

Air Quality Index (AQI) Analysis

MTH208 - Team 7

Anusha Gupta, Shreya Tarafdar,
Sruthi Subramanian, Saroj Karwal

Introduction

- **Aim:**

Exploring the variations in global air quality, identifying key environmental and socio-economic factors that may influence it, and checking the correlation of the same.

- **Methodology:**

Data Scraping and Cleaning, Correlation Analysis, and Interactive Visualization with R Shiny.

Key Research Questions

1. Geographic Variation:

How does air quality vary across different regions of the world?

2. Socio-Economic Link:

Is there a relationship between **population density** and AQI levels? Does **GDP per capita** influence air quality?

3. Climatic Drivers:

How do factors like **temperature, humidity, precipitation, and sunshine hours** relate to AQI?

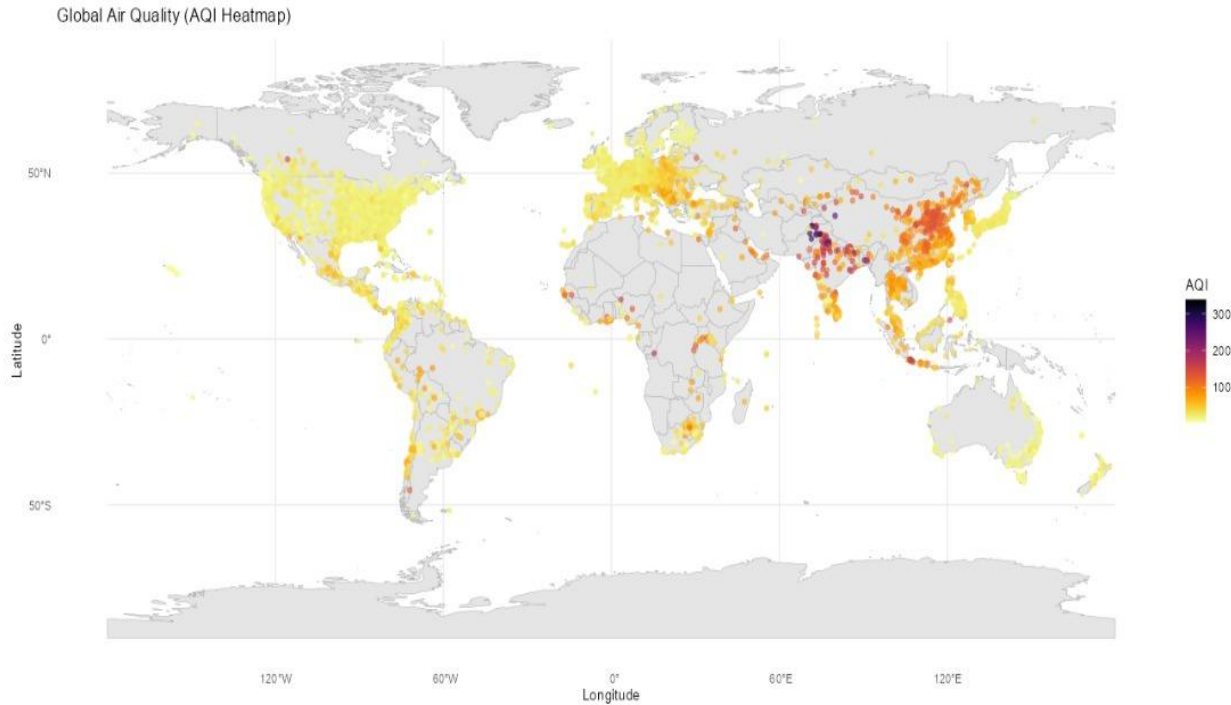
Data Acquisition and Cleaning

1. **AQI Data:** City-level AQI measurements scraped from **IQAir** (World's Most Polluted Cities).
2. **Web Scraping Tools:** **rvest** and **chromote** were used to scrape data from 6 different sources.
3. **Integration:** Merged AQI data with:
 - Climatic parameters (Temperature, Humidity, Precipitation) from weather databases.
 - Demographic (Population) and Economic (GDP per capita) data.
4. **Challenge:** Integrating diverse data based on country/city names and handling inconsistencies with such large amount of data.

