Learn By Example: Seaborn

Overview

Visualizing relationships

- Univariate and bivariate relationships
- Histograms, KDE curves, scatter plots
- Facet grid and pair grid

Plot aesthetics and style

- Themes, color palettes

Software and Skills

Be very comfortable programming in Python (Python 3)

Be comfortable working with Jupyter notebooks

Understand high school matrix operations

Demo

Installing Seaborn

Exploring Pokemon Dataset

Matplotlib

Tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code.

matplotlib.org

Seaborn

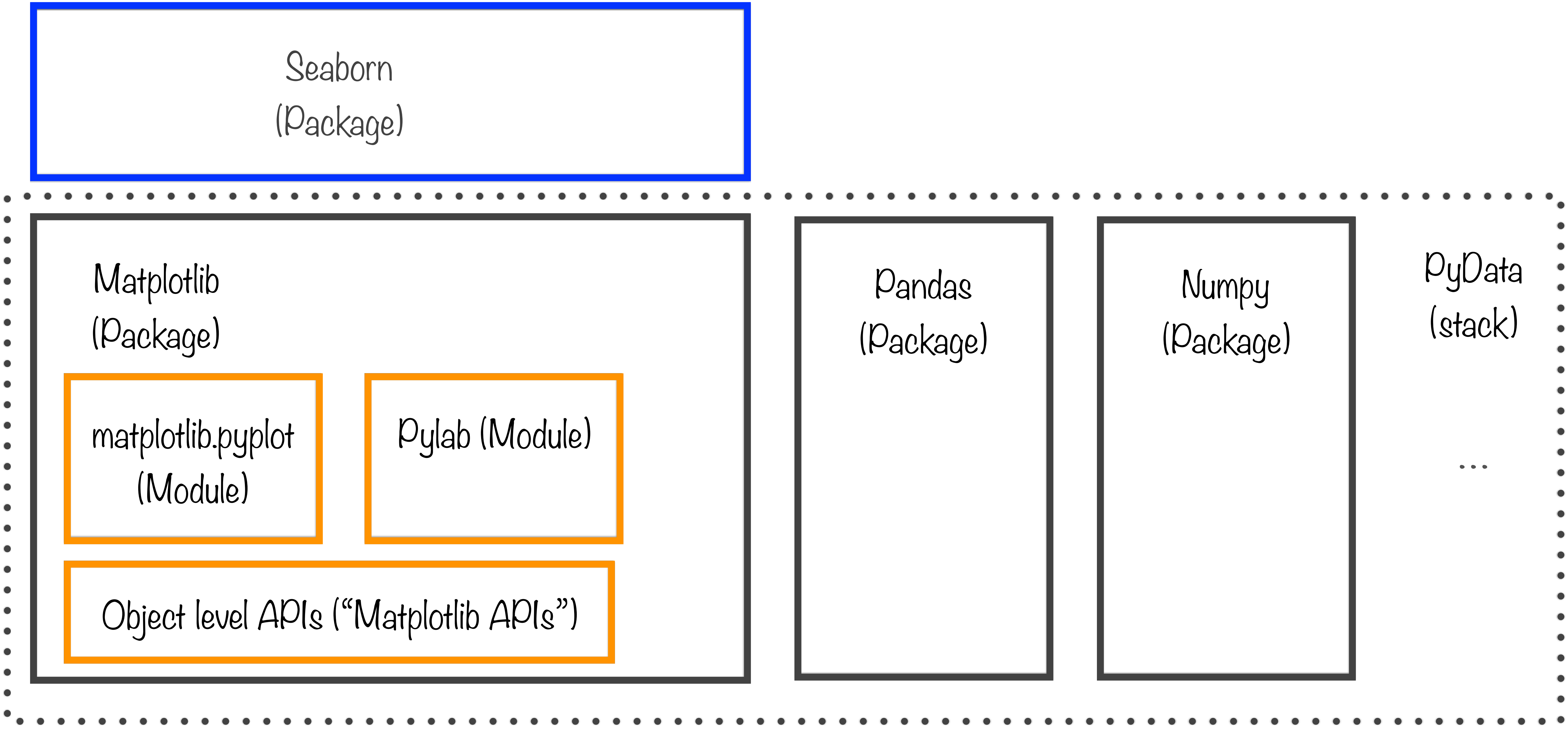
Built on top of matplotlib and tightly integrated with the PyData stack, including support for numpy and pandas data structures and statistical routines from scipy and statsmodels.

seaborn.pydata.org

Seaborn For "Production Plots"

Matplotlib	Seaborn
Part of "Pydata" - open data science stack	Built atop Matplotlib and tightly integrates with Pydata
Provides fine-grained control so that pretty much everything is possible	High level, easy-to-use abstractions for common use cases
Two APIs - Matplotlib API (low-level) and Pyplot (higher level)	Even higher level than Pyplot (used alongside it)
Production-level aesthetics possible, but need use of Matplotlib API	Production-level aesthetics without need for low-level API

Matplotlib and Seaborn



Seaborn
(Package)

Matplotlib
(Package)

matplotlib.pyplot
(Module)

Pylab (Module)

Object level APIs (“Matplotlib APIs”)

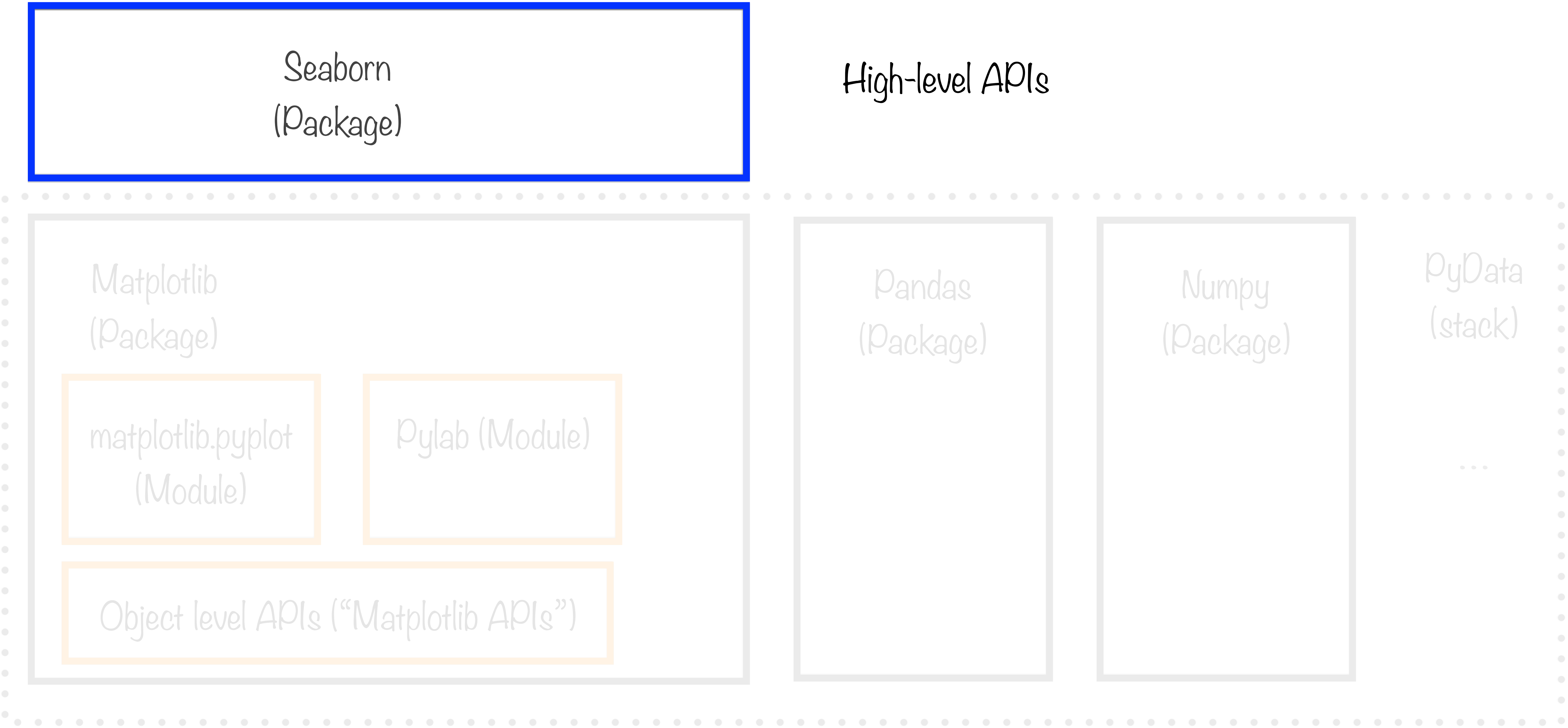
Pandas
(Package)

Numpy
(Package)

PyData
(stack)

…

Matplotlib and Seaborn



Seaborn
(Package)

High-level APIs

Matplotlib
(Package)

matplotlib.pyplot
(Module)

Pylab (Module)

Object level APIs ("Matplotlib APIs")

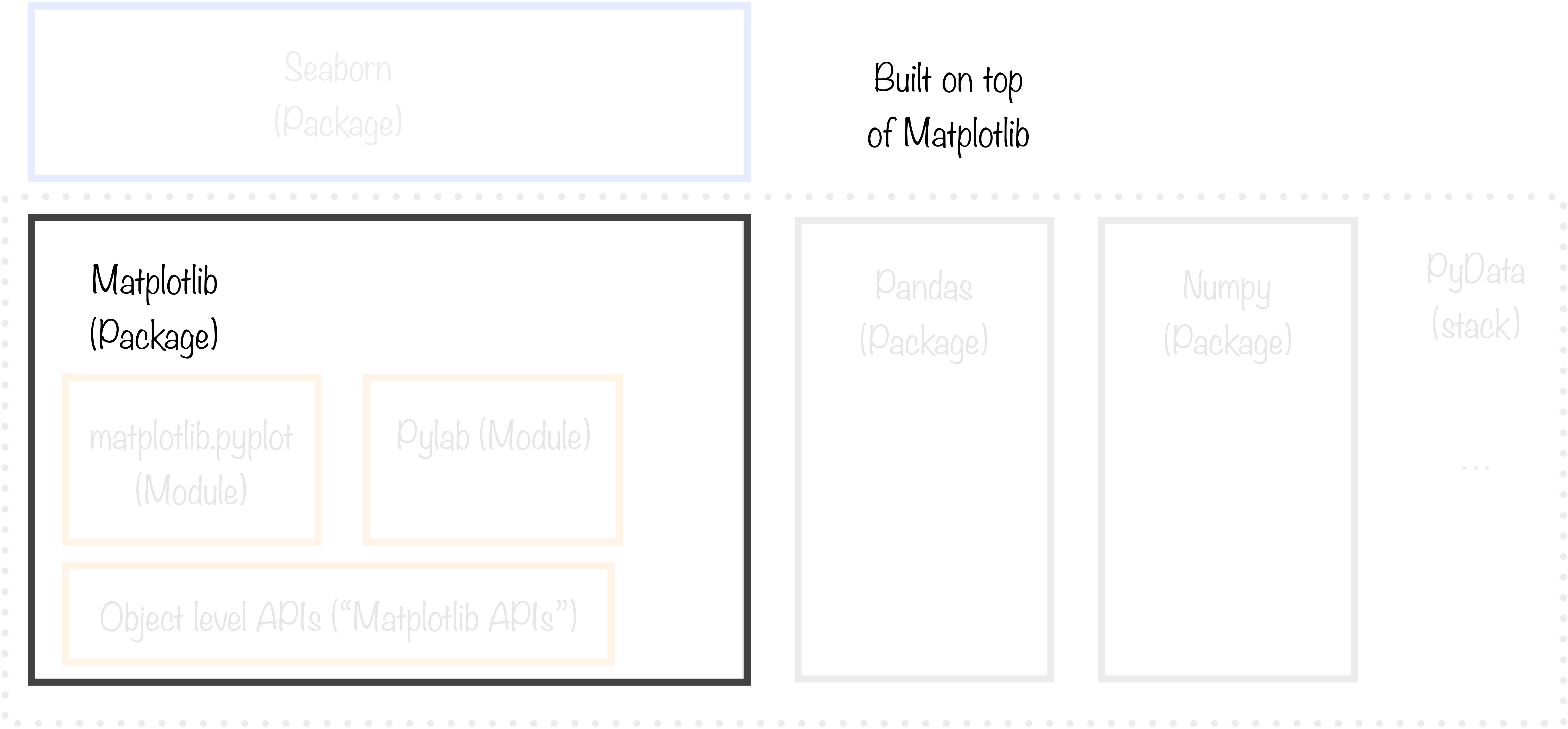
Pandas
(Package)

Numpy
(Package)

PyData
(stack)

....

