

PPHA 30537: Data and Programming for Public Policy I (Python)

Jeff Levy
levyjeff@uchicago.edu
Keller 3101

Spring Quarter, 2022

Course Information

March 28th - May 27th

Section 1: T & Th, 11:00 - 12:20, Keller TBD

Section 2: T & Th, 3:30 - 4:50, Keller TBD

In-person attendance to scheduled class times is mandatory and graded. Per university guidelines, requests for remote attendance must come through student disability services or the Title IX office.

This class will have three primary elements:

1. **Asynchronous lectures** posted to Canvas every Monday and Wednesday around noon, which will introduce new content. I aim to keep these around 30 minutes in length.
2. The scheduled lecture times will be used as **live labs**, in which we delve deeper into the content introduced in the lectures, and work through examples in groups.
3. Weekly optional **office hours** for the professor and TAs, where individuals can get one-on-one help with questions.

Course Objectives

While proprietary platforms such as Stata (or even Excel) continue to play an important role in public policy research, newer open-source languages like Python and R have grown rapidly in usage. A good researcher in these fields must be able to adapt, changing tools (languages) as called for by the project.

As one of the most utilized (3rd), most loved (6th), and most desired (1st) programming languages in the world, Python is an excellent choice for a new researcher to focus on. Python emphasizes a clear syntax, making code easy to learn and easy to read, while remaining both powerful and flexible. This makes it an ideal platform in which to learn the basics of rigorous data analysis in a way that applies to any programming language.

Note that this course is also offered in the R programming language as PPHA 30535, and that both language tracks fulfill the first course in the STEM-eligible Harris School Certificate in Data Analytics.

Office Hours (1hr each)				
Tuesday	10:00	AM	Keller 3101	Jeff Levy
Tuesday	6:00	PM	Zoom	Phoebe Collins
Wednesday	9:30	AM	Zoom	Yangzhou Ou
Wednesday	3:00	PM	Zoom	Justin Berg
Wednesday	7:00	PM	Zoom	Ziyang (Ryan) Wang
Thursday	12:30	PM	Zoom	Valeria Balza Pineda
Thursday	6:00	PM	Zoom	Ruyang (Sherry) Shao
Friday	1:00	PM	Zoom	Yuetong (Yolanda) Feng
By appointment				Jonas Heim

This programming and data course is geared toward public policy students who have either **no past programming experience**, or **some experience** primarily in proprietary data platforms like Stata.

Technical goals:

- Learn to write basic Python
- Gain a deeper understanding of how Python works
- Learn the tools of data analysis in Python

Non-technical goals:

- Practice good programming and data principles that are relevant to working in other languages, such as R, Stata, or SAS
- Understand how good programming relates to reproducible research
- Develop skills that apply directly to summer internships working with data

Software and Resources

There are no required text books for this class. Python is extremely well supported online. I expect students will primarily be using the official Python documentation and StackOverflow, which will be discussed in class.

For the data sections of the class, the text Python for Data Analysis 2nd Edition by Wes Mckinney may be helpful as both a quick reference or when read comprehensively as a guide, but will not be referenced directly in class.

There are two pieces of software that are required for this class, both of which are free:

- The Anaconda Python distribution
- The GitHub Desktop application

Homework, Exams, and Grading

There are no exams or projects in this class. Assignments are given on Mondays and due on GitHub Classroom just before midnight on Sundays.

Your grade will be calculated as **90% weekly assignments, 10% attendance to scheduled class times** for your section. This class requires a 60% or above to pass, and is not curved. All passing grades will use the following intervals:

A	[96% - 100%]
A-	[91% - 96%)
B+	[86% - 91%)
B	[81% - 86%)
B-	[60% - 81%)

Late Assignments

Every student has **four 12-hour extensions** available to them during the quarter. Those extensions will be automatically applied to any late work, and require no excuse to be given. **To turn in late work, commit your final code as usual, then email harris.data.skills.class@gmail.com with your GitHub ID, so the graders know to pull your latest code.** These extensions are used in complete blocks of time - e.g. turning an assignment in 12 hours and 30 minutes late will use two tokens.

Once your extensions are used up for the quarter, all assignments will be penalized at a rate of 5% per 12-hour block. Only issues of sufficient magnitude that academic affairs is involved in the discussion can qualify for exceptions. Once sample solutions have been posted to the class (generally Wednesday), no further assignments may be turned in.

Other Information

1. See the academic integrity policy and the grading rubric on the Canvas course site.
2. Find information about the Harris Student Affairs office here.
3. Learn more about accommodations for students with disabilities here.

Course Outline

This outline may be subject to change. The dates next to homework assignments are in the form: date given - date due.

Week 1: Introduction - No homework

- March 29th: Introduction, software review and setup
- March 31st: Setup, GitHub basics

Week 2: Python basics - Homework 1 (Apr 4th-Apr 10th)

- April 5th: Data types
- April 7th: Logic control statements and loops

Week 3: Python functions and classes - Homework 2 (Apr 11th-Apr 17th)

- April 12th: Functions and lambdas
- April 14th: Classes and methods

Week 4: Python functions and classes - Homework 3 (Apr 18th-Apr 24th)

- April 19th: More functions and classes
- April 21st: More functions and classes: creating a game

Week 5: The Pandas dataframe - Homework 4 (Apr 25th-May 1st)

- April 26th: Pandas I
- April 28th: Pandas II

Week 6: More Pandas - Homework 5 (May 2nd-May 8th)

- May 3rd: Pandas III
- May 5th: Pandas IV

Week 7: Data visualization - Homework 6 (May 9th-May 15th)

- May 10th: Matplotlib I
- May 12th: Matplotlib II

Week 8: Web scraping - Homework 7 (May 16th-May 22nd)

- May 17th: Using data APIs; introduction to html
- May 19th: Requests and BeautifulSoup

Week 9: Advanced topics - No homework

- May 24th: NumPy and Statsmodels
- May 26th: Data transformations and missing values