

If you walked in here feeling sick,
please turn around and walk back out.
Podcast recording of this class is turned on

In-person illness policy

Please do not attend any in-person activity (lecture/section/office hours) if you are feeling ill, especially if you are sneezing/coughing and have a fever. If you feel mildly ill but without sneezing/coughing, or if you have bad allergies, then you may come to in-person events while wearing a well-fitting mask.

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>

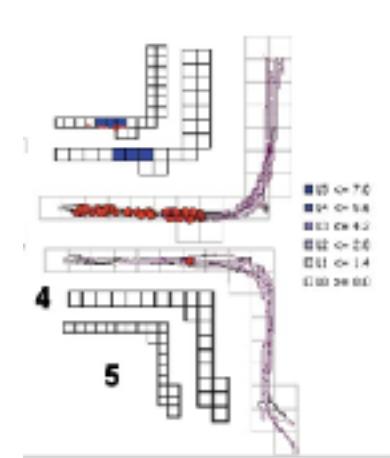
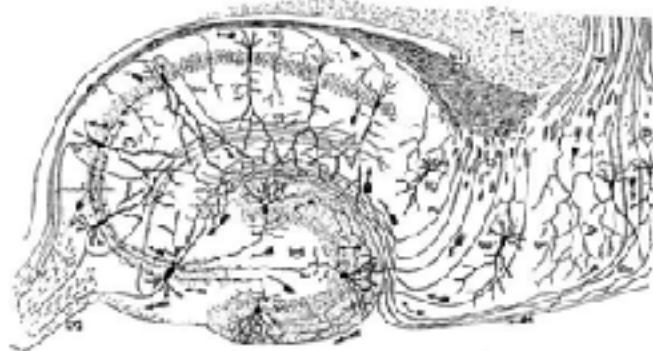
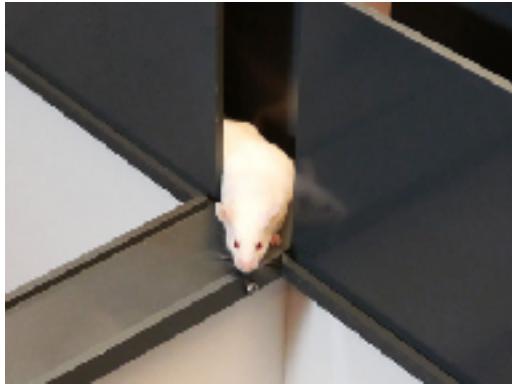
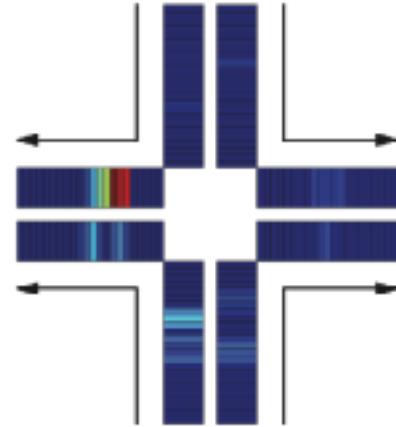
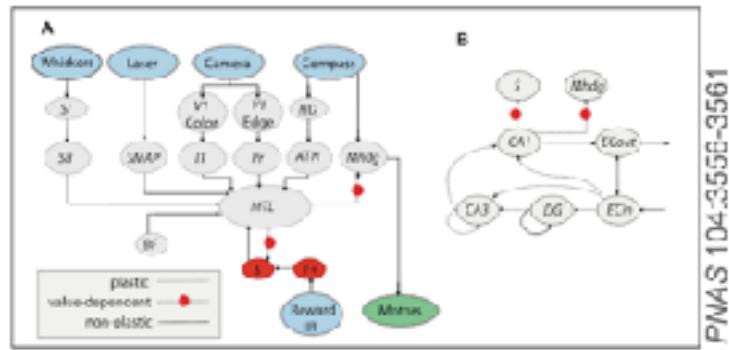
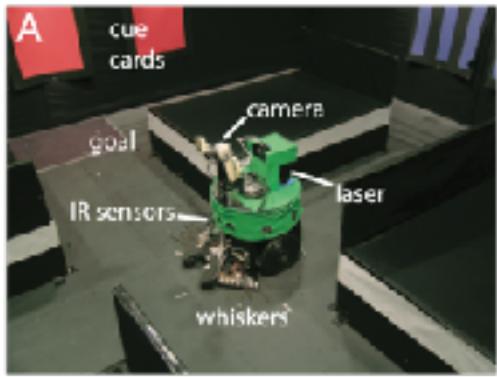


