Jacob Watkins

Data Collection Methods

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.

Literature Review

Using Twitter data for public opinion research has only recently become popular, with methods and strategies still evolving. Some advantages and disadvantages with the use of Twitter data in this context are necessary to understand before diving into new research.

While nobody has specifically utilized Twitter data as I intend to do so in the upcoming analysis, plenty have provided useful insight in their own research studies involving Twitter data. As Goritz, Kolleck and Jorgens recount, “Recently, Twitter has become a popular online platform for political actors to interact with each other. Politically influential individuals and organizations are increasingly using Twitter as a communication platform” (Goritz et al. 2019: pg. 2). The way Twitter has evolved in recent years, specifically through increased use of political actors to push their messages, provides an easy and relatively untapped faucet of public opinion to analyze. Compared with traditional methods of accruing data (focus groups, surveys), Twitter data is “found” data, it does require the time or resources that is typical of traditional data collection, and also avoids some of its pitfalls (Stephens-Davidowitz 2017). For instance, acquiescence bias does not exist, as the subjects (Twitter or social media) users rarely know or care that their posts could be used as data. In addition, other traditional survey biases such as demand characteristics bias, extreme responding bias, and question order bias do not exist with Twitter data collection.

Although some biases are eliminated when using Twitter data, others still exist. Social desirability bias remains prevalent, because even though Twitter users might not be aware or worried that their posts can be used as data, there is still a tendency for humans to post opinions about what they believe their friends (in this case followers) find desirable. They want to look good to their followers, even if that means the posts are not truthful or representative of their actual opinions. It would not be absurd to argue that social desirability is even more of a problem with Twitter data than it is in traditional data collection methods. Stephens-Davidowitz argues that the only online data platform that escapes this issue is Google search (since nobody else can specifically see what you search, there is no incentive to lie) (Stephen-Davidowitz 2017).

While utilizing Twitter data certainly has its benefits and drawbacks, the ultimate convenience makes it very attractive. Hours of time that would be spent collecting data can instead be harnessed elsewhere in research design or implementation for any given product. It is important, however, to keep in mind the challenges and possible biases in each step throughout the scientific process.

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.

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.

The first of these datasets (aka “Trump dataset”) is a congregation of Donald Trump tweets about the economy that were posted during his presidency. It consists of 513 observations (tweets) with 88 variables (including text, date-time posted, # of likes, # of comments, etc). The data was collected using Twitter API in conjunction with an R package called rtweet. Gaining access to Twitter API required that I apply for developer access through Twitter. This allowed me to query search Twitter’s full archive, setting filters to only collect data that fell within the appropriate time bounds. The keywords included in the query were “economy” and “jobs”. If a Trump tweet within the time period included one or both of those words, it was included in the dataset. The other obvious condition required to restrict the dataset was to only collect observations that were posted by the username “@realDonaldTrump”, which is President Trump’s Twitter username.

Because of Twitter’s restrictions on download capacity and its pricey subscription service, the second dataset (aka “general dataset” had to be collected through other means. This dataset includes tweets from any user on Twitter that mentions the economy within the time period. The ultimate resulting dataset contains 433,757 observations (tweets) of 12 variables (text and date-time included). I used a Python package called GetOldTweets3 to accumulate this data.. The keyword search here was “U.S. economy”, and any tweet that mentioned that phrase within the time period was included in the dataset. I then omitted the tweets included in this dataset from “@realDonaldTrump”, along with all retweeted Trump tweets, so that they would not skew the results of my sentiment change analysis.

The last dataset (aka “Trump overall dataset”) collected was a relatively minor one, which I use to test my Hypothesis 1 that Trump tweets more positively about the economy than he tweets overall. In order to test this, I had to gather a dataset that included all of Trump’s tweets during the given time period rather than only the ones on the economy. I again used GetOldTweets3 to accumulate this dataset, this time collecting all tweets from “@realDonaldTrump” during the time period without limiting by any keywords. The resulting dataset included 8050 observations (tweets) with 12 variables (text and date-time created included). I then omitted the observations that were already included in the Trump economy tweet dataset, so that I could compare Trump tweets about the economy to the rest of his tweets. As mentioned previously, Trump has tweeted over 11,000 times during his presidency. The dataset instead consists of only 8,050 tweets because all Trump tweets that were retweets were excluded (this is automatically done when using GetOldTweets3).