Matplotlib and Seaborn



Seaborn
(Package)

Built on top
of Matplotlib

Matplotlib
(Package)

matplotlib.pyplot
(Module)

Pylab (Module)

Object level APIs ("Matplotlib APIs")

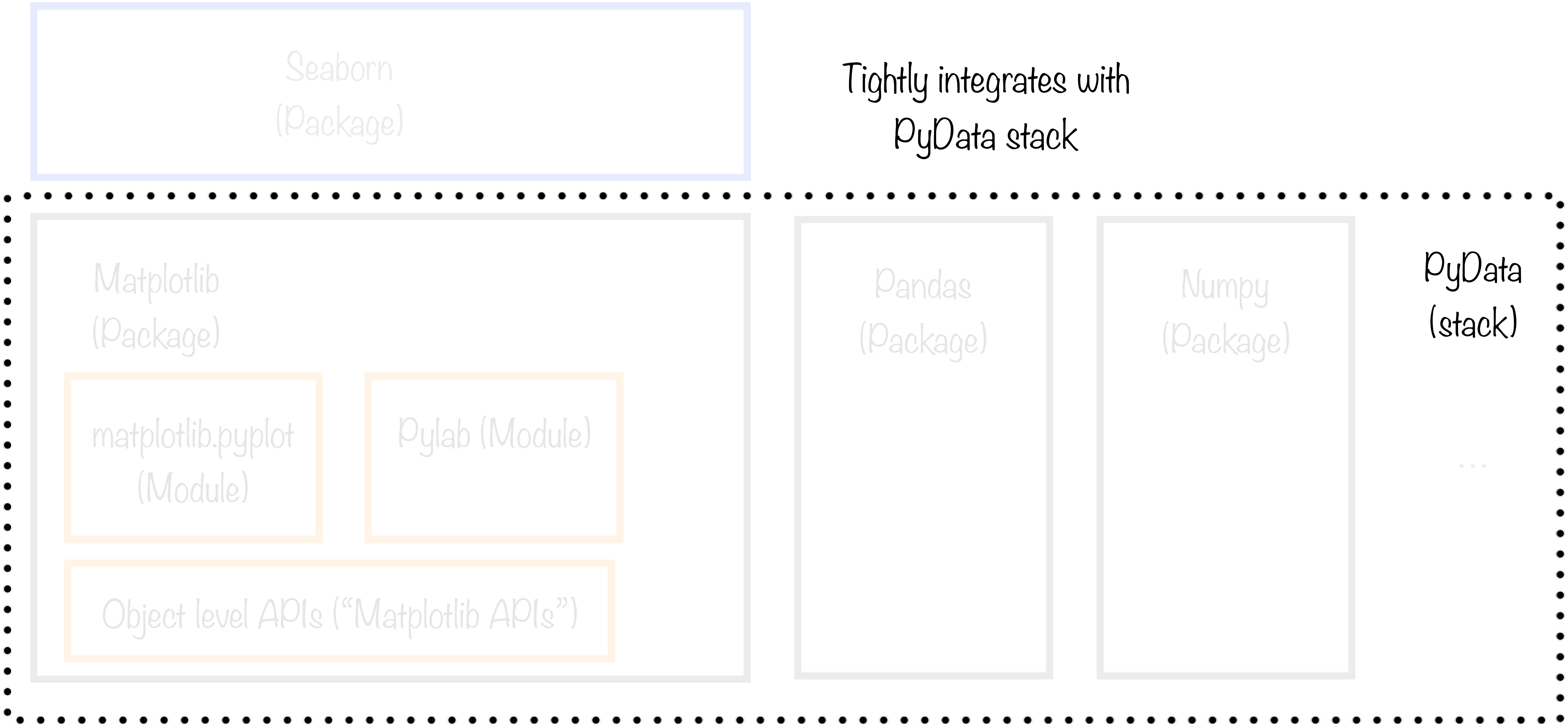
Pandas
(Package)

Numpy
(Package)

PyData
(stack)

...

Matplotlib and Seaborn



Seaborn
(Package)

Tightly integrates with
PyData stack

Matplotlib
(Package)

matplotlib.pyplot
(Module)

Pylab (Module)

Object level APIs (“Matplotlib APIs”)

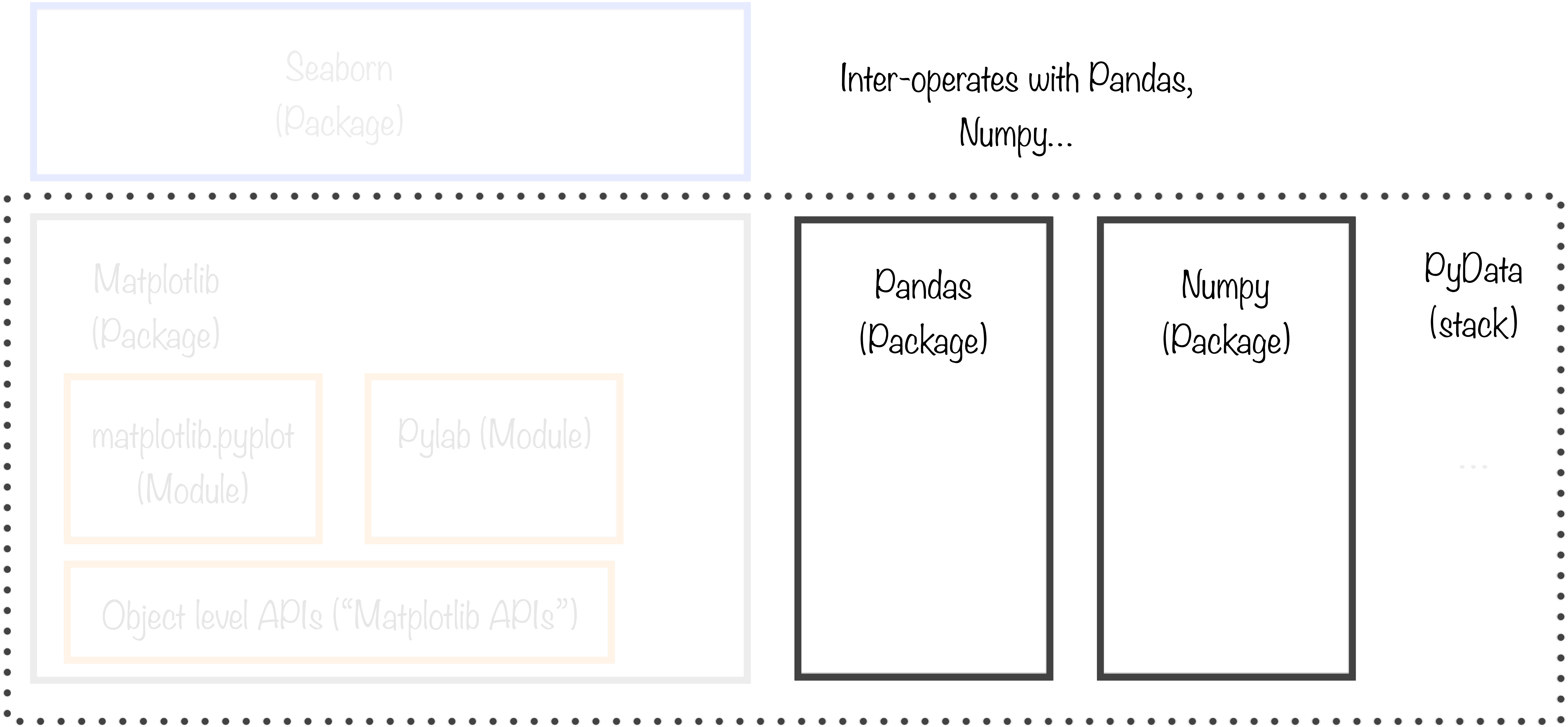
Pandas
(Package)

Numpy
(Package)

PyData
(stack)

...

Matplotlib and Seaborn



Seaborn
(Package)

Inter-operates with Pandas,
Numpy...

Matplotlib
(Package)

matplotlib.pyplot
(Module)

Pylab (Module)

Object level APIs ("Matplotlib APIs")

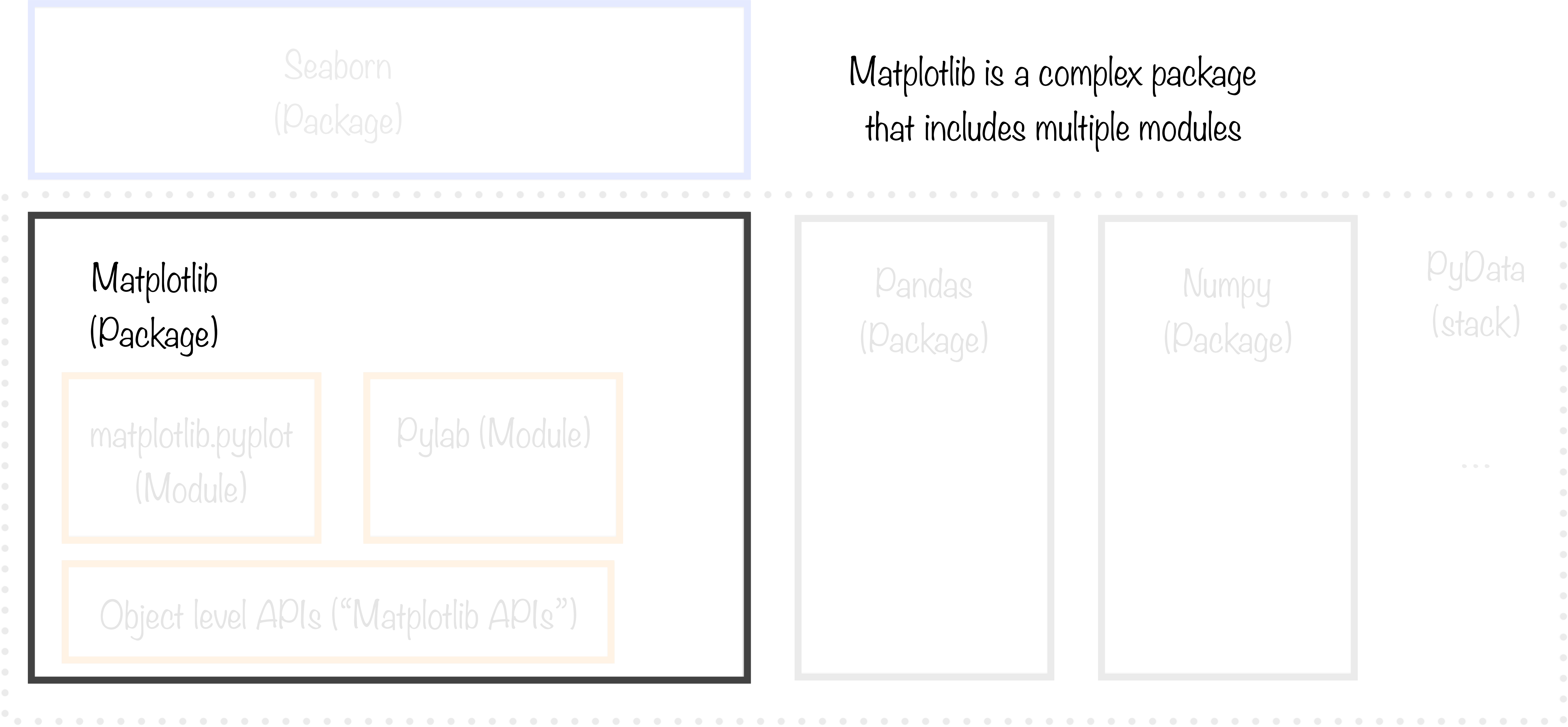
Pandas
(Package)

