5/2/2023

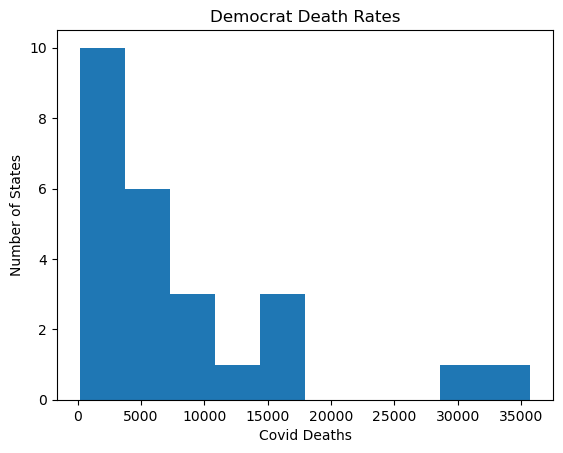
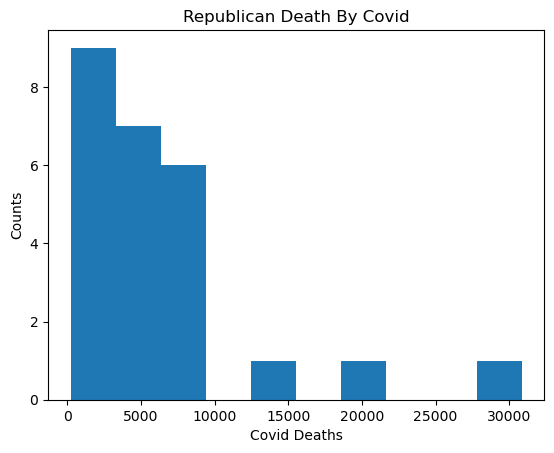
Marquis Sills

Project One

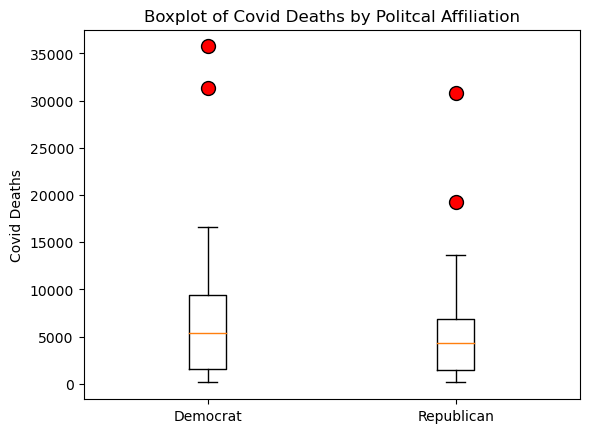
The goal is to determine the death rates of covid in the United States. To determine this, we gather as much relevant information as we can. I gathered information on demographics and education per USA state. The first question was does political affiliation affects likelihood of contracting covid. I performed web-scrapping techniques from a Youtuber named, “Python Simplified” video, <https://www.youtube.com/watch?v=oF-EMiPZQGA>.

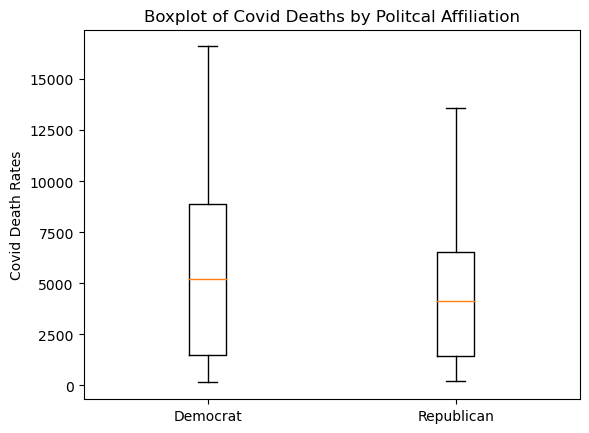
Using the technique, I extracted the number of votes per state. Using python coding I divided the states on two parameters, states that voted for Joe Biden the most vs states that voted for Donald Trump the most. Once the groups have been separated, I performed statistical analysis. Before performing the two-sample t-test we must test to see if the distribution among the groups is normally distributed.

