AY 128: Astronomy Data Science Lab

UC Berkeley, Spring 2025

> This course consists of three data-centric laboratory experiments that draw on a variety of tools used by professional astronomers. Students will learn to procure and clean data (drawn from a variety of world-class astronomical facilities), assess the fidelity/quality of data, build and apply models to describe data, learn statistical and computational techniques to analyze data (e.g., Bayesian inference, machine learning, parallel computing), and effectively communicate data and associated scientific results. This class will make use of data from facilities such as *Kepler*, *Gaia*, the *Sloan Digital Sky Survey*, and the *Hubble Space Telescope* to explore the structure and composition of the Milky Way, stars, and galaxies throughout the local and distant Universe. There is a heavy emphasis on software development in the Python language, statistical techniques, and high-quality communication (e.g., written reports, oral presentations, and data visualization).

Classroom: In person Tuesday and Thursday 2:00 to 3:30 PM, 131 Campbell. Attendance and engagement with lectures is very important. Please have your computer ready for interactive demos and to work on your labs in class.

Website: https://ucb-datalab.github.io

Office Hours: see course website.

Labs: This class features 3 labs plus one warm-up activity (Lab 0). We will introduce the topic and contents of each lab throughout the semester. You will generally have 3-4 weeks to complete each lab and write up a report. There are required weekly checkpoints in each lab. Labs and checkpoints will be submitted via Gradescope.

Quizzes and Readings: There are no formal exams for this class.

Communication: Our preferred method of communication is through the class Ed Discussion site. You will receive a link to sign up for our channel. Email is OK for logistical items (e.g., scheduling a meeting time with an instructor).

Recommended Textbooks:

"Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data" by Željko Ivezić et al.

"An Introduction to Modern Astrophysics" by Bradley W. Carroll & Dale A. Ostlie

"Foundations of Astrophysics" by Barbara Ryden & Bradley Peterson

and select readings to be provided.

Prerequisites:

- This class assumes that you have completed introductory astrophysical instruction (Astro 7A and 7B, or higher such as Astro 160/161) as well as knowledge of calculus (including Math 53) and linear algebra (Math 54 or Physics 89)
- You should have proficiency or fluency in the Python programming language. This class heavily emphasizes software development, and is **not** the place to learn Python for the first time.

Lab Reports (60% of total grade):

- Submitted as lab reports along with a polished, executable Jupyter Notebook.
- due **before** specified due date/time.

- Late Labs & "Slip days":
  - Each person is allocated 8 slip days for the semester.
    - These are days you can use to turn in a lab report late without any penalty by submitting the Slip Day Form which you can find here. You must submit your slip day form no more than 24 hours after the normal lab due date and time. You are not able to retroactively change the number of slip days used on an assignment once the form is submitted.
    - You can allocate them as you see fit throughout the semester.
    - **NO** extensions for labs will be granted beyond the 8 slip days under any circumstances.
    - Slip days do not apply to checkpoints, only to labs.
    - Once all slip days are used, then late lab grading policy is -10% x raw score x number of days late. For example, a lab submitted 2 days late that got a raw score of 90% would be reduced by -.10 x 90 x 2 = -18 pt, yielding a total of 72 points for the lab.
    - Each unused slip day at the end of the semester is worth 0.25% extra credit toward your final class grade.
- Collaborate (talk, draw pictures, analyze data) with your lab mates, but you **MUST** implement separately (your own equations, code, plots, writing)
- **Lab 0: Introduction to ADQL and Gaia Data** [50 points]
- **Lab 1: Gaia, RR Lyrae stars, and Galactic Dust** [100 points]
- **Lab 2: Modeling Stellar Spectra** [100 points]
- **Lab 3: Galaxy Morphology with Convolutional Neural Nets** [100 points]

Lab Checkpoints (30% of total grade):

- Turning in checkpoints (11 2 point checkpoints for 50% of checkpoint grade)
  - They are graded on completeness and not on accuracy (i.e., it's OK to submit work-in-progress). Though simply submitting filler material without *any* substance will be considered incomplete.
- Presenting at least one check point to the class and leading related discussion (50% of checkpoint grade).
- The **two** lowest checkpoint scores will be dropped from your final grade. No checkpoint extensions are allowed.

Class Participation (10% of total grade):

- Participation can take many forms including asking questions during lecture, leading group discussion, posting questions AND answers to others questions on Ed Discussion, attending office hours, and more.
- The key is for you to be actively engaged in the class.

Reading:

- Lab instructions and topical handouts linked on the class webpage

Materials:

- Each student will be given access to Datahub (https://astro.datahub.berkeley.edu) to execute their code.

Class Philosophy:

This class is effectively an introduction to astronomical research from a data-centric perspective. The labs are designed to be challenging, yet manageable mini-research projects. They will be time-consuming.

Often when doing research, you will not have all the necessary technical skills to complete a project. It is normal to have to learn entirely new skills along the way. A main aim of this class is to provide exposure

to those skills, in a setting that requires you to think and act like a researcher, but with extra guidance and resources. You should also get used to asking questions that seem basic (sometimes embarrassingly so) in hindsight, and working in open and collaborative environments.

The weekly lectures will provide you with broad background, but are not intended to teach all the nitty-gritty technical skills you need. Much like real research, these are skills you learn on your own, in groups, and/or in consultation with an advisor. The equivalent with this class are questions via Ed Discussion and office hours, particularly with the GSIs.

Research is fun, but challenging. It usually does not progress linearly, and you often have to take a few steps backward before ultimately moving forward.

Class Conduct:
This is a work-intensive class. You are going to spend significant time on your own in the lab with minimal supervision. At all times, you are expected to abide by the UC Berkeley Code of Conduct (https://sa.berkeley.edu/code-of-conduct), acting with respect to your peers, GSIs, and instructor. Should you experience any form of harassment or discrimination, we maintain a list of resources that can help you decide how to respond. (https://astro.berkeley.edu/department-resources/reporting-harassment). GSIs and instructors are non-confidential reporters; we have a legal obligation to act on any reports of harassment. Please know that we take our responsibility seriously.

Collaboration Policy (don't cheat):
The Data Lab is a collaborative class. To repeat what is written previously in the syllabus: while you should talk with others about the labs and we strongly encourage collaboration, we require that you write your solutions and reports individually.

Use of AI:
You are not allowed to use AI (e.g., ChatGPT) to aid with or generate submitted material for this class. Such use of AI will result in a zero and, depending on the situation, may also be considered grounds for academic misconduct. You may use AI in other ways outlined in the department AI use policy, which you can find here. **When in doubt, ask for clarification from the instructors.**

DSP Accommodations:
UC Berkeley is committed to creating a learning environment that meets the needs of its diverse student body including students with disabilities. If you anticipate or experience any barriers to learning in this course, please feel welcome to discuss your concerns with me.

If you have a disability, or think you may have a disability, you can work with the Disabled Students' Program (DSP) to request an official accommodation. The Disabled Students' Program (DSP) is the campus office responsible for authorizing disability-related academic accommodations, in cooperation with the students themselves and their instructors. You can find more information about DSP, including contact information and the application process here: dsp.berkeley.edu. If you have already been approved for accommodations through DSP, please meet with the course instructors so we can develop an implementation plan together. Note that per University policy, we can only provide DSP accommodations if an official letter is provided by the DSP office.

Your success!
We want you to succeed in this class. We know that Berkeley can at times be an overwhelming place. If you find yourself struggling to keep up, please come talk to us, we're here to help you succeed.