STAT 542, Computing for Data Science – Spring 2025

Instructor:
David Hitchcock, professor of statistics
202 LeConte College
Phone: 803-777-5346
Email: hitchcock@stat.sc.edu
Course Web Page: http://people.stat.sc.edu/hitchcock/stat542.html
(Also accessible via Blackboard)

Class Meeting Times: MWF 9:40 - 10:30 am, LeConte College 103

Office Hours: Monday, Wednesday, Friday 10:45-11:45 a.m., Tuesday 9:30-10:30 a.m., or by appointment

Textbook: *Modern Data Science with R*, 2nd edition, by Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton. CRC Press, 2021.
  Online version available at: https://mdsr-book.github.io/mdsr2e/

Prerequisite: A grade of C or higher in STAT 301 or STAT 509 or STAT 515.

Course Outline: Chapters 2-8, 14-16 and (time permitting) 19 of the Baumer, Kaplan, and Horton textbook. Topics covered include data visualization, the grammar of graphics, data wrangling, tidy data, iteration in R, data ethics, interactive data graphics, database querying and administration with SQL, and (time permitting) analysis of text data.

Learning Outcomes:
Upon successful completion of this course, students should be able to:
- Perform basic and interactive data visualizations and graphics
- Understand the concepts of tidy data and the grammar of graphics
- Undertake data wrangling, database querying and database administration
- Recognize ethical questions involved with data collection and management
- Be able to effectively communicate in written and/or oral formats the information obtained from a data analysis

Grading: The course grade will be based on attendance (7%), homework average (10%), the two midterm exams (25% each), the project (8%), and the final exam (25%). The overall course average will result in the following grades: 90-100 = A, 87-89 = B+, 80-86 = B, 77-79 = C+, 70-76 = C, 67-69 = D+, 60-66 = D, 59 and below = F.

Class Lectures / Attendance Requirement:
You are urged and expected to attend the lectures live on MWF 9:40 - 10:30 am in LeConte Room 103. If you are forced to miss a class or if you would like to review material from a class, I plan to ATTEMPT TO record the lectures on Zoom and post them in Panopto in the Blackboard STAT 542 course page. This is not a replacement for attendance and some aspects of the class will not show up on the recorded version. In addition, I make **no guarantees** that the recordings will always be successful. In short: Please come to class!

Since this is an in-person class, you are expected to attend at least 80% of the class sessions in person. Attendance will be taken each class, and your grade on the attendance component will be 1.25 times the percentage of class sessions that you attend live (with a maximum of 100% for the attendance grade). For example, if you attend 60% of the class sessions in person, your attendance grade (which

is 7% of the overall course grade) will be 75%. If you attend 80% or more of the class sessions in person, your attendance grade will be 100%.

Homework: Homework problems will be assigned periodically. You may work with other students in this class on these problems, but you should write your answers independently. Test problem(s) will sometimes be similar in nature to assigned homework problems. Therefore you are personally responsible for knowing how to solve the homework problems!

Use of AI: Use of AI programs such as ChatGPT for out-of-class work is strongly discouraged since the material generated by these programs may be inaccurate, incomplete, or otherwise problematic, and work turned in with such flaws will receive a poor grade. Using AI on assignments will also likely hinder your own independent thinking, creativity, and learning. You will better prepare yourself for the in-person exams by working (and perhaps struggling) with the out-of-class assignments on your own rather than by using AI as a crutch. *If you include material generated by an AI program, you must cite it like any other reference material* (with due consideration for the quality of the reference, which may be poor).

Graduate Students: The university requires that 500-level classes be more rigorous for graduate students than for undergraduates. Homeworks will occasionally involve extra problems that will be required only for graduate students. These will be graded separately for graduate students, and a grade of 70% or higher must be earned on the combined extra sections across the problem sets or a letter grade (A→ B, B+→ C+, etc…) penalty will be applied to the final course grade.

Disabilities: Any student with a documented disability should contact the Student Disability Resource Center at 777-6142 to make arrangements for appropriate accommodations.

Exams: There will be two midterm exams and a final exam. All these exams will consist of multiple-choice (and possibly short-answer) questions about the concepts studied in the class, plus a short free-response coding problem. The midterm exams will be given in the classroom during the regularly scheduled class time. If you have a documented reason for missing an exam, you must contact the instructor ahead of time (or at that first possible opportunity) to make arrangements to take the exam.

Data Science/Computing Project: There will be a required group project that will be worked on outside of class during the semester and will be completed near the end of the semester. Groups should discuss their project topics ahead of time with the instructor; examples of possible project ideas include the collection, wrangling, and manipulation of some data set of interest and using computing tools to extract and present valuable information about the data. Students will turn in a typed project report as a group, and groups will make an oral presentation to the class near the end of the semester. The last few class periods will be devoted to the groups' oral presentations.

Computing: Use of a computer is required for this computing class. The examples in class will be done using R. No previous knowledge of R is assumed. Everyone is encouraged to download a free copy of R (see the course page for downloading instructions). All necessary analyses will be able to be done in R, at least --- you may use other software if it accomplishes the appropriate task and if you know how to use it.

There may be occasional opportunities during the semester to work on coding in class with other students and the instructor, so I encourage you to bring a laptop to class when possible.

During Class: No cell phones may be on during class. In the classroom, laptops and tablets (e.g., iPads) may be used only for note-taking and other course-related activities. Students may not use laptops and tablets to look at unrelated web pages, play games, etc.