The estimation of the US 2020 election: Donald Trump has a higher chance of winning

Jiayue Li, Yiyun Sun, Mengying Li, Yifei Li

Nov.2, 2020

Code and data supporting this analysis is available at: https://github.com/jlkuee/Sta304-problemset3

Model

We are interested in predicting the vote outcome of the 2020 American federal election (2020 United States presidential election, 2020, October 30). To do this we are constructing two logistic regression models based on the survey data and employing a post-stratification technique according to the census data. In the following subsections, we will discuss the model specifics, the post-stratification and the additional information.

Model Specifics

First of all, we use the variable vote_2020 from the raw survey data to create two new binary variables which are called vote_trump and vote_biden if "Donald Trump" shows in vote_2020, then vote_trump will be 1, otherwise, it will be 0. Similarly, if "Joe Biden" shows in vote_2020, then vote_biden will be 1, otherwise, it will be 0. We choose vote_trump / vote_biden as our response variable, and we will be using two logistic regression models to analyze the proportion of voters who will vote for Donald Trump or Joe Biden because we have a binary response variable. Then, we choose gender, race, household_income, education, state and age to be our explanatory variables. We select these variables because they will affect the results according to our research (FairVote.org. n.d.). Trump and Biden have different perspectives on race, education, household income, etc. People with various traits may have different preferences for voting president. As well, we also can find these variables in the census data for doing post-stratification in the next step. The only numerical variable is age, others are categorical variables, we don't use age as age-groups (categorical variable), for the reason that people with different ages may have various opinions, if we use age-groups instead, it will reduce the accuracy of outcomes. We use the "glm()" function to build the logistic regression models and RStudio to run our models.
The logistic regression model for Donald Trump we are using is:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 Male + \beta_2 RA_{ba} + \beta_3 RA_c + ... + \beta_8 RA_w + \beta_9 HI_1 + \beta_{10} HI_2 + ... + \beta_{i-1} HI_{23} + \beta_i ED_a +$$

$$\beta_{i+1} ED_{cd} + ... + \beta_{j-1} ED_o + \beta_j ST_{al} + \beta_{j+1} ST_{ar} + \beta_{j+2} ST_{az} + ... + \beta_n x_{age}$$

Where RA, HI, ED and ST refer to race, household_income, education and state, respectively. $log(\frac{p}{1-p})$ is a logit function, p represents the proportion of voters who will vote for Donald Trump. Similarly, β_0 represents the intercept of the model, and is the log odds of the probability of voting for Donald Trump at gender is Female, race is American Indian or Alaska Native, household_income is from \$100,000 to \$124,999, education is 3rd Grade or less, state is AK and age is 0. Additionally, β_1 represents the change in log odds, when the voter is a male. Male, RA, HI, ED and ST are all dummy variables which mean they are either equal to 0 or 1. For instance, β_2 represents the change in log odds, when the voter is Black or African American, RA_{ba} is 1, then other RA variables are equal to 0. Equivalently, β_9 represents the change in log odds, when household income of the voter is between \$125,000 to \$149,999, HI_1 is 1, then other HI variables are equal to 0. This is the same for other dummy variables. However, β_n represents the slope of the model, for every one-unit increase in age, we expect an β_n increase in the log odds of the probability of voting for Donald Trump. In terms of similarity, the logistic regression model for Joe Biden is the same as Donald Trump's except for the response variable.

