Understanding Twitter's Role in the 2018 Midterm Election

Katie Ta, Charles Tan

Goal

Our goal is to find out how Twitter plays a role in US politics, specifically in the 2018 midterm elections. Looking at the results of the primary elections and tweets from the respective candidates during their campaign, we used topic modeling and sentiment analysis to determine if a relationship exists between winning or losing, Tweet sentiment, and Tweet topic frequency. We hope that the absence or existence of any type of relationship will offer insights as to what type of Tweets lead, or do not lead, to electoral success. Additionally, we wanted to see if there were significant differences between Republican and Democratic candidates and their Twitter behaviors.

Methodology & Results

Data Collection

In order to examine the outcomes of the 2018 midterm elections, we used the FiveThirtyEight primary-candidates-2018 dataset to determine who the candidates and winners were running for each office type (Governor, Senator, and Representative) in each state.[1] A "successful" election is measured by an advancement through the primary election, which in the data, corresponds to the "won_primary" column. The dataset separates candidates by party, Democratic or Republican, and includes the name of the candidate, the office they're running for, and their primary result among other information we choose ignore. There are 812 Democratic candidates and 775 Republican candidates in the dataset.

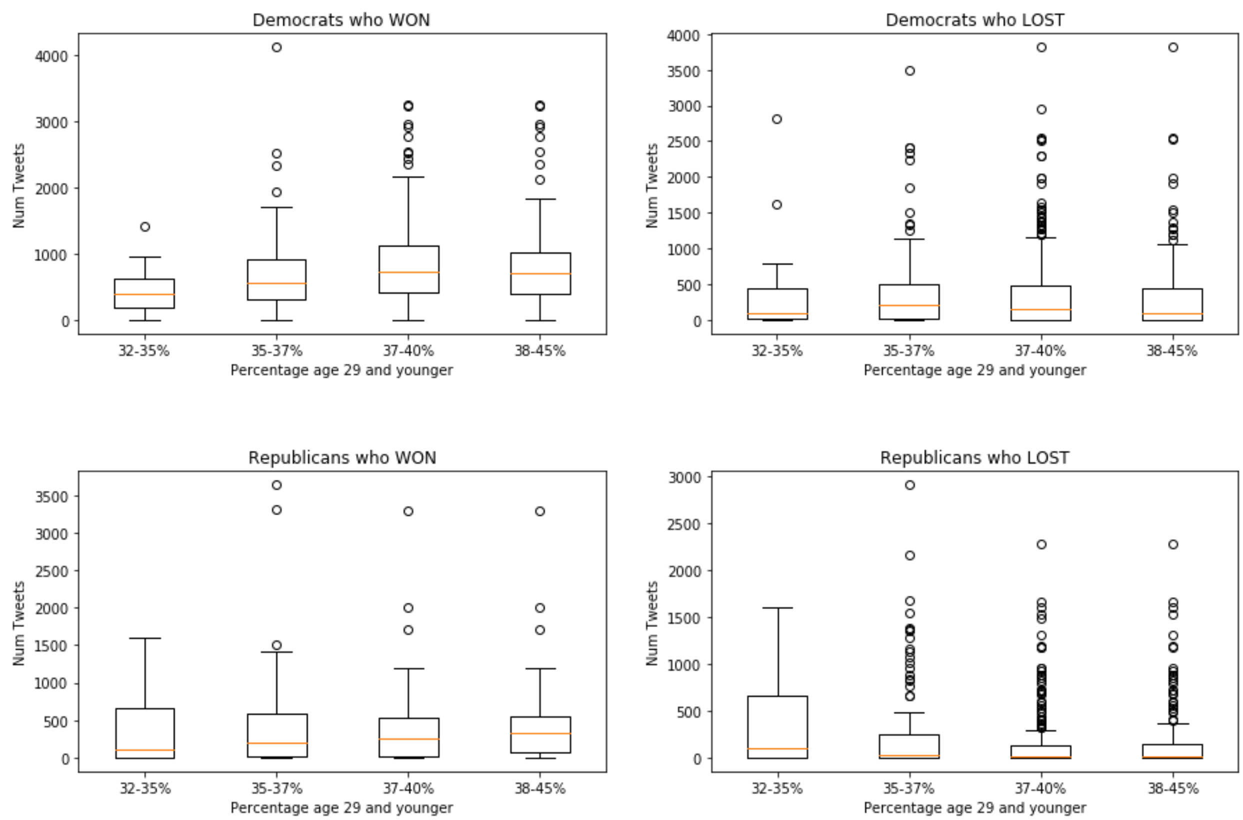
To gather the tweets per candidate, we manually searched for each candidate's Twitter handles, including up to two Twitter handles if they had multiple accounts. The Twitter API limits tweet searches to tweets up until one week prior, so we used a public Tweet scraper available on Github to search for the tweets per candidate.[2] We gathered all tweets per candidate from the 365 day period leading up to the election date on November 6 2018. In total, we collected 690,588 tweets, with 439,335 tweets from Democratic candidates and 251253 from Republican candidates. It's important to note that the corpus didn't include Retweets unless the user Retweeted **and** added their own text as commentary.

Finally, we removed all data with candidates who ran uncontested as that their Tweeting habits would add unnecessary noise and false information to our results. After

[1] https://github.com/fivethirtyeight/data/tree/master/primary-candidates-2018
[2] https://github.com/Jefferson-Henrique/GetOldTweets-python

removing these tweets, we had 557,549 total tweets, 387,581 of which were Democratic and 169,968 Republican. For the sake of consistency, the Tweet corpus we refer to throughout this write-up excludes all uncontested candidate tweets.

We also found a dataset that included demographic data per county, such as percentage of individuals under 29 years old, median household income, rural percentage, racial breakdowns, etc. However, because candidate information is segmented by state for Governor and Senate candidates and by District for Representatives, we were only able to aggregate state demographic data from our dataset. More specifically, we wanted to look at how total number of tweets is influenced by the portion of young people in a candidate's target voter audience. We found that candidates who won, both Republican and Demographic, tweeted more if a larger percentage of their state was age 29 and younger.



Sentiment Analysis

We perform sentiment analysis on the tweet data we've collected in order to score the general emotion each candidate and each candidate's tweets have. To perform this analysis, we train a model to predict the sentiment of tweets by utilizing the Sentiment140 dataset[3] and Google's BERT pre-trained model[4].

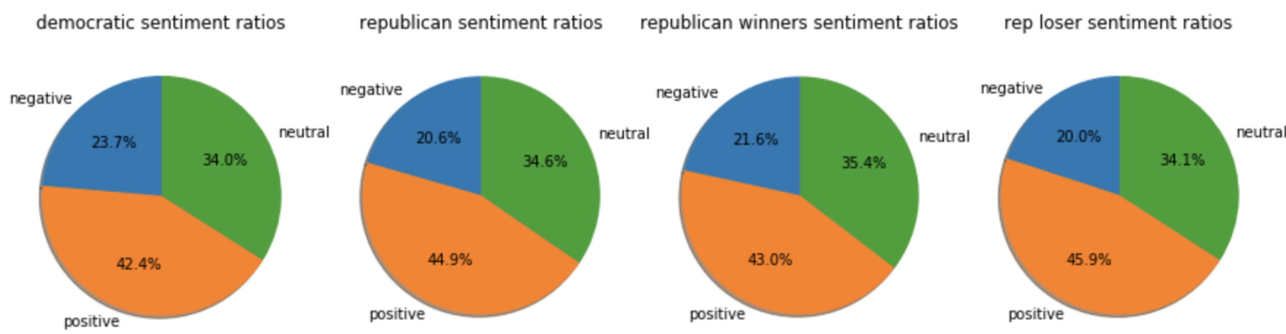
---

[3] http://help.sentiment140.com/for-students/
[4] https://github.com/google-research/bert

The Sentiment140 dataset includes 1.6 million labeled tweets, 1.4 million used for training and 200 thousand used for testing. The labels represent the sentiment for a particular tweet, 0 for a negative tweet and 4 for a positive tweet. Sentiment140 was created autonomously through machine learning techniques and by using emojis as heuristics to label tweets as positive or negative. We chose to use this dataset due to the dataset's size and simplicity in labeling tweets as either positive or negative.

Google's BERT pre-trained model is a language representation that has been trained on Wikipedia (2,500 million words) and BooksCorpus (800 million words). The pre-trained model can be fine tuned to perform sentiment analysis by adding an additional layer to the model. This outer layer consists of training the model further on the Sentiment140 data for 5 epochs. Using this method, we were able to predict the sentiment of tweets at with 87.55% accuracy.

Our fine tuned BERT model returns probabilities that the Tweet should be labeled as positive or negative. Using a 70% threshold for these probabilities (if the logits returned had a probability of 70% for positive or negative, then the Tweet would be labeled as positive or negative respectively, if the probabilities were both less than 70% then the Tweet would be labeled neutral), we found our data had the following distribution:



Unpacking our results further, we can see qualitatively that our model performs much better for deciphering negative tweets than it does for positive tweets.

Here are 5 random tweets which have a 0.7 to 0.8 probability of being negative:

1. `I just read some of the claims made by @JamesADamore toward google. The fact that Google managers are actually bias against conservative or libertarian voices is disgusting. @sundarpichai and his company are the newest threat to American Freedom.`
2. `My Democratic opponents have turned this race into everything thats wrong w/ politics, & none of it is helping struggling families. Heres my pledge: Ill keep our budget balanced, invest in affordable housing, fight for universal health care, & stand up to @realDonaldTrump.`
3. `Destroying #NetNeutralty on Dec.16 Circa March 2017: White House spokesman Sean Spicer said that next up on President Trumps telecom agenda is to roll back the FCCs 2015 Open Internet net neutrality rules. @AjitPaiFCC @potus #BlueWave2018`

4. `Thank you. We have 22 vet's a day taking their own life. When elected to Congress, I'll work to stop this trend and to ensure the services our Vet's need are there. Too much sacrifice and blood for these men & women to be ignored.`
5. `I also support legislation that would ban bump stocks, a device that enables mass shooters to increase their rate of fire that was used in the Las Vegas mass killing last year. We also must concentrate on improving preventive measures such as mental health and criminal history.`

Here are 5 random tweets which have a 0.7 to 0.8 probability of being positive:

1. `Bolton is a very dangerous choice. His hawkish views may add fuel to @realDonaldTrump's delusions of unbridled power.`
2. `People are more important than things.`
3. `Surely as a lawyer you know something about the nuances of libel.`
4. `When a responder responds, it is in that very instant that the entire situation is defined. Our goal is to make sure responders have the tools to appropriately respond so the situation doesn't escalate. #thinkDIFFERENTLY`
5. `The Koch bros supported apartheid in South Africa. They also funded Reason, which dedicated an entire edition to Holocaust denial. They are the Louis Farrakhan of the right. And they are far more powerful.`

We calculate sentiment score per candidate as being the average of all their Tweet's sentiment values. However, based on our uneven distribution of positive and negative tweets and the ambiguity in our model of deciphering positive from neutral tweets, we decided to calculate sentiment as a negativity value. So, tweets with a greater than or equal to 0.7 probability of being negative as produced from the BERT are labeled as 1, and all other tweets are labeled as 0. Since our fine tuned BERT predicts negative tweets more precisely and ambiguity was introduced in determining a threshold between neutral and positive tweets, we found the negativity method to be less noisy than our original sentiment calculation.

Topic Model

After tokenizing and lemmatizing the Tweet corpus, we used the gensim library to create a bag of words model to use in the topic model.[5]

In running the topic model a few times, we found that there were a significant number of tweets that didn't have definitive topics that may have been convoluting our topic results, giving us nonsense topics. We decided to filter out tweets with fewer than 4 words to remove noise, which produced reasonable, decipherable topics. We also attempted running the topic

---

[5] https://radimrehurek.com/gensim/

model on random subsets of the corpus to see if we could also produce a diverse set of topics. Ultimately, we found that the corpus of tweets with > 4 words produced better topics than the random subsets. Successful results were measured on whether they represented actual topics or categories. For example, `('0.072*"thanks" + 0.069*"great" + 0.036*"morning" + 0.024*"ready"')`, would be considered a poor result, while `('0.039*"school" + 0.039*"healthcare" + 0.035*"child" + 0.028*"public"')` pointed at comprehensive topics such as education and healthcare.

Below are results from combined and separated Republican and Democratic topic models, with most decipherable topics highlighted.

Top 15 Republican & Democratic Tweet Topics, Tweets > 4 words

```
(0, '0.062*"thank" + 0.058*"campaign" + 0.052*"congress" + 0.052*"support"')
(1, '0.068*"health" + 0.037*"vote" + 0.032*"forward" + 0.028*"world"')
(2, '0.030*"business" + 0.029*"enough" + 0.027*"economy" + 0.027*"worker"')
(3, '0.073*"change" + 0.034*"action" + 0.030*"violence" + 0.022*"problem"')
(4, '0.107*"candidate" + 0.045*"debate" + 0.031*"forum" + 0.030*"class"')
(5, '0.054*"woman" + 0.038*"family" + 0.025*"honor" + 0.025*"proud"')
(6, '0.039*"school" + 0.039*"healthcare" + 0.035*"child" + 0.028*"public"')
(7, '0.045*"voter" + 0.044*"county" + 0.033*"today" + 0.023*"voting"')
(8, '0.074*"trump" + 0.050*"democrat" + 0.041*"republican" + 0.033*"president"')
(9, '0.072*"thanks" + 0.069*"great" + 0.036*"morning" + 0.024*"ready"')
(10, '0.667*"SCREEN_NAME" + 0.035*"virginia" + 0.015*"proud" + 0.010*"dreamer"')
(11, '0.078*"texas" + 0.063*"people" + 0.033*"money" + 0.032*"would"')
(12, '0.043*"please" + 0.040*"senate" + 0.031*"social" + 0.028*"really"')
(13, '0.049*"election" + 0.049*"check" + 0.044*"primary" + 0.027*"watch"')
(14, '0.046*"democratic" + 0.026*"call" + 0.025*"administration" + 0.023*"office"')
```

<u>Top 15 Democratic Tweet Topics, Tweets > 4 words (365,207 tweets)</u>

```
(0, '0.062*"congress" + 0.041*"right" + 0.039*"trump" + 0.032*"country"')
(1, '0.038*"issue" + 0.030*"state" + 0.026*"community" + 0.021*"business"')
(2, '0.086*"healthcare" + 0.047*"health" + 0.031*"child" + 0.027*"spend"')
(3, '0.056*"people" + 0.035*"voting" + 0.033*"election" + 0.033*"money"')
(4, '0.067*"would" + 0.027*"elect" + 0.020*"things" + 0.019*"people"')
(5, '0.589*"SCREEN_NAME" + 0.060*"texas" + 0.021*"proud" + 0.016*"great"')
(6, '0.043*"family" + 0.039*"washington" + 0.038*"friend" + 0.036*"voice"')
(7, '0.058*"campaign" + 0.055*"tonight" + 0.047*"night" + 0.034*"debate"')
(8, '0.087*"every" + 0.039*"endorsement" + 0.025*"receive" + 0.020*"person"')
(9, '0.081*"change" + 0.040*"never" + 0.039*"register" + 0.036*"voter"')
(10, '0.050*"great" + 0.038*"district" + 0.037*"today" + 0.035*"county"')
(11, '0.144*"support" + 0.114*"thank" + 0.040*"please" + 0.027*"campaign"')
(12, '0.041*"school" + 0.039*"woman" + 0.030*"student" + 0.030*"public"')
(13, '0.057*"years" + 0.054*"still" + 0.043*"veteran" + 0.038*"violence"')
(14, '0.036*"candidate" + 0.033*"party" + 0.029*"question" + 0.029*"running"')
```

<u>Top 15 Republican Tweet Topics, Tweets > 4 words (159,538 tweets)</u>

```
(0, '0.036*"border" + 0.030*"healthcare" + 0.028*"school" + 0.026*"child"')
(1, '0.027*"record" + 0.026*"change" + 0.022*"things" + 0.021*"reason"')
(2, '0.038*"people" + 0.024*"government" + 0.023*"money" + 0.021*"problem"')
(3, '0.030*"illegal" + 0.029*"immigration" + 0.024*"talking" + 0.024*"liberal"')
(4, '0.055*"campaign" + 0.043*"please" + 0.034*"debate" + 0.031*"tomorrow"')
(5, '0.066*"senate" + 0.052*"support" + 0.047*"conservative" + 0.040*"state"')
(6, '0.041*"question" + 0.041*"could" + 0.040*"folks" + 0.029*"agree"')
(7, '0.087*"candidate" + 0.066*"republican" + 0.035*"party" + 0.035*"congress"')
(8, '0.057*"election" + 0.040*"primary" + 0.040*"vote" + 0.039*"democrat"')
(9, '0.091*"great" + 0.074*"thank" + 0.037*"county" + 0.034*"today"')
(10, '0.057*"senator" + 0.035*"national" + 0.031*"important" + 0.027*"security"')
(11, '0.105*"president" + 0.096*"trump" + 0.053*"happy" + 0.025*"remember"')
(12, '0.058*"right" + 0.045*"american" + 0.039*"america" + 0.036*"freedom"')
(13, '0.038*"spending" + 0.025*"going" + 0.024*"years" + 0.022*"amaze"')
(14, '0.694*"SCREEN_NAME" + 0.032*"texas" + 0.017*"interview" + 0.014*"video"')
```

Based on these results, the topic model identified healthcare, business / economy, violence, school / education, and Trump to be common topics shared between Democratic and Republican candidates. Interestingly, "Texas" was also identified as a common topic, which could possibly be because of the large number of Texan candidates or because Texas had one of the largest, closest Senate races this year.[6]

---

[6] https://www.nytimes.com/elections/results/texas-senate

While the shared Democratic and Republican topic words were rather broad, we could see how these topics were more closely addressed in the separate results for the Republican and Democratic candidates. For example, “violence” appeared as a shared topic, however, Republican topics related to violence included “illegal immigration” and “border” coming up in the Republican results with few related topics coming up in Democratic results. Regarding “education,” Democrats had topics including “school” and “public” while Republicans had no relevant education topics. Top of all results, however, was “healthcare” and “trump,” clearly divisive topics, but common ground nonetheless, for both parties.

More generally, candidates both seemed to Tweet about “change”, “voting”, “campaigns”, “debates”, and many “thank you’s”, which point to Twitter being used as a platform primarily to promote their campaign and make announcements and perhaps less so for sharing opinions and ideas.

Topic Classification

To combine our findings from our sentiment analysis and topic modeling, we created a topic classification model to label our corpus with topics. By labeling each tweet with its respective topics, we could segment sentiment and tweet activity by topic.

Using the results from our topics model, we identified topics we to label in addition to other topics we were interested in seeing results for. Our list of topics included:

- gun control
- health care
- kavanaugh
- jobs
- education
- abortion
- environment
- trump
- immigration
- freedom
- lgbt
- women

We used a logistic regression analysis to incrementally build up a reliable set of topic labels. After tokenizing and lemmatizing the tweet corpus, our labelling methodology per topic went as follows:

1. Gather & label tweets that include explicitly relevant keywords
    a. i.e. for “gun control” we labeled tweets that contained the words gun control
    b. The point of this step was to build a small dataset from which the Logistic Regression could learn to label tweets with topics
2. Create a train/test dataset on which the model could train
    a. Combine previously labeled tweets containing the topic and an equal number of tweets not containing the topic into dataset

    b. Split into 90% training data, 10% testing data
3. Train the model & evaluate the features, or words in this case, with the highest coefficients
    a. pick out observably relevant words with the highest coefficients and add them to set of keywords for labeling
    b. the words with the highest coefficients are essentially keywords that the model finds related to the topic & keywords
4. Run predictions on the remaining tweets not selected for the train/test dataset
5. Determine a probability confidence threshold for which to add the predicted labels to the train/test dataset
    a. do so by randomly sampling predicted labeled tweets & determining whether or not the labels are correct
6. Add newly labeled tweets to dataset & repeat until no new tweets can confidently be added to labeled dataset

In all, we found the following topics to be the most popular:

Republican Candidates

topic	avg_count	total_classified
trump	19.53	13818
immigration	6.15	4339
jobs	5.74	4124
gun_control	4.23	3086
health_care	3.77	2956
freedom	3.48	2418
education	2.47	1738
abortion	1.96	1405
kavanaugh	1.67	1180
environment	0.56	435
women	0.21	157
lgbt	0.08	74

Democratic Candidates

topic	avg_count	total_classified
trump	28.24	21582

health_care	23.49	17831
gun_control	15.01	11455
jobs	13.25	10071
education	9.55	7292
environment	7.65	5855
immigration	7.04	5379
kavanaugh	2.74	2088
lgbt	2.39	1826
freedom	2.16	1651
women	2.03	1548
abortion	1.93	1474

After attempting to label as many tweets as possible under our desired topics, we found that we were unable to label a significant number of tweets for a few of the topics and decided to omit them from our dataset altogether. Overall, topic frequencies matched those in our topic model, but gave a more specific ranking of importance to both parties. While both parties tweeted the most about Trump, healthcare, and gun control, and jobs, but more notably, Republicans seemed to tweet more about immigration and abortion (proportionally) than Democrats.

We also found that certain topics were difficult to label, namely gun control, lgbt, and women. While trying to label these topics, we found that certain auxiliary topics were being wrongly associated with the topics we were trying to label. For example, any tweet speaking of violence in general was labeled with gun control as a topic, most likely because many of the tweets regarding gun control were related to the various shootings that happened during the year. At the same time, lgbt and women/gender rights were being associated with each other, with tweets about feminism being as ‘lgbt’ and vice versa. As a result, the thresholds for these topics to be confidently labeled had to be fairly high, which might have contributed to the low number of labeled tweets.

Republicans in general also tweeted far less than their Democratic counterparts, also explaining why it was so difficult to gather a substantial number of topic-labeled tweets on the Republican side.

Mixed Linear Model

We use a mixed linear model to determine the role tweet sentiment and topic play towards election success. For each candidate, the variables we use include the number of tweets the candidate posted in our 1 year time window, the sentiment score (described in the "Sentiment Analysis" section of the paper), and the number of tweets for each topic chosen from results in our topic modeling. Our results are shown below.

Figure 1 – Democrat MixedLM:

```
             Mixed Linear Model Regression Results
==================================================================
Model:                    MixedLM    Dependent Variable:    won_primary
No. Observations:         606        Method:                REML
No. Groups:               598        Scale:                 0.1487
Min. group size:          1          Likelihood:            -501.5674
Max. group size:          2          Converged:             Yes
Mean group size:          1.0
------------------------------------------------------------------
                        Coef.  Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------------
Intercept                0.424    0.062  6.882 0.000  0.303  0.544
candidate               -0.000    0.000 -1.743 0.081 -0.000  0.000
num_tweets               0.000    0.000  1.234 0.217 -0.000  0.000
sentiment               -0.568    0.213 -2.670 0.008 -0.985 -0.151
gun_control_tweet_count -0.002    0.001 -2.057 0.040 -0.004 -0.000
health_care_tweet_count  0.002    0.001  2.793 0.005  0.001  0.004
kavanaugh_tweet_count   -0.004    0.003 -1.661 0.097 -0.010  0.001
jobs_tweet_count         0.002    0.001  1.617 0.106 -0.000  0.005
education_tweet_count    0.001    0.002  0.530 0.596 -0.002  0.004
abortion_tweet_count    -0.004    0.005 -0.943 0.345 -0.014  0.005
environment_tweet_count -0.001    0.001 -1.224 0.221 -0.002  0.000
trump_tweet_count       -0.001    0.001 -1.480 0.139 -0.002  0.000
immigration_tweet_count  0.003    0.002  1.724 0.085 -0.000  0.007
freedom_tweet_count      0.012    0.005  2.233 0.026  0.001  0.022
Group Var                0.094    0.369
==================================================================
```

Figure 2 – Republican MixedLM:

```
             Mixed Linear Model Regression Results
==================================================================
Model:                    MixedLM    Dependent Variable:    won_primary
No. Observations:         413        Method:                REML
```

```
No. Groups:           403        Scale:                 0.0018
Min. group size:      1          Likelihood:            -280.3402
Max. group size:      2          Converged:             Yes
Mean group size:      1.0
------------------------------------------------------------------
                        Coef.  Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------------
Intercept                0.275    0.055  5.020 0.000  0.168  0.383
candidate                0.000    0.000  1.544 0.123 -0.000  0.001
num_tweets               0.000    0.000  2.589 0.010  0.000  0.000
sentiment               -0.468    0.205 -2.287 0.022 -0.869 -0.067
gun_control_tweet_count -0.004    0.002 -2.043 0.041 -0.008 -0.000
kavanaugh_tweet_count    0.012    0.003  3.302 0.001  0.005  0.018
jobs_tweet_count         0.006    0.001  4.368 0.000  0.003  0.008
abortion_tweet_count    -0.000    0.004 -0.004 0.996 -0.007  0.007
environment_tweet_count -0.004    0.009 -0.412 0.680 -0.020  0.013
immigration_tweet_count  0.000    0.002  0.278 0.781 -0.003  0.004
education_tweet_count   -0.002    0.002 -0.779 0.436 -0.007  0.003
health_care_tweet_count  0.001    0.001  0.844 0.399 -0.001  0.003
freedom_tweet_count     -0.003    0.002 -1.389 0.165 -0.007  0.001
trump_tweet_count       -0.002    0.001 -3.134 0.002 -0.003 -0.001
Group Var                0.178    2.632
==================================================================
```

By removing the topics whose p-values are higher than 0.1 from Figures 1 and 2, we get a more focused mixed linear model result shown below.

Figure 3 – Democrat Filtered MixedLM:

```
                Mixed Linear Model Regression Results
==================================================================
Model:                MixedLM    Dependent Variable:    won_primary
No. Observations:     606        Method:                REML
No. Groups:           598        Scale:                 0.1554
Min. group size:      1          Likelihood:            -477.2160
Max. group size:      2          Converged:             Yes
Mean group size:      1.0
------------------------------------------------------------------
                        Coef.  Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------------
Intercept                0.335    0.059  5.658 0.000  0.219  0.452
candidate                0.000    0.000  1.459 0.144 -0.000  0.000
num_tweets               0.000    0.000  1.250 0.211 -0.000  0.000
sentiment               -0.665    0.205 -3.250 0.001 -1.066 -0.264
```

```
gun_control_tweet_count -0.003    0.001 -2.812 0.005 -0.005 -0.001
health_care_tweet_count  0.003    0.001  4.917 0.000  0.002  0.005
kavanaugh_tweet_count   -0.007    0.002 -3.082 0.002 -0.012 -0.003
immigration_tweet_count  0.002    0.002  1.224 0.221 -0.001  0.005
freedom_tweet_count      0.014    0.005  2.698 0.007  0.004  0.024
Group Var                0.089    0.389
===================================================================
```

Figure 4 – Filtered Republican MixedLM:

```
             Mixed Linear Model Regression Results
===================================================================
Model:                   MixedLM    Dependent Variable:    won_primary
No. Observations:        413        Method:                REML
No. Groups:              403        Scale:                 0.0015
Min. group size:         1          Likelihood:            -249.8288
Max. group size:         2          Converged:             Yes
Mean group size:         1.0
-------------------------------------------------------------------
                        Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------------
Intercept                0.250    0.053  4.682 0.000  0.145  0.355
candidate                0.000    0.000  2.213 0.027  0.000  0.001
num_tweets               0.000    0.000  2.547 0.011  0.000  0.000
sentiment               -0.432    0.188 -2.293 0.022 -0.801 -0.063
gun_control_tweet_count -0.005    0.002 -2.737 0.006 -0.008 -0.001
kavanaugh_tweet_count    0.013    0.003  3.760 0.000  0.006  0.019
jobs_tweet_count         0.005    0.001  4.465 0.000  0.003  0.007
trump_tweet_count       -0.002    0.000 -3.549 0.000 -0.003 -0.001
Group Var                0.177    2.765
===================================================================
```

After removing topics/factors that add noise, our new mixed linear model shows the remaining variables to be highly significant (shown in bold in figures 3 and 4). These variables have been stress-tested over several variations of adding/removing different variables to give confirm with more certainty that these variables are actually significant. The results of the model suggest that sentiment and the volume of tweets about particular topics are significantly correlated with success in elections. Further, the topics that are significant for Democrats differ from those of Republican, also as demonstrated by our topic explorations.

We can see sentiment is a very significant factor for both Democrats and Republicans, with Democrats showing a higher significance (0.001 in Figure 3 vs 0.022 in Figure 4) and correlation coefficient (-0.665 in Figure 3 vs -0.432 in Figure 4). The negative correlation coefficient indicates that the higher the sentiment score (higher sentiment means more negative), the more likely the candidate is to lose their election. However, sentiment for both

Democrats and Republicans have high standard deviations (0.205 in Figure 3 and 0.188 in Figure 4), so the magnitude of the effect sentiment plays is still uncertain. Regardless, it's notable to point out even on the lower end of their confidence intervals, sentiment still has the largest effect.

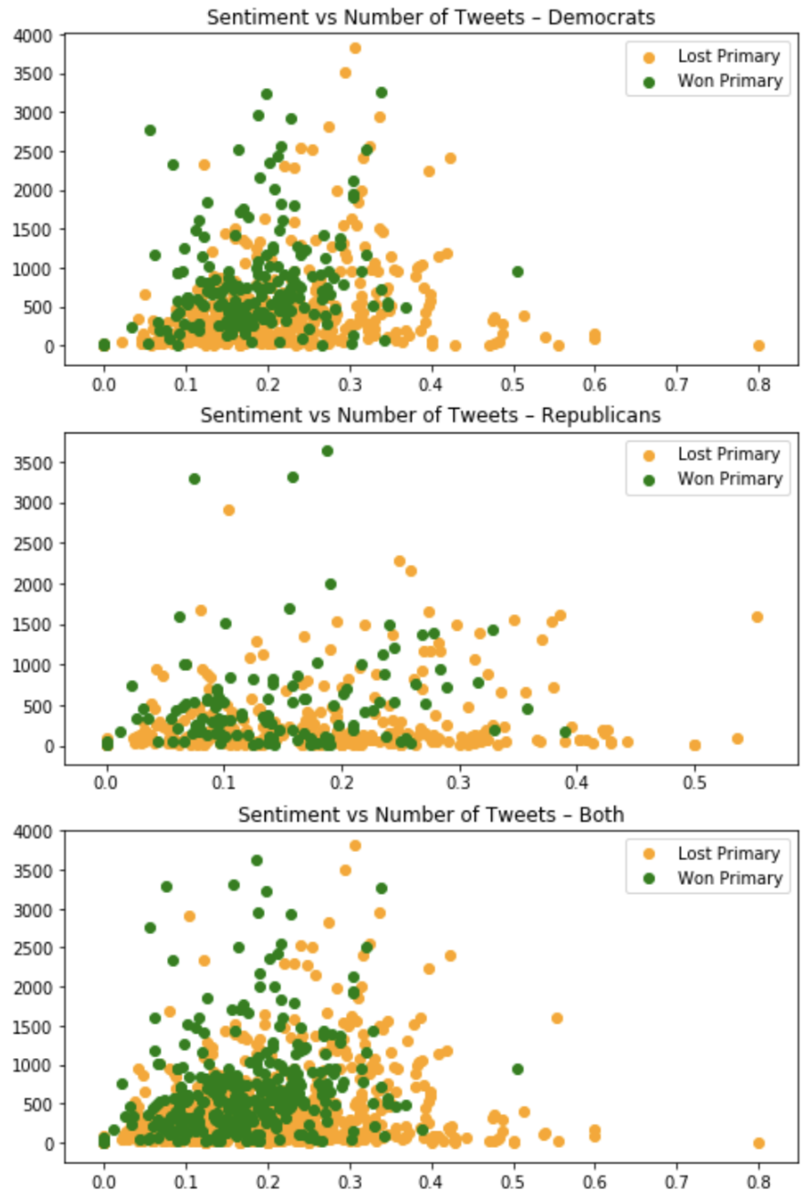
Both Democrats and Republicans had 4 topics that were shown to be significant from the mixed linear model, and each of these topics had relatively low correlation coefficients. While these topics were significant, there wasn't a large magnitude of effect that these topics had in determining the overall success of a candidate. Through our stress-testing of our variables, healthcare for Democrats and jobs for Republicans were the two most poignant variables, as their effect remained the least variable. The more Democrats tweet about healthcare and the more Republicans tweet about jobs are correlated with success. Considering stereotypes for the two parties, this isn't particularly surprising that these correlations exist. Tweets about Brett Kavanaugh and gun control were the two overlapping topics that were significant to both parties during this election. While correlation with gun control was more or less the same for both parties, the correlation coefficient for the count of Kavanaugh tweets was negative for Democrats and positive for Republicans. Kavanaugh was clearly a hot topic during the 2018 midterm elections, and given that Kavanaugh is a Republican it's reasonable to see a positive correlation for Republicans and a negative correlation for Democrats.

While there are many topics that weren't shown to be significant in the mixed linear model, it doesn't necessarily mean that they aren't important features, rather they're features that don't set candidates apart from their parties.

Further Analysis on Significant Values From Mixed Linear Model

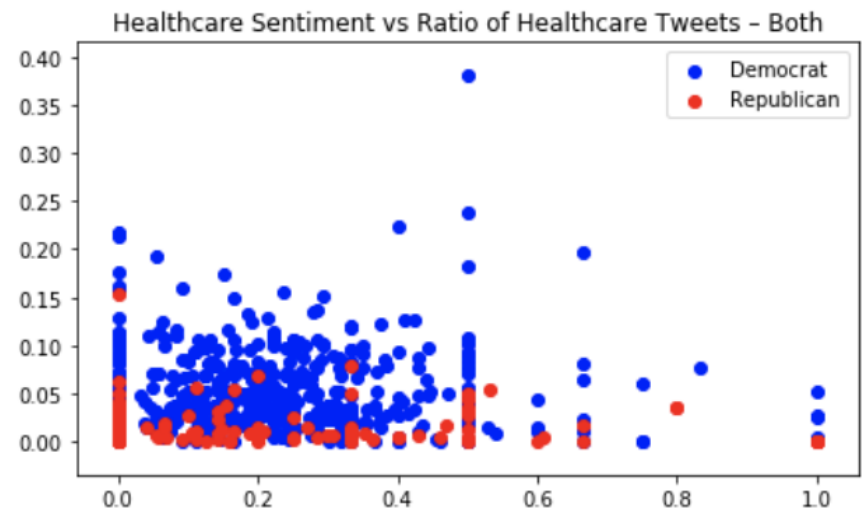
Sentiment

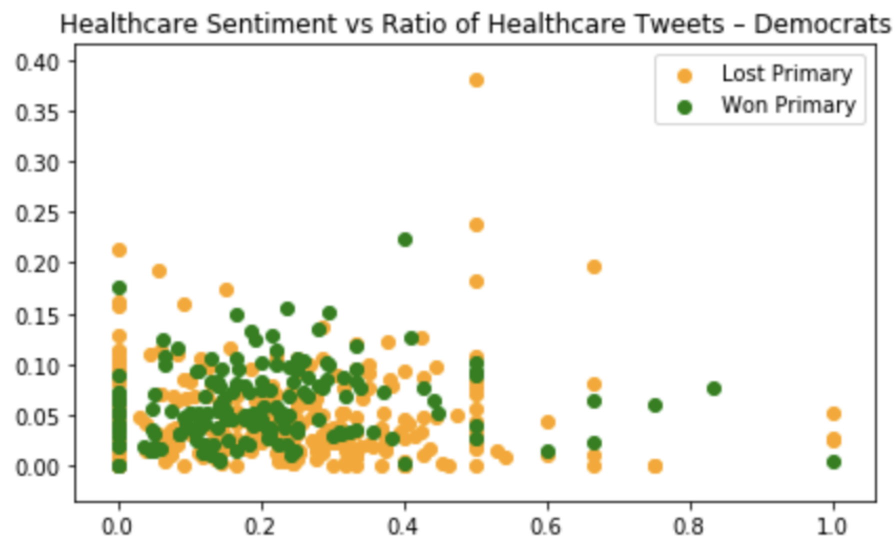
Sentiment was had the highest correlation coefficient from the results of the mixed linear model. In the graphs shown below, we use sentiment in the x-axis and number of tweets in the y-axis. The green points represent candidates who won their primary and the orange dots represent candidates who lost their primary. As shown in the graph, the average sentiment score (higher sentiment score means the candidate's tweets were more negative on average) for those who win their primary tends to be lower than for those who lost their primary, and this trend holds up for both parties.



Democratic Relationship to Healthcare

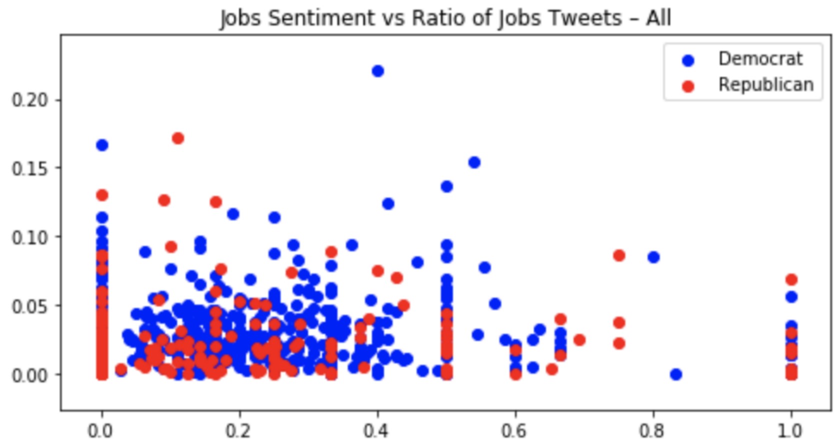
Looking at the "healthcare" topic, we can see that Democrats tweet more about healthcare relative to other things than republicans do. This is consistent with the mixed linear model result of healthcare being a significant factor for Democrats and not republicans. Looking further into this result, we can see that the ratio of tweets about healthcare was significantly higher for Democrats who won their primary. By running a z-test, we were able to confirm that the ratio of healthcare tweets to the total number of tweets for candidates who won was significantly higher than for candidates who lost (z-score = 2.729, p-value = 0.00635). While it appears tweet sentiment related to healthcare for Democrats may be significant, a z-test reveals there is no statistical significance (z-score = 0.119, p-value = 0.906).

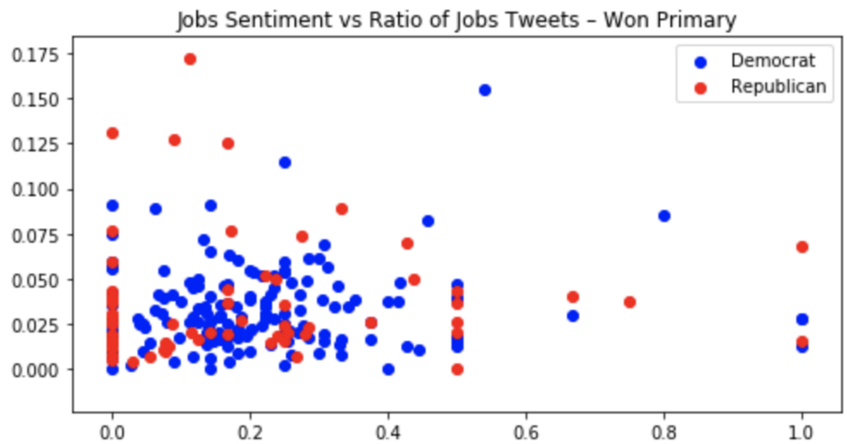


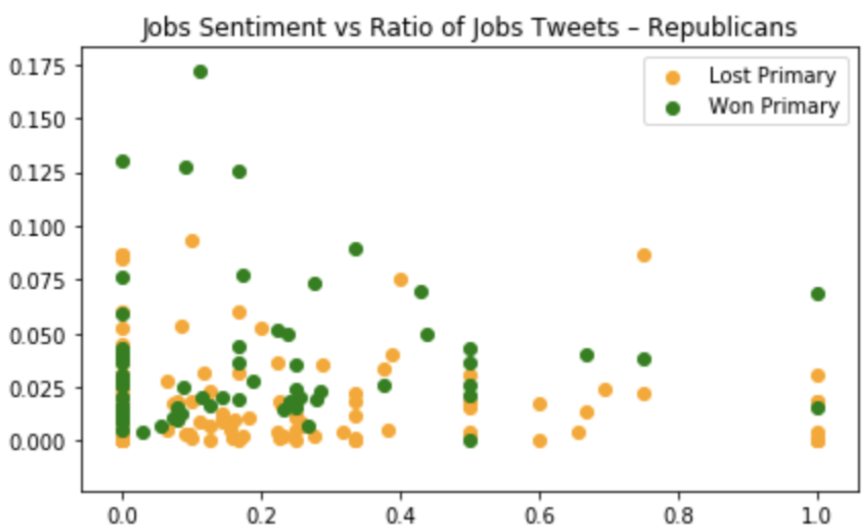


Republican Relationship to Jobs

The "jobs" topic was significant in the mixed linear model for republicans. We can see that Democrats tweet more about jobs relative to other things than republicans do, but when we look exclusively at candidates who won their primaries, the ratios are much closer. Similar to the healthcare analysis, we can see that the ratio of tweets about jobs was significantly higher for republicans who won their primary compared to republicans who lost. By running a z-test, we were able to confirm that the ratio of jobs tweets to the total number of tweets for candidates who won was significantly higher than for candidates who lost (z-score = 7.189, p-value = 6.514e-13). Tweet sentiment related jobs however, wasn't significant for republicans (z-score = -0.410, p-value 0.682).





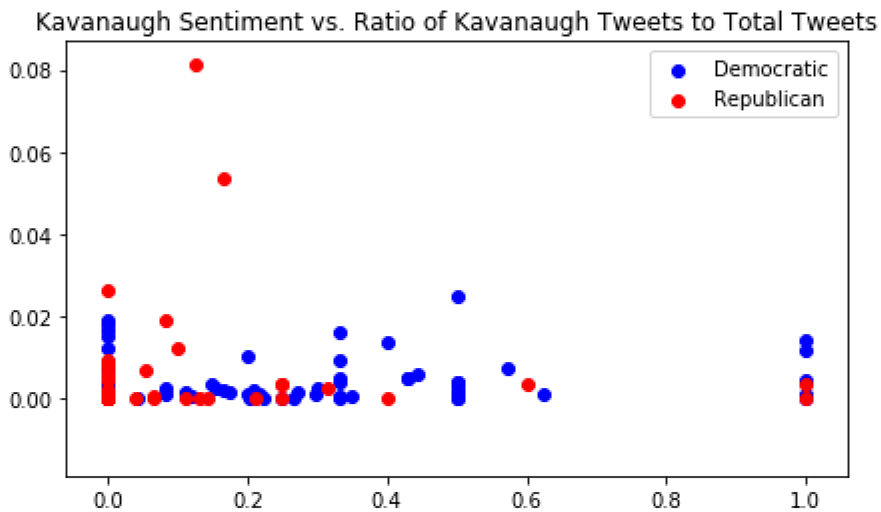


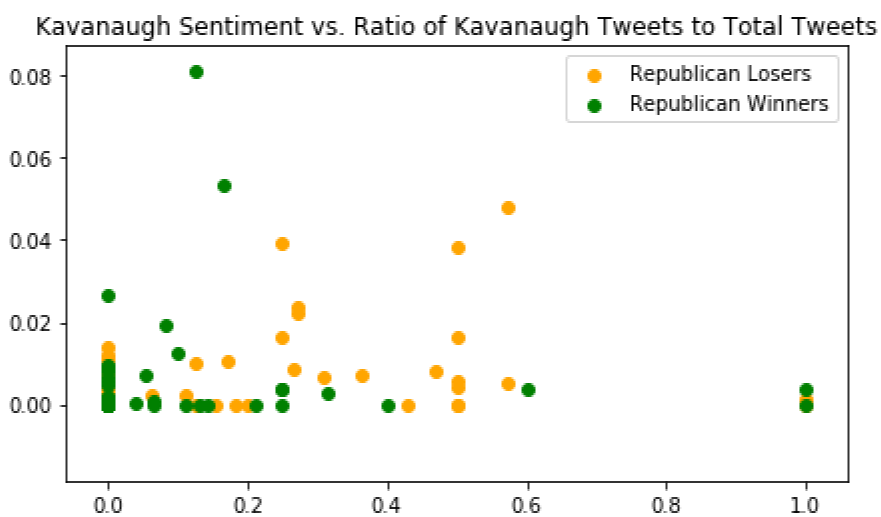
Kavanaugh

While there was sparse tweet data on the "kavanaugh" topic, we found a statistically significant difference (p-value = 0.056, z-score = 1.91) between the tweeting habits of the Republican and Democratic candidates who won their elections. Unsurprisingly, Republicans tweeted less negatively about Republican Supreme Court Justice Brett Kavanaugh than their Democratic counterparts.

Among Republican candidates, however, those who lost their candidacy tended to tweet more negatively than those who won (p-value = 0.04, z-score = -2.02).

The Kavanaugh topic is interesting in particular, because the events of the controversy occurred within the weeks and leading up to the election date, making it very possible that it had a significant impact on voter impressions because of its recency.





Conclusions/Main Findings

Based on our various experiments, models, and observations, we've recognized a few overarching findings:

Overall candidate sentiment matters
Looking at the results of the Mixed Linear Model, a candidate's overall sentiment significantly influences electoral success more than any other factor we considered. Given that our sentiment scores are based on negativity, having more negative tweets is correlated with losing a primary election. It's also interesting to point out that this correlation is more statistically significant for Democrats than for Republicans. Further, the sentiment correlation coefficient is higher for Democrats, suggesting that negative sentiment has a greater negative impact on Democratic candidates than Republicans. However, the standard deviation associated with sentiment for both Democrats and Republicans is high, making it difficult to determine the magnitude of effect sentiment plays. These results could possibly also have been influenced by the fact that there are significantly more Democratic tweets than Republican. Nonetheless, when observing Twitter behavior and candidate success, negative sentiment is statistically significant in determining election outcomes.

Democratic and Republican candidates care about different topics – which topics they tweet about matters
As demonstrated in our topic modelling and topic classification results, both parties have different issues they find worth tweeting about. We identified a few common topics through tweet frequency and topic modeling, such as healthcare and gun control. However, the mixed linear model revealed that there are other common topics, such as "kavanaugh," that weren't tweeted about frequently but had significant impact on electoral results for both Democrats and Republicans.

It's also interesting to consider the topics deemed statistically significant by the mixed linear model and the actual tweet counts per topic. Given that "kavanaugh" was an important feature and on average, Democratic winners tweeted about Kavanaugh 3.47 times and losers tweeted 2.55 times, had the Democratic losers tweeted about Kavanaugh one extra time, it might have benefited their campaign. It's unclear one tweet would have significantly impacted election results, but its significance is certainly one worth considering.

Candidates don't necessarily use Twitter to express their opinions
From our topic classification results, a significant portion of the tweet corpus cannot be classified under a comprehensive topic. This unclassifiable subset points to the fact that during elections, candidates tweet more frequently to promote their campaign through announcements and public acknowledgements & endorsements ("thank you's"), and less frequently use it to voice their political stances. This is reinforced by our topic modelling results, as nearly half of all identified topics were related to campaigning, tagging other users, making announcements, and thanking voters. In total, we were only 15% of our total tweet corpus using the range of topics we curated through our topic model and by our own manual observations. While we sometimes associate Twitter as a platform in which high profile individuals informally interface with audiences and share their opinions (as frequently done by our President), based on our findings, Twitter is still largely used as an extension of curated political campaigns.

Shortcomings & Future Steps

While our results show trends and correlations between political candidates' Tweets and the results of their election, this project is an exploration of the relationship between Twitter habits and political outcome as opposed to a conclusive study. A few places our project falls short include limitations on our sentiment analyzer, an incomplete sample of all the races that took place during the 2018 midterm election, inability to gather retweets, and a general lack of Twitter activity coming from Republicans.

Although our BERT based sentiment analyzer performed pretty well, when we qualitatively look at the data, there was some clear ambiguity from the program as to what qualifies as a positive Tweet. This is furthered by the fact that many Tweets display a range of emotions. For example, a Tweet can be both sad yet hopeful, and even at the human level it's unclear if such a Tweet should be labeled as positive, negative, or neutral. Some future steps on this front could be to decipher the different emotions each Tweet has and tally up the occurences of each emotion to get a better sense on sentiment. As NLP technology and research continues to develop, our sentiment analysis will inevitably improve as well.

Due to the vast number of races that occurred during the 2018 midterm election, it was simply infeasible to collect data for each race. However, our study could be improved if we were able to collect a more comprehensive dataset.

Republicans tended to be less active on Twitter. They tweeted less often than Democrats and fewer Republicans had a Twitter accounts compared to Democrats. This materialized in our mixed linear model result where the number of tweets were a significant indicator to the success of the candidate for Republicans, but not Democrats. It appeared that more Democratic candidates were active Twitter users regardless of how hard they were campaigning, whereas many of the Republican candidates only had a Twitter if they were campaigning. Thus, many of the less serious candidates simply didn't have a Twitter that could be found. This potential issue of sparsity was furthered by our inability to gather retweets. Due to the way the Twitter API works, we weren't able to collect retweets which is a common way for candidates to stay active on Twitter. A sparse Republican dataset could have affected our results, and more research should be done to determine if sparsity is affecting our results and by how much.

We also believe that our analysis would greatly benefit from additional contextual data at more local district levels to further explain certain trends or correlations we found in our exploration. It would be interesting to see how much topic frequency is influenced by a candidate's target constituency. This would be effective in measuring how well candidates cater their tweet habits to their respective voting populations and ultimately how much that has an influence on their chances of winning.

It would also be interesting to see how topic classification, sentiment, and Twitter usage in general differ along race types, Senator, Representative, and Governor, as each race involves varying levels of locality and very voting demographics. It might be that candidates with larger constituents have a greater incentive to use Twitter and therefore tweet about more concrete issues, accounting for a lack of data we could find topic labels for.