Final Report: Election Prediction

Alexander Bendeck, Lynn Fan, Cathy Lee, Alice Liao, Justina Zou

2 November 2020

Introduction

The United States is getting closer to the 2020 presidential and Congressional elections on November 3, 2020. All 435 seats in the United States House of Representatives, 35 of the 100 seats in the United States Senate, and the office of President of the United States are up for election (Wikipedia, 2020). With the current polarizing political landscape, the election outcomes are significant to determine the next stage of this country. Polarization also makes election prediction both less and more difficult. It is less difficult because election results are less subject to election or candidate-specific factors, and it is more difficult as elections become increasingly close and fierce competitions, especially for presidential elections (Gelman, 2020). Therefore, we believe 2020 is a particularly challenging yet interesting year for statisticians to predict election outcomes and compare results with other predictions and the actual outcome to reflect on methodologies and unaccounted predictors. These election predictions not only inform the public about campaign trends and political sentiments but also help political strategists make decisions on allocating campaign resources for different candidates (Linzer et al.). American political pundits have been spending countless hours obtaining and analyzing relevant data to predict the election outcomes, and historical models and predictions and pre-election polls are two of the most important sources of information (Linzer et al.). In this report, similarly, we will also be using both literature review and polls for 2020 to build our prediction models.

We plan to build prediction models for the 2020 U.S. presidential election and the Senate election nationwide. Also, among all states, we decide to further predict the U.S. House election outcome for North Carolina, as it has been a swing state in presidential and Congressional elections for decades. Since 1996, the Republican statewide vote share in Congressional elections has varied "from a low of 45% in 2008 to a high of 55% in 2014" (Perrin et al.). To summarize, this report aims to use statistical models to predict (1) the outcome of the presidential election, (2) whether the US Senate remains in Republican control, (3) the electoral college vote, (4) the outcomes of all NC Congressional elections (the 13 federal Representatives to Congress), and (5) the outcome of the NC Senate election, including characterization of uncertainty in predictions.

Data Description

In literature, election prediction relies on polling data as well as the fundamentals, which are economic indicators (The Economist, 2020) and voter turnout by demographic groups (Hansford, et al.). To forecast the outcomes of presidential election, Senate election and House election (for North Carolina only), we obtained 2020 presidential polling data from *The Economist*, 2020 senate and house polling data as well as partisan lean data from *FiveThirtyEight*. Partisan lean data refers to the average difference between how a state or district votes and how the country votes overall, based on results from the past two presidential elections as well as statewide elections in 2018.

The fundamentals data were retrieved from various online sources, including Andrew Gelman's presidential election prediction Github repository (for correlation across states and historical incumbent party's June approval ratings), Federal Reserve Economic Data (for second quarter real income growth), and the NC Board of Elections website (for 2020 NC registered voter demographics). Please refer to Appendix B for a detailed description of all data sets used.

Data Processing and Missing Data

For both the president and Senate election models, we had to choose which states we considered to be "battlegrounds" to include in the models. We chose states for the models separately, since some states have a competitive race for presidential election but not for Senate, and vice versa.

For the presidential election data, we choose to use polling data within 30 days before the election, as this aligns with Andrew Gelman's observation that polling data becomes more predictive as the day of poll becomes closer to election day (Gelman, 2020). Note that state-wide polls 30 days or less are unavailable for 15 states, including Oregon and Idaho. But these states are not swing states in recent elections, so we did not include those in our modeling process. For modeling, we will filter out those states with obvious party preferences and focus on those states showing percentages swinging right above or below 50%. For the presidential model, the states chosen are Arizona, Florida, Georgia, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Texas, and Wisconsin. These thirteen states are rated as either "Toss Up", "Lean Republican", or "Lean Democrat" races (the three most competitive categories) by the Cook Political Report, an nonpartisan elections newsletter, as of October 28 (Cook Political Report, 2020). These states are also projected to be competitive in the presidential race by both the FiveThirtyEight and Economist models. We select poll responses from likely voters only, since it is known that using all responses may overestimate support for the Democratic party, according to FiveThirtyEight (Silver, 2014).

For the Senate model, the states chosen are Alaska, Arizona, Colorado, Georgia, Iowa, Kansas, Maine, Michigan, Montana, North Carolina, South Carolina, and Texas. Note that state-wide polls 30 days or less are unavailable for 28 states. One example is Oregon, but this state has strong historical party preferences, and the missing data does not hurt our modeling process as we are only focusing on the swing states. There is also a special election for the Senate in Georgia in addition to the regularly scheduled election. These thirteen races are also rated as either "Toss Up", "Lean Republican", or "Lean Democrat" (the three most competitive categories) by the Cook Political Report as of October 29 (Cook Political Report, 2020). and are projected to have competitive Senate races by both the FiveThirtyEight and Economist models.

For the U.S. House election, due to the scarcity of poll responses for the NC House elections, we included all poll responses within 115 days of the election. We will supplement the polls with the voter turnout results from our interim report to predict the House election for NC.

Exploratory Data Analysis

After initial data cleanup for analysis, Figure 1 aims to explore the polling data we have for the 2020 presidential election within 30 days of election, after filtering for the states of interest. The left plot shows that most states, as the polls indicate, have a vote share higher than 50% for Biden, averaging on all the polls available for each state respectively. Ohio is the only state here with a light red color. The right plot shows that within 30 days to election, variations exist for Biden support rate among polls within individual states. Combined with abundant literature that takes time into account for outcome prediction models, we will take days to election as a part of our model prediction as well (Gelman, 2020).

Similarly, Figure 2 visualizes the polling data we have for the 2020 U.S. Senate election, also within 30 days of election and after filtering for the states of interest. The left plot shows that most states, as the polls indicate, have a vote share higher than 50% for the Democratic Party, averaging on all the polls available for each state respectively. Texas is the only state with a bright red color in this data set. The right plot shows that within 30 days to election, variations exist for the Democratic party vote share among polls within individual states, such as the trend seen for North Carolina as Democratic Party vote share decreases getting closer to the election. We will take days to election as a part of our model prediction as well.

Figure 3 shows the different Democratic Party vote shares predicted by our interim report for all congressional districts in North Carolina. After predicting for voter turnout in the interim report, we use the party registration for predicted voters to get the vote share for the Democratic Party. First, we grouped predicted voters for congressional districts. There are three categories in party registration: Democratic, Republican, and other (indicating a third party or unaffiliated). Within each congressional district, we split the `other` population in half, and add it to both the Democratic Party vote count and the Republican Party vote count, as FiveThirtyEight
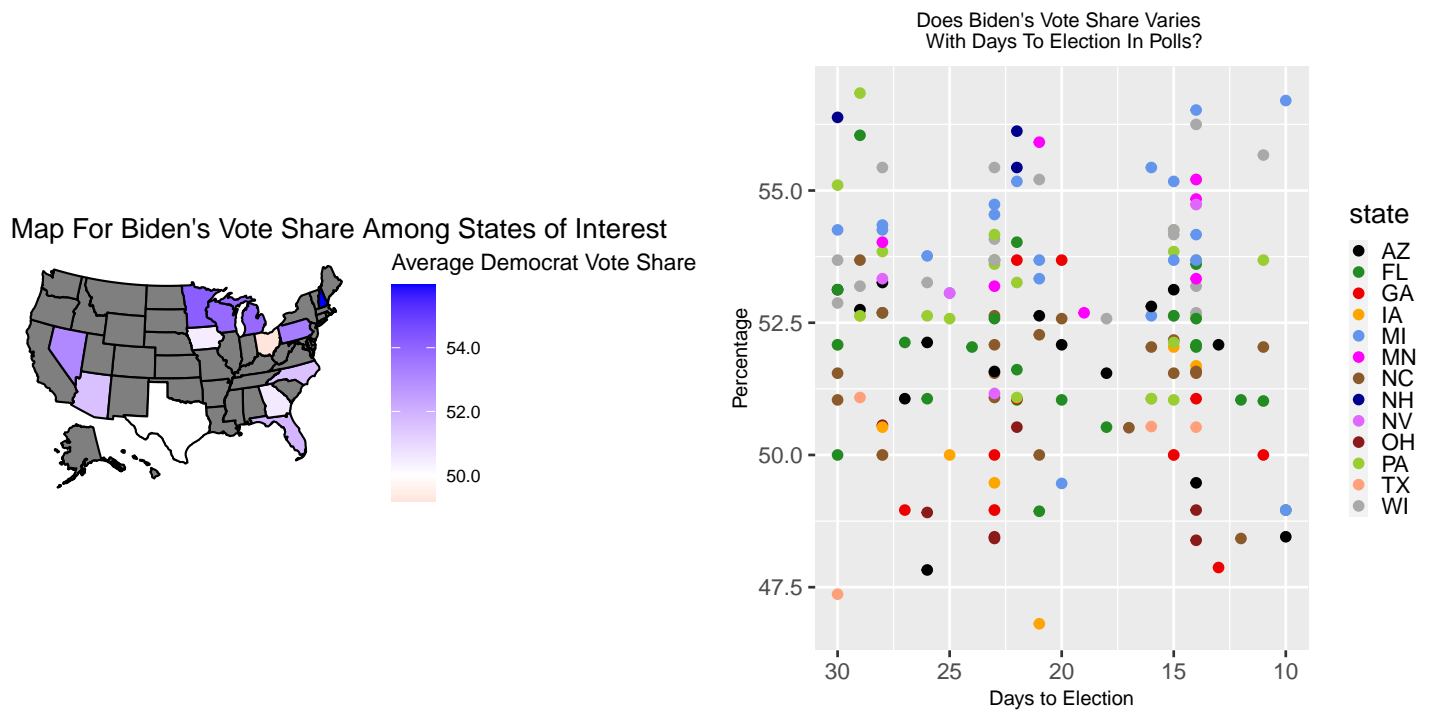
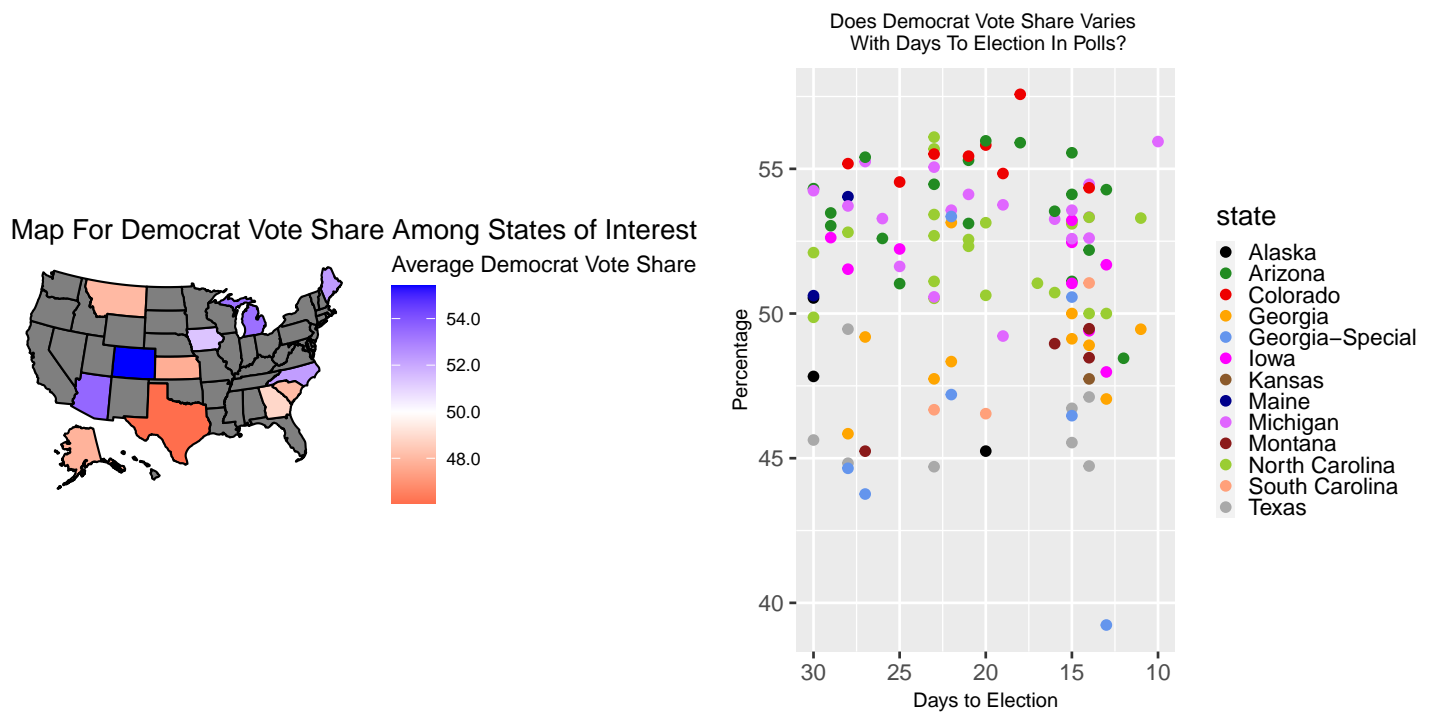Figure 1: Presidential Election Data Visualization

