

Lab 1: Question 2

Hassan Saad, Chandler Haukap, Courtney Smith

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

(Exploratory, ungraded question: Were the types of reported difficulties the same, or different for respondents who identify with the different parties? We're not going to grade this, and it isn't a fully formed question here, but you [like us] might be interested in exploring this while you're writing up your report.)

Importance and Context

The Covid-19 pandemic was clearly a great cause for concern during the 2020 presidential election. Setting aside the fact that the two majority party leaders took different stances on what actions were necessary to quell the virus, Covid-19 greatly factored into US citizens' ability to vote in a safe manner. We can expect that last year especially there was an increased likelihood for voters to express difficulty getting to a polling site. In addition, there was clearly an increased amount of animosity between supporters of Democratic candidate Joe Biden and Republican candidate Donald Trump. Several media reports warned voters of potential violence and anger at polling locations, further decreasing propensity to vote in-person. Many turned to alternative methods such as mailing ballots (which also had its inherent difficulties to achieve) to avoid potential exposure to the virus or other danger.

Through this question, we explore how likely respondents of each party were to indicate that they had difficulty voting. We focus less on the cause(s) and more on the amount of difficulty reported by people of each major political affiliation.

Description of Data

As our base data set, we collected a report from the 2020 American National Election Studies (ANES). It includes pre- and post-election data consisting of 1381 variables and a sample size of 8280 respondents.

For this question, we selected columns of the original data that were specific to political affiliation and respondent-reported difficulty in casting a vote. The experienced difficulty was best represented by the original column V202119, which we renamed Difficulty_Voting. This data follows a Likert scale from 1 'Not Difficult at All' to 5 'Extremely Difficult'.

The relevant column of the original data set related to political affiliation seems to be V201231x, which we rename to 'Political_Affiliation'. A value of 1 in the response corresponds to a voter being a "strong democrat" and 7 a "strong republican". We eliminate values of 4, -8 and -9 from our data set since these correspond to supporting an independent party, not knowing, or refusing to respond, respectively. Through this method we develop the following assignments of voter political affiliations shown here:

Table 1: Meaning of Numerical Political Affiliations

Number	Affiliation
1	Strong Democrat
2	Not Very Strong Democrat
3	Independent Democrat
5	Independent Republican
6	Not Very Strong Republican
7	Strong Republican

We then take the approach of creating ‘binary’ data among the political parties, and we group every respondent that answered 1-3 or 5-7 into either ‘Democrat’ or ‘Republican’ respectively to produce the following quantities of respondents in each group:

Table 2: Distribution of Respondents for Each Major Political Party

Major Party	Count
Democrat	3130
Republican	2702

This classification of political parties is used for the remainder of this question. It’s the first significant change we made in terms of organizing the data for this question. It relies on the assumption that a respondent who feels more strongly about their political affiliation offers no more statistical significance in our study as somebody who is a little more ‘centrist.’ At first this seemed like too big of an assumption to make. For example, if the majority of the Republican party were made up of “Independent Republicans” while the majority of Democrats were “Strong Democrats,” we’d probably have to rethink this strategy.

Then we delved into the distribution of each party affiliation and noticed a helpful pattern. It nicely trends towards higher values on the far sides of the political spectrum with an approximately symmetric valley towards the “centrist” tendencies. This means that both major political parties have approximately even proportions of people who are ‘strong’ versus more ‘independent’ supporters within our sample.

Table 3: Political Affiliation and Count of Respondents

Affiliation	Count
1	1692
2	665
3	773
5	674
6	628
7	1400

Before carrying out an analysis, it was important for us to get some intuition behind the data distribution. We plotted a bar chart that shows the percentage of respondents from each party that experienced difficulty voting. Note that because we have different numbers of Democrats and Republicans overall in the available data, it’s important to account for this in the chart (i.e. it would not make sense to just plot the *number* of people per level of difficulty reported). Simply put, it displays the proportion of people that expressed each difficulty level in voting. We couple this with a table showing the same data numerically.

