Welcome to COGS 108!
Data Science in Practice

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

Lectures : https://github.com/COGS108/Lectures-Sp20

hello
my name is
Professor Ellis

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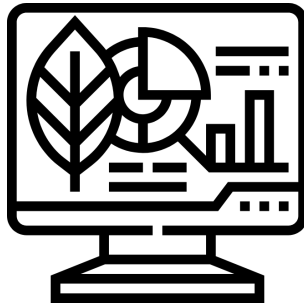
How we'll approach learning about *and doing* data science in COGS 108

Course links

GitHub	https://github.com/COGS108	course materials & final projects
DataHub	https://datahub.ucsd.edu	assignment completion & submission
Piazza	http://piazza.com/ucsd/spring2020/cogs108	questions, discussion, and regrade requests
Canvas	https://canvas.ucsd.edu/courses/12630	grades, video quizzes, video links
Anonymous Feedback	Submit via Google Form	if I ever offend you, use an example you hate, or to provide general feedback

COGS 108: General Plan

Week	Topic(s)
1	Data Science & Ethics
2	Version Control, Data, & Python
3	Data, Dataviz, & Data Wrangling
4	Intro to Analysis, Descriptive & Exploratory Analyses
5	Inference
6	Text Analysis
7	Machine Learning
8	Nonparametric Analysis & Geospatial Analysis
9	Dimensionality Reduction & DS Jobs
10	DS Communication & Future of Data Science

Programming Prerequisite

- CSE 7 - MATLAB
- CSE 8A or 11 - Java
- COGS 18 - Python

Bottom line: we will assume programming knowledge.

Python will be used for all assignments.

No programming experience (or you forget 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*

General grading:

	% of Total Grade
Content Engagement	20
(6) Assignments	45
(1) Project Proposal	8
(1) Project Survey	2
(1) Final Project	25

Letter Grade vs P/NP: Will be able to decide later in the quarter. If I were taking this course, I'd take letter grade for now and change to P/NP if needed later in the quarter.

Content Engagement: Video Quizzes

- 50+ video quizzes (due every Mon, Wed, & Fri - starting with Wed of Week 1)
- Goal: to help you keep on top of the material covered in pre-recorded lecture videos
- Why?: experience + student feedback
- How:
  - Taken on Canvas
  - Single Attempt
  - 1-3 videos per "lecture"
  - 3-15 questions/quiz
  - Timed : 1 min/question
  - 20 points 'free'
  - Each week's videos and quizzes will be posted by Friday night of the preceding week
  - Meant to test concepts from each lecture video

Intended Plan: Whenever COGS 108 class time is for you, you sit down, complete that day's videos (trying to remain focused for each) and then take corresponding quiz. Together, video + quizzes should take class time: < 50 min)

There <u>is</u> a corresponding video quiz for today's lecture material

It will be posted under Monday on Canvas.

quiz name: 01_01_welcome

It is not due until Wednesday.

A note on video naming:
- 01_ (the beginning) corresponds to the week
- _01_ (the second one) corresponds to the video in that week
- _welcome Is the name of the video

<u>Be sure you're opening the correct quiz.</u>

Each week's content will be available by the Friday of the previous week.
This week's (week 1) content will all be available by noon today.

(6) Assignments

Assignments 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 four 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

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

DATA SCIENCE / MACHINE LEARNING PLATFORM

UC San Diego

jupyterhub

Log In

*Registered Users
"username@ucsd.edu"*

UC San Diego Jupyterhub (Data Science) Platform

By the end of the week: log onto datahub & have a working installation of Jupyter on your computer

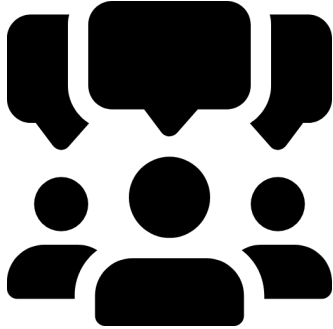| Week	Topic(s)	Due (Fri @ 11:59 PM)
1	Data Science & Ethics	--
2	Version Control, Data, & Python	--
3	Data, Dataviz, & Data Wrangling	A1 - git + python
4	Intro to Analysis, Descriptive & Exploratory Analyses	Project Proposal*
5	Inference	A2 - pandas
6	Text Analysis	A3 - Data Exploration
7	Machine Learning	A4 - Data Privacy
8	Nonparametric Analysis & Geospatial Analysis	A5 - Data Analysis
9	Dimensionality Reduction & DS Jobs	A6 - NLP
10	DS Communication & Future of Data Science	Final Project Survey

