Presidential Speeches and Job Performance:

Don’t be Neutral

John Kellogg

CUNY School of Professional Studies

12/10/2021

**Abstract**

The words of the United States President often shape or demonstrate the path of the office holder. It is those words we will study in this paper. This project has two questions: 1) Can we determine if the words of the president effect the people of the nation? 2) If we can detect the effect of the words, can we track any useable trends and potentially make predictions for future presidents, speechwriters, or researchers?

The methodology for addressing these questions was applying sentiment analysis procedures across all official speeches given by US presidents from Harry Truman to Barack Obama. The speeches were matched up with Gallup Job approval poll numbers and analyzed for trends. A few very useful trends were found and expanded on; however, no prediction model could be created due to the specificity of the main trend.

**Introduction**

The institution of the American Presidency is a globally respected office. The United States does not have a single presidency, it has an institution with discrete administrations. As leaders of the country, they are the US people’s voice both to Congress and to the rest of the world. Their words have power, they have meaning, they have purpose. Sometimes the word of the president is required to calm a frightened nation, sometimes to drive the county’s purpose toward a goal, and sometimes to simply deliver information on recent events and the president’s mindset in dealing with the event.

**Statement of purpose**

Can we use the latest analysis techniques to judge how the United States population sees the effectiveness of the president by what the president says? If someone can track the reception of a speech and any trends, the future speeches can be geared to fill in any gap. The language of the next speech can be altered to effectively push the US citizenry toward approval and endorsement of the officeholder.

**Donald J Trump**

At the time of this study, the country is still heavily divided by partisanship centered around the tenure of President Donald Trump. We felt it would be difficult to keep the biasness out of the project, both in ourselves and the data. Gallup, Inc does an exceptional job trying to keep biasness out of their polls. However, it was questioned weather the poll questionees already had their mind set on President Trump’s ability to perform the job before the poll was started and any question asked. As such, the last president included in this study is President Barack Obama.

**Literature Review**

Sentiment Analysis (SA), through techniques like Natural Language Processing (NLP), is now a widely used data technique. Companies and individuals reply on SA for powerful insights and apply affective subjective information into their growing text-based data. SA is often applied to add context to the customer’s voice during items such as reviews, surveys, and social media posts.

NLP techniques are now being applied to live communications. In areas like health care (Siwicki, 2021), NLP mines text in documents and transcripts of doctor’s notes to identify facts, relationships and assertions which would not have been seen before. The extracted data can be further analyzed for trends and gain the impact of social, economic, and environmental factors leading to the individual health and quality of life.

In another example, companies and states are running the transcripts of legal cases through NLP engines to gain close to the same insights as in the medical example. Parole hearing transcripts (Hong & Voss, 2021) can be searched to answer key factors such as when the last disciplinary infraction was and what was the cause. These models can speed up and raise the accuracy of important legal proceedings.

The question we are discussing fits between those two models, can we run SA and NLP over the transcript of a formal written speech by the United States President and gain useful information. Next, can that information be used to track changes in their job performance as seen through the Gallop polls? Is it only specific ‘Defining Presidential Moments’ which can correlate between the speech and the job approval numbers? Are there other factors which can have associations or links to approval number changes?

Schubert, Stewart, and Curran (Schubert, Stewart, & Curran, 2002) created a study in 2002 to judge the effects and rallies of nationalism after a major event driven by presidential speeches, specifically the effects of George W. Bush in the days after the 11 September attack. Their results proved a dramatic jump in approval after his speech to congress the day of the attack. Later, in this analysis, we will see if our data backs up these results.

Putting presidential speeches through SA models was also performed by John Paul Miranda and Rex Bringula in 2020 (Miranda & Bringula, 2021) and by Renalyn Banguis-Bantawig in 2018 (Banguis-Bantawig, 2019). Miranda and Bringula used the State of Nation Address (SONA) of the past 13 Philippine presidents the team analyzed the general sentiment of the addresses during specific periods of Philippine governments. Their findings showed the positive and negative sentiments followed the general state of the period. For instance, the speeches with the lowest sentiment scores were during the martial law periods around 1974. Banguis-Bantawig used 54 speeches of selected Asian Presidents to analyze their methods of reaching across the podium to sway the opinions of the audience. The results of both studies prove the importance of a presidential “speech” as a venue to discuss and demonstrate the nation’s state and direction.

Matthew Eshbaugh-Soha (Eshbaugh-Soha, The Impact of Presidential Speeches on the Bureaucracy, 2008) states “...*presidential speeches* are an effective means of influence over bureaucratic activity, and that this influence is contingent on the direction of the president’s policy signals” (p. 129). Positive signals in the speech reinforce the mission and motivate bureaucrats to do more. In contrast, negative signals have little to no impact on the same mission or motivation. He does not go into the effect of the third aspect of SA, the neutral signal. We will need to delve into its effects.

**Model**

Sentiment Analysis works by applying a score to specific words. The scores are complied to give us general direction of the text.

In this paper we will use the VADER (for Valence Aware Dictionary for Sentiment Reasoning) (Hutto & Glibert, 2014) through TextBlob. While the VADER model is specifically geared to social media text, we chose this method as a speech, while written, is still created to be spoken. Social Media is also another form of same process, it’s written but tends to the less formal aspects and structures of literature.

TextBlob (Loria, 2020) gives us 5 key values per text set: Positive, Negative, Neutral, Polarity, and Subjectivity. Positive, Neutral and Negative scores are the percentage ratio of words in the text which fall into those 3 buckets; as an example, a text can have 20% Positive, 77% Neutral, and 3% Negative words leading us to summarize the text as more positive than negative. Since we are analyzing speech, we will always have a neutral value as it would be very difficult to effectively communicate while only using positive and negative words. We can potentially see the effects of a high ratio of neutral words on the sentiment and the job scores. Polarity is a single value score which lies between [-1,1] where a -1 indicates a highly negative sentiment and a 1 indicates a highly positive sentiment. Negation words reverse the polarity. Subjectivity is a single value score which lies between [0,1]. The closer the number is to zero the more subjective the text. In reverse, the closer to 1, the more objective the text is.

**Data and Data Preparation**

**Gallup Job Approval Numbers**

Gallup, INC is an American analytics and advisory company which operates from Washington D.C. Founded in 1935, the company is best known for its public opinion polls conducted worldwide and has become one of the accepted sources of party neutral US politics public opinion. Since 1935, the company has performed the same style of opinion poll to generate the US public’s opinion on how well the President is doing in his job, the Presidential Job Approval Ratings. For each president, the poll shows the percentage from 100 of those who *Approved*, *Disapproved*, or were *Unsure* of the President’s ability to perform the job and/or how well they thought the president was doing in the job.

Gallup was chosen because they have a clearly defined published polling process ( Gallup, Inc., 2010) which has measures to reduce any bias. Each survey, yhe company randomly polls 1,000 US national adults across the United States with a margin of error of +/- 4%. In the early days of the poll, it was conducted in person. The poll was then manually complied and published. In the mid/late 1980s, the poll became conducted over the phone. The time to perform and compile the in-person polls will be come a factor for us.

To make use of this data, we had to find how much the data changed over time. A percentage change formula was applied for each of those columns to capture the change in sentiment from the previous poll. The formula took a rolling median of the current value and the next two values. *([Original value]+[Next Value 1]+[Next Value 2]).Median)* The rolling median was then applied to a percentage change formula *([Rolling\_Value] - [Value])/[Rolling\_Value] \*100)*. The method allows us to see how much change, if any, occurred between the polls by percentage.