Data Collection Problems/Missing Data Problems

The methods used for collecting the three datasets certainly have their flaws. The first and main flaw is inherent in all social media sampling. Twitter provides an endless stream of messy data, and it is impossible to download or even know the whole sampling frame that you wish to sample from. Part of this issue was solved by placing this study and the datasets within a certain time period, but in doing so data is surely missed out on, specifically data that has accumulated after 9/20/19 when I first began to extract my data. However, even if I went back and added observations from 9/20/19 up until the current date, some observations would still be missed out on, as people post on Twitter about the economy every couple of seconds on average.

Building upon this idea, the datasets collected are but a sample of an unknown sampling frame. The Trump dataset and general dataset seek to capture opinions about the economy, but the economy is a broad subject, and not all posts that discuss the economy will be captured by keyword searches such as “economy” or “jobs”. How representative are these tweets that mention these keywords of the overall population of the discussion about the economy that we wish to capture? In addition to this, since not all Twitter users post their opinion about the economy, it is hard to use these datasets to draw inferences of how Trump is changing *opinion*. The concepts “sentiment” and “opinion” are undoubtedly related, as sentiment seeks to quantify written opinion. Opinion, however, is not always written or recorded, and we do not know whether written opinions actually represent overall opinion.

Regarding key word searches to accumulate data, there is also ample opportunity for extraneous data to be included in our datasets that should not. We can search for tweets that only include certain words, but that does not preclude other topics that do not relate from being included. This proposes a problem when attempting to use sentiment analysis, because you want to use data that is exclusive to your subject so that sentiment scores will not be skewed. It is virtually impossible to limit all extraneous data, but if you can identify which unrelated topics are frequently included in your data, it is possible to exclude them by keyword. Topic modeling would help to do identify main topics in the three datasets, and the possibility for that analysis in future research will be discussed more thoroughly in coming sections to address this limitation.

This particular problem was especially present when trying to accumulate the general dataset. The Trump dataset was able to accumulate fairly relevant data using the keywords “economy” and “jobs”, but when the same words were used to collect data for the general dataset, a mess of irrelevant tweets were present. When Trump uses the phrases “jobs”, he is usually speaking about job creation, but that is not the case for other Twitter users. They might speak of how they hate their job, but that does not specifically relate to the economy or their opinion of such. The keyword “economy” proved to be even more problematic. Even when the query search was limited by geographic location (using –near “United States, United States“ specification in GetOldTweets3), a fair portion of the data discussed the world economy, or the economy of another nation rather than the United States. The ultimate fix to this problem was to use “U.S. economy” as the keyword to accumulate the general dataset, but that decision did not come without cost. Firstly, many valuable tweets that mention just the keyword “economy” and actually discuss the U.S. economy were lost, because the inclusion of U.S. is very formal for most Twitter posts. Building on this, I believe that the general dataset could overrepresent news companies’ Twitter posts as well as the Twitter posts of political figures, because they are much more likely to use the whole “U.S. economy” phrase than the average citizen discussing the economy on Twitter is. Ultimately, some data was lost by using this keyword search, but the alternative was including a wealth of extraneous data about the economies of numerous other nations, which would skew the sentiment irreparably.

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. The tweets within each dataset were broken down into words and were then assigned a sentiment score using the “Bing” lexicon. Words were either assigned to be either “positive” (+1 in numeric value) or “negative” (-1 in numeric value). The words in each tweet could then be summed into an overall tweet sentiment score, which could be further manipulated (averaged, grouped, etc.) to perform analysis.

