



**DATA 8**  
Spring 2022

# Lecture 40

---

Conclusion

# Announcements

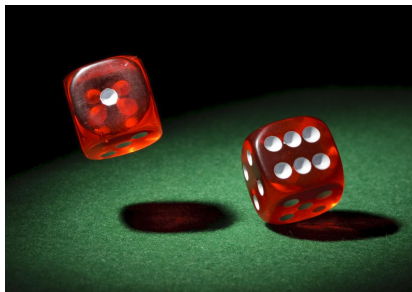
---

- Project 3 due tonight 11:59pm PT
- Homework 13 due Thurs, 05/05
- Final Exam on Tue, 05/10 3-6pm PT
  - Please fill out [final exam conflict form](#) by Mon, 05/02 11:59pm PT

# **Course Overview Slides**

**What's Next?**

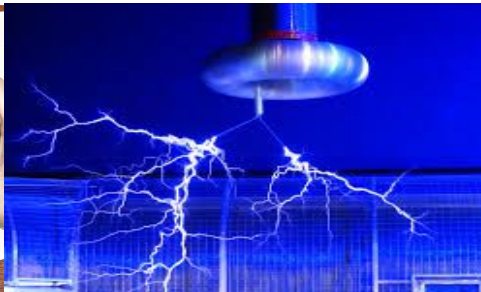
# Fall 2022 Connector Courses



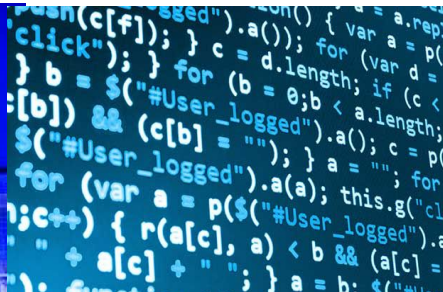
**Data C88S (Stat 88)**  
Prob and Stats in Data Science



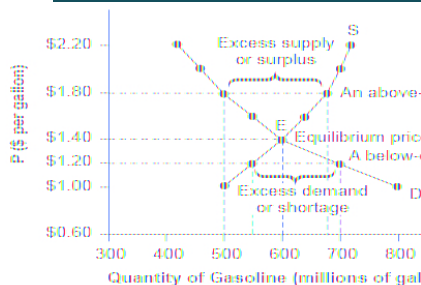
**UGBA 88**  
Data and Decision



**PHYSICS 88**  
Data Science Applications in Physics



**Data C88C (CS 88)**  
Computational Structures



**DATA 88E**  
Economic Models



**POLISCI 88**  
Scientific Study of Politics



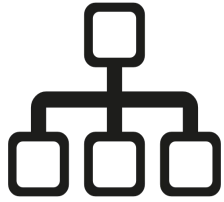
**EPS 88**  
Python and Earth Science



**LEGALST 88**  
Taking Measure of the Justice System

# DSUS Student Teams

*Hone your skills as an educator and data scientist by working with Data Science Undergraduate Studies*



## Infrastructure

Improve autograding and DataHub software to support courses across campus



## Peer Consulting

Help fellow undergrads with data research, academic work, and data science technology.



## External Pedagogy

Create a national community of practice for institutions to work with and learn from each other.

# DSUS Student Teams

---

*Hone your skills as an educator and data scientist by working with Data Science Undergraduate Studies*



## **Connector Assistants**

Help instructors of Data Science Connector courses deliver and teach material.



## **Modules**

Create curriculum materials for Connectors, Data-Enabled Courses, or short explorations into DS (modules).



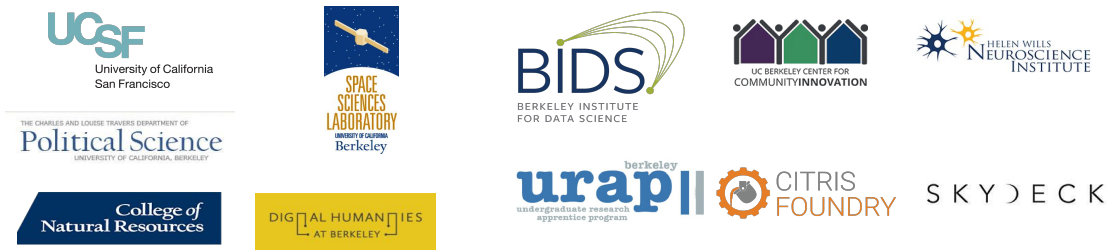
## **Human Context and Ethics**

Integrate critical thinking about ethical issues in relation to technology into the Berkeley data science program and community.

