



**THE UNIVERSITY OF CHICAGO**  
**HARRIS SCHOOL**  
**OF PUBLIC POLICY**

**PPHA 30537: Data and Programming  
for Public Policy I - Python**

**Dr. Christopher Clapp**

**Syllabus, Spring 2023**

Class:

Days: T & Th

Section 01 - 9:30am-10:50am

Section 02 - 11:00am-12:20pm

Locations:

Keller 1002

Keller 1002

Professor: Chris Clapp (he/him)

Office Hours: Th 4:00-5:00pm

or by appointment

Email: [cclapp@uchicago.edu](mailto:cclapp@uchicago.edu)

Location: Keller 3039

Head TA: Jonas Heim (he/him)

Email: [jonas.heim@uchicago.edu](mailto:jonas.heim@uchicago.edu)

TAs:<sup>1</sup>

**Pavan Prathuru** (he/him)

**Victor Perez** (he/him)

**Michael Wagner** (he/him)

**Sai Kandukuri** (he/him)

**Fadhil Muhammad** (he/him)

**Jhony Pulido** (he/him)

## Course Description

In this course, aspiring data analysts, researchers, programmers, and policymakers will study rigorous data and programming using Python. As one of the **most utilized** (3rd) and **most desired** (1st) programming languages in the world, Python is an excellent choice for a new data analyst 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 on which to learn the basics of data analysis in a way that applies to any programming language. While proprietary platforms such as Stata and SAS 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 analyst in these fields must be able to adapt by changing tools (languages) as called for by the project.

Generations of researchers and practitioners have grown up in a computing environment dominated by this small number of proprietary computing platforms while relying on ad hoc coding skills acquired through trial and error. This imposes real costs, including the inability to collaborate with analysts using other platforms, difficulty picking up new skills, trouble finding needed functions that only exist in a different language, and worst of all, **mistakes that taint results** while hiding in sloppy code and bad practices.

**This programming and data course is geared toward public policy students who have either no past programming experience, or minimal experience in other platforms.** While the course covers basic programming, the focus wherever possible will be on applications to real-world data and research. It is designed to continue seamlessly into PPHA 30538: Data and Programming for Public Policy II – Python Programming in the autumn, which will culminate in a final research project covering topics from both classes.

---

<sup>1</sup>TA office hours are listed on the Canvas Zoom page.

## Modes of Learning

Instruction for this class will have four primary elements:

- New content will be introduced in **asynchronous lectures** posted to Canvas around noon on the day before scheduled class time. I will aim to keep those lectures to less than 30 minutes in length.
- 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.
- Weekly optional **office hours** for the professor and TAs, where individuals can get one-on-one help with questions.
- An optional **discussion board** for questions and discussions outside of office hours and class.

## Learning Objectives: What's My Incentive for Taking This Course?

Technical goals:

- Learn to write basic Python and understand its syntax.
- Apply the tools of data analysis in Python.
- Gain a deeper understanding of how Python works “under the hood.”

Non-technical goals:

- Practice good programming and data principles that are relevant to working in other languages, such as R, Stata, or SAS, and how to make informed choices between them and Python.
- Understand how good programming practices relate to collaboration and reproducible research.
- Become comfortable looking up new programming skills and information using online resources.
- Develop skills that apply directly to summer internships working with data.

## Evaluation

Your progress towards achieving the learning objectives via the modes of learning and your final grade in this course will be related to performance in two areas. The weight placed on each component will be as follows:

In-Class Quizzes	15%
Take-Home Assignments	85%

Quizzes: Each class will have a brief (5 minute, 2-3 question) quiz on Canvas that will cover a core skill or concept from lecture. These will provide important instant feedback on the material.

Assignments: There will be coding assignments assigned during the second through eighth weeks of the class that ask students to use class concepts to solve research programming questions. Assignments will test your ability to work on a question with a starting place and a broad goal, mimicking real-world data analysis tasks wherever possible.

## Grades

This class requires a 60% or above to pass and is not curved. All passing letter grades will be determined based on the following intervals used in the Data Science Certificate sequence:

A [95% – 100%] | A- [90% – 95%) | B+ [85% – 90%) | B [80% – 85%) | B- [60% – 80%)

Pass/Fail (P/F), Withdrawal, and Incomplete grade requests will be handled in accordance with University and Harris policy. Students who wish to take the course pass/fail rather than for a letter grade must use the Harris P/F request form (<https://harris.uchicago.edu/form/pass-fail>) and must meet the Harris deadline, which is generally 9am on the Monday of the 5th week of courses. To earn a P grade, students taking the course P/F must: submit all seven assignments and earn a grade that is overall equivalent to at least a B- letter grade.

