Influence of Party Affiliation on Reported Voting Difficulty

Kevin Lustig, Rebecca Nissan, Anuradha Passan, Giorgio Soggiu

2/29/2022

Contents

1 Importance and Context

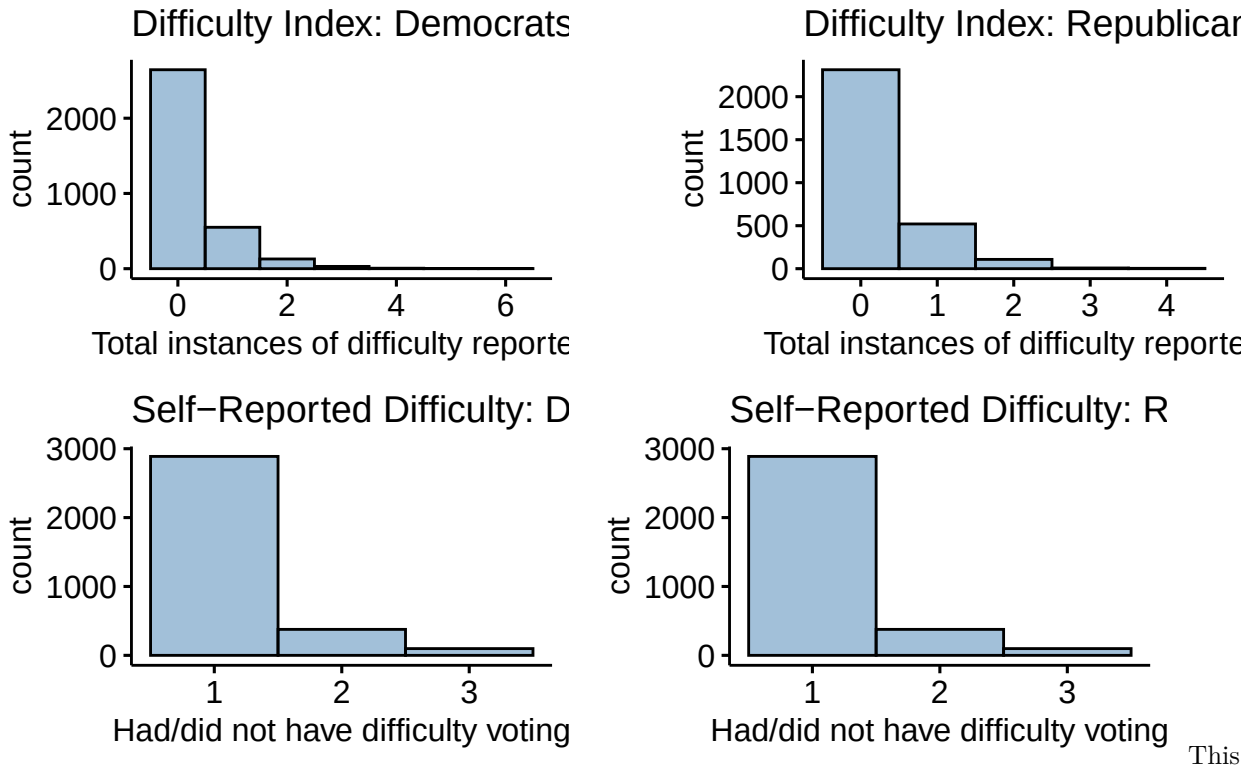
The topic of obstacles to voting has been the focus of considerable discussion in the United States in recent years, including notably before, during, and since the 2020 federal election. The debate over the facility with which Americans are able to vote has elicited claims both that voting in practice is too restrictive and challenging to allow for participation by all would-be voters and contrarily that it is so open and unrestricted that fraudulent voting can and does occur with limited recourse. Perhaps unsurprisingly, these assertions have largely arisen along partisan lines, and have been used as evidence in support of (often contradictory) changes to voting laws by elected officials and their supporters in both major political parties. Those affiliated with the Democratic Party are more likely to feel that it is too difficult to vote and support efforts to increase access to the polls, while Republican Party supporters are more likely to support restrictions on voting access. Naturally, the partisan nature of this debate means that voting behavior itself has been impacted, for example with Democratic voters significantly more likely to vote early and by mail than Republican voters in 2020, both methods widely discussed as fraught with error and likely to be fraudulent by prominent members of the Republican party.

Naturally, this raises the question of whether the actual experience of difficulty voting is related as well to party affiliation. Perhaps attitudes toward measures to loosen or tighten voting regulations are driven by real or perceived differences in personal experience with voting difficulty. This analysis therefore attempts to examine the linkage, if any, between political party identification and difficulty voting. We seek to address the following research question:

Did Democrats or Republicans have more difficulty voting in the 2020 election?

