

Syllabus

Stats 32: Introduction to R for Undergraduates

Spring Quarter 2024, Stanford University

Logistics

This course meets on Tuesdays and Thursdays, for the first five weeks of the quarter (April 2 to May 2), from 12 noon to 1:20 pm in Mitchell Earth Sciences B67. See a detailed class schedule below.

Instructors and office hours

Instructor: Harrison Li. hli90722@stanford.edu. TA: Yujin Jeong. yujinj@stanford.edu

Office hours: Mondays, 11 am - 1 pm, Sequoia Hall 207 (Yujin) and Tuesdays, 9 am - 11 am, Sequoia Hall 220 (Harrison)

Please only use email for private, personal concerns not related to the material of the course or assignments. For all other course-related inquiries, please use the Ed discussion forum.

Course goals

The course aims to provide a fast-paced introduction to the R programming language for statistical computing. No prior experience in programming or statistics is assumed.

Our focus is not just on writing usable code but also on how to leverage R effectively to do good *data science*. This means that you will be challenged to apply the concepts covered in class sessions to solve real-world data problems, in a way that may require substantial critical thinking and persistence beyond simply learning the syntax of R. The hope is that this will enhance your ability to think deeply and precisely about data.

Assignments and grading

Grades in this course are determined by **in-class participation** (20%) and **weekly problem sets** (80%).

In-class participation will be graded based on interactive labs, to be done in groups and graded on a 0/1 completion scale. The lowest two lab scores will be dropped. The first lab on April 2 will not count.

There will be 4 problem sets, equally weighted, released on Thursdays and due the following Thursday at 11:59 am (before class). Please see the detailed course schedule below.

No late homeworks will be accepted.

Please submit each assignment on Gradescope as a single PDF file knitted from RMarkdown. Assignments submitted in any other format will not be accepted.

You are encouraged to discuss the problems with other students in the class, but every individual student must submit their own code.

Note that Internet sources **are** allowed for consultation on homeworks. This includes any non-human AI tools such as ChatGPT. Anyone programming in the real world will constantly be searching things up on the Internet. But please cite any outside sources you use (i.e. anything that is not in the class materials), and note all homework questions can be solved using the tools presented in lecture, although some may require clever thinking in terms of how to apply them.

All code you write must be original and your own work. That is, you may not substantively copy portions of your classmates' work. You may not use Internet forums to ask for answers to the assignments. Failure to adhere to these guidelines may constitute a violation of the Honor Code, and appropriate disciplinary action may be taken as a result.

Class meeting structure

Each class meeting will have 2 halves: a lecture portion, followed by an interactive “lab” session where students will work in small groups with the course instructors walking around the room to help guide the learning. Labs must be submitted before the end of class for participation credit.

You are **required** to bring a laptop computer to class. This is so you can work on the labs. Attendance is an integral part of this course.

Detailed course schedule

Unit	Lecture	Date	Topics	Released	Due
Fundamentals	1	4/2/2024	Introduction to R and RStudio. Basic data structures.		
Fundamentals	2	4/4/2024	Data frames, functions, packages, and the tidyverse.	HW 1	
Data manipulation	3	4/9/2024	File I/O, paths, logical and comparison operators.	HW 2	
Data manipulation	4	4/11/2024	Introduction to dplyr. Summary statistics, grouping, joins. Categorical and quantitative data variables.		HW 1 due
Data visualization	5	4/16/2024	Introduction to ggplot2. Univariate visualizations.	HW 3	
Data visualization	6	4/18/2024	More ggplot2. Multivariate visualizations.		HW 2 due
Data analysis	7	4/23/2024	Simple linear regression: Correlation, prediction, and graphical diagnostics.	HW 4	
Data analysis	8	4/25/2024	Multiple linear regression. Polynomial regression and locally weighted regression.		HW 3 due
Data analysis	9	4/30/2024	Hypothesis testing and A/B testing.		
Data analysis	10	5/2/2024	Model evaluation and selection. Cross validation. Prediction vs. inference.		HW 4 due

Class platforms

We will use Gradescope for assignment submission, and Canvas for course materials.

The Ed discussion forum will be used to facilitate collaboration between students in understanding the course material and homework assignments. The instructors will check periodically to help answer questions. You can use email or private posts on Ed to ask the instructors logistical questions about the course. Students who ask good questions and/or consistently give helpful responses to other students on Ed may receive a small boost (up to 5%) to their final grades.

OAE accomodations

Stanford is committed to providing equal educational opportunities for disabled students. Disabled students are a valued and essential part of the Stanford community. We welcome you to our class.

If you experience disability, please register with the Office of Accessible Education (OAE). Professional staff will evaluate your needs, support appropriate and reasonable accommodations, and prepare an Academic Accommodation Letter for faculty. To get started, or to re-initiate services, please visit oea.stanford.edu.

If you already have an Academic Accommodation Letter, we invite you to share your letter with us. Academic Accommodation Letters should be shared at the earliest possible opportunity so we may partner with you and OAE to identify any barriers to access and inclusion that might be encountered in your experience of this course.