Individual Assignment 2: Visualizing Text

Parker Quinn

Northwestern University

Problem Summary

This report analyzes and visualizes text from twitter posts of three U.S. Presidential candidates – Bernie Sanders, Hillary Clinton, and Donald Trump. Twitter is a common communication platform for public figures and is frequently used as a way to explore beliefs and sentiments they hold. The purpose of this analysis is to reveal insights relating to the campaigns and personalities of each candidate.

Methodology and Programming

There are three modeling techniques that are used in this analysis. First, we use data visualization to display a high-level view of the frequently used terms used in twitter posts by each candidate. Specifically, we show a word cloud for each candidate, with word frequency determining its size in the word cloud. Second, we use a form of unsupervised machine learning – clustering algorithms – to find groupings of twitter posts and display the most frequent terms from each "cluster" to find common themes. Finally, we use sentiment analysis to assign the words used by the candidates an associated emotion (anger, anticipation, disgust, fear, joy, sadness, surprise, or trust) and overall sentiment (positive or negative). Sentiment analysis helps describe the emotional tone of each candidate.

Programmatically, this involved a lot of pre-processing of the original data. Primarily, this includes separating each post into individual terms, removing non-alphanumeric terms, links, numbers, punctuation, stop words ("the", "and", "it", etc.), and converting words to lower case. Part of the output includes a document-term-matrix with frequency calculations for each term. For analysis, sparse terms – those that do not appear in at least 0.5% of the twitter posts – were removed. Furthermore, the number of

clusters is different for each candidate because there were a different number of natural groupings in the posts (these were determined visually). To assign emotions and sentiments to each term, we use the "Emotion Lexicon" created by the National Research Council Canada.

Results

The word clouds in Figures 1-3 show some of the favorite terms for each candidate, like "people" (Sanders), "need" (Clinton), and "great" (Trump). Figures 4-6 show the clusters that were found for each candidate, the ten most common words in each cluster, and a short description of the associated theme. We can see that the democratic candidates share some themes, like healthcare and campaign participation, while Trump tends have themes of gratitude and poll results. We can also see that Trump and Clinton spend more time attacking each other and other candidates, while Sanders features more pointed discussions on policy and ideology. The results from the sentiment analysis (Figures 7-12) show a similar distribution of emotion and positivity from each candidate. We see that Trump tends to use more words associated with sadness, surprise, and disgust, and fewer words associated with trust and anticipation than the democratic candidates. Finally, we can see that the democratic candidates have a slightly higher percentage of positive words than Trump. Many of these findings agree with the public narrative of the 2016 campaign, but also points to the many similarities in the candidates' communication with the public.

References

Chang, W. 2013. *R Graphics Cookbook.* Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1449316952]

Mohammad, Saif. "NRC Word-Emotion Association Lexicon." *Researcher in Computational Linguistics, National Research Council Canada.* 2015. http://saifmohammad.com/.

