STOR 320-002
Introduction to Data Science
Spring 2023

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.

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 Details

Number	STOR 320-002
Title	Introduction to Data Science
Credit hours	4
Course Format	The course format will include 3 lectures per week. 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 or 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	Dr. Mario Giacomazzo Office: Hanes 134 Phone: 480-489-1398 (Cell) Email: mgiacoma@email.unc.edu Office Hours: TTh, 8AM-11AM
Lab Instructors	Ian Ferer (320.404, 320.405) Email: iferer@email.unc.edu Office Hours: TBD Charles Zhao (320.406, 320.407) Email: charlesz@unc.edu Office Hours: TW, 4:30PM – 5:30PM
Course Website	https://supermariogiacomazzo.github.io/STOR320_WEBSITE/
Class Days, Times, Location	MWF, 10:10AM – 11AM, Gardner 105
Lab Days, Times, Location	320.404: T, 9:30AM – 10:20AM, Hanes 107 320.405: Th, 9:30AM – 10:20AM, Hanes 107 320.406: T, 8AM – 8:50AM, Dey 209 320.407: Th, 8AM – 8:50AM, Alumni 205
Zoom Links	Dr. Mario Office Hours Ian Office Hours Charles Office Hours
Course Texts	*R for Data Science*, Hadley Wickham. Legally free online, but can be purchased for less than $40 on Amazon.

Course Assessments

Assignments	Percentages
Labs	25%
Homework	10%
Analyses	25%
Final Project	30%
Group Involvement	10%

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. If you are there for the entire class, you will receive 10 points. During this period, students are required to work on 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. These assignments are to be completed using RMarkdown and submitted as an HTML file on Canvas by 8:00AM on the following Tuesday. 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. You will need to get university approved absence and notify your lab instructor to prevent a loss of points in these weekly labs if you miss class. No late lab assignments will be accepted.

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. Each homework assignment will be worth 10 points. These assignments are to be completed using RMarkdown and submitted as an HTML file on Canvas. 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 cannot work with any other student on these assignments or you will receive a 0 if caught cheating. Each analysis will be worth 40 points. These analyses are to be completed using RMarkdown and submitted as an HTML file on Canvas. 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 Canvas as soon as they are completed.

Final project:
The final project is done in groups of 5-6 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, is worth 10 points and will be due sometime in the beginning of the semester after groups have been designated. The second part, the Exploratory Data Analysis, is worth 20 points and will be due approximately in the middle of the semester after the Project Proposal has been completed. The third part, the Final Paper, is worth 40 points and must be submitted on Canvas by 11:59PM on Friday, April 28. The fourth part, the Final Presentation, is worth 30 points and will take place during our designated final exam time according to the university calendar. For our class, this is 8:00AM to 11:00AM on Friday, May 5. Slides must be submitted by 8AM on May 5 to Canvas.

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).

Course Policies and Resources

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.
Etiquette for Zoom	These are my five expectations of you regarding Zoom. • Have your camera turned on. • Mute your microphone unless answering or asking a question. • Communicate by unmuting yourself or using the chat feature. • Be mindful of background noise when not muted. • Limit your distractions and avoid multi-tasking.
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	No right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:
	    1. Authorized University activities
	    2. Disability/religious observance/pregnancy, as required by law and approved by Accessibility Resources and Service and/or the Equal Opportunity and Compliance Office (EOC)
	
	Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students, Gender Violence Service Coordinators, and/or the Equal Opportunity and Compliance Office (EOC).

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.
Honor Code Statement	Students are bound by the Honor Code in taking exams and in written work. The Honor Code of the University is in effect at all times, and the submission of work signifies understanding and acceptance of those requirements. Plagiarism will not be tolerated. Please consult with me if you have any questions about the Honor Code.
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 | Dr. Mario 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. |