This question has the potential to illuminate important aspects of the debate on measures to change voting laws. Namely, is it possible that Democrat-associated policies to ease voting regulations are based on an actual increased level of difficulty voting for Democrats as compared to Republicans (and likewise for Republican efforts to further restrict voting?) What is the scale of the actual voting difficulty for members of each party, and who would be most affected and in what way by changes meant to address the current ease or difficulty of voting?

2 Data and Methodology



This study will use data from the 2020 American National Election Studies (ANES), specifically their 2020 Time Series Study, which includes survey responses for a representitive sample of voters from both before and after the 2020 election.

In order to answer the research question *Did Democrats or Republicans have more difficulty voting in the 2020 election?*, three conceptualization questions need to be answered: 1. Who or what is a **voter**? 2. Who is a Republican and who is a Democrat; i.e. How do we define **party affiliation**? 3. What is **difficulty in voting**?

For this study, the below definitions and assumptions have been applied:

Voter We define a "voter" as anyone in the ANES survey sample who registered to vote - regardless of whether or not they voted in the 2020 election. Since this study examines **difficulty** voting, it is important to consider the subset of people who implied some intention to vote (by registering) but ultimately did not vote - perhaps because of some difficulty. In order to determine who is a "registered" voter, we use a post-vote status variable because it is possible some people registered after the pre-election survey but before the election. Eliminating anyone for whom we did not have voter status info as well as those who were not registered by the time of the post-election survey drops 1,227 individuals from our sample, leaving 7,053 "voters."

Party Affiliation In order to determine party affiliation, we use a self-reported party identification variable. Since this study focuses specifically on Democrats and Republicans (and does not break down results by the strength of one's affiliation to their party), we classify anyone who reported leaning Democrat as a Democrat and anyone who reported leaning Republican as a Republican. The 719 individuals who self-identified as Independents with no party lean are classified as Independents; they remain in our sample for exploration purposes, but the difficulty analysis focuses only on Democrats and Republicans. Another possible way to define somebody's party is based on who they voted for: but in the 2020 election especially, many people did not vote for their party's candidate. Therefore, we define party affiliation solely based on the self-reported variable.

Difficulty in Voting Given the subjective nature of the term "difficulty", there are various ways difficulty in voting could be conceptualized. The first important distinction we need to make among registered voters is whether or not they actually voted in the 2020 election. We have different information available for these two groups.

For **people who voted,** the ANES survey specifically whether they had encountered specific difficulties in trying to vote (e.g. long lines, failing to provide proper registration papers, etc). Participants could indicate more than one type of difficulty, if they encountered multiple. For the group who voted, the survey also asked directly how much difficulty they encountered, if any. For **people who did not vote,** the ANES survey asked for the main reason they did not vote, and respondents could choose one form of difficulty.

1. **Difficulty Index:** This is a synthetic index that was created to measure how many different types of obstacles registered voters faced in trying to vote. We assume a) that having at least one reported type of difficulty (rather than none) is a valid measure of whether somebody experienced any difficulty and b) that reporting a greater number of types of difficulty implies that somebody had more difficulty. This assumption is based on the premise that if each obstacle is equally challenging, then facing a greater number of obstacles implies that one faced an overall greater challenge when trying to vote.

To generate the difficulty index for **people who voted,** the research team created separate binary variables for each type of difficulty: registration issues, not having the correct form of ID, difficulties in receiving absentee ballots, confusion about the ballot or machine, difficulties in getting to the polling place (i.e. transportation issues), long wait times, difficulties with work schedules, bad weather, issues mailing ballots, and other. After creating the binary variables, the sum of them all was taken to create the difficulty index for each voter. There were no weights applied to this index because based on the data given, there was no way to perceive qualitative differences between types of difficulty.

To generate an index for **people who did not vote,** the research team followed a similar approach: we generated a binary variable for each form of difficulty listed under the "main reason you did not vote" and checked the sum of all these variables for each respondent (although it is important to note that the maximum index value for this group is 1, because they only selected one main reason). The reasons for not voting that we considered forms of difficulty were: lacking the correct form of ID, being sick or disabled, bad weather, transportation, long lines, not allowed to vote at the polls, requested but did not receive absentee ballot, did not know where to vote, and other. We excluded reasons for not voting that were not due to difficulty (e.g. "Out of town").

2. **Ordinal Difficulty Variable:**

Since the maximum possible index value for non-voters is 1 (they could only report one reason), the index value has a few shortcomings: first, the maximum value is not the same for voters versus nonvoters, which makes it hard to compare levels of difficulty across these two groups (although this is somewhat mitigated by the fact that nearly 80% of people who voted and reported any type of difficulty only reported one type). The second challenge with the index is that it does not account for the fact that non-voters were presumably the *most* impacted by the experience of difficulty - so much so that they did not complete the voting process.

