



Dealing with Data: Spring 2016

*Analysis of Public Sentiment on Twitter Versus Debate
Performance of Top GOP Candidates*

May 16, 2016

Team 2: Ankur Chaudhary, Christina Wei, Janul Hernandez, and Peter Meijer

Executive Summary

The 2016 GOP race has become one of the most talked about topics of the year, garnering the public's attention with the debates becoming a televised spectacle. Fox News' debate in Detroit is the highest viewed debate of the year at 16.9 million viewers. This is no surprise with the interest Donald Trump has gathered from the general public. Trump's brash manner and lack of civility has ran counter to traditional political campaigns. Our project tried to see if there was a correlation between how Trump using negative rhetoric during the debates helped his standing with the public in contrast to his rivals.

We took a sentiment analysis of all GOP candidates participating in the last three debates of the GOP nominee trail and compared it to Twitter's response of each candidate. To our surprise we found that Trump's debate performances were not as negative in sentiment as his running mates. Additionally, our expectation for positive public sentiment on Twitter was wrong as well. What we did find was that Trump dominated the digital space in sheer volume of tweets and re-tweets.

Background

Our hypothesis was that Trump using negative rhetoric on the debate stage helped him garner positive sentiment from the public on social media specifically Twitter. Our datasets included:

Debate transcripts from the Washington Post website:

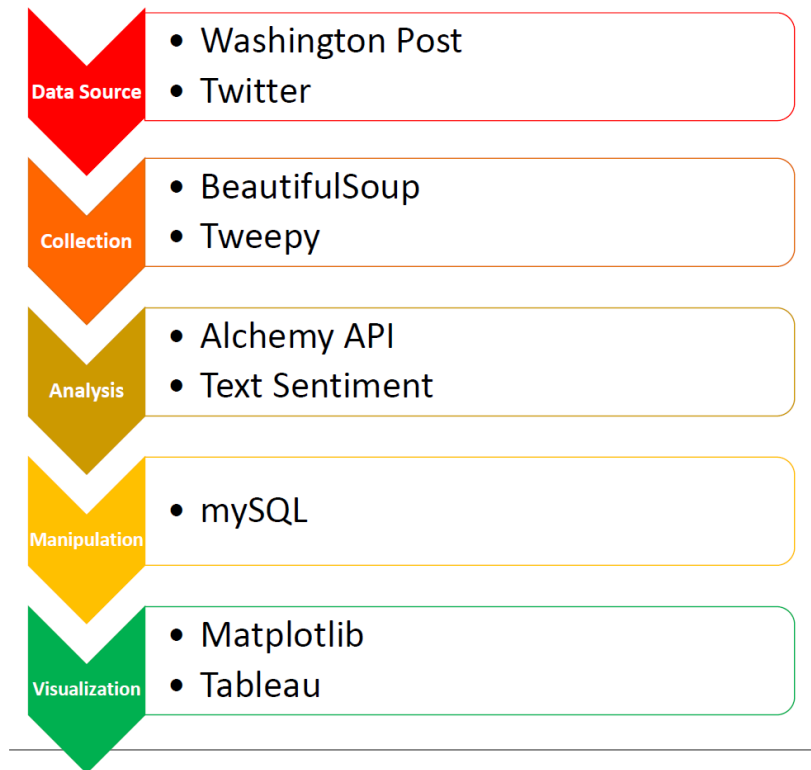
- 1) Debate 1 (Houston, TX, 02/25/2016):
<https://www.washingtonpost.com/news/the-fix/wp/2016/02/25/the-cnntelemundo-republican-debate-transcript-annotated/>
- 2) Debate 2 (Detroit, MI, 03/03/2016):
<https://www.washingtonpost.com/news/the-fix/wp/2016/03/03/the-fox-news-gop-debate-transcript-annotated/>
- 3) Debate 3 (Coral Gables, FL, 03/10/2016):
<https://www.washingtonpost.com/news/the-fix/wp/2016/03/10/the-cnn-miami-republican-debate-transcript-annotated/>

Twitter data was extracted for different candidates using Tweepy API.

