# DATA 11800: Introduction to Data Science

## Instructor: Dr. Amanda R. Kube Jotte

### Autumn 2023

E-mail: akube@uchicago.edu

Office Hours: WTh 1:00-2:00 pm in Ryerson 256

Location: Ryerson Phys Lab 251

Class Meetings: MWF 9:30-10:20 am

---

## Course Description

Data has become increasingly available in recent years, not just to scientists but to the average citizen as well. With this information at our fingertips, there is increased need for training in how to use and understand this data. In this course, we will grapple with how to use data to study the world around us. We will learn how to collect and store data, use Python and Jupyter Notebooks to explore data, and study methods for making inferences about that data. In addition to learning to manipulate and test data, together we will discuss when and how to use data and whether data is always neutral. How might data exacerbate existing biases? When is the use of data appropriate? How can we be more critical consumers of the data that is constantly being presented to us? We will learn to see data in our everyday lives and work with datasets that are exciting and relevant to ourselves and our community.

## Course Objectives

1. Introduce students to the data science pipeline.

2. Develop students' abilities to be informed and critical readers of quantitative, data-based arguments.

3. Enable students to perform data analysis using Python.

4. Develop students' abilities to integrate code, background information, results and interpretation in a Jupyter Notebook to communicate data analyses effectively.

5. Help students gain flexible problem-solving and programming skills applicable to a large variety of problems independently.

6. Understand foundational concepts of probability and statistics.

7. Develop skills to apply probability and statistics for data analysis.

# Student Learning Outcomes

1. Students will be able to code and plot in Python using Jupyter Notebooks and packages such as NumPy, Pandas, and Matplotlib.

2. Students will be able to think critically about the use and collection of data.

3. Students will be able to clean, filter, group, and visualize datasets and use these methods to better understand and explain their data.

4. Students will be able to use simulation to study difficult mathematical or computational problems.

5. Students will gain an understanding of statistical methods including hypothesis testing, confidence intervals, and bootstrap.

6. Students will be able to use concepts of statistical inference to engage with research questions.

7. Students will be informed and critical readers of quantitative data-based arguments.

# Course Structure

The goal of this course is to introduce students to all aspects of a data analysis process, help students to be informed and critical readers of quantitative data-based arguments, provide skills that enable them to perform data analysis using Python and integrate code, background information, results and interpretation in a Jupyter Notebook and to help students gain flexible problem-solving skills applicable to a large variety of problems.

This course will cover topics such as basic python coding, data manipulation, and statistical inference, and more. These topics will be taught and reinforced through lectures, coding activities, labs, homework assignments, a midterm project and a final exam.

**You need NO prior programming or statistics experience to take this course**, although we expect you to be able to pick up a lot of things on your own through labs and assignments. If you are a CS or Stats major this course is NOT for you as it will likely be covering many topics you have already seen.

This course can be taken standalone, but is only an introduction. You should plan on also taking 11900 next quarter. This is also the gateway course for the Data Science minor and major. It should give you enough context to pick up other tools later on your own.

# Course Policies

## Course Materials and Announcements

**Textbooks** (both freely available):

- Intro to Data Science for 11800 and 11900
  https://ds1.datascience.uchicago.edu/intro.html

- (Optional) Python Data Science Handbook by Jake VanderPlas, O'Reilly Media
  https://jakevdp.github.io/PythonDataScienceHandbook/

The Intro to Data Science for 11800 and 11900 textbook was written by the team of Data Science instructors here at UChicago. This is the first release and chapters for DATA 119 are still under revision. We welcome your comments! Please add any typos found or any constructive comments to the Textbook discussion thread on Canvas.

**Software:**
You will need the ability to work with Jupyter Notebooks to complete assignments and view lectures for this course, so you must have access to a computer with Python3 and Jupyter Notebooks installed or access to Google Collab to open and edit the .ipynb files.

