

Visual Analytics of Air Quality and Pollution-Related Health Impacts in India

Abhishek Srivastava, Anay Sinhal, Arth Banka, Happy Pachori, Pankaj Nirmal

Riktेश Singh, Saransh Shivhare, Tanmay Chhapola, Vaibhav Nalotia

I. INTRODUCTION

Air quality has become a critical environmental and public health concern in India. With rapid urbanization and industrialization, the levels of air pollution have escalated, carrying significant implications for the health and well-being of its population. To understand and address these challenges, we propose a visual analytics project aimed at dissecting the intricate layers of air quality data and its correlation with respiratory health outcomes.

The corpus of our study is a meticulously curated dataset from Kaggle, encompassing historical AQI data for the years 2021-2023 across various Indian states and territories. This data, presenting a periodic tabulation of critical pollutants like ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, and particulate matter (PM10 and PM2.5), becomes the linchpin for our analysis. Our visual analytics system will leverage this data to provide an intelligent temporal and spatial analysis of pollution metrics, paralleling the respiratory death toll attributable to this invisible menace.

We will embark on this project with a team of diligent members, each vested with distinct responsibilities ranging from data curation to system development, ensuring a collaborative yet specialized approach to achieving our objectives. Through this endeavour, we envisage not only visualizing India's AQI data but also underscoring the imperative for urgent intervention in the battle against air pollution and its lethal ramifications.

II. DATASET DESCRIPTION

Title: AQI data of Indian States (2021-2023)

Source:

<https://www.kaggle.com/datasets/rudravpatel/aqi-data-of-india-2021-2023>

This dataset details the Air Quality Index (AQI) across Indian states and union territories from 2021 to 2023. This dataset, sourced from Kaggle, comprises periodic AQI measurements and other air quality parameters within individual Excel files for each region. It covers a wide array of pollutants, such as Ozone, CO, SO₂, NO₂, PM10, and PM2.5, alongside metadata including the date range, state, city, and monitoring station.

The dataset captures over 3 years of data, with each file offering a fine-grained temporal resolution that captures daily pollution levels. These details permit a study of air quality trends over time and across different locations.

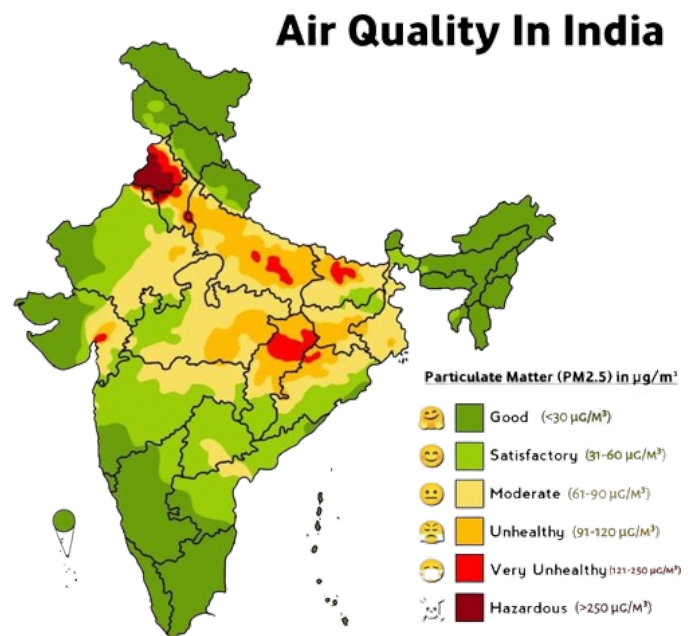


Figure 1. AQI Choropleth Map of India

Other files include:

- <https://www.kaggle.com/datasets/shrutibhargava94/india-air-quality-data>
- <https://data.gov.in/resource/year-wise-number-incidence-lung-cancer-cases-both-sexes-are-incr>

III. GOALS

To offer a detailed view of the temporal distribution of pollutants across different states, we will develop a dynamic visualization quartet:

- A Choropleth Map to geographically represent varying AQI levels across India.
- A Time Series Plot to track the AQI over time.
- Parallel Coordinates to facilitate comparative analysis of different pollutants within a state.
- A Bar Chart to depict the number of respiratory-related deaths.

To assess the combined effect of pairs of pollutants on health across various states using:

- A Scatter Plot to compare pollutants against each other.

