Python

These slides are primarily used to give structure to the live-coding lecture: Week 1 - Monday & Wednesday (Python lecture) - YouTube

R vs Python

- Researchers use both/either
- Both open-source
- Both hugely popular
- Magic is code packages from community members
  - library(tidyverse) **vs** import numpy
- Python → general purpose code
- R → statistics

- R → data visualization
- Python → large scale deployments

Which to pick?

- If you know one → easy to learn the other
- Who will you work with?

Setup comparison

- R setup is widely standardized (RStudio IDE)

- Python setups vary widely
  - Install Python + code editor (e.g., VS code)
  - Install Python + IDE (e.g., Pycharm)
  - Install Anaconda

- Conflicts may arise

Anaconda installation

Download and install the .exe file from this location:
https://www.anaconda.com/products/individual

Run the .exe to install the program. Remember the destination folder, e.g.
“C:\Users\[Your Name]\Anaconda3” on Windows

“/Users/[Your Name]/opt/anaconda3” on Mac

Keep “Register Anaconda3 as my default Python 3.8” selected, we will use Python 3.8 for this course, not Python 2.7

PyCharm installation

Download and install the **Community** .exe file from this location:
Windows: https://www.jetbrains.com/pycharm/download/#section=windows
Mac: https://www.jetbrains.com/pycharm/download/#section=mac
Linux: https://www.jetbrains.com/pycharm/download/#section=linux
(Do not download the Profession version unless you are willing to pay for it in 30 days)

- Start a new “Pure Python” project with Base interpreter: Python 3.8
- OR: Make sure the we are using the correct verison of Python by going to File -> Settings -> Project: -> Python Interpreter -> Python 3.8 at either “C:\Users\[Your Name]\Anaconda3”\python.exe” or “~/opt/anaconda3/bin/python”
    - If you don’t see it here: First make sure you installed Anaconda Python 3 (see last slide)
      If you have: nagivate to Settings symbol on the same page -> System Interpreter -> Navigate to one of the location, either: “C:\Users\[Your Name]\Anaconda3\python.exe” or “/Users/[Your name]/opt/anaconda3/bin/python

Scripts, comments

Variables & data types

Testing and transforming data types

Packages

- library(data.table)
- fread(“data.csv”)

- import pandas
- pandas.read_csv(“data.csv”)

Methods inside objects

Vectors, Lists

- R
- my_atomic_vector = c(1,2,3,4,5)
- my_list= list(1,2,3, “John Smith”, TRUE)

- Python
- my_list = [1,2,3, “John Smith”, True]
- my_tuple = (1,2,3, “John Smith”, True)
- my_atomic_vector = numpy.array([1,2,3,4,5])

Vector, matrix operations

- Adding
- Multiplying
- Matrix multiplication

- Appending
- Deleting

Indexing / Subsetting

- R
- my_list= list(1,2,3, “John Smith”, TRUE)
- my_list[4]

- Python
- my_list= [1,2,3, “John Smith”, TRUE]
- my_list[3]

Missing data types

- None, NaN vs NULL, NA, NaN, Inf

Saving and loading data

- save()
- pickle.dump()

- load()
- pickle.load()

String manipulation

- Finding stuff
  - my_planets = c("Earth", "Jupiter")
  - grep("Jupiter", my_planets)

  - my_planets = ["Earth", "Jupiter"]
  - my_planets.index("Jupiter")

- Replacing stuff
  - gsub(pattern ="World", replacement ="Jupiter", "Hello World")
  - “Hello World”.replace(“World”, “Jupiter”)

Random data

- sample(c("earth","mars","jupiter"), 2)
- random.sample(["mars", "earth", "jupiter"], 2)

- rnorm(100, 100, 10)
- numpy.random.normal(100, 10, 100))

Careful with copies!

- Python v R: Important Differences To Be Aware Of — Practical Data Science

Observed differences R vs Python

- Different setup
- Different packages
- Some differences in brackets/parentheses
- Counting starts at 1 vs 0
- Dots mean nothing vs something
- Careful with overwriting copies
- Indentation matters