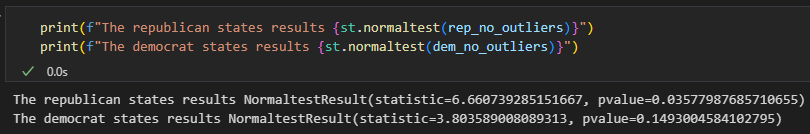
The mean covid death rates for republican are 5913.52. The median covid death rates for republican are 4336.0. The standard deviation of covid deaths for republican are 6621.29. The mean covid death rates for democrats are 8086.48 The median covid death rates for democrats are 5432.0. The standard deviation of covid deaths for democrats are 8970.08. The mean and the median aren’t close to each other and the standard deviation is large. The large standard deviation means that covid death rates among states will have large variability.



The normal distribution for both Republican and Democrat, are right skewed indicating that neither distributions are normally distributed. This indicates an outiler is present for both democrats and republicans for the death rates.



The boxplots for democrats and republicans have two outliers; the outliers represents the states in the USA. The outlier states for democrats are California and New York. The two outliers for republicans are Florida and Texas. These four states have large populations, which might be the reason why they are outliers. Once those outliers are removed, we test to see if the distributions are normal. 

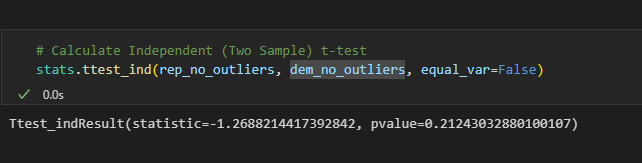


Even after removing the outilers the republican states covid deaths yield a not so normal distribution but the democrat states covid deaths are normally distributed.

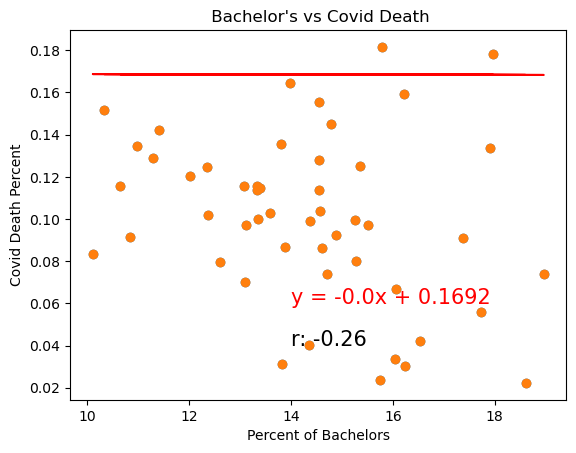
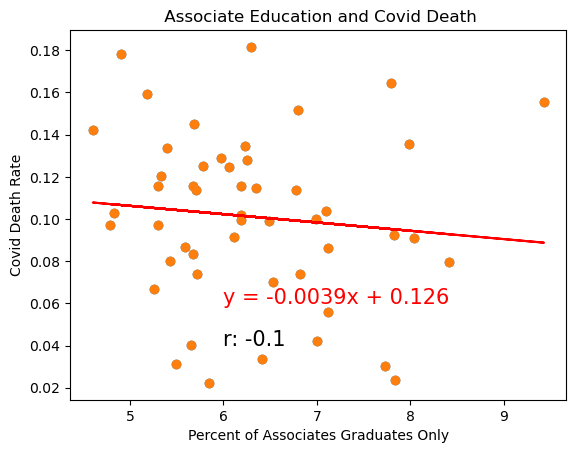
Before performing the two-sample t-test, we must write down the Hypothesis test.