Through the University of California, Santa Barbara, we received the combined datasets for all presidents Truman through Trump (Peters, 2021).

**Presidential Speech Text**

The Miller Center at the University of Virginia has a mission to preserve the information for prosperity the American institution of the Presidency. Their lofty and wonderful goal is to help guide the future of American democracy by generating scholarship using a unique nonpartisan tradition. One of their missions is to capture and record the transcript of every written speech from President George Washington to our current President Joseph Biden.

The text of the speeches was scraped from the Miller Center webpage (UVA Miller Center, 2021) along with the date the speech was performed. We ran each transcript through the model, described earlier, to produce the sentiment numbers for each speech. With the sentiment numbers in hand, we matched the Gallup poll numbers.

The dataset covers all speeches from a president even those given before and after being in office. As the poll numbers are only generated during their time in office, we centralized the dataset to focus the speeches to the same timeframes per president.

**Roadblocks**

As stated earlier, prior to 1980’s the poll was conducted in-person and at the person’s home. As such, the poll, was not run on a regular basis. While we have almost weekly poll numbers for Presidents George W Bush and Barrack Obama, earlier Presidents, such as Dwight Eisenhower, we have large gaps between the poll dates. There was no clean method to programmatically match the poll numbers and their changes for each speech; a manual matching process was followed. For sections of this research, we needed to find a reproducible method which allowed for the poll participants to listen to the speech before being asked their opinions. Where possible, we followed two methods which allowed for the speech to have time to be heard and a poll organized to capture the public mood:

* A speech was matched up with the corresponding poll numbers of the week following the speech (+/- 10 days).
* If the poll started the day of the speech, we felt there was ample time in these cases.

**Analysis**

**General observations**

Except for Dwight D. Eisenhower and Lyndon B. Johnson, most presidents gave a median of 22 speeches during each term; Eisenhower gave 6 speeches and Johnson gave 49. Those who had two full terms, gave more speeches during their first term over their second.

Table 1

Presidential Overview

|  |  |  |  |
| --- | --- | --- | --- |
| President | Entered office | Left Office | Number of Speeches |
| Harry S. Truman | 4/12/1945 | 1/20/1953 | 19 |
| Dwight D. Eisenhower | 1/20/1953 | 1/20/1961 | 6 |
| John F. Kennedy | 1/20/1961 | 11/22/1963 | 37 |
| Lyndon B. Johnson (Partial Term) | 11/22/1963 | 1/20/1965 | 20 |
| Lyndon B. Johnson (Term 2) | 1/20/1965 | 1/20/1969 | 49 |
| Richard M. Nixon | 1/20/1969 | 8/9/1974 | 22 |
| Gerald Ford | 8/9/1974 | 1/20/1977 | 14 |
| Jimmy Carter | 1/20/1977 | 1/20/1981 | 19 |
| Ronald Reagan (term 1) | 1/20/1981 | 1/20/1989 | 33 |
| Ronald Reagan (term 2) | 1/20/1985 | 1/20/1989 | 22 |
| George H. W. Bush | 1/20/1989 | 1/20/1993 | 21 |
| Bill Clinton (Term 1) | 1/20/1993 | 1/20/1997 | 25 |
| Bill Clinton (Term 2) | 1/20/1997 | 1/20/2001 | 12 |
| George W. Bush (Term 1) | 1/20/2001 | 1/20/2005 | 21 |
| George W. Bush (Term 2) | 1/20/2005 | 1/20/2009 | 18 |
| Barack Obama (Term 1) | 1/20/2009 | 1/20/2013 | 29 |
| Barack Obama (Term 2) | 1/20/2013 | 1/20/2017 | 17 |

**Sentiment numbers**

Table 2

Median Sentiment Numbers

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| President | Polarity | Subjectivity | Negative | Neutral | Positive |
| Harry S. Truman | 0.154 | 0.472 | 12.63% | 58.19% | 29.19% |
| Dwight D. Eisenhower | 0.134 | 0.463 | 11.88% | 55.43% | 32.68% |
| John F. Kennedy | 0.131 | 0.466 | 11.63% | 60.60% | 27.79% |
| Lyndon B. Johnson | 0.139 | 0.437 | 11.29% | 63.77% | 24.93% |
| Richard M. Nixon | 0.144 | 0.450 | 12.26% | 60.55% | 27.20% |
| Gerald Ford | 0.130 | 0.418 | 9.59% | 62.29% | 28.12% |
| Jimmy Carter | 0.121 | 0.435 | 10.57% | 60.85% | 28.60% |
| Ronald Reagan | 0.110 | 0.435 | 12.15% | 61.92% | 25.94% |
| George H. W. Bush | 0.112 | 0.438 | 11.88% | 61.15% | 26.98% |
| Bill Clinton | 0.120 | 0.438 | 11.94% | 62.29% | 25.75% |
| George W. Bush | 0.127 | 0.450 | 15.09% | 57.50% | 27.40% |
| Barack Obama | 0.105 | 0.438 | 12.14% | 63.35% | 24.50% |

To recap, Polarity is a single value score which lies between [-1,1] where a -1 indicates a highly negative sentiment and a 1 indicates a highly positive sentiment. The polarity score does not reflect political polarization or the swaying of a speech toward or from a binary political ideology. The text engine is completely apolitical. Subjectivity is a single value score which lies between [0,1]. The closer the number is to zero the more subjective the text. In reverse, the closer to 1, the more objective the text is. Positive, Neutral, and Negative is the ratio from 100 of the word counts from each category.

It comes as no surprise the polarity scores tends toward zero and the neutral scores are the highest. The President may be the leader of a particular party he is still the single person representing all people of the US. Keeping your speech neutral reduces the risk of offending either political side.

It is surprising to see polarity, subjectivity, positivity, and negativity have generally dropped over time. In response, neutral has risen, apart from Lyndon Johnson who had the lowest Positive and highest neutral. We do not see the return of high numbers of neutral and low numbers of positivity until President Obama.

Johnson’s presidency was fraught with civil unrest due to the civil rights movement, the assassination of President Kennedy, and the ramp up of tensions in Vietnam. It makes sense in that light to have calming speeches and not swaying too far into positive or negative sentiments. For Obama, his presidency was also touched with heavy tensions from foreign issues. However, I suspect the growing trend of partisanship to have more effects on the reasons for the sentiment numbers and for the same reasons as Johnson. He may have wanted to calm the nation.

**Gallup Data**

Figure 1

Chart, line chart

Description automatically generated

The Gallup data does not show a definitive cyclical trend but does show major events. The major movement around the middle of figure 1 (Red circle) starts on September 11, 2001. The data backs up the study discussed earlier by Schubert, Stewart, and Curran (Schubert, Stewart, & Curran, 2002) where they studied the dramatic jump in approval after his speech to congress the day of the attack. We also see, as Schubert, Stewart, and Curran state in their study, the effect does not last long. By December 12, 2001, the disapproval numbers start rising and do not come back down during President Bush’s tenure.