Therefore, we created an additional difficulty variable which describes registered voters ordinally based both on whether they experienced difficulty and whether they were able to cast a vote. A registered voter who did not vote is considered to have faced difficulty if they reported a difficulty-related reason for not voting (we use the same reasons as we did when creating the index). A registered voter who did vote is considered to have faced difficulty if they responded to the question of "how difficult was it to vote" by choosing "a little difficult," "moderately difficult," "very difficult," or "extremely difficult." Note that we do not differentiate here between scales of difficulty, in part because there is no objective way to numerically define the distance between each level of difficulty on the Likert scale.

This variable has three categories: *Category 1:* Those who did not experience difficulty (regardless of whether they voted). *Category 2:* Those who experienced difficulty and did vote. *Category 3:* Those who experienced difficulty and did not vote.

3 Analysis & Results

Test 1 Our analysis begins with an examination of the voter-difficulty index across all registered voters surveyed. We test the difference, if any, between the number of reported types of difficulty encountered by Republican- and Democrat-identified voters:

```
t.test(diff_index~party,alternative = "two.sided",data = data_fin_reduced)
```

```
##
##  Welch Two Sample t-test
##
## data:  diff_index by party
## t = 1.2208, df = 6313.7, p-value = 0.2222
## alternative hypothesis: true difference in means between group D and group R is not equal to 0
## 95 percent confidence interval:
##  -0.01085447  0.04669076
## sample estimates:
## mean in group D mean in group R
##       0.2807018       0.2627836
```

The result of this test provides no evidence that there is a significant difference in reported experience of voting difficulty based on political identification (t=1.2208,p=0.2222). This is unsurprising because Democrats reported a mean experience of .2807 types of voting difficulty, while Republicans reported a mean of .2628 types That is, voters of both parties collectively experienced, on average, far less than even one incident of difficulty; in fact, our ordinal variable makes it easy to see that 86.56% of voters reported no difficulty at all (Table 1).

Table 1: Self-Reported Difficulty and Party

	Did not report difficulty	Reported difficulty
Democrat	0.457	0.075
Republican	0.408	0.059

Table 2: Self-Reported Difficulty and Voting Ability

	Did not report difficulty	Reported difficulty
Didn't Vote	0.000	0.034
Voted	0.866	0.101

Test 2 Using the ordinal difficulty variable, we test whether there are differences by party in the ability to complete the process of casting a vote among those who experienced difficulty (people in categories 2 and 3). In other words, we are defining "not voting due to difficulty" as a more severe form of difficulty than "facing difficulty but still managing to vote," and checking whether Republicans or Democrats fall closer to category 3 - the most severe case.

```
wilcox.test(diff_subj2~party,alternative = "two.sided",data = data_fin_difficulty, conf.int=TRUE )
```

```
##
##  Wilcoxon rank sum test with continuity correction
```

```
##
## data:  diff_subj2 by party
## W = 79601, p-value = 0.0005455
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -7.103201e-05 -6.340510e-06
## sample estimates:
## difference in location
##          -6.410914e-05
```

In this result, we find evidence of a difference in the distribution of voting difficulty between the two parties, with a small but significant decrease in the category value for Republican voters as compared to Democratic voters (p=0.01746). Following the principle that inability to vote due to difficulty indicates greater voting difficulty, this suggests that Democratic voters had a significantly harder time voting than Republican voters. Based on this result, we reject the null hypothesis that there is no difference in voting difficulty between the political groups.

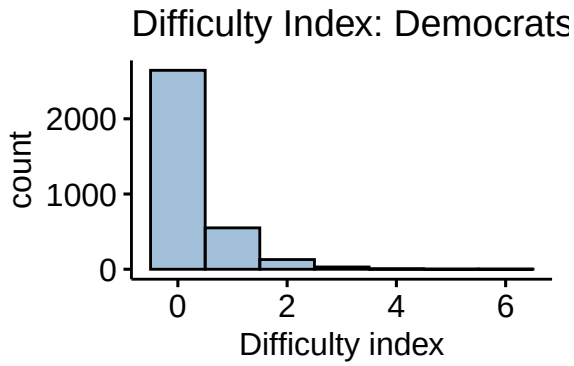
It is important to note some key limitations of this approach. Factors that are not considered in this analysis are quite likely to have an impact on the experience of voting difficulty, such as method of voting and voting location. Further analysis that attempts to analyze differences across methods and locations would be a valuable addition to these findings. Moreover, it is critical to note that the method of questioning voters in the ANES survey asks only for self-reported and therefore subjective descriptors of difficulty. It does not attempt to validate or weight these incidents. Thus, this study does not attempt to analyze an objective measure of how difficult the voting experience was for any voter. In reality, two voters may regard the same experience as more or less difficult, and may choose to behave differently at the polls because of that perception. Different data than that found in the ANES dataset will be needed to form a clearer and less subjective picture of the nature of obstacles voters face when attempting to cast a vote.

