Analysis of President Trump's Tweets

Campbell Alden, Oliver Fishstein, Dylan Pincus, Kimberly Whitney

College of Engineering, Northeastern University
Emails: { alden.ca, fishstein.o, pincus.d, whitney.k } @husky.neu.edu

Introduction

President Donald J. Trump has been using a Twitter account to voice his thoughts, feelings, opinions and more since May of 2009 - far before he became President of the United States of America. Although he was given an official Presidential Twitter account upon his inauguration in 2017, he still uses his personal account for tweeting his response or opinion to everything. The official @POTUS account primarily consists of retweeted content from his personal. Having tweeted over 33,450 times total and 3,118+ of those since becoming President, Trump's account has become the way for Trump to bypass typical means of presidential statements and speak directly to foreign leaders, citizens, the press, etc. himself [5]. Since more than 2/3rds of Americans are getting at least some of their news from social media, Trump's twitter tirades have a much larger impact on the public than any president before him [3]. In addition to the public, Trump regularly mentions his meetings with or makes direct statements or insults about foreign leaders and countries in his tweets [10]. With such consequence depending on frequent 280-character statements, analyzing what messages or trends come directly from him can help assess the validity or motivation behind his statements. In this paper, we will be discussing and displaying our approach to visualizing different aspects of Trump's tweets including sentiment analysis over time, device used throughout the day, geographical locations of places mentioned, keywords, and more. The resulting visualizations can be found hosted online in an interactive format or in Appendix A.

Motivation / Real-World Applications

Since the early days of Barack Obama's campaign, Twitter has steadily grown to become a powerful platform for politicians to reach their constituents. Never before has it been easier for a candidate to reach millions of voters, simply by typing on their phones. Leaders across the world have embraced Twitter, including Vladimir Putin, Benjamin Netanyahu, and David Cameron [7].

Throughout his campaign and presidency, Twitter has been Trump's primary platform for all public discussion and announcements. The Trump administration has released approximately 2450 public statements throughout Trump's 15 months in office. By comparison, Trump tweets approximately 375 times per month, or 5625 times in the same 15 month span [8]. This massive amount of public statements from the president has a huge impact on politics and economics on a global scale. On multiple occasions, Trump has tweeted complaints about individual companies, including Toyota, Amazon, Nordstrom and many others. These callouts have a direct response on the open market, typically causing around a 1% fall in stock price for the targeted company. Trump's numerous tweets on Mexico have had caused immediate drops in the currency exchange rate, plummeting the price of the Mexican Peso [7].

Trump tweets about issues ranging from foreign affairs to healthcare, with seemingly no topics too sensitive for the public platform. He routinely tweets information about public policy initiatives, communicates with foreign leaders, and directly criticizes or mocks anyone he cares to. It is difficult to keep up with the torrent of comments made by the sitting president, and even more difficult to determine

how much any of the individual tweets really matter. With our visualization work we hope to provide insights on Trump's activity and impact, allowing people to easily see the major trends and characteristics of his Twitter persona. The visualizations show Trump's tweeting tendencies and focuses, which help to shed light on the mind of our president and what major decisions he may soon make.

Related Work

As the President of the United States, everything Trump does receives a great deal of attention and scrutiny, and his tweets are no exception. Several studies and analyses have already been conducted, and several other data visualization have also been created off of this work. Some of the most notable work includes a visualization of Trump tweets characterized by category, time, and retweets by Teen Vogue [2], an analysis of Tweet motivation (self-promotion, critique, and opinion) by Quartz magazine [4], and a graphic analysis of Trump tweets focused on policy issues by Politico [9]. Our visualization work sought to augment and extend the previous and ongoing studies done on Trump's tweets, to add extra dimensions of analysis to further extract impact and meaning from the data.

Data Source and Background

Donald Trump's Twitter account is publicly accessible on their website [6]. A website titled Trump Twitter Archive created by Brendan Brown makes gathering all of the data from each tweet incredibly easy [1]. His site polls Trump's account for new content every 1 minute and saves the data every hour. Since the account is checked so frequently, the dataset even contains tweets that Trump deleted (starting after September 2016). Having scraped the data from Twitter itself, the archive contains the time the tweet was sent, the text content device the tweet was tweeted from, the number of retweets and favorites the tweet received, and if the tweet is a retweet of someone else or not. The last binary parameter proved extremely helpful in allowing us to filter out retweets since they may not directly reflect his personal thoughts, feelings, or words.

