Syllabus

Stat 133: Concepts in Computing with Data

Andrew Bray

An introduction to computationally intensive applied statistics. Topics will include organization storage of data, visualization and graphics, statistical learning and data mining, model validation procedures, and the presentation of results. This course uses R as its primary computing language.

Class Format

This is an in-person class with three lectures each week and one two-hour lab led by a GSI. This class is designed assuming you'll be in class. If you miss class, you'll be able to find most materials from class - slides, snippets of code - published on the course wesite.

Group tutoring

Tutors will offer group tutoring sessions several times each week. This is an opportunity to finish up any assignments that you've started in class or review any topics that are confusing for you. You're welcome to attend any session that works well for your schedule. Check the Office Hours page to see the group tutoring schedule.

Group tutoring is a great place to go to meet other students and collaborate on assignments with tutors on hand to help you get unstuck.

Office Hours

Office hours are an opportunity to chat one-on-one with your instructor and GSIs. Please come to office hours! Coming to office hours does not send a signal that you are behind or need extra help. On the contrary, coming to office hours early and often tends to co-occur with success in the course. We're happy to chat about the course material, statistics in general, careers in statistics, and whatever other statistics or data science topics are on your mind! Please check the Office Hours page to see the times of the various office hour/group tutoring sessions.

Materials

Computational tools

Each student will need a laptop in order to engage with the computing aspects of this class. Throughout the semester you'll be installing several different software tools, all of which are free. We will be mainly using the computing and programming environment R (via RStudio) to analyze data in this class. We will also ask you to use a command line interface to interact with your operating system. You do need your own computer to use R and do the assignments. No need to install these ahead of time; we'll guide you through how to get set up.

About half of the students in 133 have some coding experience in Python or R and about half do not. For those who come from Data 8 or some previous coding experience in python, the first labs will help you transfer that knowledge over to R. Both languages are excellent platforms for analyzing data, are widely used in data science, and

have their individual strengths. R has been developed within the statistics community specifically for data analysis, while python is a general-purpose programming language but has large data analysis capabilities.

Supplementary texts

Your primary reference throughout the course will be the notes that you take in class but there are also several terrific books that are free and online that can expand your knowledge.

- Advanced R, Hadley Wickham
- R for Data Science, Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund
- R packages, Hadley Wickham and Jenny Bryan

Course communication

bCourses

We don't use bCourses for much. It is primarily used to disseminate important announcements for the entire class such as the initial welcome announcement, final exam information etc.

Discussion forum

The official discussion forum for the class will be hosted on Ed. Ed is a forum to ask and answer questions with your fellow students and course staff. It's a useful resource for learning from your peers and seeking help from tutors and instructors.

If you have a question about the material or assignments (or anything related to the course), create a new post to ask your question on Ed. If it is about something personal, then mark it as "private", and only course staff will be able to see your post. This is the best way to contact us if you have a personal concern, as it ensures the fastest response. If your question does not include personal information and can be answered by other students, make sure it is public.

In a course this large, the instructors have a difficult time responding to individual emails, so please use the class forum or visit office hours if you wish to contact us.

Assignments, Exams, and Grading

Turning-in assignments

You will be turning in your assignments on a platform called Pensieve, the younger sibling of Gradescope. This is also the platform where your assignments will be graded, so you can return there to get feedback on your work. You are welcome to file a regrade request if you notice that we made an error in applying the rubric to your work. Note that regrade requests will need to be submitted by a deadline, which will usually be about a week the grades are released.

Problem Sets

Problem sets provide you with essential practice on the ideas and techniques discussed in class. They are generally due each week and are graded credit / no credit.

Projects

Projects are open-ended assignments assignments designed to apply the concepts from the lecture notes in the cause of doing an analysis of real data. This will involve both writing code and communicating your thoughts and findings in English. We'll be working through some problems from the projects in lab section, but you will be spending time on them outside of class time.

Projects are also graded credit / no credit.

Quizzes

Paper-and-pencil quizzes reinforce the most important concepts from the lecture and provide you the opportunity to work through misunderstandings of concepts with peers and the instructor. They cover the material from lecture, problem sets, and projects up until the time of the quiz.

There is both an individual and group component to each quiz.

The individual component will last about 40 minutes. You are allowed one, A4, one-sided handwritten sheet of notes with your name in the upper right-hand corner. The group component will take place immediately after the individual component has been completed and will last about 25 minutes. Your final (composite) quiz grade will be the average of your group and individual quiz scores.

