**How Do States Vote, and Do Governors Have an Impact?**

**Ben Harwood**

**Introduction**

The United States of America constitutionally protects the right of the people to speak freely. The ability to express oneself freely and openly was so important to the Framers of the Constitution – politicians, mostly – that they listed it before the remainder of the first 10 amendments commonly referred to as the Bill of Rights. Indeed, it is this right that allows for protest – preferably peacefully – to take place, which has led to monumental change throughout the history of this Nation. From Martin Luther King Jr. to the current climate in response to the death of George Floyd in Minneapolis, free speech has defined us as a Nation. But free speech does not infer that everyone speaks for themselves. Unfortunately, some people seem to think, speak, and act a certain way in order to be representative of a group. Call it optics, and no group of people in the country exemplifies this more than politicians.

Anecdotally, it seems as if politicians are likelier than ever to toe the party line. While it is true that party lines shift and change over time, certain fundamental beliefs seem to persist within each party. In a previous work, it was hypothesized by the author (and one other) that the tone, context, and language of the party members is pervasive enough to be identified with natural language processing techniques. Indeed, it was shown that the word usage in gubernatorial inaugural speeches could predict the political affiliation of the speaker with nearly 70% accuracy.

Consider, however, that while governors are elected state-level leaders, they do not always reflect how a state votes in a presidential election. Kentucky, for example, consistently elects democratic governors but has not voted for a democrat in a presidential election since Bill Clinton and has elected Mitch McConnell to the senate for almost as long as the present author has been alive. Notwithstanding, it is not unreasonable to think that governors can impact how their states vote, not directly of course, as that could be considered election tampering. Given the tumultuous and polarizing election[[1]](#footnote-1) in 2016, it seems a good opportunity to investigate this possibility.

**Section 1: Analysis and Models**

The primary goal of this study is to determine if governors have any impact on how their state votes in presidential elections[[2]](#footnote-2). This will be done by first examining the state of the state (or commonwealth where appropriate) speeches using natural language processing, then comparing the sentiment of those speeches with how the states voted. Additionally, election data will be examined to see how the states and individual counties voted in the past two elections (2012 and 2016) in the “raw” sense and relative to their governor.

**Section 1.1: Data**

This section will examine both collections of data mentioned in the beginning of this section.

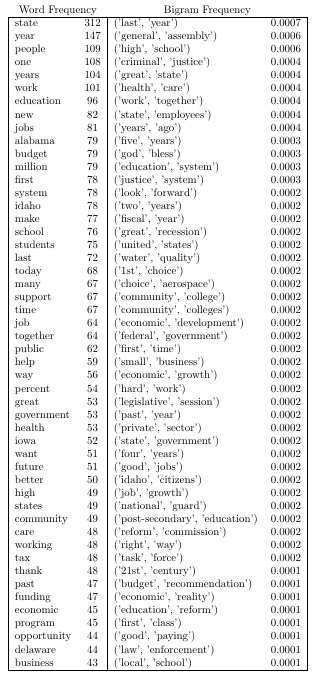
It goes without saying that since the sentiment of governor speeches is of central interest to this investigation, said speeches were needed. The search for the speeches was… arduous, as there does not seem to be any sort of repository of governor speeches, so each speech had to be manually found. This is when it was discovered that some governors did not deliver a speech in 2016 (specifically, the governors for Arkansas, Montana, Nevada, North Carolina, North Dakota, Oregon, and Texas). Additionally, Alaska was not included for reasons that arise later, and six speeches simply could not be found (Colorado, Kentucky, New Hampshire, Utah, Wisconsin, and Wyoming). Thus, the study will focus on the remaining 36 states whose speeches were acquired. Below is a wordcloud showing the 100 most frequently used words from the 36 speeches.

