📘 Glossary – Describing and Visualizing Data

1️⃣ Types of Data

- **Quantitative Data:** Numeric values that can be measured (e.g., height, income).
- **Qualitative (Categorical) Data:** Non-numeric categories or labels (e.g., color, gender).
- **Discrete Data:** Countable values (e.g., number of students).
- **Continuous Data:** Measurable values within a range (e.g., weight, time).
- **Nominal Data:** Categories without order (e.g., eye color).
- **Ordinal Data:** Ordered categories (e.g., small, medium, large).

2️⃣ Summary Statistics

- **Mean:** The arithmetic average of a dataset.
- **Median:** The middle value when data are ordered.
- **Mode:** The most frequent value in a dataset.
- **Range:** Difference between the maximum and minimum values.
- **Variance:** Measure of data spread around the mean.
- **Standard Deviation (σ):** Average distance from the mean.
- **Quartiles & IQR:** Divide data into quarters; IQR = Q3 – Q1.
- **Outlier:** A value significantly higher or lower than most of the data.

3️⃣ Matplotlib Basics

- **Matplotlib:** A core Python library for creating static, animated, and interactive plots.
- **Figure:** The overall window or page that everything is drawn on.
- **Axes:** The area where data is plotted (contains x/y labels, ticks, etc.).
- **plot():** Draws line plots of data.
- **bar():** Creates a bar chart.
- **hist():** Displays a histogram of numeric data.
- **scatter():** Creates a scatter plot for relationships between two variables.
- **xlabel(), ylabel():** Label axes for clarity.
- **title():** Adds a chart title.
- **legend():** Displays label identifiers for plotted data.

4️⃣ Seaborn Basics

- **Seaborn:** A Python visualization library built on Matplotlib for attractive statistical graphics.
- **sns.barplot():** Displays mean values with confidence intervals.
- **sns.histplot():** Shows data distribution with flexible bin control.
- **sns.scatterplot():** Creates scatter plots with style and hue options.
- **sns.boxplot():** Shows data distribution through quartiles and outliers.
- **sns.violinplot():** Combines boxplot and density visualization.
- **sns.heatmap():** Visualizes correlations or matrices with color gradients.
- **Style & Theme:** Seaborn allows quick aesthetic control (e.g., `sns.set_style('whitegrid')`).

5️⃣ Visualization Mini-Lab

- **Trend:** General direction of data movement in a chart.
- **Outlier:** A point that differs greatly from other observations.
- **Correlation:** Statistical relationship between two variables (positive, negative, or none).
- **Distribution:** The pattern of frequency of data points.
- **Visualization Best Practice:** Always label axes, include units, and provide readable titles.
- **Exploratory Data Analysis (EDA):** Process of visually exploring datasets to summarize main characteristics.