Project 1 Analysis

Intro

For this project, my task was to see if there was any correlation between a country’s non-renewable energy use and its air quality. Using the cleaned data from all the datasets we used (as well as separate Dataframes by country that were derivative of the original Dataframe), I conducted the following analysis.

First Slide

The first thing that I did was compare the total non-renewable energy usage to the total particulate matter by each country for each year. After plotting the resulting data on a scatterplot, I performed a Linear Regression test to determine the correlation, and a t-test to determine the statistical significance.

The results of the t-test gave us a p-value that was much less than .05 and a t-statistic of a little over 6.79, which tells us that the Non-Renewables Value and Particulate Matter variables are significantly related and there’s a very low probability that the relationship between the two is due to chance alone. However, the results of the linear regression gave us an r-value of just over .03. This means that overall, there is a significant, but very weak positive correlation between the Non-Renewables Value and Particulate Matter variables.

Second Slide

Next, I created Dataframes for 6 selected countries from our original Dataframe and performed the same linear regression test and t-tests for all of them. I then plotted 6 different scatterplots and grouped them in the same grid for better visualization.

After all tests were performed, the results were quite different than the previous visual. Every r-value for each country was over .7 and the p-value for every country was below .05, which tells us that all selected countries showed a statistically significant and strong positive correlation.

Insights Slide

There are a couple conclusions we can draw from the visualized data. Firstly, to the surprise of no one, it’s safe to say the countries that use more non-renewable energy sources have worse air quality, except if your name is America.

America is quite the outlier when compared to the other selected countries because despite the high use of non-renewable energy sources in the United States, the total amount of particulate matter in the country’s air from 2010 to 2019 was significantly less than that of the other selected countries. This is probably due to a number of external factors such as strict new regulations on sources of non-renewable energy pollution, a massive shift to renewables in recent years and the way the country is laid out geographically (since urban areas are typically more polluted than the vast open areas of the country such as the Great Plains.