For this project, we collected, analyzed, and visualized Trump's tweets since January, 2014 - around 20,000 tweets total. This allowed us to compare the sentiment or popularity of Trump's statements before he declared his candidacy in 2015 and before he became president in 2016. Although this excludes a few years of data before 2014, we were limited to 20,000 API calls in our analysis and determined that the importance of his most recent tweets outweigh some statements he may have made before he was on the political radar. The rest of the data used to create our visualizations was collected through the APIs mentioned in our analysis based on these tweets alone.

Analysis

Due to the sheer volume of the corpus, the analysis of Donald Trump's tweets was necessarily done computationally. Using Python and a collection of libraries, we created a number of scripts for processing the data from the twitter archive. Our goal was to analyze both the content and sentiment of the tweets so that we could give a varied and deep impression of Trump's online persona. This analysis was comprised of three steps: preprocessing, analysis, and data aggregation.

The data from the twitter archive is stored in a comma separated value (csv) notation. In order to get the data into a format we could work with in Python, we generated object representations for each of

the tweets by reading in the csv file and mapping the values objects using a custom class. Using this representation, the tweets were fed into our analysis function.

Our scripts are designed in such a way that any form of analysis can be performed on the data provided you match a certain input and output contract. We designed analysis functions to utilize the Indico.io API to find the sentiment, political leaning, organizations and places mentioned, and a twitter engagement score for each tweet. After each analysis step, the new data is stored with its respective tweet.

The final step in the data analysis step was to prepare the data to be analyzed by our chosen tools. We set out to visualize the data using word charts, maps, and graphs. The frequency of words and organizations were counted by looking through each tweet and corresponding sentiment data to create histograms. Likewise, the map visualizations relied on counting how often locations were mentioned. The remaining information was collected into a table and written to a CSV file for easy importation in Tableau.

Visualization Techniques

Our visualizations focused on three topics: word frequency, location, and time. Our two word frequency visualizations were designed to quickly and easily demonstrate the major focuses found throughout all Trump tweets. The word clouds were shaped to resemble elements of President Trump so that it would be immediately apparent that the content they displayed was connected to him. The first word cloud featured Trump's most tweeted words across all of his tweets. It serves as the first visualization we present, designed to give viewers a first impression of the topics and ideas that Trump deems most important. The second word cloud focuses on which entities Trump references most frequently in his tweets. This visualization allows viewers to see the news and political organizations Trump seems to care about most. In this cloud, entities are colored by their function, such as news organizations being red.

Our next set of visualizations displayed the locations referenced by Trump tweets. This data is displayed as geographic visualizations and shows, either by color density or size, the frequency with which he mentions both domestic and foreign locations. For the domestic visualizations, we show a comparison of this tweet density map along with a map showing typical party affiliation to allow viewers to compare the two.

Our final set of visualizations analyzed the date and time of Trump's tweets over multiple dimensions. Several visualizations were made displaying Trump tweets over time, each modeling a different element of sentiment analysis. These visualizations also included marks to show when important events in the Trump campaign occured, to explain certain changes in the analysis. Visualizations were also made showing the most frequent time of day that Trump tweets were sent, as well as tracking the source of these tweets. The tweets came from multiple different devices, some Trump's personal devices and some run by his staff, and so a comparison of the time of day each device allows viewers to see which sentiments are coming directly from President Trump.

Challenges and Solutions

The bulk of the challenges with this project were the limitations of existing text analysis tools. Sentiment analysis is still far from perfect and works best when analyzing the sentiment of text that is

very cut and dry positive or negative. Unfortunately, working with the tweets from Donald Trump we learned that most sentiment analysis services still struggle with analyzing sarcasm. While Indico.io does a good job overall, it was still evident in the scoring of some tweets as positive that the sarcasm wasn't being analyzed properly. This can only improves as the sentiment analysis techniques continue to improve.

Another issue was that when analyzing tweets for location, the location probability was not always correct. When running a tweet through location analysis, it would return the location and the probability that it is actually a location. In order to ensure that we had enough data to work with, we selected a threshold of 50% certainty that it was a location to use the data, but still found that the Indico.io API was recognizing people or other things as locations. In order to handle this, manually post processing of the data was needed to remove data points that did not refer to locations.

The last issue was the timestamp format provided by Twitter in the JSON data set was abnormal and not recognized by built-in date and time parsing services. An example of the date and time format provided by Twitter is "Mon Jan 01 13:37:52 +0000 2018", and while most of it is easy to parse, the inclusion of the email time zone indicator ("+0000") made it so we had to remove that information from each timestamp before converting to a DateTime object to be used by Tableau. This object was used to provided breakdowns of the data by time of day and specific date with averages per day, week, month or year [Appendix A].

