Muhammad Abdullah Umar, Netid: mumar4
Joshua Herman, Netid: jherma20
Hamza Shahid, Netid: hshahi2

418 Project 1

2. How many variables does the dataset have? What is the type of these variables? Are there any missing values? If so, how will you deal with these missing values?

How many Variables?

The dataset has 89 variables.

Type of the variables

The datatype of the variables is object(16), float64(43), int64(30).

Deal with missing values

There are missing values, two columns have NaN for all elements. We deleted these columns because they don't affect the data since they are NaN.

6. Compute the mean population in 2014 for Democratic counties and Republican counties. Which one is higher? Perform a hypothesis test to determine whether this difference is statistically significant at the $\alpha = 0.05$ significance level. What is the result of the test? What conclusion do you make from this result?

Results

Population of Democratic counties = 185786.724
Population of Republican counties = 77988.653.

Ttest_indResult(statistic=4.544147193705178, pvalue=6.584164225051351e-06)

Higher mean?

Population of Democratic counties at 185786.724 which is greater than 77988.6453 for the population of Republican counties

Hypothesis test:

Null Hypothesis:
Mean population of Democratic counties <= Mean population of Republican counties

Alternate Hypothesis:
Mean population of Democratic counties > Mean population of Republican counties

Significance

P-value < α
6.584164225051351e-06 < .05
This is a statistically significant result therefore we reject the null hypothesis that the mean population of Democratic counties is less than or equal to the mean population of republican

counties. We accept the alternative hypothesis where the mean population of Democratic counties is greater than the mean population of Republican counties.

7. Compute the mean median household income for Democratic counties and Republican counties. Which one is higher? Perform a hypothesis test to determine whether this difference is statistically significant at the $\alpha = 0.05$ significance level. What is the result of the test? What conclusion do you make from this result?

Results

Mean median household income for Democratic counties= 45643.87192982456
Mean median household income for Republican counties= 45268.803236245956

Ttest_indResult(statistic=6.116308071019074, pvalue=1.7811433799586417e-09)

Higher mean?
The mean median household income is higher for Democratic counties as it is 45643.87192982456 which is greater than that for Republican counties which is 45268.803236245956

Hypothesis test

Null Hypothesis:
Mean median household income for Democratic counties <= Mean median household income for Republican counties

Alternate Hypothesis:
Mean median household income for Democratic counties > Mean median household income for Republican counties

Significance

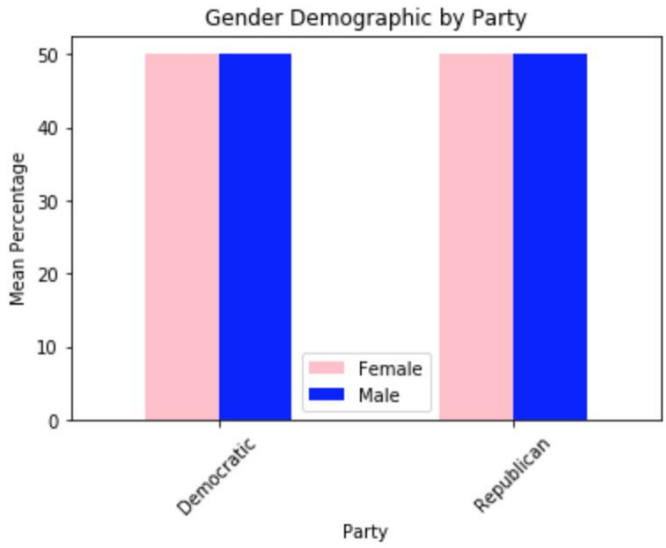
P-value < α

1.7811433799586417e-09 < .05

The result is statistically significant therefore we reject the null hypothesis which is the mean median household income for Democratic counties less than or equal to the mean median household income for Republican counties. This means we accept the alternative hypothesis which is the mean median household income for Democratic counties is greater than Mean median household income for Republican counties.

