

Democrats vs. Republicans: Who Experiences More Difficulty Voting?

Datasci 203, Section 4

Pavan Sabnis, Zane Brown, and Addy Kim

Contents

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

1 Importance and Context

Voter turnout helps determine the course of a nation. However, the ease with which citizens can cast their votes has often been a point of concern, sparking debates on voter suppression and accessibility. The American National Election Studies (ANES), a reputable project ongoing since 1948 and federally funded since 1977, compiles data through surveys which shed light on various aspects of voter experiences in the United States. This repository of data serves as a window into the American electoral ecosystem, enabling researchers to delve into various questions. We aim to answer the question:

Do Democratic voters or Republican voters experience more difficulty voting?

Answering this question could provide meaningful insight to future political campaigns for either party, ultimately influencing the nation’s voter representation. For this report, we focused on registered voters who answered the survey question “How difficult was it for you to vote?” because it directly aligns with the research question.

2 Data and Methodology

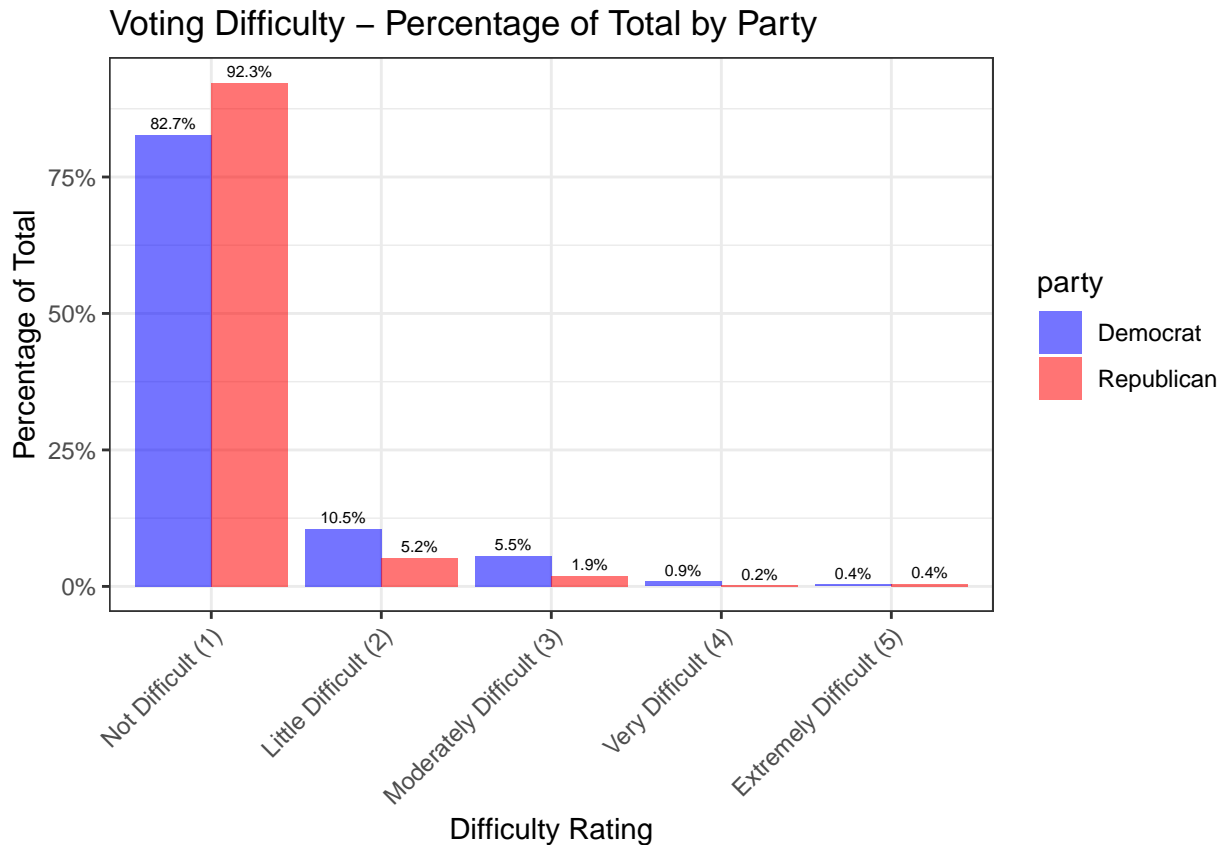
To evaluate whether Democratic voters or Republican voters experience more difficulty voting, we leveraged the 2022 data from ANES’s Pilot Study. The purpose of the study was to test potential new questions for the ANES 2024 Time Series Study and collect voting and public opinions after the 2022 midterm elections.

The 2022 ANES dataset contained 1585 observations. We conducted an exploratory data analysis (EDA) to understand these observations, identify potential data quality issues, and derive assumptions for answering the research question. Upon the EDA, we noticed a data quality issue in which 7 unregistered voters were able to vote. We dropped these 7 records. Then, to conceptualize our research question, we defined our key terms: Democrat, Republican, Voters, and Voting Difficulty. In Table 1, we defined voters as those who are registered to vote. Any individuals not registered were dropped. We defined voters as such because we wanted to include only the individuals who have the ability to vote, hence have the ability to experience voting. Voting difficulty is defined by its pre-existing definitions, with 1 indicating no difficulty and 2 through 5 indicating increasing levels of difficulty. Additionally, we presented a column titled ‘Party’, which categorizes data points from the ‘pid_x’ variable. Data points with values of 1 (strong democrat), 2 (democrat), and 3 (lean democrat) are grouped under the label ‘Democrat’, while values of 5 (lean republican), 6 (republican), and 7 (strong republican) are grouped under the label ‘Republican’. Any other data points within the ‘pid_x’ variable were excluded from the analysis. Finally, in the Votehard column, we dropped -1 cases (inapplicable, legitimate skip) because we only wanted to conduct the test on those who had experience with voting. For exact column value terms, please visit ANES’s codebook.

