

Lab 1: Voter Patterns in the 2020 General Election

Maria DiMedio, Austin Jin, Jillian Luicci

1. Is the average age of self-identified Democratic voters different from self-identified Republican voters in 2020?

1.1 Importance and Context

In many ways, the 2020 Presidential election created much greater partisan polarization than recent elections. The nation was particularly divided by its views on the government’s handling of the COVID-19 pandemic and the history of systemic racial injustice. Since much of the Presidential campaign and the Presidential debates occurred during state-mandated lockdowns, many citizens used their time at home to focus on the election.

Understanding the demographics of American voters is critical to predicting the outcomes of elections and describing the views of generational groups. When voters self-identify with a political party, the voter tends to select their party’s candidate in upcoming Presidential elections (Petrocik). Additionally, grouping voters by generation allows us to describe generations in terms of their political priorities. A better understanding of the voters who are affiliated with the two largest political parties in the United States could help us understand the political attitudes and behaviors of Americans. Further, the government could use this information to provide more targeted support to its citizens.

1.2 Description of Data

To answer this question, we will use the data from the 2020 American National Election Studies (ANES) Time Series Study. In this dataset, all observations are individual respondents to the survey.

To ensure the sample included only ‘voter’ respondents, this portion of this report limits the data to individuals that stated that they either voted early or on election day. Further, the dataset was narrowed to respondents who completed the post-election survey to maintain the integrity of testing. Lastly, this report only contains analysis on survey respondents who provided their age. For improved readability, we operationalized the data by replacing the numeric values for the parties with the naming descriptors ‘Democratic’ and ‘Republican’ as defined in the ANES Time Series Data codebook.

Figure 1.1: Voter Party of Registration in 2020

Democratic	Republican	None	Other
44.71%	31.61%	23.03%	0.65%

As reported in Figure 1, roughly 75% of voters identify as either Democratic or Republican. We will focus this analysis on this majority population going forward. Party of registration was determined using pre-election survey question. The post election party of registration survey only contained a response of ‘Inapplicable’ for the voter population defined above; as such, the post-election party of registration field was not considered in this analysis.