# Data Science Discovery Research

---

*Be a **student researcher** in a program that connects students with hands-on data science research- non-profits, start-ups, institutions, etc. Students from underrepresented minority groups and first-time researchers receive priority.*



**<https://data.berkeley.edu/discovery>**

**Project Showcase Next Tuesday 5/3 2pm-4pm: [bit.ly/discoverysp22](https://bit.ly/discoverysp22)**



# Programming

---

The programming content in Data 8 is part of what you'll learn about programming in CS 88 or CS 61A.

What's left?

- How to write larger programs and think about them.
- The concepts and language features that support writing larger programs.
- How programming languages are executed.

CS 88 is 3 out of 4 units of CS 61A, but with more connections to data science in the examples.

---

# Human Contexts and Ethics

---

Data science studies the real world, and there are important ethical considerations in doing so.

- The impact of data collection and analysis
- Fairness and bias in both data collection and prediction
- Institutions that use data, such as companies & gov't
- The relationship between data and the law
- Frameworks for reasoning about these complex issues

Data C104 and Info 188 are the most popular courses for students continuing from Data 8.

---

# Probability

---

The probability content in Data 8 is part of what you'll learn about probability in a lower-div probability course:

Data C88S (Stat 88), CS 70, Math 10B or 55, CivEng 93

While the Data Science major does not require a lower-division probability course, taking one is a good idea.

- Understanding random events and probabilities for both categorical and numerical variables.
  - Concepts for reasoning about randomness.
  - Characteristics of commonly encountered distributions.
-

# Data 100

---

Prerequisites: Data8 & programming

Co-requisite: Linear algebra

John recommends taking linear algebra before.

John recommends taking lower-division probability before.

Very much a sequel to Data 8:

- Data manipulation and visualization
  - Linear regression, but with multiple variables
  - Prediction and inference
-

**Staff AMA**

# Data Science

# Why Data Science

---

- Unprecedented access to data means that we can make new discoveries and more informed decisions
  - Computation is a powerful ally in data processing, visualization, prediction, and statistical inference
  - People can agree on evidence and measurement
  - Data and computation are everywhere: understanding and interpreting are more important than ever
-

# Limitations of Data Science

---

- Evidence and measurements are critical ingredients for good decision-making
    - ...but they're not enough by themselves!
  - Data science is a powerful complement to qualitative analysis
    - ...but it's not a replacement!
-



# How to Analyze Data

---

Begin with a question from some domain, make reasonable assumptions about the data and a choice of methods.

Visualize, then quantify!

*Perhaps the most important part:* Interpretation of the results in the language of the domain, without statistical jargon.

---

# How *Not* to Analyze Data

---

Begin with a question from some domain, make reasonable assumptions about the data and a choice of methods.

Visualize, then **quantify!**

*Perhaps the most important part:* Interpretation of the results in the language of the domain, without statistical jargon.

---

# How to Analyze Data after Data 8

---

Begin with a question from some domain, make reasonable assumptions about the data and a choice of methods.

Visualize, then quantify! Do both using computation.

*Perhaps the most important part:* Interpretation of the results in the language of the domain, without statistical jargon.

---

# The Design of Data 8

---

- Table manipulation using Python
  - Working with whole distributions, not just means
  - Decisions based on sampling: assessing models
  - Estimation based on resampling
  - Understanding sampling variability
  - Prediction
-

# **One Last Thought**

# My Journey Through Cal

---

2009: Joined Cal 🐻

2011: Added Statistics major 📊

2013: Started PhD 🗽

2015: Interned as a Data Scientist ⚡

2017: Joined Autopilot 🚗

2019: Left to work on a startup 🎾

+ Returned to Cal as a lecturer 🐻

---

# Thank you!

---





# **Complete the NSF Survey about Data 8!**

You received a link in your email