Voting Difficulty in the 2020 U.S. Presidential Election

Brodie Deb, Ethan Moody, and Eugene Oon

June 2022

Contents

1 Overview: Importance and Context

The 2020 U.S. presidential election shattered many norms. It was just the second time in U.S. history when voters cast their ballots during a major pandemic. It also saw the highest turnout rate since 1900, with over 150 million votes tallied in the final count.[1] For the first time in three presidential cycles, the incumbent president was defeated after a single term. By all accounts, the election made history.

Among the most consequential forces in this election was the mail-in ballot, and both major political parties were quick to offer their judgments.[2] Republicans argued that the widespread use of absentee and early election voting represented a potential threat to election integrity. Democrats claimed that a surge in these voting methods exposed issues and obstacles with traditional voting practices in elections. Both groups decried foul play from elected officials, government workers, and polling workers, claiming that voters had experienced undue difficulties (and pressures) with trying to make their voices heard.

We seek to explore this point further. Our aim is to add clarity around voter difficulty in the 2020 election using data as our guide. In doing so, we intend to investigate the following research question: *Did Democratic voters or Republican voters experience more difficulty voting in the 2020 election?* We believe an analysis of this question offers broad, non-partisan benefits. To that end, we hope our review encourages unity of resolve to tackle voter difficulty issues – in whatever form they take – rather than further stokes the flames of political division.

2 Operationalization: Data and Methodology

To analyze this research question, we use data from the "ANES 2020 Time Series Study" – the latest in a series of survey-based datasets published by the American National Election Studies (ANES) to measure public sentiment in U.S. presidential elections.[3] Survey respondents include a mix of individuals who participated in the ANES 2016 survey, as well as a new sample of participants drawn from a cross-section of the U.S. population. A total of 8,280 responses were collected via pre-election interviews between August and November 2020, while 7,449 responses were collected via post-election interviews between November and December 2020.

To supplement each study, ANES provides a comprehensive dataset and "codebook" (containing background details on the surveyed sample and survey questions) for public review. We make extensive use of this codebook to define the following key terms from our research question:

- **Voter:** We define "voter" as *any survey respondent who was at least 18 years of age* at the time of the 2020 election, and use "respondent age" (V201507x) to classify respondents as voters or non-voters. We do *not* restrict our definition to just those respondents who actually cast a ballot in the 2020 election, because we acknowledge that difficulty experienced in the voting process could have prevented an individual from voting.
- **Democrat or Republican:** We define "Democrat" as *any respondent who identified as a Democrat pre-election* and "Republican" as *any respondent who identified as a Republican pre-election.* These definitions include Independents who "lean Democrat" or "lean Republican", as research by Petrocik (2009) suggests that Independent-leaners behave like partisan voters [4]. We use pre-election "party ID" (V201231x) to classify respondents as Democrat or Republican, since it represents a composite of several party-related measures (e.g., V201228z, V201229, V201230, V201156, V201018 and V201157). Notably, *Independents with no leaning* are left out of scope.
- **Difficulty:** We define "difficulty" as *any difficulty faced by the respondent related to voting in the 2020 election.* This definition reflects our view that perceptions of difficulty are inherently subjective. We prefer the straightforward question of "how difficult was it to vote" (V202119) above all other measures of voting difficulty within the survey (e.g., V202120 a/b/c and others). This measure presents survey responses as ordinal data, ranked on a five-point scale from "not difficult at all" (1) to "extremely

---

[1]Pew Research Center. "Turnout soared in 2020 as nearly two-thirds of eligible U.S. voters cast ballots for president." (2021).
[2]United States Census Bureau. "Majority of Voters Used Nontraditional Methods to Cast Ballots in 2020." (2021).
[3]American National Election Studies. 2021. ANES 2020 Time Series Study Full Release [dataset and documentation]. July 19, 2021 version. www.electionstudies.org.
[4]Petrocik, J. R. "Measuring party support: Leaners are not independents." (2009).

difficult" (5). We recode these responses to a binary set of values – "no difficulty" (0) and "difficulty" (1) – to simplify our analysis with metric data and align with our earlier contention that any perceived difficulty is true difficulty. While we acknowledge that this transformation is "lossy," we believe the benefits of analyzing the research question from a more *general* position (i.e., which voters experienced more difficulty *in general*) outweigh the loss of some information granularity.

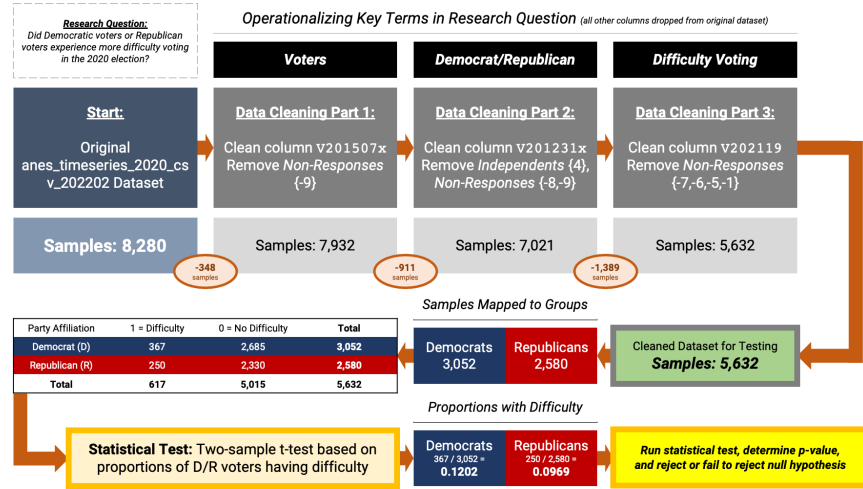These definitions underlie the data-cleaning steps laid out in Figure 1 below.