After the datasets went under this transformation, 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 comparing the top 10 positive and negative words by both Trump about the economy on Twitter as well as general Twitter.

The main analysis attempts to reject the hypothesis 3 null. 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 variables (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.

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”.

Main Analysis

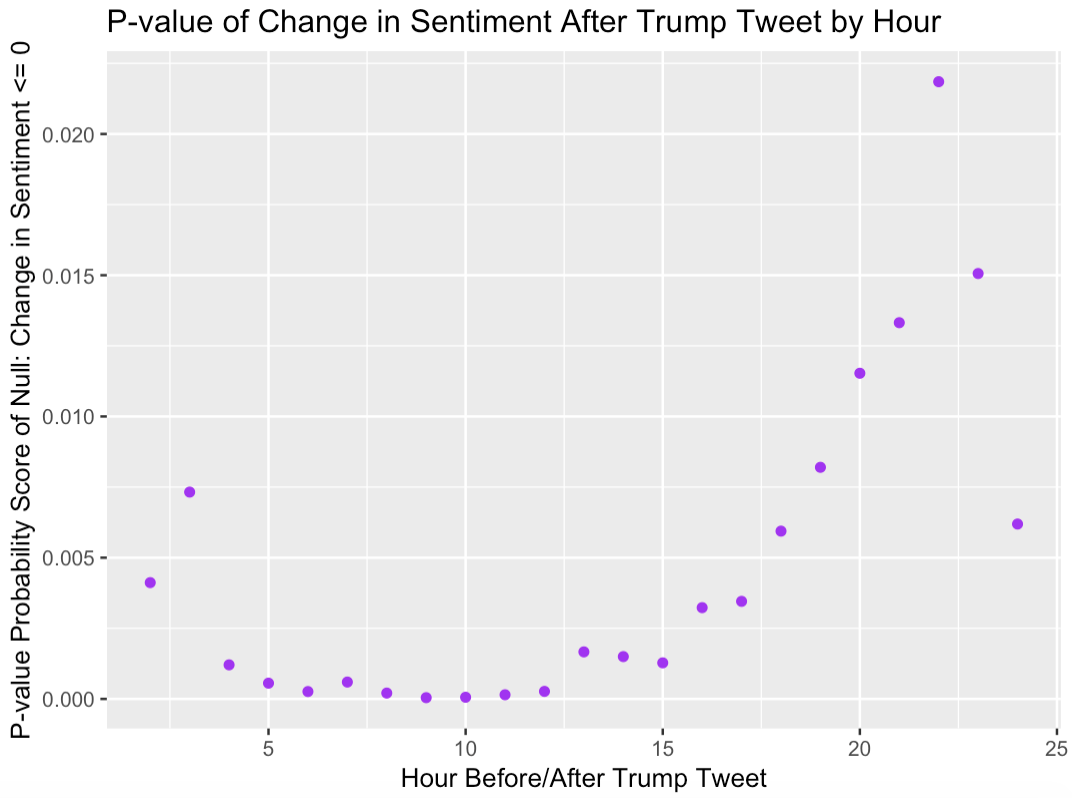
*Overall Analysis Testing Hypothesis 3:*

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 points after a post, 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. 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.

Conclusions

Through the implementation of the analysis and significance testing, the study presented evidence supporting the 3 laid-out hypotheses. 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 supports 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 any given 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.

As previously mentioned, some topic modeling could be performed to evaluate the accuracy of the datasets in gauging only sentiment related to the economy. Based on the words and verbiage of the Trump dataset 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

Goritz, Alexandra, Kolleck, Nina and Helge Jorgens. 2019. “Analyzing Twitter Data:

Advantages and Challenges in the Study of UN Climate Negotiations.” *Sage*

*Journals.*

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/>

Stephens-Davidowitz, Seth. 2017. *Everybody Lies: Big Data, New Data and What the*

*Internet Can Tell Us About Who We Really Are.* New York, NY: HarperCollins

Publishers.