

Election Forecasting

By Benjamin Kreiswirth, Theo Schiminovich, Mario Tutuncu-Macias

Introduction

The COVID-19 has suspended traditional campaign events and overshadowed the 2020 elections. Despite this, we have created an election forecast to predict likely scenarios of this year's presidential race.

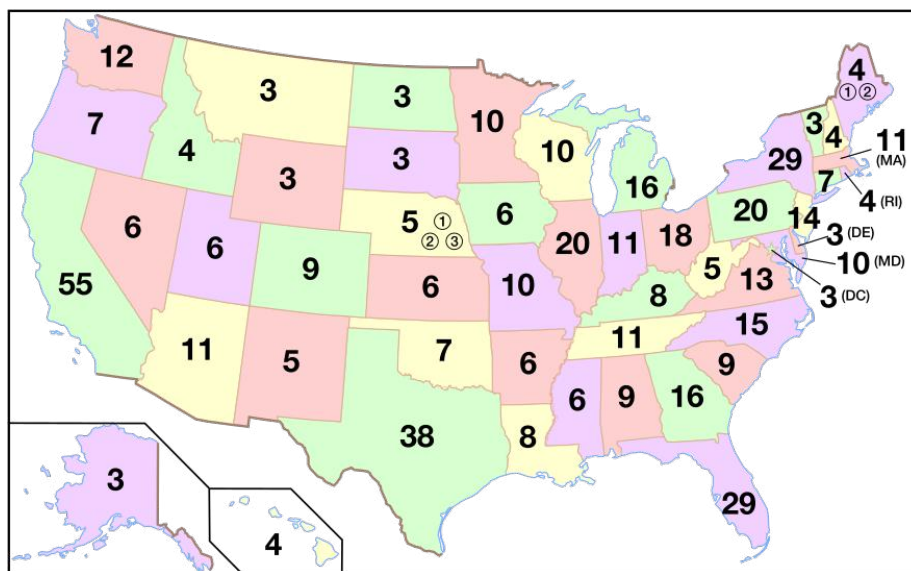
The following election forecast is based on the predicted outcome of each state. Notably, without a lot of technology or modelling, many states (at least 30) are easy to predict, as they tend to always vote for the same party.

Terminology and Background Info

Electoral College

Each state is given a certain number of electoral votes based on its population. This is determined by the 2010 census (the 2020 census will apply starting with the 2022 midterms, assuming COVID-19 does not disrupt the standard process). A majority of the electoral votes are required to win the election, and since there are only 2 major parties (the Democratic and Republican Parties) that win states (sorry Libertarians and Green Party), it amounts to winning 270 electoral votes out of the total 538 (with 269 leading to a complex congressional decision that would ultimately end in a Republican victory because there are more GOP states).

A map of the states by electoral votes is shown below. On a side note, notice how this map satisfies the 4 color theorem.



The States that matter

Many of these states are not a contest. The Democrats will win New York and Oregon, while the Republicans will win Arkansas and South Carolina. This wasn't always the case; during the mid 20th century, candidates could nearly sweep the entire country. However, today, the job of election forecasters, comes down to predicting the results in these states:

Arizona, Florida, Maine, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Virginia, and Wisconsin

You could argue that other states belong on this list, or that some states on this list don't belong here, but this is a baseline grouping of the swing states. Determining the swing states is a job in of itself, that all campaign managers will have to do when allocating their advertising budget. Some other states whose results we were wary of while creating our model included Georgia, Iowa, and Texas. However, in the end, we decided they were not going to be swing states.

Professional Modelling

Naturally, many organizations and companies want to predict the results of elections, just as we do. One of the most popular sites during the 2016 election cycle was fivethirtyeight.com. This site became famous partly for correctly predicting every state (and D.C.) in the 2012 presidential elections. They were not as accurate in 2016, correctly predicting out 46 of the 51 contests and gave a final 71.4% chance to Clinton winning.

Sites like FiveThirtyEight have taken into account far more factors than we have, but we take an in-depth look at how that affected our modelling in the last section.

Professional modellers quantify uncertainty with advanced modelling. In addition to that, uncertainty increases with time. Additionally, debates, VP announcements, COVID-19, and a host of other possibilities can create substantial changes in the upcoming election.

Primary Modelling

The primaries are arguably harder to predict than the general election, with more candidates, an election spread out over several months, and highly unpredictable voter turnouts.

FiveThirtyEight released a widely popular model. They have a fairly comprehensive article on how they forecast [here](#).

