Jacob Watkins

Knowledge Mining

Dr. Karl Ho

12/5/2019

How effective is Trump on Twitter?: An Analysis of Donald Trump’s Tweets About the Economy

Introduction

Donald Trump’s use of Twitter as a platform to influence public opinion is unprecedented for a U.S. president. How does Trump use his Twitter account to gain popular support and move public opinion in his favor, and how effective is it? It is possible to examine these questions through aggregation and analysis of Twitter data. Public opinion metrics on any given subject can be gauged using Twitter data through text and sentiment analysis over time. We can then analyze how that sentiment metric changes after Trump tweets about a given subject. An analysis such as this gives us insight to the true effectiveness of Trump on Twitter in swaying opinion on various subjects, one of which will be further examined in the following sections.

While many have written about and performed sentiment analysis of Trump on Twitter, I have not seen another attempt to perform an analysis of this nature, which seeks to quantify an actual change in public sentiment on Twitter after Donald Trump makes a post. There are a multitude of possible topics to explore, considering the volume and breadth of Trump tweets since inauguration (he has tweeted over 11,000 times now since entering office) (Harris et al. 2019). Some of these topics include immigration, the Russian investigation, gun control, the trade war with China, healthcare reform, and most recently impeachment.

For the purpose of simplicity, I have limited my study to exploring sentiment change after Trump tweets about the economy. Using this topic allows a relatively large sample of Trump tweets to analyze, as he perhaps tweets about the economy in some form or fashion more than any other topic. Speculatively, it seems that Trump’s tweets about the economy are very positive in sentiment. If this is indeed the case, I expect to see a positive reaction from general Twitter sentiment after a Trump economy post. Lastly, using the economy as the topic for this analysis should incubate my correlations more than other topics. For instance, when considering a controversial subject such as gun control, it is often the case that Donald Trump might make a post about gun control shortly after a mass shooting, and it would be fairly impossible to prove that Trump’s post influence any change in public Twitter sentiment, rather than the event itself contributing to the change. It is less likely that we have polarizing and controversial events about the economy that could both spark a Trump tweet and substantially change opinion, which allows my findings from this study to better hold up to scrutiny.

Research Question and Hypotheses

The upcoming analysis tests various hypotheses through several means. These hypotheses each contribute to help answer one overall research question: What are the nature of Donald Trump’s tweets about the economy, and how do they affect the overall sentiment of Twitter users? The first part of this question is explored through preliminary research, using methods such as word clouds, sentiment analysis over time, and most common words contributing to sentiment. The latter portion required much deeper analysis, which includes comparing averages of sentiment scores of general Twitter about the economy before and after Trump tweets at various hours. My hypotheses are listed below.

H1: Trump tweets more positively about the economy than his overall twitter sentiment mean.

H1 null: Trump tweets are equal in sentiment or more negatively than

his overall twitter sentiment mean.

H2: Trump tweets more positively about the economy than the general sentiment

 on Twitter does.

H2 null: Trump tweets equally in sentiment or more negatively about the

 economy than the general sentiment on Twitter does.

H3: General twitter sentiment changes after a Trump tweet.

H3 null: General twitter sentiment remains equal after a Trump tweet.

If H3 null is rejected, I also present the following hypotheses:

H4: The direction of change in general sentiment is positive after Trump posts a positive tweet about the economy.

H4 null: The direction of change in general sentiment is negative or equal after

Trump posts a positive tweet about the economy.

H5: The direction of change in general sentiment is negative after Trump posts a

negative tweet about the economy.

H5 null: The direction of change in general sentiment is positive or equal after

Trump posts a negative tweet about the economy.

Data Collection Methods

The analysis involves the collection of three datasets. Each of these datasets were set within the date bounds of January 21, 2017 and September 20, 2019. This time period aligns with the beginning of Trump’s presidency and the day the analysis began. Each of these datasets were collected through the same time period in order for them to be easily comparable over time. One dataset includes 513 Trump tweets about the economy within the period (aka “Trump dataset”), the second contains 433,757 tweets about the economy in general (aka “general dataset”), and the last aggregates 8050 Trump tweets about any topic posted during the time period (aka “Trump overall dataset”).

