Seaborn

Seaborn is a popular library for data visualization in Python built on Matplotlib, offering a user-friendly way to create visually appealing and informative statistical graphics. It is used by many people, including data scientists, statisticians, and business analysts.

**Seaborn vs Matplotlib**

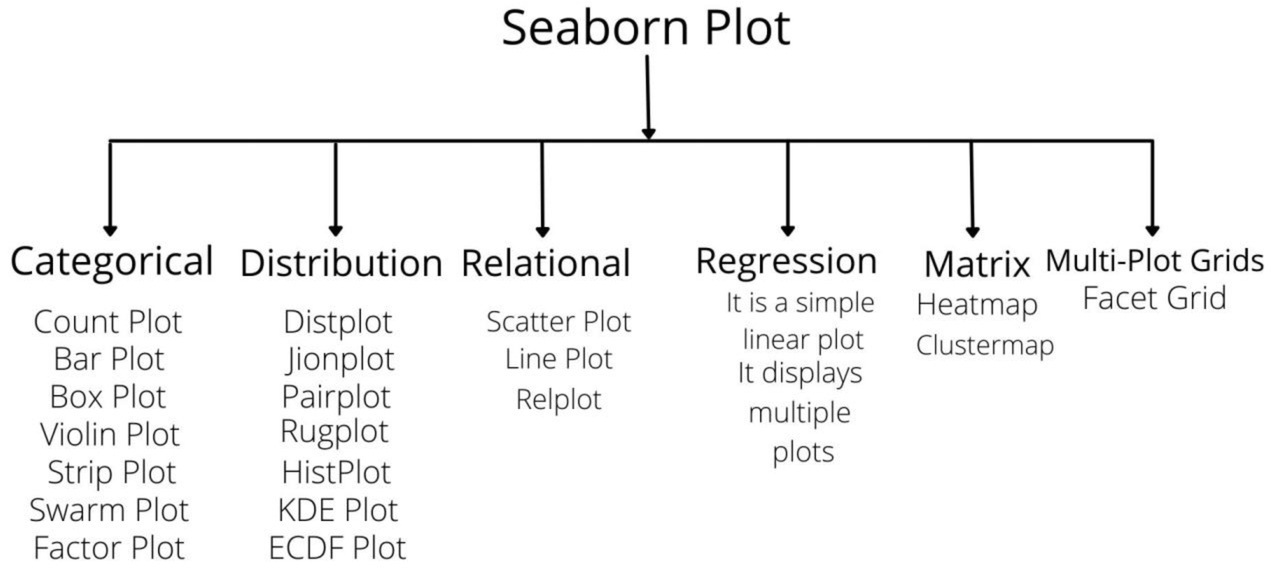
|  |  |
| --- | --- |
| **Seaborn** | **Matplotlib** |
| * Seaborn provides numerous default themes and a diverse range of schemes for statistical visualization. | * Matplotlib is commonly used for creating simple charts. |
| * Seaborn automates the creation of multiple figures more easily than Matplotlib such as the “FacetGrid” function | * Matplotlib does not create multiple figures more easily than Seaborn |
| * Seaborn is the best for statistical data visualization. | * Matplotlib is more suitable for customization needs. |

* Seaborn and Matplotlib both make graphs, but Seaborn lets you make them fancier and cooler. With Seaborn, you can change things more and your graphs look fancier and they look better.
* When you want to make graphs in Matplotlib, you might need to write some complicated Python code. Seaborn makes this easier and is friendlier for beginners. It's like Seaborn speaks a simpler language, making it easier to learn and use.

Before delving into the exploration of different plot functions and graphs, it's important to ensure that we have a clear understanding of the hierarchy.

**Seaborn graphs have six broad categories.**

1. Categorical Plot
2. Distribution Plot
3. Relational Plot
4. Regression Plot
5. Matrix Plot
6. Residual Plot
7. Multi Plot



**Here are several common types of plots that are often utilized**

* Histogram
* Scatter Plots
* Box Plots (Box-and-Whisker Plots)
* Violin Plots
* Bar Plots
* Line Plots
* Heatmaps
* Pair Plots
* Correlation Matrix Heatmap
* Pie Charts