Project Description

The workflow process was as follows:

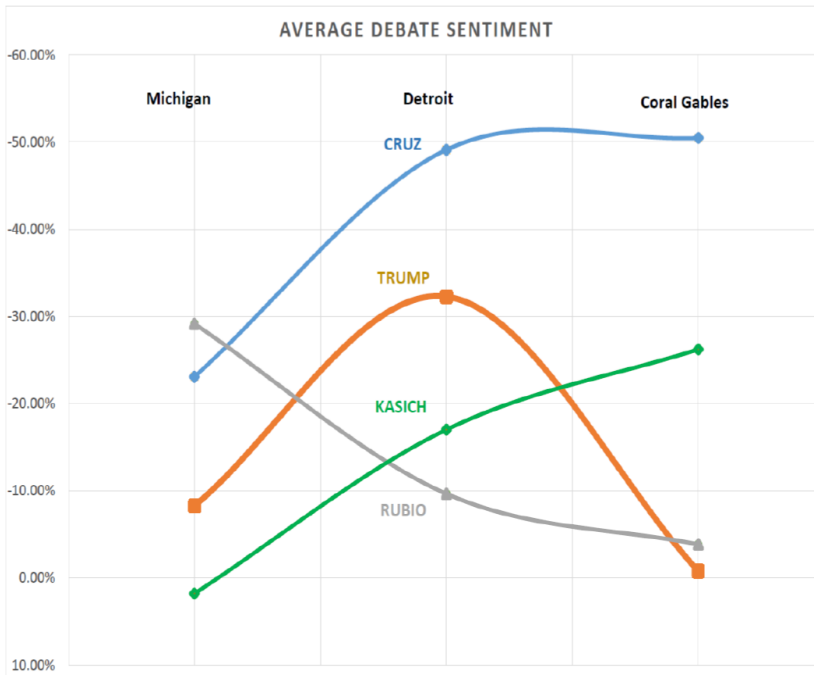
Figure 1: Workflow



- 1) We used transcripts from the Washington Post of three previous Republican Debates (Houston, Detroit, and Miami). With the help of Beautiful Soup (an HTML parser) we scraped these sites to build dictionaries of all words said by specific candidates during each debate.
- 2) The dictionaries were placed in a relational database to find keywords and then sent through Alchemy Sentiment API to get a confidence and sentiment score.
- 3) We made a Twitter app and used its consumer and access keys to extract tweets using query words and hashtags. Twitter only allowed us to access the last 10 days of data free of charge. So we extracted relevant tweets using Tweepy cursor along with API.search. This data set was also sent through a Sentiment API.
- 4) After both data sets were collected we assigned a weighting factor and then combined using SQL. From the sets we performed data visualization using pandas (MATPLOTLIB) and Tableau to showcase the trends and our findings.

What did the Debate Data Say?

Figure 2: Debate Sentiment



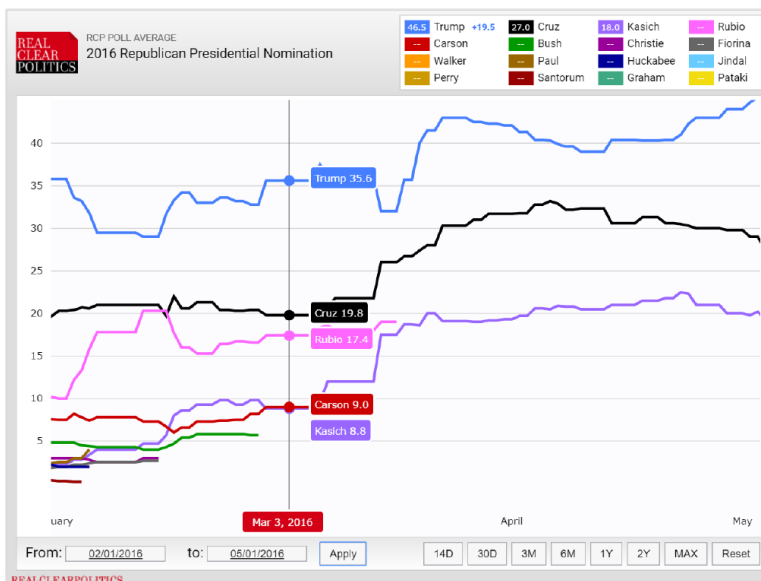
$$Sensitivity = \frac{\sum_{i \rightarrow l} (i_{pos} - i_{neg})}{\sum i}$$

where i = relevance

- Cruz and Kasich get more negative as they get desperate.
- Trump has an outlier negative performance in Detroit
- Rubio continues to be neutral

Figure 2 shows the average debate sentiment of all participating GOP candidates through the 3 debates. One thing that immediately is surprising is the fact that Trump was not the most negative in sentiment. We did notice an outlier performance in Detroit, this was due to the fact that he was being attacked by all 3 remaining GOP nominees. So much so that it dominated the headlines the next day (see figure 3).

