

PPHA 30536: Data and Programming for Public Policy II (Python)

Jeff Levy
levyjeff@uchicago.edu
Keller 3101

Fall Quarter, 2021

Course Information

September 27th - December 3rd

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

Section 2: T & Th, 2:00-3:20, Keller TBD or Zoom

All in-person and synchronous content for this class is optional; you may take this class even if the scheduled time conflicts with another course, or if you are in a timezone that makes it impractical.

This class will have three primary elements:

1. **Asynchronous lectures** posted to Canvas every Monday and Wednesday before noon, which will introduce new content. I will strive 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, often in groups. You can attend and participate in these labs in-person in the Keller Center, or live over Zoom. If you cannot attend at the scheduled time, the content will be recorded on Canvas. All students in the class will be responsible for the material covered, which will be accompanied by a short quiz (1-3 questions) on Canvas covering the content, due that same day.
3. Weekly optional **office hours** for the professor and TAs, where individuals can get one-on-one help with questions.

Office Hours					
Thursday	3:30	PM	Keller 3101 or Zoom	Jeff Levy	
?	?	?	Zoom	TA1	ta1@uchicago.edu
?	?	?	Zoom	TA2	ta2@uchicago.edu
?	?	?	Zoom	TA3	ta3@uchicago.edu
?	?	?	Zoom	TA4	ta4@uchicago.edu

Prerequisites

The course PPHA 30535, Data and Programming for Public Policy I, is required to take this course. If you did not take PPHA 30535, or took PPHA 30535 sections taught in R, you should email me for approval, and to demonstrate proficiency with Python and Pandas.

Note that Data and Programming for Public Policy II in the R language will be offered Winter quarter 2022.

Course Objectives

This course will build directly on the material covered in PPHA 30535. 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 30535 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

There are no exams for this class. Your grade will consist of four assignments, twice-weekly lab quizzes, and a project. Quizzes will be on Canvas, while all code must be turned in using GitHub Classrooms. Dates below are listed as *date given - date due*:

- Homework 1: Coding and Data - Oct 4 - Oct 10
- Homework 2: Natural Language Processing - Oct 11 - Oct 20
- Homework 3: Data Visualization - Oct 20 - Oct 31
- Homework 4: Spatial Data - Nov 1 - Nov 7
- Final Project: Sep 28 - Dec 7

Your grade will be calculated as 45% assignments, 45% final project, and 10% quizzes. 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 or quizzes.

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

- September 28th - Introduction, project discussion
- September 30th - Parsing data from PDF documents

Week 2

- October 5th - Code generalization and organization, functions, PEP8
- October 7th - Irregularly shaped data

Week 3

- October 12th - Natural Language Processing 1
- October 14th - Natural Language Processing 2

Week 4

- October 19th - Natural Language Processing 3
- October 21st - Data visualization 1

Week 5

- October 26th - Data visualization 2
- October 28th - Data visualization 3

Week 6

- November 2nd - Spatial data 1
- November 4th - Spatial data 2

Week 7

- November 9th - Intro to Machine Learning in Python 1 (tentative)
- November 11th - Guest speaker

Week 8

- November 16th - Group project presentations
- November 18th - Group project presentations

Week 9

- November 23rd - Thanksgiving Break
- November 25th - Thanksgiving Break

Week 10

- November 30th - Other common data languages
- December 2nd - Creating good code samples, wrap-up, additional topics