ECON 370: Economic Applications of Data Science

Alex Marsh

Fall 2022

E-mail: alex.marsh@unc.edu
Web: alexmarsh.io/teaching
Class Hours: TuTh 3:30 PM - 4:45 PM
Office Hours: TuTh 1:00 PM - 2:00 PM
Class Room: Murphy Hall 104
Office: Gardner Hall 416

Course Description

ECON 370 is intended to provide a broad-based introduction to numerical and data-science methods commonly used in economics. The course will first introduce students to the R programming language, assuming no prior experience. Subsequent lectures, using R, will provide students an opportunity to apply this knowledge on real-world data to achieve an economic objective. The methods used in these applications will include (but are not limited to): collecting, cleaning, merging, processing, and visualizing data, descriptive analysis, optimization, and supervised/unsupervised statistical learning. In addition, the course has an experiential component that connects students with industry leaders in economic applications of data-science through a series of on-campus events.

Course Goals

My teaching goals for this course are as follows:

1. Teach students how to competently program in R with good style,
2. Teach students how to think and approach problems from a computational perspective,
3. Teach students basic data science skills including data visualization and basic models,
4. Imbue students with a desire to learn more about econometrics and data science.

Learning Objectives

Upon successfully completing ECON 370, students should be able to do the following:

1. Be able to write functioning, readable, and aesthetically pleasing code in the R programming language,

2. Given raw data, be able to manipulate the data into the correct format needed for an analysis,
3. Given data and a research question, be able to create a exploratory data visualization in the right format to get at answering the question,
4. Be able to communicate results to a non-technical audience.

Prerequisites and Requirements

ECON 101 and a declared economics major.

Course Materials

- **Recommended Supplemental Textbooks:** I will be pulling readings from various free textbook online.
  - *Learning R* by Richard Cotton (RC)
  - *Hands-On Programming with R* by Garrett Grolemund (GG)
  - *R for Data Science* by Hadley Wickham and Garrett Grolemund (WG)
  - *Hands-On Programming with R* by Garrett Grolemun (G)
  - *An Introduction to Statistical Learning with Applications in R* by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (JWHT)
  - *Advanced R* by Hadley Wickham (HW)
- **R and RStudio:** We will be using R as our primary programming language of choice for this class. R is a great language to learn on and has an excellent Integrated Development Environment (IDE) called RStudio. To install R, go to **https://cran.r-project.org** and download the correct distribution for your machine. After installing R, you can install Rstudio by going to **https://www.rstudio.com/products/rstudio/download/** and downloading the free version of RStudio Desktop.
- **Recorded YouTube Recitation Sessions:** I wish this class could have a recitation section; there is a lot of material that needs to be covered that is not suited for a traditional lecture. However, this course is not big enough (nor should it be big enough) for the Graduate School to allow us a full TA. As such, I have recorded a bunch of short lectures of the type of material we would cover in a recitation section that you are responsible for watching. These are apart of the required readings, and you should watch them. These videos will get down in the weeds and cover details that I don't want to waste class time on. To find the videos, please follow **this link.**
- **Sakai:** All announcements, materials, assignments, grades, etc will be posted on the course's Sakai site. Please visit **sakai.unc.edu**. However, I will also be posting all materials on the **Github repository** for the course. You may access them however you wish, but assignments will need to be submitted on Sakai.

Course Content

The course will cover the following topics using illustrative economic applications. While this is the broad outline of the course, the exact details are subject to change based on interest and background.

Module 1: Introduction (week 1)

- Introduction to course, the syllabus, and R

Module 2: R programming (weeks 2-5)

- Data types and structures
- Objects and the environment
- Logic, loops, and control flow
- Functions and miscellaneous topics (dates, regular expressions, etc)

Module 3: Data Acumen (weeks 6-9)

- Structured and unstructured data
- Loading, cleaning, validation, merging, and processing data

Module 4: Data Science and Visualization (weeks 10-15)

- Descriptive analysis
- Visualizing and plotting data (spatial, etc)
- Clustering, classification, etc

Assessments

The following items will contribute to your overall grade:

- Problem Sets (60%): A total of five homework assignments counting equally. Assignments are due before class on the due date. These problem sets will likely be time intensive, so please plan for enough time to complete the assignment. I will make sure you have enough time between release date and submission date. There will be a 15 minute grace period in case there are any submission issues. The lowest assignment will be dropped. Late assignments will receive a score of zero.
- Participation (10%): Regular participation in class discussions and attendance at speaker events. I will be tracking participation at the start of each class. If you attend 75% of the lectures, you are guaranteed at least half (5% total) of the participation points. The rest are determined via a linear scale between 75% and 100% plus or minus how well I believe you are participating in class. This is a small class in which I can get to know all of you, and I want that to be the case. Asking questions, coming to office hours, and engaging with the material will serve you greatly both in this class, and for your time after this course.

- Final Project (30%): in groups of 5 or less, students will present their findings on an economic application of data science (of their choosing) during the scheduled final time. More details on this to come.

