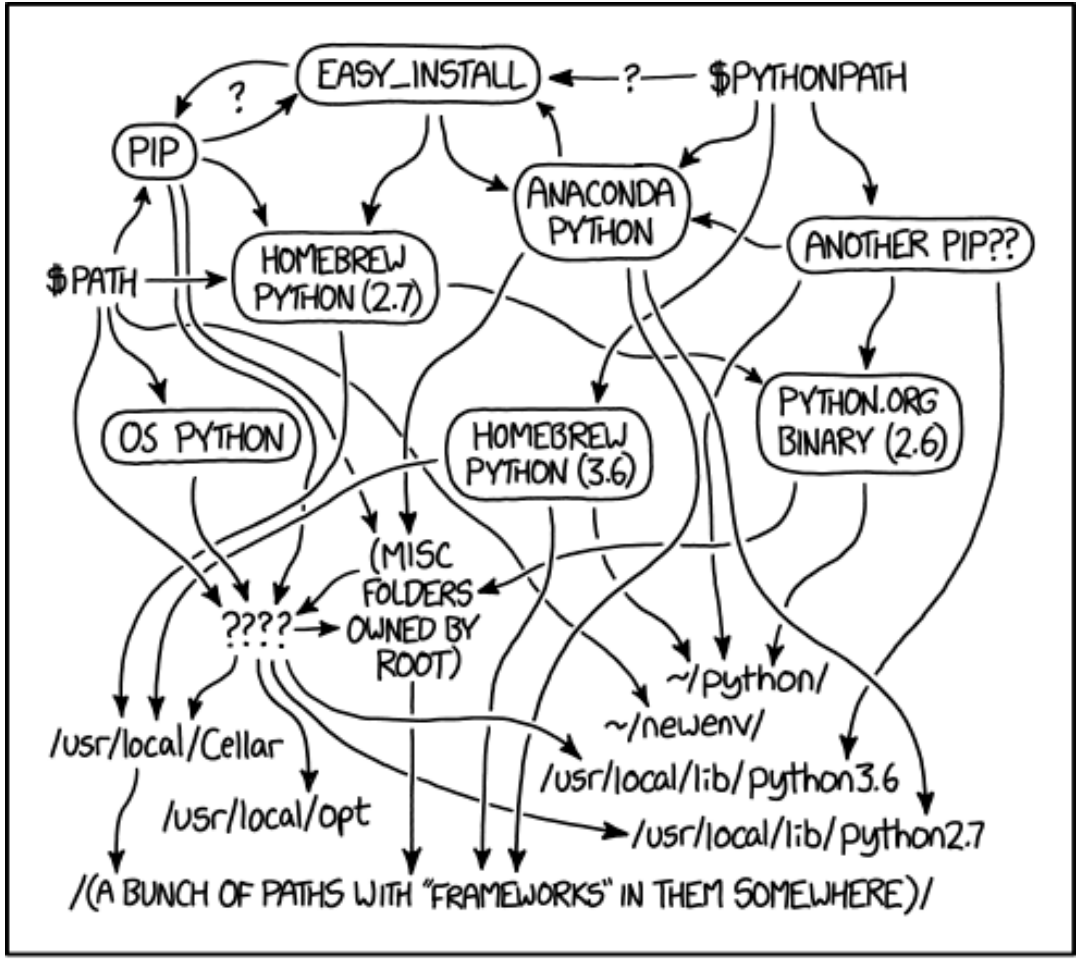


MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

CS 1671 / CS 2071 / ISSP 2071
Human Language Technologies

Session 3: Python for data science

Michael Miller Yoder

January 21, 2025

School of Computing and Information

Course logistics

- Reading for next class, Mon Jan 26
  - Jurafsky + Martin 2-2.6, 2.8, 2.10
- Please remind me of your name before asking or answering a question