Figure 1.2: Party of Registration

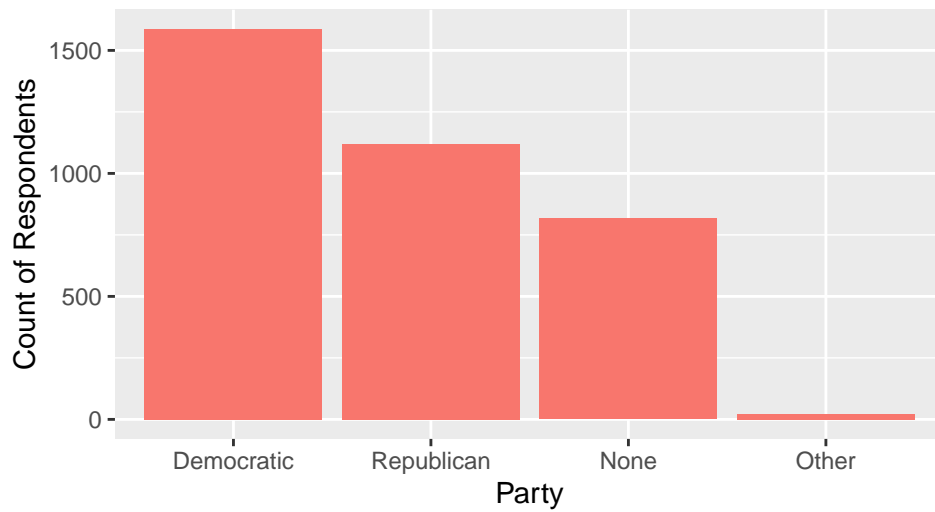
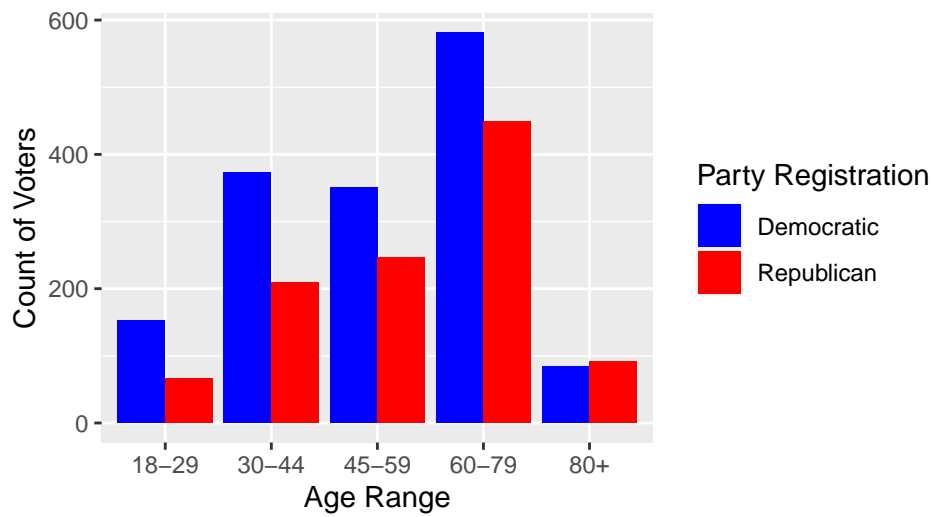
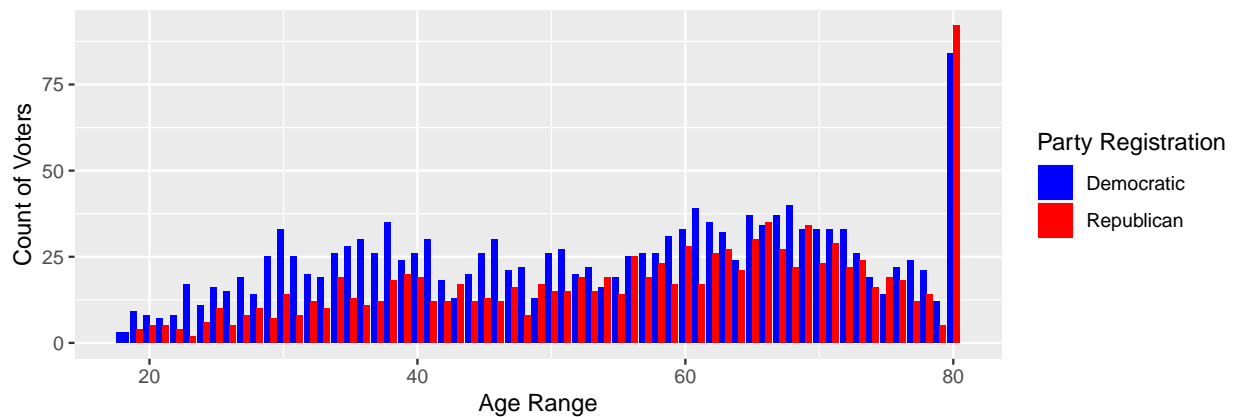


Figure 1.3: Democratic and Republican Age Range



Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defa

Figure 1.4: Democratic and Republican Age Range



As seen in Figure 4, there are fewer voters affiliated with either the Democratic and Republican parties in the 18-29 and 80+ ranges. The age range with the most party affiliation is the 60-79 range. The ages of Democratic and Republican voters have a slightly negative skew. As seen in Figure 3, there is a spike of voters aged 80. This is due to a limitation in the survey data; all voters 80 and above are considered to be 80 for the purposes of this report.

1.3 Most Appropriate Test

We have two groups, Democrats and Republicans. Each group's age is measured on a metric scale. Therefore, the most appropriate test is a two-sample t-test.

The two-sample t-test requires the following assumptions to be true. First, data must be on a metric scale. Generally, age is on a metric scale since each voter responded with a numeric age. It is important to note that all voters aged 80+ are considered 80. We believe this slight violation of this assumption is acceptable given the 80+ bucket only represents 6.5% of all Democratic and Republican voters. We do not believe using 80 for this group of people will significantly impact the means and corresponding results. Second, the data should be independently and identically distributed. Age and political party are both independent variables. Since the individuals were randomly sampled, this criteria is sufficiently met. Lastly, there should be no deviations from normality. We measured the skew to be $-.27$, but the sample size is large enough to rely on the Central Limit Theorem ("CLT") which implies a normal distribution of the mean.

1.4 Test, Results and Interpretation

The null hypothesis that we are testing is that the age of democratic voters is the same as republican voters. The alternative hypothesis is that the age of democratic voters is not the same as the age of republican voters.

