Algorithms and Data Structures

Python Packages for Data Science

Agenda

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

Matplotlib

- Matplotlib is a graph plotting library
- Specifically designed for data visualization
- Open, free, mostly in Python.

Import

- import matplotlib
- Or:
- import matplotlib.pyplot as plt (most often)

Simple Plot

- Create two lists of data:
  - x for values on x-axis (by default, arange(size of y))
  - y for value on y-axis
  - plt.plot(x, y) for a 2D trend of x and y, or data points as (x, y)
  - Mark:['o', '*', '.', ',', 'x', 'X', '+', 'P', 's', 'D', 'd', 'p', 'H', 'h', 'v', '^', '<', '>', '1', '2', '3', '4', '|', '_']
  - Line: ['-',':','--','-.']
  - Color: ['r','g','b','c','m','y','k','w']
  - Multiple lines

Plot Settings

- Labels:
  - plt.xlabel('labelx')
  - plt.ylabel('labely')
- Title:
  - plt.title('Title')
- Grid
  - plt.grid(), plt.grid(axis = 'x'), plt.grid(axis = 'y')
- Legend
  - plt.legend()

Subplots

- plt.subplot(row, column, position)
- Each subplot should have separate settings (labels, titles, legends)

Scatter plots

- plt.scatter(x, y)
- ColorMap (3rd dimension of data)
  - plt.scatter(x, y, camp)
  - There are many colormaps ready for use
- Dot size (3rd dimension of data)
  - plt.scatter(x, y, size)
  - You can setup the transparency using alpha
  - You can combine colormap and dot size for more information.

Other plots

- Bars
- Histgrams
- Pie
  - Labels
  - Legends
  - Explode

Data Visualization

- We want to deliver information - efficiently and effectively.
- Data Visualization aims at information delivery
- A classic video to watch:
  - Data Visualization by Hans Rolling (https://www.youtube.com/watch?v=RUwS1uAdUcI)
- A good example to read:
  - Tracking Omicron and Other Coronavirus Variants: https://www.nytimes.com/interactive/2021/health/coronavirus-variant-tracker.html

Thank you!