



Data Science Newsletter

Greetings from Data Science Advising. Welcome to the ninth issue of our Newsletter! If you missed previous issues, make sure to check them out [here](#).

Major Updates



Data Science Honors Program 2019-2020 Deadline

The Data Science Honors Program for research-focused students in the Data Science BA Program is a year-long research and thesis project conducted over 2 consecutive semesters. Students will also complete 2 courses as part of the honors program. Please see the [application packet](#) for more information and details.

This is a reminder that applications for the 2019-2020 academic year are due by **11:59pm on Sunday, July 7, 2019**. Applications should be submitted to ds-advising@berkeley.edu or in person to 436 Evans Hall.

Fall 2019 Data Science Major Upper-Division Units Minimum

Effective in Fall 2019, the Data Science BA requires a minimum of **28 upper-division units**. The major still requires a minimum of 7 upper-division units in the Computational & Inferential Depth requirement.

Students graduating in Fall 2019 or later may follow the new requirement. Students graduating prior to Fall 2019 are held to the previous requirement of **30 upper-division units** (please consult with Data Science advisor if you have questions or concerns about your own situation).

New Data Science Connector Courses!

Connector courses weave together core concepts and approaches from Data 8 with complementary ideas or areas. Along the way, students gain additional experience, broader insights, or deeper theoretical or computational foundations.

Offered by faculty across many departments and fields of study, connectors are optional but **highly encouraged** and are designed to be taken at the same time or after the Foundations course.



Check out two of our awesome, *new* Fall 2019 [Connector Courses](#):

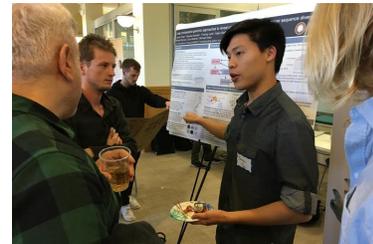
- [Data 88 001 - Economic Models](#)
 - This Data Science connector course will motivate and illustrate key concepts in Economics with examples in Python Jupyter notebooks. The course will give data science students a pathway to apply python programming and data science concepts within the discipline of economics. The course will also give economics students a pathway to apply programming to reinforce fundamental concepts and to advance the level of study in upper division coursework and possible thesis work.
- [Data 88 002 - Data Science in Genetics and Genomics](#)
 - Recent years have witnessed a rapid expansion in the creation and utilization of genetic and genomic data across diverse domains such as business, biological research, and medicine. In this Data 8 connector course we will survey relevant questions of interest and employ the methods frequently relied upon by analysts to derive insights from genetic and genomic data. Topics will include the comparison of DNA sequences, dimension reduction, the characterization of transcriptomes, and genome-wide association studies, among others. In addition to hands-on work with data, we will also consider the history of the genetic and genomic sciences and their intersection with current events, ethics, and modern medicine. Students should exit with an understanding of the central role played by data in the fields and an appreciation for the remaining challenges in light of ever-increasing degrees of personalization of, and access to, these sciences. No biological background is required.

Student Opportunities

Discovery Research Program Fall 2019 Project Application

In light of the growing demand for research opportunities, we are now allowing undergrads to **pitch their own research ideas** to lead a project in the fall! We are working on a pipeline to connect students like yourself to faculty and grad students who will review your proposals.

While we can't guarantee that your project will be selected, we are committed to providing as much support as we can.



If your project is selected and vetted, you will be asked to present it at a **Student Pitch event on August 26th** (location and time TBA) open to all Discovery Students. It will be an opportunity for Discovery Applicants to form teams immediately and bypass the typical interview process. If a project team is formed, then it will be part of the Fall program and a member from the review board will be assigned to provide mentorship to you throughout the semester. If this interests you, then we encourage you to **apply with a project idea by July 26th**. Fall project applications are now open!

Find the Application [HERE](#).

Please email ds-discovery@berkeley.edu if you have any questions.

Fall 2019 Cal-ADAR Program Application (Deadline Extended)

Fall 2019 deadline extended! [Cal-ADAR](#), Berkeley's mentored research program in demography of aging for underrepresented undergrads is now accepting applications for its fall 2019 cohort. Demography of aging uses data to examine life course – how early life events or situations influence later life – and population studies – births, deaths, migrations. For the fall 2019 semester, we will accept applications through **August 15, 2019**. Through financial support, mentoring, internships and other opportunities, Cal-

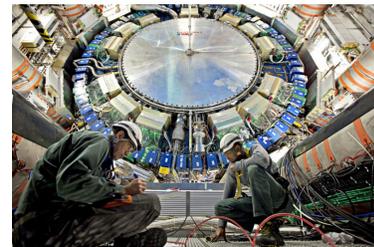
ADAR will provide you a frame of reference on graduate school in some area of population research: [sociology](#), [public health](#), [economics](#), [public policy](#), [geography](#), or another field. You will also gain marketable skills and experience for after graduation. To start, you need to meet eligibility requirements, be interested in graduate school as a possible goal, and want to learn how to work with data.

Cal-ADAR provides students with a community in which to develop quantitative social science skills and use them to study socially relevant issues. Each year, **up to 6 Cal-ADAR scholars** are selected for the program, which takes place over their junior and senior years and the summer in between (seniors with at least 2 semesters left are also eligible). Scholars are given **generous financial support (partial tuition/fees, up to 3 semesters)**, a **paid summer internship**, travel stipend for conference travel and support and mentoring from faculty members and Cal-ADAR staff through graduation and beyond. Cal-ADAR scholars have gone on to work in local government, finance, academic research. Many are actively pursuing graduate school.

