CDSS 170: Data Discovery Research Practicum

Spring 2026

Mon 3-4pm, Evans 10 | Mon 4-5pm, North Gate 105

Instructor
Emily Remirez, PhD
eremirez@berkeley.edu
Office: Warren 130

Program Director
George Obaido, PhD
gobaido@berkeley.edu
Office: Warren 130

Course description and goals

Students accepted to the Data Discovery program match with an industry or faculty mentor to work on a project that applies data skills to real-world problems. The project exposes you to the challenges of problem solving beyond the classroom. Complementing and scaffolding that work, the course builds skills for working with data, communication, teamwork, and project management across four interrelated themes:

1. **Contextualizing your work** within the larger framework of the project and integrating your experience with your studies and professional goals.
2. **Communicating** clearly and genre-appropriately about your data, methods, and interim results with different audiences.
3. **Working as part of a team** towards collaborative learning, project management, and problem-solving.
4. **Developing personal processes** around time management, resourcefulness, and independent work.

Time commitment

Data Discovery is a serious commitment. Unlike a typical course, there are real world stakes to the work you're doing. We're paying attention to your engagement by checking in during and outside of class, reading your lab notebook, and communicating with your mentor.

This is a 3-unit course, so you're committing to 9 hours of work a week: 6 hours for project tasks from your mentor, 1 hour for in-person meetings in class, and 2 hours for class assignments and

preparation outside of class. In weeks where there are less than 2 hours of work for CDSS 170, that time should be spent working on your project or developing skills relevant to your project or presentation: attending a workshop or research talk on campus, doing some background reading on a relevant topic, or getting ahead on project tasks.

Expectations for the project

The details of working together will be specific to your team and should be negotiated and revisited collaboratively. There are a few things we expect of everyone:

- **Weekly project hours:** You should spend 6 hours working directly on your project each week. This is an average across weeks, not a strict minimum per week. Your mentor shouldn't require you to work on non-instructional days or weekends, or to meet outside of typical work hours. If you need to reduce your hours for a given week, you're responsible for communicating with your mentor and teammates with as much advance notice as possible. You may be asked to make up the hours another week.
- **Proactive communication with mentor and team:** Keep your group updated as issues arise (e.g., if you aren't able to complete a task as expected or have a schedule conflict).
- **Academic and intellectual honesty**: Any work you submit to your mentor should reflect your genuine contributions, and you should never take credit for work that isn't yours. Do not use generative AI for any part of your project without clearing it explicitly with your mentor.

Course staff, office hours, and getting in touch

- **"No wrong door."** If you contact us with a question, we will do our best to help you get connected! When in doubt, email **cdss-datadiscovery@berkeley.edu**. Note that this is a shared inbox.
- Email and bCourses are good for quick questions, but for complex technical questions or sensitive interpersonal manners, please schedule office hours
- Please allow at least 24 hours (excluding weekends) for a response. If you still haven't heard from us and need a response, feel free to send a reminder.
- **Before asking your question, check your resources!** Have you searched bCourses and the slides? Have you asked a teammate?

Office hours

Course staff are available to meet with you about your project and the course outside of class. Office hours are on Zoom by default and can be scheduled in 15-minute increments. You can come to office hours alone or with your team to discuss any aspect of the course, get technical help for your project,

or talk about your individual goals and needs as a student. If none of the appointments available work or if you prefer to meet in person, please get in touch over email or bCourses. Office hours are subject to change for individual weeks–please refer to the calendar for the most up-to-date offerings.

Staff	**OH Schedule**	**Set an appointment**
Emily	Thursdays 2:30-4, Zoom	Emily's office hours
George	Tuesdays 2-4, Zoom	George's office hours
Claire	Tuesdays 10-11, Zoom	Claire's office hours
Rohan	Mondays 11 - 12, Google Meet	Rohan's office hours
Table 1: Staff office hours		

Staff (bios coming soon)

Dr. Emily Remirez (Emily, she/they)
CDSS 170 Instructor and Data Discovery Co-Director | eremirez@berkeley.edu
Ask me about: Assignments and grading; social science and humanities; qualitative data; human subjects research; lit reviews; writing and communication; interpersonal concerns

Dr. George Obaido (George, he/him)
Data Discovery Director | gobaido@berkeley.edu
Ask me about: Python, Machine Learning, SQL, Data Science, LLM, Choice of ML Algorithms / Tools, Fairness AI/Bias, Responsible AI, etc.

