**Text Mining 2016 Election Tweets**

**Description**

In the following analysis, we examine Twitter data pertaining to campaign tweets from Hillary Clinton and Donald Trump during the 2016 election. First, examining tweet frequency by candidate shows their general use pattern. Then, word frequency analysis reveals the most common words used by each candidate, demonstrating their messaging focus. Next, we compare candidate word usage - the differences in the language used by the candidates and their distinct rhetorical strategies. Then, we examine the changes in top word use over time, highlighting shifting campaign priorities. Lastly, we analyze favorites and retweets, indicating how well their messages resonated with viewers.

**Conclusions**

***Tweet Frequency***

***A screenshot of a graph

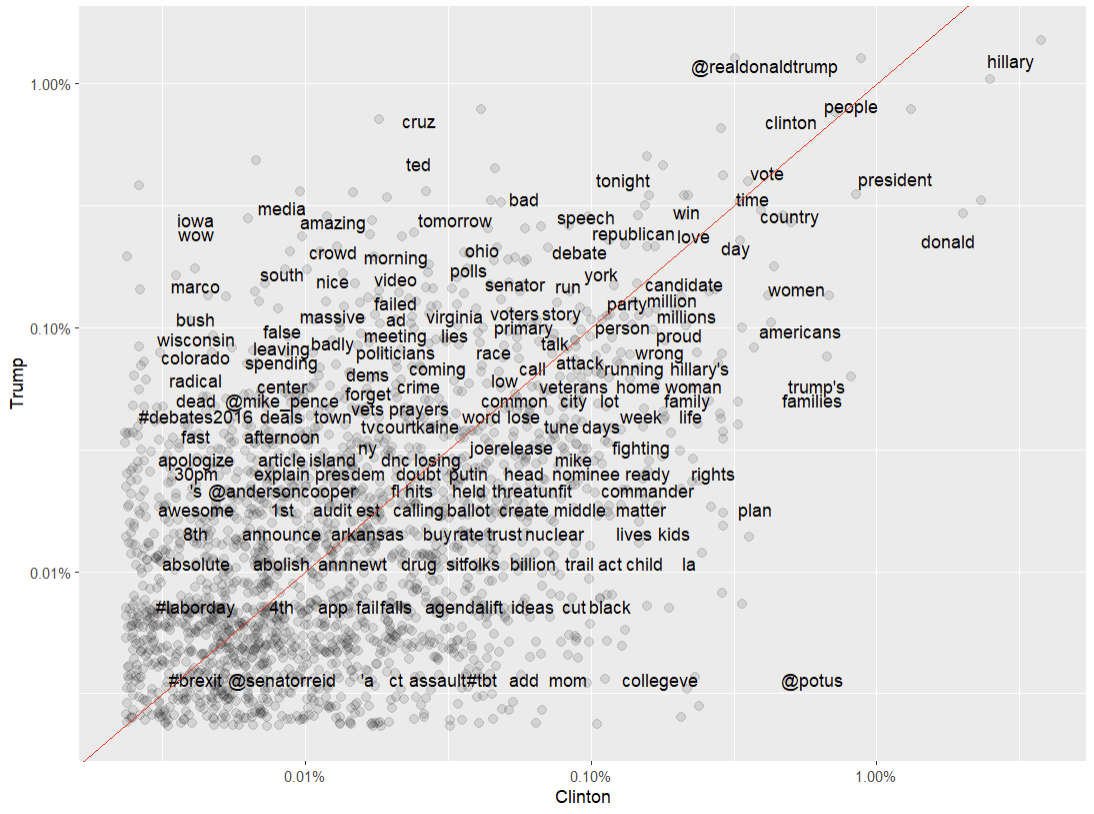
Description automatically generated***

**Figure 1: Tweet Frequency**

Trump maintained consistent tweet volume over the year, with minor spikes in the spring and fall, as depicted in Figure 1. In contrast, Clinton's tweet activity surged starting in June, peaking in July and again in October. This analysis showcased Trump’s steady use of Twitter, while Clinton increased her presence as the election approached. In other words, Trump was active on Twitter long before the presidential campaign season began.

***Word Frequency Analysis***

Figure 2 depicts the results of word frequency analysis from 2016 election tweets. Words to the right of the red line are more associated with Clinton, while those to the left are more likely to be associated with Trump. Words appearing on or near the line appear with roughly equal frequency from both candidates. The distribution of words demonstrates clear differences in messaging and campaign rhetoric between the two candidates.

******

**Figure 2: Word Frequency Analysis**

***Comparison of Word Usage***

The first clear difference to emerge between the two candidates is that Clinton maintained active Spanish language outreach efforts on Twitter during her campaign. The initial version of Figure 3 included a handful of Spanish stopwords as words most likely to be associated with Clinton. Removing these stopwords ("de", "en", "los", "el", "para", "ve", "es", "por", "la") resulted in Figure 3. This adjustment provides a more accurate representation of the English language terms most emphasized by each campaign.