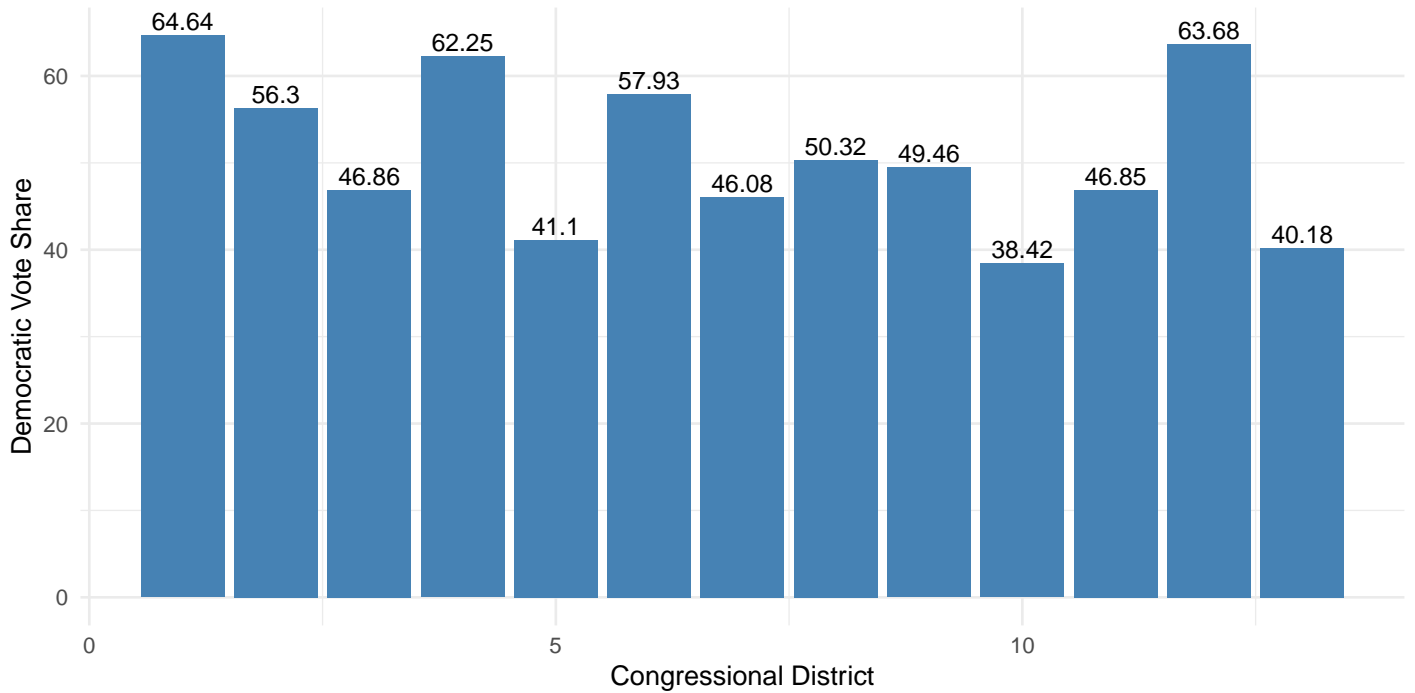Figure 2: U.S. Senate Election Data Visualization

3

Figure 3: Democratic Vote Share Predicted By The Interim Report

used the same approach to get vote share for their House predictions, shown in[https://projects.fivethirtyeight.com/2020-election-forecast/house/]. Among 13 districts, 6 of them have a Democratic Party vote share higher than 50%.

Method

Similar to prior work (Linzer 2013), we use a hierarchical model structure for our election models to overcome the limitation that not every state is polled on every day, allowing the model to borrow data across states (or congressional districts). This also helps account for the fact that polls from each individual state are correlated data observations (Gelman, 2020). We believe that vote share should be modeled as a normal distribution because, historically, vote shares among candidates have mode toward the center of the distribution (landslides are unlikely, especially in the case of a presidential election). In addition, we chose for our random walk to be lag-1 because it's more reasonable to have updated predictions dependent on the previous prediction as opposed to a longer time period (see model specification below).

To specify priors for the mean two-party vote share for the Democratic Party in each state or district, we used a combination of state partisanship, economic fundamentals, candidate incumbency, and projected voter turnout based on demographics. First, we started with state or district partisan lean calculated by FiveThirtyEight (see "FiveThirtyEight's Partisan Lean" in references). Partisan lean is the average difference between how a state or district votes and how the country votes overall, based on results from the past two presidential elections as well as statewide elections in 2018. For example, North Carolina has a partisan lean of R+4, meaning that in a 50/50 political environment nationwide, a Republican would be expected to win North Carolina by 4 points. For each partisan lean, we turn it into a part of our prior mean of Democratic vote share by adding or subtracting half the partisan lean from 50%. North Carolina is a Republican-leaning state, so we subtract 2 from 50% to get a partisan lean prior of 48% (note that the Republican would then get 52% of the two-party vote, resulting in an R+4 margin of victory).

Presidential Election Model

For the presidential model, as discussed previously, the states chosen are Arizona, Florida, Georgia, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Texas, and Wisconsin.

The response we will use is Y_k, Biden's two-party vote share using the following model where the poll k (of $K=3361$ polls) ended t days (of $T = 30$ days) before the election and was conducted on state j (of $J=13$ states of interest).

$Y_k \sim N(\theta_{jt}, \sigma_{yj}^2)$
$\theta_{...t} \sim MVN(\theta_{...t-1}, \Sigma)$
$\Sigma \sim Wishart(S, J + 1)$ where S is the state covariance matrix (obtained from Andrew Gelman's model) and J is the number of states in the model.
Priors for σ_{yj}^2: $\setminus \sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
Priors for θ_{j1}: $\setminus \theta_{j1} \sim N(\mu_j, \sigma^2)$
$\sigma^2 \sim InvGamma(0.5, 0.5)$
$\mu_j \sim N(h_j, 7.5^2)$
$h_j = 0.1 * \text{Presidential Fundamentals}_{2020} + 0.9 * \text{Vote Share from Partisan Lean}_j$
$\text{Presidential Fundamentals}_i = 100 - \text{Incumbent Party Vote Share}_i$
$\text{Incumbent Party Vote Share}_i = \gamma_0 + \gamma_1 \text{June Approval Rating for Incumbent Party}_i + \gamma_2 \text{Three Month Stock Growth}_i + \gamma_3 \text{2nd Quarter Real Income Growth}_i + \epsilon_i$
$\epsilon_i \sim N(0, \phi^2)$

For our presidential model, we wanted to incorporate state-level correlations that differ between each pair of states. For example, Wisconsin is much more similar to its Great Lakes neighbor Michigan both geographically and demographically than it is to Arizona, which is in the Southwest and has a much larger Hispanic population. So, we would expect that Wisconsin and Michigan have a higher correlation in their election results than Wisconsin and Arizona. In order to incorporate this into our presidential model, we used a multivariate normal distribution to model two-party vote share in each state, which allowed us to specify a covariance matrix for the states. We used a covariance matrix which was downloaded from *The Economist* forecast model (Gelman, 2020). These similarities between states were calculated by comparing their demographic and political profiles, such as the state's share of white voters and how urban/rural the state is.

In essence, the presidential model used fundamentals and vote share from partisan lean as a starting point for the expected vote share for the Democratic party, and updated the priors with presidential election polling data. We first predicted incumbent party's national vote share (Incumbent Party Vote Share$_i$) based on the corresponding election year i's economic data and June net approval rating of the incumbent party. Since in the 2020 presidential election Biden's Democratic Party is not the incumbent party, Presidental Fundamentals$_{2020}$, the national level prior of Biden's vote share, is 100 minus Incumbent Party Vote Share$_{2020}$. When computing the state specific prior, 10% weight is allocated to the fundamentals based on economic data and 90% weight to the state's partisan lean because the effect of economic indicators has shrunk over the years as electorates became increasingly polarized (The Economist, 2020). In addition, economic indicators are particularly volatile in 2020 due to the COVID-19 pandemic (Gillespie et al. 2020). For the hyperparameters, we chose high variance to reflect the uncertain nature in the political arena in recent years.

Presidential elections in the United States are decided by the Electoral College, so estimating percentage support in each state alone does not tell us who wins the election. For each set of MCMC samples of two-party vote share by state, we subsequently use the predicted winner in each state to add up the electoral votes of each candidate. For states not in the model, their electoral votes are allocated assuming they vote the same way as in 2016. The probability of President Trump winning re-election is then the probability that he receives 270 or more electoral votes across simulations. Note that although Maine and Nebraska allocate some electoral votes by Congressional District, only two electoral votes are competitive due to this wrinkle, so for simplicity we simply allocate all electoral votes to the statewide winner of each respective state.

Senate Election Model

For the Senate model, the states chosen are Alaska, Arizona, Colorado, Georgia, Iowa, Kansas, Maine, Michigan, Montana, North Carolina, South Carolina, and Texas.

In addition to the 13 competitive races in our model, there are 22 additional races that are not considered competitive. Then there are also 65 other Senate seats that are not up for re-election this year. Note that the Georgia special election actually has multiple Republicans and Democrats running on the same ballot. If no candidate wins over 50% of the vote, which is considered likely, the top two finishers will advance to a one-on-one runoff election in January (Ballotpedia, 2020). For simplicity, our model simply sums up the support for Republican candidates and compares that to the sum of the support for Democratic candidates, then treats this election like the others. By summing the Republican and Democratic support, we essentially assume that partisan support will consolidate around one candidate for each party and will remain roughly balanced between the two parties. (This is not the ideal way to simulate this election, but the difficulty of predicting turnout and the political environment for the January runoff election made this decision seem like the best one to make.) Once again, we only included poll responses from likely voters that responded within 30 days of the election.

The response we will use is Y_k, the Democrat candidate's share of the two-party vote using the following model where the poll k (of K=_fill in_ polls) ended t days (of $T = 60$ days) before the election and was conducted on state j (of J=13 states of interest).

$Y_k \sim N(\beta_{jt}, \sigma_{yj}^2)$
$\beta_{jt} \sim N(\beta_{jt-1}, \sigma_{\beta j}^2)$
Priors for σ_{yj}^2: $\backslash \; \sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
Priors for $\sigma_{\beta j}^2 \backslash \; \sigma_{\beta j}^2 \sim InvGamma(\nu_\beta, \nu_\beta \tau_\beta))$
$\nu_\beta \sim Uniform(0, 100)$ and $\tau_\beta \sim Uniform(0, 100)$.
Priors for $\beta_{j1} \backslash \; \beta_{j1} \sim N(\mu_j, \sigma^2)$
$\sigma^2 \sim InvGamma(0.5, 0.5)$
$\mu_j \sim N(h_j, 7.5^2)$
$h_j = $ Vote Share from Partisan Lean$_j$ + Incumbency Advantage$_j$

For the U.S. Senate model we also take the incumbency advantage into account. Analysis by FiveThirtyEight (Rakich, 2018) found that incumbent senators get a 2.6-point boost. For each incumbent was running a Senate race, depending on whether they were a Democrat or Republican, we either added or subtracted this incumbency adjustment with the prior mean computed from partisan lean.

For the NC-specific prior, we supplemented partisan lean and incumbency adjustment with voter turnout predicted by the model in our Interim Report. When computing the prior for NC only, 10% weight is allocated to predicted Democratic vote share from turnout and demographics, and 90% weight is allocated to partisan lean and incumbency adjustment. The turnout is weighted much less because there is considerable uncertainty surrounding turnout because of the COVID-19 pandemic. See Appendix A.2 for more specifics on how the voter turnout model was used to compute predicted Democratic vote share.

For predicting control of the U.S. Senate, things are complicated slightly by the fact that a 50/50 split of Senate seats is a possible outcome of the elections. In this case, the Vice President breaks the tie, so the party that wins the Presidential election will only need 50 Senate seats for a Senate majority, while the party that loses will need 51. Since our President model gives Democrat Joe Biden a very high chance of winning (discussed more below), we assume in the Senate model that Democrats will control the Senate in the case of a 50/50 split.

House Election Model

The response we will use is Y_k, the Democrat candidate's two-party vote share using the following model where the poll k (of K= 8 polls) ended t days (of $T = 30$ days) before the election and was conducted on district j (where j is in districts $1, \cdots, 11, 13$ of North Carolina). For district 12, the vote share is coded as 100 because

there is only one candidate, and she is a Democrat. Note that, due to the scarcity of poll responses for the NC House elections, we included all poll responses within 115 days of the election.

$Y_k \sim N(\beta_{jt}, \sigma_{yj}^2)$
$\beta_{jt} \sim N(\beta_{jt-1}, \sigma_{\beta j}^2)$
$\sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
$\sigma_{\beta j}^2 \sim InvGamma(\nu_\beta, \nu_\beta \tau_\beta))$
$\nu_\beta \sim Uniform(0, 100)$ and $\tau_\beta \sim Uniform(0, 100)$.
$\beta_{j1} \sim N(h_j, 7.5)$
$h_j = 0.9 * \text{Vote Share from Partisan Lean}_j + 0.1 * \text{Expected Vote Share from Voter Turnout}_j$

According to FiveThirtyEight (Rakich, 2018),incumbent members of the House get a 2.7-point margin boost in their favor. For the U.S. House races, the incumbency adjustment was calculated by the same method as it was for the Senate races. Expected Vote Share from Voter Turnout$_j$ was calculated from the model in the Interim Report. Please see Appendix A.3 for more details.