Future Work

As long as Donald Trump is around, we are always going to continue getting new content from him. Sentiment analysis, impact, and keywords are only a few areas we can see trends or analyze impact out of the entire realm of text-analyzation possibilities. Since this project only analyzes Trump's tweets without consideration of what one would consider "normal," comparing against a larger sample set will be our next step in this project. First, we will analyze Trump's sentiment, impact, and keywords with those of previous presidents (e.g. Obama) or other world leaders on Twitter. With this, we could see if his overall sentiment is similar to the average, or if he is traditionally more negative or positive. We can also generalize this analysis to include the sentiment of Democrats vs. Republicans, sentiment of statements on foreign and domestic tragedies, and more.

Once we have baseline data to compare Trump's statements to, we plan to next work on finding trends between different entities. For example, we know he mentions China a lot in his tweets so we can then analyze the sentiment of these China tweets and see if these mentions are more largely negative or positive. To add a third dimension, we can account for time. Has he always mentioned China in a positive/negative light? If not, was there a certain event that caused this change? Ultimately, our goal at the end of this project is to have enough mapped trends in his tweets to be better prepared and for Trump's next impactful Twitter tired and hope that informing the public of these predictable abnormalities will help prevent large financial or societal events as a result of a tweet.

Division of Work

All team members contributed equally to the two presentations and report. For the rest of the project, we divided work into three stages for this project: data acquisition, data processing, and data visualizations. Dylan and Kimberly located and downloaded the initial dataset of Donald Trump's tweets since 2014 to allow us to analyze his tweets throughout the campaign process into him taking the

presidency. From there, we split the different parts of data processing in order to hit our visualizations. Using the Indico.io API, Campbell wrote the initial script in order to process all of the tweets after loading them into a local python "Tweet" object that Oliver helped developed based off the important attributes to retain from the initial JSON object. Dylan, Kimberly, and Campbell ran the tweets through the Indico.io API in order to collect data on sentiment, political leaning, organizations and places mentioned, and the twitter engagement score for each tweet. Oliver worked on parsing the date and time into a usable DateTime object to enable analysis of tweets over time and throughout the day. In order to combine all of this data, Campbell and Oliver worked on a script to combine the results from all Indico.io runs.

For the data visualizations, we divided them up into three categories: word frequency, location, and time. Campbell and Dylan did the visualizations showing the most used words and mentioned organizations. Kimberly did the visualizations showing the US states Trump has mentioned and the World Places. Lastly, Oliver worked on the five Tableau visualizations showing the sentiment, engagement and political sentiment of Trump's tweets over time and frequency of his tweets throughout the day including a breakdown by device.

Conclusion

As long as Donald Trump is still tweeting, there is information to be gained by analyzing his tweets. With over 51.1 Million followers and an even wider reach worldwide due to news reports of his statements, the different meanings, sentiments, words, organizations, and more will have an impact on how citizens of every culture interpret them. We learned in this project that his high engagement with Twitter mainly consists of using words to describe himself or news organizations he favors or dislikes greatly. He also tweets mainly after watching or listening to the news himself, typically from a personal cell phone when it's late at night. His statements have maintained a conservative stance over time but starkly increased during important election dates implying his deliberate intent on catering to his platform. Additionally, he focuses more on mentioning Democratic states and countries that oppose America - implying his obsession with those who dislike him or America. Whether this data can be used for action or just for informational knowledge is up to the viewer: either way, the official political statements of the leader of America should be analyzed rather than taken at face value.

References

[1] B. Brown, "Trump Twitter Archive", Trumptwitterarchive.com, 2016. [Online]. Available: http://www.trumptwitterarchive.com/. [Accessed: 11- Apr- 2018].

[2] B. McNamara, “We Analyzed A Full Year of Trump Tweets So You Wouldn't Have To,” Teen Vogue, 22-Jan-2018. [Online]. Available: https://www.teenvogue.com/story/trump-tweet-analysis. [Accessed: 22-Apr-2018].

[3] E. Shearer and J. Gottfried, "News Use Across Social Media Platforms 2017", Pew Research Center's Journalism Project, 2017. [Online]. Available: http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/. [Accessed: 10- Apr- 2018].

[4] N. Sonnad, “Linguists reveal the evolution of Donald Trump's tweets over time,” Quartz, 20-Dec-2017. [Online]. Available: https://qz.com/1153080/trump-tweets-analyzed-by-linguists-reveal-five-characteristics/. [Accessed: 22-Apr-2018].

