JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SPATIAL AND TEMPORAL VISUALIZATION OF ENVIRONMENTAL AIR QUALITY DATASETS: A STEP-BY-STEP GUIDE

[1] Kumar Pushkar[1], [2]Dr. Mahua Banerjee

[1] Student, Department of management, Birla Institute of Technology Mesra, Lalpur Unit, Ranchi city, 834001, Jharkhand, India.

[2] Assistant Professor, Department of Computer Science, Birla Institute of Technology Mesra, Lalpur Unit, Ranchi city, 834001, Jharkhand, India

***Abstract**:* Harnessing the power of data visualization is crucial for representing the dynamic and multifactor environmental datasets concerning air quality on various scales. This guide illustrates a structured approach to spatial and temporal visualization, enabling a better understanding of air quality and pollution patterns. By integrating diverse data sources from monitoring stations, satellite imagery, and meteorological records—and employing geospatial and temporal techniques, this article can reveal pollution hotspots and their evolution over time. The spatial and temporal techniques and guidelines helps to understand how pollutant and air data changes, evolves, or behaves over different time scales and geographical regions.

Emphasizing data visualization, this methodology empowers environmentalist, researchers, or person to identify trends, analyses correlations, and assess the impact of air pollution on public health and the environment. Through interactive dashboards and dynamic visualizations, it transforms raw data into a compelling narrative, fostering informed decisions for effective environmental management and public safety.

Keywords: Data visualization, environmental study, air quality, geographical representation, temporal representation

I. INTRODUCTION

The escalating complexities of urban environments and industrial activities have led to a surge in air pollution, demanding meticulous monitoring and analysis. Air Quality Index (AQI) datasets, often vast and multifaceted, provide crucial insights into the spatial and temporal distribution of pollutants. However, raw data alone offers limited understanding. Visualizing these datasets transcends mere data presentation; it transforms abstract figures into tangible, actionable knowledge [4]. Effective visualization, through techniques like interactive maps and dynamic time-series charts, enables stakeholders to discern pollution hotspots, identify temporal trends, and comprehend the impact of environmental factors.

This paper empowers communities, policymakers, and researchers to address air quality challenges with precision and representing the data through various forms of graphs, maps and charts. Comprehending air quality data is essential for making educated decisions and formulating effective policies within communities. However, managing raw data without visual representations can be difficult, particularly when handling extensive datasets. This process is crucial for informed decision-making, enabling targeted actions, and ultimately protecting public health. Therefore, visualizations are essential since they assist in converting intricate AQI datasets into clear and engaging visuals.

II. VISUALIZATION TOOLS

2.1 Microsoft™ Power BI

Power BI plays a pivotal role in transforming complex environmental data into actionable insights through interactive visualizations [9]. Its ability to seamlessly integrate diverse data sources of various formats from sensor networks to satellite imagery, allows for comprehensive analysis. By leveraging its robust geospatial capabilities, Power BI enables the creation of dynamic maps and dashboards that reveal air quality patterns and trends.

- **Interactive Mapping***:* Power BI allows for the creation of choropleth maps, heatmaps, and point maps to visualize pollution distribution across geographical areas.
- **Time Series Analysis***:* It facilitates the creation of line charts and area charts to illustrate pollution trends over time, enabling the identification of seasonal variations and long-term patterns.
- **Data Filtering and Drill-Down***:* Users can easily filter and drill down into specific data points to explore underlying details and identify contributing factors.
- **Real-time Dashboards:** allows continuous monitoring of environmental data, providing immediate insights into changing conditions.
- **Integration with GIS:** Integration with Geographic Information Systems (GIS) to provide advanced spatial analysis, combining environmental data with geographical context.

2.2 ArcGIS™

ArcGIS™ by ESRI™ offers significant advantages for data visualization by providing a robust platform to transform complex spatial air quality data into insightful and interactive visual representations [10]. Its comprehensive suite of tools allows users to create not just static maps, but also dynamic web maps, 3D scenes, charts, and dashboards that reveal patterns, trends, and relationships often hidden in raw data tables. This capability stems from its strong foundation in geographic principles and advanced analytical functions, enabling visualizations that are both aesthetically compelling and analytically sound

3 Approaches to Air Quality data visualization

This article highlights the usage of spatial and temporal visualization of air quality datasets as the primary precursor towards representation of data to various target audiences and presenting the findings of analysis.

