

Joe Biden projected to win Popular Vote in 2020 US Election with 51% of Vote

Data given with 4 percent margin of error

Alen Mitrovski, Xiaoyan Yang, Matthew Wankiewicz

November 2nd, 2020

Abstract

It seems like everyone has been waiting for the 2020 election, almost as soon as Donald Trump won in 2016. In our report, we use a logistic regression model along with multilevel regression with post-stratification in order to predict who will win the election. According to our model, we predict that Joe Biden will win the popular vote over Donald Trump, 51% to 49%. These results are promising for the people of the United States, as they may be getting the change in President they need.

Keywords: Forecasting, US 2020 Election, Trump, Biden, multilevel regression with post-stratification;

1 Introduction

The 2020 United States presidential election will be on November 3rd, 2020. The Republican party is represented by, current president, Donald Trump. Likewise, former vice-president, Joe Biden is representing the Democratic party. This year's election is particularly interesting due unforeseen circumstances of the COVID-19 pandemic and the civil unrest sparked by the killing of George Floyd.

For our report, we have looked at which candidate won the popular vote (ie. had the higher percentage of votes). In addition, we have calculated the total votes within each of the 50 American states in order to see how many points each candidate can accumulate for the electoral college. The electoral college is a system that was implemented in the United States to ensure that states with higher populations do not overwhelm the vote and provide greater power to individuals in historically smaller states (West). The presidential candidate with the most votes in a particular state wins all the electoral college votes for that state (West). The only exceptions to this rule are Nebraska and Maine who allocate their votes between both candidates based on the congressional district method (West). The electoral college has historically been viewed as a critical part of American democracy. However, it's modern application has come into question in 2016 when Republican candidate Donald Trump won the election by accumulating the most electoral college votes despite losing the popular vote.

The survey data for this study was obtained from the Democracy Fund Voter Study Group; an organization that took the collaborative efforts of many scholars and analysts to get a better understanding of the evolving behavior of American voters (Voter Study Group). The post-stratification data for this study was obtained from the Integrated Public Use Microdata Series (IPUMS). IPUMS is an organization that provides survey and census data through the collaboration of 105 national statistical agencies (IPUMS). For the purpose of this report, we utilized the data from the 2018 American Community Survey. We selected state, race, age, sex, education level and whether an individual was Hispanic or not as key variables to use for the post-stratification of the dataset.

The following paper will outline the methods, data, results, discussion and weaknesses of the research. Our hypothesis is that Joe Biden will win the election based on the popular vote. Our null hypothesis is that Joe

Biden will not win the election based on the popular vote. Our report is carried out using R (R Core Team [2019]), and the library Tidyverse (Wickham et al. [2019]) was used the most during the report.

2 Data

We utilized the U.S. presidential election 2020 survey data from Nationscape conducted on June 25, 2020 as the study data to predict the 2020 election winner between Trump and Biden. The census data from IPUMS America Census Service was used as the post-stratification data for the survey data weighting adjustment. In the following sections, we will introduce the variable selection basis and the main data differences between the survey and the stratification data.

2.1 Survey Data

The survey data was provided by Nationscape. Nationscape is a featured project of Democracy Fund Voter Study Group. Democracy Fund + UCLA Nationscape is one of the largest public opinion surveys ever conducted - interviewing people in nearly every county, congressional district, and mid-sized U.S. city in the leadup to the 2020 election [Tausanovitch and Vavreck, 2020]. The survey data from June 25, 2020 was selected as the study data. It had 6479 responses and 265 variables. The survey was designed to cover age, gender, race, education, religion, remarks on Trump's performance in his first term, economy, propensity to the major political parties, their leaders, or candidates, ethnicity inequality, gun policy/issues, environment issues, green energy, health care/insurance, income tax, college tuition, abortion, international trade, immigration, legalize marijuana, and COV-19 pandemic. Our goal is to select variables which can explain and predict the winner of the U.S. presidential election 2020 by applying a logistic multilevel regression model. It is crucial to understand the most important voting issues, especially the differences between Trump and Biden's supporters, and hence the selected variables must have the power to distinguish the two camps.

According to the research (1 [2020]) released on August 13, 2020, there are stark differences in how registered voters who support Donald Trump and Joe Biden view the importance of top voting issues. Foreign policy and Supreme Court appointments are comparable shares of Biden supporters and Trump supporters view as very important. The top five voting issues from Trump supporters are economy (88%), violent crime (74%), immigration (61%), gun policy (57%), and foreign policy (57%). By contrast, the largest shares of Biden supporters view health care (84%), coronavirus outbreak (82%), and racial and ethnic inequality (76%) the top three voting issues. There are substantial differences between Trump and Biden supporters on the importance of most issues, the widest gaps are on climate change (57%) and racial and ethnic inequality (52%). Based on these facts, we selected age, gender, race, hispanic, education, and states as the explanatory variables in the model. A binary variable with values "Trump" or "Biden" was used as the respondent variable. The values were extracted from survey variable `vote_2020`. `vote_2020` has 5 selections: "Donald trump", "Joe Biden", "I am not sure/don't know", "I would not vote", and "Someone else". We kept the data with "Trump" or "Biden" only and removed the remaining data with other selections. Age is one of the most important variables in our study. According to [Frey, 2020], Americans age 45 and older comprise a plurality of 2020 eligible voters (at 56%) and actual voters (61%), estimated by applying 2016 turnout rates to 2020 eligible voters. Between 2012 to 2020, seniors' share (persons age 65 and older) of eligible voters rose from 17% to 23%, while their share of actual voters rose from 22% to 26% [Frey, 2020]. From 2004 through 2016, these older groups showed a preference toward Republican candidates in presidential elections. But in the October 15 to 18, according to the New York Times/Siena College poll, they shifted toward Democrats [Frey, 2020]. Comparing 2012 and 2016, the 2020 poll shows modest gains (from 13% and 14% to 17%) in favor of Democrats from the 18-to-44 age group. But for the 45-to-64 age group, the 2020 poll shows a Democratic swing of 7 and 11 points (from -4% and -8% to 3%) from previous elections. For the 65-and-older age group, the 2020 poll shows a Democratic swing of 23 and 18 points (from -12% and -7% to 11%) from previous elections.

Another distinct feature of the 2020 election is that the young generation shows strong willingness to vote. According to [Janfaza, 2020], 51% of registered voters ages 18 to 34 say that they are extremely or very

enthusiastic to vote in 2020, compared to the 30% of registered voters in that age group who said the same in 2016. Likewise, a preview of the Harvard Institute of Politics Fall 2020 Youth Poll [Janfaza, 2020] found that 63% of young Americans say they will “definitely” vote in November as compared to the 47% who said the same in 2016. Overall, the results of Harvard’s survey project the highest youth voter turnout in 12 years. Based on our research, we divided the age group into 4 categories: ≤ 35 , 36-50, 51-65, and 65+. We believe these four categories should have strong differential power for voting issues concentrated on different age groups.

Another important variable is race. In the survey data, variables `race_ethnicity` and `Hispanic` represent people with no Hispanic and Hispanic background respectively. Variable `race_ethnicity` includes 6 categories and 15 options. The five categories include “White”, “Black, or African American”, “American Indian or Alaska Native”, “Asian”, “Pacific Islander”, and “Some other race”. “Asian” category includes 7 options: “Chinese”, “Japanese”, “Asian Indian”, “Filipino”, “Korean”, “Vietnamese”, and “Other”. According to [Budiman, 2020], More than 11 million Asian Americans will be able to vote in 2020, making up nearly 5% of the nation’s eligible voters (for this analysis, U.S. citizens ages 18 and older). The estimated number of Native Hawaiian and Other Pacific Islander alone-or-in-combination residents of the United States in 2018 is 1.6 million [5, 2020]. There is no research results or evidence to show that within Asian or Pacific Islander communities, specific voting issues have split voters into opposing tendencies. We decided to merge “Asian” and “Pacific Island” to one option. In the end, we have four options for the race variable: “white”, “black”, “asian”, and “other race, nec”

People with Hispanic heritage were classified to have a separate variable `Hispanic`. In the survey data, `Hispanic` was subdivided into 15 groups. In 1976, the U.S. Congress passed the only law in this country’s history that mandated the collection and analysis of data for a specific ethnic group: Americans of Spanish origin or descent. The language of that legislation described this group as “Americans who identify themselves as being of Spanish-speaking background and trace their origin or descent from Mexico, Puerto Rico, Cuba, Central and South America, and other Spanish-speaking countries.” This includes 20 Spanish-speaking nations from Latin America and Spain itself, but not Portugal or Portuguese-speaking Brazil. Standards for collecting data on Hispanics were developed by the Office of Management and Budget in 1977 and revised in 1997. However, the Census Bureau does not apply this definition when counting Hispanics. Rather, it relies entirely on self-reporting and lets each person identify as Hispanic or not [Lopez et al., 2020]. The U.S. Hispanic population reached a record 60.6 million in 2019, up 930,000 over the previous year and up from 50.7 million in 2010, according to newly released U.S. Census Bureau population estimates. Between 2010 and 2019, the Latino share of the total U.S. population increased from 16% to 18%. Latinos accounted for about half (52%) of all U.S. population growth over this period. They are the country’s second largest racial or ethnic group, behind white non-Hispanics. [Noe-Bustamante et al., 2020]. In 2019, the top five states which have the most Hispanic populations are California (15.57 million), Texas (11.52 million), Florida (5.66 million), New York (3.75 million), and Arizona (2.31 million) [Duffin, 2020]. Professor Raul Madrid in the Department of Government at UT who specialized in latino American politics said “Cubans are traditionally more conservative, more likely to vote Republican. Mexicans are more liberal, more likely to vote Democrat. The same is true of Puerto Ricans. Central Americans in general tend to vote more Democrat. But whereas some people from some South American countries like Venezuela tend to be more conservative” [de Leon, 2020]. Based on the findings stated above, we aggregated hispanic into five groups: “no hispanic”, “mexican”, “puerto rican”, “cuban”, “other”.