Other items of note in recent history are:

* The major movement in the last 3rd of figure 1 (Light Blue circle) in both the approval and disapproval charts happens in the days after President Obama’s inauguration.
* On the middle chart which represents the *Unsure* data, the major spike in the middle of the data (Blue Circle) is on January 24, 1989; 4 days after the inauguration of President George H.W. Bush.
* The *Unsure* spike just behind the January 24th spike happens on January 30, 1981 (Green Circle) which coincides with President Ronald Reagan’s first press conference given the day earlier.

Looking at the *Positive* and *Negative* numbers, we see they mostly mirror each other. The *Unsure* data stays mostly constant. While it is a given, a change in the *Approve* or *Disapprove* numbers influences the opposite statistic, since the Unsure is so constant, any change in the *Unsure* polling number has a major effect on the two other stats, not necessarily on both at the same time.

**Comparing Presidents**

President Carter and Obama make for good comparison of the sentiment analysis concepts. President Carter is as close as we can get to the median of all five metrics of SA. President Obama is on the extreme end of the Polarity, Neutral and Positive numbers.

Table 3

Comparison chart of specific presidents and key events

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Polarity | Subjectivity | Negative | Neutral | Positive |
| Median of all Speeches | 0.128 | 0.438 | 0.119 | 0.610 | 0.273 |
| Median of Jimmy Carter | 0.121 | 0.435 | 0.106 | 0.608 | 0.286 |
| Median of Barack Obama | 0.105 | 0.438 | 0.121 | 0.634 | 0.245 |
| Obama's Inaugural Address | 0.056 | 0.467 | 0.149 | 0.574 | 0.277 |
| W. Bush's address on 9/11 | 0.058 | 0.506 | 0.256 | 0.428 | 0.316 |
| Reagan's first Press Conference | 0.123 | 0.410 | 0.084 | 0.734 | 0.182 |
| H.W. Bush's Inaugural Address | 0.223 | 0.540 | 0.068 | 0.555 | 0.377 |

In the sections below, we will take a closer look at the speeches themselves and see the change in Job Approval numbers. Each speech, when available, is paired up to the change in the Gallup poll numbers (Approve, Disapprove, Unsure) for each of the five key speech indicators. We should start to understand which of the five key indicators has potential effects in poll number trends.

**Jimmy Carter**

Jimmy Carter was the 39th President of the United States and held office from January 20, 1977, through January 20, 1981. President Carter employed Hendrik Hertzberg as his chief speechwriter, who wrote most of the speeches over the President’s four years in office. Hertzberg is credited with writing the worst presidential speech of all time (Dale, 2009); the “Malaise” speech on July 15, 1979. Both Carter and Hertzberg wanted to present an educated but solemn presence from the White House. What they received was openly criticized even by former speechwriters as the “Passionless Presidency” (Fallows, 1979). Thus, he makes the perfect central line to start the analysis.

Figure 2 Chart, line chart

Description automatically generated

President Carter started off well in the polls but progressively got worse as he was accused of mishandling multiple international issues. He was also the president during the worst energy crisis the country had seen until that point. Fallows says when the presidential honeymoon ended the country saw the contrast between the promise and popularity and the disappointment felt later (Fallows, 1979). His breakdown follows the trend seen earlier; Approving and Disapproving mirror each other.

Figure 3

Chart, bar chart

Description automatically generated

Figure 4

Chart, line chart

Description automatically generated

The *Polarity* and *Subjectivity* across all the speeches are mostly flat. The Negative sentiment for all speeches is the same. Except for the ‘Panama Canal’ speech in September 1977, the *Polarity* and *Subjectivity* are never more than +/- 0.1 point from each other. The ‘Diplomatic Relations with China’ speech in December 1978 is the only other speech where the negative sentiment moves significently. The Panama Canal Speech in 1977 generates a high Polarity score, marking it as one of the most postive speeches of his term. In contrast, the ‘Diplomatic Relations with China’ speech in 1978 also has a very low negative sentiment but the Polarity is offset by the high amount of Neutral sentiment. Even the famed “Malaise” Speech (Energy and the National Goals - A Crisis of Confidence, July 15th 1979) does not move the numbers.

Looking at the change in poll numbers compared to the *Polarity* and *Subjectivity* numbers we can see heavy outliers but do see movement. We do know the Unsure metric is magnitudes lower than the other two statistics. However, we should not count out movement here as those moving out of unsure will end up in one of the other two buckets.

Figure 5

Chart, line chart

Description automatically generated

Figure 6

Chart

Description automatically generated

In both metrics, we see as the *Polarity* and *Subjectivity* rise, we see a downward trend in the Unsure and Disapproval numbers and a rise in Approval numbers. It indicates, for Carter, as he became more Positive the amount of change in Disapproval and Unsure went down. As he became more Objective, the amount of people changing their Disapproval rating went down, more people approved and chose a side (unsure change).

Using that same methodology, we can compare the other three key indicators; Positive, Neutral and Negative sentiment in the words of the speech themselves.

Figure 7

A picture containing text, sky, bunch, day

Description automatically generated

Figure 8

Chart, line chart, scatter chart

Description automatically generated

Figure 9Chart, line chart, scatter chart

Description automatically generated

While we still have heavy outliers on both sides of the graphs, they are not weighted heavily enough to shift the trends of any of the key indicators. There is not enough correlation between Positive, Neutral or Negative sentiment and a change in the polls.

Let’s see how the extreme of the five key indicators looks in comparison to these numbers.

**Barack Obama**

Barack Obama entered office as the 44th president 28 years after Jimmy Carter on January 20, 2009 and served two terms until 2017. The president brought into the White House Jonathan Favreau who was his Director of Speechwriting while the future president was a Senator from Illinois. Favreau freely admits he draws influence (Walker, 2013) from Robert Kennedy, Michael Gerson [George W. Bush’s speechwriter] and Peggy Noonan [Ronald Reagan]. Favreau’s speeches and Obama’s oratory skills delivered famous memorable moments. How do these speeches look when compared to Jimmy Carter?

Figure 10

Chart, line chart

Description automatically generated

Barrack Obama’s breakdown follows the same trend we’ve seen so far, the Approving and Disapproving numbers mirror each other. We do see a lot less unsure poll entries overall from President Carter. The major difference between the two is all three categories in President Obama’s numbers stay relativity straight.

Figure 11

Chart, bar chart

Description automatically generated

Figure 12

Chart, line chart

Description automatically generated

Again, we see Polarity and Subjectivity numbers with in +/- 0.1. The lowest polarity speech is on September 10, 2013 (red outlines) where the president addressed the nation concerning the growing situation in Syria. I can understand why that speech would want to be as negative as possible, however, it does not have a corresponding upwards spike in Subjectivity. In fact, it has the opposite and has the 3rd lowest value in Subjectivity. The speech is more subjective than objective while still being low in polarity.

Figure 13

A picture containing text, sky, different

Description automatically generated

Polarity is where we start to get a glimpse of the eventual trend. As Polarity gets larger, the rate of change in the Unsure metric rises. The more polarizing, the more people are change their minds. The trend in the unsure metric follows in almost all the other five key indicators.

Figure 14

Timeline

Description automatically generated with medium confidence

President Carter had more change in the Subjectivity trend than President Obama. We will need to investigate the three other SA characteristics and potentially other presidents to ensure any trends.

Figure 15

A picture containing text, sky, bunch

Description automatically generated

Figure 16

A picture containing text, sky

Description automatically generated

Figure 17

A picture containing text, sky, day

Description automatically generated

As we look at the other three indicators, we can clearly see the trend in the Unsure metric. As the indicator reaches the top of the scale, there is less movement in the Unsure change rate. As a speech has higher ratio of negative, the rate of unsure change drops. As a speech has more positive the rate of change rises, however slightly.