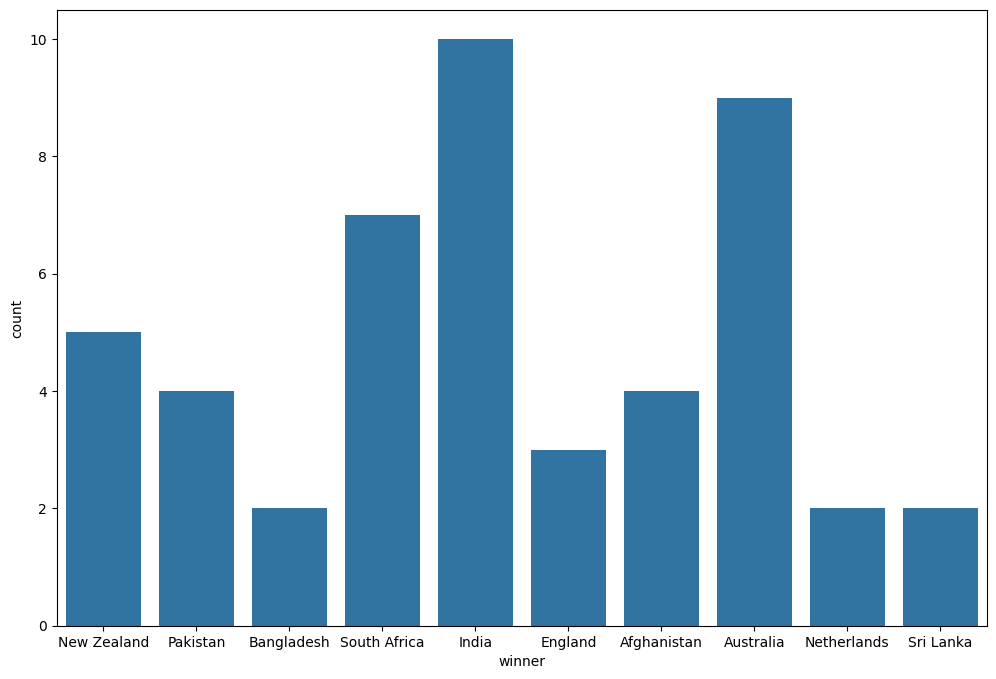
1. **Categorical Plot**

Categorical plots are used to visualize the distribution and relationships between categorical variables, it is a way to represent the distribution of a variable or the relationship between two variables across categories. Some Seaborn Categorical Plots

**Seven plots belong to the Categorical Plot category.**

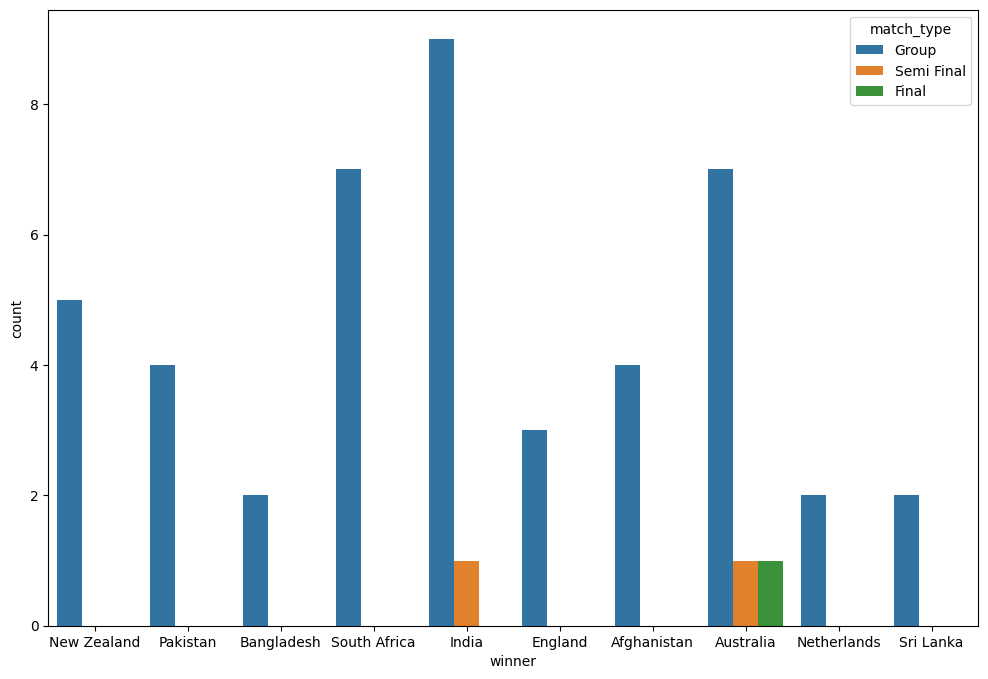
1. **Count Plot**

A count plot is a type of graph that displays the number of occurrences of categorical data. Categorical plots help us see how different categories are spread out and compare how often each category appears in our data.



