SINGAPORE INSTITUTE OF TECHNOLOGY

Visualization PM2.5 Concentrations in the World (2017-2019)

Ng Wei Herng, Timothy Zoe Delaya, Clarence Agcanas, Yeo Song Chen, Lee Ru Yuan

I. Original Data Visualization in News Media

The visualization titled "Comparing PM2.5 Concentrations in Capital Cities" created by Pallavi Rao (2023) and published on "The Visual Capitalist", presents a snapshot of PM2.5 air pollution levels in various capital cities around the world for the year 2022. PM2.5 refers to particulate matter that is less than 2.5 micrometers in diameter, which is small enough to penetrate the lungs and enter the bloodstream, posing significant health risks.

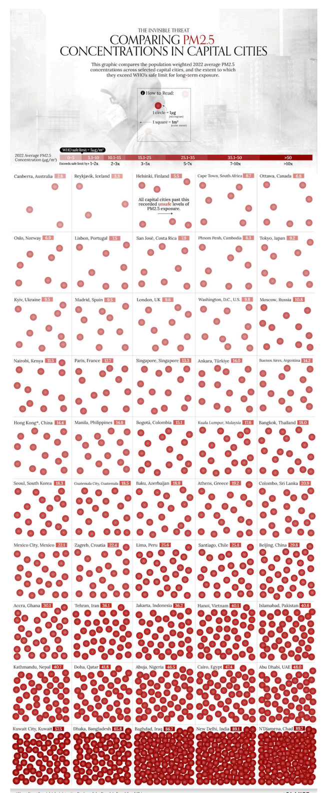

Figure 1: Visualized: Air Quality and Pollution in 50 Capital Cities (IQAir 2022 World Air Quality Report)

This visualization uses a series of red circles to represent the PM2.5 concentrations in each capital city. The number of circles corresponds to the level of PM2.5 concentration that exceeds the World Health Organization's (WHO) safe limit for PM2.5, which is 5 µg/m³. Any value above this indicates a higher risk for adverse health effects.

i. Issues with the Original Visualization

The original visualization clearly shows air pollution in popular cities but lacks additional factors like population size and GDP, which could explain the pollution levels. The use of multiple red circles makes it overwhelming and confusing, hindering quick comprehension of relative pollution levels between cities.

II. Dataset

The data set we have chosen to use comes from the World Health Organization (WHO) which provides data on air quality for various countries from a wider year range. The data set contains information on PM2.5 concentrations for different countries and years and is also more precise as WHO has a 60% inclusion requirement whereby the recorded data require annual data availability of at least 60% of the total number of hours in a year to be included. Alongside this data set, we have also chosen to use 3 additional data set for the purposes of enhancing the visualization, as well as to improve on the data engineering and data cleaning aspect of the WHO data set.

1. Country Codes (2024)
2. Population Data (2024)
3. GDP Data (2024)

i. Data cleaning and preparation

We began by reading in the PM2.5 concentration dataset and filtered it to include data from 2017 to 2019. For GDP data, we skipped metadata, transformed it from wide to long format, renamed columns for clarity, and imputed missing values using previous years' data. Unnecessary columns were dropped, and the data was filtered to include 2017-2019 before saving to a new CSV file. The population data was similarly cleaned and filtered for 2017-2019, with missing values imputed from previous years. We merged the GDP and population data with the PM2.5 dataset on matching country names and years, dropped any remaining rows with null values, and saved the final complete dataset to a new CSV file for further analysis.

