



PER SCHOLAS

Data Visualization With Seaborn



Prerequisites

Learners should be proficient in Python programming and Python Pandas, and should install Python and Jupyter Notebook on their machines, and be able to access a cloud-based notebook.



Table of Contents

- Overview of Seaborn Library
- Introduction to Seaborn Library
- Installing and Getting Started
- Seaborn Plot Types
- Seaborn Functions
- Common list of Seaborn Functions
- load_dataset() method in Seaborn
- Seaborn vs. Matplotlib

Learning Objectives

By the end of this lesson, learners will be able to:

- Describe the Seaborn library, including its purpose, features, and advantages over other visualization libraries.
- Install the Seaborn library.
- Create plots using Seaborn library.
- Explain the various plot types available in Seaborn and know when to use each type.
- Identify the different functions provided by Seaborn for data visualization.
- Identify common Seaborn functions and their applications in visualizing data.
- Use the `load_dataset()` method in Seaborn for loading sample datasets.
- Compare Seaborn with Matplotlib, including the differences and similarities in data visualization capabilities.

Overview of Seaborn Library

- Seaborn is another data visualization library, built on top of Matplotlib.
- Seaborn offers a range of built-in themes and color palettes that improve the visual appeal of your plots with minimal effort.
- The key benefit of Seaborn is its ability to create visually appealing plots with minimal code. With Seaborn, you can generate plots using five times less code compared to Matplotlib.
- Seaborn integrates well with statistical libraries such as NumPy and SciPy, allowing you to easily combine statistical analysis with visualizations.
- Seaborn provides several functions to customize the charts. In addition, you can also use matplotlib functions for further customizations.

Introduction to Seaborn library

- Seaborn works well with Pandas DataFrames so you can load your data into a DataFrame using Pandas.
- **Choose a plot type:** Seaborn provides a variety of plot types for different kinds of data and visualization needs. Some popular plot types include scatter plots, line plots, bar plots, histograms, and heatmaps.
- **Customize your plot (optional):** Seaborn allows for extensive customization of plots to meet your specific requirements. You can adjust colors, styles, axis labels, titles, and more.
- **Show or save your plot:** Once your plot is ready, you can display it directly in your Python environment or save it to a file for later use.
- Additionally, Seaborn offers a range of built-in statistical functions, allowing users to easily perform complex statistical analysis with their visualizations.

Installing and Getting Started

Official releases of Seaborn can be installed from PyPI, by using below command:

```
pip install seaborn
```

The library is also included as part of the Anaconda distribution, and it can be installed with conda:

```
conda install seaborn
```

Import Seaborn: Once installed, you can import Seaborn in your Python script or notebook:

```
import seaborn as sns
```

```
import seaborn.objects as so
```

Mandatory dependencies:

- numpy
- pandas
- matplotlib

- **Note:** If you are working in a Jupyter notebook or an IPython terminal with matplotlib mode enabled, you should immediately see the plot; otherwise, you may need to explicitly call `matplotlib.pyplot.show()`.