hello
my name is
Jason Fleischer



Wikimedia Commons
Sunlight on Colorado National
Monument.jpg
By Meelmouse







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

Science Advances 12 Dec 2018;
Vol. 4, no. 12, eaat6780
DOI: 10.1126/sciadv.aat6780

Article

Figures & Data

Info & Metrics

eLetters

PDF

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 biologically determines 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 4.2% increase in the median grades of the students and an improvement in attendance.

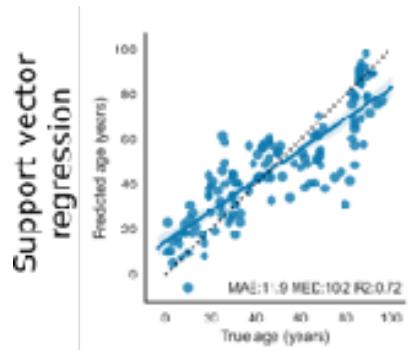
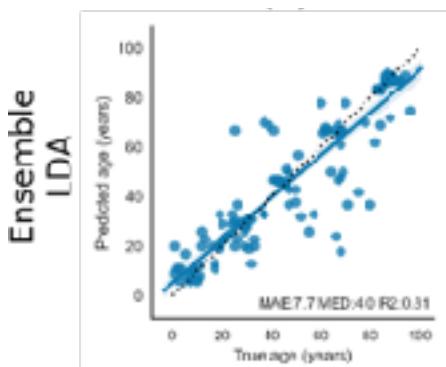
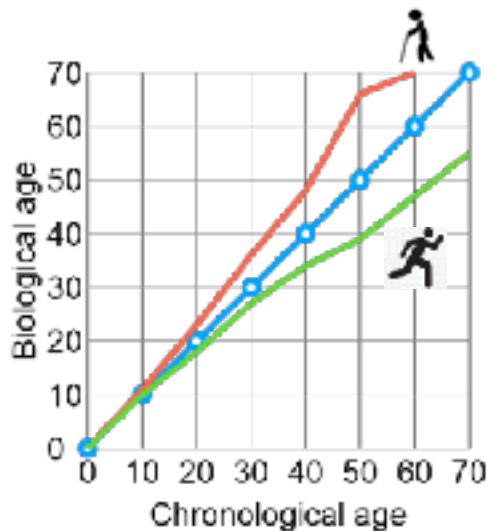
Cell Metabolism

Clinical and Translational Report



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

Michael J. Wilkinson,^{1,2} Emily N.C. Mancoglio,^{3,4} Adina Radulescu,¹ Hanru Li,¹ Ravindra Pakhouja,⁵ Amrit Singh,² Xintao Wang,¹ Jason G. Fleischner,⁶ Saxon Hwang,² Sudhirendra Ponda,^{2,4,7} and Pam R. Taub^{1,2}



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







WATCH PARTY

SATURDAY, SEPTEMBER 30 | 7:30PM KICKOFF
FAIRPLAY | 4026 30TH STREET



San Diego Wave FC @sandiegowavefc · 4h
picture this

- you buy a ticket for October 15
- you score a sweet jersey giveaway
- you cheer on the wave in their final 2023 regular season home game

pretty good day if we do say so



3

6

44

3,582



San Diego Wave FC @sandiegowavefc · 4h
and hey tickets start at 10 bucks!



[ticketmaster.com](#)

San Diego Wave FC vs. Racing Louisville FC Tickets ...