In addition to applying many of their features from above, the primary model also accounts for endorsements. For example, they predicted months before the actual event occurred that if Buttigieg were to drop out, he would likely endorse Biden.

Here are their steps, in exact words, from their site:

1. "Calculate national and state polling averages and translate them into a polling-based forecast
2. Calculate a non-polling forecast for each state based on demographics, geography and state fundraising
3. Begin simulating the rest of the primary process, starting with day-to-day movement in the polls
4. Simulate state and district results — accounting for uncertainty — and allocate delegates
5. Simulate bounces (or crashes) from winning (or losing) primaries
6. Forecast the probability that candidates drop out."

Our Modelling

We spent a lot of time working on our model, but relative to professional models, we have limited resources. That is to say, do not put your faith entirely into what our model says, especially when compared to professional models. Our goal was to emulate such models as well as we could, with a budget of \$0 and access only to publicly available census data and past voting data. In short, there is no way that our model would have the level of precision professional polls show.

Our modelling involved multiple steps, including factoring in demographics, past election results, national polling, and state-wide polling. We will go through how we used each of them in detail over the next few sections, including a final prediction and its many caveats.

Computing Demographic Leans

Our first step in modelling the election outcome was to look at a variety of demographic data for each state. This included race, college education, age, and sex. For each state, we took the percentage of the population that fell into each demographic group and multiplied it by the percentage of that demographic expected to vote for Democrats based on past elections. The sum of these products results in an expected percentage of the population of a particular state that will vote for Democrats. By comparing this to what we would expect if we performed the same process using demographic breakdowns for the entire country, we are able to compute a predicted demographic lean.

Let's do an example using breakdown by religion in New Jersey.

	New Jersey %	% Democrats	% Percent Democrats
Protestant	33%	42.9%	14.1%
Catholic	34%	50.5%	17.2%
Jewish	6%	82.3%	4.9%
Other	9%	76.0%	6.8%
Unaffiliated	18%	71.4%	12.9%
Total	100%	-	55.9%

We do the above for each state, then do the same for the United States as a whole to find a baseline.

	United States %	% Democrats	United States % Democrats
Protestant	49.4%	42.9%	21.2%
Catholic	20.6%	50.5%	10.4%
Jewish	1.9%	82.3%	1.6%
Other	5.5%	76.0%	4.1%
Unaffiliated	22.6%	71.4%	16.2%
Total	100%	-	53.5%

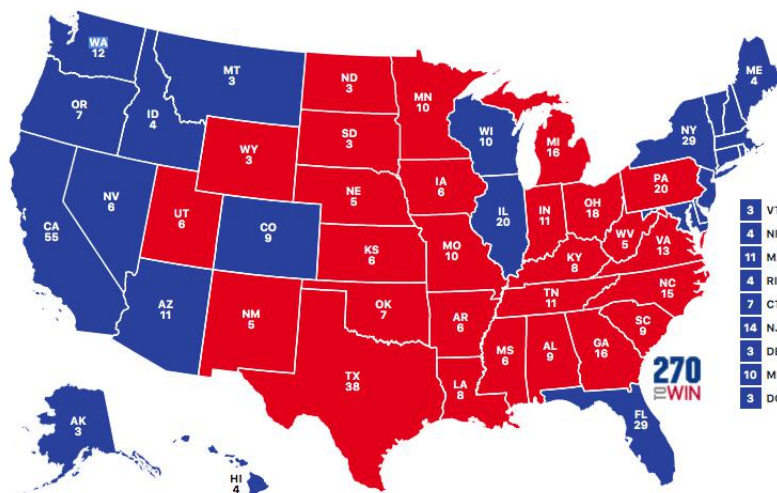
This gives us an expected religion lean for New Jersey of $55.9\% - 53.5\% = 2.4\%$, or D+2.4.

We can repeat this process for each of the demographics in each of the states to determine additional leans. Note that for the other demographics we also considered varying voter turnout among groups, which required additional computation and increased accuracy of resulting leans (e.g., old people and white people tend vote more). Also note that the maps in the following sections do not represent predictions. They merely represent parts of our investigation using the method we just discussed.

