



**IN THE UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF MARYLAND**

<hr/>		)	
WIKIMEDIA FOUNDATION,		)	
		)	
Plaintiff,		)	
		)	
v.		)	No. 1:15-cv-0662 (TSE)
		)	
NATIONAL SECURITY AGENCY, <i>et al.</i> ,		)	
		)	
Defendants.		)	
<hr/>		)	

## 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.”<sup>1</sup> My review analyzes the data, methodology, and conclusions presented in the Penney Declaration.<sup>2</sup>
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.

---

<sup>1</sup> Penney Declaration, paragraph 10.

<sup>2</sup> The Penney Declaration, in paragraphs 12 through 21, describes research on chilling effects theory. The Penney 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.”<sup>3</sup> 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.”<sup>4</sup> These three overlapping datasets contain page views for Wikipedia articles from January 2012 through August 2014 (“study period”).<sup>5</sup> 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.”<sup>6</sup> 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 “Hammas.” I will call this dataset “Terror 48 without Hammas.” 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.<sup>7</sup>

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”);<sup>8</sup> 25 domestic-security related articles (“Security 25”); 34 infrastructure articles (“Infrastructure 34”); and 26 popular (“Popular 26”) articles.<sup>9</sup>
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.<sup>10</sup>
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

---

<sup>3</sup> Penney Declaration, paragraph 10.

<sup>4</sup> Penney Declaration, paragraph 31.

<sup>5</sup> Penney Declaration, paragraph 34.

<sup>6</sup> 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.

<sup>7</sup> 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.

<sup>8</sup> 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.

<sup>9</sup> Penney Declaration paragraphs 52-56 describe the Popular, Infrastructure, and Security articles.

<sup>10</sup> 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<sup>11</sup> 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.<sup>12</sup>

**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.<sup>13</sup> 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.”<sup>14</sup>
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.

---

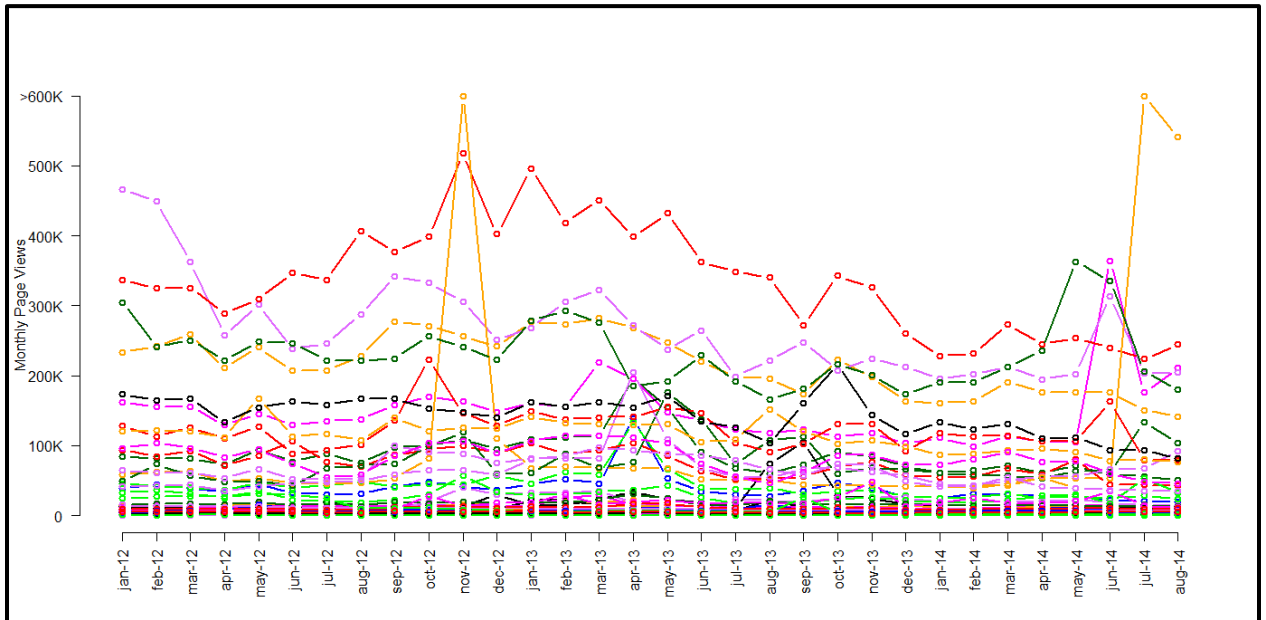
<sup>11</sup> The analysis covers all 48 articles but the conclusions made in the Penney Declaration apply only to 47 (the Terror 48 minus Hamas set of articles) and 31 (the High Privacy 31) of those articles.

<sup>12</sup> See Penney Declaration, paragraph 11.

<sup>13</sup> 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.

<sup>14</sup> 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.

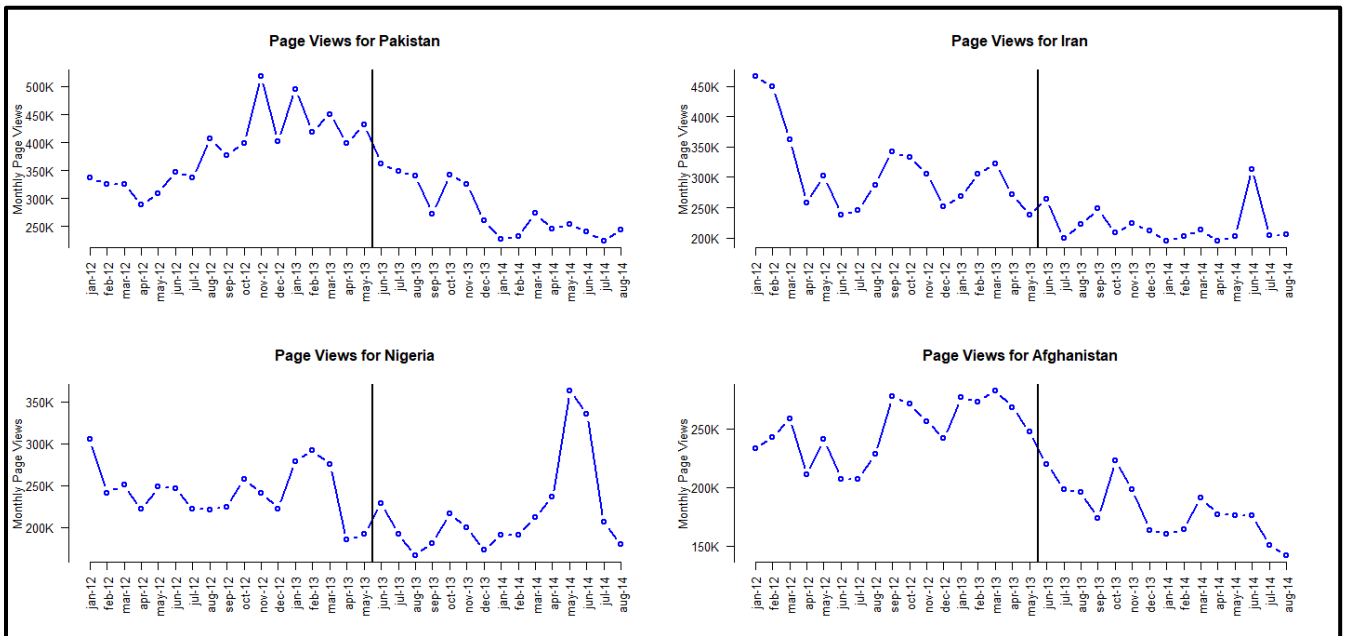
**Figure 1: Individual Page Views for Each of the Articles Within the Terror 48, Which The Penney Declaration Hypothesized Show an Immediate Decline Beginning in June 2013**



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.<sup>15</sup> Given those vast differences, it is not statistically appropriate to combine them for the purposes of analysis, as Dr. Penney did in his analysis.

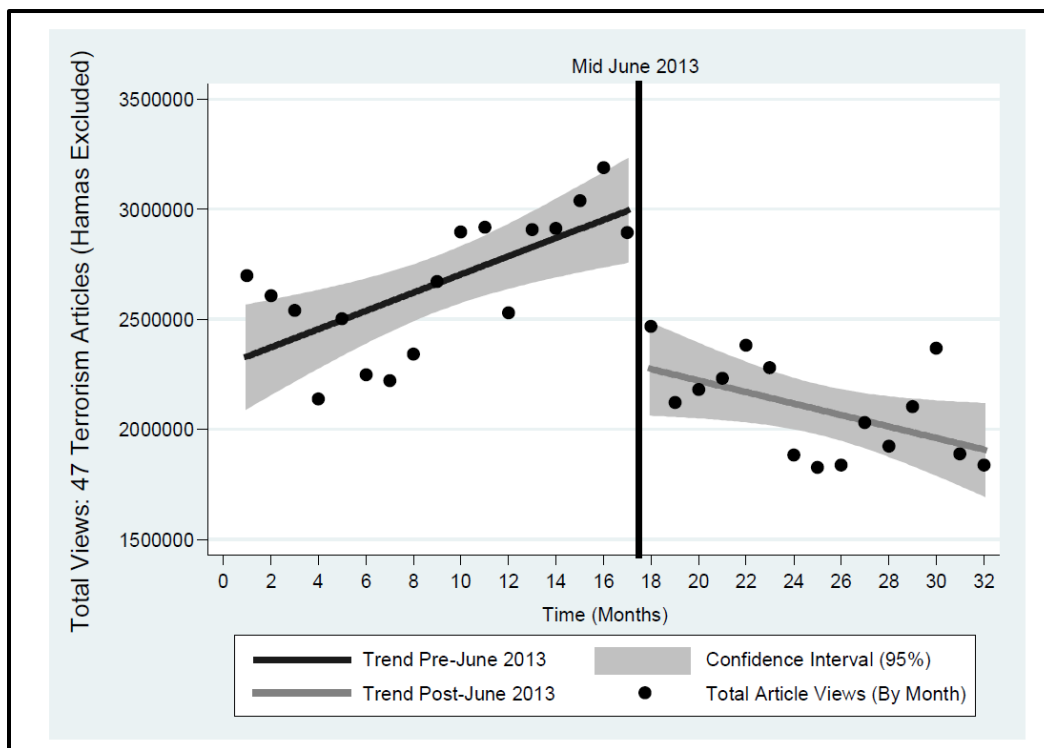
17. As I explain in Section F below, ignoring these differences biases the model and renders it invalid. The simple reason is that such aggregation masks the individual differences in page views. Although aggregation can be appropriate in instances where most of the data tell a consistent and similar 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.

<sup>15</sup> 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**

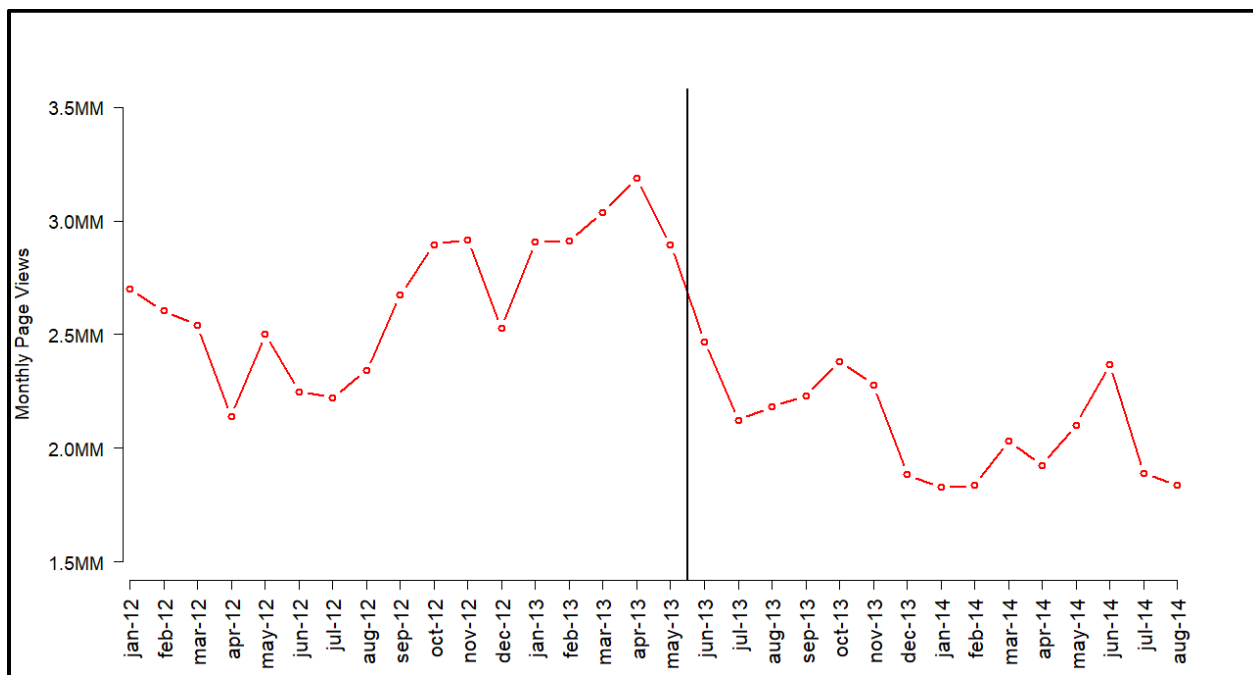


19. The suggestive trend lines in the Penney Declaration’s figure give the impression of a steady increase followed by a decrease, but the points, representing individual months, reveal otherwise. Careful attention to Figure 2 in the Penney Declaration reveals that the page views went up and down several 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.

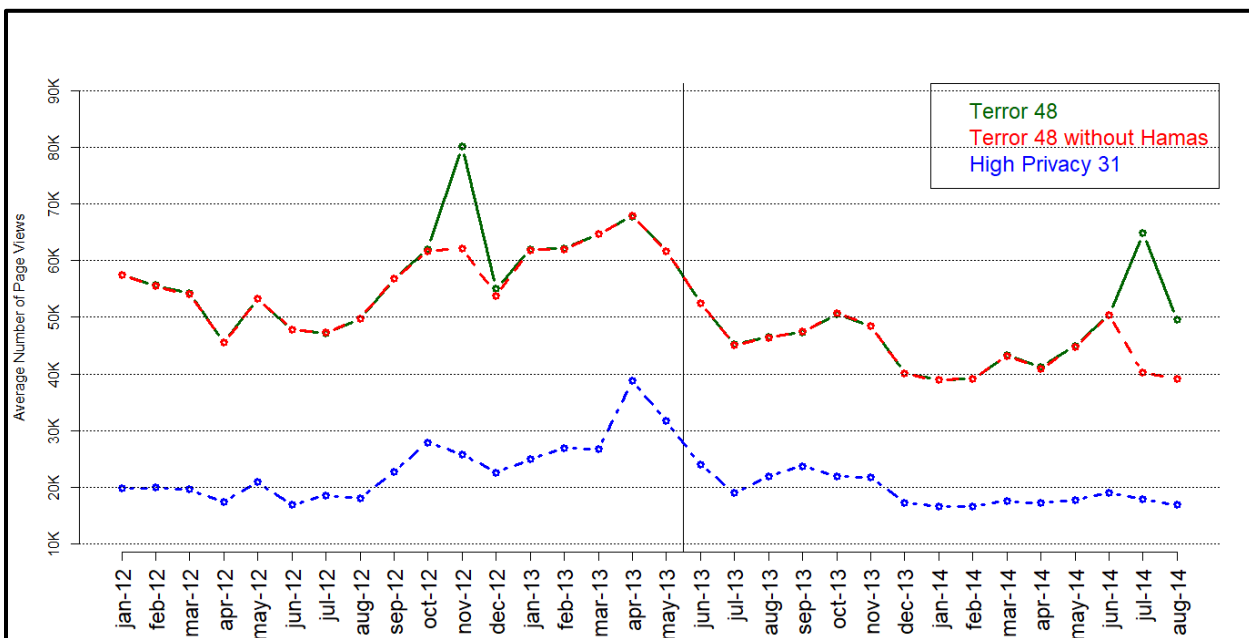
**Figure 4: Terror 48 Without Hamas Dataset Without the Penney Declaration “Trend” Lines**



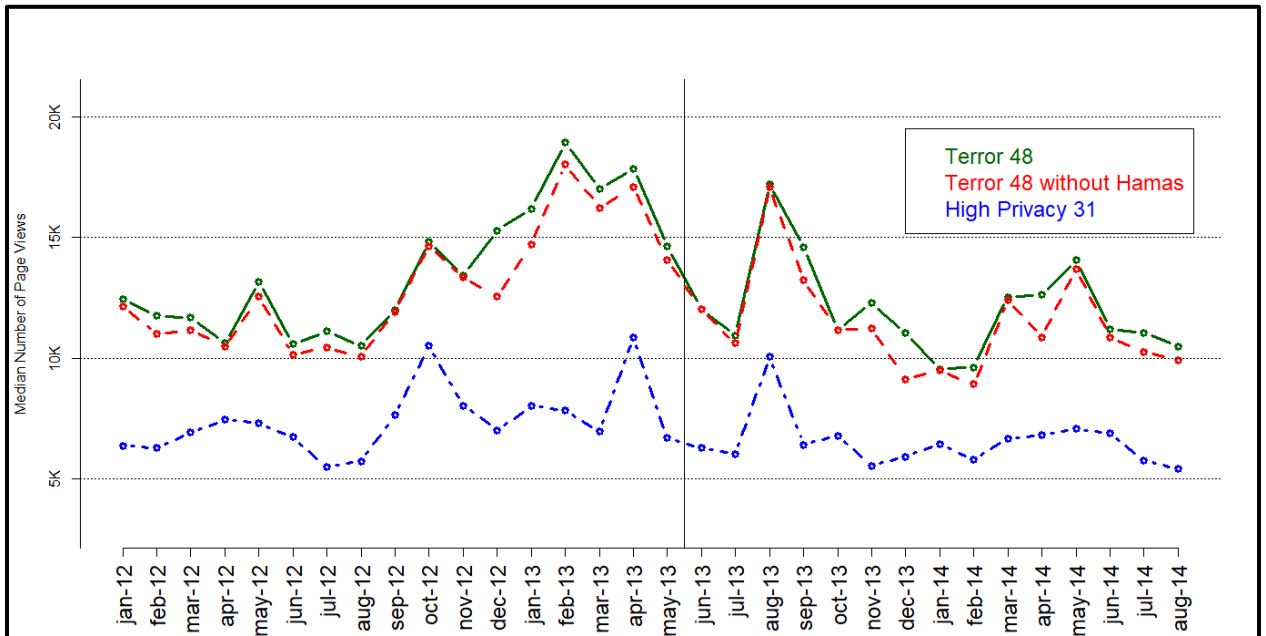
23. Figure 5 below adds the other two datasets analyzed (Terror 48 and High Privacy 31) to the Terror 48 Without Hamas dataset graphed above, and I used the average page views per article rather than the sum.<sup>16</sup> Once again, Figure 5 indicates that the peak is in April 2013 (and prior to April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

<sup>16</sup> 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.<sup>17</sup>

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”).<sup>18</sup>

<sup>17</sup> 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.”

<sup>18</sup> 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."<sup>19</sup>
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.<sup>20</sup>

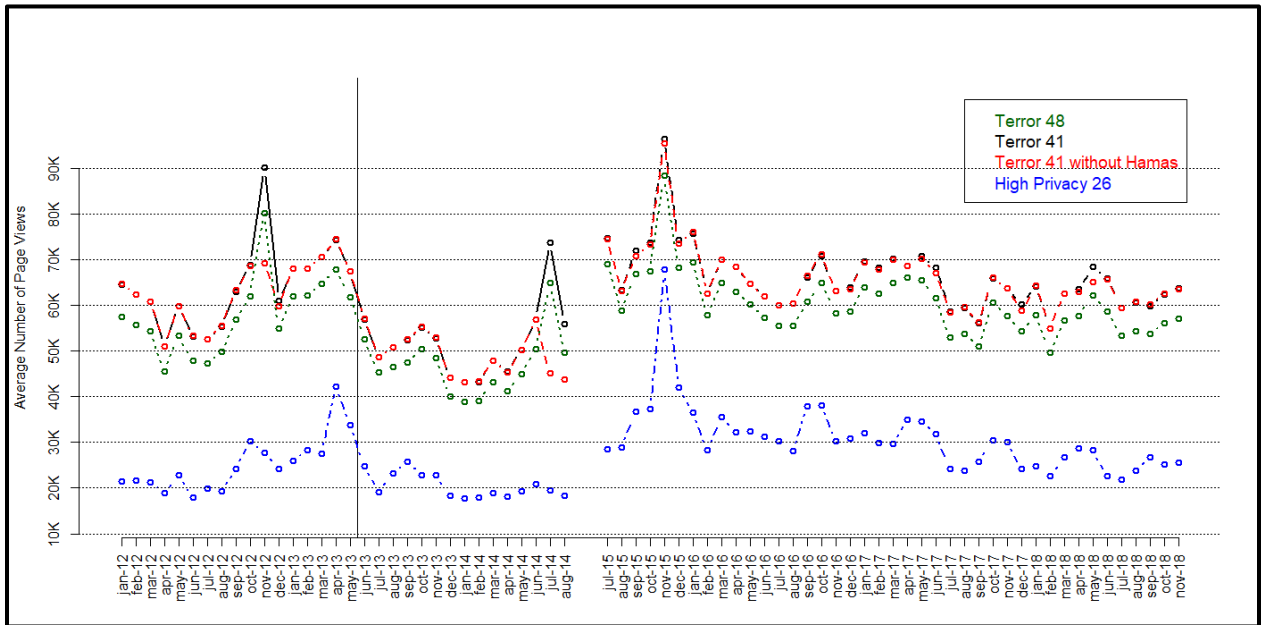
---

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).

<sup>19</sup> 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.

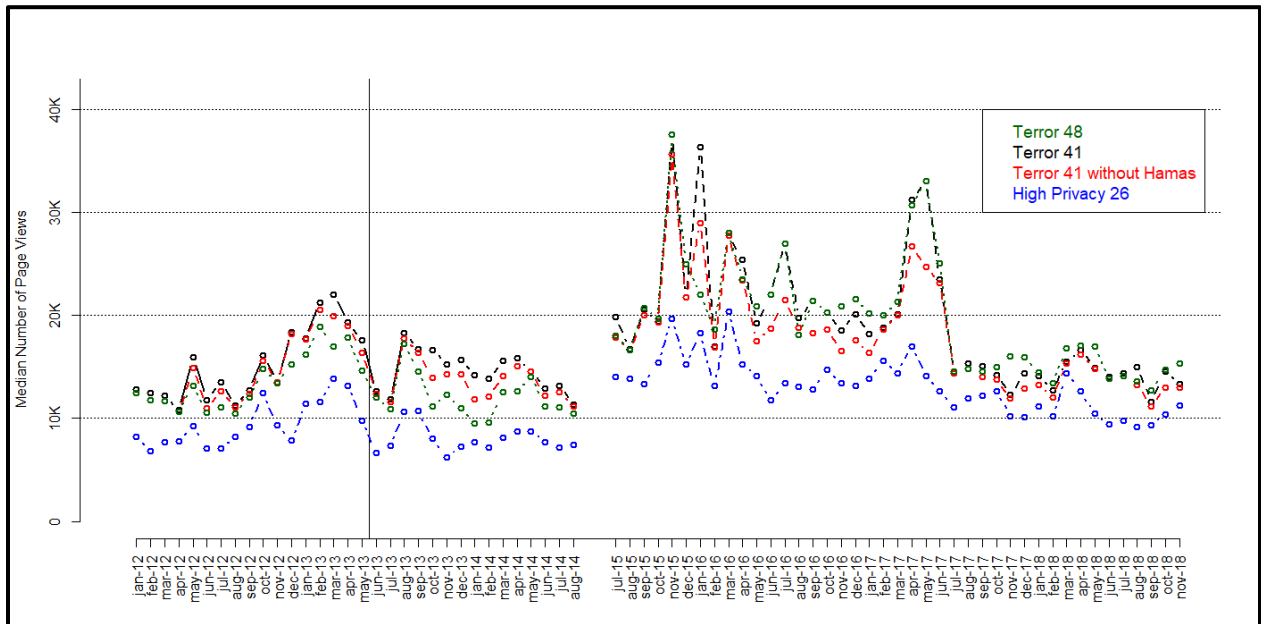
<sup>20</sup> As I will explain further below, the behavior of the aggregate data need not be indicative of the behavior of the individual article data. For example, the aggregate averages have a peak near the November 2015 Paris terror attacks, but that does not mean that all or most of the individual 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.<sup>21</sup> 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

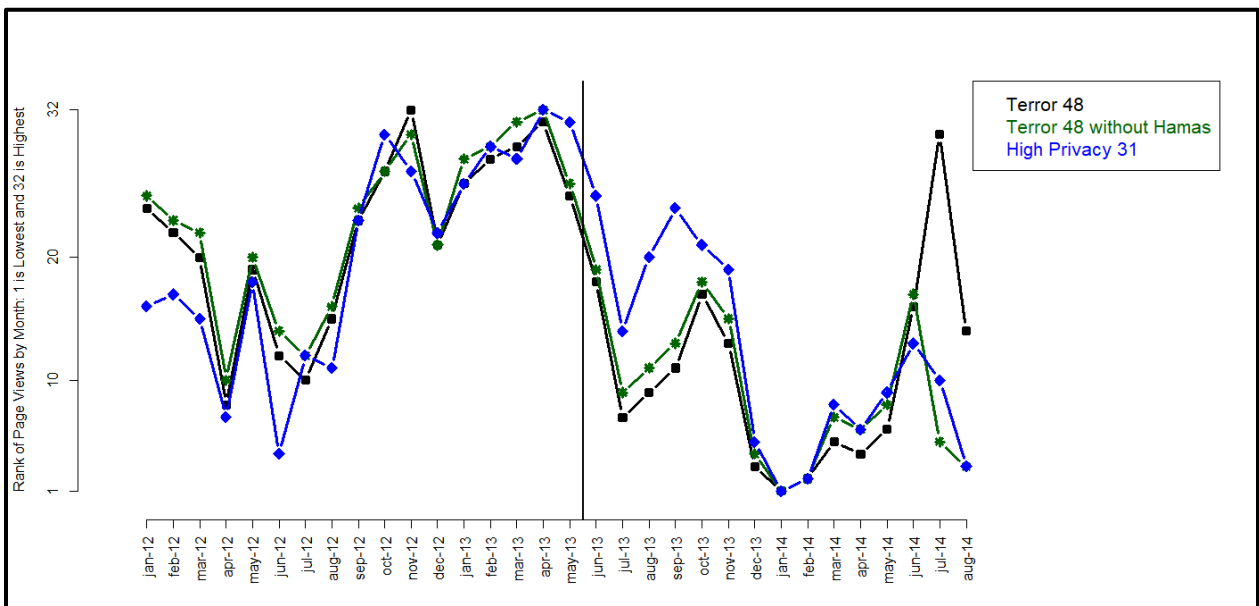
<sup>21</sup> These five datasets consist of “three comparator article groups” cited in paragraph 53 of the Penney Declaration as well as the two global view datasets of Wikipedia home page views used in the Penney Declaration. See my description of these datasets, above, in paragraph 8.



way over time), the estimated effects derived using such comparison groups could be wrong.<sup>22</sup>

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.<sup>23</sup> 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

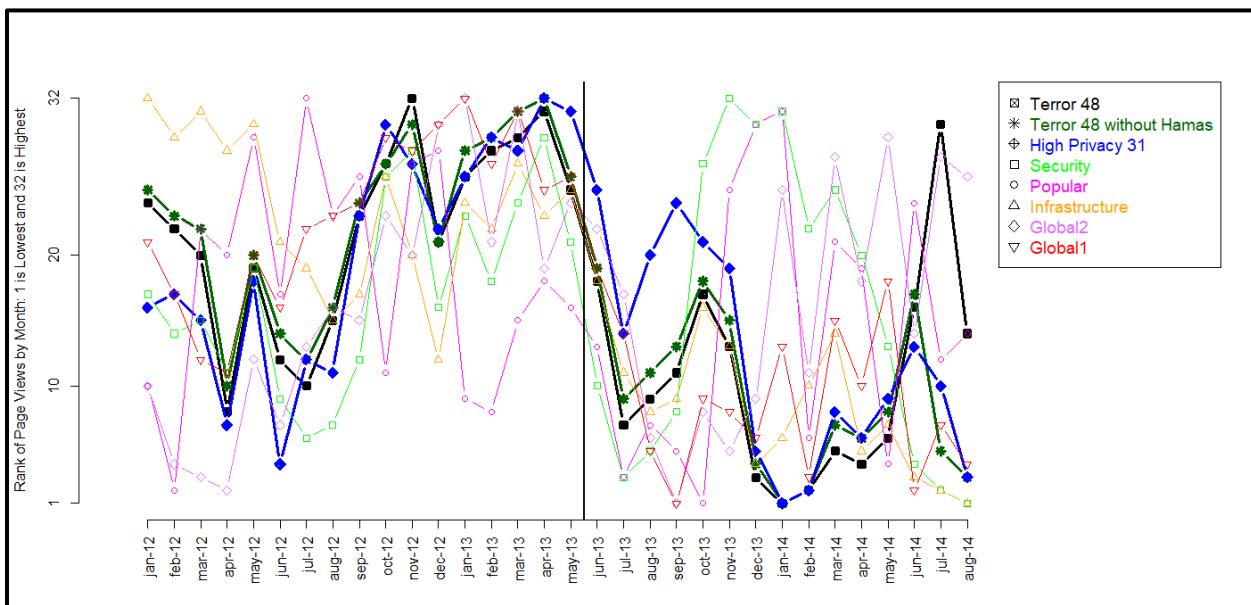
<sup>22</sup> See, for example, “Campbell, Donald, and Stanley, Julian C., Experimental and quasi-experimental Designs for Research, 1963, Houghton-Mifflin, p. 55-57. This issue is also discussed in Salzberg, Alan J., “Removable Selection Bias in Quasi-Experiments,” The American Statistician, 1999, pp. 103-107.

<sup>23</sup> 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	<b>41,420.51**</b> <i>p=0.00</i>	<b>-693,616.9**</b> <i>p=0.00</i>	<b>-67,513.1**</b> <i>p=0.00</i>	Yes <i>F=0.00</i>
25 Security Articles	11,135.0 <i>p=0.187</i>	-24,638.34 <i>p=0.84</i>	-20,465.87 <i>p=0.12</i>	No <i>F=0.45</i>
34 Infrastructure Articles	<b>-11,079**</b> <i>p=0.00</i>	-12,721.0 <i>p=0.77</i>	2,431.84 <i>p=0.61</i>	Yes <i>F=0.00</i>
26 Popular Articles	-48,458 <i>p=0.798</i>	-1,716,643 <i>p=0.53</i>	177,324.7 <i>p=0.551</i>	No <i>F=0.79</i>

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.<sup>24</sup> Both of these datasets show an increase through June 2013 followed by a decline after June 2013.<sup>25</sup> 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

<sup>24</sup> These are the datasets identified as Global 1 and Global 2 in paragraph 8, above.

<sup>25</sup> 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.<sup>26</sup> 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.<sup>27</sup> 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

---

<sup>26</sup> As with the terror datasets, the decline actually begins *before* the hypothesized month of June 2013.

<sup>27</sup> 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.<sup>28</sup> 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.<sup>29</sup> 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.<sup>30</sup> 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.<sup>31</sup> Thus, for example, the

---

<sup>28</sup> 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.

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

<sup>30</sup> The model also does not allow for there to be multiple peaks in the data.

<sup>31</sup> 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.<sup>32</sup> 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.<sup>33</sup>
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.<sup>34</sup> 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.<sup>35</sup>

---

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.

<sup>32</sup> 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.

<sup>33</sup> 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.

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

<sup>35</sup> For an example of this, see Gujarati, Damodar N., Basic Econometrics, 3<sup>rd</sup> 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.<sup>36</sup>
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.<sup>37</sup> 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

---

<sup>36</sup> 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.

<sup>37</sup> 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.<sup>38</sup>
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.<sup>39</sup> 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.<sup>40</sup> 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 undoubtedly 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

---

<sup>38</sup> 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>.

<sup>39</sup> 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](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).

<sup>40</sup> 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<sup>41</sup> 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

---

<sup>41</sup> 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 datel=date(viewsdate,"MDY")
format datel %d
gen month1=month(datel)
gen year1=year(datel)
*
* 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 datel 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 datel=date(dateorig,"MDY")
gen month1=month(datel)
gen year1=year(datel)
sort datel
gen monthindex=_n
* account for skipped 11 months
replace monthindex = monthindex + 10 if year>=2015
gen intervention=1
replace intervention=0 if datel<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 datel 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 datel

*
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=datel
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 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_* {
  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				Prob > F	=
0.0002	3.1498e+12	3	1.0499e+12	R-squared	=
0.4953	3.2091e+12	28	1.1461e+11	Adj R-squared	=
0.4413				Root MSE	=
3.4e+05	6.3590e+12	31	2.0513e+11		

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]

	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					

```
-----+-----
0.4672                               Adj R-squared =
      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
-----+-----
20.87                               F(3, 28)      =
0.0000                               Prob > F      =
  Model |  5.1404e+11         3   1.7135e+11  R-squared     =
0.6910                               Adj R-squared =
0.6579                               Root MSE     =
  Residual |  2.2989e+11        28   8.2102e+09
-----+-----
Total |  7.4392e+11        31   2.3998e+10
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
-----+-----
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
-----+-----
0.0084
      Total |  8.5020e+11          31   2.7426e+10   Adj R-squared     =
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
-----+-----
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
-----+-----
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
-----+-----
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
-----+-----
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
-----+-----
20.64
      Model |  10062791.9           3  3354263.97  Prob > F          =
0.0000
      Residual |  4549258.31          28  162473.511  R-squared         =
0.6887
-----+-----
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
-----+-----
16.30
      Model |           4   F(4, 27)          =

```

```

      Model | 5.2604e+11      4 1.3151e+11  Prob > F      =
0.0000
      Residual | 2.1789e+11     27 8.0700e+09  R-squared     =
0.7071
-----+-----
0.6637                               Adj R-squared  =
      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
-----+-----
18.49                               F(4, 27)       =
      Model | 3.5157e+12      4 8.7893e+11  Prob > F      =
0.0000
      Residual | 1.2835e+12     27 4.7538e+10  R-squared     =
0.7326
-----+-----
0.6929                               Adj R-squared  =
      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
-----+-----+-----+-----+-----+-----
10.06
      Model |   6663270.2            3   2221090.07      Prob > F           =
0.0001
      Residual |   6180561.8           28   220734.35      R-squared           =
0.5188
-----+-----+-----+-----+-----+-----
0.4672
      Total |   12843832           31   414317.161      Adj R-squared       =
469.82
      Root MSE
-----+-----+-----+-----+-----+-----

```

```

-----+-----
-
  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

```

interv~y	monthi~x	postsl~e	postsl~l	postsl~y	interv~n	interv~l
0   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



9.	9	0	0	0	0	0
0						
10.	10	0	0	0	0	0
0						
-----						
11.	11	0	0	0	0	0
0						
12.	12	0	0	0	0	0
0						
13.	13	0	0	0	0	0
0						
14.	14	0	0	0	0	0
0						
15.	15	0	0	0	0	0
0						
-----						
16.	16	0	1	0	0	1
0						
17.	17	0	2	1	0	1
1						
18.	18	1	3	2	1	1
1						
19.	19	2	4	3	1	1
1						
20.	20	3	5	4	1	1
1						
-----						
21.	21	4	6	5	1	1
1						
22.	22	5	7	6	1	1
1						
23.	23	6	8	7	1	1
1						
24.	24	7	9	8	1	1
1						
25.	25	8	10	9	1	1
1						
-----						
26.	26	9	11	10	1	1
1						
27.	27	10	12	11	1	1
1						
28.	28	11	13	12	1	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	=
-----+-----					
Model	9.6611e+11	2	4.8306e+11	F(2, 29)	=
Residual	5.3928e+12	29	1.8596e+11	Prob > F	=
-----+-----					
Total	6.3590e+12	31	2.0513e+11	R-squared	=
-----+-----					
				Adj R-squared	=
				Root MSE	=

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

```
. 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	=	
-----+-----					F(2, 29)	=
Model	2.9173e+11	2	1.4586e+11	Prob > F	=	
Residual	4.5220e+11	29	1.5593e+10	R-squared	=	
-----+-----					Adj R-squared	=
Total	7.4392e+11	31	2.3998e+10	Root MSE	=	

```

-----
-
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	=	
-----+-----					F(3, 28)	=
Model	5.1404e+11	3	1.7135e+11	Prob > F	=	
Residual	2.2989e+11	28	8.2102e+09	R-squared	=	
-----+-----					Adj R-squared	=
Total	7.4392e+11	31	2.3998e+10	Root MSE	=	

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	=	
-----+-----					F(3, 28)	=
Model	4.5452e+11	3	1.5151e+11	Prob > F	=	
Residual	2.8941e+11	28	1.0336e+10	R-squared	=	

```
-----+-----
0.5693                               Adj R-squared   =
      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
      postslope   | -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
-----+-----
12.16                               F(3, 28)         =
      Model |  4.2092e+11          3  1.4031e+11  Prob > F         =
0.0000
      Residual |  3.2300e+11         28  1.1536e+10  R-squared        =
0.5658
-----+-----
0.5193                               Adj R-squared    =
      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
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
-----+-----+-----+-----+-----+-----
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
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
-----

```

```
-----+-----
0.6378                               Adj R-squared =
      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
  postslopeamay | -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
-----+-----
14.09                               F(3, 28)      =
      Model |  2.8871e+12          3   9.6236e+11  Prob > F      =
0.0000
      Residual |  1.9122e+12         28   6.8291e+10  R-squared     =
0.6016
-----+-----
0.5589                               Adj R-squared =
      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
-----+-----
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
-----+-----
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
-----+-----
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

```

```
-----+-----
0.6378                               Adj R-squared =
      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
      postslope    | -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
-----+-----
14.09                               F(3, 28)      =
      Model |  2.8871e+12          3  9.6236e+11  Prob > F      =
0.0000
      Residual |  1.9122e+12         28  6.8291e+10  R-squared     =
0.6016
-----+-----
0.5589                               Adj R-squared =
      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 datel 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				Prob > F	=
0.4470	7.5795e+10	3	2.5265e+10	R-squared	=
0.0891	7.7441e+11	28	2.7657e+10	Adj R-squared	= -
0.0084				Root MSE	=
1.7e+05	8.5020e+11	31	2.7426e+10		

```

-----
-
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
-----+-----
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
-----+-----
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   =
-----+-----
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
-----+-----
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	=	
-----+-----					F(3, 28)	=
Model	2.9756e+12	3	9.9188e+11	Prob > F	=	
Residual	3.1438e+12	28	1.1228e+11	R-squared	=	
-----+-----					Adj R-squared	=
Total	6.1195e+12	31	1.9740e+11	Root MSE	=	

	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 datel 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					

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
totview31					
-----+-----					
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 datel
```

```
.
. *
. 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~1)	med(rankv~1)
N(rankvie~1)				
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   -1.5910936
6.656134
      4 |   -0.5020325  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
-----+-----
7.50
      Model |   40176.52           11   3652.41091   Prob > F         =
0.0000
      Residual |  741743.313        1,524   486.708211   R-squared         =
0.0514
-----+-----
0.0445
      Total |  781919.833        1,535   509.394028   Adj R-squared     =
22.061
      Root MSE

```



```

-----
-
rankview$all |      Coef.   Std. Err.      t    P>|t|    [95% Conf.
Interval]
-----+-----
-
      month1 |
      2 |   -0.8854167   2.599969   -0.34   0.733   -5.985312
4.214478
      3 |    3.173611   2.599969    1.22   0.222   -1.926284
8.273506
      4 |   -0.4930556   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 |   -0.984375   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 rankview$early==maxrank$early

```

```

      |      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		Total		
	0	1			
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=datel

. 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$artnames!="Hamis",], .(monthindex, intervention, postslope), summ
arise, mean1=mean(rankviewsall),
median1=median(rankviewsall), meanviews=mean(pageviewall), medviews=median(page
viewall))
sum2018_31noissue<-ddply(artincl2018noNA[artincl2018noNA$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/", "terror48plum
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
Hamam","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,lwd=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(sum2014_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$pageviewall/1000),label=paste0(pretty(tmpplot$pageviewall/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(compop26)
for (i in 1:26) {
  tmpploty<-compop26[,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.xlsx* – 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.

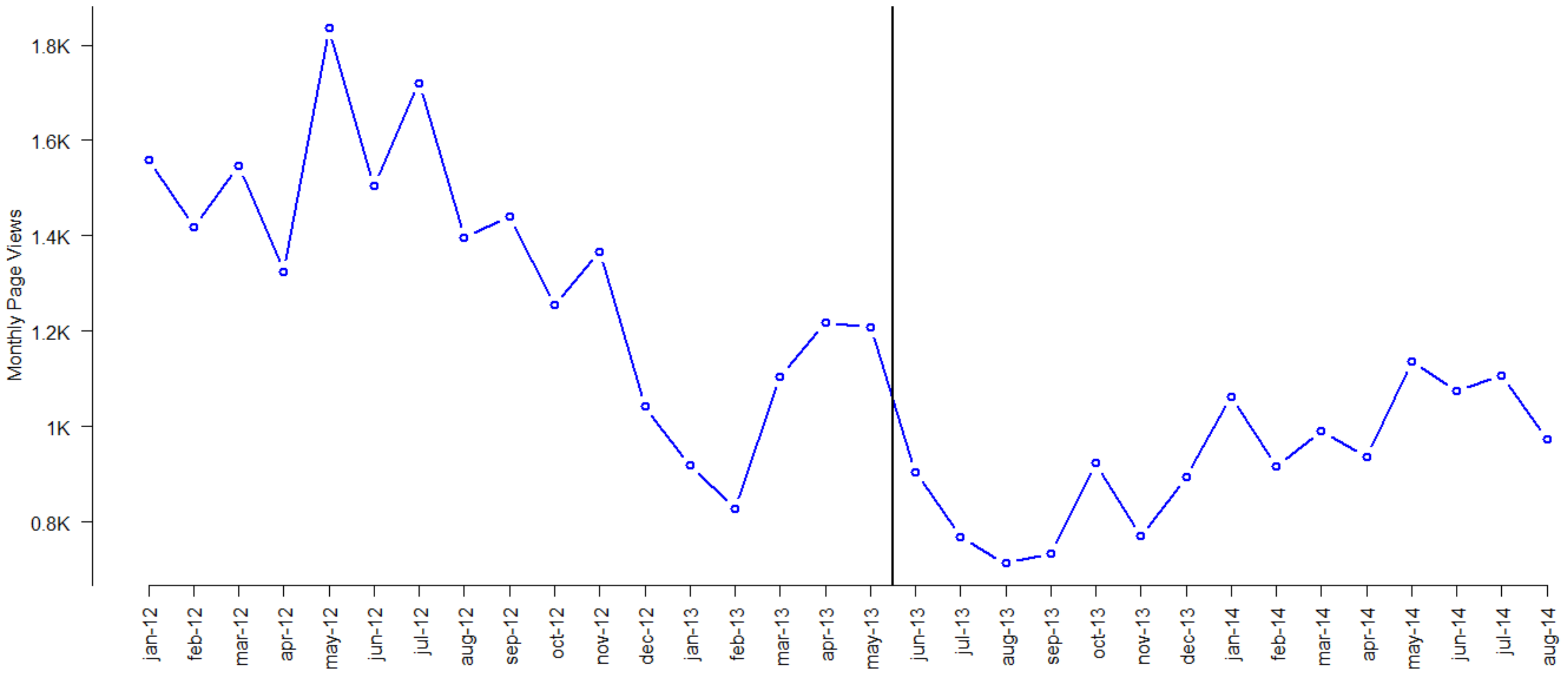
Enjoy ultimate Frisbee, basketball, biking, hiking, running, tennis, chess, and bridge.

#### FOUR YEAR TESTIMONY HISTORY

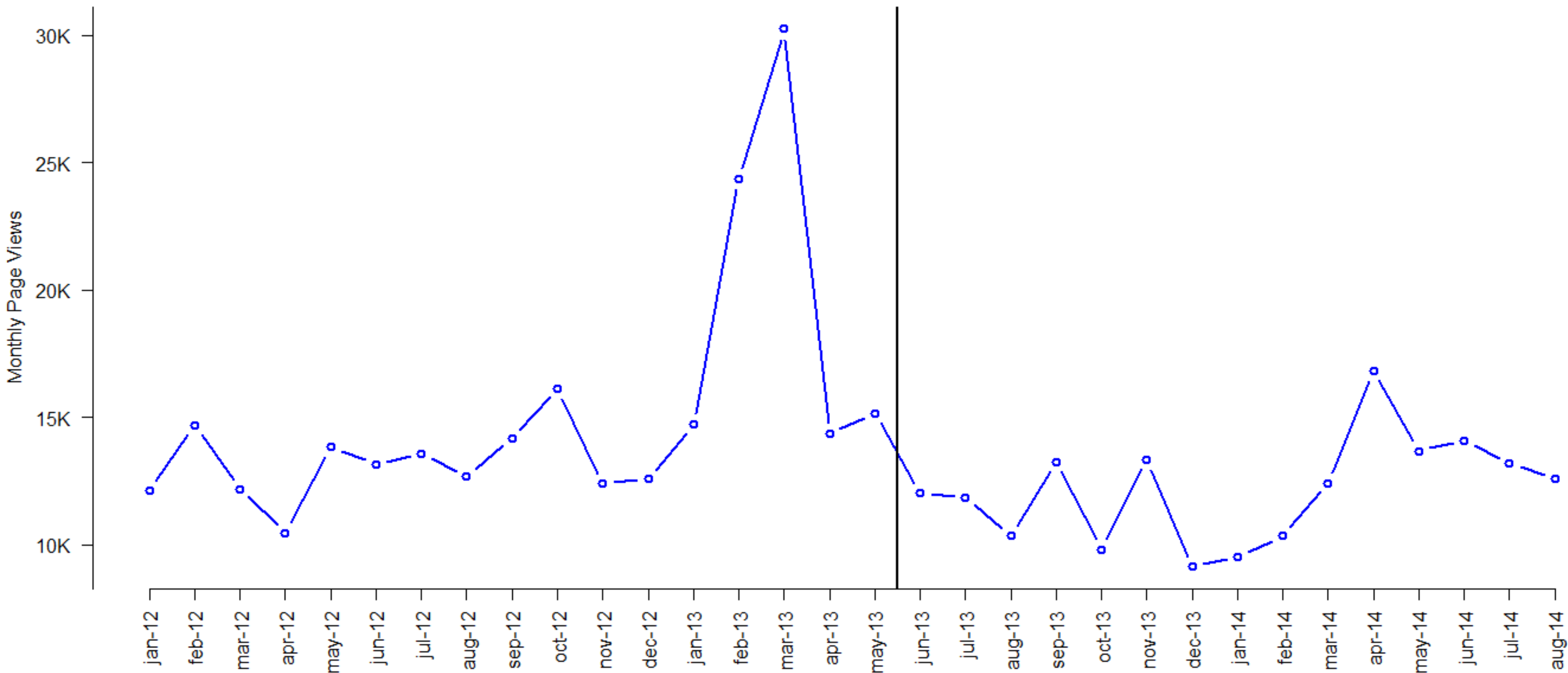
1. [Federal court] Bayer Healthcare LLC, v. Baxalta, et al, 2019.
2. [Federal court] Steward, et al, v. State of Texas, 2018.
3. [deposition] Center for Independence of the Disabled, et al, v. Metropolitan Transit Authority, et al, 2018.
4. [deposition] Bayer Healthcare, LLC, v. Baxalta Inc., et al, 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**

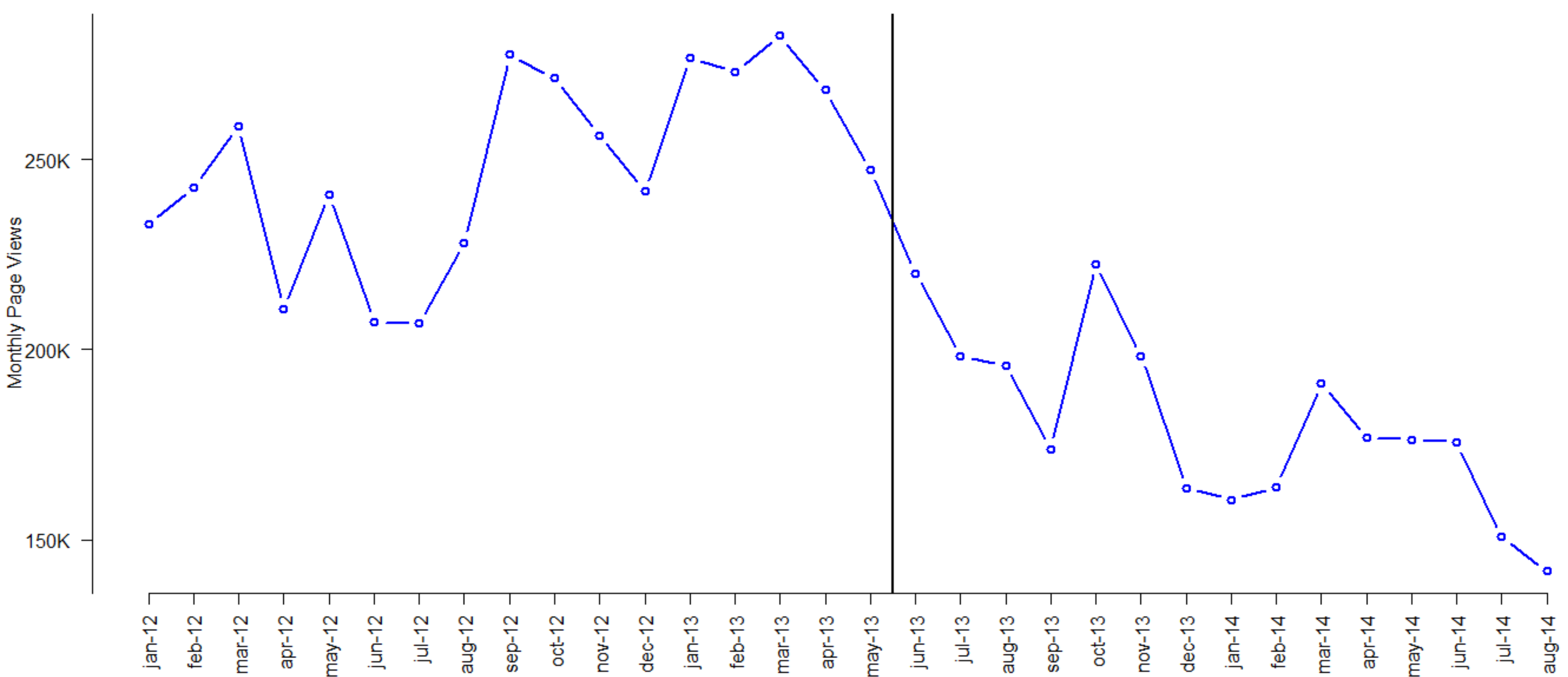
# Page Views for \_Euskadi\_ta\_Askatasuna



