Lab 1: Democratic Voter Age vs Republican Voter Age :

Ria Mahajan, Shane Kramer, Viswanathan Thiagarajan

07/01/2021

Contents

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

Importance and Context

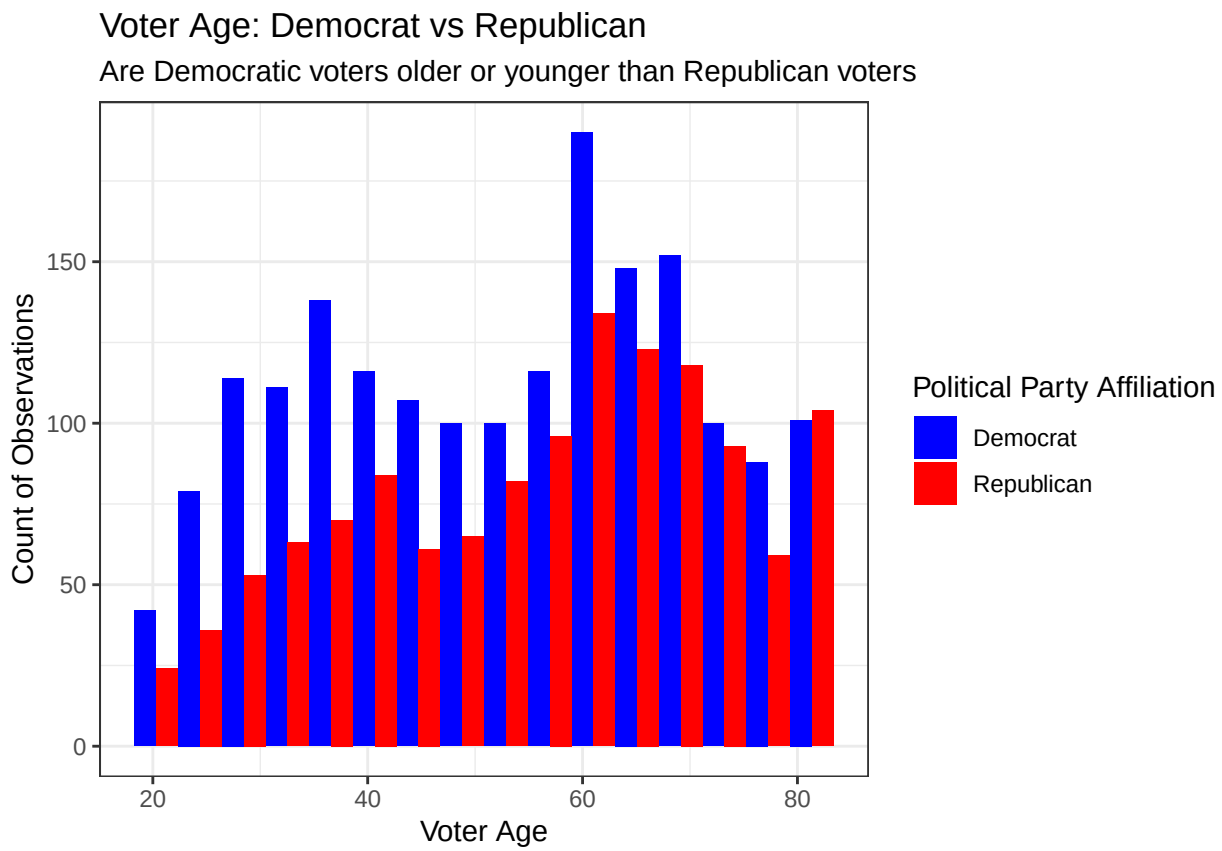
They say elections are decided by those who show up to vote. For question 1, we are trying to identify the ages of Democrat and Republican voters. In general, there is a link between age and voting behavior. As people age, they are more likely to be at the top of their earnings, so they are more likely to favor traditional conservative policies such as lower taxation on higher earners. Also, younger people when it comes to voting care about getting a job, education loan forgiveness, and the environment as compared to older voters who care more about health policies, retirement, social security, and other benefits. Because of this, we want to analyze if in the 2020 Election the Democratic voters were older or younger than Republican Voters.

To help us answer this question, we used the dataset anes_timeseries_2020_stata_ 20210324. The data set includes different attributes (per column) and includes values for age, voter registration status, and party affiliation. For this question we looked at variables V201507x (which was the Respondents Age) and V201018 (which was the Party of Registration). We created a new subset called "Registration" to extract data relevant for the voter age analysis.

