Introduction to Data Science (Fall 2025)

CS 1090a, AC 2090a, Stat 109a, or CSCI E-109a

Instructors

- Pavlos Protopapas (SEAS)
- Kevin Rader (Statistics)

Preceptor

- Chris Gumb (SEAS)

Meeting Times

- **Lectures:** Mon & Wed 10:30am-11:45am (Room TBD)
- **Sections:** See registration for times (Rooms TBD)
- **Office Hours:** TBD

Course Introduction

Welcome to **CS1090a/AC209a/STAT109a**, also offered by the DCE as **CSCI E-109a**, Introduction to Data Science. The course will focus on the analysis of messy, real-life data to perform predictions and inferences using machine learning and statistical methods.

Material covered will integrate four key facets of an investigation using data:

1. **Data wrangling** – web scraping, data cleaning
2. **Exploratory data analysis** – generating hypotheses and building intuition
3. **Prediction, inference, or statistical learning**
4. **Communication** – justifying decisions and analyzing results

This course is the first in a two-part series. In the spring semester, the curriculum builds upon the content of the fall course, diving deeper into unsupervised learning, deep neural networks for computer vision and language modeling, transformers, and generative models. Students are strongly encouraged to enroll in both the fall and spring courses within the same academic year.

This course fulfills the Quantitative Reasoning with Data (QRD) requirement, with the condition that for the Class of 2029 and later, it must be taken for a letter grade to fulfill this requirement.

Prerequisites

A foundational knowledge of Python programming is required for this course. Specifically, students should be comfortable with the following Python concepts:

- Conditionals, loops, and data structures (e.g., lists, dictionaries, etc.)
- Functions
- File I/O and string parsing
- Classes, methods, attributes, and general OOP principles

Additionally, students should possess an understanding of the following:

- Random variables and common probability distributions (e.g., normal and binomial)

- Basic concepts in probability (e.g., independence, joint and conditional probabilities)
- Calculus at an introductory level (*Note: Multivariable calculus is not required*)

If you are unsure about your preparation, HW0 will help you assess readiness for the course.

Topics

- Web scraping
- Manipulating tabular data: Pandas
- Exploratory data analysis (EDA)
- kNN & linear regression
- Multiple & polynomial regression
- Model selection & cross-validation
- Regularization (LASSO and Ridge)
- Maximum likelihood estimation (MLE)
- Bootstrap, confidence intervals, & hypothesis testing
- Bayesian inference, sampling, and hierarchical models
- High dimensionality & principal component analysis (PCA)
- Causal inference
- Missing data & imputation
- Classification & logistic regression
- Decision trees
- Ensemble methods:
  - Bagging
  - Random forests
  - Boosting
  - Mixture of experts
  - Meta-models and model stacking

Differences Between CS1090A & AC2090A

- **Readings:** There may be further readings assigned to 209A students.
- **Assignments:** Homework assignments for 209A students may have *additional* components.
- **Exams & Coding Assessments:** The exam questions will differ slightly between 1090A and 2090A students.
- **Projects:** Project groups with one or more 209A students will perform some self-directed outside reading to inform their projects. The goal is to make use of some approach or method not explicitly covered in class and communicate an understanding of the method(s) and its applicability to the problem of focus through the final report and presentation.

Course Components

Lectures will be live-streamed for Extension School students and can be accessed through the 'Zoom' section on Canvas.
Recordings will be made available to all registered students within 24 hours and can be accessed through the 'Course Videos' section on Canvas.

Lecture

The class meets for lectures twice a week (Mon & Wed). Attending and participating in lectures is a crucial component of learning the material presented in this course. Students may be asked to complete short

readings before certain lectures.

On-campus students are expected to attend all lectures and all of their assigned sections. See the 'Grades' section below for information on how attendance relates to students' grades.

For every 4 classes attended, on-campus students earn a late day. See the homework section below for more information on late days.

> **Note:** This policy does **not** apply to **DCE (Extension School)** students.

