STA 3100: Programming with Data

Summer A 2023

Instructor

Matias Shedden
Email: sheddenmatias@ufl.edu
Preferred contact: Email

Office Hours: 1:45-2:45 p.m. W/F (as needed)
Office: FLO 234

Class information
Classroom: FLO 0100
Class Hours: 12:30 pm - 1:45 pm MTWRF

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 four quizzes, two tests, and three homework assignments. The class will be weighted as follows:

Item	Weight
Homework	60 %
Quizzes (highest 3)	15 %
Exam 1	12.5 %
Exam 2	12.5 %

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

Attendance and Make-Ups

This class will move very fast. 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 Grolemund 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 (Grolemund 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.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/.

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
- Logic
- Generating data
- Introduction to statistical analysis in R
- Quiz 1
- Homework 1 assigned

Week 2

- Vectors, matrices, and lists (and related functions, indexing, matrix algebra)
- Algorithms
- Quiz 2
- Homework 1 due

Week 3

- Importing data
- Data frames
- Tidyverse, dplyr packages
- Pipes
- Homework 2 assigned
- Quiz 3

Week 4

- Merging data frames
- Tidy data
- Pivoting
- Dates
- Test review
- Homework 2 due
- Homework 3 assigned
- **Test 1**

Week 5

- Linear Regression in R
- Hypothesis testing in linear regression
- Linear models with categorical data and interactions
- Quiz 4

Week 6

- String matching
- Catch up
- Review
- Homework 3 due
- **Test 2**

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.