Interactive Visualization using R Shiny

- The project used the **R Shiny** library to build an app with interactive graphs.
- **Goal:** Allow users to explore complex relationships dynamically and promote data-driven awareness.
- **Key Visualizations:**
 - Global AQI Heatmaps (Geographic Distribution).
 - Lollipop Plots (Correlation Overview).
 - Dynamic Scatter Plots (Parameter vs. Mean AQI).

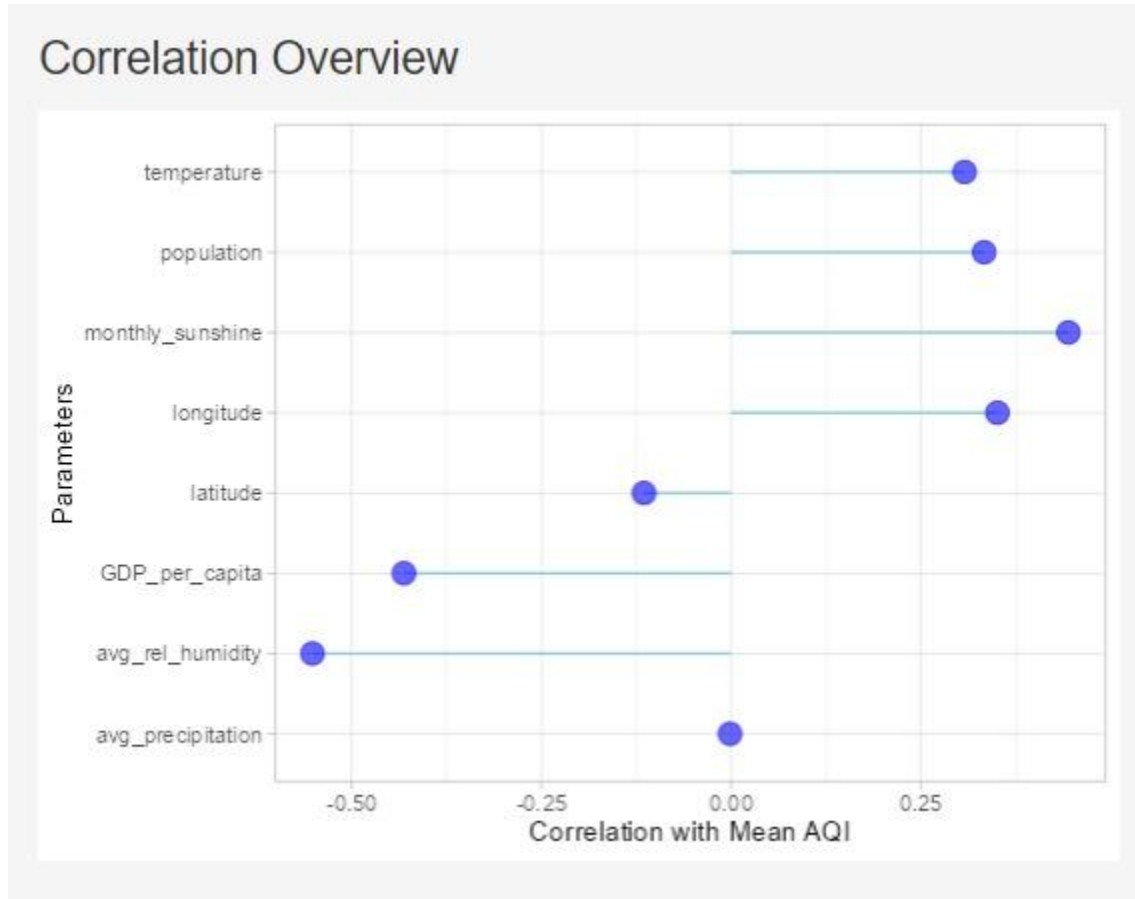
Global AQI Distribution (Heatmap)



Global Air Quality (AQI Heatmap)

- The global heatmap visualization showed significant variation.
- Air quality is poorest in parts of Asia, with Northern India exhibiting the **highest AQI values**, aligning with CPCB (Central Pollution Control Board) reports.
- **Visualization Method:** Points plotted on world maps using ggplot2 and sf, colored by AQI value using the viridis scale.

