DHIRUBHAI AMBANI UNIVERSITY

DS612: INTERACTIVE DATA VISUALIZATION

COURSE PROJECT

Prof. Bhaskar Chaudhary

Visualising Air and the Pollutants it contains

Group 11:

Jalak Vyas (202418062)

Vedant Dave (202418014)

Kumar Sanskar (202418027)

Saarthak Jindal (202418046)

Introduction:

Air pollution is one of the most pressing environmental challenges of our time. It poses serious risks to human health, ecosystems, and the climate. Fine particulate matter, toxic gases, and other airborne pollutants can lead to respiratory diseases, cardiovascular conditions, and premature deaths. According to the World Health Organization (WHO), air pollution is responsible for millions of deaths annually, particularly in urban areas where vehicular emissions, industrial activities, and poor urban planning contribute to deteriorating air quality.

Urbanization is a major contributor to increasing pollution levels across the globe and understanding it is crucial.

Monitoring and analyzing air quality data has become essential for policymakers, researchers, and the public to take informed actions. In this context, accessible, reliable, and real-time air quality data plays a pivotal role.

OpenAQ is a non-profit open-source project that collects and unifies air quality data from governments, research institutions, and environmental agencies worldwide. It aims to promote transparency and drive awareness by providing free access to high-quality, standardized air pollution data. It maintains a collection of air quality data throughout the globe collected through sensors owned by government and private institutions.

Through this visualization project, we will dive deeper into the air pollution scenarios in different time frames, and how it depends on the location and urban conditions.

Dataset Description:

The dataset used in this project is sourced from OpenAQ and contains measurements of key air pollutants—including PM2.5, PM10, NO_2, SO_2, CO, and O_3—recorded at various time intervals from different monitoring stations worldwide.

The analysis is based on the dataset, which contains structured air quality measurements from various cities and monitoring stations. Key attributes include:

Location: Monitoring station name

City: The city where the reading was taken

Country: Country of origin (in ISO format)

Parameter: Pollutant measured (e.g., PM2.5, NO_2)

Value: Measured pollutant concentration

Unit: Unit of measurement (typically $\mu g/m^3$)

Datetime: Local timestamp of measurement (UTC format)

Latitude / **Longitude**: Geographic coordinates of the station

Geographically, the data is available for the majority of countries. The number of sensors throughout a country or city varies hugely.

Frequency-wise, the data is available for different time stamps based on locations/ sensors.

Project Goals:

The primary goal of this project is to identify and analyze patterns in air pollution related to the time of day, human activity, and seasonal changes. We aim to test three hypotheses using real-time environmental data.

Utilize a comprehensive dataset from OpenAQ, enriched with temporal features (hour, day, month, season) and derived pollution categories, to provide a detailed understanding of how urban activities and environmental factors interplay to affect air quality.

This project aims to explore and visualize global air quality trends using the OpenAQ dataset. Through a series of Tableau dashboards and interactive visualizations, we seek to test:

- **Hypothesis 1:** Traffic and Urban Activity Drive Pollution Peaks During Rush Hours
- **Hypothesis 2:** Major cities will have higher pollution levels than smaller towns
- **Hypothesis 3:** Seasonal Variations in Air Quality Due to Prolonged Exposure

Context and Problem Statement:

Urban air quality is a critical issue that has significant implications for public health, urban planning, and environmental sustainability. Rapid urbanization, coupled with increased vehicular traffic and industrial activity, has resulted in elevated levels of pollutants in cities worldwide. Data from organizations such as OpenAQ provide valuable insights into the spatial and temporal distribution of these pollutants, yet many underlying patterns remain to be fully understood.

In modern urban environments, the variability in air quality is influenced by both anthropogenic activities and natural atmospheric processes. For example, rush hour traffic is often associated with short-term spikes in pollution, while seasonal changes affect atmospheric dispersion and pollutant accumulation.

Temporal Dynamics:

When do urban pollution levels reach their peak? While it is evident that commuter traffic leads to increased pollutant concentrations during specific hours, it is less clear whether air quality degrades cumulatively throughout the day. Understanding the daily progression of pollution is essential for assessing exposure risks and informing mitigation measures.