Post-Stratification

In this part, we estimate the winning-election proportion of Donald Trump and Joe Biden, respectively, by using post-stratification analysis on American Community Surveys data. Poststratification analysis is a useful method to correct inevitable errors between survey data and census data. The first step of the post-stratification analysis is to divide the population into different cells. Specifically, we create our own cells by combining all predictor variables we have: gender, race, household_income, education, state and age because every variable here has significant influences on proportion variables and has corresponding variables in American Community Surveys data, so dividing in this way can give us the most accurate result. Thus we partition 3214539 observations into 1062953 cells. Whether it's predicting Trump's election rate or Biden's election rate, our predictions are all based on these cells. Next, we use the first logistic model we build to estimate the winning proportion of Trump and use the second one to estimate the winning proportion of Biden within each cell. Finally, we need to use a formula to weight each cell according to each cell's own population size. The weight-formula of post-stratification is $\hat{y}^{PS} = \frac{\sum_{j=1}^{j} N_j \hat{y}_j}{\sum_{j=1}^{j} N_j}$, where $\hat{y}_j$ is the estimate vote population in each cell, N_j represents the population size of the j^{th} cell and $\hat{y}^{PS}$ is the winning-election proportion of each candidate. Thereby, the weighting process is to sum the multiplication of each cell's population size and the estimated proportion value in each cell firstly and then divide the entire population size which is 2965169 in our analysis.

Addtional information

Besides the logistic model estimating the proportion of voters who will vote for Donald Trump, we also build another model to estimate the proportion of voters who will vote for Joe Biden. The goal of building two models is to compare the vote proportion of two candidates under the same condition because during the survey process, many participants did not choose anyone which caused the sum vote proportion of Trump and Biden is not equal to 1. If we filter all useless participants, the size of the sample will not be big enough, so we build two models to predict the proportion of voters respectively. The model for Joe Biden is really similar to the model for Donald Trump, where the response variable changes to vote_biden, and the predictor variables – gender, race, household_income, education, state and age all remain the same. In order to make the result fit to reality, we not only did the post-stratification analysis on the whole census data, but also on each state. We calculate the predicted winner-proportion of Trump and Biden within every state to see who gets the support of more states.

To make the analysis more accurate, we did a cleaning data step as well. For preparation of the survey data– Democracy Fund + UCLA Nationscape' Full Data Set, we construct two new variables: vote_trump and vote_biden, in order to build a logistic model to estimate the election-winning proportion of two candidates. Then we filter out the participant who didn't register to vote which is pointless to our analysis. Next, we filter out the participants whose highest level of education is "Completed some graduate, but no degree" since there is no match categorize in census data – American Community Surveys data. For the preparation of the American Community Surveys data, we firstly select the voters whose education level is not "n/a". Then we delete the voters whose household income is 9999999 to filter worthless leverages and select the voters whose household income is at least 0 to make the model significant. Next, we remove the voters whose age is less than 18, since only U.S. citizens over 18 are eligible to vote and contribute to our models. After these steps, we change the name of each variable and the name of their categories in the American Community Surveys data to make the data correspond to the survey data so that we can do post-stratification analysis. Finally, we remove all NA rows which are useless in both survey data and census data.

