

STA 3100: Programming with Data

Spring 2024

Instructor

Matias Shedden

Email: sheddenmatias@ufl.edu

Preferred contact: Email

Office Hours: 11:30 am - 12:30 pm MWF

Office: FLO 101A

Class information

Classroom: FLO 0100

Class Hours: 8:30 am - 9:20 am MWF

Teaching Assistant

Minxuan Wu

Email: wuminxuan@ufl.edu

Office Hours: 1:00-4:00 pm Wednesday

Office: On Zoom, Room Number 99603410309

Sections: All

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

Course goals

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.

Grades

The grades will be determined according to the following scale:

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

In total, there will be five quizzes, two exams, and four homework assignments. The class will be weighted as follows:

Item	Weight
Homework	60 %
Quizzes (highest 4)	15 %
Exam 1	12.5 %
Exam 2	12.5 %

The lowest of your five quiz scores will be dropped, and only the highest four will factor into your final grade.

Attendance and Make-Ups

Attendance is expected and will be essential for performing well in the class. There is however, no attendance grade.

See university attendance policies:

<https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/>

Late homework will be graded with a 15% deduction per day late. More than 3 days late will receive a 0.

Recommended Textbooks

There is no textbook for the class. The following free, online texts, however, may be referred to in class and be otherwise useful as a reference:

- r4ds : R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Wickham and Golemund 2016) <https://r4ds.had.co.nz/>
- rp4ds : R Programming for Data Science (Peng 2016) <https://bookdown.org/rdpeng/rprogdatascience/>
- hopr : Hands-On Programming with R : Write Your Own Functions and Simulations (Golemund 2014) <https://rstudio-education.github.io/hopr/>
- ggplot2 : ggplot2: Elegant Graphics for Data Analysis (Wickham, Navarro, and Pedersen 2022) <https://ggplot2-book.org/>
- advr : Advanced R (2nd Ed) (Wickham 2019) <https://adv-r.hadley.nz/>

UF Grading Policies

<https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>

Evaluations

Students are expected to provide professional and respectful 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.ua.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.ua.ufl.edu/public-results/>.

Accommodations

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Recordings

This class is 100% in-person, there will be no recordings of the lectures.

Weekly Schedule

Here is a tentative weekly schedule for the class:

Week 1

- Introduction to R, Rstudio, and Rmarkdown
- Directories
- Introduction to computer programming
- Basic R functions
- Vectors
- Logic

Week 2

- Distributions
- Generating data
- Monday off
- Homework 1 Assigned
- Quiz 1

Week 3

- Algorithms
- Plotting
- Functions
- Homework 1 Due

Week 4

- Statistical analysis, testing
- Testing Simulations
- Quiz 2

Week 5

- Matrices, arrays, contingency tables

Week 6

- Importing data
- Data frames
- Quiz 3

Week 7

- Tidyverse, dplyr
- Pipes
- Homework 2 Assigned

Week 8

- More tidyverse
- Joins
- Homework 2 Due

Week 9

- Review
- Exam 1

Week 10

- Spring Break

Week 11

- ggplot2
- Homework 3 Assigned

Week 12

- Sampling
- Classification
- Quiz 4
- Homework 3 Due

Week 13

- Linear Regression
- Hypothesis testing in linear regression
- Linear models with categorical data and interactions

Week 14

- Regression continued
- Homework 4 Assigned
- Quiz 5

Week 15

- Logistic Regression
- Strings, Regex
- Homework 4 Due

Week 16

- Review
- Exam 2
- Friday off

If this document is updated, an announcement will be made in class, and the new version will be uploaded to the course website in canvas.