Claire Chen (Claire, they/she)
Tutor | claire.chen@berkeley.edu
Ask me about: Data Science, SQL, Python, Pandas, NumPy, data visualization, LLMs; effective writing & communication; interpersonal dynamics, program logistics

Rohan Bhagat (Rohan, he/him)
Tutor | rohanb@berkeley.edu
Ask me about: Python, Pandas, effective data visualizations, Plotly, Scikit-learn, LLMs, SQL, ML, matplotlib, AWS, NumPy, Seaborn, writing, communication, interpersonal concerns

Computing resources

- Computing resources are available for you to complete your work on the project
    - You have access to CDSS DataHub for large-scale data processing, AI, and machine learning tasks
    - You can request access to Savio, Berkeley's supercomputing infrastructure.
- Both platforms include pre-installed Python and R data science libraries, with options for additional software
- More details can be found at these links:
    - DataHub: https://berkeley-cdss.github.io/discovery-handbook/computing/datahub/
    - Savio: https://berkeley-cdss.github.io/discovery-handbook/computing/savio/

Please reach out to George if you have questions or concerns about accessing these resources.

Academic honesty and generative AI (GenAI)

- As with all your classes, you must abide by the UC Berkeley Code of Student Conduct and Honor Code, which prohibit cheating, plagiarism, and presenting false information.
    - Do not misrepresent any work and ideas from other sources as your own. Doing so constitutes plagiarism. This includes your Data Discovery teammates, your GSI, Wikipedia, a journal article, a podcast, and GenAI.
    - Do not present false information. You are solely responsible for any inaccuracies, including those copied from GenAI.
- We expect that most of your assignments will be a collaborative effort among your team. Someone may contribute more to an assignment one week and less another, but the division should balance out over the semester. Everyone should be credited for their contributions.
- If you reference outside sources, acknowledge the source of the information with a citation following any major style guide. Include a link to any source accessed online or wherever possible.
- If you choose to use generative AI (GenAI; e.g., large language models, including chatbots like Chat-GPT, Gemini, Claude, etc), **you must:**
    - Include a description of the tool you used and how you used it.
    - Ensure that all ideas come from you and your prompt, with the AI serving as a tool for organizing your ideas and refining your writing.
    - Refer to campus guidance on using genAI in your work.
- Do not use GenAI in your project work without explicit permission from your project mentor.
- Written work should include a contribution statement including how each team member contributed to this specific assignment; whether any outside sources (including people!) were

consulted, with citations where appropriate; whether and how any generative AI tools were used, including the name of the tool and what you included in the prompt.

Assignments and grading

Expectation	**Description**	**% of grade**
Participation and attendance	Come to class and engage in relevant discussion with your team. Share out to the larger group at least twice	40
Lab notebook assignments	Complete 5 short written assignments as a team to document your progress and prepare for your final presentation	30
Poster or final presentation	Present a poster at the Symposium (or comparable final presentation) reporting on your work this semester	15
Mentor evaluation	Maintain commitments to your project mentor	10
Program assessments and surveys	Respond to surveys to help us check in on your team and assess our program	5
Table 2: Grading breakdown and point distribution		

General assignment and grading policies

- **Submit on Gradescope:** Written assignments will be submitted via Gradescope. The course number is **WNDG4P**. Most assignments will be group submissions you complete with your team. Refer to Gradescope's guide on adding group members here.
- **Communication is key!** If you have questions or ideas about your assignment, please check in at least 48 hours before the deadline (or as soon as possible). We're happy to work with you, but we need to be on the same page, and you need to be proactive.
- **Late work:** All work will be accepted within 24 hours of the deadline without a grading penalty. After 24 hours, 50% will be deducted. After 96 hours, no late work will be accepted. (That is, if work is due on Wednesday, you can submit on Thursday "for free" or on Friday, Saturday, or Sunday for a penalty.)

