The Sentiment of U.S. Presidential Elections on Twitter

Zahra Masood, Hannah Peterson, Sravya Voleti

Abstract:

Political tensions in the United States came to a head in 2020 as the public responded to various major events such as the onset of the COVID-19 pandemic and the murder of George Floyd, as well as the 2020 presidential election. Here we investigate if there is evidence of increasing polarization and negativity in regards to politics among the American public on social media by analyzing Twitter data related to the 2016 and 2020 presidential elections. Using publicly available datasets of tweets for each election, we perform sentiment analysis on the text of tweets to quantify their degrees of negativity and subjectivity. We also identify political leanings of tweets by analyzing their hashtag usage and identify "dialogue" occurring between and amongst left- and right-leaning users by analyzing the tweets' user mentions. We then conduct permutation testing on these various groupings of tweets between the two years to determine if there is statistical evidence of increased polarization and negativity on social media surrounding the U.S. presidential election from 2016 to 2020, both generally and between and within political parties. We find that election-related tweets in 2020 generally used less neutral language than in 2016 but were not conclusively more positive or negative in sentiment.

Introduction:

The 2020 presidential election in the United States was polarizing for the American people, perhaps even historically so, as Americans grappled with casting their votes for Joe Biden or President Donald Trump in light of major events that took place in 2020, including the onset of the COVID-19 pandemic at the beginning of the year and the racial protests and riots sparked by the murder of George Floyd in May. Happenings such as these have led the American people to regard the year 2020 as generally negative, public discourse being rife with statements such as "I just want 2020 to be over." There is also talk amongst the media and general public that, politically, Americans are "more divided than ever," largely in reference to conflicting opinions about Donald Trump's controversial presidency. We are curious if there is empirical evidence of Americans' attitudes towards politics becoming increasingly negative and polarized as suggested by these common sentiments.

The aim of this investigation is to determine if there is evidence of increasing polarization and negativity in regard to politics in recent years among the American public on social media, which we do by performing sentiment analysis and permutation testing on tweets related to the 2016 and 2020 presidential elections. We choose to analyze Twitter data (tweets) because Twitter has become a popular platform for voters and candidates to express their beliefs and sentiments. Social media has been the focus of many studies on sentiment analysis in recent years given its rise popularity as a medium for public discourse and the unique nature of text on such platforms, with its heavy use of slang, acronyms and abbreviations, emoticons, etc. requiring a different analytical approach from traditional text. Notably, in 2015 researchers at Georgia Tech developed a lexicon and rule-based sentiment analysis tool named VADER (for Valence Aware Dictionary for sEntiment Reasoning) that is specifically attuned to analyze social media text with high accuracy, outperforming individual human raters at classifying the sentiments of tweets, for example [1]. VADER is backed by a lexicon of 7,500 words, acronyms and emoticons commonly used in microblogs that are scored on a scale from "Extremely Positive" to "Extremely Negative" assessed via a wisdom-of-the-crowd approach. It also factors in textural features such as exclamation points and capitalization to determine the intensity of text snippets. We use VADER to determine the sentiments of tweets in this investigation.

In terms of investigating politics on Twitter, in December 2019, Knight Foundation released a study analyzing the political dynamics of 86 million tweets from 2017 [2]. In this study, they binned Twitter users into four political segments—extreme left, center left, center right, and extreme right—and found that the political spectrum on Twitter skews heavily center left, with 10% of users falling in extreme left segment, 57% in the center left, 8% in the center right, and 25% in the extreme right, also indicating that conservative representation in Twitter is dominated by the far right. These are important considerations to factor into interpreting the results of our investigation. Our work bears some similarity to Knight Foundation's in that it is also based on assigning political leanings to users, however, our approach refers to tweets' hashtag usage to do so, whereas Knight Foundation's assigns leanings to users based on the type of well-known, politically involved users they follow. Additionally, our investigation is focused on how polarization and negativity on Twitter may have changed over time, which is not a factor of the Knight Foundation study.

We use two different datasets of tweets in this investigation, one for each election. The dataset of tweets related to the 2016 election was compiled by researchers at Harvard [3]. It was collected through data-driven keyword search using Twitter's API and contains approximately 280 million ids of tweets related to the 2016 presidential election from between July 13, 2016, and November 10, 2016. These tweet ids are further broken down into subcategories based on different events: the election day, debates, and conventions. Each tweet id maps to a unique tweet, the full content of which we fetch (in a process called "hydration") using Twarc, a Python package that is linked to Twitter's API. Each hydrated tweet contains data fields such as the full text of the tweet, the hashtags it uses, the users it mentions, if it was an original tweet or retweet, etc., as well as information about the Twitter user who authored the tweet. The dataset for the 2020 election we use was compiled by researchers at the University of Southern California and contains over 800,000,000 tweet ids from May 12, 2020, through December 2020 [4]. The ids were collected by tracking various election-related keywords and tweets from politically-affiliated accounts using Twitter's API. We randomly sample from these datasets and perform some additional filtering to compile our own datasets of 414,713 tweets for 2016 and 500,000 tweets for 2020. Though the two datasets we use were collected by different groups, they both were collected through comprehensive, unbiased keyword searches that would have captured the majority of political discussion on Twitter surrounding the two elections, making them appropriate for our analyses here. We hypothesize that within this data there will be a shift towards more negative sentiment and less neutrality from 2016 to 2020.

