

## Project 1 Report

### Ques 2:

**No of Variables:** The Merged dataset has 89 variables in total.

**Types:** There are 43 float variables, 30 integer variables and 16 object(String) variables.

**Missing variables:** There are two variables with missing/NaN values namely **No Preference Party** and **Uncommitted Party**

**Dealing with Missing values:** We **drop** the variables consisting of missing values using **dropna** function.

### Ques 6:

Mean Democratic county population: 185786.7245

Mean Republican county population: 78017.0336

**Democratic county population is higher.**

**Hypotheses Test statistics: ( $\mu_1 = \mu_2$ ) VS ( $\mu_1 > \mu_2$ )**

- **Test statistic: 4.5427**
- **P-Value: 3.32 e-06 (right tailed)**

Clearly p-value < 0.05 hence there is **statistically significant** evidence that the mean democratic county population is greater than the mean of republican county population.

### Ques 7:

Mean democratic county median house income: 45643.871

Mean republican county median house income: 45265.621

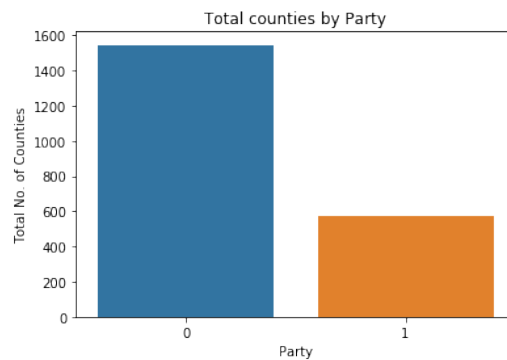
Mean democratic county house income is slightly greater.

**Hypotheses Test Statistics: ( $\mu_1 = \mu_2$ ) VS ( $\mu_1 \neq \mu_2$ )**

- **Test Statistic: 0.5414**
- **P-Value: 0.5883**

Clearly p-value > 0.05 and hence **there is no statistically significant evidence** to prove the median house incomes were different.

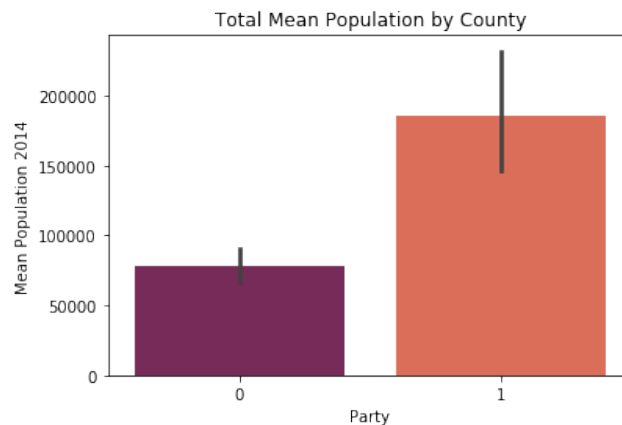
### Ques 8:



Clearly there were significantly more republic counties than democratic counties in 2014 midterm election

### Party Vs Population

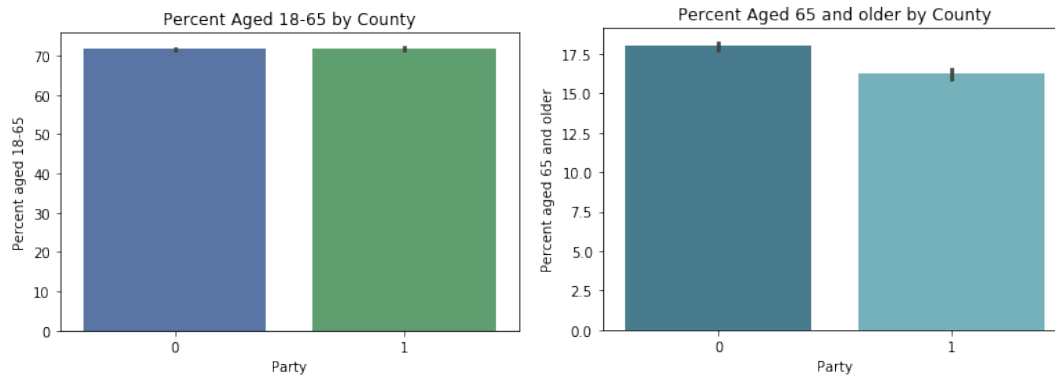
statistic=4.542, p-value=3.312e-06



Clearly the **mean population of Democratic Counties is higher** than Republican counties

