Syllabus
Introduction to Data Analysis in R

University of Massachusetts Amherst
Spring 2025

COURSE DETAILS

- CICS 109: Introduction to Data Analysis in R
- 1 credit. No prerequisites. Pass/fail grading for undergraduates.
- Thursdays, 9:00-9:50 AM
- Computer Science Building, Room 142

Catalog Description

An introduction to data analysis in the open-source R language, with an emphasis on practical data work. Topics will include data wrangling, summary statistics, modeling, and visualization. Will also cover fundamental programming concepts including data types, functions, flow of control, and good programming practices. Intended for a broad range of students outside of computer science. Some familiarity with statistics is expected.

Learning Objectives

After completing this course, I expect you will have a solid foundation of skills to pursue practical data work in any discipline using R. You will learn the syntax of the R language and how to execute a variety of common data operations. More importantly, you will gain general knowledge of how to approach data-oriented problems using a programming language, including how to recognize common errors and how to discuss and present your work. (Statistics is outside of our purview though.)

Instructor

Jasper McChesney
Senior Data Analyst
Manning College of Information and Computer Sciences

Office: Computer Science Building 126 (generally available Tuesdays and Thursdays)
Email: jmcchesney@umass.edu (please include "CICS109" in all subject lines)

When I'm not on campus, I'm happy to have a call with you, probably via Zoom.

Required Materials

- **Canvas** will be used to disseminate course information and assignments.
- You will need to install **R** itself and the programming environment **RStudio** on your local machine. Both are available for MacOS, Windows, and Linux.
    - R: https://cran.case.edu/ (select OS and the most-recent installer from the top)
    - RStudio: https://posit.co/download/rstudio-desktop/#download
- There is no textbook

Accommodations

I intend this course to be a welcoming environment for all kinds of students to learn programming and data analysis regardless of previous experience, academic major, or personal characteristics.

The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (https://www.umass.edu/disability/)

FORMAT AND EXPECTATIONS

Lectures

Class time will primarily be devoted to "lecture" where I introduce topics and walk you through writing real R code. We will use a variety of artificial datasets designed to illustrate key topics. I may occasionally ask you to try briefly attempting a problem, to force you to grapple with the ideas on your own.

I do not take attendance, but quizzes may occur on certain days (see below).

Homework Sets

Between lectures there will be problem sets for you to attempt. Depending on your background, these could take 20 minutes to an hour to complete (hopefully somewhere in the middle).

These will not be collected or graded, and answer keys are provided. Their purpose is for you to practice doing analysis in R yourself (you can't learn simply by watching me do it in class). Students who take time to complete these usually succeed; those who don't will typically fail.

I recommend you attempt each problem, check your answer against the key, and then proceed to the next problem. Think carefully about why you did something differently from me, and if that's a problem (sometimes it's not: there are multiple ways to approach anything). Bring any questions about homework to class, or feel free to email me.

Quizzes

I will have some short surprise quizzes at the beginning of some classes. These will tend to focus on integrating new material with earlier material. They will be fairly general and will be more conceptual than code-oriented. Most questions will be open-ended.

If you have a reasonable excuse for missing a quiz, I will allow you to complete it remotely at a designated time or give you an alternate assignment, at my discretion.

Final Project

The majority of your grade will be a final project. I will furnish a dataset (possibly you will have some choice between a few), and you will undertake an analysis from start to finish, exploring the dataset, loading it, and analyzing it on your own. I will prompt you with a limited number of questions: about half will be narrow and specific, requiring mostly technical expertise, and the other half will be more open-ended, allowing you to tell me something about the data as you see fit. I will also ask you to produce graphs showing some of your key findings.

BASIS OF GRADING

For undergraduates, this course has a mandatory pass/fail grading basis, while graduate students will receive a normal letter grade unless auditing. A 60% is required to pass, with the assignments being worth the following:

- Quizzes: 25% (around 3-5% each)
- Final exam: 75%

You could certainly pass without attending any classes, but it would be dangerous to risk it.

Collaboration

The homework problems are not graded, and you are free to work with others, though I strongly suggest you attempt them on your own first.

The in-class quizzes must be done on your own, without any outside help or references.

The final project must be completed by yourself. You may refer to reference material, obviously including any material from the class. I do not recommend searching the internet for help, as it's likely to use different libraries and be confusing to a new R user—but you are allowed to, if you like. You may not, however, use generative AI (like chatGPT) to aid in completing the project, either to generate code or to examine the dataset. The purpose of the project is to see what you can do, not what AI can do.

General Academic Honesty Statement

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

SCHEDULE OF TOPICS

The days of class meetings are listed below but the topics covered may shift as needed.

#	*Date*	*Notes*	*Topics*
1	30-Jan		Introduction, expressions, vectors
2	6-Feb		Data frames, file I/O, multi-variable graphs
3	13-Feb		Graphing multiple variables
	20-Feb	*Monday Schedule*	
4	27-Feb		Logical comparisons
5	6-Mar		Ordering and sorting
6	13-Mar		Aggregations
		Spring Break	
7	27-Mar		Working with text
8	3-Apr		Dates and times
9	10-Apr		Linear models
10	17-Apr		Approaches to modeling
12	24-Apr		Cleaning and combining files
13	1-May		Data wrangling
14	8-May		More visualization