Question 1 - Are Democratic voters older or younger than Republican voters in 2020?

Matt Pribadi, MaKenzie Muller, Mon Young

Contents

1 Are Democratic voters older or younger than Republican voters in 2020?

1.1 Importance and Context

Are Democratic voters older or younger than Republican voters in 2020?

The 2020 election was unlike any other prior elections. The election took place during the COVID-19 pandemic, and it was well known that people over the age of 65 were at a much higher risk of death than people that were younger. This raises an interesting question of whether the voters' age played a role in the 2020 election. In this report, we plan to answer the question, "are Democratic voters older or younger than Republican voters in 2020?" This question could provide helpful insights and background information for political parties to target specific age group participants to increase their vote counts.

We will address this question by using data from the 2020 American National Election Studies (ANES). This sample data used various tactics to collect information, including online surveys, live video interviews, and telephone interviews, in response to the COVID-19 pandemic. The population was a subset of the 231 million U.S. citizens aged 18 or older living within the United States.

1.2 Description of Data

The 2020 ANES study consists of 1,381 variables and 8,280 respondents. However, we are primarily interested in the voter's party affiliation and the voter's age to answer our research question. Hence, it would make sense to narrow down these variables that we are interested in.

To determine voter's party affiliation, we considered multiple variables that can be used to determine voter's party affiliation. Because this ANES data is based on user survey responses, it included pre-election and post-election survey data. Therefore, analyzing both the Pre Party of Registration and the Post Party of Registration variables will give us the most appropriate voter's party affiliation information.

- V201018 - PRE: PARTY OF REGISTRATION
- V202064 - POST: PARTY OF REGISTRATION

By removing incomplete, non-Democratic party and non-Republican party entries in the Pre Party Registration answers, we reduced our data size and focused on significant data. In addition, 124 respondents answered "-1. Inapplicable" in the Pre Party of Registration question and later responded to their party affiliation in the Post Party of Registration. We make sure that we have done our exploratory data analysis properly and include these data entries. This resulted in 3,271 from the original 8,280 total respondents. We also examined the V202066 - POST: DID R VOTE IN NOVEMBER 2020 ELECTION variable, since the voters that voted can be significant as the COVID-19 pandemic may have caused older voters to not participate in in-person votings. This resulted in 2,598 from the original 8,280 total respondents.

The V201506 - RESTRICTED: PRE: BIRTHDATE: YEAR OF BIRTH could be the best variable to determine voter's actual age. Unfortunately, after further examining the data in this variable, all entries are answered in **-3. Restricted**, so we are not able to use this variable. The next best variable to use is the Age: V201507x.

- Age: V201507x
    - -9. Refused
    - 80. Age 80 or older
   - 18-79. Actual value

One concern that may occur is that the variable is set to 80 when the voter's age is 80 or older. Practically this makes sense as voters at the age of 95 or 100 will be rare and might seem like an outlier and skew the data. Hence, we decide to keep the data and remove the "-9. Refused" entries. The following histogram with the voter's age shows no outliers nor skew by keeping "80. Age 80 or older."

Voter's Age
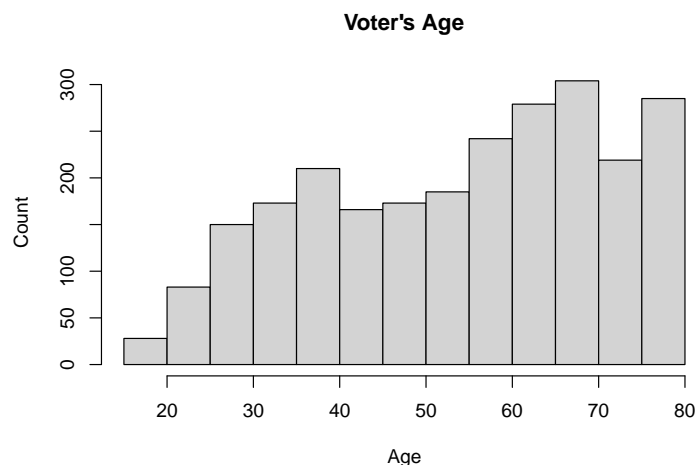
After further exploratory data analysis and data wrangling, we have our sample size at 2,497 with the following two variables for our analysis:

- Voters: categorical variable of Democratic or Republican
- Age: metric variable of actual age or 80

The following histogram shows a distribution of voter's age vs. voter's party affiliation. This histogram shows there might be age differences between Democratic voters and Republican voters.

