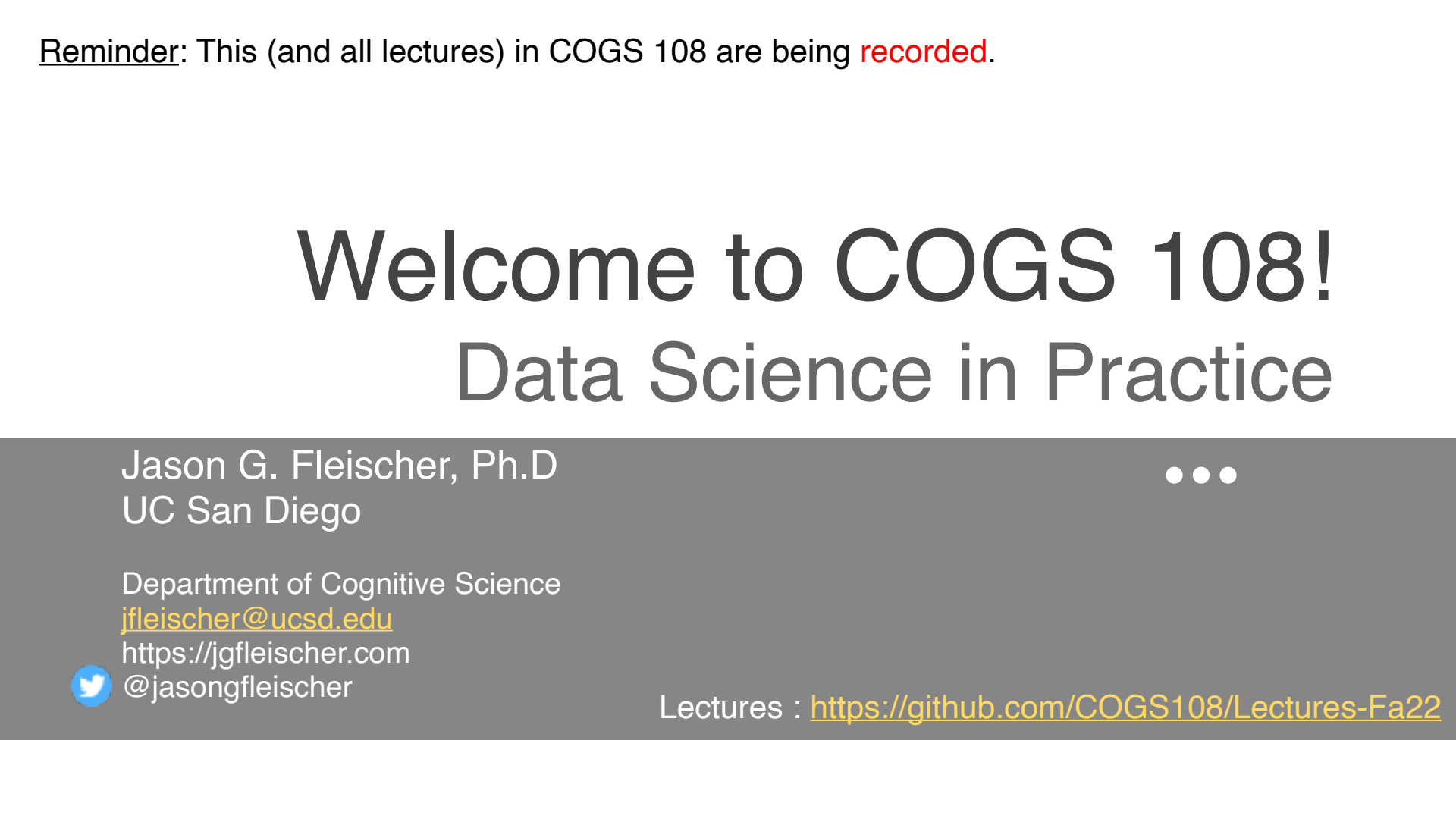
Reminder: This (and all lectures) in COGS 108 are being recorded.

Welcome to COGS 108!