Note, in appendix A, we have produced the all the graphs above for all presidents in this study.

**Correlation**

Correlation or more specifically the correlation coefficient is a measure which determines the strength of the relationship between two variables. When the coefficient is greater than zero (>0) and up to one (1) it indicates a positive relationship. As one number goes up the corresponding number also goes up. When the coefficient is less than zero (<0) and down to negative one (-1) it indicates a negative relationship. When one number goes up the corresponding number goes down.

For the purposes of cleanliness, the columns were renamed following the bullets below:

* APP = Approving
* DIS = Disapproving
* UNS = Unsure/NoData
* APPC = Approve\_change
* DISC = Disapprove\_change
* UNSC = Unsure\_change
* POL = Polarity
* SUB = Subjectivity
* NEG = Negative
* NEU = Neutral
* POS = Positive

**Corpus – All paired speeches**

Figure 18

Chart

Description automatically generated with medium confidence

Creating a correlation matrix across all the speeches which have matching poll numbers, nothing stands out between the five key speech indicators and the Poll numbers. In fact, it is so near zero on all the indicators, it makes us suspicious. We have already seen cases where, for instance, Polarity affects the Unsure Change statistic.

**Presidential Breakdown**

Figure 19

Chart

Description automatically generated

Carter’s correlation plot is what we expected to see from the corpus. There are a few items to highlight:

* Disapproving has a large negative correlation to Neutral
* Approving has the exact opposite correlation to Neutral
* Positive Sentiment has some correlation with Approval and Disapproval
* None of those have significant correlation to their sister change over time number.

Figure 20

Chart

Description automatically generated

We do not see the same effects on the Neutral statistic in the Obama correlation plot. In the Carter Correlation plot, we saw a significant downward trend in Unsure Change to Negative. For the key speech indicators, there is no statistically significant correlation.

If we look at the correlation plots next to the Approval statistic, we start to see an interesting correlation trend.

Figures 21-22

Chart

Description automatically generatedChart, line chart

Description automatically generated

Figures 22-23

Chart

Description automatically generatedChart

Description automatically generated

From President Truman forward to present day, when a president has a sustained downward approval rating of over 20 percentage points (such as in the example above with Richard Nixon) their Neutral coefficient has a significant correlation to their Approval and Disapproval coefficient. Most cases, but not all, the same conditions create a significant correlation between the Approval and Disapproval coefficients and Positive sentiment.

For Presidents, such as President Obama where the Approval percentage stayed relatively constant, there is no correlation between the Gallup poll generated coefficients and the speech coefficients. All the correlation numbers stay between -0.3 to 0.3.

The effect holds partially true for President Eisenhower who did see a drop in his approval numbers but not over 20 percentage points. His plot only shows a significant correlation between Disapprove and Neutral coefficients. With the required matching process between speech and Gallup poll date, only 4 of his speeches made it through. There is not enough data for his speeches to affect the stated trend.

We are sure there are a lot of factors surrounding each of the approval numbers for each president. Each one is a different story. However, the data shows if the approval numbers start to drop, there is significant correlation between either Positive and/or Neutral speech and the Approval/Disapproval coefficients.

**Conclusions and next steps**

**Overview**

This project started with two questions:

1) Can we determine if the words of the president effect the people of the nation?

2) If we can detect the effect of the words, can we track any useable trends for future presidents, speechwriters, or researchers?

For the first round of analysis, we investigated 409 speeches of twelve United States Presidents for their sentiment. Using TextBlob, we generated the value of Polarity and Subjectivity, also gathering the ratio of Positive, Neutral, and Negative words.

In the second round, we analyzed the job approval polls from Gallup Inc. for spikes and dips. Taking the dates of those events, we found the corresponding speech. In most cases, we found major events or speeches on or near those dates.

Next, we reduced the available pool of speeches to those which closely matched up with dates of Approval polls. The reduction left us with 275 speeches. The remaining speeches were then analyzed for trends using correlation tables and plots and scatterplots.

First, in some sense, Polarity seems to have effects on the rate of people who are unsure. In most cases, higher polarity equals higher values of unsure. This effect does not show up in the correlation plots for any of the presidents and may be an item of future research.

More importantly, we can determine there is some correlation between Neutral type words and the Approval/Disapproval numbers under specific conditions. When the poll statistic of Approving is continuously dropping, there is a high correlation with the use of Neutral words. As Approval goes down, the rate of Neutral words rises and vice versa. In most cases, the effect is also on the Disapproving statistic but in reverse, as Disapproval rises so does Neutral words.

**Final Thoughts**

After this research, if I was to find myself talking with the speechwriter for any public figure, I would advise them to use more strength words. Positive and Negative words have return on investment. If the speech is full of what I refer to as zero sum words (Neutral), the reception of the speech may not be what you expect.

**Areas of additional research**

* Investigate potential correlation between Polarity and the change in Unsure
* Create a specific sentiment library geared to public speakers, does the new engine change this research?
* Create clinical scenarios to have the SA statistics before a speech is given and monitor the polls immediately after.
* Determine if the Neutral/Approving ratio is still highly correlated in other scenarios such as when the Approval number has a large curve analysis could be:
  + From the peak of Approval to the base of the valley and from the base of the valley to the next peak

# Bibliography

Gallup, Inc. (2010). *How Are Polls Conducted?* Retrieved from Gallup.com: https://www.gallup.com/file/poll/125927/How%20Are%20Polls%20Conducted%20FINAL.pdf

Banguis-Bantawig, R. (2019). The role of discourse markers in the speeches of selected Asian Presidents. *Heliyon*, e01298.

Bermingham, A., & Smeaton, A. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. *SAAIP* (pp. 2-10). Chiang Mai: IJCNLP.

Ceron, A., Curini, L., & Lacus, S. (2015). Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence Fromthe United States and Italy. *Social Science Computer Review, 33*(1).

Dilin Liu, L. L. (2018). *The appeal to political sentiment: An analysis of Donald Trump’s and Hillary Clinton’s speech themes and discourse strategies in the 2016 US presidential election.* Amsterdam: Elsevier.

Eshbaugh-Soha, M. (2008). The Impact of Presidential Speeches on the Bureaucracy. *SOCIAL SCIENCE QUARTERLY, Volume 89, Number 1,*, 116-132.

Eshbaugh-Soha, M. (2010). The Politics of Presidential Speeches. *Congress & the Presidency, 37*, 1-21.

Haselmayer, M., & Jenny, M. (2014). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *EPSA Confrence.* Edinburgh: Springer Link.

Hong, J., & Voss, C. (2021, 05 03). *HAI Stanford University*. Retrieved from Is Natural Language Processing Ready to Take on Legal Hearings?: https://hai.stanford.edu/news/natural-language-processing-ready-take-legal-hearings

Hutto, C., & Glibert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.* Palo Alto, California: Association for the Advancement of Artificial Intelligence.

Lauderdale, B., & Herzog, A. (2016). Measuring Political Positions from Legislative Speech. *Political Analysis, 24*(3), 374-394.

Loria, S. (2020, 04 26). *TextBlob*. Retrieved from TextBlob: Simplified Text Processing: https://textblob.readthedocs.io/en/dev/

Lowe, W. (2008). Understanding Wordscores. *Political Analysis Vol 16*, pp. 356-371.