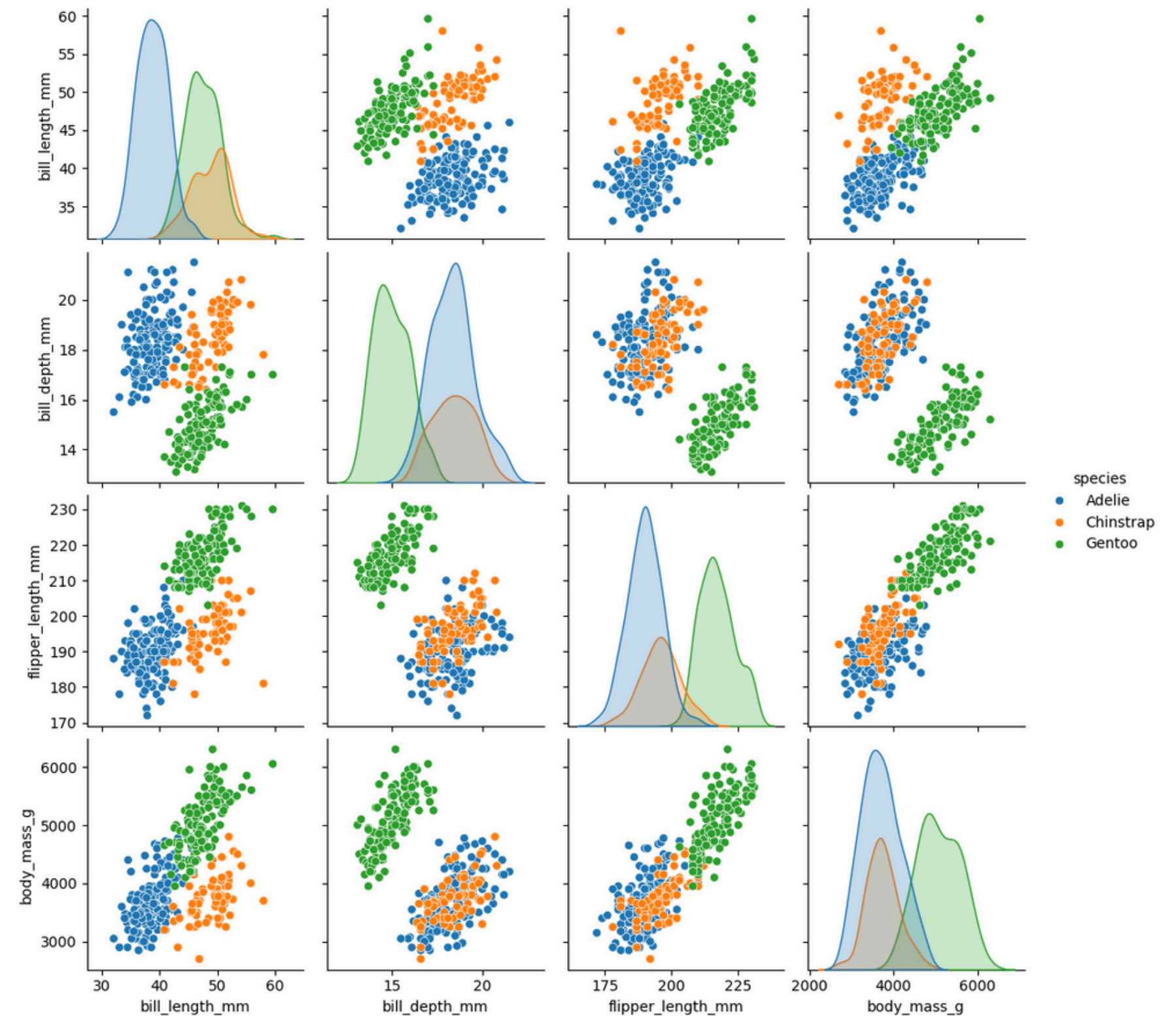
Reference: <https://seaborn.pydata.org/installing.html>

Basic Example

- `import seaborn as sns`
- `df = sns.load_dataset("penguins")`
- `sns.pairplot(df, hue="species")`

Note: If you are working in a **notebook** or an **IPython** terminal with matplotlib mode enabled, you should immediately see the plot. Otherwise, you may need to explicitly call `matplotlib.pyplot.show()` function, as shown below:

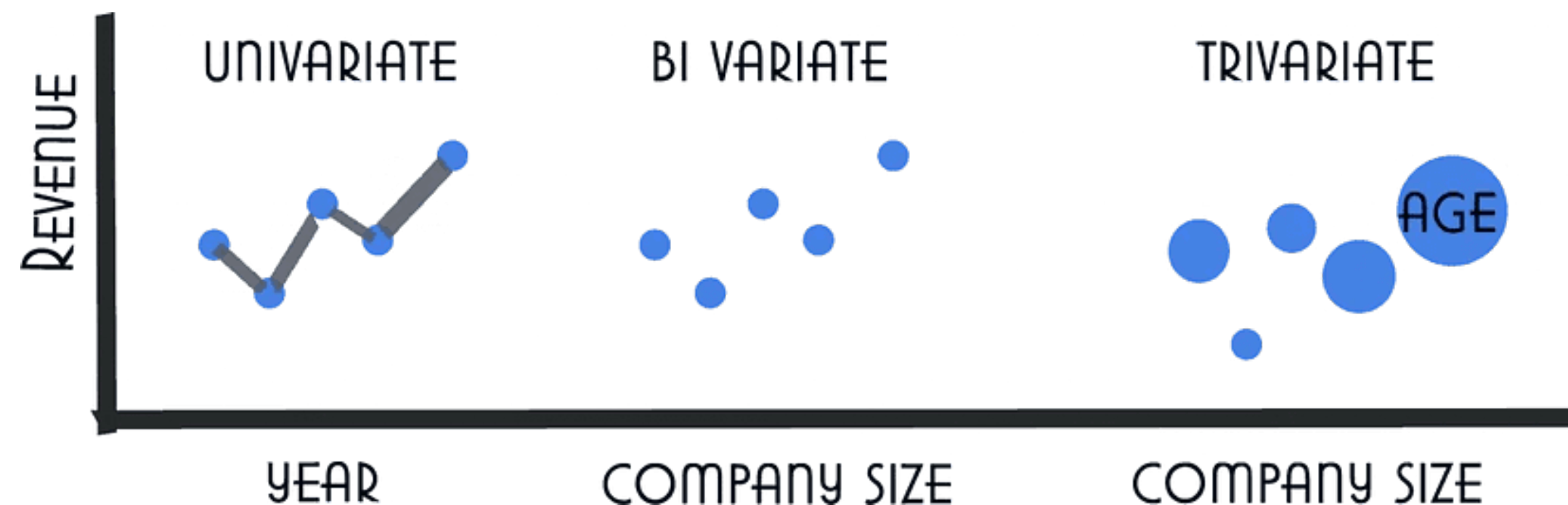
```
import matplotlib.pyplot as plt  
plt.show()
```