Gender plays an important role in U.S. presidential elections. Although the constitution was amended in 1920 to give women the right to vote, women were discouraged by their families from participating in politics for decades. The situation changed in the baby boomer generation. Women grew up expecting a college education and a work-life outside the home. In 1980, with the nomination of Reagan for president, the Republic Party moved sharply to the right on a handful of issues relevant to women. The party dropped into support of the Equal Rights Amendment, embraced an anti-abortion position, and courted conservative Christians who lamented the effect of working women on “traditional” families [pew, 2018]. In the next few decades, conservative elites sorted into the republican Party, while liberal elites sorted into the Democratic Party. Voters followed. This alignment had specific ramification for the gender gap. Men had consistently held more conservative positions than women on a range of issues. As the parties became more ideological

the gender gap kept growing-from 8% in 1980 to 12% in 2000, to 13% in 2016 [pew, 2018]. Back to 2020 presidential election, the pollings conducted by multiple polling organizations from June 2, 2020 to October 27, 2020 have shown that women presented strong gender gap in favoring Biden. The Center for American Women and Politics (CAWP) is tracking gender differences in support for the major party candidates in the 2020 presidential election at the national level and in battleground states [12, 2020]. The polling results show that at the national level, women had consistent support to Biden over Trump. The difference ranged from 8% to 15%. However, in the battleground states, the difference was not obvious. From June 8, 2020 to October 28, 2020, 81 polls were conducted in the battleground states. 3 pollings show (-2% OH, -3% Tex, -1% OH) of women who less favored Biden. 6 polling results were flat. The remaining results were that women voted Biden over Trump. Gender is definitely a winner of variables in our model.

The education gap is an important factor shaping two distinct and increasingly antithetical worldviews in the USA. In the 2016 election, whites with a four-year college degree or more education made up 30% of all validated voters. Among these voters, far more (55%) said they voted for Clinton than for Trump (38%). Amount the much larger group of white voters who had not completed college (44% of all voters), Trump won by more than two-to-one (64% to 28%) [12, 2020]. But many pollsters did not adjust their samples - a process known as weighting - to make they had the right number of well-educated or less educated respondents. Well-educated voters are much likelier to take surveys than less educated ones. About 45% of respondents in a typical national poll of adults will have a bachelor's degree or higher [Cohn, 2017]. According to the US education attainment census [Cohn, 2017], in 2019, high school was the highest level of education completed by 28.1% of the population age 25 and older and 22.5% finished four years of college. 36% of the population age 25 and older with a bachelor's degree or higher. In the survey data, education was grouped into "3rd Grade or less", "Completed some high school", "High school graduate", "Other post high school vocational training", "Completed some college, but no degree", "Associate Degree", "College Degree (such as B.A, B.S)", "Completed some graduate, but no degree", "Masters degree", and "Doctorate degree" 11 groups. We regrouped education into "High school or lower", "Completed some college, but no degree", and "Post Secondary or Higher" three groups to suit our model. The weights for these three groups are 28.94%, 20.15%, and 50.90% respectively. 50.90% is much higher than the census data 36%. The gap will be adjusted in the post-stratification procedure.

2.2 Post-Stratification Data

The post-stratification data was downloaded from IPUMS America Census Service (ACS). The ACS is a project of the U.S. Census Bureau that has replaced the decennial census as the key source of information about American population and housing characteristics. The IPUMS database contains samples from the 2000-2019 ACS [Ruggles et al., 2020]. We assumed the census data approximated the true distribution of the U.S. population and selected the most recent ACS 2018 data as the post-stratification data set.

The post-stratification has 2,543,459 records. To maintain unity with the survey data, we selected age, gender, race, hispanic, education as the target variables. We cleaned and regrouped these variables to align with the survey data. Figures 1 and 2 show the corresponding variable comparisons between the survey and the post-stratification data. Solid line and dot line represent the survey and the post-stratification data respectively. The differences in the following paragraphs are the absolute differences. For the age group comparisons, the weights for 50-65 groups are very close. The survey data over represents the 18-35 and 36-49 groups. The discrepancies are 3.2% and 7.9% respectively. On the other hand, the senior group (65+) the survey data is under represented. The difference is 8.2%. The gender comparisons do not show large differences. The differences are within 2%. For the Hispanic and non-Hispanic population comparison, the weights are almost the same. For the non-Hispanic population, the percentages for white are very close. The difference is 3.3%, the white population is under represented in the survey data. Black population is slightly over-represented in the survey data. The difference is 1.4%. The Asian weight in the survey is slightly less than the post-stratification data. The difference is 1.9%. For the population in other race category, the weight in the survey data is 3.8% higher than the weight in the post-stratification data. Education is the main adjustment in the post-stratification procedure. For the post secondary or higher education group, the weight is 12.1% over represented in the survey data. For the high school or lower group, the survey data

is 10.4% lower than the post-stratification data. The weights are close for the some post secondary group from the two data sets. These observations further proved that well-educated voters are much likelier to take surveys. According to surveys and 2016 election results, voters with four-year college degrees or more education have higher propensity to vote for Biden. The population who have not completed college has a higher percentage to support Trump. The last graph in Figure 2 illustrates the population weights by states. Overall, the curves present very similar patterns. In most states, the weights are very close. The noticeable gaps are in Florida, New York, and Texas. The survey data over represent the weights in Florida and New York by 1.5% and 2.8% respectively. In Texas, the survey data presents the population by 1.1% compared with the post-stratification data. In general, by comparing with the survey data and the post-stratification data, the largest gap is in education variables. Age ranks the second one. The other variables do not show large percentage discrepancies between the two data sets. The post-stratification method is applied to make the survey results more representative for the whole population.

We have used R (R Core Team [2019]), specifically Tidyverse (Wickham et al. [2019]) for data analysis

Data is from ACS (Ruggles et al. [2020]) and from Voter Study Group (Tausanovitch and Vavreck [2020]).

