

Introduction to Programming in R

Spring 2020, STAT 33A

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.

Instructor:

Nick Ulle
nick.ulle@berkeley.edu

Office Hours:

M 11am-12pm 387 Evans
W 2-3pm 387 Evans

Lecture:

M 9-10am 50 Birge

GSI:

Seema Shet
seema_shet@berkeley.edu

Office Hours:

T 2-3pm 432 Evans
Th 2-3pm 432 Evans

Labs:

W 9-10am 342 Evans
W 10-11am 342 Evans
W 11-12pm 342 Evans
W 12-1pm 342 Evans

Reference Materials:

- [N] [An Introduction to Programming with R](#) by D. Nolan
 - [W] [R for Data Science](#) by H. Wickham
-

Grade Breakdown

Homeworks (30%): There will be 6 homework assignments. The homework assignments review concepts covered in lecture and lab. Your lowest homework score will be dropped. Late homeworks will not be accepted.

Lab Assignments (30%): There will be 7 lab assignments. Lab assignments will generally be due by midnight on Saturdays, but are meant to be finished in your lab section. Lab grading emphasizes completion rather than correctness. Your lowest lab score will be dropped. Late labs will be penalized 10% per day.

Quizzes (20%): There will be 4 quizzes (on paper), given during the lab session. Quiz questions are generally multiple-choice or fill-in-the-blank. Your lowest quiz score will be dropped. No make-up quizzes will be offered.

Final Exam (10%): There will be a final exam. The format will be similar to a quiz, but longer. See CalCentral for the exact time and location.

Attendance (10%): You must attend at least 9 of the 12 lectures to receive full credit. At some point during each lecture, there will be an attendance question. Your answer counts for attendance (even if it is incorrect). Suspended after 3/6 due to COVID-19.

Tentative Schedule (as of Mar 15)

This schedule will be updated throughout the semester.

W	Date	Lecture	References	Lab	HW Due
1	Jan 20	No Lecture (MLK, Jr. Day)		No Lab	
2	Jan 27	Syllabus; R Syntax; RMarkdown	N 1-2	Install R	
3	Feb 3	Types; Special values; Vectorization	N 3	Lab 1	
4	Feb 10	No Lecture (Cancelled)		Lab 2	HW 1
5	Feb 17	No Lecture (President's Day)		Quiz 1	
6	Feb 24	Paths; Data frames	N 4-5	Lab 3	
7	Mar 2	dplyr; Graphics with ggplot2	N 6; W 3, 5	Quiz 2	HW 2
8	Mar 9	Exploratory data analysis		Lab 4	
9	Mar 16	Subsets		Quiz 3	HW 3
S	Mar 23	No Lecture (Spring Break)		No Lab	
10	Mar 30	Reshaping data; Merging data	W 12-13	Review	
11	Apr 6	Writing functions; Conditionals	W 19	Lab 5	HW 4
12	Apr 13	Apply functions	Note	Lab 6	
13	Apr 20	Iteration	W 21.1-21.3	Lab 7	HW 5
14	Apr 27	Code Organization		Quiz 4	
R	May 4	Review Week			HW 6