

What is the Best Primary State?

Jaeyeon Won, Issac Barrash, Hunter Blum, Kyung Seo

Group 7, 480 Final Project

May 07, 2020

Abstract

Iowa, a fly-overstate notable for its centrality, corn, and lack of notable locations; every four becomes the focus of the nation as the first state to pick candidates who will for the next year, battle for presidency of the most powerful country on earth. No contest passes without criticism, however, the 2020 caucus brought criticism of Iowa's first-place position to a new height with many questioning whether Iowa should be first at all noting its lack of diversity, organization, and electoral significance. Much work has been done to discover how voters choose candidates in primaries, what qualities make a good candidate, and how to strategically apply resources to win a party race, however, the literature on how primaries themselves should be run is surprisingly sparse. We seek to contribute to this field by presenting a methodology for finding the aggregate primary state based on swing-state demographics and answer the question of which state should vote first.

Introduction

Why only one? Notably pushed by Rep. Richard Hobson many have suggested holding a national primary where everyone across the country votes for their party's candidate at the same time (Kurtzleben, 16). While simplifying the process, it is doubtful that this method is the best, as with such a wide area to cover, only candidates with large amounts of name recognition or fortunes would be able to fund a campaign that could reach all party voters. This barrier to entry does not necessarily help select the best candidate either, notably many candidates beginning without name recognition or initial funding have been able to win competitions through superior campaigning skills (Hull, 07). Most recently Barack Obama. Candidates in the most recent

primary, such as Pete Buttigieg, Amy Klobuchar, Kamala Harris, and Cory Booker were able to at least compete against figures with national name recognition like Bernie Sanders, Joe Biden, and Michael Bloomberg because of the limited number of contests at the start of the primary season. Although none won the primary, it is hard to argue the party was made weaker by giving them a greater chance. It remains clear in the staggered approach to primaries is the best, which begs the question, which states should go first? This is a very consequential question because candidates winning the first states get strong "bounces" in their national polling making it much more likely that these early winners can maintain their lead especially if the win is unexpected. Thus, the state which goes first is a sought-after position as it has disproportionately powerful weight in who becomes the party nominee.

Several articles have been published on the reformation of the nomination process based on demographics. (Skelley, 19) creates a hypothetical primary order based on the make-up of democrat voters and finds Iowa to be unrepresentative. Similarly (Khalid, 16) created an index based on which state is most nationally representative in race, education, age, income, and religion. She concludes by putting Iowa in the middle with it being representative of income, and religion, but importantly far off from race. (Lewis-Beck and Squire, 09) investigate whether Iowa is representative of the nation socially, economically, and politically. Their analysis puts Iowa at 12th, with age and race holding it back from a better spot. This spread of results shows the methodology of determining which factors to score states on is almost as important as the state demographics themselves so we will next explain in depth our process for choosing our demographic weighting scheme, score aggregation method, and prove robustness by running models with scoring modifications.

Description of Data

For our analysis, we used the *election-context-2018* dataset from the MIT Election Data Lab which contains 3114 county-level observations of vote totals in the 2012 and 2016 general elections, combined with 17 demographic variables on economics and race from the 2010 census. This dataset does not include data for Alaska nor American territories; even though these participate in primaries, none are strongly contested or carry significant electoral college weight, so we presume our conclusions to be accurate to high standards despite this omission, as the previously discussed articles also omitted these. In this report, we aggregated this

original data to the state level and dropped 3 counties with missing data. We merged an additional column which shows the number of electoral college votes for each state to the aggregated data. Also, the *Importance* variable that shows how important each state is was created based on the electoral votes and the closeness of the 2012 and 2016 elections for each state.

Analysis

1. Swing State Voters and State Aggregation

Here we must dig into what exactly we are trying to find, because the question isn't as straightforward as it first seems. As in the articles discussed so far, we could try to find the state most representative of the country (Khalid, 16), or we could find the state most representative of the democratic party (Skelley, 19). While these goals admirably focus on a democratic aggregate, they miss the pragmatic purpose of party nomination processes. Like a publicly owned company whose sole purpose is to generate shareholder profit, political parties seek first and foremost to place politicians in their party into powerful positions in order to pass policies in line with its voters. We construct our method based on the mission of a political party to win the presidential election and so attempt to order states which would choose the strongest candidate for a party.

Winning an American election isn't as simple as getting the most votes, because we use a complex system of winner-take-all state elections, and aggregate these contests to a national level in a system called the electoral college. Not all voters are created equally in

the electoral college, small states do better. A California vote translates to about 27% of a Wyoming vote, however neither political party would spend a cent to win either of these votes. That is because neither state is a swing-state: a state which could go to either party in an election. Which way these swing-states go ultimately decides every election. The greatest capability to win these swing-states is what determines the best party nominee. A party cannot account for the particulars of how each individual votes, and must focus on demographics who have correlated preferences across state lines.

1.1 Data Manipulation

Following this theory, we first needed state-level data. To get this we aggregated county-level data weighted by population. Next, to weight states according to importance, we created a score based on the two factors of importance: how many electoral college ballots a state is worth, and how close it's presidential elections were in 2012 and 2016. We gave a score here equal to how many percentage points off 50-50 the split of major party votes was. We weighted the closeness of the 2016 election with twice the weight of 2012 here because of recency and unprecedented voting patterns of the 2016 election (Lind, 16).