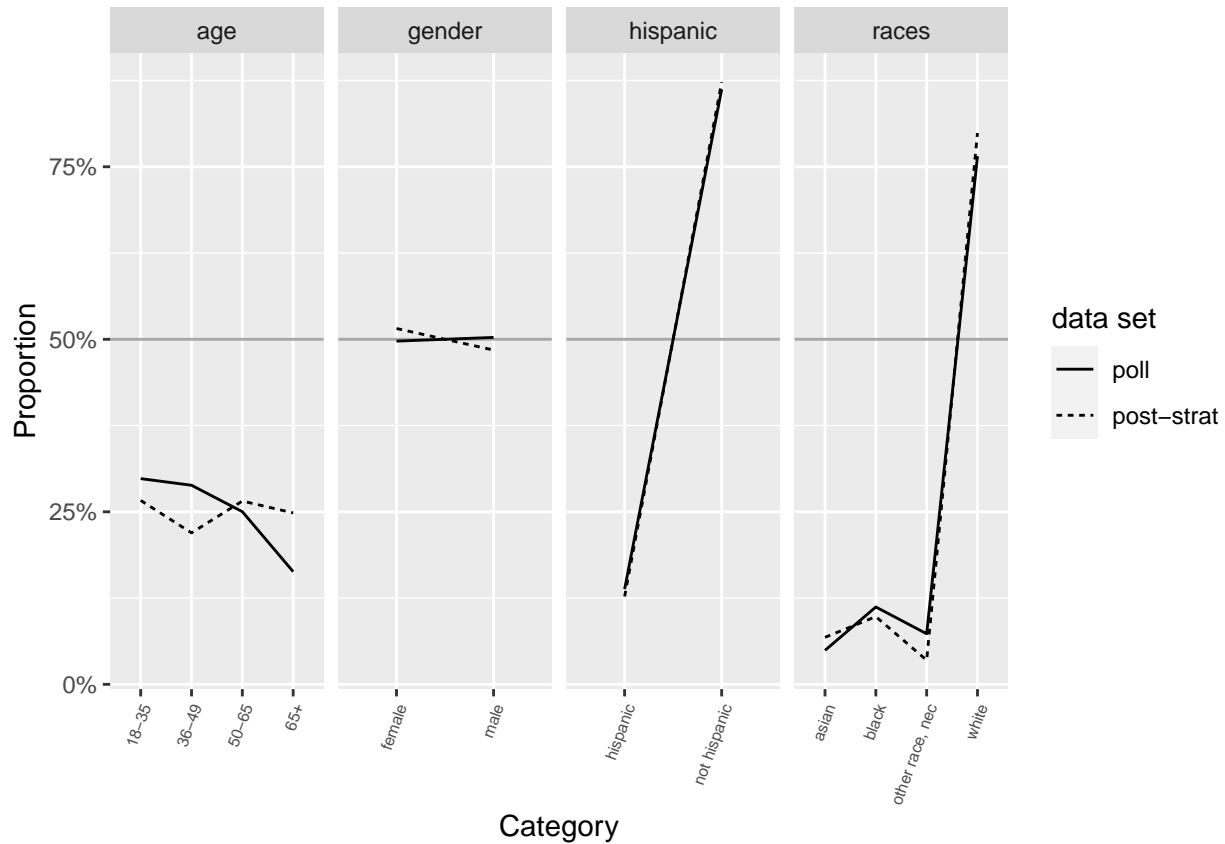


Figure 1: Demographics of Sample and Population

Figure 1 show us the voter demographics from the VSG data (Tausanovitch and Vavreck [2020]) vs the ACS data (Ruggles et al. [2020]).

Table 1: (#tab:polling props)Who decided voters plan to support (Polling Data)

Candidate	Number of Respondents	Proportion (%)
Donald Trump	2481	48
Joe Biden	2719	52

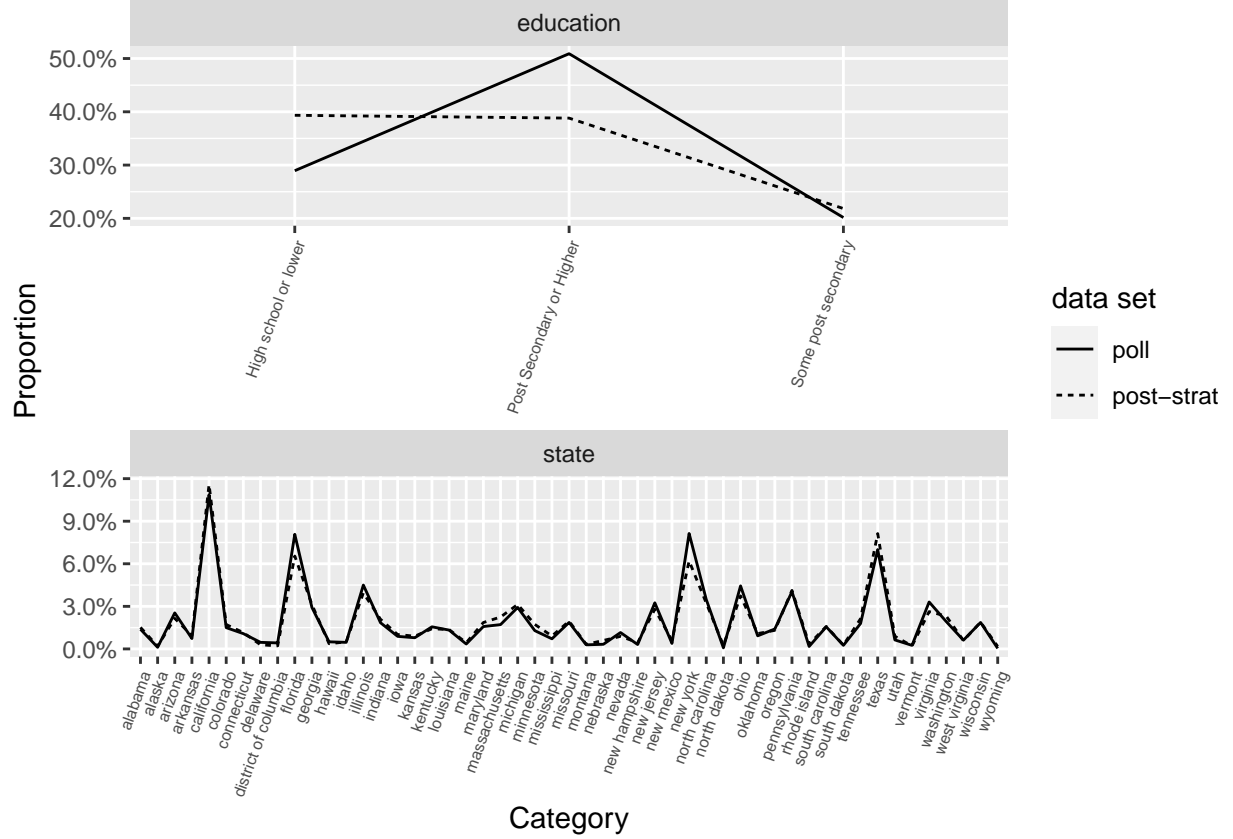


Figure 2: More Demographics of Sample and Population

Figure 2 shows us more of the voter demographics from the VSG data (Tausanovitch and Vavreck [2020]) vs the ACS data (Ruggles et al. [2020]).

Table @ref(tab:polling props) shows the proportion of decided voters who plan to vote for Donald Trump or Joe Biden. The data used to create this table is from the Voter Study Group (Tausanovitch and Vavreck [2020]).

3 Model

For our analysis, we plan to use multilevel regression with post-stratification. Multilevel regression with post-stratification (MRP) is a type of analysis where we fit a model using a smaller data set, in this case our polling data and then use the results of the model to apply it to a larger population.

The main steps for MRP are: Find the data set you want to use to create your model. For our scenario, we used the polling data from the Voter Study Group (Tausanovitch and Vavreck [2020]). Next, you must create a model using your smaller sample. We used logistic regression and the data used was the polling data. Our

equation takes the form of equation (1), as seen below. Once you have your model, you must apply it to your larger data set to give an idea of the population. For our report, we used the Census data from IPUMS (Ruggles et al. [2020]).

MRP is extremely useful not only because of how simple it is but also because it allows us to estimate preferences of a population using individual responses from surveys. Also, as Kennedy and Gelman discuss in their paper “Know your population and know your model: Using model-based regression and post-stratification to generalize findings beyond the observed sample”, MRP works best when you have a population and variables of interest and want to apply these variables using two data sets with different characteristics (Kennedy and Gelman). Also, in relation to surveys, since you only use one survey to create your model, you don’t encounter issues with having to have multiple surveys for each region/state and lets under-sampled populations to be represented through post-stratification.

MRP does encounter some weaknesses as well. Once again we can look to Kennedy and Gelman’s paper where they say that MRP is dependent on how accurate the model is (Kennedy and Gelman). If the generated model makes incorrect predictions or assumptions, your post-stratification will be incorrect and will hurt your results.

To predict whether or not a person plans to vote for Joe Biden or Donald Trump, we plan to build a logistic regression model using data from the Voter Study Group (Tausanovitch and Vavreck [2020]) and then post-stratify it using Census Data (Ruggles et al. [2020]). Since logistic regression only works for binary response variables, we created a variable called `vote_biden` which returns a 1 if the respondent plans to vote for Joe Biden and a 0 if they plan to vote for Donald Trump.

The logistic regression model takes the form of:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 x_{sex} + \beta_2 x_{agegroup} + \beta_3 x_{race} + \beta_4 x_{state} + \beta_5 x_{income} + \beta_6 x_{hispanic} \quad (1)$$

Once we have our regression model, we will use the `predict` function in R (R Core Team [2019]), to apply our model to the Census data (Ruggles et al. [2020]). We do this by grouping the Census data by the demographics we plan to analyze (sex, race, age, education level, hispanic or not, state), and then applying the model to each of those groups. After applying the model, we will receive probabilities that a person in that group will vote for Joe Biden. Once the predictions are complete, we can use them to find out who will win the popular vote or how many electoral colleges a candidate will win. We also can use a 95 percent confidence interval, which means that we are 95 percent certain that true value for the population (in this case popular vote percentage), is within the range we obtain. From this we can also conclude that our results will be accurate between +/- 4%.

In equation (1), each β represents a coefficient that the regression model will compute for us. As for our variables, we have chosen to use sex, age, race, income, state, and whether the respondent is hispanic. We decided to use the first 3 because they are generally strong predictors of which candidate a person would support, such as how some states tend to vote republican year after year while some states flip between democratic and republican almost every election. Next, we decided to choose income, because Joe Biden has made claims to increase taxes on the rich, which may influence their support for him. Lastly, we wanted to focus on whether the respondent was hispanic and if so, where they were from. This variable was important for our predictions because we know how poorly Donald Trump has spoken of hispanic people and we believe they could have a strong impact on the election.

