Modern Data Science with R

3-credit course (DSCI101)

Instructor

Gregory J. Matthews, Ph. D.
Loyola Hall 107
Chicago, IL
Voice: 83561
Email: gmatthews1@luc.edu
Office hour: By appointment (https://calendly.com/statsinthewild/office-hours-dsci101)

Class Time

Class Time: Tuesday and Thursday 2:30pm-3:45pm
Classroom: Cudahy Library - Room 318

Twitter

Course Twitter Account: @StatsClass
Course TikTok: @statsinthewild
Course Hashtag: T.B.D.

Course Description

This course provides students with an introduction to data science using the R programming languages covering such topics as data wrangling, data visualization, interacting with databases, principles of reproducible research, building simple statistical models/machine learning and data science ethics.

Course Outcomes

Students will obtain an extensive background in the basic tools used in the field.

Course Evaluation

It is a professional expectation that all students participate in course evaluations to guide ongoing program improvement.

Prerequisite

None.

Book

- Modern Data Science with R (2nd edition). Baumer, Kaplan, and Horton
  PDF of book is available here: https://beanumber.github.io/mdsr2e/index.html

Grade

- 8 Homework Assignments (25%)
- One Individual project (25%)
- Exam 1 (25%)
- Final Exam (25%)

Final Letter Grades

- $[93, \infty)$ = A
- $[90, 93)$ = A-
- $[87, 90)$ = B+
- $[83, 87)$ = B
- $[80, 83)$ = B-
- $[77, 80)$ = C+
- $[73, 77)$ = C
- $[70, 73)$ = C-
- $[67, 70)$ = D
- $[60, 67)$ = D+
- $(-\infty, 60)$ = F

Exams

All exams will be take home exams.

Homework

Homework is due approximately every other week. Discussion between classmates is encouraged; however, the final work should be independent. Homework must be submitted through Sakai.

Homework turned in after the due date will receive no credit. To help your final grade, please do avoid late homework.

Project

The individual project will require students to find a raw data set, wrangle the data into a useful format, perform some interesting analysis, and present results in a written report following the principles of reproducible research. All code must be version controlled through github (or repository of your choice) and a link to the repository must be submitted along with the final report. The project is due the same day as the scheduled final exam, which is Saturday, May 6, 2023.

- Make-up policy: Missed Exams: No make-up exams will be given. If you miss one exam and have a valid and verifiable personal or medical excuse (such as an illness requiring hospitalization or a death in the immediate family), your final exam score will be pro-rated to make up the missing points. All medical excuses must be signed by your physician. Original documents must be shown to the instructor; no photocopied or scanned documents will be accepted. Students who do not provide appropriate documentation for missing an exam will receive a zero on the exam. No more than one exam may be missed. Please note that vacations and travel plans are not valid reasons for missing an exam.

- Academic Dishonesty: Many practice problems will be given to be worked on outside of class; discussing practice problems with others is encouraged and in-class discussions are also encouraged. It is presumed and required that students do their own work on the exams. Cheating includes, but is not limited to, illegal collaboration, copying, using materials not permitted, and assisting others on tests. Anyone found cheating will not be permitted to withdraw and will receive an F grade for the course. Your academic dean will be informed and a statement will be placed in your permanent file.

- For more information on academic integrity, please see:
https://www.luc.edu/academics/catalog/undergrad/reg_academicintegrity.shtml

Week	Topic	Chapter	Due Dates
1	Introduction to R. Introduction to IDE's (e.g. RStudio). File types (.csv, .txt, etc) and command prompt. Version control and github. Principles of reproducible research	1	
2	Data Visualization	2	HW1
3	Data Summarization	3	HW2
4	Data wrangling for a single table (select, filter, mutate, arrange, summarize) Data wrangling for multiple tables (inner join, outer join, etc.)	4 and 5	HW3
5	Working with tidy data (i.e. rectangular data files) (data input/output and reshaping data)	6	HW3
6	Iteration (vectorized function, map family, etc.)	7	HW4
7	Data Ethics	8	HW5
	Spring Break		
8	Sampling distributions and the bootstrap	9	Exam 1
9	Statistical Modeling	9	Project
10	Introduction to Predictive Modeling	10	HW6
11	Supervised and Unsupervised Learning	11 and 12	
12	Basic simulation Interactive Data Visualization (e.g. Shiny, plotly, matplotlib, seaborn, etc)	13	HW7
13	Thanksgiving Break		
14	SQL and querying databases	15	HW8
15	***Final Exam Week***		Final Exam