Computing Workshop: Introduction to Python

Matt Ingram, Nakissa Jahanbani, and Cesar Rentería
mingram@albany.edu
Rockefeller College of Public Affairs & Policy
University at Albany, SUNY
Oct. 4, 2018

Overview

Workshop assumes no prior knowledge of Python (or familiarity with any statistical software), so we'll start and build from there.

After the workshop you should have basic answers to the following questions:

(1) what is Python?

(2) what is Python used for?

(3) why might I want to use Python?

(4) how do I get Python?

(5) how do I use Python?

(6) what are some best practices in using Python?

(7) where can I learn more about Python?

What is Python?

- Free, open-source, general-use programming language that also has an excellent set of tools for data analysis

See:

- Python: https://www.python.org/

What is Python used for?

- Lots of things!
  - Data management
  - Data analysis (statistical analysis)
  - And
    - Web development
    - Software and app development
    - Network management
    - May other applications …

- Examples in data science (see also documentation at https://www.python.org/about/apps/)
  - Management:
    - Creating your own data set
    - Cleaning a data set
    - Re-organizing data
    - Transforming existing data into new variables
  - Analysis
    - Descriptive, summary statistics
    - Regression analysis
    - Mathematical modeling and simulation

Why might I want to learn Python?

- Your own data management or analysis (e.g., paper, thesis, dissertation, research for work)
- Widely used and growing fast in academia, especially social sciences
  - http://r4stats.com/articles/popularity/
  - See also IEEE ranking (Python #1 in 2017 and 2018), TIOBE index and PyPL Index
  - Do your own research on Google Trends
- So, you might find that:
  - Your advisor uses Python
  - You want to collaborate with an individual or organization that uses Python
  - You are at a professional event (e.g., conference or workshop) where Python is being used
  - You're asked to review work that uses Python.
- Free, open-source software (more on next slide)

***** I strongly recommend learning free, open tools (Julia, Python, R) *****

Rockefeller College requires Stata in core methods courses, but good to know free, open alternatives.

What does it mean to be free/open?

- Think of freedom of expression or information, not free of cost (though also free of cost)
- Part of FREE, LIBRE, OPEN, FLOSS software movement
  - See: http://www.gnu.org/ and links there
- Benefits:
  - No cost
  - Can retain, reuse, revise, remix, and redistribute freely
  - Can examine existing code to learn/understand
  - Can revise, remix, and redistribute code
  - More intellectual autonomy
  - Large, active, responsive user community
  - Lots of events
- Potential Costs
  - No glossy manuals or documentation
  - No tech support

How do I get Python?

- Two general options: campus computer or own computer
- Campus computer
  - Python should already be installed on all computers in computer labs and library common areas
- Own computer
  - Can download here: https://www.python.org/downloads/
  - Recommendation:
    - Download full Anaconda distribution: https://www.anaconda.com/download/

How do I use Python?

Several ways to use Python:

- conventional: via .py scripts or Integrated Development Environment (IDE) like Spyder
- Recommended: Jupyter
- Always use command files (scripts, i.e., .py files)

Tour

- Open Python in Spyder and Jupyter and take a quick tour of each
- Shift focus to .py files in Jupyter for rest of workshop

Some best practices in Python?

- Use .py files (command file)
- Use two or more .py files, e.g.:
  - Data management
  - Data analysis
- Structure your .py files (header, environment setup, directory setup, etc.)
- Use ProjectTIER protocol “trifecta”
  - Hierarchical directories
  - Set working directory
  - Relative file paths
- Comment extensively (# is comment character in Python)
- Advanced:
  - Examine code underlying functions
  - Try adapting or generating your own functions

Sources/Resources

Getting Started in Anaconda: http://docs.anaconda.com/anaconda/user-guide/getting-started/

UCLA Sandbox "Introduction to Python 2018": https://sandbox.idre.ucla.edu/sandbox/introduction-to-python-2018

SciPy (scientific Python): https://www.scipy.org/

Keras (machine learning): https://keras.io/

DataCamp Free Course, Introduction to Python for Data Science: https://www.datacamp.com/courses/tech:python

User groups: Cross-Validated (https://stats.stackexchange.com/), Stack Overflow (https://stackoverflow.com/questions/tagged/python); note both are from Stack Exchange

ProjectTIER: https://www.projecttier.org/