Lab 1 Part 2

Amy Jung, Varun Dashora, Jenna Morabito

10/15/2021

Introduction

The 2020 election came with more challenges than any other election prior: a global pandemic, the increasing use of technology, and the previous 2016 election that polarized the nation — all came into play. The COVID-19 pandemic put not only the United States, but also the entire world into a state of isolation, with borders closing and stay-at-home orders put into place. Nowhere to go, many individuals inevitably turned to technology to maintain connection with their friends, family and coworkers. In this technology dependency, advancements in the social media space tightened their grip on the American population. Social media algorithms keenly observing, learning and outputting recommendations that closely align with the user may have also impacted the polarizing shift in political ideologies across the nation.

But beyond social media platforms and polarizations, a true democracy in the United States can only stand if participation in the presidential elections is equally accessible to all eligible voters, regardless of party affiliation. With differing state or local government orders and opinions on the pandemic, voters may have experienced varying degrees of voting difficulty. In this report, we explore if reported levels of difficulty in voting in the 2020 election differ for Democrats and Republicans.


```
# statistical test
wilcox.test(agg_dat$n[1:5], agg_dat$n[6:10], paired=FALSE, alternative="two.sided")
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  agg_dat$n[1:5] and agg_dat$n[6:10]
## W = 15, p-value = 0.6905
## alternative hypothesis: true location shift is not equal to 0
```

Hypothesis Testing

To investigate the research question about whether Democrats or Republicans had more difficulty with voting in the 2020 election, we needed to compare the voting difficulty of 2 voting samples: Democrats and

Republicans. Since we cannot take averages due to the fact that voting data is ordinal, we must compare overall distributions between Democrat and Republican voting difficulty to determine whether or not there is a significant difference. In addition, the data involves unknown population variance and a 2-group comparison, the Wilcoxon rank-sum test would be most appropriate for this situation.

The following is an examination of the assumptions being made in a Wilcoxon rank-sum test:

1. Ordinal Scale: The data is categorical with numbered categories imposing an order, meeting the ordinal scale condition.

2. Data are IID: The samples are drawn from the total population, meaning they must be identically distributed. In addition, the data collectors are limited to one sample per household, so household independence is minimized. The other major potential for independence is community-based political grouping, but there are so many different communities in the United States that there is no real danger of oversampling from one community and setting up a skewed and unrepresentative data collection. This assumption is sufficiently met.

Null Hypothesis: The probability that a randomly selected Democratic voter faces more difficult voting conditions than a randomly selected Republican voter is statistically the same as the probability of the opposite situation occurring; namely, a Republican voter facing more adverse voting conditions than a Democratic voter.

Alternative Hypothesis: There is a statistically significant chance that one group has more voting difficulty than the other.

Considering the results of the Wilcoxon rank-sum test, we accept the null hypothesis.

Most did not face difficulties voting
No significant difference in difficulty found between parties
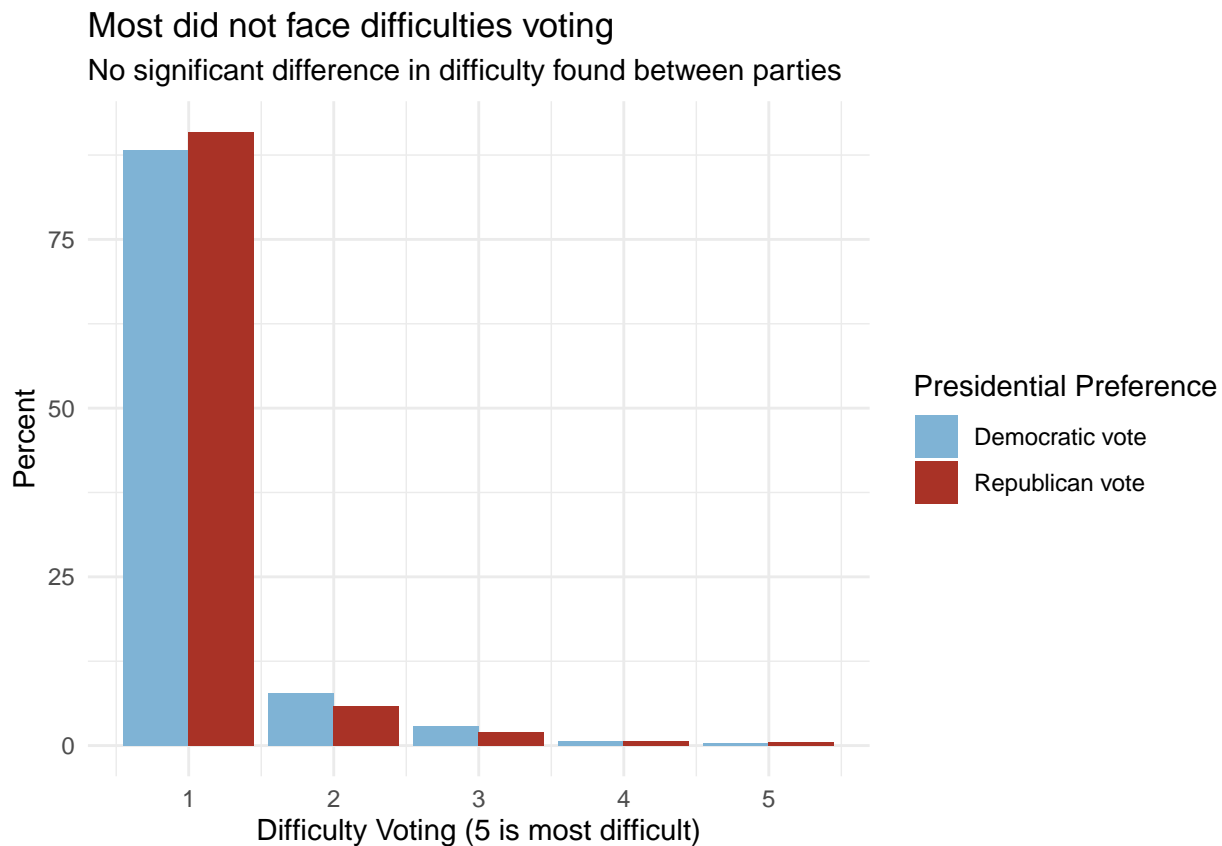
A visual inspection confirms that the percent of Democrats who had trouble voting about equals the percent of Republicans who had trouble voting. In this case, all participants ultimately voted.