Seaborn Plot Types

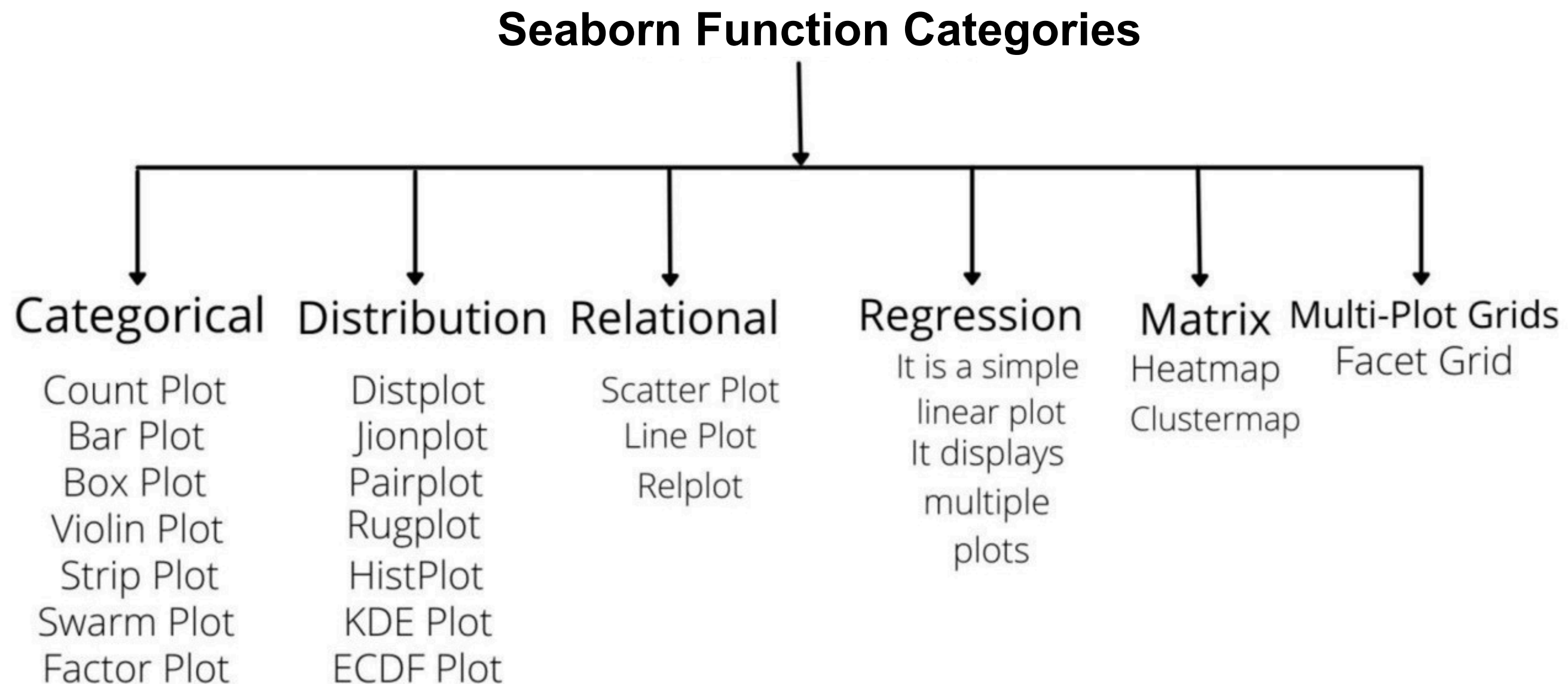
Seaborn offers a diverse range of plot types for data visualization and exploratory data analysis. These visualizations can broadly fall into one of three categories:

1. Univariate - Utilizing only one axis of information (x).
2. Bivariate - Utilizing two axes of information (x and y).
3. Trivariate - Utilizing three axes of information (x, y, and z).



