STOR 320-001
Introduction to Data Science
Fall 2022

Course 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, web scraping, 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.

Course Goals and Learning Objectives

This course will enable you to:

- Establish proficiency in the statistical programming language R making the student immediately competitive in the data science job market.
- Acquire both structured and unstructured data for the purpose of gathering insight on well-crafted research hypotheses.
- Clean, transform, and merge datasets.
- Summarize data using professionally developed tables and various visualizations.
- Model relationships between variables using a variety of techniques, including linear regression, nonlinear regression, logistic regression, and various machine learning techniques.
- Evaluate and compare predictive accuracy from competing modeling methods using cross-validation.
- Honestly interpret results from predictive analytics using creative visualizations and tables.
- Effectively communicate insights both verbally and in writing to a non-technical audience.
- Share data science ideas with a worldwide audience using HTML web pages and dynamic web applications using RMarkdown and RShiny.

As part of the General Education curriculum, this course will satisfy the following focus capacities:

Quantitative Reasoning

Students learn to comprehend and apply mathematical concepts in authentic contexts, developing tools for reasoning with data, logic, and quantitative methods.

Number	STOR 320.001
Title	Introduction to Data Science
Credit hours	4
Course Format	In-person unless otherwise advised. In the case of a transition to a virtual environment, lectures will continue synchronously via zoom. Lectures will be supplemented with in-class programming and practical discussion. Students will also be required to register for a weekly lab with required attendance in person.
Prerequisites	STOR 120, STOR 155 or Exemption
Target Audience	1st or 2nd year undergraduates seeking a quantitative reasoning (QR) course with a required experimental investigation lab (EIL) and interested in learning about the process of data science including data acquisition, data visualization, data analysis, and technical communication.
Instructor	Andy Ackerman Office: Hanes B-7 Email: atacker@email.unc.edu Office Hours: TWTH 3:30-4:30 pm
Lab Instructors	Parvathi Meyyapan Email: pmeyyappan@unc.edu Office Hours: Announced in Lab Michael Nisenzon Email: mnisenzon@unc.edu Office Hours: Announced in Lab
Course Website	https://sakai.unc.edu/portal/site/70fc1af5-d61c-442d-85be-01496ed4f996/page/10800003-46eb-491c-be36-98178803504e
Lecture Details	MWF 2:30-3:20 pm; Hanes 120

Course Texts	*R for Data Science*, Hadley Wickham. Legally free online, but can be purchased for less than $40 on Amazon.

Empirical Investigation Lab

One Focus Capacity course must include or be associated with a one-credit Empirical Investigation Lab. In such labs, students participate in measurement, data collection and analysis, and hypothesis testing connected to the course content. An Empirical Investigation Lab is not usually a separate class; ordinarily it is a fourth credit attached to another Focus Capacity class.

1. Take empirical measurements using appropriate apparatus.
2. Generate and test hypotheses.
3. Gather, store, and organize data.
4. Analyze and report on data and hypothesis testing.

Course Assessments

Assignments	Percentages	Frequency
Labs	25%	1/Week
Homework	15%	1/Week
Analyses	25%	4 Total
Final Project	30%	
Group Involvement	5%	

Grading 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	93 to 100	B	83 to 86.99	C	73 to 76.99	D	60 to 66.99
A-	90 to 92.99	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		

Assignment Descriptions

Labs:

Attendance to all labs is mandatory. Every week, your lab instructor will take attendance. During this period, students are required to complete a lab assignment. 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. These assignments are to be completed using RMarkdown and submitted as an HTML file on Sakai by 5:00PM on the day of the lab. A lab

instructor will be provided to help students in the completion of the lab and to facilitate group work. Take heed, no late lab assignments will be accepted. You will need to get a university excused absence to prevent a loss of points in these weekly labs if you miss class.