The output of the logistic regression model will give us a probability of whether or not a voter plans to vote for Joe Biden or not. In order to find this probability, we take the sum of the right side of equation (1) and plug it into the equation below:

$$\frac{e^{sum}}{1 + e^{sum}} \quad (2)$$

Equation (2) is just a manipulation of equation (1), where e is the exponential equation and sum is the sum of the right side of equation (1). We see that as the sum of the right side increases, the probability that a person will vote for Joe Biden increases as well. We are running our regression model using the `glm()` function in R (R Core Team [2019]). The decision to run this model over other models like linear regression was made by the fact that we were predicting a binary variable about a voter's decision. Since there are only two possible options our data will likely follow an S shape and a straight line equation will not be helpful to model this relationship. Another strength present for logistic regression is that when combined with post-stratification it allows us to take information from under-represented populations and it allows their views to be accounted for more greatly. For example, our polling data (Tausanovitch and Vavreck [2020]), includes only 2 observations from Wyoming, but using multilevel regression with post-stratification, we can have that expanded to over 3000 people!

Our model does have some weaknesses, since the output must be binary, we cannot account for other candidates or a person deciding not to vote. This issue isn't too large because our main goal is to determine which of the two main candidates will be chosen by the people of America. Another weakness we do encounter with our model and multi level regression with post-stratification is that it has a strong dependence on the survey data. This is a weakness because if the survey has any gaps or there are any tweaks we need to make, it can change the course of results.

Table 2, shows the estimates for the coefficients that will fit into our logistic regression equation. These coefficients will fit into Equation (1), and were calculated using data from the Voter Study Group (Tausanovitch and Vavreck [2020]). The table is made using `kable` from `knitr` (Xie [2020]) and is formatted using `kableExtra` (Zhu [2020])

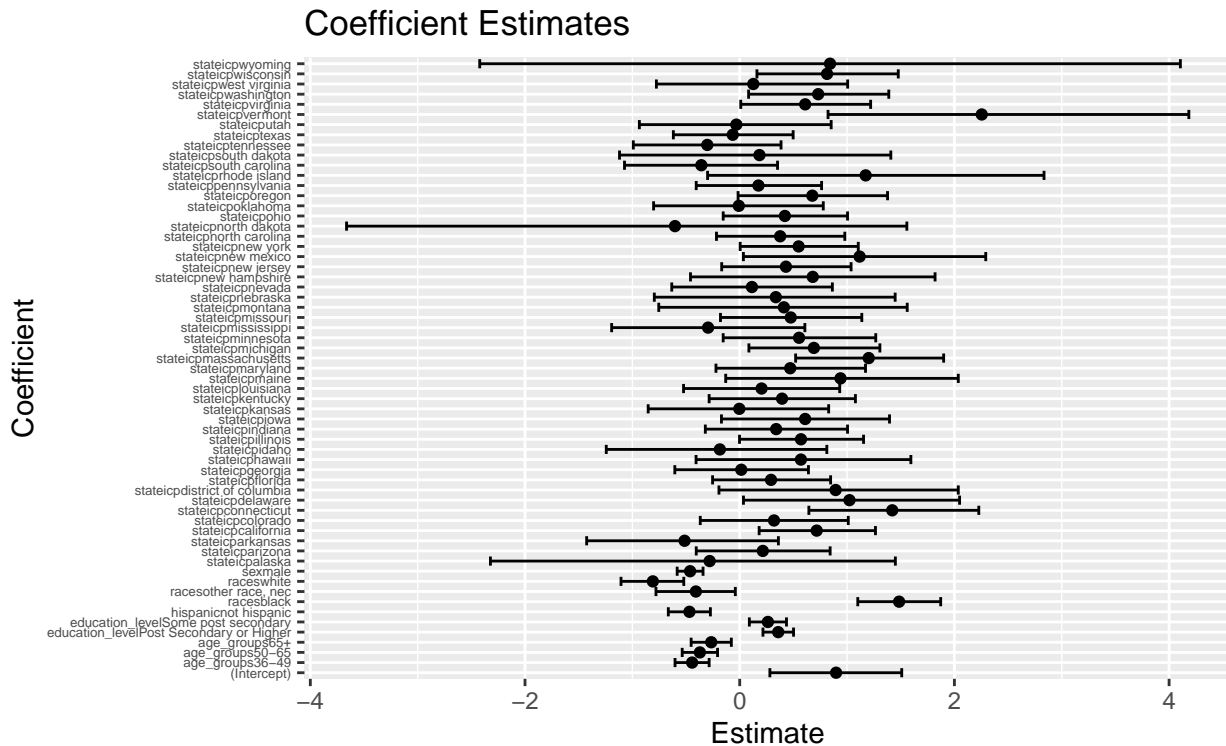


Figure 3: Coefficient Estimates

Figure 3 shows us the coefficients that would fit into equation (1) using the polling data (Tausanovitch and Vavreck [2020]). We also have error bars present, which show the upper and lower estimates for the coefficients. What we have to look out for in this scenario is that coefficients with negative values would mean that the person is more likely to vote for Donald Trump (with that characteristic) and positive values

Table 2: Coefficients from the Model

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.8980072	0.3126492	2.8722518	0.0040756	0.2811482	1.5092669
sexmale	-0.4619733	0.0614678	-7.5156968	0.0000000	-0.5825849	-0.3416100
age_groups36-49	-0.4433787	0.0810235	-5.4722267	0.0000000	-0.6023717	-0.2847185
age_groups50-65	-0.3709067	0.0839278	-4.4193534	0.0000099	-0.5355720	-0.2065317
age_groups65+	-0.2650972	0.0954808	-2.7764449	0.0054957	-0.4524197	-0.0780729
racessblack	1.4848649	0.1967429	7.5472352	0.0000000	1.1000293	1.8722214
racessother race, nec	-0.4082380	0.1884919	-2.1658125	0.0303255	-0.7800952	-0.0406482
racesswhite	-0.8091757	0.1489996	-5.4307243	0.0000001	-1.1057163	-0.5208546
stateicpalaska	-0.2797864	0.9132895	-0.3063502	0.7593381	-2.3221840	1.4499433
stateicparizona	0.2150895	0.3175737	0.6772900	0.4982220	-0.4050470	0.8423335
stateicparkansas	-0.5136530	0.4533527	-1.1330098	0.2572102	-1.4256850	0.3603696
stateicpcalifornia	0.7171878	0.2756199	2.6020894	0.0092658	0.1809687	1.2646193
stateicpcolorado	0.3196581	0.3509959	0.9107174	0.3624443	-0.3674600	1.0111726
stateicpconnecticut	1.4206903	0.4024846	3.5298003	0.0004159	0.6444034	2.2272548
stateicpdelaware	1.0223226	0.5096224	2.0060394	0.0448520	0.0349976	2.0485310
stateicpdistrict of columbia	0.8934468	0.5630684	1.5867464	0.1125701	-0.1932745	2.0361631
stateicpflorida	0.2917968	0.2796465	1.0434486	0.2967406	-0.2525567	0.8467523
stateicpgeorgia	0.0145529	0.3169457	0.0459162	0.9633770	-0.6044370	0.6403815
stateicphawaii	0.5701168	0.5070301	1.1244241	0.2608331	-0.4067535	1.5945679
stateicpidaho	-0.1849658	0.5193924	-0.3561196	0.7217510	-1.2431768	0.8115242
stateicpillinois	0.5709498	0.2940343	1.9417793	0.0521638	-0.0017038	1.1536467
stateicpindiana	0.3397257	0.3373038	1.0071802	0.3138482	-0.3203070	1.0044955
stateicpiowa	0.6094353	0.3982495	1.5302852	0.1259462	-0.1700387	1.3956818
stateicpkansas	-0.0036380	0.4273321	-0.0085133	0.9932075	-0.8526640	0.8287237
stateicpkentucky	0.3944897	0.3468740	1.1372710	0.2554250	-0.2847029	1.0777918
stateicplouisiana	0.2037260	0.3703378	0.5501086	0.5822449	-0.5232388	0.9314653
stateicpmaine	0.9391237	0.5468677	1.7172777	0.0859284	-0.1304090	2.0362073
stateicpmaryland	0.4704849	0.3547184	1.3263617	0.1847199	-0.2211046	1.1719732
stateicpmassachusetts	1.2019368	0.3510437	3.4238951	0.0006173	0.5210273	1.8998570
stateicpmichigan	0.6908228	0.3106500	2.2237975	0.0261621	0.0855832	1.3058171
stateicpminnesota	0.5531700	0.3618500	1.5287273	0.1263321	-0.1539583	1.2674587
stateicpmississippi	-0.2945026	0.4578634	-0.6432107	0.5200874	-1.1928848	0.6069939
stateicpmissouri	0.4757005	0.3352768	1.4188291	0.1559488	-0.1792418	1.1375350
stateicpmontana	0.4116022	0.5823474	0.7067984	0.4796918	-0.7540098	1.5605872
stateicpnebraska	0.3361727	0.5657196	0.5942391	0.5523522	-0.7953709	1.4480437
stateicpnevada	0.1138363	0.3808846	0.2988735	0.7650365	-0.6330333	0.8632322
stateicpnew hampshire	0.6805873	0.5742896	1.1850941	0.2359802	-0.4582967	1.8199862
stateicpnew jersey	0.4305098	0.3068634	1.4029365	0.1606358	-0.1677030	1.0377356
stateicpnew mexico	1.1169726	0.5688036	1.9637230	0.0495622	0.0341497	2.2920748
stateicpnew york	0.5489828	0.2798485	1.9617146	0.0497957	0.0043994	1.1044931
stateicpnorth carolina	0.3769283	0.3043149	1.2386125	0.2154890	-0.2163184	0.9791604
stateicpnorth dakota	-0.6028355	1.2029598	-0.5011269	0.6162818	-3.6624093	1.5571343
stateicpohio	0.4205613	0.2948223	1.4264906	0.1537268	-0.1539062	1.0045259
stateicpoklahoma	-0.0082254	0.4019938	-0.0204615	0.9836752	-0.8020673	0.7784033
stateicporegon	0.6762432	0.3544891	1.9076559	0.0564357	-0.0152077	1.3771460
stateicppennsylvania	0.1745544	0.2972761	0.5871793	0.5570833	-0.4051598	0.7628937
stateicprhode island	1.1732738	0.7752391	1.5134348	0.1301693	-0.2992149	2.8349608
stateicpsouth carolina	-0.3574239	0.3629406	-0.9848000	0.3247223	-1.0732722	0.3525165
stateicpsouth dakota	0.1840415	0.6330361	0.2907282	0.7712592	-1.1196390	1.4078076
stateicptennessee	-0.3020013	0.3503869	-0.8619082	0.3887380	-0.9912078	0.3850810
stateicptexas	-0.0649951	0.2838591	-0.2289698	0.8188924	-0.6179683	0.4977429
stateicputah	-0.0312335	0.4535913	-0.0688583	0.9451024	-0.9346965	0.8520835
stateicpvermont	2.2547412	0.8171449	2.7592918	0.0057927	0.8231091	4.1840250
stateicpvirginia	0.6100694	0.3079872	1.9808268	0.0476107	0.0099348	1.2197413
stateicpWASHINGTON	0.7311564	0.3325406	2.1986981	0.0278994	0.0828534	1.3888521
stateicpwest virginia	0.1268231	0.4520245	0.2805668	0.7790427	-0.7761741	1.0056192
stateicpwisconsin	0.8136348	0.3352005	2.4273077	0.0152114	0.1605671	1.4770183
stateicpwyoming	0.8415398	1.4394702	0.5846177	0.5588048	-2.4218041	4.1050804
education_levelPost Secondary or Higher	0.3587122	0.0726763	4.9357504	0.0000008	0.2165005	0.5014293
education_levelSome post secondary	0.2626752	0.0881652	2.9793539	0.0028886	0.0900105	0.4356705
hispanicnot hispanic	-0.4683979	0.0999464	-4.6864901	0.0000028	-0.6650514	-0.2731151