Appendix

WORD CLOUDS

Figure 1 – Sanders word cloud

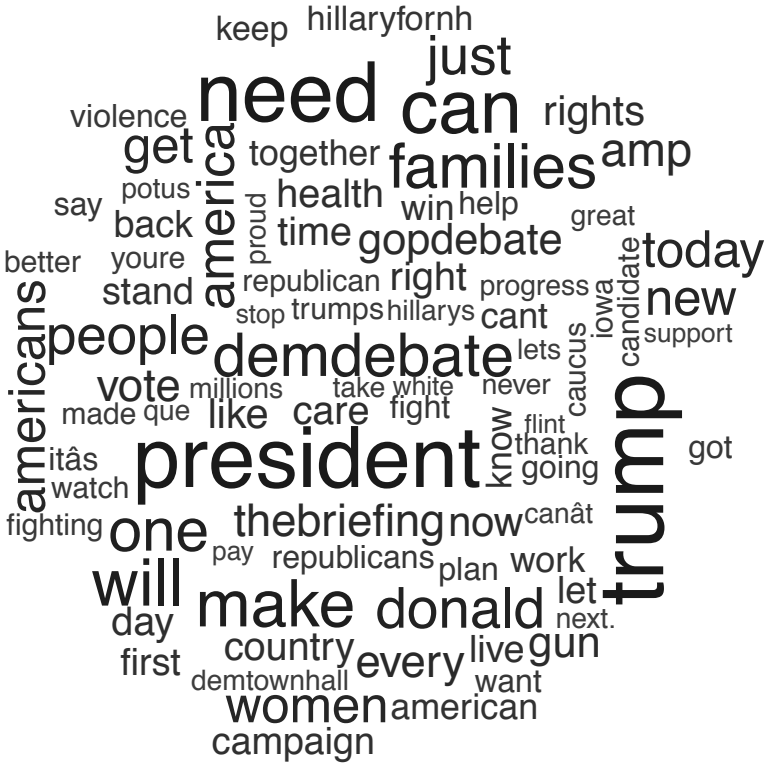

Figure 2 – Clinton word cloud

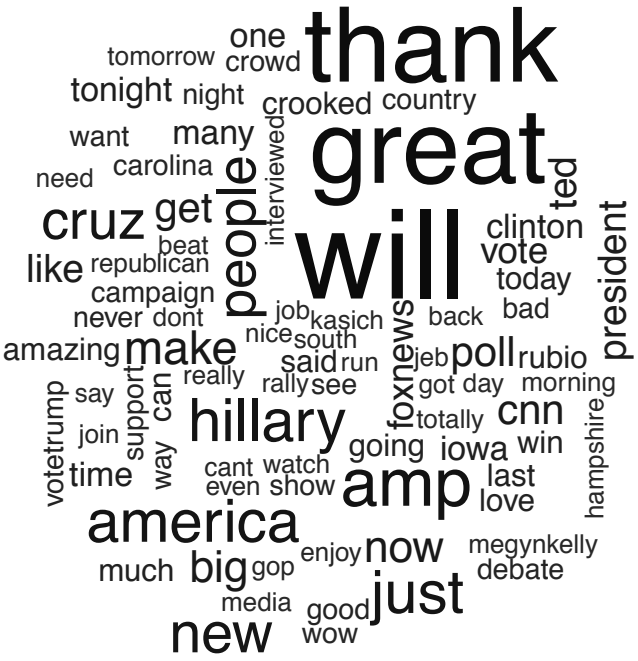

Figure 3 – Trump word cloud

CLUSTERING

Cluster	Top 10 terms	Description
1	People – campaign – must – need – will – country – can – change – just – together	Together, the people can change the country
2	Wall – street – will – big – greed – economy – millions – people – Americans – congress	Wall street greed affects millions
3	Health – care – right – plan – people – every – country – major – guarantee – Americans	Every major country guarantees health care, except the U.S.
4	political – revolution – vote – family – friends – ready – grab – join – like – people	Reach out to friends and family to join the political revolution
5	vote – today – new – make – win – get – register – York – people – can	Make sure you vote
6	live – watch – rally – tune – msnbc – campaign – believe – speaking – future – stage	Tune in to my speeches

Figure 4 - Sanders clusters

Cluster	Top 10 terms	Description
1	can – one – will – demdebate – need – people – just – today – new – get	Tune in to the debate
2	Need – families – make – gun – America – violence –will – keep – can – sure	Keep families safe from fun violence
3	President – court – supreme – next – republican – will – can't – united – Americans – let	Can't let Republicans control the Supreme Court
4	Trump – Donald – president – thebriefing – just – says – hes – take – Americans - running	Do not trust Trump
5	Care – health – affordable – child – act – womens – Americans – demdebate – millions - women	Affordable health care is important, especially for women/children

Figure 5 - Clinton clusters

Cluster	Top 10 terms	Description
1	Great – will – amp – people – just – thank – now – like – president – America	The people will make America great
2	Thank – great – votetrump – support – new – poll – nice – get – words – will	Thanks for the support
3	Hillary – Clinton – crooked – will – bad – beat – Bernie – can – women – isis	Do not trust Clinton or Sanders
4	Will – enjoy – interviewed – great – tonight – back – amp – morning – soon – stop	Watch my recent interview
5	Great – make – America – will – big – thank – safe – vote – amp – get	Thanks for voting to make America great
6	New – poll – Hampshire – York – thank – just – big – Cruz – lead - cnn	New poll in NH/NY shows me in the lead
7	Cruz – Ted – Rubio – Marco – lyin – amp – just – lightweight – senator – win	My opponents are weak/untrustworthy

