STATS 306 200 FALL 2023

Introduction to Statistical Computing

This is an introductory statistical computing course based on the R programming language. Topics covered include data cleaning, visualization, basics of programming in R, and basics of statistical modeling.

Class Information

Time
MW 4p-5:30p.
Location
1324 EH

Important Websites

- Canvas: You should check the Canvas website for this course frequently. All course-related announcements, homework assignments and submissions, and grades will be posted on Canvas.
- Lecture recordings: Lectures will be recorded and posted here automatically.
- Piazza: Questions about the course, lectures, problem sets, etc. should be posted to Piazza. Any course-related question that does not require confidentiality should go here.
- R for Data Science: The textbook for this class. Note that we are using the second edition, not the first.
- Google Colab: Easy way to access the Jupyter notebooks during lectures.

Instructor Information

Name: Jayashree Ravi
Email: jayaravi@umich
Office: 258 West Hall
Office hours
Tu/Th 10a-11a (In person)
Tu 1p-2:30p (Zoom)

GitHub Course Repo

GSI Information

All GSI office hours are held in G219 Angell Hall.

Name: Zehua Wang
E-mail: wangzeh@umich.edu
Office Hours: Tuesday 2:30 - 4:00 PM
Lab Session 201 (Tuesday 4:00 - 5:30 PM @ B760 EH)

Name: Yijia Wang
E-mail: yijiaw@umich.edu
Office Hours: Tuesday 3:00 - 4:30 PM
Lab Session 202 (Tuesday 5:30 - 7:00 PM @ B760 EH)

Name: Junting Wang
E-mail: juntingw@umich.edu
Office Hours: Wednesday 9 - 10:30 AM
Lab Session 203 (Tuesday 8:30 - 10:00 AM @ 269 WEISER)

Name: Kevin Christian Wibisono
E-mail: kwib@umich.edu
Office Hours: Friday 11:30 AM - 1:00 PM
Lab Session 204 (Tuesday 10:00 - 11:30 AM @ 2080 SKB)

Waitlist / Enrollment Policy

Waitlisted students are expected to attend the lecture and submit assignments on time as though they were enrolled. Failure to do so may result in being removed from the waitlist. If you are on the waitlist and cannot access the course website in Canvas, e-mail the instructor so that you can be manually added. Waitlisted students are expected to turn in problem sets at the same time as everyone else, beginning with PS1.
I do not manage the wait list for this course and will not issue manual overrides. Enrollment for STATS 306 is handled by the statistics undergraduate advisers (statsugradprogram@umich). Please allow several days for them to respond to enrollment-related issues during busy times of the semester.

Grading

The final grade in the course will be determined by your scores in participation, homework, and exams using the weights given below

- Participation (15%).
    - 5%: Doing the assigned reading on Perusall.
    - 10%: iClicker participation on in-class quizzes. Join using this link: https://join.iclicker.com/ZDAW
    - I will award extra participation credit to students who consistently post high-quality questions or answers on Piazza.
- Homework (25%). There will be about ten weekly problem sets. Most weeks, they will be posted on Sunday and due the following Sunday. Late submissions are not accepted, but we will drop you lowest two scores when calculating your homework average. You may discuss the homework with other students, but you must write up your own solutions.
- Exams:
    - Midterm Exam (25%). The midterm will be a take-home midterm. You will have 24 hours to complete it.
    - Final Exam (35%). Administered in class on **Thursday, December 14th, 4pm-6pm.** You must be physically present to take the exam. It cannot be taken at any other time or location.

The grading scale is:

- > 90: A-/A/A+
- 80-90: B-/B/B+
- 70-80: C-/C/C+
- <70: D-E

If necessary, I will adjust these thresholds downwards (i.e. to give more A's), but will not adjust them upwards. The overall grade distribution will be similar to past semesters

Schedule of Lectures

A tentative schedule of lectures and homework is shown below. All readings are from R4DS unless otherwise indicated. I may alter this schedule as the semester progresses depending on the interests and abilities of the class.