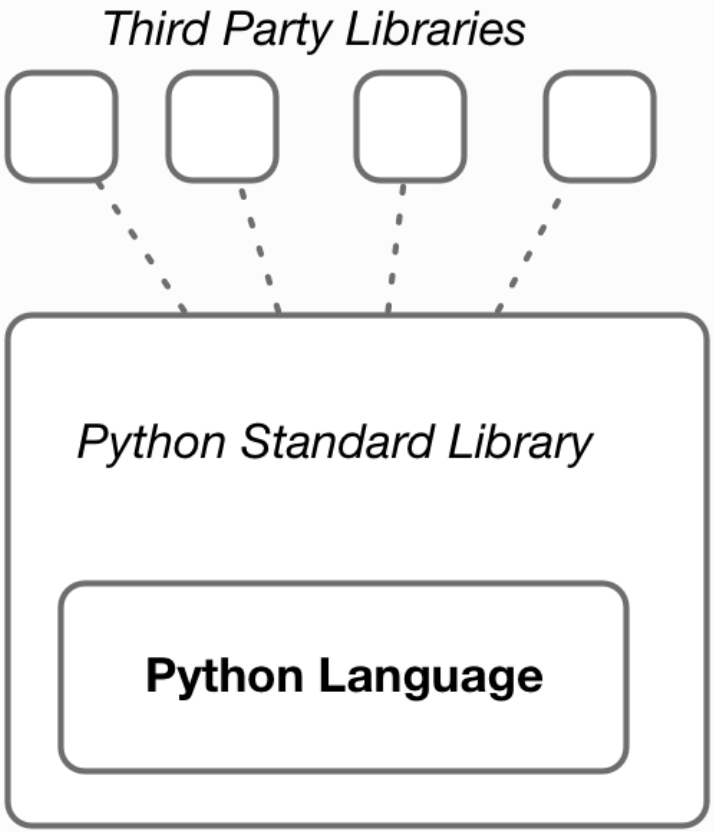
Learning objectives: Python for data science

Students will be able to:

- Create and export Python virtual environments with Anaconda
- Load data into a Pandas DataFrame
- Describe different data types in Pandas
- Select and subset data from Pandas DataFrames
- Write Python functions
- Apply vectorized functions to Pandas DataFrame columns

Python packages and virtual environments

Python packages



Anaconda

- Manages Python packages
  - You can download specific versions of Python packages that are hosted on different “channels”
  - The ‘defaults’ channel contains older, sometimes out-of-date packages
  - conda-forge has newer versions, uploaded by the community
- Manages Python virtual environments
  - Can create environments with specific versions of packages all available
  - You can include a specific version of Python itself within an environment
- Is one of many options for managing Python packages and environments
  - uv, venv, pip are other options

Creating Anaconda environments

- It’s best to create a new environment for each project/course/whatever unit
  - I usually download Miniconda and then just install packages that are needed into each environment
- Two ways of creating Anaconda environments
  - conda create -p *environment_path*
    - Creates an environment at a specific path (preferred)
  - conda create -n *environment_name*
- Activate your conda environment
  - conda activate *environment_path_or_name*
    - If the environment is a file path, put ‘/’ at the end of it

Exporting and copying Anaconda environments

- Export your environment
  - conda env export > *environment*.yml
  - For example, there is an export of the class conda environment at **/ix1/cs1671-2026s/cs1671-2026s_environment.yml**
- Load a new environment from an export (copy it)
  - conda create –p *new_environment_path* –f *environment*.yml

Coding activity: Pandas and Python basics

Load in-class notebooks

1. Go to this nbgitpuller link (also available on course website)
2. Log in with your Pitt username if necessary
3. Start a server with **TEACH – 6 CPUs, 48 GB**
4. Load custom environment at **/ix1/cs1671-2026s/class_env**
5. This should pull the cs1671_spring2026_jupyterhub folder into your JupyterLab
6. Open **session3_pandas.ipynb**