Plaintiff's allegations, in paragraphs 76 and 110 of the Amended Complaint, that "Upstream surveillance has resulted and will result in some foreign readers, editors, contributors, and volunteers being less willing to read, contribute to, or otherwise engage with Wiki Projects," including but not limited to the allegations that "Wikimedia users have

expressed reluctance to continue participating in the Wikimedia movement because of" Upstream surveillance, and that Upstream surveillance reduces the likelihood that individuals will share information or communicate with Wikimedia's staff, or otherwise contribute to or read its Projects.

RESPONSE TO INTERROGATORY NO. 18:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is improperly compound in that it contains multiple subparts.

Numerous Wikimedia users around the world have expressed their reluctance and concern with respect to participating in the Wikimedia movement and to sharing information or communicating with Wikimedia's staff because of NSA surveillance, including Upstream surveillance, in a variety of ways and settings, including but not limited to, Wikimedia community forums and discussion groups, communications with Wikimedia employees, and responses to peer-reviewed academic studies.

Wikimedia hosts a number of community forums and other similar web pages in which Wikimedia users can and do converse on a range of topics, including their use of, and participation in, Wikimedia projects. Users in these conversations have discussed U.S. government surveillance and how that surveillance deters users from participating in Wikimedia projects. Indeed, because of this surveillance, many Wikimedia users feared not only participating in Wikimedia projects as contributors or editors but also even reading or visiting Wikimedia pages; this collected information could be used by the U.S. government to reveal users' identities, to identify their political or social activism, or to detect anti-American bias. Users were concerned that they could suffer adverse consequences as a result of this government surveillance.

Additionally, Wikimedia staff have had numerous conversations with Wikimedia users outside of the United States who have voiced substantial concerns with NSA surveillance

activities, including Upstream. Many of these people are involved in political or social activism and live or work in geopolitical areas that are a special focus of the U.S. government's counterterrorism or diplomatic efforts, such as Iran, Russia, Egypt, Ukraine, India, and China. And these individuals have engaged in repeated acts of self-censorship vis-à-vis Wikimedia because of NSA surveillance: some refuse to discuss sensitive political topics on which they once spoke candidly; some will now only speak in person rather than over email or other communication channels they used to use; and some will only speak through intermediaries. These individuals have censored their speech in part or altogether because they fear that the U.S. surveillance could, among other things, serve to identify them, jeopardize or undermine the political or social movements they work in, or otherwise harm themselves or their families. Many of them were especially concerned because of the NSA slides showing that the NSA has expressed interest in surveilling Wikimedia's communications.

As one specific example, due in part to concerns about U.S. government surveillance, including Upstream surveillance, some of Wikimedia staff's international contacts have refused to communicate certain information to Wikimedia over the Internet. These refusals directly affect Wikimedia's ability to carry out its work. For example, Wikimedia historically required individuals seeking particular administrative privileges to provide Wikimedia staff with photo identification. However, several European users declined to transmit photo identification to Wikimedia via the Internet because of concerns about U.S. government surveillance.

Moreover, academic studies have explored in quantitative and qualitative fashion the negative effects of NSA surveillance on Wikimedia users' participation in, and interaction with, Wikimedia projects. *See* J. W. Penney, Chilling Effects: Online Surveillance and Wikipedia Use, 31 Berkeley Tech. L.J. 117 (2016); A. Forte, N. Andalibi, R. Greenstadt, *Privacy, Anonymity, and*

Perceived Risk in Open Collaboration: A Study of Tor Users and Wikipedians, Proceedings of Computer-Supported Cooperative Work and Social Computing (2017).

INTERROGATORY NO. 19:

Please state or, if necessary, estimate the decline in readership of Wikimedia's Projects (in terms of lost page views), in contributions to Wikimedia's Projects or related websites and pages (in terms of lost one-way messages from contributors or editors), and in other communications from outside individuals to Wikimedia's staff (in terms of one-way messages) that Plaintiff contends has occurred because of Upstream surveillance, as described in paragraph 110 of the Amended Complaint, stating the basis of that contention.

RESPONSE TO INTERROGATORY NO. 19:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is a contention Interrogatory that is premature at this stage in the litigation. Plaintiff additionally objects that this Interrogatory requests information that may the subject of expert discovery. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery. Plaintiff further objects that this Interrogatory is improperly compound in that it contains multiple subparts.

Wikimedia staff have had numerous conversations with Wikimedia users outside of the United States, including those residing or working in geopolitical areas that are a special focus of the U.S. government's counterterrorism or diplomatic efforts, such as Iran, Russia, Egypt, Ukraine, India, and China. These individuals have engaged in repeated acts of self-censorship vis-à-vis Wikimedia because of NSA surveillance, including Upstream – among other things, they have refused to communicate with Wikimedia staff over email and they have curtailed or abstained from working on Wikimedia projects.

As one specific example, due in part to concerns about U.S. government surveillance, including Upstream surveillance, some of Wikimedia staff's international contacts have refused to communicate certain information to Wikimedia over the Internet. For instance, Wikimedia historically required individuals seeking particular administrative privileges to provide Wikimedia staff with photo identification. However, several European users declined to transmit photo identification to Wikimedia via the Internet because of concerns about U.S. government surveillance.

Moreover, academic studies have explored in quantitative and qualitative fashion the negative effects of NSA surveillance on Wikimedia users' participation in, and interaction with, Wikimedia projects. *See, e.g.*, J. W. Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 Berkeley Tech. L.J. 117 (2016); A. Forte, N. Andalibi, R. Greenstadt, *Privacy, Anonymity, and Perceived Risk in Open Collaboration: A Study of Tor Users and Wikipedians*, Proceedings of Computer-Supported Cooperative Work and Social Computing (2017).

Finally, Wikimedia staff have censored their electronic communications with users and avoided communicating with users regarding certain sensitive topics because of concerns about U.S. government surveillance, including Upstream surveillance.

INTERROGATORY NO. 20:

For each reduction in readership of, contributions to, or other engagement with Wikimedia Projects (or related websites and pages), and reduction in communications or information shared with Wikimedia's staff, identified in Plaintiff's response to Interrogatory No. 19, above, please state the basis of Plaintiff's contention that the reduction is attributable to Upstream surveillance rather than any other NSA surveillance activity, any other U.S. government surveillance, any foreign government surveillance, any surveillance by non-government entities, protection against

computer viruses or other computer-crime activities, or other factors.

RESPONSE TO INTERROGATORY NO. 20:

In addition to the General Objections above which are incorporated herein, Plaintiff objects that this Interrogatory is a contention Interrogatory that is premature at this stage in the litigation. Plaintiff therefore specifically reserves the right to supplement and amend its response based on further investigation and discovery. Plaintiff further objects that this Interrogatory is overbroad, unduly burdensome and improperly compound in that it contains multiple subparts.

Wikimedia bases its contention that the reduction in readership of, contributions to, or other engagements with Wikimedia Projects (or related websites and pages), and reduction in communications or information shared with Wikimedia's staff, identified in Wikimedia's response to Interrogatory No. 19, is attributable to Upstream surveillance on a variety of sources, including but not limited to, as discussed in detail in response to Interrogatory No. 19, conversations and communications with Wikimedia users outside of the United States, such users' behavior and conduct, scholarly articles such as Jonathon Penney's Chilling Effects, and the NSA slide showing that the NSA has expressed interest in surveilling Wikimedia's communications.

Dated: January 11, 2018

/s/Ashley Gorski
Ashley Gorski
AMERICAN CIVIL LIBERTIES UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
agorski@aclu.org

Counsel for Plaintiff Wikimedia Foundation, Inc.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 10

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,) No. 1:15-CV-00662-TSE
Plaintiff,)
v.)
NATIONAL SECURITY AGENCY,)
et al.,)
Defendants.)
)

- - -

THURSDAY, APRIL 12, 2018

- - -

30(b)(6), Topic 4, Deposition of WIKIMEDIA FOUNDATION, by and through its designee, JAMES ALEXANDER, taken at the offices of Cooley LLP, 1299 Pennsylvania Avenue, NW, Ste 700, Washington, D.C., beginning at 10:00 a.m., before Nancy J. Martin, a Registered Merit Reporter, Certified Shorthand Reporter.

APPEARANCES:

U.S DEPARTMENT OF JUSTICE
CIVIL DIVISION, FEDERAL PROGRAMS BRANCH
BY: JAMES J. GILLIGAN, ESQ.
TIMOTHY A. JOHNSON, ESQ.
OLIVIA HUSSEY SCOTT, ATTORNEY AT LAW
20 Massachusetts Ave., N.W., Room 6102
Washington, D.C. 20001
(202) 514-3358
james.gilligan@usdoj.gov
Representing the Defendants

COOLEY LLP
BY: DEVON HANLEY COOK, ATTORNEY AT LAW
101 California Street
Fifth Floor
San Francisco, California 94111
(415) 693-2116
dhanleycook@cooley.com
Representing the Plaintiff

ALSO PRESENT:
PATRICK TOOMEY, ACLU, STAFF ATTORNEY
ASHLEY GORSKI, ACLU STAFF ATTORNEY

INDEX

	PAGE
TESTIMONY OF JAMES ALEXANDER	
BY MR. JOHNSON	5
BY MS. HUSSEY SCOTT	102
BY MR. JOHNSON	127
BY MS. HANLEY COOK	192

EXHIBITS

NUMBER	DESCRIPTION	MARKED
Exhibit 1	Notice of Deposition Pursuant to Fed.R.Civ.P. 30(b)(6), 3 pages	7
Exhibit 2	Wikimedia Statistics, 1 page	97
Exhibit 3	Wikipedia Report Card: Summaries for 50 most visited languages, 101 page	101
Exhibit 4	Wikitionary Report Card: Summaries for 50 most visited languages, 8 pages	110
Exhibit 5	Wikimedia Traffic Analysis Report, Wikipedia Page View Per Country, WIKI6369 - 6374, 6 pages	118
Exhibit 6	Wikimedia Foundation Quarterly Report, WIKI5978 - -6131, 156 pages	127
Exhibit 7	Access to Nonpublic Information Policy, WIKI6389 - 6392, 4 pages	171

EXHIBITS

NUMBER	DESCRIPTION	MARKED
Exhibit 8	Talk: Access to Nonpublic Information, policy/Archives/2013, WIKI6410 - 6452, 43 pages	173
Exhibit 9	The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity is Causing its Decline, 20 pages	178
Exhibit 10	Intelligent Machines, The Decline of Wikipedia, 17 pages	180

WASHINGTON, D.C., THURSDAY, APRIL 12, 2018; 10:00 A.M.

- - -

JAMES ALEXANDER,
having been first duly sworn/affirmed,
was examined and testified as follows:

EXAMINATION

BY MR. JOHNSON:

Q. Okay. Well, welcome. I'm Tim Johnson, representing the government defendants in this matter. With me are my colleagues, Jim Gilligan and Olivia Hussey-Scott. I'll probably be doing most of the talking on our side, but they may occasionally jump in with questions.

A. Okay.

Q. Would you please state your full name for the record.

A. James Alexander.

Q. And have you ever been deposed before?

A. I have not.

Q. Okay. So fun times for all. I'll give you a few basic guidelines, but if you have any questions about procedure, obviously you can ask counsel or just feel free to stop me and ask me to clarify. Glad to.

So please keep your answers verbal. No nods

Page 6

or "uh-huh" just so the court reporter can get every answer down. And please, for the same reason, try to speak slow and clearly. I'll try my best to do the same.

On that note, if you don't hear me or don't understand my question, please let me know. I'll be happy to rephrase or ask it again.

If you answer a question, I'll assume that you've heard the question and understood it. So, again, any confusion just let me know.

Along those lines, I know that these questions might raise some technical issues. If a technical answer is you believe the most accurate and correct way to answer the question, feel free to give it, though I will very likely follow up with some clarifying questions, trying to put it in layman's terms both for us and for anyone who might be reading the transcript.

A. Okay.

Q. If you realize you've made a mistake, forgotten something, want to return to any question, feel free to just let me know and do so.

If you'd like a break at any point, that's fine. Just let me know. I would ask you to finish answering whatever question has been asked before we

Page 7

take a break.

A. Uh-huh.

Q. Do you have any concerns about what I've just said, any questions?

A. No, that seems to make sense.

Q. Great. So are there any physical or mental conditions, any drugs or alcohol you've consumed recently or anything else that might affect your ability to testify truthfully today?

A. No, there are not.

Q. So to the best of your knowledge, you're able to testify truthfully and accurately today?

A. Correct.

MR. JOHNSON: Now I'd like to add one exhibit just pro forma.

Could you mark this as Government Exhibit No. 1, please.

(Deposition Exhibit 1 was marked for identification.)

BY MR. JOHNSON:

Q. Have you seen this document before?

A. I have seen the document.

Q. And what is the document?

A. The document is the "NOTICE OF DEPOSITION" and the topics that would be discussed.

Page 8

Q. Thank you. And could, for the record, you read the topic listed as No. 4.

A. Sure. It reads, "Plaintiff's assertion 'that Upstream surveillance has resulted and will result in some foreign readers, editors, contributors, and volunteers,' among others, 'being less willing to read, contribute to, or otherwise engage with Wikimedia's Projects' or to 'share information or communicate with Wikimedia staff,' as alleged in the 1st Amended Complaint 76, 110, as set forth in response to DOJ's Interrogatory Nos. 18-20, and as argued in support of Plaintiffs' standing in" Plaintiff's Motion of Opposition at 41.

Q. Thank you. I just want to confirm, are you appearing as plaintiff, Wikimedia Foundation's, designated witness on this topic?

A. I am.

Q. And are you prepared to testify regarding this topic today?

A. I am.

Q. Thank you. I'd just like to start off with some general background. Who is your current employer?

A. My current employer is the Wikimedia Foundation.

Page 9

Q. What's your position with the Wikimedia Foundation?

A. I'm the manager for trust and safety.

Q. How long have you been the manager for trust and safety, roughly?

A. In this specific role, about three to four years.

Q. And what were you doing before you took this position?

A. I had some lower-level positions working on similar topics, as well as working on the fundraising team.

Q. How long altogether have you been with the Wikimedia Foundation?

A. It will be eight years as of August.

Q. And could you just briefly describe your general duties at present.

A. Currently I supervise a team of three people directly and am part of a team of eight now. My main focuses are liaising and working with community members with especially trusted responsibilities. Especially users who have access to private information or private data. I've been elected into those positions by the community, as well as liaising with law enforcement and working on threats of harm,

Page 42

They also had -- we've also had conversations or that they have discussed their concerns about getting specialized access, especially if that specialized access would require sending private information or private concerns to us, especially through electronic means.

So as an example, they are the identification concern that we had earlier. There were quite a few who were very concerned about sending them -- sending us their identification unless we completely changed our policy to not require that, and that they would either -- that they would either refrain from putting themselves up for election or resign their position if we did not change our policies to not require them to send in that private information.

That's all I can think of off the top of my head right now, but I think that others, in some of what we've written, include a little more.

Q. Thank you. How many individuals does the Wikimedia Foundation know to a certainty refrain from using Wikimedia projects based on their concerns about NSA surveillance?

A. When you say, "refrain," do you mean refrain completely, refrain partially?

Q. Well, let's start with refrain completely.

Page 43

A. I think knowing for certainty, as you phrased, is difficult. I don't -- it is difficult to 100 percent say that somebody left because of this. There were some users who mentioned that they might or they would, especially as we were discussing what we could or would change to make them more comfortable, and then did, in fact, leave. However, they didn't send us a letter that explained exactly why they were leaving. People do leave for different reasons.

And so being able to point to that and know for certain that they left, indeed, because of the concerns that they had given us earlier is difficult. It is also difficult to say for certain that they did not come back in a means that we are unable to tell. But we certainly have had people who have stated that and then, in fact, did not come back.

Q. Do you have any -- excuse me.

Does Wikimedia Foundation have any estimate or ballpark of about the number of individuals it specifically expressed concerns about NSA surveillance which you understood to include upstream and then left?

A. For individuals who explicitly presented it to somebody within the Wikimedia Foundation and then left, I would say four to six. However, given the

Page 44

context and concerns in general, I think that they, on a personal basis and from talking to others within the Wikimedia Foundation more broadly, it is also likely that many of them were not willing to discuss that with us because they would have had to -- would have had to say it in such a way that itself could have been seen.

And so there is a good chance that a portion of those people who left around that time, or since then, have done it because of surveillance in general, NSA surveillance, specifically.

MS. HANLEY COOK: Tim, we've been going about an hour and I could use a break. So whenever you're at a good place.

MR. JOHNSON: This is a perfectly good stopping point.

MS. HANLEY COOK: Okay. I didn't mean to --

MR. JOHNSON: That's fine.

(A recess was taken from 11:02 a.m. to 11:18 a m.)

BY MR. JOHNSON:

Q. So we were discussing the interactions between Wikimedia Foundation personnel and users who were concerned about upstream and related NSA surveillance.

Page 45

A. Correct.

Q. So in those conversations, interactions, did the users explain specifically why NSA surveillance was of concern to them?

A. Specifically, during the interactions when they were talking?

Q. Yes.

A. Yes. So, in general, they would explain that they were concerned both on that -- sorry. Just to clarify, do you mean the sort of -- either why they believe they would be targeted, why they think it -- like how it would affect them?

Q. Mine was just a general question that subsumed, basically, all of those. So feel free to start wherever you feel most appropriate.

A. So, in general, many of them believed that there was a concern that everything that they had, that they were doing could be seen, could possibly be saved. And so could be of concern if -- either now or later. So there were worries that they sort of could come back to haunt them or could be taken out of context.

So, for example, fears of what it would look like if you just took a small slice of the articles that they were viewing or reading or editing,

Page 46

especially in the editing context because they might be administering or adjusting articles that are not necessarily actually about their personal belief. So their personal topics. They will try to keep an article neutral even though it's about somebody they are very not neutral about or a topic they are not very neutral about.

So it could even include, for example, just as an imaginary example, taking an article about a former Nazi or a current Nazi but a dead Nazi, and taking out quotes or long pieces that made them look bad because it was taking up huge amounts of the space and was making the article slanted more than it should be, given the context. If something like that was taken out of context, it could make them look like they were more favorable to the individual when they were not. Or when in their country, being favorable towards that person could even be illegal.

The same thing could happen on the LGBT topics, on local current politics topics talking about the history of their government or their country or about somebody else. And so there are worries about present day concerns or something that they did now that can then be taken out of context years down the road if that information was stored for one reason or

Page 47

another.

They -- many of the individuals are -- at least the ones who spoke to me and who spoke to some of the other staff members and Foundation staff who I spoke to, staffing contractors who they are involved in activities locally that may be of concern both to either to the United States, to their local government or both.

For example, they were involved in revolutionary activities, in human rights activism or activism in general. For many of them in their country just operating on Wikipedia or the Wikimedia projects in general could be considered activism or can be considered problematic. The right to free knowledge that is very important to the Wikimedia Foundation and frequently very important to sort of the United States or to western citizens is not always seen the same in other countries, and even specific topics can be of -- can be of concern or can be seen as a concern for citizens of those countries.

So Europeans, for example, have very -- many of the European countries have very specific laws about speaking about Nazism or about Fascism or hate crimes that are very different to the United States or vice versa. Different topics that would be sensitive.

Page 48

And they had concern that the information that they, the NSA, either in upstream or in other surveillance programs, could take can then be shared -- could either be used by the United States, either now or in the future, or it could be shared with their own government and then be used against them by their own government.

Again, just giving the breadth of this, there are many other examples there that I may not be thinking about.

Q. I'm not trying to put words in your mouth. So correct me to the degree I'm misstating anything. But there is a concern that NSA would collect their information, pass it on to foreign governments, and that would lead to prosecution or other adverse action by the foreign government in the country that the individuals lived in. That's one concern?

MS. HANLEY COOK: Objection. Misstates prior testimony.

BY MR. JOHNSON:

Q. To the degree it misstates your prior testimony, please correct my characterization.

A. That is certainly one concern of, I think, many. So it would also include the NSA or the U.S. government using surveillance that they collected in

Page 49

upstream ourselves or as the United States, either, for example, when they were coming to the United States, either as a visitor, either within -- as a "Wikimedian" -- so for conferences or meetings that we would have here, for Visa applications for the same or as a tourist or in their day-to-day work, since most of these people -- most of these individuals have day jobs that may require travel here or elsewhere, or that it could be used against them by the United States in a foreign country, as well as, I imagine, other concerns on their part.

Q. Did these concerns evolve over time or have they been relatively consistent?

A. I think it depends on the individuals as well as the individuals' context. They certainly evolved early on. There is -- there was very little conversation, as I said earlier, about U.S. surveillance until sort of June 2013 when awareness became one vault. There were small blips, but in general, what was discussed before that was considered to not be significantly affecting the Wikimedia Foundation and its projects specifically, while some of the things that came out, especially upstream, were seen as something much more directly affecting our projects.

Page 74

numberswise, it would require additional research. I know that there was some that have been done by researchers who have more knowledge than I have.

Q. Just so I'm clear, I want to close the loop, make sure I don't misunderstand you. Putting aside academic studies, statistical breakdown, is there any other evidence that you're speaking of a representative from Wikimedia Foundation is aware of that would demonstrate that participation in the Wikimedia projects have decreased based on upstream surveillance?

A. So I can give specific examples, if necessary, of conversations. Most of them -- most of what I know is conversations both documented and undocumented that they have given us or that they've had with us, and there may be others that I don't know or other examples that I'm not thinking of.

Q. I'm sorry. Just to be clear, the "they" in response --

A. The users -- the users have communicated with us.

Q. You mentioned "examples." Are there any examples beyond the examples we've already discussed today?

A. Sure. I can give you a couple if you'd like.

Page 75

Q. Yes. That would be very helpful. Thank you.

A. So I'm just trying to think of not the entire breadth of every example I've heard but some representative options. For example, we had a user who was very concerned that they, as a U.S. citizen who lived abroad, would be a representative target and be much more interested -- interesting to U.S. surveillance.

They specifically sent us a message that -- they actually sent it directly to the stewards, to our trusted community members who were elected to do this, amongst other things, asking for permission to be able to use Tor or virtual private networks. Not other. They're sort of separate technologies, but in order to hide their true location and their true IP address when editing and when viewing the projects.

When viewing is -- generally, you would still be able to view, but you would not be able to edit from any of those projects through Tor or through any open or closed proxy. They asked for that permission explicitly stating that they were doing so because they felt at higher risk as a U.S. person outside of the United States because of NSA surveillance and because of -- how they described it, I believe off-stream surveillance because they were specifically

Page 76

asking right around the consultations around upstream surveillance and because they talked about sort of in the background conversation.

They were granted that permission, but that is a cumbersome process that would require action on their part. So that every time they want to edit, it's not the normal way to view pages or to work with our sites. And even though they do have permission to do that, it requires them to take special action in order to continue editing on our sites.

We also had a number of users. We've already talked about some of the users with the identification policy and concern about sending identification. We have had conversations about whether or not we would be willing to allow people to hide their IP address by default, something that we were not completely willing to do, which I imagine could very well cause some of them to back off without telling us exactly why.

We had a user more recently, I think 2017, similar to the last person I talked to from 2013, who is a non-U.S. person, but this person was arrested into the Philippines, specifically asking permission to use virtual private networks or other proxies in order to be able to edit and hide his IP address because they were afraid that the NSA would surveil

Page 77

their information and turn it over to those Philippine -- the Filipino government, who they felt we were closely tied with and were willing to share information with. So, again, we're asking for the specific knowledge of that.

I've been told stories of Chinese users who -- especially Chinese users who were more western focused and had come to the United States for education, to attend school and then went to -- went back to China and are now on mainland China, who believe they may be especially focused on -- a special focused on NSA surveillance because they sort of came onto their screen while they were sort of in the United States but now that they've moved back, and because of that have been wary of communicating with us about grants.

They got grants from us when they were in the United States. Or communicating with us -- with our servers and reading or editing topics that may be of interest to the United States and to their activities on mainland or adjacent to Hong Kong or Macau. So Chinese controlled areas where many of them are engaged in prodemocracy or antigovernment behavior. And so they felt that might be of interest to U.S. surveillance in addition to probably understandably

Page 78

concerns about foreign surveillance. But they did have specific concerns.

There have been similar concerns that have been relayed to me through other stuff from other Asian users. For example, we had a contractor and a long-term editor from Vietnam who expressed a strong desire to avoid specific pages, and while he was helping us with translation and communication in Vietnamese, to refrain from contacting certain people because of their connection with antigovernment groups. They're a more publicly known connection, and a perceived -- or at least perceived understanding that their government, who they had big concerns with, was getting closer to the United States and would be, perhaps, more likely to receive information.

As part of that, they also asked my permission to hide their name and to normally -- while we don't require regular users to provide their real name, if they want to be a contractor, if they want to work for us, they have to use their public name, their real name on their work accounts or anything public facing. They did not want to do that because they were afraid of that, especially to communicate through electronic communication.

Many users in Asia specifically go through

Page 79

intermediaries. And so I will either find out through another community member, if they want to contact me, or how other behaviors that have happened, they'll usually go through somebody perhaps in Taiwan or somebody in a position that they feel like will be a little bit easier to get ahold of us. Or they go through staff members that they know are not in the United States.

Q. Okay.

A. The -- let's see. Other general -- so specifically I'm thinking about editing withdrawal on this?

Q. At this point I was interested in anything at all that we hadn't discussed that would shed light or provide a basis for why Wikimedia Foundation has concluded that there's been a decrease in engagement, be that with editors, users, or anyone else that participates in the Wikimedia project.

MS. HANLEY COOK: Objection. Vague and ambiguous.

BY MR. JOHNSON:

Q. To the degree you don't understand my question, I can clarify, please. Let me know how I can clarify.

A. So, again, as we talked earlier, putting

Page 80

aside the possibility of academic research, we also have had some specific -- multiple specific sort of sensitive incidents and issues. So my work, for example, on investigations of sensitive topics, we had where we've had to do a lot of work to try to make people feel comfortable and make ourselves feel comfortable, and we were talking to them about government surveillance or about government actions in their local government and their concern that the United States would be listening in on that. That would include --

Q. I'm sorry. So these are foreign individuals --

A. Yes.

Q. -- concerned about foreign government actions?

A. Yes.

MS. HANLEY COOK: Objection. Misstates the prior testimony.

THE WITNESS: So, yes. For those individuals, these are foreign individuals but interacting with Wikimedia Foundation staff --

BY MR. JOHNSON:

Q. Okay.

A. -- or us attempting -- the Wikimedia

Page 81

Foundation trying to reach out to foreign individuals to ask them questions or ask their evidence in incidents that have happened in multiple -- in both the United States and internationally. And in many of those cases they have expressed concern communicating with us, interacting with us.

I already spoke about some of the privacy policy conversations and identification there, and conversations with users who stated that they may -- that they wanted to reduce their interaction. At the moment, that's the big sort of topic, sort of general areas. There are multiple examples in most of those areas that we could go in, but that covers most that I can think of for right now.

Q. Obviously, if you think of something else that's relevant, we appreciate you letting us know later during this discussion.

Sort of along those same lines, we've been focusing on users. Is it Wikimedia Foundation's position that upstream surveillance has similarly caused Wikimedia staff and contractors to decrease their participation, engagement in Wikimedia projects?

MS. HANLEY COOK: Objection. Outside the scope of the topic for which this witness is designated, but I'll let him answer in his personal

Page 82

capacity.
MR. JOHNSON: This topic I will just briefly state, this was covered in Wikimedia's response to, it's --
MR. GILLIGAN: Interrogatory 19.
MR. JOHNSON: Which was part of Topic 4.
Q. But proceed.
A. I think it's definitely true that Wikimedia Foundation staff and contractors have had to reduce or change our interactions or communication with community members, and at times with the general public because of concerns with this.
Q. Have -- would that include any self-censorship communications that you would have sent but didn't send because of upstream surveillance?
A. Yes, there have definitely been times where both I personally and other members of the Wikipedia Foundation staff have decided to either change what they were sending or not send something because of concerns about surveillance.
Q. Can you provide some examples?
A. Sure. So one specific example, which actually led to some ongoing concerns, was a case we had involving a torture in Azerbaijan.
This was originally a complaint that came to

Page 83

us from multiple community members in an Azerbaijani language project, that they had -- they had an individual who was -- who had been tortured -- or was being tortured by local government, as well as under the -- what they believe was the direction of a local government official who was an elected administrator on the project. That was obviously of great concern to us, and so we wanted to investigate that. We did not completely understand the topic. We were speaking through language difficulties.
They -- both themselves and staff, we were concerned about communicating about the topic directly in the open. Some of that was concern for local government surveillance, but especially on our part there was also concern about U.S. surveillance because we were talking about a sort of sensitive area of the world. We knew that there was lots of interest both in the U.S. government and from other governments in that region and that we were talking about specific actions that were being done by the local government, sort of in retaliation and against us and the press in general.
So because of that, we had to proceed quite slowly and carefully, and doing this investigation despite knowledge that some of this behavior may have

Page 84

been ongoing and highly problematic for users. That involved quite a few conversations early and ongoing with human rights organizations and other groups that had more knowledge and understanding in how to do these types of investigations and the behavior of both the United States and other countries.
It involved having to work through intermediaries, community members that we knew were either able to speak the language orally and so could contact individuals behind the scenes and have discussions and then communicate with us orally as well through either encrypted voice chat or through encrypted written conversations with a preference towards encrypted voice chat and other nondocumented methods.
It also involved multiple staff members outside the United States having interviews with the individuals who had reached out to us to try to more fully understand the -- exactly what was happening. So, for example, at the very beginning we had a general belief that their word "tortured" actually meant much more of a theoretical sense, that somebody was being stressed because of interactions. It took us a while to realize they meant physical torture.
We were eventually able to take some action,

Page 85

including banning and removing the administrator involved, but it took us almost a month of attempting to sort of methodically go through this while communicating as little as we could with individuals outside, and for what communication we had to do, taking some burdens, some steps to try to keep that as private as possible for their safety.
And in the end, we couldn't do everything. We sort of had to do the most we could. That included what became an ongoing sort of monthly -- before that it had been sort of ad hoc -- oral and encrypted, where possible, meetings with our larger steward group, with our elected global users from around the world, to be able to brief them on what was happening, why, and to what we were doing in our direction.
Those were conversations that we did not want to be overheard by anybody, both U.S. and overseas because that group includes a large amount of people who are involved in -- who are involved in behavior that we felt could be interesting both to their local government and to the United States. 30 to 35 people, for example, who are in 25 different countries sort of spread out.
And we, in past before that we had had -- we had sort of avoided -- we'd either had written

Page 86

documented meetings, especially prior to 2013, sort of ad hoc, occasional meetings that we would have in documented IRC channels or conversations that we just happened to have in open E-mail.

We then moved to trying to just avoid sensitive topics with them and talk about topics that we were fine with being viewed. Because of the great usefulness that they presented to us in that investigation, they -- we wanted to be able to continue to talk to them about sensitive topics, and so had to start setting up regular meetings using encrypted forms of communication to be able to continue to have that, but then also to keep focusing those sensitive topics to very specific times when we were able to do so securely. We also had a couple in-person meetings, not on that topic but with that group because of that.

There have been a number of times, both personally and as the -- within the Foundation, that we have refrained from sending notices or warnings to people that we knew were in sensitive locations. For example, in China, specifically where we knew that there were individuals that had been surveilled that they had been presented evidence that there was surveillance. We felt that there was a good chance

Page 87

that they were targets of U.S. and foreign surveillance. So they would be of interest.

We -- again, there are many examples, but another specific example would be we had some Iranian users who we work very closely with, one of which ran into problems in Iran where they had been picked up multiple times, had threats made to them by the local government, and because of that they wanted to leave Iran, and we were going to help them, along with our affiliates in Germany.

And they were very concerned and we were concerned about surveillance from both the U.S. and internal, domestic surveillance in Iran. So they sent us -- they were willing to send certain documentation, especially after the fact or when it was in the open, but during the actual sort of most sensitive periods of that movement they wanted to talk to somebody private and outside the United States.

So I had one of my staff members who lives in Greece talk to them, and then the communication with Germany -- with our German affiliate happened through encrypted E-mail between myself and the liaison there, specifically to keep that restricted.

Q. Thank you. Is it possible to estimate how many times Wikimedia Foundation staff have had to

Page 88

completely forgo communications based on concern of upstream surveillance?

A. I think it would be impossible to estimate a full number. There have certainly been dozens of times that I know of, but given that there are many times that people sort of decide to do it differently or to not do it at all, it never gets to a point where somebody I've talked to or myself would know about it. Knowing for certain all of those occasions would not be possible. And there are certainly other examples that I may be unaware of that may have already been things that we've turned over or maybe something that no one has yet told me about, somebody I've talked to.

Q. Thank you. When Wikimedia staff engaged in such censorship, what did they fear would happen if the NSA intercepted their communications?

A. So I can't speak for everybody. For those I've talked to myself, I think the biggest concern was for the individuals we were communicating with, and that that information could be used to -- either directly by the United States now or in the future if it was seen to be of interest. Many of these individuals were either to our knowledge or could be where they were involved in activities that would be of interest in the United States and to allied

Page 89

countries.

And there was concern that their identification and information about what they -- about their actions and travel and communication would put them at risk, would allow them -- would make them be more easier to pick up, to be talked to easier, to blackmail or to be asked to do certain other activities or would -- well, could be used against them in general while asking them to do certain activities I think was the biggest one.

And, again, as I talked about some of the community concerns earlier, there was concern about this sort of out-of-context questions about it, either purposefully or not, that if you only see a slice of activity, it can be very -- it can look very differently than it is intended to. I think I am unaware of -- yeah. I'm unaware of any specific incident where we thought the United States would or should be worried about an individual we were talking to, but that given the wide variety of things they were doing, any small snippet could make it appear to be of concern. And so that was a worry that that would then be seen as a problem -- as a person who is a problematic individual or -- and a need of action to be taken because of that.

Page 90

Q. Obviously, I don't want to mischaracterize what you're saying, but am I understanding you correctly -- and correct me to the degree I'm wrong -- the concern was that the U.S. government or foreign governments might take action against your users as opposed to Wikimedia staff themselves?

MS. HANLEY COOK: Objection. Misstates prior testimony.

BY MR. JOHNSON:

Q. To the degree I did, please correct me.

A. I think that is one concern is that that communication could then be used against the users. I believe there would be -- both the witness tomorrow and some of our others, there's also concern of staff that it could be used against them. Many of our staff are international citizens or international residents.

From a completely personal capacity, I know that there are multiple Visa holders that have been worried about communication that could then be used against them as citizens who are residing in the United States, sort of on approval from the United States, that could be used against them in order to remove them or to cause them to do something in order to stay.

We also have many staff members and

Page 91

contractors -- or actually, all contractors, but sort of related individuals who are outside the United States and work with us who are also -- who are then worried that they would be targets communicating back and forth.

Q. Thank you.

A. I should clarify there may be other examples that I'm not thinking about right now.

Q. Of course. Thank you.

You had mentioned that -- and again, please correct me to the degree you disagree with any of this characterization -- that one basis for Wikimedia Foundation's conclusion that upstream surveillance was decreasing -- or had played a role in decreasing engagement was academic studies.

In the discovery responses two particular studies I mentioned are entitled "Showing Effects, On-line Surveillance on Wikipedia Use," and Privacy, Anonymity, and Perceived Risk in Open Collaboration, A Study of Tor Users and Wikipedians."

I can provide additional information of those articles to the degree that's unclear, but to the degree you understand the articles I'm referring to, are those the articles that you've mentioned as one basis.

Page 92

MS. HANLEY COOK: Objection. Misstates prior testimony in the beginning of that long question.

Go ahead.

THE WITNESS: I do not know, just as not writing it, how much these were relied upon in the original filings for future, just that I believe that academic studies are going to be much better at giving specific stats or specific facts about changes related to some upstream surveillance and other surveillance in, especially, readership and editing at large because there are so many different things that need to be controlled for that requires specialized knowledge and specialized research.

I am aware of those two -- of the two studies that you mentioned, or at least believe I am from how you described them. I have not read every word of both of those studies but had perused them in preparation for this deposition.

BY MR. JOHNSON:

Q. Thank you. Are you aware of any other academic studies, excluding any expert testimony that the Wikimedia Foundation might offer in this case, on which the Wikimedia Foundation is relying to demonstrate a decrease in engagement?

MS. HANLEY COOK: Objection. Outside the

Page 93

scope of the topic noticed. He'll answer in his personal capacity. It might also call for expert testimony.

MR. JOHNSON: I obviously disagree.

Q. Please answer.

A. I am not aware of any specific studies that have been done other than that. I imagine that if we are aware, we would give them to you or they would be made aware otherwise. It would surprise me if there were more, but I do not know any off the top of my head.

Q. Of the two studies aforementioned, did the Wikimedia Foundation support or facilitate these studies in any way?

MS. HANLEY COOK: Objection. Outside the scope of the topic noticed. He'll answer in his individual capacity.

MR. GILLIGAN: We disagree.

MR. JOHNSON: We disagree.

MS. HANLEY COOK: If you want to save time, we can stipulate that you always disagree with me when I make those objections.

MR. JOHNSON: Sure. Fine. Thank you.

THE WITNESS: I'm not aware of any specific support which gave for either of those. I am aware

Page 146

a total.
BY MR. JOHNSON:
Q. Have Wikimedia readers or editors complained about the article quality on Wikimedia sites?
MS. HANLEY COOK: Objection. Same objection.
THE WITNESS: Again, that is a topic that has come up sort of off and on for a while. I think in my personal experience and knowledge from others, that is a complaint that has become less and less prevalent. It used to be a very frequent question or concern either from editors or from readers or from the mass media and from others outside of the movement. That has become a significantly less concern to the point that, in general, most comments that we hear now are the opposite.
And so, for example, lots of large companies using our content either directly -- so, for example, Facebook shows -- if you go to a page about a company that has not created a page on its own, it will attempt to show you the Wikipedia articles that you can still see information about that individual or -- that individual or that company. There is much more -- in the more recent news there is use of articles on news companies on Facebook recently to try to look at the idea of fake or incorrect news.

Page 147

YouTube is using it to try to present people with information and knowledge around different means or different topics that may be confusing to people. That has become an increasing -- an ever increasing thing that I think is a sign that people are trusting our content. A lot of different search engines, for example, also use our content, both Wikipedia content as well as content coming from Wikipedia Commons, a repository with data -- our repository to present that data, for example, in the little info box on the side of Google or Bing. I know uses it in some of their presentations. That is an ever increasing thing, which I think sort of goes against the older concerns about quality.
BY MR. JOHNSON:
Q. Are there any particular changes that Wikimedia Foundation has made that have increased quality?
MS. HANLEY COOK: Objection. Vague and ambiguous. Overbroad. Beyond the scope of the topic noticed.
He can answer in his individual capacity if he knows.
THE WITNESS: So the Wikimedia Foundation itself, just to be clear, does not control content.

Page 148

That is much more in the realm of the editors, except for extreme situations where there is a legal reason or a safety reason or something of that level that we have to come in on. So the editors create content policies. They help to manage the content. They have done so and have obviously been perfecting and adjusting those policies throughout the time.
We have also -- the Wikimedia Foundation has certainly either assisted or run programs to try to push for better content. So part of the gender diversity push, which focuses on -- for example, there were pushes mostly led by the community but with support from the Wikimedia Foundation on articles about female scientists, or similar, to try to push for better representation there.
We have also tried to support the development of the new project Wikidata, which is used by a lot -- as sort of a data repository, and it's freely available for anybody, both commercially and noncommercially to use. That is a relatively recent project that we have put a lot of time and effort and money into developing, as well as trying to ensure that the view, both through APIs for third parties and through our websites, is easier for that content.
BY MR. JOHNSON:

Page 149

Q. Do Wikimedia sites experience any seasonal variations in their traffic?
MS. HANLEY COOK: Objection. Beyond the scope of the topic noticed.
He can answer in his individual capacity to the extent he knows.
THE WITNESS: In general, I think there is some seasonal variation. The -- exactly what seasonal variation can adjust depending on the projects -- which project we're talking about or areas of the projects that individual articles may or may not have seasonal variation depend on the subject involved. Different languages may have a difference based on the population that is using them.
One example is that our global user base, especially in English Wikipedia, tends to have a bit of a dip during the summer, just because there are people out of school, and a lot of people use it in school or when they are studying. And then that will come back up. So that, obviously, needs to be taken into account.
BY MR. JOHNSON:
Q. Have Wiki users or editors complained about foreign government censorship of Wikimedia projects?
MS. HANLEY COOK: Objection. Beyond the

Page 186

excuse us, I know it's easier for us to step out briefly. We just want to discuss and make sure that we haven't neglected any important areas of inquiry, but otherwise, we're just about done.

MS. HANLEY COOK: I'm going to run to the bathroom.

MS. HUSSEY SCOTT: Let's go off the record.

(A recess was taken from 4:55 p m. to 5:01 p m.)

BY MR. JOHNSON:

Q. Okay. Really, just to wrap things up, are there any answers to my questions you've given today that you'd like to change before I stop asking questions?

A. Not that I can think of specifically other than to just clarify our methods of communication. I know a lot of the time we were talking about communication I focused on, sort of person-to-person written communication, E-mails and chat programs, and that includes VPN or like private chat channels, apps that may be encrypted, allow voice chat, allow text chat, E-mail, encrypted E-mail, phone conversation, and the like.

But it also includes with the Wikipedia Foundation as a whole, a lot of server traffic that

Page 187

can -- that is also a significant amount of the communication, obviously, between users especially, our readers, and editors whenever they're interacting with our website.

Q. Any information responsive to any of my previous questions that you didn't remember when I asked you but that you've since recalled?

A. Not that I can think of, but it's certainly possible. But, yeah, nothing that I can think of at the moment.

Q. Anything else you'd like to add to what you've told us today so that we can better understand Wikimedia Foundation's perspective on this issue?

MS. HANLEY COOK: Objection. Vague and ambiguous.

BY MR. JOHNSON:

Q. The topic is the topic notice of the deposition.

A. Only to point out sort of the breadth of the fears that people can have. And so there are very explicit and specific fears that information collected by the NSA will be used against them now and in the future, and that it will be used as sort of a chip against them or that it could be -- that it could be harmful to them in the future. However, there's also

Page 188

the sort of underlying and ongoing concern of the violation of privacy in general, sort of the looking over your shoulder fears. The harm that comes from feeling that you're always watched.

That has been an ongoing conversation, an ongoing concern that has gone on in individual one-on-one, person-to-person conversations, especially at Wikimedia events. I even talked to one staff member who got stopped at one point on the street by a reader who was concerned and asking questions about who could view the information that they were sending, like what articles they were reading and similar.

Editors certainly have a lot of that concern. Some of the communication we've gotten from readers is like the fear that just sort of everything that they're doing is being watched. I think that was a strong underlying fear and harm from everything as they were going on, especially after more and more of the revelations happened.

The original awareness was -- awareness increased originally around more electronic sharing between like coming from a service provider on a specific requests to the government, to the U.S. government. That was always seen as a lower level concern because it sort of meant that there was --

Page 189

they were being targeted. And so any individual who did not feel they had a reason to be targeted did not feel they had a reason to be afraid.

The upstream surveillance changed that fear very significantly because suddenly, they're worried about the mass collection or the mass viewing of their data about sort of somebody always looking. And so always having to be careful that what you're doing could be taken out of context or could be seen differently, or one mistake could suddenly come back to haunt you later on when they may not have even realized it was a mistake.

I think that has been another ongoing one even if at times we focused on a specific incident or a specific fear at one individual point in time. I don't think that always gets to that broader underlying concern.

Q. So am I understanding you to be saying that the Wikimedia users complained -- who expressed concerns about upstream surveillance understood it to be a mass surveillance program?

MS. HANLEY COOK: Go ahead.

THE WITNESS: It is my understanding from a lot of the communication that I've received from editors, I viewed from editors and the communication

Page 190

I've received from other staff, that most of the people they discussed NSA surveillance with, upstream surveillance with, they saw it as a mass collection program that was not -- that could be targeted but was not always targeted.

And so they were unsure of how much that would be filtered and how much was going to be viewed and for how long.

BY MR. JOHNSON:

Q. Just to clarify, you mentioned a fear that the United States would use users' data as a chit against them, I believe was the phrase. Could you elaborate on what you mean by that?

A. To clarify the term "chit," that was definitely my own wordage.

Q. Of course not a technical term, but what you meant in context.

A. That it could be used as either blackmail or as leverage against them, that if they made -- if they were viewing, say, articles of political significance or of concern, that that could be shown to them. It could be threatened to be given to others, that articles that they were writing or editing that they felt were private, for example, something that really revealed that they may have been gay or transgender or

Page 191

a lesbian or something that revealed that they may hold political views that are unacceptable or problematic in their region.

Whether that's a region in the United States or a region in the world, that that could be used as a negative in their favor or against them.

MR. JOHNSON: Okay. As a technical housekeeping matter, we would like to hold this deposition open for now simply because --

MS. HANLEY COOK: Have you said you have no further questions?

MR. JOHNSON: I have no further questions at this time.

MS. HANLEY COOK: So I'm going to step outside and figure out if I have any redirect to clean up the record at all, but go ahead.

MR. JOHNSON: I just want to make sure that I note for the record that we're holding it open simply because we need to review the additional documents that were produced last night. If those don't bring any further questions, then we'll be happy to officially close the deposition.

MS. HANLEY COOK: Okay. Let's go off the record. I'll try and keep it quick. Let's just circle out for one second. Let me figure out if

Page 192

there's anything I need to clean up.

(A recess was taken from 5:10 p m. to 5:13 p m.)

EXAMINATION

BY MS. HANLEY COOK:

Q. So, James, earlier today you were asked what types of private information users expressed fear about sending to Wikimedia Foundation due to upstream surveillance. I believe you said IP address, sending pictures of their photo or government ID, personal information related to attendance at events, and I think you just mentioned web requests. Were there any other kinds of private information users expressed a fear of sending to the Wikimedia Foundation due to upstream surveillance?

A. So web requests were like HTP requests specifically, contain a bunch of information themselves that they would possibly be adding, much of which would be considered private information. It would include the actual pages that who they are viewing and who they are, their IP address, what pages they're requesting specifically. And it would also include information about their computer. What we would call a "user agent," but also things like the

Page 193

size of their monitor and what browser they're using, what OS information that can be used to identify -- that could be used to identify them compared to, say, other people who were at the IP address, and they could appoint specific laptop or desktop phones that they were on.

It could -- it also could include other information that could be used to identify them. For example, information from our cookies that could help somebody to connect to the user who was actually viewing those pages, or the person that is viewing those pages, potentially information from other cookies or other sites that would be over the same, as well as the site that they were coming from, which would reveal something about them.

For the web address, I think that's the majority. There also may be pieces of it that I may not be thinking about. In addition, the identification or identifying information sent to us is not just for events. It could also include identifying information because they want to join one of our programs. It could be information that they are required to present in order to be on a committee, to be on -- to ask us for assistance, or they also have to frequently send information to identify

themselves or to reidentify themselves to the community but to our servers for the ticketing system that I was suggesting to release -- if they want to release copyright images or if they want to confirm copyright text that they already hold and they're trying to donate, or if they want to verify who they are, that they own an account.

For example, if they are a notable individual, they have to -- who says that they are a noted individual, they will have to send us identifying information to compare and connect their account to their individual, and while we keep it private, it still comes through our serves and would still be readable and accessible. There may be others that I'm not thinking about.

They also, of course, not infrequently, will tell us what some of their fears are, which can then reveal information about them, specifically what could be used to target them.

MS. HANLEY COOK: Great. I have no further questions.

MR. JOHNSON: Okay. No further questions from us either.

MS. HANLEY COOK: Okay. Great.

Thanks, Nancy. I will figure out who can

scan this.

(Witness excused.)

(Deposition concluded at 5:17 p.m.)

CERTIFICATE

I do hereby certify that the aforesaid testimony was taken before me, pursuant to notice, at the time and place indicated; that said deponent was by me duly sworn to tell the truth, the whole truth, and nothing but the truth; that the testimony of said deponent was correctly recorded in machine shorthand by me and thereafter transcribed under my supervision with computer-aided transcription; that the deposition is a true and correct record of the testimony given by the witness; and that I am neither of counsel nor kin to any party in said action, nor interested in the outcome thereof.

Nancy J. Martin, RMR, CSR

Dated: April 16, 2018

(The foregoing certification of this transcript does not apply to any reproduction of the same by any means, unless under the direct control and/or supervision of the certifying shorthand reporter.)

INSTRUCTIONS TO WITNESS

Please read your deposition over carefully and make any necessary corrections. You should state the reason in the appropriate space on the errata sheet for any corrections that are made.

After doing so, please sign the errata sheet and date it. You are signing same subject to the changes you have noted on the errata sheet, which will be attached to your deposition. It is imperative that you return the original errata sheet to the deposing attorney within thirty (30) days of receipt of the deposition transcript by you. If you fail to do so, the deposition transcript may be deemed to be accurate and may be used in court.

Page 198

- - - - - - - -

E R R A T A

- - - - - - - -

PAGE	LINE	CHANGE
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____

Page 199

ACKNOWLEDGMENT OF DEPONENT

I, JAMES ALEXANDER, do hereby certify that I have read the foregoing pages, _____ to _____, and that the same is a correct transcription of the answers given by me to the questions therein propounded, except for the corrections or changes in form or substance, if any, noted in the attached errata sheet.

_____

DATE SIGNATURE

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 11

IN THE UNITED STATES DISTRICT COURT FOR THE
DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, INC.,
Plaintiff,
v. Civil Action No.: 1:15-cv-00662-TSE
NATIONAL SECURITY AGENCY, et al.,
Defendants

DEPOSITION OF MICHELLE S. PAULSON
(Corporate Designee for Wikimedia Foundation)
Washington, D.C.
Friday, April 13, 2018
9:53 a.m.

Reported by: Sheri D. Hayhurst-Smith, RPR

DEPOSITION OF MICHELLE S. PAULSON, held at the Offices of:

COOLEY, LLP
1299 Pennsylvania Avenue, NW
Suite 700
Washington, DC 20004

Pursuant to Notice, before Sheri D. Hayhurst-Smith, Registered Professional Reporter and Notary Public in and for the State of Maryland.

A P P E A R A N C E S

ON BEHALF OF THE PLAINTIFF:
DEVON HANLEY COOK, ESQUIRE
COOLEY, LLP
101 California Street
5th Floor
San Francisco, California 94111-5800
(415) 693-2116

ASHLEY GORSKI, ESQUIRE
ASMA PERACHA, ESQUIRE
AMERICAN CIVIL LIBERTIES UNION FOUNDATION
125 Broad Street
18th Floor
New York, New York, 10004
(212) 549-2500

A P P E A R A N C E S
(C o n t i n u e d)

ON BEHALF OF DEFENDANTS:
JAMES J. GILLIGAN, ESQUIRE
TIMOTHY JOHNSON, ESQUIRE
U.S. DEPARTMENT OF JUSTICE
Civil Division
20 Massachusetts Avenue, N.W.
Room 6102
Washington DC 20001
(202) 514-3358

C O N T E N T S

EXAMINATION OF MICHELLE S. PAULSON PAGE
By Mr. Gilligan 7
By Ms. Hanley Cook 177
By Mr. Gilligan 179

EXHIBITS: (ATTACHED) PAGE
Exhibit No. 11 - Notice of Deposition 12
Exhibit No. 12 - Wikimedia Foundation Inc.'s Responses and Objections to U.S. Department of Justice's First Set Of Interrogatories 16
Exhibit No. 13 - Privacy Policy 17
Exhibit No. 14 - Wikimedia Foundation Report, January, 2012 21
Exhibit No. 15 - Launching a Privacy Policy Built the Wiki Way - Wikimedia Blog 24
Exhibit No. 16 - Access To Nonpublic Information Policy 63
Exhibit No. 17 - Data Retention Guidelines 65

Page 6

CONTENTS
(Continued)

Exhibit No. 18 - Requests For User Information Procedures & Guidelines 67

Exhibit No. 19 - Securing Access To Wikimedia Sites with HTTPS, Wikimedia Blog 90

EXHIBITS: (ATTACHED) PAGE

Exhibit No. 20 - The Future of HTTPS On Wikimedia Projects - Wikimedia Blog 94

Exhibit No. 21 - Wikimedia Engineering Report 2011, October 100

Exhibit No. 22 - Wikimedia Village Pump (Technical) Archive 138 124

Exhibit No. 23 - Document entitled HTTPS - Meta 130

Exhibit No. 24 - Document entitled Enable IPSec Between Datacenters 151

Exhibit No. 25 - Job Description of Traffic Security Engineer 161

Page 7

MICHELLE S. PAULSON, having been sworn, testified as follows:

EXAMINATION

BY MR. GILLIGAN:

Q Good morning. Could you please state your name for the record.

A Michelle Sarah Paulson.

Q Sarah Paulson?

A Paulson, P-A-U-L-S-O-N.

Q S-A-R-A?

A H.

Q Oh, that's your middle name?

A Yes.

Q So, Ms. Paulson then, for the record?

A Yes.

Q Okay. Very good. Ms. Paulson, my name is Jim Gilligan. I'm an attorney with the Department of Justice, representing the defendants in this case, including the National Security Agency and others.

We're here to take your testimony in a lawsuit in which the Wikimedia Foundation asserts that a

Page 8

form of NSA surveillance, known as Upstream, violates Wikimedia's rights under the law.

I'm going to be asking you questions regarding certain allegations that Wikimedia has made in the case, and your answers will be given under oath. And they will be transcribed by the court reporter here, and then it will then become part of the record in the case.

Throughout the deposition, I'll try to make my questions clear to you, but if you don't understand any question I've asked, please feel free to ask me to clarify it, and I will attempt to do so.

Let me start off with a question, actually, that I never enjoy asking.

Are you suffering today from any physical or mental impairment that would prevent you from testifying truthfully and accurately?

A No.

Q Could you then, please, state for me your current employment.

A I am a contractor with the Wikimedia Foundation.

Page 9

Q And what are your duties and responsibilities as a contractor for Wikimedia?

A To assist with the NSA case, this case.

Q What sort of assistance are you providing in connection with the case?

A Appearing here today, amongst other things.

Q That is your sole responsibility with respect -- I'm sorry. Did you say there are other?

A In relation to this case.

Q In relation to this case?

A (The witness nods head up and down).

Q I'm sorry. I apologize. I lost the thread there.

You said that your duties as a contractor with the Wikimedia Foundation are to assist with this case?

A Yes.

Q And that they include appearing here today?

A Yes.

Q Are there any other duties and

Page 10

responsibilities that you have with respect to this case?
A I started in my contract capacity in December of last year and have just helped on an as-needed basis, because during my long-term employment with the Foundation that ended in May of 2017, I had done a substantial amount of work on this case. I was leading the case in-house for Wikimedia.
And, therefore, they thought I would be best suited to continue at this next stage of the case.
Q So, until May of 2017, you were employed by the Wikimedia Foundation?
A Yes.
Q And by whom have you been employed since May, 2017?
A I decided to take time off for a year, and I've been traveling.
Q Okay. Excellent. What were your duties and responsibilities -- well, what was your -- let me start here.
What was your position at the Wikimedia

Page 11

Foundation that you last held before leaving in May of 2017?
A I was legal director and interim general counsel.
Q And how long did you hold that position?
A As far as I can remember, the interim general counsel was approximately nine months, and legal director was probably approximately two years.
But I need to go and double check dates to be exact.
Q It's close enough. And how long were you employed by the Wikimedia Foundation all together?
A Since 2009.
Q And were you employed during that entire time in a capacity as an attorney?
A In a variety of states. I started as a legal intern and worked my way up.
Q Did you have any full-time employment prior to joining the Wikimedia Foundation in 2009?
A No. I was just finishing school.
Q And speaking of which, can you please state your educational background briefly.

Page 12

A Yes. I have a JD from University of California, Hastings College of the Law, and a Bachelor's degree from University of California Berkeley.
MR. GILLIGAN: I'd like to mark this as what -- as I recall -- or as I'm told, I should say, since I wasn't here, we ended at Government's Exhibit 10 yesterday?
MS. HANLEY COOK: That's correct.
MR. GILLIGAN: So, we'll start with 11.
MS. HANLEY COOK: Is that the notice?
MR. GILLIGAN: Yeah.
MS. HANLEY COOK: I think we marked it yesterday. If you have it, we can use it again. But I'm happy to mark it again as well.
MR. GILLIGAN: Let's mark it again, yes. Many more trees will be felled in this case than over this little notice.
(Government's Exhibit No. 11 was subsequently marked for identification and attached hereto).
MR. GILLIGAN: Here are several copies for you folks. You can give that to the witness.

Page 13

BY MR. GILLIGAN:
Q Ms. Paulson, we've just marked as Government's Exhibit 11, the Government's notice of deposition pursuant to Federal Rule of Civil Procedure 30(b)(6).
Are you familiar with this document?
A Yes.
Q Did you review it in preparation for this deposition today?
A Yes.
Q I direct your attention to the third topic of deposition listed on Page 2.
A Yes.
Q You're familiar with that?
A Yes.
Q Are you the person that the Wikimedia Foundation has designated to testify on the matters described in Topic 3?
A Yes.
Q And do you understand that in so far as my questions today concern this third topic, the answers you give will be on behalf of Wikimedia and

Page 14

not yourself personally?
A Yes.
Q And do you understand that as far as my question is concerned, Topic 3, the answers you give are not to be limited by the extent of your own personal knowledge, but are to include all information known or reasonably available to Wikimedia Foundation?
A Yes.
Q Can you please tell me what you did in order to prepare for the deposition today?
A Sure. I met with members of Cooley who are present here --
Q Uh-huh.
A -- as well as a couple of other members who are not, for a prep meeting on Monday that lasted, I think, seven to eight hours maybe, and then met again on Tuesday briefly.
Q When you say "briefly," does that mean an hour? Two hours?
A I would guess two hours, but I would need to look at my records.

Page 15

Q That's all right. And did you review any documents in preparation for the deposition today?
A I did.
Q Could you give me an estimate of how many documents you reviewed?
A It's not an exact number, but probably, roughly, 60 or so, a variety of documents.
Q Do you know whether these are documents that were produced to the defendants in this litigation?
A Yes.
Q All of them?
A Yes. As far as I know.
MR. GILLIGAN: May I ask you, Devon, if that's accurate or not?
MS. HANLEY COOK: Yeah.
MR. GILLIGAN: Okay. Thank you.
MS. HANLEY COOK: With the exception of filings in this case, which are not produced.
MR. GILLIGAN: Filings, we'll give you a break on.
MS. HANLEY COOK: Thanks, Jim.

Page 16

THE WITNESS: I should also say that I reviewed the documents on my own without Cooley as well in preparation.
BY MR. GILLIGAN:
Q In addition to the Monday and Tuesday sessions?
A Yes.
Q How much time do you think you've spent reviewing the documents on your own?
A Again, I would need to check my records, but maybe ten.
Q Ten hours or so?
A (The witness nods head up and down).
MS. HANLEY COOK: Can we go off the record for one second?
(A discussion was held off the record).
MR. GILLIGAN: Okay. I'd like to mark this document as Government's Exhibit 12.
(Government's Exhibit No. 12 was subsequently marked for identification and attached hereto).
BY MR. GILLIGAN:
Q Ms. Paulson, we've marked as Government's

Page 17

Exhibit 12, Wikimedia Foundation, Inc.'s responses and objections to United States Department of Justice's first set of interrogatories.
Are you familiar with this document?
A Yes.
Q Did you review this document in preparation for the deposition today?
A Yes.
MR. GILLIGAN: Okay. Then I would like to mark, next, Tim, it's going to be -- it should be Number 7. Mark that as Government's Exhibit 13, I believe.
(Government's Exhibit No. 13 was subsequently marked for identification and attached hereto).
MR. GILLIGAN: We've just marked, Ms. Paulson, as Government's Exhibit 13, a document produced to the Government in the case, bearing on the front page, Bates Stamp Number Wiki -- that's W-I-K-I -- 0006674. And that goes through Bates Number Wiki0006696.
MS. HANLEY COOK: Well --
MR. GILLIGAN: I'm sorry -- 6697. Thank

Page 34

Internet backbone and scooping up all communications that way without Wikimedia's knowledge. They were wondering what kind of impact it would have to readers and editors and whether there would be a chilling effect on use of the projects. They were wondering -- they were generally not very comfortable with the idea that the government would be watching what they're reading or editing. They were wondering what we could do to safeguard communications of the Foundation and the community and whether our present security measures were enough and what could be done to better safeguard those communications.

But, again, I would need to review it to make a more exhaustive list.

Q So, among the concerns, they expressed a concern that the U.S. Government, through the NSA, was, as you put it, reading their communications with the Wikimedia Foundation?

MS. HANLEY COOK: Objection. Misstates the prior testimony.

Q You may answer.

Page 35

A Sorry. Could you repeat your question?

Q Having listened to your answer there, I understood you to say -- but please correct me if it's wrong -- that "readers expressed concern that the U.S. Government was," as you put it, "watching what they were reading?"

MS. HANLEY COOK: Same objection.

Q You may answer.

A I believe they were concerned about the general tapping of the Internet backbone and being able to scoop up communications that would include communications with Wikimedia, that may include reader information.

Q But communications of theirs?

A I'm sorry?

Q The members of the community were concerned that the NSA was intercepting communications of theirs?

A As well as Wikimedia's.

Q As well as Wikimedia's. And what sort of information were they concerned that the NSA was acquiring of theirs?

Page 36

A This isn't an exhaustive list, and there may be people at the Foundation that have additional examples.

From what I can recall, there were some concerns expressed about sharing forms of identification that some select community members were previously required to send us from the previous version of the access to nonpublic information policy. At that time, it was called access to nonpublic data policy.

There were also concerns about communicating sensitive information over e-mail or other electronic forms. There was some concern about getting access to otherwise nonpublic information, such as IP addresses that we consider to be private and identifying information, which, in turn, could indicate reading habits of users.

Q In addition to readers, did individuals who contribute to or edit project websites express concerns of this nature?

A Yes. As far as I can tell you, again, if you can look at the documents that I referenced

Page 37

earlier of the discussion pages for the two consultations regarding surveillance and the privacy policy, you might be able to find particular examples.

Q And if the NSA obtained this information, what were they concerned would happen?

I'm trying to figure out if I'm speaking in a grammatically correct fashion.

If the NSA obtained this information, what is it they were concerned would happen?

A I certainly can't speak to what every community member or staff member feared would happen.

Q Right.

A That would be a great deal of speculation on my part.

But from, personally, in my conversations with users and what I saw on mailing lists and in the discussion pages --

Q Uh-huh.

A -- they were concerned that there could be a chilling impact. They were concerned that the

Page 38

government would have that kind of information on them. There aren't that many people out there in the world that are thrilled for anybody to know what sites that they visit and what their browsing habits are. They have no idea what the NSA does with this information and who it's shared with. We don't know that.

And not knowing how your information is used and information about how you live your life can be scary to some people. And they don't know what kind of consequences could happen as a result of those practices.

Q Did some users, at least, express a concern that the information you were describing, if acquired by the NSA, would somehow be used against them by the U.S. Government?

A I don't recall. But other people at Wikimedia might have specific examples, and there may be instances of that in the discussion pages or mailing lists that we turned over.

Q Did some users express concern that the information you were describing, if acquired by the

Page 39

NSA, would be shared with their own governments?

A There were a couple of instances that I was personally aware of, where there was concerns about information that couldn't be possibly intercepted being shared. However, there may be more instances that other people in the Foundation would be aware of that I was not consulted on.

Q And did these users express concern that if this information was shared with their government, they would suffer adverse consequences of some kind as a result?

A Yes.

Q What kind of adverse consequences?

MS. HANLEY COOK: Objection. Beyond the scope of the topic as noticed. The witness will answer in her personal capacity.

MR. GILLIGAN: The reasons that should become clear soon. We disagree. And we'll take that as a standing objection to save time -- or a standing response to your objection, I should say --

MS. HANLEY COOK: Great.

Page 40

MR. GILLIGAN: -- that we disagree.

MS. HANLEY COOK: If you would like to save time, I can also say "beyond the scope of the topic" to shorten my objection, if we can stipulate that that means that the witness is answering in her individual capacity.

MR. GILLIGAN: And two people suffering with allergies agree to speak less.

MS. HANLEY COOK: Also, there's a question pending, but I do have a logistical matter we should stop for one second and discuss at some point.

BY MR. GILLIGAN:

Q Well, are you prepared to answer?

A Sorry. Could you repeat the question for us?

MR. GILLIGAN: Madam Court Reporter, could you please repeat the question for us?

(Question read back).

BY MR. GILLIGAN:

Q That is to say, what kind of adverse consequences did community members state they feared

Page 41

they would suffer if the information that you were describing was intercepted by the NSA and shared with their home governments?

A This is not a definitive list, but it's what I can remember off of the top of my head. Again, if you speak with other Wikimedia employees who might have more.

There were some concerns about being investigated or harassed by local authorities, detained, having an adverse impact on their job or family, prevented from leaving the country, and in one case, potentially tortured.

MS. HANLEY COOK: Can we go off the record for one second?

MR. GILLIGAN: Oh, yes. I'm sorry. Yes. You had a logistical issue.

(A discussion was held off the record).

MR. GILLIGAN: Back on the record.

BY MR. GILLIGAN:

Q Okay. Ms. Paulson, can I direct your attention, please, to Exhibit 12, Wikimedia's responses to the DOJ interrogatories. And could I

Page 74

practices, and it contributed to them being reluctant to talk to us over certain mediums.
Q And that's the full extent of your recollection of what they said?
A That is what I recall now. I might recall something later in the deposition.
Q Did they express any concern about sensitive information being transmitted by electronic means somehow being used against them?
MS. HANLEY COOK: Objection. Asked and answered.
A I do not recall whether or not they specified that. However, I believe that they -- I don't know whether they -- I can't remember whether or not they specified by the U.S. Government, but I do recall there being concerns about surveillance being potentially leading to negative consequences for them.
And as we discussed earlier, there were the potential ways that they could be retaliated against, including being investigated, harassed, detained, tortured, adverse impact on the job, and

Page 75

prevented from leaving the country.
Q And these were consequences, as I recall us discussing, that would have flowed from the sharing of their sensitive information with their home governments?
MS. HANLEY COOK: Objection. Misstates the testimony.
A From what I recall we discussed earlier, these were potential consequences that certain individuals feared would happen by their own government, because they don't know what NSA practices are in regards to sharing of information or using such information.
Q So, returning to the interrogatory response, the second sentence, as I read before, states, "Fears over NSA surveillance of international text-based Internet communications meant that Wikimedia was required to increasingly rely on telephone and in-person communications and encrypted messaging systems."
Why was Wikimedia required to increasingly rely on these systems? To make it easy, is it because

Page 76

the users were reluctant or unwilling to use other forms of communication?
A The Wikimedia Foundation works hand in hand with its community. It's what makes us a successful organization and have successful projects.
So, maintaining open lines of communication with our community is vital to our existence and the way we function. And if there were community members that were not comfortable communicating in one way, we preferred not to forego the communication all together. So, we would try to find other methods of communication that they were more comfortable with that may not have been as insecure as an encrypted e-mail.
And as I stated before, a few examples of these messaging systems that we used --
Q Could you do this slowly this time?
A Yes. Of course. Signal, one that I personally recommend and love, telephone, in-person meetings, telegram, iMessage, IRC, Messenger.
Q Is that it?

Page 77

A I believe that's all I said before. Again, this isn't a completely exhaustive list.
Q Okay. If you will indulge my ignorance, can you tell me what kind of messaging system Signal is?
A It's an SMS messaging system.
Q Texting?
A Yes. An encrypted texting.
Q And when you said "telephone," were you referring to just picking up the phone and dialing and calling somebody, or were you referring to something else?
A It can -- when I said "telephone," it can mean a traditional telephone. It can mean a cell phone. It could also mean voiceover IP methods of communication. Would you like examples of those?
Q No. Thank you. That far into the 20th century -- 21st Century -- excuse me. I made it. I almost just proved my own point there.
Telegram, is that a reference to traditional Telegram communications, or is that something different? Now, we're back in the 19th Century.

Page 110

A To the best of my recollection, there may be more information in the discovery documents that we provided or through the personal knowledge of other employees of the Wikimedia Foundation.

However, it's my recollection that they spoke about NSA Upstream surveillance intercepting our -- sorry, not intercepting -- but the surveillance impacting Wikimedia communications in between users and the Wikimedia Foundation. And by the people who tend to come to Wikimedia or talk about these issues in person or participate in mailing lists are some of the most active members of Wikipedia on the other projects, meaning that they are contributors to it.

So, when they do talk about something affecting the projects and communications, they're frequently talking about their own as well as the work of others.

Q And in what way did they state that NSA surveillance might impact Wikimedia communications?

A Some community members -- and there may be other examples of this that other Wikimedia staff know or appear in the discovery documents or that I

Page 111

might recall later.

But some of them, based on their learning from credible news sources, that the NSA's dragnet surveillance program is where we're sweeping up their nonpublic communications with Wikimedia, which could include information that they deemed to be personal and nonpublic, such as IP addresses or reading habits.

Q And in addition to that, was there some concern that --

A There was also mention of potential chilling effects on participation of both reading and contributing to Wikimedia projects.

Q If the NAS intercepted their communications?

MS. HANLEY COOK: Objection. Vague and ambiguous.

A That there would be chilling effects as the result of this new knowledge of the extent and sophistication of the NSA's dragnet surveillance, in which it was tapping into the backbone of the Internet and through practices that were not

Page 112

previously known by the Foundation itself and not by the users until the media attention.

Q Did they explain how or state how they believe that knowledge or that belief would lead to a chilling effect on Wikimedia communications?

A From my recollection, there were instances in which people talked about a possible chilling effect as a result of NSA surveillance. There may be other examples that foundation staff or the discovery documents provide. But I remember there being a conversation about people tending to self-censor when they believe their actions are being monitored, especially when it's nonconsensual and by a large and powerful entity that is governmental in nature, such as the NSA.

There were also reasonable concerns around what the NSA's practices are and who they shared that information with, which could again contribute to a chilling effect.

Q Because of a fear of adverse consequences as a result of who the NSA might share the information with, correct?

Page 113

A I would say it was a reasonable concern for those users, as I'm sure you're aware of the U.S. is not a country in isolation, and it has partnerships with countries around the world, and intelligence sharing is part of those practices.

While it is not common knowledge exactly the nature and extent and methodology of the sharing, I think it's a perfectly reasonable suspicion that some information gets shared with other organizations. And without knowing what that is, they want to be able to protect the nature of those communications.

Q Okay. Moving further down then in Exhibit 20 to the second paragraph, it states, "Our current architecture cannot handle HTTPS by default, but we've been incrementally making changes to make it possible. Since we appear to be specifically targeted by XKeyscore, we'll be speeding up these efforts."

Do I understand correctly the reference to, in the second paragraph, XKeyscore to be referenced to the so called NSA XKeyscore program alluded to in

Page 122

sites with HTTPS", okay, the document is dated -- the announcement, I should say, is dated June 12th, 2015.

And at the very end, before you get to the comments section on page WIKI0007109, it says, "Today, we are happy to start the final steps of this transition, and we expect completion within a couple of weeks."

Did Wikimedia achieve that objective? In other words, did they complete the final steps of the transition, at least as that term is used in this document, within a few weeks of June 12th, 2015?

A I believe the authors of this blog post were referring to this step in this particular part of the transition, in that HTTPS by default implementation would occur within the next couple of weeks.

However, as we discussed earlier, this is an ongoing process. We still have those two ways where HTTP traffic could still occur, one of which is in our control, and the other is not. As I mentioned before, the browsers, the old browsers, we don't

Page 123

have control over, but over time, those browsers would be presumably updated or fallen out of use, for the most part.

The part that we do have active work on is making the noncanonical sites secure, and that is an ongoing process.

So, to say that the transition is totally complete would be inaccurate at any point, and especially when you also consider that security standards change, and upgrades will have to continue to be made in order to maintain industry standards around our TLS termination servers will need to be updated.

So, this is an ongoing process and one that we're committed to in order to continue to protect against NSA surveillance, particularly upstream.

Q Thank you. But I'm not sure, in your response there, that it necessarily got to my question.

Insofar as this document -- let me try it this way: The document in the last sentence there before the comments section talks about a transition.

Page 124

And I'm understanding the qualifications, the qualifiers in the answer you just gave, it says that "we expect completion", at least with the transition that this announcement is talking about, "within a couple of weeks."

And, so, all I'm asking is is whether the transition that this document is talking about was, in fact, completed within a couple of weeks of June 12th, 2015?

A I would need to check other documents or talk to someone at Wikimedia to ensure it. But as far as I know, it did complete within the couple of weeks stated.

But, again, this is only for this particular step of the transition.

MR. GILLIGAN: Very good. Okay. Let's take a break. Off the record, please.

(A brief recess was taken).

MR. GILLIGAN: All right. I would ask the court reporter to mark this as Government Exhibit 22.

(Government Exhibit No. 22 was subsequently

Page 125

marked for identification and attached hereto).

BY MR. GILLIGAN:

Q Ms. Paulson, we have just marked as Government Exhibit 22 a document produced by Wikimedia to the defendants in the litigation, bearing Bates Stamp Numbers WIKI0006872 through -- can you tell me the last one or show me -- 6938.

Ms. Paulson, if you know, can you tell us what a Wikipedia village pump is?

A A village pump is a certain area of Wikipedia that is frequently used for discussions amongst community members and sometimes with interactions with staff, usually relating to topics on Wikipedia as opposed to other Wikimedia projects.

Q And are you familiar with the document titled "Wikimedia Village Pump (technical)/Archive 138?"

A Yes.

Q Did you review this in preparation for the deposition?

A I did.

Q Is it, in fact, what it purports to be,

Page 126

Wikimedia Village Pump (technical)/Archive 138?
A I haven't read all 138 pages, but it appears to be that on first glance.
Q There's a table of contents on the first page. Do you see where I'm referring to?
A I do.
Q Listing various, I guess, discussion topics in the pump, for lack of a better term. One of which is HTTPS by default; do you see that, about a third of the way down the list?
A Yes.
Q If you would turn to Page 7 of the document, bearing bates stamp Number WIKI0006878, there is a heading "HTTPS by default," as indicated in the table of contents. And there's a -- tell me if this is not the proper terminology -- but a message that begins, "Hi everyone, over the last few years, the Wikimedia Foundation has been working towards enabling HTTPS by default for all users, including the anonymous ones, for better privacy and security for both readers and editors."
And then it goes on after several sentences to

Page 127

say, "This has finally been implemented on English Wikipedia, and you can read more about it here," and it gives a link to a website.
It then continues "Most of you shouldn't be affected at all. If you edit as a registered user, you have already had to log in through HTTPS. We'll keep an eye on this to make sure everything is working as it should. Do get this touch with us if you have any problems logging in or editing Wikipedia after this change, or contact me if you have any other questions."
And following the end of the message there, it appears the name -- I don't know how you say it -- Johan or Johan -- I'm not sure which -- followed by the initials, WMF, in parenthesis, the word, "talk", and then a time and date stamp of 12:43 on June 12th, 2015.
Are you familiar with a Wikimedia staff person named Johan?
A I am.
Q Would that Johan, as in Johnson?
A Yes.

Page 128

Q J-O-N-S-S-O-N?
A I believe so.
Q So, does the appearance of his name here at the end of this message indicate that he is the person who wrote that message?
A It is an indication that he is the one that posted that message. I can't say whether or not he is the one that was the original drafter.
Q But he posted it?
A Yes.
Q And, so, then a couple lines down, there is another message. It says, "To Johan," followed by the initials, WMF, "you have to know what a real drag this is."
And the individual goes on to say, after several sentences, "I want to be able to choose whether or not I'm on the HTTP server or the HTTPS server."
And whether or not that's the person's real name, it's signed Paine, P-A-I-N-E, with a date stamp of June 12th, at 1421 hours.
What then follows under that, is a message that

Page 129

again, ends with the name, Johan, followed by the initials, WMF.
Is that an indication then that the message underneath Paine's message was posted by Johan from the Wikimedia staff?
A Yes. Well, it is possible that it was not posted by him, because sometimes things go out of alignment if you don't format it properly. This indicates that it was an insert by Johan.
Q Thank you. And it states there, does it not, "The answer I was given when I was asked about this is that any form of opt-out would also leave potential security risks in our implementation, which would make it difficult to safeguard those who do not opt out."
Do you have any understanding of what the security risks are there that are being referred to?
A As I stated, I'm not a technical expert, and I am not here to answer questions in that capacity.
Unfortunately, I do not recall the particular security risks that Johan is referring to, but there

Page 130

might be others in Wikimedia Foundation that would know which he's referring to, as he does seem to have asked someone in particular for this information. It may also appear in other parts of discovery documents. I might recall later.

MR. GILLIGAN: All right. Let's go to the next one, Tim.

Let's mark this as Number 23.

(Government Exhibit No. 23 was subsequently marked for identification and attached hereto).

BY MR. GILLIGAN:

Q Ms. Paulson, we are marking as Government Exhibit 23, a document produced to us in the litigation by Wikimedia, bearing the Bates Stamp Numbers WIKI0006536 through 6540.

Is this a document you are familiar with?

A Yes.

Q Is this a document you reviewed in the course of your preparation for the deposition?

A Yes.

Q All right. It says at the very top of the page, "HTTPS Meta."

Page 131

Does that indicate it was posted on the site, meta.Wikimedia.org?

A Yes. As indicated at the bottom of the page.

Q Indeed. And is this then a post of some kind on that website by Wikimedia on this topic as HTTPS?

A The author, unless I'm missing it somewhere, is not indicated, and I would need to look at the revision history of the document in order to see whether it was a Wikimedia employee, a series of Wikimedia employees or a combination of Wikimedia employees and community members, because Meta is open to editing by the community as well as the foundation.

Q Well, it says at the --

A If you would like to direct me to it.

Q Yes. I indeed would. It says at the bottom of Page 3, with Bates Stamp Number 6538, under the heading "Effect on Unregistered Contributors," "Once HTTPS is switched fully on, there will be no option to disable, and all users

Page 132

have will have to use it.

And then it says in parenthesis, "This is necessary to prevent the SSL-stripping attack."

A Okay.

Q Okay. But you say you can't tell, on the base of this document, whether that particular statement was made by Wikimedia or by third party user?

A Based on the document that appears before me, without an editing history or a signature, I cannot confirm with certainty that it was written by a Wikimedia Foundation employee.

Q Where would I be able to find the editing history of the document?

A If you go to the address at the bottom of the page that we referred to earlier, HTTPS:/media -- sorry -- meta.wikimedia.org/Wiki/HTTPS, there will be a tab that you can click on the top of the page that says, "History," and that will provide you with dates and identify who, either by IP or user name, made edits and what it is that they made.

Page 133

Q When edits are made by Wikimedia personnel, is there any distinctive format to be made?

A Sometimes. Not always. And in earlier years, it was not as uniform as it is now. So, there may still be employees that don't have the parenthesis, WMF, that you saw in the previous exhibits that we went through.

Now, most of them have it, but not all of them.

Q But if that was used, that would be indicative --

A Yes.

Q -- that the edit was done by Wikimedia personnel?

A Yes.

MR. GILLIGAN: I want to redirect your attention to the interrogatory responses, but -- and you can certainly do that on your own. I, at the moment, however, cannot find my own copy of the interrogatory responses.

And, you know what, if you'll indulge me, can we just go off the record for a second. I'm going

Page 178

names.
You didn't say that those were the only two ways that HTTP traffic could occur, did you?
MR. GILLIGAN: Before the witness answers, I would just like object that that misstates the prior testimony.
MS. HANLEY COOK: I wasn't sure.
THE WITNESS: I don't know exactly what I said previously, but for clarification, these are two of the ways. There could be others, but those were the two that I recalled at this time.
Other people at the foundation might know other ways or other instances or examples, and there might be more information in the documents. As you stated earlier, there's quite a lot of them.
MS. HANLEY COOK: And then we can go off the record. I had need to take a quick break, holding this open five minutes to check something. And then I'll come back on the record, if that's okay?
MR. GILLIGAN: Okay.

Page 179

(A brief recess was taken).
MS. HANLEY COOK: I have no further questions.
MR. GILLIGAN: Just one question on redirect. And you may have already said this, but I don't recall. And, so, I just want to be sure.

EXAMINATION

BY MR. GILLIGAN:
Q So, Ms. Paulson, you were asked on redirect by Wikimedia's counsel, whether there were perhaps other ways that the two of you had testified to that Wikimedia might still have some HTTP traffic, not withstanding it having transitioned to HTTPS by default.
And as I recall, you said it's possible there are other ways that might happen, other than the two you testified to; is that correct?
A As far as I can recall, I thought that I said that these were two that I remember. There could be others. It might have been worded differently. But, yeah, I think that's it.

Page 180

Q Okay. That's fine. And you put your finger on just what I want to make sure is clear, at least in my mind, that although you say there may be others, you don't know what they are, at least as you sit here now?
A I do not recall anything else.
MR. GILLIGAN: All right. That's all, subject to our position about holding the deposition open, as we discussed before.
MS. HANLEY COOK: Right. And mine as well.
MR. GILLIGAN: Yes.
(Signature having not been discussed, the deposition of Michelle S. Paulson was concluded at 5:33 p.m.)

Page 181

CERTIFICATE OF SHORTHAND REPORTER - NOTARY PUBLIC

I, Sheri D Smith, Registered Professional Reporter, the officer before whom the foregoing deposition was taken, do hereby certify that the foregoing transcript is a true and correct record the testimony given; that said testimony was taken by me stenographically and thereafter reduced to typewriting under my direction and that I am neither counsel for, related to, nor employed by any of the parties to this case and have no interest, financial or otherwise, in its outcome

IN WITNESS WHEREOF, I have hereunto set my hand and affixed my notarial seal this 19th day of April, 2018

My commission expires November 4, 2020

REGISTERED PROFESSIONAL REPORTER
NOTARY PUBLIC FOR THE STATE OF MARYLAND

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 12

GOVERNMENT EXHIBIT 19 4-13-18

FOUNDATION, LEGAL, PLATFORM ENGINEERING, TECHNOLOGY, WIKIPEDIA

Securing access to Wikimedia sites with HTTPS

By Yana Welinder
Victoria Baranetsky, Wikimedia Foundation
Brandon Black, Wikimedia Foundation
June 12th, 2015

The Wikimedia Foundation is happy to announce that we are implementing HTTPS to encrypt all traffic on Wikimedia sites. With this change, nearly half a billion monthly visitors on Wikipedia and its sister projects will be able to share in the world's knowledge more securely.

To be truly free, access to knowledge must be secure and uncensored. At the Wikimedia Foundation, we believe that you should be able to use Wikipedia and the Wikimedia sites without sacrificing privacy or safety.

Today, we're happy to announce that we are in the process of implementing HTTPS to encrypt all Wikimedia traffic. We will also use HTTP Strict Transport Security (HSTS) to protect against efforts to 'break' HTTPS and intercept traffic. With this change, the nearly half a billion people who rely on Wikipedia and its sister projects every month will be able to share in the world's knowledge more securely.

The HTTPS protocol creates an encrypted connection between your computer and Wikimedia sites to ensure the security and integrity of data you transmit. Encryption makes it more difficult for governments and other third parties to monitor your traffic. It also makes it harder for Internet Service Providers (ISPs) to censor access to specific Wikipedia articles and other information.

HTTPS is not new to Wikimedia sites. Since 2011, we have been working on establishing the infrastructure and technical requirements, and understanding the policy and community implications of HTTPS for all Wikimedia traffic, with the ultimate goal of making it available to all users. In fact, for the past four years, Wikimedia users could access our sites with HTTPS manually, through HTTPS Everywhere, and when directed to our sites from major search engines. Additionally, all logged in users have been accessing via HTTPS since 2013.

GET CONNECTED

GET OUR EMAIL UPDATES

Your email address

Subscribe

MEET OUR COMMUNITY

The one-man band writing the history of Nigerian cinema into Wikipedia

Wikipedia is a to "plant and harvest" free knowledge: Ar Aghayan

More Community Profiles

MOST VIEWED THIS MONTH

'Monumental' winners from th world's largest photo contest showcase history and heritage

The top fifteen images from Wiki...

Türkiye'den Vikipedi'ye erişim engeli halen devam ediyor

Vikipedi'nin tüm dil sürümleri, Nisan ayını

New monthly dataset shows w people fall into Wikipedia rabl holes

The Wikimedia Foundation's Analytics tea

ARCHIVES

FEBRUARY 2018

JANUARY 2018

Over the last few years, increasing concerns about government surveillance prompted members of the Wikimedia community to push for more broad protection through HTTPS. We agreed, and made this transition a priority for our policy and engineering teams.

We believe encryption makes the web stronger for everyone. In a world where mass surveillance has become a serious threat to intellectual freedom, secure connections are essential for protecting users around the world. Without encryption, governments can more easily surveil sensitive information, creating a chilling effect, and deterring participation, or in extreme cases they can isolate or discipline citizens. Accounts may also be hijacked, pages may be censored, other security flaws could expose sensitive user information and communications. Because of these circumstances, we believe that the time for HTTPS for all Wikimedia traffic is now. We encourage others to join us as we move forward with this commitment.

The technical challenges of migrating to HTTPS

HTTPS migration for one of the world's most popular websites can be complicated. For us, this process began years ago and involved teams from across the Wikimedia Foundation. Our engineering team has been driving this transition, working hard to improve our sites' HTTPS performance, prepare our infrastructure to handle the transition, and ultimately manage the implementation.

Our first steps involved improving our infrastructure and code base so we could support HTTPS. We also significantly expanded and updated our server hardware. Since we don't employ third party content delivery systems, we had to manage this process for our entire infrastructure stack in-house.

HTTPS may also have performance implications for users, particularly our many users accessing Wikimedia sites from countries or networks with poor technical infrastructure. We've been carefully calibrating our HTTPS configuration to minimize negative impacts related to latency, page load times, and user experience. This was an iterative process that relied on industry standards, a large amount of testing, and our own experience running the Wikimedia sites.

Throughout this process, we have carefully considered how HTTPS affects all of our users. People around the world access Wikimedia sites from a diversity of devices, with varying levels of connectivity and freedom of information. Although we have optimized the experience as much as possible with this challenge in mind, this change could affect access for some Wikimedia traffic in certain parts of the world.

In the last year leading up to this roll-out, we've ramped up our testing and optimization efforts to make sure our sites and infrastructure can support this migration. Our focus is now on completing the implementation of HTTPS and HSTS for all Wikimedia sites. We look forward to sharing a more detailed account of this unique engineering accomplishment once we're through the full transition.

Today, we are happy to start the final steps of this transition, and we expect completion within a couple of weeks.

Yana Welinder, Senior Legal Counsel, Wikimedia Foundation
Victoria Baranetsky, Legal Counsel, Wikimedia Foundation
Brandon Black, Operations Engineer, Wikimedia Foundation

40 Comments on Securing access to Wikimedia sites with HTTPS

Uityyy 7 months

How do I manually force unencrypted access on an old mobile browser that does not support HTTPS?
I have one that 's failing to access en.m.wikipedia.org, apparently because of this, and I see no solution here.
Any magic "en.insecure.wikipedia.org"?

Share

DECEMBER 2017

NOVEMBER 2017

OCTOBER 2017

OLDER POSTS 28

WORK AT WIKIMEDIA

Work with the foundation that supports W
and its sister projects around the world. A
and join us

Frushi 8 months

HTTPS is a 'must have' in present internet. When Google said it's gonna take a closer look for a website that don't use SSL it become clear that even websites which don't need them (because they don't have any secure infomation) will have to go to HTTPS from old http.

Share

Tom 1 year

Following the huge fail of the french ISP Orange redirecting wikipedia.fr and others, why wikipedia.fr is not protect with https/HSTS ?
http://www.theregister.co.uk/2016/10/18/orange_blows_up_french_gov_website/

Share

bart 1 year

Google usually has an alternate (cache) for each wiki link.
I just use these cache pages.

Share

Rodion 2 years

I also want there is a way to use wikipedia with plain HTTP if necessary. Currently there is a stupid debate between our government and local wiki representatives (I could not decide which of them is more stupid, I'm sorry) about restricting access to certain pages (about drugs). Providers can do this for single page if it is accessed with HTTP, but they need to deny access to whole website if it is accessed via HTTPS.

So it would be good if we have some fallback, perhaps with banner explaining "all horrible consequences" of reading wiki in plain HTTP. In my personal opinion being super-obsessed with security measures may sometimes create unwanted problems to other people :(

Share

Creg 3 years

Flo said

"Concerning privacy: when you browse Wikipedia the URLs contain the topic you are reading thus any sniffer can track what you are currenly reading. Only the *contents* is encrypted, but the contents is visible by anybody anyway (in contrast to the content of my bank account)."

False. The root domain (wikipedia.org) can be inferred from the IP address of the server during the TCP/IP request but the complete URL and exact page you're reading cannot.

Read the article on https.

Share

Flo 3 years

Is there *any* way to use Wikipedia *without* https?
I have an old device which is not capable of using https. And please don't tell me to buy new hardware or software.
So please offer a possibility to read Wikipedia *without* forced https!!!!

BTW: I cannot follow the reasons to *enforce* https:
Concerning privacy: when you browse Wikipedia the URLs contain the topic you are reading (e.g.:

https://en.wikipedia.org/wiki/CMAC) thus any sniffer can track what you are currenly reading. Only the "contents" is encrypted, but the contents is visible by anybody anyway (in contrast to the content of my bank account).

Concerning "integrity of data": nobody will guarantee that the content of Wikipedia is accurate because everybody can contribute to it. Thus I do not "fully" rely to anything I read in Wikipedia.

Share

omtim 3 years

Great step for sure, actually, in digital world https is more imperative

Share

Gary Smith 3 years

All the points are explained very clearly, Great source of information. Thanks for en-lighting us with your knowledge, it is helpful for many of us.

Share

Sports Fan Stan 3 years

All well and good to force everyone to use https. Would it be too much to ask to employ a real SSL certificate that doesn't rely on a wildcard. At present, we can't even use Wikipedia anymore because we can't trust the website. Uggghhh...

Share

astrodevamm 3 years

Very good step indeed, in fact, in cyber world https is more important because of security issues. Know a days users check website also they check that website https not. If they found https is not they click on cut button and skip from website...

Share

Pushpendra Pal 3 years

Great move team. Web is becoming a tool for governments and enforcement agencies to surveillance on citizens. SSL helps website visitors to send and receive encrypted data.

I also want to move my website http://careervendor.com from HTTP to HTTPS. I am fearing about loosing traffic, backlink and ranking. Can anyone please suggest a way for proper migration.

Share

astrodevamm 3 years

Very good step indeed, in fact, in cyber world https is more important because of security issues. Know a days users check website also they check that website https not. If they found https is not they click on cut button and skip from website...

Share

Ron 3 years

> There are two reasons someone might ask for any form of downgrade or opt-out to be permitted:

Make that three reasons.
I run in DOS, and I like to keep the functionality of Arachne.

Yes, I also run Links, Elinks and Lynx in DOS, but Arachne is more versatile than all of them – except for a lack of SSL.

Share

zzo38 3 years

I *really* want the ability to connect without HTTPS. I want to avoid the overhead required by HTTPS please.

Share

Mat2 3 years

"Because then a man in the middle can replace anyone's user agent details with another user agent, and bingo, nobody any longer has any encryption at all. Invisibly and undetectably."
Such an attack is already possible with tools such as sslstrip. Therefore user-agent sniffing doesn't decrease security for other users out there: it will make life easier neither for criminals nor for companies that want to monitor traffic.
Wikipedia is going to use HSTS and add itself to HSTS preload lists in browsers: that will block downgrade to HTTP for new browsers.

"Upgrading from IE6 to a secure browser is entirely possible for every single user on the planet. There is no sane reason for anyone, anywhere, to use an insecure browser."
Not every computer user can do this, unfortunately.

Google makes sure that IE6 still works:
https://www.ssllabs.com/ssltest/analyze.html?d=google.com&s=74.125.239.96&hideResults=on
Wikipedia is such an important site on the internet.

Share

dewimorgan 3 years

"Wouldn't it be possible to add some user-agent sniffing" NO! No it would not. Because then a man in the middle can replace anyone's user agent details with another user agent, and bingo, nobody any longer has any encryption at all. Invisibly and undetectably. Why would wikimedia hand attackers such a gift on a plate?

Upgrading from IE6 to a secure browser is entirely possible for every single user on the planet. There is no sane reason for anyone, anywhere, to use an insecure browser. The very worst smartphone and smartwatch in the world can browse securely. Even Lynx can handle secure browsing, and that's been ported to just about everything.

There are two reasons someone might ask for any form of downgrade or opt-out to be permitted: 1) they are grievously uninformed; or 2) they are maliciously requesting the downgrade on behalf of some organization which wants a MitM attack to work.

One wonders how many of each group is commenting here.

Share

Mat2 3 years

Now all IE6 users will be cut off from using Wikipedia:
https://www.ssllabs.com/ssltest/analyze.html?d=en.wikipedia.org

Wouldn't it be possible to add some user-agent sniffing so that these browsers could still access Wikipedia? They are usually used by poorer people.

Share

Ron Clarke 3 years

Steve,
> Why now adding a SSL/TLS support to that browser instead, is this really something very hard to do, or just not a priority?

Adding SSL to Arachne would be wonderful, and we wish we could. But ....we have a lack of suitably skilled coders with an interest in DOS browsers, and Arachne in particular.

Any volunteers ?

Share

dewimorgan 3 years

@Glenn McCorkle and Ron Clarke:

"Ron & I are active developers of DOS Arachne"

This ship has sailed.

Every single .gov domain will be HTTPS-only by next year. Many already are.

For active developers of web browsers which don't support HTTPS, implementing it should have been the number one priority for the last few years, because other browsers – even other command-line browsers that can run on legacy hardware – support it just fine. Like an FTP program without FTPS or SFTP, or an email program without STARTTLS, you'll lose market share and relevance.

Oh, and IPv6 URLs are a thing now, too.

Share

MORE COMMENTS

Comments are closed.

WIKIMEDIA FOUNDATION
The Wikimedia Foundation, Inc is a nonprofit charitable organization dedicated to encouraging the growth, development and distribution of free, multilingual content, and to providing the full content of these wiki-based projects to the public free of charge. Get Involved | Log In

WIKIMEDIA PROJECTS

The Wikimedia Foundation operates some of the largest collaboratively edited reference projects in the world.

WIKIPEDIA COMMONS MEDIAWIKI WIKIBOOKS
WIKIDATA WIKINEWS WIKIQUOTE WIKISOURCE
WIKISPECIES WIKIVERSITY WIKIVOYAGE WIKTIONARY

WIKIMEDIA MOVEMENT AFFILIATES

The Wikimedia projects have an international scope, and the Wikimedia movement ha already made a significant impact throughout the world. To continue this success on a organizational level, Wikimedia is building an international network of associated organizations.

WIKIMEDIA CHAPTERS THEMATIC ORGANIZATIONS WIKIMEDIA USER GROUPS

This work is licensed under a Creative Commons Attribution 3.0 unported license. Some images under CC BY-SA
Read our Terms of Use and Privacy policy. | Powered by WordPress.com VIP

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 13

GOVERNMENT EXHIBIT #22 4-13-18

WIKIPEDIA

Wikipedia:Village pump (technical)/Archive 138

< Wikipedia:Village pump (technical)

Village pump (Policy · Technical · Proposals (persistent) · Idea lab · Miscellaneous)

Village pump (technical) archive

This page contains discussions that have been archived from Village pump (technical). Please do not edit the contents of this page. If you wish to revive any of these discussions, either start (https://en.wikipedia.org/w/index.php?title=Wikipedia:Village_pump_(technical)&action=edit&section=new) a new thread or use the talk page associated with that topic.

< Older discussions · Archives: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK, AL, AM, AN, AO, AP, AQ, AR, AS, AT, AU, AV, AW, AX · 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162

Contents

No Firefox favicon for section redirect
AutoEd
Help with css
Watchlist announcements leaving a gap
  Watchlist legend
List of contributors
Problem on WP:AFD
revision history statistics "link"
Should I raise a bug for this error?
|R parameter for magic words
Percent encoding
Conversion to PDF
Tech News: 2015-28
HTTPS by default
  Not such a difficult fix
  Who loses
  And another problem: No browser history!
  How about a 'in the clear' sub-wiki?
  We have had a discussion
  Please enable HTTP mode
  Note
Link to talk page in Mobile Wikipedia
Page ID
Force desktop version?
Arunanshu abrol thanking DumbBOT
API calls just starting throwing SSL/HTTPS (?) errors
Most frequently used words with 6 or more characters on the English Wikipedia
Content Translation, the new article creation tool is now available as a beta-feature
Images not showing up
file_get_contents on wmflabs?
Need some testers
Changing wiki strings
Revision scoring IEG goes for a second round
Moving article - history lossed
Nesting footnotes
How could a 2015 news article be cited in a 2013 version of Wikipedia article?
AfD issue?
User Contribution Search
Thanks given and received
Cannot link to other Wikis
"Back to top" link/button
Pending changes - bad link in dialog screen
Error 503 Service Unavailable
Tables
Can't edit Wikipedia mobile
WGT Baseball:MLB deleted

Loss of session data
File upload problems
New contribution?
Number of page watchers
Template failed to substitute on page creation
Is there a way to "walk" categories in WP?
Manage TemplateData
Cached?
Redirects and includeonly
Apparent glitch in mobile browser
Tech News: 2015-29
Admin Edit Flag
Interwiki transclusions
Nm in Keegan's contributions
Meta-tags
Page curation
My search box is wider today
How to properly hide/show rows in table for a template?
Template rename server effect
Phabricator bug: Registering a Phabricator account with an incorrect email address
Citation issues
My first mapping effort
Tfd2 and Tfm2
Green marker in watchlist for pages that have been changed
Grayish bold
Side question
Replacing talk pages: Why not just adapt an exist webboard?
Too much dependence on volunteers
Format price problem
Do you have ideas on how to improve Wikipedia?
Template using white text even when not specified
Encoded identifier in article to serve searches?
Bullets appear in the wrong place
Reproducing the problem below
The title creation blacklist had a minor glitch in functionality...
apostrophe problem
Pages moved without leaving redirects
"helpme-helped" makes text small
Protecting an entire article against automatic spell-miscorrection
Proposed software change: Show the reference list when section editing
Subdividing long lists alphabetically
CSS required to prevent layout from breaking
Various tools are down
Mobile editing
Problem with a reference
Article prefix/namespace bug
transwiki
Tech News: 2015-30
Is the Compact Personal Bar gone permanently?
Music
Music files
Article rename
Another watchlist proposal: Symbol to replace (0) when net effect is no change at all
POTD template: direction of wide image scrolling
Discrepancies in search result numbers
Template glitch
Change the "Edits by user" external tool link
Private incident reporting and tracking system for admins
French Wikipedia has Global contributions Special page
Wikipedia:Dispute resolution noticeboard/request not working
Description of articles traffic rankings
Proposal to create PNG thumbnails of static GIF images
Tag log
Database problem
Page number weirdness
edit button not showing in mobile wikipedia

photos in an info box
Thumbnail not displaying properly
Infobox image
Article creation improvements by WMF
Table formatting List_of_mayors_of_Bremen
Navboxes in mobile
Display oddity
Usage
Problem on WikiProject Physics page
Move shown twice
DIFF highlighting glitches ?
A tool for dimming references in diffs?
Coordinate Templates and maps
'Abd al-'Aziz al-Wafa'i
Nonexistent existing userpage
Wikipedia File Upload Wizard (Fair use files)
Search page meddling
Tool for fixing malformed wikilinks
Tech News: 2015-31
Single characters
history revision statistics (alternate tool)
Private drafts?
Infobox image/text justification problem
Issues over the past few days
edit count language
Moving over talk page
Default edit summary
Authorlinking in German template
citedoi templates
Reading Non-acsii characters via mwclient
Sorting search results
English Wikipedia is extremely slow
File upload problem
Citation now spam
What just happened to the watchlist?
Use of addresses such as: "https://en.wikipedia.org/wiki/Talk/Xxxx" or "https://en.wikipedia.org/wikitalk/Xxxx" for searchability etc.
Main page on mobile
Annotations in small images
Categorization
Search links not appearing
What does a Healthy Community look like to you?
Why get involved?
More information
Module:Citation/CS1 incorrectly adding pages using edition=revised to tracking category
Citations are weird
stats.grok.se broken yet again?
AfD Statistics
Wikimedia email
Upload file from WikiEditor...to Commons
Wikitables - unable to disambiguate links
Tech News: 2015-32

No Firefox favicon for section redirect

I see no W favicon in Firefox 39.0 tabs for redirects to sections, for example Wiki spam and others tested in Category:Redirects to sections and the five other languages listed there. I see the favicon in Wiki spam in IE, Chrome, Safari and Opera. In Firefox I see the favicon for Spamdexing (direct link to the page) and Spamdex (redirect to the page but not to a section). Do others have the Firefox issue? PrimeHunter (talk) 02:26, 3 July 2015 (UTC)

I couldn't see it for the Wiki spam page in 38.0.5. But then when I went to verify the version number, the browser auto-updated to 39.0 and now that page does show a favicon. Regards, ORANGE SUEDE SOFA (talk) 02:45, 3 July 2015 (UTC)

Are you still on the page saying "(Redirected from Wiki spam)" when you see the favicon in 39.0? We use url redirection now so if the page is reloaded then you get the redirect target Spamdexing#Wiki spam where I do see the favicon and no "Redirected from". PrimeHunter (talk) 02:51, 3 July 2015 (UTC)

Yes, I see the favicon on the "Redirected from..." page. When I reload the page, the "Redirected from..." disappears as expected and the favicon remains. ORANGE SUEDE SOFA (talk) 03:04, 3 July 2015 (UTC)

Thanks. If others aren't missing the favicon then I'm not filing it in Phabricator. PrimeHunter (talk) 09:55, 3 July 2015 (UTC)

AutoEd

Requesting Volunteers to correct my AuotoEd page

https://en.wikipedia.org/wiki/User:Silver_Samurai/common.js

According to this instruction

https://en.wikipedia.org/wiki/Wikipedia:AutoEd#Installation_guide

--Silver Samurai 05:56, 3 July 2015 (UTC)

@Silver Samurai: What you've done looks fine to me. Are you not seeing the "auto ed" item in the "More" dropdown menu? — *Mr. Stradivarius* ♪ talk ♪ 06:17, 3 July 2015 (UTC)

User:Mr. Stradivarius I am seeing it now.--Silver Samurai 06:41, 3 July 2015 (UTC)

Help with css

So today I finally started my css page. I've never done that before because I find css very confusing.

Here's what I'm trying to do: hide certain templates. I pulled some css from Template:Humor/doc, though I don't want to hide those templates in particular. I added one I did want to hide, which didn't work, and upon looking at the template codes I further edited it to this. It still didn't work, so I removed all code (because it might compromise my account, booga-booga.) So, anyone know why this is happening? Eman235/*talk* 13:43, 3 July 2015 (UTC)

@Eman235: I think your problem is that you need a comma after all but the last class selector (You're missing one after `.ombox-humorantipolicy`). That CSS code is essentially looking for a `not-a-forum` box *inside* a `humorantipolicy` box, which won't exist. Adding the missing comma should solve your problem. /~huesatlum/ 14:27, 3 July 2015 (UTC)

Haha! Brilliant. I missed the comma. It works now. Eman235/*talk* 15:09, 3 July 2015 (UTC)

Watchlist announcements leaving a gap

Since a few hours ago, there's a significant vertical gap left between the line containing "Clear the watchlist" and the "You have *n* pages on your watchlist..." line. All that with dismissed announcements and "Mark all pages as visited" button hidden with some custom JavaScript. Any clues? It used to be all neat and tidy. — Dsimic (talk | contribs) 04:54, 2 July 2015 (UTC)

Fixed, there were a couple of watchlist notices that had expired today. Sam Walton (talk) 14:06, 2 July 2015 (UTC)

Watchlist legend

- While we're on the subject, could someone update the watchlist *Legend* so we mortals will know what the circles and arrow and bullets and colors mean? And BTW, I've been meaning to ask for some time (though I haven't seen this in a week or two): changed-since-my-last-visit articles usually show up in a deep bold blue, but someones one or two up them are in a sort of grayish blue. Anyone want to explain (or, again, update the legend)? EEng (talk) 14:18, 2 July 2015 (UTC)

@Samwalton9: Thank you! — Dsimic (talk | contribs) 20:30, 2 July 2015 (UTC)
@EEng: Hm, I'm a bit confused as I see no such things (circles, arrows and different shades of blue) on my watchlist. Out of curiosity, which watchlist-related options are enabled in your preferences? Oh, and I have some custom CSS forcing bold page names for modified watchlist entries. — Dsimic (talk | contribs) 20:35, 2 July 2015 (UTC)

At *Preferences* > *Gadgets* I've got these two checked:

- Display green collapsible arrows and green bullets for changed pages in your Watchlist, History and Recent changes
- Display pages on your watchlist that have changed since your last visit in bold.

Sometimes they're green and sometimes they're blue, and now there are little green bullets sometimes. It's all very entertaining. EEng (talk) 21:11, 2 July 2015 (UTC)

What's even more confusing, I also have those two options checked, and really haven't seen any fancy watchlist inconsistency. Which skin do you use? I'm using the default Vector skin, while viewing everything in Firefox. — Dsimic (talk | contribs) 21:20, 2 July 2015 (UTC)

Vector, Chrome (just checked in IE11 and it's the same). EEng (talk) 21:35, 2 July 2015 (UTC)

Thanks. Hopefully others will use all this as debugging information. — Dsimic (talk | contribs) 03:15, 3 July 2015 (UTC)

You'll see collapsible items (the arrows) when you have the enhanced watchlist enabled. -- [[User:Edokter]] [[talk]] 20:41, 3 July 2015 (UTC)

Green indicates pages you haven't visited yet since they were updated (which is also explained at the top of your watchlist). An arrow indicates a collapsed item, which you can expand. -- [[User:Edokter]] [[talk]] 20:41, 3 July 2015 (UTC)

List of contributors

I recently made a post to the help desk asking "is there an easy way to get a list of all the users (and IPs) that have contributed to article X, and maybe even list them in order of number of edits to the page?" and I was directed here. I think there's an xTools page ([1] (https://tools.wmflabs.org/xtools-articleinfo/index.php?&lang=en&wiki=wikipedia)) that's supposed to accomplish this but it doesn't work for me. I need the list fairly soon because the page (QI (A series)) is up for deletion and the page history is a bit long to go through manually. — **Bilorv**(talk)(c)(e) 22:01, 3 July 2015 (UTC)

You could use the API's prop=contributors. Anomie⚔ 23:24, 3 July 2015 (UTC)

Thank you very much. — **Bilorv**(talk)(c)(e) 07:48, 4 July 2015 (UTC)

Problem on WP:AFD

The page WP:Articles for deletion has been vandalised somehow. The problem is in template {{Deletion debates}}, but I haven't been able to pin it down any further. JohnCD (talk) 14:19, 4 July 2015 (UTC)

Investigating. Jo-Jo Eumerus (talk) 14:22, 4 July 2015 (UTC)
If you were seeing a large red screen with the text "nice meme", that was added by 120.50.54.81 (talk · contribs · WHOIS (https://tools.wmflabs.org/whois/gateway.py?lookup=true&ip=120.50.54.81)) to Module:Dynkin. The module was subsequently added to a number of templates by Keastes (talk · contribs). Everything has now been reverted. SiBr4 (talk) 14:23, 4 July 2015 (UTC)

Looks like Keastes is back in control of their account (diff (https://en.wikipedia.org/w/index.php?title=Wikipedia:Administrators%27_noticeboard/Incidents&diff=prev&oldid=669924237)). Conifer (talk) 15:06, 4 July 2015 (UTC)

See recent history of Template:Hlist (https://en.wikipedia.org/wiki/Template:Hlist?action=history) for the problem specific to {{Deletion debates}}. --Redrose64 (talk) 17:30, 4 July 2015 (UTC)

revision history statistics "link"

this has been down for a very long time,(its important for articles like Dyslexia, Ebola/west Africa,..)is there any idea when it will be working? thank you--Ozzie10aaaa (talk) 22:12, 4 July 2015 (UTC)

The whole xtools suite appears to be in a state of flux. The basic problem is that the people who made it are no longer active. Per the recent watchlist notice, some are trying to assemble a new team to rewrite it from scratch. But it is not clear whether they will succeed or how long that would take. Out of curiosity, what statistics specifically are you looking for?--Anders Feder (talk) 05:16, 5 July 2015 (UTC)

dyslexia article...1 edits per user...2. bytes added per user--Ozzie10aaaa (talk) 11:56, 5 July 2015 (UTC)

There's an alternate tool available here (http://vs.aka-online.de/cgi-bin/wppagehiststat.pl). -- Diannaa (talk) 15:40, 5 July 2015 (UTC)

Should I raise a bug for this error?

Database error
From Wikipedia, the free encyclopedia
• Javascript-enhanced contributions lookup 0.2 enabled. You may enter a CIDR range or append an asterisk to do a prefix search.

A database query error has occurred. This may indicate a bug in the software.

Function: IndexPager::buildQueryInfo (contributions page filtered for namespace or RevisionDeleted edits)
Error: 2013 Lost connection to MySQL server during query (10.64.32.25)

All the best: *Rich Farmbrough*, 19:47, 26 June 2015 (UTC).

First of all, steps to reproduce would be welcome so someone else could try to see that error too. :) --AKlapper (WMF) (talk) 08:51, 29 June 2015 (UTC)

Search my contribs with namespace "Module" or "Mediawiki" if this fails to provoke an error (due to caching) try another rarely (or never?) edited namespace, or another user with many edits such as User:Koavf.
All the best: *Rich Farmbrough*, 15:54, 29 June 2015 (UTC).

It's slow but works for me, for example for you on TimedText which gave "No changes were found matching these criteria.", on a correctly looking page without error messages. PrimeHunter (talk) 16:06, 29 June 2015 (UTC)

I'm not surprised, the second time I did it with "Module" it worked, though whether this is a result of smart caching or variance in database load I cannot tell. Nonetheless I would think that the software could deal with these queries, either by chunking or increasing the timeout. All the best: *Rich Farmbrough*, 22:42, 29 June 2015 (UTC).

Here's another one:

- https://en.wikipedia.org/w/index.php?title=Special%3AWhatLinksHere&limit=5000&target=Template%3ACOIUL&namespace=4&invert=1

All the best: *Rich Farmbrough*, 20:21, 5 July 2015 (UTC).

|R parameter for magic words

{{PAGESINCATEGORY|Featured articles|R}} should give the number without commas. Strangely this is not working - 5209. Anyone know why? All the best: *Rich Farmbrough*, 15:51, 29 June 2015 (UTC)

Your syntax calls Template:PAGESINCATEGORY which hasn't implemented R. The magic word is `{{PAGESINCATEGORY:Featured articles|R}}` which gives 5209. PrimeHunter (talk) 15:55, 29 June 2015 (UTC)

Thanks, template fixed. All the best: *Rich Farmbrough*, 22:39, 29 June 2015 (UTC).

mw:Help:Magic words#Statistics shows several possible values of a second parameter, and if others than R are used then R can be a third parameter. If you want the template to be similar to the magic word then you could just pass everything on with `{{{2|}}}|{{{3|}}}`. It appears the magic word just ignores the extra parameters if they are empty. PrimeHunter (talk) 23:09, 29 June 2015 (UTC)

Thanks, good idea, done (and in other languages). All the best: *Rich Farmbrough*, 20:21, 5 July 2015 (UTC).

Percent encoding

I noticed some strange percent encoding, every time an IP made a mobile edit, ref names were getting another layer of % encoding - [2] (https://en.wikipedia.org/w/index.php?title=Marion%2C_Ohio&type=revision&diff=665010079&oldid=665006023). Is this a known bug, or a one off glitch we can ignore? All the best: *Rich Farmbrough*, 18:51, 5 July 2015 (UTC)

Most of those user's edits are not tagged mobile. And the one that is, doesn't show this problem. I'm guessing that the user's browser has an extension that is buggered up by installed extensions or something. —TheDJ (talk • contribs) 19:58, 5 July 2015 (UTC)

Looks more like vandalism to me. The IP labelled that edit as "fixed the page". The IP has been blocked by the way. Tvx1 20:08, 5 July 2015 (UTC)

The substantive edits seem ok. The IP was blocked as an open proxy, thanks for pointing the block out. All the best: *Rich Farmbrough*, 20:17, 5 July 2015 (UTC).

I assumed they had overridden the tag - I guess that can't be done? Perhaps they were trying not to make that problem. Anyway its something to watch out for. All the best: *Rich Farmbrough*, 20:17, 5 July 2015 (UTC).

Conversion to PDF

A reader reported problems converting two articles to PDF. I just tried each of them and can confirm the same problem. In each case I received the following error:

Status: Bundling process died with non zero code: 1

The articles:

- Jacobean era
- Animism

-- S Philbrick (Talk) 15:43, 3 July 2015 (UTC)

This is filed as T104708 (https://phabricator.wikimedia.org/T104708). HTH, --Elitre (WMF) (talk) 16:03, 3 July 2015 (UTC)
WMF ops found the cause of the problem, I've deployed the fix and it all seems better now. --Krenair (talk • contribs) 17:18, 3 July 2015 (UTC)

Another email was sent to OTRS reporting problems with PDF renderings. The two articles mentioned were:

- Poseidon - Status: Rendering process died with non zero code: 1
- Twelve Olympians -Status: ! LaTeX Error: Something's wrong--perhaps a missing \item.

I had hoped to respond that the problem has been resolved but I tried both of these and both failed. I placed the error message after the article name -- S Philbrick (Talk) 14:23, 4 July 2015 (UTC)

@Sphilbrick: Those articles seem to have different error messages? "Poseidon" says "Rendering process died with non zero code: 1" which is phab:T94308. "Twelve Olympians" says "Status: ! LaTeX Error: Something's wrong--perhaps a missing \item." which welcomes a bug report. --AKlapper (WMF) (talk) 08:28, 6 July 2015 (UTC)

Good catch, I glossed over the single word difference in the error message. This means that the prior problem, which is reported solved, is solved and this is a different issue. You identified the bug report for the first error. The second one does generate a different error. It has already

read been reported as T88890 (https://phabricator.wikimedia.org/T88890). That report talks about a problem working with collections. I added this specific instance to show that it can be generated with a single article.-- S PHILBRICK (TALK) 13:27, 6 July 2015 (UTC)

Tech News: 2015-28

Latest tech news from the Wikimedia technical community. Please tell other users about these changes. Not all changes will affect you. Translations are available.

Recent changes

- On the mobile site, you now see more information when you search for a page. It now shows the description from Wikidata. [3] (https://phabricator.wikimedia.org/T94713)

Problems

- The code of long pages is not colored any more. You may see this problem on the pages of long gadgets. [4] (https://phabricator.wikimedia.org/T104109)

Changes this week

- The new version of MediaWiki will be on test wikis and MediaWiki.org from July 7. It will be on non-Wikipedia wikis from July 8. It will be on all Wikipedias from July 9 (calendar).
- The "Page information" tool shows how many users watch the page. You can now see how many are active. [5] (https://lists.wikimedia.org/pipermail/wikitech-ambassadors/2015-July/001205.html) [6] (https://phabricator.wikimedia.org/T51506)
- You can now translate articles into English with the new translation tool. [7] (https://phabricator.wikimedia.org/T94123)

Meetings

- You can join the next meeting with the Editing team. During the meeting, you can tell developers which bugs are the most important. The meeting will be on July 7 at 19:00 (UTC) (http://www.timeanddate.com/worldclock/fixedtime.html?hour=19&min=00&sec=0&day=07&month=07&year=2015). See how to join.

Tech news prepared by tech ambassadors and posted by bot • Contribute • Translate • Get help • Give feedback • Subscribe or unsubscribe.

15:13, 6 July 2015 (UTC)

HTTPS by default

Hi everyone.

Over the last few years, the Wikimedia Foundation has been working (http://blog.wikimedia.org/2013/08/01/future-https-wikimedia-projects/) towards enabling HTTPS by default for all users, including anonymous ones, for better privacy and security for both readers and editors. This has taken a long time, as there have been different aspects to take into account. Our servers haven't been ready to handle it. The Wikimedia Foundation has had to balance sometimes conflicting goals, having to both give access to as many as possible while caring for the security of everyone who reads Wikipedia. This has finally been implemented on English Wikipedia, and you can read more about it ~~[link to blog post here]~~ here (https://blog.wikimedia.org/2015/06/12/securing-wikimedia-sites-with-https/).