Figure 6 - Donald Trump twitter clusters

WORD EMOTIONS

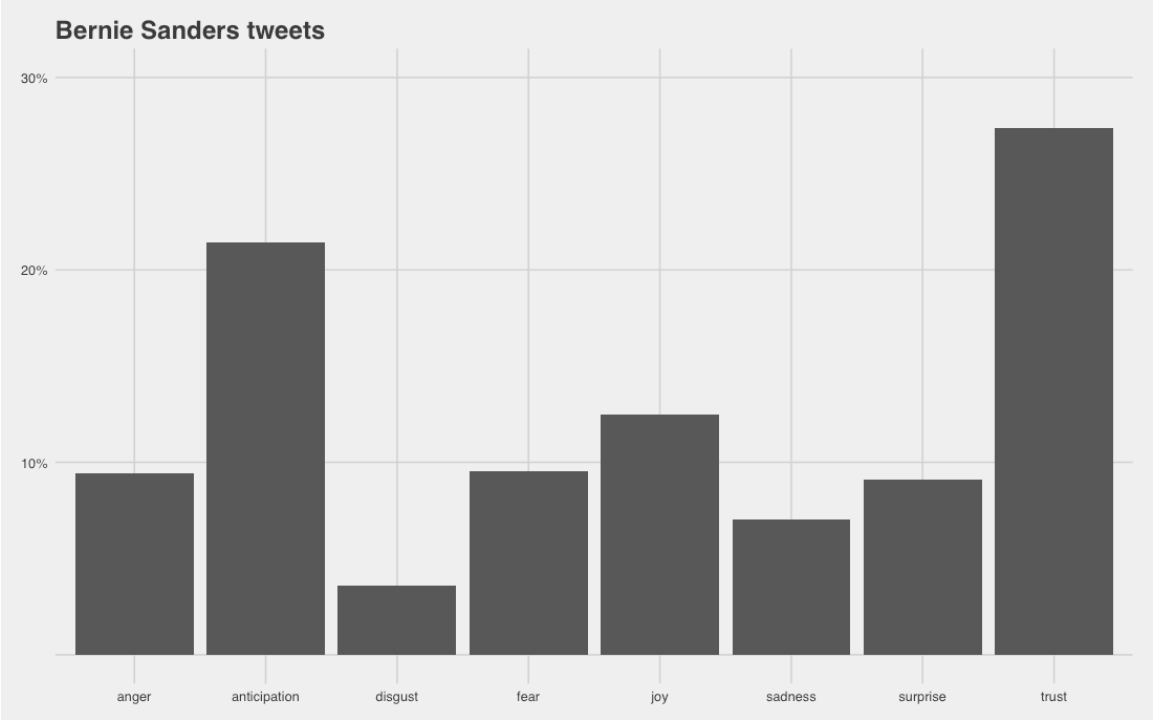

Figure 7 – Sanders emotions

