

# PPHA 30536: Data and Programming for Public Policy II - R Programming)

Jeff Levy  
levyjeff@uchicago.edu  
Keller 3101

Fall Quarter, 2022

## Course Information

September 27th - December 5th

Section 1: M & W, 9:00-10:20, Keller TBD

Section 2: M & W, 10:30-11:50, 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 Sunday and Tuesday, 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 30535, Data and Programming for Public Policy I - R Programming, is required to take this course. If you did not take PPHA 30535, or took PPHA 30537 taught in Python, you should email me for approval, and may be required to demonstrate proficiency with R.

Note that the Python equivalent for this course is offered concurrently as PPHA 30537.

## Course Objectives

This course will build directly on the material covered in PPHA 30535. We will assume a grasp of the R 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 to introduce R and ggplot as tools for data analysis, and to prepare students for internship-level policy research positions, the goals of this course will be to:

1. Go from applying R to structured questions with clearly-defined answers, to using R to solve broad research questions
2. Deepen existing skills; for example, we will spend more time using dplyr and functions
3. Broaden into new skills that require a higher level of R proficiency
4. Prepare for the post-graduation job market

## Software and Resources

*Most of the software and resources for this class are identical to PPHA 30535.* There are three pieces of software that are required for this class, all of which are free:

- The R Studio platform
- The GitHub Desktop application
- The Anaconda Python distribution (for the two lectures on Python)

## Homework, Exams, and Grading

Your grade will consist of four assignments and a project. **All code must be turned in as plain R files (R-Markdown is not accepted) using GitHub Classrooms**, no exceptions. Dates below are listed as *date given - date due*:

- Homework 1: Coding and Data *Oct 2-Oct 9*

- Homework 2: Spatial Data and Shiny *Oct 10-Oct 23*
- Homework 3: Text Processing *Oct 24-Nov 6*
- Homework 4: Beginning Python *Nov 7-13*
- Final Project: *Sep 27-Dec 7*

Assignments are due before midnight on the date listed. **Your final 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](mailto:harris.data.skills.class@gmail.com) with your GitHub ID, so the graders know to pull your latest code.** Do not use this address for any other purpose. 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 28th - Introduction, project discussion
- September 30th (Friday makup class) - Style, functions, and loops

## **Week 2**

- October 3rd - Difficult and irregularly-shaped data 1
- October 5th - Difficult and irregularly-shaped data 2

## **Week 3 - Homework 2 Material**

- October 10th - Spatial data 1
- October 12th - Spatial data 2

## **Week 4**

- October 17th - Interactive plotting 1
- October 19th - Interactive plotting 2

## **Week 5 - Homework 3 Material**

- October 24th - Text and language processing 1
- October 26th - Web scraping

## **Week 6**

- October 31st - Text and language processing 2
- November 2nd - PDF parsing and text and language processing 3

## **Week 7 - Homework 4 Material**

- November 7th - Intro to Python 1
- November 9th - Intro to Python 2

## **Week 8 - Presentations**

- November 14th - Group project presentations
- November 16th - Group project presentations

## **Week 9 - Thanksgiving Break**

- November 21st - No Class
- November 23rd - No Class

## **Week 10 - Final Topics**

- November 28th - Data transformations
- November 30th - Code samples and job prep