The UNC grading scale will be used. I reserve the right to curve grades if needed, but it will only ever be in the benefit of the student. The table below shows the grading scale, which corresponds to the traditional UNC undergraduate grading scale. To read the table, the percentage in each cell corresponds to the cut-off (or minimum percentage) for each letter grade minus/neutral/plus combination. These values are *inclusive*. So an A- corresponds to a percentage greater than or equal to 90% and *strictly less than* 93%. A B+ corresponds to a percentage greater than 87% and *strictly less than* 90%. And so on.

Cut-Offs For Letter Grades			
Letter Grade	(-)	( )	(+)
A	90%	93%	NA
B	80%	83%	87%
C	70%	73%	77%
D	NA	60%	65%
F	NA	0%	NA

Course Policies

Communication Channels

Feel free to contact me with any questions. I will try to respond as soon as possible. If I do not respond within twenty-four hours, feel free to send a follow-up email. Please write your email in a professional manner with a greeting, body, and closing statement. When addressing me, “Alex" is fine. I am not a professor (only a graduate student), and I am not yet a doctor.

Please make sure to notice the deduction in the communication policy time frame. ***I am not obligated to respond to emails regarding assignments that are sent within twenty-four hours of the deadline.*** There are multiple reasons behind this policy. The first is for your benefit as a student. Coding can take much longer than you initially thought and you must plan your time accordingly. Unless you are already a very proficient programmer, you are unlikely to be successful if you start working on an assignment the night before it is due. The second reason behind the policy is that I am also busy and a student. I try to have a life outside of my work as much as possible. It is unlikely that I will be checking my email if I am out with friends the night before one of your assignments is due.

During Class

It is strongly encouraged that you bring a computer to every lecture and be programming along with me and the slides. The only way you learn programming is by doing. I cannot stress this enough. I’ve heard some mathematicians say that math is not a spectator sport. While that is definitely true for math, it is even more true for programming. In order to learning how to program, you have to program! So please, follow along with me during lecture.

Submitting Assignments as a Group

You are allowed to work in groups of ***three*** for the homework assignments. If you do, ***please only submit one copy of the assignment and clearly state who was in your group.*** This 1) makes grading easier on our side, 2) guarantees consistency in grading across groups members, and 3) makes clear who worked with whom.

Names on Assignments

I will be giving you a unique identifier to put on your assignments. *Please do not put your real name anywhere in the file or in the file name.* This is to remove any implicit bias from names in the grading process. This is to be fair to you and also to protect my grader.

Regrade Policy

Regrading request must be sent via email within **one week** of the assignment being returned.

Attendance Policy

As participation is a part of your grade, attendance is *required*. Not attending lecture will hurt both your grade and your understanding of the course. If you have a University Approved

Absence, please let me know so that we can make plans accordingly

Academic Integrity and Honesty

You are required to follow the UNC Honor Code as stated. If you are unfamiliar with the honor code, please see me or visit: **https://catalog.unc.edu/policies-procedures/honor-code/**. Any violations of the honor code will be reported accordingly.

Accommodations for Disabilities

UNC accommodates reasonable requests 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**.

Counseling and Psychological Services

CAPS is committed to addressing the mental health needs of the UNC community. Please do not hesitate to reach out: **https://caps.unc.edu**

Discrimination and Title IX

I value the perspectives of individuals from all backgrounds reflecting the diversity of our students. I broadly define diversity to include race, gender identity, national origin, ethnicity, religion, social class, age, sexual orientation, political background, and physical and learning ability. I strive to make this classroom an inclusive space for all students. Please let me know if there is anything I can do to improve, I appreciate suggestions.

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.

Preferred Name & Preferred Gender Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records.

Additional Resources

- **The Learning Center:** The UNC Learning Center is a great resource both for students who are struggling in their courses and for those who want to be proactive and develop sound study practices to prevent falling behind. They offer individual consultations, peer tutoring, academic coaching, test prep programming, study skills workshops, and peer study groups. If you think you might benefit from their services, please visit them in SASB North or visit their website to set up an appointment: **http://learningcenter.unc.edu**.
- **EconAid Center:** Additional help can be obtained through the EconAid Center. More information can be found at **https://econ.unc.edu/undergraduate/econaid/**.

Syllabus Changes

The professor reserves the right to make changes to the syllabus, including project due dates and test dates. These changes will be announced as early as possible.

Important Dates

Please check the registrar's page for important dates: add/drop, breaks, course final, etc.

- Monday, August 15, 2022: First Day of Class
- Friday, August 19, 2022: Last Day for Late Registration
- Friday, August 26, 2022: Last Day to Drop Class (No Record)
- Monday, September 5, 2022: Labor Day - No Class
- Thursday, October 7, 2022: Last Day to Drop Class (On Record)
- Thursday, October 7, 2022: Last Day to Declare P/F
- Thursday, October 20, 2022 - Friday, October 21, 2021: Fall Break
- Wednesday, November 23, 2022 - Friday, November 25, 2021: Thanksgiving Break- No Class
- Wednesday, November 30, 2022: Last Day of Class
- Saturday, December 3, 2022: Final Exam Time