Miranda, J., & Bringula, R. (2021). Exploring Philippine Presidents’ speeches: A sentiment analysis and topic modeling approach. *Cogent Social Sciences*, 7: 1932030.

Peters, G. (2021). Presidential Job Approval [Data Set]. *American Presidency Project.* UC Santa Barbara. Retrieved from https://www.presidency.ucsb.edu/statistics/data/presidential-job-approval

Schubert, J., Stewart, P., & Curran, M. (2002). A Defining Presidential Moment: 9/11 and the Rally Effect. *Political Psychology , Sep., 2002, Vol. 23, No. 3, Special Issue: 9/11 and Its Aftermath: Perspectives from Political Psychology*, 559-583.

Siwicki, B. (2021, 02 19). *Using natural language processing to unlock SDOH in unstructured EHR data*. Retrieved from https://www.healthcareitnews.com/: https://www.healthcareitnews.com/news/using-natural-language-processing-unlock-sdoh-unstructured-ehr-data

UVA Miller Center. (2021). Presidental Speeches [Data Set]. *Famous Presidental Speeches*. UVA Miller Center.

Code Index

Python

import nltk

import nltk.corpus

from nltk.corpus import stopwords

nltk.download("stopwords")

nltk.download('punkt')

stop\_words = set(stopwords.words('english'))

from nltk.tokenize import word\_tokenize

from nltk.corpus import PlaintextCorpusReader

import random

random.seed(42)

from sklearn.pipeline import Pipeline

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import TfidfTransformer

from sklearn.naive\_bayes import MultinomialNB

from sklearn.linear\_model import LogisticRegression

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.svm import SVC

from sklearn.metrics import classification\_report, confusion\_matrix

import matplotlib.pyplot as plt

%matplotlib inline

plt.style.use('ggplot')

import seaborn as sns

import plotly.offline as py

py.init\_notebook\_mode(connected=True)

import plotly.graph\_objs as go

import plotly.tools as tls

import plotly.express as px

import plotly.io as pio

from plotly.subplots import make\_subplots

pio.renderers.default='browser'

import pandas as pd

import numpy as np

import datetime as dt

import string

#pd.set\_option("display.max\_rows",500)

import warnings

warnings.filterwarnings('ignore')

#from nltk.sentiment import SentimentIntensityAnalyzer as sia

from textblob import TextBlob

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

analyzer = SentimentIntensityAnalyzer()

import requests

import io

all\_speechs\_data = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/presidential\_speeches.csv"

obama\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/obama.csv"

bush43\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/bush43.csv"

bush41\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/bush41.csv"

clinton\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/clinton.csv"

reagan\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/reagan.csv"

carter\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/carter.csv"

eisenhower\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/eisenhower.csv"

ford\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/ford.csv"

johnson\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/johnson.csv"

kennedy\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/kennedy.csv"

nixon\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/nixon.csv"

truman\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/truman.csv"

trump\_rating = "https://raw.githubusercontent.com/kelloggjohnd/Final\_project/main/data/approval\_ratings/trump.csv"

speech\_numbers = "https://github.com/kelloggjohnd/Final\_project/blob/main/data/presidential\_speeches\_numbers.csv"

def data\_pull(url):

download = requests.get(url).content

df = pd.read\_csv(io.StringIO(download.decode('utf-8')))

return df

def rolling\_change(df,President):

df['President'] = President

df['Approve\_rolling']= df['Approving'].shift(-1).rolling(2, 2).median().shift(-1)

df['Disapprove\_rolling'] = df['Disapproving'].shift(-1).rolling(2, 2).median().shift(-1)

df['Unsure\_rolling'] = df['Unsure/NoData'].shift(-1).rolling(2, 2).median().shift(-1)

df['Approve\_change'] = (df ['Approve\_rolling'] - df['Approving'])/df ['Approve\_rolling'] \*100

df['Disapprove\_change'] = (df ['Disapprove\_rolling'] - df['Disapproving'])/df ['Disapprove\_rolling'] \*100

df['Unsure\_change'] = (df ['Unsure\_rolling'] - df['Unsure/NoData'])/df ['Unsure\_rolling'] \*100

return df

def sentiment\_engine(president\_name):

df = all\_speechs[all\_speechs["President"] == president\_name]

df['Transcript']=df['Transcript'].str.lower()

df['Transcript']=df['Transcript'].str.strip().str.replace('[^\w\s]','')

pat = r'\b(?:{})\b'.format('|'.join(stop\_words))

df['Transcript'] = df['Transcript'].str.replace(pat, '')

transcript\_blob = [TextBlob(desc) for desc in df['Transcript']]

df['Polarity'] = [b.sentiment.polarity for b in transcript\_blob]

df['Subjectivity'] = [b.sentiment.subjectivity for b in transcript\_blob]

df['compound'] = [analyzer.polarity\_scores(v)['compound'] for v in df['Transcript']]

df['Negative'] = [analyzer.polarity\_scores(v)['neg'] for v in df['Transcript']]

df['Neutral'] = [analyzer.polarity\_scores(v)['neu'] for v in df['Transcript']]

df['Positive'] = [analyzer.polarity\_scores(v)['pos'] for v in df['Transcript']]

return df

def week\_number\_approve (df):

df['End Date'] = pd.to\_datetime(df['End Date'])

df['week\_number'] = df['End Date'].dt.week

df['year'] = df['End Date'].dt.year

df['week\_year'] = df['week\_number'].map(str)+'-'+df['year'].map(str)

return df

def week\_number\_speeches (df):

df['Date'] = pd.to\_datetime(df['Date'])

df['week\_number'] = df['Date'].dt.week

df['year'] = df['Date'].dt.year

df['week\_year'] = df['week\_number'].map(str)+'-'+df['year'].map(str)

return df

def data\_merge (speeches, approval):

df = speeches.merge(approval,how='inner', left\_on='week\_year', right\_on='week\_year')

df = df[['Date','Polarity','Subjectivity','compound','Negative','Neutral','Positive', 'Approve\_change','Disapprove\_change','Unsure\_change','week\_year']]

return df

def long\_form (merge\_df, column\_value):

df = merge\_df[[column\_value,'Approve\_change', 'Disapprove\_change','Unsure\_change']]

df = pd.melt(df, id\_vars=column\_value, value\_vars=['Approve\_change','Disapprove\_change','Unsure\_change'])

def long\_form\_adjust (parentdf, President, column\_value):

df = parentdf[parentdf['President'] == President]

df = df[[column\_value,'Approve\_change', 'Disapprove\_change','Unsure\_change']]

df = pd.melt(df, id\_vars=column\_value, value\_vars=['Approve\_change','Disapprove\_change','Unsure\_change'])

return df

all\_speechs = data\_pull(all\_speechs\_data)

