

STA 4930: Programming with Data

Spring 2020
MWF 12:50–1:40 pm

Section: 1B95
Class #: 23702

Contact Information

Instructor

Name: Brett Presnell
Email: presnell@ufl.edu
Web: <https://www.stat.ufl.edu/~presnell/>
Office: FLO 225
Office Hrs: Refer to instructor's web page

TA

Name: Yue Bai
Email: baiyue@ufl.edu
Office: FLO 101A
Office Hrs: TBD

Course Description

An introduction to statistical computing and programming with data. Topics include basic programming in R; data types and data structures in R; importing and cleaning data; specifying statistical models in R; statistical graphics; statistical simulation using pseudo-random numbers; reproducible research and the documentation of statistical analyses.

Prerequisites

STA 3032 (B-) or STA 2023 (B) or AP Statistics (4).

Course Objectives

You will learn to do the following

1. Write simple functions in R making effective use of data structures and control structures.
2. Import data from various sources into R and prepare the data for analysis.
3. Determine statistical graphics appropriate to a statistical analysis and produce them using R.
4. Formulate statistical models in the R language.
5. Perform and document a basic statistical analysis.
6. Carry out simple statistical simulations and report the results.

Source Materials

We will use a variety of on-line texts and other resources. Class notes and other materials will be made available on the [course website](#). The following two texts will be our primary references and both are available on-line for free:

- [Hands-On Programming with R: Write Your Own Functions and Simulations](#) (Grolemund 2014)
- [R for Data Science: Import, Tidy, Transform, Visualize, and Model Data](#) (Wickham and Grolemund 2016)

Course Policies

Grading

- 35% Homework and programming assignments
- 25% Exam 1
- 25% Exam 2
- 10% Quizzes
- 5% Attendance

Attendance

Attendance will be recorded for each class meeting. Anyone arriving 15 minutes or more after the beginning of class will be counted as absent. Arriving late but within 15 minutes of the beginning of class will count as 1/2 of a full absence. Students with 7 or more unexcused absences will receive a 0 for attendance; students with 2 or fewer unexcused absences will earn full credit for attendance; the attendance grades of those in between will be by linear interpolation. Absences will be excused only if they are documented and conform to the attendance policies of the university as described in the [Undergraduate Catalog](#). If you know that you will have to miss class for an excused reason, please inform the instructor in advance.

Quizzes and Exams

There will be no make-ups for missed exams or quizzes. If you will be unable to attend an exam or quiz, please let the instructor know well in advance and they will try to make other arrangements for you to take the exam. In case of an (unavoidable) excused absence, suitable accommodations will be arranged.

Homework Assignments

Homework must be submitted on time. No late assignments will be accepted under any circumstances.

Students are expected to work independently, unless otherwise specified in writing. Offering and accepting solutions from others is an act of plagiarism, which is a serious offense and all involved parties will be penalized according to the [UF Honor Code](#). Discussion amongst students is encouraged, but when in doubt, direct your questions to the instructor.

Accommodations for Students with Disabilities

Students requesting accommodation for disabilities must first register with the [Dean of Students Office](#). The Dean of Students will provide documentation to the students who must then provide this documentation to the instructor when requesting information. You must submit this documentation prior to submitting any assignments or taking any exam or quiz for which you are requesting accommodation.

Academic Misconduct

Students will be held accountable to the [UF Honor Code](#).

Course Evaluations

Students are expected to provide feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

Class Schedule

Outline

This is an aspirational schedule for the class. This is the first time that this course has been taught at UF, so we will be adapting this schedule as we go along to fit the needs, abilities, and interests of the students and the instructor.

Week 1

- A Brief History of Statistical Computation
- Installing R
- Using the R Console
- Using RStudio
- Other interfaces to R
- Extending R: the R package system

Week 2

- Introduction to programming in R

Week 3

- Probability via computer: generating data via simulation
- Examining simulation results with summary statistics and simple graphics

Weeks 4-5

- Types of variables and data
- Data types in base R
- Importing data from files and other programs
- Data structures in base R

Week 6

- Review of basic statistical inference and graphics
- Basic statistical inference and graphics in R

Week 7

- Introduction to data scraping
- Working with character strings in R
- Cleaning, transforming, and organizing data

Week 8

- The concept of tidy data
- Introduction to the Tidyverse: modern data structures in R

Week 9

- Review of simple linear regression
- Simple linear regression in R

Week 10

- Introduction to multiple regression
- Model formulas: specifying models in R
- Fitting linear models in R

Week 11

- Introduction to logistic regression
- Logistic regression in R

Week 12

- The grammar of graphics: ggplot2

Week 13

- Reproducibility in science
- The practical advantages of a reproducible statistical analysis
- Literate data analysis: documenting a statistical analysis

Week 14

- Version control and collaboration: an introduction to git

References

- Grolemund, Garrett. 2014. *Hands-on Programming with R: Write Your Own Functions and Simulations*. Sebastopol, CA: "O'Reilly Media, Inc.". [%22https://rstudio-education.github.io/hopr/%22](https://rstudio-education.github.io/hopr/).
- Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, CA: "O'Reilly Media, Inc.". [%22https://r4ds.had.co.nz/%22](https://r4ds.had.co.nz/).