Introduction to Data Science for Economists

Spring 2026

Course Information

Course Number: ECON 4970
Meeting Times: M 9:30am-12:00pm
Location: Classroom South 200
Instructor: M. Cade Lawson
E-mail: mlawson18@gsu.edu (send all communication to e-mail, not through iCollege message!)
Office Hours: Mondays 12:15-1:15pm and by appointment
Office Location: 55 Park Place 728 (desk has my name on it)

Course Description

This course introduces students to the field of data science and its applications using the R data programming language, an open source platform that has become an industry standard because of its flexibility and power. It has a strong focus on using data to understand social issues in areas such as immigration, education, poverty, healthcare, and the environment. This course offers a practical, tools-based approach that is designed to build strong foundations for people that want to work as economists, policy analysts, industry data scientists, or data-driven journalists. The course will cover data programming fundamentals, data visualization, fundamentals of causal inference, and emerging topics such as machine learning and artificial intelligence (AI). The course is analytically rigorous, but no prior programming experience is assumed.

Learning Objectives

I want you to leave this class feeling a sense of agency over the social issues you care about, which is how I felt when I first learned data science. We will apply the tools we learn to a wide range of modern problems and discuss how data can be used to solve them. By the end of this course, you will be able to:

1. Demonstrate proficiency with the `R` programming language and `tidyverse` suite of commands.
2. Produce compelling data visualizations and conduct statistical analyses of large datasets.
3. Formulate a research question and locate the data needed to answer it.
4. Describe the importance of randomization for estimating causal effects and explain the methods economists use to disentangle correlation from causation.
5. Use large language models (LLMs) such as Google Gemini and GPT for the analysis of text data.
6. Effectively communicate technical results to a broad audience.

Textbook and Materials

We will primarily use two textbooks for this class. Both are required. **Don't rush to buy anything, I'll go over the cheapest options to access them on the first day of class**. They are:

1. *Quantitative Social Science: An Introduction in Tidyverse* (by Imai and Webb Williams) [abbreviated "QSS"]
2. *R for Data Science, 2nd Edition* (by Wickham, Cetinkaya-Rundel, and Grolemund) [abbreviated "RDS"]

R for Data Science (RDS) is a free, open-source textbook that you can access here. It is written by a few developers of the R programming language, and is widely considered one of the best resources for learning R.

Quantitative Social Science: An Introduction in Tidyverse (QSS) uses R to introduce how economists and other social scientists approach data science, with a focus on disentangling correlation and causation (causal inference). Make sure to get the version titled "An Introduction in Tidyverse" (this one)! Other versions are nearly identical but use slightly different code than what I'll teach in class, which will be harder to follow.

Some of what I'll teach you is so new that there is no textbook for it yet. For that stuff, I'll post separate notes or handouts on iCollege as necessary.

Piazza

We will use Piazza as a forum to answer questions in this class. You can access Piazza through the iCollege page for this course. Piazza allows you to post questions, which can be answered by myself or other students. **Please post all questions related to the course material (errors in your code, deadline questions, etc.) on Piazza instead of emailing me!** This allows the answer to be seen by other students who may have the same question. I also encourage you to answer your peers' questions on Piazza if you can. If I see that you are helpful and engaged on Piazza, I'll take that into account when deciding final grades.

Attendance Policy

This is a hands-on class where students are expected to follow along as we code together. Parts of the class will involve students working independently on data-related tasks with my assistance. Because of this, and since the class only meets once per week, attending every single session is very important.

I will take attendance every class, before and after our mid-class break. All students are allowed **2 unexcused absences** over the course of the semester without affecting their course grade. After that, each absence will result in a **5% deduction from the final grade**. Any student who misses more than 6 classes will receive an automatic grade of F for the class. Remember, we only meet 14 times total!

Only absences that are verified as "excused" by GSU will not count against your grade. Examples include military service, university-related travel, and emergencies (illness, family emergency, etc.). If you miss class for a documented reason, you must submit your excuse directly to the university using this link as soon as possible. Absences are verified by the Office of the Dean of Students, which then informs me. You do not need to send me personal information or details related to your absence! Just let me know that you will be absent and are coordinating with the Dean of Students to have it excused. Once I receive notice from them, I will correct your attendance record. Note that emergency absence requests must be submitted to GSU no later than 7 days after the emergency.

Grading

Your final grade will be determined as follows:

Assignment	Weight
Swirl Exercises	5%
Problem Sets (3 total)	35%
Quizzes (3 total)	15%
Mini-Project	15%
Final Project	30%

Swirl Exercises: Swirl is a learning resource for first-timers that is built into R. It walks you through some of the basics of coding in R step-by-step. In the first two weeks of the course, you'll go through the introductory Swirl exercises. This will be graded on completion and should take less than 2 hours. Further instructions will be given in class and on iCollege.