#speech\_poll\_numbers = data\_pull(speech\_numbers)

speech\_poll\_numbers = pd.read\_csv (r'C:\Users\renje\Documents\GitHub\Final\_project\data\presidential\_speeches\_numbers.csv')

obama\_approval = data\_pull(obama\_rating)

bush41\_approval = data\_pull(bush41\_rating)

bush43\_approval = data\_pull(bush43\_rating)

clinton\_approval = data\_pull(clinton\_rating)

reagan\_approval = data\_pull(reagan\_rating)

carter\_approval = data\_pull(carter\_rating)

eisenhower\_approval = data\_pull(eisenhower\_rating)

ford\_approval = data\_pull(ford\_rating)

johnson\_approval = data\_pull(johnson\_rating)

kennedy\_approval = data\_pull(kennedy\_rating)

nixon\_approval = data\_pull(nixon\_rating)

truman\_approval = data\_pull(truman\_rating)

obama\_approval = rolling\_change(obama\_approval, "Barack Obama")

bush41\_approval = rolling\_change(bush41\_approval, "George H.W. Bush")

bush43\_approval = rolling\_change(bush43\_approval, 'George W. Bush')

clinton\_approval = rolling\_change(clinton\_approval, 'Bill Clinton')

reagan\_approval = rolling\_change(reagan\_approval, 'Ronald Reagan')

carter\_approval = rolling\_change(carter\_approval, 'Jimmy Carter')

eisenhower\_approval = rolling\_change(eisenhower\_approval, 'Dwight D. Eisenhower')

ford\_approval = rolling\_change(ford\_approval, "Gerald Ford")

johnson\_approval = rolling\_change(johnson\_approval, 'Lyndon B. Johnson')

kennedy\_approval = rolling\_change(kennedy\_approval, 'John F. Kennedy')

nixon\_approval = rolling\_change(nixon\_approval, 'Richard Nixon')

truman\_approval = rolling\_change(truman\_approval,'Herry S. Truman')

gallup\_totals = pd.concat([obama\_approval,bush43\_approval])

gallup\_totals = pd.concat([gallup\_totals,clinton\_approval])

gallup\_totals = pd.concat([gallup\_totals,bush41\_approval])

gallup\_totals = pd.concat([gallup\_totals,reagan\_approval])

gallup\_totals = pd.concat([gallup\_totals,carter\_approval])

gallup\_totals = pd.concat([gallup\_totals,ford\_approval])

gallup\_totals = pd.concat([gallup\_totals,nixon\_approval])

gallup\_totals = pd.concat([gallup\_totals,johnson\_approval])

gallup\_totals = pd.concat([gallup\_totals,kennedy\_approval])

gallup\_totals = pd.concat([gallup\_totals,eisenhower\_approval])

gallup\_totals = pd.concat([gallup\_totals,truman\_approval])

gallup\_totals = gallup\_totals[['Start Date','President','Approving', 'Disapproving', 'Unsure/NoData']]

gallup\_totals = pd.melt(gallup\_totals, id\_vars=['Start Date','President'], value\_vars=['Approving', 'Disapproving', 'Unsure/NoData'])

president\_list = ['Harry S. Truman',

'Richard M. Nixon',

'Dwight D. Eisenhower',

'John F. Kennedy',

'Lyndon B. Johnson',

'Ronald Reagan',

'Gerald Ford',

'Jimmy Carter',

'George H. W. Bush',

'Bill Clinton',

'George W. Bush',

'Barack Obama']

all\_speechs['Transcript'] = all\_speechs['Transcript'].astype(str)

president\_speeches = all\_speechs[all\_speechs.President.isin(president\_list)]

truman\_speeches = sentiment\_engine('Harry S. Truman')

truman\_speeches = truman\_speeches[~(truman\_speeches['Date'] < '1945-04-12')]

truman\_speeches = truman\_speeches[~(truman\_speeches['Date'] > '1953-01-20')]

eisenhower\_speeches = sentiment\_engine('Dwight D. Eisenhower')

eisenhower\_speeches = eisenhower\_speeches[~(eisenhower\_speeches['Date'] < '1953-01-20')]

eisenhower\_speeches = eisenhower\_speeches[~(eisenhower\_speeches['Date'] > '1961-01-20')]

kennedy\_speeches = sentiment\_engine('John F. Kennedy')

kennedy\_speeches = kennedy\_speeches[~(kennedy\_speeches['Date'] < '1961-01-20')]

kennedy\_speeches = kennedy\_speeches[~(kennedy\_speeches['Date'] > '1963-11-22')]

johnson\_speeches = sentiment\_engine('Lyndon B. Johnson')

johnson\_speeches = johnson\_speeches[~(johnson\_speeches['Date'] < '1963-11-22')]

johnson\_speeches = johnson\_speeches[~(johnson\_speeches['Date'] > '1969-01-20')]

nixon\_speeches = sentiment\_engine('Richard M. Nixon')

nixon\_speeches = nixon\_speeches[~(nixon\_speeches['Date'] < '1969-01-20')]

nixon\_speeches = nixon\_speeches[~(nixon\_speeches['Date'] > '1974-08-09')]

ford\_speeches = sentiment\_engine('Gerald Ford')

ford\_speeches = ford\_speeches[~(ford\_speeches['Date'] < '1974-08-09')]

ford\_speeches = ford\_speeches[~(ford\_speeches['Date'] > '1977-01-20')]

carter\_speeches = sentiment\_engine('Jimmy Carter')

carter\_speeches = carter\_speeches[~(carter\_speeches['Date'] < '1977-01-20')]

carter\_speeches = carter\_speeches[~(carter\_speeches['Date'] > '1981-01-20')]

reagan\_speeches = sentiment\_engine('Ronald Reagan')

reagan\_speeches = reagan\_speeches[~(reagan\_speeches['Date'] < '1981-01-20')]

reagan\_speeches = reagan\_speeches[~(reagan\_speeches['Date'] > '1989-01-20')]

bush41\_speeches = sentiment\_engine('George H. W. Bush')

bush41\_speeches = bush41\_speeches[~(bush41\_speeches['Date'] < '1989-01-20')]

bush41\_speeches = bush41\_speeches[~(bush41\_speeches['Date'] > '1993-01-20')]

clinton\_speeches = sentiment\_engine('Bill Clinton')

clinton\_speeches = clinton\_speeches[~(clinton\_speeches['Date'] < '1993-01-20')]

clinton\_speeches = clinton\_speeches[~(clinton\_speeches['Date'] > '2001-01-20')]

bush43\_speeches = sentiment\_engine('George W. Bush')

bush43\_speeches = bush43\_speeches[~(bush43\_speeches['Date'] < '2001-01-20')]

bush43\_speeches = bush43\_speeches[~(bush43\_speeches['Date'] > '2009-01-20')]

obama\_speeches = sentiment\_engine('Barack Obama')

obama\_speeches = obama\_speeches[~(obama\_speeches['Date'] < '2009-01-20')]

obama\_speeches = obama\_speeches[~(obama\_speeches['Date'] > '2017-01-20')]

truman\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\truman.csv', sep=',')

eisenhower\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\eisenhower.csv', sep=',')

kennedy\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\kennedy.csv', sep=',')

johnson\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\johnson.csv', sep=',')

nixon\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\nixon.csv', sep=',')

ford\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\ford.csv', sep=',')

carter\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\carter.csv', sep=',')

reagan\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\reagan.csv', sep=',')

bush41\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\bush41.csv', sep=',')

clinton\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\clinton.csv', sep=',')

bush43\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\bush43.csv', sep=',')

obama\_approval.to\_csv(r'C:\Users\renje\Documents\GitHub\Final\_project\data\approval\_ratings\Adjusted\obama.csv', sep=',')

truman\_merge = truman\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

eisenhower\_merge = eisenhower\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

kennedy\_merge = kennedy\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

johnson\_merge = johnson\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

nixon\_merge = nixon\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

ford\_merge = ford\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

carter\_merge = carter\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

reagan\_merge = reagan\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

bush41\_merge = bush41\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

clinton\_merge = clinton\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

bush43\_merge = bush43\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

obama\_merge = obama\_speeches[['Date', 'President', 'Polarity', 'Subjectivity','compound','Negative', 'Neutral','Positive']]

