STOR 320-002: Introduction to Data Science
Spring 2021

Instructor: Dr. Yao Li
E-mail: yaoli@email.unc.edu

Office Hours: Wednesday and Friday 9:00AM – 10:00AM

Assistant: Taylor M Petty
E-mail: tmpetty@live.unc.edu
Office Hours:

Pavlos Zouboulоglou
E-mail: pavlos@live.unc.edu
Office Hours:

Sam Booth
E-mail: slbooth@live.unc.edu
Office Hours:

Lectures: Monday, Wednesday and Friday 8:00AM – 8:50AM

Labs: Sam Booth
Taylor M Petty
Pavlos Zouboulоglou

Course URL: Website: https://liyao880.github.io/stor320/
Assignment Submission: https://sakai.unc.edu/ and login with your Onyen

Zoom Links: Due to the pandemic, lectures and office hours will be hosted live online via Zoom. Lectures will also be recorded and linked to course website.

Lectures: https://unc.zoom.us/j/93501182856
Lab 320.400:
Lab 320.401:
Lab 320.402:

Instructor Office Hours: https://unc.zoom.us/j/99751369535
Taylor's Office Hours:
Pavlos's Office Hours:
Sam's Office Hours:

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: Lab Attendance (10%)
Labs (15%)
Homework (45%)
Final Project (30%)

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 Sakai. 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 submitted less than 24 hours from when it was due will have its score reduced 50%.
- Homework later than 24 hours 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. Every week, your lab instructor will take attendance. If you are there for the entire class, you will receive 10 points. At the end of the semester, the lab attendance grades will be curved by 10 points allowing you to miss a single lab and receive a 100% on your lab attendance grade. If you show up to every lab, you will get above 100% on your lab attendance grade.

During the lab session, students are required to complete a lab assignment that will be due 30 minutes after the lab ends.

- Deadline for 320.400:
- Deadline for 320.401:
- Deadline for 320.402:

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 be provided to help students in the completion of the lab and to facilitate group work. Every lab is worth 10 points and no late lab assignments will be accepted. Lab assignment must be submitted as the **HTML** output from an R Markdown file on Sakai.

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 **February 17**.
- Part II: **Exploratory Data Analysis**, is worth **20 points** and must be submitted on Sakai by **11:55PM** on **March 26**.
- Part III: **Final Paper**, is worth **40 points** and must be submitted on Sakai by **11:55PM** on **Wednesday, May 5**.
- Part IV: **Final Presentation**, is worth **30 points** and will take place during the last three lectures**.** Slides must be submitted before **11:55PM** on **April 29h**.

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: 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

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

Community Standards in Our Course and Mask Use:

This fall semester, while we are in the midst of a global pandemic, all enrolled students are required to wear a mask covering your mouth and nose at all times in our classroom. This requirement is to protect our educational community — your classmates and our instructional assistant – as you learn together. If you choose not to wear a mask, or wear it improperly, the instructional assistant will ask you to leave immediately submit a report to the Office of Student Conduct. At that point you will be disenrolled from this course for the protection of our educational community. An exemption to the mask wearing community standard will not typically be considered to be a reasonable accommodation. Individuals with a disability or health condition that prevents them from safely wearing a face mask must seek alternative accommodations through the Accessibility Resources and Service. For additional information, see Carolina Together.

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.