USF Tampa Graduate Theses and Dissertations   USF Graduate Theses and Dissertations

November 2017

The Road to the White House: A Correlational Analysis of Twitter Sentiment and National Polls in the 2016 Election Cycle

Melissa G. Pelletier
University of South Florida, melisss222@yahoo.com

The Road to the White House:

A Correlational Analysis of Twitter Sentiment and National Polls in the 2016 Election Cycle

by

Melissa G. Pelletier

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
with a concentration in Media Studies
The Zimmerman School of Advertising and Mass Communications
College of Arts and Sciences
University of South Florida

Major Professor: Kelli S. Burns, Ph.D.
Roxanne Watson, Ph.D.
Scott Liu, Ph.D.

Date of Approval:
October 20, 2017

DEDICATION

I dedicate my thesis to my family and friends. A special feeling of gratitude goes to my mother Gail Pelletier, as you have been my biggest cheerleader throughout my graduate school endeavor. You pushed me to move forward when I did not think it was possible. Your love, encouragement, and endless support will forever be remembered. My grandmother Hazel Missig also provided much needed encouragement throughout this process on several occasions, and for that I am grateful. I would also like to dedicate this work in loving memory of my father Kenneth Pelletier and grandfather Gerald Missig. I looked to you for spiritual guidance many times during this process, in which I always received in one way or another. I know you two are looking down upon me smiling. A special thank you to my sister Kristen for constantly reminding me that I can do anything I want through hard work and dedication. I also dedicate this thesis to my friends who helped through this process with caring words and inspiration. Rickey, thank you for always wiping the tears and telling me I can finish. Andrea, thank you for your drill sergeant tactics which inspired me to keep going. Kelly thank you for always listening when I would complain and reassuring me that I cannot be stopped. Cousin Morgan I appreciate your technical expertise. Lastly, Bryce, Xay, Morgan,and Murewa, thank you for being the best teammates, friends, and proofreaders anyone could ever hope to have.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

In this thesis, the author examines the last 131 days of the 2016 election cycle. This analysis focuses on how sentiment is present on Twitter when people engage in political communication on social media. With the increasing online political discussions created on social media such as Twitter, an analysis of sentiment is critical. The data could be obtainable for candidates to estimate the electorate's opinion of each candidate. A shift of sentiment offers a deeper insight into tracking changing attitudes toward candidates. Because Twitter only allows each tweet to be 140 characters there is a simplicity that offers statements to be concise. Trends for each candidate throughout the final days of the election cycle are correlated with national polls to assess if there is a relationship present. This study applies sentiment to recognize trends that may estimate a candidate's chance of winning the election and offers indications as to how the intended electorate may vote when a relationship is established between sentiment and national polls.

CHAPTER 1:

INTRODUCTION

President Barack Obama and other political candidates running for public service offices began using social media during the 2008 presidential election cycle to broaden the means through which they could spread political messages. In fact, President Obama sent a tweet after he won the election in 2008 stating, "We just made history." All of this happened because you gave your time, talent, and passion. All of this happened because of you. Thanks" on Twitter from his personal Twitter account. The use of social media altered the means by which the game of political communication occurred while also changing political campaigning forever. As of 2012, there were 500 million Facebook users and accounts, with 100 million Twitter users, which increased to 170 million Twitter users in 2013 (Stieglitz & Dang-Xuan, 2012; Wladarsch, & Neuberger, 2013). These numbers show tremendous growth within a span of one calendar year and continue to increase yearly. It is important for candidates to stay current with technology and with the presumed electorates regardless of the method of social media they choose to utilize.

1.1: 2016 Election Campaigns

Political communication occurs on Twitter from political players and candidates daily. The election cycle of 2016 was no exception. However, the usage of social media differ from that by candidates in prior election cycles, as the candidates President Donald Trump and Secretary Hillary Clinton used social media as an avenue for hurling insults and personal attacks against each other. The electorate joined in and insults were traded between party loyalists and

candidates. Self-efficacy of politics was present as issues and agendas were discussed between social media users, highlighting evidence of electorate involvement never seen in such a public way.

Presidential debates and other events of the past offer material for social media users to discuss on Twitter as well as other social media outlets. The first presidential debate in 2012 recorded 10 million tweets shared by 170+ million users (Flynn, 2016). The 2016 election cycle was similar, except it recorded more viewers than in 2012. Sixty-seven million viewers, an estimated number, watched the first debate in 2012 via television networks, and 80 million viewers watched it in 2016 (Flynn, 2016). The same phenomenon can be observed on social media sites as some 369,000 users watched the debate via live stream on Twitter during the 2016 election cycle (Kafka & Wagner, 2016). This probably is because Twitter currently has more than 313 million users to date, which continues to grow daily and has increased from 185 million users in 2012, essentially increasing the amount of political chatter in 2016 (Flynn, 2016).

Neil Young's *Rockin' in the Free World* played stridently at the Trump Tower on June 16, 2015 as Donald Trump rode down the golden escalator to announce his bid for presidency. This was the beginning of the most tumultuous American presidential election cycle to date. No one living within our borders or abroad will soon forget the 2016 presidential cycle. Controversy occurred daily extending from the primaries until Election Day on November 8, 2016. Media frenzy occurred in every newspaper, social media site, and on every televised network locally, nationally, and internationally. The presidential hopefuls utilized social media sites as a platform to reach the voting public. The primaries in the election cycle of 2016 were as chaotic and tumultuous as the general election, considering the bid for both Republican and Democratic nominations were up for grabs. Seventeen Republicans eventually dove into presidential race of

2016 beginning with Ted Cruz on March 23, 2015 (Bialik, 2016). Only three candidates remained in the primaries as of March 16, 2016 until the Republican National Convention when Trump was nominated (Bialik, 2016).

Hillary Clinton was the last woman standing as she eagerly awaited entry back into the White House, this time not as First Lady, but as President of the United States. History was being made as Clinton won primaries in California, New Jersey, South Dakota, and New Mexico securing the Super Delegates she needed to move forward on June 8, 2015 (Collinson, 2016). The Democrats had a total of six candidates enter the race with the first candidate being Clinton who entered on April 12, 2015. After the Iowa caucus on February 1, 2016, only two candidates remained in the race up until the Democratic National Convention when Clinton was nominated. After the Republican and Democratic National Conventions, there were two remaining candidates vying for a seat at the desk in the oval office.

1.2: Social Media Usage

Twitter is a micro-blogging social media site that allows 140 characters per tweet. Twitter is used in many ways and for many reasons. Generally, users share thoughts on various subjects such as foods, movie stars, and other newsworthy topics (Ronsenstiel, Sonderman, Loker, Ivancin, & Kjarval, 2015). Some use Twitter for recreation, others use Twitter to promote business through marketing. Twitter experts, marketing firms, and business leaders, often use Twitter for brand management and brand awareness (Lonoff Schiff, 2013). Twitter is also a social media tool that can be used to make others aware of businesses and services (Lonoff Schiff, 2013). Many use Twitter to tweet feelings about sports or whatever may be occurring in their daily activities. Twitter has become a viable avenue for the access and consumption of news

as Ronsenstiel et al. (2015) conducted a study that concluded that 81 percent of the participants checked their news daily on Twitter.

Presidential candidates Donald Trump and Hillary Clinton have both used social media more than candidates during any other prior election cycle. President Trump and Clinton also used social media in a different manner than President Obama and Mitt Romney in the 2012 election cycle. The continuing growth of the population on Twitter and the more in depth usage by political candidates offered a platform for political discourse in addition to political communication. The election cycle of 2016 was unique in the ways social media was leveraged. Twitter ran a live feed of all three presidential debates in the general election for the first time. This allowed a population that may not watch news networks to tune into the debate. In addition, Twitter users could interactively tweet during the debates. In addition, candidates displayed their response of pleasure or displeasure to certain debate questions tweeting comments and rebuttals after debates ended which continued the political communication and fueled discourse. Twitter created an official hashtag and emoji for all three of the live debates in the general election: #Debates, #debatenight, and #Debates2016 (Flynn, 2016). The 2016 presidential candidates used social media differently and more frequently than past candidates, attempting to tweet their way into the public sphere while spreading political messages.

During political election cycles, campaign managers traditionally use polls to drive important aspects of the campaign while interested people use it to predict which candidate will gain the electoral votes from certain states or regions. Due to the evolution of current technology, individuals can communicate sentiment toward candidates on social media during important events of an election cycle and interested parties have the ability through software to track these sentiments. Certain events often occur during a political race that cause concern or satisfy

individual voters. The electorate then express their approval or contempt, which could be positive, negative, or neutral (absent of negative or positive emotion). By tracking the sentiment of electorates, candidates can adjust their talking points and pivot to speak more on particular issues that the polls show as most important to the electorates. Candidates can also begin to address issues raised by the electorates in order to gain approval and win the election. Other interested people also use polls to predict which candidate will gain the electoral votes from certain states during the election cycle. Polls are also used during the last hours of the Election Day to predict the winner of the election as eager electorates leave the precinct after casting their vote.

Researching political communication and political constructs of the recent candidates in addition to social media sites such as Twitter set a litmus test for upcoming political races. This litmus offers a helpful guide for upcoming political boxing matches. Sentiment analysis in addition to opinion mining of the voting electorate demonstrates the usage of social media in which an exchange of political communication occurs. Political communication often serves the purpose of framing messages in a certain way that ensures a certain candidate or personal favorite is viewed more positively or negatively. Messages via social media are framed differently depending on the opinion of the particular electorate. Candidates may find that opinion mining or sentiment analysis produce a helpful guide for candidates to follow in future elections.

1.3: Polling

Polls are often difficult to conduct. They require a significant amount of labor for activities like making telephone calls or going door-to-door. Therefore, polling demands human

resources, this often contributes to the high financial costs, related to the conduction of polls. However, polling social media users, especially Twitter users, lowers the cost of polling and increases the sample population. Pollsters can easily poll people from every area of the country. The issue with narrowing a population to one certain social media site would be getting an ideal sample population that would accurately represent the electorate, including both parties and all demographics. However, great advantages become present through social media and modern technology, as interested individuals can poll electorates who communicate sentiments for political candidates at any time throughout the election cycle. Polls are also used during the last hours of the Election Day to predict which candidate will be victorious. Exit polls attempt to gauge which candidate is leading the political race as electorates exit voting locations, which at times can be inaccurate for different reasons. Electorates may not be truthful during exit polls, as they do not care to share their vote with others. Most exit polls were inaccurate in 2016 and attributed to the "shy Trumper" hypothesis, which states that those who voted for Trump did not want to say for fear of backlash (Mercer, Deane, & McGeeney, 2016). Furthermore, early voting is not accounted for in exit polls. Mercer et al., (2016) stated, "Statisticians say that exit poll data, while well-intentioned, is inherently flawed as a way to predict final vote totals. Due to the need to compile nearly instantaneous results, exit pollsters rely on statistical models that may be outdated by the time an election rolls around" (Mercer, Deane, & McGeeney, 2016, para 4).