## Materials

### Textbooks

There are no required textbooks for this class, as Python is extremely well supported online. I expect students will primarily use the official [Python documentation](#) and [StackOverflow](#), which will be discussed in class.

The text *Python for Data Analysis*, 3rd Edition, by Wes McKinney (ISBN-10: 109810403X) may be helpful as both a quick reference and when read comprehensively as a guide, but will not be referenced directly in class.

### Data Analysis and Statistical Software

We will use Python software in this class, and you are required to code all assignments in Python.<sup>2</sup> There are two pieces of software that are required for this class, both of which are free and can be installed on all operating systems:

- The [Anaconda Python](#) distribution, and
- The [GitHub Desktop](#) application.

### Harris Tutoring Program

Harris offers 10 hours of free tutoring support for coding in Python, Stata, and R. Tutoring will be available to Harris students starting Week 3 of the quarter. You can read more about the program on the Harris Student Handbook Canvas site (<https://canvas.uchicago.edu/courses/42004/pages/harris-tutoring-program>). Any questions should be directed to your academic advisor or [harrisdeanofstudents@uchicago.edu](mailto:harrisdeanofstudents@uchicago.edu).

## Course Policies

### General

There is no **attendance** requirement, but regular attendance is necessary (but not sufficient) to do well in the class. Quizzes can only be completed in the classroom. Your lowest two quiz scores are dropped automatically to account for unexpected absences or illness.

---

<sup>2</sup>Note that there is an analog of this class that is taught with R software by another instructor (PPHA 30535).

Assignments must be **turned in** using GitHub and Gradescope, a process we will cover in class. General feedback according to an assignment-specific rubric will be provided through Gradescope approximately one week after the due date.

Every student has four **12-hour late tokens** available to them during the quarter. These extensions will be automatically applied to any late take-home assignments and require no excuse to be given. These extensions are used in complete blocks of time – e.g. turning in an assignment 12 hours and 30 minutes late will use two tokens. Once your late tokens are used up for the quarter, all assignments will be penalized at a rate of 5% per 12-hour block. These tokens are intended to cover ordinary illness, family events, and so on – only issues of sufficient magnitude that academic affairs is involved in the discussion can qualify for exceptions. Once solutions have been posted to the class (generally Wednesday), no further assignments may be turned in.

**Regrade requests** must be submitted on Gradescope with a (polite) explanation, which will then be re-evaluated by the original grader. Continued disagreement may be escalated to the head TA first, and finally to the professor. Regrade requests handled by the professor will result in the entire assignment being regraded, and a new score, which may be higher or lower, being awarded. Please see the *assignment guidelines* document on Canvas for additional details.

The class webpage is available through the **Canvas** portal. I will use it to post announcements, assignments, and grades. Email, Canvas postings, and the discussion board are the official means of communication for out-of-class messaging. Please check your UChicago email account and the Canvas site regularly.

I will **record** asynchronous lectures and post them only to Canvas in accordance with University and Family Educational Rights and Privacy Act (FERPA) guidelines. The University has developed specific policies and procedures regarding the use of video/audio recordings that are explicitly described in the University’s student manual (<https://studentmanual.uchicago.edu>). You are expected to abide by these policies.

Email is inefficient. If you have a question about the class or the material, others probably do too! Questions and answers (knowledge) are public goods, so post your question to the **discussion board**, and feel free to answer questions your classmates ask. The TAs and I will monitor and respond as well.

If you have a question or concern about something you don’t want to discuss publicly, feel free to email me. I will respond to email within 2 business days (Monday-Friday, 9:00am-5:00pm). I teach multiple classes, so please include “DAP1:” as a prefix to your subject.

Any and all results of in-class and out-of-class assignments and examinations are data sources for research and may be used in published research. All such use will always be anonymous.

Please see the *academic integrity policy* and a general *grading rubric* on the Canvas course page.

## **COVID-19 Pandemic**

Students are expected to abide by the University’s health protocols. Note that the protocols, which address masking, self-monitoring, testing, reporting, and isolating requirements, represent evolving guidance and are subject to change (<https://goforward.uchicago.edu/>).

The class will be taught in-person. Should changing pandemic conditions necessitate, we will switch to holding class remotely according to University policy and my discretion. **If you are experiencing COVID-19 symptoms or are required to quarantine, please do not attend class!** Given the live-lab nature of the class, a remote (dual-modality) option for students needing temporary accommodations for short-term absences is (unfortunately) not feasible. However, those examples will be available on Canvas, and I will be happy to answer any questions you may have about the in-class examples.

