Is Party Affiliation Correlated with Voting Difficulty?

Datasci 203: Lab 1 Part 2

Team Royal Peacock - Darren Benemelis | Kara Christensen | Mian Haseeb | Varun Naidu

July 5th, 2023

Contents

1 Importance and Context

With a record-breaking turnout of voters, the 2020 United States election saw participation levels that had never been seen before. Despite the large number of voters, there were debates and worries about any potential voting challenges, particularly those related to party membership. To win the support of their respective bases, both Democrats and Republicans made intensive campaigning and mobilization efforts. Concerns about voting rights, accessibility, and integrity were at the center of the public conversation, with claims of voter suppression and attempts to influence election results by altering voting procedures. These discussions took place against a backdrop of a sharpening partisan division, with each side asserting that its supporters were encountering more substantial barriers to voting. The overall difficulty faced by Democrats and Republicans in the 2020 election could be attributed to factors such as: the security of voting machines, early voting periods, voter ID requirements, access to mail-in voting, and so on. This analysis aims to identify if there existed an imbalance in difficulties between voters affiliated to either parties, with the following research question:

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

The answer to this question could validate the claims and discussions concerning unequal barriers that aided the partisan division, and assess the larger consequences for democratic participation and electoral integrity in the United States.

2 Data and Methodology

Our analysis leverages data from the American National Election Studies (ANES) 2020 Time Series Study. This study features a fresh cross-sectional sample of U.S. eligible voters, and contains data from two interviews - pre-election and post-election.

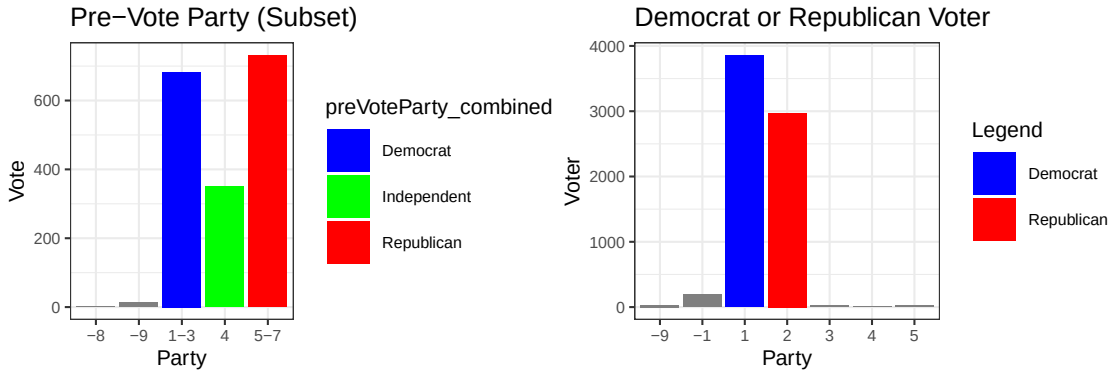
As our question is specifically focusing on voters, we begin by filtering out participants who are not voters. But who is a voter? Because we want to capture voting difficulty, we cannot just look at people who voted in the 2020 election as our entire population of voters. This perspective would miss data regarding voters who had so much difficulty voting that they were unable to cast a vote.

Therefore, for our research, a voter is someone who either voted in the 2020 presidential election, or who expressed they were either slightly likely, moderately likely, very likely, or extremely likely to vote. This captures everyone who expresses intent to vote as a 'voter', and assumes if they ended up not voting it was due to an extreme difficulty that kept them from doing so. After filtering to only capture participants who fall into our definition of a voter, we are left with 7,900 participants.

Now that we have our pool of 7,900 electorate who either voted or had at least a slight intent on voting, we want to determine who can be considered a Democrat or a Republican. According to the data gathered from this ANES sample size, 77.50% of respondents stated that they voted for either a Democratic or Republican candidate in the 2020 presidential election. However, it is worth noting that 19.50% of respondents did not participate in a post-election survey, and the remaining 3% either refused to answer or voted for a third party. To implement the concept of voter party distribution and get a more accurate reading, we decided to look into voters' pre-survey data, but only from the citizens who did not vote for Democrat or Republican in the 2020 presidential election. Given the challenges encountered during this election, the rationale behind filtering the group in this manner is to account for potential voters who had the intention to vote for one of the two primary parties but were either unable to do so or did not provide a follow-up response. For participants who did vote in 2020 presidential election, we assigned them to the party they voted for. The bar graph below on the left reveals a significant observation: a considerable proportion of voters included in the post-election survey of non-Democrat or Republican parties had originally intended to vote for one of the two major political parties.

Taking this information into consideration, we made a deliberate choice to incorporate the individuals who lacked post-election results indicating Democrat or Republican affiliation into their pre-survey responses. This approach proved highly effective in divulging the actual distribution of Democrat and Republican voters, enabling us to determine with better accuracy which party faced greater voting challenges. The

following distribution demonstrates this updated breakdown, with 95.37% of our sample now classified as either Democrat or Republican voters.