Most of you shouldn't be affected at all. If you edit as registered user, you've already had to log in through HTTPS. We'll keep an eye on this to make sure everything is working as it should. Do get in touch with us if you have any problems logging in or editing Wikipedia after this change or contact me if you have any other questions. /Johan (WMF) (talk) 12:43, 12 June 2015 (UTC)

> There's a blog post (https://blog.wikimedia.org/2015/06/12/securing-wikimedia-sites-with-https/) at the Wikimedia Foundation blog now. /Johan (WMF) (talk) 13:09, 12 June 2015 (UTC)
>
> > To Johan (WMF): – You *have* to know what a real drag this is. Not only do I want a CHOICE in the matter, and would continue to choose HTTP as long as the edit summary field's autofil function does not work when I'm on the HTTPS server, you should also consider what Redrose64 said above, that some users are unable to use HTTPS connections. The part in the blog post about "all logged in users have been accessing via HTTPS by default since 2013" is just not true, either. We've been given a choice up until now, and I for one do not want to give that up. I want to be able to CHOOSE whether or not I'm on the HTTP server or the HTTPS server. – *Paine* 14:21, 12 June 2015 (UTC)
> >
> > > Yes, we do know. The answer I was given when I asked about this is that any form of opt-out would also leave potential security risks in our implementation which make it difficult to safeguard those who do not opt-out. Because of this, we've made implementation decisions that preclude any option to disable HTTPS, whether logged in or not. This renders the current opt-out option ineffective, and the option will be removed at a later date after we've completed the transition process. /Johan (WMF) (talk) 14:27, 12 June 2015 (UTC)
> > >
> > > > You have had to use HTTPS to access the site when logging in as it's been used for the login process, though. /Johan (WMF) (talk) 14:30, 12 June 2015 (UTC)
> > > >
> > > > > It's evidently a weighty issue. And I do realize that I don't edit WP in a vacuum, that I must eventually accept this situation for the good of all. And frankly, I don't have a problem with having to stay on HTTPS as pertains to the "big picture". My problem is very basic and concerns the fact that I no longer have a drop-down list from which to pick my edit summaries, because that function is thwarted by my IE-10 when I am on any HTTPS server. If that little quirk could be fixed, I'd be a happy camper whether I'm on a secure server or not. – *Paine* 15:47, 12 June 2015 (UTC)
> > > > >
> > > > > > I'm not very familiar with IE myself, but I'll ask around and see if anyone knows a simple fix. /Johan (WMF) (talk) 16:12, 12 June 2015 (UTC)
> > > > > >
> > > > > > > @Johan (WMF): IE10 won't enable autocomplete on HTTPS pages when the "Cache-Control: no-cache" HTTP header is set (which Wikipedia does). Changing it from "no-cache" to "must-revalidate, private" would allow autocomplete, but may have other unintended consequences. --Ahecht (TALK PAGE) 16:34, 12 June 2015 (UTC)
> > > > > > >
> > > > > > > > @Paine Ellsworth: It seems like IE 11 does not have this problem, and all users would eventually be required to update to it by the end of the year (by Microsoft). Did you try IE 11? Tony Tan · talk 02:09, 14 June 2015 (UTC)
> > > > > > > >
> > > > > > > > > Yes, Tony Tan, I upgraded to Win8.1 and IE-11 yesterday and was pleased to pass it on that it has given me back what I had lost with the older browser and Windows software. Thank you very much for your kind thoughts and *Best of Everything to You and Yours!* – *Paine* 02:26, 14 June 2015 (UTC)

I also see I am struck with using HTTPS, which is nuisance and a bother as I longer have a drop-down list from which to pick my edit summaries. How can a drop-down list be re-implemented? It was the only degree of automated help we had in what is otherwise an unfriendly article editing environment. Hmains (talk) 17:44, 12 June 2015 (UTC)

> So how do I use the website in http then? I do not want extra security to protect me. I don't need protecting. This is a nonsense. Why am I being forced to use https even though I don't want to use it? There was an opt out. The opt out has been removed despite the fact that those using the opt out very clearly want to opt out. — Preceding unsigned comment added by 86.18.92.129 (talk) 19:46, 12 June 2015 (UTC)

> > Hi, the reason explanation I've been given is that any form of opt-out would also leave potential security risks in our implementation which make it difficult to safeguard those who do not opt-out. /Johan (WMF) (talk) 19:53, 12 June 2015 (UTC)

> > I'll try to figure out if there is a solution to that, Hmains. /Johan (WMF) (talk) 19:53, 12 June 2015 (UTC)

> > > Johan (WMF), Re: *"the reason explanation I've been given is that any form of opt-out would also leave potential security risks in our implementation which make it difficult to safeguard those who do not opt-out"*, would you be so kind as to ask for a one-paragraph explanation as to why they believe this to be true and post it here? Not a dumbed-down or simplified explanation, but a brief, fully technical explanation for those of us who are engineers? Thanks! --Guy Macon (talk) 20:49, 12 June 2015 (UTC)

> > > > Sure. Just so you know, they're getting a lot of questions at the moment, as well as handling the switch for the hundreds of Wikimedia wikis that aren't on HTTPS yet, but I'm passing on all questions I get that I can't answer myself. /Johan (WMF) (talk) 21:18, 12 June 2015 (UTC)

> > > > > The engineering-level explanation is that in order to help prevent protocol downgrade attacks, in addition to the basic HTTPS redirect, we're also turning on HSTS headers (gradually). The tradeoff for HSTS's increased protections is that there's no good way to only partially-enforce it for a given domainname. Any browser that has ever seen it from us would enforce it for the covered domains regardless of anonymous, logged-in, logged-out, which user, etc. Once you've gone HSTS, opt-out just isn't a viable option. /BBlack (WMF) (talk) 21:56, 12 June 2015 (UTC)

> > > > > > @Jason Quinn: see the answer above. /Johan (WMF) (talk) 22:12, 12 June 2015 (UTC)

> > > To Johan (WMF): I don't see what the problem is: create a cookie named something like `IAcknowledgeThatHttpIsInsecure` which can be set from a dedicated page: if this cookie is set, do not send the `Strict-Transport-Security` (HSTS) header and do not force redirect to HTTPS. Yes, people who have received the `Strict-Transport-Security` header will get a browser error, but I assume all browsers that implement HSTS allow some way for the user to manually override or ignore it (something like "I know what I'm doing", then set a security exception); and the users can be warned in advance on the dedicated page that sets the cookie. If you're afraid an attacker will set the cookie on an unsuspecting user (through a fake Wikipedia page) and thus bypass HSTS, please note that (1) this attack always exists anyway, because an attacker who can do this can setup a fake HTTP `wikipedia.org` proxy domain anyway (in both cases, it will impact those users who did not receive the HSTS header), and (2) you can mitigate the attack by letting the cookie's content contain a MAC of the client's IP address (or some other identification string), with a MAC key that Wikimedia keeps (and the cookie is honored only if the MAC matches). You might also display a warning in the HTML content if the cookie is set, reminding of its existence and impact, and giving a link to remove it should the user change their mind. The performance cost of all of what I just described should be completely negligible in comparison with the performance cost of doing HTTPS in the first place. And this should all be very simple to implement. **On a personal note:** I promise to donate 150€ to the Wikimedia foundation (adding to the 100€ I donate about once a year) if and when a way to access it through HTTP using the former URLs is brought back; conversely, until this happens, I will be too busy to consider how I can work around this inconvenience to contribute either financially or by editing articles. (I could also go on to emphasize how, as a cryptographer, I think the idea of forcing users to go through HTTPS to read publicly accessible and publicly editable information is absolute idiocy, but the cryptophile zealots have made up their mind already.) --Gro-Tsen (talk) 19:43, 13 June 2015 (UTC)

> > > > Keyed MAC of client IP address is not going to work due to dynamic IPs that change (And I'm not sure that there exists any other unique identifier that would be appropriate. Keep in mind, for your scheme to work, the browser cannot receive an HSTS header even once). Note that deleting an HSTS setting from your browser is actually much more hidden then you'd normally think, and are generally not meant to be user overridable. While you're correct that HSTS cannot prevent a malicious proxy if the user has never visted wikipedia before (unless we do HSTS preloading, which we do not yet), your scheme weakens the protection of HSTS, since a malicious proxy only has to set a cookie for wikipedia, not necessarily catch the user at the first visit. Furthermore, in order for the redirect not to take place, the cookie must be non-secure. Hence the malicious proxy might as well just pretend to be some fake subdomain, e.g. http://fake.wikipedia.org (Which since its fake, does not have HSTS, unless we set the includeSubDomains flag for HSTS, which we don't currently, and would prevent us from ever hosting a non-secure service on any subdomain), use some method to load traffic from that address (easy), and then set your *IAcknowledgeThatHttpIsInsecure* cookie with the domain field set to .wikipedia.org. Last of all, your scheme is also incompatible with HSTS preloading, which presumably the WMF is eventually going to pursue. Bawolff (talk) 00:53, 14 June 2015 (UTC)

> > > > > OK, I'll give up on trying to solve other people's problems with HTTPS and focus on mine: to this effect, do you (or anyone else) knows if there at least exist some reliable transparent Wikipedia mirror on HTTP (perhaps something like "wikipedia-insecure.org") that allows both reading and editing and that I could use (by spoofing my DNS to point there) without the trouble of setting up my own? (I hope we can agree that a mirror served under a different domain cannot weaken security since anyone can set up such a thing.) I'll find a way to disable HSTS on my browser somehow. --Gro-Tsen (talk) 23:02, 14 June 2015 (UTC)

It's worth giving some background here to understand the need for security. One of last year's revelations was that Wikipedia editors *were* being targeted by the NSA. So if you weren't using HTTPS (and probably even if you were), you were likely helping to build a database profile on your reading habits. But worse, your e-mail and other communications were probably also targeted for follow-up simply because you edit Wikipedia. What difference does it make? Nobody in the general public knows! The collected information is used in secret fashion in secret ways by undisclosed people. But there are real dangers to you. Supposedly, the information is being used only for national security related to terrorism. That's not true, however, because it is known from the same leaks that it is being used for more than that, for instance, in the war on drugs. And, it is also known that collected information is sometimes abused by those who have access to it for personal reasons. The use could also include (and probably is) helping to decide whether you get security clearance for future dream job. it could potentially even be used to sabotage a hopeful's political career or in general help silence people with oppositional points of view. In other words, this information has the potential to be used by people now or in the future to negatively affect your life and destiny without you even knowing. The WMF has decided (and rightfully so) that there's a need to protect users from dangers that they might not even be aware of. When it comes to this, many people say things like "I'm not doing anything wrong" or "I've got nothing to hide" but the problem is that you can't say you're doing nothing wrong because it's third parties who determine that, not you. And you *do* have stuff to hide even if you are completely a law-abiding citizen. This issue that affects you even if you think it doesn't. People are talking above about certain countries that do not allow HTTPS and how IP users there should be not be forced to use HTTPS because Wikipedia would be blocked

for them. Well, those are great examples where governments being able to see what you are reading could get you arrested, imprisoned, or worse. The use of HTTPS is only a minor step in combating the abuse of government-level surveillance but it's a step in the right direction. @Johan (WMF), it'd be interesting to know *why* the implementation cannot safely handle an opt-out because naively I don't see why the one should affect the other. Maybe this exposes a flaw in the implementation. Jason Quinn (talk) 21:17, 12 June 2015 (UTC)

> Hi Jason Quinn, thanks. I'm passing on the question to someone better suited to answer it than I am. /Johan (WMF) (talk) 21:20, 12 June 2015 (UTC)
> On January 12, 2016, Windows 7 users will be required to install Internet Explorer 11 and Windows 8 users will be required to update to Windows 8.1 anyway, so you don't need to worry about the autocomplete problem in IE10. That problem doesn't occur in IE11. GeoffreyT2000 (talk) 21:26, 12 June 2015 (UTC)

> > Wikipedians were **NEVER** targeted by the NSA, why would they be? I don't know where you people are getting your information from and if some wikipedian came along and said that s/he was being targeted, then s/he was either being paranoid (like 90% of americans) or s/he is doing soemthing "illiegal" so its the best interest of wikipedia to report that person to NSA, not ENFORCE this stupid idea....Again Wikipedia is an **INTERNATIONAL** website, its NOT only for AMERICA....why should the rest of the world have to pay for the fears of a few paranoid psychopaths that are better off in jail..oh and BTW, HTTPS has and will **NEVER** be secure, the "s" in https never stood for secure...@Jimbo Wales:, Why would you allow this?--*Stemoc* 21:43, 12 June 2015 (UTC)

> > > At the right is the main slide itself so you and others can decide for themselves what it means. The slide explicitly uses Wikipedia as an example of the websites that they are "interested in" and confirms that they are interested in "typical users" of such websites. Given the context of the slide (exploiting HTTP for data collection), it is unreasonable to assume readers and editors were not being targeted. We *all* were targeted and all our traffic to and from Wikipedia would have been caught up in the named NSA collection programs. It would be naive to think otherwise. If there is one thing that's been learned in the last year, it's that "if it can be done, it is" kind of summarizes what's been going on and "mitigated" does not described their collection techniques. As for other countries being denied access by the global removal of HTTP support, that is a point that should be debated. But I already mentioned that there are countries were the use of HTTP might literally allow Wikipedia readers to be executed for readings the "wrong" stuff. The meaning of a "free" encyclopedia would have to be discussed and the dangers of access in these countries would have to be considered and weighed in such a debate. And, regardless of how you perceive the US, it's possible the US could become as bad. Jason Quinn (talk) 22:30, 12 June 2015 (UTC)
>
>

> > > It is certainly a bit of a backtrack by @Jimbo Wales:.Blethering Scot 22:43, 12 June 2015 (UTC)

> > The real win here (imo) is making Firesheep style attacks totally impossible and thwarting non-state sponsored, and lower budget state sponsored adversaries. One assumes that the NSA will probably just slurp up the unencrypted inter-data center links (For those of you not close enough to use eqiad directly. Imagine a world where the sum of human knowledge fully deployed IPSec). Given the funding level of the NSA, I expect that they probably have traffic analysis capabilities to be able to tell who is visiting a page of interest (especially for a site like wikipedia, which imo seems like the perfect target for a traffic analysis type of attack against a TLS secured connection). However https does make it much harder to simply collect it all, and any measure that increases the cost of ubiquitous surveillance should be applauded. Bawolff (talk) 22:50, 12 June 2015 (UTC)

> > > All I see Jason is a bunch of American websites.....Mate, if NSA want to spy on you, it **WILL** SPY on you, you don't have to eff up wikipedia for them to stop and basically, by forcinghttps onto the wikipedia, would you not think that it will make NSA more interested? because only a person with something to hide would do this ..So Jimmy loses his battle with NSA in terms of NSA and this is what he comes up with? moving to https which honestly is just as secure as http...After this was defeated last year, i honestly felt like we lived in a democracy where the voice of the people was heard and adhered.....back to communist wikipedia we go..yeah Jason, executed for reading the wrong stuff on wikipedia like How to build a Bomb or How to join ISIS.....oh right, we don't have those pages cause wikipedia is NOT a terrorist organization...--*Stemoc* 22:59, 12 June 2015 (UTC)

> > > > (a) Non-Americans arguably have got more to fear from NSA surveillance; the legal framework allows for the collection of great swathes of foreign data. (b) The decision was made by Wikimedia, which is in no way a democracy. (c) Do actually read up on the issues you're arguing. Alakzi (talk) 23:29, 12 June 2015 (UTC)
> > > > Yeah, you really shouldn't let your anger and/or frustration allow such bullshit from your fingers and keyboard, *Stemoc*. "Communist Wikipedia"? no more than an airline practices communism when they check for bombs and weapons as we board – no more than when we have to pass through a building security point that helps to protect us while we're on the premises - is it communism to own a .357 and be ready to shoot a criminal who tries to steal from you? or to hurt your loved ones? Privacy, security, if you don't try to work with structures that protect them, then you're no better than the criminal, terrorist or agency that tries to circumvent them. *Best of Everything to You and Yours!* – *Paine* 00:03, 13 June 2015 (UTC)

> > > > > Calm down lady, this is just an Encyclopedia, not your ebay, paypal, bank account or your social networking sites where privacy is a MUST NEED for safety reasons.. the MAIN reason this site was created was to allow users to browse and edit anonymously so no one really knows your true identity or location, if you are using your real name and stuff, I'd advice you to invoke the 'Vanish' policy and start anew or get your account renamed, I think people keep forgetting that this is NOT like every other site they visit, infact wikipedia is based on facts and if you are scared to write down fact on articles because you fear the NSA then i really really pity you... only crooks fear the government....let that be known...and p.s, I'm brown and I don't give a shit about the NSA...as usual, the wiki revolves around America...pathetic.--*Stemoc* 02:47, 13 June 2015 (UTC)

> > > > > > @Stemoc: Out of curiosity, what do you think about the following hypothetical situation: Someone (Lets say Alice) thinks she might have <insert weird disease here>. Alice wants to look it up on Wikipedia, but is worried that her ISP is tracking which websites she visits, and will sell the information to her insurance company (or whomever else is the highest bidder), who in turn will jack up the price of her insurance beyond what she can afford, on mere suspicion of having the disease, even if she doesn't have that. Is that a legitimate reason to want/need privacy when browsing Wikipedia? You may trust the government (For some reason), but do you really trust your ISP? What about the guy sitting across the room at the starbucks on the same wifi network? Bawolff (talk) 06:12, 13 June 2015 (UTC)

Bawolff, again, another "american" problem....I have an IDEA, why not make a us version for https?, brilliant now, e.g, anyone that wants to be logged in on https, log in at https://us.en.wikipedia.org and everyone else at the old link at http://en.wikipedia.org, this will solve the problem once and for all, why "force" everyone onto https, its the same as pushing everyone over the cliff and telling them to swim instead of building a bridge to get across, those who can't swim or having health (ISP) problem will surely drown..I fought this the last time it happened and I will fight it yet again..--*Stemoc* 11:43, 13 June 2015 (UTC)
+1. Live in a country with universal health care...Or has privacy laws...I am an IT professional with a Computer Science Degree and 30+ years of experience. I know the implications of not using HTTPS, an I also know the NSA can bypass that easily if they care to. This (not allowing an opt-out) is total garbage and a false sense of security...ꓱƎ ∀ H Ↄ I ꓤ (talk) 11:56, 13 June 2015 (UTC)

Now cut that out, buddy, or I'll hit you with my purse! Hey, waitasec – how did you know I'm a "lady"? You been hackin' into my HTTP??? – *Paine* 12:46, 13 June 2015 (UTC)

Little old me? hacking? NEVAH!.... .--*Stemoc* 17:01, 13 June 2015 (UTC)

@Bawolff: We're not done with all of our plans for securing traffic and user privacy. This will be covered in deeper detail in a future, engineering-focused blog post after the initial transition is complete. But just to hit some highlights in your post: we do have an active ipsec (https://phabricator.wikimedia.org/T81543) project, which is currently testing on a fraction of live inter-DC traffic. We're also looking into what we can do for some forms of traffic analysis, e.g. mitigating response length (https://phabricator.wikimedia.org/T92298) issues. We plan to move forward as soon as possible on heavier HTTPS protection mechanisms like HSTS Preloading, HPKP, etc as well. We're committed to doing this right, we're just not done implementing it all yet :) -- BBlack (WMF) (talk) 01:53, 13 June 2015 (UTC)

@BBlack (WMF): I appreciate there's more to come, and I'm happy to see that its (finally) happening. However I think its important to give our users the full picture, the good, and the bad. HTTPS is great for confidentiality and integrity. Its somewhat ok for providing privacy, particularly against a weak adversary, and it makes bulk matching of packets against fixed strings in the body of the request impossible (Which is quite important given the selective censorship threat wikipedia faces). But its questionable how well it would hold up against a powerful nation state actor trying to simply de-anonoymize you. It certainly wouldn't hold up against a targeted attack, and its questionable if it would prevent a more broad attack (albeit it would certainly make a broad attack quite a bit more expensive to pull off). I'm also quite doubtful you can really foil traffic analysis with padding TLS sessions, unless you use extreme amounts of padding, far past what is acceptable performance wise. p.s. The ipsec project link is limited to those in the WMF-NDA group, so I can't see it (I'm in the security nda group only). However I can certainly see in puppet that IPSec is enabled on a small number of servers, and I noticed it was mentioned when I was reading the WMF quarterly report. Bawolff (talk) 03:03, 13 June 2015 (UTC)
@BBlack (WMF): It is great to see that the WMF is finally switching to HTTPS by default. I look forward to seeing Wikipedia send HSTS (includeSubDomains, long max-age, preload) and HPKP headers! However, phab:T81543 seems to have restricted access. Thanks, Tony Tan · talk 02:39, 14 June 2015 (UTC)

One of the nice things I just noticed that is really nice is that ru.wikipedia.org has an A+ on the SSL labs test [8] (https://www.ssllabs.com/ssltest/analyze.html?d=ru.wikipedia.org&latest). Here's to looking forward to that for all Wikimedia domains, once HSTS is turned up :D Bawolff (talk) 05:20, 14 June 2015 (UTC)

Not such a difficult fix

Just want to make sure that everyone catches what contributors TTO (at phab:T55536) and GeoffreyT2000 (above) have been kind enough to share with us. Several of the above users may be happy to hear that I can confirm what TTO and GeoffreyT2000 say about Win8.1 and IE-11. I just upgraded, and the new software thus far seems to work a lot better under HTTPS than my old Win8.0 and IE-10 did. Forms do indeed autofil, which means that my old drop-down boxes with my edit-summary choices do show up again. I still sympathize with all the users above who feel they've lost something with this change, however, like I said, we don't edit in a vacuum any more than we become passengers on aircraft all by ourselves. As an analogy, airport security can be a real hassle and a serious time cruncher on occasion, but compare that to what has happened, and still could happen, and there should be none of us who would not want that security to keep our flying times safe. Same for the conversion to HTTPS – it is quite the hassle for some, but the very real need to protect our privacy and security is an overwhelming priority, in my humble opinion. So, /Johan (WMF), you don't have to find an IE fix for me, and I greatly appreciate the fact that you said you would! I also deeply thank the rest of you for your enlightening responses here. *Best of Everything to You and Yours!* – *Paine* 23:32, 12 June 2015 (UTC)

Thank you. I'll still at least ask around to see if there's anything I can do. We want editing Wikipedia to be as simple as possible, no matter which browser people use. If one is OK with upgrading to IE 11, that's probably the best solution, though. /Johan (WMF) (talk) 01:25, 13 June 2015 (UTC)

So, here's what I got on this issue so far. Yes, there appears to have been an open Phabricator ticket since 2013 reporting this issue, and no, given the number of tickets, the team that dealt with the transition wasn't aware of it. We'd obviously have preferred to be. Sorry, and I really mean it. Causing trouble for people who edit Wikipedia is the opposite of what we want to achieve. We're still in the process of transitioning (English Wikipedia was one of the first to switch over, and there are more than 800 Wikimedia wikis) and I haven't found an easy fix so far (except for upgrading to Internet Explorer 11), as this isn't so much a bug as how Internet Explorer 10 intentionally behaves. The team will be able to focus more on this as soon as the HTTPS transition is complete. We're not ignoring the issue. /Johan (WMF) (talk) 12:10, 16 June 2015 (UTC)

This broke my bot :(I'm using RestClient (https://github.com/rest-client/rest-client) library to make API requests, and it apparently is unable to verify the certificate. Getting the error `SSL_connect returned=1 errno=0 state=SSLv3 read server certificate B: certificate verify failed (RestClient::SSLCertificateNotVerified)` Surely that's an issue on my end? I can force it to not verify the certificate but then what's the point of using HTTPS? — *MusikAnimal* talk 18:32, 13 June 2015 (UTC)

I took a quick look, and it seems that this library has a way to pass to the SSL library the CA certificates to be used for verification. It probably just doesn't have a default set of CA certificates. The solution would be to give it a copy of the correct root certificates to use. --cesarb (talk) 21:26, 13 June 2015 (UTC)

Who loses

@Johan (WMF): Hi, while you are here I would like to have something specific clarified. As always with these sorts of major changes, most people win and some people lose. I personal am iffy about the distribution of relative ideological and technical interest and need for this particular project, but I accept that that merely puts me in the middle of the Wikipedian spectrum, between people like TomStar81 who wants nothing to do with the ACLU and people like Jason Quinn who thinks it keeps us from being roasted on an open flame.

However in these sorts of changes I care less about who wins, because that's obvious. I can read the spam-ish blog post to find that out. I am more interested in the question: who loses?

Who does HTTPS hurt? Can we come to an understanding of this? Surely every change, no matter the size, hurts some stakeholders. **Res**Mar 03:58, 13 June 2015 (UTC)

1. Can someone clarify what is going on with the IE 10 issues? Was the WMF aware of this problem? Is it really that significant?
2. Can someone clarify what the effect will be in mainland China? Can you quantify the impact there?

Thank you. **Res**Mar 04:00, 13 June 2015 (UTC)

> Hi, good question that deserves a good answer, not just what I can come up with on the top of my head. I'll ask around about a few things to make sure I (or maybe someone else, I'll spend much of this weekend travelling) can reply properly. /Johan (WMF) (talk) 04:19, 13 June 2015 (UTC)
>
> > Great! Thank you. I think this discussion so far has been high on posturing, low on content (speaking about the community response here), and I'd love to see a frank cost-benefit analysis from the WMF on this matter, and an associated community critique. After all, *this* is the communication that the volunteers so crave. Not, frankly, blog announcements. ResMar 04:44, 13 June 2015 (UTC)
> >
> > > I'd also like to see my transparency on WMF's the analysis. Everything seems to be shrouded in unnecessary secrecy. On the subject of China. I'm not that familiar with the situation, but according to https://en.greatfire.org/search/wikipedia-pages - ~~HTTPS is currently not blocked~~ There seems to be conflicting info on if HTTPS is blocked. The greatfire website says https is not blocked, but there actual test data seems to suggest that both normal http and https on zh is blocked starting may 19 [9] (https://en.greatfire.org/https/zh.wikipedia.org) (The switchover for zh to https happened on June 9, so change in blocking status seems unrelated) but en is fine (both https and non-https). There are about 324 pages that are censored on the HTTP version, mostly on zh, however on en we had Students for a Free Tibet, Tiananmen_Papers, Tiananmen_square_massacre, Tibetan_independence_movement blocked. Switching to HTTPS forces china to decide either to block all of wikipedia or none of wikipedia (Possibly they can distinguish between languages and block say all zh, but not en. I'm not that familar with SNI, but my impression is the domain is sent in the clear). FWIW, greatfire strongly advocates switching to https on zh wikipedia [10] (https://en.greatfire.org/blog/2013/jun/wikimedia-foundation-says-it-doesnt-hold-chinese-readers-any-less-regard-we-disagree), although they are obviously a special interest group that believes Chinese censorship needs to be fought tooth and nail. I imagine the situation is similar for Russia, which rumor had (Although I've not seen direct sources for this) was trying to censor pages related to Ukraine on ru, but can't anymore due to https. The other impact, is that it makes harder (but certainly not impossible depending on their traffic analysis capabilities) for China to generate lists of people who try to visit certain politically sensitive topics (Its unclear if they actually do that. I haven't heard of any evidence that they do, but it wouldn't surprise me). Other potential things to keep in mind, in the past China has DDOS'd websites (GitHub) that host material China finds objectionable, but cannot be censored selectively due to HTTPS and are too popular to block outright (However, I consider it very unlikely they would do something like that to Wikipedia. Wikipedia has a low enough popularity in China, that they would probably just block it totally if they decided to do something about Wikipedia). Bawolff (talk) 05:18, 13 June 2015 (UTC)
> > >
> > > > Regarding secrecy, or at least part of it: yeah, we didn't really enjoy springing this on the community, though the WMF has publicly been talking about the intent to switch to HTTPS for the past years. The reason we didn't say anything about the specific deadlines or make public the transition until it was in progress was because public statements opened us to possibility of man-in-the-middle attacks. Letting everyone know meant letting bad actors, so to speak, know our plans and timeline. We couldn't have this debate in public without telling the world what we intended to do, which could have compromised the security and privacy of readers and editors in certain areas. We'd have preferred not having to worry about that, obviously. /Johan (WMF) (talk) 19:33, 16 June 2015 (UTC)
> > > >
> > > > > @Johan (WMF): But this discussion and these plans (https://phabricator.wikimedia.org/tag/https-by-default/) were open and public, where any "bad actors" could surely have followed them. Surely that workboard was missing an item relating to fixing bots that didn't operate on wmflabs.org. I can only do so much to stay tuned to such things, and a proactive heads up, perhaps by email, would have been appreciated. I asked about this last December on the Village Pump, and never got a response. How am I supposed to know about venues such as m:HTTPS, where I might have gotten help last December? Wbm1058 (talk) 16:50, 20 June 2015 (UTC)
> > > > >
> > > > > > An example I've been given is that not knowing our time plan made it much more difficult to e.g. hack DNS and traffic at a border, proxy traffic as HTTPS back to us but make it seem to everyone they're connected to us, as HSTS support in modern browsers will prevent the downgrade and warn about it. I'd have loved to be able to give everyone this would cause trouble for a heads up, and we do understand it has caused more work for people we don't wish to cause any unnecessary work for. We'd definitely have preferred to not found ourselves having to choose between either, as we saw it, putting user security in certain areas at risk or not having proper, open communication.
> > > > > > Are you still having the problems you had last December? /Johan (WMF) (talk) 13:37, 22 June 2015 (UTC)

@Resident Mario: Other people that HTTPS could potentially hurt which we know about (Personally I think this is an acceptable hurt): People who use IE6 on windows XP will not be able to view any page on wikipedia. (IE6 on XP is incompatible with modern best practices for HTTPS). ~~People on very old browsers which don't support SNI (e.g. Android 2.3.7, IE 8 / XP, Java 6u45), will get a certificate error when visiting a sister project (But wikipedia will be fine).~~ Bawolff (talk) 20:02, 13 June 2015 (UTC)

> @Bawolff: Sounds reasonable. **Res**Mar 20:21, 13 June 2015 (UTC)
> @Bawolff: The Wikimedia certificate uses subjectAltName, not Server Name Indication. SAN is supported by IE6. LFaraone 05:27, 14 June 2015 (UTC)
>
> > @LFaraone: IE6 doesn't work because it only supports SSLv3, and we require at least TLS1.0 (To prevent downgrade attacks/POODLE). We use both subject alt name, and SNI and wildcard certs. ~~If no SNI is sent you get a certificate for *.wikipedia.org with an alt name of wikipedia.org. Which is great if you're browsing wikipedia. Not so great if your browsing wiktionary.org.~~ Bawolff (talk) 05:45, 14 June 2015 (UTC)
> >
> > > @Bawolff: Browsing wiktionary.org works fine even if the browser doesn't send SNI. If the SNI is absent, the server sends a diffenert certificate whose subject alternative names include domain names of all sister projects. 191.237.1.8 (talk) 06:42, 14 June 2015 (UTC)
> > >
> > > > Oh, you're absolutely right, users get a uni cert when they don't have SNI. I saw the SNI behaviour of switching certificates and just assumed it would be broken without SNI. My bad. Bawolff (talk) 11:09, 14 June 2015 (UTC)
> > > >
> > > > > @Bawolff: I just checked my IE6, it has TLS 1.0
> > > > > —Telpardec TALK 20:57, 16 June 2015 (UTC)

> Yes, but its disabled by default. The type of people who use internet explorer are probably not messing with the TLS settings. When I was running IE6 under wine, enabling TLS1.0 didn't seem to help anything, but that was probably just wine not working great. Bawolff (talk) 04:19, 17 June 2015 (UTC)

To Resident Mario: The switch to HTTPS will badly hurt those who chose to change their browser's default list of certification authorities and who, specifically, do not trust GlobalSign (the root authority from which Wikipedia's certificate emanates). At the very least, they will be forced to add security exceptions for all Wikipedia domains, and quite possibly will be locked out of Wikipedia altogether because browsers do not always allow security exceptions on HSTS sites. In effect, the switch means that users are forced to either trust *everything that GlobalSign signs* if they wish to use Wikipedia, whereas so long as HTTP transport was permitted, one could at least read Wikipedia on HTTP if one does not care about the security of public information on Wikipedia but doesn't want to trust GlobalSign. (I can't explain the problem with GlobalSign because I don't want to risk being sued for libel, but let's say that one might not necessarily wish to trust all, or any, certificate authorities.) So the irony is that this change, which is supposed to protect the "security" of users, actually forces security-conscious users to downgrade theirs, in effect a Trojan horse kind of attack. (In all fairness, Web browsers and HTTPS in general should be blamed for having an absurdly rigid approach to security: one can't restrict a certificate authority to certain domains, or things like that, so I can't say "I trust GlobalSign only for signing certificates in the wikipedia/wikimedia/wiktionary/etc.org domains".) --Gro-Tsen (talk) 21:15, 13 June 2015 (UTC)

> For real? Any person who intentionally messes with their root certificate store, should be technically competent enough to make their own trust decisions of Wikimedia certs, by say verifying them in some other way. If you're not, you have no business removing CAs from your trust store. Bawolff (talk) 21:45, 13 June 2015 (UTC)
> About 10% of HTTPS websites use GlobalSign (http://w3techs.com/technologies/details/sc-globalsign/all/all), so it is not a Wikipedia-specific issue. One could say the same for any other CA that the WMF may decide to use. Moreover, Bawolff makes a great point that someone technically competent enough to mess with trusted roots would be able to work around this as well. They must know how to do so already, since there are numerous other sites using GlobalSign! If someone really lost faith in the CA system, they should try using Convergence, Perspectives (https://addons.mozilla.org/en-US/firefox/addon/perspectives/), or Certificate Patrol (https://addons.mozilla.org/en-US/firefox/addon/certificate-patrol/). Tony Tan · talk 03:07, 14 June 2015 (UTC)
> @Resident Mario: To answer your second question, according to zh:Template:Wiki-accessibility-CHN, zh.wikipedia.org is currently completely blocked in China using DNS poisoning. HTTPS versions of all other Wikimedia projects are not blocked. @Gro-Tsen: If you manually remove GlobalSign root certificates from your browsers' trust stores, you can manually add Wikipedia's leaf certificate to the trust store so that your access to https://en.wikipedia.org/ is not blocked by your browsers. 191.237.1.8 (talk) 05:09, 14 June 2015 (UTC)

To Resident Mario: In short: **HTTPS everywhere hurts everyone**. HTTP was built with network proxy and caching servers to decrease page load times. These are intermediate servers run by your ISP to reduce backbone data requests. Australians will bemost affected since 87 ms away from our Virginia data centers (https://www.wolframalpha.com/input/?i=Virginia+to+Perth), so they'll have a 200 ms ping. Due the design of HTML, these requests can stack meaning that 200ms could bloat to 2 seconds. Now <100 ms is considered ideal, 1 sec users become frustrated, and at 10 seconds they'll look for something else. (Proponents will weasel around this by saying *your* browser caches content, which helps if you don't back to the Google search results)

Additionally, anyone who say this'll stop the $53 billion a year NSA is delusional. Methods for the NSA to get the WMF private keys: Court order (ala Lava Soft who shutdown over this), intercepting and backdooring hardware (Cisco routers, Hard drives), to recruiting/bribery of employees. This basically leaves ISP spying on users (Verizon Wireless adds a advertizing tracking ID to all HTTP requests), but considering how willing WMF is to toss aside net neutrality... — Dispenser 15:08, 17 June 2015 (UTC)

> On that first part: well yes and no. Most browsers now support SPDY and/or HTTP2/0, for which https is a requirement and which will give you a 20-700% speed boost. Especially this last part is probably going to significantly increase the speed for the majority of the users in those areas. Second, that area is served from the San Francisco caching center, so it's slightly closer then Virginia at least, though still so far away, that there is a good point. I do know that WMF is watching the performance impact of this change around the world, and I think they were already considering adding another caching center for Asia/Oceania regardless, so if it really does drop measurably, then that consideration might get higher priority. —TheDJ (talk • contribs) 01:09, 18 June 2015 (UTC)
>
> > We send anti-caching headers (Because people edit and then things become outdated). ISP level caching servers that conform to the http spec should not be caching wikipedia pages whatsoever. So HTTPS won't really affect caching efficiency. Well lots of people go on and on about NSA, I really think the threat that this move is more designed to address is someone like China or Russia, altering pages in a Mitm fashion to make articles less NPOV. Bawolff (talk) 02:23, 18 June 2015 (UTC)
> >
> > > This isn't an anti-NSA measure, it's due to security and privacy concerns on a number of different levels, not all of them related to governments. /Johan (WMF) (talk) 13:37, 22 June 2015 (UTC)

And another problem: No browser history!

@Johan (WMF): - In addition to losing the *drop-down edit summaries* (as mentioned above), I've also lost the *browser history* for all newly-visited Wikipedia pages. Why the exclamation point?? Because this is absolutely *crucial* -- in fact, *integral* -- to my ability to work on Wikipedia. I totally depend on having those page links, which give me quick & easy access to all recently-visited pages.

Johan, you said above, "We want editing Wikipedia to be as simple as possible, no matter which browser people use." (I am using IE 8.) Please tell me there is going to be a technical fix for this problem ASAP. Because if there isn't, there is a *very real possibility that I will have to give up editing*. I am a long-time (since 2006), very conscientious editor, with nearly 60,000 edits. So I ***truly hope*** that does not become necessary. Cgingold (talk) 09:11, 13 June 2015 (UTC)

> P.S. - I raised the very same issues a couple of years ago during the last discussion on this subject, which was resolved to my satisfaction when I learned that it was possible to **opt out**. So this is really a sore point for me. It sure would have been nice if you guys at least had the consideration to place a banner at the top of all pages for a week or two giving all of us a heads up about the impending change. Matter of fact, I believe I made the same point *last time*! :-(Cgingold (talk) 09:20, 13 June 2015 (UTC)
>
> > Best advice I can give is to use IE11 or another non broken browser. —TheDJ (talk • contribs) 10:19, 13 June 2015 (UTC)
> >
> > > Yup, this is a problem for me too that is admittedly a considerable annoyance. I always opted out previously for this reason. Connormah (talk) 11:21, 13 June 2015 (UTC)
> > >
> > > > @Cgingold: If you mean that you lost your browser history for all of the `http` domains, I would say: deal with it yourself. It's a petty issue. You will regenerate the URLs soon enough as you visit the new pages again; it's no different than if you were to clear your browser history. If you have lost the *ability* to generate new URLs in your URL history, then that is a problem. I hope it can be fixed, but if it cannot...wouldn't it be easier for you to move up to an Internet browser that's less than six years old? **Res**Mar 13:48, 13 June 2015 (UTC)

> Even if it was "merely" the loss of older browser history that I was referring to -- which it wasn't -- that would hardly be "petty", my friend. You might want to check your attitude at the door before you trivialize another editor's problem. But of course, I *was* talking about the fact that my browser no longer generates new URL links in the browser history. And it is indeed a very serious problem. Cgingold (talk) 21:29, 13 June 2015 (UTC)

> Petty? The switch to HTTPS is petty. It is stark raving mad to switch to https to avoid NSA surveillance. I cannot believe the reasoning there, some people need to take their tin foil hats off. I bet if anyone were to read this at the NSA then would have a right good laugh at us all. Even if they were inclined to mine data off this site then the switch to https would be of little impediment to a body of that resources. Why do we not only operate on tor and demand VPN usage if we are trying to protect the hypothetical drug smugglers, money launderers and terrorists that apparently have abandoned the onion sites in favour of WP talk pages? There is no benefit for this change in policy and the reasoning behind is deranged.--EchetusXe 17:46, 13 June 2015 (UTC)

> > I am not here to hear your opinion, I am here to assess the damage. ReeMar 19:39, 13 June 2015 (UTC)

> @Connormah: As a sysop, you should probably use HTTPS. Otherwise, your account is at risk of being hijacked in a Firesheep-style attack, especially when you use a public network. A sysop account would be really useful for someone intending harm. :(If there are big issues, upgrading your browser to a newer version of IE, Chrome, Firefox, etc. should help. Tony Tan · talk 03:15, 14 June 2015 (UTC)

Cgingold, I just wanted to say that, yes, we really do care about your problems, we appreciate all the work you're doing, and I will ping you personally as soon as I have good answer or solution. /Johan (WMF) (talk) 12:15, 16 June 2015 (UTC)

For reference, IE < 11 represents about 5.5% of our traffic [11] (http://stats.wikimedia.org/wikimedia/squids/SquidReportClients.htm). Bawolff (talk) 18:54, 13 June 2015 (UTC)

How about a 'in the clear' sub-wiki?

Like http://ite.en.wikipedia.org which just reflects the normal wiki. Then all users of 'normal' wikipedia get HTTPS, but people who want/need HTTP have to specifically ask for it.ן ɜ ʁ H ↄ I ll (talk) 09:38, 13 June 2015 (UTC)

> It would more likely be http://en.insecurewikipedia.org, but I don't think there would be many fans to maintain such a system.. We will have to see about what kind of case can be made for that, but I think it is unlikely that it will happen. —TheDJ (talk • contribs) 10:25, 13 June 2015 (UTC)

> > Anyone could setup a proxy to do this (e.g. http://crossorigin.me/https://en.wikipedia.org [maybe that's a bad example, as it doesn't fix the links]. Anyways, point is that it is trivial to set up an independent proxy to an HTTPS site. Allowing edits might be trickier, but not impossible). Bawolff (talk) 18:28, 13 June 2015 (UTC)

We have had a discussion

Just a note that we have had a discussion at the village pump about this earlier this year (WP:VPR/HTTPS). The discussion was closed as WP:CONEXCEPT due to the highly technical nature of the issue.

From my point of view, this move to HTTPS-by-default is the correct one. Mozilla (https://blog.mozilla.org/security/2015/04/30/deprecating-non-secure-http/) (Firefox), Chromium (https://www.chromium.org/Home/chromium-security/marking-http-as-non-secure) (Chrome), the IETF (https://datatracker.ietf.org/doc/rfc7258/), and W3C TAG (http://www.w3.org/2001/tag/doc/web-https) are all behind moving websites on the Internet in general to HTTPS and deprecating insecure HTTP.

HTTPS guarantees the authenticity of content sent from Wikipedia servers as it travels through the Internet, prevents tampering (whether it is censorship in another country or your internet service provider injecting ads (http://arstechnica.com/tech-policy/2014/09/why-comcasts-javascript-ad-injections-threaten-security-net-neutrality/) or adding invasive tracking headers (http://www.forbes.com/sites/kashmirhill/2014/10/28/find-out-whether-this-privacy-killing-super-cookie-is-on-your-phone/)), and curbs mass surveillance (by a gov't or an internet provider) by making it difficult and expensive to monitor articles being read or written by individuals.

Regarding the potential negative effects of switching to HTTPS for older clients/browsers, we should be able to find a workable solution for them fairly quickly. A lot of the issues mentioned are software bugs that can be fixed without going back to HTTP. Google uses HTTPS by default, and there does not seem to be an issue with anyone using Google. Tony Tan · talk 20:43, 13 June 2015 (UTC)

> Thank you so much, Tony, for pointing out that *Google* doesn't cause these kinds of problems! Somehow, I hadn't even noticed that -- I guess precisely because it *doesn't* cause any problems... SHEESH!! If these issues are, in fact, entirely unnecessary, then WHY WERE THEY IGNORED by WMF's tech people when they had been explicitly pointed out on this very page a couple of years ago??? Inexcusable. I am sitting here literally shaking my head in disbelief... Cgingold (talk) 21:48, 13 June 2015 (UTC)

> > Well, google (the search engine anyways, not counting other sites google runs) does its own auto-complete with javascript based on what it thinks you want to search for. It does not use the built in remember what I typed previously browser feature. ~~You used the word "issues" in the plural. As far as I'm reading, the old version of IE disables auto-complete on HTTPS is the only actual issue reported in this thread that could possibly not affect Google (Or for that matter, is a reasonable complaint imo). Am I mistaken?~~ Edit: I guess you're also complaining about browser history, so that makes 2 issues. All things considered, both are essentially minor inconveniences, both are experienced only by a relatively small number of users, and the autocomplete one has an easy way of mitigating (update your browser). Not exactly what I'd call the end of the world. Bawolff (talk) 04:45, 14 June 2015 (UTC)

Please enable HTTP mode

Hi. I'm from Iran. After WP enabled https as default (and no access to http), we have a lot of problem to access WP due to Internet censorship. Because Iranian government abuses https protocol. It's very slow and pages do not load properly. Time-out error happens frequently. Editing is not easy anymore. Please enable HTTP option for restricted countries again. Wikipedia is a great contribution to humanity. Thanks. –188.158.107.24 (talk) 10:41, 14 June 2015 (UTC)

> All people everywhere possess the inalienable right to have access to information of any and every kind. And they should be able to express that right without intervention by any company, organization or government, to include suppression, censorship and secret monitoring. The sole exception would be information that is kept secret for reasons of national security. What I don't understand is why any government would suppress and censor this right by committing abuse of HTTPS and not also commit abuse of HTTP? Is HTTP really that much harder to abuse? to suppress and to censor? Since many of the problems that have erupted since Wikipedia converted to HTTPS-only are shown to be due to users using older versions of

software, and perhaps older hardware as well, maybe if you upgraded to recent versions you would find that rather than governments being the problem, usage of non-recent versions of hardware and software is the problem? – *Paine* 16:06, 14 June 2015 (UTC)

> They try to block HTTPS and other encrypted traffic because they can't see what you're doing. Cleartext traffic like HTTP can be examined. They want to give people some access to the Internet, because they know it's generally a lost cause to try to block Internet access completely, and trying to do so might spark a revolt, but they want to retain the ability to block some content, and keep tabs on what you're doing. For instance, China's "Great Firewall" selectively blocks access to information on things like the Tienanmen massacre through multiple techniques, including a blacklist of certain sites, and traffic analysis. --108.38.204.15 (talk) 22:33, 14 June 2015 (UTC)
>
> > @Legoktm: you might know who to pass this concern onto. Magog the Ogre (t • c) 22:35, 14 June 2015 (UTC)
> >
> > > I think I understand what it feels like to be faced with Internet censorship; I spend half my time in China, where the Great Firewall disrupts access to websites that are commonly used in countries like the U.S. It is very, very frustrating. What I do want to point out, however, is that by enabling forced HTTPS encryption, governments like that of Iran will be forced to make the decision to either **block all of Wikipedia or none of it**, instead of being able to selectively filter by the topic of individual articles. While in the short term users may find access to be unstable or even impossible, the government may eventually be forced to stop interfering with Wikipedia traffic if it decides that access to the "good" information is more important than filtering the "bad" information. So in the long run, it may be better to keep Wikipedia HTTPS only if users eventually end up having access to **all of Wikipedia**, without censorship. There is no guarantee, but I think we should at least wait and see. Tony Tan · talk 01:50, 15 June 2015 (UTC)
> > >
> > > > @108.38.204.15: Out of curiosity, do you have a source for information about the great firewall using traffic analysis? Most of the things I read seem to suggest they mostly use deep packet inspection and DNS posioning. And I'd be really interested in reading any publicly available info about how their system works. Bawolff (talk) 02:10, 15 June 2015 (UTC)
> > > >
> > > > > I'm suspicious that HTTPS will do nothing to stop spying by the NSA or GCHQ, but has been introduced to make it much harder for whistleblowers to sit in the middle and see who they are spying on. It seems we're stuck with it though, and if you're using ancient browsers such as IE8, you'll just have to upgrade. Akld guy (talk) 06:24, 15 June 2015 (UTC)
> > > > >
> > > > > > That doesn't really make sense to me. What realistic opportunities would a whistleblower ever have to be in the middle of an NSA/GCHQ communication? And even if they were in such a position, the transport security of Wikimedia would be rather irrelevant. To the best of my knowledge, no whistleblower has ever intercepted communications in transit over the internet in order to release for the public interest. Whistleblowers are usually in a trusted position, and legitimately have access to the data which they decide to divulge. Bawolff (talk) 07:47, 15 June 2015 (UTC)
> > > > > >
> > > > > > > I want to clarify one thing that's turned up a couple of times in the general discussion (and I'm not replying to any specific user here). There have been a number of comments regarding the NSA. We know that the NSA has targeted Wikipedia traffic, and the Wikimedia Foundation doesn't believe Wikipedia readers and editors ought to be targeted, but while this may have been tangentially related to concerns over the NSA, it wasn't the driving force. There are other governments and private actors to take into account, and, for example, the Firesheep style attacks that Bawolff has mentioned. Rather, it was driven by concern for the privacy and security of editors and readers all over the world, which means there are many different problems to consider. /Johan (WMF) (talk) 08:00, 15 June 2015 (UTC)
> > > >
> > > > I am with Tony Tan on this one. Our concern is not the NSA or GCHQ spying on users (that can be done even inspite of HTTPS), but its governments like Iran, China, and others that (with HTTP) could filter out certain content from Wikipedia without the majority of people noticing. HTTPS forces them to either block *.wikipedia.org entirely, or just let go. They will probably chose the latter, since the former will cause protest sooner or later. Compare, by the way, to what Russia has been doing with the Internet Archive (http://arstechnica.co.uk/tech-policy/2015/06/wayback-machines-485-billion-web-pages-blocked-by-russian-government-order/) because of their recent HTTPS-by-default policy: they had to block the entire domain. Of course they would love to only filter out LGBT-related (http://www.theguardian.com/world/2013/sep/01/russia-rise-homophobic-violence) topics etc. but it is a good thing they cannot. And this is why we have to make HTTPS the *only* option. --bender235 (talk) 09:38, 2 July 2015 (UTC)

- Just to add my 5c, I do remember using a university Internet network a year ago that completely banned HTTPS (so I could use Wikipedia only in HTTP). I do not know the origin of this block (this should be definitely a setting by university network administrator), and I do not know if that block is still there (I haven't used it since then), but I would like to inform that such networks do exist, and I don't think there is a way to track them — NickK (talk) 09:16, 15 June 2015 (UTC)

> Such networks probably exist, but I think it would be up to the network administrators to whitelist Wikipedia's servers if they believe access to Wikipedia is important. They would probably do it after realizing that it is no longer possible to access Wikipedia on plain HTTP. Tony Tan · talk 05:26, 16 June 2015 (UTC)
>
> If Iran blocks HTTPS, there's no way Wikipedia/WMF will be changing their minds by blocking access to Wikipedia for Iranians through HTTP, which is probably a desirable outcome for the regime anyways. WMF should set up additional HTTP servers for static access to Wikipedia (no-edit access) then with a disclaimer stating that the content may be modified by third party man-in-the-middle vandalism in big banner statements at the top and bottom of every page. -- 70.51.203.69 (talk) 04:44, 17 June 2015 (UTC)
>
> > It would be trivial for the men-in-the-middle to remove the disclaimers. (talk to) TheOtherGaelan('s contributions) 06:16, 17 June 2015 (UTC)
> >
> > > Yes, it would, however, it would *reenable access to populations who are completely blocked form using HTTPS*. If the governments in question actively block HTTPS, then we are just falling into their hands by removing access to Wikipedia from their populations, to limit their populations access to information by voluntarily falling into the schemes of their governments to censor the internet by removing access to Wikipedia completely, as they filter out HTTPS. -- 70.51.203.69 (talk) 11:31, 18 June 2015 (UTC)
> > >
> > > > Never really saw the logic behind moving to https...so it either stops the governments from snooping the accounts of say 10,000 wikipedians (people who browse and randomly edit the wiki) or by moving to https, it blocks 1.2bn-2bn users from **COMPLETELY** accessing the website..If i was the guy incharge of making the decision, I'll choose the latter. I'd rather have a billion users being able to access this site than help 10,000 users from "hiding" behind closed doors and randomly attacking their government and making this site look bad....sadly, I don't work for the site and I sympathize with those that can no longer access the site..if WMF had actually done their research before doing this, they would realise it was those users who contributed a lot to the website than those 10,000 who use the site for their own personal agendas...alas...the weak shall inherit the wiki..and for the 1000th time, enwikipedians demands supersedes the demands of other language wikis--*Stemoc* 11:53, 18 June 2015 (UTC)

> Billion? Do you have a citation for that? Before anyone says China, China is not currently treating https access to Wikipedia any differently then http access. I'm keenly interested in who this actually blocks, so if anyone has actual information about people who are blocked... please say so. Bawolff (talk) 21:39, 18 June 2015 (UTC)

If the governments that currently block HTTPS really intended to completely remove their citizens' access to all of Wikipedia, they would have already done so over HTTP. Precisely because they still see value in some of Wikipedia's content, they chose to filter instead of block. HTTPS removes the filter option, so they will have to either allow or block all traffic to Wikipedia. When they made the decision, Wikipedia was still available over HTTP, so they chose to block HTTPS and filter HTTP, achieving their purpose of allowing access to some information while blocking others.
Now that Wikipedia can only be accessed on HTTPS, they are forced to re-evaluate their decision. They are now forced to decide between blocking all of Wikipedia, or allowing all of it. While all of Wikipedia is blocked as of now (due to their earlier decision based on a situation that has since changed), they may eventually be forced to allow it if they think public access to certain resources is important. This was the case for GitHub. When GitHub switched to HTTPS-only, China eventually decided to allow all GitHub traffic because of its importance to software development, even though there were other information on there that the gov't wanted to censor. It may be a while before HTTPS becomes unblocked; perhaps the governments are waiting for Wikipedia to enable HTTP access again, which would make it unnecessary for them to allow HTTPS and give up filtering. Tony Tan · talk 07:34, 21 June 2015 (UTC)

> Or they could tell people to use Baidu Baike, or similar local service. -- 70.51.203.69 (talk) 12:33, 23 June 2015 (UTC)
> On that note, does that mean that Wikipedia has a TOR address? (Does Iran successfully block TOR?) -- 70.51.203.69 (talk) 12:36, 23 June 2015 (UTC)

> > You do not need a website to have a "TOR address" to use Tor to access the website. You can use Tor to access any website that does not block Tor exit node IPs. .onion addresses are used for concealing the location of the web server. Tony Tan · talk 20:43, 23 June 2015 (UTC)

Note

Google has been mentioned. While Google defaults to https it can be (easily) persuaded to use http. All the best: *Rich Farmbrough*, 16:54, 7 July 2015 (UTC).

Link to talk page in Mobile Wikipedia

The talk page link in the mobile version of Wikipedia should be shown for non-logged in users as well. GeoffreyT2000 (talk) 03:36, 3 July 2015 (UTC)

> @GeoffreyT2000: Thanks for the feedback. I'd suggest sending to the mobile mailing list, mobile-l@lists.wikimedia.org, to keep discussion centralised and give everyone a chance to participate. Thanks again! --Dan Garry, Wikimedia Foundation (talk) 20:09, 7 July 2015 (UTC)

Page ID

So... I have page ID (i simply have it). Is there some simple way to find out, to which page this ID belongs? Not using API query, SQL quarry... Can't I go to search and do some search like "id:XXXXX", or maybe LUA people can do some work? As I understand, Module:Page can't do that. --Edgars2007 (talk/contribs) 13:40, 3 July 2015 (UTC)

> My user page has page ID 26096242; this URL `//en.wikipedia.org/w/index.php?curid=26096242` is another way to load it. -- John of Reading (talk) 13:55, 3 July 2015 (UTC)

> > OK, but what about not touching URL. The perfect solution would be `{{some template|26096242}}`, which would give User:John of Reading. --Edgars2007 (talk/contribs) 14:13, 3 July 2015 (UTC)

> > > I don't know know a way to display the page name. https://en.wikipedia.org/?curid=26096242 is a shorter url. It can be used in {{querylink}} where `{{querylink|qs=curid=26096242|Unknown page}}` produces Unknown page (https://en.wikipedia.org?curid=26096242). PrimeHunter (talk) 15:13, 3 July 2015 (UTC)

> > > > This is easy from Lua: `mw.title.new(26096242).prefixedText`. It would be easy to set up a module to do this if necessary. Jackmcbarn (talk) 15:21, 3 July 2015 (UTC)

> > > Fwiw... there's already a Special: page search utility that can take a PAGID and give you the associated page (albeit a bit clunky as well as poorly labelled) -- just go to the Redirecting Special Pages section and select *Redirect by file, user, page or revision ID*. Don't forget to switch the input selector menu value from *User ID* to *Page ID* before you send your request.
> > > You can also build a template based on that special page's syntax and the previous id given above like: Special:Redirect/page/26096242 .
> > >
> > > Unfortunately, as it stands today, the output is not listed as an optional wikilink for you to follow if need be but automatically takes you to the target article, revision or user in question instead. I'm sure amending the app to display the target as a clickable wikilink rather than automatically opening the target page is the better solution here. Maybe providing a checkbox indicating not to take you to the target as an alternative maybe? -- George Orwell III (talk) 00:15, 4 July 2015 (UTC)

> > > > Thanks, guys--Edgars2007 (talk/contribs) 09:54, 7 July 2015 (UTC)!

Force desktop version?

I read WIkipedia on an iPad, and the screen is large enough that don't need the awful mobile version. Yet I can't stop my Chrome browser from constantly serving up the mobile version, constantly forcing me to tap the "Request desktop version" button. Is there a way to force Wikipedia to give me the desktop version by default? --Calton | Talk 21:21, 5 July 2015 (UTC)

> Bookmarking "en.wikipedia.org" (without the .m.) and always starting from there serves as a workround, since once you've requested desktop site once it *should* remember it as long as you don't give it the chance to go to the mobile site. I completely agree about the shittiness of the mobile site, and whoever thought it should be default should be summarily fired—even on phones, let alone tablets, it's far less friendly than the desktop site. – iridescent 21:29, 5 July 2015 (UTC)

> > @Iridescent: If you have some constructive feedback on what you don't like about the mobile view for reading, then the Reading Department would welcome it. You can give that feedback on the mobile mailing list, mobile-l@lists.wikimedia.org. That said, I would note that if you phrase your feedback to the list in the extremely combative manner that you did here, people will likely avoid engaging with you. Please keep things as constructive as possible, both on- and off-wiki. --Dan Garry, Wikimedia Foundation (talk) 20:07, 7 July 2015 (UTC)

I've found that bookmarking "en.wikipedia.org" is not sufficient (on Android Chrome) as the server will detect your platform and redirect to the mobile site anyway. However, once the mobile page loads, if you scroll to the bottom and click the "Desktop" link, then the server remembers your choice. I'm not sure if it only lasts until you close the tab or if it lasts as long as your login session, but it expires eventually. Also, there is no equivalent way to switch back to mobile, you have to add the ".m" into the URL to get back. Ivanvector ♣ (talk) 20:12, 7 July 2015 (UTC)

There is a link to the mobile version on the bottom of every desktop page. -- [[User:Edokter]] ([talk]) 21:13, 7 July 2015 (UTC)

Arunanshu abrol thanking DumbBOT

Normally, bots cannot be thanked. Why did Arunanshu abrol thank DumbBOT? GeoffreyT2000 (talk) 01:01, 6 July 2015 (UTC)

@GeoffreyT2000: Well, nobody knows except Arunanshu abrol (talk · contribs) themselves. Have you asked them why? But if you mean "how", it's very easy. All you need is the revision ID: for example, the last edit made by DumbBOT (talk · contribs) is Special:Diff/670158689 so try visiting Special:Thanks/670158689. –Redrose64 (talk) 07:44, 6 July 2015 (UTC)

I get "Thank action failed. Please try again. " YMMV. All the best: *Rich Farmbrough*, 17:02, 7 July 2015 (UTC).

OK... I just thought to look for the thanks in question. There's only one logged (https://en.wikipedia.org/w/index.php?title=Special%3ALog&type=thanks&page=User:DumbBOT) as sent to DumbBOT, and it's timed at 07:48, 29 October 2013 (at first, I had assumed that the incident was recent). Might it be that it was possible to thank bots at the time, but the software has since been amended? --Redrose64 (talk) 17:48, 7 July 2015 (UTC)

API calls just starting throwing SSL/HTTPS (?) errors

For context, I run WP:STiki, which scores every en.wp edit in near real-time. In the last couple of hours, this process has hit the fan. In the last several days, I have implemented changes to handle the HTTPS switchover and the new continuation procedure for queries returning long result sets. As of ~15 hours ago, everything was running perfectly smoothly. Now my (Java) API code is throwing errors like this at every API call (but does succeed in browser):

```
Error: HTTP error at URL: https://en.wikipedia.org/w/api.php?
action=query&prop=revisions&revids=670300218&rvtoken=rollback&rvprop=ids|timestamp|user|comment|tags&format=xml
javax.net.ssl.SSLException: java.lang.RuntimeException: Could not generate DH keypair
        at sun.reflect.GeneratedConstructorAccessor15.newInstance(Unknown Source)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:532)
        at sun.net.www.protocol.http.HttpURLConnection$6.run(HttpURLConnection.java:1458)
        at java.security.AccessController.doPrivileged(Native Method)
        at sun.net.www.protocol.http.HttpURLConnection.getChainedException(HttpURLConnection.java:1452)
        at sun.net.www.protocol.http.HttpURLConnection.getInputStream(HttpURLConnection.java:1106)
        at sun.net.www.protocol.https.HttpsURLConnectionImpl.getInputStream(HttpsURLConnectionImpl.java:254)
        [snip some lines]
Caused by: javax.net.ssl.SSLException: java.lang.RuntimeException: Could not generate DH keypair
        at sun.security.ssl.Alerts.getSSLException(Alerts.java:208)
        at sun.security.ssl.SSLSocketImpl.fatal(SSLSocketImpl.java:1697)
        at sun.security.ssl.SSLSocketImpl.fatal(SSLSocketImpl.java:1660)
        at sun.security.ssl.SSLSocketImpl.handleException(SSLSocketImpl.java:1643)
        at sun.security.ssl.SSLSocketImpl.startHandshake(SSLSocketImpl.java:1224)
        at sun.security.ssl.SSLSocketImpl.startHandshake(SSLSocketImpl.java:1201)
        at sun.net.www.protocol.https.HttpsClient.afterConnect(HttpsClient.java:440)
        [snip some more]
```

Clearly something is going on at the SSL handshake between my server and the WMF one. Given my things were working on my end and I have not intervened, this heavily suggests something was changed on the WMF side. Any pointers? I'll note that CBNG also went down parallel to my service, I believe. Thanks, West.andrew.g (talk) 03:57, 7 July 2015 (UTC)

CBNG isn't feeding on IRC either. Ж (Cnemaster) T/C 06:26, 7 July 2015 (UTC)

This is probably related to phab:T104281. You probably need to update your Java version. —TheDJ (talk • contribs) 11:34, 7 July 2015 (UTC)

moving to java 7 or higher will solve your issue. Matanya (talk) 11:47, 7 July 2015 (UTC)

Most frequently used words with 6 or more characters on the English Wikipedia

How can I find someone who knows how to do a statistical analysis on a Wikipedia dump? I would like to have a list of the most frequently used words on Wikipedia that contain 6 or more characters. For more info about why I want such a list please click here. Thanks! The Quixotic Potato (talk) 11:55, 7 July 2015 (UTC)

I have done a bigram analysis before, let me dig though my archive. All the best: *Rich Farmbrough*, 14:53, 7 July 2015 (UTC).
Oh... and the reason this is *theoretically* not a sufficient tool for working on typos, is that the statistical nature of each dump is post-correction for certain typos. For example I fixed all (5 or 6) occurrences of "chruches" a few days ago. A current dump would indicate that this is never misspelled thus. All the best: *Rich Farmbrough*, 14:57, 7 July 2015 (UTC).
Here (https://meta.wikimedia.org/wiki/User:Rich_Farmbrough/en:wp:words-10k) is some data from 2010. It might be a good test set. All the best: *Rich Farmbrough*, 15:28, 7 July 2015 (UTC).

I have coded this up, latest dump is downloading.... Moving convo to User:The_Quixotic_Potato's talk page. All the best: *Rich Farmbrough*, 16:58, 7 July 2015 (UTC).

Content Translation, the new article creation tool is now available as a beta-feature

Hello, Content Translation has now been enabled as an opt-in beta feature on the English Wikipedia for logged-in users. To start translating please enable the Beta feature in your preferences. Visit Special:ContentTranslation or go to your contributions page to open the tool. You can follow the instructions in the User Guide on how to get started. You can also find more information in our earlier announcement in The Signpost.

Since this is the first time we have installed the tool on this wiki there are chances that there may be some problems or service disruptions which we are not yet aware of. We will be monitoring the usage to check for any failures or issues, but please do let us know on the Content Translation talk page or through Phabricator if you spot any problems. Thank you. On behalf of the Wikimedia Foundation's Language Engineering Team: --Runa Bhattacharjee (WMF) (talk) 17:06, 7 July 2015 (UTC)

@Runab WMF: Why do we need this? This is the English Wikipedia; pages are written in English. If we want to translate a page to, say, German, we edit the German Wikipedia. --Redrose64 (talk) 17:36, 7 July 2015 (UTC)

@Redrose64: This isn't my initiative, but you don't seem to understand what Content Translation does. Can you read the *Signpost* article linked above? Best, Ed Erhart (WMF) (talk) 17:39, 7 July 2015 (UTC)
@Redrose64: Research (http://brenthecht.com/papers/bhecht_chi2010_towerofbabel.pdf) shows that even between big Wikipedias such as English and German the overlap is about 51%. That means that half of the German Wikipedia could be translated into English. That does not mean that all those articles are relevant to English Wikipedia, but there are some valid opportunities for translation into English. In this ticket (https://phabricator.wikimedia.org/T94123) we collected the requests from the community to enable the tool in English Wikipedia.Pginer-WMF (talk) 18:20, 7 July 2015 (UTC)

How to use Content Translation - a short video)

Images not showing up

If you add an image to a page and then upload the image, the page will fail to display the image (and put the page into Category:Articles with missing files) until the job queue catches up, or until you edit the page. In the last 24 hours, I've had a variant of this problem. I added an image to Rawson, Ohio before it had finished uploading to Commons, so I made a null edit, because that fixes ordinary purging problems. The result? No change. I had to make (https://en.wikipedia.org/w/index.php?diff=670362343) a dummy edit (while using the wrong terminology) to get things to work, and then self-revert. Meanwhile, I'd added an image to Benton Ridge, Ohio (not hitting Save until the image was visible on Commons), and the same thing occurred: it looked like a nonexistent image, despite a null edit, until I'd made a dummy edit. Ditto at Cass Township, Hancock County, Ohio. Lately I've noticed that articles occasionally don't display changes after an edit (e.g. I'll add a paragraph, save, and the paragraph's not there), but this problem is normally resolved by refreshing and always by a null edit, and anyway this problem shows no changes whatsoever, not all-changes-except-image. Can anyone explain what's going on, why I have to edit a page twice to get the image to appear? I've been adding photos to other articles as well, and none of the others had problems, even though with most images (e.g. the one at Portage Township, Hancock County, Ohio) I saved the edit as soon as the photo was done uploading. Nyttend (talk) 17:59, 7 July 2015 (UTC)

Did you bypass your own browser cache after the null edit? A month ago that was often necessary after edits as discussed at Wikipedia:Village pump (technical)/Archive 137#Post not showing up immediately. It hasn't happened to me lately but I don't know whether it makes a difference that it is null edits. PrimeHunter (talk) 20:13, 7 July 2015 (UTC)

This is different from "Post not showing up immediately". It goes from showing nothing in the infobox image space to showing a nonexistent image, and after the first edit to each page, all of them were in Category:Articles with missing files — besides the category appearing at the bottom of the page, the article names appeared when I went to the category and looked through its contents. Nyttend (talk) 21:15, 7 July 2015 (UTC)

file_get_contents on wmflabs?

Hi, trying to retrieve this page from labs returns false:

```
$url ="http://tools.wmflabs.org/catscan2/catscan2.php?language=de&categories=%catenc%0D%0A&other_cat_enc&doit=1&format=csv&all_namespaces&depth=15";
$csv_list = file_get_contents($url);
```

Any clue, why this happens? Works well on FF with this url (https://tools.wmflabs.org/catscan2/catscan2.php?language=de&categories=Physik%0D%0A&doit=1&format=csv&ns%5B-2%5D=1&ns%5B0%5D=1&ns%5B2%5D=1&ns%5B4%5D=1&ns%5B6%5D=1&ns%5B8%5D=1&ns%5B10%5D=1&ns%5B12%5D=1&ns%5B14%5D=1&ns%5B100%5D=1&ns%5B828%5D=1&ns%5B-1%5D=1&ns%5B1%5D=1&ns%5B3%5D=1&ns%5B5%5D=1&ns%5B7%5D=1&ns%5B9%5D=1&ns%5B11%5D=1&ns%5B13%5D=1&ns%5B15%5D=1&ns%5B101%5D=1&ns%5B829%5D=1&depth=15). Thanks, --Flominator (talk) 19:32, 7 July 2015 (UTC)

Well the url has variable names still in it.. that would be one reason :) —TheDJ (talk • contribs) 20:58, 7 July 2015 (UTC)

Also doesn't work with http://tools.wmflabs.org/catscan2/catscan2.php ;) --Flominator (talk) 06:52, 8 July 2015 (UTC)

Need some testers

Please see User:Howcheng/sandbox and User:Howcheng/sandbox2. We are trying to add captions to the main page images to solve the longstanding complaint of when the images don't go with the top item in ITN and OTD. I've checked this in Vector and Monobook skins on Win 7 and 8 with latest versions of Chrome, Firefox, and IE. I need some people to verify it using a Mac, using iPad (not worried about iPhone/iPod as smaller resolutions will get the mobile version, which does not include ITN and OTD), and from Android tablets with stock browser and Chrome. Additionally, if anyone has suggestions for other image types to test with, feel free to edit as needed. Thanks! —**howcheng** {chat} 20:20, 7 July 2015 (UTC)

Works just fine on Safari for Mac, and on iPads and iPhones (w00t new responsive design mode of Safari 9 for testing exactly this !). Android will be a lot more work to test. There's a lot of rendering differences between all the minor versions of Android. Android 4.4 == Chrome 30.0.0, that I know. —TheDJ (talk • contribs) 20:57, 7 July 2015 (UTC)
That's a whole lot of nesting divs with contradictory classes (floatright vs. floatnone for example) which basically do nothing. There is a lot of fat to trim. But I like the basic approach, though I would like to advocate using a separate class for main page images instead of using the inline-styled thumb classes (which look weird with the 'new image thumb' gadget enabled). -- [[User:Edokter]] {{talk}} 21:27, 7 July 2015 (UTC)
I can second what TheDJ has written. Tvx1 21:43, 7 July 2015 (UTC)

@Edokter: I started from {{plain image with caption}}. I'm unfortunately not that familiar with the classes in Wikimedia's CSS files. —**howcheng** {chat} 03:03, 8 July 2015 (UTC)

I'll do a little cleanup later. -- [[User:Edokter]] {{talk}} 05:50, 8 July 2015 (UTC)

Changing wiki strings

Is it possible to change the strings of certain things on the wiki? Like, if I wanted to change, for example, the text of "Talk" to say "Discussion" for me, for some reason. I read somewhere you could change it via the skin .css somehow, but I don't know if that's possible or how you'd do it. If it is, how do I find the, like, string id of a thing I wanna change? Thanks. -- Srdan 09:38, 8 July 2015 (UTC)

@Srdjan m: You can only change text values in CSS by using ugly hacks like these (http://stackoverflow.com/questions/7896402/how-can-i-replace-text-through-css). However, it's quite easy to change text values with JavaScript. Just add the code `$( '#ca-talk a' ).text( 'Discussion' );` to your common.js page. The "#ca-talk a" part is the CSS selector for the "Talk" link at the top of every page. The drawback of using JavaScript for this

is that it doesn't load straight away, so there will be a split-second when you load the page when it still says "Talk" instead of "Discussion". — *Mr. Stradivarius* ♪ talk ♪ 10:12, 8 July 2015 (UTC)

Thanks. Btw, would using that hack on css (even though I really didn't get what to do on that page though) actually get rid of the split second delay or would it still be a thing? It's not a huge deal, but it's just a minor drawback that I wanna see if I can eliminate. :P -- Srđan 10:26, 8 July 2015 (UTC)

Revision scoring IEG goes for a second round

Hey folks,

About 6 months ago, we posted here to notify you of an IEG-funded project we've been working on: Revision scoring as a service. Today, I'm posting to ask for your feedback on our plan for a second round of IEG funding. In the first 6 months of our project, we've met our goals. We stood up a production level service for retrieving revision scores. (Test it out right now at this link: http://ores.wmflabs.org/scores/enwiki/?models=reverted|wp10&revids=638307884|642215410) We have 5 languages running (English, French, Portuguese, Turkish and Persian) and two models ('reverted' == probability that the edit will need to be reverted & 'wp10' == WP 1.0 Assessment). We've had a set of tools and bots pick up the service. See ScoredRevisions (https://github.com/he7ig/mw-gadget-ScoredRevisions) and Reports bot 3.

In the next 6 months we plan to do some more interesting stuff.

1. Add an edit type classifier
2. Expand language support to new languages like Spanish and German and projects like Wikidata
3. Extend our WP:Labels service to allow editors with autoconfirmed accounts to create their own labeling campaigns.

If you have a moment, we'd appreciate your feedback or endorsement on our project renewal plan. Thanks. --Ebe123 (talk • contribs) 14:38, 4 July 2015 (UTC)

I've been using the 1.0 assessment model for WikiProject Medicine, and I'm pretty satisfied. In particular, it's nice to have something make a list of things tagged as stubs that probably aren't (and vice versa). Whatamidoing (talk) 16:04, 8 July 2015 (UTC)

Moving article - history lossed

I had moved page Gülnar to Gülnar (province)
then I made {{disambig}} from Gülnar
Bkonrad moved page Gülnar (province) to Gülnar
when I wanted to see the editing history of Gülnar - I found that the history has lost!

which is completely WRONG! and should not be happened! as if someone made a full article "blah-blah", and someone else made a stub "blah_blah", and then if somebody rename "blah_blah" to "blah-blah" all history of "blah-blah" will be lost! (Idot (talk) 15:35, 7 July 2015 (UTC))

When you reverted Bkonrad's move the history was moved back to the "(province)" page - with Bkonrad's move revert. Jo-Jo Eumerus (talk, contributions) 15:57, 7 July 2015 (UTC)

now I have moved page Gülnar to Gülnar (district), and there is no previous history of Gülnar (Idot (talk) 15:59, 7 July 2015 (UTC))

There are three deleted edits at Gülnar. Only one seems relevant, and that's the one where you added additional links to the page. Is that the history you are looking for? I've dropped the text of that page onto your talk page. There is no other history that I can find, but these deleted edits seem consistent with the moves that show up in the history. UltraExactZZ Said ~ Did 16:21, 7 July 2015 (UTC)

I've asked for the district's page to be moved back to Gülnar, as it is the WP:PRIMARY topic for Gülnar, and there is no need for a disambiguation page per WP:TWODABS. A hatnote for Gulnar Hayitbayeva can be placed at the top of the district's page. -*Niceguyedc* Go Bucks! 16:24, 7 July 2015 (UTC)

answered Talk:Gülnar (district)#diambig (Idot (talk) 16:41, 7 July 2015 (UTC))

anyway how about technical issues? (Idot (talk) 16:41, 7 July 2015 (UTC))

The top of page histories has a link saying "View logs for this page". The link for Gülnar is [12] (https://en.wikipedia.org/w/index.php?title=Special:Log&page=G%C3%BClnar) which includes: "23:30, 6 July 2015 Bkonrad (talk | contribs) deleted page Gülnar *(G6: Deleted to make way for move)*". Deletions can only be made by administrators like Bkonrad. When a page has been deleted, the page history is only visible to administrators. The software did as requested so there is no technical issue. It's common to delete a page with no necessary content to make way for a move. Bkonrad made a judgment call that the content in the page history was not necessary. It's possible but tricky to keep the old content somewhere when a page is moved to an already used name. As mentioned above, Ultraexactzz has copied the deleted content to your talk page. PrimeHunter (talk) 16:54, 7 July 2015 (UTC)

if the history had not been deleted by an administrator, would it be been visible? (Idot (talk) 16:58, 7 July 2015 (UTC))

Yes, but an administrator has to jump through some hoops to avoid deleting the page history when a new page is moved to a used title without the old page being moved elsewhere. And if the two page histories are merged into a single page history then the result can be quite confusing. WP:HISTMERGE may give an idea of the complications involved in history merging. We usually try to avoid it. PrimeHunter (talk) 17:30, 7 July 2015 (UTC)

OK! Thanks for the answers! (Idot (talk) 15:49, 8 July 2015 (UTC))

Nesting footnotes

Within the past four weeks, there appears to have been a software change that has solved many of the problems with nested footnotes - you can now have the main note and subnote in the same list. However, a new problem has appeared in that subnotes using {{refn}} and {{efn}} now fail within an outer note that uses <ref></ref> tags. To see examples, please look at WP:Nesting footnotes#What does not work; cases 2 and 3 displayed correctly a month ago. Can anyone investigate and report as necessary?: Noyster (talk), 19:53, 8 July 2015 (UTC)

As far as I remember, cases 2 & 3 never worked - the outer `<ref>...</ref>` always blocked expansion of certain constructs inside, which included `{{#tag:ref|...|}}` which is what is inside `{{efn}}`. --Redrose64 (talk) 20:21, 8 July 2015 (UTC)

I lack the energy to add this to the WP:, but {r} works very nicely inside {efn}. EEng (talk) 00:37, 9 July 2015 (UTC)

How could a 2015 news article be cited in a 2013 version of Wikipedia article?

This has me puzzled. (https://en.wikipedia.org/w/index.php?title=French_detainees_at_Guantanamo_Bay&oldid=546314409) How could "June 2015" be discussed in 2013? See footnote 2 at the link.Anythingyouwant (talk) 00:17, 9 July 2015 (UTC)

Looks like a simple typo. As the note on the footnote says, the date is in the URL, and the date was 2005, not 2015. I suspect the date in prose is also a typo, they probably meant 2012 instead of 2015, but that is also an easy error to make since 5 is above 2 on a numeric keypad. Resolute 00:21, 9 July 2015 (UTC)

The article has a template, `{{RemainingAtGuantanamo}}`, that has a 2015 citation. Old copies of articles use current versions of templates that they transclude.

—Trappist the monk (talk) 00:24, 9 July 2015 (UTC)

Did you click above where it says "This has me puzzled" and look at footnote 2? The article in footnote 2 is this one (http://www.bloomberg.com/news/articles/2015-06-13/u-s-govt-transfers-6-guantanamo-bay-prisoners-to-oman) which is clearly from 2015, right?Anythingyouwant (talk) 00:31, 9 July 2015 (UTC)
Oh, I see, the template did it, thanks.Anythingyouwant (talk) 00:32, 9 July 2015 (UTC)

AfD issue?

Recently, I've noticed that when someone (myself included, otherwise I may not have noticed this) nominates a page to Articles for Deletion, the system sometimes decides the result of the discussion is keep when either Twinkle or Page curation puts the relevant code on the page, nullifying the link to the discussion (the discussion is still there, it's just not being linked to properly) on the template and the user warning I keep having to remove the |result=keep to get the links working. Here (https://en.wikipedia.org/w/index.php?title=Visa_requirements_for_Somali_citizens&type=revision&diff=670447957&oldid=670435938) is an example (note that I didn't nominate this page, so it's not my computer) Have those articles been nominated before? (if so, then the nominations aren't being handled properly) Has anyone else encountered this? Adam9007 (talk) 00:50, 9 July 2015 (UTC)

It doesn't "decide" keep and it doesn't "nullify" any link. The code is in comments `<!-- ... -->` so it has no effect while it's there. If the decision turns out to be delete then the talk page is also deleted so there will be no delete decision to record there. The code is only needed for non-deletions and keep is the most common result in those cases. In some cases it should be manually changed to something else like "No consensus" when it's copied to the talk page. The link to the AfD page is sometimes red initially when tools are used for the nomination. This is because the tools can edit so quickly that when the nomination template is placed on the article, the MediaWiki software hasn't yet registered that the AfD page has been created. This is fixed by purging the article, or making any edit like you did but that is not necessary. PrimeHunter (talk) 01:33, 9 July 2015 (UTC)

- ~~It does seem to have an effect, the same thing just happened to me again here (https://en.wikipedia.org/w/index.php?title=Sarah_Bilingual_Experimental_High_School&type=revision&diff=670612526&oldid=670612476)~~ I just put the |result=keep in again and it had no effect, it seems you're right in that I sometimes have to make an edit. It's still annoying though. Adam9007 (talk) 02:22, 9 July 2015 (UTC)

No edit is required. You only have to purge it. See Wikipedia:Purge. PrimeHunter (talk) 03:09, 9 July 2015 (UTC)

User Contribution Search

When you search with User Contribution Search (https://tools.wmflabs.org/usersearch/index.html), the URLs say "http://https". Instead, they should say "https://". Geoffrey/T2000 (talk) 04:40, 9 July 2015 (UTC)

Tracked in Phabricator
Task T104812

They seem fine to me. I notice that after running a query, it says "Bugs, suggestions, questions? Contact the author at User talk:Scottywong" at the bottom. --Redrose64 (talk) 10:56, 9 July 2015 (UTC)

See phab:T104812. --AKlapper (WMF) (talk) 11:21, 9 July 2015 (UTC)

I have also posted to User talk:Σ/Archive/2015/August#Usersearch bugs. PrimeHunter (talk) 11:49, 9 July 2015 (UTC)

Thanks given and received

Is there anywhere one can go to see the thanks one has given and received on Wikipedia? DuncanHill (talk) 13:22, 9 July 2015 (UTC)

@DuncanHill: You're looking for Special:Log/thanks. — *Mr. Stradivarius* [talk] 13:23, 9 July 2015 (UTC)

Thank you. As a corollary - if one clicks "give thanks for this edit", then clicks "no" when it asks to give public thanks, does it give private thanks? DuncanHill (talk) 13:27, 9 July 2015 (UTC)

And another - Special:SpecialPages doesn't list Special:Log/thanks - should it, and if so, how do we get it to? DuncanHill (talk) 13:29, 9 July 2015 (UTC)

You cannot give "private" thanks. "no" cancels the thanks and leaves no trace anywhere. Special:Log/thanks is part of Special:Log which is linked on "Logs" at Special:SpecialPages#Recent changes and logs. There shouldn't be a direct link. PrimeHunter (talk) 13:44, 9 July 2015 (UTC)

(edit conflict) As to your first question, no, it doesn't thank the user at all. And as to your second, you can get there through the "All public logs" dropdown menu at Special:Log (it's labelled as "Thanks log"). — *Mr. Stradivarius* [talk] 13:46, 9 July 2015 (UTC)

Thanks all. DuncanHill (talk) 13:51, 9 July 2015 (UTC)

Cannot link to other Wikis

I wrote an article Charlotte Dubreuil and am unable to link it to the French counterpart. When I click on the add link button, it just says "error"? SusunW (talk) 17:34, 9 July 2015 (UTC)

Problem seems to be resolved. Thanks! SusunW (talk) 17:42, 9 July 2015 (UTC)

You resolved it correctly. --Izno (talk) 17:51, 9 July 2015 (UTC)

"Back to top" link/button

I noticed this was proposed a while back, but I think Wikipedia would greatly benefit from a back to top button. This button could be a round one which could be situated beyond the margin of the text, or if that can't be avoided, within the text if if can fade slightly so that the underlying text is shown. I find this very frustrating when I'm scrolling down long articles and that I have to manually scroll all the way back up to the top if I want to go there. I use an Apple Macbook and an iPhone 5 and this would greatly benefit Wikipedia. Sam.gov (talk) 16:31, 9 July 2015 (UTC)

Especially on mobile devices, I'm guessing? Eman235/*talk* 16:53, 9 July 2015 (UTC)

Yes, especially on mobile devices; they're generally smaller. Sam.gov (talk) 16:57, 9 July 2015 (UTC)

Small tip in that case. On iOS, you can always double tap the statusbar, to bring you to the top (in any iOS app) —TheDJ (talk • contribs) 18:03, 9 July 2015 (UTC)

Thanks, just now remembered that. Thanks for the reminder. Sam.gov (talk) 20:33, 9 July 2015 (UTC)

- Please, no RfCs at VPT. This is a WP:VPR matter. --Redrose64 (talk) 18:40, 9 July 2015 (UTC)

Ok, Thanks for letting me know. I'll bring an RfC to WP:VPR next time. Sam.gov (talk) 20:33, 9 July 2015 (UTC)
I just now removed the RfC template. Sam.gov (talk) 20:39, 9 July 2015 (UTC)

Thank you --Redrose64 (talk) 21:52, 9 July 2015 (UTC)

User:Numbermaniac/goToTop
http://codepen.io/rdallaire/pen/apoyx - NQ (talk) 18:48, 9 July 2015 (UTC)

I'll try this script. Sam.gov (talk) 20:33, 9 July 2015 (UTC)

Pending changes - bad link in dialog screen

If you have an article on your watchlist which is under pending changes, say Malala Yousafzai, and you consider that a particular series of revisions (for example these eight (https://en.wikipedia.org/w/index.php?title=Malala_Yousafzai&action=history&offset=20150709080000&limit=8)) are not acceptable, and decide to revert the whole group of eight by using the Unaccept revision feature, something rather odd happens.

You get a screen which lists the eight revisions, in reverse chronological order with dates and times, which is OK, the list is preceded by a note that uses MediaWiki:Revreview-reject-text-list as a framework. This MediaWiki: message is constructed such that $2 is supposed to be for the name of the page. What is actually fed in through $2 is the revision ID of the second edit in the group being reverted, but with commas to separate the number into groups of three digits. So in the above example, the second edit is that of 05:58, 9 July 2015, whose revision ID is 670635145, and so it shows "... from the following revisions of **670,635,145**:" when it should be showing "... from the following revisions of **Malala Yousafzai**:"

How can I find out what other values are fed into MediaWiki:Revreview-reject-text-list so that I can fix that $2 to something else? --Redrose64 (talk) 08:33, 9 July 2015 (UTC)

I saw that myself a while ago and made https://gerrit.wikimedia.org/r/#/c/222211/ . Aaron Schulz 16:45, 10 July 2015 (UTC)

Apparently it's in mw:MediaWiki 1.26/wmf13 which went live here on 9 July 2015. Presumably it was after my post above. --Redrose64 (talk) 17:59, 10 July 2015 (UTC)

Error 503 Service Unavailable

I was getting a 503 error when browsing Wikipedia for a minute. What happened? Gparyani (talk) 17:31, 9 July 2015 (UTC)

Also, when I try to view older notifications at Special:Notifications by clicking "More", I get "An error occurred while fetching results". Gparyani (talk) 17:35, 9 July 2015 (UTC)

Hmm, I cannot reproduce that problem. If you get an explicit error message, it's welcome (though please remove any personal information such as your IP address). --AKlapper (WMF) (talk) 09:46, 10 July 2015 (UTC)

@AKlapper (WMF): That problem was just gone after I posted that message. The 503 errors were occurring for a minute, then as soon as they stopped, I posted it. You don't have any logs? Gparyani (talk) 18:14, 10 July 2015 (UTC)

Tables

Hello everyone! For this year, the WikiProject Formula One has introduced new tables for its race reports. See 2012 Brazilian Grand Prix and 2015 British Grand Prix for the difference. The new format seems to make problems on Firefox. As you can see on my screenshot, the borders ofter don't appear, which makes the tables hard to read. Any idea why that happens? Zwerg Nase (talk) 14:07, 7 July 2015 (UTC)

Tracked in Phabricator
Task T105596

They use the obsolete "border" attribute, which is no longer supported by modern browsers. Use CSS. —TheDJ (talk • contribs) 14:47, 7 July 2015 (UTC)

It's weird though, there already is a CSS fallback defined, and that should take precedence I think. Might be a FF bug. —TheDJ (talk • contribs) 14:51, 7 July 2015 (UTC)

Borders on the table element do not cascade to the cells and never have with inline CSS. --Izno (talk) 14:56, 7 July 2015 (UTC)

Rigth, where border attributes do, because the attribute also affects the value of the rules attribute, which does provide borders between the cells. As in HTML4 tables spec (http://www.w3.org/TR/html401/struct/tables.html#h-11.3.1). But inline CSS doesn't, so you should just use wikitable. —TheDJ (talk • contribs) 15:17, 7 July 2015 (UTC)

Frankly, please stop using custom CSS for the tables. I see no reason not to use the `wikitable` class. --Izno (talk) 14:56, 7 July 2015 (UTC)

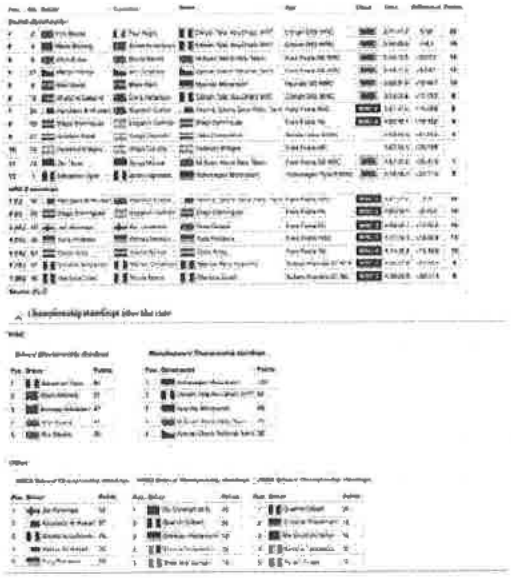
I think it has to do with them wantint to waste less horizontal whitespace, then wikitable allows them. But yes, as proven right here, such an approach is not maintainable.. —TheDJ (talk • contribs) 15:17, 7 July 2015 (UTC)

@Zwerg Nase: I can't find any table which resembles your screenshot in either 2012 Brazilian Grand Prix or 2015 British Grand Prix - which sections are they in? --Redrose64 (talk) 15:24, 7 July 2015 (UTC)

@Redrose64: Sorry, the screenshot is from 2015 Formula One season, but its the same sort of table with the same problem. Anyway, thank you for clearing that up, I will propose to the Project to go back to the wikitables. Zwerg Nase (talk) 15:40, 7 July 2015 (UTC)

They were coded that way because tables that use the "wikitable" class have barely visible outlines on the mobile site. Like in this example:

A a group of wikitables in a rally article on the desktop site



The same tables on the mobile site

@Zwerg Nase: can you please tell me the exact version of FF and operating system that you are using please ? i'd like to keep an eye on this but i've not yet found a version of FF for my Mac with this same problem. —TheDJ (talk • contribs) 15:54, 7 July 2015 (UTC)

@TheDJ: FF 38.0.5 on Win 7 SP1. Zwerg Nase (talk) 16:00, 7 July 2015 (UTC)

"barely visible outlines on the mobile site" -> Then you need to submit a ticket to get the mobile site fixed, not introduce arbitrary styling. Fix the root cause, not the symptom. --Izno (talk) 16:31, 7 July 2015 (UTC)

I did already launch a proposal to fix this, but it failed to get the problem understood. Tvx1 16:40, 7 July 2015 (UTC)
Izno, I could really use some advice here. I already attempted to have this root cause fixed but was met with fierce opposition. I agree wholeheartedly with your advice to fix the root cause, not the symptoms. However when I tried to I got the exact opposite advice, namely to fix the symptom, not the root cause. Two of the users who replied to my proposal back then have replied to this section as well (Redrose64 and TheDJ). If want to go through this hassle again I want to make sure that I make the extent of the problem clear. As TheDJ knows, I'm on Phabricator as well so would be prepared to go there if that's the best approach. Tvx1 15:24, 9 July 2015 (UTC)

The table at 2015 British Grand Prix#Qualifying uses attributes `border="2" cellpadding="2" cellspacing="0"` all of which are obsolete. The `border=` attribute is retained in HTML5, but only with the specific value `border="1"` (https://www.w3.org/TR/html5/tabular-data.html#attr-table-border). As for 2012 Formula One season#Drivers' standings and 2015 Formula One season#World Drivers' Championship standings, these both use obsolete attributes - but different ones. The 2012 table uses the `valign=` and `align=` attributes, both are obsolete in HTML5; the 2015 table uses `border="2" cellpadding="2" cellspacing="0"` with the same issues as noted earlier. --Redrose64 (talk) 16:02, 9 July 2015 (UTC)

And what are the HTML5 replacements for these obsolete parameters? I know `align=` should be replaced by `style="text-align:"` (and I have replaced quite a few of them, although it's an unmanageable task to replace them across the whole 'pedia), but I don't know the others. Tvx1 16:13, 9 July 2015 (UTC)

There aren't any, not in HTML5, anyway. Like most other attributes (and elements) that are concerned purely with visual appearance without any semantic meaning, they're no longer part of HTML, since in virtually all cases CSS can do the same thing, in a uniform way (see this doc (http://www.w3.org/TR/html5/obsolete.html#non-conforming-features), where it lists `cellpadding` and `cellspacing` with a whole bunch of others, at the bottom of which it says "Use CSS instead."). You *can* put CSS inside a HTML `style="..."` attribute, but that's tedious, error-prone and bloaty. Far better to do it through a class, with the styling set up somewhere like MediaWiki:Common.css. So, what is basically wrong with the existing `wikitable` class? --Redrose64 (talk) 16:55, 9 July 2015 (UTC)

> It's specifically the mobile `wikitable` class. The table outlines described in that are barely visible, as you can see in the picture above (specifically the one on the right). And for instance if you just compare tables at about just about any article using them. Like for instance this desktop version and its Mobile (http://en.m.wikipedia.org/wiki/2010_British_Grand_Prix) equivalent. Tvx1 17:50, 9 July 2015 (UTC)

The opposition you met was with the specific solution you proposed, which was to make the tables render *precisely* the same. There are some problems you noted as what I would personally evaluate as accessibility-related and so should be changed (borders at least, potentially background). My advice is thus that you submit a phab ticket for what might be considered real problems with the mobile skin. --Izno (talk) 16:59, 9 July 2015 (UTC)

> In that case I was misunderstood. My main concern was and still is the readability issues of the mobile "wikitables". I just used at the desktop table as the norm because it thought it was preferable to limit the differences between skins, without them being *precisely* the same. Tvx1 18:12, 9 July 2015 (UTC)
> (edit conflict) Izn, which way do you think the accessibility problem works? The dark lines on desktop make it hard for some people (e.g., with dyslexia) to separate letters from borders (and therefore to read the contents), or the faint lines on mobile make it difficult for some other people to keep track of which line they're reading across a wide page? (Maybe we should go back to green bar style.) Whatamidoing (talk) 18:17, 9 July 2015 (UTC)

> > I know this question is not direct at me, but the observations I have seen while working in my area of editing all point at the second problem. I don't think the lines on desktop are dark enough to cause problems. They are clearly lighter than the text. Tvx1 19:02, 9 July 2015 (UTC)
> > @Whatamidoing: *g* I'm simply pointing out that there's probably a legitimate concern about accessibility of table borders hiding somewhere in there. My inclination as a normal-sighted user (never mind the probably larger audience of visually disabled [not dyslexic] and normal-sighted people), I would expect the larger problem to be faint lines, not dark lines simply by number of people. (To: Wald (WMF):) Has anyone done an analysis on the usability team of *either* the mobile or desktop viewing, specifically regarding tables? --Izno (talk) 19:16, 9 July 2015 (UTC)

> > > They've done usability testing in general; I don't know if they've specifically done testing regarding tables. I'll ask the Design Research team, and let you know if I find out anything interesting. While we're on the subject, you should all consider signing up for mw:Wikimedia Research/Design Research's studies. Problems identified in their research are systematically reported to the team responsible for the problem. Whatamidoing (WMF) (talk) 20:28, 9 July 2015 (UTC)

> > > > Ok, I went on and filed a ticket for this problem. Tvx1 14:47, 11 July 2015 (UTC)

Can't edit Wikipedia mobile

I am using Chrome 43.0.2357.61 with iOS 8.3. When I try to edit a page in mobile mode, the "Save" button disappears after I make my edit. Help!!! Grover cleveland (talk) 06:56, 11 July 2015 (UTC)

WGT Baseball:MLB deleted

Hi! I noticed that the article I originated, WGT Baseball:MLB, was deleted without as far as I know, even a VFD or Speedy Deletion tag. I saw on the deleted link it was taken off by a bot. I was wondering what happened and maybe someone can re-post it? I mean if this is a case of just a bot randomly deleting an article then that could be a problem in the future for other articles. *Antonio Wonder Why Martin* (dime aca) 04:09, July 10, 2015 (UTC)

> @AntonioMartin: What happened was this:
>
> 1. 30 May 2015: WGT Baseball:MLB was moved to a new title, WGT Baseball: MLB, by User:Anarchyte, with the edit summary "Space after colon". This left a redirect page at "WGT Baseball:MLB" pointing to "WGT Baseball: MLB".
> 2. 5 June 2015: Anarchyte nominated "WGT Baseball: MLB" for deletion via WP:PROD, with the reason "The only references that aren't primary are press releases. May not be notable".
> 3. 12 June 2015: User:Ged UK deleted "WGT Baseball: MLB", repeating Anarchyte's PROD rationale. This broke the redirect at "WGT Baseball:MLB", as it was then pointing to a non-existent page.
> 4. 13 June 2015: User:AnomieBOT III automatically deleted the broken redirect.
>
> The bot is functioning as it is supposed to: it looks like you are just being confused by the page move. If you would like the page restored or moved to your userspace, the best thing to do is ask Ged UK for advice, as they deleted the article. Best — *Mr. Stradivarius* [talk] 05:28, 10 July 2015 (UTC)

> > @AntonioMartin: As the article WGT Baseball: MLB was deleted as an expired WP:PROD, you can use WP:REFUND to get that undeleted, but it's not likely that the bot-deleted redirect WGT Baseball:MLB (for which WP:CSD#G8 applied) will also be undeleted. --Redrose64 (talk) 11:54, 10 July 2015 (UTC)

Loss of session data

When replying to people on talk pages (which can sometimes take more than just a couple of minutes), I'm increasingly often getting an apology for not processing my save because of a 'loss of session data'. Trying to save again (with my edit clearly there in the edit window) does no good. I have to copy my edit, go back to the previous version, open up and paste. This isn't due to anyone else editing the page and causing a conflict. It's starting to get irritating. Is something going on (or going wrong)? Peridon (talk) 20:25, 11 July 2015 (UTC)

> How long does it take for you to type out your posts? When I do spend some time collecting sources and text before saving, it does make that same error for me. Jo-Jo Eumerus (talk, contributions) 20:32, 11 July 2015 (UTC)

> > I tend to make fairly detailed posts, but I always have. I've even left the edit window open while I've dealt with phone calls and so on before. Now, it's coming up every day, or so it seems. Before, once a year or so. Peridon (talk) 20:42, 11 July 2015 (UTC)

- It's been going on for a good month. See the phabricator ticket (https://phabricator.wikimedia.org/T102199). Length of time to complete the edit has not been relevant in my experience. Risker (talk) 21:21, 11 July 2015 (UTC)

> > This is Wikipedia:Village pump (technical)/Archive 137#"Loss of session data" error on Save page and Wikipedia:Village pump (technical)/Archive 137#Session data loss message. It happens on about 5-10% of my edits, is not restricted to discussion pages but can happen in any namespace, and although it is more likely for those where I've had the editing window open for some time, it has happened when no more than ten seconds have elapsed between "edit" and "save page". --Redrose64 (talk) 21:33, 11 July 2015 (UTC)

Happened to me yesterday, once. Edit time <20 mins, Win7, IE11. I simply resubmitted by clicking "Save page" again and was successful. Akld guy (talk) 05:22, 12 July 2015 (UTC)

I'm on XP Pro, FF20 and Monobook. That doesn't work for me - I've tried Save six times or more on one edit. Peridon (talk) 11:40, 12 July 2015 (UTC)

File upload problems

The following discussion is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

Problem solved. (non-admin closure) Erpert blah blah 06:08, 12 July 2015 (UTC)

I have tried to upload an album cover twice, and both times I received an error stating *"invalid token"*. Anyone know what that means? Erpert blah blah blah 03:51, 12 July 2015 (UTC)

It must have been some sort of glitch, but everything is fine now. Erpert blah blah blah 06:08, 12 July 2015 (UTC)

The discussion above is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

New contribution?

I noticed a set of new buttons just appearing at the top of my contributions page. Was this just added? Gparyani (talk) 06:36, 12 July 2015 (UTC)

I don't see anything different. Local toolbar maybe? -- [[User:Edokter]] [[talk]] 10:00, 12 July 2015 (UTC)

That sounds like the Content Translation beta feature. It's at the top of my contributions page as well, because I have ticked the checkbox in my beta features preferences that says "Automatically enable all new beta features". — *Mr. Stradivarius* talk 14:01, 12 July 2015 (UTC)
I've added a couple of screenshots so people can more easily understand what Gparyani and I are talking about. — *Mr. Stradivarius* talk 14:15, 12 July 2015 (UTC)

ContentTranslation popout menu. This appears when you hover over the "Contributions" link in the top right.

ContentTranslation "new contribution" menu. This appears at the top of Special:Contributions.

Number of page watchers

Per the announcement above, appending `?action=info` to the URL for this page now includes the second line:

- Number of page watchers 2,837
- Number of page watchers visiting recent edits 647

What does the second line mean? Why not "with recent edits"? The gerrit description (https://gerrit.wikimedia.org/r/#/c/193838/) is general and I'm wondering what values apply here. The description mentions a right to view the info, but when I tried from an IP (not logged in), the same info was displayed—does everyone have the right? Johnuniq (talk) 02:32, 10 July 2015 (UTC)

The code (https://gerrit.wikimedia.org/r/#/c/193838/15/includes/actions/InfoAction.php) is a bit more specific: if I've read it correctly, it refers to the number of watchers that visited the page since about 26 weeks before the latest edit. — *Mr. Stradivarius* talk 02:48, 10 July 2015 (UTC)

Yes, translatewiki:MediaWiki:Pageinfo-visiting-watchers/qqq says: "the number counts how many users have last visited the page 26 weeks or less (by default) before the latest edit to the page; in other words, watching users who may see a future edit within about 6 months". PrimeHunter (talk) 02:58, 10 July 2015 (UTC)

Thanks all, but does that mean there is a table showing who visited a page (in the last 26 weeks)? How else could it count the number of unique editors who view a page? The last proposal I recall to monitor which pages an editor *visits* was rejected on the basis of privacy concerns. Johnuniq (talk) 05:02, 10 July 2015 (UTC)

Well, it's not just "visited" a page, it's (a) has it on their watchlist AND (b) visited the page. I'm finding this uncomfortable - if for no other reason than that I thought this level of user-specific information was not kept in the database for longer than 3 months. Risker (talk) 05:34, 10 July 2015 (UTC)

To be exact, it is (a) has it on their watchlist, (b) visited the page and (c) the page has not changed since then. What can be derived, then, is whether or not the latest revision has been visited by the user. That is not to say it does not have privacy implications.--Anders Feder (talk) 06:14, 10 July 2015 (UTC)

I'm not sure about (c), or the number of active users would bounce around wildly after every edit to a page, from 0 immediately afterward to (potentially) hundreds or thousands of watchers who look at the page before it gets changed again. On articles that are (a) heavily watched and (b) being edited at a high rate of multiple edits per hour, the effect would be to keep resetting the active user number to 0 every time someone edited, whether or not there are dozens of editors who have the article on their watchlist and are viewing the page at that precise moment. Jimbo's talk page, the ANI and AN noticeboards, current events like election days, World Cup, major tragedies....would all have low "active watchers" because they're constantly edited. Meanwhile, it seems the data retention guidelines specifically mentions retention of page-browsing data, and says it would be a maximum of 90 days. This is where the disconnect is. Risker (talk) 16:09, 10 July 2015 (UTC)

Right now Page information (https://en.wikipedia.org/w/index.php?title=Wikipedia:Village_pump_(technical)&action=info) says "Number of page watchers visiting recent edits 644". Let's see what it changes to after I save this. --Redrose64 (talk) 16:31, 10 July 2015 (UTC)
"Number of page watchers visiting recent edits 644". So a page edit doesn't reset it. --Redrose64 (talk) 16:34, 10 July 2015 (UTC)
Yeah, it isn't too straightforward, but see mw:Manual:Watchlist_table#wl_notificationtimestamp. Unless I read it incorrectly, it tracks the timestamp of the latest revision the user has not yet visited.--Anders Feder (talk) 16:27, 10 July 2015 (UTC)

just one small comment/correction: the "26 weeks" referred above is actually a configuration settings (i.e., a value in LocalSettings.php), so it might be 26 weeks, or it might be something else for each individual wiki. the way i understand it, it means "any watcher who visited the page since last page edit, or no more than X time ago" (for active pages, only the 2nd part is meaningful, for stale ones, the 1st part may trump it). as to the "privacy" concerns: this is the same datum that allows the system to track whether you read a watched page since its last edit. i do not think the "privacy" is a legitimate concern: after all, the list of page you are watching is much more revealing than the timestamp of the last time you visited any specific one of them, so by "agreeing" to maintain a watchlist, you practically give up this piece of privacy.
i do not know it for a fact, but it's my understanding that those parts of the database (watchlists etc.) are censored out of the publicly accessible copies. peace - קיפודנחש (aka kipod) (talk) 22:07, 10 July 2015 (UTC)

> I think perhaps you've missed my point. When the data retention guidelines say that this information would be kept for no more than 90 days, one would expect it was kept no more than 90 days. It's quite possible that the right hand didn't recognize that the left hand had written this rule, although that guideline was created with the active participation of WMF developers who ought to know what is and is not available. Risker (talk) 22:29, 10 July 2015 (UTC)
>
> > i am not sure what "this information" you refer to is. my understanding is that your watchlist falls into the "Account settings" bracket, so the retention rule is "Until user deletes/changes the account setting." as far as i know, your actual browsing history is not kept at all - here we are talking about your watchlist. the fact that this "watchlist" contains a timestamp does not change that (as far as i understand, this timestamp does not represent your last visit to the page - it represent the timestamp of the oldest edit of the page you _did not_ read once you read the page, the timestamp disappears - it will be added again when someone edits the page, and will be cleared when you'll read it again, and so on. i may be wrong here). this may look like hair-splitting to some, but in my view it's simply a part of the watchlist, and nobody expects their watchlist to begin evaporating after 90 days. peace - קיפודנחש (aka kipod) (talk) 23:06, 10 July 2015 (UTC)

After a quick scan of the code linked above and mw:Manual:Watchlist table, I think the situation is as kipod describes—there is no privacy problem. A feature of watchists is that a user preference allows an email to be sent to the user if a page on their watchlist changes. That is governed (apparently—I know very little about MediaWiki) by the wl_notificationtimestamp field in the watchlist table. If an editor views a page on their watchlist, that timestamp is cleared. If someone else changes the page later, the timestamp is set to the time/date of that edit, and setting the timestamp will also send a notification email if enabled in preferences. The new "Number of page watchers visiting recent edits" is calculated by counting the number of people watching that page with a timestamp for the page less than 26 weeks old, or with a cleared timestamp. That counts all people who watch the page and who have viewed it either after the last edit, or less than 26 weeks before the last edit. The timestamp retains a tiny piece of information about the editor, but the database has to remember that the editor is watching the page, and the timestamp provides very little extra information. Johnuniq (talk) 03:42, 11 July 2015 (UTC)

> "26 weeks" is more precisely 180 days, the default at mw:Manual:$wgWatchersMaxAge. It's not changed for any Wikimedia Wikis in https://noc.wikimedia.org/conf/highlight.php?file=InitialiseSettings.php or https://noc.wikimedia.org/conf/highlight.php?file=CommonSettings.php. PrimeHunter (talk) 11:27, 11 July 2015 (UTC)
>
> > This is a brand new parameter added specifically for this extension/function; given it's less than a month old, it's no surprise that it hasn't changed. Risker (talk) 21:25, 11 July 2015 (UTC)
> >
> > > As one of the people who has been requesting something like this for a long time, I'm happy. If you want to know why this matters, then please take a look at the inactive vs active watchers on this very old WikiProject (https://en.wikipedia.org/w/index.php?title=Wikipedia:WikiProject_Contents&action=info). It has almost two thousand watchers! But 99% of them haven't edited in a long time, and only eight (8) of them have actually looked at the page during the last 180 days.
> > > Risker, there is no *universal* 90-day timer. If there were, then the watchlist feature wouldn't work for any page that you haven't visited within 90 days. On day 89, it would be saying that you haven't looked at Boring since it was last edited, and on day 91, it would be saying that you had, just because it was more than 90 days ago and it "forgot". Whatamidoing (talk) 03:28, 13 July 2015 (UTC)

Template failed to substitute on page creation

I just created Wikipedia:Articles for deletion/Scarus zufar and substituted Template:U into it, but the template was never substituted, nor transcluded (it was there as text only). After a null edit, the problem seemed to have fixed itself, and the template substituted itself correctly (Special:Diff/671069450). Gparyani (talk) 06:01, 12 July 2015 (UTC)

> You used code of form `{{subst:afd2 | pg=PageName | cat=Category | text=Why the page should be deleted}} ~~~~` with a second `subst` in the `text` parameter, but only the outer `subst:afd2` is performed on the first save. It's a bit complicated but see Help:Substitution if you are curious. PrimeHunter (talk) 11:35, 12 July 2015 (UTC)
>
> > @PrimeHunter: I wasn't aware that I was substituting into another substitution, as I used Twinkle to create the page. I was expecting it to substitute as I could only see one subst: (Twinkle was hiding the outer one). Perhaps I can get this "bug" fixed by asking the Twinkle developers to have the AfD module expand template substitutions before calling the API to save the edit. Gparyani (talk) 16:45, 12 July 2015 (UTC)

Is there a way to "walk" categories in WP?

What I'm talking about is IMSLP's Category Walker, which actually does many things. My favorite use, however, is intersecting categories -- for example, (http://imslp.org/index.php?title=Category:Scores_featuring_the_guitar&intersect=Folksongs&transclude=Template:CatIntro). I'd find this useful on Wikipedia, specifically for finding pages in Category:Example Topic that are also in Category:Wikipedia articles needing copy edit from January 2014, Category:Wikipedia articles needing copy edit from February 2014, and Category:Wikipedia articles needing copy edit from March 2014. Any tools/gadgets that can do stuff like this? Eman235/*talk* 18:25, 12 July 2015 (UTC)

> You want something that can intersect but already have something to intersect with? I'm confused. I know that Wikipedia:CatScan does intersection also. I think autolist (http://tools.wmflabs.org/autolist/index.php) can also do it but I haven't played at all with it. --Izno (talk) 19:53, 12 July 2015 (UTC)
>
> > Sorry, what I meant was I can intersect on IMSLP but not here. CatScan works great, thanks :) Eman235/*talk* 21:13, 12 July 2015 (UTC)

Manage TemplateData

Every time I open a template edit page, first the page loads and then, after a short time delay, the following pops in at the top of the page:

Manage TemplateData Information about TemplateData	Tracked in Phabricator Task T105685

Both are links and when they pop in, they push the rest of the page down several lines. I wonder if this drives anybody else nuts, too? There I am doing redirects in a sort of "fast edit mode", and each time I load an edit page I must adjust my "readiness" several lines downward. I wonder if there is a bug in phab that is set to fix this and make the above TemplateData links load the same time the rest of the page loads? Why is there a time delay? – *Paine* 16:40, 12 July 2015 (UTC)
PS I've looked in Prefs and there doesn't appear to be any box to check or uncheck to manage this. PS added by – *Paine* 16:43, 12 July 2015 (UTC)

> There's a delay because TemplateData is JavaScript, basically. The quick and dirty fix would be to reserve some space for TemplateData in your stylesheet, e.g. by adding a top margin to `.mw-editnotice-10` and `position: absolute` to `.tdg-editscreen-main`. Alakzi (talk) 16:48, 12 July 2015 (UTC)
> Lines 11 to 19 in User:Alakzi/common.css. Alakzi (talk) 16:57, 12 July 2015 (UTC)
>
> > Thank you so much for saving my sanity, Alakzi, and *Best of Everything to You and Yours!* – *Paine* 18:14, 12 July 2015 (UTC)
> >
> > > @Paine Ellsworth: Alternatively, if you never add templatedata, you can hide it as I did here. --Redrose64 (talk) 07:27, 13 July 2015 (UTC)
> > >
> > > > Yes, that works very well, too! Thank you very much, Redrose64! *Joys!* – *Paine* 07:40, 13 July 2015 (UTC)

It's annoying indeed. I've opened a bug report on this issue. --TheDJ (talk • contribs) 11:35, 13 July 2015 (UTC)

Cached?

I have a page set up, User:Resident Mario/godate, whose content is taken in via a script. The substance of the page is merely:

```
BOF |{#time:Y-m-d|Sunday - 14 days}| EOF
```

I am retrieving this information via `pywikibot`, but I get back a `21 June 2015`, even though the current archive is on `28 June 2015`. I think it's a caching issue: hard-purging the page in my web browser causes it to now correctly give me the `June 28 2015` date (specifically, `2015-06-28`). Is there a quick workaround? ResMar 19:17, 12 July 2015 (UTC)

> You could purge the page with mw:Manual:Pywikibot/touch.py. Alakzi (talk) 19:24, 12 July 2015 (UTC)
>
> > Alakzi: What method can I call within a script to do so? ResMar 19:43, 12 July 2015 (UTC)
> >
> > > You could use `os.system("python touch.py ...")`. Or you could rip the relevant bits of code. Alakzi (talk) 20:08, 12 July 2015 (UTC)
> > >
> > > > I can't figure out what the relevant bit of code is at a glance, it's a bit buried in the class. ResMar 20:36, 12 July 2015 (UTC)

I've never used pywikibot before, but it appears to be as simple as:

```
from pywikibot import Page, Site
godate = Page(Site("en", "wikipedia"), "User:Resident Mario/godate")
godate.purge()
```

You don't need to use `pagegenerator` or any of the factory classes since we're only working with the one page. Alakzi (talk) 20:51, 12 July 2015 (UTC)

> That is correct. When you have your page instance you can just call `purge()`. No need to call the touch script. Actually if you look in that script it'll show you how it purges pages. — xZise (disc) 21:15, 12 July 2015 (UTC)
>
> > Thank you, xZise. Alakzi (talk) 22:16, 12 July 2015 (UTC)
> >
> > > XZise, Alakzi: Thanks. I'm thinking I'll have to rewrite the method to use `pywikibot` request infrastructure and implementing this instead of using requests directly: see this bug (http://stackoverflow.com/questions/31375022/purging-the-cache-of-the-requests-library-in-python?noredirect=1#comment50729494_31375022) that I left on stackoverflow. Any idea what's going on there? ResMar 14:45, 13 July 2015 (UTC)
> > >
> > > > Not sure - probably something to do with how Varnish is configured. You can retrieve the generated text with `expand_text` in pywikibot. Alakzi (talk) 15:04, 13 July 2015 (UTC)

Redirects and includeonly

What happens if a template page containing an instance of <includeonly>#REDIRECT [[Target]]</includeonly> is transcluded or substituted? GeoffreyT2000 (talk) 02:19, 13 July 2015 (UTC)

> @GeoffreyT2000: I'm not sure, but my guess is that substitution would work and transclusion wouldn't. Why don't you give it a try? You can use the test wiki to avoid adding clutter here if you want. — *Mr. Stradivarius* ♪ talk ♪ 05:33, 13 July 2015 (UTC)
> Transclusion inserts the text "REDIRECT [[Target]]", but only when the page is rendered. Substitution inserts the text "#REDIRECT [[Target]]", which creates a redirect if it's the first thing in the article. --Unready (talk) 09:10, 13 July 2015 (UTC)

Apparent glitch in mobile browser

While recently viewing the 2015 Formula One season article, I noticed two uses of the {{refn|group=N}} template. This seems to have caused a glitch where clicking on one of the footnotes doesn't do anything. It's a lot like having two subsections with the same name on one article. The problem doesn't exist when using any other format. Prisonermonkeys (talk) 09:26, 13 July 2015 (UTC)

> I see only one refn there .. ? —TheDJ (talk • contribs) 11:29, 13 July 2015 (UTC)
>
> That should be solved by using a different group name for each of the instances the ref template is used. Tvx1 21:12, 13 July 2015 (UTC)

Tech News: 2015-29

Latest **tech news** from the Wikimedia technical community. Please tell other users about these changes. Not all changes will affect you. Translations are available.