Numpy
(Package)

PyData
(stack)

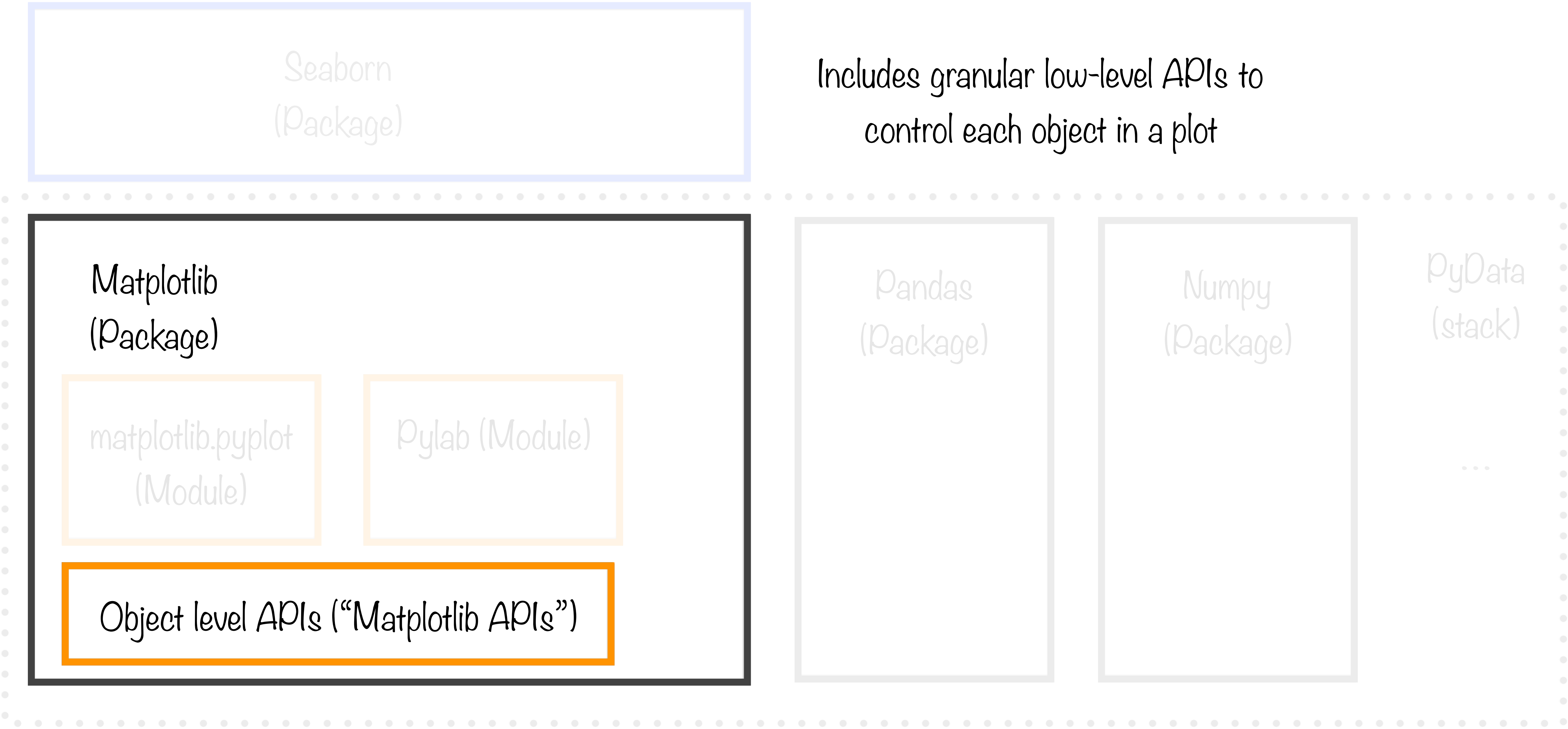
...

Matplotlib and Seaborn



Matplotlib is a complex package
that includes multiple modules

Matplotlib and Seaborn



Seaborn (Package)

Matplotlib (Package)

matplotlib.pyplot (Module)

Pylab (Module)

Object level APIs (“Matplotlib APIs”)

Pandas (Package)

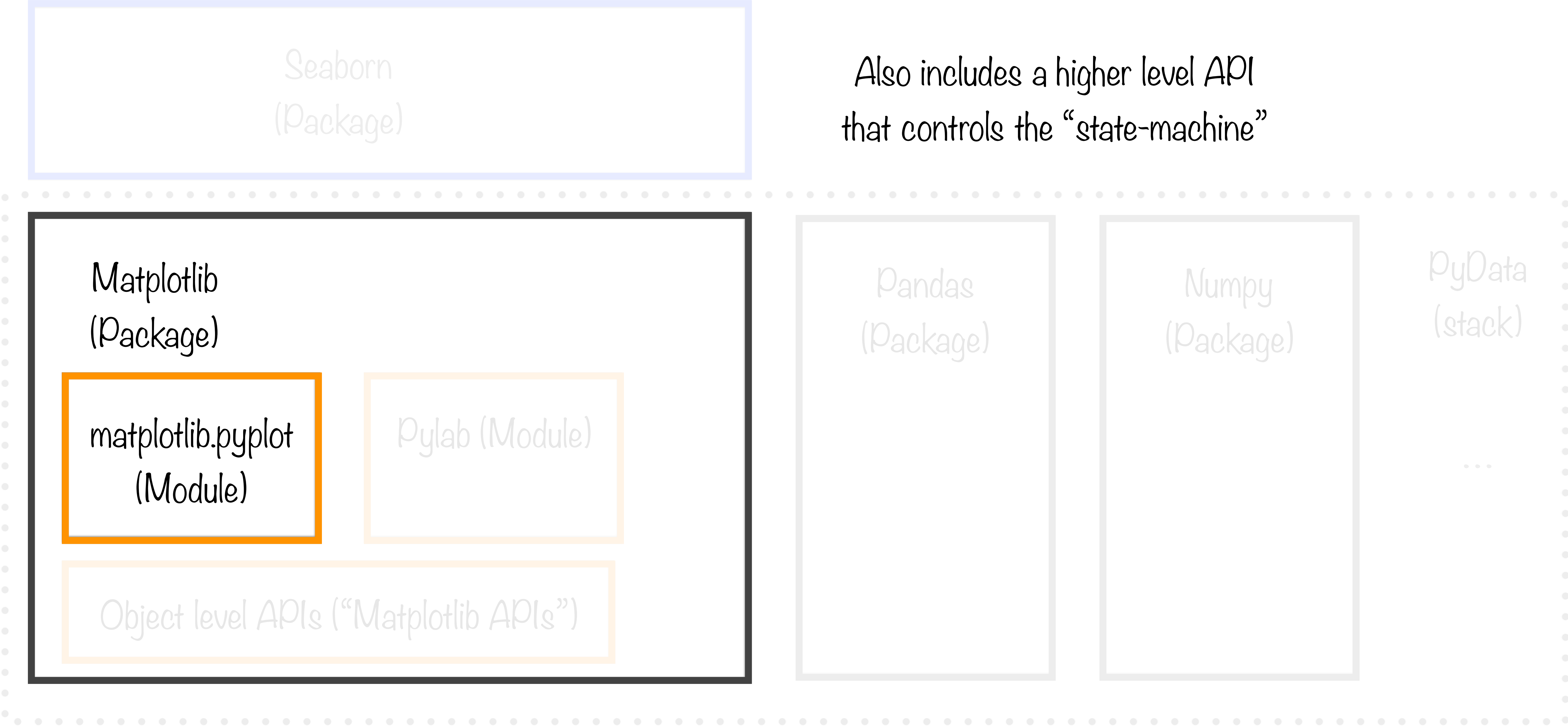
Numpy (Package)

PyData (stack)

. . .

Includes granular low-level APIs to control each object in a plot

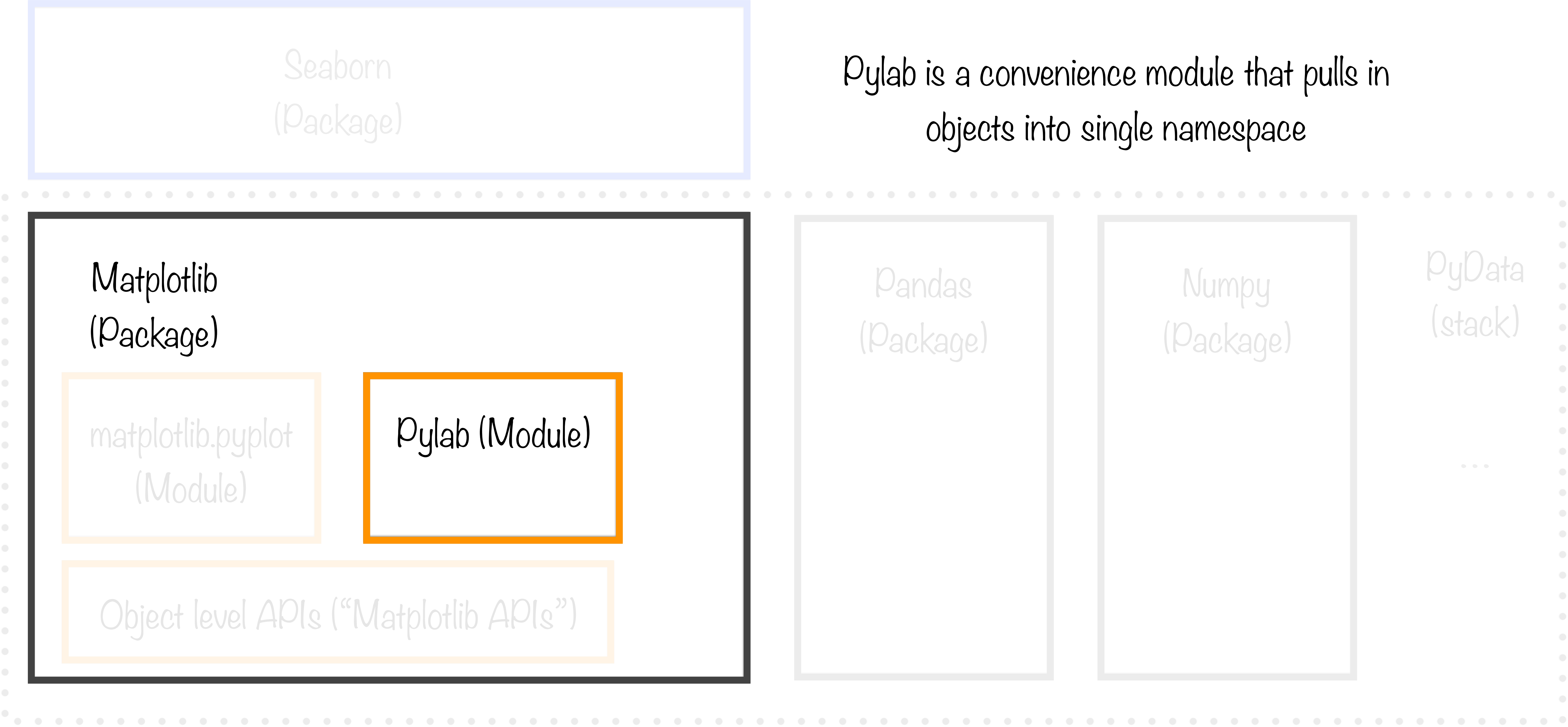
Matplotlib and Seaborn



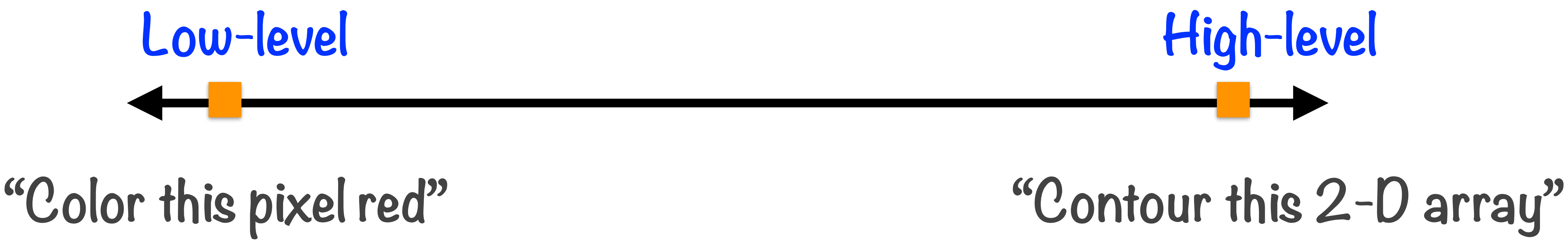
Also includes a higher level API
that controls the "state-machine"

