Lab 1: Question 2

Peter Morgan, Bruce Lam, Mia Yin

Did democratic voters or republican voters experience more difficulty voting in the 2020 election?

Importance and Context

Did democratic voters or republican voters experience more difficulty voting in the 2020 election?

Throughout most of American history, there has been voter suppression in some regard. Women were not allowed to vote until 1919. In addition, African Americans have been the target of voter suppression due to unfair literacy tests and poll taxes up until 1965.

As a nation, America has come a long way in terms of equitable voting access. However, many people still report having trouble voting to this day and understanding why people have trouble voting can help us continue to improve voting equity in America. Answering the research question will give us insight as to whether republican or democratic voters experience more difficulty voting and from there, we can create more research questions depending on the results.

Description of Data

The dataset used to address this question will be from the 2020 American National Election Studies (ANES). This dataset is a random sample of around 8,000 participants meant to represent the population of 231 million non-institutional U.S. citizens aged 18 or older living in the 50 US states or the District of Columbia.

The first column of interest in this dataset is our grouping variable, Party of Registration. This is a simple and elegant way to filter between Republicans and Democrats which are the two groups of interest in our study. This data was collected by asking a participant “What political party are you registered with if any?” The answers available to the participant are refused, don’t know, inapplicable, democratic party, republican party, none or independent, or other. Table 1 below shows how this variable is distributed in this sample of data.

Table 1: Party Affiliation Distribution, Unfiltered

Refused	Don’t know	Inapplicable	Democrat	Republican	None/Independent	Other
9	2	4010	1861	1336	1029	33

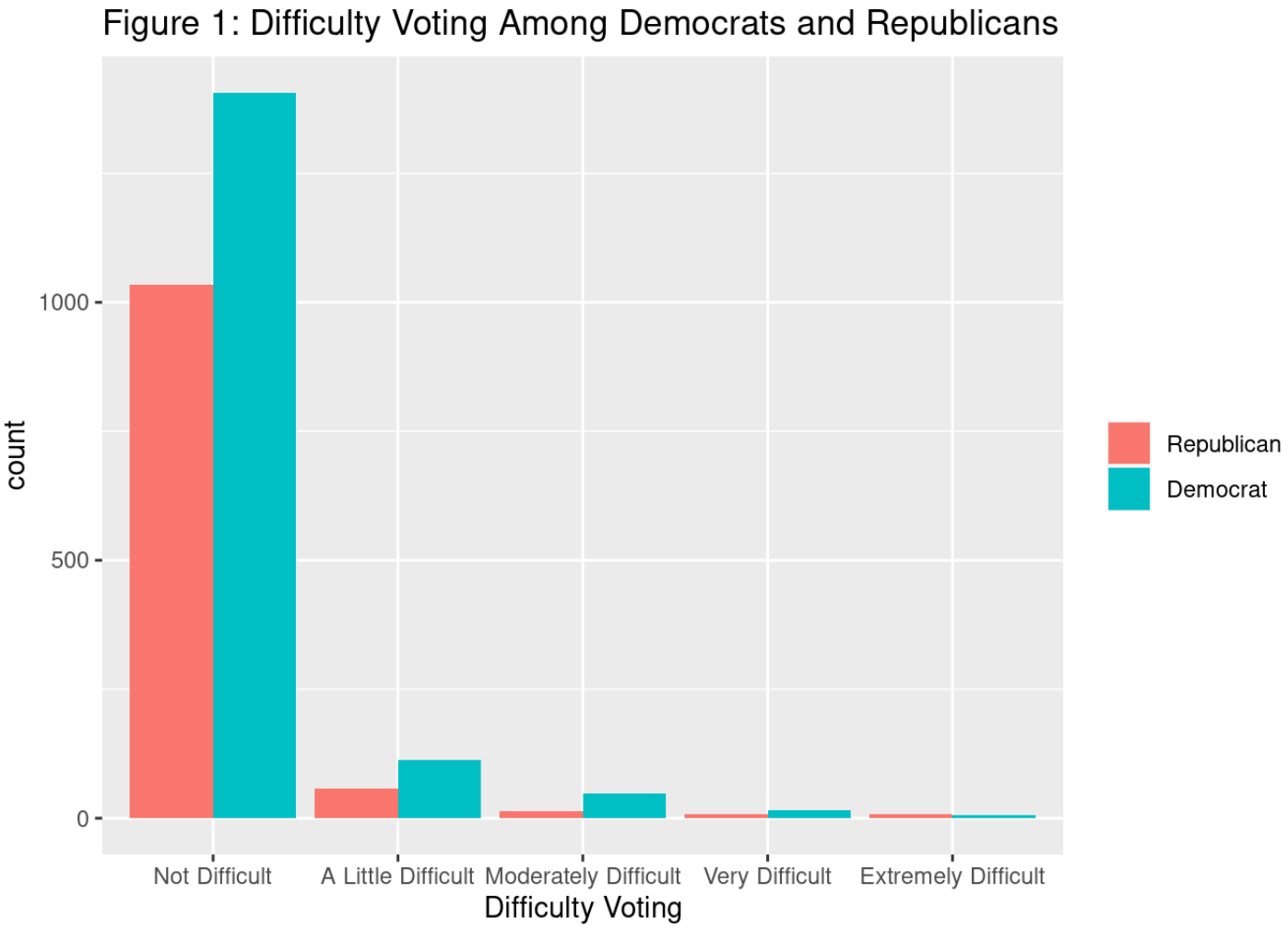
As you can see, a lot of this data is irrelevant to our question as we are only interested in comparing the Republican and Democrat groups. Because of this, we will filter out all data where the Party Affiliation is not democrat or republican. The next column of interest is Difficulty Voting. The reason this column was chosen is because it can easily be split up into a Bernoulli distribution with the options being there is difficulty voting, and there is no difficulty voting. This distribution will make it easy to perform a paired binomial test to answer our research question. The data for this variable was collected by asking the participant “How difficult was it for you to vote in this election?” The options are as follows. Refused, deleted due to incomplete interview, no post-election interview, interview breakoff, inapplicable, not difficult, a little difficult, moderately difficult, very difficult, and extremely difficult. Table 2 below shows the distribution of the difficulty voting variable.