- A Choropleth Map focused on visualizing the distribution of deaths by geography.

To dive into the intrinsic characteristics of the AQI dataset and unravel underlying patterns through statistical methods and their corresponding visual interpretations:

- A PCA (Principal Component Analysis) Plot to identify the primary factors in the dataset, thereby simplifying the complexity of multivariate data.
- A Scree Plot to determine the number of principal components to consider, offering a precise visual method for dimension reduction.
- A Squared PCA Loadings Table to quantitatively convey the significance of different pollutants in the PCA.
- An MDS (Multidimensional Scaling) Plot to visualize the distances or dissimilarities between states based on their AQI profiles.

IV. WHAT WILL OUR SOLUTION LOOK LIKE?

Our solution will be an interactive web-based visualization platform, enabling users to explore and analyze the AQI data and its health implications across India with ease and precision, refer to Figure 4.

V. WORK DIVISION

Member 1: Visualization of Distribution of pollutant AQI for each pollutant per year on India map, refer to Figure 1.
 Member 2: Visualization of relative density for each pollutant per year on the India Map.
 Member 3: PCA scatter plot and data preprocessing.
 Member 4: MDS plot and data preprocessing.
 Member 5: Scree plot.
 Member 6: Distribution of Deaths from air pollution leading to lung cancer and other diseases.
 Member 7: Scatter plot of Pollutants, refer to Figure 3.
 Member 8: Time Series Analysis.
 Member 9: Comparison of each pollutant across years, refer to Figure 2.

VI. EXPECTED RESULTS

- Comprehensive Visualization: We expect to create a set of interactive visualizations that will allow users to intuitively understand and explore the AQI data.
- Insightful Correlations: By comparing AQI data with mortality statistics, we expect to find the correlations that highlight the human cost of air pollution.
- Data Quality: Through data preprocessing, we aim to enhance the quality and reliability of the dataset.
- Research Contribution: The PCA, MDS plots, and scree plots are expected to contribute significantly to the academic research, offering a statistical analysis of pollution data.
- Educational Tool: The platform could be used as an educational resource, fostering a deeper understanding of environmental data analysis and its implications among public.

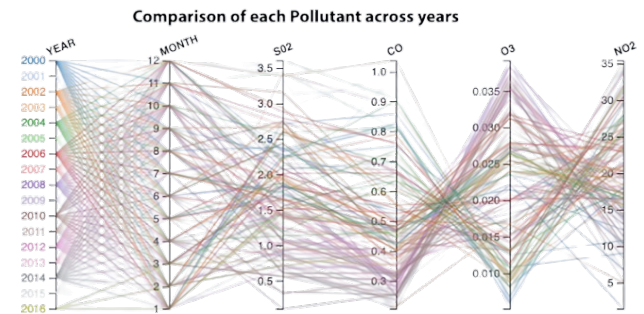


Figure 2. Pollutant Comparison

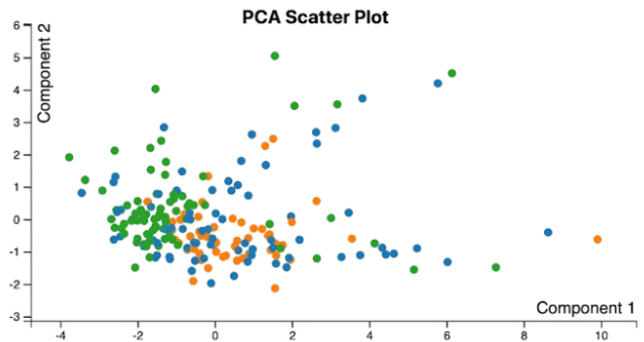


Figure 3. PCA Scatter Plot

VII. REFERENCES

- [1] "Home — Open Government Data (OGD) Platform India," Data.gov.in, Jan. 21, 2022.
<https://data.gov.in/dataset-group-name/Air%20Quality>.
- [2] "Real time Air Quality Index — Open Government Data (OGD) Platform India," Data.gov.in, Jan. 21, 2022.
<https://data.gov.in/catalog/real-time-air-quality-index>.
- [3] 262588213843476, "choropleth map of India," Gist, Apr. 14, 2022.
<https://gist.github.com/anilnairxyz/11190f144a89b54c6698699f3a83b315>.



Figure 4. Visualisation Dashboard