Syllabus

## DATA 200: Introduction to Data Science

### Course Information

**Term:** Fall 2023

**Instructor:** Isaac Quintanilla Salinas

**Email:** isaac.qs@csuci.edu

**Office Location:** BTE 2840

**Office Hours:**

| OH | Course | R Programming | TEA Math/Stat Hour |
| --- | --- | --- | --- |
| Day | MW | Tue | Th |
| Location | BT 1462 | LRC | BTE Courtyard |
| Time | 4:30-6:00 PM | 10-12 AM | 12-1:30 PM |

Or by Zoom appointment.

**Lecture:** Monday/Wednesday 10:00-11:50 AM (Sec 01) or 2:00-3:50 PM (Sec 02) in BT 1462

### Course Description

Foundations of the practice of data science. Exploration and implementation of the data science life cycle in locally relevant contexts. Emphasis on statistical thinking and computer programming required to collect, prepare and analyze data numerically, visually and algorithmically, and ethically communicate findings. Introduction to machine learning.

### Learning Outcomes

Statistical Thinking

1. Statistical Thinking
   1. Identify, calculate, and interpret multiple statistical measures (e.g. mean, standard deviation, percentiles).
   2. Construct visual representations of (tabular) data and assess which are most effective for displaying data and drawing conclusions given the context and purpose.
   3. Apply simulation to explore characteristics of different probability distributions (e.g., shape) and conjecture implications for real data sets.
   4. Utilize simulations to draw conclusions from data using inferential statistical methods (e.g., confidence intervals, hypothesis tests) and ethically reporting the results.
2. Computational Thinking
   1. Identify and implement fundamental data management operations (e.g., import, merge, sort) from a variety of sources to prepare (e.g., clean, wrangle) data for analysis and glean insights to answer realistic questions in a variety of contexts.
   2. Build a notebook environment to execute fundamental commands (e.g., import libraries, organize code, perform arithmetical operations, declare variables).
   3. Understand the logical steps of written code, including loops, conditional statements, functions, input, control flow, and output.
   4. Articulate a rationale for the ethical use of artificial intelligence (AI) to write and debug code and demonstrate an understanding of its limitations and possible unethical uses.
3. Machine Learning
   1. Differentiate among fundamental machine learning terms (e.g., features and labels, training and testing data, and accuracy and precision).
   2. Demonstrate an understanding of the difference between regression and classification in machine learning and be able to apply the most appropriate technique, aligning their choice with the nature of the data and the desired outcomes of the analysis.

### Textbook

This course will use Coursekata, a Canvas-embedded textbook with built-in R functionality.

### Recommended Software

For this course, we will use R and RStudio. You will have access to them online, but you may download and install them on your computer.

* **R** is a free statistical software program that is available for download at: <https://www.r-project.org/>.
* **RStudio** provides free and open source tools for your data analysis in R: <https://www.rstudio.com/>

### Embedded Peer Educator (EPE)

This course will have two embedded peer educators:

An EPE is there to help you understand the material and succeed in the course. They will be leading in class assignments. hold special tutoring sessions, and support you in your R programming. These educators are trained with study skills that will set you up with success in college.

### Course Grading

| Category | Percentage |
| --- | --- |
| Reading Assignments | 20% |
| Lab Assignments | 15% |
| Weekly Quizzes | 15% |
| Participation | 5% |
| Exam 1 | 15% |
| Exam 2 | 15% |
| Exam 3 | 15% |

At the end of the quarter, course grades will be assigned according to the following scale:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A+** | 98 – 100 | **B+** | 87 – <90 | **C+** | 77 – <80 | **D+** | 67 – <70 |  |  |
| **A** | 93 – <98 | **B** | 83 – <87 | **C** | 73 – <77 | **D** | 63 – <67 | **F** | < 60 |
| **A–** | 90 - <93 | **B-** | 80 – <83 | **C–** | 70 – <73 | **D–** | 60 – <63 |  |  |

#### Participation

Participation is based on short writing assignments conducted in class. There will be no make ups for these writing prompts.

