

STA 3100: Programming with Data

Spring 2022
MWF 1:55–2:45 pm

Section: DATA
Class #: 22465

Contact Information

Instructor

Name: Brenda Betancourt
Email: bbetancourt@ufl.edu
Office: Griffin-Floyd 220
Office Hrs: Wed, 3:00 - 4:30pm

TA

Name: Sourav Mukherjee
Email: souravmukherjee@ufl.edu

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.

- Download R from <http://cran.us.r-project.org/>
- Download RStudio from <http://rstudio.org/download/desktop>

Prerequisites

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

Course Objectives

You will learn to do the following

1. Import data into R and prepare the data for analysis.
2. Write functions in R making effective use of data structures and control structures.
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 basic simulations.
7. Document and report the results of data analyses and simulations in a reproducible way.

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 (free, on-line) texts will be our primary references:

- [R for Data Science: Import, Tidy, Transform, Visualize, and Model Data](#) (Wickham and Grolemund 2016)
- [OpenIntro Statistics \(4th Ed\)](#)(Diez, Cetinkaya-Rundel, and Barr 2019)

Course Policies

Due to the ongoing COVID-19 pandemic, masks are expected in the classroom.

Grading

- 100% Homework/Projects

Emphasis will be on homework assignments in which your knowledge will be put to use.

Projects and Homework Assignments

Homework must be submitted on time. Late assignments will only be accepted in exceptional circumstances.

Students are expected to work independently, unless otherwise specified in writing. Offering and accepting solutions from others is an act of plagiarism. 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. 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

- Using the R Console
- Using RStudio
- Reproducibility in science
- Using R Markdown for documenting statistical analysis

Week 2

- Introduction to programming in R
- Types of variables and data in base R

Weeks 3-4

- Review of basic descriptive statistics
- Summary statistics and simple graphics in base R
- Introduction to Tidyverse: the grammar of graphics (ggplot2) and basic data manipulation (dplyr).

Week 5-6

- Review of probability
- Probability distributions in R
- Probability via computer: generating data via simulation

Week 7-8

- Review of basic statistical inference for numerical and categorical variables.
- Basic statistical inference and graphics 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 12-13

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

Week 14

- Version control and collaboration: an introduction to git

References

Diez, David, Mine Cetinkaya-Rundel, and Christopher D Barr. 2019. *OpenIntro Statistics*. 4th ed. <https://www.openintro.org/>.

Wickham, Hadley, and Garrett Golemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, California: O'Reilly Media, Inc. <https://r4ds.had.co.nz/>.