**Discussion:**
We will use Ed and Canvas for all questions and discussions related to the class. Please post questions on Ed rather than sending an email. This serves multiple purposes. First, others may have the same question. Posting to Ed allows us to clarify the issue for everyone at the same time. Second, we are much more likely to respond in a timely manner if you post on Ed as both instructors and TAs will see the post! All announcements related to the class will be made in class, on Canvas, or on Ed. **I will assume that any announcement made on Canvas or on Ed is known to everyone in class within one business day of it being posted.** It is important to check Ed and Canvas regularly! **You are responsible for all announcements made in lecture or online.**

**Labs**

There will be weekly lab assignments posted to Canvas. These labs are not required, but they are your main source of practice for concepts we cover in class. Not completing your lab will not have a negative effect on your grade, however **completing and submitting a week's lab session will earn you extra credit.** If you complete all labs, you will earn an extra 4% (about a half letter grade) toward your final grade. Completing any portion of the labs will earn you that portion of the extra 4% (ie if you complete half of the labs you will earn 2% or half of the extra credit). We encourage you to complete your lab during one of the weekly TA office hours so that you can work together with peers and get help from the TAs when you need.

**Grading Policy**

Your course grade will be calculated as follows:

- Homework - 30%

- Midterm Project - 30%

- Final Exam - 40%

Please submit all assignments on time! You will want feedback on your work before completing the next assignments as most topics in this course build on one another. For this reason, **late assignments will not be accepted.** However, I welcome you to talk to me or to ask for extensions

if you are having issues. Because I do not allow late assignments, **I will drop your lowest homework grade.** This is meant to account for any unforeseen issues in the submission process or other problems you may encounter. It is best if you save this dropped submission for situations like this! I will not drop additional homework grades.

## Use of Generative AI

You will not be allowed to use ChatGPT, Google Bard, or any similar large language models on your homework assignments. Doing so will be treated as a violation of academic integrity. The reason for this, is that it is important you learn to understand and write code on your own in order to be able to properly use such resources in the future.

However, I will allow you to use these on the midterm if you would like. You are allowed to use outside sources in general on your midterm, but traditional rules of writing apply: ie you need to cite your sources including ChatGPT if you do use it and you should not copy and paste straight from those sources. This is a good opportunity for you to learn how to best utilize models like ChatGPT for this kind of work.

## Topics to be Covered

Topics we will discuss throughout the quarter will include but are not limited to:

- Programming in Python and Jupyter Notebooks

- Data Structures

- Visualization

- Data Collection

- Probability

- Simulation

- Randomness and Sampling

- Statistical Inference

- Hypothesis Testing

A more detailed calendar will be posted to Canvas and updated as needed. This calendar is subject to change.

## Academic Integrity

Acting with academic integrity means, in brief, not submitting the statements, work, or ideas of others as one's own. Students are expected to comply with University regulations regarding honest work. If you are in doubt about what constitutes academic dishonesty, speak with me before the assignment is due. Failure to maintain academic integrity on an assignment will result in a penalty befitting the violation, up to and including failing the course and further University sanctions. For more information, consult the student manual [https://studentmanual.uchicago.edu/academic-policies/academic-honesty-plagiarism/](https://studentmanual.uchicago.edu/academic-policies/academic-honesty-plagiarism/).

## Accommodations

**Accessibility:** Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please meet with me to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations. For more information, visit disabilities.uchicago.edu.

**Accommodations based upon sexual assault:** The University is committed to offering reasonable academic accommodations to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If a student comes to us to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if we otherwise observe or become aware of such an allegation, we will keep the information as private as we can, but as faculty members of University of Chicago, we are required to immediately report it to a Department Chair or Dean or directly to the University's Title IX Coordinator. If you would like to speak with the Title IX Coordinator directly, Bridget Collier can be reached at bcollier@uchicago.edu or 773-834-6367. Additionally, you can report incidents or complaints to the Sexual Assault Dean-on-Call (SADoC) by calling 773-834-HELP, or by contacting UCPD at (773)702-8181 or your local law enforcement agency. See https://studentmanual.uchicago.edu/university-policies/the-university-of-chicago-policy-on-title-ix-sexual-harassment/.

**Bias Reporting:** The University has a process through which students, faculty, staff and community members who have experienced or witnessed incidents of bias, prejudice or discrimination against a student can report their experiences to the University's Bias Education and Support (BEST) team. See: https://diversityandinclusion.uchicago.edu/resources/reporting-incidents/

**Mental Health:** Student Wellness' Mental Health professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See: https://wellness.uchicago.edu/mental-health/

**Preferred Name and Gender Inclusive Pronouns:** In order to affirm each person's gender identity and lived experiences, it is important that we check in with others about pronouns. This simple effort can make a profound difference in a person's experience of safety, respect, and support. See: https://inclusion.uchicago.edu/lgbtq-student-life/lgbtq-resources/