Data Science in Practice

Jason G. Fleischer, Ph.D
UC San Diego

Department of Cognitive Science
jfleischer@ucsd.edu
https://jgfleischer.com
@jasongfleischer



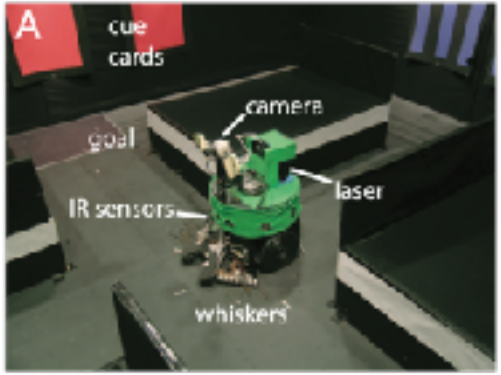
Lectures : https://github.com/COGS108/Lectures-Fa22

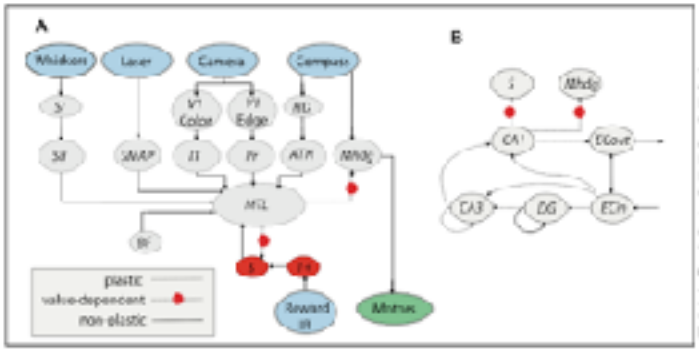
hello

my name is

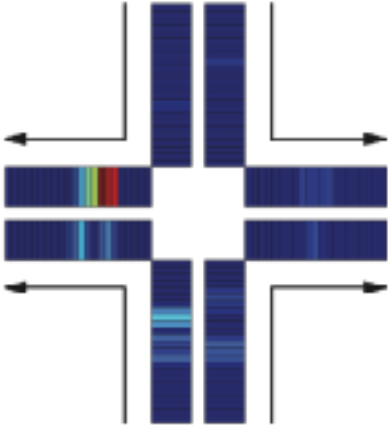
Prof Fleischer

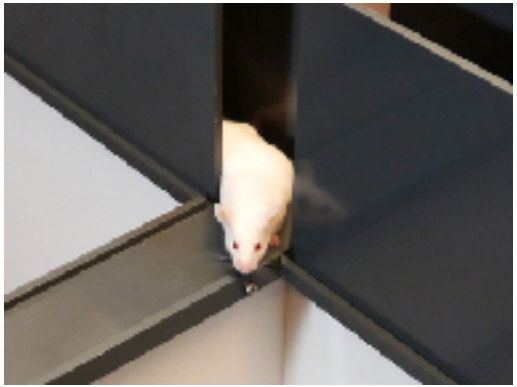
Wikimedia Commons
Sunlight on Colorado National Monument.jpg
By Meelmouse

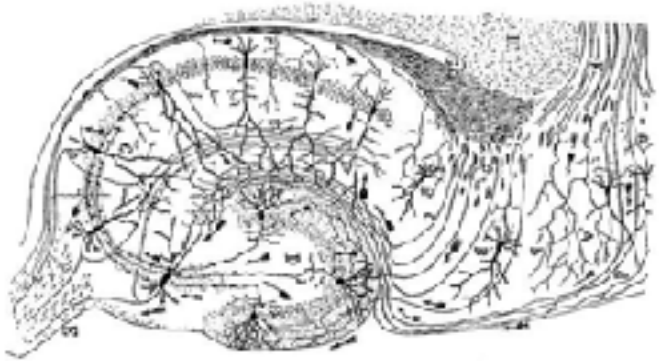


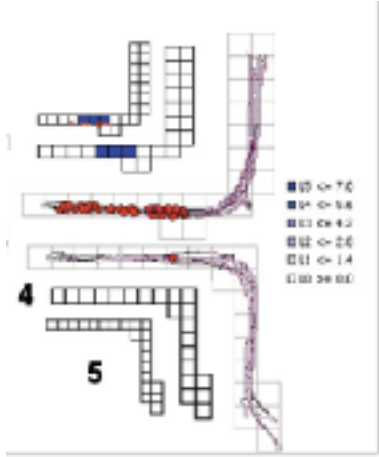


PNAS 104:3556-3561









RESEARCH ARTICLE | DEVELOPMENTAL NEUROSCIENCE

Sleepmore in Seattle: Later school start times are associated with more sleep and better performance in high school students