Readings To help you get the most out of lectures, we will provide short **pre-readings** to familiarize you with key concepts before class. We **strongly recommend** reviewing these ahead of time, as they offer valuable context for the material we'll cover.

Alongside the pre-readings, we include short **self-assessment quizzes**. These are not graded and are purely for your own learning—to help you check your understanding and highlight important ideas.

After class, we may share **post-readings** for those who want to explore topics in more depth. These are completely optional and not required for the course.

Overall, while none of the readings are graded, engaging with them will make lectures more accessible and meaningful.

Quizzes As opposed to previous iterations of the course, we will **not** have post-lecture quizzes this year. Instead, students will take **two in-section graded quizzes**, each lasting 30 minutes. These will focus on concepts introduced in lecture.

- **Quiz 1:** Week of September 28
- **Quiz 2:** Week of November 16

Students will take the quiz during their **assigned section time** that week. These quizzes are designed to evaluate your understanding of core material and prepare you for the midterm and final exams.

DCE students will complete section quizzes in their assigned online sections.

Homework

- **HW0** will be released in July and is due on September 9. It serves to evaluate students' preparedness for the course, covering the prerequisites listed above. Students finding this material too challenging should consider registering for the course in a future semester, after having spent more time with the prerequisites.
- After HW0, there will be 5 homework assignments, each due two weeks after being assigned.
- For these assignments, students may collaborate and submit their work in pairs (see the collaboration policy below).

Late policy details are below.

Exams

The course will include two exams: a **midterm** and a **final**. Both exams will consist of two components:

1. A **paper** component of **multiple-choice** and **open-ended questions** designed to assess conceptual understanding, reasoning, and practical skills.

2. A **coding** component to varify students proficiency in implenting course concepts in Python. These coding components will include test cases students can run repeatedly to confirm their implementation is successful.

The paper exams will be **closed-book**, but students may bring **a limited number of pages of personal notes** (details to be announced closer to each exam).

- **Midterm:** Paper component held in section during the week of **October 20.** Coding portion will be a take-home assignment.
  - Paper Duration: 75 minutes
  - Coding Duration (at home): TBD
- **Final Exam:** During the official exam period (exact date and time **TBD**; entire exam time ~3 hours)
  - Paper Duration: Perhaps slightly longer than the paper midterm (details TBA)
  - Coding Duration (in-person after paper component): TBD

Further details, including permitted materials and note-sheet guidelines, will be announced as the dates approach.

For remote/DCE students, proctoring differs and we will provide details.

Missed Section Quiz or Exam Policy

If you are unable to attend your **regular section** during a quiz or midterm week, please notify the course staff **at least 48 hours in advance**. Since sections are held throughout the week, we can usually arrange for you to take the quiz or midterm in another section.

*Makeup options outside of regularly scheduled section times are **not guaranteed** and will only be considered in cases of documented emergencies or serious illness. We strongly encourage you to plan ahead and coordinate early if a scheduling issue arises.*

If a conflict or illness arises suddenly, contact staff ASAP and accommodations will be considered as needed.

Project

Students may propose a project topic or adapt an example provided by the teaching staff. Students self-assemble into groups and submit their preferences over the approved set of project topics. Groups are assigned a project and a mentor from the teaching staff who will provide guidance and feedback along the remaining milestones.

Milestones

0. Project proposals
1. Preferences & group formation
2. Data check, refined problem statement, and course-of-action
3. Data processing, EDA, and baseline analysis
4. Minimum viable project
5. Final Presentations & Report

Details regarding each of the milestones will be released during the first few weeks of the course. A rubric will be made available outlining expectations for the final presentation and report.

Grading

Grade Breakdown

Final grades for the course will be computed using the following weights:

Assignment	Final Grade Weight
Homework 0	1%
Homework 1–5	29%
Section Quizzes (2 total)	10%
Midterm	18%
Final Exam	22%
Project: Milestones 1–4	5%
Project: Presentation Video	5%
Project: Final Report	10%
Total	**100%**

Attendance

Attendance is **required for all on-campus students** and will be recorded throughout the semester. To be eligible for certain letter grades, students must meet the following minimum attendance requirements (lectures and sections combined):

- **A** requires at least **66% attendance**
- **A–** requires at least **50% attendance**
- **B+** requires at least **33% attendance**

Students below these thresholds may still pass the course but are **not eligible** for the corresponding higher letter grades, regardless of numerical performance.