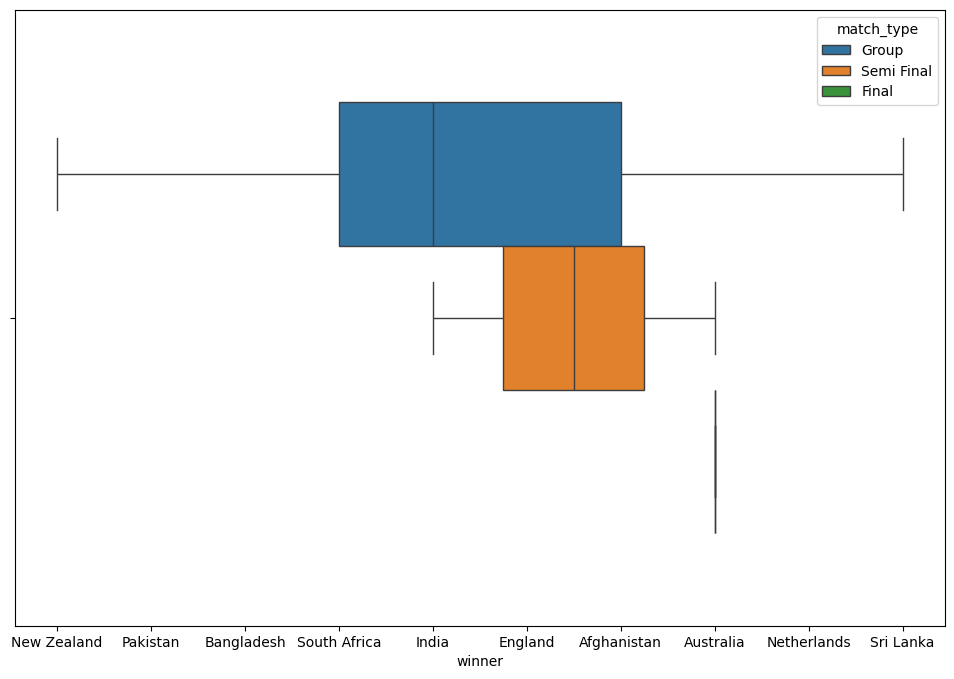
1. **Bar Plot**

A bar chart is like a picture made of rectangles. It's used to compare things, usually in different categories. A bar plot is a simple way to show how often different categories occur in data or how they relate to each other.



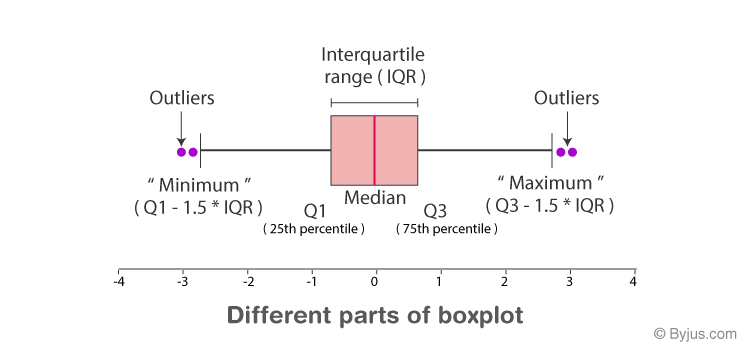
1. **Box Plot**

A box plot, or box-and-whisker plot, shows the spread and key details of a dataset, helping to visualize its distribution and summary statistics simply.



**Parts of Box Plots**

The key feature of a Box Plot is its depiction of the five-number summary: minimum, first quartile (Q1), median (second quartile, Q2), third quartile (Q3), and maximum. This is especially useful for identifying outliers in a dataset.



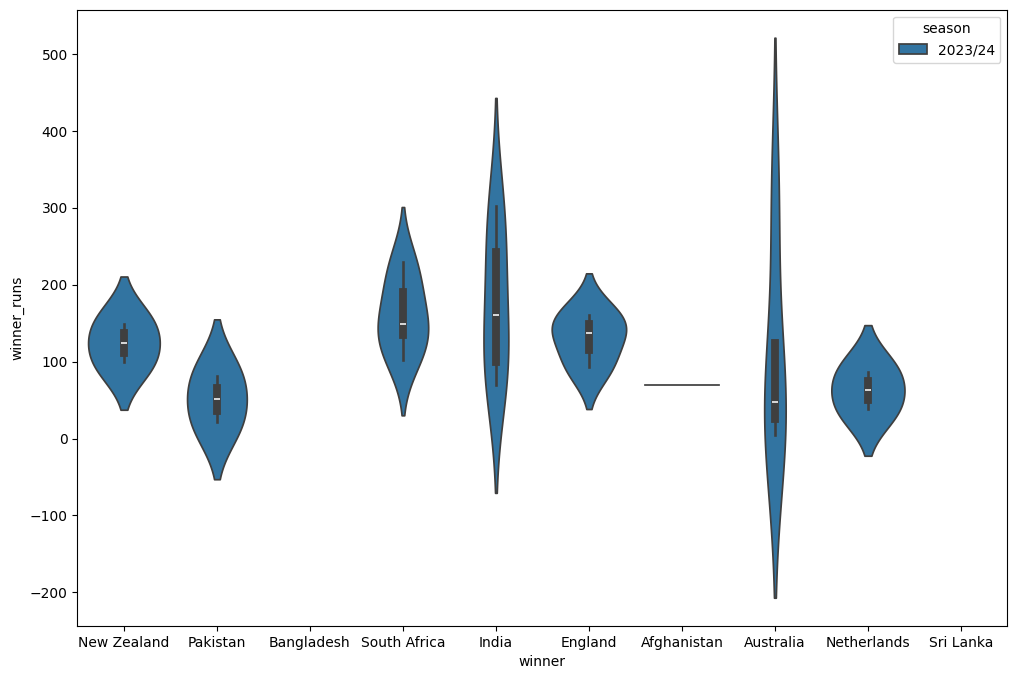
* **Minimum** The smallest value in the dataset.
* **Maximum** The largest value in the dataset.
* **First Quartile (Q1)** The median of the lower half of the dataset, representing the 25th percentile.
* **Median** The middle value of the dataset, dividing it into two halves; also known as the second quartile.
* **Third Quartile (Q3)** The median of the upper half of the data, representing the 75th percentile

