

Course overview

BIOSTAT 213: Introduction to Computing in R

Fall 2019

Course objectives

This course provides an introduction to [R](#), a language for statistical analysis. After successful completion of the course, students should be able to:

- assign objects in R and identify R object types
- save R work
- import data files into R
- conduct basic statistical analysis in R
- manipulate data objects in R
- write functions in R and conditionally or iteratively apply R code and functions
- generate reports using R Markdown

In addition, students who successfully complete the course should be able to undertake [TICR courses](#) that require prior experience with R.

This is *not* a statistics or epidemiology course, and should not be used as a substitute for statistical/epidemiological training.

Faculty

Course director:

- Yea-Hung Chen, PhD, MS (yea-hung.chen@ucsf.edu)

Course assistant:

- Pingyang Liu, PhD (pingyang.liu@ucsf.edu)

Format

Time and location:

- Mondays, 2:45–5 pm
- Genentech Hall 114 (exception: November 4, in Mission Hall 1400)

Each class session will begin with roughly 1 hour of lecture, followed by roughly 1 hour of hands-on exercises.

Office hours

To request office hours, please email the course director.

Communication

Communication from faculty to students:

- most communication will occur via the course forum in the [Collaborative Learning Environment](#)
- by default, students should receive an email notification for each post in the forum (please contact TICR staff if you notice this not to be the case)

Communication from students to faculty:

- if you believe the entire class (or most of the class) would benefit from the communication, please post on the course forum
- otherwise, please email the course director

Assignments

The course will include 4 homework assignments, assigned every other week and due the Wednesday of the following week. Collaboration with other students in this course is permitted on these assignments. However, you must write all work—including R code—in your own voice, as much as possible, and acknowledge your collaborators in your submitted assignments.

Reading

All reading material will be posted on, or linked to from, the [Collaborative Learning Environment](#). This course does *not* have a required textbook.

Some optional (and free) resources for learning R include:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
- Grolemund G, Wickham H. *R for Data Science*. r4ds.had.co.nz; 2018.
- Heeren TC, Milton JN. *Basic Statistical Analysis Using the R Statistical Package*. Boston, MA: Boston University School of Public Health; 2016.
- Machlis S. [Beginner's guide to R](#). *Computerworld*. August 18, 2017.

Software

The course will be using [R](#) and [RStudio](#), both of which are available free of cost. Instructions for downloading and installing the software are available on the [Collaborative Learning Environment](#).

Grading

Students will receive a course score, out of 100, derived by equally weighting the homework assignments but dropping the lowest score.

At the end of the term, we will use our overall expectations and the distribution of the students' course scores to assign letter grades (A, B, C, D, or F). We do not use predetermined cutoffs, but are happy to disclose the cutoffs after we have assigned the grades.

Electronic devices in the classroom

Laptops, cellphones, tablets, and other electronic devices should not be used during class time for non-course activities (examples include: social media, online shopping, texting or chatting, or watching movies or other videos).

Devices should not emit sound during class time. Please adjust your device settings accordingly.

Disability support

If you need disability-related accommodations in this class, please contact [Student Disability Services](#).

Schedule

Please check the course website throughout the term for possible changes.

Module 1: fundamentals (Monday, September 23, 2019)

Topics:

- introduction to R and RStudio
- mathematical operators and functions
- functions
- object assignment
- scripts
- vectors
- quitting

Suggested reading:

- Aragón T.J. *Population Health Data Science with R*. bookdown.org: 2018.
 - [Chapter 1](#) (you may consider skipping 1.3)
 - [Chapter 2](#) (you may consider skipping 2.3, 2.4, and 2.5)
- Grolemund G, Wickham H. *R for Data Science*. r4ds.had.co.nz: 2018.
 - [Chapter 6](#)
- Thieme N. R generation. *Significance*. 2018;15(4):14–19.

Module 2: vectors and data frames (Monday, September 30, 2019)

Topics:

- numeric objects
- character objects
- useful functions for vectors
- working directories
- data frames

Assignment:

- Assignment 1 assigned, due Wednesday, October 9

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
 - [Chapter 2](#) (you may consider skipping 2.3, 2.4, and 2.5)
 - [Section 3.2](#)

Module 3: univariate analysis (Monday, October 7, 2019)

- factors
- summarizing continuous variables
- counts and percents
- univariate hypothesis tests

Suggested reading:

- Heeren TC, Milton JN. *Basic Statistical Analysis Using the R Statistical Package*. Boston, MA: Boston University School of Public Health; 2016.
 - [Chapter 1](#) and [Subsection 2.2.1](#)
- Machlis S. [Beginner's guide to R](#). *Computerworld*. August 18, 2017.

Module 4: bivariate and multivariate analysis (Monday, October 14, 2019)

Topics:

- continuous versus continuous
- continuous versus categorical
- installing and using packages
- *ggplot2*
- linear regression

Assignment:

- Assignment 2 assigned, due Wednesday, October 23

Suggested reading:

- Heeren TC, Milton JN. *Basic Statistical Analysis Using the R Statistical Package*. Boston, MA: Boston University School of Public Health; 2016.
 - [Subsection 4.1.1](#), [Subsection 2.2.2](#), [Section 2.5](#), [Section 2.6](#), and [Subsection 4.1.3](#)
- Kabacoff R. *Data Visualization with R*. rkabacoff.github.io; 2018.
 - [Chapter 1](#) through [Chapter 4](#)
- Grolemund G, Wickham H. *R for Data Science*. r4ds.had.co.nz; 2018.
 - [Chapter 3](#)

Module 5: data manipulation (Monday, October 21, 2019)

Topics:

- logical objects
- missing values
- categorization
- indexing
- date objects

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
 - [Subsection 2.2.4](#), [Section 4.8](#), and [Section 4.9](#)

Module 6: more data manipulation (Monday, October 28, 2019)

- indexing, continued
- `subset()`
- indexing and assignment
- `ifelse()`

Assignment:

- Assignment 3 assigned, due Wednesday, November 6

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
 - [Subsection 2.2.5](#), [Section 4.4](#), [Subsection 4.5.6](#)

Module 7: data reshaping (Monday, November 4, 2019)

Topics:

- `rbind()` and `cbind()`
- importing other data files
- merging data frames
- reshaping data

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
 - [Subsection 4.1.2](#) and [Section 4.6](#)

Module 8: programming (Monday, November 18, 2019)

Topics:

- `if()`
- writing functions
- the apply functions

Assignment:

- Assignment 4 assigned, due Wednesday, November 27

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
 - [Chapter 5](#) and [Subsection 3.2.6](#)