Digital polling, on the other hand, especially with Twitter users, lowers the cost of polling and increases the sample population. Social media polls, especially those conducted on Twitter users, allows pollsters to reach a larger geographical reach. Pollsters can easily poll people from every area of the country and outside the United States. Because of the ease of use or possibility of limited funding, interested individuals can poll electorates about their sentiments

for political candidates continuously throughout the election cycle. However, polling a Twitter population has its disadvantages. Minus the question of validity of social media polls, pollsters have to put the age distribution of the sample into context. Approximately 24% of adults who use social media use Twitter (Greenwood, Perrin, & Duggan, 2016). Greenwood, Perrin, and Duggan (2016) assessed "Younger Americans are more likely than older Americans to be on Twitter. Some 36% of online adults ages 18-29 are on the social network, more than triple the share among online adults ages 65 and older" (para, 10). The authors also found through Pew Research Center surveys that Twitter is used more by adults with college educations (29%) as opposed to those (20 %) with a high school diploma or less (Greenwood, Perrin, & Duggan, 2016). It may not yet be possible to represent the electorate fully. While an overall assessment can be made, the question remains whether Twitter sentiment analyses is a productive way to poll the electorate and can we confidently ascertain that sentiment analysis is viable as a blaring signal? Altogether, we may be unable to actually predict the winner of an election. However, a story may be told by collecting sentiment over a specific time-period of the election.

CHAPTER 2:

Literature Review

2.1: Social Media and Political Communication

Stieglitz and Dang-Xuan (2012) suggest microblogging on Twitter use may increase political participation for some users. Stieglitz and Dang-Xuan (2012) assess Twitter as being different from other social media sites. Facebook usually consists of communities of friends who are familiar with each other in real life settings and much of the time have actually met in person, whereas Twitter users connect by using hashtags and similar preferences by entering key words preceded by hashtags in search bars within Twitter. Following individuals is different on Twitter than on other social media sites and often people follow a population that has similar views to their own. Pennacchiotti and Popescu (2011) state, "Intuitively, Democrats are more likely to follow the accounts of Democratic politicians and Republicans those of Republican politicians" (p. 9).

Dang-Xuan et al. (2013) believe through research social media is a legitimate avenue for shared political information. Social media has broadened the scope of gate keeping and agenda setting. In previous presidential elections, network television performed the roles of gate keeping and agenda setting. Media and political communication amongst electorates can occur every minute of every day, as opposed to waiting on the nightly news to share political information. This allows electorates to participate in their own agenda setting. Presently, micro-blogging through Twitter allows several million authors to frame messages in ways that apply to personal agendas. Dang-Xuan et al. (2013) ascertain that contemporary democracies develop avenues

through social media to engage with constituents before, during, and after political campaigns. Citizens are able to spread political messages or information through retweets on Twitter (Dang-Xuan et al., 2013).

Twitter studies have produced empirical results for several paradigms that link political communication to candidates and electorates. Journalists also find Twitter to be useful for political communication as a tool for spreading messages or setting agendas in many other instances. Broersma and Graham (2012) conducted a content analysis of Twitter messages concerning the 2010 British and Dutch elections. The goal of the study was to understand who was using Twitter, and if Twitter contributed to the print and online news headlines (Broersma & Graham, 2012). Broersma and Graham (2012) found evidence that tweets were used in headline news stories for journalists in newspapers and for politicians in tweets when scandals appeared to develop on Twitter. Evidence of tweets triggering newsworthy stories and headlines occurred more in the British election than the Dutch election (Broersma & Graham, 2012). The tweets were classified by those who authored them and by their purpose in tweeting. The category of authors was segmented into politician, expert, or cultural producer. The functions of tweets were coded as either triggers or sources of news' headlines generation (Broersma & Graham, 2012). The researchers looked at mainstream media such as respected newspapers in the U.K. and Netherlands.

2.2: Prediction based on Twitter Sentiment

Researchers have explored whether Twitter sentiment emotions have predictive power. These relationships are by no means causal. Research completed in these areas help to ascertain the usefulness of Twitter sentiment for game observations, which in essence, may work in

political paradigms (Diakopoulos & Shamma, 2010). Bollen, Mao, and Zeng (2010) conducted studies for predicting stock market value rise/fall. The authors found predictive values relating to emotions on Twitter more than measurements of negative and positive sentiment. As previously stated the prediction can only truly represent a relationship, as emotions cannot be a causation of market rise and fall.

Intensive work has been done in the past in measuring predictive factors on social media in the political arena. Tumasjan, Sprenger, Sandner, and Welpe (2011) analyzed 100,000 tweets during a German election in 2009. An aggregated sentiment analysis found that the population in fact used Twitter to deliberate opinions of candidates. Tumasjan et al. (2011) stated,

> Our results provide evidence supporting our theory that microblogging
>
> forums provide a mechanism for weighing information and that, despite
>
> individual biases, errors can cancel each other out. The predictive accuracy is
>
> even more impressive when compared to the track record of the IEM, a
>
> prediction market set up with the explicit purpose to predict election results
>
> (pg. 414).

Prediction values tended to coincide with traditional polling as researchers suspected. Nonetheless, Tumasjan et al. (2011) suggest that predictive values of Twitter are not a stand-alone method and should only compliment traditional polling as opposed to replacing polling entirely.

O'Connor, Balasubramanyan, Routledge, and Smith (2010) composed a time series sentiment analysis of public opinion while comparing the sentiment of candidates on Twitter to the sentiment in traditional polling. The authors implemented a forecasting value that would

indicate what polls would offer in future predictions. In essence, text sentiment proved to be a superior predictor at a certain milestone in the empirical work cited. Certain issues and periods offered predictive values through textual analysis as an increase of electoral confidence occurred, which also is evidence of a linear relationship (O'Connor et al., 2010).

On the other hand, some feel that predicting election results with Twitter may be an impossible feat. Burch (2015) conducted a primary study using Sysomos machine learning to mine opinions in a sentiment analysis. The question arose repeatedly throughout this literature review about whether social media sentiment analysis was a more effective prediction tool than traditional polls in recent elections. In this study, volumes of mentions as well as sentiment were analyzed, along with traditional polling, which had been the focus in many previous studies mentioned. Several periods were relevant in collecting data, as this was a two-fold experiment, which continued with the goal of predicting the primary winners in several states. The evidence in this study to date displayed a larger following for some candidates such as Bernie Sanders on Twitter, with Sanders continuously ahead of Clinton on social media when measuring volume (Burch, 2015). According to Burch (2015), Clinton was ahead in traditional polling by an overwhelming 54% to Sanders' 33%. However, looking at the Twitter conversation volume around each candidate, Sanders led the way and did consistently over the time analyzed.

2.3 Sentiment Studies

Researchers have also compared the outcome of sentiment studies that run congruently to traditional surveys. Mitchell and Hitlin (2013) conducted a yearlong research study comparing results of sentiment from Twitter and traditional surveys, using eight political events that occurred throughout the 2012 election cycle as a measuring tool. Mitchell and Hitlin (2013) ascertained that depending on whether the topic was considered more liberal or conservative, the

sentiment rated higher or lower for social issues attached to certain party lines. In some instances, such as gay marriage rulings, sentiment is altered according to social settings or exposure. People often pretend to be more liberal on Twitter but more conservative when actually answering a traditional survey (Mitchell & Hitlin, 2013). These findings are hypothesized differently because of the population tweeting about the specific topic changes according to the topic as mentioned above. Topics that leaned more conservative were tweeted more by conservatives and vice versa with topics that leaned more liberal. They disproved the belief that Twitter polls result in more liberal results than surveys. Twitter conversations occurring about the presidential candidates Obama and Romney in 2012 were overwhelmingly more negative than positive. However, Romney had a larger negative sentiment percentage in national polling and in Twitter sentiment in most instances except by the first debate (Mitchell & Hitlin, 2013). Obama and Romney sparred in the first debate bringing back hope to conservatives as the Obama stumbled several times (MacAskill, 2012). CNN conducted a poll that evening with 67 percent saying Romney was a clear winner (MacAskill, 2012). Twitter sentiment varied after President Obama was re-elected. Twitter represented a more positive sentiment than did polling by Pew Research polls (2013). Mitchell and Hitlin (2013) explained that limitations were present in their study because, "those who get the news on Twitter and those who tweet news are very different demographically from the public" (p. 1).

Obviously, everyone who is tweeting is not necessarily participating in political communication. However, many were in the past three election cycles and sentiment analysis offered predictive nuances for debate winners. Cody et al. (2015) states,

> Twitter has also been used to examine human sentiment through analysis of
> variations in the specific words used by individuals. Dodds et al. develop the

12

"hedonometer" a tool for measuring expressed happiness—positive and negative

sentiment—in large-scale text corpora (p. 2).

With these emotive expressions of happiness or disappointment through negative and positive

sentiment researchers can hypothesize as to which candidates are in the lead in the debates and at

different moments throughout the election cycle due to sentiment.

Pew researchers Rosenstiel and Jurkowitz (2011) conducted another detailed analysis of

Twitter in the presidential election of 2012 that differed some from that done by Mitchell and

Hitlin (2013). Rosenstiel and Jurkowitz compared Twitter sentiment to the blogospheres, which

was "more voluminous, more fluid, and even less neutral" (2011, p. 2). After the first

comparison of blogs to tweets occurred, a second comparison was analyzed by comparing

blogs/tweets to mainstream news such as network televisions. The sample contained 20 million

tweets that fluctuated according to certain events throughout the 2012 presidential election cycle.

The authors used two methods for analyzing data and coding tweets and blogs for sentiment.

First, a content analysis ensued to ascertain the quantity of exposure on Twitter and blogs.

Secondly, Crimson Hexagon technology (computer coding) was used, which allows a computer

to code a large data set containing millions of tweets while also coding a small number of tweets

in the beginning manually to ensure categories are mutually exhaustive (Rosenstiel & Jurkowitz,

2011, p. 28). Both political blogs and tweets were run through the Crimson Hexagon to gauge

sentiment. Because blogs often contain several assertions, only statements that contain the

candidate's names were utilized for sentiment (Rosenstiel & Jurkowitz, 2011). Findings in this

particular study displayed a greater negative sentiment on social media such as Twitter and blogs

and less negative sentiment on television broadcasts concerning the candidates (Rosenstiel &

Jurkowitz, 2011). The presidential election of 2012 was similar to 2016 with several GOP

candidates in the primaries. Rosenstiel and Jurkowitz (2011) measured sentiment on the three outlets for all candidates and only found a positive sentiment being greater than a negative sentiment on a couple occasions between May 2 and November 27 for the two candidates. Of course, negative and positive sentiment differed according to what milestones were occurring during each candidate's campaign.