- **Extensions:** Small extensions (≤1 week) for written assignments can be requested over email before the assignment due date. Please include the names of everyone requesting the extension and the date you expect to submit the work.
    - Think carefully about how much time you actually need when making your request. The due date you set will become the new due date for the assignment, and late work policies will apply from that date.
    - This policy also applies to extensions related to DSP accommodations (unless specified in your letter).
- **Revise and resubmit:** For some assignments, you may be able to revise and resubmit your work for a higher grade. Please reach out to Emily if you're interested in revising your work.
- **Written assignments should include a contribution statement**, including:
    - Who on your team contributed what to this particular assignment?
    - Were any outside sources used? This includes any people you talked to, so you'll often be listing your mentor. For sources someone else could access, include a citation in any major style guide.
    - Was generative AI used? If so, how? Include the name of the tool, what you provided in the prompt, and what you wanted to accomplish.

Participation and attendance

- Participating in class means showing up every week ready to work with your team on project tasks and discuss the day's topic. It also means paying respectful attention and not being disruptive during lectures or presentations. This is a laboratory course: The work happens here and now, so attendance is crucial.
- Course staff will be taking note of attendance and participation during group discussions by circulating and checking in with each team. If no one is able to check in with you during a given week, we may follow up asynchronously or ask to schedule an office hours meeting.
    - For full credit for attendance, be marked present for at least 75% of meetings (about 3 free absences). After that, 4% of your grade will be deducted for each unexcused absence.
- During the last part of class meetings, teams will share out with the larger group to hear what other teams are doing and practice talking about their work.
    - Each person should share out to the larger group at least twice during the semester, including a brief, informal presentation along with your team during weeks 8-11. 3% of your grade will be deducted for each missing report. (Multiple team members can share at the same time, but sharing is tracked on an individual basis.)
- Plan to bring a device (e.g., your laptop) to class every week to work on your project and CDSS 170 assignments.
    - Please keep devices out of sight when not directly relevant and remain engaged

- **Do not come to class if you are ill**. You need rest to recover, and you protect the entire community by self-isolating when you might be contagious. You're encouraged to wear a mask to class, especially if you may have been exposed to Covid or another illness.

Team meetings

- The middle part of each class will be dedicated to team meetings. During this time, you should actively engage with your teammates to make progress on your project.
    - This could include working on project assignments from your mentor, individually or collaboratively; seeking advice from course staff or other classmates; or working on upcoming CDSS 170 assignments.
    - If you need further technical support, schedule an office hours appointment.
- In addition to these in-class meetings, you should be meeting as a team with your project mentor at least once a week. You may also want to schedule a student-only meeting outside of class, depending on your project needs.

Lab notebook assignments

- Your team will keep a lab notebook to document your progress.
- Lab notebook assignments are designed to help you work together, contextualize your work with your mentor, and scaffold your poster presentation
- Assignments can be completed with any type of Data Discovery project, and specifics can be tailored for your project. If you have any questions about how to complete an assignment or ideas for modifications that would help you, please check in at least 48 hours before the assignment deadline.
- There are 5 assignments planned for the semester (More details to follow; subject to change)
    1. The Mentoring Agreement ensures your group is on the same page and set up for success this semester
    2. The Project Description documents the broader context for your work and what your goals are for the semester
    3. The Mid-Term Memo frames your work so we can see your progress
    4. A first assignment targeted directly to the poster
    5. A revision of that first plan
- In addition to submitting the written assignment to Gradescope, you may be asked to talk about your submission in class. You'll get credit for sharing as part of your attendance grade.