Buy San Diego Wave FC vs. Racing Louisville FC tickets at the Snapdragon Stadium in San Diego, C...

Waitlist information

I know this matters to you and is a source of stress and I hate it and wish it could be different!

But please do not email me asking questions about the waitlist... I generally get more than a dozen of these per day during the beginning of quarter. That's annoying and sucks up a lot of time if I reply. Time I can't spend helping students.

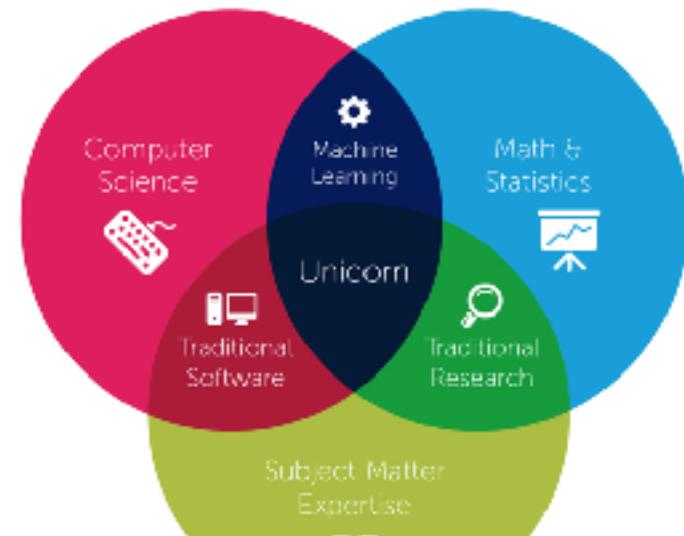
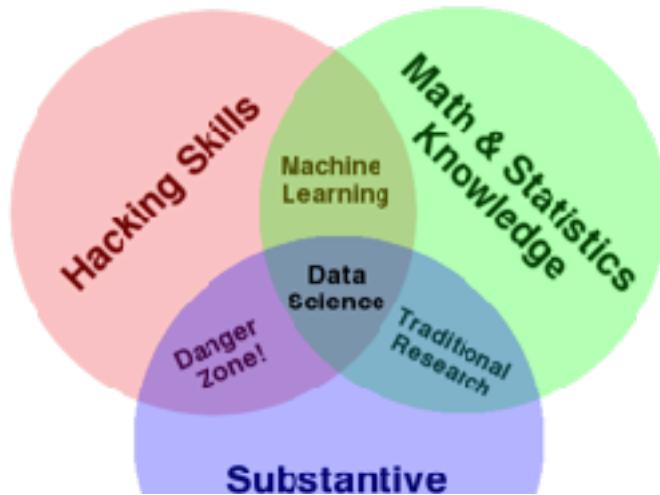
I will put everything that I know about the waitlist here:

- I have no direct control over the waitlist. If you have questions contact cogsadvising@ucsd.edu
- A few people in each section typically get off the waitlist, but that number varies each quarter.
- We have about 600 students enrolled in 2 lecture sections and ~11 lab sections.
- There are >200 on the waitlist at last look
- We will likely admit up to about 700 total enrolled. So not everyone. And the exact number admitted will depend on very complicated circumstances.
- We will prioritize students for admission in the following order:
 - COGS students, starting with those with senior status
 - Senior status and graduate students
 - After that its waitlist position
- Your wait list position is in your LAB section. There are 11 sections right now. So if you're 4th on the waitlist of your section, you can expect there are up to 44 people (11 sections x 4 people) in front of you once we admit all the priority students.
- People will be let off the waitlist in the 2nd week. Some people will be let off the waitlist each day that week. The waitlist closes at the end of week 2, if you're not off by then you will not be admitted.
- **If YOU REALLY WANT TO BE HERE KEEP UP WITH THE WORK WHILE ON THE WAITLIST!** There will be no chances to hand in things that are past deadline just because you didn't do them while on the waitlist

If you have a question not covered above, or if you think there's a good reason why you should be admitted regardless of the waiting list, then you should email cogsadvising@ucsd.edu not me.