Research Design

As previously touched on, the research is broken down into two sections. The first of those is a preliminary analysis to explore the data (especially the Trump economy dataset). This analysis seeks to test the first two hypotheses. In order to understand how Trump’s tweets affect the general sentiment on Twitter, it is important to understand the sentiment of those tweets themselves and how they relate to the general sentiment in the first place. The second section implements a for loop in R to identify mean sentiment scores before and after the Trump economy tweets, in order to calculate average change in sentiment after those tweets over time.

However, before we can delve further into the design of those two analyses, it was first necessary to create the sentiment scores for each tweet (in all 3 datasets), because those sentiment scores serve as the primary metric of the analysis for most of the study. Using the tidytext R package and text mining strategies laid out in *Text Mining with R,* by Julia Silge and David Robinson (2019), each of the two datasets went under a similar transformation. Each tweet was assigned a number beginning at one and ascending by date (1-513 for Trump dataset, 1-433,757 for general dataset, etc.). The variable “text” from each dataset was then transformed from type factor into type character, and likewise the “date created” variable from each dataset was transformed from type character into type date-time. These preliminary changes helped make the datasets compatible with the functions required to implement sentiment analysis.

The “text” variable for each dataset was then unnested, forming a new variable “word”. This change broke down the “text” variable (which was the whole text content of the tweet) into single word observations. Because we previously labeled each tweet by number, each word still had a corresponding number assigned to it, which represents which tweet it originally came from. After broken down into words, all instances of the word “Trump” were removed from the analysis. This is because the word “trump” rates positively with the lexicon used to assign sentiment values to the words, and since Trump is used as a name in almost all of its appearances in the tweets, it skews the sentiment positively when it should not.

The next step in creation of tweet sentiment scores was to perform an inner join with each of the datasets and the “Bing” lexicon, which contains 6786 words along with a corresponding “positive” or “negative” value for each. Once this was done, each of my datasets contained numerous observations with 4 variables: date-time created, word, sentiment (positive/negative), and tweet # (1-513 or 1-433757). The sentiment variable for each was then recoded into a variable called “value”, with all positive observations being relabeled a numeric type “1” and all negative observations labeled as numeric type “-1”.  This gave each word in each dataset a sentiment score, but having sentiment scores for each tweet was the end goal. To achieve this, the “group by” function was used to group each word with its corresponding tweet #, and then the “summarize” function was performed to sum the “value” variables for each of the groups. After that step, each of the datasets contained three variables: tweet #, date-time, and sentiment score of the tweet (numeric type, with only integers represented). The original “text” variable was added back using the “cbind” function to complete the datasets.

After the datasets were transformed to include sentiment scores, preliminary analysis was able to begin. The first analysis performed was comparison of the Trump dataset and the Trump overall dataset. A two-sample t-test was performed using the “value” variables from each of the datasets, using Hypothesis 1 (mean of Trump dataset “value” variable is greater than mean of Trump overall dataset “value” variable).

The same type of t-test was then performed again, this time comparing the Trump dataset with the general dataset “value” variables and testing Hypothesis 2 (mean of Trump dataset “value” variable is greater than mean of general dataset “value” variable). In addition to the significance testing, the two datasets were also compared visually in several ways. First, the datasets were further grouped and summarized to ascertain the mean sentiment values for each day within the time period along with the mean sentiment value for each month, and then those values were each displayed as column charts over time side-by-side for comparison.

Further comparison of these two datasets represent the only portion of analysis in this study that does not utilize tweet sentiment scores as its metric. To understand the differences in word selection that Trump uses compared to the general sentiment on Twitter, the datasets each underwent the “count” function (which counts the # of appearances of each word in each dataset). This allows us to see which words are most frequently used by both Trump and general Twitter in their tweets about the economy, and how those two groups are distinct from one another. It is important to note that only words ranked as positive or negative by the “Bing” lexicon are counted in this analysis, and any word commonly used by either party that is considered neutral or not included in the lexicon will not appear. For this reason, we can call this analysis “top words contributing to sentiment”. This analysis is visualized by using bar charts (for top 10 words only) and word comparison clouds (for top 100 words).