For every **4 sessions attended** (lecture or section), students earn **1 late day**, which can be used on homework (up to 2 per assignment).

DCE students are exempt from attendance tracking, as they engage with the course asynchronously. However, active participation through lectures, quizzes, and assignments remains essential for success.**DCE students are automatically granted 4 late days** at the beginning of the semester.

> **Note:** Any effort to falsify attendance records will be treated as an academic integrity violation and reported accordingly.

Grade Definitions

This course uses the grading framework outlined in the FAS Student Handbook as a general reference. The following descriptions offer high-level expectations for each grade category:

- **A, A–:** Awarded for work of excellent quality that demonstrates full mastery of the subject. A grade of A typically reflects work of extraordinary distinction.
- **B+, B, B–:** Represents a solid understanding of course material, strong command of required skills, and full engagement with course activities.
- **C+, C, C–:** Indicates adequate comprehension of material and fulfillment of basic course requirements.
- **D+, D, D–:** Denotes minimal understanding and participation, but sufficient to merit course credit.
- **E:** Given for work that fails to meet minimum standards for course credit.

These grade descriptions provide a broad guide. Throughout the semester, we will share more detailed learning objectives, performance expectations, and grading rubrics for each assignment, exam, and project milestone.

Resources

Recommended Textbook

The book for the course is *An Introduction to Statistical Learning* and is available at statlearning.com.

Software

Our primary software tools will be Python 3, Jupyter notebooks, and various 3rd party Python libraries. A set-up guide will be included in the to-be-released HW0. SEAS also provides the FASOnDemand service which is accessible through Canvas. This is a remote Python environment where students can run Jupyter notebooks. The environment has all packages used in the course pre-installed.

Late Work Policy

Extension School Late Days **Extension School** students are allocated a maximum of **6 late days** with **at most 2 days applied to any single homework**.

On-Campus Students Late Days **On-campus students** begin the semester with **0 late days** but can acquire them through attendance (see attendance policy below). **At most 2 late days can be applied to any single homework**.

General Late Day Policies There is no need to ask in advance before using one of your late days.

If a student has **exhausted all their late days**, **late homework will not be accepted**. The late day policy is designed to provide maximum flexibility up front and is intended to cover common disruptions such as illness, travel, or busy schedules. As such, additional late days will **not** be granted for these reasons.

In cases of **serious or extended circumstances** (e.g., prolonged illness, family emergency), students should contact the course staff as soon as possible. In such cases, we may consider alternate arrangements, but please note that extended absences may make it difficult to meet the learning objectives of the course within the semester.

Late days **cannot** be applied to HW0, quizzes, in-section assessments, or project milestones.

If you have a medical or personal emergency, contact us as soon as possible—do not wait until the deadline passes.

Attendance Policy

Attendance at lectures and sections is **required for all on-campus students**. The teaching staff will record on-campus attendance. For every 4 sessions attended (i.e., lecture or section), on-campus students will earn 1 additional late day. **Any effort to misrepresent attendance will be considered a violation of the honor code and be dealt with accordingly.**

See the note on attendance in the "Grading" section for how attendance can impact final grades.

Academic Integrity

We expect you to adhere to the Harvard Honor Code at all times. Failure to adhere to the honor code and our policies may result in serious penalties, up to and including automatic failure in the course and reference to the ad board.

DCE Academic Integrity Policy

If you are an Extension School student, you are responsible for understanding Harvard Extension School policies on academic integrity here and how to use sources responsibly. Stated most broadly, academic integrity means that all course work submitted, whether a draft or a final version of a paper, project, take-home exam, online exam, computer program, oral presentation, or lab report, must be your own words and ideas, or the sources must be clearly acknowledged. The potential outcomes for violations of academic integrity are serious and ordinarily include all of the following: required withdrawal (RQ), which means a failing grade in the course (with no refund), the suspension of registration privileges, and a notation on your transcript.

