

Difficulty Voting in the 2020 Presidential Election

Team03: Savita Chari, Denny Lehman, Tymon Silva

10/16/2021

```
## # A tibble: 5,649 x 8
##   Case_Id Sample_Type Eligibility_Conce~ Eligibility_Outco~ Validation_Study_0~
##   <dbl>      <dbl>      <dbl>          <dbl>          <dbl>
## 1 200039         2         0             -2             -2
## 2 200046         2         0             -2             -2
## 3 200060         2         0             -2             -2
## 4 200084         2         0             -2             -2
## 5 200114         2         1              4              3
## 6 200121         2         0             -2             -2
## 7 200138         2         0             -2             -2
## 8 200152         2         1              4              3
## 9 200169         2         0             -2             -2
## 10 200190         2         0             -2             -2
## # ... with 5,639 more rows, and 3 more variables: Did_Vote <dbl>,
## #   Political_Party <chr>, Voting_Level_Difficulty_Code <dbl>
```

1 Introduction

The 2020 election, held on November 3, recorded the highest voter turnout in our election history as well as the highest number of mail-in voters.¹ And, although more participation is generally seen as a good thing, the 2020 voting experience left many of us concerned and troubled, casting a shadow on the safety, reliability, and structure of the voting process. Videos of citizens showing up armed with weapons at the voting booths, stories of locals intimidating and harassing voters standing in line, and reports of issues with mail-in ballot boxes, seemed all too common this last general election. Is this just the media hyping the drama around the voting process or did certain groups of voters really experience increased difficulty voting this last election? The situation has spurred many discussions around the voting process and how it will look in the future.

In this report, we answer a basic, but relevant question:

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

We focus our analysis not only around the frequency at which Democrats and Republicans experienced difficulties voting but the level of difficulty expressed among these two groups as well. In answering this question, we can contribute to the conversation and help guide strategies that reform and/or improve the voting process. We must protect our voters to ensure a fair election takes place.

2 Data and Methodology

Our data comes from the 2020 Time Series Study performed by the survey firm, Westat, and generated by the American National Election Studies (ANES). The data consists of responses from both a pre and post election survey, in which participants were interviewed prior to election date and contacted again to

¹<https://www.newsweek.com/election-2020-voter-turnout-67-percent-highest-120-years-1544552>,
<https://www.newsweek.com/highest-voter-turnouts-history-us-elections-1545235>

Table 1: Distribution of Political Party

Party	Frequency	Proportion
Democrat	3031	0.5365551
Republican	2618	0.4634449

capture their experiences and opinions shortly after the election had concluded. Addresses across the US were randomly drawn from a USPS computerized delivery sequence file and the selected addresses were sent a series of letters to recruit a household member. All mailers offered financial incentives to participate in the survey. The study involved three modes of response: web -only, mixed web, and mixed video (live video interviewing via Zoom). The pre-election survey was conducted between August 18, 2020 and November 3, 2020. The study re-interviewed the same respondents in the post-election survey between November 8, 2020 and January 4, 2021. The overall response rate for the pre-election survey and the post-election survey was 40.9% and 87.9%, respectively.