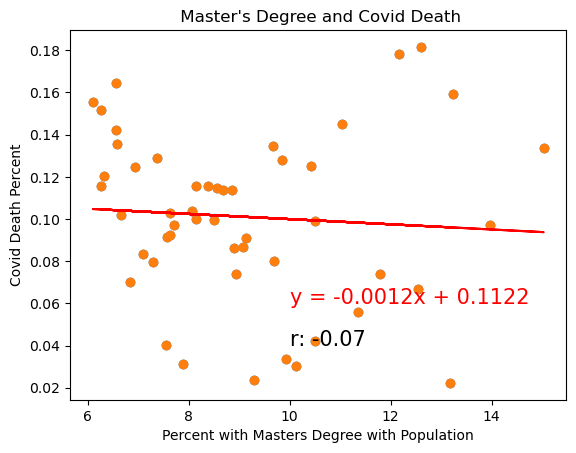
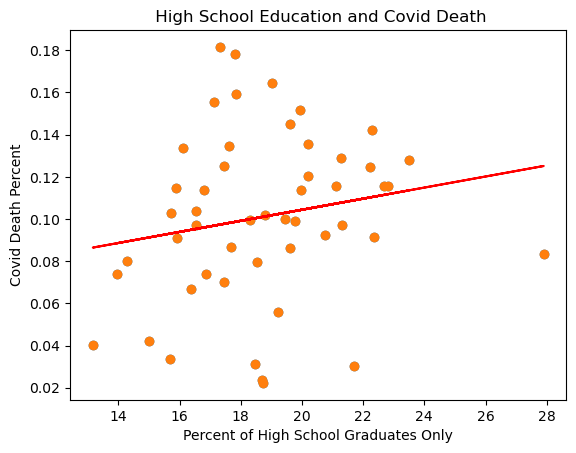
Ho: The number republican’s vs democrat’s death rates are the same.

Ha: The number republican’s vs democrat’s death rates are not the same.

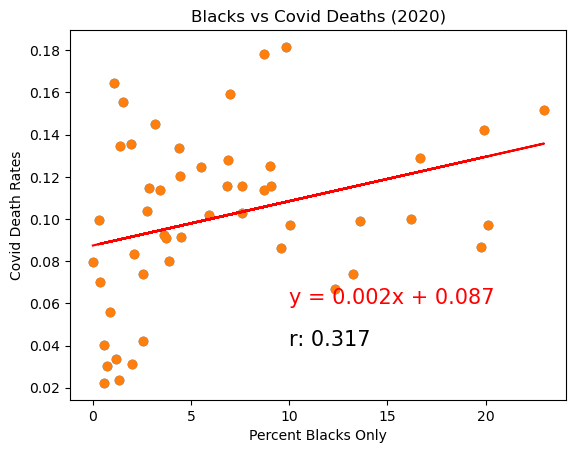
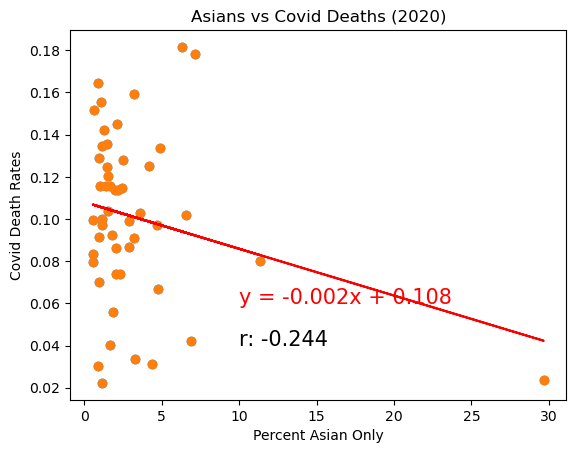


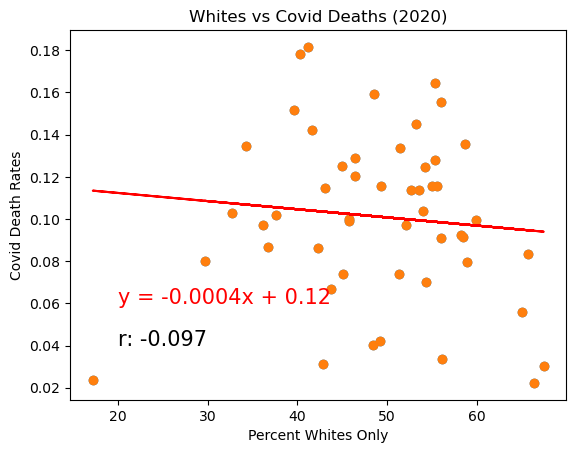
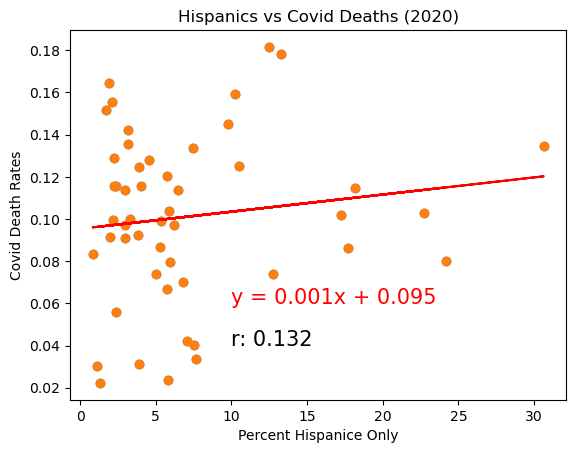
Based on our analysis, we fail to reject the null hypothesis therefore political affiliation is not factor for covid deaths. Since political affliation is not a factor for covid deaths we need to find other potenial variables. Perhaps demographics or education levels might be a factor for covid deaths.Instead of performing statistical test per variable we check the correlation between the death rates and the variables. The first varaible we test is education.





