**Exploration of Election data from 2018**

We are given two datasets which contains data pertaining to 2018 elections in American counties.

The First dataset, electrion\_train provides an overview on the dataset, whereas The Second dataset, demographics\_data provides a detailed look at the collected data. Since the two datasets needed to be merged for to carry out further analysis, reshaping was necessary. This was achieved by reshaping the data to the wide format, using pivot\_table function of pandas. On reshaping the data, we got a dataset of 1205 X 6 columns.

Before merging the datasets however, we have to make sure the data is consistent. The first inconsistency is present in the column State. We created a dictionary with the key as the acronym for the state and the values are the full form of the states and stored it in a file named state\_map

{

Exact process of merging the datasets

}

After merging the datasets, we observed that there are 21 variables, the type of these variables are object int64 and float64. There are also irrelevant or redundant variables in the dataset. Year has a value of only 2018, and no other value. Hence, this is an irrelevant/redundant variable. We deleted the Year and Office column and inserted the year 2018 and US Senator in the table header. There are missing values in Democratic and Republican columns. Also, Citizen Voting-Age Population has values mentioned as 0.

We removed the Citizen Voting-Age Population since it has over 50% of data with the value 0. We removed the 5 entries of Democratic and Republican since a small observation won't impact the data analysis.

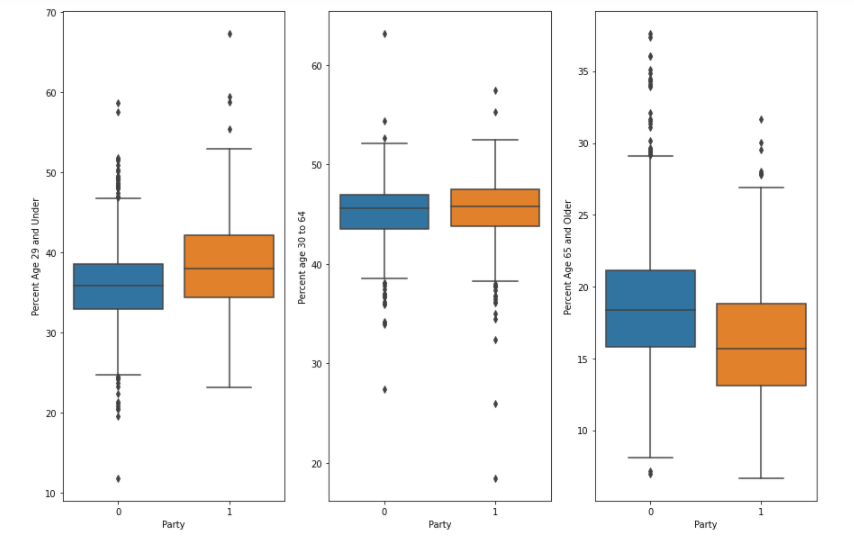
The mean median household income for Democratic counties is higher.

Since p-value less than the significance value we have sufficient evidence to reject the null hypothesis

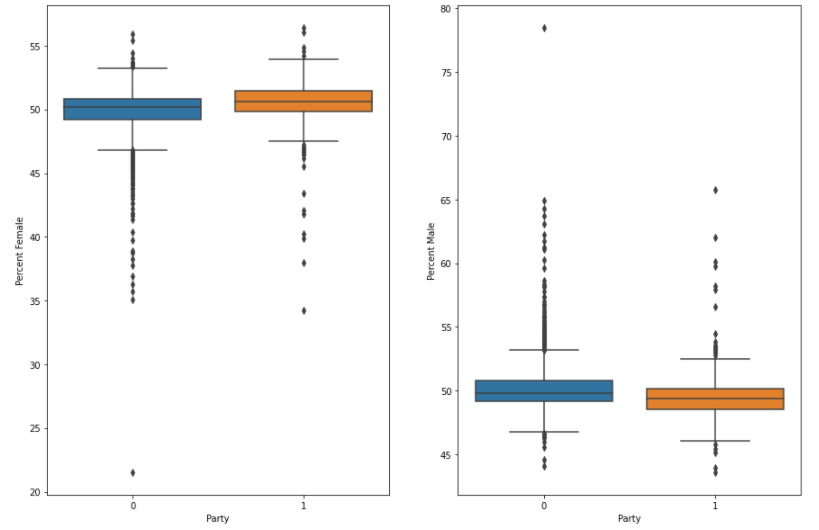
The mean population is higher for Republican Counties.

Since p-value is less than the significance value we have sufficient evidence to reject the null hypothesis.

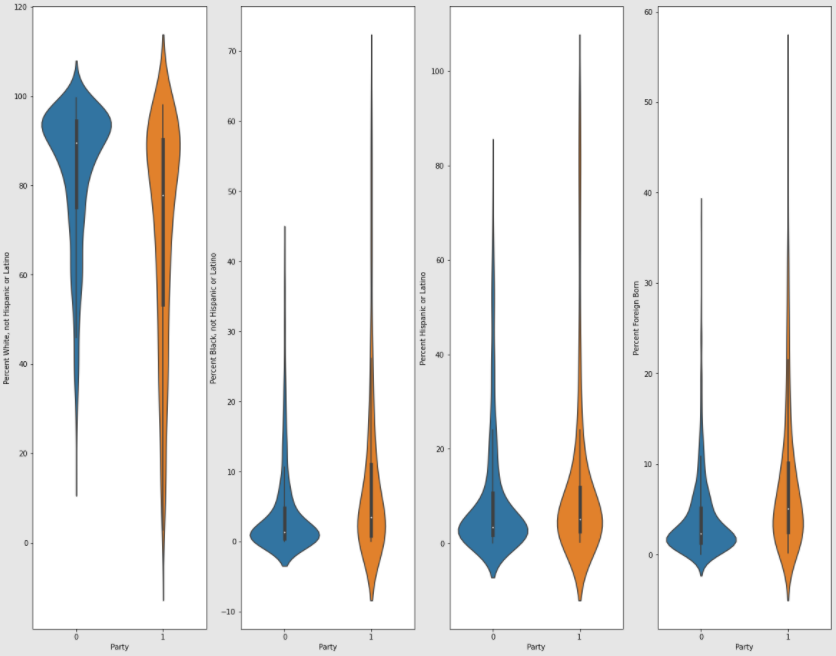
In terms of Age:



In terms of Gender:



In terms of ethnicity:



Total population is one of the important variable to determine whether a county is Republican or Democratic because the mean population of democratic counties is a lot higher than the republican counties which means the higher total population counties are inclined towards Democrats.  
Education level(Percent Less than Bachelors Degree) and Age(Percent Age 29 and Under,Percent Age 65 and Older) are also important variables because according to the plots the values of democrats and republicans in these variables vary a lot.