Results

Table 1: Summary of the Trump model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.210	1.098	0.191	0.849
as.factor(gender)Male	0.403	0.065	6.174	0.000
as.factor(race)Black, or African American	-2.134	0.340	-6.275	0.000
as.factor(race)Chinese	-1.328	0.504	-2.634	0.008
as.factor(race)Japanese	-0.668	0.712	-0.938	0.348
as.factor(race)Other asian or pacific islander	-0.411	0.358	-1.147	0.251
as.factor(race)Other race	-0.752	0.336	-2.236	0.025
as.factor(race)White	-0.018	0.308	-0.059	0.953
as.factor(household_income)$125,000 to $149,999	-0.046	0.165	-0.280	0.779
as.factor(household_income)$15,000 to $19,999	-0.647	0.184	-3.514	0.000
as.factor(household_income)$150,000 to $174,999	-0.059	0.198	-0.295	0.768
as.factor(household_income)$175,000 to $199,999	0.286	0.236	1.210	0.226
as.factor(household_income)$20,000 to $24,999	-0.335	0.183	-1.824	0.068
as.factor(household_income)$200,000 to $249,999	0.697	0.222	3.134	0.002
as.factor(household_income)$25,000 to $29,999	-0.344	0.179	-1.922	0.055
as.factor(household_income)$250,000 and above	0.158	0.229	0.691	0.489
as.factor(household_income)$30,000 to $34,999	-0.368	0.175	-2.100	0.036
as.factor(household_income)$35,000 to $39,999	-0.413	0.180	-2.293	0.022
as.factor(household_income)$40,000 to $44,999	-0.510	0.192	-2.660	0.008
as.factor(household_income)$45,000 to $49,999	-0.310	0.181	-1.714	0.087
as.factor(household_income)$50,000 to $54,999	-0.307	0.171	-1.797	0.072
as.factor(household_income)$55,000 to $59,999	-0.254	0.212	-1.201	0.230
as.factor(household_income)$60,000 to $64,999	-0.424	0.210	-2.022	0.043
as.factor(household_income)$65,000 to $69,999	-0.304	0.230	-1.323	0.186
as.factor(household_income)$70,000 to $74,999	-0.318	0.204	-1.563	0.118
as.factor(household_income)$75,000 to $79,999	-0.037	0.204	-0.182	0.856
as.factor(household_income)$80,000 to $84,999	-0.476	0.249	-1.914	0.056
as.factor(household_income)$85,000 to $89,999	-0.314	0.276	-1.136	0.256
as.factor(household_income)$90,000 to $94,999	-0.104	0.275	-0.379	0.705
as.factor(household_income)$95,000 to $99,999	-0.440	0.219	-2.009	0.045
as.factor(household_income)Less than $14,999	-0.702	0.151	-4.637	0.000
as.factor(education)Associate Degree	-0.505	0.790	-0.639	0.523
as.factor(education)College Degree (such as B.A., B.S.)	-0.471	0.785	-0.600	0.549
as.factor(education)Completed some college, but no degree	-0.395	0.785	-0.503	0.615
as.factor(education)Completed some high school	-0.070	0.791	-0.089	0.929
as.factor(education)Doctorate degree	-0.208	0.809	-0.257	0.797
as.factor(education)High school graduate	-0.082	0.785	-0.104	0.917
as.factor(education)Masters degree	-0.418	0.789	-0.529	0.597
as.factor(education)Middle School - Grades 4 - 8	0.047	1.091	0.043	0.966

	Estimate	Std. Error	t value	Pr(>|t|)
as.factor(state)AL	-0.072	0.777	-0.093	0.926
as.factor(state)AR	0.105	0.822	0.127	0.899
as.factor(state)AZ	-0.191	0.752	-0.254	0.799
as.factor(state)CA	-0.733	0.735	-0.997	0.319
as.factor(state)CO	-0.487	0.775	-0.628	0.530
as.factor(state)CT	-1.215	0.793	-1.532	0.126
as.factor(state)DC	-0.429	0.911	-0.471	0.638
as.factor(state)DE	-1.177	0.886	-1.327	0.184
as.factor(state)FL	-0.359	0.737	-0.488	0.626
as.factor(state)GA	0.068	0.754	0.090	0.928
as.factor(state)HI	-0.722	0.878	-0.822	0.411
as.factor(state)IA	-0.718	0.800	-0.898	0.369
as.factor(state)ID	0.282	0.869	0.325	0.745
as.factor(state)IL	-0.619	0.744	-0.832	0.406
as.factor(state)IN	-0.374	0.764	-0.489	0.625
as.factor(state)KS	0.247	0.814	0.303	0.762
as.factor(state)KY	-0.383	0.770	-0.498	0.619
as.factor(state)LA	-0.177	0.782	-0.227	0.821
as.factor(state)MA	-1.330	0.773	-1.721	0.085
as.factor(state)MD	-0.530	0.775	-0.685	0.494
as.factor(state)ME	-0.646	0.887	-0.729	0.466
as.factor(state)MI	-0.614	0.750	-0.819	0.413
as.factor(state)MN	-0.297	0.782	-0.380	0.704
as.factor(state)MO	-0.531	0.761	-0.699	0.485
as.factor(state)MS	-0.036	0.816	-0.044	0.965
as.factor(state)MT	-0.258	0.913	-0.282	0.778
as.factor(state)NC	-0.310	0.748	-0.415	0.678
as.factor(state)ND	0.282	1.466	0.192	0.848
as.factor(state)NE	-0.549	0.888	-0.618	0.537
as.factor(state)NH	-0.955	0.913	-1.046	0.296
as.factor(state)NJ	-0.515	0.749	-0.687	0.492
as.factor(state)NM	-1.119	0.911	-1.229	0.219
as.factor(state)NV	-0.114	0.787	-0.145	0.884
as.factor(state)NY	-0.524	0.737	-0.710	0.477
as.factor(state)OH	-0.521	0.744	-0.700	0.484
as.factor(state)OK	-0.335	0.792	-0.423	0.672
as.factor(state)OR	-0.690	0.769	-0.898	0.369
as.factor(state)PA	-0.406	0.743	-0.546	0.585
as.factor(state)RI	-1.482	1.103	-1.344	0.179
as.factor(state)SC	0.090	0.769	0.117	0.907
as.factor(state)SD	-0.034	0.960	-0.035	0.972
as.factor(state)TN	0.192	0.769	0.249	0.803
as.factor(state)TX	-0.041	0.738	-0.055	0.956
as.factor(state)UT	-0.561	0.814	-0.689	0.491
as.factor(state)VA	-0.617	0.750	-0.822	0.411
as.factor(state)VT	-2.465	1.291	-1.909	0.056
as.factor(state)WA	-0.774	0.760	-1.019	0.308
as.factor(state)WI	-0.964	0.766	-1.258	0.208
as.factor(state)WV	0.002	0.821	0.002	0.998
as.factor(state)WY	-1.216	1.456	-0.835	0.404
age	0.012	0.002	5.636	0.000

