

Influence of Party Affiliation on Reported Voting Difficulty

Kevin Lustig, Rebecca Nissan, Anuradha Passan, Giorgio Soggiu

2/29/2022

Contents

1	Importance and Context	1
2	Data and Methodology	1
3	Analysis & Results	3
4	Discussion	4
5	Appendix 1: Tables and Figures	5
6	Appendix 2: R Code	6

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, unsurprisingly, been divided along partisan lines. Democrats are more likely to feel that voting is too restrictive to allow for participation by all would-be voters, while Republicans generally assert that voting is so open and unrestricted that fraudulent voting can and does occur. Elected officials and their supporters in both major parties use claims about difficulty to guide voting laws; Democrats tend to support efforts to increase access to the polls while Republicans tend to support restrictions on voting access.

This raises the question of whether the actual experience of difficulty voting is related 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 seeks 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 about 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 uses data from the 2020 American National Election Studies (ANES), specifically their 2020 Time Series Study, which includes survey responses for a random sample of US Citizens from both before and after the 2020 election. The full dataset includes 8,280 observations. Our study focuses only on “voters” who are categorized as either Democrats or Republicans. We operationalize the concept of a “voter” and define party affiliation as follows:

We consider a “*voter*” to be 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.”

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. 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. After dropping the 719 individuals who self-identified as Independent and the 18 people who did not provide a party affiliation, we are left with a total of 6316 observations.

We next operationalize the concept of “*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. The ANES survey provides different information for both these groups.

For **respondents who voted**, the ANES survey specifically asks 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 (up to nine), if they encountered multiple. For the respondents who voted, the survey also asked directly how much difficulty they encountered, if any (e.g. none, little, moderate, extreme). For **people who did not vote**, the ANES survey asked for the two main reason they did not vote, meaning that respondents could choose only two potential forms of difficulty¹, compared to the nine for voters who did end up voting.

Given this context it was decided to look at two variables that could represent difficulty: difficulty ranking and difficulty category.

Difficulty ranking

Only respondents who did vote in the 2020 election were posed the question “How difficult was it to vote?” with the following possible responses (and given value in the ANES survey): not difficult at all (1), a little difficult (2), moderately difficult (3), very difficult (4), extremely difficult (5).²

Difficulty Category

As the difficulty ranking variable accounts only for respondents classified as voters, this variable will not be able to give insight on difficulty faced by voters (as per the definition mentioned above) who tried to vote, but encountered so much difficulty that in the end they were not able to vote. In order to incorporate this, a synthetic variable was created which would encompass 3 categories: (1) voters who did not experience difficulty when trying to vote (regardless of whether they ended up voting or not), (2) voters who experienced difficulty and still managed to vote, and (3) voters who experienced difficulty and were not able to vote.

Category 2 consists of voters (those who voted) who responded *a little difficult* until *extremely difficult* in the difficulty ranking variable. For those who voted and selected *not difficult at all*, they were categorized into group 1.

In the ANES survey, respondents who did not end up voting were asked to identify the main reason why.³ For the respondents who were asked this question the possible responses included a mix of reasons as to why a voter was not interested in voting and a difficulty that kept the voter from voting.⁴ For this category, only the following responses that indicated a difficulty preventing the person from voting (and given value in the ANES survey: (6) “I did not have the correct form of identification”, (8) “sick or disabled”, (9) “transportation”, (10) “bad weather”, (11) “the line at the polls was too long”, (12) “I was not allowed to vote at the polls, even though I tried”, (13) “I requested but did not receive an absentee ballot”, (14) “I did not know where to vote”, (16) “Other”. These mentioned responses were allocated into category 3 as they experienced difficulties that played a role in preventing them from voting (compared to those who chose not to vote because they forgot or weren’t interested).⁵ Those who chose an ineligible category 3 response, would be moved to category 1 - a voter who did not experience difficulty voting.