Seasonal Variability:

How do seasonal atmospheric conditions affect air quality? Variations in temperature, humidity, and wind patterns can significantly influence the dispersion and accumulation of pollutants. Identifying seasonal trends is crucial for designing targeted interventions during periods of heightened risk.

Intended Outcomes:

Enhanced Understanding of Temporal Patterns:

Through rigorous data analysis, the project will pinpoint the specific hours during which urban air pollution peaks, thereby explaining the role of daily human activities, particularly rush hour traffic in shaping pollutant concentrations.

Insight into Daily Cumulative Exposure:

The analysis will reveal whether and how air quality deteriorates progressively over the course of a day, providing crucial insights into the cumulative impact of pollutant exposure on public health.

Identification of Seasonal Variations:

By examining data across different seasons, the project will determine the extent to which meteorological factors such as temperature inversions, humidity, and wind patterns influence air quality. This will highlight critical periods during which pollutant accumulation is most severe.

Foundation for Future Research:

The methodologies and insights derived from this analysis will serve as a basis for further studies, enabling deeper exploration into the interplay between human activities, atmospheric conditions, and urban air quality.

Research Hypotheses:

Hypothesis 1:

Major cities will have higher pollution levels than smaller towns.

Objective:

Assess whether pollution accumulates over the day.

Approach:

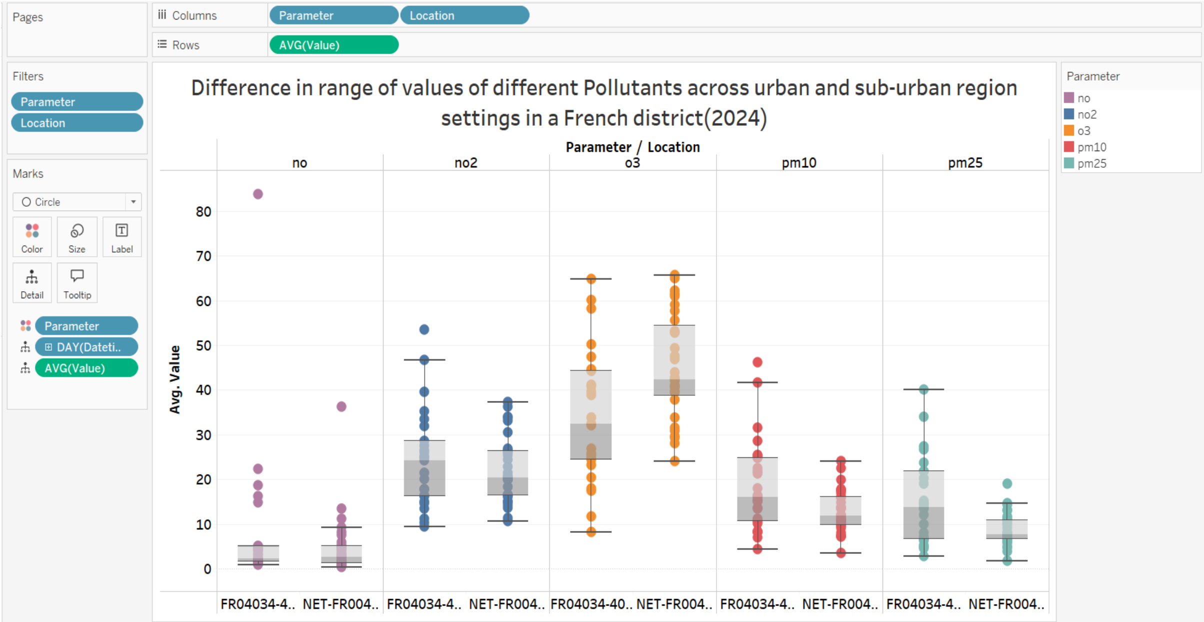
- Aggregate average pollution values for each hour throughout the day.
- Use cumulative graphs or boxplots by hour.

Expected Visualization:

- Cumulative line chart showing hourly pollution buildup.
- Boxplot per hour to examine the spread of values.

Expected Results:

- Gradual increase in pollutants like PM2.5 across the day.
- Possible drop at night due to reduced activity.



