**Introduction**

Air pollutants are derived from many sources, but are not always visible to the naked eye. According to the United States Environmental Protection Agency (EPA), air pollution can be defined as a combination of gases and particles in the atmosphere that can reach harmful concentrations and have a negative effect on the health of people, animals, and the environment. There have been considerable air pollution prevention and control measures put in place over the years. The focus of this data study is to follow the trend of pollution in the United States over a 17-year period to determine if there are significant differences in pollution among U.S. regions, and if there is an upward trend in pollution levels in the U.S.

**Hypotheses**

**TEST 1:**

**Null Hypothesis**: There is not a linear trend in pollution levels in the U.S. between the years 2000 to 2015.

**Alternative Hypothesis**: There is a linear trend in pollution levels in the U.S. between the years 2000 to 2015.

**TEST 2:**

**Null** **Hypothesis**: There is not a linear trend in pollution among U.S. regions between the years 2000 to 2015.

**Alternative** **Hypothesis**: There is a linear trend in pollution among U.S. regions between the years 2000 to 2015.

**Motivation and Questions**

Our group was motivated by the reports of significantly decreased air pollution in major cities around the globe after the implementation of social distancing and shelter-in-place orders. This led us to research the trends in air pollution to answer the questions; Is air pollution decreasing year over year? And is there a difference in air pollution by region?

**Data Summary**

The data set that we analyzed was sourced from data.world (<https://data.world/data-society/us-air-pollution-data>). With a file size of 411.49MB, the dataset contains 1,746,661 rows and 28 columns of air pollution data documented by the U.S. EPA on a daily basis from 2000 to 2016. The dataset contains information about four pollutants: Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2), Carbon Monoxide (CO) and Ozone (O3). It describes the following information for each of the pollutants in 5 different columns: units measured, mean of concentration, air quality index, maximum value of concentration in a given day, and hour when the value was recorded. In addition, the dataset includes the state, city, address, and date of monitoring.

**Data Exploration & Cleanup**

The U.S. Air Pollution data was generally clean and did not require a lot of manipulation. We removed “Country of Mexico” and all related data as our study focused solely on the United States. We also made the decision to choose 1 pollutant to focus on in the interest of time. We chose to focus on the pollutant Nitrogen Dioxide as it is the most harmful to humans. High concentrations of Nitrogen Dioxide in the atmosphere can irritate the airways in the human respiratory system causing coughing, wheezing or difficulty breathing. Prolonged exposure to Nitrogen Dioxide can make the lungs more susceptible to infection, asthma and respiratory disease. In addition to this, Nitrogen Dioxide can mix with other gases in the atmosphere to produce haze, acid rain and contribute to nutrient pollution in the coastal waters (epa.gov).

Data was grouped two ways for analysis of each hypothesis. First, data was grouped by year to completed the year-over-year analysis. Next, data was grouped by regions to complete the regional analysis. States were grouped into regions based on the regional chart located on the U.S. Embassy & Consulate in the Republic of Korea website. The four regions were; West, Midwest, South and Northeast.

**Analysis**

Analysis was done in 2 phases. First, we analyzed the data in accordance with hypothesis 1 (year-over-year). Then we analyzed the data in accordance with hypothesis 2 (regional).

**Conclusions**

Based on the analysis, we found that there is a downward linear trend in pollution year over year. NO2 pollution appears to be decreasing overall.

Visually we can conclude that the South region is different than the Northeast and West regions. However, the results of the T tests revealed that, with a significant value of 0.05%, we can statistically reject the null hypothesis. The regions are not the same in terms of NO2 pollution average.

**Implications**

There are some limitations with this analysis. The first being that the data did not include observations for the states of Mississippi, Montana, Nebraska and Vermont. The second limitation is that the downward trend cannot continue pass 0.

Citations

“Nitrogen Dioxide (NO2) Pollution.” *EPA*, Environmental Protection Agency, 13 June 2019, [www.epa.gov/no2-pollution](http://www.epa.gov/no2-pollution).

“Nitrogen Dioxide (NO2).” *Department of Agriculture, Water and the Environment*, [www.environment.gov.au/protection/publications/factsheet-nitrogen-dioxide-no2](http://www.environment.gov.au/protection/publications/factsheet-nitrogen-dioxide-no2).

[[<https://kr.usembassy.gov/education-culture/infopedia-usa/travel-usa/regions-united-states/>]]