Results
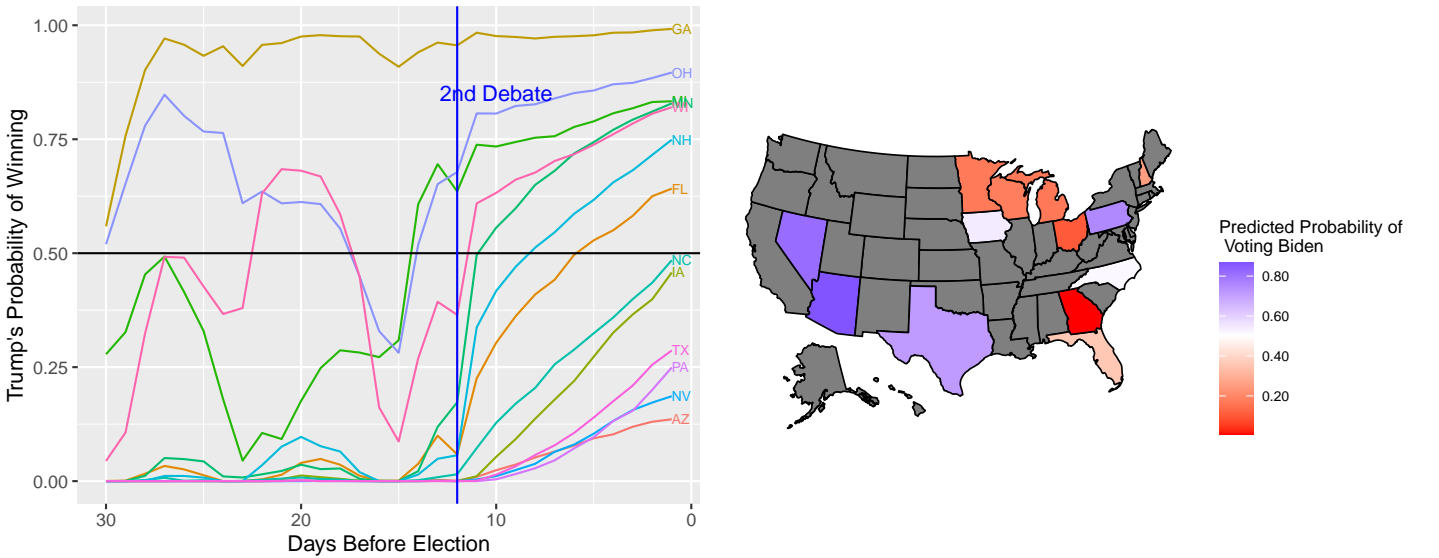
Presidential Election Model

Figure 4: Plot of Predicted Presidential Result for States of Interest on Election Day

Figure 4 shows how Trump's probability of winning the election changes over the 60 days before the election in each state of interest. It appears that his probability of winning increased after the last presidential debate. *check for final results* From Figure 4 we also see that swing states such as Florida and Ohio are predicted to vote for Trump while other swing states such as Pennsylvania and North Carolina are predicted to vote for Biden. *check for final results*

Table 1: Predicted Probability of Winning and Electoral Vote Count

	Win Probability	Estimated EC Votes	2.5% Quantile EC Votes	97.5% Quantile EC Votes
Trump	0.23	238	168	325
Biden	0.77	300	213	370

The probability that Trump wins re-election is 0.23. The predicted electoral colleges votes for Trump is 238 with a 95% confidence interval of (168, 325).
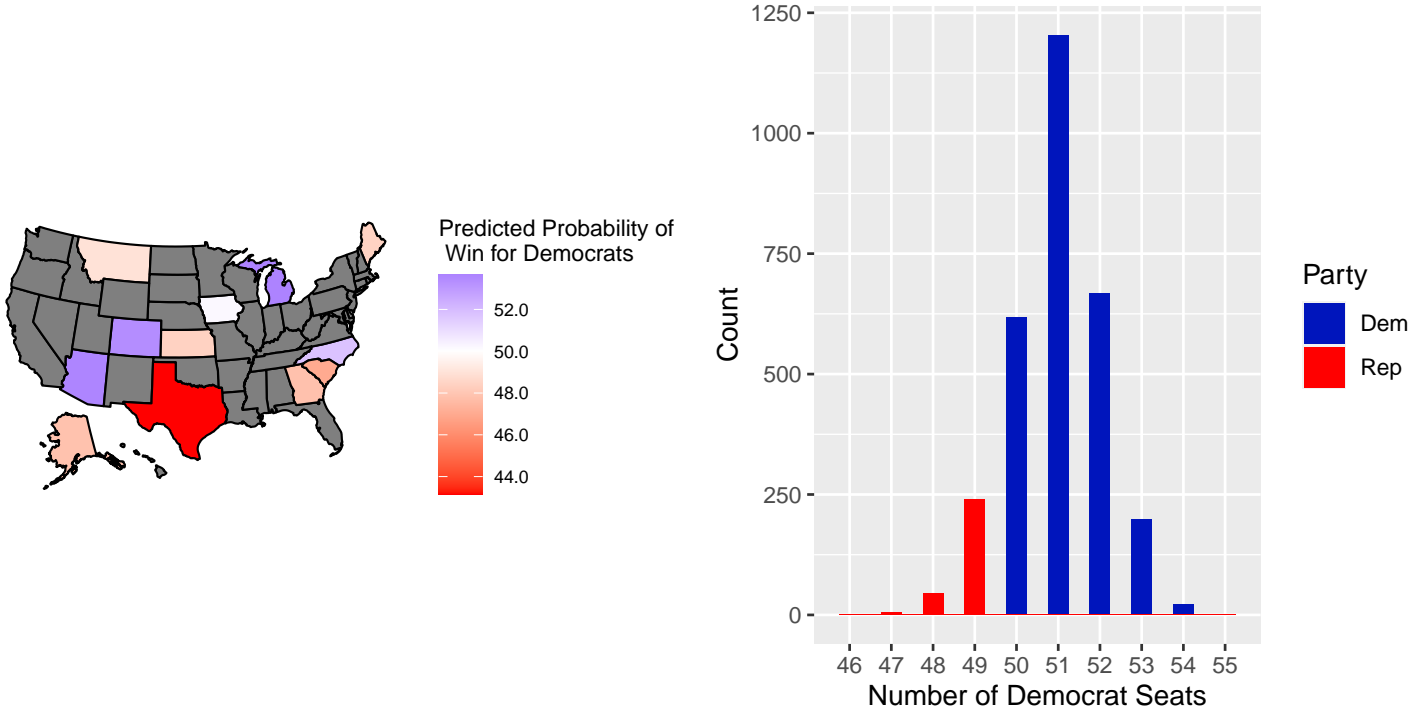
Senate Election Model

Figure 5: Plot of Predicted Senate Results for States of Interest on Election Day

From Figure 5 we see that for the Senate race, states such as Texas are predicted to lean heavily towards the Republican candidate and Democrats have a great than 50% chance of winning states such as Arizona and Michigan. From the histogram in Figure 5 we see the distribution of the predicted number of Democrat seats point to a Democrat majority in the Senate, which is in agreement with the low probability that Republicans retain control of the Senate (0.1) in our model.

Table 2: Predicted Vote Share for Tillis-Cunningham Race

	Estimate	2.5% Quantile	97.5% Quantile
Tillis	48.22	45.01	51.80
Cunningham	51.78	48.20	54.99

For the NC Senate race (Tillis vs. Cunningham), the predicted vote share for Cunningham is 51.78% with a 95% confidence interval of (48.2, 54.99).

House Election Model

Table 3: Predicted Republican Vote Share in NC's 13 Districts

	Estimate	2.5% Quantile	97.5% Quantile
District 1	41.69	36.32	47.02
District 2	40.75	35.29	46.17
District 3	63.61	58.32	68.56

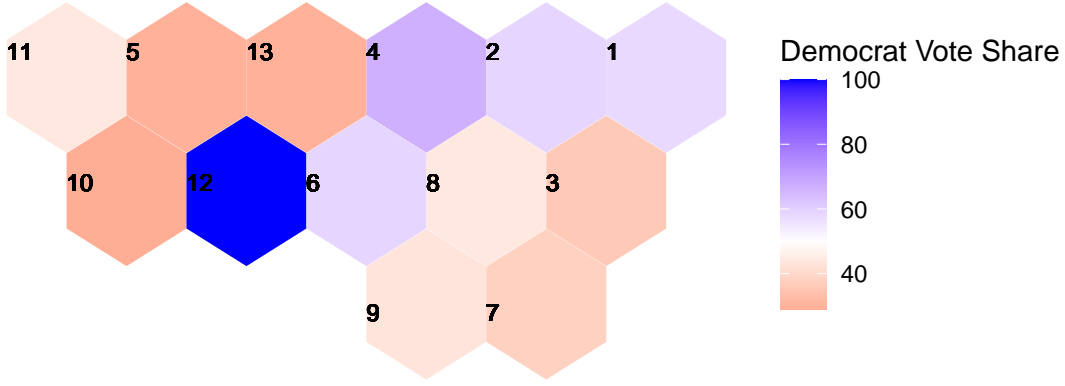|	Estimate	2.5% Quantile	97.5% Quantile
District 4	33.29	27.84	38.48
District 5	69.54	64.20	74.93
District 6	41.19	35.87	46.64
District 7	61.81	56.40	67.43
District 8	55.74	50.67	61.01
District 9	56.79	51.40	61.90
District 10	70.68	65.16	75.96
District 11	56.12	51.86	60.59
District 12	0.00	0.00	0.00
District 13	69.60	64.12	75.01

Figure 6: Plot of Predicted Senate Results for States of Interest on Election Day

From the table of predicted Republican vote share in the House elections for each district, it appears that the Republican candidate will win in districts 3, 5, 7, 8, 9, 10, 11, and 13 (where the lower bounds of the 95% confidence intervals for predicted vote share are also all above 50). From Figure 6 we can see that the vote share for districts 1, 2, 4, and 6 lean toward Democrat (note that the only candidate on the ballot for district 12 is a Democrat).

Model Validation and Sensitivity Analysis

For model validation, we trained our 2020 presidential election model on 2016 polling data and adjusted fundamentals accordingly. For instance, instead of predicting 2020 economic indicators, we removed the year 2016 from the fundamentals training data and predicted 2016 economic indicators. This 2016 model produced a 7% probability that Trump would win the presidency, which is within the range of predictions by respected models available in 2016 (The Economist, 2016). Polls were skewed in 2016, so it is expected that the model would not be entirely accurate.

For sensitivity analysis, we adjusted the relative weights placed on the components prior to the presidential election model. Instead of 10% fundamentals and 90% partisan lean, we ran a model with 50% fundamentals and 50% partisan lean multiple times and the predictions for Trump's re-election probability was generally higher (approximately 0.25) than that from our main model, which makes sense since economic fundamental model contains stock growth and is slightly favored toward Trump. However, the sensitivity analysis models still point to unlikely re-election.

We also adjusted the prior parameters for sensitivity analysis. In both the senate and house model, changing the priors from $\nu_y, \tau_y, \nu_\beta, \tau_\beta \sim Uniform(0, 100)$ to $\nu_y, \nu_\beta \sim Uniform(0, 10), \tau_y, \tau_\beta \sim Uniform(0, 1000)$ to increase the variance had little change on the estimates. Changing the relevant uniform priors in the presidential model resulted in a slightly higher probability of Trump's re-election (the probability was about 0.23), but
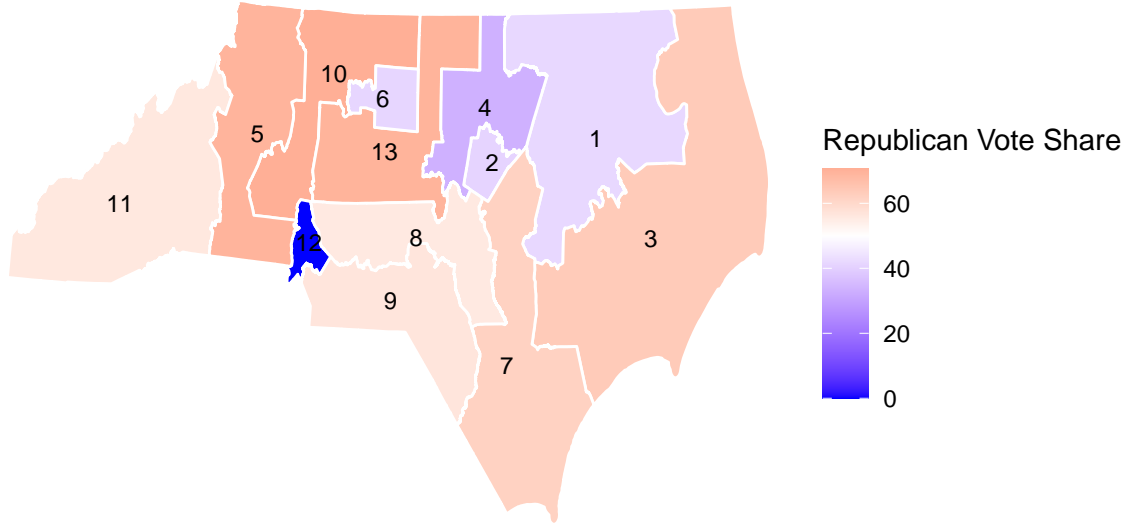
Figure 7: Plot of Predicted Senate Results for States of Interest on Election Day

again, this still indicates Using a stronger prior for the variance for the estimate on the day of the election ($\sigma^2 \sim InvGamma(10, 0.5)$) for the senate and presidential models resulted in comparable performance to our main senate model while the presidential model had higher win probabilities and slightly narrower credible intervals for Trump compared to our main model (the win probability was about 0.34). Similarly, reducing the variance for the house model ($\beta_{ji} \sim N(h_j, 1)$) also resulted in a similar performance, though the quantiles were more narrow, as expected (Linzer, 2013). To account for shock events, we could modify the priors to increase the variance associated with Y_k to reflect the increased uncertainty due to the shock event in each relevant model.

