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

Datasci 203: Lab 1

1 Importance and Context

The issue of voter suppression is extremely important for the current stage of the democracy of the US. In this political climate, the ability to influence the voting turnout for voters of desired party might as well decide the results of election. That's why it's extremely important to prevent legislation that would discriminate against certain classes of population.

Voter suppression laws differ by state and, in some cases, by county.

Experts debate that the voter suppression against historically marginalized groups in the U.S. is a common and widespread problem right now. Among others they determine following problems:

- voter registration problems
- voter purging
- increasing ballot requirements
- confusing process
- poll closures and long lines

As a first step, this analysis aims to address the following research question:

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

The answer to this question could provide insight into the severity of voter suppression problems and serve as a jumping start for further research. To help us better understand the experience of voting in the 2020 election we will ask following sub-questions:

- Did wait time at the poll differ for Democratic and Republican voters?
- What is the difference between rate of encountering specific impediments to voting between parties?

2 Data and Methodology

Our analysis leverages data from the 2018 American National Election Studies (ANES). This is an observational dataset, based on a sample of respondents drawn from the YouGov platform. The Yougov panel is not nationally representative, and consists of participants who sign up to complete questionnaires in exchange for rewards. This dataset includes 8280 individuals.

As we report in Table 1, only 44% of ANES respondents of both parties report that they did not experience any difficulty voting in 2020 elections. As the difficulty of voting is subjective and can mean difficult things for respondents we're not going to use it for analysis and only present here for illustration purposes. Another issue worth noting is that the "difficulty voting" data was gathered only for people who ended up voting so it's subject to "survivorship bias" which is another form of selection bias.

To operationalize the concept of voter difficulty, we will identify individuals in the dataset that experienced problems voting related to the democratic process such as problems with registration and long wait times. Then we will analyze if there is any difference between those numbers for democrats and republicans.

Table 1: Number of people reporting problems

political_party	Democrat	Republican
Total sample size	3133	2677
Registration problems	53	52
Absentee ballot problems	53	24
Identification problems	25	16

The survey includes questions about problems that respondents experienced while during 2020 election including: registration problems, long wait lines etc. Data is reported on binary scale (1: problem mentioned, 0: not mentioned), and the same question is asked to each respondent that voted. People who wanted to vote but didn't end up voting are not presented in the sub-set with problems.

```
## Warning: Removed 3068 rows containing non-finite values (stat_count).

## Warning: Removed 3068 rows containing non-finite values (stat_count).
```

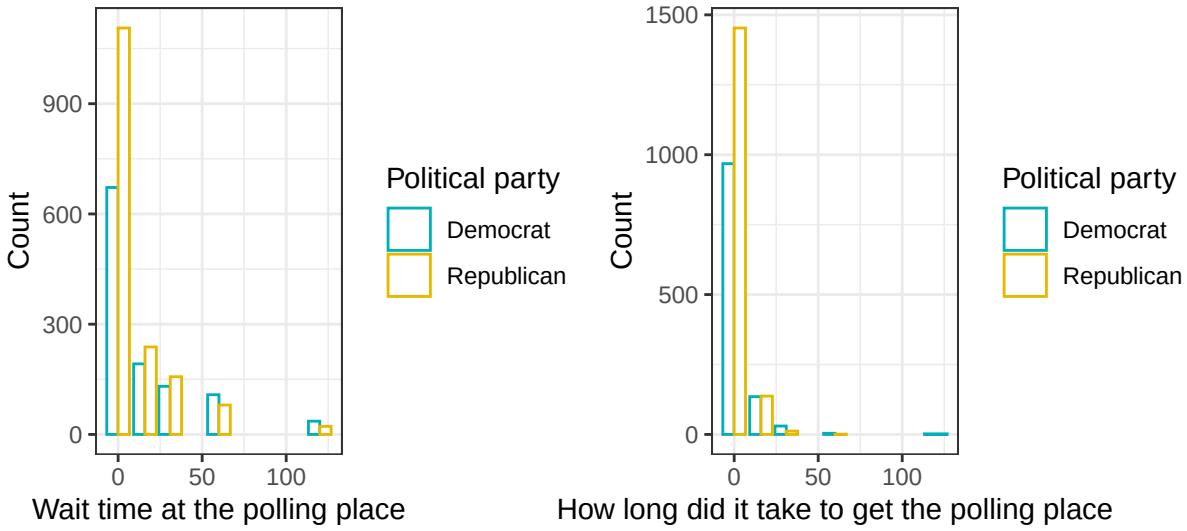

Figure 1: Waiting and travel times comparison

The left panel of Figure 1 plots <...>

To compare the difficulties, we identify the respondents who had problems while voting and respondents who planned on voting but ended up not voting. The right panel of Figure 1 shows <...>

Both our grouping variable and our outcome variable are measured at the binary level. In these circumstances, common tests could include a two-sample proportion test and Fischer's exact test. We proceed with a two-sample t-test to demonstrate tools used in Datasci 203. Given the large sample sizes, the loss of accuracy from the t-test will be negligible. The null hypothesis for the t-test can be expressed as follows:

> **Null Hypothesis:** *The probability that a member of democratic party experienced difficulty voting in 2020 election is the same that of republican party.*

The t-test requires the following assumptions to be true:

1. **i.i.d. data** First, data must be generated via an iid process. The ANES 2018 pilot uses a panel of individuals from the YouGov platform. There is a possibility that this introduces dependencies. For example, participants may tell friends or family members about YouGov, resulting in a cluster of

individuals that give similar responses. Nevertheless, YouGov claims to have millions of users, which suggests that links between individuals should be rare.

2. **Metric scale** A binary variable qualifies as metric as there is only a single interval, which goes from zero to 1.
3. **Sufficient normality** The small number of people who experienced problems <...> Nevertheless, the large sample size suggests that the sampling distribution of the statistic should be approximately normal via the Central Limit Theorem.

3 Results

```
wait_time_test <- t.test(wait_time_num ~ political_party, data = anes)

travel_time_test <- t.test(travel_time_num ~ political_party, data = anes)

registration_problem_test <- t.test(registration_problem ~ political_party,
                                    data = anes)

id_problem_test <- t.test(id_problem ~ political_party, data = anes)

wait_time_problem_test <- t.test(wait_time_problem ~ political_party,
                                 data = anes)

absentee_ballot_problem_test <- t.test(absentee_ballot_problem ~ political_party,
                                       data = anes)
```

The test yields no evidence that there is any difference in perceived voting difficulty between democrats and republicans (t = 6.09, p = 0). From a practical perspective, <. . . >.

Several limitations of our test affect the conclusions that may be drawn from it. <. . . > Additionally, the ANES data is not nationally representative, suggesting that our results may not generalize to the US population.

4 Discussion

This study found evidence <. . . > Are we measuring propensity to vote or how reliable people from different parties are? We see a correlation here but it doesn't imply causation.

Our results may be of key interest to <. . . >.