Seaborn Functions

- The Seaborn functions for the plots are categorized into six categories as shown in the image below.



img src: seaborn.pydata

Common List of Seaborn Functions

Here are some of the most common Seaborn functions:

- **Scatter Plot** - Seaborn's `scatterplot()` function provides a simple way to create scatter plots.
- **Line Plot** - Seaborn's `lineplot()` function provides a simple way to create line plots.
- **Histogram** - Seaborn's `histplot()` function provides a simple way to create histograms.
- **Box Plot** - Seaborn's `boxplot()` function provides a simple way to create box plots.
- **Heatmap** - Seaborn's `heatmap()` function provides a simple way to create heatmaps.
- **Pairplot** - Seaborn's `pairplot()` function provides a simple way to create pairplots.
- **Linear Regression** - `seaborn.regplot()` function is used to plot data and draw a linear regression model
- **fit and seaborn** - `residplot()` function is used to plot residual data of a linear regression.
- **Density plots:** `kdeplot()` function is used to plot univariate or bivariate plots using kernel density estimation.

[Click here to see complete list of the Seaborn function.](#)

load_dataset() method in Seaborn

The **load_dataset()** method is used to load **built-in datasets** from the Seaborn library, allowing users to load sample datasets directly into their Python environment. These datasets are often used for demonstration purposes and can be conveniently accessed for practicing data visualization with Seaborn. The function returns a Pandas DataFrame containing the specified dataset.

Example: Loading built-in Datasets:

```
iris = sns.load_dataset("iris")
tips = sns.load_dataset("tips")
exercise = sns.load_dataset("exercise")
titanic = sns.load_dataset("titanic")
flights = sns.load_dataset("flights")
fmri = sns.load_dataset("fmri")
```

--> [Click here to see entire list of built-in datasets.](#)

Seaborn Example

You can find following labs on Canvas under the Assignment section:

- [Guided LAB - 344.4.1 - Example Python Seaborn - Start Visualizing Data.](#)
- [Guided LAB - 344.4.2 - Python Seaborn Line Plot and Create Data Visualizations.](#)
- [Guided LAB - 344.4.3 - Python Seaborn Regression Plot.](#)

Seaborn vs. Matplotlib

- Seaborn is built on top of Matplotlib and provides a higher-level interface for creating statistical graphics. While Matplotlib is a general-purpose plotting library, Seaborn is specifically designed for statistical data visualization.
- Seaborn offers several advantages over Matplotlib, including simpler syntax for creating complex plots, built-in support for statistical visualizations, and aesthetically pleasing default settings that can be easily customized.
- Additionally, Seaborn offers several specialized plot types such as violin plots and swarm plots that are not available in Matplotlib.

Knowledge Check

1. What is Seaborn, and how does it differ from other data visualization libraries like Matplotlib?
2. Explain the main features and advantages of using Seaborn for data visualization?
3. What are the different plot types available in Seaborn, and when would you use each type?
4. What are regression plots in Seaborn, and how do they help in visualizing relationships between variables?

Summary

- Seaborn is a powerful Python data visualization library. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn offers a variety of functions for creating different types of plots, including scatter plots, line plots, bar plots, histograms, box plots, violin plots, and more. These plots are designed to handle both categorical and numerical data effectively.
- Seaborn allows for extensive customization of plots including colors, styles, markers, and labels. Additionally, it provides built-in themes and color palettes to make plots visually appealing and consistent.
- Seaborn excels at creating statistical visualizations that helps in exploring relationships between variables, identifying patterns, and understanding distributions within datasets. It offers functions for visualizing univariate and bivariate distributions, regression relationships, and matrix plots.
- Seaborn seamlessly integrates with Pandas DataFrames

References

<https://www.analyticsvidhya.com/blog/2020/06/guide-geospatial-analysis-folium-python/>
<https://medium.com/@sawsanyusuf/data-visualization-with-python-10-choropleth-maps-df7ab3118c3a>

Questions?



PER SCHOLAS

UNLOCKING POTENTIAL

CHANGING THE FACE OF TECH

