

YData: An Introduction to Data Science

Tentative Course Syllabus

Instructor: Shivam Sharma

1 Information

- 2025 Yale Summer Session B (SB25): *June 30 - August 1, 2025*
- Course number: S&DS S123 01/ CPSC S123 01/ PLSC S351 01
- Class Meeting Time: Tuesdays & Thursdays 9 AM - 12:15 PM
- Location: TBD

2 Course Description

Computational and statistical skills are increasingly important in our data-driven world. These skills are key to opening doors for research and career opportunities during and beyond your years at Yale. This course aims to enhance your knowledge and capabilities in fundamental ideas and skills in data science, especially computational and programming skills along with inferential thinking. YData is an introduction to Data Science that will emphasize the development of these skills while providing opportunities for hands-on experience and practice.

YData is designed to be accessible to students with little or no background in computing, programming, or statistics. The course is based on the Python programming language.

3 Course Format

The class will meet two times a week. During class meetings new material will be presented, and there will be opportunities to work through exercises related to the class content.

4 Enrollment Cap & Prerequisites

There is no enrollment cap on this course and there are no prerequisites. So *everyone is welcome to take the class*. Instructor permission is not needed for this class, so feel free to sign up on your own.

5 Required Course Materials

There is no required textbook for this course that you need to purchase. Weekly reading on data science analyses will be posted to Canvas.

Readings will also be taken from the following (related) online textbooks:

- Adhikari A, DeNero J (2022). [Computational and Inferential Thinking: The Foundations of Data Science - 2nd Edition](#)
- Spirling A, Jones-Rooy A (2022). [Data Science for Everyone: course text](#)
- Brett M (2020). [Data science for everyone](#)

The following online books will also be useful resources:

- Downey (2024). [Elements of Data Science](#)
- McKinney (2022). [Python for Data Analysis, 3E](#)

6 Assignments and Grading

Your grade in the class will be a weighted average of the following components:

- 9 Homework Assignments (40%)
- Final Project (15%)
- Midterm Exam (15%)
- Final Exam (25%)
- Attendance & Answering Questions on Ed Discussions (5%)

More details:

1. **Homework:** The largest component of work for this class is the homework assignments (9 total). These assignments will involve answering questions about concepts and methods discussed in class, and analyzing data in Python. You will have (tentatively) two homework assignments every week. Homework assignments will be available after classes on Tuesdays and Thursdays. They will be due at 11 pm the next day. Lowest homework grade will be dropped at the end of the semester. Late Homework will only be accepted with a Dean's Excuse.

2. **Final Project:** The class final project consists of analyzing a data set you find on your own using methods discussed in the class.
3. **Exams:** The midterm exam will be taken during class (60 minutes exam, plus 10 minutes to turn in the exam) and the final exam will be during the last class (2.5 hours exam, plus 30 minutes to turn in the exam).
4. **Attendance & Answering Questions on Ed Discussions:** Summer semester is extremely fast paced and it is important that you attend every class. Attendance will be taken (twice) during each class. Additionally, you should participate and answer questions on Ed Discussions.

Regrade requests: Although we strive for consistency and accuracy in grading, we understand that grading errors can occur. If you notice a mistake in the grading you can submit a regrade request on Gradescope. Please note the following rules surrounding regrade requests:

- Regrade requests must be submitted within one week after the assignments have been graded.
- Regrade requests must be accompanied by a written statement carefully highlighting and explaining the items that were misgraded.

7 Honor Code

You are encouraged to be helpful to your classmates and to work together, **but the work you turn in must be your own**. Any student who turns in work for credit that is identical, or similar beyond coincidence, to that of another student may face appropriate disciplinary action at the department, college, or university level. *Cheating and/or plagiarism will not be tolerated.*

If you get ideas or words from a website, journal article, book, another person, etc., cite the source in your work at the location where you use the idea. Then include a bibliography or list of sources cited at the end of your document.

The use of chatGPT and other LLMs: You can use chatGPT (or other LLMs) to answer general questions about Python and other general concepts. For example, it is fine to ask chatGPT “What does the np.sum() function in Python do?” **You are not allowed to use it to answer homework questions.** For example, it is **NOT Okay** to cut and paste a homework question into chatGPT. Violations of this policy will be treated as incidence of academic dishonesty and will face disciplinary action.

8 Course Schedule

Tentative Course Schedule	
Date	Topics
July 1 (Tuesday)	<ul style="list-style-type: none">- Class overview and what is Data Science?- Introduction to Python
July 3 (Thursday)	<ul style="list-style-type: none">- Descriptive statistics and plots- Array computations (Part 1)
July 8 (Tuesday)	<ul style="list-style-type: none">- Array computations (Part 2)- Tables and data manipulation (Part 1)
July 10 (Thursday)	<ul style="list-style-type: none">- Midterm Exam- Tables and data manipulation (Part 2)
July 15 (Tuesday)	<ul style="list-style-type: none">- Data visualization- Mapping and Geospatial data
July 17 (Thursday)	<ul style="list-style-type: none">- For Loops and Writing functions in Python- Intro to Statistical Inference
July 22 (Tuesday)	<ul style="list-style-type: none">- Hypothesis tests and confidence intervals- Machine learning: Supervised learning
July 24 (Thursday)	<ul style="list-style-type: none">- Machine learning: Unsupervised learning- Ethics
July 29 (Tuesday)	<ul style="list-style-type: none">- Web Scraping- GitHub Website Hosting- Large Language Models (LLMs)
July 31 (Thursday)	<ul style="list-style-type: none">- Final Exam