Age and Political Party

voter age in democratic party vs republican parrty
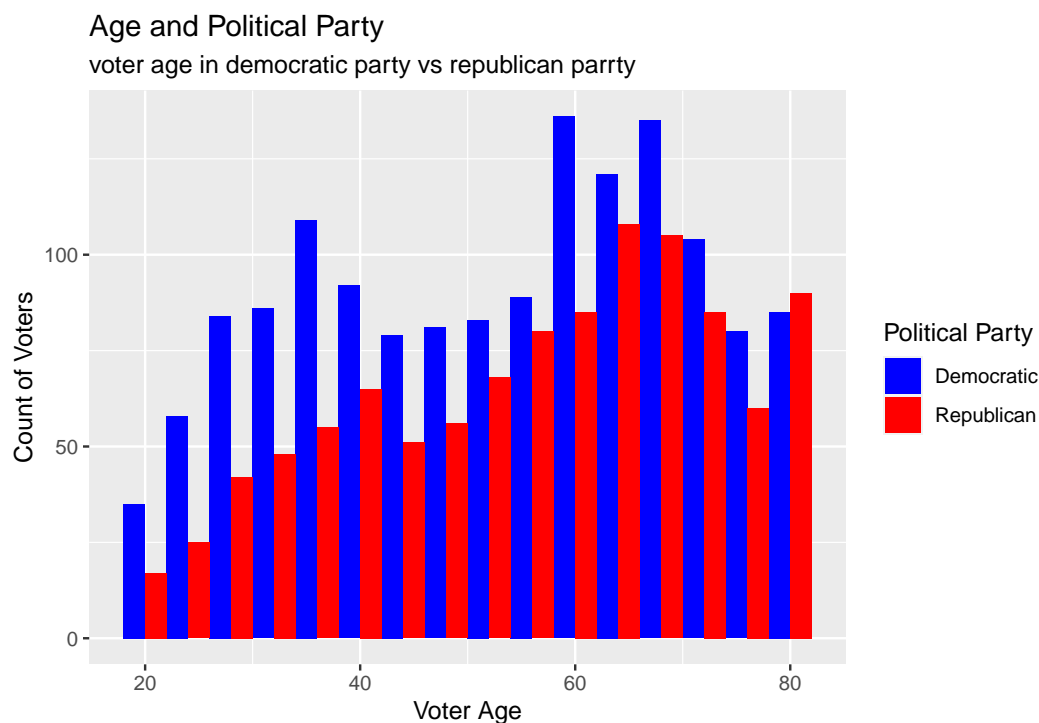
1.3 Hypothesis

To understand if Democratic voters are older or younger than Republican voters in the 2020 election, we have our null hypothesis H_o : there is no age difference between Democratic voters and Republican voters. Our alternative hypothesis H_a : there is an age difference between Democratic voters and Republican voters.

1.4 Most appropriate test

The voter's age is a metric variable, and the voter's party affiliation is in one of these two groups, the Democratic party or the Republican party. Therefore, to analyze the relationship of voter's age with voter's party affiliation, a parametric test is appropriate. Furthermore, the data is unpaired because these two party groups are independent of each other. Based on these properties, we determined using a two-sample t-test would be appropriate for this analysis. The two-sample t-test requires the following assumptions to be true:

1. Metric scale: Met. The age variable is a metric variable.
2. IID data: Met. Voters in this survey sample are independent of each other. In rare cases, family members doing this survey together could result in a cluster. However, this is unlikely, especially with a large sample size.
3. No significant deviation from normality: Met. The age histogram does not show a highly skewed distribution. Therefore, there is no significant deviation from normality. In addition, the sample size is large, at n = 2,497, which suggests that we are leaning towards a normal distribution with CLT.

```
t.test(Age ~ Voter, data = filteredVoters)
```

```
##
##  Welch Two Sample t-test
##
## data:  Age by Voter
## t = -4.7044, df = 2297.1, p-value = 2.697e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.501275 -1.852676
## sample estimates:
## mean in group Democratic mean in group Republican
##                 53.27591                 56.45288
```

1.5 Test, results and interpretation

Here we have implemented a two-sample t-test in R using t.test. In the statistical significance, the p-value for this analysis is 2.697e-06, which is less than 0.05. Based on the evidence that we have, we can reject the null hypothesis of "there is no age difference between Democratic voters and Republican voters." In the practical significance, based on my interpretation of the data that Democratic voter's average age is 53 and Republican voter's is 56. This suggests Democratic voters are younger than Republican voters in the 2020 election. To further support this, we ran a correlation in R, and the cor = 0.09 suggests there is a very small positive, close to no association between these two parties voters' ages. Thus this supports our suggestion that Democratic voters are younger than Republicans.

```
cor(filteredVoters$Age, as.numeric(filteredVoters$Voter))
```

```
## [1] 0.09306007
```

1.6 Overall Effects

We have conducted this analysis based on the data available in the ANES 2020 dataset. After finding an appropriate analysis of the data on hand, this analysis suggested the Democratic voters are younger than the Republican voters in the 2020 election. The mean comparison suggests Democratic voters are 3 years younger on average. Does this finding suggest younger voters played a role in Biden winning the 2020 election? In addition, does this tell us why Democratic voters are younger than Republican voters? These findings are interesting and make us consider if older voters do not want to risk going outside to vote during the pandemic or if states preferring the mail-in ballots than in-person votings can be related to voter's age that impacts the election results. This is insightful information, and further studies can help us understand if voter's age can influence future elections.