This test produces enough evidence to reject the null hypothesis. The p-value for this test is very small, and falls within our rejection range. However, we note that the mean age of democrats is only ~ 3.6 years less than that of republicans. While there is statistical evidence that we should reject the null hypothesis, the difference of means is practically insignificant. Further, since we consider voter age demographics in terms of generational buckets, the expected age of both democrats and republicans is in the Boomer generation.

1.5 Test Limitations

We have conducted this test based on the 2020 ANES Time Series Data. There were two major limitations in the test described in this report. First, we were limited to using pre-election survey data for each voter's party of registration. Any changes to voter party affiliation after the pre-election survey was done will not be reflected in this report. This data represents the voters party affiliation at a point in time, but this affiliation may change at any point following the pre-election survey. Second, all respondents 80 and above were considered to be 80. This has some undefined impact on the means we used to conduct our two-sample t-test. While our test was able to reject the null hypothesis given the data in the ANES survey, it is possible that using precise age figures for the 80+ respondents could change the results of our test.

2. Did Democratic voters or Republican voters report experiencing more difficulty voting in the 2020 election?

2.1 Importance and Context

The 2020 General Election was an anomaly in terms of voter turnout. The high sociopolitical tensions leading up to the election caused millions of Americans to feel strongly about casting their vote. Simultaneously, the country was in the middle of battling Coronavirus, and the pandemic made health safety a priority when holding the election. This engendered new protocols for casting mail-in ballots, and increased the overall number of ballots cast. The American National Election Studies (ANES) survey conducted polled American voters on how difficult of an experience they had casting their votes in the 2020 election. Since the election, efforts have been made by legislators to restrict voting access and introduce additional protocols that many believe will increase how difficult it is to vote, largely in Republican-led states. To the extent that this may be a result of election outcomes, it may help to understand the reported voting difficulties in the 2020 General Election, comparing across political parties.

2.2 Description of Data

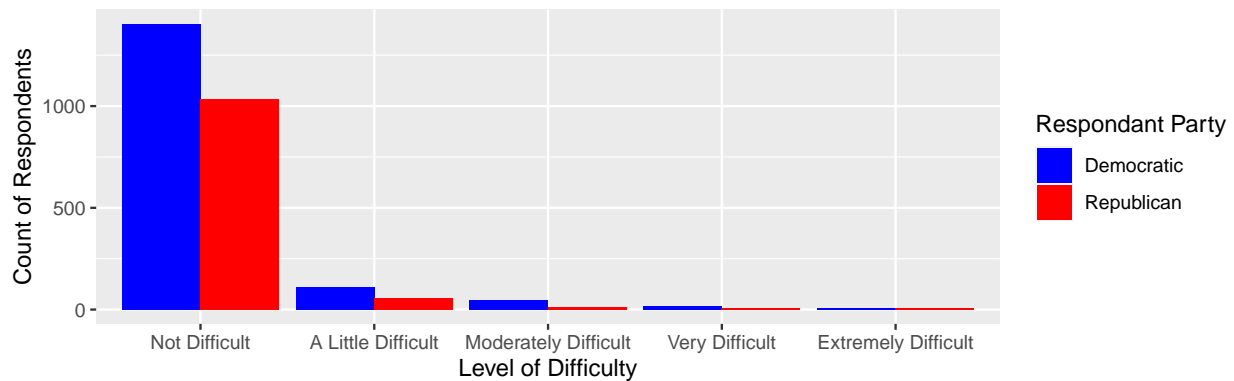
By our definition, we will refer to Democratic and Republican voters as those who report being registered to either party, and who report casting a vote in the 2020 election. In this dataset, all observations are individual respondents to the survey. In order to refine this sample to Democratic and Republican voters, the dataset was filtered by two survey data fields. First, to ensure the sample included only ‘voter’ respondents, a filter was added on the post-election survey question (‘V202116’) asking when the respondent voted. This ensures that all observations are of individuals who report having cast their vote. Second, party affiliation was determined by the pre-election survey question asking respondents which party they were registered with (‘V201018’), extracting Democratic and Republican registered respondents.

Additionally, the dataset was filtered by removing those respondents who did not respond to the post-election survey. Because we are using pre and post survey questions, this filter was applied to maintain the integrity of testing. To operationalize this research question, we will use the party affiliation variable to group Republicans and Democrats, and then will use the ordinal variable on level of voting difficulty to compare voting difficulty across parties.

