The 2020 Twitter Primary

Exploratory sentiment analysis and topic modeling of tweets

about the 2020 Democratic presidential candidates

Contribution statement:

Grace Wright, Minyoung Do, Steven Cao, Soowon Jo, and Audrey Glaser conceived of and designed the research project. M.D., S.C., and S.J. collected the Twitter data. M.D. wrote codes, cleaned the data, and ran the sentiment analysis. A.G. ran the statistical analysis of the sentiment analysis results. S.C. ran the topic modeling analysis. S.J. wrote the introduction and literature review. G.W. wrote the discussion section. M.D. designed the slides for the presentation.

INTRODUCTION

2020 is an important year for United States politics because it is a presidential election year for the nation. Although the bulk of the contention is usually between the Republican and Democratic nominees as they vie for presidency, attention is presently focused on which Democratic candidate will win more pledged delegates toward the party nomination; the Republican primary has not indicated the same degree of contention since it is likely that Donald Trump will run as the Republican candidate in the election. Moreover, February 2020 polls have shown a few Democratic candidates such as Former Vice President Joe Biden, Senator Bernie Sanders, Michael Bloomberg, Mayor Pete Buttigieg, Senator Elizabeth Warren, and Senator Amy Klobuchar beating President Donald Trump in hypothetical general election match-ups.[1] To explore how people are discussing such trends, we decided to analyze opinions posted on Twitter toward these Democratic candidates.

Twitter is one of the most popular microblogging social platforms in the United States for up-to-the-minute status updates with 330 million monthly active users as of August 2019.[2] Twitter users post short messages called tweets which have less than 280 characters, forcing Twitter users to disseminate their thoughts and ideas in a brief manner. In addition, these messages often express the user’s emotions towards whichever topics and ideas they wish to highlight. With the current presidential candidate nomination in mind, there have been many tweets - and consequently, many emotional expressions both positive and negative - which have been created in response to the various Democratic candidates. And given the increasingly-polarized nature of modern U.S. politics between party lines, we had found a strong

[1] Retrieved from: https://www.realclearpolitics.com/epolls/latest_polls/general_election/
[2] Retrieved from: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

impetus to explore the degree to which politics is polarized *within* party lines: in this case, within the Democratic party and their presidential candidates. Because expressions on Twitter may be able to capture some of the sentiment from this political atmosphere, we decided to embark on an exploratory data analysis in which we explore the emotional aspect of these tweets.

For this study, we will be using sentiment analysis to extract a sentiment score for the text within each tweet. We will explore overall sentiment towards each 2020 Democratic candidate to look at differences between each candidate. If each candidate's distribution of sentiment scores contain relatively more tweets with a high magnitude of sentiment (positive or negative), then this could indicate that Twitter users' political opinions are polarized. Additionally, if certain candidates have an overall sentiment score tending towards the positive or negative, then this could also indicate polarization within the same party. In addition, we did a topic modeling analysis to see which topics were most salient in relation to the candidates, with the goal of adding semantic context to the results of the sentiment analysis.

One caveat of our approach, however, is that we cannot straightforwardly generalize our findings about Twitter users towards the entire U.S. population, because Twitter users as a group often have a distinct expression of opinions in the public when it comes to political values.[3] Although sentiment analysis of social media discussions about political candidates has not yet been shown to be a strong predictor of candidate support or election results, our method may still provide useful information about the shifts in social sentiment surrounding such candidates, which in turn can ultimately help shape their performance at the polls.

[3] Retrieved from: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

LITERATURE REVIEW

In sentiment analysis of Twitter data, previous researchers have mainly focused on classifying tweets as positive or negative to detect emotions within tweets.[4] One of the approaches involved web-based text mining for detecting the emotions of an event embedded in English sentences. In this approach, the researchers found that the emotion-sensing problem was context-sensitive.[5] Another study by Tumasjan et al. on the German federal election supported the finding that microblogging message content is a valid indicator of political sentiment. Once they analyzed 10,000 messages that had a reference to politicians, they found that the majority of messages mentioning a party reflects the election result.[6] Yet another study has built a model that monitors political sentiment among Twitter users and predicts their voting intentions during the Irish General Election.[7] The primary methodological approach for this study was to combine sentiment analysis using supervised learning and volume-based measures. For sentiment analysis, they used a classifier trained on annotated data which consisted of multiple sentiment classes (positive, negative, mixed, neutral).[8] As a result, they found that volume played the single most significant indicator for predicting the election outcome and identified a dramatic shift in

