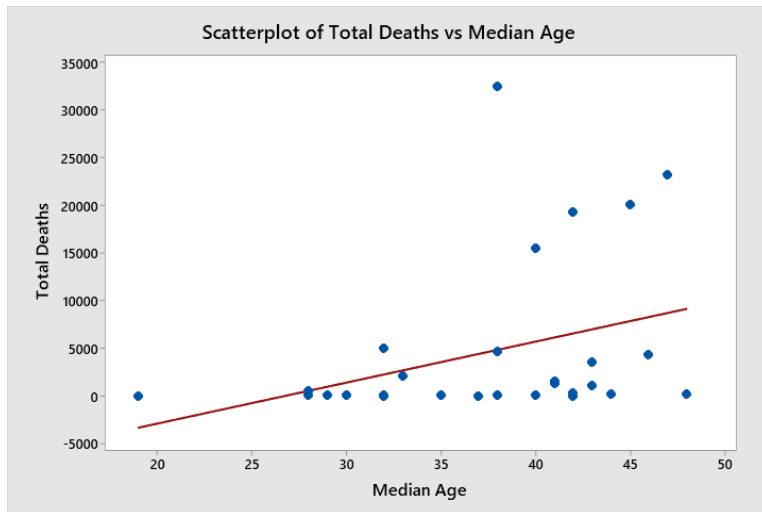


COVID-19 April 20, 2020

When I looked at the data I wanted to see if there was a differences between the sample population between the mortality rates on April 4<sup>th</sup> and April 19<sup>th</sup>. I ran a 2 sample T-Test on these two variables. Looking directly at the test, the mortality rate as of April 19<sup>th</sup> is much higher than April 4<sup>th</sup>. The mean as of April 4<sup>th</sup> was 3.57%, while the mean as of April 19<sup>th</sup> was 4.88%. With this I wonder if the Virus is more deadly because of the widespread, or is the virus just becoming more of a deadly disease.

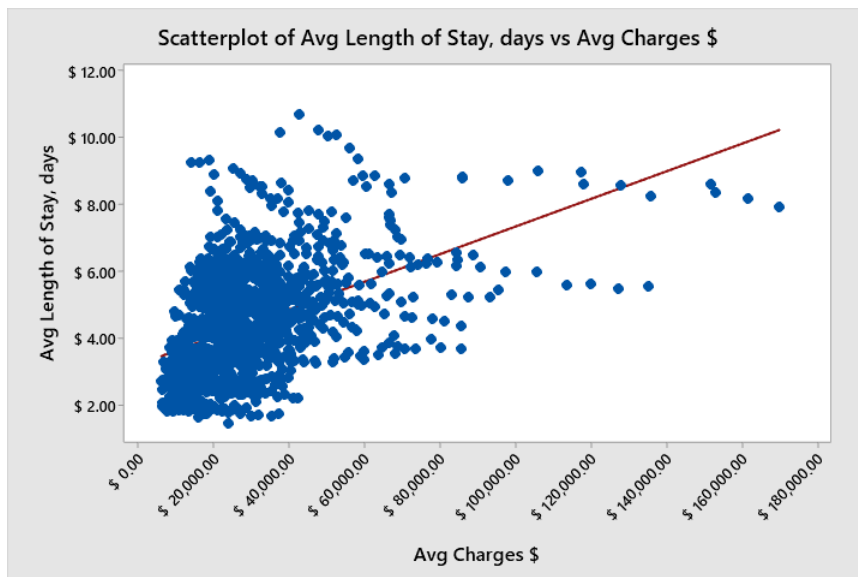
I also wanted to see if there was a correlation between the population density and confirmed cases. I made the hypothesis that as the population density increases the total cases for the area will also increase. I thought that if there are more people in an area then the virus would spread more easily. To test this, I made a scatterplot graph with regression. The data looked to have no correlation with only two outliers.

Next, I wanted to look at the correlation of the median age and the total deaths. I know that the virus is affecting older people more than younger people. I wanted to see to what extent. I made a scatterplot with regression with the Median Age on the x-axis and the total deaths on the Y-Axis. The regression is showing that there is slight increase in deaths as the median age of a country increased. Another point that I found interesting is that the country Senegal has a median age of 19 and has a death count of 3.

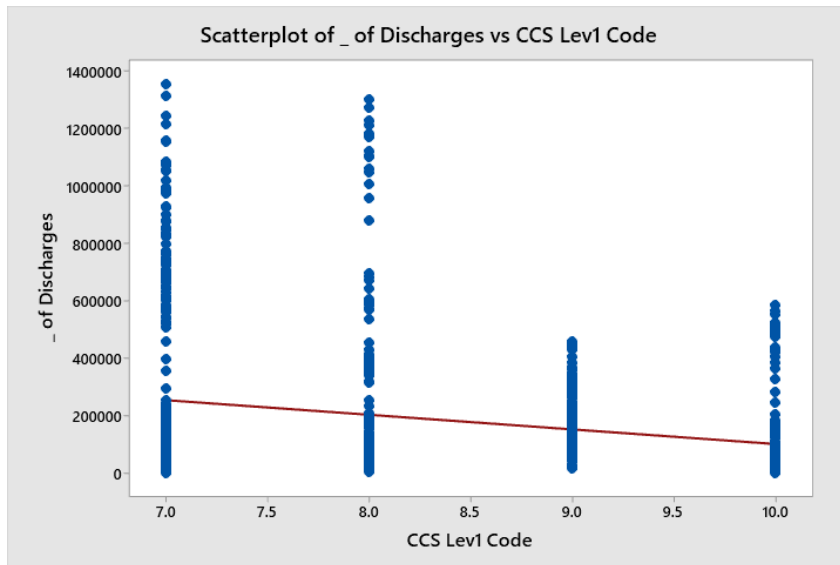


Hospital Diagnosis Discharge and Cost Data 2000-2014:

First, I looked into the correlation between Number of days spend in the hospital and the cost of the visit. I ran a scatterplot with regression to show that as the number of days people spend in a hospital increases so does the price. This makes sense because the hospital will charge people more the amount of time that they spend in the hospital and the type of treatment that they are getting.



Next, I looked at the different code in correlation to the number of discharges. I ran a scatterplot with regression showing the level code on the x-axis and the number of discharges on the y-axis. I noticed that as the level increases the average number of discharges decreases. This could prove that a higher level could show different severity of the problem.



Finally, I wanted to see if there was a correlation between the number of discharges and the average length of stay in days. I ran a scatterplot with regression with the average length on the x-axis. Looking at the data I can conclude that there is no correlation between the two variables.

