

# Stats 167 Lec 1: Syllabus

Introduction to Databases  
UCLA, Spring 2025

## Course Description

Introduction to concepts, tools, and technologies used in working with databases. Prepares students for applied data science work. Topics include the basics of data engineering; relational and non-relational database paradigms; an introduction to SQL and NoSQL languages; cloud storage platforms; connections to R and Python.

## Course Staff

**Instructor:** Michael Tsiang

**Pronouns:** He/Him

**Email:** [mike@stat.ucla.edu](mailto:mike@stat.ucla.edu)

**Office Hours:** Wednesdays 2:30–4pm and Fridays 3:30–5pm, or by appointment

**Physical Office:** Math Sciences 8105J

**Virtual (Zoom) Office:** Meeting ID: 579 439 842, Passcode: 982973

<https://ucla.zoom.us/j/579439842?pwd=TE1oTjJONFJTSzdKcUtJNHRjbjJ3UT09>

**Teaching Assistant:** Stella Huang

**Email:** [stellahyh@ucla.edu](mailto:stellahyh@ucla.edu)

**Office Hours:** Wednesdays 5–6:50pm

**Physical Office:** Math Sciences 5117 (5–5:50pm) and Boelter 5436 (6–6:50pm)

**Virtual (Zoom) Office:** TBA

## Learning Objectives

1. Demonstrate an understanding of basic data engineering concepts
  - (a) Identify whether data is structured, semi-structured, or unstructured
  - (b) Describe the differences between a database, data warehouse, data lake, and data lakehouse
  - (c) Describe issues related to data storage, consistency, and efficiency in big data systems
  - (d) Explain the components of an ETL data pipeline
2. Demonstrate mastery of basic programming in SQL
  - (a) Establish a connection to a hosted relational database (e.g., AWS, Azure, GCP)
  - (b) Read basic SQL code and describe its intended purpose
  - (c) Use SQL to query data in preparation for analysis and demonstrate the ability to filter, select, subset, merge, and summarize data
3. Demonstrate the ability to use R to perform basic queries on relational databases
4. Demonstrate the ability to use Python to perform basic queries on relational databases
5. Demonstrate an understanding of the difference between relational and non-relational databases

- (a) Debate the pros and cons of storing data in a relational or non-relational database
  - (b) Explain why ACID-compliance is important for database transactions
6. Demonstrate an understanding of the differences between non-relational database paradigms
- (a) Explain the structure of data that would be stored in each non-relational database paradigm (e.g., key-value, wide column, document, and graph)
  - (b) Explain how the CAP Theorem can affect database design
7. Demonstrate basic programming in some NoSQL languages (e.g., Redis, CQL, MongoDB)

## Course Materials

**Course Website:** <https://bruinlearn.ucla.edu/courses/210002>

**Discussion Forum:** <https://edstem.org/us/join/4QNWhY>

We will be using Ed Discussion (EdStem) for all class Q&A and discussions. Please use it to ask any class-related questions. You are encouraged to discuss and collaborate with each other and answer each other's questions. The instructor and TA will check EdStem to answer unresolved questions.

**Prerequisites:** Stats 20 and Stats 102A; Stats 21 or PIC 16A recommended

**Required Textbook:** None. Course notes and other resources will be provided on Bruin Learn.

**Required Programming Languages:** R and Python. No experience with SQL or NoSQL is assumed.

### Recommended Textbooks/Resources:

- Huo, K. and Singh, N., Ace the Data Science Interview, 2021
- Forta, B., Sams Teach Yourself SQL in 10 Minutes, 5th Edition, 2019
- Stephens, R., Sams Teach Yourself SQL in 24 hours, 7th Edition, 2021
- Teate, R. M. P., SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis, 2021
- Viescas, J., SQL Queries for Mere Mortals, A Hands-On Guide to Data Manipulation in SQL, 2018
- Kaufmann, M. and Meier, A., SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management, 2nd Edition, 2023
- Introduction to dbplyr Vignette
- SQLAlchemy 2.0 Unified Tutorial
- Sadalage, P. and Fowler, M., NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley, 2012
- DataStax Academy
- MongoDB University

**Whine and Cheese Club:** Every other Friday (Weeks 2, 4, 6, 8, and 10), 7:30–9pm Pacific

Physical Location: Math Sciences 8105

Virtual (Zoom) Location: Meeting ID: 967 5963 5259, Passcode: 908362

<https://ucla.zoom.us/j/96759635259?pwd=UHVhY042VWF0VzhvTStsMkxuZ2hWQT09>

The Whine and Cheese Club is intended to be an informal meeting place for community building and an open discussion about the topic du jour. Stop by to ask questions (related or unrelated to course content), vent about the week, get to know your classmates/instructor, or just come listen. Turning video on is encouraged but not required.