---

[4] Khairnar, J., & Kinikar, M. (2013). Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, 3(6)
[5] Balahur, A., Hermida, J.M., & Montoyo, A. (2012). Detecting implicit expressions of emotion in text: A comparative analysis. Decision Support Systems, 53(4), 10.1016/j.dss.2012.05.024
[6] Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*.
[7] Bermingham, A., & Smeaton, A.F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results.
[8] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. HLT/EMNLP

sentiment towards the parties in two days before polling day, having possibly hinted at the election outcome. Together, these studies help set the stage for analyzing the relationship between tweet sentiments and political outcomes. Additionally, topic modeling will be performed on the tweets in order to investigate any potentially interesting semantic contexts underlying tweet sentiments.

DATA & METHODS

Data Collection

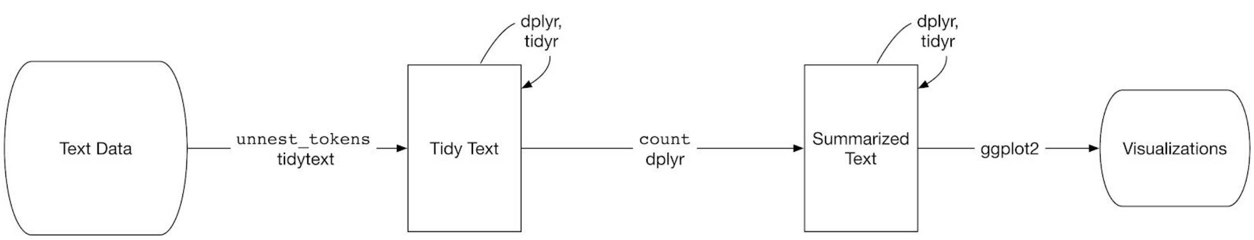
The Twitter data for this report was collected using the Python library GetOldTweets3. This Python library allows users to scrape Twitter via its browser-based "Advanced Search" function, instead of using the official API which only allows users to request tweets from the past 7 days. GetOldTweets3 allowed us to specify the search parameters of keyword, timeframe, and location.

We searched tweets from users in the United States mentioning the six top-performing candidates in the Democratic primary: former Vice President Joe Biden, Senator Bernie Sanders, Senator Elizabeth Warren, Senator Amy Klobuchar, Mayor Pete Buttigieg, and Michael Bloomberg. We obtained a total of 60,000 tweets per candidate. We assume that time is an explanatory variable in Twitter sentiment, given changes in polling, media coverage, debates, and other fluctuations in the trajectory of the race. To control for time in the data collection process, we planned to obtain 1,000 tweets per day over the same 60 day period. However, in practice, we were forced to stagger collection, so we standardized the time window across candidates after collection.

Data Cleaning & Pre-processing

Text cleaning is an essential preprocessing step in text analysis which involves removing extraneous components in the text data before tokenizing it. Our raw Twitter data contains URLs, retweet entities, quotes, punctuations, white spaces, stop words, etc. - because they typically add noise and provide little information relevant to sentiment analysis, they were removed. Afterwards, the cleaned Twitter data was tokenized, i.e. turned into keywords for use in sentiment analysis. In this project, we adopted a tidy text workflow for text cleaning and analysis as presented below.

Figure 1.1: A flowchart of a typical text analysis using tidy data principles.



Note: Reprinted from *Text mining with R: A tidy approach,* by Silge, J., & Robinson, D. Retrieved from https://www.tidytextmining.com/tidytext.html

Prior to the cleaning process, we extracted tweets for each candidate produced during the same 56-day time frame from 1/7/2019 to 2/26/2020, since the raw Twitter data for each candidate covered slightly different time windows. We then proceeded by cleaning the data, i.e. removing all unnecessary elements, including URLs, retweet marks, Twitter handles, quotation marks, punctuations, and white spaces by using the `gsub` function. The next step was to parse and tokenize the tweets in order to take out all the stop words and single letters (e.g. “a”, “how”, “and”, etc.) using the `tidytext` package. Finally, we wrapped up preprocessing by removing numbers from all tweets. After removing all extraneous components, tokenized words were re-merged into their corresponding individual tweets/sentences.

RESULTS

Sentiment Analysis

We used an R package called `sentimentr` to conduct sentiment analysis. Many researchers have demonstrated that `sentimentr` is the only sentiment analysis package that properly accounts for negators.[9] `sentimentr` is designed to efficiently calculate text polarity sentiment at the sentence level, while allowing sentiment to be aggregated via higher-level grouping variables. It also utilizes an augmented dictionary lookup, taking into account "valence shifters": words and phrases that affect the polarization by negating, amplifying, or de-amplifying the polarity of other words.

