STOR 320-002
Introduction to Data Science
Fall 2020

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 three remote synchronous lectures. 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.
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: Wherever I Am Phone: 480-489-1398 (Cell) Email: mgiacoma@email.unc.edu Office Hours: T, 9AM-3PM
Lab Instructors	Pavlos Zouboulloglou (320.403) Email: pavlos@live.unc.edu Office Hours: M,9AM-10AM F,2:25PM-3:25PM Sam Booth (320.404) Email: slbooth@live.unc.edu Office Hours: TBD Kevin O'Connor (320.405 & 320.406) Email: koconn@live.unc.edu Office Hours: TBD
Course Website	https://supermariogiacomazzo.github.io/STOR320_WEBSITE/
Class Days, Times, Location	MWF, 1:20PM – 2:10PM, Remote
Lab Days, Times, Location	320.403: Th, 8:00AM – 8:50AM, Philips 365 320.404: Th, 8:00AM – 8:50AM, Genome Science 10 320.405: Th, 6:30PM – 7:30PM, Remote 320.406: Th, 7:30PM – 8:20PM, Remote
Zoom Links	Lecture: https://unc.zoom.us/j/95493131315 Dr. Mario Office Hours: https://unc.zoom.us/j/96354599519 Pavlos Office Hours: https://unc.zoom.us/j/2903979720 Sam's Office Hours: Kevin's 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
Lab Attendance	10%
Labs	15%
Homework	15%
Analyses	30%
Final Project	30%

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	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		

Assignment Descriptions

Lab Attendance:
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.

Labs:
As mentioned, each student is required to register for 1 weekly lab in conjunction with this course. During this period, students are required to complete a lab assignment that will be due at the end of the hour. 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.

Homework:
Homework will be based on problems from the course textbook, *R for Data Science*. Each homework assignment will be worth 20 points. These assignments are to be completed using RMarkdown and submitted as an HTML file on Sakai. Expect homework assignments to be due at 5PM on Fridays. If you submit your homework 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.

Analyses:
Analyses are constructed using customized problems from real life data sets. These analyses allow you to practice the techniques learned from lab assignments. Each analysis will be worth 40 points. These analyses are to be completed using RMarkdown and submitted as an HTML file on Sakai. Expect analyses

to be due at 5PM. 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.

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, is worth 10 points and will be due sometime in the middle of the semester after groups have been designated. The second part, the Exploratory Data Analysis, is worth 20 points and will be due approximately 1 month after the Project Proposal has been completed. The third part, the Final Paper, is worth 40 points and must be submitted on Sakai by 5:00PM on Monday, November 16. 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 TBD on Monday, TBD. Slides must be submitted by TBD on TBD on Sakai.

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. • Understand How COVID-19 Spreads • Wash Your Hands Often • Practice Physical Distancing • Wear a Face Mask • Maintain Clean, Safe Spaces • Engage in Smaller Group and Virtual Settings • Follow Immunization Recommendations • Protect the community • Provide Medical Return Clearance 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.
Community Standards	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 me – as we learn

	together. If you choose not to wear a mask, or wear it improperly, I will ask you to leave immediately, and I will 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.
Delivery (Lecture)	The lectures for this course are classified as a remote course. More detail of this delivery method is found at https://carolinatogether.unc.edu/ and provided below: • Students participate remotely for the entire semester and do not attend any in-person sessions in the classroom. • Remote classes will have a scheduled class time but no physical location. Students can participate from a residence hall, campus study space, off-campus residence, or from far away from campus. • The instructor will determine whether the class will be taught synchronously, asynchronously, or a combination of the two. Types of courses that may be taught with this mode include large classes where physical distancing could not occur.
Delivery (Lab)	The labs for this course are classified as face-to-face/hybrid. More detail of this delivery method is found at https://carolinatogether.unc.edu/ and provided below: • Students and the instructor attend required in-person ("face-to-face") sessions in the classroom at a scheduled time. • The instructor may also include "hybrid" components, such as rotating which students are in the classroom on a given day or holding full class meetings remotely. • The classroom environment will be configured to support compliance with University guidelines, and students and instructors are required to wear a face mask and practice physical distancing. Types of courses that may be taught as Face-to-Face/Hybrid include high impact courses, first-year seminars and experiential education courses.
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.