Exploratory Data Analysis:

For our exploratory data analysis, we first found the 50 most common hashtags for 2016 and for 2020. For 2016 the most common hashtag used was trump, with almost 8,000 occurrences (Figure 1). For 2020 the most common hashtag was again trump but with only over 2,000 occurrences (Figure 2). We then created a new column called baseline which contained the baseline occurrence for each tweet in the dataset. We found it by dividing the number of occurrences of the hashtag by the total number of tweets in our dataset. For the most common hashtag in 2016, ‘trump’, the baseline rate was 0.0189 (Figure 3). In 2020, the most common hashtag, ‘trump’, has a baseline occurrence rate of 0.002 (Figure 4). We then plotted a histogram that showed the distribution of the number of posts per user. For 2016, this distribution was right-skewed with its largest spike around 50 (Figure 5). For 2020, the distribution was also right skewed but with a spike around 40 (Figure 6). We then plotted two more histograms which showed the distribution of the number of retweets per user. For 2016 and 2020, the distributions are right-skewed with spikes around 0.0 (Figures 7 & 8).

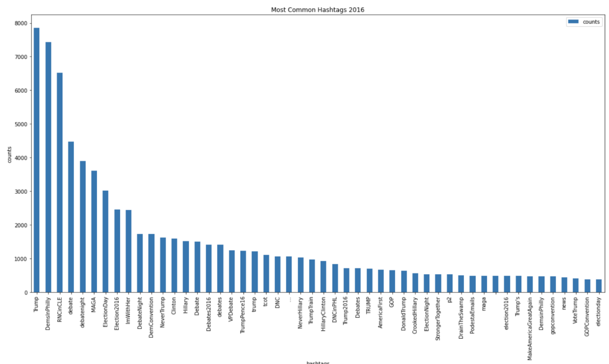

Figure 1

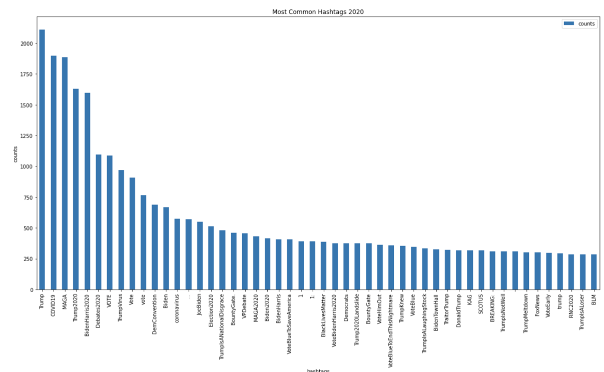

Figure 2

	hashtags	counts	baseline
0	Trump	7853	0.018959
1	DemsInPhilly	7425	0.017926
2	RNCinCLE	6528	0.015760
3	debate	4471	0.010794
4	debatenight	3903	0.009423

Figure 3

	hashtags	counts	baseline
0	Trump	2109	0.002425
1	COVID19	1900	0.002185
2	MAGA	1885	0.002168
3	Trump2020	1630	0.001874
4	BidenHarris2020	1596	0.001835

Figure 4

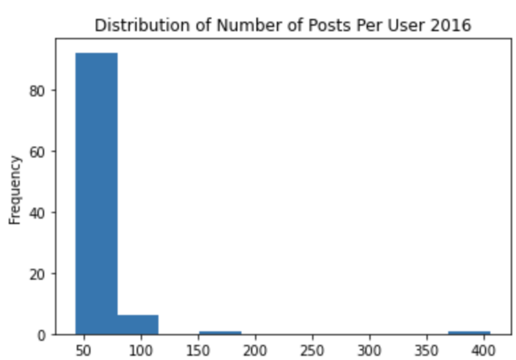

Figure 5

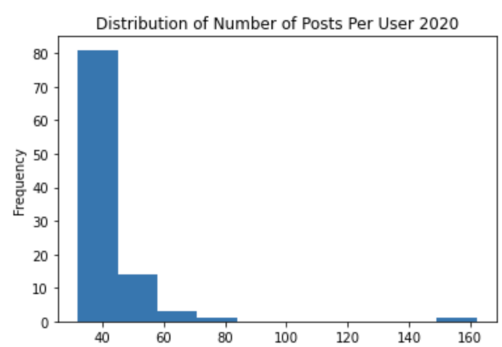

Figure 6

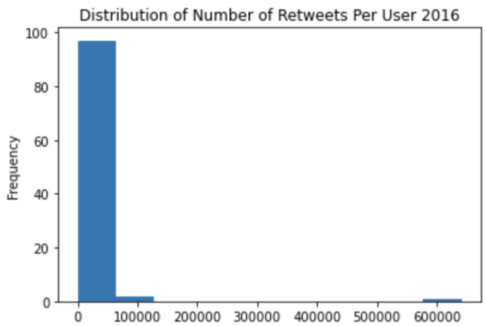

Figure 7

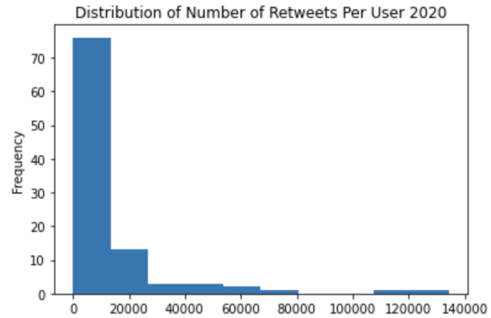

Figure 8

Methods:

Data downloading and cleaning process:

The 2016 and 2020 raw datasets consist of text files of tweet ids, which we randomly sample from to get datasets of workable sizes. For the 2016 dataset, we sample 1 out of 300 tweet ids, and for the 2020 dataset, we sample 1 out of 30. We then use twarc to "hydrate" the ids, fetching the content of their corresponding tweets. To maintain uniformity for sentiment analysis, we filter out tweets in languages other than English. Additionally, the two datasets were collected by different groups that used different keyword searches to gather election-related tweets, so for our analysis to not be affected by this, we have to unify them under a universal keyword search. The keyword search used for the 2020 dataset was more exhaustive and broad than that used for the 2016 dataset, so we filter the 2020 dataset using the 2016 keywords, changing the names of the candidates to align with the 2020 election. Our final datasets consist of 414,713 tweets from 2016 and 500,000 tweets from 2020.

Left and right separation:

We assign political leanings to a subset of the tweets that are authored by partisan news sources or politicians or that use politically-charged hashtags. We compile lists of right- and left-leaning Twitter accounts of partisan news sources by selecting those determined to be far-left or far-right by statistical analysis of popular news outlets performed by AllSides.com. We add to these lists the accounts of all of the candidates listed on the 2016 and 2020 Democratic and Republican Party presidential primaries' Wikipedia pages. With this, we have (far from exhaustive) lists of relatively influential Twitter users whose political leanings are well-defined.

To otherwise identify left- and right-leaning tweets among tweets, we refer to their hashtag usage. We hand-pick clearly left- and clearly right-leaning hashtags among commonly used hashtags within the sets of tweets for both years. For 2016, for example, left-leaning hashtags include "voteblue" and "donthecon" and right-leaning hashtags include "maga" and "lockherup". We do not include hashtags such as "trump" in our search because they could be used to speak both favorably and unfavorably about their subjects. Then, for a given year, we make left and right subsets of tweets by selecting ones that include at least one hashtag from the respective lists, excluding those that include hashtags from both lists. We then make lists of right- and left-leaning users by collecting the screen-names from each subset. We remove the intersection of users from these lists, interpreting such users' political leanings to be inconclusive considering they authored tweets that included both left- and right-leaning hashtags. With this, we now have a list of screen names for each political leaning, with no overlap between them. We combine these with the respective lists of politicians and partisan news sites accounts and then gather all of the tweets from the dataset authored by these accounts to arrive at our final left- and right-leaning subsets of the following numbers of tweets:

	Left-leaning	Right-leaning
2016	28,282	34,942
2020	17,349	17,132

Table 1: Counts of left- and right-leaning tweets by year

Dialogue separation:

Having assigned political leanings to some of the tweets, we can identify dialogue happening between and within the two political spheres by analyzing their user mentions. For a given year, we gather the left- and right-leaning tweets that mention at least one user and classify the dialogue occurring as either L-L, L-R, R-L, or R-R depending on the leaning of the author of the tweet and that of the user(s) they mention. For example, if a tweet is authored by a left-leaning user and mentions one or more right-leaning users, it is classified as L-R. If the leaning of one or more of the users mentioned is unknown, that is, they are not present in our lists of left- or right-leaning users, the dialogue is not classified. Through this process, we identify the type of dialogue of the following numbers of tweets:

	L-L	L-R	R-L	R-R
2016	4,317	854	1,073	5,628
2020	1,992	2,126	951	1,754

Table 2: Counts of tweet dialogue types by year

Sentiment analysis:

In order to conduct our sentiment analysis on the full text of tweets from 2016 and 2020, we used the Vader Python library. Initially, we conducted sentiment analysis using the Python library named Textblob. Textblob references a lexicon library and assigns polarity and subjectivity to all of those specific words and averages them to find the polarity/subjectivity of the entire text. However, as will be discussed further in the discussion, upon further research and analysis of our Textblob results, we decided to move forward with a different library called Vader since it was more appropriate for the specific social media data that we are analyzing. Vader uses a dictionary in order to map lexical features to sentiment scores, which are emotion intensities. By summing up the intensity of each word in a text, Vader can obtain the overall text's sentiment score. Vader is intuitive, in that it understands the implications of capitalization and punctuation. It also will take the usage of negative words such as “not” or “no” into account. Vader has four classes of sentiments that it assigns these scores to. These classes include *positive*, *negative*, *neutral*, and *compound*. The compound class is an aggregated score of the first three classes and it ranges from -1.0 to +1.0. These compound scores can tell us whether or not our text was expressing a positive, negative, or neutral opinion. A text with an overall positive sentiment would have a compound score of greater than 0.05. On the other hand, a text with an overall negative sentiment would have a compound score of less than -0.05. A neutral sentiment text would have a compound score of somewhere in between.