Table 2: Proportion of voting each candidates within each state

state	alp_predict_trump	alp_predict_biden	winner
AK	0.5896104	0.2124815	Donald Trump
AL	0.4962734	0.4040141	Donald Trump
AR	0.5690859	0.2917680	Donald Trump
AZ	0.5311757	0.3675304	Donald Trump
CA	0.3846767	0.4929810	Joe Biden
CO	0.4729903	0.4473317	Donald Trump
CT	0.3147153	0.5228987	Joe Biden
DC	0.3486986	0.6477961	Joe Biden
DE	0.2994037	0.5400885	Joe Biden
FL	0.4744017	0.4046218	Donald Trump
GA	0.5073322	0.3876282	Donald Trump
HI	0.3613008	0.5252351	Joe Biden
IA	0.4276588	0.4195480	Donald Trump
ID	0.6538786	0.2366863	Donald Trump
IL	0.4202152	0.4408204	Joe Biden
IN	0.4875651	0.3880858	Donald Trump
KS	0.6347300	0.2484836	Donald Trump
KY	0.4891701	0.4336005	Donald Trump
LA	0.4520322	0.4634037	Joe Biden
MA	0.2923471	0.5392662	Joe Biden
MD	0.3979273	0.5095175	Joe Biden
ME	0.4550130	0.4943586	Joe Biden
MI	0.4278402	0.4789704	Joe Biden
MN	0.5294521	0.4457679	Donald Trump
MO	0.4495088	0.3871672	Donald Trump
MS	0.4491337	0.4095446	Donald Trump
MT	0.5401387	0.3846083	Donald Trump
NC	0.4557148	0.4566354	Joe Biden
ND	0.6642051	0.0000040	Donald Trump
NE	0.4586065	0.4224955	Donald Trump
NH	0.3929673	0.5053561	Joe Biden
NJ	0.4532605	0.4209474	Donald Trump
NM	0.3214367	0.4864602	Joe Biden
NV	0.5214498	0.3342331	Donald Trump
NY	0.4254313	0.4764477	Joe Biden
OH	0.4499274	0.4083762	Donald Trump
OK	0.4866793	0.2805607	Donald Trump
OR	0.4261983	0.4478229	Joe Biden
PA	0.4940110	0.3561775	Donald Trump
RI	0.2635675	0.5314896	Joe Biden
SC	0.5281481	0.3153420	Donald Trump
SD	0.5905187	0.2449838	Donald Trump
TN	0.5904454	0.3310056	Donald Trump
TX	0.5425849	0.3463227	Donald Trump
UT	0.4540134	0.2692018	Donald Trump
VA	0.4101793	0.4914564	Joe Biden
VT	0.1267518	0.8927031	Joe Biden
WA	0.4018569	0.4772621	Joe Biden
WI	0.3753693	0.4509014	Joe Biden
WV	0.5935487	0.3341350	Donald Trump

state	alp_predict_trump	alp_predict_biden	winner
WY	0.3184163	0.3826301	Joe Biden

