Name:	Chaitya Arun Dobariya
UID:	2021600017
Course:	Advanced Data Visualization

Experiment 8

Aim:	To design interactive dashboards and create visual storytelling using D3.js on a dataset related to Environment/Forest cover, covering basic and advanced charts. Objectives: To understand how to use D3.js for data visualization. To implement basic charts like Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, and Bubble plot. To implement advanced charts like Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, and Jitter. To draw observations and insights from each chart. To create an interactive storytelling dashboard using the above visualizations. Expected Outcomes: Ability to create various types of visualizations using D3.js. Interactive dashboards demonstrating different types of charts. Insights from the Environment/Forest cover the dataset through visual storytelling.

1. **Dataset**

You can find the dataset here.

Description

The Global Air Quality Data dataset provides an extensive compilation of air quality measurements from various prominent cities worldwide. This dataset includes crucial environmental indicators such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and ozone (O3), along with meteorological data like temperature, humidity, and wind speed. With 10,000 records, this dataset is ideal for researchers, data scientists, and policy makers looking to analyze air quality trends, understand the impact of pollution on health, and develop strategies for environmental improvement.

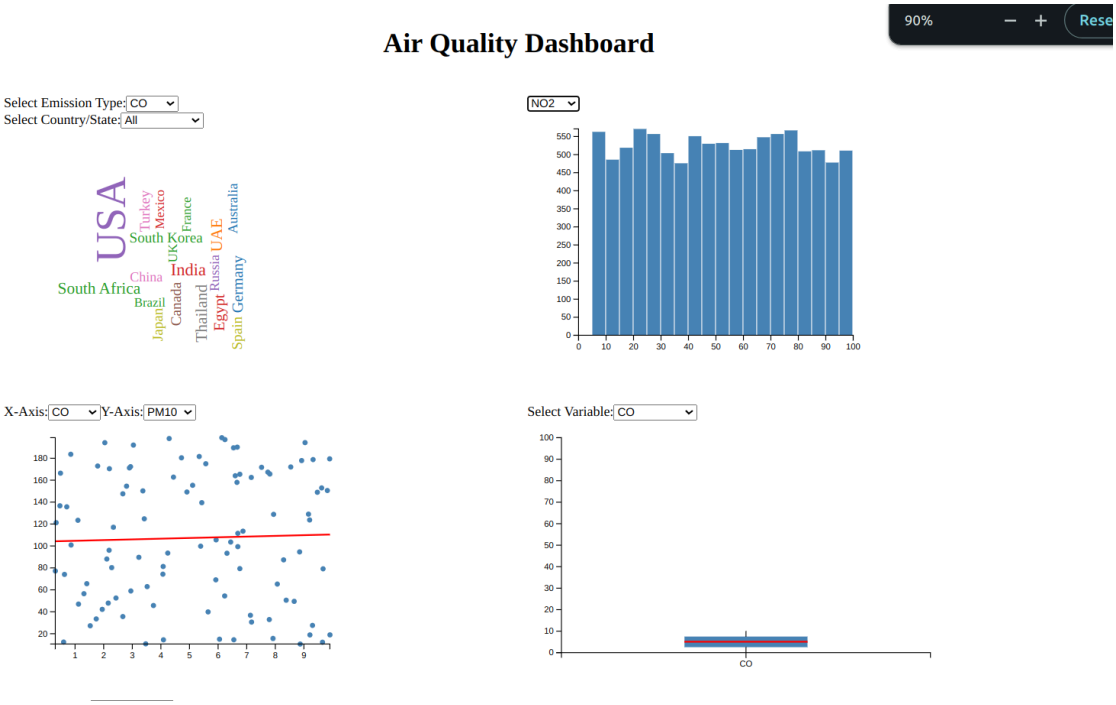
The dataset is composed of the following columns:

City: The name of the city where the air quality measurement was taken.
Country: The country in which the city is located.
Date: The date when the measurement was recorded.
PM2.5: The concentration of fine particulate matter with a diameter of less than 2.5 micrometers (μg/m³).
PM10: The concentration of particulate matter with a diameter of less than 10 micrometers (μg/m³).
NO2: The concentration of nitrogen dioxide (μg/m³).
SO2: The concentration of sulfur dioxide (μg/m³).
CO: The concentration of carbon monoxide (mg/m³).
O3: The concentration of ozone (μg/m³).
Temperature: The temperature at the time of measurement (°C).
Humidity: The humidity level at the time of measurement (%).
Wind Speed: The wind speed at the time of measurement (m/s).

