

**Correlation Methods and Findings**

Is there a correlation between global temperature changes and CO2 fossil fuel emissions?

The source temperature and CO2 data were cleaned, made format compatible, and filtered to reflect the same date range, approximately 1744-2012. It should be noted that not all countries began reporting on either data source at the same time. Then several correlative and linear regression plots were applied.

A correlation coefficient measures the extent to which two variables tend to change together. The coefficient describes both the strength and the direction of the relationship.

The Pearson correlation evaluates the linear relationship between the two continuous variables for temperature and CO2. The range of a Pearson plot is -1 to +1.

Our Pearson shows a medium strong correlation coefficient of .56588 when date range starts at 1744 for both data sources.

A Lag Plot checks whether a data set is random or not. Our lag plots for each data source by year(linear) shows a tight positive diagonal correlation.

Lag plots for the same data sources grouped by country instead of year present a scattered much weaker correlation. We attribute this to the fact that data grouped by country has no clear linear definition therefore no correlation pattern.

Autocorrelation is another plot to validate data source. Both data sources present the same arc and curve around the central axis. This is another data correlation indicator.

Histograms are used to plot the frequency of scored occurrences in a continuous data set that has been divided into classes, called bins.

A frequency distribution table is a table that shows how often a data point or a group of data points appears in a given data set. To make a frequency distribution table, first divide the numbers over which the data ranges into intervals of equal length. Then count how many data points fall into each interval.