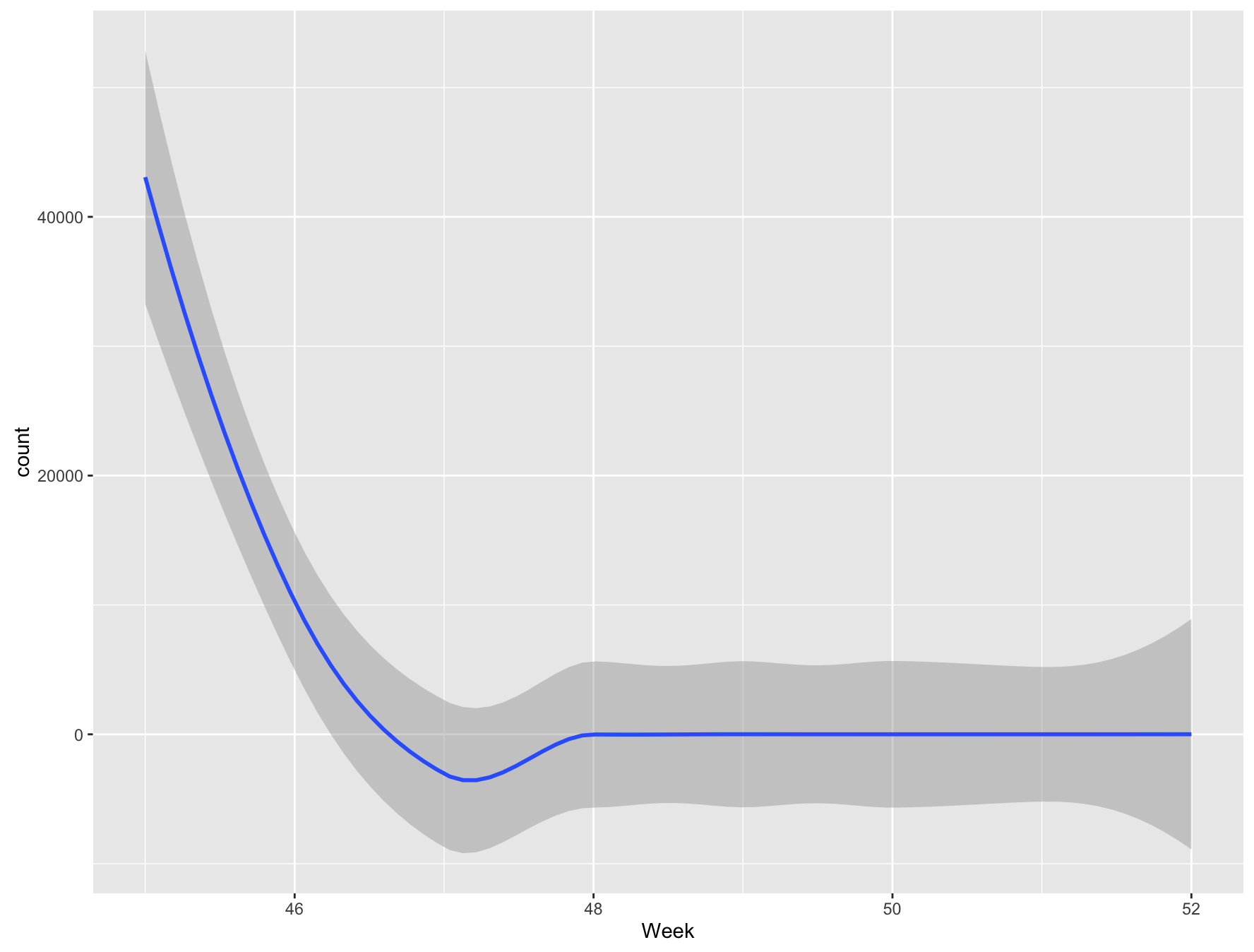
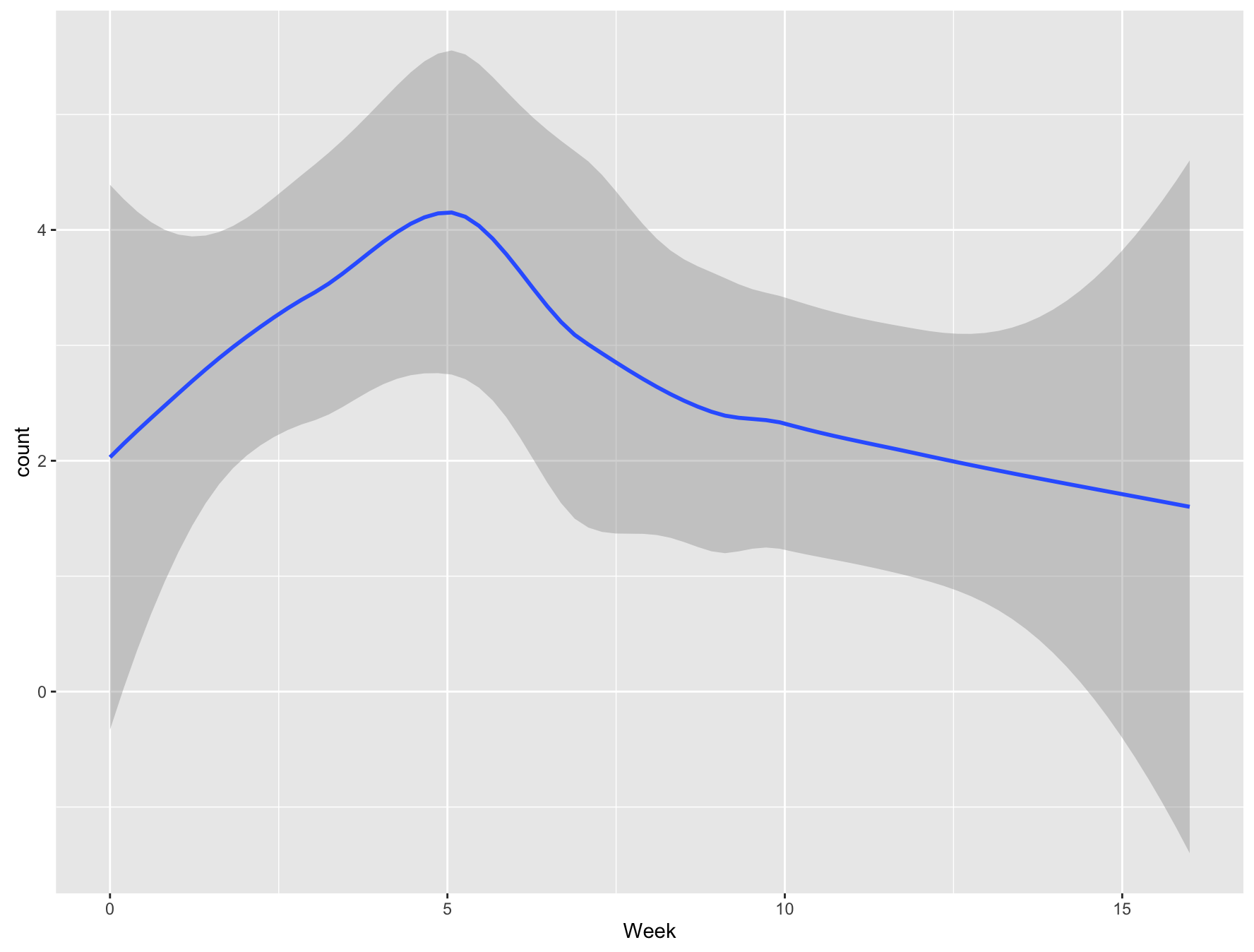