Religion

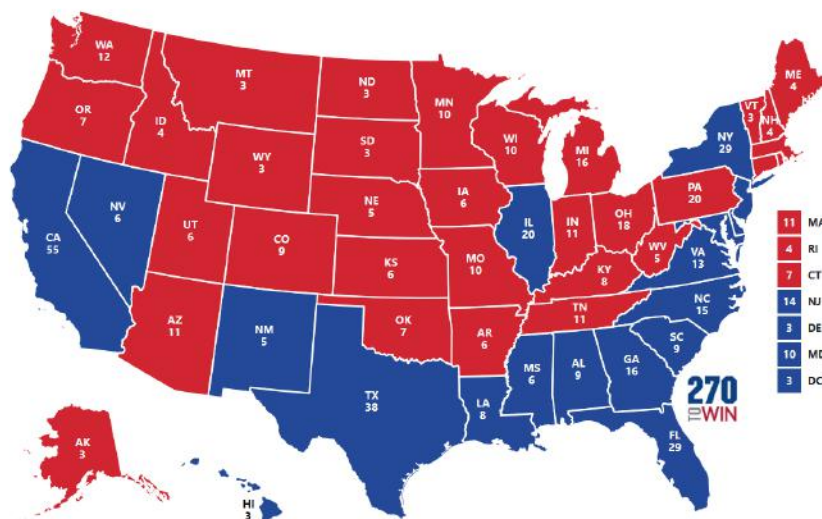
We found that the most predictive demographic was religion. This is not to say that the other demographics were not useful; we will discuss how they were used in detail. However, if we had blindly decided to predict

an election based purely on religious affiliation statistics, we would have had pretty accurate results. The map below shows what religion predicts alone:



Race

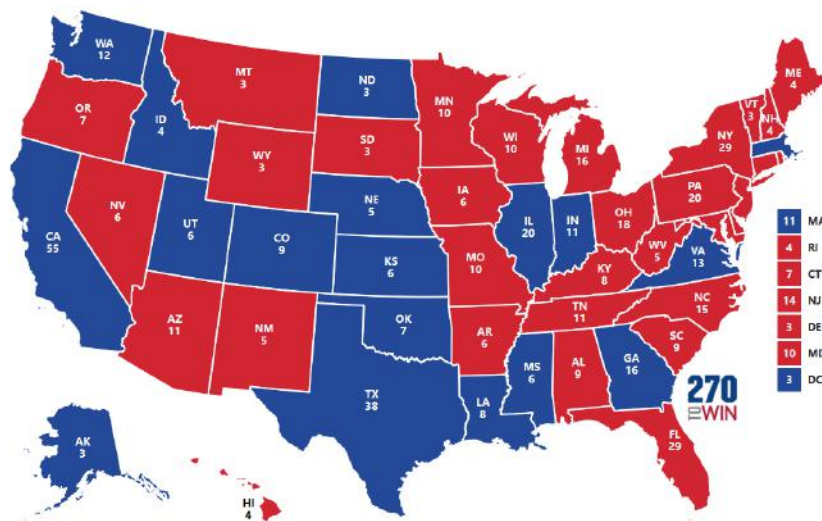
Another demographic we investigated was race. Black and Latino voters vote overwhelmingly for the Democratic party, while whites tend to have a preference for the Republican party. However, we also distinguished between non-college-educated and college-educated whites. They also have significant differences in voting preferences. In the end, even accounting for this, our race numbers predicted the South would go Democrat due to its high black population, while some heavily white Northern states would go Republican.



Age

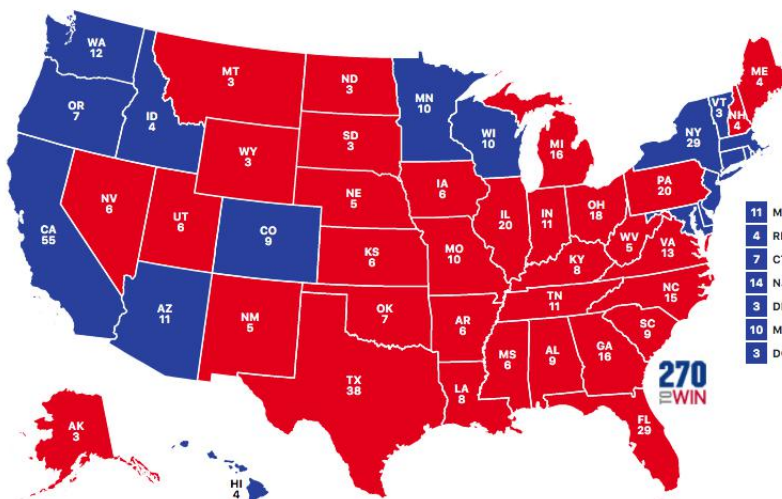
Age seems to be relatively impactful. If one just considers age across the entire country, there are stark differences between wildly Republican older voters and wildly Democratic younger voters. However, it turns out that the difference between the age breakdown among states is remarkably insignificant. One hypothesis for this is that although Republicans tend to be older, Republican states tend to have a higher fertility rate.