mean the person is more likely to vote for Joe Biden. Table 2 shows a numerical view of Figure 3, along with p-values.

Using the outputs of the logistic regression model, we can get an equation that follows the form of (1), but with the β values filled out. This equation is difficult to write out because of the many variables, but in short, if the person's characteristic fits in with a certain variable, it is used in the equation. Then the equation is summed up and the probability is found using equation (2).

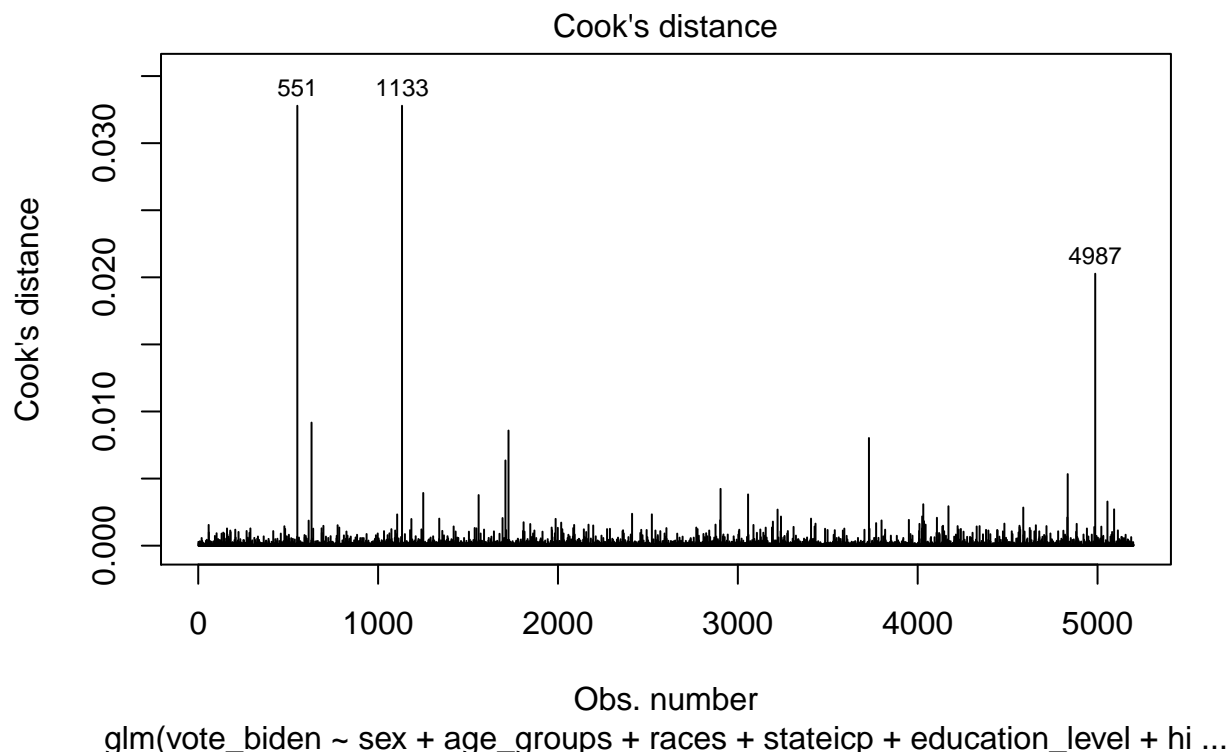


Figure 4: Cook's Distance Plot for Model

Figure 4 shows the cooks distance for observations in our polling data set. Cook's distance is useful for checking if a model is working correctly because it tells us how far a point is from the predicted value of it, telling us which points negatively impact our model. In our scenario, the distance between values would be the true value using our `glm` model vs what the `predict.glm` function returned. Figure 4 takes all observations of our polling data (Tausanovitch and Vavreck [2020]), and calculates the Cook's distance for it, showing which points can hurt the results of our model. We find that out of the 5200 observations, there aren't too many points that deviate from what we predict. This makes sense for our model because in the real world, there are some people who will support Donald Trump or Joe Biden, even if they don't "fit" the usual voter demographic for the candidate. Since the number of observations with large Cook's Distances are low, we can conclude that our model is fairly strong.

