Narrative

Elisa Bayoumi

12/10/2020

Brief substantive background / goal

On May 26th, 2020, Twitter flagged two tweets from @RealDonaldTrump as false or misleading information. On November 3rd, 2020, just days after the presidential election, Twitter flagged more tweets from the President where he claimed that he had won the election though many ballots were still being counted in important states. This flagging process has increased in recent months, but little research academic statistical research been done on the nature of flagged vs. unflagged tweets as there is no way to account for a flag on a tweet using Twitter's current API. I therefore set out to create a dataset that included noting whether tweets are flagged or not on Trump's main Twitter account, and preforming preliminary analysis in the form of time series data, popularity of flagged tweets compared to unflagged tweets, and word association of flagged tweets compared to unflagged tweets.

Collecting data

Collecting the data happened in two steps, which I then combined in cleaning/preprocessing.

First, I used RTweet to scrape the first 1000 tweets from @RealDonaldTrump's timeline, including retweets. 1000 was an arbitrary number from another study scraping Trump's timeline. A bug in RTweet returns a random number of tweets at a time for Trump's twitter. I therefore had to rerun the code several times to get a maximum number of tweets - 997 - which I then saved to another variable in case I reran the RTweet code and ended up without my original complete dataset.

Second, I went through @RealDonaldTrump and manually copied and pasted the URLs of each of his tweets that had been flagged by Twitter. I did not discriminate between claims from Donald Trump that had been flagged and retweeted content of another user that had been flagged. I then wrote a function to save only the status IDs of the flagged tweets from the end of the URLs, cutting the beginning of the hyperlink.

Cleaning / pre-processing data

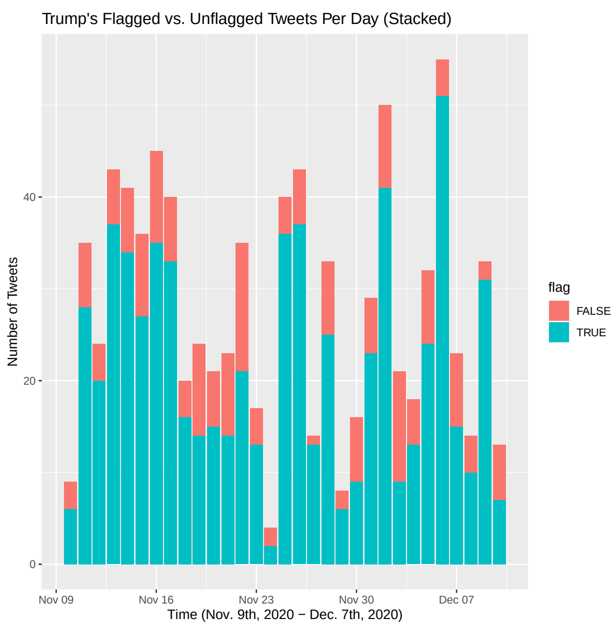
First, I mutated the dataset of Trump's tweets by adding a column of whether the status id of each tweet existed in the list of status ids of flagged tweets. I then reversed this boolean to make it clearer for the reader or user that the tweet was flagged AS false or AS true. In this process, I also selected the columns necessary for analysis, though these columns can be adapted for future researchers if they want to research different areas.

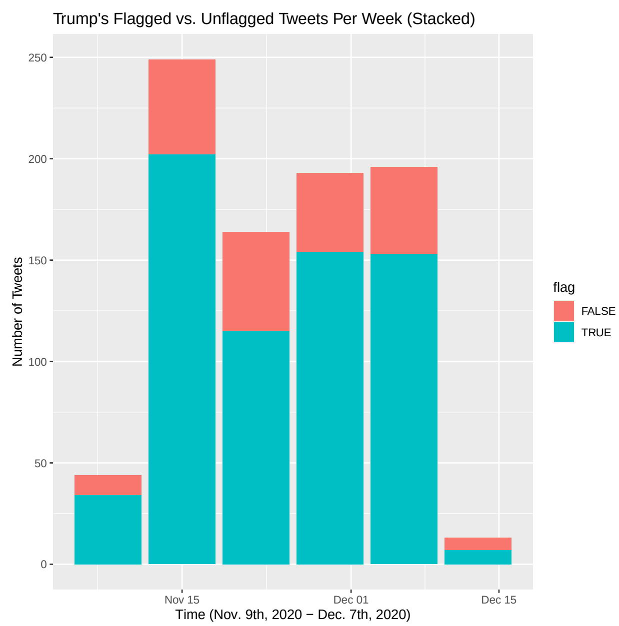
After getting the flag column, I had to account for only being able to evaluate 28 days worth of Trump's tweets. I got the date that the last flagged tweet was created at and removed all tweets that were posted before that time as I could not check whether Twitter flagged them or not. This worked as the last tweet I saw was flagged, but the status ID of the last non-flagged tweets can also be copy-and-pasted into the code.