Because we're trying to create a measure of how relatively likely a state is to swing, we need a more complex link function; a state that voted 25-75 isn't half as likely to swing as a 50-50 state, in fact, 25-75 is a landslide state that we should assign almost no importance to. Based on our knowledge that states close to 50% are certainly swingable, but this ability to shift falls off fast indicates the link function we are looking for is an S-curve. Because we will be multiplying this by electoral college ballots, we want all scores to be between 0

and 1, as a state with small numbers of electoral college ballots should always have importance bounded by their size regardless of how close they are. The formula we decided on is:

$$\begin{aligned} & \text{Closeness} \\ &= .001 * \frac{(.5 - \text{margin}/2)^4}{(.5 - \text{margin}/2)^4 + 20} \end{aligned}$$

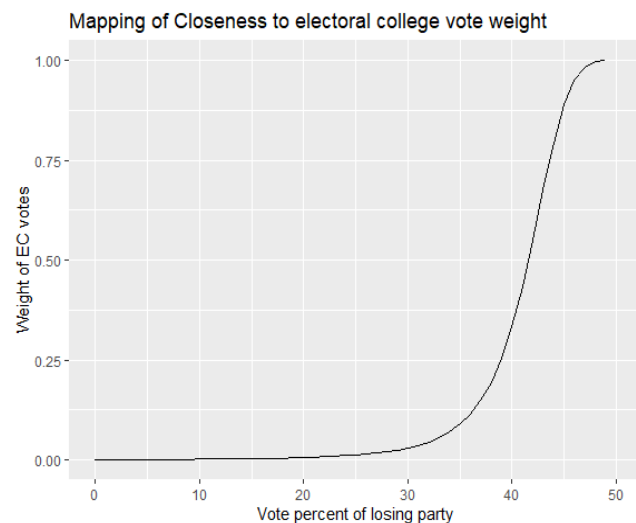


Figure 1: Closeness to Electoral College weight link function

This link function creates the curve in figure 1 above bounded between 0 and 1 for all inputs. Afterwards, we multiplied this closeness score and the number of electoral college ballots of a state to get an importance score. The distribution of states scores is in figure 3 below, which unsurprisingly places the usual suspects of Florida, Pennsylvania, Ohio, Michigan, North Carolina, Virginia, and Wisconsin all in the top ten. Notably, Texas-conventionally thought to be a state that leans relatively republican took the spot of number one importance indicating the importance our model places on the relatively close 55-45 election of the 2016 election.

We used these importance scores to weight a second aggregation. To compound state

demographics into a single set of percentages detailing what demographics are most similar to aggregate swing-states populations.

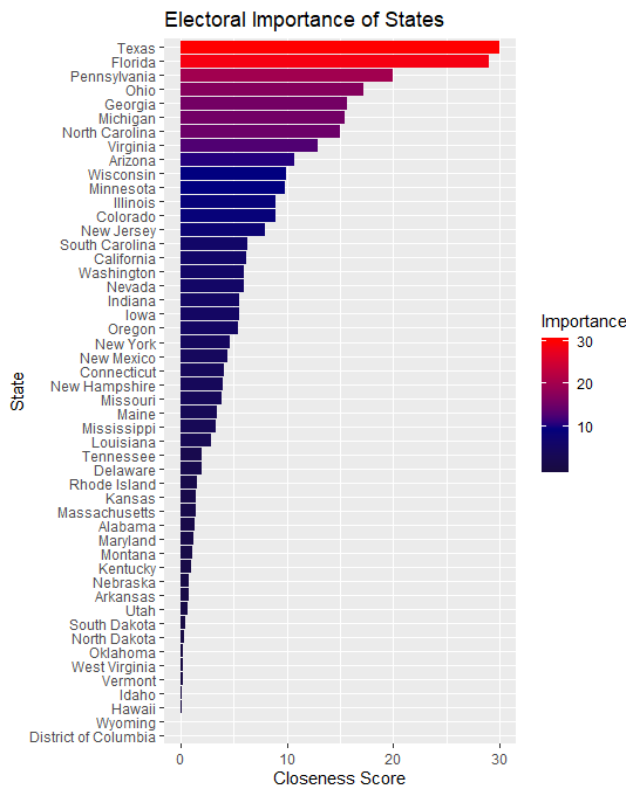


Figure 2: State Electoral Importance Chart

1.2 Results

The demographics we found to be most representative of swing states are recorded in the table below. Interestingly, we found swing states tend to be particularly diverse with populations only 65% of the population being categorized as white as opposed to the national average of 73%. Notably, most of the difference comes from more Black voters with Hispanic populations also being slightly lower than the national average. In future steps, we will use the demographics we found here to determine which states are closest to the aggregate swing state.

Aggregate Swing State Statistics	
Demographic	Percent
White	65.01
Black	15.45
Hispanic	15.46
Other	6.79
Foreign Born	11.19
Female	50.79
Age 29 and Under	39.41
Age 65 and Older	14.90
Household Income	49663
CLF Unemployment	7.38
Less than High School	12.30
Less than College	70.18
Less than HS White	7.63
Less than College White	65.29
Rural	20.63

Figure 3: Aggregate Statistics of Weighted States (Household Income is in dollars)

2. The Best Primary State

We draw from the well-established principle of the median voter theorem(Downs, 57) and apply it to demographics concluding that if we line all voters from demographics most likely to vote republican to demographics most likely to vote democrat. We posit that the key to winning is to be more persuasive than the opposing candidate to the voter in the exact middle of the state’s demographic spectrum, as the candidate will win voters more extreme on their side of the demographic spectrum, and will split the middle voters favorably to win at least a majority. We created a general scoring index that seeks to find the state which is most representative of these swing-states, as a candidate which does the best in this “generic swing-state” will do the best across all swing-states and have the best chance at winning the election.

2.1 Data Manipulation and Methods

As demographic trends are influential factors in party affiliation, determining the best primary states for Republican Party and Democratic Party was achieved by comparing the relative differences of the average demographic characteristics for each state from the aggregate demographic figures. The following descriptions provide details on this process for each party.

Considering the gap among demographic groups in partisanship, we used different modifications to the aggregate demographics found in Section 1 for each party, depending on their significance. The demographics that each party has trouble with were weighted higher by one standard deviation while the demographics that each party does good with were weighted lower by one standard deviation. Firstly, for the generic primary, we did not give any adjustments as we wanted to determine what the general best state demographics are like. Secondly, for the Democratic Party, non-white, female, and young voters are more likely to affiliate with or lean toward it. Hence, the variables “black_pct”, “hispanic_pct”, “other_pct”, “female_pct”, and “age29andunder_pct” that are less important to the Democratic Party were decreased by one standard deviation. On the other hand, they have trouble with lower-educated people, white voters, and rural areas, we increased “lesshs_pct”, “lesscollege_pct”, “lesshs_whites_pct”, “lesscollege_whites_pct”, and “rural_pct” by one standard deviation. Conversely, the demographics that the republicans have trouble with (“black_pct”, “hispanic_pct”, “other_pct”, “female_pct”, and “age29andunder_pct”) were weighted higher by one standard deviation whereas the demographics that they do good with

(“lesshs_pct”, “lesscollege_pct”, “lesshs_whites_pct”, “lesscollege_whites_pct”, and “rural_pct”) were weighted lower by one standard deviation.