Problem Sets: There will be 3 problem sets (longer homework assignments) over the course of the semester, which should be completed independently. It is important to start these early so you can ask for help when needed. They will be graded based on the accuracy of your responses and the quality of your code, both of which will be submitted to iCollege.

Quizzes: There will be 3 in-class quizzes over the course of the semester, each of which will be 10 minutes long. They are designed to make sure you understand the course material and are not falling behind. Think of them as a check-in rather than a high-stakes exam; if you have been attending class and completing all assignments, you should get a high score with just a little bit of preparation the night before. If you find yourself failing them, check in with me to make a plan for future success!

Mini-Project: In the mini-project, I will provide you with a dataset and give you some rough guidelines for how to explore it. You will work in groups of 3 or 4 to analyze the data, produce visualizations, and research the causes of the trends you find. Each group will present in class on the assigned due date. It is designed to give you a sense of what I am looking for in your final project, which will be more open-ended.

This semester's mini-project will explore data on refugees, asylum-seekers, and internally-displaced people from the United Nations Refugee Agency. Each group will analyze global immigration and refugee resettlement patterns from a different decade. If your group wants to choose a dataset rather than using the one I assign, just send me an email.

Final Project: In the final project, you will find a dataset related to an issue you care about and conduct an exploratory analysis using the tools we learned in class. This is open-ended by design and should showcase what you've learned! You will come up with a research question and use data to find interesting trends, produce nice visualizations, and conduct statistical analyses that help us learn more about the issue. The data can be related to healthcare, poverty, sports, economics, the environment, or anything else you care about.

The final project, worth 30% of your grade, is broken down into the following components:

- 5%: a two-paragraph project proposal that tells me the topic you've chosen, the dataset you'll use, and what you plan to learn from it. I'll give feedback on the feasibility of the idea and some interesting things to explore.
- 25%: your final submission, which will be a PowerPoint presentation. These will be either presented in class or uploaded as a recording to iCollege, time permitting. You must also submit your code files.

This may sound vague now, but you'll see many examples as the class progresses. I want you to come away with a nice project you can show off as part of a portfolio, and to feel like you learned more about an issue you are passionate about. Final projects can be individual or done in pairs if you find a partner with similar interests. Pairs will be expected to conduct a more extensive analysis than students working alone.

Grading Scale:
Grade percentages will be assigned to a letter grade according to the following scale:

Grade	Percent	Grade	Percent	Grade	Percent	Grade	Percent
A+	> 97	B+	87 - 89	C+	77 - 79	D	60 - 69
A	94 - 97	B	84 - 86	C	74 - 76	F	< 60
A-	90 - 93	B-	80 - 83	C-	70 - 73		

Course Policies

Late Submissions: Late assignments will not be accepted. If you need additional time to submit an assignment, contact me no less than 24 hours before the assignment due date with your request.

Makeup Quizzes: Quiz makeups will only be allowed for university-verified absences. If you expect to miss a quiz, it is very important to reach out to me immediately and start the process of verifying your absence with the university using this link. No makeups will be allowed if an excuse is received after the quiz has already been administered. No quiz retakes will be allowed.

Regrade Requests: If you think I've made a mistake in grading, notify me no later than 2 weeks after you receive your grade. No regrades will be done after this period.

Instructor Response Times: Please allow up to 36 hours for me to respond to your email before sending another. I will almost always respond much faster than this.

Extra Credit: I may offer extra credit opportunities throughout the semester. Some or all may be available only to students who attend class on that day.

Additional Policies and Information

Syllabus Disclaimer: The course syllabus provides a general plan for the course; deviations may be necessary.

Basic Needs Statement: Students who face challenges securing their food or housing and believe this may affect their performance in a course are urged to contact the Dean of Students for support. Notify the professor if you are comfortable in doing so. This will enable us to provide resources that we may possess. The Embark Program at GSU provides resources for students facing homelessness.

Course Evaluations: Your constructive assessment of this course plays an indispensable role in shaping education at Georgia State. Upon completing the course, please take the time to fill out the online course evaluation.

Students with Accommodations: Students who wish to request accommodation for a disability may do so by registering with the Access and Accommodation Center. Students may only be accommodated upon issuance by the Access and Accommodation Center of a signed Accommodation Plan and are responsible for providing a copy of that plan to instructors of all classes in which accommodations are sought.

For more information, contact the GSU Access and Accommodations Center (AACE) Phone: (404) 413-1560, Email: access@gsu.edu, Website: access.gsu.edu/ Address: Student Center East, Suite 205, 55 Gilmer Street, Atlanta, GA 30303.

Remote Academic Coaching: The Access and Accommodations Center (AACE) also offers free remote academic coaching. To learn more about these services visit this link.