## Assignments and Evaluation

**Grading Policy:** Your grade will be based on the following components:

- 70% Homework
- 30% Final project (due Wednesday of Finals Week)

**If everyone does well, everyone will get a good grade.** Grades *start from* an absolute scale (90% is an A, 80% is a B, etc.), but the cutoffs can be flexible to accommodate distributions that do not fit that very well. Both absolute scale (e.g., raw percentage) and relative standing (e.g., *z*-scores and percentiles) are considered when determining final grades. Plus and minus cutoffs will be determined at the end of the quarter.

### Homework:

There will be weekly homework assignments posted on the course website. Please follow the guidelines in each assignment, and include your first and last name and student ID on every homework assignment.

### Final Project:

The final project will assess understanding of database design, schema, and/or structure to conduct a data analysis through a scenario at the level of a technical job interview for a data analyst position. A real-world database will be provided, and data manipulation and querying will be required to analyze the data. A final report document including all written analysis and code (in SQL, R, Python, NoSQL, or a combination of some/all) will be submitted. More specific details will be given when the project is assigned.

## Course Policies

### Copyright Policy:

All course materials (lecture notes, assignments, etc.) are intended for personal use only for students who are enrolled in Stats 167 Lec 1 in Spring 2025. **Do not post, share, or distribute any portion of any course materials to anyone or anywhere, either electronically or physically, without explicit written consent by the instructor, *even after the quarter is over*.** Failure to comply is a violation of academic integrity and copyright infringement against the author(s) of the course materials.

**Recording Policy:**

To respect and protect the right to privacy of the instructor, the TA, and each student, students are strictly prohibited from any and all photographic, audio, and/or video recording by any means (e.g., phone, camera, screenshot, etc.) in all class related sessions (lecture, discussion, or office hours), whether live, livestreamed, or pre-recorded. Any content recorded by the instructor is solely for use by currently enrolled students and may not be downloaded or shared without explicit permission.

**Email Policy:**

Any email to the instructor or TA must include your enrolled lecture number (e.g., “Stats 167”) in the subject heading (in addition to the subject of the email). Any emails without this information may be ignored without warning.

Please allow at least 24 hours (more on weekends and holidays) to expect a reply before sending a follow up email.

All homework assignments must be properly turned in through the respective submission portals on Bruin Learn. An attachment to a comment on a submission portal is not acceptable. Email submissions will not be accepted: **Do not attach assignments to email.**

**Assignment Policies:**

There will be a 24-hour grace period will be granted for any late submissions. No credit will be given for assignments submitted after the grace period. Exceptions to the late penalty are considered on a case-by-case basis to accommodate extenuating circumstances.

Submissions must be reasonably presentable, properly indented, well-formatted, and easily readable to the grader. Any submissions that are difficult for the grader to read will receive a penalty for style/readability.

**Collaboration Policy:**

This course is meant to be a collaborative learning opportunity to develop skills and build a resource pool for independent learning after the course ends. Thus, open collaboration and searching online for resources is allowed and encouraged. However, students are still expected to adhere to the university policies on academic integrity. In particular, citation of sources and attribution of work that is not wholly original is required. **Always give credit where credit is due.**

**Regrade Policy:**

Requests for regrades on homework should be sent to the TA, who will address the grading concern. Homework submissions will be regraded in their entirety, and any grade adjustments (whether an increase or decrease) will be considered final.

**Course Grade Changes:**

It is the student’s responsibility to check grades on Bruin Learn in a timely manner so that any grade issues (e.g., missing or incorrect grades) are resolved well before the quarter is over. All grades are considered final 24 hours after the final project deadline and cannot be appealed.

After course grades have been submitted to the Registrar, grades are final. Grade changes will only be considered if there has been a clerical or procedural mistake. Students have one quarter to make requests for a grade change. Graded exams and other materials will be kept for one quarter. After one quarter, course grade changes will not be made.

## University and Departmental Policies

**Academic Integrity:** As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. All students must uphold University of California Standards of Student Conduct as administered by the Office of the Dean of Students. Students are subject to disciplinary action for several types of misconduct, including but not limited to: cheating, multiple submissions, plagiarism, prohibited collaboration, facilitating academic dishonesty, or knowingly furnishing false information. For more information about academic integrity, please go to <http://www.deanofstudents.ucla.edu/>.