Three additional columns were created in this dataset after subsetting on the above filters. One column indicating the political party the respondent was affiliated with, and a second for the level of difficulty they experienced when voting. This was done to appropriately visualize the numeric data fields. The third is a variable which represents respondents either had some level of difficulty of voting, or no difficulty, regardless of the degree of difficulty they reported. This was done to analyze overall the quantity of experienced voting difficulty in this sample.

The majority of Democratic and Republican voters did not experience difficulty voting, as seen in the barplot below. When omitting those that answered this question with ‘not difficult’ response, the distribution of response appears to indicate that most Democrats experienced little difficulty, while those voters that experienced the most difficulty were Republican.

Figure 2.1 Voting Difficulty in the 2020 Election by Party
Majority of respondents did not experience difficulty voting.



Having created a dummy variable to compare some level of difficulty to no difficulty in voting, the below plot shows the summary statistics for Democratic and Republican respondents' overall reaction to voting challenges.

Figure 2.2 Respondents With No Difficulty vs. Some Level of Difficulty Voting

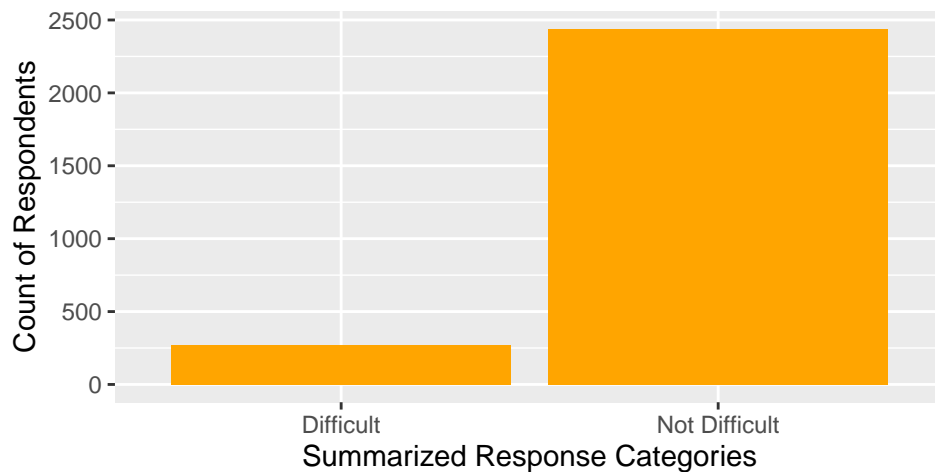
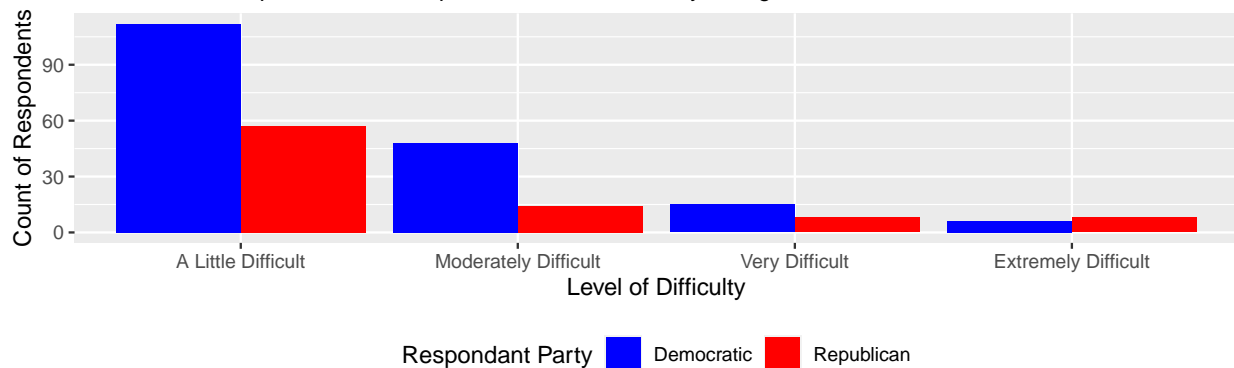


Figure 2.3 Voting Difficulty in the 2020 Election
Distribution of respondents who experienced some difficulty voting.



Interestingly, the trend in distribution of data in the Figure 2.3 appears to indicate that for those Democratic

respondents who responded that they experiences some level of difficulty, their level of difficulty tended to be low on the scale (2-5) of ranked difficulty. In contrast, not many Republican respondents indicated that they had experienced as much difficulty, except for in the most extreme category, in which more Republicans than Democrats indicated they had an extremely difficult time casting their vote. This can also be seen in Table 2.