The main analysis attempts to reject the hypothesis 3 null, and if it succeeds in doing so, tests hypotheses 4 and 5. Through use of a for loop, a new dataset is generated that contains the average sentiment scores both before and after a Trump economy tweet for each hour from 2-24. For instance, the general sentiment score two hours before each of the Trump Tweets are recorded and averaged, and likewise for 2 hours after each Trump tweet. This is done for each hour integer for 2-24 hours creating 23 datasets of two variables (sentiment before and sentiment after) and 513 observations. Additionally, one data set of two variable (mean sentiment before and mean sentiment after) with 23 observations is created. For each of the 24 overall datasets, the “sentiment before” variable values are subtracted from the “sentiment after” variable values of each observation to create a new variable: “change in sentiment”. For the initial 23 data sets, we can run a t-test for the significance of each hour against the hypothesis: change in sentiment is not equal to zero. Rejecting the null for any of these 23 t-tests lend support to Hypothesis 3. The last dataset can be plotted over time to examine the trend of growth/decay in sentiment change over time in the general sentiment after a Trump economy tweet has been posted.

Depending on the results of this analysis, I design a test for hypotheses 4 and 5. This involves separating the Trump dataset into two subsets: Trump economy tweets that have positive sentiment scores and Trump economy tweets that have negative sentiment scores. I can then replicate the immediately previous analysis using these two subsets and use the results to plot two additional scatter plots over time which examine the growth/decay and direction of the change in general sentiment of the subsets.

Preliminary Analysis

*Trump Dataset vs. Trump Overall Dataset Sentiment Mean Comparison:*

*A black sign with white text

Description automatically generated*

The resulting mean sentiment score for the Trump dataset tweets is .9866 and the mean score for the Trump overall datasets tweets is .4138. On their face, these results suggest support for our Hypothesis 1 (mean sentiment of Trump tweets about the economy are greater than the mean sentiment of his other tweets). However, a two-sample t-test is necessary to prove that the resulting different means are unlikely to come about randomly.

A close up of a screen

Description automatically generated

The t-test returns a p-value of .00000002899, a remarkably small value that allows us to reject null hypothesis 1 with well greater than 99.99% certainty. There is support from these two datasets that Trump tweets more positively about the economy than he does other topics, which makes since considering that Trump addresses many subjects he views negatively as opposed to the economy such as: the House of Representatives, the impeachment inquiry, The Russian investigation, and his numerous political opponents. While a difference in mean scores of around .5 may seem small, it actually represents a relatively major difference in sentiment, considering the small text size for each observation as well as the way each word is valued. Another important point of analysis in this comparison is that both datasets average a positive sentiment. Overall, Trump’s tweets use a majority of words that score positively.

*Trump Dataset vs. General Dataset Sentiment Mean Comparison:*

*A black sign with white text

Description automatically generated*

The mean sentiment value for the Trump dataset is again .9866, but this time it is being compared to a much lower general sentiment mean of .1298. On average, the sentiment score of general Twitter about the economy is slightly positive. Comparing these means seems to support our Hypothesis 2 (mean of Trump sentiment about the economy is greater than mean of general Twitter sentiment about the economy) but let’s evaluate with significance testing before reaching any conclusions.

A close up of a screen

Description automatically generated

As should be expected, this time the resulting p-value is even smaller at 2.2x10^-16, meaning that we can again reject our null hypothesis 2 with almost 100% certainty. The rejection of the null at this level of certainty fosters strong support that Trump tweets about the economy are greater in sentiment score on average than general tweets about the economy.

