SHADOWFOX DATA SCIENCE INTERNSHIP

VISUALIZATION LIBRARY DOCUMENTATION

BY MEGHANA P R

1. Library Overview

1.1 Seaborn

Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn integrates well with Pandas, making it easy to visualize data directly from dataframes. It is widely used for statistical analysis and exploratory data visualization.

Key Features:

- Built-in themes for aesthetically pleasing graphs
- Supports complex visualizations such as heatmaps and pair plots
- Works well with Pandas DataFrames
- Provides statistical plotting capabilities

1.2 Bokeh

Bokeh is a powerful interactive visualization library for modern web browsers. It allows users to create interactive and dynamic visualizations easily. Unlike Matplotlib and Seaborn, Bokeh is designed for building web-based visualizations and dashboards.

Key Features:

- Supports interactive visualizations
- Can handle large datasets efficiently
- Provides tools for linking multiple plots together
- Generates HTML and JavaScript outputs for web deployment

2. Graph Types

2.1 Seaborn Graphs

1. **Line Plot**

   Description: A line plot is used to visualize trends in data over time.

   Use Case: Tracking stock prices, temperature variations, or sales trends.

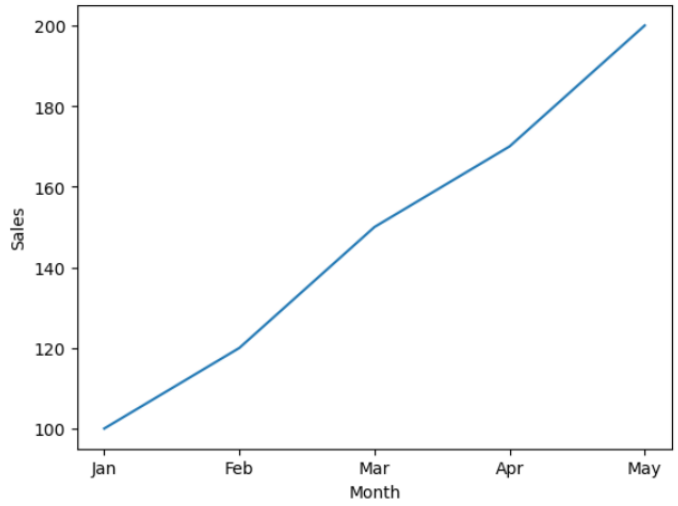
   Example:

```
In [1]: import seaborn as sns
        import matplotlib.pyplot as plt
        import pandas as pd

        data = pd.DataFrame({
            'Month': ['Jan', 'Feb', 'Mar', 'Apr', 'May'],
            'Sales': [100, 120, 150, 170, 200]
        })

        sns.lineplot(x='Month', y='Sales', data=data)
        plt.show()
```

```
C:\Users\megha\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\megha\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
  with pd.option_context('mode.use_inf_as_na', True):
```



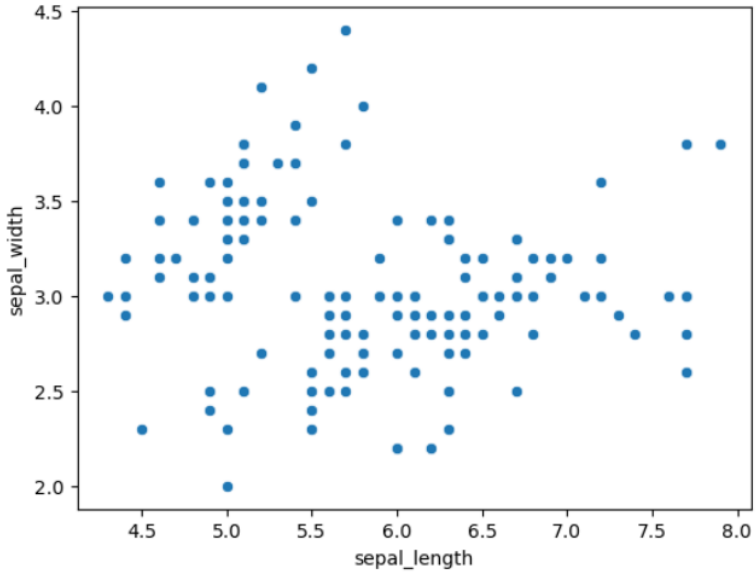
2. Scatter Plot

Description: Used to show relationships between two numerical variables.

Use Case: Analyzing correlations in datasets.