Recent changes

- You can now see a list of pages with errors in code coloring. [13] (https://phabricator.wikimedia.org/T103586)

Problems

- There was a problem with editing on Thursday. Some tools like bots and VisualEditor were broken on all wikis for 10 minutes. [14] (https://wikitech.wikimedia.org/wiki/Incident_documentation/20150709-poolcounter)
- There was a problem with images on Thursday. They were broken on all wikis for 15 minutes. [15] (https://phabricator.wikimedia.org/T105304)

Changes this week

- The new version of MediaWiki will be on test wikis and MediaWiki.org from July 14. It will be on non-Wikipedia wikis from July 15. It will be on all Wikipedias from July 16 (calendar).

Meetings

- You can join a technical meeting at Wikimania in Mexico City this week. [16] (https://wikimania2015.wikimedia.org/wiki/Hackathon)

Tech news *prepared by tech ambassadors and posted by bot • Contribute • Translate • Get help • Give feedback • Subscribe or unsubscribe.*

15:06, 13 July 2015 (UTC)

Admin Edit Flag

OK, so we've had some discussions about Administrative Actions. This comes up now and again, sometimes in the context of Arbitration Enforcement, sometimes in the context of administrator inactivity (Do edits count, or just actual, official administrator actions like deleting pages or blocking?), and other times in the context of closing discussions and requests (like declining unblock requests, for example). The trick is that the system, at present, only logs administrator actions that require the tools - protecting and unprotecting pages, blocking or unblocking editors, deleting or undeleting edits or pages, etc. So, sometimes it is confusing to tell whether an editor who has the administrator tools is actually acting in their official capacity as an admin, or just chiming in as an editor. Different answers to that question can carry different weight - taking an admin action does not make an editor involved in the discussion, but commenting as an editor might. And whether the editor is involved or not can be relevant, especially in a hot dispute.

So I had an idea. We can flag our edits as "minor" if we wish, simply by ticking a box in the edit dialogue. Would it be possible to add a similar flag for Administrative Actions? Obviously, we would need rules to govern its use, just as we have policies in place that require proper use of the "minor edit" flag (and can result in blocks for repeated misuse). But if used properly, an "Admin" flag might give administrators a way to clarify when they are "on duty" and acting in their official capacity.

Two questions, then - 1) Is this a technically workable idea? It would seem so, but I don't know how simple it would be to implement. 2) Is this something the community would have an interest in, either as a short-term trial or a long-term experiment? Thoughts? UltraExactZZ Said ~ Did 13:35, 8 July 2015 (UTC)

> My concern here is that admins are not moderators *de jure*, merely *de facto*; nor do admins have editorial precedence over other users. Implying that edits can be "admin actions" seems contradictory to those principles. {{Nihiltres|talk|edits}} 16:24, 8 July 2015 (UTC)
> This sounds like a good idea on paper, but the thing is that it's not really up to the choice of the admin to decide whether something is an "admin action" or not. The very nature of the edit is what determines that. So I can't really see that something like this would be worth it; if we take the label at face value, it's extremely prone to abuse, and if we don't, then I don't really see it providing any real benefit. Writ Keeper ☺♕ 16:38, 8 July 2015 (UTC)
>
> > This edit (https://en.wikipedia.org/w/index.php?title=Ariana_Grande&diff=prev&oldid=670519921) was an admin action; this edit (https://en.wikipedia.org/w/index.php?title=Ariana_Grande&diff=prev&oldid=670531536) 99 minutes later was not, although it was to the same page. Why do I describe them so, when only the *second* involved the {{pp-vandalism}} template? Because when I made the first edit, the article was under full protection; by the time that I made the second one, the prot had been reduced to semi, so any confirmed user could have done the same. These two edits are recorded similarly in my contribs (https://en.wikipedia.org/w/index.php?title=Special:Contributions/Redrose64&offset=20150708154500&limit=12&namespace=0) and the page history (https://en.wikipedia.org/w/index.php?title=Ariana_Grande&action=history&offset=20150708154500&limit=7); to know that the first one was when the page was full-prot, and therefore an admin action, you need to look at the prot log for the page (https://en.wikipedia.org/w/index.php?title=Special%3ALog&type=protect&page=Ariana+Grande&offset=20150708154500&limit=2) as well. --Redrose64 (talk) 16:57, 8 July 2015 (UTC)
> >
> > > Sure, I understand that, and I see how there could be value there; like I say, it sounds good on paper. But I'm worried about the implications. A hypothetical admin in an edit war could label all their reversions as admin actions in an attempt to immunize themselves from 3RR, and even from INVOLVED when they finally block the other hypothetical participant--after all, they've only been involved with that user in an administrative capacity, just look at those edits all labeled "admin"! That'd be a misuse of the function, of course, but it provides that much more "justification" to abuse of admin tools, and it'd be indelible. As you point out, Redrose, it's difficult to tell whether an edit is an admin action or not without context, but that swings both ways; it also makes it hard to *verify* an applied "admin action" label, and the consequences of an edit falsely labeled as an admin action are likely more severe than the consequences of an unlabeled admin action. I suppose my initial reply was a false dichotomy, but I still think that this would ultimately be more trouble than it's worth. YMMV Writ Keeper ☺♕ 17:19, 8 July 2015 (UTC)

Edit summaries do rather nicely for this sort of thing. Chillum 17:20, 8 July 2015 (UTC)

> True - and nothing whatsoever stops me from making an edit with the summary "Minor edit - grammar and punctuation". The software just lets me identify it as such with an additional piece of data attached to that edit. This would be similar. UltraExactZZ Said ~ Did 20:31, 8 July 2015 (UTC)

Yes, it is technically possible. The implementation would end up looking something like mw:Extension:StaffEdits. Legoktm (talk) 02:40, 14 July 2015 (UTC)

Interwiki transclusions

Are interwiki transclusions possible? I appreciate that they would be generally frowned upon, but had the idea of having the same user page across my wikimedia accounts with the syntax {{w:User:Name/UserSubPage}}. DMOUSE 22:36, 11 July 2015 (UTC)

You can have a Wikimedia-wide userpage with this but transcluding crosswiki I don't think is possible (save for files on Commons). Jo-Jo Eumerus (talk, contributions) 22:41, 11 July 2015 (UTC)

There are three more that work: your userpage, global.js, and global.css pages at Meta. If you want the same userpage across Wikimedia accounts, then create that page at m:User:U+003F, and it will happen automagically. Whatamidoing (WMF) (talk) 03:32, 13 July 2015 (UTC)

No, interwiki transclusions are not possible. The 3 examples above (global.css, global.js and Files from Commons) aren't even transclusion :p
^demon[omg plz] 01:35, 14 July 2015 (UTC)

Nm in Keegan's contributions

The minor edit checkbox has been disabled on creations of new pages. Since when did this happen? However, Keegan has some edits with edit summary {{checkuserblock}} that are marked as both creations and minor. GeoffreyT2000 (talk) 02:49, 13 July 2015 (UTC)

Tracked in Phabricator Task T105763

I have no idea. The CheckUser block form does not offer the selection of marking an edit as minor or not. Keegan (talk) 04:50, 13 July 2015 (UTC)

I've just tried to use the API to force a couple of page creations to be marked as minor edits, but they weren't marked as minor in either case. So that rules out the hypothesis that all page creations from API calls using `action=edit` and `minor=` result in the edit being marked as minor. @GeoffreyT2000: It would help if we could see the actual edits in question - could you link to them? — *Mr. Stradivarius* ♪talk♪ 05:29, 13 July 2015 (UTC)

@Mr. Stradivarius: this post reminded me to follow up on the {{checkuserblock}} that likely inspired this, and the same result occurred (https://en.wikipedia.org/w/index.php?title=User:Make_up_your_time&action=history). Keegan (talk) 05:36, 13 July 2015 (UTC)

So it's in how the CheckUser form processes the option to mark user/usertalk pages. Extension:CheckUser likely needs to be updated. Keegan (talk) 05:37, 13 July 2015 (UTC)

It's some years since "This is a minor edit" checkbox was removed from page creations. --Redrose64 (talk) 16:32, 13 July 2015 (UTC)

Now tracked as phab:T105763. — *Mr. Stradivarius* ♪talk♪ 01:48, 14 July 2015 (UTC)

Meta-tags

Hi there. I added this talk page comment to the Chevrolet Suburban article. My comment will invite editors to this draft article. My concern is that the draft article is almost identical to the real article, and I'm afraid some search engine may mistakenly pick up the draft article. I don't want someone to Google "Chevrolet Suburban" and end up at the draft article. Is there some meta-tag that could be added to the draft article to prevent this from happening? Thanks! Magnolia677 (talk) 11:00, 14 July 2015 (UTC)

Greetings. You want {{Userspace draft}} for this job. Jo-Jo Eumerus (talk, contributions) 11:09, 14 July 2015 (UTC)

Thanks! Magnolia677 (talk) 12:08, 14 July 2015 (UTC)

Page curation

When you minimize the page curation toolbar, the word "Curation" appears upside down. GeoffreyT2000 (talk) 02:48, 12 July 2015 (UTC)

Tracked in Phabricator Task T105846

To me in Firefox it appears top-down, rotated 90 degrees clockwise from normal left-to-right writing. This is intentional. Do you mean it rotates 180 degress for you? In which browser? PrimeHunter (talk) 02:56, 12 July 2015 (UTC)

I can reproduce the issue in Internet Explorer 11 (File:Page curation upside down.PNG). @GeoffreyT2000: Are you using IE11? Gparyani (talk) 06:28, 12 July 2015 (UTC)

Verified - same flipping of the text when PC toolbar is minimized under IE 11. Never mind using a paragraph tag <p> to hold the text `Curation` is a bad idea to begin with since it has top & bottom margin lengths defined by default (which in essence become margin left & right when rotated btw), `pt#mwe-pt-toolbar-vertical` has a `[ms-]transform` setting to rotate 90 degrees (so + 90 degrees) in addition to an inline style for `writing-mode` set to *tb-rl* (+ another 90 degrees = 180 degrees hence the complete flip as in the screenshot). Apparently, that additional inline setting was meant specifically for IE 8 and lower (https://git.wikimedia.org/blob/mediawiki%2Fextensions%2FPageTriage/HEAD/modules%2Fext.pageTriage.views.toolbar%2Fext.pageTriage.toolbarView.js#L113) when detected but they're not using the corresponding ms- prefixed attribute for writing-mode (https://msdn.microsoft.com/library/ms531187(v=vs.85).aspx) at the same time for IE8 (or IE9) so I don't know what is going on there exactly.
~~Personally, I'd stop using the inline styling of `writing-mode` set to *tb-rl* altogether, let `[ms-]transform` rotate the text block per its 90 degrees setting and switch the current use of a paragraph element to a block level span or div instead to accomplish this. Of course, YMMV.~~ Obviously, something is crossed with the IE version detection and/or the appropriate attribute to use in each instance. -- ~~George Orwell III (talk) 09:18, 12 July 2015 (UTC)~~ George Orwell III (talk) 14:16, 12 July 2015 (UTC)

@George Orwell III: Can you file a Phabricator ticket, then? Gparyani (talk) 17:09, 14 July 2015 (UTC)

My search box is wider today

Did the search box get wider today? I have the "Widen the search box in the Vector skin" checkbox unchecked in Special:Preferences. (On a side note, within Special:Preferences itself, the search box is the normal length.) Gparyani (talk) 17:05, 14 July 2015 (UTC)

Yes, per discussion at Mediawiki talk:Vector.css#Improve sizing for search field. --Izno (talk) 17:08, 14 July 2015 (UTC)
As for your note, that is expected behavior. --Izno (talk) 17:12, 14 July 2015 (UTC)

@Izno: Enabling the preference does nothing on wide screens, but keeps the box wide on smaller screens (i.e. the size adjustment doesn't kick in when the preference is enabled). Gparyani (talk) 17:19, 14 July 2015 (UTC)

So turn the preference off. :D (I noted the presence of the preference at the MT:Vector.css.) --Izno (talk) 17:28, 14 July 2015 (UTC)

"So turn the preference off" – what do you mean? For us, it was never activated (and still isn't). The gadget needs to be removed now that the box is already bigger, and I would like to know how to return it to its previous size. Jared Preston (talk) 18:11, 14 July 2015 (UTC)

Yes, I have suggested a change that by default gives more space for the search field if you have a wide screen, and automatically takes up significantly less width on very narrow screens. I think it is an interesting experiment and I'm wondering how people will receive it. I suspect it might make it easier to find the search for most anonymous users, and i know for sure that it will improve the experience for people with narrow screens. My suggestion it that we evaluate after a month or so ? —TheDJ (talk • contribs) 19:22, 14 July 2015 (UTC)

@TheDJ: What will we do with the user preference, then? Gparyani (talk) 19:57, 14 July 2015 (UTC)

The gadget you mean ? Let's see after the month. If you remove or disable gadgets, it's really disruptive and it makes rolling back difficult. —TheDJ (talk • contribs) 20:01, 14 July 2015 (UTC)

So will you or someone else give us a way of shortening the box in the meantime? It's way too large for my requirements. Jared Preston (talk) 21:18, 14 July 2015 (UTC)

Your common.css allows you to override whatever you want to.

```
/* I want a 12.5em wide search box */
div#simpleSearch {
        width: 12.5em !important;
}
```

—TheDJ (talk • contribs) 22:02, 14 July 2015 (UTC)

You shouldn't need the `!important` because the selector has the same specificity as the one used in this edit (https://en.wikipedia.org/w/index.php?title=MediaWiki:Vector.css&diff=prev&oldid=671422869). –Redrose64 (talk) 22:59, 14 July 2015 (UTC)
I agree that `!important` isn't required. User css loads after global/site css, and last one to load wins. It would be required to override inline HTML styles, which this isn't. FWIW, I think the old size was 12.6em, but that's just a nit. --Unready (talk) 23:06, 14 July 2015 (UTC)

How to properly hide/show rows in table for a template?

I'm currently creating a template to replace the results tables being used at esports competition pages such as Evolution Championship Series and Apex (tournament series). The intention of the table is that it can show anywhere from 4 to 16 rows of data, but I can't get it to show anything past the first rows, with any data beyond the 5th row being shown as raw data above the intended table.

The template is currently at User:NeoChaosX/sandbox/doubleelimtemplate and I'm testing it on User:NeoChaosX/sandbox, where I have 6 rows of data filled in currently. What am I doing wrong here that's causing this not to work? NeoChaosX (talk, edits) 22:14, 12 July 2015 (UTC)

Alakzi appears to have fixed it by replacing {{|}} with {{!}}. PrimeHunter (talk) 22:55, 12 July 2015 (UTC)

Yeah, it's usually easier to develop templated tables with full HTML template markup, and then try to convert it to wikicode, in a sandbox, step-by-step (or not bother with the conversion). Wikicode's operator-overloading of the "|" character, its only-partial insensitivity to whitespace, and its nitpicky requirements for certain things (including template coding bits like |+, !, etc.) to be at the beginnings of lines, can make this a slow and frustrating process. – SMcCandlish ☺ ☎ ¢ ≽^•⊥•^≼ 06:54, 15 July 2015 (UTC)

Template rename server effect

In order to make a template name more concise and consistent with its sister template, {{R from diacritics}}, I have proposed that template {{R from title without diacritics}} be renamed to {{R to diacritics}}. An administrator, Diannaa, has raised the issue that template *R from title without diacritics* has more than 388,000 transclusions and that such a page move might have an adverse effect on our server load. Diannaa suggested that I raise this issue here to discuss this and find out more about the effects on the servers. So the question would be, will it be okay to move template *R from title without diacritics* to (presently one of its aliases) template *R to diacritics*? – *Paine* 22:15, 12 July 2015 (UTC)

It will not have an adverse effect. It might take a long time before all the uses are updated. Most important with such templates is that you shouldn't edit them too often, because each time you are dumping 388000 articles into a queue, which isn't something to do willy nilly. But overall the redirect or the change shouldn't matter too much. —TheDJ (talk • contribs) 11:25, 13 July 2015 (UTC)

That's good, and thank you, TheDJ! The main page to be moved was last edited on 19 June, and the redirect, which has been transcluded only about 440 times, was last edited about 18.5 hours ago. Should we wait a bit longer? Please forgive my ignorance of the timing range within which it is best to wait. Is there a guide on MW somewhere? Seems something this important should have a policy, don't you think? – *Paine* 15:16, 13 July 2015 (UTC)

@Paine Ellsworth:, that's just fine. Doing it 2 times in a single day would be suboptimal, doing it 10 times in a week also not really helpful, and edit warring would be really bad, other than that you are good. There is not really a guide, but as soon as you pass like 100000 transclusions you should simply prepare properly and make sure that you get it right in one edit, if at all possible. The site won't explode if you make multiple edits, not even with a 1.2 million transclusions, but it will introduce a large delay for all the other background updates, which editors sometimes find annoying. —TheDJ (talk • contribs)

Yes, thank you – I could find no policy nor guideline on MW except that they have two places in the Terms that allude to not being disruptive toward the servers, but no details. Quite possibly they would rather not put bad ideas in people's minds, for while the vast majority wouldn't dream of being disruptive, MW does seem to have its fair share of disgruntled ex-editors from the past who would just love to be able to edit a high-risk template or two. Anyway, thanks again and *Best of Everything to You and Yours!* – *Paine* 15:49, 13 July 2015 (UTC)

Thanks for that explanation, TheDJ. I think this accounts for odd delays I've sometimes experienced, with changes not showing up until after waiting a while and doing a purge. It happens much less today than it did only a few years ago, I supposed due to a combination of people being more careful and (probably more so) a more robust back-end, what with all of WMF's funding. – SMcCandlish ☺ ☎ ¢ ≽^•⊥•^≼ 06:50, 15 July 2015 (UTC)

Phabricator bug: Registering a Phabricator account with an incorrect email address

Tracked in Phabricator Task T105352

Hi!

If you have an account that has never used the Phabricator, and you go to phabricator.wikimedia.org, click the "Login or Register MediaWiki"-button and use an incorrect email address (I used idonothaveanemailadress@idonothaveanemailadress.sorry) then you are:

-unable to change your email address to a correct email address (you cannot access the email settings (https://phabricator.wikimedia.org/settings/panel/email/))

-unable to log in

I know that this is the wrong place to report this problem, but I cannot file a bugreport on the Phabricator... :-D

Can someone fix this bug and delete my email address so that I can chose another one?

The Quixotic Potato (talk) 18:14, 8 July 2015 (UTC)

> I don't really see a bug here but I can imagine that it's surprising behavior, indeed. Not sure if any clarification is needed in the Phabricator Help? In general, feel free to bring such topics and this specific case up on mw:Talk:Phabricator/Help - for example, I have no idea yet about the Phabricator username to reset. :) Thanks! --AKlapper (WMF) (talk) 11:29, 9 July 2015 (UTC)
>
> > @AKlapper (WMF): It is my username: "The Quixotic Potato". We do not need clarification in the Help, we need to change the acount registering procedure so that you can change your email address if you've entered a wrong email address. The Quixotic Potato (talk) 11:45, 9 July 2015 (UTC)
> >
> > > @The Quixotic Potato: Alright, I've filed phab:T105352. --AKlapper (WMF) (talk) 15:30, 9 July 2015 (UTC)
> > >
> > > > @AKlapper (WMF): You must be aware that that is not a solution to this problem. There are more than a hundred open bugs there, and some of the bugs that are open have been open for a very very long time (2014). You assigned the bug to no one, and you set the priority to low. Running that command is not enough, the acount registering procedure should be changed so that you can change your email address if you've entered a wrong email address. Is SCFC using the email address tim<at>tim-landscheidt.de? mw:User:MModell_(WMF) says he uses mmodell<at>wikimedia.org. Are you going to email them or should I do it? The Quixotic Potato (talk) 10:39, 14 July 2015 (UTC)
> > > >
> > > > > I know it's frustrating to not be able to login. Filing the bug is only the first step in the process. The status is meant to indicate actual reality: at the moment, there really is nobody working on it (or about to start working on it very soon). Therefore, no one is "assigned to" work on it (right now), even if the team already knows who is most likely to fix the bug in the end. That will change when someone is preparing to start work on it. (They do this so that other volunteers will know that this one isn't being actively worked on yet, and that therefore they can jump in and do it themselves without fear of wasting their time.)
> > > > > The traditional priority labels are a bit misleading for some teams. For this team, "high", "medium", and "low" mean something closer to "current work", "next up", and "later". "Later" could be as soon as a couple of weeks from now. What this setting means is only a statement of reality: at the moment, and compared to the other problems in the list, this is – purely from a practical, realistic standpoint – probably not going to be one of the next bugs to get fixed by the team (although anyone outside the team is welcome to do so at any time). Whatamidoing (WMF) (talk) 16:20, 14 July 2015 (UTC)
> > > > >
> > > > > > Thank you, I have spent over 7 years of my life working with various bug tracking systems (e.g. Trac) so I know how they work, which is why I wrote what I wrote. If I didn't understand how bug tracking systems work then I would probably be happy with AKlapper's reply. The Quixotic Potato (talk) 18:58, 14 July 2015 (UTC) p.s. phab:T105352 is not something someone "outside the team" can work on. Its not even something that requires any work. Someone needs to run a single command (phabricator/ $ ./bin/remove destroy @The_Quixotic_Potato). Only a very small group of people have the ability to run that command. If AKlapper would've wanted to help me he would've contacted one of them (e.g. via talkpage/email/IRC) instead of creating phab:T105352. I am not the only one with this problem, see phab:T99455.
> > > > > >
> > > > > > > I'm sorry this takes longer. I personally cannot fix this because I don't have sufficient permissions, and as can be seen in phab:T105352 Mukunda is subscribed to that task (and is one person who could fix this), hence he received a notification. So I *did* contact one of them. --AKlapper (WMF) (talk) 19:55, 15 July 2015 (UTC)
> > > > > > >
> > > > > > > ✓ **Done** Thanks. The Quixotic Potato (talk) 00:31, 16 July 2015 (UTC)

Citation issues

I am working on an article and have encountered 2 citation issues. The page is here: User:SusunW/Inter-American Commission of Women

- One issue is that a University journal included a reprint of a 1929 magazine article, thus I have a citation within a citation and I cannot figure out how to make it work. Technically, I think the proper rendition should be something like Lee, Muna. "The Inter-American Commission of Women". *Pan-American Magazine* (1929) contained in Cohen, Jonathan (ed). "A Pan-American Life: Selected Poetry and Prose of Muna Lee" Madison. Wisconsin, University of Wisconsin Press (2004). Can someone tell me how to put this properly in the citation template?
- 2nd issue is that for another reference my PDF file has a [] in it, which Wiki is seeing as code, rather than as part of the url. The PDF file name is http://www.oas.org/en/CIM/docs/PIA [EN].pdf which results in a "file not found" error if you attempt to access the link. Any suggestions?

Thanks for the help! SusunW (talk) 16:43, 13 July 2015 (UTC)

> The second issue can be solved by percent-encoding the brackets as `%5B` and `%5D`:
>
> - `[http://www.oas.org/en/CIM/docs/PIA[EN].pdf Non-escaped brackets]` → [EN (http://www.oas.org/en/CIM/docs/PIA).pdf Non-escaped brackets]
> - `[http://www.oas.org/en/CIM/docs/PIA%5BEN%5D.pdf Escaped brackets]` → Escaped brackets (http://www.oas.org/en/CIM/docs/PIA%5BEN%5D.pdf)
>
> $SiBr_4$ (talk) 16:50, 13 July 2015 (UTC)
>
> > Cool! I knew there had to be some way around the issue. Thank you!SusunW (talk) 17:02, 13 July 2015 (UTC)
>
> The usual solution I see to the first problem is just using two citations and manually connecting them, i.e. `<ref>{{cite magazine |…}} [Verb]ed in {{cite book |…}}</ref>`. {{Nihiltres|talk|edits}} 17:43, 13 July 2015 (UTC)

@Nihiltres: Thanks. I am pretty "technically" illiterate. If I cite two sources as you indicate in the text portion, why would not that be rendered as two separate citations in the "reflink" References section? I must be missing something, but since there is not anything physically "in" the References section, there is no way to type anything there. I'll try it and see what happens, as I see that your <ref> beginning and ending brackets are around the whole. SusunW (talk) 17:55, 13 July 2015 (UTC)

This would be only one entry in the references list but would render as "two citations". See below for what is probably the better solution. --Izno (talk) 18:06, 13 July 2015 (UTC)

That the article is a reprint in the new journal is more-or-less irrelevant to the citation since the only article that you looked at was in the context of the (more-)recent journal. I would use {{cite journal}}:

- {{cite journal |last=Lee |first=Muna |orig-year=Reprinted from 1929 |date=2004 |title=The Inter-American Commission of Women |editor-last=Cohen |editor-first=Jonathan |journal=A Pan-American Life: Selected Poetry and Prose of Muna Lee |publisher=University of Wisconsin Press |location=[[Madison, Wisconsin]]}}
- Lee, Muna (2004) [Reprinted from 1929]. Cohen, Jonathan, ed. "The Inter-American Commission of Women". *A Pan-American Life: Selected Poetry and Prose of Muna Lee*. Madison, Wisconsin: University of Wisconsin Press.

--Izno (talk) 18:02, 13 July 2015 (UTC)

@Izno: Its the other way around. The only part of the article I used was Muna Lee's 1929 text. I have no idea what else is in the 2004 volume and it is irrelevant to my article.SusunW (talk) 18:10, 13 July 2015 (UTC)

That doesn't make sense. Are you suggesting you have the 1929 text or the 2004 text? Only cite whichever you have access to. --Izno (talk) 18:12, 13 July 2015 (UTC)

@Izno: Actually it makes perfect sense. Someone put only the section of Cohens' piece that contained the Muna Lee article into a pdf format. It is quite clear that it was written in 1928 and published in 1929 from her description of events. However, since I do not have access to the actual Lee article, and have no idea how I would obtain it, the proper citation is indeed to give Cohen credit for the piece and a double citation is necessary, because the pdf shows that it is copyrighted in 2004 as part of Cohen's work. SusunW (talk) 18:29, 13 July 2015 (UTC)

If you have the book or magazine in your hand then that is what you should cite. If you are citing the pdf, then that is what you should cite. The pdf is clearly a transcription so it is not the magazine article nor is it the book. WP:SAYWHEREYOUGOTIT. Use {{cite web}}:

```
{{cite web |last1=Lee |first1=Muna |title=The Inter-American Commission of Women: A New International Venture |url=http://www.uhmc.sunysb.edu/surgery/IACW.pdf |website=Stoneybrook School of Medicine |accessdate=13 July 2015}}
```

→Lee, Muna. "The Inter-American Commission of Women: A New International Venture" (http://www.uhmc.sunysb.edu/surgery/IACW.pdf) (PDF). *Stoneybrook School of Medicine*. Retrieved 13 July 2015.

—Trappist the monk (talk) 18:41, 13 July 2015 (UTC)

@Trappist the monk: Thank you. Are you saying that if I am reading a book on-line I am to cite it as a web citation and not a book? That makes no sense to me, no place to put an ISBN code, etc. I can use that cite on on Lee piece, though if you access the document, it shows the double citation. SusunW (talk) 19:00, 13 July 2015 (UTC)

@Trappist the monk and SusunW: This is why I recommended the multiple-citation approach. The ideal here (and I believe WP:SAYWHEREYOUGOTIT supports me) is to cite the source read, and then indicate the chain back to the original publication. Here we'd ideally cite the PDF ({{cite web}}), list the PDF as citing/quoting the book ({{cite book}}), and list the book as citing/quoting the magazine article ({{cite magazine}}). A little convoluted, to be sure, but it makes the origin of the cited facts clear and gives readers multiple options for verification. {{Nihiltres|talk|edits}} 19:11, 13 July 2015 (UTC)

@Nihiltres: I concur. In your citation all parties are acknowledged for whatever copyrights exist. SusunW (talk) 19:15, 13 July 2015 (UTC)

If you are reading a book at Google books or Internet Archive, then cite the book as a book when those images that you see on your screen are facsimiles (scans) of the book. If you can read *A Pan-American Life: Selected Poetry and Prose of Muna Lee* online then you can cite that as a book. At Google books, the best we can get is snippet-view (https://books.google.com/books?id=n7xIAAAAMAAJ&focus=searchwithinvolume&q=The+Inter-American+Commission+of+Women) but that is sufficient to show that the pdf is not a facsimile but rather a transcription. So, if you have access only to the pdf, then that is the thing that you should cite and as such cite it using {{cite web}}. If all we have is the pdf, we do not know what editorial license was taken when the magazine article was made ready for publication in *A Pan-American Life*.

The snippet-view problem is only true of copyrighted works Google doesn't have license to distribute freely in full text; Google Books does in fact provide full-text PDF facsimiles of a large number of public-domain books; I use it to get them frequently. —SMcCandlish ☺ ☎ ≼^•ω•^≽ 22:18, 15 July 2015 (UTC)

It is not the function of a citation to state the publication history of a work. When citing Dickens' *Oliver Twist*, for example, we cite the particular book by publication date; we don't backtrack to the original *Bentley's Miscellany* publication of the serial parts. The purpose of a citation is to identify the source material that an editor is using to support a particular claim in an article. WP:SAYWHEREYOUGOTIT does not require editors to provide publication histories in citations. —Trappist the monk (talk) 19:58, 13 July 2015 (UTC)

Also, the copyright status of the work you read has nothing to do with the information a bibliographic citation should include. Parties get acknowledged in a citation for their intellectual contributions, not for their legal rights. Whatamidoing (talk) 00:02, 14 July 2015 (UTC)

@Nihiltres: @SiBr4: Brilliant! It worked. Except that [verb]ed rendered just like that showing the brackets, so I just typed contained in and it looks perfect. Two authors, two titles, two dates! Exactly what I needed. Thank you both, you are amazing and a valuable asset to us mere mortals without technical skill. SusunW (talk) 18:10, 13 July 2015 (UTC)

@SusunW: Right. I just used "[Verb]ed in" for the example because the relationship could vary: "quoted in", "contained in", "cited in" et cetera. Anyway, glad to help. :) {{Nihiltres|talk|edits}} 19:03, 13 July 2015 (UTC)

- Concur with Nihiltres. SAYWHEREYOUGOTIT wants you to cite the actual excerpt you found it in, which is a particular work. WP:RS wants to know why this is actually a reliable source, so you need to indicate the "chain" of citation back to the original, without misleading anyone into thinking you have on-hand either the original publication or the journal that republished it that lead eventually to the third-party excerpt you have. This has nothing to do with online vs. dead-trees copies of books. It has to do with editorial sources. Someone edited out the excerpt from a journal which republished. That's a chain of three, not two, editorial processes. The last in that chain could have introduced independent errors, which is why you have to cite where you actually got the material. By contrast, if my publisher used the final TeX draft of my book to produce a paper copy and an e-book for simultaneous release, as is typical today, that is a single editorial process, and you can use `{{cite book}}` to cite the online copy of the book (otherwise that template would have no `|url=` parameter, of course). If Project Gutenberg, Google Books, or Archive.org provide a PDF copy of a book scanned from the hardcopy in a library, you can cite it with `{{cite book}}`, too; it's a photographic facsimile (even if it also includes embedded OCR text that helped you find it and which makes it content-searchable by the user, at least it if it was done right); the URL says where you got it. Some would argue that you need to use the scanning date as the publication date and use `|origyear=` for the original book, and some might even make arguments about what goes in the `|publisher=` field, but I think that's splitting hairs for no reason. By stark contrast, if a text approximation of the book in .mobi or .epub form, produced purely through OCR (which is also common at all three of those sites), this is a separate editorial process (one that introduces usually quite numerous errors), and you should cite it specifically as a PG/Archive.org/Google Books e-book edition (including Google Books when it provides snippet view only), with its modern release date, and indicate the date of the original only with `|origyear=`; you cannot cite it as the original publication. That error rate alone is a very good reason to always prefer the PDF scan when you have a choice, which is usually the case [except for non-free stuff you get snippet view for]. An e-book textual approximation is arguably not really a reliable source due to this problem. At the very least, the article using such a source would be at the mercy of the trust level that GA/FA assessors have in the proofreading at PG/Archive.org/Google Books, and most of them surely know that the QA in this regard is very low where it even happens at all. Some of the OCR'd e-books you can download from these places are almost unreadable. I've totally given up on them, and work strictly from PDFs at these sites. – SMcCandlish ☺ ☎ ✍ >^‿‿^< 08:44, 15 July 2015 (UTC)

My first mapping effort

I would like to duplicate the map found here (http://www.nuclearabms.info/HSentinel.html) in a format useful for the Wiki. I was wondering if any map experts might help me, or point me to the right starting place? Maury Markowitz (talk) 12:03, 15 July 2015 (UTC)

@Maury Markowitz: Please clarify: Are you looking for a map expert to make the map, or a map expert to help you get started making your own maps? Both are probably available. —*Mendruss* ☎ 12:12, 15 July 2015 (UTC)
@Maury Markowitz: Try Wikipedia:Graphics Lab/Map workshop. --Redrose64 (talk) 12:27, 15 July 2015 (UTC)

Exactly what I was looking for RR! Maury Markowitz (talk) 20:10, 15 July 2015 (UTC)

Tfd2 and Tfm2

Recently, Alakzi has edited Template:Tfd2 and Template:Tfm2 so that they display a bulleted list. However, the bullet is on a blank line above the links. GeoffreyT2000 (talk) 22:24, 15 July 2015 (UTC)

Please name your browser and link to a section where you see the problem. PrimeHunter (talk) 22:51, 15 July 2015 (UTC)

I use Internet Explorer 11. The problem appears on the template pages themselves. GeoffreyT2000 (talk) 23:08, 15 July 2015 (UTC)
@PrimeHunter: Screenshot: File:Bullets on wrong line.PNG Gparyani (talk) 23:13, 15 July 2015 (UTC)

Thanks. I don't have the problem in Firefox but see it in IE 9.0. It's the same for the section below, no problem in Firefox but misplaced bullets in IE 9.0. PrimeHunter (talk) 23:20, 15 July 2015 (UTC)

Alakzi has fixed it for me with [17] (https://en.wikipedia.org/w/index.php?title=Template:Tfd_links&diff=prev&oldid=671626661). PrimeHunter (talk) 23:27, 15 July 2015 (UTC)

Green marker in watchlist for pages that have been changed

Not sure if this is the place to make this comment so feel free to move this comment to a more appropriate place if necessary.

The use of a green marker in the watchlist is a real concern for me and for the significant percentage of the population with some form of colour vision deficiency.

Wikipedia:Manual of Style/Accessibility#Color recommends that colour alone should not be used to mark important information. This should equally apply to non-article pages as well. The use of green in particular is a real accessibility no-no.

Can this be relooked at, please? -- Mattinbgn (talk) 23:01, 12 July 2015 (UTC)

@Mattinbgn: In Preferences > Gadgets under the Watchlist heading there's a "Display pages on your watchlist that have changed since your last visit in bold." option which should hopefully help :) Sam Walton (talk) 23:02, 12 July 2015 (UTC)

That was quick, thanks! -- Mattinbgn (talk) 23:09, 12 July 2015 (UTC)

This should probably be made the default. I'm not colourblind, but I've got real trouble parsing the watchlist with just the bullets. Alakzi (talk) 23:13, 12 July 2015 (UTC)

Silly me -- I thought it *was* the default. BTW once in a while I see a kind of grayish bold instead of the usual deep blue bold. I've never been able to figure out what that means. Anyone know? Also, can someone please update the collapsible "Legend" inside the "Watchlist options" box, so it will include all the new bells and whistles and colors and stuff? EEng (talk) 23:47, 12 July 2015 (UTC)

I'm not sure I've seen a gray bold, but I'd be happy to update the legend - what bells and whistles are you referring to? The green marker is a bit of a long explanation so probably best left as a sentence above unless you can think of a succinct way to write it. Sam Walton (talk) 00:03, 13 July 2015 (UTC)

Not gray, but a kind of medium-light blue/grayish blue. It's only about once every few days. As to the bells and whistles, I guess didn't really see the bit about "with a green symbol", which I think is new (?), so I get it now. But why not put the four symbols -- green bullet, green collapse arrow, blue bullet, blue collapse arrow -- in the legend, along with the explanation of bold? When one sees a legend, one

naturally gravitates to it, thinking it's the key to everything. I guess I just assumed that if the legend didn't explain the green/blue thing, nothing on the page did, and stopped looking around. EEng (talk) 00:41, 13 July 2015 (UTC)

> https://en.wikipedia.org/wiki/Special:Watchlist?uselang=qqx shows the legend is not made in one place but is part of the software. We are only meant to edit the text of each existing entry. "(wlheader-showupdated)" higher up refers to MediaWiki:Wlheader-showupdated which is displayed above the "Mark all pages as visited" button on the watchlist. There is logic in having those together. I suppose we could stuff a line break and another legend description into MediaWiki:Recentchanges-legend-heading but I don't support messing with the interface in that way. We have already added a help link to a section with more details. PrimeHunter (talk) 01:32, 13 July 2015 (UTC)
> There is another issue. The legend is also displayed at Special:RecentChanges which doesn't have markings of visited pages. I don't know whether MediaWiki:Recentchanges-legend-heading can use `{{PAGENAME}}` to see it's used on a Watchlist and only display additional content there, but this is getting further into messing with the interface. PrimeHunter (talk) 01:40, 13 July 2015 (UTC)
>
> > I leave this to your judgment, but thanks for the info. Listen, I have a question... In an earlier thread I suggested adding to the popups menu a "Mark this page as visited" option. I was given a not-very-convincing reason that this wouldn't be possible except by silently fetching the article and throwing it away. I had completely forgotten about the "Mark all pages visited" button -- given its existence, it seems it shouldn't be too hard to do that on a page-by-page basis. What think you? EEng (talk) 15:33, 13 July 2015 (UTC)

People like me fussed about 'bold' because we hated it...Smarkflea (talk) 00:16, 13 July 2015 (UTC)

> How can you not like bold? We've even got a policy encouraging it. EEng (talk) 00:43, 13 July 2015 (UTC)

A general comment - Wikipedia is not particularly good with accessibility for readers/editors with colour vision deficiency. This is at least partly to do with the volunteer nature of the project - it relies on editors taking into account an issue they are likely to be unfamiliar with and there is no co-ordinated editing approach to highlight issues where they arise. Nearly every chart/map that uses color to convey information creates an issue for color vision deficient readers. Some (most?) are just completely indecipherable for me - and my deficiency is quite mild. Map makers especially love to use red and green to convey "yes" and "no" - I just throw my hands up in the air! Not sure how the issue can be addressed but it is a real issue. -- Mattinbgn (talk) 01:38, 13 July 2015 (UTC)

> The default configuration of MediaWiki uses bold. Back in the day, it was turned off at en.wp (only) because of server load problems. We had a widely publicized CENT-listed RFC at the Village Pump with unanimous support from dozens of editors to change it back to the default... and the day the devs did exactly what the community asked for, solely because we asked for it, there were dozens of people saying that they hated it and starting another RFC to demand that the config change to be reversed immediately. The green dot was conceived of as a way to provide that signal to more people, without having the bold that some people hate.
> Someone like UserEdokter will know for sure, but I believe the gadget to provide the default configuration is turned on for new accounts, but not for existing accounts. So if your account is more than about five or six years old, then it would be off by default. Whatamidoing (talk) 03:42, 13 July 2015 (UTC)
>
> > Yes that seems like a fair representation of the status quo. I think everyone is ope to change this, we just need good ideas that will keep the watch list usable for everyone. P.S. Mattinbgn, if you ever need something addressed in accessibility of the website, am happy to assist you. I made several changes to improve the overall accessibility of the website over the years, I'm willing to assist in further developments. But as you stated accurately, the problem is mostly that in order for the content to be fully accessible, we basically need to turn every editor into an accessibility expert, which is unlikely to happen :) —TheDJ (talk • contribs) 11:22, 13 July 2015 (UTC)
> >
> > > The green arrows have always been default-on for everyone since its conception (except for the enhanced watchlist). You'de have to turn off the gadget explicitly to revert to the software default (bold). However, it the default were to be turned *off*, then only those who changed the setting would retain it. The rest would follow the default setting of the gadget. -- `{{User:Edokter}} {{talk}}` 17:35, 13 July 2015 (UTC)

Grayish bold

Right now Jacob Bigelow is showing up in my watchlist in this different color. Maybe if "you" (i.e. anyone interested) add it to your watchlist right now you'll see it too. EEng (talk) 04:52, 13 July 2015 (UTC)

Side question

Has anyone worked out a way to make it easier to spot single-character changes? Sometimes someone only removes a space, or changes as "," to a ",". While normally no one cares, this can make a big difference in a page full of source code, and tracking down such changes is maddeningly tedious. -- SMcCandlish ☺ ☎ ¢ ≽^•⩊•^≼ 06:46, 15 July 2015 (UTC)

> wikEdDiff, available via the gadgets page in your preferences. Regards, ORANGE SUEDE SOFA (talk) 06:49, 15 July 2015 (UTC)
>
> > @SMcCandlish: you can try also one tool from German colleague. It's quite cool, here is a screenshoot (http://www.bildites.lv/images/h5wq576twfwqugiet.png). --Edgars2007 (talk/contribs) 12:58, 15 July 2015 (UTC)
> >
> > > Thanks! Both of those would work. Kind of a power vs. learning curve trade-off. — SMcCandlish ☺ ☎ ¢ ≽^•⩊•^≼ 11:44, 16 July 2015 (UTC)

Replacing talk pages: Why not just adapt an exist webboard?

The last few years of discussion about and efforts toward producing a replacement for the talk page system has led it to mw:Flow, which doesn't seem to have a lot of buy in, and mw:Extension:LiquidThreads which garnered even less. There's also mw:Extension:MediaWiki Bulletin Board, which I have not examined yet, and various other things listed at mw:Category:Discussion and forum extensions.

Wouldn't it make more sense to just use one of the existing freeware Web discussion board (forum) packages that most of us are already familiar with? This would require doing one of the following:

1. Replacing its native bbcode parser with calls to the wikicode parser
2. Creating a plugin for it that calls the wikicode parser

With the latter approach, we could leave the bbcode features alone (other than disabling the feature to inline images from offsite), and just add to it a `[wiki] ... [/wiki]` feature, with

anything inside that being parsed as wikicode.

Obviously this would take a non-trivial amount of work, but from a how-to-approach-it standpoint, it makes more sense to me than the current wheel-reinvention approach. I would almost be surprised if this hasn't been done somewhere already (probably without the wikicode parsing) at some other MediaWiki installation, using apache (or whatever) URL rewriting to redirect requests for talk pages to forum threads. That part of the process seems comparatively trivial to implement.

A variant of this idea would be to have a preferences setting for using wiki talk pages or a "traditional" Web forum interface, so people can post either way, and scripts would translate between them on the fly during the save process. The principal problem with that would probably be that wiki talk page editors could still insert comments into threads arbitrarily, but webboard users would only be able to append them as new material (either new topics, or new replies). All convenience comes at a cost, I guess. Another bot might even be able to fix that, though, by examining indent levels, and "translating" interpolated talk page posts into message board replies at the right level.

Anyway, just thought I'd throw that out there. —SMcCandlish ☺ ☎ ✍ ≽^•⊖•^≼ 06:00, 15 July 2015 (UTC)

> A number of discussions on Wikipedia don't have exactly an one linear thread per forum format. I also wonder about security (the website I work on, TV Tropes) once had such a forum that turned out to be too unsafe against hacking attempts, thus we now use a homebrew code that is of similar appearance as Flow. That was years back, though) and integration with MediaWiki. Jo-Jo Eumerus (talk, contributions) 13:03, 15 July 2015 (UTC)
> There is no need to replace talkpages, and all alternatives that have been proposed so far are inferior. The Quixotic Potato (talk) 18:11, 15 July 2015 (UTC)
>
> Speaking from the technology side, I think you are underestimating the work required. I count:
>
> - Replacing the parser (HUGE)
> - Replacing the login system
> - Replacing the account management
> - Replacing the skin
> - Implement transclusions
> - Integrating a new notification system
> - Deal with integrating back and frontend caching between the two systems.
> - etc etc etc.
>
> You would have butchered up the original so much (and you would have to maintain that butchered variant and keep it in sync with the original) that there is little difference between that and starting from scratch. —TheDJ (talk • contribs) 09:33, 16 July 2015 (UTC)
>
> A lot of the good web boards are published under licenses incompatible with MediaWiki's license, and the ones that are compatible with its license are not that great. I agree with WMF's decision to implement a new discussion forum system, which looks very good at the moment yet is published under a compatible license. Gparyani (talk) 19:10, 16 July 2015 (UTC)

Too much dependence on volunteers

Sockpuppet

The following discussion has been closed. Please do not modify it.

Volunteers like Cyberpower TParis Hedonil music animal nakon, x! and others have a life of their own. Everyday some tools malfunction. Can't the wikimedia employ atleast five people to make these tools work 24/7. I agree with Kudpung that these important tools shouldn't be maintained by volunteers. And if any Wikimedia people reading this just think about it. Free encyclopedia should not become bugged encyclopedia.--Captain Doverman (talk) 05:30, 16 July 2015 (UTC)

> The last thing we need is fingerpointing. I would like to thank them: X!, Hedonil, TParis, Cyberpower678, MusikAnimal, et. al., for building, maintaining, and for picking up the pieces. xTools is currently at the mercy of its environment. Repeatability is an issue. Right now, we need a DevOps project to isolate the factors which are killing xTools. --Ancheta Wis (talk | contribs) 06:09, 16 July 2015 (UTC)

Thanks for twisting my words. I was supporting them. I understand their pressure. I was accusing Wikimedia for putting too much burden on them. We don't know their real names and their pictures. Technical 13 is banned. They will contribute and retire, we will know them with their username not the human behind that name. Yes i was finger pointing against Wikimedia foundation. So what?--Captain Doverman (talk) 07:02, 16 July 2015 (UTC)

> By definition, being a volunteer is voluntary. WMF is not a nation state, and can not impose duties on volunteers they have not themselves taken upon them. If you have concrete suggestions for how to retain valued volunteers, that is a different discussion.--Anders Feder (talk) 07:47, 16 July 2015 (UTC)
>
> > I think their argument was that many of the tools are so useful or important that the WMF should be providing some support towards their maintenance. Sam Walton (talk) 08:08, 16 July 2015 (UTC)
> >
> > > Thanks for understanding a bit. Captain Doverman (talk) 08:44, 16 July 2015 (UTC)
> > > As I said, what support are you suggesting?--Anders Feder (talk) 08:26, 16 July 2015 (UTC)

Yes i know about voluntary. These are very important tools and WMF shouldn't just sit and relax putting the entire burden on volunteers. Article creation is different, tool maintenance is different. Even a 12 year old girl can understand this, which some experienced wikipedians failed to understand. What next--- Wikipedia security related issues will also be thrust upon volunteers? There must be some mechanism in selecting them. As check user identity is verified. **The tool creators must submit their professional and educational qualifications to the WMF. There must interview to check their capability. I don't have to explain that only software engineers and IT professionals or computer science students maintain tools.** History students, arts students can create articles, but they don't do these things. Captain Doverman (talk) 08:44, 16 July 2015 (UTC)

> And what are you offering the volunteers in return for all this bureaucracy you want to impose on them?--Anders Feder (talk) 08:55, 16 July 2015 (UTC)

I understand what you are trying to say. But... you have to consider that 'more stuff', means more knowledge and worse, when it comes to volunteer 'experiments', knowledge that cannot easily be institutionalized. The trick to efficiently scaling up is that you create a 'common' base of knowledge, that you can reuse across multiple projects. With the volunteer projects, this is hardly possible, because they are often 'improvised hacks' that 'do their own thing'. This is actually what makes them powerful, because people can divert from the set out conventions and find wonderful new approaches. Now WMF is already spending a lot of time building infrastructure to sustain this internally and through API and labs externally, but I do think that WMF should for instance, give more workshops to volunteers to build up knowledge about how to scale their own work. They need to encourage people to do more things like the shared static resources (http://tools.wmflabs.org/static-browser/) offering between tools. To help them with source code management and building development teams

around tools. To help them improve bug reporting and documentation etc. Because that is what will make things scale the next five years and create fewer problems long term for everyone. The problem is just that we have a few critical tools that are 10 years old that will have to find a way to catch up with that. But just tossing five expensive resources (that would have to be hired from where exactly... ???) will not fix the problem with these tools any time soon either and worse if they did work on this, we would have the exact same problems five years into the future. —TheDJ (talk • contribs) 09:08, 16 July 2015 (UTC)

> @ Anders Feder, which part of my sentence"I am not accusing the volunteers"you don't understand. I am accusing WMF for putting some volunteers to develop these important tools, which they are supposed to develop. Don't they have any responsibility?. Is it so difficult to hire only five software engineers and computer/IT professionals. Wikipedia is a free encyclopedia. Cyberpower, music animal, they have their own professional/student life. Whenever tools don't work, they come under pressure from all quarters. As volunteers, we are putting too much pressure on them. What do they get in return from WMF ?Captain Doverman (talk) 09:33, 16 July 2015 (UTC)
>
> > WMF haven't "put some volunteers" to anything. The volunteers have chosen to make the tools on their own initiative in order to solve a problem they have themselves defined.--Anders Feder (talk) 10:22, 16 July 2015 (UTC)
> >
> > > @Captain Doverman:, I think with a few more tools, it should be possible to debug where the problem sits in the code, which is visible on github. Hedonil and Cyberpower678 put the X! code up on github for anyone to see. But Cyberpower678 has observed that his test cases are not getting executed. I propose we use xdebug extension for PHP (http://xdebug.org/docs/install) to isolate where the thread is getting stepped on. But I believe this is a DevOps situation, which is bigger that the xTools (https://tools.wmflabs.org/xtools-articleinfo/index.php?) code itself. So we need cooperation from others, such as the Redis specialists (in WmF?). --Ancheta Wis (talk | contribs) 13:09, 16 July 2015 (UTC)
> > >
> > > > @Captain Doverman: The Foundation is currently assembling a team to work on tools for the community. Stay tuned. Ryan Kaldari (WMF) (talk) 15:04, 16 July 2015 (UTC)

Can admins get paid next? Chillum 15:08, 16 July 2015 (UTC)

> With all due respect, I think I'd prefer the present set-up to an organised WMF team... If the OP is still watching this thread (but he's currently unable to contribute to it, so he might not be), nobody here gets 'assigned' to tasks. I came to Wikipedia to remove some rubbish, and I'm still doing that. Others have technical knowledge and skills that I don't - they create tools and keep the basic framework of the place working. There are paid WMF people for positions that volunteers couldn't really be expected to handle - maintaining servers, dealing with legal matters and so on. Peridon (talk) 15:40, 16 July 2015 (UTC)

Format price problem

At H. H. Asquith#Asquith's final years and death the Template:Format price produces "about £500 thousand today using CPI". It should of course say "about £500,000 ...". I do not know how to fix this. Can anyone help please? DuncanHill (talk) 15:13, 16 July 2015 (UTC)

> That's what {{Format price}} does; if you want to keep the digits, use `|fmt=c` with {{Inflation}}. Alakzi (talk) 15:27, 16 July 2015 (UTC)
> (edit conflict) Saying "500 thousand" in £520 thousand is the whole point of {{Format price}}. If you don't want it then don't use the template. {{Inflation}} can insert a comma with `|fmt=c`: £520,000. PrimeHunter (talk) 15:30, 16 July 2015 (UTC)
>
> > Why would anyone want to say "£500 thousand"? Anyway, @Alakzi: has kindly fixed it, for which my thanks. DuncanHill (talk) 15:41, 16 July 2015 (UTC)

Do you have ideas on how to improve Wikipedia?

The WMF is currently assembling a team to work on tools for the community. Please post your ideas on this page: meta:Community_Tech_project_ideas. The Quixotic Potato (talk) 17:33, 16 July 2015 (UTC)

Template using white text even when not specified

I made a template in this sandbox (https://en.wikipedia.org/w/index.php?title=Template:Xi&oldid=671741240), and I'm wondering, why is it using white text even when the parameter is not specified? nyuszika7h (talk) 18:08, 16 July 2015 (UTC)

> @Nyuszika7H: That looks fine to me. What are you seeing that's wrong? Jackmcbarn (talk) 18:29, 16 July 2015 (UTC)
>
> > @Jackmcbarn: The permalink didn't work because the transclusion uses the current version anyway. Alakzi fixed it in the meantime (thanks!) nyuszika7h (talk) 18:32, 16 July 2015 (UTC)

Encoded identifier in article to serve searches?

In chemistry, there is an identifier named InChI that is an encoding for the full (3D) molecule structure. For example acetate has `InChI=1S/C2H4O2/c1-2(3)4/h1H3,(H,3,4)/p-1`. My question is: how should we best add this code to the article? It is illegible, but it is an ID for sure.

Notes:

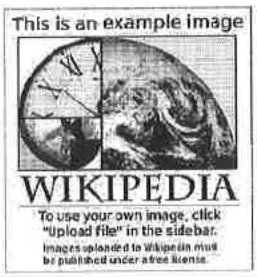
- Publishing the InChI code in an article will help outside searches.
- Per chemical substance, there exists just one *normalised* code. However, multiple non-normal codes can exist in parallel. For these, the same question exists: add to article because of external search?
- InChI code can be long, like 500 characters. (It is also hashed into a shorter Key, which is not unique. Let's forget for now).
- Today, the illegible code is folded in a collapsed "hide" box. However, as we know in mobile view it is not hidden.

I don't expect a conclusive answer here, because it is multi-faceted (readers view & ID & search). But can people give me hints on where to read & learn & detail this? Any earlier wisdom? -DePiep (talk) 20:40, 16 July 2015 (UTC)

Bullets appear in the wrong place

In Internet Explorer 11, bullets always appear on the left of the page, without regard to whether or not an image is embedded on the left side. See File:Misplaced bullets.PNG. Gparyani (talk) 22:34, 15 July 2015 (UTC)

Reproducing the problem below



An image

- A bullet
- Another bullet
- A third bullet
- Some more bullets...
- One final bullet

@Gparyani: AFAIK this is an Internet Explorer-problem, and not fixable on Wikipedia. The Quixotic Potato (talk) 23:31, 15 July 2015 (UTC)

Looks fine in IE8. --Redrose64 (talk) 23:38, 15 July 2015 (UTC)

True, but in IE9 it looks like this (http://i.imgur.com/T2WznFN.jpg). The Quixotic Potato (talk) 23:45, 15 July 2015 (UTC)

Seems like a regression. Can you, or someone else, please file a Phabricator ticket? They usually accept browser-specific issues. Gparyani (talk) 03:14, 16 July 2015 (UTC)

I don't think that it is useful to file a ticket on the Phabricator because I believe that this problem is related to Internet Explorer, and that it is not fixable on Wikipedia. AFAIK the only (good) solution is to change how Internet Explorer works. AFAIK there is no fix for this problem that could be implemented on Wikipedia. The Quixotic Potato (talk) 03:27, 16 July 2015 (UTC)

@The Quixotic Potato: If it can't be fixed, it can at least be worked around. Gparyani (talk) 19:07, 16 July 2015 (UTC)

The raw HTML doesn't look too bad:

```
<div class="thumb tleft">
  <div class="thumbinner" style="width:222px;">
    <a href="/wiki/File:Example.jpg" class="image">
      <img alt="" src="//upload.wikimedia.org/wikipedia/en/thumb/a/a9/Example.jpg/220px-Example.jpg" width="220" height="238" class="thumbimage"
srcset="//upload.wikimedia.org/wikipedia/en/a/a9/Example.jpg 1.5x,  //upload.wikimedia.org/wikipedia/en/a/a9/Example.jpg 2x" data-file-width="275" data-file-
height="297" />
    </a>
    <div class="thumbcaption">
      <div class="magnify">
      <a href="/wiki/File:Example.jpg" class="internal" title="Enlarge">
      </a>
    </div>
    An image
  </div>
</div>
</div>
  <ul>
  <li>A bullet</li>
  <li>Another bullet</li>
  <li>A third bullet</li>
  <li>Some more bullets...</li>
  <li>One final bullet</li>
  </ul>
  <p>
    <br clear="all" />
  </p>
</dl>
```

The issue may be the CSS associated with those divs being interpreted differently by different browser versions. (easy enough to check: turn off CSS) but I suspect the issue is because Wikipedia uses the old style `<br clear="all" />` instead of the more modern `<br style="clear:both;" />` or the external CSS equivalent. --Guy Macon (talk) 03:30, 16 July 2015 (UTC)

@Guy Macon: If you modify the code in Developer Tools to use the new-style code, does it work? Gparyani (talk) 03:33, 16 July 2015 (UTC)

Float:left on the <li> seems to work, but that still isn't perfect. A better solution is to set the display style of the UL element to table. Please don't ask me why, but that seems to work. Or <ul style="list-style-position:inside;">. I hate Internet Explorer. The Quixotic Potato (talk) 03:43, 16 July 2015 (UTC)

This is a bug in IE9. However, to some degree this is a problem with all lists next to floating content. It's just aggravated by the bug. We have {{Flowlist}} to deal with specially this combination of content, a list next to a left floating element. I suspect it will also fix the IE9 issue. Please do try it out. —TheDJ (talk • contribs) 09:17, 16 July 2015 (UTC)

Fwiw... to rectify this behavior (under at least IE 11), the UL or OL needs to be set to *display: inline-block;* rather than the default, *display:block;*.

An image

- A bullet
- Another bullet
- A third bullet
- Some more bullets...
- One final bullet

Since wiki markup is being used I'm guessing wrapping the list in a DIV set to the same *display:inline-block;* should work as well. -- George Orwell III (talk) 20:56, 16 July 2015 (UTC)

@George Orwell III: It displays just fine, except it's a little bit to the right from where it should be. Can you file a ticket in the Phabricator and have this fixed? Gparyani (talk) 22:56, 16 July 2015 (UTC)

@Gparyani:. The amount of distance between the image's right edge and the bullet marker items is due to the [re]use of the default UL *margin:* css settings because the assumption has always been list-items would be the left-most content in any given block of text being applied. There is not much that can be done about that except for dropping the use of wiki-markup for such lists and then overriding the defaults by adding your own modified inline css styling(s) for the opening UL tag.

As for filing a request to "correct this" across all the wiki-projects; that's not likely to be accepted -- at least not without thorough vetting against all browser versions and multiple usage scenarios that is. Since the behavior seems to be specific to IE 9 and higher, I'm betting the response to asking for such a change would be to apply templates such as {{Flowlist}} or similar workarounds (like my DIV wrapper one) on a case-by-case basis. I'm afraid I'd have to agree with such a position; there are just too many variables and possible instances of usage-type in play to craft a universal solution given the fact *float:* and DIVs are used for [thumbnail] image alignment to begin with. -- George Orwell III (talk) 23:22, 16 July 2015 (UTC)

I can confirm that this would likely be the result. Flowlist itself is the result of such a bug report. This is just one of those situations, where HTML has trouble delivering what we want it to do. —TheDJ (talk • contribs) 14:37, 17 July 2015 (UTC)

The title creation blacklist had a minor glitch in functionality...

Just wanted to point out that the software that enforces the title creation blacklist had a malfunction about five minutes ago. I cannot replicate the issue at the present time, given that the blacklist enforcement software is functioning again. (I know that there was a bug filed for this and resolved in the past, but I cannot find it right now.) *Steel1943* (talk) 21:02, 16 July 2015 (UTC)

For future reference, please describe the actual glitch that you were facing. Thanks! :) --AKlapper (WMF) (talk) 11:36, 17 July 2015 (UTC)

apostrophe problem

Only lately, I've noticed a few pages not loading for me in Opera12, though they load in new Opera and Firefox. At first it was the 2015 FIFA Women's World Cup, but 2011 & 2019 worked. Now I notice they don't, and I've also found John O'Brien (priest) doesn't work either. I suspect now it has something to do with the apostrophe, but replacing it with %27 doesn't work either. New Opera uses that and works, FF uses apostrophe and works.

I've noticed this is called a bug, but I didn't see how to make it work with my browser. Overall, I guess it's not essential, but if you know any fixes I'd like to know. Thanks.. Smarkflea (talk) 22:25, 16 July 2015 (UTC)

2011 FIFA Women's World Cup, 2015 FIFA Women's World Cup, 2019 FIFA Women's World Cup and John O'Brien (priest) all work for me in Opera 12.17 on Windows Vista. What happens when you click the links here? Does the same happen for https://en.wikipedia.org/wiki/John_O'Brien_(priest) and https://en.wikipedia.org/wiki/John_O%27Brien_(priest)? Does it also happen when you are logged out? PrimeHunter (talk) 01:10, 17 July 2015 (UTC)

I tried all those thing, incl. logging out, and I get the same thing (Win8.1). Only the address bar loads; nothing at all on the screen but white.Smarkflea (talk) 01:59, 17 July 2015 (UTC)

Pages moved without leaving redirects

Why do pages that are moved without leaving redirects have a red deletion notice? GeoffreyT2000 (talk) 17:18, 17 July 2015 (UTC)

For example, User:Username~enwiki/draft. I think it's because there's a log entry, so the software knows that a page *was* at that title in the past, although it isn't there now, so it assumes a deletion. --Redrose64 (talk) 18:51, 17 July 2015 (UTC)
Because moving w/o redirect is technically copying an article to a different title and deleting the original? Is there anything undesirable about the notice?--Anders Feder (talk) 02:21, 18 July 2015 (UTC)

"helpme-helped" makes text small

At User talk:Exoplanet Expert#References are messed up:, why has the addition of {{helpme-helped}} made the text small? This does not show on "Show preview", only after the edit is saved. JohnCD (talk) 17:31, 17 July 2015 (UTC)

There's an unclosed <small> tag in the preceding section. Alakzi (talk) 17:36, 17 July 2015 (UTC)

Thanks, that explains it. I should have spotted that. It was in the user's signature, but they seem to have gone back to a standard signature now. Cheers, JohnCD (talk) 17:47, 17 July 2015 (UTC)

Protecting an entire article against automatic spell-miscorrection

(Dthomsen8 has asked me to address this problem because I'm a linguist.)

Is there a straightforward way to protect an entire article from automiscorrection? AWB tries to "correct" some words in Catalan verbs to English words that they resemble. How can this be prevented without {{sic}}cing every Catalan word in the article, of which there are a great many? Please {{Ping}} me to discuss. --Thnidu (talk) 23:48, 14 July 2015 (UTC)

@Thnidu: A whole article shouldn't be protected from spelling corrections. Foreign words and text should be marked for this and other reasons. See Wikipedia:Manual of Style/Accessibility#Other languages and Template:Lang/doc#Rationale. PrimeHunter (talk) 23:56, 14 July 2015 (UTC)
@Thnidu: If you use the {{sic}} template on a page then AWB will display the following warning in a messagebox: "This page contains a 'sic' tag or template, please take extra care when correcting typos". Many AWB users will simply skip the article after receiving that warning. If you do not want to show it, you can use "{{sic|hide=y}}". I don't know if AWB knows that it should ignore anything inside the template {{lang}} if the language is different than the language of the typo-regexps. The Quixotic Potato (talk) 00:05, 15 July 2015 (UTC)

What if you replaced a character in the word with a Unicode equivalent that AWB won't recognize? **bd2412 T** 00:28, 15 July 2015 (UTC)

@BD2412: That would almost definitely mean that the word wouldn't be read out properly for screen reader users such as myself. Graham87 10:50, 15 July 2015 (UTC)

It might be interesting to have AWB be changed so that it also pops a warning when {{lang}} is used. Might submit a phab for that regardless and then the AWB devs can tell us whether it works like that already. --Izno (talk) 00:52, 15 July 2015 (UTC)

People who blindly correct spelling with AWB - changing dozens of instances of "movie" to "movi.e.", for example - shouldn't have access to it at all. There's been a very strong consensus to reject unsupervised spelling bots since forever precisely because of this sort of problem. —Cryptic 00:52, 15 July 2015 (UTC)

@PrimeHunter: My point is that it is *massively* impractical to {{sic}}-tag every non-English word in a sizeable article *about another language*, and it is clearly unreasonable to expect an editor to do so. The Quixotic Potato and Izno have made useful suggestions, and Cryptic has made a very good point and mentioned a consensus or policy that I was not aware of. --Thnidu (talk) 01:26, 15 July 2015 (UTC)

In general, AWB does not change anything within a template such as {{lang}}. It never does so for its several thousand typo rules. An AWB user can force the program to change text within a template for rules that they have themselves provided, but they should not do so unless they really know what they're doing. The problem is not within the program. No editor should save any change if they do not understand why that change is being suggested. If any user (with or without AWB) makes repeated erroneous edits, you should tap them on the shoulder and inform them of the disruptions, and then undo the damage. If they persist, have them banned from editing. If anyone knows the details of 'dozens of instances of "movie" to "movi.e."', please provide those details here or on AWB's talk page so we can deal with that situation; my guess is that an unreliable editor has made up a rotten rule. But please don't post vague complaints here in an effort to place restrictions on the use of AWB. **Chris the speller** yack 02:30, 15 July 2015 (UTC)

Other than here the only example of movi.e was here (https://en.wikipedia.org/w/index.php?title=Royal_Space_Force:_The_Wings_of_Honn%C3%AAamise&diff=next&oldid=671051736) and the IP before me also corrected the same error in a different part of the page. CambridgeBayWeather, Uqaqtuq (talk), Sunasuttuq 02:46, 15 July 2015 (UTC)
@Chris the speller: I am pretty sure no one has posted "*vague complaints here in an effort to place restrictions on the use of AWB*". The Quixotic Potato (talk) 02:47, 15 July 2015 (UTC)

@Izno:@Thnidu: I did some experimenting, and AWB does ignore stuff that is inside a {{lang}} template, even when I claim that the language is English. It also ignores stuff inside the templates {{sic}}, {{typo}}, {{notatypo}}, {{As written}} & {{Proper name}}. But for some reason I couldn't get AWB to show me a messagebox with the warning I mentioned earlier... I do not understand why. I am certain this used to work with older versions of AWB, here is the code (from main.cs):

```
        // check for {{sic}} tags etc. when doing typo fixes and not in pre-parse mode
        if (chkRegExTypo.Checked && !preParseModeToolStripMenuItem.Checked && TheArticle.HasSicTag)
            MessageBox.Show(@"This page contains a 'sic' tag or template, please take extra care when correcting typos.", "'sic' tag in page", MessageBoxButtons.OK,
MessageBoxIcon.Warning);
```

It is easy to find the message in older versions of the source code (if you search for "take extra" you can find it here (https://github.com/svn2github/autowikibrowser/blob/master/tags/REL_4_5_2/AWB/Main.cs) for example. But for some reason I am unable to find "take extra" inside the most recent version of the same file (http://sourceforge.net/p/autowikibrowser/code/HEAD/tree/AWB/AWB/Main.cs).
I do not understand why, but I think that this messagebox-warning has been removed for some reason.

The Quixotic Potato (talk) 02:47, 15 July 2015 (UTC)

Speculating, you can silently ignore text inside T:Lang without worry and any langs in general because text is always inside lang (and presumably that's the text that shouldn't be changed), but you can't do the same since sic doesn't always contain text within, which is why you need the FYI to the user. --Izno (talk) 02:51, 15 July 2015 (UTC)

@The Quixotic Potato: Oh brother! I'm taking this to Phabricator. Thanks very much for testing this out. --Thnidu (talk) 02:53, 15 July 2015 (UTC)

- I believe {{bots|deny=AWB}} will do what the OP wants. EEng (talk) 02:54, 15 July 2015 (UTC)

EEng is correct, I just tested it, {{bots|deny=AWB}} works. I think adding {{sic|hide=y}} isn't a bad idea, because it does show a warning in AWB, but the attention-grabbing messagebox is gone. The Quixotic Potato (talk) 03:00, 15 July 2015 (UTC)

"How can this be prevented without [*sic*]cing every Catalan word in the article, of which there are a great many?" Simple: put a {{lang}} template around only the words or phrases that draw AWB's attention. I have done this, and it only took a few minutes. The next time you have a question about AWB, try the talk page for AWB. The editors there are very knowledgeable, helpful and friendly. You'll get help a resolution faster, and it won't look like you're trying to sneak up and

gang up on the AWB users. Is it too soon to put a {{Resolved}} template on this section? **Chris the speller** [yack] 02:58, 15 July 2015 (UTC)

> It doesn't look like he is "trying to sneak up and gang up on the AWB users". Maybe you've misread something. I am an AWB user. The Quixotic Potato (talk) 03:02, 15 July 2015 (UTC)
>
> > The section starts out with an attempt to stop AWB from ever touching the article, not asking how to prevent AWB from damaging the article. Yes, I may have misconstrued Cryptic's purpose for mentioning "movi.e.". But I stand by my statement that the AWB talk page would be a better forum. **Chris the speller** [yack] 03:23, 15 July 2015 (UTC)

- @Thnidu: Regarding the Catalan verbs article, I've fixed several English missoellings and added/fixed some italics to protect some Catalan text. Chris the speller added a few {{lang}} templates to protect the rest of the Catalan text. The net effect of the changes can be seen here (https://en.wikipedia.org/w/index.php?title=Catalan_verbs&diff=671499472&oldid=669566483). Could you please add the missing end parenthesis in the last paragraph of the Catalan verbs#2nd conjugation (-re, -er) verbs section?
- @EEng: IMHO, it's better to fix the underlying problems than to protect an incorrect article. Thanks! GoingBatty (talk) 03:02, 15 July 2015 (UTC)

> I wasn't taking a position on what the best approach is, just pointing out a technical fact. But since you bring it up, I personally have been vexed many times over many years by mindless "fixes" wrought by script-kiddies in a hurry to make themselves feel they're helping WP by changing careful writing and markup to lower-functioning vanilla that looks like what they've seen in other articles, and which they therefore have concluded must be the only acceptable way to do things. EEng (talk) 03:12, 15 July 2015 (UTC)
> @Thnidu and Chris the speller: To clarify my post above, please note that AWB also doesn't fix the spelling of any text inside of quotation marks or italics. (I've just clarified this on Wikipedia:AutoWikiBrowser/Typos#AutoWikiBrowser (AWB).) By adding italics for the book titles in the Catalan verbs#Bibliography section, the {{lang}} templates were not necessary there. Not only did fixing another set of italics earlier in the page made another {{lang}} template redundant, but it then allowed AWB to identify some real English typos, which I then corrected. Hope this helps! GoingBatty (talk) 03:23, 15 July 2015 (UTC)

- @Cryptic and CambridgeBayWeather: Looking carefully at the edit summary for this edit (https://en.wikipedia.org/w/index.php?title=Royal_Space_Force:_The_Wings_of_Honn%C3%AAamise&diff=668048190&oldid=665072683) to Royal Space Force: The Wings of Honnêamise, it appears that the incorrect "ie → i.e." fix was the user's manual find & replace rule, not one of the defined typo fixes. I've reverted the remaining incorrect replacements from that edit. I'm glad Cambridge already reached out to the user who made the incorrect edit. GoingBatty (talk) 03:38, 15 July 2015 (UTC)

> I tried but couldn't duplicate changing movie to movi.e. When I saw your edit summary (https://en.wikipedia.org/w/index.php?title=Royal_Space_Force%3A_The_Wings_of_Honn%C3%AAamise&type=revision&diff=671501606&oldid=671498215) I thought that I must have made the wrong change and that movi.e. was some special term used in anime. CambridgeBayWeather, Uqaqtuq (talk), Sunasuttuq 03:44, 15 July 2015 (UTC)
>
> > @CambridgeBayWeather: Sorry my edit summary wasn't clear. The editor who accidentally changed "movie" to "movi.e.," four times made a mistake. Your action to change "movi.e.," to "movie.," was also an accidental mistake. In my edit summary, I meant to indicate that I was fixing all of these accidental (but good faith) changes. GoingBatty (talk) 03:54, 15 July 2015 (UTC)

- @Thnidu:: Proper use of `{{lang}}` is how to fix this. Yes, it's some amount of work to do it, but it should be done anyway, even aside from AWB issues, and is the semantically correct way to do it; putting `{{Sic}}` around all the Catalan content is the same amount of work, but a worse-than-pointless approach. In the interim, as someone else pointed out, even including a single `{{sic}}` will send up an AWB red-flag. You could thus use it at the bottom of the page thusly: `{{sic|hide=y|reason=Do not remove this. It is a signal to AWB users that this page contains a lot of non-English text that can produce false positives in AWB's automated spell-checking routines.|<!--Intentionally blank content.-->}}`, ~~or better yet, put that all in a template, e.g. `{{Do not auto-correct spelling with AWB}}`.~~ PS: Using `{{bots|deny=AWB}}` is also a wrong-headed approach, because AWB does many things besides spellchecking. Abuse of that tag is a minor form of WP:OWN and interferes with others' editorial rights, albeit not in a huge way. More importantly, it interferes with the ability to include Catalan-related articles in legitimate cleanup efforts. —SMcCandlish ☺ ☎ ¢ ≽^ტ_ტ^≼ 06:08, 15 July 2015 (UTC) Struck out my brain-fart. —SMcCandlish ☺ ☎ ¢ ≽^ტ_ტ^≼ 07:04, 15 July 2015 (UTC)
  - @SMcCandlish: A template such as `{{Do not auto-correct spelling with AWB}}` wouldn't work unless AWB was re-coded to recognise it. When deciding whether to display "Contains 'sic' tag" in its "Alerts" box, AWB only looks at the article text, not inside any templates it uses. -- John of Reading (talk) 06:46, 15 July 2015 (UTC)
    - D'oh! I need coffee. Oh well, then just putting an empty, hidden `{{sic}}` at page bottom should do it. My main concern was people abusing `{{sic}}` to "protect" non-English words instead properly using the same amount of work to identify non-English text with `{{lang}}`. —SMcCandlish ☺ ☎ ¢ ≽^ტ_ტ^≼ 07:03, 15 July 2015 (UTC)

> The {{sic}} template is very very rare, and the amount of typos approaches infinity. The sic template is used in ~12.500 locations (including outside the article namespace) and there are 4,917,656 articles in total. The Quixotic Potato (talk) 09:57, 15 July 2015 (UTC)
>
> > How does that relate? —SMcCandlish ☺ ☎ ¢ ≽^ტ_ტ^≼ 22:14, 15 July 2015 (UTC)
> >
> > > I tried to point out that if there are people abusing `{{sic}}` to "protect" non-English words instead of identifying non-English text with `{{lang}}` then that must be very rare, because the sic template is not frequently used. The Quixotic Potato (talk) 23:58, 15 July 2015 (UTC)

- Essentially when doing a spelling correcting run with AWB one has two options available when encountering a false positive:
  1. Skip the spelling/article (both easy)
  2. Tag the correctly-spelled-in-this-context word with a Lang, Sic, Typos, Not a typo, or Proper noun templates.

  Anyone who does the first will run into the same false positives on the next run, so it is I imagine pretty normal to do the second (I certainly did). Neither should cause problems to other editors.
  All the best: *Rich Farmbrough*, 18:15, 17 July 2015 (UTC).

@**Dthomsen8**: Does this discussion help? --Thnidu (talk) 19:42, 19 July 2015 (UTC)

Proposed software change: Show the reference list when section editing

I often find little errors in references, and usually hit the edit section button to correct these. The problem is, I have no way to preview them when I do this. I suggest that the reflist be automatically added below the edit summary window if the edited area doesn't already have one. This would allow me to preview my reference corrections, instead of having to do another edit to correct it again. Qiyabepsy (talk) 04:54, 17 July 2015 (UTC)

- **Support** - For the reason given, I usually avoid section edit for this purpose. The problems with that: Sometimes a little harder to find the spot I want to edit (although browser Find usually gets me there with a good choice of search text), increased chance of edit conflict, and edit summary doesn't show the section name. An alternative solution would be to make the ref tooltips work in section edit preview (or do the two go hand-in-hand?). Shouldn't this be at WP:VPR? —*Mandruss* ☎ 06:55, 17 July 2015 (UTC)
- Another possibility could be to enable the existing automatic reflists in preview mode. SiBr4 (talk) 07:44, 17 July 2015 (UTC)
  - I had this on my idea list as well. I'm not sure if it's entirely possible. The biggest problem however is that you are guaranteed to have big red errors as soon as you have named and reused references. It's a bit jarring to novices.. —TheDJ (talk • contribs) 13:36, 17 July 2015 (UTC)

Anomie's Ajax Preview script (https://en.wikipedia.org/wiki/User:Anomie/ajaxpreview.js) adds this functionality.(Details (https://en.wikipedia.org/w/index.php?title=User_talk:Anomie/ajaxpreview.js&diff=prev&oldid=300980242&oldid=300954335)) - NQ (talk) 08:28, 17 July 2015 (UTC)

- This has been tried in a hackish way via Template:Reflisto. A not-unworkable design would be something like this: as a "last step" in a section preview, do the equivalent of {reflist|group=foo} for all nonempty groups. These would not necessarily show the refs the same as they'd be seen in a full-article preview (# columns, etc.) but would be enough e.g. to proofread citation templates. "Referenced but not defined" errors could simply be surpressed -- or if that's hard just leave them. If all this comes at the very end of the page those who don't find it useful could ignore it all. EEng (talk) 22:50, 18 July 2015 (UTC)
  - I used to be annoyed by this and was delighted to discover Template:reflistp. Now I use it routinely. --Thnidu (talk) 19:37, 19 July 2015 (UTC)

Subdividing long lists alphabetically

I've been doing some work on List of people from Pennsylvania, which is quite lengthy and is subdivided by, I guess, reason for notability: Actors, Artists, Athletes... Recently IP user 173.52.75.38 broke up the four or five longest sections into alphabetical chunks (diff (https://en.wikipedia.org/w/index.php?title=List_of_people_from_Pennsylvania&diff=672054577&oldid=671657958)), which I thought was an excellent idea (their talk page). To reduce the need for scrolling still further, I added subTOCs— see for example § Athletes there— but I used anchors instead of subTOC wikicode because there's more than one "A–B" sub§, etc. I think this could be very useful for many long lists.

I wrote a Perl script, User:Thnidu/anchor-alphas, to semiautomate the process, but the output still has to be copy-pasted into the page. I've done that already for List of people from Pennsylvania. Can anyone automate that step and create a generally more useful wikitool?... --Thnidu (talk) 19:18, 19 July 2015 (UTC)

CSS required to prevent layout from breaking

This issue began at Wikipedia:Help desk#"I Have a Dream" formatting issue. `<blockquote>...</blockquote>` text was overrunning a `{{Listen}}` box in Martin Luther King, Jr., seen in this screenshot (https://phabricator.wikimedia.org/F193411). The solution turned out to be the addition of the inline `overflow: initial;` rule. This was done in this edit (https://en.wikipedia.org/w/index.php?title=Martin_Luther_King,_Jr.&diff=671735206&oldid=671711125) (while incidentally converting `<blockquote>...</blockquote>` to `{{Quote}}`).

My understanding from that HD thread is that this was made necessary by a change to MediaWiki:Common.css.

We sometimes see CSS used in wikitext to *enhance* layout, but this is the first time I've seen it required to prevent layout from breaking. Basically we're saying that average editors will have to know this workaround solution just to make basic layout work; to the extent they do not, there will be ongoing related issues at Help Desk, Teahouse, etc. Is this the best we can do?

Pinging the two experts from the HD thread, TheDJ and Alakzi, in case they would care to weigh in. —*Mandruss* ☎ 11:48, 17 July 2015 (UTC)

> The offending declaration block should simply be removed. I don't see why anybody would need to change the background of block quotations, and, even if they had to, they'd be wise to do so *systematically*, using a template. Alakzi (talk) 11:57, 17 July 2015 (UTC)
> See Wikipedia:Village pump (technical)/Archive 129#WikiWand, images and blockquotes for context. Alakzi (talk) 12:09, 17 July 2015 (UTC)
> Edokter has removed the CSS (https://en.wikipedia.org/w/index.php?title=Martin_Luther_King,_Jr.&diff=671877848&oldid=671735206) as "not supported by IE" and moved the `{{Listen}}` box down to avoid the `{{Quote}}`. This is no better for editors, and, as in this example, will sometimes force poor layout. —*Mandruss* ☎ 01:36, 18 July 2015 (UTC)
>
> > If you need to reset the overflow property, use `inherit` instead, as `initial` is ignored by IE. But it has been pointed out that we sometimes expect behaviour from HTML that is not always possible. In such cases, consider alternative layout; don't force the layout you want using CSS hacks. That tends to backfire. — `[[User:Edokter]] {{talk}}` 07:34, 18 July 2015 (UTC)
> >
> > > The CSS hack here is the blanket overflow in Common.css. Floats work fine without it. Alakzi (talk) 08:17, 18 July 2015 (UTC)
> > >
> > > > No they don't. It was placed there to prevent overlapping. — `[[User:Edokter]] {{talk}}` 08:48, 18 July 2015 (UTC)
> > > >
> > > > > Please identify what it is with the current version of the page that's not working. Alakzi (talk) 08:55, 18 July 2015 (UTC)
> > > > >
> > > > > > The reason it was placed in Common.css: "*Avoid collision of background with floating elements*". Note that this mainly applies to templated uses of `<blockquote>`. — `[[User:Edokter]] {{talk}}` 09:18, 18 July 2015 (UTC)
> > > > > >
> > > > > > > So there's nothing wrong with it. If it mainly applies to templates, then it should be placed in templates. Alakzi (talk) 09:23, 18 July 2015 (UTC)
> > > > > > >
> > > > > > > > What's wrong is that it's there. I agree the blanket declaration for `<blockquote>` is misplaced; it should be a template class. But let's not hide that by applying these hackish work-arounds.
> > > > > > > > — `[[User:Edokter]] {{talk}}` 09:42, 18 July 2015 (UTC)
> > > > > > > >
> > > > > > > > > Well, it is a temporary measure; I would not support applying the same fix/hack everywhere. Alakzi (talk) 09:46, 18 July 2015 (UTC)

Would the Common.css change have been made if it had been known it would introduce this issue? —*Mandruss* ☎ 10:11, 18 July 2015 (UTC)

Well, it is a temporary measure Alakzi, if this is temporary, where is the permanent? Is this something to be filed away in the back of one or two people's minds, for possible later attention as time permits? I again ask people to look at this from the average editor's perspective, and I'm still interested in an answer to my preceding question, which was not rhetorical. —*Mandruss* ☎ 02:52, 20 July 2015 (UTC)

You're preaching to the choir. I can't answer your question; I wasn't the one who made that decision. It will be fixed permanently if and when it pleases His Majesty Edokter, Protector of All Stylesheets. Alakzi (talk) 09:17, 20 July 2015 (UTC)

Various tools are down

It's known here that Article History tool hasn't been working for quite a while now (around a month). But now, things are getting worse: the Sigma tools (such as this (https://tools.wmflabs.org/sigma/usersearch.py) and [18] (https://tools.wmflabs.org/sigma/summary.py)) are now displaying a 500 error when you try to use them. Now what? Narutolovehinata5 tccsdnew 02:21, 18 July 2015 (UTC)

I can't reproduce such an error on either of those links. Have they been resolved on your end too?--Anders Feder (talk) 02:25, 18 July 2015 (UTC)

(edit conflict) My tools were briefly down for maintenance. As for WikiHistory, a replacement from the German Wikipedia is in the works, as far as I know. —Σος. (Sigma) 02:27, 18 July 2015 (UTC)

(edit conflict) @Anders Feder: I'm still getting a 500 error. Here's a direct link (https://tools.wmflabs.org/sigma/usersearch.py?name=Narutolovehinata5&page=Mami+Kawada&server=&max=) at an attempt to use. Narutolovehinata5 tccsdnew 02:30, 18 July 2015 (UTC)

You didn't specify that the server was enwiki. Try https://tools.wmflabs.org/sigma/usersearch.py?name=Narutolovehinata5&page=Mami+Kawada&server=enwiki&max= instead. —Σος. (Sigma) 02:37, 18 July 2015 (UTC)

Thanks. Weird, because in the past the tool automatically filled up the server for you. Narutolovehinata5 tccsdnew 02:51, 18 July 2015 (UTC)

It appears that MediaWiki:Histlegend should update `//tools.wmflabs.org/usersearch/index.html?page={{FULLPAGENAMEE}}` to someting like `//tools.wmflabs.org/sigma/usersearch.py?page={{FULLPAGENAMEE}}&server=enwiki`. PrimeHunter (talk) 03:10, 18 July 2015 (UTC)

Apologies for article info not working. Our team of developers cannot figure out what the problem is. As an alternative, per what Sigma said, I am working on making the German version articleinfo, called wikihistory, available for the english wikipedia. I should have it up soon.— [redacted] 14:22, 18 July 2015 (UTC)

It's not clear from the comments if this alternative is meant to be a temporary solution until the issues with revision history statistics are solved. Don't see an issue with a temporary alternative, something is better than nothing, but if it is meant as a permanent replacement this surely requires a RfC to get community opinion and consensus.--Wolbo (talk) 12:56, 20 July 2015 (UTC)

yes it seems that is possible as well--Ozzie10aaaa (talk) 13:19, 20 July 2015 (UTC)
Whether you want to use any tool temporarily or permanently is your own call, and as such does not require any consensus.--Anders Feder (talk) 13:53, 20 July 2015 (UTC)

Mobile editing

Has something changed recently with the mobile editor to make inadvertent deletions of the lead like this (https://en.wikipedia.org/w/index.php?title=Memphis_Depay&diff=prev&oldid=671937473) more likely? I'm seeing more of them, all mobile edits. --NeilN talk to me 14:45, 18 July 2015 (UTC)

I don't know statistics or an answer but will just mention that section blanking has always been common in the desktop version. However, desktop doesn't have an edit link for the lead (unless you have an account and enable a gadget), so in desktop it isn't the lead which is blanked. Mobile does have an edit link for the lead, and mobile editing is probably increasing in general. PrimeHunter (talk) 00:41, 19 July 2015 (UTC)
Indeed, tack "?useskin=minerva" at the end of any Wikipedia url, for example this page (https://en.wikipedia.org/w/index.php?title=Wikipedia:Village_pump_(technical)&useskin=minerva). The only edit link I see in my browser is the one to edit the lead, but I haven't played around with any Minerva preferences. --Unready (talk) 16:23, 19 July 2015 (UTC)

That's because this page is messy. Proper articles have a section edit button when you uncollapse the button. It depends on the ability to automatically detect sections, which is fragile (because wiktext has no proper sections, just headers). —TheDJ (talk • contribs) 11:10, 20 July 2015 (UTC)

Problem with a reference

Chrome	References [edit]
Opera	References [edit source] [edit]
Firefox	References
IE	References [edit]

Article: Central Belters. Any ideas, or have I done something stupid? Black Kite (talk) 15:17, 20 July 2015 (UTC)

The {{Reflist}} has a `colwidth=30em` in it that is stretching the citation. Jo-Jo Eumerus (talk, contributions) 15:23, 20 July 2015 (UTC)

Thank you, fixed. I borrowed the article layout from another article and didn't notice that. Black Kite (talk) 15:36, 20 July 2015 (UTC)

It's not stretching it, it's splitting it into shorter lengths in an attempt to display it in columns. Multi-column reflists are not normally useful when the number of references is small (10 or less) and the refs themselves are longer than about half the page width. --Redrose64 (talk) 16:50, 20 July 2015 (UTC)

Article prefix/namespace bug

We seem to have number of articles in namespace 0 with prefixes indicating they should not be. Tracked in Phabricator Task T87645

```
SELECT *
FROM enwiki_p.page
WHERE page_namespace = 0
AND page_title LIKE 'Wikipedia:%';
```

lists 34 rows. I'm guessing these are being created through an API that does not interpret namespace prefixes properly. Possibly related, there also appears to be an instance of User_talk:War_wizard90 in namespace 0 also - see this query (http://en.wikipedia.org/w/api.php?action=query&prop=info&pageids=45227522&inprop=url). - TB (talk) 19:45, 19 July 2015 (UTC)

phabricator:T87645. Jackmcbarn (talk) 19:50, 19 July 2015 (UTC)
Also Wikipedia:Village pump (technical)/Archive 134#API namespace issue / page reported existing in two namespaces and other threads linked from there. --Redrose64 (talk) 14:31, 20 July 2015 (UTC)

Cheers. We've a total of 69 misplaced articles as a result of this bug from 27th January this year - they can be seen at User:Topbanana/Temp. All have corresponding articles in the correct namespace and are almost entirely inaccessible through MediaWiki. I'll see what can be done about tidying them up. - TB (talk) 08:49, 20 July 2015 (UTC)

@Topbanana: These can be deleted through the API, using just the page ID, not the title. It should be quite easy to create a script that deletes all of them automatically if we use the list of known bad page IDs. — *Mr. Stradivarius* ♪ talk ♪ 14:20, 20 July 2015 (UTC)

All deleted now; the simplest solution was to correct the deletion confirmation form being produced by the MediaWiki software to retain the *curid* parameter when passed one rather than always using *title* to identify the subject of the operation. - TB (talk) 08:12, 21 July 2015 (UTC)

transwiki

Hi,

I'd like to transwiki (most of) my user pages and sub-pages to m:User:Jc37. I'm an admin here, but not currently an admin there. Is there a way to do this besides copy/paste?

Am I better off asking a steward (or someone else) to either do it for me or to ask a steward grant me Importer per m:Help:Import for this?

Basically, I'm not sure of my feet here and would like advice :) - jc37 22:40, 20 July 2015 (UTC)

@Jc37:. You need a Meta administrator - they can import from en.wiki to Meta. Ask at m:Meta:Requests for help from a sysop or bureaucrat. QuiteUnusual (talk) 14:30, 21 July 2015 (UTC)

Thank you very much :) - jc37 15:25, 21 July 2015 (UTC)

Tech News: 2015-30

Latest tech news from the Wikimedia technical community. Please tell other users about these changes. Not all changes will affect you. Translations are available.

Changes this week

- The new version of MediaWiki will be on test wikis and MediaWiki.org from July 21. It will be on non-Wikipedia wikis from July 22. It will be on all Wikipedias from July 23 (calendar).
- You now see more warnings in the image viewer. They tell you to be careful when using the image, for example if it shows a person. [19] (https://phabricator.wikimedia.org/T102693)

Meetings

- You can join the next meeting with the VisualEditor team. During the meeting, you can tell developers which bugs are the most important. The meeting will be on July 21 at 19:00 (UTC) (http://www.timeanddate.com/worldclock/fixedtime.html?hour=19&min=00&sec=0&day=21&month=07&year=2015). See how to join.

Future changes

- Soon you won't be able to use MathJax to display math. [20] (https://phabricator.wikimedia.org/T99369)

Tech news prepared by tech ambassadors and posted by bot • Contribute • Translate • Get help • Give feedback • Subscribe or unsubscribe.

03:06, 21 July 2015 (UTC)

Is the Compact Personal Bar gone permanently?

Ever since the official discontinuation of the Compact Personal Bar, I've force-enabled it by adding special code to my vector.js file. However, the bar seems to no longer load. I tried about an hour ago, and it worked fine. Has it been permanently removed even for those who force-enabled it? Gparyani (talk) 18:58, 19 July 2015 (UTC) Tracked in Phabricator Task T104659

I thought that was phab:T104659, which does not appear to have happened yet. Whatamidoing (WMF) (talk) 03:31, 20 July 2015 (UTC)

@Whatamidoing (WMF): Actually, it's been updated since you posted that comment; back then, it was outdated. See the linked bug report in the bottom reply of the one you linked (phab:T87489). Gparyani (talk) 07:29, 20 July 2015 (UTC)

Yes, it looks like Ori has removed it from the servers. The WMF devs seem to be spending some time figuring out what they can support and removing things that are at risk for bitrot due to lack of resources for maintenance. I was personally never very fond of that one, but I know that it had a couple of staunch fans. While it might be possible to mimic a few of its features in CSS, it looks like the tool itself is gone for good. Whatamidoing (WMF) (talk) 18:58, 20 July 2015 (UTC)

> @Whatamidoing (WMF): Could you get someone to fix the "critical" bugs holding this feature back? I was long aware of them, but I continued to use it because I had never ran into these bugs during the time it was official, and I was willing to accept the risk of running into them later on (I never ran into any one of them afterwards). Gparyani (talk) 19:29, 20 July 2015 (UTC)
>
> > @Gparyani:: "due to lack of resources".... —TheDJ (talk • contribs) 20:10, 20 July 2015 (UTC)
> >
> > > Yeah, with a mere 65 million USD in planned spending this year, the Foundation really has to tighten its belt and cut back to the bare bone. --108.38.204.15 (talk) 07:20, 21 July 2015 (UTC)
> > > I'm not getting into that argument, the fact is that the current resource allocation doesn't allow to support this. Also, it was a beta, primarily to explore an idea, not to deliver it. —TheDJ (talk • contribs) 18:15, 21 July 2015 (UTC)

I don't want to argue the fiscal aspects here either but do take issue with the idea this couldn't be done per a lack of resources or whatever. It seems to me that the approach -- basically a drop-down menu bulleted with "icons" for each label a la OOUI -- was the only avenue ever experimented with. And, once that approach failed due to some extraneous gadget interaction issues, it seems the entire premise (a suitable compact replacement of Personal toolbar's default text) was deemed not worth pursuing at all.

One would have thought alternatives would still continue to be sought out in spite of that one and only approach's failure -- especially if the availability of the 'work-in-progress' is being permanently removed at the end of the day -- instead of being dropped entirely in spite of the apparent demand for something compact. -- George Orwell III (talk) 20:29, 21 July 2015 (UTC)

Compact another way

> What is the "apparent demand" we are talking about here? What number of users who regret this trial being discontinued are we looking at?--Anders Feder (talk) 21:00, 21 July 2015 (UTC)
>
> > I can't say overall but on en.wikisource, the premise of some sort of "compact" personal tool bar (along with side-bar elimination and/or relocation) was in the majority as far as frequent/regular contributors were concerned. In short, CPB was "popular" until it was no longer a matter of just opting in like most typical Beta offerings. The resulting drive for alternatives vary; the thumbnail depiction is but one avenue taken.
> > Besides that, I'm under the impression a redesign of the personal (and again, side-bar) toolbar(s) is more of a design necessity than an actual response to demand or polling per the stated infobox blurb for CPB and it's pointer to the proof-of-concept (https://www.mediawiki.org/wiki/Winter) example page. Either way, folks who jump between mobile, tablet and desktop platforms seem to appreciate the 'less is more' (i.e. compact) design direction "we" seem to be on imho. -- George Orwell III (talk) 21:32, 21 July 2015 (UTC)
> >
> > > It's the same as with most changes, overall readers love it, but getting something like this production deployed is massively expensive. Think community communication, multiple variant testing, aftercare due to teething problems, rewriting extensions that use such a thing, rewriting the testcases, making it accessible for the visual impaired, fixing all the user scripts on hundreds of wikis. You would have to invest at least some 25-30 times more than the original hours of the experiment. —TheDJ (talk • contribs) 22:04, 21 July 2015 (UTC)
> > >
> > > > Points well taken here and can't argue with the rationale behind them but not every "redesign" needs to be in the form of an OOUI, .menupokey driven drop-down background-image bulleted menu. My only point was if the given "Winter" prototype design is any basis for what the future really holds for us, simply switching from text to clickable icons (or background images / mw-buttons if you like) as depicted would have helped meet the new "space" requirements for the current personal toolbar while solving the occasional gadget integration issue(s) without de-railing the entire notion & development of a "flat" or "fixed" article/skin header redesign (Vector-beta) in the process. Now, without CPB as a component, the entire endeavor seems stalled along with losing any chance of somebody fresh "stumbling in" with even better alternatives or refinements to the personal tools redesign question; be that compact or otherwise & all while moving forward. That seems short-sighted imo. -- George Orwell III (talk) 22:32, 21 July 2015 (UTC)

Music

In this table there's a column that says "What links to this file". There are many other tables with identical columns. Instead of discovering and then typing the names of individual Wikipedia articles in that column, is it possible to type some identical code together with the file name? Then the code would automatically display whatever Wikipedia articles link to the file. I think that would make things much easier for me, user:Raul654, User:Ravedave, User:Antandrus, User:Graham87, User:La_Pianista, user:Violaruley, and User:Ftiercel. Thanks for any reply.Anythingyouwant (talk) 05:52, 20 July 2015 (UTC)

> As a first approximation, {{Special:Whatlinkshere/File:Oh holy night.ogg|namespace=0}} produces:
>
> You might be able to get cleaner output with a Lua module. —Cryptic 06:16, 20 July 2015 (UTC)
>
> > Interesting, thanks very much. I have updated the table accordingly, and it works well.Anythingyouwant (talk) 06:26, 20 July 2015 (UTC)

Music files

Is there an easy way to figure out which files in this category at Wikimedia Commons are not yet listed in this table at Wikipedia?

And here's a second question: is there an easy way to make a list of files in this category at Wikimedia Commons that have not yet been included in any article at English Wikipedia? Anythingyouwant (talk) 06:26, 21 July 2015 (UTC)

> With AWB's list comparer I made these lists (https://en.wikipedia.org/w/index.php?title=User:SiBr4/AWB_Reports&oldid=672399339#.22Ba.22_music_files), which show there are 473 files in subcategories of the Commons category that are not linked on the project page.
> As for the second query, I thought using {{#if}} on a transcluded WhatLinksHere page could determine whether or not the list of usages is empty (i.e. `{{#if:{{Special:WhatLinksHere/Foo}}|1|0}}`), but it seems the special page is only expanded after the parser function is evaluated, as the #if always returns 1 even though the result of transclusion for an unlinked page is an empty string (also, interestingly, `{{#ifeq:{{Special:WhatLinksHere/Foo}}|1|1|0}}` returns 0). SiBr4 (talk) 08:59, 21 July 2015 (UTC)

Thanks, the lists made using AWB will be helpful, and I'll see if I can make similar lists for other letters of the alphabet. It will be quite tedious to add so many items from the Ba category to the Ba table, and it sure would be nice if it could be done automatically. Can it? Cheers.Anythingyouwant (talk) 17:06, 21 July 2015 (UTC)

I figured out how to make a merged document from your list of 473 songs, and then pasted the merged list into the "Ba" table. It's kind of a crummy and incomplete set of info that I pasted, but it is way better than nothing, I think.Anythingyouwant (talk) 03:55, 22 July 2015 (UTC)

Article rename

Could someone help rename an article? The article 9-Man should be called "*9-Man* (documentary)". The title should be italicized, and it should be noted that it is a documentary, or it will easily be confused with 9-man. Every time I change an article name I mess it up. Thanks! Magnolia677 (talk) 10:29, 22 July 2015 (UTC)

@Magnolia677: I have moved the article to 9-Man (film) since I agree that relying on the capitalization difference to distinguish it from the game 9-man is confusing. For your reference, italics are not truly *part* of article titles but are instead applied with the template {{Italic title}} in the body of the article (usually at the top). This template automatically leaves the parenthetical disambiguation in roman (ie. not italicized). Cheers —jameslucas (" " / +) 13:04, 22 July 2015 (UTC)

Another watchlist proposal: Symbol to replace (0) when net effect is no change at all

In the case where a sequence of edits has resulted in no net change to the source text at all (not just no net change to the *length* of the source text) how about replacing the (0) length-delta with something else, perhaps (∅)? It's useful to be able to recognize this special case at a glance. EEng (talk) 14:02, 2 July 2015 (UTC)

Collapse interesting conversation which basically came up with the idea of using *()* instead of null-set symbol

It's a good idea to distinguish between ε and 0, but the numeric field probably isn't the place to do it. Perhaps an "=" sign (or ==, or ===, depending on your preference...) after the numeric thus "(0)=" . I think it would break a smaller number of applications. All the best: *Rich Farmbrough*, 18:43, 2 July 2015 (UTC)

Good point. Maybe it could be worked into whatever it is you folks are cooking up with the green and blue arrows and dots and whatnot. EEng (talk) 19:40, 2 July 2015 (UTC)

~~Since the containing the "(0)" has its own CSS class, called "mw-plusminus-null", changing the text to something else can be done with some relatively simple user-specific JavaScript (I've tested it (https://en.wikipedia.org/w/index.php?title=User:SiBr4/common.js&diff=669672498)).~~

~~This uses the "∅" empty set symbol (U+2205 (http://unicode-table.com/en/2205/)) surrounded by parentheses, as above; this can be replaced with any wanted string. SiBr4 (talk) 19:22, 2 July 2015 (UTC)~~

~~Misread the question. SiBr4 (talk) 19:30, 2 July 2015 (UTC)~~

~~Yeah, I was wondering just where in that code the test for "no net change" was, but I thought, "Well, these Village Pump gnomes must know something I don't" and went to try it. Guess what? You did misread the question. But it does a beautiful job of turning zeros into "empty-sets". Thanks for the effort, though. EEng (talk) 19:40, 2 July 2015 (UTC)~~

- ~~Any chance on someone doing this? EEng (talk) 12:21, 7 July 2015 (UTC)~~

~~There's not much we can do here, other than using JavaScript to retrieve the two versions and making a comparison, but that would be rather slow - assume that your watchlist shows 50 edits, that means that 100 pages (50 pairs) need to be retrieved, and for one of the pages in each pair, every byte compared against the corresponding byte in the other page of the pair. Functions exist to compare strings of bytes, not sure if they'd handle strings that were several hundred K in length without breaking into substrings. If nobody is willing to try it in JavaScript, you could file a feature request at phab:. —Redrose64 (talk) 12:34, 7 July 2015 (UTC)~~

~~(You're not sure there are functions to compare long byte strings? Are you kidding???) Retrieving and comparing the actual text is obviously out of the question, but with all due respect I question the accuracy of your analysis. When I hover over e.g. *4 changes* (or whatever), where someone made a bunch of changes and someone else reverted them, it easily pops up with an empty diff, and that isn't happening by two full versions being retrieved and compared on the fly — something somewhere knows, without too much trouble, that these two versions of have a null diff. EEng (talk) 12:51, 7 July 2015 (UTC)~~

~~@EEng: A page diff is run on the Wikimedia servers and has access to all sorts of functions. Any javascript that customises display for a particular user is run client side, and so any functions and data that are used must be available to the client. —Redrose64 (talk) 15:19, 7 July 2015 (UTC)~~

~~That doesn't explain why, if the javascript can request the two pages themselves (to do its own diff), it can't just as easily request the diff directly. But anyway, since the hashes appear to be available, this is moot (unless we want to improve the performance of the hover diffs, which would be a good idea — no wonder they're so slow!). EEng (talk) 15:26, 7 July 2015 (UTC)~~

~~The null diff might perhaps use a symbol for unit type, sometimes denoted '()'. It's not the same symbol as a 'nothing at all'; it symbolizes 'action which leaves what you care about unchanged', such as adding zero to your number, or multiplying it by one. —Ancheta Wis (talk | contribs) 13:08, 7 July 2015 (UTC)~~

~~If I'm understanding you you're suggesting omitting the zero, so that *(0)* becomes *()*. That's a great idea, and might (at least partially) address RF's concern about breaking existing applications, since (one hopes) empty string will be interpreted as zero, for those applications that just want the length. EEng (talk) 13:30, 7 July 2015 (UTC)~~

~~@EEng: If you are using Popups to generate diffs when you hover over a diff link, it actually does retrieve two full versions and compare them on the fly. Popups contains its own diff generator, separate from the MediaWiki one, which is why it sometimes says "diff truncated for~~

~~performance reasons" but the native MediaWiki one doesn't. You can see the code for it by searching for "Javascript Diff Algorithm" in MediaWiki:Gadget-popups.js. Also, if you open the part of your browser console that monitors network requests, you can see the requests for each of the two pages being sent each time you hover over a different diff link. — *Mr. Stradivarius* talk 14:54, 7 July 2015 (UTC)~~

> ~~My apologies. I clearly underestimated the potential depths of implementation insanity. Why in the world don't they get Mediawiki to generate a diff and just use that? Luckily, as seen below, this can all be shortcut via hashes, as seen below. Any thoughts about the interface change to *()*? EEng (talk) 15:03, 7 July 2015 (UTC)~~

~~It would seem that a SHA-1 hash is generated for every page revision – see mw:Manual:Revision table#rev_sha1. Now, whether the hash is exposed through the JavaScript API, I've no idea. Alakzi (talk) 13:37, 7 July 2015 (UTC)~~

~~It is indeed exposed by mw:API:Revisions with `prop=revisions&rvprop=sha1`. So, it is simply a matter of retrieving the hashes of the first and last revision in a series and doing a string comparison. [21] (https://en.wikipedia.org/w/api.php?action=query&revids=670002091&prop=revisions&rvprop=sha1)[22] (https://en.wikipedia.org/w/api.php?action=query&revids=670074851&prop=revisions&rvprop=sha1) Alakzi (talk) 13:54, 7 July 2015 (UTC)~~

> ~~Yippee! That's even better than a very efficient diff, because it's obviously available for free. So before we recruit some knowledgeable gnome to implement this, what do we need to do to make sure everyone who might care is OK with the suggested output format change i.e. *()*? Paging Rich Farmbrough. EEng (talk) 14:49, 7 July 2015 (UTC)~~

> > ~~It doesn't need to be backward-compatible with anything if we're writing a user script. If you want this to be changed in core, just open a ticket on Phabricator; it seems fairly straightforward, so it might even be implemented before the turn of the century. Alakzi (talk) 15:24, 7 July 2015 (UTC)~~

> > > ~~It's an interesting dilemma: a user script would get it sooner *for me*, but this seems like something that would benefit most users, but they won't get that benefit if they have to know about a script to install, so soon that way it's better to request it in core — but that will probably delay *my* getting it. So should I go for my own selfish interests, or the greater good? EEng (talk) 15:30, 7 July 2015 (UTC)~~

> > > > ~~Why not both? ;-) Alakzi (talk) 15:38, 7 July 2015 (UTC)~~

> > > > > ~~Yeah, I thought of that too, but how much you want to bet, if there's a user script available,, that implementing it in core gets deferred because "there's already a user script for people who want this". Honestly I'm amazed this feature wasn't in there from the very beginning — it's such an obviously important special case. In fact, it really ought to be integrated into the overall grammar of bullets, arrows, bolding, coloring, and so on of the watchlist, since the idea is to help people filter out the unimportant (including the null) and focus on actual changes they care about. EEng (talk) 16:39, 7 July 2015 (UTC)~~

> > > > > > ~~I like that idea! Maybe a ring in place of the bullet? Alakzi (talk) 16:41, 7 July 2015 (UTC)~~

> > > > > > > ~~Part of the reason I brought it up is I have the impression that stuff's being changed right now (or maybe I just haven't been looking closely for a while). How do we get the right person's attention? EEng (talk) 16:51, 7 July 2015 (UTC)~~

> > > > > > > > ~~First find out who the right person is! All the best: *Rich Farmbrough*, 22:42, 10 July 2015 (UTC).~~

Is that, like, a riddle? EEng (talk) 23:28, 10 July 2015 (UTC)

I'm thinking about how to implement something so that we could visually identify net-null pairs in revision lists. The SHA-1 hashes from the API make the comparisons trivial, but I'm not sure how best to represent the results of the comparisons in the revisions list. In particular, two problems:

- I'm not sure how to model cases like back-and-forth reverts. For instance, if we had a case where the unique hashes were `[A,B,C,B,C,D]`, in order with "D" the most recent hash, how should I show the relationships between both the two "B" revisions and the two "C" revisions? My first thought was a "layered stripes" system with lines linking the first, last, and any middle instances of matches, but that could easily end up with as many as $n/2$ layers for n revisions (or more? Haven't thought it through…) with obvious nonideal ordering like `[A,B,C,D,C,B,A]` or `[A,B,A,C,B,D,C,D]`, so it's probably not a good design.
- I'm not sure how to represent pairings visually. I want to do it in a way that a) avoids modifying the revision list much, for compatibility and such and b) doesn't rely on colour if possible (using it is OK, but ideally it wouldn't be necessary for comprehension).

I'll think about it some more, but input would be helpful to solve or sidestep those problems. There's got to be something more elegant, but I'm tired and it's escaping me. {{Nihiltres|talk|edits}} 06:12, 11 July 2015 (UTC)

> Thanks for taking this on! I think the more complex output you're envisioning would be great if we could find an elegant format, but it still might be best left to the future. As a first cut I think the most bang for the buck comes from just comparing the start and end hashes for "today" (corresponding to the two versions from which the size difference is computed) and if they're the same, changing *(0)* to *()* as someone suggested above (assuming no one sees a compatibility problem with that). BTW, if retrieving the hashes is a separate step from retrieving the sizes, then we can skip getting the hashes in the very common case that change-in-length is not 0.
> I thought about "unbolding" when hashes match (and certain other conditions hold) but it starts to get not so clear what should happen when there have been changes over more than one day, when the user has visited sometime in the middle of today's sequence of changes, etc. I'll give some thought to an "ABC"-type interface you're proposing but for now I wanted to get the above posted. EEng (talk) 00:58, 12 July 2015 (UTC)

> > The problem I see for the watchlist version of the idea is that I don't see a good plan for selecting the old version would be for comparison. The ideal would be to get revision IDs based on the last visit, probably based on whatever the "updated since your last visit" system uses, but I don't see a way to do that. Another way is to select some relatively arbitrary date in the past and then pull revision hashes from then, which would likely give inconsistently useful results. Alternatively, I could rig up some system to store "most recent revisions on last visit" locally, but that seems really flaky for a whole bunch of reasons. Another idea would be to simply check the previous 10 or so revisions of each page for identical hashes and add some note or other about the timestamp of the matching hash, but that'd be really inefficient, especially for large watchlists.
> > My idea was more to augment the *history pages*, because seeing net-null pairs would be really useful and be a useful model for improving history pages (and thus probably also the watchlist) in MediaWiki proper—I don't think the watchlist idea is useful right now. The catch is, as I mentioned above… the problem of how to lay out the results sanely. Does my line of thinking make more sense now?
> > {{Nihiltres|talk|edits}} 17:48, 12 July 2015 (UTC)

I missed where you said *revision lists* -- sorry. That's an interesting idea, but if you'll excuse my selfishness I'd like to stick with the watchlist idea for a minute, because it's really quite simple. The hash comparison is meant to check for a special case of the *(0)* length-change indication i.e. the special case where, not only did the length not change (net), but nothing at all changed in the byte image of the article source (net). The the versions to be hash-compared are precisely the same as the two versions whose lengths were subtracted to produce the length-change i.e. the version at the end of "yesterday" (how ever that's defined -- UTC and so on) vs. the current version.

The intent is simply to ameliorate a frequent nuisance: a little-edited article pops up on your watchlist. It has two changes, and the net length-change is *(0)*. Now, that almost always turns out to a vandal's edit followed by someone's reversion, and so it's tempting to just assume it is indeed that. But to really tell, you have to hover over the *2 changes* and wait for the diff popup, which is annoyingly slow, not to mention inefficient for everyone. As a result, I usually just assume the net change is null, and don't check.

It's not a big deal, but it bugs me. This feature would make it immediately obvious when the net change is null, because the summary line for the article will show *()* instead of *(0)*. With these clarifications wouldn't htis be pretty easy to do?

Once again, I promise to think about ABC, but I wanted to post this. EEng (talk) 19:35, 12 July 2015 (UTC)

> If this ever gets implemented, it would be nice to have an option to have no-net-effect changes just not show up at all in the watchlist. More room for the actual changes. Regards, ORANGE SUEDE SOFA (talk) 19:41, 12 July 2015 (UTC)
>
> > I thought about this (a similar idea is to "unbold" the article) but you run into a lot of problems about what to do when e.g. the last visit was somewhere inside today's sequence of edits. EEng (talk) 22:44, 12 July 2015 (UTC)
>
> @EEng: If I simplify it down to "last 2 changes are net-null" it'd be workable, but probably >90% of the time it'd be redundant to a revert's edit summary. I'll toy around with some code for now and keep you updated. {{Nihiltres|talk|edits}} 01:56, 13 July 2015 (UTC)
>
> > No, not the last two changes! (That was just an illustration I gave.) What we want are the two changes on which the net size difference is based i.e. the first and last change *today*. Do you understand what I'm saying? Otherwise, it doesn't make sense to change *(0)* to *()*. EEng (talk) 04:47, 13 July 2015 (UTC)
> >
> > > Uh, aren't all the size differences displayed just from the most recent (single) edit? Either way, I don't see an efficient way to do that through the API. {{Nihiltres|talk|edits}} 16:02, 15 July 2015 (UTC)
> > >
> > > > We're obviously talking at cross-purposes here. Could we discuss this via IRC? I have no idea how to do that, but this would be a good time to learn. EEng (talk) 18:56, 15 July 2015 (UTC)
> > > >
> > > > > @EEng: WP:IRC explains a bit, if you do not want to download an IRC client you can use Freenode's Webchat (https://webchat.freenode.net/). It asks you to fill in a channel, the main Wikipedia channel is called #wikipedia. The Quixotic Potato (talk) 19:28, 15 July 2015 (UTC)
> > > > >
> > > > > OK, thanks for the advice, QP. How about it, Nihiltres? EEng (talk) 02:20, 18 July 2015 (UTC)
> > > > >
> > > > > <bump>, Nihiltres? EEng (talk) 02:52, 20 July 2015 (UTC)
> > > > >
> > > > > Anyone? Or must this one die on the vine like so many others? EEng (talk) 04:45, 23 July 2015 (UTC)

POTD template: direction of wide image scrolling

I have scheduled File:Chen Rong - Nine Dragons.jpg for POTD on August 3. The work is supposed to be viewed from right to left, and in the article proper the template is coded to do so:

{{Panorama |image = File:Chen Rong - Nine Dragons.jpg |height = 230 |alt = |caption = |dir = rtl }}

Is there a way to orient the POTD image like that as well? — Chris Woodrich (talk) 23:59, 14 July 2015 (UTC)

> Sure, one way is to do the same as {{Panorama}}, passing on dir = rtl. Edit {{Wide image-noborder}} and add this right after `style="overflow:auto;` (and before the ending "): `{{#ifeq:{{{dir|}}}|rtl|direction: rtl;}}`. Then add `|dir = {{{dir|}}}` when {{Wide image-noborder}} is called in {{POTD default}} so the parameter can be passed on, and add `|dir = rtl` to the call in {{POTD/2015-08-03}}. PrimeHunter (talk) 00:41, 15 July 2015 (UTC)
> Oh, {{Wide image-noborder}} already has an undocumented dir parameter so no edit there is needed. PrimeHunter (talk) 00:44, 15 July 2015 (UTC)
> I have made the other edits.[23] (https://en.wikipedia.org/w/index.php?title=Template:POTD/2015-08-03&diff=prev&oldid=671485950)[24] (https://en.wikipedia.org/w/index.php?title=Template:POTD_default&diff=prev&oldid=671486212) PrimeHunter (talk) 00:54, 15 July 2015 (UTC)
>
> - Thank you. Nice to know that we've got the same functionality. I do wish the POTD templates were better documented. — Chris Woodrich (talk) 01:54, 15 July 2015 (UTC)
>
> > I wish almost *all* templates were better documented. Not to mention better-named... (https://en.wikipedia.org/?oldid=681233152#Idiotic_naming) EEng (talk) 03:17, 15 July 2015 (UTC)
> >
> > So is it documented *now*? --Thnidu (talk) 19:40, 19 July 2015 (UTC)
> >
> > > I have documented dir=rtl in {{Wide image-noborder}} and Wikipedia:Picture of the day/Guidelines#Template parameters. PrimeHunter (talk) 16:58, 22 July 2015 (UTC)

Discrepancies in search result numbers

I've been using AWB to add missing commas to month-day-year-formated dates that appear in the middle of sentences. At present, I'm focusing on the 20th century, so my search parameter is `insource:/[yhletr] *[1-3]?[0-9], *19[0-9][0-9] [a-z]/` (https://en.wikipedia.org/w/index.php?title=Special:Search&search=insource%3A%2F%5Byhletr%5D+%2A%5B1-3%5D%3F%5B0-9%5D%2C+%2A19%5B0-9%5D%5B0-9%5D+%5Ba-z%5D%2F&ns0=1&fulltext=Search). If I search for this parameter using the site's native search engine, I get some number of results, typically between 800 and 1100. If I run the same search through AWB, I get 1049 results added to my list, and when I've edited them all I can get 1049 more, which implies that the number of relevant hits may be very large. Is there a reason that the built-in search would return only an inconsistently sized subset of search results? Is there any efficient way to find the true total number of matches so that I can get a sense of the scope of my undertaking? Cheers, —jameslucas (" / +) 12:56, 22 July 2015 (UTC)

> Probably because regexp's on all the content we have are a VERY expensive operation. So expensive, that some users (possibly AWB users doing things similar to you) actually brought the entire search cluster down, and seatbelts had to be installed to protect everyone. You might have run into one of the seatbelts. If you want to do efficient searches, you should use something on toollabs that talks to local copies of the database and is built

specifically to do this. Quarry (http://quarry.wmflabs.org) does this for SQL queries for instance. I'm not sure if there is a tool that does it for source text matching. Regardless, it will probably be slow :) —TheDJ (talk • contribs) 16:40, 22 July 2015 (UTC)

AWB can use regexps to search within a downloaded database dump. A query such as yours would take about an hour to run on my laptop. I'd offer to run it for you, but my latest download is about two months old so the results would be out of date. -- John of Reading (talk) 16:53, 22 July 2015 (UTC)

I think we're on the right track. I revised my search to ", 19" insource:/[yhletr] *[1-3]?[0-9], *19[0-9][0-9] [a-z]/ (https://en.wikipedia.org/w/index.php?title=Special:Search&search=%22%2C+19%22+insource%3A%2F%5Byhletr%5D+%2A%5B1-3%5D%3F%5B0-9%5D%2C+%2A19%5B0-9%5D%5B0-9%5D+%5Ba-z%5D%2F&ns0=1&fulltext=Search), since (if I understand correctly) the first search term is easier to search and therefore removes many, many hits before they are subjected to the regex term. I got ~1600 hits, which I suspect means I managed to get more searches in before I triggered a "seatbelt". John of Reading, would you be willing to run the query on the out-of-date dump? I'm looking for an order of magnitude more than anything. The machine I have available has only a small solid-state drive and I can't handle the dump myself. Most appreciatively —jameslucas ("" / +) 17:17, 22 July 2015 (UTC)

@JamesLucas: Your first "insource" here includes the comma, so it matches articles containing, say, "May 6, 1921". These are the ones that *don't* need correcting, is that right? I aborted the scan after it found the first 1,000 articles; it was going to find about 30,000 in all. I've begun a scan for `[yhletr] *[1-3]?[0-9] +19[0-9][0-9] *[a-z]` which is heading for about 1,000 articles; I will post them in a sandbox somewhere when it is done. -- John of Reading (talk) 17:59, 22 July 2015 (UTC)

I wrote the example to the right when discussing the work with The Quixotic Potato the other day. The commas between the month and the year and the city and the state are usually present, and when they're not, they're picked up by a number of other methods—rules in AWB, bots, mindful editors. The commas after the year and the state are often missed, and they're the ones I'm going after. Thanks so much for the anaylsis! 30,000 is in line with my expectations. —jameslucas ("" / +) 18:26, 22 July 2015 (UTC)

> On July 14, 2015, a Wikipedian operating out Brooklyn, New York, began a tiny crusade against missing commas. jameslucas

Ah, *those* commas. If you'd like up-to-date results, stick User:John of Reading/Latest download on your watchlist so you can see when I've downloaded a fresh copy, and then ask me on my talk page. -- John of Reading (talk) 18:41, 22 July 2015 (UTC)

Stalking commenced! —jameslucas ("" / +) 18:56, 22 July 2015 (UTC)

No such seatbelt? Insource:/<1900-1999>/ prefix:A (https://en.wikipedia.org/w/index.php?title=Special:Search&search=insource%3A%2F%3C1900-1999%3E%2F+prefix%3AA&ns0=1&fulltext=Search) returns 44 thousand. — CpiralCpiral 19:41, 22 July 2015 (UTC)

insource:/[yhletr] *<1-31>, *<1900-1999> [a-z]/ (https://en.wikipedia.org/w/index.php?title=Special:Search&search=insource%3A%2F%5Byhletr%5D+%2A%3C1-31%3E%2C+%2A%3C1900-1999%3E+%5Ba-z%5D%2F&ns0=1&fulltext=Search) gives 1049. THe other doesn't match `01`. — CpiralCpiral 19:41, 22 July 2015 (UTC)

No doubt, simplifying the criteria can prolong the search before the safeguards clamp down. I'll certainly be tempted to spend time improving the parameters, but I'll try to remember that the edits are the true and (comparatively) meaningful objective! —jameslucas ("" / +) 02:37, 23 July 2015 (UTC)

The numbers are changing for some reason you have to discover. I'm not primarily addressing that, (although I believe most regexp searches can always be improved). I'm encouraging *you* to discover some reasons. But it can't be the regex engine's safeguards throttling down the user's observed "number of matches" (shown to the far right of the search box query). The entire search domain must be searched every time and the full quantity always reported, or else how could one *set* a search domain (using a namespace or a prefix) from one query to another? — CpiralCpiral 03:15, 23 July 2015 (UTC)

Template glitch

There is an error in List_of_current_NCAA_Division_I_women's_basketball_coaches I can't quite seem to track down.