Gideon P. Dunster[1], Luciano de la Iglesia[1], Miriam Ben-Hamo[1], Claire Nave[1], Jason G. Fleischer[2], Estibaliz...
• See all authors and affiliations

Science Advances 12 Dec 2018:
Vol. 4, no. 12, eaau6200
DOI: 10.1126/sciadv.aau6200

Abstract

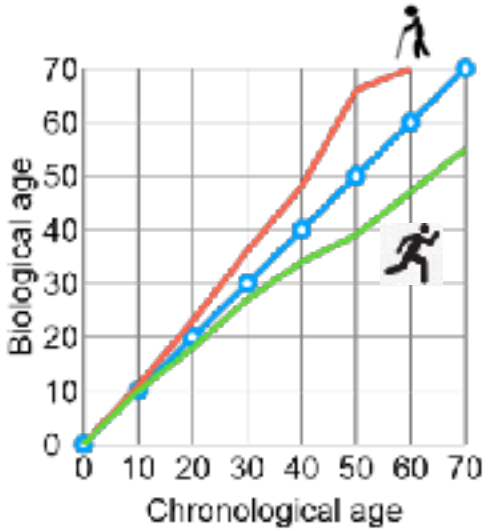
Most teenagers are chronically sleep deprived. One strategy proposed to lengthen adolescent sleep is to delay secondary school start times. This would allow students to wake up later without shifting their bedtime, which is biologically determined by the circadian clock, resulting in a net increase in sleep. So far, there is no objective quantitative data showing that a single intervention such as delaying the school start time significantly increases daily sleep. The Seattle School District delayed the secondary school start time by nearly an hour. We carried out a pre-/post-research study and show that there was an increase in the daily median sleep duration of 34 min, associated with a 4.5% increase in the median grades of the students and an improvement in attendance.

Cell Metabolism

Clinical and Translational Report

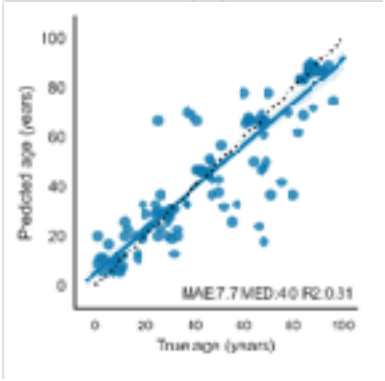
Ten-Hour Time-Restricted Eating Reduces Weight, Blood Pressure, and Atherogenic Lipids in Patients with Metabolic Syndrome

Michael J. Wilkinson,[1,3] Emily N.C. Manoogian,[1,3] Adena Zadourian,[1] Hannah Lo,[1] Savannah Fakhouri,[1] Amrin Shoghi,[1] Xinran Wang,[1] Jason G. Fleischer,[1] Sarah Navlakha,[2] Satchidananda Panda,[1,4,*] and Pam R. Taub[1,4]