air_quality.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	City	Country	Date	PM2.5	PM10	NO2	SO2	CO	O3	Temperatı	Humidity	Wind Speed
2	Bangkok	Thailand	19-03-2023	86.57	25.19	99.88	30.63	4.46	36.29	17.67	59.35	13.76
3	Istanbul	Turkey	16-02-2023	50.63	97.39	48.14	8.71	3.4	144.16	3.46	67.51	6.36
4	Rio de Jan	Brazil	13-11-2023	130.21	57.22	98.51	9.92	0.12	179.31	25.29	29.3	12.87
5	Mumbai	India	16-03-2023	119.7	130.52	10.96	33.03	7.74	38.65	23.15	99.97	7.71
6	Paris	France	04-04-2023	55.2	36.62	76.85	21.85	2	67.09	16.02	90.28	14.16
7	Los Angel	USA	05-01-2023	121.28	196.39	86.17	28.74	2.64	46.86	18.53	43.01	3.96
8	Tokyo	Japan	17-09-2023	118.76	15.69	12.99	39.05	3.04	15.75	-1.37	17.76	17.66
9	New York	USA	23-11-2023	72.52	23.43	92.01	15.26	4.57	106.5	32.14	76.57	9.42
10	Tokyo	Japan	21-09-2023	131.46	145.57	10.13	25.1	9.08	181.72	-6.2	71.42	16.81

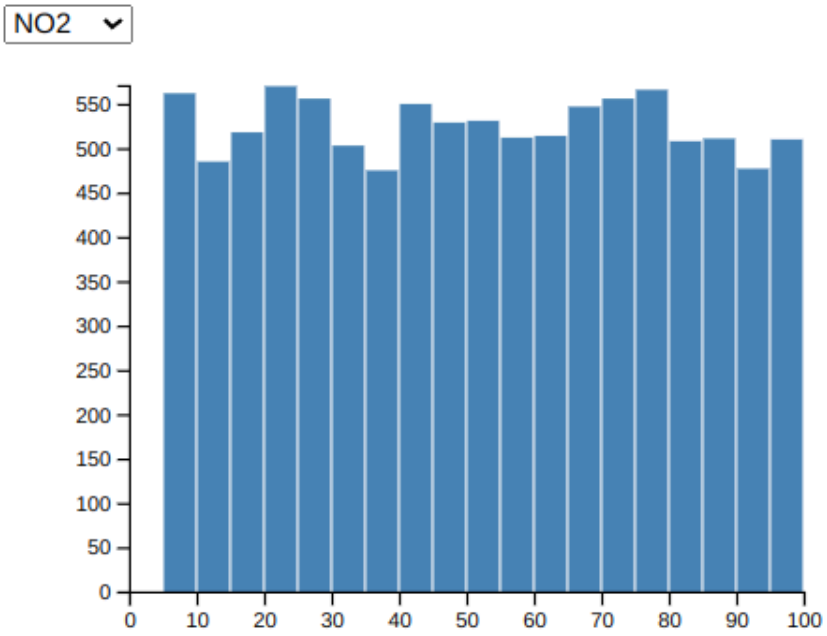
2. Charts & Plots



Select Emission Type: PM2.5

Select Country/State: All

The larger size of certain countries like China, India, and the USA indicates that they might have higher levels of air pollution.



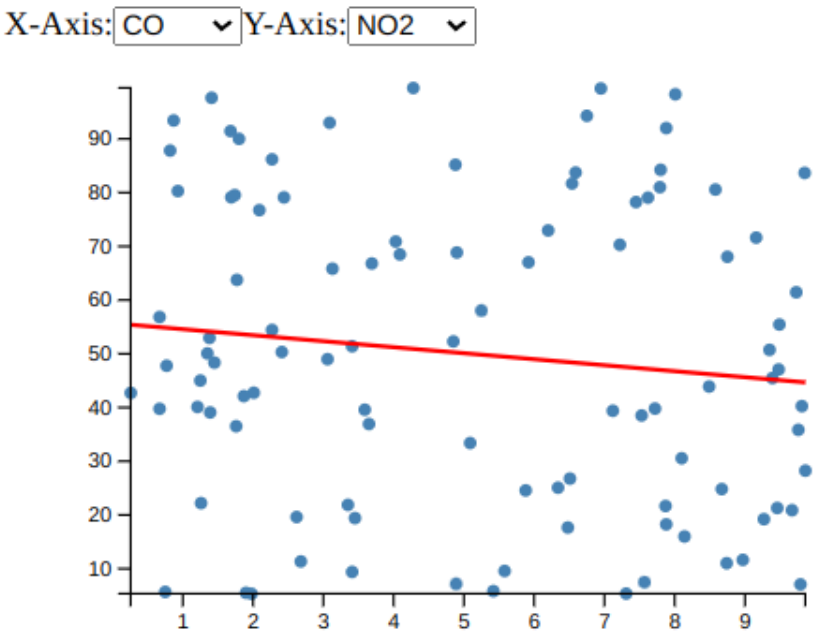
Observation:

1. Range of NO2 Concentrations: The x-axis shows the range of NO2 concentrations, likely in parts per billion (ppb).
2. Frequency: The y-axis represents the frequency or count of occurrences of each concentration range.
3. Distribution Shape: The shape of the distribution can provide insights into the nature of the data.