A screenshot of a cell phone

Description automatically generated*Donald Trump Dataset vs. General Dataset Sentiment Scores Over Time:*

The charts above visualize the comparison between average sentiment over time for the Trump dataset and for the general dataset (orange is Trump dataset, green is general dataset). The top charts display the sentiment over time averaged by date, and the bottom displays the sentiment over time averaged by month for each data set. We can see a positive trend in sentiment for each of these charts, which can be reinforced by the means we calculated for the two datasets (.9866 and .1298). Especially when looking at the monthly charts, we can see that Donald Trump’s Twitter sentiment about the economy trends more positively than does general Twitter sentiment of the economy. Daily, Trump’s sentiment scores experience way more variability than does general Twitter. There is no apparent trend over time that can be seen from either of these graphs. On its face, neither general Twitter nor Donald Trump’s tweet have changed substantially over time.

*Trump Dataset vs. General Dataset Top Words Contributing to Sentiment*

*A screenshot of a social media post

Description automatically generated*Trump Dataset Top Words Contributing to Sentiment

From the chart above, we first can observe a large discrepancy between the frequency of top positive words and top negative words. The most common negative word’s frequency ranks only 6th overall, with the positive words “great’, “best”, “good”, “better”, and “strong” all being used more frequently. The top overall word, “great”, is also used way more frequently than any other word, with greater usage than the next three most common words *combined*. It also is important to note that some of the top negative words, namely “fake”, “phony”, “crime” and “collusion”, are not words commonly associated with the economy. It could be possible that many of the negative words in Trump’s tweets about the economy could be used when talking about a secondary topic within the same tweet, which is something that should be looked into as research into Donald Trump’s tweets progresses. How does Trump’s word choice when posting about the economy compare to the general sentiment on Twitter?

A picture containing screenshot

Description automatically generatedGeneral Twitter Top Word Contributing to Sentiment

The main difference in this comparison is that the top words contributing to general sentiment on Twitter are much more balanced between negative and positive words. In addition, “great” drops down to the 3rd most commonly used positive word to describe the economy, and 5th overall. We also don’t see the same strange negative words appear for the general dataset top words analysis that we did in Trump’s. They are replaced by words that seem to be more related to the economy, such as: “hurt”, “debt”, “slow”, and “crisis”.

*Trump Dataset vs. General Dataset Word Comparison Clouds:*

A screenshot of a cell phone

Description automatically generatedTrump Economy Tweet Comparison Cloud

General Economy Tweet Comparison Cloud

A screenshot of a cell phone

Description automatically generated

Much of the analysis between the comparison clouds and the top words visualizations are the same, so they won’t be repeated. However, the word clouds allow us to paint a better picture in the separation between negative and positive words between the two datasets. By comparing the two clouds, you can see that the cloud from the Trump dataset consists of a larger portion of blue words (positive words) than does the cloud for the general Twitter word usage . We can also see some more strange word usage from Trump that doesn’t seem to relate to the economy and isn’t represented in the other cloud. Words such as “illegal”, “stupid”, “hoax”, and “impeach” fall into this category.

