

PPHA 30538: Data and Programming for Public Policy II - Python Programming

Jeff Levy
levyjeff@uchicago.edu
Keller 3101

Fall Quarter, 2022

Course Information

September 27th - December 5th

Section 1: T & Th, 2:00 - 3:20, 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, 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.

Office Hours (1hr each)					
By Appointment				Jeff Levy	
By Appointment				Jonas Heim	
TBD	TBD	TBD	Zoom	TBD	TBD@uchicago.edu
TBD	TBD	TBD	Zoom	TBD	TBD@uchicago.edu

Prerequisites

The course PPHA 30537, Data and Programming for Public Policy I - Python Programming, is required to take this course. If you did not take PPHA 30537, or took sections taught in R, you should email me for approval and will need to demonstrate proficiency with Python and Pandas.

Note that the R equivalent for this course is offered concurrently as PPHA 30536.

Course Objectives

This course will build directly on the material covered in PPHA 30537. We will assume a grasp of the Python skills from the previous class at the start, so that we can focus on practical applications to research. Whereas the goal of the first class was introduce Python as a tool for data analysis, and to prepare students for internship-level policy research positions, the goals of this course will be to:

1. Go from simply applying Python to solve research questions, to applying Python professionally, in a way that supports code maintenance, collaboration, efficiency, and readability
2. Deepen existing skills; for example, we will go from learning to create plots to discussing the principles of creating good plots
3. Broaden into new skills that require a higher level of Python proficiency
4. Prepare for the post-graduation job market

Software and Resources

The software and resources for this class are identical to the Python sections of PPHA 30537 in the spring. There are two pieces of software that are required for this class, both of which are free:

- The Anaconda Python distribution (or similar)
- The GitHub Desktop application

Homework, Exams, and Grading

Your grade will consist of four assignments and a project. All code must be turned in using GitHub Classrooms, no exceptions. Dates below are listed as *date given - date due*:

- Homework 1: Coding and Data *Oct 3-Oct 10*
- Homework 2: Data Visualization *Oct 11-Oct 24*

- Homework 3: Natural Language Processing *Oct 25-Nov 7*
- Homework 4: TBD *Nov 8-Nov 14*
- Final Project: *Sep 26-Dec 7*

Your grade will be calculated as 50% assignments, 50% final project. A minimum of 60% is required to pass this course. Among those who pass, final grades will use the following curve: 1/3 A, 1/4 A-, 1/4 B+, 1/12 B, 1/12 B-.

Late Assignments

Every student has **two 12-hour extensions**. 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.** Extensions are used in complete blocks of time - e.g. turning an assignment in 12 hours and 30 minutes late will use two extensions.

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. Late tokens may not be used for the final project.

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. Given date is for the live lab - pre-recorded lectures will be posted the day before.

Week 1 - Homework 1 Material

- September 27th - Introduction, project discussion
- September 29th - Code generalization and organization, functions, PEP8

Week 2

- October 4th - Difficult and irregularly-shaped data 1
- October 6th - Difficult and irregularly-shaped data 2

Week 3 - Homework 2 Material

- October 11th - Data visualization
- October 13th - Interactive plotting

Week 4

- October 18th - Spatial data 1
- October 20th - Spatial data 2

Week 5 - Homework 3 Material

- October 25th - PDF parsing and text processing
- October 27th - Natural Language Processing 1

Week 6

- November 1st - Natural Language Processing 2
- November 3rd - Natural Language Processing 3

Week 7 - Homework 4 Material

- November 8th -
- November 10th -

Week 8 - Presentations

- November 15th - Group project presentations
- November 17th - Group project presentations

Week 9 - Thanksgiving Break

- November 22nd - No Class
- November 24th - No Class

Week 10 - Final Topics

- November 29th - Other common data languages
- December 1st - Code samples and job prep