Table 2: Comparison Table of Some Level of Voting Difficulty Across Parties

	2	3	4	5
Democratic	0.42	0.18	0.06	0.02
Republican	0.21	0.05	0.03	0.03

2.3 Most appropriate test

In order to test this research question, the Wilcoxon rank-sum test was selected. This is because the data of interest is unpaired and nonparametric, and because the response variable we are evaluating is an ordinal categorical scale indicating the respondent’s level of difficulty voting. We will implement this test in R, with the null hypothesis of comparisons being that there was no difference between probability that Republicans experience more difficulty than Democrats, and the probability that Democrats experience more difficulty. As an alternative hypothesis, we propose that there is a difference in the probability distribution for how likely it is that Republicans or Democrats report experiencing voting difficulty. The rejection criteria for this test will be an alpha of 0.05.

The assumptions that must be met for the Wilcoxon rank-sum test are: 1. The data is non-parametric and unpaired: The data in this sample analysis is unpaired, as we are comparing across two groups (Republicans and Democrats), and This data meets the first assumption. 2. The variable is ordinal: The variable being compared is column ‘V202119’ an ordinal response variable ranking level of difficulty on a scale of 1-5, which meets the second assumption. 3. The sampling is done using I.I.D.: We assume that random sampling has taken place when the ANES survey was operationalized, and we assume that the sample is distributed identically to the population, which in this case would be voting Americans. Independence is met by the random sampling of respondents as well.

2.4 Test, results and interpretation

The results of this test are statistically significant. From this test, the p-value result is 0.00175, which is less than the alpha 0.05 established as a rejection criteria. With this result, we will reject the null hypothesis that the probability of Republican and Democratic voters to experience difficulty voting was the same. We can see from this result that the probability of a Democratic voter to have some level of difficulty voting is greater than the probability of a Republican voter to have any difficulty voting.

In considering the effect size, here we calculate an r correlation effect size and get a result of approximately 0.084. On the scale of effect sizes according to Cohen’s classification, this is a relatively small effect size.

The practical significance of this outcome indicates that all voters do not experience the same amount of difficulty when casting their vote. Democrats tend to be more likely to report experiencing these challenges than Republicans, which raises additional questions regarding why that would be the case. Additionally, for government bodies, it is important to review this information and understand that there is a potential for inequitable voting patterns to emerge or influence election outcomes. Given the large number of respondents who experienced no difficulty voting, and the small effect size of this result, the practical significance while meaningful, does not indicate that there is an overwhelming amount of party-line voting challenges. The median response for the democratic response and the republican response to this survey question is both 1.

2.5 Test Limitations

Given the statistical significance of this result, there is a difference between the likelihood of voting difficulty across political parties in the US. The outcomes of this test are limited, however, by the fact that this election was held in a very tumultuous time in American politics. We cannot guarantee that response data is an accurate representation of experiences rather than emotions. For example, there could be respondents who feel frustrated by their government and report more difficulty voting than actually was experienced. Additionally, we cannot draw causality between party membership and level of voting difficulty. As we have seen, most of the respondents experienced no difficulty whatsoever, in both parties.

3. Are people who believe that science is important for making government decisions about COVID-19 more likely to disapprove of the way their governor is handling the pandemic?

3.1 Importance and Context

The COVID-19 pandemic emerged in the United States at a moment of extremely high levels of political polarization on the one hand and extremely low levels of trust in government institutions on the other. With this backdrop, government agencies pushed a set of health mandates for all Americans to adapt, such as social distancing and vaccine injections, to limit the transmission of the virus. These orders were unprecedented in how they required all individuals to alter their everyday routines to reduce the spread of the disease and help protect public health. With that being said, it became crucial to analyze data on how a certain segment of scientifically-aligned individuals are responding to the government's methods on handling the pandemic. The data analysis and results would examine the extend of the relationship between political worldviews of individuals who believe in the importance of science for making COVID-19 decisions and their trust in the way governors have been handling the pandemic by drawing conclusions through a unique data set posted by ANES. Understanding how politics shape health behaviors and risk perception is especially pressing when we consider that the pandemic emerged among anxieties about the trustworthiness of government institutions.