The following map is wildly incorrect, but the magnitude of the Republican and Democratic leans produced by age are very small and do not end up being factored significantly into the data.



Income

The income map is pretty accurate-looking:



In reality, all the colors shown above are the reverse of the actual result our process produced. This is because individuals who are poorer on average vote Democrat (and that is the metric we used in our process), but richer states on average vote Democrat. This means that we found a counter-intuitive but strong negative correlation for income. This was also accounted for in our final tabulations. To learn more about this strange phenomenon, there are whole books about it, such as *Red State, Blue State, Rich State, Poor State*.

Computing State Leans

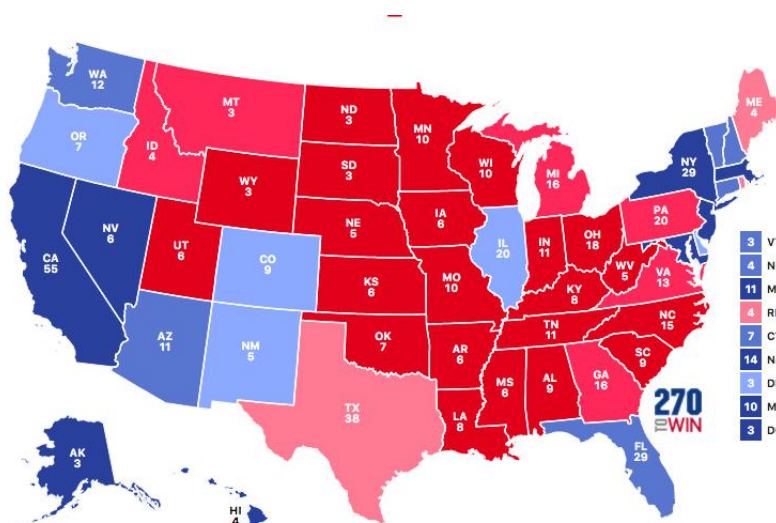
We computed final partisan leans for each state by factoring in demographic leans and 2016 results. The final partisan leans are an estimate of how many points more Democratic or Republican a state should be

than the national average. This involved two major steps in which we had to weight the various different factors in a mathematically sound manner.

Weighting the Demographics

Our primary hypothesis weights income, religion, age, and race. However, there are many ways to weight them which would all produce the same optimal number of correct predictions. To distinguish between them, we had two secondary factors: (1) find weightings that also made the magnitude of the demographic leans match closely expected magnitudes; (2) find weightings that made the standard deviation of leans similar but slightly larger than 2016.

Once we worked for an extended period of time attempting to optimize these factors, we settled on weightings which correctly predicted 36 of 38 non-swing states, and also were of approximately the right magnitude. The following map is nothing like our final prediction, but is one of the major inputs into our final prediction. This map depicts a demographic lean of 8 or more as dark red/blue, 4 to 8 as medium red/blue, and 0 to 4 as light red/blue. Note that these are estimated leans from the popular vote, so it will look a lot more red than a typical map because in recent elections the popular vote has normally favored Democrats.



This is only obviously incorrect in Rhode Island and Alaska. However, we also note that Florida and Arizona are too Democratic while Minnesota and Illinois are too Republican even considering the expectation of it being redder than usual. However, setting these edge cases aside, this weighting does a remarkably good job predicting the outcome. It will help account for demographic change from 2016 to today.

Weighting Past Elections

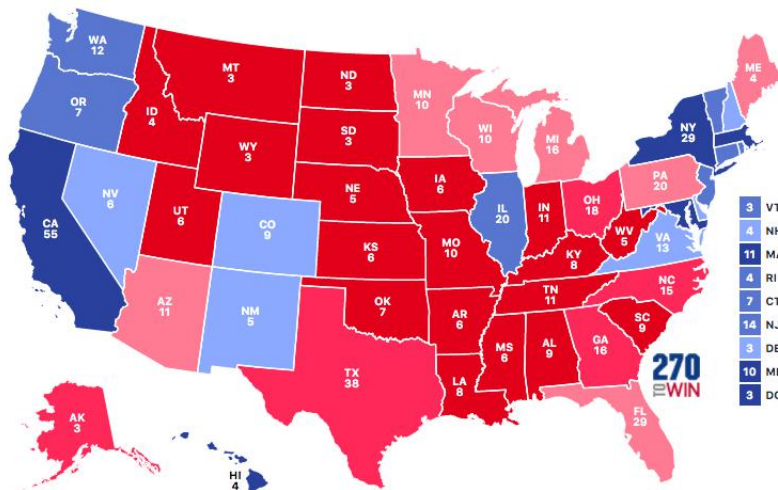
At first, we were interested in incorporating 2018 results into the forecast. However, there are at least three issues with this. First off, many congressional elections are simply uncontested, so in a state that is overwhelmingly one way or the other we wouldn't end up seeing that in total vote count for 2018. Secondly, congressional districts are heavily gerrymandered so taking them into account would basically amount to taking into account control of the state legislatures ten years ago. Finally, even if the prior two issues didn't exist, voter turnout is both far lower and differently distributed among the population in midterm elections than in presidential elections.

