

Special Topics: Data analysis using R

Date: 05 August 2019

Instructor: Dr. Robert Leaf

Office: GCRL Oceanography 119

Office Hours: Friday A.M. and by appointment

Email: robert.leaf@usm.edu

Course Meeting Day and Time: MW, 4:00 to 5:15 PM

Course Description and Objectives

This course examines the fundamental concepts and techniques for programming in the R statistical programming language. Contemporary data analysis, data manipulation, data visualization, and reproducible research necessitates command of quantitative tools. Although there are many specialized and general programming languages, the R programming language offers exceptional utility for analysis and is used widely in academia, industry, and by federal and state scientific groups. The demand for skilled data analysis practitioners is rapidly growing and this course prepares you to tackle real-world data analysis challenges.

The primary components of the course:

- 1) Introduce the basics of R programming using real world data sets

The course will introduce stereotypical programming concepts, in particular code modularisation, writing and using functions, and code re-usability. We will focus on understanding software engineering concepts such as project build and code testing. Participants will establish a working knowledge of R, R Studio, and relevant packages

- 2) Review aspects of project organization

A typical data analysis project involves several many components, each including several data files and different binary scripts with code. Keeping these files organized can be challenging and requires a suite of analytical tools.

- 3) Perform operations on vectors and understand how to use advanced functions

Learn how to wrangle, analyze and visualize data using base R operations and specialized packages (e.g. tidyverse and ggplot2)

- 4) Promote a reproducible research workflow

Finally, we will examine how to write markdown documents for high throughput data presentation which permits you to incorporate text and code into a document.

At the conclusion of this course:

Students will be able to recognize problems that can be solved using statistical programming and reproducible research approaches. The skills of sharing, automation, and organization enable making research more reproducible. By practicing and reinforcing the use of quantitative tools, participants will be better able to make insights that would otherwise be hidden.

Course Materials

R for Data Science by G. Grolemund and H. Wickham (<https://r4ds.had.co.nz/>)

Course Workload Statement

Students are expected to invest considerable time outside of class in learning the material for this course. The expectation of the University of Southern Mississippi is that students should spend approximately 2 to 3 hours outside of class each week for every hour in class working on reading, assignments, studying, and other work for the course. Time management is thus critical for student success. All students should assess their personal circumstances and talk with their advisors about the appropriate number of credit hours to take each term. Resources for academic support can be found at <https://www.usm.edu/success>.

Course Evaluation

Percentage	Letter Grade
93-100	A
90-92	A-
86-89	B+
83-85	B
80-82	B-
76-79	C+
73-75	C
70-72	C-
66-69	D+
63-65	D
60-62	D-
< 60	F

Grading scale

Evaluation type	Number	Points per item	Total points
HW	5	20	100
Mid-semester Project	1	50	50
Final Project	1	100	100

Academic Support Resources

If a student believes that they have a disability which is covered by the Americans with Disabilities Act (ADA) and makes them eligible to receive classroom accommodations, they should contact the Office for Disability Accommodations (ODA) for information regarding the registration process. Disabilities covered by the ADA may include but are not limited to ADHD, learning disabilities, psychiatric disabilities, physical disabilities, chronic health disorders, temporary illnesses or injuries and pregnancies. Students should contact ODA if they are not certain whether their documented medical condition qualifies for ODA services. Students are only required to disclose their disability to the Office for Disability Accommodations. All information submitted to ODA by the student is held with strict confidentiality.

Mental Well-Being Statement

I recognize that students sometimes experience challenges that make learning difficult. If you find that life stressors such as anxiety, depression, relationship problems, difficulty concentrating, alcohol or drug problems, or other stressful experiences are interfering with your academic or personal success, consider contacting Student Counseling Services on campus at 601-266-4829. More information is also available at <https://www.usm.edu/student-counseling-services>. All students are eligible for free, confidential individual or group counseling services. In the event of emergency, please call 911 or contact the counselor on call at 601-606-HELP (4357).

Nondiscrimination Statement

The University of Southern Mississippi offers to all persons equal access to educational, programmatic and employment opportunities without regard to age, sex, sexual orientation, disability, pregnancy, gender identity, genetic information, religion, race, color, national origin, and/or veteran status pursuant to applicable state and federal law.