[5] T. Collins, "Trump’s itchy Twitter thumbs are redefining ‘modern” presidents", *CNET*, 2018. [Online]. Available: https://www.cnet.com/news/donald-trump-twitter-redefines-presidency-politics/. [Accessed: 10- Apr- 2018].

[6] "Donald J. Trump (@realDonaldTrump) | Twitter", Twitter.com, 2009. [Online]. Available: https://twitter.com/realDonaldTrump. [Accessed: 09- Apr- 2018].

[7] “How Does President Trump's Twitter Use Impact Forex, Markets And Stocks? - FXCM,” FXCM Insights, 23-Aug-2017. [Online]. Available: https://www.fxcm.com/insights/president-trumps-twitter-impact-forex-markets-stocks/. [Accessed: 09-Apr-2018].

[8] “Trump on Twitter: A history of the man and his medium,” BBC News, 12-Dec-2016. [Online]. Available: http://www.bbc.com/news/world-us-canada-38245530. [Accessed: 10-Apr-2018].

[9] “I. You. Great. Trump.*,” *POLITICO Magazine*, 26-Apr-2016. [Online]. Available: https://www.politico.com/magazine/gallery/2016/04/donald-trump-twitter-account-history-social-media-campaign-000631?slide=4. [Accessed: 22-Apr-2018].

[10] "What has Trump said about your country?", BBC News, 2018. [Online]. Available: http://www.bbc.com/news/world-us-canada-39732845. [Accessed: 07- Apr- 2018].

Source Code Repo: https://github.com/kimception/TrumpTweetsVis

Google Sites of Visualizations: https://sites.google.com/view/trumptweets/home

Appendix A: Visualizations of Trump's Tweets

Figure 1 - Organizations and Entities Most Frequently Mentioned in Trump's Tweet.