“The scientific process of extracting value from data”

What is data science?



Thinking clearly with help
from data

Data scientists ask
interesting questions &
answer them with data

Why this course?

You need to know how to make good decisions based on empirical data because you're studying to be:

economist

journalist

small business owner

medical professional

accountant

engineer

Amazon delivery manager

data scientist

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 of the [50 Best Jobs in America](#) to help people find jobs they'll love. Each job stands out for its earning potential, job satisfaction, and job openings. Are you considering a new position? Check out the list to see what jobs made the list this year, and view open jobs at companies across the country.

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

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

Statistics

$$d = \text{distance}$$

$$\sigma_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}$$

$$\rho = \frac{\sigma_x \sigma_y}{\sigma_x \sigma_y}$$

$$\hat{y} = a + bx$$

$$P = \frac{n}{2}$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$$\mu = \frac{1}{n} \sum x_i$$

$$b = r \frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

$$\hat{p} = \frac{n}{2}$$

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

$$P(A \cap B) = P(A) \cdot P(B)$$

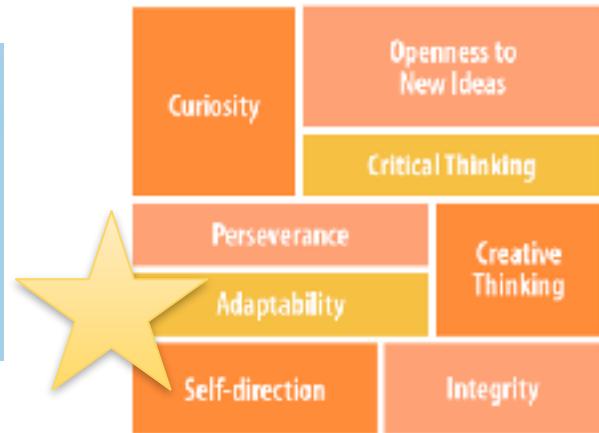
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\sigma = \sqrt{\frac{1}{n-2} \sum (x_i - \bar{x})^2}$$



HABITS OF MIND

We develop habits of mind such as...



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

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](https://www.codecademy.com) or [LearnPython.org](https://www.learnpython.org)
 - [Python Data Science Handbook](https://jakevdp.github.io/PythonDataScienceHandbook/)

Lets go through the syllabus

<https://github.com/COGS108/Overview/blob/master/COGS108-Syllabus.md>

Course links

GitHub	https://github.com/COGS108	lecture/section materials & final projects
datahub	https://datahub.ucsd.edu	assignment submission
CampusWire	https://campuswire.com/p/ G160521FD	questions, discussion, and regrade requests. Group join code: 9411
Canvas	https://canvas.ucsd.edu/courses/48733	grades, lecture videos
Anonymous Feedback	https://forms.gle/yAYq38iCXAfx8pp9	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 Monday of Week 1!

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

Weekly Lecture Quizzes:

- (9) weekly quizzes (first one due Monday 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

8 Discussion Lab exercises

Completed individually and graded partly manually (for effort and good thinking) and partly programmatically (for correctness).

- These are meant to get you practice programming around the topics covered in class.
- You will have to look some stuff up on your own. This is by design.
- Instructions must be followed perfectly to receive credit.
- You'll have the opportunity to practice in discussion section.

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

75% credit if submitted less than 3 days after deadline.

**5 LATE DAYS allowed per person without penalty
to be used for Discussion Labs + Assignments**

(4 + 1 practice) Assignments

Completed individually and graded almost completely programmatically.

- These are meant to get you practice programming around the topics covered in class.
- The first two are much simpler/shorter, the last two are harder/longer.
- You will have to look some stuff up on your own. This is by design.
- Instructions must be followed PERFECTLY to receive credit.
- You'll have the opportunity to practice in discussion section.

Assignments will be due on Wednesdays by 11:59 PM

75% credit if submitted less than 3 days after deadline.