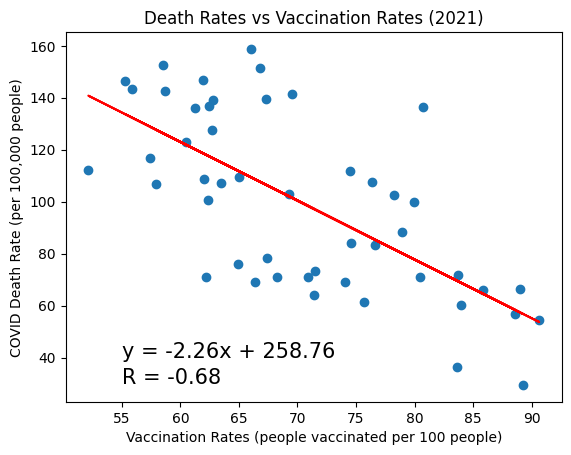
The graphs show in y-axis are the covid deaths divided by the population and the x-axis is the edcaution atainment divided by the population. For people with a high school education there is a weak positive correlation meanwhile people with assiocitates bachelors and master’s degree have a weak negative correlation. It might appear that education could play a part in covid deaths but the problem is that for all education atainment levels the correlation is weak. Therefore education level is not a factor for covid deaths. Lets see if there is a correlation between covid deaths and the demographics.





As with the previous analysis the y-axis is the population that died of covid in each state divided by the population of each state. The x-axis the total ethnic group’s population of each state divided by the population of each state. Much like the previous analysis there are weak correlations regardless or not they are positive or negatively correlated. Therefore, we can discount both education obtainment and demographics as a factor for covid deaths.

Thanks to Chris Finklea analysis we found the best variable to determine covid death rates.



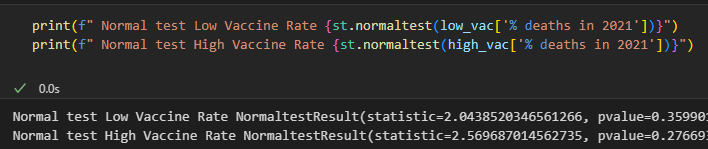
The correlation between vaccination and covid death rates negatively correlated. The higher the vaccination rate the lower the covid death rates will be. With this in mind we perform a second statistical analysis.

We picked 10 states with the lowest vaccination and picked 10 states with the highest vaccination rate. According to the correlation graph the more vaccinated the population is, the lower the death rate will be. The boxplot confirms these findings.

Before performing the two-sample t-test, must check if the distribution are normal. The Hypothesis Test gos as follow.

Ho: The given data is normally distributed.

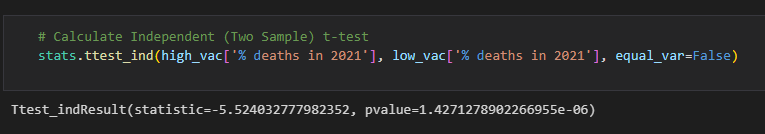
Ha: The given data is not normally distributed.



Both groups “Low Vaccine Rate” and “High Vaccine Rate” are normally distributed. Which means we perform the two-sample t-test. The Hypothesis test goes as followed.

Ho: The death rates between the unvaccinated and the vaccinated people are the same.

Ha: The death rates between the unvaccinated and the vaccinated people are not the same.



The p-value is less than 0.05, thus we reject null in favor the alternative hypothesis; The death rates between the unvaccinated and the vaccinated people are not the same. Thus, vaccines or people who vaccine have a lower covid death rate.