Main Analysis

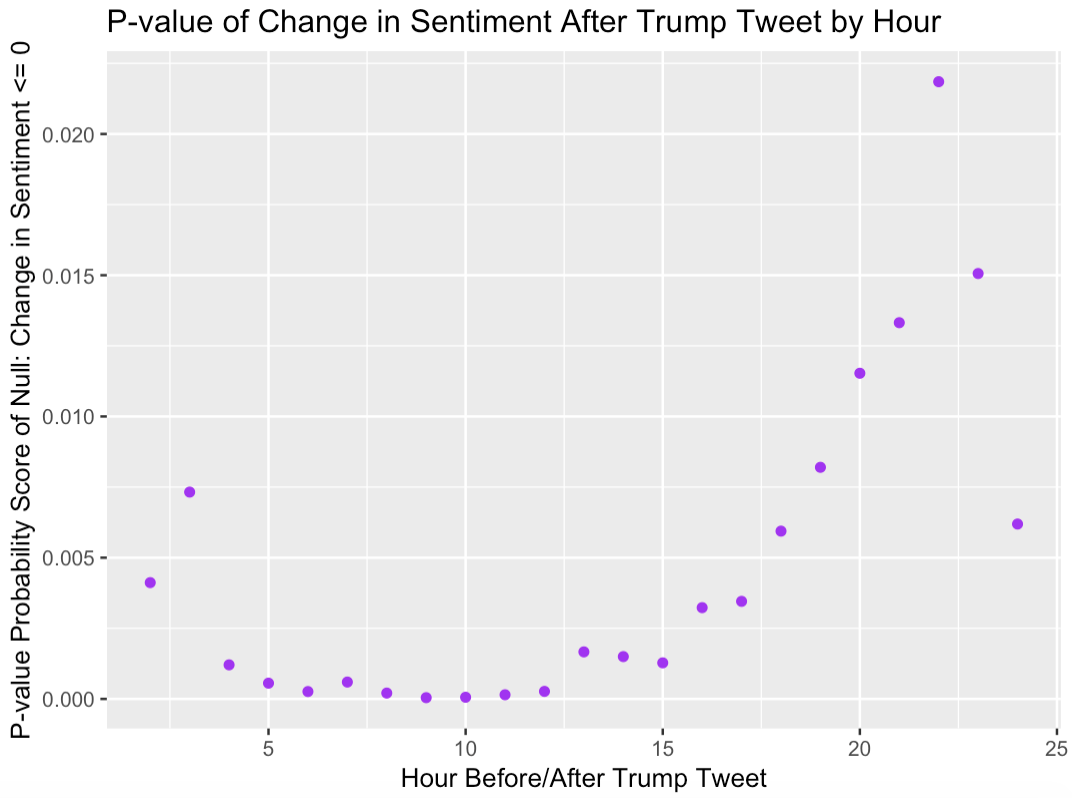
*Overall Analysis Testing Hypothesis 3:*

Change in Sentiment of General Twitter Over Time After Trump Economy Tweet

A close up of a device

Description automatically generated

After plotting the results of the for loop, it is observed that from each hour after a Trump tweet about the economy, the general Twitter sentiment surges slightly upwards. It is amazing that this result is true for all 23 hours, even though the change is fairly small. However, a .05-.09 trend upwards in sentiment can actually translate to Trump’s tweet changing sentiment for many people considering the scale of Twitter posts. In addition, it appears that the change in sentiment decays over time after a Trump tweet, with some variability in that decay. For this analysis, two significance tests are necessary. The first tests Hypothesis 3, that some change in the general sentiment exists (two-tailed t-test) each hour after Trump tweets about the economy. The second examines the significance level of the decay over time in sentiment change after Trump has posted (first differencing). For the first method of significance testing, the p-value for each hour t-test was calculated and plotted accordingly.