4 Results

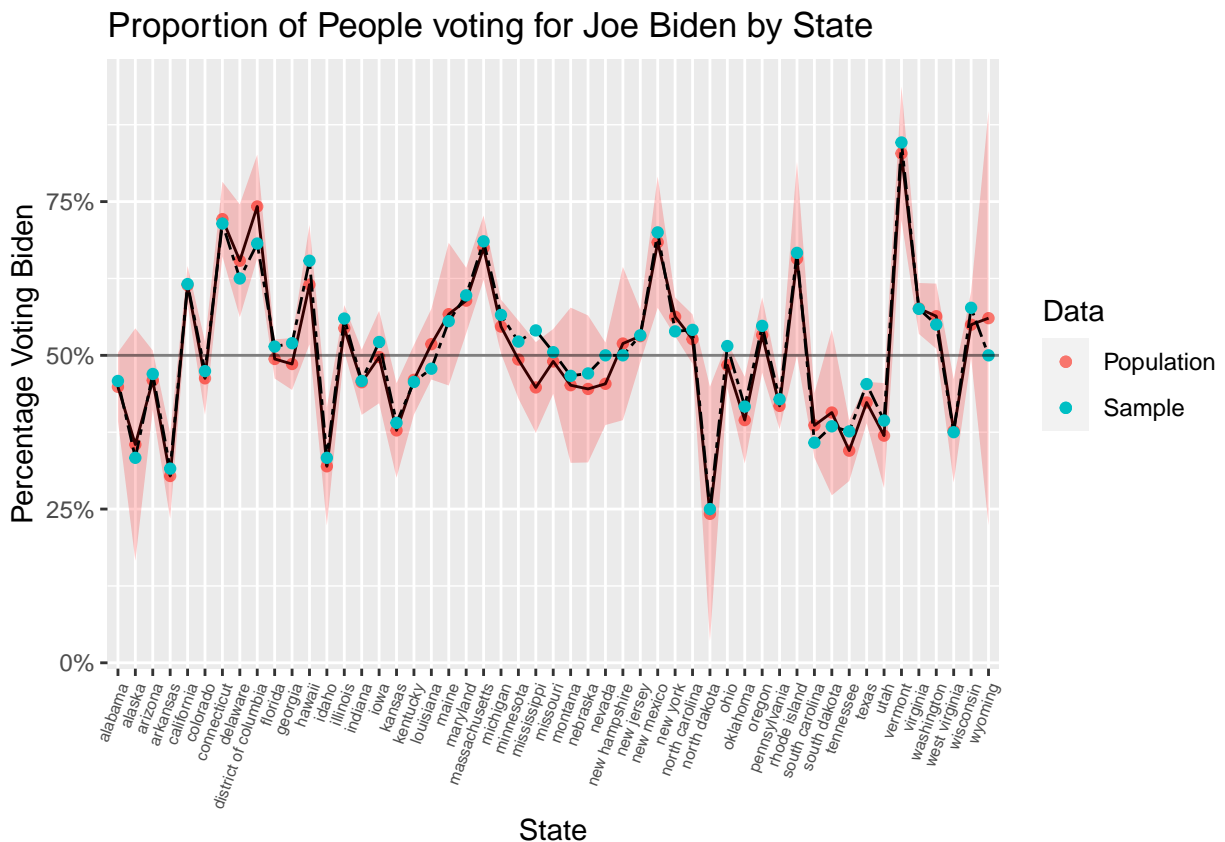


Figure 5: Proportion of each State Voting for Biden

Figure 5 shows us the proportion of respondents voting for Joe Biden broken down by state, along with the predicted proportions using MRP. We have also included the errors for each predicted value as well. We can see that the predicted proportions are fairly close to what the polling data showed. But, we also have to acknowledge the errors present, we can see that many states' errors cross over the 50% line which could be pivotal for each candidate because of the winner take all nature of the electoral college. We can see that some states like Florida and Georgia have small errors but there are some states who have very large ones like North Dakota and Wyoming.

Table 3: Joe Biden Voting Result Estimates

	Lower Estimate	Mean Estimate	Upper Estimate
Number of Colleges	191.0	260	423.0
Proportion of Vote (%)	46.1	51	55.9

Map of the USA and which states plan to support Joe Biden or Donald Trump

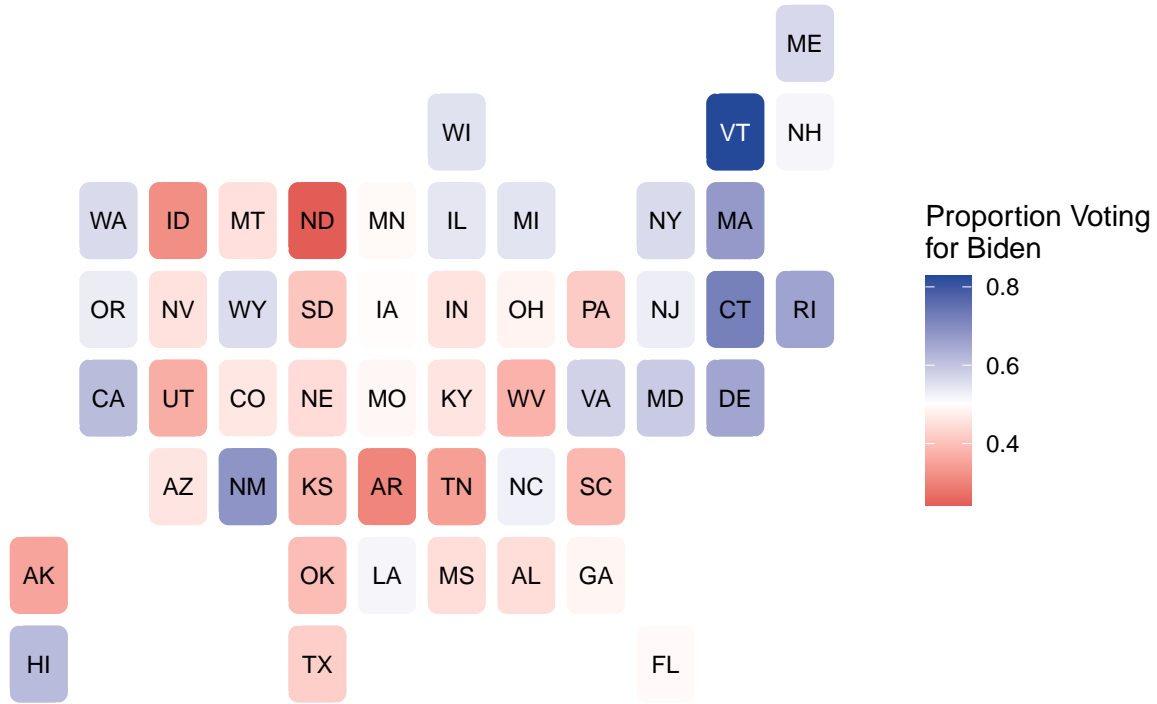


Figure 6: Proportion of Voters from Each State Voting Joe Biden

Figure 6 is an amazing view of all the states and which candidate they are leaning towards voting for using `statebins` (Rudis [2020]). We see that states like Vermont, Connecticut and California are some of the more “blue” states, meaning they plan to vote for Joe Biden, while North Dakota, Arkansas and Idaho are the “red” states, planning to vote for Donald Trump. The states that are more white in colour can be the most important for the race when it comes to deciding an actual winner through the electoral college. We see that Florida, Louisiana and New Hampshire are some of the more undecided states, and a switch in these states, can influence the election greatly.

Table 3 shows the lower, mean and upper predictions for the results of the election for Joe Biden. We see that on the lower end, Biden can expect to get 46.1% of the popular vote while only getting 191 electoral colleges. We also see that our middle estimate says Biden will get 51% of the popular vote while still losing the election by getting only 260 colleges. Lastly, on the upper estimate for Biden’s results, he can get 55.9% of the popular vote while getting 423 colleges! This truly shows how close this election is, as we can see in Figure 5, many states are hovering around the 50% mark, which shows the colleges can go either way. The number of electoral colleges by state were found on the Britannica Encyclopedia’s website (of Encyclopaedia Britannica).

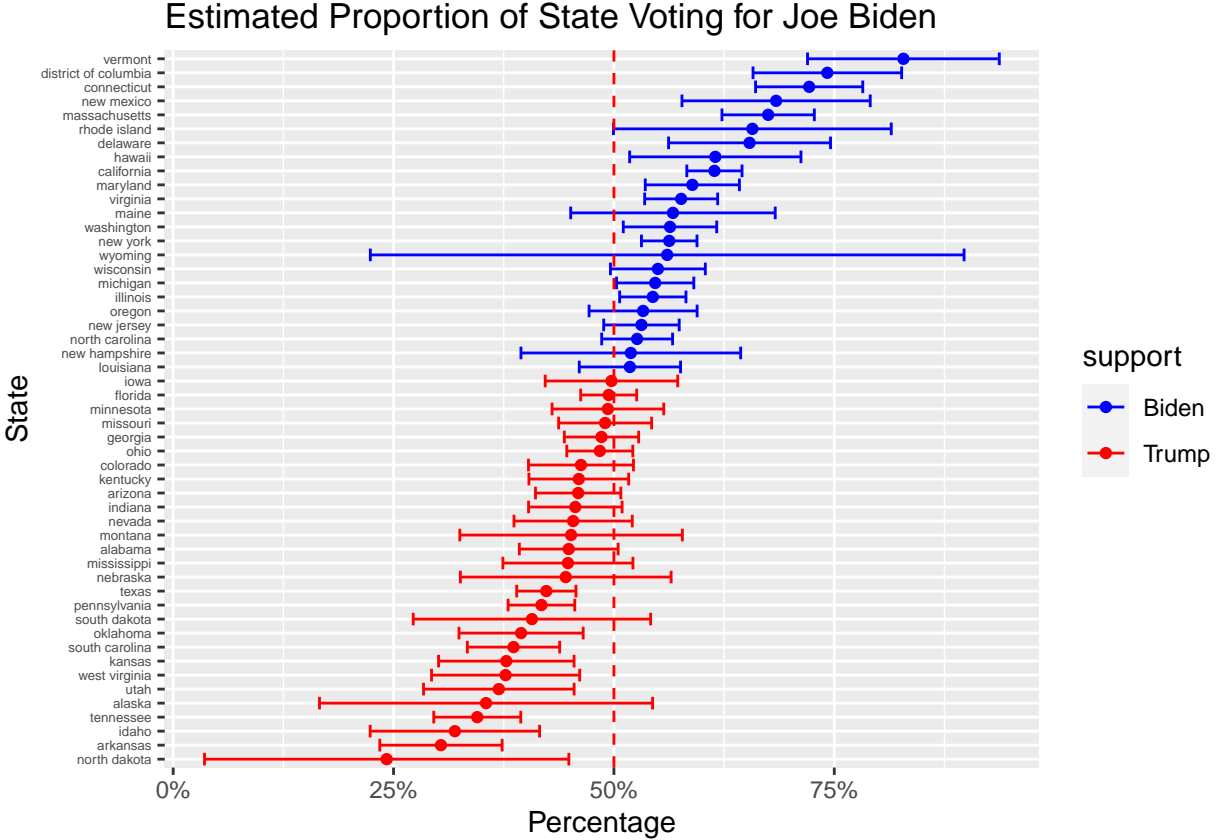


Figure 7: Proportion of Biden votes by state

Figure 7 is a different view of Figure 5, this time only focusing on the predictions for each state with the errors included. We see that many states have error bars overlapping the 50% line, showing that many states are a toss up, given the nature of the electoral college.