Bollen, Mao, and Pepe (2011) also conducted a sentiment analysis in the latter part of 2008 using tweets as data, along with a specific timeline from August 1 to December 1. These particular authors chose to compare socio-economic events alongside mood patterns mined during a sentiment analysis. Social and economic indicators could be events such as the presidential elections, Twitter mood, and stock market fluctuations, or the death of a favorite celebrity (Bollen, Mao & Pepe, 2011). About 9,664,952 million tweets were compiled and compared to profile of mood (POMS-ex), a psychometric scale that originates from POMS (Bollen, Mao & Pepe, 2011). "POMS measures six individual dimensions of mood, namely tension, depression, anger, vigor, fatigue, and confusion, not intended for a large scale textual analysis" (Bollen, Mao & Pepe, 2011, pg. 451). Measuring POMS normally occurs through a questionnaire format given to live subjects. POMS-ex differs in the way data is collected and received. POMS-ex acquires a large amount of text virtually through social media or electronic media. Questionnaires are not administered to human participants (Bollen, Mao & Pepe, 2011). This analysis was more about proving that machine mining as well as machine learning produces accurate results from large data. However, researchers did find that significant events occurring that are political in nature could be correlated with several mood dimensions that fluctuate throughout events (Bollen, Mao & Pepe, 2011).

Sentiment continues to be relevant when analyzing microblogging and political communication. As candidates increase their following via Twitter, public opinion mining becomes more prevalent to understand how a candidate is performing on the campaign trail. As with any campaign, there will be highlights and lowlights that alter positive and negative feelings from the vocal social media users. Wang, Can, Kazemzadeh, Bar, and Narayanan (2012) conducted a real-time Twitter analysis during the 2012 presidential election cycle. Many Twitter sentiment studies account for positive, negative, and neutral opinions toward candidates. Wang et al. (2012) included a category, named unsure, that normally is not included in most sentiment analyses. Therefore, instead of classifying tweets in the three categories normally used in twitter sentiment, four categories were utilized. The stated goal and approach of this empirical endeavor was to combine real-time statistical sentiment through modeling, while gaining an understanding of social and political praxes through social media, especially on Twitter (Wang et al., 2012). This study found that special events have the ability to increase tweet volume and the proposed sentiment model is sufficient to evaluate public sentiment during real-time events (Wang et al., 2012).

2.4: Studies Comparing Polls to Twitter

Anuta, Churchin, and Lou (2017) conducted an experiment to first gauge whether or not the polls of the 2016 election were biased toward one candidate over the other, and secondly, to research whether Twitter would be useful as a less biased predictor for the last presidential cycle than polls. Data was gathered from several polls and several states throughout the election cycle to analyze if polls were biased regarding the popular vote. The authors created a prediction model that would detect bias for the popular vote only. Anuta, Churchin, and Lou (2017) chose nine states in total to analyze. The authors assumed the states that leaned liberal, conservative,

and finally that were battleground or swing states might produce superlative results. For the popular and electoral data concerning Twitter sentiment, the authors used specific tweets from Twitter API generated from certain areas of the United States identical to the states used for polling information. A sentiment analysis was completed on 750,000 tweets using a program named Python, which contains the sentiment tool VADER (p. 4). The results of this study yielded biases from the eight named sources of media. Anuta, Churchin, and Lou (2017) stated,

> In the 2016 U.S. election, the media (as encapsulated by our 8 sources) was, quantifiably biased against Donald Trump by -2.0% in the popular vote and -1.6% in the state based votes over the entire election period. Towards the end of the election (in the 3-month period before Election Day), the popular vote bias decreased slightly to a -1.0% bias against Donald Trump (p. 10).

Twitter encapsulated results that were far more biased on the electoral and popular vote. There was a filter bubble on tweets that were against Clinton and for Trump.

Stecanella (2016) conducted a sentiment analysis on MonkeyLearn from July 2016 until Election Day. Millions of tweets were processed to gauge each day throughout the time span. The author of this experiment was actually an engineer who created a social media tool that showed changing sentiment graphs for the timespan as mentioned. Results returned more negative sentiment overall for each candidate than positive sentiment on a daily basis. MonkeyLearn is said to have 70% accuracy for reading sentiment, whereas Sysomos is said to be 86% accurate when human accuracy normally falls between 70 and 85% (Stecanella, 2017; Bowers, 2017 para 2).

2.5: Problem with Polls

In the recent election, polls were not as accurate or reliable as they have been in the past. Shirani-Mehr, Rothschild, Goel, and Gelman (2016) decomposed the margin of error in surveys that were given during statewide elections. This study used polling data from more than 4,000 polls for 608 state presidential elections, political races for senators, and gubernatorial elections for nearly two decades (p. 3). Shirani-Mehr et al. (2016) attempted to calculate biases while explaining the margin of error. The findings of this study resulted in considerable election-level bias and superfluous variance. Shirani-Mehr et al. (2016) estimated a standard absolute bias is "1.8 percentage points for senate races, 2.1 percentage points for gubernatorial races , and 1.0 percentage point for presidential races" (p. 22). Polls in past presidential elections displayed small excess variance. However, results yielded a larger standard error of .08 percent in senatorial and gubernatorial races (Shirani-Mehr, 2016). Williams (2015) also agrees that national polling is in somewhat of a crisis. Williams quoted a University of Michigan political science professor who specializes in political polling of elections who stated that "polling is a very important element of democracy and polls give the public an independent voice that's not generally present" (2015).

2.6: Presentation of Study

Given the abovementioned literature, the present study seeks to assess the sentiment from Twitter messages regarding each candidate separately throughout the presidential election cycle ending on November 8, 2016. This study also evaluates if there is a relationship present between Twitter sentiment and FiveThirtyEight polls.

2.7: Research Questions

RQ 1: What are the trends in Twitter sentiment for President Trump and Secretary Clinton throughout the campaign?

RQ 2: How do the trends in sentiment for President Trump and Secretary Clinton compare to polls?

CHAPTER 3:

Methodology

A descriptive analysis was done to examine trends in Twitter sentiment for Trump and Clinton between July 1, 2016 and November 8, 2016 using Sysomos. A descriptive analysis was also done using the data collected from the election forecast from FiveThirtyEight. Sentiment results were compared to polls results using the Pearson product-moment correlation. Tweet volume was also considered within the data set.

To conclude this study by answering the central questions we must examine polls and Twitter from numerous angles. There were a few candidates running in the presidential election. This study is only concerned with the candidates from the Republican Party and Democrat Party, and uses polls as a comparison that subtract for the third-party candidate. The third party is excluded from the sentiment analysis statistics and adjusted for in the polls conducted by FiveThirtyEight.

3.1: Polls for Comparison

FiveThirtyEight uses several daily polls from all 50 states. An average is then calculated for a final daily percentage displaying the popular vote, Electoral College, and chance of winning for each candidate. FiveThirtyEight offers a poll-only forecast, which contains information from polls and does not factor in any other facets such as the economy or past elections. It also has a now-cast, which gives the outcome if the election were to occur on that particular day. Polls-only and now-cast are used for comparison in the current study. FiveThirtyEight includes polls that were rated through an intense rating system for accuracy and integrity and belong to the National

Council on Public Polls (NCPP) or the American Association for Public Opinion Research (AAPOR) (Silver, 2016). Polls are excluded and placed on a banned list if the manager of FiveThirtyEight believes they have used fake data in the past or participated in unethical conduct such as robocalls to cellphones without the inclusion of live interviewers (Silver, 2016). FiveThirtyEight takes several steps to assure the best accuracy possible. Firstly, FiveThirtyEight adjusts its results by accounting for five major effects that could alter accuracy if not factored. The five effects are likely voter adjustment, convention bounce adjustment, omitted third-party candidate adjustment, trend line adjustment, and house effects adjustment. Secondly, poll outcomes are combined with other data that measure and account for third-party voting, undecided voters, projection of popular vote, national vs state polls, partisan voting index (PVI), demographic regression, and blending polls with regression, and state elasticity scores (Silver, 2016). Lastly, they simulate the election, as uncertainty normally tends to decrease closer to Election Day. FiveThirtyEight also accounted for national error, state-based error, and finally, regional or demographic error for the 2016 election forecast (Silver, 2016).

3.2: Sysomos MAP

Sysomos was the analytic tool used in the current study to acquire data from Twitter. Sysomos is an analytic tool that performs machine analysis. Sysomos Map contains an exclusive contextual sentiment engine where the entire text becomes classified mechanically through machine learning-based algorithms (Sysomos). Sysomos (2017) states, "The sentiment engine has been trained on over 200,000+ human-tagged samples to understand and classify keywords as having negative, positive, neutral, or none" (para. 2). A four-step process is used to look for keywords, phrases, and language constructs associated with positive and negative meanings to determine sentiment. Sysomos (2017) claims, "The MAP sentiment engine has been

benchmarked at an accuracy rate of 85% (+/- 5%) however, it should be noted that assessing

sentiment is a difficult task for a machine" (para. 3). Firstly, words must go through a

qualification phase that filters the several languages that Sysomos is able to read. Secondly, an

extraction of keywords that have passed through the qualification phase are extracted according

to what Sysomos filters are set. Thirdly, Sysomos sends all inquiries through the POV (point of

view) verification, which is analyzing objectivity of the query requested, only sending subjective

mentions to phase four. Lastly, after the query passes the previous steps, the query is classified as

negative, neutral, or positive. Media Analysis Platform (MAP) was used to conduct a sentiment

analysis and record volume of tweet mentions for Trump and Clinton. Sysomos has the

capability to track archived data for a total of one year or in real-time and has access to 100

percent of all tweets (Twitter Firehose) within a search criteria (Ampofo, Simon, O'Loughlin,

Chadwick, Halfpenny & Proctor, 2015). Sysomos is able to filter data in several different ways

by demographics, country of origin, and state of origin if necessary. This study only analyzes

tweets that originate in the United States, filtering out every other location to understand how

trends changed for candidates through possible electorates.

3.3: Data Collection

Collection of data began on July 1, 2016 and concluded on November 8, 2016 using

Sysomos analytics. A total of 655,500 tweets were collected. Each calendar day was reported

between these dates. In total 131 days were reported in this study. The time filter was set to run

from 12:00 a.m. through 11:59 p.m. for every day included in the sample. To acquire data for

each candidate, altering the search queries according to opposing candidates was necessary. This

helped eliminate Twitter noise referring to family members or certain words that may be directly

associated to either candidate. Sysomos only allows for the collection of 5,000 tweets daily.