Figure 3: Where were we before Detroit?



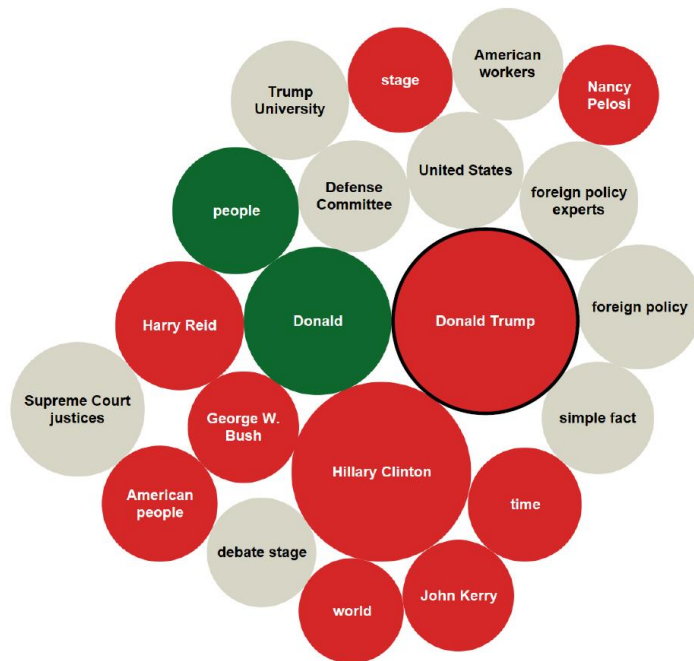
NEWS HEADLINES

- **The Washington Post**
"Trump was attention of attacks at GOP debate"
- **The New York Times**
"Taking On Donald Trump, at Their Own Peril"
- **Rolling Stone**
"Watch every time FOX News attacked Trump at GOP debate."

Due to this increased attack it was no surprise that when we focused on keywords of the three candidates excluding Trump the word "Donald Trump" was the most relevant when aggregated and

had a negative sentiment. The focused attack was so much so that “Hillary Clinton” came in second for relevance and negative sentiment (*Figure 4*).

