# PSTAT 100: Data Science Concepts and Analysis

## **Course Policies**

## Ethan Marzban

Welcome to PSTAT 100: Data Science Concepts and Analysis! I am very excited to join you as your instructor this quarter. Our journey together this quarter will take us through the basics of Data Science, and aims to prepare you for your future endeavors in the field (whether they be in classes, industry, or academia). Here's to a great quarter! — Ethan

# **Course Staff and Lecture Logistics**

Instructor: Ethan P. Marzban (he/him)

# Lecture Times and Location

T, R from 9:30am - 10:45am, in BUCHN 1940.

Teaching Assistants: TBD, TBD

# Sections:

- M 2 2:50pm in PHELPS 1518 (TBD)
- M 3 3:50pm in PHELPS 1518 (TBD)
- M 4 4:50pm in PHELPS 1517 (TBD)
- M 5 5:50pm in PHELPS 1513 (TBD)

## **Course Description**

As stated in the catalog description:

Overview of data science key concepts and the use of tools for data retrieval, analysis, visualization, and reproducible research. Topics include an introduction to inference and prediction, principles of measurement, missing data, and notions of causality, statistical "traps", and concepts in data ethics and privacy. Case studies will illustrate the importance of domain knowledge. Credit units: 4.

Indeed, this course is designed to be a hands-on introduction to Data Science for intermediatelevel students with some exposure to probability and basic computation skills, but with few or no upper-division courses in statistics.

## **Prerequisites:**

- PSTAT 120A (Probability at a calculus-based level)
- MATH 4A (a first pass at Linear Algebra)
- Prior experience with a programming language (e.g. Python through CMPSC 9 or CMPSC 16).

Note on Programming: the primary programming language of this course will be R, though you are not expected to necessarily have prior experience coding with R (so long as you have experience coding in another language, e.g. R, Julia, etc.)

# Textbooks/Readings

There are two "required" textbooks this quarter (required in the sense that readings will be assigned from them), both of which are freely available at the links below courtesy of the authors:

- R for Data Science (2e), by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund
- Learning Data Science, by Sam Lau, Joey Gonzalez, and Deb Nolan

If you are new to programming in R, you may also find this textbook useful:

• An Introduction to R, by Alex Douglas, Deon Roos, Francesca Mancini, Ana Couto & David Lusseau

There will also be a handful of articles which will be assigned as reading. All required reading can be found in the Weekly tabs of the Materials page.

# **Learning Outcomes**

By the end of this course, you should be able to:

- critically assess data quality and sampling design
- retrieve, inspect, and clean raw data
- understand the basics of exploratory, descriptive, visual, and inferential techniques
- interpret and communicate results in context

## **Assessments**

- Labs: short, structured coding assignments designed to introduce programming concepts and skills. Labs will be assigned weekly, and are designed to be mostly completed during Section on Monday but won't be due until 11:59pm on Wednesdays.
- **Homeworks:** slightly larger in scope than labs; will contain both programming and theoretical/conceptual concepts. We will have a total of 3 homework assignments released throughout the quarter; typically you will have 2 weeks to work on the homeworks and you are encouraged to start as early as possible!
- Mini Projects: more open-ended than homeworks and labs, and are designed to more closely simulate real-world data science projects and endeavors. We will have a total of 3 mini projects (released in non-exam weeks where there are no homeworks due) throughout the quarter, and you are encouraged to work collaboratively.
- Final Project: a final, comprehensive project that will be due during finals week. You will be required to work in teams on the Final Project.

## Final Project

The final project will be due by 11:59pm on Tuesday, June 11.

• 2 In-Class Assessments (ICAs) will be administered (see below for dates). I hesitate to call these "exams" because they are not intended to be as high-stakes as exams, however they are designed to test your retention of course material. More information will be released as we approach the date of the first ICA.