Table 1: Terms & Definitions

Term	Definition	No.of.Records.Post.Clean.Up
Voter	*reg* with values: 1, 2	1389
Voting Difficulty	*votehard* with values: 1, 2, 3, 4, 5	1159
Democrat	*pid_x* with values: 1, 2, 3	561
Republican	*pid_x* with values: 5, 6, 7	480

From these definitions, we plotted the percentage of respondents from either party (Democrats or Republicans) that provided each difficulty rating from 1-5. As shown below, most respondents from either party selected 1, with about 92% of Republicans selecting 1 (no difficulty) and 82% of Democrats selecting 1. Conversely, the subsequent increasing difficulty rates were selected by roughly twice as many Democrats as Republicans. However, this gap shrinks as the difficulty rates increase.



To operationalize our research question, we explored how we can measure our concepts. First, we determined that the Party and Votehard variables would be the best indicators of our research question. We considered other variables such as vharder_x but found them to be answering a slightly different question: “Why are voters experiencing difficulty?” Therefore, we chose Party and Votehard because we ultimately want to know whether Democratic or Republican voters (indicated by Party) had more difficulty voting (indicated by Votehard). Given that the two variables are not obviously paired and neither variable’s distributions are easily recognizable, we used the hypothesis of comparisons version of the Wilcoxon Rank-Sum Test. We considered other tests, such as the unpaired t-test, but we ultimately could not identify a distribution with our desired variables. We also considered the Wilcoxon Signed-Rank Test but could not prove our variables to be paired, so we chose the Wilcoxon Rank-Sum Test.

The Wilcoxon rank-sum test (hypothesis of comparisons version) has two assumptions: (1) Ordinal scale, and (2) Independent and Identically Distributed (IID) data. Because Votehard is ordered from -1 to 5 (excluding -1, which was dropped), it lacks a true zero point and has unequal intervals between the categories. Because of this, Votehard satisfies the ordinal assumption. The Party data is used solely for grouping purposes and does not need to adhere to an ordinal scale. Therefore, the ordinal assumption is satisfied. Additionally, both the Votehard data and the Party data meet the IID assumption. Votehard represents individual voters’ difficulty levels, independent of others, and Party data categorizes voters into Democrats or Republicans, irrespective of others’ affiliations. Each data point in Votehard is drawn from the same distribution, and the same holds true for Party data. Votehard and Party are independent of each other because they were collected independently; one was not influenced by the other. In the case of a paired test, both Votehard and Party would have been drawn from the same records. However, since this is not the scenario here, we can assume that Votehard and Party are independent variables.

The null hypothesis of our test can be phrased as follows:

Null Hypothesis: *The probability that a Democratic voter had more difficulty in voting than a Republican voter did is equal to the probability that a Republican voter had more difficulty in*

voting than a Democratic voter did.

3 Results

```
wilcox.test(democrats$votehard, republicans$votehard)

##
## Wilcoxon rank sum test with continuity correction
##
## data: democrats$votehard and republicans$votehard
## W = 147644, p-value = 3.742e-06
## alternative hypothesis: true location shift is not equal to 0

cles_func <- function(group1, group2) {
  cles <- pnorm(0, mean(group1) - mean(group2), sqrt(sd(group1)^2 + sd(group2)^2), lower.tail = FALSE)
  return(cles)
}
cles_func(democrats_cles$votehard, republicans_cles$votehard)

## [1] 0.5819442
```

The test resulted in a Wilcoxon rank-sum test statistic of 147644 and a p-value of 3.742e-06. This very low p-value ($\ll 0.05$) indicates a highly statistically significant result, allowing us to reject the null hypothesis. Given this rejection, we explored which group experienced more difficulty. From the Voting Difficulty - Percentage of Total by Party table above, we observed a higher percentage of Democrats experiencing more difficulty than Republicans, but reaching even by Votehard 5. To further explore and demonstrate practical significance, we calculated the Common Language Effect Size (CLES). All levels of difficulty (Votehard 2 through 5) were categorized as 1, and no difficulty (Votehard 1) was categorized as 0. We then compared Democrats' Votehard against Republicans' Votehard. The resulting CLES value was 0.582, indicating a 58.2% chance that if we randomly selected an observation from Democrats and one from Republicans, the Democrat's experience of greater difficulty would prevail. We argue that the result is practically significant considering margins in elections can be razor thin. Even the smallest increase in voter turnout can influence the outcome of elections.

Our study has several limitations. First, the Wilcoxon rank-sum test, although significant, might have less power compared to paired, parametric tests, potentially affecting result strength. Additionally, our dataset grouped political parties into two categories, Democrat and Republican, while voting party affiliation falls on a spectrum, a nuance our test didn't capture. Lastly, while the ANES is reputable, it cannot fully represent the entire US population, and we cannot be entirely confident in the survey result's independence.

4 Discussion

The study established statistical significance in suggesting that the probability of a Democratic voter having more difficulty voting did not equal that of a Republican voter. When analyzing the percentage of total Votehards, Democrats were found to face more difficulty. The CLES result supports this argument and suggests practical significance. Given more time, a deeper analysis would be interesting to understand why Democrats experienced more challenges. This could involve examining variables such as 'vharder_x' and employing a test that accounts for the spectrum of party affiliation, contributing to a more comprehensive analysis. Such insights could help parties comprehend voting behavior better and make informed decisions. Additionally, these findings could signal parties to prioritize voter education initiatives alongside traditional strategies like fundraising to enhance voter participation and education efforts.