Variables Used:

- Directly from Dataset: Location, Pollutant, Pollutant Value

Marks:

- Point (0D) & Line (1D)

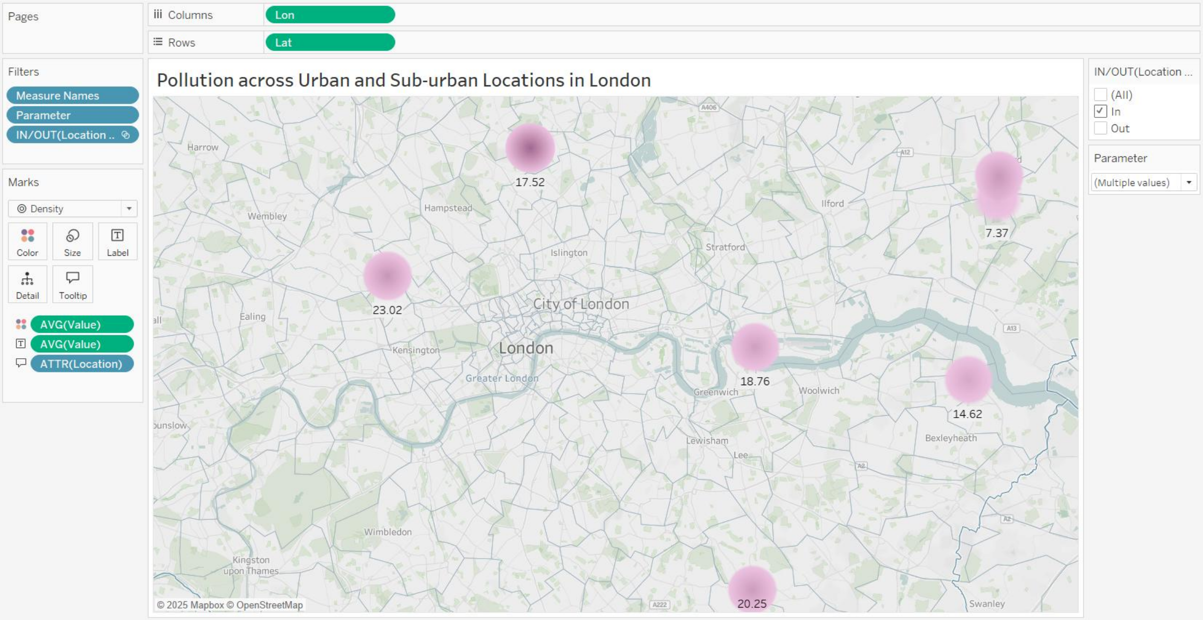
Channels:

- Color hue, vertical & horizontal positioning

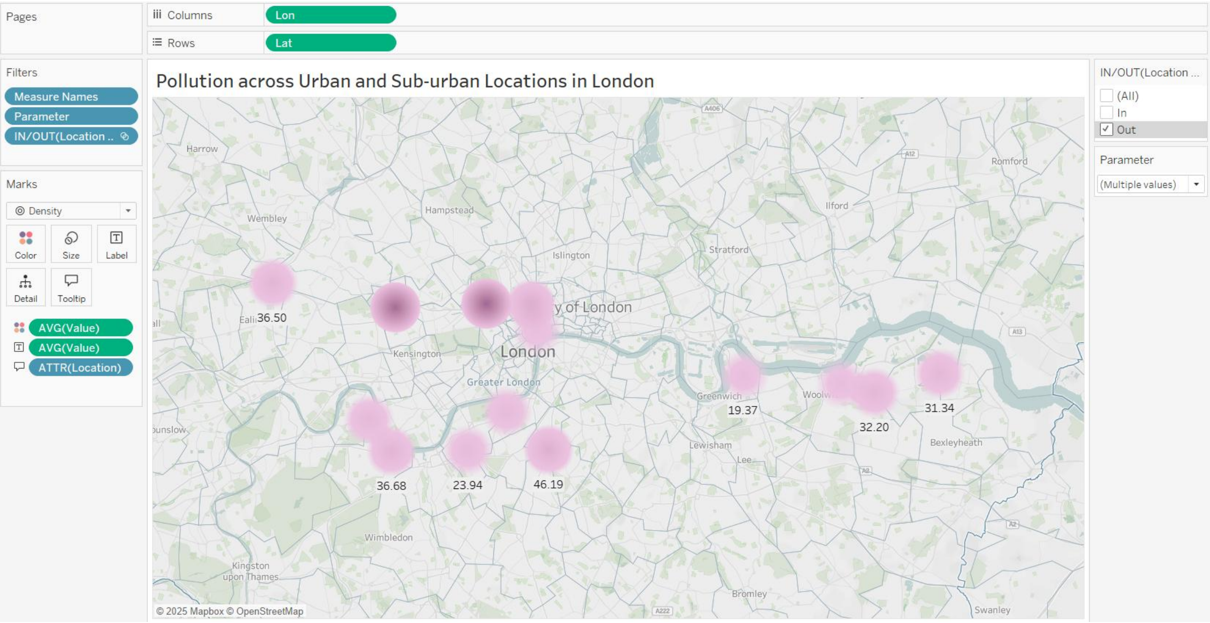
Key Insights:

- Sub-urban locations have low averages and tighter interquartile ranges, suggesting more consistency and cleaner air.
- Urban areas have lower O_3 concentrations, possibly due to NO reacting with ozone and reducing its levels (common in high-traffic zones).

Pollution across Suburban locations in London county:



Pollution across Urban locations in London county:



Variables Used:

- Directly from Dataset: Location, Pollutant, Pollutant Value
- Derived from Dataset: Urban & Suburban areas based on location

Marks:

- Area (2d) & Point (0d)

Channels:

- Horizontal & Vertical Positioning, Density

Key Insights:

- The sub-urban areas which are marked in the map above show a comparative decline in pollution density w.r.t their urban counterparts.

Hypothesis 2:

Seasonal Variations in Air Quality Due to Prolonged Exposure

Objective:

Investigate long-term exposure risks and climate effects.

Approach:

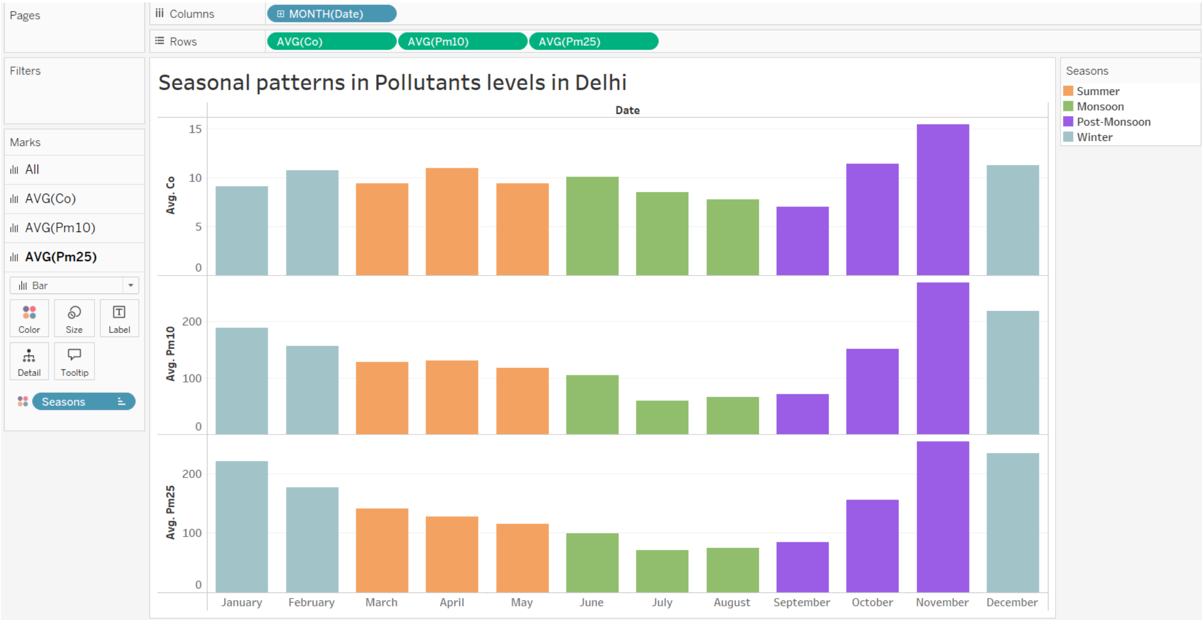
- Categorize each row into a season (based on month):
  - Summer: March - May
  - Monsoon: June - August
  - Post-Monsoon: September - November
  - Winter: December - February
- Calculate average pollution by season and pollutant.
- Compare across cities if multiple are included.

