Syllabus for STOR 320-001
Introduction to Data Science
Fall 2019

Instructor: Dr. Mario Giacomazzo   **Office hours**: T   10:00AM - 11:00AM
E-mail: mgiacoma@email.unc.edu   Th 12:15PM - 1:15PM
Phone: 480-489-1398
Office: Hanes 136

Assistant: Dhruv Patel
E-mail: dhruvpat@live.unc.edu
Office Hours: TBD in Hanes B-54

Lectures: MWF from 12:20PM – 1:10PM in Hanes 120

Course URL: Website: https://supermariogiacomazzo.github.io/STOR320_WEBSITE/
Assignment Submission: https://sakai.unc.edu/ and login with your Onyen

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.

Textbook: *R for Data Science*, Hadley Wickham. Legally free online, but can be purchased for less than $40 on Amazon. Additional suggested texts are provided on the website. All texts used in this course are free and downloadable from course website.

Software: A personal laptop with a working copy of **R** and **R Studio** is required. You are required to bring your laptop to every class. Directions for free downloads of these materials will be provided.

Prerequisites: STOR 155 (Introductory Statistics) or equivalent. If you haven't had a programming course, then you really need this course or something similar and you should expect this course to be extremely time demanding for the first 6 weeks.

Final Grade: Attendance (10%)
Labs (20%)
Analyses (40%)
Final Project (30%)

Attendance: Attendance is mandatory. Once per week at random, I will take attendance at the beginning of class in the form of an interactive game worth 0-7 points. Only students in class will be eligible for the points. Any students who are absent will be penalized by receiving 0 points. At the end of the semester, the homework grade will be curved so students who had perfect attendance get 100% on their homework grade.

Labs: Labs are constructed using problems from the course textbook, *R for Data Science*. Each lab will be worth 20 points. These labs are to be completed using RMarkdown and submitted as an HTML file on Sakai. Expect labs to be due at **5PM**. If you submit your lab 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 at least 4 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 sometime towards the end 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 Sakai by **5:00PM** on **Wednesday, December 4**. 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 **12:00PM** on **Friday, December 13.** Slides must be submitted by **12:00PM** on **Friday, December 13** on Sakai.

Grade Scale: The below grading scale is guaranteed. The "+" and "-" designations will only be used in your favor. At the end of the semester, a comprehensive inspection of all the class grades will determine the cutoff points for "+" and "-" designations.

A [90,100]
B [80,90)
C [70,80)
D [60,70)
F [0,60)

More Help: Programming can be incredibly frustrating and take some time to get used to. Before you email your instructor or IA please spend some time trying to solve/Google the problem on your own. Then, proceed to find someone in the class to ask**.** It is highly possible someone else has experienced the same problem. Finally, if you have not solved the problem, email your instructor or IA.

Ethics: It is my belief that effective programming can involve the heavy use of the work of others. That said, in this course and in your professional work, you must give attribution when using the work of others. This is true if you are borrowing a part of a friend's code, and it is true if you are dragging code off a website. Excessive use of the work of others (my judgment) will impact your grade but cannot have honor code implications if you give attribution.

For avoiding excessive use of others' code, one good rule of thumb is that it is generally ok for another person to look at your code and make suggestions. If you are copying someone's code from this class (or who took STOR 320 last spring), this is excessive use.

You will learn best by suffering through your own programming with the help of others.

Note: This section is long because this can be a touchy topic in programming courses. My real concern is that you are transparent about what you do for graded assignments. I will also try to be transparent about what is expected.

Honor Code: All students are expected to follow the guidelines of the UNC honor code. In particular, students are expected to refrain from "lying, cheating, or stealing" in the academic context. If you are unsure about which actions violate that honor code, please see me or consult honor.unc.edu.

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.

The University of North Carolina at Chapel Hill has had a student-administered honor system and judicial system for over 100 years. The system is the responsibility of students and is regulated and governed by them, but faculty share the responsibility. If you have questions about your responsibility under the honor code, please bring them to your instructor or consult with the office of the Dean of Students or the Instrument of Student Judicial Governance. This document, adopted by the Chancellor, the Faculty Council, and the Student Congress, contains all policies and procedures pertaining to the student honor system. Your full participation and observance of the honor code is expected (honor.unc.edu).

Legal: Dr. Mario reserves the right to make changes to the syllabus, including all lab and analyses due dates, when unforeseen circumstances occur. These changes will be announced as early as possible so that students can adjust their schedules.