Lab 1: Question 1

Patrick Old, Xander Snyder, Tal Segal

Importance and Context

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

There is a long held belief that the younger generations tend to lean democratic while the older generations tend to lean republican. Given that older people tend to turnout to vote at much higher rates, that belief could greatly affect election results. Age is also particularly relevant in the 2020 Election, as both candidates, Joe Biden at 78 and Donald Trump at 74 are the oldest presidents in the history of the United States.

The differences in party age could affect the candidate selection as it is possible that individuals will prefer candidates closer to them in age. There is also a hypothesis that younger people prefer political candidates that are viewed as ideologically extreme and iconoclasts to the traditional party establishment, while older people prefer more moderate and traditional mainstream candidates. A better understanding of whether Democratic voters are truely older or younger than Republican voters could help those aiming to appeal to party voters better understand their audience.

Description of Data

The data that we will be using to answer this question is drawn from the 2020 American National Election Studies ANES[1]. The target population for the study was the 231 million non-institutional U.S. citizens age 18 or older living in the 50 US states or the District of Columbia. The samples were randomly drawn, with all included residential addresses having equal probability of selection.

We will be using this dataset to extract the subset of respondents who are A) Willing to reveal their age B) Voters C) Either democratic or republican. After all filtering, our subset contained 61.5% of the total dataset.

A) Ages were extracted using V201507x which asked for the respondent's age. If a respondent refused to give their age, they were excluded from our subset. Approximately 95.7% of responders were willing to give their age.

B) The subset of voters was taken as people who either answered that they are registered voters or that they intend to register to vote. Approximately 91.3% of respondents fell into our subset of voters. To capture as many potential voters as possible, we used three key questions to filter voters:

- V201019 which asks respondents if they intend to registered to vote, if they replied that they did intend to we included them in our sample. Approximately 4.3% of our subset were respondents who intended to register to vote.

- V201008 which asks respondents which address they are registered to vote. As an address is required to register to vote, if they replied that they were registered to vote at either their pre-loaded address or at a different address we included them in our subset. Approximately 86.0% of our subset were respondents who were registered to vote at their own address and 9.7% were respondents who intended to vote at a different address.

- V201009 which asks if respondents who were answering the survey online did not have an address if they are registered to vote. If they replied that they are registered to vote without an address we

---

[1]American National Election Studies. 2021. ANES 2020 Time Series Study Preliminary Release: Pre-Election Data [dataset and documentation]. February 11, 2021 version. www.electionstudies.org)

included them in our subset. Less than 0.1% fell into this category.

C) Party affiliation was filtered using V201228, which asked if respondents think of themselves as a Democrat, Republican or other party. 67.4% of respondents categorized as voters were either Democratic or Republican. 36.2% of them identified as being Democrats. 31.3% of them identified as being Republican. The data for both Democrats and Republicans leans towards older voters. The mean age for a Democratic voter is 52 and the mean age for a Republican voter is 55.

Both distributions of Republican and Democratic voters had a left skew (figures 2,3), indicating that voters of both parties tend to be on the older side. These plots align with what is currently assumed of older being more likely to vote. Comparing the box plots of both parties (figure 1), Democrats appear to have a younger median value compared to the Republican party. The 25% and 75% points of the Democrats are wider apart, but both are lower in value than the 25% and 75% points of the Republican party, indicating a potential difference in ages.
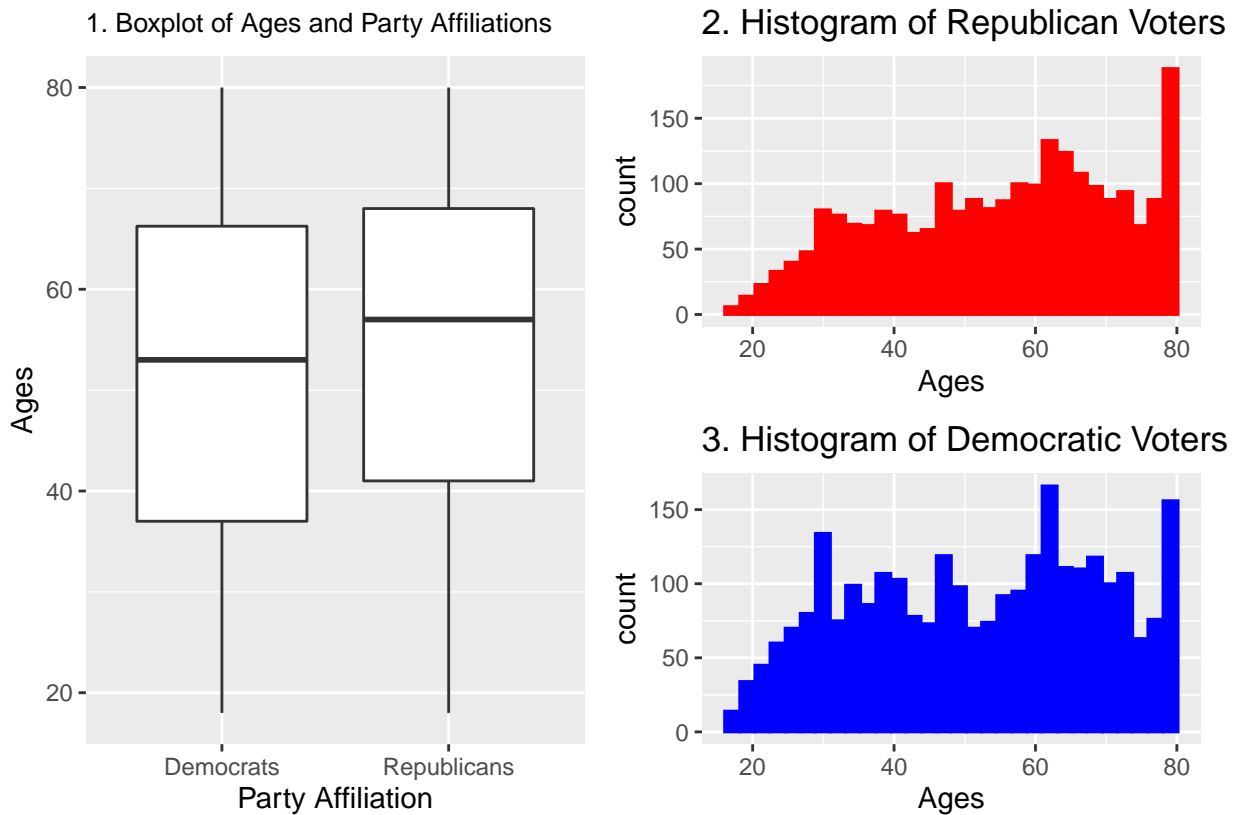