Compiling air quality data presents difficulties such as variations, quality issues, timing, and the sheer volume of information. To address these challenges, it is essential to establish stringent data cleaning processes to eliminate discrepancies and inaccuracies, thereby guaranteeing the data's dependability and precision for analysis. This entails removing outlier readings and calibrating sensor outputs. Additionally, it is crucial to ensure that all data is transformed into a uniform format to enable smooth integration and comparison. This process requires employing standardized measurement units and consistent data formats [8].

Air quality shows variation both spatially and over time as a result of localized emissions, climatic conditions, and geographical elements, creating a challenge for accurate representation. Fixed maps might not reflect the dynamic alterations happening during the day or throughout different seasons. Recognizing the relationship between air pollution and external influences is essential for pinpointing sources and developing effective mitigation strategies [5].

Visualization of Air quality datasets have two dimensional aspects
(a) Objectives
(b) Nature of data

3.1 OBJECTIVES OF STUDY

The objectives of study determine the goal of the study that needs to be achieved. There can be multiple objectives within the same study that direct the steps for visualization process. These can be related to pollutant analysis, communication to public awareness, analysis of temporal trends, identifying geographical hotspots, scientific research and forecasting.

3.2 NATURE OF DATA

The nature of data should cater the objectives and lays the foundation for the type of visualization that needs to be utilized for air quality datasets

Air quality datasets contain variety of variables related to concentration of pollutants such as PM2.5, PM10, Ozone (O_3), Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), Carbon Monoxide (CO), Ammonia (NH_3) and meteorological factors like Wind Speed, Wind Direction, Temperature, Relative Humidity (RH), Atmospheric Pressure, Solar Radiation, Rainfall, Timestamp, Location, as well as derived variables like Air Quality Index (AQI) [2].

- Geospatial data requires presence of coordinates (latitude and longitude of data points), location data, GIS data, and other geographical information like name of country, city or landmarks.

- Temporal data requires data to encompass a range of time as the time range can be hourly, monthly, yearly or even spans multiple decades across centuries. Example pre- and post-industrialization AQI of a region.

4 Spatial data visualization

Spatial representation refers to the visual depiction of data based on geographical locations, helping to identify patterns in regions divided over continents, country, intercity as well as on other physiological and demographical factors such as topology, forest area and population across different areas. It allows for an intuitive understanding of how variables are distributed spatially.

Spatial visualization of air quality data helps in illustration of huge volume of data that is often difficult for comprehension, trend analysis and pattern identification.

4.1 Elements in spatial data visualization

Selection of geographical area for visualization depends on the scope of the study, such as urban versus rural comparisons. Selecting an appropriate geographical area ensures relevance and accuracy in data interpretation.

- Irregular Spatial Representation: This method provides an alternative approach to traditional boundary-based visualizations based on segment areas data distribution rather than administrative boundaries.
- Polygon Representation: Polygon-based mapping represents different zones with varying data levels, allowing visualization of intensities in specific regions. This technique is commonly used for representing clustered or regionalized information.

4.2 Basemap types and selection

Basemaps provide reference layers for overlaying data, with choices like satellite imagery or street maps. The selection of a basemap influences the interpretability and usability of spatial visualizations.

- **Aerial:** This type of basemap uses imagery captured from satellites or aircraft (aerial photography). It provides a realistic, top-down view of the Earth's surface, showing features like buildings, roads, vegetation, and water bodies as they appear from above.
- **Dark:** This describes a basemap style with a predominantly dark color scheme – typically featuring dark gray or black backgrounds for land and water, with lighter colors used for text, labels, and important linework (like roads) to ensure visibility.
- **Grayscale:** This refers to a basemap style that uses only shades of gray (from black to white), removing all color information. This can be applied to both aerial/imagery basemaps and vector basemaps. Grayscale is often used to provide a neutral background that allows overlaid thematic data.
- **Street map:** A standard road map that displays detailed street networks, neighborhoods, and administrative boundaries. Useful for urban air quality data representation e.g. OSM.
- **Light gray:** Shows minimal geographic features in light gray tones that subtly focuses on data points.