Matplotlib and Seaborn

Pylab is a convenience module that pulls in objects into single namespace



Hierarchy of Plotting Operations



Low-level

High-level

"Color this pixel red"

"Contour this 2-D array"

Low-level operations act on specific plot elements, high-level operations act on plot as a whole

This hierarchy is formalized in the Matplotlib codebase

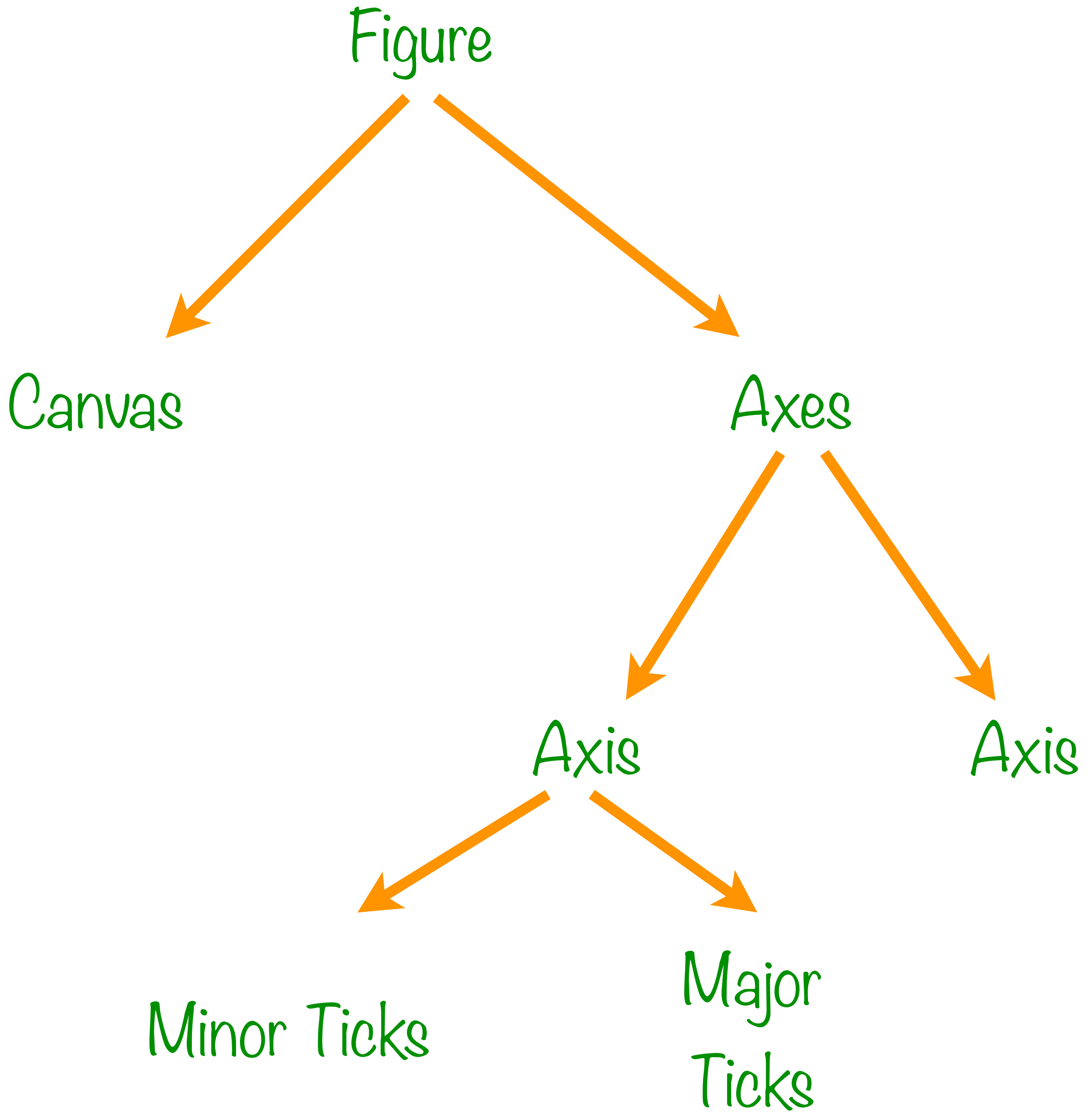
Hierarchy

Everything is an "Artist"

Artists are arranged in a hierarchy

Artist is an abstract base class

Figure is a container class

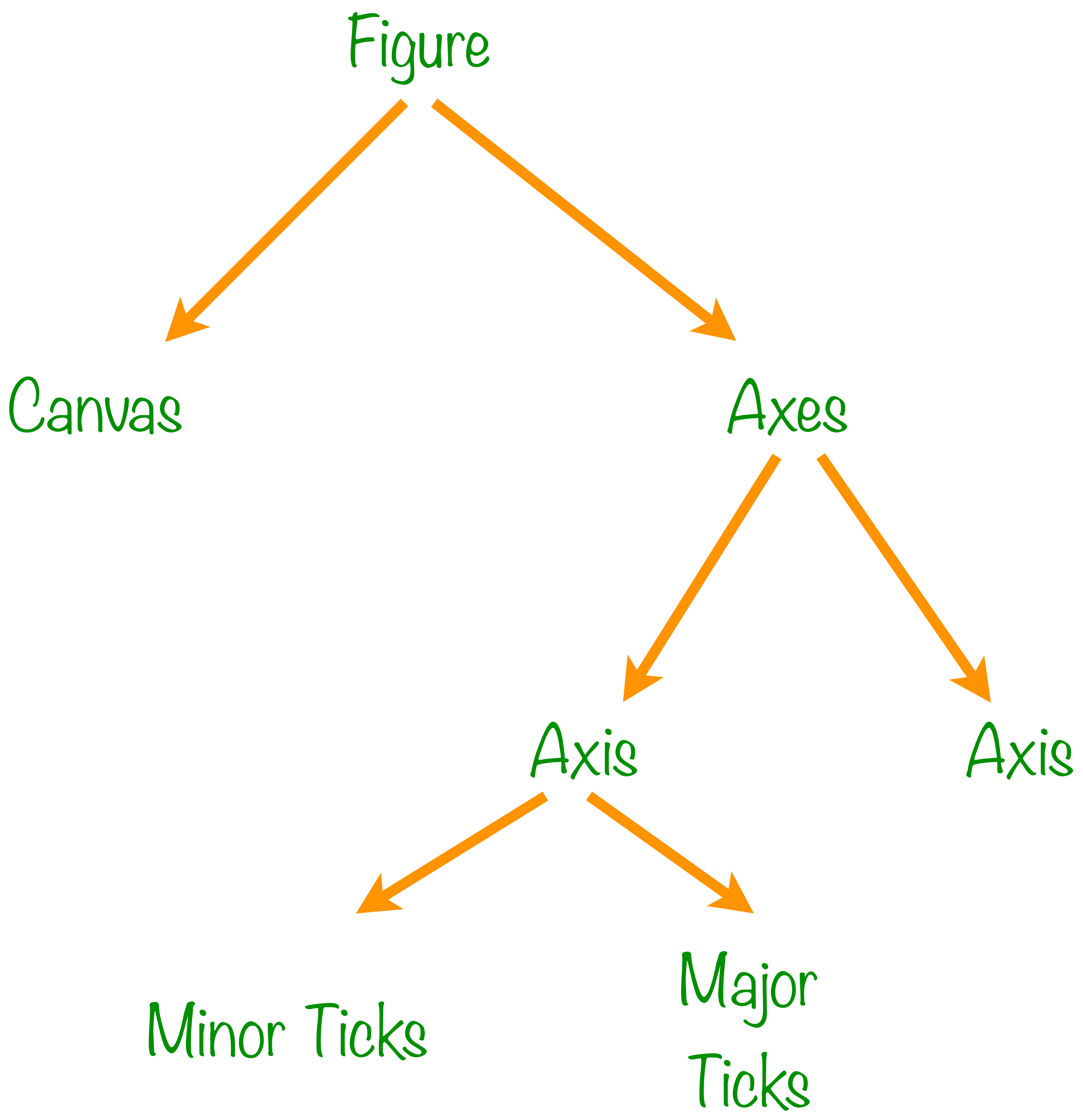


Hierarchy

Figure is a top-level container

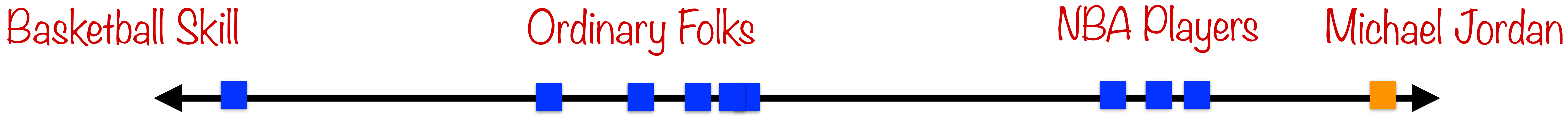
PyPlot APIs operate at higher levels

Matplotlib APIs at lower levels



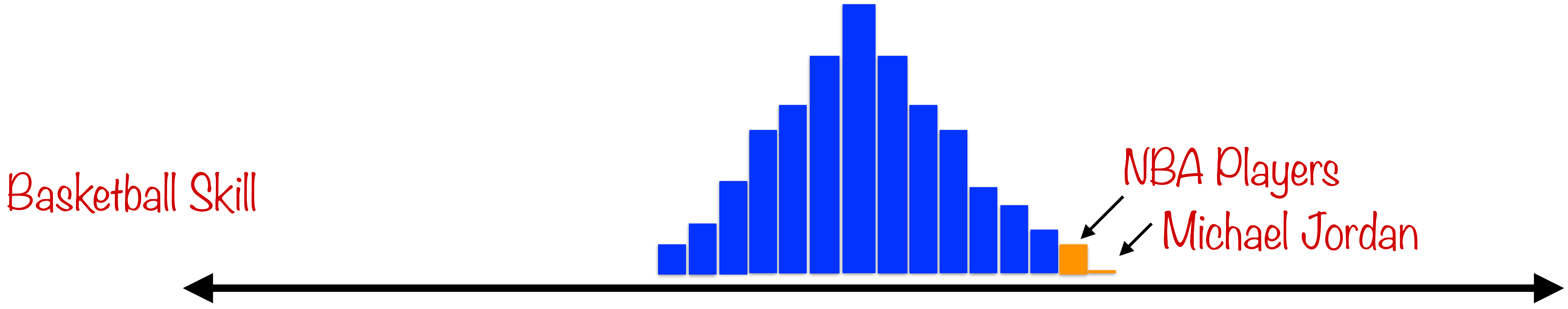
"Michael Jordan is a once-in-a- lifetime player"