#### Lab Assignments

Lab assignments are designed to expand your statistical knowledge. These will be completed in JupyterHub which can be accessed from Canvas. There are approximately 2 lab assignments every week that you can complete during class time. The lab assignment will be due either on Tuesday or Thursday at 11:59 PM following the lab day it was assigned.

#### Reading Assignments

Reading assignments are designed to teach you different statistical concepts and R Programming. As the course progresses, many of the concepts build on each other. Therefore, the assignments encourage you to read each chapter in an appropriate amount of time. You must read the chapter and answer the questions by the Sunday night at 11:59 PM. The 2 lowest reading assignments will be dropped.

#### Quizzes

Starting on Week 2, there will be a weekly quiz every Friday. The quiz will become available for you to take on Canvas at 12AM to 11:59 PM every Friday. The quiz is designed to test whether you recall the material from the previous week. You are expected to complete the quiz with out using any material such as your notes, textbook, or internet. Additionally, you will get 90% of the quiz grade for simply taking the quiz. The remaining 10% is based on getting the question correct. There are a total of 12 quizzes in the semester. The lowest 3 quizzes will be dropped from your class grade. There are no make up quizzes.

#### Exams

There will be three in exams. Exam #1 will on Sept 25, Exam #2 will be on Nov 8, and Exam #3 will be on Dec 6 at 8-10 AM (Sec 01) or 1-3 PM (Sec 02). While the exams are not considered cumulative, the material builds on each other. Developing a strong understanding of the material through out the course is important for your success. This course will operate under a zero-tolerance policy. Talking during the time of the exam, sharing materials, looking at another students’ exam, or not following directions given will be subject to the University’s academic integrity policy.

### Class Schedule

The following outline may be subject to change. Any changes will be announced in class.

| Week | Topic |
| --- | --- |
| 1 | Introduction to Programming/ Cause and Effect |
| 2 | Summarizing Statistics |
| 3 | Visualizing Data |
| 4 | Data Preparation |
| 5 | Sampling Methods |
| 6 | Exam #1/ Statistical Models |
| 7 | Simulations |
| 8 | Hypothesis Testing |
| 9 | Confidence Intervals |
| 10 | Linear Models |
| 11 | Linear Models |
| 12 | Exam #2 |
| 13 | Text Analysis |
| 14 | Image Analysis |
| 15 | Classifications |
| 16 | Exam #3 |

### University Policies

1. **Academic Honesty**:

* Please conduct yourself with honesty and integrity. Do not submit others’ work as your own. For assignments and quizzes that allow you to work with a group, only put your name on what the group submits if you genuinely contributed to the work. Work completely independently on exams, using only the materials that are indicated as allowed. Failure to observe academic honesty results in substantial penalties that can include failing the course.

1. **Disabilities:**

* If you are a student with a disability requesting reasonable accommodations in this course, you need to contact Disability Accommodations and Support Services (DASS) located on the second floor of Arroyo Hall, via email accommodations@csuci.edu or call 805-437-3331. All requests for reasonable accommodations require registration with DASS in advance of need: <https://www.csuci.edu/dass/students/apply-for-services.htm>. Faculty, students and DASS will work together regarding classroom accommodations. You are encouraged to discuss approved.

1. **Emergency Procedure Notice to Students:**

* CSUCI is following guidelines and public orders from the California Department of Public Health and Ventura County Public Health for the COVID-19 pandemic as it pertains to CSUCI students, employees and visitors on the campus. Students are expected to adhere to all health and safety requirements as noted on the University’s Spring 2023 Semester [website](https://www.csuci.edu/news/campus-updates/covid-19/index.htm) or they may be subject to removal from the classroom.

### Important note about a possible work stoppage during the semester

The California Faculty Association (the labor union of Lecturers, Professors, Coaches, Counselors, and Librarians across the 23 CSU campuses) is in a difficult contract dispute with California State University management. It is possible that we will call a strike or other work stoppage this term. I promise to promptly inform you of any schedule disruption. Our working conditions are your learning conditions; we seek to protect both. For further information go to www.CFAbargaining.org.