Permutation testing:
We conducted permutation testing on the differences in mean compound score and neutrality between 2020 and 2016 for different groupings of our data: the data overall; the left and right subsets; and the left-left, left-right, right-left and right-right dialogue subsets. We chose to perform permutation tests to be able to assess if the results of our sentiment analysis and determine if the differences in these sentiments between the two election cycles were statistically significant. We stated that the null hypothesis was that there was no change in the compound score or neutrality score between the two years. We chose to use the mean as the test statistic as we concluded that using the mean score would be a meaningful statistic to represent the compound score and neutrality score for each year and an effective way to compare the two distributions. To perform the permutation test, we found the observed difference between the two distributions (the difference in means) and randomly sampled the two distributions without replacement. We then found the p value as the proportion of sampled differences greater than our observed difference and used this p value to accept or reject our null hypothesis.

Results:

Figures 1 and 2 show the results of the Compound scores achieved via sentiment analysis of the full text of tweets of the various subsets of data for 2016 and 2020, respectively. Recall that a Compound score of -1 indicates extreme negative sentiment and +1 indicates extreme positive sentiment., with Vader assigning a default score of 0.0 if it does not recognize enough of the text to assign a polarity. Across all subsets, a Compound score of 0.0 was most common. Visually, there is a high degree of similarity between the distributions for the two years. The most noticeable difference is that L-R dialogue became more left skewed in 2020, which would indicate an increase in negative sentiment within this group.

Figure 1: Distributions of 2016 Compound scores

Figure 2: Distributions of 2020 Compound scores

Similarly, Figures 3 and 4 show the results of the Neutrality scores achieved via sentiment analysis of the full text of tweets of the various subsets of data for 2016 and 2020. Recall that a Neutrality score of 0 indicates that none of the text was neutral and 1.0 indicates that all of the text was neutral, which is the default score Vader outputs if it does not recognize enough of the text to assign a polarity. Across all subsets, a Neutrality score of 1.0 was most common. Visually, there is a high degree of similarity between the distributions for the two years, but there seems to be a general trend of the data becoming slightly more left skewed in 2020, indicating less use of neutral text, a change that is particularly evident in the L-R dialogue subset.