III. Improved Visualisation

To improve upon the original visualization, decided that we would create three visualizations to provide a more comprehensive and insightful analysis of PM2.5 concentrations globally. These visualizations include:

- **Improved Visual Appeal**: All of our visualizations are designed to be visually appealing and easy to understand, making it easier for the audience to interpret the data.
- **Additional Data**: We incorporated population and GDP data to provide context and insights into the factors contributing to air pollution levels in different regions. This additional information enhances the audience's understanding of the underlying causes of pollution.
- **Interactive Elements**: All of our visualizations are interactive, allowing users to explore the data further by hovering over data points or selecting specific countries to view detailed information.

i. Choropleth Map

The choropleth map displays PM2.5 concentrations in countries globally, with darker colors representing higher PM2.5 levels. This visualization allows for quick and clear comparison of air quality across countries. Unfortunately, due to the lack of data of air pollution, most countries are greyed out. However, the countries with data are shown in the map below.

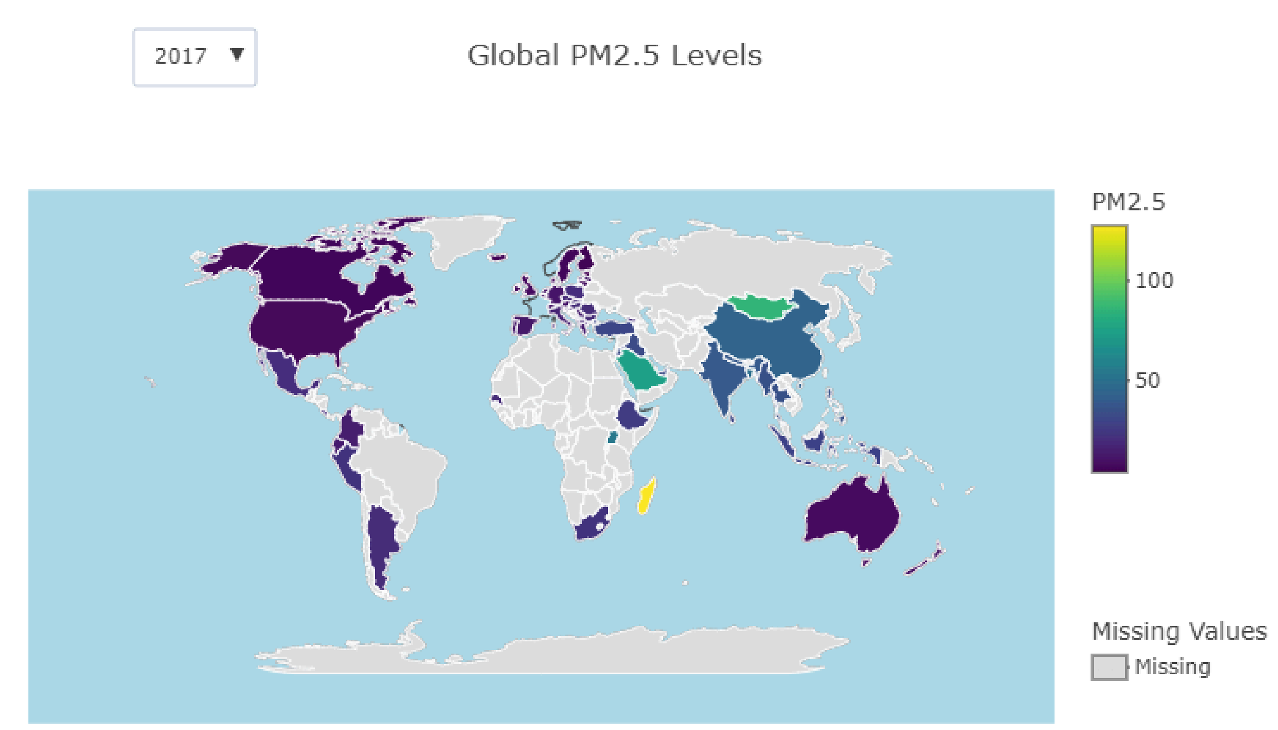

Figure 2: Choropleth Map showing the Global PM2.5 Levels

ii. Scatter Plot, PM2.5 vs GDP

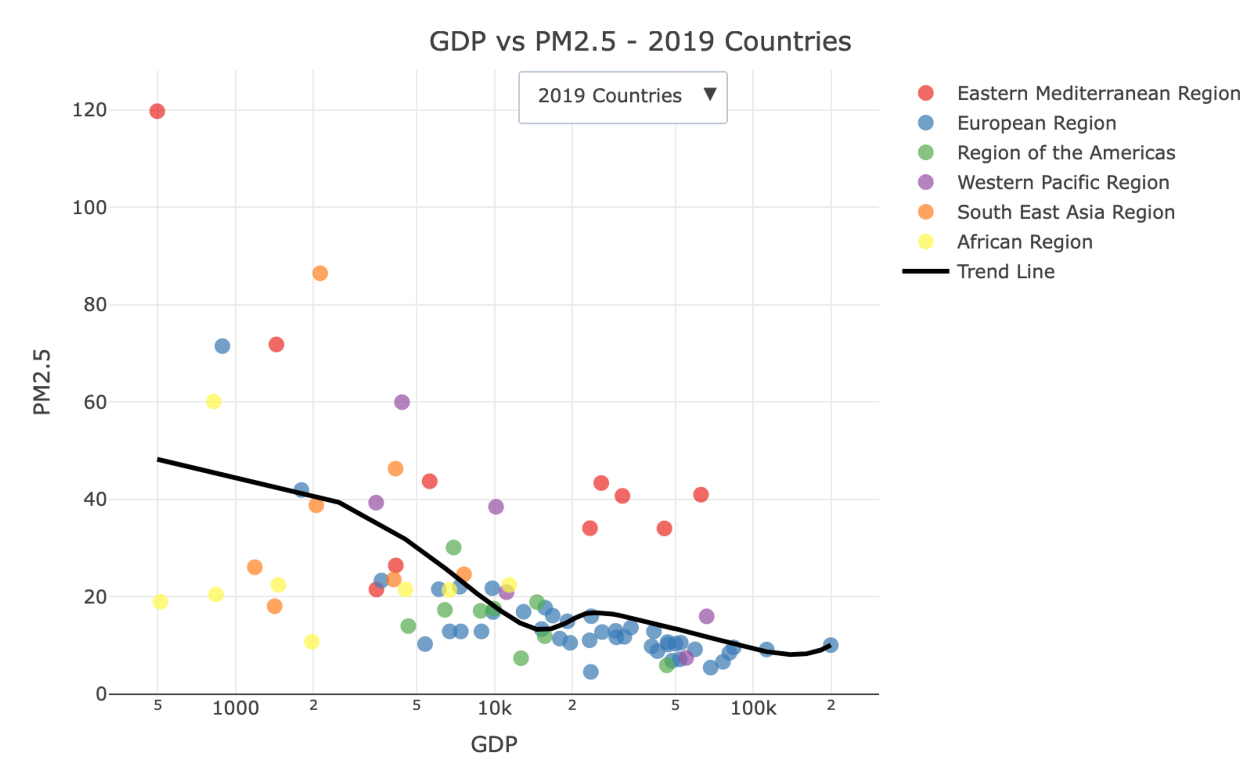

Figure 3: Global PM2.5 vs GDP

- **Averaging of Cities to Get Country Air Pollution**

To derive country-level air pollution data from city-level data, we aggregate the city-level PM2.5 concentrations by taking the average for each country. This process involves grouping the data by country and year, and then computing the mean PM2.5 concentration for each group. By doing this, we obtain a representative value of PM2.5 concentration for each country, which can then be used for further analysis.