Example:

```
In [2]: sns.scatterplot(x='sepal_length', y='sepal_width', data=sns.load_dataset('iris'))
        plt.show()
```



3. Bar chart

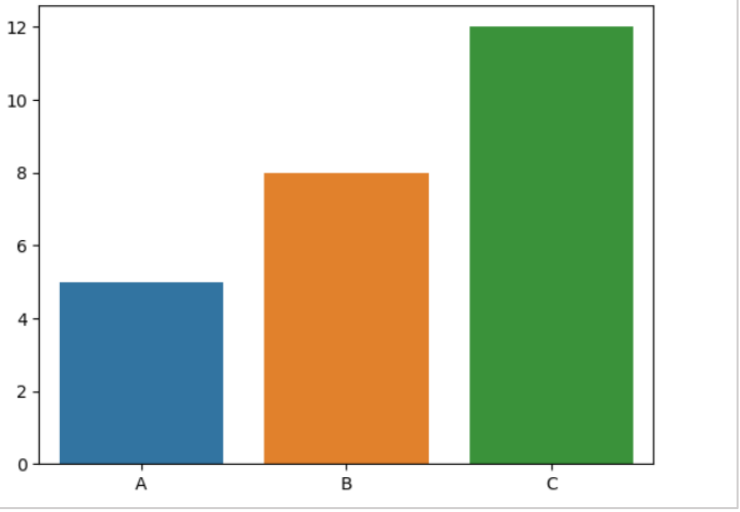
Description: A bar chart displays categorical data using rectangular bars.

Use Case: Comparing sales data for different products.

Example:

```
In [3]: sns.barplot(x=['A', 'B', 'C'], y=[5, 8, 12])
        plt.show()
```

```
C:\Users\megha\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1765: FutureWarning:
  order = pd.unique(vector)
```



4. Histogram

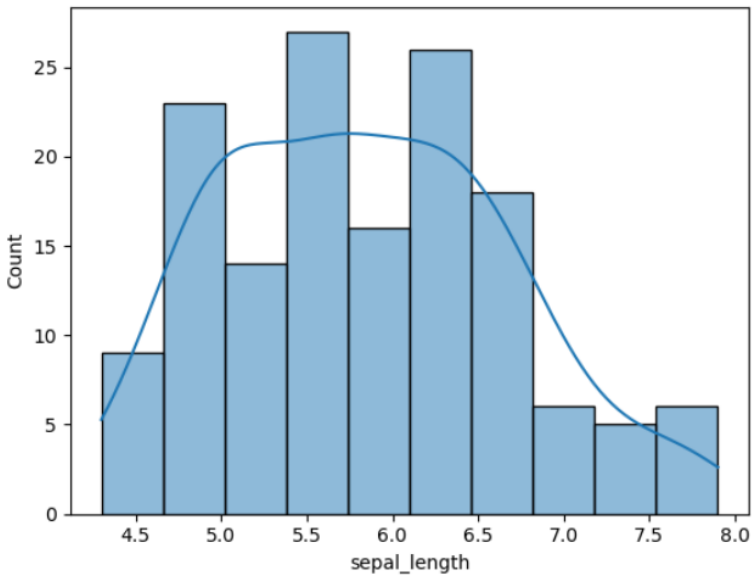
Description: A histogram represents the distribution of numerical data.

Use Case: Understanding the distribution of exam scores or salaries.

Example:

```
In [4]: sns.histplot(sns.load_dataset('iris')['sepal_length'], bins=10, kde=True)
        plt.show()
```

```
C:\Users\megha\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
  with pd.option_context('mode.use_inf_as_na', True):
```



2.2 Bokeh Graphs

1. Line Plot

Description: Similar to Seaborn, used to show trends over time.

Example:

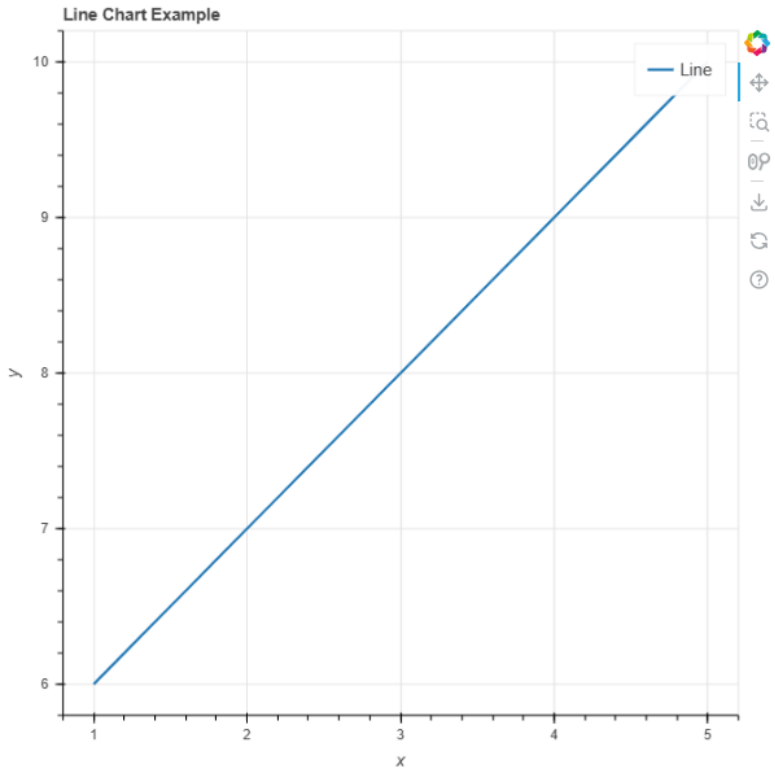
```
In [5]: from bokeh.plotting import figure, show
        from bokeh.io import output_notebook

        output_notebook()

        p = figure(title='Line Chart Example', x_axis_label='x', y_axis_label='y')
        p.line([1, 2, 3, 4, 5], [6, 7, 8, 9, 10], legend_label='Line', line_width=2)

        show(p)
```

BokehJS 3.3.4 successfully loaded.

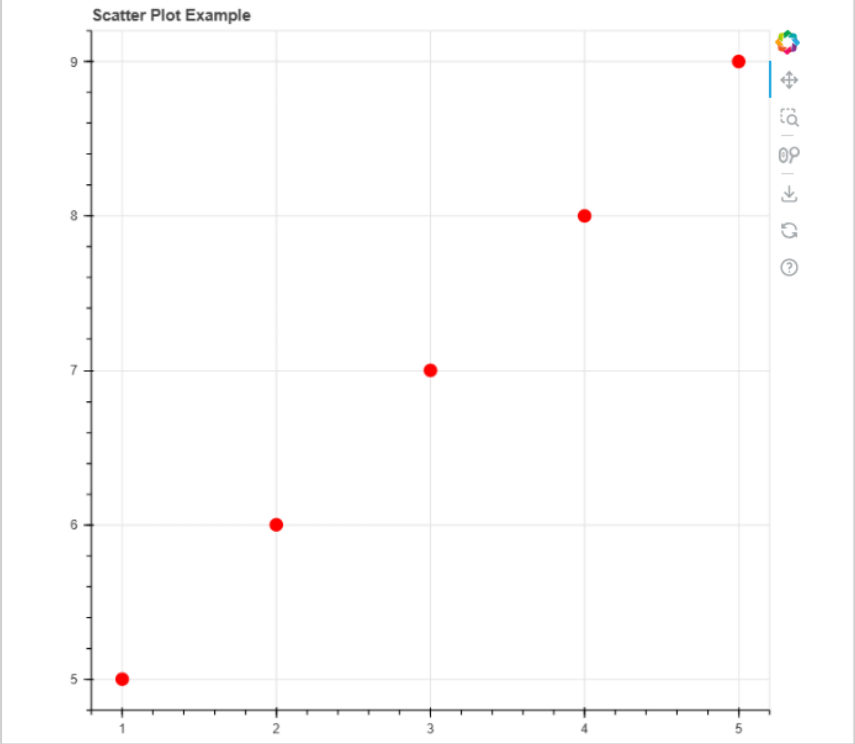


2. Scatter Plot

Description: Shows relationships between numerical variables.