This means that we have ended up sticking primarily to 2016 results, while also augmenting them with our demographic lean. Our exact process was to compute the partisan lean of each state in 2016 by comparing the popular vote margin within the state to the popular vote margin in the entire country. For example,

Minnesota went for the Democrats by a margin of 0.8%, but since the U.S. went for the Democrats by a margin of 1.1%, we say that Minnesota has a 2016 lean of 0.3% for the Republicans. Combining the 2016 lean with the demographic lean gives us a final 2020 lean for each state.

Final 2020 Leans

The following map depicts the final 2020 leans. Once again, more than 8% counts as a strong lean, 4 – 8% counts as a medium lean, and less than 4% counts as a weak lean. It requires the same disclaimer as the demographic leans - it should look redder than expected given that Democrats have won the popular vote in the past 3 elections.



Final Predictions

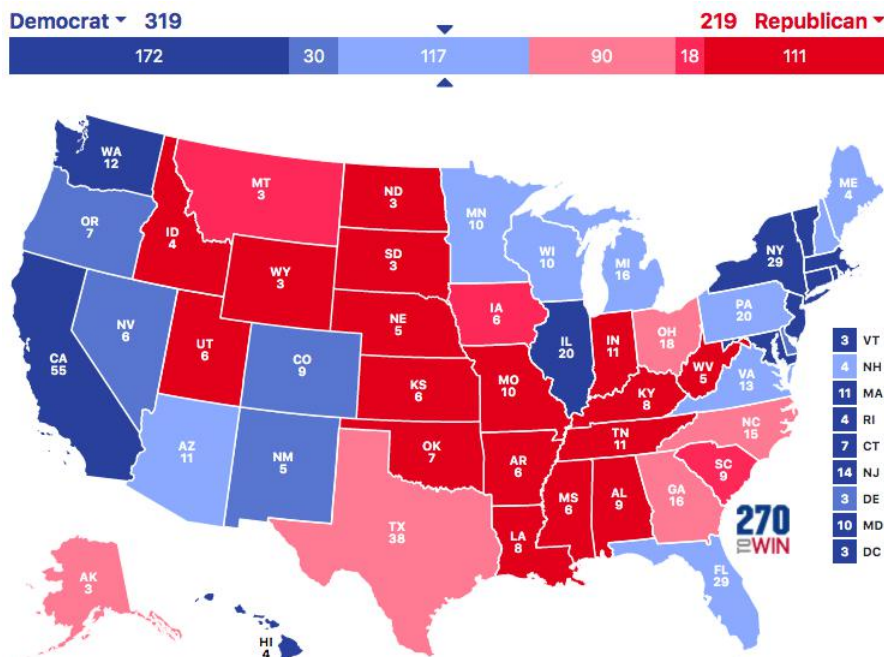
To make final predictions, we need to take the most important step - including current polling. Once we do this, we can make some final adjustments and then produce our completed map.

National Polling

Polling at the time of writing, that is, late March and early April, will not be very predictive of the actual result come Election Day. However, it is all that we have to work with so we have to use it to the best of our ability. We sourced our polls from a comprehensive list of polls FiveThirtyEight compiles. However, we did not use any aspect of FiveThirtyEight's poll rating system or any of FiveThirtyEight's general election modelling - all of our results are coming straight from raw data.

At risk of being too simplistic, we took an straight average of national polls performed in the last month. Given that we have no way of knowing the reliability of each pollster without intensive research, this was the best option available to us. In the end, this gave us an average of 53.5% Biden to 46.5% Trump. This is a pretty hefty lead for Biden, but may not spell Trump's doom, due to the quirks of the Electoral College. After all, in two of the past five elections, the candidate with the most votes has lost.