Confidentiality and Mandatory Reporting

As an instructor, one of my responsibilities is to help create and maintain a safe learning environment. I have a mandatory reporting responsibility related to my role as a faculty member. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on USM's campus with certain University officials responsible for the investigation and remediation of sexual misconduct. The information will remain private and will only be shared with those officials necessary to resolve the matter. If you would like to speak in confidence, resources available to students include Confidential Advisors with the Shafer Center for Crisis Intervention, the Counseling Center, and Student Health Services. More information on these resources and University Policies is available at <https://www.usm.edu/sexual-misconduct>.

Course Scheduling

Day	Activity	Readings and Assignments
Wednesday, January 22, 2020	Goals and objectives for the course	Leaf Shortcourse
Monday, January 27, 2020	Introduction to R and R Studio	Leaf Shortcourse
Wednesday, January 29, 2020	Programming Syntax 01	Leaf Shortcourse, HW01
Monday, February 03, 2020	Programming Syntax 02	Leaf Shortcourse
Wednesday, February 05, 2020	Programming Syntax 03	Leaf Shortcourse, HW02
Monday, February 10, 2020	Programming Syntax 04	Leaf Shortcourse
Wednesday, February 12, 2020	Programming Syntax 05	Leaf Shortcourse
Monday, February 17, 2020	Project organization - projects in R Studio	Leaf Shortcourse
Wednesday, February 19, 2020	Vector functions - tidyverse 01	R for Data Science “Explore”
Monday, February 24, 2020	Mardi Gras Holiday	
Wednesday, February 26, 2020	Vector functions - tidyverse 02	R for Data Science “Program”
Monday, March 02, 2020	Mid-semester project presentations	
Wednesday, March 04, 2020	Vector functions - tidyverse 03	R for Data Science “Wrangle”
Monday, March 09, 2020	Vector functions - tidyverse 04	R for Data Science “Model”, HW03
Wednesday, March 11, 2020	Data visualization 01	Tutorial, ggplot2,
Monday, March 16, 2020	Spring Break Holiday	
Wednesday, March 18, 2020	Spring Break Holiday	
Monday, March 23, 2020	Data visualization 02	Tutorial, ggplot2
Wednesday, March 25, 2020	Data visualization 03	Tutorial, ggplot2
Monday, March 30, 2020	Data visualization 04	Tutorial, ggplot2, HW04
Wednesday, April 01, 2020	Introduction to Reproducible Research	R for Data Science “Communicate”
Monday, April 06, 2020	Reproducible research in Rmarkdown 01	
Wednesday, April 08, 2020	Reproducible research in Rmarkdown 02	
Monday, April 13, 2020	Reproducible research in Rmarkdown 03	
Wednesday, April 15, 2020	Easter Holiday	
Monday, April 20, 2020	Introduction to Shiny 01	Shiny Tutorial
Wednesday, April 22, 2020	Introduction to Shiny 02	Shiny Tutorial
Monday, April 27, 2020	Introduction to Shiny 03	Shiny Tutorial
Wednesday, April 29, 2020	Introduction to Shiny 04	Shiny Tutorial, HW05
Monday, May 04, 2020	Final Project Presentation in-class work	
Wednesday, May 06, 2020	Final Project Presentation in-class work	
Monday, May 11, 2020	Final Exam Week - Project Presentation	All HW Due

Academic Integrity

All students at the University of Southern Mississippi are expected to demonstrate the highest levels of academic integrity. Forms of academic dishonesty include cheating (including copying from others’ work), plagiarism (representing another person’s words or ideas as your own; failure to properly cite the source of your information, argument, or concepts), falsification of documents, disclosure of or use of test material or other assignment content to another student, submission of the same paper or other assignment to more than one class without the explicit approval of all faculty members involved, unauthorized academic collaboration with others, conspiracy to engage in academic misconduct.

Engaging in any of these behaviors or supporting others who do so will result in academic penalties and/or other sanctions. If a faculty member determines that a student has violated our Academic Integrity Policy, sanctions ranging from resubmission of work to course failure may occur, including the possibility of receiving a grade of “XF” for the course, which will be on the student’s transcript with the notation “Failure due to academic misconduct.”

Course algorithm

```
print.stat.1 <- "You should consider enrolling in this Special Topics course"
print.stat.2 <- "we will cover many fundamental data topics
                relevant to your science and data workflow"

if (user == "Wants to improve their data analysis skills.") {
  print(print.stat.1) }
if (user == "Is a Python/Matlab/SQL/Excel User and that is good enough for me.") {
  print(paste0(print.stat.1, ", because ", print.stat.2))
}
```