Outliers



A once-in-a-lifetime player is an outlier, a point far from the pack

Outliers

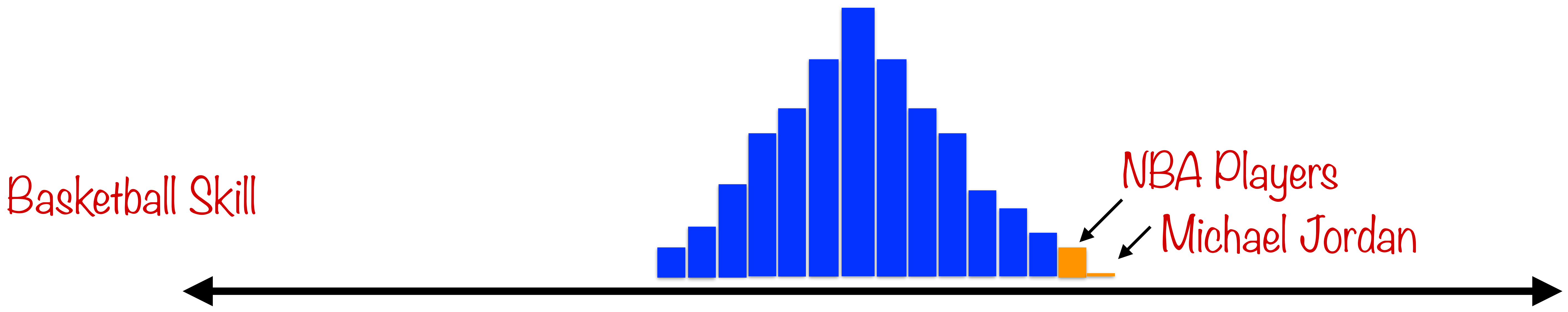


In reality, most ordinary folks would be clustered around an average level of skill

The NBA players would be outliers

Michael Jordan would be an even greater outlier

Outliers

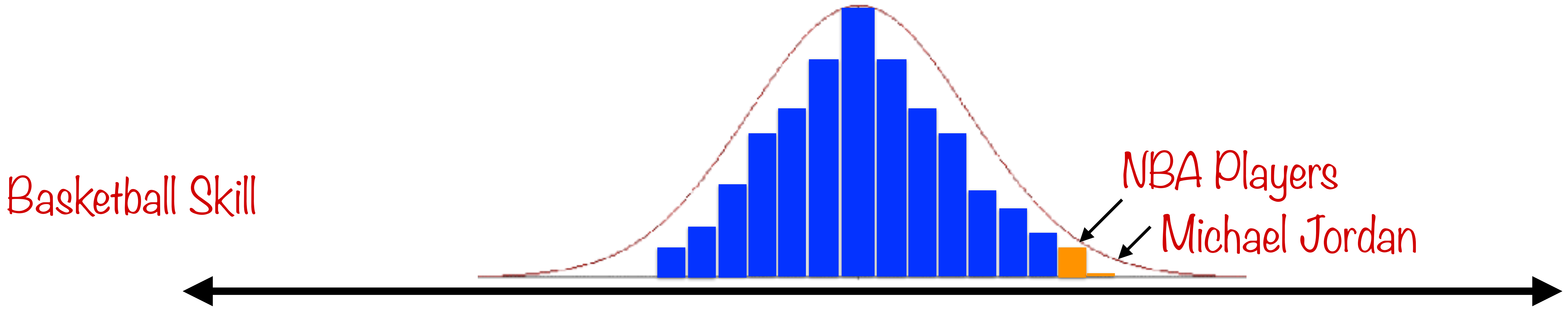


This chart above tells us how common a specific level of skill is

The shape of this chart resembles a bell

This is a Normal Probability Distribution

Outliers

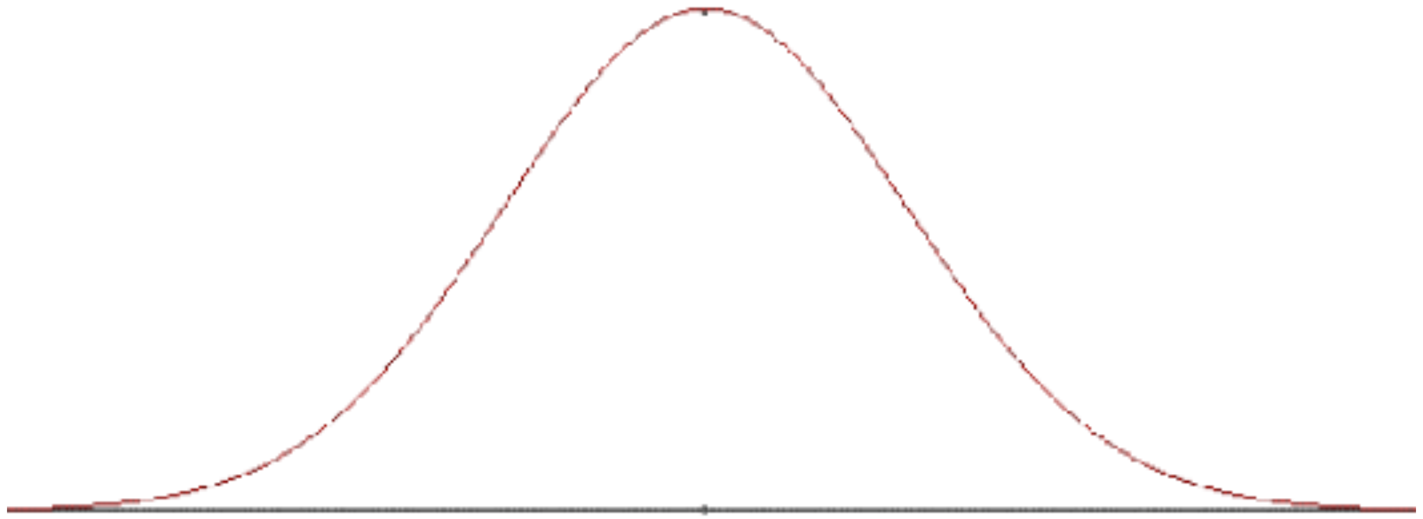


This chart above tells us how common a specific level of skill is

The shape of this chart resembles a bell

This is a Normal Probability Distribution

Outliers

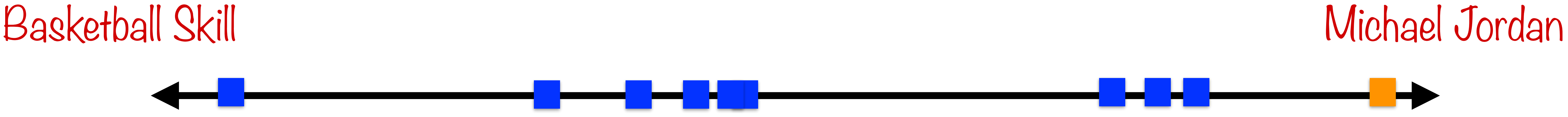


Average is common

Very high and very low are both unusual

The bell curve occurs everywhere in nature

Outliers



What is the probability of any specific value x occurring in the data?

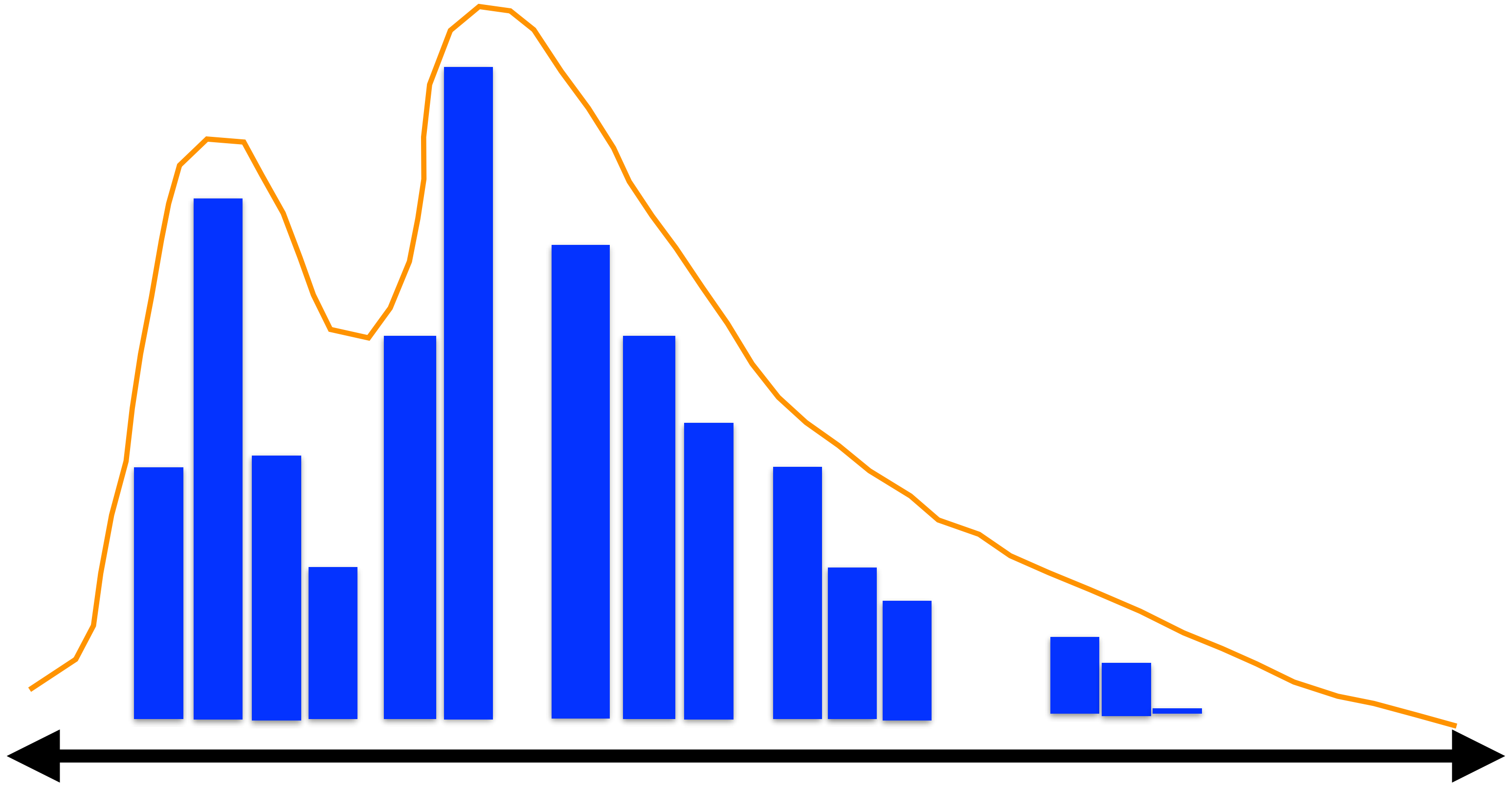
The answer lies in a **probability distribution function**

