Data Manipulation

Using Python

Dr. Rathachai Chawuthai

Department of Computer Engineering
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang

Python

- List
  - [1, 2, "a"]
  - [1, 2, 3] + [10, 10, 10] = ?
  - [1, 2, 3] * 10 = ?
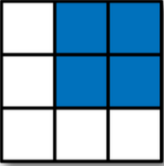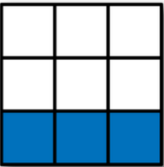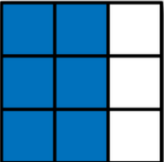- Dictionary
  - { "a":1, "b":2 }
- Tuples
  - (1, 2, "a")

Numpy

NumPy (or Numpy) is a Linear Algebra Library for Python, the reason it is so important for Data Science with Python is that almost all of the libraries in the PyData Ecosystem rely on NumPy as one of their main building blocks.

- Array

- Matrix

- Index

- Selection

- Operation

Pandas

- pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

  - Series

  - DataFrame

  - Operation

  - Join DataFrames

  - Input and Output

  - GroupBy

```
In [2]:    1 iris = sns.load_dataset('iris')

In [3]:    1 #np.r_ is useful when trying to construct arrays
           2 np.r_[0:5, -5:0]

Out[3]: array([ 0,  1,  2,  3,  4, -5, -4, -3, -2, -1])

In [4]:    1 iris.iloc[np.r_[0:5, -5:0]]

Out[4]:
```
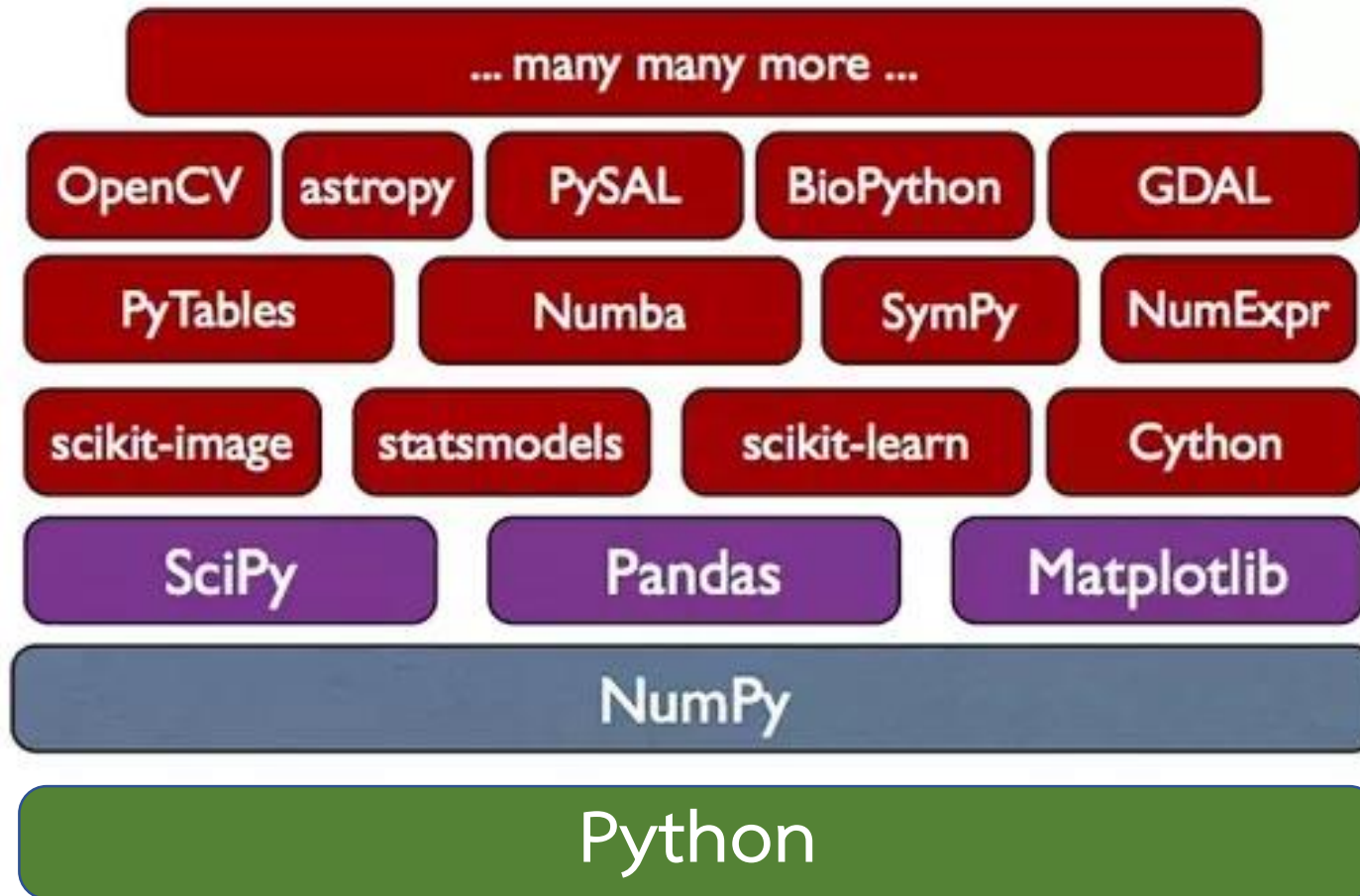
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica

Data Analytics Ecosystem

Next

INTRODUCTION TO DATA ANALYTICS

Data Manipulation using Python

Dr. Rathachai Chawuthai

Department of Computer Engineering
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang

Demonstration

> Data is the sword of the 21st century,
> those who wield it the samurai.

Jonathan Rosenberg

つづく