Figure 1: Ages of Democratic and Republican Voters

Limitations to our dataset:

1) By definition, a preliminary release is an early version of a dataset, and is subject to change in the full release.

2) Weights were intended to be used on this dataset to accurately represent the population, weights were not applied due to time constraints

3) All people 80 and over were clustered into the same age in survey (80), losing the value of their exact age and artificially skewing the data. Approximately 6.0% of our subset fell into this category.

4) Respondents who were unable to answer questions regarding their registration status, age or party affiliation were filtered out

Most appropriate test

We used the two sided Welch's t-test at a 95% confidence level to see if there is a difference between the mean Republican and mean Democrat ages.

Our null hypothesis H_0 is that mean Republican age is identical to mean Democrate age $H_0 : \mu_D = \mu_R$ Our alternative hypothesis H_a is that there is a difference between the means $H_A : \mu_D \neq \mu_R$

Assumptions of the Welch's t-test:

1) While our data is not normally distributed, the sample size is large enough for Central Limit Theorem to kick in.

2) Sample is i.i.d: The response rate overall was 36.7% thus there is a potential for nonresponse bias. The pre-election study was carried out from August 2020 up to election day on November 3rd, thus depending on the timing in which the responder was surveyed they may have different answers. Nonetheless, the data is relatively independent and identical, cross-sectional samples were randomly drawn with all included residential addresses across the all states having an equal probability of selection.

3) The data type of the variable we are testing, age, is metric, in this case ratio data.

4) Variance of the is unknown, since the Levene test rejected the hypothesis that the variances are equal, we used the Welch's t-test, which does not require sample variances to be the same.

We also tested the effective size of our mean difference using the Cohen's D test. The test assumes normally distributed, continuous metric data, which we are estimating to using the Central Limit Theorem.

```
#conducting Levene's test
vareq = t.test((repub-mean(repub))^2,(dem-mean(dem))^2,alternative="two.sided", var.equal=TRUE)
#since the variability is still different (p<0.05), we will need to use Welch's test
t <- t.test(repub, dem,alternative="two.sided", var.equal=FALSE)
#using Cohen's D to calculate effect size
cohens_d <- function(x, y) {
    csd <- ((length(x)-1)*var(x)+(length(y)-1)*var(y))/((length(x)-1)+(length(y)-1))
    cd  <- (abs(mean(x) - mean(y)))/(sqrt(csd))}
cd <- cohens_d(repub,dem)
```

Test, results and interpretation

The p-value of the Welch's t-test was 3.35^{-10}, thus we reject the null hypothesis that the means between the two parties are equal. As the mean Democrat age of our subset is 52 and the mean Republican age is 55, we find that Democrats are younger on average than Republican voters. These results align with the lower value distributions for Democrats seen on the boxplot. Although, the Cohen's D value of the magnitude of the difference was only 0.176, which corresponds to a little to no effective size. The practical significance of mean difference is thus indicated to be rather minor. This aligns with the mean age values between the two parties being only 3 years apart. The significance of the p-value in our Welch's t-test could be indicative of our relatively large sample size of 5094 test points. From both tests we conclude that Democratic voters in the 2020 election on average, are younger than Republican voters, but the overall effect of this difference is minor.