Lab 1: Voter Patterns in the 2020 General Election

Maria DiMedio, Austin Jin, Jillian Luicci

Contents

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. 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. 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.

Table 1.1: Voter Party of Registration in 2020

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

As reported in Table 1.1, 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 a 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: Democratic and Republican Age Range

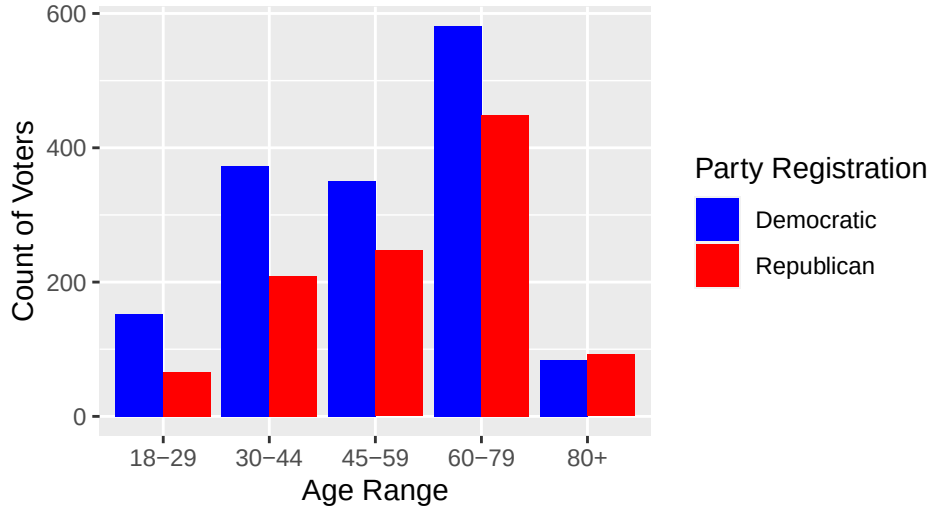
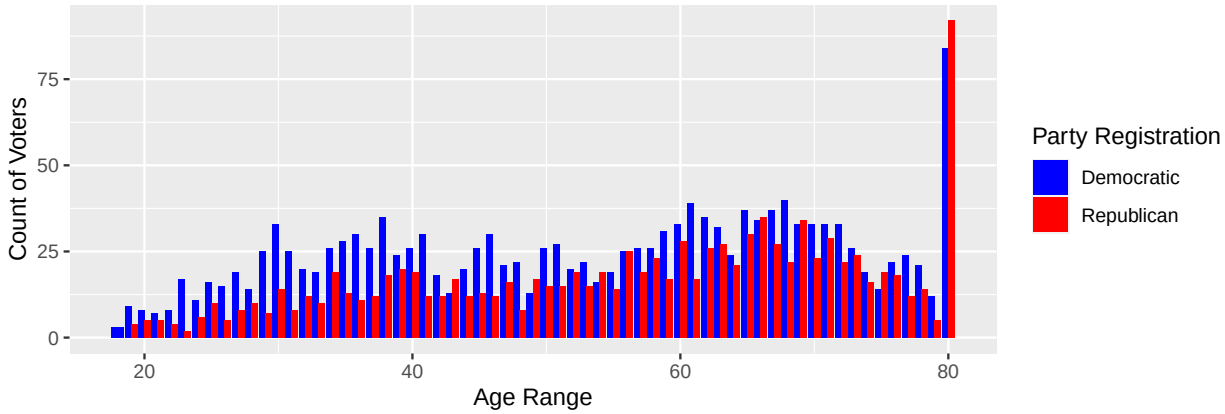


Figure 1.3: Democratic and Republican Age Range



As seen in Figure 1.2, 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 1.3, 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

In this research question, there are two groups, Democrats and Republicans, which we are interested in comparing. The variable of interest in this case, 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 considered to be 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 mean age of Democratic voters is the same as Republican voters. The alternative hypothesis is that the mean age of Democratic voters is not the same as the mean age of Republican voters.

```
t_test <- t.test(V201507x ~ political_party, data = voters_dem_rep_only_18orOlder,
                 na.rm=TRUE)
```

This test produces enough evidence to reject the null hypothesis. The p-value for this test is very small (6.177e-08), and falls within our rejection range, which is an alpha of 0.05. However, we note that the mean age of Democrats is only about 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, in a practical sense, we typically consider voter age demographics in terms of generational buckets. In this case, 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 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, 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 Figure 2.1. 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. Having created a dummy variable to compare some level of difficulty to no difficulty in voting, Figure 2.2 shows the summary statistics for Democratic and Republican respondents' overall reaction to voting challenges.

