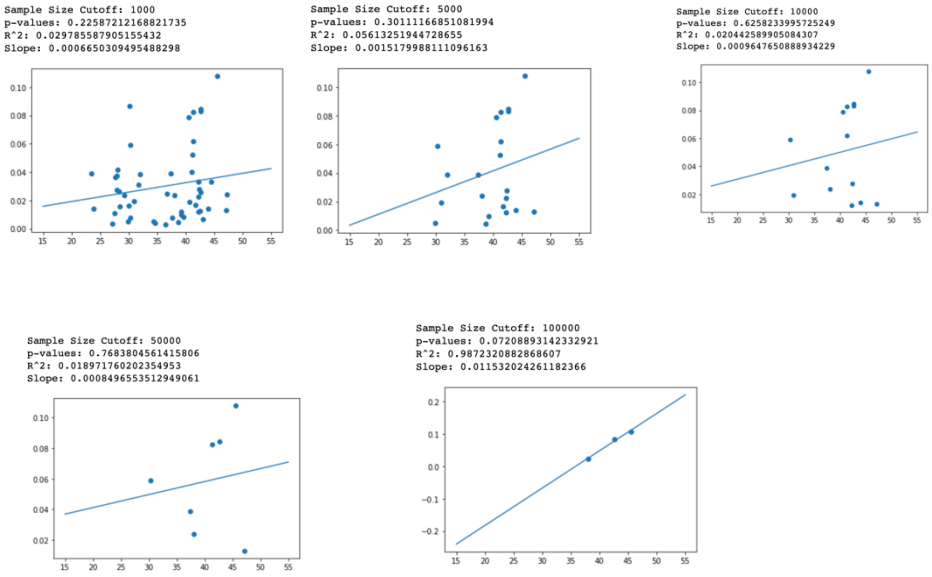
TASK 1: Sample Cleaning

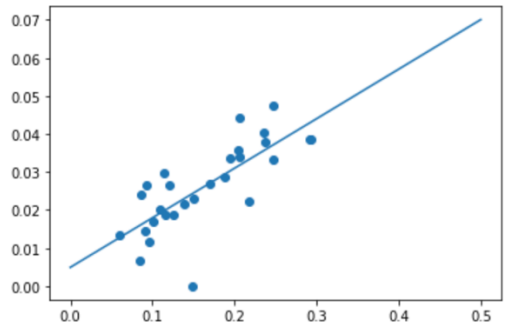
When changing the sample size cutoff, I noticed that the p and R^2 values fluctuated with little pattern. By the time that I got a correlation anywhere near significant I was dealing with only 3 countries. I am of the opinion that sample size is incredibly important in this situation, but there are not enough countries with the amount of cases we would need to guarantee an accurate representation of reality. Below are 5 regressions with varying cutoffs for the amount of cases.



Task 2: Independent Investigation

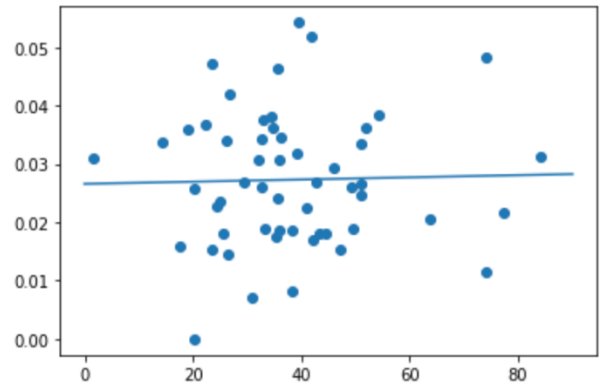
The first investigation that I wanted to do was see if some states are better at treating their patients than others. This can be expressed by the relationship between death rate and hospitalization rate. If we see that there is no relationship between the rates, this means that individual states health care systems are performing differently. However, we see that there is statistically significant moderate correlation between hospitalization rate and death rate. This means that someone with the same severity of infection is likely no better off in a different state.

```
p-values: 3.0794866822778493e-06
R^2: 0.5735428376787193
Slope: 0.13028239401087635
```



A widely held belief regarding COVID is that it will dissipate in the summer despite many scientists are warning us that this is likely not the case. I gathered the temperature in 54 US states and territories in March 2020 and ran a regression between that and the death rate. I found that there was very low correlation. This included all states that had deaths, however, and the sample size was very low for many of these states. I wouldn't conclusively say that temperature has no effect on the severity of COVID, but there is not much evidence for the affirmative at this point.

```
p-values: 0.8518668193525989
R^2: 0.0006768399092157327
Slope: 1.8580551420878277e-05
```



Task 3:

I was on the fence about taking the full version of this class at the start of the semester but opted out in favor of a class in the area of my humanities concentration. Given that I am lucky enough to have much more free time now, and that my interest in machine learning remains, enrolling in this class was an easy decision. I hope to learn how these big data tools are used and how they work, and to pick up some practical experience using pandas and jupyter notebooks.

This assignment (including time spent reading all of the pandas and jupyter documentation) took around 2.5 hours.