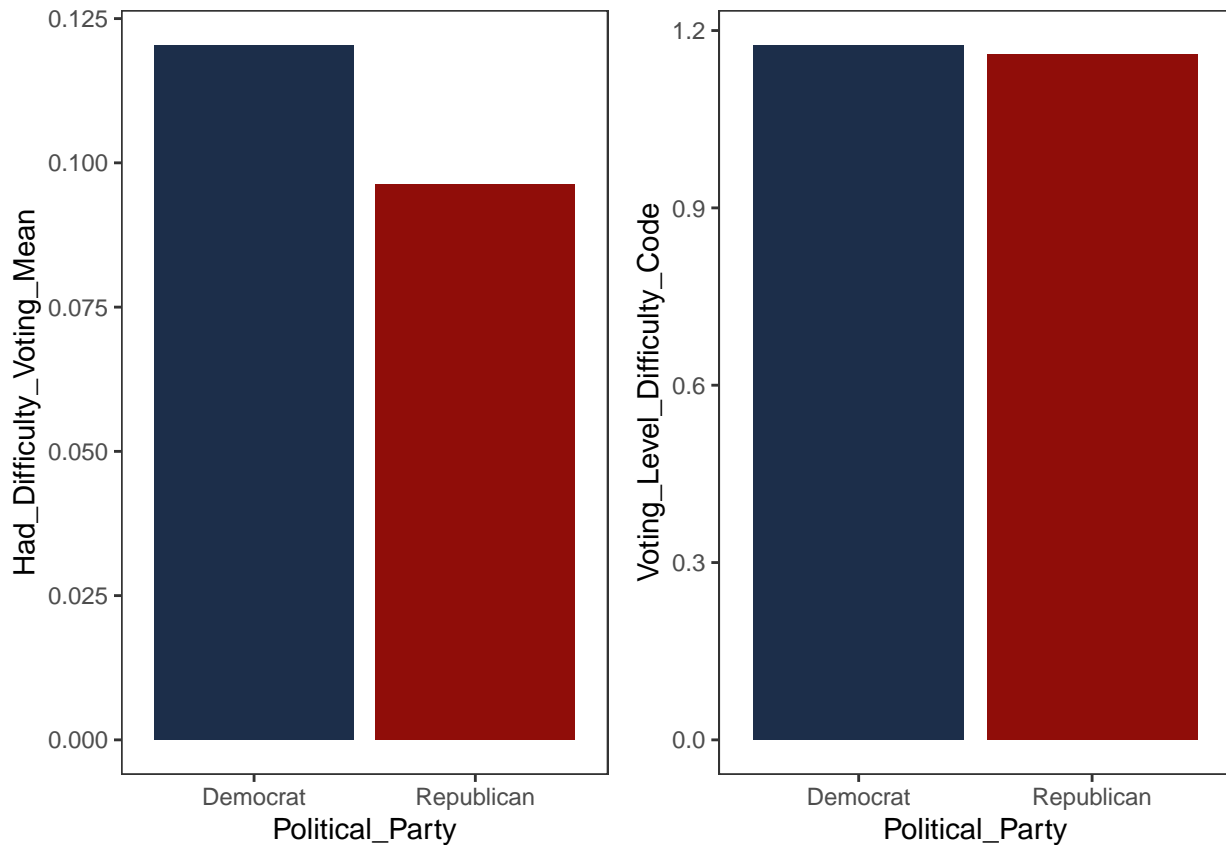
Additionally, all sample members who completed the post-election survey from the ANES 2016 Time Series Study were invited to participate in the 2020 survey online. This survey data was merged using the case ID variable to join the two datasets.

In order to operationalize the concept of difficulty voting, we first identify the Democratic and Republican voters. Per guidance from an article written by political scientists², we group voters into political parties by the direction they “lean.” In other words, a voter that indicates they “lean democratic” is classified as a Democrat and vice versa. The tables below describe these distributions. ## Add concern with sample universe for difficulty question ### Insert Distribution Table Here

Table 1 shows the frequencies of Democrats and Republicans in our dataset. And, Table 1 shows the proportion of each group, which is roughly 54% Democrat and 46% Republican.

The key variable for our analyses is the question pertaining to how difficult it was for the respondent to vote. Responses were collected on a 5-point Likert scale that ranged from 1 (“Not difficult at all”) to 5 (“Extremely difficult”). In order to operationalize the frequency in which Democrats had experienced difficulty voting compared to Republicans that experienced difficulty, we grouped voters in a binary fashion; as either had difficulty voting or had no difficulty at all and compared their means.

²<https://www.sciencedirect.com/journal/electoral-studies>



First, let's determine if one party is more likely to experience difficulty voting than the other. To test this, we can use a two-sample t-test. The null hypothesis for this t-test can be expressed as follows:

Null Hypothesis: *The probability that a Democratic voter experiences some difficulty during their voting process is equal to the probability that a Republican voter experiences some difficulty during their voting process.*

For this t-test, the following assumptions must be true:

1. **i.i.d. data** First, data must be generated via an iid process (Independent and Identically Distributed). The ANES 2020 survey merged in data from a panel, which was comprised of participants in the ANES 2016 survey. This is definitely a concern when validating that the data to be identically distributed, since the year in which these surveys were administered differs greatly. However, for the purposes of this study, we will assume the impact of this possible violation is negligible.
2. **Metric scale** Even though our raw data includes ordinal variables, we have transformed them to binary form. A binary variable qualifies as metric as there is only a single interval, which goes from zero to 1.
3. **Sufficient normality**

3 Results

```
t.test(ANES_mean$Voting_Level_Difficulty_Code ~ ANES_mean$Political_Party)

test <- wilcox.test(ANES$Voting_Level_Difficulty_Code ~ ANES$Political_Party)
```

The results of the Wicoxon Rank-Sum test conclude that the probability of the level of difficulty voting for Democrats is less the the level of difficulty voting for Republicans IS NOT equal to the probability of the

level of difficulty voting for Democrats is greater than the the level of difficulty voting for Republicans ($W = 4058825$, $p\text{-value} = 0.005804$). In practical terms, this means there was a statistically significant difference in the experienced level of difficulty voting between the two political parties.