Intro to Python

Columbia Biostatistics Computing Club
Safiya Sirota & Paula Wu
Fall 2022

What is Python?

- An easy-to-use, popular programming language
- An alternative (or supplement) to R with many applications:
  - Data cleaning, visualization, analysis
  - Web development
  - Desktop development
  - Machine Learning

Why learn Python?

Used in a wide range of fields:

- Game Development
- Business Analytics
- Scientific fields
  - Astronomy
  - Neuroscience
  - Biology
  - Mathematics
  - Biostatistics!

Oct 2022	Oct 2021	Change	Programming Language
1	1		Python
2	2		C
3	3		Java
4	4		C++
5	5		C#
6	6		Visual Basic
7	7		JavaScript
8	10	↑	Assembly language
9	9		PHP
10	8	↓	SQL
11	12	↑	Go
12	14	↑	R
13	29	⇈	Objective-C
14	13	↓	MATLAB

Source: https://www.tiobe.com/tiobe-index/

Python compared to R

- **Which should you use?**
  - It's a matter of context and opinion
  - Both are popular and easier to learn with many useful libraries and tutorials
- **Different environments may favor one over the other**
  - Academic or research settings may favor R
  - Tech, finance, and consulting companies may favor Python
- **It's good to be familiar with both languages.**
- **If you know one it is easier to pick up the other over time!**

How to download Python

- Many computers come with versions of Python already installed. Most Mac OS and Linux operating systems already have it.

- **Here is how to check your Python version:** https://phoenixnap.com/kb/check-python-version

- **If you don't have it yet, here is where you can download Python:** https://wiki.python.org/moin/BeginnersGuide/Download

Where to write Python

- **Using a text editor and running Python through your terminal**
  - (e.g., Atom, Vim, Visual Studio Code, Notepad++)

- **Google Colab Notebook**
  - https://colab.research.google.com/
  - This is the most like R Studio, no need to download anything!

- **Jupyter Notebook**
  - https://jupyter.org/
  - Also like R Studio (However, need to download more things).

What is a Terminal?

- **In simple terms it is a command line system that...**
  - allows you quick access to all your files,
  - lets you run those files, and
  - lets you control your operating system.
- **To run a file you edited in a text editor you will need to go to that file's directory and then run it using Python.**
  - Example command: python3 hello.py
  - Tutorial: https://www.datacamp.com/community/tutorials/running-a-python-script

List of Python tutorials:

- Tutorialspoint: https://www.tutorialspoint.com/python/index.htm
  - This is a great online resource! (Runs Python through terminal)
- Keras: https://keras.io/about/
  - This has a lot of code examples. It is an open source machine learning platform.
  - https://keras.io/getting_started/intro_to_keras_for_researchers/ Intro for researchers
- Python tutorial: https://www.learnpython.org/
- DataCamp: https://www.datacamp.com/?utm_source=learnpython_com&utm_campaign=learnpython_tutorials
- Video Tutorial: https://www.youtube.com/watch?v=rfscVS0vtbw

Common Libraries

- ***numpy*** -> library for some optimized data structures
  - import numpy as np
- ***pandas*** -> library for nice data frames (like tibbles)
  - import pandas as pd
  - https://pandas.pydata.org/
- ***seaborn*** -> library for plotting
  - import seaborn as sns
- ***sklearn*** -> library popular for machine learning
  - import sklearn as sk
  - https://scikit-learn.org/stable/index.html

Questions?