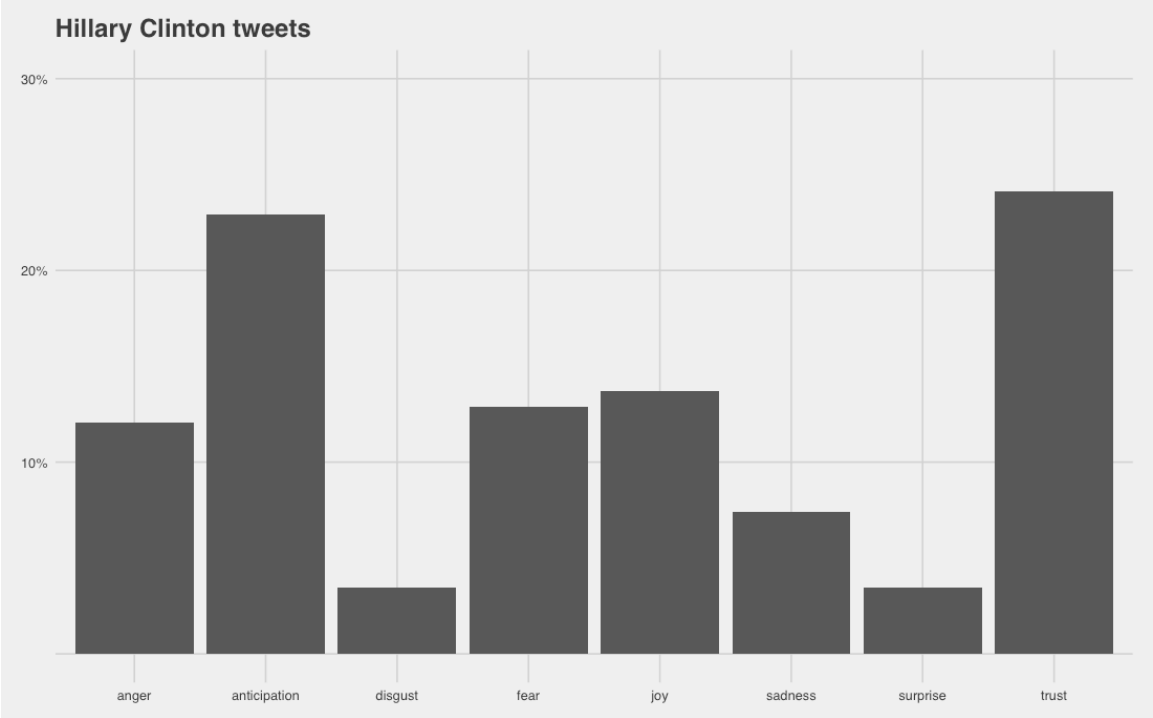

Figure 8 – Clinton emotions

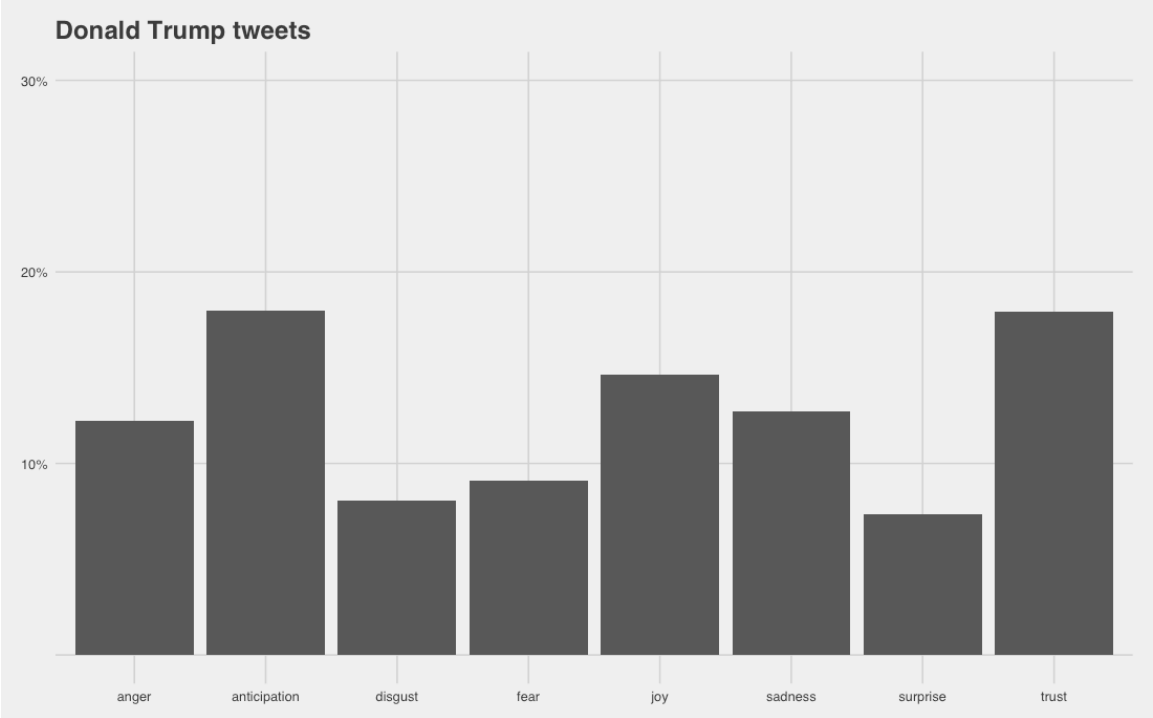

Figure 9 – Trump emotions