Using sources responsibly here is an essential part of your Harvard education. We provide additional information about our expectations regarding academic integrity on our website. We invite you to review that information and to check your understanding of academic citation rules by completing two free online 15-minute tutorials that are also available on our site. (The tutorials are anonymous open-learning tools.)

Student Collaboration

If you work with a partner on an assignment, make sure both parties solve all the problems. Do not divide and conquer. You are expected to be intellectually honest and give credit where credit is due. In particular:

- If you work with a fellow student and want to submit the same notebook, you need to form a group prior to the submission. Details in the assignment. Not all assignments will permit group submissions.
- You need to write your solutions entirely on your own or with your collaborator.
- If you worked with a fellow student on a paired assignment but decide in the end to submit different notebooks individually, include the name of the other student as a comment at the top of your notebook.
- You are welcome to take ideas from code presented in lecture or sections, but you will need to change it, adapt it to your style, and ultimately write your own. Simply copying verbatim will rarely be successful.
- If you use code found on the internet, books, or other sources, you need to cite those sources.
- You should not view any written materials or code created by other students for the same assignment.
- You may not provide or make available solutions to individuals who take or may take this course in the future.
  - If you are using a remote git repository such as GitHub to work on your assignments **you must make it private.**

Note: If you're unsure about what constitutes fair collaboration, please ask the instructors.

Use of AI Models

Purpose of Policy: This policy outlines the acceptable use of AI models, including but not limited to ChatGPT, in completing assignments for this course.

Policy Guidelines:

1. **Original Work:** Students are expected to complete assignments using their original thoughts and interpretations. AI models can be used to help understand concepts, generate ideas, or learn about different perspectives, but they should not write or complete assignments for students.
2. **Collaboration with AI:** Students may use AI models for brainstorming or generating preliminary ideas, but the final work submitted must be substantially their own. Students should be able to explain their reasoning, logic, and conclusions without relying on the model's output.

3. **Restrictions for Specific Assignments:** There may be specific assignments (e.g. quiz part of the midterms) or parts of the course where the use of AI models is entirely prohibited. These restrictions will be clearly stated in the assignment guidelines.
4. **Ethical Considerations:** Students are encouraged to approach the use of AI with ethical considerations in mind, including issues related to privacy, bias, and authenticity.

Consequences for Non-Compliance: Failure to adhere to this policy may result in academic penalties as outlined in the course's academic integrity policy.

Questions and Clarifications: If students have questions about the appropriate use of AI models in an assignment, they should consult the course instructor or teaching assistants before proceeding.

Please refer to the University's policy for further information.

Accommodations for Students with Disabilities

Harvard students needing academic adjustments or accommodations because of a documented disability must present their Faculty Letter from the Accessible Education Office (AEO) and speak with the professor by the end of the second week of the term, (fill in specific date).

Auditing

To request permission to audit the course, please email **cs1090a2025@gmail.com** with the following:

- Your **Harvard University ID (HUID)**
- A brief **statement of agreement** to the auditing terms outlined below

Important:
You must **not** be currently enrolled in the course when making this request. If you are listed as an enrolled student in Canvas, you must drop the course before you can be added as an auditor.

Auditors must agree to the following policies:

- **In-person attendance is mandatory.** This is a Harvard-wide policy. Auditors who do not confirm their presence during the **first week** of in-class instruction will be removed from the course.
- **Academic honesty standards apply.** Auditors are expected to follow the same academic integrity rules as enrolled students. Sharing or distributing homework or solutions is strictly prohibited and will be reported to the Harvard Administrative Board.
- **No future credit.** Auditors may **not** take this course for credit in the future.
- **No submissions.** Auditors should **not submit homework, midterms, or participate in course projects**.
- **Limited access to course resources.** Auditors may not use resources intended for registered students, including **Ed Discussion, FASOnDemand, or TF office hours**.