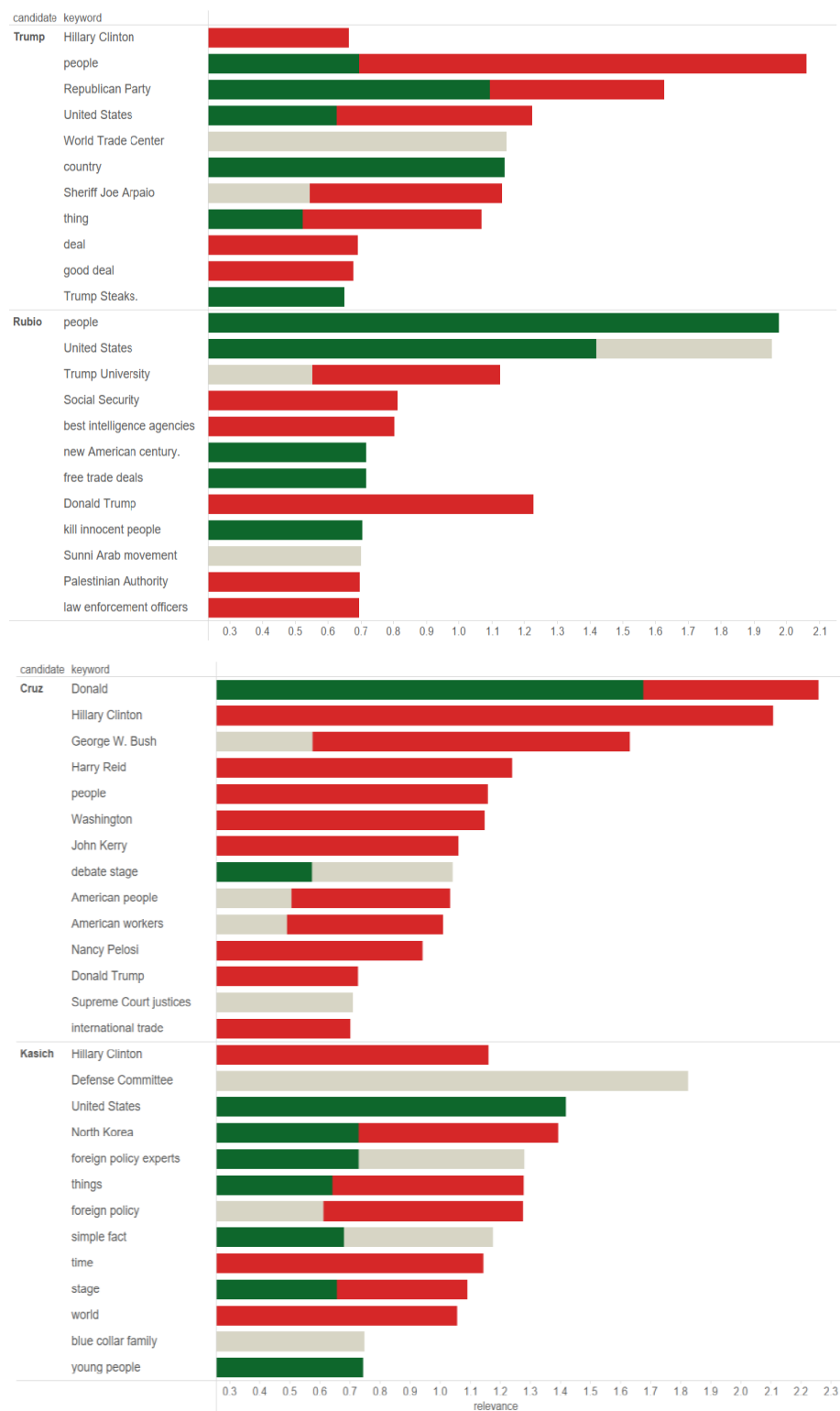
Figure 4: Keywords and sentiment during Detroit debate. (green=positive, red=negative, grey=neutral)



Trump proving he doesn't have small hands

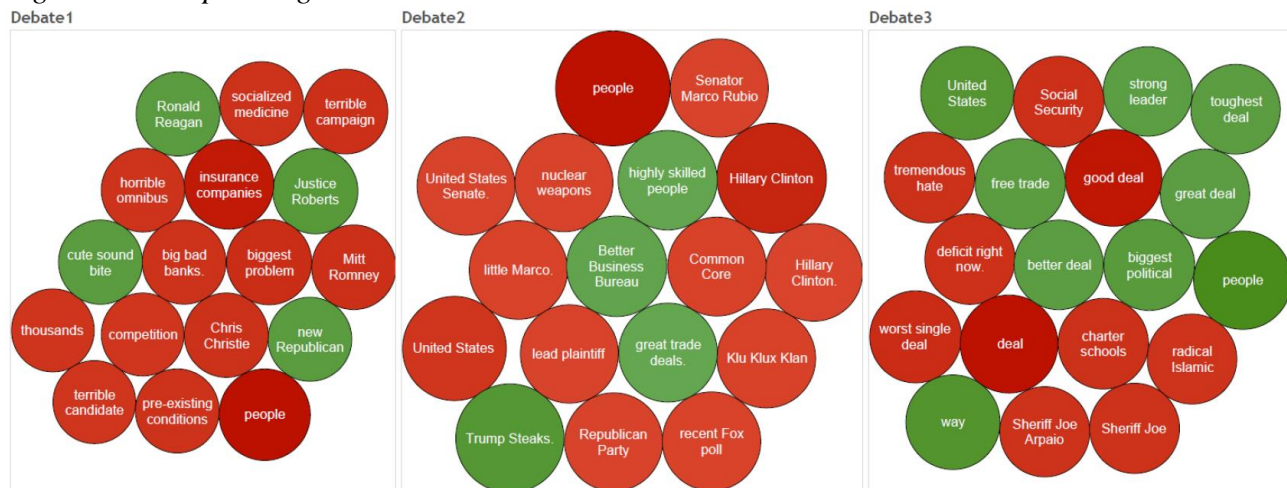
We then looked at keywords the candidates used through all three debates to see if they were continually attacking Trump and what we found further supported our theory that Detroit was an outlier. (*see figure 5 below*)

Figure 5: Keywords of each candidate during all 3 debates.



