

Syllabus

CSci 39542: Introduction to Data Science

Department of Computer Science

Hunter College, City University of New York

Fall 2021

Description: *3 hours, 3 credits:* This topics course focuses on computational methods and statistical techniques to analyze data and make inferences. Topics include data collection and cleaning, exploratory data analysis and visualization, and statistical inference and prediction. Students will acquire a working knowledge of data science through hands-on projects with real-world data. Basic proficiency in statistics and Python programming is assumed, as well as experience with matrix algebra or abstract data structures.

Prerequisites: CSci 127, Stat 213, and one of: Math 160, Math 260, or CSci 335.

Instructor: [Dr. Katherine St. John](#), professor ([office hours](#)).

Tech-in-Residence Fellow: [Susan Sun](#).

Meetings: Mondays, Thursdays, 2:45-4pm, on-line.

Grading Policy

Course Format: This course is taught in an **on-line** format with all lectures, quizzes, and exams given on-line:

- Synchronous lectures are Mondays and Thursdays, 2:45-4pm,
- Asynchronous assessments (usually due 24 hours after lecture, depending on holidays),
- Individual programming assignments and cumulative project, and
- 2 hour final exam during the time slot assigned by the registrar.

Expectations: Completing homework is an essential part of the learning experience. Students are expected to learn both the material covered in class and the material in the textbook and other assigned reading.

Honor Code: You are encouraged to work together on the overall design of the programs and homework. However, for specific programs and homework assignments, all work must be your own. You are responsible for knowing and following Hunter College's [Academic Integrity Policy](#):

Hunter College regards acts of academic dishonesty (e.g., plagiarism, cheating on examinations, obtaining unfair advantage, and falsification of records and official documents) as serious offenses against the values of intellectual honesty. The College is committed to enforcing the CUNY Policy on Academic Integrity and will pursue cases of academic dishonesty according to the Hunter College Academic Integrity Procedures.

All incidents of cheating will be reported to the Office of Student Conduct in the Vice President for Student Affairs and Dean of Students office.

Lecture Participation: Participation in lecture is measured by attendance and related questions on that day's quiz. There are 28 lectures, and the highest 25 participation grades are counted toward your final grade (that is, the lowest three grades are dropped).

Quizzes: After every class, there will be an quiz on the lecture notes, code demonstrations, classwork, and submitted programs.

- As the semester progresses, quizzes will include review questions as well as short programming exercises based on the homework.
- Quizzes are usually due 24 hours after lecture.
- There are no make-up quizzes. Instead, your score on the final exam will replace missing quiz grades (the final exam will also replace a quiz grade when you take the quiz but do better on the final exam).

Homework: Programming exercises are posted on the class website, usually three weeks before the due date. They reinforce concepts covered in lecture and lab and serve as building blocks for the classwork and semester-long project. To receive full credit for a program, the program must perform correctly, must include comments, be written in good style, and be submitted via [gradescope](#). You can miss up to 5 programming assignments without affecting your grade (if you turn in all the programming assignments, we will drop the lowest 5 scores). **No late homework is accepted.**

Project: A final project is required for this course. The grade for the project is a combination of grades earned on the milestones (e.g. deadlines during the semester to keep the projects on track) and the overall submitted program.

Final Exam: The final exam is required. It is comprehensive, covering all the material of the course. Sample exam questions will available on the course webpage. **You must take and pass the final to pass the course.**

Grades: The grading for the course will be based on:

- Participation: 10%.
- Homework: 25%.
- Project: 25%.
- Quizzes: up to 20% (if you miss a quiz, or do better on the final exam, your grade on the final will replace that quiz grade).
- Final Exam: 20% (up to 40% if you do better on the final exam than on quizzes).

Materials, Resources and Accommodating Disabilities

This is a *zero cost course*. All textbook materials are freely available to enrolled students.

Textbook & Readings: The following free on-line books are required for the course:

- UC Berkeley's Data Science 8 (DS 8): [The Foundations of Data Science](#) By Ani Adhikari and John DeNero.
- UC Berkeley's Data Science 100 (DS 100): [Principles and Techniques of Data Science](#) By Sam Lau, Joey Gonzalez, and Deb Nolan.

Additional readings and tutorials are available on the [resources](#) page.

Technology: This is a programming-intensive course in the Python programming language. See the [resources](#) page for obtaining the Python program and the packages used, links for the submitting assignments and assessments. All software used is freely available.

Computer Access: A computer is needed to attend lecture, complete the on-line assessments, and programming assignments and projects. Hunter College is committed to all students having the technology needed for their courses. If you are in need of technology, see [Student Life's Support & Resources Page](#).

Accommodating Disabilities: In compliance with the American Disability Act of 1990 (ADA) and with Section 504 of the Rehabilitation Act of 1973, Hunter College is committed to ensuring educational parity and accommodations for all students with documented disabilities and/or medical conditions. It is recommended that all students with documented disabilities (Emotional, Medical, Physical, and/or Learning) consult the [Office of AccessABILITY](#). For further information and assistance, see their [contact page](#).

Hunter College Policy on Sexual Misconduct: In compliance with the CUNY Policy on Sexual Misconduct, Hunter College reaffirms the prohibition of any sexual misconduct, which includes sexual violence, sexual harassment, and gender-based harassment retaliation against students, employees, or visitors, as well as certain intimate relationships. Students who have experienced any form of sexual violence on or off campus (including CUNY-sponsored trips and events) are entitled to the rights outlined in the Bill of Rights for Hunter College.

- **Sexual Violence:** Students are strongly encouraged to immediately report the incident by calling 911, contacting NYPD Special Victims Division Hotline (646-610-7272) or their local police precinct, or contacting the College's Public Safety Office (212-772-4444).
- **All Other Forms of Sexual Misconduct:** Students are also encouraged to contact the College's Title IX Campus Coordinator, Dean John Rose (jtrose@hunter.cuny.edu or 212-650-3262) or Colleen Barry (colleen.barry@hunter.cuny.edu or 212-772-4534) and seek complimentary services through the Counseling and Wellness Services Office, Hunter East 1123.

See [CUNY Policy on Sexual Misconduct Link](#).

Course Objectives

At the end of the course, students should be able to:

1. Acquire data sets from multiple sources and write programs that can extract (scrape) the data into a usable form.
2. Use data mining to extract new insights about the data.
3. Understand basic storage techniques and constraints.
4. Analyze data using standard techniques from statistics and linear algebra.
5. Visualize data using popular Python modules.