# VISUALIZATION

Visualization is the process of creating representations of different concepts. Data Visualization is when you create representations of data to better understand it.

Python has various libraries which make performing visualization easier. There are multiple libraries like matplotlib, seaborn, plotly etc.

We shall discuss about the two of the most prominent libraries for visualization in python: Matplotlib and Seaborn.

## Matplotlib

Matplotlib, introduced in 2002 by John Hunter, is a powerful library for creating static, animated, and interactive visualizations in Python. It is primarily used for plotting 2D plots and is built on top of NumPy.

It offers a wide range of plot types and customization options.

### Line Chart

A line chart is a basic plot where data points are connected by a continuous line. It is suitable for visualizing trends over time.

A screenshot of a computer program

Description automatically generated

A blue line graph with numbers

Description automatically generated

### Bar Graph

A bar graph uses rectangular bars to represent data. It is effective for comparing quantities or frequencies across different categories.

A screen shot of a computer code

Description automatically generated

A graph of a bar graph

Description automatically generated with medium confidence

### Histogram

A histogram displays the distribution of a continuous variable by dividing the data into bins and counting the number of observations in each bin.

A screen shot of a computer code

Description automatically generated

A diagram of a histogram

Description automatically generated

### Scatter Plot

A scatter plot uses dots to represent the relationship between two or more variables. It is useful for identifying patterns or correlations in data.

A number on a white background

Description automatically generated

A graph with blue dots

Description automatically generated

### Box Plot

A scatter plot uses dots to represent the relationship between two or more variables. It is useful for identifying patterns or correlations in data.

A computer screen shot of a program code

Description automatically generated

A diagram of a square with a line in the middle

Description automatically generated

### Pie Chart

A scatter plot uses dots to represent the relationship between two or more variables. It is useful for identifying patterns or correlations in data.

A computer screen shot of a number

Description automatically generated

A colorful circle with white text

Description automatically generated

## Seaborn

Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative plots.

### Relational Plots

Relational plots are used to visualize the relationship between variables. They are effective for exploring correlations and trends in data.

Examples:

* Scatter Plot: Scatter plots are used to visualize the relationship between two continuous variables. Each point represents an observation in the dataset, and the position of the point on the plot corresponds to the values of the two variables.

A computer code with black text

Description automatically generated

A graph with blue dots

Description automatically generated

* Line Plot: Line plots are used to visualize trends or patterns over time or another continuous variable. They are particularly useful for visualizing time series data or data collected at regular intervals.

A screenshot of a computer code

Description automatically generated

A graph with a line

Description automatically generated

* Relplot: Relational plots are a generalization of scatter plots and line plots. They can display the relationship between two variables using different markers or lines, depending on the value of a third variable.

A screen shot of a computer code

Description automatically generated

A graph with numbers and dots

Description automatically generated

### Categorical Plots

Categorical plots are used to visualize the distribution of categorical variables. They are useful for comparing groups or categories within the data.

Examples:

* Bar Plot: Bar plots are used to visualize the distribution of a categorical variable by displaying the frequencies or proportions of each category as rectangular bars.

A computer screen shot of a program

Description automatically generated

A graph of different colored rectangular shapes

Description automatically generated with medium confidence

* Box Plot: Box plots, also known as box-and-whisker plots, are used to visualize the distribution of a continuous variable within different categories. They provide a visual summary of the central tendency, spread, and skewness of the data.

A black text with red and blue text

Description automatically generated

A graph of different colored boxes

Description automatically generated

* Violin Plot: Violin plots are used to visualize the distribution of a continuous variable across different categories. They combine elements of a box plot and a kernel density plot to provide insights into the data distribution.

A screenshot of a computer program

Description automatically generated

A blue and black line graph

Description automatically generated

### Distribution Plots

Distribution plots are used to visualize the distribution of a single variable. They help understand the underlying distribution of the data.

Examples:

* Histogram: Histograms are used to visualize the frequency distribution of a continuous variable by dividing the data into bins and displaying the count or density of observations within each bin.

A computer screen shot of a penguin

Description automatically generated

A graph of a body mass

Description automatically generated

* Kernel Density Estimation (KDE) Plot: KDE plots are used to estimate the probability density function of a continuous variable. They provide a smooth curve that represents the distribution of the data.

A white background with black text

Description automatically generated

A graph of different species

Description automatically generated

### Heatmap

Heatmaps are used to visualize two-dimensional data in a matrix format. They use colors to represent the values of the data, making it easy to identify patterns and relationships.

A screen shot of a computer code

Description automatically generated

A screenshot of a grid

Description automatically generated

### Pair Plot

Pair plots are used to visualize pairwise relationships between variables in a dataset. They create a grid of scatterplots for each pair of variables, along with histograms for each variable on the diagonal.

A screen shot of a computer program

Description automatically generated

A group of graphs showing different sizes of data

Description automatically generated with medium confidence

# Comparison between matplotlib and seaborn libraries

## Matplotlib

**Core library:** Matplotlib is a comprehensive plotting library with low-level control over plot elements, making it highly customizable.

**Versatility:** It offers a wide range of plotting functions and styles, allowing users to create virtually any type of plot, including line plots, scatter plots, bar charts, and many more.

**Steep learning curve:** Matplotlib has a steep learning curve, especially for beginners, due to its extensive functionality and complex API. Users may need to invest time in learning the library’s syntax, concepts, and best practices for creating effective plots.

**Advantages of Matplotlib:**

* High degree of customization and control over plots.
* Wide range of plot types and styles.
* Strong community support and extensive documentation.
* Well-suited for creating publication-quality figures and graphics.

## Seaborn

**Integration with pandas:** Seaborn seamlessly integrates with pandas data frames, allowing users to directly pass data from pandas structures to seaborn plotting functions.

**Attractive defaults:** Seaborn comes with aesthetically pleasing default styles and color palettes, making it easy to create visually appealing plots without extensive customization.

**High-level interface:** Seaborn’s high-level API simplifies the process of creating complex statistical plots by abstracting away the tedious details of data manipulation and plot configuration.

**Advantages of Seaborn:**

* Simplified syntax and high-level functions for creating complex plots.
* Built-in support for statistical plotting and data exploration.
* Attractive default aesthetics and color palettes.
* Seamless integration with Pandas data frames for data visualization.
* Ideal for exploratory data analysis and quick visualization of relationships in data.

In summary, Matplolib provides low-level control and extensive customization options, making it suitable for creating highly customized plots. On the other hand, seaborn offers a high-level interface and attractive defaults, making it ideal for quickly creating complex statistical visualizations with minimal code.