Example:

```
In [6]: p = figure(title='Scatter Plot Example')
        p.scatter([1, 2, 3, 4, 5], [5, 6, 7, 8, 9], size=10, color='red')
        show(p)
```



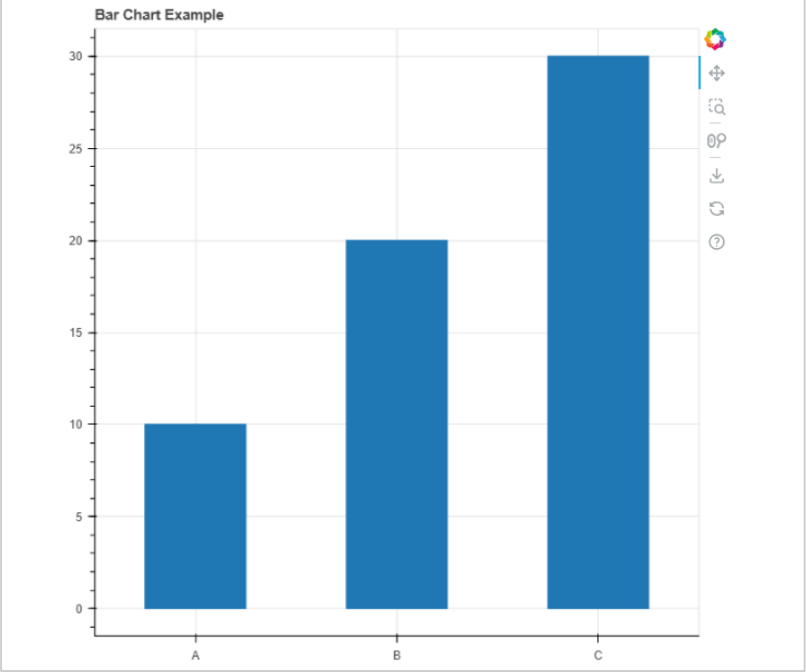
3. Bar chart

Description: Displays categorical data as bars.

Example:

```
In [7]: from bokeh.models import ColumnDataSource

        source = ColumnDataSource(data={'categories': ['A', 'B', 'C'], 'values': [10, 20, 30]})
        p = figure(x_range=['A', 'B', 'C'], title='Bar Chart Example')
        p.vbar(x='categories', top='values', width=0.5, source=source)
        show(p)
```



4. **Histogram**

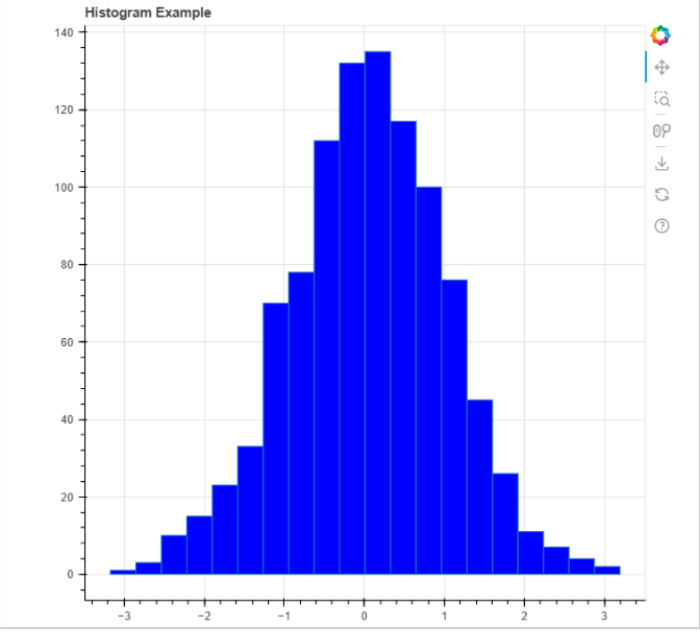
Description: Displays the distribution of numerical data.

Example:

```
In [8]: import numpy as np
        from bokeh.models import ColumnDataSource

        data = np.random.randn(1000)
        hist, edges = np.histogram(data, bins=20)

        p = figure(title='Histogram Example')
        p.quad(top=hist, bottom=0, left=edges[:-1], right=edges[1:], fill_color='blue')
        show(p)
```



3. Comparison

Feature	Seaborn	Bokeh
Ease of Use	Simple syntax, works well with Pandas	Requires more setup but allows interactive plots
Customization	Limited customization compared to Bokeh	Highly customizable
Interactivity	Static images	Interactive plots
Performance	Works well with moderate datasets	Handles large datasets efficiently
Web Integration	Not designed for web use	Generates HTML & JavaScript for web deployment

Conclusion

- Seaborn is excellent for statistical analysis and quick visualizations, especially when working with Pandas.
- Bokeh is ideal for building interactive visualizations, dashboards, and handling large datasets.
- If you need interactive visualizations, choose Bokeh. For quick and beautiful statistical plots, go with Seaborn.