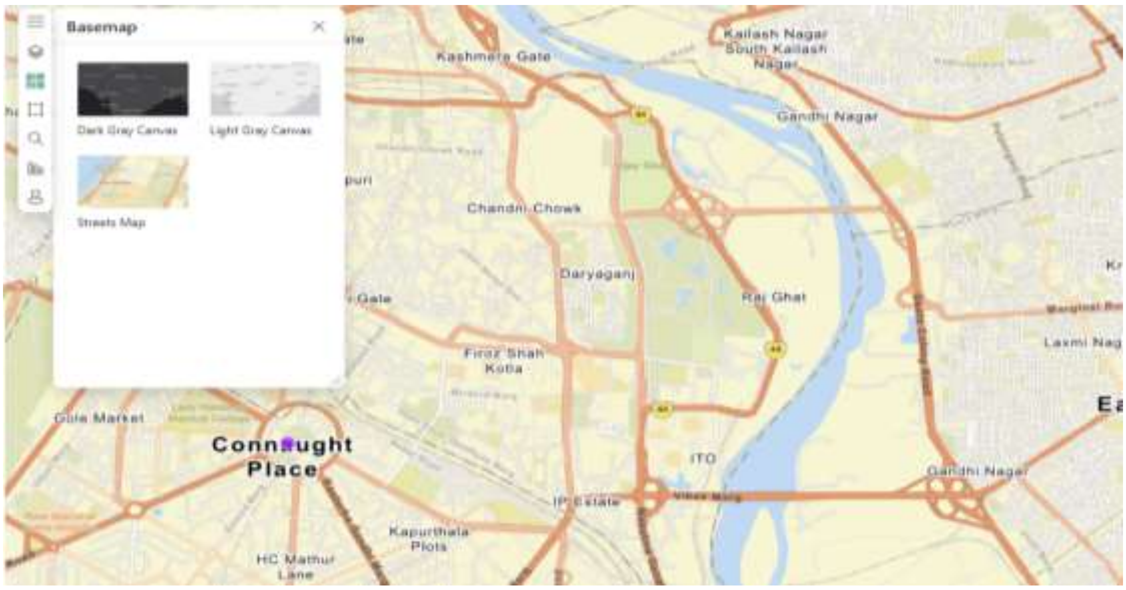

Figure 1 Basemaps types and selection

4.3 Scaling

Appropriate scaling ensures accurate representation of data variations, avoiding misinterpretation. The choice of scale affects how details and trends are perceived within a visualization. Smaller scales can tunnel vision the overall perception during the visualization process diluting the overall focus at the same time larger scales may reduce the effectiveness over an area or geographical boundary [3].

4.4 Symbols

Symbols and legends help distinguish different data levels, often using shapes or colors to represent variations. These elements enhance readability and navigation on the basemap. Selection of symbols should be based on relevance of variables in order to ensure consistency and proper differentiation when multiple variables are involved. Size of symbols should be in proportion to the size of basemap ensuring clutter free visuals [3]. Legend should include proper concise information about different symbols and textual references should be provided as such.

Figure 2: Colored symbols highlighting different AQI category

4.5 Color gradient

When dealing with more intricate data, "diverging" color schemes can be utilized, featuring two opposing colors to represent values that are above and below a central point. The use of gradients should effectively differentiate varying outcomes. Gradients should be selected to emphasize important thresholds. In the case of air quality data, the common preference is for “sequential" color schemes that move from lighter to darker shades or shift from one color to another [2]. It is generally advisable to use colorblind-friendly palettes and keep colors consistent, using warmer hues for higher values and cooler hues for lower values.