**In addition to these, another important term is**

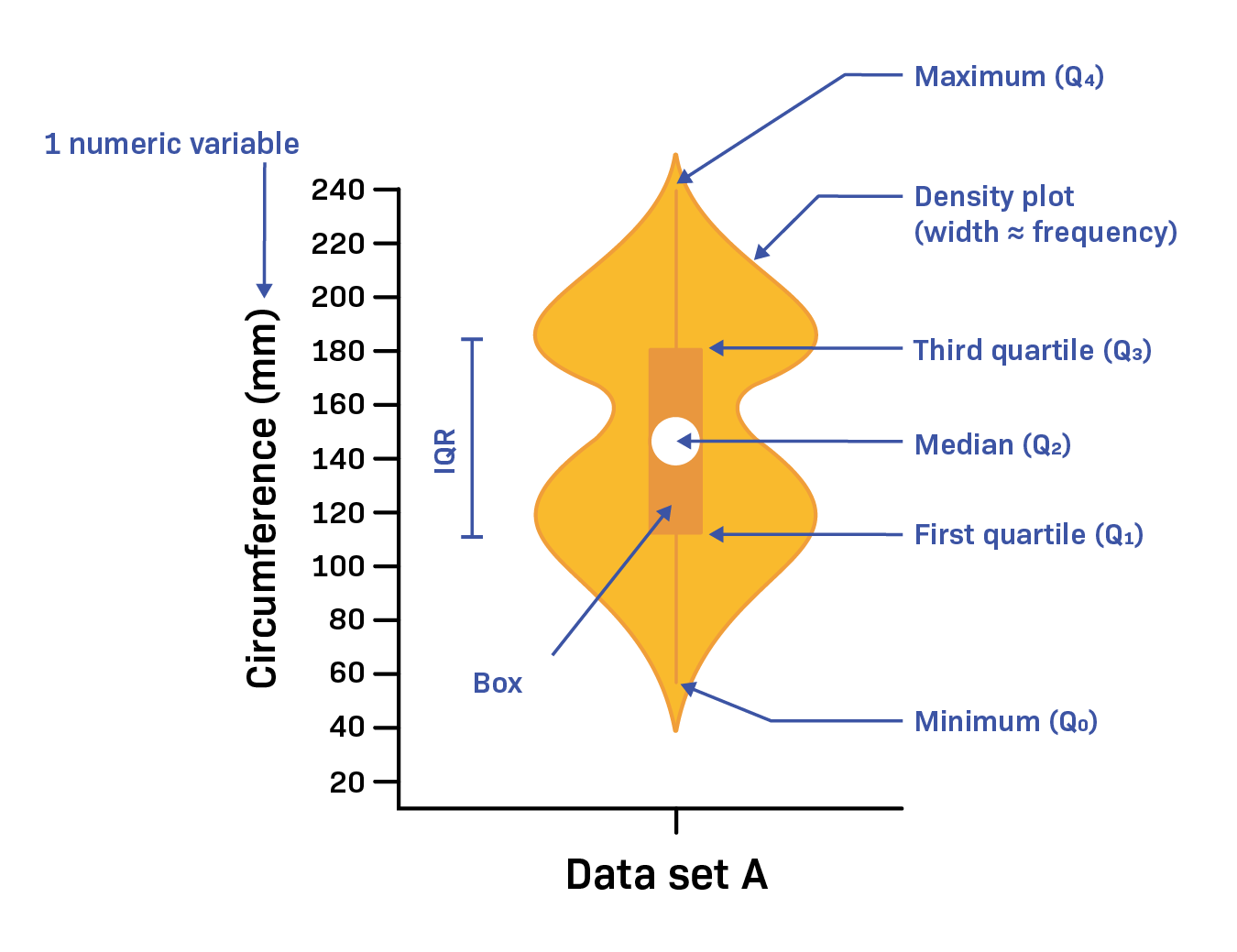
* **Interquartile Range (IQR)** The range between the third quartile (Q3) and the first quartile (Q1), calculated as IQR = Q3 - Q1. It provides a measure of the spread of the middle 50% of the data and is useful for identifying the variability in this central portion of the dataset.

1. **Violin Plot**

It is similar to the box plot. It holds more information as compared to the box. Violin plots are useful for comparing the distributions of different categories and identifying patterns in the data.