Then, we can use our partisan leans for each state computed in the past section to come up with a baseline prediction for each state. Note that these predictions do not factor in statewide polling. The following map once again splits colors at 8% and 4%, but now it represents how much above / below even will each state be at this point in our predictions.

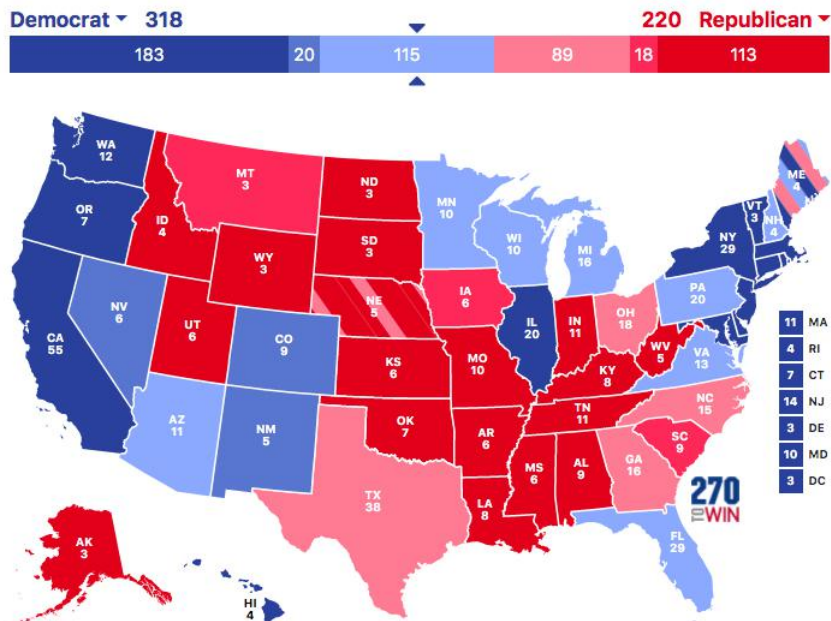


State Polling

State polling provides a finer level of detail which so far is missing from the model. It is scarce in non-swing states, but this is not much of an issue since their results are already pretty much set. However, we factored in the state polling where it existed in such a way that the more polls were available, the more they counted relative to the numbers using only national polling shown above.

State polling is weighted into the final map that can be found in the next section.

Final Map



Caveats

Probability of our Correctness

We attempted to find the probability that our model was right. First, we found the "middle state" - that is, the state such that if all the more Democrat-leaning states went Democrat, and all the more Republican-leaning states went Republican, it would be the deciding vote. In our model, the state happened to be Arizona. Then, since Arizona would flip the election we found the probability that it would go for Biden. In the end, we got a 78.3% chance that Biden would win Arizona, and by our assumption, the election.

There are flaws with using this number for our uncertainty, given that we don't know for sure our model correctly orders the states' partisan leans. To try to correct this problem, we also tried finding the same probabilities for 13 swing states, and used a program to compute the probability of each combination occurring. This gave a probability that Biden would win of 96.2%. However, this number is also flawed, because state results are correlated with each other. For example, if Republicans win Michigan, Republicans are more likely to win Wisconsin as well, because it is likely that if the polls had a Democratic bias in one of these states, it would have a similar bias in the other.

In addition, these uncertainty numbers still assume the election is happening right now. We will discuss various future events that could have an impact on the results and mean that our uncertainty should be much higher.