POSITIVE VS. NEGATIVE WORDS

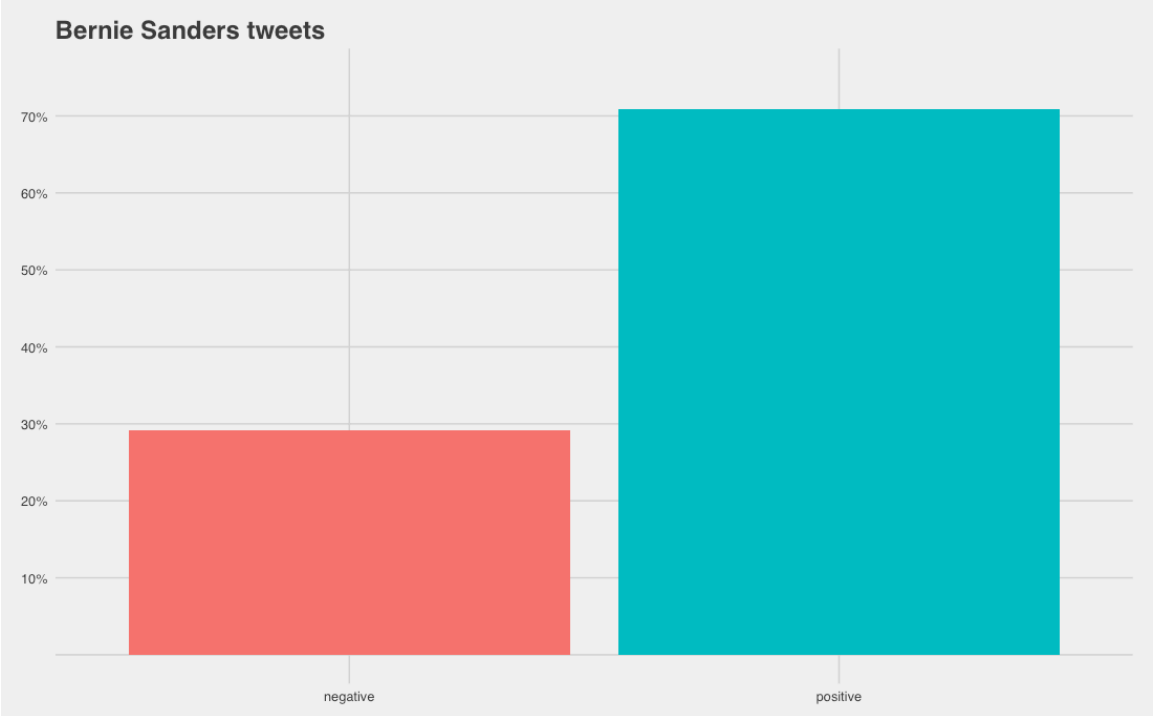

Figure 10 – Sanders positive vs. negative

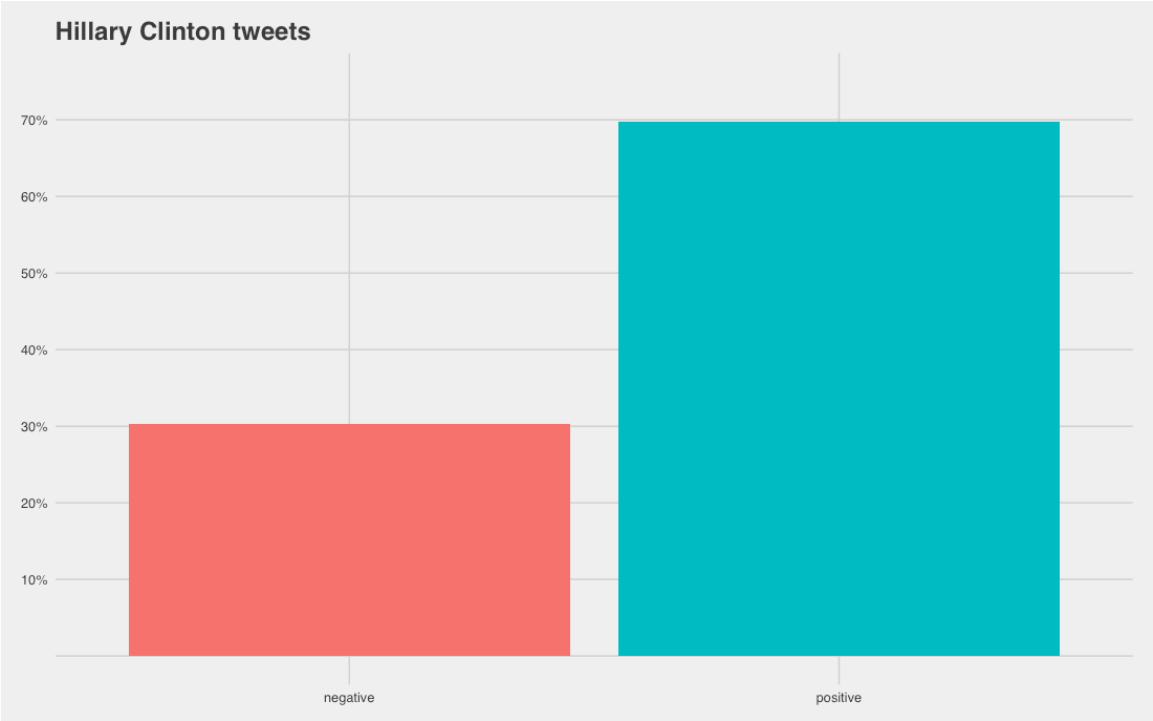