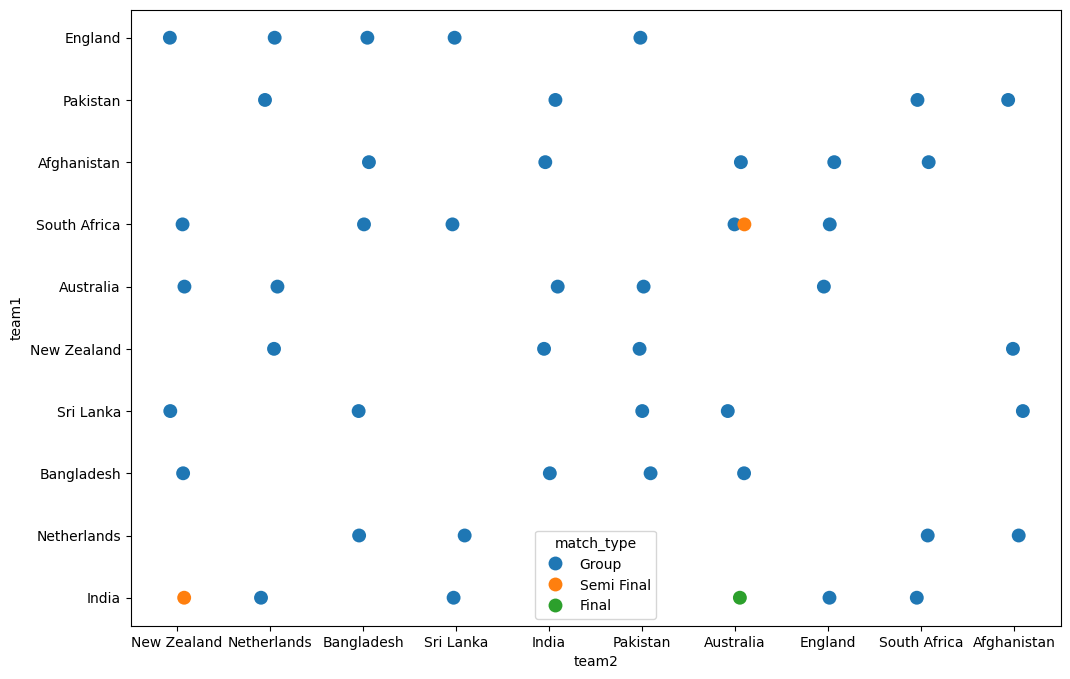


**Parts of Violin Plot**



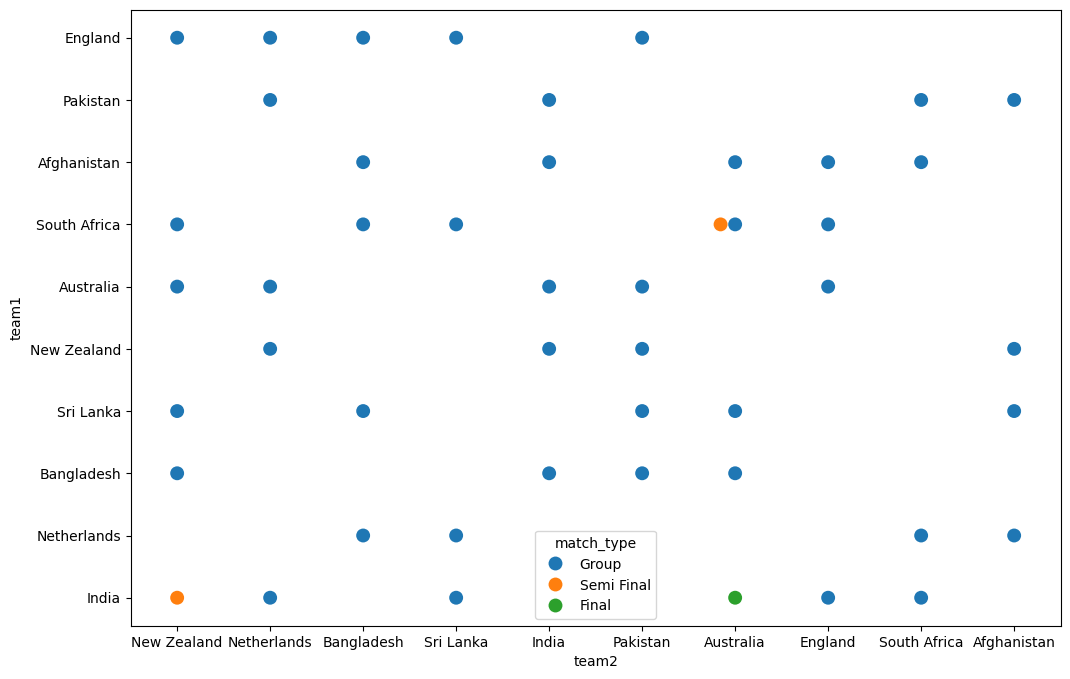
1. **Strip Plot**

A strip plot is a type of categorical scatter plot that displays individual data points along a single axis. A strip plot is a type of scatter plot with a single axis, commonly used to display the distribution of individual one-dimensional values.



1. **Swarm Plot**

It is similar to a strip plot, but it avoids overlapping points by adjusting their positions along the categorical axis. A swarm plot is a type of categorical scatter plot that displays all individual data points, helping to visualize the distribution and density of the data more effectively than traditional scatter plots.



