STOR 320-001: Introduction to Data Science
Fall 2024

Instructor: Dr. Yao Li
E-mail: yaoli@email.unc.edu
Office: Hanes 334
Office Hours: W 2:00 – 4:00 PM

Assistant: Morgan Smith
E-mail: smithmor@email.unc.edu
Office: Zoom
Office Hours: T 10:00 – 11:00 AM, Th 3:00PM – 4:00PM

Yuhao Zhou
E-mail: yuhaoza@live.unc.edu
Office: Hanes B50
Office Hours: M 10:00 – 11:00 AM, Th 2:00PM – 3:00PM

Lectures: TTH 8:00AM – 9:15AM Hanes 120

Labs: 320.400 by Morgan Smith: W 3:30PM – 4:20PM, HN107
320.401 by Morgan Smith: W 5:00PM – 5:50PM, CH104
320.402 by Yuhao Zhou: F 10:10AM – 11:00AM, HN107
320.403 by Yuhao Zhou: F 3:30PM – 4:20PM, DE203

Course URL: Website: https://liyao880.github.io/stor320/

Description: This course is an application-driven introduction to data science. Statistical and computational tools are valued throughout the modern workplace from Silicon Valley startups to marine biology labs, to Wall Street firms. These tools require technical skills such as programming and statistics. They also require professional skills such as communication, teamwork, problem solving, and critical thinking.

You will learn these tools and hone these skills through hands-on experience working with datasets provided in class and downloaded from certain public websites. During the first part of the semester, we will focus on R programming skills and data visualization. Later topics will include exploratory data analysis, data wrangling, modeling, and effective communication of results.

Plan to come to every class with your computer and ready to work with others. Using resources around you is a key component of successful data analysis. This includes the internet and people.

Textbook: **R for Data Science**, *by Hadley Wickham*.

available free online https://r4ds.had.co.nz/

Prerequisites: STOR 155 or an equivalent introductory statistics course.

Final Grade: Class Participation (5%)
Labs (20%)
Homework (45%)
Final Project (30%)

Participation: Participation points can be earned by answering two questions during class. To ensure your participation is recorded, please send me an email after class with your name and the questions you answered. While I won't be able to reply to these emails, rest assured that I will update the participation spreadsheet, which is available on the class website.

Starting from the lecture after fall break, once you have answered two questions in class, you will have the opportunity to earn extra credit by asking or answering additional questions. For each question, you will earn 0.5 extra credit points. You can earn up to 5 extra credit points by participating in up to 10 additional questions.

Homework: Homework will be based on problems from the course textbook, *R for Data Science*. Each homework assignment will be worth 20 points. Data analysis homework are constructed using customized problems from real life data sets. Each analysis will be worth 40 points. These analyses allow you to practice the techniques learned from the course.

- You may discuss homework with classmates and teaching staff. But you must submit your own work.
- You may and often should search online for solutions to coding problems. This is perfectly fine and encouraged.
- However, copying responses from students who have taken the course, including from sources online, is unacceptable and could be treated as an honor code violation.
- Homework must be submitted as the **HTML** output from an R Markdown file on Canvas. In other words, your homework submission must be a .html file with all code and writing, as produced in R Markdown. Submissions that do not 'knit' to html will not be accepted. Such cases most often result from errors in the code, which students must correct before submission.
- Late homework or a failure to adhere to the rules above will result in a score of zero for that assignment.

Labs: **Attendance to all labs is mandatory**. Each week, your lab instructor will take attendance. During the lab session, students will be required to complete a lab assignment, which must be submitted no later than 30 minutes after the lab ends.

Each lab assignment will be based on the topics discussed in lecture or related to your final project. Students are responsible to turn in their own labs but are encouraged to work in teams and help each other. A lab instructor will help students in the completion of the lab and to facilitate group work. Each lab is worth 20 points, and late submissions will not be accepted. Lab assignments must be submitted as **HTML** output from an R Markdown file via Canvas. **At the end of the semester, the lowest lab grade will be dropped.**

Final Project: The final project is done in groups of **5** and worth a total of 100 points. There will be **4 parts** of varying point values submitted throughout the semester.

- Part I: **Project Proposal**, is worth **10 points** and will be due on the designated date (find it on the class website).
- Part II: **Exploratory Data Analysis**, is worth **20 points** and must be submitted on Canvas by the designated date (find it on the class website).
- Part III: **Final Paper**, is worth **40 points** and must be submitted on Canvas by the designated date (find it on the class website).
- Part IV: **Final Presentation**, is worth **30 points** and will take place during the last three lectures**.** Slides must be submitted before the presentation.

Grade Scale: Your final grade is based on a weighted average according to the previously addressed breakdown. Curving on individual/group assessments should not be expected. A curve may be applied to the final grades depending upon the class average. Conversion to a letter grade will be based on the table below:

A	94 to 100	B	83 to 86.99	C	73 to 76.99	D	60 to 66.99
A-	90 to 93.99	B-	80 to 82.99	C-	70 to 72.99	F	0 to 59.99
B+	87 to 89.99	C+	77 to 79.99	D+	67 to 69.99		

These are hard break lines and no rounding will be applied to push an individual student up to a more desirable letter grade.

Lectures: Students can earn class participation grades (5 points in total) by answering questions in class or asking questions in final presentation (2.5 points each time and 5 points at most).

Core programming and data science skills

- R Markdown
- data frame creation and manipulation
- summary statistics
- visualization
- exploratory data analysis
- 'tidy' and relational data
- functions and functional programming
- string manipulation and regular expressions

Modeling

- cross-validation
- linear and generalized linear models
- classification techniques
- clustering

Advanced topics

- Shiny

- more advanced modeling with support vector machines and tree-based methods
- web scraping

Attendance: Regular class attendance is a student obligation, and a student is responsible for all the work, including tests and written work, of all class meetings. No right or privilege exists that permits a student to be absent from any class meetings except for excused absences for authorized University activities or religious observances required by the student's faith. If a student misses three consecutive class meetings, or misses more classes than the course instructor deems advisable, the course instructor may report the facts to the student's academic dean. (See details at https://catalog.unc.edu/policies-procedures/attendance-grading-examination/#text)

AI Tools: All homework and analysis assignments must be completed individually. Assistance from other students, AI tools (e.g., ChatGPT), or using previously uploaded work from other sources (e.g., CourseHero) is strictly prohibited. This policy also applies to project work; AI tools are not allowed to aid in the completion of any projects. Violations of this policy will result in a grade of 0 for the assignment or project. Additionally, any alleged violations will be reported to the University of North Carolina (UNC) for further review and potential disciplinary action.

Honor Code: http://instrument.unc.edu/

Accessibility: https://ars.unc.edu/

Counseling: https://caps.unc.edu/

Title IX: Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance (Adrienne Allison – Adrienne.allison@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.