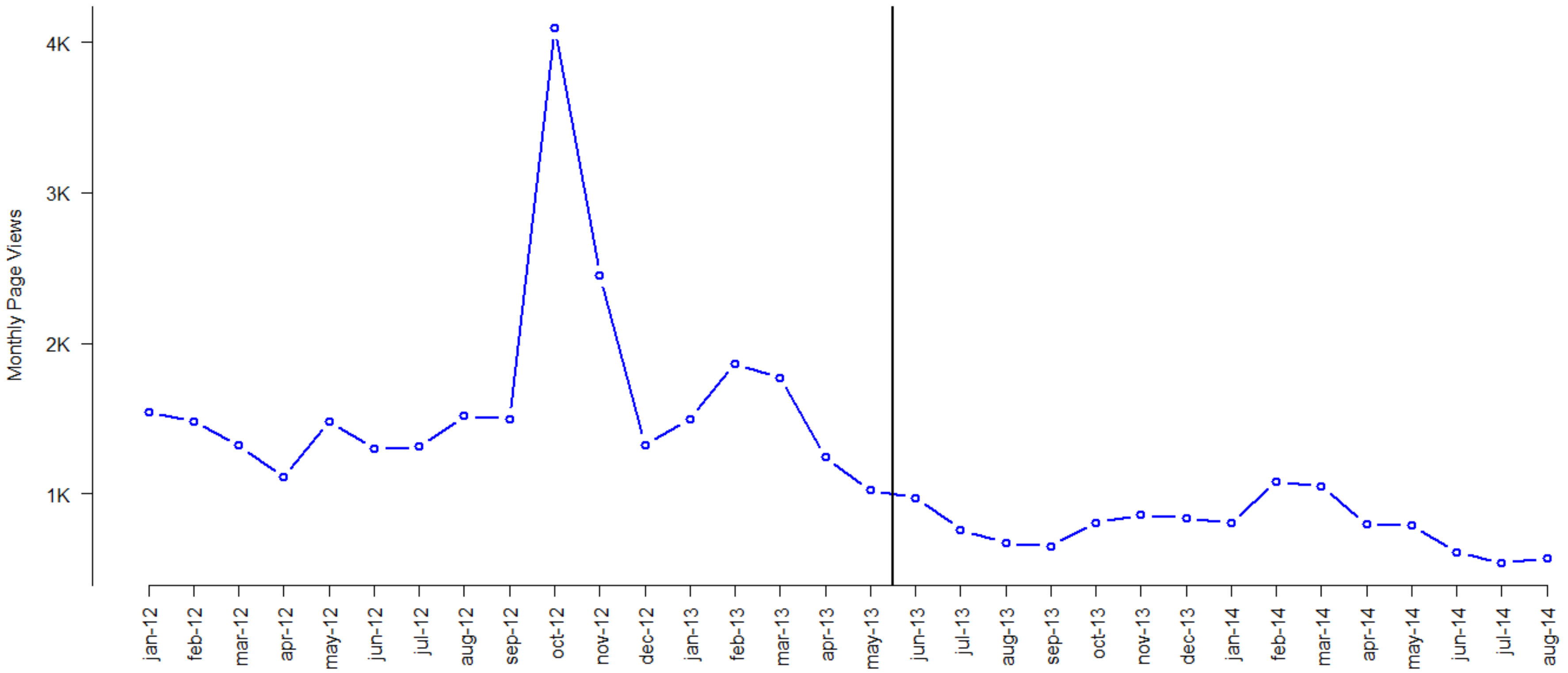
### Page Views for Abu\_Sayyaf



### Page Views for Afghanistan

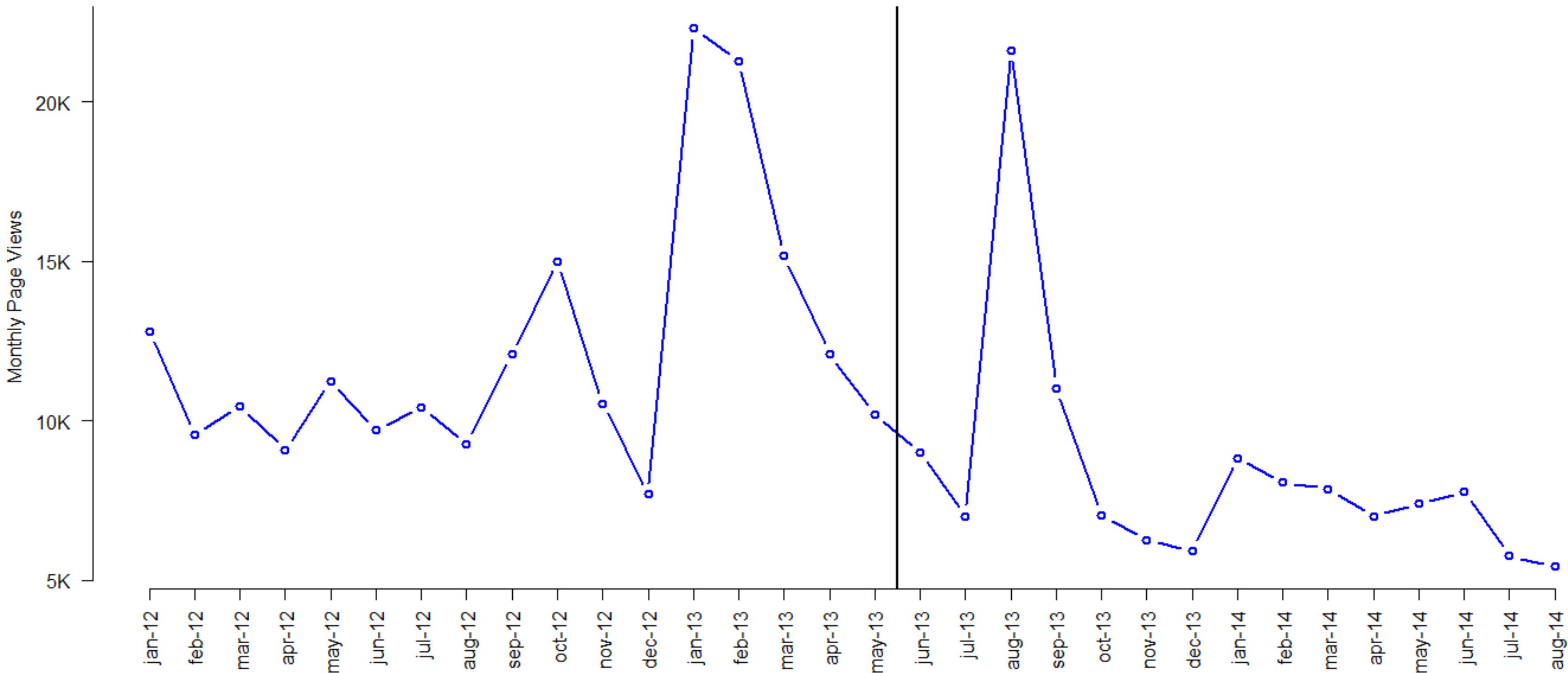


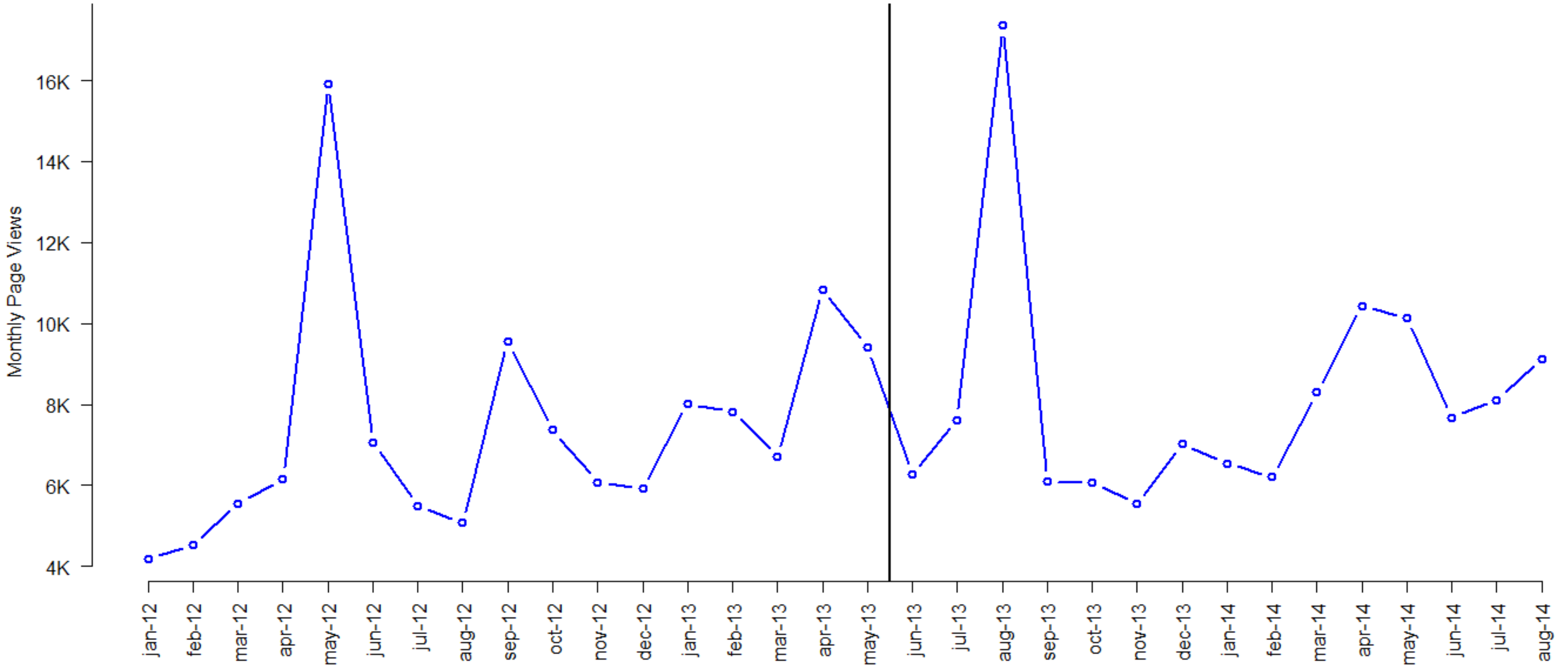
### Page Views for agro

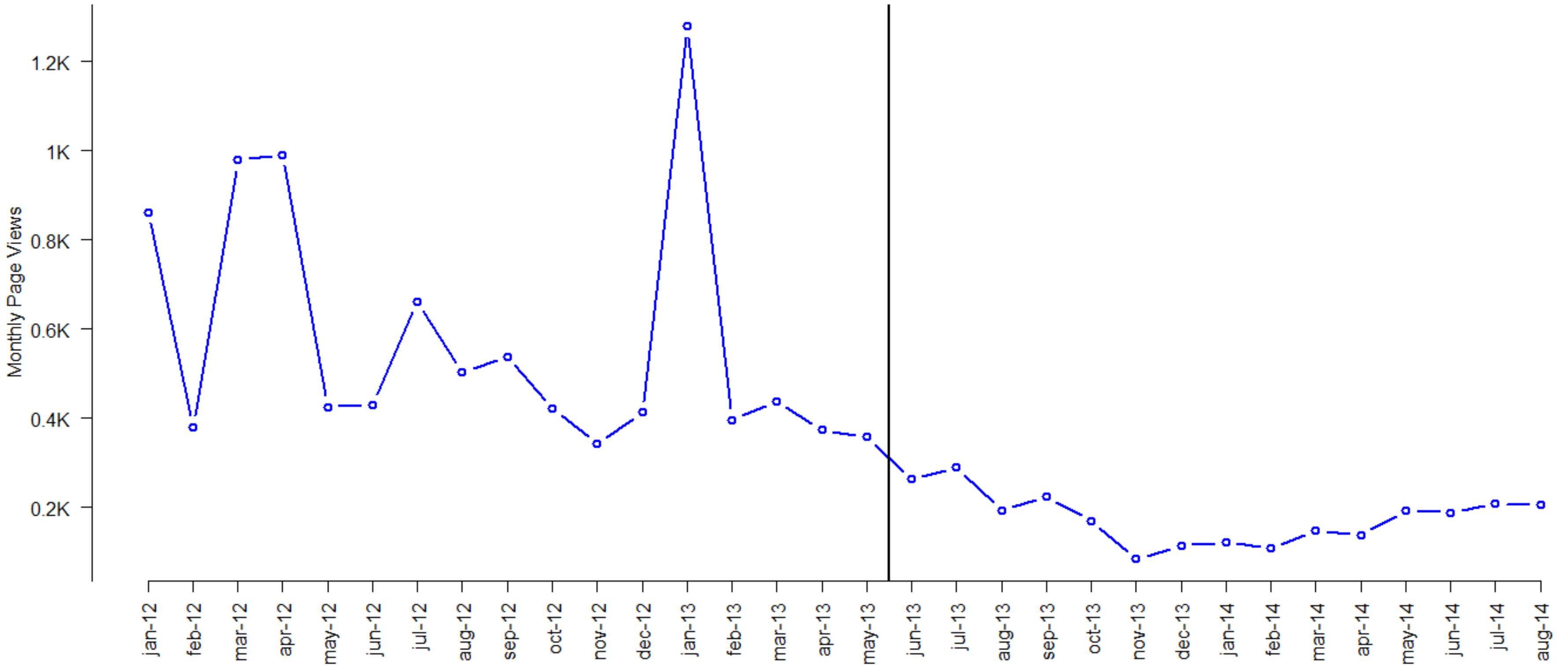




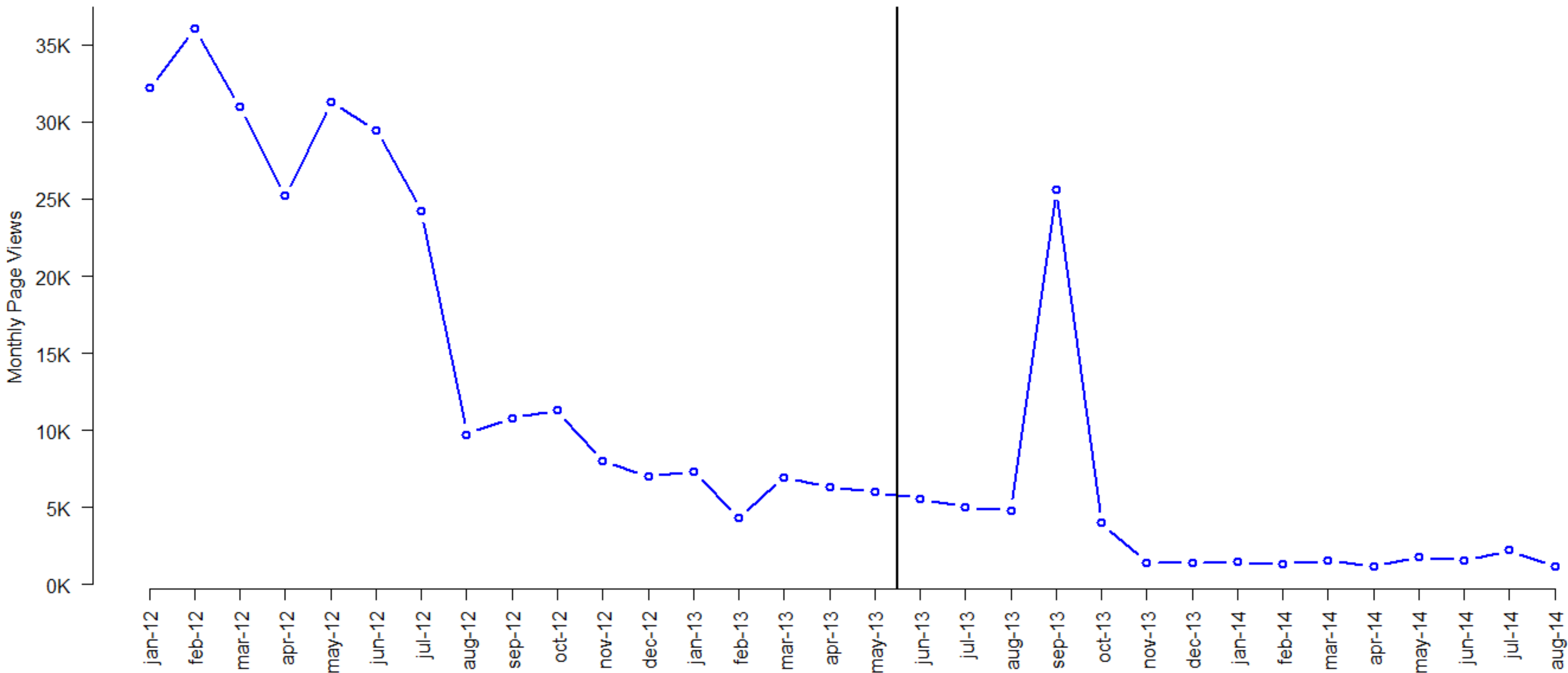
### Page Views for Al\_Qaeda



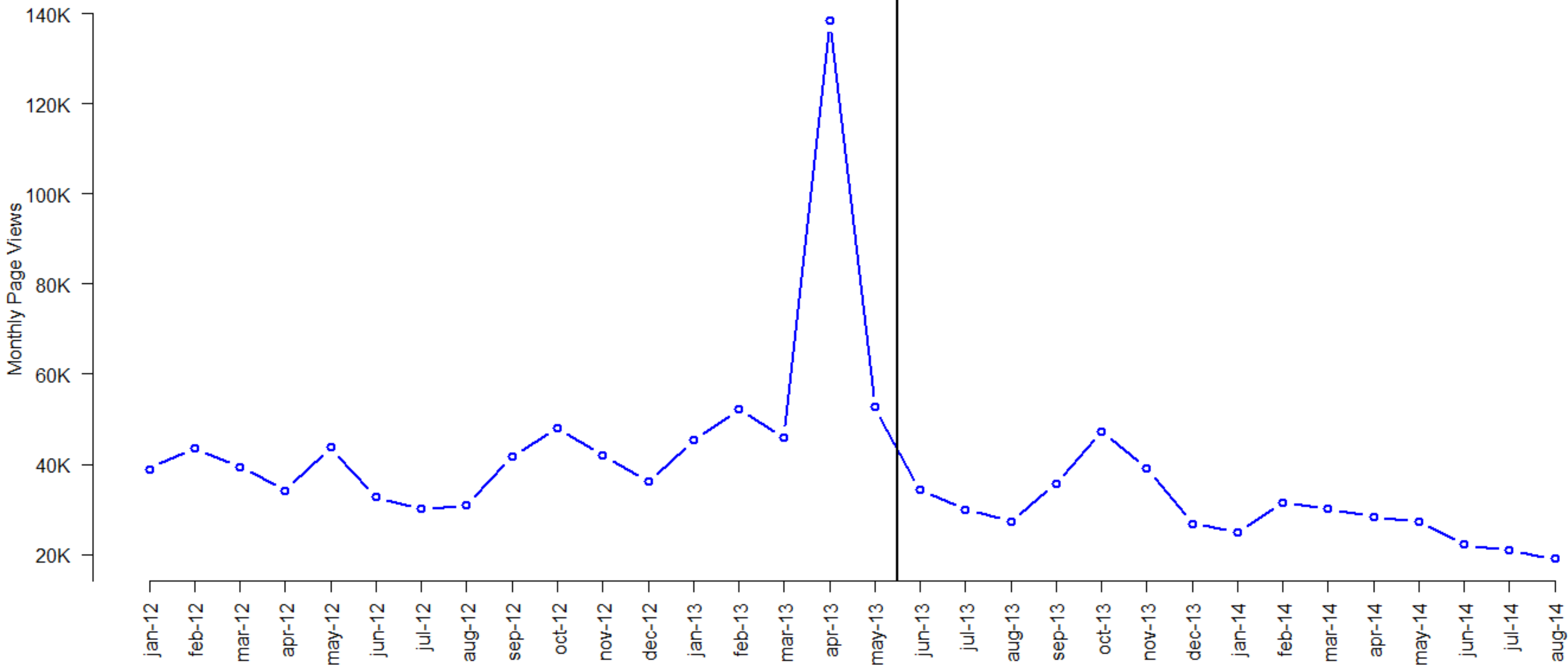




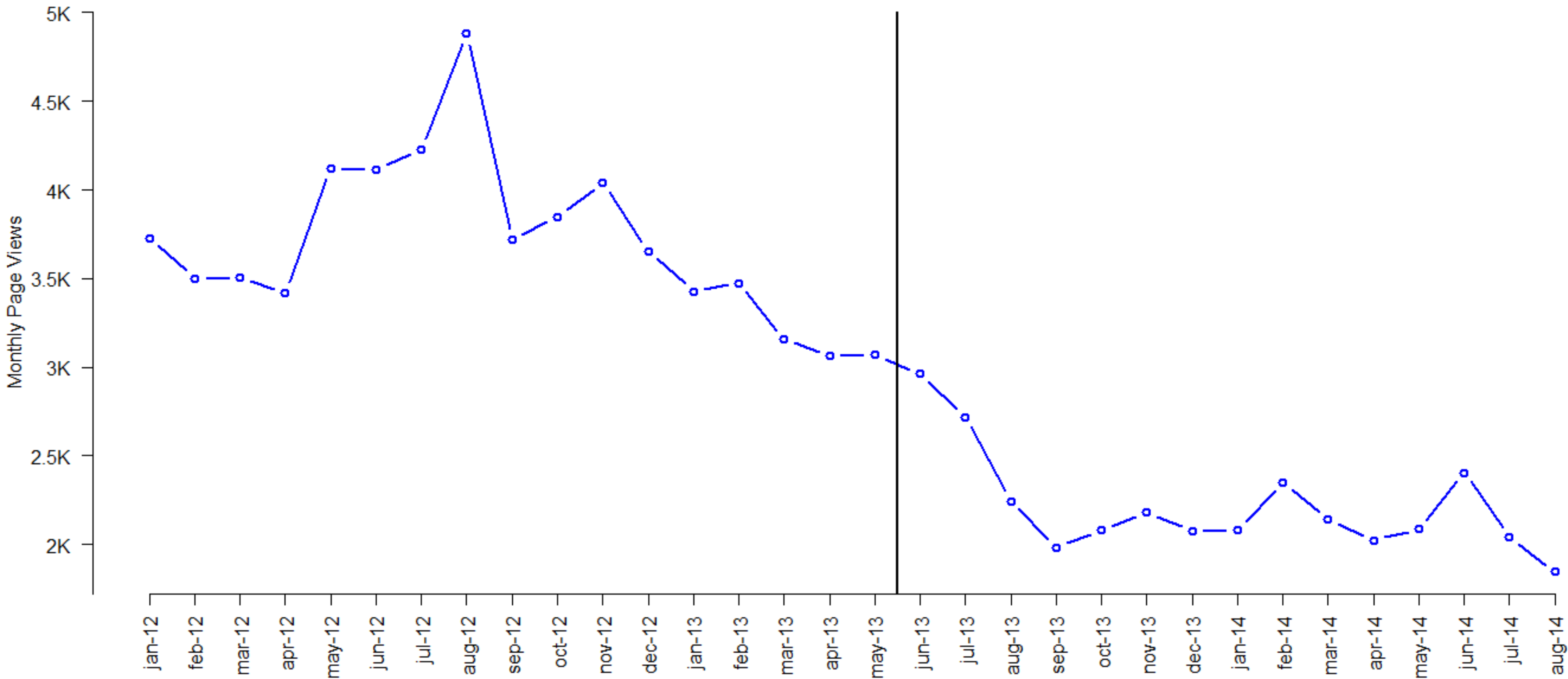
### Page Views for Al\_Shabaab



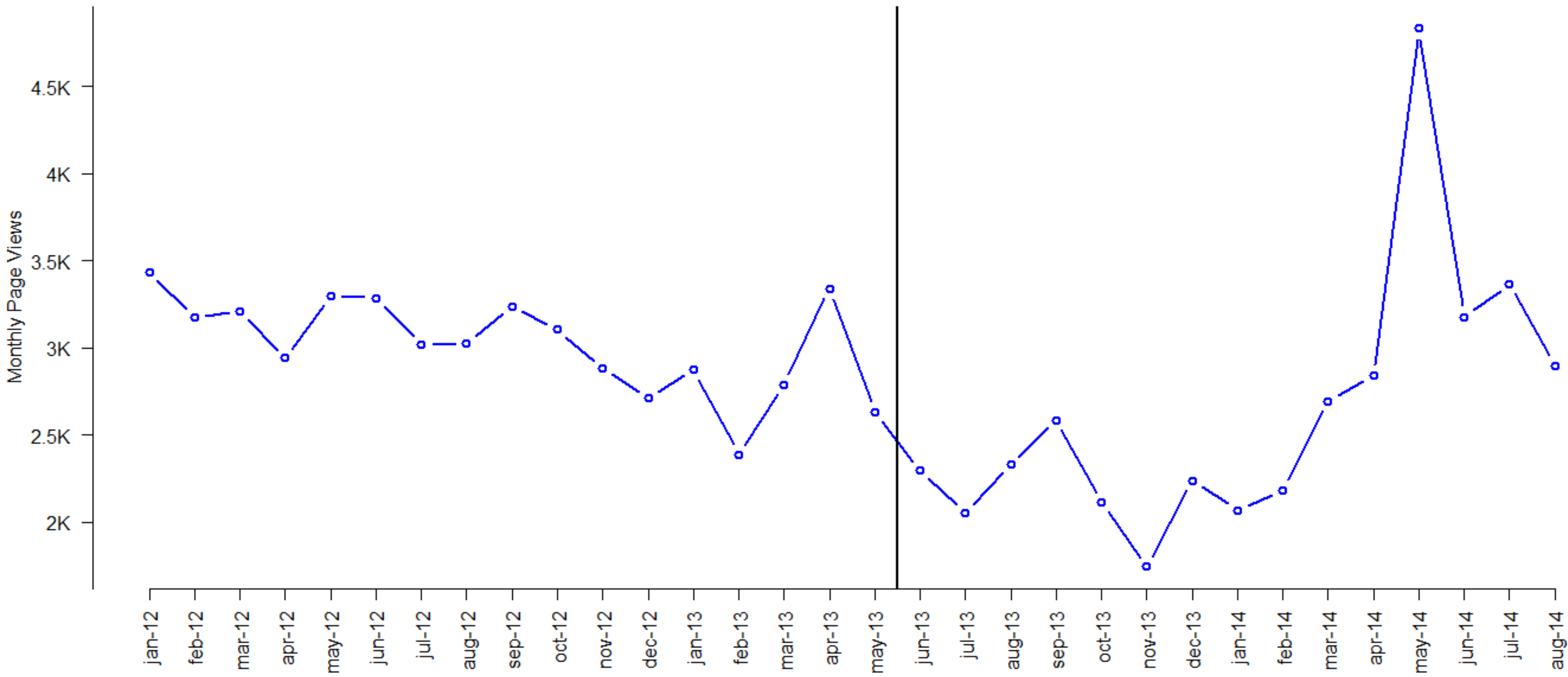
# Page Views for Ammonium\_nitrate



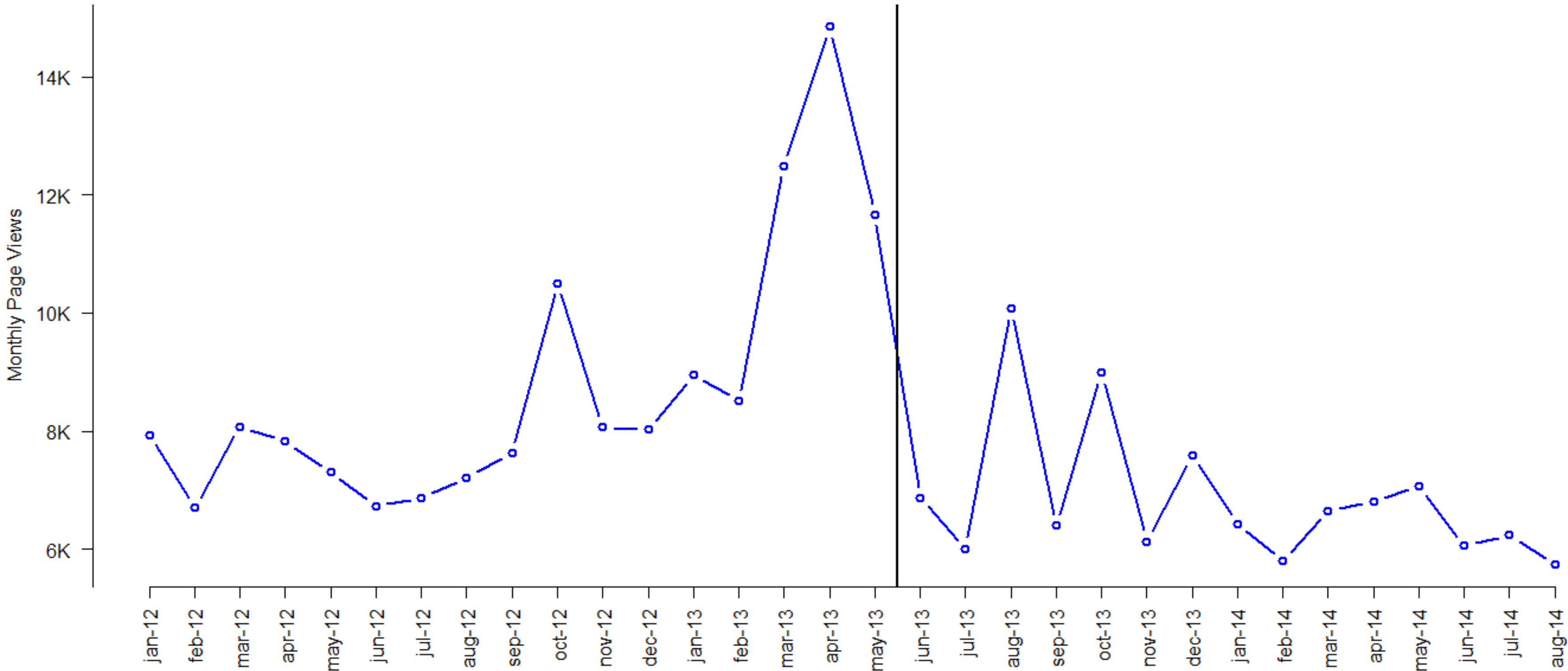
### Page Views for attack



### Page Views for Biological\_weapon

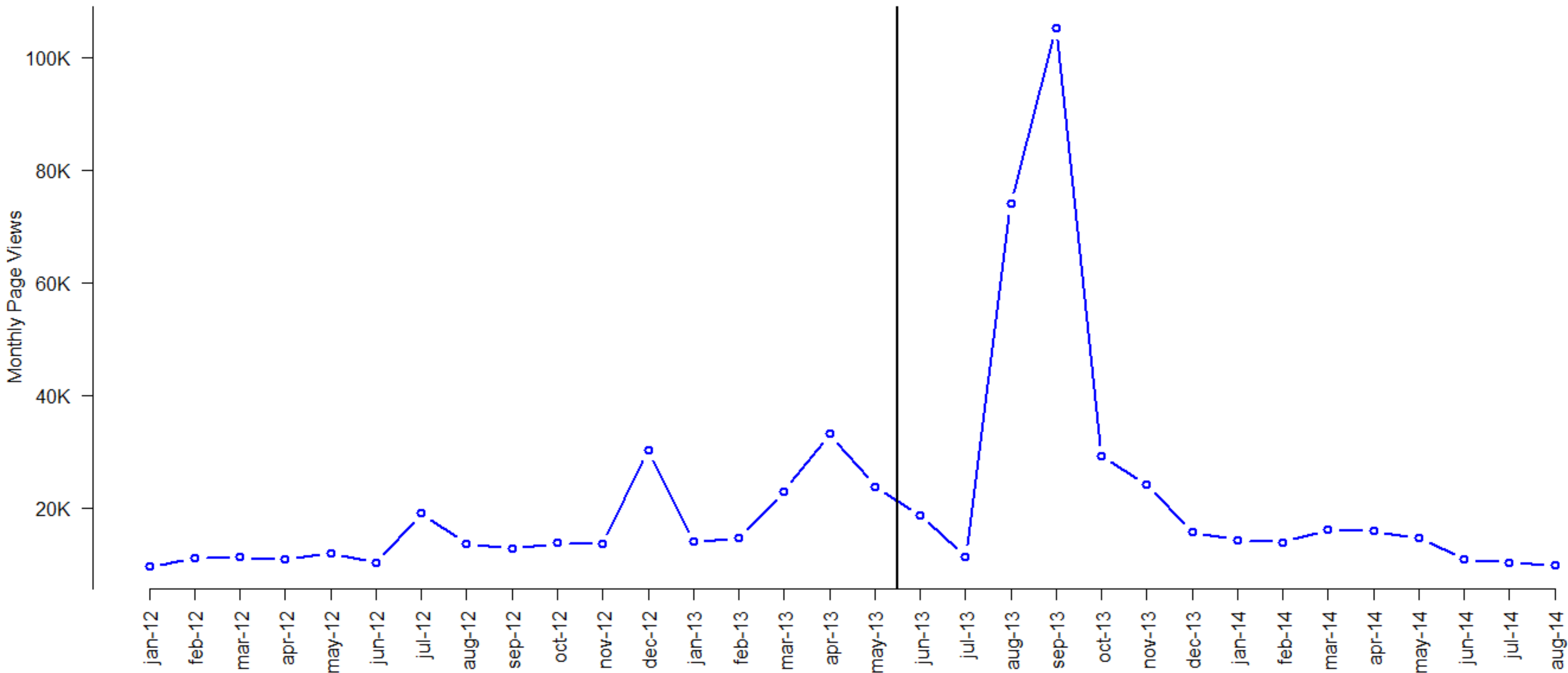


### Page Views for Car\_bomb

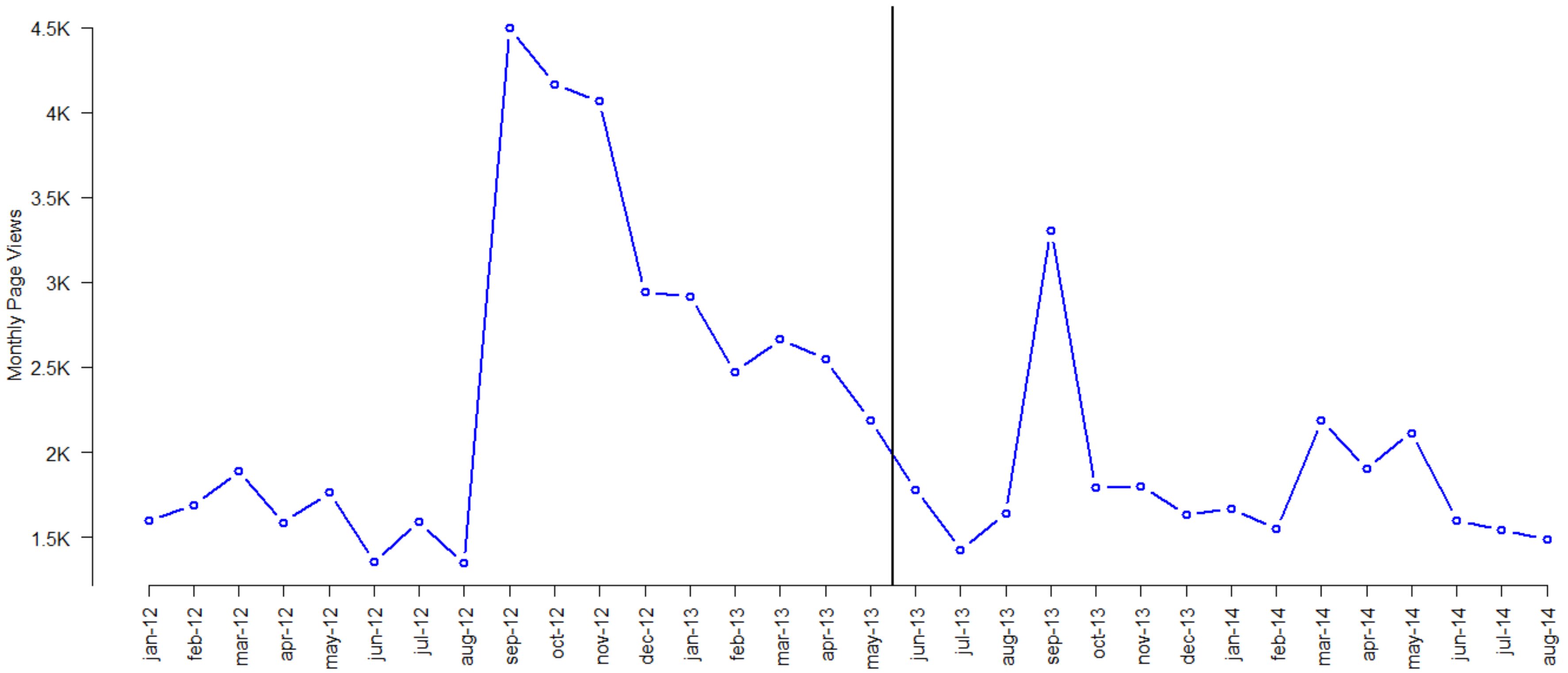




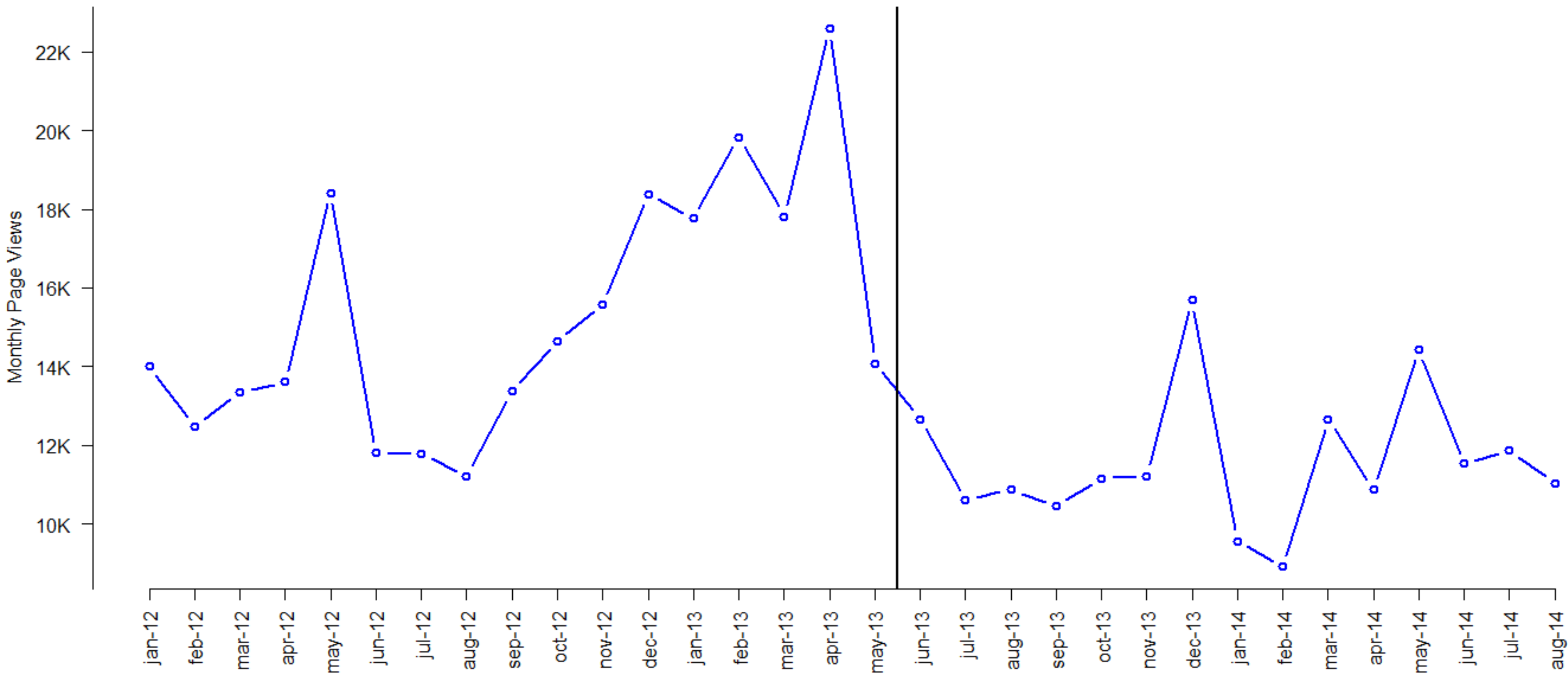
### Page Views for Chemical\_weapon



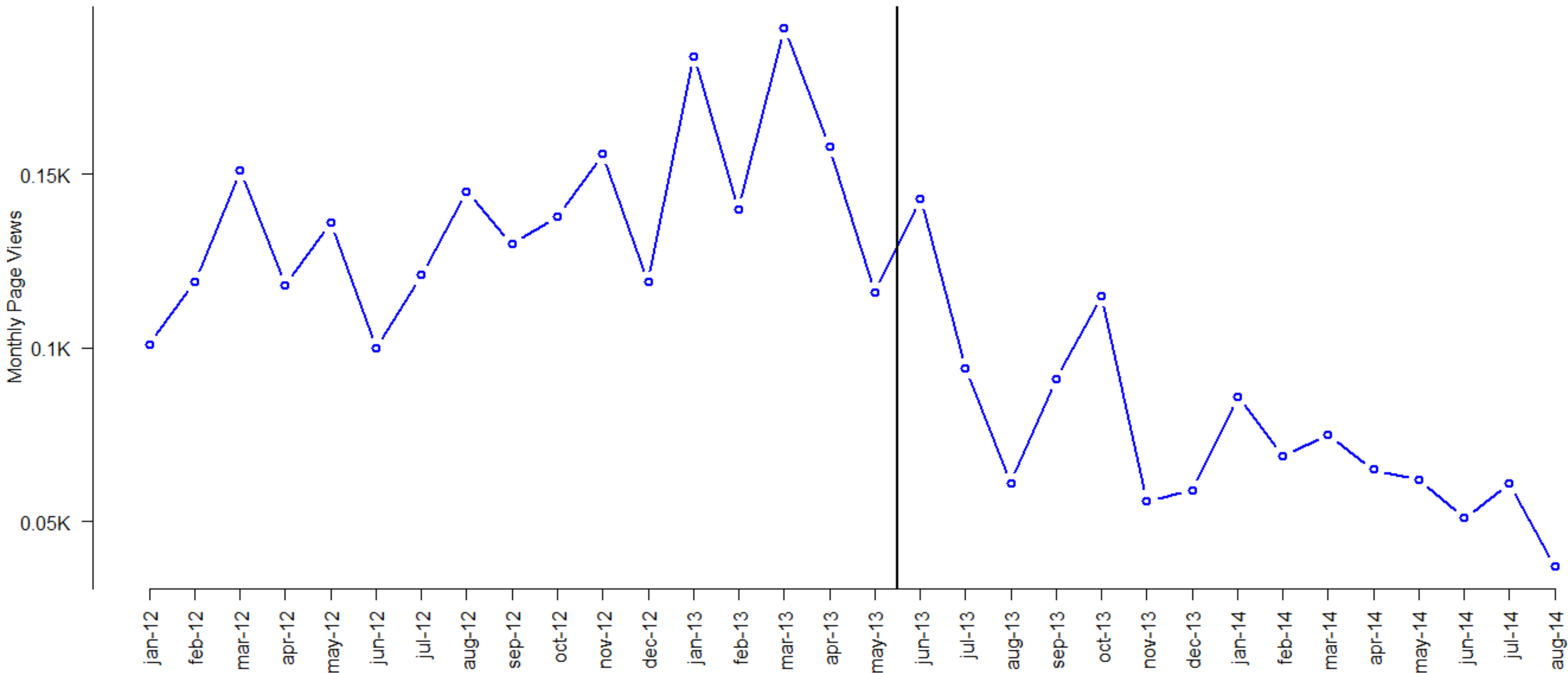
### Page Views for Conventional\_weapon



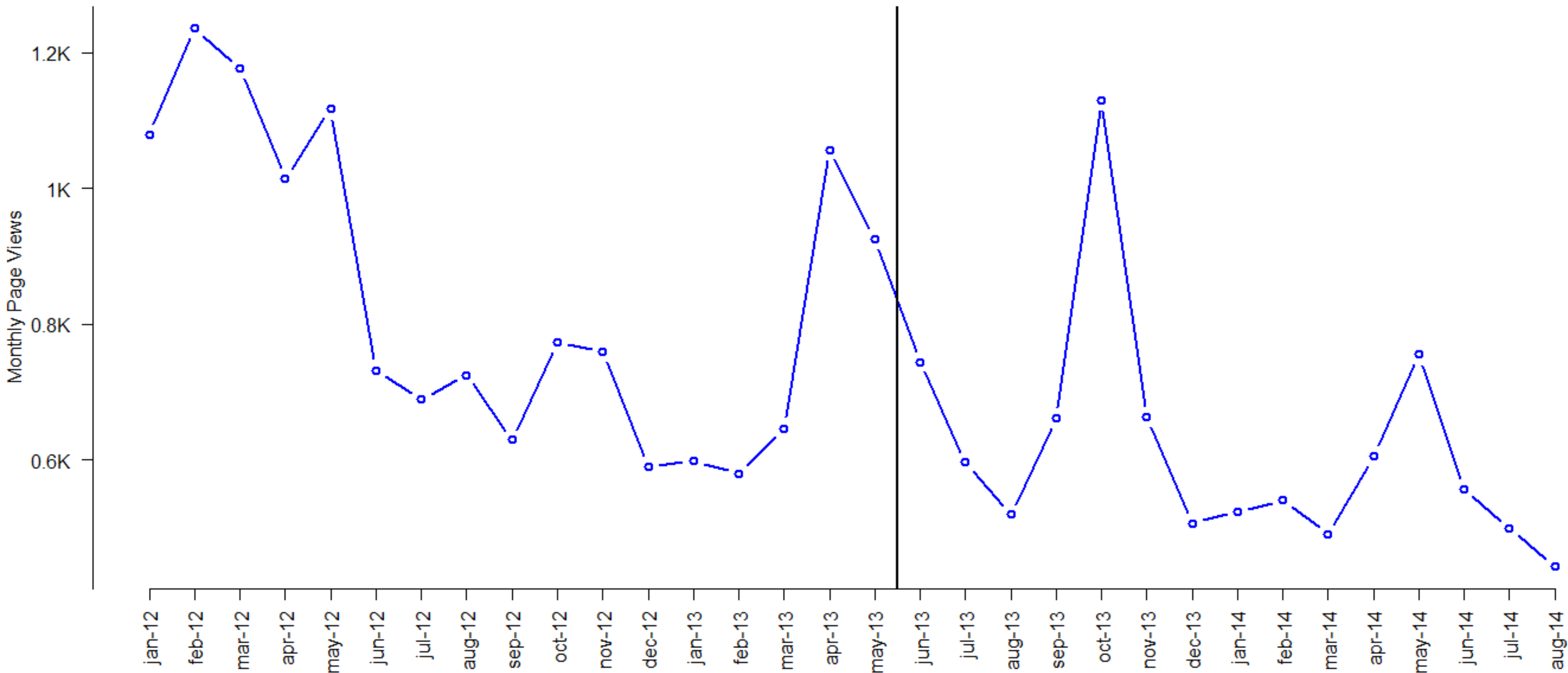
### Page Views for dirty\_bomb



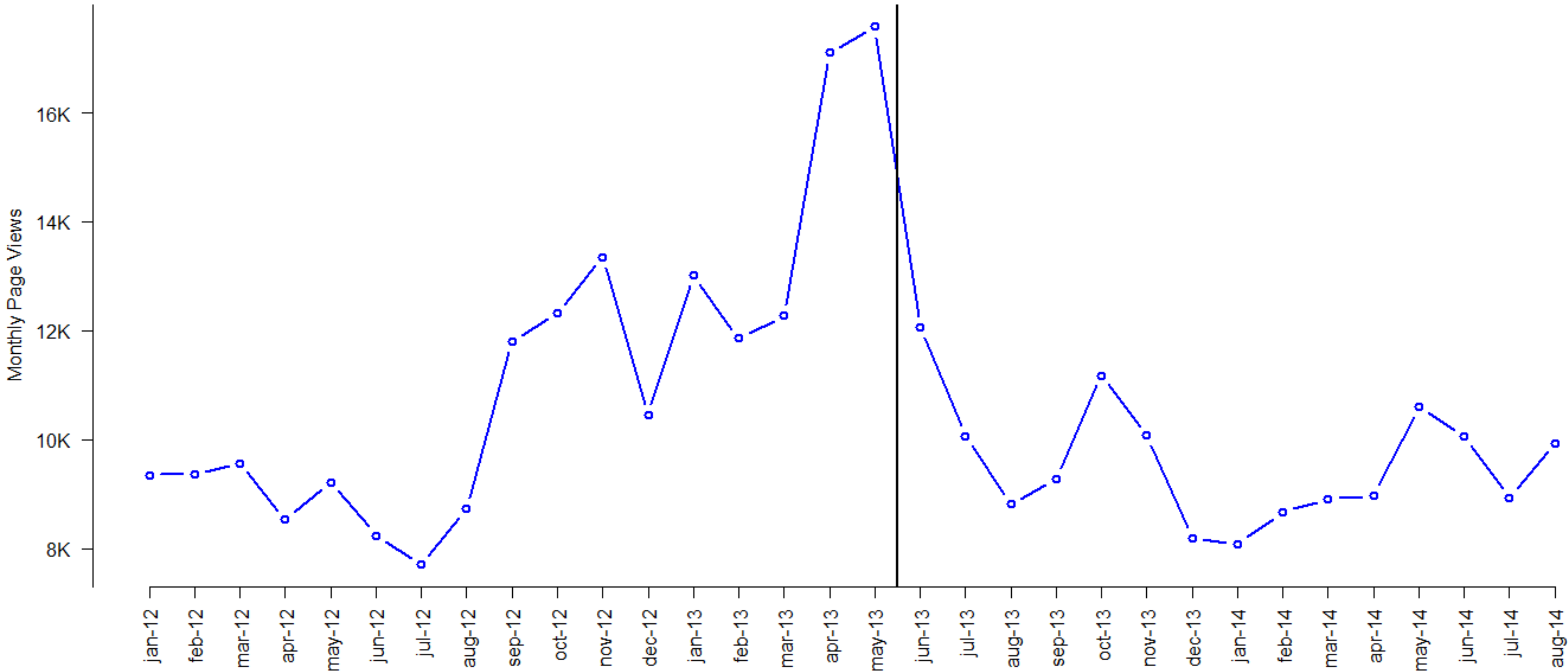
### Page Views for Eco\_terrorism



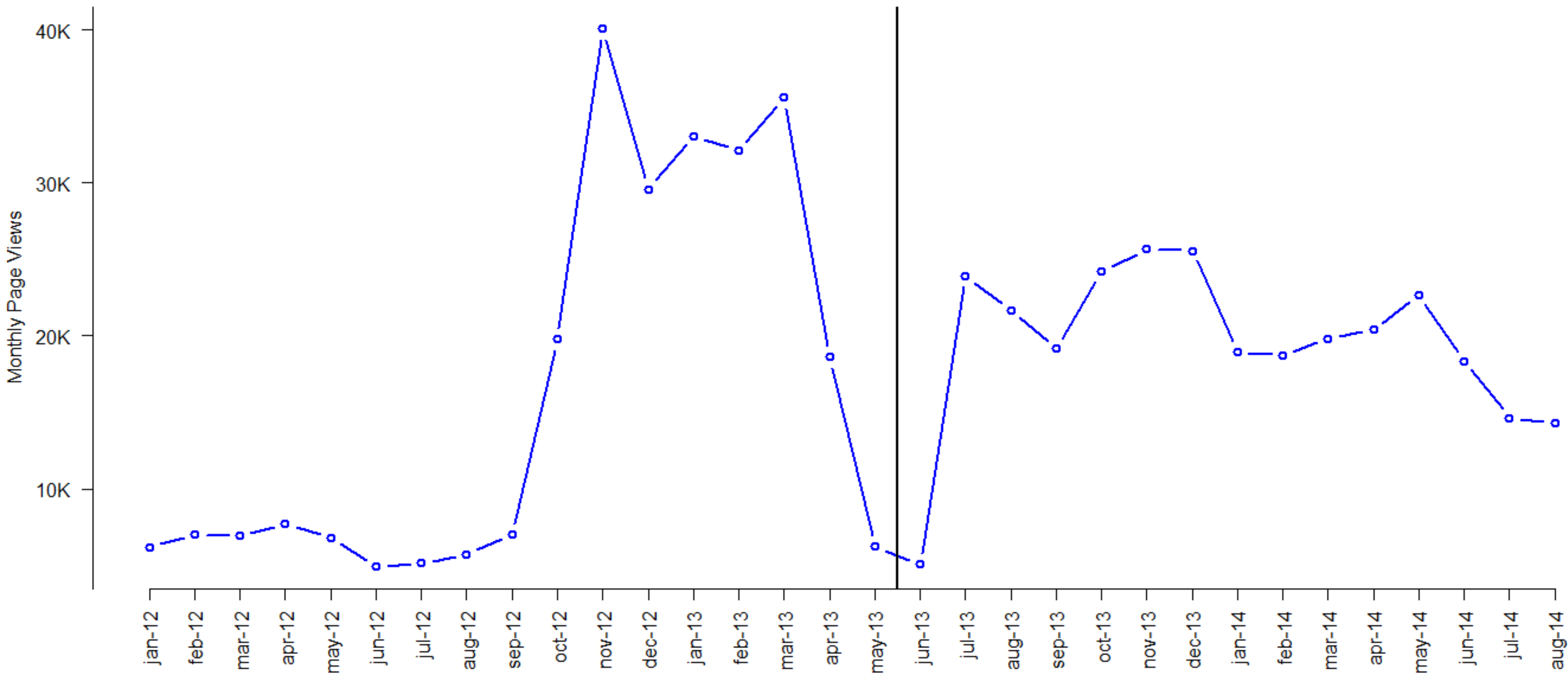
### Page Views for Environmental\_terrorist\_NA\_ism



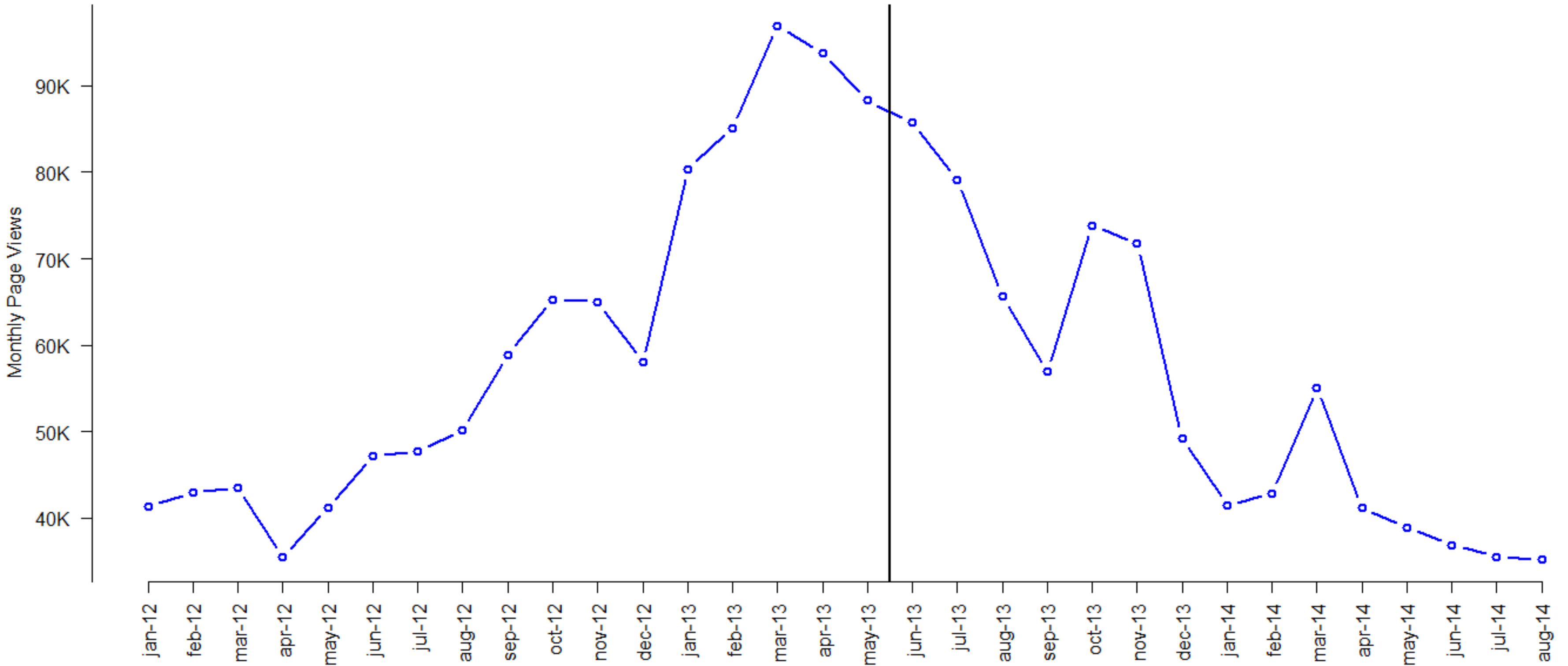
### Page Views for Extremism



### Page Views for FARC

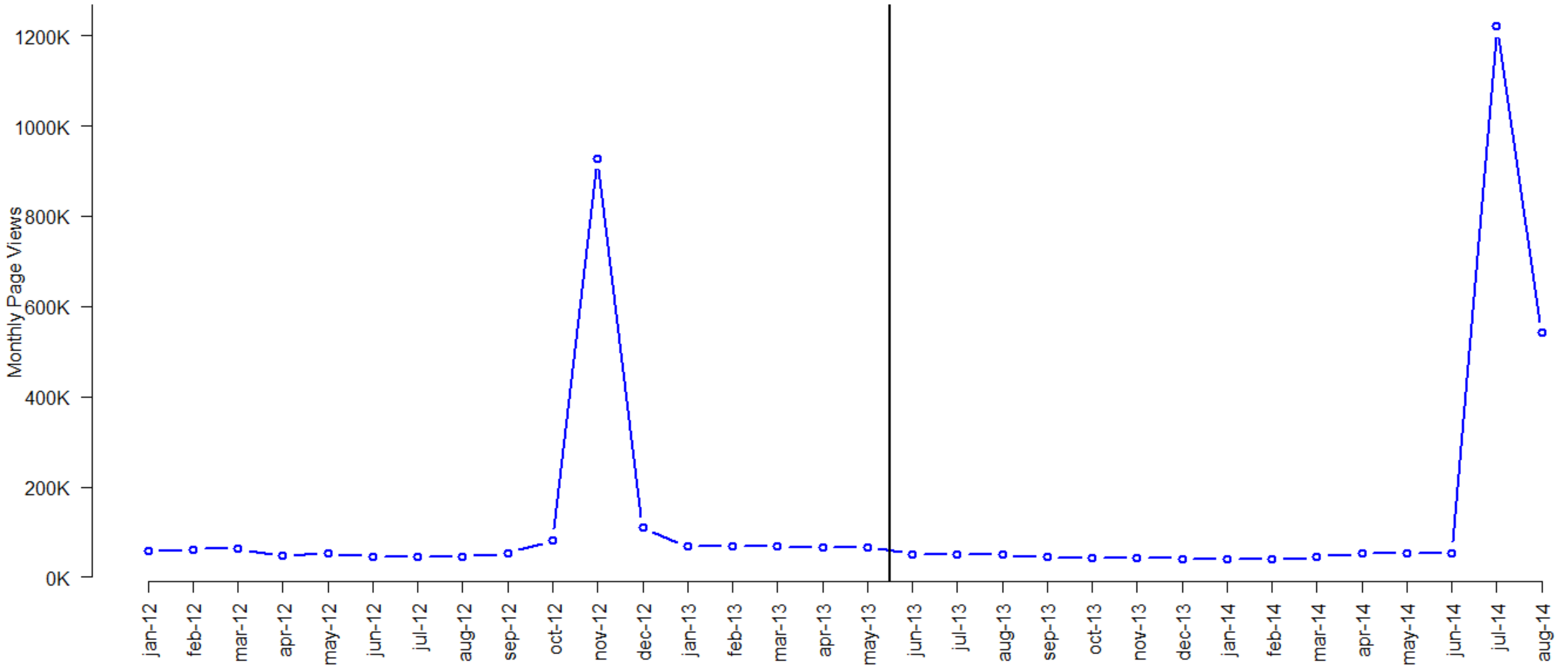


### Page Views for Fundamentalism

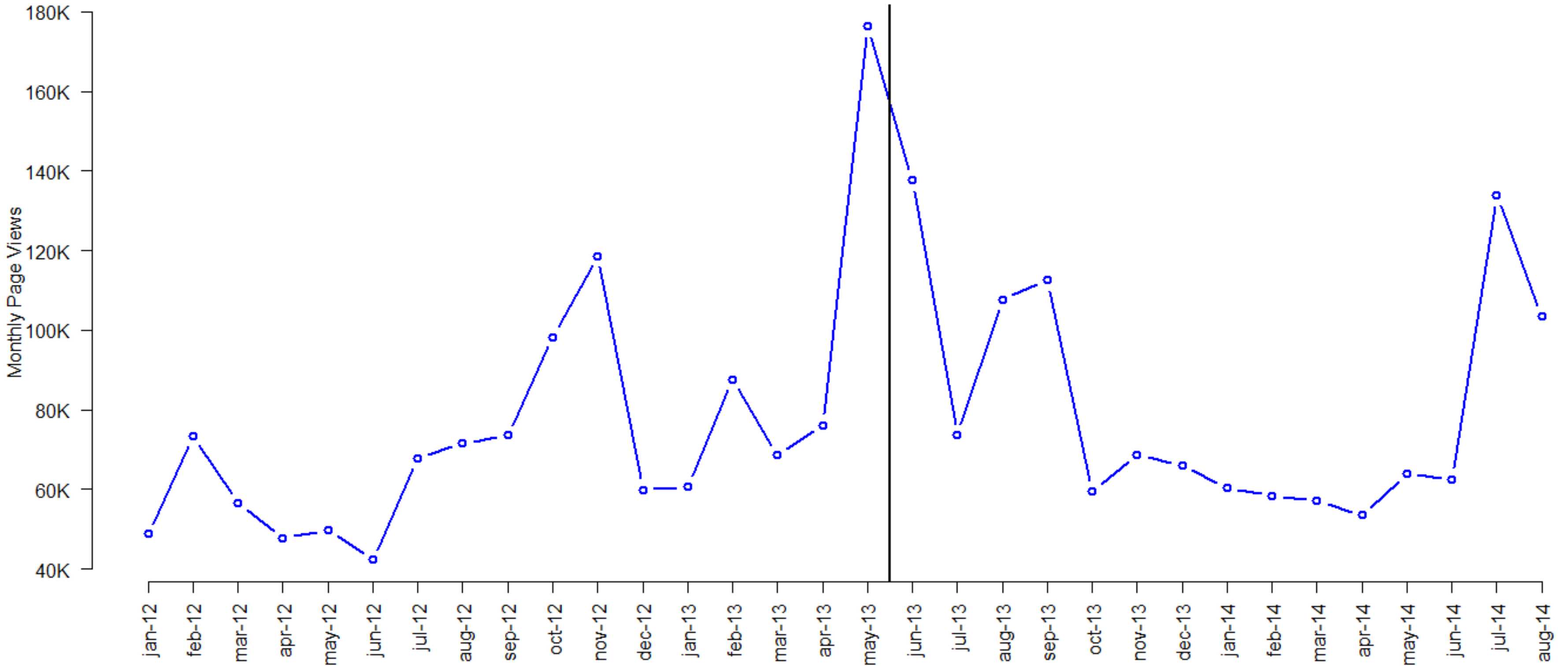




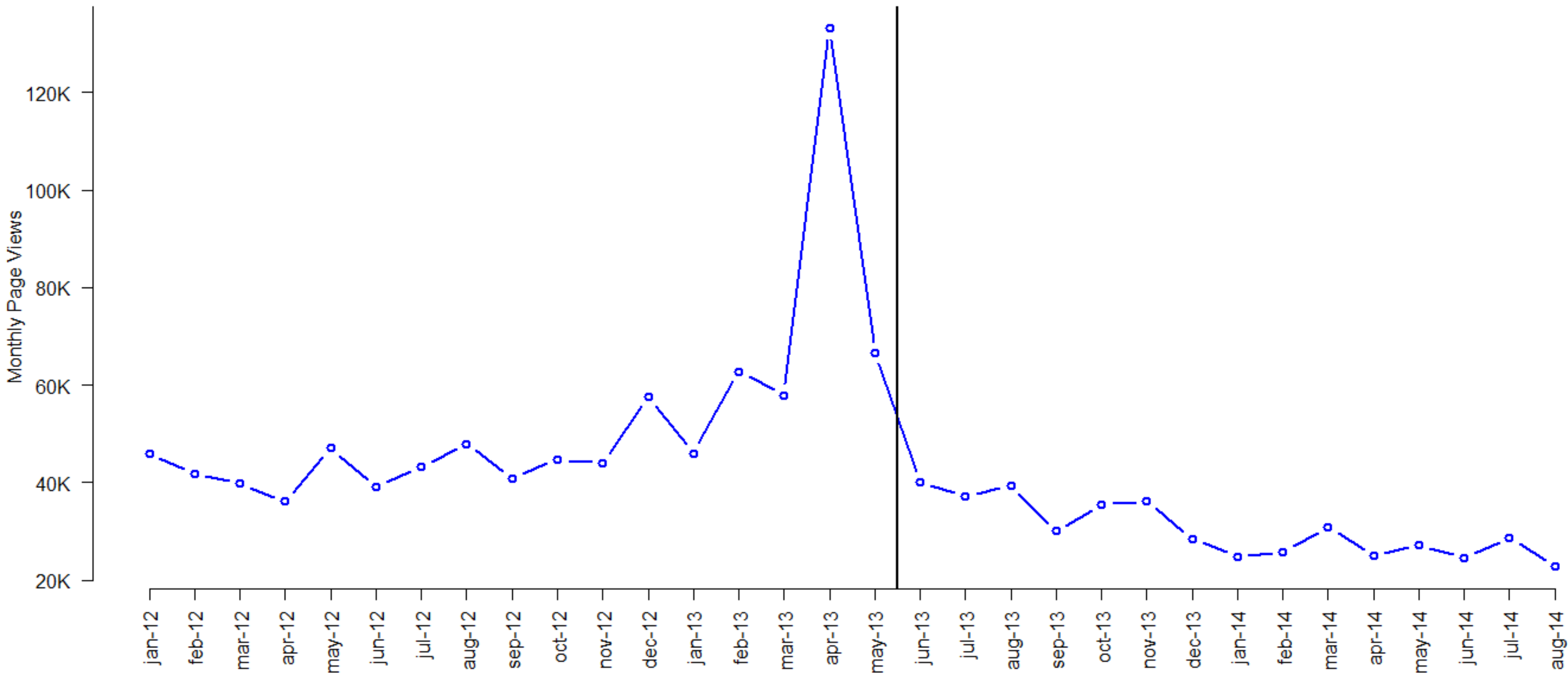
# Page Views for Hamas



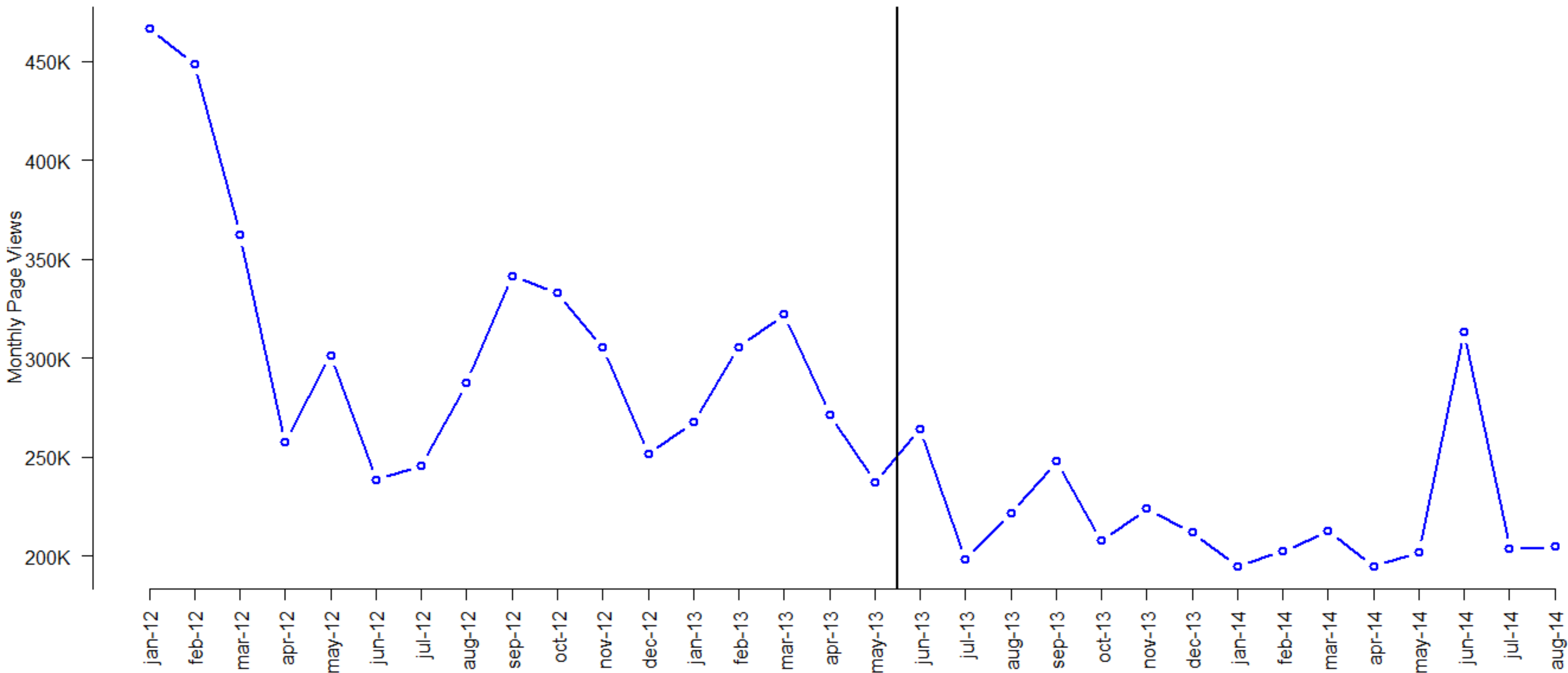
### Page Views for Hezbollah



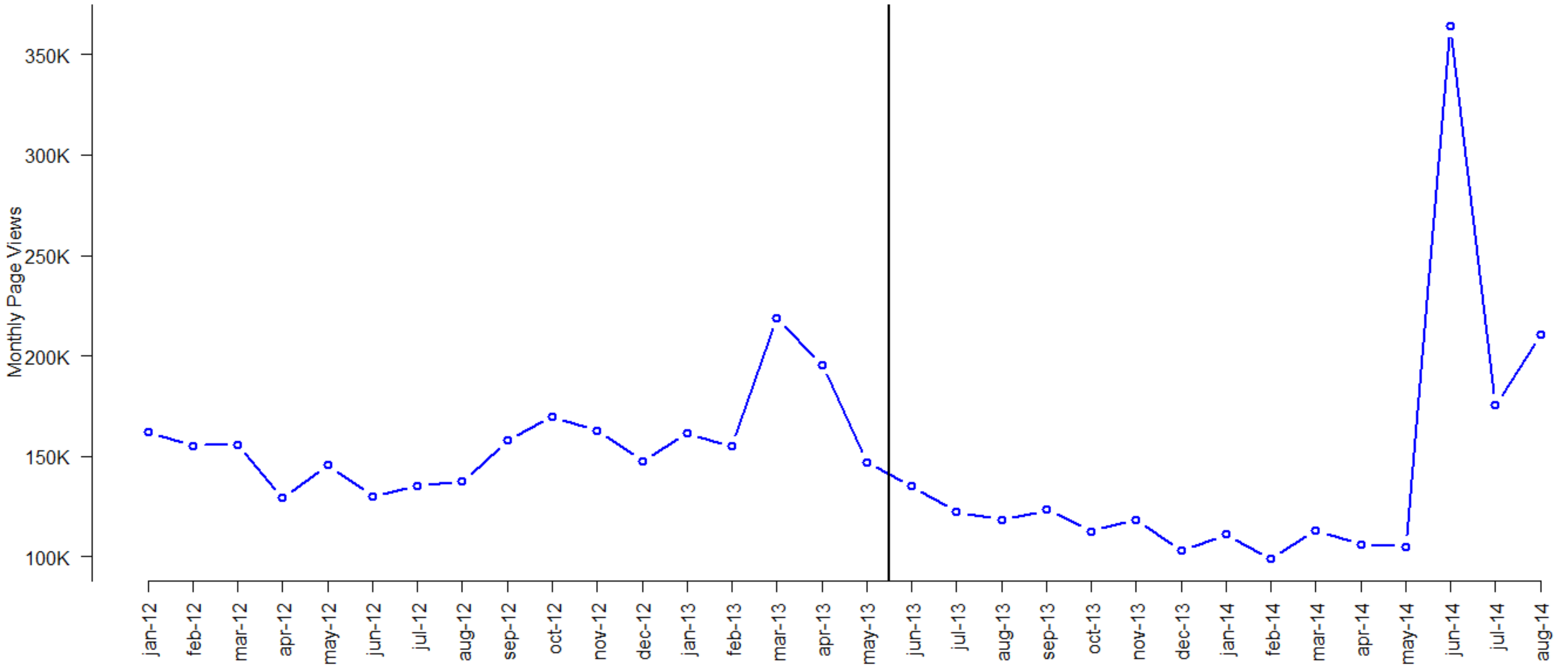
# Page Views for Improvised\_explosive\_device



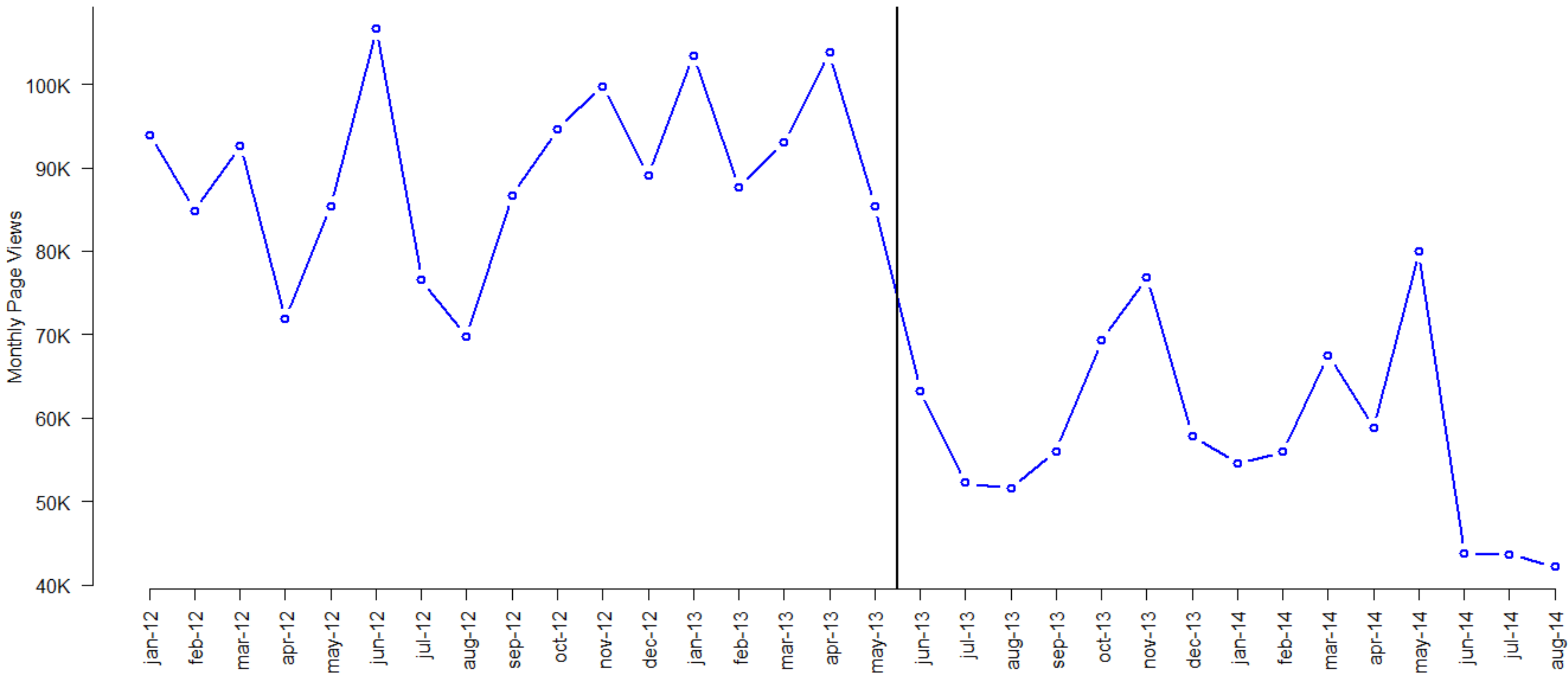
### Page Views for Iran



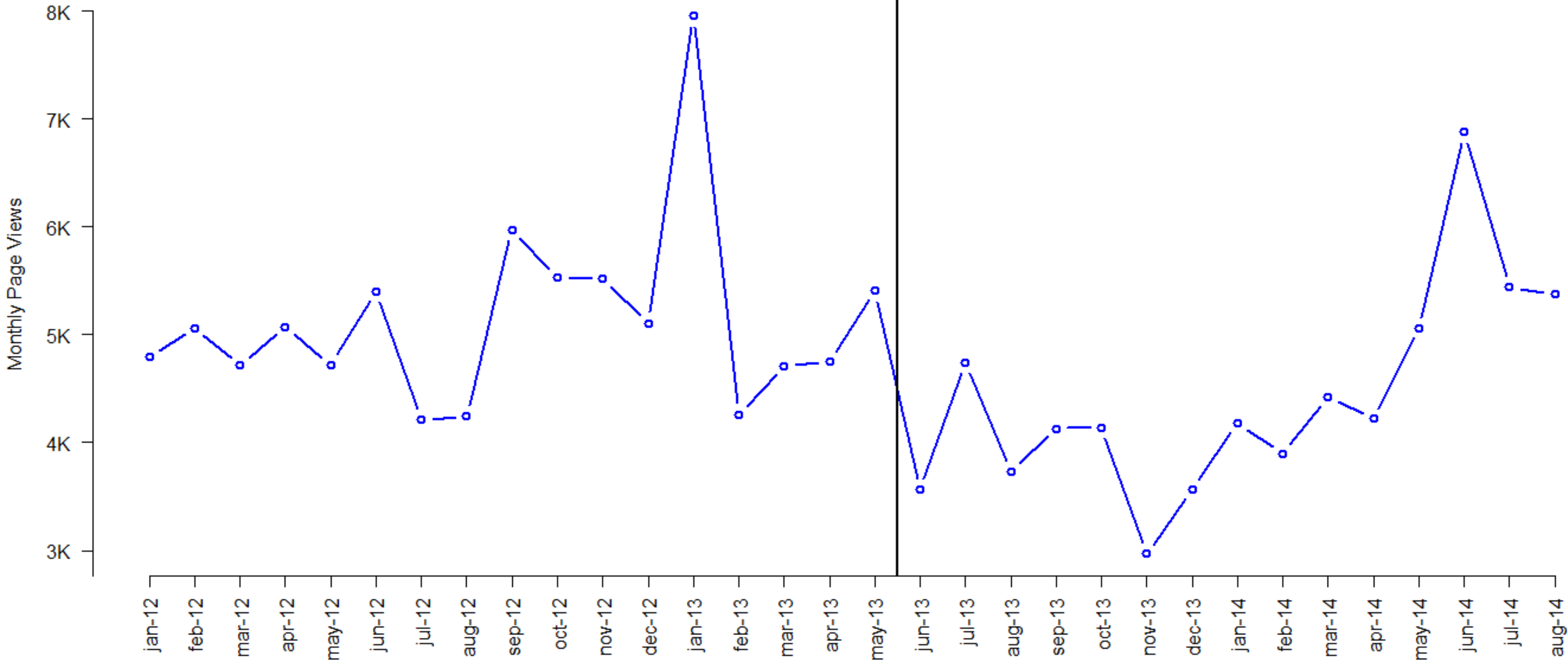
### Page Views for Iraq



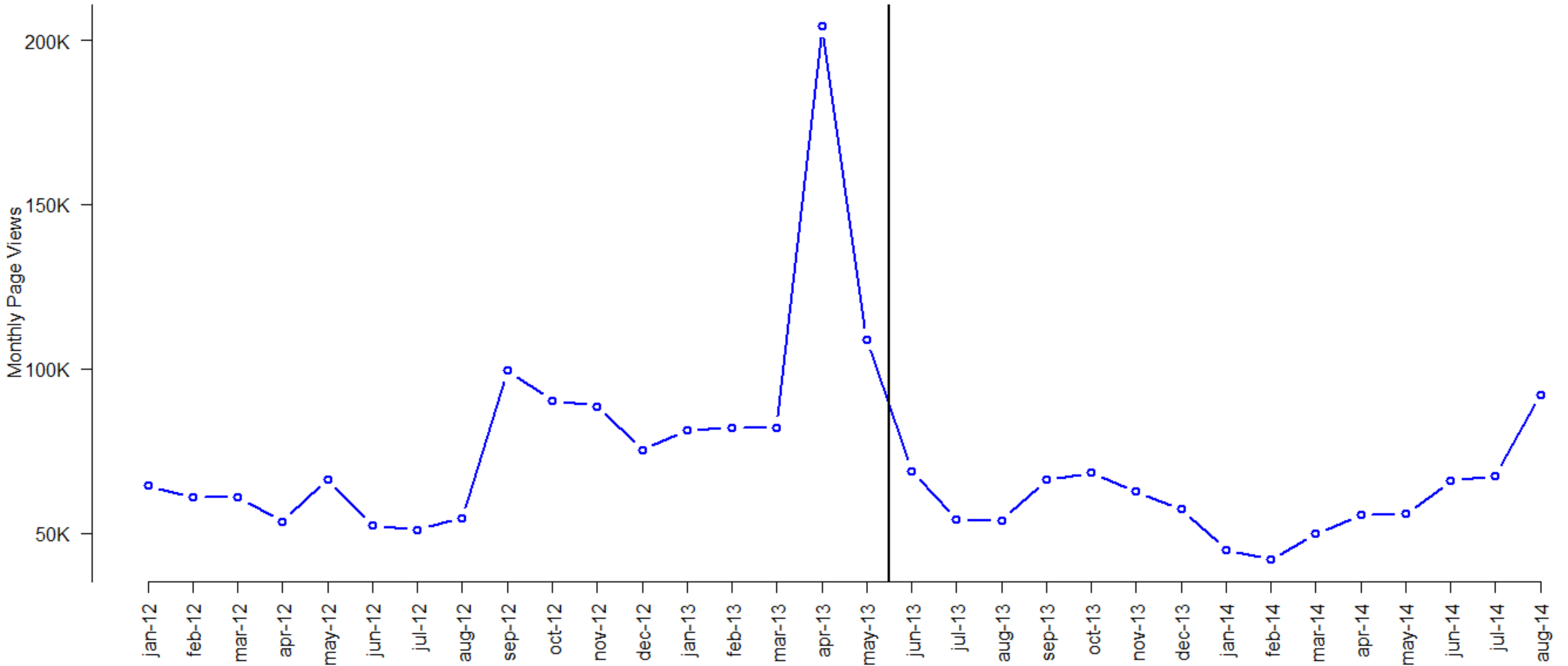
### Page Views for Irish\_Republican\_Army



### Page Views for Islamist

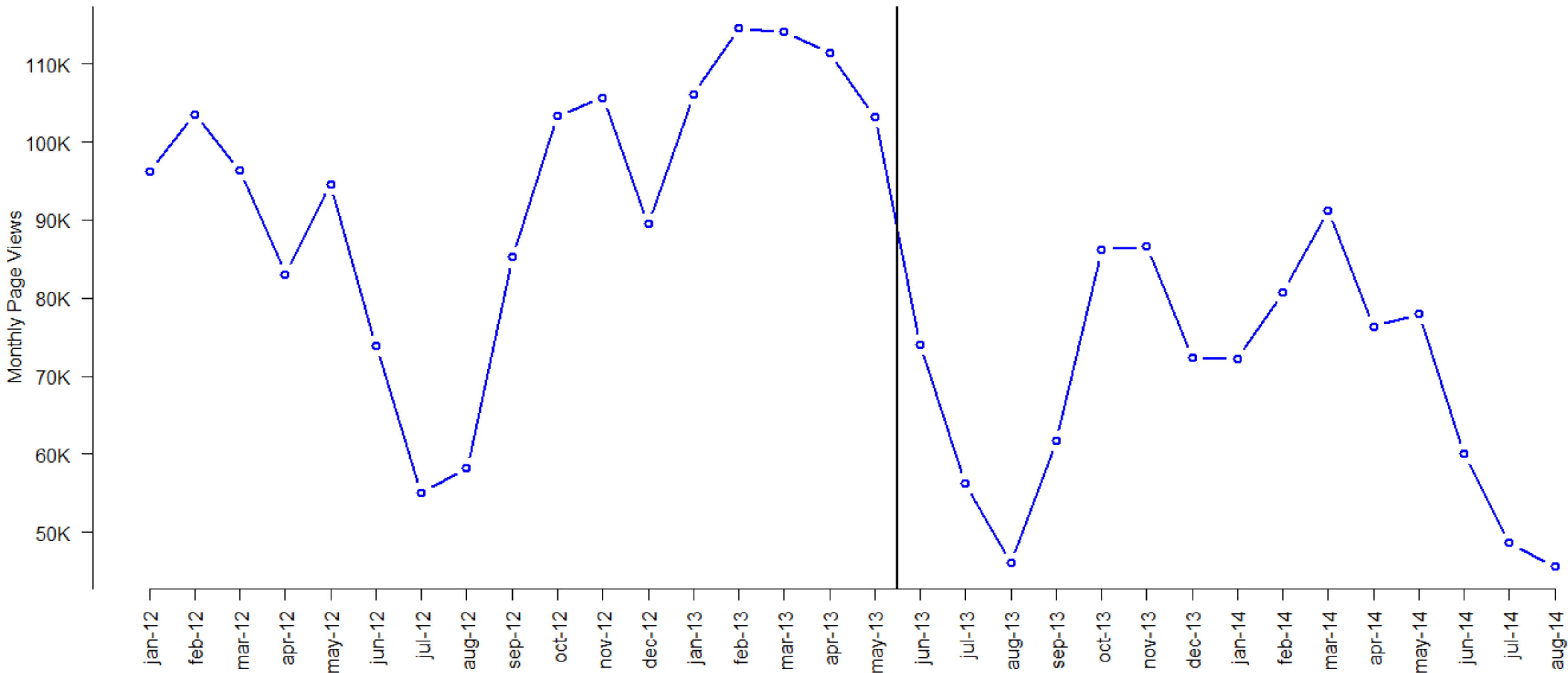


### Page Views for Jihad

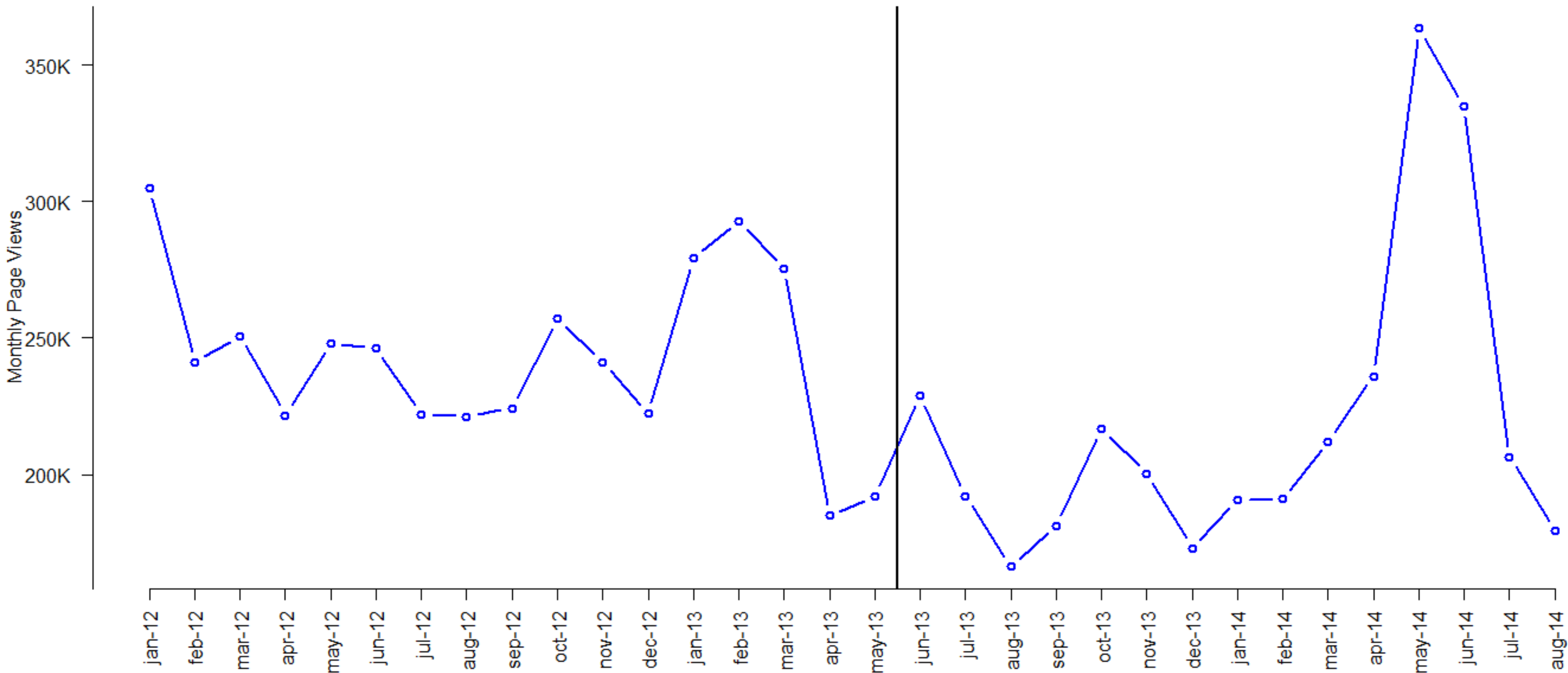




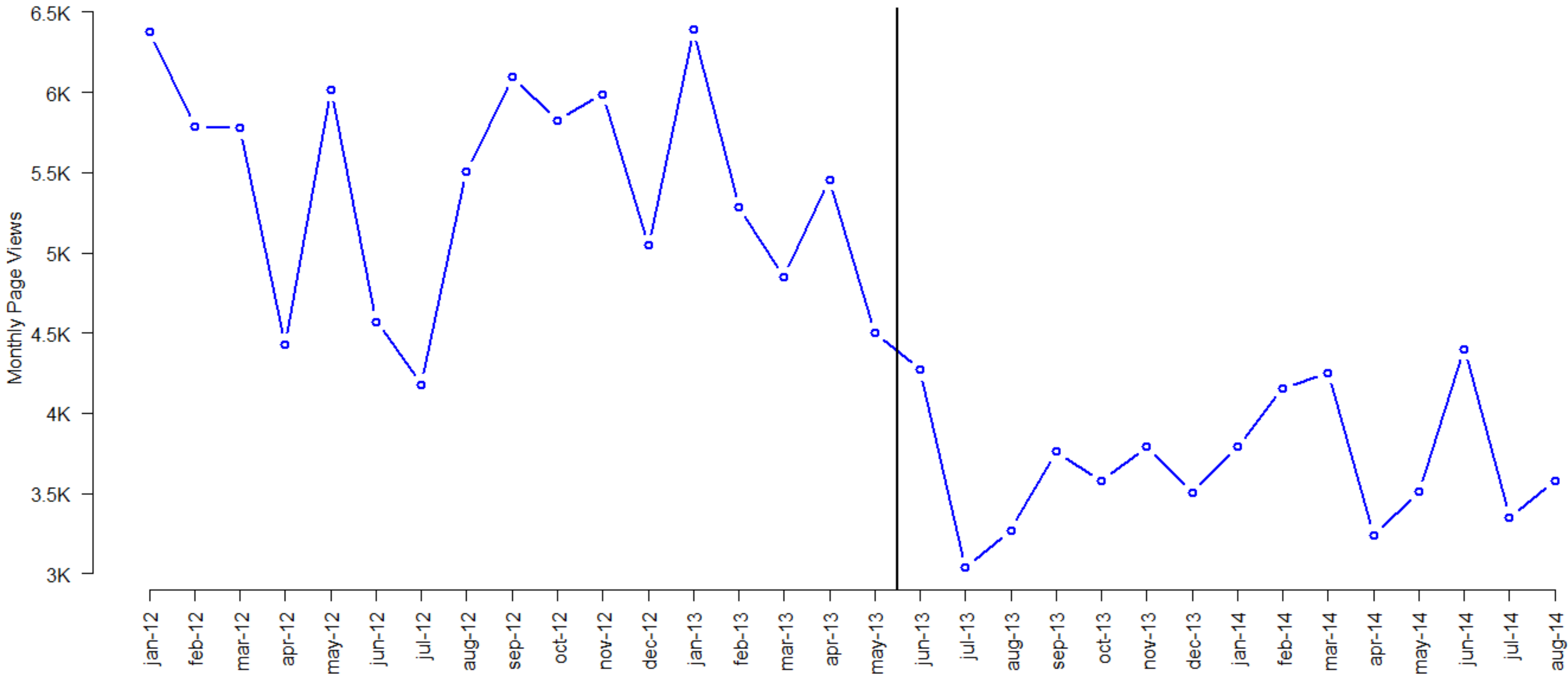
### Page Views for nationalism



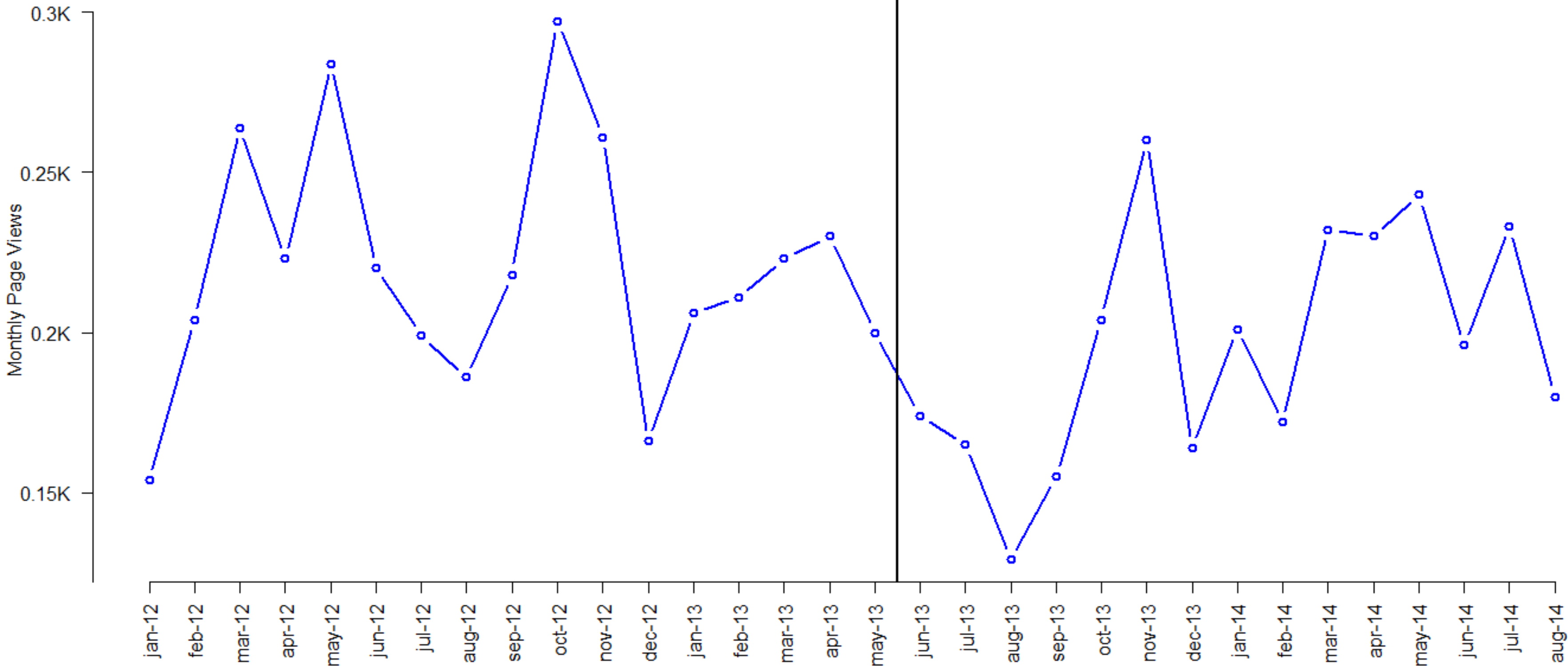
### Page Views for Nigeria



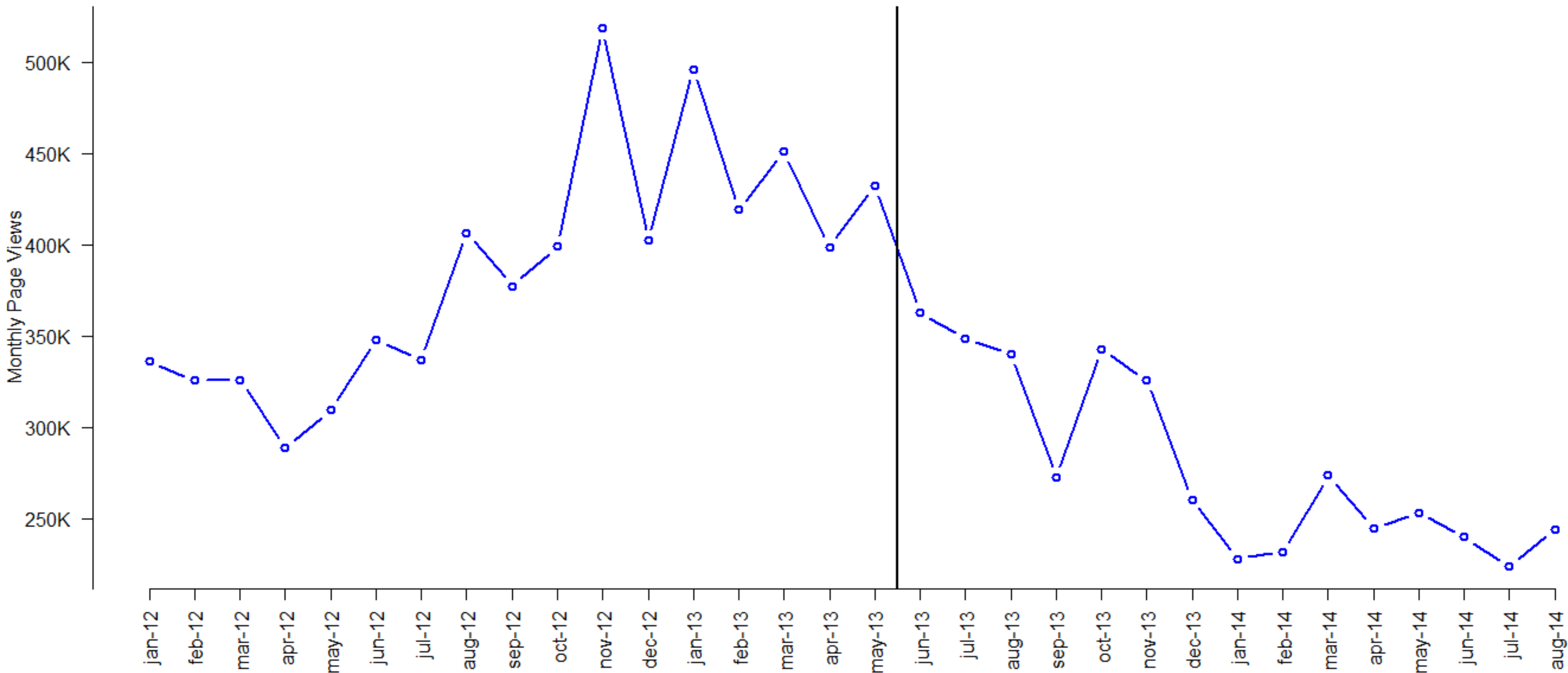
### Page Views for Nuclear



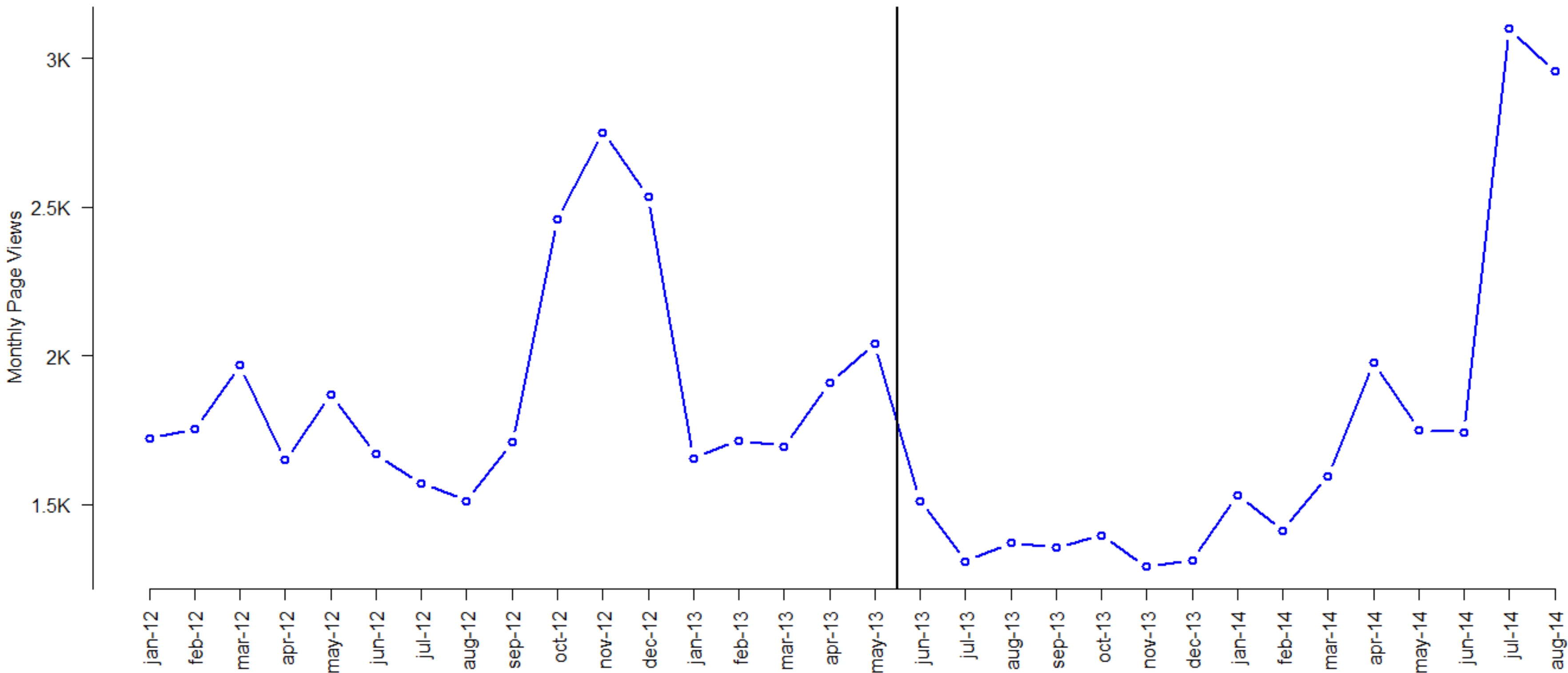
### Page Views for Nuclear\_Enrichment



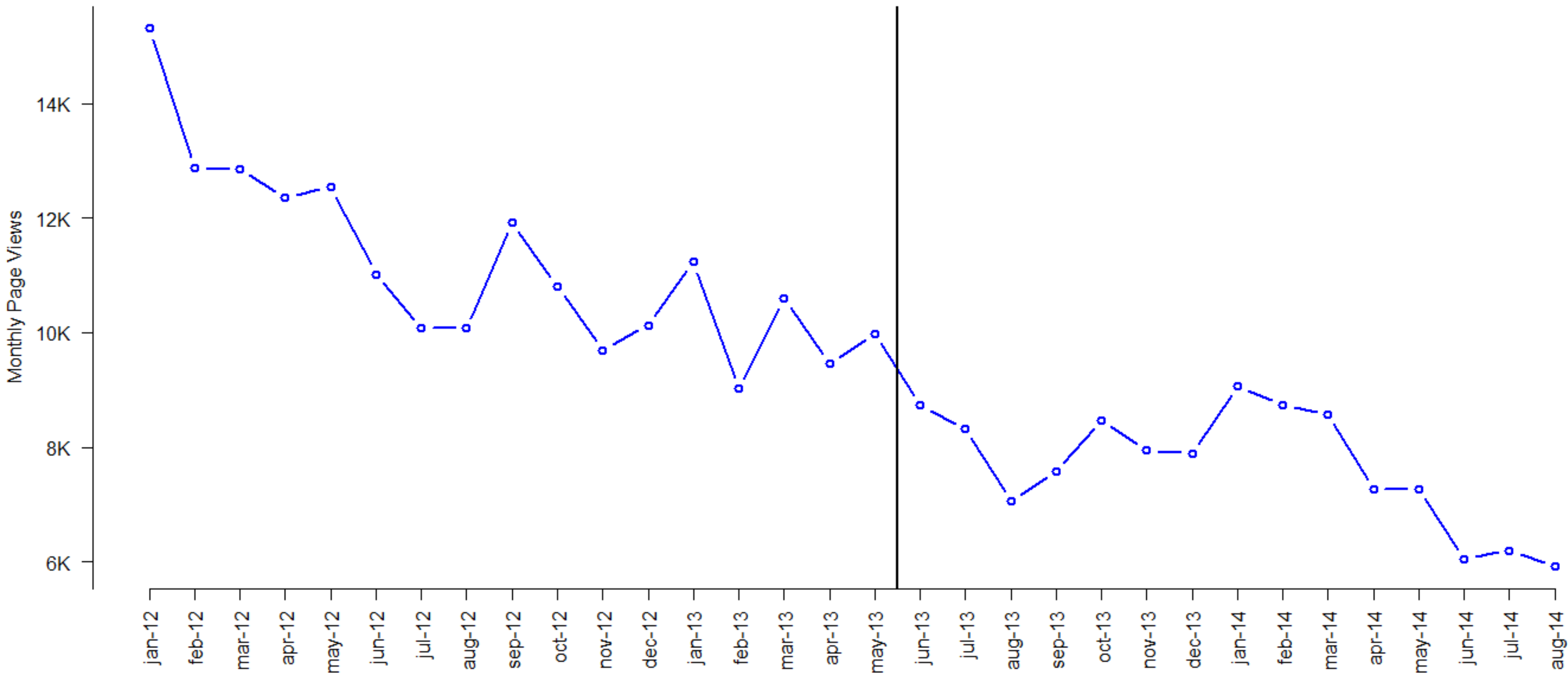
### Page Views for Pakistan



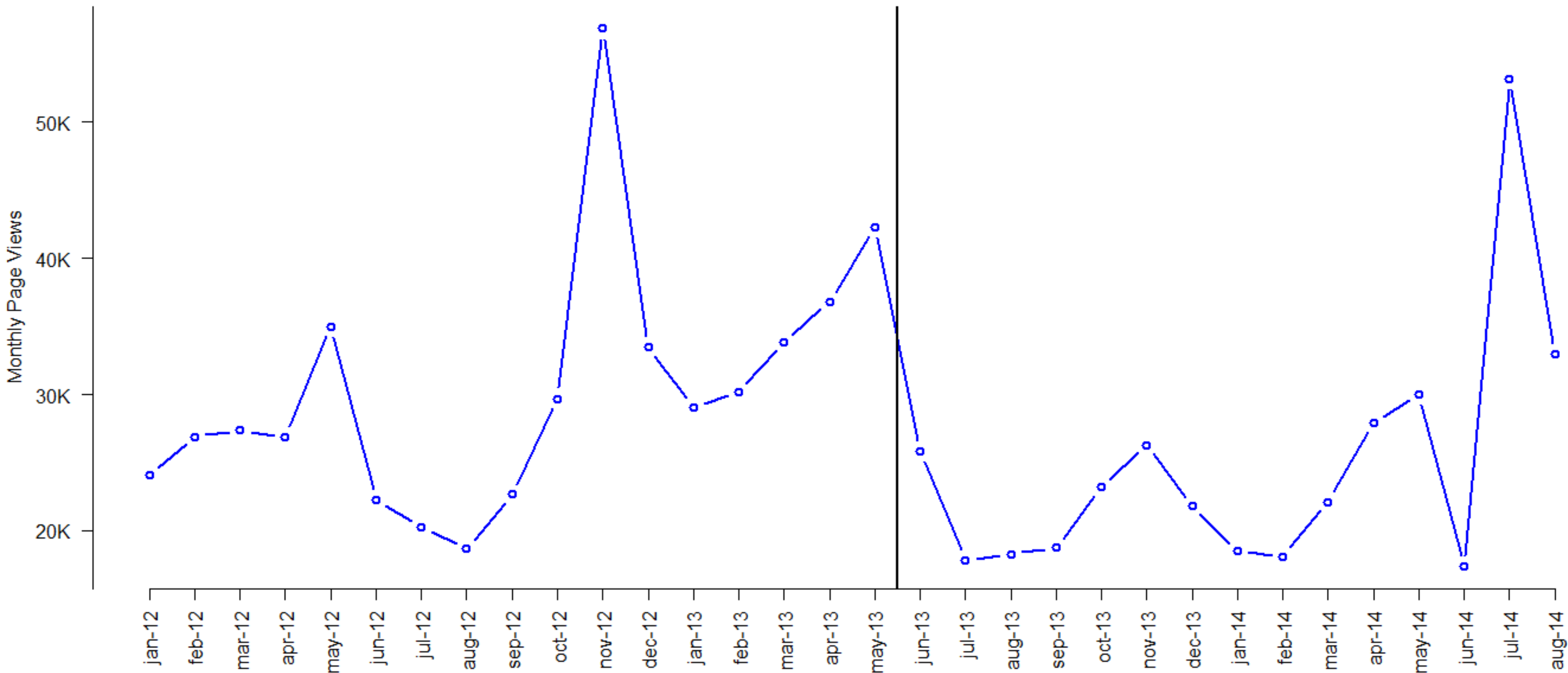
### Page Views for Palestine\_Liberation\_Fron



