

Algorithms and Data Structures

Python Packages for Data Science



Agenda

- What are Python packages
- Packages specifically for Data Science
 - Numpy
 - Pandas
 - Matplotlib
 - Seaborn

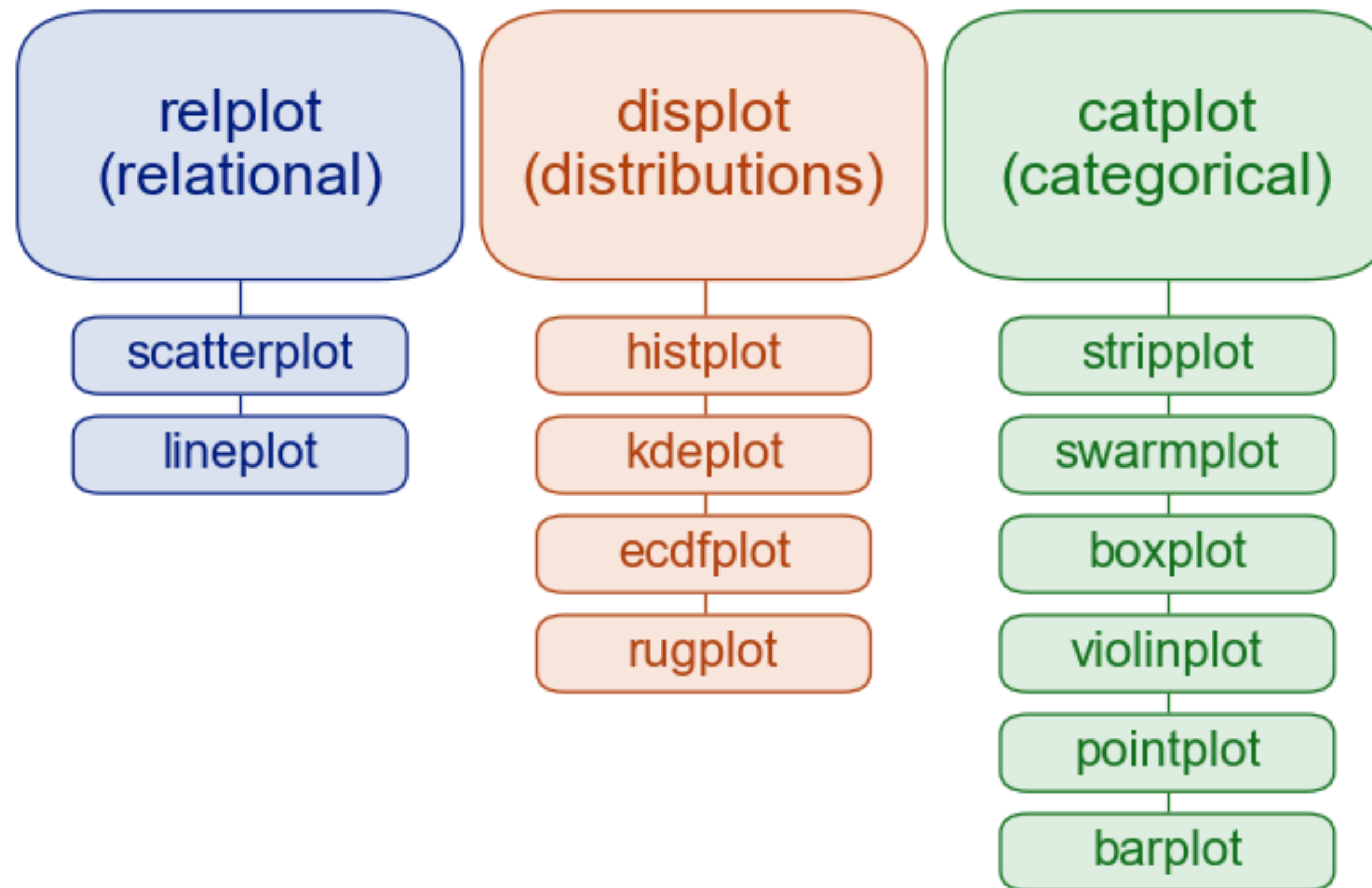


Seaborn

- Seaborn is based on Matplotlib
- It is also focusing in Data Visualization
- It provides high-level interface for drawing attractive / informative graphs.
- Import the library
 - `import matplotlib.pyplot as plt`
 - `import seaborn as sns`



Seaborn Overview



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

Relational Plots

- Relational plots show how two or more attributes are related.
- `scatterplot()`
- `lineplot()`
- `relplot()`
 - `kind = 'scatter'`, default, same as `scatterplot()`
 - `kind = 'line'`, same as `lineplot()`
 - Details: <https://seaborn.pydata.org/generated/seaborn.relplot.html>



Distributions Plots

- Distributions plots show how the data values are distributed.
 - `displot()`
 - `histplot()`
 - `kdeplot()`
 - `ecdfplot()`
 - `rugplot()`



Categorical Plots

- `catplot()`
- `boxplot()`
- `violinplot()`
- `barplot()`



Multiple Plots

- `jointplot()`
 - Kind can be ['scatter', 'hist', 'hex', 'kde', 'reg', 'resid']
- `pairplot()`
 - Create pair-wise plot for numeric attributes

API Details and Examples

- More details for all the plots:
 - <https://seaborn.pydata.org/api.html#distribution-api>
- More examples for all kinds of plots:
 - <https://seaborn.pydata.org/examples/index.html>



Thank you!