Proposed Testing We propose to conduct two hypothesis tests - one for each difficulty variable. For the data ranking variable we propose a two sample t-test, and for the difficulty category variable we propose a Wilcoxon Rank Sum Test. Both the difficulty variables are independent and identically distributed due to the data collection and transformation protocols that were followed by those conducting the ANES survey - satisfying one of the three conditions for both tests. Given that the difficulty rankings seem to have an understood equal distance between each value (no unexpected neutral jumps), this variable could qualify

¹Variable numbers: V202123 and V202124

²In the codebook this is variable number V202119; responses such as “no post election data” or “no post election interview would have been filtered out beforehand

³Variable number: V202123

⁴In the codebook this is variable number V202119; responses such as “no post election data” or “no post election interview would have been filtered out beforehand

⁵A follow up question was asked to those who did not vote to provide a second reason as to why they did not vote. Given that this category aims to weed out those who experienced legitimate difficulty in trying to vote, it was sufficient to pull from the main question asked. In the cases where a respondent would answer the follow up question to give an additional reason as to why they did not vote, if they selected an eligible category 3 response the first time, then that respondent would already have been included in category 3. If the first time they selected an ineligible category 3 response, and then selected an eligible category 3 response, in the end they would count as someone who in the end did have enough interest to put in enough effort to vote. The ineligible category 3 responses include those such as “did not like candidates” or “I forgot”, etc.

as a metric variable, satisfying another condition for a two sampled t-test. The difficulty category variable satisfies the possibility of having an ordinal variable for a Wilcoxon rank sum test. Regarding the normality condition for conducting a t-test, while the histograms for the difficulty ranking variable for both parties do not seem to show a normal distribution (Appendix: Figures 1 and 2), due to the large sample size ($n = 6316$), it is likely the central limit theorem (CLT) can be applied here. The CLT states that the distribution of the sample means converge to a normal distribution as the sample size gets larger. Given the context, it is likely this can be applied here and satisfy the normality condition - and the final condition needed to conduct a t-test. The final condition to conduct a Wilcoxon rank sum test is that the distribution and spread of the two samples (Democratic and Republican) be similar. Figures 3 and 4 (Appendix) clearly show a similar shape and distribution.

3 Analysis & Results

Ranked voting difficulty

Our analysis begins with an examination of the voting difficulty index across all registered voters surveyed. We test the difference, if any, between the level of difficulty encountered by Republican- and Democrat-identified voters:

```
t.test(diff_ranked~party,data = data_fin_reduced)
```

The result of this test provides no evidence that there is a significant difference in reported experience of voting difficulty based on political identification ($t=1.2604, p=0.2076$). This is unsurprising, as discussed previously; Democrats reported a mean level of voting difficulty of 1.176151, while Republicans reported a mean of 1.157778 (1 = “Not difficult at all”). That is, voters of both parties collectively experienced, on average, very little to no difficulty; in fact, 86.56% of voters reported no difficulty at all (Appendix: Table 1). We note again that the tested variable also does not incorporate the experience of those who were not able to cast a vote after experiencing difficulty, as these voters did not provide a ranked difficulty survey response.

Categorical voting difficulty and voting status

Therefore, we focus our analysis on only that subset of attempted voters who experienced difficulty in order to detect any difference the experience of difficulty for Republicans and Democrats. To do so, we test our secondary variable categorizing voters by their report of having or not having difficulty and voting status in categories 2 (difficulty; voted) and 3 (difficulty; didn’t vote).

```
wilcox.test(diff_categorical~party,data = data_fin_difficulty, conf.int=TRUE )
```

In this result, we find evidence of a difference in the distribution of difficulty and voting status between the two parties, with a small but significant lower shift in the category value for Republican voters as compared to Democratic voters ($p=0.0005455$). Because it excludes those who did not have difficulty voting, this test is not a true test of level of difficulty, but rather of the distribution of those that did or did not vote after having difficulty. It therefore indicates that Democratic voters were significantly less likely to complete the act of voting after experiencing difficulty. Following the principle that *inability to vote* due to difficulty indicates the most severe form of voting difficulty, we propose that this provides some indication 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 ranked level of self-reported difficulty showed no significant difference between Republican and Democrat voters; this is primarily due to the fact that neither party group as a whole 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 not complete the act of voting 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.