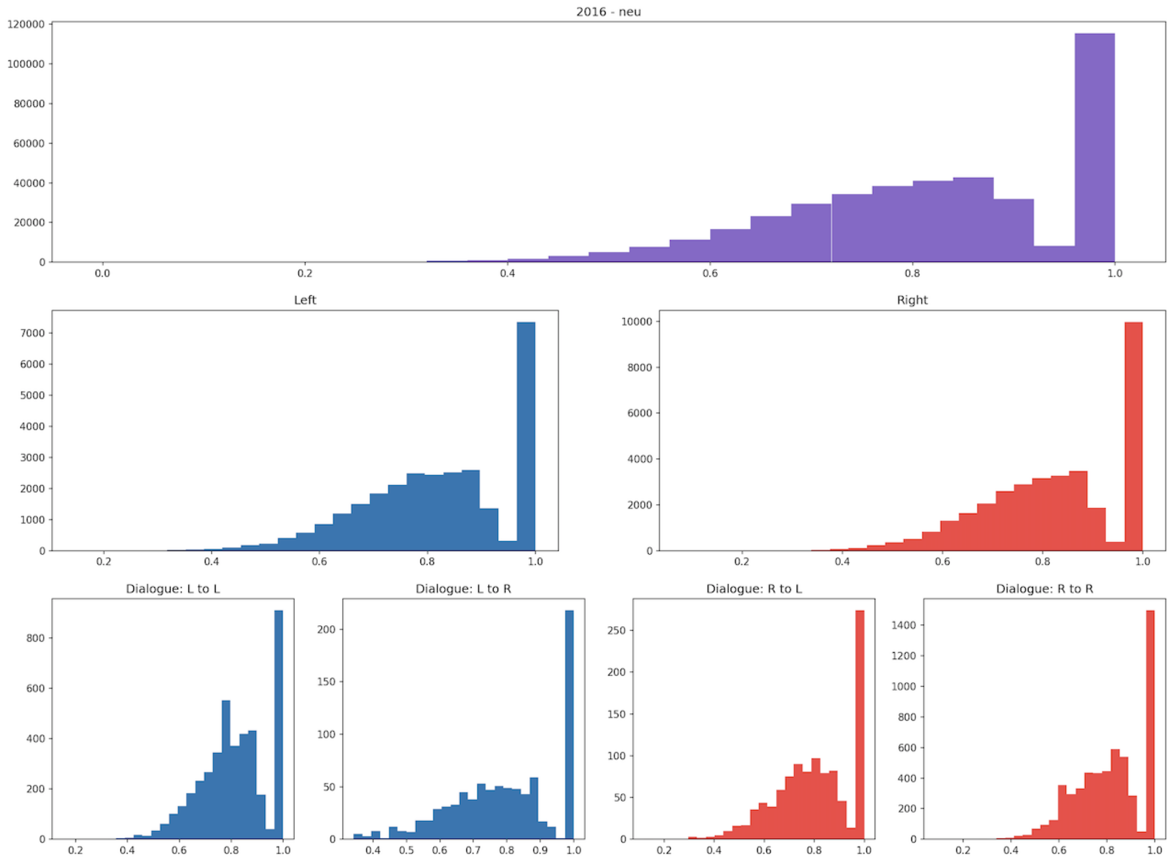

Figure 3: Distributions of 2016 Neutrality scores

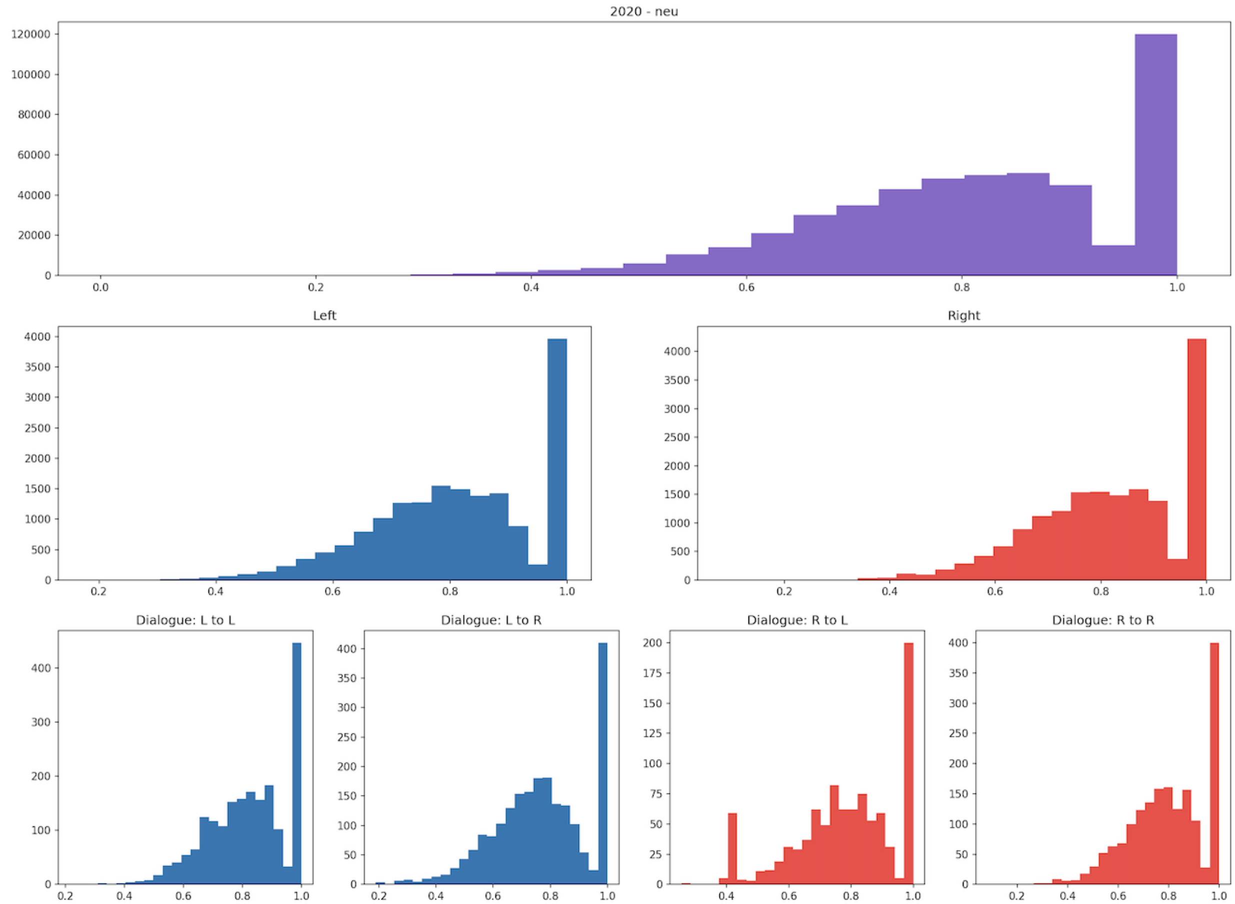

Figure 4: Distributions of 2020 Neutrality scores

For empirical evidence of change in sentiment of tweets between the 2016 and 2020, we refer to the results of permutation tests on the differences in mean scores. Table 3 provides a breakdown of the observed differences in means and the p-values attributed to these differences as a result of the permutation tests, with statistically insignificant results based on a 95% confidence level highlighted in red. Figures 5 and 6 provide graphical representations of these test results for the Compound and Neutrality scores, respectively, with the observed differences between the 2020 and 2016 mean scores represented by dotted red lines. Aligning these results with the language of our hypothesis, a negative Compound difference indicates an increase in negativity and a negative Neutrality difference indicates a decrease in neutrality.

