

Analyzing Voting Difficulty

Datasci 203: Lab 1 Part 2

Zachary Fenton, Rob Horrobin, Yoni Nackash

Contents

1	Importance and Context	1
2	Data and Methodology	1
3	Results	3
4	Discussion	3

1 Importance and Context

Voting is an extremely vital component of a healthy democracy. A substantial electorate ensures that as many citizens’ opinions and voices are represented in the government as possible. As the United States is becoming increasingly more aware of systemic biases within our country, the topic of success versus struggle when voting is an important and interesting topic to be analyzed. There may in fact be inherent biases in our voting system that cause more difficulty to certain kinds of voters than others. We are interested in answering the following question:

Do Democratic voters or Republican voters experience more difficulty voting?

Answering this question will help provide insight into potential systemic biases in the structure and operations of elections in the United States. These insights will be key for both local governments as well as the federal government to make positive changes to our voting process, and reduce the difficulty voters experience.

2 Data and Methodology

We used data from the American National Election Studies (ANES)¹ to study the challenges faced by Democratic and Republican voters and how these experiences may vary based on political party preference. This dataset has been collected since 1948, and we focused on the 2022 Pilot Study, which examines voting and public opinion after the 2022 midterm elections in the United States.

The 2022 Pilot Study contains 1585 cases, which are a subset of over a million volunteers who responded to an internet survey on YouGov for a small compensation. These 1585 cases were selected by study designers to match the overall U.S. population in terms of gender, age, race, and education. While study designers calculated weightings to represent other characteristics, for this analysis, we didn’t consider these weightings.

We filtered our data based on responses to several survey questions:

- (1) Affiliation: respondents with strong, weak or leaning preference to the Democrat or Republican parties.
- (2) Voter Identification: based on survey responses, who do we consider a voter?
- (3) Voting Difficulty: evaluating the level of difficulty voters experienced in the 2022 election.

To meet the first criterion, we analyzed survey responses where respondents claimed to identify as a Democrat, Republican, Independent, or other. We also considered the strength of their affiliation, ranging from *Lean* to *Weak* to *Strong* as self-reported. Research by Petrocik (2009)² suggested that lean or weak persuasions indicate a clear preference for a particular party, so we treated *Lean* and *Weak* preferences as a reliable indicator of political preference. Considering our question focuses on Democrats and Republicans, we removed respondents who identify as Independents, other, or NA. Out of the 1585 cases, this results in 1310 respondents identifying as either Democrat or Republican. Reference Table 1 for a summary of political affiliation.

Table 1: Political affiliation vs strength

Self Identified Party	Number of Respondents
Strong Democrat	371
Weak Democrat	183
Lean Democrat	141
Lean Republican	159
Weak Republican	172
Strong Republican	284

¹American National Election Studies, 2022, “2022 Pilot Study”, <https://electionstudies.org/data-center/2022-pilot-study/>

²John Richard Petrocik, (2009, July 2). Measuring party support: Leaners are not independents. Electoral Studies. <https://www.sciencedirect.com/science/article/abs/pii/S0261379409000511>

Second, we are interested in analyzing respondents that voted in 2022. To identify these respondents, we included those who claimed to have either definitely voted, most likely voted, and probably voted in 2022. Out of the 1310 respondents that identify as either Democrat or Republican, 726 respondents claimed to have definitely voted, 318 respondents claimed to have most likely voted, and 3 respondents claimed to have probably voted. This leaves us with a sample size of 1047 voters.

Third, we determined a parameter to measure difficulty. We opted to focus on the general question “*How difficult was it for you to vote?*” as our target variable, rather than attempt to combine various questions of difficulty into a new parameter. We believe a combination of several questions could have reduced practical significance, as we would be unable to deconvolute the new scale of this parameter. The respondents answered the general question of difficulty on Likert scale from 1 - 5, where 1 corresponds to *Not difficult at all* and 5 corresponds to *Extremely difficult*. These responses are subjective and, like the rest of the ANES data is self-reported without direct measurement. Thus, we must evaluate our results in the context that perception of respondents may play a factor and may not be objectively conclusive.

As seen in Figure 1 below, the majority of voters in both parties identified with ranking 1, claiming to experience very little difficulty in casting their vote.

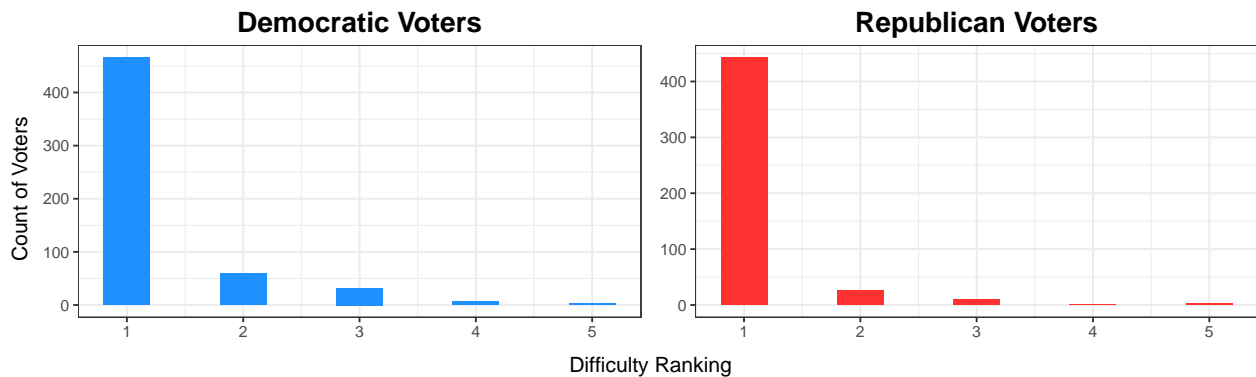


Figure 1: Distribution of voting difficulty by party

As stated above, we have identified two random variables - the difficulty in voting on a Likert scale, for both Democrats and Republicans. Because these data are ordinal and non-parametric, we selected the Wilcoxon Rank-Sum Hypothesis of Comparisons test over other two sample tests, such as the Two-Sample T-Test. We also acknowledge that based on the large sample size ($n=1047$), the central limit theorem would suggest that the sampling distribution of the statistic should be approximately normal, which would allow us to use a parametric test. We decided against this route due to the data being ordinal.