### Page Views for Pirates

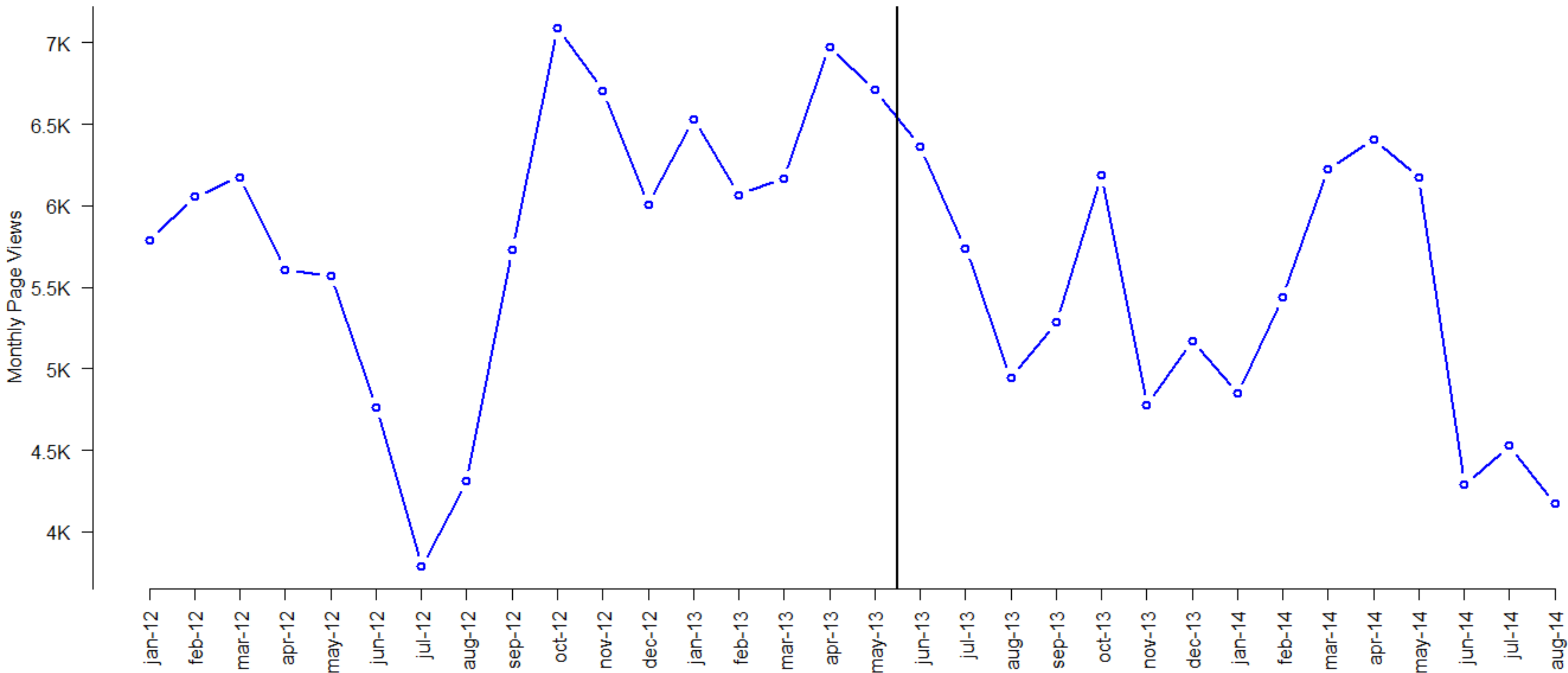


### Page Views for PLO

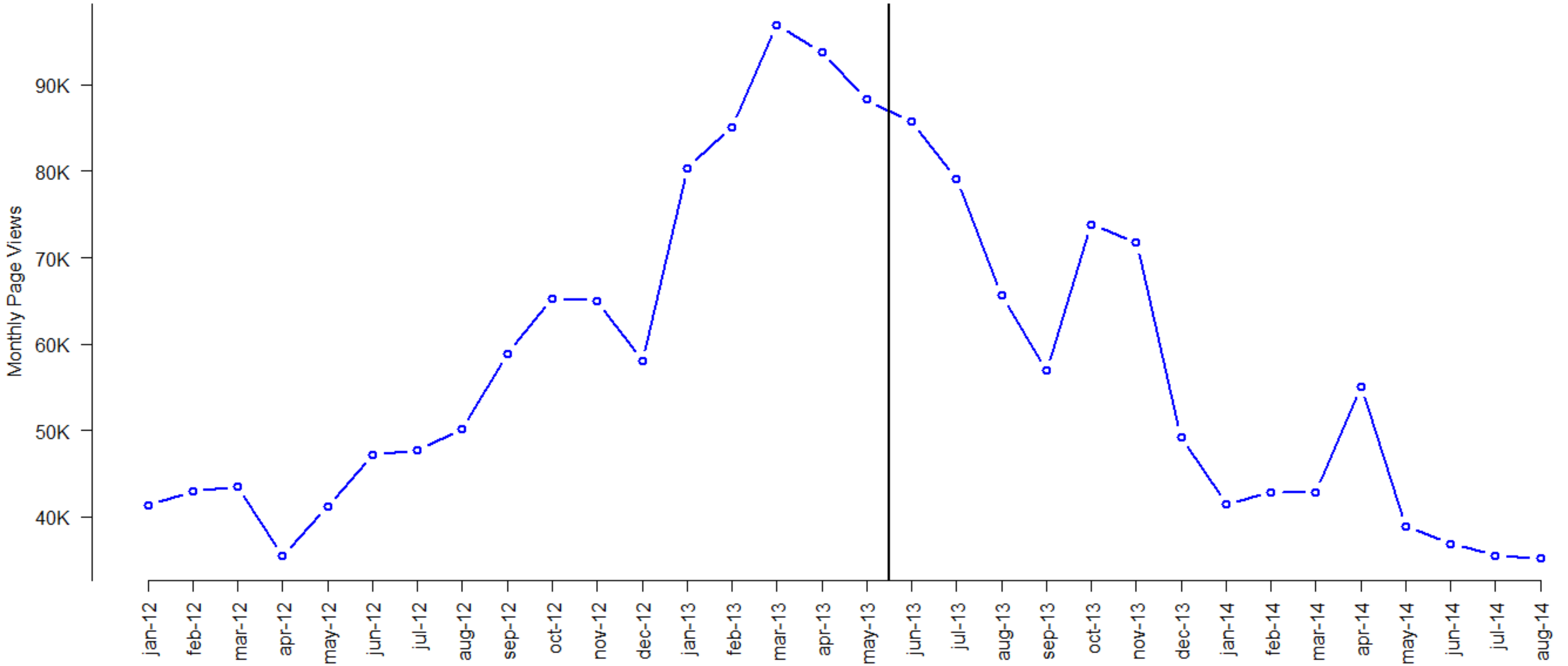




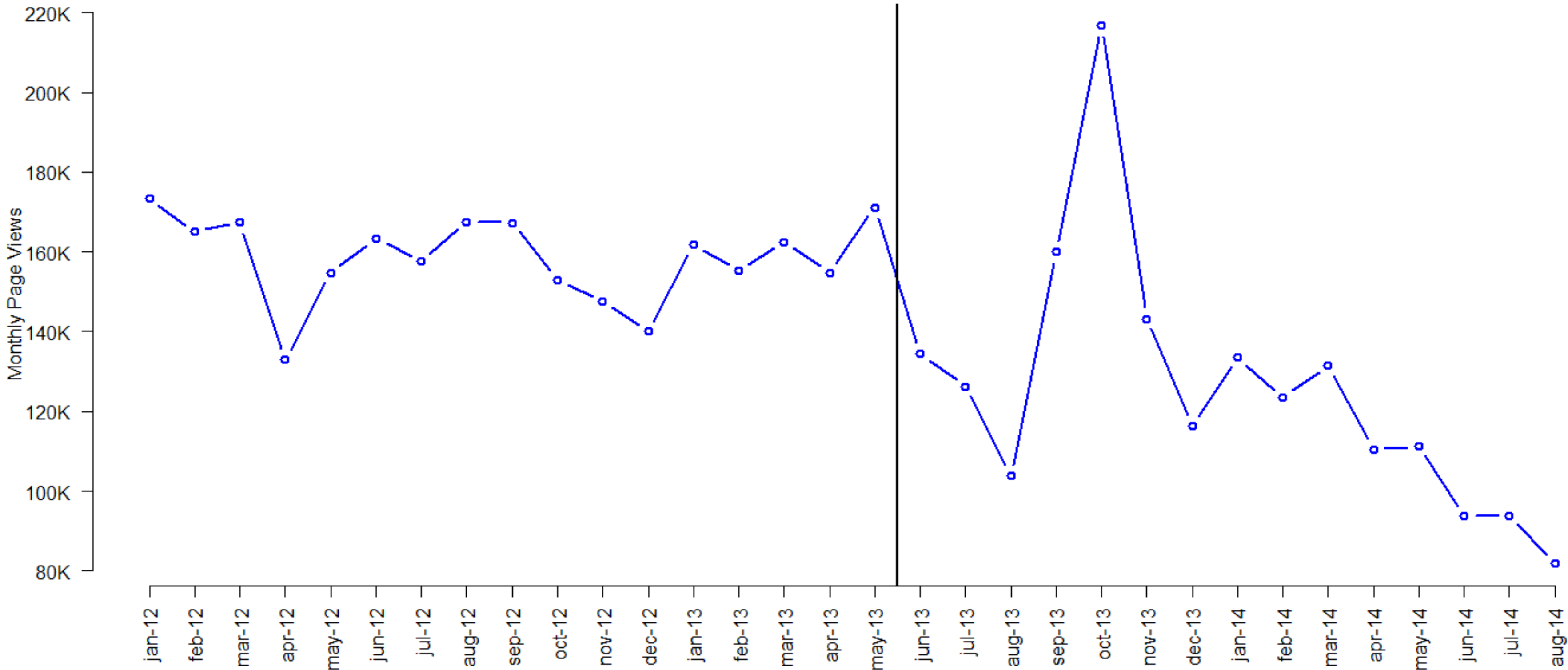
### Page Views for Political\_radicalism



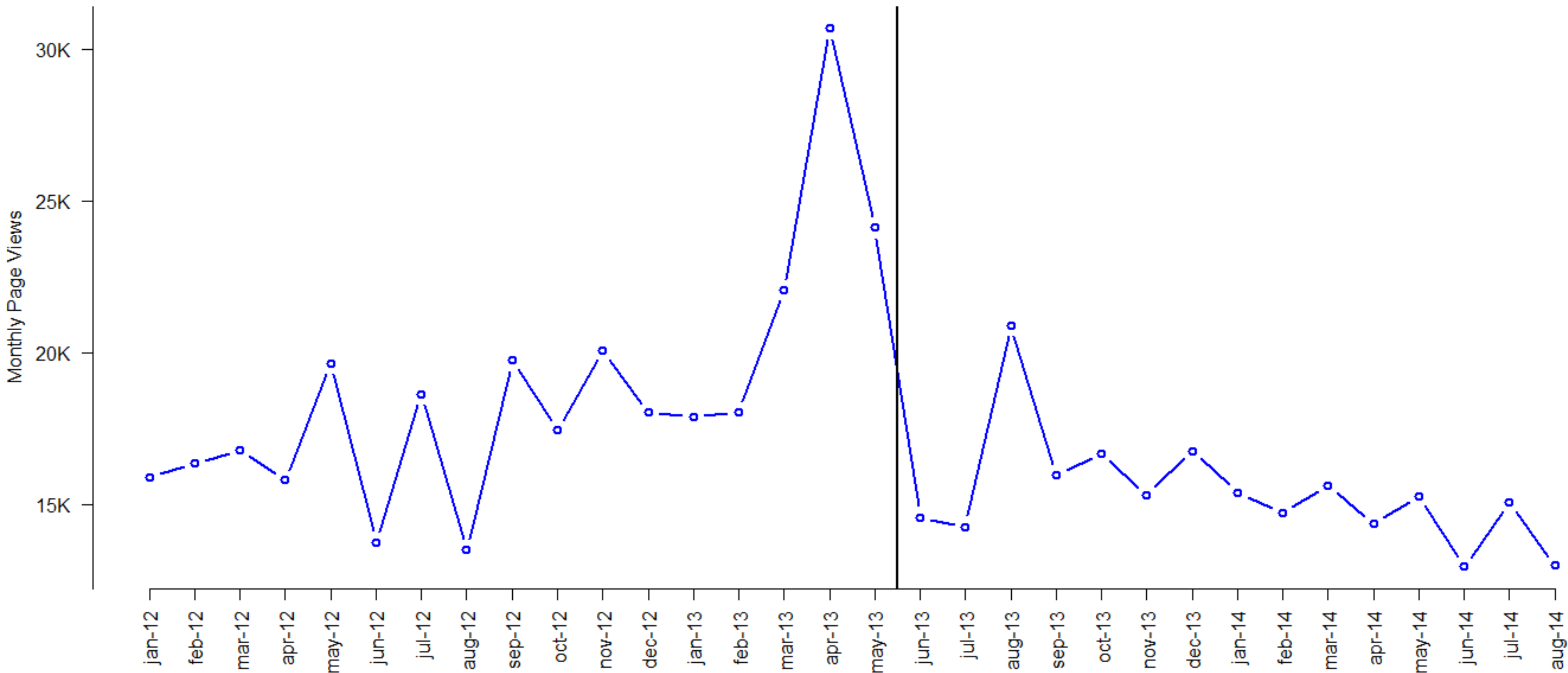
### Page Views for Recruitment



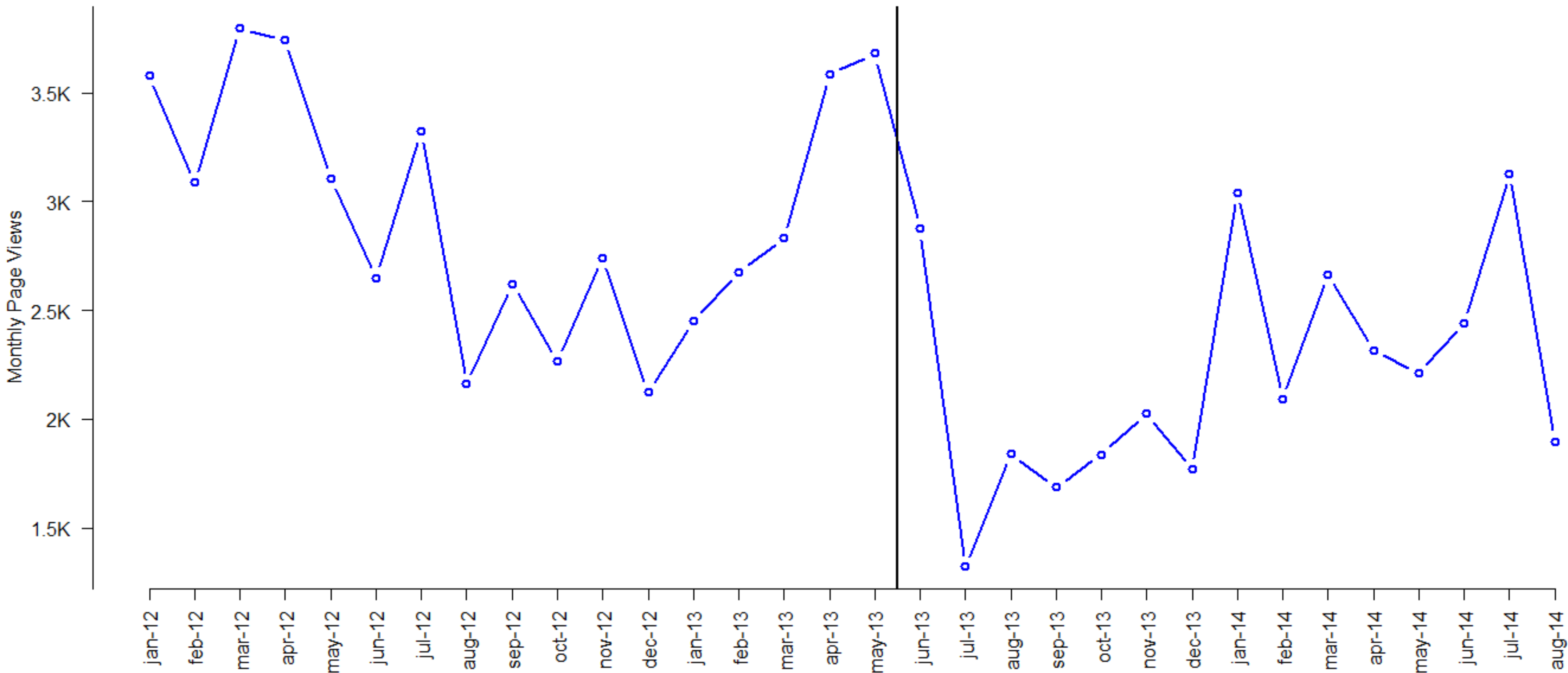
### Page Views for Somalia



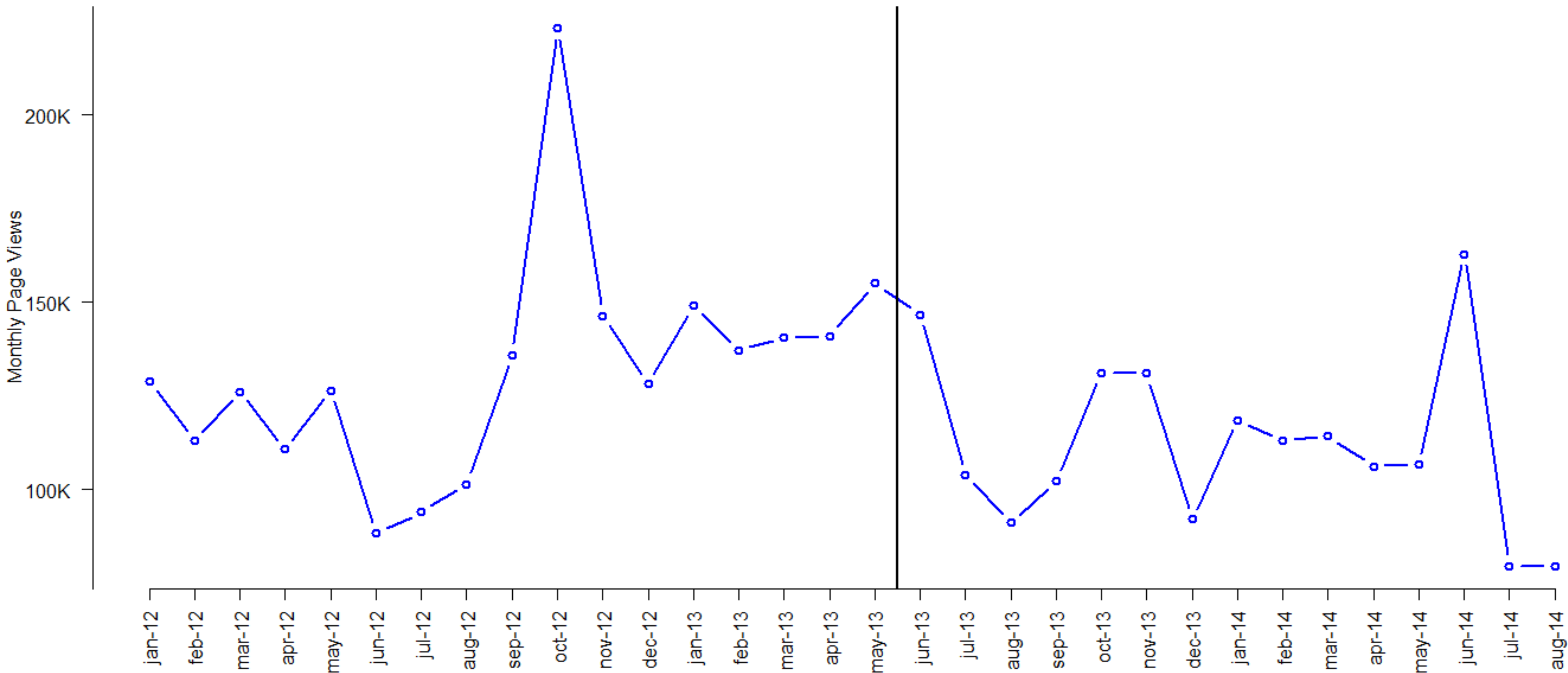
### Page Views for Suicide\_attack



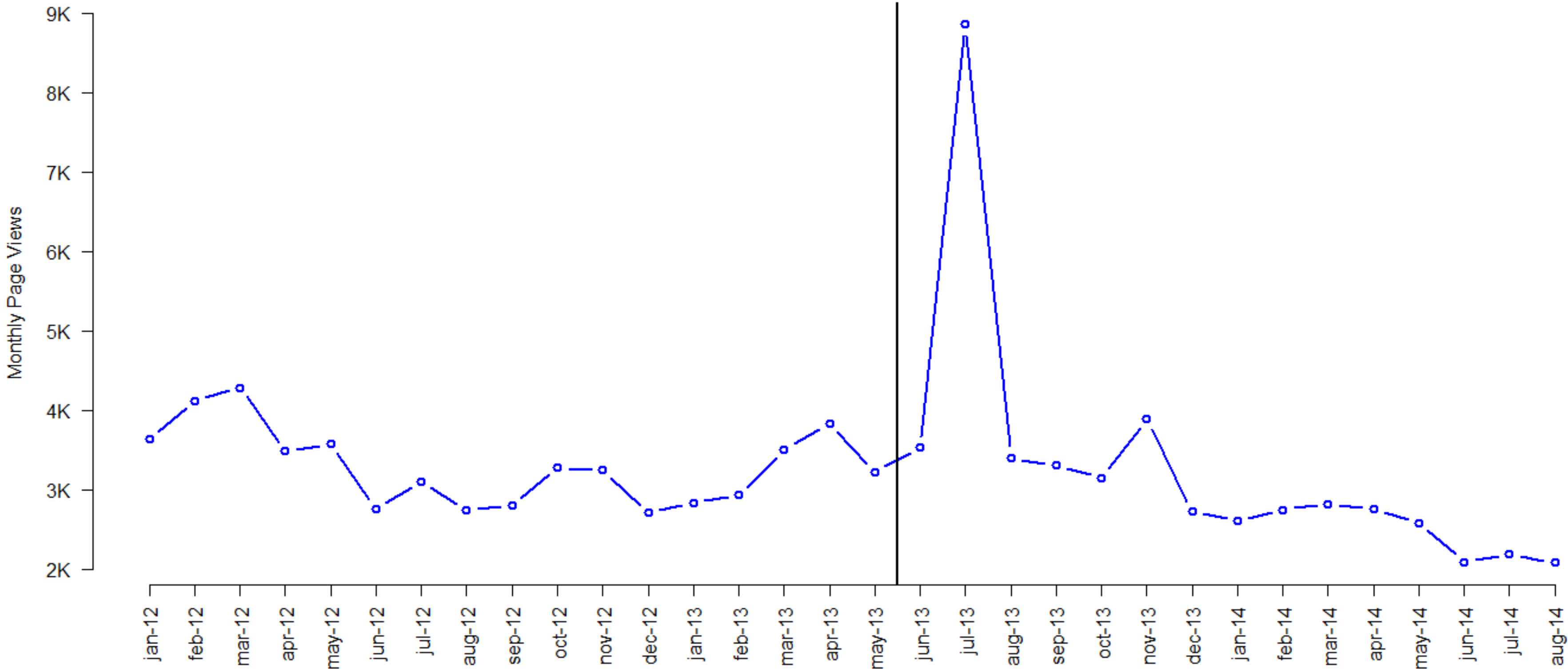
### Page Views for Suicide\_bomber



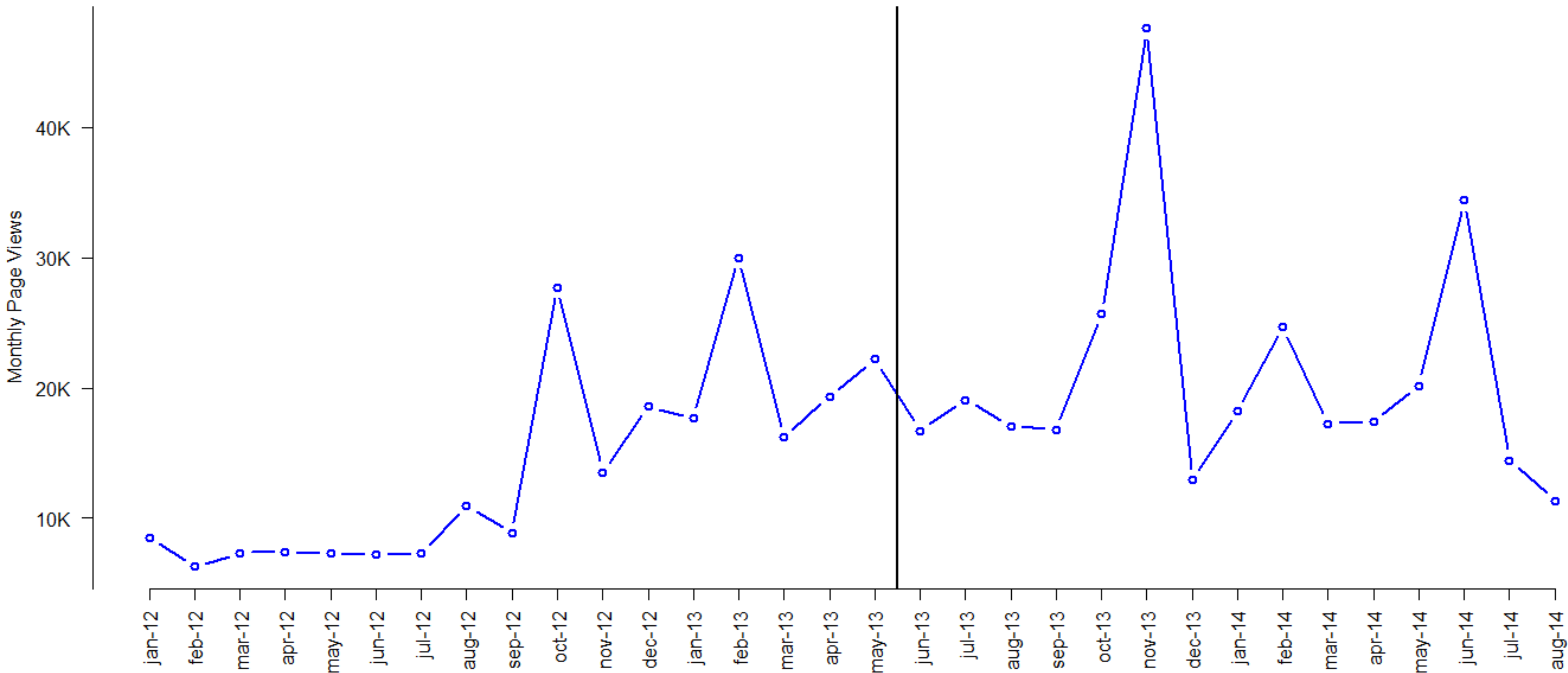
### Page Views for Taliban



### Page Views for Tamil\_Tigers

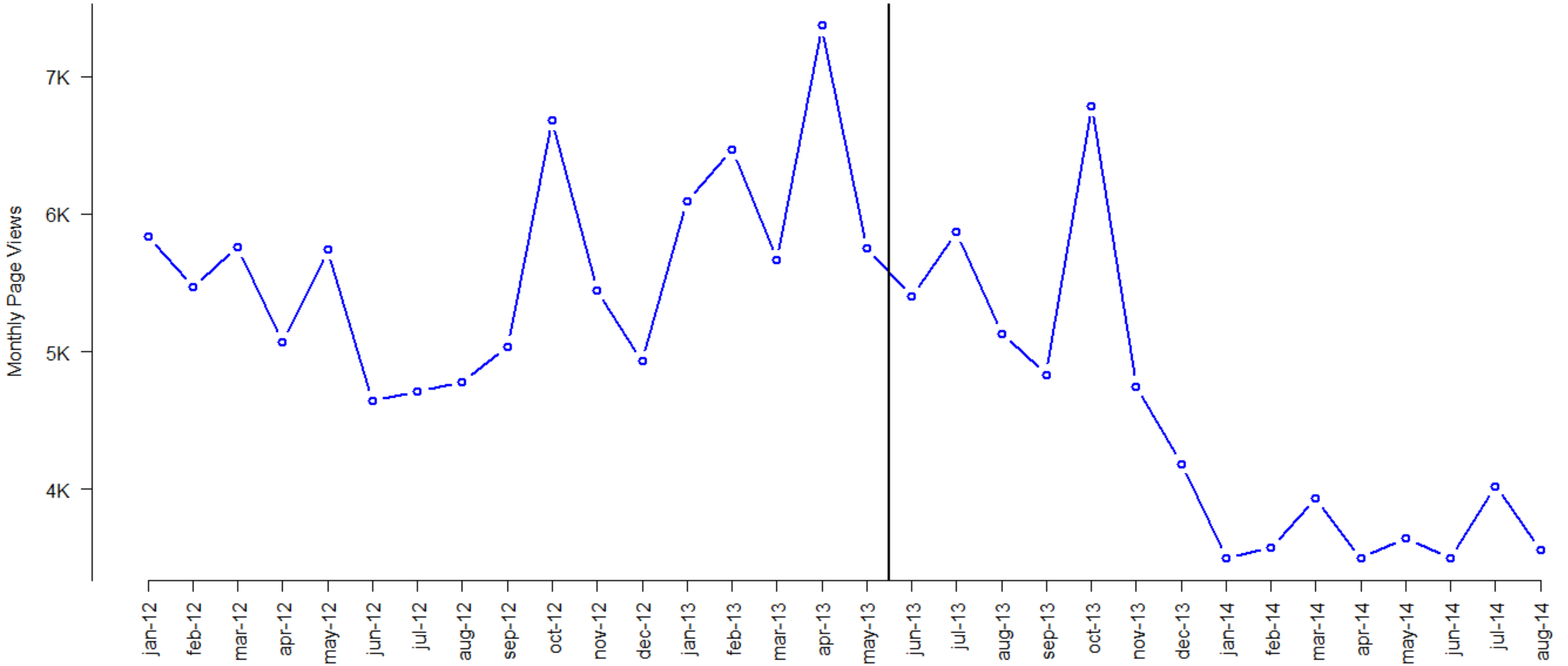


### Page Views for Tehrik\_i\_Taliban\_Pakistan

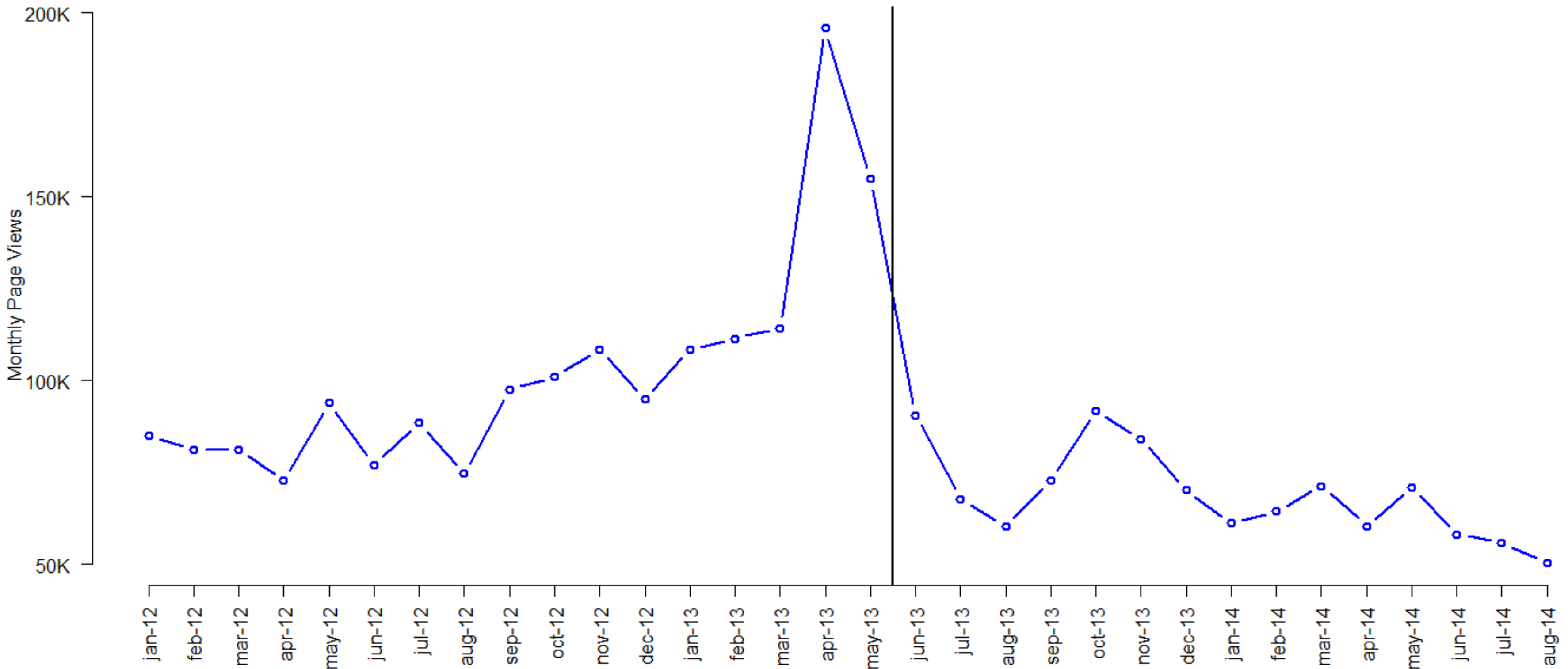




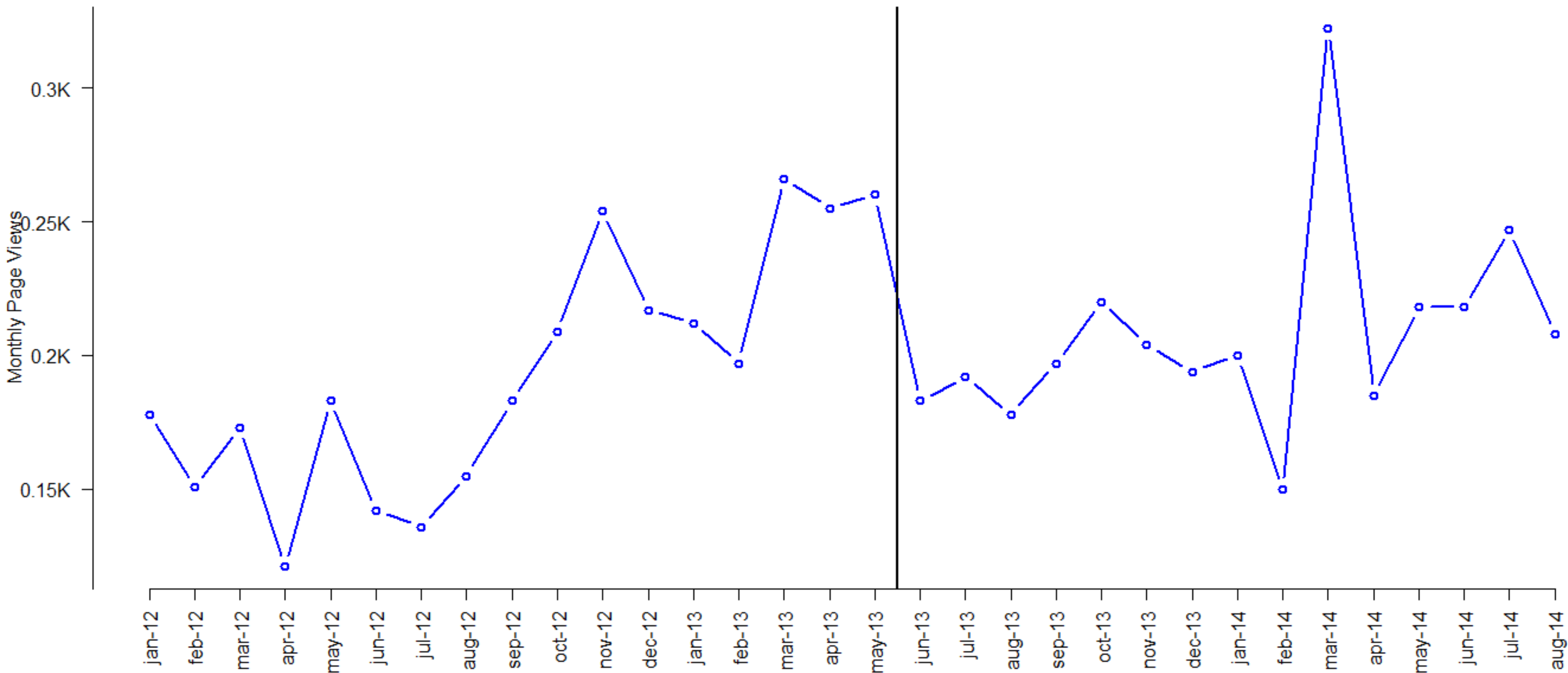
### Page Views for terror



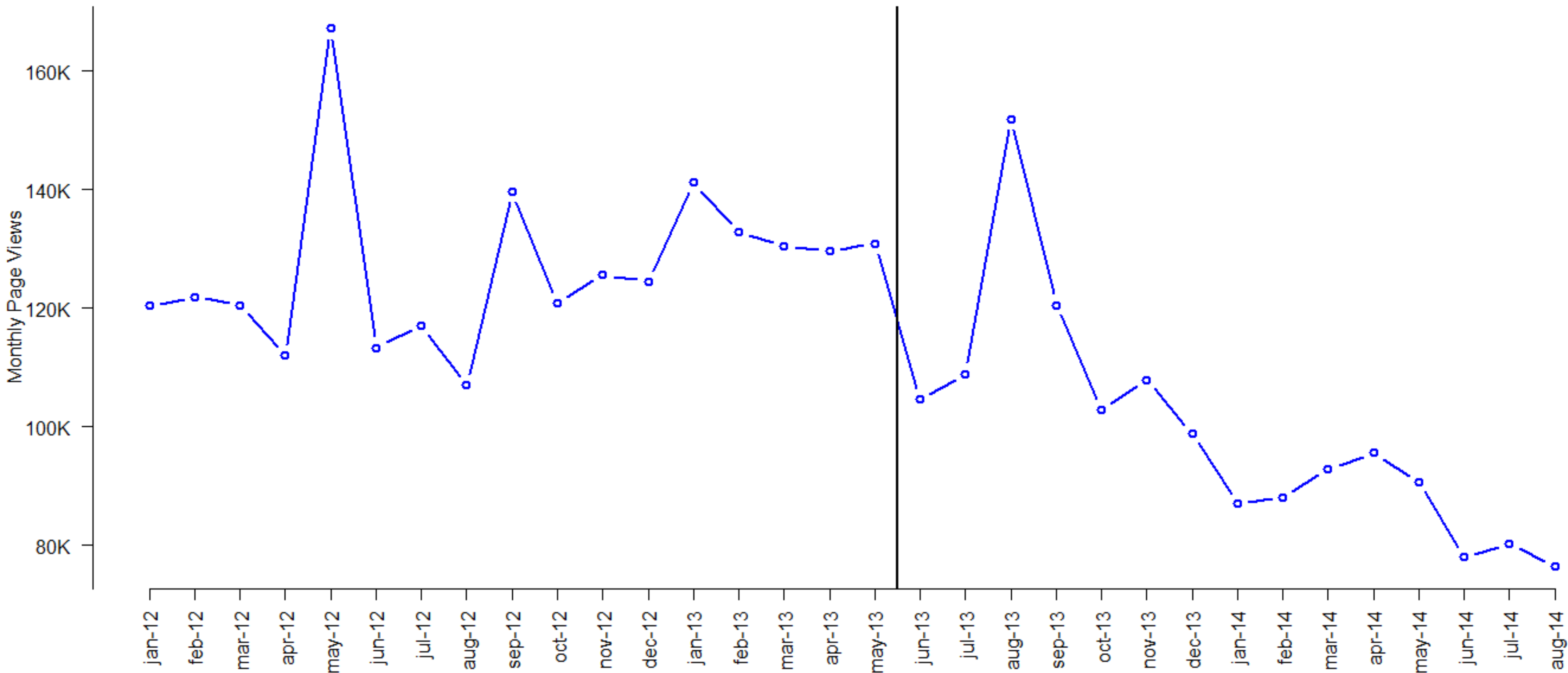
### Page Views for terrorism



### Page Views for Weapons\_grade

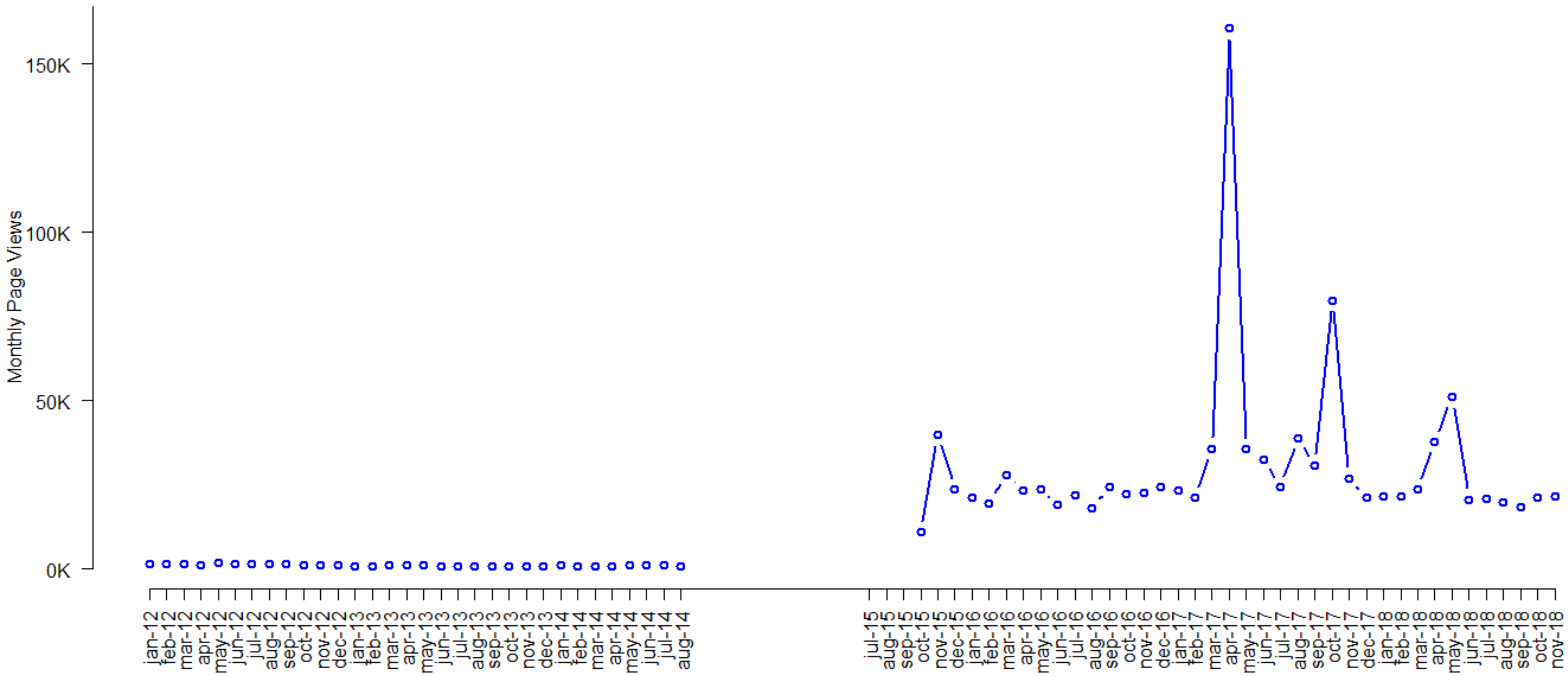


### Page Views for Yemen

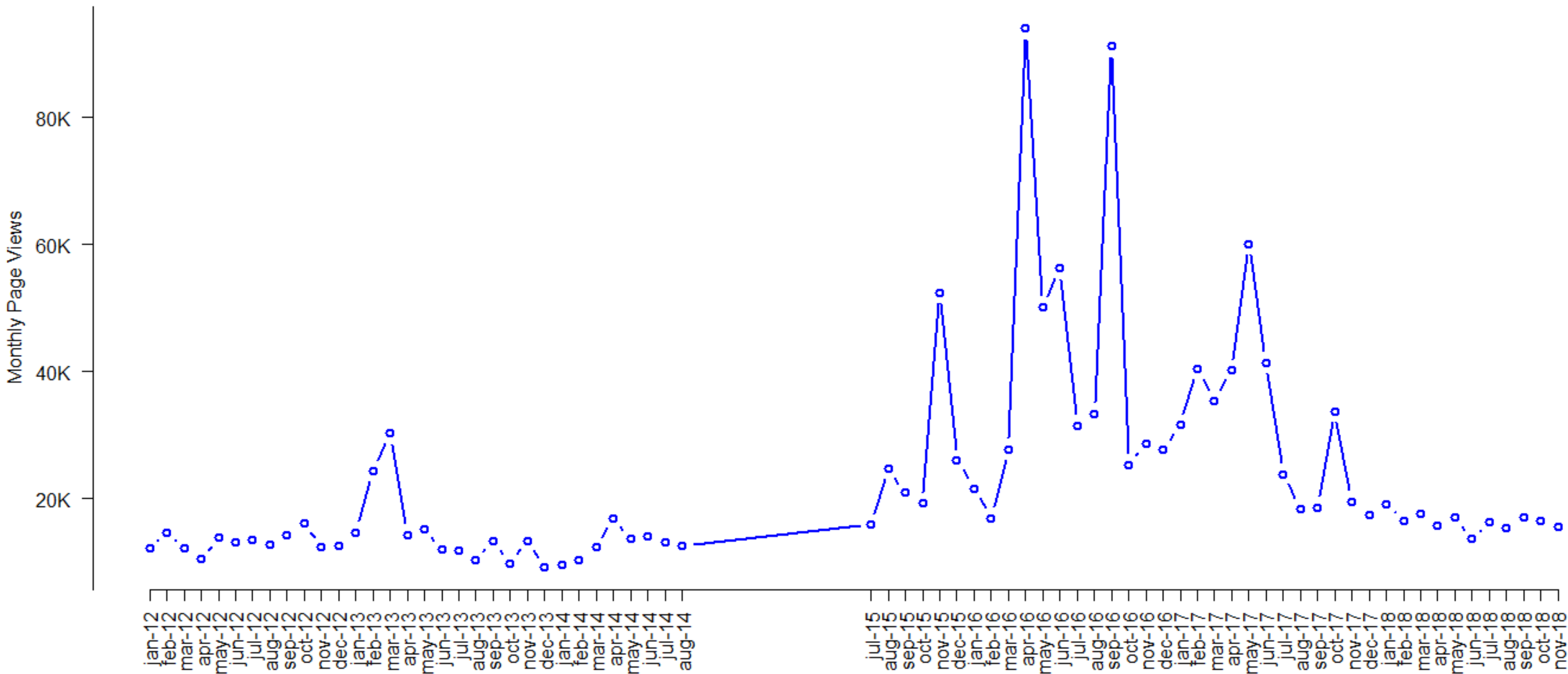


**APPENDIX V: Page Views for 48 Terror Articles, Extended Time Period**

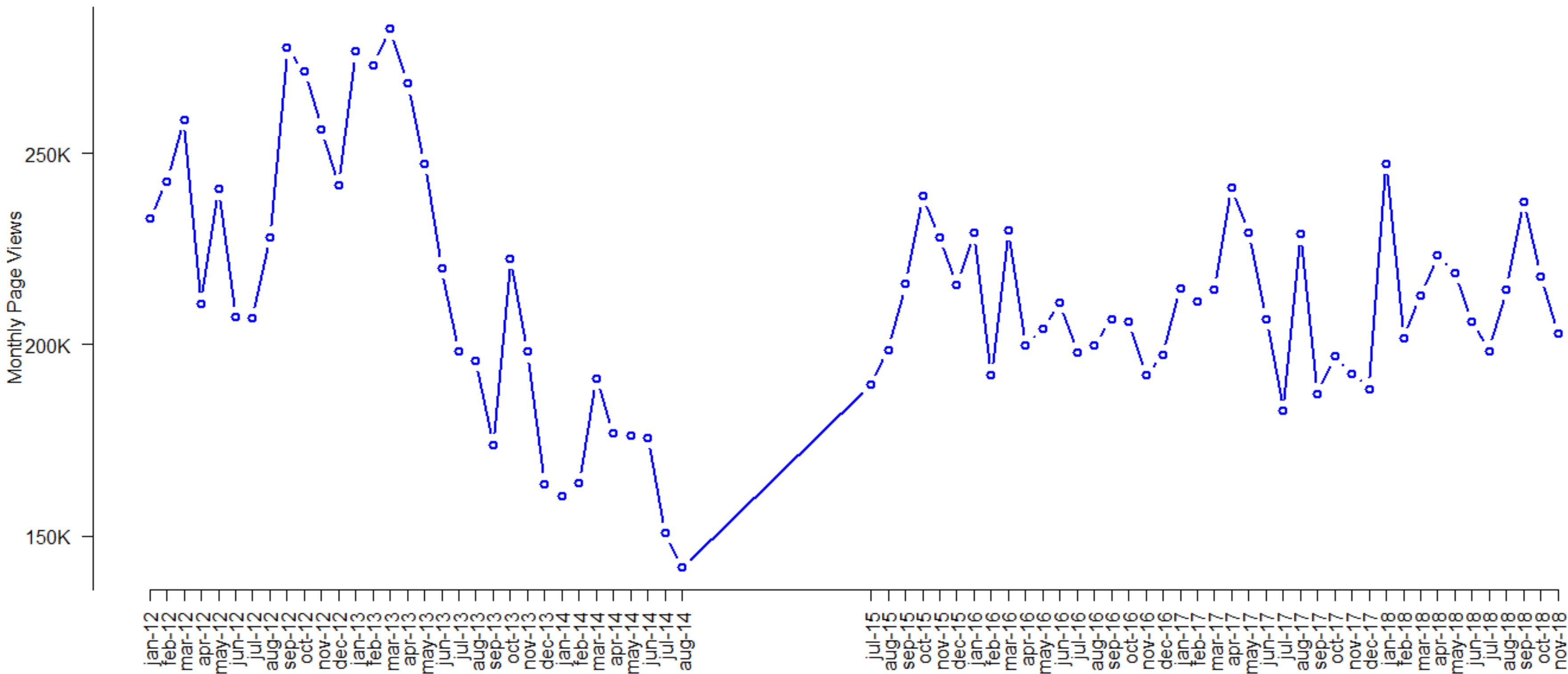
### Page Views for \_Euskadi\_ta\_Askatasuna



### Page Views for Abu\_Sayyaf

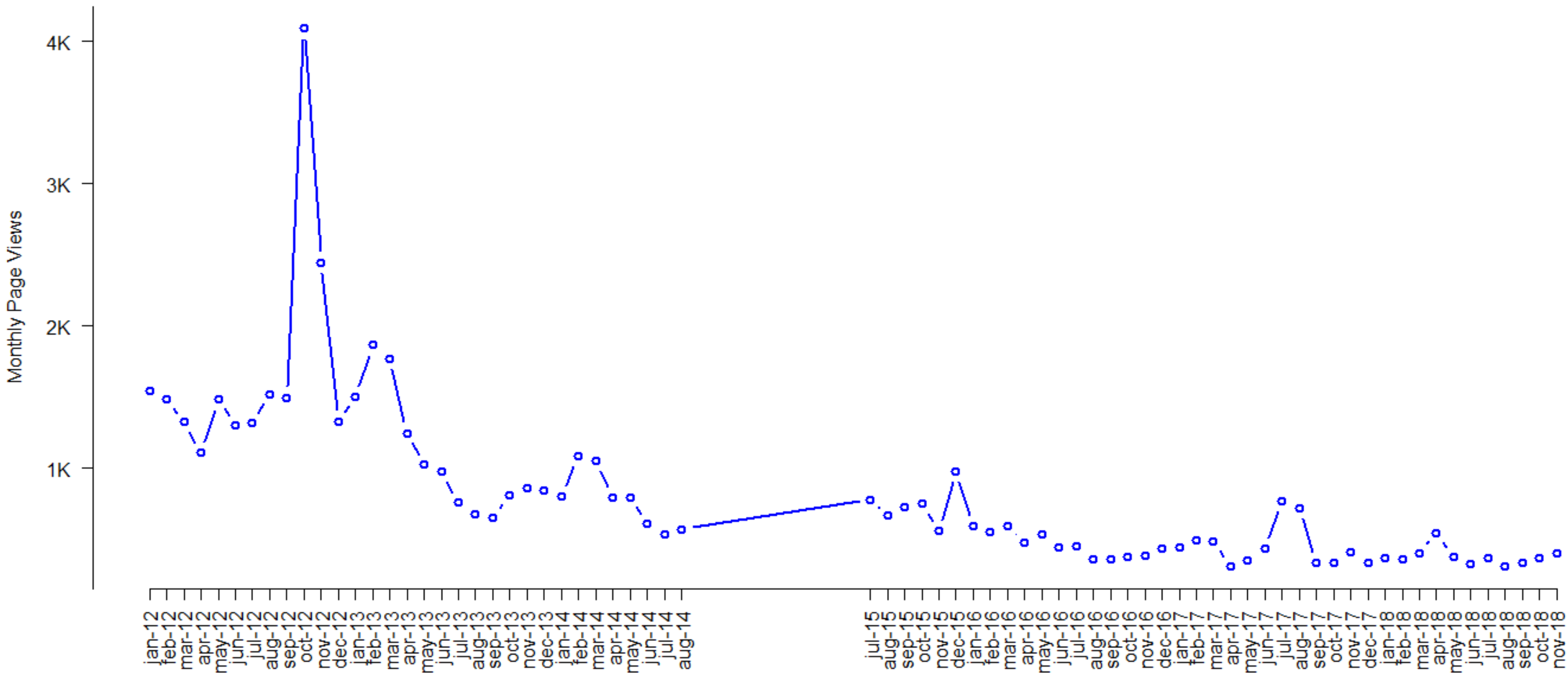


### Page Views for Afghanistan

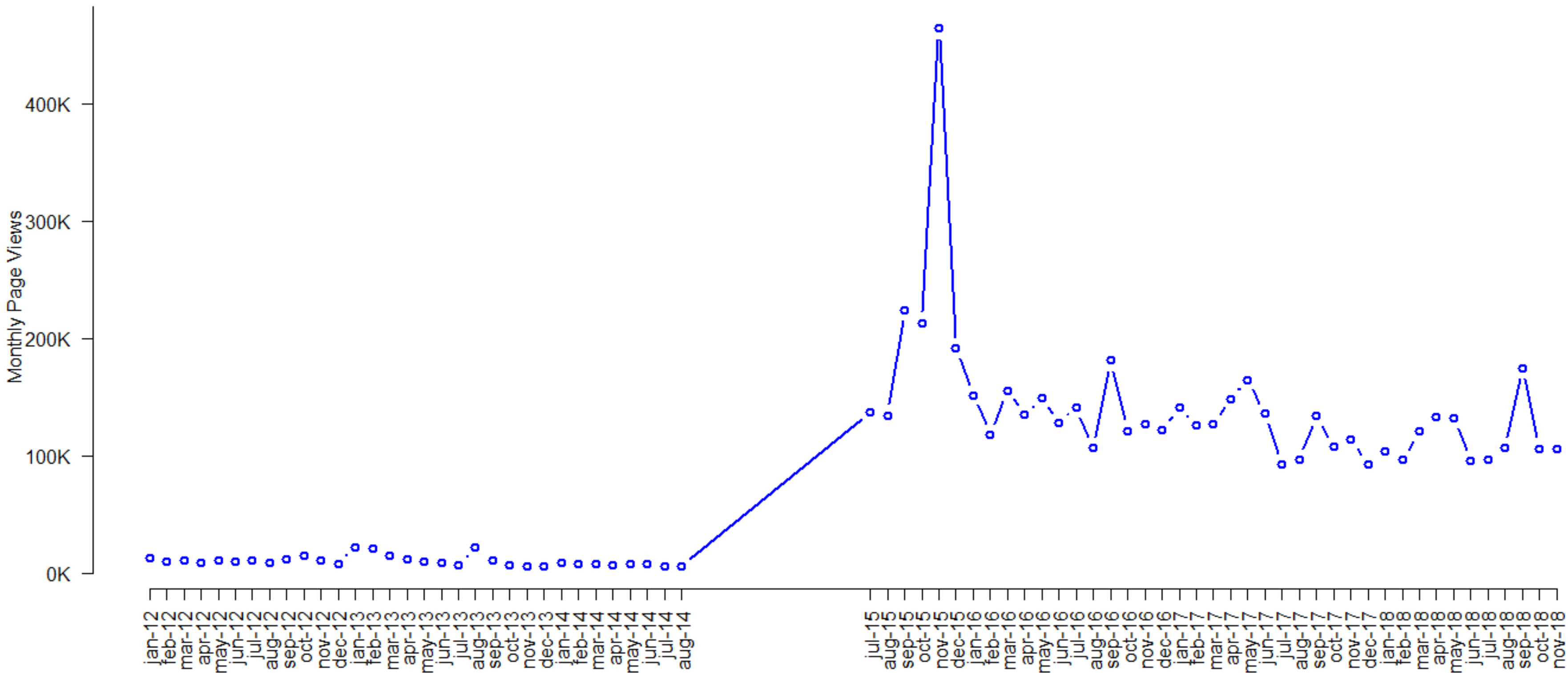


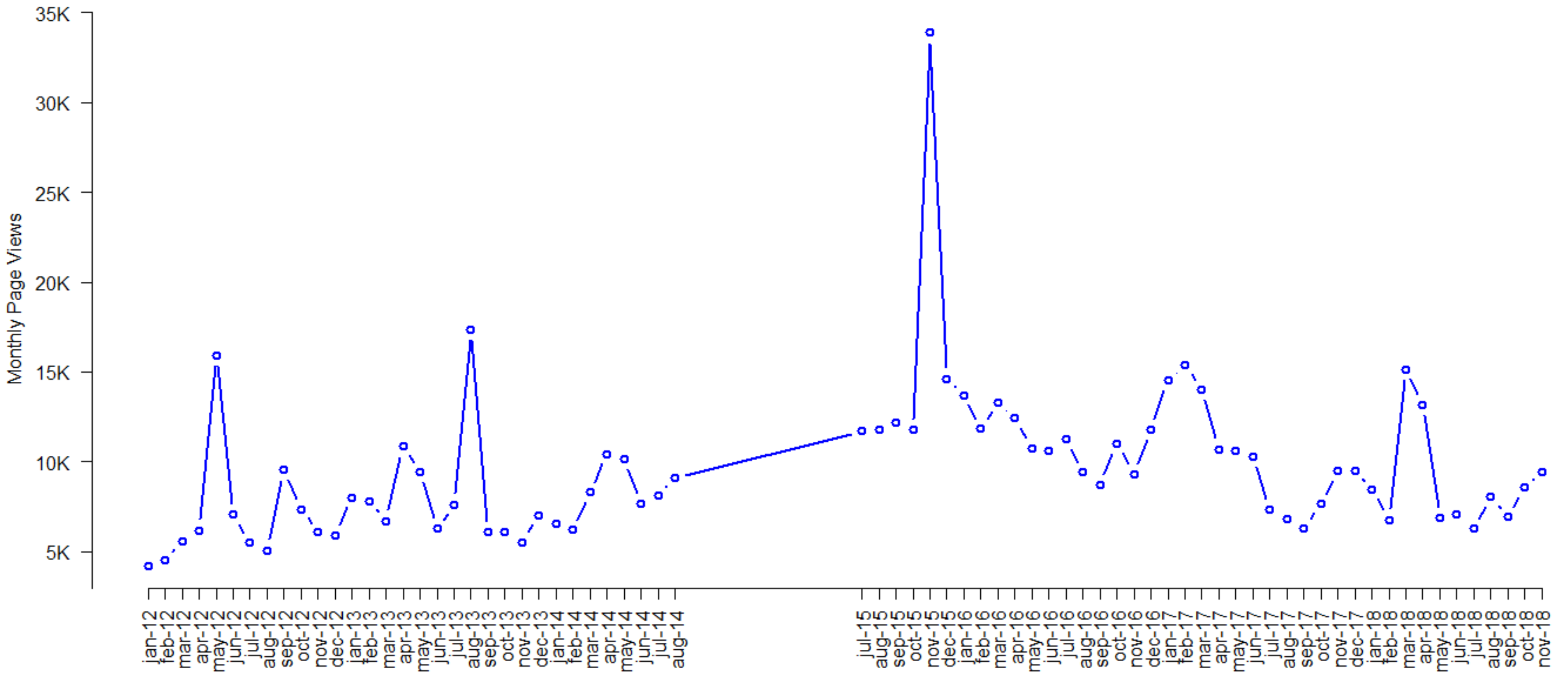


### Page Views for agro

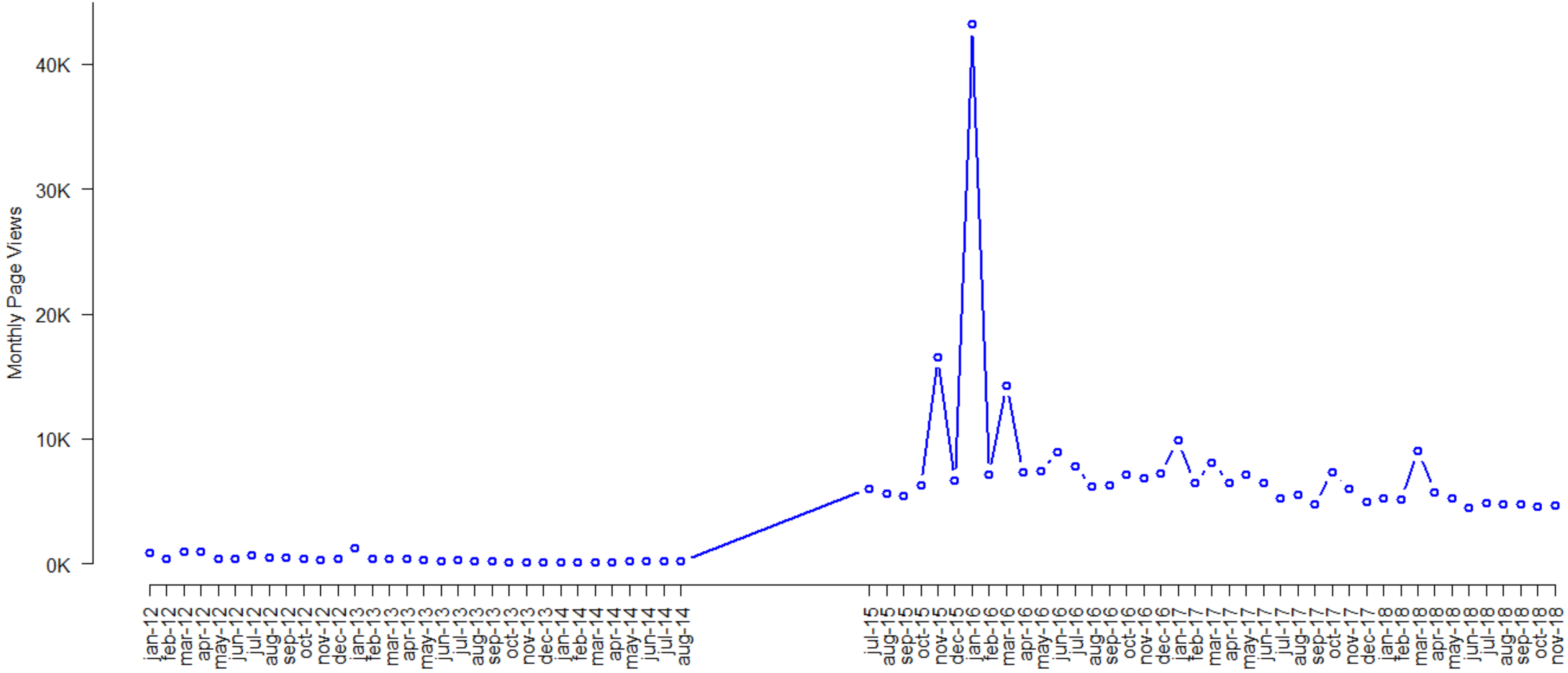


### Page Views for Al\_Qaeda

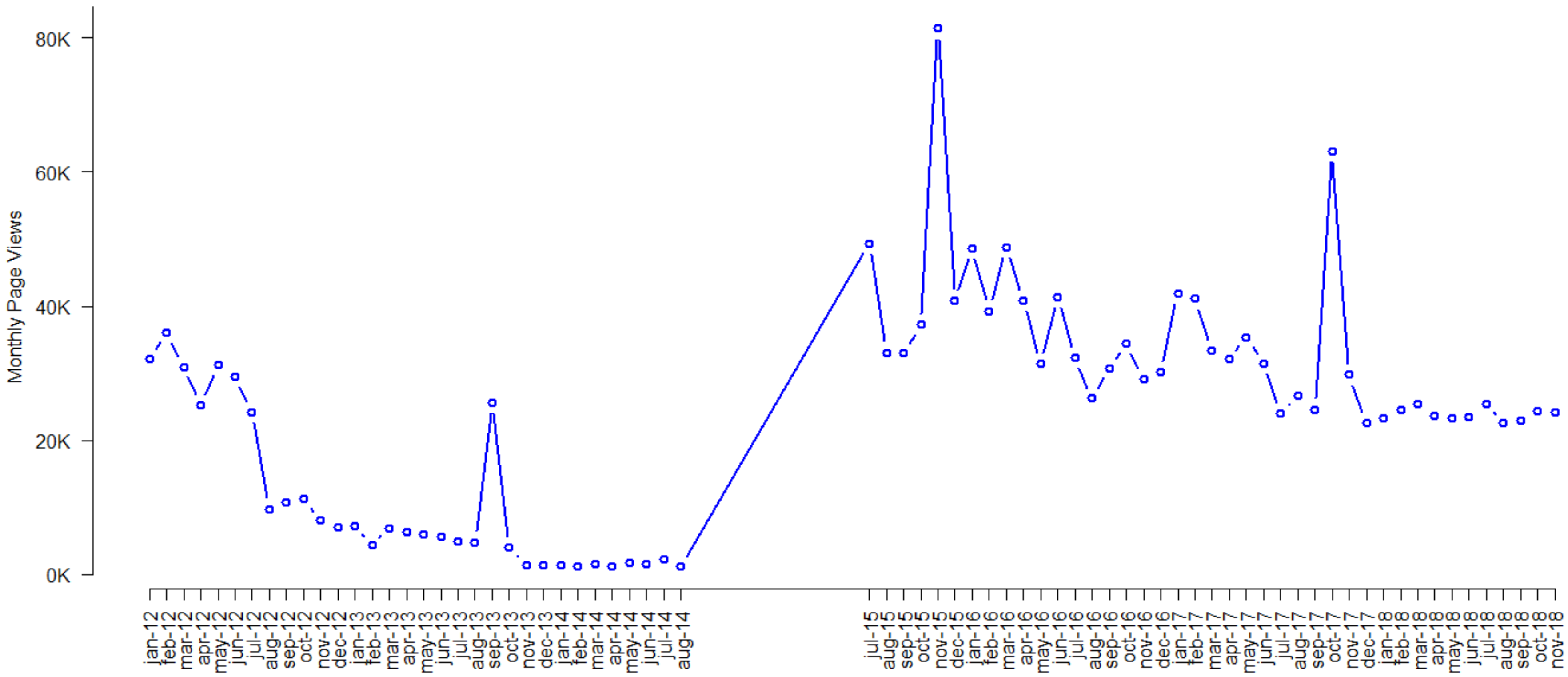




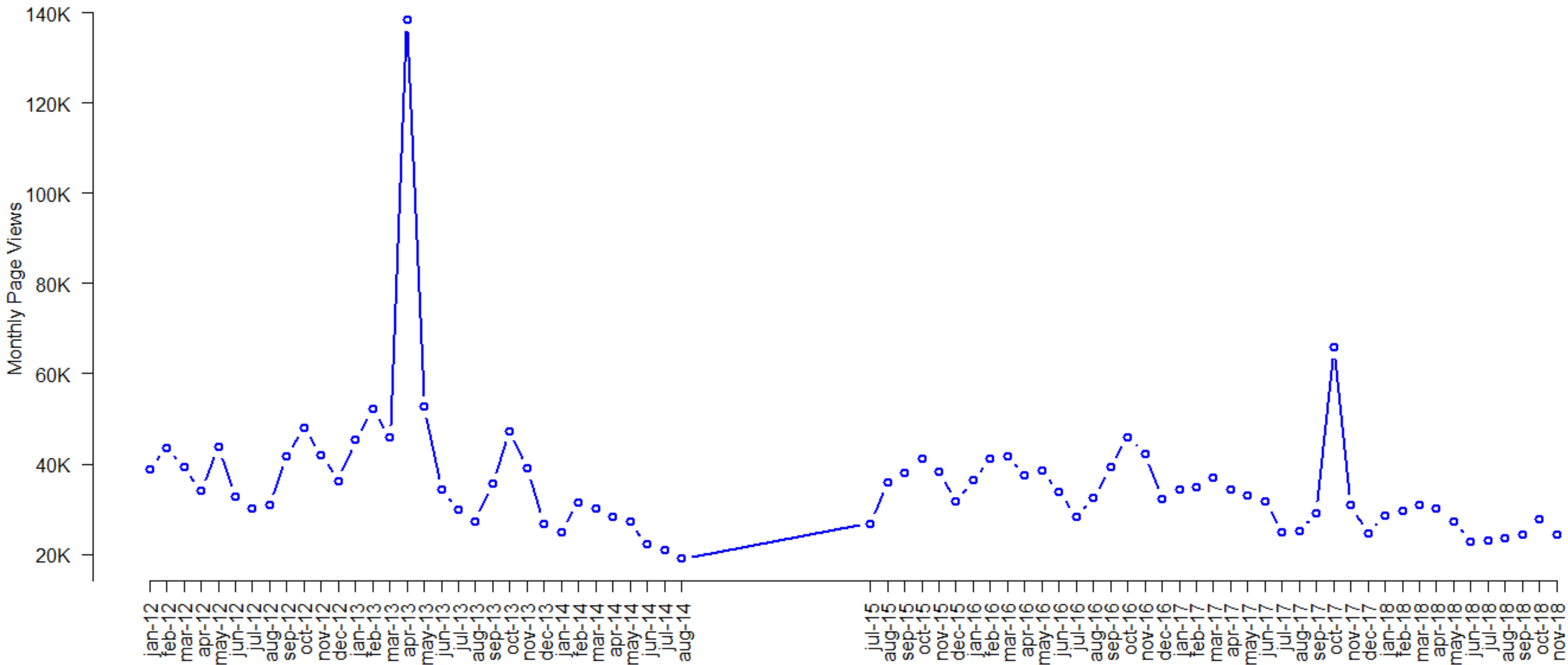
# Page Views for Al\_Qaeda\_in\_the\_Islamic\_Maghreb