We finally tried to look through our data set to see what Trump was focusing on that has made him a runaway success. We found nothing astonishing other than the common theme of people not being happy and his general denouncement of the GOP establishment.

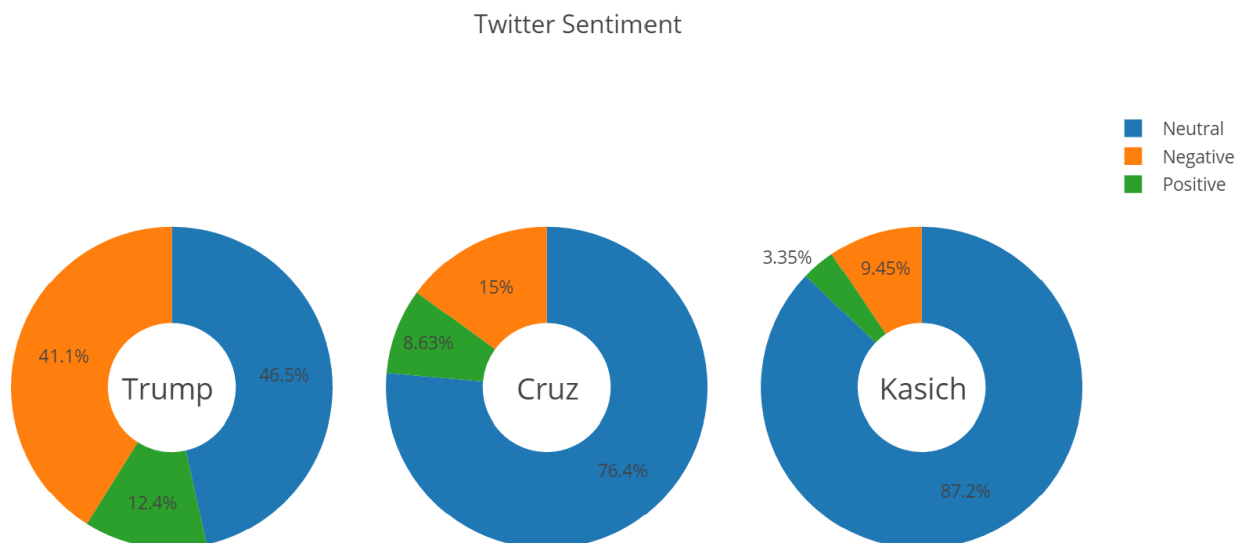
Figure 6: Trump through 3 debates.



What did the Twitter Data Say?

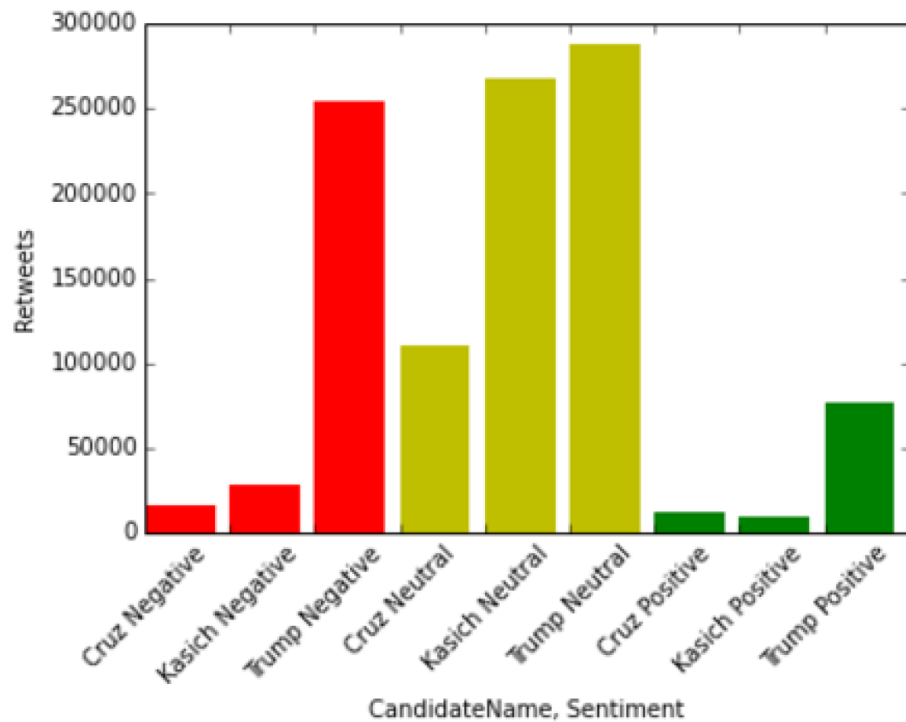
Due to the limitation on Twitter of 10 days to get historical data it was very hard to get a correlation between debate performance and public sentiment. What we were able to see was that Trump contrary to our hypothesis was not being received on Twitter in a positive light. Overall Twitter sentiment was dominated by neutral (46.5%) and negative (41.1%).

