**Does Higher Population Lead to Greater CO2 Emissions?**

At the end of my exploratory data analysis of CO2 emissions in relation to population, I can conclude that there is a positive relationship between emissions and population. However, causation was not proved and there are many other factors which have a statistically significant effect on CO2 emissions as well, such as GDP. This leaves open the possibility that there is/are an underlying reason(s) aside from population that leads to greater CO2 emissions. Given the incredibly high correlation between energy consumption and emissions, further analysis should investigate the relationship that economic production has on predicting CO2 emissions compared to the impact of individual persons.

One thing I feel I missed during the analysis was a time series analysis. I think looking at emissions over the course of the past forty years and comparing that to population changes would have proved useful in the analysis. However, such a task is beyond my current capabilities as it involves comparing time series with different units of measurements that change at different rates.

I think a variable that gave a percentage of energy produced coming from renewables would have been helpful. With such a variable, in combination with a time series analysis, I could investigate the cumulative impact of population on emissions historically compared to contemporaneously.

One assumption I made that I feel may have been incorrect was that not removing the outliers would aid my analysis. It would have been good to try removing the outlier cases and observing changes in emissions at the lower end of the distribution (between 0 and 2500 mmt CO2). Perhaps population plays a greater role in emissions on a smaller scale compared to emissions at the extremes. I fear that in leaving the outliers alone I may have exaggerated some of the statistical effects. I should have tried transforming the data to minimize the effects of outliers.

I had a really difficult time with the modeling distributions. I still do not understand how to find one that fits the distribution best, though that stems from a further problem of struggling to grasp the underlying statistics at play. I also struggled when creating plots to make the data appear visually appealing and helpful. The data was so skewed it was difficult to show relationships clearly.

Overall, while I faced a few challenges when it came to the technical parts of the project, I feel I found some interesting statistical effects and that my EDA, while incomplete and only part of a much larger analysis, was successful.