For information about benefits, eligibility, requirements, and applying please visit [Cal-ADAR's Website](#). Contact Dr. Mao-Mei Liu (mao-mei_liu@berkeley.edu) with questions.

ATLAS Experiment Paid Summer Internship

The [ATLAS experiment at CERN](#) is a playground for Big Data. Over the past few years, it has produced 250 Pb of data that is being constantly analyzed by physicists all over the world. The typical workflow consists of running simple operations to select the most interesting points from the very large input dataset. Doing this efficiently requires clusters designed for High Throughput Computing (HTC). However, the recent trend in information technology is towards High Performance Computing (HPC). An example is the [Cori supercomputer at Berkeley Lab](#). It consists of thousands of multi-core nodes, with either Haswell or Knights Landing CPUs, with very limited IO bandwidth.



We are searching for a student interested in data processing algorithms to help the Berkeley ATLAS group leverage the most from Cori. The project would include studying the resource requirements of the ATLAS group, developing a job packing algorithm to run single-threaded programs on multi-core CPUs and testing the performance of different CPU architectures for typical workflows. This position is open until filled.

Please contact: Zachary Marshall (zmarshall@lbl.gov) Karol Krizka (kkrizka@lbl.gov) and Haichen Wang (haichenwang@berkeley.edu) for more information.



Division of Student Affairs

Data Analytics and Business Intelligence Assistant (Work Study Position)

The Division of Student Affairs is the largest division on campus, comprising more than 3000 employees and key student services, such as Cal Student Central, Financial Aid & Scholarships, Career Center, University Health Services, Recreational Sports, Cal Housing, and Cal Dining, among others.

Student Affairs Data Analytics provides decision support, reporting, and process improvement solutions to the Division. The unit's main functions are to (a) build and maintain a data analytics infrastructure, (b) develop dashboards for standardized reporting and business intelligence, (c) conduct data analysis to support data-informed decision making, and (d) create automation solutions for process improvement.

Two positions are available for Fall 2019 and Spring 2020 for students with Work Study financial aid awards. A start date in Summer 2019 may be possible; however, these are not summer-only positions.

The work of the Data Analytics and Business Intelligence Assistant may include the following:

- Clean, reshape, merge, and compare data sets using Excel, Python, and other

tools

- Validate data by comparing final reports and dashboards to source systems and intermediate data sources
- Write scripts in Python and Google Apps Script to automate processes, such as API calls, report generation, approval workflows, and reconciliation
- Debug scripts and propose solutions
- Analyze and visualize data for decision support using Tableau, Excel, Google Sheets, Python, and related technologies
- Review the work of other analysts for accuracy and efficiency and provide feedback for improvement
- Conduct testing of scripts and reports for functionality and accuracy
- Update data sets manually when automation is not possible
- Research options for meeting project requirements and present alternatives
- Gather requirements and respond to inquiries from functional partners
- Write project and code documentation
- Assist with administrative tasks of the unit, such as reviewing file and system permissions, responding to emails, organizing files, and other work as assigned.

Expectations:

- Commit to 10-15 of work hours per week, with some flexibility, especially around exam times
- Be fully engaged with job-related tasks during work shifts
- Meet deadlines and achieve project milestones and deliverables
- Arrive on time to all scheduled shifts
- Inform supervisor of planned absences in advance
- Maintain strict confidentiality and objectivity
- Continued employment is contingent upon a satisfactory job performance evaluation by the supervisor

Qualifications:

- Strong attention to detail and commitment to accuracy
- Advanced knowledge of programming in Python, Google Apps Script (Javascript), or similar
- Knowledge of data structures and systems (e.g., tabular data, relational databases, JSON)
- Ability to read and transform data using spreadsheet and scripting techniques, such as pivoting, aggregating, reshaping, and joining data
- Ability to analyze data using intermediate statistical techniques, interpret results, and clearly communicate findings visually and in writing
- Ability to understand business processes and project requirements
- Excellent verbal, written, and interpersonal communications skills
- Willingness and ability to learn new technologies and programming techniques
- Ability to follow specific and general instructions
- Upper-division undergraduate student or graduate student

Compensation: The pay rate for this position is \$17.50 per hour.

Application: <https://forms.gle/8Vg5bKq4qCC8hEdT6>

- (Apply **ASAP** if you're interested! Review of applications will begin soon.)

Questions: studentaffairsdataanalytics@berkeley.edu

Fall 2019 Courses

PSYCH 101D - Data Science for Research Psychology

This Python-based course builds upon the inferential and computational thinking skills developed in the Foundations of Data Science course by tying them to the classical statistical and research approaches used in Psychology. Topics include experimental design, control variables,



reproducibility in science, probability distributions, parametric vs. non-parametric statistics, hypothesis tests (t-tests, one and two way ANOVA, chi-squared and odds-ratio), linear regression and correlation.

Please note that students will not receive credit for Psych 101D after having completed Psych 10 or Psych 101. Deficient grade in Psych 10 can be replaced with Psych

101D. Deficient grade in Psych 101 can be replaced with Psych 101. Find more information on [Psych 101D's course listing](#).

ENERES 131 - Data, Environment and Society

This course will teach students to build, estimate and interpret models that describe phenomena in the broad area of energy and environmental decision-making. Students leave the course as both critical consumers and responsible producers of data-driven analysis. The effort will be divided between (i) learning a suite of data-driven modeling and prediction tools (including linear model selection methods, classification and regression trees and support vector machines) (ii) building the programming and computing expertise to use those tools and (iii) developing the ability to formulate and answer resource allocation questions within energy and environment contexts.



We will work in Python in this course, and students must have taken Data 8 before enrolling. The course is designed to complement and reinforce Berkeley's data science curriculum. Find more information on [ENERES 131's course listing](#).