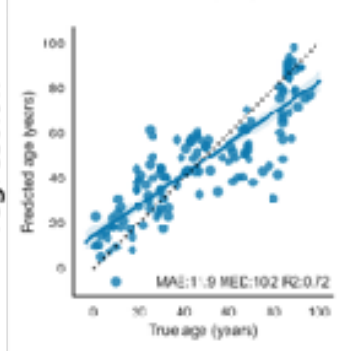


Hallmarks of Aging, López-Otín et al. Cell. 2013 Jun 6; 153(6): 1194–1217

Ensemble LDA



Support vector regression



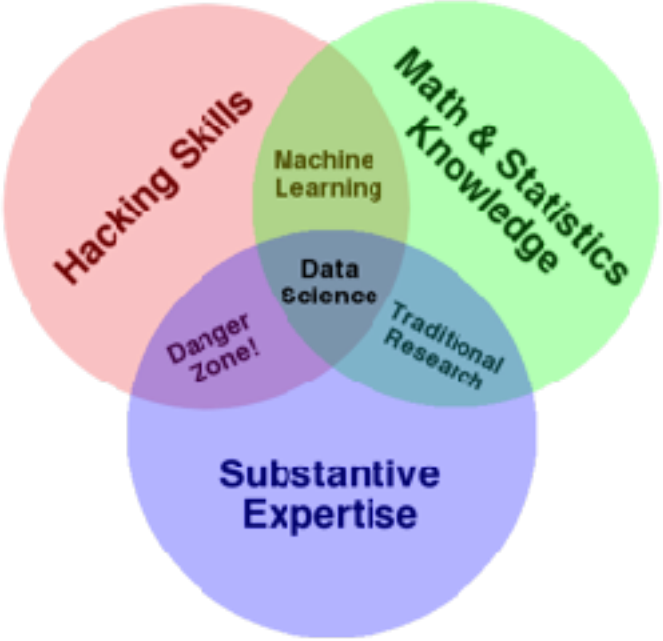
3

FINAL

0

SAN DIEGO WAVE FC

What is data science?



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Why this course?

You are going to be analyzing lots of data because you're studying to be a:

economist
journalist
neuroscientist
small business owner
sports analyst
biologist
or ?????

50 Best Jobs in America for 2022

Best Places to Work | Top CEOs | **Best Jobs** | Best Cities for Jobs | Highest Paying Jobs

2022 | United States

Discover Glassdoor's Best Jobs in 2022

Using Glassdoor's unique data on jobs, salaries, and companies, we compiled a list o
America to help people find jobs they'll love. Each job stands out for its earning pot
job satisfaction, and job openings. Are you considering a new position? Check out th
to see what jobs made the list this year, and view open jobs at companies across the

	Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1	Enterprise Architect	$144,997	4.1/5	14,021
#2	Full Stack Engineer	$101,794	4.3/5	11,252
#3	Data Scientist	$120,000	4.1/5	10,071

Data scientist is actually MANY jobs

https://hbr.org/2018/11/the-kinds-of-data-scientist

A final piece of advice for those hiring data scientists: Look for people who are in love with solving problems, not with specific solutions or methods, and for people who are incredibly collaborative. No matter what kind of data scientist you are hiring, to be successful they need to be able to work alongside a vast variety of other job functions — from engineers to product managers to marketers to executive teams. Finally, look for people who have high integrity. As a society, we have a social responsibility to use data for good, and with respect. Data scientists hold the responsibility for data stewardship inside and outside the organization in which they work.

Data science for humans

Data science for computers

Data scientists ask interesting questions & answer them with data

The goal in COGS 108 is to *do* data science.

Course Objectives

- Formulate a plan for and complete a data science project from start (question) to finish (communication)
- Explain and carry out descriptive, exploratory, inferential, and predictive analyses in Python
- Communicate results concisely and effectively in reports and presentations
- Identify and explain how to approach an unfamiliar data science task

How we'll approach learning about *and doing* data science in COGS 108

Scheduling & Staff

Lecture: MWF 10-10:50pm
Discussion Sections: M & F various times
Office Hours: BOOK on gcal!

TAs	IAs
Akshay Nagarajan	Albert Kong
Chaolin Lin	Christian Kim
Heeket Mehta	Cody Li
Shanay Shah	Jason Chen
	Winnie She

COGS 108: General Plan

Week	Topic(s)
1	Data Science, Python, & Version Control
2	Data Intuition & Wrangling
3	Data Ethics & Questions
4	Data Visualization & Data Analysis
5	Inference
6	Text Analysis
7	Machine Learning
8	Nonparametric Analysis
9	Geospatial Analysis
10	Data Science Communication & Jobs

Programming Prerequisite

- MAE 8 - MATLAB
- CSE 8A or 11 - Python/Java
- COGS 18 - Python
- DSC 10 - Python

Bottom line: we will assume programming knowledge.