Figure 1: Process flow of data cleaning and subsetting steps

As Figure 2 indicates, we *do* see evidence of difficulty experienced by voters of both parties within the data. The left-side plot shows the distribution of voter difficulty ratings for each party. The right-side plot depicts higher-level differences across party lines. Because both plots show voter *counts*, it is not immediately apparent which party's voters faced more or less difficulty. To that end, we must first restate the counts in the right-side plot as the *proportions of voters experiencing difficulty within each party.*
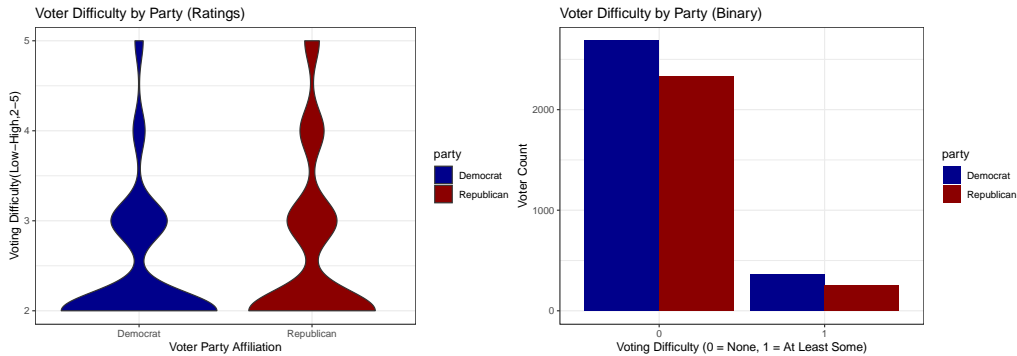

Figure 2: Democratic and Republican voters experienced varying levels of difficulty in the 2020 election

3   Results: Statistical Test and Findings

Table 1 below expands on the preceding discussion. This table standardizes the difficulty levels observed in Figure 2 by showing the proportions of voters from each party who experienced either no difficulty or at least some difficulty. Immediately, we can see that a higher proportion of Democratic voters (0.120 or 12.0%) than Republican voters (0.097 or 9.7%) experienced difficulty voting.

Table 1: Proportions of 2020 Voters Experiencing Difficulty

	Difficulty Experienced	No Difficulty Experienced
Democrat	0.120	0.880
Republican	0.097	0.903

To evaluate if this difference is significant, we must employ a statistical test. Given the presence of a grouping variable (voter political party), our range of available testing options includes only *unpaired tests.* Due to the original ordinal nature of the ANES difficulty data and our binarized transformation of this data, both non-parametric options (such as the Wilcoxon rank sum test) and parametric options (such as the two-sample Welch's t-test) are available to us. Since parametric tests offer more power than their non-parametric counterparts and our response variable (voting difficulty) is analyzable under *metric* assumptions, we prefer the parametric option. We address the necessary conditions for a two-sample Welch's t-test below, after stating our null hypothesis and alternate hypothesis:

> *Null Hypothesis: $H_0 : \mu_D = \mu_R$, where μ_D represents the proportion (or mean) of Democrat voters who experienced difficulty voting and μ_R represents the proportion (or mean) of Republican voters who experienced difficulty voting.*

> *Alternate Hypothesis: $H_A : \mu_D \neq \mu_R$*

Our choice of the two-sample Welch's t-test is supported by: 1) the analyzed data drawn from an i.i.d. sample – the sampling method employed by ANES to survey the large ($> 7{,}000$) number of respondents is i.i.d, as the administration of different survey modes to members of the same household make it unlikely that answers are interrelated or represent just one sub-group of U.S. citizens; 2) the response variable is on a metric scale – the binary transformation of the ANES voting difficulty field makes it a metric representation; 3) a grouping variable is present – Democratic and Republican voters; 4) the sample size of each group is large enough ($>=30$) to assume the Central Limit Theorem (CLT) – as shown by the bar plot in Figure 2.

```
test_result <- t.test(voting_difficulty ~ party_affiliation, data = testdf)
#Result: Reject the null hypothesis (p < 0.025 for two-tailed test)
```

The test yields a significant result ($t = 2.8191, p = 0.0048$), indicating that the proportions of Democrat and Republican voters who experienced difficulty voting in the 2020 election are not statistically equivalent. Therefore, we reject the null hypothesis. More plainly, the test result shows that *Democratic voters experienced more difficulty voting in the 2020 election than Republican voters.*

While the p-value tells us *that* a significant difference in voting difficulty exists, it does not tell us the *size* of that difference. We can assess this effect size by computing Cohen's $d = 0.07$, which suggests a small effect size ($d < 0.20$).

4 Discussion: Relevance to Today and Tomorrow

Our findings indicate that while voters from both parties experienced difficulties voting, Democratic voters experienced statistically yet not practically more difficulty than Republican voters. In closing, we highlight two key limitations of our review and offer one hopeful remark for the future. First, we recognize that the exclusion of *non-leaning* Independent voters and non-respondents limits our ability to generalize findings to the broader U.S. voter population. Second, we acknowledge that because our review does not touch on the *causes* of difficulties faced by voters, sources of voting difficulty must be investigated separately. Finally, we hope our analysis invites dialogue and change. We remain optimistic that conversations across party lines can serve as the catalyst for eliminating voting difficulties in the future.