A close up of text on a black background

Description automatically generated

Also of note, the governors used 39,176 total (4,948 unique) words in their speeches, 48.38% of which were stop words. The 50 most frequent words and 50 most frequent bigrams appear on the following page. Notice that the word “state” was overwhelmingly the most frequent word. This is not terribly surprising since these are speeches given by governors. Notice also that the bigram “last year” was the most common, indicating that governors consistently talk about what happened in their state in the previous year. Again, this is not terribly surprising. Of interesting note, Idaho is the only state that occurred as part of a top 50 bigram. Trigrams were also looked at, but none of the trigrams were especially interesting.

Turning to the election data, as mentioned there are 159 variables included for each of the 3,143 counties in the U.S. In the interest of space (and lack of necessity) only the variables utilized are listed here:

* State
* Democrats 12 (Votes) – how many votes the state cast for the democrat in the 2012 election (Barack Obama)
* Republicans 12 (Votes) - how many votes the state cast for the republican in the 2012 election (Mitt Romney)
* Votes16 Clintonh – how many votes were cast for Hillary Clinton in 2016
* Votes16 Trumpd – how many votes were cast for Donald Trump in 2016
* FIPS – Federal Information Processing Standards, each county has a unique Fips id associated to it in the dataset, used for matching and referencing purposes
* County
* Median Earnings 2010

These variables were used to define the following variables:

* Counties – the number of counties in a specific state
* D12 – the number of counties in a state that voted democrat in 2012
* D16 – the number of counties in a state that voted democrat in 2016
* DD – the number of states that voted democrat in both elections
* R12 - the number of counties in a state that voted republican in 2012
* R16 – the number of counties in a state that voted republican in 2016
* RR – the number of states that voted republican in both elections
* Vote16 – categorical variable identifying which party a state voted for in 2016.
* AvgMedInc – average median income in the state
* DemJumpers – the number of counties that voted democrat in 2012 and republican in 2016
* RepJumpers – the number of counties that voted republican in 2012 and democrat in 2016
* DJumpRatio – the percentage of counties switching from democrat to republican
* RJumpRario – the percentage of counties switching from republican to democrat
* G-Party – The political party of a state’s governor

These created variables were collected into a data frame to be used for the rest of the investigation.

**Section 1.2: Analysis**

The data frame described in the previous subsection was used to answer a number of questions even before determining and incorporating the speech sentiment. The following table describes a question of interest and how the program answered the question.

|  |  |
| --- | --- |
| **Question** | **Method** |
| How many counties in each state switched parties in the 2016 election, more specifically what percentage of one party’s counties switched? | DemJumpers, RepJumpers, DJumpRatio, and RJumpRatio variables |
| Does the average median income in a state influence how a state votes, and/or how many counties in the state switched parties? | Linear regression with AvgMedInc predicting jump counts |
| How many counties and states did not vote along the party line of their governor? | Simple comparison |
| Is there a relationship between the party of the governor and how the state voted in 2016? | Chi-squared test |

Additionally, by calculating and incorporating the sentiment of the governor speeches into the data frame, additional questions were able to be answered:

|  |  |
| --- | --- |
| **Question** | **Method** |
| What was the average sentiment for each party? | Straightforward mean calculation |
| What states had the most and least positive speeches? | Straightforward |
| Which states (if any) had negative sentiment? | Straightforward |
| Does speech sentiment influence jump ratios? | Linear regression |

While the answers to the more detailed questions are deferred to Section 2, it seems this a good time to present some answers (if nothing else to keep the reader interested). One function built into the program allows the user to select a party and then a state and then returns information of the following sort:

“There are 62 counties in NY, with 36 of them voting democrat in 2012, and 16 of them voting democrat in 2016. Additionally, 16 of the 36 counties that voted democrat in 2012 also voted democrat in 2016.”

Clearly in this case the user entered democrat and the state of NY. For comparison:

“There are 62 counties in NY, with 26 of them voting republican in 2012, and 46 of them voting republican in 2016. Additionally, 26 of the 26 counties that voted republican in 2012 also voted republican in 2016.”