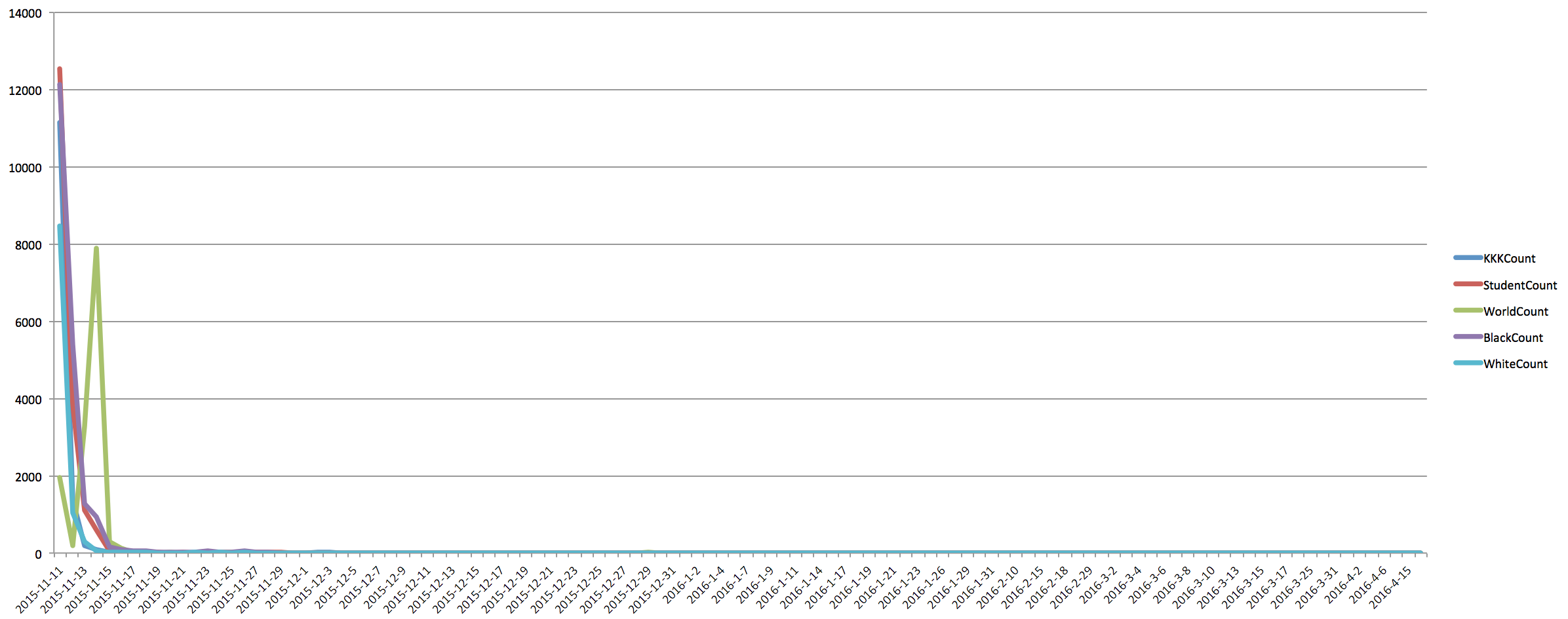