Final Project*: due Wed June 10th of finals week by 11:59 PM
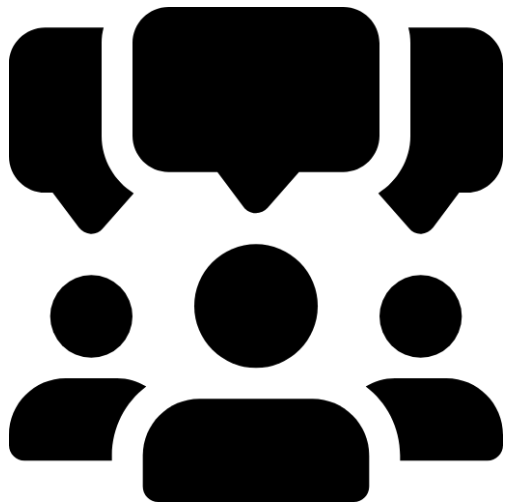
Discussion Section

- Goal:
  - help with technical aspects of the course
  - assignment help
- No section during week one
- Attend the one you registered for, if possible
- Attendance is neither mandatory nor incentifized

Office Hours

- Will be held on Zoom
- Discussion Section & Office Hour zoom links will be shared on Canvas

Prof Ellis will hold office hours week 1.
TA Office Hours start week 2.

Week	Topic(s)	Discussion Section	Due (Fri @ 11:59 PM)
1	Data Science & Ethics		--
2	Version Control, Data, & Python	intro + datahub + git (A1)	--
3	Data, Dataviz, & Data Wrangling	project groups/proposals	A1 - git + python
4	Intro to Analysis, Descriptive & Exploratory Analyses	pandas I (A2)	Project Proposal*
5	Inference	pandas II (A2)	A2 - pandas
6	Text Analysis	dataviz/seaborn (A3)	A3 - Data Exploration
7	Machine Learning	web pages + python (A4)	A4 - Data Privacy
8	Nonparametric Analysis & Geospatial Analysis	inference (A5)	A5 - Data Analysis
9	Dimensionality Reduction & DS Jobs	text data (A6)	A6 - NLP
10	DS Communication & Future of Data Science	advice	Final Project Survey

Final Project*: due Wed June 10th of finals week by 11:59 PM

Final Project: Spring 2020

Option 1: Group Final Project
- Completed throughout the quarter
- Groups of 4-5 people
- Carry out data science process on topic of your choosing

Option 2: Individual Final Project
- Completed in 48 hours during finals week
- Completed individually
- dataset/topic are chosen for you

You will have to decide (and be in a group) by the end of week 3. So...start thinking about this now, please!

(1) Project Proposal (8%)

These are **due Friday (11:59 PM) of Week 4**.

Option 1: These include a proposal of what question you are setting out to answer for your final project and lays out what data you would need to complete this project.

Option 2: This proposes what question you would ask and what data you would need on a topic provided to you.

(1) Final Report (25%) + (1) Survey (2%)

The <u>final project</u> is a Jupyter notebook completed (with your assigned group) carrying out a *(complicated) data science project* from start to finish.

The <u>survey</u> is completed individually*. It reports information about working with your group throughout the quarter (option 1) or your individual experience working on the project (option 2).

*Individuals who do not contribute equally are at risk of receiving a lower grade than other members of the group

Course Confusion

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

    - *ask during section*
    - *post on Piazza*
    - *ask a classmate*
    - *attend office hours*

In technical classes, **Piazza** is a particularly helpful resource

1. No duplicates.
2. Include Assignment & Question in Summary line.
3. *Public* posts are best.
4. Helping one another is encouraged.
5. No assignment code in public posts.
6. Regrade requests "regrades" tag as *private* post to "Instructors"
7. We're not robots.

CLASS CONDUCT

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

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. Ellis, your TAs, or IAs. If you are uncomfortable doing so, the <u>OPHD</u> and/or <u>CARE</u> 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
   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.
   a. Number of people who drop the class
   b. Number of staff members

4. The waitlist settles after week 2.

5. Our staff (cogsadvising@ucsd.edu) take care of this. They'll be able to help you more than I can.

A Note about this quarter...

1. I'm excited to have you all in COGS 108.
2. This quarter is going to be different.
3. I've built flexibility in to account for the remote nature of this quarter.
4. I'm here to support you.

Survey : http://bit.ly/cogs108_sp20

EC if completed by 11:59 PM **Friday**

COGS 108 Student Survey (Spring 2020)

This survey is used to help me get to know you a bit better! Thanks in advance for your participation!

If you complete before Friday (4/3) at 11:59 PM, there is an opportunity for a little bit of extra credit.

If any of these data are used/displayed in class, the data will be anonymized. Please answer as truthfully as possible. How you respond will NOT affect how you do in this class. Also, many questions are NOT required. Please do not answer anything that makes you uncomfortable.