Figure 2 - Words Most Often Used in Trump's Tweets Since January 1, 2014

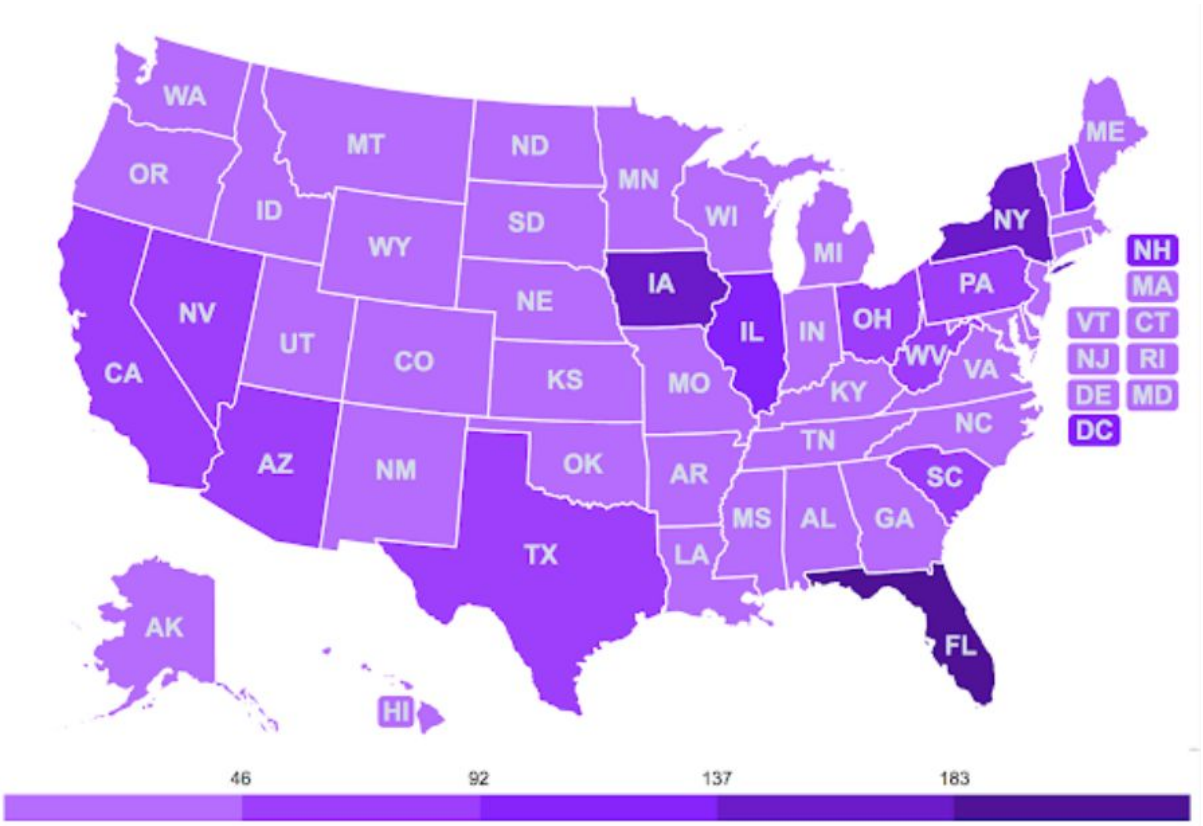

Figure 3 - States Mentioned in Trump's Tweets, Darker by the Amount of Times Mentioned