Figure 2.1 Voting Difficulty in the 2020 Election by Party

Majority of respondents did not experience difficulty voting.

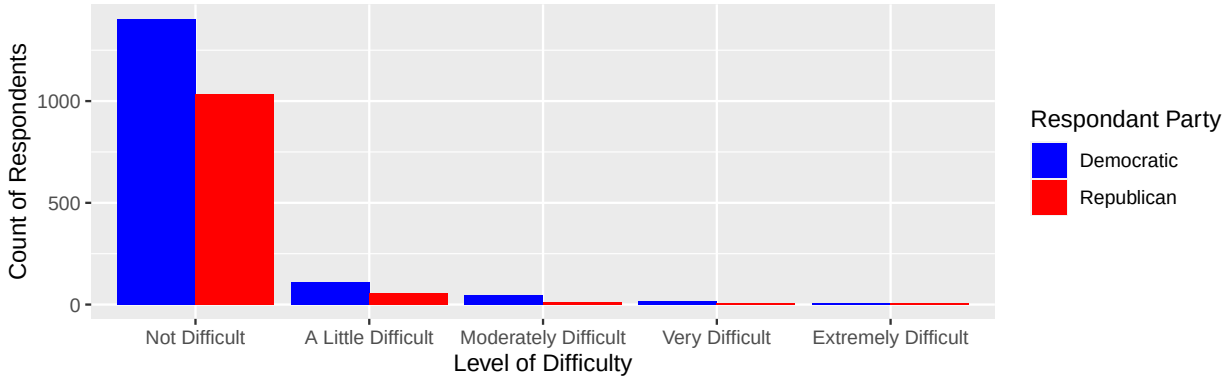


Figure 2.2 Respondents With No Difficulty vs. Some Level of Difficulty Voting

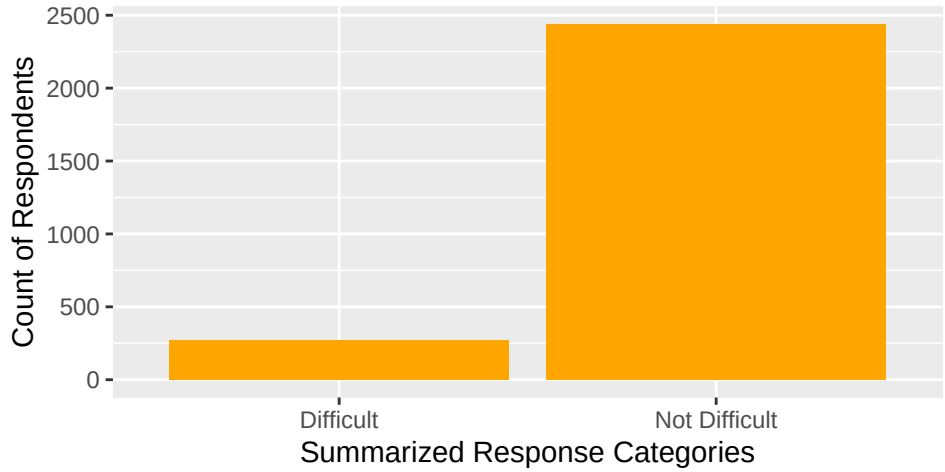
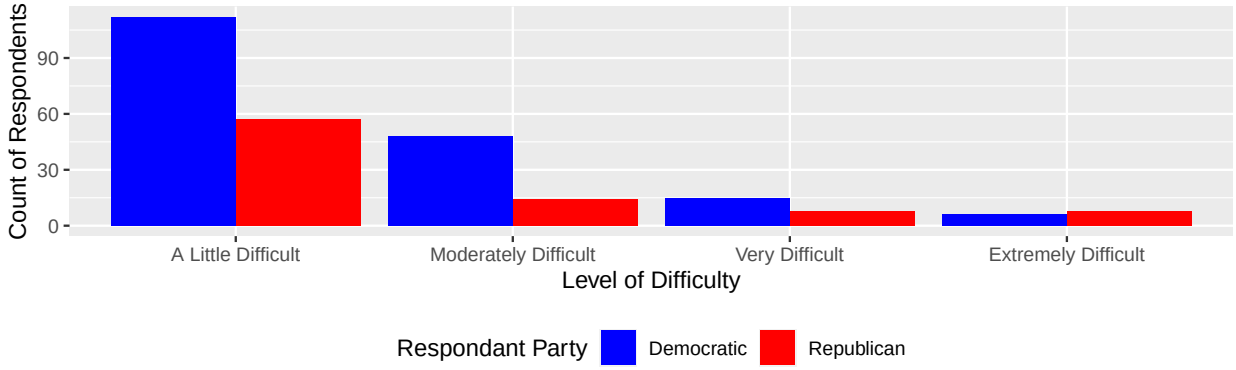


Figure 2.3 Voting Difficulty in the 2020 Election

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 the level of difficulty Republicans or Democrats report experiencing. 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; 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), so the data meets the first assumption. Second, 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. Last, the sampling is done using I.I.D.: Random sampling has taken place when the ANES survey was operationalized, and we can therefore 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

```
res <- wilcox.test(V202119 ~ political_party , data = voters_dem_rep_only,
                   alternative = 'two.sided', exact = FALSE)
```

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, or "No Difficulty".