Voting Difficulty

Proportion of Respondents that Expressed Each Level of Difficulty Voting

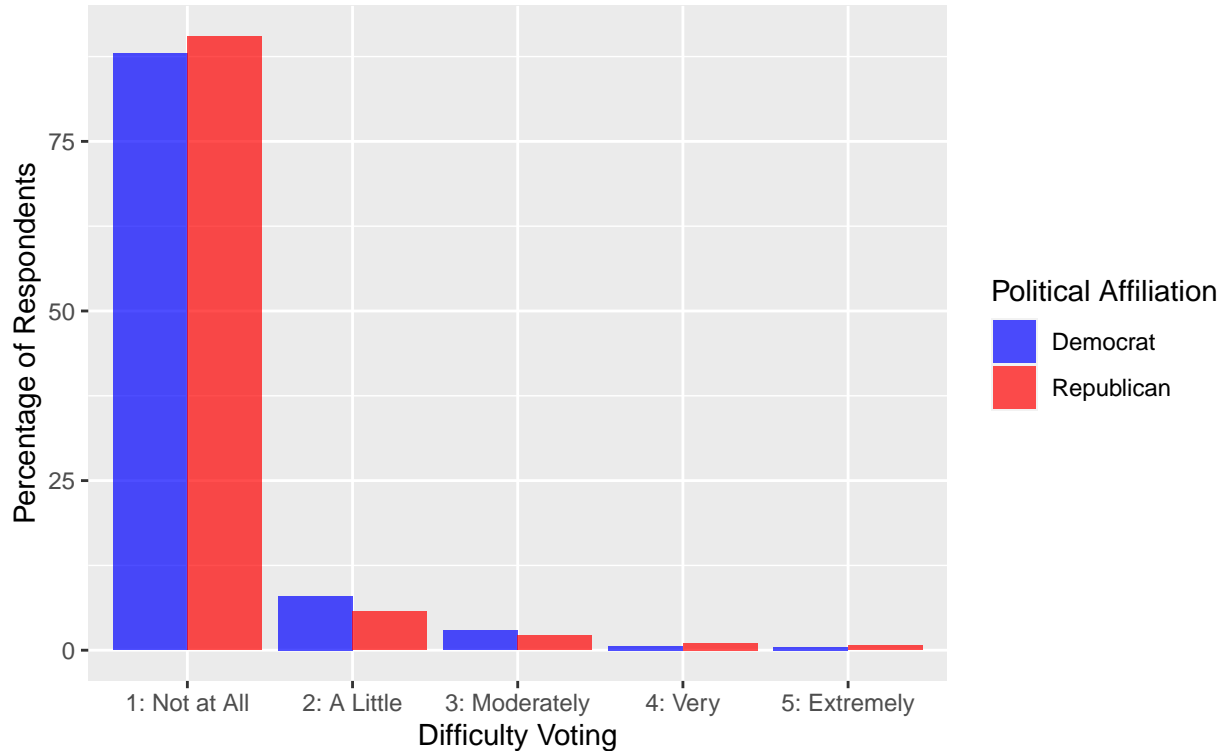


Table 4: Percentage of Respondents who Expressed Each Level of Difficulty Voting

Ordinal_Difficulty	Democrat	Republican
1: Not at All	87.96	90.45
2: A Little	8.02	5.70
3: Moderately	2.94	2.15
4: Very	0.64	1.04
5: Extremely	0.45	0.67

Most appropriate test

For this question, the most appropriate test is like a Wilcoxon rank sum test. First, this is due to the fact that we are comparing the mean difficulty level (Likert scale 1-5) for each party. We have no baseline measurement for difficulty level, which means we can't establish an understanding of how large the delta is between two lowest difficulty levels compared to the delta between the two highest. Therefore, we cannot classify the data as metric. We also do not believe this is parametric data with an underlying normal distribution, so a t-test would not be the appropriate solution here. Lastly, due to the fact that we are separating the groups by political party, we inherently do not have paired data and cannot rely on a signed-rank test.

In terms of the null hypothesis, it's not really easy to gauge from the above plot what comparison to establish. Again, we know we'll be comparing the mean voting difficulty level between Republicans and Democrats, but we cannot get a sense of whether or not one will be greater than the other. Therefore, we can rely on a two-tailed test to simply determine whether or not the means are equal based on our sample.

We can use a conventional critical value of $\alpha = .05$ for this test as well.

Null Hypothesis $H_0 : \mu_{democrat} \neq \mu_{republican}$

```
wilcox.test(Difficulty_Voting ~ Party, data = q2_data)

##
## Wilcoxon rank sum test with continuity correction
##
## data:  Difficulty_Voting by Party
## W = 4329723, p-value = 0.003526
## alternative hypothesis: true location shift is not equal to 0
```

Test, results and interpretation

After running a Wilcoxon test on the sample, we get a P-value of 0.003526. This number is well below the established critical value, and it indicates that only 0.3526% of samples are at least as contradictory to the null hypothesis than the one present in our data set. This is therefore a highly statistically significant sample, and we can reject the null hypothesis that the means are equivalent.

Ideally, we could go back and choose a one-sided test for this question. Looking over the numbers in Table-4 again, it might make sense to formulate a null hypothesis that Democrats had more overall difficulty voting. We could assume that since just under 88% said they had no difficulty voting, around 12% did have difficulty, compared to just under 10% of Republican respondents. However, because we're working with a Likert scale, this is not a substantial claim off of which to base our hypothesis. For example, 1.04 % of Republicans said their voting experience was '4: Very Difficult' compared to 0.64% of Democrats. It's possible that this difference could carry more weight than that which is present between parties regarding a difficulty level of '2: A Little Difficult'. Therefore, we did not have enough intuition behind the data to make an argument for a one-sided test.

The outcome would be practically significant if we could show that the mean values of difficulty voting expressed by Democrats and Republicans were very different. However, this is not the case. While the means are not equivalent as determined by the Wilcoxon test, we could probably only consider this difference to be practically significant if this test could show that they correspond to different values on the ordinal scale. We would need to somehow attach a metric value to the level of difficulty experienced. Then we'd need to be able to identify the fact that, for example, Republicans in general had 'A Little' difficulty voting while Democrats in general had a 'Moderately' difficult voting experience.