Final presentations and Data Discovery Symposium

- Everyone will deliver a presentation about their work this semester, with both visual and oral components. Most teams will fulfill this requirement by presenting a poster during the Symposium. This is **15% of your grade**.
- The presentation should be the original work of your current student team. That is, you shouldn't use a poster your mentor or past student teams created. The focus should be the specific work of your team, not your mentor's larger project.
- Your mentor is welcome and encouraged to provide support, and you should give them a chance to review your presentation for accuracy.
- We consider all projects to be "works in progress," and your project does not need to reach a particular point of "doneness" to be presented. For example, you could describe a problem you encountered and how you tried to address it, or you could show off something you built.
- There are 3 interrelated objectives that every final presentation, regardless of format or audience, should achieve:
    1. **Clarity of message.** You've thoughtfully chosen a topic related to your Data Discovery project and appropriate to present to an audience not directly familiar with your work. You've followed best practices to communicate clearly.
    2. **Structure and organization.** Your presentation has an internal structure with a beginning, middle, and end. It's easy to follow, well-balanced, and organized to enhance understanding of your message.
    3. **Visual communication.** Visuals (graphs, charts, figures) should feature prominently in your presentation. They're thoughtfully chosen and designed to work together with written and oral communication.

More on the Data Discovery Symposium

Save the Date for the Spring 2026 Data Discovery Symposium!
Thursday, May 7, 1:30 - 3:00 pm, Hearst Memorial Mining Building Lobby

- The Symposium is our chance to connect as a community and celebrate all the work you've done this semester. In addition to poster presentations, we'll have snacks, awards, and a faculty keynote.
- **All Data Discovery participants are welcome and strongly encouraged to present a poster at the Data Discovery Symposium during RRR week!** This is both the most straightforward and most fun way to fulfill the final presentation requirement.
- Your poster should be represented by at least one person from your team from set-up (beginning at 12:30pm) until break-down (ending around 3:30).

- Mentors have been asked to cover the cost of printing your poster (about $30-50). If your mentor is unable to cover the cost, please get in touch. There are funds we can use to print your poster if we make arrangements in advance. You may be eligible for campus programs like low cost or free printing from the Jacobs Institute.
- You can use any print service you prefer, but we'll arrange an optional group order for convenience and a small discount. (Posters must be ordered and paid for individually through the printer's order form.) Posters should be 24x36 inches, either landscape or portrait.
- The event is open to the public. Please invite your mentor to join us! We also encourage you to invite friends, family, instructors and other mentors, and anyone else who might like to see your presentation.

Other final presentations

Two kinds of alternate final presentations will be accepted with prior approval. Details about how to seek approval will be provided later in the semester.

1. If you're unable to present a poster publicly during the symposium (e.g., confidentiality concerns, major conflicts, or a poster being an unsuitable format), you can record a video or arrange a synchronous presentation with course staff. Your presentation must include *both* audio narration and visuals (e.g., a slideshow or digital poster). No standalone written reports will be accepted. Your request should include your plan for the presentation and your rationale.
2. If you're giving a comparable presentation elsewhere (e.g., at a conference or to your mentor's organization), you can petition to use this for your final presentation requirement. Your request should include the date, audience, and format, and an overview of the content. We'll also need direct confirmation, in writing, from your mentor or another sponsor.

Feedback from your mentor

- Your project mentor will be asked to share feedback on your engagement with us
    - You may be penalized if your mentor or teammates report serious and/or persistent concerns about your commitment or engagement. Especially serious concerns, such as "ghosting" your mentor, can result in a failing grade in CDSS 170 and being barred from the program in the future.
    - You will not be penalized for things like failure to achieve a desired result in your analysis, needing extra time to learn something new, missing a week of work due to an emergency, or deviating from the initial timeline as the project develops.
- Your mentor is also being asked to provide constructive feedback on your work directly to you around the middle of the semester