Using the core scoring function `sentiment()` with the Jockers-Rinker and WordSentiNet sentiment lexicons, we wrote a custom function which obtained sentiment scores for each unique tweet in our data set. Both sentiment lexicons use semi-supervised learning classifiers; randomly selected words were manually labelled with a sentiment category or value. The algorithms assign a sentiment score based on the polarity of each sentence by searching and comparing words in each sentence to the sentiment dictionary, before calculating the overall polarity for each sentence (or tweet). The sentiment scores are assigned on a scale from -5 to 5 (although none of the tweets had a score below -3 or above 3). Along with the Jockers-Rinker sentiment lexicon, we also calculated sentiment scores using the WordSentiNet 3.0 sentiment dictionary, which

---

[9] Naldi, M. (2019). A review of sentiment computation methods with R packages. *arXiv preprint arXiv:1901.08319*.

assigns scores ranging from -1 to 1. By re-scoring the data with another sentiment lexicon, we hope to cross-validate the polarity scores.

After generating sentiment scores for each tweet, we plotted the distributions of tweet sentiments per candidate. The density plots of distribution are presented below.

Figure 2.1: Jockers-Rinker, Plots of candidate-level sentiment score distributions.

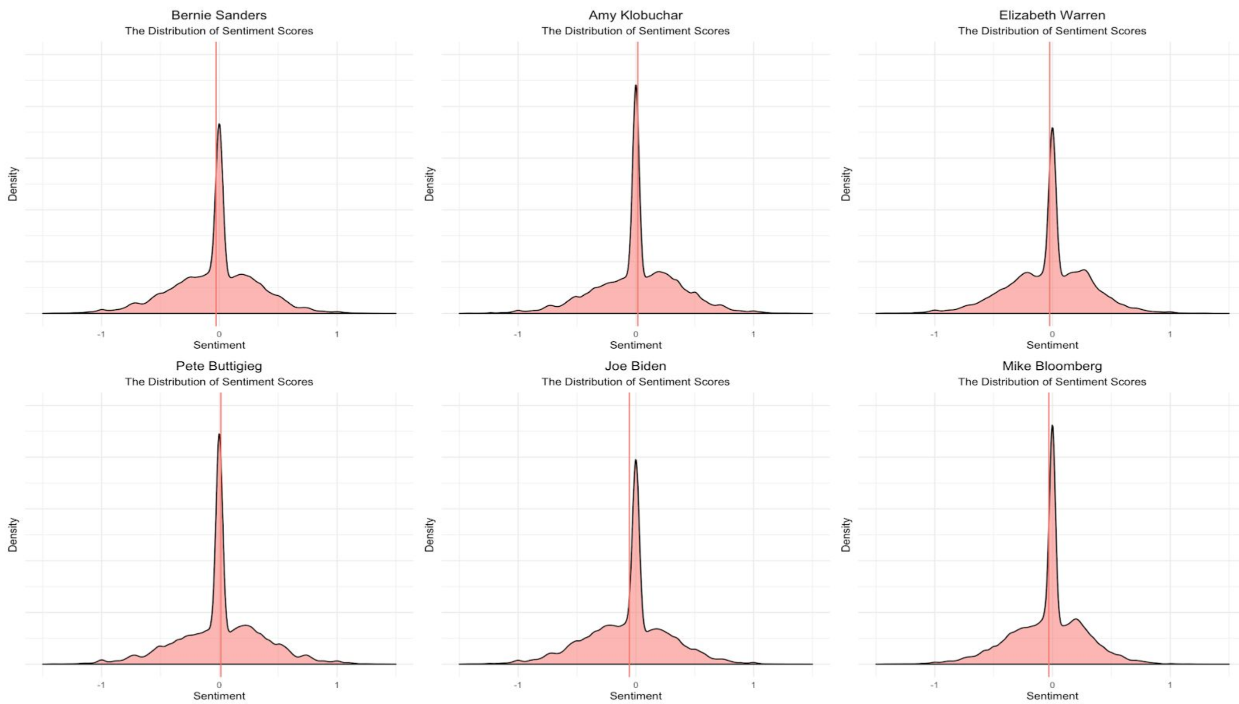
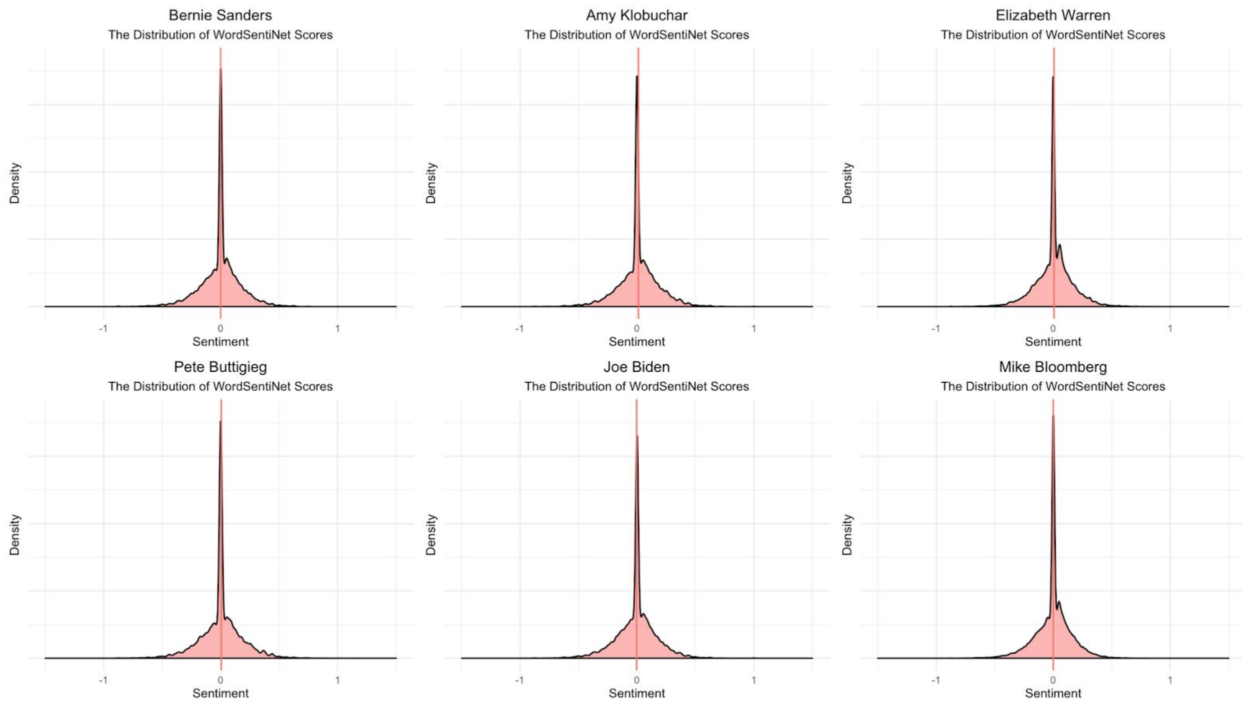


