

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](mailto:jfleischer@ucsd.edu)  
<https://jgfleischer.com>  
@jasongfleischer

Lectures : <https://github.com/COGS108/Lectures-Sp23>

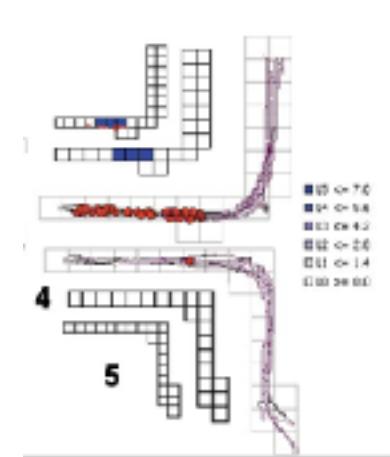
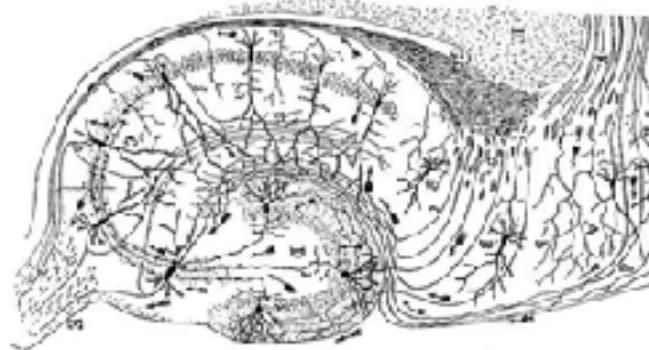
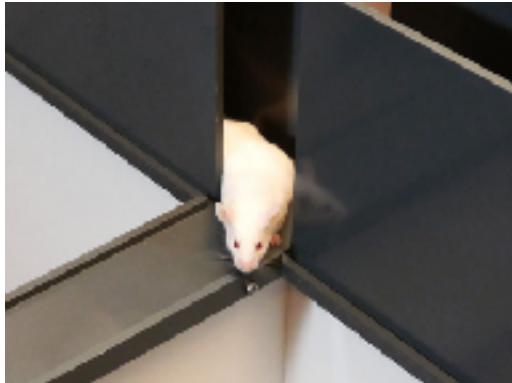
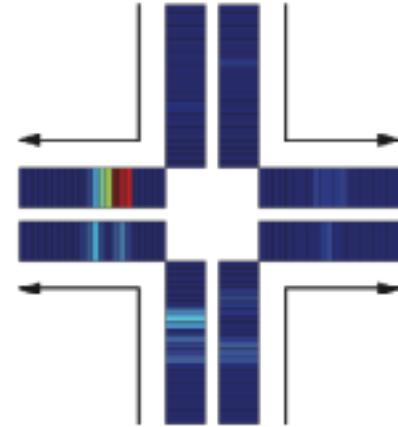
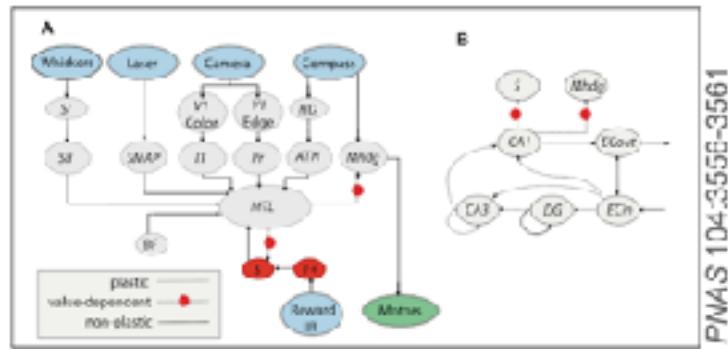
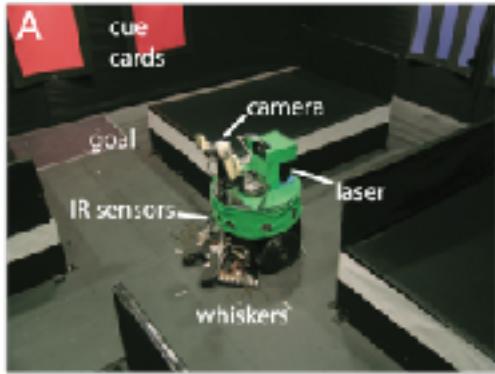






Wikimedia Commons  
Sunlight on Colorado National  
Monument.jpg  
By Meelmouse







# 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, eaar6580  
DOI: 10.1126/sciadv.aar6580

Article

Figures &amp; Data

Info &amp; 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.

