**Section A: Introduction/Background**

**Overview of the Dataset**

The dataset used in this project contains environmental and demographic data collected from 60 metropolitan areas in the United States. The primary focus of the dataset is to analyze factors associated with mortality rates in these areas. The total age-adjusted mortality rate from all causes (MORT) represents the number of deaths per 100,000 people. Although MORT itself is not included as a variable in the dataset, a grouping variable, **MORTRANK**, is derived from it. This grouping allows for classification into three categories:

• **MORTRANK = 1**: Areas where mortality rates are below 912 deaths per 100,000 people.

• **MORTRANK = 2**: Areas where mortality rates range between 912 and 968 deaths per 100,000 people.

• **MORTRANK = 3**: Areas where mortality rates are 968 or higher per 100,000 people.

The purpose of this dataset is to analyze whether environmental and socioeconomic factors contribute to higher or lower mortality rates across different metropolitan regions.

**Description of Variables**

The dataset includes the following variables:

• **MORTRANK** *(Grouping Variable)*: Represents the classification of metropolitan areas based on their mortality rates.

• **PRECIP** *(Mean Annual Precipitation in inches)*: This variable measures the average annual precipitation in each metropolitan area. Higher or lower precipitation levels might influence environmental factors related to health.

• **EDUC** *(Median Number of School Years Completed for Individuals Aged 25 and Older)*: This represents the median education level in each area, which could be a key indicator of socioeconomic status and access to health-related information.

• **NONWHITE** *(Percentage of Population That is Non-White)*: This variable represents the racial composition of each metropolitan area. Demographic factors may be correlated with health outcomes due to social determinants of health.

• **NOX** *(Relative Pollution Potential of All Nitrogen Oxides)*: This metric represents the estimated pollution impact of nitrogen oxides in each metropolitan area. Higher NOX levels might be associated with respiratory and cardiovascular diseases.

• **SO2** *(Relative Pollution Potential of Sulfur Dioxide)*: This measures the estimated pollution potential of sulfur dioxide. Like NOX, higher SO2 levels could indicate poorer air quality, potentially leading to adverse health outcomes.

**Purpose of the Analysis**

The goal of this project is to investigate the relationships between these environmental and socioeconomic variables and mortality rates. Specifically, we aim to determine which factors might contribute to higher or lower mortality rankings (MORTRANK). By analyzing these variables, we can assess whether pollution levels, education, racial composition, and climate factors are significant determinants of health outcomes in urban settings.