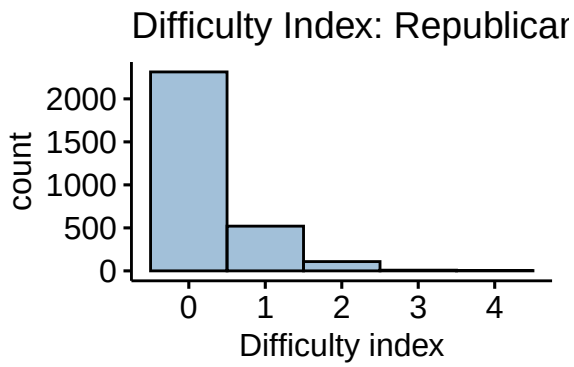
4 Discussion

Our analysis finds, first, that the reported incidence of voting difficulty is quite low for all voters. Our analysis using the difficulty index showed no significant difference in difficulty between Republican and Democrat voters; this is likely at least in part due to the fact that neither party experienced much difficulty voting. However, in examining the subset of voters who did self-report difficulty, we propose that it is meaningful to know how many were not able to ultimately cast a ballot due to that difficulty. With regard to this measurement, we do find that Democrats who experienced difficulty at the polls were slightly but significantly more likely to be unable to vote than Republicans.

This study suggests a number of interesting questions that merit further research. Are there significant differences in the kind or amount of difficulty faced by just those voters who did experience difficulty? Do Democrat and Republican voters report or experience different kinds of difficulty at different rates? Does the incidence of difficulty voting change from election to election? We believe our results are most interesting in the context of these and other answers in helping to form an evidence base for ongoing debates about voting access. The right mix of policy solutions for changing (or preserving) current voting practices depends on an understanding of the real ease or difficulty of voting, and particularly whether some groups find voting to be more or less accessible to them than other groups do. We hope that our findings can help to point the way toward additional answers that will contribute to effective and evidence-based policy-making that ensures fair access to the polls for all voters with a minimum of obstacles.

4.0.1 Working code





Code: Wilcoxon #### coding style 1

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  sub_dem$diff_index and sub_rep$diff_index
## W = 4963227, p-value = 0.9654
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -5.518342e-05  4.393156e-05
## sample estimates:
## difference in location
##          -3.712453e-05
```

4.0.1.1 coding style 2

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  diff_index by party
## W = 4963227, p-value = 0.9654
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -5.518342e-05  4.393156e-05
## sample estimates:
## difference in location
##          -3.712453e-05
```

4.0.2 Code: t-test

```
##
##  F test to compare two variances
##
## data:  diff_index by party
## F = 1.2802, num df = 3362, denom df = 2952, p-value = 5.31e-12
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.193665 1.372838
## sample estimates:
## ratio of variances
##           1.280224
```

- ratio of variances different than one.
- lower and upper bounds (from the confidence interval) are higher than 1 -> variances are really different -> not sure we can use the t-test (old version) then, nevertheless we can use the Welch's version of the t-test (apparently, default test in R)

```
##
##  Welch Two Sample t-test
##
## data:  diff_index by party
## t = 1.2208, df = 6313.7, p-value = 0.2222
## alternative hypothesis: true difference in means between group D and group R is not equal to 0
## 95 percent confidence interval:
##  -0.01085447  0.04669076
## sample estimates:
## mean in group D mean in group R
##       0.2807018       0.2627836
```

4.1 Considering diff_subj

4.1.1 Which test to use?

paired/unpaired? - we are using two random variables: diff_subj affiliated to Republicans and diff_index associated to Democrats. - Republicans and Democrats are two different populations -> unpaired

parametric/non parametric? - data distrbution is not normal -> non parametric test

distributions: -distri of diff_index_rep and diff_index_dem are really similar

data type? - diff_subj is binary /not metric -> Wilcoxon again -> independent sample t test?

4.1.2 Hypothesis

null hypothesis of comparisons H0:P(diff_subj_dem<diff_subj_dem) = P(diff_subj_dem>diff_subj_dem)

4.1.3 Code Wilcoxon

```
##
##  Wilcoxon rank sum test with continuity correction
##
```

```
## data:  diff_subj by party
## W = 4955678, p-value = 0.7298
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -4.145534e-05  4.854247e-05
## sample estimates:
## difference in location
##          -5.152669e-05
```