#### Assessment Dates

- In-Class Assessment 1: Thursday, April 25, 2024 in BUCHN 1940 (our Lecture classroom) starting at 9:30 (our regular class time)
- In-Class Assessment 2: Thursday, May 23, 2024 in BUCHN 1940 (our Lecture

classroom) starting at 9:30 (our regular class time)



## Warning

There will be no ICAs offered at alternate times, for any reasons. So, please plan accord-

#### **Final Course Grades**

Your final course grade will be computed according to the following weights:

Assessment	Weight
Labs	10%
$_{ m HW}$	15%
Mini Projects	20%
Final Project	25%
In-Class Assessments	30%

Your final letter grade will be issued according to the following scheme (cutoffs between plusses and minuses will be calculated at the end of the quarter):

- $A^- A^+$ : 90 100%
- $B^- B^+$ : 80 89.99%
- $C^- C^+$ : 70 79.99%
- $D^- D^+ : 60 69.99\%$
- F: 0 59.99%

I have elected to adopt an uncurved grading scheme to eliminate any sense of "competition" among students; I highly encourage you all to collaborate with and uplift each other. Having said that, I will certainly consider adjusting the cutoffs at the end of the quarter if necessary.

## **Policies**

#### Late Submissions

You are allowed **two** late submissions across homeworks and labs (so, 2 in total; not 2 each), that must be submitted within 48 hours of the original deadline. No work will be accepted beyond 48 hours after the original deadline.

#### **Email**

There are two primary means of communication outside of scheduled class meetings: office hours and a discussion board. If you are unsure of how to reach out to appropriate parties, please consult the following table:

Topic	Redirect to
Troubleshooting codes	Discord
Checking answers	Office hours or Discord
Clarifying assignment content	Office hours or Discord
Assignment submission	Gradescope
Re-evaluation request	Gradescope
Question about missing grades	Fill Out This Form

#### Some additional comments:

- Please note that we (the course staff) request you refrain from emailing us except in case of extreme emergency (it is up to you to decide what is an 'emergency'). Please bring all of your questions to the course staff during either Office Hours or after Lecture/Section. (Also, please note that DMs [Direct Messages] sent to the course staff through Discord will be ignored. Additionally, non-emergency emails [e.g. about content, missing grades, etc.] will be ignored.) Thank you!
- If you have questions or concerns about missing grades, please use the form linked in the table above. You are allowed to submit the form multiple times, but we ask that you please wait at least 48 business hours before submitting follow-ups. Thank you!

#### **Collaboration and Academic Integrity**

Data Science (as we will see) is an inherently collaborative field. As such, you are not only allowed but also *encouraged* to collaborate on assignments (be they lab, homework, or project). However, there are limitations to collaboration:

- collaboration on the ICAs is strictly prohibited
- do not copy other people's work and try to pass it off as your own
- if/when you work in groups, include the names of all group members on the assignment

Anyone found guilty of academic misconduct will be reported to the Academic Senate, and will receive at minimum a failing grade on the assignment in question; further actions may also include failing the course, and marks being made on permanent records. Depending on the severity of the infraction, expulsion is also a possibility.

Basically, don't cheat- please! If you're ever struggling with course material, please come talk to me or the TA's. We are truly here for you, and want only the best for you.

## **Section Switching**

Lab Sections take place in special "Collaborate Classrooms" which are equipped with laptops. There are only a fixed number of seats and laptops in these classrooms, meaning we cannot under any circumstance over-enroll sections. Therefore, if you want to switch section unofficially (we do not have the ability to switch your official enrollment through GOLD), please follow the steps at this link. Any requests to switch sections that do not adhere to the guidelines posted at that link will be ignored.

# Disabled Students Program (DSP)

If you have a disability, or otherwise require accommodations for the exams and/or quizzes please reach out to the Disabled Students Program (DSP) ASAP to ensure your request(s) for accommodation can be processed. We ask that all requests be logged at least a week in advance, to ensure the system enough time to process. Please note that we cannot grant any requests for accommodations unless they come to us from DSP directly.

# **Technology Needs**

As a part of this course, you will be required to program in Python. Though the Lab Sections take place in specially designed classrooms that come equipped with computers, your homework and quizzes may cover Python-related questions, which means we expect you to have access to a laptop capable of connecting to the internet. If you do not currently possess such a laptop, please check out UCSB's Basic Needs Resource page on Technology Resources to try and acquire one.