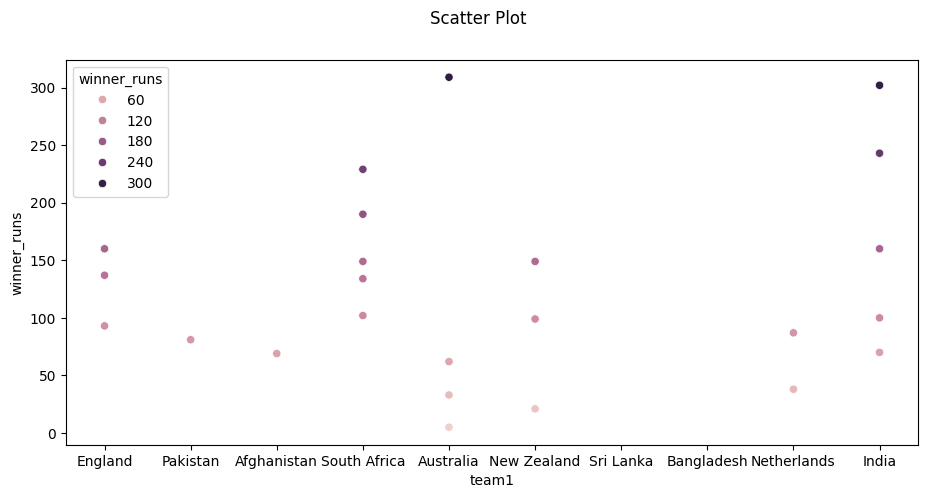
* **Relational Plot**

A relational chart or graph is a type of data visualization that displays the relationship between two or more variables. The plots help analysts and researchers understand patterns, trends, and correlations in the data.

**Three plots belong to the Relational Plot category.**

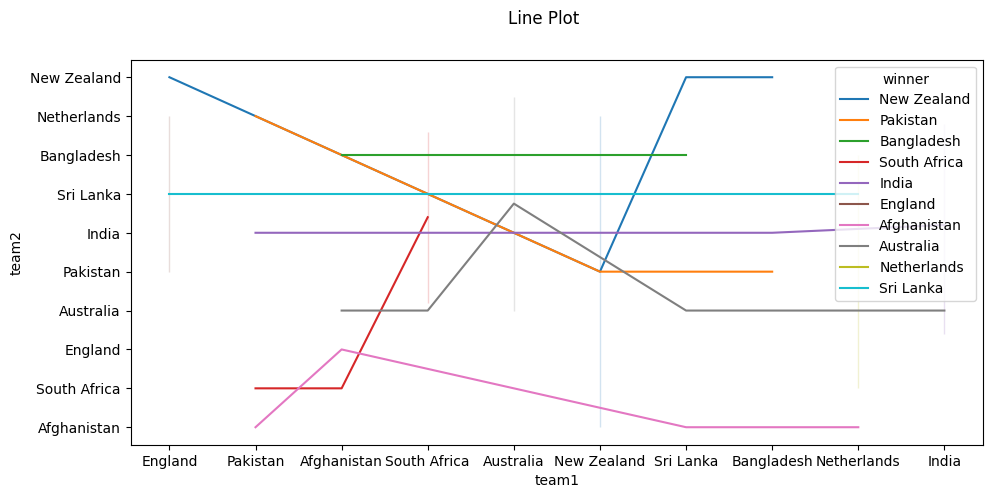
1. **Scatter plot**

A scatter plot is a popular type of relational plot that shows individual data points on a two-dimensional plane. Scatter plots are used to visually explore and understand the relationship between two continuous variables in a dataset.



1. **Line plot**

A line plot connects data points with a line, illustrating the relationship between x and y. We can display this connection using colors, sizes, and styles, making it easier to see how they relate in different sets or groups.



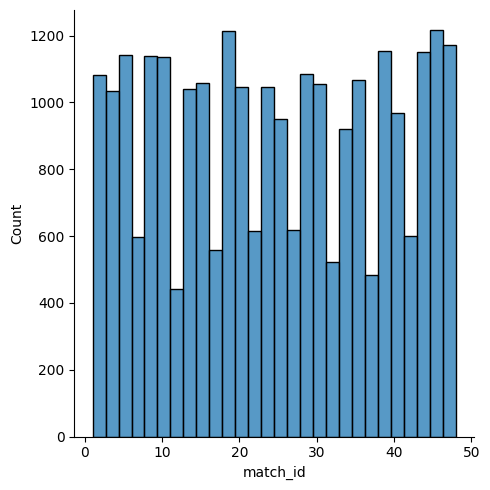
* **Distribution Plot**

A distribution plot, also known as a histogram or density plot, is a graphical representation of the distribution of a dataset. A distribution plot is a data visualization that provides insights into the distribution of a continuous variable

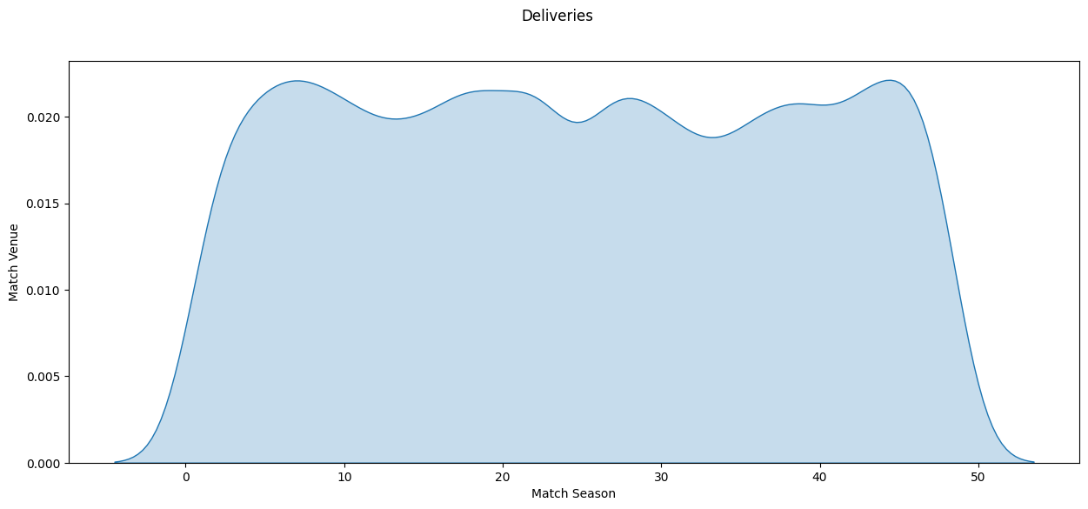
.