With these weighted demographics, we performed the same process for the generic primary, the Democratic Party, and the Republican Party. We took the absolute values of the difference between the average state demographics and the weighted aggregate demographics for each. Before combining these calculated numbers into a scoring criterion, we divided the absolute values of the “average_median_hh_inc” variable by 3000 in order to bring this in line with other variables since this gave a skewness in the scoring criterion. In addition, we multiplied the “foriegnborn_pct” variable by 0.45 as only 45% of the foreign born population can vote. After these adjustments, adding up the absolute values for all 15 demographic indicators for each state, we created the score variable that ranks the states based on their closeness to the aggregate state.

The states with lower scores are considered the best states as it means the demographics of those states are close to the weighted aggregate demographics.

2.2 Results

• Unadjusted Primary

With the unweighted aggregate state demographics, we found the best states for the general primary. The state scores based on swing state demographics, which are rounded to the nearest integer, are shown in Figure 4 with the abbreviated state names. The states with a lower score are visualized in darker shades as they represent more importance. The state scores for the generic primary range from 31.2 to 178.1. The best primary state was Illinois with a score of 31.2. Additionally, Rhode Island, Delaware, Florida, and Connecticut are ranked in the top 5 states with scores of 43.4, 45.6, 49.1, and 52.1 respectively. New York, Washington, New Jersey, Arizona, and Oregon followed the next 5 best states with scores of 53.2, 53.7, 55.0, 56.9, and 58.4. From the relatively wide

gap of 12.2 between the best state score and the second-best state score, Illinois is considered as a significantly representative state for this unadjusted generic primary.

• Republican Party

The best states for the Republican Party were determined by increasing the preferred demographics that the republicans have trouble with. Figure 5 provides a visualization of the state scores for the Republican Party that are rounded to the nearest integer with the top 5 states highlighted. On the scale of 42.4 to 192.9, Illinois received the lowest score of 42.4 (being the best state), which is sequentially followed by New Jersey, New

