



STA 3100: Programming with Data

Fall 2023

Class Number (Section): 25151 (6736)

Meets: MWF 3:00 – 3:50 PM (Period 8)

Location: FLO 100

Web: <https://ufl.instructure.com/courses/494742>

Instructor: Dr. Brett Presnell

Teaching Assistant: Dipshi Roychowdhury

Contact Information

Dr. Brett Presnell

Email: presnell@ufl.edu

Web: <https://www.stat.ufl.edu/~presnell/>

Office: FLO 225

Virtual Office: Zoom 940 1233 3509

Office Hrs: MW 4-5 PM

Dipshi Roychowdhury

Email: droychowdhury@ufl.edu

Office: FLO 117D

Virtual Office: Zoom 658 792 6980

Office Hrs: Tue 3-4:30 (in person)

Fri 1-2:30 (online)

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. 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. Formulate statistical models in the R language.
4. Perform, document, and interpret common statistical analyses.
5. Carry out statistical/probabilistic simulations.
6. Determine statistical graphics appropriate to a statistical analysis and produce them using R.
7. Document and report the results of data analyses and simulations in a reproducible way.

Text Books and Other 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](#). Most readings will be taken from the following (free, on-line) texts, which students are encouraged to peruse on their own:

- r4ds2e : [R for Data Science \(2e\): Visualize, Model, Transform, Tidy, and Import Data](#)
- rp4ds : [R Programming for Data Science](#)
- hopr : [Hands-On Programming with R : Write Your Own Functions and Simulations](#)
- ggplot2 : [ggplot2: Elegant Graphics for Data Analysis](#)
- advr : [Advanced R \(2nd Ed\)](#)

Bibliography

- Grolemund, Garrett. 2014. *Hands-on Programming with R: Write Your Own Functions and Simulations*. Sebastopol, CA: O'Reilly Media, Inc. <https://rstudio-education.github.io/hopr/>.
- Peng, Roger D. 2016. *R Programming for Data Science*. 5+ ed. Lulu.com. <https://bookdown.org/rdpeng/rprogdatascience/>.
- Wickham, Hadley. 2019. *Advanced R*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://adv-r.hadley.nz/>.
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd ed. Sebastopol, California: O'Reilly Media, Inc. <https://r4ds.hadley.nz/>.
- Wickham, Hadley, Danielle Navarro, and Thomas Lin Pedersen. 2022. *Ggplot2: Elegant Graphics for Data Analysis*. 3rd ed. Springer. <https://ggplot2-book.org/>.

Course Policies

Grading

There will be regular online quizzes to help you refine your knowledge and understanding of the course material. Homework assignments and projects will put this knowledge to use. These will be weighted in the final course average (percentage) as follows:

- 80% Homework/Projects
- 20% Quizzes

Letter grades in the course will be determined from the final course average according to the following scale:

A	A-	B+	B	B-	C+	C	D	E
94-100	90-93	87-89	83-86	80-82	77-79	67-76	50-66	0-49

Further information may be found in the university's [grades and grading policies](#).

Late Submissions

Homework and projects must be submitted on time, and it is the student's responsibility to allocate sufficient time to complete each assignment by the due date.

Late assignments will be accepted in cases of documented emergency or illness, but you must inform the instructor in advance of any illness which may lead to a late submission.

In all other cases, acceptance of late assignments will be at the discretion of the instructor. Scores on late submissions which are accepted will be reduced by 10% plus an additional 5% for each additional day between the due date and the time of submission.

Nota bene, it is the student's responsibility to correctly submit their work for every assignment, so *always double check that you have submitted the correct file(s) for each assignment*. **If you have not submitted the correct file(s) by the due date, then subsequent submission will be treated as a late submission.**

Academic Misconduct

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

Unless otherwise specified in writing by the instructor, students are expected to work independently or in *assigned* groups. General discussion of the course material is encouraged, but offering or accepting solutions from others is plagiarism. When in doubt, direct your questions to the instructor or TA.

Recording of Class Sessions

As in all courses at UF, unauthorized recording and unauthorized sharing of recorded materials by students or any other party is prohibited.

Except in special circumstances, class sessions will not be recorded by the instructor. In case a class session is recorded, students who participate with their camera engaged or who utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voice recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the “chat” feature, which allows students to type questions and comments live. The chat will not be recorded or shared.

Accommodations for Students with Disabilities

Students requesting accommodation for disabilities must register with UF’s [Disability Resource Center](#). The DRC 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.

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.bluer.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 course. It may be altered or rearranged to adapt to the backgrounds, abilities, and interests of the students in the class. There are 41 scheduled class meetings.

Week 1 (Aug 23 – Aug 25)

- Getting started
- Vectors and Vectorized Operations

Week 2 (Aug 28 – Sep 1)

- Distributions and Descriptive Statistics
- Writing Your Own Functions

- Introduction to R Markdown

Week 3 (Sep 6 – Sep 8)

- Matrices and Arrays

Week 4 (Sep 11 – Sep 15)

- Lists
- Data frames (and tibbles)
- Importing and Exporting Data

Week 5 (Sep 18 – Sep 22)

- Column and row operations on data frames
- Pipes and more operations on data frames
- Joining/Merging Data Frames

Week 6 (Sep 25 – Sep 29)

- Dates and times in base R
- The lubridate package

Week 7 (Oct 2 – Oct 4)

- Tidy Data and Pivoting
- Character strings and the stringr package
- String matching with regular expressions

Week 8 (Oct 9 – Oct 13)

- Detecting string matches
- Extracting string matches

Week 9 (Oct 16 – Oct 20)

- String replacement and string splitting
- An extended example with character strings
- Introduction to Data Scraping

Week 10 (Oct 23 – Oct 27)

- Factors in base R
- The forcats package

Week 11 (Oct 30 – Nov 3)

- Elementary statistical inference
- Simple linear regression
- Multiple regression

Week 12 (Nov 6 – Nov 8)

- Factors and dummy variables in regression
- Interactions
- Simple logistic regression

Week 13 (Nov 13 – Nov 17)

- Multiple logistic regression
- More graphics in R

Week 14 (Nov 20 – Nov 20)

- Working with lists: the purrr package

Week 15 (Nov 27 – Dec 1)

- More on data scraping and/or simulation.

Week 16 (Nov 27 – Dec 1)

- Catch up and review