5 Appendix 1: Tables and Figures

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

Figure 1: Difficulty Ranking of Voters Who Voted, Democrats

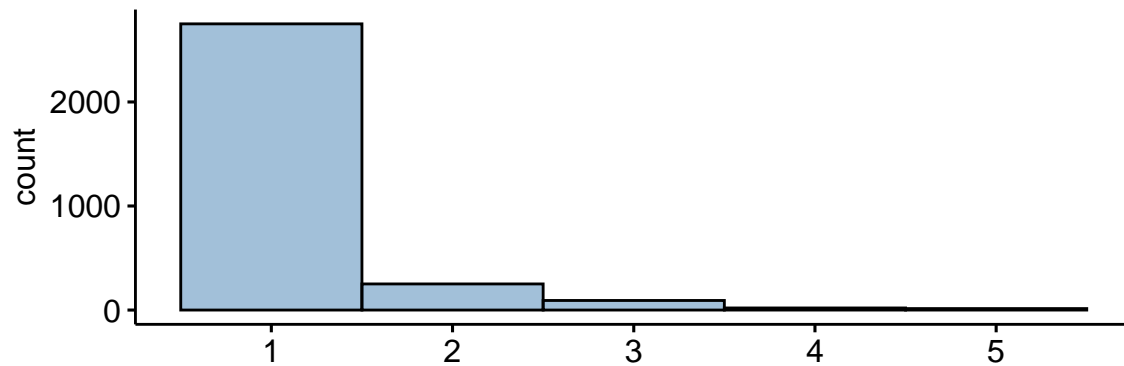


Figure 2: Difficulty Ranking of Voters Who Voted, Republicans

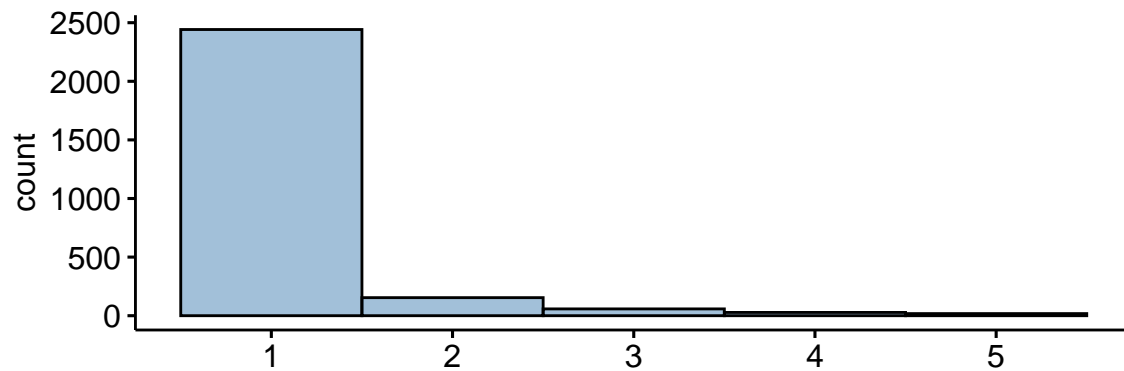


Figure 3: Difficulty Categories, Democrat Voters

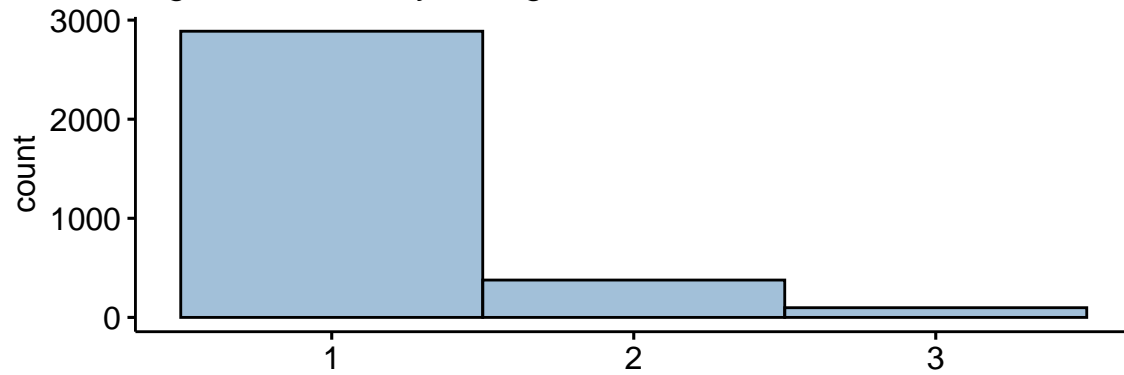
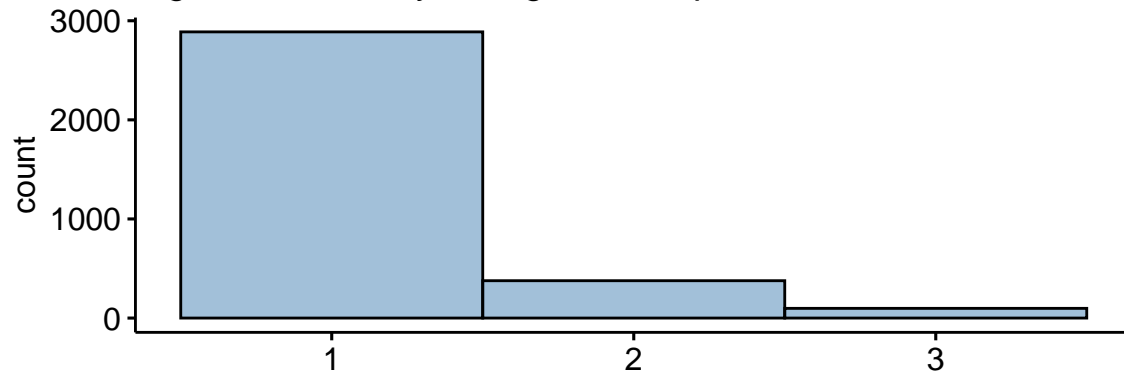