Possible Interpretations:

Without specific details about the data collection period and location, we can make some general interpretations based on the graph's shape:

1. Normal Distribution: If the bars form a bell-shaped curve, it suggests a normal distribution. This implies that most of the NO2 concentrations are clustered around the mean, with fewer occurrences at higher and lower values.
2. Skewed Distribution: If the bars are skewed to one side, it indicates that the data is not normally distributed.
○ Positive Skew: If the tail of the distribution extends to the right, it means there are more high values.
○ Negative Skew: If the tail extends to the left, it means there are more low values.



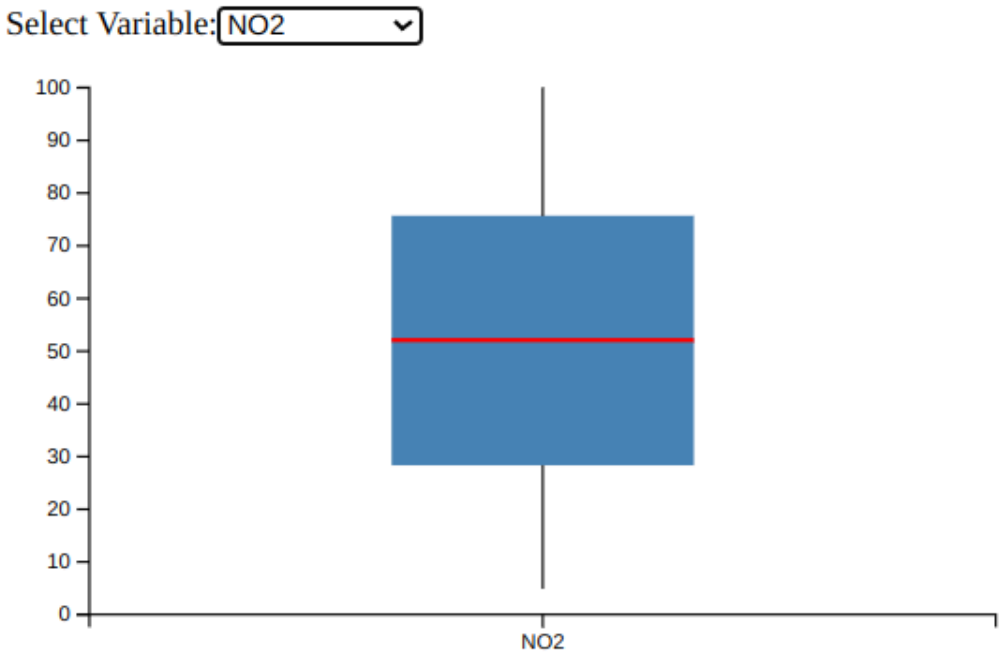
Possible Interpretations:

● Common Source: The positive correlation suggests that both CO and NO2 may originate from similar sources, such as vehicle emissions, industrial activities, and residential heating.

● Chemical Reactions: Certain atmospheric reactions can lead to the formation of NO2 from CO, further contributing to the observed correlation.

● Meteorological Factors: Weather conditions, such as wind speed, temperature, and humidity, can influence the dispersion and concentration of both pollutants.

● Variability: The scatter in the data indicates that other factors, not captured in this plot, may also influence the levels of CO and NO2.



Air Quality: The NO2 concentrations appear to be relatively high, suggesting potential air quality issues in the area.

Variability: The IQR indicates that there is some variability in NO2 levels, which might be influenced by factors like traffic, industrial activity, and meteorological conditions.

Outliers: The absence of outliers suggests that the data is relatively consistent and free from extreme values.

Conclusion

Data visualization is a crucial tool for understanding complex datasets and conveying insights effectively. The various types of plots—such as pie charts, box plots, violin plots, jitter plots, and word clouds—each offer unique perspectives on the data, making them suitable for different analytical tasks.

1. Pie Charts allow for quick assessments of proportions within categorical data, helping stakeholders grasp the overall composition at a glance.

2. Box Plots (or Whisker Plots) provide detailed insights into the distribution and spread of data, making them invaluable for identifying outliers and understanding variations across categories.
3. Violin Plots combine the strengths of box plots and density plots, offering a more nuanced view of data distributions and their shapes, particularly useful when comparing multiple categories.
4. Jitter Plots mitigate overplotting in scatter plots, allowing for clearer visibility of dense data points and enhancing the exploratory analysis of distributions.
5. Word Clouds visually represent textual data, highlighting the most significant terms and themes, making them effective for text analysis and presentations.