Homework:
Homework will be based on problems from the course textbook, *R for Data Science*. You can work with each other on this assignment, but the work you submit should be your own. Any copying of solutions will result in a 0 if caught cheating. These assignments are to be completed using RMarkdown and submitted as an HTML file on Sakai. No late homework assignments will be accepted.

Analyses:
Analyses are constructed using customized problems from real life data sets. These analyses allow you to practice the techniques learned from lab assignments. You should not work with any other student on these assignments or you will receive a 0 if caught cheating. These analyses are to be completed using RMarkdown and submitted as an HTML file on Sakai. If you submit your analyses late, expect a 25% deduction for less than 1 day late, 50% deduction between 1 and 2 days late, 75% deduction between 2 and 3 days late, and 100% deduction more than 3 days late. All late assignments must be submitted to Sakai as soon as they are completed.

Final project:
The final project is done in groups of 4-5 students and worth a total of 100 points. There will be 4 parts of varying point values submitted throughout the semester. The first part, the Project Proposal, will be due sometime in the beginning of the semester after groups have been designated. The second part, the Exploratory Data Analysis, will be due approximately in the middle of the semester after the Project Proposal has been completed. The third part, the Final Paper, must be submitted on Sakai by 11:59 PM on Monday, November 21. The fourth part, the Final Presentation, will take place during our designated final exam time according to the university calendar. Slides must be submitted by 8AM on the day of the final to Sakai.

Group Involvement:
Since the final project is a group project that is worth a tremendous amount of points, it is very important that each group member fulfills their obligation to their group. Four times during the semester, there will be a survey sent out to the class, where you will score your group members on a scale from 0 (Terrible) to 5 (Excellent).

Rough Course Outline (updated 9/3/22)

Class	Day	Date	Topics	Assignments Due (Roughly)
1	M	15-Aug	Syllabus and Introduction to R	
2	W	17-Aug	Data Visualization	
3	F	19-Aug	Data Visualization	
4	M	22-Aug	Data Wrangling	HW 1
5	W	24-Aug	Data Wrangling/Final Project 1	

6	F	26-Aug	Data Wrangling	
7	M	29-Aug	Exploratory Data Analysis (EDA)	
8	W	31-Aug	EDA	HW 2
9	F	2-Sept	EDA	
	M	5-Sept	*Labor Day*	
10	W	7-Sept	Programming	HW 3
11	F	9-Sept	Programming	Analysis 1
12	M	12-Sept	Programming	
13	W	14-Sept	Programming	
14	F	16-Sept	Tidy Data	HW 4
15	M	19-Sept	Tidy Data/ Final Project (2)	Proposal
16	W	21-Sept	Joins	
17	F	23-Sept	Joins	
	M	26-Sept	*Wellness Day*	
18	W	28-Sept	Modelling 1	
19	F	30-Sept	Modelling 1	HW 5
20	M	3-Oct	Modelling 1	Analysis 2
21	W	5-Oct	Modelling 1	
22	F	7-Oct	Modelling 2/ Final Project (3)	HW 6
23	M	10-Oct	Modelling 2	
	W	12-Oct	*University Day*	
24	F	14-Oct	Modelling 2	EDA
25	M	17-Oct	Modelling 2	
26	W	19-Oct	Modelling 3	HW 7
	F	21-Oct	*Fall Break*	
27	M	24-Oct	Modelling 3	
28	W	26-Oct	Modelling 3	Analysis 3
29	F	28-Oct	Modelling 3	
30	M	31-Oct	Modelling 4	HW 8

31	W	2-Nov	Modelling 4	
32	F	4-Nov	Modelling 4	
33	M	7-Nov	Modelling 4	HW 9
34	W	9-Nov	Modelling 5	
35	F	11-Nov	Modelling 5	Analysis 4
36	M	14-Nov	Modelling 5	
37	W	16-Nov	Modelling 6	
38	F	18-Nov	Modelling 6	HW 10
39	M	21-Nov	Modelling 6	**Final Paper**
	W	23-Nov	*Thanksgiving*	
	F	25-Nov	*Thanksgiving*	
40	M	28-Nov	Data Ethics	
41	W	30-Nov	Data Ethics	
			Final Exam	**Oral Presentations**

