Syllabus
Introduction to Data Analysis in R

University of Massachusetts Amherst
Spring 2024

COURSE DETAILS

- CICS 109: Introduction to Data Analysis in R
- 1 credit. No prerequisites. Pass/fail grading for undergraduates.

Meeting Time

- Mondays, 2:30-3:20
- Hasbrouck Lab Addition, room 113

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.

Instructor

Jasper McChesney
Senior Data Analyst
Manning College of Information and Computer Sciences
jmcchesney@umass.edu (please include "CICS109R" in the subject line)

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
- We will have readings from this **textbook**:
    - *R for Everyone: Advanced Analytics and Graphics*
      Lander, Jared P.
      ISBN: 9780134546926
      Addison-Wesley

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 problems, and how to discuss and present your work. (Statistics is outside of our purview.)

Diversity and 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

The lectures will be focused on a real-world data set—about mountain climbing in the Himalayas. We will imagine typical problems a client might bring to us as data consultants; and by tackling these, illustrate the core skills of doing data analytics in R.

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.

Take the problem sets seriously! They're the main way you'll practice the material you see in the lectures. Only by wrestling with problems can you really internalize what you see—because it is too easy to nod your head in a lecture but find yourself unable to do anything later! Experience has shown that even computer science majors must do the problems to succeed.

I recommend you attempt each problem, check your answer, and then proceed to the next one. 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 any problem). Bring any questions about homework to class, or feel free to email me.

Quizzes

At the start of most class periods, I will ask you a few questions to answer inside of a Canvas quiz. These will not be coding problems per se, but more explanatory. They'll usually pertain to the homework from the night before but might also pertain to other material in the class. I will give only about 5 minutes to answer, so your comfort and familiarity with the homework are expected. They will be "closed book" in all ways.

Final Exam

You will demonstrate your progress by working in a more open-ended way with modest dataset during the final exam period. As with the lecture, you will imagine you're answering questions for a client. The emphasis here won't be technical sophistication, but rather applying a few techniques of your choice in an appropriate way, and being able to explain what the result means in context. Because of this, it will be "open book" in that you can refer to your text, R's help system, homework problems, your own notes, etc.; but not to any other internet resources.

- It's possible the final may be May 15 at 3:30, though this is not yet fixed.

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. The final will primarily demonstrate your abilities but is likely not enough on its own — given that a 60% is required to pass.

- Quizzes: 40%
- Final exam: 60%

Missing Class and Work

If you need to be absent for an excusable reason, I will either discount any quiz that was given that day, or arrange for you to take it remotely, at my discretion. Please get in touch with me as soon as you can. If you have an extended absence, we will find a solution if possible (in my experience, though, incompletes seldom work).

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 exam must obviously be done entirely on your own, but you may refer to approved reference material, namely: the textbook, homework assignments (including answer keys and your own work), R's internal help system, and your own notes. You may not search for help on the internet or use generative AI (like chatGPT).

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, and potential topics – these may change without advance notice.

#	*Date*	*Notes*	*Topics*
1	5-Feb		Introduction, expressions, vectors
2	12-Feb		Data frames, file I/O, multi-variable graphs
	19-Feb	*President's Day*	-
3	22-Feb	*Thu on Mon Schedule*	Graphing multiple variables
4	26-Feb		Logical comparisons
5	4-Mar		Ordering and sorting
6	11-Mar		Aggregations
		Spring Break	-
7	25-Mar		Working with text
8	1-Apr		Dates and times
9	8-Apr		Linear models
10	12-Apr	*Fri on Mon Schedule*	Approaches to modeling
	15-Apr	*Patriot's Day*	-
12	22-Apr		Cleaning and combining files
13	29-Apr		Data wrangling
14	6-May		More visualization

Final exam date TBD