Student Name: Zhou Chen

Student ID: 112214117

Source of Dataset: [Election, COVID, and Demographic Data by County](https://www.kaggle.com/datasets/etsc9287/2020-general-election-polls)

Lab 1 Report

For this lab , the dataset used is election, covid, and demographic data by county. The dataset compiles county statistics for the 2016 and 2020 presidential elections, COVID cases and deaths, demographic numbers, and economic numbers. In this lab I randomly sample 500 observations from the total dataset. The features that are being visualized in this project are primarily focused on the county statistics for 2020, which is the year of the presidential election.

The list of features that being visualized in this lab includes total number of vote, total population, gender, race and ethnicity, Covid-19 cases, Covid-19 deaths, median household income, number of voting age citizens, poverty rate, total number of employment and unemployment, work types, and transportation tools, total number of people in two political party.

The feature number of people in two political parties, gender, race and ethnicity, total number of employment and unemployment, work types, and transportation tools are not originally in the dataset. These features are aggregated statistics of each county through the whole dataset.

The feature total number of votes represents the accumulated sum of the number of people who participated to vote during the presidential election in each county. The feature gender represents the total number of male and and females that correspond to the county population of the dataset. The race and ethnicity represents the total number of people in different races and ethnicities. It includes the number of hispanic/latino, asian, white, black, native, and pacific. The feature Covid-19 cases represents the total number of cases being recorded during the time period of presidential election. The feature Covid-19 deaths represents the total number of deaths being recorded during the time period of presidential election. The median household income represents the median annual household income of each county. The feature voting age citizens represents the number of citizens that are eligible to vote during the presidential election in each county. The feature poverty rate represents the percentage of population that is under the poverty line of each county. The total number of employment and unemployment represents the total number of people who are employed and unemployed regarding the dataset, which is aggregated through the number of employed and unemployment of each county. The feature work types represent the total number of people whose job categories are professional, service, office, construction, and production regarding the whole sample dataset. The feature transportation tools represent the transportation tool of the total number of people that took to work regarding the dataset. There are 5 categories for this feature: Drive, Carpool, Transit, Walk, OtherTransp and WorkAtHome. The feature number of people in two political parties represents the total number of people who voted for Trump or Biden during the presidential election.

With the usage of this dataset, we can look for a connection between the presidential election with the Covid-19 and demographic statistics. More precisely, we can visualize the factors that may influence the result of an election. Or associate the race and ethnicity of residents in a county and economic factors to their political stands. Therefore, we can generalize the pattern and use it to predict on the similar dataset.