**There are common types of distribution plots**

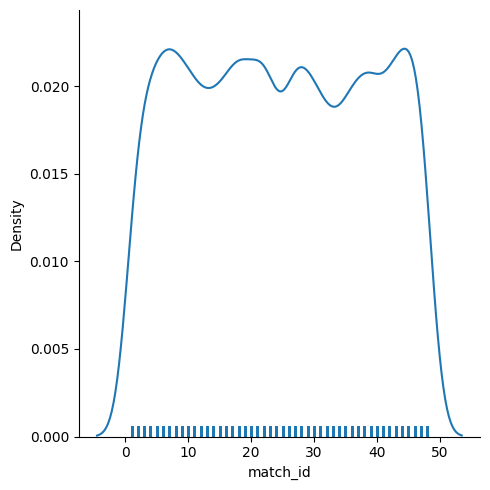
1. **Histogram** A histogram is a bar graph that represents the distribution of a continuous variable.



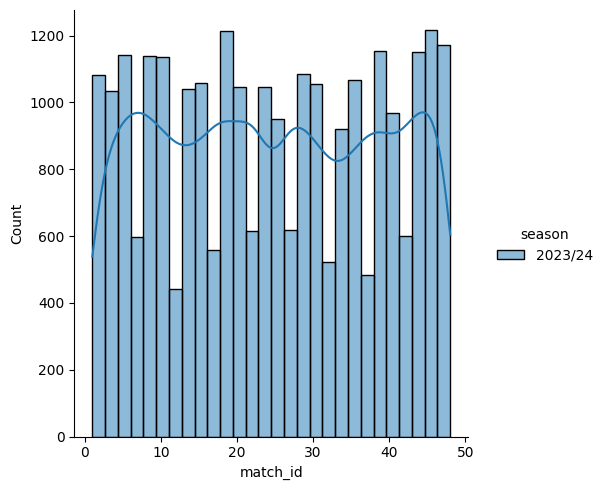
1. **KDE Plot (Kernel Density Estimate)** A KDE plot is a smoothed curve that estimates the probability density function of a continuous variable.



1. **Rug Plot** A rug plot is a one-dimensional scatter plot that places a small tick for each data point along an axis.