### Party Vs Age

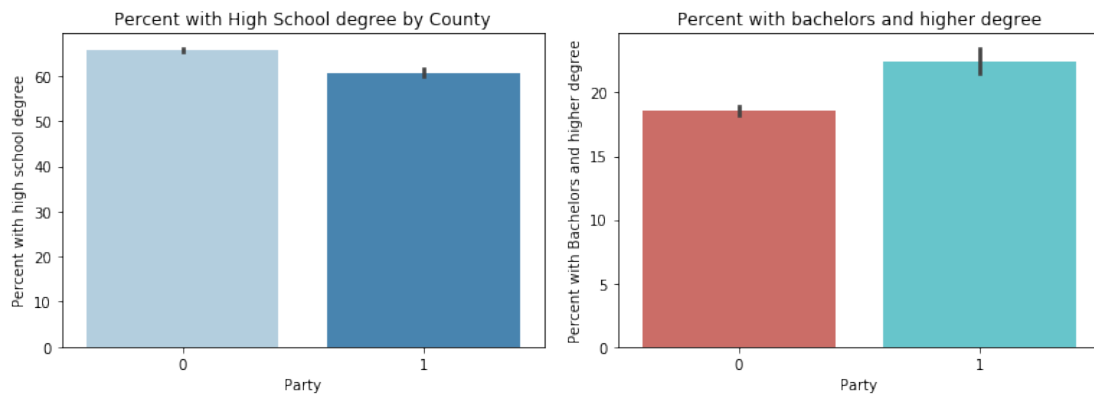
**18-65 Years Vs 65 and above (legal age to vote)**



There were **close to equal percentage of people** in both Republican and Democratic counties with the legal right to vote

## Party Vs Degree

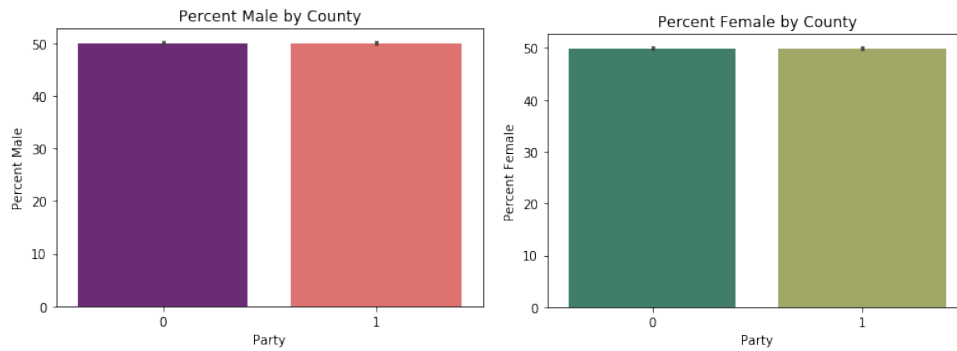
### (High School) VS (Bachelor and higher)



Clearly There **wasn't any significant difference** in the education levels of people in both counties.