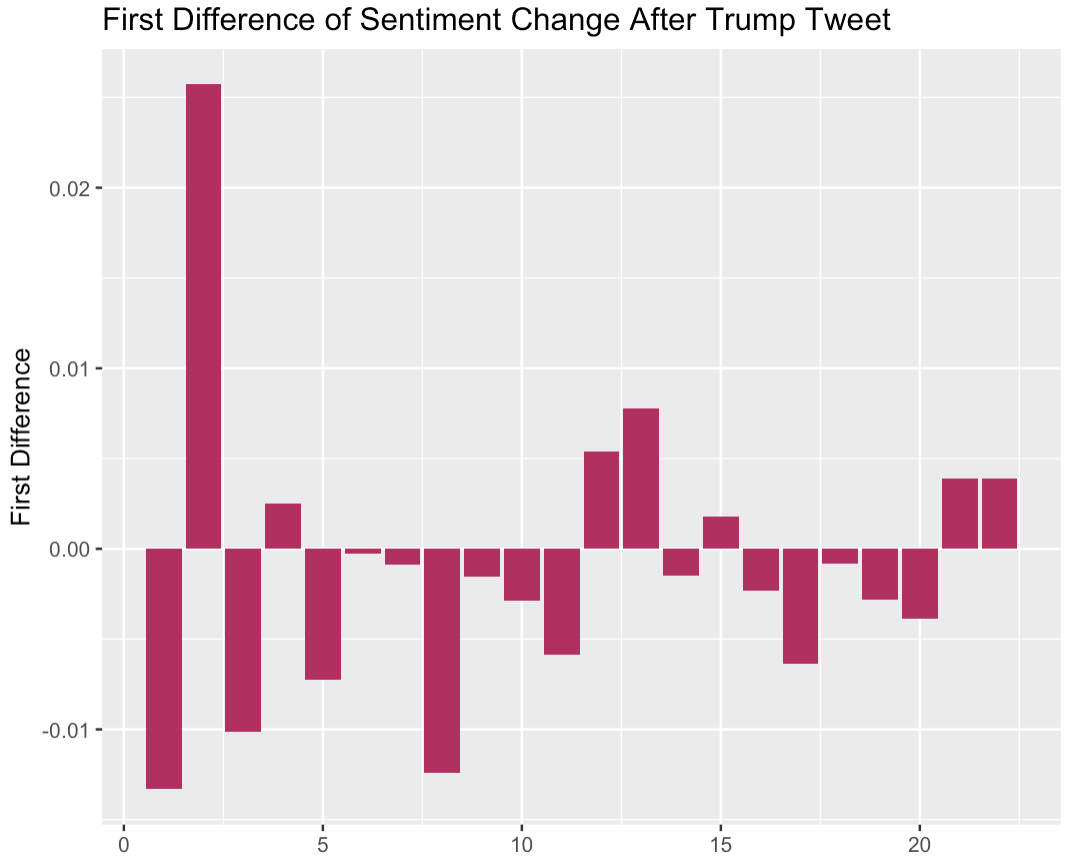


Although the p-value does increase over time, it does not climb of above .05, or even .03 for that matter. For each hour, we can be at least 97% confident in rejecting null hypothesis 3, and the ability to do so fosters great support Hypothesis 3 (mean of sentiment change after Trump tweet is not equal to zero). The upward trend up p-values makes sense, because as the sentiment change grows closer to zero over time, we are less confident in rejecting our null hypothesis that the sentiment change is different from zero.

A screenshot of a cell phone

Description automatically generated In examining the decay over time in sentiment change, I ran a test for stationarity (unit root test). With a p-value result of .4898 I was unable to reject the null, meaning that there is no evidence to support that the time-series exhibits stationarity.

Absent evidence of stationarity, we can perform first differencing on the plot to prepare for the Ljung-Box test, which will allow us to evaluate whether the apparent downward trend in sentiment change over time can be attributed to serial correlation, or rather just randomness. The results of first differencing are plotted below, along with the results from the Ljung-Box test (Hyndman and Athanasopoulos 2018).

A close up of a screen

Description automatically generated

From the graph, we can tell that most of the first differences are negative. This is because the differences are calculated by subtracting each data point from the one before it (can’t be done for first observations, so there are only 22 first differences). Since most of the points in our sentiment change over time means are smaller than the ones immediately before it, the first differences are mostly negative values. The Box-Ljung test evaluates the significance of the decay over time. With a p-value of .03428, we can say that we are over 95% confidence that we can reject the null hypothesis, that the data are independently distributed. With a relatively strong p-value, there is support that the sentiment change over time data exhibit serial correlation. Essentially, we are over 95% confident that the decay over time cannot be attributed to randomness.

*Positive Analysis Testing Hypothesis 4:*

A picture containing indoor

Description automatically generated

After performing the for loop again, but this time only including the tweets from the Trump dataset that score a positive sentiment, we can see that the resulting means are again overwhelming positive in sentiment change. No hour represents a negative average sentiment score, but the decay trend seems to erode when only including positive Trump tweets about the economy. This visualization seems to support Hypothesis 4 that positive Trump tweets about the economy sway general sentiment positively, but no significant test was performed to provide statistical evidence of this trend. The 23 for loops and subsequent t-tests were too time intensive to perform for the positive and negative subsets, as was the time-series analysis.