Expected Visualization:

- Bar chart of average pollution per season.
- Heatmap of pollution by month and pollutant.

Expected Results:

- Higher pollution in winter due to atmospheric inversion.
- Cleaner air in summer due to better ventilation and rainfall.



Variables Used:

- From Dataset Directly: Pollutant type, Avg. Pollutant value
- Derived from Dataset: Month and Season from Date Variable

Marks:

- Line Type (1D)

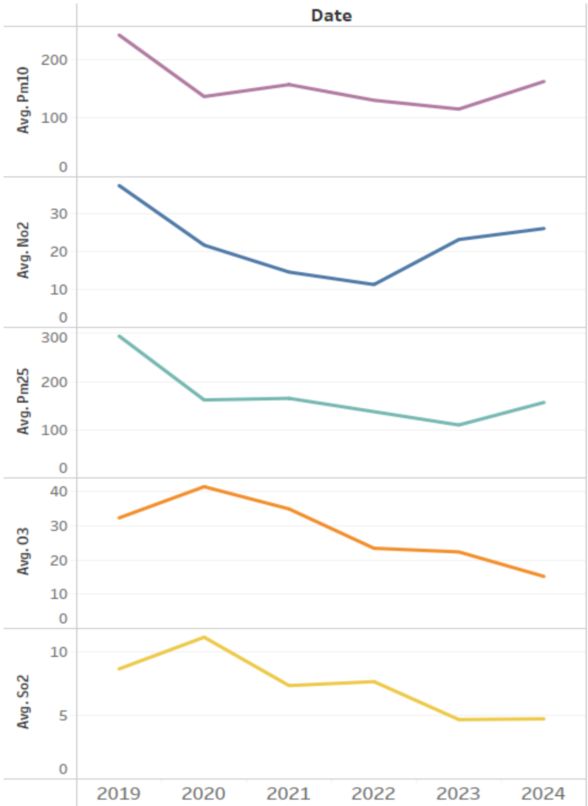
Channels:

- Horizontal Position, Color Hue

Key Insights:

- All 3 pollutants are significantly decreasing during the rainy season, but Carbon Monoxide (CO) as it is in the gaseous state is less affected by rain than PM10 and PM2.5.

How pollutants levels have changed over the years in Delhi(2019-2024)



Variables Used:

- From Dataset Directly: Location(Delhi), Pollutants, Avg. Pollutant Values.
- Derived from Dataset: Year from date values

Marks:

- Point Type (0D)

Channels:

- Horizontal & Vertical Position, Color Hue

Key Insights:

- From the visualization, we can see a significant dip during the years (2020 - 2022) due to "COVID-19", but the major change was in NO_2 (majorly caused by vehicle emission), and as there was less commute during that time and WFH(work from home) measures imposed by companies, which caused a major dip in NO_2 values.

Hypothesis 3:

Traffic and Urban Activity Drive Pollution Peaks During Rush Hours

Objective:

Check if pollution levels peak during morning and evening commute times.

Approach:

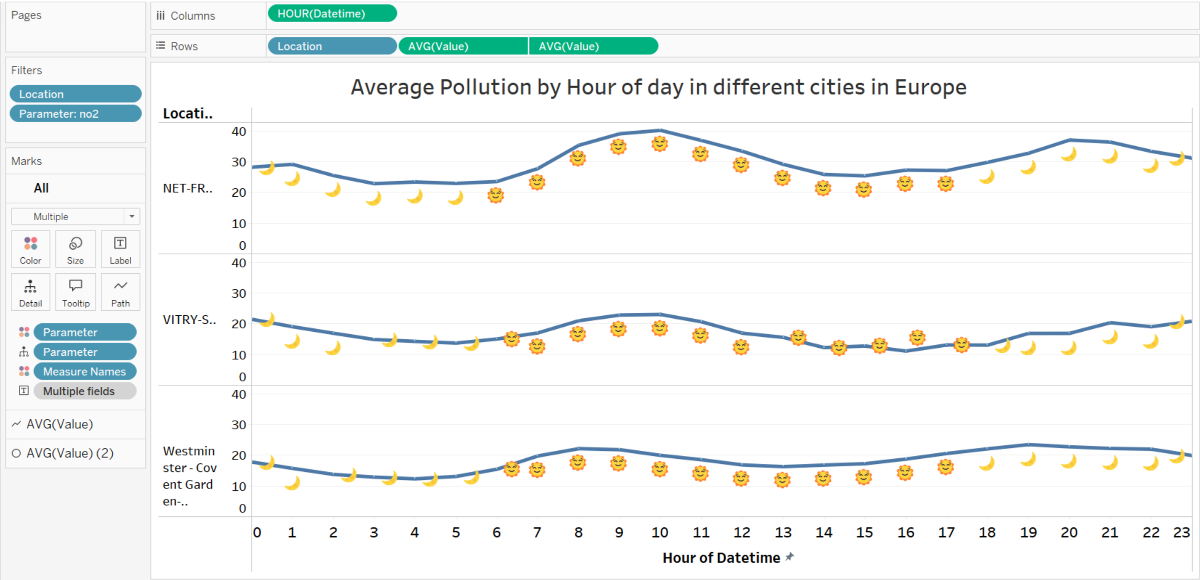
- Group data by hour and calculate the average pollution level by pollutant.
- Split by weekday vs weekend to observe activity-based differences.
- Plot a line graph of pollution level vs. hour (for PM2.5, NO_2, etc.).

Expected Visualization:

- Line chart with hours (x-axis) and average PM2.5/NO_2 (y-axis).
- Separate lines for weekdays and weekends.

Expected Results:

- Two distinct peaks: ~8 AM and ~6 PM on weekdays.
- Flatter trend on weekends due to less traffic.



Variables Used:

- Directly from dataset: Location (London & Paris), Pollutant, Pollution Value.
- Derived from Dataset: Hour of day from Date

Marks:

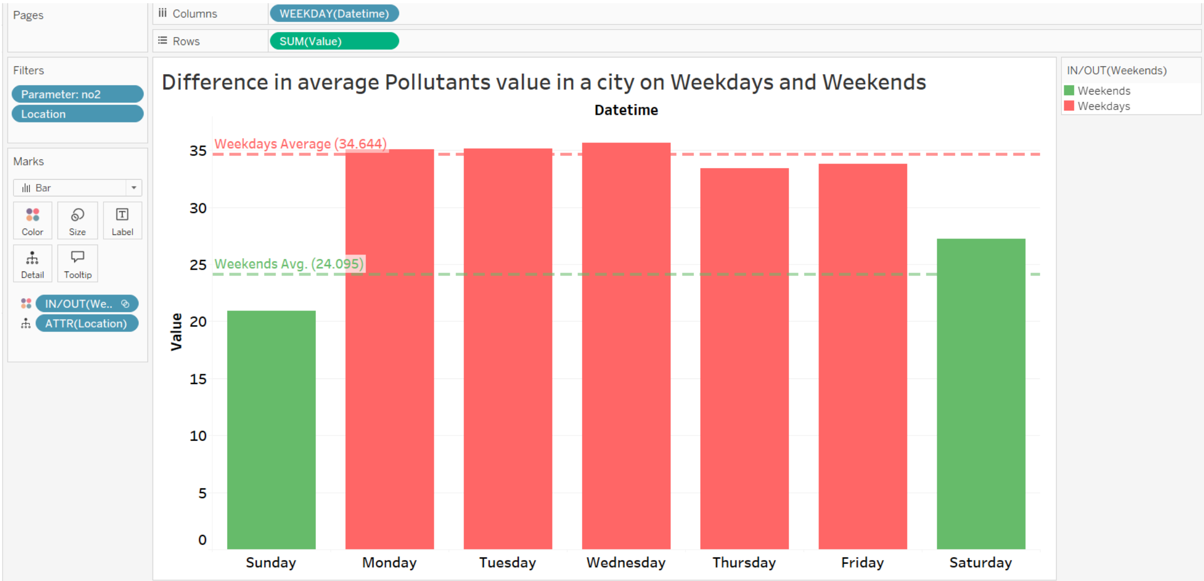
- Point (0D)

Channels:

- Horizontal & Vertical Positioning

Key Insights:

- Europe's major cities (London & Paris) generally follow the same pattern due to similar working culture & geography, with a spike in pollution during peak commute times (7-12 am) and (5-9 pm).



