Introduction to Data Visualization with Seaborn

AxesSubplot: The Matplotlib object representing a single subplot or axes, returned by single-plot Seaborn functions and used to set titles, labels, and other axis-level properties

Bar plot: A categorical plot that displays a summary statistic (usually the mean) of a quantitative variable for each category, often with confidence intervals to show uncertainty

Box plot: A plot that summarizes the distribution of a quantitative variable using the median, interquartile range (box), whiskers, and outliers to compare distributions across groups

Categorical variable: A variable that takes on a limited, discrete set of category values (e.g., "Male", "Female", "Yes", "No") and is used to group or split data in plots

catplot(): A high-level Seaborn function for creating categorical plots (count, bar, box, point, etc.) that accepts parameters like kind, hue, col, and row and returns a FacetGrid for subplots

Confidence interval (CI): A shaded region or error bar around an estimated statistic (commonly the mean) that quantifies uncertainty, with Seaborn typically showing a 95% CI by default.

Count plot: A categorical plot that shows the number of observations in each category as bars, useful for visualizing frequency distributions of a categorical variable

hue: A Seaborn parameter that maps a third variable to color, enabling subgroup differentiation within a single plot and automatically adding a legend

Line plot: A relational plot that connects summary values (often means) across ordered x-values, commonly used to track the same measurement over time

Matplotlib: A foundational Python plotting library that offers low-level control over figures, axes, and rendering, which Seaborn builds upon to produce visualizations

palette: A named Seaborn or Matplotlib color scheme or a custom list/dictionary of colors used to control the mapping from data values to visual colors

pandas: A Python library for data manipulation and analysis that provides data structures like Series and DataFrame commonly used as inputs to Seaborn plots

Point plot: A categorical plot that shows summary points (e.g., means or medians) and their confidence intervals for each category, often used to compare trends and subgroup differences

Quantitative variable: A numeric variable that measures a quantity and can be meaningfully ordered and used for mathematical operations, typically plotted on continuous axes

relplot(): A high-level Seaborn function for creating relational plots that accepts parameters like kind, hue, col, and row and returns a FacetGrid to support subplots

Scatter plot: A relational plot that displays individual observations as points on an x–y plane to visualize the relationship between two quantitative variables

Seaborn: A high-level Python visualization library built on Matplotlib that provides convenient functions for creating attractive statistical graphics with minimal code

Tidy data: A data organization principle where each row represents one observation and each column represents one variable, which is required for Seaborn to interpret columns as variables correctly