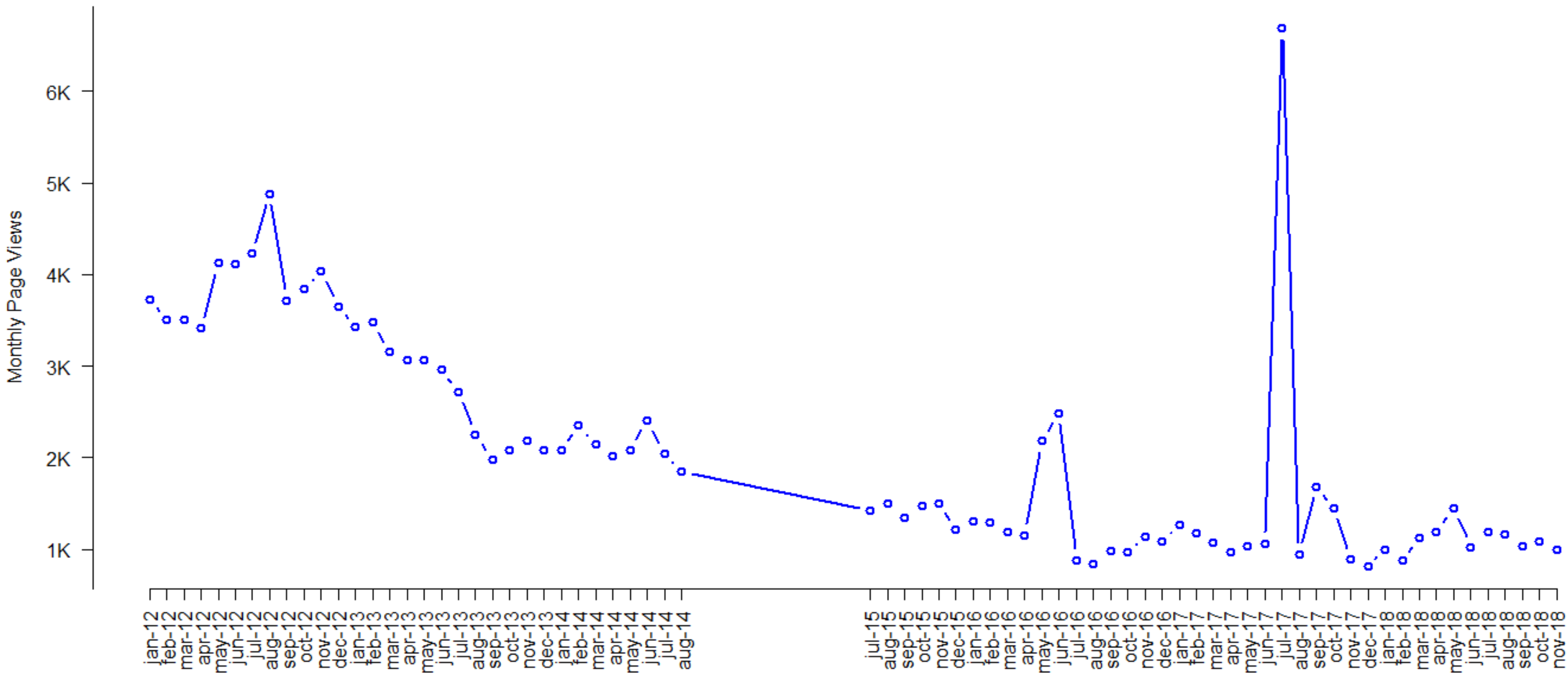
### Page Views for Al\_Shabaab



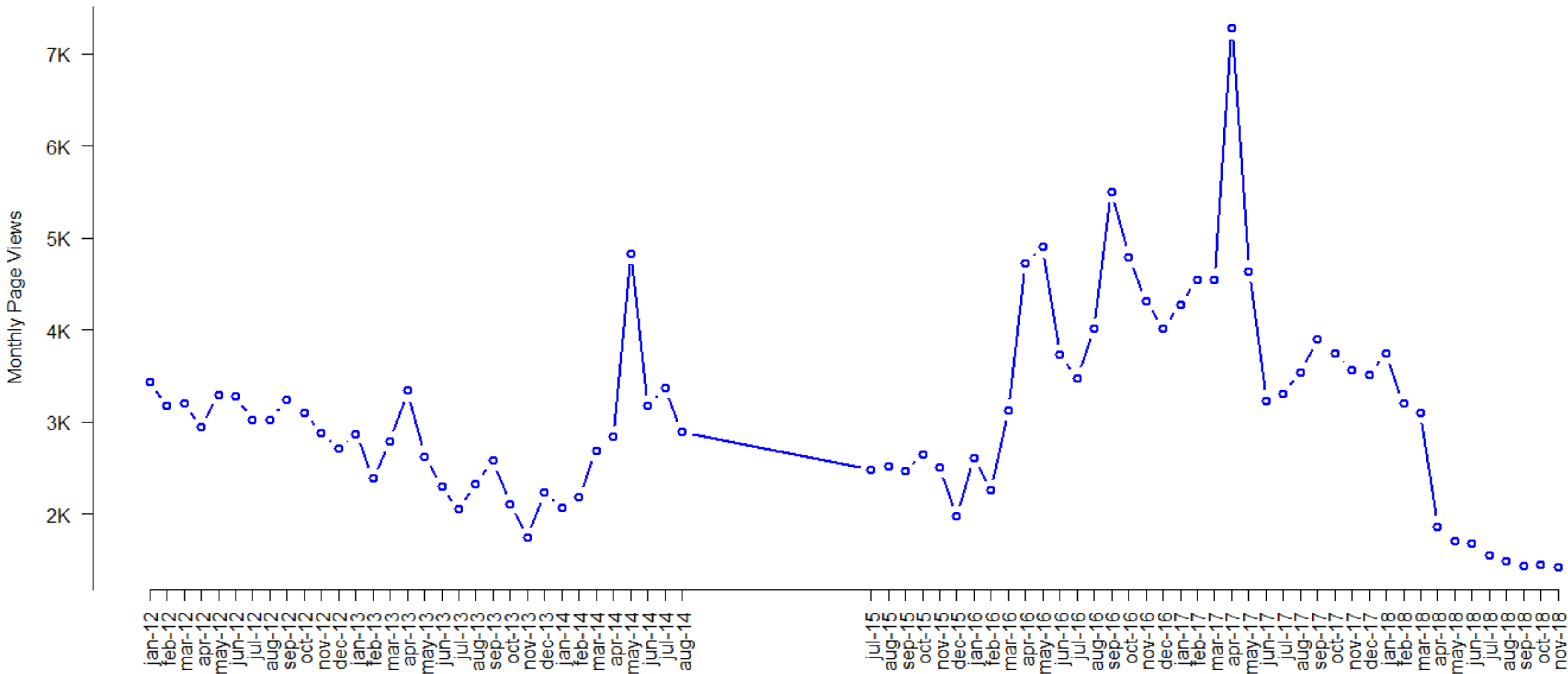
# Page Views for Ammonium\_nitrate



### Page Views for attack

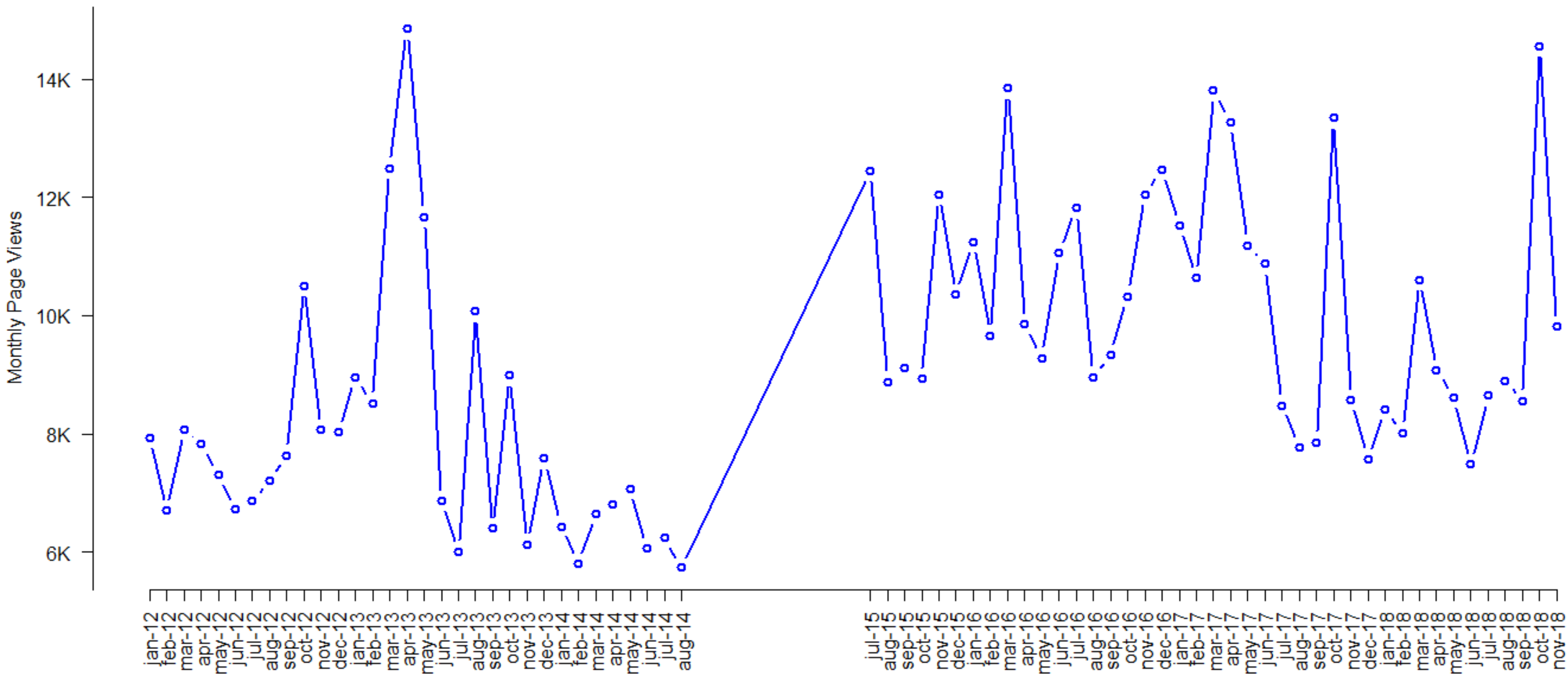


### Page Views for Biological\_weapon

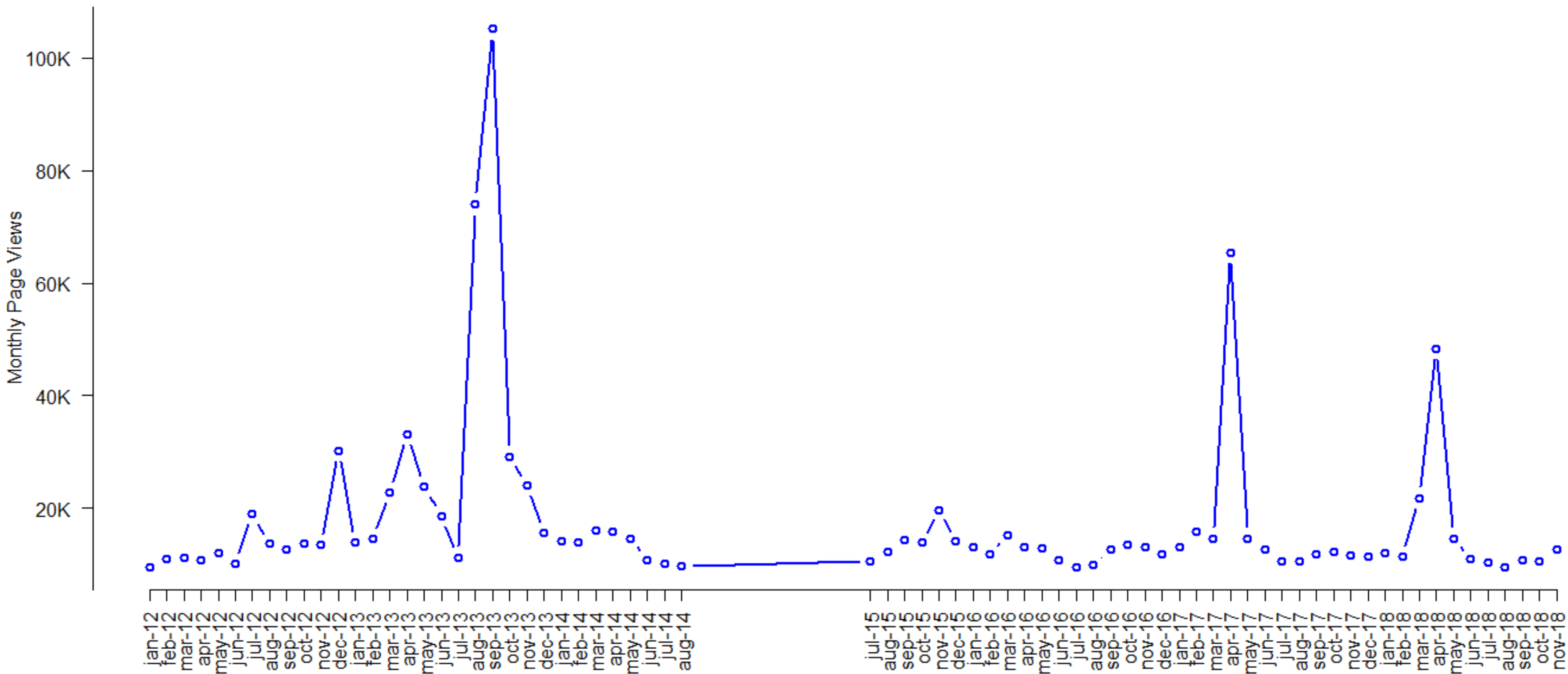




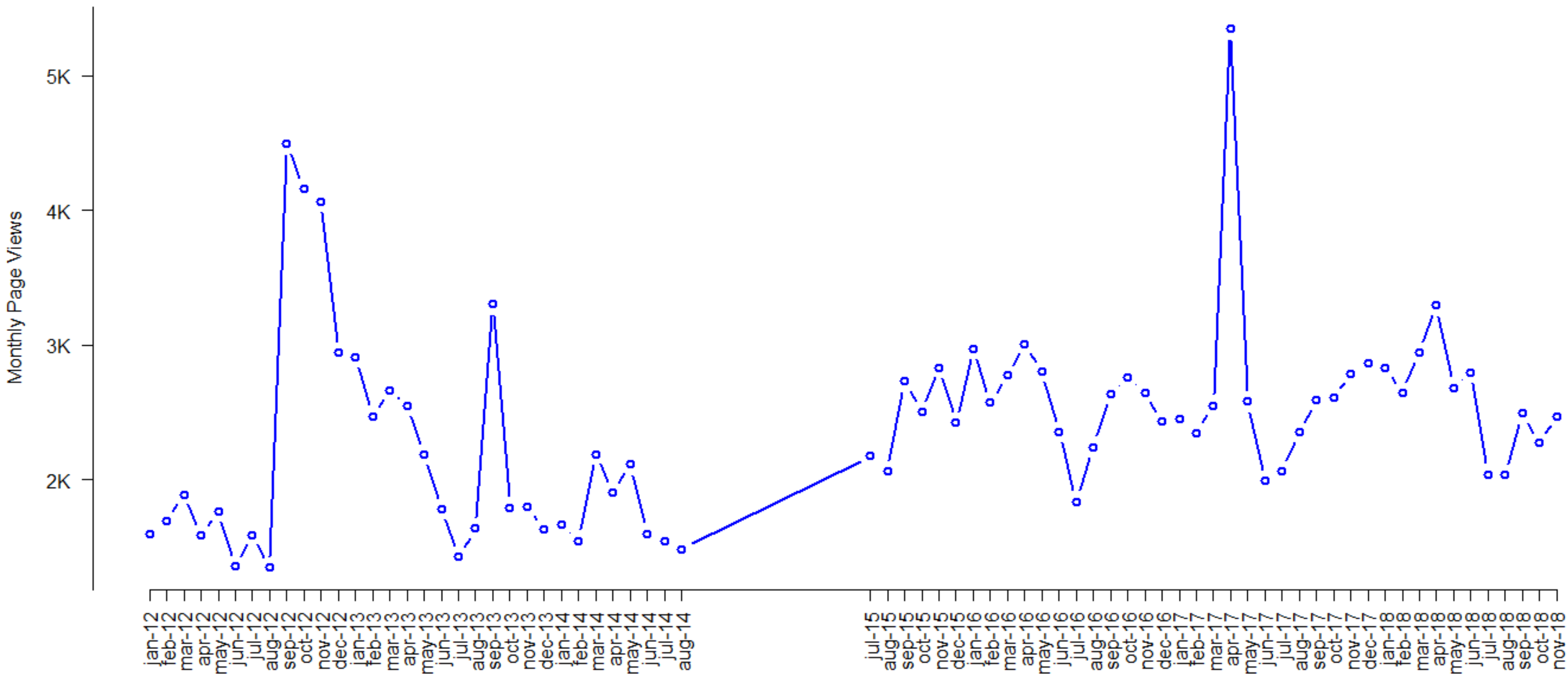
### Page Views for Car\_bomb



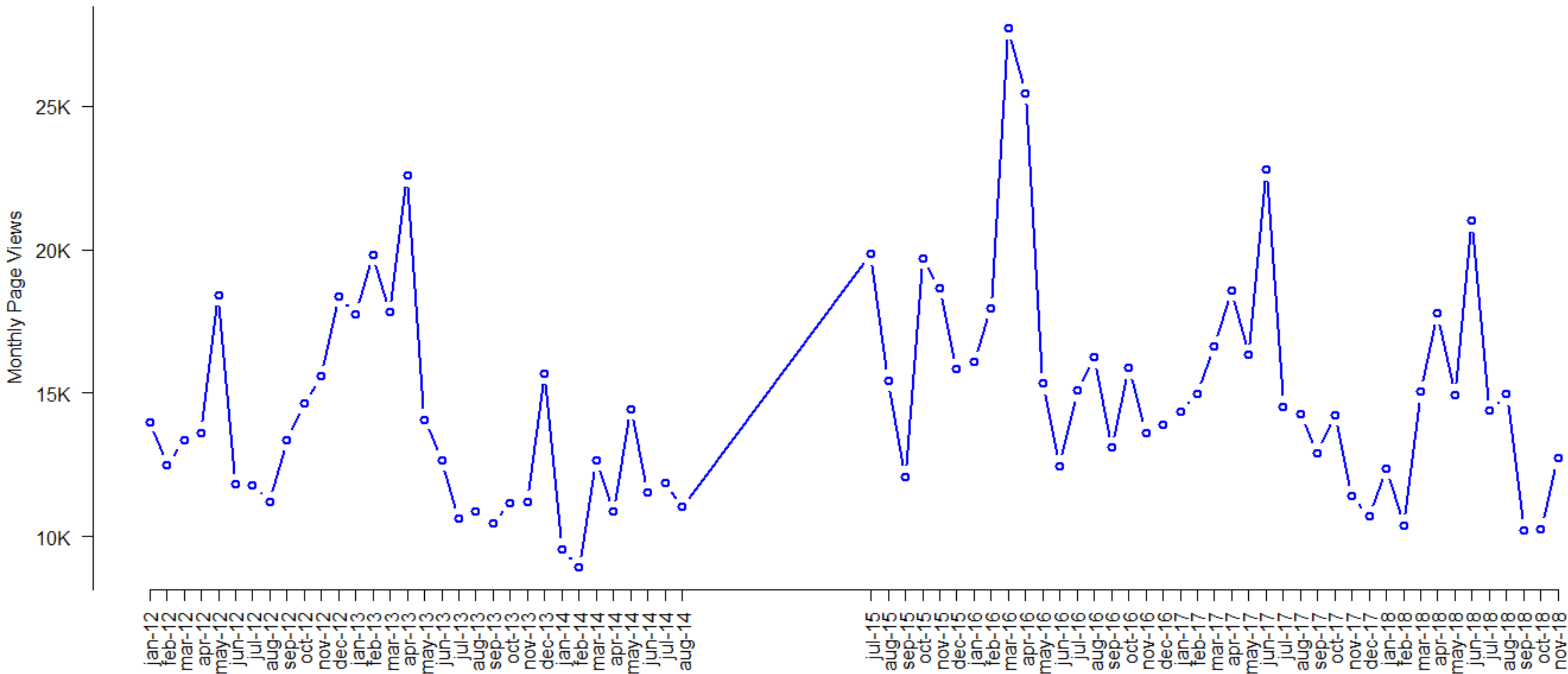
### Page Views for Chemical\_weapon



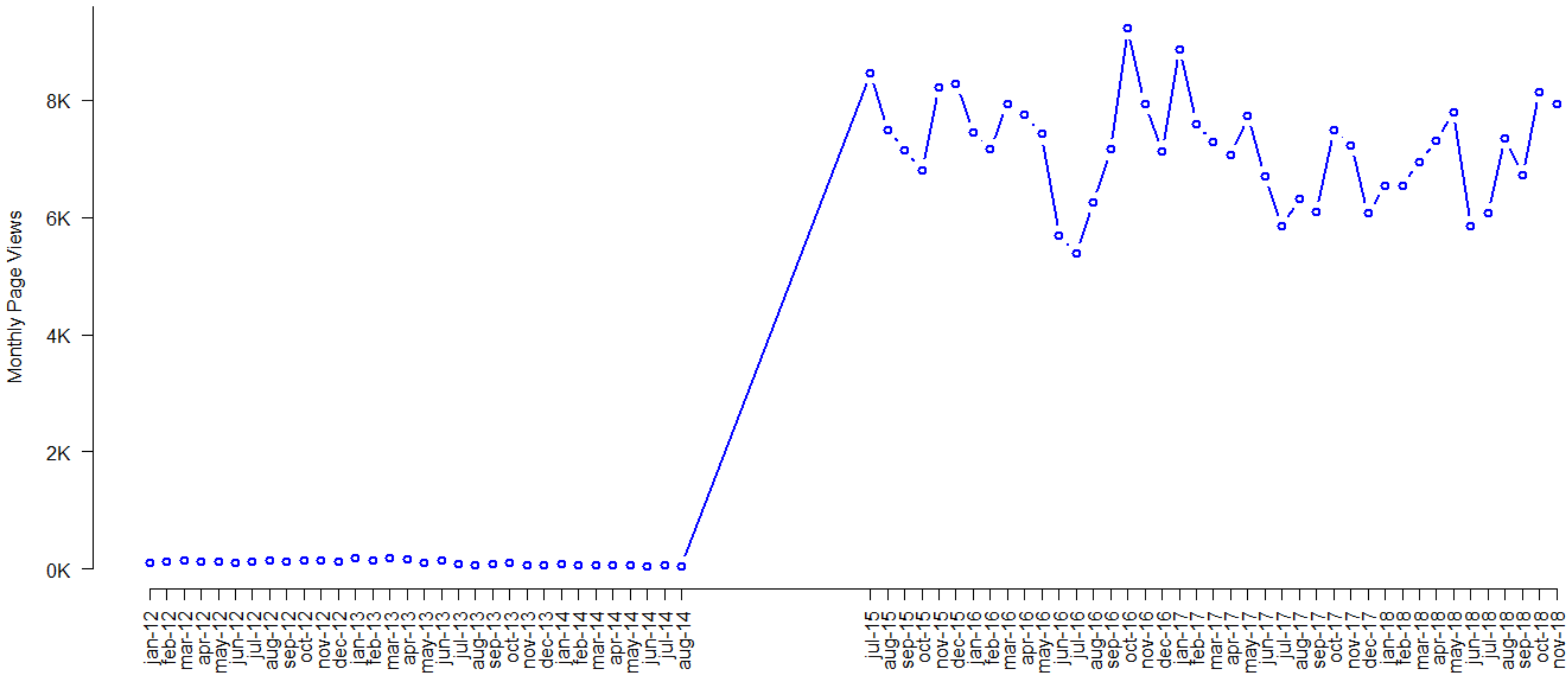
### Page Views for Conventional\_weapon



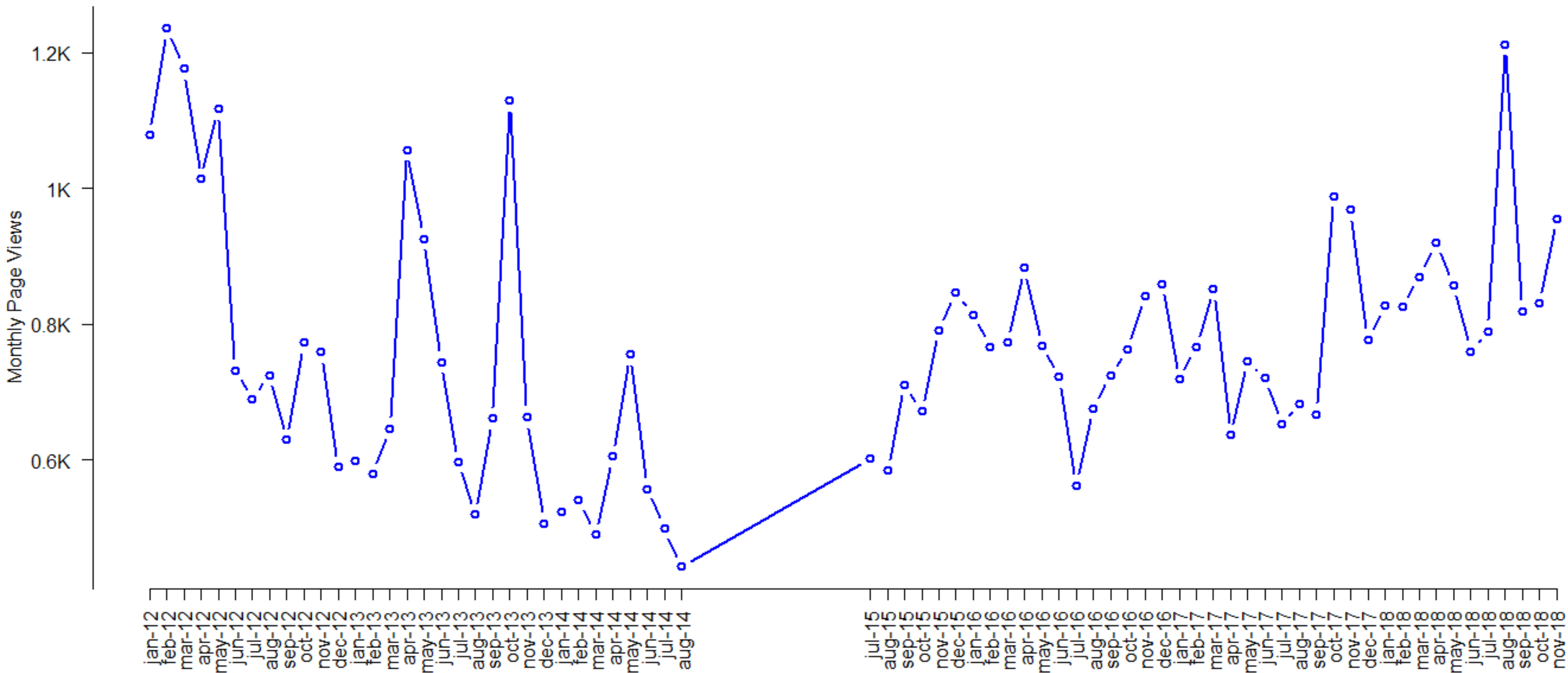
### Page Views for dirty\_bomb



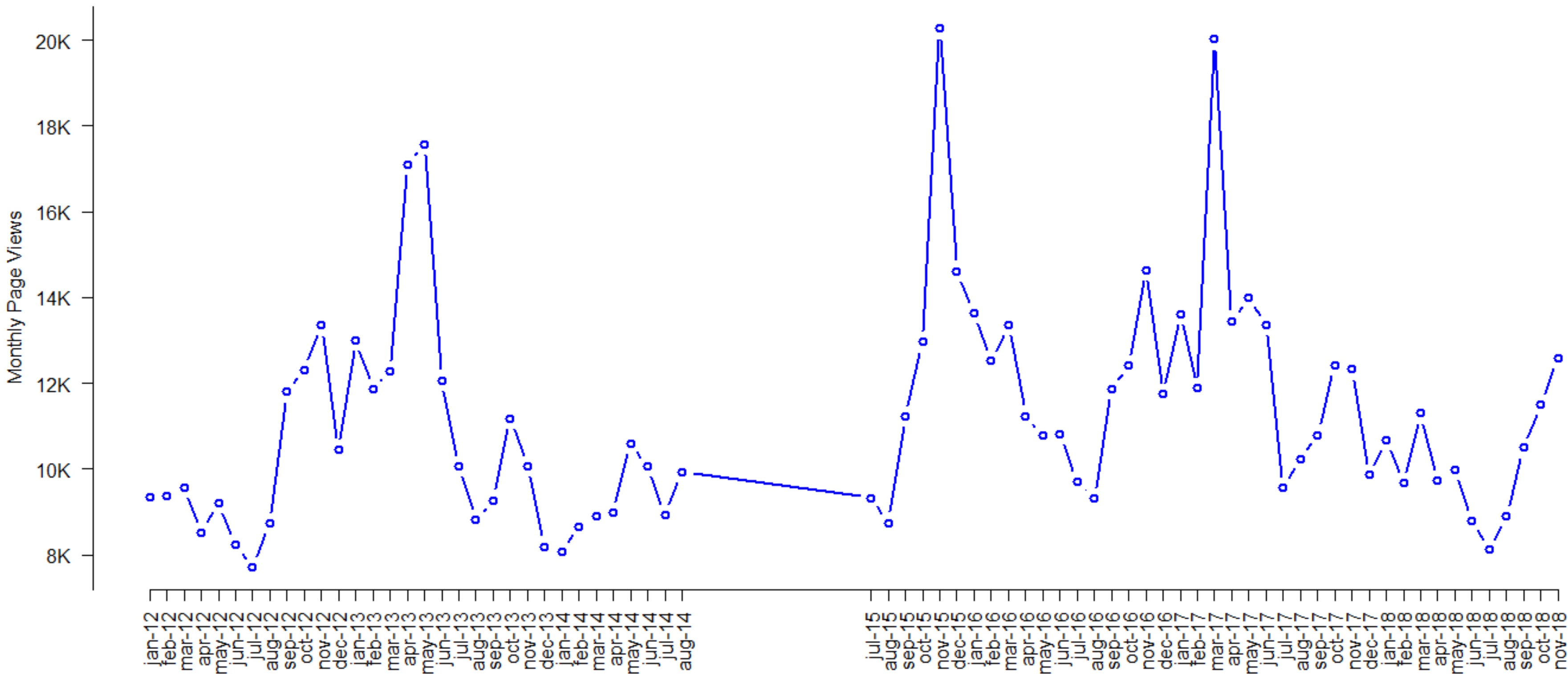
### Page Views for Eco\_terrorism



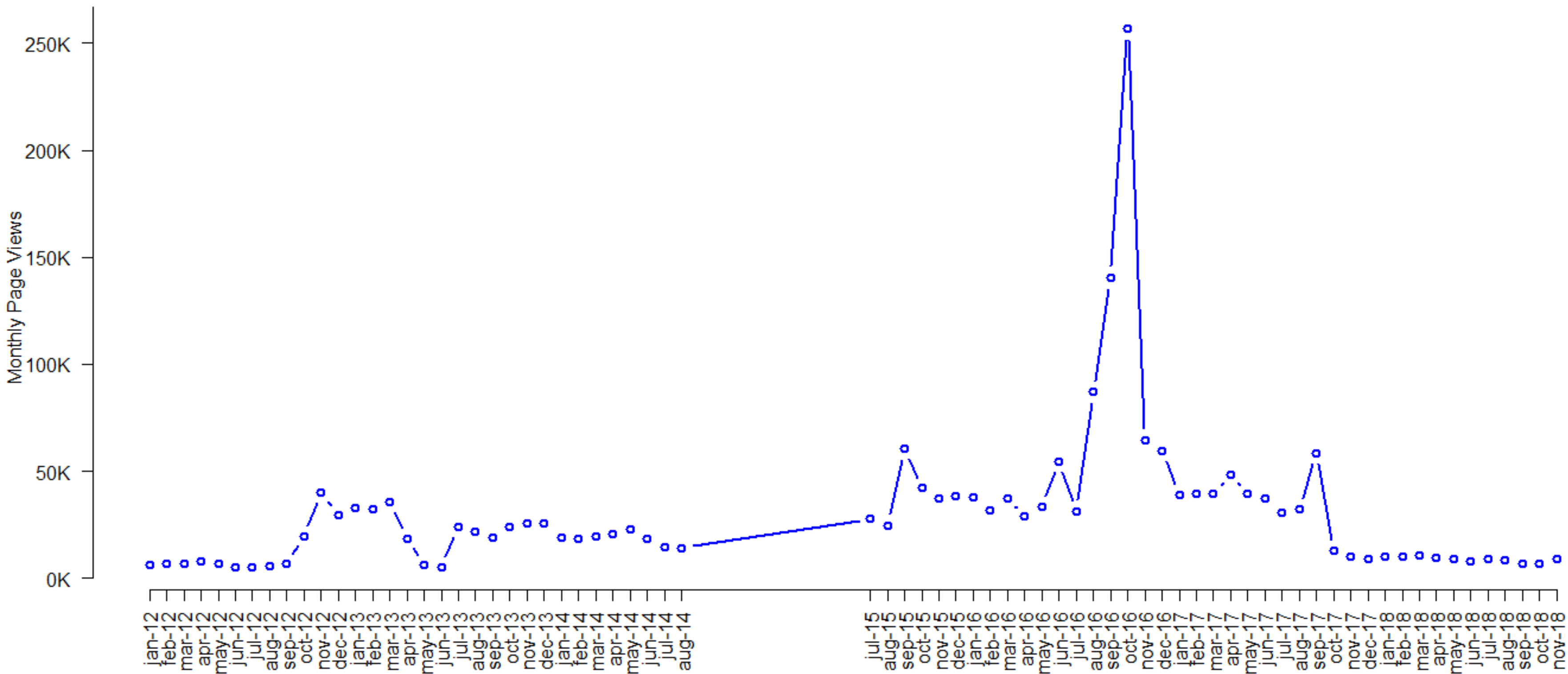
### Page Views for Environmental\_terrorist\_NA\_ism



### Page Views for Extremism

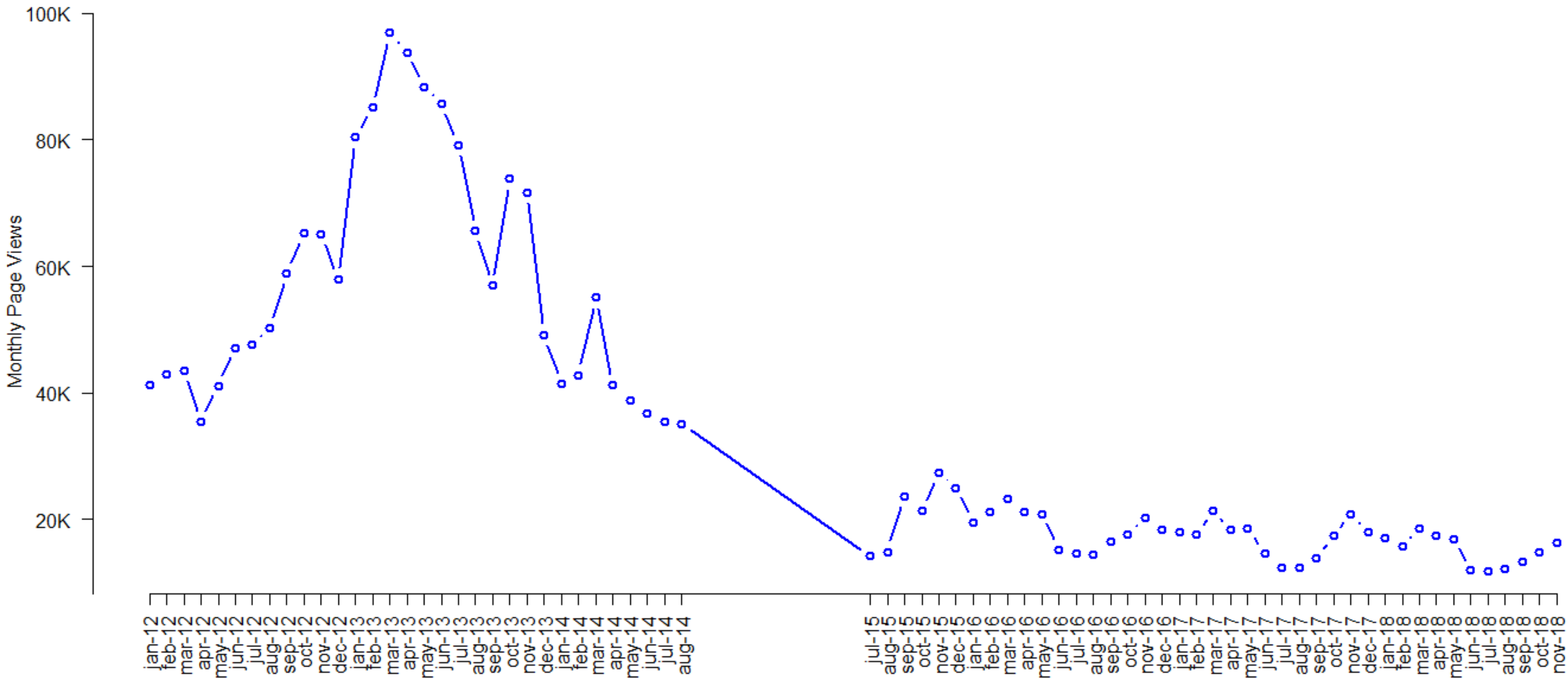


# Page Views for FARC

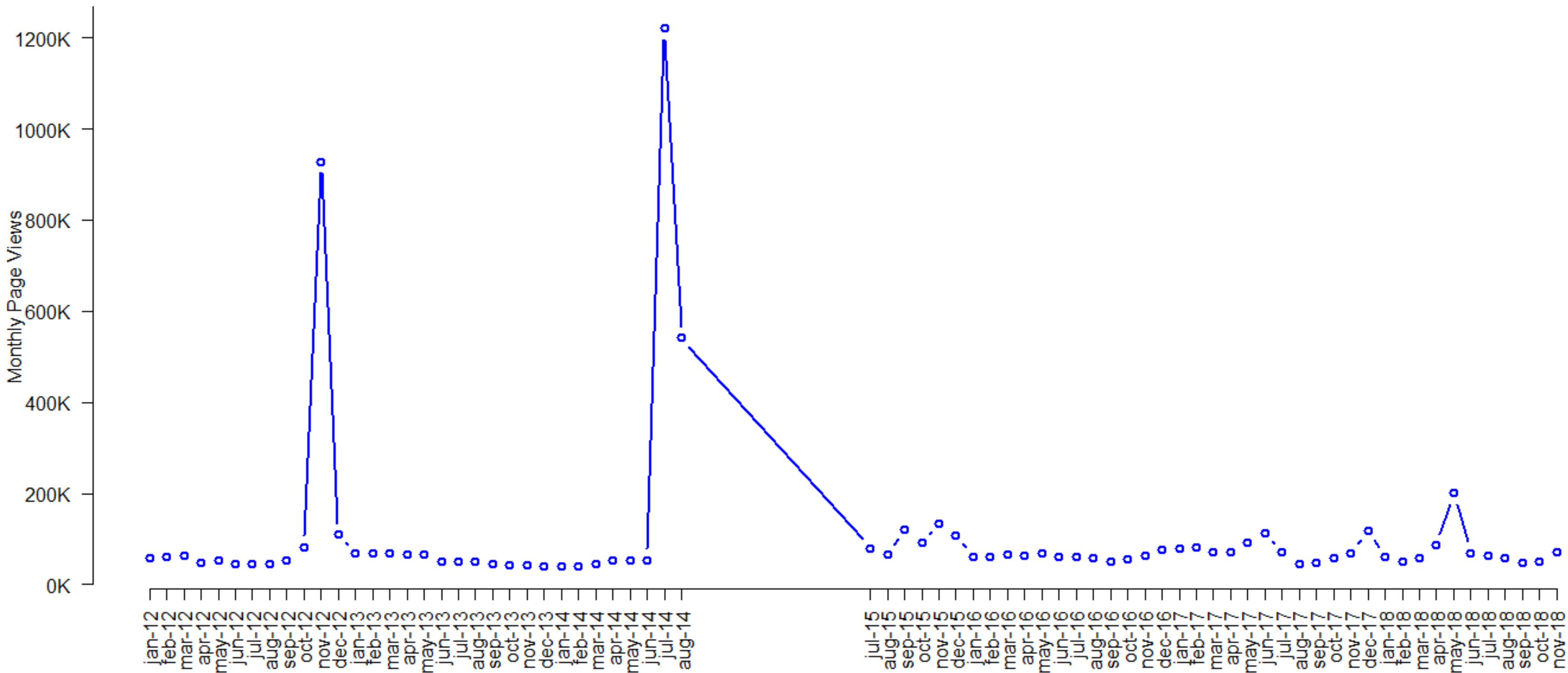




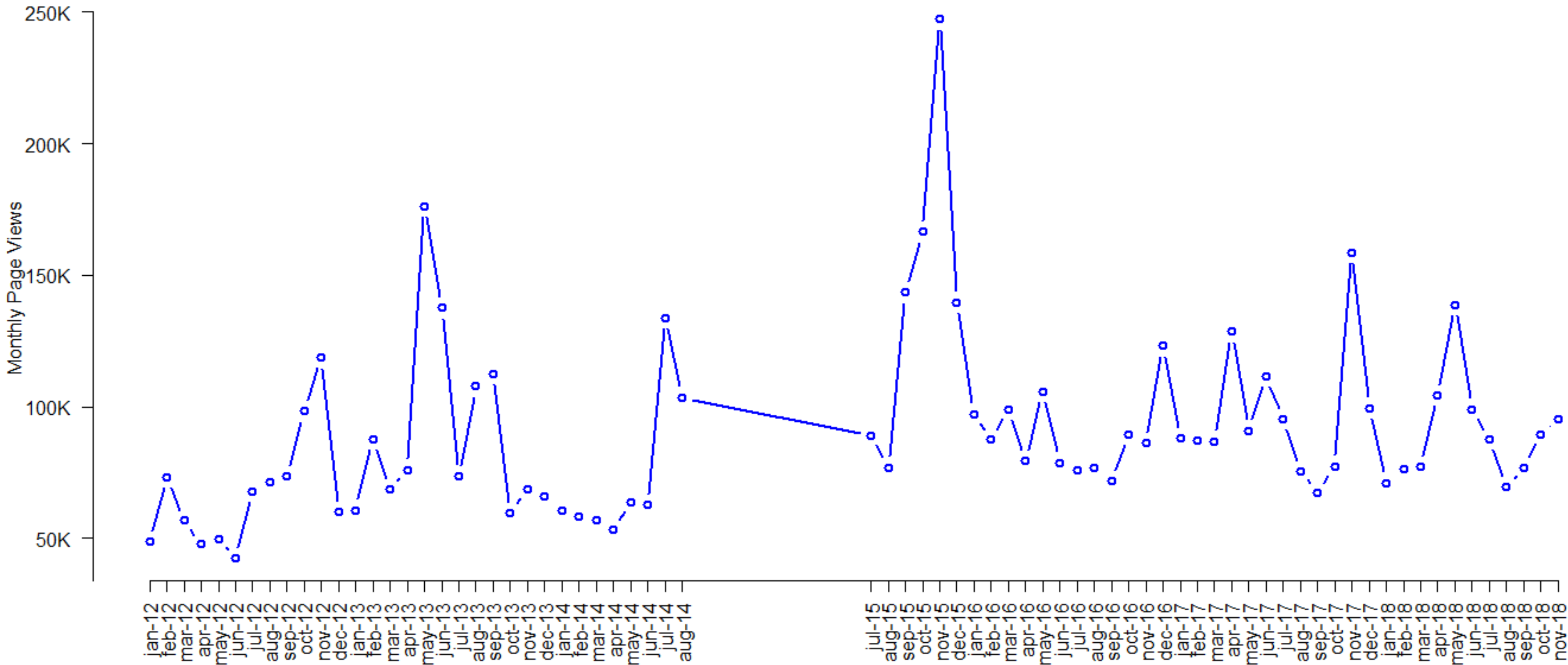
### Page Views for Fundamentalism



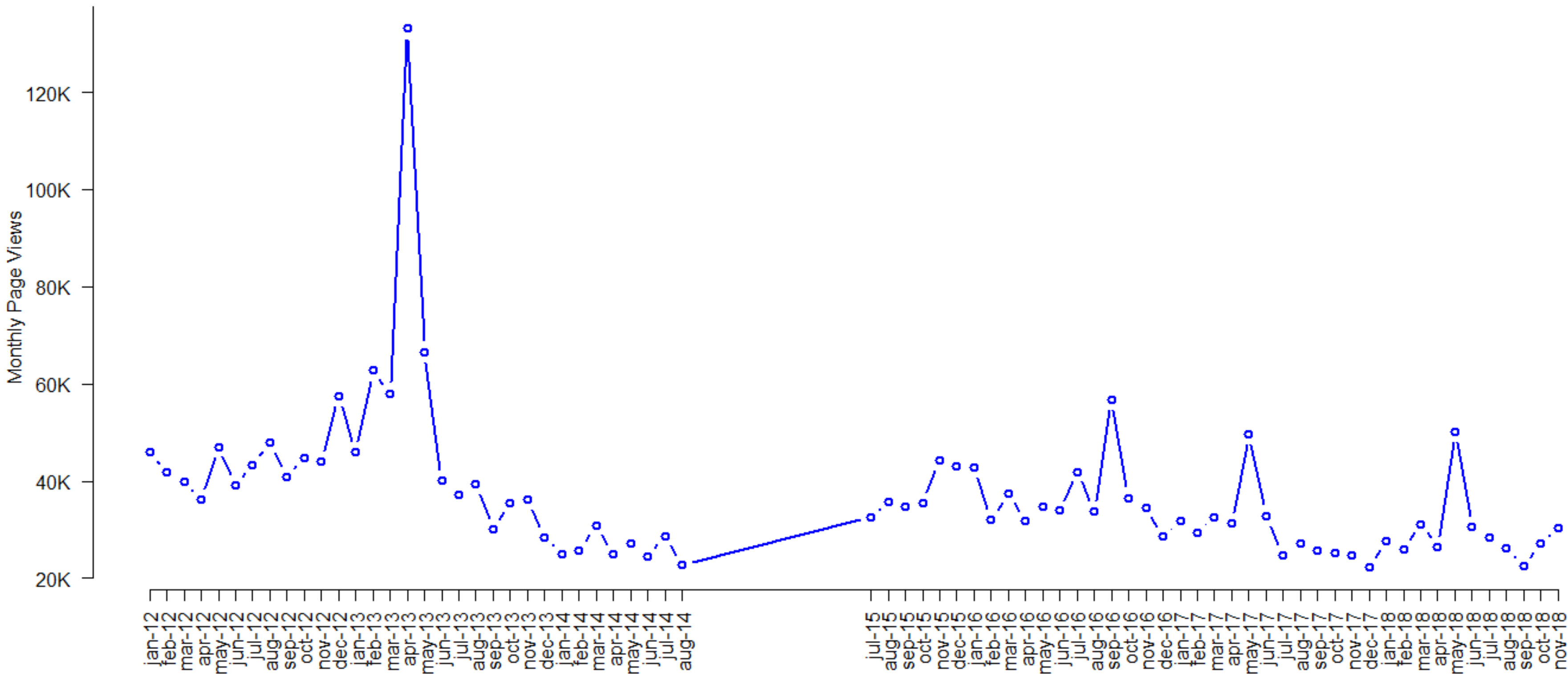
### Page Views for Hamas



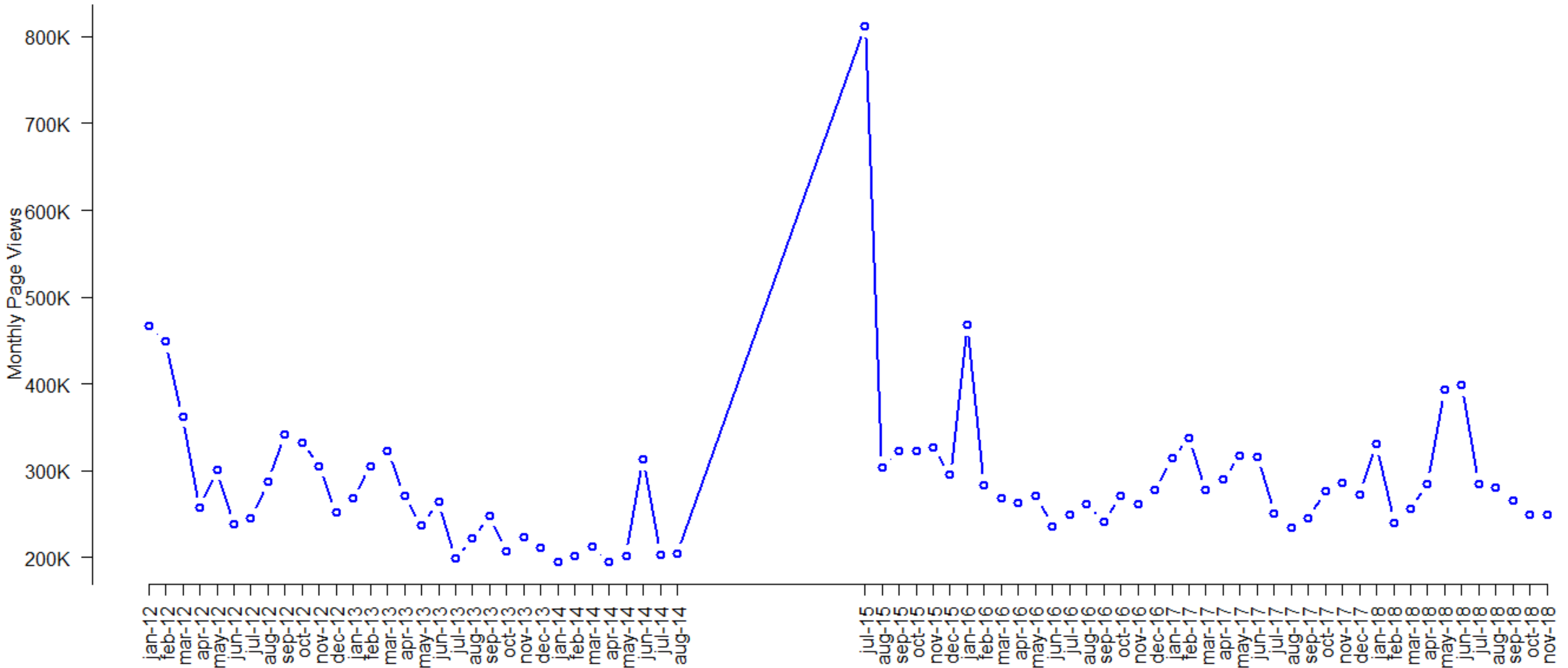
### Page Views for Hezbollah



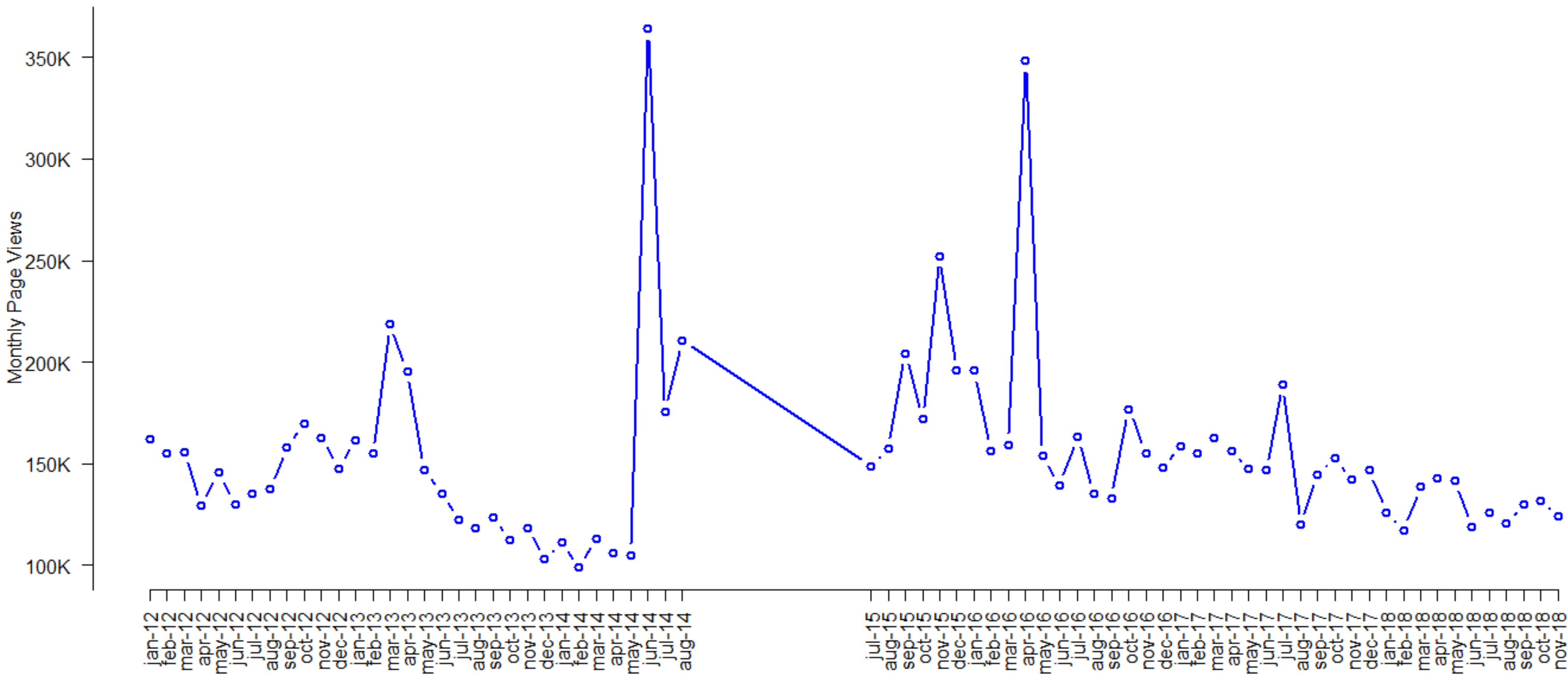
### Page Views for Improvised\_explosive\_device



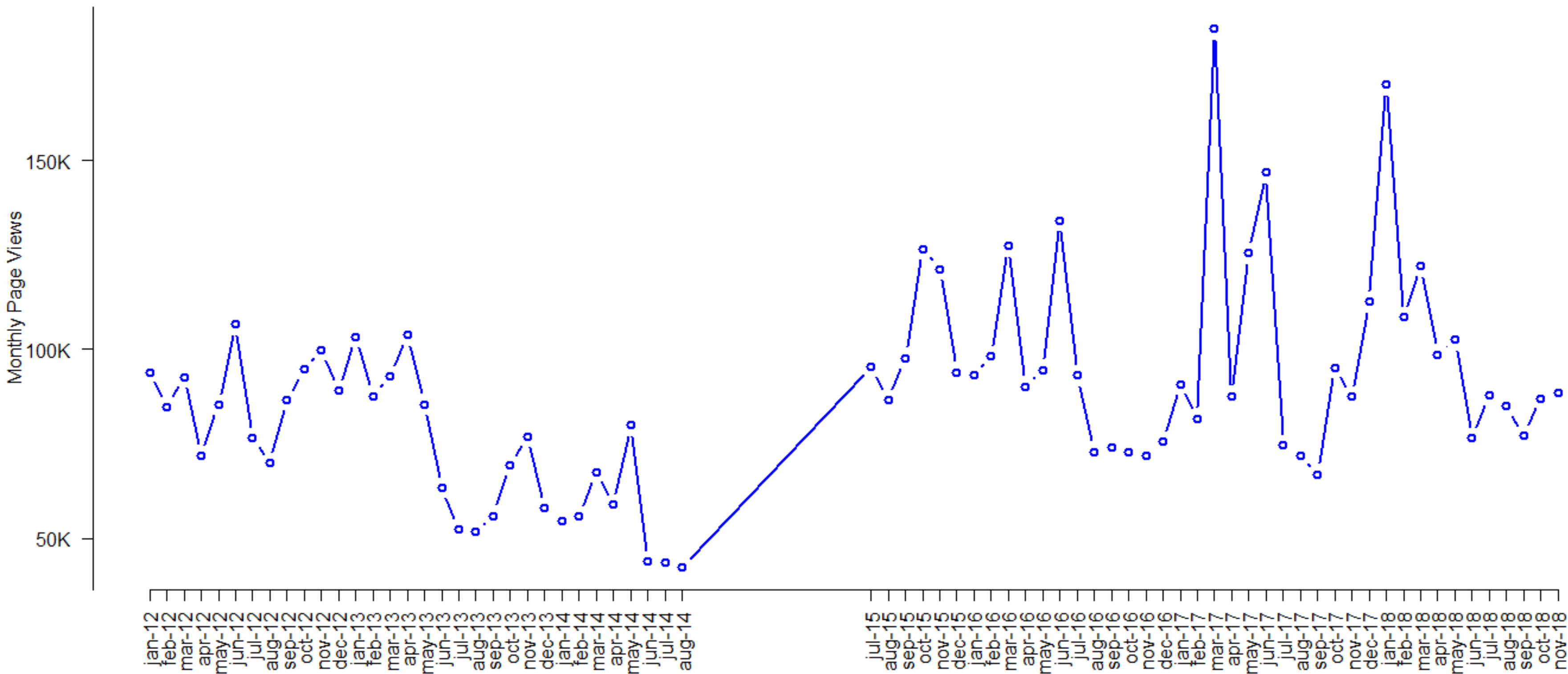
### Page Views for Iran



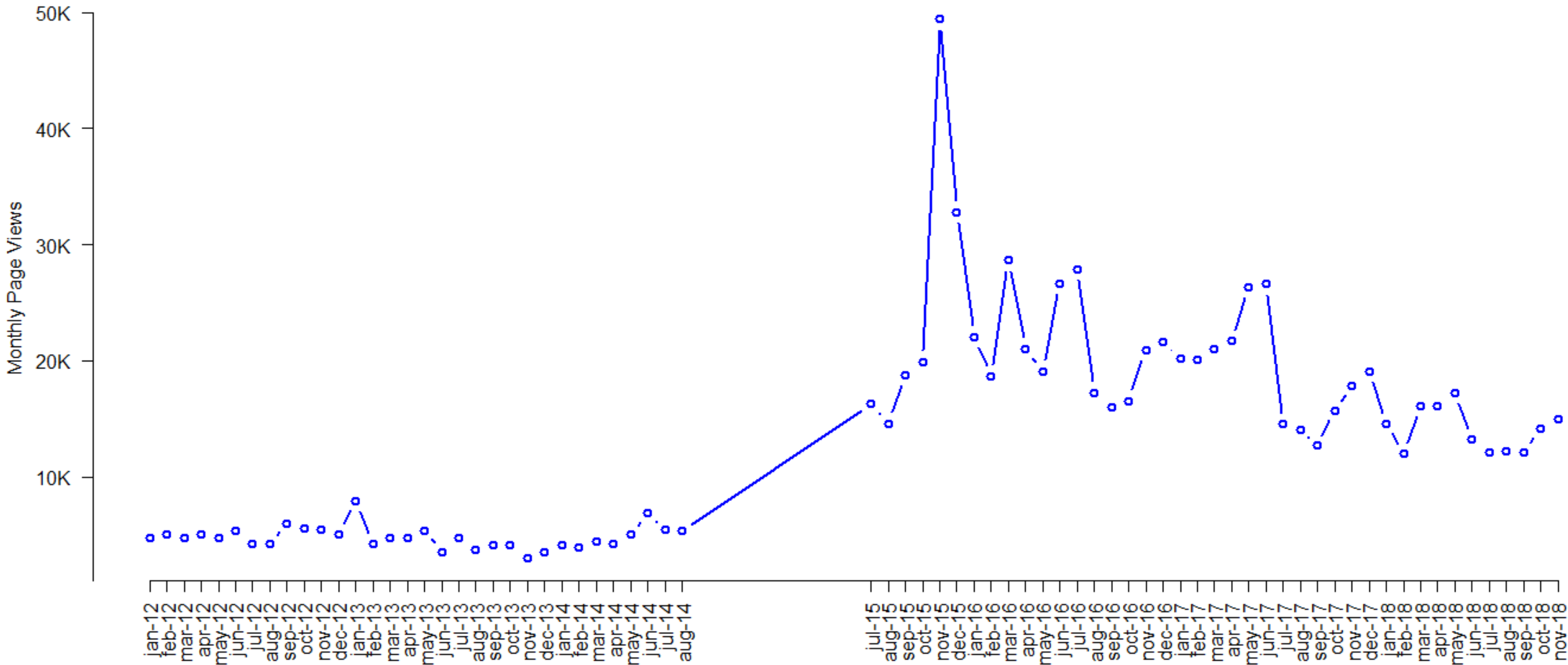
### Page Views for Iraq



### Page Views for Irish\_Republican\_Army

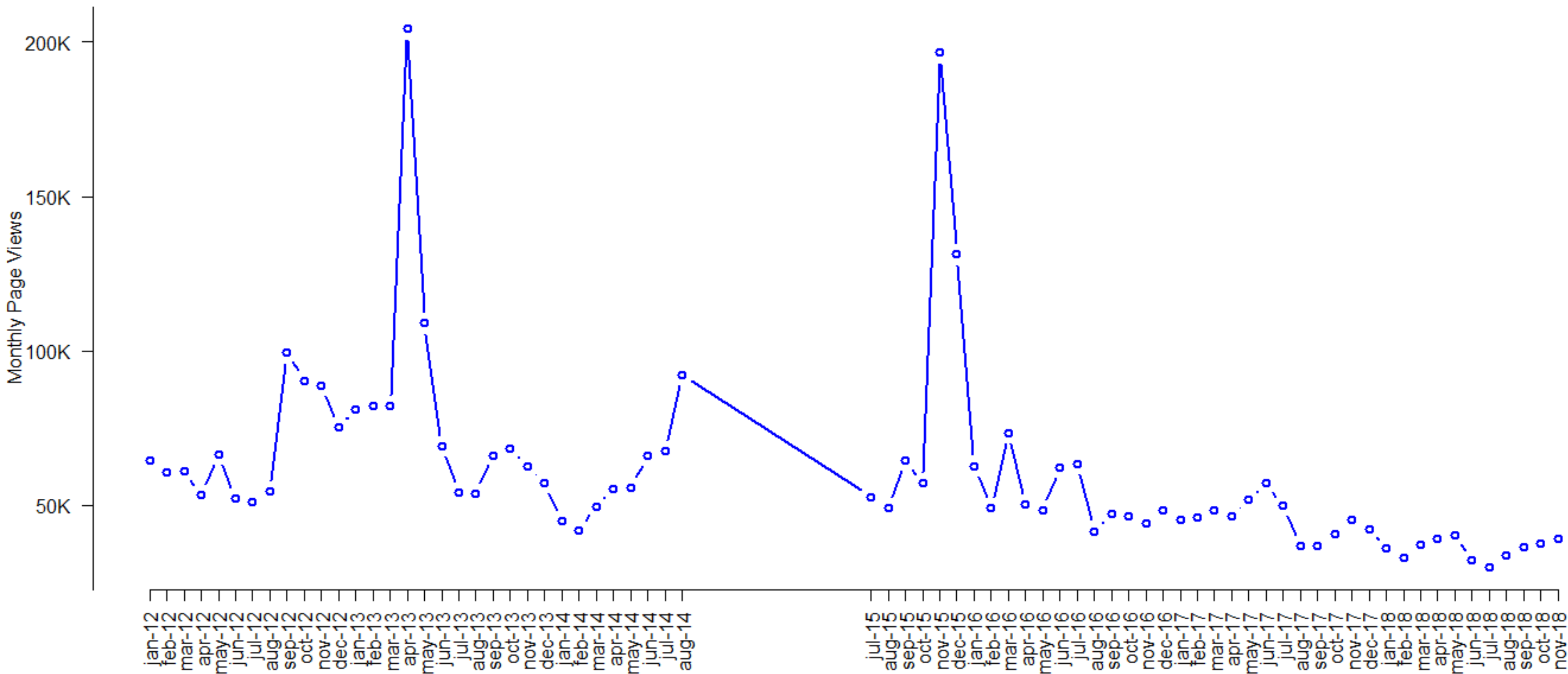


### Page Views for Islamist

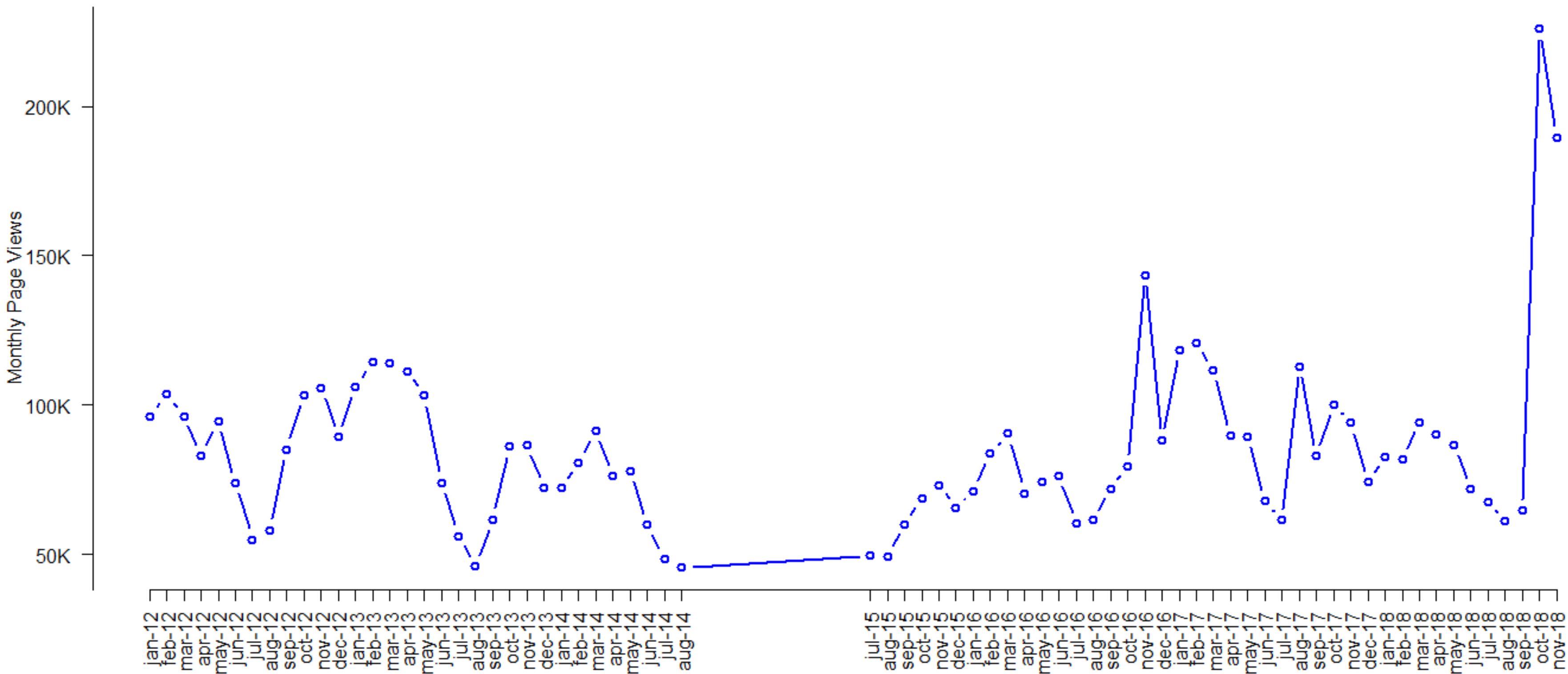




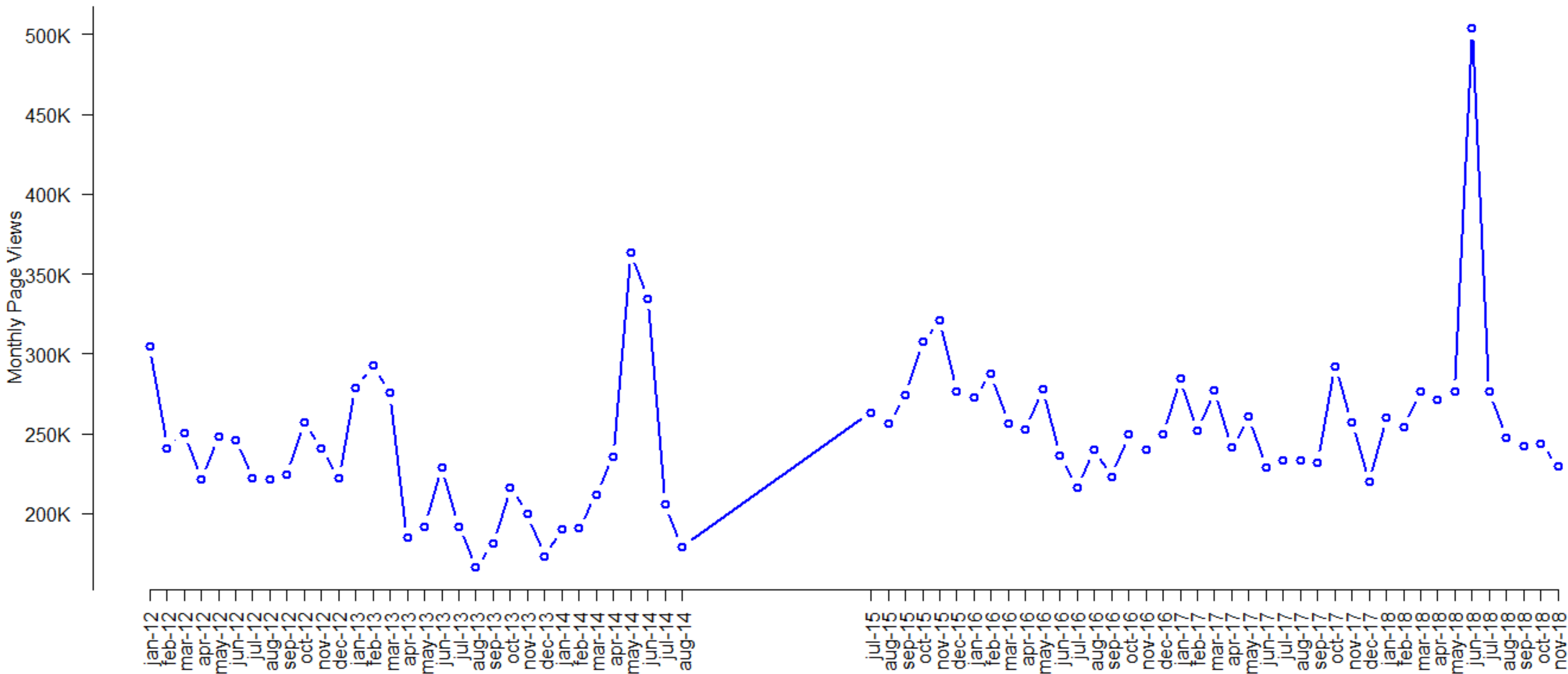
### Page Views for Jihad



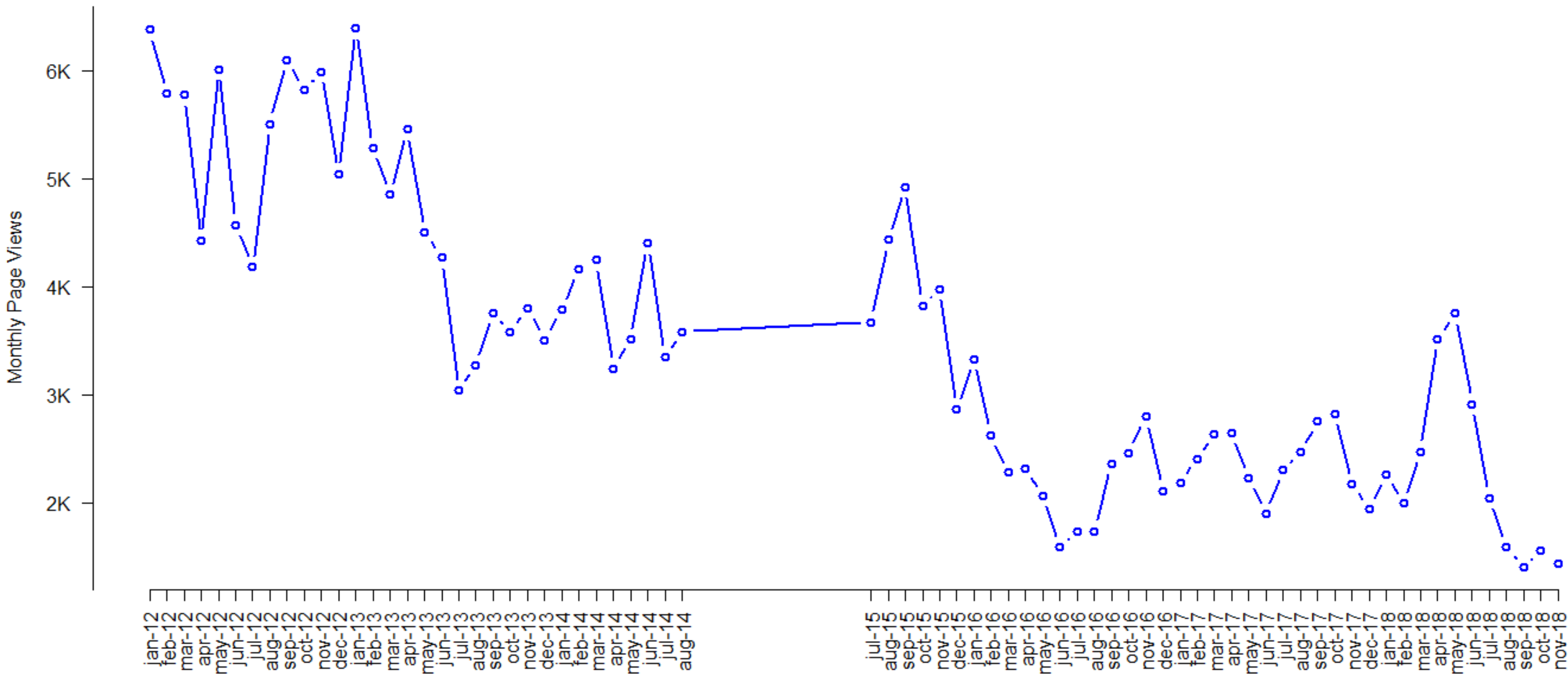
### Page Views for nationalism



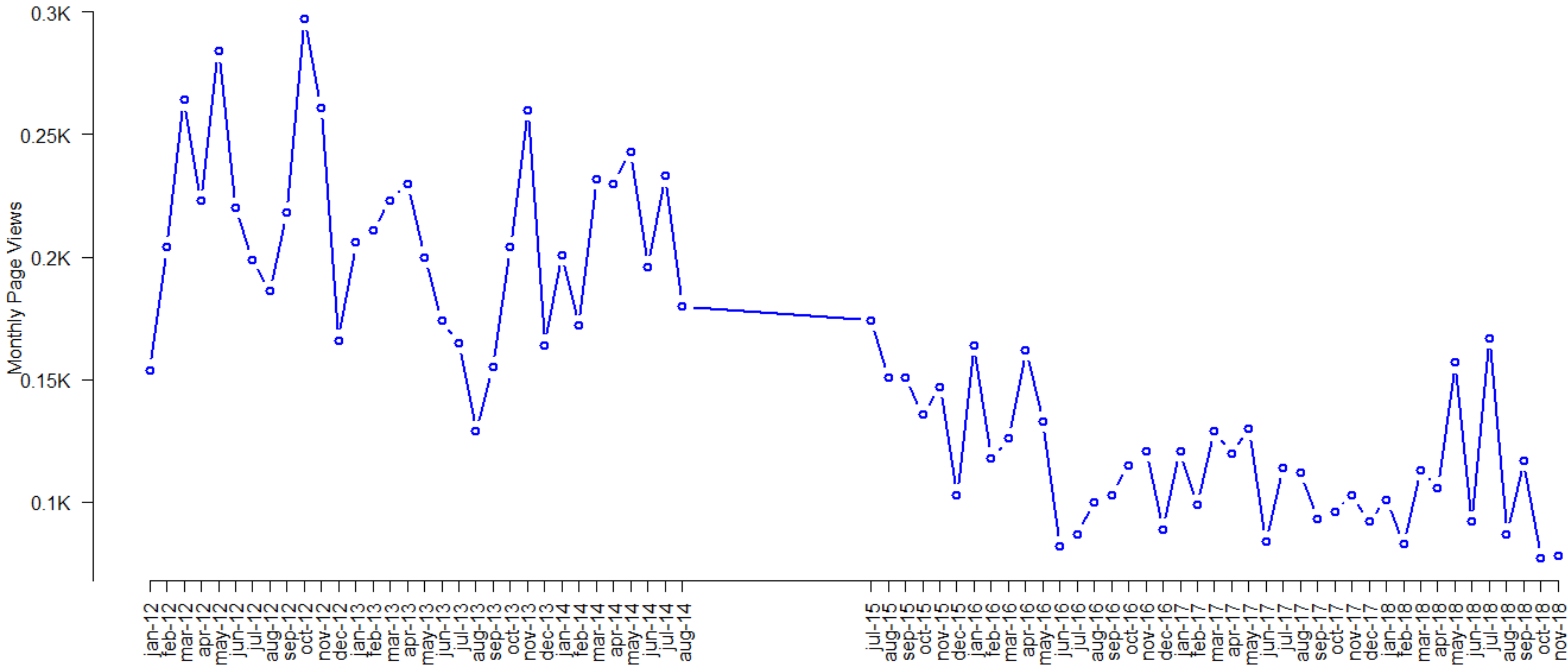
### Page Views for Nigeria



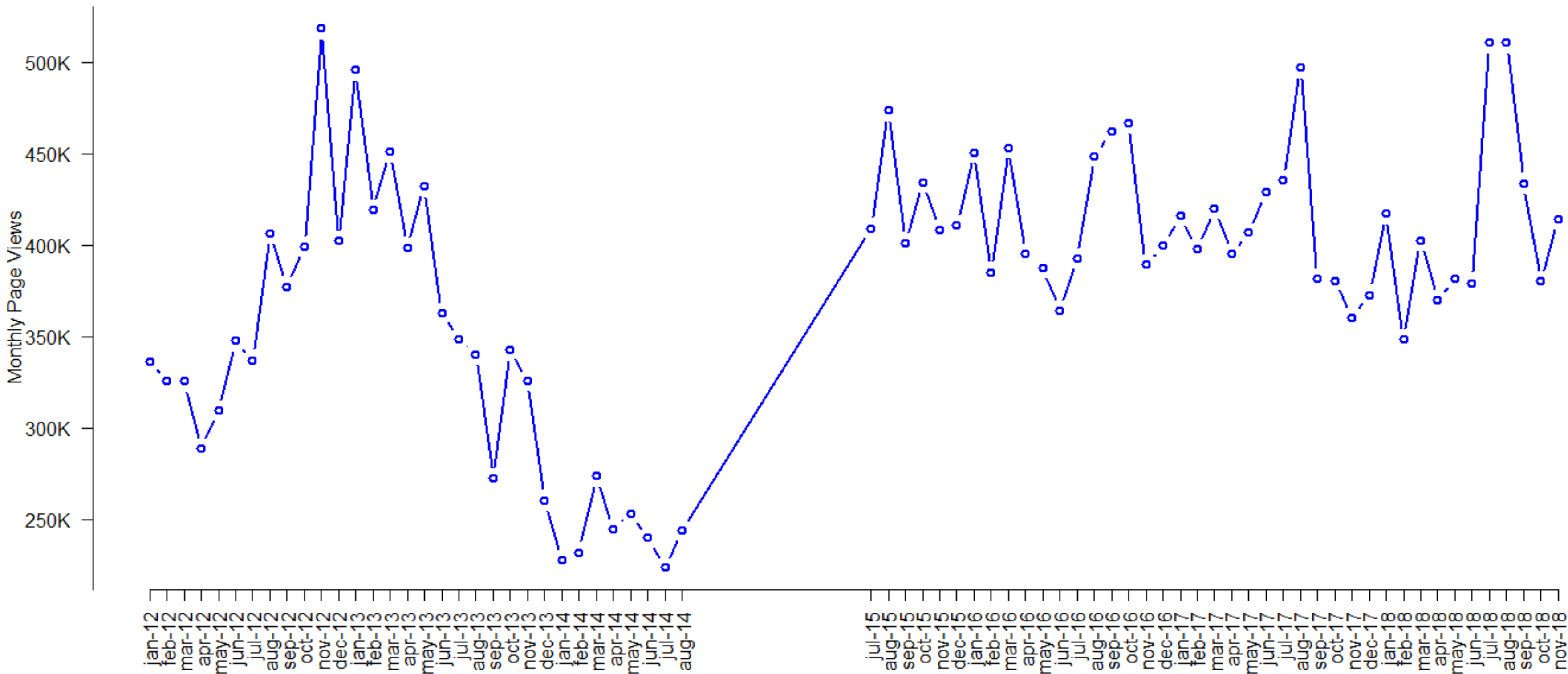
### Page Views for Nuclear



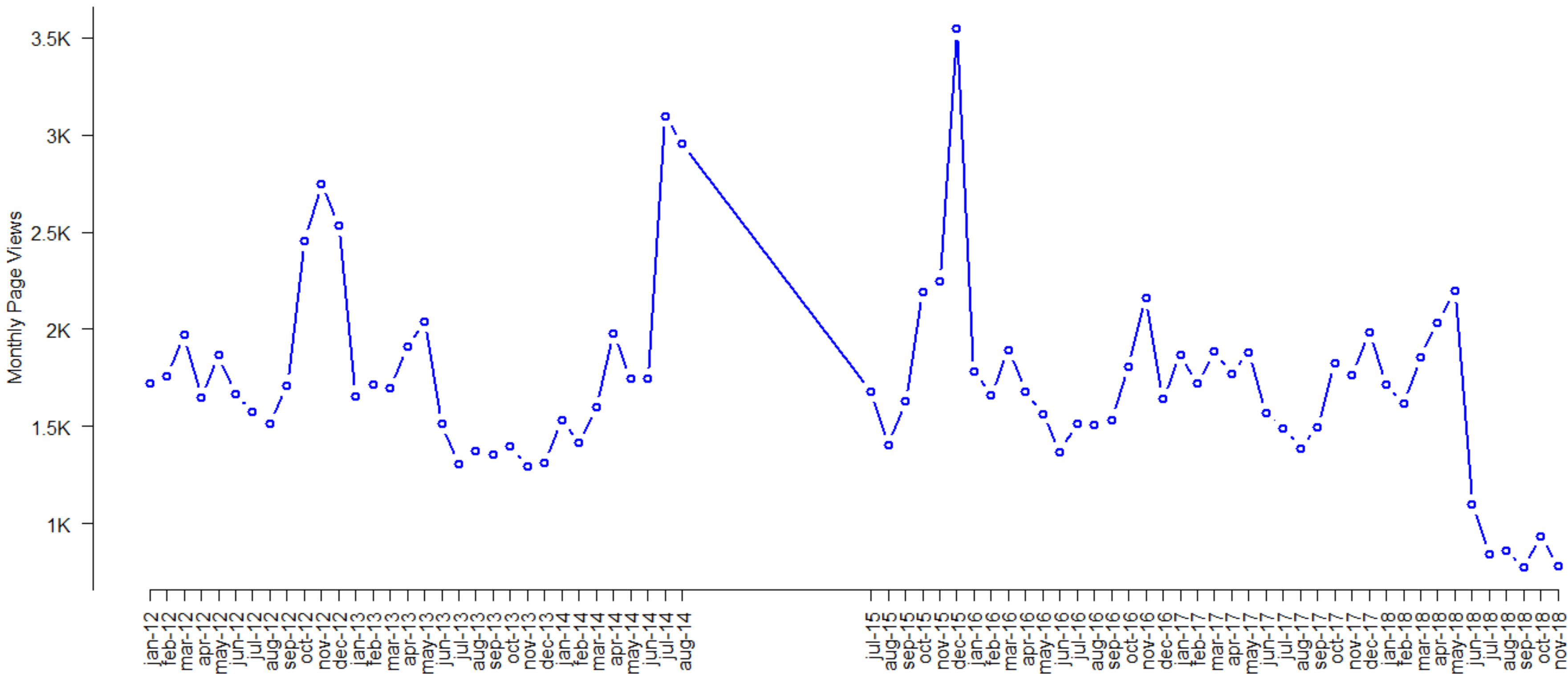
### Page Views for Nuclear\_Enrichment



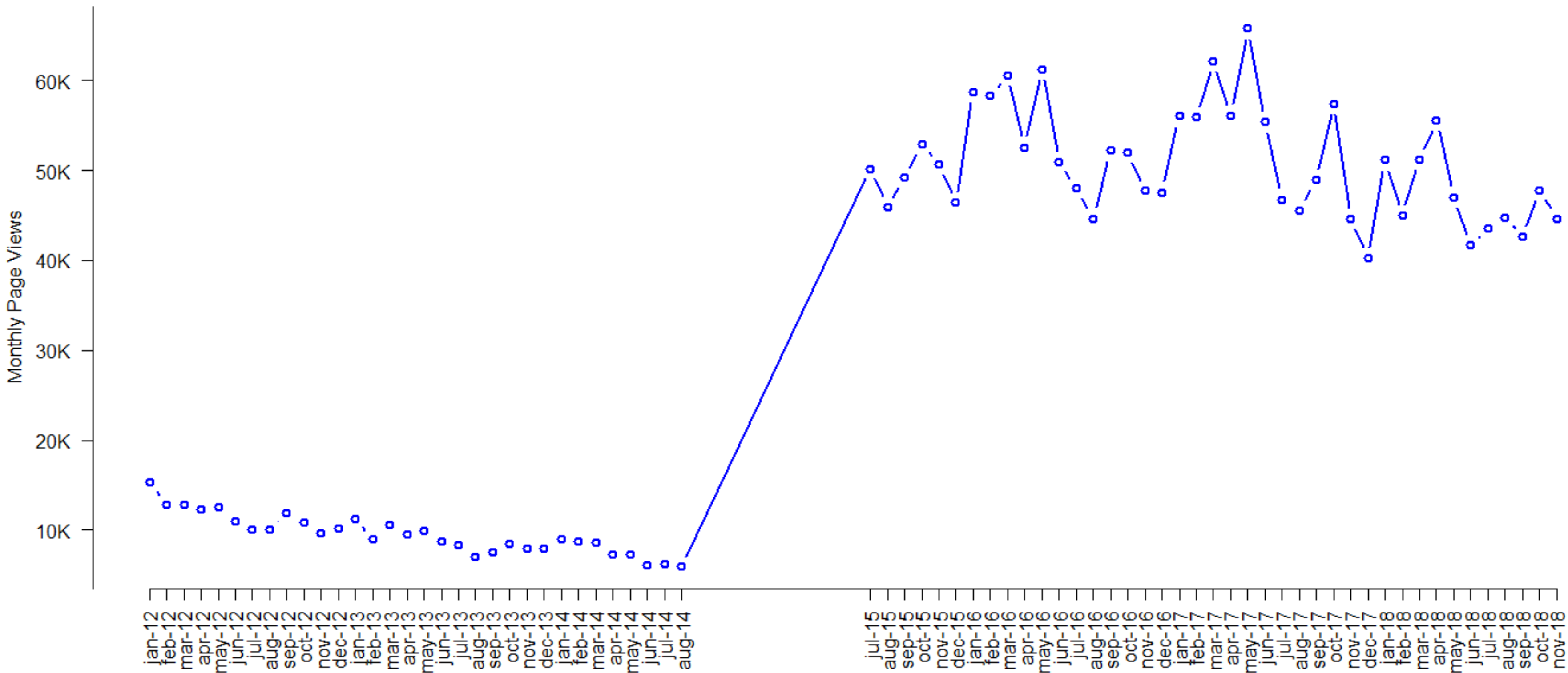
### Page Views for Pakistan



### Page Views for Palestine\_Liberation\_Fron

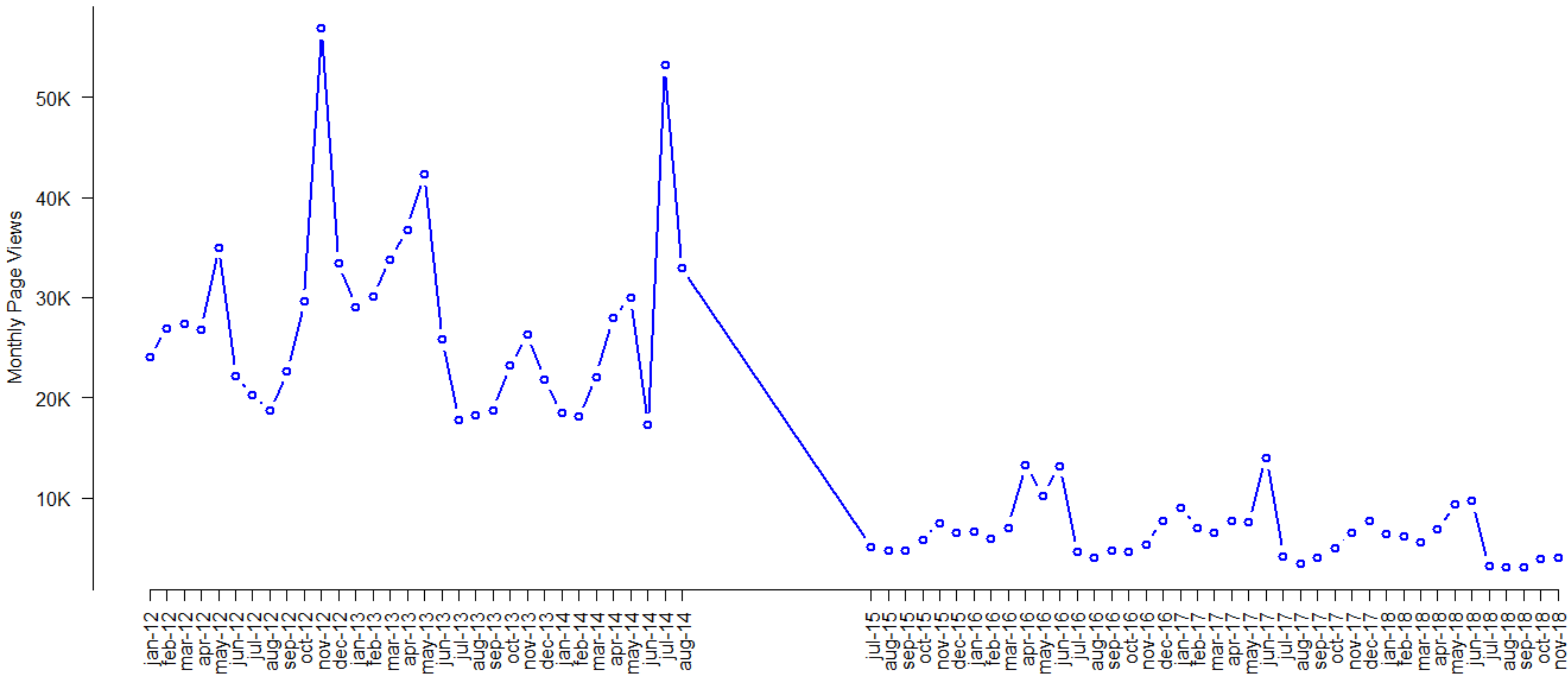


### Page Views for Pirates

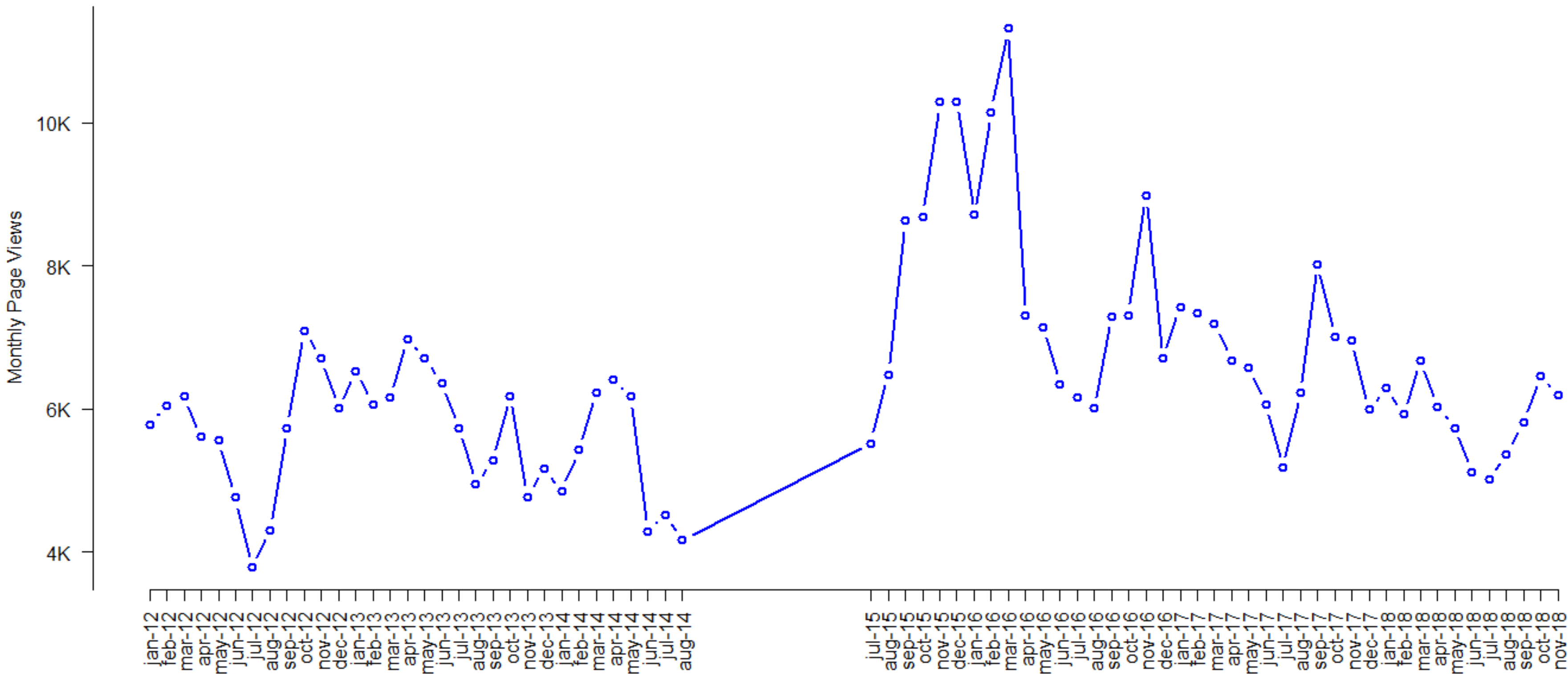




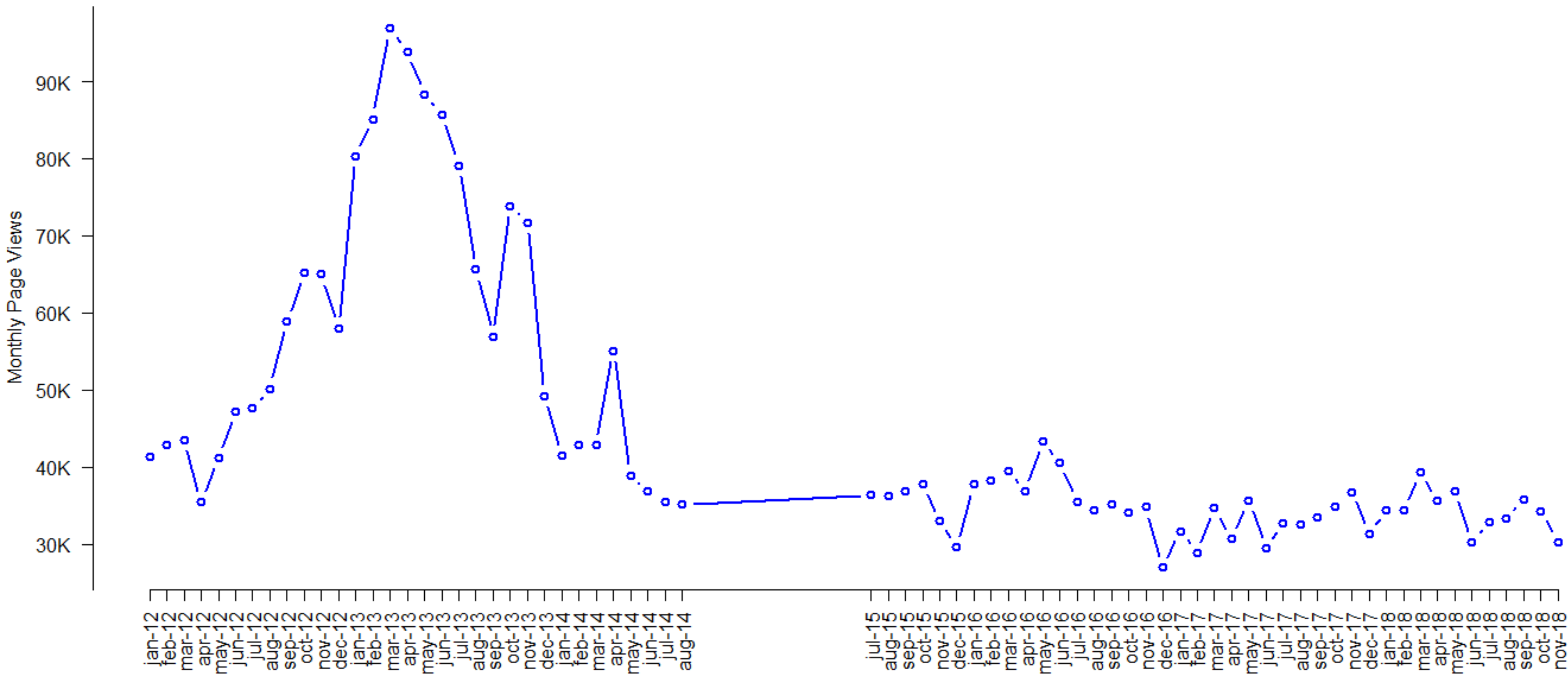
### Page Views for PLO



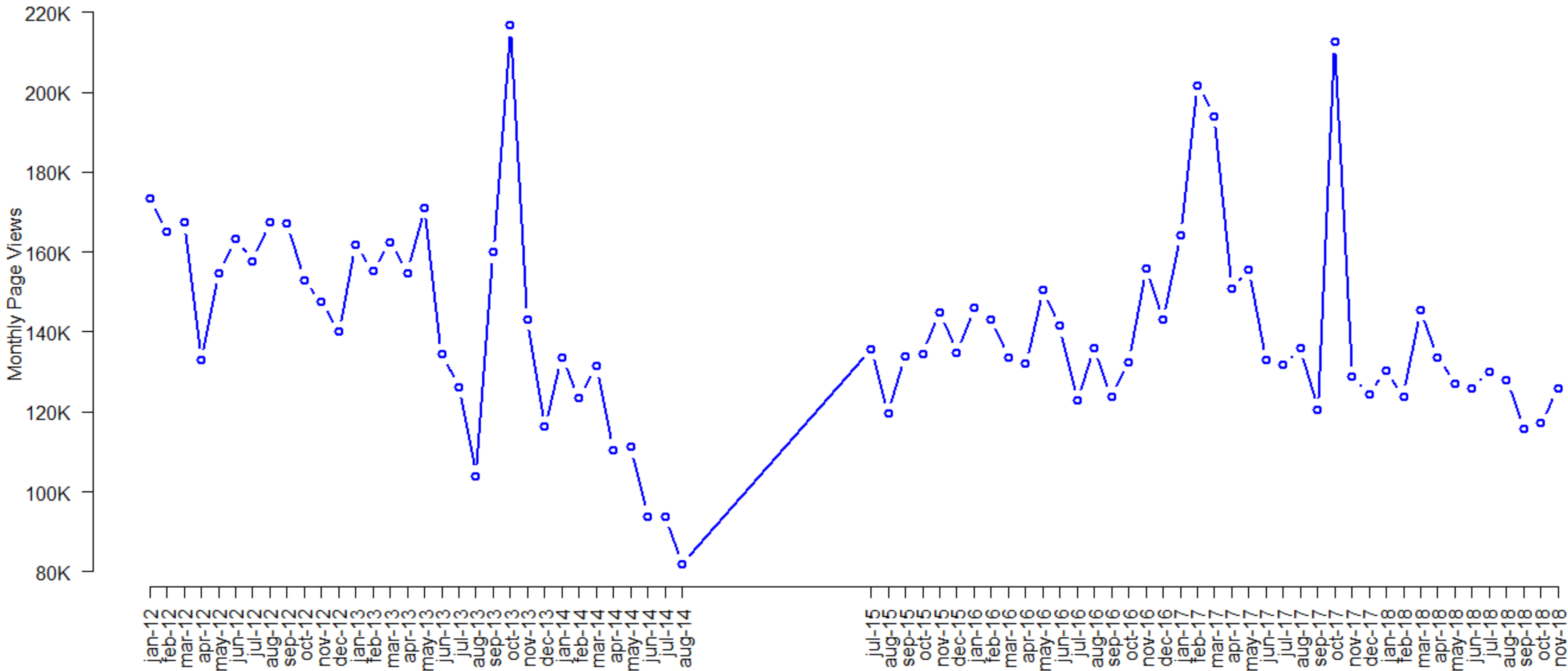
### Page Views for Political\_radicalism