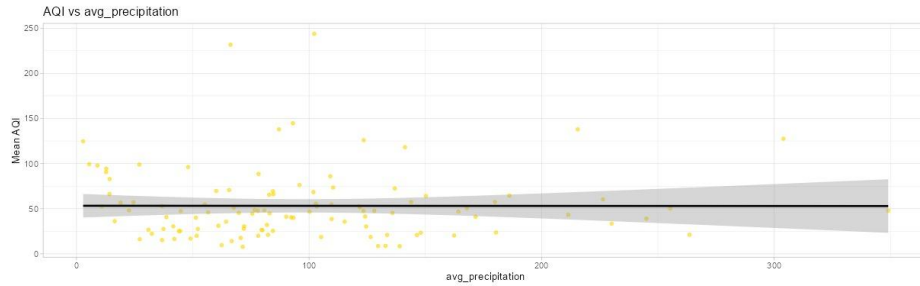
Correlation Highlights



- Significant positive correlation with sunshine.
- Slight positive correlation with temperature and population.
- Significant negative correlation with GDP per capita and average relative humidity.
- Near zero correlation with precipitation.

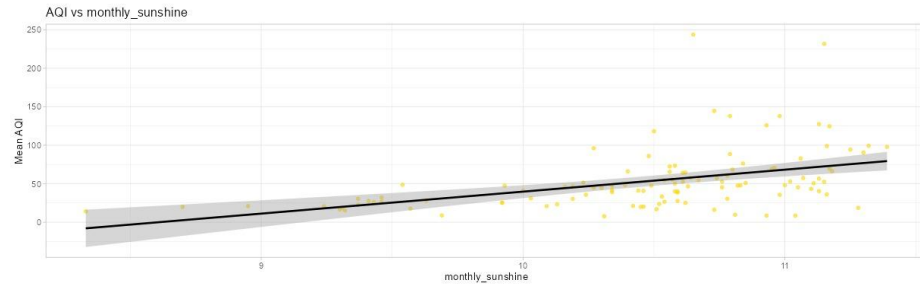
Correlation Highlights

Parameter vs Mean AQI



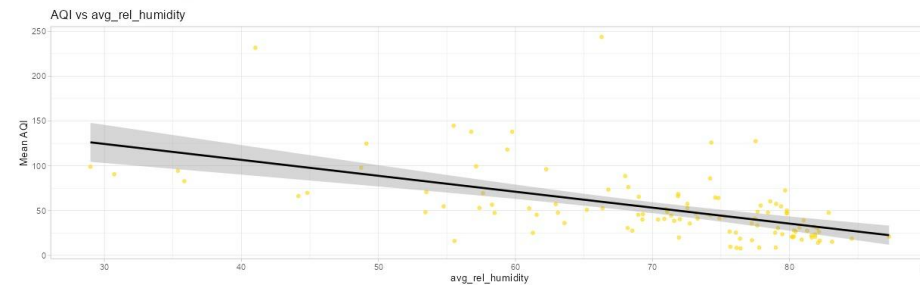
- Near zero correlation with precipitation (rainfall aids in pollutant washout but not consistently)

Parameter vs Mean AQI



- Strong positive correlation with sunshine (UV rays increases creation of ozone and smog)

