

STA 3100: Programming with Data

Summer A — 2024

Instructor Information

Name: Lingxiao Zhou

Office: FLO 218

Office Hours: TRF 2-3 PM(in person)

Email: zhou.l@ufl.edu

Phone: (352) 392-1941

Preferred contact: Email

Class Information

Time: MTWRF 12:30-1:45 PM

Classroom: FLO 0100

TA Information

Name: Patrick Mellady

Email: pmellady@ufl.edu

Office: FLO 105

Office Hours: M 2-3 PM (in person)

W 2-3 PM ([Zoom](#))

F 3-4 PM ([Zoom](#))

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

- Import data into R and prepare the data for analysis.
- Write functions in R making effective use of data structures and control structures.
- Determine statistical graphics appropriate to a statistical analysis and produce them using R.
- Formulate statistical models in the R language.
- Perform and document a basic statistical analysis.
- Carry out basic simulations.
- Document and report the results of data analyses and simulations in a reproducible way.

Grading

The course grade is determined by the following components:

Homework	70%
Final Project	30%

Grade Scale

Final grades will be assigned according to the following scale:

A	94 – 100	C+	77 – 79
A-	90 – 93	C	67 – 76
B+	87 – 89	D	60 – 66
B	83 – 86	E	0 – 59
B-	80 – 82		

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 and project will not be accepted.

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 Policy

<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.afl.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/>.

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.

Lecture Format

This class is 100% in-person, there will not be online options for lectures.

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. 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.

Tentative Schedule

The following is a *tentative* schedule for the course.

Week 1

- Introduction to R, Rstudio, and Rmarkdown
- Basic R functions
- Vectors
- HW 1 assigned

Week 2

- Algorithms
- Plotting
- Functions
- HW 1 due, HW 2 assigned

Week 3

- Matrices, arrays, contingency table
- Dataframes
- Tidyverse, dplyr
- HW 2 due, HW 3 assigned

Week 4

- Join
- More data types
- More on dplyr
- HW 3, and HW 4 assigned
- Final project assigned

Week 5

- Linear regression
- Hypothesis testing in linear regression
- Linear models with categorical data and interactions
- Nonlinear regression
- HW 4 due, HW 5 assigned

Week 6

- More on regression
- HW 5 due
- Final project due