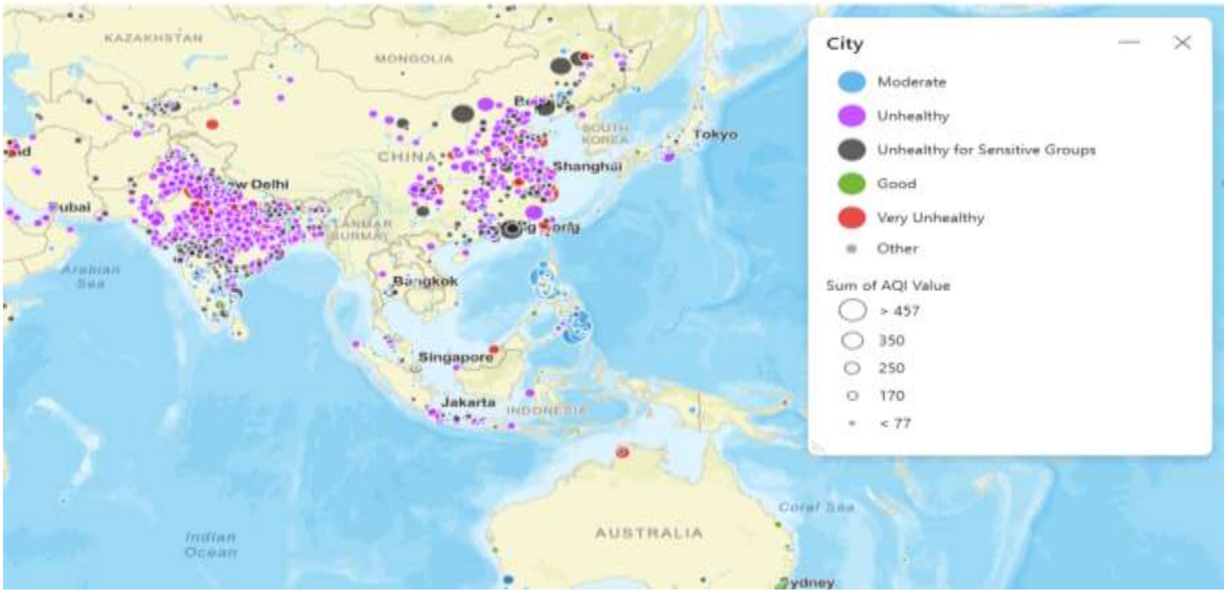

Figure 3: Different symbol sizes and color gradient for city wise representation of air quality

4.6 Representation types

- **Filled Representation**: Filled maps use color gradients to illustrate data intensity, such as concentration levels or density. This method effectively highlights distribution patterns and gradients.
- **Point Representation**: Various symbols or icons on symbol maps are positioned at specific locations, and their size or color varies to represent different levels of pollution or types of pollutants.

- **Heatmaps:** It shows the concentration or intensity of measurements or occurrences across a map. heatmaps make it easy to visually identify patterns, clusters, and areas of high or low intensity at a glance, which might be difficult to see in raw data tables or simple point maps.

Figure 4 pollution intensity in different cities across the world represented through heatmap with color-blind friendly colors

5 Temporal data visualization

Temporal visualization refers to the techniques and methods used to visually represent data points that are ordered by time. Its main goal is to help understand how data changes, evolves, or behaves over different time scales (hours, days, months, years).

5.1 Line graph

A line graph depicts points that are joined by lines (known as trend lines) to illustrate the variations in a dependent variable in relation to an independent variable. An independent variable is, as its name suggests, not influenced by other factors, while the dependent variable is determined by changes in the independent variable. In time-based visualizations, time is consistently the independent variable, placed on the horizontal axis, with the dependent variable displayed on the vertical axis.

Figure 5: Monthly line chart showing pollution in 4 major cities

5.2 Area chart

An area chart resembles a line chart in that it features points linked by straight lines on a two-dimensional graph. It also places time as the independent variable along the x-axis and the dependent variable on the y-axis. In contrast, an area chart has multiple variables "stacked" upon one another, with the space beneath each line filled with colors to signify each variable.