To see the error, go to the page and click on the team column to sort alphabetically. The error will pop to the top. I looked at the entry for Julie and didn't see anything wrong with it. I thought it might be a problem in the template above the Southland template, and I thought I found a problem in it and fixed it but that did not solve the problem.

I tried re-creating the entry for Julie by copying another entry and bring in the information for Julie. Oddly that corrected row now appears as a second row but I don't know where the first row is coming from. --S PHILBRICK (TALK) 16:01, 22 July 2015 (UTC)

@Sphilbrick: fixed (https://en.wikipedia.org/w/index.php?title=Template:Wbb_coaches/Mountain_West_Conference&diff=672598290&oldid=671004459) - NQ (talk) 16:08, 22 July 2015 (UTC)

I think I've fixed it. Needed a noinclude doc at {{Wbb coaches/Southland Conference}}. --Jules (Mrjulesd) 16:10, 22 July 2015 (UTC)

thanks for the quick response. I thought I looked there, but sometimes with this sequence of templates the problem is in the preceding template and I found a problem there which I fixed so assume that was it. I now see a problem with the Mountain West but I think I can fix it-- S PHILBRICK (TALK) 16:16, 22 July 2015 (UTC)

Change the "Edits by user" external tool link

Clicking on the current link goes to a page that says that the tool is defunct and that this tool [25] (https://tools.wmflabs.org/sigma/usersearch.py) should be used instead. Can the link on Wikipedia history pages be changed to this one? Gparyani (talk) 21:02, 22 July 2015 (UTC)

Oh, and it doesn't prefill the "page" field with the page I came from when clicking on the link to go to the new one. Can that be fixed? Gparyani (talk) 21:06, 22 July 2015 (UTC)
Are you talking about one of the links on MediaWiki:Sp-contributions-footer? Jo-Jo Eumerus (talk, contributions) 21:07, 22 July 2015 (UTC)

@Jo-Jo Eumerus: No; it's at the top of every page history where it says "external tools". Gparyani (talk) 21:13, 22 July 2015 (UTC)

Ah. Then it's MediaWiki:Histlegend instead? Jo-Jo Eumerus (talk, contributions) 21:14, 22 July 2015 (UTC)

@Jo-Jo Eumerus: Yes. Gparyani (talk) 21:16, 22 July 2015 (UTC)

I have updated the link [26] (https://en.wikipedia.org/w/index.php?title=MediaWiki:Histlegend&diff=672638350&oldid=672522228) as I suggested at #Various tools are down without getting comments. PrimeHunter (talk) 21:34, 22 July 2015 (UTC)

Private incident reporting and tracking system for admins

First of all ANI/Incidents is currently the most frequently edited page. Second of all, it seems that it's possible to report things privately to arbcom, but not to the administrators as a whole. There's a number of pros and cons to having a private means to report incidents to administrators. There's two potential approaches, one is all admins having access to this tool, a second is only admins IDed to WMF have access.

Pros:

1. Good for shy editors who aren't comfortable drawing attention to themselves.
2. Good for issues that editors might be uncomfortable bringing up in public, such as sexual harassment.
3. Would make ANI/Incidents less of a huge mess.
4. Would reduce burden on arbcom potentially, as minor issues regarding some privacy wouldn't have to go through them.
5. Less drama, canvassing opportunities, etc.
6. Might be good for other language wikis, that may have different cultures regarding reporting incidents.
7. Most other websites have a means to privately report issues to administrators, so it's what many people are used to.
8. Might make the lives of admins easier, as a wiki is not an ideal issue tracking system.

Cons:

1. Potential legal issues if admins not identified to WMF have access. Editors might assume issues reported privately to administrators is private information, even though such information is not legally.
2. Potential off site drama with wp admins with ill intent leaking private incident reports.
3. Potential for misuse (like any new tool).
4. Might make ANI/Incidents (which this would certainly not replace) get less admin attention.

Personally I'd support such a tool, as it would be tremendously helpful to editors who might have difficulty "raising their voice" so to speak. --ScWizard (talk) 17:43, 22 July 2015 (UTC)

Having some concern about such a tool attracting a lot of reports that don't need to be private or don't need to be targeted at administrators. That's my expectation based on reports I handle on other websites, for the record. Jo-Jo Eumerus (talk, contributions) 17:49, 22 July 2015 (UTC)

A report that doesn't need to be private, but also doesn't need to be public, would be more quickly resolved through a proper incident tracking system than through a wiki, I think. As for reports that don't need to be targeted at admins, you're right that there are likely to be reports in the fashion of "this article is wrong!!!" by individuals who are completely unfamiliar with wikipedia. However such a report might give the wiki an opportunity to engage people who wouldn't think of editing the wiki otherwise. --ScWizard (talk) 17:56, 22 July 2015 (UTC)

There are experienced editors who are active on ANI but who are not administrators. Any change that hides the sort of cases we see on ANI from those editors denies Wikipedia a useful tool for dealing with abuse. --Guy Macon (talk) 20:49, 23 July 2015 (UTC)
I think having one "court" that can work semi-privately is enough (Arbcom). Don't need more WP:CABAL accusations. --NeilN talk to me 20:59, 23 July 2015 (UTC)

French Wikipedia has Global contributions Special page

I just noticed that French Wikipedia has the Special:CentralAuth tool which allows you to query user contributions and rights across all Wikimedia Projects. See for example: https://fr.wikipedia.org/wiki/Sp%C3%A9cial:CentralAuth/Sadads . Is there a reason we don't have this implemented? Can we implement it? It would be insanely useful for things like whether to ping someone else on another wiki, checking whether to give editors rights or when blocking editors, for WP:The Wikipedia Library screening if editors have sufficient contributions to qualify for partnerships access, etc. I can't think of a reason not to have it enabled. Sadads (talk) 20:30, 22 July 2015 (UTC)

We have as well, actually. Jo-Jo Eumerus (talk, contributions) 20:34, 22 July 2015 (UTC)
We link to it on "accounts" at the bottom right of user contributions, at least if you have either the default English or British English as interface language at Special:Preferences. The link is made by MediaWiki:Sp-contributions-footer. Does the French Wikipedia link to it in a place you think we also should? PrimeHunter (talk) 21:13, 22 July 2015 (UTC)

To be precise the one named "accounts" at bottom right of contribs is m:Special:CentralAuth/Sadads which goes through meta:. But the action is the same as Special:CentralAuth/Sadads on en.wp and fr:Special:CentralAuth/Sadads on fr.wp. Any differences that you may see are down to your user settings - such as interface language, time zone or skin. --Redrose64 (talk) 22:32, 22 July 2015 (UTC)

Right, we link the meta version on user contributions. It gives the same for most accounts. Special:CentralAuth/Thisfeelsawesome versus m:Special:CentralAuth/Thisfeelsawesome shows an example where meta adds global account changes. PrimeHunter (talk) 22:52, 22 July 2015 (UTC)

Nevermind on not having it, I must have had a typo in trying to find it. However, we don't link it anywhere obvious (for example at the top of Special:Contributions/Sadads). It seems silly to be sending users to meta, without a significant difference (its probably really disorienting for English only contributors). Moreover, I seem to remember a ton of different templates linking to this tool (https://tools.wmflabs.org/guc/?user=Sadads&blocks=true), which is always clunky and slow, but not to CentralAuth. Could we add it to the top of Special:Contributions? And I didn't even know that box was at the bottom of Special:Contributions. Is there a way we could move that up the page, make it discover-able --- it seems like a lot of valuable information that shouldn't be hidden at the bottom of the page (maybe a side bar on the right?). French Wikipedia has it in a bunch of different default user templates. I will try to find some of the templates (like Template:User8 where we could make Special:CentralAuth more visible, instead of sending users to an outside tool). Sadads (talk) 23:14, 22 July 2015 (UTC)

Based on uselang=qqx (https://en.wikipedia.org/wiki/Special:Contributions/Sadads?uselang=qqx) an alternative placement on user contributions would be the currently blank MediaWiki:Contributions-summary which is displayed above the "Search for contributions" box at Special:Contributions/Sadads. That is probably *too* prominent. MediaWiki:Sp-contributions-explain is displayed inside the "Search for contributions" box but that place should probably only be used for a help link like now, or information about the fields in the box. The names "Contributions-summary" and "Sp-contributions-explain" also hint that the messages aren't intended for something like this. The French Wikipedia [27] (https://fr.wikipedia.org/wiki/Sp%C3%

A9cial:Contributions/Sadads?uselang=fr) uses MediaWiki:Sp-contributions-footer at the bottom like us. I think that is the best. PrimeHunter (talk) 23:42, 22 July 2015 (UTC)
@Sadads: The reason for linking to Meta's version of Special:CentralAuth instead of ours is that only Meta's version shows log entries for global account actions (like locking). Jackmcbarn (talk) 19:59, 23 July 2015 (UTC)

Sadads, for your purposes, I think that you will benefit from adding some popular user scripts to your account. Here's my list of recommendations, complete with the code you need to add them to your .js file:

```
// Gives some useful links on user, user talk, and user contribution pages
// by [[m:User:Hoo man]] <http://meta.wikimedia.org/wiki/User:Hoo_man/Scripts/Useful_links>
mw.loader.load('//meta.wikimedia.org/w/index.php?title=User:Hoo_man/useful_links.js&action=raw&ctype=text/javascript');
if(typeof(usefulLinksConfig) == 'undefined') usefulLinksConfig = {};
usefulLinksConfig.toolLinkMethod = 'p-cactions';

// Useful script to tell if a user is currently blocked, by striking out the username
mw.loader.load('//ru.wikipedia.org/w/index.php?title=MediaWiki:Gadget-markblocked.js&action=raw&ctype=text/javascript');

// [[File:userinfo.js]]
mw.loader.load('//en.wikipedia.org/w/index.php?title=User:PleaseStand/userinfo.js&action=raw&ctype=text/javascript');
```

Just paste that into m:User:Sadads/global.js and reload the page to run it. You will then have quick access to links on all user pages via the "More" dropdown menu that normally houses only "Move" (CentralAuth will be listed as "CA"), a visible indicator if any person is blocked (on any page, every time the user page is linked), and a list of user rights, account age, number of edits, and how recently the editor has edited on all user pages, immediately underneath the page title. Also, if you add it to your global.js page at Meta, it should work on all the Wikipedias. That should save you a bit of time. WhatamIdoing (talk) 19:08, 23 July 2015 (UTC)

Wikipedia:Dispute resolution noticeboard/request not working

The Wikipedia:Dispute resolution noticeboard/request page doesn't work for me. Previously I had that problem only with Firefox, but on my new computer this page doesn't work with Internet Explorer either. At the time I posted at Wikipedia_talk:Dispute_resolution_noticeboard/Archive_20#Technical_problem, and was advised to seek help here. Debresser (talk) 20:34, 22 July 2015 (UTC)

Wikipedia:Dispute resolution noticeboard/request requires JavaScript in your browser. The content is made by MediaWiki:Gadget-DRN-wizard.js which is enabled by default as "Form for filing disputes at the dispute resolution noticeboard" at Special:Preferences#mw-prefsection-gadgets. It's also enabled for unregistered users. If your browser doesn't have JavaScript or JavaScript is disabled then you see a blank page like in the mobile version https://en.m.wikipedia.org/wiki/Wikipedia:Dispute_resolution_noticeboard/request. PrimeHunter (talk) 20:51, 22 July 2015 (UTC)

But I do have Java 8 Update 51 installed on my computer. Debresser (talk) 20:56, 22 July 2015 (UTC)

Java and JavaScript are unrelated (yes, the similar names are confusing, not our fault). JavaScript comes with the browser but may be disabled. Do you have a [show]/[hide] to the right of "Contents" at #toc? That also requires JavaScript. PrimeHunter (talk) 21:03, 22 July 2015 (UTC)

Yes, I do have the [hide] option next to the TOC. Debresser (talk) 22:15, 22 July 2015 (UTC)

@Debresser: Can you check if the "Form for filing disputes at the dispute resolution noticeboard" option is enabled in your preferences under gadgets? - NQ (talk) 22:18, 22 July 2015 (UTC)

It wasn't, and now that I checked it, it works. Thanks. Debresser (talk) 18:08, 23 July 2015 (UTC)

Description of articles traffic rankings

I would just like to draw Wikipedians attention to the way the page view statistics are presented. In each and every Wikipedia page one would click 'View history', then 'Page view statistics' and he will get the statistics page. Now, the headline reads - '(Wikipedia page) has been viewed xxxxx times in the last 30 days. (and, for 10,000 Wikipedia pages, also -) This article ranked xxx in traffic on en.wikipedia.org.' One may get the wrong impression that the ranking is an all time one. As I understand it, the ranking is merely the page's rating for the month of March 2014 (see the figures here (http://stats.grok.se/en/top)). This can be a bit misleading.

Thanks a lot in advance for your help -- Limitless undying love (talk) 16:23, 23 July 2015 (UTC)

Yes, the rank is only for March 2014, for example for http://stats.grok.se/en/latest/Malaysia_Airlines_Flight_370 which disappeared that month and peaked in page views there. http://stats.grok.se is made and controlled by a single volunteer editor who can be contacted at User talk:Henrik but hasn't edited since August 2014. The editors of the English Wikipedia have decided to link the tool in page histories but it's not an "official" tool run by the Wikimedia Foundation which runs Wikipedia. PrimeHunter (talk) 16:43, 23 July 2015 (UTC)

Proposal to create PNG thumbnails of static GIF images

There is a proposal at the Commons Village Pump requesting feedback about the thumbnails of static GIF images: It states that static GIF files should have their thumbnails created in PNG. The advantages of PNG over GIF would be visible especially with GIF images using an alpha channel (compare the thumbnails on the side)

The thumbnail of this gif is of really bad quality.

This change would affect all wikis, so if you support/oppose or want to give general feedback/concerns, please post them to the proposal page. Thank you. --McZusatz (talk) & MediaWiki message delivery (talk) 05:07, 24 July 2015 (UTC)

Tag log

How a PNG thumb of this GIF would look like

Why is the tag log empty? Will it ever become nonempty? GeoffreyT2000 (talk) 00:51, 23 July 2015 (UTC)

The associated feature was not popular at Wikipedia:Village pump (technical)/Archive 136#Edit Tags. The English Wikipedia has no tags which can be manually added or removed, but I think admins could create such tags. The tag log would show when users added or removed the tags. Special:Log/tag is empty but fr:Special:Log/tag is not. French history pages like [28] (https://fr.wikipedia.org/w/index.php?title=Example&action=history&uselang=en) have buttons saying "Edit tags of selected revisions". Apparently the only such tag they currently have is called "Test balise" which means Test tag. I don't know whether they use it for anything other than just testing the feature. PrimeHunter (talk) 01:55, 23 July 2015 (UTC)

MediaWiki talk:Tag-OneClickArchiver#Protected edit request on 11 May 2015 was an RfC to see if there was consensus to set up tags for a particular userscript. --Redrose64 (talk) 15:27, 24 July 2015 (UTC)
The obtrusive UI issue should be addressed by [29] (https://gerrit.wikimedia.org/r/#/c/218353/), although we could also tweak common.css/js. Cenarium (talk) 17:39, 24 July 2015 (UTC)

Database problem

[30] (https://en.wikipedia.org/wiki/Special:AbuseFilter):

Tracked in Phabricator Task T106798

A database query error has occurred. This may indicate a bug in the software.

Function: IndexPager::buildQueryInfo (AbuseFilterPager)
Error: 2013 Lost connection to MySQL server during query (10.64.48.28)

This has been unavailable, at least to me, for some time.

All the best: *Rich Farmbrough*, 21:13, 23 July 2015 (UTC).

We're looking into this in the #wikimedia-operations channel. Legoktm (talk) 10:14, 24 July 2015 (UTC)
Should be fixed now, the problematic change was reverted. Legoktm (talk) 10:26, 24 July 2015 (UTC)

Thanks. All the best: *Rich Farmbrough*, 01:05, 25 July 2015 (UTC).

Page number weirdness

Check out Realm of Impossibility, look especially at the references. Although it seems I am using the same format for the cite tags throughout, some of them render the pages as if they are the "issues", with a colon. Any ideas? Maury Markowitz (talk) 16:51, 24 July 2015 (UTC)

You aren't exactly using the same format throughout. The two "colon-nized" citations are {{cite journal}} while the two that display the "p."-style are {{cite news}}.--Anders Feder (talk) 17:04, 24 July 2015 (UTC)

@Maury Markowitz: Yes, in a ref constructed as

```
{{cite journal |url=http://www.atarimagazines.com/rom/issue4/interview.php |title=Interview: Mike Edwards |journal=ROM Magazine |date=February/March 1984
|page=12 |first=Peter |last=Ellison}}
```

which displays as
Ellison, Peter (February/March 1984). "Interview: Mike Edwards" (http://www.atarimagazines.com/rom/issue4/interview.php). *ROM Magazine*: 12. Check date values in: |date= (help)
the template is {{cite journal}}, which for as long as I can remember (six years) has never displayed "p." or similar before page numbers. Note that issue numbers differ from page numbers in that they get parentheses -

```
{{cite journal |url=http://www.atarimagazines.com/rom/issue4/interview.php |title=Interview: Mike Edwards |journal=ROM Magazine |date=February/March 1984
|page=12 |first=Peter |last=Ellison |issue=12345 }}
```

displays as
Ellison, Peter (February/March 1984). "Interview: Mike Edwards" (http://www.atarimagazines.com/rom/issue4/interview.php). *ROM Magazine* (12345): 12. Check date values in: |date= (help)
It's explained at Template:Cite journal#csdoc_page, note that |journal= is an alias for |work=. You might like to fix those red errors by using |date=February–March 1984. --Redrose64 (talk) 19:41, 24 July 2015 (UTC)

So is the ":12" the expected outcome? It seems odd compared to what we want from an sfn. Is there any reason for this difference? Maury Markowitz (talk) 21:49, 24 July 2015 (UTC) I read the "In-source locations", and it seems to suggest p. 12 is the expected outcome? What am I missing here? Maury Markowitz (talk) 21:54, 24 July 2015 (UTC)

The things that are missing are |volume= and |issue=. {{cite journal}} is 'optimized' for academic journals which usually include both of those parameters:

```
{{cite journal |author=Author |title=Article title |journal=Prestigious Journal |volume=1 |issue=2 |page=25}}
```

Author. "Article title". *Prestigious Journal*. 1 (2): 25.

In the example, the numbering flows largest element to smallest and is consistent with how academic journals identify pages in a journal issue. For the case of *ROM Magazine*, the value in |url= shows that these citations are to articles in issue 4 (no volume that I can tell) so adding that bit of information to the template:

```
{{cite journal |url=http://www.atarimagazines.com/rom/issue4/interview.php |title=Interview: Mike Edwards |journal=ROM Magazine
|date=February/March 1984 |page=12 |first=Peter |last=Ellison}}</source> which displays as ::{{cite journal
|url=http://www.atarimagazines.com/rom/issue4/interview.php |title=Interview: Mike Edwards |journal=ROM Magazine |date=February
-March 1984 |issue=4 |page=12 |first=Peter |last=Ellison}}
```

Ellison, Peter (February–March 1984). "Interview: Mike Edwards" (http://www.atarimagazines.com/rom/issue4/interview.php). *ROM Magazine* (4): 12.

There has been some discussion at Help talk:Citation Style 1 about tweaking the way {{cite journal}} renders page numbers so that when |volume= and |issue= are not provided, Module:Citation/CS1 uses the p. and pp. prefixes.
—Trappist the monk (talk) 22:47, 24 July 2015 (UTC)

@Maury Markowitz: You say "what we want from an sfn" - but the article Realm of Impossibility doesn't use {{sfn}}. It uses Citation Style 1 templates wrapped in <ref>...</ref> tags. --Redrose64 (talk) 22:53, 24 July 2015 (UTC)

edit button not showing in mobile wikipedia

I think it's been a month since I have been seeing this kind of problem with mobile wikipedia version. I have seen seen this kind of problem in the beginning of this year but all went normal after some time but now the problem is as it is for about a month.

Now the problem is that the edit button (that looks like a pencil) is not coming on any mobile Wikipedia page (though the same thing is not happening in desktop mode). I am logged in with Wikipedia and I tried clearing cache but all was useless. Then I checked the same thing on other browsers n the same thing happened again on ucmini and ucbrowser and chrome also. After this I faced the same problem with another phone.

I think it's a bug in Wikipedia s mobile site. (Anyone else is facing the same problem??)

Plz Help!! -- चक्रपाणी (talk) 02:37, 20 July 2015 (UTC)

@चक्रपाणी: Are you using mobile interface beta? Check your settings. That might affect it. --Thnidu (talk) 03:31, 20 July 2015 (UTC)

Also, what kind of browser are you using when accessing the mobile website... Perhaps it doesn't support SVG or something, causing the images not to show ? —TheDJ (talk • contribs) 11:05, 20 July 2015 (UTC)

Surely SVG images are always converted to PNG server-side, because not all browsers support them? For example, this image appears in the `<img />` tag as `src="//upload.wikimedia.org/wikipedia/commons/thumb/8/84/Example.svg/20px-Example.svg.png"` --Redrose64 (talk) 16:47, 20 July 2015 (UTC)

The pencil icon that forms the "edit" links in the mobile skin is actually directly shown as an SVG for me (Chrome). It is added by the following CSS rule, which seems to be added to the page using JavaScript (it is not visible using "View source", only with the "Inspect element" tool):

```
.mw-ui-icon-edit-enabled:before {
 background-image: url(//en.wikipedia.org/w/load.php?modules=skins.minerva.icons.images.js&image=edit-enabled&format=rasterized&/* ... */);
 background-image: -webkit-linear-gradient(transparent,transparent),url(data:image/svg+xml,/* ... */);
 background-image: linear-gradient(transparent,transparent),url(data:image/svg+xml,/* ... */);
 background-image: -o-linear-gradient(transparent,transparent),url(//en.wikipedia.org/w/load.php?modules=skins.minerva.icons.images.js&image=edit-
enabled&format=rasterized&/* ... */);
}
```

I'm guessing the `background-image` property is set to four different values to let it fall back to the last working one if any variation is not supported by the used browser. An SVG file with code directly embedded on the page (replaced with a comment here) is used in two cases, and a PNG version from a different page in the other two. $SiBr_4$ (talk) 17:51, 20 July 2015 (UTC)

Please let me know your device, operating system and browser (including version) that you are using and I will look into this. Jdlrobson (talk) 00:18, 26 July 2015 (UTC)

photos in an info box

Did something change the ability to load photos in an info box? Either they are HUGE, or if you put pixel or thumb notices they do not show up at all? Examples: Ana Rosa Tornero or María Rivera Urquieta. SusunW (talk) 22:30, 24 July 2015 (UTC)

Tracked in Phabricator Task T106895 RESOLVED

Issue with all the recently uploaded files. See Special:NewFiles -NQ-Alt (talk) 22:33, 24 July 2015 (UTC)

Tracked in Phabricator Task T106740 OPEN

So what does that mean? We no longer can add photos? Someone is working on the problem? I should not upload the other 4 I wanted to because they won't work either? SusunW (talk) 22:37, 24 July 2015 (UTC)

At the moment it appears that there is some sort of thumbnail generation problem on en.wikipedia.org for recently uploaded files (you saw yourself that displaying full size images is fine). German Wiki etc and Commons do not appear to have the same issue. You can continue to add photos. You can continue to upload them. We don't know if anyone is working on the problem. Also, you may find it useful to follow the {{infobox person}} template instructions and change the syntax of your image additions in the infobox (and add the image_size parameter) so that they will display properly when the problem is resolved or passes. E.g. change

```
| image       = [[File:Ana_Rosa_Tornero.jpg|thumb|right|Ana Rosa Tornero in ''[[Wara Wara]]'' (1930)]]
| alt         =
| caption     =
```

to

```
| image       = Ana_Rosa_Tornero.jpg
| image_size  = 200px
| alt         =
| caption     = Ana Rosa Tornero in ''[[Wara Wara]]'' (1930)
```

(image size is a guess at what may be suitable) Hopefully tech are looking at the underlying issue already. Nanonic (talk) 23:03, 24 July 2015 (UTC)

FYI, I added two existing Tasks that seemed relevant to this issue. -- George Orwell III (talk) 23:16, 24 July 2015 (UTC)
with or without the thumb notices they don't work. If you leave off "file" all together, the photo took up almost the entire page. And I see that RedRose was trying to get it to work and now it shows no image even without "file". No point in spending a lot of time with "what if" parameters. They may not need any adjustments if they fix the problem. If I upload fair use images, and they don't fix the problem quickly I'll have to do it again because they'll be deleted for not being tied to a file within 7 days. I'll just save the links and see if it works later. Thanks! SusunW (talk) 23:23, 24 July 2015 (UTC)

As an additional data point, seen here, some **|upright** values work while others break the thumb. My default thumb size is 220px; changing it to 300px causes different upright values to break the thumb. This image was uploaded in 2007. –*Mandruss* ☎ 04:11, 25 July 2015 (UTC)

Added tests of fixed px values. For me, 300, 270, 250, 240, 220, 200, 180, 150, and 120 work; 290, 280, 260, 230, 210, 190, 170, 160, 140, and 130 do not. But File:Sands Hotel 1950s.jpg fails at 250, so apparently all of that is image-dependent. –*Mandruss* ☎ 09:13, 25 July 2015 (UTC)

Don't think it is only enwiki problem. We (at lvwiki) don't see images, too. --Edgars2007 (talk/contribs) 09:11, 25 July 2015 (UTC)

Somebody get it fixed asap!! ♦ *Dr. Blofeld* 10:11, 25 July 2015 (UTC)

Appears to be fixed now. –*Mandruss* ☎ 10:34, 25 July 2015 (UTC)

Thumbnail not displaying properly

The following discussion is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

SELF-TROUT:
D'oh! Didn't see the thread right above this one. (non-admin closure) Erpert blah, blah, blah... 04:18, 25 July 2015 (UTC)

I have tried twice to upload an album cover to the article *Chicago '85... The Movie*. While the image (https://en.wikipedia.org/wiki/Chicago_%2785..._The_Movie#/media/File:Dave_Hollister_-_Chicago_85_the_Movie_album_cover.jpg) and its fair use information does show up properly when it is clicked on, the thumbnail itself appears broken in the article. Anyone know what could be wrong? Erpert blah, blah, blah... 04:17, 25 July 2015 (UTC)

The discussion above is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

Infobox image

Can somebody figure out why the image in Sands Hotel and Casino won't appear? ♦ *Dr. Blofeld* 08:11, 25 July 2015 (UTC)

> Very possibly related to one or more of the three image-related threads immediately preceding this one. —*Mandruss* ☎ 08:17, 25 July 2015 (UTC)
>
> > I thought originally it was a glitch with one of the photos I uploaded so I tried another one and it did the same thing! It won't even show when not in the infobox either. Strange. Hope it's fixed soon! ♦ *Dr. Blofeld* 08:38, 25 July 2015 (UTC)

Article creation improvements by WMF

After WMF refused to implement the autoconfirmed account creation trial, they promised that in lieu of it, there would be better tools for new page patrollers and users creating articles. Page Curation was launched three years ago, but it doesn't seem like any progress on Wikipedia article creation has been made since 2013. Is that still a project under development by WMF? **Conifer** (talk) 11:54, 25 July 2015 (UTC)

Table formatting List_of_mayors_of_Bremen

Hello, I could use some help with table formatting in List_of_mayors_of_Bremen please. It's a bit hard to describe, but I'll try: Look at the last 2 rows about Jens Böhrnsen and the new mayor Carsten Sieling. The 2nd Mayor Karoline Linnert will probably stay in office (per the German senate site). So I would like to display Karoline Linnert's data in 2-row columns stretching over both 1st Mayors (as she was 2nd Mayor with *both* 1st Mayors). If A are all columns for a 1st Mayor (the first 4 columns) and B are all columns for a 2nd Mayor (the last 4 columns), I want something like:

- A1 B1
- A1 B2
- A2 B2, where A1 and B2 should only be 4 single boxes stretching over 2 rows. I have tried fiddling with rowspan (in preview mode) but failed. Are such alternating, "overlapping" 2-row columns possible in Wiki-tables? GermanJoe (talk) 15:17, 25 July 2015 (UTC)

> @GermanJoe: like this (https://en.wikipedia.org/w/index.php?title=List_of_mayors_of_Bremen&type=revision&diff=673029059&oldid=673022991) ? I've kept the split on the last row intact, but it can be easily removed as well. —TheDJ (talk • contribs) 15:34, 25 July 2015 (UTC)
>
> > @TheDJ: I think, the problem occurred when I tried to remove the last split as well (sorry for missing that detail). I simply removed the last empty cell descriptions, and then the formatting broke somehow. Would you mind removing the last split too please (the last 4 small empty cells)? Aside from that, it is exactly what I looked for. GermanJoe (talk) 15:41, 25 July 2015 (UTC)
> > (tested again). Deleting the last 4 cell elements is indeed my problem. It breaks the display of Karoline Linnert's cells - they no longer extend across 2 rows afterwards. GermanJoe (talk) 16:17, 25 July 2015 (UTC)
> >
> > > What you've got here is essentially the same problem as at Help talk:Table/Archive 5#Governors-General of the Philippines and Help talk:Table#uneven rowspans. Browsers will display a row (including a spanned row) only as high as it needs to be. To get the spacing right you need at least one column which has no rowspans. See Template:Rail line three to two which looks like it has four rows; in fact there are six, columns 2 & 4 (out of 5) each contain six unspanned cells. --Redrose64 (talk) 16:34, 25 July 2015 (UTC)
> > >
> > > > Yeah, that worked (I added the pseudo column at the beginning just to be sure). Many thanks to both of you for your help. GermanJoe (talk) 17:57, 25 July 2015 (UTC)

Navboxes in mobile

Navboxes do not show up on the mobile Wikipedia. GeoffreyT2000 (talk) 23:04, 25 July 2015 (UTC)

> I don't think they ever did. It's down to the classes associated with the template. --Redrose64 (talk) 23:15, 25 July 2015 (UTC)
>
> > *Confirmed.*— Win 8.1 / IE 11. Compared...
> >
> > - https://en.wikipedia.org/wiki/Angela_Eagle#External_links (desktop view)
> > - https://en.wikipedia.org/w/index.php?title=Angela_Eagle&useskin=minerva#External_links (desktop view with Minerva skin applied)
> > - https://en.m.wikipedia.org/wiki/Angela_Eagle#External_links (mobile mode)
> >
> > ... and the Navboxes only showed in desktop mode. I know navboxes are not suppose to "appear" when printed out (class="noprint") but I doubt that has anything to do with them not rendering in 'Mobile Mode' (if they ever did that is as Redrose64 pointed out). -- George Orwell III (talk) 23:28, 25 July 2015 (UTC)

It's by purpose. The majority of navboxes do not render well at by 320px and unfortunately our current tech stack (templates) doesn't make it possible to style them differently on mobile and desktop. Given that on desktop the HTML markup associated with navboxes are huge it would be a great idea for us to all rethink them (possibly using JavaScript make them more interactive). Jdlrobson (talk) 00:22, 26 July 2015 (UTC)

Display oddity

In Firefox 39 (current production version) in Win 7, the following, which uses <code><pre><nowiki>, nested in that order,:

```
| foo
| bar
```

displays with the first line indented by about half an em. I'm sure this hadn't used to be the case. Andy Mabbett (Pigsonthewing); Talk to Andy; Andy's edits 12:57, 26 July 2015 (UTC)

> Confirmed; same config. —*Mandruss* ☎ 13:01, 26 July 2015 (UTC)
> Me too, also same config. --Izno (talk) 13:11, 26 July 2015 (UTC)

Usage

Don't nest <pre> in <code>. What is the point? -- [[User:Edokter]] {{talk}} 14:01, 26 July 2015 (UTC)

Without <code>:

```
| foo
| bar
```

Without <pre>:

| foo | bar

So omitting <code> gives the desired display. But looses the semantic meaning. Andy Mabbett (Pigsonthewing); Talk to Andy; Andy's edits 14:06, 26 July 2015 (UTC)

> @Pigsonthewing: Why is the semantic meaning lost? the `code` element (http://www.w3.org/TR/html5/text-level-semantics.html#the-code-element) "represents a fragment of computer code"; whereas the `pre` element (http://www.w3.org/TR/html5/grouping-content.html#the-pre-element) "represents a block of preformatted text" such as "fragments of computer code". --Redrose64 (talk) 15:51, 26 July 2015 (UTC)
>
> <pre> is block, <code> is inline. I keep explaining you cannot nest block inside inline elements (despite what HTML5 allows); MediaWiki (Read: HTML Tidy) does not allow that. Also, <pre> invokes <nowiki> by default, so no need to specify that, unless you want to actually show the tag. -- [[User:Edokter]] {{talk}} 15:52, 26 July 2015 (UTC)

Problem on WikiProject Physics page

Hi. There is an odd problem on Wikipedia:WikiProject Physics. Some flag colored yellow, red and green floats over the content in the "Current status of physics articles" section. I am using Google Chrome 44 under Windows 8 with Vector skin. --Meno25 (talk) 15:24, 26 July 2015 (UTC)

> The displayed file is File:New.png, which was already included on the page but was recently overwritten with an image of the flag of Syrian Kurdistan by Iraqi man10. I've reverted the image. SiBr4 (talk) 15:31, 26 July 2015 (UTC)
>
> @SiBr4: Thank you for the quick help. --Meno25 (talk) 15:39, 26 July 2015 (UTC)

Screenshot of the problem

Move shown twice

Why is Michael Hardy's move from Uncorrelated to Uncorrelated random variables shown twice in the move log? GeoffreyT2000 (talk) 01:27, 23 July 2015 (UTC)

> The log is [31] (https://en.wikipedia.org/w/index.php?title=Special%3ALog&page=Uncorrelated). The page history of the target [32] (https://en.wikipedia.org/w/index.php?title=Uncorrelated_random_variables&action=history) also shows it twice, with consecutive page revisions 669814688 and 669814689. I guess it's just a glitch. PrimeHunter (talk) 02:05, 23 July 2015 (UTC)
>
> "Just a glitch" is usually what comes back in a few months or years as "OMG, I wish we'd fixed that problem when we first noticed a symptom. All the best: *Rich Farmbrough*, 00:27, 27 July 2015 (UTC).

DIFF highlighting glitches ?

Not sure if this is a glitch or not. My personel opinion is that this is a logic flaw in the diff engine.

GLITCH DESCRIPTION:

> Adding an inline citation immediately after an existing inline citation results in a flawed diff highlight as follows:

ORIGINAL:

- article text blah blah.<ref>{{cite ...|ref=ALPHA}}</ref>

EDITED:

- article text blah blah.<ref>{{cite ...|ref=ALPHA}}</ref><ref>{{cite ...|ref=OMEGA}}</ref>

INTUITIVELY WHAT SHOULD BE HIGHLIGHTED:

- article text blah blah.<ref>{{cite ...|ref=ALPHA}}</ref><ref>{{cite ...|ref=OMEGA}}</ref>

REALITY OF WHAT IS HIGHLIGHTED:

- article text blah blah.<ref>{{cite ...|ref=ALPHA}}</ref><ref>{{cite ...|ref=OMEGA}}</ref>

ALTERNATELY -- This is also true if you insert a new citation in the middle of multiple existing citations:

- article text blah blah.<ref>{{cite ...|ref=ALPHA}}</ref><ref>{{cite ...|ref=OMEGA}}</ref><ref>{{cite ...|ref=DELTA}}</ref>

HOWEVER -- The diff highlighting is correct if you insert the new citation *BEFORE* any existing citation(s).

- article text blah blah.<ref>{{cite ...|ref=OMEGA}}</ref><ref>{{cite ...|ref=ALPHA}}</ref>

Could this be some issue with giving parse priority to template tags (curly braces) maybe ? 172.68.146.9 (talk) 13:50, 23 July 2015 (UTC)

> Diff isn't an exact science, and in my experience Wikipedia's implementation frequently highlights unintuitively. I wouldn't hold much hope that this can be resolved, except by grafting on some ugly hack, which probably would not be well-received by the developers.--Anders Feder (talk) 15:39, 23 July 2015 (UTC)
> Registered users can enable wikEdDiff at Special:Preferences#mw-prefsection-gadgets. wikEdDiff often gives a better diff when the default diff is poor. PrimeHunter (talk) 16:48, 23 July 2015 (UTC)
>
> > I believe that the diff feature starts at the right-hand end of the line and works towards the left. So when adding a `<ref>...</ref>` after one that is already there, the diff feature notices that each line ends with a `</ref>` which is considered to match. Then as it works its way to the left, it will come across another `</ref>` on the later version which is considered not to match, so it gets the highlight. --Redrose64 (talk) 08:37, 24 July 2015 (UTC)
>
> I actually can't see what you mean in the above example. A smarter diff would certainly recognise where delimiters package entities and match for example "<<bar>> " instead of "bar>> <<" in the following, assuming that "<<bar >>" had been inserted.
>
> foo <<bar>> <<baz>>
>
> All the best: *Rich Farmbrough*, 00:33, 27 July 2015 (UTC)

A tool for dimming references in diffs?

Hi there, is anyone aware of any tool (or possibly willing to make one) that would dim/remove references from view upon command while looking at a diff?

In these edits (https://en.wikipedia.org/w/index.php?title=Balochistan%2C_Pakistan&type=revision&diff=672896078&oldid=672191451) I'm interested in seeing what the net change to the prose was in the big blue block, but with all the long references, it's difficult to read the prose and see what changed. I'm thinking of something like WikEdDiff that, with a press of a button, would turn anything between <ref></ref> tags light gray or something. Naturally this would have no effect on the article. It would only affect what a user sees on their screen. Thanks. Cyphoidbomb (talk) 16:34, 24 July 2015 (UTC)

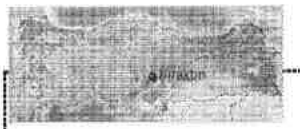
Was thinking about related script (feel free to move this to new section). When somebody moves a text lower for some rows **and** makes some other changes in that text, in diff window you can't see what really has been done in that text, because it only shows you that whole paragraph or so has been moved somewhere else (specifically, some rows up or down). I think I'm not the only one, who would like to know, what else has been done in that edit. I could really live with the fact, that I can't see in diff window, that text has been moved, just show me the real changes in the text. Mission impossible (such script)? --Edgars2007 (talk/contribs) 10:19, 25 July 2015 (UTC)

> @Edgars2007: wikEdDiff, linked above by Cyphoidbomb, is the solution to that problem. —*Mandruss* ☎ 10:26, 25 July 2015 (UTC)
>
> > Thanks! It looks like we will be good friends with wikEdDiff :) --Edgars2007 (talk/contribs) 10:45, 25 July 2015 (UTC)
>
> WP:LDR is you friend. All the best: *Rich Farmbrough*, 00:45, 27 July 2015 (UTC)

Coordinate Templates and maps

Hi all, I mostly translate articles from other wikis into English and templates are the bane of my life. I was wondering if anyone could explain how I should perform a particular translation. The German version of the **coord** template can be made to produce a map (see an example here: de:Felsrelief von Firaktin). Could anyone suggest how I might get the same effect on English wikipedia? Cheers, Furius (talk) 13:40, 25 July 2015 (UTC)

> @Furius: I don't know of anything on en-wiki that has coding similar to that "Coordinate" template, but see if `{{Location map}}` will suffice for your purposes. —*Mandruss* ☎ 15:05, 25 July 2015 (UTC)
> The template is de:Vorlage:Coordinate which passes most of its parameters through to either of two subtemplates: de:Vorlage:CoordinateComplex or de:Vorlage:CoordinateSimple. --Redrose64 (talk) 15:06, 25 July 2015 (UTC)



Location of Firaktin in Turkey

The German code is `{{Coordinate |article=/ |map=left |maptype=relief |name=Firaktin |NS=38/16/18/N |EW=35/37/54/E |type=landmark |region=TR-38 }}`. A similar map can be made by {{Location map}} with some work. Here I used:

```
{{Location map | Turkey
| float = left
| relief = yes
| label=Firaktin
| caption = Location of Firaktin in [[Turkey]]
| lat_deg = 38 | lat_min = 16 | lat_sec = 18 | lat_dir = N
| lon_deg = 35 | lon_min = 37 | lon_sec = 54 | lon_dir = E
}}
```

> Some infoboxes like {{Infobox settlement}} can also be coded to include a map. PrimeHunter (talk) 12:42, 26 July 2015 (UTC)

'Abd al-'Aziz al-Wafa'i

Should the title of 'Abd al-'Aziz al-Wafa'i start with ', or should it be "Abd al-'Aziz al-Wafa'i" instead?--DThomsen8 (talk) 15:02, 25 July 2015 (UTC)

> I think that the initial apostrophe is OK. We do have 's-Hertogenbosch after all. --Redrose64 (talk) 15:09, 25 July 2015 (UTC)

I went through 's-Hertogenbosch on my way to Liège many years ago, but the name has nothing in common with Arabic names. I will take your advice on that name.--DThomsen8 (talk) 22:32, 26 July 2015 (UTC)

The apostrophe is the transliteration of ayn. Please see WP:MOSAR. Alakzi (talk) 00:55, 27 July 2015 (UTC)

Nonexistent existing userpage

User:Green Giant doesn't exist: I get the little "View or restore 4 deleted edits?" link at the top, just as I do with a previously deleted page that doesn't currently exist, and the tabs at top say "Create this page" and "undelete 4 edits" instead of "edit this page" and "history". All very nice, except for the obvious fact that the page exists! All the links go directly to Commons, and it's identical to Commons:User:Green Giant. Has some software weirdness happened, or am I just unaware of a new feature/bug that transcludes your Commons userpage if your en:wp userpage doesn't exist? Nyttend (talk) 15:35, 26 July 2015 (UTC)

See the note at the bottom of the page: "What you see on this page was copied from //meta.wikimedia.org/wiki/User:Green_Giant." It is a global user page.--Anders Feder (talk) 15:38, 26 July 2015 (UTC)

Hehe, aye it is the global user page; if you don't have a user page on any particular wiki, it will show whatever is on your Meta user page, but your user talk pages are not affected. When it became operational, I requested deletion of several of my user pages on other wikis because I want to migrate from being a Wikipedian to being a Wikimedian. The ENWP and Commons user pages were more complex so I didn't get round to getting them deleted until yesterday. I believe there will also be global notifications and watch-lists in the future. Green Giant (talk) 19:47, 26 July 2015 (UTC)

Wikipedia File Upload Wizard (Fair use files)

I would like to ask two questions about the Wizard, specifically the non-free files form, and more specifically: Step 3 > *"This is a copyrighted, non-free work, but I believe it is Fair Use."* > *"This is an historic portrait of a person no longer alive."*

- A) Why am I able to confirm that *"the image will be shown as a primary means of visual identification at the top of the article dedicated to the person in question"* by simply ticking a box, but I have to write a text 'explaining" that *"a free alternative to this image cannot be found"*, *"our use of the file will not harm any commercial opportunities of its owner"* and *"the use of this file will be minimal"*? The third one, especially, could be very easily replaced by a box-to-tick.
- B) Why is the Wizard sequence **Free alternative-Commercial opportunities-Minimal use** while the Non-free rationale template sequence is **Free alternative-Minimal use-Commercial opportunities**?

Thank you in advance.

--The Traditionalist (talk) 17:12, 26 July 2015 (UTC)

I suppose you can always write "✓" in the text box. But I think our fair use policy could do with a slightly more liberal overhaul. All the best: *Rich Farmbrough*, 12:46, 27 July 2015 (UTC).

I always add the same text. It looks, however, like an attempt to train editors to write creatively, which would be laughable
Could a template editor fix what I address at my question B)? It is most likely a mistake.--The Traditionalist (talk) 13:52, 27 July 2015 (UTC)

Search page meddling

Recently the Wiki search page has been spoilt, by adding a migraine headache in the search box and a promiscuous drop down menu. I've tried to stop it by altering the settings in Preferences to no avail, can anyone help please.Keith-264 (talk) 11:58, 27 July 2015 (UTC)

Which skin are you using? (It seems relatively unchanged to me, though I see from comments above it has some regex ability.) All the best: *Rich Farmbrough*, 12:43, 27 July 2015 (UTC).

Monobook ([33] (https://en.wikipedia.org/wiki/Special:Preferences#mw-prefsection-rendering))Keith-264 (talk) 13:02, 27 July 2015 (UTC)

Oh you mean the search box? Yes that dropdown that hides the search key is sometimes irritating, but also sometimes useful. The Migraine - maybe someone can offer some css to turn that off? All the best: *Rich Farmbrough*, 13:29, 27 July 2015 (UTC).

The post is about the big search box at Special:Search and not the smaller box on all pages. In Firefox, each time I type a character in the big box I briefly see annoying flickering tilted grey lines in the whole box. The drop-down with search suggestions is bigger and more attention seeking than for the small box, and it covers "Multimedia Everything Advanced" so I have to click somewhere else on the page before I can use those options. "Disable the suggestions dropdown-lists of the search fields" at Special:Preferences#mw-prefsection-gadgets works on the small box but not the big. PrimeHunter (talk) 13:34, 27 July 2015 (UTC)

It's not the little one at the side of the article page (that's annoying enough) but the big one on the page that the search goes to if there isn't an obvious wikipage. The drop down is never useful and the pattern in the box comes from the imagination of a !"£$%^&*(). Don't the people who do these things ever ask first?Keith-264 (talk) 13:38, 27 July 2015 (UTC)

You can use...

```
#searchText.oo-ui-pendingElement-pending input {
    background-image: none;
}
```

...to hide the scrolling stripes of doom, and...

```
#searchText > div {
    display: none;
}
```

...to hide the glamorous dropdown. Alakzi (talk) 14:01, 27 July 2015 (UTC)

Thanks but where do I put them....Keith-264 (talk) 14:09, 27 July 2015 (UTC)

your CSS. PrimeHunter (talk) 14:15, 27 July 2015 (UTC)

Thanks, I didn't know I had one, it seems to be working.Keith-264 (talk) 14:19, 27 July 2015 (UTC)

- Is there any way to disable this feature altogether? For eg. if I intend to search in Special:Search for *Lorem*, it fills in *Lorem Ipsum* by default when I press the Enter key to search. I'd have to press the Tab key and then the Enter key every time to get the desired search query which is really cumbersome. NQ-Alt (talk) 14:20, 27 July 2015 (UTC)

I don't suppose Wiki will force anyone who changes it to add a "change it back" button to every change? That *would* be useful.Keith-264 (talk) 14:25, 27 July 2015 (UTC)

- Fixed T106273 - NQ-Alt (talk) 01:35, 28 July 2015 (UTC)

Tool for fixing malformed wikilinks

Is there a tool that can quickly convert multiple instances of [https://en.wikipedia.org/wiki/Article_title Article title] to proper wikilinks? (Asking again here as I got no joy at the HD.) Roger (Dodger67) (talk) 13:33, 27 July 2015 (UTC)

I guess WP:AWB could do it, but I just resort to doing it manually - it's a pet peeve of mine too (among many).--ukexpat (talk) 13:56, 27 July 2015 (UTC)

It sounds like user:js/urldecoder does what you want. You could also use the find and replace tool on the right side of the Advanced menu of the edit toolbar: paste `\[{{?:https?:)?\/\/en\.wikipedia\.org\/wiki\/.+?) (.+?)\]` into the search box and `{{subst:u2wl1=$1|2=$2}}` into the replace box, and check "Treat search string as a regular expression", and click "Replace all". This replaces all external Wikipedia links with {{subst:urltowiki}}, which converts them on page save. /~huesatlum/ 14:22, 27 July 2015 (UTC)

Thanks, I've installed and tested urldecoder - like it's "one click and it's done" operation. Roger (Dodger67) (talk) 18:34, 27 July 2015 (UTC)

At one time it was possible to use Special:LinkSearch and enter `en.wikipedia.org` - that facility was removed last year. LinkSearch has recently been altered again, so that it assumes http:// unless you explicitly specify https:// which means that when hunting down spam links, you now need to do twice as many searches as you used to. --Redrose64 (talk) 18:54, 27 July 2015 (UTC)

Tech News: 2015-31

Latest tech news from the Wikimedia technical community. Please tell other users about these changes. Not all changes will affect you. Translations are available.

Recent changes

- You can now use redirects to link to JavaScript pages. [34] (https://phabricator.wikimedia.org/T73200)
- You cannot use the compact user bar any more. [35] (https://phabricator.wikimedia.org/T104659)
- You can change your options to see bigger images. [36] (https://phabricator.wikimedia.org/T65440)
- You can watch short videos about how to use VisualEditor. [37] (https://www.mediawiki.org/wiki/VisualEditor/GIFs)
- You can now edit pages linked in "what links here" more easily. [38] (https://phabricator.wikimedia.org/T97269)

Problems

- There was a problem with some Lua modules on July 22 and 23. Some pages using them did not list them in "what links here". You can fix those pages with a null edit: edit and save the page without making any change.
- There was a problem with the abuse filter page on big wikis on July 23. It was due to a code error. [39] (https://phabricator.wikimedia.org/T106798)

Changes this week

- The new version of MediaWiki will be on test wikis and MediaWiki.org from July 28. It will be on non-Wikipedia wikis from July 29. It will be on all Wikipedias from July 30 (calendar).
- JavaScript authors: You cannot use `wgNoticeUserData` to get edit counts anymore. [40] (https://phabricator.wikimedia.org/T85984)

Meetings

- You can join the next meeting with the VisualEditor team. During the meeting, you can tell developers which bugs are the most important. The meeting will be on July 28 at 19:00 (UTC) (http://www.timeanddate.com/worldclock/fixedtime.html?hour=19&min=00&sec=0&day=28&month=07&year=2015). See how to join.

Tech news prepared by tech ambassadors and posted by bot • Contribute • Translate • Get help • Give feedback • Subscribe or unsubscribe

15:05, 27 July 2015 (UTC)

Single characters

Why do the history pages for pages with only a single character in them show as "index" in the browser history? GeoffreyT2000 (talk) 17:03, 27 July 2015 (UTC)

That would be an odd browser feature. Which url did you visit, what is the browser, exactly what does it say in the browser history, and what does it say for a page with more characters like Example? PrimeHunter (talk) 17:13, 27 July 2015 (UTC)

history revision statistics (alternate tool)

hi, ive posted on this topic here before Wikipedia:Village_pump_(technical)/Archive_136#revision history statistics "link" on this matter (as well as the maintainers talk) *however*, (since it seems we are getting nowhere) after a month or more of the tool being down/not working...id like to know if there is an alternate tool (for *revision history statistics*) that gives the **same** information?,,thank you--Ozzie10aaaa (talk) 16:23, 14 July 2015 (UTC)

Yes, there is, but it only works on the German Wikipedia for the time being. I'm working on making it work for the english wikipedia, and should have it up and running soon. I will post a link once I have it running.—[illegible] 14:17, 18 July 2015 (UTC)

thank you--Ozzie10aaaa (talk) 23:38, 20 July 2015 (UTC)

C678, WikiHistory is a poor and rudimentary replacement for the history revision statistics page. As a stopgap it is fine, better to have something than nothing, but work should continue to get the original tool up-and-running again. The process and communication regarding the status of this tool is subpar.--Wolbo (talk) 11:26, 22 July 2015 (UTC)

@Wolbo: You are free to continue work on getting the original tool up-and-running again - you can find the source code here (https://github.com/x-Tools/xtools/). Other editors have been collaborating to that end at Wikipedia talk:XTools. In the meantime, C678 may dispense his time the way he find most beneficial. If you need anything beyond that you should instead direct your comments to the m:Community Tech team.--Anders Feder (talk) 12:25, 22 July 2015 (UTC)
I'm sorry Wolbo, but there is only so much I can do. For now this will have to do until we can get the code rewritten.—[redacted] 14:23, 23 July 2015 (UTC)

[redacted] I saw the new link at Ebola_virus_epidemic_in_West_Africa very impressive, **congrats**...(BTW will the individuals "bytes" be shown as before?) thank you again--Ozzie10aaaa (talk) 13:47, 23 July 2015 (UTC)

This is a different tool that we borrowed for now. We are still working on bringing the original code back up.—[redacted] 14:23, 23 July 2015 (UTC)

Here is a sample link to get to the substitute tool for articleinfo: it's called Wikihistory. I found that you have to capitalize the first letter of the article name, and substitute underscores '_' for blanks. http://tools.wmflabs.org/xtools/wikihistory/wh.php?page_title=Software_agent Thank you for your work, Cyberpower678. --Ancheta Wis (talk | contribs) 15:24, 26 July 2015 (UTC)

Good to read that work is proceeding to bring the 'history revision statistics' back to live (that wasn't entirely clear from the communications I read on the subject). Opinions may differ but in my view both the information it contained as well as the way the information was displayed was excellent so hopefully the attempts to get it going again will be successfull.--Wolbo (talk) 22:32, 28 July 2015 (UTC)

Private drafts?

Where are we at with private drafts? Tonight I found myself in need of such a thing (long unfinished reply to a contentious conversation, didn't want to save it in userspace, ended up saving it in an offline text editor). I found this discussion (https://en.wikipedia.org/wiki/Wikipedia:Village_pump_%28proposals%29/Archive_94#Allowing_users_to_keep_private_drafts_of_their_work) from a few years ago, which pointed to this bug (https://phabricator.wikimedia.org/T39992), which looks like it got mired in a bunch of what-if navel gazing. Did this ever get enabled, and if so, how do I use it? Regards, Orange Suede Sofa (talk) 04:23, 21 July 2015 (UTC)

I actually agree that such a thing would be useful, and really a standard part of any modern web-based authoring interface. But, at the same time, the concerns raised in the bug are legitimate. The lack of eyeballs on anything "private" means it is easy for malicious users to abuse.--Anders Feder (talk) 09:53, 21 July 2015 (UTC)

How would we keep people from using this namespace to store all kinds of non-encyclopedic crap? – SMcCandlish ☺ ☎ ¢ ≽^•⊕•^≼ 16:06, 22 July 2015 (UTC)

Indeed. I think a client-based solution, as suggested in the bug, would be better than nothing, though.--Anders Feder (talk) 16:08, 22 July 2015 (UTC)

If I have a page that absolutely nobody else can see (and therefore there are no "eyeballs" on it), then how exactly would I go about using it for abuse? I'm trying to understand what "abuse" means when absolutely nobody except me can see the contents of the page. Whatamidoing (talk) 16:11, 22 July 2015 (UTC)

Nobody could see it without a password. People wishing to abuse the system could share passwords. AndyTheGrump (talk) 16:17, 22 July 2015 (UTC)

Sharing the password to my whole account? I don't think that's a realistic scenario. For that matter, there are already opportunities for doing that, at least for short messages. If I wanted to leave you a secret message in my account, I could type it into the sig field in prefs, and then give you the password. But I really can't imagine anyone wanting to go to that much trouble, when free private web boards are so easy to get. Or I could get a free e-mail account from any one of thousands of providers, and do the same thing by saving a draft of an e-mail message. This does not sound like a significant problem. Whatamidoing (talk) 18:38, 22 July 2015 (UTC)

Just to be clear: If you "can't imagine" any problems, is anything here preventing you from forwarding the request?--Anders Feder (talk) 18:59, 22 July 2015 (UTC)

It depends on the exact implementation, but just consider resource consumption: some sad pubescent kid could probably find great pleasure in making a program that automatically created 700 yottabyte worth of drafts featuring ASCII art lolcats, just for the sake of crashing Wikimedia servers.--Anders Feder (talk) 16:18, 22 July 2015 (UTC)

I suspect that $wgMaxArticleSize would prevent that. Whatamidoing (talk) 18:38, 22 July 2015 (UTC)

Really? Do you also have a constant to limit the number of drafts created by each user accounts? And once you have, do you also have a constant to limit the number of user accounts created by each physical person? If so, how do you plan on enforcing it?--Anders Feder (talk) 18:49, 22 July 2015 (UTC)

This particular possible issue aside, wouldn't this just make loads of work for the WMF, who would be the only ones able to patrol these pages for issues related to problematic content on their servers? Sam Walton (talk) 19:02, 22 July 2015 (UTC)

Yes. Unless Whatamidoing has some magic solution I am as of yet oblivious to.--Anders Feder (talk) 19:10, 22 July 2015 (UTC)

The proposal is for a single draft (per account), not for an infinite number. There is already a system in place that limits the size of a page. There is already a system in place that limits the number of accounts you can create (per computer/IP address during a given time period). Your scenario is definitely implausible.
If **nobody can see it** except the one logged-in user, then why would the page need to be patrolled in the first place? This feels like "we have to make a note about what color the invisible unicorn is, because we check the color of every animal that can be seen in public!" Well, yes: we do try to check the content of every page that is visible to the public. But this one would not be visible to the public, so why should the visible-to-the-public rules need to be applied to an invisible-to-the-public page? Whatamidoing (talk) 18:51, 23 July 2015 (UTC)

> "The proposal is for a single draft (per account)". Lies. There is no such proposal. And if all concerns raised about the suggestion are "definitely implausible" anway, what are you waiting for? Why don't you go ahead and prod the engineering team to implement it instead of arguing with people here who have no power over the installation anyway?--Anders Feder (talk) 19:38, 23 July 2015 (UTC)

Let me present a more focused scenario. Let's say that I'm drafting some sensitive non-article text; like a response to an RfA or Arbcom thing with lots of diffs. I'm in the middle of doing so and I need to step away for a few hours. I'm not done with my text but I want to save it because I'm putting a lot of effort into collecting and formatting all of those diffs and the power might go off or my browser might crash or whatever. And I don't want to save it in my sandbox because it's not cool to just post stuff concerning other users unless I'm sure it's ready for others to see. Today I have to copy that text to an offline text editor and copy it back later. What I'd like is just a button that says "save draft" and when I come back to the article, I can "resume draft". That's it. No namespaces, no unlimited storage of pictures of my board game collection.
Given that this has been a feature of practically all web-based content systems since forever (webmail, blog software, etc.) the objections presented so far are surprising. For example, I'm pretty sure a script kiddie wouldn't need the draft extension turned on if they wanted to DoS the system. In fact, this scenario happens rarely enough for me that I would be satisfied with being limited to a single draft with a fixed size limit if that addresses some of the concerns. Or a client-side solution, as was proposed. Regards, ORANGE SUEDE SOFA (talk) 17:25, 22 July 2015 (UTC)

> There isn't anything "surprising". Wikipedia isn't "all web-based content systems". It's a specific system working under its own constraints, and there is no two ways about having to address those constraints if you want to deploy this or any other new feature.--Anders Feder (talk) 17:59, 22 July 2015 (UTC)

> > What constraints do you have in mind? So far I've heard "nobody has created that yet" and "people might violate the terms of use by posting their account passwords on the web". Neither of these seem especially relevant to the question of whether it might be useful enough, for legitimate purposes, to be worth requesting. Whatamidoing (talk) 18:38, 22 July 2015 (UTC)

> > > Who have questioned whether it "might be useful enough"? Do you also see the exact words "I actually agree that such a thing would be useful" above or are they something I am imagining?--Anders Feder (talk) 18:44, 22 July 2015 (UTC)

> > @Orange Suede Sofa: You mention "drafting some sensitive non-article text; like a response to an ... Arbcom thing with lots of diffs" - there was a recent arbcom case where the accused was doing precisely that. It didn't go down at all well. You may notice that they haven't posted to this page for over a month now. --Redrose64 (talk) 19:30, 22 July 2015 (UTC)

> > > Redrose, this is confusing. The editor can't have been accused of "doing precisely that", because "precisely that" is technologically impossible at this point in time. Was the editor accused of drafting a reply in public, i.e., precisely the thing Orange Suede Sofa wants to avoid? Or of drafting the reply offline (which nobody has any business caring about)? Whatamidoing (talk) 18:51, 23 July 2015 (UTC)

> > > > By "doing precisely that" I mean that they were drafting their replies in userspace; that user subpage was quite lengthy, and contained a number of allegations against another Wikipedian which were sufficiently libellous for them to be removed and revdelled very quickly, and not long after were oversighted. --Redrose64 (talk) 08:06, 24 July 2015 (UTC)

> > > > > This sounds like an argument in favor of having a private space for such efforts. Private notes, unseen by anyone else are never libel. (Libel requires publication.) With luck, the editor would have kept editing until the contents were legal; if not, then it would be no worse than what happened. Whatamidoing (talk) 00:10, 29 July 2015 (UTC)

I would oppose having pages that only one editor could see. At the very least, there should be no spaces in the project that can not be seen by admins (who can see, for example, deleted page content and redacted revisions). Any editor who wants to draft things in a private space can already do it offwiki. **bd2412** T 19:44, 22 July 2015 (UTC)

> There is a type of draft that is semi-private. It is a draft in user space. Anyone can see it, but is unlikely to see it unless they either search for it or are directed to it. For composing something off-line that is completely private, why not just use a word processor or text editor? Robert McClenon (talk) 20:01, 22 July 2015 (UTC)

> > "Anybody in the entire world can read this" is what we call "not at all private". Security through obscurity is no security at all.
> > BD2412, I'd be interested in knowing *why* it's a problem to have a page that only the logged-in editor can see. Whatamidoing (talk) 18:51, 23 July 2015 (UTC)

> > > First, I consider it a WP:NOT problem. The policy says that Wikipedia is not a web hosting service, which is what an option like this could easily become. Unsavory characters could even use such a capacity as a space to communicate criminal plans. **bd2412** T 19:03, 23 July 2015 (UTC)

> > > > Definition from the article: "A web hosting service is a type of Internet hosting service that allows individuals and organizations to make their website accessible via the World Wide Web." That pretty much rules out a page that *nobody else can see*. Also, I'm not sure how one "communicates" anything, criminal or otherwise, when *nobody else can see it*. I asked why it's a problem to have a page that only the logged-in editor can see. You have replied with a concern that it could be bad if other people could see it. I agree, but that's not an answer to my question. What is the inherent problem in having a small page of text that *only one person* can look at? Whatamidoing (talk) 19:17, 23 July 2015 (UTC)

> > > > > First, it's not within Wikipedia's mission, so not worth the time doing the programming needed to create it. Second, people wanting to use this for illicit communication would only need to share the password with each other. Third, "a small page of text"? How small? How do we know how small it is, if it can't be looked at? Will the page history be publicly viewable? **bd2412** T 22:21, 23 July 2015 (UTC)

> > > > > > Because there's already a limit on page sizes. It's immediately enforced by software, so no human needs to look at any page to know that it's being done. Whatamidoing (talk) 00:10, 29 July 2015 (UTC)

- Copying and pasting only takes a few extra seconds if your favorite text editor is decent. If you really can't afford waiting those seconds, then why not use a pastebin? *Esquivalience* t 00:09, 23 July 2015 (UTC)
  - Among other problems: You might not be at your usual computer, or you might be planning to finish it elsewhere (e.g., you start at work and you finish at home). Saving on your computer doesn't so well work if your access is via an internet cafe or a borrowed computer. Pastebins aren't necessarily private.[41] (http://pastebin.com/trends) Copying and pasting on mobile and tablet devices is often difficult. Text editors have a tendency to create curly quotes, which mean that when you wanted *italics*, you end up with "a mess" instead. And that's just off the top of my head; other people could presumably add to this list. Whatamidoing (talk) 18:51, 23 July 2015 (UTC)
    - Isn't the wiki markup source code open source? As far as I know, anyone can download it and create their own wiki public or private. If you really want a draft space with wiki markup, that's an option. **bd2412** T 22:24, 23 July 2015 (UTC)

> Word processors have a tendency to create curly quotes. Text editors normally do what they're told. --Redrose64 (talk) 12:44, 24 July 2015 (UTC)

It looks like my original question has been answered. I appreciate the responses. Orange Suede Sofa (talk) 00:18, 23 July 2015 (UTC)

> To be honest, they didn't answer your question. They raised some valid concerns. Given, for example, that we had one draft page, limited to, say 64k or 128k, and viewable by admins most of those concerns would vanish. All the best: *Rich Farmbrough*, 00:18, 27 July 2015 (UTC)

Infobox image/text justification problem

I don't know when it exactly happened, I can't find the code where it happened. So did some css or module get updated, because over the past month or so I've constantly ran in to text and image justification problems. What used to be center justified is now left justified. Here's four examples (http://s1342.photobucket.com/user/BRaysball/talk/library/Wikipedia) I could find. I know I had to add some code (https://en.wikipedia.org/w/index.php?title=Template:Infobox_Wrestling_event&diff=669802150&oldid=663036065) to one template to fix it on that particular one. Did someone break something? TrueCRaysball (talk) 06:06, 25 July 2015 (UTC)

> @CR90:, that album is private. —TheDJ (talk • contribs) 07:25, 25 July 2015 (UTC)
>
> > @TheDJ: Sorry about that, it's fixed now. TrueCRaysball (talk) 07:54, 25 July 2015 (UTC)
> >
> > > @TrueCRaysball: It is still private :) --Edgars2007 (talk/contribs) 09:07, 25 July 2015 (UTC)
> > > Infobox alignment was changed in [42] (https://en.wikipedia.org/w/index.php?title=MediaWiki:Common.css&diff=668317639&oldid=668315752), discussed at MediaWiki talk:Common.css#Alignment of infobox labels. PrimeHunter (talk) 10:28, 25 July 2015 (UTC)
> > >
> > > > @TrueCRaysball: It's private for me too. It also takes ages to "leave" that page to come back here. I suspect a high level of javascript and advertising. If what you've posted there are screenshots, it's better to WP:WPSHOT than use an external service. --Redrose64 (talk) 14:58, 25 July 2015 (UTC)
>
> @Edgars2007 and @Redrose64: I tried redoing the privacy settings again. I used an external service because my screenshot has non-free logos in them to illustrate the point and I can't upload that to Commons. Should be fixed now. If it comes to it, I'll link each individually. @PrimeHunter: So IE9 wasn't interpreting code correctly so lets break the tables for everyone who updates their browser. Makes perfect sense. TrueCRaysball (talk) 19:40, 25 July 2015 (UTC)
>
> > By the way, I just finished update the template code for those infoboxes to fix their alignment. There are others, I'm sure, still affected by this. Wouldn't it have been easier to to just add `text-align:inherit;` to MediaWiki:Common.css rather than moving that code like what happened? TrueCRaysball (talk) 20:30, 25 July 2015 (UTC)
> >
> > > I did finally get through to that photobucket page - which took over a minute to load because the advertising included a video (of a potter's wheel, for some reason) which soaked up all the CPU cycles and caused the mouse and keyboard to stop responding - and then it took a further three mins to get out again and get to this edit screen, most of which was Windows cleaning up its cache, swapfile etc. to make space for Wikipedia's javascript to load again. If the logos are not relevant to the problem (they probably aren't, as it's about text alignment) crop them off or paint them out. --Redrose64 (talk) 21:06, 25 July 2015 (UTC)
> > >
> > > > One compound word: AdBlock. TrueCRaysball (talk) 23:05, 25 July 2015 (UTC)
> > > >
> > > > > For me (XP and latest Firefox) AdBlock isn't working. But OK, I understand that it's offtopic, so here it's just a note, that AdBlock doesn't work for everybody. If you have some suggestions (except upgrading the system :)), then you're welcome to my talkpage. --Edgars2007 (talk/contribs) 09:18, 28 July 2015 (UTC)

Issues over the past few days

- Session data seems to be lost almost every edit.
- Saving often displays the pre-edited version of the page. This was almost unheard of previously.

All the best: *Rich Farmbrough*, 21:03, 27 July 2015 (UTC).

> - In regards to the Saving, I'm finding it 100% of the time the last few days. I have to do a "refresh" after saving to see the changes.
> - Loss of session data happens sporadically with me, but not all the time.
>
> — Maile (talk) 21:10, 27 July 2015 (UTC)
> @Rich Farmbrough: Wikipedia:Village pump (technical)/Archive 137#"Loss of session data" error on Save page and Wikipedia:Village pump (technical)/Archive 137#Post not showing up immediately respectively. Other threads exist. When are you next in Oxford? --Redrose64 (talk) 22:53, 27 July 2015 (UTC)
>
> > With a little luck and a following wind, August. All the best: *Rich Farmbrough*, 23:01, 27 July 2015 (UTC).
> >
> > > Session data lost: phab:T102199. --AKlapper (WMF) (talk) 08:50, 28 July 2015 (UTC)

edit count language

When I go to my user contributions and I select Edit Count I see that that my edits are broken down according to type of namespace. (Talk, User space, ect.) The top category, which should be mainspace, or Articles, or something like that, displays these foreign (perhaps Korean) symbols: 일반 문서. What's wrong? -- Naytz (talk) 20:20, 28 July 2015 (UTC)

> It's a known bug in XTools (or some external service it relies upon). See Wikipedia talk:XTools#Label for article namespace in edit counter is in Korean.--Anders Feder (talk) 20:28, 28 July 2015 (UTC)
> Me too. Hawkeye7 (talk) 20:29, 28 July 2015 (UTC)
> User talk:MusikAnimal/Archive 16#anomoly > tools.wmflabs.org and https://github.com/x-Tools/xtools/issues/60 NQ-Alt (talk) 20:33, 28 July 2015 (UTC)

Moving over talk page

Apologies if this has been covered before, but I couldn't find anything in the archives. Currently when you move a page (as an admin) if the target location has a non-negligible history (i.e. anything more than a single edit redirecting to the page you're moving) you get a really helpful screen that tells you the target has history, gives a link to that history and then gives you a tick box to delete that page so you can proceed with the move if you want. The problem is that often the talk page will also have a non-negligible history, but you don't get a prompt or any real warning for this. Once the move is completed there is a line at the bottom of the page that tells you whether moving the talk page (and any archives) was successful or not, but it's very easy to miss. So my question is, would it be feasible to somehow have the move feature detect when, if the "move accompanying talk page" box is ticked, the talk page that is being moved to has a non-negligible history and give you a tick box option to delete it all in the same process? If that is too difficult or complicated, would it at least be possible to make the notice somehow more prominent (bigger and in red, say?) when moving the talk page is unsuccessful? Thanks, Jenks24 (talk) 08:17, 22 July 2015 (UTC)

Anyone? Jenks24 (talk) 14:28, 23 July 2015 (UTC)

Yes but I think it's a developer job to do it properly. Possibly you could use javascript to suppress the "real" move tab and create a "fake" one, and Lua to do the extra checks, but it would be better to make it a feature IMHO. All the best: *Rich Farmbrough*, 00:20, 27 July 2015 (UTC).

Hey Rich, many thanks for responding. I agree that a developer making this a site-wide feature, rather than just a personal hack for me, would be the optimal outcome. The following might be a silly question, but how do I actually contact the developer team with a suggestion like this? Jenks24 (talk) 05:46, 27 July 2015 (UTC)

Through the Phabricator (https://phabricator.wikimedia.org/). Jo-Jo Eumerus (talk, contributions) 08:27, 27 July 2015 (UTC)
+1. All the best: *Rich Farmbrough*, 12:40, 27 July 2015 (UTC).

Thank you both. This has prompted me to finally sign up for Phabricator (can't remember if I had an account for the old one, whatever that was called). Turns out this has been a bug since 2007 [43] (https://phabricator.wikimedia.org/T12814). I added a comment to it, but no idea if that will do anything. Jenks24 (talk) 13:55, 29 July 2015 (UTC)

Default edit summary

Is it possible with js or css code in my personal js/css subpage to add some default edit summary for edits? *&summary=* isn't answer this time, because I'm using regular wikilinks, not URLs. Sometimes I'm too lazy to add edit summary if I'm making some mass edits to many pages where I'm doing the same thing, like adding template, DEFAULTSORT etc. --Edgars2007 (talk/contribs) 11:14, 28 July 2015 (UTC)

You could use a bookmarklet: `javascript:document.getElementById("wpSummary").value = "New summary"; void(0);`. Alakzi (talk) 11:27, 28 July 2015 (UTC)

I assume you tested that. Not bad, especially if your browser supports buttons on a bookmark toolbar (do they all?). You would still have to remember to click the button, but it would be slightly better than a copy-and-paste from an open Notepad window. Little use for less tech-savvy editors, unless there was a page with well-written usage instructions for each of the major browsers. Ideal solution: A check box at the bottom of the edit window, "Save edit summary", but I won't hold my breath on that one. —*Mandruss* ☎ 11:42, 28 July 2015 (UTC)

OK, bookmarklet does the job well. Thanks! But if that code can be modified to put it here, then I would love it more, as Mandruss said - still have to remember to click the button. Like the idea about check box :) --Edgars2007 (talk/contribs) 12:12, 28 July 2015 (UTC)

You can do it like that:

```
$(function() {
        $( "#wpSummary" ).val( "New summary" );
});
```

The summary will be filled each time the page is loaded. Darkdadaah (talk) 12:48, 28 July 2015 (UTC)

Thanks! Works good. --Edgars2007 (talk/contribs) 13:11, 28 July 2015 (UTC)

I use a free programme called AutoHotKey to create macros, so I can type any regularly-used string with just three keystrokes. I wrote a blog post explaining how (http://pigsonthewing.org.uk/using-autohotkey-macros-make-typing-life-easier/). I reserve the combination `aa` for temporary strings. I also take advantage of my browser (Firefox)'s autocomplete function. Andy Mabbett (Pigsonthewing); Talk to Andy; Andy's edits 12:44, 28 July 2015 (UTC)

This one looks interesting, thanks! --Edgars2007 (talk/contribs) 13:11, 28 July 2015 (UTC)

User:Equazcion/CustomSummaryPresets allows you to define custom edit summaries, which appear in a drop-down menu below the edit summary line. - Evad37 (talk) 01:23, 29 July 2015 (UTC)

Authorlinking in German template

I would like to introduce authorlinks into citations in Biber-Danube interglacial and Biber glaciation to the authors Lorraine Lisiecki and Maureen Raymo. Unfortunately, the citations use a template "literatur" which appears to be German and doesn't seem to accept "authorlink=". Any help gratefully received. DuncanHill (talk) 12:38, 29 July 2015 (UTC)

a) introduced a new parameter `|Authorlink=` (i know it's not German), so you can use it
b) probably nobody would blame you, if you use wikilinks in the `|Autor=`, because I see, that there are multiple persons
c) General question for everyone - why not use {{cite book}} (which is based on Lua) as metatemplate for {{Literatur}}? --Edgars2007 (talk/contribs) 12:58, 29 July 2015 (UTC)

Thanks, I've gone for wikilinking in the "Autor" field, together with a plea in the edit summary. DuncanHill (talk) 13:18, 29 July 2015 (UTC)
T:Literatur is not a CS1 template *directly* though it does use Citation/core ~~which relies on the same Module:Citation as CS1~~. It is a copy fast I checked for the German template of the same name. In general, where it is found in an article it should be *replaced* with the appropriate CS1 template (IMO) since its only use is in copying/transwiking an article *from* German *to* English. --Izno (talk) 16:36, 29 July 2015 (UTC)

I concur with regards to replacement. `{{Literatur}}` does not use Module:Citation (nothing does, I think), rather, `{{Literatur}}` uses `{{citation/core}}` which used to support both Citation Style 1 and Citation Style 2 until both of those migrated to Module:Citation/CS1.

—Trappist the monk (talk) 16:46, 29 July 2015 (UTC)

citedoi templates

Some references call a "citedoi" template, and do not seem to have a way of authorlinking. Is there any way in which the authors of a work cited in this way can be linked? An example would be the ref name=" LisieckiRaymo" in the article Pastonian Stage. Thanks, DuncanHill (talk) 13:21, 29 July 2015 (UTC)

> Click the Edit tab to see a list of transcluded templates at the bottom of the edit window. Pastonian Stage has the code `{{Cite doi|10.1029.2F2004PA001071}}` which transcludes Template:Cite doi/10.1029.2F2004PA001071. You can manully edit that page. It uses Template:Cite journal which has documentation for making author links. PrimeHunter (talk) 13:38, 29 July 2015 (UTC)
>
> > I had never noticed the list of templates at the bottom, probably because it was collapsed! Thanks, hopefully will be able to manage from here on. DuncanHill (talk) 13:41, 29 July 2015 (UTC)
> >
> > Again thanks - was much simpler than I had feared! I shall endeavour to remember for future reference. DuncanHill (talk) 13:46, 29 July 2015 (UTC)
> >
> > > DuncanHill, please keep asking questions. We'd rather have you productive than stuck on something, especially when we can answer your questions quickly. You can also post at WP:Help desk if you have less-technical questions about editing. Whatamidoing (talk) 16:41, 29 July 2015 (UTC)

Reading Non-acsii characters via mwclient

I am currently experiencing an issue when I try to read from pages with non-ascii characters that I lose that character. For instance, if I have prime symbol it is changed to a '?'. Has anyone experienced similar issues? I am utf-8 encoding everything once I have the text, but I have already lost the non-ascii characters prior. Any suggestions or support resources would be greatly appreciated.Julialturner (talk) 06:40, 28 July 2015 (UTC)

> Some minimal testcase might be welcome - how do you invoke mwclient? How did you set utf-8 encoding? Which underlying operating system is this about? --Malyacko (talk) 08:52, 28 July 2015 (UTC)
>
> Sure. I invoke mwclient like so:

```
    def connection(self):
        useragent = 'Protein Box Bot, Run by The Scripps Research Institute: nanis@scripps.edu'
        connection = mwclient.Site(('https', settings.BASE_SITE),clients_useragent=useragent)
        connection.login(self.username, self.password)
        return connection
```

> utf-8 is set like:

```
self.wikitext().encode('utf-8')
```

> My os is Ubuntu 14.04. Thanks,Julialturner (talk) 20:39, 29 July 2015 (UTC)

Sorting search results

Is there any way to sort search results by date the page was modified? --NeilN talk to me 14:27, 27 July 2015 (UTC)

> Don't think so. Using *prefer-recent:* is the only alternative. More: T40403, T64879 - NQ-Alt (talk) 14:54, 27 July 2015 (UTC)
>
> > "Adding in a feature to sort by date or alphabetically by title will, for the reasons explained above, result in degraded performance for the vast majority of users. It's for this reason that search engines like Google don't allow you to sort by date or alphabetically by title; it degrades the quality of the service. I'm WONTFIXing this bug accordingly, as I cannot justify adding features to CirrusSearch that degrade the experience for the vast majority of its users." Gotta love the arrogance of some of the development team, telling users "no, no, we think you won't understand what a 'sort by date' button *really* does." --NeilN talk to me 16:06, 27 July 2015 (UTC)
> >
> > > @NeilN: If you wish to constructively discuss this request, then please reach out to me privately, or discuss the request here, and I'd be more than happy to talk to you about it. If, on the other hand, you wish to continue in this unconstructive manner, attacking others rather than discussing the matter at hand, then I will not engage with you further. Thank you. --Dan Garry, Wikimedia Foundation (talk) 03:30, 28 July 2015 (UTC)
> > >
> > > > @Deskana: I see the discussion that took place in the phabricator report. Will repeating the points do *any* good? You've already made your decrees based on very misleading statements ("degrade the experience for the vast majority of its users"). --NeilN talk to me 03:41, 28 July 2015 (UTC)
> > > > @Deskana: What the heck. I'll give it a shot. Please justify your comparison that Wikipedia pages are the "web" and Wikipedia search is "Google". This is a little grandiose and ignores the fact that Wikipedia pages have structure and Wikipedia does not contain a billion pages of garbage. Given the less than stellar parts of the current UI, please justify your assumption that a clearly marked "Sort by date" button would "result in degraded performance for the vast majority of users". Please justify your statement that "I've already outlined that sorting by date will, for the vast majority of users, generate meaningless results." You've repeated your assumption, you haven't justified it. --NeilN talk to me 04:10, 28 July 2015 (UTC)
> > > >
> > > > > @NeilN: That isn't exactly a positive start to the conversation. There's very little to be gained by debating the past, especially given your combative way of asking these questions. What would be productive is for us to work together to identify what it is you're trying to do, so that I can see if I can help support it. Why don't we start by you walking me through what task you're trying to accomplish? Then I can see if we can support it. Does that sound like something you'd be interested in doing? (P.S. Please don't ping my volunteer account with messages relating to my work, as I'm unlikely to see them; please ping User:Deskana (WMF) instead) --Dan Garry, Wikimedia Foundation (talk) 04:34, 28 July 2015 (UTC)
> > > > >
> > > > > > User:Deskana (WMF), but you've already dismissed my use case (looking at articles containing a term which have recently changed) using the assumptions I've listed above. Now I'm asking you to justify them. --NeilN talk to me 04:41, 28 July 2015 (UTC)

@NeilN: I have already done so. You are within your rights to disagree that I have. However, that does mean that this is no longer a productive conversation, so I must discontinue it so I can get back to my work. Best wishes. --Dan Garry, Wikimedia Foundation (talk) 04:59, 28 July 2015 (UTC)

Not pinging Deskana as it's clear he no longer wishes to participate but can someone else look at the phab reports and point out where he's actually justified his assumptions? --NeilN talk to me 05:05, 28 July 2015 (UTC)

Why didn't you just ask him to explain the parts of the justification you don't understand?--Anders Feder (talk) 14:49, 28 July 2015 (UTC)

Anders Feder, *what* justification? All I see is a bunch of unsupported assertions. Kind of odd for a group that loves A/B testing. --NeilN talk to me 13:36, 30 July 2015 (UTC)

English Wikipedia is extremely slow

I am writing to report that English Wikipedia is extremely slow right now (I can not access WP:VPT), and sometimes gives an error:

"This page can't be displayed

•Make sure the web address https://en.wikipedia.org is correct. •Look for the page with your search engine. •Refresh the page in a few minutes. •Make sure TLS and SSL protocols are enabled. Go to Tools > Internet Options > Advanced > Settings > Security"

Thanks, --Jax 0677 (talk) 18:16, 27 July 2015 (UTC)

That seems to be a problem with your internet connection. It works fine for me. Tvx1 18:18, 27 July 2015 (UTC)

Reply - Actually, I can access Spanish Wikipedia and Simple English Wikipedia just fine. --Jax 0677 (talk) 18:43, 27 July 2015 (UTC)

Did you follow the last step in the instructions given in the error message? --Malyacko (talk) 08:46, 28 July 2015 (UTC)
This sounds like it could be a load-balancing problem in one of the WMF datacenters - see this thread for a previous example. Another reason may be slow JavaScript. Try logging out (or browsing in private mode, which essentially logs you out) and see if you still experience slowness. If things are still slow when you are logged out, it is a good indicator that it is a load-balancing problem rather than a JavaScript problem. If this is the case, please let us know roughly where you are in the world, as load-balancing problems are often limited to a specific geographical area. — *Mr. Stradivarius* ♪ talk ♪ 10:08, 28 July 2015 (UTC)

Reply - The speed of English Wikipedia is back to normal. --Jax 0677 (talk) 13:10, 30 July 2015 (UTC)

File upload problem

The following discussion is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

Problem fixed. (non-admin closure) Erpert blah blah blah... 23:55, 30 July 2015 (UTC)

It looks like the "upload failed: invalid token" error message is showing up again when trying to upload files. Erpert blah blah blah... 04:15, 30 July 2015 (UTC)

Is this about https://en.wikipedia.org/wiki/Wikipedia:File_Upload_Wizard ? Which file types have you tried? Which browser is this about?
--AKlapper (WMF) (talk) 10:11, 30 July 2015 (UTC)

Yes, it was the file upload wizard. And I tried to upload a jpeg via Google Chrome. But everything is working fine now, so I'll close this. Erpert blah, blah, blah... 23:55, 30 July 2015 (UTC)

The discussion above is closed. Please do not modify it. Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.

Citation now spam

I found a citation to a website that is now squatted to a generic spam search. Luckily I was able to find an archive to link to. However I didn't want to leave the spam link, nor did I want to remove the url - I think I took the protocol identifier off, and left it at that. Is there a consensus way tot deal with these links? All the best: *Rich Farmbrough*, 19:55, 29 July 2015 (UTC).

Check whether the link was recently replaced - some spammers replace links in citations, especially broken ones. Otherwise, standard editing will have to serve, along with link blacklisting if it happens repeatedly. Jo-Jo Eumerus (talk, contributions) 20:03, 29 July 2015 (UTC)

Rich, that sounds like a good solution. Why don't you document it at WP:DEADREF, in case anyone else encounters the same problem?
Whatamidoing (talk) 18:56, 31 July 2015 (UTC)
Well given that the archive version was good, it obviously wasn't the type of "dead link spam" SEO'ers have been doing recently. And link blacklisting is no help against what happened several years ago, far better to generate a list of all external links to that domain and add archive urls where possible (a nice little job for automation, which, of course I cannot do!).
But thanks for the suggestion. All the best: *Rich Farmbrough*, 19:12, 31 July 2015 (UTC).

What just happened to the watchlist?

I have MonoBook skin, not some beta-testing Mobile thing. The box at the top of the Watchlist, with various options, has just gone all Facebooky, grey and unreadable with lots of blank space. How can I switch it back to how it was? --Redrose64 (talk) 18:37, 30 July 2015 (UTC)

Tracked in Phabricator Task T107311

"Invert selection" and "Associated namespace" only apply if a namespace has been selected so they are grey before that.
Do you see other grey parts, or are they still grey after a namespace selection? PrimeHunter (talk) 18:43, 30 July 2015 (UTC)
(edit conflict) I too have MonoBook. Multiple buttons are very large all of a sudden, and the Invert selection/Associated namespace checkboxes have a bit of excessive whitespace around them. *Dustin* (talk) 18:45, 30 July 2015 (UTC)

Why is "Mark all pages as visited" so big? I never use that button so I don't know why it needs to take up so much vertical space. Sam Walton (talk) 18:46, 30 July 2015 (UTC)

Yes, that needs fixing. Lugnuts Dick Laurent is dead 18:56, 30 July 2015 (UTC)

Yup. A waste of space - poor ergonomics. 18:57, 30 July 2015 (UTC) — Preceding unsigned comment added by AndyTheGrump (talk • contribs)

@PrimeHunter: The obviously-visible items are: the word "Namespace"; the word "all" below that; two grey squares (which *may* be checkboxes - without the familiar inset border it's hard to tell); a "Go" button, which is much bigger than it used to be - and with a background of light blue instead of silver. There is also some barely-readable grey text; dragging my mouse over it, I see that it's "Invert selection" and "Associated namespace". I also find that the namespace selector has a border that is so pale that it's even less noticeable than that grey text. Going away and coming back I find that the text starts off black but quickly turns grey, like there's some javaScript going on. --Redrose64 (talk) 19:07, 30 July 2015 (UTC)

Seems to be fixed. Back to how it was before. - NQ (talk) 19:10, 30 July 2015 (UTC)

Yes, with black text, very little superfluous space, the namespace selector and checkboxes white within inset border, and a silver button. --Redrose64 (talk) 19:15, 30 July 2015 (UTC)
This looks like gerrit:211131, which was the patch for task phab:T99256. I assume someone just rolled it back... — *Mr. Stradivarius* /talk/ 19:13, 30 July 2015 (UTC)
I see it was reverted by Legoktm in gerrit:228046 just now. The spacing problems with the new patch are tracked in phab:T107311 if people are interested. — *Mr. Stradivarius* /talk/ 19:21, 30 July 2015 (UTC)

Hi, the change of watchlist interface was indeed deployed prematurely (and now reverted), and we missed some of the display issues with it (mostly there wasn't meant to be that much whitespace, and it behaved strangely on small screens). I didn't author or accept it myself, but I reviewed it and didn't flag them. We're going to be trying again, with feeling this time, probably next-next week (week of 10 August). Please watch phab:T99256 for updates (you can "Subscribe" if you have a Phabricator account to receive updates by e-mail), I'll make sure there's a testing wiki with the patch set up and linked from that patch at least a few days earlier, for everyone to play with and comment. Matma Rex talk 20:11, 30 July 2015 (UTC)

Use of addresses such as: "https://en.wikipedia.org/wiki/Talk/Xxxx" or "https://en.wikipedia.org/wikitalk/Xxxx" for searchability etc.

At present talk pages for articles have addresses in formats such as: "https://en.wikipedia.org/wiki/Talk:Foo"

If, however, the page has a title starting say with: "Category:Foo" the associated talk page is assigned its address in the format: "https://en.wikipedia.org/wiki/Category_talk:Foo"

Would it be possible/practical to change talk page address formatting to "https://en.wikipedia.org/wiki/Talk/Xxxx" for article talk pages and, for instance, "https://en.wikipedia.org/wiki/Talk/Category:Foo" for category talk pages. I would also be interested to know how possible it might be to use addresses such as "https://en.wikipedia.org/wiki/category/Foo" and "https://en.wikipedia.org/wiki/talk/category/Foo" either with or without the initial capitalisation of the words "talk" and "category"?

The type of changes mentioned, I understand, would greatly increase the internet search-ability of talk pages as this would facilitate the use of search terms such as: site:www.en.wikipedia.org/wiki/Talk/ search term/s.

GregKaye 10:11, 30 July 2015 (UTC)

Wikipedia data structure

Namespaces			
Subject namespaces		Talk namespaces	
0	(Main/Article)	Talk	1
2	User	User talk	3
4	Wikipedia	Wikipedia talk	5
6	File	File talk	7
8	MediaWiki	MediaWiki talk	9
10	Template	Template talk	11
12	Help	Help talk	13
14	Category	Category talk	15
100	Portal	Portal talk	101
108	Book	Book talk	109
118	Draft	Draft talk	119
446	Education Program	Education Program talk	447
710	TimedText	TimedText talk	711
828	Module	Module talk	829
2300	Gadget	Gadget talk	2301
2302	Gadget definition	Gadget definition talk	2303
Virtual namespaces			
-1	Special		
-2	Media		

This may have been a good suggestion in the early days of the MediaWiki software, twelve or so years ago, but it's far too late to change now - too much depends on the existing pagename format and URL structure. --Redrose64 (talk) 10:20, 30 July 2015 (UTC)

And you can use keywords in Google like intitle and inurl to get the results you need. Graham87 11:24, 30 July 2015 (UTC)

I'm puzzled. This is a search in talkspace (https://www.google.com/?gws_rd=ssl#safe=off&q=site:en.wikipedia.org%2Fwiki%2Ftalk+birds) and this is one in category talkspace (https://www.google.com/?gws_rd=ssl#safe=off&q=site:en.wikipedia.org%2Fwiki%2Fcategory_talk+birds). More fundamentally, consider using Special:Search with Advanced since it works well these days. --Izno (talk) 12:02, 30 July 2015 (UTC)

I oppose the suggested change but will just note about the above Google search that Google doesn't actually index "Talk:". They do index some talk pages they picked up as "Talk%3A" where "%3A" is a percent-encoded ":" and our servers produce the same content. I don't know why Google drops "Talk:". I don't see anything relevant in https://en.wikipedia.org/robots.txt, and there is no noindex in the html of the pages (if there were then it should also be in the %3A versions). I don't want Google to index talk pages, I just wonder what stops them. Did Google decide on their own that talk pages are too uninformative to deserve the high placement they would get? PrimeHunter (talk) 12:36, 30 July 2015 (UTC)

Google has special cleverness for Wikipedia. As far as I know they have not shared what it is. I am pretty sure they do not honour __NOINDEX__ though. All the best: *Rich Farmbrough*, 19:35, 31 July 2015 (UTC).

This wouldn't work because it would clash with Main namespace pages containing slashes in the title. An article named "Talk/Foo" would have the same URI as the Talk page for the page named "Foo". Slashes in URIs are already an issue in MediaWiki, since MediaWiki uses them as part of the page title for subpages, which raises the same issue of potential URI clashes. So, MediaWiki has a setting that allows you to disable subpages on a per-namespace basis, with them off by default in the Main namespace (which is the setting Wikipedia uses). You also touched on initial capitals in page titles, which are another pain point. By default the first character in a page title is case-insensitive, so [[Foo]] and [[foo]] will go to the same page, which is what people tend to expect. Wiktionary has this setting toggled so it can have different articles on, e.g., Rock and rock, but this means you always have to pay attention to the initial capital in wikilinks, and I think it messes with searching too. Basically there's never a perfect solution for anything in (software) engineering. It's all about what tradeoffs you choose. --108.38.204.15 (talk) 22:36, 31 July 2015 (UTC)

Main page on mobile

Breaking with centuries of tradition I today views the mainpage from a mobile phone. A lot of the content, for example DYK was not visible. All the best: *Rich Farmbrough*, 19:36, 31 July 2015 (UTC).

It's by design. At Talk:Main Page/Archive 182#Link to full Main Page for mobile users I suggested an option for mobile users to see the full main page without having to switch to desktop. PrimeHunter (talk) 19:53, 31 July 2015 (UTC)

On the german mobile WP main page (https://de.m.wikipedia.org/wiki/Wikipedia:Hauptseite) you get it all ("Für die mobile Hauptseite wurden bisher die Rubriken Artikel des Tages, Was geschah am?, In den Nachrichten, Kürzlich Verstorbene und Schon gewusst? aktiviert.") Youu can decide it, and see https://www.mediawiki.org/wiki/Mobile_Gateway/Mobile_homepage_formatting --Atlasowa (talk) 20:26, 31 July 2015 (UTC)

Annotations in small images

I don't know how many people have set their preferences to be able to view image annotations from Commons, but the feature can cause problems in small resolutions, where it is far more annoying than useful. It's not so much the plethora of tiny yellow boxes that can obscure an image without highlighting anything visible, as it is the text that appears below it—"This file has annotations. Move the mouse pointer over the image to see them."—which can take up more space than the image itself. I noticed this phenomenon in the display templates for good topics (where I initially tried to solicit opinions, before giving it a try here), but I imagine it can affect small images everywhere, such as in navigation templates.

The situation hasn't changed much in two-and-a-half years (except perhaps the increased likelihood of coming across an image with annotations), and my question is this: is there a way to suppress the feature when displaying an image? And if not, could one be devised? Waltham, *The Duke of* 14:32, 31 July 2015 (UTC)

@The Duke of Waltham: I've gone to Preferences → Gadgets and enabled "ImageAnnotator: view image notes and comments on file description pages". Where can I see these tiny yellow boxes and the text below it? --Redrose64 (talk) 19:19, 31 July 2015 (UTC)

I for example see it for File:The death of general warren at the battle of bunker hill.jpg at Wikipedia:Featured topics/Boston campaign. It's not in thumbs but without thumb and with at least 89px I get the yellow boxes and "This file has annotations. Move the mouse pointer over the image to see them." The first version displayed here is 88px and the second 89px. PrimeHunter (talk) 19:47, 31 July 2015 (UTC)

Nope. Is it skin- or browser-specific? I use MonoBook and Firefox 39. --Redrose64 (talk) 20:45, 31 July 2015 (UTC)

In Firefox 39 I see it in both MonoBook and Vector. The text is made with JavaScript and is added shortly after page load. The yellow boxes are only visible when hovering over the image. I don't know whether the 89px limit depends on anything. I have added a 400px version where I also see the text and yellow boxes. PrimeHunter (talk) 20:57, 31 July 2015 (UTC)

Aha. I needed to also disable "Redirect image links to Commons for files hosted there". Seems that they can't coexist. --Redrose64 (talk) 21:45, 31 July 2015 (UTC)

Thank you for the detailed investigation. It makes sense that there would be a size limit, but it really is arbitrary because it depends on the specific image and its level of detail. And as I've said, the text takes up a lot of space.
It turns out that a closer study of the documentation reveals the possibility of turning annotations off, at least in Commons. There's a template there, ImageNoteControl, which incorporates the feature and that could be transferred here. No, wait; it's already here, though there is no interwiki link on the Commons page so it's not immediately apparent. {{ImageNoteControl}} in en.wikipedia is primarily transcluded in File pages themselves, and... little else (https://en.wikipedia.org/w/index.php?title=Special:WhatLinksHere/Template:ImageNoteControl&namespace=6&limit=50&invert=1), which is also curious. The page is practically an orphan (https://en.wikipedia.org/w/index.php?title=Special:WhatLinksHere/Template:ImageNoteControl&hidetrans=1); no wonder most people have probably never heard of it in its five-year history.
Although my programming skills are extremely limited, I think I've managed to copy the template's relevant command here and suppress annotations in the medium-sized image on the right. (I have no idea if there is any difference between span and div, though; they're both in the template.) If someone knows how to make this—the command rather than the entire ImageNoteControl template—part of {{Featured topic box}}, that would solve the immediate problem. For other small images, more publicity for {{ImageNoteControl}} might be desirable. Waltham, *The Duke of* 06:09, 1 August 2015 (UTC)

I think that it's "cleaner" to have three classes instead of just one: `class="wpImageAnnotatorControl wpImageAnnotatorOff wpImageAnnotatorCaptionOff"` As for `<div>...</div>` versus `<span>...</span>` it depends upon the context. For images used as block elements (as with all examples so far) div is correct; for images used inline, like this then span is correct. --Redrose64 (talk) 11:01, 1 August 2015 (UTC)

@The Duke of Waltham: In the case of `{{featured topic box}}` the image is inline, so `<span>...</span>` is correct. This edit (https://en.wikipedia.org/w/index.php?title=Template:Featured_topic_box/sandbox&diff=prev&oldid=674084073) should do it; compare Wikipedia:Featured topics/Boston campaign/sandbox with Wikipedia:Featured topics/Boston campaign. --Redrose64 (talk) 11:19, 1 August 2015 (UTC)

Capital! This is exactly the desired effect. Thank you very much for your trouble—and for the impromptu HTML lesson along the way. It would feel wrong for me to implement your edit, especially considering the possibility that you might still want to tinker with it, so I'll leave it for you to proceed with that step whenever you are ready. Other

than that... I'll see if I can find some image-related help page in the English Wikipedia where inserting a mention of {{ImageNoteControl}} would be productive. Waltham, *The Duke of* 11:59, 1 August 2015 (UTC)
@Redrose64: Apparently, notifications work in a very specific way. This time it ought to work. Waltham, *The Duke of* 12:16, 1 August 2015 (UTC)

✓ Done --Redrose64 (talk) 13:04, 1 August 2015 (UTC)

Categorization

Recently I've notice a user added Vido to Category:Islands of Greece. It is indeed an island in Greece, but the article is already in the Category:Islands of the Ionian Islands (region), which is then contained in the Category:Islands of Greece. This way, the article is included in the Category:Islands of Greece twice. Is it possible to make some software limitation that would prevent this? That would prevent an article being thrown into a category to whose subcategory it already belongs? Or, if this is not possible, is it that possible to make a bot that would remove such redundant categories? Vanjagenije (talk) 23:43, 1 August 2015 (UTC)

@Vanjagenije: It's not clear-cut, see WP:CATDIFFUSE and WP:DUPCAT. --Redrose64 (talk) 00:13, 2 August 2015 (UTC)

Precisely. This is, frankly, an ambiguous provision that cannot easily be decided and makes no real sense. We'd be much better off enforcing the Commons policy, which prohibits the inclusion of parent categories with a single easy-to-identify exception. Nyttend (talk) 01:34, 2 August 2015 (UTC)

Search links not appearing

I just discovered that all my search links that are limited to searching the article text (eg. `{{search link|text="buggy"}}`) stopped rendering sometime in the ~~past 48 hours~~ past week (apparently my earlier conversation on this very board and with John of Reading on his userpage didn't use the *text* parameter). I've checked instances logged-in and logged-out, on Chrome and on Firefox, and I get the glitch in all cases. This does not seem to affect instances of {{Search link}} that do *not* have that limitation. I have not had a chance to test various alternates. Test matrix:

plaintext using <nowiki> tag – regular wikitext
{{search link|text="buggy"}} – "buggy" (https://en.wikipedia.org/w/index.php?title=Special:Search&search=%22buggy%22&ns0=1&fulltext=Search)
[restored as of 22:34 (UTC)]
{{search link|"buggy"}} – "buggy" (https://en.wikipedia.org/w/index.php?title=Special:Search&search=%22buggy%22&ns0=1&fulltext=Search)

Insights will be welcome! —jameslucas (" " / +) 21:20, 28 July 2015 (UTC)

User:Cpiral changed the parameter names in [44] (https://en.wikipedia.org/w/index.php?title=Template:Search_link&diff=672627706&oldid=671097083) without allowing the old names as aliases. That's problematic for an old template with many uses. I see Cpiral updated some uses of the old names. Was that all of them or are there still many? In either case I suggest allowing the old names. PrimeHunter (talk) 21:35, 28 July 2015 (UTC)

Was problematic. With {{Template usage}} now anyone can now find *all* template usage and directly removing obsolete parameter usage from the wikitext, avoiding the need for backward compatible code. — CpiralCpiral 23:03, 28 July 2015 (UTC)

What does {{Template usage}} have got to do with breaking existing transclusions of another template? Alakzi (talk) 23:40, 28 July 2015 (UTC)

Cpiral has been doing some strange things recently, see their edits to Help:Template over the last two weeks (https://en.wikipedia.org/w/index.php?title=Help:Template&action=history&offset=20150714000000&dir=prev). --Redrose64 (talk) 21:51, 28 July 2015 (UTC)

It was *necessary* to evolve {{search link}} for {{regex}} which was necessary for Help:Searching/Draft. My work on the {{Val}} family got me to create {{Template usage}}, which got me interested in improving Help:Template. — CpiralCpiral 23:03, 28 July 2015 (UTC)
Replacing parameter names is a common newbie mistake. What about their edits to Help:Template? Alakzi (talk) 22:38, 28 July 2015 (UTC)

@Alakzi: They have been rewriting whole sections, much of their new text is barely comprehensible. In this edit (https://en.wikipedia.org/w/index.php?title=Help:Template&diff=prev&oldid=671371357), for example, terms like "parameter" and "argument" are used almost interchangeably; and although they state early on that there are two kinds of parameter: named and unnamed. Soon after, we find that there is a third kind, the positional parameter, which is apparently not the same as an unnamed parameter. Have a look at each edit individually - they really are difficult to follow. The most recent large edit produced the paragraph

To improve readability many programming languages ignore much of the whitespace, so programmers can add newlines and indent almost at will. Because of the nature of transcluding text in place, seemlessly, MediaWiki software is very sensitive to whitespace, only allowing it around some places, but in most places newlines for code-readability are treated by the software as content, so the template code uses <!-- comments --> as a work around, adding <!-- before each newline character and --> after it.

which really is *not* an improvement in readability. --Redrose64 (talk) 23:32, 28 July 2015 (UTC)

Right, I see what you mean. The documentation of {{Template usage}} is difficult to follow as well. Alakzi (talk) 23:40, 28 July 2015 (UTC)

Awh, let's go ahead and sully the talk page at Help:Template. I've started a conversation there about the changes. You can refer to me in first person now. Thanks. -- CpiralCpiral 00:16, 29 July 2015 (UTC)

There's another big issue with updating parameter names: Many other projects rely on the template infrastructure of the English Wikipedia. Breaking stuff makes it much harder to adopt updates. All the best: *Rich Farmbrough*, 19:58, 29 July 2015 (UTC).

I think I understand your concerns, but can you specify? I am not convinced that template infrastructure need develop differently than the way I am developing it. I can achieve new-feature parity for any template and avoid the need for carrying any backward compatible code, by directly changing every instance of obsolete parameter usage on the wiki, then changing the template. The documentation was updated. Was I supposed to change these few on the wiki in User space? (https://en.wikipedia.org/w/index.php?title=Special:Search&search=hastemplate%3A%22search+link%22+insource%3A%2F%5C%7B%5C%7B+%2A%5BSs%5Dearch+link+%2A%5C%7C%5B%5E%7D%5D%2A%28link%

7Ctext%29%3D%
2F&ns0=1&ns1=1&ns2=1&ns3=1&ns4=1&ns5=1&ns6=1&ns7=1&ns8=1&ns9=1&ns10=1&ns11=1&ns12=1&ns13=1&ns14=1&ns15=1&ns100=1&
— CpiralCpiral 02:36, 3 August 2015 (UTC)

What does a Healthy Community look like to you?

Hi,
The Community Engagement department at the Wikimedia Foundation has launched a new learning campaign. The WMF wants to record community impressions about what makes a healthy online community. Share your views and/or create a drawing and take a chance to win a Wikimania 2016 scholarship! Join the WMF as we begin a conversation about Community Health. Contribute a drawing or answer the questions on the campaign's page.

Why get involved?

The world is changing. The way we relate to knowledge is transforming. As the next billion people come online, the Wikimedia movement is working to bring more users on the wiki projects. The way we interact and collaborate online are key to building sustainable projects. How accessible are Wikimedia projects to newcomers today? Are we helping each other learn?
Share your views on this matter that affects us all!
We invite everyone to take part in this learning campaign. Wikimedia Foundation will distribute one Wikimania Scholarship 2016 among those participants who are eligible.

More information

- All participants must have a registered user of at least one month antiquity on any Wikimedia project before the starting date of the campaign.
- All eligible contributions must be done until **August 23, 2015 at 23:59 UTC**
- Wiki link: **Community Health learning campaign**
- URL https://meta.wikimedia.org/wiki/Grants:Evaluation/Community_Health_learning_campaign
- Contact: Maria Cruz / Twitter: @WikiEval #CommunityHealth / email: eval@wikimedia . org

Happy editing!

MediaWiki message delivery (talk) 23:42, 31 July 2015 (UTC)

A healthy community is definitely one where people get blocked for hate speech (https://en.wikipedia.org/w/index.php?title=Special:Log&offset=20060301000000&limit=2&type=block&user=Carnildo&page=&tagfilter=&hide_patrol_log=1&hide_tag_log=1&hide_review_log 7C).Codeofdusk (talk) 07:18, 3 August 2015 (UTC)

Module:Citation/CS1 incorrectly adding pages using edition=revised to tracking category

Module:Citation/CS1 automatically adds "ed." after the value of the edition parameter. Citations that explicitly use something like "2nd ed." are added to the hidden maintenance category Category:CS1 maint: Extra text to allow fixing these values more easily. However, the module probably just checks whether the value ends with "edition", "ed" or similar. This leads to issues on pages like Pi, where a reference (No. 79, *The Penguin Dictionary of Curious and Interesting Numbers*) gets incorrectly marked as erroneous because it uses "edition=revised", which ends with "ed". Is this a bug? —Mathe314 (talk) 20:04, 2 August 2015 (UTC)

Yes, a bug that has been fixed in the sandbox. Questions and concerns about this module are best addressed at Help talk:Citation Style 1.

—Trappist the monk (talk) 20:32, 2 August 2015 (UTC)

Citations are weird

I just added a citation to same-sex marriage in Mexico for Puebla:

In November 2014 a lesbian couple filed for an amparo and were granted an injunction to marry. The state appealed the decision. 10 July, 2015, the Appellate Court upheld the ruling in favor of the couple. Their wedding, which was the first same-sex marriage in the state of Puebla<ref name="1st marriage">{{cite news|last1=Hernández Alcántara|first1=Martín|title=Mañana se celebrará el primer matrimonio gay en la historia de Puebla|url=http://www.lajornadadeoriente.com.mx/2015/07/31/manana-se-celebrara-el-primer-matrimonio-gay-en-la-historia-de-puebla/|accessdate=31 July 2015|publisher=La Jornada de Oriente|date=31 July 2015|location=Puebla, Mexico|language=Spanish}}</ref> took place on 1 August 2015.<ref name="1st marriage">{{cite news|last1=Fernández|first1=Tuss|title=Se celebra en Puebla la primera boda de personas del mismo sexo|url=http://ladobe.com.mx/2015/08/se-celebra-en-puebla-la-primera-boda-de-personas-del-mismo-sexo/|accessdate=3 August 2015|publisher=La Dobe|date=2 August 2015|location=Puebla, Mexico|language=Spanish}}</ref> As you can plainly see the text is different, but the numbers are the same: 254 is the citation number for both. Edit was made at 19:03 and it is still the same at 19:18? SusunW (talk) 00:18, 3 August 2015 (UTC)

@SusunW: This has happened because you've given both the references the same name, "1st marriage", so MediaWiki thinks your second reference is actually a reuse of the first one rather than a totally new reference. If you change the name of one of the references, it'll be fixed. Hope that helps. --(ıɐlʞ) ɐuɐʎsǝp 00:24, 3 August 2015 (UTC)

Thank you. Some days one can't see the forest for the trees ;) SusunW (talk) 00:27, 3 August 2015 (UTC)

stats.grok.se broken yet again?

- There have been no updates since the 24th. For example, this (http://stats.grok.se/ja/latest90/川田まみ). Narutolovehinata5 (contribs) 03:34, 28 July 2015 (UTC)
- http://stats.grok.se Stats from June 16 and all dates beginning on July 25 are missing even though the raw data exists.--TonyTheTiger (T / C / WP:FOUR / WP:CHICAGO / WP:WAWARD) 03:58, 28 July 2015 (UTC)
  - July stats are up, but June 16 is still missing --TonyTheTiger (T / C / WP:FOUR / WP:CHICAGO / WP:WAWARD) 04:35, 30 July 2015 (UTC)

Cannot reach stats.grok.se, name cannot be resolved. DNS lookup failed. — Preceding unsigned comment added by Dstone1029 (talk • contribs) 16:41, 3 August 2015 (UTC)

AfD Statistics

The AfD Statistics tool ([45] (https://tools.wmflabs.org/afdstats/afdstats.py?name=Vanjagenije&max=500&startdate=&altname=)) shows my vote at WP:Articles for deletion/Bob Girls discography as "keep", abut in fact I proposed the article for deletion. Why is that? Vanjagenije (talk) 09:21, 30 July 2015 (UTC)

You commented below someone else's support, which the tool probably picked up as "support". It should not do that, someone commenting on someone else's !vote is usually not a vote, or a contestation. Not sure why it ignored the nomination, though. Jo-Jo Eumerus (talk, contributions) 10:03, 30 July 2015 (UTC)

Is there somebody who can fix that? The author (Scottywong) is retired. I don't know whom to ask. Vanjagenije (talk) 10:55, 30 July 2015 (UTC)

I was bold and have tweaked the page. What's probably going on is that you used the incorrect list type to start your comment (please read WP:Accessibility#Lists). Check in a day or two to see if that fixed it. --Izno (talk) 11:58, 30 July 2015 (UTC)

@Izno: Thanks a lot for your time, but it is still the same. The problem is not fixed. Vanjagenije (talk) 00:29, 3 August 2015 (UTC)

Made a couple more tweaks. Wait a couple more days. --Izno (talk) 01:33, 3 August 2015 (UTC)

@Izno: Thank you very much. I think I found what's wrong. User:4minute lover signed his "keep" vote with a signature that contains a link to my talk page ([46] (https://en.wikipedia.org/w/index.php?title=Wikipedia:Articles_for_deletion/Bob_Girls_discography&diff=613569205&oldid=613521768)). That's very weird. He probably didn't know how to sign, but copied my signature. I changed it, and the AfD Statistic is now OK. I took a look into 4minute lover's edits, and I found that here (https://en.wikipedia.org/w/index.php?title=Wikipedia:Articles_for_deletion/CLC_(band)&diff=prev&oldid=653162295) he signed his post with a signature of User:Gene93k. Very strange behavior. Vanjagenije (talk) 11:30, 3 August 2015 (UTC)

WP:SIGFORGE. --Redrose64 (talk) 11:44, 3 August 2015 (UTC)
That one with Gene looks like he was moving the deletion sorting notice, not signing the page with a new comment. Which it's weird, but separately weird. --Izno (talk) 13:56, 3 August 2015 (UTC)

This is what is so ugly about talk pages. They are messy and unstructured. WP:Flow would be much easier for tools like this to work with. But a lot of reactionary people wants it to fail.--Anders Feder (talk) 11:08, 30 July 2015 (UTC)

Wikimedia email

Is Wikimedia email currently working? I sent an email to another user about an hour ago, ticked the box to receive a confirmation copy to my registered email address, got the on-screen confirmation that the email had been sent, but never received the confirmation email. Can this be independently checked? Dirtlawyer1 (talk) 02:37, 3 August 2015 (UTC)

The easiest way to check would be to have a different account with a different email address, and try to email that account. עוד מישהו Od Mishehu 02:50, 3 August 2015 (UTC)

Good idea, OM. I should have thought of that. I have three alternative accounts that I use to maintain separate large watch lists. I will try emailing one of those accounts. Cheers. Dirtlawyer1 (talk) 02:57, 3 August 2015 (UTC)

@Od Mishehu: About 45 minutes ago, I tried to send an email from my alternate account, Dirtlawyer2: Olympics, to my primary user account, Dirtlawyer1. I have not received either the email sent to the recipient account, or the confirmation email that the email system is supposed to generate for the sender account. About 10 minutes ago, I also sent a test email to you through the Wikipedia email system. Please let me know if you do or don't receive it. Thanks. Dirtlawyer1 (talk) 03:39, 3 August 2015 (UTC)

Some mail providers block Wikimedia mail due to the way it is sent. See for example Wikipedia:Village pump (technical)/Archive 136#Email is not working and phab:T66795. PrimeHunter (talk) 03:47, 3 August 2015 (UTC)

Better to direct people to a thread that has less of Technical 13's scary misinformation, like Wikipedia:Village pump (technical)/Archive 129#Is "Email this user" on the blink? and the threads linked back from there. --Redrose64 (talk) 10:23, 3 August 2015 (UTC)

In my opinion, it would be better to direct people to http://www.ietf.org/mail-archive/web/ietf/current/msg87153.html with a brief explanation that Yahoo mail is broken and that there is little we can do about it except possibly block Yahoo mail. --Guy Macon (talk) 12:51, 3 August 2015 (UTC)

As I've suggested a few times since this problem presented itself, we could resolve this issue by correctly and accurately identifying the sender (albeit by proxy) of the email as Wikimedia and provide the email address of the Wikipedia user who requested Wikimedia send the email in the reply-to field.
Alternatively, each user with email enable could be assigned virtual email address that Wikimedia would route to their actual email. xeno@users.wikimedia.org, for example. --xeno(talk) 13:00, 3 August 2015 (UTC)

Upload file from WikiEditor...to Commons

Hi there, Wikipedians!