	Compound		Neutrality	
	Observed Diff.	p-val	Observed Diff.	p-val
Overall	0.01733	0.000	-0.00570	0.000
Left	2.03099e-05	0.490	-0.01107	0.000

Right	0.02678	0.000	-0.00634	0.000
Dialogue: L-L	0.07682	0.000	0.00571	0.052
Dialogue: L-R	-0.10595	0.000	-0.02790	0.000
Dialogue: R-L	0.05880	0.002	-0.02032	0.002
Dialogue: R-R	-0.05066	0.000	-0.01868	0.000

Table 3: Observed differences in mean Neutrality and Compound scores (2020 minus 2016)

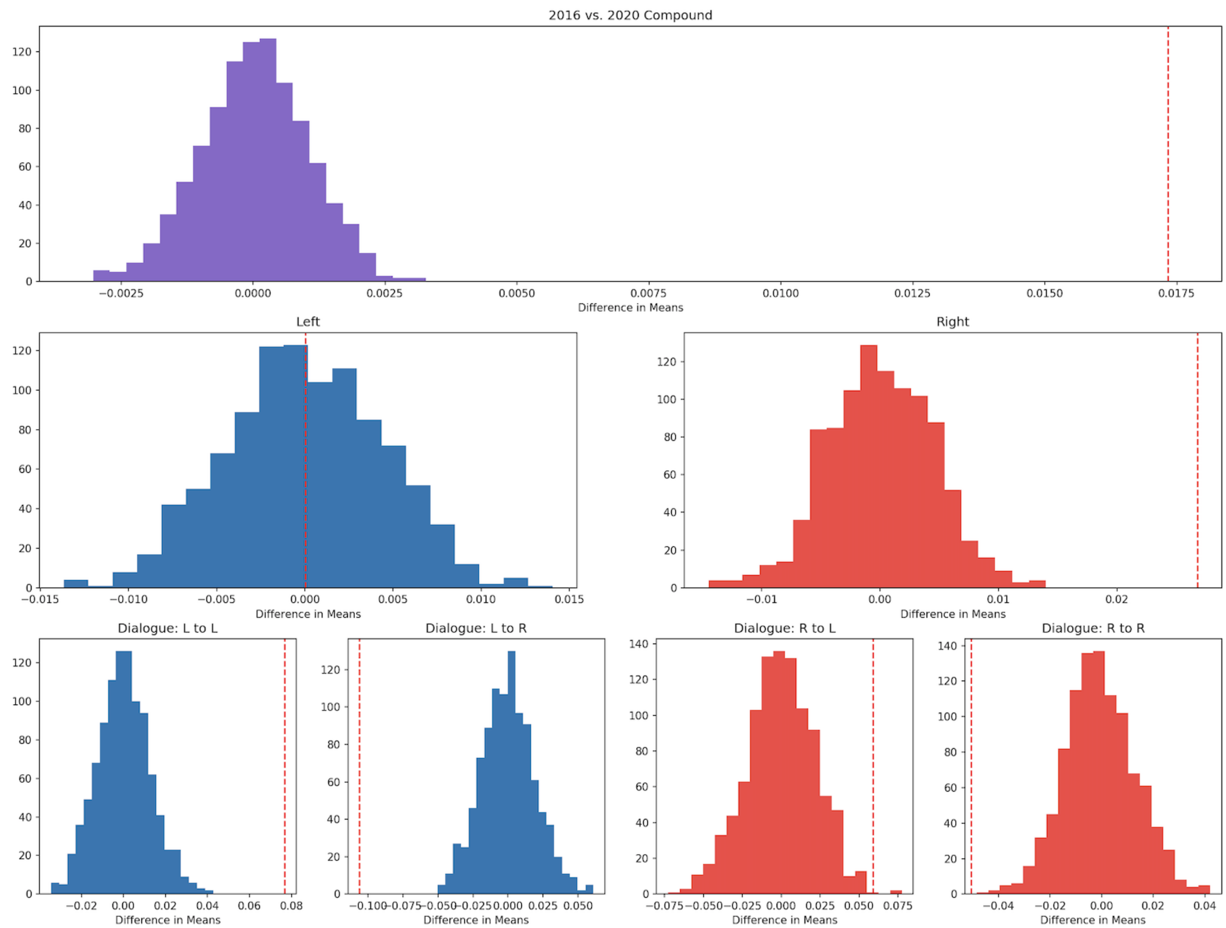

Figure 5: Differences in mean Compound scores (2020 minus 2016) from permutation tests

