

STA 3100: Programming with Data

Spring 2021
MWF 1:55–2:45 pm

Section: DATA and OLTA
Class #: 24935 and 26197

Contact Information

Instructor

Name: Brenda Betancourt
Email: bbetancourt@ufl.edu
Office: Zoom
Office Hrs: TBD

TA

Name: Somnath Bhadra
Email: somnath.bhadra@ufl.edu

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.

- Download R from <http://cran.us.r-project.org/>
- Download RStudio from <http://rstudio.org/download/desktop>

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. 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)
- [OpenIntro Statistics \(4th Ed\)](#)(Diez, Cetinkaya-Rundel, and Barr 2019)

Course Policies

Attendance for lectures is strongly recommended. Students enrolled in class number 24935 are encouraged to attend the live lectures in Turlington Hall 1105. Students enrolled in class number 26197 are expected to join the live lectures through Zoom and are NOT allowed in the live lectures because of social distancing rules in place due to COVID-19. We will likely be recording the lectures and posting them to the course website, but this is our first semester working with HyFlex teaching, so please attend the live session in case recordings are not available.

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

Grading

- 100% Homework/Projects

Emphasis will be on homework assignments in which your 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.

COVID related practices for face to face attendance

- You are required to wear approved face coverings at all times during class and within buildings. Failure to do so will lead to a report to the Office of Student Conduct and Conflict Resolution. UF's policy on masking and physical distancing can be found here (<https://policy.ufl.edu/policy/masking-and-physical-distancing/>). Note that the policy clearly states that "Masks with exhalation valves or vents allow air to be expelled and are prohibited"

and “neck gaiters and bandanas should not be worn unless no other face covering is available”. Students without masks will be asked to leave.

- This course has been assigned a physical classroom with enough capacity to maintain physical distancing (6 feet between individuals) requirements. Utilize designated seats and maintain appropriate spacing. Do not move desks or stations.
- Sanitizing supplies are available in the classroom if you wish to wipe down your desks prior to sitting down. Please clean your work area thoroughly when the class is completed.
- Practice physical distancing when entering and exiting the classroom.
- IMPORTANT: If you are experiencing COVID-19 symptoms, please use the UF Health screening system and follow the instructions on whether you are able to attend class.

Accommodations for Students with Disabilities

Students requesting accommodation for disabilities must first 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.bluer.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

Recording of Class Sessions

Our class sessions may be audio-visually recorded for students in the class to refer back and for enrolled students who are unable to attend live. Students who participate with their camera engaged or 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. As in all courses, unauthorized recording and unauthorized sharing of recorded materials by students or any other party is prohibited.

Class Schedule

Outline

This is an aspirational schedule for the class. 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

- Using the R Console
- Using RStudio
- Reproducibility in science
- Using R Markdown for documenting statistical analysis

Week 2-3

- Introduction to programming in R
- Probability via computer: generating data via simulation

Weeks 4-5

- Types of variables and data
- Data types and structures in base R
- Summary statistics and simple graphics
- The grammar of graphics: ggplot2

Week 6-7

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

Week 8

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

Week 9-10

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

Week 11

- Review of simple linear regression
- Simple linear regression in R

Week 12

- Introduction to multiple regression
- Model formulas: specifying models in R
- Fitting linear models in R

Week 13

- Introduction to logistic regression
- Logistic regression in R

Week 14

- Version control and collaboration: an introduction to git

References

Diez, David, Mine Cetinkaya-Rundel, and Christopher Barr. 2019. *OpenIntro Statistics*. 4th ed.

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/>.