*Negative Analysis Testing Hypothesis 5:*

A close up of a white background

Description automatically generated

Surprisingly, the change in general sentiment remains positive even when Trump’s tweets about the economy score negatively. As has been touched on before, this can possibly be explained by subtopics or secondary topics being included in these tweets. Trump could be using positive words specifically about the economy, but that tweet could still score negatively if Trump uses negative words about another subject in the same tweet. This opens up an avenue for further research and possible topic modeling to analyze these strange results, as will be discussed later in the study when considering future research. The results from this analysis on its face refute and castsserious doubt upon Hypothesis 5, at least until follow-up research is performed. Interestingly, the decay over time is present in this subset analysis, while it was absent with the positive tweets subset.

Conclusions

Through the implementation of the analysis and significance testing, the study presented evidence supporting 4 of the laid-out hypotheses (Hypotheses 1-4). It has been uncovered that Trump’s tweet about the economy differ in substance and sentiment greatly from both the rest of his tweets and general tweets about the economy. Even more importantly, evidence has been presented that support the hypothesis that Trump’s tweets about the economy have an impact on the sentiment of tweets about the economy after. Further, that change in sentiment caused by his tweets about the economy decays over time. It seems that the sway in opinion has a short impact, but a well-timed tweet could realistically impact opinion before a vote or government event in Trump’s favor. A change in sentiment of .05-.1 points may seem to be insignificant, but with the scale of users on Twitter and the magnitude of people who sees his posts, it is reasonable to believe that even a small increase in sentiment change actually represent a significant number of peoples’ opinions affected. While this original study is far from perfect, there is room to continue and expand research. Discussion of limitations of this research study and future research direction is the first step towards that.

Limitations

The design of this research study has several limitations which could cast some doubt on the otherwise strong findings. One of these limitations has to do with the inherent sample bias when examining public opinion only on Twitter. Twitter users are not representative of the actual American public opinion about the economy. Most likely, the sentiment on Twitter about the economy comes predominantly from younger individuals, because they are more likely to use Twitter. Because of this, the findings of this study cannot be extended and generalized to support the stance that Trump’s statements sway public sentiment about the economy in his favor. We can only examine and draw inferences based on the population that our sample represents, which in this case is Twitter users.

Building on this idea, one also has to consider the propensity for individuals to post on Twitter at any given time. Many Twitter users will never create a post about the economy, and because of that fact sentiment analysis (and text analysis in general) does not tell the whole story of how Trump’s tweets actually affect the opinion of users who read it. The fact is that a large number of Twitter users will read a Trump tweet about the economy and not create a post about it. The question we find ourselves asking then, is how do the users that post about the economy truly represent all Twitter users. Perhaps not well, because those likely to post about the economy could align with ideological extremes on both sides, with moderates not caring enough to share their input on the topic. In addition, it is possible that a Trump post about the economy only encourages his supporters to post their already preconceived notions about the economy, skewing the sentiment change score. It is hard to tell if this is truly the case, and if so, do his posts similarly inflame his opposition to post their opinion? Further, if those who see his posts are more likely to post their own opinion about the economy, how many of his supporters see his post versus how many of his opposition (one would guess that most of his followers are his supporters).  We can only hope that the sentiment change captured by the analysis of actual posters after a Trump tweet does not stray too far from the true sentiment change of Twitter users.

Another limitation previously touched on in the main analysis is the possibility for extraneous discussion in Trump’s economy tweets. Not all Trump’s tweets cover one topic, and many touch on multiple different subjects and agendas that Trump wishes to promote. Unfortunately, this can be extended to our Trump dataset. Just because Trump mentions the words “economy” or “jobs” in his tweet does not mean that it is the only topic of discussion within the tweet, or even the main one. In future research, is it possible to filter out some of these Tweets so that they don’t contribute sentiment about other topics when we are trying to capture sentiment only on the economy? Speculatively, it seems counterintuitive for Trump to want to speak negatively in sentiment about the economy, considering that many Americans believe that the state of the economy is a main responsibility for the President. If anything, it is in his best interest to speak of the economy in an overwhelmingly positive light. As will be discussed in the upcoming section of future research, implementing some topic modeling in this research study could help us evaluate this limitation and respond accordingly.