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

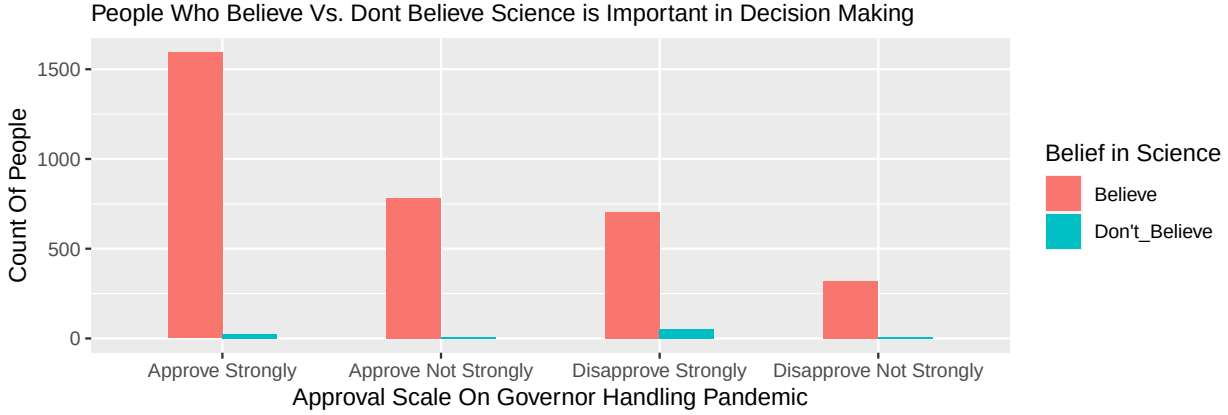
The COVID-19 pandemic emerged in the United States at a moment of extremely high levels of political polarization on the one hand and extremely low levels of trust in government institutions on the other. With this backdrop, government agencies requested all Americans to adapt social distancing and lock-down protocols 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 extent 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 of interest here is specifically focused on 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 analysis to use 5 (extremely important) and 1 (not important at all), filtering the data to utililize two binary groups of people who emphatically believe 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. We also operationalized this question by focusing on the variable V201147x, which reports ordinal level of approval for a Governor's handling of the pandemic. This variable scaled from 1 to 4 (1 being approve strongly, 2 being approve not strongly, 3 being disapprove not strongly, and 4 being disapprove strongly). Using this variable, we filtered the dataset to only include respondents who selected 1 through 4, excluding non response to this survey question.

In summary, this test specifically contains data on respondents who provided valid rankings for both support of science and governor approval.

Figure 3.1 Belief In Science and Pandemic Approval Levels



As shown in Figure 3.1, a majority of those who believe in science approve of how the governor is handling the COVID-19 pandemic. More specifically, those who believe in science tend to "Approve Strongly" of how the governor is handling the pandemic whereas those who don't believe in science happened to "Disapprove Strongly" of how the governor is handling the pandemic which is logically correct in terms of a direct relationship. In Figure 3.1, it can also be seen that the size of each subgroup is quite different; many more respondents believe in science than those that don't. Due to the differences in size between the two groups of those who believe in science versus those who don't, the separated graphs are presented in Figures 3.2 and 3.3 to independently graph out the distribution of responses.

