

Policies

Course Description

Foundations of Data Science combines three perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze that data so as to understand that phenomenon? The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. It delves into social issues surrounding data analysis such as privacy and design.

Prerequisites

This course does not have any prerequisites beyond high-school algebra. The curriculum and format is designed specifically for students who have not previously taken statistics or computer science courses. Students with some prior experience in either statistics or computing are welcome to enroll, though some parts of the course will be slow. Students who have taken both statistics and computer science courses should instead take a more advanced course.

Materials & Resources

Our primary text is an online book called Computational and Inferential Thinking: The Foundations of Data Science (<http://inferentialthinking.com/>). This text was written for the course by the course instructors.

The computing platform for the course is hosted at datahub.berkeley.edu (<http://datahub.berkeley.edu/>). Students find it convenient to use their own computer for the course. If you do not have adequate access to a personal computer, we have machines available for you; please contact the instructor.

Support

You are not alone in this course; the staff and instructors are here to support you as you learn the material. It's expected that some aspects of the course will take time to master, and the best way to master challenging material is to ask questions. For online questions, use Piazza (<http://piazza.com/>). We will also hold office hours (office-hours.html) for in-person discussions.

Labs

Weekly labs are a required part of the course and should be submitted during your lab session. To receive credit, you must attend lab, work on the lab assignment until you're finished or the lab period is over, and get checked off by a course staff member. Labs will be released on Monday night. If you don't want to attend lab physically, you may complete a lab assignment remotely, but you must complete it by Tuesday at 11:59pm to receive credit. Note that if you attend lab, you can still get credit even if you don't finish all of the lab problems. However, if you choose to work remotely, you must finish the entire lab to receive credit. Each person must submit each lab independently, but you are welcome to collaborate with other students in your lab room.

Sections

Small-group tutoring sections will be available to a subset of students who sign up for them during the second week of classes. For students who have not programmed before, these sections will be an excellent use of your time. Details about sign-ups will be shared in lecture and posted here.

Assignments

Data science is about analyzing real-world data sets, and so a series of projects involving real data are a required part of the course. You may work with a single partner on all projects; your partner must be from your lab section.

Weekly homework assignments are a required part of the course. Each student must submit each homework independently, but you are allowed to discuss problems with other students.

Exams

The midterm exam will be held in class (during the lecture period) on Friday, October 6. The final exam will be held from 3 p.m. to 6 p.m. on Tuesday December 12. Unless you have accommodations as determined by the university or permission from the instructor, you must take the midterm and the final at the dates and times provided here. Please check your course schedule and make sure that you have no conflicts with these exams.

Participation

Participation points can be earned in one of two ways: attending lecture or completing a final independent data exploration. Lecture attendance will begin to count in week 3; the first two weeks are optional. Students who have previously taken both computer science and statistics courses cannot receive participation credit for attendance; they must complete a final independent data exploration.

Details:

- The *final independent data exploration* can be completed in pairs and is expected to require 10-20 hours of work. The purpose of the exploration is to apply what you've learned to some data set of your choosing. Explorations will require the following elements. Correct execution of every element is sufficient for full credit. You will present your exploration during RRR week to the course staff.
 1. Choose and describe a data set that includes at least two tables, some quantitative variable, and some categorical variable.
 2. Visualize some quantitative variable(s) of the data in a way that summarizes the data effectively and write a short observational description.
 3. Visualize some categorical variable(s) of the data in a way that summarizes the data effectively and write a short observational description.
 4. Summarize some aspect of the data in a table that involves grouping or pivoting and write a short observational description.
 5. Summarize some aspect of the data in a table that involves joining the two tables and write a short observational description.
 6. State a hypothesis related to the data and the corresponding null hypothesis.
 7. Perform a statistical test for the hypothesis and write a short conclusion.
 8. Describe a prediction problem related to the data.
 9. Apply a prediction technique to the problem and briefly justify your choice of approach.
 10. Evaluate the prediction technique quantitatively and write a short conclusion.
- The requirement for *attending lecture* is to attend at least 2 lectures in each of at least 10 weeks, starting with week 3. Every Monday, Wednesday, or Friday in which lecture is not held (holidays, midterm, etc.), all students will receive credit as if they had attended lecture. Therefore, if a week has only 2 lectures because of a holiday, you only need to attend 1 of 2 to receive credit. In the event that wireless connectivity interrupts the attendance process, the instructor may announce that all enrolled students receive attendance credit. Not attending could potentially save you 10-20 hours; just enough time to complete your data exploration. Attendance will be taken via a Google Form. Yes, it's possible to fake attendance, but doing so is quite silly given that you have another option. Students caught faking attendance will fail the course. If you are intending to earn attendance credit but can't attend due to unforeseen circumstances, just fill out the excuse request form (<https://goo.gl/forms/XVtX9q53p3PZH3bZ2>).
- Students who have substantial prior experience in both statistics and computer science are only eligible for the data exploration option. That's not to punish you; it's to ensure that this course actually furthers your education. Both of the following must be true for this policy to apply to you.

