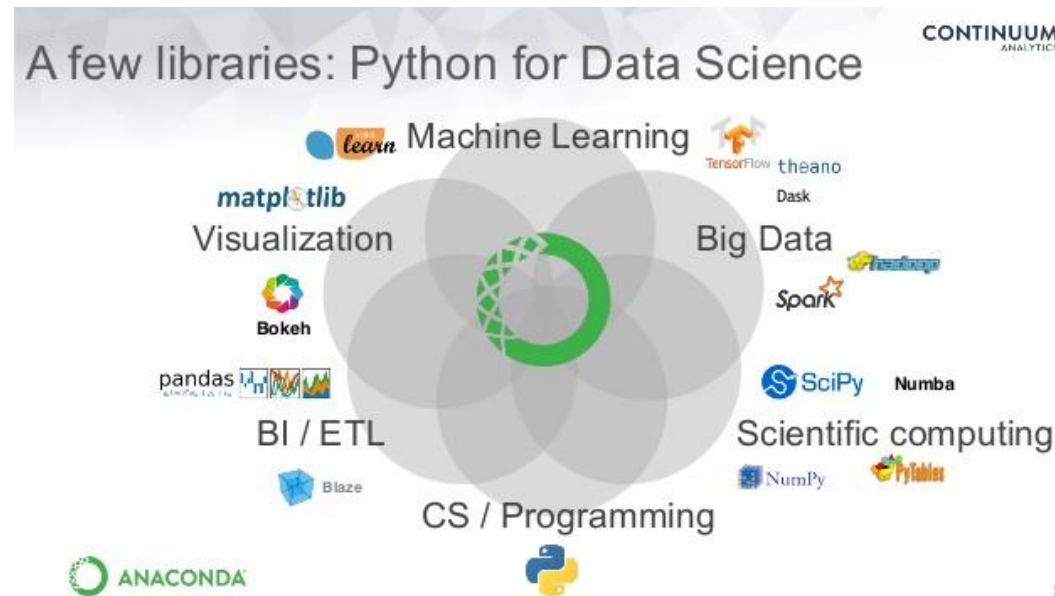# Python Packages for Data Science

JOUR7280/COMM7780

Big Data Analytics for Media and Communication
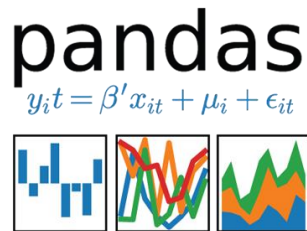
Instructor: Dr. Xiaoyi Fu

Python Libraries

- A Python library is a collection of functions and methods that allow you to perform lots of actions without writing any code.

- The libraries usually contain built-in modules providing different functionalities which you can use directly

A few libraries: Python for Data Science

Scientific Computing Libraries

- Pandas (data structures and tools)

  - The primary instrument of Pandas is the two dimensional table consisting of column and row labels, which are called a data frame.

    - It is designed to provided easy indexing functionality.

- NumPy (arrays & matrices)

  - Can be extended to objects for matrices and with minor coding changes, developers can perform fast array processing

- SciPy

  - Functions for some advanced math problems as well as data visualization.

Visualization Libraries

- Matplotlib (plots & graphs)

  - Most well known library for data visualization

- Seaborn

  - Based on Matplotlib

  - Very easy to generate various plots such as heat maps.

Algorithmic Libraries

- Scikit-learn
  - Machine learning: regression, classification, clustering …
- Statsmodels
  - Explore data, estimate statistical models, and perform statistical tests.

This content is copyright protected and shall not be shared, uploaded or distributed.

Thank You