Hypothesis: Countries with higher urban populations have larger yearly increases in temperature.

The outcome of the EDA was inconclusive. The correlation coefficient between the two variables showed only a weak positive linear relationship and no evidence of a non-linear relationship. The R-squared value showed that only a small amount of the variance in temperature was accounted for by urban population.

It was determined to keep all outliers in the population data at the beginning of the analysis. After reviewing the results, it seems that those outliers could have been handled in a different manner. If it were done over, the very lowest population values (zero) would be removed and possibly re-examined after an analysis of the remaining data points. It would also be helpful to search for possible correlation before choosing a variable to increase the likelihood of finding potential causation.

Historical data of temperatures could have been helpful. Having temperatures from times earlier than the 1990s could have helped to determine a baseline for how much yearly variation in temperature already occurred before urban populations boomed.

It was assumed in the beginning that all emissions values would be positive. After analyzing the total emissions variable, it was revealed that some emissions were actually negative. Further study of how emissions could be negative might help reveal information about temperature changes.

Discovering that the data for both urban and rural populations had such a large spread with some countries reporting values of zero was challenging. I was unsure how to handle those values and although I found a way, I feel there is likely a better way to handle it. I also did not fully understand how the p-value could be significant with the correlation coefficient and R-squared values being so low. A better understanding of autocorrelation would have helped me.