Schedule

Specific topics subject to change

Week	Date	Topic	Due this week
1	**Jan 26**	Welcome and orientation	
2	**Feb 2**	Working as a team	Mentoring Agreement
3	**Feb 9**	Understanding the broader context	
4	**Feb 16**	*Holiday - no class*	
5	**Feb 23**	George guest lecture	Project Description
6	**March 2**		
7	**March 9**	Communicating with visualizations	
8	**March 16**	Mid term memo presentations	Mid-term Memo
9	**March 23**	*Spring break*	
10	**March 30**	Poster presentations/mid term memo presentations	
11	**April 6**	Storytelling and finding a narrative within your work	
12	**April 13**	Visual communication; posters as an (audio)visual medium	Lab notebook assignment 4: Finding the story for your poster
13	**April 20**	Poster work time; clarity, concision, context	
14	**April 27**	In class poster review	Lab notebook assignment 5: Draft poster + critique
15	**May 4**	*RRR Week*	Symposium poster presentations
16	**May 11**	*Finals week*	Final final deadline for all outstanding work May 13

Table 3: Week-by-week schedule

Accommodations and resources

Our well-being is a prerequisite to our learning. If you're struggling this semester academically or personally, we can work together to mitigate the effects. As instructors, there are some forms of violence and abuse that we are legally required to report. Whenever we are not required by contract to report information, we will always respect your privacy.

Disability accommodations

Diverse perspectives improve the learning environment for everyone, and all students have a right to participate equitably in class. Students with disabilities are legally entitled to reasonable accommodations to make this course and program accessible. The Disabled Students Program (DSP; https://dsp.berkeley.edu/home) is the campus office responsible for authorizing disability-related academic accommodations.

An *accommodation* mitigates access barriers without interfering with the core functions of learning or assessment. Accommodations are a collaborative effort that involves students, instructors, and disability specialists. Examples of accommodations include, but are not limited to, extra time or extensions for assignments; a quiet or separate environment for testing; excused absences; adaptive furniture; captioning, transcription, or note-taking services; alternate formats for course materials; and accompaniment by a service animal.

If you anticipate needing disability accommodations this semester, please work with your DSP specialist to send your letter to us as soon as possible. In addition to the letter, you are responsible for communicating with us about using your accommodations. For example, if you have an accommodation for extensions, you must request the extension and set a new due date in advance.

Confidential resources

- For SVSH and intimate partner violence concerns, contact PATH to Care: care.berkeley.edu; for urgent support, call 510-643-2005
- For mental health concerns, The Tang Center (https://uhs.berkeley.edu/caps) provides both individual and group counseling for students, and has drop-in hours during the week (https://uhs.berkeley.edu/counseling/urgent)

- For other concerns, such as interpersonal conflicts with someone else on campus, call the ombuds office (https://studentaffairs.berkeley.edu/ombuds-office-for-students-postdoctoral-appointees/)

Other campus resources

- Accessing resources:
    - Basic Needs Center: https://basicneeds.berkeley.edu/pantry
    - Open computing facility: https://www.ocf.berkeley.edu/
    - Student Technology Equity Program: https://studenttech.berkeley.edu/step
    - Campus-provided software downloads: https://software.berkeley.edu/
- Community & campus spaces
    - Centers for Educational Justice and Community Support: https://cejce.berkeley.edu/ Links to Undocumented Students' Program, Gender Equity Resource Center, Multicultural Community Center, and several identity-specific community resources and groups
- Academic tutoring: Student Learning Center: https://slc.berkeley.edu/home
- Responding to distress or conflict
    - Resource from AAPI student development: "Maintaining Your Grades in a Mental Health Crisis"
    - Restorative Justice Center: https://rjcenter.berkeley.edu/

