

Lab One, Part Two

Matt Lauritzen, Torrey Trahanovsky, Sophie Yeh

03/01/2022

Contents

1	Importance and Context	2
2	Conceptualization (Data) & Methodology	2
3	Results	3
4	Discussion	4

1 Importance and Context

The 2020 election was a tense election with a heated political environment and threats to citizen safety and the US economy at play. The overall impact was that while past elections may have had polarizing issues or candidates affecting voter turnout, the 2020 election occurred during social unrest with an actual deadly virus threatening voters from turning out at the polls amid lockdowns.

Also, during November of 2020, vaccines were not prevalent, with only clinical participants receiving the required dosages to have potential immunity against the virus. Thus, the first portion of our analysis seeks to answer which side of the political spectrum experienced more difficulty voting during the 2020 election. The specific research question our analysis will answer is:

“Did Democratic or Republican voters experience more difficulty voting in the 2020 election?”

By answering this question with an understanding of the environment in which voters were casting their ballots, and creating clear definitions as to what a “Voter”, “Democrat” and “Republican” are. It’s desired that the data and analysis can help readers understand more about how voters who side with a particular party can experience potentially increased difficulty in voting.

2 Conceptualization (Data) & Methodology

Our analysis took a detailed look at the dataset from the 2020 American National Election Studies (ANES) which utilized data from the Westat survey firm. The population was selected from an equal population of citizens, using the USPS computerized delivery sequence file (C-DSF), selecting from the population those who were at least 18 and resided at these addresses. Once selected, an invitation was mailed to potential surveyees within the 50 states and Washington D.C.. This population selection method for surveyees ensured that surveyees would more likely be eligible to vote and represent the differing voting practices among states.

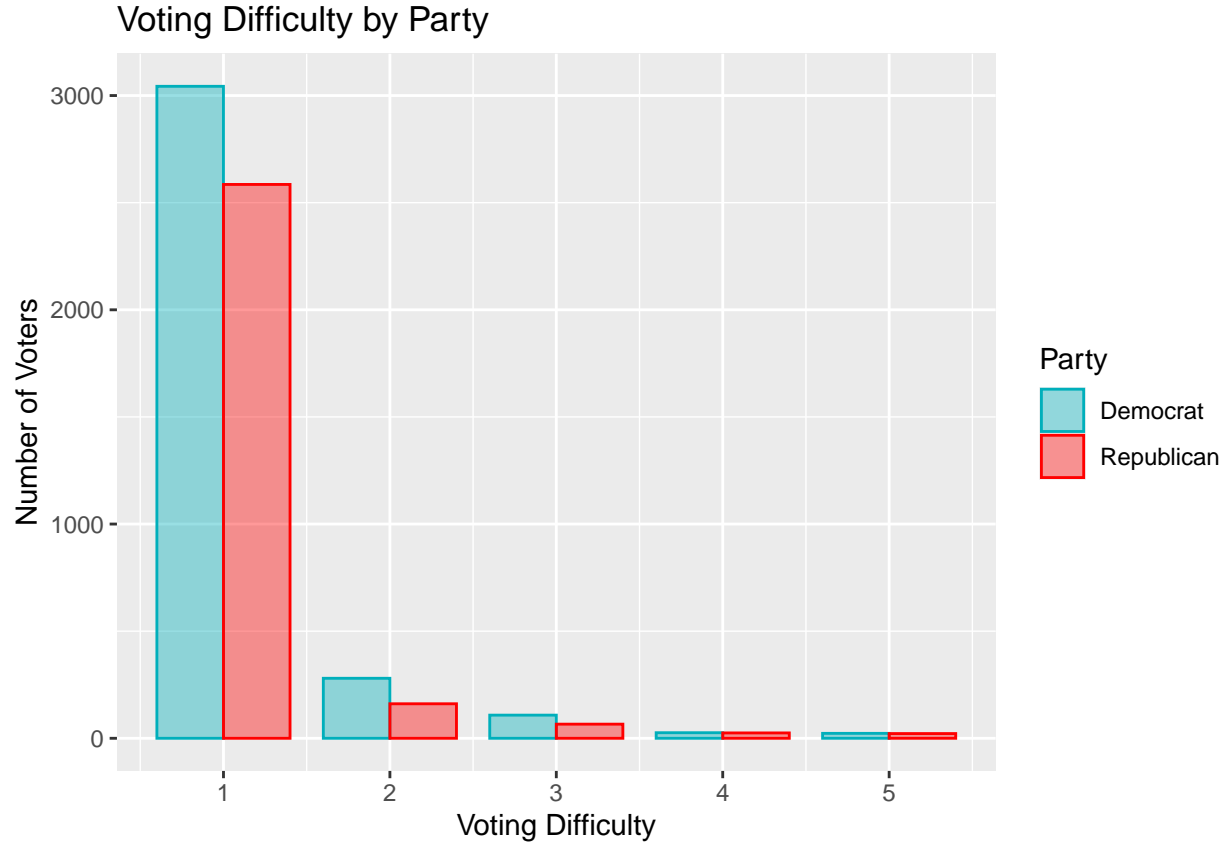
Given the limited number of 8,280 survey respondents it was crucial for the analysis to maintain the largest sample size from the dataset while maintaining accuracy in grouping our sample.