Table 2: Voting Difficulty Unfiltered

Refused	No Data	No Interview	Interview Breakoff	N/A	None	Little	Moderate	Very	Extreme
1	77	750	2	1045	5675	449	179	56	46

Table 2 shows that there is a higher proportion of relevant data to our question compared to the data in Table 1. We will filter out any data that does not give us insight into difficulty voting. This includes data labeled refused, no data, no interview, interview breakoff, or inapplicable (N/A). After cleaning both variables, we now have only 2709 observations remaining compared to our initial sample of 8,000.

To get an idea of how the data is distributed after cleaning, we turn to figure 1. Figure 1 shows a histogram of depicting the difficulty of a voter filtered by political party.



As you can see, both parties experience a low proportion of overall voting difficulty. In addition, figure 1 shows that our sample contains more democrats than republicans. Now it is time to determine if there is a statistically significant difference in voter difficulty between democrats and republicans.

Most appropriate test

Before proceeding with the test, we need to manipulate the data a little bit more. We will transform the difficulty voting variable into a bernoulli variable with 0 corresponding to no trouble voting, and 1 corresponding to some amount of trouble voting. This manipulation does not consider the amount of trouble a voter had voting. This means that a response of “no difficulty” corresponds to the value 0 and the responses “little difficulty”, “moderate difficulty”, “very difficult”, and “extreme difficulty” all correspond to the value 1.

The first step in selecting a test is determining if we will use a paired or unpaired test. For this research question, we will use an unpaired test because there is only one measure per sample. The next step in determining the best test is to determine if we have metric or ordinal data. In this case we have ordinal data because the measurements are categorical like “no difficulty” and “extreme difficulty”. This eliminates any possibility of a t-test and leaves us with the option of performing a wilcoxon rank-sum test. Since we have ordinal data, we must use the hypothesis of comparison version of the test.