First, this is not a very polished script, so forgive the omissions and thrown-together UI.

At Wikimania, I wrote a gadget that uploaded from enwiki beta to Commons beta. And it worked. And it was brilliant.

Now, the code required for that gadget is on enwiki proper. So I ported the script over to enwiki.

All you need is the following line in your common.js to enable it:

```
importScript( 'User:MarkTraceur/editPageUploadTool.js' );
```

After that, a second "insert file" icon will appear in WikiEditor after a second or two, and its title text will be "upload file" or something.

Note, this is going to Commons, under a CC-BY-SA 4.0 license, so by reading this and installing my script you're agreeing to those terms when you upload...see, I told you it was unpolished. In the Future, it will have a license disclaimer.

This is going to be the very rough basis for our upcoming tools, which will live both in WikiEditor and in VisualEditor. They will definitely have proper license disclaimers.

In this grand tradition, you can create subclasses of my mw.Upload object(s), and my mw.UploadDialog object(s), and write specialized upload tools for various purposes. Want to add a special category to images in a class of articles that you edit often? Easy! Subclass mw.CommonsUploadForEditDialog to return a subclass of mw.CommonsUploadForEdit which adds the category in automatically on creation. This is only one example of the cool, specialized stuff you can do.

See the documentation for mw.Upload on doc.wikimedia.org: https://doc.wikimedia.org/mediawiki-core/master/js/#!/api/mw.Upload

Happy hacking! --MarkTraceur (talk) 15:07, 3 August 2015 (UTC)

Wikitables - unable to disambiguate links

Wikitables on the following articles Athletics at the 1962 British Empire and Commonwealth Games – Women's 100 yards, Athletics at the 1968 Summer Olympics – Women's 400 metres, Athletics at the 1962 British Empire and Commonwealth Games – Women's 220 yards, and Australia at the 1962 British Empire and Commonwealth Games generate links to Joyce Bennett, but they should link to Joyce Bennett (athlete) (a legitimate redlink). I cannot see how to edit the wikitable to make them generate the correct links. Help please! DuncanHill (talk) 15:27, 3 August 2015 (UTC)

> Like so (https://en.wikipedia.org/w/index.php?title=Athletics_at_the_1962_British_Empire_and_Commonwealth_Games_%E2%80%93_Women%27s_100_yards&type=revision&diff=674350850&oldid=605861193), per the Template:Sortname documentation. --Izno (talk) 15:31, 3 August 2015 (UTC)
>
> > You can also do this using `|dab=athlete` as opposed to my solution for each of the template uses. I think my solution is marginally easier to understand. YMMV. --Izno (talk) 15:33, 3 August 2015 (UTC)
> >
> > > Many thanks - I went with the first method. DuncanHill (talk) 15:46, 3 August 2015 (UTC)

✓ Resolved

Tech News: 2015-32

Latest tech news from the Wikimedia technical community. Please tell other users about these changes. Not all changes will affect you. Translations are available.

Recent changes

- The ContentTranslation extension has been updated:
  - You can now type in list paragraphs on wikis with right-to-left scripts. [47] (https://phabricator.wikimedia.org/T103504)
  - Some user interface updates changed how details in the ContentTranslation extension look. [48] (https://phabricator.wikimedia.org/T106241) [49] (https://phabricator.wikimedia.org/T105799) [50] (https://phabricator.wikimedia.org/T106643) [51] (https://phabricator.wikimedia.org/T106437)
- The error page you see when the sites are not working is now simpler and easier to read. It also shows the Wikimedia logo. [52] (https://gerrit.wikimedia.org/r/#/c/223012/)
- MediaWiki now supports redirects for CSS pages. [53] (https://phabricator.wikimedia.org/T73201)
- Bots can't guess captchas unlimited times anymore. [54] (https://phabricator.wikimedia.org/T92376)

Problems

- There was a problem with thumbnails on wikis with local images on July 24 and 25. It was due to a code error. [55] (https://phabricator.wikimedia.org/T106895)
- We can't search by file type on Wikimedia wikis now. [56] (https://phabricator.wikimedia.org/T107265)

Changes this week

- The new version of MediaWiki will be on test wikis and MediaWiki.org from August 4. It will be on non-Wikipedia wikis from August 5. It will be on all Wikipedias from August 6 (calendar).

Meetings

- You can join the next meeting with the VisualEditor team. During the meeting, you can tell developers which bugs are the most important. The meeting will be on August 4 at 19:00 (UTC) (http://www.timeanddate.com/worldclock/fixedtime.html?hour=19&min=00&sec=0&day=04&month=08&year=2015). See how to join.

Tech news prepared by tech ambassadors and posted by bot • Contribute • Translate • Get help • Give feedback • Subscribe or unsubscribe.

15:51, 3 August 2015 (UTC)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Wikipedia:Village_pump_(technical)/Archive_138&oldid=721364514"

This page was last edited on 21 May 2016, at 09:59.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 14

The New York Times

OP-ED CONTRIBUTORS

Stop Spying on Wikipedia Users

By Jimmy Wales and Lila Tretikov

March 10, 2015

SAN FRANCISCO — TODAY, we're filing a lawsuit against the National Security Agency to protect the rights of the 500 million people who use Wikipedia every month. We're doing so because a fundamental pillar of democracy is at stake: the free exchange of knowledge and ideas.

Our lawsuit says that the N.S.A.'s mass surveillance of Internet traffic on American soil — often called "upstream" surveillance — violates the Fourth Amendment, which protects the right to privacy, as well as the First Amendment, which protects the freedoms of expression and association. We also argue that this agency activity exceeds the authority granted by the Foreign Intelligence Surveillance Act that Congress amended in 2008.

Most people search and read Wikipedia anonymously, since you don't need an account to view its tens of millions of articles in hundreds of languages. Every month, at least 75,000 volunteers in the United States and around the world contribute their time and passion to writing those articles and keeping the site going — and growing.

On our servers, run by the nonprofit Wikimedia Foundation, those volunteers discuss their work on everything from Tiananmen Square to gay rights in Uganda. Many of them prefer to work anonymously, especially those who work on controversial issues or who live in countries with repressive governments.

These volunteers should be able to do their work without having to worry that the United States government is monitoring what they read and write. Unfortunately, their anonymity is far from certain because, using upstream surveillance, the N.S.A. intercepts and searches virtually all of the international text-based traffic that flows across the Internet "backbone" inside the United States. This is the network of fiber-optic cables and junctions that connect Wikipedia with its global community of readers and editors.

As a result, whenever someone overseas views or edits a Wikipedia page, it's likely that the N.S.A. is tracking that activity — including the content of what was read or typed, as well as other information that can be linked to the person's physical location and possible identity. These activities are sensitive and private: They can reveal everything from a person's political and religious beliefs to sexual orientation and medical conditions.

You have 4 free articles remaining.
Subscribe to The Times

The notion that the N.S.A. is monitoring Wikipedia's users is not, unfortunately, a stretch of the imagination. One of the documents revealed by the whistle-blower Edward J. Snowden specifically identified Wikipedia as a target for surveillance, alongside several other major websites like CNN.com, Gmail and Facebook. The leaked slide from a classified PowerPoint presentation declared that monitoring these sites could allow N.S.A. analysts to learn "nearly everything a typical user does on the Internet."

The harm to Wikimedia and the hundreds of millions of people who visit our websites is clear: Pervasive surveillance has a chilling effect. It stifles freedom of expression and the free exchange of knowledge that Wikimedia was designed to enable.

During the 2011 Arab uprisings, Wikipedia users collaborated to create articles that helped educate the world about what was happening. Continuing cooperation between American and Egyptian intelligence services is well established; the director of Egypt's main spy agency under President Abdel Fattah el-Sisi boasted in 2013 that he was "in constant contact" with the Central Intelligence Agency.

So imagine, now, a Wikipedia user in Egypt who wants to edit a page about government opposition or discuss it with fellow editors. If that user knows the N.S.A. is routinely combing through her contributions to Wikipedia, and possibly sharing information with her government, she will surely be less likely to add her knowledge or have that conversation, for fear of reprisal.

And then imagine this decision playing out in the minds of thousands of would-be contributors in other countries. That represents a loss for everyone who uses Wikipedia and the Internet — not just fellow editors, but hundreds of millions of readers in the United States and around the world.

In the lawsuit we're filing with the help of the American Civil Liberties Union, we're joining as a fellow plaintiff a broad coalition of human rights, civil society, legal, media and information organizations. Their work, like ours, requires them to engage in sensitive Internet communications with people outside the United States.

That is why we're asking the court to order an end to the N.S.A.'s dragnet surveillance of Internet traffic.

Privacy is an essential right. It makes freedom of expression possible, and sustains freedom of inquiry and association. It empowers us to read, write and communicate in confidence, without fear of persecution. Knowledge flourishes where privacy is protected.

Jimmy Wales, the founder of Wikipedia, is a board member of the Wikimedia Foundation, of which Lila Tretikov is the executive director.

A version of this article appears in print on March 10, 2015, on Page A21 of the New York edition with the headline: Stop Spying on Wikipedia Users

READ 403 COMMENTS

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 15

Today, the Wikimedia Foundation is filing suit aga
National Security Agency (NSA) and the Departme
Justice (DOJ) of the United States. The lawsuit ch
the NSA's mass surveillance program. (..)

FOUNDATION, FREE CULTURE, FROM THE ARCHIVES, LEGAL, WIKIMEDIA V. NSA

Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance

By Michelle Paulson, Wikimedia Foundation
Geoff Brigham, Wikimedia Foundation
March 10th, 2015

Today, the Wikimedia Foundation is filing suit against the National Security Agency (NSA) and the Department of Justice (DOJ) of the United States. The lawsuit challenges the NSA's mass surveillance program. (..)

THIS ARTICLE IS AVAILABLE IN:
ENGLISH 日本語 中文 FRANÇAIS DEUTSCH ITALIANO
JĘZYK POLSKI PORTUGUÊS DO BRASIL ESPAÑOL

Justice presides with her scale and sword in Frankfurt am Main.
Photo by Roland Meinecke, licensed under a CC BY 2.0 license.

Today, the Wikimedia Foundation is filing suit against the National Security Agency (NSA) and the Department of Justice (DOJ) of the United States [1]. The lawsuit challenges the NSA's mass surveillance program, and specifically its large-scale search and seizure of internet communications — frequently referred to as "upstream" surveillance. Our aim in filing this suit is to end this mass surveillance program in order to protect the rights of our users around the world. We are joined by eight other organizations [2] and represented by the American Civil Liberties Union (ACLU). The full complaint can be found here.

GET CONNECTED

GET OUR EMAIL UPDATES

Your email address

Subscribe

MEET OUR COMMUNITY

Meet the scientist working to increase the number of underrepresented scientists and engineers on Wikipedia

For these Wiki editors, ancien Egyptian pyra are more than symbol

More Community Profiles

MOST VIEWED THIS MONTH

Türkiye'den Vikipedi'ye erişim engeli halen devam ediyor
Vikipedi'nin tüm dil sürümleri, Nisan ayını

Love is strange: ten weird Valentine's facts from Wikiped
Roses are red Divorce is contractual...

Introducing the unique devices dataset: a new way to estimat reach on Wikimedia projects
With the unique devices dataset, we've...

ARCHIVES

AUGUST 2018

JULY 2018

"We're filing suit today on behalf of our readers and editors everywhere," said Jimmy Wales, founder of Wikipedia. "Surveillance erodes the original promise of the internet: an open space for collaboration and experimentation, and a place free from fear."

Privacy is the bedrock of individual freedom. t is a universal right that sustains the freedoms of expression and association. These principles enable inquiry, dialogue, and creation and are central to Wikimedia's vision of empowering everyone to share in the sum of all human knowledge. When they are endangered, our mission is threatened. If people look over their shoulders before searching, pause before contributing to controversial articles, or refrain from sharing verifiable but unpopular information, Wikimedia and the world are poorer for it.

When the 2013 public disclosures about the NSA's activities revealed the vast scope of their programs, the Wikimedia community was rightfully alarmed. In 2014, the Wikimedia Foundation began conversations with the ACLU about the possibility of filing suit against the NSA and other defendants on behalf of the Foundation, its staff, and its users.

Our case today challenges the NSA's use of upstream surveillance conducted under the authority of the 2008 Foreign Intelligence Surveillance Act Amendments Act (FAA). Upstream surveillance taps the internet's "backbone" to capture communications with "non-U.S. persons." The FAA authorizes the collection of these communications if they fall into the broad category of "foreign intelligence information" that includes nearly any information that could be construed as relating to national security or foreign affairs. The program casts a vast net, and as a result, captures communications that are not connected to any "target," or may be entirely domestic. This includes communications by our users and staff.

"By tapping the backbone of the internet, the NSA is straining the backbone of democracy," said Lila Tretikov, executive director of the Wikimedia Foundation. "Wikipedia is founded on the freedoms of expression, inquiry, and information. By violating our users' privacy, the NSA is threatening the intellectual freedom that is central to people's ability to create and understand knowledge."

The NSA has interpreted the FAA as offering free rein to define threats, identify targets, and monitor people, platforms, and infrastructure with little regard for probable cause or proportionality. We believe that the NSA's current practices far exceed the already broad authority granted by the U.S. Congress through the FAA. Furthermore, we believe that these practices violate the U.S. Constitution's First Amendment, which protects freedom of speech and association, and the Fourth Amendment, which protects against unreasonable search and seizure.

Additionally, we believe that the NSA's practices and limited judicial review of those practices violate Article III of the U.S. Constitution. A specialized court, the Foreign Intelligence Surveillance Court (FISC), hears issues related to foreign intelligence requests, including surveillance. Under U.S. law, the role of the courts is to resolve "cases" or "controversies" — not to issue advisory opinions or interpret theoretical situations. In the context of upstream surveillance, FISC proceedings are not "cases." There are no opposing parties and no actual "controversy" at stake. FISC merely reviews the legality of the government's proposed procedures — the kind of advisory opinion that Article III was intended to restrict.

In 2013, the U.S. Supreme Court dismissed a previous challenge to the FAA, *Amnesty v. Clapper*, because the parties in that case were found to lack "standing." Standing is an important legal concept that requires a party to show that they've suffered some kind of harm in order to file a lawsuit. The 2013 mass surveillance disclosures included a slide from a classified NSA presentation that made explicit reference to Wikipedia, using our global trademark. Because these disclosures revealed that the government specifically targeted Wikipedia and its users, we believe we have more than sufficient evidence to establish standing.

Wikipedia is the largest collaborative free knowledge resource in human history. t represents what we can achieve when we are open to possibility and unburdened by fear. Over the past fourteen years, Wikimedians have written more than 34 million articles in 288 different languages. Every month, this knowledge is accessed by nearly half a billion people from almost every country on

JUNE 2018

MAY 2018

APRIL 2018

OLDER POSTS 27

WORK AT WIKIMEDIA

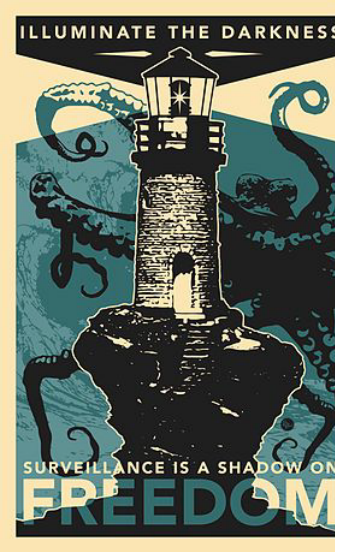
Work with the foundation that supports W and its sister projects around the world. A and join us

earth. This dedicated global community of users is united by their passion for knowledge, their commitment to inquiry, and their dedication to the privacy and expression that makes Wikipedia possible. We file today on their behalf.

For more information, please see our op-ed, *Stop Spying on Wikipedia Users*, by Wikipedia founder Jimmy Wales, and Wikimedia Foundation executive director Lila Tretikov, in the March 10 edition of *The New York Times*. [3]

*Michelle Paulson, Senior Legal Counsel, Wikimedia Foundation **
Geoff Brigham, General Counsel, Wikimedia Foundation

* The Wikimedia Foundation and its co-plaintiffs are being represented by the American Civil Liberties Union (ACLU) in this suit. We would like to thank them, in particular Patrick Toomey, Ashley Gorski, and Daniel Kahn Gillmor for their work and dedication throughout this process.



Spread the word about inappropriate surveillance. Art by Rich Black, CC BY 3.0.

References

1. Other defendants include: Michael Rogers, in his official capacity as Director of the National Security Agency and Chief of the Central Security Service; Office of the Director of National Intelligence; James Clapper, in his official capacity as Director of National Intelligence; and Eric Holder, in his official capacity as Attorney General of the United States.
2. Today, we're proud to bring this lawsuit alongside a coalition of organizations from across the ideological spectrum, including The National Association of Criminal Defense Lawyers, Human Rights Watch, Amnesty International USA, Pen American Center, Global Fund for Women, The Nation Magazine, The Rutherford Institute, and Washington Office on Latin America. We believe the wide variety of perspectives represented in this lawsuit demonstrates that the defense of privacy and freedom of expression and association is not defined by partisanship or ideology.
3. To read more about our opposition to mass government surveillance, please see our previous blog posts on PRISM, opposing mass surveillance on the internet, and transparency in the use of surveillance.

Frequently Asked Questions

Q What does this lawsuit challenge?
A: Our lawsuit challenges the NSA's unfounded, large-scale search and seizure of internet communications, frequently referred to as "upstream" surveillance. Using upstream surveillance, the NSA intercepts virtually all internet communications flowing across the network of high-capacity cables, switches, and routers that make up the internet's "backbone." This backbone connects the Wikimedia global community of readers and contributors to Wikipedia and the other the Wikimedia projects.

Q What is the U.S. government's legal justification for this program?
A: The U.S. government has used the Foreign Intelligence Surveillance Act (FISA) Amendments Act of 2008 (FAA) (see 50 U.S.C. § 1881a) to justify broad, "upstream" mass surveillance. Under the FAA, "the Attorney General and the Director of National Intelligence may authorize jointly, for a period of up to one year from the effective date of the authorization, the targeting of [non-US] persons reasonably believed to be located outside the United States to acquire foreign intelligence information." The statute only requires "reasonable belief" that a

non-US person is located outside the United States. There is no need to show that target is a foreign agent, much less a terrorist. The purpose of the statute is to acquire "foreign intelligence information"– a very general concept. We believe the broad interpretation of this statute that allows for upstream surveillance is unconstitutional.

Q How does surveillance or the fear of surveillance affect readers and editors of Wikipedia and its sister projects?

A: Mass surveillance is a threat to intellectual freedom and a spirit of inquiry, two of the driving forces behind Wikimedia. Wikipedia is written by people from around the world who often tackle difficult subjects. Very frequently they choose to remain anonymous, or pseudonymous. This allows them to freely create, contribute, and discover, without fear of reprisal. Surveillance might be used to reveal sensitive information, create a chilling effect to deter participation, or in extreme instances, identify individual users. Pervasive surveillance undermines the freedoms upon which Wikipedia and its communities are founded.

Q How does surveillance affect Wikipedia as a knowledge resource?

A: Wikipedia is a living resource for knowledge. It is written by volunteers around the globe, in hundreds of languages. It reflects the world around us and changes to embodies current events, notable individuals, evolving theories, emerging art, and more. Wikipedia relies on the contributions of editors and the support of readers to evolve and grow. If readers and editors are deterred from participating in Wikipedia because of concerns about surveillance, the health of Wikipedia as a resource to the world is jeopardized.

Q What kind of Wikimedia communications could the NSA be intercepting?

A: Wikipedia and its sister projects is created entirely by volunteer editors. More than 75,000 editors each month edit Wikipedia, amounting to more than 33 million articles. These editors not only contribute content, but also discuss and share information on discussion pages and elsewhere within the project. Privacy and free expression are core values of the Wikimedia community. When volunteer editors contribute to Wikipedia, they expect it to be a safe, open space in which creativity and knowledge can thrive.

Q Why is it important that the Wikimedia Foundation ensures privacy and anonymity for its users?

A: Privacy is a core value of the Wikimedia movement. From the beginning, Wikipedia has allowed for users to maintain private identities through the use of anonymous or pseudonymous editing. This has been reinforced by the Wikimedia Foundation's firm commitment to protecting the privacy and data of its users through legal and technical means.

Privacy makes freedom of expression possible, sustains freedom of inquiry, and allows for freedom of information and association. Knowledge flourishes where privacy is protected.

Q Why is the NSA interested in the communications of innocent Wikimedia users?

A: You would have to ask them. One could guess, however, that they are trying to amass as much information as possible into their databases, and, as with other websites, they may believe there is value in the data, conversations, and personal information on Wikipedia and in the Wikimedia community.

Q How do you know Wikimedia has been singled out for surveillance by the NSA?

A: One of the NSA documents revealed by whistle-blower Edward Snowden specifically identifies Wikipedia for surveillance alongside several other major websites like CNN.com, Gmail, and Facebook. The previously secret slide declares that monitoring these sites can allow NSA analysts to learn "nearly everything a typical user does on the Internet."

Q Has the Wikimedia Foundation taken any measures to protect its users' privacy?

A: The Wikimedia Foundation takes privacy very seriously, which is why we find the NSA's upstream mass surveillance so troubling. You do not need to create an account or login to read or edit Wikipedia or the other Wikimedia sites. If you do decide to create an account, you can choose any username you like — we don't require real names, email addresses, or any other personally identifying information, and we never sell your data.

Q Why did Wikimedia join this lawsuit against the NSA?

A: Our role at the Wikimedia Foundation is to protect Wikipedia, its sister projects, and the Wikimedia community of users. This means providing our users with the right conditions to facilitate their work, and protecting them

when necessary. Defending the privacy of our editors, readers, and community is paramount to us. We believe privacy is essential to facilitating and advancing free knowledge.

You can also find this FAQ here on Wikimedia.org.

29 Comments on Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance

arielsbecker 4 years

This is very good news! And a surprise too. Congratulations, and all my support, from Argentina.

Share

Philippe Verdy 4 years

Note that France is currently in the process of adopting a system similar to the NSA, by placing the surveillance under the direct autority of the government, without real control by the justice.
There will be NO parties in a judiciary suite, so nobody will be able to defend his case. Everyone will be monitored for various purposes and not just for national security : it will include spying on commercial communications, profiling any one on any communication network (whever in France or abroad), by wiretappinbg also directly on the backbones and in all internal network of network providers or security providers (those that emit encryption keys for the PKI infrastructure).
Today, string encruption is allowed in France only for the strong authentication of users (but this is not a problem for spying agencies, given that strong authentication allows identifying anyone), but not for the content of their communications (which use weak encryption mechanisms with authorized “key escrows”): if users are strongly authentified but their communication are not really encypted, they can be profiled in all their online activities associated to this strong authentication.

The mechanisms used in HTTPS (and PSEC) have also backdoors, allowing a third party to reduce the strength of engryptions, or substituting encryption keys by others, provided that they have an unlimited access to the backbones and internal networks used by commercial network service providers (Internet, mobile phone, electronic payment systems, externalized billing systems, cloud storage or computing providers). They will be able to track each new account creation on these services, recording a unique identifier of the user that will be usable later to decrypt his tapped communications.

There’s also no limit of time for keeping the records and judges won’t have access to the contents of these records (because it will be only under the direct supervision and authority of the government, which will be able to oppose a “secret défense” to a judiciary injunction for getting access to the records.

Here also, there will be a pseudo-judiciary control with an adhoc advisory committee similar to the “FLIC” in USA, where not all parties will be heard (in fact there will be no judiciary suite, and none of the advices will be enforceable (the governement consults, but does not have to follow any decision, it will allow its administration to do everything it wants and will protect all their action, maintaining the secret).

Funny side node: “FLIC” in French is the common popular word for designating a member of the police (national police, or gendarmerie). It is no longer a “slang” word, even the police uses it. It is perfectly legal in France to call a policeman a “flic”. I wonder if the US acronym is not made on purpose to reuse this very common French term !

But clearly its meaning has nothing in common with the justice, it is part of the executive administration system (under control of the political party in the current majority of government, or the current presidential majority for some branches of the police in France: if these two majorities are not the same, each one has his own privileges and can act independantly without any real judiciary control, or even any parliamentary control for the presidential branch).

What this means is that the democratic separation of powers (theoretically claimed and protected by the French constitution) has another breach: this is a real threat against journalists, lawyers, political parties not in the current majority, and against all citizens or legal foreign residents in France (including EU nationals). That law will also permit spying on other EU member countries, for their domestic activities or

No. 20-1191

UNITED STATES COURT OF APPEALS
FOR THE FOURTH CIRCUIT

WIKIMEDIA FOUNDATION,

Plaintiff–Appellant,

v.

NATIONAL SECURITY AGENCY, *et al.*,

Defendants–Appellees.

On Appeal from the United States District Court
for the District of Maryland at Baltimore

JOINT APPENDIX—VOLUME 7 OF 7 (JA3879–JA4124)

H. Thomas Byron III
Joseph Busa
Michael Shih
U.S. DEPARTMENT OF JUSTICE
950 Pennsylvania Ave. NW
Washington, DC 20530
Phone: (202) 616-5367
Fax: (202) 307-2551
h.thomas.byron@usdoj.gov

Counsel for Defendants–Appellees

Patrick Toomey
Ashley Gorski
Charles Hogle
AMERICAN CIVIL LIBERTIES
UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
ptoomey@aclu.org

Counsel for Plaintiff–Appellant
(Additional counsel on next page)

Alex Abdo
Jameel Jaffer
KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY
475 Riverside Drive, Suite 302
New York, NY 10115
Phone: (646) 745-8500
alex.abdo@knightcolumbia.org

Deborah A. Jeon
David R. Rocah
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF MARYLAND
3600 Clipper Mill Rd., #350
Baltimore, MD 21211
Phone: (410) 889-8555
Fax: (410) 366-7838
rocah@aclu-md.org

Benjamin H. Kleine
COOLEY LLP
101 California Street, 5th Floor
San Francisco, CA 94111
Phone: (415) 693-2000
Fax: (415) 693-2222
bkleine@cooley.com

***Wikimedia Foundation v. National Security Agency, et al.*,**
No. 20-1191 (4th Cir.)

JOINT APPENDIX
Table of Contents

VOLUME 1

U.S. District Court for the District of Maryland, Docket Sheet,
Case No. 1:15-cv-00662 ..... JA0001

Plaintiff Wikimedia Foundation's Amended Complaint
(June 22, 2015), ECF No. 72 ..... JA0036

Exhibits to Wikimedia Foundation's Motion to Compel

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation
(Mar. 26, 2018), ECF No. 125-3 ..... JA0096

Exhibit 1: Chart Identifying Discovery Requests at Issue on
Wikimedia Foundation's Motion to Compel,
ECF No. 125-4..... JA0101

Exhibit 2: Wikimedia Foundation's Requests for Admission
and attachments (Nov. 7, 2017), ECF No. 125-5 ..... JA0118

Exhibits to Defendants' Opposition to Wikimedia Foundation's Motion to Compel

Declaration of Daniel R. Coats, Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-2 ..... JA0170

Declaration of Lauren L. Bernick, Senior Associate Civil Liberties
Protection Officer in the Office of Civil Liberties, Privacy, and
Transparency at the Office of the Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-3 ..... JA0190

Notice of Filing Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141 ..... JA0199

Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141-1 ..... JA0201

Exhibits to Wikimedia Foundation's Reply in Support of Its Motion to Compel

Declaration of Ashley Gorski, Counsel for Wikimedia Foundation (May 18, 2018), ECF No. 143-1 ..... JA0270

Exhibit 1: Chart Identifying Deposition Questions at Issue on Wikimedia Foundation's Motion to Compel, ECF No. 143-2 ..... JA0272

Exhibit 2: Transcript of Deposition of NSA's Designated Witness, Rebecca J. Richards, Pursuant to Fed. R. Civ. P. 30(b)(6) (Apr. 16, 2018), ECF No. 143-3 ..... JA0286

Opinion & Order Denying Wikimedia Foundation's Motion to Compel

Memorandum Opinion (Aug. 20, 2018), ECF No. 150 ..... JA0689

Order Denying Plaintiff's Motion to Compel Discovery Responses & Deposition Testimony (Aug. 20, 2018), ECF No. 151 ..... JA0716

Exhibits to Defendants' Motion for Summary Judgment

Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Nov. 13, 2018), ECF No. 164-4 ..... JA0719

Declaration of James Gilligan, Counsel for Defendants (Nov. 13, 2018), ECF No. 164-5 ..... JA0818

Exhibit 3: Wikimedia Foundation's Amended and Supplemental Responses and Objections to NSA's First Set of Interrogatories (Mar. 23, 2018), ECF No. 164-6 ..... JA0821

Exhibit 4: Wikimedia Foundation's Amended Responses and Objections to ODNI's Interrogatory No. 19 (Apr. 6, 2018), including Technical Statistics Chart, ECF No. 164-7 ..... JA0861

Exhibit 5: Wikimedia Foundation's Responses and Objections to NSA's First Set of Interrogatories (Jan. 11, 2018), ECF No. 164-8..... JA0876

VOLUME 2

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment

Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Dec. 18, 2018), ECF No. 168-2 ..... JA0920

Appendices A through Z to Declaration of Scott Bradner (Dec. 18, 2018), ECF Nos. 168-3 to 168-4 ..... JA1067

VOLUME 3

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Appendices AA through FF to Declaration of Scott Bradner (Dec. 18, 2020), ECF No. 168-5 ..... JA1791

Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Dec. 18, 2018), ECF No. 168-6 ..... JA2151

Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-7 ..... JA2218

Declaration of James Alexander, Former Manager for Trust and Safety and Former Legal and Community Advocacy Manager at Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-8 ..... JA2244

Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Dec. 18, 2018), ECF No. 168-9 ..... JA2253

Declaration of Emily Temple-Wood (Dec. 18, 2018), ECF No. 168-10 ..... JA2268

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-11 ..... JA2278

Exhibit 8: Wikimedia-hosted email list discussing NSA slide with Wikimedia logo, from July to August 2013, ECF No. 168-12 ..... JA2283

Exhibit 9: Wikimedia "Talk page" discussing its non-public information policy, from September to December 2013, ECF No. 168-13 ..... JA2305

Exhibit 10: "OTRS" ticket showing Wikimedia user requesting Tor permissions in September 2013, ECF No. 168-14 ..... JA2349

VOLUME 4

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 11: Wikimedia webpage showing Wikimedia user requesting Tor permissions in September 2017, ECF No. 168-15 ..... JA2353

Exhibit 12: Wikimedia document compiling German-user-

community appeal concerning privacy in 2013,
ECF No. 168-16 ..... JA2357

Exhibit 13: Wikimedia "Talk page" discussing NSA
surveillance from June to December 2013,
ECF No. 168-17 ..... JA2363

Exhibit 14: Wikimedia Technical Statistics Chart & Supporting
Exhibits A-G, ECF No. 168-18 ..... JA2396

Exhibit 15: Privacy & Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* (July 2014), ECF No. 168-19 ..... JA2434

Exhibit 16: FISC Memorandum Opinion, [*Redacted*], 2011 WL
10945618 (Oct. 3, 2011), ECF No. 168-20 ..... JA2631

Exhibit 17: Office of the Director of National Intelligence, *DNI Declassifies Intelligence Community Documents Regarding Collection Under Section 702 of FISA* (Aug. 21, 2013),
ECF No. 168-21 ..... JA2717

Exhibit 18: Defendant NSA's Objections and Responses to
Plaintiff's First Set of Interrogatories (Dec. 22, 2017),
ECF No. 168-22 ..... JA2721

Exhibit 19: FISC Submission, *Clarification of National Security Agency's Upstream Collection Pursuant to Section 702 of FISA*
(May 2, 2011), ECF No. 168-23 ..... JA2743

Exhibit 20: Office of the Director of National Intelligence,
Statistical Transparency Report Regarding Use of National Security Authorities, Calendar Year 2017 (Apr. 2018),
ECF No. 168-24 ..... JA2748

Exhibit 21: FISC Memorandum Opinion & Order
(Apr. 26, 2017), ECF No. 168-25 ..... JA2790

VOLUME 5

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 22: FISC Submission, *Government's Response to the Court's Briefing Order of May 9, 2011* (June 1, 2011), ECF No. 168-26 ..... JA2890

Exhibit 23: *Big Brother Watch & Others v. United Kingdom*, App. Nos. 58170/13, 62322/14, 24960/15, Eur. Ct. H.R. (2018), ECF No. 168-27 ..... JA2932

Exhibit 24: NSA Director of Civil Liberties & Privacy Office, *NSA's Implementation of FISA Section 702* (Apr. 16, 2014), ECF No. 168-28 ..... JA3145

Exhibit 25: *Legal Issues Relating to the Testing, Use, and Deployment of an Intrusion-Detection System (EINSTEIN 2.0)*, 33 Op. O.L.C. 1 (Jan. 9, 2009), ECF No. 168-29 ..... JA3157

Exhibit 26: Minimization Procedures Used by the NSA in Connection with Acquisitions of Foreign Intelligence Information Pursuant to Section 702 of FISA (July 2014), ECF No. 168-30 ..... JA3193

Exhibit 27: Glenn Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet,"* Guardian, July 31, 2013, ECF No. 168-31 ..... JA3209

Exhibit 28: NSA slide, excerpted from Exhibit 27 (Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet"*), ECF No. 168-32 ..... JA3220

Exhibit 29: Morgan Marquis-Boire, et al., *XKEYSCORE: NSA's Google for the World's Private Communications*, Intercept, July 1, 2015, ECF No. 168-33 ..... JA3222

Exhibit 30: NSA slide deck, *XKEYSCORE for Counter-CNE*, published in The Intercept on July 1, 2015, ECF No. 168-34 ... JA3237

Exhibit 31: Wikimedia, *Founding Principles* (accessed Mar. 14, 2018), ECF No. 168-35 ..... JA3259

Exhibit 32: Yana Welinder, *Opposing Mass Surveillance on the Internet*, Wikimedia Blog (May 9, 2014), ECF No. 168-36 ..... JA3262

Exhibit 33: Wikimedia Public Policy, *Privacy* (accessed Mar. 14, 2018), ECF No. 168-37 ..... JA3266

Exhibit 34: Wikipedia, *Sock Puppetry* (accessed Mar. 14, 2018), ECF No. 168-38 ..... JA3273

Exhibit 35: Wikimedia, *Privacy Policy* (accessed Feb. 14, 2018), ECF No. 168-39 ..... JA3286

Exhibit 36: Ryan Lane, *The Future of HTTPS on Wikimedia Projects*, Wikimedia Blog (Aug. 1, 2013), ECF No. 168-40 ..... JA3311

Exhibit 37: Yana Welinder, et al., *Securing Access to Wikimedia Sites with HTTPS*, Wikimedia Blog (June 12, 2015), ECF No. 168-41 ..... JA3317

Exhibit 38: Wikimedia email describing Tech/Ops goals and the importance of HTTPS (May 23, 2014), ECF No. 168-42 .... JA3325

Exhibit 39: Wikimedia document discussing IPsec implementation, including July 8, 2013 statement from a Wikimedia engineer, ECF No. 168-43 ..... JA3328

Exhibit 40: Wikimedia job posting for Traffic Security Engineer (accessed Feb. 8, 2018), ECF No. 168-44 ..... JA3364

Exhibit 41: Michelle Paulson, *A Proposal for Wikimedia's New Privacy Policy and Data Retention Guidelines*, Wikimedia Blog (Feb. 14, 2014), ECF No. 168-45 ..... JA3367

Exhibit 42: Wikimedia's Supplemental Exhibit C in response

to NSA Interrogatory No. 8 (volume of HTTP border-crossing communications by country), ECF No. 168-46 ..... JA3375

Exhibit 43: Wikimedia's Supplemental Exhibit D in response to NSA Interrogatory No. 8 (volume of HTTPS border-crossing communications by country), ECF No. 168-47 ..... JA3388

Exhibit 44: Wikimedia analytics document showing monthly unique visitors to Wikimedia by region, from December 2007 to May 2015, ECF No. 168-48 ..... JA3400

Exhibit 45: Press Release, NSA, *NSA Stops Certain Section 702 "Upstream" Activities*, Apr. 28, 2017, ECF No. 168-49..... JA3404

VOLUME 6

Exhibits to Defendants' Reply in Support of Their Motion for Summary Judgment

Second Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Feb. 15, 2019), ECF No. 178-2 ..... JA3407

Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Feb. 15, 2019), ECF No. 178-3 ..... JA3452

Second Declaration of James Gilligan, Counsel for Defendants (Feb. 15, 2019), ECF No. 178-4 ..... JA3725

Exhibit 9: Wikimedia Foundation's Responses and Objections to DOJ's First Set of Interrogatories (Jan. 11, 2018), ECF No. 178-5..... JA3728

Exhibit 10: Relevant Portions of the Deposition of James Alexander, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-6..... JA3761

Exhibit 11: Relevant Portions of the Deposition of Michelle

Paulson, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-7 ..... JA3777

Exhibit 12: Wikimedia Foundation, *Securing access to Wikimedia sites with HTTPS*, June 12, 2015 (WIKI0007108-7114), ECF No. 178-8 ..... JA3791

Exhibit 13: Wikipedia: Village pump (technical)/Archive 138 (WIKI0006872-6938), ECF No. 178-9 ..... JA3800

Exhibit 14: Jimmy Wales and Lila Tretikov, "Stop Spying on Wikimedia Users," N.Y. Times, Mar. 10, 2015, ECF No. 178-10 ..... JA3869

Exhibit 15: Wikimedia Foundation, *Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance*, Mar. 10, 2015, ECF No. 178-11 ..... JA3873

VOLUME 7

Exhibits to Wikimedia Foundation's Sur-reply in Opposition to Defendants' Motion for Summary Judgment

Second Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Mar. 8, 2019), ECF No. 181-1 ..... JA3879

Second Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Mar. 8, 2019), ECF No. 181-2 ..... JA3940

Second Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Mar. 8, 2019), ECF No. 181-3 ..... JA4006

Second Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Mar. 8, 2019), ECF No. 181-4 ..... JA4012

Second Declaration of Emily Temple-Wood (Mar. 8, 2019),
ECF No. 181-5 ..... JA4015

Exhibits to Defendants' Sur-reply in Support of Their Motion for Summary Judgment

Third Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University
(Mar. 22, 2019), ECF No. 182-2 ..... JA4019

Second Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Mar. 22, 2019), ECF No. 182-3 ..... JA4048

Opinion & Order Granting Defendants' Motion for Summary Judgment

Memorandum Opinion (Dec. 16, 2019), ECF No. 188 ..... JA4073

Order Granting Defendants' Motion for Summary Judgment
(Dec. 16, 2019), ECF No. 189 ..... JA4123

Wikimedia Foundation's Notice of Appeal

Notice of Appeal (Feb. 14, 2020), ECF No. 191 ..... JA4124

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 1

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

REPLY DECLARATION OF SCOTT BRADNER

TABLE OF CONTENTS

I. MY CONCLUSION HAS NOT CHANGED. 1

II. INTRODUCTION 1

III. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY CONFLICTS WITH WHAT IS PUBLICLY KNOWN ABOUT UPSTREAM COLLECTION. 5

A. The four key foundations of my conclusion in my original declaration 5

1. Foundation 1: To conduct upstream collection of international public Internet communications traversing a circuit, the NSA must be copying, reassembling, and reviewing transactions on that circuit. 5
2. Foundation 2: Wikimedia communications are transported on all international circuits originating or terminating in the United States. 6
3. Foundation 3: The NSA conducts upstream collection on at least one international Internet link. 6
4. Foundation 4: The NSA is copying, reassembling, and reviewing Wikimedia communications on the international Internet links it is monitoring. 9

B. Why the NSA is copying, reassembling, and reviewing all communications on the international Internet links it is monitoring 10

1. The NSA has acknowledged that, on international Internet links it is monitoring, it does not apply IP filters. 10
2. The PCLOB has explained that upstream collection has been implemented in a technological manner designed to "comprehensively" acquire the communications of the NSA's targets. 13
3. Other technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications. 17

IV. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY ASSUMES THE IMPLAUSIBLE USE OF WHITELIST AND BLACKLIST FILTERS. 19

A. Whitelist and blacklist filters 19

1. Filters on international Internet circuits. 20
2. Whitelist filters are incompatible with the government's public descriptions of the upstream collection program. 20
   a. Whitelist filters are almost useless in the upstream collection program because using them would require that the NSA know unknowable information. 21
   b. Dr. Schulzrinne's proposed whitelisting is based on other assumptions or simplifications that are also inconsistent with what is known about the upstream collection program. 23
      i. *Number of NSA targets.* 24

ii. *Impossible targeting*.....25
iii. *Whitelist complexity and dynamism*.....27
c. Whitelist filters assume that the NSA wants to avoid almost all of the world's communications.....28
3. Blacklist filters are incompatible with the government's public descriptions of the upstream collection program.....29
a. Improbable non-targeting.....29
b. Blacklist filters would not guarantee that the NSA would avoid copying, reassembling, and reviewing Wikimedia communications.....31
4. Still copying packets.....33
5. Probing for blind spots.....33
B. "About" communications.....35

V. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY IS IMPLAUSIBLE FOR NUMEROUS OTHER TECHNICAL AND PRACTICAL REASONS.....37
A. Copy-then-filter vs. in-line filter.....37
1. Fiber-optic splitter.....38
2. Configuring the ISP router to mirror communications.....40
3. All packets are copied.....41
4. Configuring the ISP router to filter before mirroring.....41
B. Collecting "web activity".....44
C. Collecting web communications.....45
D. Collecting encrypted communications.....46
E. GCHQ surveillance.....48
F. ISP-operated copy device.....52
G. EINSTEIN.....52

VI. SUMMARY.....53

1. My name is Scott Bradner. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-006622-TSE (D. Md.), to provide an expert report addressing the government's reply to the plaintiff's brief and to my declaration, both of which were dated December 18, 2018. My qualifications to express an opinion in the case as well as my compensation and CV are as stated in my previous declaration.

2. A list of the documents provided to me by plaintiff's counsel was attached as Appendix B to the previous declaration.

I. MY CONCLUSION HAS NOT CHANGED.

3. I have carefully reviewed Dr. Schulzrinne's reply declaration as well as the government's reply brief. These documents do not change my basic conclusion in this case that "*it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications*."

II. INTRODUCTION

4. My conclusion rests on four basic foundations:

   i. The NSA is copying packets, reassembling them into communications and then reviewing the communications for the presence of selectors as part of the upstream program.

   ii. Wikimedia's traffic traverses every circuit carrying public Internet traffic into and out of the country (i.e., "international internet links").

   iii. The NSA is monitoring at least one such circuit under the upstream collection program.

   iv. On any circuit it is monitoring, the NSA must be copying, reassembling, and reviewing transactions, including Wikimedia communications, to find those communications that are to or from its targets.

5. The government and Dr. Schulzrinne do not dispute the first two foundations, and they do not seriously dispute the third foundation. But they do dispute the fourth foundation.

6. Dr. Schulzrinne disputes the fourth foundation of my conclusion primarily by describing what I will call a "Wikimedia-avoidance theory"—a hypothetical architecture for an upstream collection program that intentionally avoids Wikimedia's communications (and potentially many other types of communications), rather than having as its goal comprehensively collecting the communications to and from the NSA's targets. This hypothetical architecture is deliberately designed *not* to be comprehensive—because it is designed to avoid entire categories of Internet communications on the off chance that there might be Wikimedia communications present. In offering his Wikimedia-avoidance theory, Dr. Schulzrinne is effectively ignoring the inescapable technical implications of the government's own descriptions of the upstream collection program. Dr. Schulzrinne does not cite any evidence in either of his declarations that the NSA is actually using the extensive filters he describes, nor does he cite any evidence that the NSA is actually avoiding every one of the billions of Wikimedia communications.

7. I disagree with Dr. Schulzrinne and believe that his hypothetical architecture is inconsistent with what the government has disclosed about the upstream collection program. His architecture conflicts with the government's definitive statement that the NSA "*will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being*

monitored by NSA."[1] Dr. Schulzrinne's hypothetical architecture also conflicts with the NSA's goal "*to comprehensively acquire communications that are sent to or from its targets*."[2] The architecture also conflicts with other technical and practical necessities of conducting a program that has collected millions of communications to or from tens of thousands of targets dispersed around the world. Each of these conflicts independently disproves Dr. Schulzrinne's speculation that the NSA is using his Wikimedia-avoidance theory in its upstream collection program, and each independently supports my conclusion concerning the NSA's monitoring of Wikimedia communications.

8. Dr. Schulzrinne argues that his hypothetical architecture, based on extensive whitelist and blacklist filters, does not conflict with government disclosures about the upstream collection program. The bulk of Dr. Schulzrinne's reply declaration is a set of nuanced discussions that are not relevant to the first two conflicts between his hypothetical architecture and the government's disclosures about the upstream collection program. He devotes little space to showing that his hypothetical architecture is consistent with the government's definitive admission that the NSA will acquire wholly domestic communications under some conditions or with the NSA's described goal to comprehensively acquire the communications to or from the NSA's targets. His few arguments relating to these two conflicts are either technically incorrect or consist basically of pleas to ignore the plain meanings of the government's disclosures.

---

[1] Appendix P (FISC Opinion at 45 (Oct. 3, 2011), *available at* ECF No. 168-4 at 562-643 ("FISC Opinion")).

[2] Appendix F (Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* at 10 (July 2, 2014), *available at* ECF No. 168-3 at 199-395 ("PCLOB Report")).

9. As I explain below, Dr. Schulzrinne is also incorrect in his response to the other technical and practical necessities that support my conclusion.

10. Before getting to these explanations, I want to respond to a thread that runs through Dr. Schulzrinne's declaration. He insists that it is impossible for him or me to know the NSA's practices, priorities and capabilities, or the relative likelihood of different technical implementations of the upstream collection program.[3] While absolute assurance may be difficult, the NSA must operate in the real world and deal with the technical and operational limitations inherent in the Internet and in the telecommunications providers it compels to assist it. This need to operate in the real world constrains the ways the NSA could have implemented and be operating the upstream collection program and enables informed deduction of the NSA's actual implementation. Where different implementations present certain technical or practical trade-offs, I have tried to clearly state the degree of certainty or confidence I have in my conclusions.

11. The government has made numerous disclosures relating to the upstream collection program over the years. Many of these disclosures have specific technical implications. The descriptions of the upstream collection program in my first declaration as well as in this declaration are not speculative; instead they are based on the application of my technical expertise to analyze the government's disclosures to understand those technical implications.

[3] Schulzrinne Reply Decl. ¶ 3.

III. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY CONFLICTS WITH WHAT IS PUBLICLY KNOWN ABOUT THE UPSTREAM COLLECTION PROGRAM.

12. Dr. Schulzrinne's theory of how the NSA might be conducting its upstream collection program conflicts with what the government has publicly said about the upstream collection program. To see why, it is helpful to begin with my original declaration. The final conclusion that I reached in my original declaration was based on four key foundations. I provided support for each of the foundations, summarized here:

A. The four key foundations of my conclusion in my original declaration

1. *Foundation 1: To conduct upstream collection of international public Internet communications traversing a circuit, the NSA must be copying, reassembling, and reviewing transactions on that circuit.*

13. The first foundation for the final conclusion in my previous declaration was "*the NSA must be copying packets, reassembling them into communications and then reviewing the communications for the presence of selectors as part of the upstream program*."

14. I discussed this foundation in my original declaration.[4]

15. In summary, in order to determine if one or more selectors are present in a communication, the NSA must first copy the packets that make up a communication, then reassemble the packets into a copy of the communication. Only after the communication has been reassembled can the NSA review the contents of the communication to determine if the communication contains one or more selectors. The NSA must do this whether or not selectors are present.

[4] Bradner Decl. ¶¶ 250-320, ECF No. 168-2.

16. Neither the government nor Dr. Schulzrinne disputes the requirement that packets be copied and reassembled before they could be reviewed for the presence of selectors.

2. *Foundation 2: Wikimedia communications are transported on all international circuits originating or terminating in the United States.*

17. The second foundation for the final conclusion in my previous declaration was "*Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries*" (i.e., "international internet links").

18. I discussed this foundation in my original declaration.[5]

19. In summary, considering the volume of Wikimedia's international communications and the fact that there are users of Wikimedia's U.S.-based services in all of the world's inhabited continents and islands, there must be Wikimedia communications traversing all of the international Internet links connecting the U.S. to the rest of the world.

20. Neither the government nor Dr. Schulzrinne disputes this foundation.

3. *Foundation 3: The NSA conducts upstream collection on at least one international Internet link.*

21. The third foundation for the final conclusion in my previous declaration was "*the NSA is monitoring at least one such circuit under the upstream collection program*."

22. I discussed this foundation in my original declaration.[6]

[5] Bradner Decl. ¶¶ 336-50.

[6] Bradner Decl. ¶¶ 150-153, 222-28, 260-64, 291, 331-35.

23. In summary, based on disclosures the government has made, the NSA is monitoring at least one international Internet circuit that is transporting Wikimedia communications as part of the upstream collection program.

24. Dr. Schulzrinne does not dispute this foundation. The government disputes this foundation and makes, in my opinion, an unpersuasive argument on pages 5 and 6 of their reply brief as to whether government disclosures confirm that the NSA has monitored at least one international Internet circuit.

25. I draw support for my conclusion that the NSA is monitoring at least one such circuit under the upstream collection program from:

a. the NSA's response to plaintiff's Interrogatory No. 12, in which the NSA acknowledges that the "Internet backbone" includes international Internet circuits or links that convey Internet traffic "*internationally via terrestrial or undersea circuits*,"[7]

b. the FISC Opinion of October 3, 2011, in which the FISC stated that "*the government readily concedes that NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*,"[8]

[7] Appendix D (NSA Response to Plaintiff's Interrogatory No. 12, at 18 (Dec. 22, 2017), *available at* ECF No. 168-3 at 77-98).

[8] Appendix P (FISC Opinion at 45).

c. NSA representative Rebecca J. Richards's April 16, 2018 deposition, in which she testified that the "*will acquire*" sentence in the FISC Opinion "*is accurate*,"[9]

d. the PCLOB Report of July 2, 2014, which includes the same "*will acquire*" concession by the government,[10]

e. the NSA's 2014 targeting procedures, which make it clear that, at least in some circumstances, the NSA does not make use of an IP filter to discard wholly domestic communications,[11]

f. the explanation in my first declaration that it is logical and unsurprising that, in designing a program to intercept international Internet communications, the NSA would monitor international Internet links.[12]

26. Taken individually and together, these references make it clear that the NSA has monitored at least one international Internet circuit. Because Wikimedia communications are present on all international Internet circuits, the NSA has monitored at least one international Internet circuit that carries Wikimedia communications.

[9] Appendix K (Transcript of Deposition of Rebecca J. Richards 160:4-17 (Apr. 16, 2018), *available at* ECF No. 168-4 at 105-507 ("Richards Dep.")).

[10] Appendix F (PCLOB Report at 41 n.157).

[11] Appendix T (NSA Section 702 Targeting Procedures at 2 (2014), *available at* ECF No. 168-4 at 1062-1071).

[12] Bradner Decl. ¶¶ 222-24, 293, 332.

4. *Foundation 4: The NSA is copying, reassembling, and reviewing Wikimedia communications on the international Internet links it is monitoring.*

27. The fourth foundation for the final conclusion in my previous declaration was that on any circuit it is monitoring, the NSA must, for a variety of technical reasons, be copying, reassembling, and reviewing all transactions, including Wikimedia communications, to find those communications that are to or from (or about) its targets

28. I discussed this foundation throughout my original declaration.[13]

29. In summary, based on the foundations discussed above, Wikimedia international communications will be transported over at least one of the international Internet circuits the NSA is monitoring. As I explain in detail below, there are several independent reasons why, in the process of monitoring such a link, the NSA must be copying, reassembling, and reviewing at least all international communications transported on the link.

30. Both the government and Dr. Schulzrinne dispute this foundation. Dr. Schulzrinne maintains that the NSA could be using whitelist filters (that is, filters that enumerate the specific IP addresses and/or protocols the NSA wants to review) and/or blacklist filters (that is, filters that enumerate specific IP addresses and/or protocols that the NSA does not want to review) to avoid copying, reassembling and reviewing Wikimedia communications. I disagree that the use of such filters would be technologically consistent with the government's public descriptions of the upstream collection program, and I also disagree that the use of such filters would avoid Wikimedia communications. I will discuss my objections below.

[13] Bradner Decl. ¶¶ 36-48, 272-89, 293-94, 301-18, 333, 335.

31. I will focus on Dr. Schulzrinne's objections, leaving it to counsel to address any of the government's objections that are not based on Dr. Schulzrinne's objections.

B. Why the NSA is copying, reassembling, and reviewing all communications on the international Internet links it is monitoring

32. My conclusion that the NSA is copying, reassembling, and reviewing all communications on at least some of the circuits it is monitoring is supported by at least three independent bases. Each of these bases shows that it is a virtual certainty that the NSA is copying, reassembling, and reviewing Wikimedia's communications. I will now provide a short explanation of each of these bases and of the way in which Dr. Schulzrinne's Wikimedia-avoidance theory conflicts with what is publicly known about the upstream collection program.

1. *The NSA has acknowledged that, on international Internet links it is monitoring, it does not apply IP filters.*

33. The government has disclosed that it does not always apply filters to the traffic on circuits it is monitoring. On at least some circuits, the government has acknowledged that it does not rely on filters to eliminate wholly domestic communications.[14]

34. The government has been quite clear about one of the circumstances in which it does not apply filters. It has acknowledged that it does not rely on filters on the international links it is monitoring.[15]

[14] Bradner Decl. ¶ 291.

[15] Bradner Decl. ¶¶ 292-300.

35. In particular, the NSA has acknowledged to the FISC that it "*will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*."[16] Note that this disclosure does not say "*may acquire*" or "*might acquire*." Instead, the statement is definitive—the NSA *will* acquire wholly domestic about communications, i.e. wholly domestic communications that include one or more selectors, if they are routed through an international Internet link the NSA is monitoring.[17]

36. Dr. Schulzrinne quotes this statement in ¶ 56 of his declaration. But two paragraphs later he seems to dismiss the categorical nature of the FISC statement and says "*wholly domestic communications of the kinds described above* ***could*** *still be copied and scanned by the NSA*" (emphasis added).[18]

37. There are multiple reasons to believe that the FISC statement should be taken at face value.

38. **Reason for the FISC Opinion**. Considering the circumstances that led to the decision that this quote is part of, it is hard to imagine that the FISC is not being as precise as it possibly could be. The Opinion was the result of a sequence of multiple hearings that were called after the FISC learned of previously undisclosed surveillance activity by the NSA. The hearings involved multiple submissions to the FISC explaining the details of the surveillance activity. It is clear from this that the FISC considered the

[16] Appendix P (FISC Opinion at 45).

[17] Bradner Decl. ¶¶ 292-294.

[18] Schulzrinne Reply Decl. ¶ 58.

situation to be very important and deserving of a very careful Opinion. Thus, there is every reason to believe that the FISC was very careful in what it wrote.

39. **Precise use of language**. It is clear that the FISC was purposeful in its choice of the phrase "*will acquire*." The FISC used a different, less emphatic phrase in a different circumstance only a few paragraphs away.[19]

40. **Richards deposition**. NSA representative Rebecca J. Richards confirmed the sentence in the FISC Opinion was accurate during her April 16, 2018 deposition.[20]

41. The excerpt from the FISC Opinion specifically discusses "about" collection, but that discussion shows that the NSA does not employ IP filters at least in some circumstances.

42. As a technological matter, the only way the NSA *will* acquire wholly domestic "about" communications on the international Internet links it is monitoring, as the government has disclosed, is if it is not applying any filters to that traffic before reviewing the communications to see if they contain one or more selectors. If the NSA were applying, for example, a filter that discarded web (ports 80 and 443) communications, the NSA would miss web-based "about" communication, which would not be consistent with the FISC's statement that all "about" communications "*will*" be acquired. Therefore the NSA must not be applying any filters at these locations. Thus, even in the improbable case where the NSA were deploying the whitelist and blacklist filters of Dr. Schulzrinne's Wikimedia-avoidance theory, based on the FISC Opinion, the filters would not be deployed at the international Internet links being monitored by NSA.

[19] Appendix P (FISC Opinion at 35).

[20] Appendix K (Richards Dep. 160:4-17).

43. The FISC disclosure concerning wholly domestic "about" communications is consistent with other government disclosures that IP filters are not always used, as discussed above. The lack of all filters on these international Internet links means that the NSA would copy, reassemble and review all communications on the link, including all Wikimedia communications that happen to be on the link, for the presence of selectors. This is how the NSA would find the "about" communications. Note that, since it is undisputed that there are Wikimedia communications on every international Internet circuit, there will be Wikimedia communications on any such international Internet circuit that the NSA is monitoring.

44. The definitive statement in the FISC Opinion does not provide any room for any filters, such as the whitelist or blacklist filters Dr. Schulzrinne hypothesizes could be used, because the use of such filters would discard some wholly domestic about communications.

45. Thus, at least on the international Internet links—where it does not employ filters—the NSA must be copying, reassembling and reviewing all communications, including any Wikimedia communications that traverse the link, in order to determine which communications contain targeted selectors.

2. *The PCLOB has explained that upstream collection has been implemented in a technological manner designed to "comprehensively" acquire the communications of the NSA's targets.*

46. The government has said that the aim of the NSA is "*to comprehensively acquire communications that are sent to or from its targets*."[21]

[21] Appendix F (PCLOB Report at 10, 123, 143); Bradner Decl. ¶ 333.

47. Dr. Schulzrinne summarily dismisses the government's own statement as an unrealistic or unrealized goal. He also implies that we should not accept that the PCLOB meant what it said when it used the term "*comprehensively*."[22]

48. I find no reason to doubt the PCLOB's use of the word "*comprehensively*," which it used in describing the technical reasons why the NSA collects "about" communications. Specifically, the PCLOB states that this collection is "*an inevitable byproduct of the government's efforts to comprehensively acquire communications that are sent to or from its targets*."[23] A corollary statement is that the NSA would not be collecting "about" communications if it was not striving to be comprehensive in its collection of the communications of its targets. The PCLOB meant the term "*comprehensively*" to explain the need for the NSA's specific technological implementation of the upstream collection program, and so I find it an appropriate basis on which to explain the technological implementation of the upstream collection program.

49. Even if Dr. Schulzrinne were correct that the NSA is not being comprehensive in which international Internet circuits it is monitoring, it would not follow that the monitoring on the circuits it *does* monitor would not be comprehensive. In fact, the FISC's description of the improper collection of "about" communications only makes sense if the NSA were comprehensively monitoring individual circuits.

[22] Schulzrinne Reply Decl. ¶¶ 72-74.

[23] Appendix F (PCLOB Report at 10).

50. It may be useful to note that the NSA worked "*extensively*" with PCLOB while the PCLOB was preparing the Report to ensure the Report's accuracy.[24] Because of the obvious care that was taken in preparing the Report, it is appropriate to understand the term "*comprehensively*" literally, as the PCLOB obviously intended it to be taken. It is also useful to note that Ms. Richards confirmed in her deposition that the NSA reviewed every sentence of the PCLOB Report, that it identified any inaccuracies in the Report to the PCLOB, and that it does not believe there to be any inaccurate statements of fact in the Report.[25]

51. The use of whitelist and/or blacklist filters as Dr. Schulzrinne hypothesizes is incompatible with a goal of comprehensively acquiring communications that are to or from NSA targets. Such filters work against a goal of being comprehensive. Any such filter inevitably discards communications that might include communications that are to or from the NSA's targets. For example, Dr. Schulzrinne hypothesizes a blacklist filter that would discard traffic to or from Wikimedia servers.[26] He imagines that such a blacklist would discard all communications accessing information on Wikimedia websites by an NSA target. As Dr. Schulzrinne states, I do not know the NSA's surveillance priorities, practices, and capabilities insofar as they are unstated or not inferable based on the NSA's extensive public disclosures.[27] But any such blacklist would, by definition, be incompatible with the government's *stated* goal of completeness.

[24] Appendix K (Richards Dep. 105:20-106:13); Appendix F (PCLOB Report at 3-4).

[25] Appendix K (Richards Dep. 101:22-102:5, 105:7-12, 105:20-106:13, 107:1-5, 108:11-15, 145:9-12).

[26] Schulzrinne Reply Decl. ¶¶ 12-13, 39-42.

[27] Schulzrinne Reply Decl. ¶ 3.

52. The use of whitelist filters, as imagined by Dr. Schulzrinne, is even less compatible with the concept of completeness. At least with a blacklist, one is specifying the relatively few addresses or protocols that are not of interest among the billions of possible addresses and thousands of possible protocols. In that case, communications to or from unknown addresses or using unknown protocols will still be reviewed. With a whitelist, however, one is specifying the relatively few addresses or protocols that are of interest. This means that communications to or from most of the billions of possible addresses or using most of the thousands of possible protocols will be discarded and not reviewed to see if they are from or to NSA targets. It is certainly technically possible to design an Internet surveillance program using a whitelist, but doing so would purposefully ignore most of the Internet, and it would be inconsistent with the publicly known details about the upstream collection program.

53. If the NSA were using a whitelist, or even a blacklist, of the sort that Dr. Schulzrinne speculates, the PCLOB would not have been able to say that the NSA's goal was "*to comprehensively acquire communications that are sent to or from its targets*." I will discuss a number of other flaws in Dr. Schulzrinne's concept of using whitelist and/or blacklist filters below.

54. Based on the technical detail in the PCLOB Report, the NSA must be comprehensively copying, reassembling, and reviewing all communications on a circuit that it is monitoring if the NSA is to meet its goal "*to comprehensively acquire communications that are sent to or from its targets*." This goal is technologically incompatible with any significant use of filters other than those IP filters that ensure that at least one end of a communication is outside the U.S., and, as discussed above, the

government has disclosed that IP filters are not used on international Internet circuits. Since there are Wikimedia communications on every international Internet circuit, there will be Wikimedia communications on any international Internet circuit that the NSA is monitoring and, if the NSA is comprehensively copying, reassembling, and reviewing all packets on circuits that it is monitoring so that it can comprehensively acquire communications that are sent to or from its targets, the NSA will be copying, reassembling, and reviewing Wikimedia communications.

3. *Other technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications.*

55. There is no reliable way for the NSA to know if an individual packet on a circuit is part of a communication that contains one or more selectors without reviewing a reassembled communication containing that packet. Thus, the NSA cannot know in advance which packets need to be copied because they are part of a communication that contains selectors. Thus, the NSA must be copying, reassembling, and reviewing at an absolute minimum those communications it wishes to scan for the presence of selectors.[28]

56. The NSA could not be making use of extensive whitelist and/or blacklist filters, such as the Wikimedia-avoidance architecture Dr. Schulzrinne imagines, because of the technical inability to know in advance which packets on a circuit are part of communications to or from the NSA's targets. Dr. Schulzrinne speculates that the NSA could know its targets' IP addresses and communications protocols in advance, but as I explain at length in the next section, that is not possible given all that we know about the scale and purposes of the upstream collection program.

[28] Bradner Decl. ¶¶ 236-48, 30-18, 333, 335.

57. The only way that the upstream collection program could possibly avoid all of Wikimedia's ubiquitous communications is if the NSA had actively strived to eliminate them, and Dr. Schulzrinne presents no evidence that the NSA has ever attempted to do so or any plausible explanation for why it would do so. Even setting this fact aside, the whitelist and blacklist filters Dr. Schulzrinne imagines would not, in fact, guarantee that the NSA would be able to avoid all Wikimedia communications. Even if they could do so, there is no plausible technical or practical reason why the NSA actually would want to avoid all Wikimedia communications. Some Wikimedia communications, for example, those communications that could reveal what the NSA's foreign intelligence targets are reading and writing on Wikimedia websites such as Wikipedia, would provide information that the NSA could consider to be potentially of interest. In any case, it is almost inconceivable that the NSA went out of its way to try to specifically ensure that upstream collection would not encounter even a single Wikimedia communication.

58. Dr. Schulzrinne's Wikimedia-avoidance architecture is pure speculation, unsupported by any of the government's disclosures concerning the upstream collection program or by any plausible technical or practical consideration. My explanations of the upstream collection program are based on the public record about it and on my expert analysis of that record and of the technology of Internet surveillance. Dr. Schulzrinne's theory, in contrast, is merely a thought experiment, conducted without any consideration of whether the architecture he has imagined would remotely satisfy the purposes of the upstream collection program.

59. I will address Dr. Schulzrinne's responses to this basis for my opinion in much greater detail in the next section.

60. The above three bases show that the NSA is not employing the types of ubiquitous whitelist or blacklist filters that Dr. Schulzrinne imagines.

IV. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY ASSUMES THE IMPLAUSIBLE USE OF WHITELIST AND BLACKLIST FILTERS.

61. As I explained above (¶¶ 32-60), there are three independent technological bases that support my final conclusion in this case. Dr. Schulzrinne and the government focus most of their attention on the third basis—that technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications—and do not seriously challenge the first two bases. In the previous section, I focused on the first two bases, and in this section I will mostly focus on Dr. Schulzrinne's responses to my third basis. I will also address at least some of the points he attempts to make concerning fine points in the other bases.

62. Most of Dr. Schulzrinne's response to my third basis involves his Wikimedia-avoidance theory for the upstream collection program—an architecture based on whitelist and blacklist filters.

A. Whitelist and blacklist filters

63. Dr. Schulzrinne's declaration largely focuses on the proposition that the NSA could use whitelist and/or blacklist filters to limit the scope of the upstream collection program at least enough to avoid copying, reassembling, and reviewing Wikimedia communications but still be compatible with government disclosures on the operation of the upstream collection program.[29] He also says that the use of such filters would ensure that Wikimedia communications were not among the communications the

[29] See, e.g., Schulzrinne Reply Decl. ¶¶ 50-51.

NSA copies, reassembles and reviews in the process of searching for communications to or from its targets.

64. I disagree on both suggestions. The use of some filters may be compatible with the government disclosures that discuss the use of IP filters to discard wholly domestic communications, but the use of whitelist or blacklist filters is not compatible with the disclosures such as the ones in the FISC Opinion and PCLOB Report that describe monitoring without the use of IP filters.[30] As I explain below, there are other reasons why whitelisting and blacklisting filters are technologically incompatible with what is publicly known about the upstream collection program, and anyway, such filters would not reliably avoid Wikimedia communications.

1. *Filters on international Internet circuits.*

65. Even if it were the case that the NSA was making use of whitelist and/or backlist filters in some circumstances, for example where filters are also used to discard wholly domestic communications, it does not follow that the NSA would add special whitelist or blacklist filters where it is not using IP filters, such as international Internet circuits as noted in the FISC Opinion. (See above at ¶¶ 33-45.)

2. *Whitelist filters are incompatible with the government's public descriptions of the upstream collection program.*

66. As noted above, a whitelist filter is a filter that enumerates the specific IP addresses and/or protocols that the NSA wants to review. All incoming packets that do not have an IP source or destination address that matches an IP address in the filter list, or are using a protocol that is not listed in the filter list of protocols, will be discarded.

[30] Appendix P (FISC Opinion at 45); Appendix F (PCLOB Report at 36-37).

67. There are multiple reasons that the use of whitelist filters would be incompatible with the public descriptions of the NSA's upstream collection program.

a. Whitelist filters are almost useless in the upstream collection program because using them would require that the NSA know unknowable information.

68. As I said in my previous declaration: Whitelisting requires knowing in advance all of the IP addresses that might be used by each of the NSA's targets as well as assuming that those targets are not moving around and thereby changing their IP addresses.

69. Basically, the underlying assumption inherent in the use of whitelist filters is that the NSA has up to date, comprehensive and accurate information on where its targets will be, what sites they will be communicating with and what protocols they will be using in advance of the start of any such communications.[31] This assumption would be impossible to meet for communications to or from the NSA's targets, but even harder to meet for "about" communications because the NSA would have to (1) know which non-targets will be talking about targets and (2) have comprehensive and accurate information on which IP addresses these non-targets will be using, what sites they will be communicating with, and what protocols they will be using, in advance of the start of any such communications. In other words, the use of whitelist filters involves an assumption of precognition.

70. Dr. Schulzrinne seems to think that developing and maintaining whitelists is easy, but it would be virtually impossible to do so for a surveillance program meant to capture the communications of thousands of targets, and in fact impossible to do so for a

[31] Bradner Decl. ¶ 366(d).

program meant to capture the communications of *unknown non-targets* about targets.[32] The examples he provides for an IP address-based whitelist include unspecified individual IP addresses or blocks of IP addresses,[33] the IP addresses of VPN and e-mail servers,[34] and the IP addresses of selected web servers, webmail or chatroom sites.[35] Dr. Schulzrinne's list focuses on servers rather than clients. This makes sense in the same way as looking under a streetlight to find your lost keys makes sense. As long as the NSA's targets use these servers, the NSA will intercept the targets' communications. But there are billions of IP addresses and countless e-mail, web, and other servers in the world, and it is trivial to set up even more of these kinds of servers. Anyone can set up new servers of the kinds Dr. Schulzrinne lists. For example, I have both an e-mail server and a web server in my house. They were easy to set up. If the NSA were restricting itself to the IP addresses of known servers, it would be deliberately foreclosing its ability to capture large amounts of target traffic. Communications making use of new or temporary servers, including those that have been temporarily set up to facilitate communication by NSA targets, would escape the NSA's upstream collection program.

[32] Dr. Schulzrinne speculates that the NSA might obtain "about" communications exchanged by individuals using whitelisted IP addresses (Schulzrinne Reply Decl. ¶ 58), but the PCLOB has made clear that "about" collection permits the NSA to acquire communications between entirely unknown non-targets (Appendix F (PCLOB Report at 121, 126)). It is not possible to whitelist the IP addresses of unknown non-targets in way that would reliably acquire about communications traversing circuits being monitored by the NSA.

[33] Schulzrinne Reply Decl. ¶¶ 8, 43.

[34] Schulzrinne Reply Decl. ¶ 58.

[35] Schulzrinne Reply Decl. ¶¶ 35, 37.

71. Note that Dr. Schulzrinne admits that an IP whitelist that does not include Wikimedia IP addresses would not exclude Wikimedia communications if someone using a whitelisted address communicated with Wikimedia. [36]

72. In theory, protocol-based whitelist filters would not be quite as useless as IP address-based ones for conducting upstream-*style* collection. But, as I pointed out in my previous declaration, there is nothing that restricts Internet users to using assigned port numbers for their applications.[37] Thus, protocol-based whitelist filters could easily miss a lot of communications that the NSA would otherwise want to review, including those using new, non-public or ad hoc protocols—for example, ad hoc protocols used to facilitate the communications of the NSA's targets.

73. That said, the only example Dr. Schulzrinne provides for a protocol whitelist is one for web protocols.[38] From government disclosures, it is already known that the NSA copies, reassembles and reviews web communications, so it is clear that the NSA is not using protocol-based whitelist (or blacklist) filtering to exclude the web protocols.[39] (See ¶ 130 below.)

b. Dr. Schulzrinne's proposed whitelisting is based on other assumptions or simplifications that are also inconsistent with what is known about the upstream collection program.

74. Dr. Schulzrinne makes a number of significant assumptions or simplifications as he argues that the NSA could be using whitelists and blacklists to

[36] Schulzrinne Reply Decl. ¶ 12.

[37] Bradner Decl. ¶ 109.

[38] Schulzrinne Reply Decl. ¶ 10.

[39] Bradner Decl. ¶¶ 314-315.

implement its upstream collection program without copying, reassembling or reviewing Wikimedia communications.

i. *Number of NSA targets.*

75. As I explained in my last declaration, one reason that whitelisting of the sort Dr. Schulzrinne describes is not remotely possible for the upstream collection program has to do with the number and mobility of the NSA's targets.[40] Dr. Schulzrinne responds by implying that the NSA might only have a few targets for the upstream collection program. As he points out, the government has indicated that the NSA has over 120,000 Section 702 targets, but has not stated explicitly that all of these targets are part of the upstream collection program.[41]

76. While that is true, there must be a significant number of upstream collection program targets or they must be prolific communicators—the government has disclosed that 26 million communications were collected under the upstream collection program in 2011.[42] For example, if there were only a thousand upstream collection targets, they would have to average 26,000 communications captured under the upstream collection program each per year in order for the NSA to have collected 26 million communications per year. While the actual number has not been publicly released, based on how many communications were collected in 2011 there are almost certainly tens of thousands of targets for the upstream collection program.

[40] Bradner Decl. ¶ 366(d).

[41] Bradner Decl. ¶ 334; Schulzrinne Reply Decl. ¶ 46.

[42] See, e.g., Appendix F (PCLOB Report at 37); Appendix P (FISC Opinion at 26, 30-34, 73).

ii. *Impossible targeting.*

77. Another reason that whitelisting of the sort Dr. Schulzrinne describes is not remotely possible for the upstream collection program has to do with the requirement that the NSA know, in advance, the IP addresses of its targets or the services the targets are using, even when the targets move around and their IP addresses change.[43] In particular, it would also, as a general rule, be difficult to identify IP addresses that are exclusively used by particular individuals or groups, including NSA targets, which would be required if the aim is to limit the possible copying, reassembly and review to NSA targets.

78. In addition, not all communications with targets will contain a target's IP address. For example, the IP addresses of targets do not appear in the communications when a target is using an intermediary or a communications service that involves multiple hops.[44] Nor do IP addresses associated with targets appear in "about" communications (¶¶ 108-109). In the former case, a whitelist would miss communications to or from the target, and in the latter case, communications about a target. If the whitelist included the IP addresses of intermediary or communications services that involve multiple hops, it would sweep in the communications of all other users of the services, which could include Wikimedia communications.

79. Dr. Schulzrinne argues that the movements of the NSA's targets could be limited to a "*given geographical area*." He notes that the NSA could use a whitelist that

[43] See, e.g., Bradner Decl. ¶¶ 171, 173-74, 229-30, 366(d).

[44] Bradner Decl. ¶¶ 244-7; Appendix P (FISC Opinion at 33-35).

includes a "*set of IP addresses . . . associated with geographical areas where the target is believed to be located*."[45]

80. The use of ranges of IP addresses in a whitelist can simplify the whitelist creation and maintenance, but their use would broaden the range of addresses that would be accepted by the whitelist. Considering the broad geographic distribution of Wikimedia users, the broader the range the more likely that the whitelist would result in the copying and review of Wikimedia communications.

81. Dr. Schulzrinne's suggestion that target movements could be limited to a given geographic area is theoretical at best. There is no particular reason to think that the NSA targets are so limited.

82. But, even if the target's mobility were limited, that does not mean that their use of the Internet would be restricted to any particular range of IP addresses. Because a target could use different ISPs at different times and at different nearby locations and, because the IP addresses that ISPs use are not geographically assigned, a target could move from IP address range to IP address range as they moved around.[46] Moreover, it is reasonable to infer that the NSA's targets are widely distributed across a number of geographic regions, given that the foreign intelligence, counterterrorism, weapons proliferation, and cyber-security uses of Section 702 surveillance implicate foreign governments, organizations, and actors around the world, so that the IP address ranges would be numerous and varied.[47]

---

[45] Schulzrinne Reply Decl. ¶ 47.

[46] Bradner Decl. ¶¶ 159-60.

[47] *'Section 702' Saves Lives, Protects the Nation and Allies*, NSA/Central Security Service (Dec. 12, 2017), https://perma.cc/3JAL-WVV2 ("*'Section 702' Saves Lives*").

83. In general, I do not think it is possible to reliably predict how a user's IP address may change over time or as they move, much less the IP addresses of thousands of users. This is made even more complicated by the fact that the actual users of a given IP address can change over time.

84. Dr. Schulzrinne also suggests that the NSA could be "*whitelisting the IP addresses of websites, webmail services, and/or chatrooms of interest.*" I will note that a webmail service is a website. In addition, many websites and services are now making use of content distribution networks, which, by design, can have different, and changing, IP addresses in different parts of the world. Many websites and services are also making use of cloud-based services, such as Amazon AWS, which also can have multiple and changing IP addresses. Keeping track of the set of IP addresses in use by a particular service at any point in time is, at best, difficult.

iii. *Whitelist complexity and dynamism.*

85. As discussed above in ¶ 84, Dr. Schulzrinne imagines that the NSA could get by with a very simple set of whitelist rules. But he does not mention the fact that the rules would have to be frequently updated as NSA targets were added or removed, or as they changed their locations or methods of operation. He paints a picture of only having to list the IP addresses of some servers along with some IP addresses of a few individuals and ranges of IP addresses.

86. Dr. Schulzrinne implies that the targets of the NSA's upstream collection program do not include individuals.[48] The government's own public documents indicate that the targets include individuals.[49]

[48] Schulzrinne Reply Decl. ¶ 47.

87. Dr. Schulzrinne suggests the NSA could be using whitelists that do not include Wikimedia addresses.[50] A whitelist that included all of the possible IP addresses or IP address ranges that NSA targets could be using yet excluded Wikimedia addresses would be very large indeed since Wikimedia is using only a handful of the approximately 4 billion possible Internet addresses. (There are a bit more than 4 billion possible IP version 4 addresses, as well as billions and billions of times more IP version 6 addresses. For this declaration I will focus on IP version 4 addresses because those are the addresses in most common use.)

88. The only way that a whitelist used in the upstream collection program could be simple while still ensuring that the NSA is comprehensively collecting communications to and from its targets is if the whitelist included most non-U.S. IP addresses and most protocols, in which case there is little reason to have a whitelist in the first place. Of course, any such whitelist would unquestionably include IP addresses and protocols used by Wikimedia users and thus would unquestionably include Wikimedia communications.

c. Whitelist filters assume that the NSA wants to avoid almost all of the world's communications.

89. If the NSA is using a whitelist filter, it means that the NSA is only interested in the people and processes it already knows about and that it has decided to actively ignore everything else. The use of a whitelist in the NSA upstream collection program would be the equivalent of deciding to only look at the material coming through

[49] See, e.g., Office of the Director of National Intelligence, *Section 702 Overview* at 5-7, https://perma.cc/J9X6-YME6 ("*Section 702 Overview*"); *'Section 702' Saves Lives*, *supra* note 47: Appendix F (PCLOB Report at 36); Appendix N (FISC Submission at 4 (Aug. 16, 2011)).
[50] Schulzrinne Reply Decl. ¶ 12.

a few select holes of a sieve. Since whitelisting is specifically designed to ignore most of the Internet, it would be extraordinarily easy for the NSA's targets to avoid being monitored.

3. *Blacklist filters are incompatible with the government's public descriptions of the upstream collection program.*

90. As noted above, a blacklist filter is a filter that enumerates the specific IP addresses and/or protocols that the NSA does not want to review. All incoming packets that have an IP source or destination address that matches an IP address in the filter list or is using a protocol that is listed in the filter list of protocols will be discarded.

91. The government has disclosed that the NSA does impose one type of blacklist filter in at least some circumstances. Wholly domestic communications are filtered out, at least at some—but not all—locations (¶¶ 33-45). This type of filtering can be done with a blacklist that discards packets whose source and destination IP addresses are both within the U.S.

92. There are multiple reasons that any additional use of blacklist filters would be incompatible with the public descriptions of the NSA's upstream collection program and why the use of blacklist filters would not mean that the NSA was avoiding Wikimedia communications. I will now review them.

a. Improbable non-targeting.

93. Dr. Schulzrinne suggests that the NSA could be using blacklist filters to avoid all communications that are to or from the IP addresses of Wikimedia servers.[51] In hypothesizing that the NSA is using such a filter, Dr. Schulzrinne presupposes that the

[51] See, e.g., Schulzrinne Reply Decl. ¶ 12.

NSA would have had a reason to deliberately avoid Wikimedia communications in the upstream collection program.[52] This is different than in a whitelist filter where the NSA is deciding which communications it wants to look at.

94. Dr. Schulzrinne does not provide any evidence, let alone any creditable reason that the NSA would have specifically decided that it did not want to include Wikimedia communications in the upstream collection program. It would be difficult for him to do so because, as he points out, he does not have any specific knowledge of the NSA's priorities. But, based on the available public information about the operation of the upstream collection program and the program's intelligence-gathering purpose, I find it impossible to infer that the NSA would have singled out Wikimedia communications, among the vast array of communications on the Internet, as communications that should be ignored.

95. Dr. Schulzrinne does talk about reducing the load on the devices the NSA is using in the upstream collection program, but he does not indicate why a desire to reduce load would have led the NSA to exclude Wikimedia communications.[53]

96. Dr. Schulzrinne notes that the inter-regional Internet capacity is very large these days—as much as 295 terabits per second.[54] To put the volume of Wikimedia communications into context, in the six-month period between August 1, 2017 and January 31, 2018, Wikimedia engaged in approximately 760 billion international communications.[55] This works out to about 48 thousand communications per second.

---

[52] Bradner Decl. ¶ 367(a).

[53] Schulzrinne Reply Decl. ¶¶ 20-22.

[54] Schulzrinne Reply Decl. ¶ 20.

[55] Bradner Decl. ¶ 346.

Assuming that an average Wikimedia communication is 81 packets,[56] and that the average packet length is 1,500 8-bit bytes, that means Wikimedia communication averages about 47 Gbps per second. Thus, Wikimedia represents about 0.016% of the inter-regional Internet capacity. Even if I were way off in my estimate of an average communication length of 81 packets, the percentage of inter-regional Internet capacity, measured in bits per second, represented by Wikimedia communications is still extraordinarily low. Blacklisting the Wikimedia IP addresses would not make any measurable difference to the load experienced by the NSA's upstream collection system. Thus, it is very unlikely that the NSA would have decided to specifically blacklist Wikimedia communications to reduce the load on the upstream collection program systems, even if, as Dr. Schulzrinne suggests, it might be easy to do.[57] Note also that although Dr. Schulzrinne says that, according to a statistics website, one of the Wikimedia websites is the 5th most popular website globally, it does not follow that Wikipedia generates a significant amount of traffic measured in bits per second compared to the inter-regional Internet capacity.

b. Blacklist filters would not guarantee that the NSA would avoid copying, reassembling, and reviewing Wikimedia communications.

97. I noted a number of situations under which NSA's use of a blacklist filter designed to block the communications to or from the IP addresses of Wikimedia servers would not guarantee that the NSA would avoid copying, reassembling and reviewing

[56] Bradner Decl. ¶¶ 144-45.

[57] Schulzrinne Reply Decl. ¶ 41.

Wikimedia communications in my previous report.[58] Dr. Schulzrinne responded to the discussion in my previous declaration in his reply declaration.[59] In his response, Dr. Schulzrinne addressed three of the cases in which I said that Wikimedia communications could still be copied, reassembled and reviewed even if there were a blacklist filter in place that discarded communications to or from the IP addresses of Wikimedia servers. Dr. Schulzrinne responds to my MCT, e-mail and VPN examples.

98. For each of these three cases, Dr. Schulzrinne creates a list of conditions that he says must be true for Wikimedia communications to be copied, reassembled and reviewed.

99. For example, in each of his sets of conditions, he says that the relevant communication would have to *not* be blacklisted in order for that communication to be acquired, as though that would be a difficult condition to meet. But, in reality, the possibility that these communications *would* be blacklisted in each example is far-fetched. For the MCT example, in which a Wikimedia communication is enclosed in an MCT traversing an international Internet link, Dr. Schulzrinne says that the MCT would have to not be blacklisted. Such an MCT could be one to or from a mail server. In the e-mail example, Dr. Schulzrinne says that the e-mail itself would have to not be blacklisted. In this case the communication would also be to or from an e-mail server. And, in the VPN example, he says that the VPN server communications would have to not be blacklisted.

[58] Bradner Decl. ¶ 367(b).

[59] Schulzrinne Reply Decl. ¶¶ 78-87.

100. It is technically possible for the NSA to block communications to and from e-mail servers and to or from VPN servers but it is hard to understand why the NSA would do this. These types of servers are exactly the types of services that NSA targets could be using, and so a world-wide blacklist blocking communications to and from e-mail and VPN servers would block just the type of sites bad actors would be using and the NSA would have an incentive to target.

101. All three of Dr. Schulzrinne's sets of conditions include the obvious requirements that either the user or the server be located outside the U.S., but not both, that communications between the user and server transit an international Internet link the NSA was monitoring, and that the users must be communicating with Wikimedia. These are all requirements I assumed when I described the cases. In my opinion, these requirements are obvious and would likely be frequently met.

4. *Still copying packets.*

102. As I discuss below and in my original declaration, even if the NSA were employing some sort of filter that discarded packets that were part of Wikimedia communications, the NSA would most likely be copying those packets before discarding them.[60]

5. *Probing for blind spots.*

103. In my original declaration I mentioned that if the NSA were using blacklist or whitelist filters to ignore protocols or IP address ranges, targets could probe to see if they could discover the lapses in coverage.[61] Dr. Schulzrinne questions the

[60] See ¶¶ 114-122; Bradner Decl. ¶¶ 269-279.

[61] Bradner Decl. ¶ 366(b).

possibility that the NSA's targets are sophisticated enough to probe for gaps in the NSA's coverage.[62] I will note that foreign intelligence officers and services are among the people and groups that are lawful targets for the upstream collection program,[63] and they are among our most sophisticated adversaries. I will also note that such probing is more likely to require process rather than technical sophistication. For example, it could involve purposefully communicating information about an information resource over a protocol you suspect the NSA is not monitoring and then monitoring the information resource to see if it is accessed, or purposefully conveying actionable information, such as the identity of a foreign agent, over such a protocol and seeing if action is taken against the agent.

104. There is another significant risk to the NSA's use of whitelists and/or blacklists to limit what its surveillance devices copy, reassemble, and review, and it would not require probing.

105. To implement the kind of whitelisting and blacklisting that Dr. Schulzrinne hypothesizes in the way that he hypothesizes, ISPs would need to configure their routers with the whitelists and blacklists. It is standard practice for ISPs to backup their router configurations in their network management systems, so that they can quickly deploy, modify or re-deploy the configurations as needed. There have been too many cases where ISPs' network management systems have been compromised, and so whitelisting and blacklisting of the sort Dr. Schulzrinne describes would create the unnecessary risk of compromise of the NSA's whitelists and blacklists. In addition,

[62] Schulzrinne Reply Decl. ¶ 32.

[63] 50 U.S.C. § 1881a.

multiple ISP technicians generally have access to the management stations, increasing the number of people that would have to be trusted.

106. Dr. Schulzrinne strangely suggests that the NSA might not care if foreign actors knew of ways to get around NSA monitoring.[64] I agree that I am not privy to the NSA's priorities that are not public, but I would find it quite extraordinary if the NSA did not care whether targets of its surveillance had a roadmap for evading its surveillance.

107. In fact, George C. Barnes, Deputy Director of the NSA, specifically stated that revealing information that could be used to help an adversary evade the NSA would be a problem:

> *Revealing which channels* [of communication] *are free from surveillance and which are not could also reveal sensitive intelligence methods, and thereby help an adversary evade detection and capitalize on limitations in the NSA's surveillance capabilities*.[65]

B. "About" communications

108. Dr. Schulzrinne's proposed use of whitelist and blacklist filters is also entirely inconsistent with "about" collection. About collection is a process within the upstream collection program that involves the collection of communications between two non-targets that contain one or more selectors associated with an NSA target. Dr. Schulzrinne discusses "about" communications in ¶¶ 49-52 of his reply declaration.

109. As a general rule, the IP addresses on the packets that make up the "about" communications will likely have no relation to any targets. For this reason, the use of

[64] Schulzrinne Reply Decl. ¶ 33.

[65] Barnes Decl. ¶ 57, ECF No. 141-1.

whitelist and blacklist filters of the sort that Dr. Schulzrinne describes is not compatible with "about" collection, because that kind of filtering would guarantee that the NSA's upstream collection devices miss "about" communications.

110. Dr. Schulzrinne responds to this fact by describing a two-step process for the collection of "about" communications that he claims is compatible with the use of whitelist and blacklist filters. The first step uses a whitelist IP address-based filter, but to set up this whitelist filter the NSA would have to know in advance which non-targets' IP addresses to whitelist, or what servers' IP addresses to whitelist in order to find the "about" communications.

111. Dr. Schulzrinne oversimplifies the problem by saying that "about" communications would be collected if the communication containing the "about" selector were whitelisted. Dr. Schulzrinne ignores the fact that in order to include such communications in a whitelist, the NSA would first have to know in advance which non-targets were going to be talking about targets and also know in advance what IP addresses the non-targets would be using. If the NSA were following Dr. Schulzrinne's description, they might capture an occasional "about" communication if one of the non-targets was using an IP address or service that the NSA had whitelisted for some other reason, but they could not normally capture them.

112. Under Dr. Schulzrinne's model, the only way the NSA could reliably capture about communications would be to whitelist all non-wholly domestic communications, in which case the whitelist would be guaranteed to result in the copying and review of Wikimedia communications.

V. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY IS IMPLAUSIBLE FOR NUMEROUS OTHER TECHNICAL AND PRACTICAL REASONS.

113. Dr. Schulzrinne made a number of other points in his reply brief. I will respond to some of them now.

A. Copy-then-filter vs. in-line filter

114. In my original declaration, I expressed the opinion that the NSA was most likely using a copy-then-filter architecture for its upstream collection program.[66] The government treats this as a significant concession, but the government completely misrepresents how this point relates to my ultimate conclusion. My discussion of the copy-then-filter implementation is a discussion of an independent reason to believe that the NSA is copying Wikimedia's communications as they travel across international Internet links—even if the NSA were performing filtering of the kind Dr. Schulzrinne hypothesizes. Specifically, a copy-then-filter architecture renders all of Dr. Schulzrinne's speculation about filtering irrelevant. That is because a copy-then-filter architecture involves the NSA copying all communications on a circuit independent of any filtering that might subsequently be performed on the copies of the packets. That of course includes the copying of Wikimedia's communications.

115. It bears emphasis, however, that even if the NSA is not using a copy-then-filter architecture, but instead is using the in-line filtering architecture Dr. Schulzrinne describes (which he refers to as "*filter-then-copy-and-scan*"), some Wikimedia communications will be copied. Dr. Schulzrinne's Wikimedia-avoidance architecture is entirely implausible for the many independent reasons I have explained above (including

[66] Bradner Decl. ¶¶ 265-289.

the multiple ways in which his theory conflicts with the government's public disclosures) and explain below (including the multiple ways in which his theories are at odds with the technical and practical necessities of conducting the upstream collection program as it has been described). In short, even if implemented as Dr. Schulzrinne imagines, Dr. Schulzrinne's filters will miss some Wikimedia communications and those communications will be copied, reassembled and reviewed.

116. I will now respond to Dr. Schulzrinne's points about the implications of using the one implementation (copy-then-filter) versus the other (in-line filtering). But, it continues to be my opinion that the copy-then-filter architecture is the simplest, most reliable and easiest to operate architecture for the NSA to use for the upstream collection program. If the NSA uses this architecture, all packets on a communications circuit being monitored by the NSA, including the packets that make up Wikimedia communications, are copied. It is possible that the NSA uses different architectures in different monitoring locations, but, in my opinion, the advantages of the copy-then-filter architecture mean that it is most likely the default architecture. Even if the copy-then-filter architecture is not being used everywhere, all packets on the circuits where the copy-then-filter architecture is used are copied, as are many packets on the circuits where an in-line filter architecture is used.

1. *Fiber-optic splitter.*

117. The simplest option involves the use of a fiber-optic splitter, which Dr. Schulzrinne suggested in his first declaration and I discussed in my original declaration.[67] A fiber-optic splitter produces a copy of all communications on a fiber by splitting the

[67] Schulzrinne Decl. ¶¶ 55-56, ECF No. 164-4; Bradner Decl. ¶¶ 275-77.

light on the fiber into two streams.[68] Dr. Schulzrinne notes that adding a fiber-optic splitter into a network adds a potential failure point and causes loss of optical power.[69] He is correct that a fiber-optic splitter could be a failure point, but a fiber-optic splitter is a passive device that does not include a processor or software that could have bugs or need upgrading, and it does not require any configuration or power so the probability of failure is very low and the possibility that misconfiguration or that a power failure could impact the network is nonexistent. A fiber-optic splitter does reduce the optical power that would be received by the ISP's receiving device but, as long as the ISP knows the splitter is in-line, the receiving device can be configured to compensate for the loss.

118. On the other hand, using a router's mirror function (as Dr. Schulzrinne describes), would have a much higher failure and disruption probability because the router requires power and because the router includes a computer and software that can have bugs, would need updating, and would be vulnerable to hacking. If the ISP used an existing router to filter and then copy communications, the *added* risks would be a little bit less significant, but the need to constantly reconfigure the device with updated blacklists and whitelists would create the risk of misconfiguration or overloading. Either way, the risks of failure are greater for the in-line device Dr. Schulzrinne proposes than for a fiber-optic splitter.

119. Dr. Schulzrinne notes, as I did in my original declaration, that the use of a fiber-optic splitter would mean that it would need to be coupled with an opto-electronic

[68] Bradner Decl. ¶¶ 5, 275-77.

[69] Schulzrinne Reply Decl. ¶ 27.

device to split out the channels the NSA wanted to monitor.[70] Dr. Schulzrinne paints a pretty dire picture of the requirements for this device, but, as I pointed out in my original declaration, not all channels on the cables are used to transport international Internet communications.[71] The splitting device only needs to pay attention to the circuits that are so used and, of those, only the circuits that the NSA wishes to monitor. Note that the splitting device Dr. Schulzrinne mentions is not an esoteric piece of equipment; it is the same device that ISPs routinely use to split the light on optical fiber into different channels and is normally included in the router that the optical fiber is plugged into.

2. *Configuring the ISP router to mirror communications.*

120. The other option to support the copy-then-filter architecture is, as Dr. Schulzrinne suggests, for the NSA to command the ISP to configure its router to mirror the communications on the circuits the NSA wants to monitor and send the mirrored packets to one or more NSA-operated devices.[72]

121. The operationally simplest way to do this is to command the ISP to configure its router to mirror all of the packets on the channels the NSA wants to monitor. Such a configuration is very simple and would not have to change over time. This way also means that the ISP personnel are not exposed to any of the NSA's collection methods other than the fact that data collection is being done on a particular circuit.

[70] Bradner Decl. ¶ 277; Schulzrinne Reply Decl. ¶ 21.

[71] Bradner Decl. ¶¶ 214-25.

[72] Bradner Decl. ¶ 278.

3. *All packets are copied.*

122. All packets on a circuit, including the packets comprising any Wikimedia communications, are copied with either of the above two designs.

4. *Configuring the ISP router to filter before mirroring.*

123. Dr. Schulzrinne hypothesizes that the NSA could be applying whitelist or blacklist filters to the packets *before* the router mirrors the packets to the NSA upstream collection devices.[73] He says, and I agree, that if the NSA is doing so, then the logical place to apply such filters is in an ISP router that will be processing the stream of packets entering or exiting a communications channel the NSA wants to monitor. Dr. Schulzrinne and I disagree as to whether the NSA could in fact be using such whitelist or blacklist filters on all of the circuits that it monitors, even if it uses them on some, and we disagree as to what specific configurations would be likely for whitelist or blacklist filters. I discuss this disagreement at length above.[74]

124. There are different categories of whitelist or blacklist filters that have different implications when it comes to configuring the filters. In this case, I believe it most likely that the actual router configuration is being performed by ISP personnel since, in my experience, an ISP would be unlikely to allow non-ISP personnel to configure its routers. In theory, the NSA could create some configuration data files that the ISP personnel could then load into the router—this would put the NSA directly in charge of the filter details while avoiding having the fingers of non-ISP personnel in the routers. In this case the ISP personnel would still have access to the details of the

[73] Schulzrinne Reply Decl. ¶ 28.

[74] See ¶¶ 63-107.

configuration. In any case, the more complex and detailed the filter configurations, the more often they will change. If an ISP had to implement all of the whitelist and blacklist filters Dr. Schulzrinne has suggested, the ISP would have to be updating its router configuration all the time. Operationally, this is not a good idea from an ISP's perspective because it increases the chance of human error impacting the ISP operations.

125. In his reply declaration, Dr. Schulzrinne says that he does not suggest installing an NSA-operated device in the middle of an ISP's network.[75] There are only two options for the type of in-line filter that Dr. Schulzrinne proposes: (1) having an NSA-operated device perform the filtering, or (2) exposing ISP personnel to the details of the NSA's collection program. Each of these options has significant downsides. If in-line filters are used at all, it may well be that there is no single answer. For example, the NSA might want to operate its own device in smaller ISPs or ISPs demonstrating less technical expertise but delegate the operation to ISP personnel in larger ISPs or those with more technical expertise.

126. As I discussed in my previous declaration, it is my opinion that the NSA would want to limit the exposure of the details of at least some of the types of whitelist and blacklist filters Dr. Schulzrinne suggests could be used. Dr. Schulzrinne dismisses that concern by pointing to the fact that the NSA shares target information with telecommunications providers under the PRISM program.[76] But Dr. Schulzrinne ignores the fact that the categories of information are quite different. With PRISM, the identities of one or more individual targets are exposed to the personnel of the telecommunications

[75] Schulzrinne Reply Decl. ¶ 23.

[76] Schulzrinne Reply Decl. ¶ 18.

companies it compels to provide assistance. But the detailed configuration of blacklist and whitelist filters can provide a roadmap that can be used to entirely avoid NSA surveillance.

127. Note that the NSA has to share confidential target information with telecommunications providers under PRISM, since only the telecommunications providers have the ability to retrieve the communications of interest. The NSA does not have that constraint in the upstream collection program. The NSA could:

a. use a copy-then-filter approach with a fiber-optic splitter, which would reveal no confidential information to the provider other than the fact that the NSA was monitoring one or more channels on a fiber;

b. use a copy-then-filter approach by commanding the ISP to provide the NSA with a full copy of the packets on a particular channel, which would increase the shared confidential information to include the specific channels being monitored;

c. operate its own in-line filter device which would also share the information about which channels were being monitored; or

d. command the ISP to operate an in-line filter, which would require the NSA to share the details of exactly what it is monitoring and not monitoring (i.e., the configurations for the whitelist and blacklist filters Dr. Schulzrinne hypothesizes).

128. Given the fact that the NSA has a choice, it seems reasonable to infer that the NSA would want to minimize the confidential information that it had to share to

operate the upstream collection program. (See also ¶ 107 above, citing Barnes Decl. ¶ 57.)

129. It is one thing to expose the fact that the NSA has asked for Mr. Smith's e-mail and a very different thing to expose the fact that the NSA is using a whitelist filter that discards packets to or from large chunks of the Internet or discards all packets that are not e-mail or web traffic.

B. Collecting "web activity"

130. Dr. Schulzrinne questions the idea that the NSA might be monitoring communications to or from web servers.[77] I discussed the government's disclosure that it is monitoring "web activity" in my original declaration.[78] Dr. Schulzrinne suggests that the government might have been sloppy and was referring to overall Internet activity when they wrote "web activity."[79] That seems to me to be a very tenuous argument. I have seen no indication in any of the government's released documents that they are that sloppy. In particular, the reference to "web activity" is in a formal and highly technical government submission to the FISC in response to a highly technical request from the court, hardly a place that anyone would be sloppy. And, the context in which the phrase was used makes total sense if the government was using the phrase precisely, as a reference to the world wide web protocols (HTTP/HTTPS).

131. It is also well known that terrorists make use of communications tools that use HTTP/HTTPS. Most human-to-human communications on the Internet are transported using HTTP/HTTPS, whether over older mediums (websites) or more recent

[77] Schulzrinne Reply Decl. ¶ 36(b).

[78] Bradner Decl. ¶¶ 314-15.

[79] Schulzrinne Reply Decl. ¶ 36(b).

ones (messaging services). For example, terrorists make use of the Telegram application, which can operate over port 80 (HTTP).[80] There are obviously examples of widely used protocols that do not use HTTP/HTTPS—for example, FTP/SMTP/IMAP/POP/etc. — but an increasing amount of Internet communications do use HTTP/HTTPS to increase security (HTTPS) or to bypass firewalls (both).

C. Collecting web communications

132. Dr. Schulzrinne spends some time discussing ways the NSA could be filtering out web or encrypted traffic to make it less likely that it is copying, reassembling and reviewing Wikimedia communications. I will now review his suggestions.

133. Dr. Schulzrinne suggests that the NSA might be configuring a blacklist to block both HTTP and HTTPS communications.[81] First of all, the government has acknowledged that it is capturing web activity, i.e. HTTP and/or HTTPS communications (¶ 130). The NSA cannot be both blocking all web communications and be collecting web activity.

134. Note that web traffic makes up a very large percentage of Internet traffic. For example, one study of traffic between the Japanese WIDE Project ISP and its backbone ISP showed that web traffic was 75% of the overall traffic.[82] As Dr. Schulzrinne notes, web traffic is not restricted to web sites such as cnn.com and

[80] Joby Warrick, *The 'App of Choice' for Jihadists: ISIS Seizes on Internet Tool to Promote Terror*, Wash. Post (Dec. 23, 2016), http://wapo.st/2hzoY6P; *MTProto Mobile Protocol: Transport*, Telegram, https://perma.cc/T6FL-WWP8.

[81] Schulzrinne Reply Decl. at 15 n.7.

[82] Chia-ling Chan, et al., *Monitoring TLS Adoption Using Backbone and Edge Traffic* (2018), https://perma.cc/6C8D-GWCT.

wikipedia.org; web protocols are also used for webmail and chatrooms.[83] Ignoring web traffic as Dr. Schulzrinne has suggested would exclude the vast quantities of human-to-human communications that are transported by the web.

135. As I mentioned in my original declaration, any such discarding of all web communications would leave a very large hole in the NSA's coverage[84]—contrary to any notion of completeness such as that noted in the PCLOB Report

136. Dr. Schulzrinne suggests that the NSA could be restricting its collection of web communications to a few sites such as "*specific webmail and chatroom sites*." As I discussed above, this would also leave very large holes in the NSA's coverage. (¶ 134) Any such filtering would also be contrary to the aim of the "about" collection program, which is to collect communications between unknown non-targets.

D. Collecting encrypted communications

137. Dr. Schulzrinne says that the NSA might not be collecting HTTPS communications using the authority in Section 702 to collect encrypted communications.[85] Dr. Schulzrinne makes no actual argument that the NSA is not collecting at least some encrypted communications, he just says that my arguments are not technical.

138. Multiple government disclosures make it clear that the NSA collects encrypted communications. For example:

a. The PCLOB noted: "*With respect to all of the agencies, extensions from these age-off requirements may be sought from*

[83] Schulzrinne Reply Decl. ¶ 35.

[84] Bradner Decl. ¶ 366(f).

[85] Schulzrinne Reply Decl. ¶ 36(a).

a high-level agency official. Other limited exceptions apply, such as to communications that are still being decrypted." (Note the phrase "*still being decrypted.*")[86]

b. The PCLOB also noted: "*The NSA may also retain communications beyond the normal age-off period if it is still decrypting the communication or using the communication to decrypt other communications*." (Note the phrases "*still decrypting*" and "*decrypt other communications.*")[87]

c. The NSA's minimization procedures note: "*In the context of a cryptanalytic effort, maintenance of technical data bases requires retention of all communications that are enciphered or reasonably believed to contain secret meaning, and sufficient duration may consist of any period of time during which encrypted material is subject to, or of use in, cryptanalysis*." (Note the phrase "*encrypted material is subject to.*")[88]

139. Dr. Schulzrinne hypotheses that while the NSA may collect encrypted communications it may only be doing so under PRISM.[89] This seems unlikely. The telecommunications providers assisting the NSA in the case of PRISM will frequently have direct access to the user's unencrypted communications, for example in an e-mail

[86] Appendix F (PCLOB Report at 60).

[87] Appendix F (PCLOB Report at 63).

[88] Appendix S (NSA Section 702 Minimization Procedures at 10 (2014), *available at* ECF No. 168-4 at 1046-1061).

[89] Schulzrinne Reply Decl. ¶ 36(a).

server, whereas communications across the Internet are increasingly being encrypted.[90] If the upstream collection program were to ignore encrypted communications it would be increasingly unable to collect any communications.

E. GCHQ surveillance

140. I discussed some public disclosures from the U.K.'s signals intelligence agency, Government Communications Headquarters (GCHQ), in my first declaration in order to "*reinforce my conclusions that the NSA relies on the copy-then-filter configuration to conduct the upstream collection program and that it does not selectively filter traffic prior to copying it as Dr. Schulzrinne hypothesizes it could*."[91]

141. Those include the disclosure by the GCHQ that, under a surveillance program analogous to upstream collection, "*it is necessary to intercept the entire contents of a bearer [circuit], in order to extract even a single specific communication for examination*."[92]

142. Dr. Schulzrinne dismisses the disclosures as only being "*the roughest outline*" of the process the GCHQ uses, and as being "*non-technical*."[93] I disagree and find the disclosures have enough detail for me to draw my conclusions.

143. He also says that the disclosures, even if not detailed, are "*quite comparable*" to his suggested filter first approach. I disagree with this as well.

144. Dr. Schulzrinne provides a citation describing the GCHQ's bulk interception as purported proof of his contention.[94] The citation Dr. Schulzrinne provides

[90] John Maddison, *Encrypted Traffic Reaches a New Threshold*, NETWORKComputing (Nov. 28, 2018), https://perma.cc/6VFY-YEGL.

[91] Bradner Decl. ¶ 369.

[92] Bradner Decl. ¶ 368.

[93] Schulzrinne Reply Decl. ¶¶ 59-60.

is from a filing by the U.K. government in the European Court of Human Rights. The U.K. filing, in turn, refers to a "*Bulk Powers Review*" of "*the operational case for various intelligence gathering powers.*"[95] The U.K. filing quotes the *Bulk Powers Review* to provide a summary on how the GCHQ interception program works.[96] Dr. Schulzrinne's citation is of the U.K. filing's quoting of the *Bulk Powers Review*.

145. Dr. Schulzrinne says that the U.K. filing "*actually describe*[*s*] *a collection approach quite comparable (at least at a general level) to the type of IP address and port and protocol number filtering described in my earlier declaration*."[97] I disagree with Dr. Schulzrinne's analysis of the U.K. filing he provides in support of his conclusion. I do not believe that the citation shows any evidence that the GCHQ is filtering traffic on a channel before copying the traffic; in fact, the citation shows the opposite. The paragraph in the *Bulk Powers Review* that immediately follows the outline Dr. Schulzrinne cites makes this clear:

> ***The two major processes***
>
> *2.19. A description is given in the 2015 ISC report (paras 61-73), of two major and distinct processes that apply to interception under bulk warrants. Those processes are identified in more detail in the closed version of the report, and I have been briefed on each of them. In summary:*
>
> *(a)* ***The "strong selector" process*** *(2015 ISC report, paras 61-64) operates on the bearers that GCHQ has chosen to access. As the*

---

[94] Schulzrinne Reply Decl. ¶ 61.

[95] David Anderson Q.C., U.K. Independent Reviewer of Terrorism Legislation, *Report of the Bulk Powers Review*, Cm 9326 (August 2016), https://perma.cc/V3ME-QZED ("*Bulk Powers Review*").

[96] *Bulk Powers Review*, *supra* note 101, at 23-24.

[97] Schulzrinne Reply Decl. ¶ 61.

internet traffic flows along those chosen bearers, the system compares the communications against a list of strong selectors in near real-time. Any communications which match the selectors are automatically collected and all other communications are automatically discarded. The nature of the global internet means that the route a particular communication will take cannot be predicted and a single communication is broken down into packets which can take different routes. In order to identify and reconstruct the wanted communications of subjects of intelligence interest, GCHQ's processing relies on accessing the "related communications data" (secondary data) in the bearer.

A copy of all the communications on a bearer has to be held for a short period in order to allow the strong selectors to be applied to those communications. This process accordingly requires a bulk warrant under the Bill. However, in the opinion of the ISC, "while this process has been described as bulk interception because of the numbers of communications it covers, it is nevertheless targeted since the selectors used relate to individual targets".

(b) ***The "complex query" process*** *(2015 ISC report paras 65-73) is used where GCHQ is looking to match much more complicated criteria, for example with three or four elements. This process operates across a far smaller number of bearers. These bearers are not chosen at random, as GCHQ focuses its resources on those most likely to carry communications of intelligence value. As a first step in the processing under this method the system applies an initial set of processing rules. Those rules seek to select communications of potential intelligence value while discarding those least likely to be of intelligence value. The selected communications are not available to GCHQ staff to search through at will. Further complex searches draw out the*

> *communications of intelligence value. By performing searches combining a number of criteria, the odds of a 'false positive' are considerably reduced.*
>
> *This second process is closer to true bulk interception, since it involves the collection of unselected content and/or secondary data. It permits types of analysis and selection that are not currently achievable in the near real-time environment of the strong selector process (2.19(a) above). But as with the first process, it remains the case that communications unlikely to be of intelligence value are discarded as soon as that becomes apparent.*[98]

146. The description of the "strong selector" process specifically says "*[a] copy of all the communications on a bearer has to be held for a short period in order to allow the strong selectors to be applied to those communications."* Combining this statement with the description of the third stage of collection in the extract Dr. Schulzrinne provided makes it clear that all of the communications on a bearer (GCHQ's term for a circuit) are copied and stored at least temporarily so that those communications that contain selectors, if any, can be located.

147. The description of the "complex query" process says that the process is "*closer to true bulk interception, since it involves the collection of unselected content and/or secondary data.*"[99] This statement by itself notes that the GCHQ is collecting "*unselected content*." To do so, it is copying communications that it has not checked for the presence of selectors.

---

[98] *Bulk Powers Review*, *supra* note 101, at 24-25.

[99] *Bulk Powers Review*, *supra* note 101, at 25.

148. In both cases, GCHQ is copying all the contents of a bearer, which is the point I made in my original declaration.

F. ISP-operated copy device

149. Dr. Schulzrinne suggests that a fiber-optic splitter, or even an electronic device such as a router, could be operated by the ISP and the copy of the communications created by such a device could be sent to a filtering device operated by the ISP with the output of the filtering device sent to the NSA.[100] Such an arrangement would not actually change the fact that the NSA is creating a copy, since the copy device and filter would be operated at the direction and auspices of the NSA. As I noted in my first declaration, work performed at the direction of the NSA is still work done by the NSA.[101] In addition, as with other copy-then-filter configurations, all the packets on the circuit, including packets that are part of Wikimedia communications, are copied.

G. EINSTEIN

150. I mentioned in passing the U.S. government-operated EINSTEIN 2 & 3 systems in my original declaration.[102] Dr. Schulzrinne made rather much more of the mention than I had in mind. I just mentioned EINSTEIN as an example of a deep packet inspection (DPI) device. But I will comment on Dr. Schulzrinne's discussion of EINSTEIN.[103]

151. Dr. Schulzrinne tries to differentiate EINSTEIN from the NSA's upstream collection program in two ways: (1) he says that EINSTEIN has to look at all traffic

[100] Schulzrinne Reply Decl. ¶ 64.

[101] Bradner Decl. ¶ 5.

[102] Bradner Decl. ¶¶ 259, 286.

[103] Schulzrinne Reply Decl. ¶¶ 67-69.

whereas the upstream collection program, at least as he imagines it, does not; and (2) EINSTEIN is a cybersecurity system and upstream collection program is not.

152. Relating to (1), Dr. Schulzrinne says "*cyber attacks can use any protocol, originate from any external Internet host, and can target any destination system, to be effective an intrusion-detection system must inspect all incoming traffic*."[104] Of course, the same can be said, with the exception of the discarding of wholly domestic communications, of the upstream collection program. The main theme of Dr. Schulzrinne's declaration is how to limit the upstream collection program so that it avoids Wikimedia communications at the expense of being able to capture communications from, to or about its targets which "*can use any protocol, originate from any external Internet host, and can target any destination system*."[105]

153. Relating to (2), the use of DPI is not limited to cybersecurity systems. The upstream collection program uses DPI to find the selectors in the communications it reviews. In addition, the NSA's public materials include mention of the use of the upstream collection program for cybersecurity purposes.[106]

VI. SUMMARY

154. In order for Dr. Schulzrinne to be correct in his speculation that the NSA could operate the upstream collection program without ever copying, reassembling, or reviewing even a single Wikimedia communication, every one of the following assumptions, which are prerequisites for his claim, must be true:

[104] Schulzrinne Reply Decl. ¶ 69.

[105] Schulzrinne Reply Decl. ¶ 69.

[106] See, e.g., *'Section 702' Saves Lives*, *supra* note 47; *Section 702 Overview*, *supra* note 49, at 4.

1. That the government's unambiguous concession to the FISC—"*that NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*"[107]—was false, despite the NSA's representative confirming its accuracy at her deposition.[108]
2. That the PCLOB's statement—that the technical design of the upstream collection program supports the NSA's goal "*to comprehensively acquire communications that are sent to or from its targets*"—was false.[109]
3. That the NSA decided to significantly limit the scope of its collection under the upstream collection program by doing at least one of the following:
    a. Deliberately blacklisting all Wikimedia IP addresses.
    b. Deliberately whitelisting a subset of IP addresses other than Wikimedia's IP addresses.
    c. Deliberately excluding all of the communications protocols that Wikimedia uses, through blacklists or whitelists configured to exclude:
        i. All web activity (i.e., HTTP/S), notwithstanding the government's concession that upstream collection involves the collection of "web activity," *and*

[107] Appendix P (FISC Opinion at 45).

[108] Appendix K (Richards Dep. 160:4-17).

[109] Appendix F (PCLOB Report at 10).

ii. All e-mail activity (i.e., SMTP), notwithstanding the government's concession that, under the upstream collection program, it uses e-mail addresses as selectors.[110]

4. That the NSA has limited the scope of its upstream collection in this way on *every* one of the international Internet links it monitors.

5. That the limitations on collection above are in fact entirely effective at avoiding Wikimedia's communications, even though there are multiple circumstances in which they would not be. For example, Dr. Schulzrinne's claim requires that none of the following could ever occur:

   a. If the NSA whitelisted IP addresses other than Wikimedia's IP addresses:

      i. A user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link monitored by the NSA.

   b. If the NSA blacklisted all Wikimedia IP addresses:

      i. One of Wikimedia's communications is enclosed in a multi-communication transaction (MCT) that is not blacklisted, and it traverses an international Internet link monitored by the NSA.

      ii. One of Wikimedia's communications passes through an intermediary that replaces Wikimedia's IP address (such as an e-mail server, VPN, or other communication service)

[110] Appendix F (PCLOB Report at 7).

with an IP address that is not blacklisted, and it traverses an international Internet link monitored by the NSA.

iii. A user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link monitored by the NSA.

c. If the NSA excluded the protocols that Wikimedia's communications use:

i. One of Wikimedia's communications is enclosed in a multi-communication transaction (MCT) using a protocol that is not excluded, and it traverses an international Internet link monitored by the NSA.

ii. One of Wikimedia's communications passes through an intermediary using a protocol that is not excluded, and it traverses an international Internet link monitored by the NSA.

6. That the NSA uses an in-line filter architecture to accomplish the upstream collection program in all cases, rather than ever using a copy-then-filter architecture.

155. Given these considerations and my analysis of the NSA's disclosures regarding the upstream collection program, it remains my opinion that it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Date: 3/8/2019

Scott Bradner

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 2

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

REPLY DECLARATION OF JONATHON PENNEY

TABLE OF CONTENTS

I. INTRODUCTION 1
II. DR. SALZBERG'S ANALYSIS IS FLAWED 1
III. RESPONSES TO DR. SALZBERG'S SIX METHODOLOGICAL CRITIQUES 18

I. INTRODUCTION

1. My name is Jonathon Penney. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-00662-TSE (D. Md.), to provide this reply declaration to address the Defendants' reply to the Plaintiff's brief and to my declaration, both of which were dated December 18, 2019. Results of my additional analysis in support of this reply declaration are included herein in the Appendix.

2. My qualifications and expertise are discussed in detail in my opening Declaration ("Declaration"). *See* ECF No. 168-02. Unless otherwise stated, I have personal knowledge of the facts herein.

II. DR. SALZBERG'S ANALYSIS IS FLAWED

3. In support of their Reply motion, Defendants submit the Declaration of Dr. Alan Salzberg ("Salzberg Declaration"). *See* ECF No. 178-3. Dr. Salzberg's analysis misunderstands my interrupted time series (ITS) design and study in fundamental ways and ignores relevant literature on methodological best practices for ITS studies. Furthermore, in critiquing my study, the Salzberg Declaration relies primarily on a visual inspection of data, which can often be misleading,[1] rather than formal testing mechanisms that can be verified. Salzberg's reliance on a visual inspection of the data causes him to formulate misguided critiques and conclusions as to my study's reliability and validity. In actuality, my methodology, method of analysis, and regression model is superior to any alternatives he suggests.

4. **First**, Salzberg's use of disaggregated line plots (see ¶¶ 12-14 and Figure 1 of the Salzberg Declaration) to analyze the Wikipedia page view data is an inferior method of analysis

[1] GENE S. FISCH, *EVALUATING DATA FROM BEHAVIORAL ANALYSIS: VISUAL INSPECTION OR STATISTICAL MODELS?* 54 Behavioural Processes 137, 137 (2001) (quoting Howard Wainer: "A graph is nothing but a visual metaphor. To be truly evocative, it must correspond closely to the phenomena it depicts… If a graphic depiction of data does not faithfully follow this notion it is almost sure to be misleading.").

for ITS studies, as compared to the segmented linear regression trend analysis of aggregated data that I used. Dr. Salzberg's approach ignores relevant literature on methodological best practices. Indeed, the recommended method of analysis for ITS design studies is segmented linear regression analysis, which I employed in my study, as it allows researchers to: (1) control for prior trends in the data; (2) measure the dynamics of change in response to an intervention; (3) tolerate fewer time points than alternative methods; (4) adjust for serial correlation in the data; and (4) apply these methods to aggregate level data.[2] By contrast, Salzberg's disaggregated line-plots method offers none of these strengths or advantages and is neither recommended nor discussed in relevant literature.

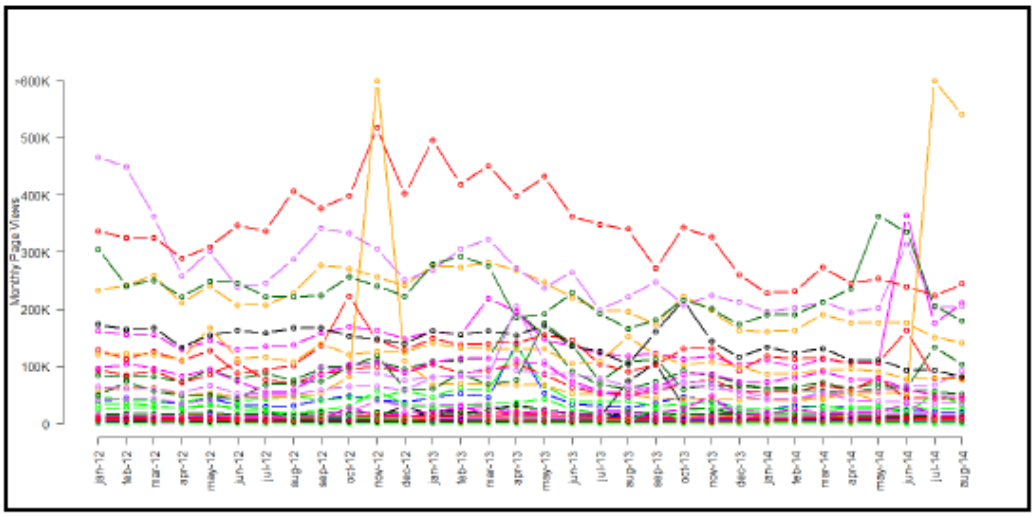
5. **Second**, Salzberg's use of disaggregated line plots adds "noise," both visual and statistical, which masks actual overall trends in the data best understood through analysis of *aggregated* monthly page view data. Consistent with my approach, a majority of ITS design studies use aggregated data.[3] This is to reduce or remove noise in the data and to allow for more

[2] The leading peer reviewed works on ITS design, methodology, and analysis recommends segmented linear regression analysis over several alternatives methods and models, including generalized estimating equations (GEE) method and autoregressive integrated moving average (ARIMA) modeling. FANG ZHANG, A.K. WAGNER, ET AL., *METHODS FOR ESTIMATING CONFIDENCE INTERVALS IN INTERRUPTED TIME SERIES ANALYSES OF HEALTH INTERVENTIONS*, 62:2 J. Clinical Epidemiology 143, 143-144 (2009) (discussing the advantages of segmented linear regression analysis for ITS designs compared to alternatives); A.K. WAGNER ET AL., *SEGMENTED REGRESSION ANALYSIS OF AN INTERRUPTED TIME SERIES IN MEDICATION USE RESEARCH*, 27 J. Clinical Pharmacy & Therapeutics 299, 299, 208 (2002) (describing segmented linear regression analysis as a "powerful statistical method or estimating intervention effects" in ITS studies, and describing "strengths"); MYLENE LAGARDE, *HOW TO DO (OR NOT TO DO) … ASSESSING THE IMPACT OF A POLICY CHANGE WITH ROUTINE LONGITUDINAL DATA,* 27:1 Health Policy and Planning 76, 79 (2012) (noting this method "controls for secular trends and can also adjust for potential serial correlation of the data"); ROBERT B. PENFOLD & FANG ZHANG, *USE OF INTERRUPTED TIME SERIES ANALYSIS IN EVALUATING HEALTH CARE QUALITY IMPROVEMENTS*, 13:6 Acad. Pediatrics S38 (2013) (discussing the advantages and limitations of employing time series analysis to understand and explore the impact of health policy changes).

[3] Wagner (2002), *id.*, at 308 ("Segmented regression typically aggregates individual-level data by time point" and noting a leading ITS study where the "unit of analysis" was a monthly aggregated data, as used in this study); JANDOC, ET AL., *INTERRUPTED TIME SERIES ANALYSIS IN DRUG UTILIZATION RESEARCH IS INCREASING: SYSTEMATIC REVIEW AND RECOMMENDATIONS*, 68 J. Clinical Epidemiology 950, 950 (2015) ("Interrupted time series methods use aggregate data collected over equally spaced intervals before and

sophisticated statistical tests and analysis.[4] Salzberg departs from this approach, which is standard in a majority of ITS studies, leading him to incorrect inferences and conclusions in his analysis. For example, Figure 1 of the Salzberg Declaration distorts and hides important trends by plotting individual line plots for the 48 Terrorism articles. Here, line plots for a majority of the 48 Articles cannot be seen as they have page views too small to be visualized with the large page view scale (0 to 600,000 page views) used on the vertical axis of the graph:

Figure 1: Individual Page Views for Each of the Articles Within the Terror 48, Which The Penney Declaration Hypothesized Show an Immediate Decline Beginning in June 2013



6. Salzberg's Figure 1 creates a false impression there are no patterns or trends for overall page views over 32 months. But those trends are easily visible when individual article page views are analyzed as aggregated monthly page views as visualized in **Figure 1a** below. This

after an intervention, with the key assumption that data trends before the intervention can be extrapolated to predict trends had the intervention not occurred"); EMMA BEARD, *USING TIME-SERIES ANALYSIS TO EXAMINE THE EFFECTS OF ADDING OR REMOVING COMPONENTS OF DIGITAL BEHAVIOURAL INTERVENTIONS AND ASSOCIATIONS BETWEEN OUTCOMES AND PATTERNS OF USAGE*, Centre for Behaviour Change (CBC) Conference, University College of London 15 (2017), https://www.ucl.ac.uk/behaviour-change/events/presentations-17/beard.pdf (noting that a "[m]ajority of studies use aggregated data". She also specifically notes that linear regression may be used for "interrupted time series design" if autocorrelation is controlled).

[4] BEARD, *id.*, at 15 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").

figure plots the median aggregated page views for the 48 Terrorism Articles with trend lines included to understand the shift in trend.

Figure 1a: Aggregated Median Page Views for 48 Terrorism Wikipedia Articles

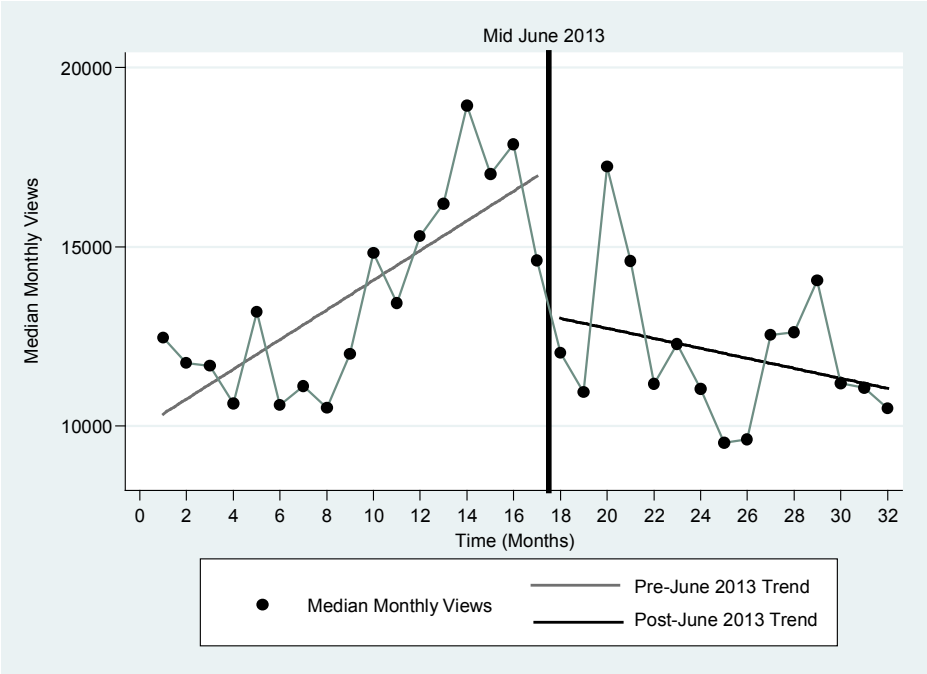
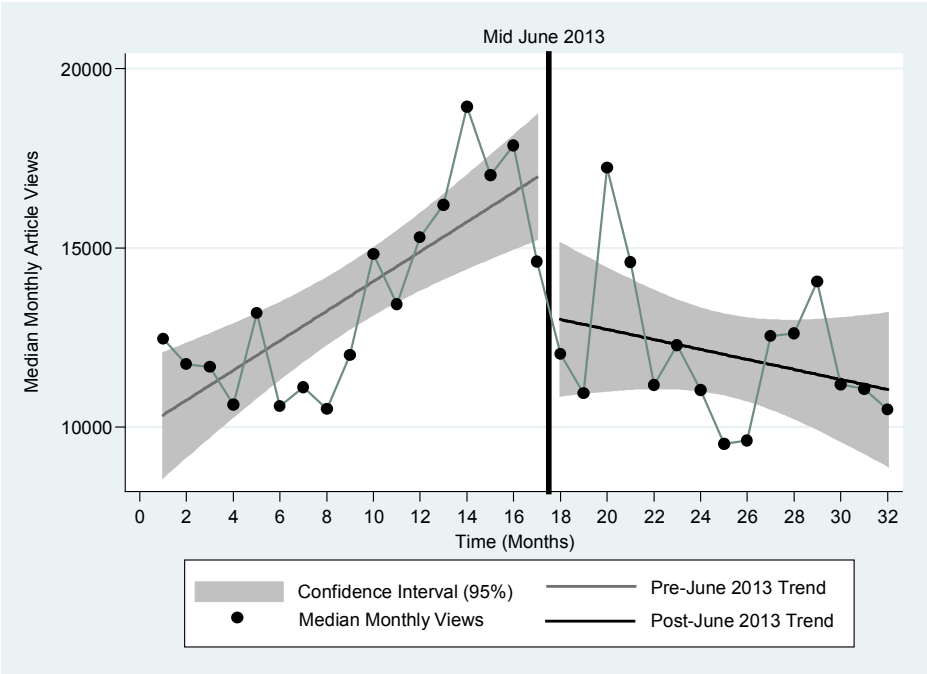


Figure 1b: Aggregated Median Page Views for 48 Terror Articles (With C.I.)

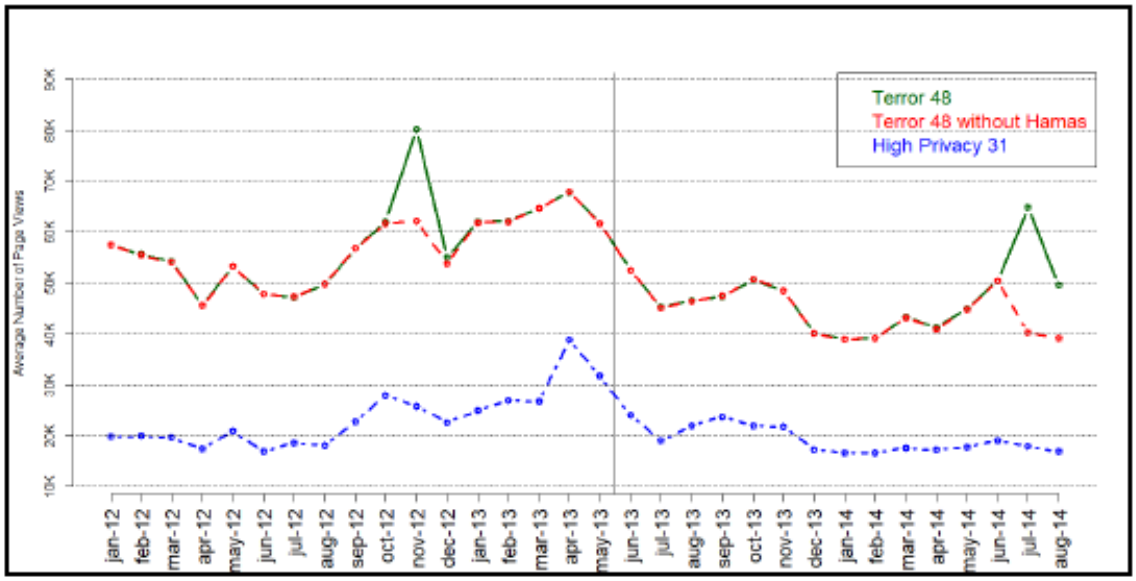


7. **Figure 1b** visualizes the same data with pre/post June 2013 trend lines and

confidence intervals (the gray shaded area), and demonstrates that the trend change before and after June 2013 was statistically significant, as there is no overlap of confidence intervals during these two periods. Specifically, this visualization of the data demonstrates a statistically significant drop in June 2013 and reduction in overall monthly page views. This is a far clearer visualization, with clear trends, compared to Salzberg's Figure 1, which masks these trends in the "noise" of 48 disaggregated line plots.

8. Figures 5 and 6 are of the Salzberg Declaration are similarly distorted. These figures present line-plots for the 48 Terror Articles, 47 Terror Articles, and 31 High Privacy Articles together. For example, Figure 5 visualizes a line plot for the average monthly page views for those three article sets:

Figure 5: Average Page Views Show a Peak in April 2013 or Before



9. Again, by presenting the data associated with these three different sets of articles on the same graph with the same scale on the vertical axis (10,000 to 90,000 average page views), the Figure distort the presentation of the data, creating a false impression that both the 47 Terror Articles and the 31 articles with more privacy sensitive ratings ("31 Higher Privacy Articles") have

a flat trend over the course of the 32 months. These distortions are easy to visualize when the average monthly page views for the 31 Higher Privacy Articles are plotted and presented with an appropriate scale as in **Figures 2a** and **2b** below.

Figure 2a: Average Monthly Page Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Drop-Off After That Month

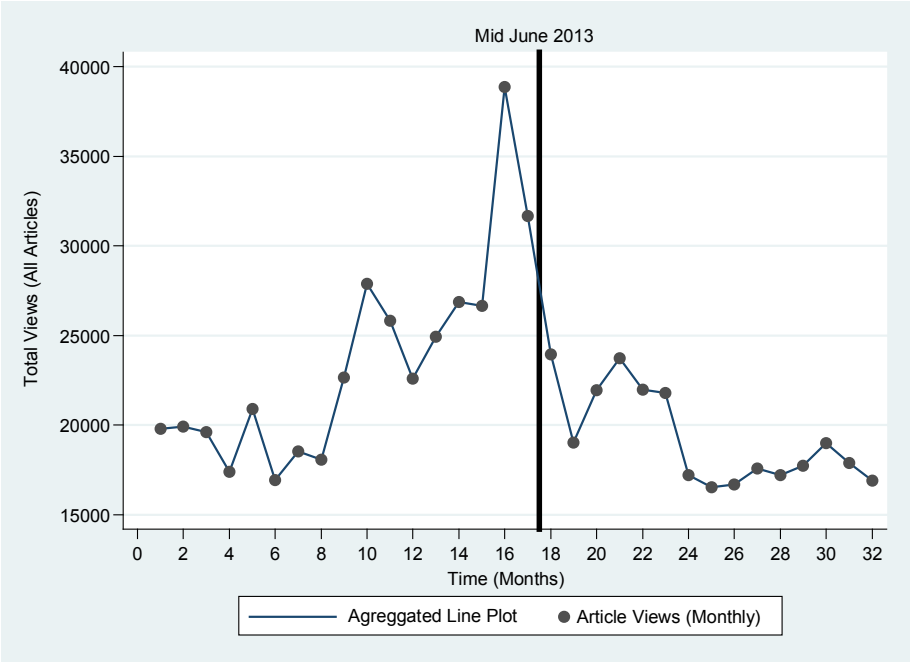
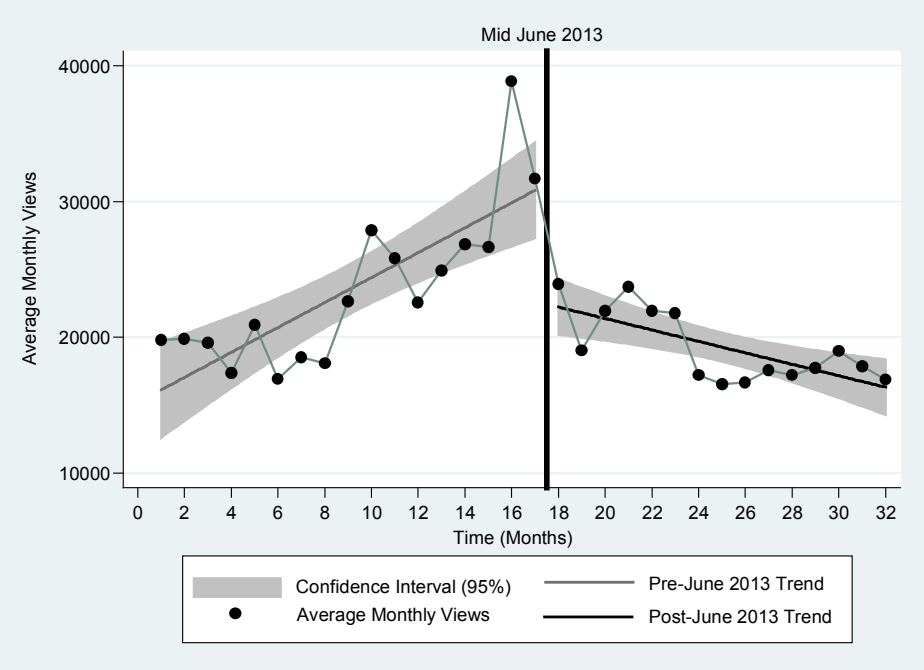


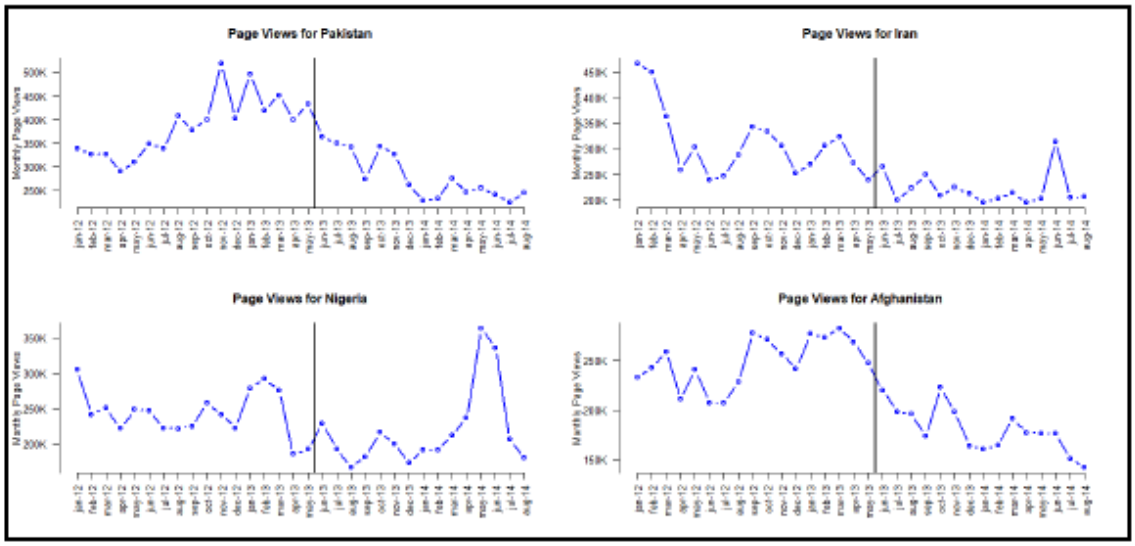
Figure 2b: Average Monthly Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Sharp Drop-Off (With Trend Lines)



10. In **Figure 2a** above, the average monthly page views for the 31 Higher Privacy Articles show a clear trend and not the flattened pattern reflected in Figure 5 of the Salzberg Declaration. With an appropriate scale on the vertical axis, page views increase until mid-June 2013 and declined thereafter. This point is even clearer in **Figure 2b**, which plots the very same data but adds trend lines and confidence intervals for clarity.

11. **Third**, Salzberg focuses his analysis on cherry-picked individual articles that obscure and mislead about actual trends in the data. For example, Figure 2 of the Salzberg Declaration visualizes monthly page view line-plots for four articles: Pakistan, Iran, Nigeria, and Afghanistan:

Figure 2: Individual Articles show no Association of June 2013 with a Decline in Page Views



12. While these four Wikipedia articles did form part of the 48 Terror Article set, they have among the lowest privacy-sensitivity scores among all articles in the set (*see* Table 12 of my Declaration). That is, these articles raised few privacy concerns for survey participants. As discussed in my opening declaration, 415 independent Internet users participated in a survey in which they provided feedback on how keywords associated with each of the 48 Terrorism Articles may raise privacy-related concerns ("Privacy Evaluation Survey").[5] In the Privacy Evaluation Survey, the combined average privacy-sensitivity rating for all 48 Terror Articles was 2.15 and the median was 2.07. The articles that Salzberg cherry-picked fell far below that mean and median: Pakistan (1.82), Iran (1.85); Nigeria (1.71); Afghanistan (1.83). Since the hypothesis that I tested in my study concerns a ***privacy***-based chilling effects theory [6]—i.e., that Wikipedia users avoided privacy-sensitive Wikipedia articles due to awareness of NSA Upstream Surveillance—it is inappropriate to rely on trends for these four articles, which fell far below the mean and median privacy-sensitive rating. As such, to the extent Salzberg isolates and relies on these four articles to

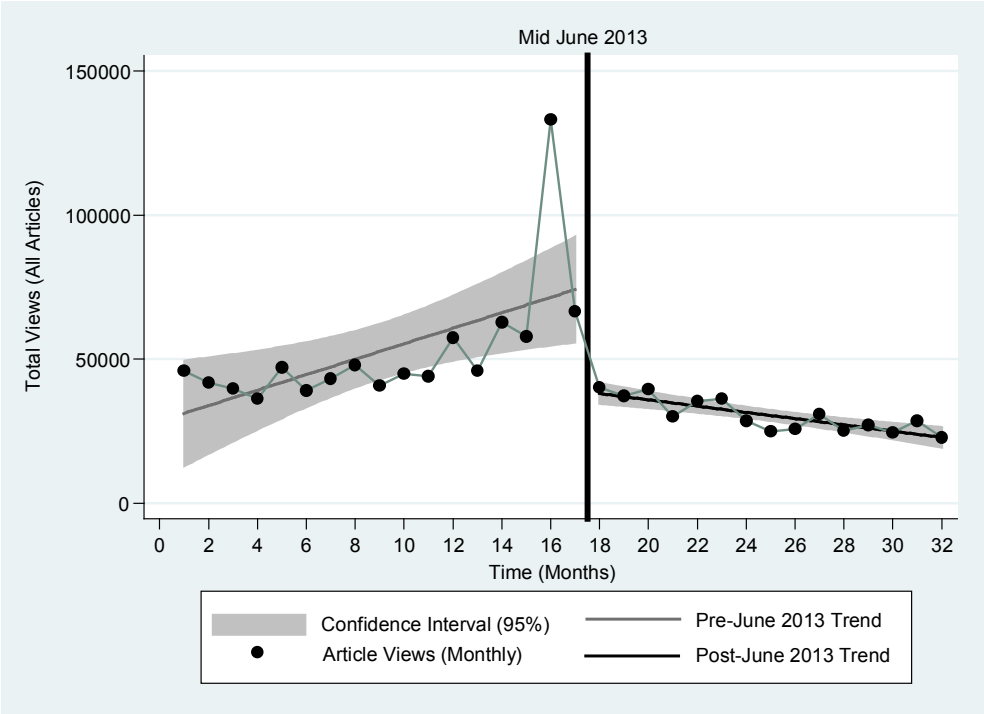
[5] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[6] See ¶¶12-21 of my Declaration dated December 18, 2018.

reject a chilling effects hypothesis (*see* paragraph 16 of the Salzberg Declaration), his analysis is misleading, unreliable, and masks actual trends found in my study.

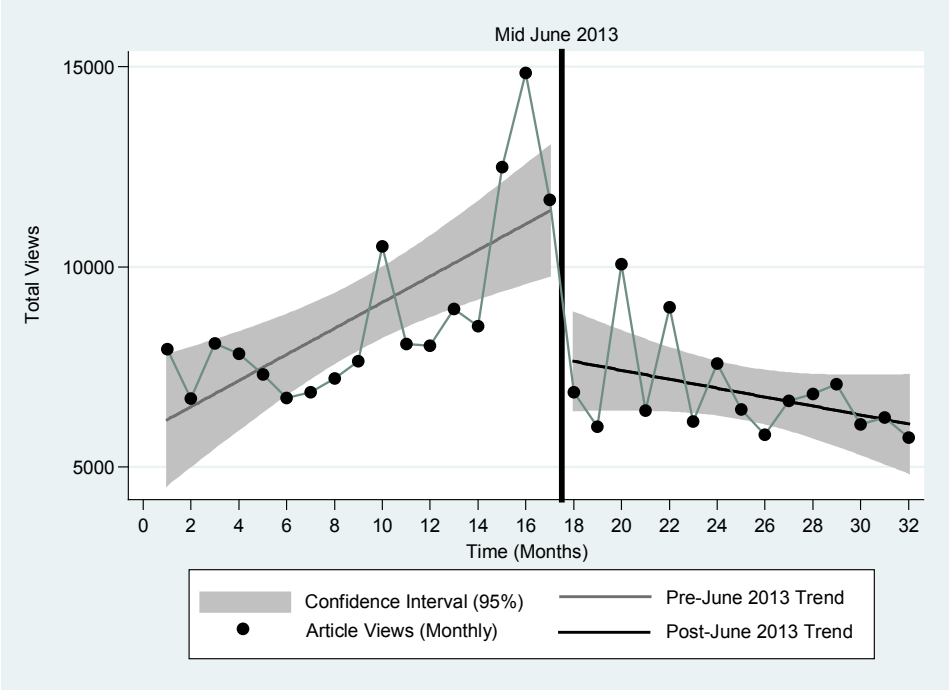
13. **Fourth**, the Salzberg Declaration focuses on disaggregated line-plots that mask aggregated data trends or less privacy-sensitive or privacy-concerning articles like the four articles noted above (Pakistan, Iran, Nigeria, and Afghanistan). This approach is inappropriate given my study tests a chilling effects hypothesis based on a privacy theory. Moreover, Salzberg's approach leads to a flawed analyses and conclusions. In fact, the page views for the four articles (among the 48 Terrorism Articles) with the very highest privacy-sensitivity scores according to the Privacy Evaluation Survey—improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), and ammonium nitrate (2.61)—are entirely consistent with a chilling effects hypothesis in June 2013. **Figure 3a** depicts monthly page views for each of these most privacy-sensitive articles. Each figure demonstrates page view trends consistent with a chilling effects hypothesis: a monthly increase in page views leading up to June 2013, an abrupt statistically significant decline and subsequent change in trend to a monthly decrease in page views.

Figure 3a: Page Views For The Four Most Privacy-Sensitive Articles Are Consistent With A Chilling Effects Hypothesis in June 2013

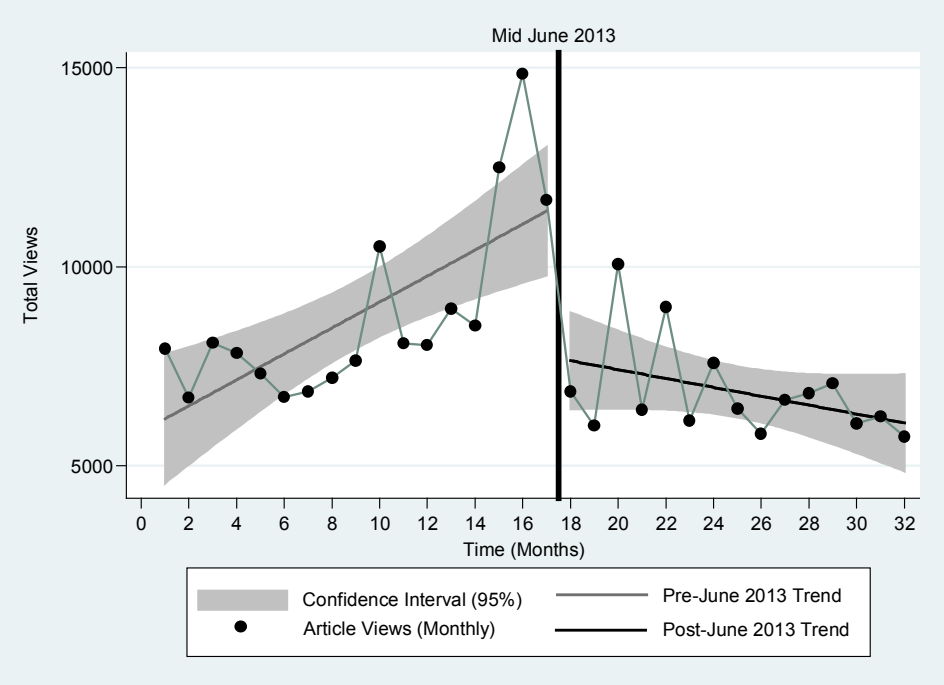
Improvised Explosive Device



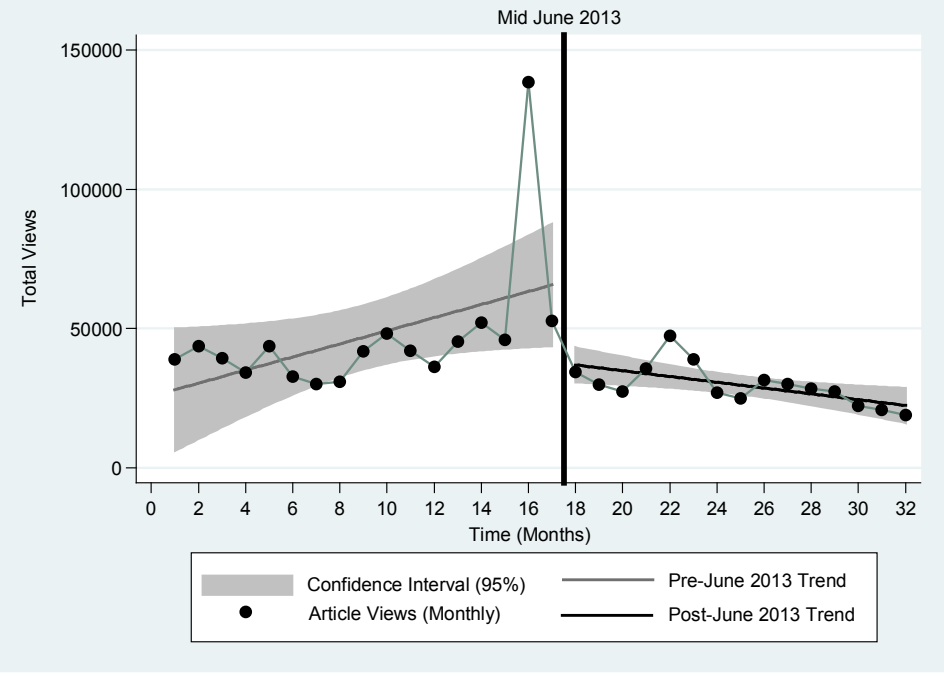
Dirty Bomb



Car Bomb

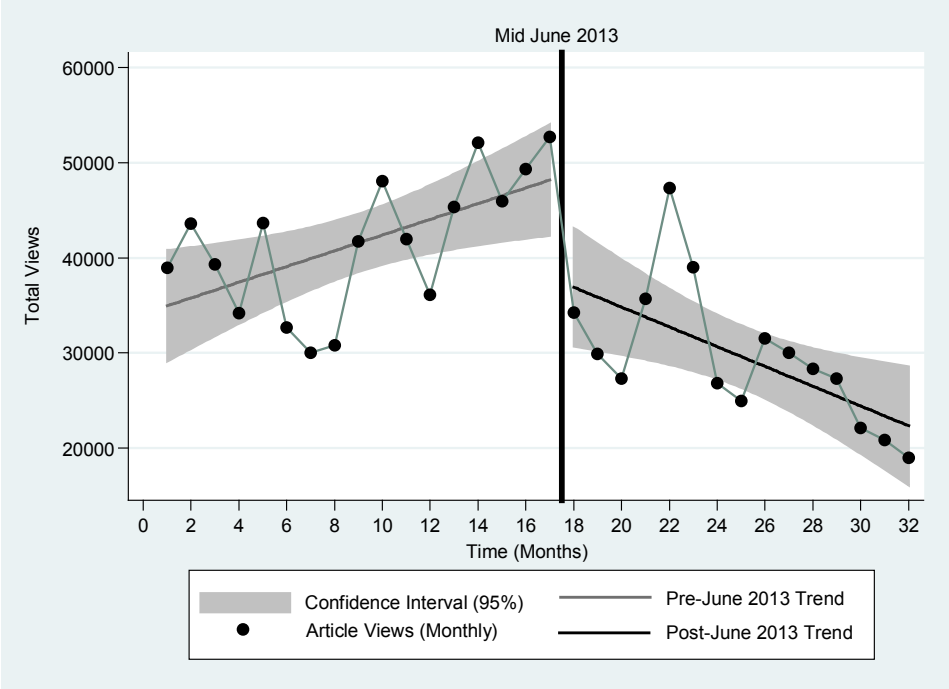


Ammonium Nitrate



14. Salzberg claims that the "ammonium nitrate" article is an outlier in its number of page views in April 2013.[7] Even assuming this is correct, if the article's page views for that month are normalized,[8] the overall trend for the ammonium nitrate article remains consistent with a chilling effect hypothesis. **Figure 3b** provides the ammonium nitrate article's normalized page views over 32 months, which are consistent with a June 2013 chilling effects hypothesis: increasing monthly articles views in the months leading up to June 2013, and then an abrupt statistically significant decline in June and a subsequent monthly reduction in views:

Figure 3b: Page Views for Normalized Ammonium Nitrate Article Are Consistent With June 2013 Chilling Effects Hypothesis



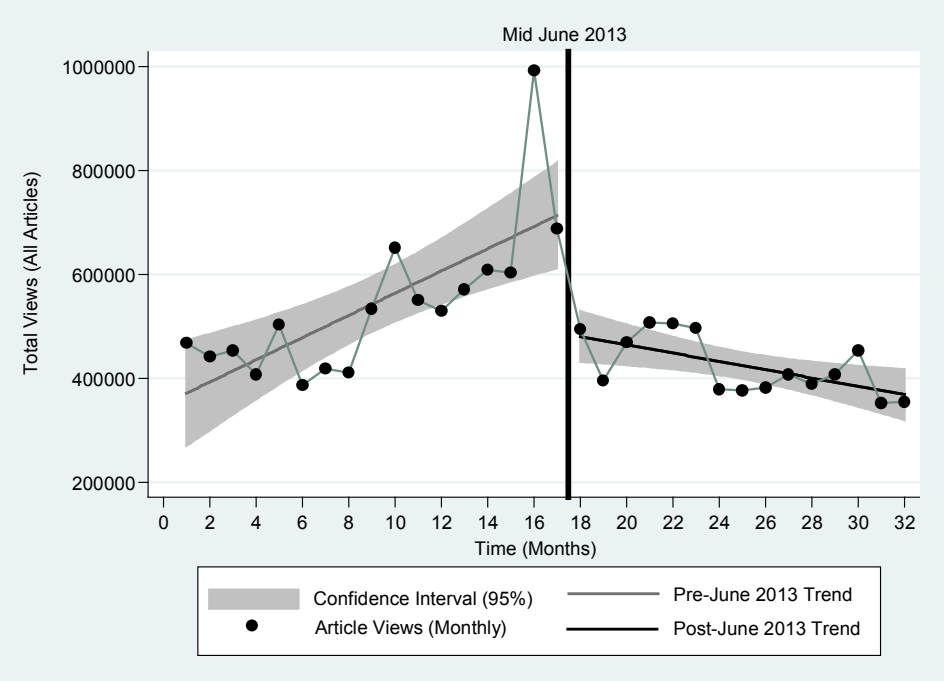
[7] See ¶ 60 of the Salzberg Declaration.

[8] I replaced the outlier value for the article in April 2013 (138363) with an average of the total page views for the article (49316) in the two adjacent months (May and March 2013). Correcting, modifying, or deleting an outlier value or observation in a data set is consistent with best-practices in dealing with outliers: HERMAN AGUINIS, RYAN K. GOTTFREDSON & HARRY JOO, *BEST-PRACTICE RECOMMENDATIONS FOR DEFINING, IDENTIFYING, AND HANDLING OUTLIERS*, Organizational Res. Methods 8, 20–23 (2014), http://orm.sagepub.com/content/early/2013/01/11/1094428112470848.abstract ("Once error outliers have been identified, the correct procedure is to either adjust the data points to their correct values or remove such observations from the dataset").

15. Furthermore, the aggregate total monthly page views, average monthly page views, and median monthly page views for the 23 most privacy-sensitive articles among the set of 48 Terrorism Articles in the study[9] are also consistent with a chilling effect hypothesis in June 2013, when measured over a 32-month period. **Figure 4** visualizes monthly median page views for these 23 most privacy sensitive Wikipedia articles:

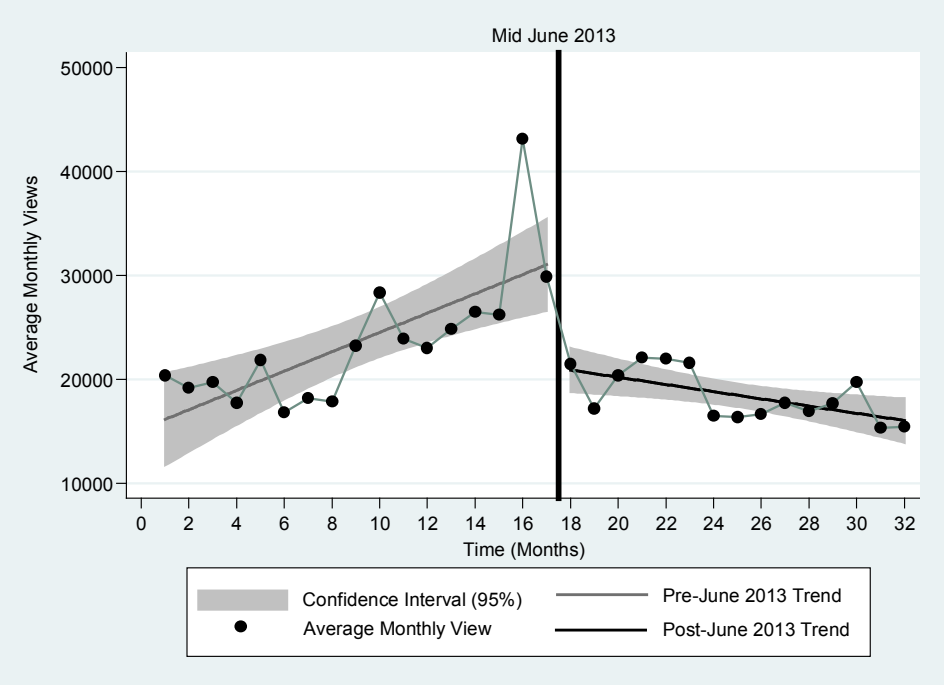
Figure 4: Total Monthly Page Views, Average Monthly Page Views, and Median Monthly Page Views for 23 Most Privacy-Sensitive Wikipedia Articles Over 32 Months Are All Consistent With A Chilling Effects Hypothesis

Raw Total Monthly Page Views: 23 Most Privacy Sensitive Articles

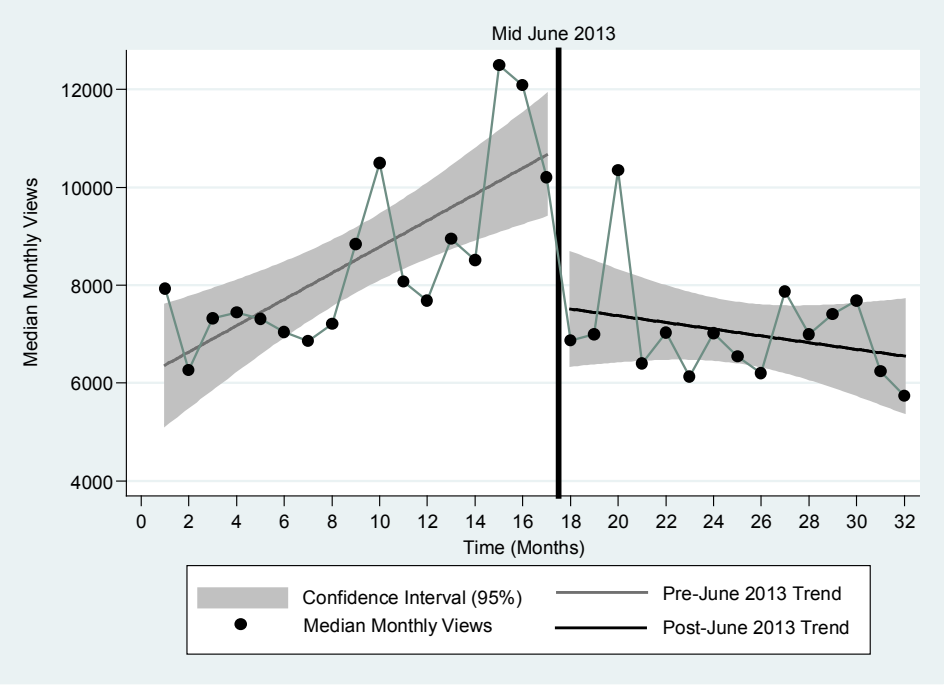


[9] This set of the 23 Most Privacy-Sensitive Article includes all articles in the 48 Terrorism Article group with a combined average privacy-sensitivity score greater than the median of those combined scores (2.07). This group includes: improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), ammonium nitrate (2.61), biological weapon (2.60), chemical weapon (2.51), suicide attack (2.50), suicide bomber (2.44), Nuclear Enrichment (2.39), environmental terrorism (2.39), eco terrorism (2.39), weapons grade (2.39), jihad (2.35), Al Qaeda (2.34), terrorism (2.30), conventional weapon (2.27), Taliban (2.22), AL Qaeda in the Arabian Peninsula (2.17), Al Qaeda in the Islamic Maghreb (2.17), terror (2.15), Abu Sayyaf (2.14), Tehrik-i-Taliban Pakistan (2.12), and attack (2.08).

Average Monthly Page Views: 23 Most Privacy Sensitive Articles



Median Monthly Page Views: 23 Most Privacy Sensitive Articles



16. Here, in each graph set out in **Figure 4**, the total raw, average, and median monthly page views for these 23 articles increase in the time period leading up to June 2013. They also

demonstrate a statistically significant decline in June 2013, as well as a statistically significant trend reversal, with monthly page views declining after June 2013. Again, these findings are entirely consistent with a chilling effect hypothesis.

17. In the end, these results—focused on the most privacy-sensitive articles, analyzed both on an individual disaggregated analysis and aggregate monthly analysis—are entirely consistent with a chilling effect hypothesis. By ignoring the privacy theory upon which the chilling effect hypothesis is based, the Salzberg Declaration is deeply flawed.

18. **Fifth,** one of Salzberg's primary critiques of my analysis rests on a false premise: that my study "assumes a single peak in May 2013." This premise is false because my study makes no such assumption. My study hypothesizes a surveillance chilling effect in June 2013. Consistent with other ITS design studies, my study analyzes an outcome variable measured at consistent intervals (monthly privacy-sensitive Wikipedia article view data) to test that hypothesis over 32 months, by examining for statistically significant changes in level and trend in that data both before and after June 2013.

19. **Sixth,** Salzberg claims that my study's model can be altered to "prove" an April 2013 peak or earlier peak (based on a theory that the Boston Marathon bombings caused the page view trend reversal). However, he cites no cross-validation analysis to compare models and results to support his claim.[10] Cross-validation analysis is an established technique for understanding whether the results of a statistical test are robust – i.e., if we leave out datapoints at random from our dataset, do the results still hold? To answer that question and disprove Salzberg's theory, I

[10] LUKE JOHN KEELE, SEMIPARAMETRIC REGRESSION FOR THE SOCIAL SCIENCES 86 (Wiley & Sons, 2008) (describing "cross-validation" as a "general technique for assessing model fit based on resampling that can be applied to most statistical models").

conducted a "leave one out" cross-validation analysis[11] on both the 23 Most Privacy-Sensitive Wikipedia Articles set, as well as the larger 47 Terrorism Article set (the 48 Terrorism Articles without the Hamas article) to compare different statistical models based on a March, April, May, or June 2013 intervention effect. For comprehensive analysis, I used three data sets for each of these article sets—raw total page views, average monthly page views, and median monthly page views. Furthermore, for the 47 Terrorism Article Set, I excluded the "fundamentalism" article, making it a set of 46 total articles. (Salzberg noted that the "fundamentalism" article had too similar values to the "recruitment" article in the broader 48 Article set, so I have excluded it from this supplemental analysis. (Salzberg Decl. ¶ 7.)) I also performed the analysis both including and excluding the "ammonium nitrate" and "jihad" articles from the sets (Salzberg claims these two articles have outlier values).

20. The results of this cross-validation analysis show that for the 23 Article set of the Most Privacy-Sensitive Wikipedia articles, a statistical model based on a June 2013 intervention effect was superior to models based on March, April, and May 2013 interventions ***in every single data set analyzed*** (raw total monthly page views, average monthly page views, median monthly page views). That is, a June 2013 statistical model resulted in fewer estimation errors (lower room mean square errors (RMSE) and mean absolute errors (MAE)) than the other models. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis.

21. For the 46 Terrorism Article Set, a statistical model based on a June 2013

[11] *Id*. at 8 (describing "leave one out" cross validation as "probably the most commonly used method" as it "works well with most any sample size"; also that with "leave-one-out cross-validation, one observation is randomly selected and then omitted from the data set. The analyst then fits one of the possible models to this slightly truncated data set and calculates measure of fit. Next, a new data point is dropped, and the measure of fit is calculated again. This process is repeated as each of the data points is removed from the data set. The cross validation score is the averaged measure of model fit and can be used to compare different model specifications.")

intervention effect was also superior to models based on March, April, and May interventions for both the raw monthly page views set and the average monthly page views data set. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis. For the median monthly page view set, a June 2013 model was also superior to models based on both April and March 2013 interventions, but not May, where the RMSE and MAE scores for both models were very close (only a 2.5% difference in the MAE and 5% difference in the RMSE scores). These results all held or without "ammonium nitrate" and "jihad" in the sets.

22. In short, these results demonstrate the strength and robustness of my June 2013 model and its findings: it proved superior to comparable models in 46 of 48 total tests, and even in those two remaining tests, the difference in results were minimal. Moreover, when focused on the most privacy-sensitive Wikipedia articles, whether including or excluding the "jihad" and "ammonium nitrate" articles that Salzberg claims were outliers due to the Boston Marathon Bombing-- my model was a better "fit" to the data than the alternative models that Salzberg proposes based on earlier interventions in every single data set and scenario.

23. **Seventh,** Salzberg's comparative analysis of recent page view data is fundamentally undermined by the fact that Wikimedia's "page view" definition has changed over time. (*See* Salzberg Decl. ¶¶ 27-32.) Wikimedia has publicly published explanations on recent changes in the page view definition. (*See, e.g.*, Wikimedia Downloads: Analytics Datasets, https://dumps.wikimedia.org/other/analytics/; Research:Page View, https://meta.wikimedia.org/wiki/Research:Page_view.) The page view data that Salzberg relies on from July 2015 through November 2018 includes data on *mobile* page views, and therefore is incomparable to the data from the time period that I studied. As I explained in my Declaration, my study "used data for English language article view counts from stats.grok.se, an online portal

that provided access to *non-mobile* Wikipedia article view count data on a daily and monthly basis." (Decl. ¶ 34 (emphasis added); *see id.* at Table 3, 8, 9 (expressly indicating non-mobile data used.) Salzberg provides an example link to the Pageviews tool that he used to gather the more recent data, which shows that he selected all "Platform" types, including mobile. (*See* Salzberg Decl. ¶ 27, n. 17.) The difference between page views with non-mobile vs. mobile data included is often very significant, and therefore Salzberg's "extended data" comparison analysis is deeply flawed at the source and should be ignored. For example, using the "Hamas" example Salzberg offers, the difference between the "All" platforms and the "Desktop" (non-mobile) data for the month of May 2018 is over 100,000 views.

III. RESPONSES TO DR. SALZBERG'S SIX METHODOLOGICAL CRITIQUES

24. In his Declaration, Salzberg presents a series of purported critiques regarding my analysis. (Salzberg Decl. ¶¶ 47-66.) I respond to these issues below.

25. **Salzberg's first critique**: Aggregation "masks the differences in the changes over time by article" and was "performed without any analysis of the individual datasets" to determine whether it was the appropriate method. Standard methods for analyzing this kind of "panel data" were ignored. (Salzberg Decl. ¶¶ 48–50.)

26. **My Response:**

(a) As noted earlier, my method of analysis to test the June 2013 surveillance chilling effect hypothesis was an ITS design using aggregated data with segmented regression trend analysis. I chose an ITS design because it is an "ideal design" for assessing the impact of a "population-wide" intervention—like the effects of mass online government surveillance—that "affects the whole population and where randomization or a control group is impossible."[12] ITS

[12] N. BRUCE BASKERVILLE, ET AL., *IMPACT OF CANADIAN TOBACCO PACKAGING POLICY ON USE OF A TOLL-FREE QUIT-SMOKING LINE: AN INTERRUPTED TIME-SERIES ANALYSIS*, 6(1) CMAJ Open E59, E64 (2016)

design studies have also been commonly used in contexts like this one, to study information systems context (e.g., computing context)[13] and the impact of media coverage.[14] Within ITS design studies, use of segmented regression to analyze aggregated data to understand pre/post intervention trends in the data is not only "standard," but the ***recommended*** method and approach.[15]

(b) Second, there is no single determinative method or factor to decide whether an aggregated or disaggregated analysis of data is appropriate. Most ITS design studies use aggregated data,[16] because such time series designs "examine aggregate effects"[17] and are "strong designs for estimating the effects of instituting uniform, full-coverage programs or the effects of

---

(discussing "interrupted time-series design" as an "ideal design for assessing the effects of a population-wide intervention"; a "robust method for the evaluation of a policy that affects the whole population and where randomization or a control group is impossible"); RICHARD MCCLEAR ET AL., DESIGN AND ANALYSIS OF TIMES SERIES EXPERIMENTS 7, 297 (2017) (describing interrupted time series designs as "the major application of time series data for causal inference" and as a "strong quasi-experimental design… when random assignment was unfeasible"); CHESTER L. BRITT, DAVID J. BORDUA, & GARY KLECK, *A REASSESSMENT OF THE D.C. GUN LAW: SOME CAUTIONARY NOTES ON THE USE OF INTERRUPTED TIME SERIES DESIGNS FOR POLICY IMPACT ASSESSMENT*, 30 Law & Soc'y Rev. 361, 361 (1996) ("Interrupted time series designs provide one of the most common means for assessing the impact of a change in law or in social policy"); D.T. CAMPBELL, *REFORMS AS EXPERIMENTS*, 24(4) American Psychologist 409 (1969) (this seminal article by Campbell was among the first to advocate for interrupted time series designs in cases where natural experiments are not possible); WAGNER ET AL., *supra* note 1, at 308 (describing ITS designs as the "strongest, quasi-experimental designs" to estimate intervention effects in "non randomized settings").

[13] *See e.g.*, S. ASGARI & NUNES BAPTISTA, *EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH IN INFORMATION SYSTEMS,* IADIS International Workshop Information Systems Research Trends: approaches and methodologies (ISRTAM 2011), 20-26 July (noting ITS designs have been "used often in the field of [Information Systems]").

[14] *See e.g.,* MELANIE A WAKEFIELD, BARBARA LOKEN,& ROBERT C HORNIK, *USE OF MASS MEDIA CAMPAIGNS TO CHANGE HEALTH BEHAVIOR*, 376 The Lancet 1261, 1262-1263, (2010) (discussing interrupted time series analyses studies in the health context); RANDY ELDER ET AL., *EFFECTIVENESS OF MASS MEDIA CAMPAIGNS FOR REDUCING DRINKING AND DRIVING AND ALCOHOL-INVOLVED CRASHES*, 27(1) Am. J. Prev. Med. 57 (2004) (ROBERTO GRILLI ET AL., *MASS MEDIA INTERVENTIONS: EFFECTS ON HEALTH SERVICES UTILIZATION*, 1 Cochrane Database of Systematic Reviews 1, 1 (2002) (providing a comprehensive review of research studying the impact of media coverage on health service use— and noting that among the "twenty studies" reviewed in the work, all used interrupted time series designs).

[15] See works cited at *supra* note 2 and accompanying text.

[16] See also works cited at *supra* note 3 and accompanying text.

[17] JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS A MODERN APPROACH 15 (5th ed., 2012) ("…time series data are often used to look at aggregate effects. An example of a time series data set on unemployment rates and minimum wages…").

making changes in such programs."[18] NSA surveillance, including Upstream surveillance, was "uniform" and "full coverage," within the meaning of this guidance. It also has aggregate effects, as the entire U.S. Internet-using population is subject to their reach. It was therefore appropriate that my study employed aggregate page view analysis, since I sought to make aggregate level inferences about large scale NSA surveillance effects.[19]

(c) Third, aggregated data and analysis is further appropriate in ITS studies where the aim is to explore national or major regional rates and trends;[20] to reduce or remove "noise" in the data;[21] and to allow for more sophisticated statistical tests and analysis.[22] All of these circumstances apply to this study. First, my study was focused on evaluating the large-scale national aggregated effects or impact of mass awareness of NSA surveillance in June 2013 and after. Thus, examining the Wikipedia article page view data in aggregate is consistent with that aim. Second, analyzing the Wikipedia articles in aggregate helped reduce "noise" in the data given that, inevitably, individual Wikipedia article page views would like fluctuate and vary widely over 32 months; if analyzed in aggregate, broader overall trends or patterns could be discerned. Analyzing aggregated data makes particular sense in this ITS study, as Wikipedia page view data has a particularly high signal-to-noise ratio—that is, where the signal or "true" patterns in data (like longer term trends due by chilling effects) may be obscured by "noise," that is, more

---

[18] PETER ROSSI ET AL., EVALUATION: A SYSTEMATIC APPROACH 352-354 (6th ed., 1999) (noting "[m]ost existing times series involve aggregated data" as they often aim to study national, state, or large regional subjects).
[19] L. LEYDESDORFF, *THE SCIENCE CITATION INDEX AND THE MEASUREMENT OF NATIONAL PERFORMANCE IN TERMS OF NUMBERS OF SCIENTIFIC PUBLICATIONS*, 1-2 Scientometrics 111, 113 (1989) ("In general, one should prefer aggregated data for inferences at the aggregated level, since otherwise methodological problems of inference may emerge.").
[20] ROSSI ET AL., *supra* note 18, at 354.
[21] BEARD, *supra* note 4, at 2 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").
[22] *Id*.

temporary variations for individual article page views for other reasons.[23] Several prior studies have observed that Wikipedia page view data has such "noise" and those studies likewise used aggregated page views for analysis.[24] Third, analyzing the Wikipedia article page view data in aggregate also allowed for the more sophisticated statistical tests and analysis. In this case, that was segmented regression trend analysis pre/post June 2013, which is the recommended method of analysis for ITS design studies. In short, an aggregated analysis of the Wikipedia article page view data was both a "standard" method and entirely justifiable.

(d) Fourth, my opinions *are* supported by an analysis of individual article page views. Specifically, I examined the page view trends for individual and smaller groups of articles with higher privacy-sensitivity scores to verify the results of my aggregate data analysis showing: (1) a statistically significant drop in June 2013, and (2) trend change in monthly page views from increasing views before that month to a monthly decline after that month. As the chilling effect hypothesis I was testing is based on privacy sensitivity, then page view trends for the most privacy-sensitive Wikipedia articles would reveal page view trends consistent with the aggregate data results. As seen from **Figures 3a, 3b, and 4**, those page views were consistent with a June 2013 chilling effects hypothesis. Furthermore, I also examined and analyzed individual article page views to identify and investigate outliers. For example, this was done with the "Hamas" article, among others. Examining individual and aggregate level data allowed me to identify overly

---

[23] N. GENEROUS, ET AL., *GLOBAL DISEASE MONITORING AND FORECASTING WITH WIKIPEDIA*, 10(11) PLoS Comput Biol 1, 8, 12 (2014)(study on using Wikipedia page view data to track and predict global diseases using aggregated data as "signal-to-noise" ratio in Wikipedia page view data may mean "[t]rue patterns in data may be obscured by noise", in this case, the "noise" being variations in page views of health information on Wikipedia unrelated to the personal diagnosis); J.D. SHARPE, ET AL., *EVALUATING GOOGLE, TWITTER, AND WIKIPEDIA AS TOOLS FOR INFLUENZA SURVEILLANCE USING BAYESIAN CHANGE POINT ANALYSIS: A COMPARATIVE ANALYSIS*, 2(2) JMIR public health and surveillance 1, 2, 4 (2016) (noting the "signal-to-noise" ratio in Wikipedia page view data can be "problematic" and also aggregating).

[24] *Id*.

influential articles. This is a standard aspect of conducting regression model diagnostics. Additionally, I used individual articles as part of my cross-validation analysis.

(e) Fifth, Salzberg describes my data as "panel data." This is accurate in the broad definition of the term—my data is "longitudinal," involving repeated measurements from a group to study the impact of an intervention (monthly measurements of individual Wikipedia article page views aggregated into the larger sets).[25] However, using more precise definitions informed by the aim and design of my study, the data is more accurately described as time series data. This is because the subject of my study is the aggregate Wikipedia page view trends both before and after June 2013 in order to test the aggregate effects of NSA surveillance on Wikipedia and its users. In short, *time* is a central unit of analysis in my study.[26] By contrast, panel data and panel studies typically follow *individuals* over time, and include multiple-dimensional observations from each individual (e.g., schooling, employment, marital status, training, child rearing, health, etc.).[27] This conclusion is supported by Woolbridge in *Introductory Econometrics: A Modern Approach*, the same text Salzberg cites. It states that a "time series data set consists of observations on a variable or several variables over time," with time an "important dimension in a time series data set."[28] As I was interested in examining the aggregate effects of media coverage of NSA surveillance introduced at a specific point in time (June 2013), and trends before and after that time, the "variable" in my study was aggregate Wikipedia page views, and a central dimension was time. Woolbridge also describes panel data as longitudinal data that follows or "attempts to follow" individuals over time, collecting multiple observations on a range of different data points,

[25] CHENG HSIAO, PANEL DATA ANALYSIS 1 (CUP, 2017); ROSSI ET AL., *supra* note 18, at 352.
[26] GREGORY B. MARKUS, ANALYZING PANEL DATA 7-8 (sage, 1979) (describing the difference between time series data and panel data, with the former having "time" as the central unit of analysis, while panel studies take individuals as the central unity of analysis).
[27] HSIAO, *supra* note 25, at 1; MARKUS, *id* 7-8.
[28] WOOLBRIDGE, *supra* note 17, at 8.

with repeated measures not necessarily uniform or taken at regular intervals.[29] This creates additional layers of complexity for analysis suited for panel studies.[30] The "standard" methods for that Salzberg recommends at paragraph 50 of his Declaration, in fact, concern panel data analysis, specifically data collected at only two or three points in time.[31] These are not applicable to my time series data set collected at regular intervals over 32 months.

27. **Salzberg's second critique**: Penney's model "assumes a single peak in May 2013" rather than "letting the data reveal where, if anywhere a peak in the data exists." Penney's model can be altered to "prove" an April 2013 peak and support the theory that the Boston Marathon bombings caused the page view trend reversal. A "polynomial model" further shows that Penney's hypothesized peaks in page views were incorrect. (Salzberg Decl. ¶¶ 51–54.)

28. **My Response:**

(a) First, my ITS design does not "assume a single peak"—it tests for the effects of real-world events happening at a particular time—June 2013. In order to test a surveillance chilling effect hypothesis in that month, I examined page view trends before an intervention point. This is a standard approach in naturalistic studies like this, where the aim is to test the impact of an intervention at a given point in time (here, NSA surveillance revelations and media coverage in June 2013). Any peak before or after June 2013 arises from the data itself, and is not any assumption or requirement in my model.

(b) Second, Salzberg's approach of "letting the data reveal where, if anywhere, a peak in the data exists," is not a sound or reliable social scientific approach and can lead to substantial bias in results. My ITS study, designed to examine a June 2013 impact due to media

[29] WOOLBRIDGE, *supra* note 17, at 448.
[30] WOOLBRIDGE, *supra* note 17, at 448.
[31] WOOLBRIDGE, *supra* note 17, at 459-474.

coverage of NSA surveillance programs, was based on a hypothesis grounded in existing empirical and theoretical research on privacy, surveillance studies, and chilling effects. By contrast, Salzberg's approach of visually inspecting data and running various statistical models, including a "polynomial model," until a "fit" showing "earlier peaks in 2013" is found, is a biased approach.[32] Any earlier "peak" or polynomial model is not grounded on any *a priori* hypothesis, theory or research. Nevertheless, even assuming a "polynomial model" estimating earlier peaks in 2013 is not biased. It also does not discount or disprove a chilling effects hypothesis in June 2013. My statistical model and results did not require or assume any such "peak" in May. A statistically significant drop in June 2013 based on prior trends, or a reduction in monthly views thereafter, would each be consistent with a chilling effect hypothesis, notwithstanding any earlier peaks in the page view data in the 32 months. My analysis found both, but neither required any "peak" in May or April 2013.

(c) Third, my statistical model based on a June 2013 intervention is a superior fit for the page view data compared to models based on March, April, and May 2013 interventions in every single data set analyzed. This is demonstrated by my cross-validation analysis on both the 23 Article Set (raw monthly page views, average monthly page views, median monthly page views) and the 28 Article Set (raw monthly page views, average monthly page views, median monthly page views).[33] My June 2013 model was a better "fit" when focused on the page view data for the most privacy-sensitive Wikipedia articles compared to Salzberg's alternative models,

---

[32] ANDREW GELMAN & ADAM ZELIZER, *EVIDENCE ON THE DELETERIOUS IMPACT OF SUSTAINED USE OF POLYNOMIAL REGRESSION ON CAUSAL INFERENCE*, Research & Politics (2015) (describing reported effects, based on a "curve fitting" polynomial model that is "statistically significant but substantively dubious, and are sensitive to model choice"); MEGAN L. HEAD ET AL., *THE EXTENT AND CONSEQUENCES OF P-HACKING IN SCIENCE*, 13(3) PLoS Biol 1 (2015) ("Inflation bias… occurs when researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results.").

[33] *See supra* ¶¶ 16-17.

notwithstanding any impact of the Boston Marathon media coverage in April or May 2013. In short, my ITS approach, statistical regression model, and method of analysis is the best way to understand the page views trends for the most privacy-sensitive Wikipedia articles.

29. **Salzberg's third critique:** Penney's model is "oversimplified, leaving out virtually all factors that could affect page views of terror-related articles from the model." For example, the model fails to account for seasonality or major news events. The model tacitly acknowledges this failure in how it handles the Hamas outlier data, which is ultimately manipulated in a way that is favorable to the hypothesis. Penney does not consider other real world variables that may not be favorable to the hypothesis, like the Boston Marathon bombings, which happened shortly before the NSA disclosures. (Salzberg Decl. ¶¶ 55–60.)

30. **My Response:**

(a) My analysis ***does*** account for various external factors that may affect page views in the model. First, while seasonality is a confounding concern in ITS designs, there is no basis to expect large seasonal effects with these page views—that, for example, Internet users tend to view terrorism-related content in the spring but not in the summer. In any event, to account for possible seasonality and seasonal effects in the data in the ITS design, I went beyond the "general recommendation" for 12 data points before and after the hypothesized intervening chilling effect (June 2013),[34] and instead collecting data for 17 data points before (January 2012 through May 2013) and 14 data points following (July 2013 through August 2014). This longer study period allows for better assessment of overall page view trends by identifying and accounting for seasonal trends. This is particularly challenging with Wikipedia page view data because of its high signal-

[34] WAGNER ET AL., *supra* note 2, at 301 ("A general recommendation is for 12 data points before and 12 data points after the intervention (8), although this number is not based on estimates of power. Rather, with 24 monthly measures, the analyst can adequately evaluate seasonal variation.").

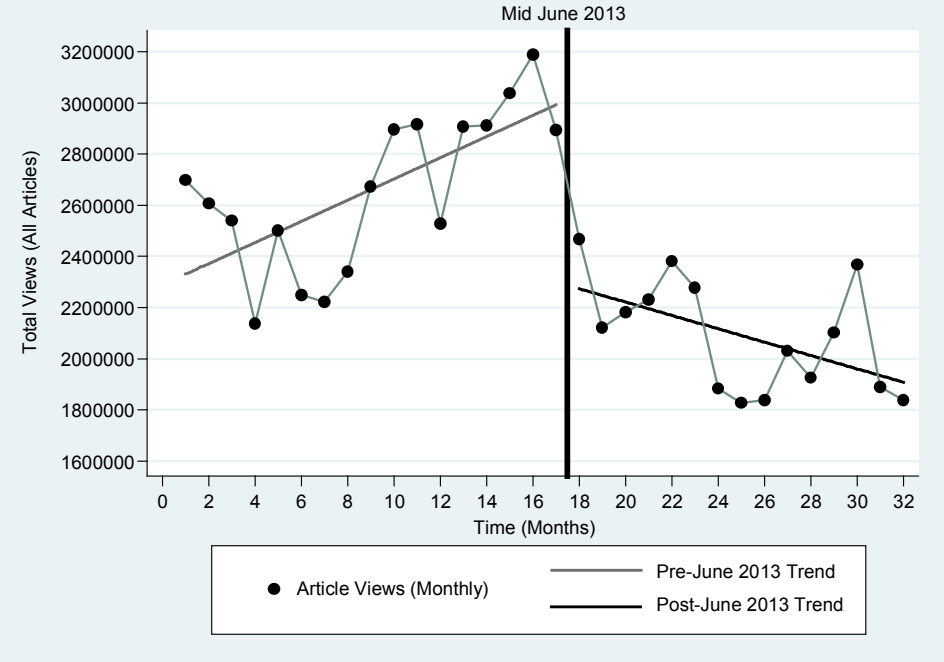
to-noise ratio, that is, variability of page views over the course of a study period due to a range of daily factors.[35]

(b) Second, the data ***was*** also analyzed for seasonal trends as well for real-world events to determine whether any such events would have an outsized effect on page views. This is precisely how I identified the "Hamas" article as an outlier and related it to real world events. This was not done by cherry-picking or manipulation, but through standard regression model diagnostics, as well as best practices for identifying and dealing with outlier observations.[36]

(c) Third, no seasonal or "real world" event-related variations identified by Salzberg explain the actual trends apparent in the aggregated data before and after June 2013. Salzberg points to a seasonal "trough" in the summer of 2012 and a "peak" due to the Boston Marathon bombing in April 2013 for page views in the 47 Terrorism articles. These are visible in **Figure 5** below:

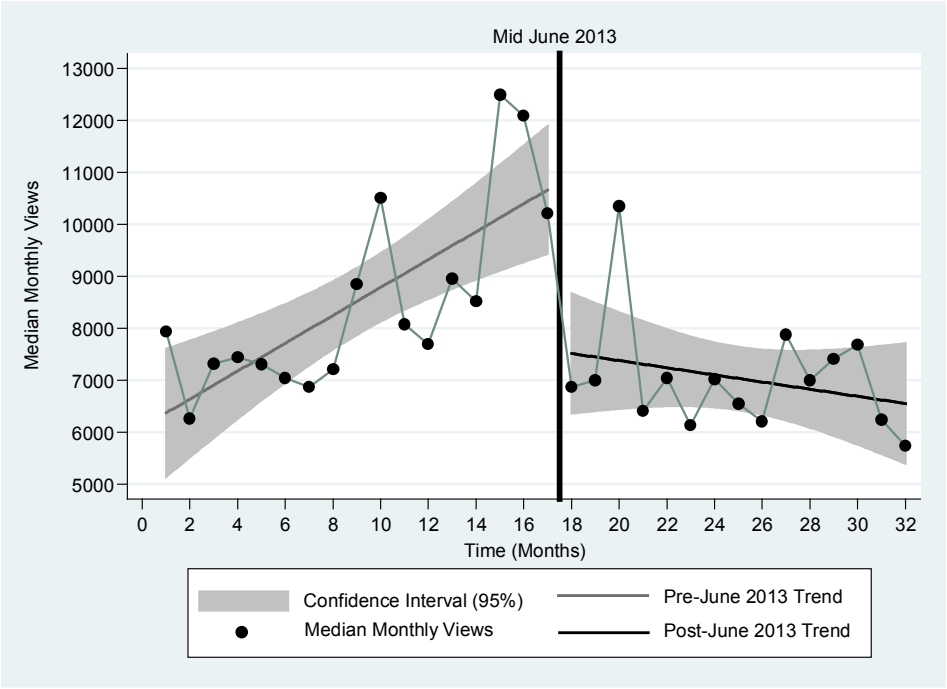
[35] See works cited at *supra* note 23 and accompanying text.
[36] The Hamas article was identified using standard regression model diagnostics, including examining cook's distance, leverage, residual values, z-scoes, among others. Handling of the influential outlier was also done according to best practices, with the outlier removed but results before and after removal conveyed. See AGUINAS ET AL., *supra* note 8, at 20-23.

Figure 5: Aggregated Total Monthly Page Views for 48 Terror Articles (Without Hamas)



However, the added trend analysis of page views in **Figure 5** demonstrates that despite these possibly seasonal and "real world" event-related variations over the course of the 17 months prior to June 2013 and 14 mounts after that period, there is no variation or "trough" comparable to the one that occurs between the total page views as of the beginning of June 2013 (2,893,553 page views) and that at the end of July 2013 (2,121,335). Indeed, during time period there was a decline of 772, 218 page views, or roughly 27%. By contrast, the decline between April and May 2013 that Salzberg highlights (294, 820 page views) is slightly greater than 9%. Nor is there any point in the entire 32-month study period where there are consistently fewer monthly page views than in the months of December 2013 through August 2014. These points are consistent with a June 2013 chilling effects hypothesis. These points are even clearer when examining the median monthly page views for the 23 Most Privacy Sensitive Articles visualized in **Figure 6:**

Figure 6: Aggregated Median Monthly Page Views for 23 Most Privacy-Sensitive Articles



(d) Here, again, there is a significant drop from the median monthly page views as of the beginning of June 2013 (12,090) and the end of July 2018 (6,864) totaling 5,226, which is slightly greater than a 45% drop in median monthly views. Nor, despite some variation, is there any other period in the 32 months where monthly median page views are trending so clearly lower over time as the months after June 2013 onto August 2014. Again, this is consistent with a chilling effects hypothesis. Salzberg offers no alternative explanation beyond identifying "peaks" and "troughs" in 2012 and 2013, which do not account or explain these findings.

(e) Fourth, in a naturalistic study outside the experimental context, it is not possible to control for all confounding factors, like the impact of all real world events on page views over 32 months. However, the ITS design was chosen for the very fact that its pre/post design *can* help control for other explanatory factors as any such known or unknown confounding variables would be present in both the pre and post measurements (monthly page views), thus any

changes after the intervention (June 2013) can be attributed to intervention itself. This dimension of the ITS design can be further strengthened by adding one or more comparators,[37] which was also done in this study. Comparators help control for confounding factors and seasonality.[38] Here, the comparator groups—security, infrastructure, popular, and Wikipedia English home-page traffic—not expected to be impacted or affected by the ITS design intervention (here, surveillance chilling effects in June 2013) can be compared to the study group (page views for privacy-sensitive terrorism related Wikipedia articles). As my results showed, the privacy-sensitive terrorism Wikipedia were impacted (statistically significant drop in June 2013 and trend reversal after that month) while none of the comparator articles showed the same effects. This is consistent with a chilling effect hypothesis.

(f) Salzberg claims that my comparator groups do not corroborate my findings as they are not "proper" controls groups that "exhibit the trend" shown by the terrorism articles before June 2013.[39] Of course, comparator groups identical to the study group are ideal but are often not feasible.[40] In fact, Campbell and Stanley, in their leading 1966 text *Experimental and Quasi-Experiment Designs for Research* that Salzberg cites, recommend that for an ITS study of a "major administrative change" that a researcher use a "similar institution" as a comparator not expected to undergo the change and upon which the same ITS design can be tested.[41] Wagner et al.'s (2002) leading article on ITS design recommends that where an identical control group is not

[37] R. BARKER BAUSELL, *THE DESIGN AND CONDUCT OF MEANINGFUL EXPERIMENTS INVOLVING HUMAN PARTICIPANTS 199* (OUP, 2015) (noting that an ITS design can be "significantly buttressed" by adding one or more comparator group); MARY A. M. ROGERS, *COMPARATIVE EFFECTIVENESS RESEARCH 94* (OUP, 2014).
[38] BAUSELL, *id.* at 200; Wagner et al., *supra* note 2, at 306.
[39] Salzberg Decl. at ¶¶ 33-46.
[40] BAUSELL, *id.* at 199; Wagner et al., *supra* note 2, at 306.
[41] DONALD CAMPBELL & JULIAN STANLEY, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH 55* (Houghton-Mifflin, 1966).

possible, a study may examine a "different but related group . . . not expected to change following the intervention, in the same group of subjects."[42] This is precisely the methodology that I adopted. Since identical or randomly sampled control groups were not possible, I used comparator groups drawn from a "different but related group" of Wikipedia articles drawn from a "different but related" group of DHS key words (security and infrastructure), as well as a set of popular Wikipedia articles. Given that security, infrastructure, and popular Wikipedia articles do not raise privacy concerns, they would not be expected to change after a surveillance chilling effect intervention in June 2013. We would also expect similar viewer audiences for these articles (e.g. someone with an interest in national security would be just as likely to view Wikipedia articles on terrorism as domestic or infrastructure security articles, the comparators in this study). My results showed that they were not impacted, while the privacy-sensitive terrorism-related articles were. This, as noted, was consistent with a surveillance chilling effect hypothesis.

31. **Salzberg's fourth critique:** The model did not take into account that the 48 terror articles chosen based on 2011 DHS list would naturally rise and decline in interest over time. In other words, the 2011 terrorism-related key words would "undoubtedly become stale over time." The same is true for the trends in the comparator article groups. (Salzberg Decl. ¶¶ 61–64.)

32. **My Response:**

(a) The articles in my study were chosen based on keywords associated with "terrorism" that DHS uses to track and monitor social media. Since the media coverage relating to the Snowden revelations framed the issue of NSA surveillance as a matter of national security and terrorism threats, it was logical for me to use DHS keywords associated with "terrorism" to create the Wikipedia articles that represent the sort of articles that users may be chilled from accessing

[42] WAGNER ET AL., *supra* note 2, at 306.

in light of privacy concerns about government surveillance. I used this approach for pragmatic methodological reasons, as there is no pre-determined list reflecting all privacy-sensitive Wikipedia articles from which to draw a random sample. Furthermore, cherry-picking a list of articles relating to certain sensitive topics (like Syria or ISIL, as Salzberg suggests) would be subject to serious selection biases. Using government keyword lists to study government surveillance is not a novel or unprecedented approach;[43] indeed, my methodology is similar to the methodology of an earlier, peer-reviewed study exploring the chilling effects associated with the NSA surveillance in Google search data, which also used these DHS keyword lists.[44]

(b) The fact that some likely privacy-sensitive articles (like Syria and ISIL) that were not included in the DHS keyword list may have recorded higher page views for some period of time during my study does not in any way undermine the overall conclusions for the privacy-sensitive Wikipedia articles examined over the entire 32-month period that I studied.

(c) Salzberg's Declaration offers no evidence for his claim that the "many of the 2011 terrorism-related keywords undoubtedly became stale over time" and thus "page views dropped."

33. **Salzberg's fifth critique:** The page view data examined only extends for 32 months through August 2014, therefore the "results do not and cannot imply that an effect of the June 2013 disclosures persists today, or did so even in 2015." (Salzberg Decl. ¶ 65.)

---

[43] JEDIDIAH R. CRANDALL & MASASHI CRETE-NISHIHATA ET AL., *CHAT PROGRAM CENSORSHIP AND SURVEILLANCE IN CHINA: TRACKING TOM-SKYPE AND SINA UC*, First Monday, July 1, 2013, http://firstmonday.org/ojs/index.php/fm/article/view/4628/3727 [https://perma.cc/M5FJ-T4D5]; JEFFREY KNOCKEL, JEDIDAH CRANDALL & JARED SAIA, *THREE RESEARCHERS, FIVE CONJECTURES: AN EMPIRICAL ANALYSIS OF TOM-SKYPE CENSORSHIP AND SURVEILLANCE*, 16:4 FOCI '11: USENIX Workshop on Free & Open Comm. on Internet (2011), https://www.cs.unm.edu/~crandall/foci11knockel.pdf [https://perma.cc/FH8H-JUBA].

[44] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR*, *IN* CAMBRIDGE UNIVERSITY HANDBOOK ON SURVEILLANCE LAW (David Gray et al. eds., 2017).

34. **My Response:**

(a) The statistically significant trend reversal from increasing monthly views prior to June 2013 to a downward trend, with a monthly reduction in page views afterwards, is indicative of a lasting chilling effect. This is supported by other research on long-term online chilling effects due to public awareness of surveillance.

(b) First, an Massachusetts Institute of Technology (MIT) study on Google search data later published a peer reviewed chapter in the *Cambridge University Handbook on Surveillance Law*, found a statistically significant reductions in privacy-sensitive Google searches after the June 2013 Snowden disclosures about NSA surveillance.[45] The findings, the authors concluded, provided "substantial empirical documentation of a chilling effect," both in the "shorter term" and "in the longer term", that "appeared to be related to increased awareness of government surveillance online."[46]

(c) Second, a 2017 peer reviewed study on Wikipedia editors found evidence, based on qualitative interviews in the spring and summer of 2015, that editors were chilled from certain activities on Wikipedia due to awareness of government surveillance.[47] For example, one Wikipedia editor stated, "for the Edward Snowden page, I have pulled myself away from adding sensitive contributions, like different references, because I thought the name may be traced back to me in some way." The fact that some Wikipedia users have avoided sensitive or controversial topics in 2015, two years after the Snowden revelations in 2013, is consistent with my findings.

---

[45] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[46] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[47] ANDREA FORTE, NAZANIN ANDALIBI, AND RACHEL GREENSTADT, *PRIVACY, ANONYMITY, AND PERCEIVED RISK IN OPEN COLLABORATION: A STUDY OF TOR USERS AND WIKIPEDIANS*, *in* CSCW 1800 (2017).

(d) Third, a recent 2018 study exploring how journalists have been impacted by "potential surveillance by government," which involved qualitative interviews with American journalists in 2015, found all seven journalists in the study indicated that "their work and lives have changed under a real or perceived threat of mass government surveillance."[48] The author concluded that "participants reported an increased awareness of mass government surveillance" and "[i]n every case, they reported adjusting their behavior to some degree."[49] This is also consistent with my findings.

(e) Fourth, a Pew Research Center survey of 475 adult Americans conducted between November 26, 2014 and January 3, 2015 found that, among the 87% of respondents aware of "government surveillance programs" due to the Snowden revelations, 34% had taken "at least one step to hide or shield their information from the government," including avoiding using "certain terms in online communications."[50] The survey also found 25% changed the patterns of their own use of various online platforms "a great deal" or "somewhat" since the Snowden revelations. These findings from a survey administered in late 2014 and early 2015 are also consistent with my conclusions.

(f) Fifth, a PEN America survey of 520 American writers in October 2013[51] found that 28% of the writers surveyed had "curtailed or avoided" certain online activities due to "fear of surveillance" and another 12% "seriously considered" doing so; 24% "deliberately avoided certain topics in phone or email conversations," and another 9% have "seriously

---

[48] STEPHENSON WATERS, *THE EFFECTS OF MASS SURVEILLANCE ON JOURNALISTS' RELATIONS WITH CONFIDENTIAL SOURCES, DIGITAL JOURNALISM,* 6:10 Digital Journalism 1294, 1310 (2018).
[49] *Id*. at 1310.
[50] LEE RAINIE ET AL., *PEW RES. INTERNET PROJECT*, Americans' Privacy Strategies Post-Snowden 4 (Mar. 16, 2015), http://www.pewinternet.org/files/2015/03/PI_AmericansPrivacyStrategies_0316151.pdf.
[51] FDR GROUP & PEN, AMERICAN CENTER, *CHILLING EFFECTS: NSA SURVEILLANCE DRIVES U.S. WRITERS TO SELF-CENSOR 3–4* (2013), http://www.pen.org/sites/default/files/Chilling%20Effects_PEN%20American.pdf.

considered it"; and 16% have refrained from "conducting Internet searches or visiting websites on topics that may be considered controversial or suspicious," and another 12% have "seriously considered it." These results are consistent with my conclusions as to a long term chilling effect.

(g) Also, as explained above, Salzberg's extended comparison analysis that relies on more recent page view data is fundamentally invalid because it compares across page view definitions—the more recent data includes "mobile" page views, while my study relied on "non-mobile" data.[52]

35. **Salzberg's sixth critique:** Penney's model fails to isolate the "particular effect of public 'awareness' about the NSA Upstream program" from potential other effects of the Snowden disclosures, including increased awareness about other NSA surveillance activities. (Salzberg Decl. ¶ 66.)

36. **My Response:**

(a) In any study of naturalistic changes in human behavior, it will not be possible to isolate the source of all causes and effects on behavior. It is enough for purposes of establishing whether Upstream likely had a chilling effect on Wikipedia users that the reporting on NSA surveillance in June 2013 included multiple references to international Internet communications monitoring, and that general public awareness of NSA surveillance grew due to media coverage after June 2013.[53] As earlier noted, a Pew Research Center survey of adult Americans conducted between November 26, 2014 and January 3, 2015 found 87% of respondents were aware of NSA surveillance programs due to the Snowden revelations.[54]

(b) Furthermore, in-line with the empirical conclusions of my Declaration,

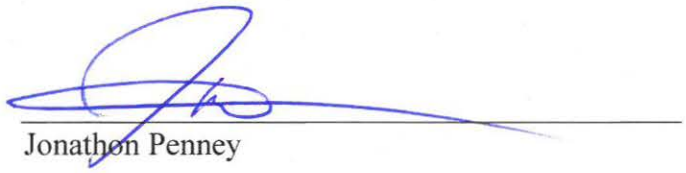
[52] *See supra* ¶ 23.
[53] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[54] *See supra* ¶ 30(e).

Wikimedia has introduced other evidence establishing the particular chilling effect that awareness of Upstream surveillance had on Wikimedia's readers and contributors. The Declarations of Michelle Paulson and James Alexander describe the chilling effect that Upstream surveillance had on the Wikimedia community at large, particularly among users abroad who engage with the platform concerning privacy-sensitive topics. (*See* Pl.'s Ex. 3, ¶¶ 41, 45, 46; Pl.'s Ex. 4, ¶¶ 4-11.) The Declaration of Emily Temple-Wood, an active Wikimedia community member, further describes first-hand the chilling effect that awareness of Upstream surveillance has had among the community of readers and contributors. (*See* Pl.'s Ex. 6, ¶¶ 20-21.)

(c) Salzberg's critique that the "particular effect" of Upstream cannot be entirely isolated is not actually a methodological critique, but rather, a general observation about a naturalistic studies. However, courts have rejected such challenges when ruling on *Daubert* motions. *See A Woman's Choice-East Side Women's Clinic v. Newman,* 904 F. Supp. 1434 (S.D. Ind. 1995) (upholding naturalistic study against *Daubert* challenge when ruling on preliminary injunction motion).

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct to the best of my knowledge and belief.

Executed on March 8, 2019 in Halifax, Canada.



Jonathon Penney

APPENDIX

23 Most Privacy Sensitive Article Set Cross Validation Analysis

RAW total monthly views

```
.
. reg HP23RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.82
       Model |  3.1487e+11     3  1.0496e+11           Prob > F      =  0.0000
    Residual |  1.8571e+11    28  6.6324e+09           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   81439

------------------------------------------------------------------------------
HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   21383.58   4031.855     5.30   0.000      13124.7    29642.46
intervention |  -224931.2   58212.11    -3.86   0.001    -344173.3   -105689.1
   postslope |  -29367.59   6320.044    -4.65   0.000    -42313.61   -16421.56
       _cons |   349787.6   41314.22     8.47   0.000     265159.3      434416
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   89506.354
Mean Absolute Errors     |   63503.274
Pseudo-R2                |      0.49622
----------------------------------------

. reg HP23RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |  2.8498e+11     3  9.4993e+10           Prob > F      =  0.0000
    Residual |  2.1560e+11    28  7.6998e+09           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   87749

------------------------------------------------------------------------------
   HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |    22020.4   4758.837     4.63   0.000     12272.36    31768.43
interventionMAY |  -152067.1   62229.88    -2.44   0.021    -279539.2   -24594.96
   postslopeMAY |  -34404.22   6730.013    -5.11   0.000    -48190.03   -20618.42
          _cons |   345966.8   46015.77     7.52   0.000     251707.7    440225.8
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   100292.59
Mean Absolute Errors     |   71401.914
Pseudo-R2                |      0.37718
----------------------------------------
```

```
. reg HP23RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |  2.4087e+11     3  8.0289e+10           Prob > F      =  0.0003
    Residual |  2.5971e+11    28  9.2752e+09           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96308

------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |    13623.8   5755.506     2.37   0.025     1834.181    25413.42
interventionAPRIL |   65634.32   68033.27     0.96   0.343    -73725.51    204994.1
   postslopeAPRIL |   -34751.7   7473.906    -4.65   0.000    -50061.31    -19442.1
            _cons |   393547.5   52329.76     7.52   0.000     286354.8    500740.1
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   108143.33
Mean Absolute Errors     |   69579.447
Pseudo-R2                |     0.28104
----------------------------------------

. reg HP23RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |  2.3801e+11     3  7.9335e+10           Prob > F      =  0.0004
    Residual |  2.6257e+11    28  9.3775e+09           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96837

------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   13450.41   6420.254     2.09   0.045     299.1154     26601.7
interventionMARCH |   88472.13   68403.98     1.29   0.206    -51647.07    228591.3
   postslopeMARCH |  -33519.11   7782.969    -4.31   0.000     -49461.8   -17576.42
            _cons |   394472.2   54666.49     7.22   0.000       282493    506451.4
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   105572.3
Mean Absolute Errors     |   69332.589
Pseudo-R2                |     0.30842
----------------------------------------
```

Average monthly views

```
.
. reg HP23AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.83
       Model |   595220203     3   198406734           Prob > F      =  0.0000
    Residual |   351044760    28  12537312.9           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |   946264964    31  30524676.2           Root MSE      =  3540.8

------------------------------------------------------------------------------
HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   929.7206   175.2961     5.30   0.000     570.6429    1288.798
intervention |  -9779.688   2530.932    -3.86   0.001    -14964.07   -4595.308
   postslope |  -1276.853   274.7814    -4.65   0.000    -1839.717   -713.9885
       _cons |   15208.16    1796.25     8.47   0.000     11528.71    18887.61
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3891.5395
Mean Absolute Errors     |   2760.9415
Pseudo-R2                |     0.49623
----------------------------------------

. reg HP23AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |   538718524     3   179572841           Prob > F      =  0.0000
    Residual |   407546439    28    14555230           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |   946264964    31  30524676.2           Root MSE      =  3815.1

------------------------------------------------------------------------------
   HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   957.4059   206.9046     4.63   0.000     533.5811    1381.231
interventionMAY |  -6611.619   2705.628    -2.44   0.021    -12153.85   -1069.391
   postslopeMAY |  -1495.838   292.6072    -5.11   0.000    -2095.217   -896.4595
          _cons |   15042.05   2000.672     7.52   0.000     10943.86    19140.24
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4360.523
Mean Absolute Errors     |   3104.3749
Pseudo-R2                |     0.37718
----------------------------------------

. reg HP23AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |   455331527     3   151777176           Prob > F      =  0.0003
    Residual |   490933436    28    17533337           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
```

```
       Total |   946264964    31  30524676.2           Root MSE      =  4187.3

------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   592.3357    250.238     2.37   0.025     79.74639    1104.925
interventionAPRIL |   2853.676   2957.952     0.96   0.343    -3205.414    8912.766
   postslopeAPRIL |  -1510.946   324.9507    -4.65   0.000    -2176.577   -845.3148
            _cons |   17110.78   2275.194     7.52   0.000     12450.26    21771.31
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4701.8614
Mean Absolute Errors     |   3025.1117
Pseudo-R2                |     0.28105
----------------------------------------

. reg HP23AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |   449922646     3   149974215           Prob > F      =  0.0004
    Residual |   496342317    28  17726511.3           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |   946264964    31  30524676.2           Root MSE      =  4210.3

------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   584.7802   279.1393     2.09   0.045     12.98935    1156.571
interventionMARCH |   3846.882   2974.063     1.29   0.206    -2245.209    9938.974
   postslopeMARCH |  -1457.347   338.3873    -4.31   0.000    -2150.502   -764.1918
            _cons |   17151.08   2376.785     7.22   0.000     12282.45    22019.7
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4590.0594
Mean Absolute Errors     |   3014.3548
Pseudo-R2                |     0.30843
----------------------------------------
```

Median monthly views

```
. reg HP23Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   11.65
       Model |  48268345.7     3  16089448.6           Prob > F      =  0.0000
    Residual |  38654921.8    28  1380532.92           R-squared     =  0.5553
-------------+------------------------------           Adj R-squared =  0.5077
       Total |  86923267.5    31  2803976.37           Root MSE      =    1175

------------------------------------------------------------------------------
  HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   268.5098   58.16923     4.62   0.000     149.3555    387.6641
intervention |  -3080.506   839.8499    -3.67   0.001     -4800.86   -1360.151
   postslope |  -337.3348   91.18187    -3.70   0.001    -524.1124   -150.5572
       _cons |   6097.706   596.0572    10.23   0.000     4876.738    7318.674
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1273.658
Mean Absolute Errors     |   1017.3205
Pseudo-R2                |      0.41399
----------------------------------------

. reg HP23Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.23
       Model |  43222658.1     3  14407552.7           Prob > F      =  0.0002
    Residual |  43700609.4    28  1560736.05           R-squared     =  0.4973
-------------+------------------------------           Adj R-squared =  0.4434
       Total |  86923267.5    31  2803976.37           Root MSE      =  1249.3

------------------------------------------------------------------------------
     HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   279.9441   67.75249     4.13   0.000     141.1594    418.7288
interventionMAY |  -2201.731   885.9788    -2.49   0.019    -4016.576   -386.8856
   postslopeMAY |  -406.6324   95.81649    -4.24   0.000    -602.9035   -210.3612
          _cons |     6029.1   655.1354     9.20   0.000     4687.116    7371.084
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1373.4931
Mean Absolute Errors     |   1147.5357
Pseudo-R2                |      0.32413
----------------------------------------

. reg HP23Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.36
       Model |  35235938.7     3  11745312.9           Prob > F      =  0.0020
    Residual |  51687328.8    28  1845976.03           R-squared     =  0.4054
-------------+------------------------------           Adj R-squared =  0.3417
```

```
       Total |  86923267.5    31  2803976.37           Root MSE      =  1358.7

------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |      234.75   81.19588     2.89   0.007     68.42777    401.0722
interventionAPRIL |  -483.0456   959.7803    -0.50   0.619    -2449.066    1482.975
   postslopeAPRIL |   -435.625   105.4382    -4.13   0.000    -651.6054   -219.6446
            _cons |      6285.2   738.2428     8.51   0.000     4772.978    7797.422
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1508.2152
Mean Absolute Errors     |   1224.239
Pseudo-R2                |      0.20017
-----------------------------------------

. reg HP23Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.84
       Model |  36767609.2     3  12255869.7           Prob > F      =  0.0013
    Residual |  50155658.3    28  1791273.51           R-squared     =  0.4230
-------------+------------------------------           Adj R-squared =  0.3612
       Total |  86923267.5    31  2803976.37           Root MSE      =  1338.4

------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   146.1165   88.73404     1.65   0.111    -35.64696    327.8799
interventionMARCH |   1425.536   945.4084     1.51   0.143    -511.0457    3362.117
   postslopeMARCH |  -402.6335   107.5681    -3.74   0.001    -622.9767   -182.2903
            _cons |   6757.912   755.5431     8.94   0.000     5210.252    8305.572
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   1433.4557
Mean Absolute Errors     |   1161.2393
Pseudo-R2                |      0.26252
-----------------------------------------
.
```

21 article set (23 set minus Ammonium Nitrate and Jihad articles)

RAW total monthly views

```
. reg HP21RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |  1.5222e+11     3  5.0742e+10           Prob > F      =  0.0000
    Residual |  6.9713e+10    28  2.4898e+09           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   49898

------------------------------------------------------------------------------
HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   14449.77   2470.296     5.85   0.000     9389.599    19509.95
intervention |    -134792   35666.25    -3.78   0.001      -207851   -61732.97
   postslope |  -22040.86   3872.258    -5.69   0.000    -29972.82    -14108.9
       _cons |   284385.7   25313.01    11.23   0.000     232534.4      336237
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   54117.269
Mean Absolute Errors     |   43122.186
Pseudo-R2                |      0.58202
----------------------------------------

. reg HP21RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |  1.3902e+11     3  4.6342e+10           Prob > F      =  0.0000
    Residual |  8.2913e+10    28  2.9612e+09           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   54417

------------------------------------------------------------------------------
   HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   14548.22   2951.169     4.93   0.000     8503.022    20593.41
interventionMAY |  -81022.03   38591.54    -2.10   0.045    -160073.2   -1970.848
   postslopeMAY |   -25025.8   4173.583    -6.00   0.000    -33574.99    -16476.6
          _cons |     283795   28536.44     9.95   0.000     225340.8    342249.3
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   60680.401
Mean Absolute Errors     |   47288.848
Pseudo-R2                |      0.47884
```

```
----------------------------------------

. reg HP21RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |  1.2601e+11     3  4.2004e+10           Prob > F      =  0.0000
    Residual |  9.5927e+10    28  3.4260e+09           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58532

------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   10747.09   3497.949     3.07   0.005     3581.866    17912.31
interventionAPRIL |   29848.38   41347.69     0.72   0.476    -54848.53    114545.3
   postslopeAPRIL |  -25421.95   4542.319    -5.60   0.000    -34726.47   -16117.43
            _cons |   305334.8   31803.78     9.60   0.000     240187.7    370481.8
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   64314.986
Mean Absolute Errors     |   43954.927
Pseudo-R2                |     0.41567
----------------------------------------

. reg HP21RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |  1.2575e+11     3  4.1916e+10           Prob > F      =  0.0000
    Residual |  9.6190e+10    28  3.4354e+09           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58612

------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   10464.43   3885.943     2.69   0.012     2504.441    18424.43
interventionMARCH |   52992.36    41402.4     1.28   0.211    -31816.62    137801.3
   postslopeMARCH |  -24766.06   4710.744    -5.26   0.000    -34415.58   -15116.54
            _cons |   306842.3   33087.61     9.27   0.000     239065.4    374619.1
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   63048.122
Mean Absolute Errors     |   43494.51
Pseudo-R2                |     0.43631
----------------------------------------
```

Average monthly views

```
. reg HP21AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |   345191835     3   115063945           Prob > F      =  0.0000
    Residual |   158083216    28  5645829.13           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |   503275051    31  16234679.1           Root MSE      =  2376.1

------------------------------------------------------------------------------
HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   688.1005   117.6342     5.85   0.000     447.1377    929.0633
intervention |   -6418.64   1698.409    -3.78   0.001    -9897.673   -2939.608
   postslope |  -1049.597   184.3949    -5.69   0.000    -1427.313    -671.881
       _cons |   13542.04   1205.392    11.23   0.000      11072.9    16011.17
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   2577.0343
Mean Absolute Errors     |   2053.4405
Pseudo-R2                |     0.58203
----------------------------------------

. reg HP21AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |   315262322     3   105087441           Prob > F      =  0.0000
    Residual |   188012729    28  6714740.31           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |   503275051    31  16234679.1           Root MSE      =  2591.3

------------------------------------------------------------------------------
   HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-----------------------------------------------------------------
           time |   692.7941    140.532     4.93   0.000     404.9274    980.6608
interventionMAY |  -3858.256   1837.694    -2.10   0.045    -7622.601   -93.91028
   postslopeMAY |  -1191.737   198.7422    -6.00   0.000    -1598.842   -784.6318
          _cons |   13513.87    1358.88     9.94   0.000     10730.34    16297.41
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   2889.5437
Mean Absolute Errors     |   2251.8857
Pseudo-R2                |     0.47886
----------------------------------------
```

```
. reg HP21AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |   285751156     3  95250385.3           Prob > F      =  0.0000
    Residual |   217523895    28  7768710.54           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |   503275051    31  16234679.1           Root MSE      =  2787.2

------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   511.7821   166.5695     3.07   0.005       170.58    852.9843
interventionAPRIL |   1421.438   1968.944     0.72   0.476     -2611.76    5454.636
   postslopeAPRIL |  -1210.601   216.3015    -5.60   0.000    -1653.674   -767.5272
            _cons |   14539.61    1514.47     9.60   0.000     11437.36    17641.86
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3062.6289
Mean Absolute Errors     |   2093.0978
Pseudo-R2                |     0.41568
----------------------------------------

. reg HP21AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |   285155599     3  95051866.3           Prob > F      =  0.0000
    Residual |   218119452    28  7789980.44           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |   503275051    31  16234679.1           Root MSE      =  2791.1

------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   498.3099   185.0451     2.69   0.012     119.2623    877.3575
interventionMARCH |   2523.736   1971.545     1.28   0.211    -1514.791    6562.263
   postslopeMARCH |   -1179.36   224.3213    -5.26   0.000    -1638.862    -719.859
            _cons |   14611.46   1575.602     9.27   0.000     11383.99    17838.94
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3002.2907
Mean Absolute Errors     |   2071.122
Pseudo-R2                |     0.43632
----------------------------------------
```

Median monthly views

```
. . reg HP21Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    5.51
       Model |  20811047.6     3  6937015.87           Prob > F      =  0.0042
    Residual |  35281531.9    28  1260054.71           R-squared     =  0.3710
-------------+------------------------------           Adj R-squared =  0.3036
       Total |  56092579.5    31  1809438.05           Root MSE      =  1122.5

------------------------------------------------------------------------------
  HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   206.3873    55.5731     3.71   0.001     92.55092    320.2236
intervention |  -1514.518   802.3669    -1.89   0.069    -3158.092    129.0565
   postslope |  -282.8123   87.11237    -3.25   0.003    -461.2539   -104.3707
       _cons |   5062.868   569.4548     8.89   0.000     3896.392    6229.343
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1218.2317
Mean Absolute Errors     |   886.83799
Pseudo-R2                |     0.18535
----------------------------------------

. reg HP21Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.20
       Model |  17413400.8     3  5804466.93           Prob > F      =  0.0142
    Residual |  38679178.7    28  1381399.24           R-squared     =  0.3104
-------------+------------------------------           Adj R-squared =  0.2366
       Total |  56092579.5    31  1809438.05           Root MSE      =  1175.3

------------------------------------------------------------------------------
     HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   185.0985   63.74119     2.90   0.007     54.53063    315.6664
interventionMAY |  -427.3015   833.5242    -0.51   0.612    -2134.698    1280.095
   postslopeMAY |  -313.7162   90.14365    -3.48   0.002    -498.3671   -129.0653
          _cons |     5190.6   616.3479     8.42   0.000     3928.068    6453.132
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1293.5702
Mean Absolute Errors     |   965.87467
Pseudo-R2                |     0.10277
----------------------------------------

. reg HP21Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.84
       Model |  19161806.3     3  6387268.76           Prob > F      =  0.0077
    Residual |  36930773.2    28  1318956.19           R-squared     =  0.3416
-------------+------------------------------           Adj R-squared =  0.2711
```

```
       Total |  56092579.5    31  1809438.05           Root MSE      =  1148.5

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |     108.275    68.6335     1.58   0.126    -32.31436    248.8644
interventionAPRIL |    1336.64    811.286     1.65   0.111    -325.2039    2998.484
   postslopeAPRIL |  -297.9931   89.12516    -3.34   0.002    -480.5577   -115.4285
            _cons |   5625.933   624.0241     9.02   0.000     4347.678    6904.189
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1233.6708
Mean Absolute Errors     |   996.36376
Pseudo-R2                |      0.16269
----------------------------------------

. reg HP21Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    3.97
       Model |  16749983.3     3  5583327.77           Prob > F      =  0.0177
    Residual |  39342596.2    28  1405092.72           R-squared     =  0.2986
-------------+------------------------------           Adj R-squared =  0.2235
       Total |  56092579.5    31  1809438.05           Root MSE      =  1185.4

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   125.9802   78.58901     1.60   0.120    -35.00206    286.9625
interventionMARCH |   1064.811   837.3191     1.27   0.214    -650.3597    2779.981
   postslopeMARCH |  -282.8265   95.26972    -2.97   0.006    -477.9776   -87.67527
            _cons |   5531.505   669.1613     8.27   0.000     4160.791     6902.22
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1264.4955
Mean Absolute Errors     |   1001.816
Pseudo-R2                |      0.12727
----------------------------------------
```

46 Article Group Cross Validation (48 minus Hamas and Fundamentalism articles)

RAW TOTAL MONTHLY VIEWS

```
reg T46NoFundaorHamasRaw time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  3.2537e+12     3  1.0846e+12           Prob > F      =  0.0000
    Residual |  1.2496e+12    28  4.4628e+10           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.1e+05

------------------------------------------------------------------------------
T46NoFunda~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   37645.21   10458.63     3.60   0.001     16221.69    59068.73
intervention |  -683829.1   151002.1    -4.53   0.000    -993142.9   -374515.3
   postslope |  -60274.25   16394.18    -3.68   0.001    -93856.21   -26692.28
       _cons |    2261895     107169    21.11   0.000      2042370     2481421
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   226571.35
Mean Absolute Errors     |   189268.95
Pseudo-R2                |     0.63861
----------------------------------------



. reg T46NoFundaorHamasRaw time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  2.9994e+12     3  9.9980e+11           Prob > F      =  0.0000
    Residual |  1.5039e+12    28  5.3711e+10           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T46NoFundaorH~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   40060.48   12568.78     3.19   0.004      14314.5    65806.46
interventionMAY |    -500903   164358.1    -3.05   0.005    -837575.4   -164230.7
   postslopeMAY |  -75642.86   17774.94    -4.26   0.000    -112053.2   -39232.55
          _cons |    2247404   121534.3    18.49   0.000      1998452     2496355
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   253115.47
Mean Absolute Errors     |   200434.46
Pseudo-R2                |     0.55147
----------------------------------------
```

```
. reg T46NoFundaorHamasRaw time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  2.6623e+12     3  8.8744e+11           Prob > F      =  0.0000
    Residual |  1.8410e+12    28  6.5749e+10           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |       34170   15323.79     2.23   0.034     2780.649    65559.36
interventionAPRIL |    -203913   181135.6    -1.13   0.270    -574952.6    167126.5
   postslopeAPRIL |   -83616.5   19898.95    -4.20   0.000    -124377.7   -42855.34
            _cons |    2280783   139325.7    16.37   0.000      1995387     2566179
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   280248.05
Mean Absolute Errors     |   226329.8
Pseudo-R2                |      0.45312
----------------------------------------

. reg T46NoFundaorHamasRaw time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.67
       Model |  2.5927e+12     3  8.6423e+11           Prob > F      =  0.0000
    Residual |  1.9106e+12    28  6.8236e+10           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   29286.11   17318.75     1.69   0.102    -6189.741    64761.95
interventionMARCH |   265.6691   184520.9     0.00   0.999    -377708.4    378239.7
   postslopeMARCH |  -84909.06    20994.7    -4.04   0.000    -127914.7   -41903.37
            _cons |    2306830   147463.8    15.64   0.000      2004765     2608896
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   281781.73
Mean Absolute Errors     |   236816.42
Pseudo-R2                |      0.44585
----------------------------------------
```

AVERAGE TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  1.5377e+09     3   512561551           Prob > F      =  0.0000
    Residual |   590556612    28  21091307.6           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
```

```
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  4592.5

------------------------------------------------------------------------------
T46NoFundH~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   818.3775    227.364     3.60   0.001     352.6434    1284.111
intervention |  -14866.27   3282.691    -4.53   0.000    -21590.56   -8141.982
   postslope |  -1310.277   356.3993    -3.68   0.001    -2040.328   -580.2265
       _cons |   49171.72   2329.788    21.11   0.000     44399.37    53944.07
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4925.5268
Mean Absolute Errors     |   4114.587
Pseudo-R2                |      0.63861
----------------------------------------

. reg T46NoFundHamasAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  1.4175e+09     3   472494167           Prob > F      =  0.0000
    Residual |   710758766    28  25384241.6           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5038.3

------------------------------------------------------------------------------
T46NoFundHama~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   870.8809    273.239     3.19   0.004     311.1762    1430.586
interventionMAY |  -10889.49   3573.063    -3.05   0.005    -18208.58   -3570.406
   postslopeMAY |  -1644.387   386.4183    -4.26   0.000    -2435.929   -852.8448
          _cons |    48856.7   2642.095    18.49   0.000     43444.61    54268.79
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5502.6127
Mean Absolute Errors     |   4357.318
Pseudo-R2                |      0.55146
----------------------------------------



. reg T46NoFundHamasAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  1.2582e+09     3   419393074           Prob > F      =  0.0000
    Residual |   870062045    28  31073644.5           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5574.4

------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   742.8179   333.1326     2.23   0.034     60.42674    1425.209
interventionAPRIL |  -4432.982   3937.811    -1.13   0.270    -12499.22    3633.259
   postslopeAPRIL |   -1817.73   432.5947    -4.20   0.000     -2703.86   -931.5995
```

```
       _cons |   49582.39   3028.882    16.37   0.000     43378.01    55786.77
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   6092.4748
Mean Absolute Errors     |   4920.2718
Pseudo-R2                |     0.45311
----------------------------------------


. reg T46NoFundHamasAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.66
       Model |  1.2253e+09     3   408423359           Prob > F      =  0.0000
    Residual |   902971189    28    32248971           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5678.8

------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   636.6527    376.502     1.69   0.102    -134.5765    1407.882
interventionMARCH |    5.58394   4011.404     0.00   0.999    -8211.405    8222.573
   postslopeMARCH |  -1845.833   456.4154    -4.04   0.000    -2780.758   -910.9087
            _cons |    50148.6   3205.799    15.64   0.000     43581.82    56715.39
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   6125.807
Mean Absolute Errors     |   5148.2146
Pseudo-R2                |     0.44584
----------------------------------------
```

MEDIAN TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.75
       Model |  76524624.7     3  25508208.2           Prob > F      =  0.0000
    Residual |  56024918.5    28  2000889.95           R-squared     =  0.5773
-------------+------------------------------           Adj R-squared =  0.5320
       Total |   132549543    31  4275791.72           Root MSE      =  1414.5

------------------------------------------------------------------------------
T46NoFundH~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   345.4044   70.02958     4.93   0.000     201.9553    488.8535
intervention |  -3798.599    1011.09    -3.76   0.001    -5869.723   -1727.474
   postslope |   -439.108   109.7733    -4.00   0.000    -663.9684   -214.2476
       _cons |   9221.419   717.5897    12.85   0.000     7751.503    10691.33
------------------------------------------------------------------------------

. cv_regress
```

```
Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1513.5372
Mean Absolute Errors     |   1253.5942
Pseudo-R2                |      0.45732
----------------------------------------

reg T46NoFundHamasMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   14.61
       Model |  80886765.8     3  26962255.3           Prob > F      =  0.0000
    Residual |  51662777.4    28  1845099.19           R-squared     =  0.6102
-------------+------------------------------           Adj R-squared =  0.5685
       Total |   132549543    31  4275791.72           Root MSE      =  1358.3

------------------------------------------------------------------------------
T46NoFundHama~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   400.9118   73.66656     5.44   0.000     250.0127    551.8109
interventionMAY |  -3519.988   963.3153    -3.65   0.001     -5493.25   -1546.726
   postslopeMAY |  -529.4309   104.1803    -5.08   0.000    -742.8345   -316.0273
          _cons |   8888.375   712.3218    12.48   0.000      7429.25     10347.5
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1436.5901
Mean Absolute Errors     |   1220.6195
Pseudo-R2                |      0.50736
----------------------------------------



. reg T46NoFundHamasMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.74
       Model |  67688935.8     3  22562978.6           Prob > F      =  0.0001
    Residual |  64860607.4    28  2316450.27           R-squared     =  0.5107
-------------+------------------------------           Adj R-squared =  0.4582
       Total |   132549543    31  4275791.72           Root MSE      =    1522

------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   381.7107   90.95623     4.20   0.000     195.3953    568.0261
interventionAPRIL |  -1824.989   1075.153    -1.70   0.101     -4027.34    377.3625
   postslopeAPRIL |  -592.4264   118.1127    -5.02   0.000    -834.3693   -350.4835
            _cons |   8997.181   826.9851    10.88   0.000     7303.179    10691.18
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1645.5625
Mean Absolute Errors     |   1376.5626
Pseudo-R2                |      0.36082
```

```
----------------------------------------


. reg T46NoFundHamasMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.37
       Model |    62686297     3  20895432.3           Prob > F      =  0.0004
    Residual |  69863246.2    28  2495115.94           R-squared     =  0.4729
-------------+------------------------------           Adj R-squared =  0.4165
       Total |   132549543    31  4275791.72           Root MSE      =  1579.6

------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |       349.2    104.726     3.33   0.002     134.6784    563.7216
interventionMARCH |  -326.1035   1115.794    -0.29   0.772    -2611.703    1959.496
   postslopeMARCH |  -609.2341   126.9544    -4.80   0.000    -869.2884   -349.1797
            _cons |   9170.571   891.7101    10.28   0.000     7343.986    10997.16
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1706.3859
Mean Absolute Errors     |   1444.0275
Pseudo-R2                |      0.31438
----------------------------------------
```

44 Article Group Cross Validation (48 minus Hamas Fundamentalism Jihad Ammonium Nitrate articles)

RAW TOTAL MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundRAW time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  2.7271e+12     3  9.0903e+11           Prob > F      =  0.0000
    Residual |  1.0771e+12    28  3.8467e+10           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.0e+05

------------------------------------------------------------------------------
T44NoHamAm~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    30711.4   9709.946     3.16   0.004     10821.48    50601.32
intervention |  -593689.9   140192.6    -4.23   0.000    -880861.5   -306518.3
   postslope |  -52947.52   15220.61    -3.48   0.002    -84125.52   -21769.52
       _cons |    2196493   99497.33    22.08   0.000      1992682     2400304
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   210080.45
Mean Absolute Errors     |   174432.3
Pseudo-R2                |     0.63222
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  2.5266e+12     3  8.4221e+11           Prob > F      =  0.0000
    Residual |  1.2776e+12    28  4.5627e+10           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.1e+05

------------------------------------------------------------------------------
T44NoHamAmmJi~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |    32588.3   11584.35     2.81   0.009     8858.842    56317.76
interventionMAY |    -429858     151485    -2.84   0.008    -740160.9     -119555
   postslopeMAY |  -66264.44   16382.74    -4.04   0.000    -99822.96   -32705.91
          _cons |    2185232   112015.3    19.51   0.000      1955779     2414685
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   232324.24
Mean Absolute Errors     |   185440.75
Pseudo-R2                |     0.55250
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionAPRIL postslopeAPRIL
```

```
      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.12
       Model |  2.3522e+12     3  7.8406e+11           Prob > F      =  0.0000
    Residual |  1.4520e+12    28  5.1857e+10           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5774
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   31293.29   13608.94     2.30   0.029     3416.651    59169.93
interventionAPRIL |    -239699   160865.2    -1.49   0.147    -569216.3    89818.35
   postslopeAPRIL |  -74286.75   17672.11    -4.20   0.000    -110486.4   -38087.08
            _cons |    2192570   123734.1    17.72   0.000      1939113     2446028
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   247561.74
Mean Absolute Errors     |   199474.92
Pseudo-R2                |     0.49301
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.67
       Model |  2.2607e+12     3  7.5358e+11           Prob > F      =  0.0000
    Residual |  1.5434e+12    28  5.5123e+10           R-squared     =  0.5943
-------------+------------------------------           Adj R-squared =  0.5508
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   26300.13   15565.88     1.69   0.102    -5585.138     58185.4
interventionMARCH |   -35214.1   165845.2    -0.21   0.833    -374932.6    304504.4
   postslopeMARCH |  -76156.01   18869.78    -4.04   0.000      -114809   -37503.01
            _cons |    2219201   132538.7    16.74   0.000      1947707     2490694
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   253436.9
Mean Absolute Errors     |   211313.23
Pseudo-R2                |     0.46859
----------------------------------------
```

AVERAGE MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  1.4086e+09     3   469536182           Prob > F      =  0.0000
    Residual |   556363558    28  19870127.1           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4457.6

------------------------------------------------------------------------------
T44NoHamAm~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   697.9975   220.6837     3.16   0.004     245.9475    1150.048
intervention |  -13493.04   3186.241    -4.23   0.000    -20019.76   -6966.322
   postslope |  -1203.355   345.9278    -3.48   0.002    -1911.956   -494.7537
       _cons |   49920.14   2261.335    22.08   0.000     45288.01    54552.27
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4774.6186
Mean Absolute Errors     |   3964.3921
Pseudo-R2                |     0.63221
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  1.3051e+09     3   435018695           Prob > F      =  0.0000
    Residual |   659916019    28  23568429.3           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4854.7

------------------------------------------------------------------------------
T44NoHamAmmJi~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |     740.65   263.2848     2.81   0.009     201.3354    1279.965
interventionMAY |  -9769.475   3442.896    -2.84   0.008    -16821.93   -2717.022
   postslopeMAY |  -1506.013    372.341    -4.04   0.000    -2268.719   -743.3073
          _cons |   49664.22   2545.843    19.51   0.000      44449.3    54879.15
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5280.1831
Mean Absolute Errors     |   4214.5736
Pseudo-R2                |     0.55248
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.12
       Model |  1.2150e+09     3   404983797           Prob > F      =  0.0000
    Residual |   750020712    28    26786454           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5774
```

```
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  5175.6

------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   711.2107   309.2991     2.30   0.029     77.64032    1344.781
interventionAPRIL |  -5447.554   3656.086    -1.49   0.147    -12936.71    2041.599
   postslopeAPRIL |  -1688.336   401.6453    -4.20   0.000    -2511.069   -865.6025
            _cons |   49831.05   2812.184    17.72   0.000     44070.55    55591.55
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5626.4981
Mean Absolute Errors     |   4533.5744
Pseudo-R2                |     0.49299
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.67
       Model |  1.1677e+09     3   389240217           Prob > F      =  0.0000
    Residual |   797251454    28  28473266.2           R-squared     =  0.5943
-------------+------------------------------           Adj R-squared =  0.5508
       Total |  1.9650e+09    31  63386196.9           Root MSE      =    5336

------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   597.7121   353.7757     1.69   0.102    -126.9645    1322.389
interventionMARCH |  -799.9311   3769.269    -0.21   0.833    -8520.929    6921.066
   postslopeMARCH |  -1730.808   428.8654    -4.04   0.000    -2609.299    -852.317
            _cons |   50436.37   3012.291    16.74   0.000     44265.97    56606.77
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5760.0175
Mean Absolute Errors     |   4802.6243
Pseudo-R2                |     0.46858
----------------------------------------
```

MEDIAN MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.36
       Model |  57879940.6     3  19293313.5            Prob > F      =  0.0000
    Residual |  35166601.3    28  1255950.05            R-squared     =  0.6221
-------------+------------------------------           Adj R-squared =  0.5816
       Total |  93046541.9    31  3001501.35            Root MSE      =  1120.7

------------------------------------------------------------------------------
T44NoHamAm~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   277.2255   55.48251     5.00   0.000     163.5747    390.8763
intervention |  -3722.905    801.059    -4.65   0.000      -5363.8    -2082.01
   postslope |  -315.3791   86.97037    -3.63   0.001    -493.5298   -137.2283
       _cons |     8611.5   568.5266    15.15   0.000     7446.926    9776.074
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1204.2531
Mean Absolute Errors     |   998.7613
Pseudo-R2                |      0.50971
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.34
       Model |  63777614.9     3    21259205            Prob > F      =  0.0000
    Residual |    29268927    28  1045318.82            R-squared     =  0.6854
-------------+------------------------------           Adj R-squared =  0.6517
       Total |  93046541.9    31  3001501.35            Root MSE      =  1022.4

------------------------------------------------------------------------------
T44NoHamAmmJi~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   336.9338   55.44788     6.08   0.000      223.354    450.5137
interventionMAY |  -3670.966   725.0751    -5.06   0.000    -5156.215   -2185.717
   postslopeMAY |  -404.5265   78.41514    -5.16   0.000    -565.1526   -243.9003
          _cons |    8253.25   536.1555    15.39   0.000     7154.985    9351.515
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1079.2187
Mean Absolute Errors     |   935.57829
Pseudo-R2                |      0.60239
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   10.04
       Model |  48209351.7     3  16069783.9            Prob > F      =  0.0001
    Residual |  44837190.2    28  1601328.22            R-squared     =  0.5181
-------------+------------------------------           Adj R-squared =  0.4665
       Total |  93046541.9    31  3001501.35            Root MSE      =  1265.4
```

```
------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   309.8536   75.62426     4.10   0.000     154.9443    464.7629
interventionAPRIL |   -1926.92   893.9207    -2.16   0.040    -3758.034   -95.80659
   postslopeAPRIL |  -468.0104   98.20312    -4.77   0.000    -669.1704   -266.8505
            _cons |   8406.705    687.585    12.23   0.000     6998.251    9815.159
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1387.7669
Mean Absolute Errors     |   1091.2975
Pseudo-R2                |     0.35549
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.08
       Model |  43173895.2     3  14391298.4           Prob > F      =  0.0005
    Residual |  49872646.7    28  1781165.95           R-squared     =  0.4640
-------------+------------------------------           Adj R-squared =  0.4066
       Total |  93046541.9    31  3001501.35           Root MSE      =  1334.6

------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   283.9912   88.48334     3.21   0.003     102.7413    465.2411
interventionMARCH |  -632.2332   942.7373    -0.67   0.508    -2563.343    1298.877
   postslopeMARCH |  -489.7167   107.2642    -4.57   0.000    -709.4374    -269.996
            _cons |   8544.637   753.4085    11.34   0.000      7001.35    10087.92
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1445.5777
Mean Absolute Errors     |   1183.4531
Pseudo-R2                |     0.30089
----------------------------------------
```

List of 48 Terrorism Articles with Privacy Sensitivity Scores (from Survey)

Topic Keyword	Wikipedia Articles	Government Trouble	Browser Delete	Privacy Sensitive	Avoidance	Average
Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Iraq	http://en.wikipedia.org/wiki/iraq	1.60	1.74	1.76	2.25	1.84
Afghanistan	http://en.wikipedia.org/wiki/afghanistan	1.61	1.71	1.75	2.23	1.83
Iran	http://en.wikipedia.org/wiki/iran	1.62	1.73	1.78	2.25	1.85
Pakistan	http://en.wikipedia.org/wiki/pakistan	1.59	1.71	1.75	2.22	1.82
agro	http://en.wikipedia.org/wiki/agro	1.51	1.80	1.76	2.29	1.84
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Nuclear	http://en.wikipedia.org/wiki/nuclear	1.84	1.97	1.91	2.55	2.07
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Hamas	http://en.wikipedia.org/wiki/hamas	1.90	1.93	1.97	2.49	2.07
FARC	http://en.wikipedia.org/wiki/FARC	1.83	1.88	1.90	2.46	2.02
Irish Republican Army	http://en.wikipedia.org/wiki/Irish_Republican_Army	1.62	1.77	1.83	2.24	1.87
Euskadi ta Askatasuna	http://en.wikipedia.org/wEuskadi_ta_Askatasuna	1.86	1.88	1.88	2.43	2.01
Hezbollah	http://en.wikipedia.org/wiki/hezbollah	1.86	1.90	1.96	2.46	2.05
Tamil Tigers	http://en.wikipedia.org/wiki/Tamil_Tigers	1.76	1.86	1.87	2.39	1.97
PLO	http://en.wikipedia.org/wiki/Palestine_Liberation_Organization	1.77	1.87	1.91	2.42	1.99
Palestine Liberation Front	http://en.wikipedia.org/wiki/Palestine_Liberation_Front	1.81	1.89	1.95	2.47	2.03
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Peninsula	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12
Yemen	http://en.wikipedia.org/wiki/yemen	1.60	1.72	1.74	2.18	1.81
Pirates	http://en.wikipedia.org/wiki/pirates	1.44	1.67	1.67	2.10	1.72
Extremism	http://en.wikipedia.org/wiki/extremism	1.64	1.90	1.86	2.40	1.95
Somalia	http://en.wikipedia.org/wiki/somalia	1.50	1.68	1.67	2.12	1.74
Nigeria	http://en.wikipedia.org/wiki/nigeria	1.48	1.66	1.64	2.07	1.71
Political radicalism	http://en.wikipedia.org/wiki/Political_radicalism	1.75	1.91	1.97	2.48	2.03
Al-Shabaab	http://en.wikipedia.org/wiki/Al-Shabaab	1.84	1.89	1.89	2.48	2.03
nationalism	http://en.wikipedia.org/wiki/nationalism	1.48	1.71	1.73	2.20	1.78
Recruitment	http://en.wikipedia.org/wiki/recruitment	1.74	1.90	1.87	2.54	2.01
Fundamentalism	http://en.wikipedia.org/wiki/fundamentalism	1.60	1.79	1.80	2.32	1.88
Islamist	http://en.wikipedia.org/wiki/islamist	1.79	1.89	1.93	2.45	2.45
					MEDIAN	2.08

- 2.08 = Median of the Average Privacy Ratings for the 48 Articles
- 23 Most Privacy Sensitive Article Set includes all articles in the 48 Terrorism Group with combined privacy rating average above the 2.08 median.

Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Penin	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 3

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 15-cv-00662-TSE

SECOND DECLARATION OF MICHELLE PAULSON

I, Michelle Paulson, declare:

1. I am a resident of San Francisco, California, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as a former employee of and current consultant to the Wikimedia Foundation, Inc. ("Wikimedia").

2. As explained in my prior declaration in this litigation and the First Declaration of Emily Temple-Wood, Wikimedia has an active and close relationship with its community—which consists of the individuals who read or contribute to the body of knowledge comprising the twelve Wikimedia Projects. Although I focus here on Wikimedia's readers, it is critical to understand that readers and contributors are not static or mutually exclusive categories of users within the Wikimedia community. A Wikimedia user can be a reader at one moment in time, contribute an edit to a Wikipedia page at another, and then return to reading pages. In other words, our users play a variety of different roles as they engage with the Projects.

3. Below, I discuss several additional examples of Wikimedia programs and policies that help maintain the strong relationship between Wikimedia and its readers. These examples also illustrate the ways in which Wikimedia's ability to fulfill its mission depends on its ability to reach and engage readers. Indeed, readers are the beneficiaries of the vast body of human knowledge and free educational content that Wikimedia strives to develop and distribute online. Just as libraries work to enrich reader experiences and to expand their base of patrons in fulfillment of their mission, Wikimedia does the same. The ultimate aim of Wikimedia's work is to attract more readers as well as more contributors to the Projects. Readers are especially important to Wikimedia because, unlike patrons at a library, they also routinely generate content for the Projects—which, in turn, helps to attract even more readers.

4. First, unlike many other websites and platforms, Wikimedia offers all of its readers the opportunity to participate directly in the development of community standards and policies. Readers of the Projects are able to participate in discussions and decision-making that help determine how the community governs itself, and thus the shape of the Projects going forward. These include discussions about matters ranging from the Wikimedia movement's strategy and values to copyright issues and search query structure. *See, e.g.*, First Decl. of Michelle Paulson ¶¶ 9-11.[1]

5. Second, Wikimedia is engaged in a continuous process of consultation with its readers, soliciting their input on a variety of different topics related to the content and operation of the Projects.

[1] *See also, e.g.*, https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017; https://meta.wikimedia.org/wiki/Values/2016_discussion; https://meta.wikimedia.org/wiki/Copyright_strategy; https://www.mediawiki.org/wiki/User_Interaction_Consultation.

a. Wikimedia assigns various teams to conduct research and surveys to learn about readers' experiences of the Wikimedia Projects. Some of these research projects are specifically designed to help increase and diversify readership. For instance, Wikimedia has a "New Readers" team, which is dedicated to learning about the preferences of readers in underserved countries with respect to Wikimedia content and their reading experiences.

b. Other research projects are designed to gather information to enrich current readers' experiences—for example, a short survey that asks readers, "Why are you reading this article?"

c. Wikimedia's research activities involve significant expenditures, including costs related to survey design, consultants, and foreign travel to interview Wikimedia readers, contributors, and others who have not yet engaged with the Wikimedia projects.

6. Third, Wikimedia devotes considerable resources to maintaining and increasing the engagement and involvement of its readers. Wikimedia has an entire Department dedicated to the engagement of its community, including readers. The Community Engagement Department is responsible for a variety of functions, including facilitating communications between Wikimedia and its community, and assisting Wikimedia and its community with studying and evaluating projects and program activities to help make them more effective. In addition, Wikimedia's Galleries, Libraries, Archives, and Museums program establishes partnerships between Wikimedia and these institutions to help incorporate their content into the Projects. By incorporating images and information from these institutions into Wikimedia pages, Wikimedia seeks to enrich the experiences of current readers and to attract new ones.

7. Fourth, Wikimedia devotes considerable resources to expanding its readership, including by making the content of the Projects available to greater numbers of readers. For example:

a. In the interests of disseminating educational content to a broader readership base, Wikimedia has dedicated substantial resources to the translation of its web pages, to help ensure that robust content is available in multiple languages.

b. Wikimedia's Communications Department runs Wikimedia-awareness campaigns around the world to increase readership. Recent campaigns involved outreach in Mexico and Nigeria. These campaigns include the creation and local dissemination of videos and other advertisements, such as the following video, Superdotada, available at https://www.youtube.com/watch?v=fsCxG0MTrJs.

c. The Wikipedia Zero program, discussed at greater length in my earlier declaration, has been primarily focused on providing free access to Wikipedia readers via mobile devices.

d. Wikimedia engineering teams have dedicated time and resources to ensuring that the Projects are available to readers using older model mobile phones or web browsers that are popular with people in emerging markets. This helps ensure that readers in all locations, even with readers with older devices or slow or limited internet access, can use the Wikimedia Projects.

8. Finally, Wikimedia has advocated for and will continue to zealously advocate for all of its readers' and contributors' free expression and privacy rights, including the rights afforded by the First and Fourth Amendments. As discussed at greater length in my earlier declaration, Wikimedia takes numerous, costly steps to protect the confidentiality of its communications,

including its communications with readers. Through policies, public statements, and guidelines, Wikimedia assures its community—including its readers—that it will reject third-party requests for non-public user information unless it is legally required to disclose that information. In keeping with these assurances, Wikimedia resists third-party demands for information—including information about readers' page views—that are overly broad, unclear, or irrelevant, and it notifies users individually of information requests when legally permitted and it is able to do so.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2018 in Los Angeles, California.

Michelle Paulson

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 4

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

SECOND DECLARATION OF TILMAN BAYER

I, Tilman Bayer, declare:

1. I am a resident of San Francisco, California, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as an employee of the Wikimedia Foundation, Inc. ("Wikimedia").

2. I am a Senior Analyst in Wikimedia's Product Analytics team, and have been a full-time employee of the organization since 2012. My responsibilities include the reporting of pageview statistics and other key web traffic metrics to Wikimedia's executives and board. I hold degrees in mathematics from the University of Cambridge (Certificate of Advanced Study in Mathematics) and the University of Bonn (diploma, equivalent to a Master's degree in the US).

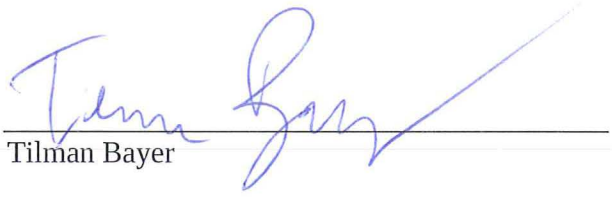
3. As explained in my previous Declaration in support of Wikimedia's Opposition to Defendants' Motion for Summary Judgment, Wikimedia provides technical infrastructure to twelve free-knowledge "Projects" on the Internet. (Pl.'s Ex. 5 ¶¶ 3-4.) Wikimedia users edit specific Wikimedia Project pages via "Edit" pages, which contain a text box with editable versions of the online article. As of March 2018, there have been approximately 3.4 billion edits over the

lifespan of the Wikimedia Projects. (*Id.* ¶¶ 12-13.) Edits are transmitted from users to Wikimedia servers as HTTP and HTTPS communications, part of the previously described *Category 1 – Wikimedia communications with its community members*. (*Id.* ¶ 26.)

4. When a user edits Wikimedia Project pages, the final changes to the master versions of Wikimedia's content databases are stored on Wikimedia's servers located in the United States, not Wikimedia's caching servers located abroad. (*See id.* ¶ 4 (noting country locations of Wikimedia's servers).) For this reason, all edits to Wikimedia Project pages from users located in foreign countries are routed to Wikimedia's servers in the United States.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2019 in San Francisco, California.



Tilman Bayer

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 5

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 15-cv-00662-TSE

SECOND DECLARATION OF EMILY TEMPLE-WOOD

I, Emily Temple-Wood, declare:

1. I am a U.S. citizen and resident of Downers Grove, Illinois, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as a Wikimedia Foundation, Inc. ("Wikimedia") community member. I am not an employee or contractor of Wikimedia.

I. Background

2. My earlier declaration in this litigation attested to the relationship between Wikimedia and its users; the importance of Wikimedia's non-U.S. readers and contributors to U.S. users like myself; the importance of anonymity to Wikimedia users; and obstacles that other users and I face in bringing suit.

3. As detailed in that declaration, since April 2007, I have been both a reader and an editor of the English Wikipedia. I have created nearly 400 articles on Wikipedia and have edited thousands more.

II. Contributing to Wikimedia While Outside of the United States

4. Throughout my time as a member of the Wikimedia community, I have traveled to numerous countries outside of the United States, including for the purpose of attending Wikimedia conferences abroad.

5. As a U.S. citizen traveling abroad, I have continued to serve as an editor of the English Wikipedia. While outside of the United States, I have edited hundreds of Wikipedia articles, including:

- Cherney incision, https://en.wikipedia.org/wiki/Cherney_incision (edited from Mexico);
- Urethral diverticulum, https://en.wikipedia.org/wiki/Urethral_diverticulum (edited from Mexico);
- X-linked recessive hypoparathyroidism, https://en.wikipedia.org/wiki/X-linked_recessive_hypoparathyroidism (edited from Mexico);
- Lucie Randoin, https://en.wikipedia.org/wiki/Lucie_Randoin (edited from Mexico);
- Birt-Hogg-Dubé syndrome, https://en.wikipedia.org/wiki/Birt%E2%80%93Hogg%E2%80%93Dub%C3%A9_syndrome (edited from Hong Kong);
- Wikipedia talk:WikiProject Women scientists: Difference between revisions, https://en.wikipedia.org/w/index.php?title=Wikipedia_talk:WikiProject_Women_scientists&diff=prev&oldid=567822352 (edited from Hong Kong); and
- Endometrial cancer, https://en.wikipedia.org/wiki/Endometrial_cancer (edited from the United Kingdom).

6. While traveling abroad, I have edited Wikipedia articles from my pseudonymous account.

7. While traveling abroad, I have removed sensitive items within Wikipedia pages from public view in my capacity as an English Wikipedia administrator.

8. I have also witnessed dozens of individuals whom I believe to be U.S. citizens or permanent residents edit Wikipedia articles while they, too, are outside of the United States.

9. I intend to continue traveling abroad on at least an annual basis. During these travels, I also intend to continue editing Wikipedia articles.

10. As a Wikipedia user, I am concerned about government surveillance, including Upstream surveillance of my communications while abroad with Wikimedia servers in the United States. *See also* First Declaration of Emily Temple-Wood ¶ 20.

III. Upstream Surveillance Impairs My and Other U.S. Users' Ability to Exchange Information with Wikimedia's Foreign Readers and Editors

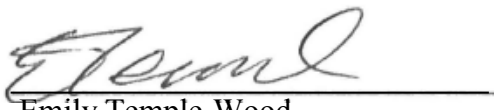
11. As detailed in my earlier declaration, my interest in contributing to Wikipedia is based in part on my ability to reach an international audience, and I also read and benefit from the contributions of non-U.S. users on a wide array of topics.

12. NSA surveillance, including Upstream surveillance, threatens the anonymity and privacy of Wikimedia community members, and that threat discourages individuals from engaging with the Wikimedia Projects.

13. The loss of foreign readers and editors is a direct detriment to me and to the Wikimedia community at large. When foreign users are less willing to read Wikimedia Project pages, I and other community members lose potential readers of our original content and other contributions to the Projects. When foreign users are less willing to contribute to Wikimedia Project pages, I and other community members are unable to read and benefit from their contributions.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2019, in Downers Grove, Illinois.



Emily Temple-Wood

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 16

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	)	No. 1:15-cv-0662 (TSE)

(THIRD) DECLARATION OF DR. HENNING SCHULZRINNE

Dr. Henning Schulzrinne, for his (third) declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

INTRODUCTION AND SUMMARY

1. I am the Julian Clarence Levi Professor of Computer Science at Columbia University in New York, New York. I previously submitted two declarations in this case, dated November 12, 2018 and February 15, 2019. I submit this third declaration at the request of the United States Department of Justice to address the conclusions reached by Mr. Scott Bradner in his reply declaration filed on March 8, 2019, including Mr. Bradner's assessment of conclusions reached in my February 15 declaration. My background and qualifications in the fields of computer science, electrical engineering, and digital communications technology; the sources of information I considered in arriving at the conclusions stated in this case (apart from those cited herein); and my compensation for my services in this matter, are all stated in my prior declarations.

2. For the reasons I detail herein, it remains my conclusion that the hypothesis advanced in this case by plaintiff Wikimedia Foundation ("Wikimedia"), that the National Security Agency ("NSA"), in the course of conducting Upstream collection, must as a matter of

technological necessity be intercepting, copying, and reviewing at least some of Wikimedia's electronic communications that traverse the Internet, is incorrect. Based on what is publicly known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications. Nothing stated in Mr. Bradner's reply declaration, or in Wikimedia's sur-reply, alters that conclusion.

3. I also adhere to the conclusions reached in my second declaration, (Second) Declaration of Dr. Henning Schulzrinne (hereinafter, "Second Decl."). In his first declaration Mr. Bradner concluded (i) that the NSA "most likely" uses his copy-all-then-scan configuration to conduct Upstream collection, and (ii) that it is "implausible" that the NSA employs a filter-then-copy-and-scan approach, using whitelisting and blacklisting techniques, such as I described in my first declaration; and (iii) that that even if the NSA uses one or more of the techniques I describe, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. *See generally* Declaration of Scott Bradner (hereinafter, "Bradner Decl."). In my second declaration, I explained that none of these conclusions has a foundation in Internet technology or engineering, and instead are based principally on assumptions Mr. Bradner makes about the NSA's practices and priorities, its resources and capabilities, and its Upstream surveillance targets, matters about which Mr. Bradner has no specialized knowledge or information. These conclusions also remain unaltered by Mr. Bradner's reply declaration, or Wikimedia's sur-reply.

4. In his second declaration, Mr. Bradner takes a somewhat different approach. The central thesis of Mr. Bradner's second declaration is that the use of traffic-mirroring techniques (that is, whitelisting and blacklisting) to filter the communications traversing a monitored link before communications of interest are copied and then scanned for selectors, would "conflict[ ]" with "the government's own descriptions" of the Upstream program. Reply Declaration of Scott Bradner (hereinafter "Bradner Reply Decl.") ¶¶ 6, 12, 30. By the "government's own

descriptions," Mr. Bradner means, first, a single statement contained in an 80-page opinion issued by the Foreign Intelligence Surveillance Court ("FISC") in October 2011, regarding the NSA's acquisition of wholly domestic "about" communications, Bradner Reply Decl., App'x P, and second, a single remark in the nearly 200-page report of the Privacy and Civil Liberties Oversight Board on the NSA's Section 702 surveillance program (hereinafter, "PCLOB Section 702 Report"), regarding the NSA's goal of "comprehensively" acquiring communications to or from its targets, Bradner Reply Decl., App'x F.

5. As I already discussed in my prior declaration, the use of a filter-then-copy-and-scan approach to Upstream collection would be entirely consistent with both of these statements. Second Decl. ¶¶ 56-58. The conflict perceived by Mr. Bradner with the FISC's statement arises not from any technical grounds but his own speculative interpretation of that statement, and his assumption that a remark made by the FISC nearly eight years ago still reflects the technical realities of Upstream collection today. The supposed conflict that Mr. Bradner attempts to seize on with the PCLOB's remark ignores the difference between a statement of aspirations on a printed page and the real-world challenges of designing, constructing, deploying, maintaining, and paying for the collection systems required to implement the kind of Upstream collection process that Mr. Bradner envisions. Moreover, he misconstrues the PCLOB's description of the NSA's prior collection of "about" communications as an aim of the program, whereas the PCLOB clearly described "about" collection as "byproduct" of the NSA's efforts to acquire communications sent to or from its Upstream surveillance targets.

6. Mr. Bradner also opines that a filter-then-copy-and-scan approach "conflicts with other technical and practical necessities of conducting" Upstream surveillance. Bradner Reply Decl. ¶¶ 7, 61-112. As I also discuss below, these so-called "technical and practical necessities" are based on the unsupported conclusions Mr. Bradner draws from the FISC and PCLOB statements, as well as many of the same non-technical, speculative assumptions made in his previous declaration about the NSA's surveillance practices and priorities, resources and

capabilities, and its targets, about which Mr. Bradner has no actual knowledge. As such, they still provide no basis in Internet technology or engineering for the conclusions he reaches.

7. Mr. Bradner remarks that I do not offer evidence that the NSA is actually using whitelisting and/or blacklisting techniques in the course of Upstream surveillance that would avoid interaction with Wikimedia's communications. Bradner Reply Decl. ¶¶ 6, 57-58. The same can be said of Mr. Bradner, of course, that he has offered no evidence—only speculation—that the NSA conducts Upstream surveillance in the manner that he describes. As I have explained before, I did not attempt in my previous declarations to reach conclusions about how the NSA actually conducts Upstream surveillance, because I was not asked by the Department of Justice to opine on that question, because the question implicates operational details of Upstream surveillance that remain classified, and because it would have required that I engage in the same sort of speculation as Mr. Bradner, concerning the NSA's actual surveillance practices, capabilities, and targets, about which neither I nor Mr. Bradner has any specialized knowledge or information. *See* Second Decl. ¶ 3. I do not engage in such speculation now, either.

8. As was the case with my first two declarations, in reaching the conclusions stated herein I have not considered, nor have I been provided with, any classified or other non-public information concerning Upstream surveillance.

MR. BRADNER'S CONCLUSION THAT WHITELISTING AND BLACKLISTING WOULD "CONFLICT" WITH THE FISC'S 2011 STATEMENT REGARDING THE COLLECTION OF WHOLLY DOMESTIC "ABOUT" COMMUNICATIONS LACKS A FOUNDATION IN INTERNET TECHNOLOGY AND ENGINEERING

9. As Mr. Bradner observes, in an October 2011 opinion concerning the legal implications of the NSA's collection of so-called multi-communication transactions ("MCTs"), the FISC stated that:

> Indeed, the government readily concedes that NSA will acquire a wholly domestic "about" communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server. See June 1 Submission at 29.

Bradner Reply Decl., App'x P at 45.

10. In his previous declaration, Mr. Bradner cited this statement in support of a very different conclusion than the one he offers now. Earlier, he cited the FISC's statement as evidence only that the NSA, at least at some monitored links, was not using "an IP filter to eliminate wholly domestic [communications] before" copying and scanning. Bradner Decl. ¶ 294; *see* Second Decl. ¶ 56. Mr. Bradner did not cite this conclusion as a ground for viewing his copy-all-then-scan configuration as "most likely," or a filter-then-copy-and-scan approach as "implausible." *See* Bradner Decl. ¶¶ 282-89, 366-67. Nevertheless Wikimedia, in its summary judgment opposition brief, argued that that the acquisition of some wholly domestic communications, even at a so-called "international Internet link," is inconsistent with the use of the traffic-mirroring techniques I have described, even though Mr. Bradner had not made any such assertion in his first declaration filed at that time. *See* Second Decl. ¶ 56. I explained, therefore, why use of whitelisting and blacklisting techniques would be consistent with the acquisition of at least some wholly domestic "about" communications, Second Decl. ¶¶ 56-58, a conclusion that Mr. Bradner does not dispute, *see* Bradner Reply Decl. ¶¶ 36, 111.

11. Only now, in his second declaration, does Mr. Bradner make an argument that implementing a filter-then-copy-and-scan approach would be inconsistent with the collection of wholly domestic about communications, as referred to in the FISC's opinion. Bradner Reply Decl. ¶¶ 32-45. Indeed, he is emphatic that the FISC's statement would have to be "false" for the NSA to employ a filter-then-copy-and-scan approach. *Id.* ¶ 154(1). This is the case, according to Mr. Bradner, because the use of whitelists or blacklists to filter out certain communications traversing a monitored link could filter out at least some wholly domestic about communications, containing targeted selectors, that the NSA might otherwise acquire. *Id*. ¶ 44. That outcome, Mr. Bradner states, "would not be consistent with the FISC's statement that *all* 'about' communications 'will' be acquired." *Id*. ¶ 42 (emphasis mine).

12. But the FISC did not state that "all" wholly domestic about communications crossing a monitored international Internet link will be acquired by the NSA. It said that "a wholly domestic 'about" communication" will be acquired if it crosses such a link, or is routed through a

foreign server. Mr. Bradner arrives at the conclusion that whitelisting and blacklisting would be inconsistent with the FISC's statement by imputing to the FISC a term, "all," that the FISC did not use. His contention that a filter-then-copy-and-scan approach would not be consistent with the acquisition of wholly domestic about communications is not based on a disagreement over Internet technology or engineering, but on his current reading of the FISC's 2011 statement, one that he did not propose in his first declaration, *see* Bradner Decl. ¶ 296.

13. There are numerous reasons why the statement in the FISC's 2011 opinion does not support Mr. Bradner's conclusion about how Upstream collection must operate today. First, as noted, the FISC simply does not use the term "all" in its statement regarding the collection of wholly domestic about communications.

14. Second, Mr. Bradner does not take into account the context in which the FISC made this remark. The paragraph in which the statement appears concerned the NSA's inability to prevent the acquisition of certain wholly domestic communications, and the Government's suggestion that these acquisitions resulted from a "failure" of the NSA's "technical means." Bradner Reply Decl., App'x P at [45]. The FISC was unwilling to accept that explanation, finding no reason to conclude that the collection of wholly domestic communications was attributable to malfunctions or failures in the NSA's collection equipment. This was the point at which the FISC remarked that the Government had conceded that a wholly domestic about communication would be collected if it crossed a monitored international Internet link, or was routed through a foreign server. In other words, it appears that when the FISC made this remark it was explaining that acquisitions of wholly domestic about communications would occur as the result of technical limitations in the equipment's normal operation (rather than as the result of a malfunction), and was not making a statement about the scope or completeness of those acquisitions.

15. Third, and telling, in an earlier portion of the FISC's opinion where it was discussing the same phenomenon—the NSA's inability to prevent the acquisition of at least some wholly domestic communications—the FISC stated that due to this limitation on the NSA's abilities, the "NSA *may* acquire wholly domestic communications." Bradner Decl., App'x P at 35 n.34

(emphasis mine). Mr. Bradner's conclusion, that according to the FISC the NSA obtained all wholly domestic about communications crossing a monitored link, is thus further undermined by the FISC's clear statement describing the NSA's acquisition of wholly domestic communications as a possibility ("may acquire"), rather than a certainty in all cases.[1]

16. Mr. Bradner also suggests that the FISC must have been "as precise as it possibly could be" when describing the acquisition of wholly domestic about communications, based on the "multiple hearings" it held and the "multiple submissions" it received from the Government before issuing its October 2011 opinion. Bradner Reply Decl. ¶ 38. Of course, this assumption is not a technical basis on which to reach conclusions about how Upstream surveillance is conducted. And I do not understand how Mr. Bradner could know exactly what was said by the Government in those submissions, what was addressed during those hearings, or in what depth, all of which I am advised by the Department of Justice remain classified, in whole or in substantial part. In the end, however, if the FISC meant to say that the NSA would acquire *all* wholly domestic communications crossing a (hypothetically) monitored international Internet link, then the most precise way of expressing that thought would have been to use the word "all." The FISC did not do so.

17. On this subject I observe finally that the FISC's statement cited by Mr. Bradner was made in October 2011, based on a June 1, 2011 submission by the Government, Bradner Reply Decl., App'x P, at 45, nearly eight years ago. Even if the FISC meant that in June 2011 the NSA would acquire all wholly domestic "about" communications crossing an international Internet

[1] Although in this earlier passage the FISC referred to the collection of "wholly domestic communications," rather than "wholly domestic about communications," as in the later passage, there is no reason to believe that it meant to suggest that the collection of wholly domestic communications in general (including both "to/from" and "about" communications) was only a possibility, but at the same time that collection of all wholly domestic "about" communications was a certainty. At least as described in the PCLOB Section 702 Report, the underlying causes for acquisition of wholly domestic communications are similar for both cases: once a communication (for example, an email), makes it past any IP filters, the scanning mechanism would be just as likely to pick up wholly domestic "to" or "from" communications that contain targeted selectors as "about" communications containing targeted selectors. *See* PCLOB Section 702 Report at 10, 37, 84-86, 123, 124, 144-45.

link theoretically monitored by the NSA, and it could be inferred, therefore, that the NSA was not then employing whitelist or blacklist filters, it cannot be taken for granted that the situation has remained the same since then. The growth of Internet traffic, even since 2011, has been enormous. Between 2012 and 2016 alone, the volume of traffic (as measured in used international bandwidth by Telegeography, Inc.) more than quadrupled from 100 terabits per second to over 400 terabits per second[2]. *See* https://blog.telegeography.com/shaping-the-global-wholesale-bandwidth-market (figure 1). Given the growth in international Internet traffic since 2011, the possibility cannot be ignored that the NSA at some point might have adopted a form of whitelisting or blacklisting at monitored links to reduce the technical and logistical challenges and costs of processing large volumes of traffic, as discussed in my second declaration, Second Decl. ¶¶ 20-21. Neither Mr. Bradner nor I can know which is the case, of course, but it cannot be taken for granted that the assumption made by Mr. Bradner is the correct one.

18. As I have noted, Mr. Bradner does not disagree with my conclusion that the use of whitelisting and blacklisting would be consistent with the acquisition of at least some wholly domestic "about" communications at international Internet links that hypothetically might be monitored by the NSA. *See* Bradner Reply Decl. ¶¶ 36, 111. Therefore, because it is speculative to suggest (i) that the FISC meant to say that the NSA would collect *all* wholly domestic "about" communications if they crossed a monitored international Internet link in 2011, and (ii) even if that was the FISC's meaning in 2011, that the NSA would continue to collect all wholly domestic about communications at such links years afterward, Mr. Bradner lacks a technical basis on which to conclude that whitelisting and blacklisting would be inconsistent with the FISC's statement.

19. In sum, I find no reason, and certainly none based in Internet technology or engineering, to conclude that whitelisting or blacklisting at international Internet links that the NSA may in theory be monitoring would be inconsistent with the 2011 FISC statement cited by Mr. Bradner.

[2] This traffic volume includes all international traffic, not just traffic entering and exiting the United States. The growth rates, however, have been similar across all major geographies.

THE GOAL OF "COMPREHENSIVELY" ACQUIRING TARGETS' COMMUNICATIONS IS NOT A TECHNICAL BASIS ON WHICH TO CONCLUDE THAT THE NSA DOES NOT EMPLOY WHITELISTING OR BLACKLISTING TECHNIQUES

20. Mr. Bradner next opines that a filter-then-copy-and-scan approach to Upstream collection process would be "incompatible" with the goal of "comprehensively acquir[ing] communications that are sent to or from [the NSA's] targets," as stated on pages 10 and 123 of the PCLOB Section 702 Report. Bradner Reply Decl. ¶¶ 46-54.[3] He goes so far as to say that the PCLOB's statement would have to be "false" for the NSA to conduct Upstream collection using the traffic-mirroring techniques I have described. *ID.* ¶ 154(2). As I have explained previously, Mr. Bradner offers no technical basis for concluding as a matter of concrete reality that the NSA "must be" copying and reviewing all communications crossing a monitored link, Bradner Reply Decl. ¶ 29, simply because the NSA might wish in the abstract to obtain as many of its targets' communications as it can. Second Decl. ¶¶ 71-75.

21. Specifically, I explained that one cannot assume away, based on a stated aim of comprehensiveness, the many technical, logistical, and financial hurdles, and competing mission priorities, that would stand in the way of designing, constructing, deploying, and maintaining the kind of collection systems envisioned by Mr. Bradner. Second Decl. ¶ 73. (See also paragraphs 20-21 of my second declaration, in which I describe some of the technical and logistical challenges of implementing Mr. Bradner's copy-all-then-scan approach.) Therefore, one cannot draw meaningful conclusions about the technical details of the NSA's Upstream collection systems with nothing more to go on than an abstract goal of comprehensiveness. Second Decl. ¶ 74.

22. Mr. Bradner does not dispute these practical realities. He himself states that "the NSA must operate in the real world and deal with the technical and operational limitations

[3] I note that in contrast the PCLOB describes the collection of "about" communications as a "byproduct" of the NSA's efforts to collect communications sent to or from its targets. PCLOB Section 702 Report at 10, 123. The PCLOB's description of "about" collection as a "byproduct" rather than an objective of Upstream collection tends to rebut Mr. Bradner's interpretation of the FISC's October 2011 statement to mean that the NSA must have configured Upstream collection in such a way as to acquire "all" wholly domestic about communications on a monitored link. *See also* paragraphs 47-49, below.

inherent in the Internet." Bradner Reply Decl. ¶ 10. Yet he continues to disregard these same realities when he attempts to draw technical conclusions about how Upstream surveillance "must be" conducted from a single statement by the PCLOB attributing a goal of "comprehensiveness" to the NSA.

23. Mr. Bradner remarks that the use of whitelists and blacklists to reduce the technical and logistical difficulties and costs of Upstream collection is "incompatible" with the goal of completeness, Bradner Reply Decl. ¶ 51, but they are entirely compatible with reducing the formidable technical and logistical burdens and costs of processing large volumes of communications traffic, and Mr. Bradner has no way of knowing, and certainly no technical way of determining, whether such "real world" constraints have convinced or compelled the NSA to compromise the goal of completeness. It is, at bottom, the hard realities of achieving Mr. Bradner's vision of Upstream collection that may have proven incompatible with that goal, just as they are incompatible with a bare assumption that the NSA has succeeded in implementing that vision simply because it wants to.

24. Mr. Bradner opines that the single term "comprehensively" provides an "appropriate basis on which to explain the technological implementation" of Upstream collection, because "the PCLOB used the term 'comprehensively' to explain the need for the NSA's specific technological implementation" of the program. Bradner Reply Decl. ¶ 48. But the PCLOB's Section 702 Report, being an unclassified public report about a highly classified foreign-intelligence gathering activity, contains no specific technical detail about how Upstream collection is conducted. It certainly includes no technical detail at the level required to support the conclusions reached about Upstream collection by Mr. Bradner, except as a matter of speculation and conjecture.

25. "If wishes were horses," as they say, then one could simply assume that the NSA had the technical, logistical, and financial wherewithal, consistent with all its other mission requirements, to deploy an Upstream collection architecture capable of comprehensively acquiring every single one of its targets' online communications, just because it would like to do

so. And perhaps it has succeeded in doing so. I (like Mr. Bradner) do not know. But the assumption has no basis in the "real world" of Internet technology and engineering, Bradner Reply Decl. ¶ 10, in which the NSA must operate.

MR. BRADNER'S OPINION THAT WHITELISTING AND/OR BLACKLISTING WOULD "CONFLICT" WITH "TECHNICAL AND PRACTICAL NECESSITIES" OF UPSTREAM COLLECTION RESTS ON SPECULATION RATHER THAN A BASIS IN INTERNET TECHNOLOGY AND ENGINEERING

26. In addition to the inferences he draws from his interpretation of the FISC's 2011 statement, and the PCLOB's use of the term "comprehensive" to describe Upstream's objective, Mr. Bradner also attempts to reach conclusions based on what he calls "technical and practical necessities" that make clear, in his view, that the NSA is copy and scanning at least some of Wikimedia's communications. As I discuss below, these "technical and practical necessities" turn out again to be assumptions by Mr. Bradner about the NSA's surveillance practices and priorities, resources and capabilities, and the nature and behavior of its Upstream surveillance targets, mixed with the unsupported inferences he draws from the FISC and PCLOB statements discussed above. They supply no basis in Internet technology and engineering for concluding that the NSA "most likely" uses a copy-all-then-scan configuration to conduct Upstream surveillance, or that a filter-then-copy-and-scan approach is "implausible."

Whitelisting IP Addresses of Interest

27. Developing and maintaining whitelists: In Mr. Bradner's view whitelisting would be unworkable for purposes of Upstream collection because it would require "knowing in advance all of the IP addresses that might be used by each of the NSA's targets," Bradner Reply Decl. ¶ 68, and maintaining "comprehensive" information on where they will be, what sites they communicate with, and the protocols they use, Bradner Reply Decl. ¶ 69. *See also id.* ¶ 88. These conclusions lack a non-speculative technical basis.

28. As a technical matter, the NSA would not need to know all of its targets' IP addresses or gather comprehensive information on their whereabouts, the sites they visit, or the types of online communications in which they engage, before it could whitelist the IP addresses that the agency already knows about. As pointed out in the Government's reply brief, the NSA

could acquire information about its targets' IP addresses from communications acquired under the Section 702 PRISM program (*see* PCLOB Section 702 Report (Bradner Decl., App'x F) at 7, 33-34), Executive Order 12,333, prior Upstream acquisitions, information obtained from other U.S. intelligence agencies, and other sources. Reply Brief in Support of Defendants' Motion for Summary Judgment at 12-13. Mr. Bradner does not dispute this. Presumably communications acquired from these sources would also provide insight into particular websites of interest that targets visit, and their preferred modes of communication, thus allowing for further enhancement of the NSA's whitelist(s).

29. Mr. Bradner suggests no technological reason why it would not be possible for the NSA to conduct Upstream surveillance in this fashion. Doing so would not mean that the NSA "is only interested in the people and processes it already knows about and that it has decided to actively ignore everything else." Bradner Reply Decl. ¶ 89. Nothing about the use of whitelists would prevent the NSA from learning about new potential targets of interest, or new information about the communications of its existing targets, either from ongoing Upstream collection or other intelligence sources (in Mr. Bradner's words, discovering which streetlights to look under for your keys, *id.* ¶ 70), and then updating its whitelists accordingly.

30. There would be no need whatsoever for the NSA to know in advance all the information Mr. Bradner refers to unless one works backward from the conclusion, as Mr. Bradner does, that the NSA's acquisition of its targets' communications must be comprehensive, *see* Bradner Reply Decl. ¶¶ 69, 88, based on the PCLOB's passing remark about the NSA's goals. As I have discussed, the PCLOB's report does not provide a technical justification for concluding that the NSA's acquisition of its targets' communications is comprehensive in fact, or drawing inferences based on that premise about how Upstream surveillance is conducted.

31. Mr. Bradner observes that whitelists would have to be updated as targets were added or removed, or changed their locations or modes of communications. Bradner Reply Decl. ¶ 85. This is true, of course, but how often that would have to occur Mr. Bradner does not say,

because he cannot know, and he certainly offers no reason why updating whitelists as needed would be beyond the NSA's capabilities.

32. Target numerosity: Mr. Bradner next repeats his argument that the NSA's targets are too numerous to make the development and maintenance of whitelists of their IP addresses practical. Bradner Reply Decl. ¶¶ 75-76; *see* Bradner Decl. ¶ 366(d). I have already explained that in so arguing Mr. Bradner is making speculative assumptions about the number, nature, and communications habits of the NSA's Upstream targets, about which Mr. Bradner has no information. First Decl. ¶¶ 45-48. He does not maintain otherwise. Instead, he offers a hypothetical in place of facts: if the NSA had 1,000 Upstream targets in 2011, when according to the FISC it acquired 26 million communications using its Upstream collection technique, then on average each target would have had to engage in at least 26,000 online communications that year. Bradner Reply Decl. ¶ 76. But if we used another randomly chosen number of Upstream targets, say 5,000, then the number of communications each target would have had to send or receive in 2011 drops to 5,200, or approximately 14 per day. That is a trivial number considering (i) that "communications" can include not only such media as email but also all the individual HTTPS (or HTTP) requests and responses made during a single visit to a website, and (ii) that the NSA's Upstream targets could include organizations as well as individuals, organizations employing dozens or hundreds of persons capable of generating thousands of communications per day. Mr. Bradner has given no factual, technical basis for concluding that the number of the NSA's Upstream targets would make whitelisting unworkable.

33. Target mobility: Mr. Bradner also repeats his arguments that NSA use of whitelists is not "remotely possible" because its targets may move around, and as a result their IP addresses could change; because targets may use intermediary communications services, such as virtual private networks (VPNs, which remove the targets' IP address from packet headers during certain legs of their journeys across the Internet); and because targets may use multiple Internet service providers (ISPs), that assign different IP addresses to their subscribers. Bradner Reply Decl. ¶¶ 77-78, 82. In my view, Mr. Bradner overstates the matter. To take just one example, if the

NSA were seeking to track a targeted individual's email, determining the source and destination IP addresses to whitelist could be automated readily. An exhaustive list of SMTP destination IP addresses for any email address can be looked up online via DNS (MX records), and the list of source IP addresses can be derived from the DNS SPF entries. These IP addresses would not be affected by user mobility, and would only depend on the sender or receiver email address, not the user's current location.

34. But again, most fundamentally, Mr. Bradner is making assumptions about the nature, mobility, and communications practices of the NSA's targets. Only the NSA knows the extent to which its targets' IP addresses change due to their mobility, the extent to which they use VPNs, or multiple ISPs, to communicate, and therefore whether whitelisting would be impractical for purposes of meeting its intelligence-collection needs. Whitelisting could not be considered technologically impossible, however, unless one started from the premise that the NSA must be "comprehensively" acquiring every single one of its targets' communications without fail, and worked backward from there. As I have explained, Mr. Bradner has offered no justification, certainly no technological justification, for adopting that premise.

35. Further, as I note in my second declaration, to the extent a target moves from place to place within a given geographic area, the NSA could whitelist a set of IP addresses, rather than just a single address, associated with the geographic region where the target is believed to be located. Second Decl. ¶ 47. Mr. Bradner does not dispute this, but points out that using a range of IP addresses could increase the chances that communications to or from Wikimedia could be copied and scanned. Bradner Reply Decl. ¶¶ 80, 82. This is true in the abstract, but Mr. Bradner cites no information about the extent to which the NSA might find it necessary to whitelist ranges of IP addresses—how many, how large, in what geographic areas—in order to reliably monitor its targets' communications. As a result he can only speculate whether whitelisting ranges of IP addresses would result in the NSA copying and scanning Wikimedia communications.

36. Combined whitelisting and blacklisting to selectively acquire web communications: I previously observed that if it wished the NSA could, as a technological matter, simultaneously whitelist the IP addresses of particular websites, webmail services, and/or chatrooms of interest while blacklisting all other HTTP and HTTPS traffic, and thus obtain access to web communications of interest without necessarily copying and scanning Wikimedia's. Second Decl. ¶¶ 35, 36(b), 37. In response, Mr. Bradner remarks that "many" websites and other web-based services are now making use of content distribution networks ("CDNs"), which can have different and changing IP addresses around the world. Bradner Reply Decl. ¶ 84. Yet again, the extent to which websites, webmail services, and chatrooms of interest to the NSA (if any) are using CDNs for their communications, and the extent to which their use of CDNs would make it difficult or unworkable for the NSA to use whitelists to track its targets' communications, are matters known to the NSA, but about which Mr. Bradner has no information, and can only speculate.

37. Whitelisting by protocol: Mr. Bradner also points out (i) that if the NSA used protocol-based whitelisting (as opposed to whitelisting by IP addresses), it could miss communications that it ideally it might want to review, Bradner Reply Decl. ¶ 72, and (ii) that the NSA could not in fact be using whitelists to exclude all communications using web protocols (HTTP and HTTPS) while at the same time collecting at least some web communications, *id.* ¶ 73. These are points that Mr. Bradner has raised before, Bradner Decl. ¶ 366(f), (g), and that I addressed in my second declaration, Second Decl. ¶¶ 34-35, 36(b), 37. For the reasons already explained in my second declaration, these observations by Mr. Bradner supply no basis in Internet technology or engineering for concluding that whitelisting by the NSA would be implausible.

38. I do not mean to suggest that developing and maintaining whitelists for purposes of NSA Upstream collection would necessarily be "easy," or that "the NSA could get by with a very simple set of whitelist rules." *See* Bradner Reply Decl. ¶¶ 70, 85, 88. But neither I nor Mr. Bradner can know how difficult or easy it would be without far more detailed information about the number, nature, and communications habits of the NSA's Upstream targets. I do not

presume, as does Mr. Bradner, that the NSA would be incapable of developing and maintaining sufficiently reliable whitelists to meet its intelligence needs. Mr. Bradner has presented no basis in Internet technology or engineering for so assuming.

Blacklisting by IP Address or Protocol

39. Mr. Bradner states that "[t]here are multiple reasons" why the use of blacklists to filter out communications before they are copied or scanned in the Upstream process "would be incompatible with the public descriptions of the NSA's upstream collection program," Bradner Reply Decl. ¶ 92, but then proceeds to offer none. Instead he contends (i) that blacklisting Wikimedia IP addresses so as to prevent copying and scanning Wikimedia communications is "improbable," and (ii) that blacklisting Wikimedia IP addresses would not "guarantee" that Wikimedia communications are not copied and scanned during the Upstream collection process. *Id.* ¶¶ 93-101. Both are points that Mr. Bradner has raised before, that I have already addressed, and that still do not constitute non-speculative, technical grounds for deeming it improbable or implausible that the NSA might blacklist Wikimedia communications.

40. Blacklisting high-volume websites, including Wikimedia's: Following the observations in my first declaration that the NSA could blacklist Wikimedia IP addresses to prevent communications to and from Wikimedia from being copied and scanned, First Decl. ¶¶ 78-87, Mr. Bradner responded that he found it "basically inconceivable" and "totally unbelievable" that the NSA would sift through millions of websites to decide which to monitor and which not. Bradner Decl. ¶ 367(a). I then explained, in my second declaration, that I had no such extreme measure in mind, but rather, the trivial task of creating a blacklist of numerous high-volume popular websites of potentially low interest to the NSA, such as Wikimedia's, to eliminate unwanted volumes of communications that would otherwise have to be processed. Second Decl. ¶¶ 39-41.

41. Now Mr. Bradner takes the opposite tack, arguing that it is "very unlikely" that the NSA would decide to "specifically blacklist Wikimedia communications to reduce the load" on its collection apparatus, in light of the relatively small percentage of inter-regional Internet capacity

that Wikimedia communications traffic represents. Bradner Reply Decl. ¶ 96. I note first that a more enlightening statistic would be the percentage of the NSA's processing capacity that Wikimedia traffic represents, but Mr. Bradner has no way of knowing that information any more than he could know, rather than speculate, whether the NSA might find reason to blacklist Wikimedia communications or not. But it is also important to note that Mr. Bradner has once again missed the point of my observations. I did not suggest, in my second declaration, a scenario in which the NSA specifically singles out Wikimedia sites for blacklisting. I pointed out that Wikimedia's websites would naturally fall on any blacklist of high-volume, popular websites, of perhaps low interest to the NSA, that the NSA might assemble in order reduce (potentially by as much as 90 percent or more) the technological, logistical, and financial burdens of processing large volumes of unwanted web traffic. Second Decl. ¶ 41. Whether or not the NSA actually does so I do not know, but Mr. Bradner offers no technological basis for dismissing the possibility as improbable.

42. Hypothetical copying and scanning of blacklisted Wikimedia communications: Second, Mr. Bradner again draws attention to three hypothetical scenarios in which communications to or from Wikimedia would be copied and scanned during the Upstream collection process, even if the NSA had blacklisted communications containing Wikimedia IP addresses. Bradner Reply Decl. ¶¶ 97-101; *see* Bradner Decl. ¶ 367(a). These involved transmission of a Wikimedia communication within a multi-communication transaction (MCT) across an international Internet link, transmission of an email to Wikimedia from abroad using a U.S.-based email service, and visits to Wikimedia websites from abroad using a U.S.-based VPN service. I explained in my prior declaration that in each of these scenarios four (or in the third scenario, five) conditions would have to be met before a communication to or from Wikimedia would be copied or scanned, in each case rendering that possibility a matter of speculation. Second Decl. ¶¶ 77-85.

43. Regarding the initial three (or four) conditions required before each scenario could come to pass, Mr. Bradner summarily asserts that these conditions "would likely be frequently

met," yet offers no supporting data or explanation to demonstrate why that would be so. Instead, he focuses his attention principally on the last condition that would have to be met in each scenario before a Wikimedia communication could be copied or scanned—that the communication not be blacklisted for other reasons. Mr. Bradner dismisses as "far-fetched" the possibility that communications of the kind he posited in his three scenarios would be blacklisted, based on conjecture about the value that the NSA might attach to them. Bradner Reply Decl. ¶¶ 99-101. But he does not address the reasons given in my prior declaration to expect that the communications he describes likely would be encrypted, and perhaps blacklisted by the NSA, therefore, if it lacked the ability to decipher them. Second Decl. ¶¶ 80, 82, 84. In any event, if the initial three (or four) conditions in each scenario are not met, whether or not the final condition is met would be of no consequence. And as noted above, Mr. Bradner gives no reason to expect that the occurrence of the initial three (or four) conditions in each scenario would be anything but speculative.

Additional Points Concerning Whitelisting and Blacklisting

44. Mr. Bradner completes his discussion of whitelisting, blacklisting, and the filter-then-copy-and-scan approach generally with additional points, most already made in his prior declaration that I will now address.

45. "Blind spots": Mr. Bradner returns to his earlier remark that if the NSA blacklisted particular types of communications by port or protocol number, then doing so would leave "blind spots" in its collection that "[s]ophisticated" targets could "easily probe" to discover and evade collection of their communications. Bradner Reply Decl. ¶¶ 103-07; *see* Bradner Decl. ¶ 366(b), (e). On this point I observed previously that Mr. Bradner had not explained what targets could "probe," or how, to discover these so-called blind spots, the level of sophistication required, or on what basis he presumed that the NSA's Upstream targets possess the needed sophistication. Second Decl. ¶ 32. At bottom, I further observed, Mr. Bradner could only speculate whether the creation of "blind spots" would be of such genuine concern to the NSA as to dissuade it from utilizing whitelisting or blacklisting techniques. *Id.* ¶ 33.

46. In his second declaration, Mr. Bradner adds to his earlier conjecture by supposing that unspecified foreign intelligence services that may or may not be Upstream targets could test a protocol (type of communication) they suspect the NSA is not monitoring by communicating "actionable" intelligence ("such as the identity of a foreign agent") over that protocol and then observing whether any responsive action is taken. Bradner Reply Decl. ¶ 103. He vaguely suggests that whitelists or blacklists shared with telecommunication service providers could be at unspecified risk of "hacking," without considering additional security measures that could be taken to mitigate those risks. *Id*. ¶ 105. Typically, however, carriers use segregated networks to manage their routers and switches and encrypt network management information. There have been no indications that I am aware of that such carrier management networks have been breached. In short, neither of these suggestions constitutes a non-speculative basis in Internet technology or engineering to support Mr. Bradner's conclusions. And he overlooks the fundamental point I made previously, that only the NSA knows whether it considers these to be genuine risks that it would be unprepared to take in order to implement a filter-then-copy-and-scan approach to Upstream collection. Second Decl. ¶ 33.

47. "About" communications: In his second declaration Mr. Bradner suggests for the first time that whitelisting or blacklisting would be "entirely inconsistent" with the NSA's now-discontinued acquisition of "about" communications. Bradner Reply Decl. ¶ 108; *see also id.* ¶ 78. (As I noted in my second declaration, this was an argument advanced by Wikimedia in its legal brief, but not by Mr. Bradner in his first declaration. Second Decl. ¶ 49.) The position now taken by Mr. Bradner is based on a complete misunderstanding of "about" collection as described in the PCLOB Section 702 report.

48. Mr. Bradner acknowledges, as I have explained, that even if the NSA were to employ whitelists and/or blacklists to implement a filter-then-copy-and-scan approach to Upstream collection, it would still be possible to acquire at least some "about" communications, that is, communications neither to nor from a target but which refer to a selector (e.g., an email address) associated with the target. *See* Second Decl. ¶ 51; Bradner Reply Decl. ¶ 111. Mr.

Bradner's point is that the use of whitelists or blacklists would be incompatible with the *comprehensive* collection of "about" communications. For example, he states (twice) that developing and maintaining whitelists for acquisition of "about" communications "would be impossible to do . . . for a program *meant to capture* the communications of unknown non-targets about targets." Bradner Reply Decl. ¶¶ 69-70 (emphasis mine). In the same vein, he later reasons that "to set up [a] whitelist filter the NSA would have to know in advance which non-targets' IP addresses to whitelist … *in order to find the 'about' communications.*" Bradner Reply Decl. ¶ 110 (emphasis mine). He ends with the conclusion that, using a filter-then-copy-and-scan model, "the only way the NSA *could reliably capture* about communications would be to whitelist all non-wholly domestic communications." Bradner Reply Decl. ¶ 112 (emphasis mine).

49. The premise of Mr. Bradner's argument, however, is incorrect. At least as publicly described by the PCLOB Section 702 Report, in the very passage reproduced by Mr. Bradner in his declaration, the NSA's collection of "about" communications was

> an inevitable *byproduct* of the government's efforts to comprehensively acquire communications that are sent to or from its targets.

Bradner Reply Decl. ¶ 48 (quoting PCLOB Section 702 Report at 10) (emphasis mine). That is to say, while the goal of Upstream collection as described by the PCLOB is to comprehensively acquire communications sent "to or from" the NSA's foreign-intelligence targets, the program was not likewise "meant to capture," Bradner Reply Decl. ¶¶ 69-70, "reliably" find, *id.* ¶¶ 110, 112, or otherwise "comprehensively acquire" "about" communications. *See also* PCLOB Report at 84-86, 123, 124, 144, 145. The incidental collection of only some "about" communications, even if the NSA were using a filter-then-copy-and-scan configuration as described in my second declaration, would be entirely consistent with the PCLOB's description of "about" collection as a byproduct rather than an objective of Upstream collection. For example, if a whitelist included a specific email server, the NSA could scan for targeted email addresses in the whitelisted communications, and would, as described in the PCLOB Section 702 Report, occasionally acquire "about" communications that referred to the targeted email address in addition to

communications to or from the email address. Thus, the "about" communication would be captured because of the way the scanning is implemented. And being incidental, "about" collection would not have required advance knowledge of all non-targets' IP addresses.

50. Collection of web communications, including encrypted HTTPS communications: Mr. Bradner again addresses the reasons why he believes the NSA is collecting web (i.e., HTTP and/or HTTPS) communications, Bradner Reply Decl. ¶¶ 130-36, 154(3)(c)(i); *see* Bradner Decl. ¶¶ 314-15m 366(f), focusing in particular on his earlier opinions (i) that blacklisting the HTTP and HTTPS protocols to prevent the copying and scanning of all web communications "would leave a very large hole in the NSA's coverage." Bradner Reply Decl. ¶ 135; *see* Bradner Decl. ¶ 366(f), (g), and (ii) that encrypted communications collected by the NSA (if any) include, specifically, HTTPS communications, Bradner Reply Decl. ¶¶ 137-39; *see* Bradner Decl. ¶¶ 325, 366(g).[4] Mr. Bradner largely ignores the observation, in my second declaration, that the NSA could use a combined whitelisting/blacklisting technique to block NSA access to all HTTP and HTTPS communications except those to or from IP addresses included on a whitelist containing the addresses of websites, chatrooms, and/or webmail services of intelligence interest. In this way the NSA could obtain access to HTTP and HTTPS communications of interest, while excluding all others, including, hypothetically, Wikimedia's. Second Decl. ¶¶ 35, 36(b), 37.

51. Mr. Bradner opines that this technique "would also leave very large holes in the NSA's coverage," Bradner Reply Decl. ¶ 136; *see also id.* ¶ 139, but whether the "holes" would be so unacceptably large as to motivate the NSA to copy and scan all HTTP and HTTPS communications (including, therefore, Wikimedia's) is a matter implicating the NSA's surveillance priorities, resources, and capabilities, about which Mr. Bradner has no information and can only

[4] I note in passing that Mr. Bradner expresses skepticism at my suggestion that the NSA could collect encrypted communications under its Section 702 authority using its PRISM acquisition method, as opposed to Upstream collection, because providers assisting in PRISM collection "will frequently have direct access to the user's unencrypted communications." Bradner Reply Decl. ¶ 139. But numerous online file-storage and webmail services now store their users' files and/or messages in encrypted formats, either by default or upon request.

speculate. Mr. Bradner also suggests that combined whitelisting and blacklisting of HTTP and HTTPS communications "would also be contrary to the aim of the 'about' collection program." Bradner Reply Decl. ¶ 136. As I discussed, however, in paragraph 49, above, there is no evidence that the NSA ever engaged in an "'about' collection program." The PCLOB Section 702 Report does not state that "about" collection was ever an "aim" of Upstream collection, merely the "byproduct" of efforts to collect communications to and from designated targets.

52. Relative complexity of the copy-all-then-scan and filter-then-copy-and-scan approaches: Mr. Bradner describes his copy-all-then-scan approach as a "simpl[er], mo[re] reliable, and easi[er] to operate architecture" than a filter-then-copy-and-scan configuration, Bradner Reply Decl. ¶ 116, because in his view, "the need to constantly reconfigure the [provider's router or switch] with updated blacklists and whitelists would create the risk of misconfiguration or overloading." *Id.* ¶ 118; *see also id*. ¶ 124.

53. To begin, it is not unusual for a provider to reconfigure its routers and switches on a routine basis to meet the evolving data transmission needs of a dynamic commercial customer base. Whether it would require more frequent reconfiguration to update an NSA whitelist or blacklist is a matter of conjecture that would depend on how frequently the NSA requested such updates, based on how often it makes additions to or deletions from its target list, and how often its targets change their modes of communication. It is also a matter of speculation, as I have observed previously, whether the NSA's targets (or, more precisely, their associated IP addresses), are so numerous that loading a whitelist would run the risk of overloading a router's or switch's processing capacity. Second Decl. ¶ 25. These are all matters about which Mr. Bradner apparently has no information. Moreover, carriers usually implement change-management protocols in which new router configurations are tested in laboratory settings before they are loaded, in order to mitigate any risk of misconfiguration or overloading.

54. The question Mr. Bradner also leaves unanswered is whether the risks he cites would outweigh the daunting technical and logistical difficulties and financial burdens that would complicate implementation of his "simpl[er]" and "easi[er]" approach, as discussed in paragraphs

20-21 of my second declaration. Mr. Bradner does not take issue with my description of these hurdles (with the exception of the marginal observation that an opto-electronic device is not an "esoteric" piece of equipment). Bradner Reply Decl. ¶ 119. When considered from a broader perspective, the asserted simplicity of Mr. Bradner's copy-all-then-scan configuration as compared to a filter-then-copy-and-scan approach becomes less apparent, to say the least, and which way the scales tip in the NSA's eyes is a matter about which Mr. Bradner, once again, can only speculate.

55. Sharing sensitive information with a provider: Mr. Bradner also repeats a point made in his prior declaration that, "in [his] opinion," the NSA would not want to implement a filter-then-copy-and-scan approach because it would require sharing with provider personnel the sensitive information about NSA targets and the scope of its surveillance that would be contained in whitelists and blacklists. Bradner Reply Decl. ¶ 126. This is a point I have already addressed. Second Decl. ¶ 18. Mr. Bradner acknowledges, as I observed previously, that the NSA shares information about its targets with Internet service providers in order to conduct PRISM collection, but suggests that in his estimation the information that would have to be shared with a provider in a whitelist or blacklist is even more sensitive. Bradner Reply Decl. ¶ 126. Whether or not that is so (a matter about which Mr. Bradner has no apparent expertise), it remains the case that the extent to which the NSA would be willing (or find it necessary) to share classified information with an assisting provider in order to conduct Upstream surveillance is a matter about which Mr. Bradner has no specialized knowledge or information.

56. Proposed "channel mirroring": Mr. Bradner also makes a passing reference to a configuration not previously proposed by him, in which a provider would configure its router or switch to copy all communications just on particular channels (circuits), designated by the NSA, that cross a monitored link. Bradner Reply Decl. ¶ 121. He describes that configuration, too, as very simple, static, and involving fewer disclosures of sensitive information to non-government personnel, *id.*, but it is not so simple a picture as Mr. Bradner paints.

57. In brief, due to carrier practices including link aggregation, among others, it is unclear how the NSA could know which optical channels would be carrying the communications of potential targets, at least without a deep understanding of the assisting carrier's routing architecture and configuration. In addition, because of a phenomenon referred to as traffic failover, traffic can move from one logical link to another, so this configuration is unlikely to be stable and unvarying. As a result, the NSA would have to convey to the carrier which IP addresses it would like to monitor, so that the carrier can map these addresses onto logical links and their optical channels. Thus, the NSA would have to convey just as much information to the carrier as with a filter-then-copy-and-scan configuration. Furthermore, the mirroring capability of at least some common routers and switches is limited to a small number of interfaces. Thus, only a small fraction of the router input or output ports could be monitored at any time if all the traffic, unfiltered, is to be copied to these ports.

58. U.K. Section 8(4) collection: Mr. Bradner returns to this subject to prove a point that I did not take great issue with in my second declaration: that the brief filed by the U.K. government in the European Court of Human Rights describing its "Section 8(4)" collection program includes references to "intercept[ing]"--which could be taken to mean copying—"the entire contents of a [circuit]" before the communications stream is filtered, and the remainder then scanned for targets' communications. Bradner Reply Decl. ¶ 140-48; *see* Second Decl. ¶¶ 61-63.[5] Notably, Mr. Bradner does not take issue with the conclusions I previously drew from the U.K. governments' filings: (i) that the U.K. government documents describe a process of filtering to winnow out communications deemed to lack significant intelligence value before communications are scanned for targets' selectors, Second Decl. ¶ 62, and (ii) that even if all communications on a circuit were copied first, the copying and initial filtering could be conducted

[5] Oddly, Mr. Bradner appears to find fault that I based my discussion of the Section 8(4) collection program on one of the same documents he relied on his declaration (the U.K. government brief), and that I failed to cite a document, *not* appended to his first declaration, that he himself did not cite previously, and that he himself now cites only for the first time in his reply declaration. *See* Bradner Reply Decl. ¶ 144; Second Decl. ¶ 61 (citing Bradner Decl., App'x EE).

by the service provider, so that only those communications meeting the filter criteria, rather than all communications, would pass into the government's (whether the GCHQ's, or hypothetically, the NSA's) control. Second Decl. ¶ 64.[6]

THE SCENARIOS ENVISIONED BY MR. BRADNER IN WHICH AT LEAST SOME WIKIMEDIA COMMUNICATIONS WOULD BE COPIED AND SCANNED, EVEN IF THE NSA EMPLOYED TRAFFIC-MIRRORING TECHNIQUES, ARE SPECULATIVE AND CONJECTURAL

59. Finally, I address Mr. Bradner's repeated contention that the filter-then-copy-and-scan configuration using whitelists and/or blacklists would not "guarantee" that the NSA avoids all interaction with Wikimedia communications during the Upstream collection process. Bradner Reply Decl. ¶ 57; *see also id.* ¶¶ 30, 64, 97-101, 115, 154(5). In support of this contention, Mr. Bradner identifies a number of hypothetical scenarios in which Wikimedia communications would be copied and scanned if all of the necessary conditions were met for these scenarios to come to pass. I have already discussed most of the scenarios outlined by Mr. Bradner above, and in my second declaration. I address the scenarios newly conceived of by Mr. Bradner below. Whether any of them has occurred or would ever come to pass is a matter of speculation for which Mr. Bradner gives no evidence.

60. Whitelisting by IP address: First, Mr. Bradner points to a situation in which a user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link (hypothetically) monitored by the NSA. Bradner Reply Decl. ¶ 154(5)(a)(i); *see also id.* ¶ 154(5)(b)(iii). I acknowledged in both my first and second declarations the theoretical possibility that Wikimedia communications of this kind could be copied and scanned during Upstream surveillance, even if the NSA used a whitelisting technique

[6] Mr. Bradner wonders why I also addressed the U.S. Government's cyber-defense system known as Einstein 2.0, when he mentioned it only "in passing" in his first declaration. Bradner Reply Decl. ¶ 150. As I explained in my second declaration, I addressed the implications of Einstein 2.0 because Wikimedia, in its legal brief, attempted to rely on Einstein 2.0 as "corroboration" for Mr. Bradner's conclusions, even though Mr. Bradner himself had not done so, Second Decl. ¶ 66, and does not do so now. For all intents and purpose, Mr. Bradner now disregards my explanation of the reasons why conclusions about the Upstream collection process cannot be drawn from Einstein 2.0. *Compare* Second Decl. ¶¶ 68-69 to Bradner Reply Decl. ¶¶ 151-53.

that otherwise excluded Wikimedia's communications. Second Decl. ¶ 43; First Decl. ¶ 81. But this scenario, like the blacklisting scenarios posited by Mr. Bradner, is conjectural. *See* paragraphs 42-43, above; Second Decl. ¶¶ 78-85. Mr. Bradner cites no evidence of the number or geographic locations of persons using whitelisted IP addresses who communicate with Wikimedia, and of course could not do so without knowing the composition of whitelists (hypothetically) employed by the NSA. There is no basis, therefore, on which to conclude that communications between Wikimedia and persons using whitelisted IP addresses would cross every international Internet link to and from the United States (as Wikimedia claims of its communications generally), or, for that matter, that they would cross one or more links (if any) that happen to be monitored by the NSA. In addition, the communications in question must not themselves be blacklisted, as might be the case if they were encrypted, *see* Second Decl. ¶¶ 80, 82, 84, rendering this scenario even more uncertain.

61. Blacklisting by IP address: Mr. Bradner next mentions three scenarios in which he states that blacklisting by IP address would not "guarantee" that the NSA would avoid all interaction with Wikimedia communications. Bradner Reply Decl. ¶ 154(5)(b). The first of these, involving the enclosure of a Wikimedia communication within an MCT, Bradner Reply Decl. ¶ 154(5)(b)(i), is the same as the first of the three scenarios he posited in his first declaration, Bradner Decl. ¶ 367(b)(1), which I have already discussed in my second declaration, Second Decl. ¶¶ 78-80, and again in paragraphs 42-43, above. The second scenario, involving the passage of a Wikimedia communication through an intermediary service such as a VPN, or an email server, Bradner Reply Decl. ¶ 154(5)(b)(ii), is a generic restatement of the second and third scenarios hypothesized in Mr. Bradner's first declaration, Bradner Decl. ¶ 367(b)(2)-(3), which I discussed in my second declaration, Second Decl. ¶¶ 81-84, and again above in paragraphs 42-43. The third blacklisting scenario suggested in Mr. Bradner's second declaration, Bradner Reply Decl. ¶ 154(5)(b(iii), is a repeat of the whitelisting scenario he referred to in paragraph 154(5)(a)(i) of his second declaration, and that I addressed in paragraph 60, above.

62. Port or protocol blacklisting: The last two scenarios hypothesized by Mr. Bradner concern blacklisting by port or protocol (type of communication) rather than by IP address. Bradner Reply Decl. ¶ 154(5)(c). He envisions two situations in which a Wikimedia communication, even if using a blacklisted protocol, would still be copied and scanned during Upstream surveillance if it crossed a monitored link and either (i) it were enclosed in an MCT using a different protocol, one not blacklisted by the NSA, or (ii) it passed through an intermediary (such as an email server) that, likewise, used a different protocol not blacklisted by the NSA.

63. Mr. Bradner gives no examples of either kind of supposed communication, in which a communication using one protocol is transported within another communication using a different protocol. Nor does he give evidence of how frequently such supposed communications could be expected to occur, or under what circumstances. Consequently, the occurrence of such a communication in the first place is itself a matter of speculation. An even greater degree of speculation would then be required to imagine (i) that the protocol used by the enclosing MCT or the intermediary service is not blacklisted by the NSA, (ii) that the IP addresses assigned to the communication within the MCT, or by the intermediary service, are not excluded to due whitelisting, and (iii) that the communication happened to cross an international Internet link monitored by the NSA (if any).

64. For all of the reasons I have explained herein (paragraphs 42-43, 60-63, above), and in my second declaration, Second Decl. ¶¶ 78-85, whether and when any of the scenarios envisioned by Mr. Bradner might come to pass at a particular international Internet link that happened to be monitored by the NSA (if any), such that the NSA would copy and scan communications of Wikimedia's, is a matter of speculation.

CONCLUSION

65. For the reasons I discuss above and in my first two declarations, it remains my opinion that, based on what is publicly known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a

number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications.

66. While I offer no opinion on the likelihood that the NSA does or does not, in fact, employ these techniques, I have previously examined, and now re-examined, the bases of Mr. Bradner's opinions (i) that the NSA, in conducting Upstream surveillance, "most likely" copies, reassembles, and scans for selectors all communications packets traversing an international Internet link that is monitored by the NSA (if any); (ii) that it is "implausible" that the NSA uses the traffic-mirroring techniques (white- and blacklisting) described in my first declaration; and (iii) that even if the NSA uses one or more of the techniques I described, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. I still conclude that these opinions lack a non-speculative foundation in Internet technology and engineering.

67. My opinions are unaltered by the statements referred to by Mr. Bradner in the FISC's October 2011 opinion and the PCLOB Section 702 Report. As I explained above, the use of traffic-mirroring techniques to implement a filter-then-copy-and-scan approach to Upstream collection would be entirely consistent with both statements. My opinions are unaltered, as well, by the so-called "technical and practical necessities" discussed by Mr. Bradner in his reply declaration. With few exceptions, they are simply reiterations of the same grounds given for the conclusions reached in his second declaration. They, too, are principally based on speculation about the NSA's surveillance practices and priorities, its capabilities and resources, and the number, nature, and communications practices of its Upstream surveillance targets, and lack a non-speculative foundation in Internet technology and engineering.

I declare of penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Executed in New York, New York on March 22, 2019.

HENNING G. SCHULZRINNE

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 17

SECOND DECLARATION OF DR. ALAN J. SALZBERG

Dr. Alan Salzberg, for his second declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

I. Introduction

1. I am the Principal (and owner) of Salt Hill Statistical Consulting. I previously submitted a declaration in this case, dated February 14, 2019 ("Salzberg February Declaration"). My February declaration commented on the "Declaration of Jonathon Penney" ("Penney December Declaration"), which was submitted in December 2018. I submit this second declaration at the request of the United States Department of Justice in response to the "Reply Declaration of Jonathon Penney" ("Penney Reply"), which was submitted on March 8, 2019. I have previously submitted my resume describing my background and qualifications in statistical sampling, analysis, and review for government and industry, as well as information regarding prior testimony and fees.

2. This report proceeds as follows. In the next section, I summarize my findings. In the third section, I detail those findings. In the fourth section, I set forth my conclusions. Finally, I have included an appendix with a program log showing the results of additional analyses.

II. Summary of Findings

3. In my February Declaration I addressed the deeply flawed model presented by Dr. Jonathon Penney in his December Declaration. Specifically, in summary I previously found as follows:[1]

   A. "The methodology used in the Penney Declaration—which purportedly shows an upward trend in page views of certain articles posted on Wikipedia through May 2013, followed by an abrupt drop and downward trend in views of those articles beginning in June 2013—is deeply flawed, inappropriate, and likely biased."

   B. "The Penney Model simply assumes that a single change occurred in June 2013, rather than letting the data identify the timing and number of changes in trends that occurred. Even though there is no consistent trend in the data, the design of the Penney Model will create the appearance that the data contain just one inflection point. And, because of its design—even though changes in trend occurred before these June 2013 disclosures—the Penney Model will find that the disclosures caused them."

   C. "Contrary to the hypothesis presented in the Penney Declaration, analysis of page views for the 48 individual articles in the privacy-sensitive group do not show a rising trend followed by an immediate and sustained drop in June 2013."

   D. "With the one exception of removing the article on Hamas, the Penney Declaration does no analysis or adjustment for factors (such as world events) affecting these individual article page views. Instead, the Penney Declaration inappropriately aggregates the vastly different page view data for individual articles, with the result that these individual differences in page views are masked."

[1] Salzberg February Declaration, paragraph 4.

E. "Even at that aggregate level, I find that the hypothesized peak in page views of "privacy-sensitive" articles in May 2013 does not exist, and the hypothesized upward and then downward trends in views of privacy-sensitive articles before and after June 2013, respectively, do not exist."

F. "Extended data through 2018 regarding page views of the privacy-sensitive articles do not indicate a long-term decline in page views from pre-June 2013 levels."

G. "A proper control dataset would exhibit similar page view behavior prior to June 2013. The comparison datasets used in the Penney Declaration do not and are thus inappropriate controls."

H. "The Penney Declaration analysis ends in July 2014. No data are presented that shed any light on whether page views at the time the Amended Complaint was filed in 2015 (or thereafter) were affected by Upstream. In other words, even if the purported effect and trends were a correct conclusion for the data examined (and they are not), the Penney Declaration analysis does not and cannot show that the effect continued years after the study ended."

I. "Even if a chilling effect occurred in June 2013, there are no data analyzed in the Penney Declaration that show any effect was due specifically to "public awareness of" the specific NSA surveillance program challenged here (known as Upstream surveillance) rather than possible inaccuracies, if any, about the program reported in the press, disclosures about other NSA programs, disclosures about other surveillance programs (e.g., surveillance by Britain), or other, unrelated events of June 2013."

4. As discussed in detail below, the Penney Reply does not raise any valid critiques of my original findings, and the additional analyses in the Penney Reply do not bolster the flawed model presented in the Penney December Declaration. In addition, the Penney Reply does not propose a new model that corrects the flawed model presented in the Penny December Declaration, and the slight modifications attempted do not address any of the issues I raised. Therefore, my findings and conclusions set forth in my February declaration remain unchanged.

III. Details of Findings

5. The Penney Reply begins with seven critiques of my analyses, in paragraphs 4 through 23 of the Penney Reply, and goes on to respond to my critiques in paragraphs 25-36. I reviewed all of the Penney Reply and in this declaration I organize my responses by topic, so as not to be repetitive. In particular, this section proceeds with the following six subsections:

A. Overview of the Incorrect Assumptions Made in the Penney Reply;

B. Spurious Statistical Conclusions from the Penney Model are Partly Due to Aggregation of the Article View Data;

C. The Penney Reply's Additional Analyses Fail to Address the Flaws in the Penney Model;

D. Data Beyond Time Period 2014 Show Article Views at About 2012 through 2014 Levels, Even When Earlier Data is Corrected for Mobile Views;;

E. Omitted Variable Bias of the Penney Model Cannot be Solved by Deleting Valid Data; and

F. The Penney Model's Failure to Isolate the Effect of Awareness of NSA's Upstream Program.

A. Overview of the Incorrect Assumptions Made in the Penney Reply.

6. Before responding to the specific claims of the Penney Reply, I first address some false assumptions the Penney Reply made regarding my critiques.

7. First, while my review of the disaggregated data provides an important, simplified explanation of many of the flaws of the Penney Model, the flaws of that model remain whether that model is applied to the aggregated or the disaggregated data. I am not suggesting that the application of the deeply flawed Penney Model to each of the 48 articles, individually, would be appropriate, nor am I suggesting that there could never be theoretical circumstances where the data could be aggregated without presenting the deeply flawed and misleading results that the Penney Model presented here.

8. Second, my February report provided no wholesale critique of the so-called ITS "Interrupted Time Series" designs or of regression models in general. My critiques instead relate to the particular methods Dr. Penney employed and the underlying data used in the Penney December Declaration.

9. Third, as I pointed out a number of times in my February Report,[2] I do not present an alternate model of page views, but I do use a number of examples and perform analyses that demonstrate the flaws in the Penney Model. Statements in the Penney Reply regarding "alternatives" that I suggest are therefore misleading.

10. Fourth, while the Penney Reply is correct in that much of my analysis uses "visual inspection" as an aid to understanding the issues with the Penney Model, I also perform statistical tests and point out many specific flaws in the Penney December models. As with the issue of aggregation and disaggregation, I am not advocating one or the other, but rather, doing both. A simple visual review of the data using charts and graphs, such as I the one performed, has long been considered a fundamental component to developing correct statistical models.

B. Spurious Statistical Conclusions from the Penney Model are Partly Due to Aggregation of the Article View Data

11. As I stated in my original declaration, a review of the disaggregated data leads to the conclusion that there is no May 2013 peak or steep drop beginning in June 2013, contrary to the Penney December Declaration's conclusion.[3] The aggregated data do not show a May 2013 peak either, but rather an April 2013 peak,[4] followed by declines beginning in May 2013. This means that both the disaggregated data and the aggregated data show that the drop in page views begins *before* the June 2013 disclosures.

[2] In my February Declaration, for example, in paragraphs 52 (note 31) and 53, I explicitly state that I am not proposing an alternative model.
[3] Salzberg February Declaration, paragraphs 11-17.
[4] As stated, while the Boston Marathon bombings are one possible reason for an April peak, such a simple model is far from adequate for many of the same reasons that the Penney Model is inadequate, but at least in the April peak model, the drop in page views comes after the purported cause and not before the purported cause.

12. The Penney Reply responds to this critique by erroneously claiming that my disaggregated review should be disregarded because aggregation is appropriate here. This somewhat misses the point, which is that neither the disaggregated data nor the aggregated data support Dr. Penney's conclusions. Nonetheless, I reviewed the literature that the Penney Reply cites in support of its claim. The literature cited does not support the Penney Reply's claim.[5] The articles cited are general articles on ITS designs rather than articles espousing aggregation.[6] Moreover, the four reasons cited in the Penney Reply paragraph four are not specific to aggregated data (they apply equally to disaggregated data).

13. In paragraph 26b, the Penney Reply acknowledges that there is "there is no single determinative method or factor to decide whether an aggregated or disaggregated analysis of data is appropriate." As I stated, by aggregating all the data prior to analysis, there is no possibility of correcting for any article-specific differences in the data or exploring whether there are important differences among article views that need to be accounted for in any model. This leads to a biased model and erroneous claims of statistical significance where, as here, such differences exist.[7] The idea of not reviewing and understanding the disaggregated data, and discarding information by inappropriately aggregating that data, is anathema to scientists, because such ignorance often leads to false conclusions. That review typically includes graphical analysis, because, as one statistician put it: "[g]raphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations."[8]

14. The Penney Reply argues that that my use of simple graphs to provide a visual inspection of the disaggregated data should be disregarded, in part, because a "visual inspection of data . . . can often be misleading," a point he makes with a quotation of one of the great proponents of graphical analysis, Dr. Howard Wainer.[9] Dr. Wainer, however, is not saying that graphs should not be used; he is only saying to be careful that they are not used in a misleading manner.[10] Ironically, by ignoring the disaggregated data and aggregating dissimilar page views to tell a

---

[5] Specifically, in the footnotes for paragraphs 4-5, the Penney Reply identifies several sources that Dr. Penney claims supports his use of aggregation in this circumstance. The only citation that even appears to support aggregation, in this type of situation, is not from a paper or textbook but from a PowerPoint presentation by Emma Beard which appears to have been presented at a conference in London (*see* footnotes 3, 4, and 21 in the Penney Reply). I reviewed the PowerPoint presentation and it presents no reasoning or data to support the claims (nor is it obvious that the author even made such claims regarding a simple regression model like the one in the Penney December Declaration). Additionally, unlike a scholarly article, a PowerPoint presented at a conference is typically accompanied by an oral portion of the presentation that may provide additional context or present the point differently than the language on the printed slides). In short, none of the cited source materials in the Penney Reply alter my conclusion that in this instance the use of aggregated data is inappropriate and misleading.

[6] The Penney Reply in paragraph 26e, takes issue with my terming the data "panel data" and not "time series" data. Panel data is a form of time series data, as the introduction to the text and chapter on panel data in my source make clear. See Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, 2012, South-Western Cengage Learning, p. 10 and 448.

[7] See Salzberg February Report, paragraphs 56-60, for example.

[8] Tufte, Edward R., The Visual Display of Quantitative Information, Graphics Press LLC, 2001, p. 13. Also, p. 9 of the same text states that: "Often the most effective way to describe, explore, and summarize a set of numbers – even a very large set – is to look at pictures of those numbers. Furthermore, of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful."

[9] Penney Reply, paragraph 3 footnote 1.

[10] Quote from Penney Reply, paragraph 3 footnote 1 (quoting Howard Wainer).

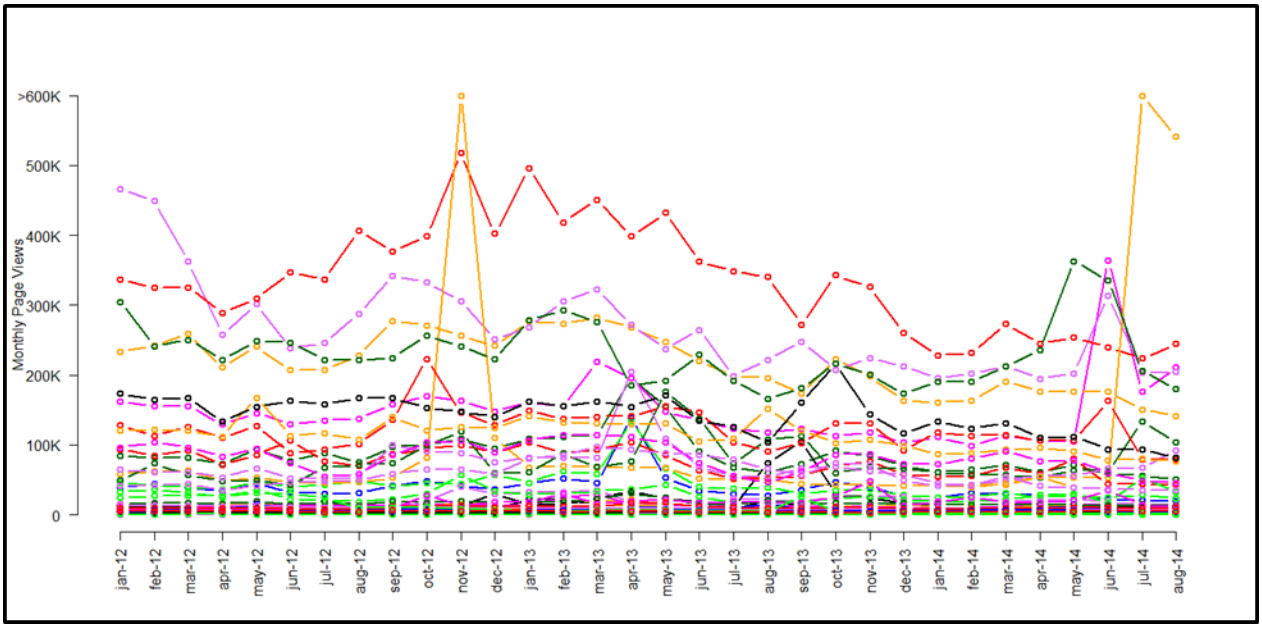
misleading story, Dr. Penney has created precisely the type of misleading graphs that Dr. Wainer is warning against. As Yale statistician Edward Tufte says: "[a]ggregations by area can sometimes mask and even distort the true story of the data"[11] and "[a]ggregations over time may also mask relevant detail and generate misleading signals."[12] Tufte concludes:"[i]f in doubt, graph the detailed underlying data to assess the effects of aggregation."[13] As I explained in my first Declaration (Paragraphs 18-26), Figure 2 of the Penney December Declaration is misleading because it inappropriately aggregates the data and shows a suggestive regression line, while obscuring the fact that the decline was not as indicated.[14]

15. I created and included (as Appendix IV to my first Declaration) graphs of each of the 48 articles' page views individually, so that all the data is available to view in a clear graphical form. I invited (and invite) review of each of those graphs. The only reasonable conclusion from a review of those graphs is that the effect supposedly found in the Penney December Declaration is spurious. I also included graphs of the aggregated data (*see* paragraphs 18-26 of my February Declaration), and those graphs also do not indicate a May 2013 peak. I did not leave out anything or "cherry-pick," contrary to what the Penney Reply states in paragraphs 11, 12, and 32(a).[15]

16. The Penney Reply claims that disaggregation adds "noise" to the data, "both visual and statistical," and points to my first graph showing all 48 articles in a single figure.[16] I showed all the data in a single figure (as well as in 48 separate figures in Appendix IV) because it provides important context and a comparison point to Dr. Penney's aggregated plot, which artificially smooths the differences.[17] As a reminder, my Figure that includes all 48 articles is below.

[11] Tufte, Edward R., Visual Explanations, Graphics Press LLC, 1997, p. 35.
[12] Ibid, p. 36.
[13] Ibid, p. 37.
[14] Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015, also cited in the Penney Reply, is also clear that graphical analysis is recommended.
[15] Penney Reply, paragraph 12.
[16] Penney Reply, paragraph 5.
[17] This method, of putting all the data into a single plot, is done in so-called spark graphs, examples of which can be found in Tufte, Edward R., Beautiful Evidence, Graphics Press LLC, 2006. p. 47-63.

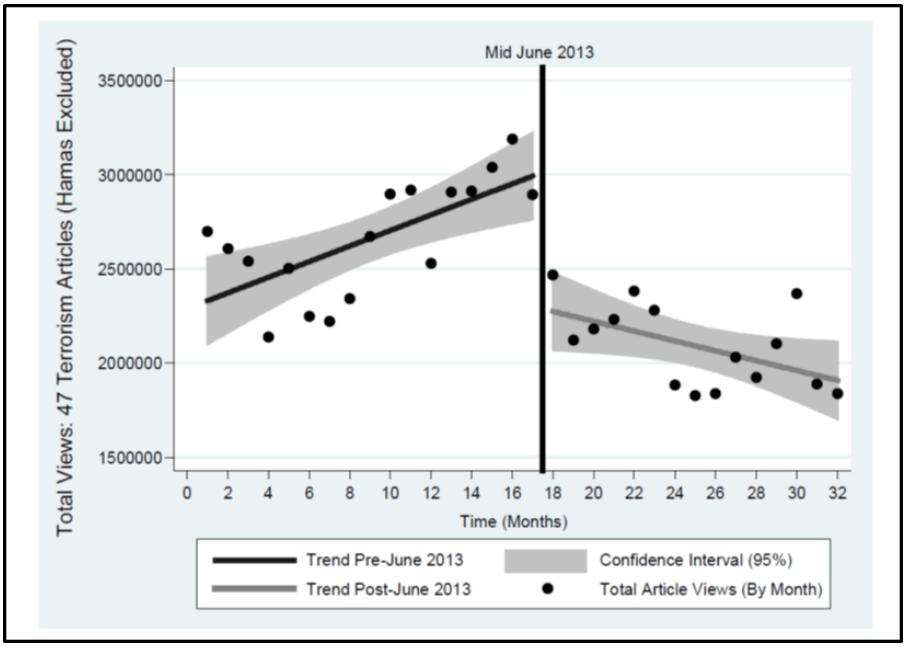
Figure 1: Page Views for all 48 Articles Considered to Suffer from a Chilling Effect beginning in June 2013



17. In contrast, Penney's Figure 2 from his December Declaration, shown below, misleadingly indicates a simple up and down movement that is belied by the individual data in Figure 1, above. The same is true for the confidence intervals drawn on Penney's Figure 2 graph itself, as I pointed out in my February declaration.[18]

[18] See Salzberg February Declaration, paragraph 20.

Figure 2: Penney December Declaration Aggregate Figure Masks Individual Differences



18. The contrast between the simple, disaggregated view and Penney's misleading aggregate view should have led Dr. Penney to question whether his aggregated model masks systematic differences at the article-level. As I stated above, I also provided each plot individually, so the reader can see what is behind the total picture of page-views shown in my first figure.

19. The Penney Reply is also wrong when it categorically states that disaggregation adds noise. The Penney Reply concerns that a disaggregated model will not allow for estimation of an "aggregate level inference about large scale NSA surveillance effects"[19] are misplaced. If the same naive model is run on both datasets, the estimated effect in the disaggregated model is exactly the same as the estimated effect in the aggregate model.[20] The statistical significance of these effects will also be the same if the disaggregation only adds noise to the model, and I show this fact through a simulation.[21]

20. However, if the disaggregated data reveal systematic differences in the data, in that the individual articles' page views do not tell the same or even a similar story as the aggregated data, then the naive model needs to be modified in order to avoid bias, whether run on aggregate or individual article data. To further support the analysis I already performed showing the model is over-simplified and perhaps mis-specified, I performed a statistical test to determine whether the

[19] Penney Reply, paragraph 26b suggests that because the question regards aggregate differences the aggregated data must be used.

[20] This fact is shown in the Appendix to this Declaration, and can be observed by noting the coefficient estimates for the Penney Model as shown in my Appendix. In the Appendix, I run the Penney Model on the averages and run the same model on the individual articles. The estimated effects (model coefficients) are exactly the same.

[21] I have included in the Appendix a simulation that shows the results of running the Penney Model on aggregated and disaggregated data are the same when the errors are statistical noise. This includes not only the regression coefficients (which will be the same whether the difference are due to noise or not, as explained above) but also the standard errors (i.e., the statistical significance) of those coefficients.

differences by article are mere noise or systematic.[22] I found, with high statistical significance, that the differences among articles are systematic (the statistical results are in the Appendix to this Declaration). This means that the model used in the Penney Reply is incorrect, *whether using the aggregated or the disaggregated data.* Only by accounting for the article-level, seasonal, and other differences can a valid model or set of models be produced. Furthermore, the model's estimates show increased error when calculated in disaggregated form.[23] This fact confirms my conclusions in my February report.[24] Because the differences are systematic and not mere "noise," the aggregation produces a result with inflated statistical significance.[25]

21. In reviewing some of the specific examples I cited to explain the fact that aggregating the data masks differences in the articles, the Penney Reply re-explains some analyses and runs additional models, but none address the issues I raised.[26] The Penney Reply presents Figures 2A and 2B, which purported show an "Increase until June 2013 and then a Sharp Drop-off."[27] This labeling is wrong. The increase is only through April, with a drop off in May and a continuation of that drop in June. This fact can be seen in Penney Reply's own Figures 2A, 2B, 3A, and 4 of the Penney Reply. Each shows an April and not a May peak, and a May and not a June start to the drop in page views. As I explain in my first Declaration, the fact that the drop in page views began *before* the June 2013 disclosures does not support Dr. Penney's conclusion that the June 2013 disclosures caused the drop in page views, and violates a basic tenet of causal models (i.e., a cause cannot occur after an effect).[28]

22. The models using the data in Figures 2, 3, and 4 of the Penney Reply suffer from the same problems as the original model in the Penney December Declaration.[29] The Penney Reply

[22] See Salzberg February Report, paragraphs 55-60 for my comments regarding the over-simplified model and omitted variable bias.

[23] As shown in my appendix, attached here, in some cases the claimed effects are not statistically significant. In other cases the statistical significance is weaker. These are further indications that the article differences are not mere noise. As my simulation (in the Appendix attached here) shows, when differences are based on mere noise, the statistical significance of the coefficients for the effects will remain unchanged when running the model on aggregated versus disaggregated data.

[24] See Salzberg February Report, paragraph 48-50 and 55-60.

[25] This is due to omitted variable bias, among other factors. I pointed this out in my February Report, paragraph 56. I do not attempt to correct for the omitted variable bias by adding additional variables, and therefore the disaggregated model is also incorrect.

[26] These re-analyses and the Penney Reply's commentary on them is found in Penney Reply, paragraphs 6-22 and paragraphs 26, 28, and 30.

[27] Penney Reply, Figures 2A and 2B.

[28] For two examples of such spurious inferences that ascribed a later cause to an earlier effect, see a source cited in the Penney Reply: McCleary, Richard, McDowall, David, and Bartos, Bradley J., Design and Analysis of Time Series Experiments, Oxford University Press, 2017. The examples are portrayed in this text in Figure 5.15 (explained on p. 214-215) and Figure 7.1 (explained on p. 275-276), and involve "interventions" and data with similarities to the data analyzed in the Penney December Declaration.

[29] The Penney Reply inexplicably discards its high-privacy group of 31 articles in favor of a new high privacy group of 23 articles for Figure 4 and some accompanying analyses. The Penney December Declaration already determined (perhaps also arbitrarily) a 31-article set that is highly privacy sensitive and this new set of 23 is a subset of those articles. Of course, re-running the same model on datasets that are nearly the same will produce results that are nearly the same, and proves nothing.

analysis ignores the large and obvious effect of events of April 2013 in its analysis of "improvised explosive device," "dirty bomb," "car bomb," and "ammonium nitrate."[30]

23. The only graph that the Penney Reply shows that appears to have a peak in May is Figure 3b (page views for so-called "normalized" Ammonium Nitrate), but that supposed "peak" is artificially created because the Penney Reply manipulated the graph to remove the April peak and replace it with the average of the March and May.[31] Removing such outliers and replacing them with averages in this way is against the practice of statisticians in general. Outlier handling is discussed in detail in an article the Penney Reply cites (at footnote 8), and this article says such adjustment is only appropriate for *error* outliers.[32] Here, the data points for Ammonium Nitrate page views are not errors and so removing the correct data point and replacing it with an average is inappropriate.[33]

C. The Penney Reply's Additional Analyses Fail to Address the Flaws in the Penney Model

24. Paragraphs 18 and 28 of the Penney Reply assert that no assumption is made in the Penney Model concerning a May peak. However, the Penney Model hypothesis is a single trend line through May 2013, and then a second line, starting in a potentially different place. The assumption is a single point of inflection, and that point is a peak in May and a drop off beginning in June.[34] While it is correct that the model can find that there is no peak at all in the data, my point is that no other month is modeled as a possibility, and that if the data goes up and down, the model finding a June peak will be statistically significant even though the peak did not occur in May and the drop did not begin in June.

25. The Penney Reply in paragraph 28 criticizes my demonstration, using a polynomial model, that the peak did not occur in May and says such an approach is biased, citing a scholarly article.[35] That article refers to higher order polynomials (which I did not use) and, even for higher order polynomials, the article does not say that such models are biased, only that they may not reduce bias.[36] Indeed, as shown in the quote below, the article brings up the same issues that I do with

---

[30] While the Boston Marathon bombings did not use ammonium nitrate and were not a "dirty bomb," this does not mean they may not have been a reason for a huge uptick in page views. Some news articles (for example https://www.theatlantic.com/technology/archive/2013/04/new-boston-bomb-parts-photos/316183/) discussed the possibility of ammonium nitrate being used. Even if the Boston Marathon bombings had nothing to do with the April uptick in page views, the complete exclusion of any cause of those changes biases the Penney Model, as I have explained.

[31] Penney Reply, paragraph 14 and footnote 8. See page 11 of the Penney Reply for the graph of Ammonium Nitrate views without April data deleted and replaced with the average of March and May 2013.

[32] The article is Aguinis, Herman, Gottfredson, Ryan K., and Joo, Harry, "Best-Practice Recommendations for Defining, Identifying, and Handling Outliers," Organizational Research Methods, 16(2), 2013, p. 270-301.

[33] Neither Dr. Penney nor I have suggested that the change in views in ammonium nitrate in April 2013 was due to an error in the archives used to collect the data.

[34] Penney December Declaration, paragraph 23, describes the design as testing for a "decrease in level and trend" beginning in June 2013.

[35] The article, cited in paragraph 28(b), footnote 32 of the Penney Reply, is Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015

[36] Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015, p. 5.

respect to simplistic linear models, saying that modeling higher polynomial effects does not necessarily fix those issues:

> "the higher-order polynomial has the effect of slightly modifying and improving the fit of the natural linear model. In criticizing the use of high-degree polynomials in RD [RD stands for Regression Discontinuity—the issue theorized in the Penney December Declaration] adjustments, we are not recommending global linear adjustments as an alternative...We recommend that any RD analysis include a plot such as Figure 1 showing data and the fitted model, and that users be wary of any resulting inferences based on fits that don't make substantive sense."[37]

26. In other words, plotting the data is recommended, and the authors are not recommending that a simple linear model is better than a polynomial one. Indeed, they preface that discussion specifically with:

> "Our point here is not to argue that the linear model is correct...Our point is rather that the headline claim, and its statistical significance, is highly dependent on a model choice that may have a data-analytic purpose, but which has no particular scientific basis. Figure 1 indicates to us that neither the linear nor the cubic nor any other polynomial model is appropriate here. Instead, there are other variables not included in the model which distinguish the circles in the graph."[38]

27. I include these extended quotes because despite the Penney Reply's misinterpretation, the article is useful in that it points out the very issue of spurious statistical significance and omitted variable bias that is at the heart of my critiques of the Penney Model in the first place.

28. Next, Paragraphs 19 through 22 of the Penney Reply describes a series of analyses of the single peak May model against other single peak models, concluding that the June model (with a May peak) is better than the others. These analyses are flawed in numerous ways.

29. First and most importantly, the entire exercise is based on a mischaracterization of my critique that implicitly assumes I am proposing a model with an April peak. I merely stated that a *naive* model such as the Penney Model could also be used to "prove" an April peak, meaning that such an analysis could also lead to spurious statistical significance. None of the Penney Reply analyses question this fact. I am not proposing that the data experienced a single change that caused the trend to abruptly reverse after April 2013 (a peak in that month and a decline thereafter). As I have stated numerous times, the data do not indicate a single change model is appropriate, whether that single change is in June 2013 or in some other month.

30. Second, in paragraph 19 of the Penney Reply, Dr. Penney attempts to complete a cross-validation analysis that uses three data sets for each of these article sets. However, two of the three models proposed in paragraph 19 of the Penney Reply, the "total page view" model and the "average total page view" model, are exactly the same statistically.[39] The total page view is simply the average page views multiplied by the number of articles. These two models are equivalent,

[37] Ibid, p. 6.
[38] Ibid, p. 3-4.
[39] Penney Reply, paragraph 19.

statistically, since regression models are invariant to changes in units.[40] For example, suppose we were trying to predict how far a person can jump according to their height in feet, and we ran a regression model that predicted someone who is 6 feet tall can jump 10 feet on average. If we use the same data but run the regression model based on inches, that new model would predict that someone who is 72 inches tall can jump 120 inches on average – in other words, the prediction is unchanged except for the expression in inches instead of feet.

31. The same is the case with running one model on the total and a second on the average, as is done in the Penney Reply (the results of which are summarized in the Penney Reply, paragraph 22). The model is unchanged but one is in terms of averages and one is in terms of totals. Therefore, the estimates for the model run on totals will be 23 times the estimates for the model run on the averages (for the Penney Reply model that has 23 articles). Thus, for example, in the Appendix to the Penney Reply showing the "23 Most Privacy Sensitive Article Set Cross Validation Analysis" (page 41), the coefficient for the variable *time* for the total model is shown to be 21,383.58. Two pages later (page 43), the same coefficient for the variable *time* in the average model is 929.72, which is exactly 21,383.58 divided by 23. The summary statistics like the t-statistic, which is 5.30, are also exactly the same.[41] The Root Mean Square Error and Mean Absolute Errors highlighted for the total model are 89,506.35 and 63,503.27 (on page 41), which, when divided by the 23 articles considered, is equal to the highlighted totals of 3,891.54 and 2760.94 shown for the average model for the highlighted RMSE and Mean Absolute Error, respectively, shown in the attachments to the Penney Reply (on page 43).[42]

32. Thus, while the Penney Reply asserts that there are 48 models (3 models by 4 datasets by 4 change points), there are really only 32 (2 models by 4 datasets by 4 change points). The four datasets also largely overlap, since the 46 article dataset includes all 44 articles in the 44 article dataset, which includes all 23 articles in the 23 article dataset, which includes all 21 articles in the 21 article dataset. In addition, the four months modeled are adjacent, meaning the regression models are very similar (this was part of my original point that the specification of the change point does not make much difference). In other words, though the Penney Reply asserts there are 48 separate models, there are only 32, and most of the 32 are highly related to one another and must produce similar results.

33. Third, the Penney Reply's use of cross validation is misplaced and performed incorrectly. In part the Penney Reply employs a "cross validation analysis."[43] This approach, which the Penney Reply uses to delete different time periods one at a time, is improper for time series models, in which the data points are related to one another.[44] In addition, the Penney Reply's cross

---

[40] See, for example, Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, 2012, South-Western Cengage Learning, p. 40-41.

[41] The r-squared and the p-values are also exactly the same.

[42] There is a slight difference due to rounding or less than 1 for each of the figures.

[43] Penney Reply, paragraph 19.

[44] This is because the data in the cross validation set, or the data "left out", is not independent of the other data. See for example, Bergmeir, Christopher, and Benitez, Jose M., *"On the use of cross-validation for time series predictor evaluation,"* Information Sciences, 2012, 192-213. This paper discusses some of the fundamental problems with traditional cross-validation in time series, primarily in Sections 3.3 and 3.4. Also, see David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig and Carsten F. Dormann, *"Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography 40: 913-929 (913-925 in particular), 2017.*

validation's purpose is to compare the June model to models with a different change point. None of the Penney Reply's cross validation analyses compare the simple single-change model to a model that accounts for other factors or otherwise corrects for omitted variables. Therefore, the Penney Reply's use of cross-validation to compare models and the attempt to show a May Peak model is better than an April peak model or other models are mere distractions that are not related to my criticism.

D. Data Beyond 2014 Show Article Views at About 2012 through 2014 Levels, Even When Earlier Data is Corrected for Mobile Views

34. In my February Declaration I pointed out that the extended view of page view data also does not indicate any long term decline. The Penney Reply, in paragraphs 23 and 34(g), responds to point out that my extended data includes mobile use while the original data presented in the Penney December Declaration did not. To address this "apples to oranges" comparison, I therefore adjusted the 2012 to 2014 data to account for mobile usage. As I describe below, this adjustment has no effect on my conclusions.

35. I also considered the effect that the non-inclusion of mobile usage and the lack of adjustment of that increasing usage had on the Penney Model. I find that such exclusion and lack of adjustment are additional flaws in the Penney Model. Specifically, the Penney December Declaration data excluded mobile page views from the terror and control data sets.[45] If these views were a constant percentage of total views, such an exclusion would likely not affect the Penney Model. However, as I explain below, there was a dramatic increase in mobile web access from January 2012, the first month of data included in the Penney December Declaration analysis, to August 2014, the last month included.

36. The data provided with the Penney December Declaration (but not used in the Penney December Declaration or the Penney Reply) indicates that in January 2012, mobile views accounted for about 12% of total page views.[46] By the end of the study period, that figure was 32%. In other words, the Penney December Declaration's exclusion of mobile views had an increasingly downward bias on total page views. This is yet another bias that affects the Penney Model, and, by not accounting for mobile visits, the Penney Model is biased toward finding an effect and toward finding a larger effect. This bias is a result of the fact that for later data the model excluded more views than for earlier data.[47]

37. In terms of my graphs of extended data as compared to earlier data, the data prior to August 2014 would be higher with mobile data. My graphs included the data as originally provided with the Penney December Declaration, which did not include mobile data for the terror articles. Using

[45] It may be that mobile views were not available, in which case an adjustment, like the one I made, could have been made; or the Penney Model could have included a factor that accounts for such usage.
[46] This is based on the difference between the global English page views non-mobile and the total global English page views, and is consistent with this article https://techcrunch.com/2016/11/01/mobile-internet-use-passes-desktop-for-the-first-time-study-finds/.
[47] While it may seem that simply using desktop views only would not cause a bias to the results, this notion is not correct. To the extent that 1) mobile use was growing during the period, and 2) individuals were using mobile instead of (rather than in addition to) desktop views, the desktop views would be depressed in the latter part of the period and thus bias the results. This has occurred to such an extent that an increasing number of people rely exclusively on mobile access. See, for example, https://techcrunch.com/2016/11/01/mobile-internet-use-passes-desktop-for-the-first-time-study-finds/.

the Penney December Declaration's global article view dataset, which provides total views as well as total views excluding mobile, I adjusted the page views for the terror articles from January 2012 through August 2014 to account for mobile views.[48] The graphs below, showing the extended average and median page views with mobile page views factored in, are consistent with my earlier graphs of the extended data in that they indicate there was no downward trend after June 2013.[49]

38. Average and median page views appear to decline some in mid-2017 but views in 2015 and 2016 appear to be at or above 2012 through 2014 levels. It is also notable in these longer data series that there are clear peaks around the times of major U.S. or European terror attacks, adding further evidence that any reasonable model would account for such attacks (and of course the Boston Marathon bombings occurred very close to the time of the alleged drop due to the Snowden disclosures).

[48] This rough adjustment is undoubtably inaccurate but captures the magnitude and pattern of the mobile views.
[49] The adjustment results in an increase in article views for each month from January 2012 through August 2014, with the amount of increase depending on the share of total Wikipedia views that were mobile.

Figure 3: Average Page Views, Adjusting Data Before 2015 to Factor in Mobile Page Views

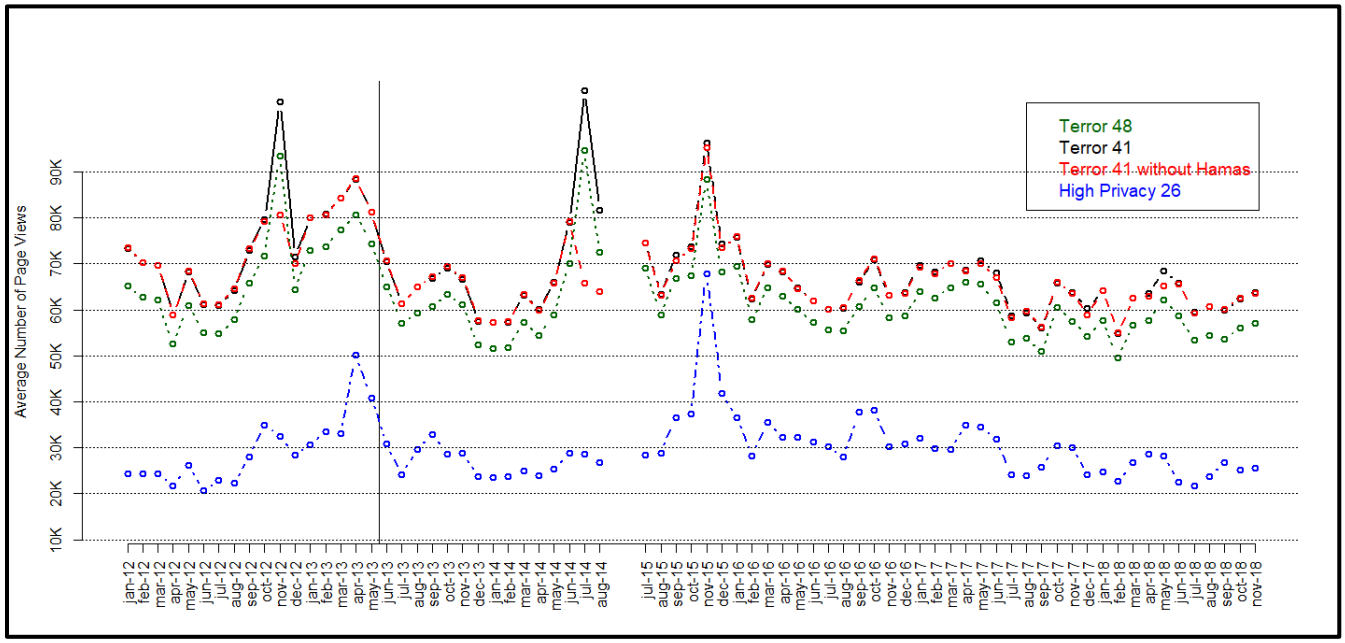
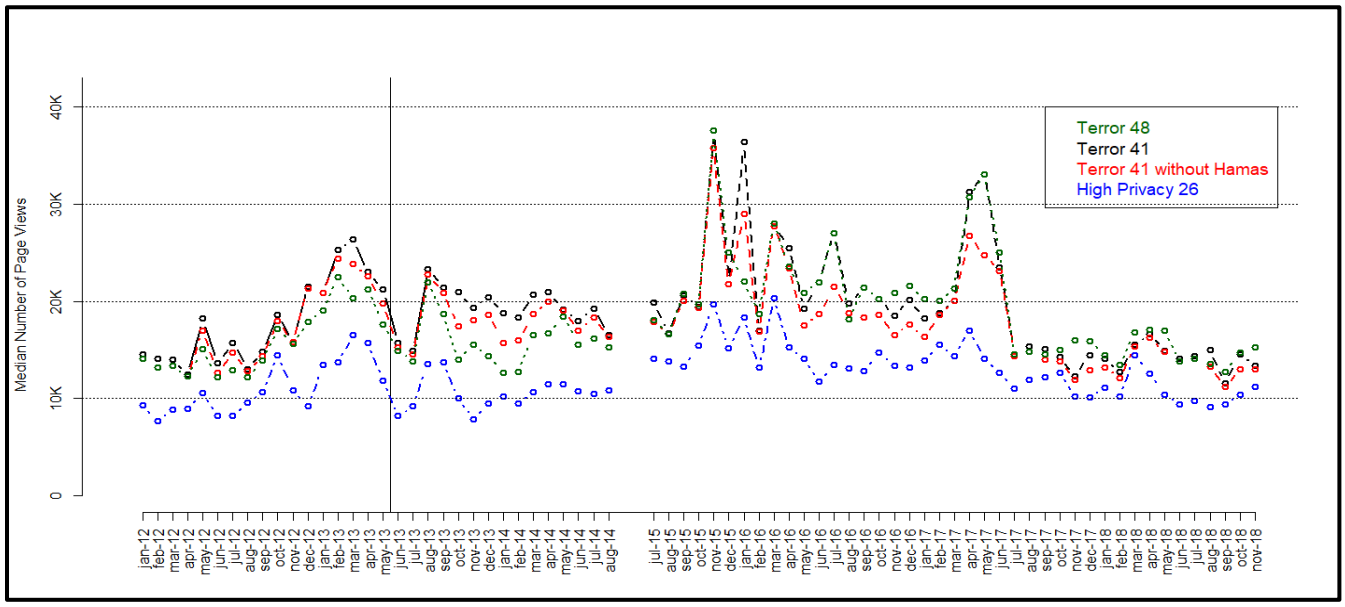


Figure 4: Median Page Views, Adjusting Data Before 2015 to Factor in Mobile Page Views



39. The Penney Reply cites some studies that purport to support the idea that the Penney December Declaration conclusions would continue beyond August 2014, but the Penney Reply neither considers (nor produced in this case) the data underlying those other studies. Even if those studies were to be based on a solid scientific and statistical grounds (and I cannot evaluate whether this is true without the underlying data), they only claim to offer conclusions applicable

to (at the latest) 2015.[50] Moreover, only one of the studies Dr. Penney cites in his reply appears to look at web data, rather than interview answers, and that study is from a working paper that was not published in a scientific journal and it expressly states it only includes data from 2013, and thus does not include any extended data. In any case, there is no way for me to evaluate the validity of those results, because I was not provided the data and it is not publicly available.

40. I do note that one article[51] cited in the Penney Reply footnote 44 adjusts for additional variables and appears to find a smaller (and not statistically significant) effect in terms of drops in searches. This finding is consistent with omitted variable bias I outlined in my first Declaration with respect to the Penney December Declaration.[52]

E. Omitted Variable Bias of the Penney Model Cannot be Solved by Deleting Valid Data.

41. In my February declaration, I pointed out a number of omitted variables that cause bias to the estimates made in the Penney December Declaration. These variables include ones associated with seasonality, individual differences in articles, and news events (the Boston Marathon bombings in particular).[53] The Penney Reply leaves these largely unaddressed but does assert it controls for seasonality because it includes more than one year of data before and after June 2013.[54] However, despite having sufficient data (barely), the Penney Model makes no correction for seasonality and includes no analysis that shows there is not such an effect. I showed such seasonal changes appear in this data and they are statistically significant.[55] In other words, though there was sufficient data, and that data shows statistically significant seasonal effects, the Penney December Declaration ignored seasonality. Wikimedia acknowledged these effects during the deposition of its designee, James Alexander: "global user base, especially in English Wikipedia, tends to have a bit of a dip during the summer, just because there are people out of school, and a lot of people use it in school or when they are studying."[56] Curiously, the Penney Reply, paragraph 30a, states that there is "no basis to expect large seasonal effects with these page views." This statement is speculation that flies in the face of the qualitative and statistical evidence.

42. The Penney declaration correctly states that "in a naturalistic study outside the experimental context, it is not possible to control for all confounding factors."[57] However, the Penney December Declaration corrects for no confounding factors. As one recent author put it: "Obviously, one cannot include in a regression every variable that might conceivably be relevant. But when a factor has a reasonable chance of being important, to exclude it from the modeling is to risk substantial distortion."[58] The Penney Reply re-asserts that the comparator datasets help

[50] Penney Reply, paragraph 34.
[51] Section 3.2 of the article Marthews, Alex, and Tucker, Catherine, "Government Surveillance and Internet Search Behavior," February 17, 2017, found at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564 .
[52] See p. 38 in Marthews and Tucker for lack of statistical significance. For my discussion of omitted variable bias, see Salzberg February Declaration, paragraphs 55-60.
[53] Salzberg February Declaration, paragraphs 55-60.
[54] Penney Reply, paragraph 30a.
[55] Salzberg February Declaration, paragraph 57.
[56] Deposition of Wikimedia designee, James Alexander, April 12, 2018, p. 145.
[57] Penney Reply, paragraph 30e.
[58] Barnett, Arnold I., Applied Statistics: Models and Intuition, Dynamic Ideas LLC, 2015, p. 582.

control for confounding factors, but this is not correct for article-specific factors and is not true when the comparator data is not comparable, as I showed in my February Declaration.[59]

43. In some of the re-analyses in the Penney Reply, articles or time periods are deleted and the Penney Model is re-run.[60] It may be that the Penney Reply does this to address particular examples of the data not following the Penney Model. However, as I said above, these re-analyses do not support the results any more than the original analysis in the Penney December Declaration. Furthermore, by deleting data that tends to disprove the Penney Model and then re-running that data rigs the results toward adoption of the flawed Penney Model.

44. The Penney Reply seems to misinterpret my remarks concerning the staleness of a 2011 DHS list.[61] I was not commenting on the objective nature of the selection, but rather that any list gets stale over time, and the list here used is no exception. For that reason, the static list has no mechanism to update the key articles and therefore a natural decline occurs. The same was not true for the comparator list of popular articles. Because the determination of which articles were popular was made after the time period studied in the Penney December Declaration, articles such as Deaths in 2014 -- which had virtually no page views in 2012 -- were part of the list.[62] On the other hand, a group like ISIL/ISIS, which gained prominence in 2014, was not on the 2011 list, as I pointed out.[63]

F. The Penney Model's Failure to Isolate the Effect of Public Awareness about the NSA Upstream Program

45. My sixth critique, discussed in my February declaration, is that "there are no data or statistical analysis offered that indicate such an effect [an abrupt decline in page views] was due to awareness of the specific NSA program at issue here rather than other related or unrelated events of June 2013."[64] The Penney Reply acknowledges that "in any study of naturalistic changes in human behavior, it will not be possible to isolate the source of all causes and effects on behavior" and that my critique is "a general observation about a *[sic]* naturalistic studies."[65] While this is correct, the Penney December Declaration analysis does not adjust for *any* of those causes, even the obvious ones like seasonality that affect summer page views.

46. Furthermore, the fact that the Penney Model may have been doomed from the start in terms of isolating the effect it intended to prove is not a reason for accepting the model; rather, it is a reason for rejecting it. Despite the passage of nearly six years since the Snowden disclosures, the Penney Reply does not cite a single study published in a peer-reviewed scientific journal that demonstrates the particular effect or even any chilling effect on Internet usage due to awareness of the actual operation of NSA programs.

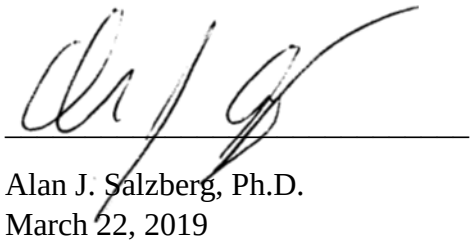
---

[59] Salzberg February Declaration, paragraph 32-46.
[60] In Figure 3b and its explanation in the Penney Reply, the key month of April 2013 is deleted. In Figure 4 and accompanying analyses in the Penney Reply, eight of the original 31 high-privacy articles are deleted for reasons that are unclear to me and unstated in the Penney Reply.
[61] Penney Reply, paragraphs 31 and 32, refer to this critique.
[62] Salzberg Paragraph 64 and database showing 26 most popular articles, which accompanied the Penney December Declaration.
[63] Salzberg Declaration, paragraph 63.
[64] Salzberg Declaration, paragraph 66.
[65] Penney Reply paragraph 36a and 36c, for the first and second quoted material, respectively.

IV. Conclusion

47. In conclusion, my original critiques, detailed in my February Declaration are unchanged by the Penney Reply. In short, the analysis in the Penney December Declaration and the Penney Reply fail to show that public awareness of the Snowden revelations regarding the NSA Upstream program caused any drop in page views of Wikipedia articles.

I declare of penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief. Executed in New York, New York on March 22, 2019.



Alan J. Salzberg, Ph.D.
March 22, 2019

Appendix: Stata Program Log

The following log shows the results of the analysis I performed and described in this declaration. The program was run using Stata, Version 14.

```
      name:  <unnamed>
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\penneyreply\regression_effects_20190318.log
  log type:  text
 opened on:  19 Mar 2019, 11:01:40

. clear

.
. insheet using orig48long.csv
(23 vars, 3,504 obs)

. drop if artnames=="Hamas"
(73 observations deleted)

. save orig48long, replace
file orig48long.dta saved

. keep if monthindex<=32
(1,927 observations deleted)

. save orig48long32, replace
file orig48long32.dta saved

.
. *
. * Simulation that shows no difference in agg v. disagg if same model is run and
issue is just noise
. *
. use orig48long32, clear

. drop if artnames=="Hamas"
(0 observations deleted)

. drop if monthindex>32
(0 observations deleted)

. * run regression to get forecast error
. * no need to show output (but will show output of this for a different purpose
below)
. regress pageviews monthindex intervention postslope, noheader notable

. predict pviewmont
(option xb assumed; fitted values)

. predict sf, stdf

. * simulate data with same forecast error and run regression on disagg
. sort artnum monthindex

. isid artnum monthindex

. * set rndnum seed so can be replicated
. set seed 20190318
```

```
. gen errsim=rnormal(0,sf)

. replace pviewmont=pviewmont+errsim
(1,504 real changes made)

. regress pviewmont monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =     1,504
-------------+----------------------------------   F(3, 1500)      =      3.35
       Model |  6.2583e+10         3  2.0861e+10   Prob > F        =    0.0185
    Residual |  9.3476e+12     1,500  6.2317e+09   R-squared       =    0.0067
-------------+----------------------------------   Adj R-squared   =    0.0047
       Total |  9.4102e+12     1,503  6.2609e+09   Root MSE        =     78941

------------------------------------------------------------------------------
   pviewmont |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   568.2692   570.0658     1.00   0.319    -549.9417     1686.48
intervention |  -11618.79   8230.636    -1.41   0.158    -27763.56    4525.991
   postslope |  -1155.076    893.594    -1.29   0.196    -2907.903      597.75
       _cons |   51521.53   5841.437     8.82   0.000     40063.28    62979.78
------------------------------------------------------------------------------

. * now aggregate, and see that regression standard errors and pvalues are about the
same
. * coeffcients are exactly the same except for rounding because they do not depend on
simulation
. * the Root mean square error is about rmse of disagg model * sqrt(47), or about 7
times as high as mean
. collapse (mean) pviewmont , by( monthindex intervention postslope)

. regress pviewmont monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      3.18
       Model |  1.3316e+09         3   443853226   Prob > F        =    0.0392
    Residual |  3.9062e+09        28   139508526   R-squared       =    0.2542
-------------+----------------------------------   Adj R-squared   =    0.1743
       Total |  5.2378e+09        31   168961239   Root MSE        =     11811

------------------------------------------------------------------------------
   pviewmont |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   568.2691   584.7501     0.97   0.339    -629.5372    1766.075
intervention |  -11618.78   8442.648    -1.38   0.180    -28912.77    5675.196
   postslope |  -1155.076   916.6119    -1.26   0.218    -3032.671    722.5181
       _cons |   51521.53   5991.905     8.60   0.000     39247.67    63795.39
------------------------------------------------------------------------------


. *
. * END Simulation
. *
.
.
. use orig48long32, clear

. drop if artnames=="Hamas"
(0 observations deleted)

. *
. * large changes in standard errors and stat. sign. with removal of a single
observation is another sign of a poor model
```

```
. *
. preserve

. keep if highprivind==1
(512 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope highpriv)

. regress pageviews monthindex intervention postslope if highpriv==1

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      2.98
       Model |  13595332.7         3  4531777.56   Prob > F        =    0.0482
    Residual |  42544868.8        28   1519459.6   R-squared       =    0.2422
-------------+----------------------------------   Adj R-squared   =    0.1610
       Total |  56140201.5        31  1810974.24   Root MSE        =    1232.7

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   123.6005   61.02594     2.03   0.052    -1.405487    248.6065
intervention |  -1336.267   881.0953    -1.52   0.141    -3141.109    468.5747
   postslope |  -189.8362   95.65985    -1.98   0.057    -385.7865    6.114118
       _cons |   6285.478   625.3298    10.05   0.000     5004.548    7566.408
------------------------------------------------------------------------------

. restore

. preserve

. * possible error since recruitment and fundamentalism have exact same page views
nearly every month
. * thus show results without as well as with
. drop if artnames=="Recruitment" | artnames=="Fundamentalism"
(64 observations deleted)

. keep if highprivind==1
(480 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope highpriv)

. regress pageviews monthindex intervention postslope if highpriv==1

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      4.55
       Model |  9185572.85         3  3061857.62   Prob > F        =    0.0102
    Residual |  18850621.1        28   673236.47   R-squared       =    0.3276
-------------+----------------------------------   Adj R-squared   =    0.2556
       Total |    28036194        31  904393.355   Root MSE        =    820.51

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   93.56127   40.62129     2.30   0.029     10.35233    176.7702
intervention |  -1379.924    586.492    -2.35   0.026    -2581.298   -178.5493
   postslope |  -117.9791     63.675    -1.85   0.074    -248.4115    12.45319
       _cons |   6070.125   416.2444    14.58   0.000     5217.487    6922.763
------------------------------------------------------------------------------

. restore

. preserve
```

```
. collapse (median) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      8.18
       Model |  84545042.9         3    28181681   Prob > F        =    0.0005
    Residual |    96510363        28  3446798.68   R-squared       =    0.4670
-------------+----------------------------------   Adj R-squared   =    0.4098
       Total |   181055406        31  5840496.96   Root MSE        =    1856.6

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   374.8799    91.9132     4.08   0.000     186.6043    563.1556
intervention |  -3299.076   1327.047    -2.49   0.019    -6017.408   -580.7433
   postslope |  -535.3763   144.0765    -3.72   0.001    -830.5036    -240.249
       _cons |   9601.022     941.83    10.19   0.000     7671.771    11530.27
------------------------------------------------------------------------------

. restore

. * now without possibly error data
. drop if artnames=="Fundamentalism" | artnames=="Recruitment"
(64 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     17.36
       Model |  72354244.3         3  24118081.4   Prob > F        =    0.0000
    Residual |  38905201.2        28  1389471.47   R-squared       =    0.6503
-------------+----------------------------------   Adj R-squared   =    0.6129
       Total |   111259446        31  3589014.37   Root MSE        =    1178.8

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   315.9363   58.35724     5.41   0.000     196.3969    435.4757
intervention |  -4298.331   842.5644    -5.10   0.000    -6024.246   -2572.416
   postslope |  -342.8184   91.47658    -3.75   0.001    -530.1997   -155.4371
       _cons |   8841.338   597.9838    14.79   0.000     7616.424    10066.25
------------------------------------------------------------------------------

. *
. * demonstrate that errors are correlated with articles, meaning disggregation or
some type of adjustment is needed
. * Also shows that stat significance does not exist for overall data
. use orig48long32, clear

. regress pageviews monthindex intervention postslope if artnames!="Fundamentalism" &
artnames!="Recruitment"

      Source |       SS           df       MS      Number of obs   =     1,440
-------------+----------------------------------   F(3, 1436)      =      3.37
       Model |  6.7546e+10         3  2.2515e+10   Prob > F        =    0.0178
    Residual |  9.5866e+12     1,436  6.6759e+09   R-squared       =    0.0070
-------------+----------------------------------   Adj R-squared   =    0.0049
       Total |  9.6541e+12     1,439  6.7089e+09   Root MSE        =     81706

------------------------------------------------------------------------------
```

```
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   752.6646   603.0018     1.25   0.212    -430.1942    1935.523
intervention |  -14970.22   8706.167    -1.72   0.086    -32048.39    2107.947
   postslope |  -1179.932   945.2219    -1.25   0.212    -3034.096    674.2313
       _cons |   49658.62    6178.93     8.04   0.000     37537.93    61779.32
------------------------------------------------------------------------------

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =     1,504
-------------+----------------------------------   F(3, 1500)      =      3.86
       Model |  7.4228e+10         3  2.4743e+10   Prob > F        =    0.0091
    Residual |  9.6056e+12     1,500  6.4037e+09   R-squared       =    0.0077
-------------+----------------------------------   Adj R-squared   =    0.0057
       Total |  9.6798e+12     1,503  6.4403e+09   Root MSE        =     80023

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   881.2874   577.8797     1.53   0.127    -252.2505    2014.825
intervention |  -14757.81   8343.452    -1.77   0.077    -31123.88    1608.265
   postslope |  -1436.449   905.8423    -1.59   0.113    -3213.301    340.4031
       _cons |   48705.37   5921.504     8.23   0.000     37090.07    60320.68
------------------------------------------------------------------------------

. predict residual1, residual

. * stat sign correlation between articles and residuals mean model is insufficient
(see p-value and f-statistic)
. anova residual1 artnum

                          Number of obs =      1,504    R-squared     =  0.9258
                          Root MSE      =    22124.8    Adj R-squared =  0.9234

                  Source | Partial SS         df         MS        F    Prob>F
              -----------+----------------------------------------------------
                   Model |  8.892e+12         46   1.933e+11    394.91  0.0000
                         |
                  artnum |  8.892e+12         46   1.933e+11    394.91  0.0000
                         |
                Residual |  7.132e+11      1,457   4.895e+08
              -----------+----------------------------------------------------
                   Total |  9.606e+12      1,503   6.391e+09

. * note same coefficients in agg results
. collapse (mean) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     24.85
       Model |  1.5793e+09         3   526437311   Prob > F        =    0.0000
    Residual |   593272771        28  21188313.2   R-squared       =    0.7269
-------------+----------------------------------   Adj R-squared   =    0.6977
       Total |  2.1726e+09        31  70083377.5   Root MSE        =    4603.1

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   881.2874   227.8862     3.87   0.001     414.4836    1348.091
intervention |  -14757.81   3290.232    -4.49   0.000    -21497.54   -8018.073
   postslope |  -1436.449    357.218    -4.02   0.000    -2168.177    -704.721
```

```
       _cons |   48705.37   2335.139    20.86   0.000     43922.06    53488.69
------------------------------------------------------------------------------

.
. *
. * show art is also stat sign for 31 high privacy
. use orig48long32, clear

. keep if highpriv==1
(512 observations deleted)

. regress pageviews monthindex intervention postslope if artnames!="Fundamentalism" &
artnames!="Recruitment"

      Source |       SS           df       MS      Number of obs   =       960
-------------+----------------------------------   F(3, 956)       =      4.20
       Model |  1.3198e+10         3  4.3994e+09   Prob > F        =    0.0058
    Residual |  1.0017e+12       956  1.0478e+09   R-squared       =    0.0130
-------------+----------------------------------   Adj R-squared   =    0.0099
       Total |  1.0149e+12       959  1.0582e+09   Root MSE        =     32369

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   823.6276   292.5762     2.82   0.005     249.4618    1397.793
intervention |  -8112.912   4224.229    -1.92   0.055    -16402.74    176.9203
   postslope |  -1145.897   458.6213    -2.50   0.013    -2045.918   -245.8766
       _cons |    14796.3   2998.014     4.94   0.000     8912.854    20679.75
------------------------------------------------------------------------------

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =       992
-------------+----------------------------------   F(3, 988)       =      5.18
       Model |  1.6582e+10         3  5.5273e+09   Prob > F        =    0.0015
    Residual |  1.0532e+12       988  1.0660e+09   R-squared       =    0.0155
-------------+----------------------------------   Adj R-squared   =    0.0125
       Total |  1.0698e+12       991  1.0795e+09   Root MSE        =     32650

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   918.8429   290.3169     3.16   0.002     349.1343    1488.552
intervention |  -8179.243   4191.609    -1.95   0.051    -16404.72    46.23536
   postslope |  -1340.458   455.0798    -2.95   0.003    -2233.492   -447.4243
       _cons |   15198.27   2974.863     5.11   0.000     9360.491    21036.04
------------------------------------------------------------------------------

. predict residual1, residual

. * stat sign correlation between articles and residuals mean model is insufficient
(see p-value and f-statistic)
. anova residual1 artnum

                          Number of obs =        992    R-squared     =  0.8558
                          Root MSE      =    12570.6    Adj R-squared =  0.8513

                  Source | Partial SS         df         MS        F    Prob>F
              -----------+----------------------------------------------------
                   Model |  9.014e+11         30   3.005e+10    190.14  0.0000
                         |
                  artnum |  9.014e+11         30   3.005e+10    190.14  0.0000
                         |
```

```
    Residual |  1.519e+11        961   1.580e+08
-------------+------------------------------------------------------------
       Total |  1.053e+12        991   1.063e+09

. * note same coefficients in agg results
. collapse (mean) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     20.87
       Model |   534899840         3   178299947   Prob > F        =    0.0000
    Residual |   239215204        28  8543400.16   R-squared       =    0.6910
-------------+----------------------------------   Adj R-squared   =    0.6579
       Total |   774115045        31  24971453.1   Root MSE        =    2922.9

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   918.8429   144.7056     6.35   0.000     622.4269    1215.259
intervention |  -8179.243   2089.266    -3.91   0.001    -12458.91   -3899.577
   postslope |  -1340.458     226.83    -5.91   0.000    -1805.099   -875.8181
       _cons |   15198.27   1482.791    10.25   0.000     12160.91    18235.63
------------------------------------------------------------------------------

.
. log close
      name:  <unnamed>
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\penneyreply\regression_effects_20190318.log
  log type:  text
 closed on:  19 Mar 2019, 11:01:40
```

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 1:15-cv-662**
	)	
NATIONAL SECURITY AGENCY/	)	
CENTRAL SECURITY SERVICE, *et*	)	
***al.*,**	)	
Defendants.	)	

MEMORANDUM OPINION

Plaintiff, Wikimedia Foundation ("Wikimedia"),[1] challenges the legality of the National Security Agency's ("NSA") Upstream surveillance data gathering efforts, one of a series of recent cases challenging the constitutionality of the NSA's surveillance programs.[2] According to the Director of National Intelligence ("DNI"), Upstream surveillance is a surveillance program authorized pursuant to § 702 of the Foreign Intelligence Surveillance Act ("FISA") that involves the targeted collection of non-U.S. persons' international Internet communications by the NSA.[3] Wikimedia alleges that the NSA has intercepted, copied, and collected Wikimedia's Internet

---

[1] This action was originally brought by nine organizations, including Wikimedia, that communicate over the Internet. The other eight organizations were dismissed at the threshold because those organizations lacked Article III standing. *See Wikimedia Found. v. Nat'l Sec. Agency*, 857 F.3d 193, 216–17 (4th Cir. 2017) (affirming in part *Wikimedia Found. v. Nat'l Sec. Agency*, 143 F. Supp. 3d 344 (D. Md. 2015)).

[2] *See Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1144 (2013) (involving a facial challenge to Section 702 of the Foreign Intelligence Surveillance Act); *Obama v. Klayman*, 800 F.3d 559 (D.C. Cir. 2015) (involving a challenge to the NSA's bulk collection of telephone metadata produced by telephone companies); *Am. Civil Liberties Union v. Clapper*, 785 F.3d 787 (2d Cir. 2015) (involving a challenge to the NSA's bulk telephone metadata collection program); *Jewel v. Nat'l Sec. Agency*, No. C 08–04373 (N.D. Cal. April 25, 2019), *appeal docketed*, No. 19–16066 (9th Cir. May 21, 2019) (involving a challenge to the NSA's interception of Internet communications); *Schuchardt v. Trump*, 2019 WL 426482 (W.D. Pa. Feb. 4, 2019), *appeal docketed*, No. 19–1366 (3d Cir. Feb. 14, 2019) (involving a challenge to the NSA's interception of Internet communications through the PRISM surveillance program).

[3] *See* Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

communications pursuant to the Upstream surveillance program and that such interception, duplication, and collection exceeds the NSA's authority under FISA and violates Wikimedia's rights under the First and Fourth Amendments of the Constitution.

At issue in this matter is defendants' motion for summary judgment. Defendants argue that judgment must be entered in their favor because Wikimedia, the only remaining plaintiff, lacks Article III standing. Defendants also argue that even if a genuine dispute of material fact exists as to Wikimedia's standing, the state secrets doctrine precludes further litigation of Wikimedia's standing, and thus requires entry of judgment in defendants' favor.

Before analyzing the parties' arguments on the issue of Article III standing and the state secrets doctrine, however, it is important to address briefly three topics: (i) the definition of Upstream surveillance and the statutory authority for the NSA's Upstream surveillance program, (ii) the procedural history of this case, and (iii) the undisputed factual record developed by the parties. After addressing these three preliminary topics, which frame all of the analysis that follows, the pertinent summary judgment standard is set forth, and the parties' arguments are analyzed under that standard. For the reasons that follow, Wikimedia has failed to establish that it has Article III standing sufficient to survive summary judgment, and further litigation of this matter is precluded by the state secrets doctrine. Accordingly, this case must be dismissed, and judgment must be entered in favor of defendants.

I.

To begin with, it is necessary to define Upstream surveillance, the NSA program at issue in this litigation, and to clarify what is meant by the term Upstream surveillance as that term is used in this litigation. The NSA conducts Upstream surveillance pursuant to § 702 of FISA, 50 U.S.C. § 1881a. The government has acknowledged that it conducts § 702 surveillance through

two programs, namely the Upstream and PRISM programs.[4] In PRISM surveillance, the government acquires communications directly from a United States-based Internet Service Provider ("ISP"). *See* PCLOB 702 Report, at 33. In contrast, the acquisition of communications via Upstream surveillance does not occur "with the compelled assistance of the United States ISPs, but instead with the compelled assistance…of the providers that control the telecommunications backbone over which communications transit."[5] *Id.* at 35. Thus, Upstream collection, unlike PRISM collection, "does not occur at the local telephone company or email provider with whom the targeted person interacts." *Id.* Instead, the collection of communications for Upstream surveillance "occurs 'upstream' in the flow of communications between communication service providers." *Id.* Only the Upstream surveillance program is at issue in this case.

As noted, the government contends that its Upstream surveillance program is conducted pursuant to FISA § 702. Specifically, § 702 permits the Attorney General and the DNI to authorize jointly, for up to one year, foreign-intelligence surveillance targeted at non-U.S. persons located abroad,[6] if the Foreign Intelligence Surveillance Court ("FISC")[7] approves the government's written certification demonstrating that the intended surveillance complies with statutory

---

[4] *See* Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act* 7 (2014) ("PCLOB 702 Report"), *available at* https://www.pclob.gov/library/702-Report-2.pdf.

[5] The telecommunications or Internet "backbone" is the network of high-capacity fiber-optic cables, switches, and routers operated by telecommunications service providers that facilitates both domestic and international communication via the Internet. This backbone primarily consists of a network of fiber-optic cables, including terrestrial cables that link areas across the U.S. and transoceanic cables that link the U.S. to the rest of the world.

[6] Importantly, the statute expressly prohibits the intentional targeting of any persons known at the time of acquisition to be in the United States or any U.S. person reasonably believed to be located outside the United States. 50 U.S.C. § 1881a(b). Section 702 does allow the government, however, to intercept communications between a U.S. person inside the United States and a foreigner located abroad who has been targeted by intelligence officials. *See id.* § 1881a(a)–(b).

[7] FISC, a tribunal composed of eleven federal judges designated by the Chief Justice of the U.S. Supreme Court, is charged with the review of applications for electronic surveillance. *See* 50 U.S.C. § 1803(a).

requirements.[8] To approve such a certification, the FISC must determine that the government's targeting procedures are reasonably designed:

> (i) to ensure that acquisition "is limited to targeting persons reasonably believed to be located outside the United States," 50 U.S.C. § 1881a(j)(2)(B)(i);
>
> (ii) to prevent the intentional acquisition of wholly domestic communications, *id.* § 1881a(j)(2)(B)(ii);
>
> (iii) to "minimize the acquisition and retention, and prohibit the dissemination, of nonpublicly available information concerning unconsenting United States persons consistent with the need of the United States to obtain, produce, and disseminate foreign-intelligence information," *id.* § 1801(h)(1); *see id.* § 1881a(j)(2)(C); and
>
> (iv) to ensure that the procedures "are consistent with…the [F]ourth [A]mendment," *id* § 1881a(j)(3)(A).[9]

In effect, FISC approval of government surveillance pursuant to § 702 means that the FISC has found that the surveillance comports with the statutory requirements and the Constitution.

The recent release of public reports and declassification of some FISC opinions have revealed additional details regarding the collection of communications pursuant to § 702. After the FISC approves a § 702 certification, the NSA designates "targets," which are non-U.S. persons located outside the United States who are reasonably believed to possess or receive, or are likely to communicate, foreign-intelligence information designated in the certification.[10] The NSA then attempts to identify "selectors," namely the specific means by which the targets communicate,

---

[8] The government must certify that a significant purpose of the acquisition is to obtain foreign intelligence information and that the acquisition will be conducted in a manner consistent with the Fourth Amendment and the targeting and minimization procedures required by statute. 50 U.S.C. § 1881a(b), (g).

[9] In addition, following the passage of the FISA Amendments Reauthorization Act of 2017, the FISC must now also find that the government's querying procedures meet the statutory requirements and are consistent with the Fourth Amendment. *Id.* § 1881a(j)(2)(D); (j)(3)(A). These provisions have been cited to the version of § 1881a in effect since January 18, 2018. All of these provisions are identical to those in the version of § 1881a effective between June 2, 2015 and January 18, 2018, but the provisions are now located within § 1881a(j) rather than § 1881a(i).

[10] PCLOB 702 Report, at 41–46.

such as email addresses or telephone numbers.[11] Importantly, selectors cannot be key words (*e.g.*, "bomb") or targets' names (*e.g.*, "Bin Laden"); rather, selectors must be specific communication identifiers.[12] The government then may issue a § 702 directive to a U.S. telecommunications service provider requiring it to assist the government in acquiring communications involving those selectors.[13]

As for the actual collection of communications containing these targeted selectors, the government has described the Upstream surveillance collection process as follows:

> [C]ertain Internet transactions transiting the Internet backbone network(s) of certain electronic communication service provider(s) are filtered for the purpose of excluding wholly domestic communications[,] and are then scanned to identify for acquisition those transactions [that contain communications] to or from . . . persons targeted in accordance with the applicable NSA targeting procedures; only those transactions that pass through both the filtering and the scanning are ingested into Government databases.

Defs.' Br. 4 (quoting Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2).[14] Thus, the Upstream surveillance collection process involves three steps—(1) filtering, (2) scanning, and (3) ingesting. As this description shows, although the government has disclosed some information about Upstream surveillance in declassified documents and

---

[11] NSA Director of Civil Liberties and Privacy Office Report, *NSA's Implementation of FISA Section 702* 4 (2014), *available at* https://www.nsa.gov/Portals/70/documents/news-features/press-room/statements/NSAImplementationofFISA70216Apr2014.pdf.

[12] *Id.*; PCLOB 702 Report, at 32–33, 36.

[13] 50 U.S.C. § 1881a(i); PCLOB 702 Report, at 32–33.

[14] Prior to April 2017, Upstream collection included Internet communications "that were to, from ***or about*** (i.e., containing a reference in the communication's text to) a selector tasked for acquisition under Section 702." FISC Mem. Op. & Order, at 16 (April 26, 2017) (emphasis in original), *available at* https://www.dni.gov/files/documents/icotr/51117/2016_Cert_FISC_Memo_Opin_Order_Apr_2017.pdf. According to the PCLOB 702 Report, under the Upstream surveillance program that included "about" collection, "a communication between two third parties might be acquired because it contains a targeted email address in the body of the communication." PCLOB 702 Report, at 119. As of March 2017, however, the NSA ceased "about" collection entirely, which a FISC judge concluded "should substantially reduce the acquisition of non-pertinent information concerning U.S. persons pursuant to Section 702." FISC Mem. Op. & Order, at 23, 25 (April 16, 2017).

unclassified reports, most technical details of the Upstream surveillance process remain classified. *Wikimedia Found. v. Nat'l Sec. Agency*, 857 F.3d 193, 202 (4th Cir. 2017) (citing *Jewel v. Nat'l Sec. Agency*, 810 F.3d 622, 627 (9th Cir. 2015)).

II.

With this statutory framework and definition of Upstream surveillance in mind, it is appropriate to turn to the procedural history of this case. On June 22, 2015, Wikimedia, along with eight other organizations,[15] filed the Amended Complaint in this suit, challenging the legality of the NSA's Upstream surveillance program. The Amended Complaint alleges that Upstream surveillance (i) exceeds the scope of the government's authority under § 702, (ii) violates Article III, (iii) violates the First Amendment, and (iv) violates the Fourth Amendment and requests (i) a declaration that Upstream surveillance violates the Constitution and § 702 and (ii) an order permanently enjoining the NSA from conducting Upstream surveillance. On August 6, 2015, defendants moved to dismiss the Amended Complaint, arguing that plaintiffs lacked Article III standing. On October 23, 2015, defendants' motion was granted on the ground that plaintiffs' allegations were too speculative to establish Article III standing. *Wikimedia Found. v. Nat'l Sec. Agency,* 143 F. Supp. 3d 344, 356 (D. Md. 2015), *aff'd in part, vacated in part, and remanded by*, 857 F.3d 193 (4th Cir. 2017).

Thereafter, plaintiffs appealed, and the Fourth Circuit affirmed in part, vacated in part, and remanded the case for further consideration. *Wikimedia Found.,* 857 F.3d at 200. Specifically, the Fourth Circuit vacated the finding that Wikimedia lacked standing, but affirmed the finding that the other plaintiffs lacked standing. *Id.* The Fourth Circuit concluded that Wikimedia had

[15] These original plaintiffs included the National Association of Criminal Defense Lawyers, Human Rights Watch, Amnesty International USA, Pen American Center, Global Fund for Women, the Nation magazine, the Rutherford Institute, and the Washington Office on Latin America.

established standing sufficient to survive a facial challenge to the Amended Complaint based on the "Wikimedia Allegation", namely the allegation "that the sheer volume of [Wikimedia's] communications makes it virtually certain that the NSA has intercepted, copied, and reviewed at least some of [Wikimedia's] communications[,]" "even if the NSA conducts Upstream surveillance on only a single [I]nternet [backbone] link." *Id.* at 202, 209 (internal quotation marks and citation omitted). Three factual allegations, accepted as true as required at the motion to dismiss stage, made the Wikimedia Allegation plausible: (i) "Wikimedia's communications almost certainly traverse every international [Internet] backbone link connecting the United States with the rest of the world[,]" (ii) "the NSA has confirmed that it conducts Upstream surveillance at more than one point along the [I]nternet backbone[,]" and (iii) "the government, for technical reasons[,] . . . must be copying and reviewing all the international text-based communications that travel across a given [Internet backbone] link upon which it has installed surveillance equipment." *Id.* at 210–11 (internal quotation marks and citations omitted).

Importantly, the Fourth Circuit rejected the "Dragnet Allegation", that is the allegation "that[,] in the course of conducting Upstream surveillance[,] the NSA is intercepting, copying, and reviewing substantially all text-based communications entering and leaving the United States, including" those of the nine plaintiffs. *Id.* at 202 (internal quotation marks and citation omitted). Plaintiffs alleged the following facts in support of the Dragnet Allegation: (i) "the NSA has a strong incentive to intercept communications at as many [Internet] backbone chokepoints as possible, and indeed must be doing so at many different [Internet] backbone chokepoints," (ii) "the technical rules governing online communications make this conclusion especially true," and (iii) "a *New York Times* article asserts that the NSA is temporarily copying and then sifting through the contents of what is apparently most e-mails and other text-based communications that cross the

[U.S.] border." *Id.* at 213 (internal quotation marks and citations omitted). The Fourth Circuit concluded that the Dragnet Allegation failed to establish standing because it did "not contain enough well-pleaded facts entitled to the presumption of truth." *Id.* at 200. As such, although Wikimedia pled sufficient facts to establish standing at the motion to dismiss stage, the other plaintiffs did not. *Id.* at 200. Thus, Wikimedia is the only remaining plaintiff.

On remand, an Order issued on October 3, 2017 directing the parties to conduct a limited five-month period of jurisdictional discovery. *See* ECF Nos. 117, 123. Both sides took depositions and served requests for written discovery and production of documents. Defendants objected to 53 of Wikimedia's 84 discovery requests on the ground that responses to the requests would reveal classified information protected by the common law state secrets privilege and related statutory privileges. Thereafter, the DNI formally asserted the state secrets privilege and the statutory privilege set forth in 50 U.S.C. § 3024(i)(1).[16] Defendants stated that the information Wikimedia sought, if disclosed, reasonably could be expected to result in exceptionally grave damage to U.S. national security.[17] Wikimedia subsequently moved to compel production of the documents. On August 20, 2018, an Order and Memorandum Opinion issued, concluding that defendants satisfied the procedural requirements necessary to invoke the state secrets privilege, that the information sought to be protected qualified as privileged under

[16] Defendants also submitted a classified declaration from George C. Barnes, the Deputy Director of the NSA. The classified declaration provided additional detail about the harm to national security that would be caused by disclosure of the information contained in Wikimedia's discovery requests.

[17] The DNI's and the NSA's assertions of privilege encompassed seven categories of information: (i) individuals or entities subject to Upstream surveillance; (ii) operational details of the Upstream collection process such as the technical details concerning methods, processes, and devices employed (including the design, operation, and capabilities of the devices); (iii) locations (and nature of the locations) at which Upstream surveillance is conducted; (iv) the specific types or categories of communications either subject to or acquired in the course of the Upstream collection process; (v) the scope and scale on which Upstream collection has or is now being conducted; (vi) the NSA's cryptanalytic capabilities or limitations; and (vii) additional categories of classified information encompassed within numerous FISC opinions and orders. *See* DNI Decl. ¶¶ 18, 21–47.

the state secrets doctrine, and that therefore, Wikimedia's motion to compel must be denied. *Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 790 (D. Md. 2018). Accordingly, the parties continued jurisdictional discovery, limited to information not protected by the state secrets privilege.

Defendants now seek summary judgment on the ground that Wikimedia lacks Article III standing to contest the legality of the NSA's Upstream surveillance program, or alternatively, that if there is a genuine issue of material fact as to the three essential elements of the Wikimedia Allegation articulated in the Fourth Circuit's remand order, the state secrets doctrine operates to preclude further litigation of Wikimedia's standing and thus requires entry of judgment in defendants' favor.

III.

Summary judgment is appropriate only where there are no genuine disputes of material fact. Rule 56, Fed. R. Civ. P. Accordingly, the material facts as to which no genuine dispute exists must first be identified. Defendants set out their statement of material facts in their brief in support of summary judgment, as required by the local rules. Plaintiff, in addition to responding to defendants' statement of material facts as required by the local rules, also offered their own separate statement of material facts in their brief in opposition to summary judgment. Neither the local rules of the District of Maryland nor the Eastern District of Virginia require plaintiff, as the non-moving party, to set forth a statement of material facts. *See generally* D. Md. Local Rules; E.D. Va. Local Civ. R. 56(B). In the interest of completeness, however, and because each party has responded to the other party's statement of material facts, all facts, and disputes as to those facts, have been considered in deriving from the record the following undisputed material facts.

1. The Internet is a global collection of networks, large and small, interconnected by

a set of routers.[18] Together, these large and small networks function as a single, large virtual network, on which any device connected to the network can communicate with any other connected device.

2. To communicate over the Internet, an individual user connects with the network of a local Internet Service Provider ("ISP"), either directly (typically for a monthly fee) or indirectly through an organization (*e.g.*, a place of business, an Internet café). In turn, the local ISP's network connects to the networks of larger regional and national ISPs, the largest of which are called "Tier 1" telecommunication service providers (*e.g.*, AT&T, CenturyLink, Cogent, Verizon).

3. Tier 1 providers and other large carriers maintain high-capacity terrestrial fiber-optic networks, known generally as Internet "backbone" networks, that use long-haul terrestrial cables to link large metropolitan areas across a nation or region. Data travel across these cables in the form of optical signals, or pulses of light.

4. The Internet backbone also includes transoceanic cables linking North and South America with each other and with Europe, Asia, the Middle East, and Africa. These undersea cables reach shore at points known as cable landing stations, from which they are linked to the terrestrial telecommunications network.

5. Tier 1 providers and other large carriers typically connect separate legs of their own networks using high-capacity switches. To allow users of different providers' networks to communicate with one another, Tier 1 providers and other large carriers typically interconnect their networks using high-capacity routers.[19]

6. Generally speaking, to send a communication on the Internet, the transmitting device (*e.g.*, a personal computer, a cell phone) first converts the communication into one or more small bundles of data called "packets," configured according to globally accepted protocols.[20]

7. When a communication is broken into separate packets, each packet includes (i) a "header," which consists of the routing, addressing, and other technical information required to facilitate the packets' travel from its source to its intended

---

[18] Routers are specialized computers that ensure that Internet communications travel an appropriate path across the Internet. Routers serve a similar role for the Internet as switches (or switchboards) do on the telephone network.

[19] Routers and switches perform similar functions, namely directing the transport of Internet communications across the network. Routers generally connect one communications service provider's network to a different communications service provider's network, whereas switches generally connect a single communication provider's network.

[20] Protocols can be thought of as electronic languages. Each protocol, or language, has its own rules and vocabulary. For example, instead of English and Spanish, there is Transmission Control Protocol ("TCP") and User Datagram Protocol ("UDP").

destination, and (ii) a "payload," which consists of a portion of the contents of the communication being transmitted.

8. A packet's header contains three relevant pieces of address and routing information: (i) the packet's source and destination Internet Protocol ("IP") addresses; (ii) the source and destination ports; and (iii) protocol numbers.

9. IP addresses, which are included in packet headers, are unique numeric identifiers assigned to particular computers, devices, or systems connected to the Internet.[21] IP addresses are used to direct data back and forth between one computer (or other online device) and another online device. IP addresses may be analogized to the destination and return addresses on a mailing envelope.

10. The IP addresses of entities with a large, fixed presence on the Internet do not change and are publicly accessible.[22]

11. Port numbers, which are also included in packet headers, are used to identify communications of different kinds (*e.g.*, webpage requests, or email) so that servers hosting multiple communications services (*e.g.*, a website and an email service) can distinguish packets destined for one service from those meant for another. Port numbers for common applications, like web-browsing and email, are assigned in a common industry registry maintained by the IANA. Whereas IP addresses can be analogized to the street address on a letter, port numbers are roughly analogous to the apartment numbers at a multi-unit dwelling.

12. Protocol numbers, which are also included in packet headers, are used by receiving devices to determine the appropriate method of interpreting data (*e.g.*, HTTP, TCP/IP). A protocol defines the actions taken upon the transmission and/or receipt of a message or other transmission. Protocols are also assigned numbers maintained in a common industry registry maintained by the IANA.

---

[21] There are circumstances, however, in which IP addresses do not uniquely identify individual Internet users. For example, residential Internet customers ordinarily get exactly one "dynamic" IP address at a time, which is assigned on a temporary basis by their ISP. Dynamic IP addresses may be assigned for a day, an hour, or some other period of time depending on the needs, resources, and business practices of a particular ISP, after which the dynamic IP addresses are assigned to other customers. Thus, although the IP addresses of business customers of ISPs almost never change, the IP addresses of individual ISP customers can change fairly often, with the same IP address subsequently being assigned to a different customer of the ISP. *See* Dr. Henning Schulzrinne Decl. ¶¶ 30, 33-34, ECF No. 162-2. As another example, the IP addresses in the packets that make up email messages sent or received by an email server on behalf of its users may have the IP address of the server as the source or destination IP address, not an IP address associated with the individual email user. In other words, the IP address in packets transmitting email messages might be the IP address of the email server (*e.g.*, Gmail, Yahoo), rather than the IP address of the individual user of the email address. Scott Bradner Decl. ¶¶ 244-46, ECF No. 168-2.

[22] Each Internet Service Provider or other large enterprise with a fixed presence on the Internet (*e.g.*, Amazon, Wikimedia) acquires blocks of "static" IP addresses assigned on a permanent basis from the appropriate regional Internet registry affiliated with the global Internet Assigned Numbers Authority ("IANA"). There are public databases that record, with very high accuracy, which address blocks are used by what entities.

13. After a communication has been broken into packets by the transmitting device, specialized computers called routers and switches ensure that the packets travel an appropriate path across the Internet to their destination IP address.

14. Each router or switch through which a packet transits scans the packet's header information, including its destination IP address, and determines which direction (path) the packet should follow next in order to reach its intended destination. The router or switch operates somewhat similarly to Google Maps, updating the fastest route to take between a user's starting point and his or her destination.

15. When packets transmitting a communication arrive at the receiving computer, smartphone, or other online device, the receiving device reassembles the packets into the original communication, such as a webpage or email.

16. Traffic "mirroring" is a technical term for a process by which a router or switch, in addition to determining where on the Internet each packet should be forwarded next, can also identify certain packets to be copied ("mirrored") and divert the designated copies off-network for separate processing. In other words, traffic mirroring can create a copy of all communications, or a subset of all communications, passing through a router or switch without interrupting the flow of those communications.

17. Traffic mirroring is accomplished by programming routers and switches with access control lists ("ACLs") to determine whether packets will be copied and collected at a certain link (the "interface") between the router or switch and another device. The criteria used in the ACL can include a packet's source or destination IP address, the port number, the protocol numbers, or other information contained in a packet header.

18. The router or switch examines the header information of each packet it processes, and compares it to the ACL for each interface, to determine which interfaces the packet may or may not pass through without mirroring (copying).

19. Tier 1 providers and other smaller service providers employ traffic mirroring in the normal course of their operations for such purposes as monitoring traffic load, conducting quality-control processes, and rejecting unwanted traffic.

20. At any link on the Internet where surveillance may be conducted, traffic mirroring with ACLs can be used in several ways to make only certain packets available for inspection by a collecting entity.[23]

---

[23] Plaintiff disputes this fact, as well as facts 22-24, to the extent the "collector" or the "collecting entity" is the NSA conducting Upstream surveillance. These facts, as stated, do not put forth that the "collector" or the "collecting entity" is the NSA. In fact, these facts simply establish that any entity, government or private, trying to collect Internet communications could, *hypothetically*, employ traffic mirroring in this manner. Plaintiff's argument that the

21. To conduct traffic mirroring, an interface (a fiber-optic link) would have to be established between the router or switch directing traffic at the selected location and the separate equipment used by the collecting entity (hereinafter, the "collector interface").

22. After the collector interface is established, communications traffic passing through the carrier's router or switch to the collector's equipment can be filtered by "whitelisting" or "blacklisting" techniques. "Whitelisting" or "blacklisting" involves configuring an ACL to allow only packets meeting the ACL's criteria to be copied and passed through the collector interface to the collector's equipment.

23. For example, the collector could configure an ACL containing a "whitelist" of specific IP addresses of interest. When the router or switch examines the header information of each packet it processes, it would then, (i) as usual, forward a copy of the packet toward its intended destination, (ii) perhaps forward additional copies through other interfaces, per the carrier's routine business practices, and (iii) if, and only if, the packet header contains a source or destination IP address on the whitelist, create an additional copy of the packet, and forward it through the collector interface into the collector's possession and control. In other words, packets containing IP addresses on the whitelist would be copied and sent through to the collector's equipment. Packets not meeting the whitelist criteria would not be copied for, or made available to, the collector's equipment for any purpose.

24. Blacklisting, conversely, involves configuring an ACL to allow all packets to be copied to the collector interface *except* those matching the ACL's criteria. With a blacklist, the router or switch would examine each packet header and (i) as usual, forward a copy of the packet toward its intended destination, (ii) perhaps forward additional copies through other interfaces, per the carrier's routine business practices, and (iii) create an additional copy of every packet and forward it through the collector interface into the collector's possession and control, *except* for those packets with source or destination IP addresses on the blacklist. In other words, if the router or switch finds that a packet header contains a source or destination IP address on the blacklist, an additional copy of that packet is not created or forwarded through the collector interface.

25. Whitelisting and blacklisting techniques can also be used to limit mirroring to particular sources of traffic, such as only cables used by specific carriers, or only cables linked to specific countries or regions.

26. In addition, ACLs can be configured to whitelist or blacklist particular types of communication based on their port or protocol numbers, such as email communications or communications from accessing websites.

NSA does not use traffic mirroring in this way when the NSA conducts Upstream surveillance is discussed at length *infra* Part V.C.

27. Wikimedia operates twelve free-knowledge projects on the Internet, including Wikipedia. Wikipedia, a free-access, free content encyclopedia, is one of the top ten most-visited websites in the world. In 2017, Wikipedia's website received visits from more than 1 billion unique devices each month.

28. Wikimedia engages in more than a trillion international Internet communications each year, with individuals in every country on the planet. This includes communications between foreign users and Wikimedia's U.S.-based servers, and communications between U.S. users and Wikimedia's foreign-based servers.

29. Wikimedia has identified three categories of its international Internet communications that it contends are subjected to Upstream surveillance collection by the NSA: (i) communications with its community members[24] ("Category 1"), (ii) internal "log" communications ("Category 2"), and (iii) the electronic communications of Wikimedia's staff ("Category 3").

30. Category 1 consists of communications with and among Wikimedia's community members, including requests from foreign and domestic users to view or download content from Wikimedia websites, and email communications sent from foreign users to Wikimedia servers.[25] All of these communications were directed to the public IP address ranges assigned to and used by Wikimedia.

31. Category 2 consists of internal log communications transmitted from Wikimedia's servers in the Netherlands to its servers in the United States. These communications are encrypted and received at one of the same public IP address ranges as Wikimedia's communications in Category 1.[26]

32. Category 3 consists of communications by Wikimedia's staff using various protocols, some of which are encrypted, some of which are not. These communications, like those in Categories 1 and 2, are sent and received from the public IP address ranges assigned to and used by Wikimedia.[27]

33. The total volume of Wikimedia's international Internet communications exceeds the number of cables transporting Internet communications between the U.S. and other countries. Moreover, Wikimedia's communications are broadly distributed, with users in every country in the world.

---

[24] Wikimedia community members are people who read or contribute to Wikimedia's twelve free-knowledge projects.

[25] According to Wikimedia, the volume of the email communications in Category 1, and the countries from which those emails are received, are unknown. Defs.' Ex. 4, Pl.'s Am Resps. & Objs. to ODNI Interrog. No. 19, Ex. 1 (hereinafter, "Technical Statistics Chart"), ECF No. 162-5.

[26] Technical Statistics Chart; Schulzrinne Decl. ¶¶ 83-84.

[27] Technical Statistics Chart; Schulzrinne Decl. ¶¶ 85-87.

34. It is "virtually certain" that Wikimedia's communications traverse every cable carrying public Internet traffic that connects the U.S. to other countries.

35. The government has described Upstream surveillance as involving three steps. First, "certain Internet transactions transiting the Internet backbone network(s) of certain electronic communication service provider(s) are filtered for the purpose of excluding wholly domestic communications." Second, these Internet transactions "are then scanned to identify for acquisition those transactions [that contain communications] to or from . . . persons targeted in accordance with the applicable NSA targeting procedures." And third, "those transactions that pass through both the filtering and the scanning are ingested into Government databases."[28]

36. Prior to April 2017, Upstream surveillance involved "about" collection (i.e., a communication containing a reference in the communication's text to a selector tasked for acquisition under § 702). "About" communications were not necessarily sent to or from the user of a § 702 tasked-selector.

37. The statement—the "NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server"—was accurate as of October 3, 2011.[29]

IV.

Summary judgment is appropriate when there is "no genuine issue as to any material fact" and based on those undisputed facts the moving party "is entitled to judgment as a matter of law." *Celotex Corp. v. Catrett*, 477 U.S. 317, 322 (1986). To serve as a bar to summary judgment, facts must be "material," which means that the disputed fact "might affect the

---

[28] Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

[29] R. Richards Dep. at 160:4-17; [Redacted], 2011 WL 10945618, at *15. Defendants' Rule 30(b)(6) witness confirmed the accuracy of this statement as of October 2011. Defendants argue that statements of fact in a judicial opinion, such as this statement from a FISC Opinion, are inadmissible hearsay, and thus, plaintiff cannot rely on such statements at summary judgment. Summary judgement evidence must either be in admissible form or capable of being rendered admissible at trial. *Humphreys & Partners Architects, LP v. Lessard Design, Inc.*, 790 F.3d 532, 538-39 (4th Cir. 2015); Fed. R. civ. P. 56(c)(2). Statements of fact in judicial opinions that are offered for the truth of the matter asserted are hearsay. *Nipper v. Snipes*, 7 F.3d 415, 417-18 (4th Cir. 1993); *see also Zeus Enter., Inc. v. Alphin Aircraft, Inc.*, 190 F.3d 238, 242 (4th Cir. 1999); *Carter v. Burch*, 34 F.3d 257, 265 (4th Cir. 1994). Even though the 2011 FISC Opinion is inadmissible hearsay, defendants' Rule 30(b)(6) witness testimony, confirming the accuracy of this specific statement as of October 3, 2011, is not hearsay. Thus, this statement is admissible, but solely this statement because it is as a statement of a party opponent.

outcome of the suit under the governing law." *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 248 (1986). Where a party "fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial," there can be no genuine issue as to any material fact. *Celotex Corp.* 477 U.S. at 322.

Article III limits the jurisdiction of federal courts to actual "Cases" or "Controversies." *See* U.S. Const. art. III, § 2, cl. 1. As the Supreme Court has made clear, one "essential and unchanging part of the case-or-controversy requirement" is that a plaintiff must establish Article III standing to sue. *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992). A plaintiff establishes Article III standing by showing that he, she, or it seeks relief from an injury that is "concrete, particularized, and actual or imminent; fairly traceable to the challenged action; and redressable by a favorable ruling." *Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1147 (2013) (quoting *Monsanto Co. v. Geerston Seed Farms*, 561 U.S. 139, 149 (2010)). In other words, a plaintiff must establish (1) an injury-in-fact; (2) a casual connection between the injury and the alleged conduct; and (3) the redressability of the injury by a court.

To establish injury-in-fact, the alleged injury must be "real and immediate," not "conjectural or hypothetical." *City of Lost Angeles v. Lyons*, 461 U.S. 95, 201 (1983). The Supreme Court has "repeatedly reiterated that '[a] threatened injury must be *certainly impending* to constitute injury in fact,' and that '[a]llegations of *possible* future injury' are not sufficient." *Clapper*, 133 S. Ct. at 1147 (quoting *Whitmore v. Arkansas*, 495 U.S. 149, 158 (1990)) (emphases in original). In some cases, injury-in-fact can also be established "based on a 'substantial risk' that the harm will occur, which may prompt plaintiffs to reasonably [sic] incur costs to mitigate or avoid that harm."[30] *Id.* at 1150 n. 5. Importantly, the standing inquiry is

[30] The parties disagree on whether the appropriate standard for determining injury-in-fact sufficient to establish

"especially rigorous when reaching the merits of the dispute would force [a court] to decide whether an action taken by one of the other two branches of the Federal Government was unconstitutional," particularly "in the fields of intelligence gathering and foreign affairs." *Clapper*, 133 S. Ct. at 1147.

Because standing is a threshold jurisdictional requirement, it may be attacked at any time, including at summary judgment. As the Supreme Court has made clear, each element of standing must be supported "in the same way as any other matter on which the plaintiff bears the burden of proof, *i.e.*, with the manner and degree of evidence required at the successive stages of the litigation." *Defenders of Wildlife*, 540 U.S. at 561. Where, as here, standing is challenged at the summary judgment stage, "'the party invoking federal jurisdiction bears the burden of establishing' standing—and…such a party 'can no longer rest on…mere allegations, but must set forth by affidavit or other evidence specific facts'" to establish standing. *Clapper*, 133 S. Ct. at 1148-49 (quoting *Defenders of Wildlife*, 504 U.S. at 561).

Thus, if a plaintiff cannot set forth, by affidavit or other evidence that will be in admissible form at trial, specific facts sufficient to show a genuine issue for trial on standing,

---

standing is a "certainly impending" standard or a "substantial risk" standard in this case. The Supreme Court has not been clear as to whether the "substantial risk" standard applies and whether that standard is distinct from the "certainly impending" requirement in cases such as this that involve government surveillance. *See Clapper*, 133 S. Ct. at 1150 n. 5. But the Supreme Court has "found standing based on a 'substantial risk' that harm will occur" in some cases. *Id.*

The Fourth Circuit has indicated that injury-in-fact may be established under either the "certainly impending" or the "substantial risk" standard, and thus, standing should be analyzed under both standards in some cases. *See Beck v. McDonald*, 848 F.3d 262, 275 (4th Cir. 2017) (after determining that the threatened harm was not "certainly impending," the Fourth Circuit stated "our inquiry on standing is not at an end, for we may also find standing based on a 'substantial risk' that the harm will occur, which in turn may prompt a party to reasonably [sic] incur costs to mitigate of avoid that harm"). Importantly, the "substantial risk" standard does not change "the common-sense notion that a threatened event can be 'reasonabl[y] likel[y]' to occur but still be insufficiently 'imminent' to constitute an injury-in-fact." *Id.* at 276.

In this opinion, both standards are applied. Moreover, the injury-in-fact standard, whether "certainly impending," "substantial risk," or both, does not impact the outcome in this case because under whichever standard applies, litigation of any remaining dispute of material fact as to Wikimedia's Article III standing cannot be further litigated without violating the state secretes doctrine, as further discussed *infra* Part VI.

then Rule 56(c) mandates entry of summary judgment against the plaintiff. *See Celotex Corp.*, 477 U.S. at 322.

V.

At this stage of the litigation, Wikimedia must present specific facts, supported by admissible record evidence, that are sufficient to show a genuine issue for trial on Wikimedia's Article III standing. In other words, Wikimedia must present specific facts which show that defendants, through the Upstream surveillance program, have copied and collected Wikimedia's international Internet communications, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection.[31]

Both parties have focused their discussion of Wikimedia's standing on the three prongs necessary to establish the Wikimedia Allegation,[32] which were enumerated in the Fourth Circuit's remand order in this case. *See Wikimedia Found.*, 857 F.3d at 210–11. The three prongs are: (A) Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world; (B) the NSA conducts Upstream surveillance at one or more points along the Internet backbone; and (C) the NSA, for technical reasons, must be copying and reviewing all the text-based communications that travel across a given Internet backbone link upon which it conducts Upstream surveillance. Together,

---

[31] *See Obama v. Klayman*, 800 F.3d 559, 562 (D.C. Cir. 2015) ("In other words, plaintiffs here must show *their own* metadata was collected by the government.") (emphasis in original); *Halkin v. Helms*, 690 F.2d 977, 999-1000 (D.C. Cir. 1982) ("[T]he absence of proof of actual acquisition of appellants' communications is fatal to their watchlisting claims.").

[32] The Wikimedia Allegation is the allegation that the sheer volume of Wikimedia's communications makes it virtually certain that the NSA has intercepted, copied, and reviewed at least some of Wikimedia's communications through the Upstream surveillance program, even if the NSA conducts Upstream surveillance on only a single Internet backbone link. *See supra* page 7.

these three prongs would establish that the NSA has copied and collected some of Wikimedia's communications in the course of the NSA's Upstream surveillance program, thereby providing Wikimedia standing to sue here.

The sufficiency of the evidence with respect to each of these prongs is discussed in detail below. The summary judgment record contains specific facts which show no genuine dispute as to the veracity of the first two prongs of the Wikimedia Allegation. With respect to the third prong, however, the summary judgment factual record contains specific facts that establish, without a genuine dispute of material fact, that the NSA, in the course of Upstream surveillance, does not need to be copying any of Wikimedia's communications as a technological necessity. Thus, the summary judgment record does not contain the facts necessary for Wikimedia to establish standing at summary judgment via the Wikimedia Allegation.

A.

The first prong of the Wikimedia Allegation is that Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world.

Wikimedia primarily supports this contention through the declarations of Scott Bradner, plaintiff's Internet expert.[33] Mr. Bradner states that "it is virtually certain that Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries." Bradner Decl. ¶ 6(d), ECF No. 168-2. Mr. Bradner supports this conclusion with evidence of the volume and global distribution of Wikimedia's communications and the relatively few international circuits connecting the U.S. to

[33] Mr. Bradner worked at Harvard University from 1966 to 2016 in a variety of technical and educational roles, including service as Harvard University's Chief Technology Security Officer for a number of years.

other countries. *Id.* at ¶¶ 346-47, 201-05, 209, 218, 220. Thus, Mr. Bradner concludes, to a virtual certainty, that every international fiber-optic cable that transports Internet communications between the U.S. and the rest of the world transports at least some of Wikimedia's international communications.

Defendants have not disputed this fact. *See* Defs.' Brief in Support of Motion for Summary Judgment, Dkt. 162 at 1 (referring to Wikimedia's standing argument as a "one-legged stool" and taking issue with the other two prongs of Wikimedia's standing argument, but not with the argument that Wikimedia's communications traverse every international Internet backbone link).[34]

Thus, there is no genuine dispute between the parties in the summary judgment record that Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world. Wikimedia has presented specific facts, supported by the conclusion of Mr. Bradner, that establish the first prong of the Wikimedia Allegation.

B.

The second prong of the Wikimedia Allegation is that the NSA conducts Upstream surveillance at one or more international Internet backbone links, all of which, as established in the first prong, some of Wikimedia's communications traverse.

Wikimedia primarily relies upon a sentence in a redacted 2011 FISC Opinion and on language describing the Internet backbone in the PCLOB 702 Report to establish this prong. The

---

[34] The government has not explicitly conceded this prong of the Wikimedia Allegation, that Wikimedia's communications traverse every international Internet backbone link connecting the United States with the rest of the world. But the government has indicated that even assuming *arguendo* that Wikimedia has presented sufficient facts to establish this first prong, Wikimedia still does not have standing in this case. *See also id.* at 21.

sentence in the 2011 FISC Opinion states: the "NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server." [Redacted], 2011 WL 10945618, at *15. Defendants' Rule 30(b)(6) witness confirmed the accuracy of this statement as of October 2011.[35] *See* R. Richards Dep. at 160:4-17. Thus, as a statement of a party opponent, this statement is admitted as part of the summary judgment record.

Based on this admission, plaintiff contends that Upstream surveillance involves monitoring "international Internet link[s]." Defendants, however, assert that the meaning of the term "international Internet link" is protected by the state secrets privilege and cannot be confirmed or denied by defendants. Defendants' Rule 30(b)(6) witness testified that "unlike the other words you had me go through in terms of definitions…[which were] what a teleco[m] expert would" provide, the "NSA has an understanding of this term [international Internet link] that is specific to how [the FISC Judge] described it, but it's classified to provide any further information." R. Richards Dep. at 160:19-161:22. Thus, the differences between the term "international Internet link" and the term "circuits," which is a colloquial term used in the telecom industry and is used to describe where along the Internet backbone Upstream collection occurs in the PCLOB 702 Report,[36] cannot be known without violation of the state secrets

---

[35] *See supra* note 29 for further detail as to why the statement in the 2011 FISC Opinion is not inadmissible hearsay in the context of this litigation as a result of defendants' Rule 30(b)(6) testimony regarding the statement.

[36] It is worth noting that the PCLOB 702 Report's reference to "circuits" does not suggest that the NSA is conducting surveillance on more than one circuit. To be sure, the PCLOB 702 Report does use the term "circuits," but it does not do so to refer to the number of sites the NSA is monitoring. Instead, the PCLOB 702 Report uses the term "circuits" in the context of defining the "Internet backbone." Specifically, the PCLOB 702 Report explains that the "Internet backbone" consists of "circuits that are used to facilitate Internet communications[.]" PCLOB 702 Report at 36-37.

privilege.[37] *See* PCLOB 702 Report, at 35-37. Moreover, that this statement was accurate on October 3, 2011 says nothing of this statement's accuracy either in 2015, when this suit was filed, or today.[38]

Rather than belabor the squabble between the parties about the meaning of this particular term from a 2011 FISC Opinion, a different, admissible record document sheds significantly more light on this prong of the Wikimedia Allegation. The Public Declaration of Daniel R. Coats, Director of National Intelligence ("DNI"), states that the United States Intelligence Community "has publicly acknowledged that Upstream surveillance is conducted on one or more points on the Internet backbone" and that the United States Intelligence Community "has publicly acknowledged that…NSA is monitoring at least one circuit carrying international Internet communications." Pub. Decl. of Daniel R. Coats, DNI, ¶¶ 30, 37, ECF No. 138-2.[39] In other words, the DNI, who oversees the United States Intelligence Community, has admitted, in the course of this litigation, that the NSA conducts Upstream surveillance on at least one point on the Internet backbone and, to the extent the terms Internet backbone and international Internet circuit are not interchangeable, on at least one circuit carrying international Internet communications.[40]

---

[37] The state secrets privilege's applicability to this case is discussed in significantly greater depth *infra* Part VI.

[38] The statement from the 2011 FISC Opinion pertains to the Upstream surveillance program's collection of "about" communications. As of April 2017, Upstream surveillance no longer involves any "about" collection. Thus, at least the conclusion of this conditional statement is no longer accurate today.

[39] Neither party has cited to these specific paragraphs of the Public Declaration of the DNI in their briefs. Nonetheless, the Public Declaration of the DNI is clearly part of the evidentiary record in this matter, as defendants have cited to other paragraphs of this declaration in their statement of undisputed material facts. Moreover, as the "oversee[r of] the United States Intelligence Community," the DNI is in a position to make such statements from personal knowledge.

[40] In this context, the terms Internet backbone and international Internet circuits both refer on some level to the transoceanic fiber-optic cables that transport Internet communications and connect the U.S. to the rest of the world.

Accordingly, the undisputed summary judgment record adequately establishes that the NSA monitors at least one circuit carrying international Internet communications in the course of Upstream surveillance and that Wikimedia's communications traverse every circuit carrying international Internet communications from the United States to the rest of the world. Thus, Wikimedia has established the first two prongs of the Wikimedia Allegation with the support of admissible record evidence and without a genuine dispute as to any material fact.

C.

With respect to the third prong, however, the summary judgment factual record contains specific facts that establish, without a genuine dispute of material fact, that it is *not* a technological necessity that the NSA has copied or collected some of Wikimedia's communications over the one circuit that the NSA admits monitoring to conduct Upstream surveillance.[41] Accordingly, the summary judgment record does not contain the facts necessary for Wikimedia to establish standing at summary judgment via the Wikimedia Allegation.

To address this prong of the Wikimedia Allegation, both parties have submitted extensive expert reports. The government's expert, Dr. Henning Schulzrinne,[42] has provided expert testimony that details a method of collecting Internet communications, which could, *hypothetically*, avoid collecting any of Wikimedia's communications. Dr. Schulzrinne Decl. ¶

---

[41] Importantly, to establish standing, Wikimedia need only prove that the NSA has copied or scanned some of its communications as part of the Upstream surveillance program, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Wikimedia has chosen to prove that it is a technological necessity that the NSA has copied or scanned some of its communications only because the government's assertion of the state secrets privilege prevents Wikimedia from posing the more direct question of whether the NSA has actually copied or scanned any of Wikimedia's communications as part of the Upstream surveillance program. *See Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 788-90 (D. Md. 2018).

[42] Dr. Henning Schulzrinne has been a professor of computer science at Columbia University since 1996 and holds a Ph.D. and a Master's Degree in Electrical Engineering.

77-88. Thus, Dr. Schulzrinne concludes that the NSA, via Upstream surveillance, does not *have* to be collecting any of Wikimedia's communications "as a matter of technological necessity." Dr. Schulzrinne 2d Decl. ¶ 2. Importantly, Dr. Schulzrinne does not provide testimony about the actual operational details of Upstream surveillance because the actual operational details of Upstream surveillance are classified and protected by the state secrets privilege, and thus, Dr. Schulzrinne does not know any of the classified operational details. *Id.* at ¶ 3-4.

On the other side, Wikimedia's expert, Scott Bradner, has provided expert testimony in which he opines, based on a combination of technical and practical factors, that the NSA "most likely" copies all communications transported across an international Internet circuit *before* filtering any of the communications. Bradner Decl. ¶ 282. As a result, Mr. Bradner concludes that "even if the NSA were monitoring only a single circuit under [U]pstream collection, it would be copying and reviewing at least some of Wikimedia's communications." *Id.* at ¶ 353.

Each expert unsurprisingly takes issue with the other's findings. Dr. Schulzrinne claims that Mr. Bradner has provided "no support, and certainly none based in Internet technology and engineering, for concluding that the NSA 'almost certainly' (Bradner Decl. ¶ 6(a)) copies and scans all communications traversing any circuit it monitors, including Wikimedia's." Dr. Schulzrinne 2d Decl. ¶ 5. And Mr. Bradner claims that Dr. Schulzrinne's conclusion that the NSA does not have to be collecting any of Wikimedia's communications as a matter of technological necessity "is simply implausible as a practical matter given everything that is known about [U]pstream collection." Bradner Decl. ¶ 362. For the reasons that follow, this dispute does not present a triable issue of fact.

To begin with, it is necessary to address the practical grounds on which Mr. Bradner reaches his conclusions. Mr. Bradner contends that the NSA could not accomplish its stated goal

of "*comprehensively acquir[ing]* communications that are sent to or from its targets" through Upstream surveillance without first copying all international communications transported over the circuit(s) that the NSA is monitoring. *Id.* at ¶ 333 (quoting PCLOB 702 Report, at 10, 123, 143 (emphasis added)); *Id.* at ¶ 335. To accomplish this goal, Mr. Bradner opines that the NSA is "most likely" copying all of the communications traveling across a circuit before later filtering those communications based on the NSA's targeted selectors. *Id.* at ¶¶ 282, 289. As the basis for this opinion, Mr. Bradner claims (i) that any other method would require the NSA to share sensitive information about its targets and/or filtering criteria with an assisting provider, which the NSA would prefer not to do, (ii) that any other method would require the NSA to place an NSA-operated device into the heart of an ISP's network, which the NSA would prefer not to do, and (iii) that the NSA has no operational incentive to reduce the number of communications it scans for selectors. *Id.* at ¶¶ 283-88.

None of Mr. Bradner's bases for this opinion, however, have a non-speculative foundation in technology. Instead, speculative assumptions about the NSA's surveillance practices and priorities and the NSA's resources and capabilities form the basis for Mr. Bradner's opinion in this regard.[43] *See* Dr. Schulzrinne 2d Decl. ¶ 73. Simply put, Mr. Bradner does not know what the NSA prioritizes in the Upstream surveillance program because that information is classified, and therefore Mr. Bradner has no knowledge or information about it.

---

[43] *See, e.g.*, *Obama v. Klayman*, 800 F.3d 559, 567 (D.C. Cir. 2015) (rejecting a plaintiff's claim that the NSA's collection must be comprehensive to be effective because "there are various competing interests that may constrain the government's pursuit of effective surveillance. Plaintiffs' inference fails to account for the possibility that legal constraints, technical challenges, budget limitations, or other interests prevented NSA from collecting metadata from Verizon Wireless."). Wikimedia has gone significantly further than the plaintiffs in *Klayman* to address the technological issues pertinent to the effectiveness of a less comprehensive surveillance system, but Mr. Bradner still takes significant speculative leaps about the NSA's practical and operational decision-making to reach these particular aspects of his conclusions. These specific conclusions require speculative leaps which are too significant to accept as the foundational basis for an expert's opinion.

As a result, Mr. Bradner's opinions as to these specific propositions are inadmissible pursuant to Rule 702, Fed. R. Evid., and the standards articulated in *Daubert v. Merrell Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).[44]

Moreover, even if Mr. Bradner's opinions on these specific propositions were admissible, any conclusions drawn from those opinions would be barred by the state secrets doctrine, as further discussed *infra* Part VI. No matter how intuitively appealing Mr. Bradner's opinions about the NSA's operational priorities may seem, courts have consistently recognized that "judicial intuition" about such propositions "is no substitute for [the] documented risks and threats posed by the potential disclosure of national security information." *Al-Haramain Islamic Found., Inc. v. Bush*, 507 F.3d 1190, 1203 (9th Cir. 2007). Importantly, defendants cannot effectively defend themselves against Mr. Bradner's speculations without disclosing information about the operational details of the NSA's Upstream surveillance program. But defendants have thoroughly documented the risks of such a disclosure in the classified declaration, explaining that to reveal such facts regarding the operational details of the Upstream surveillance collection process, even considering the public disclosures made to date, would provide insight into the structure and operations of the Upstream surveillance program and in so doing, undermine the effectiveness of this important intelligence method. Thus, even if Mr. Bradner's conclusions,

---

[44] Rule 702 provides that an expert may offer opinion testimony if "the expert's scientific, technical, or other specialized knowledge" will be helpful to understand the evidence or to determine a fact in issue, the proffered opinion is "based on sufficient facts or data," and it is "the product of reliable principles and methods…reliably applied…to the facts of the case." Fed. R. Evid. 702(a)-(d). *Daubert* explained that to meet the test of admissibility under Rule 702, an expert's testimony must rest on a reliable foundation, meaning it "must be based on scientific, technical, or other specialized *knowledge* and not belief or speculation." *Oglesby v. Gen. Motors Corp.*, 190 F.3d 244, 250 (4th Cir. 1999) (emphasis in original); *see also Nease v. Ford Motor Co.*, 848 F.3d 219, 229, 231 (4th Cir. 2017). Here, the critical propositions that form the basis for Mr. Bradner's opinion that the NSA is "most likely" copying all communications before any filtering do not meet this requirement as they are based on Mr. Bradner's speculation as to the NSA's operational priorities and capabilities, not on any technical requirements for the collection of Internet communications. Although the NSA has made some public disclosures about Upstream surveillance, Mr. Bradner's interpretations of single sentences within the public disclosures stretches those disclosures far beyond a natural reading of them, and again, is not based on any knowledge, technical or otherwise.

built off assumptions about the NSA's operational goals from the NSA's limited public disclosures, were admissible as expert opinions, the state secrets doctrine would bar any further litigation of this prong of Wikimedia's standing argument, as further discussed *infra* Part VI.

Analysis of the third prong of the Wikimedia Allegation, however, does not end with dismissal of Mr. Bradner's non-technical assumptions. Each expert has also presented technical arguments for and against the proposition that the NSA must be collecting at least some of Wikimedia's communications at the circuit(s) monitored pursuant to the Upstream surveillance program.

Dr. Schulzrinne explains how the NSA, using the technique of "traffic mirroring" in a specific manner,[45] could conduct Upstream surveillance on an international Internet circuit "without intercepting, copying, reviewing, or otherwise interacting with [the] communications of Wikimedia." Dr. Schulzrinne Decl. ¶ 77. To begin with, Wikimedia has been allocated a number of static IP addresses. *Id.* at ¶ 80. A "static" IP address is an IP address that is assigned on a *permanent* basis from the appropriate regional Internet registry. *See id.* at ¶ 32-33. Static IP addresses are generally assigned to large enterprises on the Internet so that users around the world have consistent access to their websites. Public databases record, with very high accuracy, which IP address blocks are used by what entities. *Id.* Thus, any member of the public can ascertain all of the IP addresses assigned to Wikimedia.

Through a process of "blacklisting" Wikimedia's IP addresses, the NSA could conduct Upstream surveillance without receiving access to any of Wikimedia's communications. *Id.* at ¶ 82. To do so, the NSA could blacklist all of Wikimedia's IP addresses using an access control

---

[45] Traffic mirroring, as defined in the statement of material facts in the summary judgment record, is a technical term for a process by which all communications passing through a router or switch can be copied without interrupting the flow of communications.

list, a list employed in the traffic mirroring process that determines which packets carrying Internet communications will be copied and collected at a certain circuit on the Internet backbone. By blacklisting Wikimedia's IP addresses, all Internet communications *except* those containing Wikimedia's blacklisted IP addresses would be copied and collected by the NSA. Importantly, this hypothetical does not propose that the NSA is copying all Internet communications other than Wikimedia's, but rather states that, as a technical matter, the NSA *could* blacklist certain high-frequency, low-interest IP addresses to minimize the collection of communications of little interest to the NSA and that Wikimedia's IP addresses *could* be high-frequency, low-interest IP addresses to the NSA. Thus, strictly considering the technological limitations of copying Internet communication in transit, it is possible that the NSA has not copied and collected any of Wikimedia's communications despite monitoring an international Internet circuit that transmits some of Wikimedia's communications.[46]

In response, Mr. Bradner finds this hypothetical "simply implausible" as a practical matter given everything that is known about Upstream surveillance, although Mr. Bradner does admit that selective collection is technologically possible. Bradner Decl. ¶ 362, 272(b), 280-81, 299, 325, 366. The foundation for Mr. Bradner's response is that the NSA has disclosed to the public that Upstream surveillance operates by identifying "selectors," the specific means by which the targets communicate, such as email addresses or telephone numbers.[47] Because the

---

[46] In addition to blacklisting Wikimedia's IP addresses, Dr. Schulzrinne proposes several other whitelisting or blacklisting options which would prevent the NSA from collecting Wikimedia's international Internet communications. Dr. Schulzrinne Decl. ¶ 77-88. For example, the NSA could blacklist the ports assigned to HTTP and HTTPS communications so as not to collect any web communications that involve accessing websites. *Id.* at ¶ 79.

[47] NSA Director of Civil Liberties and Privacy Office Report, *NSA's Implementation of FISA Section 702* 4 (2014), *available at* https://www.nsa.gov/Portals/70/documents/news-features/press-room/statements/NSAImplementationofFISA70216Apr2014.pdf.

NSA cannot know in advance which communications contain selectors, Mr. Bradner contends, the NSA must first copy all communications before scanning any of them for selectors. Bradner Decl. ¶ 333, 301.

Despite Mr. Bradner's arguments to the contrary, the traffic mirroring hypothetical proposed by Dr. Schulzrinne does not contradict the government's public disclosures about Upstream surveillance. Importantly, the government has described Upstream surveillance as involving three steps—(1) filtering, (2) scanning, and (3) ingesting.[48] The whitelisting and blacklisting process of traffic mirroring proposed by Dr. Schulzrinne would occur at the first step in the NSA's collection process, the filtering, prior to any copying or scanning. Thus, under Dr. Schulzrinne's hypothetical, the first step, filtering, would involve a combination of whitelisting and blacklisting to exclude wholly domestic communications *and* other low interest communications, and Wikimedia's communications may qualify as low interest communications that the NSA filters out.[49] *Second*, and only after filtering, the NSA would scan the remaining communications for "selectors," which could result in the collection of both communications to or from a targeted selector and about a targeted selector. *See* Dr. Schulzrinne 2d Decl. ¶ 50-52. This second step described in the government's public disclosures is the step on which Mr. Brander focuses. Given the distinction between the first two steps, Dr. Schulzrinne's hypothetical is consistent with government's public disclosures about Upstream surveillance.

---

[48] *See* Material Fact 35; Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

[49] It is noted that the government has not disclosed that anything other than wholly domestic communications are filtered out at the first step in the Upstream collection process. Given the government's limited disclosures about the technical details of how Upstream surveillance operates, however, this disclosure does not mean that the government does not, or could not, engage in additional filtering at the first step in the Upstream surveillance collection process. Whether or not the government actually engages in additional filtering at the first step in the Upstream surveillance collection process is a fact protected by the state secrets privilege. *See Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 789-90 (D. Md. 2018); Pub. Decl. of Daniel R. Coats, DNI, ¶ 18(B), 18(D), ECF No. 138-2.

Moreover, the hypothetical, regardless of whether it is actually how the NSA conducts Upstream surveillance, does show that there is a technological method by which the NSA could conduct Upstream surveillance on a circuit transporting International internet communications without copying, collecting, or otherwise reviewing any of Wikimedia's communications that traverse that path.

But this does not end the analysis, for there is a technological hurdle that remains. Even if the NSA used the whitelisting and blacklisting techniques proposed by Dr. Schulzrinne to filter the communications it collected via Upstream surveillance, Mr. Bradner maintains that there are three scenarios in which Wikimedia's communications would still be copied and scanned by the NSA. Bradner Decl. ¶¶ 367(b), 370. In these three specific scenarios—namely (i) communications contained within a multi-communication transaction,[50] (ii) emails to or from Wikimedia involving a person located abroad who is using an email service located in the U.S.,[51] or (iii) a person located abroad who accesses Wikimedia's websites through a U.S.-based Virtual

---

[50] A "multi-communication transaction" (MCT) is "an Internet transaction that contain[s] multiple discrete communications." NSA Response to Plaintiff's Interrogatory No. 8 (Dec. 22, 2017). When an email user logs into their email service to check his or her email, the group of all unread email messages is transmitted together as a single communication from the email service to the subscribing user's inbox. This transmission of multiple emails in a single communication might be considered an MCT. Bradner Decl. ¶¶ 67, 132, 317. In transit, an MCT of this type would contain the IP address of the email service as the sender and the IP address of the user as the recipient. If an email to or from Wikimedia were contained within the batch of emails sent as an MCT, the Wikimedia email would be transmitted to the user's inbox without Wikimedia's IP address in the individual packet headers of the MCT. Dr. Schulzrinne 2d Decl. ¶ 78. Thus, this specific type of Wikimedia communication could be transmitted from an email service to a user of the email service without Wikimedia's IP address being the source or destination IP address. And as a result, blacklisting Wikimedia's IP addresses would not prevent the NSA's collection of such an email from an international Internet circuit which the NSA is monitoring.

[51] This scenario is similar to the first MCT scenario. If (i) an email user sent an email to Wikimedia or received an email from Wikimedia, (ii) that email user was abroad, and (iii) that email user utilized a U.S.-based email service, the communication between the email user and the email service would not include Wikimedia's IP address in the packet headers and would need to traverse an international Internet circuit between the U.S.-based email service and the user located abroad. Bradner Decl. ¶ 367(b)(2); Dr. Schulzrinne 2d Decl. ¶ 81. Thus, this specific type of Wikimedia communication could be transmitted from an email service to a user of the email service without Wikimedia's IP address being the source or destination IP address. And as a result, blacklisting Wikimedia's IP addresses would not prevent the NSA's collection of such an email from an international Internet circuit which the NSA is monitoring.

Private Network (VPN),[52] Wikimedia's IP address would not appear as the source or destination IP address on the packet header traversing the international Internet circuit into or out of the U.S. *See* Bradner Decl. ¶ 367(b)(1)-(3); Dr. Schulzrinne 2d Decl. ¶ 77-87. Thus, these communications would not be blocked by the NSA's hypothetical blacklist of Wikimedia's IP addresses because the communications would not contain Wikimedia's IP address in the packet header, despite involving a Wikimedia communication.

Dr. Schulzrinne admits that each of these scenarios is "theoretically possible" but "could come to pass only in the uncertain event that particular conditions are met." Dr. Schulzrinne 2d Decl. ¶ 77. For communications in each of these three scenarios to be collected by the NSA through Upstream surveillance, at least four conditions would have to be met,[53] none of which Wikimedia has established as to any of their communications in this case. Specifically, for Wikimedia communications to exist in either of the first two scenarios, an email user in a foreign location must be downloading emails from a server located in the United States (such that the communication would traverse an international Internet circuit monitored by the NSA) *and* the email user must be sending email to and/or receiving email from Wikimedia. *Id.* at ¶¶ 78, 81. Wikimedia has not presented evidence of any such subset of its communications.[54] For

---

[52] When a user communicates via a Virtual Private Network (VPN), all of the user's communications are encrypted and first routed through the VPN server before being directed to their ultimate destination. Dr. Schulzrinne 2d Decl. ¶ 57. As a result, first, each communication's packet is assigned the VPN server's address as its destination IP address, not the IP address of the ultimate destination. *Id.* Then, once the communication has reached the VPN server (destination one), the communication travels from the VPN server to the ultimate destination (destination two), with the VPN server IP address as the source IP address, rather than the individual user's IP address. Therefore, if a person is located abroad and accesses Wikimedia's website while using a U.S.-based VPN and the first leg communication between the VPN user and the VPN server traverses an international Internet circuit that the NSA is monitoring, the NSA could collect that communication even if the NSA has blacklisted Wikimedia's IP addresses. Bradner Decl. ¶ 367(b)(3).

[53] Dr. Schulzrinne 2d Decl. ¶ 78, 81, 83.

[54] It is worth noting that Wikimedia has acknowledged that is does not know the volume of its international email communications, or the countries from which the emails are received. *See* Technical Statistics Chart. In addition to

Wikimedia's communications to exist in the third scenario, a user of a Virtual Private Network (VPN) that is based in the United States must use that VPN while abroad to visit one of Wikimedia's websites, and the NSA must monitor the international Internet circuit that transmits that communication from the user abroad to the domestic VPN. Again, Wikimedia has not presented evidence of any such subset of its communications. As a result, satisfaction of the chain of conditions necessary to establish that the NSA collected Wikimedia's communications in one of these three circumstances is too speculative to establish standing. *See Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1148, 1150 (2013) (holding that a speculative chain consisting of five contingencies was insufficient to establish standing). Thus, although it is possible that such communications exist,[55] the summary judgment record does not contain any evidence that such communications actually exist, a requirement at this stage of the litigation. *See Clapper*, 133 S. Ct. at 1148-49.

In sum, the undisputed summary judgment record does not establish that the NSA has copied any of Wikimedia's international Internet communications in the course of Upstream surveillance, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Specifically, the summary judgment record establishes that it is not a technological necessity that the NSA must copy all of the text-based Internet communications traversing a circuit that the NSA monitors while conducting Upstream surveillance. The NSA could, *hypothetically*, utilize a process of

---

the total volume and location of all of Wikimedia's international email communications being unknown, this particular subset of Wikimedia's international email communications is also unknown – in volume, in geographic diversity, or even whether such communications exist.

[55] It is worth noting that if such communications exist, they are likely to be far fewer in number than the trillions of international Wikimedia communications every year that traverse every International circuit connecting the U.S. to the rest of the world. Thus, a finding that such communications exist could trigger a re-evaluation of the first prong of Wikimedia's standing argument, *i.e.* that Wikimedia's subject international Internet communications traverse every international Internet backbone link connecting the United States with the rest of the world.

whitelisting and blacklisting to filter out low-interest Internet communications, including Wikimedia's communications, prior to scanning the Internet communications for targeted selectors. At most, there is a genuine dispute of material fact as to whether the NSA can conduct Upstream surveillance without copying Wikimedia's communications, if any, that (i) are contained within a multi-communication transaction, (ii) are emails to or from Wikimedia involving a person located abroad using an email service located in the U.S., or (iii) involve a person located abroad accessing Wikimedia's websites through a U.S.-based Virtual Private Network (VPN) *and* that traverse an NSA-monitored circuit. To the extent there is a genuine issue of material fact with respect to the NSA's collection of this currently unidentified subset of Wikimedia's international communications, that issue cannot be further litigated given the state secrets doctrine, as further discussed *infra* Part VI.

VI.

Even assuming *arguendo* that, there is a genuine dispute of material fact as to the third prong of the Wikimedia Allegation, the question remains as to how the matter should proceed consistent with Supreme Court and Fourth Circuit precedent regarding the state secrets doctrine. Wikimedia's standing cannot be fairly litigated any further without disclosure of state secrets absolutely protected by the United States' privilege. For Wikimedia to litigate the standing issue further, and for defendants to defend adequately in any further litigation, would require the disclosure of protected state secrets, namely details about the Upstream surveillance program's operations. For the reasons that follow, therefore, the standing issue cannot be tried, or otherwise further litigated, without risking or requiring harmful disclosures of privileged state secrets, an outcome prohibited under binding Supreme Court and Fourth Circuit precedent. Thus, the case must be dismissed, and judgment must be entered in favor of defendants.

A.

It is necessary first to review the well-settled Supreme Court and Fourth Circuit precedent concerning the state secrets doctrine. Settled Supreme Court and Fourth Circuit precedent make clear that "[u]nder the state secrets doctrine, the United States may prevent the disclosure of information in a judicial proceeding if 'there is a reasonable danger' that such disclosure 'will expose…matters which, in the interest of national security should not be divulged.'" *Abilt v. CIA*, 848 F.3d 305, 310-11 (4th Cir. 2017) (quoting *El-Masri v. United States*, 479 F.3d 296, 302 (4th Cir. 2007)) (quoting *United States v. Reynolds*, 345 U.S. 1, 10 (1953)). In this regard, the Fourth Circuit has recognized that the state secrets doctrine "performs a function of constitutional significance, because it allows the executive branch to protect information whose secrecy is necessary to its military and foreign-affairs responsibilities." *Id.* at 312 (quoting *El-Masri*, 479 F.3d at 303).

The Fourth Circuit has mandated a three-step analysis for resolution of the state secrets question:

> First, "the court must ascertain that the procedural requirements for invoking the state secrets privilege have been satisfied." Second, "the court must decide whether the information sought to be protected qualifies as privileged under the state secrets doctrine." Third, if the "information is determined to be privileged, the ultimate question to be resolved is how the matter should proceed in light of the successful privilege claim."

Abilt, 848 F.3d at 311 (quoting *El-Masri*, 479 F.3d at 304). Previously, an Order and Memorandum Opinion issued in this case, which concluded that defendants satisfied the procedural requirements necessary to invoke the state secrets privilege, that the information sought to be protected qualified as privileged under the state secrets doctrine, and that therefore, Wikimedia's motion to compel certain information in discovery had to be denied. *Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 790 (D. Md. 2018). The seven categories of

information determined to be privileged under the state secrets doctrine in relation to plaintiff's motion to compel discovery are the same categories of information at issue for plaintiff to establish standing via further litigation of this case.[56] Thus, as already established in the previous Memorandum Opinion and Order, the first two steps of the state secrets analysis have been resolved, and the step that remains is "how the matter should proceed in light of the successful privilege claim." *Abilt*, 848 F.3d at 311.

B.

How the matter should proceed turns on the centrality of the privileged information to the issue at hand. Whether the NSA has copied and collected any of Wikimedia's international Internet communications, or such collection is certainly impending, or there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection, is the threshold issue for Wikimedia to establish standing in this litigation. Where, as here, the privileged information is so central to the subject matter of the litigation, dismissal is the appropriate, and only available, course of action.

As the Fourth Circuit has made quite clear, "both Supreme Court precedent and our own cases provide that when a judge has satisfied himself [or herself] that the dangers asserted by the government are substantial and real, he [or she] need not—indeed, should not—probe further." *Sterling v. Tenet*, 416 F.3d 338, 345 (4th Cir. 2005). Moreover, Fourth Circuit precedent establishes that where "circumstances make clear that sensitive military secrets will be so central to the subject matter of the litigation that any attempt to proceed will threaten disclosure of the

---

[56] The seven categories of information privileged pursuant to the state secrets doctrine are: (i) individuals or entities subject to Upstream surveillance activities, (ii) operational details of the Upstream collection process, (iii) locations at which Upstream surveillance is conducted, (iv) categories of Internet-based communications subject to Upstream surveillance activities, (v) the scope and scale on which Upstream surveillance is or has been conducted, (vi) the NSA's cryptanalytic capabilities, and (vii) additional categories of classified information contained in FISC opinions, orders and submissions.

privileged matters, dismissal is the appropriate remedy." *El-Masri v. Tenet*, 437 F. Supp. 2d 530, 538-39 (E.D. Va. 2006) (quoting *Sterling*, 416 F.3d at 348), *aff'd*, 479 F.3d 296 (4th Cir. 2007).[57]

As such, "[i]f a proceeding involving state secrets can be fairly litigated without resort to the privileged information, it may continue." *El–Masri*, 479 F.3d at 306. On the other hand, "a proceeding in which the state secrets privilege is successfully interposed must be dismissed if the circumstances make clear that privileged information will be so central to the litigation that any attempt to proceed will threaten that information's disclosure." *Id.* at 308 (citations omitted).[58] Such a decision is never taken lightly, as "dismissal is appropriate '[o]nly when no amount of effort and care on the part of the court and the parties will safeguard privileged material.'" *Sterling*, 416 F.3d at 348 (quoting *Fitzgerald v. Penthouse Int'l, Ltd.*, 776 F.2d 1236, 1244 (4th Cir. 1985)) (alteration in original). Nonetheless, "dismissal follows inevitably when the sum and substance of the case involves state secrets." *Id.* at 347. In this regard, the Fourth Circuit has identified three examples of circumstances in which the privileged information is so central to the litigation that dismissal is required. First, "dismissal is required if the plaintiff cannot prove the *prima facie* elements of his or her claim without privileged evidence." *Abilt*, 848 F.3d at 313-14 (citing *Farnsworth Cannon, Inc. v. Grimes*, 635 F.2d 268, 281 (4th Cir. 1980) (en banc) (per curiam)). Second, "even if the plaintiff can prove a *prima facie* case without resort to privileged information, the case should be dismissed if 'the defendants could not properly defend themselves without using privileged evidence.'" *Id.* at 314 (quoting *El-Masri*, 479 F.3d at 309).

---

[57] Importantly, "state secrets and military secrets are equally valid bases for invocation of the evidentiary privilege." *Sterling*, 416 F.3d at 343 (internal quotation marks and alterations omitted).

[58] *See also Sterling*, 416 F.3d at 347–48 ("We have long recognized that when 'the very subject of [the] litigation is itself a state secret,' which provides 'no way [that] case could be tried without compromising sensitive military secrets,' a district court may properly dismiss the plaintiff's case." (quoting *Fitzgerald*, 776 F.2d at 1243) (alterations in original)); *Bowles v. United States*, 950 F.2d 154, 156 (4th Cir. 1991) (per curiam) ("If the case cannot be tried without compromising sensitive foreign policy secrets, the case must be dismissed.").

Third, "dismissal is appropriate where further litigation would present an unjustifiable risk of disclosure" of state secrets. *Id.* (citing *El-Masri*, 479 F.3d at 308).

C.

Given these principles and given "the delicate balance to be struck in applying the state secrets doctrine," it is appropriate to analyze the litigation at hand, namely the centrality of state secrets to Wikimedia's standing. *El-Masri*, 479 F.3d at 308. To establish standing, Wikimedia must prove (1) injury-in-fact, (2) causation, and (3) redressability. Through an extensive jurisdictional discovery process, Wikimedia has established that the NSA monitors at least one circuit carrying international Internet communications in the course of Upstream surveillance and that Wikimedia's communications traverse every circuit carrying international Internet communications from the United States to the rest of the world. Importantly, this extensive jurisdictional discovery process has resulted in the compilation of a voluminous record, including hundreds of pages of expert reports, government disclosures and declassified documents regarding Upstream surveillance, Rule 30(b)(6) testimony from an NSA representative, and extensive interrogatory responses from the parties. Thus, Wikimedia has been granted the opportunity to establish its standing without resort to privileged information, and Wikimedia has made significant progress on that front.

Nonetheless, the summary judgment record does not establish that the NSA has copied or collected any of Wikimedia's communications via Upstream surveillance conducted on an NSA-monitored circuit, that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Wikimedia has been unable to make this showing because it is not true, as a technological necessity, that the NSA must be copying every text-based communication that traverses a circuit that the NSA

monitors. Indeed, Dr. Schulzrinne has convincingly demonstrated that there are technologically feasible methods by which the NSA could hypothetically operate Upstream surveillance that would result in the NSA not copying or collecting any of Wikimedia's communications. Thus, the undisputed summary judgment record establishes that Wikimedia does not have Article III standing sufficient to survive summary judgment.

Even if Wikimedia could establish a *prima facie* case of its standing based solely on the public, unclassified record, which it has not been able to do thus far in this case, the state secrets doctrine still requires dismissal because the defendants cannot properly defend themselves without using privileged evidence. The Fourth Circuit "ha[s] consistently upheld dismissal when the defendants could not properly defend themselves without using privileged information." *Abilt v. CIA*, 848 F.3d 305, 316 (4th Cir. 2017). As in *El-Masri*, "virtually any conceivable response to [Wikimedia's] allegations [that the NSA has copied and collected some of Wikimedia's international Internet communications] would disclose privileged information." *El-Masri*, 479 F.3d at 310. Defendants have provided a detailed and persuasive explanation, in more than 60 pages of classified declarations, that disclosure of the entities subject to Upstream surveillance activity and the operational details of the Upstream collection process would (i) undermine ongoing intelligence operations, (ii) deprive the NSA of existing intelligence operations, and significantly, (iii) provide foreign adversaries with the tools necessary both to evade U.S. intelligence operations and to conduct their own operations against the United States and its allies. *Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 789 (D. Md. 2018). Accordingly, defendants could not properly defend themselves in any further litigation of Wikimedia's standing, and thus, the case must be dismissed.

Moreover, if the issue of Wikimedia's standing were further adjudicated, "the whole

object of the [adjudication]…[would be] to establish a fact that is a state secret," presenting an unjustifiable risk of disclosing privileged information. *Sterling*, 416 F.3d at 348. Courts have concluded that where, as here, the information sought to be disclosed involves the identity of parties whose communications have been acquired, this information is properly privileged. *See Al-Haramain Islamic Found., Inc. v. Bush*, 507 F.3d 1190, 1203-04 (9th Cir. 2007) (finding that the fact of a plaintiff's surveillance by the NSA was covered by the state secrets privilege); *Halkin v. Helms*, 598 F.2d 1, 9 (D.C. Cir. 1978) (upholding assertion of state secrets privilege with respect to "the identity of particular individuals whose communications have been acquired"). Accordingly, because the privileged information, namely the operational details of the Upstream collection process and whether any of Wikimedia's international Internet communications have been copied or collected by the NSA, is so central to the litigation of Wikimedia's standing, the case must be dismissed, and judgment must be entered in favor of defendants.

VII.

To avoid the conclusion that the case must be dismissed, Wikimedia revives its argument that 50 U.S.C. § 1806(f) displaces the state secrets doctrine in cases challenging electronic surveillance pursuant to FISA and provides for *in camera* review of the materials related to the NSA's Upstream surveillance program. This argument, however, has already been considered and rejected in this litigation. *See Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 786 (D. Md. 2018). Specifically, the "§ 1806(f) procedures do not apply where, as here, a plaintiff has not yet established that it has been the subject of electronic surveillance" as required by the statute. *Id.* at 780. Nonetheless, plaintiff raises two additional arguments as to why *in camera* review pursuant to § 1806(f) is appropriate in this case: (i) plaintiff has now established a

genuine dispute of material fact concerning its status as an "aggrieved person"[59] before invoking FISA's procedures and (ii) the Ninth Circuit recently held that § 1806(f) displaces the state secrets privilege in an affirmative legal challenge to electronic surveillance pursuant to FISA. *See Fazaga v. Fed. Bureau of Investigation*, 916 F.3d 1202 (9th Cir. 2019).

First, plaintiff has not established a genuine dispute of material fact concerning its status as an aggrieved person, *i.e.*, that plaintiff's communications have been the subject of electronic surveillance, as discussed *supra* Part V.C. As previously explained, "the text of § 1806(f) points persuasively to the conclusion that Congress intended § 1806(f) procedures to apply only after it became clear from the factual record that the movant was the subject of electronic surveillance." *Wikimedia Found.*, 335 F. Supp. 3d at 781. To be sure, "affirmative government acknowledgement of surveillance of a specific target is not the only means by which a plaintiff can establish evidence of his or her 'aggrieved person' status." *Id.* at 784. But here, despite the extensive jurisdictional discovery undertaken in this case, plaintiff has been unable to make a factual showing that Wikimedia was the subject of electronic surveillance using admissible record evidence. Thus, the §1806(f) *in camera* review procedures remain inapplicable to this case.

In addition, no binding authority establishes that § 1806(f)'s review procedures displace the state secrets doctrine even if a plaintiff survived summary judgment on the issue of whether plaintiff has been the target of electronic surveillance, which again is not the case here. Specifically, in *ACLU Foundation of Southern California v. Barr*, 952 F.2d 457 (D.C. Cir. 1991), the D.C. Circuit reasoned that "legitimate concerns about compromising ongoing foreign

---

[59] For the purposes of FISA, an "aggrieved person" is "a person who is the target of an electronic surveillance or any other person whose communications or activities were subject to electronic surveillance." 50 U.S.C. § 1801(k).

intelligence investigations" are more properly considered at the summary judgment stage, not upon the pleadings. *Id.* at 469. In doing so, the D.C. Circuit only considered what a party must show to establish his or her "aggrieved person" status and therefore invoke § 1806(f) review. Simply put, the D.C. Circuit did not consider whether or when § 1806(f) *in camera* review is inappropriate or unnecessary because of the state secrets doctrine.

Moreover, the Ninth Circuit's opinion in *Fazaga* does not hold that § 1806(f) displaces the state secrets doctrine in this case, despite plaintiff's arguments to the contrary. The Ninth Circuit reasoned in *Fazaga* that § 1806(f)'s procedures displace a dismissal remedy for the *Reynolds* state secrets doctrine *only where § 1806(f)'s procedures apply.*[60] *Fazaga*, 916 F.3d at 1234. Specifically, the Ninth Circuit held that for FISA's § 1806(f) procedures to apply, "[p]laintiffs must satisfy the definition of an 'aggrieved person.'" *Id.* at 1238. In this case, as previously discussed at length, Wikimedia has not established it is an "aggrieved person" as defined in § 1801(k). *See Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 780, 786 (D. Md. 2018). Thus, § 1806(f) does not apply to this case, and dismissal on state secrets grounds is appropriate, as discussed *supra* Part VI.

Notably, the only court to address this issue post-*Fazaga* held that "where the very issue

---

[60] *Fazaga* addressed a challenge to an allegedly unlawful FBI counter-terrorism investigation involving electronic surveillance. *Id.* at 1210-11. Specifically, in that case, "several sources" confirmed the identity of a confidential FBI informant and disclosed that that specific confidential informant "created audio and visual recordings" for the FBI. *Id.* at 1214. The district court dismissed all but one of plaintiff's claims at the pleading stage without further discovery based on the government's assertion of the state secrets privilege. *Id.* at 1211. The Ninth Circuit reversed, concluding that § 1806(f)'s procedures are to be used when "aggrieved persons" challenge the legality of electronic surveillance and that the district court erred by dismissing the case without reviewing the evidence. *Id.* at 1238, 1252. In remanding for further proceedings, the *Fazaga* court held that "[t]he complaint's allegations are sufficient *if proven* to establish that Plaintiffs are 'aggrieved persons.'" *Id.* at 1216 (emphasis added). Thus, the Ninth Circuit's decision reasoned that at the pleading stage of the litigation, where plaintiffs have alleged sufficient facts, assumed to be true at that stage of the litigation, to establish they are "aggrieved persons" as required for application of Section 1806(f), dismissal on the basis of the state secrets doctrine was inappropriate. This holding says nothing, however, about the relationship between § 1806(f) and the state secrets doctrine dismissal remedy where, as here, a plaintiff has not established that he, she, or it is an "aggrieved person" using admissible record evidence, after a lengthy jurisdictional review process, at the summary judgment stage of the litigation.

of standing implicates state secrets," the holding in *Fazaga* and § 1806(f) do not foreclose "dismissing [the case] on state secrets grounds" at the summary judgment stage of the litigation.[61] *Jewel v. Nat'l Sec. Agency*, No. C 08–04373, at *24 (N.D. Cal. April 25, 2019), *appeal docketed*, No. 19–16066 (9th Cir. May 21, 2019). Accordingly, because plaintiff has not established it is an "aggrieved person" as defined in the statute, and hence § 1806(f) does not apply, and because the issue of standing in this case necessarily implicates state secrets, dismissal of the case is appropriate.

VIII.

To avoid dismissal of the litigation on state secrets grounds, Wikimedia has raised several additional standing arguments separate and apart from the Wikimedia Allegation—namely (i) Upstream surveillance has impaired Wikimedia's communications with its community members, (ii) Upstream surveillance has required Wikimedia to take costly protective measures, and (iii) Wikimedia has third-party standing to assert the rights of its users. Wikimedia's arguments fail as to each of these theories of standing for the reasons discussed below.

First, Wikimedia argues it has standing because Upstream surveillance has impaired Wikimedia's communications with its community members, as evidenced by a drop in the readership of certain Wikipedia pages. In *Clapper* and *Laird*, however, the Supreme Court unequivocally held that "[a]llegations of a subjective 'chill' are not an adequate substitute for a claim of specific present objective harm or a threat of specific future harm." *Clapper v. Amnesty*

---

[61] To be sure, the district court in California did review "classified evidence submitted by Defendants in response to Plaintiffs' discovery requests" pursuant to the procedures of § 1806(f) of FISA prior to its summary judgment ruling dismissing the case. *Id.* at *24-25. That court did not, however, consider the question of whether plaintiffs were "aggrieved persons" prior to undertaking § 1806(f)'s procedures for *in camera* review. Nevertheless, that court still found that where, as here, "the answer to the question of whether a particular plaintiff was subjected to surveillance – *i.e.*, is an 'aggrieved person' under Section 1806(f) – is the very information over which the Government seeks to assert the state secrets privilege," dismissal of the case and entry of judgment in favor of the government is the appropriate action at summary judgment. *Id.* at *23, *25.

Int'l USA, 133 S. Ct. 1138, 1152 (2013) (quoting *Laird v. Tatum*, 408 U.S. 1, 13-14 (1972)). To avoid the conclusion that any drop in readership is the result of a "subjective chill," Wikimedia relies upon a statistical analysis performed by Dr. Jonathon Penny, which concludes it is "highly likely" that "public awareness of NSA surveillance programs, including Upstream surveillance,…ha[s] had a large-scale chilling effect on Wikipedia users" since June 2013. Dr. Jonathon Penney Decl. ¶ 10-11. But Dr. Penney's conclusion that Wikipedia's readership has suffered an actual chill as the result of Upstream surveillance is undermined for two principal reasons. First, Dr. Penney's data only covers a 32-month period which ends in August 2014, before this lawsuit was even filed. Thus, Dr. Penney's evidence, even if reliable, does not say anything about any ongoing harm suffered by Wikimedia that is traceable to Upstream surveillance. Second, these alleged readership effects were from public awareness of "media coverage of NSA surveillance" generally, not Upstream surveillance specifically. *Id.* at ¶ 26. Thus, Dr. Penney's findings do not demonstrate an ongoing and sustained drop in Wikimedia's readership stemming from the NSA's Upstream surveillance program.

Moreover, "a 'chilling effect aris[ing] merely from the individual's knowledge that a governmental agency was engaged in certain activities or from the individual's concomitant fear that, armed with the fruits of those activities, the agency might in the future take some other and additional action detrimental to that individual'" is insufficient to establish standing.[62] *Clapper*, 133 S. Ct. at 1152 (quoting *Laird*, 408 U.S. at 11). This is exactly the situation here—Wikimedia claims that this decreased readership is a result of individual's fear that the government might be

---

[62] It is worth noting that the Fourth Circuit and the Supreme Court have explained that "standing requirements are somewhat relaxed in First Amendment cases." *Cooksey v. Futrell*, 721 F.3d 226, 235 (4th Cir. 2013) (citing *Secretary of State of Md. v. Joseph H. Munson Co., Inc.*, 467 U.S. 947, 956 (1984)). Even though the standing requirements are somewhat relaxed in the First Amendment context, subjective and speculative fears of government surveillance, such as in this case, do not establish Article III standing at summary judgment, as the Supreme Court specifically held in *Clapper* and *Laird*. *See Clapper*, 133 S. Ct. at 1151-52; *Laird*, 408 U.S. at 10-15.

monitoring their Internet activity and might use that information at some later date. Moreover, the Supreme Court has specifically found that a claimed reluctance by third parties to communicate with a plaintiff, due to their subjective fears of surveillance, is not fairly traceable to the alleged surveillance, and is thus foreclosed as a basis for standing. *Clapper*, 133 S. Ct. at 1152 n.7. Accordingly, Wikimedia cannot establish standing under this theory given the Supreme Court's holdings in *Clapper* and *Laird*.

Second, Wikimedia argues it has standing because Upstream surveillance has required Wikimedia to take costly protective measures—namely, transitioning its Internet communications into encrypted formats such as HTTPS and IPSec, acquiring new technical infrastructure, and hiring a full-time engineer to manage the protective measures. The Supreme Court has already foreclosed this alternative theory of standing where, as here, a plaintiff has failed to establish that their communications have been collected by the government, or that such collection is certainly impending. *Clapper*, 133 S. Ct. at 1151. Applicable here is the Supreme Court's statement in *Clapper* that a plaintiff "cannot manufacture standing merely by inflicting harm on themselves based on their fears of hypothetical future harm that is not certainly impending." *Id.*

Wikimedia attempts to distinguish this case from *Clapper* by arguing that the harm Wikimedia faces from Upstream surveillance is well-established, not some "hypothetical future harm." As discussed at length *supra* in Part V, however, the summary judgment record does not establish that Wikimedia's communications have been collected by the NSA during Upstream surveillance, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Thus, any harm to Wikimedia from the Upstream surveillance program remains a purely hypothetical harm

insufficient to establish standing. As the Supreme Court has sensibly observed, to find otherwise "would be tantamount to accepting a repackaged version of [plaintiff's] first failed theory of standing," namely the Wikimedia Allegation. *Id.* (citing *Am. Civil Liberties Union v. Nat'l Sec. Agency*, 493 F.3d 644, 655–56 (6th Cir. 2007)). Accordingly, Wikimedia's alleged expenditures to protect its communications from Upstream surveillance collection do not establish its standing.[63]

Third, Wikimedia argues it has third party standing to assert the rights of its users. In the Fourth Circuit, a plaintiff must demonstrate "(1) an injury-in-fact; (2) a close relationship between [itself] and the person whose right [it] seeks to assert; and (3) a hindrance to the third party's ability to protect his or her own interests" to "overcome the prudential limitation on third-party standing."[64] *Freilich v. Upper Chesapeake Health Inc.*, 313 F.3d 205, 215 (4th Cir. 2002) (citing *Powers v. Ohio*, 499 U.S. 400, 410–11 (1991)). Wikimedia has met none of these requirements. As discussed at length *supra* in Part V, Wikimedia has been unable to establish injury-in-fact in this case. In addition, Wikimedia has not presented admissible evidence that establishes a "close relationship" between Wikimedia and its largely unidentified contributors.[65]

---

[63] Moreover, without evidence that the alleged injuries from implementing these protective measures would be redressed by the injunctive relief plaintiff seeks, these alleged injuries cannot confer standing to sue. *See Clapper*, 568 U.S. at 409; *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560-61 (1992). Given the number of other reasons that plaintiff has admitted influenced its decision to implement these protective measures, including protecting against individual computer hackers and keeping their company policies up-to-date and transparent, injunctive relief enjoining the NSA from conducting the Upstream surveillance program would not redress any alleged injury from these protective expenditures. In fact, Wikimedia began the process of switching to HTTPS as early as 2011, years before any disclosures about the NSA's Upstream surveillance program. *See* ECF No. 178-8.

[64] As the Supreme Court has appropriately warned, "[f]ederal courts must hesitate before resolving a controversy, even one within their constitutional power to resolve, on the basis of the rights of third persons not parties to the litigation." *Singleton v. Wulff*, 428 U.S. 106, 113 (1976).

[65] Close relationships that have established third-party standing in the past include lawyer-client and doctor-patient. *See Department of Labor v. Triplett*, 494 U.S. 715 (1990) (lawyer-client); *Singleton v. Wulff*, 428 U.S. 106 (1976) (doctor-patient). Wikimedia's relationship with its unidentified contributors clearly does not rise to the level of those protected, close relationships.

In fact, Wikimedia has only presented declarations from one single contributor who has edited Wikimedia's web projects while abroad, and this single contributor has stated that her "workload as a medical student" makes it "impossible" for her to bring a lawsuit as a plaintiff.[66] Such "normal burdens of litigation," however, are insufficient to satisfy the third requirement that an obstacle exists that prevents the third party from bringing the lawsuit herself or himself.[67] *See Lawyers Ass'n v. Reno*, 199 F.3d 1352, 1364 (D.C. Cir. 2000). Thus, Wikimedia has also failed to satisfy the third requirement to establish third-party standing. Accordingly, Wikimedia's third-party standing argument clearly fails.

For the reasons stated above, Wikimedia's three additional standing arguments clearly fail because Wikimedia has not established an injury-in-fact using admissible record evidence and Wikimedia has not satisfied the strict requirements to proceed on the basis of third-party standing.

IX.

In sum, Wikimedia has failed to present specific facts which show that defendants, through the Upstream surveillance program, have copied and collected Wikimedia's international Internet communications, that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. More specifically, the summary judgment record establishes that it is not a technological necessity that the NSA must copy all of the text-based Internet communications

---

[66] Temple-Wood Decl. ¶ 26, ECF No. 168-10.

[67] Thus, Ms. Temple-Wood, a contributor to Wikimedia's free-knowledge projects, also states that "serving as a plaintiff in a lawsuit would threaten the anonymity [upon which Wikimedia] users depend." Temple-Wood Decl. ¶ 27, ECF No. 168-10. Although privacy and anonymity are valid concerns, in this case a putative plaintiff would not need to reveal the contents of their communications with Wikimedia in order to serve as a plaintiff; they would only need to disclose the form in which the communications were sent (*i.e.*, sending an email or accessing or editing a web project), and the location from which the communications were sent.

traversing a circuit that the NSA monitors while conducting Upstream surveillance. Thus, there is no genuine dispute of material fact that the NSA could conduct Upstream surveillance without collecting any of Wikimedia's communications, and Wikimedia has been unable to present specific facts that establish otherwise, largely because the necessary facts are protected by the state secrets privilege.

Moreover, even if Wikimedia had established a genuine issue of material fact as to whether the NSA has copied or collected any of its international Internet communications, which Wikimedia has not done on this record, further litigation of this matter is precluded by the state secrets doctrine, which has been properly invoked by defendants. The extensive jurisdictional discovery process in this case has made clear that the very issue of standing implicates state secrets and that despite plaintiff's valiant efforts, establishing standing solely on the basis of the public, unclassified record is not possible in this case. Pursuant to Supreme Court and Fourth Circuit precedent, at this stage of the litigation, namely summary judgment post-jurisdictional discovery, dismissal and entry of judgment in favor of defendants is the appropriate, and only available, remedy because the issue of standing in this case necessarily implicates state secrets.

It is important to acknowledge the unfortunate burden that this decision places on Wikimedia. *See Abilt*, 848 F.3d at 317; *Sterling*, 416 F.3d at 348; *El-Masri*, 479 F.3d at 313 ("As we have observed in the past, the successful interposition of the state secrets privilege imposes a heavy burden on the party against whom the privilege is asserted."). Wikimedia suffers dismissal of its claim "not through any fault of [its] own, but because [its] personal interest in pursuing [its] civil claim is subordinated to the collective interest in national security." *El-Masri*, 479 F.3d at 313; *see also Abilt*, 848 F.3d at 318; *Fitzgerald*, 776 F.2d at 1238 n.3 ("When the state secrets privilege is validly asserted, the result is unfairness to individual litigants—through the loss of

important evidence or dismissal of a case—in order to protect a greater public value."). It is appropriate, however, "in limited circumstances like these, [that] the fundamental principle of access to court must bow to the fact that a nation without sound intelligence is a nation at risk." *Sterling*, 416 F.3d at 348.

Plaintiff contends that a holding which finds plaintiff does not have standing and precludes further litigation of this matter because of defendants' invocation of the state secrets doctrine leads to the result that "the Executive Branch alone controls who can and cannot challenge unlawful surveillance."[68] This contention is incorrect; the Supreme Court addressed and rejected a similar argument in *Clapper*. There, the Supreme Court explained that Section 702 surveillance orders are not insulated from judicial review because (i) the FISC reviews the government's certifications, targeting procedures, and minimization procedures for Section 702 surveillance, including whether the targeting and minimization procedures comport with the Fourth Amendment, (ii) criminal defendants prosecuted on the basis of information derived from Section 702 surveillance are given notice of that surveillance and can challenge its validity, and (iii) electronic communications service providers directed to assist the government in surveillance may challenge the directive before the FISC. *Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1154 (2013). Even if those other avenues for judicial review were not available, the Supreme Court has made clear that "'[t]he assumption that if [plaintiff has] no standing to sue, no one would have standing, is not a reason to find standing.'" *Id.* (quoting *Valley Forge Christian College v. Americans United for Separation of Church and State, Inc.*, 454 U.S. 464, 489 (1982)).

Moreover, since this litigation began in 2015, FISA Section 702, pursuant to which the

---

[68] Plaintiff's Br. in Op. to Defs.' Motion for Summary Judgment, ECF No. 168, at 2.

NSA Upstream surveillance program operates, was reauthorized by Congress. FISA Section 702 was set to expire on December 31, 2017, but Congress voted in January 2018 to extend FISA Section 702 for an additional six years (the "FISA Amendment Reauthorization Act of 2017").[69] This reauthorization process sparked significant public debate, and the FISA Amendment Reauthorization Act of 2017 enacted a number of reforms to address the public's civil liberties concerns.[70]

Thus, rather than the executive branch alone controlling who can and cannot challenge unlawful surveillance, the judicial branch provides for review and oversight via the limited avenues outlined by the Supreme Court in *Clapper*, including the significant role of the FISC, and the legislative branch provides for review and oversight via the FISA reauthorization process and the executive branch's ongoing reporting requirements to Congress. These avenues are sufficient to meet Constitutional requirements while at the same time precluding the unnecessary disclosure of state secrets.

* * *

For the reasons set forth above, this case must be dismissed, and judgment must be entered for defendants.

An appropriate order will issue separately.

---

[69] FISA Amendments Reauthorization Act of 2017, PL 115-118, January 19, 2018, 132 Stat 3.

[70] For example, the FISA Amendment Reauthorization Act of 2017 added a requirement that the DNI adopt procedures consistent with the requirements of the Fourth Amendment for querying information collected pursuant to Section 702 authority and made these querying procedures subject to FISC review. *See id.* at Sec. 101 Querying Procedures Required. The FISA Amendment Reauthorization Act of 2017 also restricted the use of U.S. person information obtained under Section 702 as evidence in a criminal proceeding and amended the mandatory reporting requirements to require the release of information on the breakdown of U.S. and non-U.S. person targets of electronic surveillance. *See id.* at Sec. 102. These represent only a few of a number of reforms enacted by the FISA Amendment Reauthorization Act of 2017. These reforms, combined with the short period of reauthorization, demonstrate the legislative branch's focused oversight of the executive branch's Section 702 authority.

The Clerk is directed to provide a copy of this Opinion to all counsel of record.

Alexandria, Virginia
December 13, 2019

/s/
T. S. Ellis, III
United States District Judge

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 1:15-cv-662**
	)	
NATIONAL SECURITY AGENCY/ CENTRAL SECURITY SERVICE, *et al.*,	)	
	)	
Defendants.	)	

ORDER

For the reasons stated in the accompanying Memorandum Opinion,

It is hereby **ORDERED** that defendants' motion for summary judgment (Dkt. 161) is **GRANTED.**

Accordingly, it is further **ORDERED** that this matter is **DISMISSED without prejudice.**

The Clerk is directed to enter Rule 58 judgment on behalf of defendants and against plaintiff and place this matter among the ended causes.

The Clerk is further directed to send a copy of this Order to all counsel of record.

Alexandria, Virginia
December 13, 2019

/s/
T. S. Ellis, III
United States District Judge

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY, *et al.*, *Defendants*.	Hon. T.S. Ellis, III Civil Action No. 15-cv-662-TSE

NOTICE OF APPEAL

Wikimedia Foundation, Plaintiff in the above-captioned case, hereby appeals to the United States Court of Appeals for the Fourth Circuit from the final judgment entered in this action on the 17th day of December, 2019, granting Defendants' motion for summary judgment.

Dated: February 14, 2020

Respectfully submitted,

/s/ Deborah A. Jeon
Deborah A. Jeon (Bar No. 06905)
David R. Rocah (Bar No. 27315)
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF MARYLAND
3600 Clipper Mill Rd., #350
Baltimore, MD 21211
Phone: (410) 889-8555
Fax: (410) 366-7838
jeon@aclu-md.org

Benjamin H. Kleine (pro hac vice)
COOLEY LLP
101 California Street, 5th Floor
San Francisco, CA 94111
Phone: (415) 693-2000
Fax: (415) 693-2222
bkleine@cooley.com

Counsel for Plaintiff

/s/ Patrick Toomey
Patrick Toomey (pro hac vice)
(*signed by Patrick Toomey with permission of Deborah A. Jeon*)
Ashley Gorski (pro hac vice)
AMERICAN CIVIL LIBERTIES UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
ptoomey@aclu.org

Alex Abdo (pro hac vice)
Jameel Jaffer (pro hac vice)
KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY
475 Riverside Drive, Suite 302
New York, NY 10115
Phone: (646) 745-8500
alex.abdo@knightcolumbia.org