Table 3: Number of state voted for Trump or Biden

winner	n()
Donald Trump	29
Joe Biden	22

Table 4: Proportion of voting each candidate

alp_predict_trump	alp_predict_biden
0.4552354	0.4238066

Table 1 shows a summary of the Trump winning model. With significant P values (extremely small), some categories of variables have statistical significance in this model, which indicates a large possibility of validation of our alternate hypothesis (statistical difference in the response variable). That is, the following variables: gender, race, household_income, age in the model are statistically significant to predict the value of our response variable, trump_vote; while education and state variables are not statistically significant in this model. Table 5 in the Appendix presents the results of the Biden winning model, which has similar results to the Trump winning model. Except for the above-mentioned variables in the Trump model, the Biden model has a statistically significant variable on VT (Vermont) state. We present the noticeable details of the two models in the following based on the coefficient estimates. First, among gender categories, male participants have a positive relationship in the Trump model and a negative relationship in the Biden model. Second, three out of seven categories of race have statistical significance. The three races are Black (or African American), Chinese, and Other races. All of them have a negative relationship in the Trump model and a positive relationship in the Biden model. Third, categories of household income that fall in the amount of less than 45,000 per year have positive effects in the Biden model. Contrarily, the category of household income between 200,000 and 249,999 has a positive effect on the Trump model and a negative effect on the Biden model. Fourth, the age variable has a positive relationship in the Trump model and (statistically insignificant) negative coefficient in the Biden model. For every one-unit increase in age, we expect a 0.012 increase in the log odds of the probability of voting for Trump. Lastly, only one state has a statically significant value in both models, which is the VT state having a positive relationship in the Biden model.
Since the U.S. presidential election is based on the votes of each state to get the final result, we calculated the proportion of voting for Trump and the proportion of voting for Biden in 51 states, which was demonstrated in Table 2. As shown in Table 3, our estimated proportion of voting for Donald Trump is higher in 29 out of 51 states, whereas the estimated proportion of voting for Joe Biden is higher in 22 out of 51 states. Moreover, we have post-stratification analyses of the proportion of voting for either Trump or Biden for the entire sample, modeled by a logistic model, which accounted for the state, gender, age, race, education levels, and household income levels. Our post-stratification analyses estimate the proportion of voters in favour of Trump to be 0.455 while the proportion of voters in favour of Biden to be 0.424, presented in Table 4. While both candidates' proportion of voting for the entire U.S is less than 0.5 (half of the votes), we can see that Trump's proportion is higher than Biden's.

Discussion

Summary

To summarize what we did earlier, it could be composed of four main parts. The first part is we obtained the raw survey data from the Democracy Fund + UCLA Nationscape' Full Data Set' (Tausanovitch, Chris and Lynn Vavreck. 2020) and raw census data from the American Community Surveys (IPUMS USA, University of Minnesota). The second part is choosing the variables we wanted from these two data and cleaning the data, as mentioned in Additional Information. The third part is we constructed two logistic regression models based on the survey data by employing the "glm()" function to analyze the proportion of voters who will vote for Donald Trump or Joe Biden. Next, we used two logistic models to make a post-stratification analysis on the American Community Surveys data. We divided it into different cells according to all of our response variables. After we weighted the estimated proportion within each cell, we converted voters' intentions to election forecasts. In the last part, we interpreted the results by comparing two logistic regression models. We noticed that gender, race, household_income, and age in the models are statistically significant to predict American election outcomes. We also explained some significant coefficients in the results section. Moreover, we compared the $\hat{y}^{PS}$ in total, which means the probability of a candidate winning on election day. Donald Trump has a 0.455 approval rating, whereas Joe Biden has a 0.424 approval rating. We also compared the approval ratings in terms of states, the number of states which support Donald Trump is 29, but Joe Biden only has 22 states that support him. In conclusion, we predicted Donald Trump has a higher chance of winning on election day. In terms of bias, our prediction was determined if the candidate has a higher approval rating in all states and owns support from more states, then he will have a higher chance of winning on election day. However, according to our research in "How do the US presidential elections work?" (CBBC Newsround), the public vote in their states; each state has a certain number of delegates representing the public opinions on election day. If a candidate wins in a state, then he will get all state's delegates. In the end, the candidate who gets more delegates wins. This is different compared with what we did, so our prediction might not be accurate.