Figure 4: Difficulty Categories, Republican



6 Appendix 2: R Code

```
knitr::opts_chunk$set(fig.width=6, fig.height=2.5)

options(tinytex.verbose = TRUE)
knitr::opts_chunk$set(echo=FALSE, message=FALSE)

library(tidyverse)
library(magrittr)
library(knitr)
library(ggpubr)
library(ggplot2)
source("data_wrangling.r")
data_fin_reduced <- data_fin %>% select(party, diff_ranked, diff_categorical)
data_fin_reduced <- data_fin_reduced %>% filter(data_fin_reduced$party == 'D' | data_fin_reduced$party == 'R')

# splitting the dataset in two
sub_dem <- data_fin_reduced[data_fin_reduced$party == 'D', ]
sub_rep <- data_fin_reduced[data_fin_reduced$party == 'R', ]
```

```

# focus on only voters who self-reported difficulty either after casting a vote or being unable to
data_fin_difficulty = data_fin_reduced %>% filter(diff_categorical > 1)

t.test(diff_ranked~party,data = data_fin_reduced)
wilcox.test(diff_categorical~party,data = data_fin_difficulty, conf.int=TRUE )
party_summary_table <- data_fin_reduced %>%
  mutate(
    party = case_when(
      party == "D" ~ 'Democrat',
      party == "R" ~ 'Republican'),
    difficulty = case_when(
      diff_categorical > 1 ~ 'Reported difficulty',
      diff_categorical == 1 ~ 'Did not report difficulty')) %%%
prop.table(
  table(
    party,
    difficulty))

voted_summary_table <- data_fin_reduced %>%
  mutate(
    vote_status = case_when(
      diff_categorical < 3 ~ 'Voted',
      diff_categorical == 3 ~ 'Didn\'t Vote'),
    difficulty = case_when(
      diff_categorical > 1 ~ 'Reported difficulty',
      diff_categorical == 1 ~ 'Did not report difficulty')) %%%
prop.table(
  table(
    vote_status,
    difficulty))

kable(
  party_summary_table,
  digits = 3,
  caption = 'Self-Reported Difficulty and Party',
  booktabs = TRUE,
  position = "float_left"
)

kable(
  voted_summary_table,
  digits = 3,
  caption = 'Self-Reported Difficulty and Voting Ability',
  booktabs = TRUE,
  position = "float_right"
)

# diff_ranked distribution for Democrats
gghistogram(sub_dem$diff_ranked, bins = 5, fill= 'steelblue', xlab='', main = 'Figure 1: Difficulty Ran
# diff_ranked distribution for Republicans
gghistogram(sub_rep$diff_ranked, bins = 5, fill= 'steelblue', xlab='', main = 'Figure 2: Difficulty Ran
# diff_categorical distribution for Democrats
gghistogram(sub_dem$diff_categorical, bins = 3, fill= 'steelblue', xlab='', main = 'Figure 3: Difficult
# diff_categorical distribution for Republicans
gghistogram(sub_rep$diff_categorical, bins = 3, fill= 'steelblue', xlab='', main = 'Figure 4: Difficult

```