Lastly, there is a possibility for a spurious relationship between our correlation in the main analysis. It might be the case that Trump is more likely to tweet about the economy after some event or news breaks related to the economy, essentially bragging about it. This could mean that the positive change in sentiment of general Twitter we see is due more to the event rather than Trump’s tweet itself. However, if this were truly the case, I suspect that we would see a larger increase in sentiment change than we actually do in the findings, casting doubt on this limitation’s plausibility. The topic of the economy was also selected mainly to negate this possibility, as major widespread news or events about the economy are less prevalent than with other topics, for example as opposed to gun control or the impeachment proceedings.

Future Research

Throughout the various steps in designing and implementing this study, various avenues have presented themselves for possible further research into this topic. The first is that this type of research can be expanded into analyzing different subjects rather than just the economy, along with different influencers rather than just Donald Trump. With the presence of Twitter and social media today, there is an endless breadth of people talking about a multitude of different topics at any given time. The strategies in this study can be generalized to examine the effect on the general sentiment that any influencer has on any given topic, and the ability to do that is powerful. In my personal opinion, it would be interesting to explore Donald Trump’s influence on the general sentiment about his own impeachment proceedings, as that topic has recently come to dominate most of the discussion of politics. Further, I also believe that these strategies could be interestingly implemented to examine how Trump’s tweets about specific political opponents (maybe Nancy Pelosi, or whoever wins the Democratic Presidential nomination) affect general Twitter’s opinion of them. These are only a couple of avenues in a seemingly endless combination of topics and influencers.

It would also be interesting to examine which characteristics of Trump tweets are most effective in changing the general sentiment positively. Could a certain time of day impact how receptive the general public is on Trump posts about the economy? Could certain verbiage used in Trump tweets be more effective in swaying Twitter opinion in his favor? These questions could be examined empirically, although it would be rather meticulous to do so.

Related to the strange findings that cast doubt on Hypothesis 5 in this study, some topic modeling could be performed to possibly explain the discrepancy. Based on the words and verbiage of the Trump dataset word clouds and most common words, it seems possible that some of the tweets that mention the economy (which are included in the Trump dataset) might not be the only topics being discussed in the tweet, which could skew sentiment scores. Topic modeling could help us understand why a Trump tweet about the economy might be scored as negative when sentiment analysis is performed, because it seems that speaking poorly of the economy works against Trump’s agenda (and 22.4% of his economy tweets are scored negatively). What other topics is Trump discussing in these tweets rather than just the economy that are contributing to sentiment scoring?

Finally, one last area of future research for this study would be the implementation of network analysis showing the connections that lead to the dispersion of Trump tweets to an audience much larger than just his own followers. This analysis would help show how Trump’s tweets have such an effect on the general sentiment of Twitter, by showing its means of dispersion and how many users his tweets truly reach.

Bibliography

Grolemund, Garrett and Hadley Wickham. 2019. *R for Data Science.* O’Reilly Media.

https://r4dz.had.co.nz/

Harris, Rich, Migliozzi, Blacki, Rosenberg, Matthew and Rachel Shorey. 2019. “How

Trump Reshaped the Presidency in Over 11,000 Tweets*.”*

<https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter->

presidency.html

Ho, Karl. Accessed 2019. “Big Data Management Workshop.”

https://datagenerations.org/bdm/

Hyndman, Rob J. and George Athanasopoulos. 2018. *Forecasting: Principles and*

*Practice*. Melbourne, Australia: OTexts.

Kumar, Shamanth, Morstatter, Fred and Huan Liu. 2014. *Twitter Data Analytics.* New

York, NY: Springer.

Lee, Jayeon and Weiai Xu. 2018. “The More Attacks, the More Retweets: Trump’s and

Clinton’s Agenda Setting on Twitter.” *Public Relations Review* 44(2): 203-213.

Silge, Julia and David Robinson. 2019. *Text Mining with R: A Tidy Approach*.

O’Reilly Media. https://www.tidytextmining.com/