This group quiz system almost always results in higher grades. We'll do a check at the end of the semester to calculate your quiz grade using just your individual quizzes. In the unlikely event that your individual grade average is higher, we'll use that grade.

Exam

The paper-and-pencil final exam will be held in person during finals week. It is a comprehensive exam and covers the material from the lecture, problem sets, and projects.

The time and date is Wednesday December 17th, 7 - 10 pm.

Grading

Your final grade in this class is, to the best of our ability, a measure of your understanding of the concepts of computing with data. As such, the lions share of your grade comes from measures of your individual understanding: quizzes and the final.

The best way to prepare for these assessments is to earnestly engage with the problem sets and proejects. Their role in the course is practice and they'll be your single most effective way to learn. We just want to see that you are indeed engaging in practice, so the problem sets and projects are both graded credit / no credit, with full credit given for evidence of earnest engagement.

- Problem Sets: 5%
- Projects: 15%
- Quizzes: 50%
- Final: 30%

The grades will not be curved (> 90% is some kind of A, 80-90 is some kind of B, etc.), so there is no limit to the number of As and Bs that are earned.

In order to provide flexibility around emergencies that might arise for you throughout the semester we will drop everyone's lowest quiz grade before calculating your quiz average at the end of the semester.

Policies

Late Work

Rigorously completing the problem sets and projects is for your own benefit. Since we grade them credit / no credit, no late work is accepted. Note that you can receive full credit for an assignment that is not complete as long as we see evidence of earnest engagment (please don't just submit a blank page).

Collaboration policy

On problem sets and projects, we strongly encourage collaboration between students. One of the best ways to learn is to talk things through with someone else. At the same time, these are your best opportunity to practice and learn, so be cautious not to lean to heavily on the understanding of a peer when working through your assignments.

AI policy

Please be thoughtful when using AI tools for your schoolwork. Doing so on problem sets and projects robs you of a valuable opportunity to learn and undermines your performance on exams. At the same time, they can be helpful in providing feedback on your writing.

At the end of the course, we'll bring in AI tools for coding and learn how to use them. They're very powerful, but you need firm mental models of the way code works before you're able to use them well. For that reason, we suggest you refrain from using them until they're introduced in class.

Accomodations for students with disabilities

Stat 133 is a course that is designed to allow all students to succeed. If you have letters of accommodation from the Disabled Students' Program, please share them with your instructor as soon as possible, and we will work out the necessary arrangements.

Staying healthy

Maintaining your health and that of the Berkeley community is of primary importance, so if you are feeling ill or have been exposed to illness, whether it's COVID-19 or something else, please do not come to class. You're encouraged to reach out to fellow students to discuss the class materials or stop by group tutoring or office hours to chat with a tutor or the instructor.

Frequently Asked Questions

1. What should I do if I'm on the waitlist?

   Attend both lecture and lab and submit all assignments on time. Unfortunately we cannot grow the size of the class any more, so places will become available only when students drop. If you have an urgent need for the course, visit the drop in advising sessions in our department: https://statistics.berkeley.edu/academics/undergrad/advising.

2. Are class sessions recorded?

   No.

3. Is attendance required?

   No.

4. Are time conflicts allowed?

No.

5. What if I join the class late?

   If you join the class within the first two weeks, read through all of the material on the course website that you've missed. Also ask a neighbor in lecture if you can borrow their notes. Finally, visit the Group Tutoring to discuss any topics you've missed. If you've missed any assignments in the first two weeks, request an extension from your GSI.

   After two weeks into the semester, you'll have too much material that you'll need to make up, so you will have to wait to a subsequent semester to take the course.

Campus Resources

UC Berkeley has a vast wealth of resources to help with all aspects of student life. You can read about them at the links below.

- Berkeley Supportal- great place to start
- Disabled Students' Program
- University Health Services
- UCB Path to Care
- Student Learning Center
- Berkeley International Office
- Ombuds Office for Students and Postdoctoral Appointees
- Gender Equity Resource Center
- Center for Educational Justice & Community Engagement
- UHS Counseling and Psychological Services (CAPS)
- Campus Academic Accommodations Hub
- ASUC Student Advocate's Office
- Basic Needs Center
- ASUC Mental Health Resources Guide