All\_merge = pd.concat([obama\_merge,bush43\_merge])

All\_merge = pd.concat([All\_merge,clinton\_merge])

All\_merge = pd.concat([All\_merge,bush41\_merge])

All\_merge = pd.concat([All\_merge,reagan\_merge])

All\_merge = pd.concat([All\_merge,carter\_merge])

All\_merge = pd.concat([All\_merge,ford\_merge])

All\_merge = pd.concat([All\_merge,nixon\_merge])

All\_merge = pd.concat([All\_merge,johnson\_merge])

All\_merge = pd.concat([All\_merge,kennedy\_merge])

All\_merge = pd.concat([All\_merge,eisenhower\_merge])

All\_merge = pd.concat([All\_merge,truman\_merge])

All\_merge['Date'] = pd.to\_datetime(All\_merge['Date'])

speech\_poll\_numbers['Date'] = pd.to\_datetime(speech\_poll\_numbers['Date'])

merged = speech\_poll\_numbers.merge(All\_merge, left\_on=['Date','President'], right\_on = ['Date','President'])

merged = merged.drop(columns=['Approve\_rolling','Disapprove\_rolling','Unsure\_rolling'])

All\_merge\_numbers = All\_merge.groupby('President').mean()

gallup\_totals = pd.concat([obama\_approval,bush43\_merge])

All\_merge = pd.concat([All\_merge,clinton\_merge])

All\_merge = pd.concat([All\_merge,bush41\_merge])

All\_merge = pd.concat([All\_merge,reagan\_merge])

All\_merge = pd.concat([All\_merge,carter\_merge])

All\_merge = pd.concat([All\_merge,ford\_merge])

All\_merge = pd.concat([All\_merge,nixon\_merge])

All\_merge = pd.concat([All\_merge,johnson\_merge])

All\_merge = pd.concat([All\_merge,kennedy\_merge])

All\_merge = pd.concat([All\_merge,eisenhower\_merge])

All\_merge = pd.concat([All\_merge,truman\_merge])

###########Graphics###########

pio.templates.default = "simple\_white"

def speech\_graphs (df, President, merge\_df, approval\_df):

df\_polarity = long\_form\_adjust(df,President,'Polarity')

fig1 = px.scatter(df\_polarity, x="Polarity", y="value", facet\_row="variable", trendline='ols', title= President+" Polarity")

df\_subjectivity = long\_form\_adjust(df,President,'Subjectivity')

fig2 = px.scatter(df\_subjectivity, x="Subjectivity", y="value", facet\_row="variable", trendline='ols', title=President+" Subjectivity")

df\_positive = long\_form\_adjust(df,President,'Positive')

fig3 = px.scatter(df\_positive, x="Positive", y="value", facet\_row="variable", trendline='ols', title=President+" Positive")

df\_negative = long\_form\_adjust(df,President,'Negative')

fig4 = px.scatter(df\_negative, x="Negative", y="value", facet\_row="variable", trendline='ols', title=President+" Negative")

df\_neutral = long\_form\_adjust(df,President,'Neutral')

fig5 = px.scatter(df\_neutral, x="Neutral", y="value", facet\_row="variable", trendline='ols', title=President+" Neutral")

df\_numbers = merge\_df[['Date','Negative','Neutral','Positive']]

df\_numbers = pd.melt(df\_numbers, id\_vars='Date', value\_vars=['Negative','Neutral','Positive'])

fig6 = px.bar(df\_numbers, x="Date", y="value", color="variable", title=President+" Speech Breakdown")

fig6.update\_xaxes(type='category')

df\_polar\_subj = merge\_df[['Date','Polarity','Subjectivity']]

df\_polar\_subj = pd.melt(df\_polar\_subj, id\_vars='Date', value\_vars=['Polarity','Subjectivity'])

fig7 = px.line(df\_polar\_subj, x = 'Date', y='value', color='variable', facet\_row='variable', title = President+" Polarity & Subjectivity")

fig7.update\_xaxes(type='category')

fig7.update\_layout(legend=dict(

yanchor="top",

y=0.99,

xanchor="left",

x=0.01

))

approval = approval\_df [['Start Date','Approving', 'Disapproving', 'Unsure/NoData']]

approval = pd.melt(approval, id\_vars='Start Date', value\_vars=['Approving', 'Disapproving', 'Unsure/NoData'])

fig8 = px.line(approval, x="Start Date", y="value", color="variable", facet\_row='variable',title=President+" Approval Breakdown")

fig1.show()

fig2.show()

fig3.show()

fig4.show()

fig5.show()

fig6.show()

fig7.show()

fig8.show()

return fig1, fig2, fig3, fig4,fig5,fig6,fig7,fig8

###########Gallup###########

gallup\_totals['Start Date'] = pd.to\_datetime(gallup\_totals['Start Date'])

gallup\_totals['Start Date']= gallup\_totals['Start Date'].dt.date

gallup\_totals.sort\_values('Start Date', inplace=True)

gallup\_fig = px.line(gallup\_totals, x="Start Date", y="value", color="variable", facet\_row='variable',title="Gallup Breakdown")

gallup\_fig.update\_xaxes(type='category')

fig.update\_layout(legend=dict(

yanchor="top",

y=0.99,

xanchor="left",

x=0.01

))

gallup\_fig.show()

############CORR Plots##############

import matplotlib

matplotlib.use("TKAgg")

corr = merged.corr()

############Obama##############

speech\_graphs(merged,"Barack Obama", obama\_merge,obama\_approval)

############Bush43##############

speech\_graphs(merged,"George W. Bush", bush43\_merge,bush43\_approval)

############clinton##############

speech\_graphs(merged,"Bill Clinton", clinton\_merge,clinton\_approval)

############Bush41##############

speech\_graphs(merged,"George H. W. Bush", bush41\_merge,bush41\_approval)

############reagan##############

speech\_graphs(merged,"Ronald Reagan", reagan\_merge,reagan\_approval)

############Carter##############

speech\_graphs(merged,"Jimmy Carter", carter\_merge,carter\_approval)

############ford##############

speech\_graphs(merged,"Gerald Ford", ford\_merge,ford\_approval)

############Nixon##############

speech\_graphs(merged,"Richard M. Nixon", nixon\_merge,nixon\_approval)

############Johnson##############

speech\_graphs(merged,"Lyndon B. Johnson", johnson\_merge,johnson\_approval)

############Kennedy##############

speech\_graphs(merged,"John F. Kennedy", kennedy\_merge,kennedy\_approval)

############Eisenhower##############

speech\_graphs(merged,"Dwight D. Eisenhower", eisenhower\_merge,eisenhower\_approval)

############Truman##############

speech\_graphs(merged,"Harry S. Truman", truman\_merge,truman\_approval)

**R Studio**

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

library(tidyverse)

library(ggplot2)

library(ggcorrplot)

library(GGally)

```

```{r}

library(readr)

merged <- read\_csv("C:/Users/renje/Documents/GitHub/Final\_project/data/approval\_ratings/Adjusted/merged.csv")

```