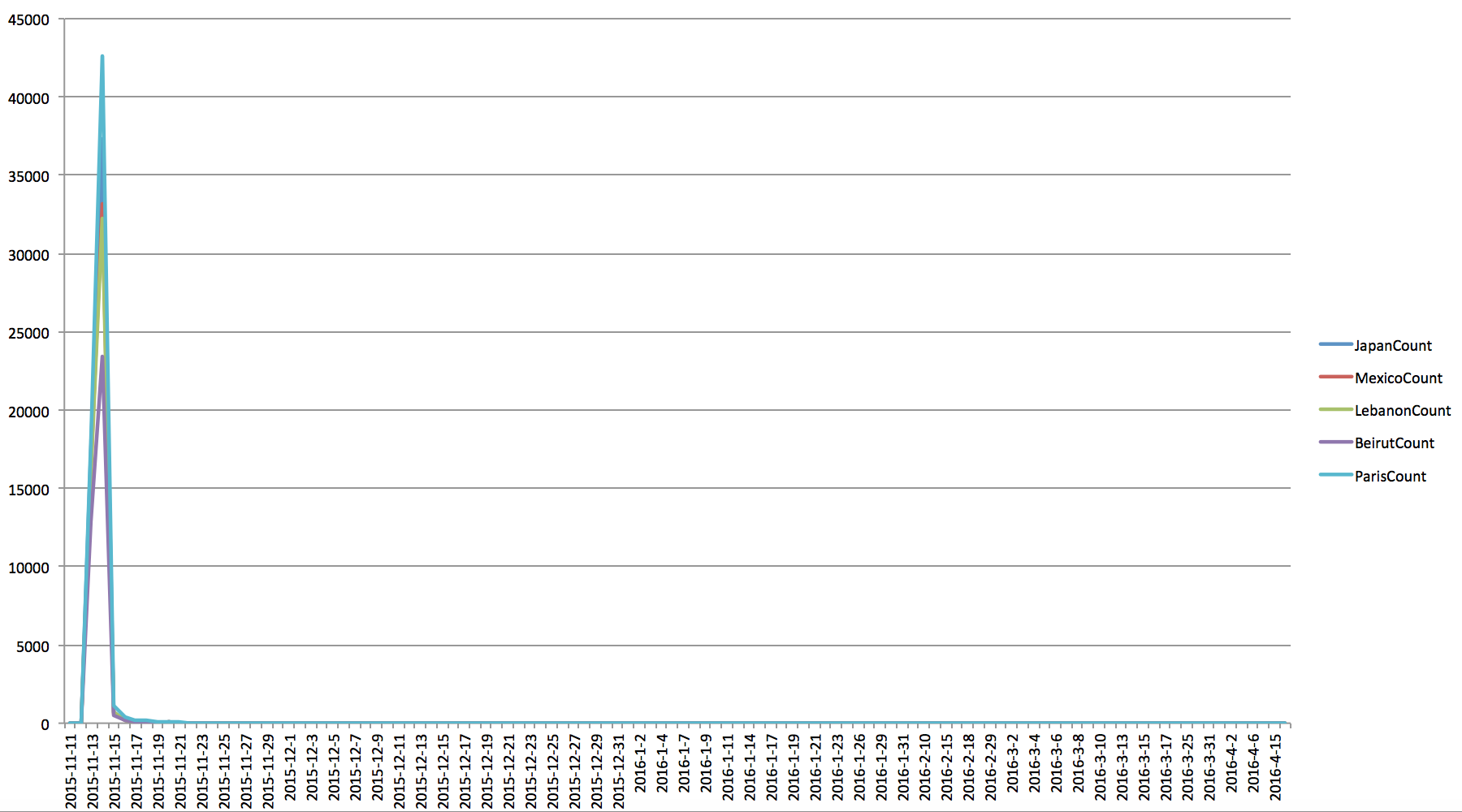
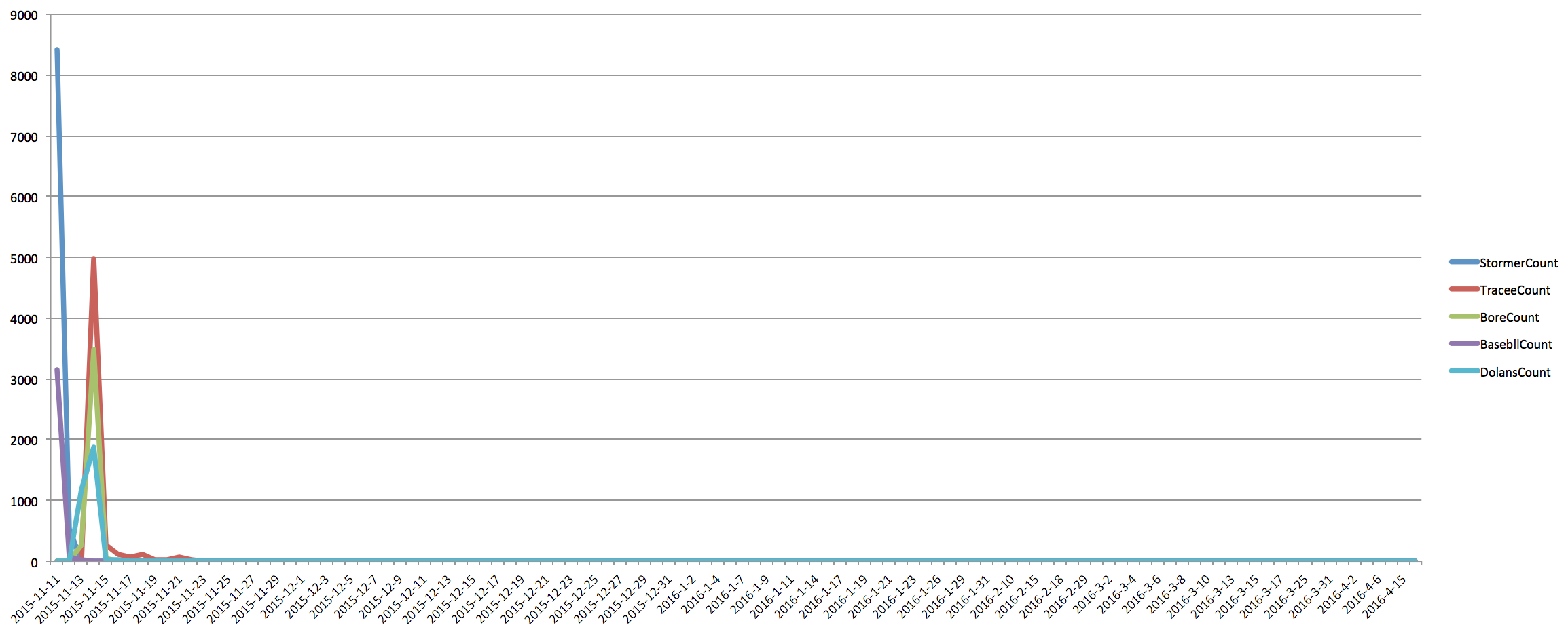
**Github Link:** <https://github.com/jbekcc/Data_Sci_Final_Proj>