This function effectively answers the first question. Overall, 69.44% of states voted for the party of their governor. Additionally, 33.33% of states with democrat governors voted for Hillary Clinton, while 87.50% of states with republican governors voted for Donald Trump. Additionally, the average speech sentiment for states led by democratic governors was 18.33% and the average speech sentiment for states led by republican governors was 19.83%. This was initially a bit surprising as it somewhat goes against some preconceived notions of the author, but upon further reflection this does make sense. And continuing with the sentiment related questions, the state with the most positive speech was Idaho at 36.34% and the state with the least positive speech was Pennsylvania at -3.22%. The average overall sentiment score was 19.33%, and both Pennsylvania and Maine had negative sentiment scores.

**Section 2: Results**

To determine if there is a relationship between governor party and the party a state voted for in 2016, a chi-squared test was performed on the “G-Party” and “Vote16” variables with the following:

*H*0: G-Party and Vote16 are independent

*H*1:G-Party and Vote16 are not independent, i.e. the two variables are related

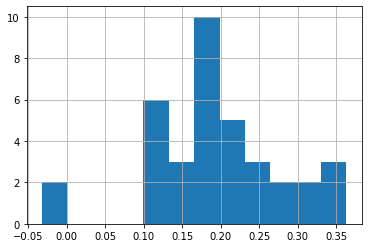
Assuming a significance level of α = 0.05, the chi-squared test is not significant, with p = 0.13652, thus the null hypothesis that governor party and voting party are independent cannot be rejected.

Linear regression was explored to measure if average median income could predict the number of counties that would jump. The following plots show median income vs jumpers.

|  |  |
| --- | --- |
|  |  |

Linear regression was performed for the DemJumpers, with *F*(1, 35) = 15.78, *p* = 0.0003, *R2* = .291, so while the model is significant, median income only accounts for 29.1% of the variation in the number of DemJumpers. Similarly, regression for the RepJumpers yields *F*(1, 35) = 6.464, *p* = 0.0156, *R*2 = 0.132. Again, a significant regression, but very little variance is accounted for. Thus, it is difficult to contend for a strong influence of median income of the number of party-jumping counties.

It is at this point that the author acknowledges not much has been said about the sentiment of the speeches (except for the brief mention at the end of section 1.2). To that end, here is a histogram showing the distribution of sentiment scores:



Here are plots of sentiment vs jumpers. Notice a potentially linear relationship for the DemJumpers (though with negative direction).

|  |  |
| --- | --- |
|  |  |

Linear regression was performed to predict DemJumpers from Sentiment. Regression yields

*F*(1, 35) = 13.47, *p* = 0.0007, *R*2 = 0.257. This is once again a significant regression, with 25.7% of overall variance accounted for. Considering the RepJumpers, regression yields *F*(1, 35) = 4.629, *p* = 0.0384, *R*2 = 0.092.

None of the previous regression models were especially strong. However, notice that both median income and sentiment had similar *R2* values when predicting DemJumpers. Linear regression was again performed, this time with sentiment and median income predicting DemJumpers. This resulted in a model with *F*(2, 34) = 7.760, *p* = 0.0017, *R*2 = 0.273, indicating that combining the two actually was less effective than just using median income.

**Conclusion**

Ultimately, how a state votes is really up to the voters in that state. Yet the results from section 2 indicate that while the party of the governor was not impactful on the party a state voted for, yet there was a relationship (albeit small) between the sentiment of the governor’s speech and how many counties changed parties.

It was initially hoped that a model could be built to use in predicting how states would vote in the upcoming election using the speech sentiment, however given the general lack of results this was not pursued further. Besides, 2020 has been a year unlike any other, especially in the time period after the state of the state speeches would have been given. Maybe the American institution of the election is immune to outside influence.

Ok that was a joke.

1. A concerted effect has been made by the author NOT to introduce any personal beliefs or party/ideological stances in the examination. Democrat variables are listed first purely due to the author’s obsession with lexicographic order. [↑](#footnote-ref-1)
2. It is acknowledged that this will not be an exhaustive investigation. [↑](#footnote-ref-2)