Conclusions

In conclusion, based on our models' results, we have the following findings: 1. a statistically significant and positive coefficient of male category in the Trump winning model indicates that a larger possibility of male voters is in favour of voting for Trump. 2. Similarly, voters in race groups include, Black or African American, Chinese, and Other races have a higher preference for Biden voting. 3. Voters in low household income levels (less than 45,000 per year) would be more likely to vote for Biden, while voters in the top household income level (200,000 - 249,999) have a preference for voting for Trump. 4. Voters' age has a positive relationship with the proportion of voting for Trump implying that a larger possibility of aging voters is in favor of voting for Trump. 5. Larger proportion of voters in Vermont state prefer voting for Biden.The estimated proportion of voters who prefer to vote for Trump being 0.455 is higher than the estimated proportion of voters favoring Biden, which is 0.424. Besides, the results of our estimated voting proportion for each state predict that Donald Trump will get a higher proportion of votes than Joe Biden in more states. Hence, we predict that Donald Trump will win the election. The results based on the post-stratification analyses are consistent with some of the findings of our models' results. We estimate that male voters would be more likely to vote for Trump while voters in minority race groups (including Black or African American, Chinese, and Other race) and voters in low household income levels (less than $45,000 per year) would be more likely to vote for Biden. Male voters are more inclined to vote for Trump, and males occupy an essential part of the total population. Hence, Trump has a large advantage over Biden in this factor. On the contrary, Whites, who occupy a large proportion of the population, do not have a statistically significant preference in our models, while the races that prefer to vote for Biden are not the majority in the United States population. The same situation applies to income levels. Therefore, our prediction of Trump winning the election is reasonable.

Weaknesses

(1)We only remain the variables that are persuasive and can match census data. However, there are still many other factors affecting the election results, such as variables in survey data: interest, ideo5 or some other variables related to coved-19. In the step of cleaning data, our group dropped "completed some graduate, but no degree" in the variable of education. Because this classification has no corresponding type in census data, it can not be classified into other categories. However, the absence of a degree does not mean that the voter does not have experience and knowledge in the relevant industry/field. Because the two data(survey & census) are classified differently, we have to do a great deal of cleaning before analyzing the data. If the variables of survey data are selected closer to the census data template, the error of the whole analysis will become smaller. (2) The number of respondents in the survey data set is insufficient. In this way, each value has a more significant impact on the overall trend. Therefore, the overall data capacity can be increased. At the same time, we use the data from June, five months away from now. There must be data updates and increases in this process, which will bring the result deviation. (3) The weighted average is not accurate enough for Post-regulation. In this step, we summarized the probability of Biden and Trump's data, respectively. After that, the weighted average for each group was calculated; that is, Biden / Trump's support rate. However, the problem ignored in this step is that the weighted average is extremely vulnerable to extreme values. A 10% approval rating may pull down the overall approval rating by several percentage points, but he represents only one vote. In other words, the weighted average does not represent the overall support rate very well. Such calculation has errors, although it seems reasonable.

Next Steps

After the real election result is announced, our group envisages that the following parts can be completed. First, combined with the actual results and find out the method to improve the data analysis accuracy. The updated method can also be used to analyze each subsequent US election/the same type of election or polling. Second, the scarcity of data in survey data has caused many inconveniences, which is very much the critical component in future attempts to recover the impact of the number of databases and increase the number of people under investigation. Third, a questionnaire sent to the general public can be established to make up for some deficiencies in the survey's calculation results.