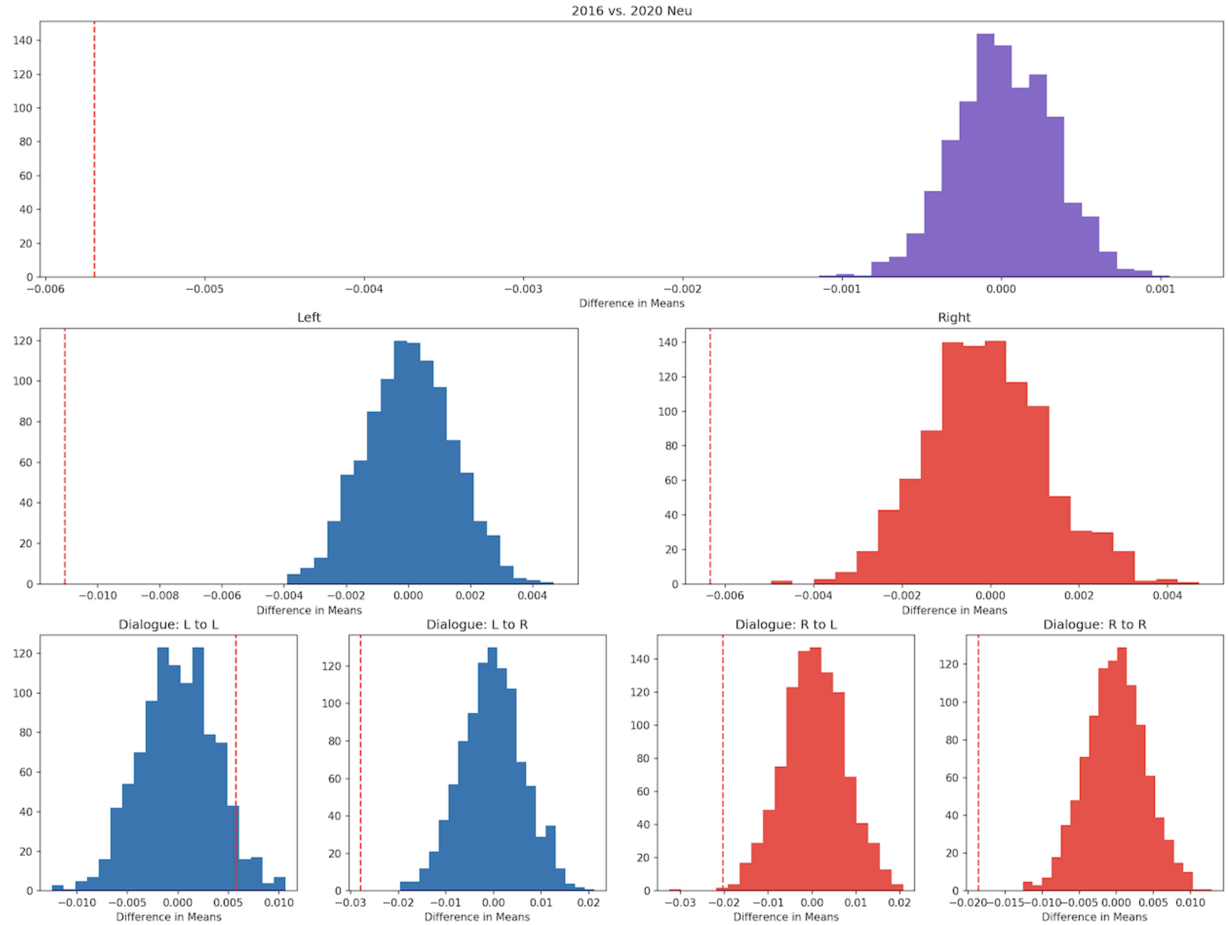

Figure 6: Differences in mean Neutrality scores (2020 minus 2016) from permutation tests

From these results, we can see that, overall, the 2020 tweets were overwhelmingly more positive and less neutral than the 2016 tweets. With the exception of the L-L dialogue subset, for which the results are inconclusive (p=0.052), tweets used less neutral language in 2020 than in 2016 across all subsets.

Results for the Compound scores are less conclusive, as the direction of change differs depending on the subset. There is no detectable difference in Compound score for left-leaning tweets (p=0.49), whereas right-leaning tweets were conclusively more positive, with a mean increase of 0.02678 in Compound scores As for dialogue, L-L and R-L dialogue became more positive from 2016 to 2020, with mean increases in Compound scores of 0.07682 and 0.05880, respectively, and L-R and R-R dialogue became more negative, with mean decreases in Compound scores of -0.10595 and -0.05066.

Discussion:

As determined by our results, we observed that in terms of neutrality the 2020 tweets were actually overwhelmingly more positive and less neutral when compared to the 2016 tweets. In terms of the compound score, right leaning tweets were more positive while left-leaning tweets had no detectable