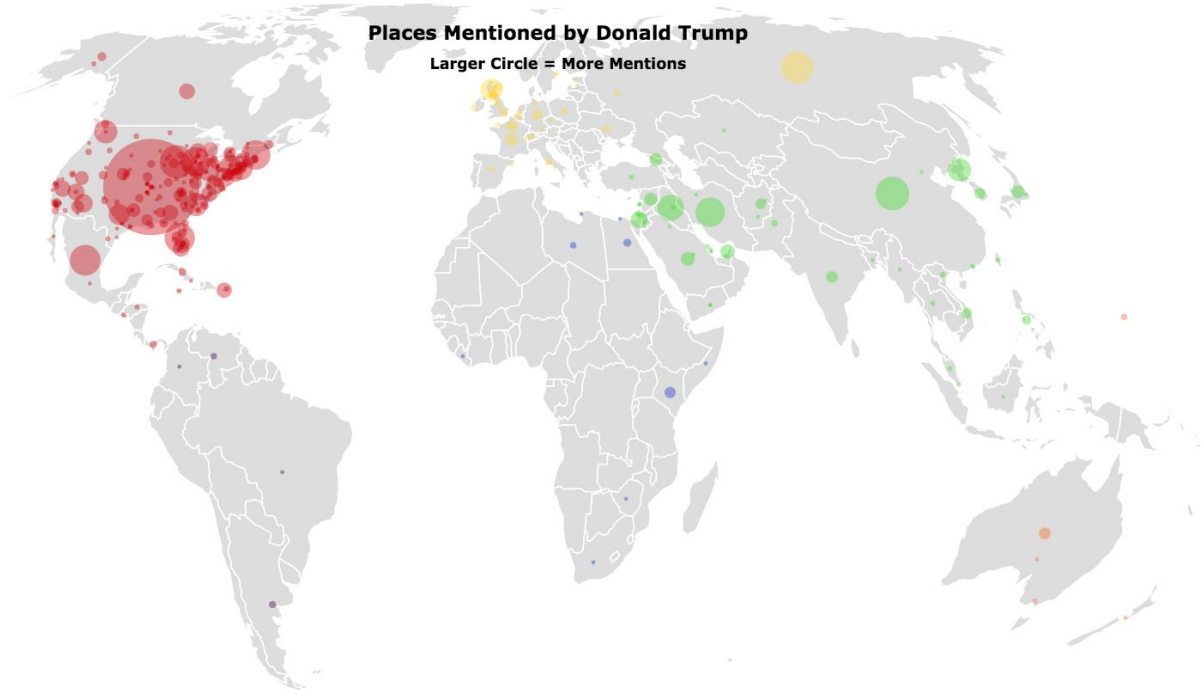

Figure 4 - World Places Donald Trump Has Mentioned

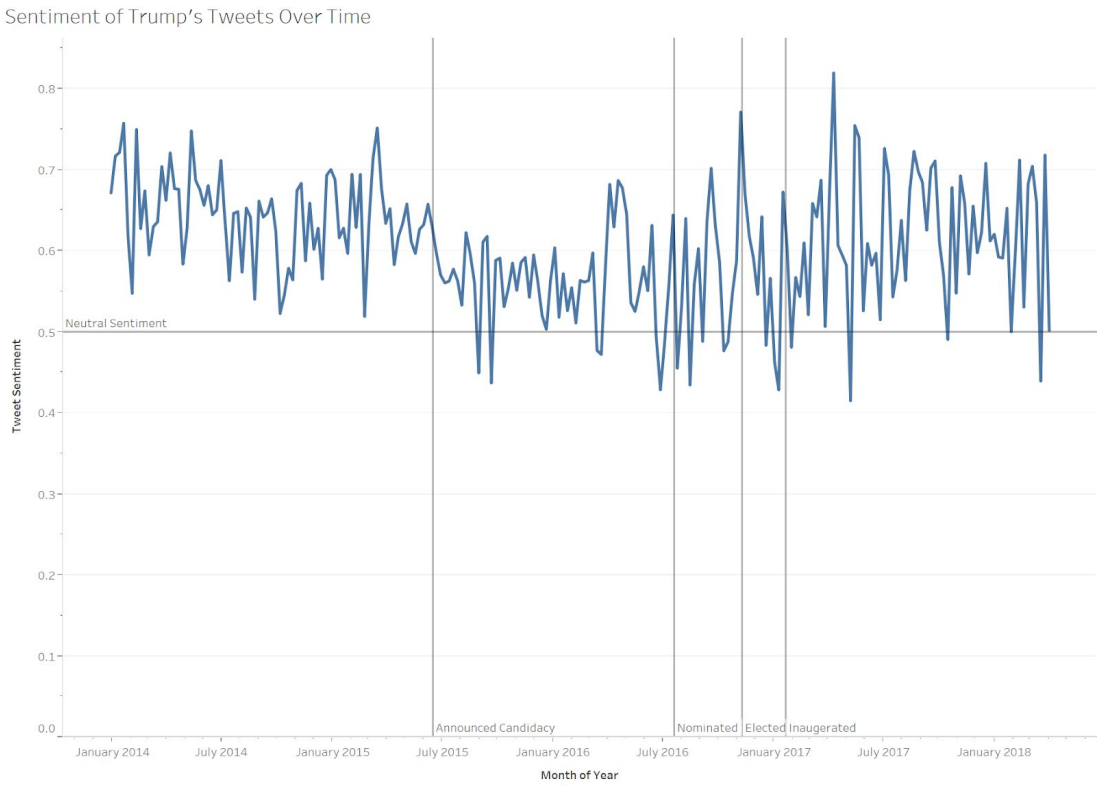

Figure 5 - Sentiment of Trump's Tweets Over Time

