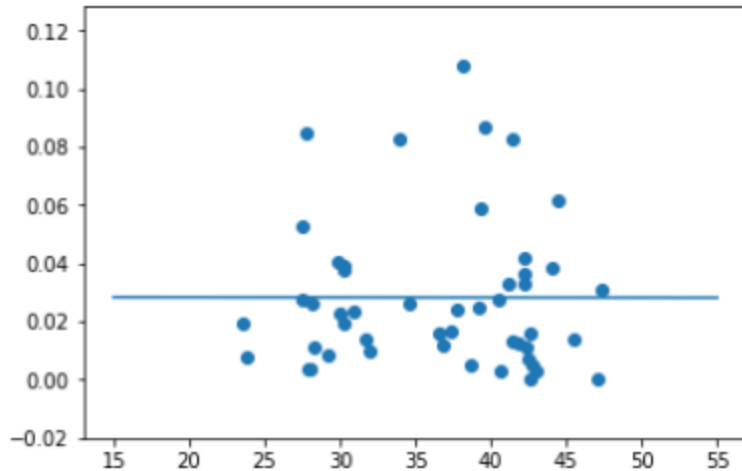


Kobe Rico  
Assignment #1

1. We can see that there is no obvious correlation between Median age and death rate for Sample Sizes of over 1000 people. However, while there is no correlation,

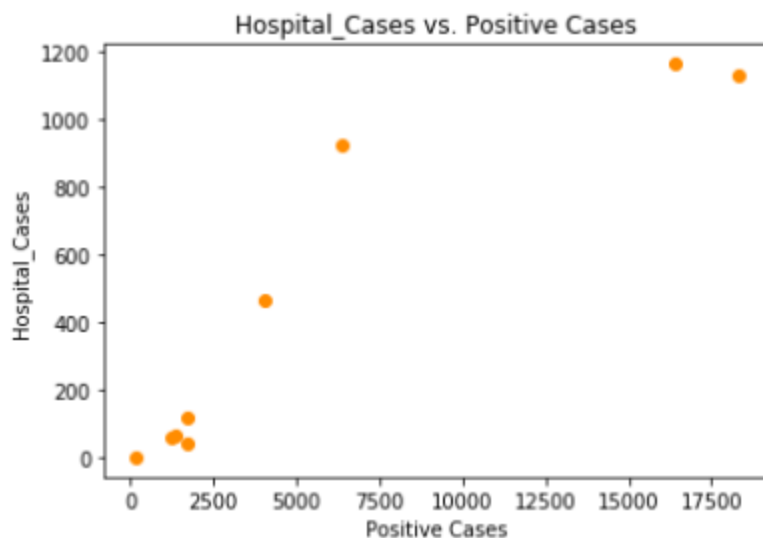
p-values: 0.9932041578331077  
R^2: 1.5600501343476618e-06  
Slope: -4.816353222423989e-06



this could directly mean that there is in fact no correlation to median age and death rate. Thus, we may be able to think that death rate and median age are not related, which is contrary to previous beliefs about the COVID-19 virus. Sharing this type of data would be imperative as many people who fall in the younger age (20's, 30's) believe they are not susceptible to COVID-19, or will not

- end up as sick as older people. While there are some spikes as we get into the late 30's and early 40's, we can still use this data in order to tell people who are of the younger age category that COVID-19 is still a serious concern for them, with the death rates being around the same for all people of the median age group.
2. My research question was what is the relationship between positive cases and hospitalization as to find out how serious positive cases are. I used the CORD-19 Data set with the United States information by state.

I first graphed all the cases of positive cases vs. hospitalization cases (by state) in a scatter plot and after seeing that there were a lot of outliers. Decided to limit positive cases to maximum 20000 cases and maximum of hospitalization cases to 1750 so that the graph was more representative of the cases.



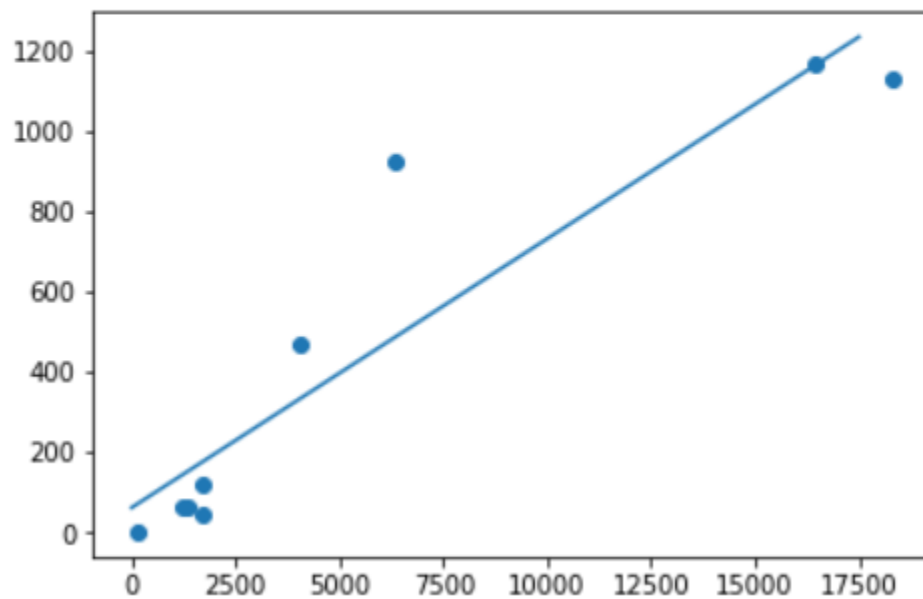
The results look promisingly linear, so we can conclude from our dataset that when there is a positive case, it is most likely to be bad in terms of health because most people appear to be hospitalized. However, these conclusions are hard to validate because of the nature of COVID-19 since some people are asymptomatic while others are not allowed to get tested in some states because they are not in the most at risk groups.

The  $R^2$  value is a small positive number implying that the dependent variable is positively related to the independent variable. Moreover, since the p value is  $< 0.05$ , this implies that the findings may be significant statistically and may be something we see in the future as well when it comes to bad cases of COVID-19 and that they are common.

p-values: 0.00030296378838843755

$R^2$ : 0.8617224876075458

Slope: 0.06696973068503236



3. I decided to take this course because I have never taken machine learning and wanted to see what it is like. Moreover, I wanted to learn how to work with data and interpret it for COVID-19 in order to contribute to the research being done. The course appears interesting and has topics that I have never been exposed to. I am unfamiliar with github and pandas and wish to learn more about both. I am hoping the tutorials and exercises will help me with the COVID-19 project I am participating in.

This assignment took me about 7 hours to complete.