differences. This disproves our hypothesis which stated that between 2016 and 2020, there will be an overall shift towards a more negative sentiment and less neutrality. We observed that, although there was an overall decrease in neutrality, there was actually a shift towards more positive sentiment on Twitter between the two election cycles. This decrease in neutrality was particularly evident when we analyzed the dialogue between the left-leaning and right-leaning users, indicating that users who are left-leaning but mention right-leaning users used noticeably less "neutral" language and became more polarized. Although there was a general positive shift between the cycles, we did notice that between these same subsets of users there was actually an increase in negative sentiment. We can conclude that users who identified as left-leaning, but mentioned right-leaning groups became more negative in their dialogue between 2016 and 2020. Thus, although we did notice an overall positive shift in sentiment, this unique subset provides us with insight into how the interactions between groups of the two parties changed between the two cycles; there seems to be an increase in political polarization and negative feelings from those who are left-identified to those who are right identified and this is reflected in stronger and more politically-charged tweets. This could be a result of the frustrations harbored during the 2016-2020 presidential cycle as well as the use of Twitter as a platform to tweet opinions more openly and aggressively towards users of opposing parties. Interestingly, the dialogue between right-leaning users who mention left-leaning users became more positive between the two cycles, indicating that although the left-leaning users might have interacted with the right more aggressively it is not the same vice versa. In addition, the dialogue within left users demonstrated a more positive shift, but that within right users became more negative, indicating that we can't make conclusive statements in terms of the sentiment shift within a certain political party, but perhaps there was more optimism within the left-leaning users with the new hope that came with the new election cycle, and more worry / pessimism within the right. Thus, although there was an overall positive shift of tweets regarding the election between the two cycles, we were able to gain interesting insights by examining the dialogue between users and the two groups.
Our work was similar, but different to the prior work completed about this topic. We cited an article where researchers classified Twitter users into four segments by assigning leanings to users based on the type of well-known, politically involved users they follow. These researchers found that the overall conservative representation on Twitter was dominated by the far right. Our results were a little different as the purpose of our project was to use sentiment analysis to come to conclusions about the change in sentiment between different users in the two election cycles. However, we used a similar approach to the researchers of classifying Twitter users into left-leaning or right-leaning, but we did so by identifying left-and right-leaning hashtags.

Textblob

As mentioned in the methods portion, we initially conducted our sentiment analysis using the Python library Textblob. To do this, we created two functions: sentiment_polarity and sentiment_subjectivity. The first function finds the polarity of the text. This polarity score is a float within the range [-1.0, 1.0]. A negative score would mean that the text has mainly negative connotations while a positive score means the opposite. The second function finds the subjectivity of the text and it returns a float within the range [0.0, 1.0] with 0.0 is very objective and 1.0 is very subjective. We applied these two functions to the full_text column in the 2016 and 2020 dataset. For the 2016 dataset, the mean polarity was 0.054 while the mean subjectivity is 0.333. This means that the tweets from 2016 have an overall positive polarity while being more on the objective side. For 2020, the mean polarity score is 0.056 while the subjectivity is 0.340. This means that the tweets from 2020 also are more objective with an overall positive polarity.

Upon further analysis of these results and research into various methods of sentiment analysis for Twitter data, we decided to use another library, Vader, so we could compare our scores and results. In the end we decided to use Vader as the tool for conducting sentiment analysis because it is geared towards social media text specifically and it can understand slang and emojis while Textblob was more commonly used for general text.

One large limitation of our approach is the size of the data that we used. After splitting our overall dataset into groups based on L-R, R-L, etc., some of these smaller groups had only a fairly small number of tweets left for us to analyze. For example, our 2016 L-R dialogue group only had 854 tweets which might not be enough to draw a substantial conclusion. Also, it is important to take the general demographics of Twitter into account when looking at our results. While we assume that Twitter has an equal number of people of all political leanings, this is not the case. As of 2020, the largest age demographic that is active on Twitter is from 25-34 years old[1], and it is also important to note that more than 50% of people in this age category are generally left-leaning[2]. This would mean that a majority of the tweets we find in our dataset would be more left-leaning than right-leaning but this is not necessarily representative of the entire United States population. Another limitation in our approach might come from the fact that our 2020 and 2016 datasets were considerably different, which required us to filter it. The way in which we conducted our filtering of the 2020 dataset in order for it to be aligned with the 2016 dataset might have been ineffective. Moreover, when we were filtering our tweets into right and left leaning, those users whose leaning was unknown were also not present in our subgroups for sentiment analysis. These filtering processes might have had us remove valuable tweets that would have been helpful for us to draw conclusions on.

There is a lot of room for expansion for our project. In the future, it would be interesting to expand our dataset to include the political tweets from a longer time period, such as over the entire four years between 2016 to 2020, so we get a better idea of how the public sentiment on Twitter changed over a continuous time period. It might be interesting to look at tweets over this entire time period due to the fact that it was very tumultuous in American history because we had a controversial president. We could use our methods of dividing the data into subsets depending on political dialogue and then perform sentiment analysis and permutation testing on these groups to see if they uphold the same conclusions that we had. It might also be interesting to expand our analysis to include classification. We could use logistic regression in order to predict what group (L-L, L-R, R-R, R-L) a user would fall in depending on the text of their tweet and compound scores.

References:

[1] VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text (by C.J. Hutto and Eric Gilbert) Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[2] Freelon, Deen. *Tweeting Left, Right & Center: How Users and Attention Are Distributed across Twitter*. Knight Foundation, 17 Dec. 2019,

[1] https://blog.hootsuite.com/twitter-statistics/

[2] https://www.pewresearch.org/politics/2018/03/20/1-trends-in-party-affiliation-among-demographic-groups/

knightfoundation.org/wp-content/uploads/2019/12/KF-Twitter-Report-Part1-v6.pdf.
[3] Littman, Justin; Wrubel, Laura; Kerchner, Daniel, 2016, "2016 United States Presidential Election Tweet Ids", https://doi.org/10.7910/DVN/PDI7IN, Harvard Dataverse, V3.
[4] Emily Chen, Ashok Deb, Emilio Ferrara. #Election2020: The First Public Twitter Dataset on the 2020 US Presidential Election. Arxiv (2020)