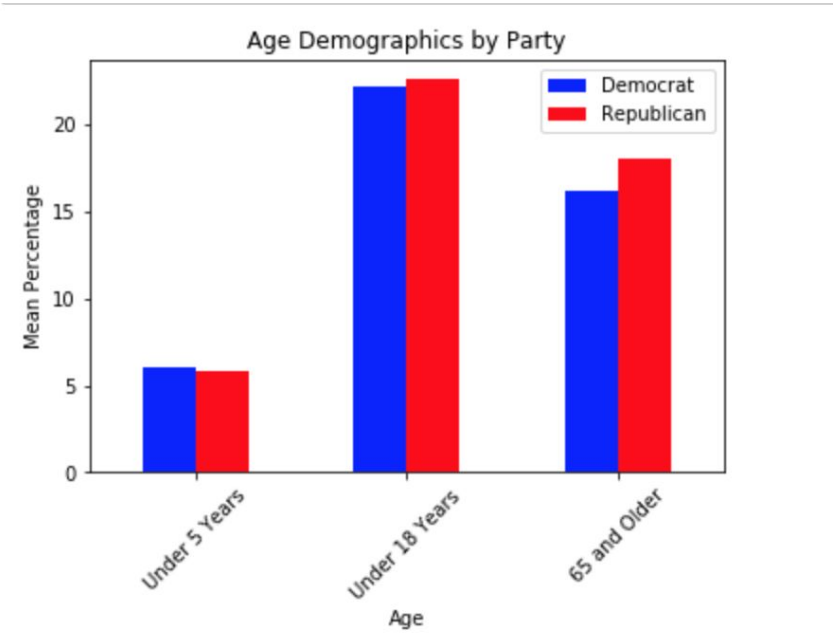
8. Compare Democratic counties and Republican counties in terms of age, gender, race and ethnicity, and education by computing descriptive statistics and creating plots to visualize the results. What conclusions do you make from the descriptive statistics and the plots?

Gender



Looking at the plots of the males and females of demographic by party there, is no significant difference between males and females that vote per party. We can clearly see that the percentage females in democratic counties is 49.96%, and in republican counties it is 49.94%. The percentage of males in democratic counties is 50.04% and in republican counties it is 50.06%. Looking at these statistics, there is no significant difference that shows that democratic or republican counties have more males than females in them or vice versa. There is an equal distribution of votes to both the parties be it male or female. Therefore, we conclude that gender of a county does not influence which political party that it votes in.

Age

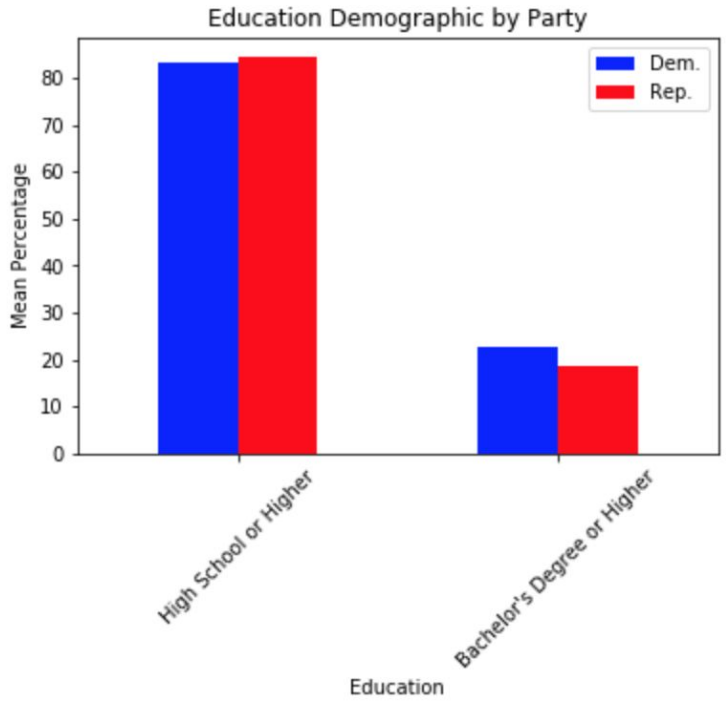


By looking at the age plots, we see that there isn't much difference between the age demographics by party for 18 and under as well as the age demographics for under 18 years.. The largest difference that is seen is for the 65 and older category which shows that counties with a higher percentage of 65 and older tend to vote for the Republican party.

The percentage age demographics for Under 5 years for democratic has a mean of 5.98 and a republican mean of 5.81. For under 18 years, the age demographics for democratic has a mean of 22.19 and a republican mean of 22.57. The age demographics for 65 and older has a democratic mean of 16.22 and a republican mean of 18.01.

As a result, we conclude that the age demographics of a county is not influenced by the age under 18, but this makes sense because no one under 18 can vote. However, for age 65 and older there is a 10% higher percentage of these counties voting for the Republican party. Therefore we conclude that counties with a higher number of 65 and older will more likely vote for the Republican party than the Democratic party.