There are two assumptions for a hypothesis of comparison wilcoxon rank sum test. First, the two groups must both be measured on the same ordinal scale. This assumption is valid and was shown during the test selection process. The second assumption is that the data is IID. This assumption is valid according to the sample design section of the ANES user guide. Now it is time to state the null and alternative hypotheses. The null hypothesis is the probability that a republican experiences difficulty voting (P(R)) equals the probability that a democrat experiences difficulty voting (P(D)). The alternative hypothesis is, the probability that a republican experiences difficulty voting does not equal the probability democrat experiences difficulty voting. These hypotheses are stated mathematically below.

$$H_0 : P(R) = P(D)$$

$$H_a : P(R) \neq P(D)$$

Test, results and interpretation

To evaluate our null hypothesis, we can perform a binomial test to determine the sample probability of voter difficulty for a democratic voter $P(\hat{D})$ and a republican voter $P(\hat{R})$. In addition, these will include a 95% confidence interval for the true probability of voter difficulty. If the confidence intervals for republicans and democrats overlap, we will fail to reject the null hypothesis. If they do not overlap, we will be able to reject the null hypothesis at a 95% confidence level. The results of the two binomial tests are shown below.

$$P(\hat{R}) = 0.078 : 95\%\ CI = (0.063, 0.095)$$

$$P(\hat{D}) = 0.114 : 95\%\ CI = (0.099, 0.131)$$

Since the two confidence intervals do not intersect, we can reject the null hypothesis with 95% certainty. This means that with 95% certainty, democratic voters report experiencing more difficulty voting in the 2020 election than republican voters. To explain how meaningful this discovery is, we will now investigate the practical significance of this.

The first method we will use to evaluate practical significance is the difference in probability method. This is simply calculated by subtracting the sample probability of voter difficulty of democrats by that of republicans. Performing this calculation gives us a difference in probabilities of 0.036, a relatively small quantity. This small practical significance could be because the large sample size of democrats and republicans which are both over 1000. Another metric we will look at to evaluate practical significance is the odds ratio (OR). The odds ratio is simply the sample probability that democrats will experience trouble voting divided by the sample probability that republicans will experience trouble voting. The odds ratio for this example is 1.46 which is medium to low magnitude. This states that democrats are 1.46 times more likely to experience trouble voting.

Types of Reported Difficulties

To wrap up this research, we will perform an exploratory analysis on the types of voting troubles experienced by republicans and democrats. To explore this, we will need to import some more data. Ten more columns were added to the data. Each of the ten columns asked if the subject encountered a specific problem in the

election. The ten problems are shown as the column headers in table 3. The possible answers for all 10 columns are the same. They are refused, no post-election data, no post-election interview, inapplicable, not mentioned, and mentioned. For this exploratory analysis we are only curious in participants that have the difficulty “mentioned” or “not mentioned”. All other data will be filtered out. Table 3 below shows the proportion of participants that mention a specific difficulty filtered out by party.

Table 3: Percentage of Voters that Experience Difficulty

	Registration Problem	Identification Card Concern	Difficulty Obtaining Absentee Ballot	Ballot Or Machine Confusion	Difficulty Getting To Polling Place
Democrats	0.0082	0.0038	0.0202	0.0107	0.0057
Republicans	0.0018	0.0089	0.0054	0.0107	0.0071

	Long Wait Times	Work Schedule	Bad Weather	Issue Mailing Ballot	Other Problem
Democrats	0.0536	0.0246	0.0132	0.0170	0.0334
Republicans	0.0589	0.0223	0.0089	0.0036	0.0214

As expected with our experiment, more categories have democrats reporting a higher frequency of difficulty. The only difficulties that republicans report at a higher frequency are identification card concerns, difficulty getting to polling places, and long wait times. This raises even more questions than answers. Particularly in the variables that have a more pronounced difference. An example would be, why do democrats report registration problems at about 4 times the rate as republicans and is this statistically significant? Another example is why do republicans report over twice the amount identity card concern and is this statistically significant? Or are there traits that republicans or democrats are more likely to have that cause different types of voting difficulty? While we do not know these answers now, these are an excellent basis to conduct follow up studies to better understand why people have difficulty voting. To answer our initial question, democrats did report more difficulty voting in the 2020 election in the context of a 95% confidence interval. However, the practical significance is not as large as we would like. Fortunately, we have a strong direction for follow up studies that have a chance to better understand why Americans have difficulty voting.