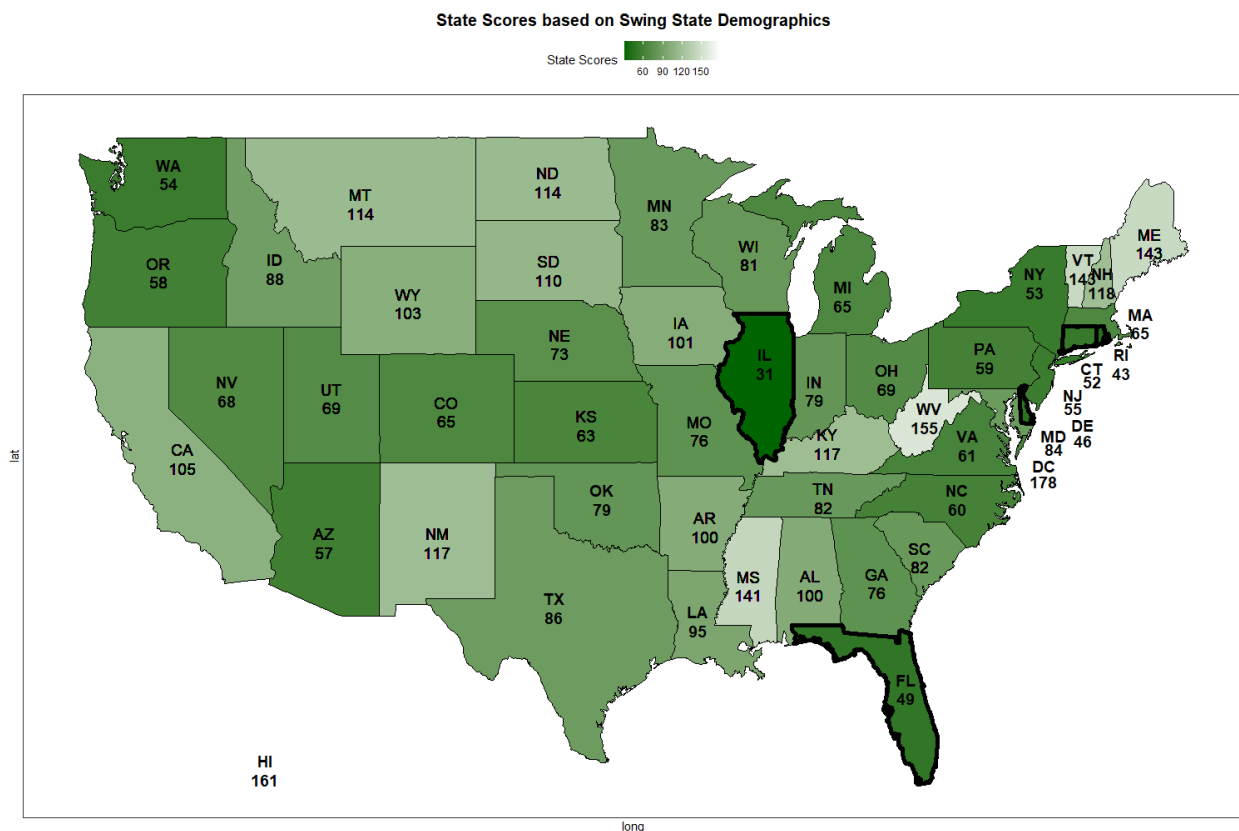


Figure 2: Representativeness Map of Generic Swing States

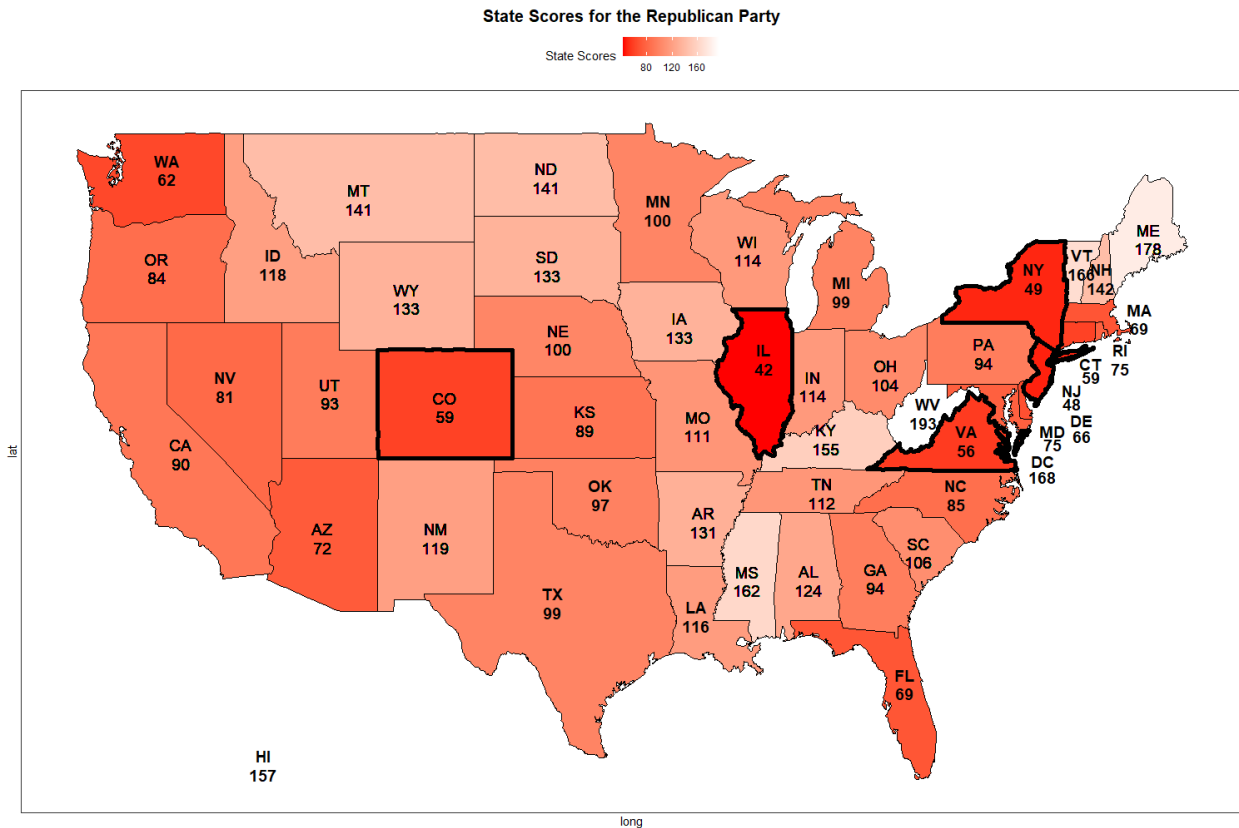


Figure 3: Representativeness Map of Republican Adjusted Swing States

York, Virginia, and Colorado with scores of 47.6, 48.9, 56.4, and 59.2 respectively. The states ranked 6 through 10 for the Republican Party are Connecticut with 59.2, Washington with 62.4, Delaware with 66.4, Massachusetts with 69.0, Florida with 69.4.

● Democratic Party

The results after weighting more to the demographic variables that are less important to the Democratic Party are shown in the map of state scores for the Democratic Party (Figure 6). Rhode Island was ranked as the best primary state for the Democratic Party with a score of 57.6. Sequentially, Delaware, Indiana, Pennsylvania, and Oklahoma ranked from 2nd to 5th with scores of 64.8, 64.9, 66.3, and 66.4. Also, Ohio, Illinois, Nevada, Oregon, and Washington were high ranked (6th to 10th) with the scores of 67.0, 69.4, 71.2, 71.3, and 71.3 respectively.

Overall, Illinois is the best state for the Generic Primary, the Republican Party, and the Democratic Party as it ranked high in all of them. Illinois ranked highest both in the Generic Primary and the Republican Party and 7th in the Democratic Party. Similarly, Delaware is an important state for all of them as it received the 3rd in Generic Primary, 8th in the Republican Party, and 2nd in the Democratic Party. Interestingly, even though the exact rankings of the Generic Primary and the Republican Party are different, they generally show a similar pattern as their top 10 rankings include many similar states. However, the Democratic Party has some states that best represent itself solely. For example, Indiana is a significant