Diagnostics

The MCMC for all three models (presidential, senate, house) have all converged, as indicated by traceplots in Appendix C. In addition, the effective sample sizes are generally above 500.
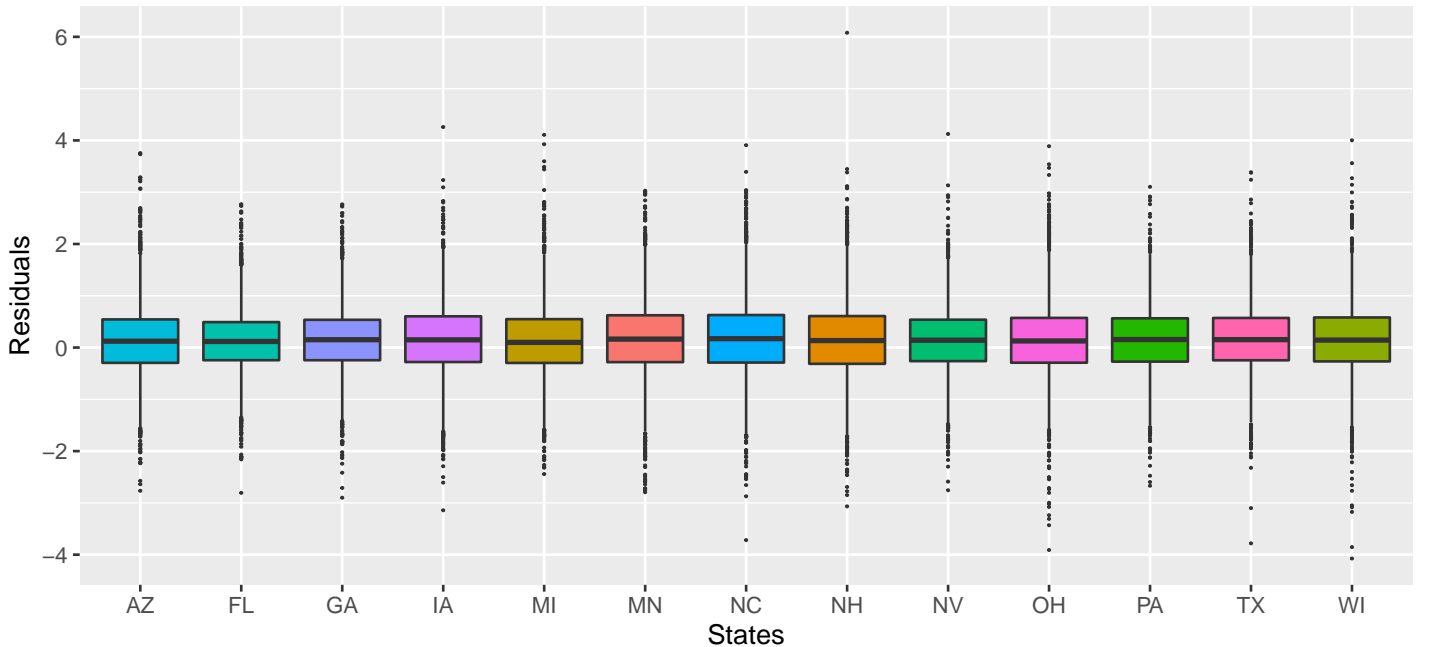

Figure 8: Presidential Model Residual Plot

As shown in Figure 8, the residuals for the estimates of the probability that Biden wins the election are distributed around 0 for all of the states of interest. Residuals for Senate and House race are in Appendix C.

Discussion and Limitations

One place where we made decisions regarding the predictive ability of variables was in the presidential election model. We followed Alan Abramowitz's "Time for Change" model where annualized second-quarter GDP growth rates and the incumbent party's June approval ratings were incorporated as the fundamentals data (Abramowitz 2008). However, given the economic impact of COVID-19, this year's annualized second-quarter GDP growth rate is abnormally low (-34%) (BEA, 2020), which subsequently resulted in a very low predicted probability of Trump's re-election (approximately 5%). Therefore we felt annualized second-quarter GDP growth rate is not an appropriate predictor and substituted it with second-quarter real income growth from one year ago and stock performance three months prior to the election date. As seen in the Results section, the updated predicted probability of Trump winning the election is more sensible.

For the NC House election model and NC Senate election model, we incorporated the voter turnout prediction based on registered voters' demographics from the Interim Report. We believe that even if we had abundant polling data, there's a difference between how people respond to a poll and whether or not they actually turn out to vote. Therefore, we included turnout information from our interim model.

A limitation of our model is that we did not explicitly weigh any polls more strongly than others. Poll-based models often weight polls based on recency of the poll, sample size, and pollster quality (Silver, 2014). Our model actually does take into account poll recency, since our prior specification assumes that polls closer to election day have less variance (noise) around the true level of candidate support. However, we do not weight based on sample size or pollster quality. While it would have been possible to do this, it would have required making many arbitrary decisions. For example, what is a "good" sample size, and how heavily should the sample size be weighted? How do we determine pollster quality, especially if we don't just want to blindly use FiveThirtyEight's pollster ratings? We decided to avoid these issues, especially since it has been observed that such adjustments rarely have a large impact on models (Silver, 2014).

There are inherent limitations to modeling elections based on polls. In some cases, as with the U.S. House races in North Carolina, there may be few or no polls for a given district. Even when polling data is plentiful, we are essentially assuming that the polls are reasonably accurate measures of candidate support. It has been well-documented that systematic polling errors during the 2016 election cycle underestimated the support of then-candidate Trump. While many pollsters have adjusted their sample weighting methods (for example, to adjust for education level), it is not impossible that we could witness a 2016-level polling error again (Skelley, 2020). A recent election prediction study shows that state polls have much higher margins of error than expected (Shirani-Mehr et al., 2018). It has also been noted that registered Democrats have been overrepresented in recent polling samples compared to their proportion of the population overall (Cohn, 2020). In addition, we acknowledge that with the COVID-19 pandemic, polling has moved from in-person interviews to phone calls, live surveys, and internet questions, which could lead to unknown biases with its lower response rates (Gelman, 2020). In addition, COVID-19 posted significant uncertainty in election outcomes. For example, rejected absentee ballots due to mailing issues were not as significant in previous elections. But in 2020, so much more voters are mailing their votes, and an unknown number of them will be rejected, potentially changing voter turnout (Rakich, 2020). We could have also used early voting data, but we excluded them because abundant literature has proven early voting data to have a strong partisan bias (McDonald, 2020). Another factor that could significantly affect election results is potential shock events - terrorist attacks, natural disasters, power outages. Although literature proves that sometimes elections could be vulnerable to those events, because our predictions are so close to the actual day of the election, we did not account for specific shock events in our analysis (Morley, 1905). However, we did sensitivity analysis by increasing variance to account for unpredictable events.

One feature that we did not include is the correlation between simulations of the Presidential, House, and Senate elections. For instance, if Joe Biden wins the Presidential election by a wide margin, we would also expect Democrats to do better in the House and Senate elections compared to if Biden did not win (Desilver, 2020).

This is especially true by state; if Biden wins North Carolina by a wide margin, this would likely help Democrat Cal Cunningham win the NC Senate race. Additionally, since the Vice President can determine Senate control in the case of a 50/50 split in seats, having model correlations could help account for this dependence between Senate control and Presidential results. Instead, all of our models simply run independently.

We do account for state-by-state correlations in our President model, but we do not for the Senate model or the House model. Although there is likely still some correlation in these cases, such races are generally a bit more localized (Gillespie et al. 2020) and incumbency advantages can contradict national trends. As a result, we treated these races as being uncorrelated. Possible ideas for addressing this in future studies could be to try to compute our own covariances between districts based on voter demographics, or for the Senate to try predicting covariance between states that is similar to that in the Presidential model but "weaker".

Furthermore, for making our prior to Democrat vote share in the House and Senate model (for NC), we assume that registered Democrats and Republicans will vote for their party's candidate, and then we split unaffiliated voters evenly between the two parties. FiveThirtyEight's models such as the 2020 President and Senate models allocate undecided voters in polls 50/50 as well, so our assumption is in line with standard approaches. However, other reasonable approaches could have been to allocate unaffiliated voters proportionally to partisan support within the state of interest or even to calculate the split based on Trump's overperformance of polls in 2016 based on winning over undecided voters (Golshan, 2016). Such approaches have been explored in the past by FiveThirtyEight (Silver, 2008) but are not the norm.

Conclusion

In conclusion, the probability that Trump wins re-election is 0.23. The predicted electoral colleges votes for Trump is 238. We predict that the US Senate is unlikely to remain in Republican control (in fact, we predict there is only a 0.1 chance that Republicans retain control) For the House elections in NC's 13 districts, it appears that the Republican candidate will win in districts 3, 5, 7, 8, 9, 10, 11, and 13 and that the Democratic candidate will win districts 1, 2, 4, and 6. For the NC Senate race (Tillis vs. Cunningham), the predicted vote share for Cunningham is 51.78%, and we predict that he will win by a small margin. *check to see if nothing changes*

Key parts of our model include the incorporation of a state covariance matrix and the use of economic fundamentals in our presidential model, the incumbency adjustment in our Senate model, and the inclusion of voter turnout information in our NC House model. As mentioned before, there are several limitations, but overall, we believe our analysis to be thorough and robust.

Appendix A

A.1 Presidential Election Model

Presidential Model Purpose and Structure

The purpose of this model is to predict the probability that president Trump will win the election, as well as to predict the electoral college vote. The states chosen are Arizona, Florida, Georgia, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Texas, and Wisconsin. These thirteen states are deemed to be competitive in the 2020 presidential race by political analysts.

The response we will use is Y_k, Biden's share of the two-party vote using the following model where the poll k (of K=3361 [insert final number] polls) ended t days (of $T = 30$ days) before the election and was conducted on state j (of J=13 states of interest). We choose to use polling data from 30 days before the election (as opposed to over 30 days), as this aligns with Andrew Gelman's observation that polling data becomes more predictive as the day of poll becomes closer to election day (The Economist).

$Y_k \sim N(\theta_{jt}, \sigma_{yj}^2)$
$\theta_{...t} \sim MVN(\theta_{...t-1}, \Sigma)$
$\Sigma \sim Wishart(S, J+1)$ where S is the state covariance matrix (obtained from Andrew Gelman's model) and J is the number of states in the model.
Priors for σ_{yj}^2: $\sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
Priors for θ_{j1}: $\theta_{j1} \sim N(\mu_j, \sigma^2)$
$\sigma^2 \sim InvGamma(0.5, 0.5)$
$\mu_j \sim N(h_j, 7.5^2)$
$h_j = 0.1 * \text{Presidential } \hat{\text{Fundamentals}}_{2020} + 0.9 * \text{Vote Share from Partisan Lean}_j$
$\text{Presidential Fundamentals}_i = 100 - \text{Incumbent Party Vote Share}_i$
$\text{Incumbent Party Vote Share}_i = \gamma_0 + \gamma_1 \text{June Approval Rating for Incumbent Party}_i + \gamma_2 \text{Three Month Stock Growth}_i + \gamma_3 \text{2nd Quarter Real Income Growth}_i + \epsilon_i$
$\epsilon_i \sim N(0, \phi^2)$

A hierarchical model was used, where Biden's share of the two-party vote for poll k on any given day t for state j is modelled as a normal distribution with mean θ_{jt} and variance σ_{yj}^2. A random walk was used to calculate θ_{jt} for each day t before the election (up until 30 days before the election) from θ_{jt-1} and Σ, a JxJ state covariance matrix. The mean h_j of the normal prior on μ_j (which is the mean of the normal prior on θ_{j1}) was calculated by multiplying the economic fundamentals Presidental Fundamentals$_i$ for $i = 2020$ by 0.1 and state partisan lean by 0.9, then summing those two values.

We predicted Incumbent Party Vote Share$_{2020}$ (i.e. Republican Party's national vote share in the 2020 election) from a linear regression model that had historical election year's June approval rating of the incumbent party, three month stock growth, and 2nd quarter real income growth as predictors. We borrowed this idea from the "Time for Change" model which assumes that presidential election is referendum on the performance of incumbent president (Alan Abramowitz, 2008). Then we used 100 minus Incumbent Party Vote Share$_{2020}$ to obtain the predicted vote share for Democrats (Presidential Fundamentals$_{2020}$).