Analysis and visualization

There were three elements to this analysis.

First, I did a date-time analysis. I wanted to develop this skill with Twitter data for future research purposes. I simplified the date with round_date() for a daily analysis and weekly analysis which I then plotted in ggplot.





Next, I did an analysis of the popularity of flagged tweets vs. unflagged tweets using favorites and retweets as a proxy for popularity. I mutated the existing retweet and favorite counts per tweets to get the average number of retweets and favorites for flagged and unflagged tweets over one day. I then manipulated the dataframe to make it easier to plot in ggplot. I additionally found a way to plot all 4 lines for favorites and retweets of flagged tweets as well as favorites and retweets of unflagged tweets while using color as a proxy for flagged vs. unflagged visualization.

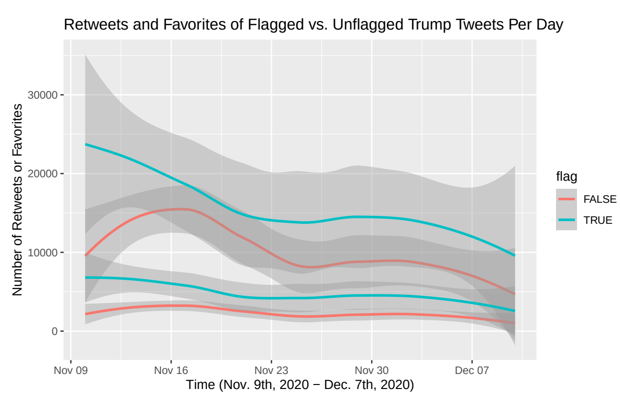


Figure 1: Popularity of Flagged Tweets

Finally, I conducted a frequency word analysis on Trump's tweets. I chose to analyze the text of Trump's tweets for frequency of words, rather than unique words, as there is only one speaker and Trump tends to repeat words in his tweets. Overall, I revealed the common words in flagged and unflagged tweets as well as commonly tagged users in flagged and unflagged tweets.

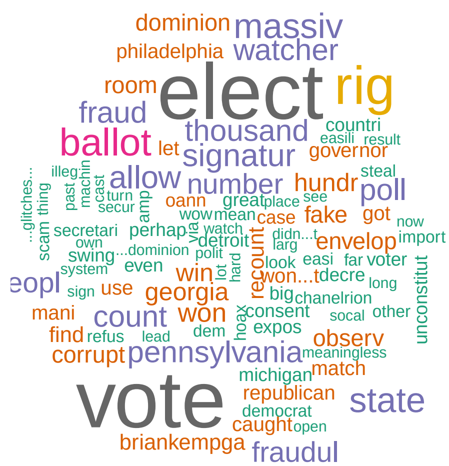


Figure 2: Word cloud for frequent words in flagged tweets

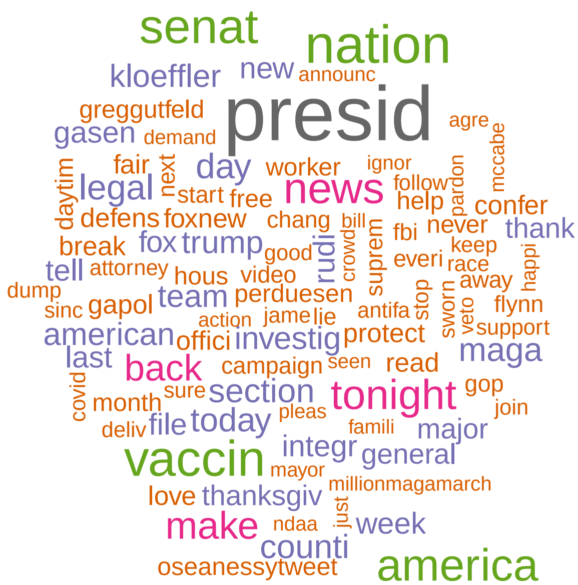


Figure 3: Word cloud for frequent words in unflagged tweets

Future work

Going forward, I would like to continue this project in three ways.

First, I would like to further analyze the dip in favorites of flagged tweets, while retweets for flagged tweets stayed as popular as ever. Preliminary analysis of the data suggested few outliers; the dip in popularity therefore could warrent more analysis to see if it is statistically significant, and if so, what is driving the dip in favoriting tweets.

Second, I would like to distinguish between @RealDonaldTrump's retweets of other users' flagged tweets and his original flagged content. This analysis could have serious implications for disinformation analysis and understanding how disinformation propogates, especially if it is extended to account for whether the original tweets are from verified or unverified accounts.

Finally, there is an obvious difference in tone and sentiment across flagged and unflagged tweets. With this surface validity, this project could be used to produce a sentiment dictionary for @realdonaldtrump based on his most frequently used words. Other mechanics like punctuation and capitalization would also be important for evaluating sentiment.