Description of Data

A Democrat is a person registered with the Democratic party, and this is a person who believes in the political or social equality of all people. A Republican is a person registered with the Republican party, and this is a person who favors or supports a republican form of government. For this analysis, we utilized the ANES 2020 Time Series Study dataset. In order to adequately answer our question, we needed data that would show us age of the voters and what registered party they are affiliated with.

Looking at Age (Variable V201507x), we saw that the range of voters was between -9 to 80. For both parties we noticed the max age was 80 (as anyone over the age of 80 was listed as 80 too). To clean up "Age", we decided to filter out any voters younger than 18, because they can not vote, and any voters that had -9 because this represents voters who did not disclose their birth date. The second variable we chose to keep is "Party of Registration" (Variable V201018) as it told us what political party each voter is registered with. Since we are wanting to look at voters who are registered with the Democratic Party and Republican Party, we can filter out all other values.



Originally we started out with 8,280 entries but after filtering out voters younger than 18/voters whos birthday did not register and only looking at the political party affiliation of "Democrat" or "Republican", we are left with 3,067 entries too analyze and compare.

```
#Summary of Republican Voters (sd= 16.664, mean = 55.89, sample size= 1,255 )
summary(Republican_Age$Pre_Summary_Resp_Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   19.00   42.00   58.00   55.89   69.00   80.00
```

```
#Summary of Democrat Voters (sd= 17.205, mean = 52.66,sample size= 1,792  )
summary(Democrat_Age$Pre_Summary_Resp_Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   19.00   38.00   54.00   52.66   67.00   80.00
```

Sample Size	Notes
8280	Initial Size
3067	Subset of Democratic and Republican Voters (Age > 18 years)
1255	Total Republican Voters (Age > 18 years)
1792	Total Democratic Voters (Age > 18 years)

Figure 1: Sample Size Audit.

Most appropriate test

Our analysis aims to determine if Democratic voters are older or younger than Republican Voters. Our group decided to use a t test since we have two groups, Democrats and Republicans. The reason we went with this test is because we have Age data which is a numeric metric scale and the data is independent and identically distributed. We can assume this is IID data because there is a large sample size for both the parties and we can see that it is close to a normal distribution. Based on that, for our Null Hypothesis, we define it as H{0}: mean age for democrat voters = mean age for republicans voters. Alternate Hypotheses is defined as H{1}: mean age for democrat voters !=mean age for republicans voters.

Additionally, since the response variable, Age, is metric data, a parametric test is appropriate as we will be comparing the mean ages of the respective parties. We plan to use the Cohen's d Test in the analysis to see how effect size plays a role.

```
#T test for Democrat Age vs Republican Age
t.test(Democrat_Age$Pre_Summary_Resp_Age,Republican_Age$Pre_Summary_Resp_Age)
```

```
##
##  Welch Two Sample t-test
##
## data:  Democrat_Age$Pre_Summary_Resp_Age and Republican_Age$Pre_Summary_Resp_Age
## t = -5.2118, df = 2773.5, p-value = 2.007e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.443537 -2.014029
## sample estimates:
## mean of x mean of y
##  52.66371  55.89249
```

```
#Cohen's D test Democrat Age vs Republican Age (Practical Significance)
cohen.d(Republican_Age$Pre_Summary_Resp_Age,Democrat_Age$Pre_Summary_Resp_Age)
```

```
##
## Cohen's d
##
## d estimate: 0.1901059 (negligible)
## 95 percent confidence interval:
##     lower     upper
## 0.1180279 0.2621839
```

Test, results and interpretation

The question we are trying to answer is, Are Democratic voters older or younger than Republican Voters in 2020? When conducting a Welch Two Sample t-test we see that the p-value is equal to 2.007e-07. This indicates a high statistical significance, and we can therefore reject the null hypothesis in which the mean age for voters in the democratic and republican party is equal at a 95% confidence interval. We see that the mean for Democratic voters is 52.6 and the mean for Republican voters is 55.89 which means that the average republican voter is about 3.5 years older than the democrat average voter's age.

To find the practical significance, our group also conducted a Cohen's D Test at a 95% confidence interval and got an effect size of approximately .1901. Since difference between two groups (Democrat Voters and Republican Voters) is less than 0.2, the effect size is small and therefore difference is negligible, even if the relationship is statistically significant. Despite the classification of the effect size, future governors and policy makers can leverage this information and can conclude that Democratic Voters are younger than Republican Voters according to our data.