Week	Date	Topic	Reading	Homework

1	Mon, 8/28/23	Introduction to R. Basic visualization.	Ch. 1-4	
	Fri, 8/30/23			
2	Mon, 9/4/23	No lecture(labor day)	Ch. 14-15	
	Wed, 9/6/23	Transforming data with dplyr.		
	Sun, 9/10/23			HW1 due
3	Mon, 9/11/23	Advanced dplyr. Pipelines.	Ch. 5	
	Wed, 9/13/23			
	Sun, 9/17/23			HW2 due
4	Mon, 9/18/23	Reshaping and tidying data. Missing data.	Ch. 6, Ch. 20	
	Wed, 9/20/23			
	Sun, 9/24/23			HW3 due
5	Mon, 9/25/23	Exploratory data analysis using ggplot.	Ch. 11-13	
	Wed, 9/27/23			
	Sun, 10/1/23			HW4 due
6	Mon, 10/2/23	Structured query language (SQL) and relational data.	Ch. 21, Ch. 23	
	Wed, 10/4/23			

	Sun, 10/8/23			HW5 due
7	Mon, 10/9/23	Importing data. Online data sources.	Ch. 22, Ch. 26	
	Wed, 10/11/23			
	Fri, 10/13/23			HW6 due
8	Mon, 10/16/23	No Class (Fall study break)		
	Wed, 10/18/23	Take home midterm		
	Thu, 10/19/23	Midterm is due at 12 pm.		
9	Mon, 10/23/23	Strings, text data, and regular expressions.	Ch. 16-17	
	Wed, 10/25/23			
	Sun, 10/29/23			HW7 due
10	Mon, 10/30/23	Factors, dates and times.	Ch. 18-19	
	Wed, 11/1/23			
	Sun, 11/5/23			HW8 due
11	Mon, 11/6/23	Lists, iteration, and functions/functional programming.	Ch. 27-28	
	Wed, 11/8/23			
	Sun, 11/12/23			HW9 due

12	Mon, 11/13/23	Statistical modeling in R. Linear regression.	LRR Ch. 1, 3, 4	
	Wed, 11/15/23			
	Sun, 11/19/23			HW10 due
13	Mon, 11/20/23	Prediction and classification	Getting started with Tidymodels , §1-2.	
	Wed, 11/22/23	No class (thanksgiving)		
14	Mon, 11/27/23	Case studies		
	Wed, 11/29/23			
15	Mon, 12/4/23	Case studies		
	Wed, 12/6/23	Final exam review		

Advice for success in STATS 306

- This is a programming course. The only way to learn how to program is by practicing. Reading the book and coming to lecture, while helpful, are not entirely sufficient. You should spend as much time as possible writing code to learn the ins and outs of R.
- Some students desire additional practice exercises beyond those offered in class and on the problem sets. Some good sources for these are:
    - https://www.datacamp.com
    - https://hackerrank.com
    - https://exercism.io/my/tracks/r
    - https://swirlstats.com
    - https://www.kaggle.com
- A surprising number of questions in office hours and on Piazza are of the form, "Will <x> work?" If you are not sure if something works, try it and see for yourself! Unlike your peers in chemistry, biology or physics, you have the luxury that failed experiments do not have the potential to result in catastrophe. If the code crashes or does not produce the

expected result, change something and try it again. Repeat until it does what you expect. Trial and error is an important part of the learning process.
- Get started on the problem sets early; they are time-consuming.
- You can miss one problem set without it affecting your grade. (See grading policy, above.) Use this wisely, if at all. The problem sets are the best source of practice questions for the exam.
- Most students have a high homework average at the end of the semester. Therefore, failing to turn in problem sets can significantly impact your final grade. Students who end up with two or more zeros on the problem sets tend to score a C or lower.

Academic Integrity

You are expected to abide by the LSA Community Standards of Academic Integrity
There is a zero-tolerance policy for cheating in this class. Any instance of cheating will be reported to the dean's office, automatically and without exception, and will result in a grade of zero for the assignment or exam in question. Examples of cheating include, but are not limited to:

- Copying homework solutions from another student.
- Turning in homework solutions from a previous semester.
- Using any form of chat or social media during an exam.
- Plagiarizing code or written answers from online sources.
- Using AI (e.g. ChatGPT) to generate homework solutions.

Exam Policies

Exams are taken in class and require a computer. It is essential that you have a working, fully charged laptop on exam day. If you are having problems getting R running on your computer, it is up to you to work with the GSIs and/or instructor beforehand to resolve them. There is no time to debug IT problems during the examination.

Test security

I take test security extremely seriously in this class. Use of any resource except those expressly authorized by the instructor is absolutely prohibited during examinations. In particular, use of any form of chat app or social media during an exam will result in automatic failure. Code plagiarism is easy to detect and will result in a substantial loss of points for all involved parties.

Regrade Policy

If, after an exam, you feel that a grading mistake has been made, you may request a regrade. In that case, the instructor will personally regrade all questions on the exam, which is not guaranteed to increase your overall score. Think carefully about this before requesting a regrade.