In addition, each student is the sole owner of their own code and work and must NOT:

- Submit work that is not original.
- Publish code or solutions online.
- Post the course questions on forums other than the designated course discussion forum. This means students cannot post questions on places like Stack Overflow, Chegg, ChatGPT, or other similar places.
- Submit someone else's work, or a modification of that work, with or without that person's knowledge.
- Allow someone else to submit their work, or a modification of their work.
- Contract course work out to others.
- Plan or execute with another student some form of cheating during an exam.
- Make use of unauthorized material during an exam.

**Zero Tolerance Policy:** Any and all issues of potential academic dishonesty will be reported to the Dean of Students without warning.

**Accessible Education:** If you have a disability that will require academic accommodation, please contact the UCLA Center for Accessible Education (CAE). Please contact the CAE as soon as possible to allow for sufficient time to coordinate accommodations.

**Title IX:** Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 1st Floor Wooden Center West, CAREadvocate@caps.ucla.edu, (310) 206-2465. In addition, Counseling and Psychological Services (CAPS) provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768. You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. Faculty and TAs are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator should they become aware that you or any other student has experienced sexual violence or sexual harassment.

**Undergraduate Mentoring:** I am an undergraduate mentor for the Department of Statistics and Data Science. This means that you may visit me during office hours (or by appointment) for an informal meeting where you can ask questions unrelated to course content, and about research opportunities, graduate studies, career paths, or any other topic pertinent to your education. You are welcome to visit and meet even after the course ends.

## Miscellaneous Advice

1. **Focus on the learning, not on the grade.** You are not a student at UCLA simply to get letters on a transcript. The point of this course is to introduce data technologies and concepts so that you are not floundering on the fundamentals in later applications when there are higher stakes and larger consequences. My main hope for you is eventually to become a strong, self-sufficient data scientist. I know that that takes more than one short quarter to happen. What is ultimately more important is to have a willingness to learn, a growth mindset, and integrity in everything you do.

No one is born knowing data science, statistics, or any other thing that you may feel “comes naturally” to some people and not others. It takes hard work. It takes practice (sometimes a LOT of it). It takes failing and trying again. It takes asking others for advice. Being comfortable with disappointment and knowing how to deal and respond to it is part of the growth process. It is hard, it often sucks, but it is ultimately necessary to become the best you can be.

2. **Ask for help.** Post to the discussion forum and attend office hours if you have questions or concerns. The teaching team holds many office hours spread out over the week to be available and accessible to you. Please let the instructor know if there are time conflicts with these.
3. **Attend lectures and communicate.** Attendance in lecture, while not strictly mandatory, is expected. Please communicate with the instructor if you are not able to attend regularly.
4. **No grade in any class, including this one, is more important than your physical well-being, your mental well-being, and your integrity.** Take time to rest, eat, exercise, go for a walk, connect with friends, speak to a counselor, or whatever it takes to take care of yourself. It might not feel like it, but you are not alone. Be mindful of others’ struggles as well.

<https://stand.ucla.edu/standforall>

“We need to remember what’s important in life: friends, waffles, work. Or waffles, friends, work. Doesn’t matter, but work is third.” – Leslie Knope (from Parks and Recreation)

## (Tentative) Topics Covered

1. Overview of Data Engineering
2. Relational (SQL) Databases
  - Basic SQL statements
  - Relational Databases in R with `dbplyr`
  - Relational Databases in Python with `sqlalchemy`
3. Introduction to NoSQL Databases
  - Key-Value databases (with Redis)
  - Wide column databases (with Cassandra)
  - Document databases (with MongoDB)
  - Graph databases (with Neo4j)

4. Further topics in databases (if time permits)

- Vector databases (with Pinecone)

## Tentative Course Schedule (Subject to Change)

Week 1	Course Logistics/Introduction
Week 2	Overview of Data Engineering
Week 3	Basic SQL
Week 4	Intermediate SQL
Week 5	SQL in R
Week 6	SQL in Python
Week 7	NoSQL Databases
Week 8	NoSQL Databases
Week 9	NoSQL Databases
Week 10	TBA

## Important Dates

- Week 7, Thursday (May 15): Self-Study Day, no lecture
- Finals Week, Wednesday (June 11): Final Project due