Syllabus
PSTAT 10 - Principles of Data Science with R
Summer Session A 2024 (6/24 – 8/3)

Lecture	MTWR 2:00 PM - 3:20 PM, Broida Hall, 1640	
Lab Sections	MW 8:00 AM - 8:50 AM at Phelps Hall, 1513	*with Cyrus Tabatabai-Yazdi*
	MW 9:00 AM - 9:50 AM at Phelps Hall, 1513	*with Cyrus Tabatabai-Yazdi*
	MW 12:00 PM - 12:50 PM at Phelps Hall, 1525	*with Daniel Silva*

Welcome

This course is designed for students intending to major in Statistics and Data Science, Actuarial Science, or Financial Mathematics and Statistics. We will explore many facets of the data science workflow through the R programming language. Through solving coding challenges, the student will be prepared to explore the R ecosystem on his own. These are the first steps required to apply for data science jobs.

Instructor

Ingmar Sturm
Email: `ingmar_sturm@ucsb.edu`
Please call me "Ingmar".

Office hours: 9 am to 11 am at Phelps 6213 or by appointment

Teaching Assistants

Cyrus Tabatabai-Yazdi
Email: `cyrustabatabai-yazdi@umail.ucsb.edu`

Office hours TBD

Daniel Silva
Email: `dcsilva@ucsb.edu`

Office hours TBD

Assessment

- 10 lab worksheets, 2 per week, total 25%
- 5 homework assignments, 1 per week, total 25%
- Midterm Exam 20%
- Final Exam 30%

The midterm exam will be an in-class exam on July 11, 2024 during the regular lecture time.

The final project will be given in week 5 and you have one week to complete it by Thursday, August 1, 2024 at 11:59 pm.

Attendance

Attendance at lectures and lab sections is mandatory. At the end of each lab section, you will hand in a lab worksheet to record your attendance.

AI Policy

Generative AI technology is quickly transforming Data Science and I encourage you to use it. However, I have strict rules for how you can use it.

You may use AI to ask questions, generate ideas, or to explain things that are hard to understand. In other words, I encourage you to use it as a learning tool.

However, under no circumstances may you use AI to create your deliverables (worksheets, quizzes, homework, exams). Doing so is a case of academic dishonesty and may lead to a failure of the class. Any code or words written should be your own and based on the material covered in class.

Tentative Course Agenda

—Subject to change—

Week 1: Introduction to R and Basic Operations

- **Monday, 6/24:** Lecture 1 - The R ecosystem, Vectors
- **Tuesday, 6/25:** Lecture 2 - Vector and Matrix Operations, Filtering, Recycling, Vectorization
- **Wednesday, 6/26:** Lecture 3 - Functions and control flow
- **Thursday, 6/27:** Lecture 4 - Problem Solving

Week 2: Data Handling and Visualization Basics

- **Monday, 7/1:** Lecture 5 - Factors, Data frames, Tibbles
- **Tuesday, 7/2:** Lecture 6 - Working with data frames (and tibbles)
- **Wednesday, 7/3:** Lecture 7 - Base R Plots
- **Thursday, 7/4:** *Holiday - Independence Day*

Week 3: Probability and Databases

- **Monday, 7/8:** Lecture 8 - Probability and Simulation
- **Tuesday, 7/9:** Lecture 9 - Random Variables and Expectation
- **Wednesday, 7/10:** Lecture 10 - Continuous Random Variables
- **Thursday, 7/11: Midterm Exam**

Week 4: Advanced Simulations and Databases

- **Monday, 7/15:** Lecture 11 - Applications of Simulation
- **Tuesday, 7/16:** Lecture 12 - Relational Databases
- **Wednesday, 7/17:** Lecture 13 - SQL Queries
- **Thursday, 7/18:** Lecture 14 - Aggregation and Joins

Week 5: Advanced Database and Visualization

- **Monday, 7/22:** Lecture 15 - More on Databases
- **Tuesday, 7/23:** Lecture 16 - Data Visualization
- **Wednesday, 7/24:** Lecture 17 - More on ggplot
- **Thursday, 7/25:** Lecture 18 - Tidy Data

Week 6: Final exam

- **Monday, 7/29:** *Study day*
- **Tuesday, 7/30:** *Study day*
- **Wednesday, 7/31:** *Study day*
- **Thursday, 8/1: Final exam**

Grading

Percentage Cutoffs for Grades

Grade	**Percentage Range**
A	Above 93%
A-	90% - 92%
B+	87% - 89%
B	83% - 86%
B-	80% - 82%
C+	77% - 79%
C	73% - 76%
C-	70% - 72%
D+	67% - 69%
D	63% - 66%
D-	60% - 62%
F	Below 60%