3.4: Volume of Mentions

To ensure the research obtained all tweets referring to Hillary Clinton, the keyword Hillary Clinton or Hillary or Clinton was used. The Clinton name is associated with family members, foundations, and numerous other buzzwords (keywords) as found in searches. Sysomos also allows researchers to filter out words that may not pertain to the candidate in question. To eliminate Tweets that combine sentiment for both candidates the opposing candidates name was removed from each search query with the expectation that the sentiment rating was solely about one candidate. For Clinton, the search query in Sysomos as follows ("Hillary"OR"Clinton")AND NOT"Bill"OR"foundation"OR"Chelsea"OR"Donald"OR"Trump"). Eliminating these words would allow for a distinct search for each candidate and offer a more concise sentiment analysis for the candidate in the data set. It was also important to filter results in order to receive the Sysomos analytics solely from the United States.

To ensure the research obtained all tweets referring to Donald Trump, the keyword Donald Trump or Trump was used. The Trump name is associated with family members, foundations, and numerous other buzzwords as found in searches. Filters in Sysomos were also used to eliminate as much noise as possible that related to the Trump name that had nothing to do with the race for presidency. To eliminate Tweets that combine sentiment for both candidates the opposing candidates name was removed from each search query with the expectation that the sentiment was just calculated for a one particular candidate. For Trump, the search query in Sysomos was as follows ("Donald"OR"Trump")AND NOT "Ivanka"OR"Melania"OR"Donald Jr."OR"Barron"OR"Tiffany"OR"Eric"OR"Tower"OR"Hillary"OR"Clinton"). Eliminating these words would offer a distinct search for Trump and offer a more concise sentiment analysis for

the candidate in the data set. Again, as stated above this data set also excluded any tweet not originating solely in the United States.

3.5: Sentiment Analysis

Sysomos analyzes everything in question such as word clouds, buzzwords, hashtags, volume of mentions, and sentiment analysis simultaneously. However, because this study was comparing trends in sentiment throughout the campaign to FiveThirtyEight polling concurrently, a sentiment analysis was done separately for each day to acquire the maximum amount of tweets. The daily maximum amount of tweets Sysomos allows for mining is 5,000. The same dates, days of the week, and search query was used for Trump and Clinton as named above.

CHAPTER 4:

Results

The overall purpose of this study was to investigate existing relationships and trends that

occurred during the presidential election cycle of 2016 by analyzing Twitter sentiment for the

two remaining candidates (Trump, Clinton) of the major political parties in the United States

along with polls that were conducted during this race. This study seeks to display helpful

information that candidates may consider in the future while on the campaign trail by

recognizing various milestones and trends that occurred during this particular race that could

possibly hinder or aid in the election process for candidates. These trends in sentiment were

recorded from Twitter and congruently from polling of FiveThirtyEight (2016). Limited research

has been done regarding trends in Twitter sentiment for presidential candidates (Trump, Clinton)

while comparing trends to polls.

4.1: Descriptive Statistics

This study used the 131 days leading to November 8, 2016, which was Election Day. As

seen in Table 1, each variable contained has been summarized. For both candidates the highest

mean and standard deviation was from the variable Twitter results which pertains to volume of

tweets from possible electorates regarding each candidate with Trump having (M=645,713.27,

SD=628,755.98) and Clinton having (M=379,903, SD=240,946.41). Figure 1 displays the trend

on Twitter volume for both candidates. Trump had a larger volume than Clinton. Neutral

sentiment as mentioned above is lacking negative or positive sentiment was also high for both

candidates however, is virtually unimportant considering the experiment is looking at trends that

would offer an indication as to which candidate would essentially win by looking at numerical parameters. The chance of each candidate winning if the election were held on one of the 131 days of the sample (n=131) according to FiveThirtyEight (2016) shows a trend that is very different for each candidate. Trump had a substantially lower chance of winning statistically as his average was *M=26.442, SD=10.91,* while Clinton had an average of *M=73.44, SD=11.05*. The intended popular vote by the electorate taken by FiveThirtyEight closed the gap some opposed to the chance of winning per day as Trump had *M=43.267, SD=1.143* and Clinton *M=49.9, SD=1.179*. Negative sentiment from Sysomos for both candidates also told a story through trends as seen if Figure 3, as in this race it seemed the candidate with the lowest negative sentiment eventually was victorious as Trump had a *M=11.769, SD=1.682* while Clinton had a *M=12.246, SD=2.119* . The lowest mean and standard deviation was calculated in the category of positive sentiment which was mined from Sysomos with Trump having *M=3.5, SD=.8025* and Clinton having *M=2.337, SD=.9681*. The trend for positive sentiment for each candidate shows in Figure 2.

Table 1 Descriptive Statistics			
	N	Mean	Std. Deviation
positive Trump	131	3.560	.8025
neutral trump	131	84.695	1.7614
negative trump	131	11.769	1.6822
Trump Intended popular Voting by Electorate	131	43.267	1.1431
trump chance of winning	131	26.442	10.9082
positive Clinton	131	2.337	.9681
neutral Clinton	131	85.442	2.3220
negative Clinton	131	12.246	2.1198
Clinton Intended popular Voting by Electorate	131	48.063	1.1797
Clinton chance of winning	131	73.444	11.0520
Twitter results Trump	131	645713.27	628755.981
Twitter Results Clinton	131	379903.51	240946.407
Valid N (listwise)	131		
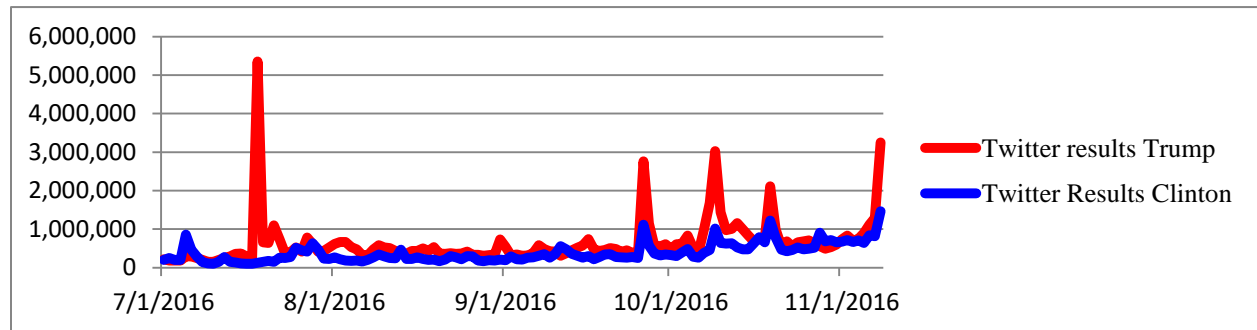
Figure 1: Twitter Results Trump vs Twitter Results Clinton
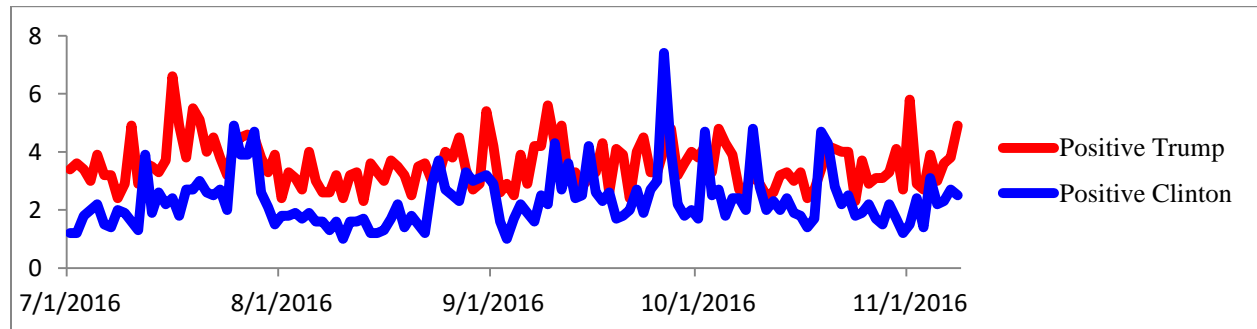
Figure 2: Positive Sentiment Trump vs Positive Sentiment Clinton

26
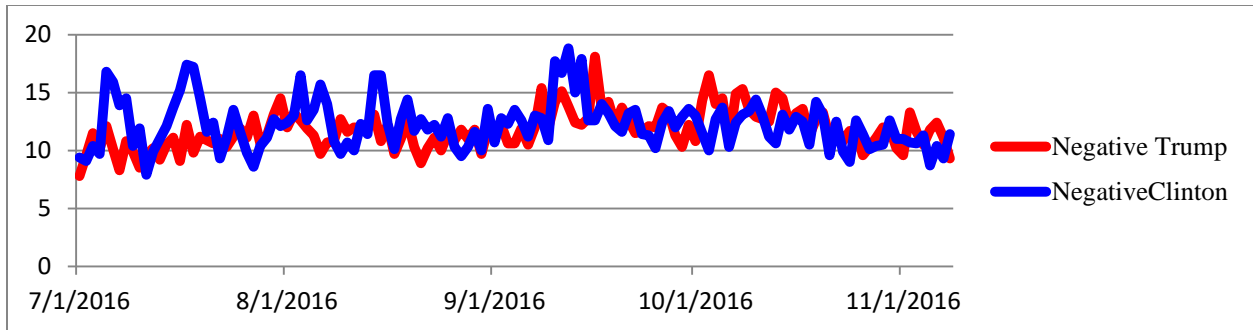
Figure 3: Negative Sentiment Trump vs Negative Sentiment Clinton

4.2: Inferential Statistics

For this particular study in order to identify trends through relationships between sentiment variables and variables from polls conducted by FiveThirtyEight (2016) (positive, neutral, negative, intended voting by electorate, chance of winning), a series of analyses were conducted using a Pearson product-moment correlation.

A Pearson product-moment correlation coefficient was computed to assess the relationship between six variables during this analyzation individually for each candidate. The six categories are as follows and can be seen in Table 2 and Table 3. Twitter results measured the volume of tweets given on each day for each candidate (Trump, Clinton). Positive (Trump, Clinton) is the measurement of positive sentiment reported by Sysomos during the sentiment analysis. Neutral (Trump, Clinton) is the measurement of sentiment in which no negative or positive sentiment was detected in Sysomos. Negative (Trump, Clinton) is the measurement of negative sentiment detected in the Sysomos sentiment analysis tool. The next two categories were from the polling of FiveThirtyEight. The first one is listed as Intended popular vote by electorate, meaning that the candidate would essentially win the popular vote from the electorate. The final category is titled chance of winning. The chance of winning variable accounts for the

candidate's (Trump, Clinton) actual chance of winning the election if it were to be held on that particular day. All of the Pearson product-moment correlations were 2-tailed.