**Data Set:**

I looked at #PrayForMizzou fro this analysis.

**Descriptive Statistics**

1. How many tweets are in the collection?
   1. 214,133, 145
2. When do they start?
   1. 11/11/2015 at 4:57:55
3. When do they end?
   1. 4/20/2016 at 9:15:07
4. What is the trend for tweet volume?
   1. 2015
   2. 
   3. 2016
   4. 
5. If you look at the most common words over the lifetime of the search, do you notice any particular trends associated with those words?
   1. 
6. What external events might correspond with the differences in the trends of most common words?
   1. The main event that affected the trends was the individual reactions to the online threats last semester (mainly on Yik Yak) against black students to the point where many people of color did not feel comfortable coming to campus the next day. For the words KKK, white, black, and student they spiked the day of the threats and then quickly died down within 2 days. World on the other hand has very little activity when the other 4 do, yet spikes 2 days later which corresponds with the spikes in the hashtags such as #PrayForParis and #PrayForMexico (among numerous others)
7. What hashtags show up as most prominent in each month of the lifecycle?
   1. 
8. Which twitter users are the most mentioned?
   1. stormer9K - 8528
   2. TraceeEllisRoth - 5749
   3. onlinebore - 3823
   4. whitebaseblldad - 3253
   5. realdolans - 3107
9. How frequently is each user mentioned during each month of the lifecycle?
   1. 
10. What is the relationship between the volume of tweets you selected and the volume of tweets for other collections in the data set?
    1. This data set is tied to a specific example (the online threats from last semester) and so has a ton of activity for the first two days before (as all social media trends do) dying out and having very little to no activity for the rest of the collection. This data has similar spikes at the beginning as ConcernedStudents1950 since the two events are related. However ConcernedStudents1950 continues to have more mentions than PrayForMizzou does after the first few days. It was interesting to note that when tweeting #PrayForMizzou people also included the various international events that were occurring such as #PrayForParis and #PrayForBeirut.