5 Discussion

As seen in table 1, the mean estimate is that Joe Biden will win the popular vote with 51% of the vote. This means that we reject the null hypothesis. However, the American election operates according to the electoral college and as such, it is possible that a candidate wins the popular vote but loses the electoral college. This has happened on two occasions: in 2000 when Al Gore lost to George W. Bush and in 2016 when Donald Trump won over Hillary Clinton (Blum [2020]). According to our results, Biden would win the popular vote but lose the electoral college vote with a mean estimate of 260 points (270 is the threshold for victory). The abolishment of the electoral college has been considered, however, professor Opal from McGill university notes that it would require a constitutional amendment to pass this unlikely change (Blum [2020]).

Through Figure 5, we are able to visualize numerous national trends of voting behavior. First, it is observable that Joe Biden and the Democrats have a strong hold of the west coast states. However, this was not always the case. Up until 1988, the Republican consistently had the support of California voters, who make up 55 electoral college points (Krishnakumar et al. [2016]). However, California had a massive spike in Hispanic and Asian immigrants during the 1990's (Krishnakumar et al. [2016]). This growing voter base was put off by the anti-immigration stance of the Republican party; Particularly in 1994, when the Republican's attempted to pass a law that would deny public services to illegal immigrants (Krishnakumar et al. [2016]). As a result

of information, it comes as no surprise that Biden has a strong lead in California in our data and collects the 55 electoral college points.

Secondly, based on Figure 5, Republican party leader Donald Trump performs very well in the American mid-west. This would follow-up Trump’s 2016 victory in the historically democratic states of Iowa, Ohio, Michigan and Wisconsin (Lauck [2017]). Many of these states had not voted Republican since the 80’s and were considered pillars of the Democratic voter base (Lauck [2017]). However, blue-collar workers in these states have flipped their vote over their dissatisfaction of the economy and manufacturing sector under Democratic leadership (Wilson [2017]). Considering the importance of an improving economy to mid-west voters, the economic growth under the Trump administration should lead to another set of victories in this part of America (News [2020]). This information would explain the results in our data for these mid-west states.

Thirdly, as seen in Figure 5, Joe Biden seems to possess a strong lead in north-eastern coastal states. This geographic area of America has historically, and consistently, been comprised of democratic-leaning states. This is the case because Democrats tend to gather a large percentage of the votes in big cities (Thompson [2019]). The north-eastern coast includes many of these populous cities such as New York, Boston, Washington D.C. and Newark. As these cities have become more ethnically diverse, less manufacturing oriented and more educated over the years, the voter base has shifted towards the Democratic party (Thompson [2019]). Like California on the west coast, these big east coast cities are focal points for incoming immigrants who tend to vote for the democratic party (Thompson [2019]). As a result, the results from Figure 4 and 5 for north-eastern coastal states fall in line with prior voting behavior and tendencies.

In Figure 5, we see that Hispanic voters are more likely to vote for, democratic nominee, Joe Biden. As previously mentioned, the Republican’s willingness to pass Proposition 187 to restrict public services to illegal immigrants was poorly received by the Hispanic community, in particular (Krishnakumar et al. [2016]). However, modern circumstances surrounding President Trump have also played a role in the aversion of the Hispanic community towards the Republican party. In the 2019 paper by Gutierrez et al., Gutierrez cites that Mexican Americans and immigrants felt threatened and targeted by Trump’s proposals and rhetoric (Gutierrez et al. [2019]). Despite comments specifically targeting illegal immigrants from Mexico, a focus group in Florida composed of Puerto Rican voters noted that they feel Trump is villainizing the entire Latino population (Gutierrez et al. [2019]). This could explain why Hispanic voters lean towards voting for Biden based on results seen in Figure 5. Gutierrez also follows-up by suggesting that the anger towards the 2016 election result, that saw Trump become president, could mobilize the Hispanic community to vote in greater numbers during the 2020 election (Gutierrez et al. [2019]).

Based on our results, white voters have a strong propensity to vote for Trump in the 2020 election (Figure 3). This is important as white voters make up over 75% of the sample population and the post-stratification data (Figure 1). In the paper by Major et al., the researchers suggest that white voters are strongly supporting the Trump campaign as a result of the changing demographics in the United States (Major et al. [2016]). As forecasted by the U.S. Census Bureau, the proportion of non-white individuals in the United States will be considerably higher than white individuals by the year 2060 (us2 [2012]). When reminded of this projection, Major et al. reports that white voters with high ethnic identification became more concerned with their declining influence and have begun supporting Trump’s anti-immigration policies, as a result (Major et al. [2016]). In contrast, our results show that black voters have a high propensity to vote for Biden (Figure 3). As suggested by researchers, this support is likely due to their resentment towards Trump. Trump’s opposition to the Black Lives Matter group and failure to condemn racism at key moments is viewed as threatening among African American voters (Towler and Parker [2018]). Black voters may represent a smaller percentage of the vote than white individuals, however, their support for Joe Biden has the potential to impact the election on a grand scale.

Based on the polling data in Figure 1, male and female voting is nearly equal at 50%. However, our results suggest that men are more likely to vote for Republican nominee Trump, while women are more likely to vote for Democratic nominee Joe Biden (Figure 3). It is important to note that in the 2016 election, a significantly greater number of women voted in the election than men (pew [2018]). Furthermore, 54% of women voted for the Hillary Clinton of the Democratic party (pew [2018]). This disposition to vote for the Democratic

party has been staple of American women for the past 40 years. When president Reagan was nominated in 1980, the Republican party adopted right-leaning viewpoints on several issues relevant to women (Thompson [2020]). In particular, the Republican party began opposing abortion rights for women and ceased supporting the Equal Rights Amendment (Thompson [2020]). Even though Reagan won that election, the decline of female voting support of the Republican party was commencing. This trend could serve as an explanation for the female support of Joe Biden in our results.

Contrary to classic convention, our results indicate that older individuals are progressively more likely to vote for Democratic nominee Joe Biden (Figure 3). One possible explanation for this change in voter behavior could be the Covid-19 outbreak that has plagued the entire world this year. As per the Centers for Disease Control and Protection, Covid-19 deaths are increasingly severe with age and disproportionately higher past the age of 55 (cdc [2020]). As a result, the Covid-19 outbreak is viewed as a critical issue in this year's election for the older percentage of voters (pew [2020]). It is possible that Trump's downplaying of the virus and refusal to wear a mask has upset the older voter population; culminating into Biden's surprising lead over older individuals in our research (Happe et al. [2020]). In addition, these voters may be impressed with Biden's plan to provide free testing to all Americans and financial support to those that lost their job due to the pandemic (Happe et al. [2020]).

Our results indicate that the higher education level completed, the more likely an individual will vote for Democratic nominee Joe Biden during the 2020 presidential election (Figure 3). A possible explanation for this general trend is that large, urban cities contain more universities and educational complexes compared to their rural counterparts. As a result, cities are more accessible for their inhabitants to complete higher levels of education. It is well understood that these large, urban cities tend to vote for the Democratic party while rural areas lean towards the Republican party (Thompson [2019]). As such, Republican-leaning states in the rural, mid-west will naturally be inhabited by less educated individuals than the urban individuals living in Democrat-voting coastal states.

We are very confident in our results for states that have coefficient estimates over 1.000 such as Massachusetts, Connecticut and Vermont to vote for Biden over Trump (Figure 3). In addition, we are also very confident in states that have coefficient estimates under 0.0 such as Idaho and Arkansas to vote for Trump over Biden (Figure 3). In contrast, we are not confident in the states that fall between these coefficient estimates as their polling results are incredibly close and various factors may tip the scales towards Republicans or Democrats when the election actually takes place. We are also not confident in states that have a standard error over 0.75 as there is a greater chance of inaccuracy for these states (Figure 3). In particular, we are not confident in the results of Alaska, Rhode Island, Wyoming and North Dakota (Figure 3).

Considering the concurrences between our data results and the historical information we just outlined, we believe that Joe Biden will win the popular vote but ultimately lose the election through the electoral college. We believe Biden will gain the support of big city residents to win the popular vote but fail to make up ground in the rural, blue-collar mid-west that Trump had won in 2016. However, it is very possible that our prediction doesn't hold due to the incredibly close results in certain states, such as Florida, Georgia and others (**Figure 4**). In addition, considering the weaknesses and limitations of this study (Outlined in the next section), the validity of the data used in this research is subject to debate.

Lastly, we can take a look at the state of Ohio. Ohio has been one of the best predictors for who will win the presidential election. Ballotpedia calls the state a "bellwether" because since 1900, Ohio has voted for the winning candidate 93.33% of the time (Ballotpedia). This is a remarkable number and as seen in 6, Ohio appears to be more red than blue which may give us evidence that although Biden won the popular vote, Trump may scrape out a victory.