The null hypothesis of our Wilcoxon Rank-Sum Test is:

Null Hypothesis: *The probability that a Democrat identifying voter finds it difficult to vote is equal to the probability that a Republican identifying voter finds it difficult to vote.*

For the Wilcoxon Rank-Sum test to be a valid test, the following must be true, or assumed true:

First, the variables being measured are on an ordinal scale. As indicated above, our target parameter answers the question “*How difficult was it for you to vote?*” with answers ranging on a Likert scale of *Not difficult at all* to *Extremely difficult*. This scale is ordinal due to the ranking of the values, as well as the difference between each value being subjective.

Second, the data is drawn from an I.I.D. sample. Voters from the same party could experience similar difficulties in voting based on a number of considerations, such as socioeconomic status or distance to polling location. We acknowledge that this could introduce dependencies within our samples. However, given that sample is drawn from a population of millions of YouGov respondents, this would suggest that dependency or links between individuals would be uncommon. Thus, for the purposes of this analysis, we assume I.I.D is satisfied.

3 Results

```
test <- wilcox.test(votehard ~ Affiliation, data = abt, exact = FALSE)
```

The final result of the Wilcoxon Rank-Sum test provides sufficient evidence to reject the null hypothesis, ($p=6.0105867 \times 10^{-6}$). This means that there is a perceived difference in the way Democratic voters experience difficulty versus Republican voters. However, through our two-tailed test alone we are unable to interpret which party may find it more difficult. To help inform this, we standardized each sample distribution to 1, allowing us to compare distribution of difficulty across parties irrespective of population size. In Figure 2 it becomes clear that while the majority of both Democrats and Republicans responded to feeling no difficulty, there is a slightly larger distribution of Democrats responding 2 *A little difficult* and 3 *Moderately difficult* than Republicans.

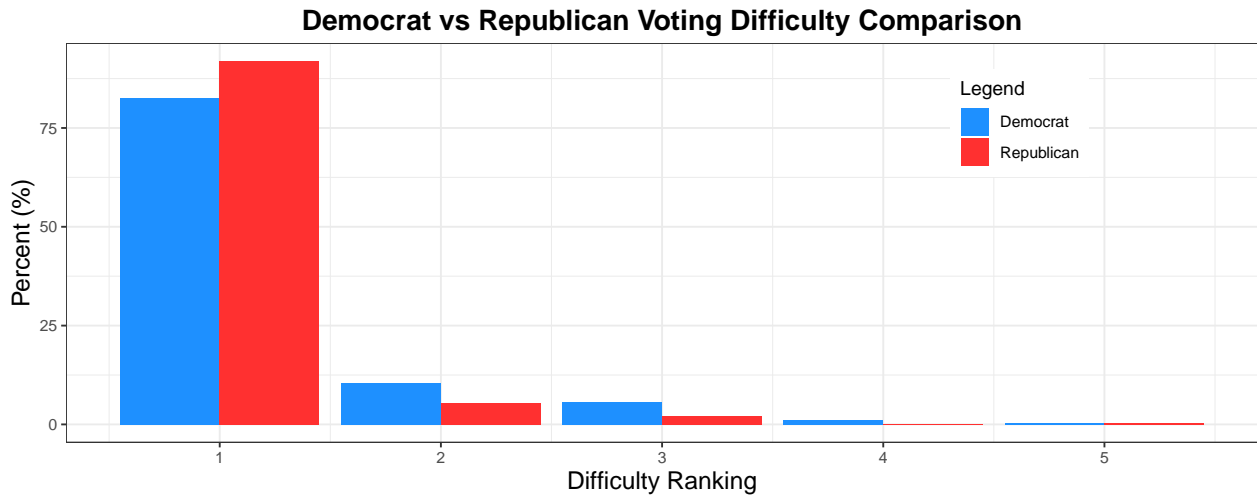


Figure 2: Comparing difficulty by party, as a percentage

4 Discussion

Our study has shown evidence that Democrats experience more difficulty voting than Republicans. While our results do not indicate direct causation between affiliated party and difficulty in voting, there is a perceived relationship between a voter's political identification and how much difficulty they experience while voting. We have found that 17.52 percent of Democratic voters experience difficulty voting, while 8.09 percent of Republican voters experience difficulty. The effect appears to be practically significant, with Democrats twice as likely to experience difficulty than Republicans. Given that the ANES 2022 Pilot Study is weighted and intended to represent the entire US population, 17.52 percent of the roughly 52 million Democratic voters from the 2022 General Election³ could represent millions of voters. This is a considerable amount of voting-eligible citizens that may be experiencing difficulty.

Our results may be of interest to key government officials, as well as election precinct organizers. While our analysis focused on whether one party experiences more difficulty than the other, future studies may be interested in identifying the root causes of difficulty. This may yield actionable insight on how the United States can restructure its voting processes, and hopefully reduce the difficulty experienced by American voters.

³MIT Election Data and Science Lab, 2017, "U.S. House 1976–2022", <https://doi.org/10.7910/DVN/IG0UN2>,