Readings and bibliography

- **Launching a team collaboration**
    - Skillshub Ltd., instructor, & Packt Publishing, publisher. (2024). *Feedback skills.* ([First edition].) [Video recording]. Packt Publishing. Chapter 1: "Feedback Skills".
    - Tannenbaum, S. I., & Salas, E. (2020). *Teams that work: the seven drivers of team effectiveness.* Oxford University Press. Chapter 11: "A Quick Refresher". Chapter 13: "Team Members: Being a Great Team Member"
    - Edmondson, A. C. (2012). *Teaming: how organizations learn, innovate, and compete in the knowledge economy.* (1st ed.). Jossey-Bass. Chapter 4: "Making it Same to Team" Chapter 5: "Failing Better to Succeed Faster".
- **Data research basics**
    - Lau, S., Gonzalez, J., & Nolan, D. A. (2023). *Learning data science: data wrangling, exploration, visualization, and modeling with Python.* (First edition.). O'Reilly Media. Chapter 1: "The Data Science Life Cycle". Chapter 2: "Questions and Data Scope". Chapter 3: "Simulation and Data Design". Chapter 4: "Wrangling Dataframes ".

    - Nolan, D. A., & Stoudt, S. (2021). *Communicating with data : the art of writing for data science.* (First edition.). Oxford University Press. Chapter 3: “Describing Data”.
- **Identifying and addressing challenges with data**
    - Jann, T. (2018) Understanding through documenting: The importance of seeing how your code fits into the bigger picture and how documentation can help facilitate that. Retrieved from https://stat198-spring18.github.io/blog/2018/04/03/understanding-through-documenting
    - Kosara, R. Data is a dish best served raw. https://eagereyes.org/data/dish-best-served-raw, Retrieved June 2009.
    - Nolan, D. A., & Stoudt, S. (2021). *Communicating with data : the art of writing for data science.* Oxford University Press. Chapter 5: “Communicating Through Code” .
    - Peng, R. (2011). Reproducible research in computational science. *Science*, 6060:1226–1227.
    - Kernigham, B.W. & Pike, R.(1999) *The Practice of Programming*. Addison-Wesley Professional, Boston, MA. Chapter 5: “Debugging”.
    - Rutta, Z. (2025, September 2). *Getting started on Stack Overflow: a step-by-step guide for students.* Stack Overflow. https://stackoverflow.blog/2025/09/02/getting-started-on-stack-overflow-a-step-by-step-guide-for-students/
    - Tannenbaum, S. I., & Salas, E. (2020). *Teams that work : the seven drivers of team effectiveness.* Oxford University Press. Appendix 2: “Team Debriefing Tips and Outline”
- **Communicating about your work**
    - Gillen, C. (2020) The data suggest: writing in the sciences. In G. Graff and C. Birkenstein, editors, They Say / I Say, pages 156–174. W. W. Norton & Company, New York, 2nd edition.
    - Herbst, E., & Kopf, S. (2024). Writing an abstract. *Arthroskopie*, *37*(4), 258–261. https://doi.org/10.1007/s00142-024-00688-5
    - Mott, N. M., Zope, M., Reynolds, I. S., Tadé, Y., Serra, G. P., Long, J. J., & Oslock, W. M. (2025). AJS virtual research mentor: Tips on writing an abstract for a conference. *The American Journal of Surgery*, *241*, Article 116009. https://doi.org/10.1016/j.amjsurg.2024.116009
    - Nolan, D. A., & Stoudt, S. (2021). *Communicating with data : the art of writing for data science.* Oxford University Press. Chapter 4: “Communicating Through Statistical Graphs”. Chapter 6: “Organizing the Story”.
    - Wainer, H. (1984) How to Display Data Badly. *The American Statistician*, 38(2), 137–147. https://doi.org/10.1080/00031305.1984.10483186
    - Wilcox, C., Brookshire, B., and Goldman, J. G., editors. (2016) *Scientific Blogging: The Essential Guide*. Yale University Press, New Haven.

- **Synthesizing your work into a final presentation**
    - Library guides: Posters, presentations & science writing: Start. (2025). Retrieved from https://guides.lib.berkeley.edu/posters
    - Faulkes, Z. (2021). *Better posters : plan, design and present a better academic poster.* (First edition.). Pelagic Publishing. Chapter 1: “Poster Design: The Short Form”. Chapter 4: “Design Thinking”. Chapter 6: “Narrative Thinking”.