Appendix

Table 5: Summary of the Biden model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.936	1.168	-0.801	0.423
as.factor(gender)Male	-0.332	0.063	-5.243	0.000
as.factor(race)Black, or African American	1.840	0.336	5.474	0.000
as.factor(race)Chinese	1.311	0.454	2.887	0.004
as.factor(race)Japanese	1.088	0.685	1.588	0.112
as.factor(race)Other asian or pacific islander	0.685	0.365	1.878	0.060
as.factor(race)Other race	0.917	0.343	2.670	0.008
as.factor(race)White	0.337	0.323	1.042	0.297
as.factor(household_income)$125,000 to $149,999	0.158	0.166	0.954	0.340
as.factor(household_income)$15,000 to $19,999	0.505	0.178	2.844	0.004
as.factor(household_income)$150,000 to $174,999	0.214	0.199	1.075	0.283
as.factor(household_income)$175,000 to $199,999	-0.182	0.244	-0.749	0.454
as.factor(household_income)$20,000 to $24,999	0.503	0.180	2.802	0.005
as.factor(household_income)$200,000 to $249,999	-0.570	0.233	-2.452	0.014

	Estimate	Std. Error	t value	Pr(>|t|)
as.factor(household_income)$25,000 to $29,999	0.352	0.176	1.999	0.046
as.factor(household_income)$250,000 and above	-0.020	0.229	-0.085	0.932
as.factor(household_income)$30,000 to $34,999	0.352	0.173	2.032	0.042
as.factor(household_income)$35,000 to $39,999	0.420	0.176	2.384	0.017
as.factor(household_income)$40,000 to $44,999	0.538	0.187	2.875	0.004
as.factor(household_income)$45,000 to $49,999	0.294	0.180	1.633	0.102
as.factor(household_income)$50,000 to $54,999	0.348	0.168	2.073	0.038
as.factor(household_income)$55,000 to $59,999	0.220	0.210	1.049	0.294
as.factor(household_income)$60,000 to $64,999	0.403	0.206	1.955	0.051
as.factor(household_income)$65,000 to $69,999	0.283	0.228	1.239	0.216
as.factor(household_income)$70,000 to $74,999	0.509	0.200	2.540	0.011
as.factor(household_income)$75,000 to $79,999	0.252	0.202	1.245	0.213
as.factor(household_income)$80,000 to $84,999	0.589	0.241	2.445	0.015
as.factor(household_income)$85,000 to $89,999	0.532	0.269	1.979	0.048
as.factor(household_income)$90,000 to $94,999	0.278	0.277	1.002	0.317
as.factor(household_income)$95,000 to $99,999	0.484	0.213	2.268	0.023
as.factor(household_income)Less than $14,999	0.407	0.146	2.794	0.005
as.factor(education)Associate Degree	-0.482	0.781	-0.617	0.537
as.factor(education)College Degree (such as B.A., B.S.)	-0.399	0.777	-0.513	0.608
as.factor(education)Completed some college, but no degree	-0.616	0.777	-0.793	0.428
as.factor(education)Completed some high school	-0.815	0.783	-1.041	0.298
as.factor(education)Doctorate degree	-0.668	0.802	-0.833	0.405
as.factor(education)High school graduate	-0.889	0.778	-1.143	0.253
as.factor(education)Masters degree	-0.339	0.781	-0.434	0.665
as.factor(education)Middle School - Grades 4 - 8	-0.782	1.062	-0.737	0.461
as.factor(state)AL	0.566	0.874	0.648	0.517
as.factor(state)AR	0.166	0.921	0.181	0.857
as.factor(state)AZ	0.636	0.854	0.745	0.456
as.factor(state)CA	1.057	0.838	1.262	0.207
as.factor(state)CO	0.999	0.871	1.146	0.252
as.factor(state)CT	1.294	0.879	1.472	0.141
as.factor(state)DC	1.369	0.982	1.395	0.163
as.factor(state)DE	1.275	0.945	1.349	0.177
as.factor(state)FL	0.695	0.840	0.827	0.408
as.factor(state)GA	0.410	0.853	0.480	0.631
as.factor(state)HI	1.076	0.946	1.138	0.255
as.factor(state)IA	0.937	0.893	1.049	0.294
as.factor(state)ID	0.050	0.974	0.052	0.959
as.factor(state)IL	0.884	0.845	1.045	0.296
as.factor(state)IN	0.721	0.864	0.834	0.404
as.factor(state)KS	0.059	0.919	0.064	0.949
as.factor(state)KY	0.937	0.869	1.078	0.281
as.factor(state)LA	0.759	0.878	0.865	0.387
as.factor(state)MA	1.370	0.863	1.588	0.112
as.factor(state)MD	1.000	0.869	1.150	0.250
as.factor(state)ME	1.273	0.970	1.313	0.189
as.factor(state)MI	1.087	0.850	1.279	0.201
as.factor(state)MN	1.045	0.879	1.189	0.234
as.factor(state)MO	0.704	0.860	0.819	0.413