4.2.1: Chance of Winning Election

Beginning with Trump, a two-tailed Pearson correlation showed various significant relationships. Strong significant relationships were reported with Trump chance of winning and Trump intended popular voting by electorate. The positive correlation between these variables indicated that when intended voting by electorate increased, so did his chance of winning, r = (131) = .788, p < .001. Clinton also had a strong positive significant relationship, as expected, also indicating the chance of winning increased as the intended voting by the electorate increased, r (131) = .924, p < .001. Trump also had a positive correlation relating to positive Twitter sentiment. The more positive Twitter sentiment from the electorates became, the more his chance of winning increased showing a moderate relationship between the two variables r(131) = .348, p < .001 which can be seen in Figure 6 . This trend did not continue with Clinton. Clinton had a moderately significant negative relationship between the variable chance of winning and positive Twitter sentiment. As her positive Twitter sentiment increased, her chance of winning the election decreased r (131) =-.340, p< .001 which can be seen in figure 8.

4.2.2: Intended Vote by Electorate

The intended vote by electorates showed relationships for both candidates with Trump Clinton. Trump had a relationship that was weak while Clinton's correlations showed a moderate relationship. The positive correlation between intended vote by electorates raised when his positive sentiment was calculated in Sysomos raised r (131) = .226, p < .001 and can be seen in Figure 7. This trend was quite different for Clinton. When the intended electorate showing in

28

FiveThirtyEight (2016) polls increased, the positive sentiment for Clinton decreased, creating a negative correlation relating to positive Twitter sentiment r(131) = -.350, p < .001 which is shown in Figure 9.

Relationships were also shown for both candidates between intended votes by electorates Twitter results, which included volumes of mentions. Trump had a positive correlation with Twitter results and the same trend occurred for Clinton. When the intended votes by electorates increased during polls from FiveThirtyEight, so did Twitter results. This happened to be a weak relationship as r (131) = .244, p < .001. This trend also repeated for Clinton. The positive correlation between Clinton intended voters by electorates and Twitter results indicated that the higher the volumes of mentions, the better chance there was that Clinton would win votes by the intended electorate r(131) = .246, p < .001.

4.2.3: Negative Sentiment

Trends for negative sentiment in Twitter were only weakly related for Trump to Twitter volume. As the positive correlation between negative sentiment and volumes of tweets indicated the more negative sentiment in tweets occurred the more mention he was getting in volume r (131) =.134, p < .001. Clinton had a negative non-significant relationship between negative sentiment and Twitter results (volume of tweets).

4.2.4: Positive Sentiment

Positive sentiment for Trump had a positive non-significant relationship with Twitter results. The positive correlation for Clinton was much more significant than Trump, but only moderately. The correlation between positive sentiment for Clinton and Twitter results specified

that the more positive the sentiment, the higher the Twitter Volume raised r (131) = .330, p <

.001.

Correlations

Table 2 Correlations Trump		Twitter results Trump	Positive Trump	Neutral trump	Negative trump	Trump intended vote by electorate	Trump chance of winning
Twitter results Trump	Pearson Correlation	1	.069	-.162	.134	.244**	.051
	Sig. (2-tailed)		.433	.065	.126	.005	.559
	N	131	131	131	131	131	131
Positive Trump	Pearson Correlation	.069	1	-.393**	-.045	.226**	.348**
	Sig. (2-tailed)	.433		.000	.611	.009	.000
	N	131	131	131	131	131	131
Neutral Trump	Pearson Correlation	-.162	-.393**	1	-.882**	-.328**	-.284**
	Sig. (2-tailed)	.065	.000		.000	.000	.001
	N	131	131	131	131	131	131
Negative trump	Pearson Correlation	.134	-.045	-.882**	1	.230**	.135
	Sig. (2-tailed)	.126	.611	.000		.008	.125
	N	131	131	131	131	131	131
Trump intended vote by electorate	Pearson Correlation	.244**	.226**	-.328**	.230**	1	.788**
	Sig. (2-tailed)	.005	.009	.000	.008		.000
	N	131	131	131	131	131	131
Trump chance of winning	Pearson Correlation	.051	.348**	-.284**	.135	.788**	1
	Sig. (2-tailed)	.559	.000	.001	.125	.000	
	N	131	131	131	131	131	131

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3 Correlations Clinton		Twitter Results Clinton	Positive Clinton	Neutral Clinton	Negative Clinton	Clinton intended vote by electorate	Clinton chance of winning
Twitter Results Clinton	Pearson Correlation	1	.330**	-.068	-.081	.246**	.078
	Sig. (2-tailed)		.000	.439	.356	.005	.378
	N	131	131	131	131	131	131
Positive Clinton	Pearson Correlation	.330**	1	-.446**	.032	-.350**	-.340**
	Sig. (2-tailed)	.000		.000	.715	.000	.000
	N	131	131	131	131	131	131
Neutral Clinton	Pearson Correlation	-.068	-.446**	1	-.902**	.292**	.163
	Sig. (2-tailed)	.439	.000		.000	.001	.063
	N	131	131	131	131	131	131
Negative Clinton	Pearson Correlation	-.081	.032	-.902**	1	-.175*	-.031
	Sig. (2-tailed)	.356	.715	.000		.046	.721
	N	131	131	131	131	131	131
Clinton intended vote by electorate	Pearson Correlation	.246**	-.350**	.292**	-.175*	1	.924**
	Sig. (2-tailed)	.005	.000	.001	.046		.000
	N	131	131	131	131	131	131
Clinton chance of winning	Pearson Correlation	.078	-.340**	.163	-.031	.924**	1
	Sig. (2-tailed)	.378	.000	.063	.721	.000	
	N	131	131	131	131	131	131

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

CHAPTER 5:

Discussion

In this study, a general examination of sentiment was done regarding presidential candidates by investigating social media. Tweets were analyzed to elicit a more in-depth understanding of associations between Twitter sentiment and national polls. Correlations were made displaying relationships to sentiment and polls from FiveThirtyEight (2016) for both candidates. The intended electorate has demonstrated that Twitter is becoming more commonplace as a viable way to discuss candidates, which created trends in sentiment for each political candidate throughout the final 131 days of the 2016 election cycle.

5.1: Findings

Changing trends for each candidate offer insights about sentiment during each candidate's presidential campaign and transpire in the results of this study when compared to national polls. First, it was found that the most significant relationships or correlations between sentiment and polls appear when there is a positive or negative sentiment present for each candidate. Secondly, moderate relationships were established between positive/negative sentiment and intended vote by electorate in different ways for each candidate. The more Trump's positive sentiment grew, the better his intended vote became from the electorate increasing his chance of winning the election. However, Clinton's results were different. The relationship became negative according with positive sentiment when relating this to intended vote, which also decreased her chance of winning the election. Trump's correlations make sense.

Clinton's negative relationships requires speculation as to why when positive sentiment was increasing, her chance of winning and intended vote by electorate was decreasing.

Table Four demonstrates the highest and lowest positive and negative sentiment, chance of winning, and intended vote by electorate for each candidate. There is a possibility that this could be attributed to certain events or situations that occurred during the 2016 election cycle. These events could reveal why trends of sentiment and polls rose or plummeted. These events were not scientifically related to sentiment but ensued on dates when sentiment was at the maximum and minimum juncture for each participating candidate. Scandal for Trump and Clinton continued during the entire campaign process.

Discussing a few dates from Table Four should allow for some clarity. Positive sentiment for Trump was lowest at the end of October 2016, and highest in mid July 2016. At the end of October, Jessica Drake, Trump's eleventh accuser of sexual misconduct, came forward, which may demonstrate why his positive sentiment was at an all-time low (Kenny, 2016). Trump's highest positive sentiment occurred in mid-July when he announced Mike Pence ad his Vice President during the RNC (Brander, Bush, & Lee, 2016). Trump's negative sentiment was lowest at the beginning of July 2016 even though accusations of sexual assault continued surfacing from an individual who said Trump assaulted her when she was thirteen. His highest negative sentiment occurred in mid-September 2016 as he called inner cities crime-ridden and jobless. In addition, Donald Trump Jr. referenced that they would be warming up the gas chamber if Clinton were a Republican in regards to her email scandal, which were noted in the media as being anti-Semitic (REPUBLICINSANITY, 2016).

Clinton suffered similar setbacks and highlights for positive and negative sentiment that also offers insights as to why her trends continued changing. Positive sentiment for Clinton was lowest in Mid-August of 2016 and highest at the end of September of 2016. Clinton's scandals differed from Trump's however, were just as damaging. The DNC email was hacked. A group called Anonymous published several emails daily. Early August when Clinton's positive sentiment was lowest emails were released exhibiting ties with the State Department workers and the Clinton Foundation suggesting that" loyal supporters of Clinton should be found a position in Washington" (Fain, 2016). Emails continued to be released during the duration of her campaign with some being more damaging than others are. Clinton's highest positive sentiment came directly after the first presidential debate as she was deemed the winner by several news outlets (Fain, 2016). The lowest negative sentiment for Clinton came in mid-July when the director of the FBI announced that it found no wrongdoing by Clinton in the investigation into her having confidential emails on her personal email account (Hartig, Lapinski & Psyllos, 2016). The highest negative sentiment occurred for Clinton in mid-September as she stumbled on the campaign trail raising questions about her health and transparency (Collinson, 2016).

Also by looking at Table Four, it is easily seen that the minimum and maximum dates for negative and positive sentiment occurred within a two-week range of each candidates highest and lowest chance of winning and intended voting by electorate. When comparing Trump and Clinton's positive and negative sentiment at a glance, it is easily seen that the candidates had similar lows and highs. However, when observing variables from national polls (intended vote by electorate, chance of winning) these figures offer a different outcome.

At the height of Clinton's positive sentiment, she was still unable to overcome her personal and political issues in the swing states and otherwise traditionally Democratic favoring

states on Election Day. One state after another turned red on election night that had remained

blue for decades from electoral votes in the past election cycles. Trump's followers, also named

the silent majority, somehow overlooked his personal issues and thus his positive sentiment had

a positive correlation to his chance of winning and intended vote by electorate. This may be one

reason he was victorious. Clinton had negative relationships for positive sentiment relating to

her chance of winning and intended vote by electorate, even though she secured the popular vote.