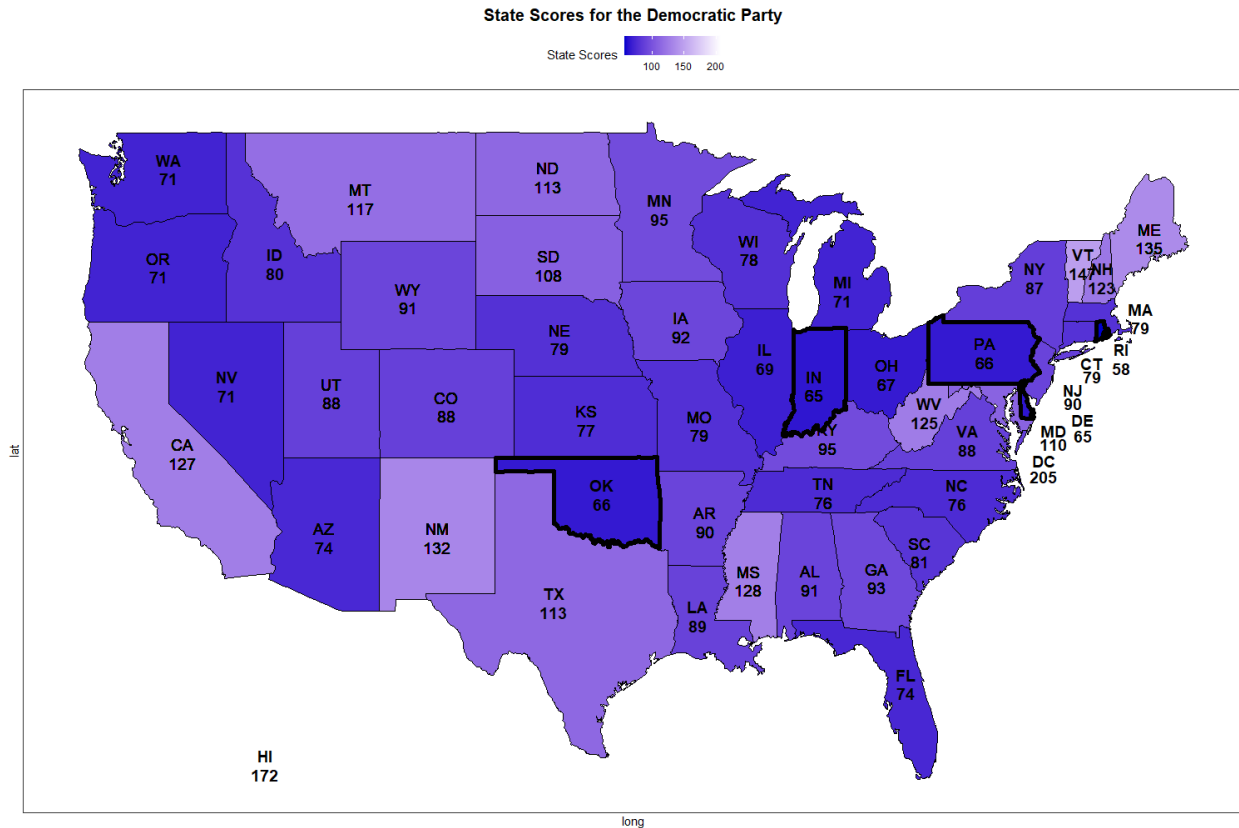


Figure 4: Representativeness Map of Democratic Adjusted Swing States

representative of the Democratic Party only as it ranked near highest (3rd) in the Democratic Party while being ranked as the 25th and 32nd state in the General Primary and the Republican Party respectively. Likewise, with its high ranking in the Democratic Party (5th) and low rankings in the Generic Primary (24th) and the Republican Party (22th), Oklahoma represents the Democratic Party solely well.

3. The Best Primary State for the Country

The best primary state for the country regardless of the bipartisan system is the state that most accurately captures the political snapshot of the nation and most accurately predicts the outcome of the election. This is an important metric to track because we want

to see how close our previous results were to the demographics of the country as a whole. By looking at all people in the country, the result here will show us the most democratic state to go first in an election. We only used the 2016 election because of the fast-changing nature of the political climate, and change our state selection method as a robustness check on the previous results in two ways. If we achieve similar findings to part 2, then we can be certain our prescriptions are resilient to unstable assumptions. Additionally, because the articles in the literature answered this question, we can also check the quality of our data by reconstructing their answer.

3.1 Data Manipulation and Methods

Before we were able to start our analysis, we had to compute the demographic of the 2016

election. This was done by combining the data we had for all 50 states and calculating the

Variable Weights After Normalization	
Demographic Variable	Weight
White	0.991
Black	1.000
Hispanic	0.968
Non-White	0.975
Foreign Born	0.975
Female	0.975
Age 29 and Under	0.975
Age 65 and Older	0.975
Household Income	0.975
CLF Unemployment	0.974
Less than High School	0.975
Less than College	0.975
Less than HS White	0.975
Less than College White	0.975
Rural	0.975

Figure 5: Normalization Weights

weighted average.

The method of calculating similarity was done by calculating the average percent difference of every state compared to the nation for all of the variables. In order to give more weight to the more important factors of the election, we used a logistic regression model (where the independent variables were the 15 variables in our dataset) and the dependent variables is the winning party for each state (where 0 represents democrats and 1 represents republicans). We then normalized the coefficients of the model from 0 to 1, and this normalized value became the weight of the variable.

Using this weight, we computed the similarity score which was the weighted average of the percent difference for each state compared to the country.

3.2 Results

The results of the logistic regression model are listed in figure 8 below.

The intercept value does not provide us with much useful interpretation because it is impossible for the input to have zeros. However, the intercept is the lowest coefficient in this model, which will be the 0 after the normalization.

	coef	se(coef)
(Intercept)	-9.062693e-01	1.051816e+01
white_pct	1.516398e-02	1.304757e-01
black_pct	2.354192e-02	7.061582e-02
hispanic_pct	-6.171023e-03	7.604820e-02
nonwhite_pct	3.819136e-05	2.142920e-04
foreignborn_pct	-1.342557e-04	6.978361e-04
female_pct	-1.538708e-07	5.300804e-05
age29andunder_pct	-1.396894e-06	1.496082e-04
age65andolder_pct	9.481614e-06	3.783133e-04
average_median_hh_inc	2.372275e-08	2.162280e-07
clf_unemploy_pct	-2.298555e-04	1.955563e-03
lesshs_pct	2.186848e-05	6.073492e-04
lesscollege_pct	1.662629e-05	1.423582e-04
lesshs_whites_pct	-1.496153e-04	1.274923e-03
lesscollege_whites_pct	8.845798e-07	4.360933e-05
rural_pct	-1.194016e-06	9.898096e-05

Figure 6: Regression Coefficients

We can see that race played the largest role in the election from the highest value by the white and black_pct. Using these values, we can find the weighted percent difference between the individual states and the country. Here are the top 10 states and the graph for

all 50 that are the most similar compared to the entire country.

Most Representative States		
Rank	State	Score
1	Illinois	9.1
2	Virginia	13.2
3	New York	13.7
4	Delaware	14.6
5	Connecticut	15.9
6	North Carolina	18.5
7	Washington	18.7
8	Florida	18.7
9	Rhode Island	19.9
10	Pennsylvania	20.3

Figure 9: Top 10 states with demographics most similar to the country as a whole.

The score can be interpreted as the average percent difference to the country. Illinois came as the state that is the most similar to the nation according to the 2016 presidential election with only a 9.07% difference, which makes them the best primary state for the country. Virginia, New York and Delaware came close being only 4-5% higher than Illinois.

The results obtained here are consistent with other research conducted in the past. An article written by Khalid from NPR summarizes a research conducted using census data to see which state is the most similar to the entire country. Although they had an additional factor of religion, they also found Illinois to be the most overall similar state, and had Virginia and Delaware being most similar in Education and Age respectively. Our results confirming previous findings gives us strong indications that our chosen dataset is both useful to our goal, and our choice of variables aligns with expectations of other researchers. The similar results over different methods shows and our conclusions are robust to criticisms of methodological assumptions.

