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 in the U.S. 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 categories of population.

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 for addressing this concern, 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 for members of each party we will ask following sub-questions:

- *Did wait time at the poll and travel time to the polling place differ for Democratic and Republican voters?*
- *Is there any difference in rates of encountering specific impediments to voting between Democrats and Republicans?*

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.

The survey includes questions about problems that respondents experienced while voting 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. Similar set of questions asked to each respondent about the reasons that prevented them from voting. Another useful set of data directly related to difficulty voting is the wait time at the polling place and travel time to the polling place. Both of those variables recorded in a range format: 0-15 minutes, 16-30 minutes etc.

It's important to note that ANES includes information on self-reported difficulty voting in 2020 elections. As the difficulty of voting is subjective and can mean different things for each respondent we're not going to

Table 1: Main reasons for not voting

	Democrat	Republican
Bad weather	0.00	0.28
Did not like the candidates	8.15	7.02
I am not registered	6.46	10.67
I did not feel that I knew enough about the choices	4.78	4.49
I did not have the correct form of identification	3.09	1.12
I did not know where to vote	0.84	0.56
I forgot	1.40	0.84
I requested but did not receive an absentee ballot	1.12	1.40
I was not allowed to vote at the polls, even though I tried	0.56	0.28
I'm not interested	3.93	3.65
Out of town	3.09	3.09
Sick or disabled	4.78	6.74
The line at the polls was too long	2.25	3.09
Too busy	6.18	8.15
Transportation	0.84	1.12

Table 2: Problems people encountered while voting

	Democrat	Republican
Bad weather	0.15	0.06
Concern about identification card	0.23	0.23
Confusion about ballot or machine	0.42	0.45
Difficulty getting to polling place	0.32	0.17
Difficulty obtaining absentee ballot	0.83	0.36
Issue mailing ballot	0.53	0.26
Long wait times	3.08	2.99
No problems at all	46.57	38.86
Other problem	1.44	0.96
Registration problem	0.57	0.26
Work schedule	0.55	0.70

use it in this analysis. 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 analyse 2 categories of data separately. First we will compare wait times at the polls and travel times to the polling place between Democrats and Republicans. Then we will identify individuals in the dataset who experienced problems related to the democratic process: problems with registration, identification problems, problems obtaining absentee ballot. We will compare rates for each of these problems separately.

Above tables represent voting problems split by whether respondent successfully voted or did not vote. To improve accuracy and remove selection bias we will pool results together from both groups of people.

For each potential impediment to voting 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 hypotheses for the t-test can be expressed as follows:

> **Null Hypothesis 1:** *The probability that a member of democratic party experienced difficulty registering in 2020 election is the same that of republican party.*
>
> **Null Hypothesis 2:** *The probability that a member of democratic party experienced difficulty providing correct identification in 2020 election is the same that of republican party.*
>
> **Null Hypothesis 3:** *The probability that a member of democratic party experienced difficulty obtaining absentee ballot in 2020 election is the same that of republican party.*

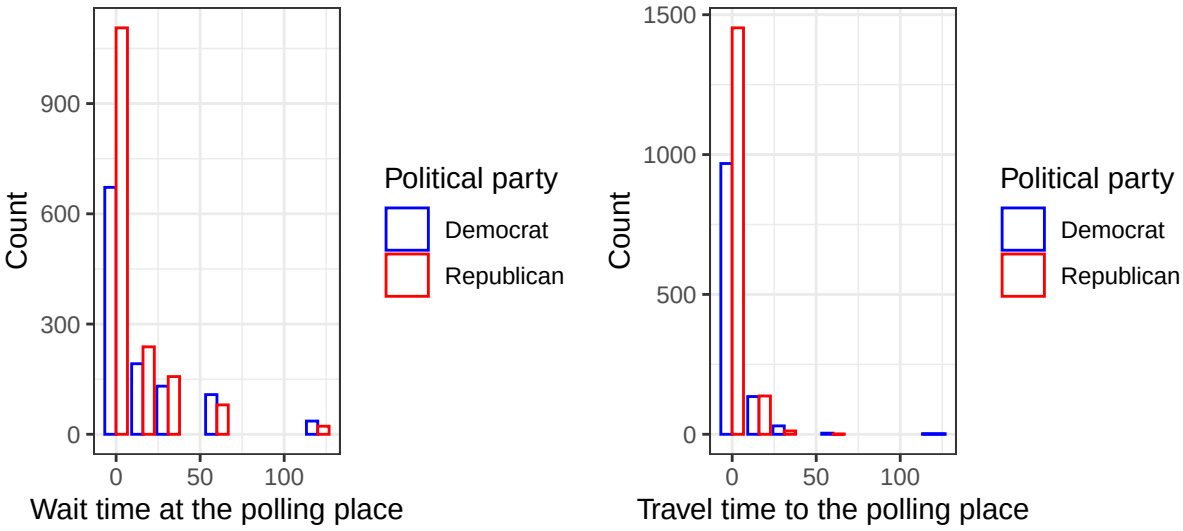

Figure 1: Waiting and travel times per political party

The left panel of Figure 1 plots amounts of respondents who reported waiting time within a given bracket. The right panel of Figure 1 shows similar data for travel time.

To test differences in waiting and travel times we will use the beginning of the time bracket as the data point. Since it's metric data and due to large sample size we can ignore the distribution shape and apply two-sample t-test with following Null Hypotheses:

> **Null Hypothesis for waiting time:** *A member of democratic party on average experiences the same waiting time at the polling place as a member of republican party.*
>
> **Null Hypothesis for traveling time:** *A member of democratic party on average takes the same amount of time to travel to the polling place as a member 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. Travel and wait times are also metric.
3. **Sufficient normality** The wait and travel times don't have normal distribution. Nevertheless, the large sample size suggests that the sampling distribution of the statistic should be approximately normal via the Central Limit Theorem. Sample sizes for individual voting problems are also large enough to apply CLT.

3 Results

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

travel_time_test <- t.test(travel_time_min ~ 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)

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

The test yield evidence that Democrats experience longer wait times (Democrats = 15.74, Republicans = 10.05, t = 6.09, p = 1.3078579×10^{-9}) and travel times (Democrats = 3.13, Republicans = 1.64, t = 4.95, p = 8.1283243×10^{-7}) than Republicans.

As for experiencing specific impediments to voting, the evidence was inconsistent:

- Registration problems: no difference detected t = -0.71, p = 0.48
- Identification problems: no difference detected t = 0.92, p = 0.36
- Problem obtaining absentee ballot: statistically significant evidence detected, t = 2.71, p = 0.01

Several limitations of our test affect the conclusions that may be drawn from it. As mentioned above, we are only able to measure association between experienced difficulties and political party, not causation. 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 of longer times for waiting at the polling place and traveling to the polling place for democratic than republican voters. Nonetheless, from practical point of view the difference in wait and travel times is relatively small (less than 6 minutes) and, probably, doesn't produce much effect on the democratic process.

As for potential issues of voter suppression, we evaluated problems with registration, identification and problems requesting absentee ballots and only found some evidence of difference in absentee ballots problem. Despite that fact, the ratio of people experiencing this problem is very small (around 1%, see tables above) and thus, the results suggest that impediments to voting were not widespread.

Our results may be of key interest to elected officials as they plan reforms in voting procedures. It's important to know that in our polarized time, Democrats and Republicans did not encounter significant voting difficulties.