Figure 11 – Clinton positive vs. negative

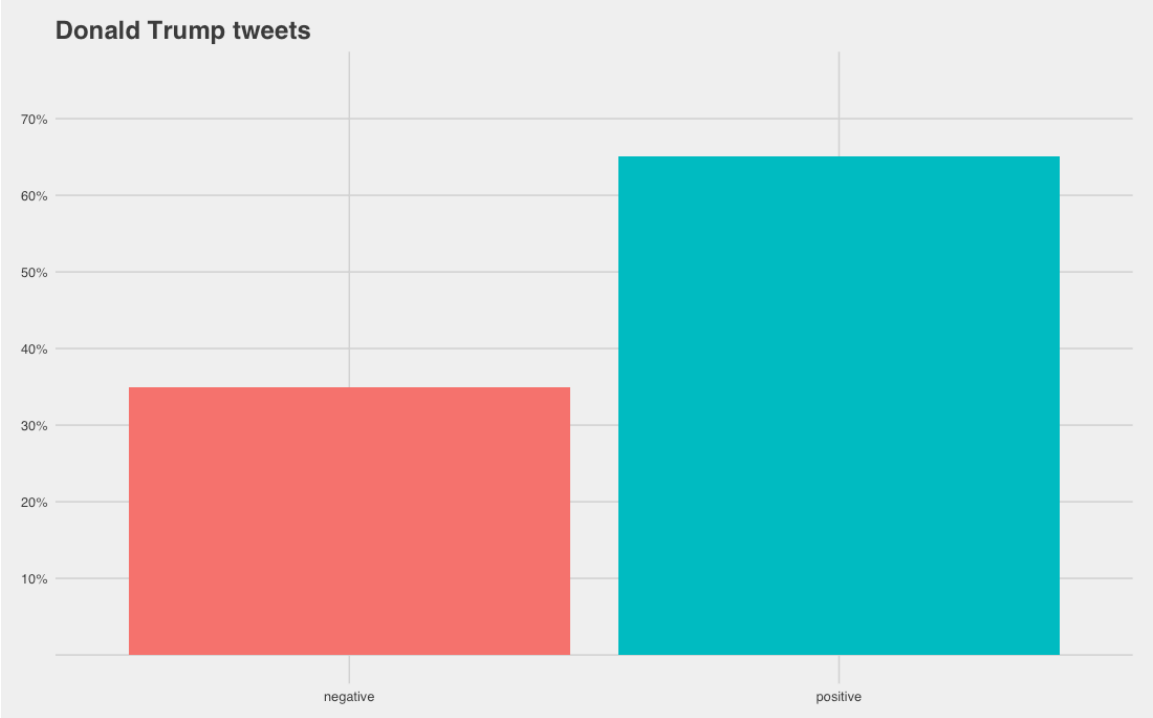

Figure 12 – Trump positive vs. negative