	Estimate	Std. Error	t value	Pr(>|t|)
as.factor(state)MS	0.379	0.906	0.419	0.676
as.factor(state)MT	0.810	0.999	0.810	0.418
as.factor(state)NC	0.832	0.849	0.980	0.327
as.factor(state)ND	-11.201	186.529	-0.060	0.952
as.factor(state)NE	0.912	0.969	0.941	0.347
as.factor(state)NH	1.334	0.983	1.358	0.175
as.factor(state)NJ	0.808	0.850	0.950	0.342
as.factor(state)NM	1.202	0.956	1.257	0.209
as.factor(state)NV	0.373	0.882	0.423	0.673
as.factor(state)NY	0.966	0.840	1.150	0.250
as.factor(state)OH	0.787	0.845	0.930	0.352
as.factor(state)OK	0.216	0.893	0.242	0.809
as.factor(state)OR	1.014	0.866	1.171	0.242
as.factor(state)PA	0.611	0.846	0.723	0.470
as.factor(state)RI	1.349	1.084	1.245	0.213
as.factor(state)SC	0.114	0.871	0.131	0.896
as.factor(state)SD	0.143	1.083	0.132	0.895
as.factor(state)TN	0.363	0.869	0.417	0.677
as.factor(state)TX	0.435	0.841	0.517	0.605
as.factor(state)UT	0.218	0.921	0.237	0.813
as.factor(state)VA	1.045	0.850	1.229	0.219
as.factor(state)VT	3.454	1.352	2.555	0.011
as.factor(state)WA	1.113	0.858	1.298	0.194
as.factor(state)WI	1.075	0.862	1.248	0.212
as.factor(state)WV	0.560	0.921	0.608	0.543
as.factor(state)WY	0.799	1.505	0.531	0.595
age	-0.004	0.002	-1.857	0.063

References

1. Tausanovitch, Chris and Lynn Vavreck. 2020. Democracy Fund + UCLA Nationscape, October 10-17, 2019 (version 20200814). Retrieved from [https://www.voterstudygroup.org/downloads?key=fbeaf64b-ea8a-48a9-be61-52eccb9b4bc4].

2. 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. https://doi.org/10.18128/D010.V10.0

3. S. (2018, October 19). Rename Data Frame Columns in R. Datanovia. https://www.datanovia.com/en/lessons/rename-data-frame-columns-in-r/

4. Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. International Journal of Forecasting, 31(3), 980–991. https://doi.org/10.1016/j.ijforecast.2014.06.001

5. FairVote.org. (n.d.). What Affects Voter Turnout Rates. Retrieved November 01, 2020, from https://www.fairvote.org/what_affects_voter_turnout_rates

6. 2020 United States presidential election. (2020, October 30). Retrieved November 01, 2020, from https://en.wikipedia.org/wiki/2020_United_States_presidential_election