**5 LATE DAYS allowed per person without penalty
to be used for Discussion Labs + Assignments**

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

DATA SCIENCE / MACHINE LEARNING PLATFORM

UC San Diego

Information Technology Services - Educational Technology Services Help Options -



UC San Diego Jupyterhub (Data Science) Platform

Before next Mon: 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 will be assigned!!

~~How to find a group:~~

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

[Questions](#)[Responses](#)[Settings](#)

Demographic information for research purposes



You are being invited to participate in a research study while you are attending COGS 108 Data Science in Practice. All students in the course are invited to participate in this study, which is being conducted by the course's instructor Dr. Jason Fleischer. **By filling out this optional section of the survey you are agreeing to participate in this study.**

The purpose of this research study is to investigate how different ways of assigning students to project groups affects student learning and satisfaction. The results of this study may help improve learning outcomes for future students in this course and others like it.

By filling out this section of the survey you agree to have the above data included in analyses that will appear in publication. Analysis of your data will be done by myself and members of my research lab. Any publications from this study will report only aggregated, de-identified data. Your personally identifiable information will never appear in public. We will avoid reporting data on demographic groups of < 3 students to avoid potentially revealing personal information through aggregate reporting. This data will be secured with access limited to a small team of researchers responsible for this study. The data will be deleted after 10 years. You may also contact us at any time to revoke your permission and have your data deleted.

Your participation in this study is completely voluntary and you can withdraw at any time. **Choosing not to participate or withdrawing will not affect your grade or relationships with the course instruction team.** If you do participate you are free to skip any question you do not wish to answer.

If you have questions about this project, have a research related problem, or wish to revoke your permission to use your data, you may contact Jason Fleischer at jfleischer@ucsd.edu. If you have any questions concerning your rights as a research subject, you may contact the UC San Diego Office of IRB Administration at irb@ucsd.edu or 858-246-4777.

By participating in this research study you are indicating that you are at least 18 years old, have read this consent form, and agree to participate in this research study.

Confusion is expected ITS OK!!!

If something in lecture, a section workbook, or an assignment is unclear:

- *ask in class*
- *ask during section*
- *post on Campuswire*
- *ask a classmate*
- *come to office hours*

Academic integrity

Don't cheat. Please review academic integrity policies here.

You will work together on projects. You should help one another learn in general. Assignments and discussion labs should be completed individually, although you may seek help from your fellow students. However you may not give answers to each other at any time. THERE IS NO ASKING QUESTIONS OF EACH OTHER ON QUIZZES AND EXAMS!!

Examples of good collaboration on assignments:

- Student posts non-working code to get help, others send a link to a good reference page in sklearn documentation, or point out the generic kind of mistake being made (e.g., you've messed up the order of operations). Nobody just writes the correct code for the student.
- Student posts a question about a theory or concept. If its not directly related to an assignment question you can choose to answer in full. However, it's generally more helpful for learning if you use the Socratic method: ask the student questions that lead them to find the answer themselves. Also doing this helps you cement your own knowledge of the subject!
- Student posts a question about an assignment problem. Others point out important principles that we have learned in class that can be used to solve it. They describe the important points, or mention important pitfalls to avoid. References to book pages, lecture slides, or lecture video times are helpful. Nobody posts the correct answer for the student.

For group projects, you will work together but every person in the group is expected to understand every aspect of the project. People may be asked to individually explain any aspect of the project and your grade may be reduced compared to the rest of the group if you are unable to do so. Projects may include ideas and code from other sources—but these other sources must be documented with clear attribution.

Know that a third of the class typically feels overwhelmed at the start of the quarter. That said, the average is quite high in this course typically (A-/B+). So, while we anticipate you all doing well in this course, if you are feeling lost or overwhelmed, that's ok! Should that occur, we recommend: (1) asking questions in class, (2) attending office hours and/or (3) asking for help on Campuswire.

Cheating and plagiarism have been and will be strongly penalized.

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 considerate 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.

What COGS 108 logistics
questions do you have?

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