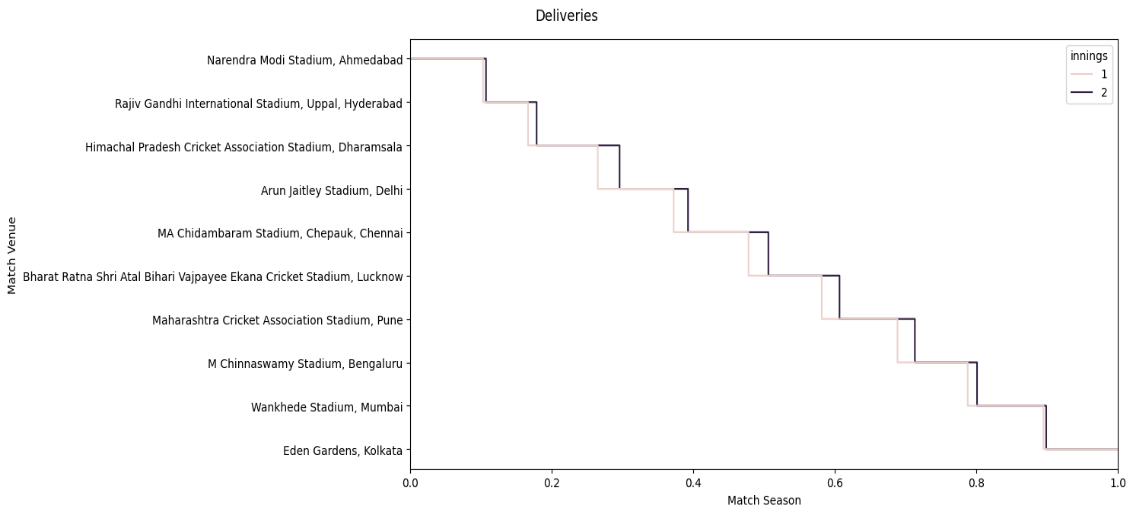
****

**A combined rug plot, kernel density estimate (KDE) plot, and histogram.**

****

1. **ECDF**

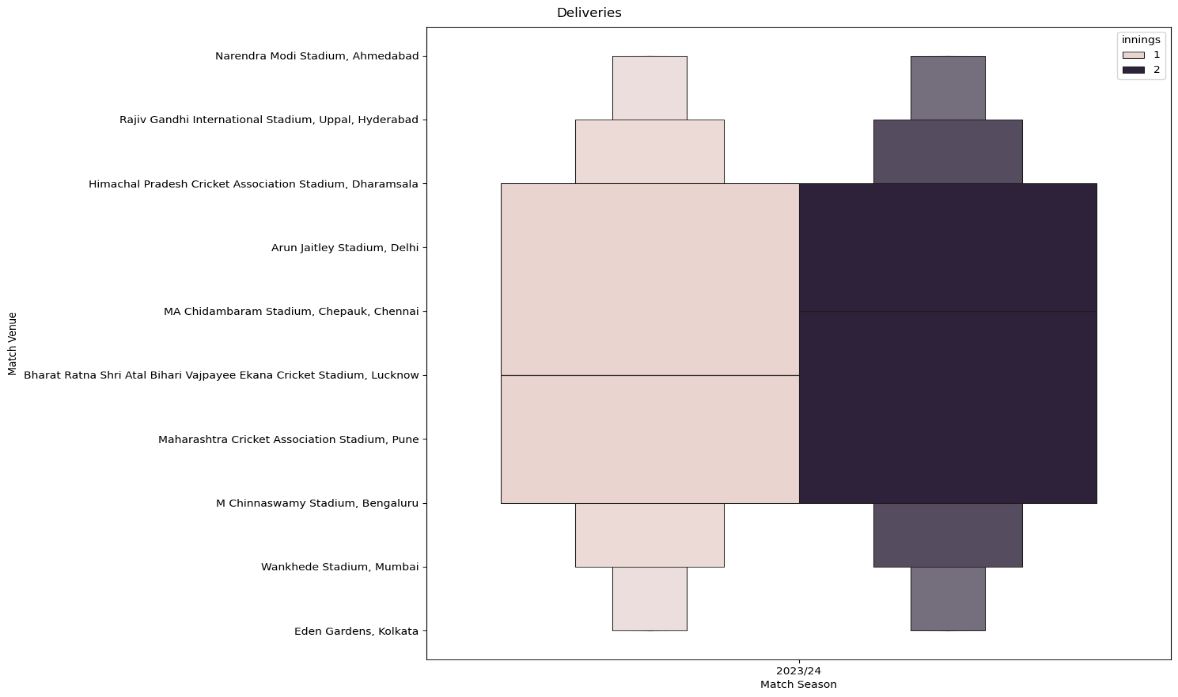
ECDF stands for “Empirical Cumulative Distribution Function. “It’s a statistical technique used to describe the distribution of a set of observations in data analysis. It helps compare different sets of data or understand the overall distribution of a single set of data. A scatter plot is advantageous compared to a histogram or KDE plot because it visualizes each data point directly, making it easier for users to interact with the plot.



1. Boxen plot

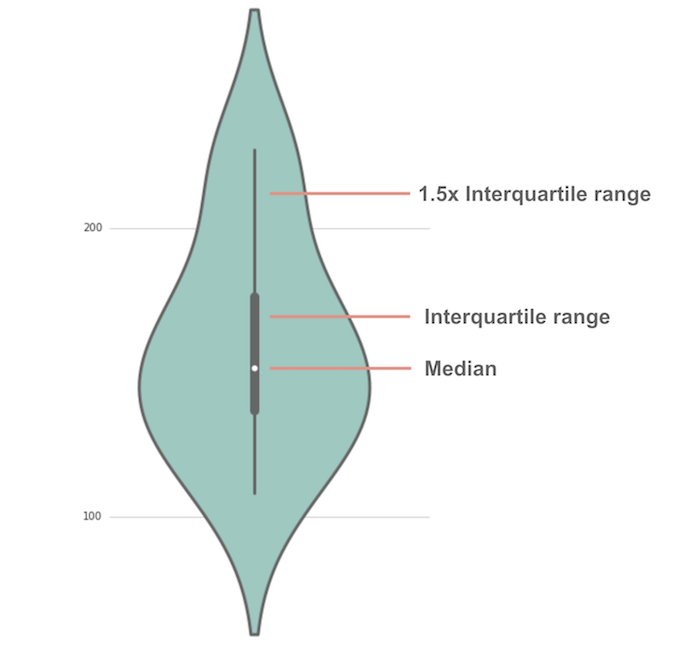
A Boxen plot is a type of graph that displays quantitative data. It's an effective and simple way to represent median and quartiles visually. **The plot is made up of four rectangles:**

Box, Whiskers, Lower Quartile (LQ), and Upper Quartile (UQ). The height of the box represents the median, and the length of the whiskers represents the interquartile range (IQR).



1. Violin plot

A violin plot, also known as a box plot or quadratic box plot, is a graphical representation of a set of ordered numerical data. . The box length shows the interquartile range, and the whiskers extend to the min and max values. It is useful for visualizing skewed or asymmetrical data distributions.



**Violin plots share similar summary statistics with box plots.**

* The median is represented by the white dot.
* The wide gray bar at the center depicts the interquartile range.
* The narrow gray line depicts the remaining distribution, excluding points identified as "outliers" based on a method involving the interquartile range.