5.1 Weaknesses and Next Steps

As seen in the 2016 election, the polls leading up to the presidential election can be inaccurate. Typically, this has not been an issue for the American polling companies. However, since becoming a presidential candidate, Donald Trump has become a polarizing figure in the media and among voters. As a result, it is very possible that some Trump voters are unwilling to show their support for the Republican candidate.

In addition, 2020 has been plagued by the COVID-19 pandemic and has changed the landscape of the presidential election. Not only has it become a key concern for voters when deciding their next president, but the virus has also changed the campaigning procedures and influencing voting mediums. For example, democratic representative Joe Biden has been holding parking lot rallies where supporters are social distancing within their cars while showing their support. Likewise, Trump rallies have seen a lower number of supporters than his previous campaign trails. However, most importantly, voters are changing their approach for voting in relation to safety precautions regarding the pandemic. 91.6 million votes have already been cast through the mail to avoid coming in contact with others at the polling booths. As a result, this election is far from usual or predictable. Various aspects of the election are critically affected by the pandemic virus and in turn, the polling results from ACS should not be considered reliable or representative of the population as they may have been in previous years.

A limitation of the survey dataset we are using is that it is from June 2020. As a result, the data is missing four months where voters may have changed their opinions and political allegiances. Healthcare, in particular, is an issue of very high importance for Americans during this election (pew [2020]). As of September 2020, four Covid-19 vaccines have entered phase 3 of clinical trials and a vaccine may arrive by early 2021 (nih [2020]). This kind of impactful development, among others, may influence voter behavior and alter the dataset if collected now.

For future studies, we would suggest allocating more of the polling company budget towards examining the education level of individuals in order to get a better estimate of the results and gain a clearer picture of what the American population is leaning towards before the election. As seen in Figure 3, the estimate for this variable is not particularly strong in favor, nor against, voting for Joe Biden. In addition, there is little separating the estimates for the various categories of education level in our research.

6 Code

Code supporting this analysis can be found at: https://github.com/matthewwankiewicz/US_election_forecast

References

- U.s. census bureau projections show a slower growing, older, more diverse nation a half century from now, Dec 2012. URL <https://www-census.gov.myaccess.library.utoronto.ca/newsroom/releases/archives/population/cb12-243.html>.
- An examination of the 2016 electorate, based on validated voters, Aug 2018. URL <https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters/>.
- Oct 2020. URL <https://www.pewresearch.org/politics/2020/08/13/important-issues-in-the-2020-election/>.
- 2020 presidential gender gap poll tracker, Sep 2020. URL <https://cawp.rutgers.edu/presidential-poll-tracking-2020>.
- Asian american and pacific islander heritage month: May 2020, Apr 2020. URL <https://www.census.gov/newsroom/facts-for-features/2020/aian.html>.
- Covid-19 provisional counts - weekly updates by select demographic and geographic characteristics, Oct 2020. URL https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm.
- Fourth large-scale covid-19 vaccine trial begins in the united states, Sep 2020. URL <https://www.nih.gov/news-events/news-releases/fourth-large-scale-covid-19-vaccine-trial-begins-united-states>.
- Oct 2020. URL <https://www.pewresearch.org/politics/2020/08/13/important-issues-in-the-2020-election/>.
- Ballotpedia. Presidential voting trends in ohio. URL https://ballotpedia.org/Presidential_voting_trends_in_Ohio.

- Benjamin Blum. Understanding the electoral college and how the u.s. president is really elected, Oct 2020. URL <https://www.cbc.ca/news/world/electoral-college-explainer-1.5768507>.
- Abby Budiman. Asian americans are the fastest-growing racial or ethnic group in the u.s. electorate, Jul 2020. URL <https://www.pewresearch.org/fact-tank/2020/05/07/asian-americans-are-the-fastest-growing-racial-or-ethnic-group-in-the-u-s-electorate/>.
- Nate Cohn. A 2016 review: Why key state polls were wrong about trump, May 2017. URL <https://www.nytimes.com/2017/05/31/upshot/a-2016-review-why-key-state-polls-were-wrong-about-trump.html>.
- Luis de Leon. 2020 election: The impact of the latino vote in texas, Oct 2020. URL <https://www.kvue.com/article/news/politics/vote-texas/election-2020-texas-latino-vote-impact/269-e27ced91-8f67-4bcf-bcc9-f0f56c5bc67f>.
- Erin Duffin. U.s. hispanic population, by state 2019, Sep 2020. URL <https://www.statista.com/statistics/259850/hispanic-population-of-the-us-by-state/>.
- William H. Frey. Older voters may secure a biden victory in 2020's swing states, Oct 2020. URL <https://www.brookings.edu/blog/the-avenue/2020/10/28/older-voters-may-secure-a-biden-victory-in-2020s-swing-states/>.
- Angela Gutierrez, Angela X. Ocampo, Matt A. Barreto, and Gary Segura. Somos más: How racial threat and anger mobilized latino voters in the trump era. *Political Research Quarterly*, 72(4):960–975, 2019. doi: 10.1177/1065912919844327.
- Mackenzie Happe, Kate Sullivan, and Max Pepper. Where trump and biden stand on major policy issues, Oct 2020. URL <https://www.cnn.com/2020/09/29/politics/trump-vs-biden-2020-election-policies/index.html>.
- Rachel Janfaza. Data suggests 2020 could see a surge in young u.s. voters, Oct 2020. URL <https://www.ctvnews.ca/world/america-votes/data-suggests-2020-could-see-a-surge-in-young-u-s-voters-1.5159607>.
- Lauren Kennedy and Andrew Gelman. Know your population and know your model: Using model-based regression and post-stratification to generalize findings beyond the observed sample. URL <https://arxiv.org/pdf/1906.11323.pdf>.
- Priya Krishnakumar, Maloy Moore, and Armand Emamdjomeh. After decades of republican victories, here's how california became a blue state again, Dec 2016. URL <https://www.latimes.com/projects/la-pol-ca-california-voting-history/>.
- Jon K Lauck. Trump and the midwest: The 2016 presidential election and the avenues of midwestern historiography. *Studies in Midwest History*, 03(01), Jan 2017.
- Mark Hugo Lopez, Jens Manuel Krogstad, and Jeffrey S. Passel. Who is hispanic?, Sep 2020. URL <https://www.pewresearch.org/fact-tank/2020/09/15/who-is-hispanic/>.
- Brenda Major, Alison Blodorn, and Gregory Major Blascovich. The threat of increasing diversity: Why many white americans support trump in the 2016 presidential election. *Group Processes amp; Intergroup Relations*, 21(6):931–940, 2016. doi: 10.1177/1368430216677304.
- BBC News. Us 2020 election: The economy under trump in six charts, Sep 2020. URL <https://www.bbc.com/news/world-45827430>.
- Luis Noe-Bustamante, Mark Hugo Lopez, and Jens Manuel Krogstad. U.s. hispanic population surpassed 60 million in 2019, but growth has slowed, Jul 2020. URL <https://www.pewresearch.org/fact-tank/2020/07/07/u-s-hispanic-population-surpassed-60-million-in-2019-but-growth-has-slowed/>.
- Editors of Encyclopaedia Britannica. United states electoral college votes by state. URL <https://www.britannica.com/topic/United-States-Electoral-College-Votes-by-State-1787124>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>.

- Bob Rudis. *statebins: Create United States Uniform Cartogram Heatmaps*, 2020. URL <https://CRAN.R-project.org/package=statebins>. R package version 1.4.0.
- Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. *IPUMS USA: Version 10.0 [dataset]*. Minneapolis, MN: IPUMS, 2020. URL <https://doi.org/10.18128/D010.V10.0>.
- Chris Tausanovitch and Lynn Vavreck. *Democracy Fund + UCLA Nationscape*. October 10-17, 2019 (version 20200814), 2020. URL <https://www.voterstudygroup.org/publication/nationscape-data-set>.
- Derek Thompson. How democrats conquered the city, Sep 2019. URL <https://www.theatlantic.com/ideas/archive/2019/09/brief-history-how-democrats-conquered-city/597955/>.
- Derek Thompson. Why men vote for republicans, and women vote for democrats, Feb 2020. URL <https://www.theatlantic.com/ideas/archive/2020/02/how-women-became-democratic-partisans/606274/>.
- Christopher C. Towler and Christopher S. Parker. Between anger and engagement: Donald trump and black america. *The Journal of Race, Ethnicity, and Politics*, 3(1):219–253, 2018. doi: 10.1017/rep.2017.38.
- Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- Reid Wilson. How the midwest slipped away from dems, Aug 2017. URL <https://thehill.com/homenews/state-watch/347414-how-the-midwest-slipped-away-from-dems>.
- Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2020. URL <https://yihui.org/knitr/>. R package version 1.30.
- Hao Zhu. *kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax*, 2020. URL <https://CRAN.R-project.org/package=kableExtra>. R package version 1.3.1.