Kernel Density Estimation

Given a set of points

Figure out their probability distribution

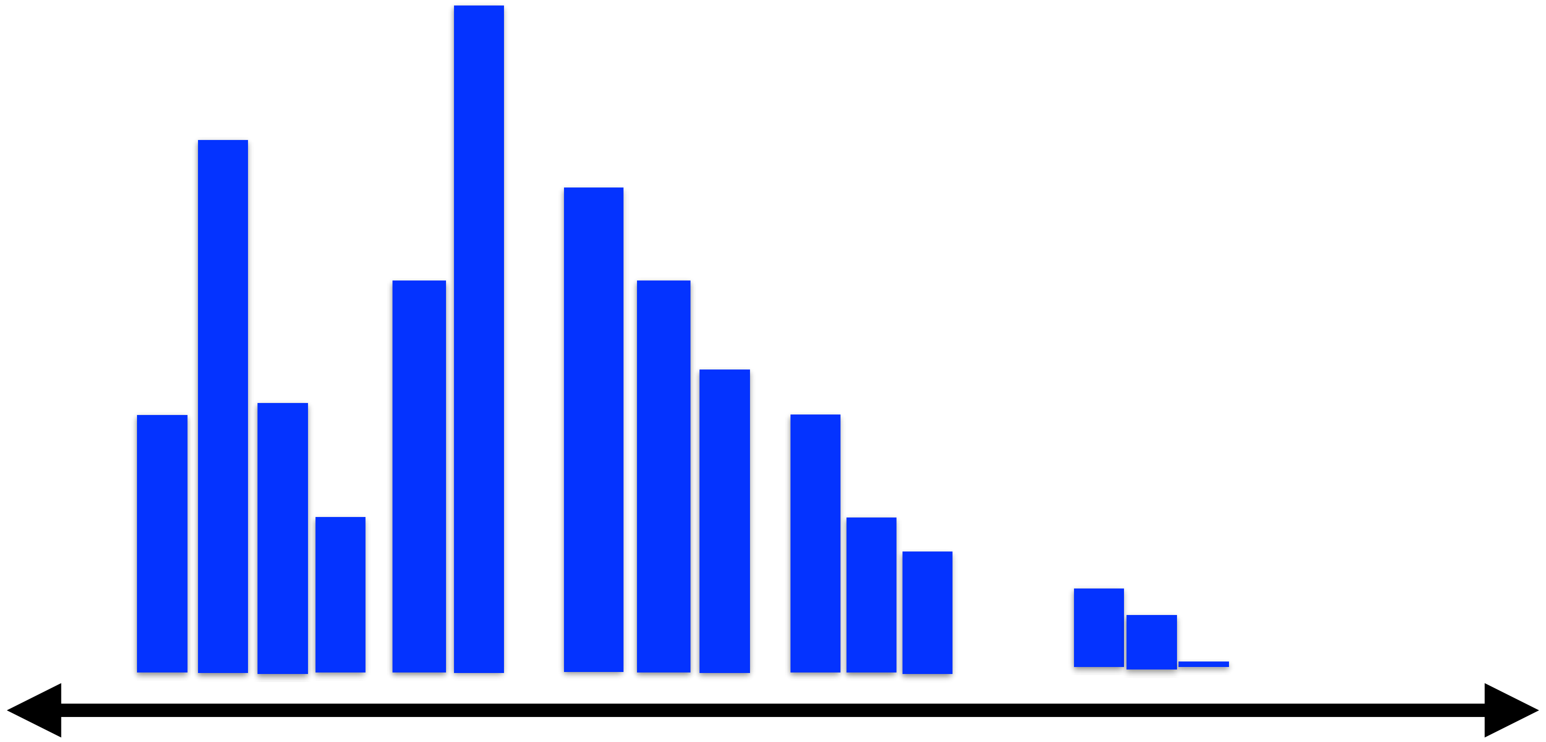
Area under curve must sum to 1



Kernel Density Estimation

KDE is a standard technique

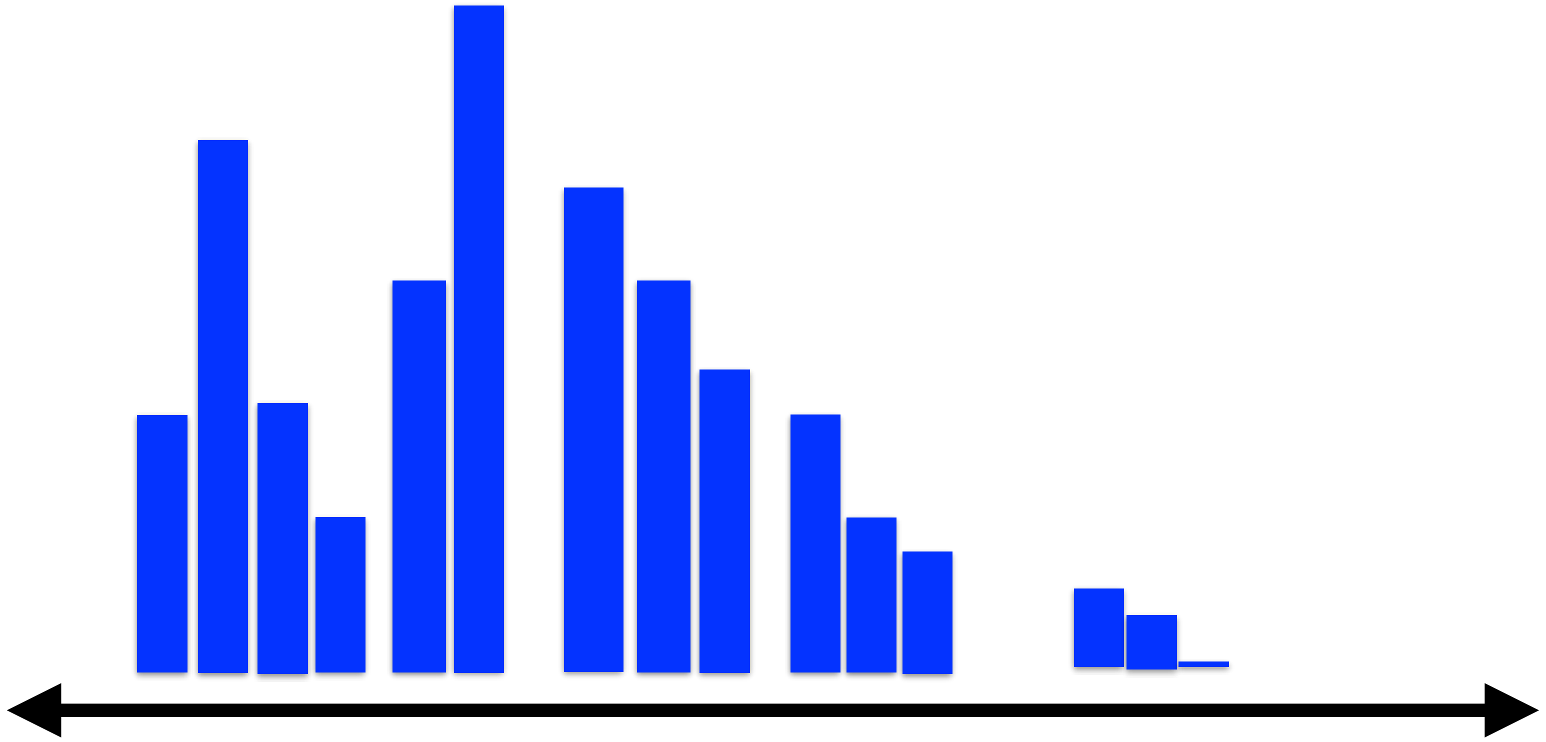
Non-parametric "Smoothing" technique



Kernel Density Estimation

Assume points have same distribution

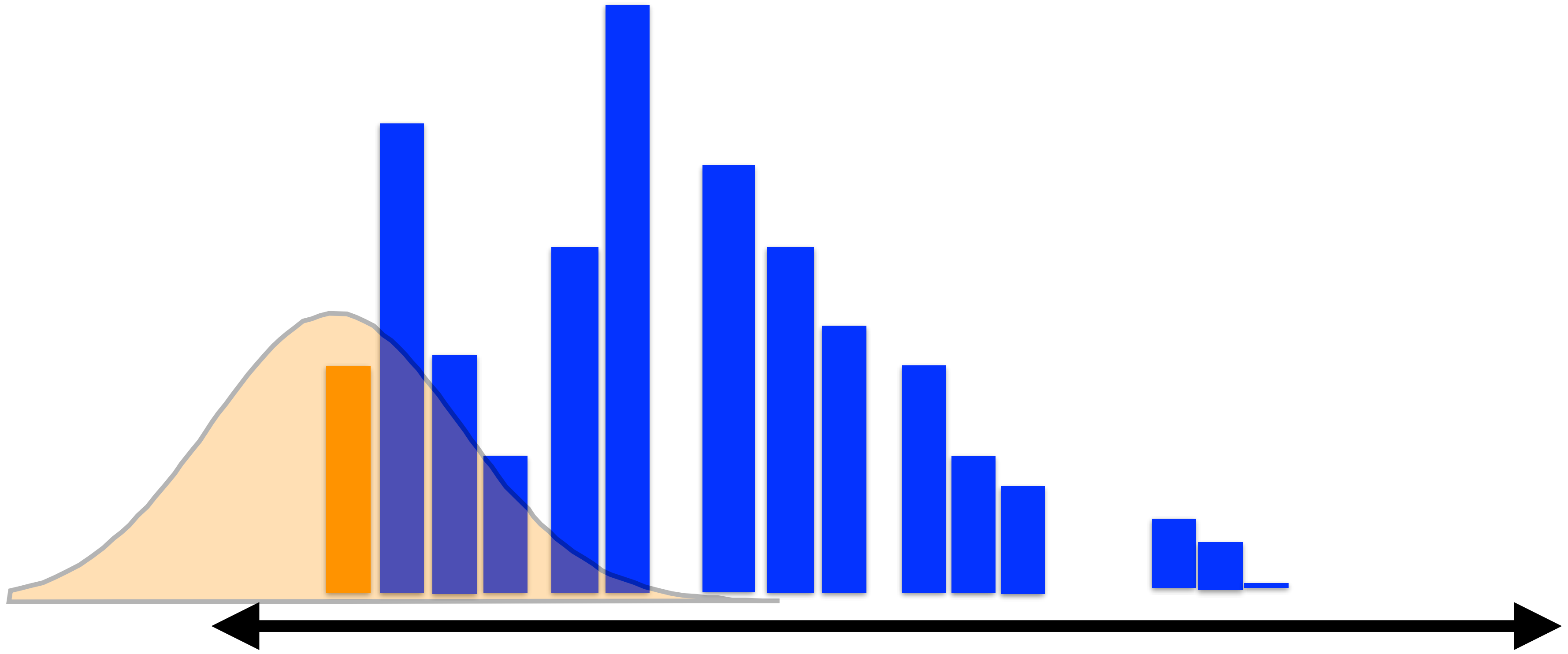
“Independent Identically Distributed”