There is not a clear or definitive explanation for this occurrence. One can only speculate as to

why these relationships occur when they seem to defy logic. Many believe that she could not

build the enthusiasm to bring people to the polls in her favor in swing states even though many

spoke positively about her on Twitter. Pew researchers Mercer, Deane, and McGeeney (2016)

assess, "Because many traditional likely-voter models incorporate measures of enthusiasm into

their calculus, 2016's distinctly unenthused electorate – at least on the Democratic side – may

have also wreaked some havoc with this aspect of measurement" (para, 8).

Another speculation from experts surrounding Clinton is the fact that many voters were

angry after the Democratic Party allegedly railroaded Bernie Sanders. Voters liked that Sanders

as well as Trump was different from the last four presidents. Even though Sanders has been a

politician for many years, his stance on policies aligned more with Trump than Clinton.

Following party lines normally suggest that whoever secures the nomination for a particular

party, those belonging to that party would vote for that individual in the general election even if

it were not their first choice. Pilkington and Chalabi (2016) assessed during interviews of 700

Sanders followers that 500 of them would make a Sanders-Trump switch. Even though this

sample is not huge it could offer evidence that a trend may occur nationwide that those who

planned to vote for Sanders in the general election had he won the primaries, would not follow

the endorsement he gave for Clinton on Election day. Pilkington and Chalabi (2016) interviewed electorates and stated, "They explained their unconventional position by expressing a variety of passionately held views on their shared commitment for protecting workers and against new wars, on their zeal for an alternative to the establishment, and on their desire to support anyone but Hillary Clinton (para, 3). The electorates in this sample felt that Trump would achieve this better than Clinton would.

Because this study relied heavily on Twitter sentiment, it is necessary to look at Twitter to cogitate why Clinton's results occurred in the manner they did. This third and final speculation may shed a light on how a heightened positive Sentiment on Twitter would correlate to a declining chance of winning and intended vote by electorate for Clinton in a negative way. Every Twitter user has probably encountered a Twitter Bot at one time or another. Leon (2017) a journalist from *The Washington Beacon* referenced an analyst site named TwitterAudit. TwitterAudit has the capability to audit whether Clinton's followers are real or fake on Twitter. 9,086,280 of Clinton's 17.9 million followers are fake, while 8,729,955 are real accounts which displays that over half of her followers were fake (Leon, 2017, para 2). This may explain the reasons why as her positive Twitter sentiment rose, her chance of winning or intended vote by electorate did not. Twitter bots can be programmed to tweet, retweet, and spread fake news, which also denotes that a positive Twitter sentiment can come from bots that are unable to vote in elections.

Table 4 Important dates that contributed to Trends				
	min	date	max	date
positive Trump	2.3	10/24/16	6.6	7/16/16
negative trump	7.8	7/1/16	18.1	9/16/16
Trump Intended popular Voting by Electorate	40.8	8/17/16	45.5	7/30/16 11/4/16
trump chance of winning	10.8	8/14/16	50.1	7/30/16
positive Clinton	1.0	8/10/16 9/3/16	7.4	9/26/16
negative Clinton	7.9	7/11/16	18.8	9/12/16
Clinton Intended popular Voting by Electorate	46.1	7/26/16 7/27/16	49.9	8/8/16
Clinton chance of winning	48.7	7/22/16	89.2	8/14/16
Twitter results Trump	147,834	7/9/16	5,351,444	7/18/16
Twitter Results Clinton	96,160	7/10/16	1,459,740	11/8/16

5.2 Linking Related Work to Current Study

During the last decade, the concentration on sentiment from Twitter has rapidly grown. This may be attributed to an increase of interest in personal opinions on various topics that users turn to Twitter to divulge. Aforementioned studies used Twitter sentiment for predicting political outcomes, stock market, and feelings about climate change. In the German election of 2009, Tumasjan (2011) associated Twitter volume served somewhat as a predictor to the winner of the election. This study although not looking for predictive factors however, found relationships between Twitter sentiment and polls just as this study revealed. Despite the awareness that these

relationships are not causal, they are able to shed some light on the possibilities of studying

Twitter sentiment for upcoming events. In the political realm, these studies continue to grow as

Wang (2012) established through his work the development of a system that performs real-time

sentiment of the entire presidential election, which was recorded on an interface that tracked

dominating keywords that were deemed positive or negative via a nave Bayes modeling system.

This study was similar in findings when relationships were established between Twitter

sentiment and tweet volume in the final days of the election cycle. Correlating polls to Twitter

sentiment just as this study has done has benefited politicians in the area of measuring public

opinion (O'Connor 2010).

CHAPTER 6:

Conclusion

6.1 Research Contributions and Practical Implications

This particular study makes several contributions to research. First, it is demonstrated that Twitter is being used by intended electorates for political communication during election cycles as shown in other previous studies by way of tweeting opinions on policy and political candidates. Secondly, this study offers an extension of current literature by probing how positive and negative sentiment adds to a candidate's chance of winning an election and the possibility as to how the intended electorate plans to vote. Furthermore, this study addresses the relationships that Twitter sentiment has on existing polls. Given the fact that social media has become relevant s an arena for political communication, the necessity for politicians to use social media is growing with each election cycle.

6.2 Limitations and Future Research

Many researchers are gauging for sentiment in different paradigms and are attempting to formulate a theory for a strong theoretical foundation. This study and many other similarly lack a strong theoretical foundation to test research questions and hypotheses. The limitation of this study also derives from the analysis of data restricted to Twitter. Twitter is not fully representative of the electorate. Firstly, not all the electorate or Twitter users tweet about politics. Secondly, the population on Twitter tends to be a younger generation therefore the sentiment does not represent the views of the older voting population as stated above in connection with a

Pew research study that offers Twitter demographics. This raises issues of generalizability. Future research may extend sentiment studies to other social media outlets and include Twitter to represent a broader electorate. Another limitation derives from using tweets because of the short length of text offered. This is another reason future research should include other social media sites or blogs when gauging public sentiment for political candidates. Lastly, limitations of this study come from using Sysomos MAP as a method for collecting. Sysomos cannot filter for sarcasm and restricts the number of tweets a researcher can obtain daily from Sysomos. Future research might include gathering tweets with and without the use of electronic mining and use human coders to code for sentiment along with using Sysomos even though Sysomos claims accuracy in upwards to 85%.

The predominant goal of this study was to gather a thorough concept of trends in sentiment on Twitter for the two remaining candidates (Trump, Clinton) during the 2016 presidential election cycle and then to compare these trends to national polls. Twitter sentiment is a viable avenue to use for researchers when gauging public opinion. Twitter sentiment has the unique capability of indicating public opinion and sentiment for political candidates when used congruently with polling and is practical for estimating which candidate is likely to win the upcoming elections. The current study aids in helping both researchers and politicians to better understand the political discourse and the function of sentiment in information dispersion on Twitter. Correlating polls with sentiment can provide important practical uses for politicians while also furthering goals in research.

References

Ampofo, L., Collister, S., O'Loughlin, B., Chadwick, A., Halfpenny, P., & Proctor, R. (2015). Text mining and social media: When quantitative meets qualitative and software meets people. Innovations in Digital Research Methods. Thousand Oaks: Sage Publications *Inc*., 161-192.

Anuta, D., Churchin, J., & Luo, J. (2017). Election bias: Comparing polls and Twitter in the 2016 US election. ArXiv Preprint arXiv:1701.06232,

Bialik, C. (2016). How the Republican field dwindled from 17 to Donald Trump . Retrieved from https://fivethirtyeight.com/features/how-the-republican-field-dwindled-from-17-to-donald-trump/

Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Icwsm*, 11, 450-453.

Bowers, J. (2017). Sysomos review. Retrieved from http://www.toptenreviews.com/services/internet/best-social-media-monitoring/sysomos-review/

Broersma, M., & Graham, T. (2012). Social media as beat: Tweets as a news source during the 2010 British and Dutch elections. Journalism Practice, 6(3), 403-419.

Burch, P. (2015). For elections, is social data more predictive than traditional polls?. *Sysosmos* Retrieved from https://sysomos.com/2015/12/23/for-elections-is-social-data-more-predictive-than-traditional-polls/

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate Change sentiment on Twitter: An unsolicited public opinion poll. *Plos One*. 10(8), 1-18. doi:10.1371/journal.pone.013Bump 6092.

Collinson. S. (2016). Hillary Clinton celebrates victory. Retrieved from http://www.cnn.com/2016/06/07/politics/primary-results-highlights/

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. *Information, Communication & Society*, 16(5), 795-825.

Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1195-1198.

Flynn, Kerry. (2016). Twitter Announces Its Debate Hashtag -- and a New Emoji. *Mashable*, Mashable, 22 Sept. 2016, mashable.com/2016/09/22/twitter-debates-hashtag/.

Hartig, H., Lapinski, J. & Psyllos, S. (2016). Hillary Clinton's lead over trump shrinks after controversial week: Poll. Retrieved from https://www.nbcnews.com/storyline/data-points/hillary-clinton-s-lead-over-trump-shrinks-after-controversial-week-n607351

Houshmand, S., Rothschild, D., Goel, S. and Gelman, A. (2016) Disentangling Bias and Variance in Election Polls, Unpublished Manuscript. https://5harad.com/papers/polling-errors.pdf.

Leon, J. (2017) Clinton, Sanders Have More Fake Than Real Twitter Followers. *The Washington Beacon*. Retrieved from http://freebeacon.com/politics/clinton-sanders-more-fake-than-real-twitter-followers/

Mercer, A.,Deane,C., & McGeeney, K. (2016). Why 2016 Election Polls Missed Their Mark. *Pew Research Center*, 9 Nov. 2016, www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/.

Mitchell, A., & Hitlin, P. (2013). Twitter reaction to events often at odds with overall public opinion. *Pew Research Center*, 4.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Icwsm*, 11(122-129), 1-2.

Pennacchiotti, M., & Popescu, A. (2011). Democrats, Republicans and Starbucks Aficionados: User classification in Twitter. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 430-438.

Pilkington, E., Chalabi, M. (2016). The Bernie Sanders voters who would choose Trump over

    Clinton. *The Guardian*. Retrieved from https://www.theguardian.com/us-

    news/2016/mar/13/bernie-sanders-supporters-consider-donald-trump-no-hillary-clinton

REPUBLICINSANITY. (2017). Donald Trump- Full timeline part three: The final run to

    election day 2016. Retrieved from

    https://www.dailykos.com/stories/2017/8/11/1688845/-Donald-Trump-Full-Timeline-

    Part-Three-The-Final-Run-to-Election-Day-2016.

Rosenstiel, T., Jurkowitz, M. (2011). Twitter and the campaign how the discussion on Twitter

    varies from blogs and news coverage and Ron Paul's Twitter triumph. *Pew Reseach

    Center*, 1-55.

