

Lab 1: Question 1

Courtney Smith, Chandler Haukap, and Hassan Saad

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

Importance and Context

The 2020 election had record rates of voter turnout in both parties. This question explores the relationship between voter age and political party. Visualizing the age distribution can help us see how the shifting demographics of the country are reflected in the outcome of elections.

Description of Data

The data for this report comes from the 2020 American National Election Studies (ANES) survey preliminary release. It contains 1,381 variables and has 8,280 respondents. To answer the question, we used variables V201231x and V201507x, renamed as ‘party_id_scale’ and ‘age’, respectively. Party affiliation is on a 7-point scale with 1 corresponding to “strong Democrat” and 7 corresponding to “strong Republican”. The scale is shown below.

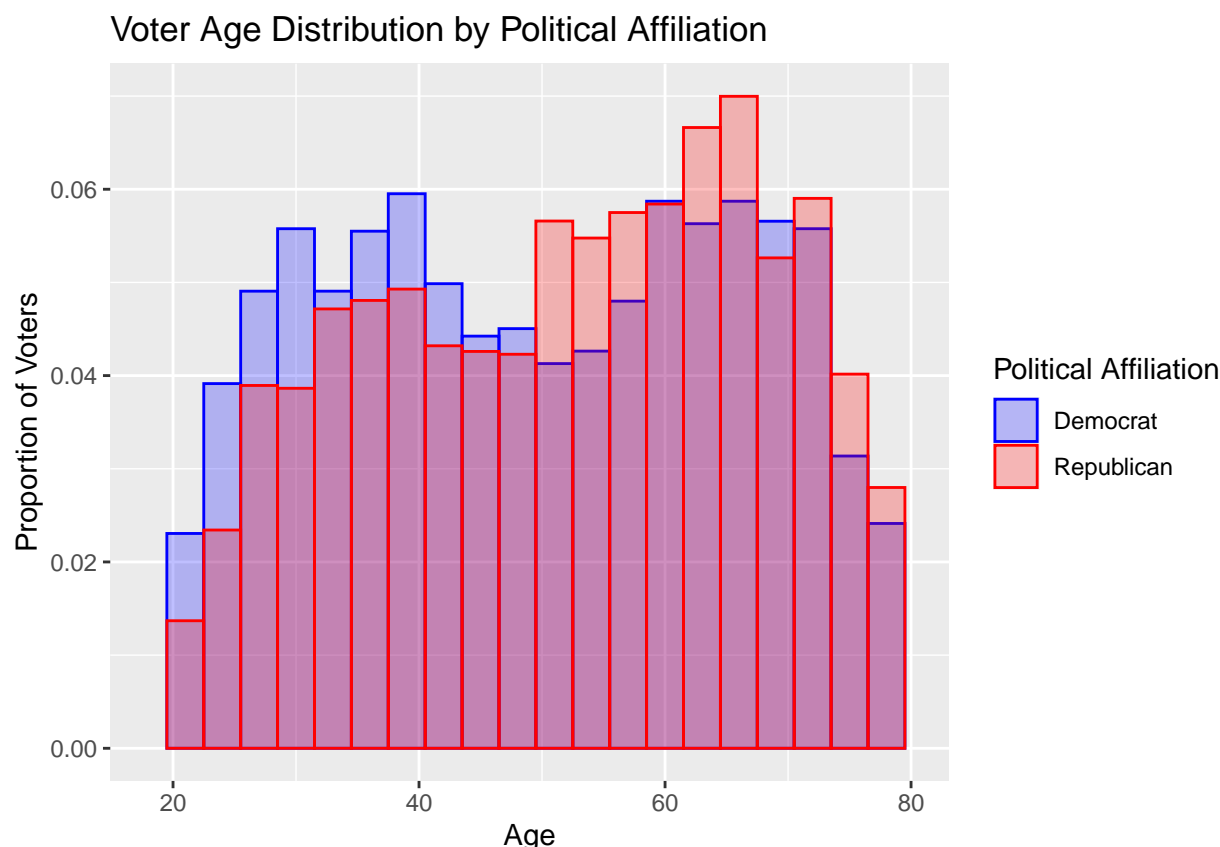
Table 1: Meaning of Numerical Political Affiliations

Number	Affiliation
1	Strong Democrat
2	Not Very Strong Democrat
3	Independent Democrat
4	Independent
5	Independent Republican
6	Not Very Strong Republican
7	Strong Republican

To operationalize this question, the party affiliation scale variable was used to create a binary party affiliation variable which classified values 1, 2, and 3 as Democrat and values 5, 6, and 7 as Republican; values 4 (Independent), -8 (Don’t know), and -9 (Refuse to respond) were filtered out of the data. The table below shows the size of the Democratic and Republican groups.

Political Party	Count
Democrat	3730
Republican	3287

The age variable was filtered to respondents of voting age and older, dropping the value -9 for respondents who did not give their age. After filtering on age and political party, there were 7,017 observations remaining. 47% of these respondents were Republicans and 53% were Democrats. The binary party affiliation variable will allow us to compare the age of voters by party to identify whether there is a significant difference in the mean age of the groups.



Most appropriate test

Age is measured on a metric scale and the data is unpaired because the groups are mutually exclusive, so an unpaired t-test is appropriate here.

The unpaired t-test requires three assumptions:

1. A metric variable: Age is metric so this assumption is met.
2. Independent and identically distributed (IID) data: Participants in this study were selected randomly, with equal probability, from all residential addresses in the United States. This generates an IID sample, but a cash incentive was used to encourage participation which may create some bias regarding who chose to participate in the survey, and survey respondents from 2016 were invited via email to participate again in the 2020 study, which could create a dependency.
3. A distribution that isn't too far from normal: Both groups have slight valleys which cause them to differ from the normal distribution, but with the large sample size of each group, the t-test is still valid.

Null hypothesis

The histogram does not clearly show a substantial difference in age distribution by party, so we will use a two-tailed t-test to evaluate the difference in mean age between the two groups. In this case, our null hypothesis is that the means are equal.

Test, results and interpretation

We will test our hypothesis with a significance level of $\alpha = 0.05$. We will reject the null hypothesis if the p-value is less than 0.05.

```
t.test(q1_df$age ~ q1_df$party_id)
```

```
##  
## Welch Two Sample t-test  
##  
## data: q1_df$age by q1_df$party_id  
## t = -7.9694, df = 6964.3, p-value = 1.85e-15  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -4.040457 -2.445135  
## sample estimates:  
## mean in group Democrat mean in group Republican  
## 50.52386 53.76666
```

The two sample t-test has a p-value less than 0.05, so the result is statistically significant and we reject the null hypothesis. The 95% confidence interval for the difference in means is [-4.04, -2.45], indicating that the mean age of Democratic voters in the underlying population is very likely lower than the mean age of Republican voters, but our hypothesis did not test a specific directional difference.

In terms of practical significance, a small difference in average age of the voting population is not an actionable insight. A more useful question may be how the proportion of Republican vs. Democratic voters varies across age groups, rather than how the age of the voting population varies by party, so organizers and politicians can identify segments for outreach.