Variables Used:

- Direct from Dataset: Pollutants, Pollutant Value, Location
- Derived from Dataset: Weekday name from Date

Marks:

- Line (1D)

Channels:

- Horizontal Positioning & Color Hue

Key Insights:

- Pollution levels throughout the weekdays are seen to be much higher than how much it is on the weekends.

Heatmap: Hour vs Day of Week
Location: Westminster - Covent Garden-238103

AVG(Value)
9.44 27.82

	Weekday						
Hour of ..	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	
12 AM	20.61 µg/m³	14.31 µg/m³	13.78 µg/m³	20.09 µg/m³	15.08 µg/m³	17.82 µg/m³	22.96 µg/m³
1 AM	18.78 µg/m³	13.35 µg/m³	12.03 µg/m³	16.14 µg/m³	12.95 µg/m³	16.61 µg/m³	20.05 µg/m³
2 AM	16.43 µg/m³	10.13 µg/m³	10.96 µg/m³	14.25 µg/m³	11.21 µg/m³	15.05 µg/m³	18.22 µg/m³
3 AM	14.83 µg/m³	9.44 µg/m³	9.84 µg/m³	14.28 µg/m³	10.79 µg/m³	14.64 µg/m³	16.29 µg/m³
4 AM	13.33 µg/m³	10.29 µg/m³	10.48 µg/m³	12.79 µg/m³	11.01 µg/m³	14.50 µg/m³	13.87 µg/m³
5 AM	13.54 µg/m³	12.25 µg/m³	11.98 µg/m³	14.08 µg/m³	11.34 µg/m³	15.70 µg/m³	12.86 µg/m³
6 AM	13.79 µg/m³	15.94 µg/m³	14.84 µg/m³	17.18 µg/m³	14.55 µg/m³	17.39 µg/m³	14.66 µg/m³
7 AM	14.61 µg/m³	21.57 µg/m³	21.20 µg/m³	22.22 µg/m³	20.43 µg/m³	22.07 µg/m³	16.75 µg/m³
8 AM	15.21 µg/m³	23.28 µg/m³	23.37 µg/m³	25.56 µg/m³	23.45 µg/m³	25.75 µg/m³	19.02 µg/m³
9 AM	15.83 µg/m³	22.53 µg/m³	21.29 µg/m³	25.09 µg/m³	24.11 µg/m³	24.09 µg/m³	20.25 µg/m³
10 AM	15.31 µg/m³	19.25 µg/m³	19.64 µg/m³	21.64 µg/m³	21.95 µg/m³	22.01 µg/m³	20.41 µg/m³
11 AM	13.93 µg/m³	17.01 µg/m³	19.29 µg/m³	19.57 µg/m³	22.15 µg/m³	18.89 µg/m³	19.48 µg/m³
12 PM	12.19 µg/m³	16.05 µg/m³	17.78 µg/m³	18.06 µg/m³	19.28 µg/m³	16.74 µg/m³	18.05 µg/m³
1 PM	12.19 µg/m³	16.10 µg/m³	17.53 µg/m³	17.35 µg/m³	17.75 µg/m³	16.68 µg/m³	16.70 µg/m³
2 PM	13.13 µg/m³	16.55 µg/m³	18.37 µg/m³	17.40 µg/m³	18.84 µg/m³	16.90 µg/m³	16.68 µg/m³
3 PM	13.59 µg/m³	18.89 µg/m³	18.89 µg/m³	17.13 µg/m³	18.84 µg/m³	17.22 µg/m³	16.38 µg/m³
4 PM	14.81 µg/m³	19.70 µg/m³	20.29 µg/m³	19.33 µg/m³	20.81 µg/m³	19.28 µg/m³	17.19 µg/m³
5 PM	16.53 µg/m³	20.41 µg/m³	22.08 µg/m³	21.45 µg/m³	22.46 µg/m³	22.30 µg/m³	18.90 µg/m³
6 PM	18.16 µg/m³	20.93 µg/m³	23.83 µg/m³	22.21 µg/m³	23.30 µg/m³	24.60 µg/m³	21.62 µg/m³
7 PM	18.40 µg/m³	21.84 µg/m³	24.43 µg/m³	22.16 µg/m³	25.64 µg/m³	27.82 µg/m³	23.94 µg/m³
8 PM	17.89 µg/m³	20.08 µg/m³	23.51 µg/m³	20.35 µg/m³	24.13 µg/m³	26.89 µg/m³	26.29 µg/m³
9 PM	16.32 µg/m³	19.88 µg/m³	23.42 µg/m³	20.57 µg/m³	23.99 µg/m³	26.44 µg/m³	24.78 µg/m³
10 PM	16.75 µg/m³	18.82 µg/m³	22.47 µg/m³	19.63 µg/m³	25.69 µg/m³	26.07 µg/m³	24.64 µg/m³
11 PM	15.52 µg/m³	16.20 µg/m³	21.64 µg/m³	17.71 µg/m³	22.65 µg/m³	24.49 µg/m³	21.45 µg/m³