Figure 3.2 People Who Believe In Science − Pandemic Approval Levels

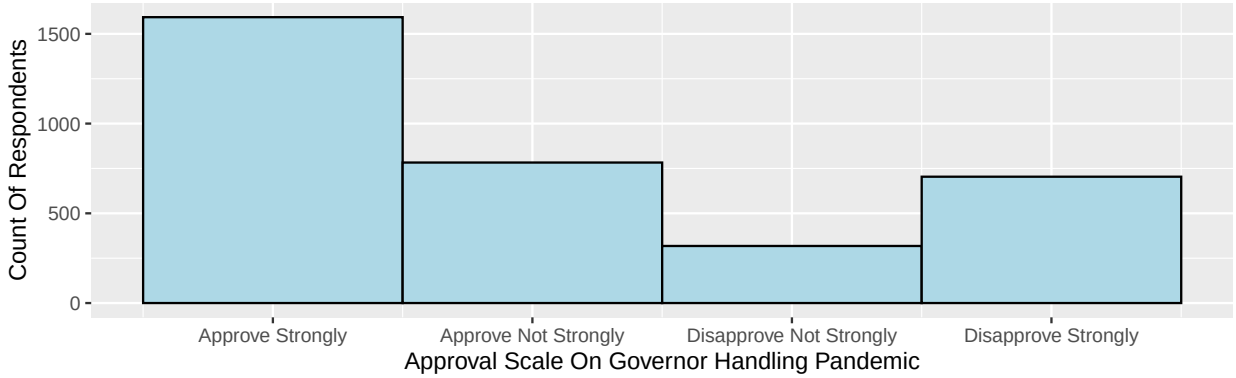
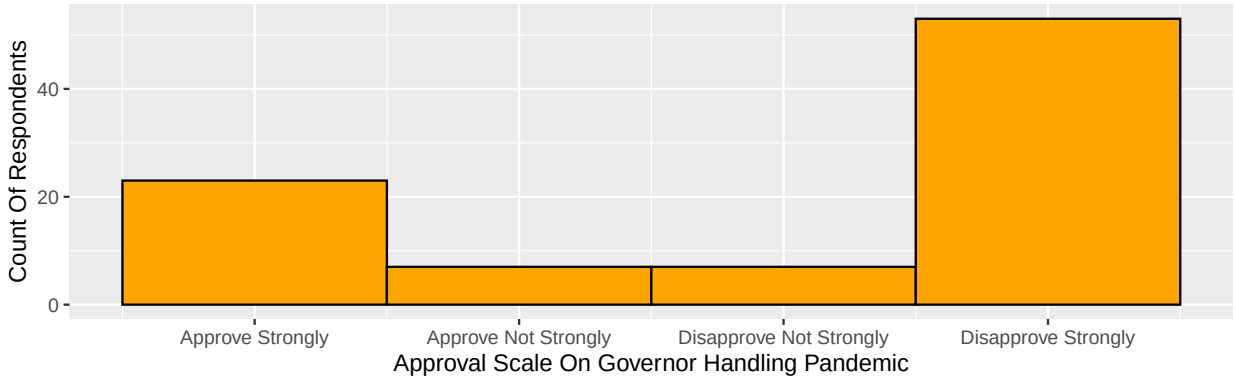


Figure 3.3 People Who Dont Believe In Science − Pandemic Approval Levels



3.3 Most Appropriate Test

In essence, the method was to first group 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. The most appropriate test to use to compare these traits is the Wilcoxon Rank-Sum test because the data of interest is unpaired and non-parametric, and because the response variable we are evaluating is an ordinal categorical scale indicating the respondent's level of approval for their Governor's handling of the Pandemic.

The assumptions that must be met for the Wilcoxon rank-sum test are: First, the data is non-parametric and unpaired: The data in this sample analysis is unpaired, as we are comparing across two groups (Those who believe in science versus those who don't), and this data meets the first assumption. Second, the variable is ordinal: The variable being compared is column 'V201147x' an ordinal response variable ranking level or approval on a scale of 1-4 which meets the second assumption. Third, the sampling must be done using I.I.D.:

The random sampling has taken place when the ANES survey was operationalized, and we can therefore assume that the sample is distributed identically to the population. Independence is met by the random sampling of respondents as well.

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 disapprove of their government's handling of the pandemic, compared to the probability that those who don't believe in science's will disapprove. The alternative hypothesis would be that there is a difference between these probability distributions for each group.

3.4 Test, Results and Interpretation

```
wilcox2 <- wilcox.test(V201147x ~ V202310, conf.int=T, conf.level=0.95, paired=F,
                       data = extremes, alternative = 'two.sided', exact = FALSE, rm.na=T)
```

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 3.607eˆ-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's importance tend to disapprove of how the governor is handling the pandemic whereas those who believe in science's importance tend to approve. Moreover, 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. It is also worth mentioning that there are a lot of factors that go into governor approval and it would be challenging to isolate belief in science as a variable.

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 worth mentioning that the overall sample size of those who don't believe in science was relatively very small compared to the size of those who believe in science which would be one of the limitations for using the ANES Times Series Data. It is also evident that the outcomes of this test could be 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 of 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. One other limitation is that there wasn't a question explicitly asking whether the respondent believes in science and that we don't know the reasons behind why a respondent may disapprove of their governor; whether it'd be aligning with their political affilation or other factors outside the scope of this analysis.