3.2 Description of Data

The two variables used from the 2020 American National Election Studies (ANES) were V201147x and V202310. Since the question asked specifically for people who believe that science is important for making government decisions about COVID-19, we utilized the V202310 variable which provided a scale from 1 to 5 (1 being not important at all to 5 being extremely important) on how important respondents selected science to be for decisions about COVID-19. We particularly focused on scoping the data analysis to just using 5 (extremely important) and 1 (not important at all) so that we could utilize filtered data that pertained to only two binary groups of people who believed in the importance of science for making COVID-19 decisions versus those who don't believe in the importance of science for making COVID-19 decisions. Out of the respondents who believed that science is extremely important and who believed that science is not important at all for making decisions about COVID-19, we added in a layer of approval levels through V201147x that scaled from 1 to 4 (1 being approve strongly, 2 being approve not strongly, 3 being disapprove not strongly, and 4 being disapprove strongly). Just like how V202310 was filtered to only show data on respondents who selected 'extremely important'/'not important at all', the 2nd variable V201147x was also filtered to only show respondents who selected 1 through 4 and excluding -2 (DK/RF in V201145 or V201146) since it seemed to have no relevance with the approval levels.

In essence, the method was to first decide how to form a group of the following two elements: People who believe in science for making decisions about COVID-19 along with their approval levels on how the governor is handling the pandemic VS. People who don't believe in science for making decisions about COVID-19 along with their approval levels on how the governor is handling the pandemic. Then the next steps were to measure the results of both groups and decide whether people who believe or don't believe in science were likely to either approve or disapprove of how the governor is handling the pandemic.

Figure 3.1 Belief In Science and Pandemic Approval Levels
 People Who Believe Vs. Dont Believe Science is Important in Decision Making

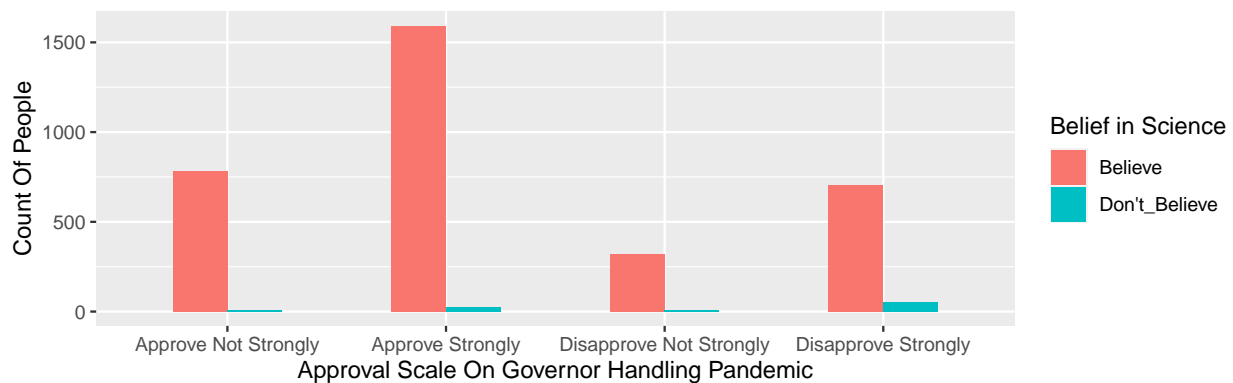


Figure 3.2 People Who Believe In Science – Pandemic Approval Levels
 Count of People Who Believe In The Importance Of Science And Their Approval Levels