Raw Model Output

```
##                      mean          sd         2.5%        97.5%
## Sigma[1,1]     0.5926277 0.4169338  0.1401836021   1.7274851
## Sigma[2,1]     0.4251874 0.3387688 -0.0002283602   1.2701249
## Sigma[3,1]     0.2440198 0.3449571 -0.2321267767   1.1510161
## Sigma[4,1]     0.3132443 0.2933798 -0.0848882218   1.0270067
## Sigma[5,1]     0.3266411 0.2721255 -0.0305401182   1.0137364
```

```
## Sigma[6,1]     0.3040325 0.2888737 -0.0476364569  1.0713749
## Sigma[7,1]     0.2275937 0.2554548 -0.1548570273  0.8166067
## Sigma[8,1]     0.2820362 0.3265992 -0.1752488754  1.0902519
## Sigma[9,1]     0.4943119 0.4183646  0.0524378585  1.5207589
## Sigma[10,1]    0.3178951 0.2708774 -0.0233583632  1.0321280
## Sigma[11,1]    0.3589893 0.2818443 -0.0119504340  1.1193115
## Sigma[12,1]    0.4009916 0.3857631 -0.0477329538  1.3872376
## Sigma[13,1]    0.3061326 0.2816699 -0.0846494342  1.0119030
## sigma2_y[1]    2.0631807 0.4200509  1.3870137066  2.9433971
## sigma2_y[2]    1.9571972 0.3755882  1.3275332882  2.8096934
## sigma2_y[3]    2.1106819 0.4542211  1.4196045776  3.2142118
## sigma2_y[4]    2.0630892 0.4452591  1.3583745536  3.1091737
## sigma2_y[5]    2.2417661 0.4562729  1.5720275705  3.2751637
## sigma2_y[6]    1.9866924 0.4300040  1.2666244420  2.9180791
## sigma2_y[7]    1.8909327 0.3703109  1.2359783462  2.6689607
## sigma2_y[8]    1.9989117 0.4777182  1.2531722490  2.9656064
## sigma2_y[9]    2.0168268 0.4603962  1.2751520466  3.0331654
## sigma2_y[10]   1.9445118 0.4149592  1.2135413636  2.8749040
## sigma2_y[11]   1.9940210 0.3782283  1.3578814364  2.8688157
## sigma2_y[12]   1.9642988 0.4210706  1.2395456762  2.8454045
## sigma2_y[13]   1.9782000 0.3692536  1.3651748113  2.8174252
## theta[1,1]    48.0194135 2.0836772 43.6142452243 51.9545187
## theta[2,1]    48.7149570 2.1791798 44.1220274640 53.1787715
## theta[3,1]    48.0160544 2.1220506 43.5233782490 51.8645864
## theta[4,1]    47.8690570 2.2645322 42.7809010828 51.4733646
## theta[5,1]    51.4117743 1.7444836 48.3715016168 54.8335647
## theta[6,1]    51.4797346 1.7345245 47.7037921433 54.2858014
## theta[7,1]    49.3298396 1.7453197 46.0742712849 52.7612071
## theta[8,1]    53.1836658 2.8757671 46.5072760843 58.1831573
## theta[9,1]    50.1423759 2.6569660 44.6985116032 55.1245100
## theta[10,1]   46.0793044 1.7337032 42.8994507747 49.2815498
## theta[11,1]   50.1685276 1.6677493 46.8360317555 53.3025424
## theta[12,1]   47.1694021 2.5359719 41.2455675411 51.5600100
## theta[13,1]   51.0595498 1.8395423 47.2871151522 54.2248898

## mean Rhat: 1.017614

## mean effective sample size: 860.6576
```

Since the outcome of the election is decided by Electoral College vote, for each set of MCMC samples of the two-party vote share by state, we use the predicted winner in each state j on the day of the election (β_{j1}) and sum the electoral votes from those states for each candidate. For states not in the model, votes were given to the candidates based on how the states voted in 2016. The probability of President Trump winning re-election is the probability that he receives 270 or more electoral votes across simulations. (Note: Maine and Nebraska do not have a "winner-take-all" framework for allocating votes, but for simplicity, we assume that they have the same vote-allocating procedures as the other states).

A.2 Senate Election Model

The purpose of the Senate model is to predict whether the U.S. Senate remains in Republican control (and to predict the outcome for the NC Senate race). The states included in the model are Alaska, Arizona, Colorado, Georgia, Iowa, Kansas, Maine, Michigan, Montana, North Carolina, South Carolina, and Texas. These states were chosen because they are deemed by political analysts as being competitive states for the Senate race in 2020. There is also a special election for the Senate in Georgia in addition to the regularly scheduled election.

For the Georgia special election, our model sums up the support for Republican candidates and compares that to the sum of the support for Democratic candidates, and treats this election in the same way as the other states.

The response we will use is Y_k, the Democrat candidate's share of the two-party vote using the following model where the poll k (of K=*fill in* polls) ended t days (of $T = 30$ days) before the election and was conducted on state j (of J=14 states of interest).

$Y_k \sim N(\beta_{jt}, \sigma_{yj}^2)$
$\beta_{jt} \sim N(\beta_{jt-1}, \sigma_{\beta j}^2)$
Priors for σ_{yj}^2: $\sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
Priors for $\sigma_{\beta j}^2$ $\sigma_{\beta j}^2 \sim InvGamma(\nu_\beta, \nu_\beta \tau_\beta))$
$\nu_\beta \sim Uniform(0, 100)$ and $\tau_\beta \sim Uniform(0, 100)$.
Priors for β_{j1} $\beta_{j1} \sim N(\mu_j, \sigma^2)$
$\sigma^2 \sim InvGamma(0.5, 0.5)$
$\mu_j \sim N(h_j, 7.5^2)$
$h_j = $ Vote Share from Partisan Lean$_j$ + Incumbency Advantage$_j$

A hierarchical model was used, where the Democratic candidate's share of the two-party vote for poll k on any given day t for state j is modeled as a normal distribution with mean β_{jt} and variance σ_{yj}^2. A random walk was used to calculate β_{jt} for each day t before the election (up until 30 days before the election) from β_{jt-1} and $\sigma_{\beta j}$. The mean h_j of the normal prior on mu_j (which is the mean of the normal prior on β_{j1}), was calculated by the sum of the Vote Share from Partisan Lean$_j$ and Incumbency Advantage$_j$.

Raw Model Output

```
##                      mean          sd        2.5%       97.5%
## beta[1,1]       46.4507584 2.2433708 41.59965446 50.6435166
## beta[2,1]       52.3060134 2.1867985 46.72905116 55.6747117
## beta[3,1]       54.3407636 1.7697115 50.52703121 57.6127336
## beta[4,1]       48.6580340 1.6538193 45.60341428 51.9550692
## beta[5,1]       45.5772068 1.9532845 41.67556604 49.6975127
## beta[6,1]       50.7226677 1.6592584 47.18653081 53.7717883
## beta[7,1]       46.2249577 2.6063537 40.90148615 51.1973216
## beta[8,1]       52.0308945 2.2519892 47.08007333 55.9982111
## beta[9,1]       53.9554866 1.7664757 50.67107557 57.9060076
## beta[10,1]      47.5874120 2.0102568 43.45416459 51.5177091
## beta[11,1]      51.7831808 1.6530492 48.20280605 54.9907635
## beta[12,1]      48.8159096 1.9127126 45.15421200 52.8553178
## beta[13,1]      45.7833702 1.8480167 42.38430159 49.9298690
## sigma2_beta[1]   0.2174465 0.1709002  0.03591539  0.6455778
## sigma2_beta[2]   0.2195665 0.1769982  0.03763168  0.6657771
## sigma2_beta[3]   0.2117813 0.1594372  0.03599774  0.6146854
## sigma2_beta[4]   0.2150078 0.1639863  0.03653461  0.6168701
## sigma2_beta[5]   0.2248891 0.1892429  0.03824851  0.6904631
## sigma2_beta[6]   0.2127312 0.1623794  0.03671041  0.6352535
## sigma2_beta[7]   0.2177799 0.1719668  0.03634232  0.6345328
## sigma2_beta[8]   0.2146970 0.1668502  0.03705036  0.6189533
## sigma2_beta[9]   0.2103161 0.1587663  0.03750246  0.6029883
## sigma2_beta[10]  0.2202669 0.1761017  0.03688560  0.6480835
## sigma2_beta[11]  0.2111939 0.1591431  0.03804373  0.5951090
## sigma2_beta[12]  0.2172667 0.1701173  0.03499755  0.6603020
## sigma2_beta[13]  0.2134587 0.1636727  0.03839452  0.6263231
```

```
## sigma2_y[1]        4.3393222 2.2137569  2.06291739  8.9778112
## sigma2_y[2]        3.8692044 1.0191342  2.27841397  6.2710692
## sigma2_y[3]        3.4111862 1.1556115  1.48423175  6.0022095
## sigma2_y[4]        3.9842974 1.2215696  2.22593396  6.8663237
## sigma2_y[5]        7.4445583 4.5573201  3.38615960 20.3325823
## sigma2_y[6]        3.7518461 1.1728339  1.95333042  6.6424013
## sigma2_y[7]        4.3641806 3.0370416  1.89056380 10.3323020
## sigma2_y[8]        4.2269202 2.2465433  1.94931827  8.4792981
## sigma2_y[9]        3.5124807 0.9368648  1.89915246  5.5939477
## sigma2_y[10]       3.9496067 1.6799443  1.69059348  7.7787986
## sigma2_y[11]       3.6043268 0.9004233  2.11599421  5.7642380
## sigma2_y[12]       4.2285806 1.8916422  1.99432151  8.2372830
## sigma2_y[13]       3.7806794 1.1815197  1.90567452  6.6652333

## mean Rhat: 1.077681

## mean effective sample size: 710.0193
```

For predicting control of the U.S. Senate, if a 50/50 split is predicted, then the VP breaks the tie. In this case, we predict Democrat majority since our presidential model gives Biden a high chance of winning (and a Democrat VP will break the tie).

A.3 House Election Model

House Model Purpose and Structure

The purpose of the House model is to predict which candidate in each of NC's 13 congressional districts will win in their respective elections.

The response we will use is Y_k, the Democrat candidate's share of the two-party vote using the following model where the poll k (of K=*fill in* polls) ended t days (of $T = 60$ days) before the election and was conducted on district j (where j is in districts $1, \cdots, 11, 13$ of North Carolina). For district 12, the vote share is coded as 100 because there is only one candidate, and she is a Democrat. Note that, due to the scarcity of poll responses for the NC House elections, we included all poll responses within 115 days of the election.

$Y_k \sim N(\beta_{jt}, \sigma_{yj}^2)$
$\beta_{jt} \sim N(\beta_{jt-1}, \sigma_{\beta j}^2)$
$\sigma_{yj}^2 \sim InvGamma(\nu_u, \nu_y \tau_y)$
$\nu_y \sim Uniform(0, 100)$ and $\tau_y \sim Uniform(0, 100)$.
$\sigma_{\beta j}^2 \sim InvGamma(\nu_\beta, \nu_\beta \tau_\beta))$
$\nu_\beta \sim Uniform(0, 100)$ and $\tau_\beta \sim Uniform(0, 100)$.
$\beta_{j1} \sim N(h_j, 7.5)$
$h_j = 0.9 * \text{Vote Share from Partisan Lean}_j + 0.1 * \text{Expected Vote Share from Voter Turnout}_j$

A hierarchical model was used, where the Democratic candidate's share of the two-party vote for poll k on any given day t for district j is modelled as a normal distribution with mean β_{jt} and variance σ_{yj}^2. A random walk was used to calculate β_{jt} for each day t before the election (up until 30 days before the election) from β_{jt-1} and $\sigma_{\beta j}$. The mean h_j of the normal prior on mu_j is computed by multiplying Vote Share from Partisan Lean$_j$ by 0.9 and Expected Vote Share from Voter Turnout$_j$ by 0.1, then summing those two values.

Interim Report Model

$$Vote_i \sim Binomial(n_i, \pi_i)$$

$$logit(\pi_i) = \alpha_0 + \Sigma_{j=1,2,3}\beta_j * I(Income_i = Income\ Category_s) + \Sigma_{j=5,6}\beta_j * I(Gender_i = Gender\ Category_g) +$$

$$\Sigma_{j=7,8}\beta_j * I(Race_i = Race\ Category_r) + \Sigma_{j=9,10,11}\beta_j * I(Age_i = Age\ Category_a) +$$

$$\Sigma_{j=12,13}\beta_j * I(Party_i = Party\ Category_p) +$$

$$\Sigma_{j=14,...,17}\beta_j * I(Gender_i = Gender\ Category_g)I(Party_i = Party\ Category_p) +$$

$$\Sigma_{j=18,...,21}\beta_j * I(Race_i = Race\ Category_r)I(Party_i = Party\ Category_p) +$$

$$\Sigma_{j=22,...,27}\beta_j * I(Gender_i = Gender\ Category_g)I(Age_i = Age\ Category_a) +$$

$$\Sigma_{j=28,...,33}\beta_j * I(Party_i = Party\ Category_p)I(Age_i = Age\ Category_a)$$

where:

$$Income\ Category_s \in \{Median\ Income\ \$46864 - \$52798,\ Median\ Income\ \$52798 - \$64509,$$
$$Median\ Income > \$64509\}$$

$$Gender\ Category_g \in \{Male, Unspecified\}$$
$$Race\ Category_r \in \{Black, Other\}$$
$$Age\ Category_a \in \{Age\ 30 - 44,\ Age\ 45 - 59,\ Age\ 60+\}$$
$$Party\ Category_p \in \{Republican,\ Other\}$$

Where π_i is the probability that subgroup i votes. $Vote_i$ is the number of people in the ith subgroup that voted in the 2016 election. The baseline is `Median Income < $46864` for income, `Female` for gender, `White` for race, `18-29` for age, and `Democrat` for party affiliation.