- **Analysis using Trend Line**

The trend line provides a visual representation of the general pattern or relationship between GDP and PM2.5 concentrations. This can help to identify whether higher GDP is associated with higher or lower levels of air pollution. We use LOESS Method because it does not assume a specific functional form (like linear regression) and can adapt to the underlying data structure, providing a more flexible and accurate representation of the relationship.

iii. Scatter Plot, PM2.5 vs Population

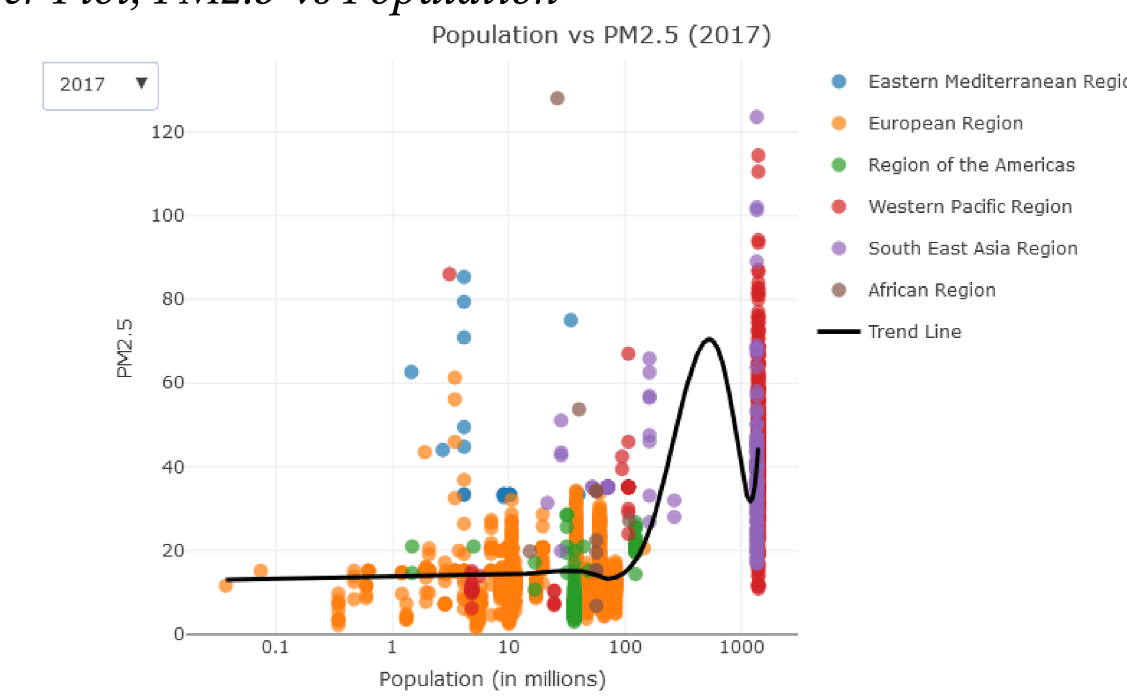

Figure 4: Global PM2.5 vs Population

This scatter plot shows the relationship between PM2.5 values and population size for each city in the country, the bubbles are color coded by their region. The trend line provides a visual representation of the general pattern or relationship between population size and PM2.5 concentrations. This can help to identify whether higher population size is associated with higher or lower levels of air pollution.

IV. Further Suggestions for Improvement

- **Incorporate More Data**: Include additional datasets such as weather data, industrial activity, and traffic congestion to provide a more comprehensive analysis of air pollution.
- **Enhance Interactivity**: Add more interactive elements such as filters, sliders, and dropdown menus to allow users to customize their viewing experience.
- **Include wider Historical Data**: Incorporate historical data to analyze trends and patterns in air pollution levels over time.

V. Conclusion

Our improved visualizations provide a more comprehensive and insightful analysis of PM2.5 concentrations globally. By incorporating additional data on population and GDP, we have enhanced the audience's understanding of the factors contributing to air pollution levels in different regions. The interactive elements allow users to explore the data further and gain deeper insights into the relationship between air pollution, population size, and GDP. Overall, our visualizations offer a more engaging and informative way to visualize air quality data and raise awareness of the importance of addressing air pollution on a global scale.