Ay128/256: Astronomy Data Science Lab

UC Berkeley, Spring 2024

> 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 software development in the Python language, statistical techniques, and high-quality communication (e.g., written reports, oral presentations, and data visualization).

Classroom: In person, 131 Campbell. Attendance and engagement with lecture 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 check-ins in each lab. We will provide details for how to submit your labs and check-ins during the first full week of class.

Quizzes and Readings: There are no formal exams for this class. However, to help with reading, we will have short interactive ungraded quizzes during class that will focus on key concepts from assigned readings and related topics.

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 (65% of total grade):

- Submitted as lab reports along with a polished, executable Jupyter Notebook.
- due **before** specified due date.
- Late Policy:
  - 1 day - 5 pts deducted

- 2 days - 7.5 pts deducted
- 3 days - 10 pts deducted
- 4 days - 12.5 pts deducted
- 7 days - 15 pts deducted
- 14 days - 25 pts deducted

- 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** [10% of lab grade]
- **Lab 1: Gaia, RR Lyrae stars, and Galactic Dust** [30% of lab grade]
- **Lab 2: Modeling Stellar Spectra** [30% of lab grade]
- **Lab 3: Galaxy Morphology with Convolutional Neural Nets** [30% of lab grade]

Lab Check Points (25% of total grade):

- Check Points are due weekly as indicated in each lab.
- 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.
- Be ready to discuss the assigned checkpoints at lecture each week.

Class Participation (10% of total grade):

- During virtual instruction participation can make many forms including asking questions during lecture, leading group discussion, interactions in chat on Zoom, 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, we ask that you write your solutions individually.

Use of AI: You are not allowed to use AI (e.g., ChatGPT) to aid with or generate submitted material for this class. The use of AI will result in a zero and, depending on the situation, may also be considered grounds for academic misconduct.

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.