Introduction to Programming in R

Fall 2020 Stat 33A

Description

This course introduces the R statistical software to students with minimal prior exposure to programming. The course aims to prepare students to carry out a data analysis and to write simple functions. The focus is on the computational model that underlies the R language with the goal of providing a foundation for further coding. Topics include: data types and structures, such as vectors, data frames and lists; the REPL evaluation model; function calls, argument matching, lazy evaluation, and environments; writing simple functions and control flow. Tools for reading, analyzing, and plotting data are also covered, such as data input/output, the formula language, and graphics.

Philosophy

This is a programming course rather than a math or statistics course. Programming is a broad topic, so it's not possible for lecture to provide you with a specific recipe for every situation you may encounter. Instead, one of the goals of this class is to help you become comfortable reading documentation and searching or asking for help.

Generally, you should spend about 10-15 minutes trying to solve a problem on your own. That may include reading documentation and experimenting with code to improve your understanding. If you're still stuck after that, it's time to look for help on the class forum, in office hours, or online (StackOverflow is a useful reference).

Assignments will sometimes ask you to learn about a topic related to but not covered in lecture. This may also happen spontaneously: if your code returns an unfamiliar error message, your first step should be to check documentation or search online to figure out what the error message means.

Whatever you do, don't get discouraged! Programming is challenging, and this course is designed to help ease you into it. The class forum and office hours are your support group.

Lecture & Lab

Lectures for this course will be asynchronous. Each week I'll post 4-6 lecture videos, each about 10-15 minutes long. I'll also post a workbook with exercises to accompany the videos. You're expected to watch a video, work through the related exercises, and repeat until finished. The entire process should take about 2 hours (1 hour of videos + 1 hour of exercises) per week.

There is no lab for this course; the workbooks use the time that would normally be used for labs.

Piazza

This course uses Piazza (www.piazza.com) as a forum and support group. You should have already received an email invite for Piazza. On Piazza, anyone in the class can answer your question, so you're likely to get an answer quickly.

Be courteous when posting! When you have a question:

- Search first in case it was already asked.
- Explain the context of the question.
- Provide a minimal example (with code) if appropriate.
- Post code as text rather than as a screenshot.

If you have trouble accessing Piazza, email me or the GSI.

Contacts

Instructor	Nick Ulle	nick.ulle@berkeley.edu
GSI	Seema Shet	seema_shet@berkeley.edu

Use email **only** for private matters: enrollment, grades, extensions, DSP accommodations, etc.

Use Piazza or office hours for questions about course materials. The office hours schedule is in a pinned post on Piazza.

Grade Policy

10% Participation: You are expected to participate on Piazza at least 2 times per week on average. Participation means posting good-faith questions, answers, or comments.

20% Workbooks: There will be 13 workbooks, each due a week after being posted. Workbooks will be graded for completion.

50% Homeworks: There will be 7 homeworks, each typically due 2 weeks after being posted. The homeworks will give you a chance to apply your skills to larger problems. Homeworks will be graded for correctness.

Late workbooks and homeworks will be penalized 20% per day.

10% Quizzes: There will be 4 quizzes. Each quiz will be time-limited to 1 hour. Your lowest quiz score will be dropped.

10% Final Exam: The format of the final exam will be similar to the quizzes, but 2-3 times longer. The final exam will be time-limited to 3 hours.

The quizzes and final exam will be available to start at any time in a 24-hour window on the scheduled day.

Collaboration

Professional statisticians and programmers collaborate and use references to solve problems. For workbooks and homeworks, I encourage you to:

- Discuss topics and exercises with your classmates.
- Search for references online and in books.
- Adapt short pieces of code you find on Piazza. 'Adapt' means you did more than simply copy and paste or change variable names.

In each case, you must cite the source (no formal citation style is required):

- Classmates: cite their name.
- Piazza: cite the post number.
- Other sources: cite the title, author, and URL.

All writing and graphics must be your own. All code must be your own, with the exception for short adaptations (see above). We may use automated tools to check for plagiarism.

For quizzes and exams, collaboration is not allowed.

Honor Code

UC Berkeley's honor code states "**As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.**"

References

The lecture notes are the primary reference for this course.

If you'd like additional references, see:

- R for Data Science by H. Wickham
- An Introduction to Programming with R by D. Nolan, N. Ulle, & C. Fitzgerald

The latter is a book we are in the early stages of developing for this course, so many parts are incomplete, and the chapter references listed on the schedule below may change.

Tentative Schedule

Quizzes will be on **Monday**.
Workbooks and homework will be due on **Thursday**.
Lecture videos will be posted on **Friday**.

W	Sunday	Topic	Optional References	Due
1	Aug 23	Setup; R Syntax	N 1-2	
2	Aug 30	Packages; Notebooks; Vectors	W 4, 6	
3	Sep 6	Data Types; Special Values	N 3	HW 1
4	Sep 13	Files; Data Frames	N 4-5; W 8, 10	Quiz 1
5	Sep 20	Subsets		HW 2
6	Sep 27	Graphics	N 6; W 3	Quiz 2
7	Oct 4	Graphics Case Study	N 6; W 3	HW 3
8	Oct 11	dplyr; If-statements	W 5	
9	Oct 18	Apply Functions	Note	HW 4
10	Oct 25	Functions	W 19	Quiz 3
11	Nov 1	Iteration	W 21.1 - 21.3	HW 5
12	Nov 8	Examples; Debugging		Quiz 4
13	Nov 15	Relational Data	W 13	HW 6
	Nov 22	Thanksgiving Break		
14	Nov 29	Tidy Data	W 12	HW 7
	Dec 6	Review Week		
	Dec 13	Finals Week		Final

W: R for Data Science
N: An Introduction to Programming with R

Accommodations for Students with Disabilities

Extra-time accommodations for quizzes and the final exam will be provided automatically as soon as I receive your DSP letter. If you need other accommodations, email me as soon as possible so we can work out the necessary arrangements.

Waitlist

This class has a long waitlist. If you decide you don't want to take the class, please drop as soon as possible to make room for others. The add/drop deadline is Sep 16th.

If you are waitlisted and need access to the bCourse or Piazza, email me.

If you have any other questions about enrollment, email the Statistics Department undergraduate advisors (stat-ugrad@berkeley.edu).

STAT 33A or 33B?

You might be wondering what the difference is between STAT 33A and 33B. The topics in 33A are a subset of the topics in 33B.

33A is designed for people that have never programmed before, the pace of certain parts is slower and more examples are provided. 33A puts slightly more emphasis on data analysis packages than 33B. The primary goal of 33A is that students finish the course with the programming skills necessary for a data analysis or implementation of a simple algorithm.

33B is designed for experienced programmers (not necessarily in R). The curriculum assumes students will already be familiar with programming concepts such as variables, if-statements, loops, and functions, so these topics are covered more quickly than in 33A. The extra time is used to cover details about how R works at a lower level. The primary goal of 33B is that students finish the course with the programming skills necessary to develop and maintain software in R.

IIn case you're unsure whether you satisfy the prerequisites for 33B, I've posted a short prerequisite test on the course website.

Letters of Recommendation

One-credit courses do not provide me with enough information about students to write a strong letter of recommendation. Therefore I am unwilling to provide a letter of recommendation based on your work in this course alone.