We fit the voter turnout model from the Interim Report on the 2020 registered voter dataset. The predicted values were the number of people that will vote for each demographic subgroup in each congressional district. We made a reasonable assumption that if someone is registered as a Democrat, that they will indeed vote for the Democratic candidate (and made the same assumption for registered Republicans). If the demographic subgroup had third/unaffiliated party, then we split their predicted number of voters evenly between the Democrat and Republican parties. From this, we calculated Democratic party vote share for each district by taking the ratio of the predicted number of Democratic voters to total number of registered voters in that district.

Raw Output for Appendix

```
##                    mean         sd        2.5%       97.5%
## beta[1,1]       58.306127   2.716740 52.97769754  63.675746
## beta[2,1]       59.253823   2.777571 53.83087333  64.707466
## beta[3,1]       36.385167   2.670935 31.43679257  41.684174
## beta[4,1]       66.711295   2.729256 61.52344576  72.155486
## beta[5,1]       30.463614   2.735714 25.07016722  35.796273
## beta[6,1]       58.811408   2.742312 53.36276143  64.131692
## beta[7,1]       38.194965   2.812420 32.56643960  43.595922
## beta[8,1]       44.256514   2.564823 38.99160944  49.328885
## beta[9,1]       43.211496   2.580971 38.10132834  48.597018
## beta[10,1]      29.324002   2.730484 24.04375613  34.836887
## beta[11,1]      43.880322   2.187492 39.40863985  48.139221
## beta[12,1]      68.969534   2.744467 63.67742984  74.298244
## beta[13,1]      30.396369   2.731562 24.98980088  35.883334
## sigma2_beta[1]   1.983584   2.619386  0.06743277   9.811024
## sigma2_beta[2]   2.123004   6.608926  0.06874763   9.154619
## sigma2_beta[3]   1.998874   2.700569  0.06760733   9.785534
```

```
## sigma2_beta[4]     1.990037     2.628090  0.06989916    9.827270
## sigma2_beta[5]     2.177180    10.277233  0.06870609    9.399325
## sigma2_beta[6]     2.009647     2.882833  0.07021266    9.610445
## sigma2_beta[7]     2.016162     2.950729  0.07050820    9.488029
## sigma2_beta[8]     1.922471     2.481571  0.07470821    9.502068
## sigma2_beta[9]     1.990683     2.643962  0.06224621    9.629270
## sigma2_beta[10]    2.011775     3.124264  0.06929920    9.668897
## sigma2_beta[11]    1.903645     2.461996  0.07902105    9.060829
## sigma2_beta[12]    2.002740     2.699599  0.06953311    9.788253
## sigma2_beta[13]    1.980313     2.623479  0.07023513    9.407093
## sigma2_y[1]      110.421134 4140.772763  3.63094568   98.700218
## sigma2_y[2]       37.433096  219.283024  3.65723244  100.172850
## sigma2_y[3]       32.606741   32.253440  3.60571168   97.711550
## sigma2_y[4]       32.953798   63.934587  3.77973266   97.786865
## sigma2_y[5]       37.302642  226.543715  3.70241168   98.689607
## sigma2_y[6]       32.556122   31.099612  3.83433114   99.768413
## sigma2_y[7]       43.575421  555.350283  3.68651670  100.398992
## sigma2_y[8]       31.245772   26.743780  3.66679587   94.702051
## sigma2_y[9]       32.297343   28.016058  3.94052101   97.735543
## sigma2_y[10]      32.578443   31.998891  3.69421505   99.966396
## sigma2_y[11]      31.093843   27.015430  3.59150699   95.332058
## sigma2_y[12]      32.599110   38.049204  3.64701778   98.692937
## sigma2_y[13]      35.672709   90.873938  3.74207455  101.658292

## mean Rhat: 1.116908

## mean effective sample size: 2018.325
```

Appendix B

Presidential Election Model:

- `2020 US presidential election polls - all_polls.csv`: 2020 U.S. Presidential Election Polls compiled by *The Economist.* Available at https://projects.economist.com/us-2020-forecast/president/how-this-works. + We only used poll responses from likely voters in the battleground states within 30 days prior to the election date. From the poll responses we calculated Biden's vote share in the two-party race in each poll (Y_k).

- `partisan_leans_538.csv`: effect of partisan lean on the Democratic candidate's presidential election vote share in each state. Available at https://github.com/fivethirtyeight/data/tree/master/partisan-lean. + 90% weight is allocated to this variable Vote Share from Partisan Lean$_j$ when computing hyperparameter h_j for each state j.

- `abramowitz_data.csv`: downloaded from Andrew Gelman's Github repository on the 2020 Presidential Election. Available at https://github.com/TheEconomist/us-potus-model. + This data set contains information on each of the 1948-2016 election year i's annualized 2nd quarter GDP performance (not used due to its abnormally low value in 2020), incumbent party's June net approval rating (June Approval Ratings for Incumbent Party$_i$) and incumbent party's national vote share (Presidental Fundamentals$_i$).

- `states_cov_matrix_full.csv`: covariance matrix of states which takes into account similarity between states based on their demographic and political profiles. Retrieved from Andrew Gelman's Github repository on the 2020 Presidential Election (https://github.com/TheEconomist/us-potus-model). + The covariance matrix corresponds to S, the scale matrix in the prior distribution of Σ

- `abramowitz_additional.csv`: we supplemented `abramowitz_data.csv` with the corresponding year i's 2nd quarter real income growth compared to one year ago (2nd Quarter Real Income Growth$_i$) and S&P stock performance 3 months prior to the election date (Three Month Stock Growth$_i$). Data sets are from https://fred.stlouisfed.org/series/A067RO1Q156NBEA and https://www.multpl.com/s-p-500-historical-prices/table/by-month. A subset of the combined data set is displayed below. *check where the table is* + We fitted a linear regression model on historical incumbent party's national vote share with June Approval Ratings for Incumbent Party, 2nd Quarter Real Income Growth, and Three Month Stock Growth as predictors. We then predicted this year's incumbent party's national vote share (i.e. Republican's national vote share) and used 100 minus the predicted value to obtain Democrat's national vote share (Presidental Fundamentals$_{2020}$). We can think of Presidental Fundamentals$_{2020}$ as a national level prior. 10% weight is allocated to the "national prior" when computing the state specific prior h_j.

Year	Incumbent Party Vote Share	Democrat Incumbency	Incumbent Party June Net Approval	2nd Quarter GDP Growth	3 Month Stock Growth	2nd Quarter Real Income Growth
2016	51.10	Yes	4	1.1	0.0033832	1.8
2012	52.00	Yes	1	1.3	-0.0338848	2.4
2008	46.30	No	-40	0.6	-0.2743827	0.0
2004	51.24	No	-1	2.9	0.0458816	2.5
2000	50.30	Yes	15	8.0	-0.0613126	5.8
1996	54.70	Yes	10	7.1	0.0900753	3.3

Senate Election Model:

- `senate_polls.csv`: Senate election polls compiled by *FiveThirtyEight.* Available at https://projects.fivethirtyeight.com/polls-page/senate_polls.csv. + Similar to the presidential election prediction model,

we only included poll responses from likely voters in the battleground states within 30 days prior to the senate election date. From the poll responses we calculated Democratic candidates' vote share in the two-party race in each poll (Y_k).

- `partisan_leans_538.csv`: same as the data set in presidential election model and we used this partisan lean in the calculation of prior hyperparameter h_j.
- We also incroporated incumbency advantage in calculating h_j. As *FiveThirtyEight* suggests, incumbency advantage is on average 2.6 for senate incumbents (https://fivethirtyeight.com/features/how-much-was-incumbency-worth-in-2018/).
- For NC Senate race only, we supplemented the prior with predicted voter turnout from the Interim Report model.

House Election Model (for North Carolina):

- `house_polls.csv`: House election polls compiled by *FiveThirtyEight*. Available at https://projects.fivethirtyeight.com/polls-page/house_polls.csv. + We only used polls related to the NC House race. Due to the lack of poll responses for the NC House elections, we included all poll responses within 115 days of the election.
- `ncvoter_1027_small.rds`: NC registered voter demographics information provided by the NC State Board of Elections (https://dl.ncsbe.gov/list.html). + Please refer to the Interim Report to see a detailed explanation of how voter turnout is modeled by registered voters' race, age, party affiliation, income and gender.
- Partisan lean and incumbency information was taken from *FiveThirtyEight* at the following links, respectively: https://fivethirtyeight.com/features/north-carolinas-new-house-map-hands-democrats-two-seats-but-it-still-leans-republican/, https://fivethirtyeight.com/features/how-much-was-incumbency-worth-in-2018/

Appendix C

Additional EDA

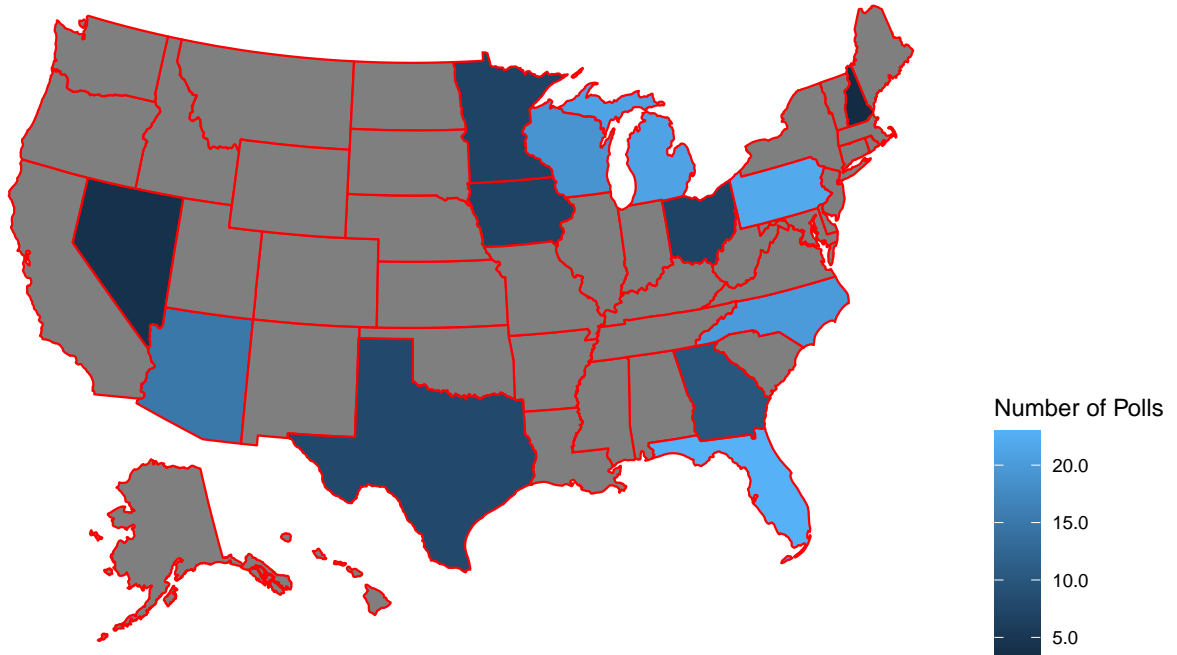Map For The Number of Filtered Polls Among States