Frequently Asked Questions (FAQ)

Attendance

Q: What counts as "attendance"? Is there a grace period if I arrive late or leave early?
A: Attendance means being physically present for the majority of the session. Arriving more than 15 minutes late or leaving more than 15 minutes early may result in being marked absent for that day.

Q: Do I have to attend my assigned section, or can I attend any section?
A: You must attend your assigned section unless you have prior approval to switch. This is important for accurate attendance tracking and balanced participation.

Q: If I attend 7 sessions, do I get 1 or 2 late days?
A: You earn 1 late day for each full set of 4 sessions attended (e.g., 4 sessions = 1 late day, 8 sessions = 2 late days). Partial sets do not count for extra late days.

Q: What if I have a documented medical or personal emergency and miss a session or quiz?
A: Our attendance policy is intentionally very flexible: you can miss up to 33% of all sessions (lectures and sections) with no penalty. This allowance is designed to cover routine illness, interviews, minor emergencies, and other unavoidable conflicts—so you do not need to provide documentation for individual absences.
If you experience a serious or prolonged emergency (e.g., extended illness, major family event) that may cause you to fall below the minimum attendance required for your target grade, contact the course staff as soon as possible. We will review such situations individually, but keep in mind that extended absences may make it difficult to meet course learning objectives.

Quizzes and Exams

Q: What if I miss a section quiz or exam due to illness, religious observance, or another conflict?
A: Notify staff at least 48 hours in advance if possible. We'll try to arrange for you to take the quiz or exam in another section. Sudden illness or emergencies should be reported as soon as possible.

Q: Will I get the same quiz as others if I take it in a different section?
A: We vary quiz content between sections and monitor for academic integrity. Sharing or discussing content is an honor code violation.

Q: How many pages of notes can I bring to the exam? What about font size?
A: Specific note-sheet policies will be announced before each exam, including page count and formatting. Details will be enforced.

Q: How are exams proctored for DCE students?
A: DCE students will follow specific online proctoring instructions. Academic integrity applies to everyone.

Homework and Collaboration

Q: How much can I borrow from online sources or classmates?
A: Consult public sources and lecture code, but write your own work. Anything you borrow must be cited. Copying or sharing solutions is not allowed.

Use of AI Models

Q: Can I use ChatGPT or other AI tools to help with homework?
A: You may use AI tools for ideas or to clarify concepts, but your submission must be in your own words and show your understanding. You must be able to explain everything you submit.

Q: What does "substantially your own work" mean?
A: Your final answer should not be mostly generated or paraphrased by AI. AI is for brainstorming, not for writing your main answer.

Q: How will you check for AI or code copying?
A: We use plagiarism tools and may ask you to explain your work in a follow-up.

Grading and Rubrics

Q: If I have a high grade but low attendance, can I still get an A?
A: No. You must meet the minimum attendance thresholds for certain letter grades. No rounding.

Q: Will there be detailed rubrics for assignments and exams?
A: Yes. Rubrics and grading criteria will be provided. You may request a regrade within a set window.

Policies and Exceptions

Q: What documentation is needed for excused absences or extensions?
A: Doctor's notes, official notifications, or other verification. Contact us early.

Q: Can late days be used for quizzes or project milestones?
A: No. Late days are for regular homework only (not HW0). Quizzes, exams, and projects have fixed deadlines.

Q: Who should I contact for emergencies?
A: Email the instructors and preceptor as soon as possible.

Auditing

Q: What if I miss the first week as an auditor due to illness or travel?
A: Auditors who miss the first week may lose access. If you have a compelling reason, contact us ASAP, but reinstatement is not guaranteed.

Q: Can I submit homework or take exams as an auditor for feedback?
A: No. Auditors may not submit graded work or join projects, exams, or office hours meant for registered students.

Still have a question not listed here?
Please reach out to the teaching team on Ed Discussion or email the instructors directly. We're here to help!