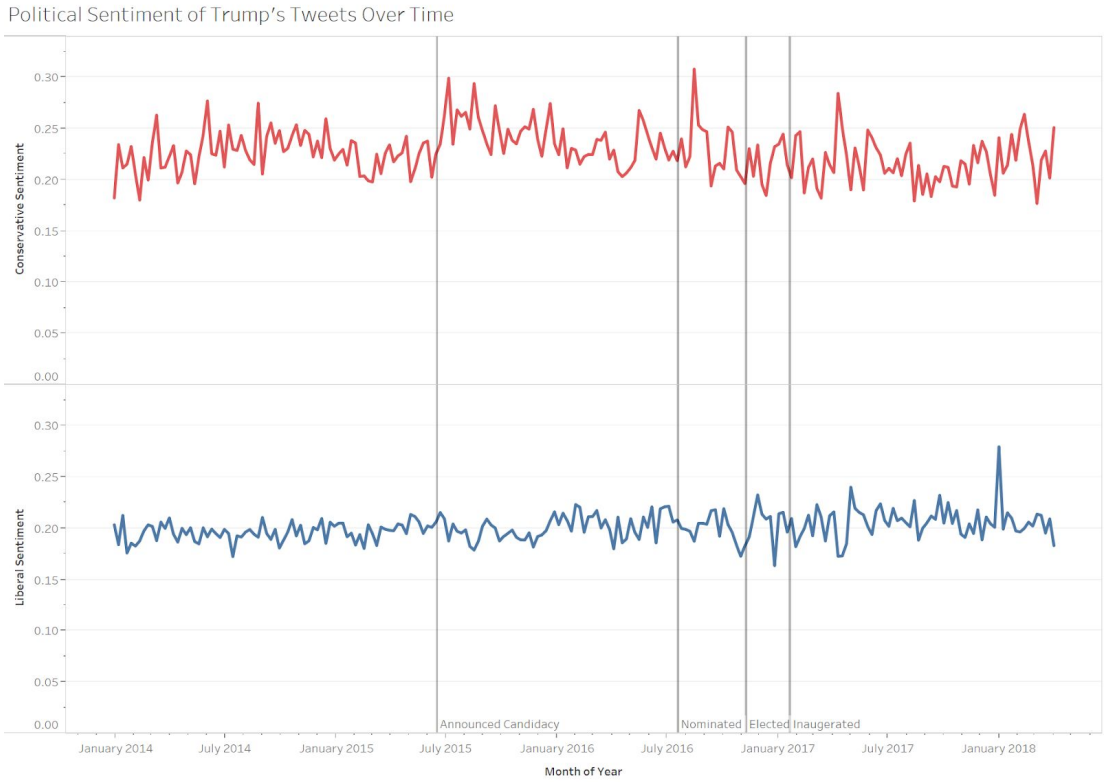

Figure 6 - Political Sentiment of Trump's Tweets Over Time

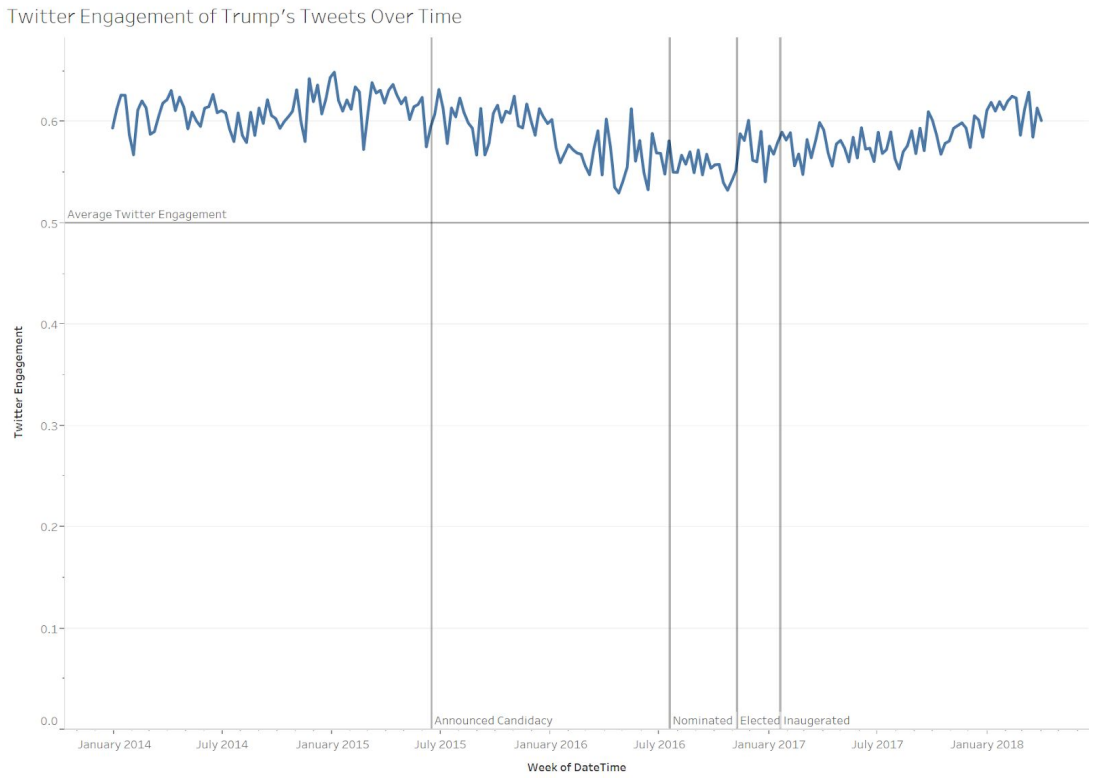

Figure 7 - Twitter Engagement of Trump's Tweets Over Time