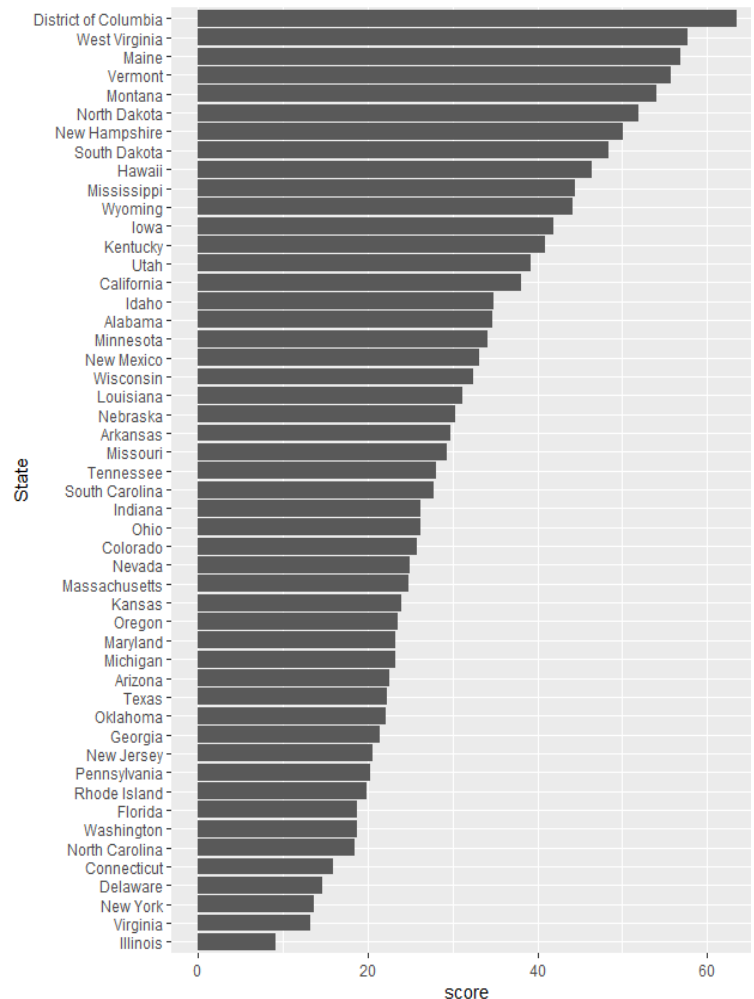


Figure 10: Ranking of states by similarity to the country as a whole.

4. What Makes the Best Third Party State?

Because “third party” encompasses a variety of political beliefs and demographics, our previous adjustment method would not be practical. Also, since third parties often do not receive electoral votes, the optimal state will draw the most attention to the third-party candidates. In order to answer this unique question, we will create models that predict the third-party votes in the 2012 and 2016 presidential elections. Then, we will compare these models with the actual elections and use

Coefficients	Estimate	Standard Error	t-value	p-value
Intercept	-1.85×10^{-4}	5.109×10^{-1}	0	.9997
Percent White	1.016×10^{-4}	4.985×10^{-5}	2.037	.0479
Percent Foreign Born	3.449×10^{-4}	1.731×10^{-4}	1.993	.0528
Percent Whites Less than High School	-6.472×10^{-4}	2.489×10^{-4}	-2.6	.0128
Percent Rural	1.347×10^{-4}	7.127×10^{-5}	1.89	.0656

Figure 7: Results of the stepwise model predicting third-party votes based on the 2012 presidential election.

real world context in order to determine both where third parties should hold their first primary if they have multiple candidates, and where they should focus their campaign to maximize votes.

4.1 Data Manipulation and Methods

To begin our analysis of the third-party data, we analyzed the results of both the 2012 and 2016 elections. We created maps for both of the elections utilizing ggplot2 in order to visualize what percent of people voted third party in each state (Wickham 2019).

To create a prediction for future elections, we constructed two models in order to determine what characteristics correlate with significant change in the percent of third-party presidential voters. To create each model, we utilized the MASS package in RStudio to perform a forward and backward stepwise regression and select the best model (Venables and Ripley 2002). To determine which states were the best based on our model, we calculated the predicted percentage of third-party voters for every state in both models and created a map in ggplot2 (Wickham 2019).

Finally, we compared our model to the actual real-world elections in order to determine their effectiveness. This allowed us to utilize our models to combine with real world

context to reach our conclusion on the best third-party state.

4.2 Results

In 2012, we found that as the percentage of the population that is white, foreign-born, or rural increases, the percent of third-party voters increases. On the other hand, as the percentage of the population that is white, has less than a high school diploma, or is female increases, the percentage of third-party voters decreases. We also found that our model was statistically significant ($p=.015$). The top states for our model consisted of Vermont (1.35%), North Dakota (1.21%), and Maine (1.15%).

Coefficients	Estimate	Standard Error	t-value	p-value
Intercept	-9.28	3.329	-2.787	.008
Percent Black	-9.764×10^{-4}	1.737×10^{-4}	-5.621	1.61×10^{-6}
Percent Under 29 and Unemployed	.0046	6.961×10^{-4}	6.614	6.48×10^{-6}
Percent of Civilian Labor Force Unemployed	.0051	.001730	2.941	.005414
Percent Less than High School	-.003	8.248×10^{-4}	-3.626	8.05×10^{-4}
Percent Less than College	-8.812×10^{-4}	3.451×10^{-4}	-2.553	.014594
Percent Rural	4.878×10^{-4}	1.347×10^{-4}	3.622	8.15×10^{-4}

Figure 12: Results from the 2016 stepwise model predicting third-party votes based on the 2016 presidential election.

For the 2016 presidential election, we discovered that as the percentage of the population that is under 29 and unemployed, unemployed and in the civilian labor force, and rural increases, so does the percentage of third-party voters. Finally, as the percentage of the population that is black, had less than a high school diploma, and had less than a college degree increased, the percentage of third-party voters decreased. Our 2016 model was also statistically significant ($p < .001$). Utah (7.81%), Idaho (5.21%), and Montana (4.89%) were the top three states for our 2016 model.