Variables Used:

- Directly from Dataset: Pollutants (NO_2), Locations, Pollutant Value
- Derived from Dataset: Hour & Weekday from Date Field

Marks:

- Area (2D)

Channels:

- Color Hue, Vertical & Horizontal Positioning

Key Insights:

- Weekdays tend to have higher pollution concentration than weekends and in peak hours of commute on weekdays, pollution concentration is the highest. During weekends pollution spikes tend to be during late evening hours as most people go for an outing in the evening.

Skills learned in this Course:

1. Foundations of Visualization

Skills Acquired:

- Understanding of why and how we visualize data.
- Recognizing the importance of human perception and cognition in designing effective visualizations.
- Differentiating between exploratory and explanatory visualization goals.

2. Data Abstraction

Skills Acquired:

- Identifying data types: categorical, ordinal, quantitative.
- Understanding dataset types: tables, networks, trees, spatial, temporal, and multivariate data.
- Ability to map real-world data to abstract types for visualization design.

3. Visual Encoding Techniques

Techniques Learned:

- Mastery of visual channels: position, color, shape, size, orientation, texture, motion.
- Mapping data attributes to visual channels based on their effectiveness and perceptual accuracy.
- Designing encodings that enhance interpretability and accuracy.

4. Marks and Channels

Concepts Covered:

- Understanding how marks (points, lines, areas) and channels (position, color, etc.) build all visualizations.
- Choosing the appropriate combination of marks and channels for given data types and tasks.

5. Task Abstraction

Skills Acquired:

- Identifying and categorizing user tasks: lookup, compare, summarize, filter, explore, etc.
- Linking visualization design choices to task requirements for better user interaction and comprehension.

6. Design Principles

Techniques Learned:

- Applying the “What-Why-How” framework to structure visualization thinking:
    - What: the data.
    - Why: the tasks.
    - How: the visual encoding, interaction, and layout.
- Utilizing design cycles: ideation, prototyping, critique, refinement.
- Awareness of design pitfalls (e.g., misleading scales, clutter).

7. Interaction Techniques

Skills Acquired:

- Implementing interactive techniques like:
  - Zooming and panning
  - Brushing and linking
  - Filtering and highlighting
  - Details on demand
- Understanding the role of interaction in facilitating data exploration.

8. Evaluation and Critique

Skills Acquired:

- Conducting formative and summative evaluations of visualizations.
- Applying heuristics to critique visualization quality (e.g., expressiveness, effectiveness).
- Understanding when to use qualitative vs. quantitative evaluation methods.

9. Data and Task-Driven Design

Skills Acquired:

- Designing visualizations that serve real-world user needs.
- Aligning task goals with visualization types (e.g., bar charts for comparison, line graphs for trends).
- Integrating user feedback into iterative visualization development.

10. Visualization Techniques for Different Data Types

Techniques Learned:

- Tabular data: bar charts, line plots, scatterplots.
- Temporal data: timelines, horizon charts.
- Hierarchical data: node-link diagrams, treemaps.
- Network data: force-directed layouts, adjacency matrices.
- Geospatial data: choropleth maps, symbol maps.

11. Storytelling with Data

Skills Acquired:

- Structuring visualizations into compelling narratives.
- Balancing aesthetics, accuracy, and clarity.
- Using annotations, highlights, and sequences to guide the viewer.

12. Ethics and Integrity in Visualization

Topics Covered:

- Avoiding misleading visual elements.
- Ensuring inclusivity and accessibility.
- Acknowledging biases in visual representation.