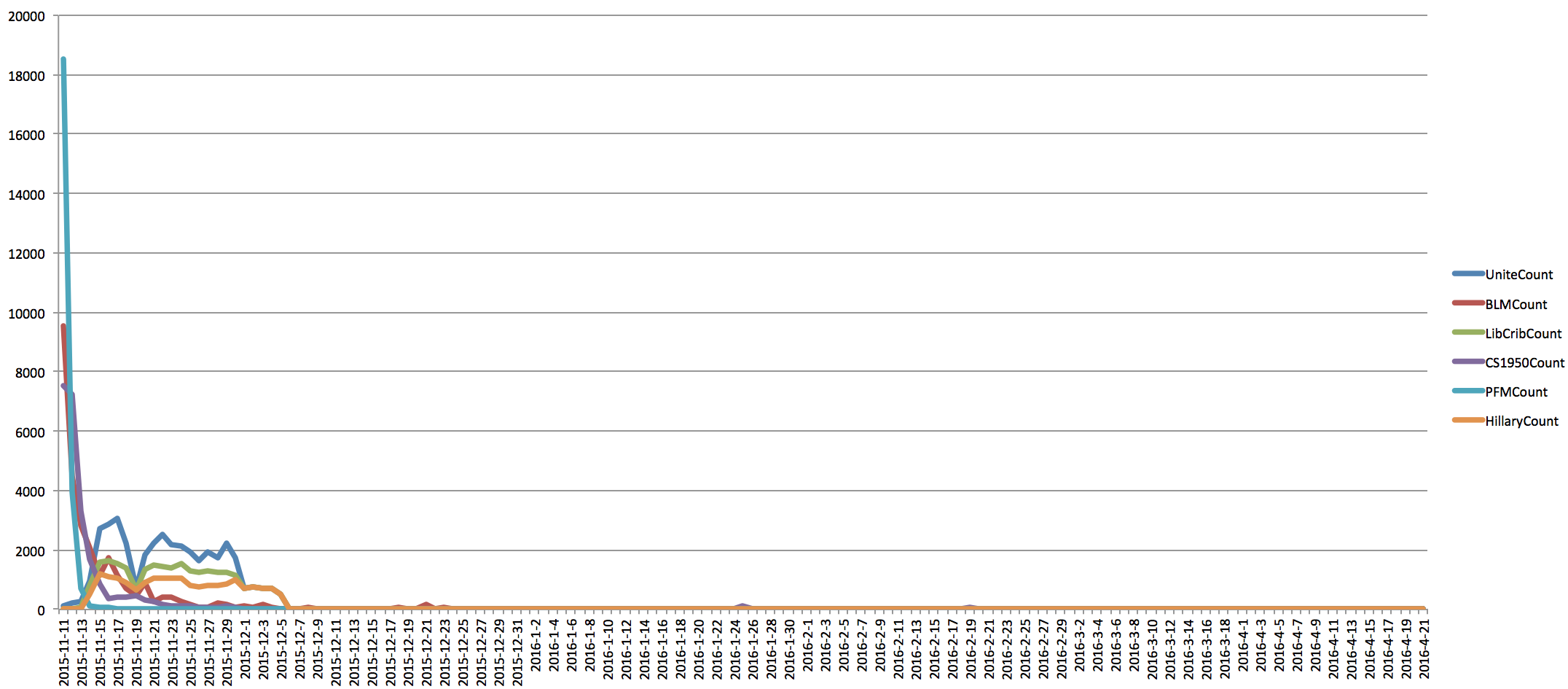
**Problems with this dataset:**

The data was not collected until November 11th which was the night of the Yik Yak threats. That is true for PrayForMizzou and MIZZOU and ConcernedStudent1950 which means that the buildup to that night and the rest of the semester is missing from this set. Since it wasn’t collected until 4:57 that evening there’s data from earlier in the day that is missing and so this is even a subset of the total tweets that were tweeted.

**Research Question:**

How does the activity in this hashtag relate to other hashtags that were used in conjunction with #mizzou

**Answer:**



Not surprisingly #BlackLivesMatter, #PrayForMizzou, and #ConcernedStudent1950 all follow the same patter (for this data set) where they are talked about a lot at the beginning and then within 3-4 days are down to very little to no activity. Oddly, as those hashtags were decreasing in activity, #Hillary2016 and #LibCrib were increasing. After some research it seems that #UniteBlue and #LibCrib are hashtags and accounts that are used to talk about liberal policies and people in government while critiquing republican policies. After looking into these hashtags, they are not hashtags that would have been used by students on campus, which means that a large part of this data is people talking about Mizzou without attending the school or being in the area. It is worth noting that those hashtags, along with #Hillary2016 all followed the same pattern of activity and didn’t rise to a noticeable amount of tweets until after the other hashtags had subsided.

#PrayForMizzou, #BlackLivesMatter, and #ConcernedStudent1950 all follow the trend of how media portrayed what occurred on campus last semester. One issue with the dataset is that the tweets are not collected prior to November 11th which for most datasets is alright however since this one is event driven (by a single event that only occurred once) this dataset is missing the build up and all previous mentions relating to the event, particularly with #BlackLivesMatter and #ConcernedStudent1950. This made the analysis a bit harder since it is harder to see the spike in tweets, outside of the fact that it occurred the first day of the collection.

I had initially wanted to do a sentiment analysis on the #PrayForMizzou data set however I ran into some issues. I have the script made (Sentiment.ipynb) however after running for 6 hours and having no output I had to shutdown the kernel. I had attempted to modify the script to make it more efficient but to no avail. The script is still in the repo so that it can be seen but it is something that I will have to experiment with more over the summer due to not having enough time this week. My original plan was that once I had the sentiment values for all tweets I could start tracking the sentiment over time both in a general sense, separated by specific hashtags, and separated by user mentions to see if any trends exist.