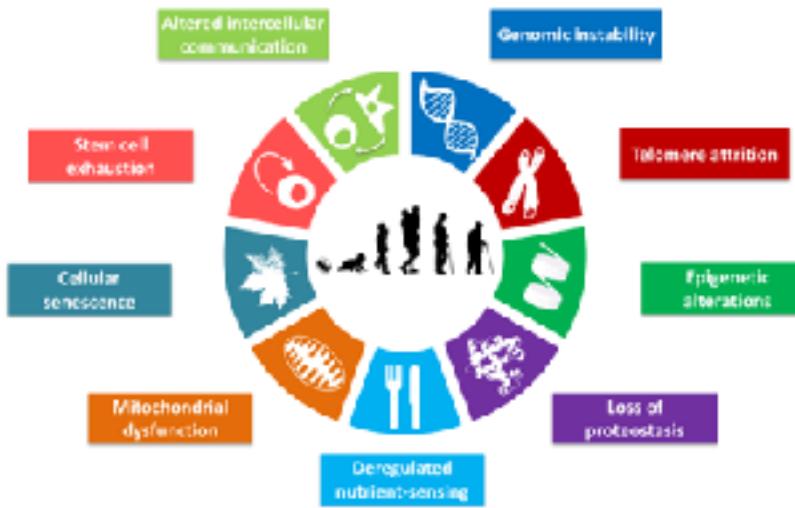
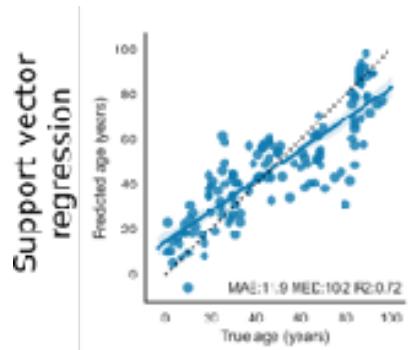
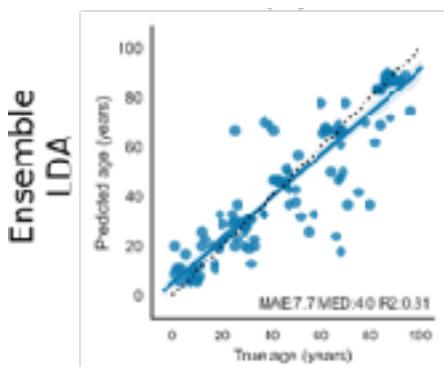
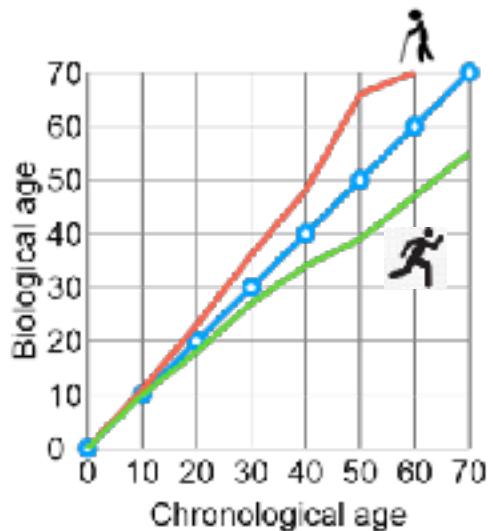
Cell Metabolism

