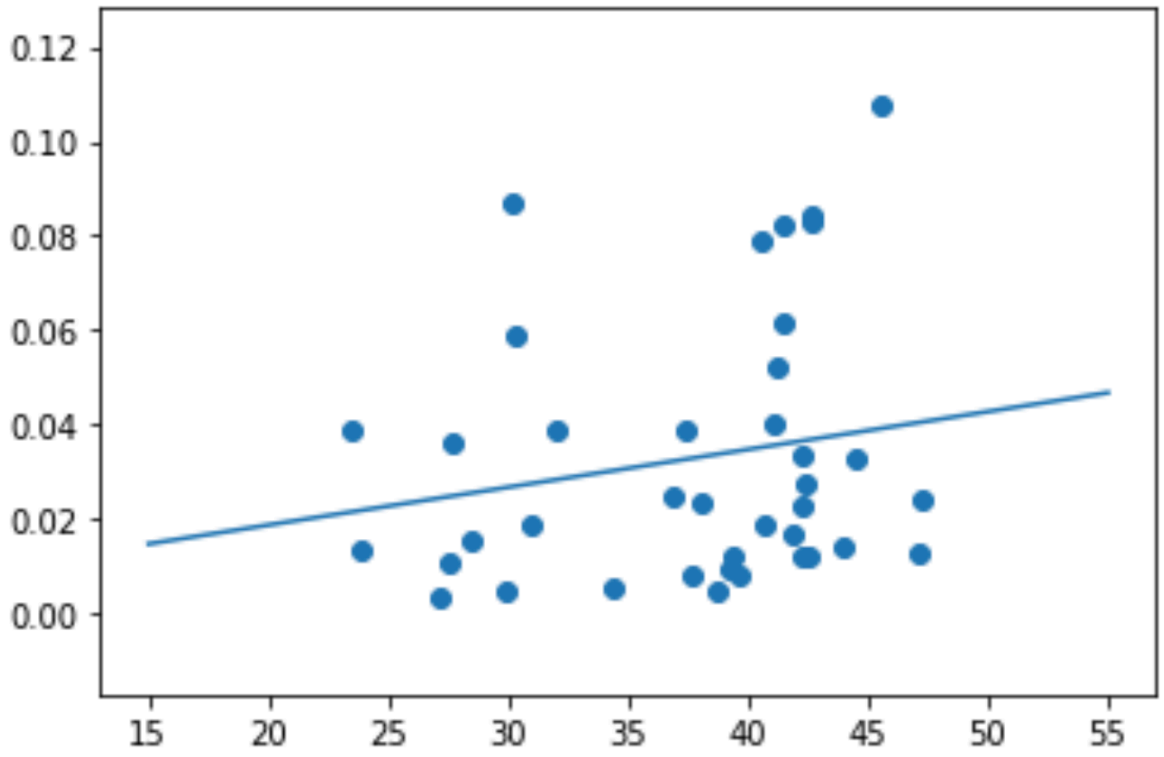
1)

```
p-values: 0.2511999715361453
R^2: 0.03542419131914649
Slope: 0.0007994940475298446
```



Results, Interpretation, and Discussion.

Results

The results of my linear regression yielded a slope of 0.0007994 death pe, r-squared of 0.034525 and p-values of 0.2512. The regression line has a positive trend.

Interpretation

The data is not very linear as we had an r-squared value of 0.035, meaning that the majority of our data strayed from my regression line. Furthermore, median age is not a good way to predict the death rate according to our data because our p-value is greater than 0.05. This means that

changes in our predictor are not associated with the changes in the response and is not statistically significant.

Discussion

I filtered the data based off of countries that had at least 1000 cases by March 28th, this may not be the best way to check whether or not our predictor works well because we only have a few days' worth of data to work with. If I used an earlier filter date and a similar sized threshold my predictor may have been more accurate. We can't conclude much from this regression.

2)

Research Question

Does temperature effect the number of corona virus cases.

- plot log of cases per day versus temperature

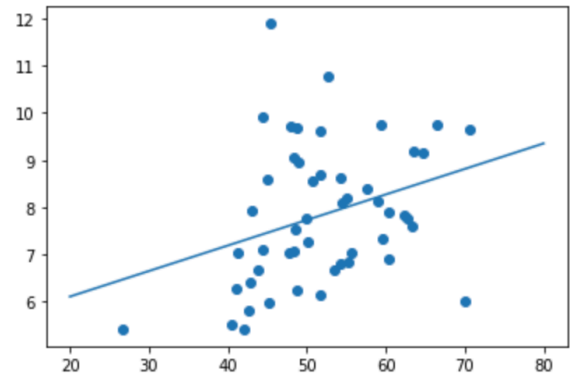
Data Source

The Covid Tracking Project's current data. This contains the current count of cases and more for each American state and territory.

Method and Results

To plot my data, I looked online for the average temperature per state. I then copied this data into a dictionary in my notebook and planned to add it into my dataframe. I then loaded the current data csv and removed the unnecessary columns as well as rows territories and the District of Colombia and grouped by state. Then I took the logarithm of each current count to get a more accurate relationship between each state, however, I am not sure whether I could do this and end up with a valid result. I then appended the average temperatures to a new column in my data frame and then plotted my data resulting in the following graph.

```
p-values: 0.024010895806925094
R^2: 0.10167452072810751
Slope: 0.054105929558857034
```



The graph has a slope of 0.054105 cases per Fahrenheit, p-values of 0.024011, and an r-squared value of 0.1016. This tells us that our predictor of temperature appears to be a good predictor in terms of the number of total data points. However, I doubt this is true due to the logarithm that I applied to the total number of cases. On top of this the squared value is very small, which indicates that most the points do not fall along the regression line. While the p-values indicate this results may be statistically significant, I do not believe my results to be true because of the questionable logarithm I applied as well as the fact that I used the average temperature of each state over the course of a year rather than the temperature of each state at the moment. This

model could be improved by using a temperature API and the historical data provided by the COVID Tracking project and plotting each temperature and increase in number of cases in each state. This would provide a more accurate model of the effect of temperature on total daily infections.

3)

I wanted to take this course because I wanted to provide means to visualize data for others and try to help inform and slow the outbreak using my skillset. I have also always been interested in Data Science and Machine Learning, so have the ability to combine this interest with the pandemic we are facing encouraged me to take the class. This assignment took me about 3 hours.