Figure 2.2: WordSentiNet 3.0, Plots of candidate-level sentiment score distributions.



We can observe that all ten density distributions take roughly the same shape, with a large majority of tweets clustering around a score of zero and long tails. However, the means appear to be slightly different. For the Jockers-Rinker scores, the Buttigieg and Klobuchar mean scores are slightly above zero while Warren, Biden, Sanders, and Bloomberg mean scores are slightly below zero. For the WordSentiNet 3.0 scores, the Buttigieg, Klobuchar, and Warren mean scores are slightly above zero while Biden, Sanders, and Bloomberg mean scores are slightly below zero. We conducted a number of statistical tests to determine the significance of these results.

First, we conducted one-sample t-tests on each candidate's Jocker's and WordSentiNet 3.0 score distributions to determine if we could reject the null hypothesis that the mean score is equal to zero for either lexicon. For all t-tests except for Mike Bloomberg's mean WordSentiNet 3.0 score, the null hypothesis was rejected.

Second, we conducted one-way ANOVA tests on both sets of mean sentiment scores--Jocker's and WordSentiNet 3.0--to determine if there are any significant differences between the mean scores for each candidate. For both ANOVA tests, the reported p-values allowed us to reject the null hypothesis that the means were equal at alpha = 0.01.

Third, to determine which pairwise differences in mean sentiment scores are significant, we conducted a multiple pairwise comparison between the mean scores of the candidates using a Tukey Honest Significant Differences (HSD) test. The tables below highlights the pairs for which there is a significant difference in means.

Table 1: Tukey multiple comparisons of means (Jockers-Rinker)

Candidates compared*	**Difference between mean scores**	**Lower CI bound (CL = 0.95)**	**Upper CI bound (CL = 0.95)**	**P-value**
Bloomberg-Biden	0.0229478930	0.016800422	0.029095364	0.0000000
Buttigieg-Biden	0.0666119764	0.060339081	0.072884872	0.0000000
Klobuchar-Biden	0.0709234522	0.064678884	0.077168021	0.0000000
Sanders-Biden	0.0267798098	0.020535213	0.033024407	0.0000000
Warren-Biden	0.0275775645	0.021365988	0.033789141	0.0000000
Sanders-Klobuchar	-0.0441436424	-0.049957876	-0.038329409	0.0000000
Bloomberg-Klobuchar	-0.0479755592	-0.053694728	-0.042256391	0.0000000
Buttigieg-Klobuchar	-0.0043114758	-0.010153393	0.001530442	0.2595264
Warren-Klobuchar	-0.0433458877	-0.049127810	-0.037563965	0.0000000
Bloomberg-Sanders	-0.0038319167	-0.009551114	0.001887280	0.3574810

Buttigieg-Sanders	0.0398321667	0.033990221	0.045674112	0.0000000
Warren-Sanders	0.0007977548	-0.004984196	0.006579706	0.9957436
Buttigieg-Bloomberg	0.0436640834	0.037916745	0.049411422	0.0000000
Warren-Bloomberg	0.0046296715	-0.001056674	0.010316017	0.1718745
Warren-Buttigieg	-0.0390344119	-0.044844200	-0.033224623	0.0000000

Note: 95% family-wise confidence level, Yellow highlight indicates significance

Table 2: Tukey multiple comparisons of means (WordSentiNet 3.0)

Candidates compared*	Difference between the mean scores	Lower CI bound (CL = 0.95)	Upper CI bound (CL = 0.95)	P-value
Bloomberg-Biden	0.0026546236	-0.0002792749	0.005588522	0.1025298
Buttigieg-Biden	0.0081639697	0.0051702121	0.011157727	0.0000000
Klobuchar-Biden	0.0160839540	0.0131037157	0.019064192	0.0000000
Sanders-Biden	0.0017591464	-0.0012211057	0.004739399	0.5436848
Warren-Biden	0.0091343589	0.0061698658	0.012098852	0.0000000
Sanders-Klobuchar	-0.0441436424	-0.049957876	-0.038329409	0.0000000
Bloomberg-Klobuchar	-0.0479755592	-0.053694728	-0.042256391	0.0000000
Buttigieg-Klobuchar	-0.0043114758	-0.010153393	0.001530442	0.2595264
Warren-Klobuchar	-0.0433458877	-0.049127810	-0.037563965	0.0000000
Bloomberg-Sanders	-0.0038319167	-0.009551114	0.001887280	0.3574810
Buttigieg-Sanders	0.0398321667	0.033990221	0.045674112	0.0000000
Warren-Sanders	0.0007977548	-0.004984196	0.006579706	0.9957436
Buttigieg-Bloomberg	0.0436640834	0.037916745	0.049411422	0.0000000
Warren-Bloomberg	0.0046296715	-0.001056674	0.010316017	0.1718745
Warren-Buttigieg	-0.0390344119	-0.044844200	-0.033224623	0.0000000

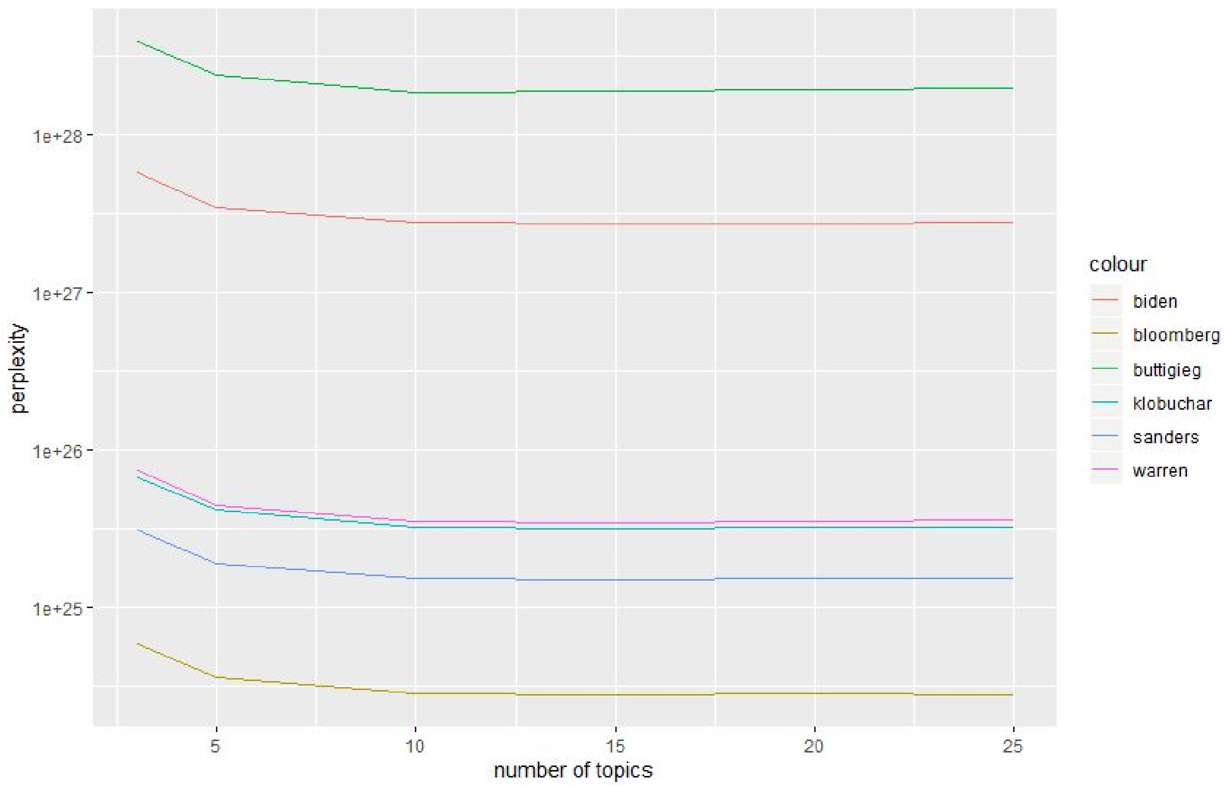
Note: 95% family-wise confidence level, Yellow highlight indicates significance

We can draw a few observations from the statistical results. First, the general clustering of scores around zero may suggest that Twitter users are not particularly polarised in their opinions of the various 2020 Democratic candidates. However, there is a statistically significant difference between candidates with positive mean sentiment scores versus negative mean scores, indicating there are measurable valences to the Twitter discussions of individual candidates. Finally, we note that the different dictionaries produce different results, suggesting that the choice of dictionary is important to the validity of the results.

Topic Modeling

Although the focus of this project is to examine the polarity of the political landscape via sentiment analysis, we have also performed some exploratory topic modeling on the Twitter data. We wanted to see if we could observe semantic context in the sentiments we were examining, and whether we could discern any interesting semantic patterns and commonalities among the various candidates being discussed. Towards this end, topic modeling was performed at both $k = 5$ and $k = 15$ using Latent Dirichlet Allocation, treating tweets surrounding/mentioning each candidate as a separate corpus. More specifically, all tweets referring to one candidate would be treated as one large body of text, and each candidate would have 5 (or 15) “topics” associated with them. We chose $k = 5$ because it was a balance between having a small enough number of topics that allows for a quick overview of the data while having an acceptable amount of perplexity, and we chose $k = 15$ because it had the best overall perplexity among all six candidates.

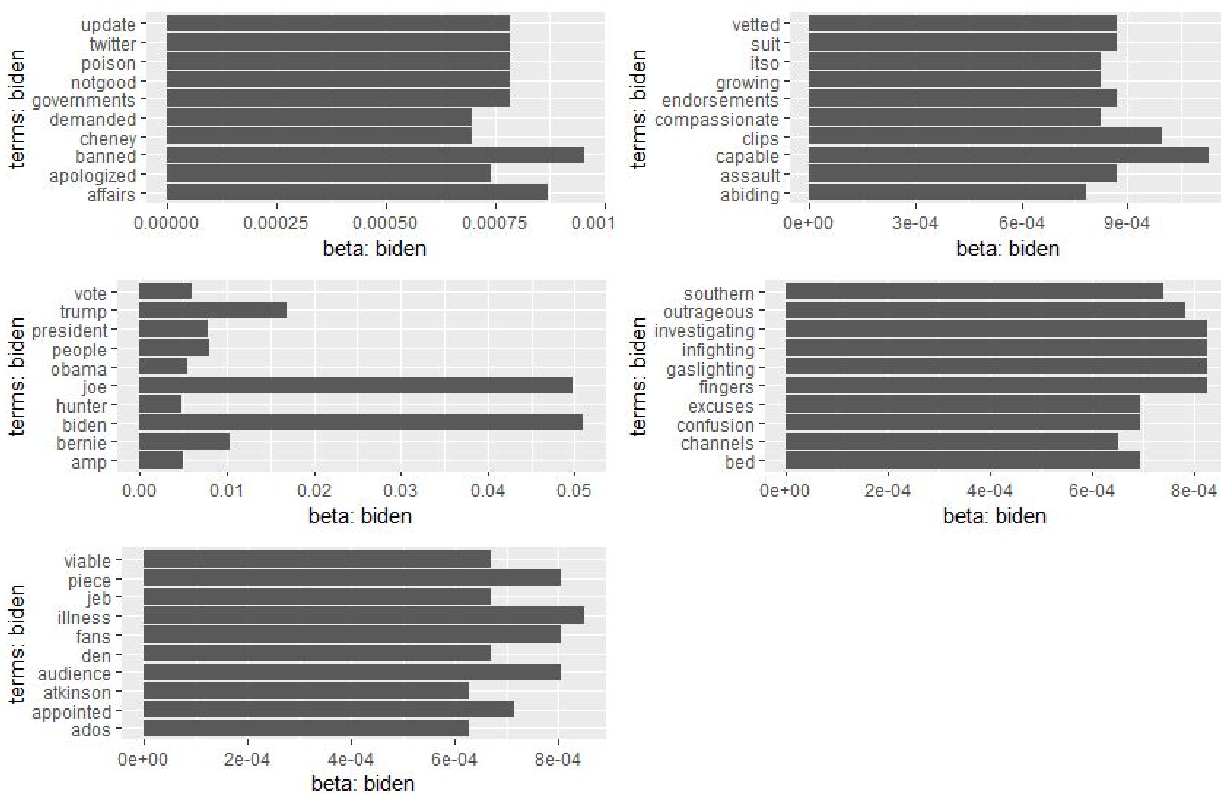
Figure 3.1: Perplexity plot of all candidates.



Note: See appendix for perplexity plots of each individual candidate; these have a better resolution, making it clearer that $k = 15$ had better perplexity between the candidates than $k = 10$ or $k = 20$.

Unsurprisingly (given how the Twitter data was fetched), the terms with the highest betas (i.e. likelihoods) were the names of the respective candidates - e.g., the terms, "bernie" and "sanders" would have the highest beta values for tweets corresponding to Sanders. All other terms have beta values that are at least one order of magnitude smaller. Additionally, the topic that involved the candidate's own name also typically involved the names of other candidates or politicians (e.g. the topic for Elizabeth Warren which mentions "elizabeth" and "warren" also mentions "trump", "sanders", "biden", etc.). Few other topics for any of the candidates had the same degree of indicating a clear-cut pattern, or exhibiting a cohesive topic. We have found these observations to be consistent between LDA at $k = 5$ and LDA at $k = 15$. (An uncommon example of a candidate with more than one cohesive topic is Biden, which includes "poison", "notgood", "banned", "apologized", etc., which may indicate Twitter users' negative reception of the candidate. An example of a non-cohesive topic under Klobuchar which has terms too vague to be helpful: "stress", "prime", "libs", "blow", "glad", etc.)

Figure 3.2: LDA output of one candidate (Biden) at *k* = 5.



At a more general level, we can make two indications about the use of Twitter in expressing thoughts over American political events. The first, specific to the current Democratic nominations, is that the focus of discussion appears to be most heavily and commonly placed on candidates and their relations to other important political figures, rather than on candidates and their particular positions, issues, opinions, etc. The second is that, due to the fragmentary nature of Twitter (where millions of *different* users contribute millions of *short* texts when referring to a person or idea), it can be difficult to discern a sustained and cohesive discussion with distinct patterns and topics at a *general* level. While it may be possible to find such patterns and topics with, for example, specific Twitter users serving as the unit of analysis, we were not able to find clear patterns of discourse at the general level.

DISCUSSION: LIMITATIONS & FUTURE RESEARCH

There are some limitations to our sentiment analysis. To begin with, it is important to note that Twitter isn't necessarily representative of public opinion. Assuming surveys are representative of public opinion, the sentiment of Twitter users is more concentrated towards extremes than the sentiment of general survey respondents.[10] However, Twitter is an accurate indicator of which issues are most salient to individuals. Thus, though our results are significant, they can only be stated to represent the sentiment of U.S.-based Twitter users and not the sentiments of any broader populations in the United States.

Though `sentimentr` does a better job than most in accounting for the context surrounding polarizing words rather than just the words themselves, there are contextual nuances it doesn't adjust for, such as cultural nuances. For example, if one tweets "mayo Pete" when discussing Buttigieg or adds snake emojis after mentioning Warren, `sentimentr` would not assume the word "mayo" or snake emojis to be derogatory. However, if one understands the cultural and political context, they would understand that that tweet conveys a negative sentiment. Sarcasm,

[10] Alamos , F., & Ganesh, B. (2019). Is Twitter a Proxy for Public Opinion?, 1–22.

likewise, falls into this category - it can be recognized by a reader who knows the context, but may not be picked up in the sentiment score.

One final notable limitation is the specific sentiment dictionary (Jocker's) used within the `sentimentr` package. Jocker's dictionary consists of 11709 words, each given one of 19 values weighted on a scale from [-2, 1]. Jocker's dictionary is commonly accepted as the best and most comprehensive dictionary for sentiment analysis, however, it should be noted that it is still limited to only the words within the dictionary. Thus, inevitably, our sentiment analysis will be restricted by the words within the dictionary. In an effort to account for this limitation, we utilized a second sentiment dictionary (WordSentiNet). However, this dictionary is even more limited than Jocker's, and thus, the limitation still stands.

One potential avenue for future research would be to examine trends in Twitter sentiment as a result of political events. A political event such as a rally or a debate could induce a change in sentiment trends wherein related political figures may be seen more positively or negatively as a result. It would be a worthwhile investigation to have a means by which one can pinpoint a particular fluctuation or trend in sentiment to a particular political event. Additionally, if said event had involved a speech or conversation, said words could be analyzed for their content in order to discern what kinds of words or topics had led up to the political outcome (e.g. the sentiment trend which was the outcome of that speech). This would have practical applications for the outcome of political campaigns.

CONCLUSION

This paper uses sentiment analysis and basic topic modeling to explore differences between sentiment in tweets pertaining to the 2020 Democratic primary candidates. The results are based on data sets of 60,000 tweets per candidate, collected over a sixty-day interval with uniform amounts of tweets per day. Overall, our sentiment distributions at the candidate-level clustered around zero, suggesting a lack of sentiment polarity in the Twitter conversations about any of the candidates. We found significant differences in mean sentiment scores between the set of candidates with average positive scores and the set of candidates with average negative scores. As for topic modeling, we found that much of the tweets had focused on each candidate relative to their competitors, but that there was otherwise no clear and coherent topic sustaining discussion about any particular candidate.

REFERENCES

Alamos , F., & Ganesh, B. (2019). Is Twitter a Proxy for Public Opinion?, 1–22.

Bakliwal, A., Foster, J., Puil, J.V., O'brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier.

Bermingham, A., & Smeaton, A.F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results.

Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alignment of Twitter Users. *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 192-199.

Feng, S., Wang, D., Yu, G., Yang, C., & Yang, N. (2009). Sentiment Clustering: A Novel Method
to Explore in the Blogosphere. *APWeb/WAIM*.

Hernandez-Suarez, A., Sanchez-Perez, G., Martinez-Hernandez, V., Meana, H.M., Toscano-Medina, K., Nakano-Miyatake, M., & Sanchez, V. (2017). Predicting political mood tendencies based on Twitter data. *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, 1-6.

Hernandez-Suarez, A., Sanchez, V., Sanchez-Perez, G., Pérez, H., Olivares, J., Toscano-Medina, K., Nakano-Miyatake, M., & Martinez-Hernandez, V. (2016). Security attack prediction based on user sentiment analysis of Twitter data. *2016 IEEE International Conference on Industrial Technology (ICIT)*, 610-617.

Laboreiro, G., Sarmento, L., Teixeira, J., & Oliveira, E.C. (2010). Tokenizing micro-blogging messages using a text classification approach. *AND '10*.

Raghuwanshi, A.S., & Pawar, S. (2017). Polarity Classification of Twitter Data using Sentiment Analysis.

Shamma, D.A., Kennedy, L., & Churchill, E.F. (2009). Tweet the debates: understanding community annotation of uncollected sources. *WSM ‘09*.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level

Sentiment Analysis. *HLT/EMNLP*.

APPENDIX

All codes, tables, plots, and data files for our project can be found at:
https://github.com/minyoungdo/mlgroup