If you need a more-permanent remote-learning accommodation, please contact the Dean of Students, Kate Biddle ([kbiddle@uchicago.edu](mailto:kbiddle@uchicago.edu)). Per Harris policy, all such requests can only be approved centrally, not by

individual instructors. More generally, if you get sick, are caring for a sick relative, or anything else that becomes an obstacle to your coursework, please inform me and your advisor as soon as you are able. We will all work together to develop appropriate accommodations.

### **Americans With Disabilities Act**

Students with disabilities needing an academic accommodation should contact UChicago's Student Disability Services (SDS). Please see their webpage for contact information (<https://disabilities.uchicago.edu>). If SDS determines a disability accommodation is appropriate, you should inform the Harris Dean of Students office by the end of the first week of class. The Harris Dean of Students office will work with the student and instructor to coordinate the implementation of the student's accommodations. Harris students are not required to submit their accommodations letter to the instructor, but please feel free to come talk to me if you are comfortable doing so. I'm happy to support your learning however I can.

### **Mental Health Services**

As graduate students, I recognize that you are all under immense pressure to achieve academic excellence while maintaining personal and often professional lives. Students differ in how much they know about mental health services. Your use of UChicago's Student Health and Counseling Services (SHCS) is free, confidential, and not linked to your academic file. If you find yourself suffering in silence, please do not hesitate to make use of the services provided by SHCS. Please see SHCS' mental health webpage for services and contact information (<https://wellness.uchicago.edu/mental-health/>). And if you are having serious mental, physical, or other problems, immediately contact the urgent medical care line at (773) 702-3625 (available 24 hours a day, 7 days a week).

### **Diversity and Inclusion**

UChicago is committed to diversity and rigorous inquiry that arises from multiple perspectives, and Harris encourages thought-provoking discourse that involves not only speaking freely about all issues but also listening carefully and respectfully to the views of others. I concur with this commitment and view the diversity that students bring to my class as a valuable resource and a benefit to learning. I expect to maintain a productive learning environment based on open communication, mutual respect, and non-discrimination. I strive to present materials in a way that is respectful of diverse student backgrounds. As there can always be a gap between intent and execution, suggestions for promoting a positive and open environment are welcomed. Please feel free to correct me on your preferred name and gender pronouns if necessary.

### **Responsible Employees (Title IX)**

All University of Chicago faculty and TAs are classified as "Responsible Employees." As such, they are required to report any discussions of sexual misconduct, dating violence, domestic violence or stalking to the Title IX Coordinator for the University. This includes the identities of the student making the complaint and alleged perpetrator. You will receive an email once a report is filed, but you are not obligated to meet with anyone or engage in the process. Alternatively, there are "Confidential Resource" employees at the University who do not have an obligation to share identifying information. For more information, including phone numbers, see the UChicago U Matter website (<https://umatter.uchicago.edu/find-support/>).

## Syllabus Change Policy

Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advance notice.

## Tentative Course Outline

The weekly coverage might change as it depends on the progress of the class. “HW” is an abbreviation for “Homework.” These take-home assignments will generally be made available each Monday and are due at 11:59pm on the following Sunday evening. Please check Canvas for the exact due dates.

<b>Tentative Course Schedule</b>				
<b>Week</b>	<b>Date</b>	<b>Day</b>	<b>Topic</b>	<b>Due</b>
1 - Introduction	03/21	Tuesday	Introduction, software review, & setup	None
	03/23	Thursday	Setup, GitHub basics	
2 - Python Basics	03/28	Tuesday	Data types	HW 1
	03/30	Thursday	Logic control statements & loops	
3 - Functions & Classes	04/04	Tuesday	Functions & lambdas	HW 2
	04/06	Thursday	Classes & methods	
4 - More Functions....	04/11	Tuesday	More functions & classes	HW 3
	04/13	Thursday	More functions & classes: creating a game	
5 - The Pandas Dataframe	04/18	Tuesday	Pandas I	HW 4
	04/20	Thursday	Pandas II	
6 - More Pandas	04/25	Tuesday	Pandas III	HW 5
	04/27	Thursday	Pandas IV	
7 - Data Visualization	05/02	Tuesday	Matplotlib I	HW 6
	05/04	Thursday	Matplotlib II	
8 - Web Scraping	05/09	Tuesday	Using data APIs, introduction to HTML	HW 7
	05/11	Thursday	Requests and BeautifulSoup	
9 - Advanced Topics	05/16	Tuesday	NumPy and Statsmodels	None
	05/18	Thursday	Data transformations & missing values	