Academic Integrity: In the event of academic dishonesty, policies stated in the Student Code of Conduct and Policy on Academic Honesty will be followed. Cheating includes presenting another's work as your own, presenting previous work as work done in this semester, copying the work of another, the use of unauthorized materials during an exam, the use of artificial intelligence (AI) or other assistance to complete assignments, etc. If you have any questions about whether an activity is cheating, refer to the Deans of Students or see the instructor. The instructor with the guidance of university policies, will determine the sanction for dishonest behavior. Academic Dishonesty may result in an F for the course and may result in suspension or expulsion from the College.

Student Disruptive Behavior: Students are expected to behave properly in the shared student learning environment – so as not to interfere with the learning environment of others in the class. Students not adhering to these rules/guidelines may be asked to leave the class and may be subject to an administrative withdrawal (depending on the severity of the infraction). For an online course, this includes disruptive behavior in the course website(s). See the Student Code of Conduct Handbook for more information on GSU's policy on disruptive student behavior in the classroom or other learning environment.

Schedule

Note: The course schedule is subject to change. The most recent version of the syllabus will be on iCollege. Throughout the class, we will work through data exercises in class. Many will be chosen based on your interests. Some specific issues we will analyze in class are:

- The Flint, Michigan water crisis [Week 3]
- Achievement gaps in the National Assessment of Educational Progress (NAEP) and K-12 education [Week 4]
- Food insecurity in the U.S. [Week 6]
- Racial discrimination in the labor market [Week 11]
- Unemployment and the minimum wage [Week 12]
- The effect of smaller class sizes on student learning (Project STAR) [Week 13]

Week 1: Jan. 12

Topics	Introduction to R
Readings	• QSS Ch. 1: Introduction to R
Due	

Week 2: Jan. 19

Topics	**No Class – Martin Luther King, Jr. Day**
Readings	
Due	*Swirl Exercises Due Friday, 1/23 at 11:59pm*

Week 3: Jan. 26

Topics	**No Class – "Ice Storm"**
Readings	• Watch iCollege video on RDS Ch. 7 and Ch. 27.
Due	

Week 4: Feb. 2

Topics	Mechanics and Logic of R
Readings	• RDS Ch. 3: Data Transformation • RDS Ch. 4: Workflow and Code Style • RDS Ch. 12: Logical Vectors • RDS Ch. 13: Numbers
Due	

Week 5: Feb. 9

Topics	Data Cleaning and Merging
Readings	• RDS Ch. 14: Strings • RDS Ch. 18: Missing Values • RDS Ch. 19: Joins
Due	*Quiz 1 (In-Class)*

Week 6: Feb. 16

Topics	Data Visualization I: Univariate and Time Series Graphs
Readings	• RDS Ch. 1: Data Visualization • RDS Ch. 9: Layers • RDS Ch. 10: Exploratory Data Analysis
Due	*Problem Set 1 Due Friday, 2/20 at 11:59pm*

Week 7: Feb. 23

Topics	Data Visualization II: Maps
Readings	• QSS Ch. 5: Discovery (Section 5.3) • Elegant Graphics for Data Analysis Ch. 6: Maps
Due	*Final Project Proposal Due Friday, 2/27 at 11:59pm*

Week 8: Mar. 2

Topics	AI and Large Language Models for Data Science, Querying Data with APIs
Readings	
Due	*Quiz 2 (In-Class)* *Mini-Project Due Sunday, 3/8 at 11:59pm*

Week 9: Mar. 9

Topics	Further Topics in R, Mini-Project Presentations
Readings	
Due	

Week 10: Spring Break (No Class)

Week 11: Mar. 23

Topics	Randomization and Randomized Control Trials
Readings	• QSS Ch. 2: Causality (Sections 2.1 - 2.4)
Due	*Problem Set 2 Due Wednesday, 3/25 at 11:59pm*

Week 12: Mar. 30

Topics	Causal Inference with Observational Data
Readings	• QSS Ch. 2: Causality (Sections 2.5 - 2.6)
Due	

Week 13: Apr. 6

Topics	Survey Sampling and Statistical Bias
Readings	• QSS Chapter 3: Measurement (Sections 3.5 - 3.6)
Due	*Quiz 3 (In-Class)*

Week 14: Apr. 13

Topics	Machine Learning and Model Validation I
Readings	*An Introduction to Machine Learning in R* by L. Gatto: • Ch. 2: Introduction to Machine Learning • Ch. 4: Unsupervised Learning
Due	

Week 15: Apr. 20

Topics	Machine Learning and Model Validation II
Readings	*An Introduction to Machine Learning in R* by L. Gatto: • Ch. 5: Supervised Learning • Ch. 6 (Section 6.2): Model Performance
Due	*Problem Set 3 Due Friday 4/24, 11:59pm*

Week 16: Apr. 27

Topics	Final Project Presentations
Readings	
Due	*Final Projects Submitted by Start of Class 4/27*

Final Exam Period: Monday, May 4th from 8:00-10:30am. To be used for final project presentations.