Univariate Analysis Structure

Numeric features

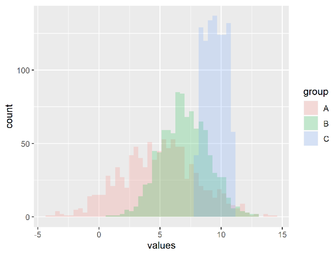
The prefix "Uni" signifies "one," so univariate analysis refers to examining a single variable at a time.

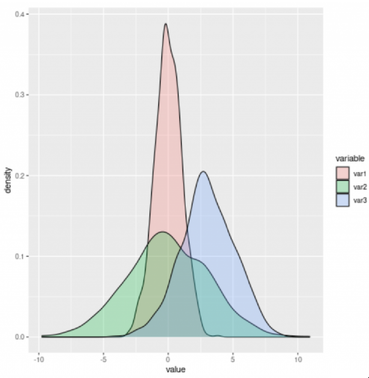
For numerical features, the goal is to understand the range of values and how frequently each value—or group of values—appears.

For categorical features, we focus on identifying the number of distinct categories and how often each category occurs

A histogram organizes data into ranges (or bins), with the height of each bar representing the number of values that fall within that range.

Example:
Histogram – Frequency distribution of a variable.





Density Plot

A density plot visualizes the distribution of a continuous variable using a kernel density estimate (KDE), which smooths the frequencies into a continuous curve. Unlike histograms, density plots do not rely on bins and provide a clearer view of the data's shape, including peaks and spread.

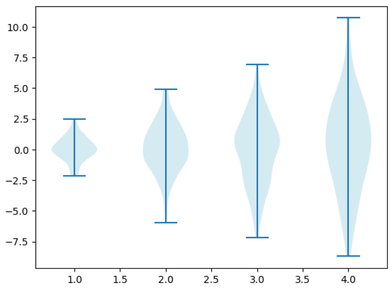
Use Case:
To observe the underlying distribution of a numeric feature, especially useful for identifying skewness or multimodal distributions.

Example:
Density Plot – A smoothed frequency distribution highlighting the shape of the data.

Violin Plot

A violin plot combines aspects of a box plot and a density plot. It shows the summary statistics (like the median and quartiles) along with the distribution shape, including peaks and skewness.

Example:
Violin Plot – Distribution with density curves.



Line Plot

A line plot is used to display the trend of a variable over time or another continuous axis.

Example:
Line Plot – Trend of a variable over time

Categorical Features
Categorical features are variables with a fixed number of distinct categories, such as gender, country, or age group.

Before visualizing them, we analyze summary statistics—like the number of unique categories per feature—to decide whether visualization is effective. Too many categories can make plots cluttered and hard to interpret.

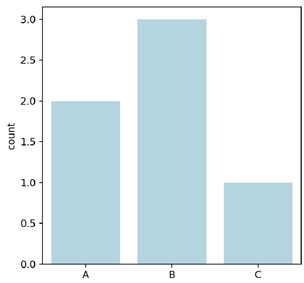
Count Plot Explained

What is a count plot?

It compares different classes of a categorical feature and how often they occur, like a bar chart showing the number of times each class appears.

When should I use a count plot?

For categorical features with fewer than 10 unique classes.

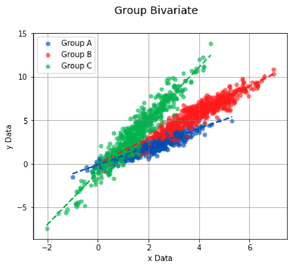


Understanding Pie Charts

Pie Chart

A pie chart represents the percentage distribution of a categorical variable using slices of a circle.

Example:
Pie Chart – Percentage distribution for categorical features with fewer than 10 classes.



Bivariate Analysis

Bivariate analysis explores relationships between two variables. This is commonly visualized using scatter plots, line plots, or correlation matrices, depending on the type of data.

Example:
Bivariate Analysis – A scatter plot showing the relationship between two variables.