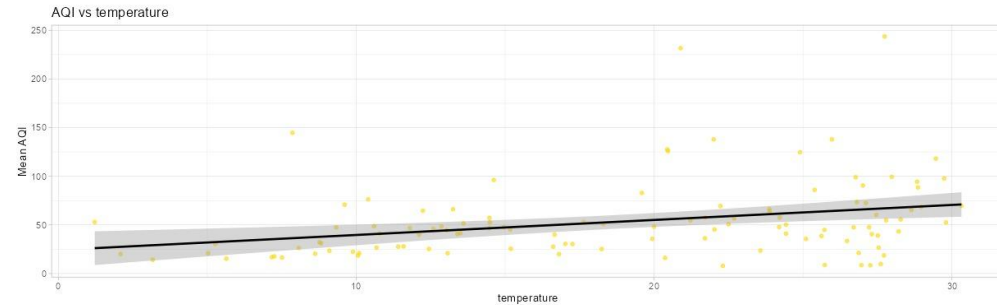
Parameter vs Mean AQI



- High negative correlation with humidity (water vapor helps tiny dust and smoke particles settle down => less in air)

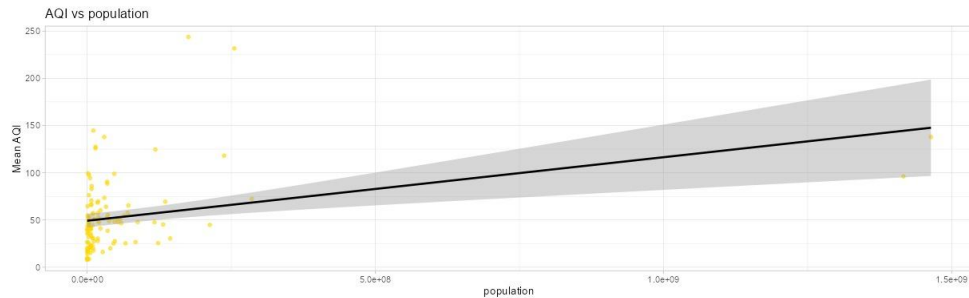
Correlation Highlights

Parameter vs Mean AQI



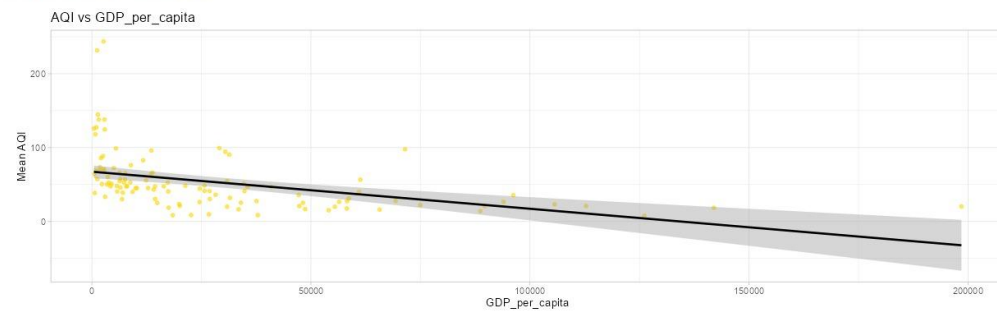
- A mild positive correlation with temperature (higher temperature causes trapping of pollutants near the ground + smog created faster)

Parameter vs Mean AQI



- Strong positive correlation with population (urbanization and vehicular emissions increases pollution)

Parameter vs Mean AQI



- Negative correlation with GDP per capita (i.e more developed countries)

Data Limitations and Bias :

- **Incomplete Coverage:** The dataset excludes many rural and less-developed regions due to a lack of comprehensive monitoring systems.
 - **Urban Bias:** Data availability is higher for developed, densely populated regions, potentially overstating correlations in urban zones while rural pollution remains underrepresented.
- **Standardization Issues:** Varying national AQI calculation methods (CPCB vs. IQAir) introduces inconsistencies.
- **Temporal Variation:** Using averages overlooks daily and seasonal fluctuations.
- **Technical Limitations:** Collected data in sessions and saved it as R was crashing when we tried to scrape all the data in one go.

Thank you