Import → Tidy → Transform → Visualise → Model → Communicate

Understand

Program

STA 3100: Programming with Data

Spring 2023

Section: 21262 (DATA)	Section: 28386 (3DTA)
MWF 1:55 – 2:45 PM	MWF 3:00 – 3:50 PM

Contact Information

Instructor	Teaching Assistants
Name: Brett Presnell	Section: 21262 (DATA)
Email: presnell@ufl.edu	Name: Zhumengmeng Jin
Web: https://www.stat.ufl.edu/~presnell/	Email: z.jin@ufl.edu
Office: FLO 225	Virtual Office: Zoom 956 3662 3618
Virtual Office: Zoom 940 1233 3509	Office Hrs: Tu,Th 3–3:30 PM
Office Hrs: 11-12 a.m. Tu,Th (Zoom)	
	Section: 28386 (3DTA)
	Name: Matias Shedden
	Email: sheddenmatias@ufl.edu
	Office: FLO 234
	Office Hrs: MWF 12:45–1:45 PM

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. 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. Formulate statistical models in the R language.

4. Perform, document, and interpret common statistical analyses.

5. Carry out statistical/probabalistic simulations.

6. Determine statistical graphics appropriate to a statistical analysis and produce them using R.

7. Document and report the results of data analyses and simulations in a reproducible way.

Text Books and Other 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. Most readings will be taken from the following (free, on-line) texts, which students are encouraged to peruse on their own:

- r4ds : R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Wickham and Grolemund 2016)

- rp4ds : R Programming for Data Science (Peng 2016)

- hopr : Hands-On Programming with R : Write Your Own Functions and Simulations (Grolemund 2014)

- ggplot2 : ggplot2: Elegant Graphics for Data Analysis (Wickham, Navarro, and Pedersen 2022)

- advr : Advanced R (2nd Ed) (Wickham 2019)

Course Policies

Unless otherwise announced, the instructor's office hours will be held online via Zoom.

Grading

There will be several online quizzes to help you refine your knowledge and understanding of the course material. Homework assignments and projects will put this knowledge to use. There will be two exams during the semester to test your understanding of course concepts. These will be weighted in the final course average (percentage) as follows:

- 50% Homework/Projects
- 15% Quizzes
- 35% Exams

Letter grades in the course will be determined from the final course average according to the following scale:

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

Further information may be found in the university's grades and grading policies.

Exams

Makeup exams will be given only for excused absences as determined by the university's policies on attendance. Please read in particular the policies on absences and the policies on illnesses. Students must notify the instructor as early as possible prior to the exam of any planned absence or illness. For unplanned absences because of accidents or emergency situations, students should contact the instructor as soon as conditions permit. Documentation of excused activities and illnesses will be required before any makeup exam is given.

Projects and Homework Assignments

Homework and projects must be submitted on time. Late assignments will be accepted in cases of documented emergency or illness. In all other cases, acceptance of late assignments will be at the discretion of the instructor and scores on such assignments will be reduced by 10% plus an additional 5% for each additional day between the due date and the time of submission.

Students are expected to work independently or in *assigned* groups, unless otherwise specified in writing by the instructor. 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.

Accommodations for Students with Disabilities

Students requesting accommodation for disabilities must register with UF's Disability Resource Center. The DRC 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/.

Recording of Class Sessions

Some 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 course. It may be altered or rearranged to adapt to the backgrounds, abilities, and interests of the students in the class.

Week 1

- Getting started
- Vectors and Vectorized Operations
- Distributions and Descriptive Statistics

Week 2

- Introduction to R Markdown

Week 3

- Writing Your Own Functions
- Matrices and Arrays

Week 4

- Lists
- Data frames (and tibbles)
- Importing and Exporting Data

Week 5

- Column and row operations on data frames

- Pipes and more operations on data frames

- Joining/Merging Data Frames

Week 6

- Dates and times in base R

- The lubridate package

Week 7

- Tidy Data and Pivoting

- Character strings and the stringr package

- String matching with regular expressions

Week 8

- Exam 1 (Monday, February 27, 2023)

- Detecting string matches

- Extracting string matches

Week 9

- String replacement and string splitting

- An extended example with character strings

- Introduction to Data Scraping

Week 10

- Factors in base R

- The forcats package

Week 11

- Elementary statistical inference

- Simple linear regression

- Multiple regression

Week 12

- Factors and dummy variables in regression
- Interactions
- Simple logistic regression

Week 13

- Multiple logistic regression
- More graphics in R
- More on data scraping

Week 14

- Working with lists: the purrr package

Week 15

- Catch up and review
- Exam 2 (Wednesday, April 26, 2023)

References

Grolemund, Garrett. 2014. *Hands-on Programming with r: Write Your Own Functions and Simulations*. Sebastopol, CA: O'Reilly Media, Inc. https://rstudio-education.github.io/hopr/.

Peng, Roger D. 2016. *R Programming for Data Science*. 5+ ed. Lulu.com. https://bookdown.org/rdpeng/rprogdatascience/.

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/.

Wickham, Hadley, Danielle Navarro, and Thomas Lin Pedersen. 2022. *Ggplot2: Elegant Graphics for Data Analysis*. 3rd ed. Springer. https://ggplot2-book.org/.