Python will be used for all labs/projects/assignments.

No programming experience (or you forgot it all)?

- *Preferred option*
  - Take a programming course first
  - COGS 18 : Introduction to Python
- *Can't wait?*
  - Use online sites like *codecademy.com* or *LearnPython.org*
  - *Python Data Science Handbook*

Course links

GitHub	https://github.com/COGS108	lecture/section materials & final projects
datahub	https://datahub.ucsd.edu	assignment submission
Piazza	**https://piazza.com/ucsd/fall2022/cogs108**	questions, discussion, and regrade requests
Canvas	https://canvas.ucsd.edu/courses/39457	grades, lecture videos
Anonymous Feedback	https://forms.gle/3xiCoRQYFaiBnMSu5	if I ever offend you, or to provide general feedback on what's going well or badly

Discussion Section

- Goals:
  - MORE chance for individual contact
  - help with technical aspects of the course
  - assignment & project help
- Can I switch sections? Yes, but stick with one for the duration
- You'll never be required to go to section.
  - Do lab exercises on your own if you feel comfortable with material
  - Questions via Piazza if you can't attend
- At least one section is always recorded

Discussion Sections start Fri of next week (week 1)!

General grading:

	% of Total Grade
(8/9) Weekly Quizzes (lecture content)	8
(8) Discussion Labs (technical)	16
(4+1) Assignments	33
Final Group Project	44
(1) Project Review*	5
(1) Project Proposal*	8
(2) Project Checkpoints*	10
(1) Final Report*	15
(1) Final Video*	3
(1) Team evaluation survey	1

Attendance is neither required nor incentivized

- All lectures will be recorded (available end of day Canvas Media Gallery)
- Simultaneously on zoom https://ucsd.zoom.us/j/99741596334
- One discussion section each week will be recorded

Weekly Lecture Quizzes:

- (9) weekly quizzes (first one due Friday of Week 2)
- Goal: to help you keep on top of the material covered in lecture
- Why?: experience + student feedback
- How:
  - Taken on Canvas
  - Single Attempt
  - ~10 Questions
  - Posted by Friday sometime after class and before midnight; due the following Mon
  - Meant to test concepts from previous week's lecture

Lecture quizzes will be due on Mon by 11:59 PM.

Lowest quiz score will be dropped.

NO LATES

(4 + 1 practice) Assignments, 8 Discussion Lab exercises

Both are completed individually and graded programmatically.

- These are meant to get you practice programming around the topics covered in class.
- The first two are much simpler than the following two and should take less time.
- You will have to look some stuff up on your own. This is by design.
- Instructions must be followed to receive credit.
- You'll have the opportunity to practice in discussion section.

Assignments will be due on Fridays by 11:59 PM

Discussion labs will be due on Wednesdays by 11:59PM

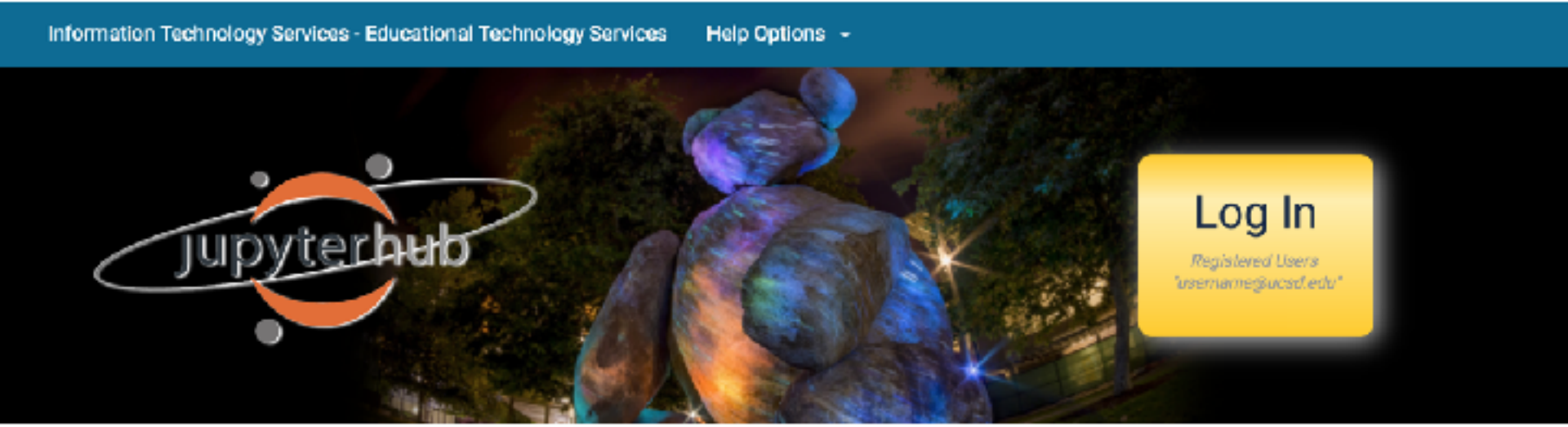
75% credit if submitted less than one week after deadline.