Reasons for not voting
Being busy, not liking the candidate, and not being registered
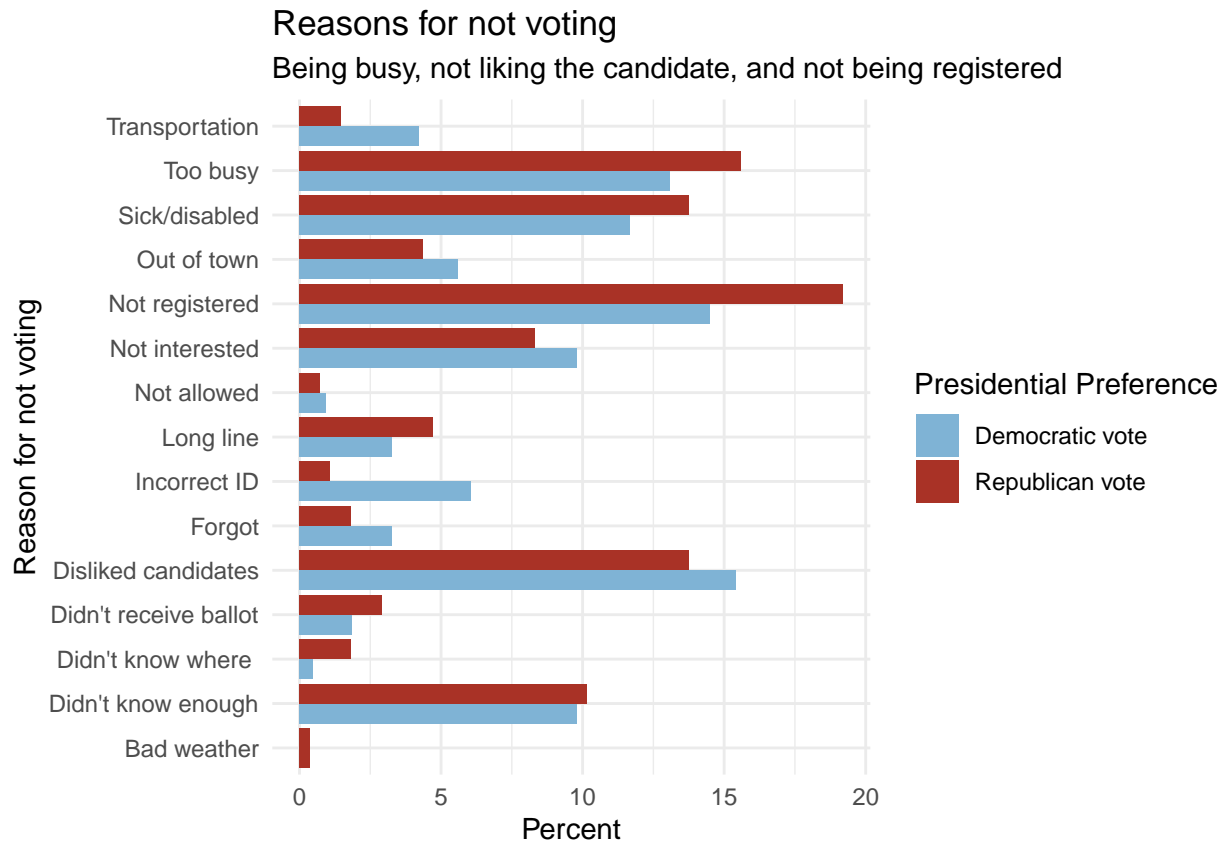
Exploring further, we examined the main reasons that Republicans and Democrats did not vote at all, but did not find statistically significant differences between the two groups:

```
wilcox.test(dem_no_vote$reason_didnt_vote, rep_no_vote$reason_didnt_vote, paired=FALSE)
```

```
## Warning in wilcox.test.default(dem_no_vote$reason_didnt_vote,
## rep_no_vote$reason_didnt_vote, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dem_no_vote$reason_didnt_vote and rep_no_vote$reason_didnt_vote
## W = 103, p-value = 0.9477
## alternative hypothesis: true location shift is not equal to 0
```

Test, results and interpretation

We did not find a significant difference in the difficulty that Democrats and Republicans had voting, and the most common reasons for not voting were the same between the two groups: being busy, not liking the candidates, not being registered, and not knowing enough about the candidates. It seems that even though there may have been differing opinions on in-person voting safety and voting option preferences between the two major political parties during this pandemic, our results indicate that there is not enough evidence with a p-value of 0.6905 to reject the null hypothesis, which was that there are equal chances that the following possibilities occur: a Republican voter had more difficulty than a Democrat; and, a Democratic voter had more difficulty than a Republican. The W-value, the Mann-Whitney statistic, for our test is 13, meaning: the sum of all ranks of the data entries for the random variable X (in this case, Democratic) is 13. We decided to aggregate and normalize the data in order to account for the fact that there are more Democratic voters sampled than Republican voters.