1. In a prior semester, you have passed one of the following courses at Berkeley: Stat 2, 20, 21, 133, 134, 135 or any upper-division course containing the words "Probability" or "Statistics" in the course title, such as EE 126 or IEOR 172.
2. In a prior semester, you have passed one of the following courses at Berkeley: Stat 133, CS 61A, CS 61B, or Engineering 7.

Grading

Grades will be assigned using the following weighted components:

Activity	Grade
Participation	5%
Lab	10%
Homework	20%
Projects	25%
Midterm	10%
Final	30%

The course will not be curved, but further details of grading criteria may not be announced until the end of the course. It is certainly possible for all students to receive high grades in this course if all of you show mastery of the material on exams and complete all assignments.

Learning Cooperatively

With the obvious exception of exams, we encourage you to discuss all of the course activities with your friends and classmates as you are working on them. You will definitely learn more in this class if you work with others than if you do not. Ask questions, answer questions, and share ideas liberally.

Since you're working collaboratively, keep your project partner and the course staff informed. If some medical or personal emergency takes you away from the course for an extended period, or if you decide to drop the course for any reason, please don't just disappear silently! You should inform your project partner, so that nobody is depending on you to do something you can't finish.

Academic Honesty

Cooperation has a limit, however. You should not share your code or answers directly with other students. Doing so doesn't help them; it just sets them up for trouble on exams. Feel free to discuss the problems with others beforehand, but not the solutions. Please complete your own work and keep it to yourself. The exception to this rule is that you can share everything related to a project with your project partner and turn in one project between you.

Penalties for cheating are severe — they range from a zero grade for the assignment or exam up to dismissal from the University, for a second offense.

Rather than copying someone else's work, ask for help. You are not alone in this course! The course staff is here to help you succeed. If you invest the time to learn the material and complete the projects, you won't need to copy any answers.

Late Submission

We don't accept late submissions of labs or homeworks after the deadline. In lieu of offering exceptions or extensions, your lowest homework score and lowest lab score will be dropped. Projects will be accepted up to 2 days (48 hours) late; a project submitted less than 24 hours after the deadline will receive 2/3 credit, a project submitted less than 48 hours after the deadline will receive 1/3 credit, and a project submitted 48 hours or more after the deadline will receive no credit.

If you have a special circumstance that you believe requires an extension, you must contact your lab TA explaining the situation at least 24 hours before the due date of the assignment.

Late Adds

If you add the class late and were not on the waitlist, we will give you a grace period for submitting assignments if you contact us right away; please contact Fahad Kamran at fhdkmrn@berkeley.edu as soon as you add the course with your name, Student ID number, Calnet ID (username), and @berkeley.edu email address.

If you are on the waitlist, please do all coursework and attend labs and lectures; all assignments are due at the normal times.

A Parting Thought

This page shouldn't end with a list of penalties for cheating or lateness, because penalties and grades aren't the purpose of the course. We actually just want you to learn. Please keep that goal in mind throughout the semester. Welcome to Data 8.