# Clinical and Translational Report



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

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



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









**NEXT  
MATCH**

VS  
**FC TULSA**

APR. 8 • 2:00PM  
TOFFERO STADIUM  
ESPN+ FOX 5

**TOGETHER**

**STAY LOYAL**  
2023 ERL FINAL TWO TEAM

*Sycuan*  
ANADAMS STADIUM

# NEXT HOME GAME



7:00 PM | APRIL 19

PRESENTED BY  
FRANCIS GILLIS &  
**QUEEN CHARLOTTE**  
A SAGETTV & NETFLIX  
NETFLIX



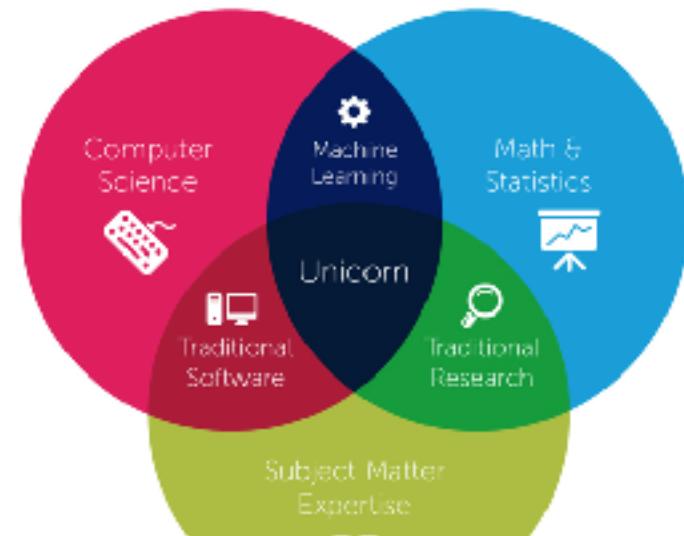
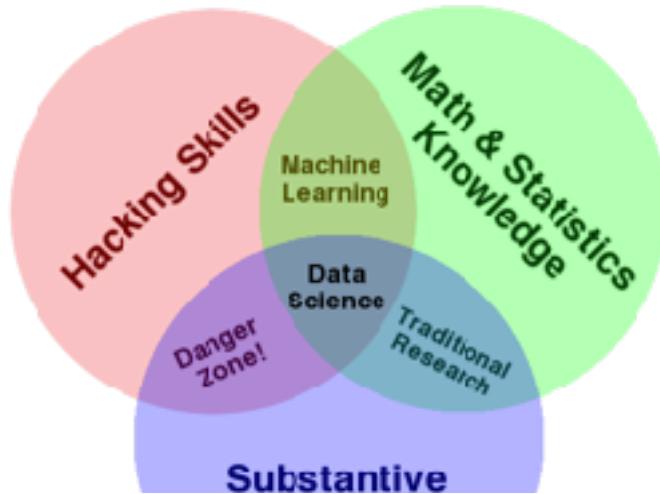


# The (dreaded) waitlist

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](mailto: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 in week 2.
8. If YOU REALLY WANT TO BE HERE KEEP UP WITH THE WORK!

“The scientific process of extracting value from data”

## What is data science?



Thinking clearly with help  
from data

# 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 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 <a href="#">Enterprise Architect</a>	\$144,997	4.1/5	14,021	[View]
#2 <a href="#">Full Stack Engineer</a>	\$101,794	4.3/5	11,252	[View]
#3 <a href="#">Data Scientist</a>	\$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

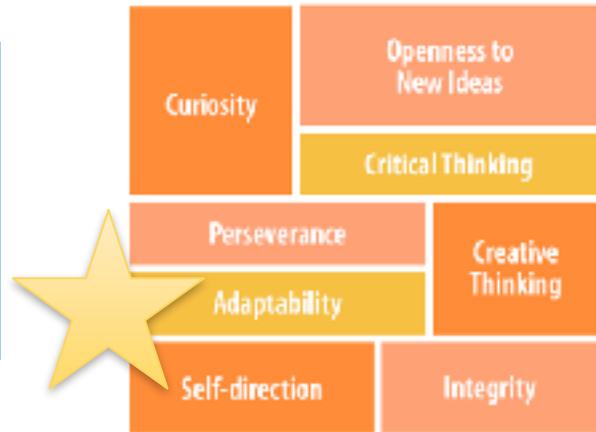
Data scientists ask  
interesting questions &  
answer them with data

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



## 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

# Scheduling & Staff

**Lecture:** MWF 10-10:50pm

**Discussion Sections:** M & W various times

**Office Hours:** BOOK on gcal!

TAs	IAs
Fuling Sun	Cindy Wang
Ruby Ying	Nathaniel Mackler
Heeket Mehta	Jinyi Zhao
Shanay Shah	

# 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/)

# Course links

GitHub	<a href="https://github.com/COGS108">https://github.com/COGS108</a>	lecture/section materials & final projects
datahub	<a href="https://datahub.ucsd.edu">https://datahub.ucsd.edu</a>	assignment submission
CampusWire	<a href="https://campuswire.com/p/G91B371ED">https://campuswire.com/p/ G91B371ED</a>	questions, discussion, and regrade requests. <b>Group join code: 0714</b>
Canvas	<a href="https://canvas.ucsd.edu/courses/45023">https://canvas.ucsd.edu/courses/ 45023</a>	grades, lecture videos
Anonymous Feedback	<a href="https://forms.gle/MpFPqBLvnRtT2ytJ6">https://forms.gle/MpFPqBLvnRtT2ytJ6</a>	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:

	<b>% of Total Grade</b>
(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 not required, BUT you will learn better!

- All lectures will be recorded (available end of day Canvas Media Gallery)

# 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

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

# COURSE SCHEDULE

	Week	Day	Topic	Section covers	Lab due	Assignment due	Lecture quiz due
Apr-3	1	M	Welcome!	--			
Apr-5	1	W	Python Review	--			
Apr-7	1	F	Version Control I				
Apr-10	2	M	Version Control II	D1			Q1
Apr-12	2	W	Data & Intuition	D1		Practice assignment, pre-course survey	
Apr-14	2	F	Data Wrangling (pandas)		D1	--	
Apr-17	3	M	Ethics	D2			Q2
Apr-19	3	W	Data Science ?s	D2		A1: Group Signup*	
Apr-21	3	F	Dataviz I		D2	--	

# Course Confusion

- 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*

# 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.

## 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.

## 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.

What COGS 108 logistics  
questions do you have?

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