Syllabus for STOR 320.2:
Introduction to Data Science, Fall 2018

Instructor: Dr. Mario Giacomazzo
E-mail: mgiacoma@email.unc.edu
Phone: 480-489-1398
Office: 136 Hanes Hall

Office hours: Tuesday 3:30-4:30 p.m.
Thursday 2:30-3:30 p.m.

Assistant: Thomas Keefe
E-mail: tkeefe@live.unc.edu
Office Hours: Monday 5:00-6:00 p.m. and Wednesday 2:30-3:30 pm. in Hanes B-26

Lectures: Monday, Wednesday, Friday 11:15 a.m.-12:05 a.m., 120 Hanes Hall

Course URL: Visit 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 through a series of modules.

Plan to come to every class and come with your computer, ready to work with others. Using resources around you is a key component of successful data analysis. This includes people.

We will be closely following the textbook described below.

Textbook: **R for Data Science**, Hadley Wickham.

- Legally free online, but I have found it very handy to go ahead and buy the paper version (less than $40 at Amazon).

Secondary Reference: R Programming for Data Science, Roger Peng.

-you can get this for free or pay for it.

Software: This course will depend heavily on the use of the R programming system along with R-Studio. Directions for free downloads of these materials will be provided. along with other software that can be

Prerequisites: Please read this section carefully. For Fall 2017, the pre-requisite for the course is "STOR 155 (Introductory Statistics) or equivalent OR a programming course" . In future semesters, the "OR" will become "AND".

If you have had no programming course, This has two implications:

1. You really need this course or something similar.
2. You should expect the course to require lots of your time for the first 6 weeks or so. Learning programming for the first time can be frustrating, but you will be a better person for the experience. ☺

Final Grade: Labs (30%) – Weekly Labs.

Longer Assignments (40%) – Data Analyses.

Final Project (30%) – Details To Be Determined

Good Ethical Practice for Graded Assignments: 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 390 last spring), this is excessive use.

You will learn best by suffering through your own programming with the help of others. I will post a resources page for human help and for online help.

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.

Exams: Our final assessment will consist of a project submitted in print via RMarkdown. The final exam will be due before 5 PM on Tuesday, December 11.

Honor Code: The honor code will be observed in this class at all times.

For the complete honor code, please visit http://instrument.unc.edu/

Getting Help: Programming can be incredibly frustrating and take some time to get used to. Before you email the IA please spend some time trying to solve/Google the problem. Once you have exhausted your resources/patience ask someone (rule of thumb: spend at least 5 minutes, don't spend more than 20 minutes stuck on one problem).

Iain's FAQ/tips for success (Iain was the creator of this course.)

- Code a little bit every day. Even if you can only find 20 minutes - write a little bit of code.
- Google is your best friend for solving programming problems. I will repeat this many times.
- If one of the readings isn't satisfying to you go find a different explanation! It often takes 3 different explanations (read multiple times) before I understand an unfamiliar, tricky concept.
- If you are new to programming the beginning of the course will be a big adjustment. Don't give up, you'll get the hang of it in a few weeks.
- Coding in general can be very frustrating. Be patient and keep hacking at the problem.
- If you are already familiar with R there is still a lot to learn. For example, if you are not familiar with the tidyverse, the beginning of course will take some adjustment.
- This class may be quite different from other STEM courses you have taken - particularly the amount of class participation.

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