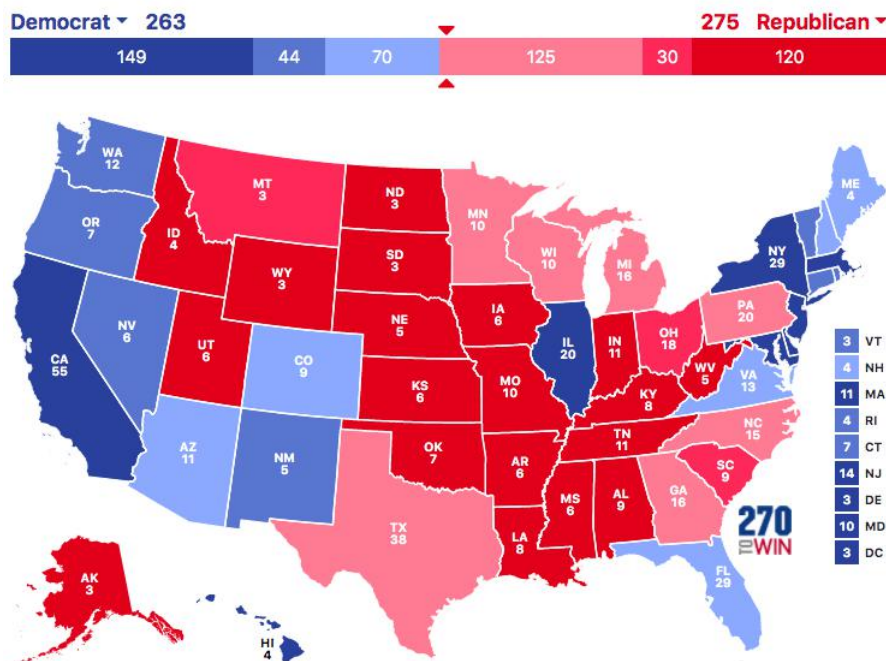
Ignoring Third Parties

During this process, we disregarded the effect that third parties have on polling numbers. We took every poll and assumed those who want third parties or who are undecided don't exist. In different areas, third parties take different portions of the vote. Furthermore, in different areas, undecided voters end up voting differently. Given their relative levels of unpopularity among their respective parties, Biden or Trump may end up losing votes to third parties to an extent that flips the result in very close states.

Trump Polling Inaccuracies

In 2016, the polls were on average more Democratic than the final election day results. It was not as inaccurate as many believed, but a number of swing states were incorrectly predicted. In one of the most extreme examples, Wisconsin polling averages gave Clinton a 5% lead, but she ultimately lost the state by 0.5%. In an attempt to correct for this, we considered polling averages for Clinton vs. Trump in 2016 around the time Sanders dropped out (so in theory they should be similarly accurate to ones at the time of writing), and compared them to actual results. We then took this discrepancy and adjusted the 2020 polls according to said discrepancy. To do this we simply compared the polling average before the election with the final election results to computer how "wrong" the polls were, and assumed the same level lean in polls for this election cycle.

After doing this, we get a new map of results which is less clear-cut:



Incumbent Polling Inaccuracies

This election is different from 2016, because Donald Trump is an incumbent. This means voters have seen how he is in office. This may mean polling numbers are more accurate for Trump than they were in 2016, because voters have had a much longer time to form an opinion on him.

Furthermore, Trump would receive an incumbency advantage, whereby the incumbent president is expected to do better than in their first election. Since 1912, the incumbent has on average increased their popular vote percentage by 3.4% in their re-election. This is generally explained by voters with not particularly strong feelings who end up voting for the president, as they are more well known and usually seen as a "safer" pick, as they represent the status quo. We did not account for this advantage in our model.

Future Events

There are also many things that could happen between now and Election Day which could have a significant effect on the outcome. Biden's vice president choice could give him a boost, depending on who he chooses; people who are excited by his vice presidential pick may be more likely to vote for him. Trump is less likely to experience this effect, because he already has a vice president.

COVID-19

There is also the elephant in the room: COVID-19. The coronavirus outbreak has upended millions of lives across America. This could have a variety of different effects, depending on how the post-COVID-19 economic revival plays out. It would probably have the biggest effect on Trump's polling, because he has more control over the situation. It could boost Trump's numbers if Americans feel he handled the coronavirus crisis well, or it could decrease his numbers if they feel he did not meet expectations.

Another potential impact of COVID-19 is called the "rally behind the flag effect." Usually, in crises throughout history, the current leader spikes in popularity. Here are some examples:

- George W. Bush went from 55% to 90% approval after 9/11

- FDR went from 70% to 85% approval after Pearl Harbor
- JFK went from 60% to 75% after the Cuban Missile Crisis