Education

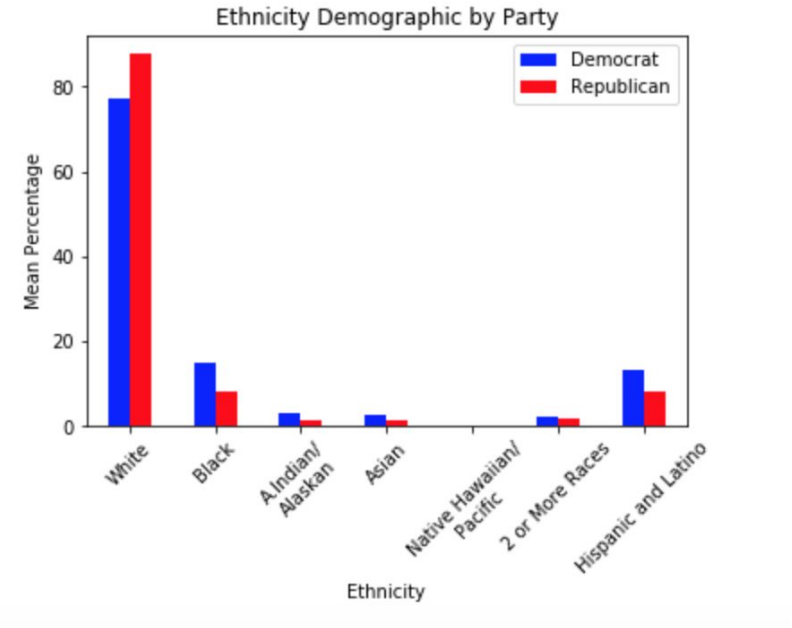


Looking at the bar graph the percentage of people who did high school or higher in the counties have a slightly higher rate of voting for the Republican party. On the contrary people in counties with a Bachelor's degree or higher have a larger difference and they favor the Democratic party.

For Democratic counties the percentage who completed high school or higher is 83.197%, and those who completed a Bachelor's degree or higher is 23.49%. For Republican counties the percentage the completed high school or higher is 84.41%, and those who completed a Bachelor's degree or higher is 18.59%. Taking this data and looking at Democratic counties with high school, but not a bachelor's degree is 59.71% and those in the Republican counties with high school, but not a bachelor's degree is 65.82%.

Based on the data shown, we conclude that there is a significance in counties where the education is high school, but not up to a Bachelor's degree tends to vote for the Republican party. On the contrary, counties that tend to have a Bachelor's degree or higher tend to vote more towards the Democratic party. Therefore, we conclude that counties with a higher level of education tend to vote towards the Democratic party, and counties with a lower level of education tend to vote towards the Republican party.

Ethnicity



Looking at the plots of the ethnicities of Democratic party there is a significant difference between the ethnicities and the parties they voted for. Specifically seen is the differences with counties that tend to vote Republican have a higher percentage of white, while a higher percentage of Democratic counties have a higher percentage of black and hispanic and latino voters.
The mean percentage of White is 77.04% in democratic counties, and in republican counties it is 87.85%. A difference of 10.81% more who voted Republican.
The mean percentage of Blacks is 15.025% in democratic counties, and in republican counties it is 7.87%. A difference of 7.16% more who voted Democratic.
The mean percentage of American Indian/Alaskans is 3.07% in democratic counties, and in republican counties it is 1.40%. A difference of 1.67% more who voted Democratic.
The mean percentage of Asians is 2.42% in democratic counties, and in republican counties it is 1.06%. A difference of 1.36% more who voted Democratic.
The mean percentage of Native American/ Pacific Islanders is 0.17% in democratic counties, and in republican counties it is 0.09%. A difference of .06% more who voted Democratic.
The mean percentage of 2 or More Races is 2.27% in democratic counties, and in republican counties it is 1.71%. A difference of .56% more who voted Democratic.
The mean percentage of Hispanic/Latino is 13.06% in democratic counties, and in republican counties it is 8.14%. A difference of 4.92% more who voted Democratic.
Based on the data we can conclude that Republican counties have a higher percentage of whites than do Democratic counties. Democratic counties have a higher percentage of blacks and hispanics

and latinos. Therefore counties with higher amount of whites tend to vote republican and those with a higher percentage of blacks and hispanics tend to vote Democrat.

9. Out of all the variables in the dataset, which ones do you think are more important to determine whether a county is labeled as Democratic or Republican? Justify your answer

Ethnicity.
White voters had the largest percentage of overall votes, the largest percentage of Republican votes, 87.85%, and the largest difference in percentage between voters who voted Democratic and those that voted Republican, 10.81% in favor of Republican. This leads us to conclude that the counties with a higher number of White voters are more likely to vote Republican.
For Democrats counties with a large number of black voters and hispanic and latino voters tend to vote towards the Democratic voters. There is a difference of 4.92% in favor of Democrats for latinos and a difference of 7.16% in favor of democrats for blacks. So it can be shown that ethnicity can be used to determine the labeling of a country as Democrat or Republican.

Education
Those with an education of a Bachelor's degree or higher tend to vote Democratic 5.35% more than Republicans do. Those with a high school degree or higher but not a Bachelor's degree tend to vote for the Republicans 6.11% more than do Democrats. So it can be shown that education can be used to determine the labeling of a country as Democrat or Republican.

Age (65 or older).
For age 65 and older there is a 10% higher percentage of these counties voting for the Republican party. Therefore we conclude that counties with a higher number of 65 and older will more likely vote for the Republican party than the Democratic party. So it can be shown that education can be used to determine the labeling of a country as Democrat or Republican.