AFRICA DATA SCHOOL

Seaborn: Statistical data visualization

Seaborn is a Python data visualization library based on Matplotlib . It provides a high-level interface for drawing attractive and informative statistical graphics.

Seaborn vs Matplotlib

Matplotlib “tries to make easy things easy and hard things possible”, Seaborn tries to make a well-defined set of hard things easy too.”

Seaborn vs Matplotlib

Characteristics	Matplotlib	Seaborn
Use Cases	Matplotlib plots various graphs using Pandas and Numpy	Seaborn is the extended version of Matplotlib which uses Matplotlib along with Numpy and Pandas for plotting graphs
Complexity of Syntax	It uses comparatively complex and lengthy syntax.	It uses comparatively simple syntax which is easier to learn and understand.
Multiple figures	Matplotlib has multiple figures can be opened	Seaborn automates the creation of multiple figures which sometimes leads to out of memory issues
Flexibility	Matplotlib is highly customizable and powerful.	Seaborn avoids a ton of boilerplate by providing default themes which are commonly used.

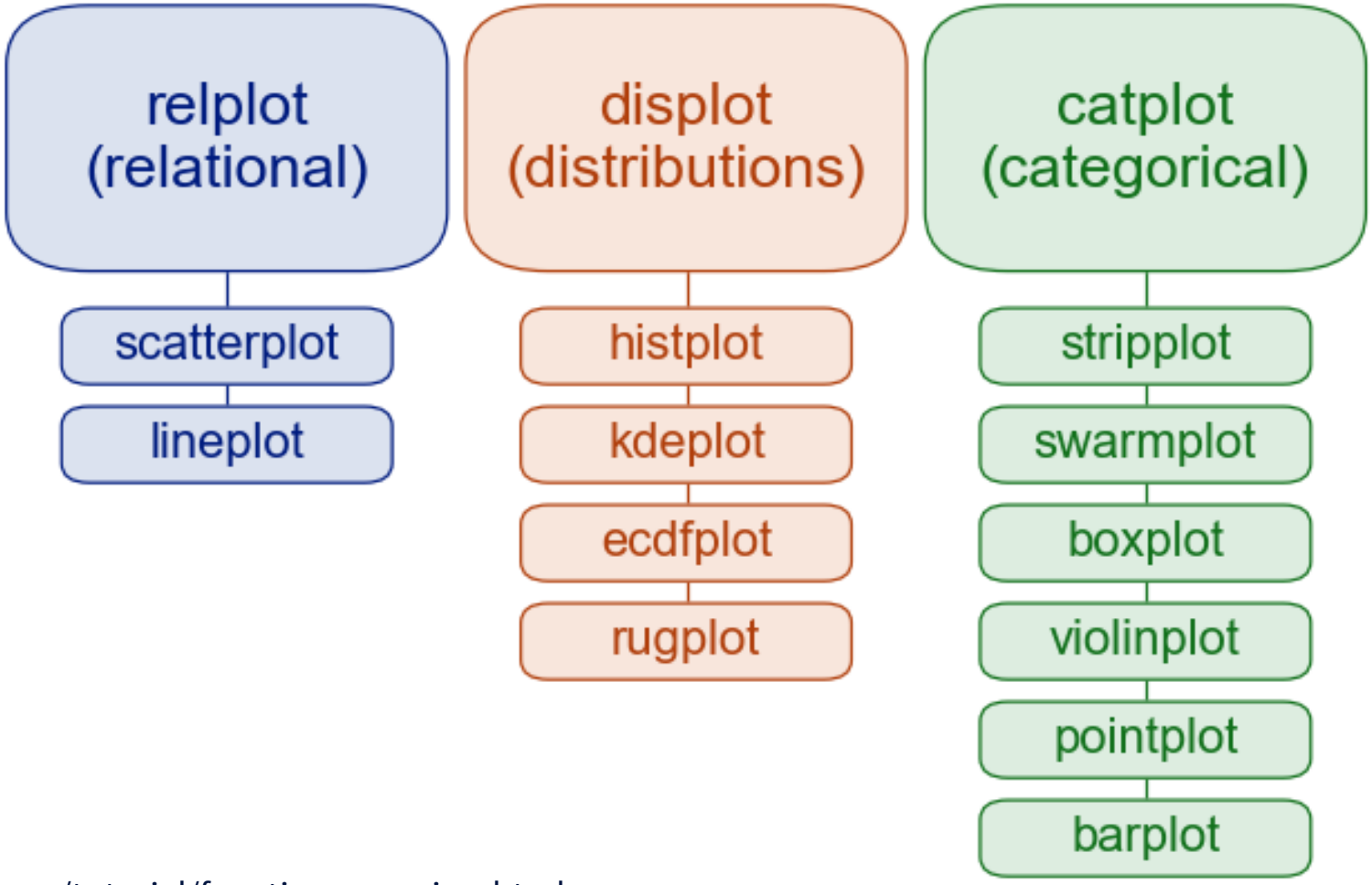
Important Features of Seaborn

- Built in themes for styling matplotlib graphics
- Visualizing univariate and bivariate data
- Fitting in and visualizing linear regression models
- Plotting statistical time series data
- Seaborn works well with NumPy and Pandas data structures
- It comes with built in themes for styling Matplotlib graphics

Installing Seaborn

```
pip install seaborn
conda install seaborn
```

Figure-level vs. axes-level functions



https://seaborn.pydata.org/tutorial/function_overview.html

Types of Plot

Relational plots¶ - Understanding how the variables in the dataset relate each other and their relationships

Distribution plots - Suitable for comparing range and **distribution** for groups of numerical data

Categorical plots - used for visualizing the relationship between variables. Those variables can be either be completely numerical or a category like a group, class or division

Regression plots - Creates a regression line between 2 parameters and helps to visualize their linear relationships

Matrix plots - A **matrix plot** is an array of scatterplots. effective when you are only interested in relationships among certain pairings of variables.