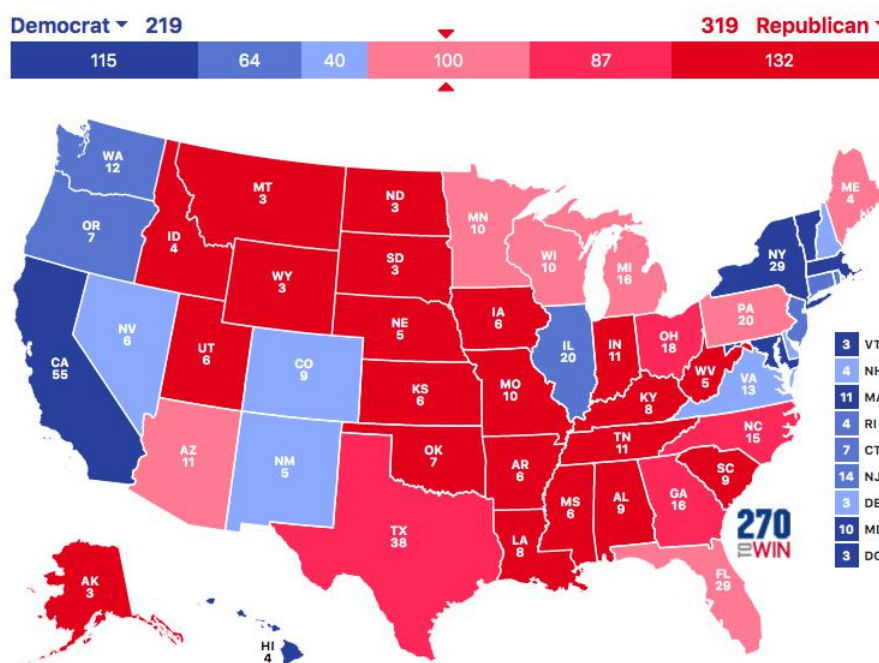
It is worth noting that Trump has not yet seen such a boost, but one could be coming that could affect the election.

A Specific Future Event

In the time between when we gave our editor this article and our editing the article, Biden's sexual assault allegation has been one of the few media stories that manages to last in the age of coronavirus. In addition to potentially affecting Biden's poll numbers and Democratic turnout, this also functions as an example of how hard it is to predict what is going to happen in November, especially at such an uncertain time.

A Final Note

In 2016, Donald Trump won the Electoral College despite losing the popular vote. We created a map demonstrating what the results of each state would be if the popular vote was exactly 50/50. We fixed the popular vote and adjusted every state according to its lean. In this scenario, Trump would be given more electoral votes than Biden:



When Biden and Trump receive an equal share of the popular vote, the Republicans are given a strong victory. The electoral college gives more votes per million people to smaller states than larger ones. Because smaller states tend to lean Republican, Donald Trump would receive 319 electoral votes compared to Biden's 219 votes.

Citations

1. Kilgore, Ed. "Trump Outperformed His Polling in 2016. Will He Again in 2020?" *Intelligencer*, 4 Aug. 2019, nymag.com/intelligencer/2019/08/trump-out-performed-his-polling-in-2016-will-he-again.html.
2. "2016 United States Presidential Election." Wikipedia, Wikimedia Foundation, 17 Apr. 2020, en.wikipedia.org/wiki/2016_United_States_presidential_election.
3. Griffin, Rob, et al. "Voter Trends in 2016." Center for American Progress, www.americanprogress.org/issues/democracy/reports/2017/11/01/441926/voter-trends-in-2016/.
4. "Religion in America: U.S. Religious Data, Demographics and Statistics." Pew Research Center's Religion & Public Life Project, 11 May 2015, www.pewforum.org/religious-landscape-study/.
5. "President: General Election Polls." FiveThirtyEight, 18 Apr. 2020, projects.fivethirtyeight.com/polls/president-general/.
6. Gelman, Andrew, and David Park. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton Univ. Press, 2010.
7. Sciupac, Elizabeth Podrebarac, and Gregory A. Smith. "How Religious Groups Voted in the Midterm Elections." Pew Research Center, Pew Research Center, 7 Nov. 2018, www.pewresearch.org/fact-tank/2018/11/07/how-religious-groups-voted-in-the-midterm-elections/.
8. "Voting in America: A Look at the 2016 Presidential Election." The United States Census Bureau, 10 May 2017, www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html.
9. "Election 2016 Exit Polls: Votes by Income." Statista, 9 Nov. 2016, www.statista.com/statistics/631244/voter-turnout-of-the-exit-polls-of-the-2016-elections-by-income/.
10. "Why the Voting Gap Matters." Demos, 23 Oct. 2014, www.demos.org/research/why-voting-gap-matters.
11. Birnbaum, Emily. "AP Poll Finds Only Group Breaking for Republicans Are Americans 65 or Older." *TheHill*, 7 Nov. 2018, thehill.com/homenews/campaign/415356-survey-only-group-breaking-for-republicans-are-americans-65-or-older.
12. Young, J.T. "Incumbency's Advantage Could Trump Democrats in 2020." *TheHill*, The Hill, 10 Jan. 2020, thehill.com/opinion/campaign/477774-incumbencys-advantage-could-trump-democrats-in-2020.