# STA 3100: Programming with Data

| | |
|---|---|
| Fall 2020 | Section: 3502 |
| MWF 12:50–1:40 pm | Class #: 27256 |

## Contact Information

| **Instructor** | **TA** |
|---|---|
| Name: Brett Presnell | Name: Mirajul Islam |
| Email: presnell@ufl.edu | Email: mirajulislam@ufl.edu |
| Web: https://www.stat.ufl.edu/~presnell/ | Web: NA |
| Office: FLO 225 (Zoom) | Office: Zoom |
| Office Hrs: Refer to instructor's web page | Office Hrs: Tues 3:00-4:00pm |

## 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 statisical analyses.

### Prerequisites

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

## Course Objectives

You will learn to do the following

1. Write simple functions in R making effective use of data structures and control structures.
2. Import data into R and prepare the data for analysis.
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)
- Advanced R (2nd Ed) (Wickham 2019)
- Hands-On Programming with R: Write Your Own Functions and Simulations (**???**)

## Course Policies

Due to the COVID-19 pandemic, all class meetings and office hours will be held online using Zoom.

### Grading

- 80% Homework/Projects
- 20% Quizzes

There will be regular online quizzes to help you test and refine your knowledge. Primary emphasis will be on homework assignments and projects in which this 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 rst 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. This is the first time that this course has been taught at UF, so 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

- A Brief History of Statistical Computation
- Installing R
- Using the R Console
- Using RStudio
- Other interfaces to R
- Extending R: the R package system

### Week 2

- Introduction to programming in R

### Week 3

- Probability via computer: generating data via simulation
- Examining similation results with summary statistics and simple graphics

### Weeks 4-5

- Types of variables and data
- Data types in base R
- Importing data from files and other programs
- Data structures in base R

### Week 6

- Review of basic statistical inference and graphics
- Basic statistical inference and graphics in R

### Week 7

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

**Week 8**

- The concept of tidy data
- Introduction to the Tidyverse: modern data structures 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**

- The grammar of graphics: ggplot2

**Week 13**

- Reproducibility in science
- The practical advantages of a reproducible statistical analysis
- Literate data analysis: documenting a statistical analysis

**Week 14**

- Version control and collaboration: an introduction to git

## References

Wickham, Hadley. 2019. *Advanced R*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://adv-r.hadley.nz/.

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