

# Course overview

*BIOSTAT 213: Introduction to Computing in R*

*Winter 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](mailto:yea-hung.chen@ucsf.edu))

Course assistant:

- Adrienne Epstein, MS ([adrienne.epstein@ucsf.edu](mailto:adrienne.epstein@ucsf.edu))

## Format

Time and location:

- Wednesdays, 1–3:30 pm
- Mission Hall 1407

Each class session will begin with roughly 1.5 hours of lecture, interspersed with hands-on exercises, followed by roughly 1 hour of additional 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 3 homework assignments. 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.

## 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 from the 3 homework assignments, weighted by the total number of points available for each assignment.

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 (Wednesday, January 16, 2019)

Topics:

- introduction to R and RStudio
- mathematical operators and functions
- functions
- object assignment
- scripts
- numeric objects
- character objects

Required reading:

- Thieme N. R generation. *Significance*. 2018;15(4):14–19.

Suggested reading:

- Aragón TJ. *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](#)

### Module 2: univariate analysis (Wednesday, January 23, 2019)

Topics:

- working directories and images
- reading CSV files
- data frames
- factors
- descriptive measures
- histograms and stem-and-leaf plots
- tables and `prop.table()`
- univariate hypothesis tests

Assignment:

- Assignment 1 posted, due the following week

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 3: bivariate and multivariate analysis (Wednesday, January 30, 2019)

Topics:

- scatter plots and `loess()`
- box plots
- introduction to *ggplot2*
- bivariate tables and `prop.table()`
- bivariate hypothesis tests
- linear regression

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](http://rkabacoff.github.io); 2018.
  - [Chapter 1](#) through Chapter 4
- Grolemund G, Wickham H. *R for Data Science*. [r4ds.had.co.nz](http://r4ds.had.co.nz); 2018.
  - [Chapter 3](#)

## Module 4: data manipulation (Wednesday, February 6, 2019)

Topics:

- logical objects
- missing values
- indexing
- `subset()`
- `ifelse()`
- date objects

Assignment:

- Assignment 2 posted, due the following week

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. [bookdown.org](http://bookdown.org); 2018.
  - [Section 4.8](#), [Subsection 2.2.4](#), [Subsection 2.2.5](#), [Section 4.4](#), [Subsection 4.5.6](#), and [Section 4.9](#)

## Module 5: advanced data manipulation (Wednesday, February 13, 2019)

Topics:

- reading other data files
- `rbind()` and `cbind()`
- merging data frames
- reshaping data and *tidyr*
- *dplyr*

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. [bookdown.org](http://bookdown.org); 2018.
  - [Subsection 4.1.2](#) and [Section 4.6](#)
- Grolemund G, Wickham H. *R for Data Science*. [r4ds.had.co.nz](http://r4ds.had.co.nz); 2018.
  - [Chapter 12](#)

## Module 6: programming (Wednesday, February 20, 2019)

Topics:

- writing functions
- `if()`
- `for()`
- the apply functions
- lists
- regular expressions

Assignment:

- Assignment 3 posted, due the following week

Suggested reading:

- Aragón TJ. *Population Health Data Science with R*. bookdown.org; 2018.
  - [Chapter 5](#), [Subsection 3.2.6](#), [Section 3.1](#), and [Section 4.11](#)
- Sanchez G. *Handling Strings with R*. www.gastonsanchez.com; 2018.

## Module 7: report generation (Wednesday, February 27, 2019)

Topics:

- R Markdown

Suggested reading:

- [Get Started](#). Boston, MA: RStudio; 2018.
- Xie Y, Allaire JJ, Golemund G. *R Markdown: The Definitive Guide*. bookdown.org; 2018.
  - [Chapter 1](#)
  - [Chapter 2](#) (you may consider skipping 2.7)
  - [Chapter 3](#)