Introduction to Data Analysis Summer School (Python)

M. Fuat Kına

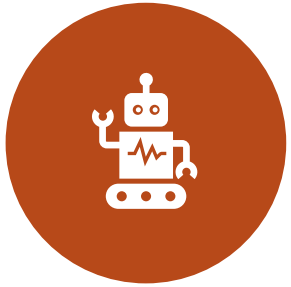
fuatkina@gmail.com

Course objectives

Learn Python 3 syntax

Understand basic programming concepts

Understand data analysis problems and the needed tools to solve them



Get ready for future tasks: Establish a basic understanding of machine learning concepts and algorithms

Why Python?

- Great for beginners and good for advanced use
  - Easily readable code
  - Online resources
- Widely used, especially in scientific computing
- Powerful
  - Advanced data analysis techniques
  - Machine learning modules
- Open-source
- Alternatives for data analysis: R, STATA, SPSS, GIS programs (ArcGIS, QGIS, Geoda) etc.

What will be covered?

- Python basics (data types, lists, sets, dictionaries, basic operations, if statements, functions, and loops)
- Data collection (web scrapping)
- Working with data, creation and manipulation (numpy, matplotlib, pandas)
- Advanced data analysis techniques
  - OLS
  - Spatial statistics
  - Bayesian statistics
  - Machine learning

Summer School Program

	Monday	Tuesday	Wednesday	Thursday	Friday
10:00 - 11:00	Welcome and Introduction to the Summer School	Python Programming: Functions	Introduction to Numpy for Numerical Data	Data Manipulation with Pandas	Basics of Statistical Analysis Using Python
11:00 - 12:00	Setting up Anaconda Navigator and Google Colab	Problem Solving with Python (Exercises)	Numpy (Exercises)	Data Manipulation with Pandas	Basics of Statistical Analysis Using Python
12:00 - 13:00			LUNCH BREAK		
13:00 - 14:00	Python Programming: Data Types, Basic Operations	Problem Solving with Python (Exercises)	Data Visualization Basics with Matplotlib	Data Manipulation with Pandas	Basics of Statistical Analysis Using Python
14:00 - 15:00	Python Programming: If Statements and Loops	Introduction to Numpy for Numerical Data	Matplotlib (Exercises)	Pandas (Exercises)	Discussion about future topics

Course materials

VanderPlas, Jake. 2016. Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media. Available at: https://jakevdp.github.io/PythonDataScienceHandbook/

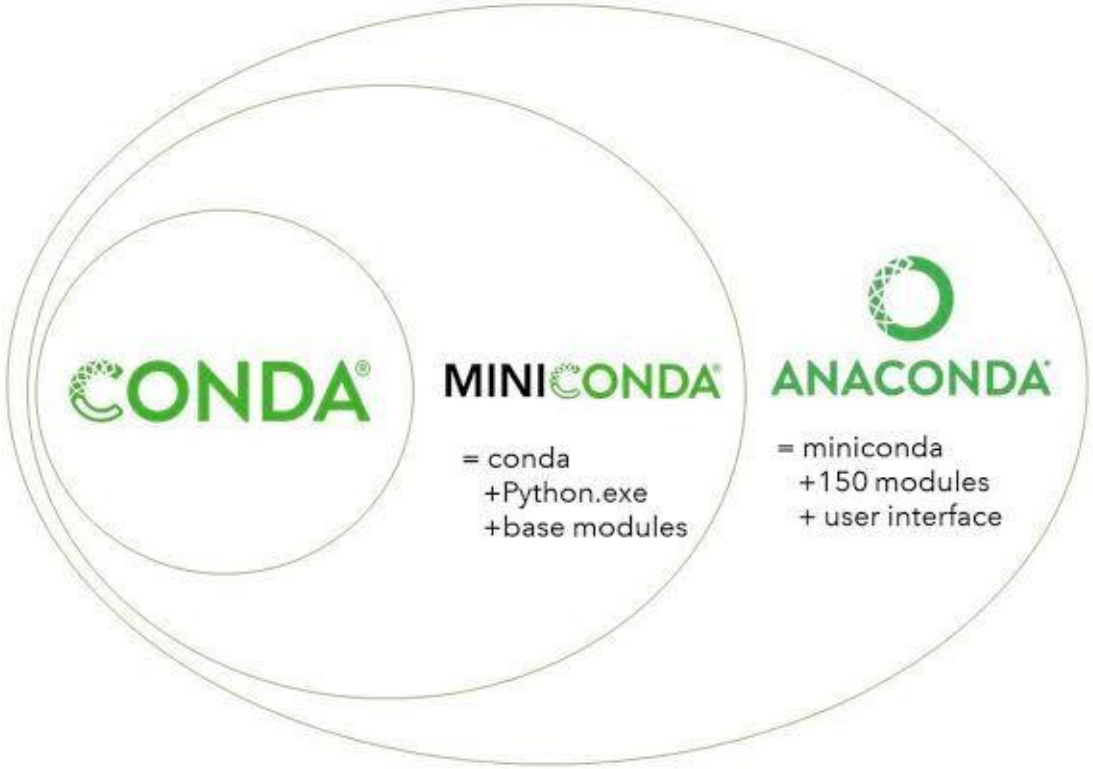
Shaw, Zed A. 2017. Learn Python 3 the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful World of Computers and Code (Zed Shaw's Hard Way Series). 1st Edition. Addison-Wesley. Available at: https://learnpythonthehardway.org/python3/

The Official Documentation for Python. Available at: https://docs.python.org/3/

Let’s look at installation guideline!

You can run python without downloading anything

https://colab.research.google.com/



Conda, miniconda, anaconda

Conda environments

- What is an environment?
  - A conda environment is a directory that contains a specific collection of conda packages that you have installed. For example, you may have one environment with NumPy 1.7 and its dependencies, and another environment with NumPy 1.6 for legacy testing.
  - https://docs.conda.io/projects/conda/en/latest/user-guide/concepts/environments.html#:~:text=A%20conda%20environment%20is%20a,NumPy%201.6%20for%20legacy%20testing.
- How to manage an environment?
  - https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#id2