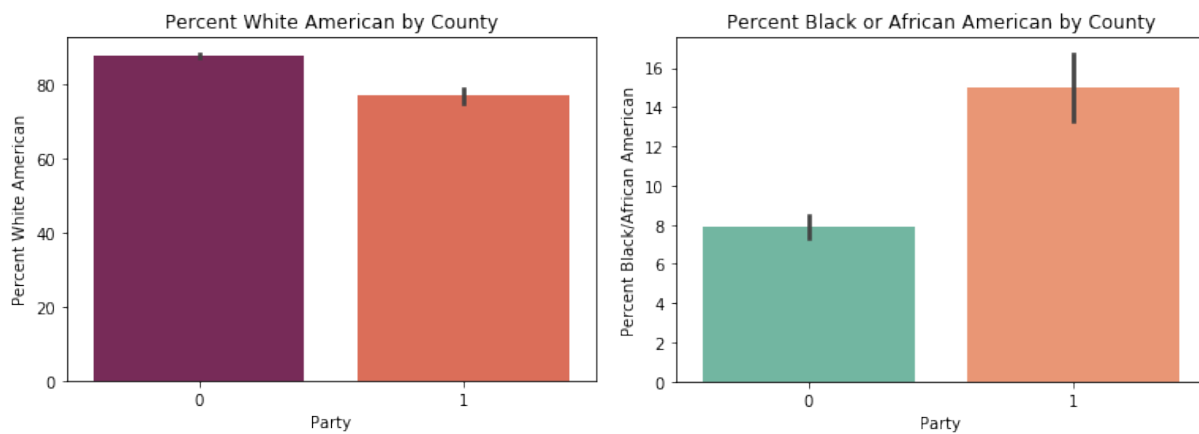
## Party Vs Sex

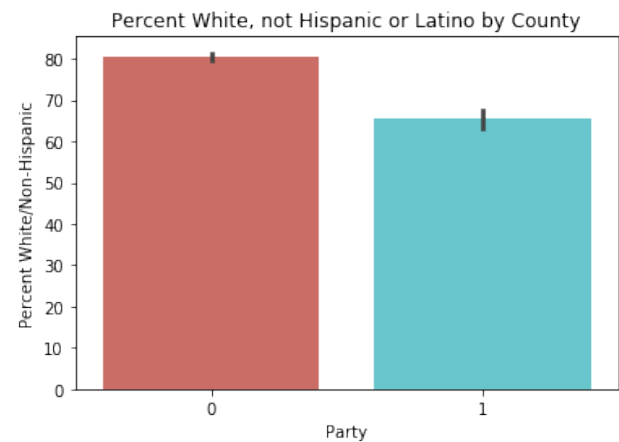
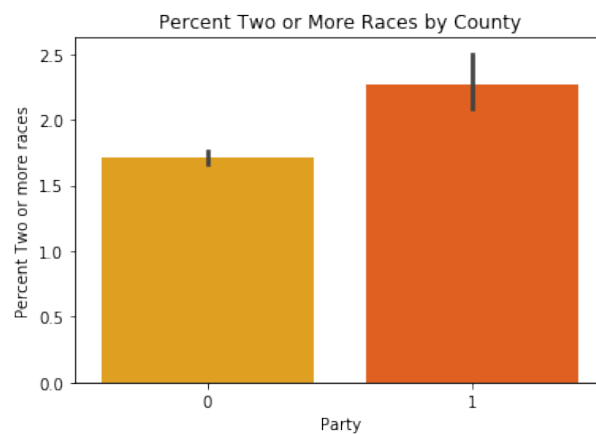
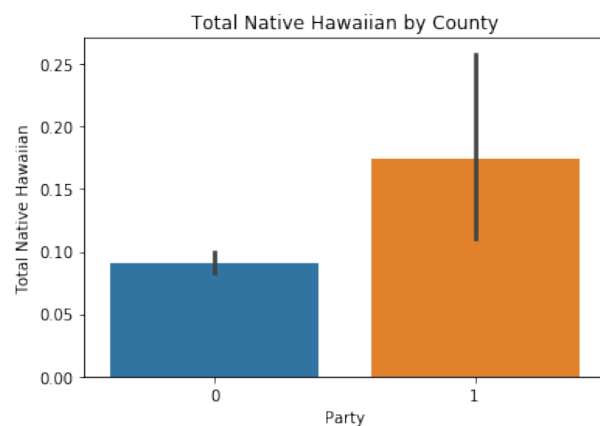
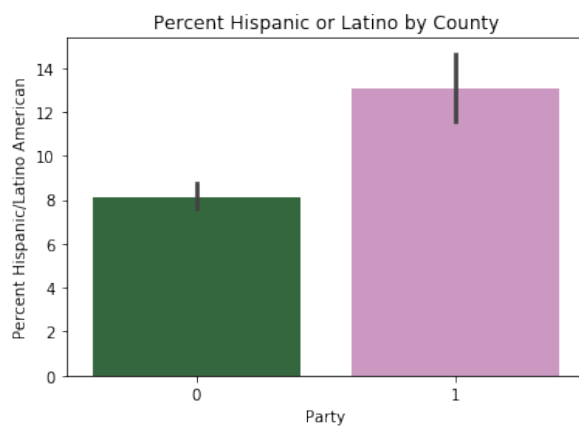
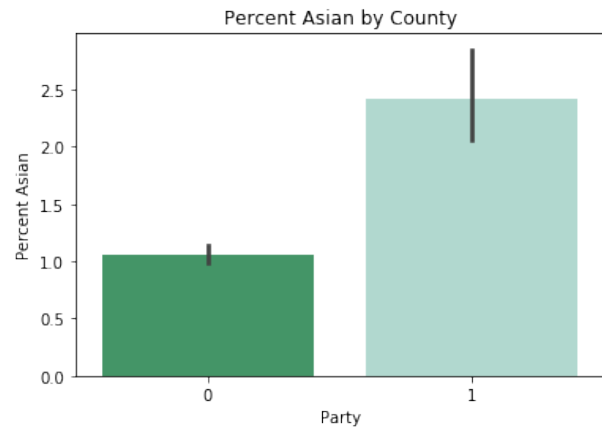
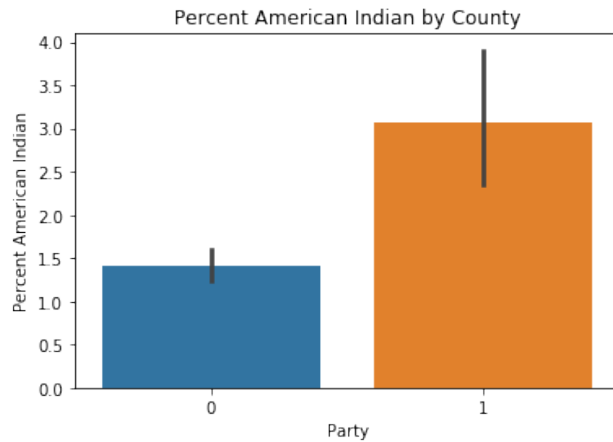
### Male Vs Female



Clearly there is **no difference** in the percentage of male and female population in both counties

## Party Vs Race and Ethnicity





Considering all the races and ethnicities in different counties the only significant differences in the percentage were found in the **White**(More in **Republican** Counties), **Black**(More in **Democratic** Counties), **Hispanic or Latino**(More in **Democratic** Counties), **White non-Hispanic**(more in **Republican** Counties)

**Ques 9:**

Based on the plots above the variables that show statistical difference in determining the counties being Democratic or Republic are:

- **Race:** Clearly higher percentage of **White Americans** were found in **Republican** Counties while higher percentage of **Black and Hispanic or Latino** were found in **Democratic** Counties.
- **Population:** Clearly the **Mean Democratic County population was significantly higher** than the Mean Republic County Population.

**Ques 10:**