4.3 Comparison to Real World

When comparing the results of the actual 2012 election to the prediction of our model, our model mainly underpredicted the true value of third-party votes that each state received. The 2012 model also did not fit the data well with an adjusted R^2 value of only 0.18. Our 2016 model fit the real world much better with an adjusted R^2 value of 0.68. To decide which state would be the best for a third party to hold their first primary in, it

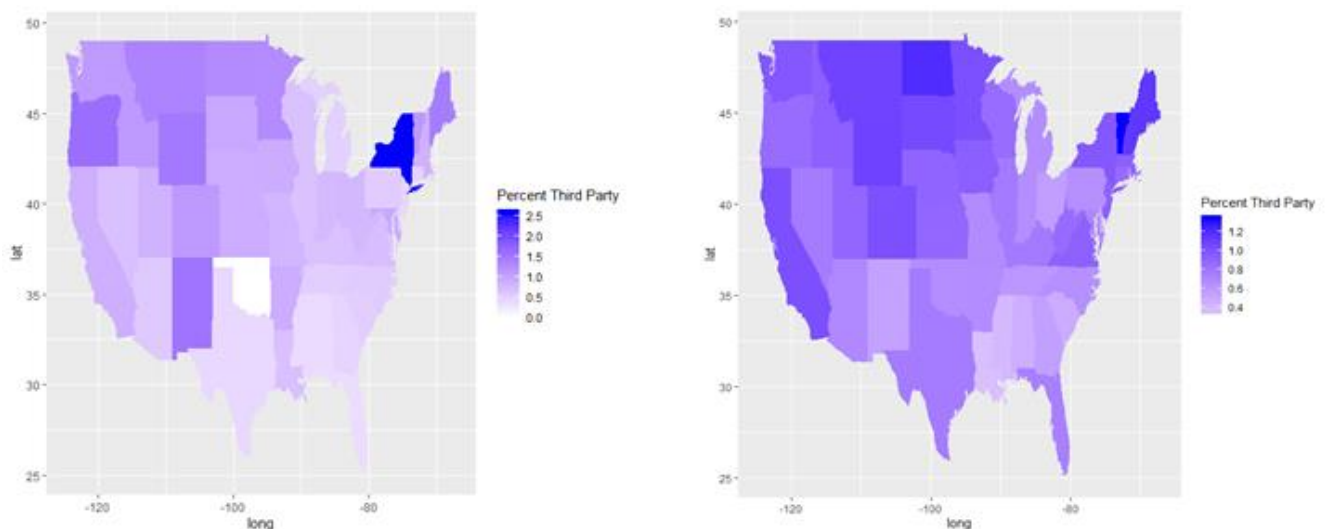


Figure 13: Maps displaying percent of third-party votes by state for the 2012 presidential election. The left is the actual election and the right is the prediction by our model.

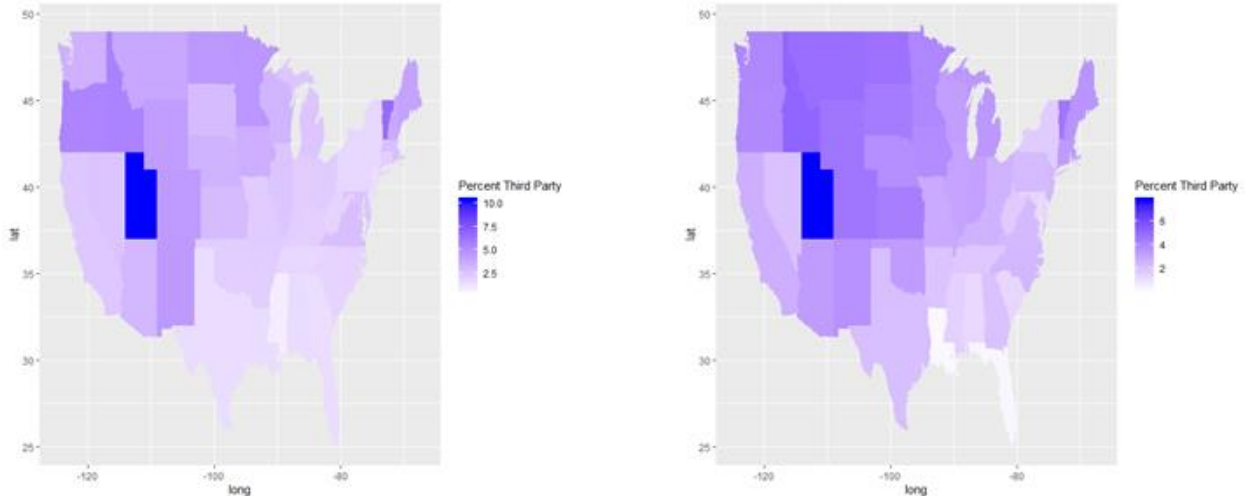


Figure 8: Maps displaying percent of third-party votes by state for the 2016 presidential election. The left is the actual election and the right is the prediction by our model.

would likely be best to hold it in the state with the highest percentage of votes in order to draw the most attention to the party platform. Both of our models predict that the states around the Rocky Mountains would have the highest percentage of votes. Unfortunately, most of these states also have a low population which would decrease the attention to the party. Because of this, we would recommend Colorado to be the first primary for a third party.

Colorado is the highest populated state in the Rocky Mountain region, with a population of over 5.5 million people. When looking into the context of elections, Colorado is also a good balance of political ideologies. Colorado is a classic example of a swing state, which will provide the symmetry to allow any third party to succeed. In both the 2012 and 2016 elections, Colorado was in the top ten for third party voting states displaying they are willing to support third party candidates more than the average state.

In the context of the two major third parties, the Green and the Libertarian parties, Colorado is also relatively even. In the past two elections, Colorado had a mostly even balance between votes. The Libertarian party

typically received a higher percentage of votes, however the Green party also received a healthy amount. Since the Libertarian candidate in both elections, Gary Johnson, was from nearby New Mexico, we anticipate the balance in votes to be even closer in future elections with different candidates.