Figure 9: Additional Presidential Election Data Visualization
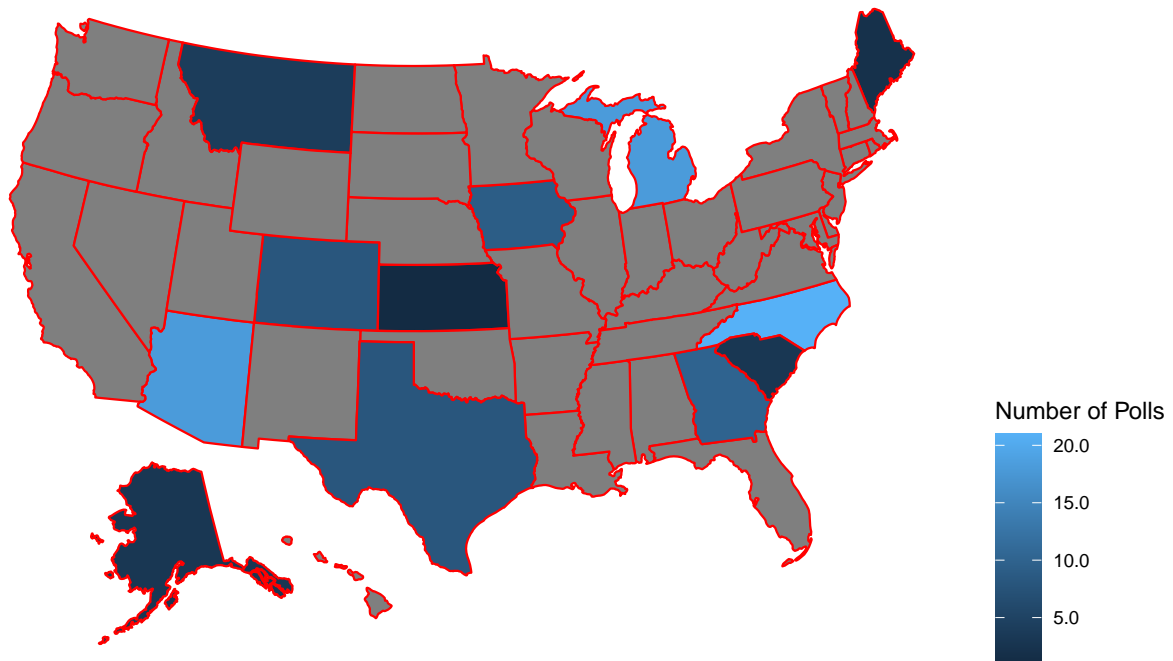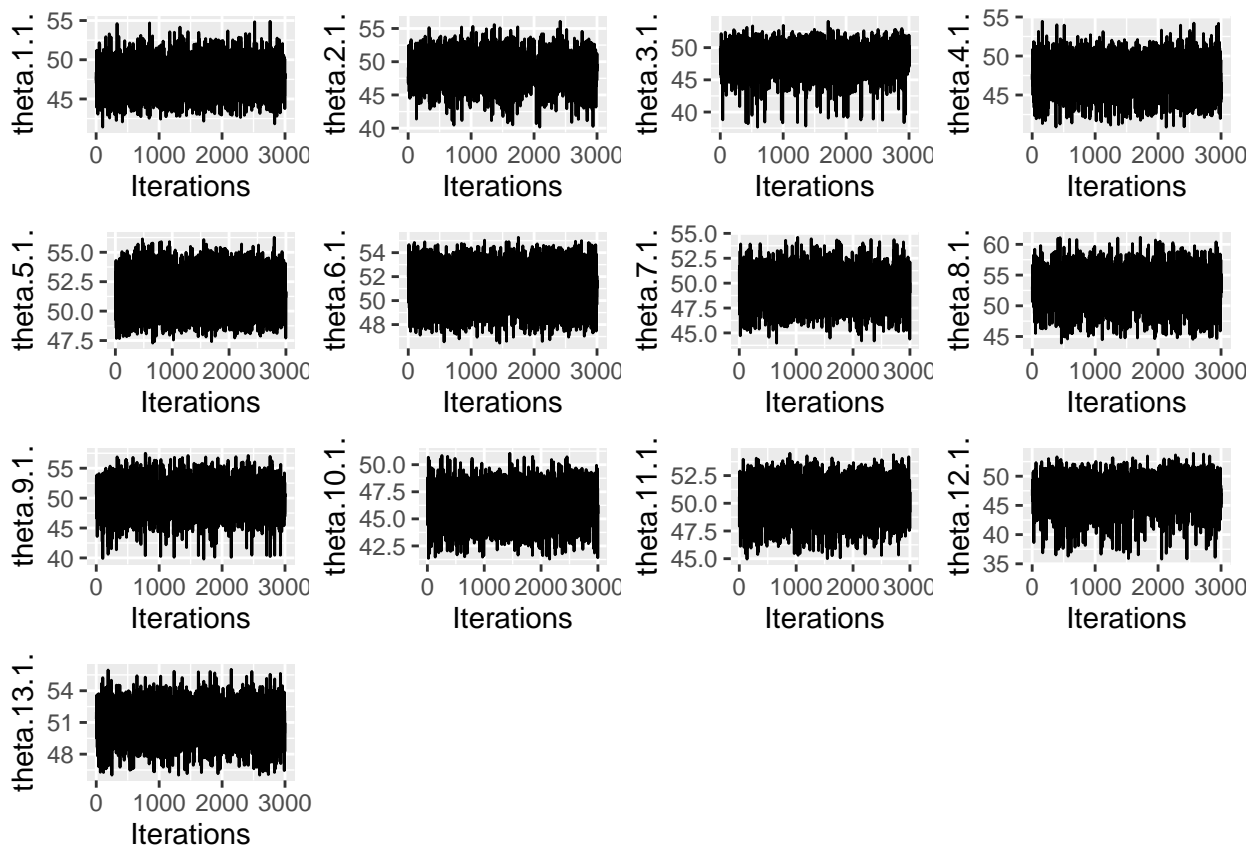
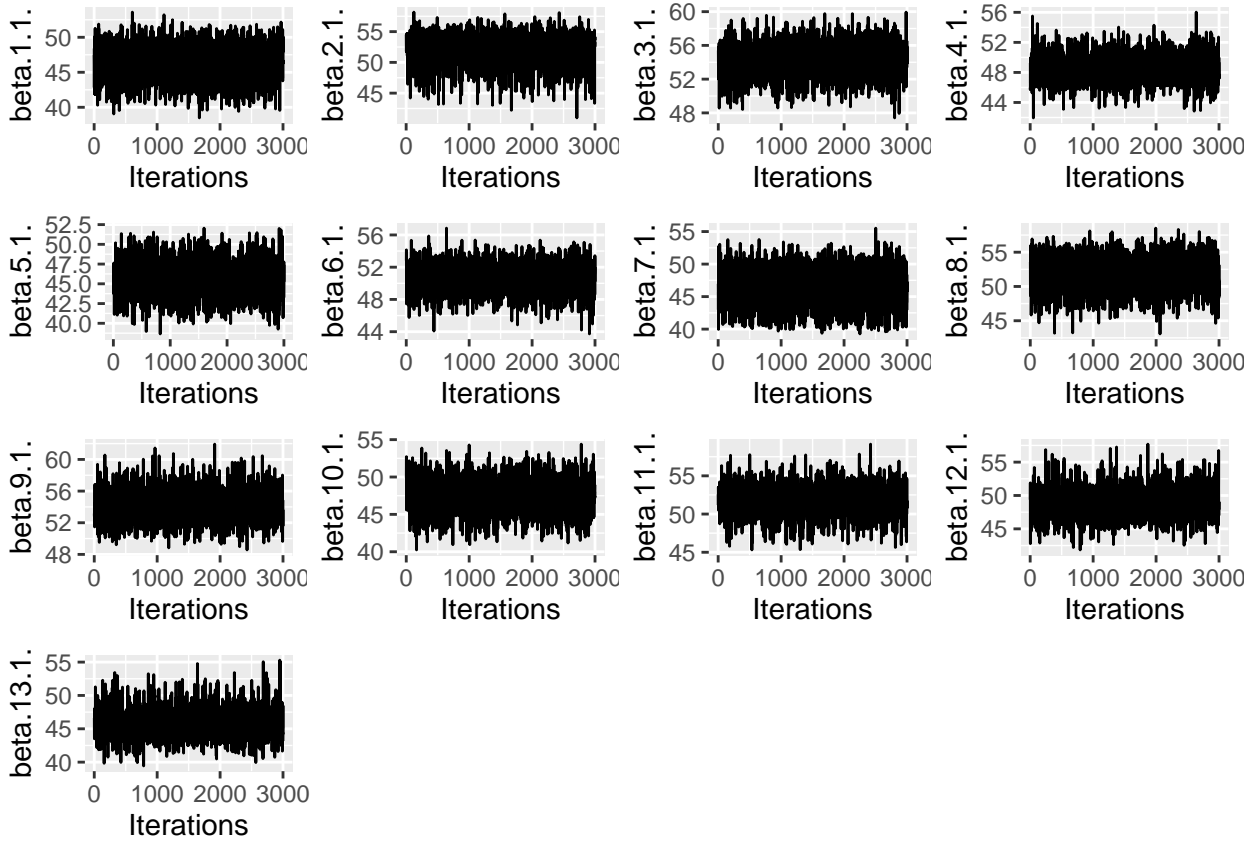Map For The Number of Filtered Polls Among States

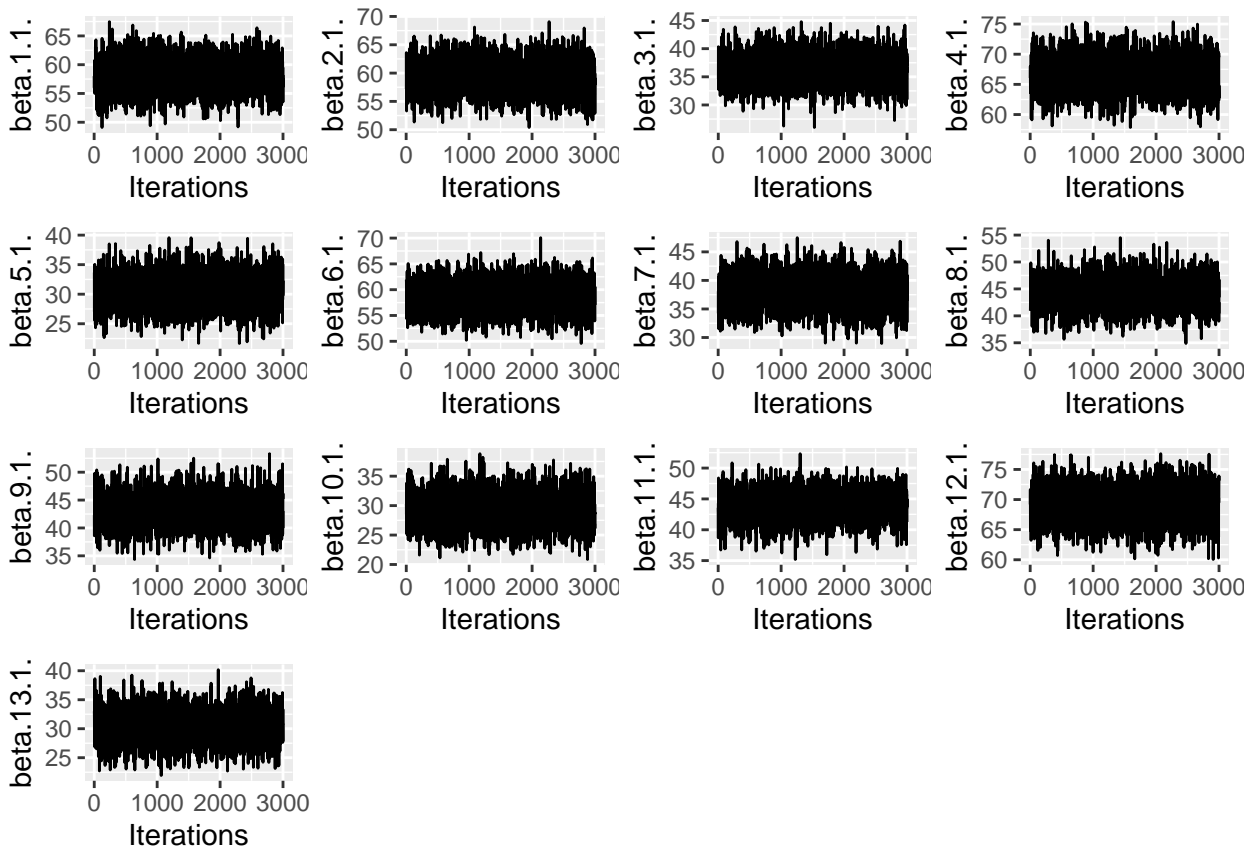Figure 10: Additional Senate Election Data Visualization

Traceplots for Presidential Model

Traceplots for Senate Model

Traceplots for House Model
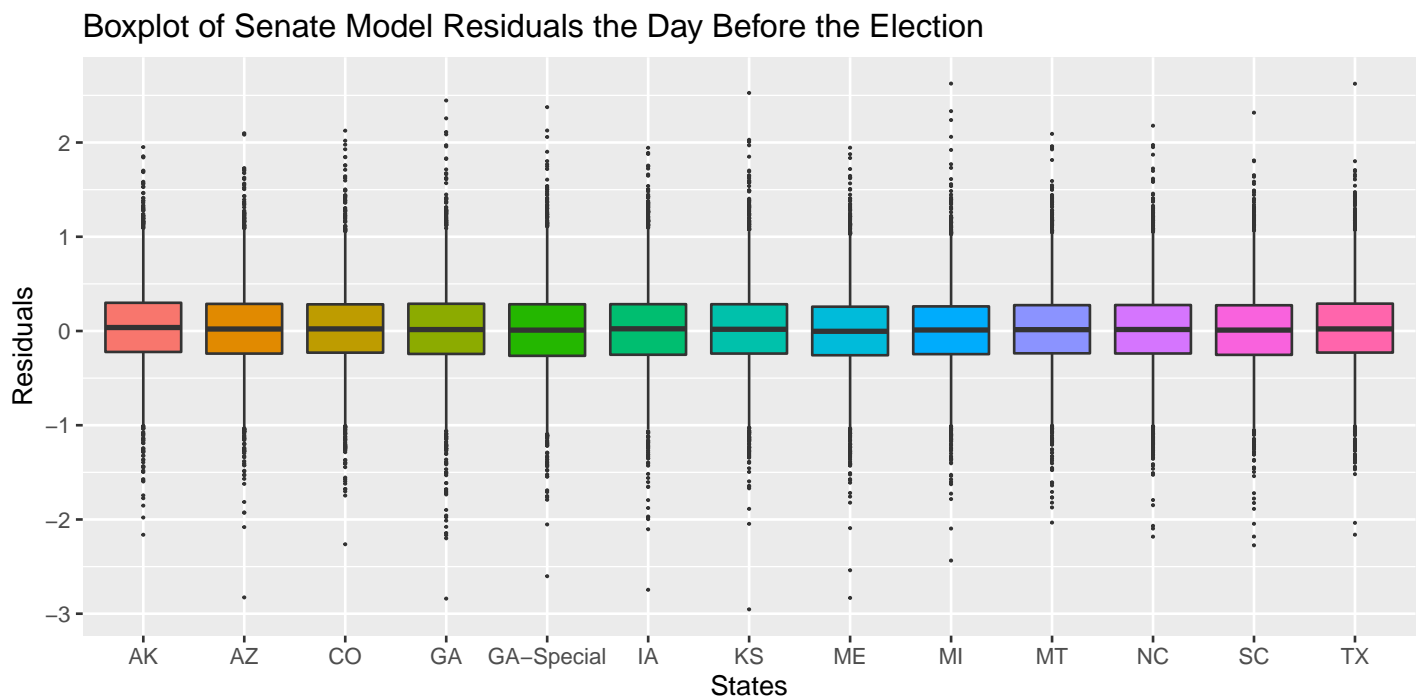
Residual Plot for Senate Model