5 LATE DAYS allowed per person this quarter without penalty

Assignment Submission @ Datahub: https://datahub.ucsd.edu

DATA SCIENCE / MACHINE LEARNING PLATFORM

UC San Diego



UC San Diego Jupyterhub (Data Science) Platform

Before next Fri: log onto datahub & have a working installation of Jupyter on your computer

Group Projects: the main focus of COGS 108

Groups of 4-5 Individuals

How to find a group:

1. go to discussion section week 1
2. post on Search for Teammates on Piazza category
3. talk to people you are sitting near after class

COURSE SCHEDULE

	Week	Day	Topic	Section covers	Lab due	Assignment due	Lecture quiz due
9/23/2022	0	F	Welcome!	--		--	--
9/26/2022	1	M	Python Review	--		--	--
9/28/2022	1	W	Version Control I	--		--	--
9/30/2022	1	F	Version Control II	D1		Practice assignment	--
10/3/2022	2	M	Data & Intuition	D1			Q1
10/5/2022	2	W	Data Wrangling (pandas)	--	L1	--	--
10/7/2022	2	F	Ethics	D2		A1; Group Signup*	--
10/10/2022	3	M	Data Science ?s	D2		--	Q2
10/12/2022	3	W	Dataviz I	--	L2	--	--
10/14/2022	3	F	Intro to Analysis			Project Review*	--
10/17/2022	4	M	Descriptive Analysis	D3		--	Q3
10/19/2022	4	W	EDA	--		--	--
10/21/2022	4	F	Inference I	D3		Project Proposal*	--
10/24/2022	5	M	Inference II	D4		--	Q4

Course Confusion

- If something in lecture, a section workbook, or an assignment is unclear:
  - *ask in class*
  - *ask during section*
  - *post on Piazza*
  - *ask a classmate*
  - *come to office hours*

Please do not use Canvas messages.

CLASS CONDUCT

In all interactions in this class, you are expected to be respectful. This includes following the UC San Diego principles of community.

This class will be a welcoming, inclusive, and harassment-free experience for everyone, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), political beliefs/ leanings, or technology choices

At all times, you should be considered and respectful. Always refrain from demeaning, discriminatory, or harassing behavior and speech. Last of all, **take care of each other**.

If you have a concern, please speak with Prof, your TAs, or IAs. If you are uncomfortable doing so, the OPHD and/or CARE are wonderful resources on campus.

The (dreaded) waitlist

1. I know this matters to you and is a source of stress (and I hate that).
2. I have no control over the waitlist. If you have questions contact cogsadvising@ucsd.edu
    a. I know in other departments profs have control of this
    b. I quite literally do not have access to the system
3. A few people in each section typically get off the waitlist, but that number varies each quarter.
4. We have 417 enrolled with 175 on the waitlist at last look
5. We will likely admit up to about 430 or 440 total enrolled. So not everyone
6. Your wait list position is in your section. There are 7 sections. So if you're 6th on the waitlist of your section, you can expect there are up to 41 people in front of you
7. The waitlist settles after week 2.

What COGS 108 logistics questions do you have?

I’m excited to have
you all in COGS 108!