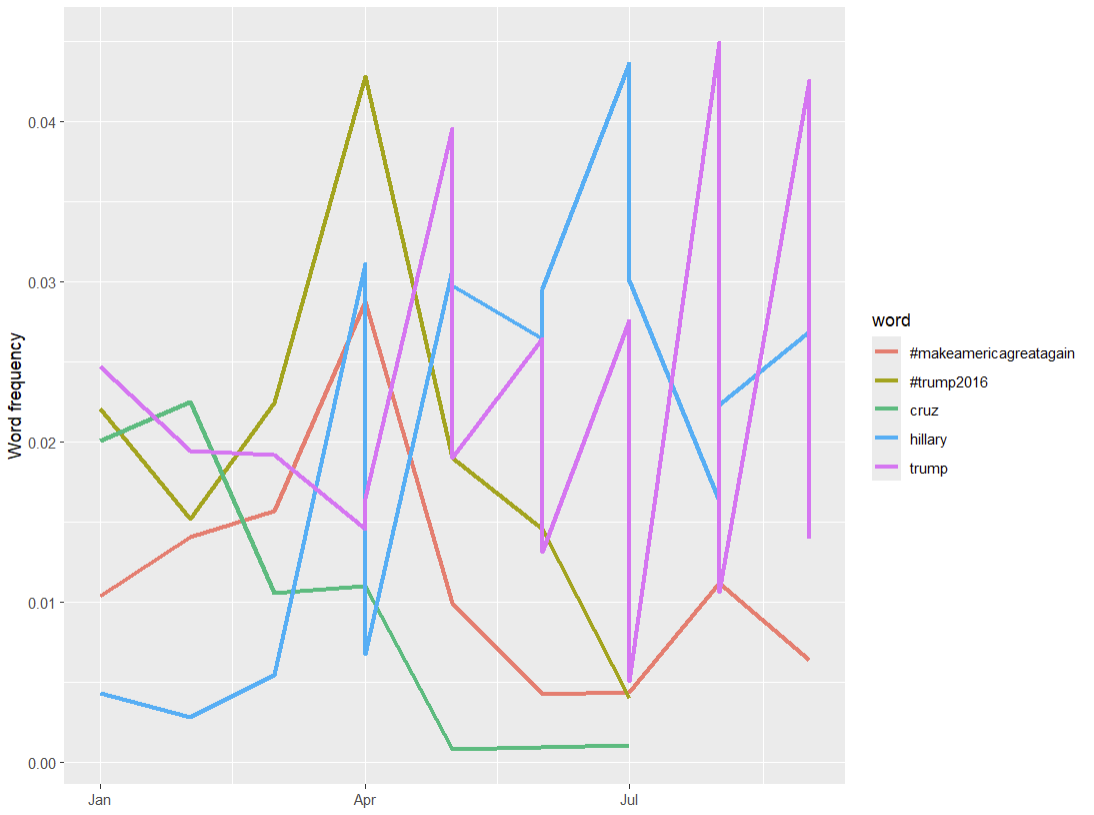
A graph of a graph

Description automatically generated with medium confidence

**Figure 3: Comparison of Word Usage**

***Changes in Word Use Analysis***

Both word use analysis graphs show how both Trump and Clinton's tweets reflected efforts to stand out from rivals. In Figure 4 (Trump), terms like "Cruz" and "#MakeAmericaGreatAgain" peak early in the year, emphasizing his distinction from Republican primary opponents. As the campaign progresses, "Hillary" and related terms rise, highlighting his focus on the general election.



**Figure 4: Changes in Word Use - Trump**

In Figure 5 (Clinton), "Trump" consistently dominates, indicating that even in her messaging, Clinton frequently referenced Trump as a key point of contrast. Terms like "college" and "rights" appear less frequently, reflecting her focus on specific policy issues in contrast to Trump’s nationalistic messaging focus.

A graph of different colored lines

Description automatically generated

**Figure 5: Changes in Word Use – Clinton**

***Favorites and Retweets Analysis***

***A screenshot of a graph

Description automatically generated***

**Figure 6: Median Retweets by Top Words**

**A graph of a graph showing a number of different colored bars

Description automatically generated with medium confidence**

**Figure 7: Median Favorites by Top Words**

Figure 6 and Figure 7 compare the median number of retweets and favorites for specific words used in tweets by Clinton and Trump. For Clinton, words like "Alicia" (referring to Alicia Machado), "release," and "praise" received high engagement in terms of retweets, while "Khan," "thrilled," and "changemaker" were prominent in favorites. This reflects her focus on specific individuals and events during her campaign.

On Trump’s side, terms such as "emails," "FBI," and "condolences" garnered the most retweets, while "condolences," "France," and "FBI" led in favorites, indicating significant public attention to security, scandals, and international affairs. These patterns suggest that both candidates sparked high engagement around different topics. Clinton focused on personal stories and debate moments, while Trump focused on controversies and international incidents.

**Closing Remarks**

Analyzing Clinton and Trump's 2016 tweets revealed a striking contrast in their online struggle for votes. As expected (for the time), Trump commanded Twitter with a relentless stream of pronouncements and attacks, while Clinton adopted a more traditional approach focused on policy. Now that eight years have passed since the 2016 election, as a politically engaged U.S. citizen, most of the insights from this dataset feel like old, unsurprising news. It’s a shame that current data from X (formerly Twitter) are no longer freely available for public research and usage without paying an exorbitant API fee.