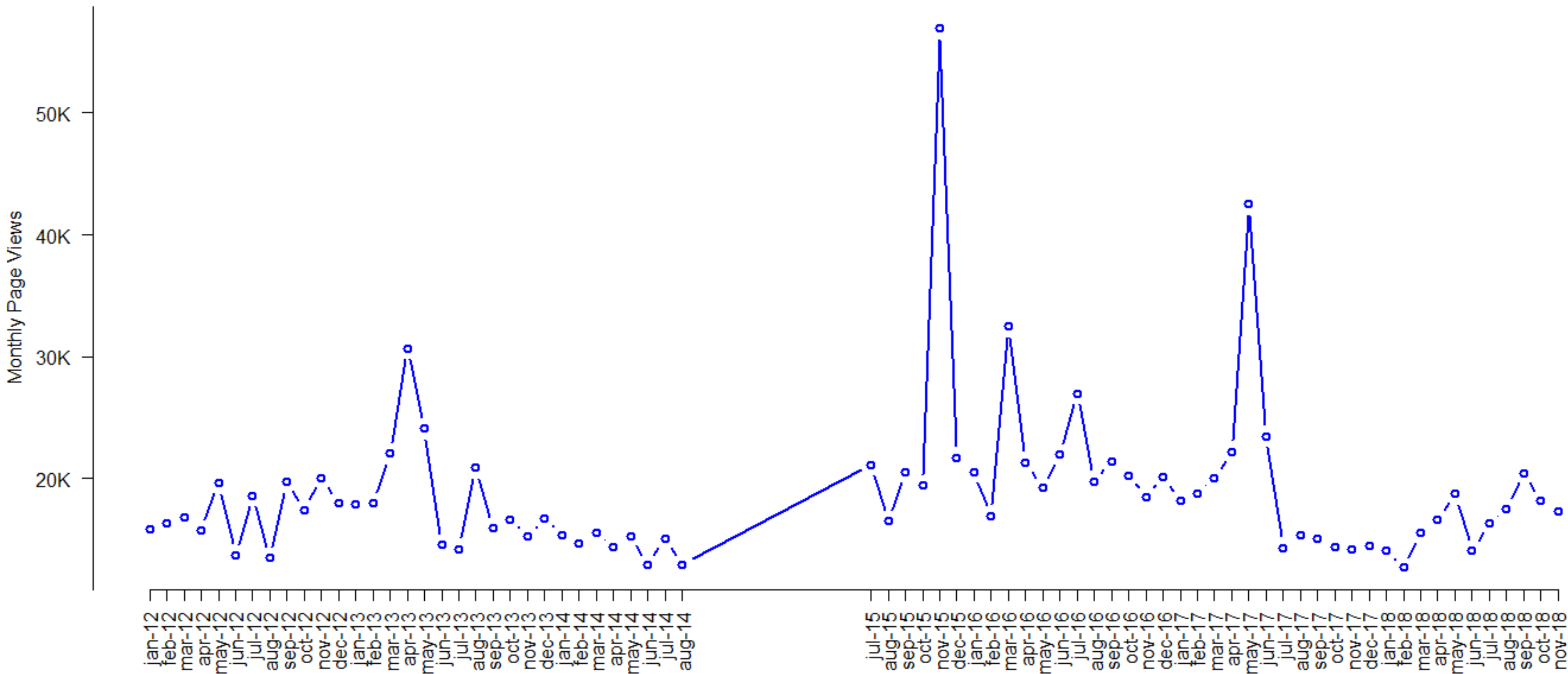
### Page Views for Recruitment



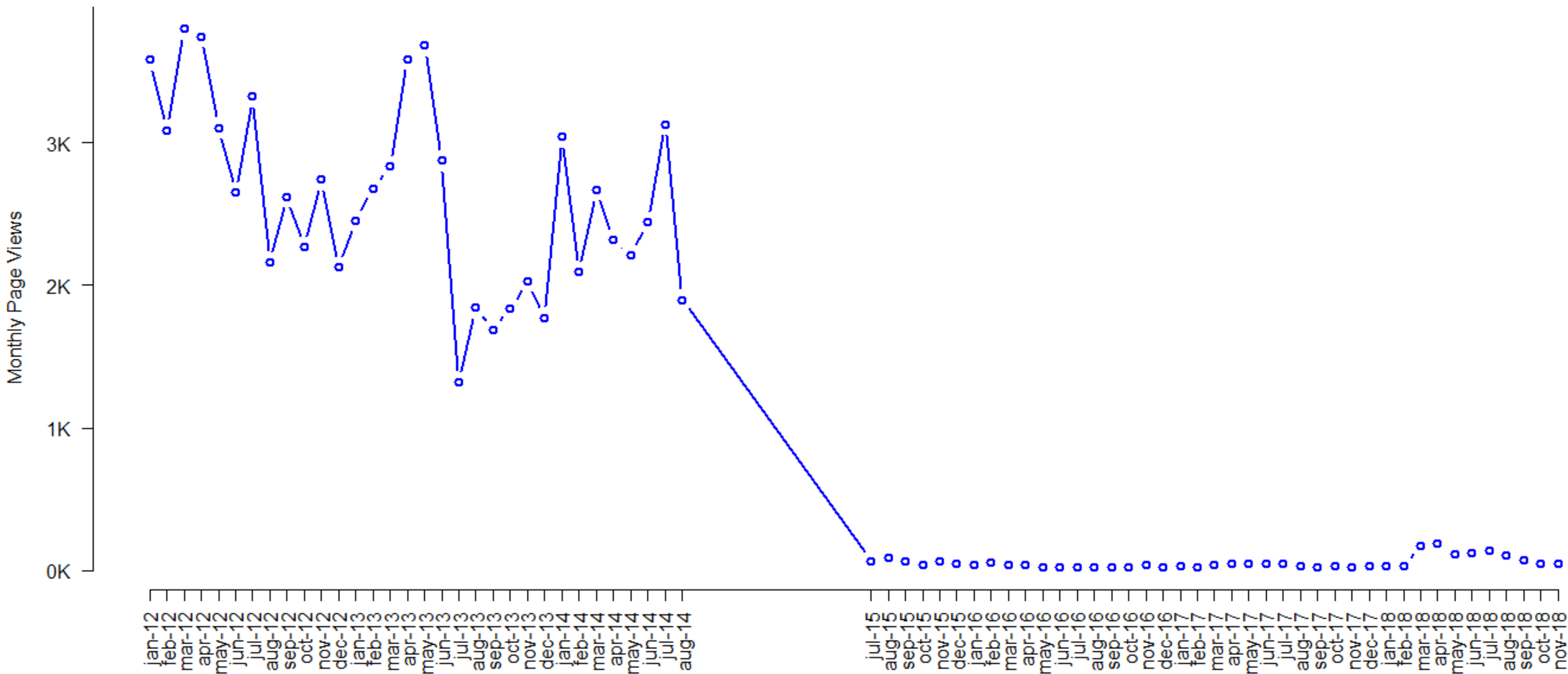
### Page Views for Somalia



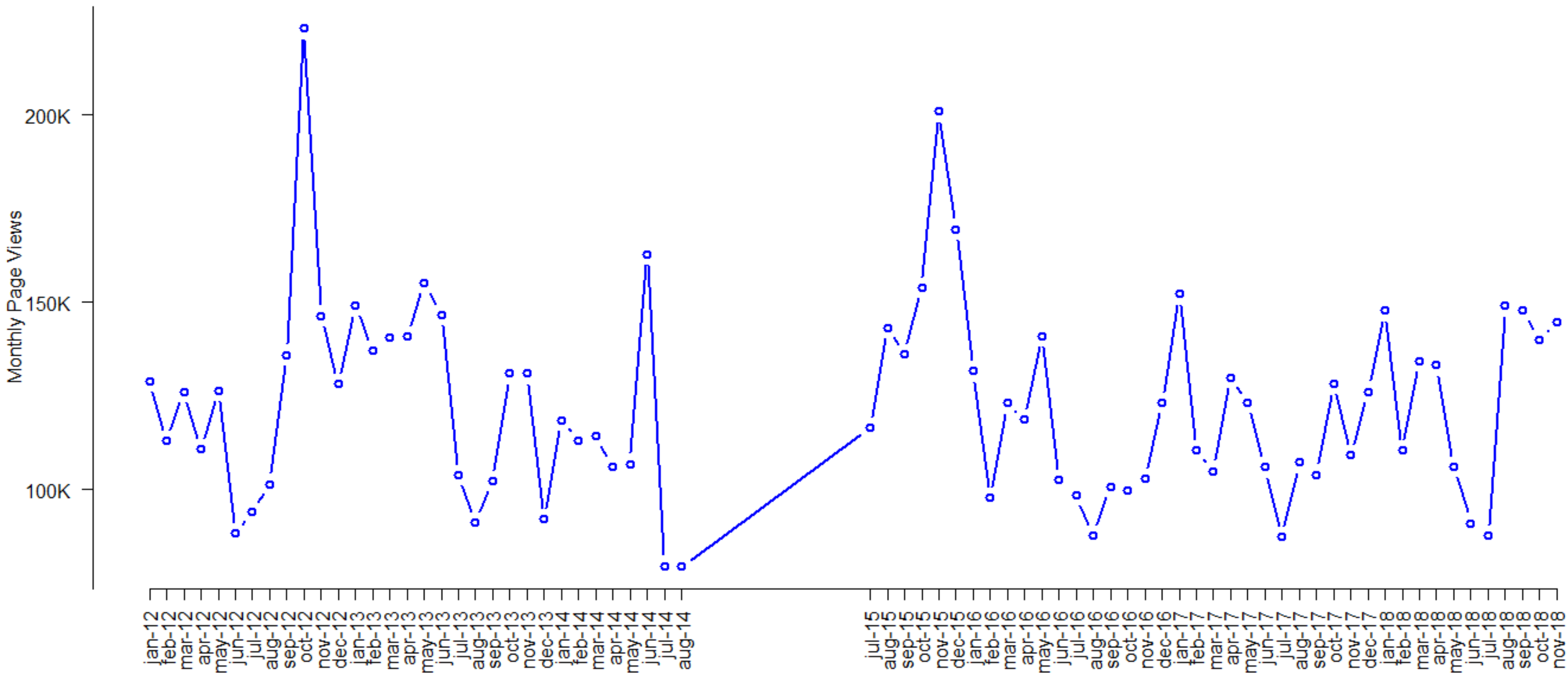
### Page Views for Suicide\_attack



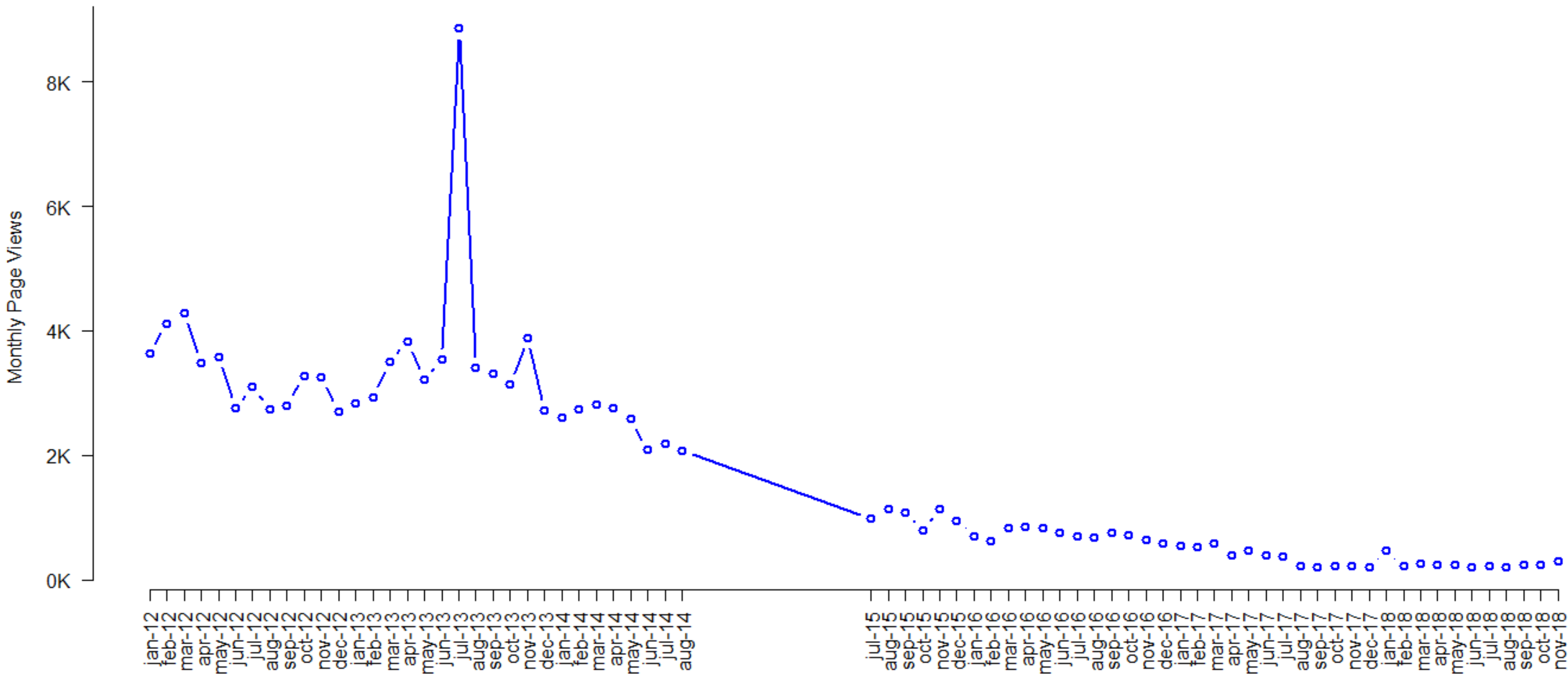
### Page Views for Suicide\_bomber



### Page Views for Taliban

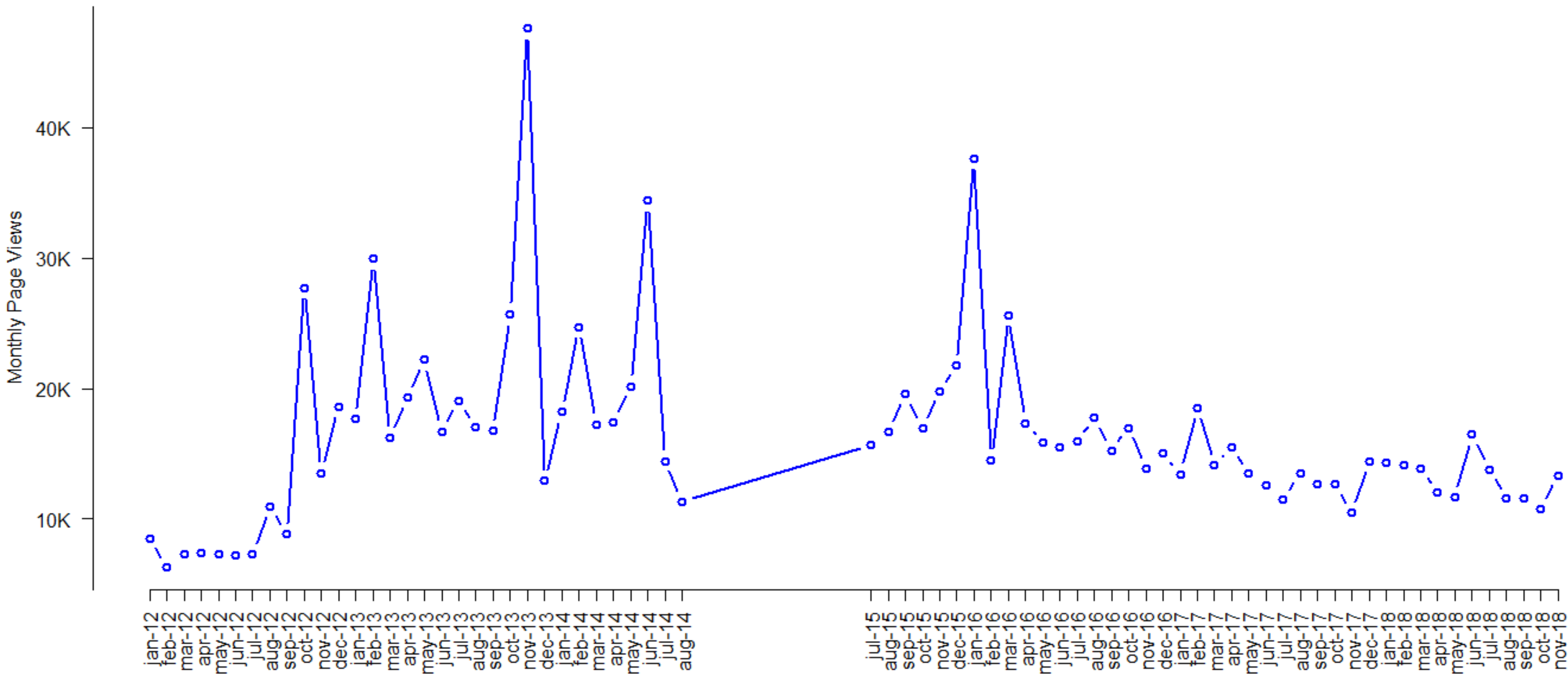


### Page Views for Tamil\_Tigers

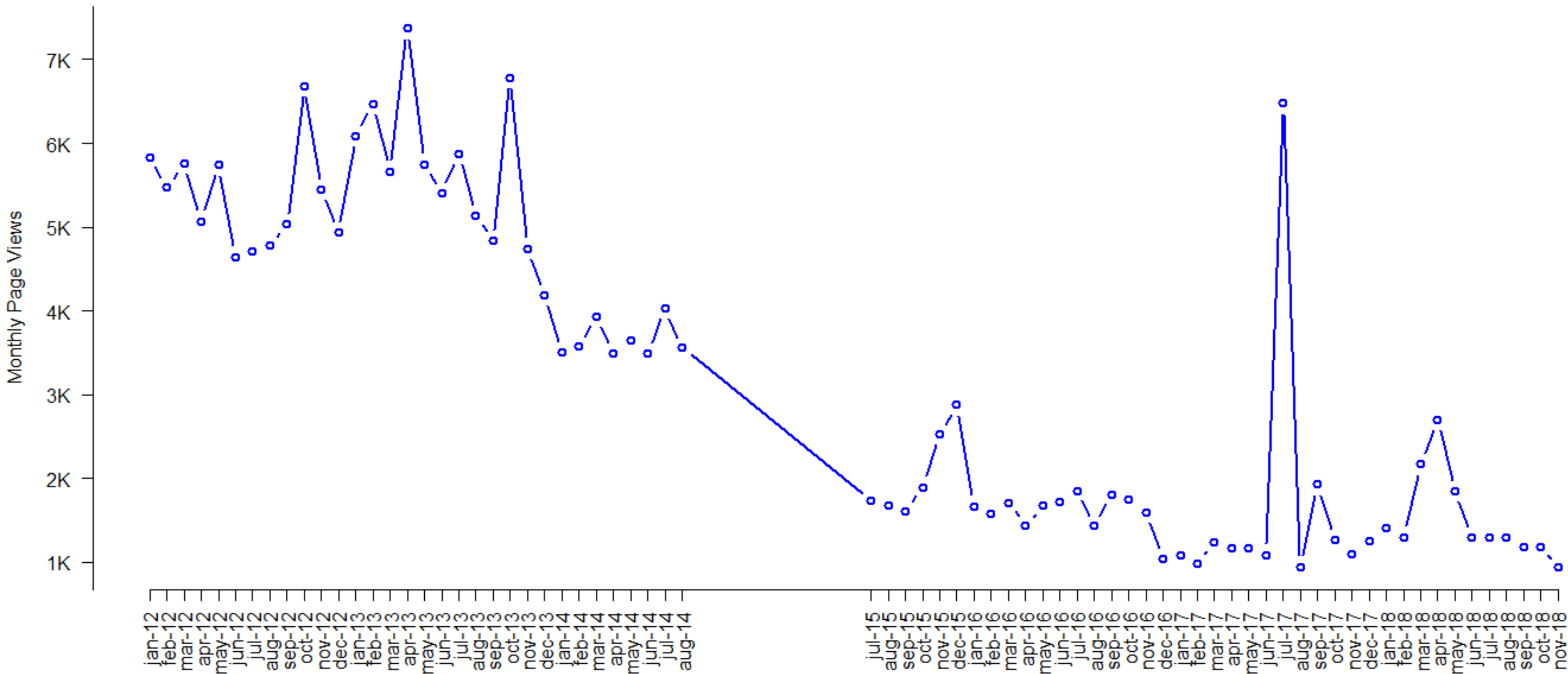




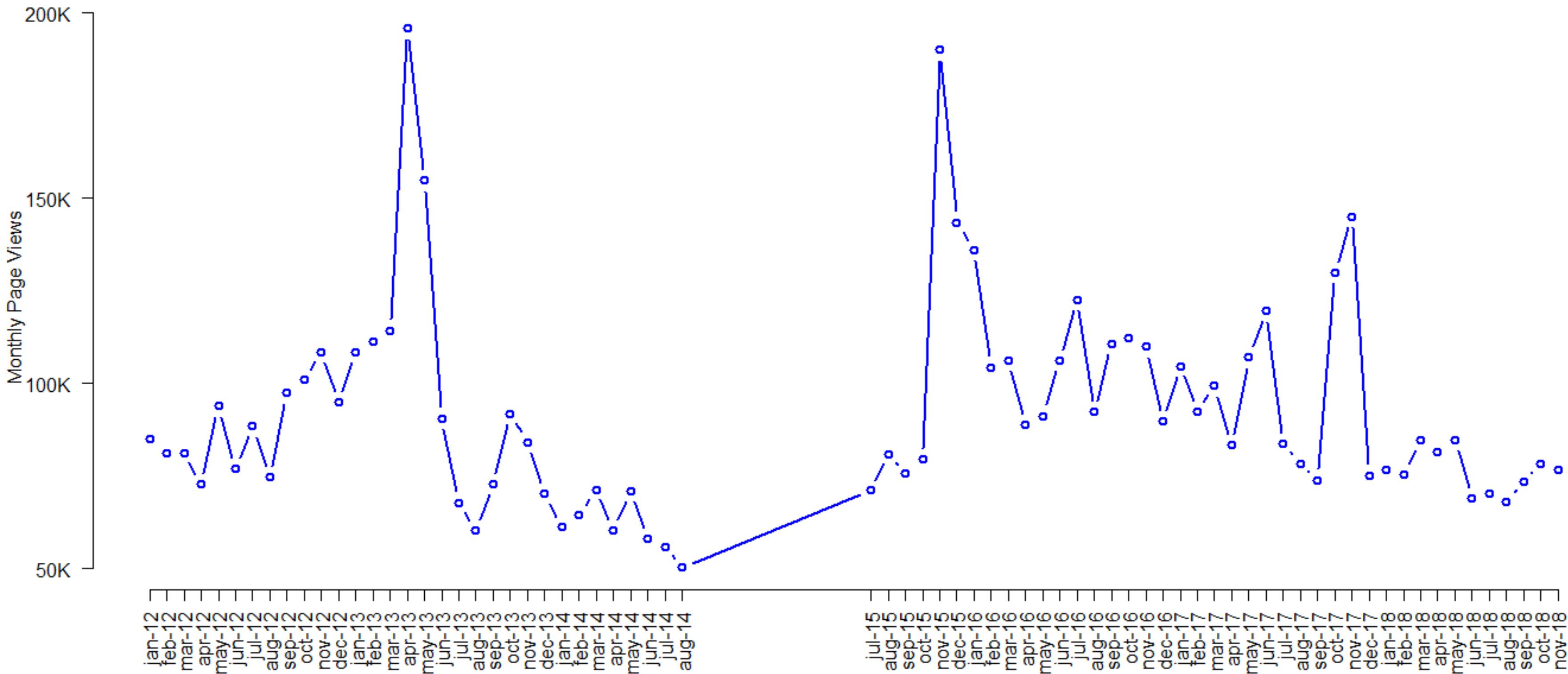
### Page Views for Tehrik\_i\_Taliban\_Pakistan



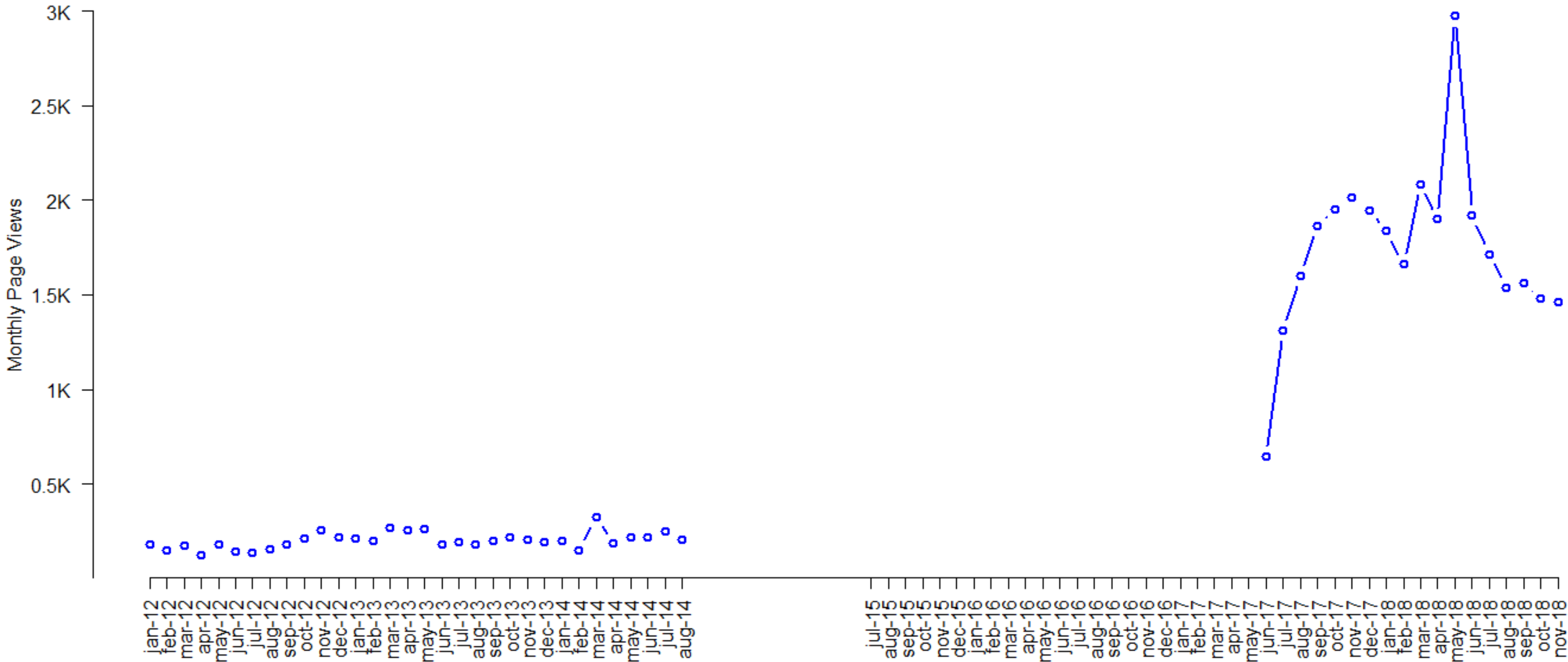
### Page Views for terror



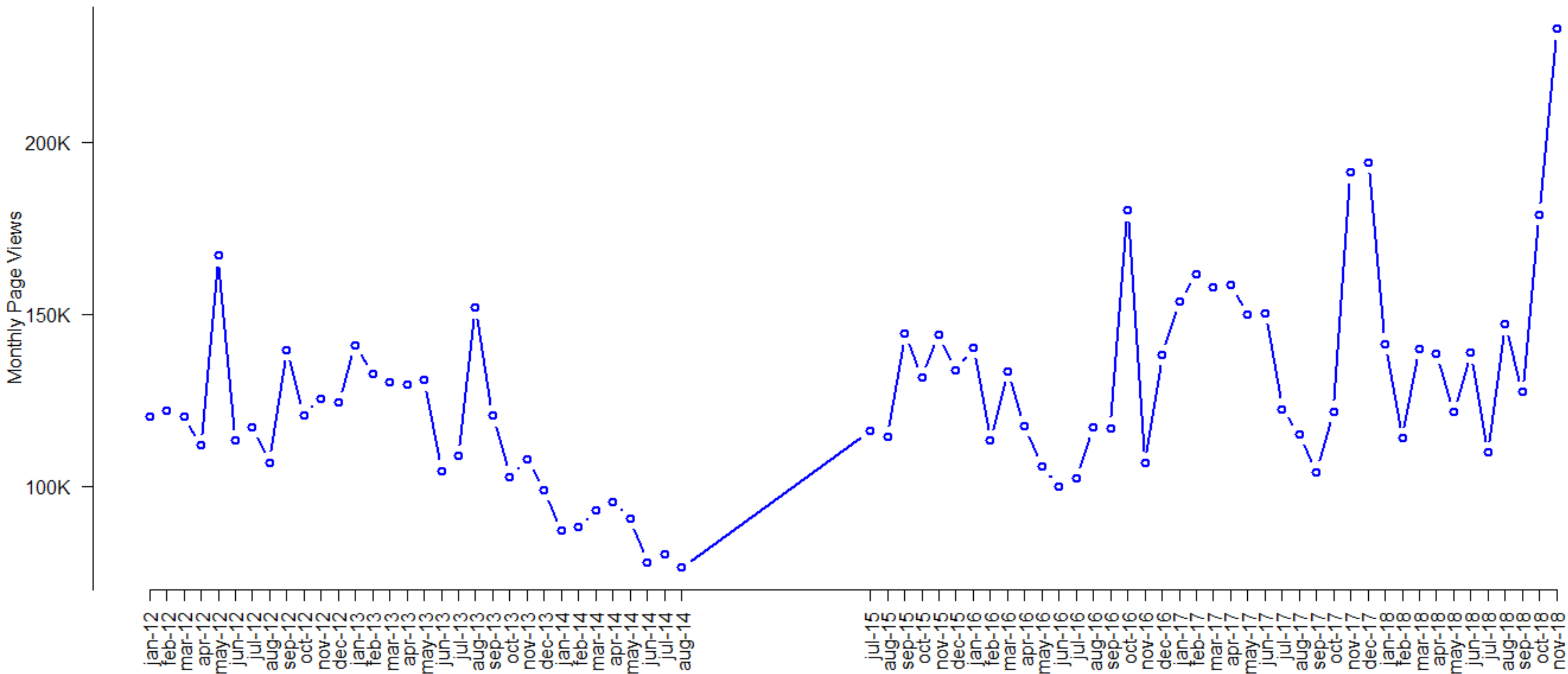
### Page Views for terrorism



### Page Views for Weapons\_grade

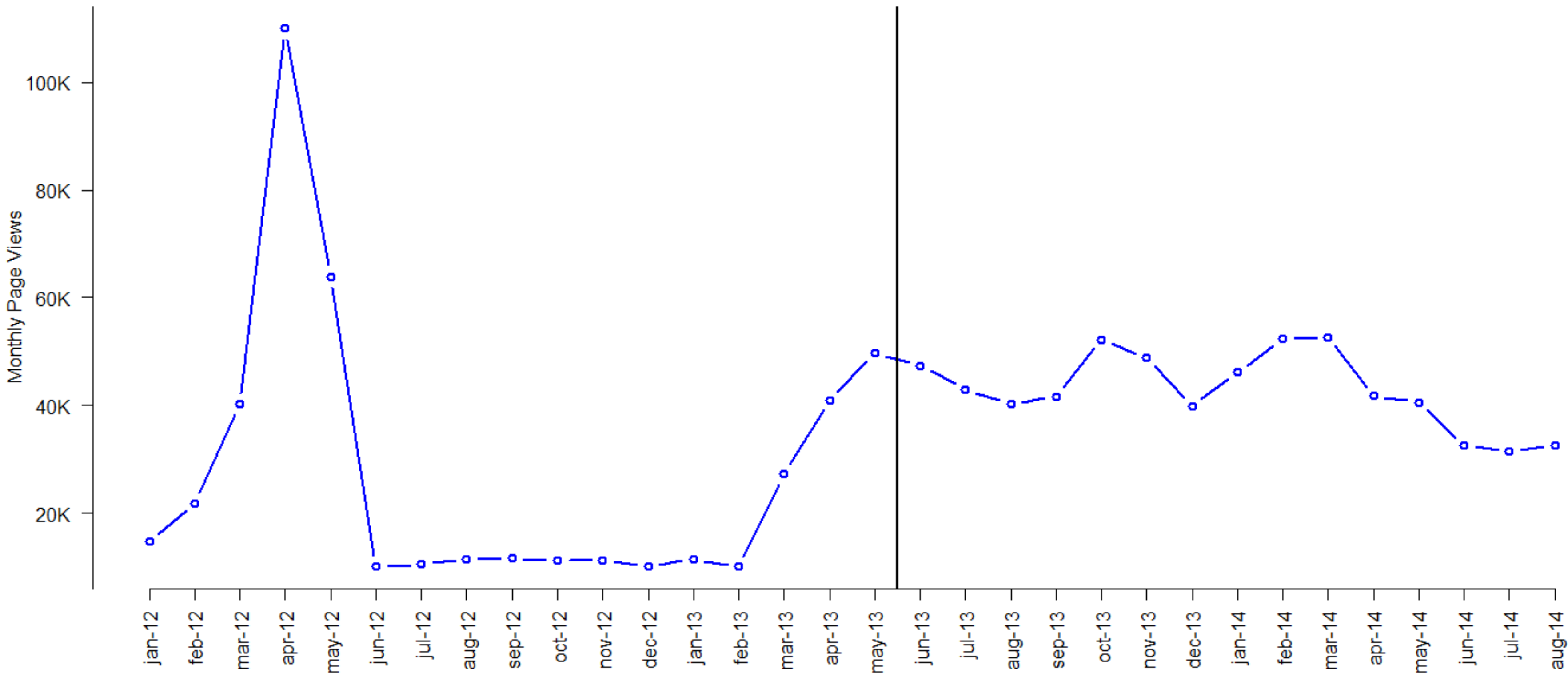


### Page Views for Yemen

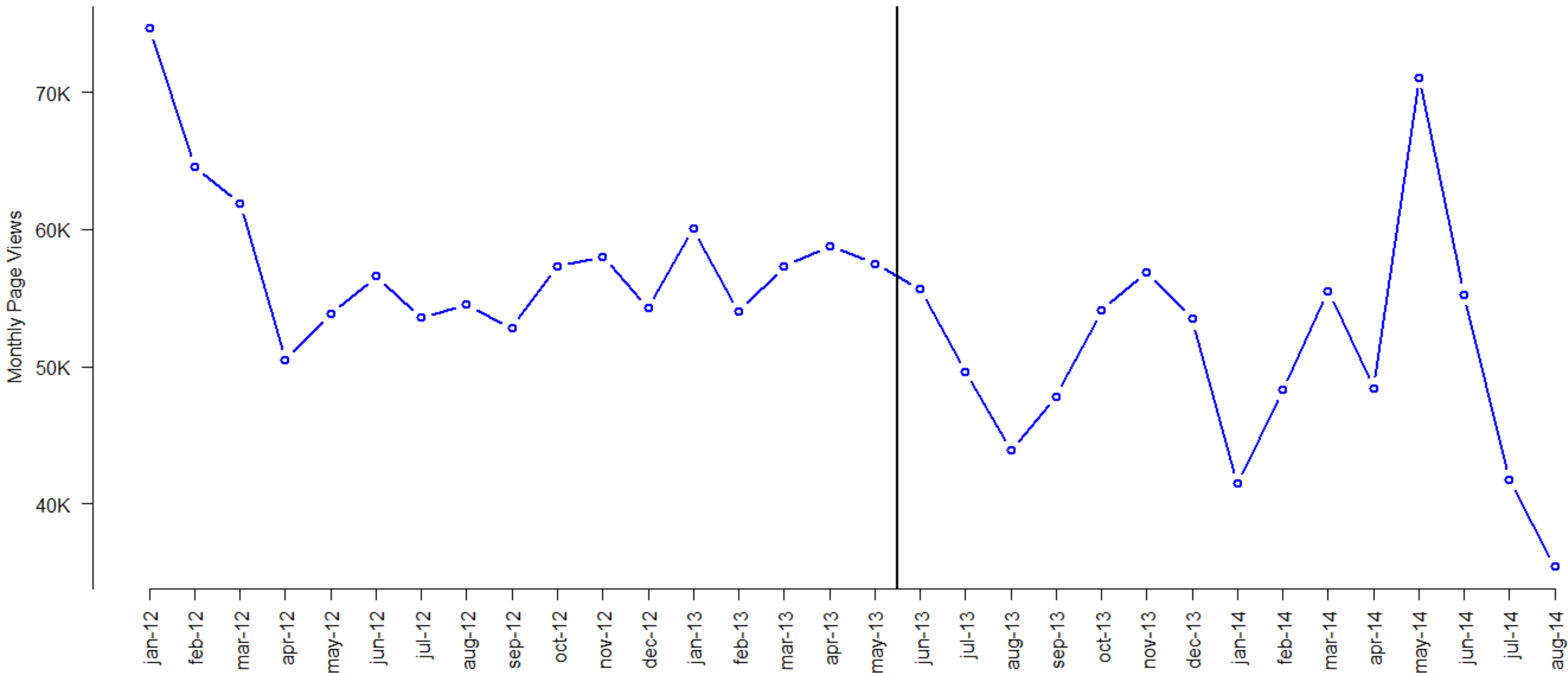


**APPENDIX VI: Page Views for 85 Comparative Articles**

### Infrastructure: Page Views for airplane

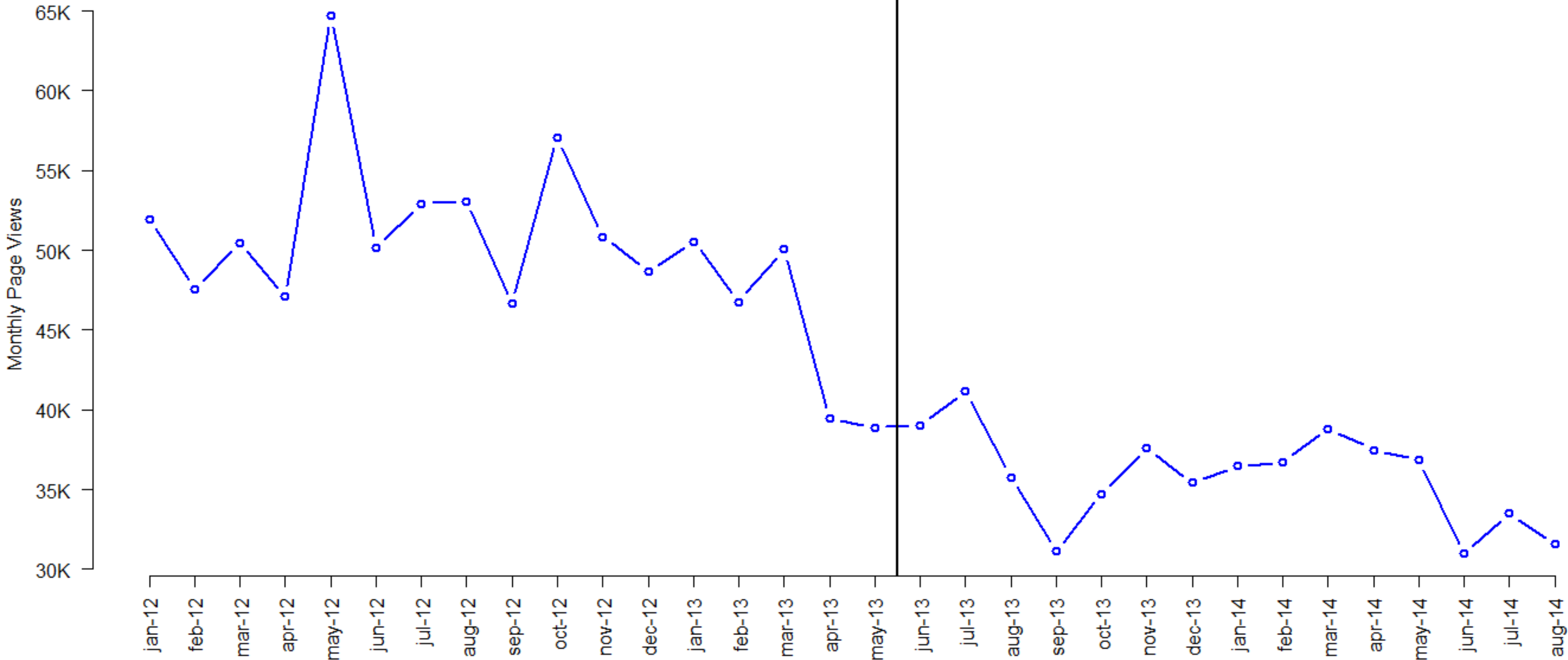


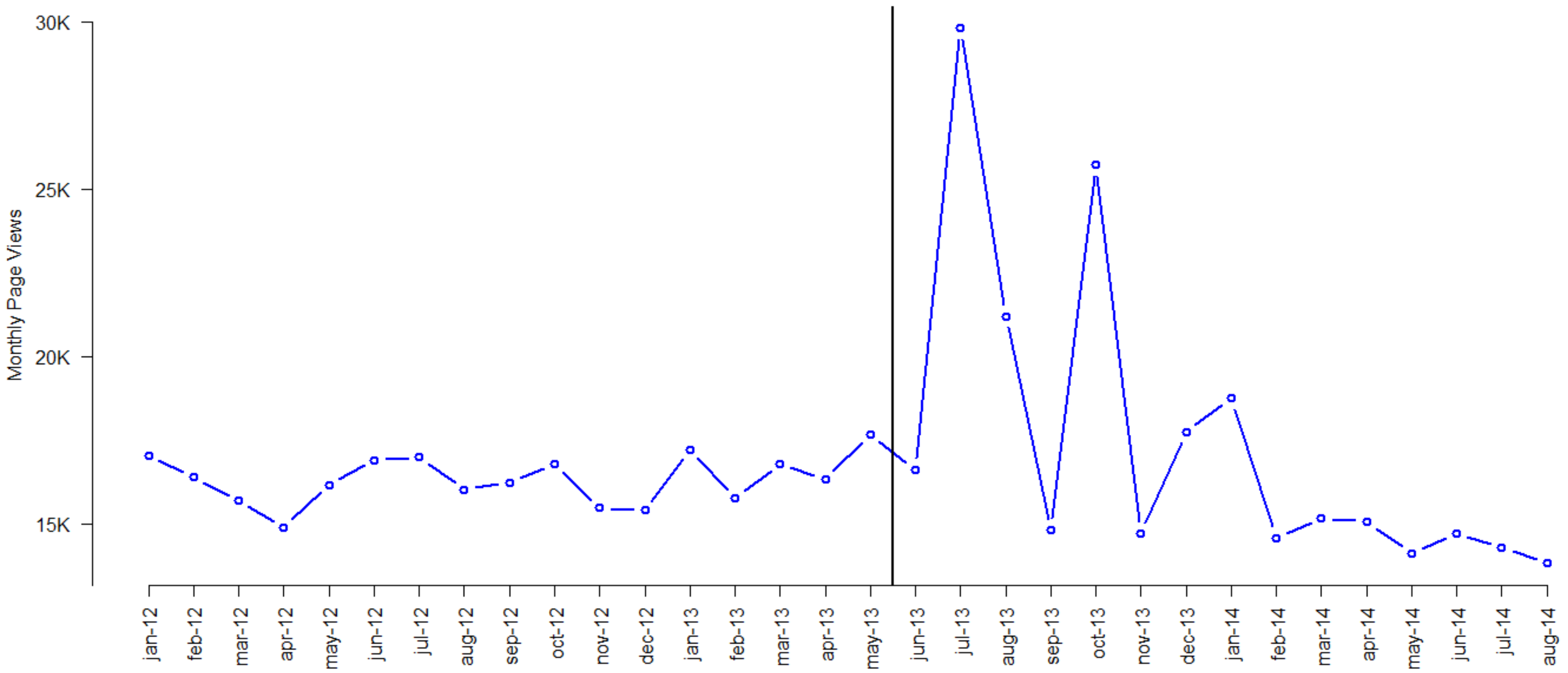
### Infrastructure: Page Views for airport



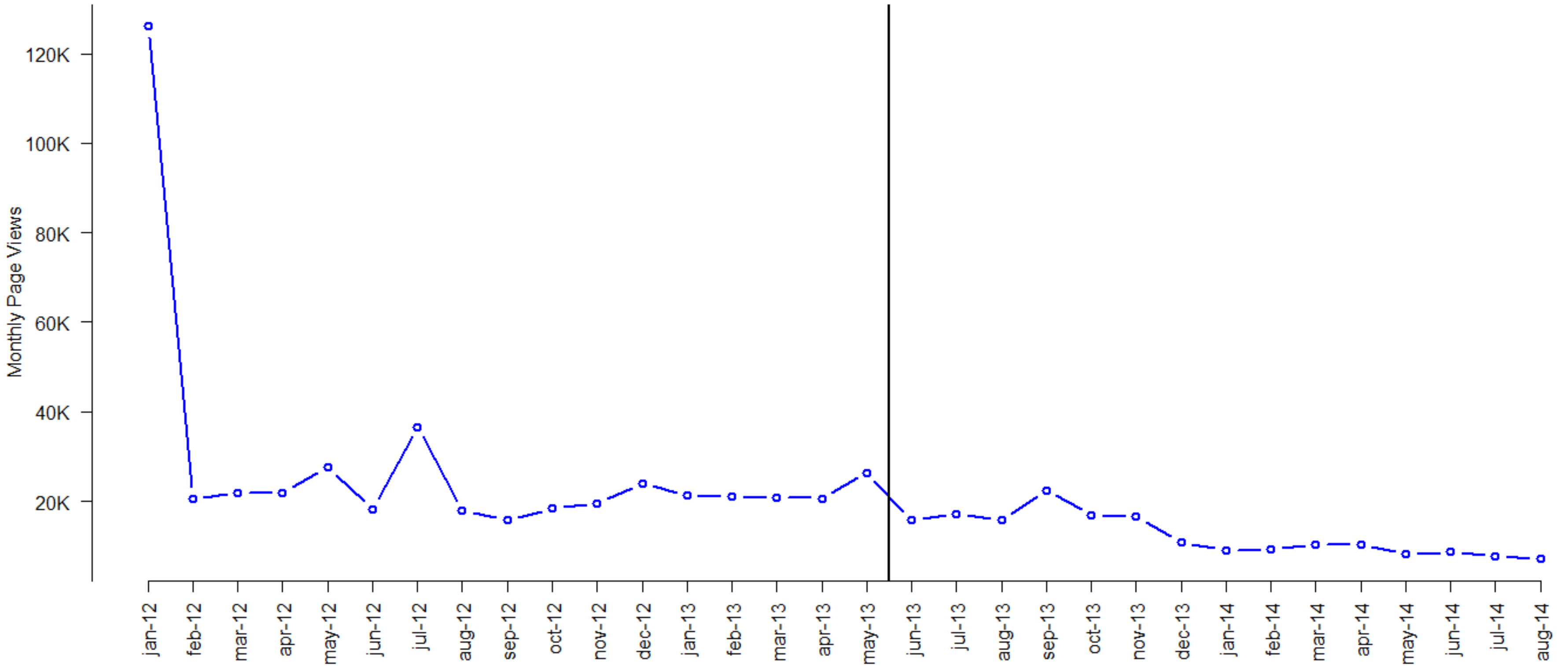


### Infrastructure: Page Views for amtrak

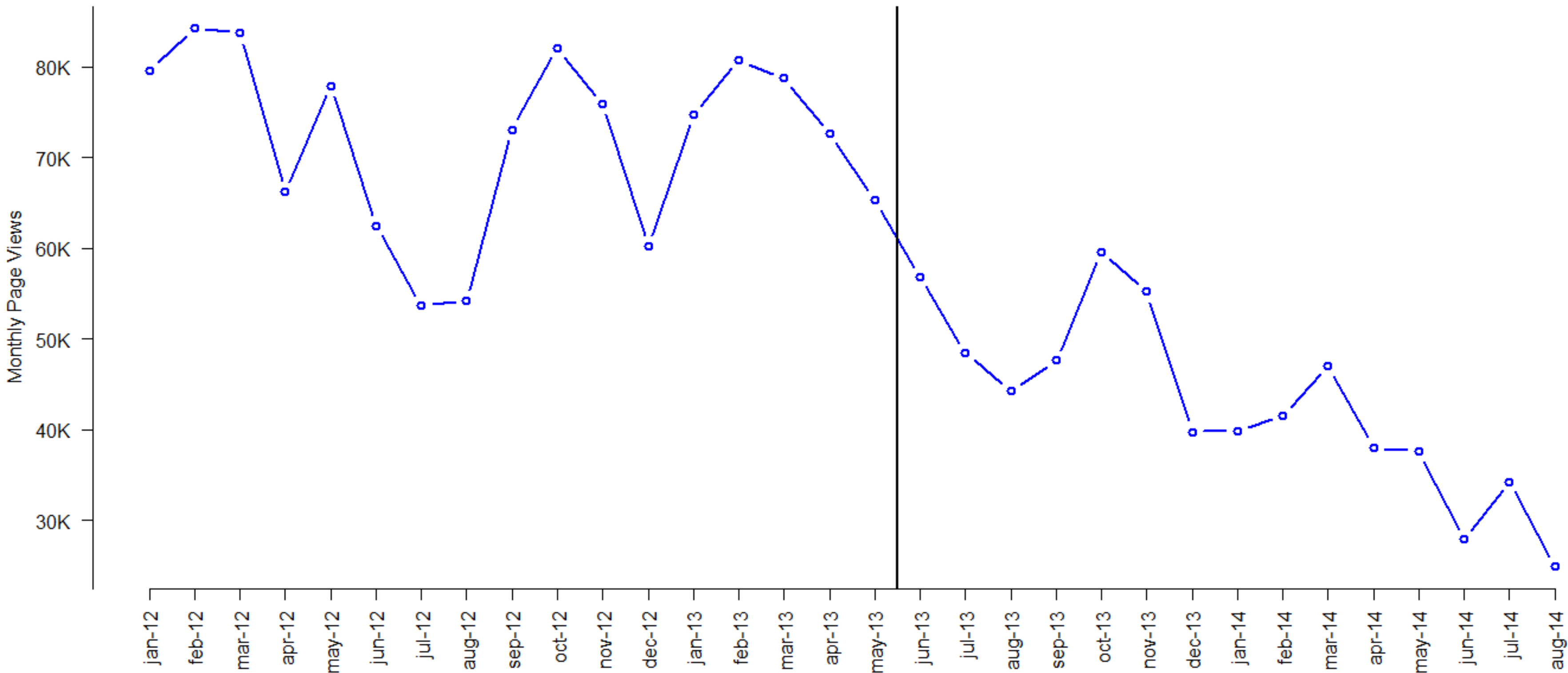




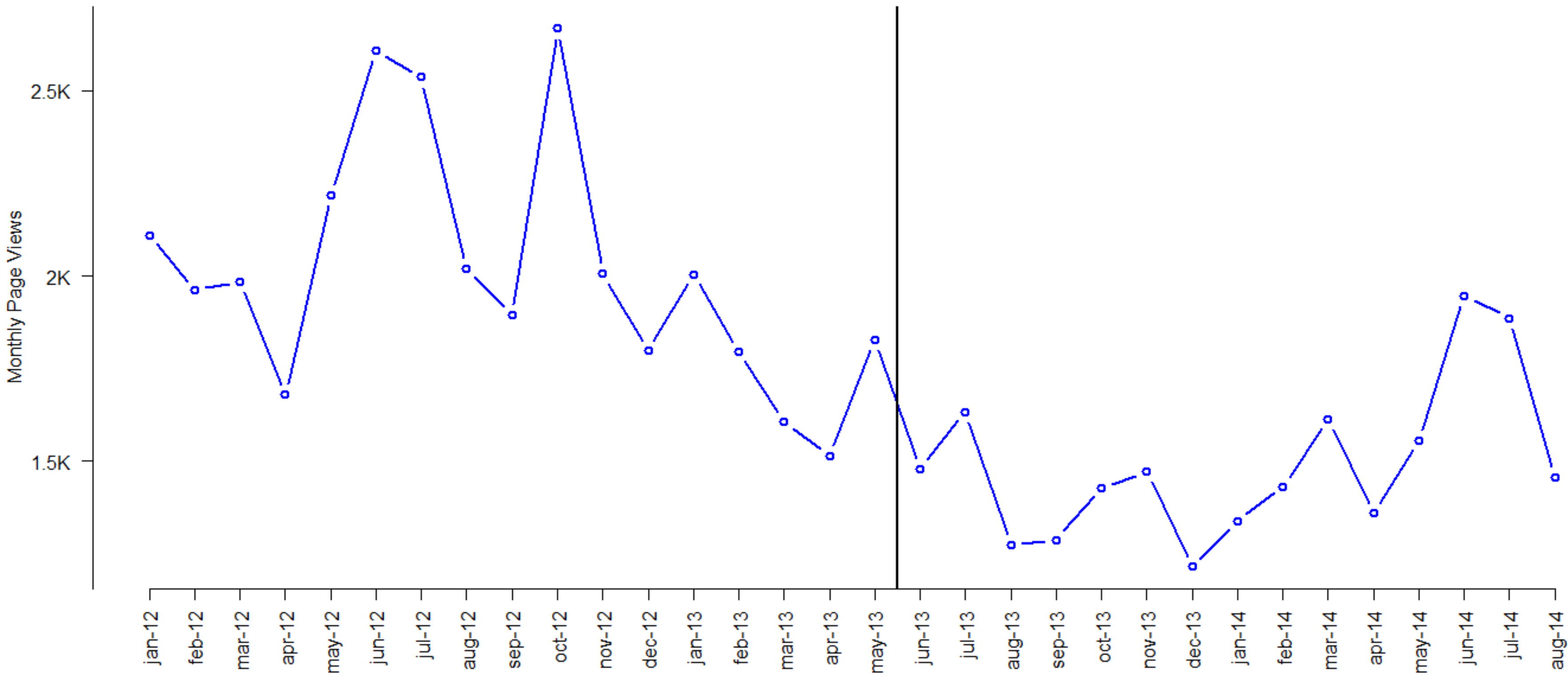
### Infrastructure: Page Views for blackout



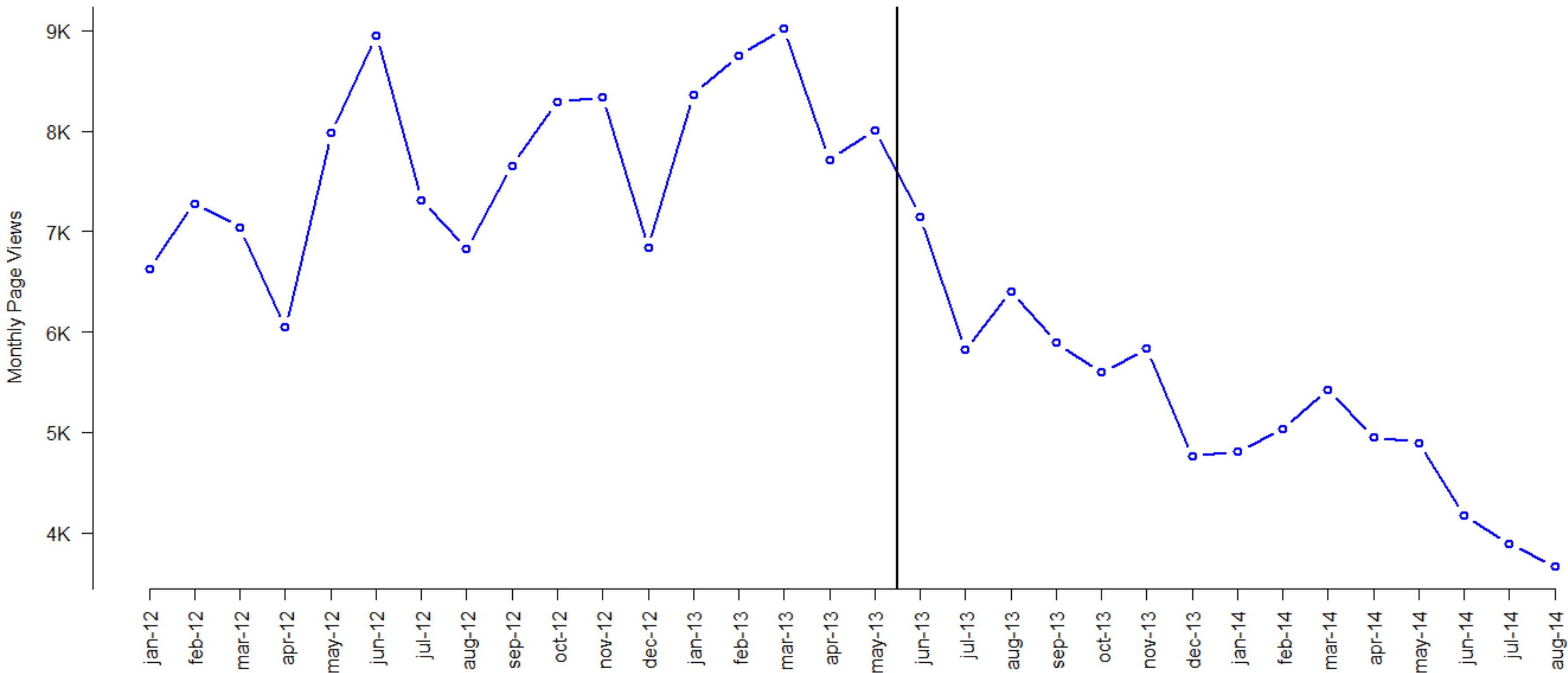
### Infrastructure: Page Views for bridge



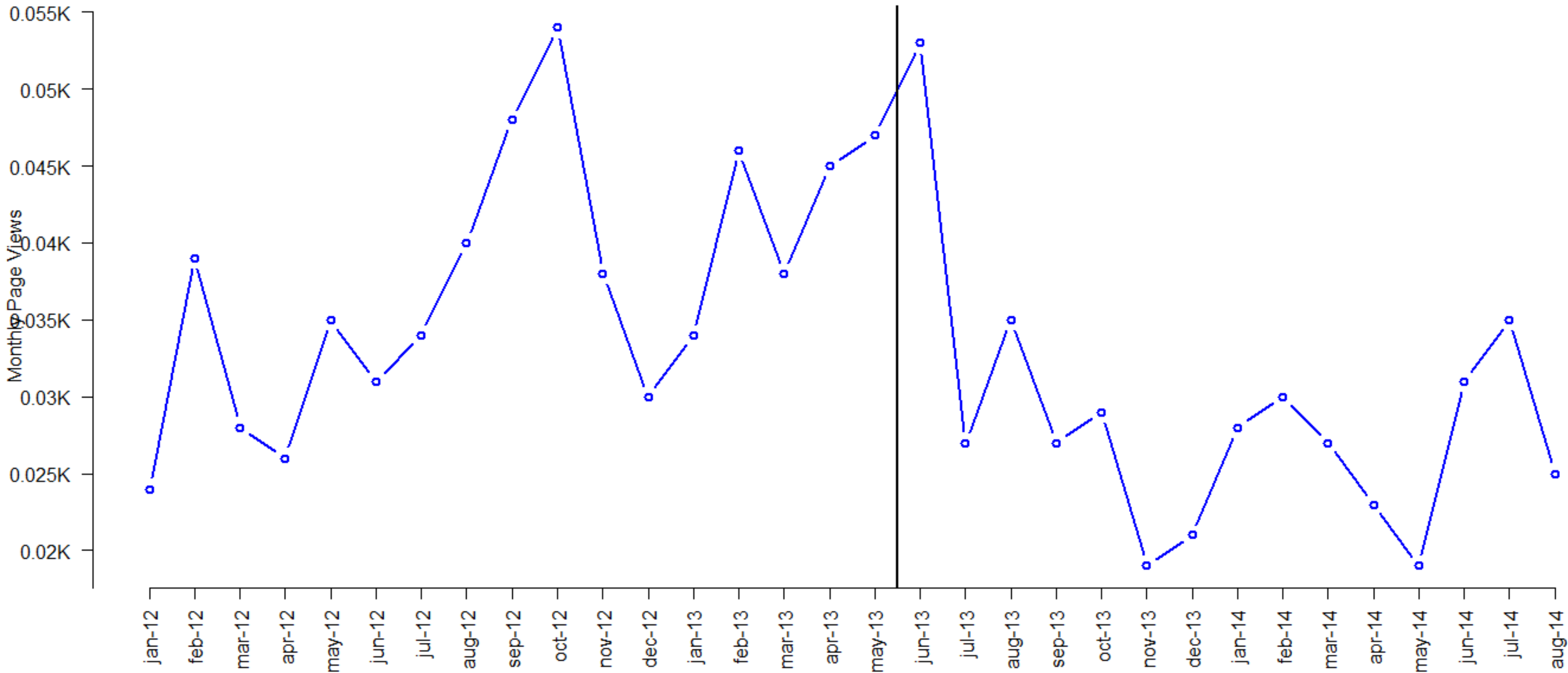
### Infrastructure: Page Views for brownout



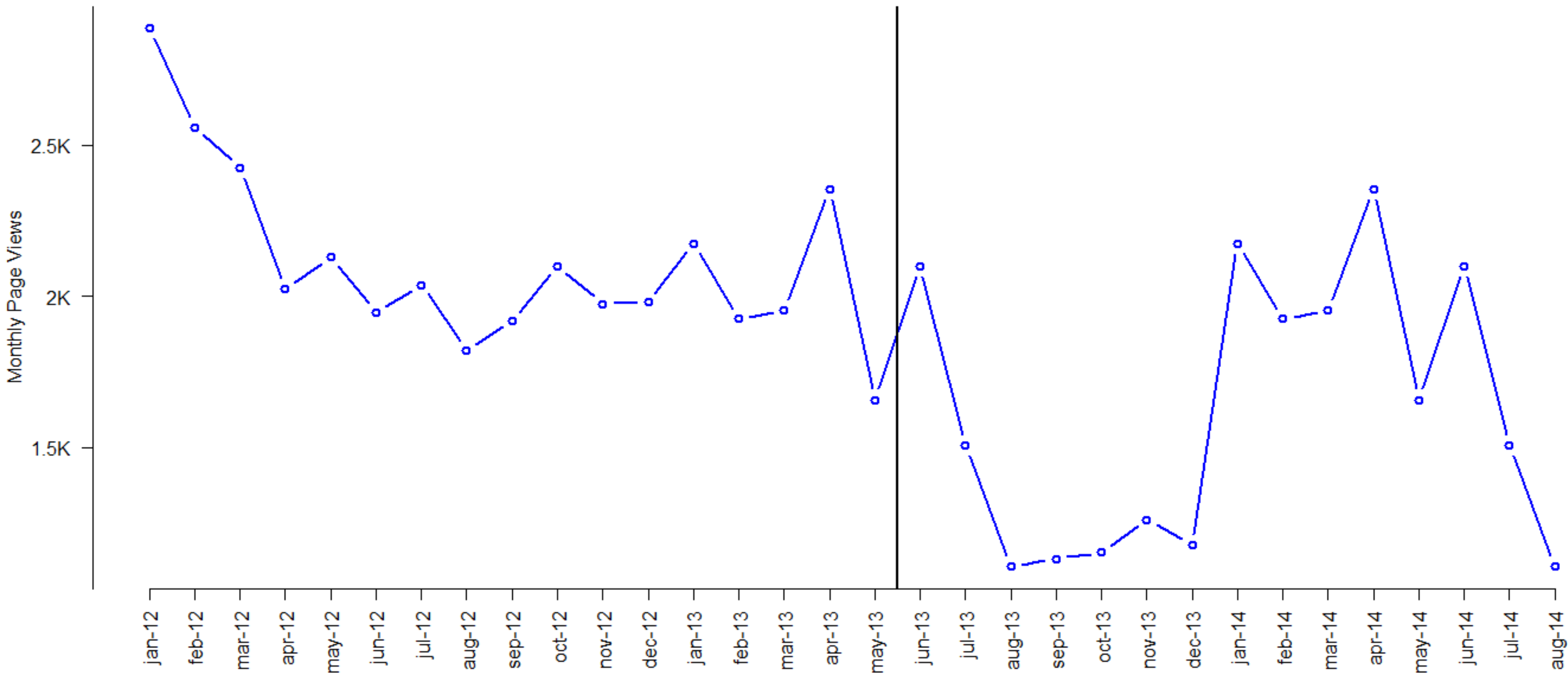
### Infrastructure: Page Views for chemical\_burn



### Infrastructure: Page Views for ckr

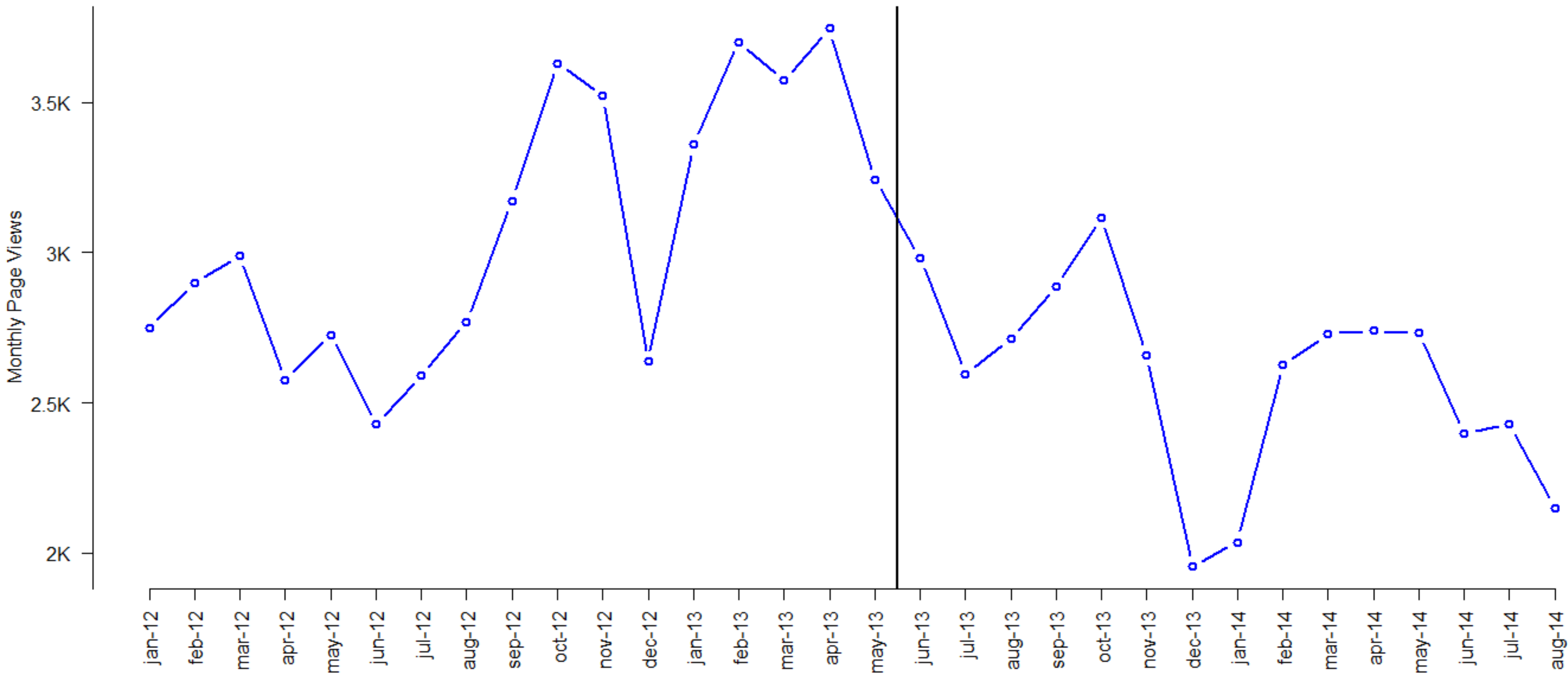


### Infrastructure: Page Views for collapse

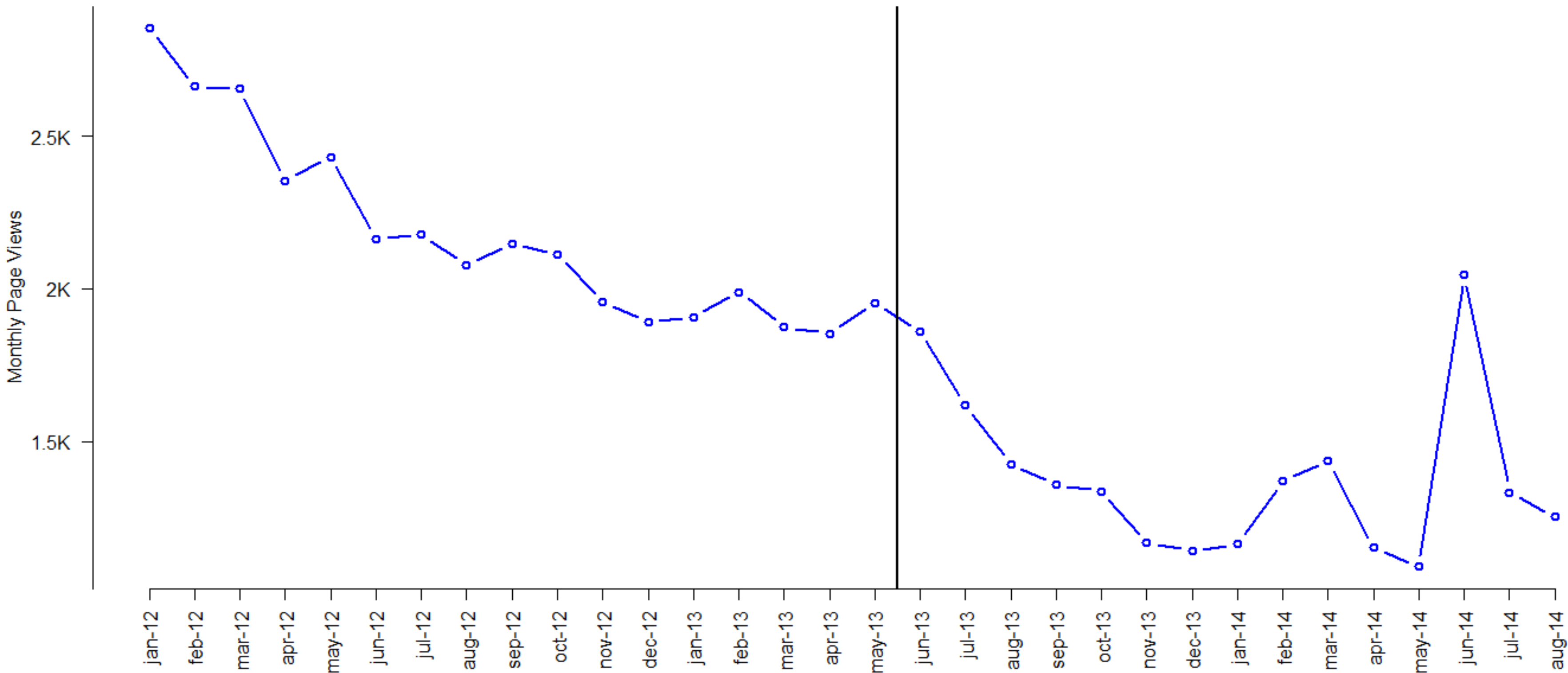




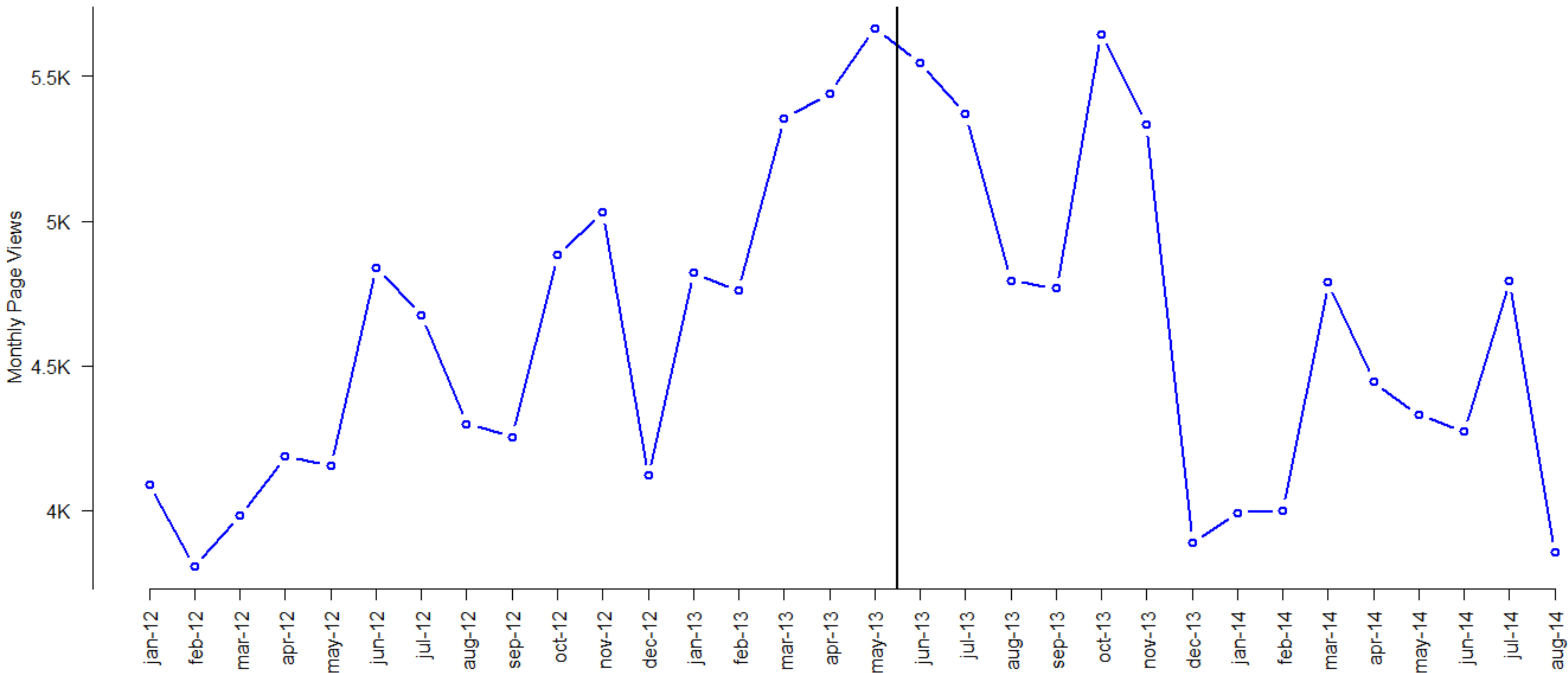
### Infrastructure: Page Views for critical\_infrastructure



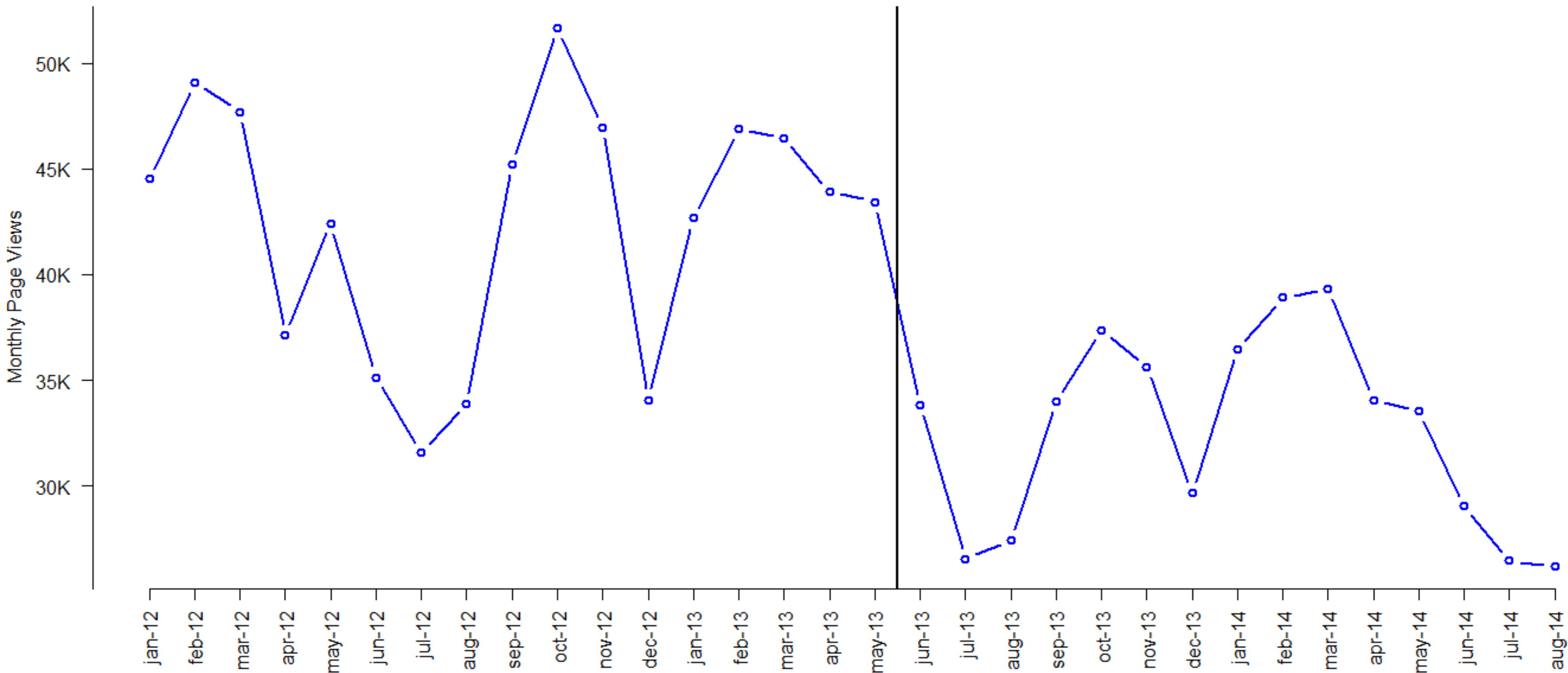
### Infrastructure: Page Views for delay

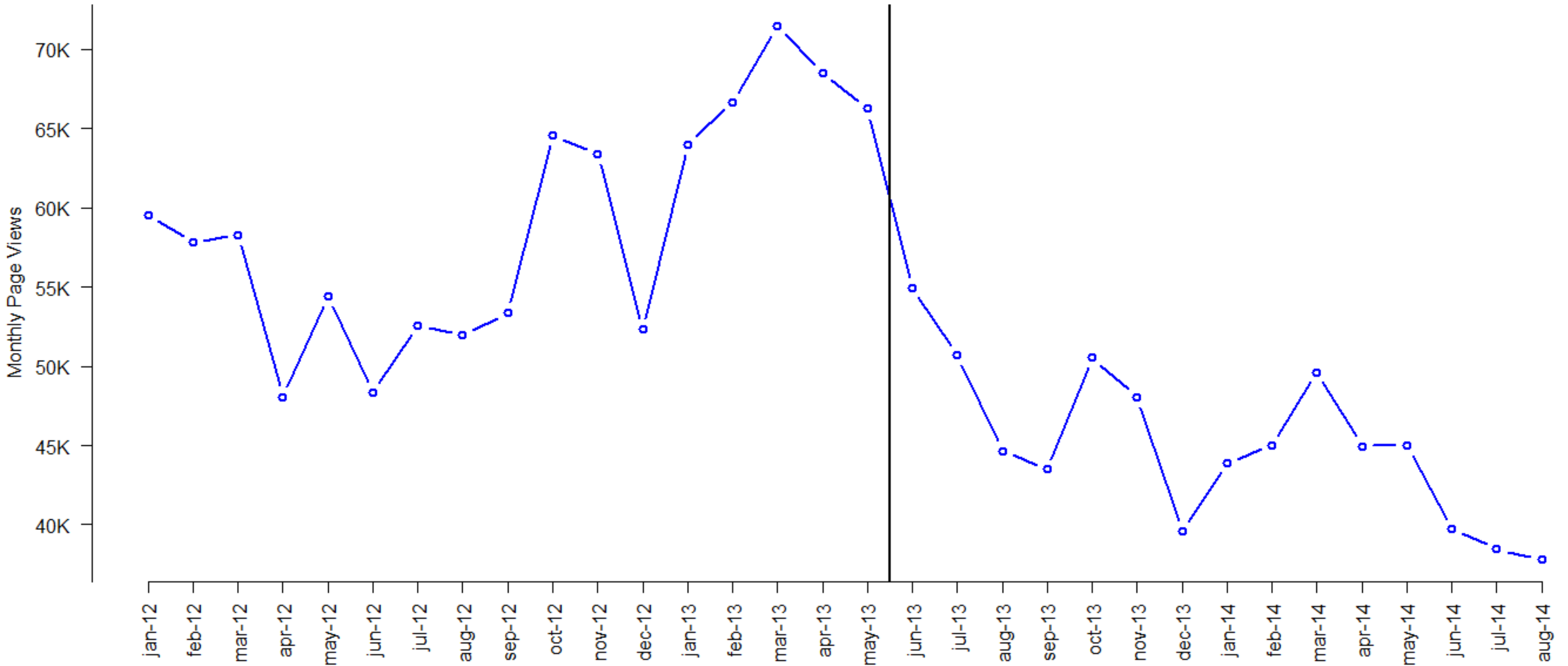


### Infrastructure: Page Views for dock\_maritime

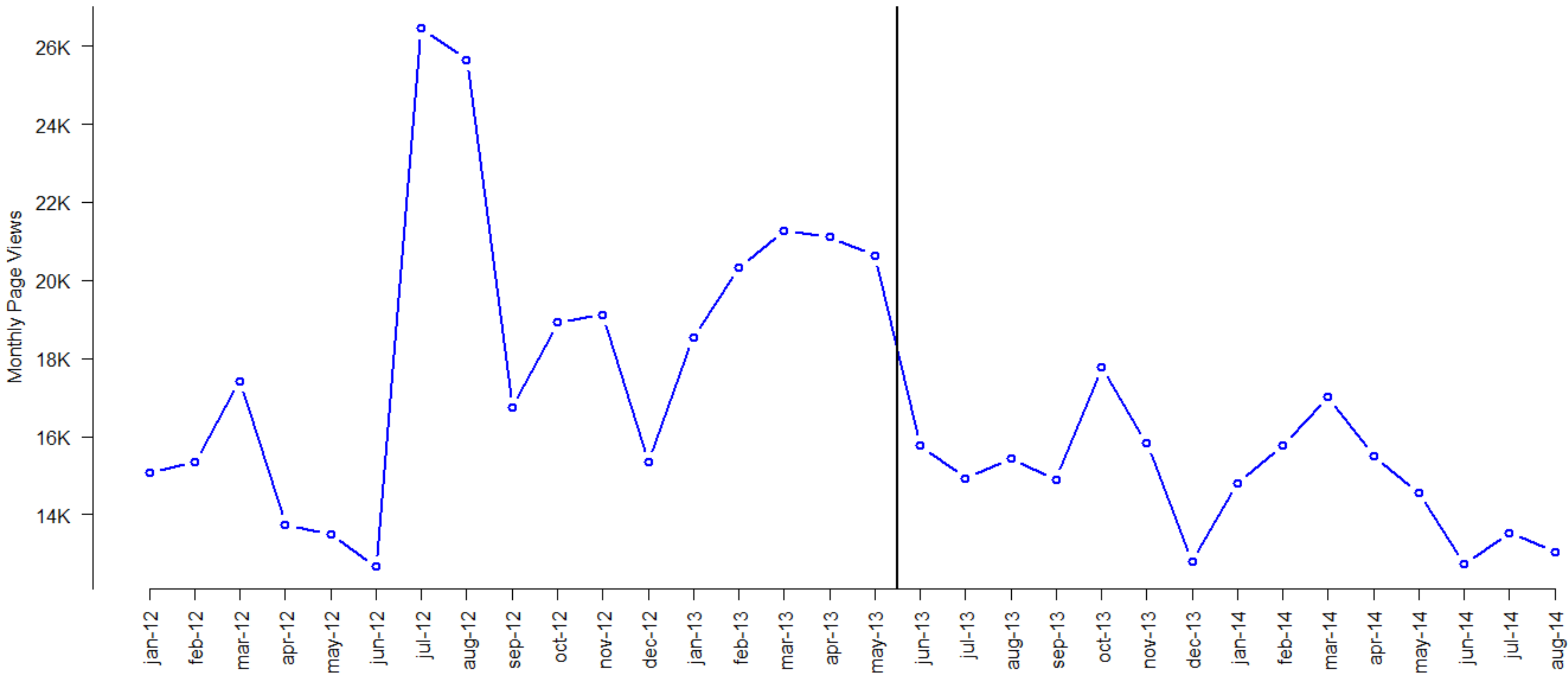


### Infrastructure: Page Views for electric\_power

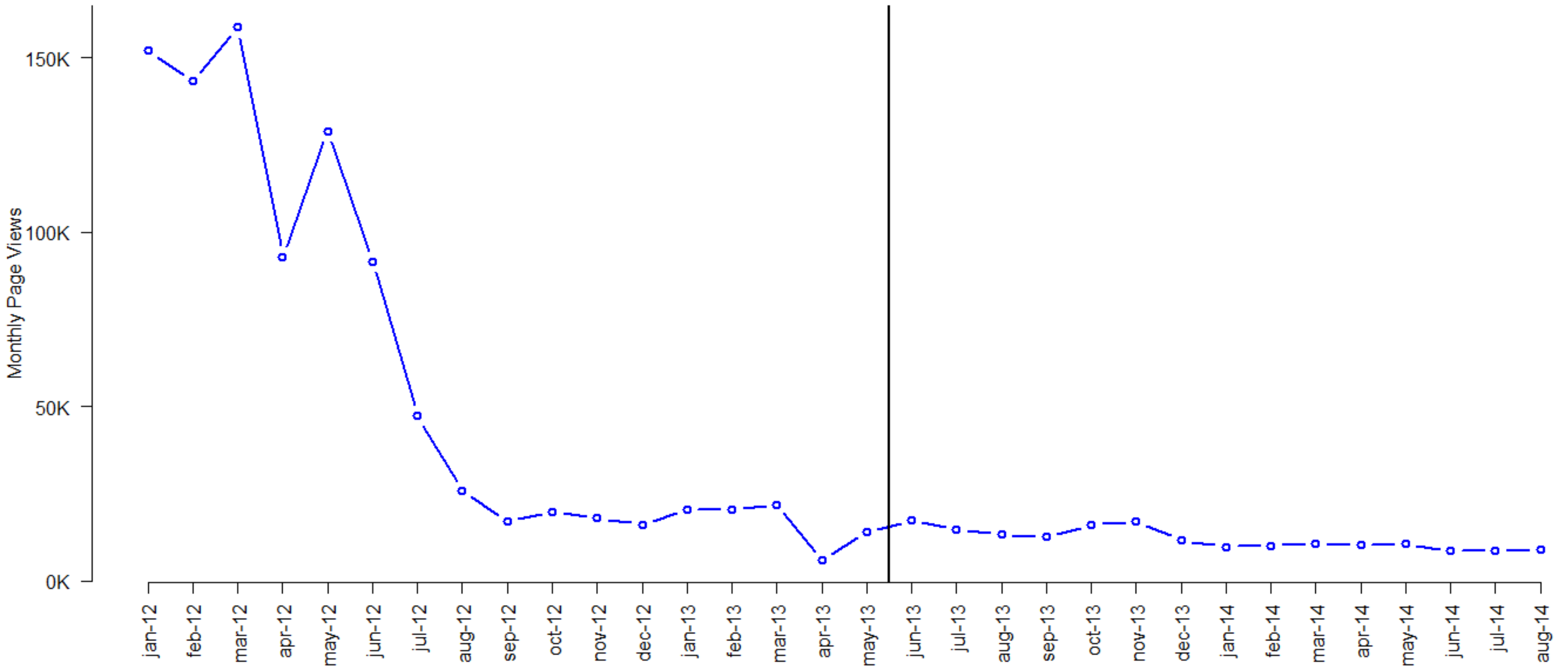


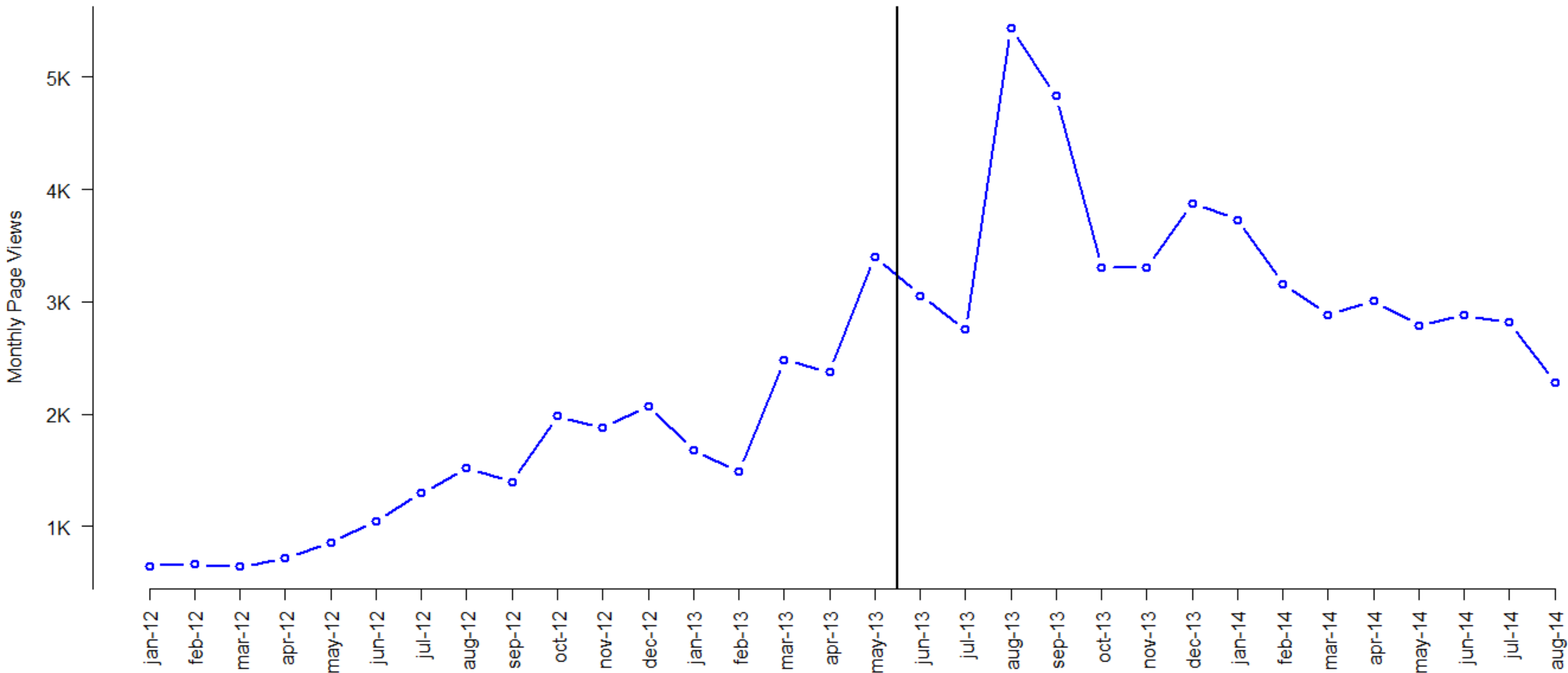


### Infrastructure: Page Views for electrical\_grid



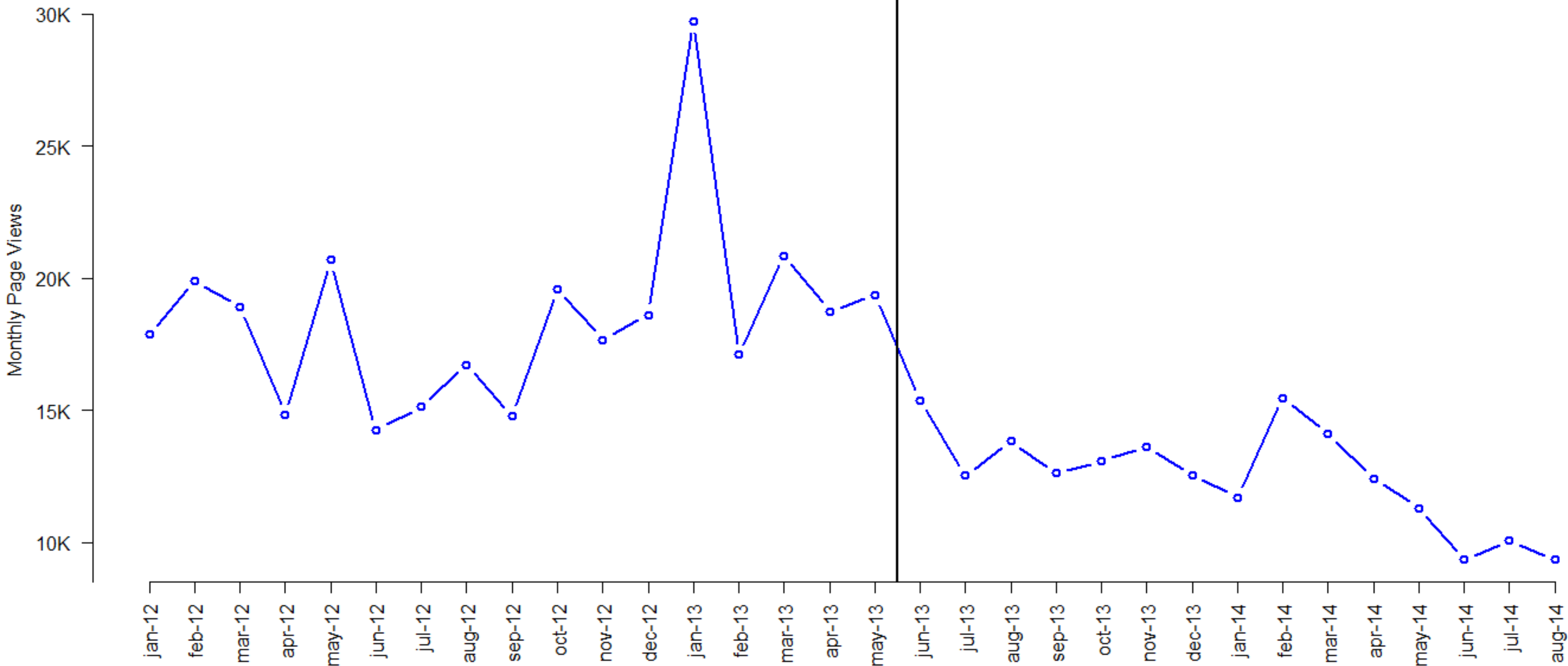
### Infrastructure: Page Views for failure

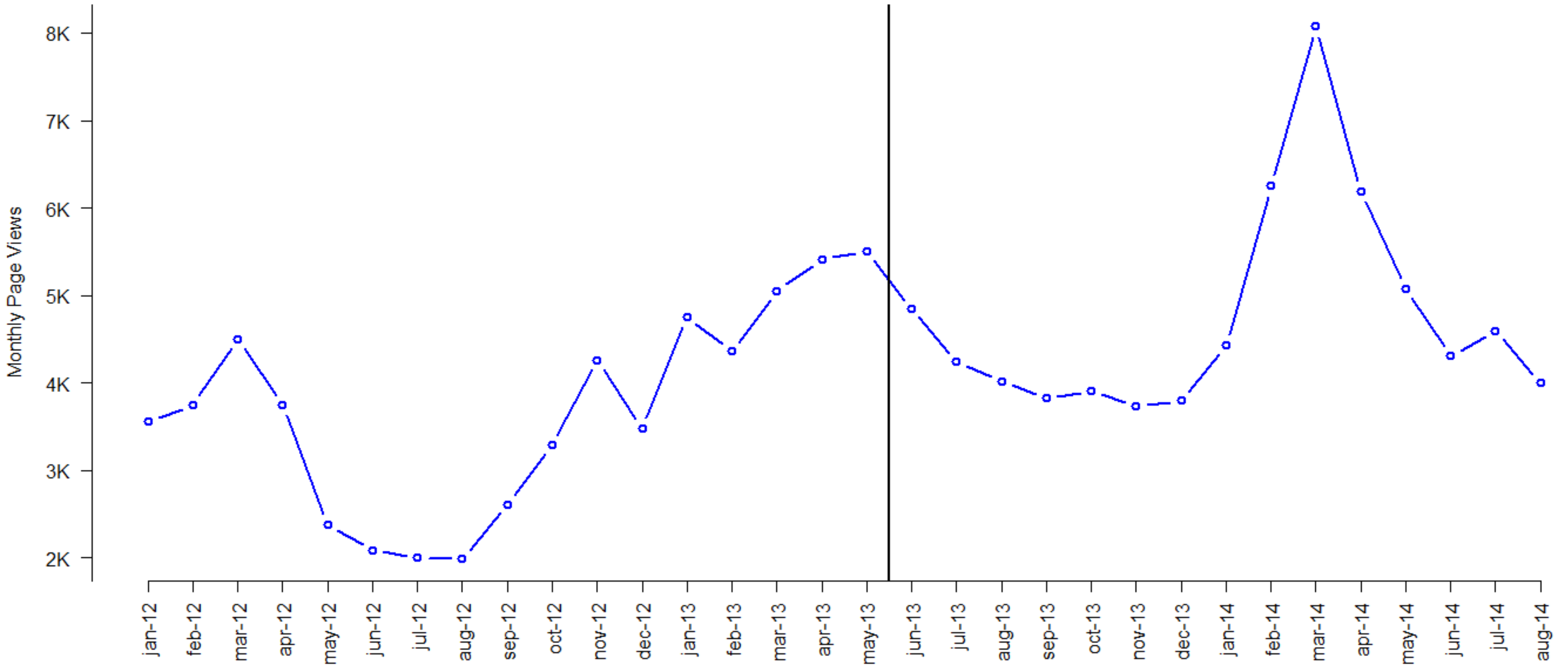


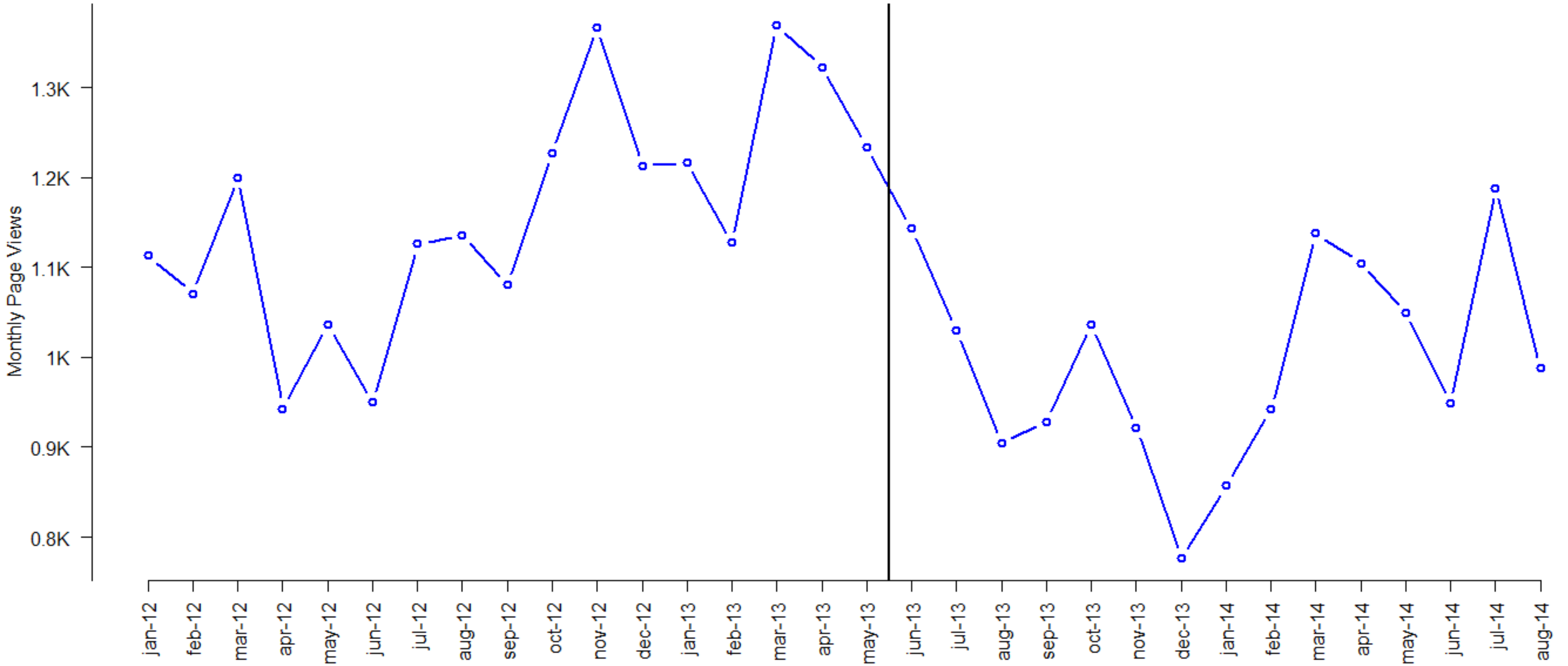




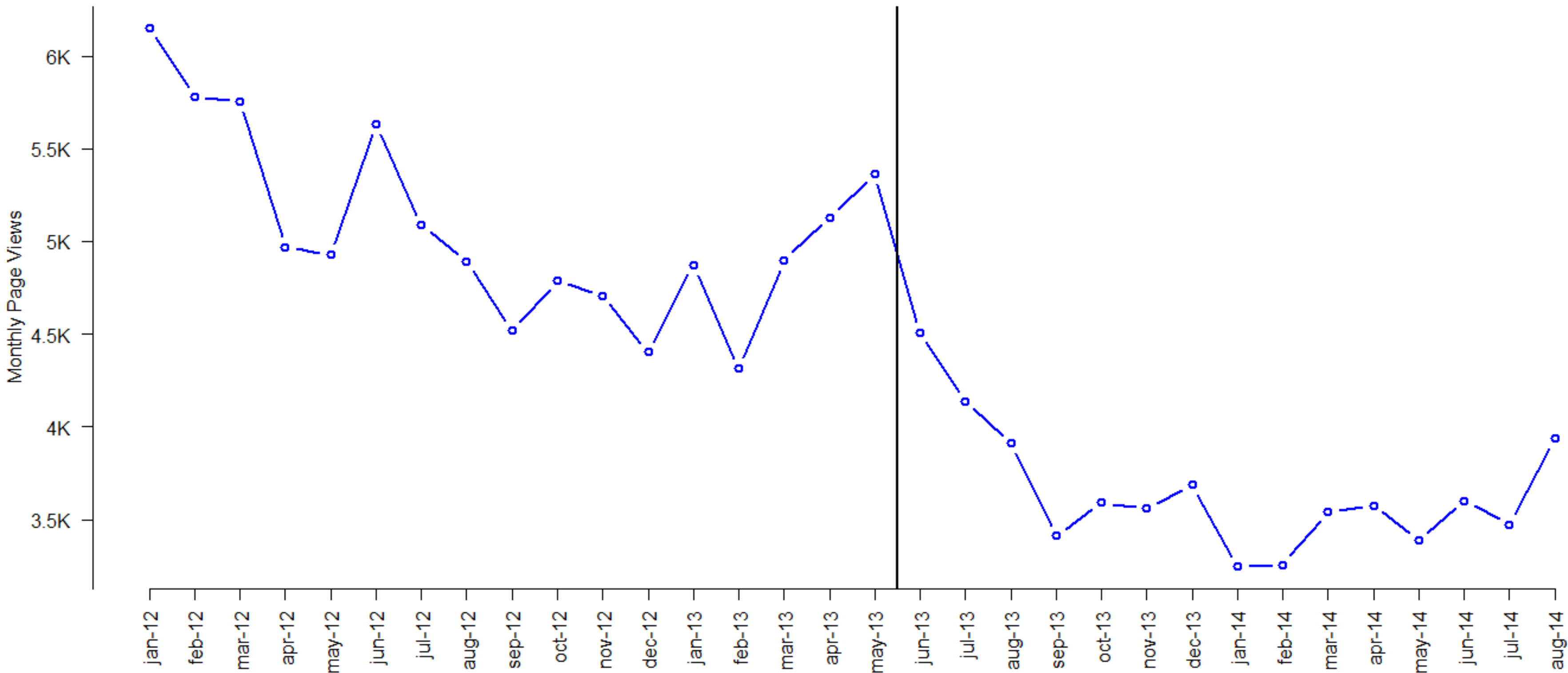
### Infrastructure: Page Views for full\_body\_scanner



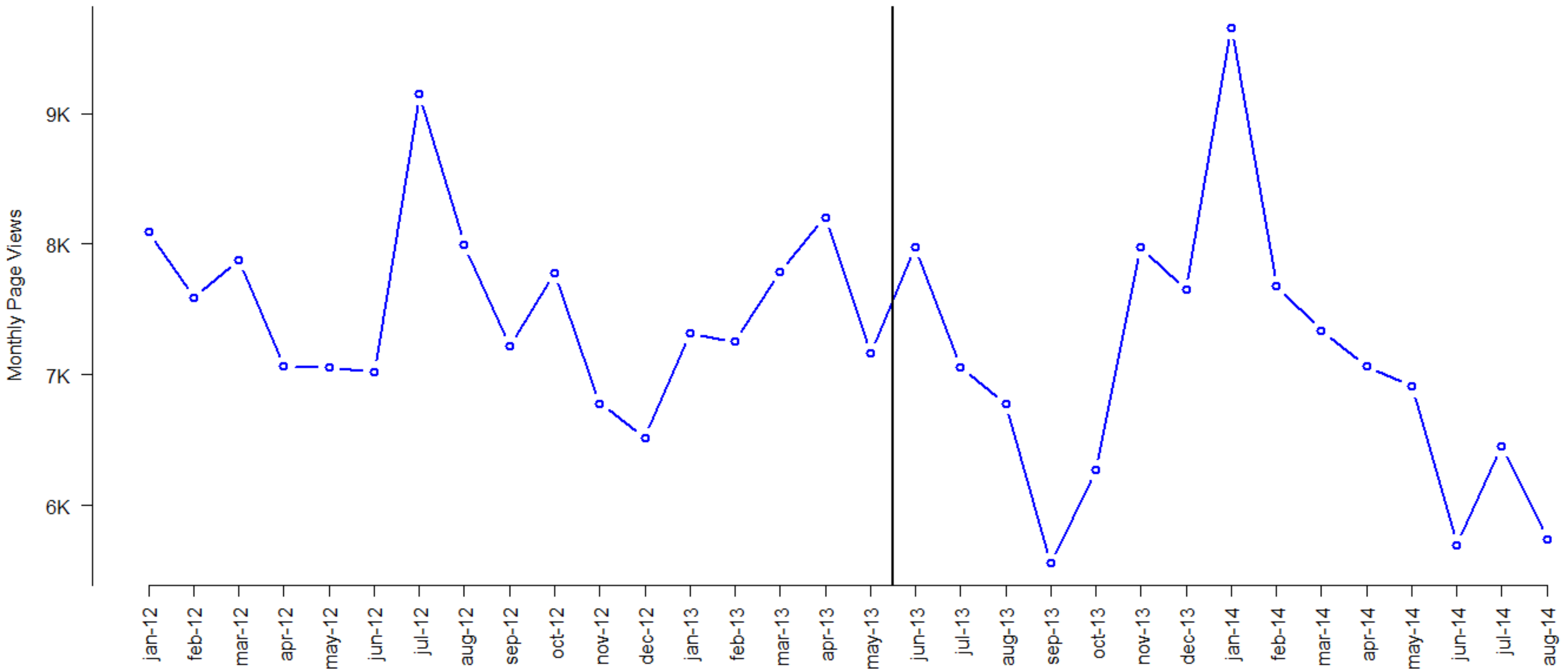




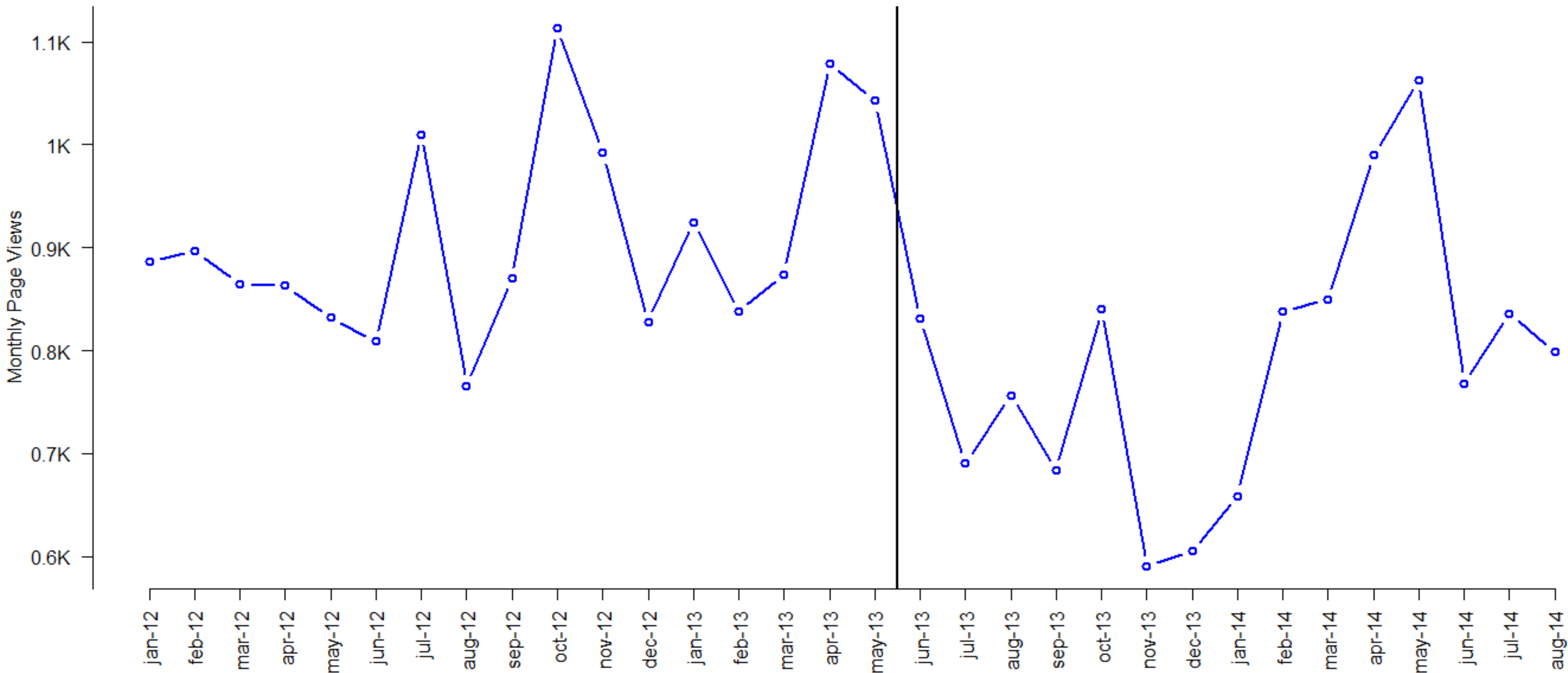
### Infrastructure: Page Views for metro\_station



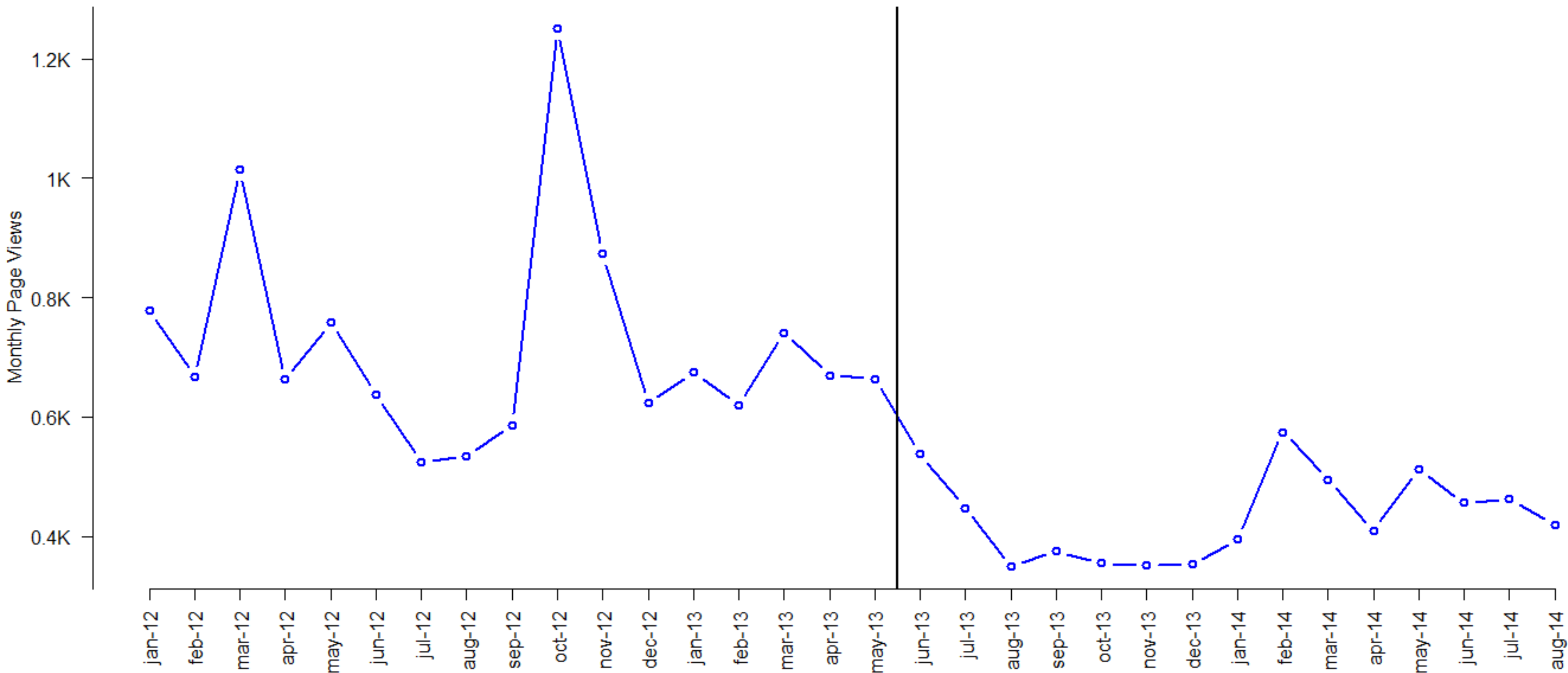
### Infrastructure: Page Views for Metropolitan\_Atlanta\_Rapid\_Trans



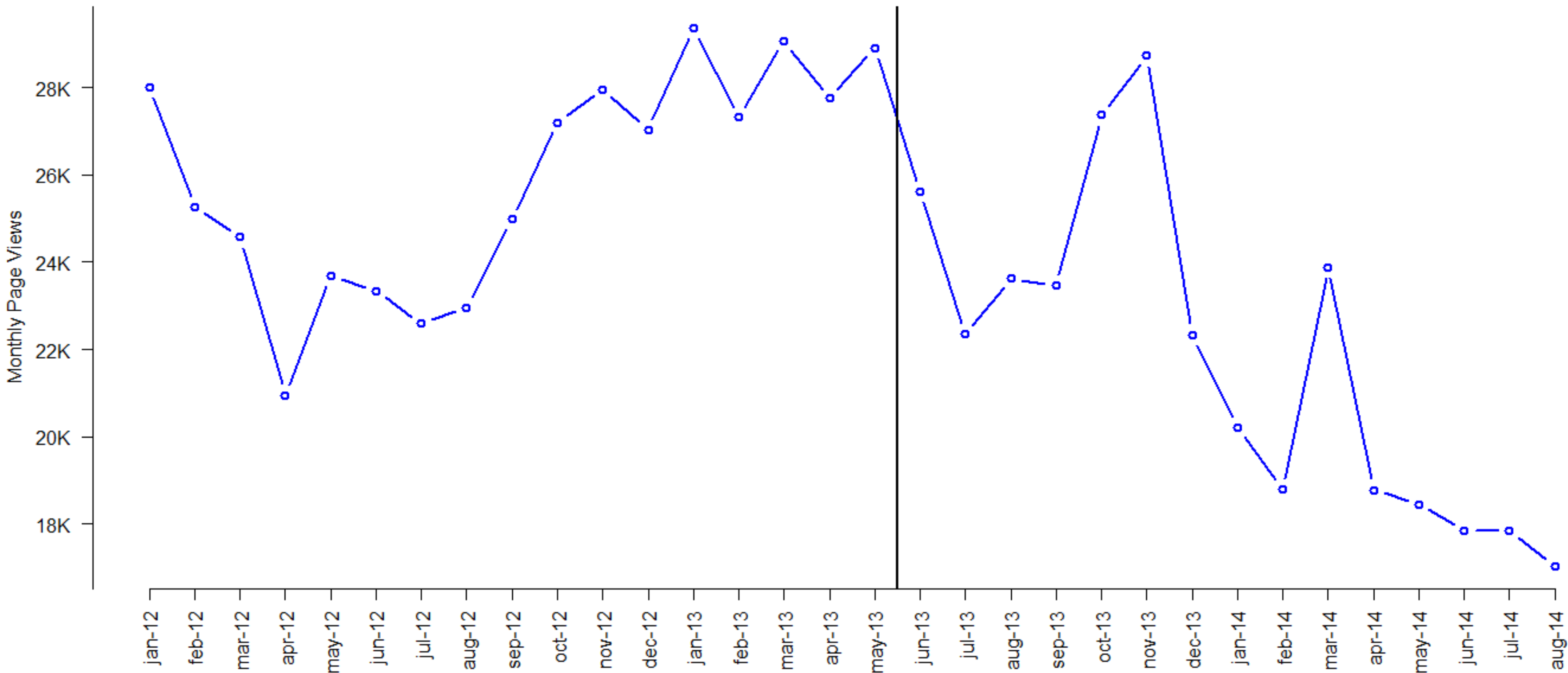
### Infrastructure: Page Views for national\_information\_infrastruct



### Infrastructure: Page Views for nbc

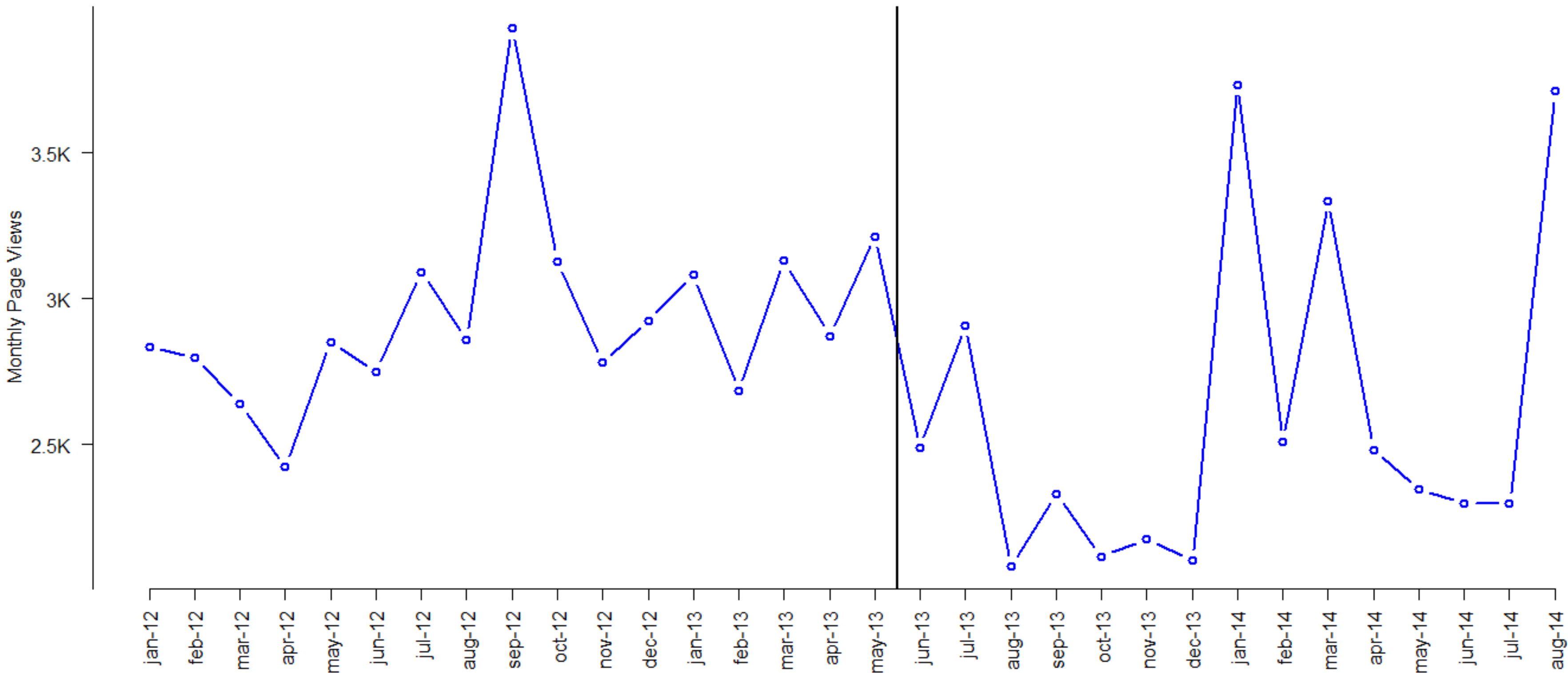


### Infrastructure: Page Views for port

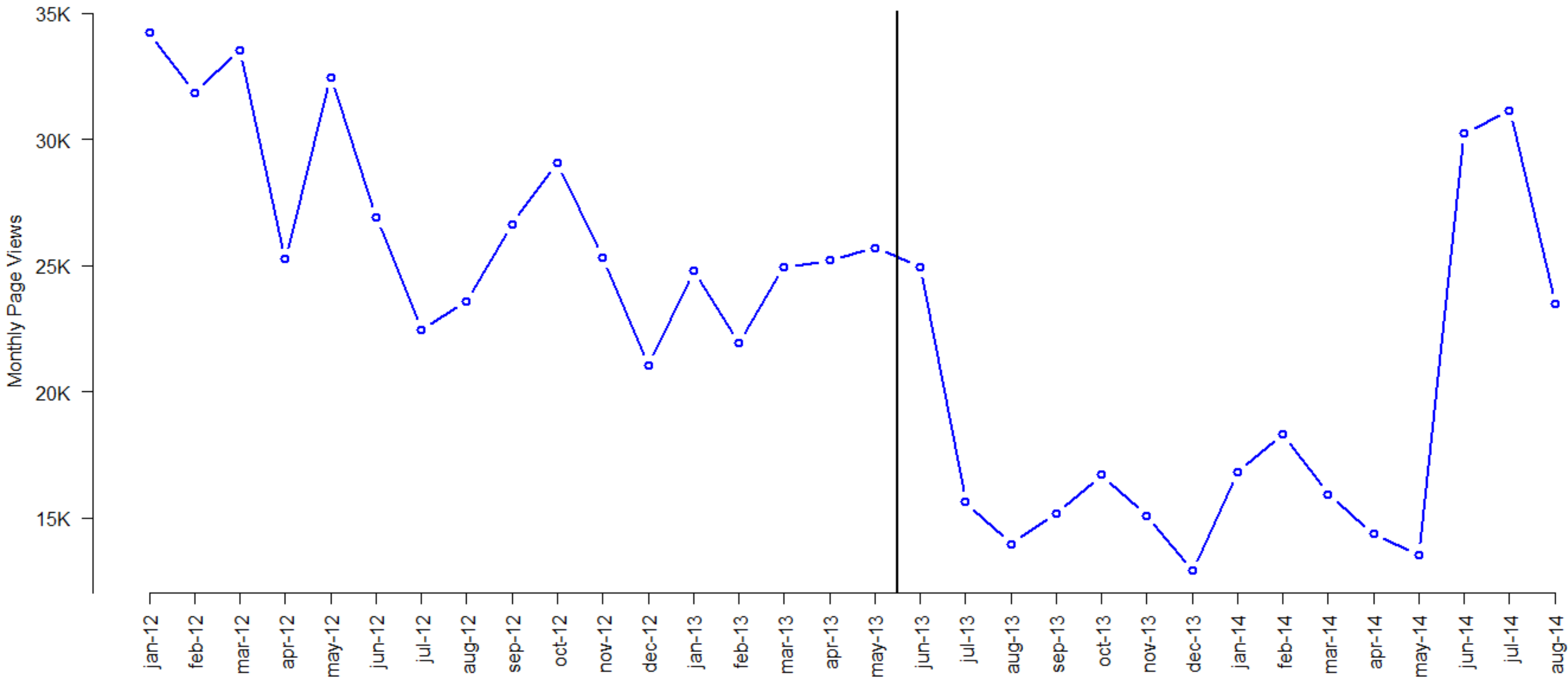




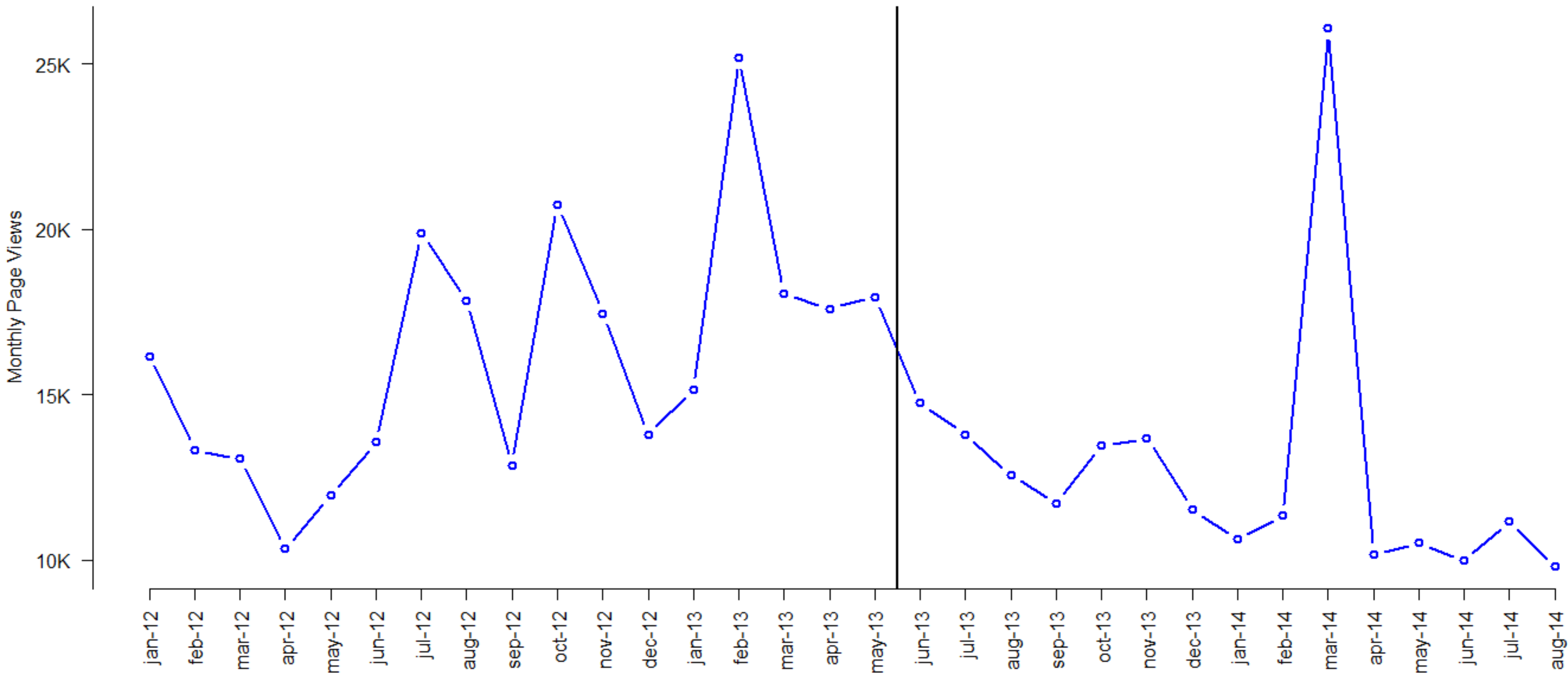
### Infrastructure: Page Views for Port\_authority



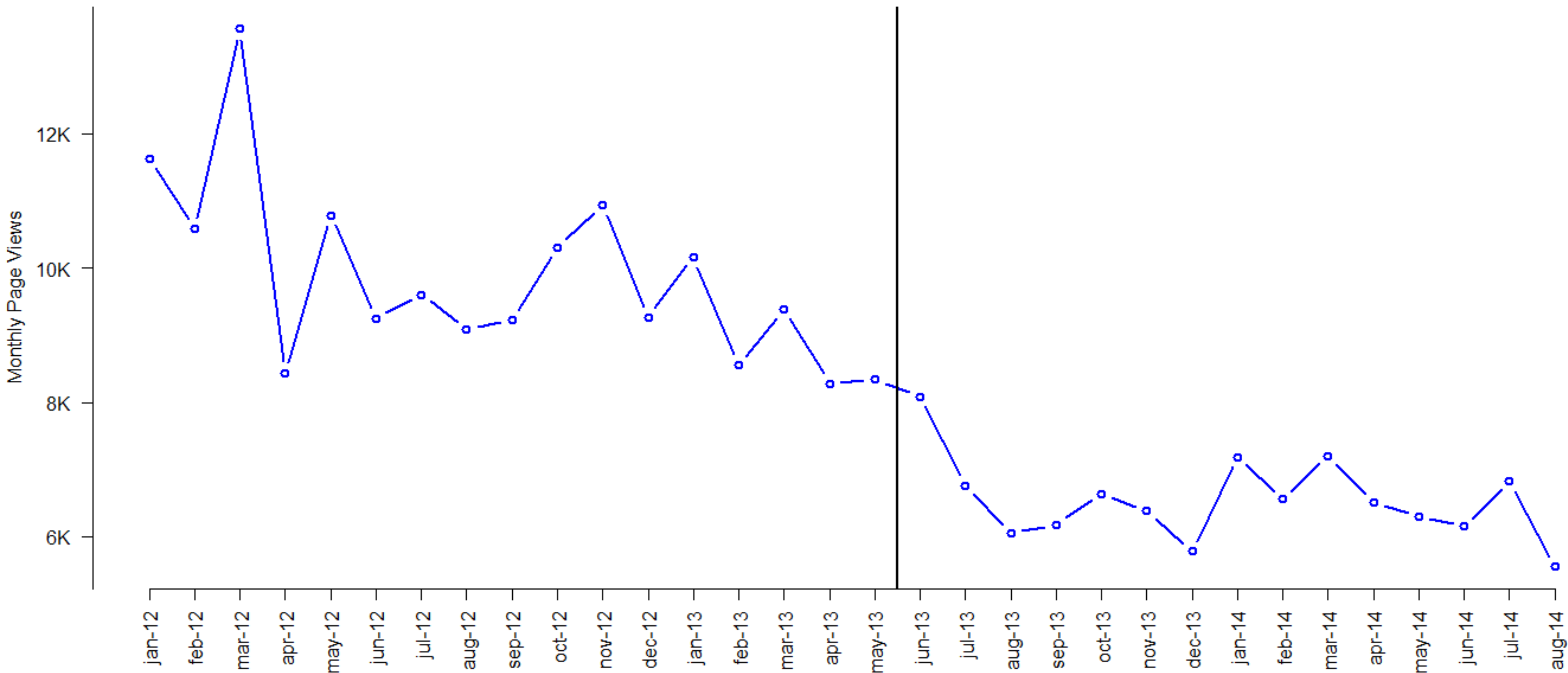
### Infrastructure: Page Views for power



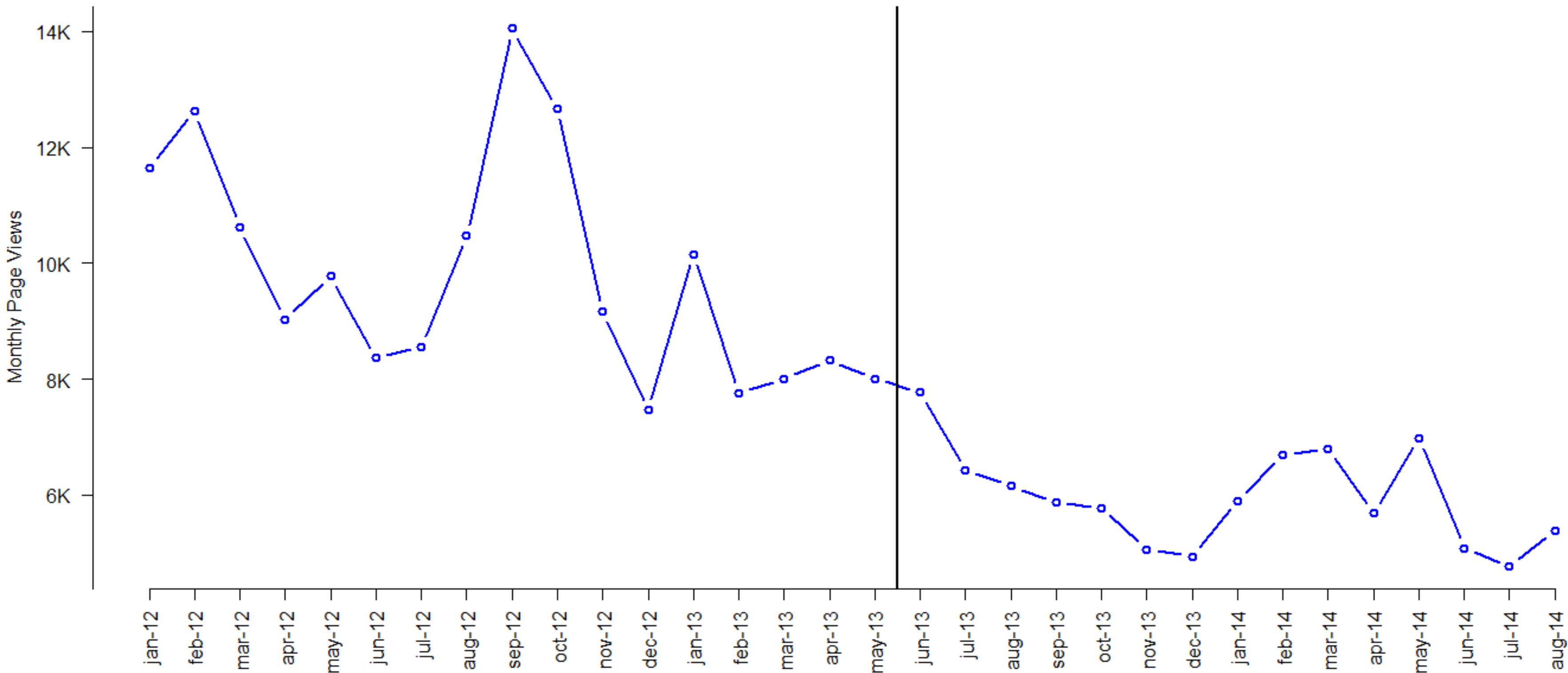
### Infrastructure: Page Views for power\_outage



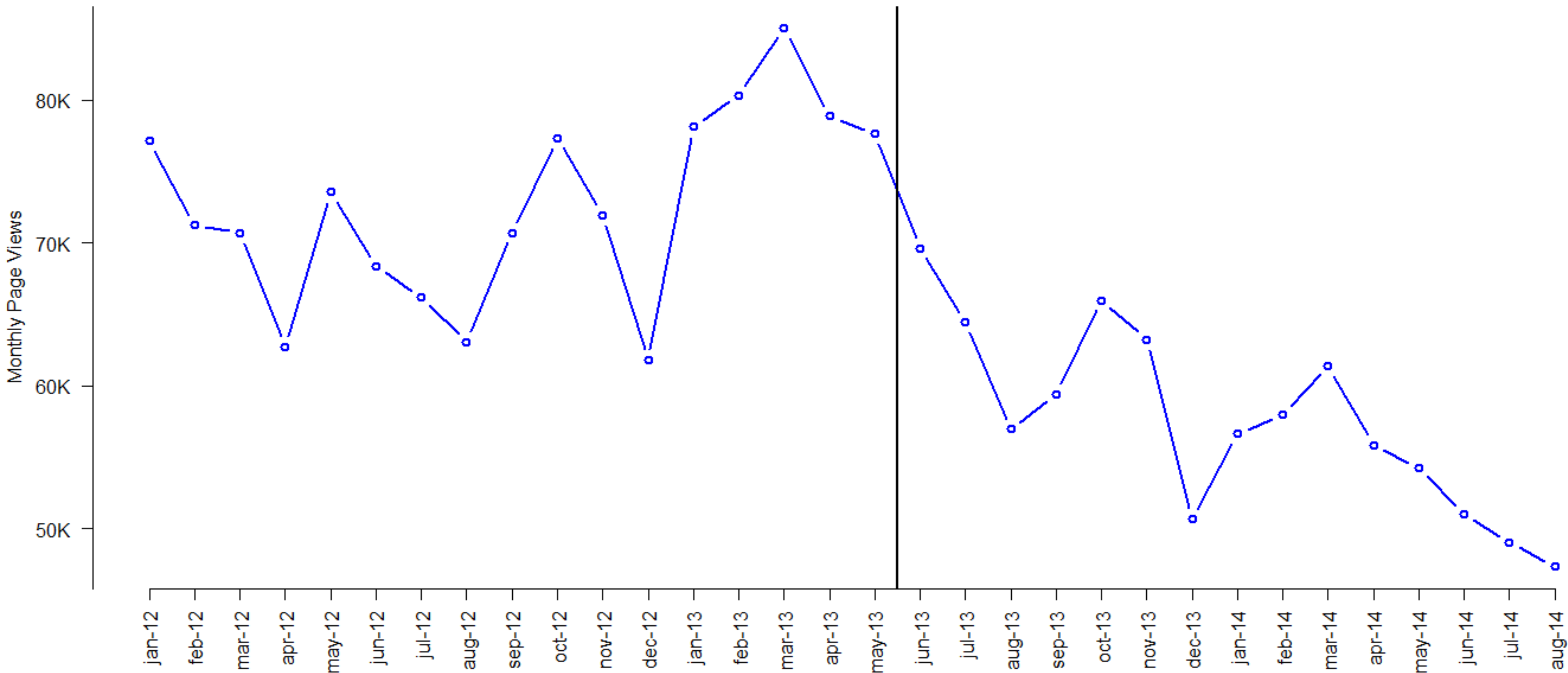
### Infrastructure: Page Views for smart



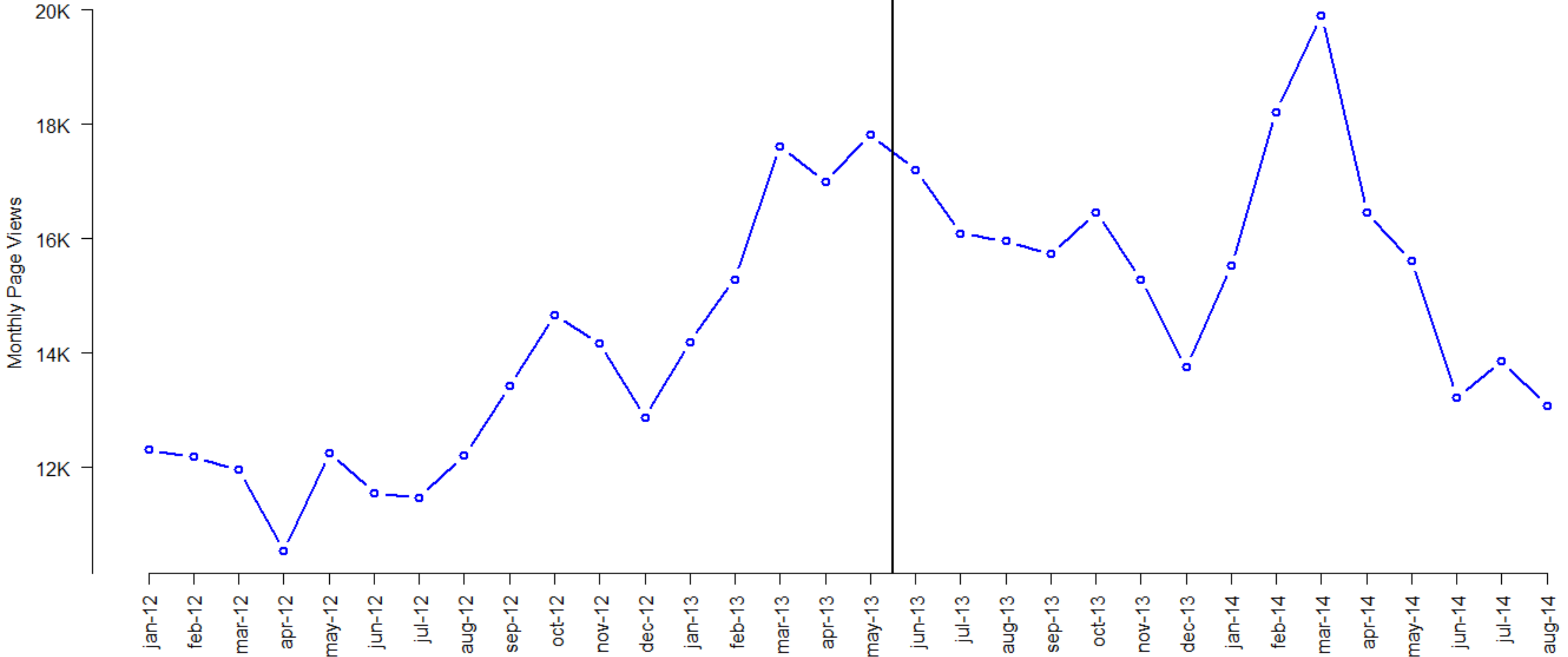
### Infrastructure: Page Views for subway



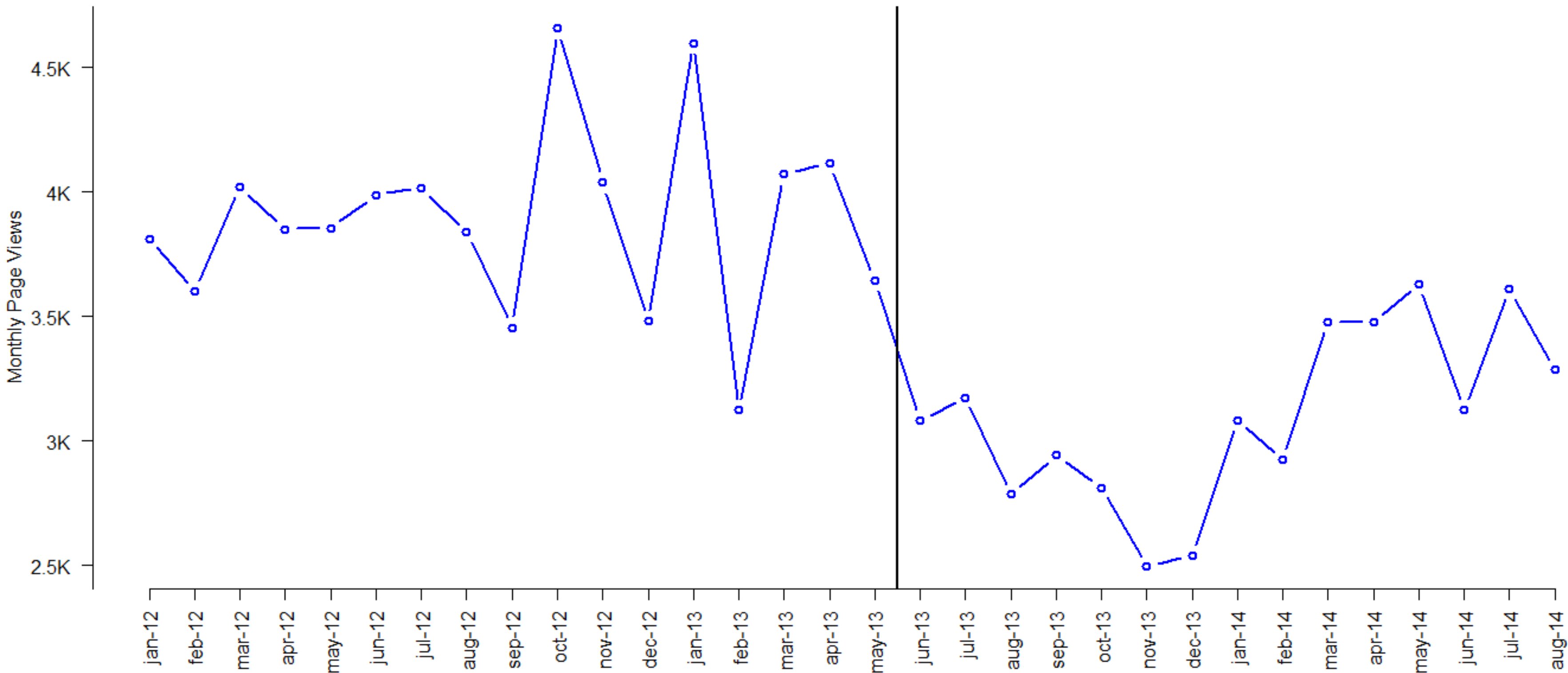
### Infrastructure: Page Views for telecommunication



### Infrastructure: Page Views for Telecommunications\_network

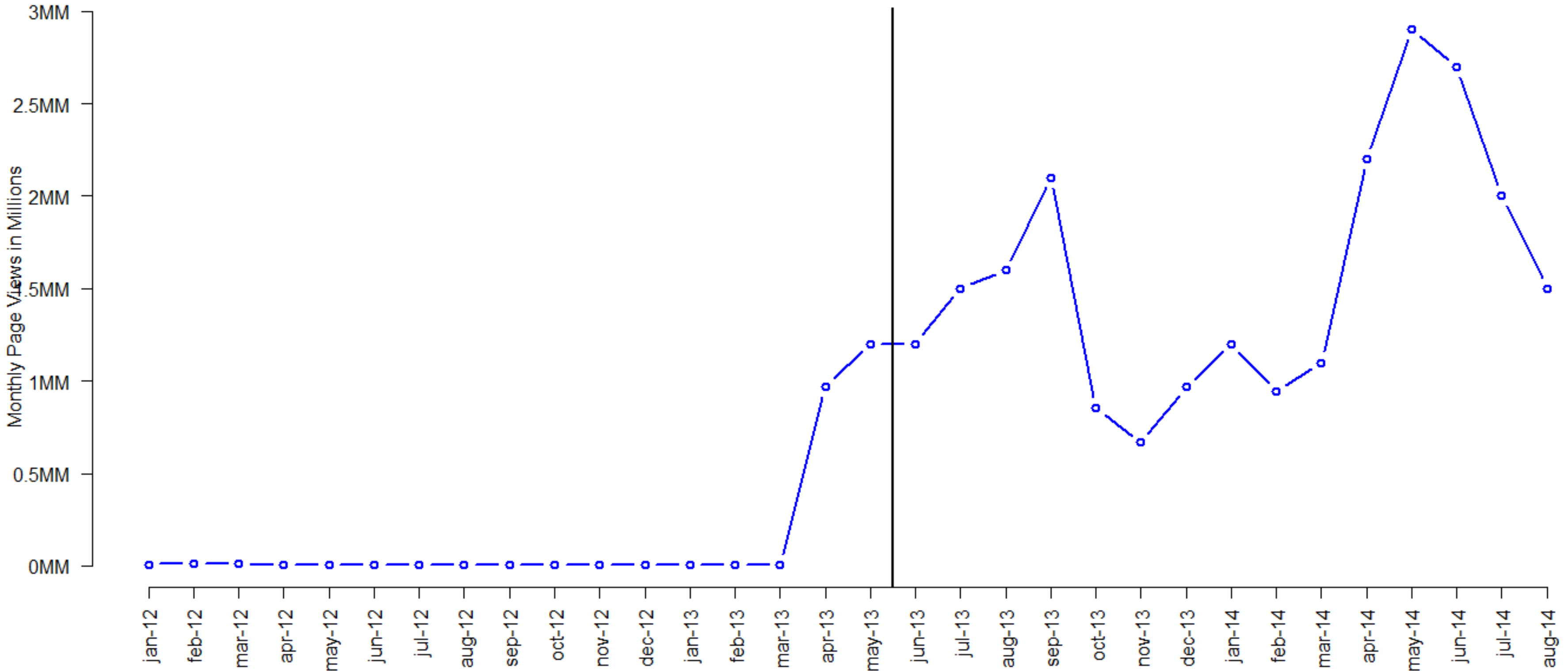


### Infrastructure: Page Views for washington\_metropolitan\_area\_tra

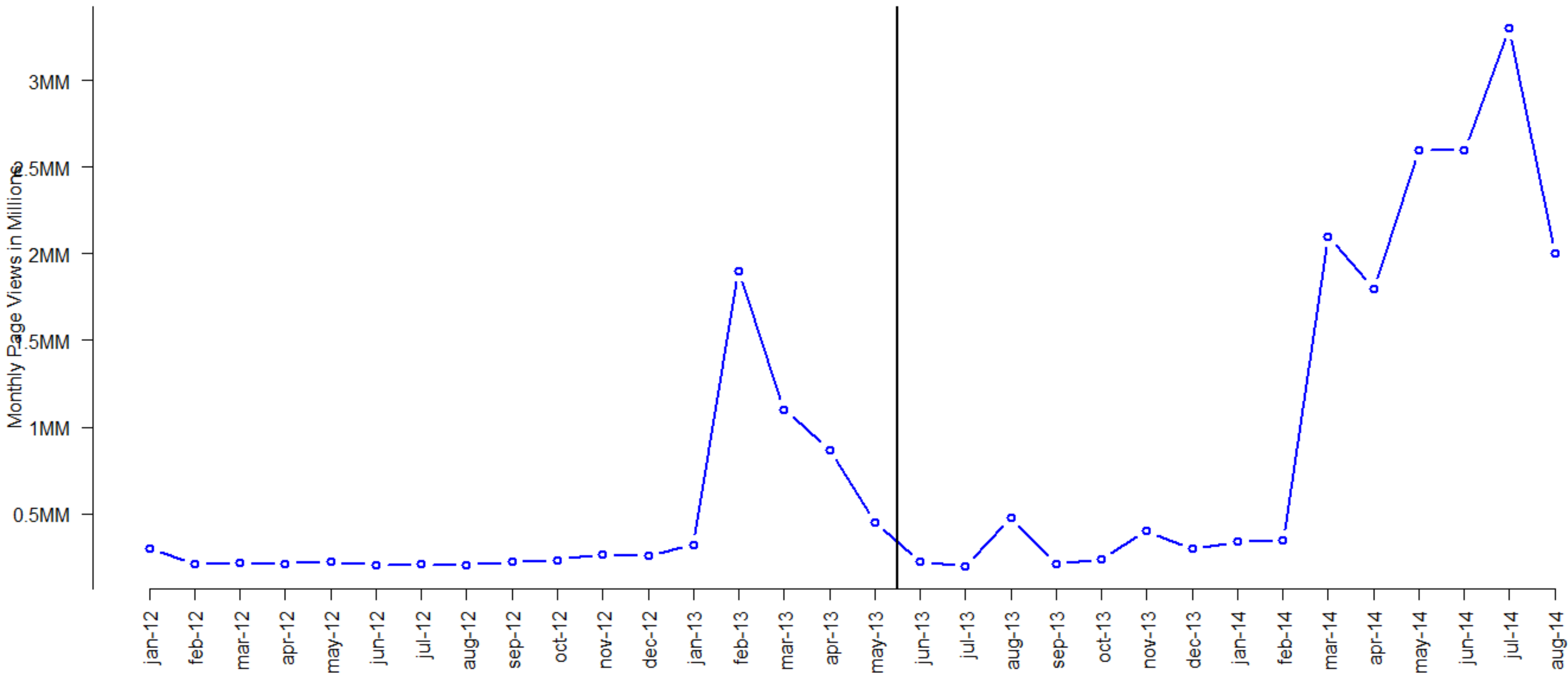




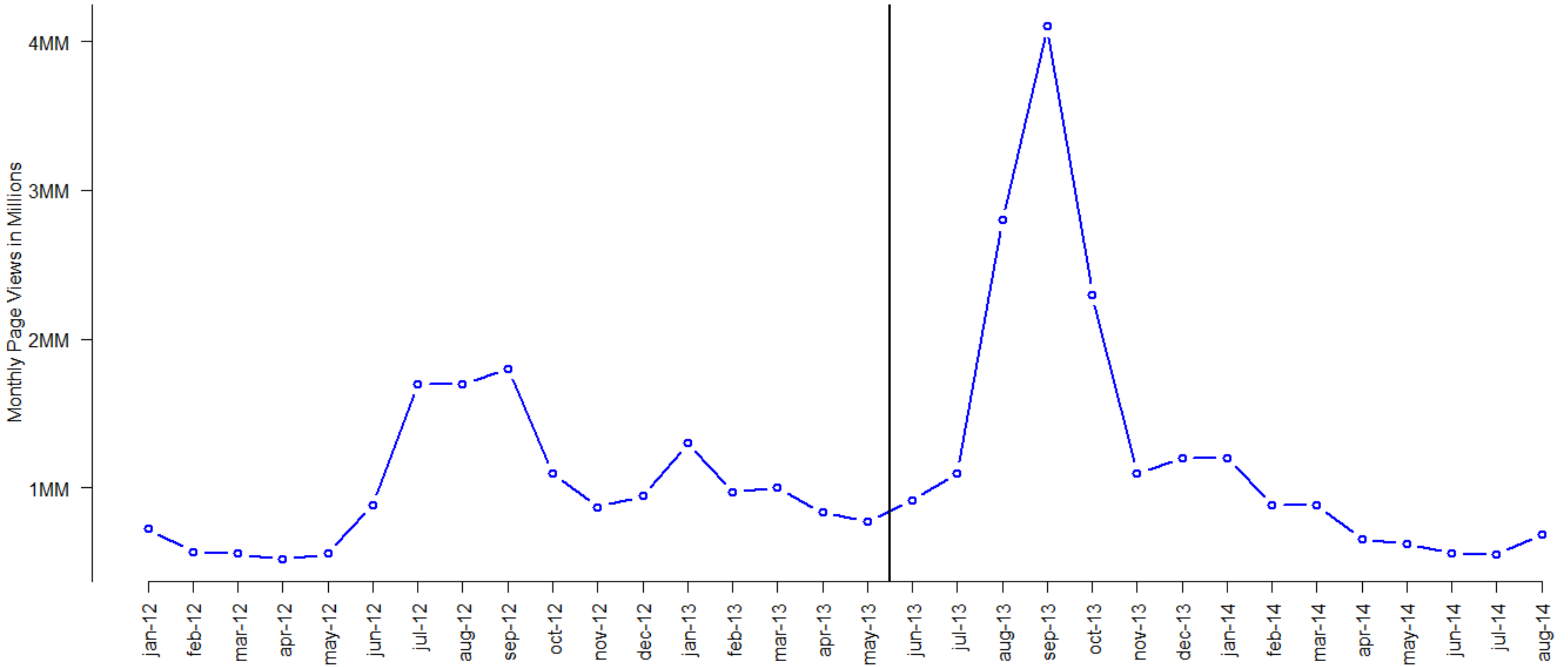
### Popular: Page Views for alive



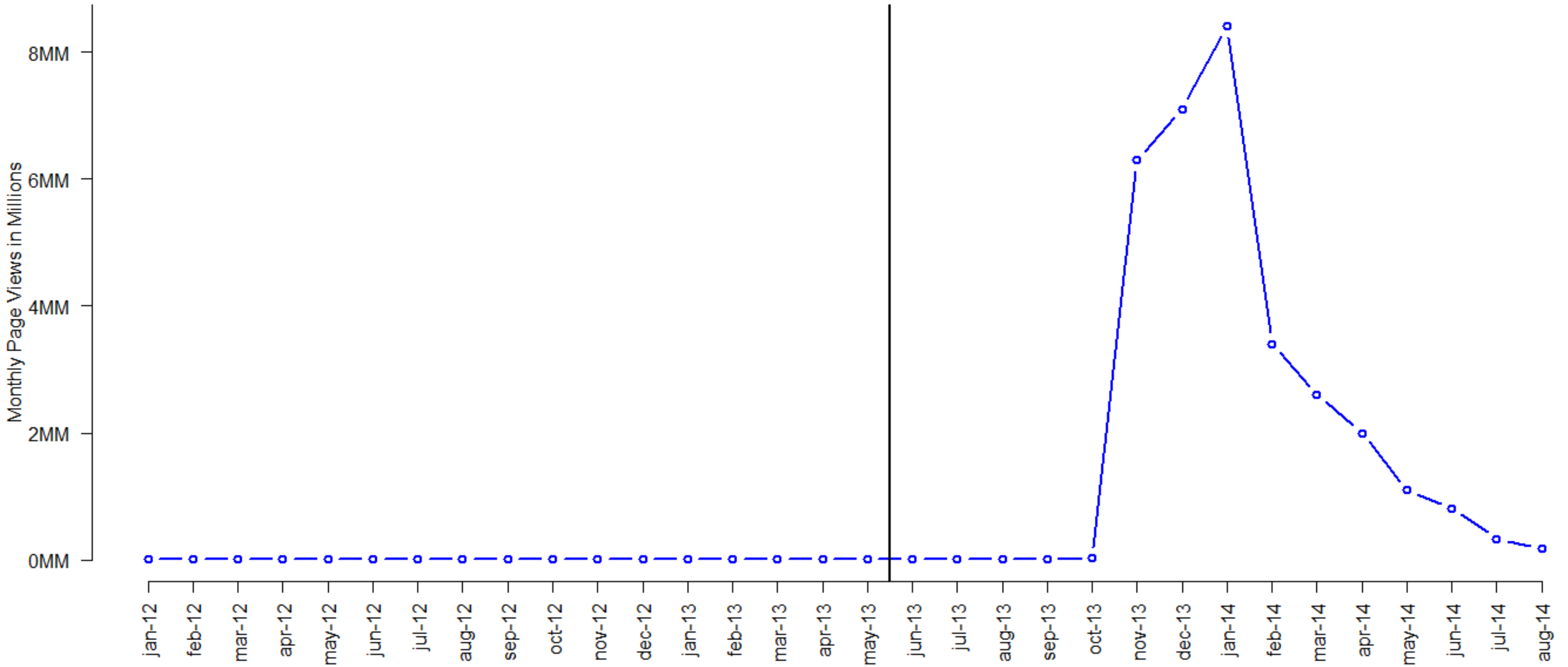
### Popular: Page Views for amazoncom



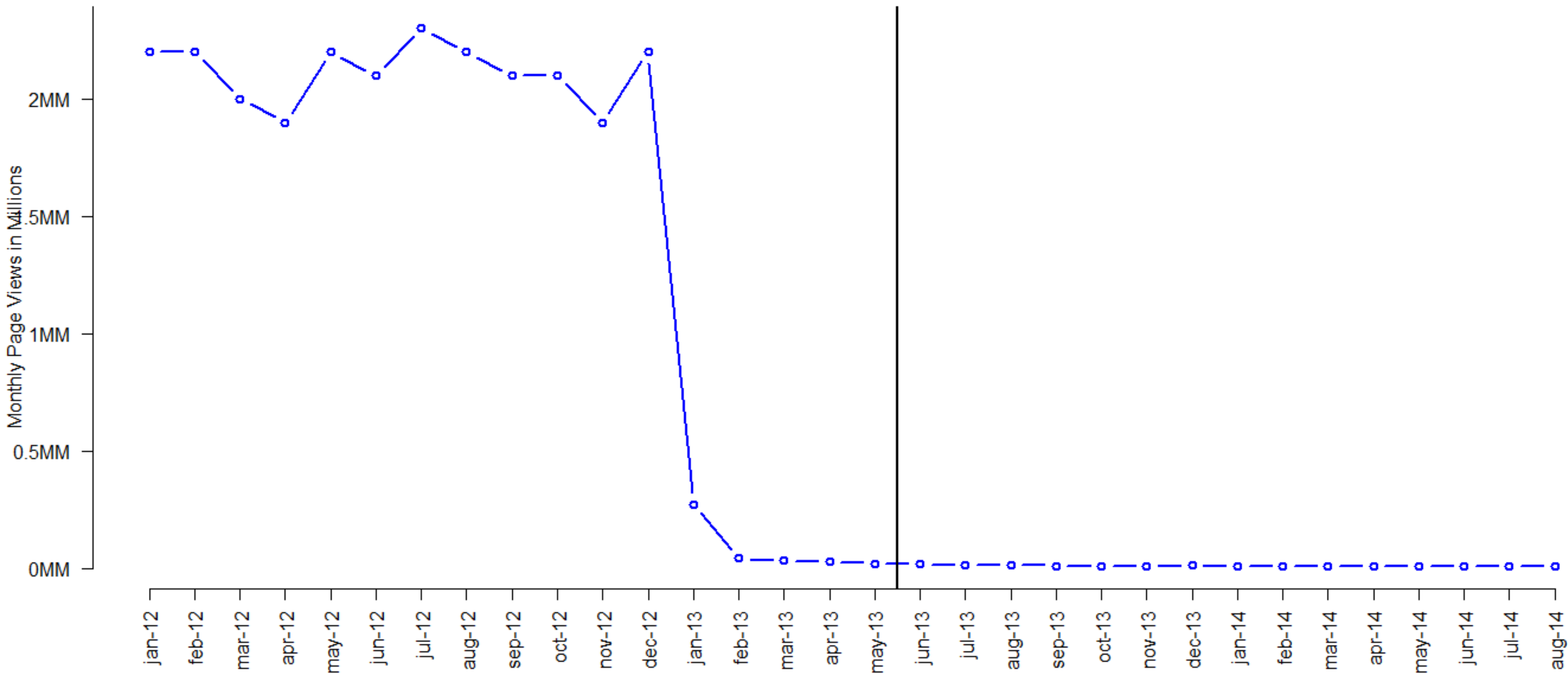
### Popular: Page Views for breaking\_bad



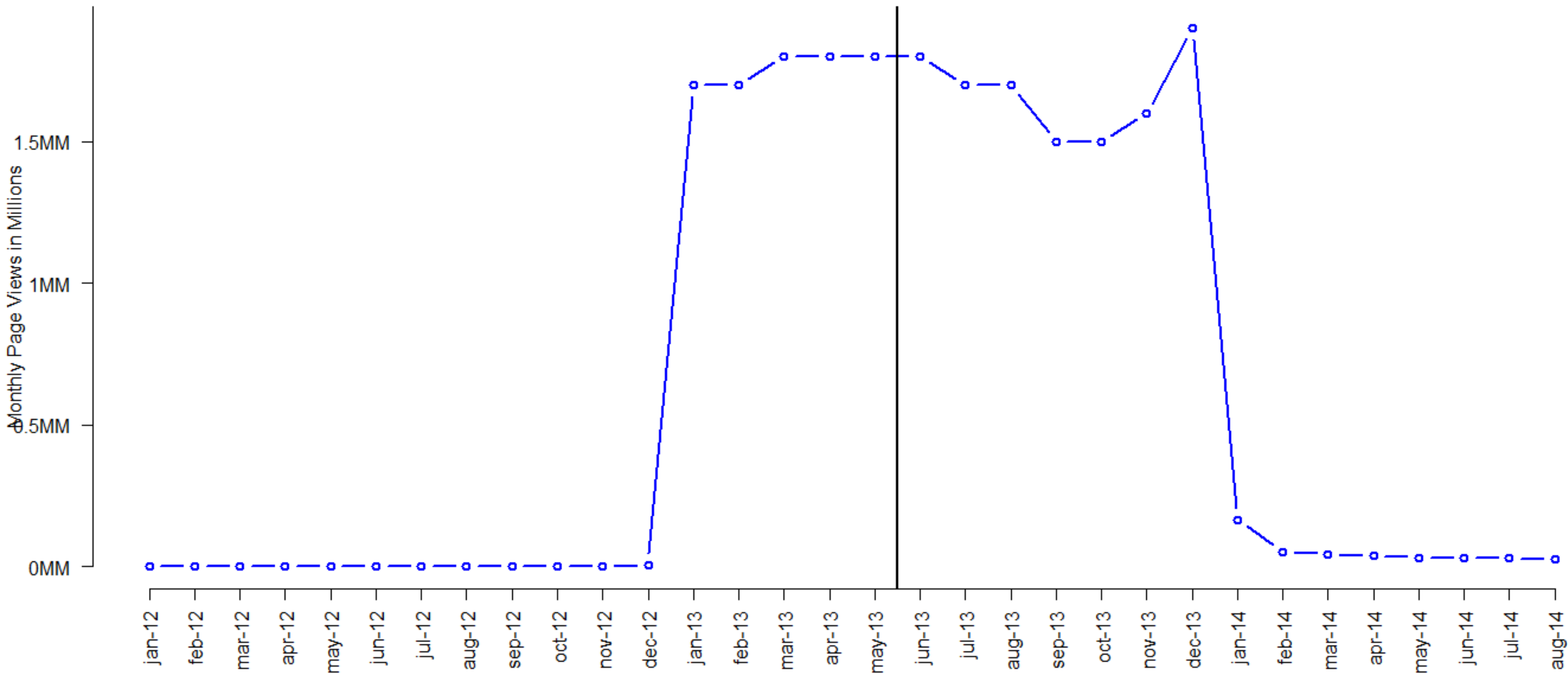
### Popular: Page Views for climatic\_research\_unit\_email\_con