Conclusions and Future Work

In exploring the 2012 and 2016 election data and voters' demographics, our analysis investigated the best primary state for a generic, Democratic Party, and Republican Party primary. We also looked at some factors for improving democracy as a whole like which state best represents the country as a whole, and what states are best for third parties to campaign in. Our research found that Illinois is the most representative of swing states overall. If we wished to have parties maintain the same primary schedules, this would be the best state to begin with. Illinois has the additional benefit of being the state most closely representative of the country as a whole. Illinois completes its hat trick being the first primary state for the

Republican Party as we predict it will pick the best republican candidates for winning a general election. For the Democratic Party, we found Rhode Island to be the best primary state. Lastly, Colorado was found to be the optimal state for the third parties to pick up larger vote shares.

Possible future works can utilize more analysis through integrating more information into our data. This includes combining the 2020 Census that would allow us to include more updated and various demographic variables. Also, we could project the best primary states into the future by adding the election data previous to 2012 and incorporating the trends in our closeness score. Last but not least, different weights can be given by obtaining the 2020 election data. This would allow us to see how likely the demographics are to vote.

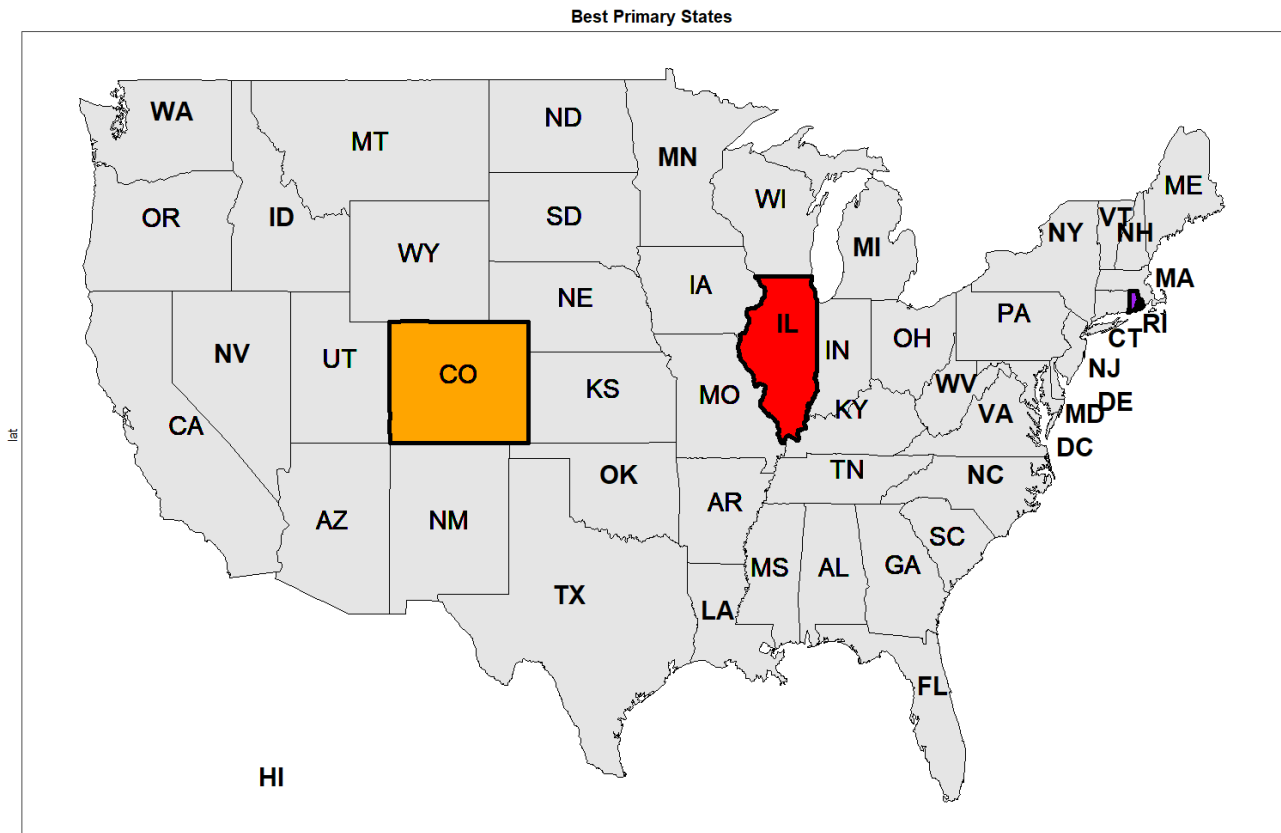


Figure 9: Map of Concluded Best Primary States. (Red-Republican and Overall, Purple-Democrat, Orange-Third Party)

Citations

- Ahmed, Washim. "USA Lat,Long for State Abbreviations." *Kaggle*, 3 Feb. 2017, www.kaggle.com/washimahmed/usa-latlong-for-state-abbreviations/data.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.
- Hull, Christopher. 2007. *Grassroots Rules: How the Iowa Caucus Helps Elect American Presidents*. Stanford, CA: Stanford University Press
- Khalid, A. (2016, January 30). The Perfect State Index: If Iowa, N.H. Are Too White To Go First, Then Who? Retrieved from <https://www.npr.org/2016/01/29/464250335/the-perfect-state-index-if-iowa-n-h-are-too-white-to-go-first-then-who>
- Kurtzleben, Danielle. "No Way To Pick A President? Here Are 6 Other Ways To Do It." *NPR*, NPR, 26 Jan. 2016, www.npr.org/2016/01/26/463870736/no-way-to-pick-a-president-here-are-6-other-ways-to-do-it.
- Lewis-Beck, M., & Squire, P. (2009). Iowa: The Most Representative State? *PS: Political Science & Politics*, 42(1), 39-44. doi:10.1017/S1049096509090039
- Lind, Michael. "This Is What the Future of American Politics Looks Like." *POLITICO Magazine*, 22 May 2016, www.politico.com/magazine/story/2016/05/2016-election-realignment-partisan-political-party-policy-democrats-republicans-politics-213909.
- Radford, Jynnah. "Key Findings about U.S. Immigrants." *Pew Research Center*, Pew Research Center, 17 June 2019, www.pewresearch.org/fact-tank/2019/06/17/key-findings-about-u-s-immigrants/.
- Skelley, Geoffrey. "We Re-Ordered The Entire Democratic Primary Calendar To Better Represent The Party's Voters." *FiveThirtyEight*, 7 Mar. 2019, fivethirtyeight.com/features/what-if-early-democratic-primary-states-looked-more-like-the-party/.
- Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>