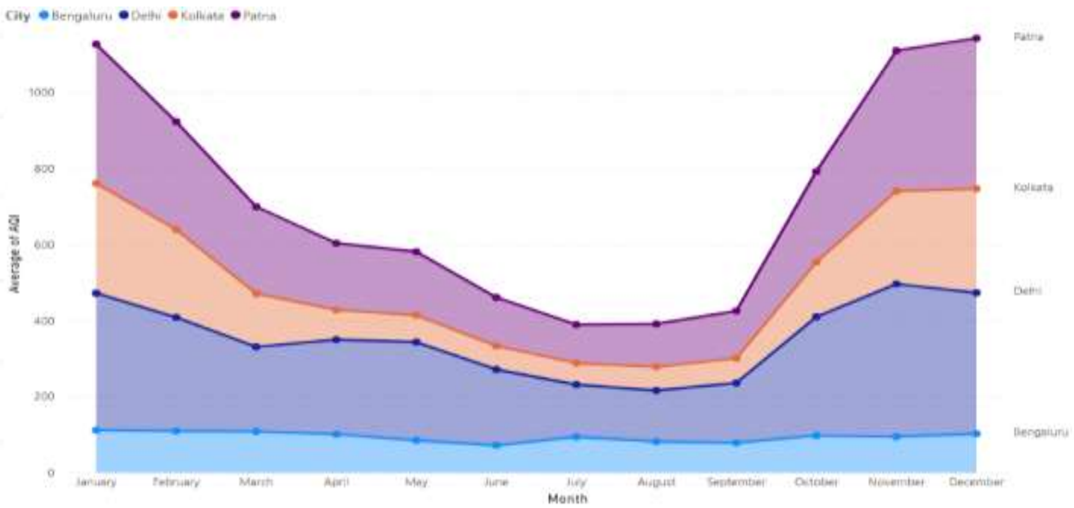

Figure 6: Monthly area chart showing pollution in 4 major cities

5.3 guidelines

- Selecting appropriate time ranges, such as hourly or seasonal data, helps analyse fluctuations and patterns. Time range selection impacts the granularity of data analysis [8].
- The most common way to visualize air quality trends is by placing time (hours, days, months) on the horizontal (x) axis.
- The vertical (y) axis would then represent the concentration of a specific air pollutant (e.g., PM2.5 in µg/m³) [2].
- A line chart would plot the PM2.5 concentration at each point in time, with the line connecting these points showing the trend of PM2.5 levels over the chosen time period (e.g., the last 24 hours, the past week, the last year). The position of each point along the x-axis clearly indicates when that measurement was taken, and its position on the y-axis shows the level of pollution at that time.
- We could have multiple lines on the same chart, each representing a different pollutant (e.g., PM2.5, PM10, O3) [2], all sharing the same time axis. Their vertical positions would allow for comparison of their levels over time in area chart.
- A palette of distinct hues to represent different categories or events occurring over time [7].
- The colors should be easily distinguishable, especially if many categories are present.
- There should be consistent color for each category across different time points or visualizations.

6 Conclusions

Air quality information to different groups presents a notable challenge due to their varying degrees of understanding. Visual representations need to be tailored to the specific audience, utilizing suitable metaphors and user-friendly interfaces. Color gradients, legends, and tooltips provide contextual insights and assistance with interpretation. Interactive visual displays enable users to explore the data further and obtain extra details.

REFERENCES

[1] Dataset utilized: https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates?resource=download
[2] https://atmotube.com/blog/air-quality-data-visualization-for-environmental-research
[3] Ashouri, M., Phoa, F.K.H., Chen, C.H., Shmueli, G. (2023). An Interactive Clustering-Based Visualization Tool for Air Quality Data Analysis. Aerosol Air Qual. Res. 23, 230124. https://doi.org/10.4209/aaqr.230124
https://pmc.ncbi.nlm.nih.gov/articles/PMC5298603/
[4] A Survey of Data Visualization Tools for Analyzing Large Volume of Data in Big Data Platform R.S. Raghav Research Scholar Department of CSE Pondicherry University vpmrags@gmail.com
https://www.researchgate.net/profile/RsRaghav/publication/315870481_A_survey_of_data_visualization_tools_for_analyzing_large_volume_of_data_in_big_data_platform
[5] Likhitha Ravi, Qiping Yan, Sergiu M. Dascalu, Frederick C. Harris, Jr." A Survey of Visualization Techniques and Tools for Environmental Data"
[6] Data Visualization Pitfalls." https://www.slideshare.net/qlik_arg/5-data-visualization-pitfalls.
[7] 8 Ways to Turn Good Data into Great Visualizations." https://www.gooddata.com/blog/8-ways-turn-good-data-great-visualizations.
[8] Skau, Drew. 2012. A Code of Ethics for Data Visualization Professionals." https://visual.ly/blog/a-code-of-ethics-for-data-visualization-professionals/
[9] PowerBi documentations https://learn.microsoft.com/en-us/power-bi/
[10] Create ArcGIS maps in PowerBI https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualizations-arcgis