```{r}

merged\_numbers <- merged %>%

select(-1, -Date, -President, -compound)

#merged\_numbers[] <- lapply(merged\_numbers, function(x) as.numeric(as.character(x)))

corr <- round(cor(merged\_numbers),1)

#head(corr[,1:6])

```

```{r}

merged\_numbers <- merged%>%

rename(APP = Approving)%>%

rename(DIS = Disapproving)%>%

rename(UNS = `Unsure/NoData`)%>%

rename(APPC = Approve\_change)%>%

rename(DISC = Disapprove\_change)%>%

rename(UNSC = Unsure\_change)%>%

rename(POL = Polarity)%>%

rename(SUB = Subjectivity)%>%

rename(NEG = Negative)%>%

rename(NEU = Neutral)%>%

rename(POS = Positive)

merged\_numbers\_all <- merged\_numbers %>%

select(-1, -Date, -President, -compound)

ggcorr(merged\_numbers\_all, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "All Presidents Coorelation Plot")

```

```{r}

ggcorrplot(corr, hc.order = FALSE,lab = TRUE, title = "All Presidents Coorelation Plot")

```

```{r}

truman\_numbers <- merged\_numbers %>%

filter(President == "Harry S. Truman")%>%

select(-1, -Date, -President, -compound)

ggcorr(truman\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Truman Coorelation Plot")

```

```{r}

Eisenhower\_numbers <- merged\_numbers %>%

filter(President == "Dwight D. Eisenhower")%>%

select(-1, -Date, -President, -compound)

ggcorr(Eisenhower\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Eisenhower Coorelation Plot")

```

```{r}

kennedy\_numbers <- merged\_numbers %>%

filter(President == "John F. Kennedy")%>%

select(-1, -Date, -President, -compound)

ggcorr(kennedy\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Kennedy Coorelation Plot")

```

```{r}

Johnson\_numbers <- merged\_numbers %>%

filter(President == "Lyndon B. Johnson")%>%

select(-1, -Date, -President, -compound)

ggcorr(Johnson\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Johnson Coorelation Plot")

```

```{r}

Nixon\_numbers <- merged\_numbers %>%

filter(President == "Richard M. Nixon")%>%

select(-1, -Date, -President, -compound)

ggcorr(Nixon\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Nixon Coorelation Plot")

```

```{r}

Ford\_numbers <- merged\_numbers %>%

filter(President == "Gerald Ford")%>%

select(-1, -Date, -President, -compound)

ggcorr(Ford\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Ford Coorelation Plot")

```

```{r}

Carter\_numbers <- merged\_numbers %>%

filter(President == "Jimmy Carter")%>%

select(-1, -Date, -President, -compound)

ggcorr(Carter\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Carter Coorelation Plot")

```

```{r}

Reagan\_numbers <- merged\_numbers %>%

filter(President == "Ronald Reagan")%>%

select(-1, -Date, -President, -compound)

ggcorr(Reagan\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Reagan Coorelation Plot")

```

```{r}

Bush41\_numbers <- merged\_numbers %>%

filter(President == "George H. W. Bush")%>%

select(-1, -Date, -President, -compound)

ggcorr(Bush41\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "H.W.Bush Coorelation Plot")

```

```{r}

Clinton\_numbers <- merged\_numbers %>%

filter(President == "Bill Clinton")%>%

select(-1, -Date, -President, -compound)

ggcorr(Clinton\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Clinton Coorelation Plot")

```

```{r}

Bush43\_numbers <- merged\_numbers %>%

filter(President == "George W. Bush")%>%

select(-1, -Date, -President, -compound)

ggcorr(Bush43\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "W.Bush Coorelation Plot")

```

```{r}

Obama\_numbers <- merged\_numbers %>%

filter(President == "Barack Obama")%>%

select(-1, -Date, -President, -compound)

ggcorr(Obama\_numbers, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Obama Coorelation Plot")

```

```{r}

Group\_1 <- c('Harry S. Truman', "Richard M. Nixon","Jimmy Carter","George H. W. Bush", "George W. Bush")

merged\_group\_1 <- merged\_numbers%>%

filter (President %in% Group\_1)

merged\_group\_1 <- merged\_group\_1 %>%

select(-1, -Date, -President, -compound)

merged\_group\_2 <- merged\_numbers%>%

filter (!President %in% Group\_1)

merged\_group\_2 <- merged\_group\_2 %>%

select(-1, -Date, -President, -compound)

ggcorr(merged\_group\_1, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Group1 Correlation Plot")

```

```{r}

ggcorr(merged\_group\_2, method = c("everything", "pearson"), label = TRUE, legend.position = "right", label\_size = 4,hjust = .5, size = 3, low = "midnightblue", mid = "white", high = "darkred")+

labs(title = "Group2 Correlation Plot")

```

Appendix A

Presidential Speech graphs

Harry S Truman

Need to fix

Dwight D. Eisenhower

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

John F Kennedy

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Lyndon B Johnson

Fix it

Richard Nixon

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated  
Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

Gerald Ford

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated



Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Ronald Reagan

Chart, line chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

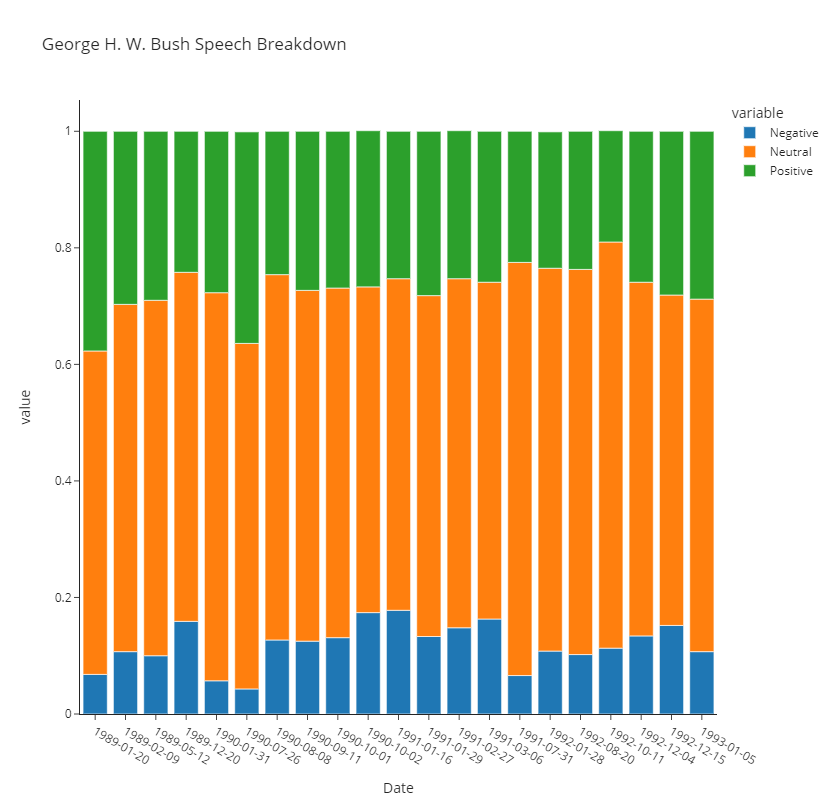
Chart, line chart, scatter chart

Description automatically generated

George H.W. Bush

Chart, line chart

Description automatically generated



Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Bill Clinton

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

George W. Bush

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated