199 Final Report

QJ

12/23/2020

Introduction

The first United States Presidential Debate was held in Chicago on September 26, 1960 to an audience of 66.4 million. Since then, the event has become a regular mainstay of both presidential (with the exceptions of 1964, 1972 and 1978) and numerous primary elections. These debates are often pointed to as key milestones in an election cycle where viewers could make their decisions whether to support or reject a candidate. In 2016, Marco Rubio's poor performance in the first February Republican Primary, where he was criticized for often repeating stock talking points, was often cited as the main reason for his 6th place finish in the New Hampshire primary and ultimate withdrawal from the race. Another historic debate moment came in 1988, where Democratic vice-presidential candidate Lloyd Bentsen famously quipped "Senator, you're no Jack Kennedy" to the Republican candidate, Dan Quayle.

However, whether debates really matter in the grand scheme of things remains up for contention. A 2019 study by Pennec and Pons (https://www.nber.org/papers/w26572) claimed that debates neither helped undecided voters make their minds up nor decided voters to switch their choices. It is perhaps telling that even in 1988, it was the Republican candidate for president, George H. W. Bush who defeated the Democrat Michael Dukakis to clinch the presidency. In recent years, controversial debates such as the first Trump-Biden debate on September 29, 2020 have also brought into question how the nature of debates have changed over time, whether it be the complexity of the arguments or the type of sentiments conveyed.

This report will attempt to demystify various aspects of the debates. More specifically, I will first explore the relationship between polling data pre- and post-debate to see how presidential or primary debates affect a candidate's support. Following that, I will also look into the complexity and sentiment of presidential debates over time to see if we can observe a statistically significant change.

Methods

I used polling data provided by FiveThirtyEight (https://github.com/fivethirtyeight/data/tree/master/polls) covering presidential races from 1968-2016, as well as primary races from 1980-2020. This estimate of polling averages was created by aggregating the results from multiple pollsters (https://fivethirtyeight.com/features/what-makes-our-new-2020-democratic-primary-polling-averages-different/), weighted by different factors such as the historical accuracy of each pollster. Within the datasets, polling averages for the primary and presidential races were provided for each week up till the Democratic/Republican Convention and Election day respectively. To look at how public opinion of presidential candidates were affected by debates, I compared the polling averages before and after the dates of the debate, and compared them to the mean change in polling average for all candidates from week to week to see if there was a significant difference. Specifically, I used the top 8 contenders (filtered by candidates with a peak estimated polling percentage of 5% or more) in the Democratic Primary race of 2020, ensuring the consistency of comparison between polling averages for debate weeks and regular weeks.

```
##          Date Party readability  days
## 1: 2016-10-09   DEM    8.586330 14626
## 2: 2016-10-19   DEM    9.189510 14636
## 3: 2016-09-26   DEM    9.638791 14613
## 4: 2012-10-22   DEM   10.731116 13178
## 5: 2012-10-16   DEM    8.488426 13172
## 6: 2012-10-03   DEM   10.377963 13159
```

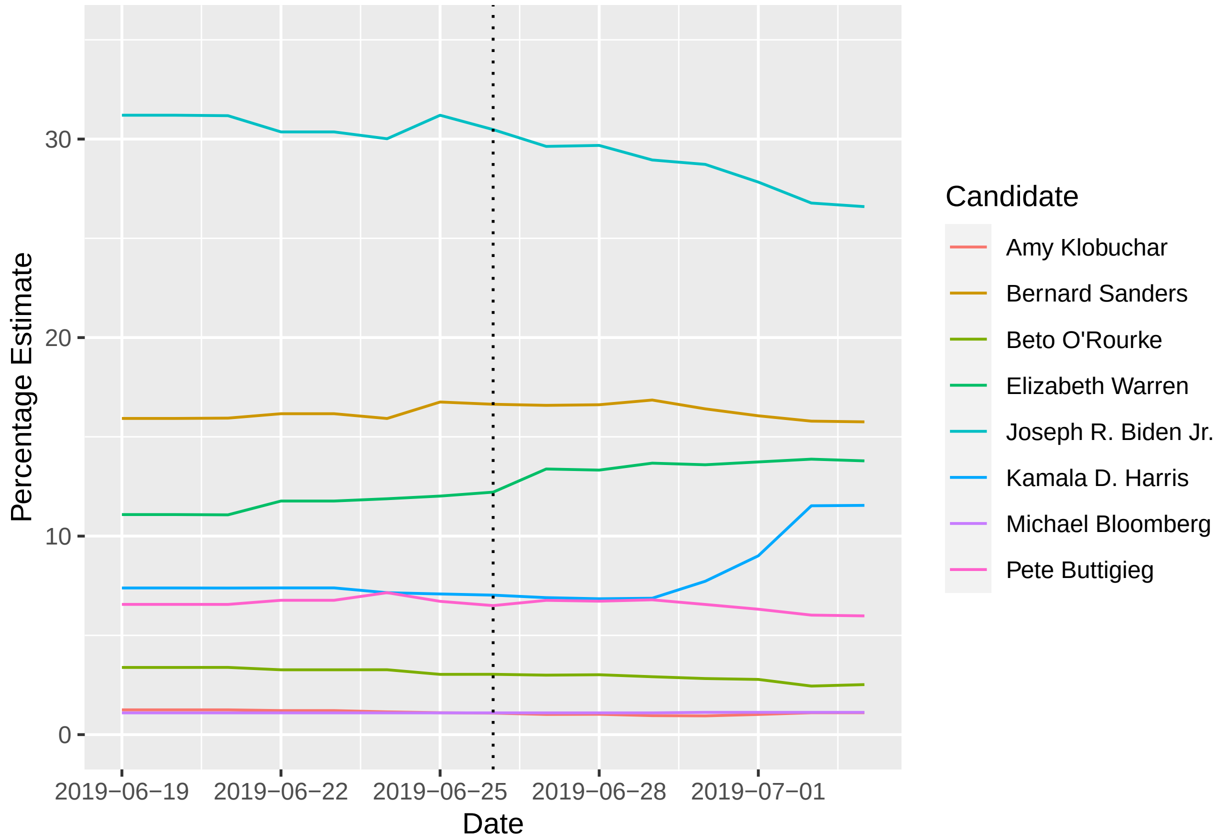
```
##    Year Party proportion sentiment
## 1: 2016   DEM  0.5221932  negative
## 2: 2016   DEM  0.4778068  positive
## 3: 2012   DEM  0.4077961  negative
## 4: 2012   DEM  0.5922039  positive
## 5: 2008   DEM  0.5837275  negative
## 6: 2008   DEM  0.4162725  positive
```

To measures changes in complexity and sentiment of debates over time, I scraped and cleaned the transcripts of every single presidential debate from https://www.debates.org/voter-education/debate-transcripts, then used the tidytext and tm packages in R to obtain a corpus of texts for each election cycle. In lieu of a recognized algorithm to determine the complexity of spoken text, I used the Flesch–Kincaid readability test to determine the grade level of each candidate's transcribed debate responses for every election cycle from 1976 to 2016. Following this process, we used a simple t-test with the null hypothesis, H_0 as $\beta_1 = 0$, and alternative hypothesis, H_a as $\beta_1 \neq 0$, where β_1 represents the number of days since the first debate used in this analysis on September 23rd, 1976. To control for possible confounding variables, I also included the party the candidate belonged to as a variable in my model to see if this was a relevant factor. Our final model looks as such: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, with the response variable Y being the reading grade level of the transcribed debates of all presidential candidates, X_1 as the number of days since the first debate, X_2 as the party the candidate belongs to, β_2 as the coefficient of the party and ϵ as the random error of the model following the distribution $N(0, \sigma_\epsilon)$.

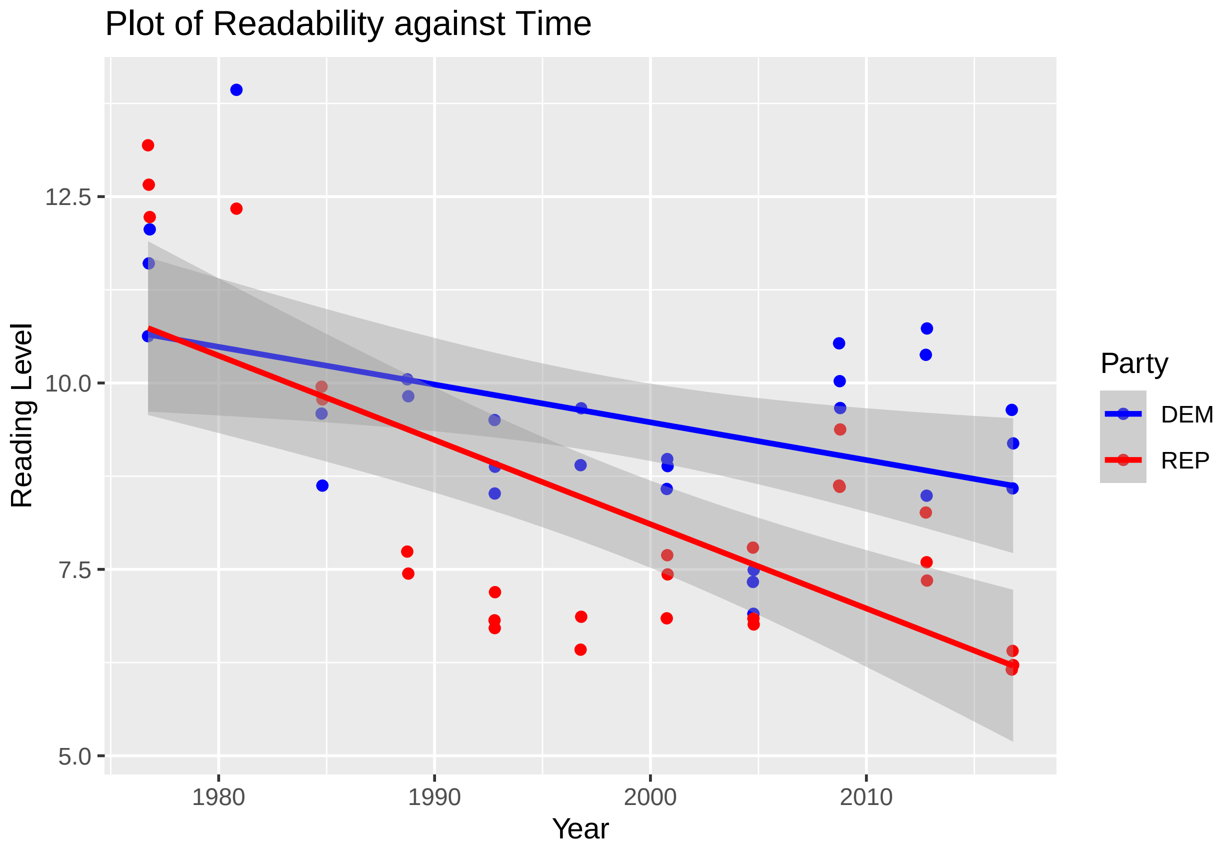
Conversely, the text sentiment of every candidate was extracted by filtering stop words from the text, then using the Bing Sentiment Lexicon and NRC Word-Emotion Association Lexicon to obtain both binary and more complex sentiments depending on the words the candidates used. Similar to the previous section, we used a simple t-test with the null hypothesis, H_0 as $\beta_1 = 0$, and alternative hypothesis, H_a as $\beta_1 \neq 0$, where β_1 represents the number of days since the first debate used in this analysis on September 23rd, 1976. Likewise, I also included the party the candidate belonged to as a variable in my model to see if this was a relevant factor. Our final model looks as such: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, with the response variable Y being the percentage of negative sentiment of the transcribed debates of all presidential candidates, X_1 as the number of days since the first debate, X_2 as the party the candidate belongs to, β_2 as the coefficient of the party and ϵ as the random error of the model following the distribution $N(0, \sigma_\epsilon)$.

In the case of the NRC Word-Emotion Association Lexicon, because of its multilevel nature, I decided to focus on a few specific sentiments. Much of the political discourse of today has centered around the idea of harnessing negative emotion, so I decided to use 'Fear', 'Anger' and 'Disgust', to see if the extent of these sentiments in the debates changed over time, with the same methodology used to measure the change in negative sentiment.

Results

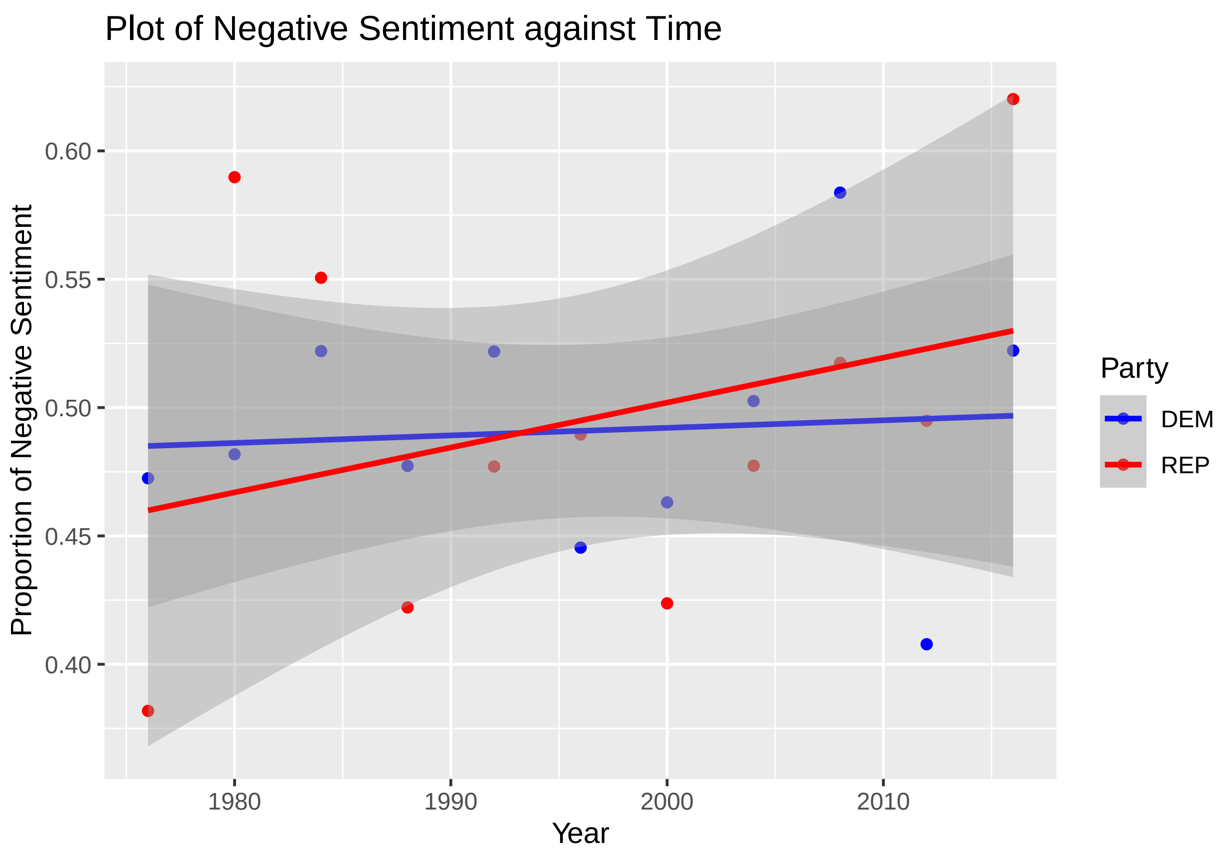


Perhaps the most prominent example of debates influencing public opinion in the 2020 election cycle was the first Democratic Primary debate in June 2019, where Kamala Harris won plaudits for her attacks on Joe Biden for his previous support for racial policies like busing. Comparing pre and post debate polling, Harris went from an estimated 7.39% to 11.5% of the vote, approximately a 55% increase in support. Conversely, Biden dropped 4.6 points from 31.2% to 26.6%. After gathering all the polling changes from week to week, I found the mean polling change for regular weeks to be 0.00285% and mean polling change for debate weeks to be 2.174%. An ANOVA test comparing the two groups found significant differences in the means of the two groups, resulting in a p-value of 1.23e-05. This points to debates indeed being a significant factor in influencing public opinion.

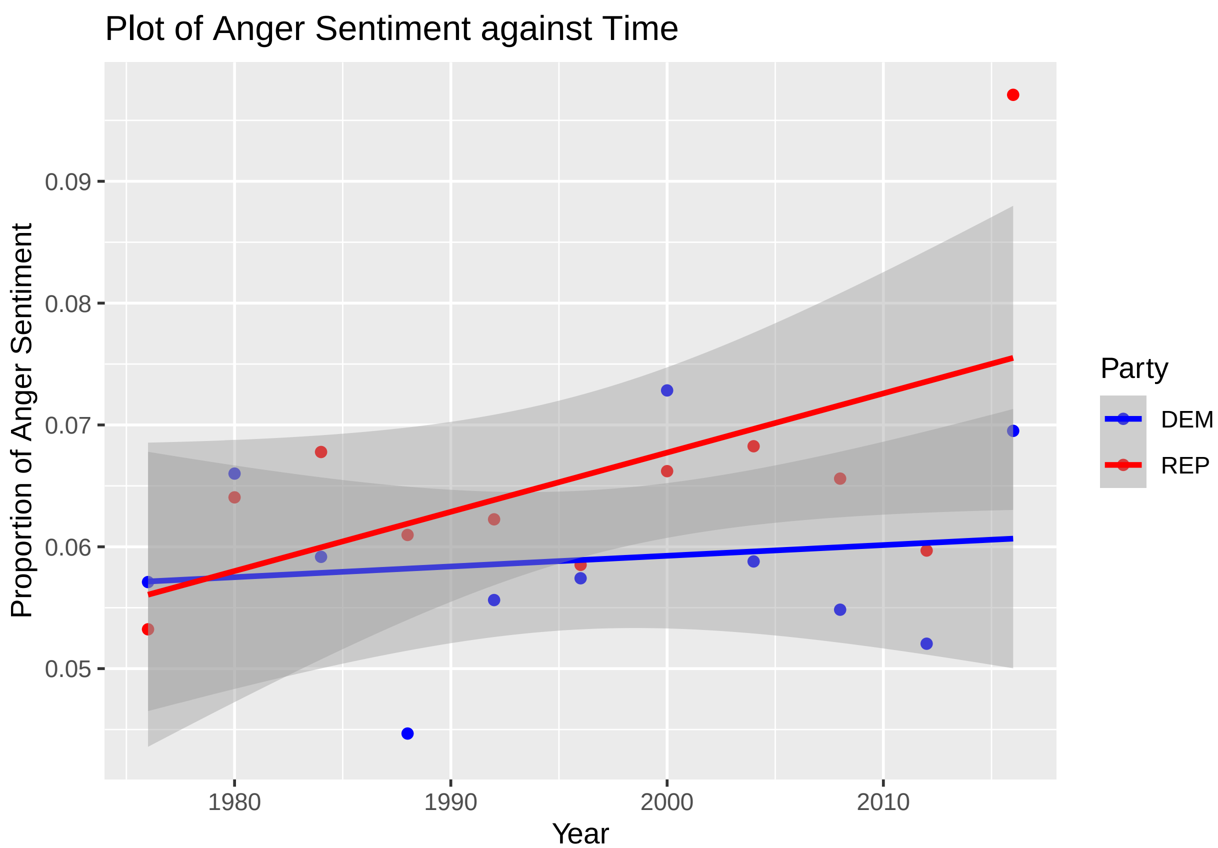


The linear model of readability over time resulted in a p-value of 2.35e-06, suggesting a statistically significant relationship between reading level and date of the debate. The coefficient of this variable was negative as well, leading me to believe that reading level has a strong negative correlation with debate date. For example, in the September debate, the transcript of Donald Trump's statements were classified at a grade level of 6.16, the lowest rating in our data. However, the highest reading level identified was 13.93, by incumbent Jimmy Carter in 1980, the second earliest year in the dataset.

Another interesting insight was that beyond the age of the election cycle, the model recognized the party of the candidates as another statistically significant factor. In the case of readability over time, While both parties had similar reading grades in 1976, it seems like the reading levels of the transcripts of Republican candidates have dropped significantly more than that of Democratic candidates.



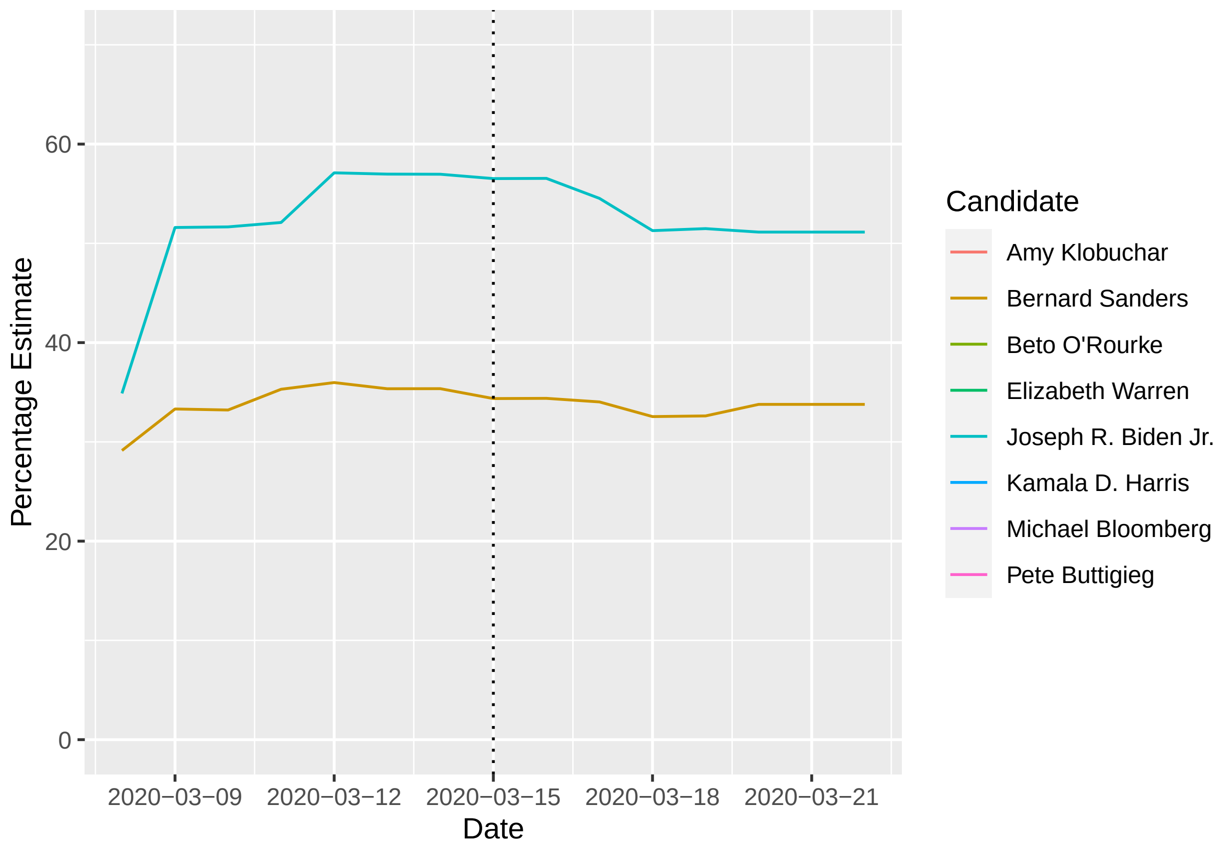
However, linear models looking at the change in sentiment in debates over time were largely inconclusive. While the coefficient of the term in the model evaluating the change in negative sentiment was indeed positive as expected, the p-value of the model was only 0.332, implying that there was a possibility that the variation in our data was a result of random error. This pattern of high p-values was the case for the models looking at the changes in the fear and disgust sentiments as well.



However, it is worth noting that in a model looking at the change of proportion over time, the p-value was 0.0815. This suggests that there was some level of correlation between debates over time and the proportion of anger sentiment in the transcripts of presidential candidates. Looking at the plot of anger over time, it seems like Republican candidates have a much higher increase in anger sentiment proportion compared to Democratic candidates. This observation is reflected in the relatively low p-value of 0.0976 in the case of the Party variable.

Conclusion and Limitations

My results appear to indicate that debates do affect polling data to an extent. However, there are some caveats that need to be kept in mind when considering this relationship. For one, the polling estimates that I used were specifically for the 2020 Democratic Primary. It is entirely possible that the relationship between debate performance and polling results could be very different when looking at the polling averages for another year or party. Similarly, whether debates affect presidential races rather than primary races remains up in the air. Something else that might also have affected my results was the decision to use week by week polling estimates as the main metric of evaluating debate performance.



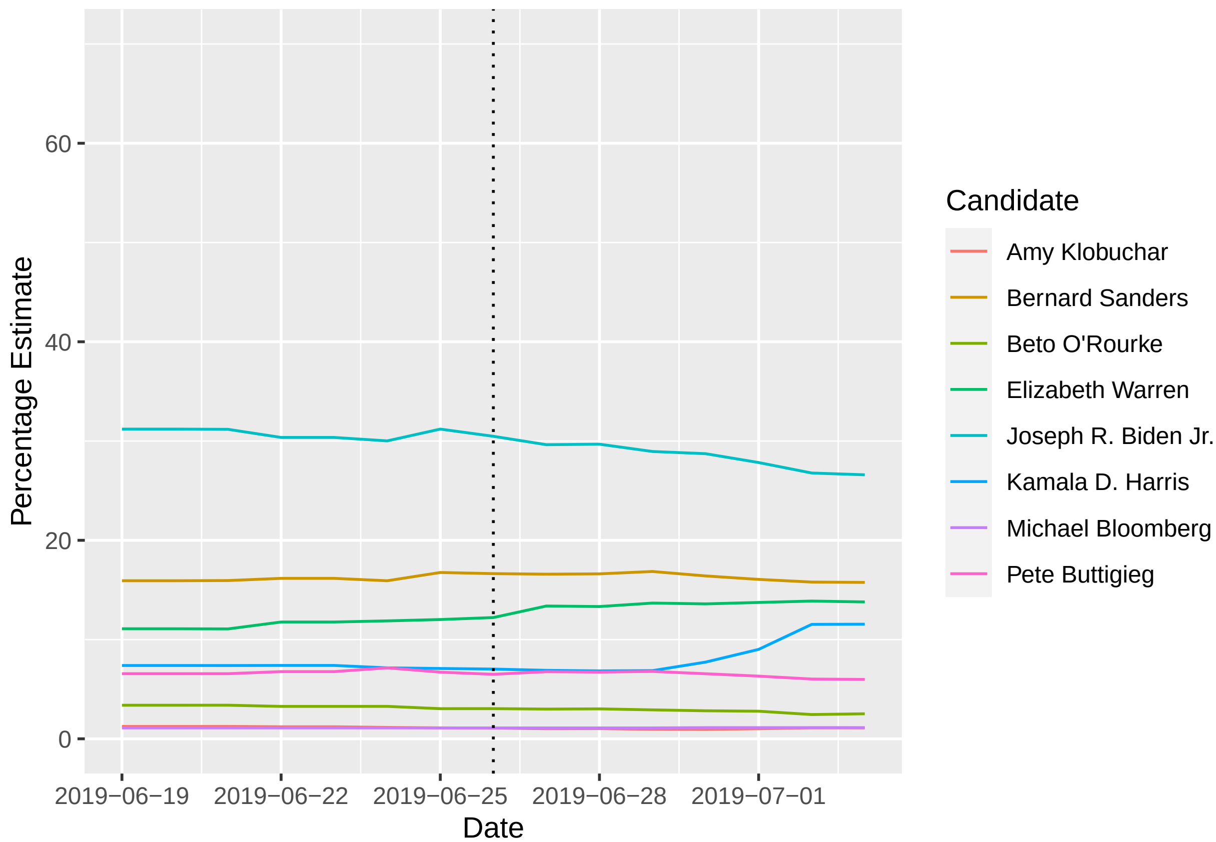
In the example above, Joe Biden experienced a large increase in voting percentage comparing the week before the final Democratic Primary and the week after. However, a closer inspection of the plot reveals that the bump of percentage share came before the debate rather than after. This could mean that the effect of debates should be studied more in their surrounding days rather than weeks. This could also have been caused by the seemingly assured eventuality of Biden's victory after Super Tuesday on 3rd March. Since Bernie Sanders' vote share did not decrease in contrast, it could be that previously undecided voters came out for Biden rather than any effect of the debates.

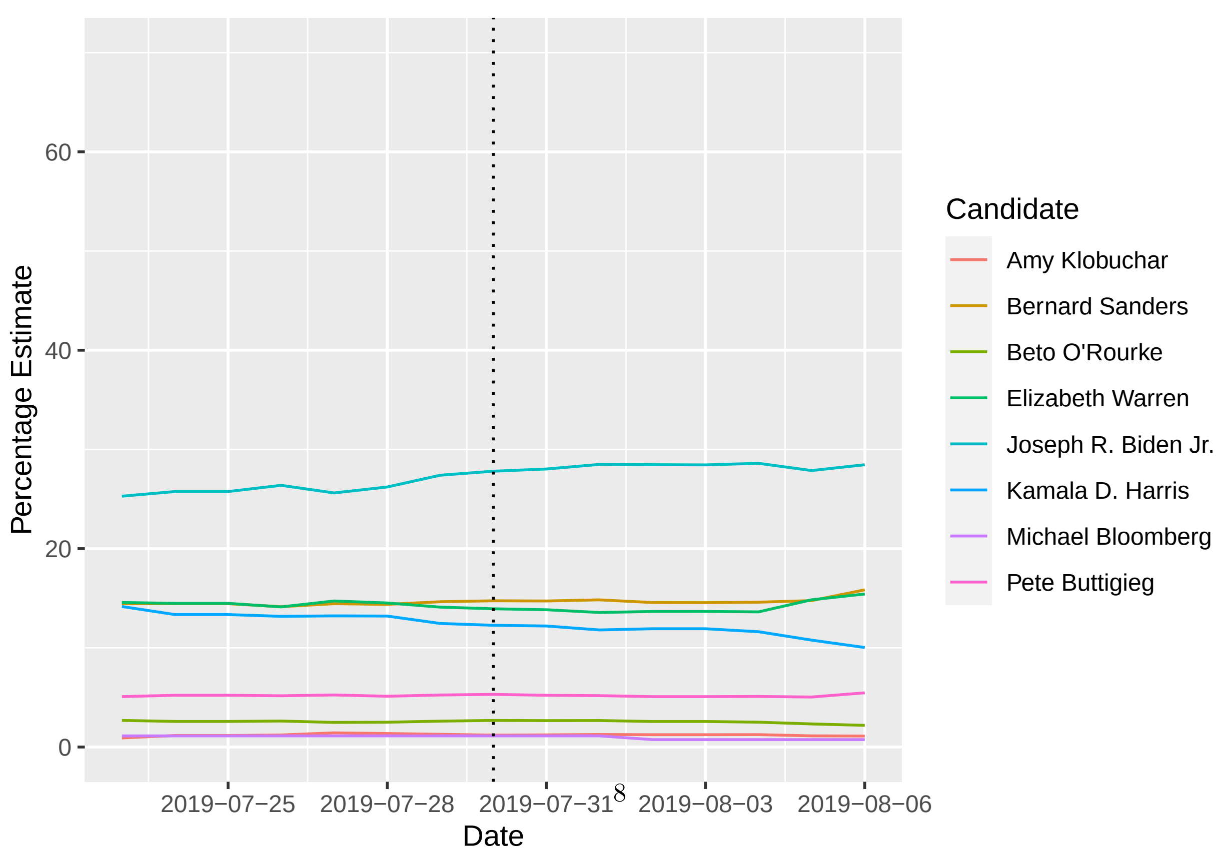
However, it seems clear that debates have gotten less complex over time, in terms of the words used by candidates. While debates are obviously spoken and not read, I believe that using the reading grade as a metric for complexity serves its purpose well, especially since the analysis uses the same Flesch–Kincaid metric continuously. Coming up with a good way to evaluate spoken word directly through measures like intonation, speed or pronunciation could allow future studies to make leaps and bounds in this field of research. On the other hand, it seems like for the most part, the overall proportion of different sentiments has not changed much over time, with the exception of anger to some degree.

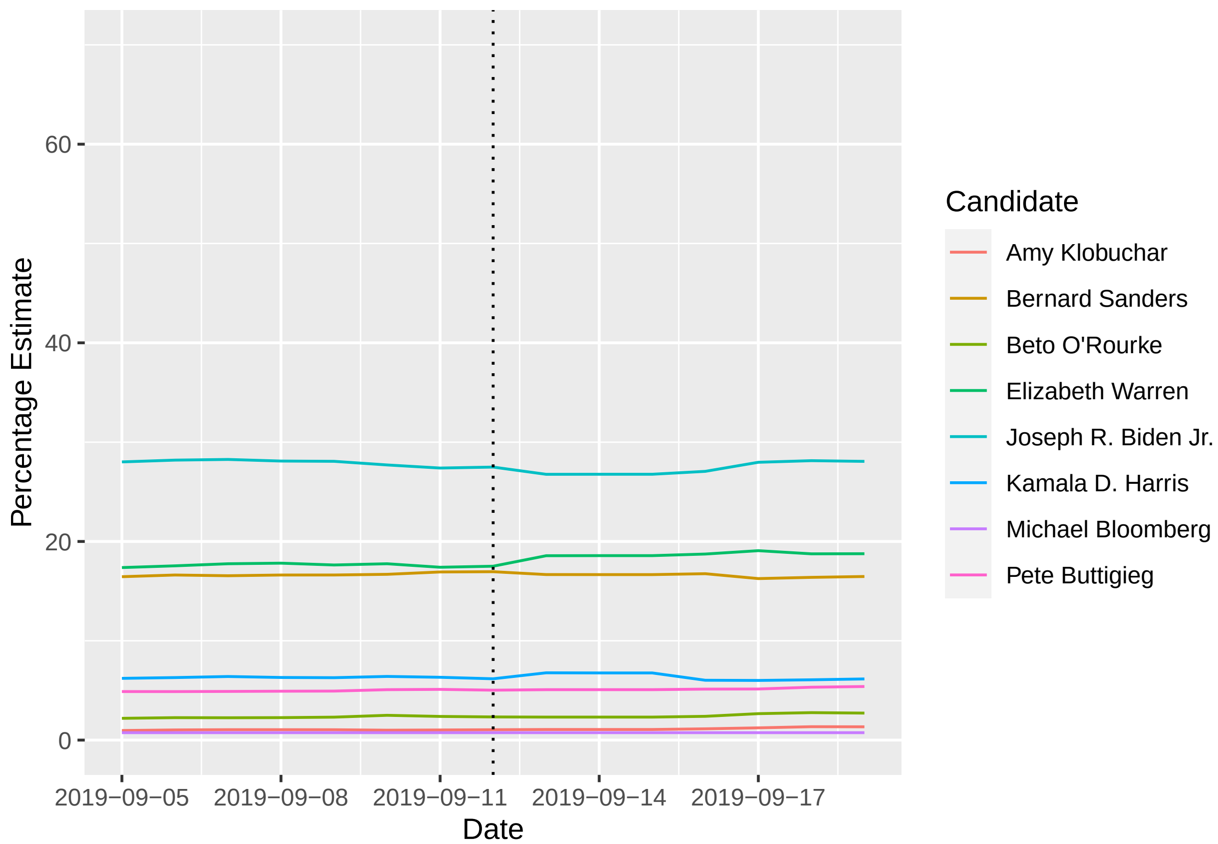
My findings do not necessarily imply that the quality of debates have become worse; if the goal of a debate was to put your ideas over to the public clearly, one would indeed expect candidates to be as succinct as possible in their choice of words. Yet the fact that Republican candidates in particular have a drastically increased rate of reading grade drop compared to Democratic candidates suggests that this is not the case, since a shift towards concise speech should reflect equally in either party. More evidence against this theory can be found in the increase of anger sentiment in the speeches of Republican candidates, which perhaps hints that rather than a progression borne of necessity, the inherent nature of presidential debate has been changing towards something more visceral.

Appendix

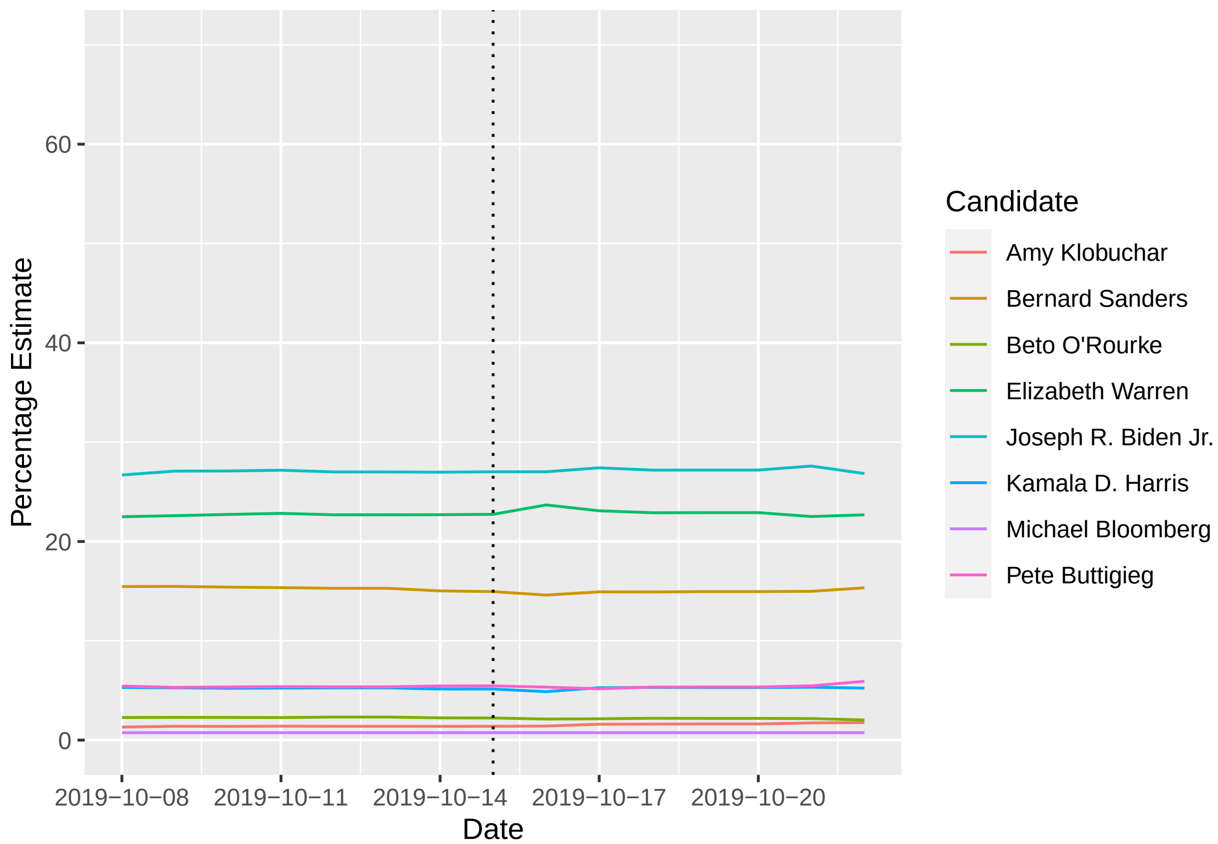
Graphs tracking polling aggregate change for the 2020 Democratic Primary

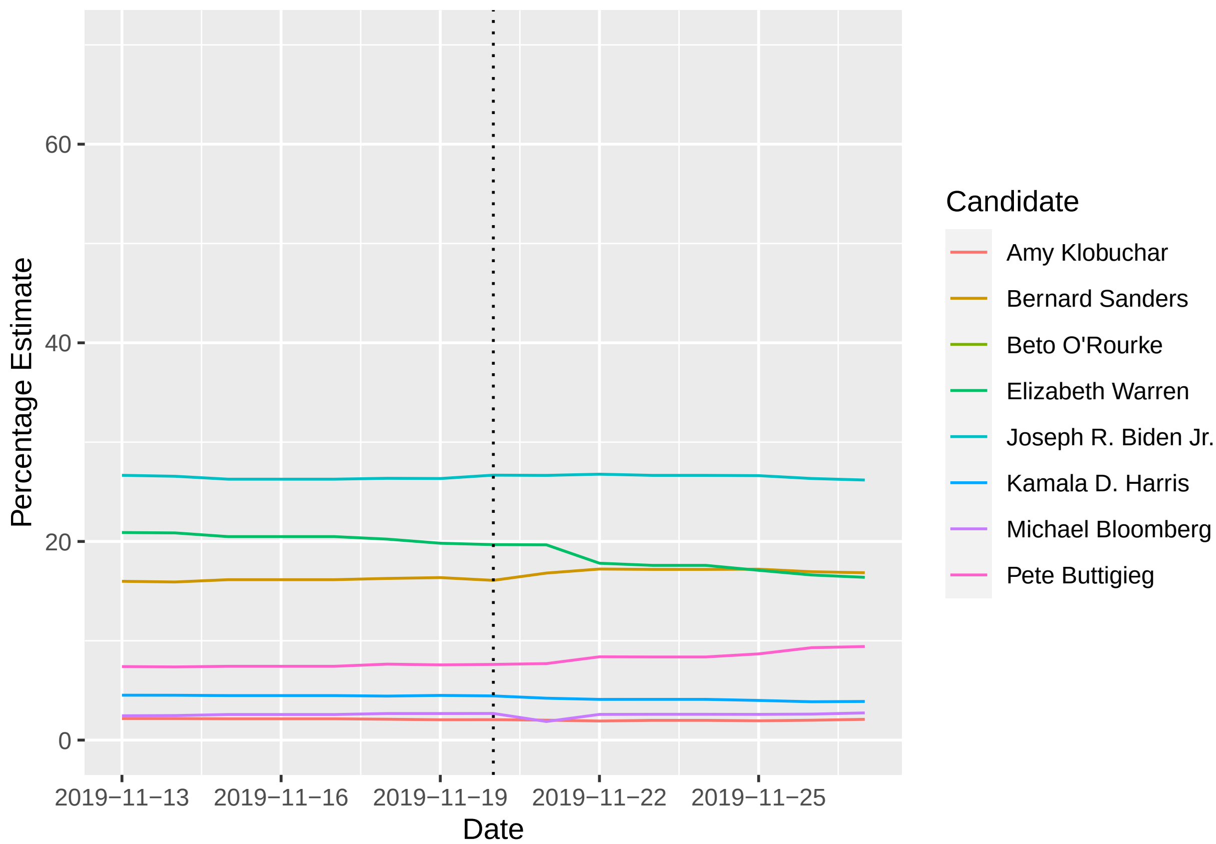


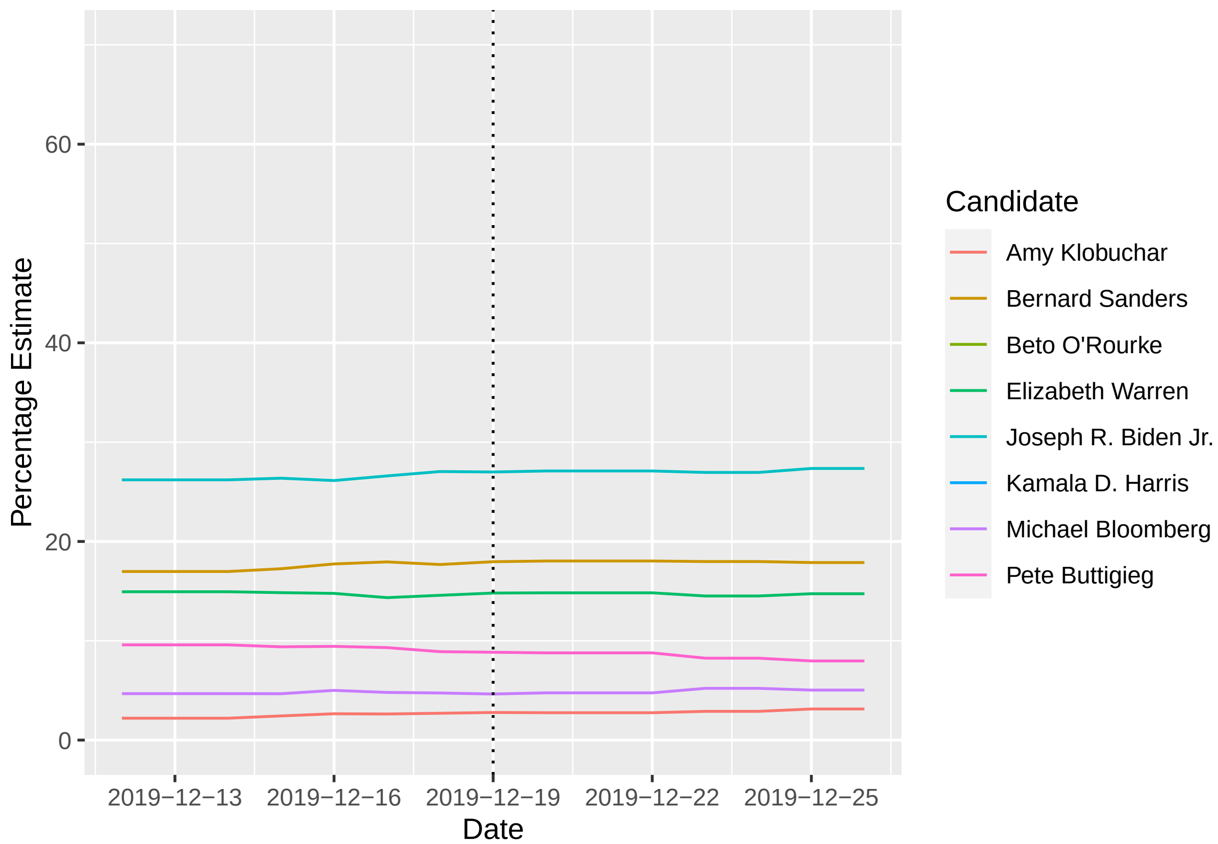




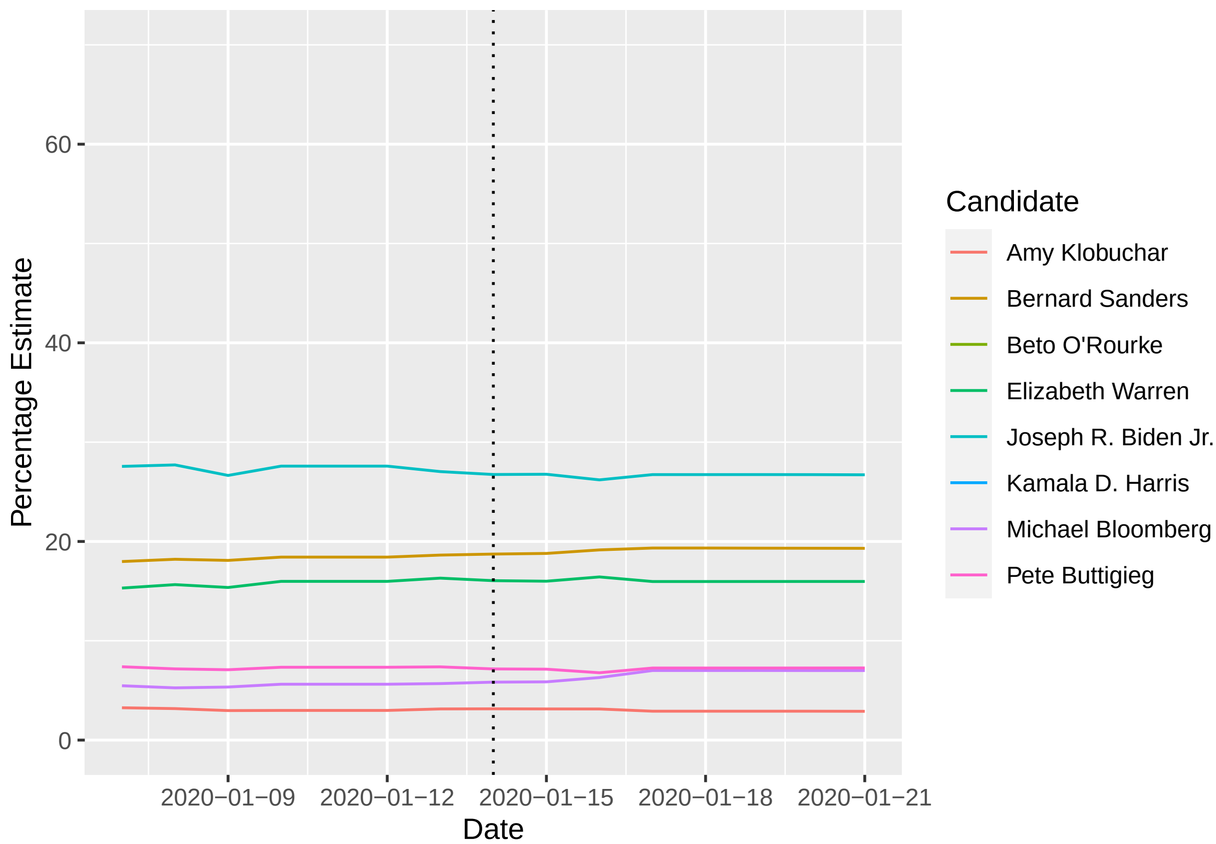
60
40
20
0
Percentage Estimate
2019−09−05
2019−09−08
2019−09−11
2019−09−14
2019−09−17
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg

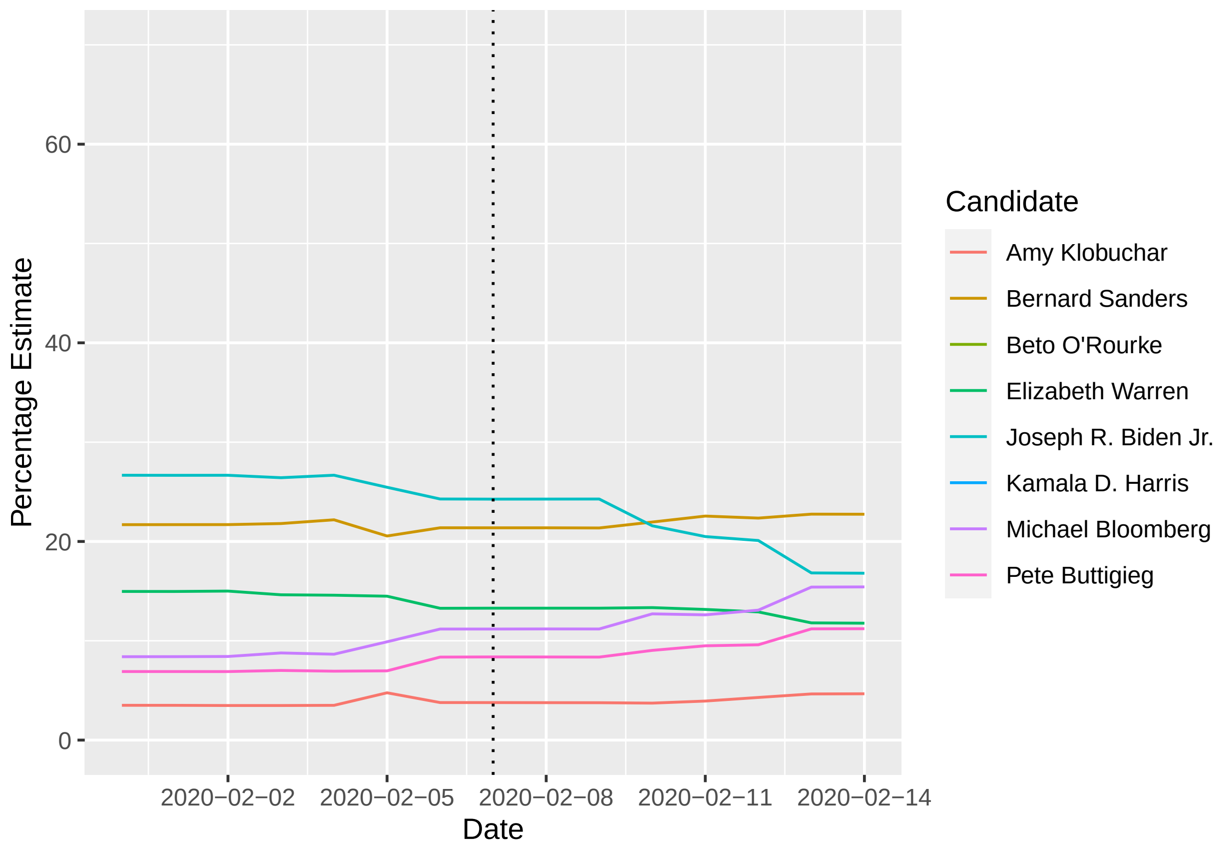

Percentage Estimate
60
40
20
0
2019−10−08
2019−10−11
2019−10−14
2019−10−17
2019−10−20
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg

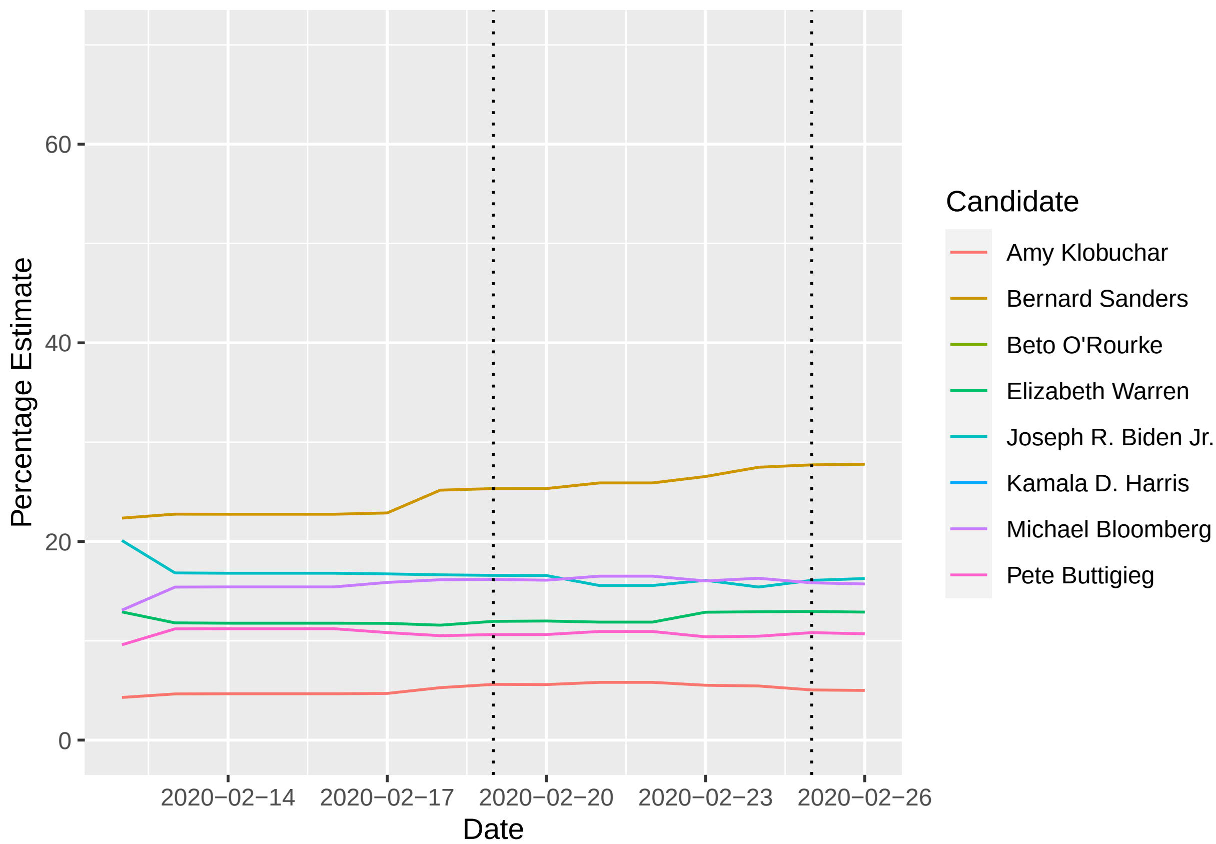

Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg
Percentage Estimate
60
40
20
0
2019−11−13
2019−11−16
2019−11−19
2019−11−22
2019−11−25
Date

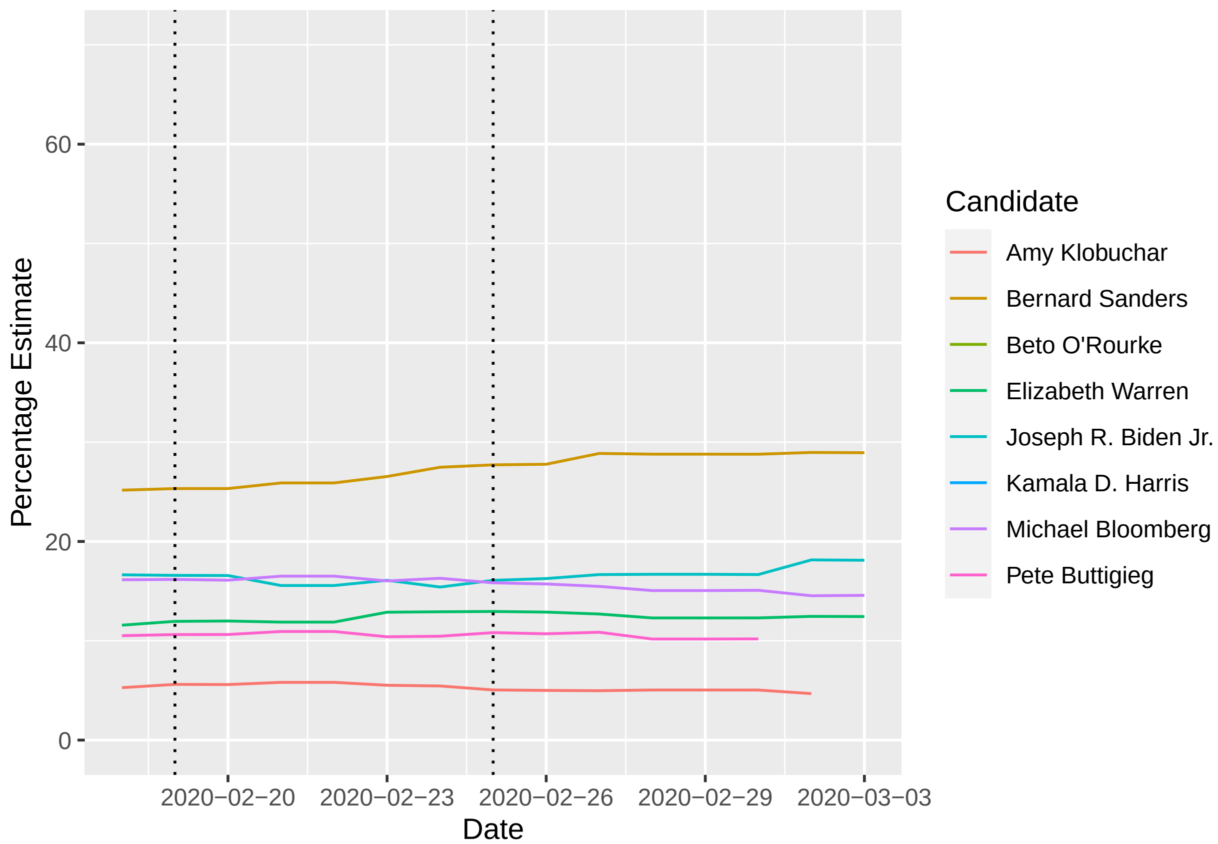


Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg
Percentage Estimate
0
20
40
60
2019−12−13
2019−12−16
2019−12−19
2019−12−22
2019−12−25
Date

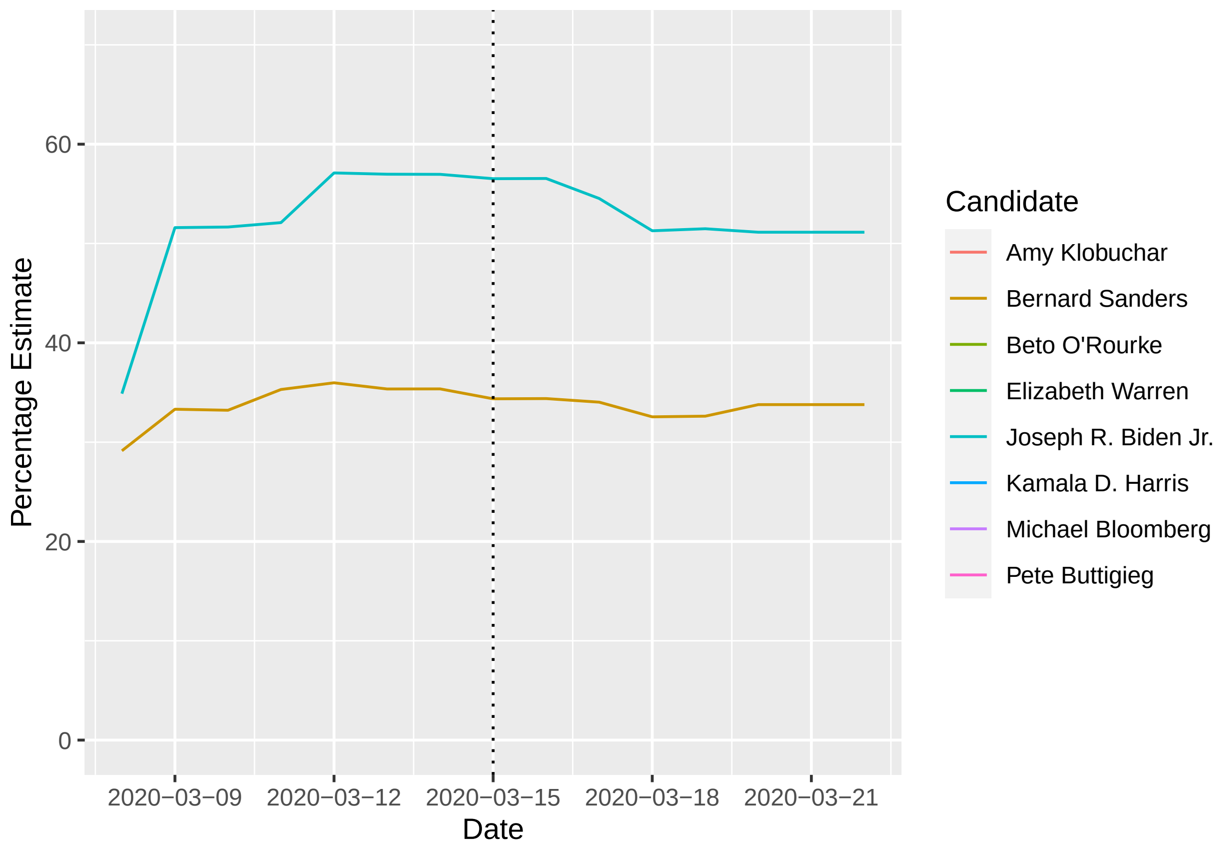

Percentage Estimate
60
40
20
0
2020−01−09
2020−01−12
2020−01−15
2020−01−18
2020−01−21
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg


Percentage Estimate
0
20
40
60
2020−02−02
2020−02−05
2020−02−08
2020−02−11
2020−02−14
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg


Percentage Estimate
60
40
20
0
2020−02−14
2020−02−17
2020−02−20
2020−02−23
2020−02−26
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg



Percentage Estimate
60
40
20
0
2020−02−20
2020−02−23
2020−02−26
2020−02−29
2020−03−03
Date
Candidate
Amy Klobuchar
Bernard Sanders
Beto O'Rourke
Elizabeth Warren
Joseph R. Biden Jr.
Kamala D. Harris
Michael Bloomberg
Pete Buttigieg



ANOVA Summary comparing Percentage Estimate between Debate Weeks and Normal Weeks

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## prepost_debate    1     33   33.44   19.18 1.23e-05 ***
## Residuals      2887   5035    1.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary of Linear Model for Readability of Debate text over time

```
##
## Call:
## lm(formula = readability ~ days + Party, data = plot_read)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1354 -1.1006 -0.3782  1.1400  3.1362
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.133e+01  4.368e-01  25.948  < 2e-16 ***
## days        -2.240e-04  4.232e-05  -5.293 2.35e-06 ***
## PartyREP    -1.281e+00  3.902e-01  -3.284  0.00182 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.46 on 53 degrees of freedom
## Multiple R-squared:  0.4226, Adjusted R-squared:  0.4008
## F-statistic:  19.4 on 2 and 53 DF,  p-value: 4.77e-07
```

Summary of Linear Model for Sentiment of Debate text over time - Bing

```
##
## Call:
## lm(formula = proportion ~ Year + Party, data = plot_sent_N)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.099476 -0.030408 -0.001718  0.028961  0.111164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.549196   2.048427  -0.756    0.459
## Year         0.001022   0.001026   0.996    0.332
## PartyREP     0.004015   0.025962   0.155    0.879
##
## Residual standard error: 0.06089 on 19 degrees of freedom
## Multiple R-squared:  0.05075,    Adjusted R-squared:  -0.04917
## F-statistic: 0.5079 on 2 and 19 DF,  p-value: 0.6097
```

Summary of Linear Model for Sentiment of Debate text over time - NRC, Fear

```
##
## Call:
## lm(formula = proportion ~ Year + Party, data = plot_sent_all_N)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0235309 -0.0135378 -0.0002526  0.0099096  0.0304290
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3400589  0.5125435   0.663    0.515
## Year        -0.0001302  0.0002568  -0.507    0.618
## PartyREP     0.0101207  0.0064960   1.558    0.136
##
## Residual standard error: 0.01523 on 19 degrees of freedom
## Multiple R-squared:  0.1238, Adjusted R-squared:  0.03157
## F-statistic: 1.342 on 2 and 19 DF,  p-value: 0.2849
```

Summary of Linear Model for Sentiment of Debate text over time - NRC, Disgust

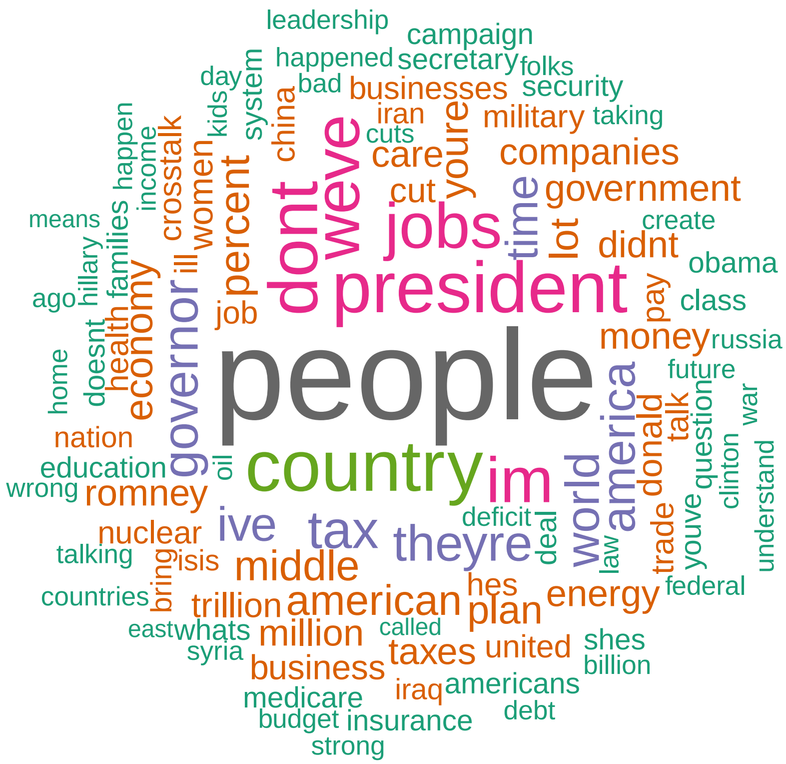
```
##
## Call:
## lm(formula = proportion ~ Year + Party, data = plot_sent_all_N)
```

```
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0145013 -0.0036297  0.0004128  0.0028331  0.0231165
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.846e-02  2.623e-01   0.299   0.7681
## Year        -2.378e-05  1.314e-04  -0.181   0.8583
## PartyREP     7.179e-03  3.325e-03   2.159   0.0438 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007798 on 19 degrees of freedom
## Multiple R-squared:  0.1981, Adjusted R-squared:  0.1137
## F-statistic: 2.347 on 2 and 19 DF,  p-value: 0.1227
```

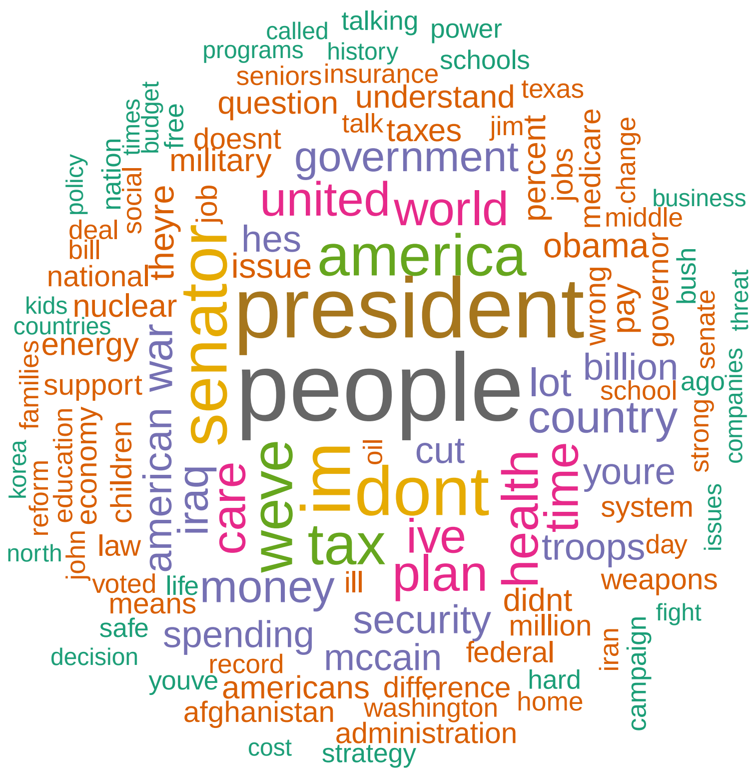
Summary of Linear Model for Sentiment of Debate text over time - NRC, Anger

```
##
## Call:
## lm(formula = proportion ~ Year + Party, data = plot_sent_all_N)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011946 -0.006023 -0.001816  0.003875  0.025572
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5138168  0.3113126  -1.650   0.1153
## Year         0.0002869  0.0001560   1.840   0.0815 .
## PartyREP     0.0068747  0.0039456   1.742   0.0976 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009253 on 19 degrees of freedom
## Multiple R-squared:  0.2526, Adjusted R-squared:  0.1739
## F-statistic:  3.21 on 2 and 19 DF,  p-value: 0.06293
```

Wordclouds by Decade: 2010s

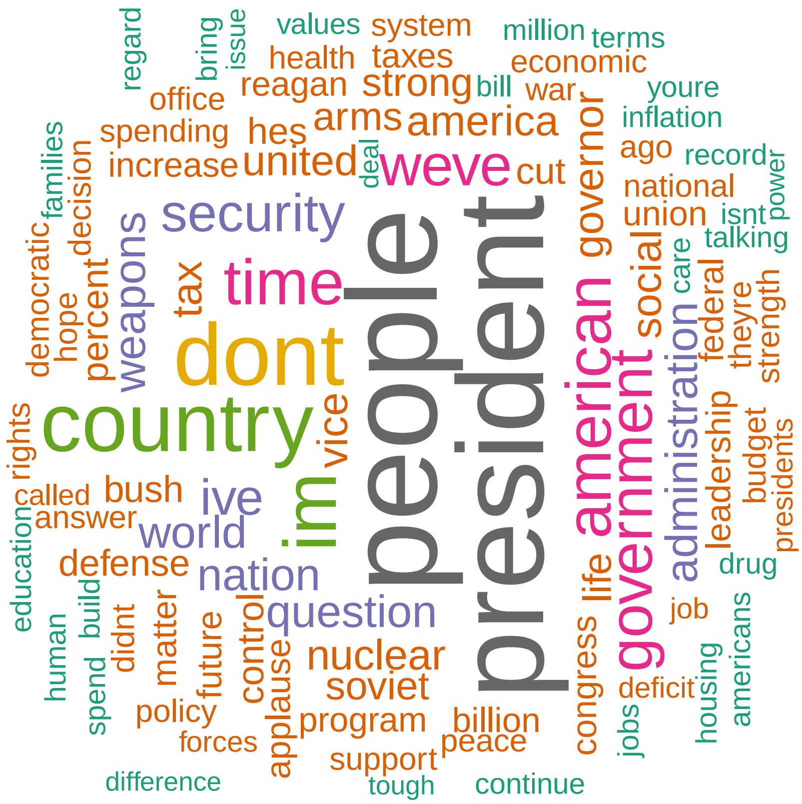


Wordclouds by Decade: 2000s



Wordclouds by Decade: 1990s

Wordclouds by Decade: 1980s



Wordclouds by Decade: 1970s