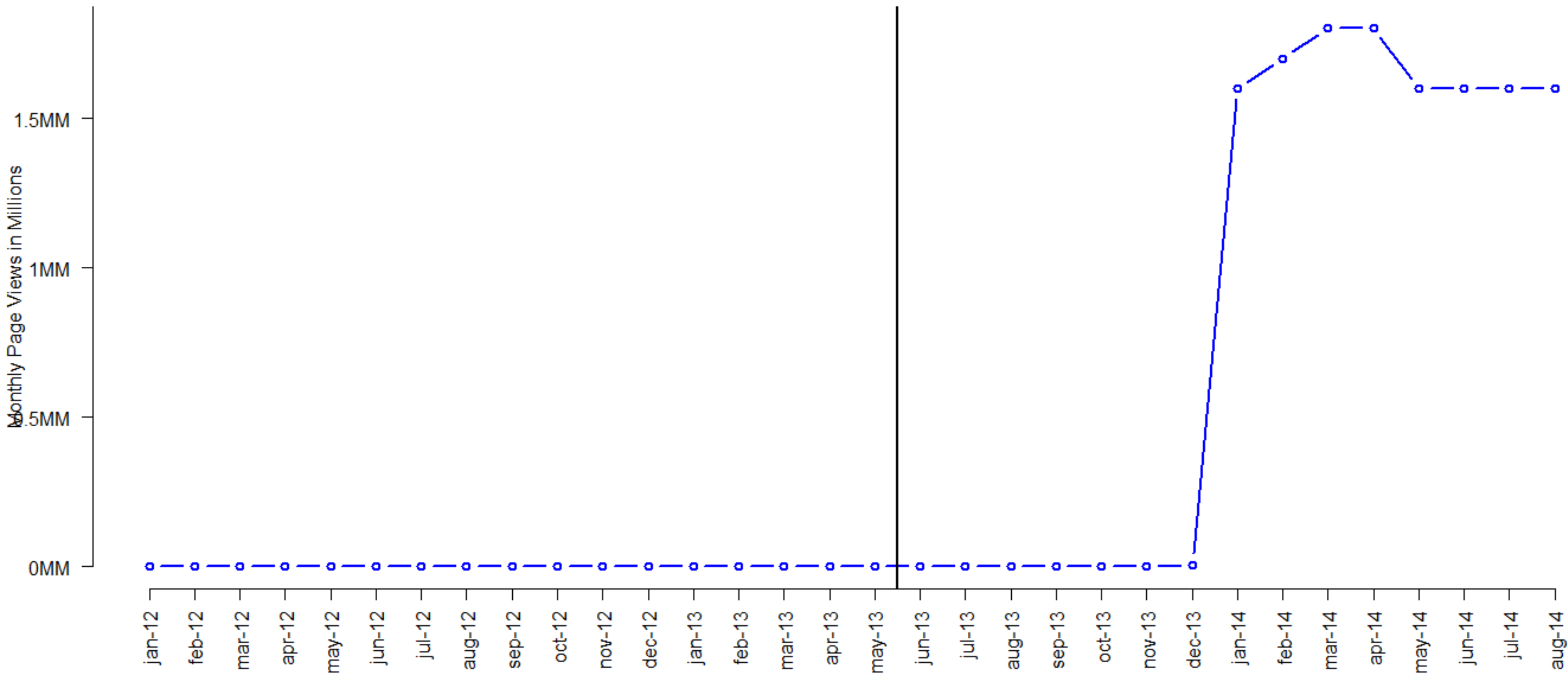
### Popular: Page Views for deaths\_in\_2012



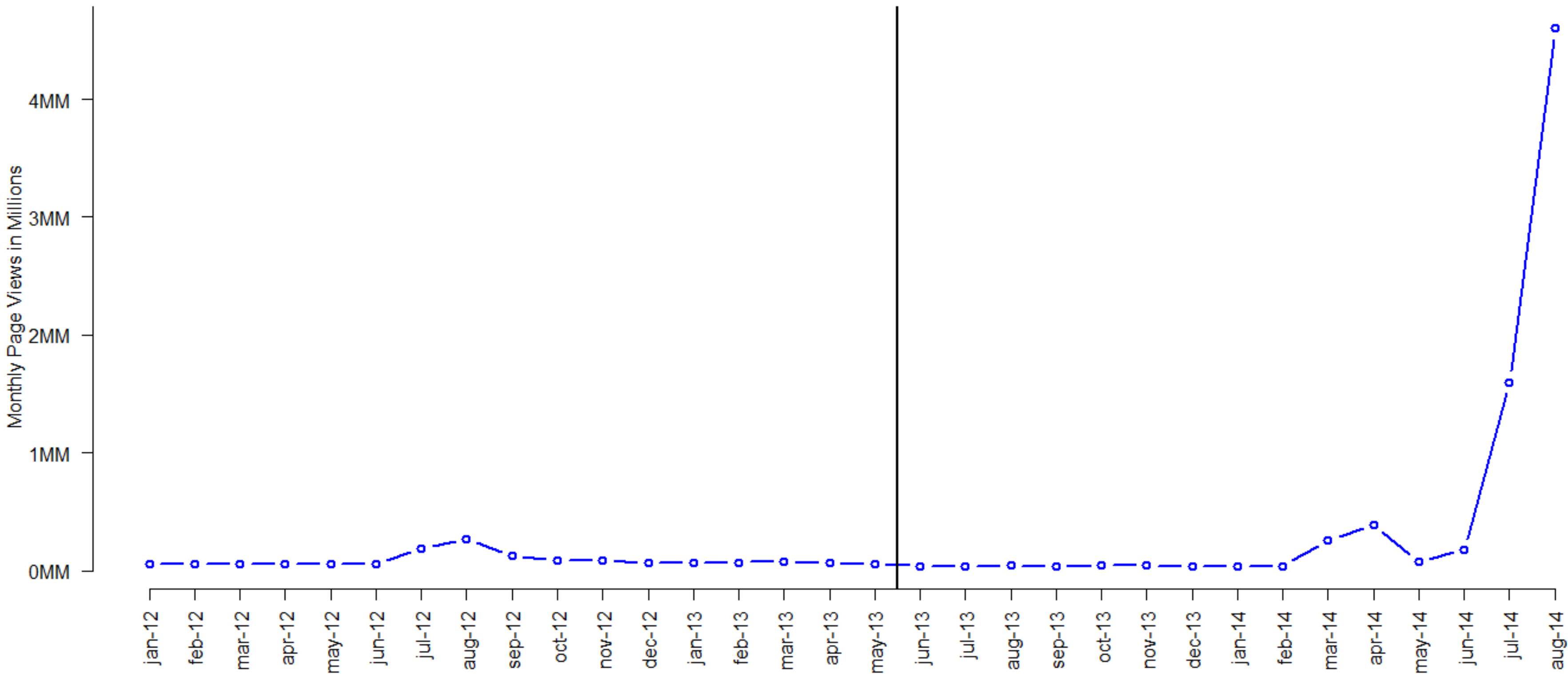
### Popular: Page Views for deaths\_in\_2013



### Popular: Page Views for deaths\_in\_2014

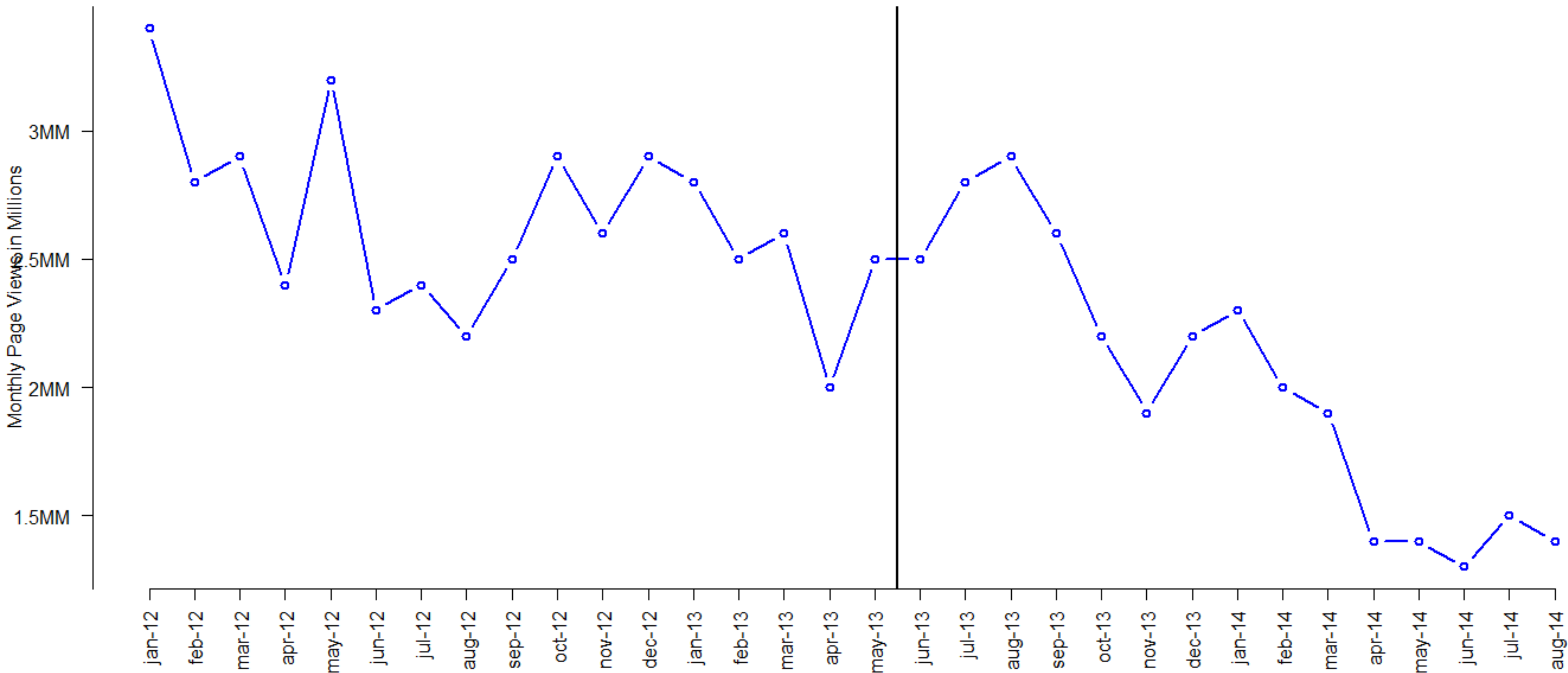


# Popular: Page Views for ebola\_virus\_disease\_

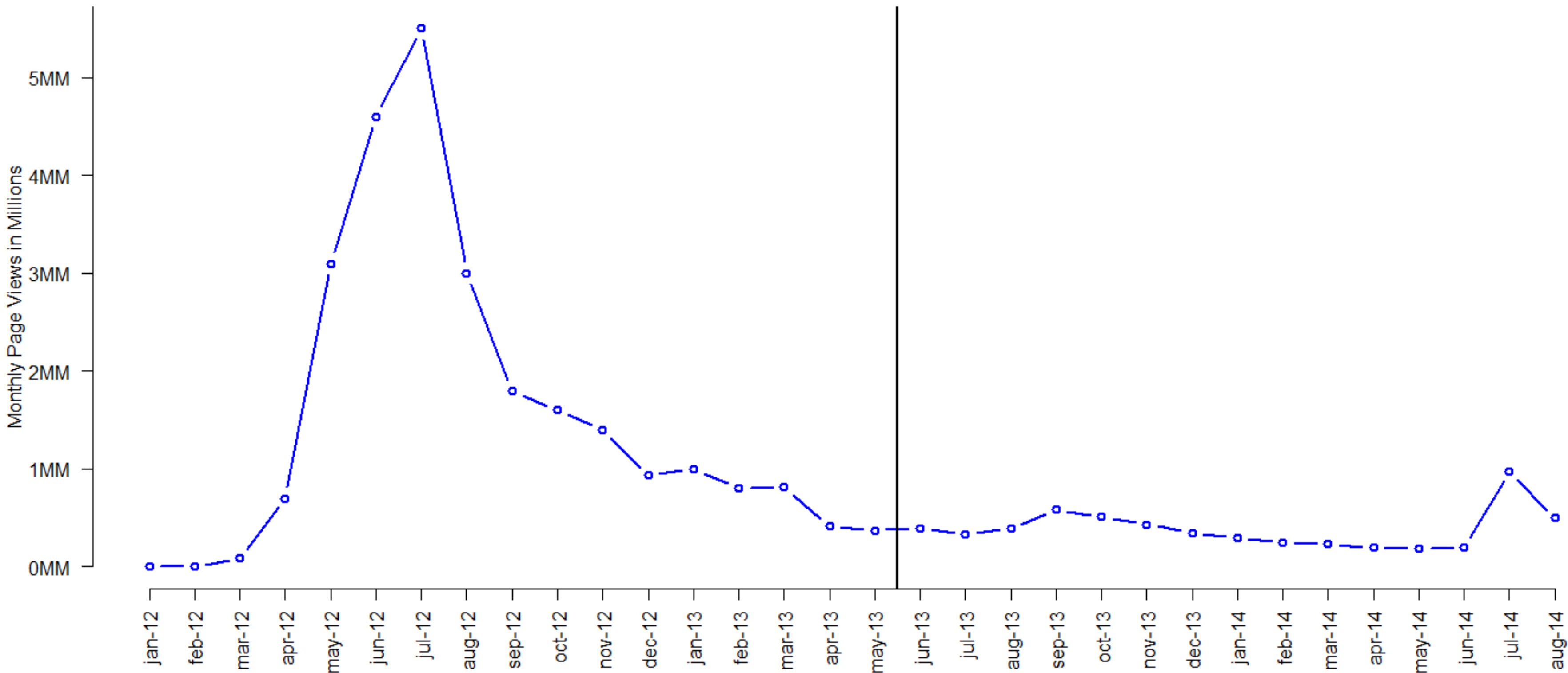




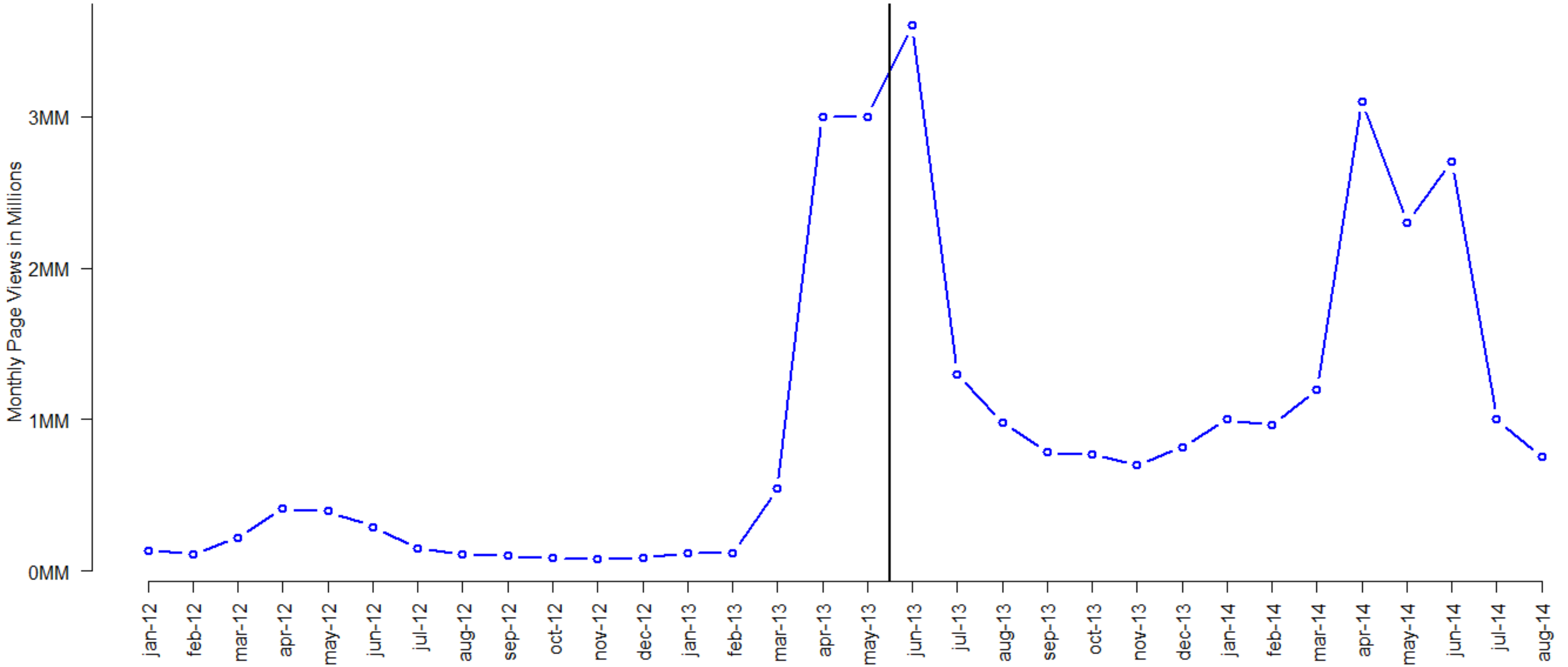
### Popular: Page Views for facebook



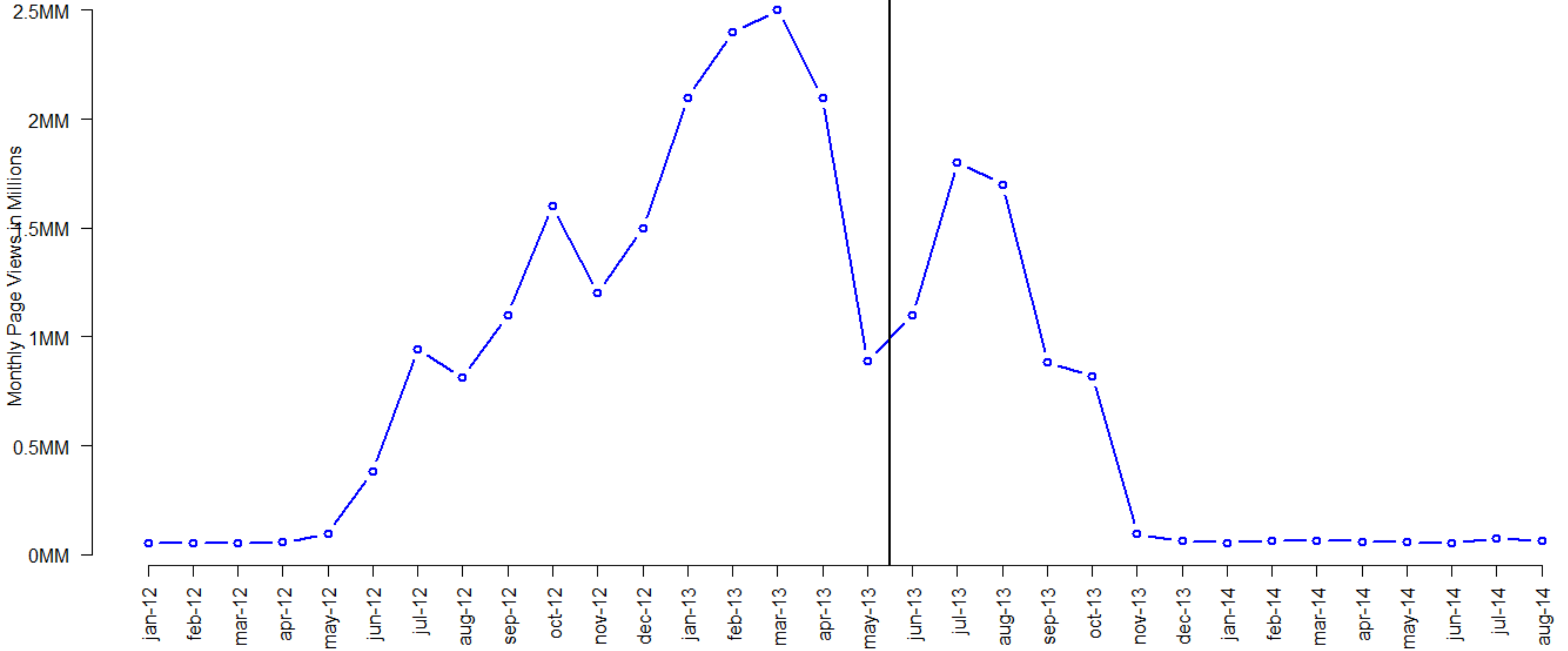
### Popular: Page Views for fifty\_shades\_of\_grey



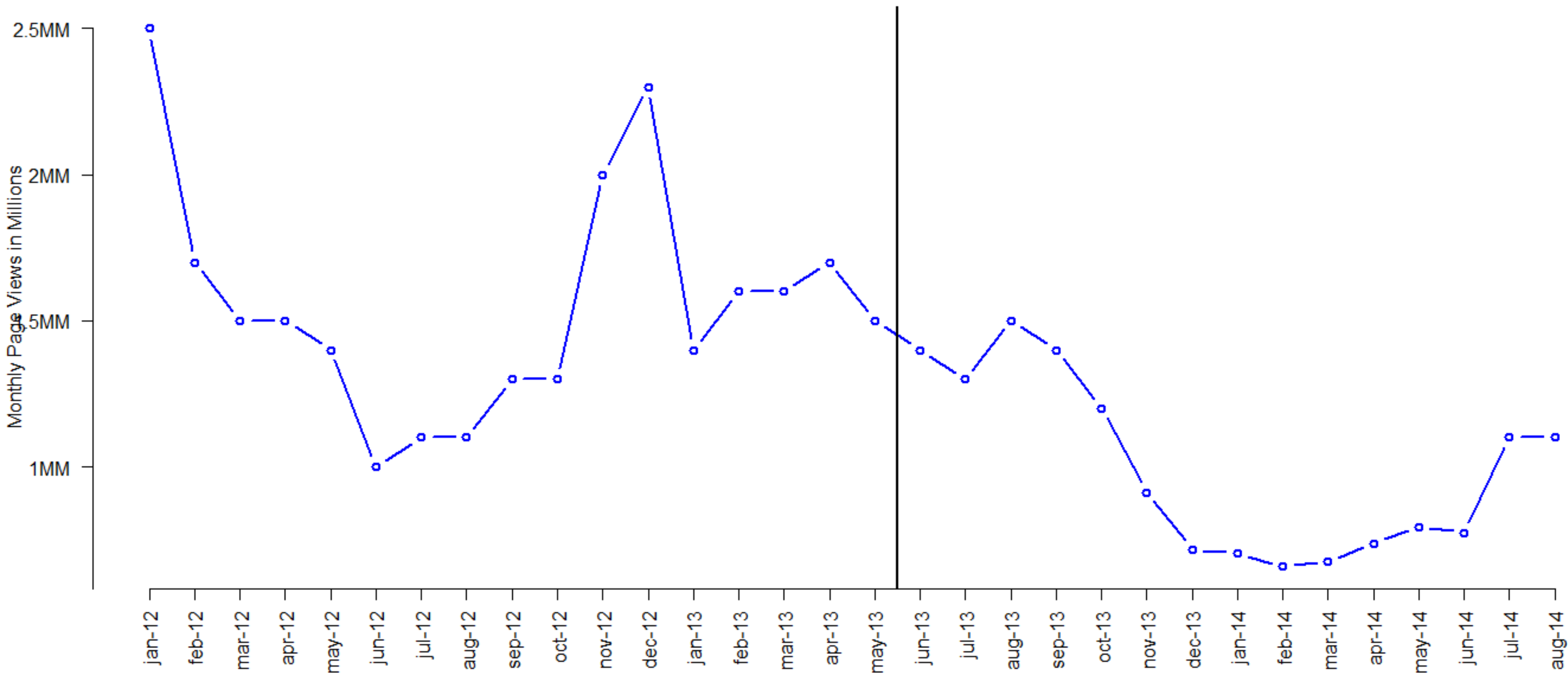
### Popular: Page Views for game\_of\_thrones



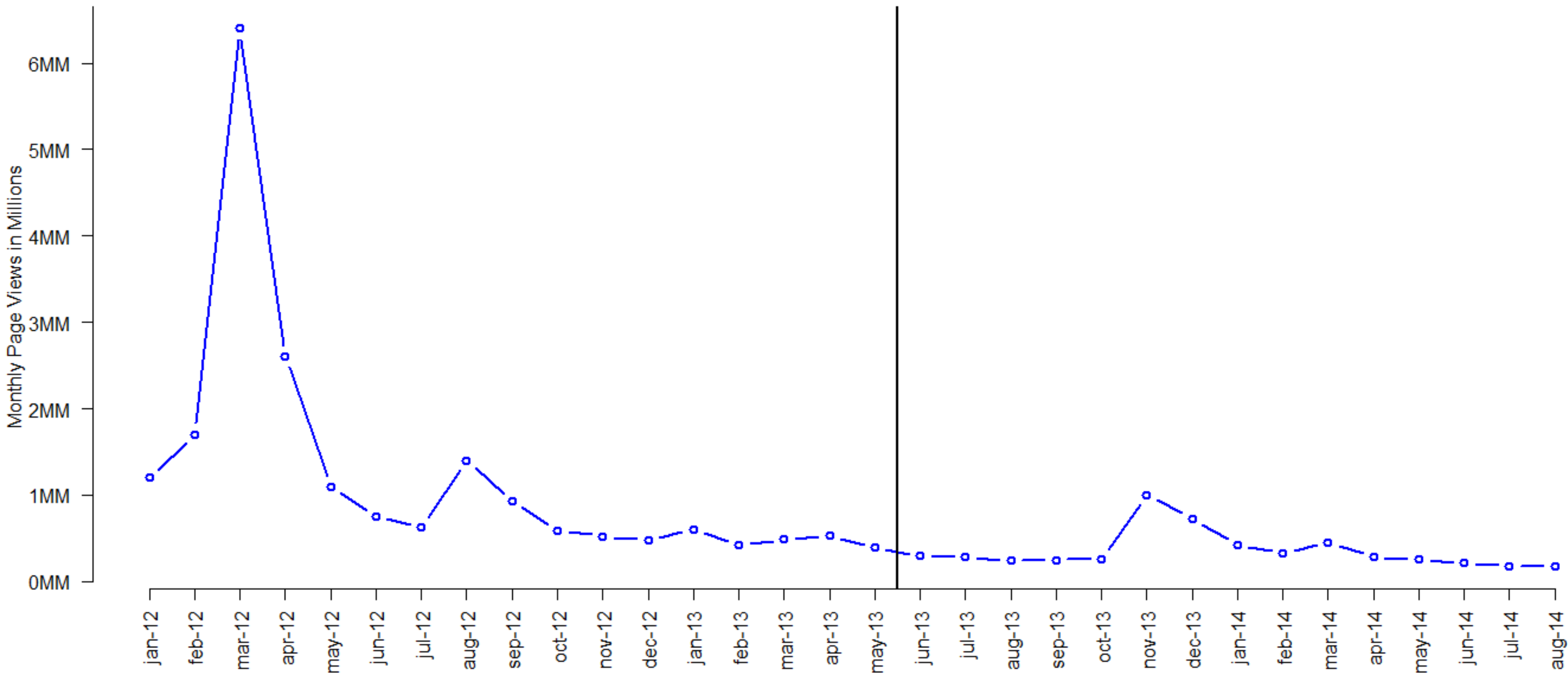
### Popular: Page Views for gforce



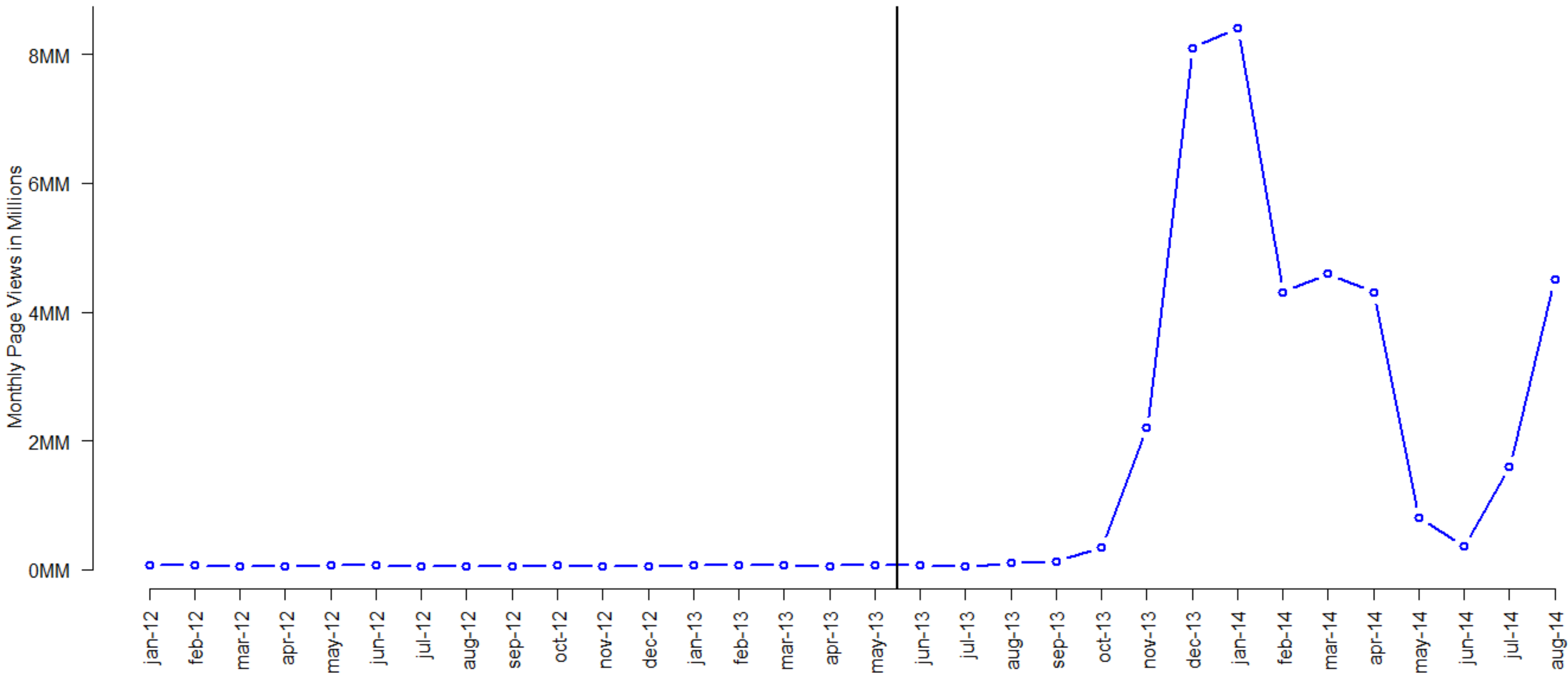
### Popular: Page Views for google

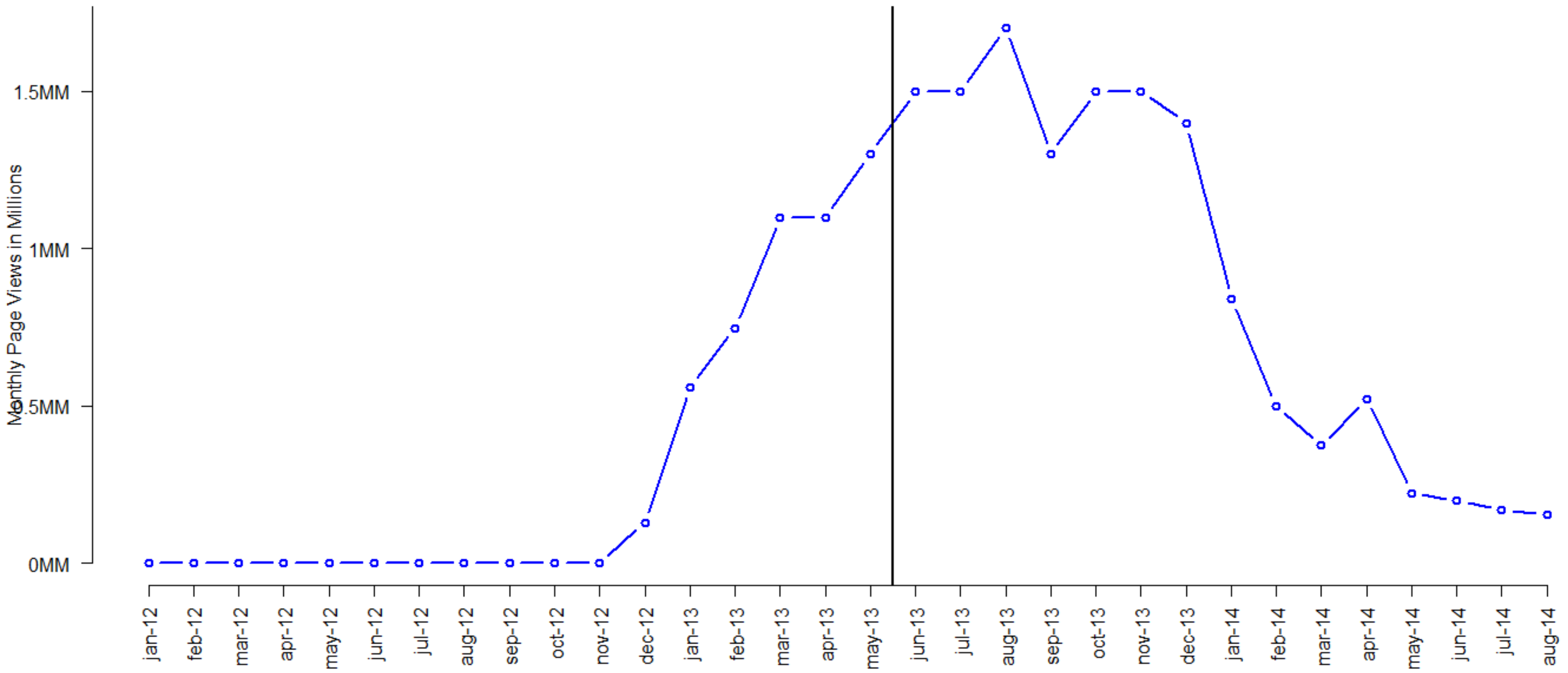


### Popular: Page Views for hunger\_games



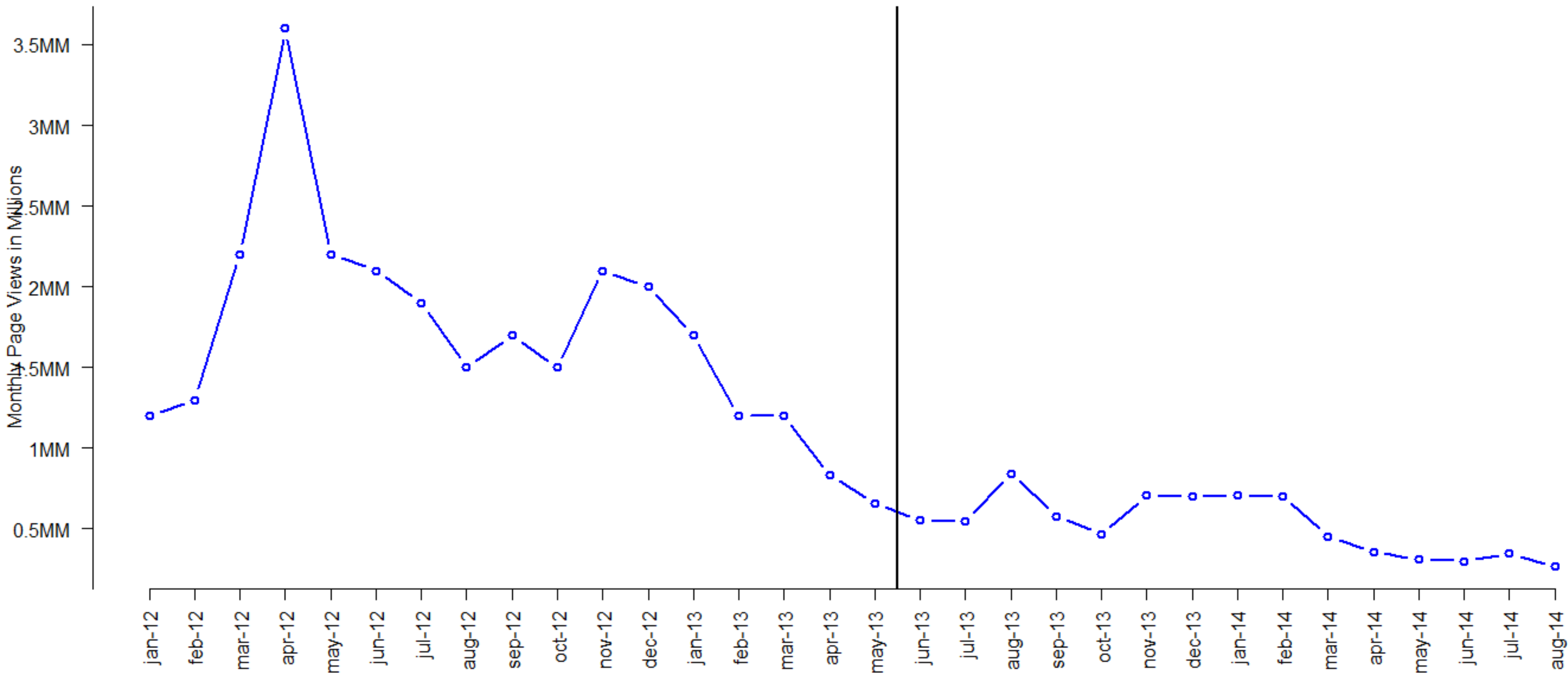
### Popular: Page Views for java



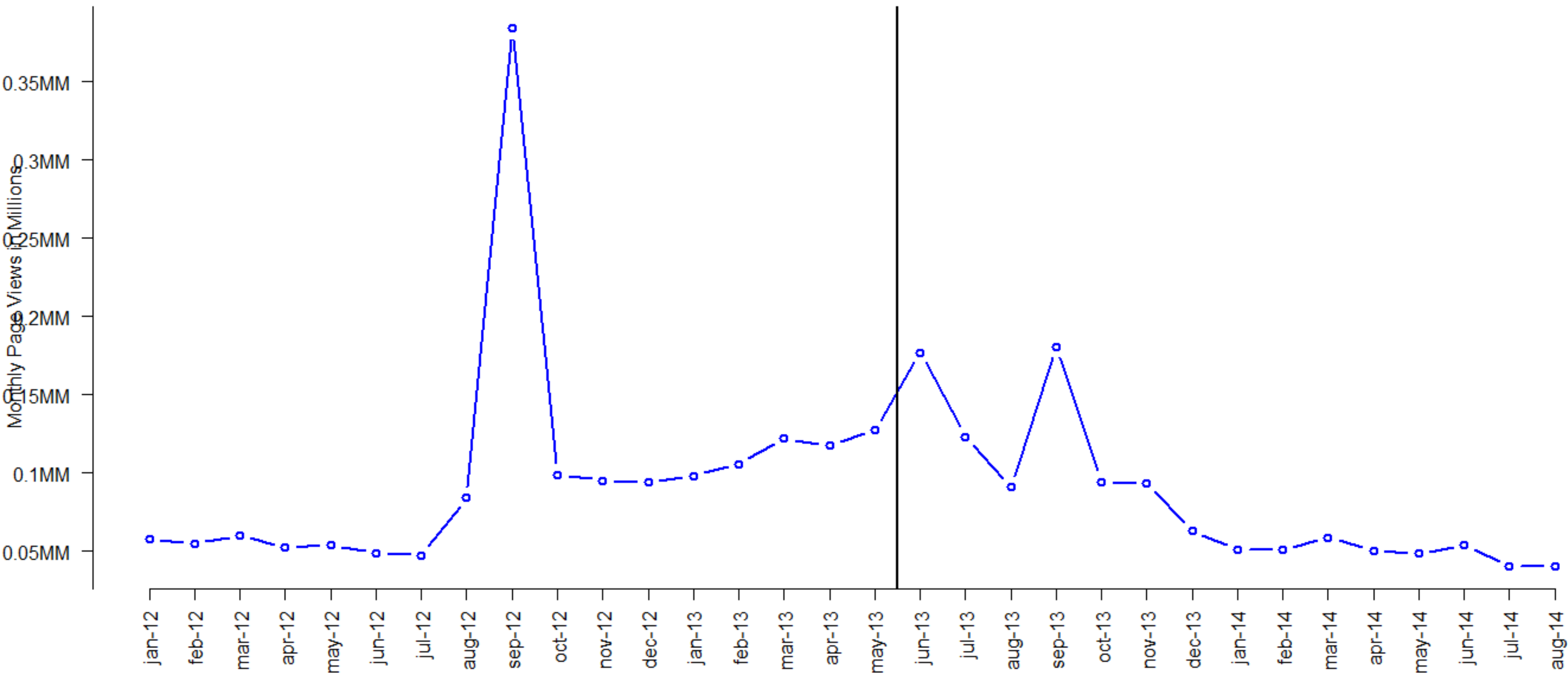




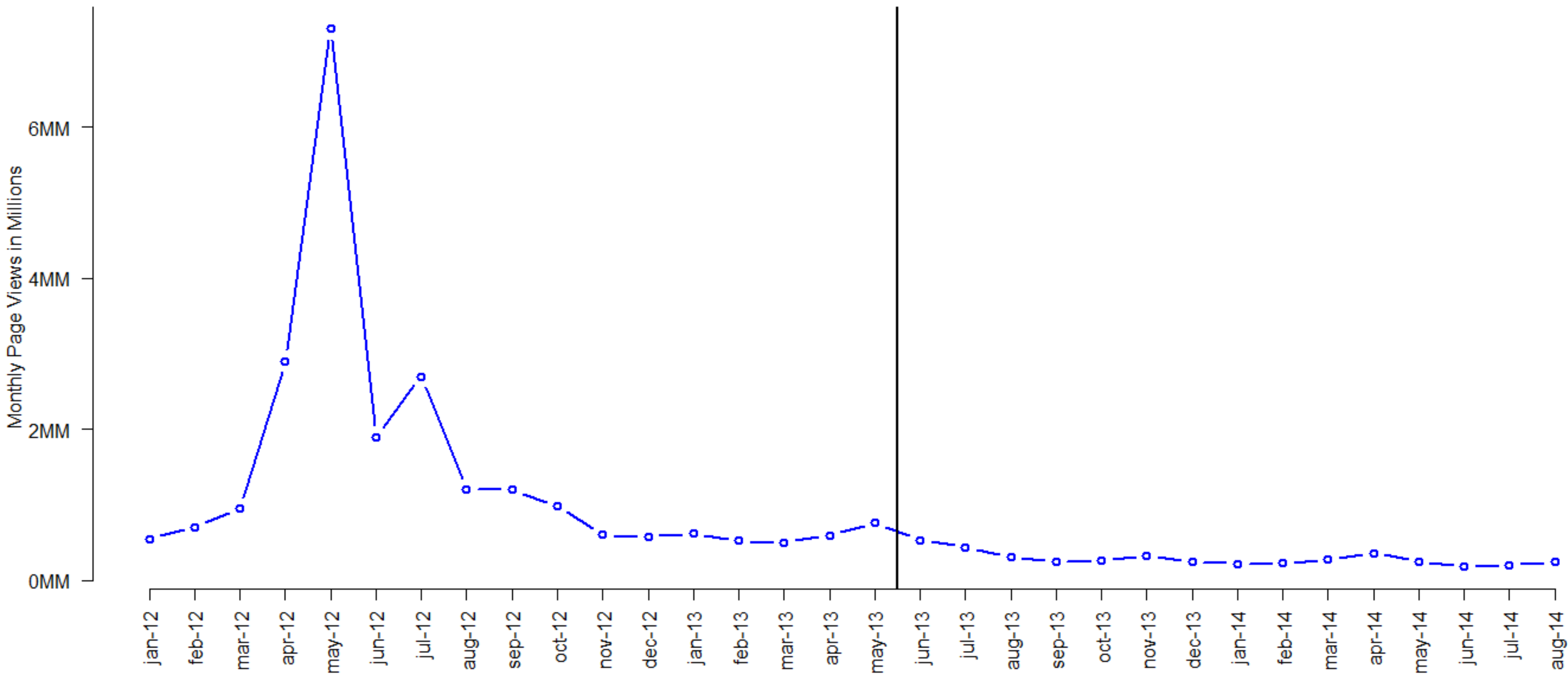
### Popular: Page Views for one\_direction



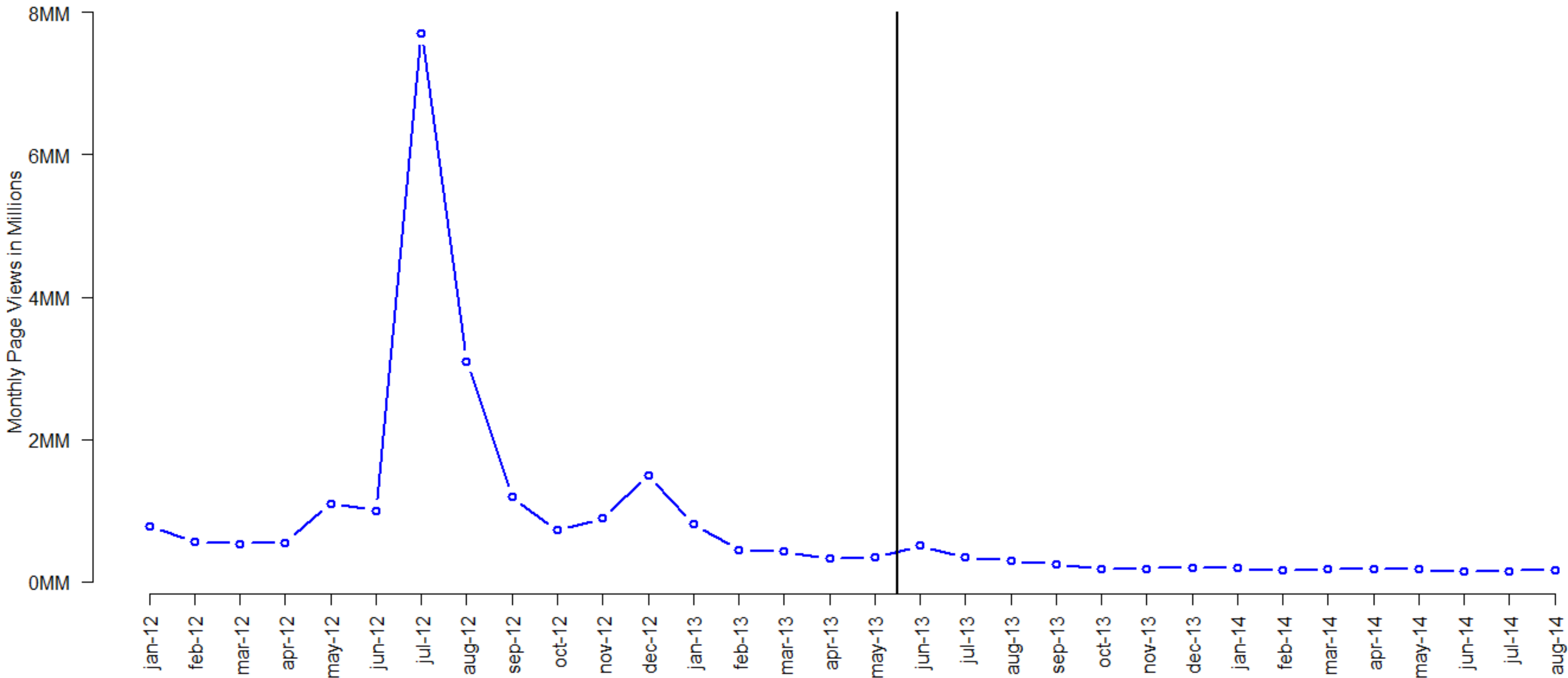
### Popular: Page Views for online\_shopping



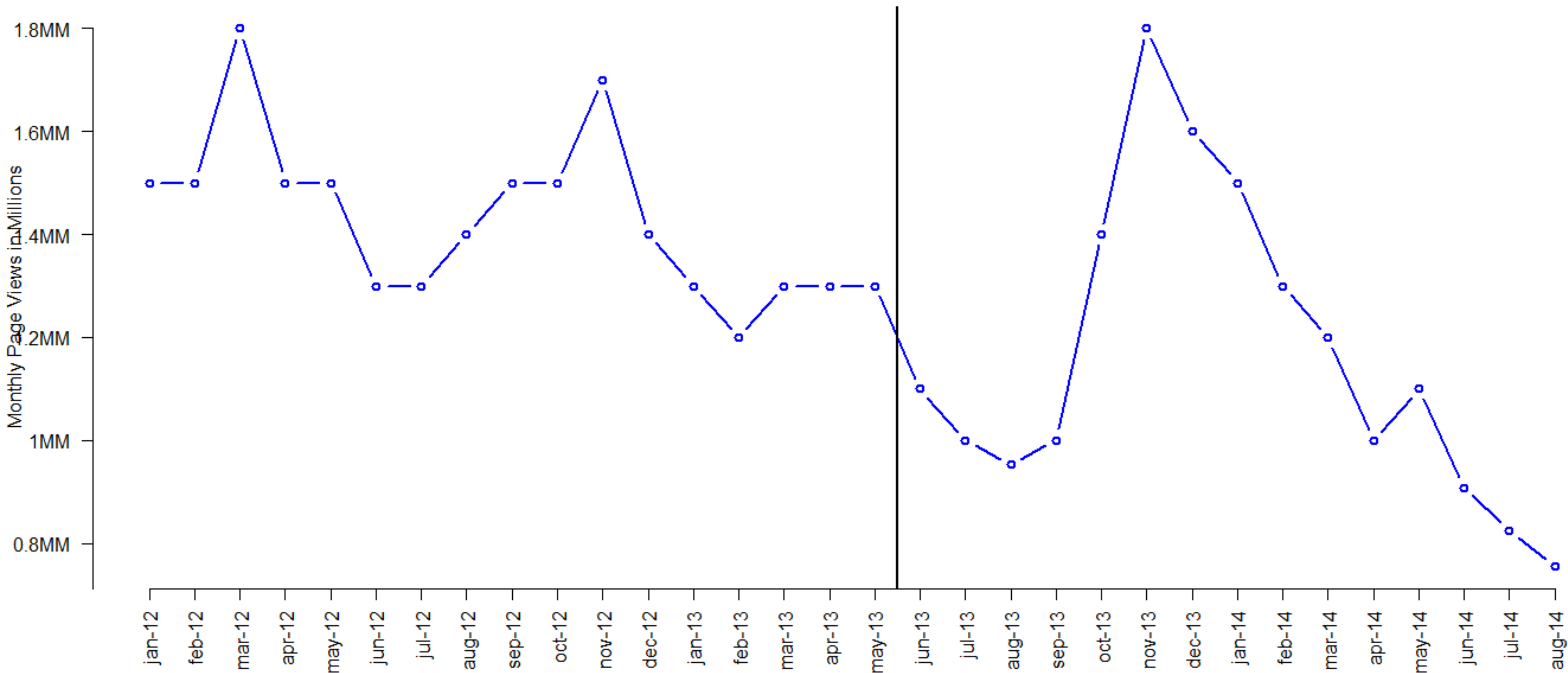
### Popular: Page Views for the \_avengers\_2012\_film



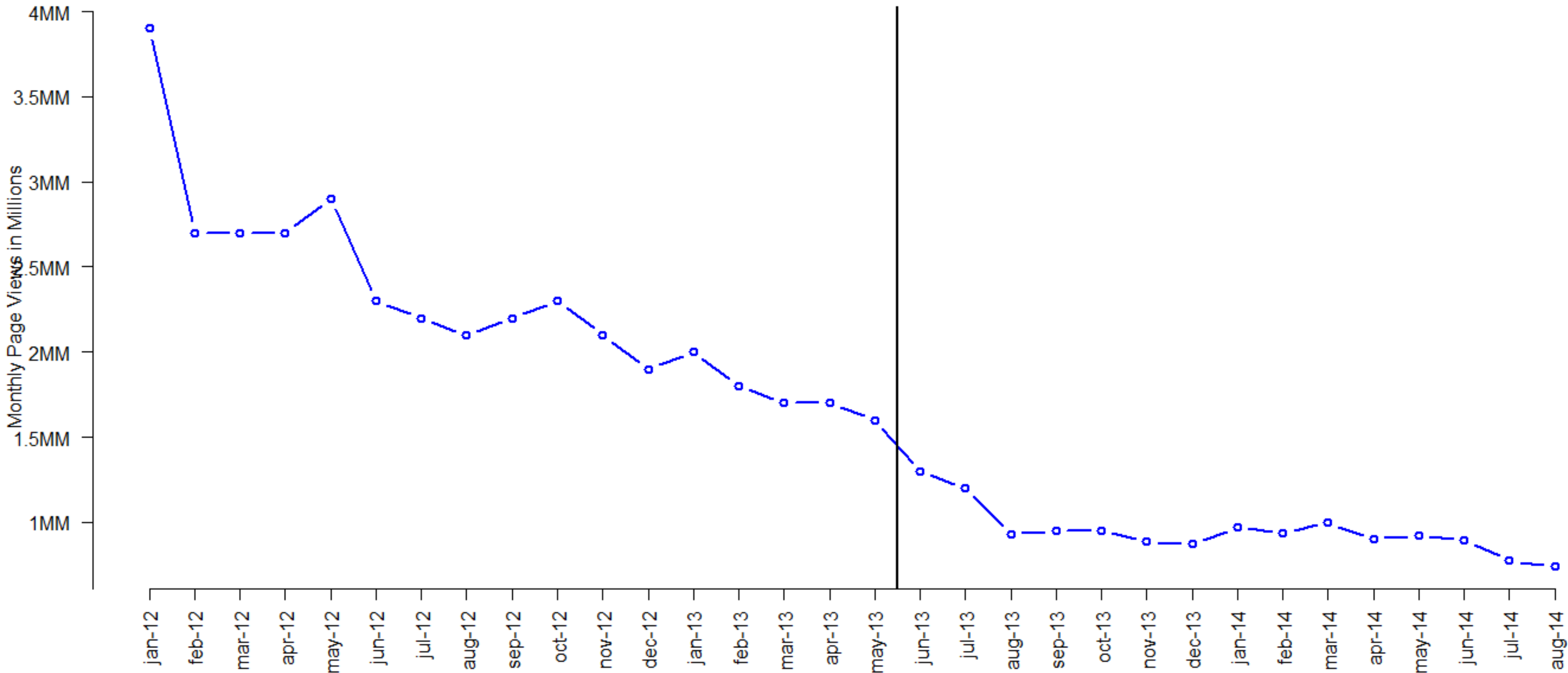
### Popular: Page Views for the\_dark\_knight\_rises



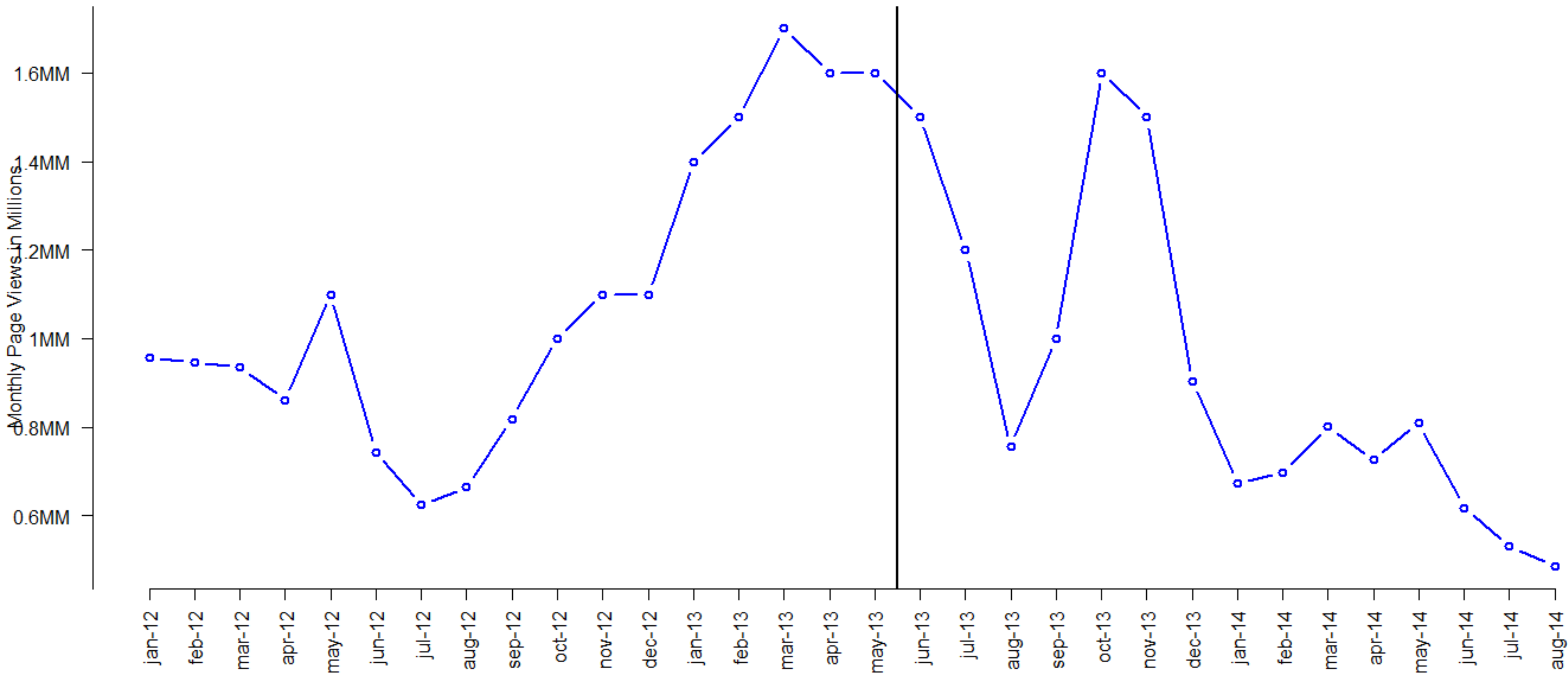
### Popular: Page Views for united\_states



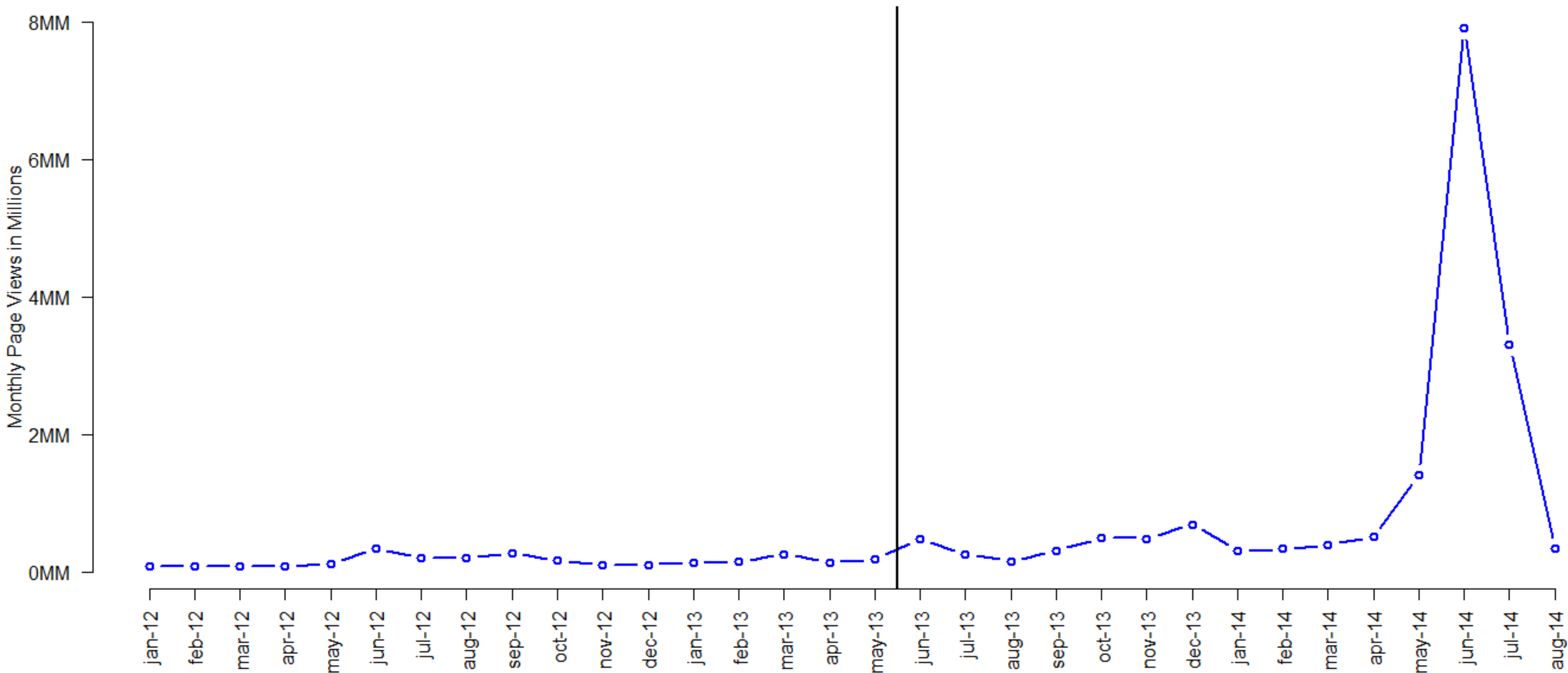
### Popular: Page Views for wiki



### Popular: Page Views for world\_war\_ii

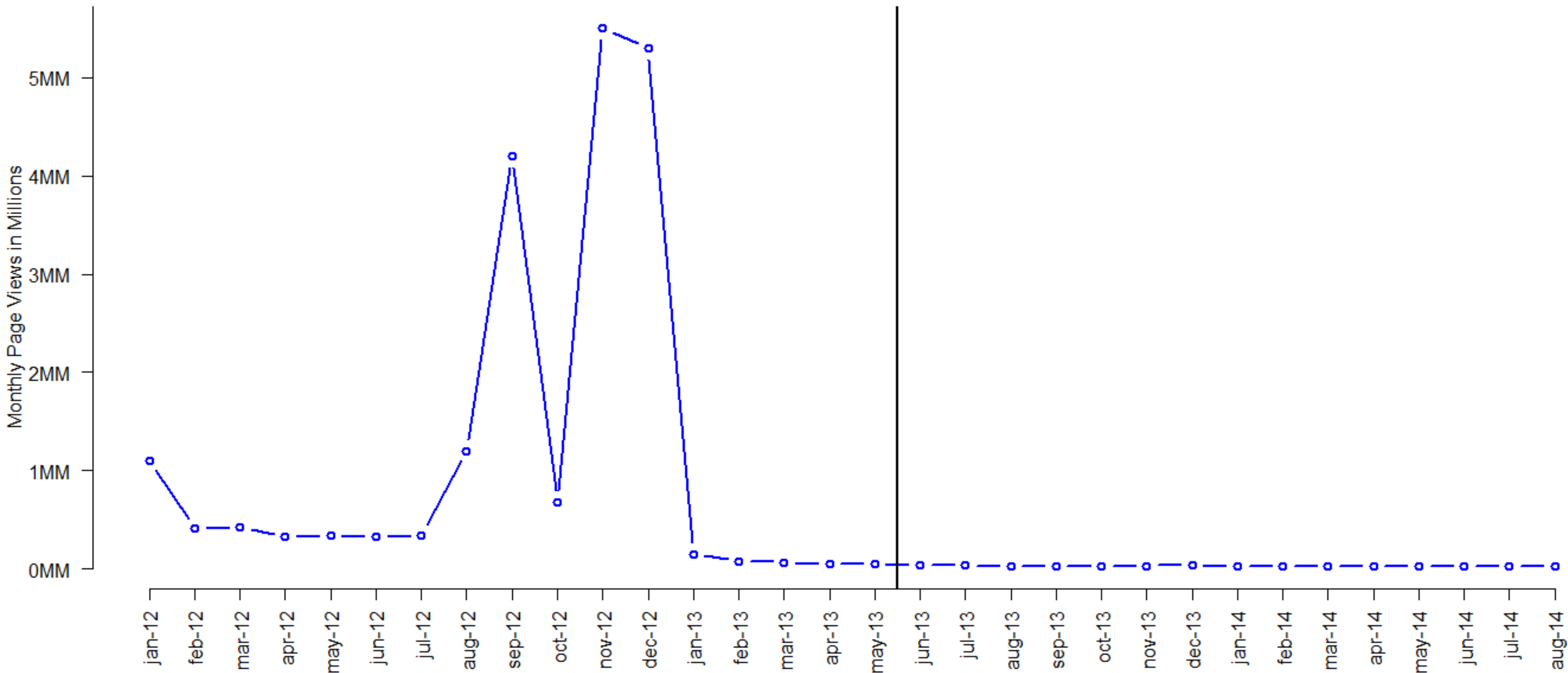


# Popular: Page Views for X\_fifa\_world\_cup

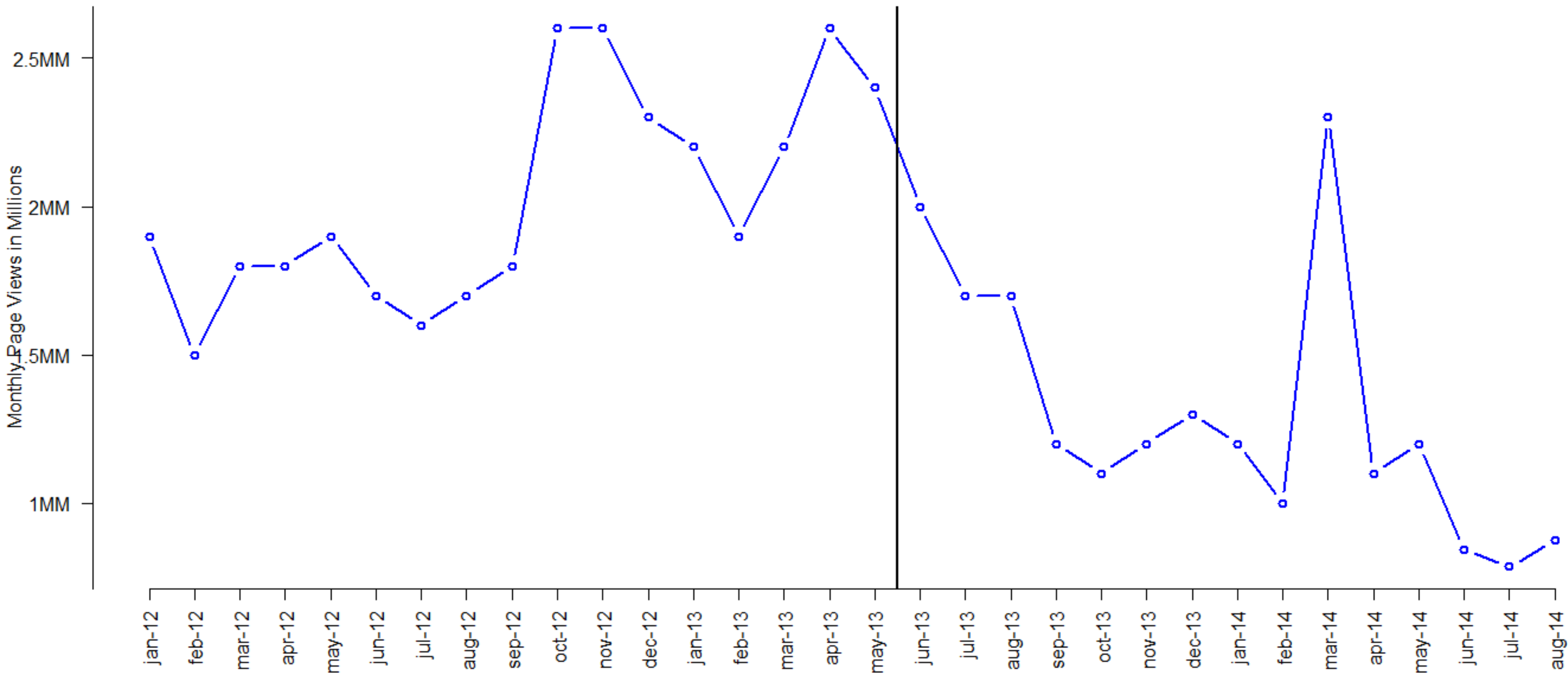




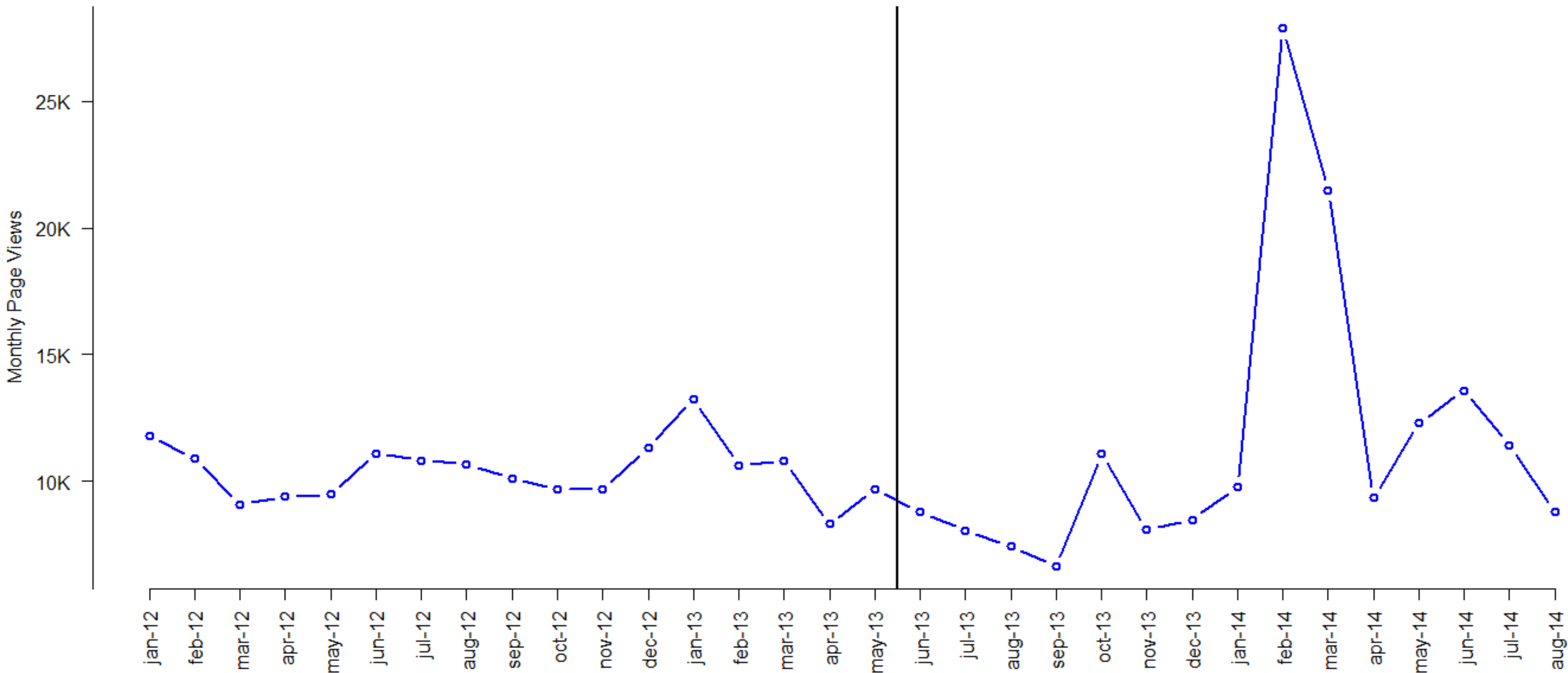
### Popular: Page Views for X\_phenomena



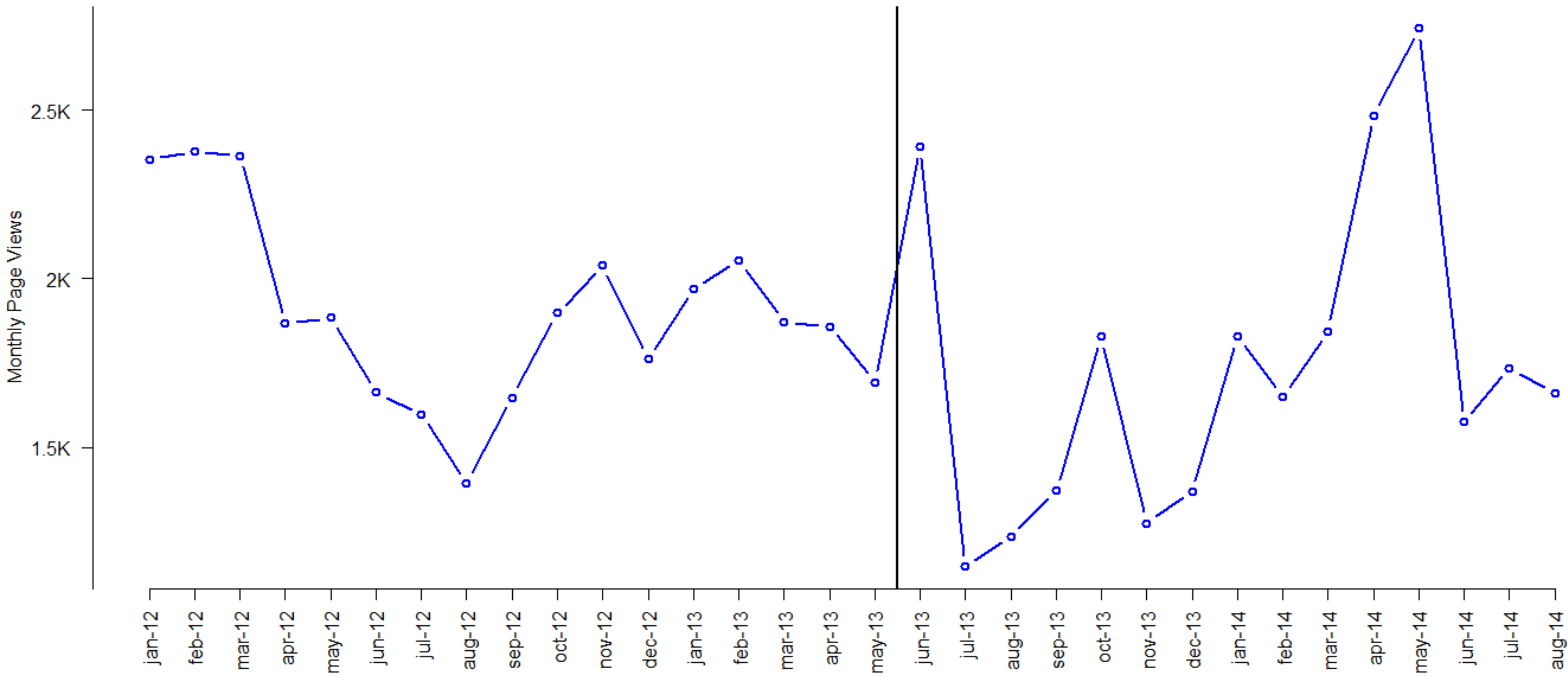
### Popular: Page Views for youtube



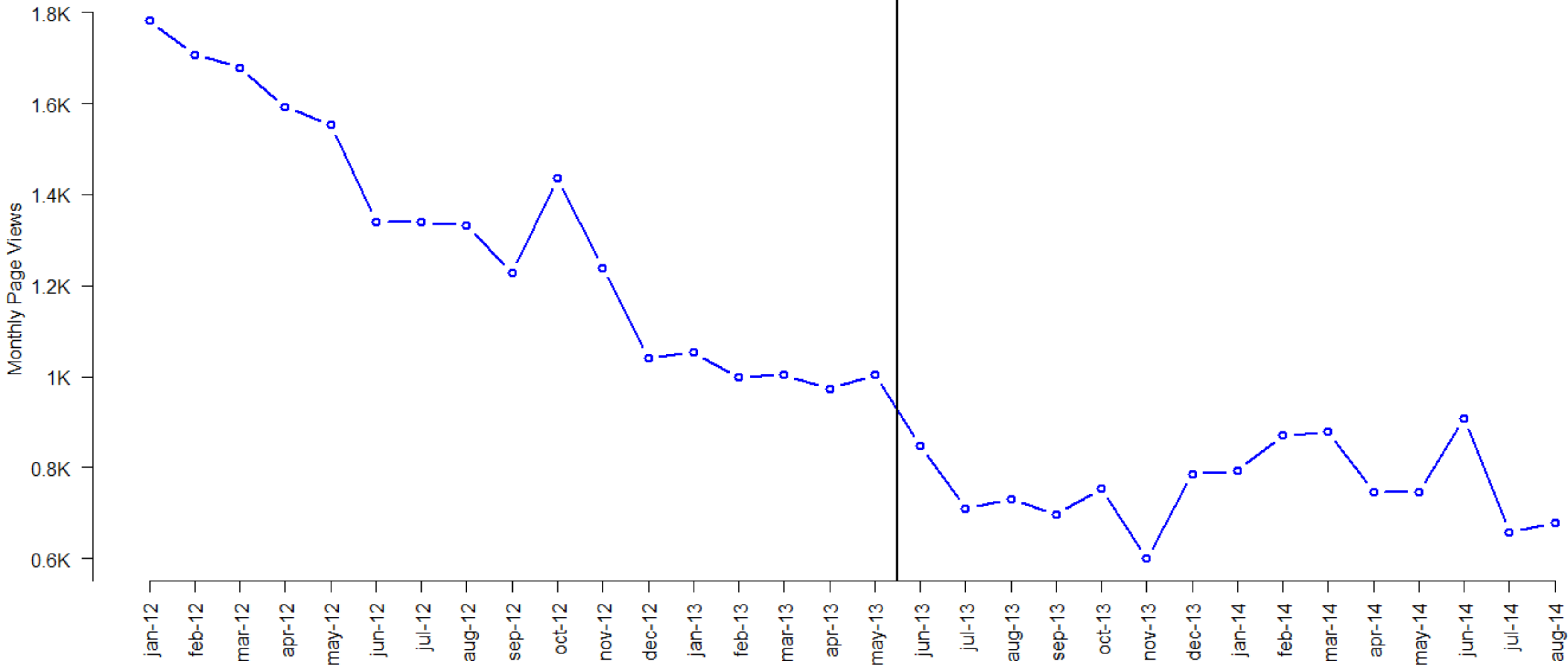
### Security: Page Views for air\_marshals