Rosenstiel, T., Sonderman, J., Loker, K., Ivancin, M. & Kjarval, N. (2015).

    How people use Twitter in general. American Press

    Institute. Retrieved from

    https://www.americanpressinstitute.org/publications/reports/survey-research/how-people-

    use-twitter-in-general/

Stecanella, B. (2016). Donald Trump vs Hillary Clinton: Sentiment analysis on Twitter mentions.

    MonkeyLearn. Retrieved from https://blog.monkeylearn.com/donald-trump-vs-hillary-

    clinton-sentiment-analysis-twitter-mentions/

Stieglitz, S.Political Communication and Influence through Microblogging--An Empirical

    Analysis of Sentiment in Twitter Messages and Retweet Behavior. (2012). *2012 45th

    Hawaii International Conference on System Sciences, System Science (HICSS), 2012

    45th Hawaii International Conference on*, 3500. doi:10.1109/HICSS.2012.476

Sysomos. (2017). Frequently asked questions: Sentiment analysis. Retrieved from

    https://support.sysomos.com/customer/en/portal/articles/1182323-sentiment-

    analysis?b_id=11647

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with

      Twitter: How 140 characters reflect the political landscape. *Social Science Computer*

      Review, 29(4), 402-418.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time

      Twitter sentiment analysis of 2012 U.S. presidential election cycle. Proceedings of the

      ACL 2012 System Demonstrations, 115-120.

Willams, J. (2015). The problem with polls. *Us News World Report* Retrieved from