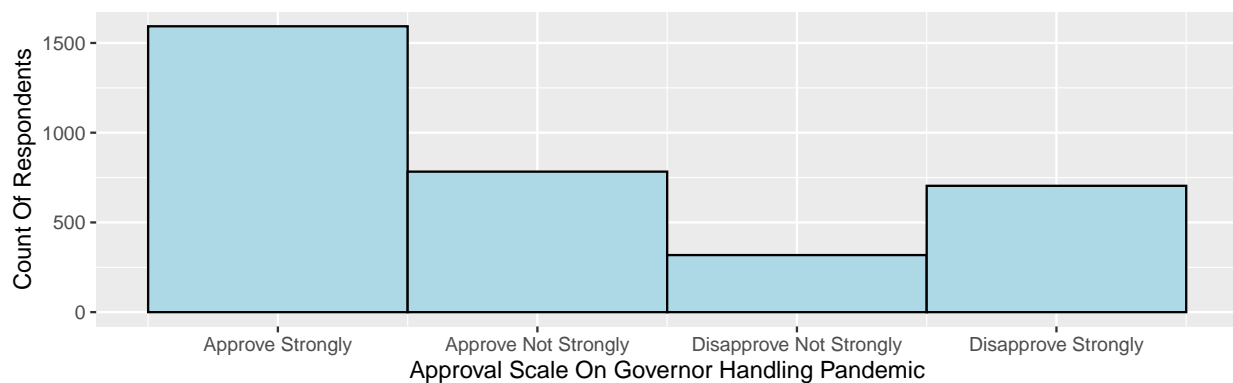
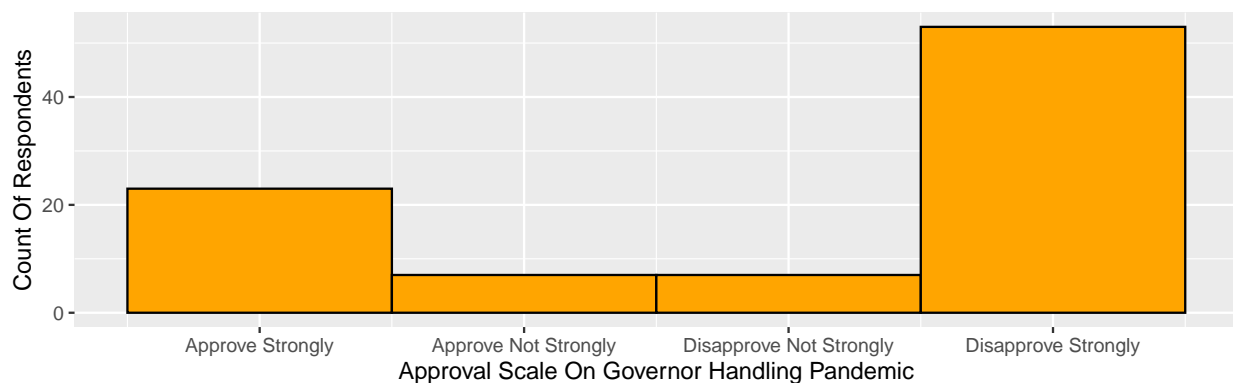


Figure 3.3 People Who Dont Believe In Science – Pandemic Approval Levels
 Count of People Who Dont Believe In The Importance Of Science And Their Approval Levels



3.3 Most Appropriate Test

The most appropriate test to use was the Wilcoxon Rank-Sum test because the two groups are unpaired and variables are ordinal. We have also made our assumptions that the data is I.I.D, as stated in previous sections of this report, and variables for respondents who either believe or don't believe in science and their approval levels were ordinal. The null hypothesis for this research question is that there is no difference in the probability that those who believe in science's importance will be more likely to disapprove their government's handling of the pandemic, compared to the probability that those who don't believe in science's

disapproval. As an alternative, we propose that there is a difference in the probability that one group have a greater disapproval than the other.

3.4 Test, Results and Interpretation

After conducting the Wilcoxon Ranked-Sum Test for both groups, the results are statistically significant since the p-value is less than 0.05 with an approximate value of $3.607e^{-11}$. This would mean that there is enough evidence to reject the null hypothesis since the p-value was less than the rejection criteria of alpha 0.05. The effect size of this test is given as 0.17, which is considered to be relatively small.

In parallel, the practical significance of this outcome indicates that the likelihood of disapproval of one's governor's handling of the pandemic will not be the same for those who believe in science and those who don't. From this analysis, those who don't believe in science tend to disapprove of how the governor is handling the pandemic whereas those who believe in science tend to approve. It is also worth noting that the difference in location which is another way of saying the median of the difference between the sample of those who believe in science and those who don't believe in science had a value of 1 which is considered small.

3.5 Test Limitations

As for test limitations - Since the two groups weren't equal in size, there could be an underlying skew in the effect size of the test being run. It is also evident that the outcomes of this test are limited due to the fact that the election was held in a very tumultuous time in American politics and we cannot guarantee that the response data is an accurate representation of typical emotions. In order to arrive at the most appropriate results, we had to perform some data cleansing to remove any variables that didn't fit the scope of people's belief in science along with their approval scales. Even so, some respondents might've disapproved of the way governors are handling the pandemic regardless of their inclination to science or not. We also cannot draw causality between people's belief in science and their disapproval counts with the way governors are handling the pandemic even though the results of the wilcoxon ranked-sum test provided a very low p-value that enabled us to reject the null hypothesis.