Thus, our team made the following assumptions regarding Democrat, Republican, and Difficulty to answer the research question. Difficulty is a subjective concept so in order to gauge the perceived difficulty experienced by voters, we used responses from the following question: “How difficult was it for you to vote in this election?” The respondents answered on a scale of 1-5 with 1 being “Not difficult at all” and 5 being “Extremely difficult”. Voters who did not respond to the difficulty question were removed from the analysis, resulting in 6,401 voters remaining. A necessary effect of using this variable was that difficulty rating was only asked in the POST survey, thus removing all voters who did not respond to the question due to only submitting the PRE survey.

To determine whether a respondent should be categorized as a Democratic voter or Republican voter, they were first checked to see if they were registered to the Democratic or Republican Party. If they were, then the analysis considered them the party of registration.

For remaining voters who did not register as Democrat or Republican, the major point of the 2009 research study by Petrocik shared in the lab Github repository was utilized in the analysis. Independents were classified as a Democratic or Republican voter based on factors other than their party membership. Thus, remaining voters were then checked to see if based on party ID summary, whether they leaned towards a particular party. If voters did, then voters were considered the party of lean. For all remaining voters without lean, the count of Democrat versus Republican candidate votes cast were tallied (numerous variables provided in data file). If there was a higher number of votes cast towards either Democrats or Republicans, then the voter was considered the party of higher votes.

For 62 voters, they did not show any lean towards either Democrat or Republican based on the above analysis and thus they were removed from the final test due to being considered truly independent voters, resulting in 6,339 voters for the final test.



The above visual shows the general distribution of the difficulty ratings for voters categorized in each of the parties. The test of a two-sample Wilcoxon Ranked Sum Test was selected due to the data, research question, and null hypothesis. The data was ordinal, as shown by the fact that responses were specific categorical type answers. Groupings occurred by having parties in the data and ties were limited with similar answers functioning as a further grouping mechanism.

Also, the research question requires answering which party had a greater difficulty. Essentially, asking whether two-samples, the sample population of Republicans and the sample population of Democrats, had an unequal difficulty voting and if so, which population had more difficulty. The Wilcoxon Ranked Sum Test allows the user to tell using medians, whether the samples are likely from the same population. In the context of the voter data, being from the same population would imply both Republicans and Democrats experienced difficulty in the same way during the 2020 election. Thus, the below Null-Hypothesis was used:

Null-Hypothesis: The probability that a Democratic voter had difficulty voting is equal to the probability that a Republican voter had difficulty voting.

3 Results

The test parameters for the `wilcox.test` function in R is as follows:

- Testing difficulty ratings in voters data.
- Paired = F means not a paired test.
- $\mu = 0$ is our null hypothesis that there is no difference between the two.
- Confidence Level = 0.95

```

# Run test
wilcox.test(difficulty~party,
             alternative = c("two.sided"),
             data=voters,
             paired=F,
             mu=0,
             conf.level=0.95,
             conf.int=T)

##
## Wilcoxon rank sum test with continuity correction
##
## data: difficulty by party
## W = 5118490, p-value = 0.0002911
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  2.808597e-06 7.832539e-05
## sample estimates:
## difference in location
##          5.281033e-05

```

Using the Wilcoxon Test, we determined that the p-value of the distribution is 0.003. The 95% confidence interval is [2.80e-06, 7.83e-05]. Finally, the difference in location is 5.28e-05. Although the difference in location is small (e-05), it has the same scale as our confidence interval.

4 Discussion

The test provides evidence to reject our null hypothesis that the two groups had equal difficulty voting with the results indicating that Democratic voters experienced more difficulty voting than Republican voters (p-value = 0.0003, difference in location = 5.281033e-05). The difference in perceived difficulty may be considered small, but in an environment of increasingly competitive national elections, any difference in difficulty of voting could prove vital in swinging the results of an election.

The Wilcoxon Rank Sum Test is advantageous for this purpose because it is easily understandable and does not require an assumption of the distribution. However, it is important to note that one of the major assumptions is that the data is independently and identically distributed. If the survey process is not genuinely random, data can be skewed.

In conclusion, the results of the study show that Democratic voters have experienced more difficulty voting than Republican voters. This is especially relevant at this moment as voting rights and changes to the voting process come to the forefront of political discussion. While this study has found potentially important results, it is not able to explain the causes of difficulty in voting so we would propose additional research to determine if the causes are related to voting logistics or other factors. By isolating individual factors, we can obtain more pronounced results and solidify our conclusion.