Kernel Density Estimation

Assume points have same distribution

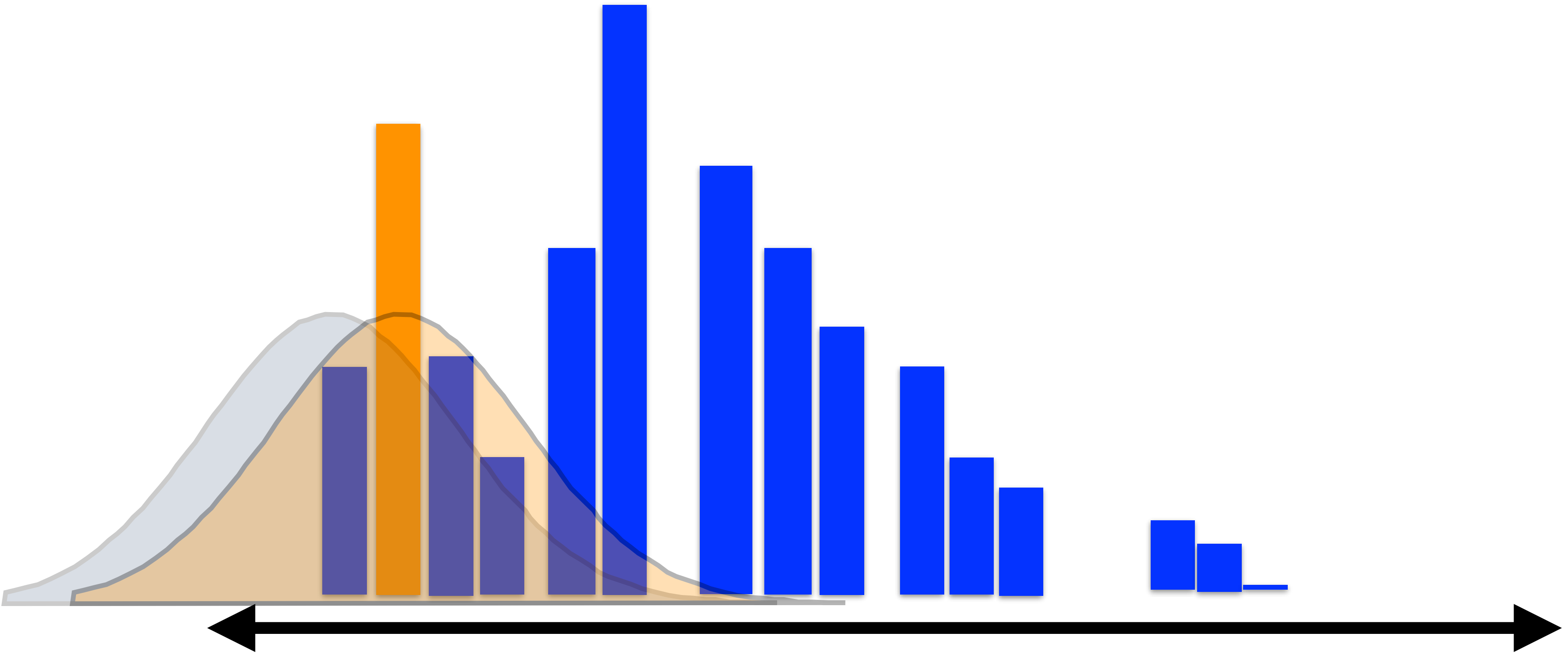
“Independent Identically Distributed”



Kernel Density Estimation

Assume points have same distribution

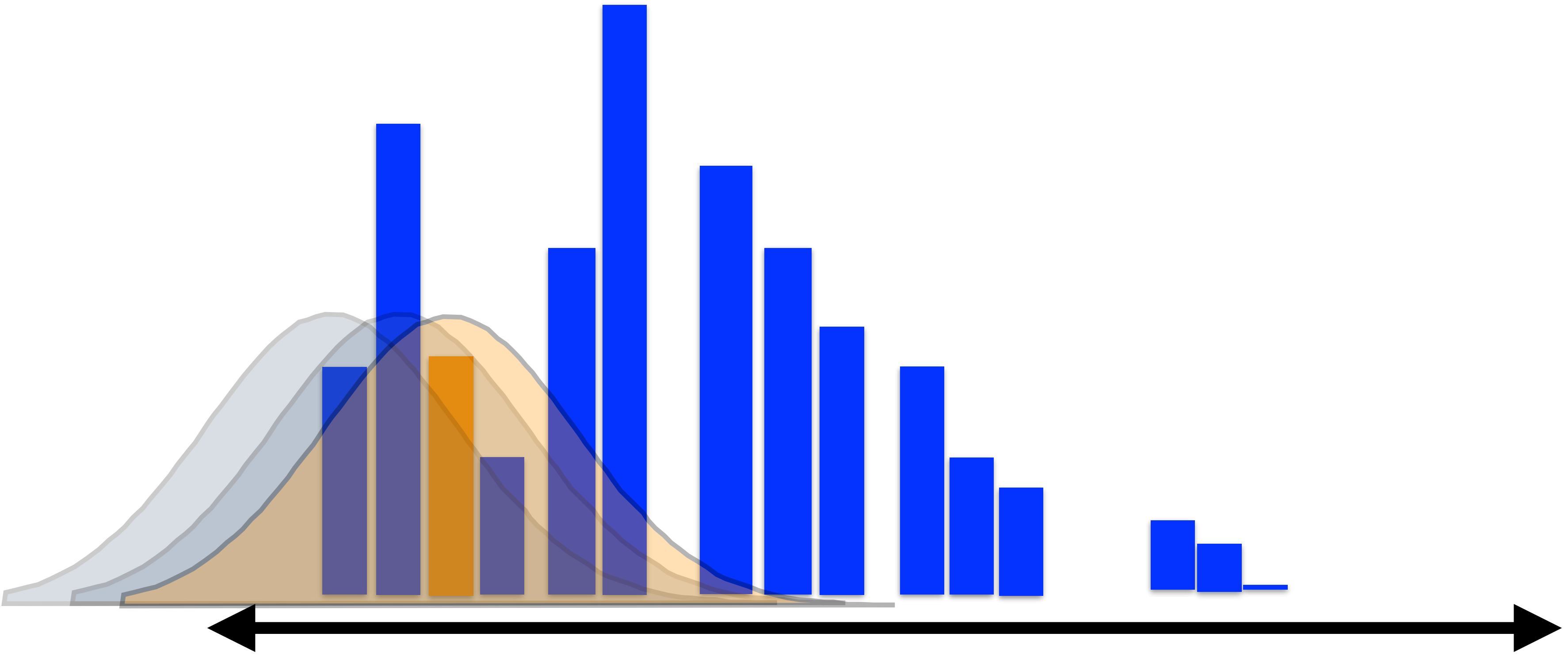
“Independent Identically Distributed”



Kernel Density Estimation

Assume points have same distribution

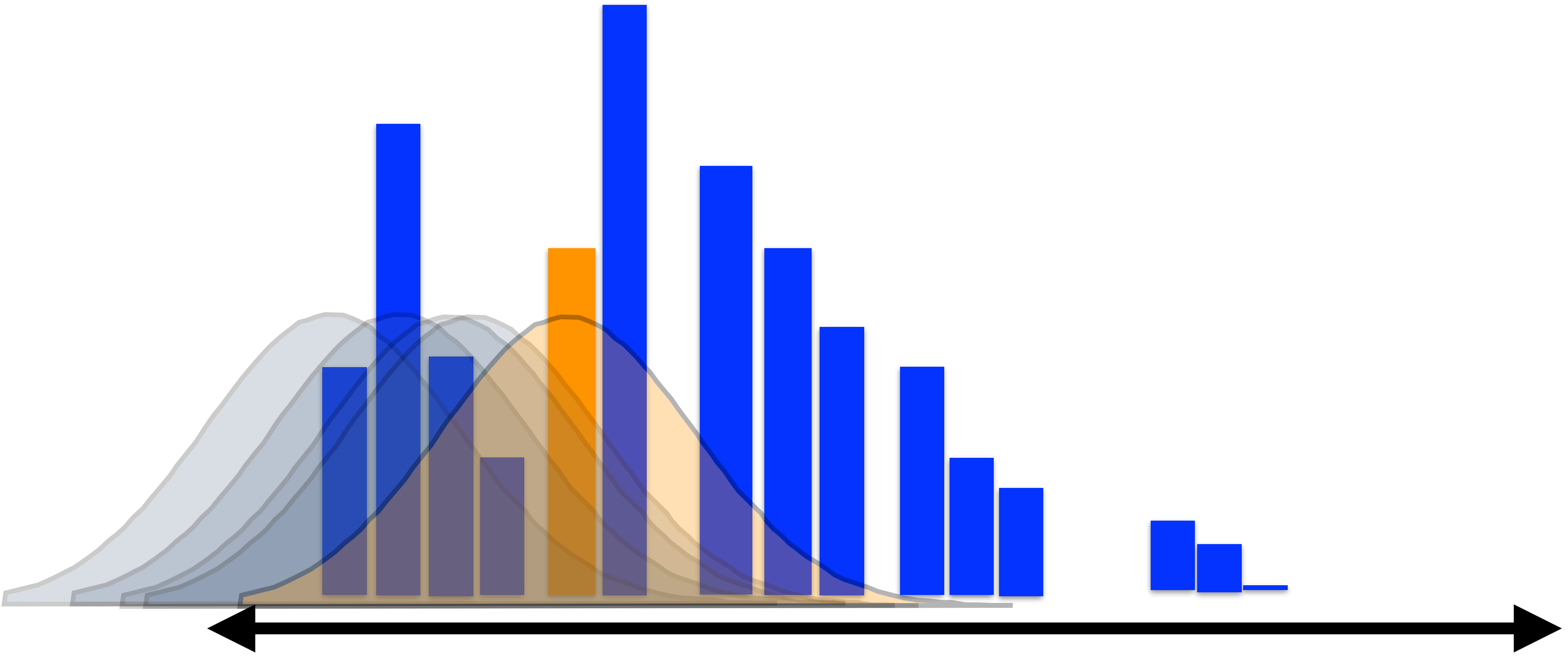
“Independent Identically Distributed”



Kernel Density Estimation

Assume points have same distribution

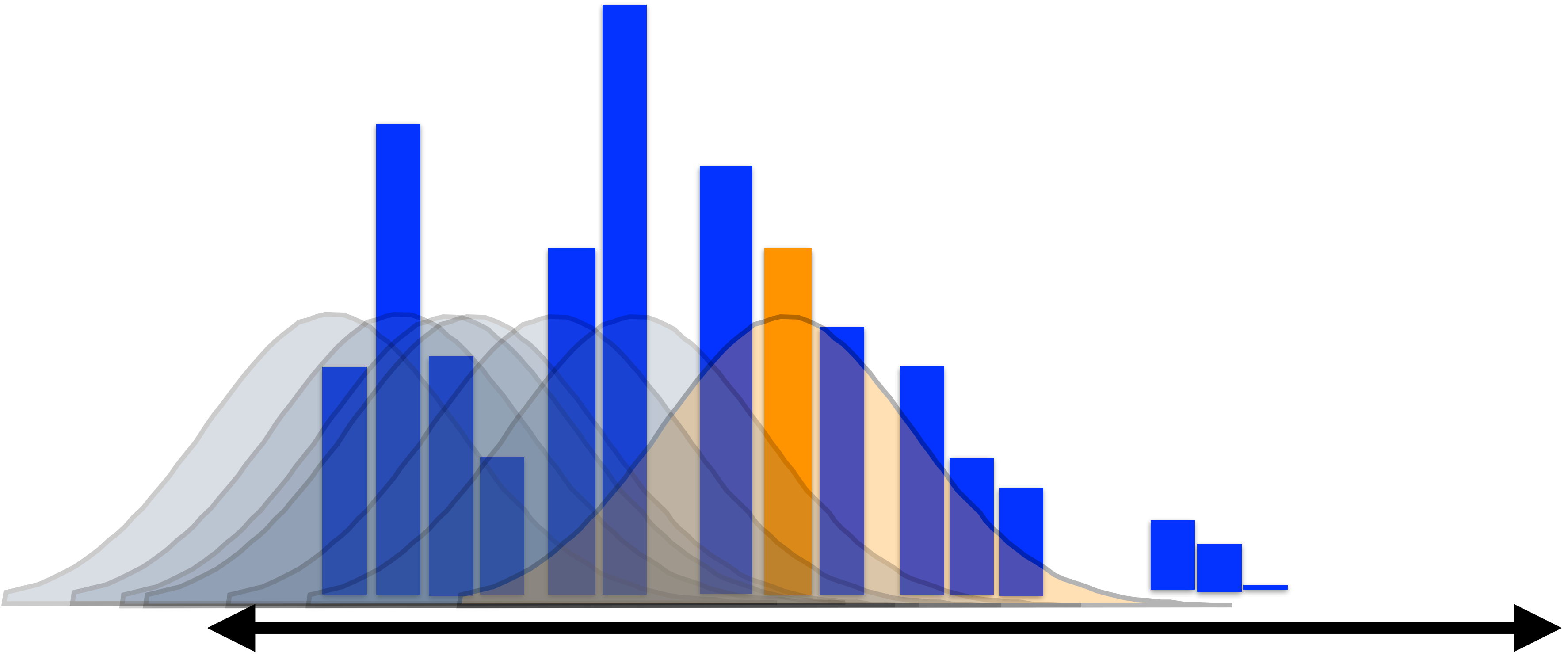
“Independent Identically Distributed”