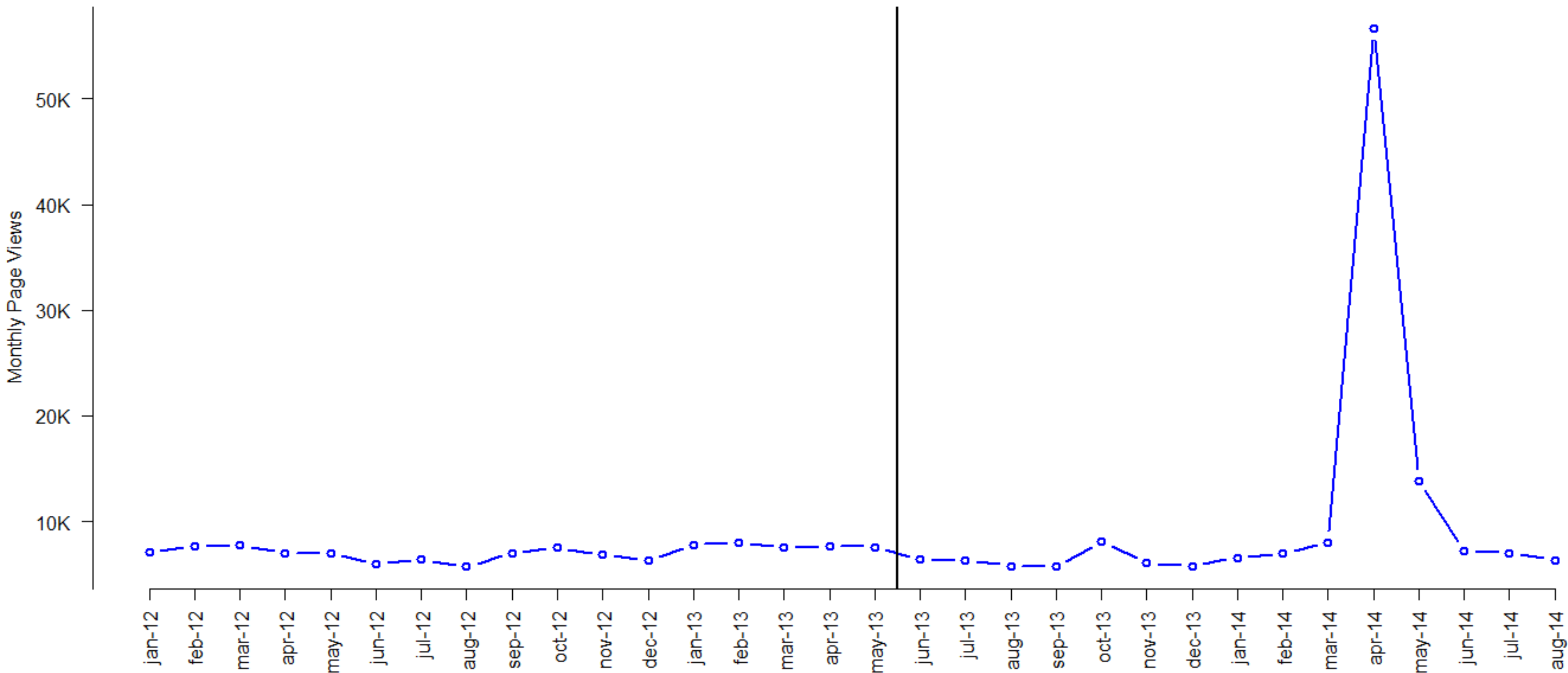


### Security: Page Views for alcohol\_and\_tobacco\_tax\_and\_trad

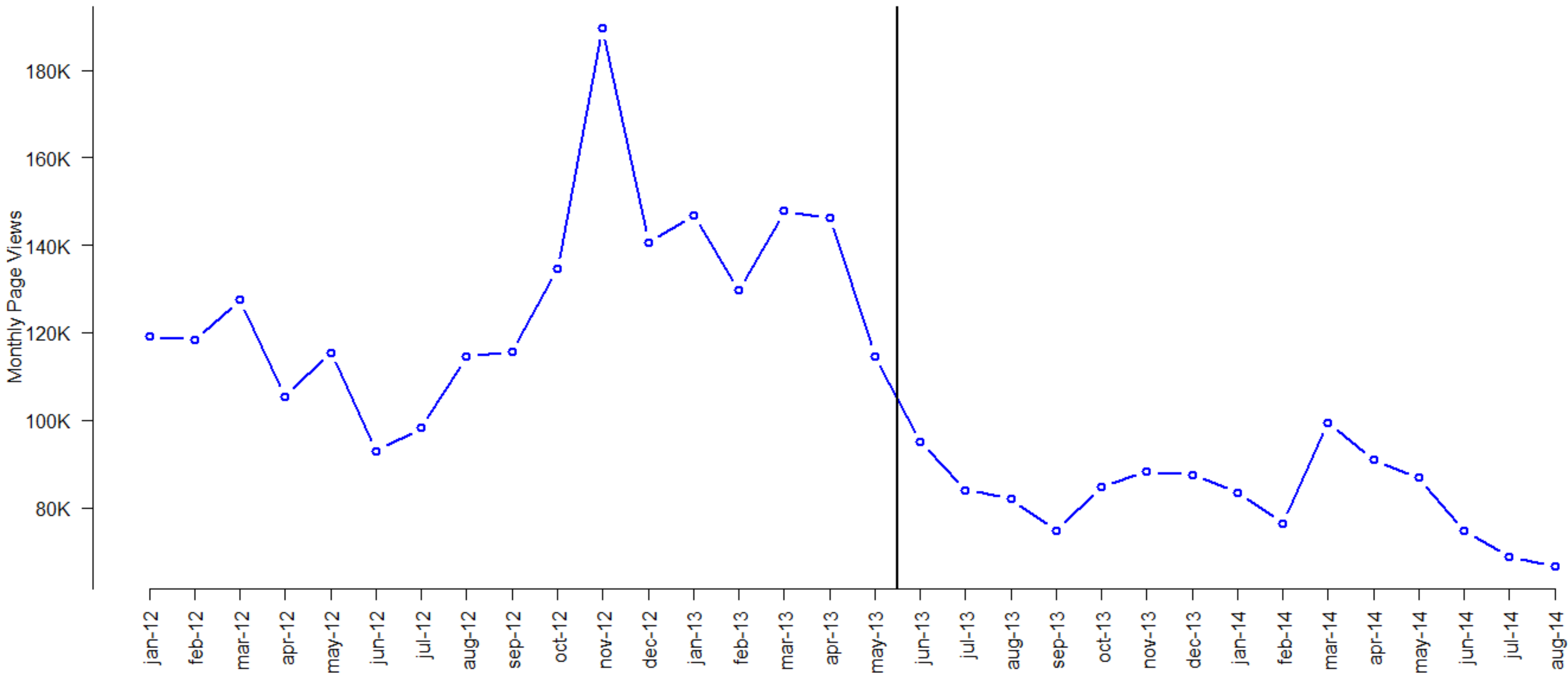


### Security: Page Views for border\_patrol

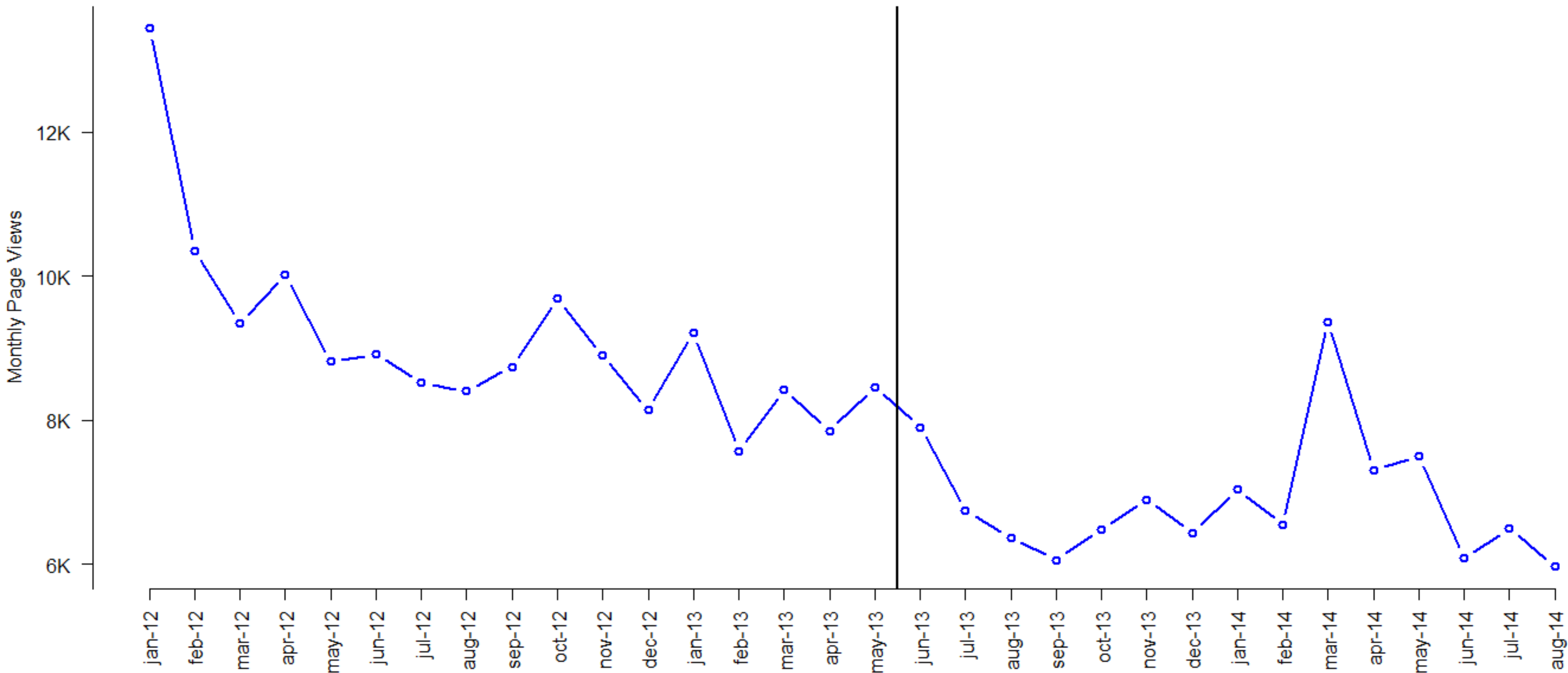




### Security: Page Views for cia

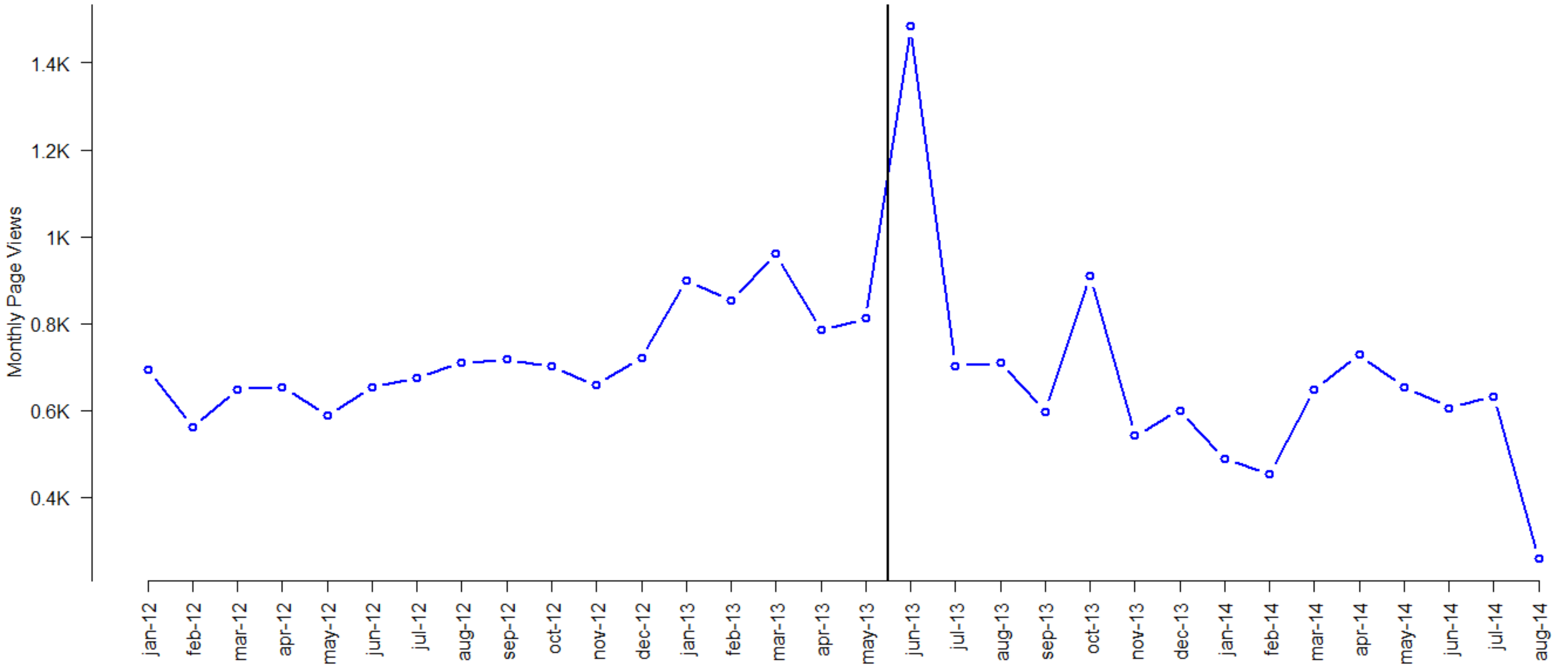


### Security: Page Views for coast\_guard

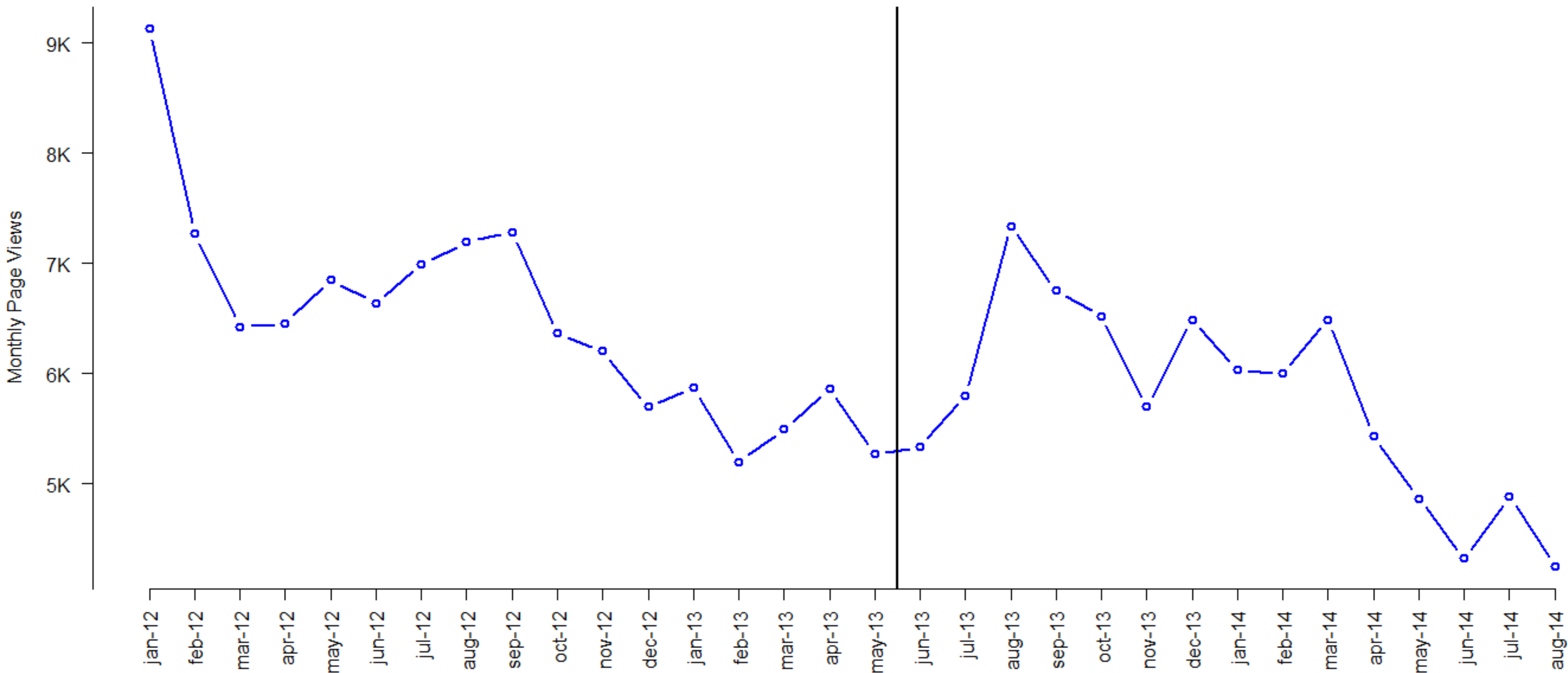




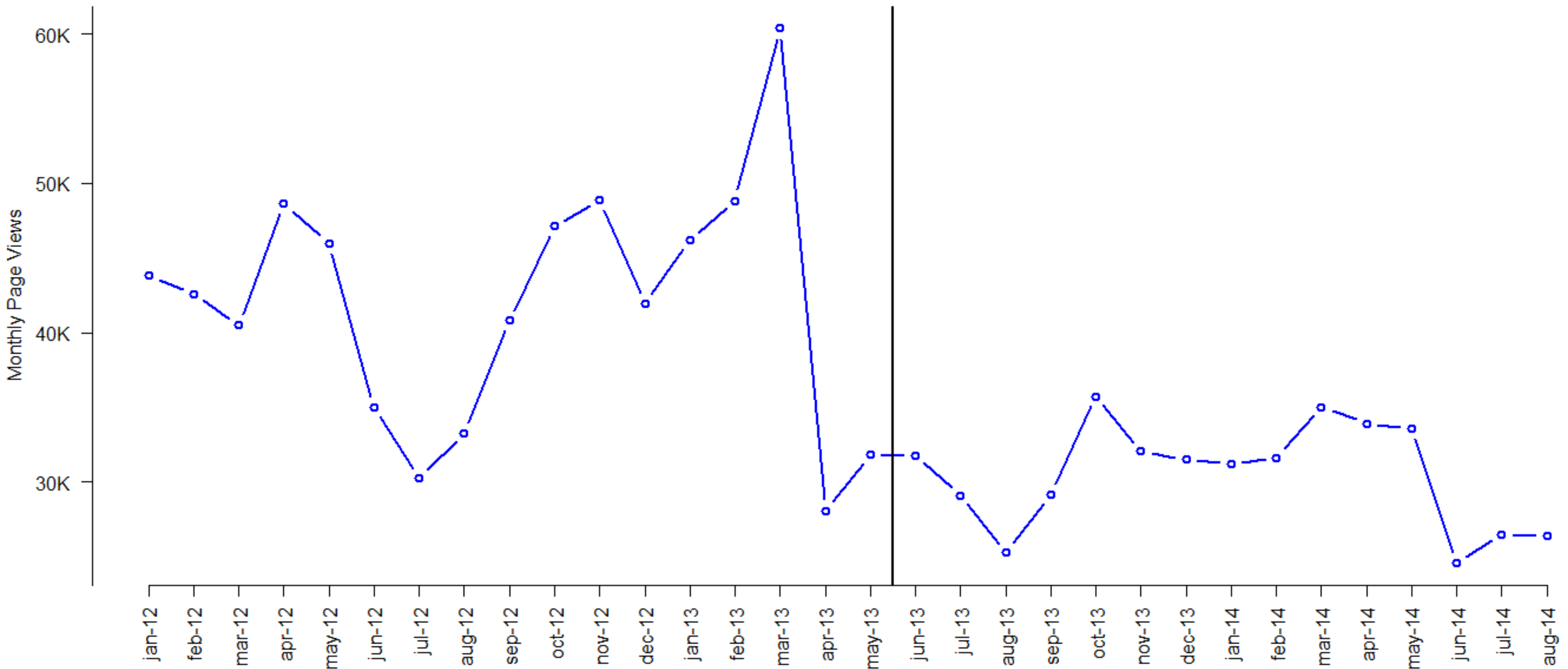
### Security: Page Views for customs\_and\_border\_protection



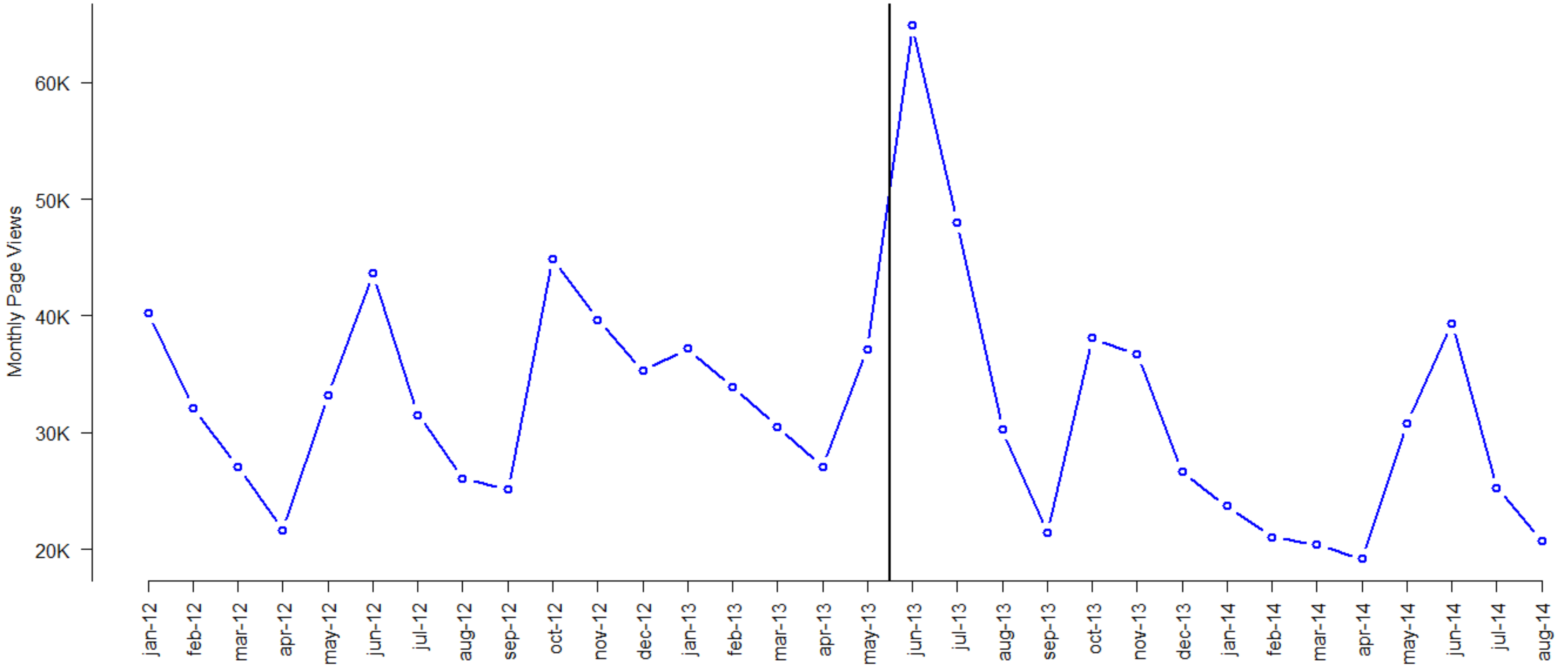
### Security: Page Views for dea



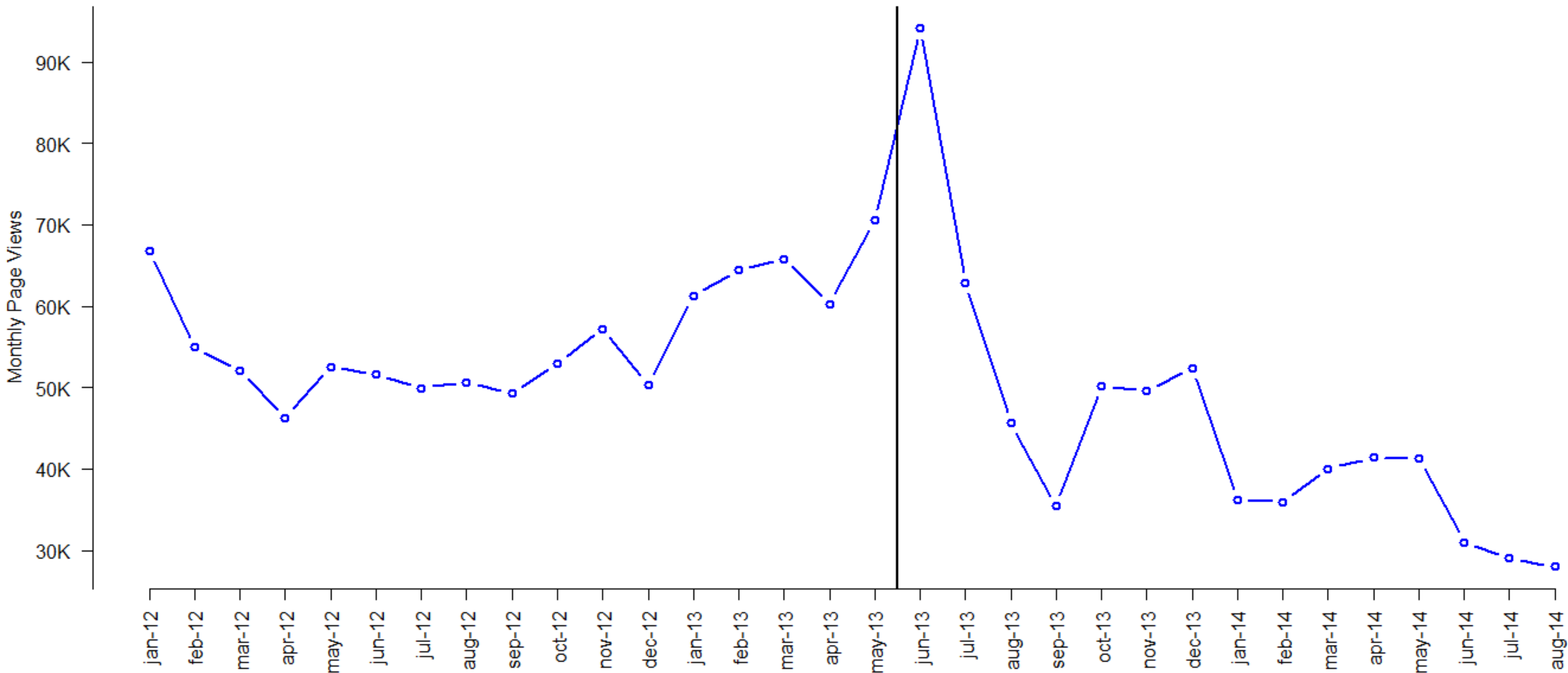
### Security: Page Views for department\_of\_homeland\_security



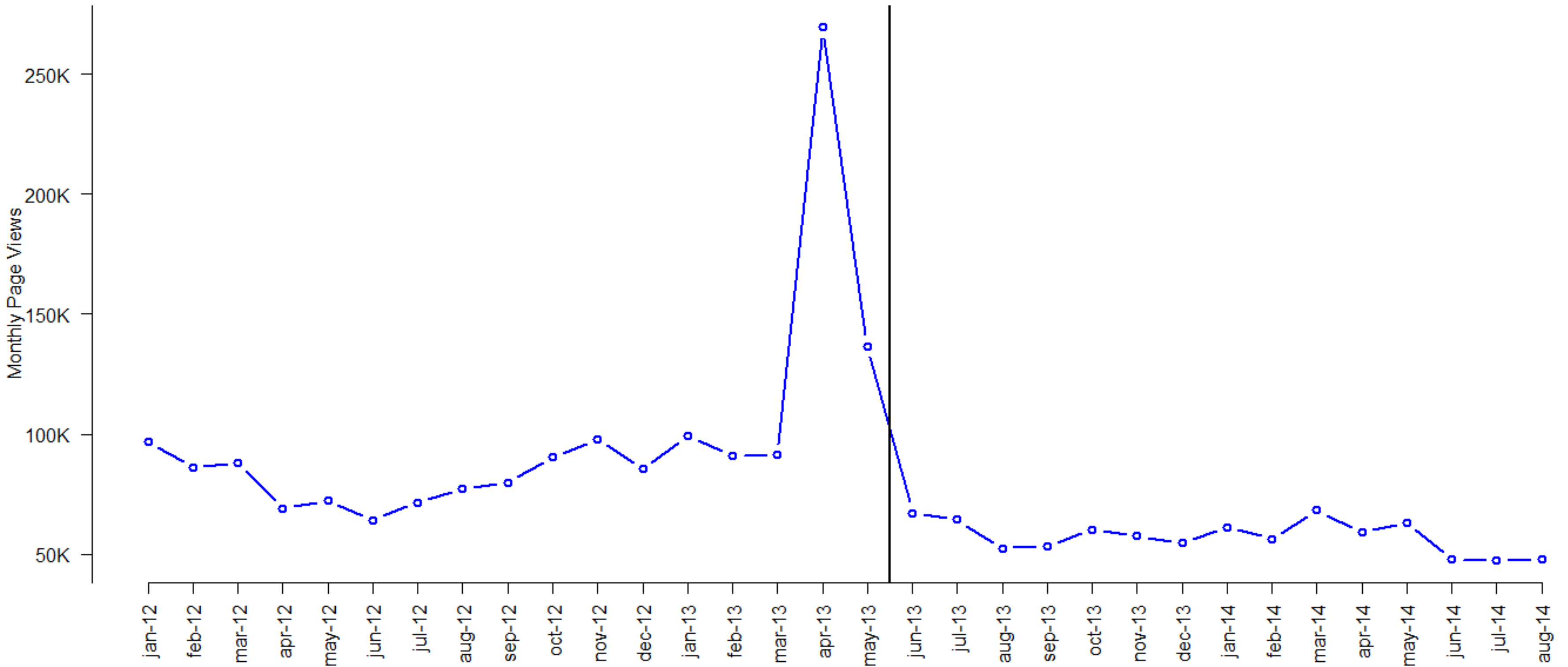
### Security: Page Views for emergency\_management

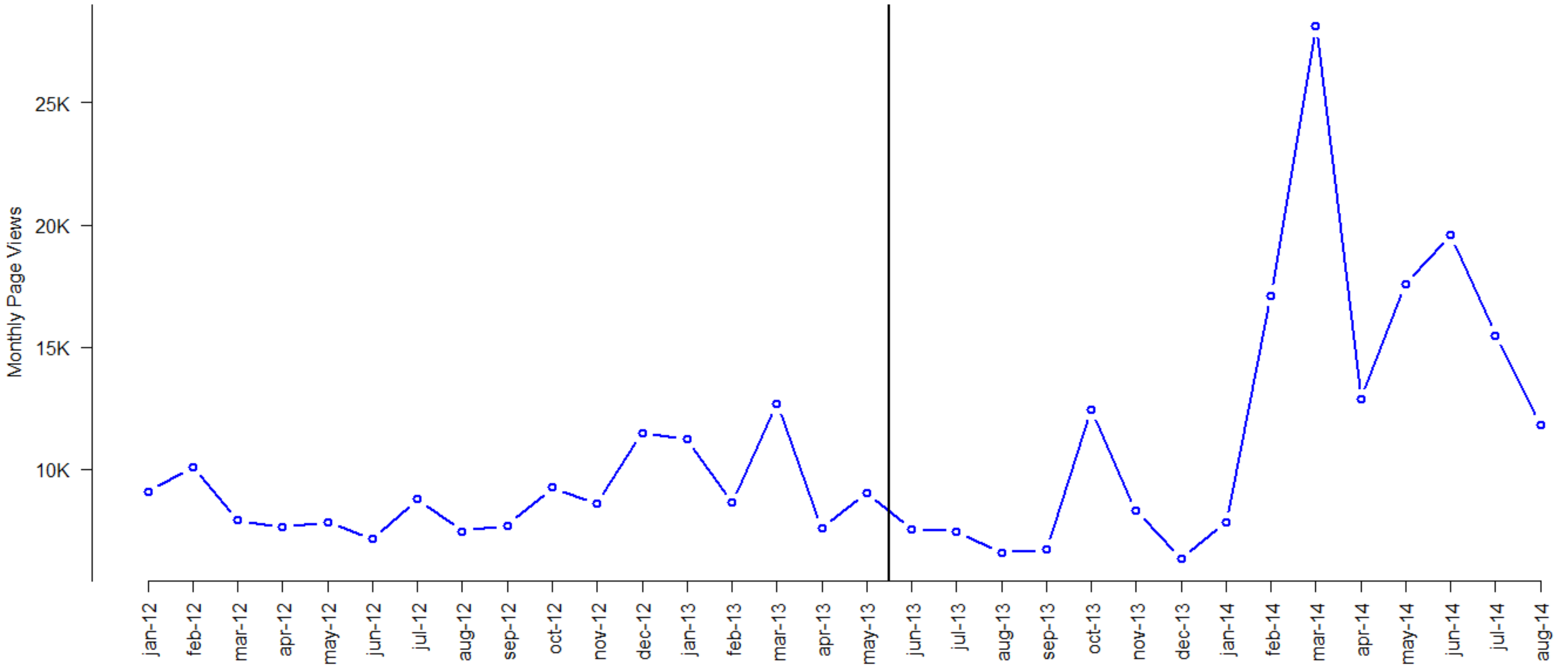


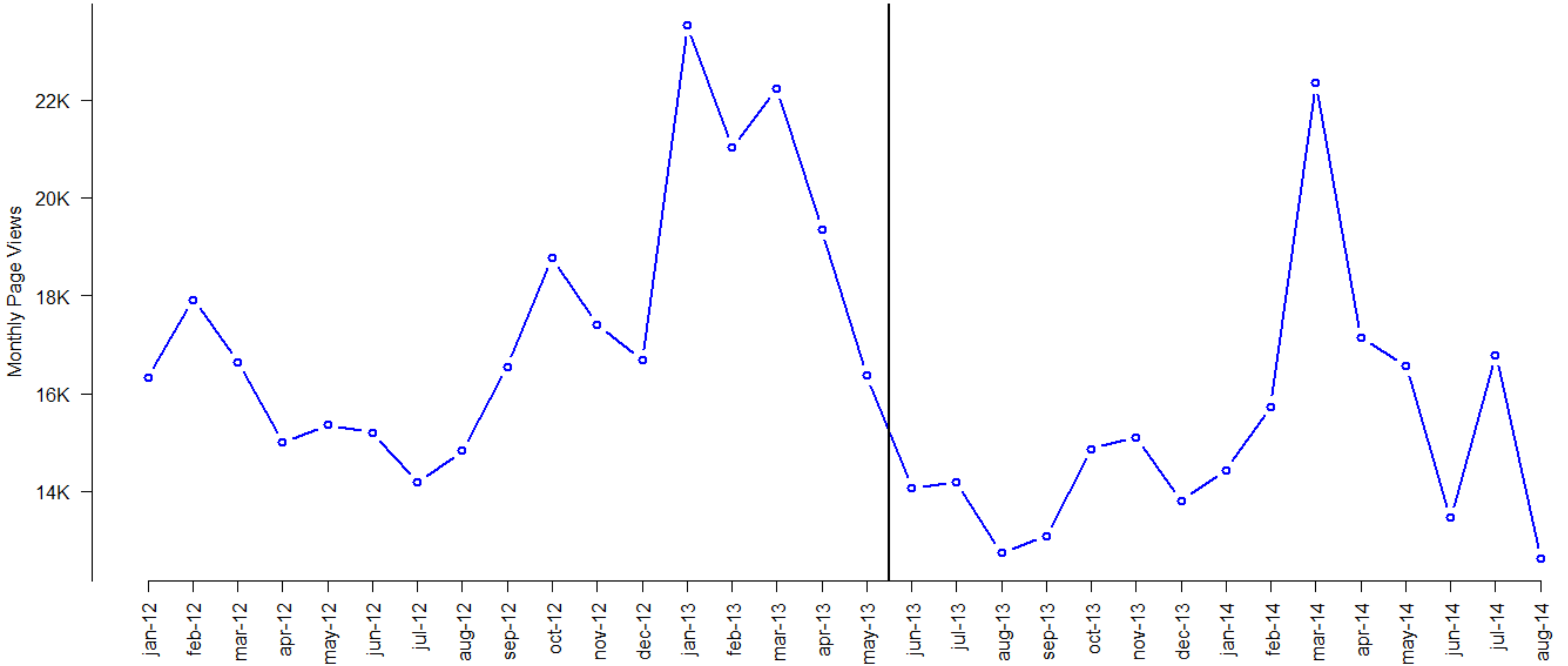
### Security: Page Views for espionage



### Security: Page Views for fbi

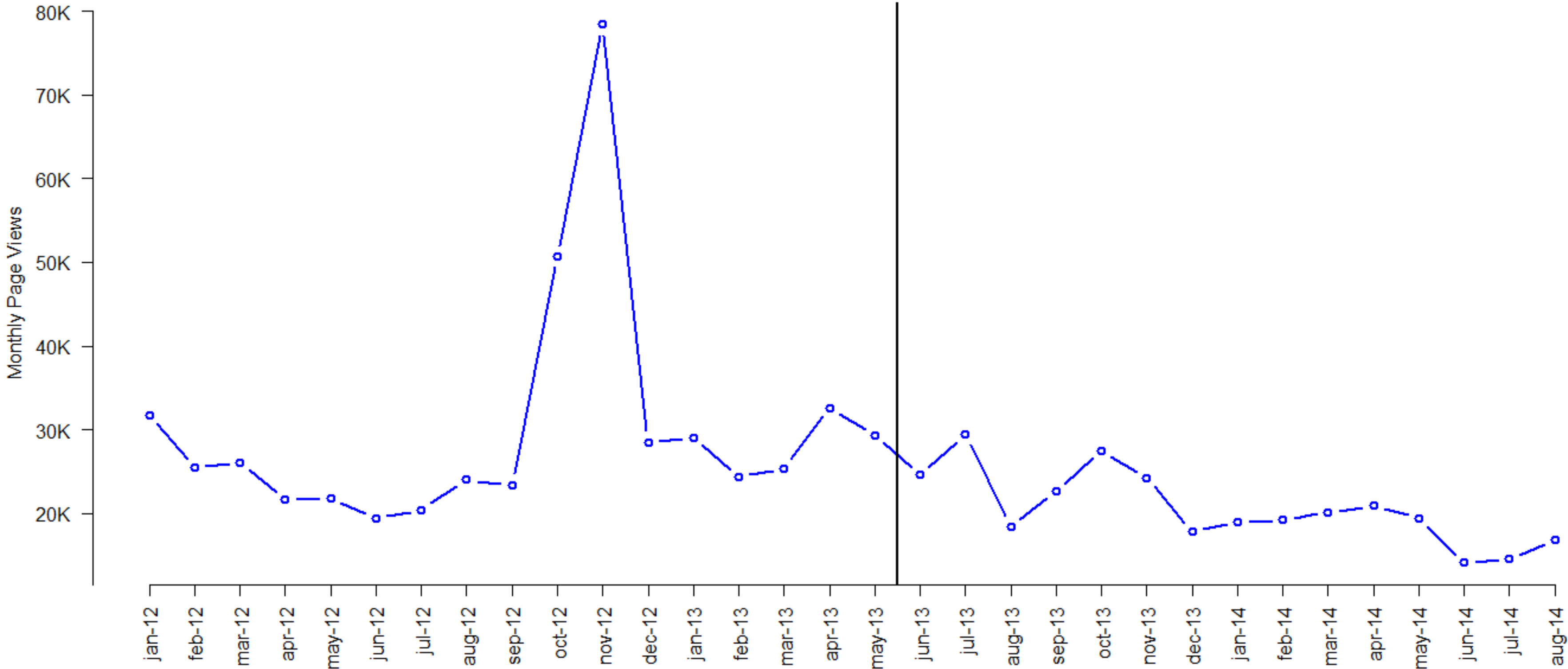




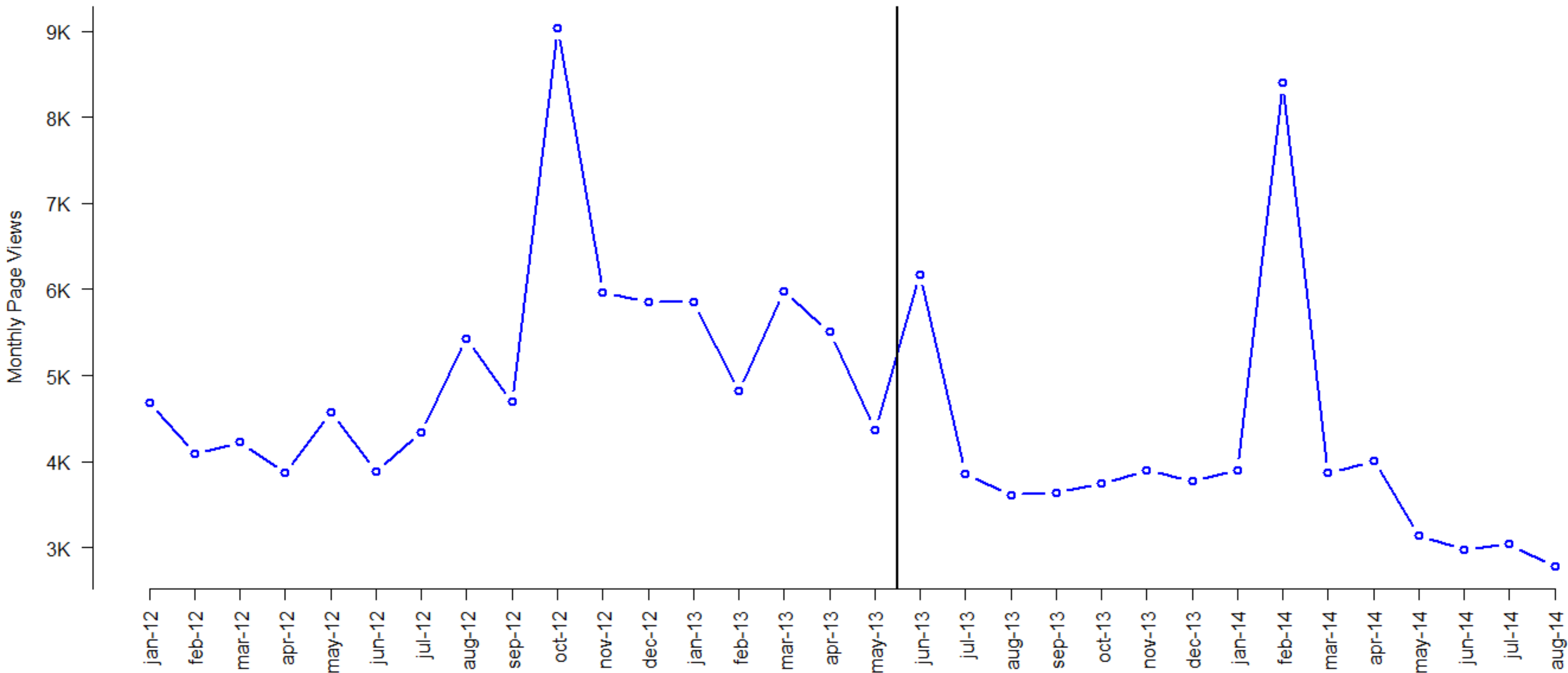




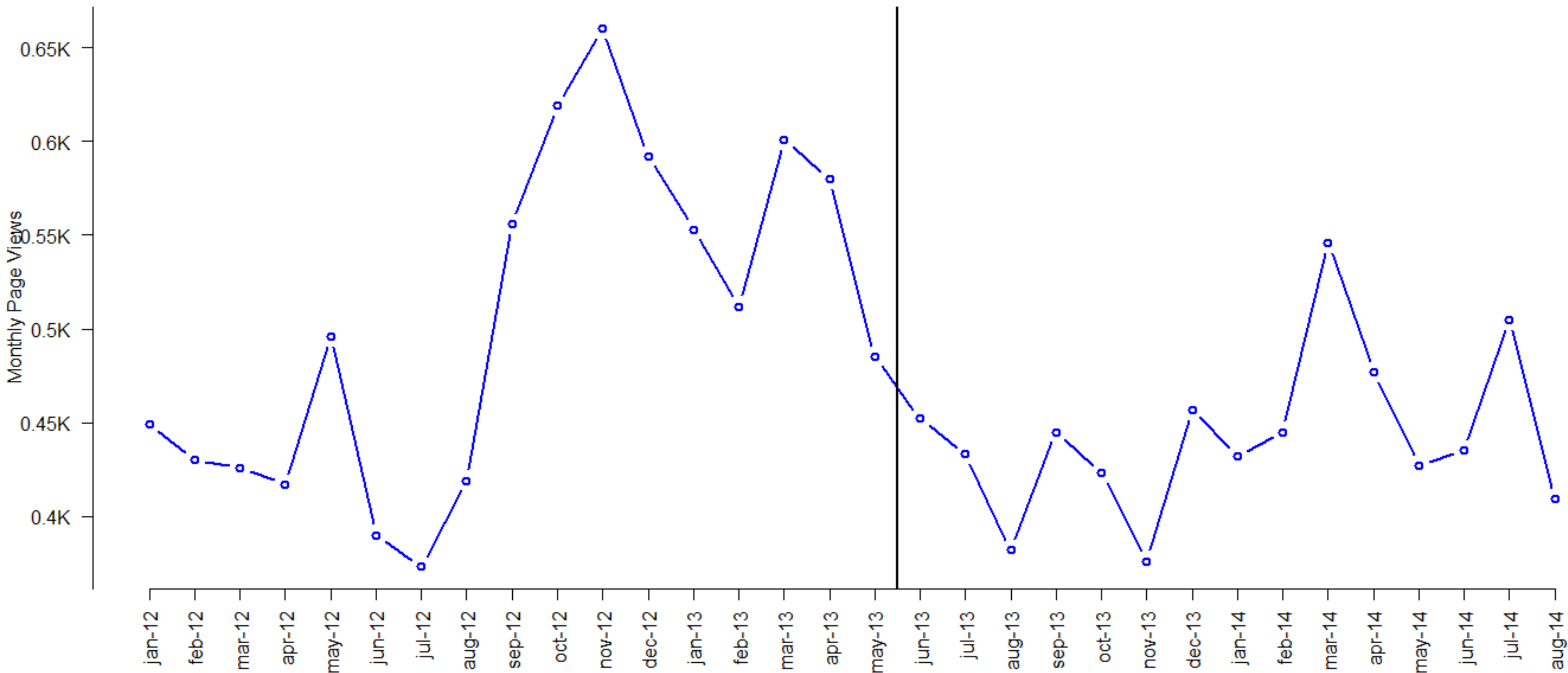
### Security: Page Views for federal\_emergency\_management\_age



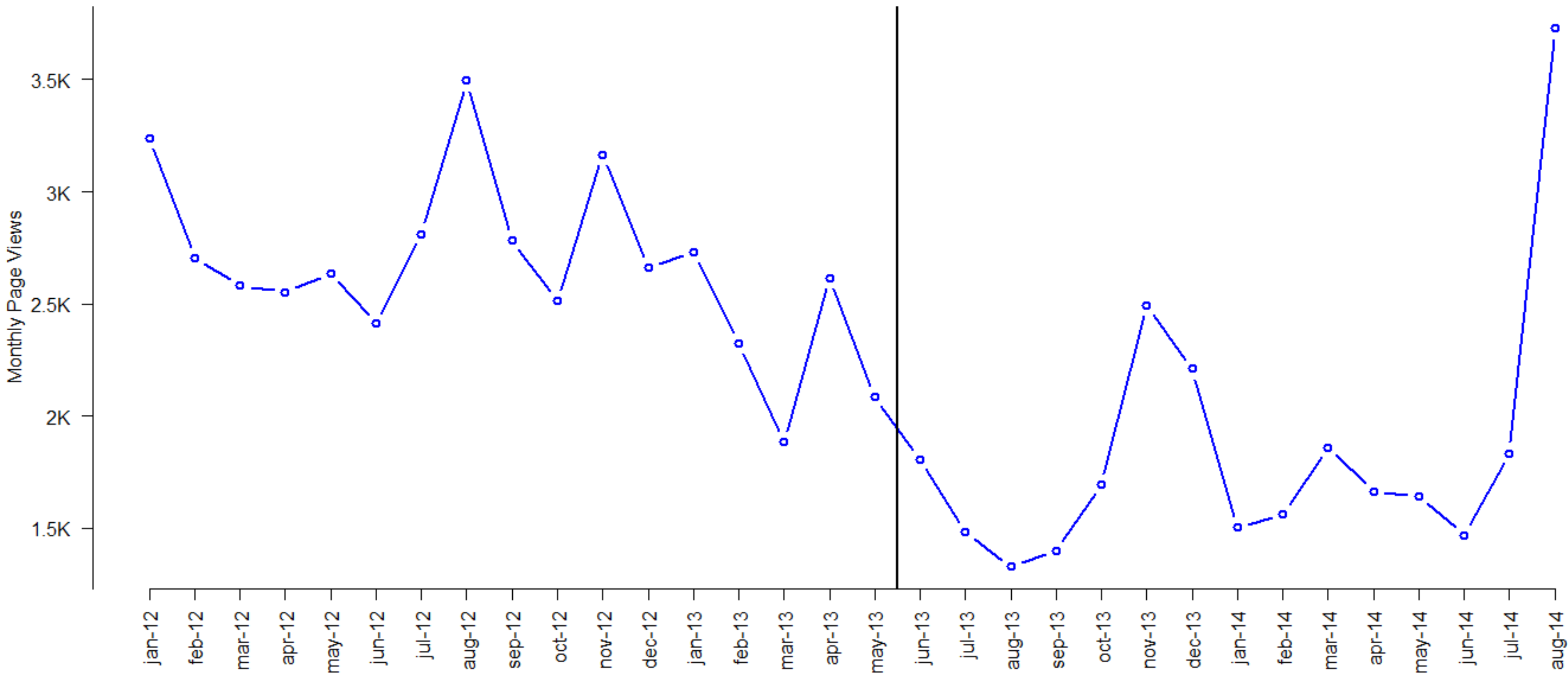
### Security: Page Views for fusion\_center



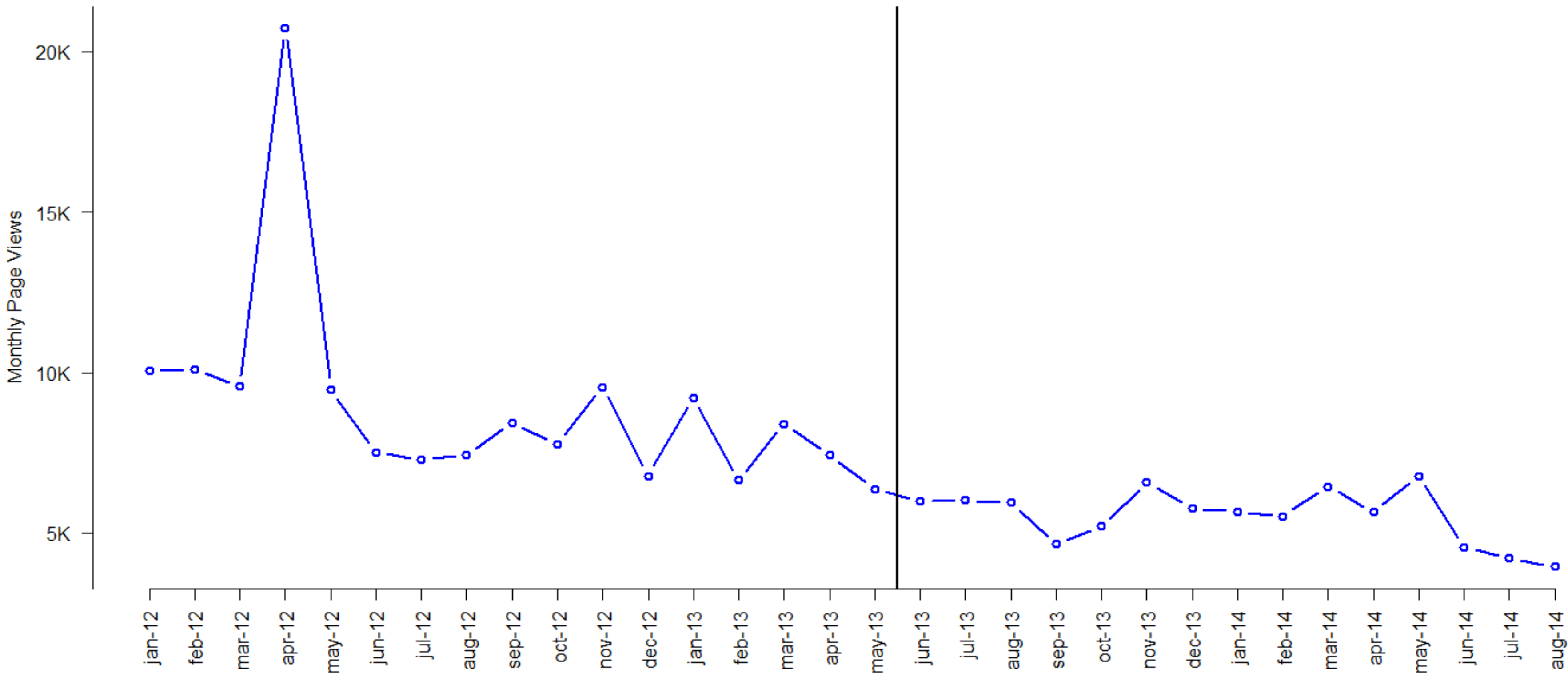
### Security: Page Views for homeland\_defense



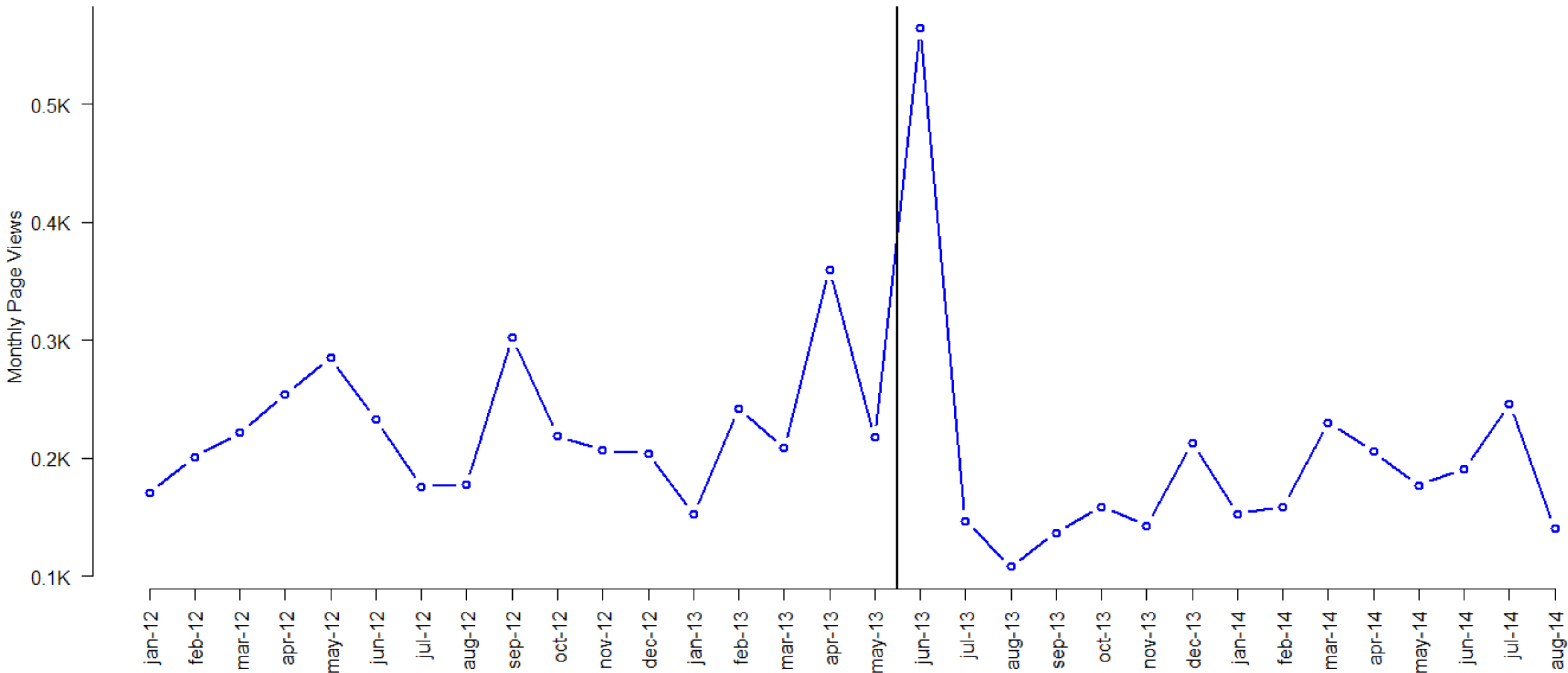
### Security: Page Views for national\_guard



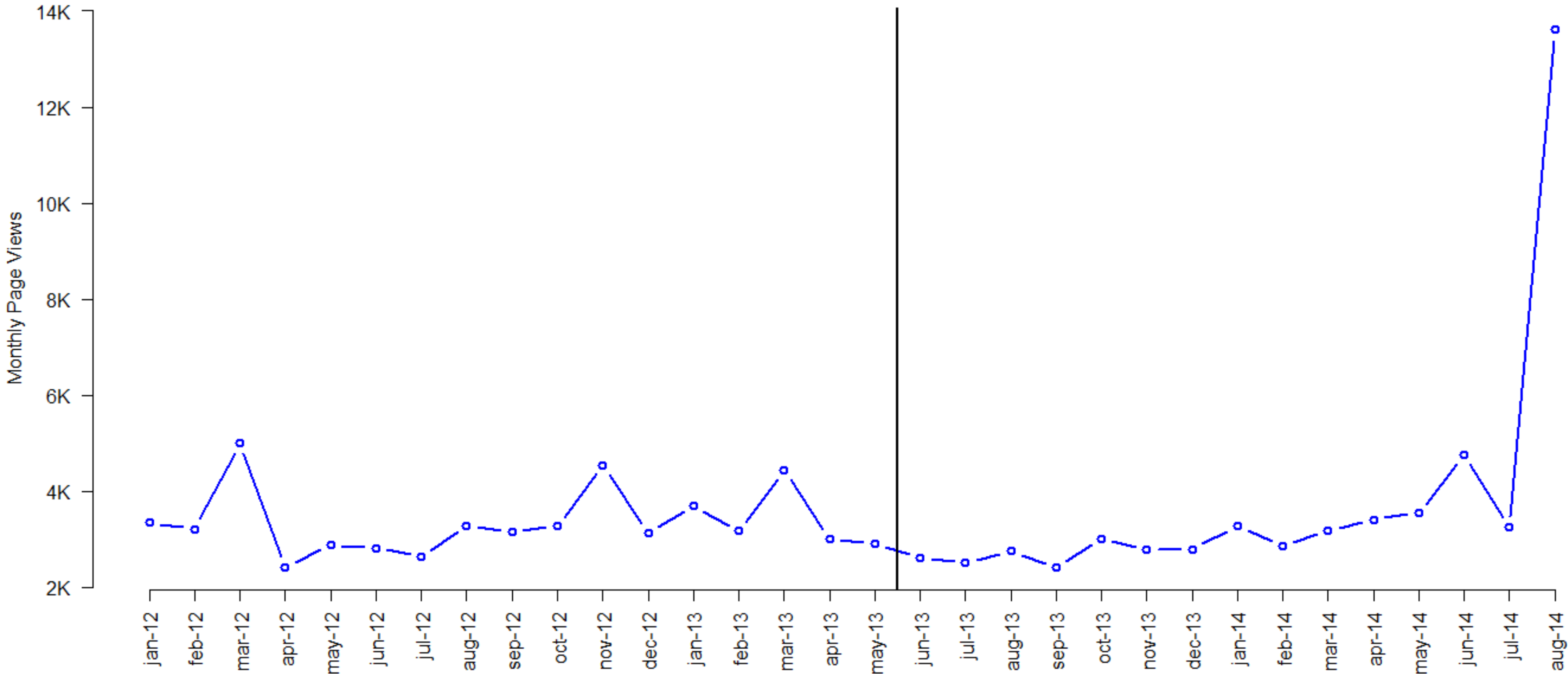
### Security: Page Views for secret\_service



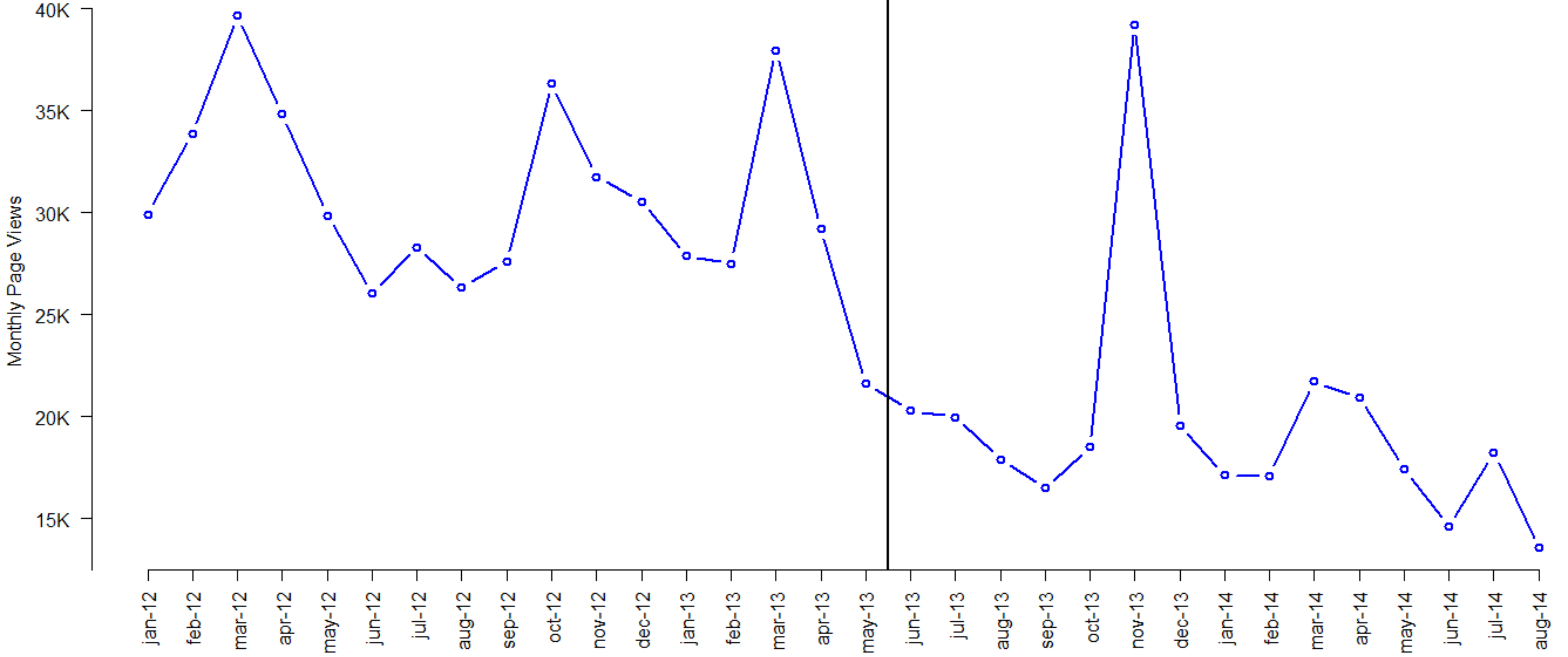
### Security: Page Views for secure\_border\_initiative



### Security: Page Views for Task\_Force\_88\_antiterrorist\_unit

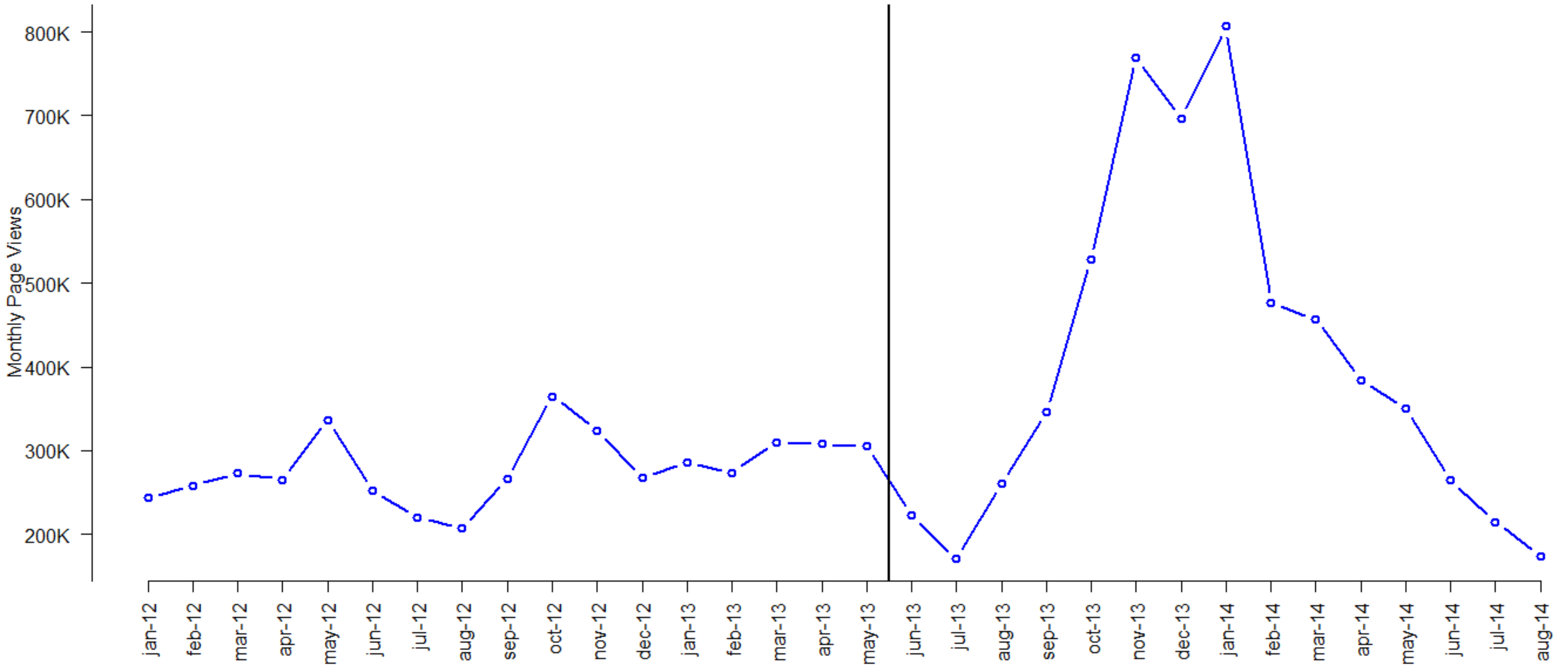


### Security: Page Views for transportation\_security\_administ

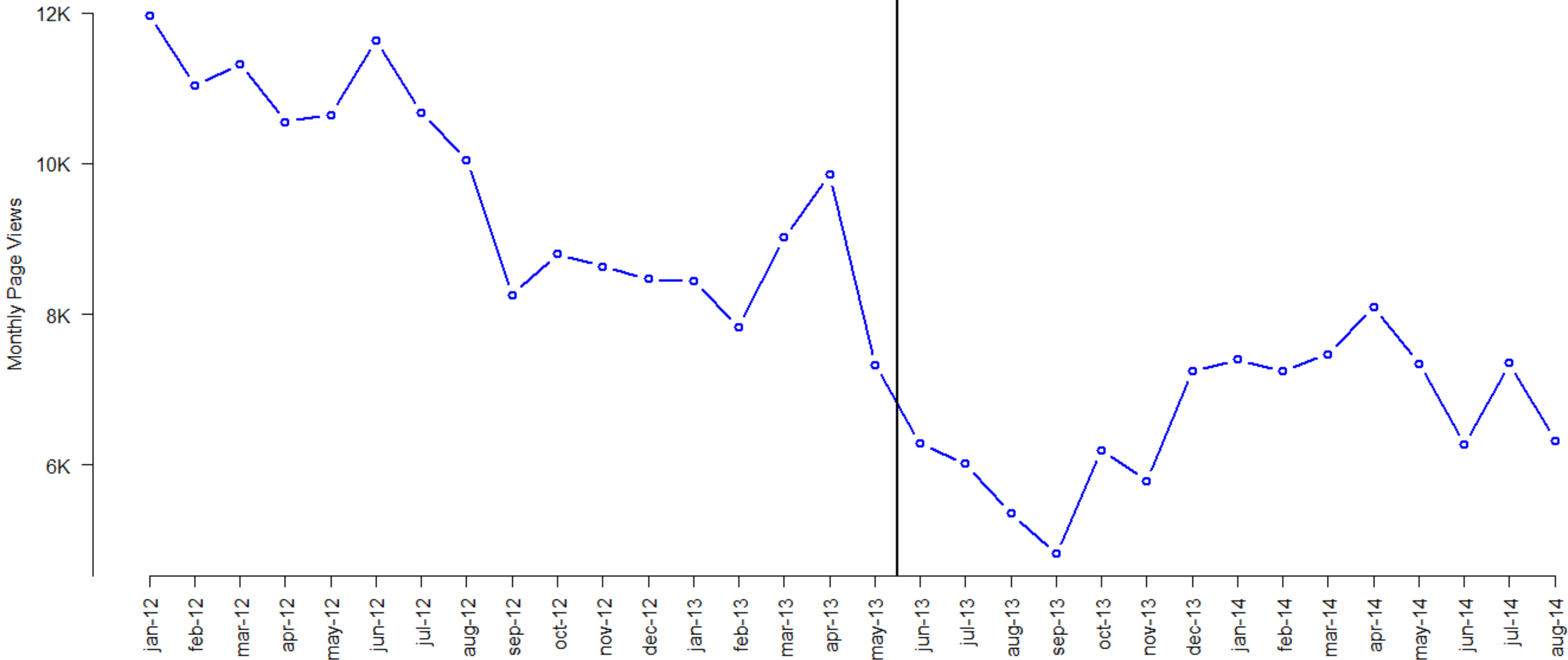


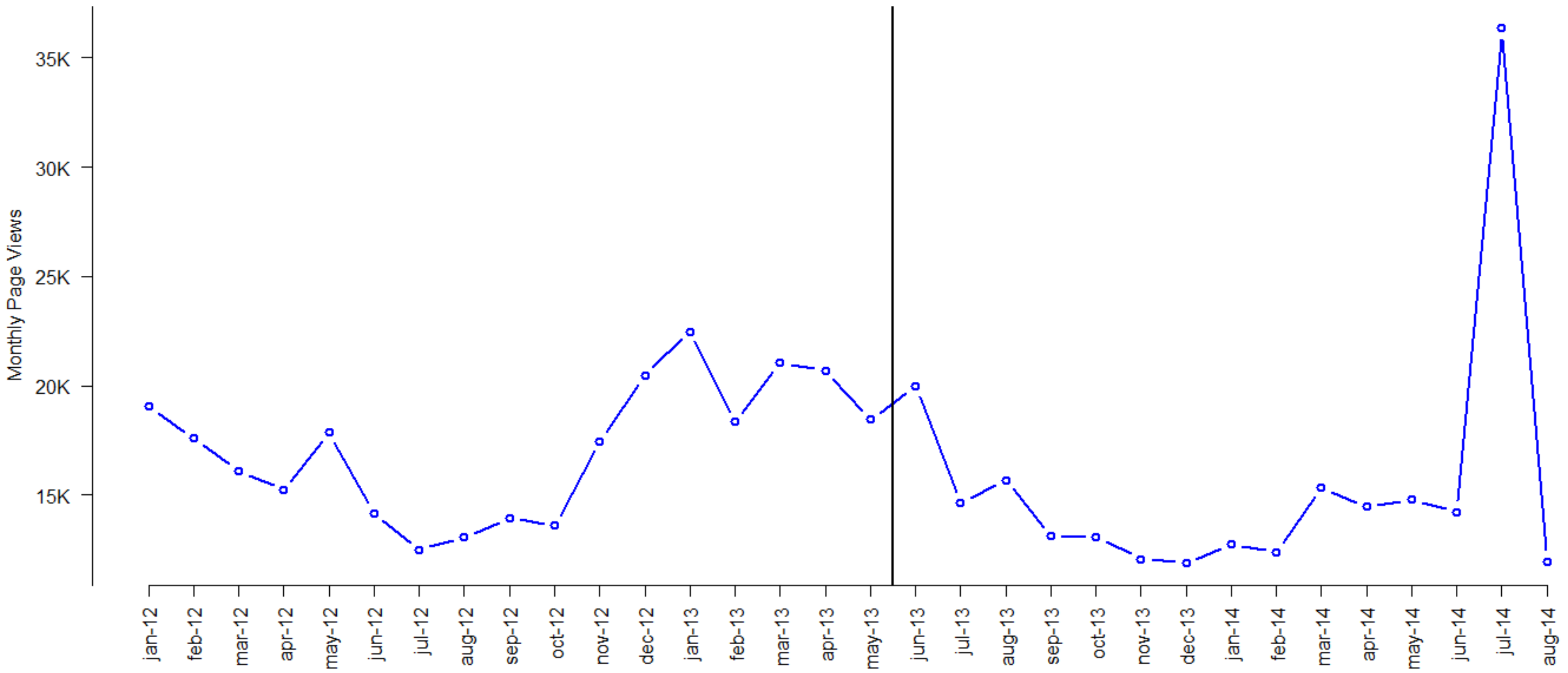


### Security: Page Views for united\_nations



### Security: Page Views for us\_citizenship\_and\_immigration\_s

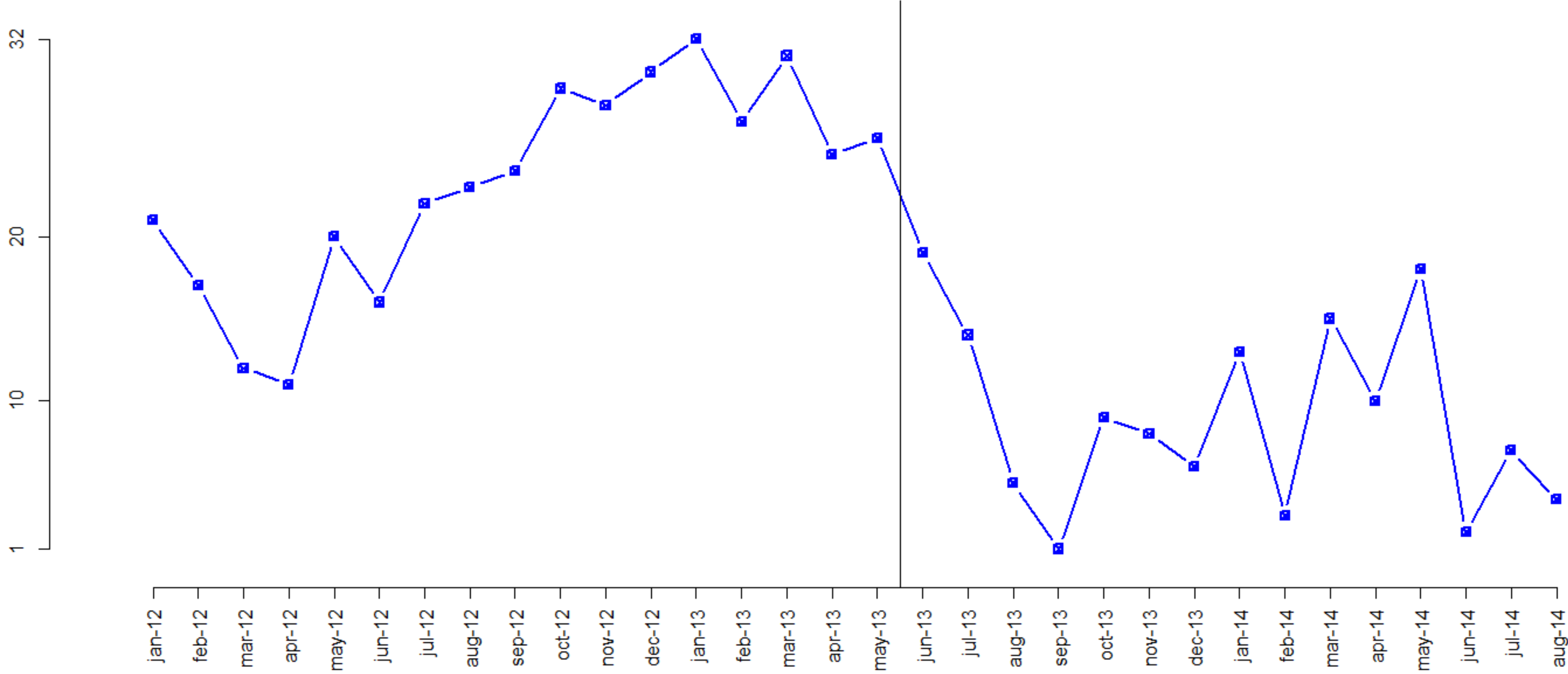




**APPENDIX VII: Page Views for Five Aggregate Comparison Datasets**

### Rank of Views by Month for Control: Global1

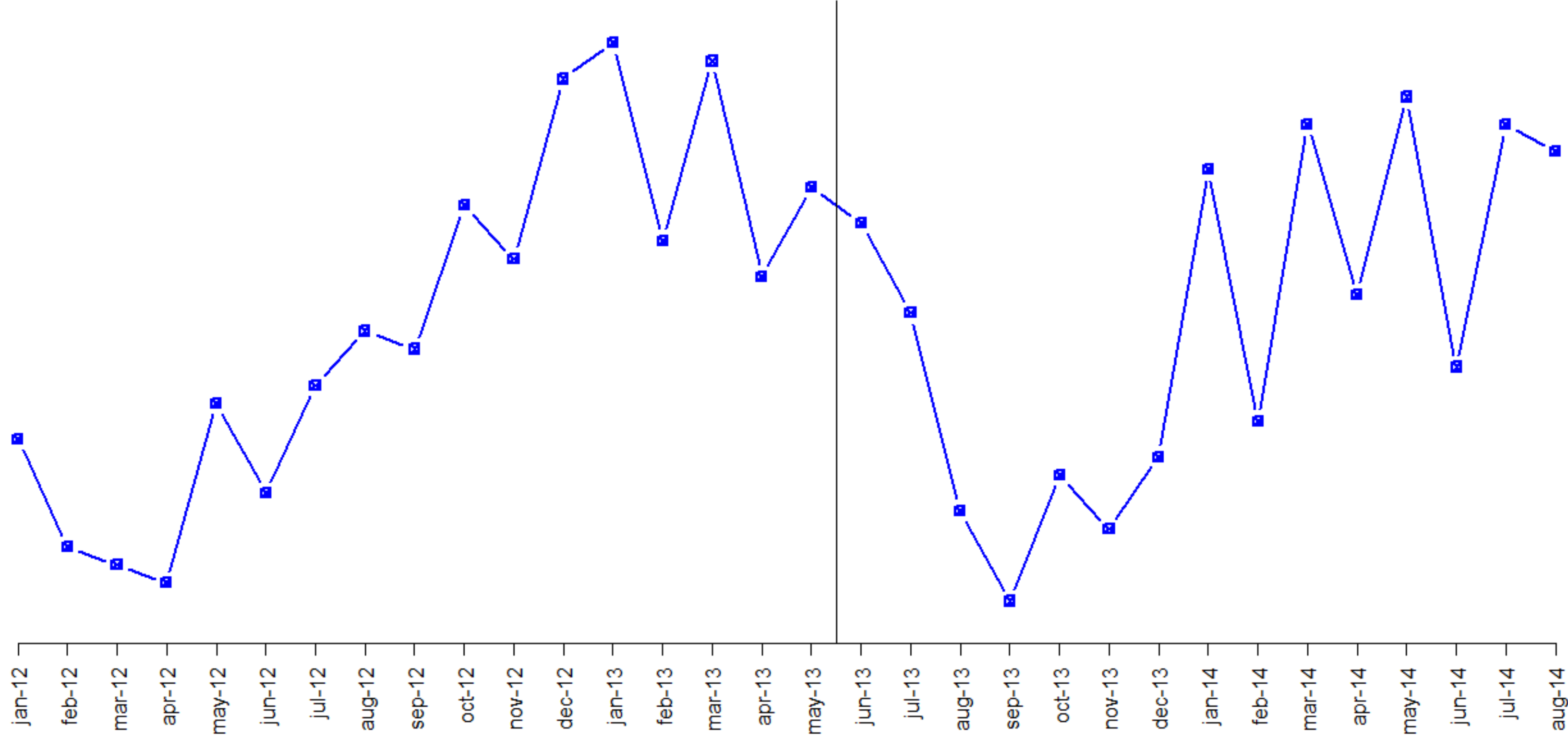
Rank of Page Views by Month: 1 is Lowest and 32 is Highest



### Rank of Views by Month for Control: Global2

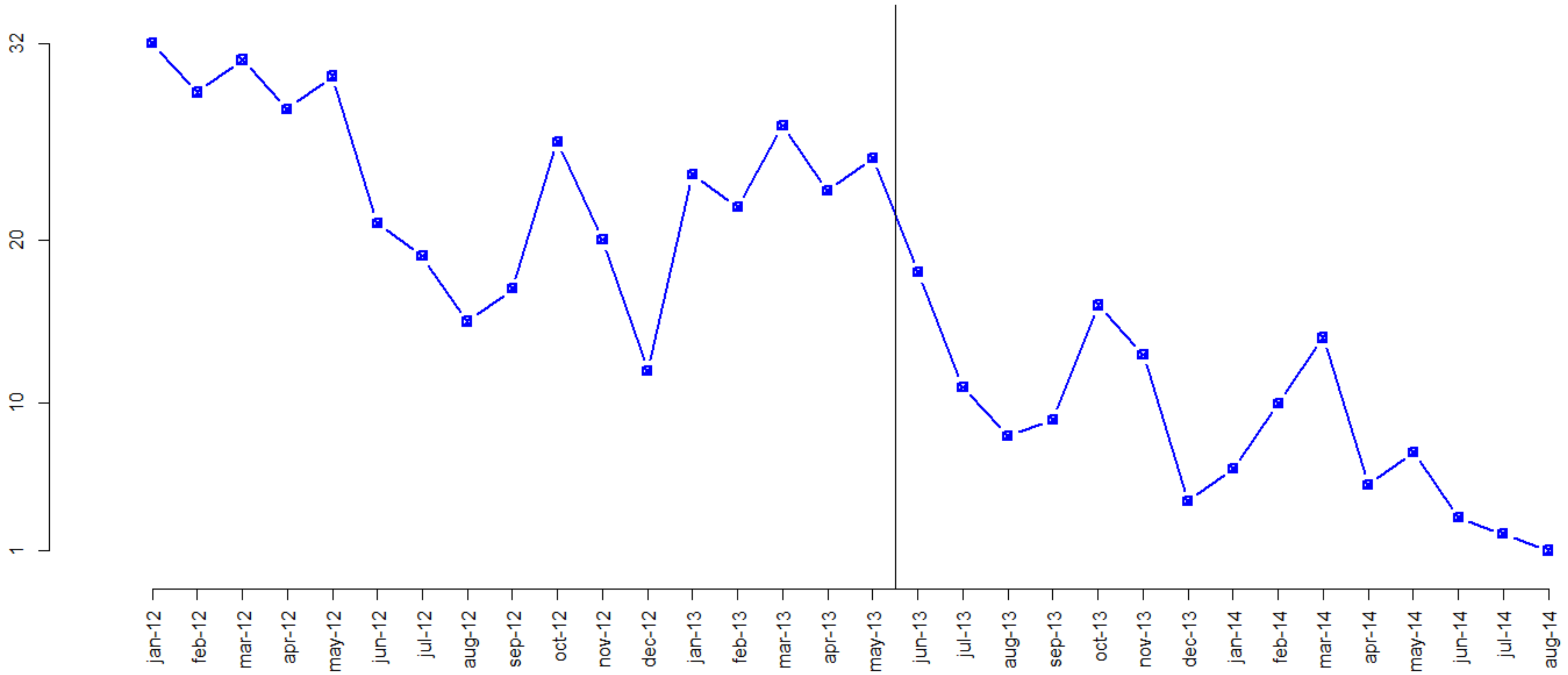
Rank of Page Views by Month: 1 is Lowest and 32 is Highest

32  
20  
10  
1



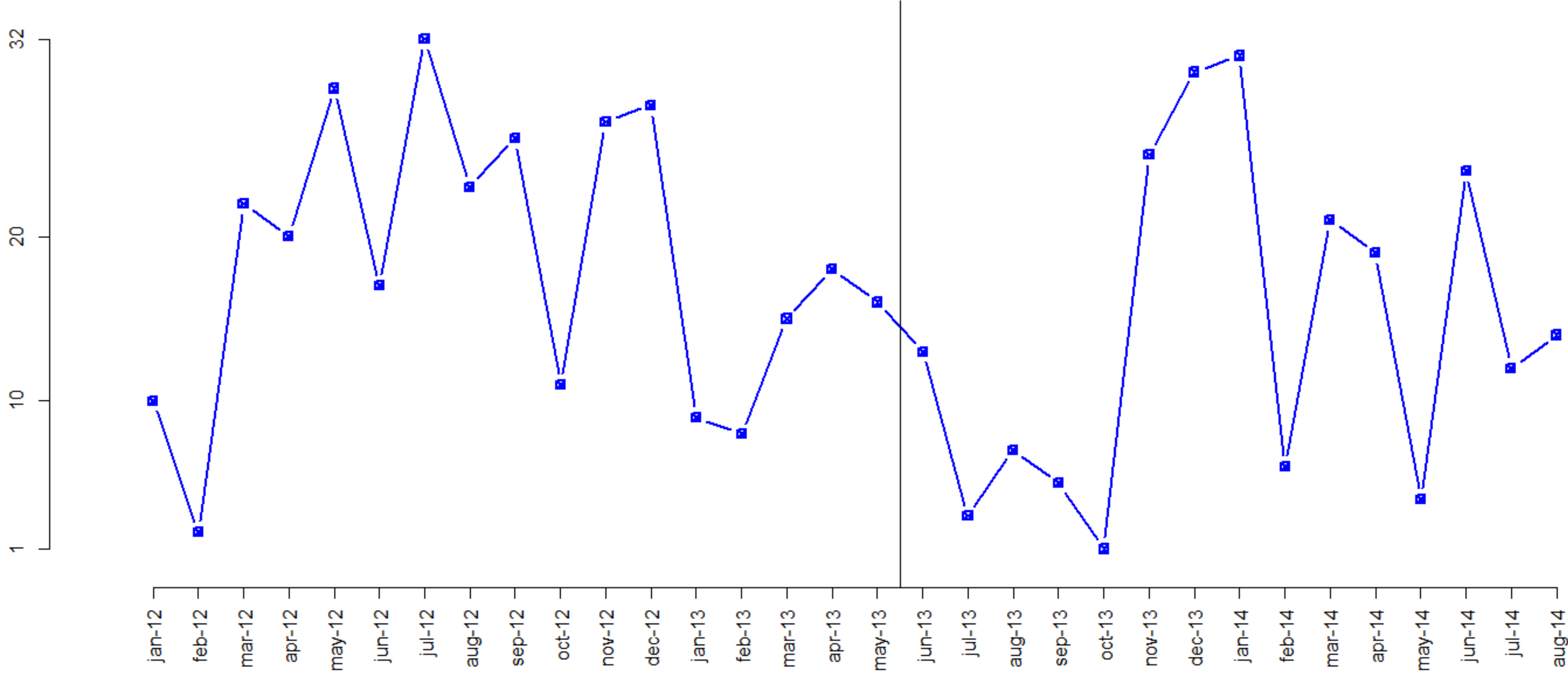
### Rank of Views by Month for Control: Infrastructure

Rank of Page Views by Month: 1 is Lowest and 32 is Highest



### Rank of Views by Month for Control: Popular

Rank of Page Views by Month: 1 is Lowest and 32 is Highest





### Rank of Views by Month for Control: Security

