Fall 2020 Data Challenge

Ran Liao, Zedong Wu

Nov, 8th, 2020

In this data challenge, we examined data-driven questions that are related to social issues like racial justice and gender inequality in the American society. Our analysis focused on showing the voter turnout across all races/genders, states, the common reasons for people in the least-participated race and the races with the largest population to not vote, and the voting methods across states.

Having imported the dataset, we extract the needed variables (columns) and change the values of variables that were represented by numbers to meaningful string literals using the *select* function and joining the datasets with the data dictionary.

In the first analysis, we grouped both datasets by race and used R functions to count the total number of voters and total population in each race. Next, a new column “*voted\_percent*” was created using the mutate function whose values are the total respondents per race divided by its voted population, which calculates the voted percentage by race. We analyzed the results in a barplot and noticed that for the races that have the largest population (White, Black, Asian), over half of the population in the dataset voted and White and Black have a much higher voter turnout than Asian. However, nearly one third of White and Black population and half of the Asian population still did not vote. When dividing the results by gender using the *facet\_wrap* function, we did not find a significant difference in voter turnout across gender.

Economics and political situations could potentially impact the voter turnout. In the second research question, we analyzed the voter turnout across all the states in the country. Here, we grouped the datasets (2016 & 2018) by states and calculated the voter turnout in each state. The results are shown in a map plot built by the *map\_data* function and the state of Colorado has the highest voter turnout. In addition, states on the east and west coast tend to have a much higher turnout than those in the south, central America and midwest, where most toss-up states are located.

We then wanted to analyze the common reasons for people to not (registered to) vote. Here, we also grouped the data by race and showed the reasons in pie charts. Despite grouping, “didn’t like candidates or campaign issues” and “not interested, felt my vote won’t make a difference” are the top two reasons for not vote; “not interested in the election or not involved in politics” and “did not meet registration deadlines” are the top two for not register among all races.

Because of Covid-19, we have seen a great number of mail-in ballots in the 2020 election, which some have suspected as a source of voter fraud. Thus, we analyzed the main voting methods across states and plotted the results on a map. We noticed that back in 2016 and 2018, in seven states including Washington, Oregon and California, the number of mail voters surpassed that of those who voted in-person, indicating that voting by mail has long been used.