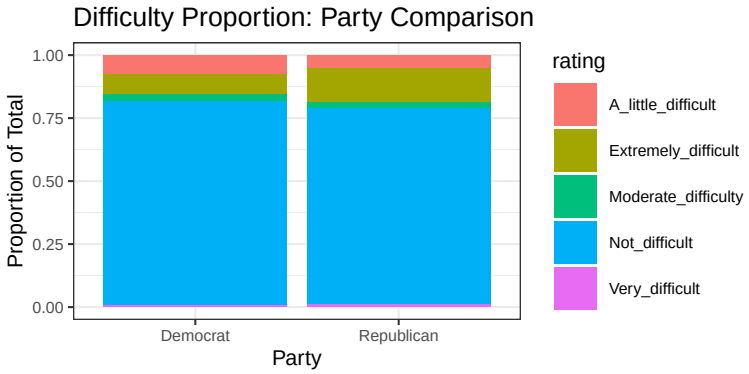


In order to come up with a measure for difficulty voting, this study used the survey question "How difficult was it for you to vote in this election?" measured in a Likert Scale of 1-5 or "not difficult" to "extremely difficult". Participants that voted in the election but didn't provide a response to this question were filtered out as no level of difficulty can be measured from those respondents.

As discussed previously, in order to capture those who had such difficulty voting that they could not vote, our collection of voters includes people who said it was likely they would vote in the pre survey, but did not actually vote. Our assumption is that those individuals had such difficulty voting that they were unable to vote. Therefore, we assigned all participants who indicated a likelihood of voting, but then did not vote, a difficulty rating of 5 - Extreme Difficulty, to capture that they experienced such a roadblock to voting that they were unable to follow through. It should be noted, however, that the true level of difficulty experienced by these individuals likely varied, but regardless of the magnitude of the difficulty experienced by these participants, it kept them from voting and can therefore be considered extreme.

```
##
##                          Democrat  Republican
##   A_little_difficult  0.074216440 0.052969502
##   Extremely_difficult 0.079243051 0.132958801
##   Moderate_difficulty 0.027202839 0.023274478
##   Not_difficult       0.813424009 0.781166399
##   Very_difficult      0.005913661 0.009630819
```

The table above provides a summary of the percentages of each difficulty measured by both democrats and republicans. It is notable that 3% more Democrats than Republicans reported no difficulty voting. In addition 5% more Republicans than Democrats reported that they experienced extreme difficulty voting. The visual on the below summarizes the proportion table into a graphic.



We formulate our null hypothesis by assuming that the Democratic voters faced more difficulty voting than the Republican voters in the 2020 election. The alternate hypothesis is the contradicting statement. To validate our null hypothesis, we are using the ***One-Directional Wilcoxon Rank-Sum Test (Hypothesis of Comparisons)*** with

'Party Affiliation' as the grouping variable and 'Difficulty Voting' as the target variable.

Ho: Democratic voters faced more difficulty voting than Republican voters in the 2020 election.

Ha: Republican voters faced more difficulty voting than Democratic voters in the 2020 election.

The data in both groups under the Party Affiliation variable are independent (as one vote/voter does not depend on another) and hence not paired, and the 'Difficulty Voting' variable of interest is ordinal. These two reasons lead us towards the Wilcoxon Rank-Sum Test as the most suitable. We will use the 'Hypothesis of Comparisons' as we are interested in comparing the difficulty faced by the two groups. We chose the one-directional version of this test as we are particularly interested in determining which group had more difficulty voting rather than just assessing if they faced equal difficulties or not.

Below, we evaluate the assumptions vis-a-vis the chosen test concerning the data at hand.

1. Ordinal scale - The data qualifies as the target variable 'Difficulty Voting' is a Likert Scale (1 - Not difficult at all, 2 - A little difficult, 3 - Moderately difficult, 4 - Very difficult, and 5 - Extremely difficult).
2. Independently and Identically Distributed (IID) - Each voter in either group should be independent of any other voter. Hence, the difficulty faced is also independent. However, some dependencies may exist as randomly selected subjects might belong to a single household. However, the sampling methodology document (https://electionstudies.org/wp-content/uploads/2022/08/anes_timeseries_2020_methodology_report.pdf) mentions that only one eligible household member was randomly selected to participate, and all participants received the same treatment at the initial recruiting and screening phase. Hence, this qualifies the variable as IID.

3 Results

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  to_test$Democrat and to_test$Republican
## W = 6064733, p-value = 1.264e-05
## alternative hypothesis: true location shift is less than 0
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  to_test$Republican and to_test$Democrat
## W = 6577183, p-value = 1.264e-05
## alternative hypothesis: true location shift is greater than 0
```

The test yielded a test statistic value (W) of 6064733 and a p-value of 0.00001264. Since the p-value is less than our significance level of 0.05, we reject the null hypothesis. The test indicates there is evidence that the amount of difficulty faced by Republican voters was greater than the amount of difficulty faced by Democratic voters.

4 Discussion

This study found evidence that Republican Voters experience more difficulties than Democrat Voters. This effect is practically significant because it touches on the core of our democracy. If, as the evidence suggests, Republican Voters are experiencing more difficulties voting than their Democrat counterparts, Democratic candidates have an unfair advantage, potentially altering the political landscape for years to come.

However' despite the extremely significant p-value, more research is necessary. There may be other factors at play beyond difficulty voting. For example, since we assigned extreme difficulty to anyone who said they were likely to vote but then didn't, maybe Republican voters are just more likely to say they are likely to vote, when they really aren't considering it. We can begin to explore if this is the case by doing more research into specific difficulties voting. Is one party more prone to having identification difficulties, or difficulties getting time off of work? This more specific testing would give us a better picture of the voting landscape, and allow us to pinpoint what difficulties people are actually facing, and who is actually facing them.