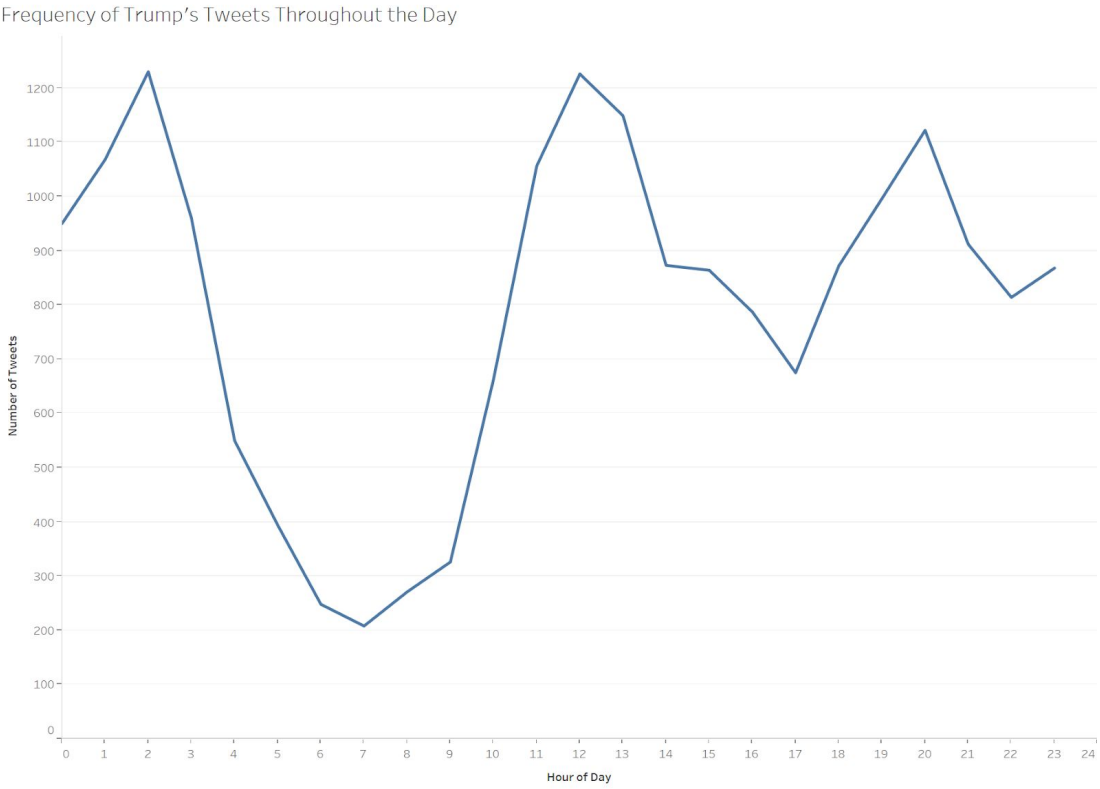

Figure 8 - Frequency of Trump's Tweets Throughout the Day

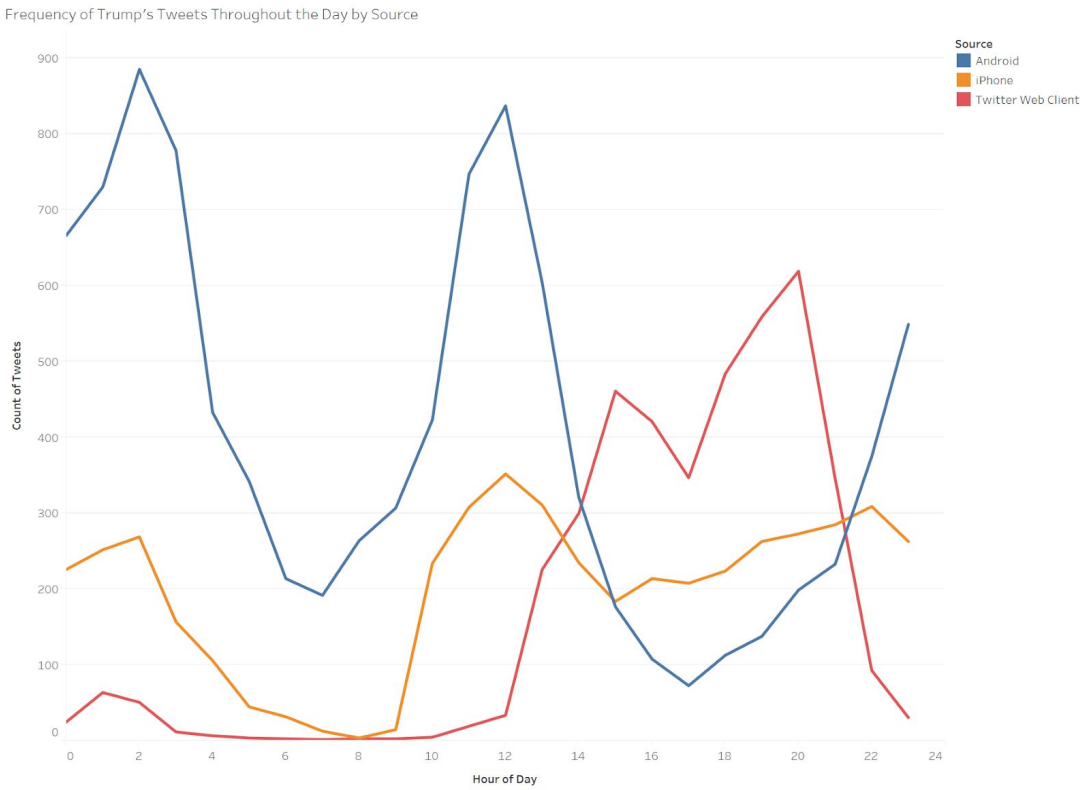

Figure 9 - Frequency of Trump's Tweets Throughout the Day By Source