Figure 7: Twitter sentiment.



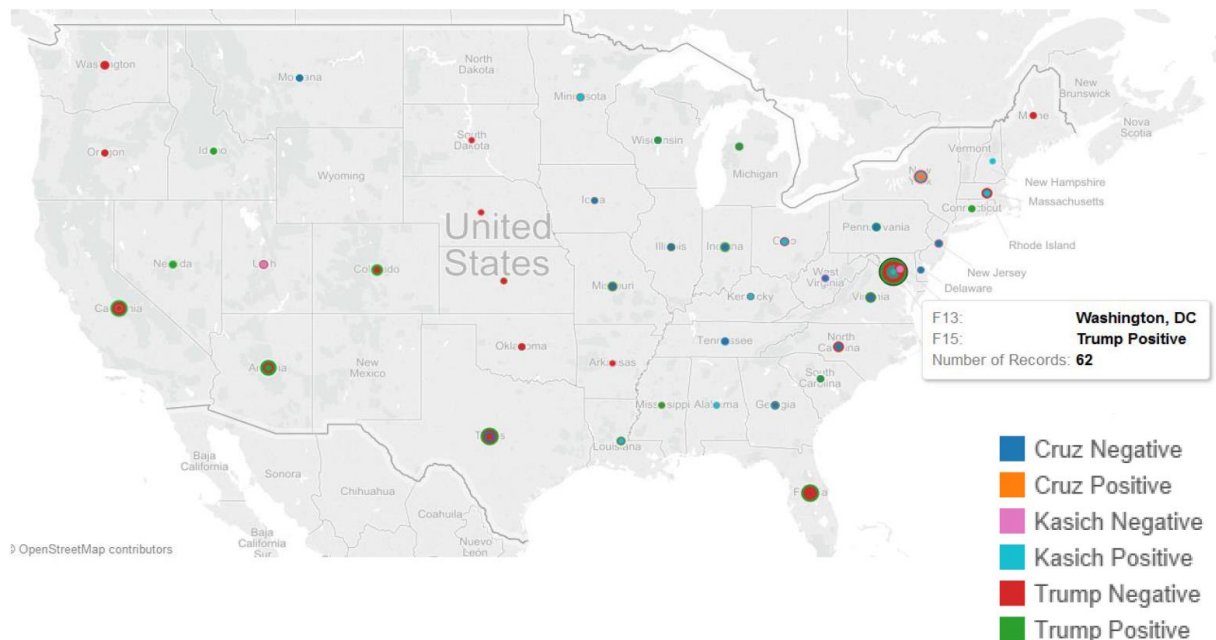
Where Trump did succeed was in the sheer volume of tweets and re-tweets he was getting, regardless of sentiment.

Figure 8: Tweet count by volume and sentiment.



We also looked at the distribution of tweets in the U.S. and found nothing telling.

Figure 9: Distribution of tweets.



Conclusion

Our hypothesis erred on two key aspects:

- Assumption: Trump most negative debate sentiment
- Assumption: Trump most positive social media support

Ultimately, Trump's debate sentiment *improved* over the course of the debates while his primary challenges Cruz and Kasich grew more negative. Trump generated the most social media, but also accounted for the **most negative** social media due to heightened general scrutiny of controversial comments

Future Opportunities

Our approach yielded additional research opportunities:

1. Using Twitter sentiment analysis and location data of tweets, examine correlation between change in tweet sentiment pre/post-debate and whether debate sentiment had an impact on polling data relative to state primary results.
2. For the debates between Donald Trump and the democratic candidate, apply a similar analysis over several debates to chart the impact of debate sentiment for a more constrained data sample.