      https://www.usnews.com/news/the-

      report/articles/2015/09/28/why-public-opinion-polls-are-increasingly-inaccurate

Appendix: A

Figure: 4 Chance of Winning FiveThirtyEight

Figure 5: Intended Vote by Electorate FiveThirtyEight
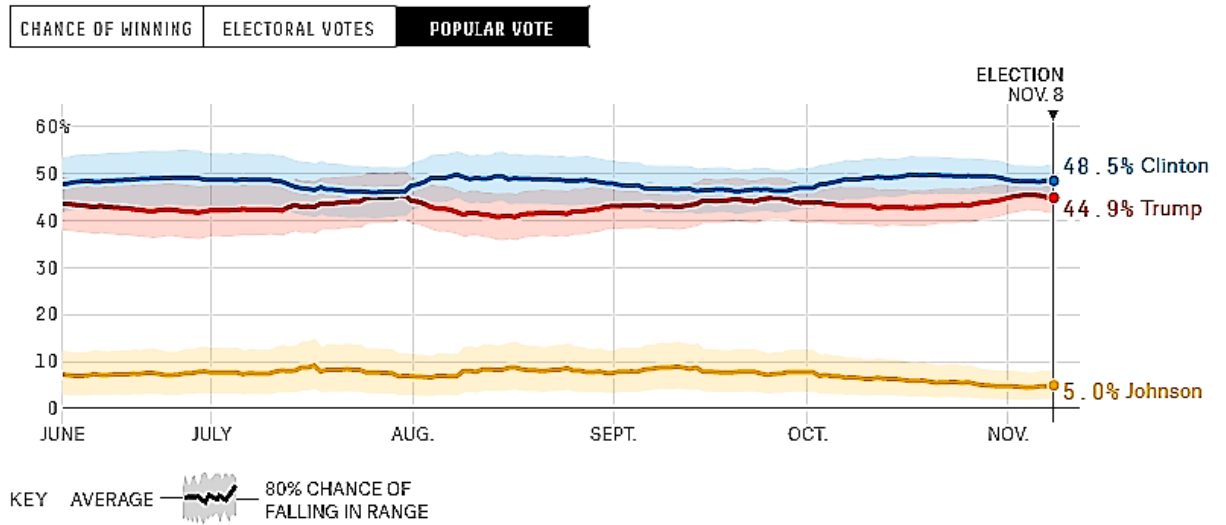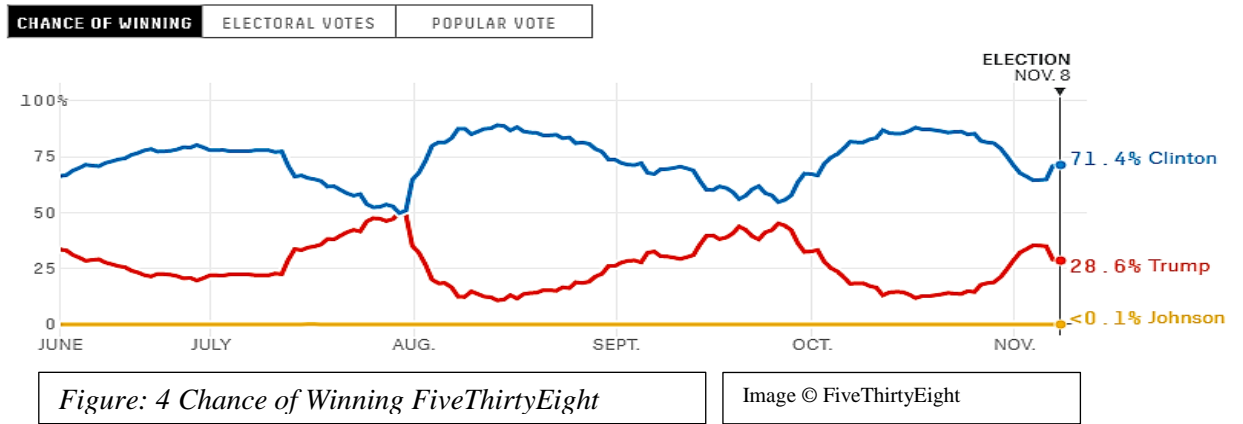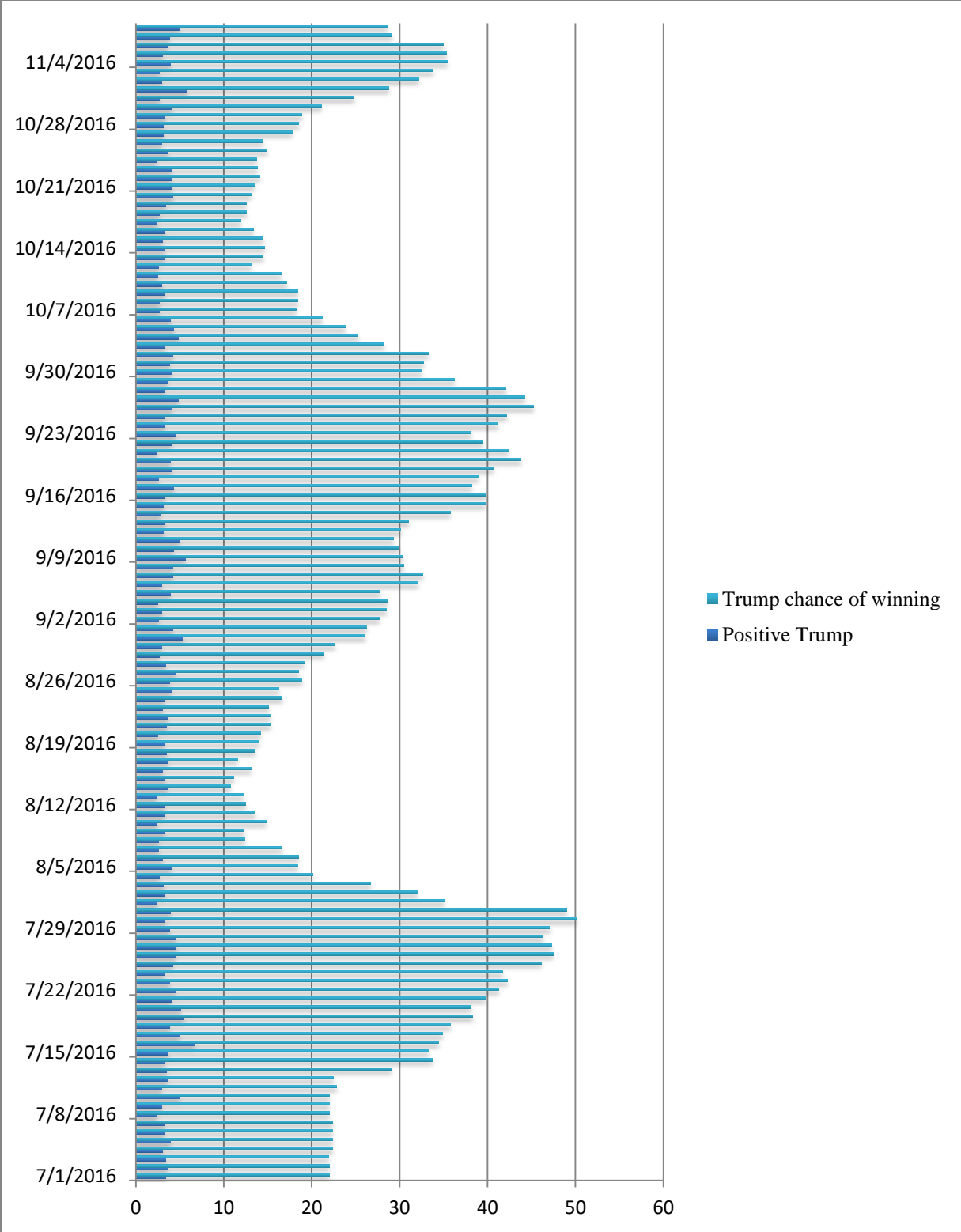
Figure 6: Trump Chance of Winning and Positive Sentiment
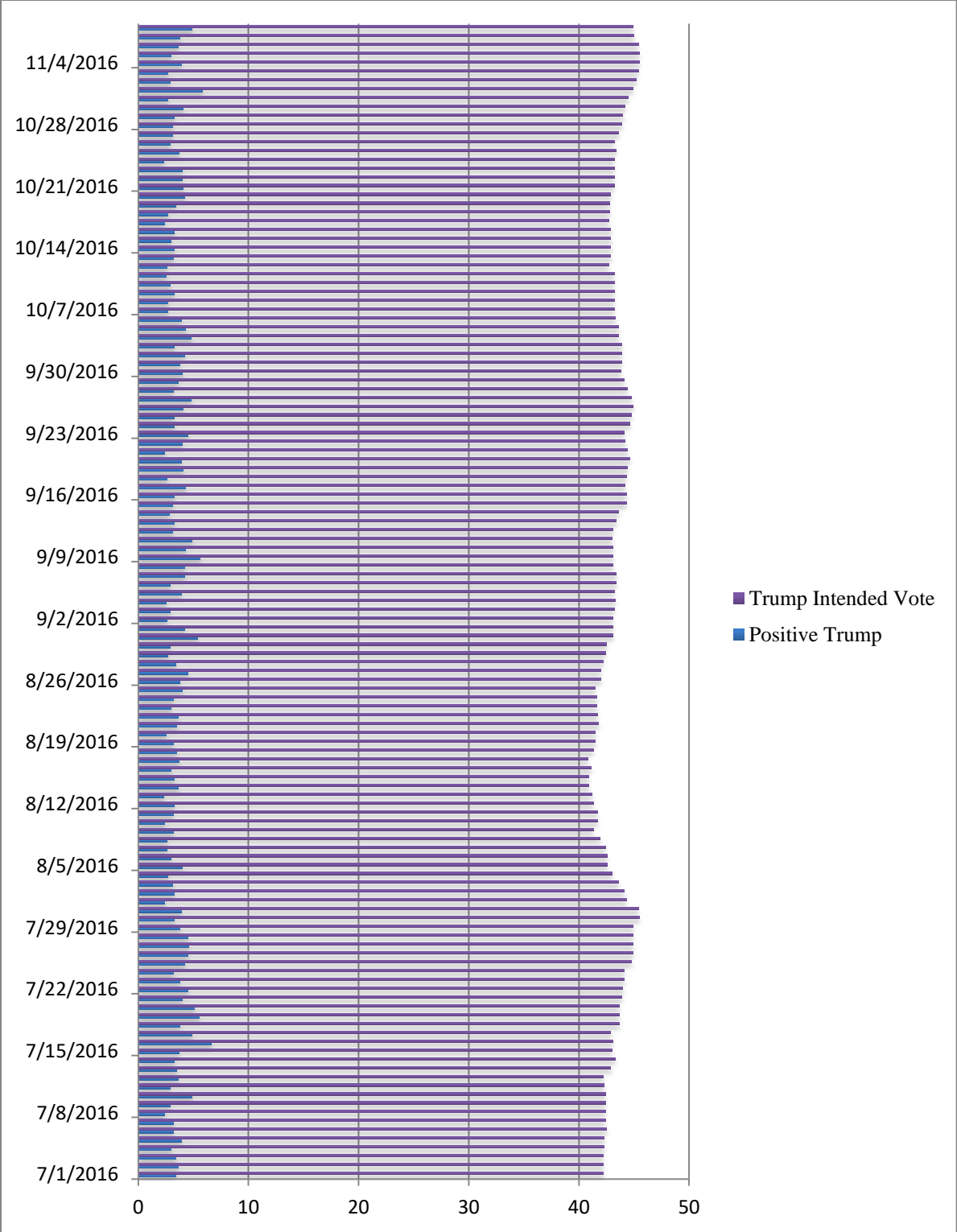
Figure 7: *Trump Intended Vote by Electorate and Positive Sentiment*
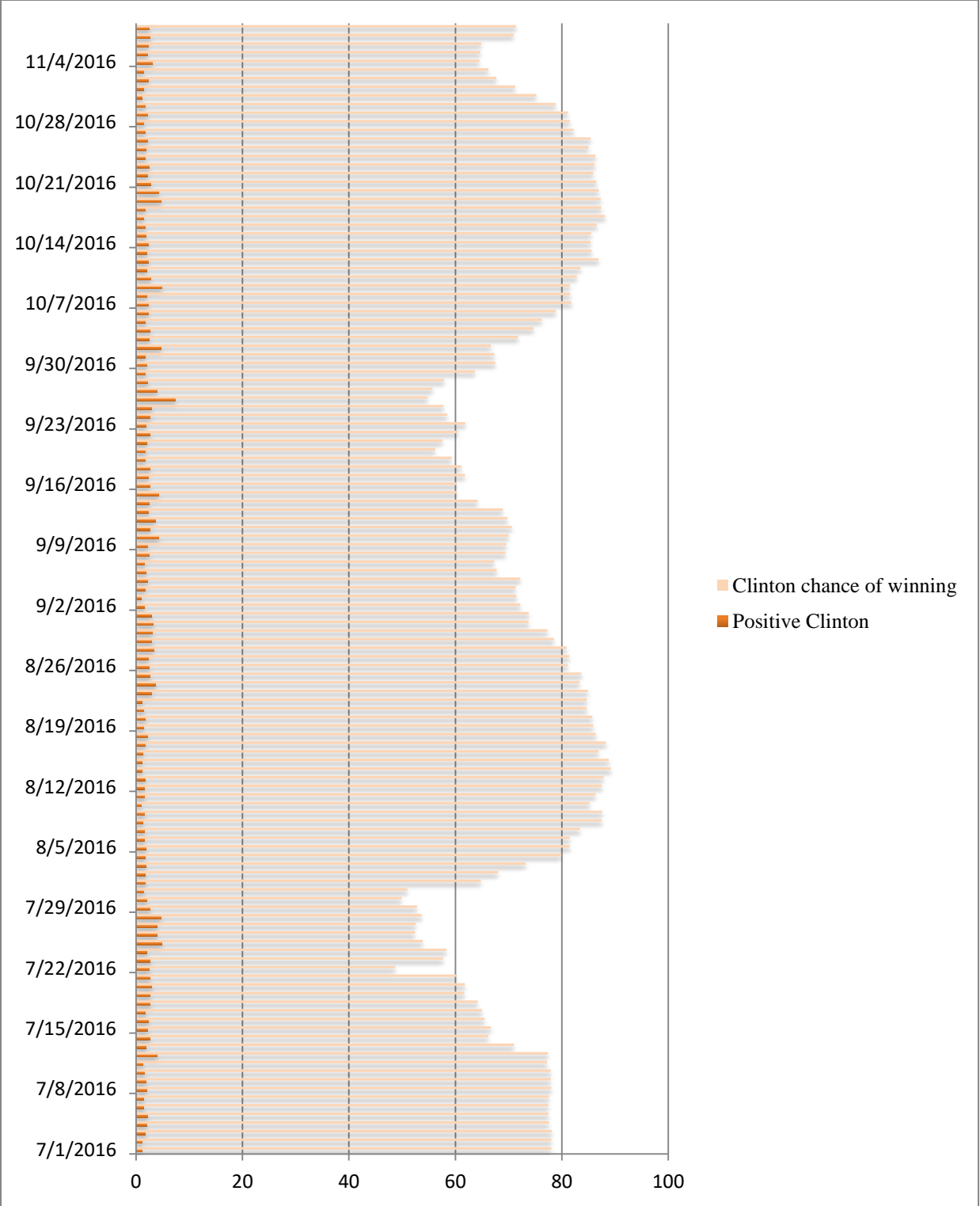
Figure 8: Clinton Chance of Winning and Positive Sentiment
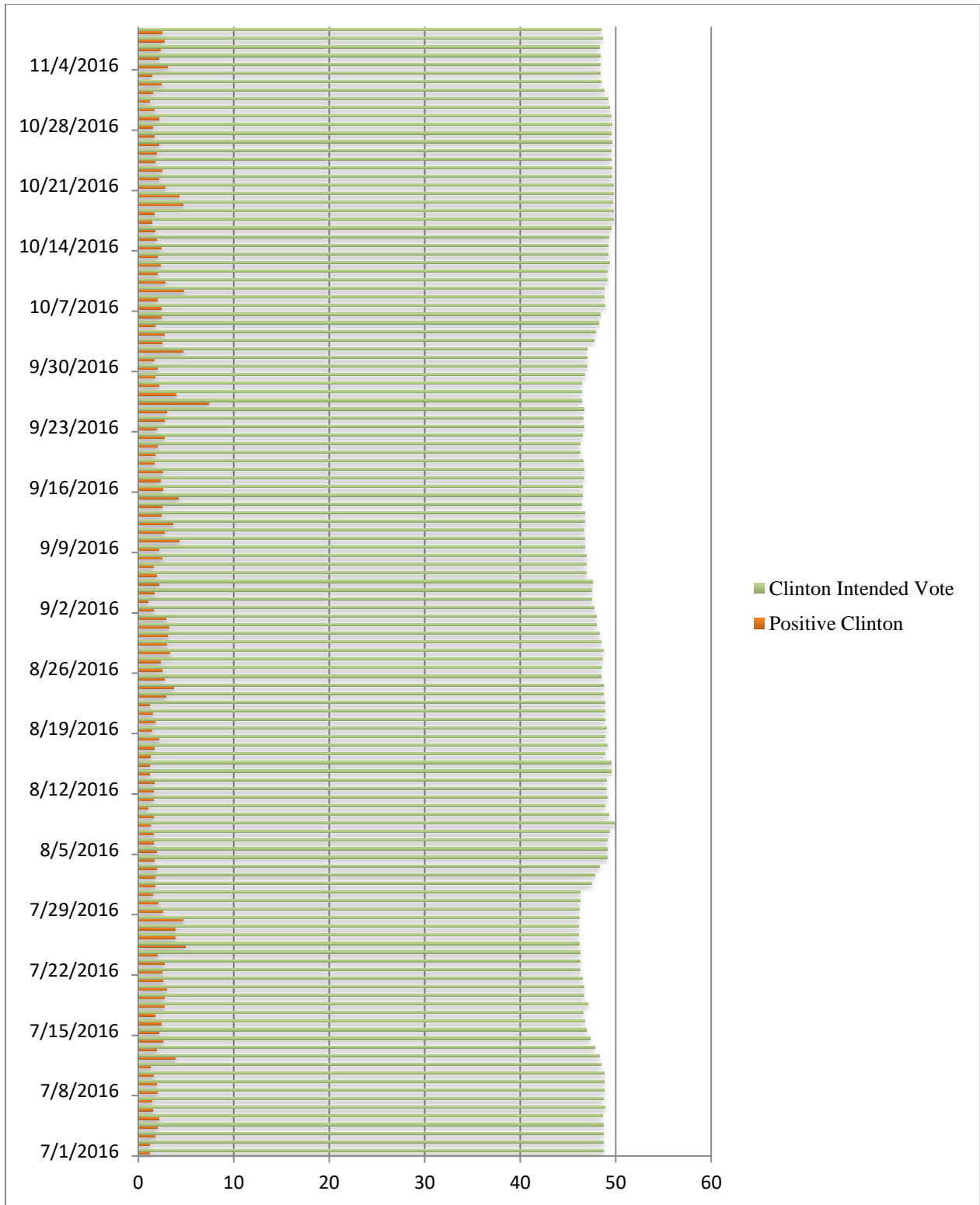
Figure 9: *Clinton Intended Vote by Electorate and Positive Sentiment*

Figure 10: Trump Positive and Negative Tweets and Sentiment

Figure 11: Clinton Positive and Negative Tweets and Sentiment