Course Policies and Resources

Teaching Philosophy	My general approach aims to make two principles abundantly clear: there are no (sincere) bad questions, and education is only one facet of your life. To the first point, if you ask me a question out of genuine curiosity or confusion, I guarantee to respect and attend to it regardless of the level of complexity. Note, I do not claim to always have the answer, and certain questions may require an outside (office hours) conversation. But my guarantee is to provide an answer when I am able and search with you when unable. The flip side of this guarantee is your willingness to keep me accountable. Ask questions. If you are confused, please reach out for help. If in-class questions are too daunting, take advantage of office hours, tutorial sessions, or electronic communication. As to the second point, while I am ultimately a very ardent advocate of the merits of education, I want to reiterate that no single facet of your life – perhaps least of which performance in STOR 320 – fully defines you as a person. This is not to say that this, or any class, is unimportant, rather that this class is meant to enable and instruct not to define.

Honor Code	It is my aim to make my expectations for acceptable and original work as clear and explicit as possible. Gray areas frustrated me as a student, so I aim to prevent them as an instructor. To that end, students are bound by UNC's honor code in taking exams and in written work, and the submission of said work signifies understanding and acceptance of those requirements. More specific to this course, collaboration is not only permitted but *encouraged* on homework and projects. Analyses and labs, by contrast, will be completed individually. I recognize this can be immensely stressful, and I do not intend to add to that burden. That said, my job is to cultivate and assess your *personal* understanding. Insofar as individual assignments serve this purpose, they will be occasionally used. Please consult with me if you have any questions about the honor code. Asking for clarification in advance is far preferred to asking for forgiveness post-facto.
COVID-19	Each of us has a responsibility to know and act on these standards and policies in a way that maximizes a safe and healthy environment for us to teach, work, learn and live. To this end, we are developing a set of community standards and policies for our students, faculty, staff and visitors. We are all in this together, and we believe that together, we can face the challenges presented by COVID-19 with resilience, determination and great support for our community. See https://carolinatogether.unc.edu/community-standards-3-2/ for a list of guidelines that we all need to follow to reduce the spread of COVID-19. See the Carolina Together Roadmap at https://carolinatogether.unc.edu/ for more information on the University's plans regarding COVID-19. This website is continuously updated and should be checked weekly. More specifically, in accordance with the current iteration of these policies, masks will not be mandatory in class.
Accessibility Resources	UNC-Chapel Hill facilitates the implementation of reasonable accommodations for students with learning disabilities, physical disabilities, mental health struggles, chronic medical conditions, temporary disability, or pregnancy complications, all of which can impair

	student success. See the ARS website for contact and registration information: https://ars.unc.edu/about-ars/contact-us
Attendance Policy	Lecture attendance is not required. That said, class participation is fundamental to success in this class. You, the student, are responsible for filling information that was missed during absences.
University Testing Center	The College of Arts and Sciences provides a secure, proctored environment in which exams can be taken. The center works with instructors to proctor exams for their undergraduate students who are not registered with ARS and who do not need testing accommodations as provided by ARS. In other words, the Center provides a proctored testing environment for students who are unable to take an exam at the normally scheduled time (with pre-arrangement by your instructor). For more information, visit http://testingcenter.web.unc.edu/.
Counseling and Psychological Services	CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: https://caps.unc.edu/ or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.
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.
Technology Use	Students are required to bring their computer to every class and lab with a working copy of R and RStudio. Directions for free downloads of this software will be provided. The professor or lab assistant will occasionally request computers to be closed for dynamic discussion and guest speakers.
Legal	Andrew Ackerman reserves the right to make changes to the syllabus, including all due dates. These changes will be announced as early as possible so that students can adjust their schedules.