Kernel Density Estimation

Assume points have same distribution

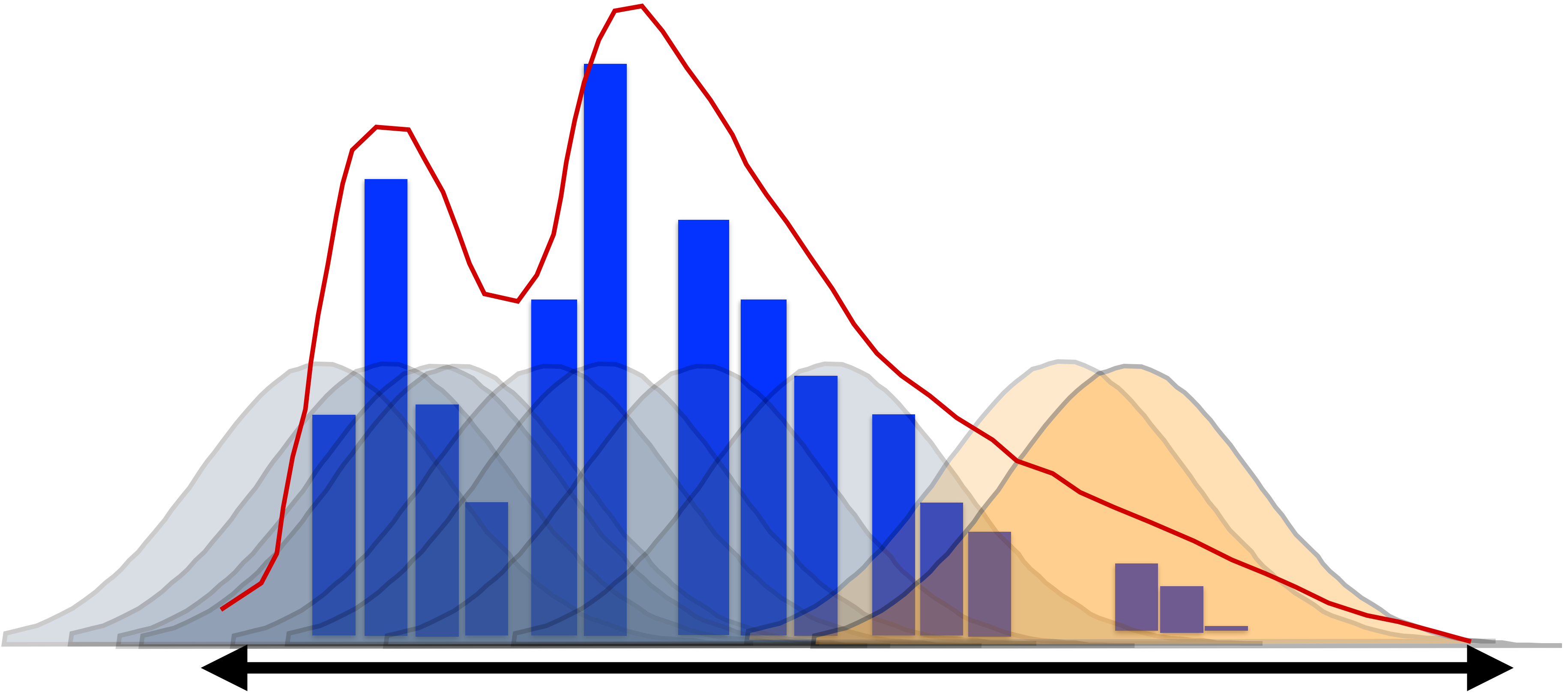
“Independent Identically Distributed”



Kernel Density Estimation

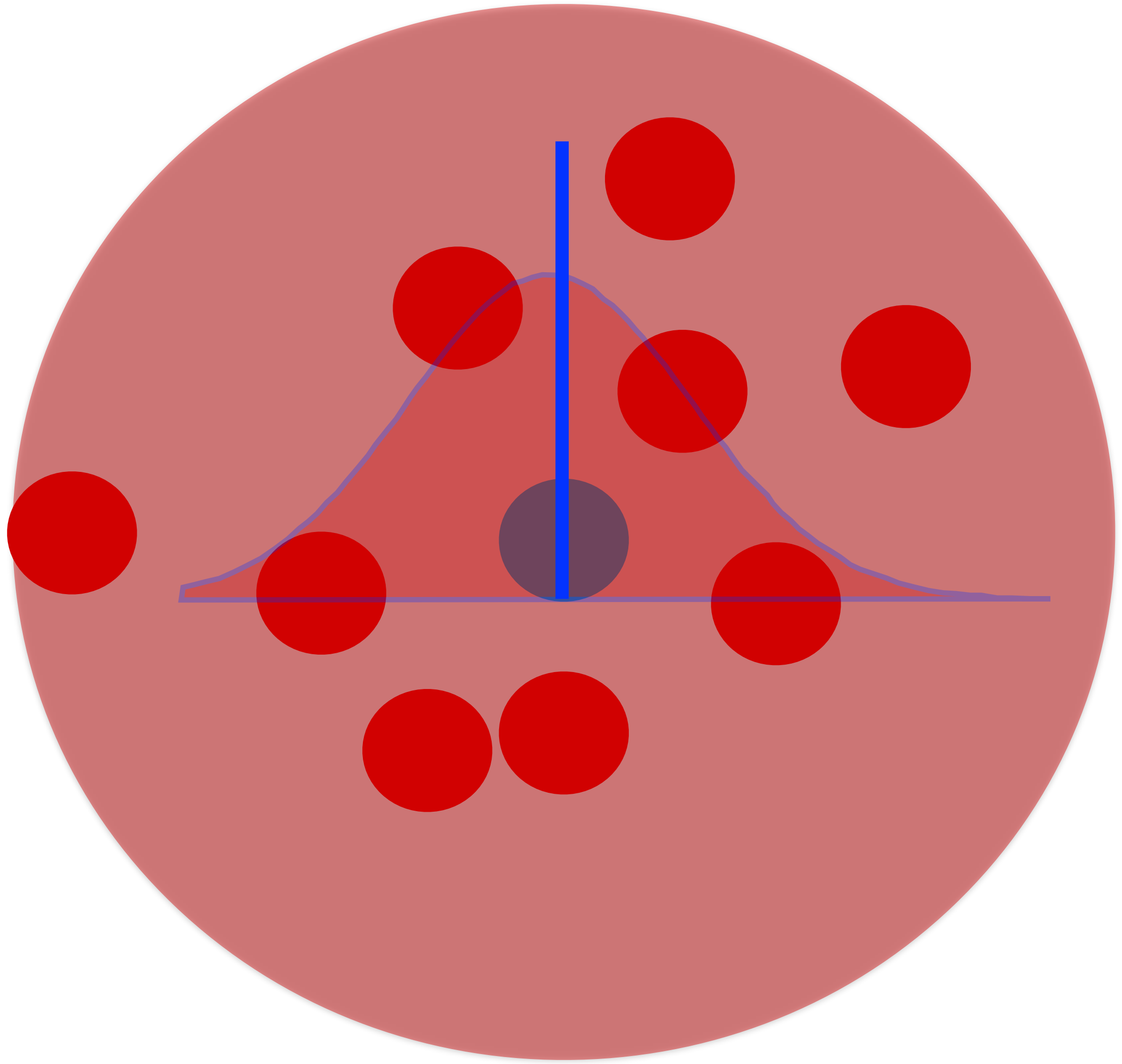
“Sum” them all up

Get resulting PDF of data

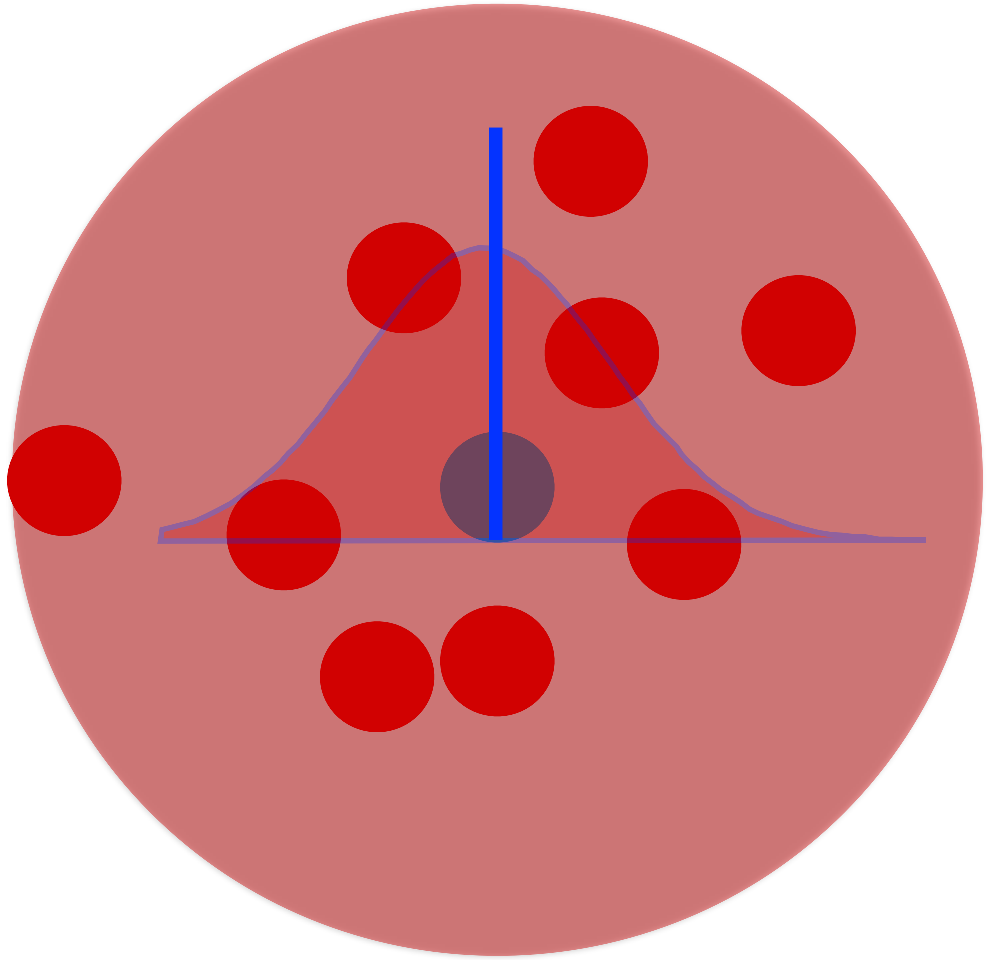


Kernel Density Estimation

Fit distribution from histogram



Gaussian Kernel



Gaussian probability distribution

Defined by

- mean μ
- standard deviation σ

Demo

Distplots

KDE plots

Demo

lmplots for linear relationships

Controlling size and shape of plots

Combination plots

Demo

Categorical plots

Wide form data

Demo

Working with FacetGrids

Demo

Customizing FacetGrids

Demo

Working with PairGrids

Demo

Exploring a car dataset

Demo

Themes and figure styles

Demo

Color palettes

Demo

Overriding styles

Summary

Seaborn is a powerful visualization library

Makes "production ready " plots

Use of histograms, KDE plots, FacetGrids, PairGrids

Specify themes to govern plot aesthetics

Utilize different color palettes