Boxplot of Senate Model Residuals the Day Before the Election

Figure 11: Senate Model Residual Plot

Residual Plot for House Model

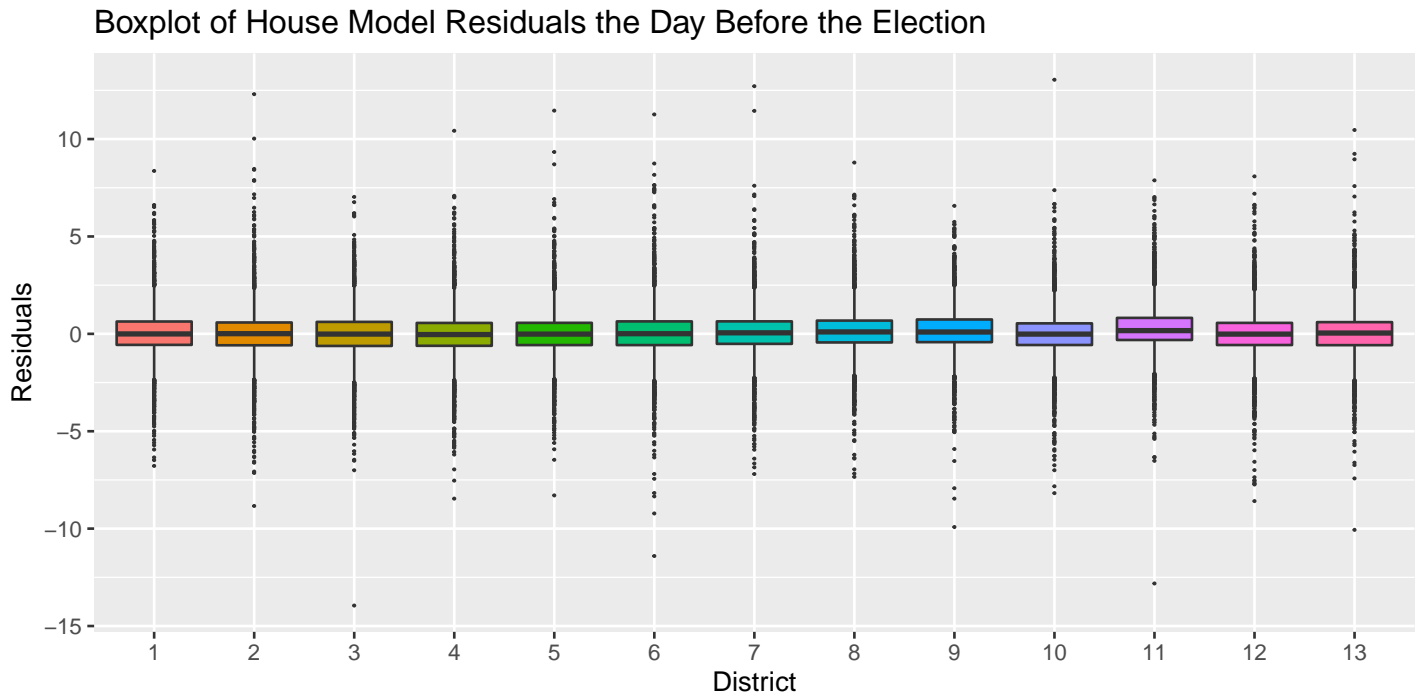Boxplot of House Model Residuals the Day Before the Election

Figure 12:   House Model Residual Plot

References

1. Linzer, D. A. (2013). Dynamic Bayesian Forecasting of Presidential Elections in the States. Journal of the American Statistical Association, 108(501), 124-134. doi:10.1080/01621459.2012.737735

2. Hansford, T. G., & Gomez, B. T. (2010). Estimating the Electoral Effects of Voter Turnout. American Political Science Review, 104(2), 268-288. doi:10.1017/s0003055410000109

3. 2020 Election. (2020, October 20). Retrieved October 20, 2020, from https://fivethirtyeight.com/politics/elections/

4. Park, D. K., Gelman, A., & Bafumi, J. (2006). State-Level Opinions from National Surveys:. Public Opinion in State Politics, 209-228. doi:10.2307/j.ctvr33bdg.17

5. Mahler, V. A., Jesuit, D. K., & Paradowski, P. R. (2013). Electoral Turnout and State Redistribution. Political Research Quarterly, 67(2), 361-373. doi:10.1177/1065912913509306

6. Uhlaner, C. J., & Scola, B. (2015). Collective Representation as a Mobilizer. State Politics & Policy Quarterly, 16(2), 227-263. doi:10.1177/1532440015603576

7. Godbout, J. (2012). Turnout and presidential coattails in congressional elections. Public Choice, 157(1-2), 333-356. doi:10.1007/s11127-012-9947-7

8. Kim, S. S., Alvarez, R. M., & Ramirez, C. M. (2020). Who Voted in 2016? Using Fuzzy Forests to Understand Voter Turnout. doi:10.33774/apsa-2020-xzx29

9. Weinschenk, A. C. (2019) That's Why the Lady Lost to the Trump: Demographics and the 2016 Presidential Election, Journal of Political Marketing, 18:1-2, 69-91, DOI: 10.1080/15377857.2018.1478657

10. Charles, K. K., & Stephens, M. (2011). Employment, Wages and Voter Turnout. doi:10.3386/w17270

11. Hills, M. (2020, September 25). US election 2020: A really simple guide. Retrieved October 20, 2020, from https://www.bbc.com/news/election-us-2020-53785985

12. Railey, K. (2016). Federal Judges Let Stand North Carolina,'s New Congressional Map. The Hotline. https://link.gale.com/apps/doc/A498010836/ITOF?u=duke_perkins&sid=ITOF&xid=119d6ad9

13. Perrin, A. J., & Ifatunji, M. A. (2020). Race, Immigration, and Support for Donald Trump: Evidence From the 2018 North Carolina Election. Sociological Forum, 35(S1), 941-953. doi:10.1111/socf.12600

14. Redistricting in North Carolina. (2020). Retrieved October 21, 2020, from https://ballotpedia.org/Redistricting_in_North_Carolina

15. Â§ 132-1. Public Records. https://www.ncleg.gov/EnactedLegislation/Statutes/PDF/BySection/Chapter_132/GS_132-1.pdf

16. Â§ 163-82.10. Official record of voter registration. https://www.ncleg.gov/EnactedLegislation/Statutes/PDF/BySection/Chapter_163/GS_163-82.10.pdf

17. FairVote.org. (n.d.). Voter Turnout. Retrieved October 22, 2020, from https://www.fairvote.org/voter_turnout

18. McDonald, M. P. (2020). Voter Turnout Demographics. Retrieved October 23, 2020, from http://www.electproject.org/home/voter-turnout/demographics

19. Y. Ghitza, A. Gelman's (2013). Deep interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups. American Journal of Political Science 57 762– 776.

20. Walker, H.L., Herron, M.C. & Smith, D.A. Early Voting Changes and Voter Turnout: North Carolina in the 2016 General Election. Polit Behav 41, 841–869 (2019). https://doi-org.proxy.lib.duke.edu/10.1007/s11109-018-9473-5

21. Pew Research Center (2020). Men and women in the U.S. continue to differ in voter turnout rate, party identification. https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

22. Pew Research Center (2016). Party affiliation among voters: 1992-2016. https://www.pewresearch.org/politics/2016/09/13/2-party-affiliation-among-voters-1992-2016/

23. J. Misra (2019). Voter Turnout Rates Among All Voting Age and Major Racial and Ethnic Groups Were Higher Than in 2014. https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

24. Killian, L. (2012). Five Myths About Independent Voters. Retrieved October 22, 2020, from https://www.washingtonpost.com/opinions/five-myths-about-independent-voters/2012/05/17/gIQAZmGyWU_story.html

25. Igielnik, R. (2020). Men and Women in the U.S. Continue to Differ in Voter Turnout Rate. Retrieved October 22, 2020, from https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

26. Akee, R. (2019). Voting and Income. Retrieved October 22, 2020, from https://econofact.org/voting-and-

income

27. Wikipedia. 2020 United States elections. (2020, October 27). Retrieved October 31, 2020, from https://en.wikipedia.org/wiki/2020_United_States_elections

28. Gelman, A., & Heidemanns, M. (2020). President-Forecasting the US 2020 elections. Retrieved November 01, 2020, from https://projects.economist.com/us-2020-forecast/president

29. Gelman, A., Hullman, J., Wlezien, C., & George, E. M. (2020). Information, incentives, and goals in election forecasts. Judgment and Decision Making, 15(5), 863-880.

30. Gelman, A. and Heidemanns, M. (2020). President–Forecasting the US 2020 Elections | The Economist. Retrieved October 31, 2020 from https://projects.economist.com/us-2020-forecast/president

31. The Cook Political Report (2020). 2020 Electoral College Ratings. Retrieved October 31, 2020 from https://cookpolitical.com/sites/default/files/2020-10/EC%20Ratings.102820.pdf

32. The Cook Political Report (2020). 2020 Senate Race Ratings. Retrieved October 31, 2020 from https://cookpolitical.com/ratings/senate-race-ratings

33. Ballotpedia (2020). United States Senate special election in Georgia, 2020. Retrieved October 31, 2020 from https://ballotpedia.org/United_States_Senate_special_election_in_Georgia,_2020

34. FiveThirtyEight (2020). FiveThirtyEight's Partisan Lean. Retrieved October 31, 2020 from https://github.com/fivethirtyeight/data/tree/master/partisan-lean

35. Rakich, N. (2018). How Much Was Incumbency Worth In 2018? Retrieved October 31, 2020 from https://fivethirtyeight.com/features/how-much-was-incumbency-worth-in-2018/

36. Gillespie, A. et al. (2020). Race for the Senate 2020: Experts' views from the states. Retrieved October 31, 2020 from https://www.brookings.edu/blog/fixgov/2020/10/30/race-for-the-senate-2020-experts-views-from-the-states/

37. Delsiver, D. (2020). Most Senate Elections Reflect States' Presidential Votes. Retrieved October 31, 2020 from https://www.pewresearch.org/fact-tank/2020/09/01/most-senate-elections-reflect-states-presidential-votes/

38. Silver, N. (2014) Registered Voter Polls Will (Usually) Overrate Democrats. Retrieved October 31, 2020 from https://fivethirtyeight.com/features/registered-voter-polls-will-usually-overrate-democrats/

39. The Economist. (2016). Hillary Clinton has got this. Probably. Very probably https://www.economist.com/graphic-detail/2016/11/08/hillary-clinton-has-got-this-probably-very-probably

40. Shirani-Mehr, H., Rothschild, D., Goel, S., & Gelman, A. (2018). Disentangling Bias and Variance in Election Polls. Journal of the American Statistical Association, 113(522), 607-614. doi:10.1080/01621459.2018.1448823

41. McDonald, M. (2020). 2020 General Election Early Vote Statistics. Retrieved November 01, 2020, from https://electproject.github.io/Early-Vote-2020G/index.html

42. Morley, M. (1905). Election Emergencies: Voting in the Wake of Natural Disasters and Terrorist Attacks, 67 Emory L. J. 545. Available at: https://scholarlycommons.law.emory.edu/elj/vol67/iss3/6

43. Abramawitz, A. (2008). Forecasting the 2008Presidential Election with theTime-for-Change Model. Political Science and Politics Vol. 41, No. 4 (Oct., 2008), pp. 691-695. https://www.jstor.org/stable/20452296?seq=1#metadata_info_tab_contents

44. Golshan, T. (2016). Survey: Trump Won Big Among Voters Who Decided in the Last 2 Weeks of the Campaign. Retrieved November 1, 2020 from https://www.vox.com/policy-and-politics/2016/12/21/14025198/trump-undecided-final-weeks-survey

45. Silver, N. (2008). Allocating the Undecideds. Retrieved November 1, 2020 from https://fivethirtyeight.com/features/allocating-undecideds/

46. Silver, N. (2014). How The FiveThirtyEight Senate Forecast Model Works. Retrieved November 1, 2020 from https://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/

47. Nilsen, E. (2020). Georgia's two super-competitive Senate races, explained. Retrieved October 31, 2020 from https://www.vox.com/21536826/georgia-senate-race-ossoff-loeffler-warnock-perdue

48. Skelley, G. (2020). What Pollsters Have Changed Since 2016 — And What Still Worries Them About 2020. Retrieved November 1, 2020 from https://fivethirtyeight.com/features/what-pollsters-have-changed-since-2016-and-what-still-worries-them-about-2020/

49. Cohn, N. (2020). Are the Polls Missing Republican Voters? Retrieved November 1, 2020 from https://www.nytimes.com/2020/07/29/upshot/polls-political-party-republicans.html

50. U.S. Bureau of Economic Analysis. (2020). https://www.bea.gov/news/glance