STAT 535 - Statistical Computing
Fall 2023
3 credits

Instructor: Shai Gorsky

Office: Newton Mount Ida Campus of UMass Amherst, School of Design 125G

Email: sgorsky@umass.edu

Class Meetings:

Wednesdays, 6:00-8:30pm
Class meetings will be held in Room 105, the School of Design, **Mount Ida Campus**

Office Hours:

Wednesdays 4:00-5:00pm, Office 125G, School of Design, Mount Ida Campus

Course Dates:

Wednesdays between September 5 and December 8, 2023
(No class: November 22, 2023, Thanksgiving break)

Course Description:

This course provides an introduction to fundamental computer science concepts relevant to the statistical analysis of large-scale data sets. Students will collaborate in a team to design and implement analyses of real-world data sets, and communicate their results using mathematical, verbal and visual means. Students will learn how to analyze computational complexity and how to choose an appropriate data structure for an analysis procedure. Students will learn and use a computing language to implement and study data structure and statistical algorithms.

Course Objectives:

At the end of the course, you should be able to:

1. Design and implement an end-to-end statistical analysis of a data set,
2. Connect notions of computational complexity to statistical methods for data analysis,
3. Collaborate effectively in a team to develop a solution to real-world problems using statistics,
4. Evaluate and critique the choice data structure for an algorithm, and
5. Communicate statistical analysis results using mathematics, verbal and visual means.

This course aims to not only give you the opportunity to learn fundamental computer science concepts critical for understanding machine learning algorithms, but also brings you into contact with real data and allows you opportunities to make meaningful quantitative contributions to problems in genetics, social science, and other disciplines.

Textbook: Class lecture notes

Reference Books:

The Art of R programming: Tour of Statistical Software Design,
2011, by Matloff. ISBN-13: 978-1593273842.
This resource is available online via UMass library's website.

R for Data Science: Import, Tidy, Transform, Visualize, and Model Data,
2016, by Wickham. ISBN-13: 978-1491910399
Freely available at: https://r4ds.had.co.nz/

Prerequisites:

Prerequisites: STAT 516 and COMPSCI 121/INFO 190S/CICS 110
Note: Prior knowledge of statistical methods and programming experience (STAT 525 or equivalent) is strongly recommended for this course.

Expectations and Requirements:

1. Attendance is expected at the time class is held. If you cannot attend at the class time, please contact the instructor.
2. Weekly assignments (except when there are exams) are due every week before class. No late homework is accepted except for legitimate reasons.
3. Discussion of homework assignments with fellow students is encouraged. However, the code and the final report must be your own.
4. For the final project, you will be presenting a topic of interest, related to statistical computing, such as a new computing package, a simulation of an advanced statistical model, or some of your own research results demonstrated using a statistical software or method.

Assessment:

Course assessments consist of ten problem sets, one exam, and a final group project. The point distribution is

Assessment	**Percentage**
Problem sets	30%
Exam	25%
Project	40%
Class participation	5%

Grading thresholds are:

	A	A-	B+	B	B-	C+	C	C-	D+	D
≥	90	87	83	79	75	71	67	63	59	55

Exam:

The midterm exam will require you to translate and solve a statistical computing problem stated in words. You must state any assumptions that you make in your solution. You may be asked to write code or modify code without a computer. It is typical in an interview situation to sketch out your solution on paper or a whiteboard and the exam replicates this kind of assessment.

Project:

For the final group project, you will be presenting a topic of interest, related to statistical computing, such as a new computing package, a simulation of an advanced statistical model, or some of your own research results demonstrated using a statistical software or method.

Class Participation:

Attendance is required for all meetings. At selected points in the semester, participation will be assessed in terms of attendance and active participation. If you are not in class for a reason that is not accommodated, you will not receive class participation credit.

Makeup Policy:

If you will be absent due to extenuating circumstances – including jury duty, military oblications, scheduled activities for other classes, the death of a family member, or verifiable health-related incapacity – you remain responsible for meeting all class requirements and contacting me in a timely fashion about making up missed work. Timely means within one week of the event. We will meet to work out a make-up or alternative to make-up work.

Policies on Academic Honesty:

> 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/).

Accommodation Statement:

The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you have a disability and require accommodations, please register with Disability Services (161 Whitmore Administration building; phone number 413-545-0892) to have an accommodation letter sent to your faculty. Information on services and materials for registering are also available on their website www.umass.edu/disability

Course Schedule:

Week 1: Introduction to coding and the IDE (Integrated Development Environment)
Week 2: Data types and structures, introduction to computational complexity
Week 3: Flow control and functions
Week 4: Strings and regular expressions
Week 5: Transformations of data, SQL
Week 6: Visualizing data
Weeks 7-8: Functional coding, optimization and model selection
Week 9: Random number generation and sampling
Week 10-11: Monte Carlo methods, Bootstrapping
Week 12: Binding lower level code
Week 13: Introduction to parallel computing

Assessment Schedule:

Number	Date	Type and Description
1	Week 2	HW 1: Computing setup and introduction
2	Week 3	HW 2: Basic data structures, computational complexity
3	Week 4	HW 3: More data structures, and functions
4	Week 5	HW 4: Regular expressions
5	Week 6	HW 5: Transformations of data, and SQL
6	Week 7	Midterm exam
7	Week 8	HW 6: Optimization
8	Week 9	HW 7: Sampling
9	Week 10	HW 8: Monte Carlo
10	Week 12	HW 9: Monte Carlo, and bootstrap
11	Week 13	HW 10: Binding lower level code
12	Finals Week	Final project presentations