

Cogs 9

Discussion Section

FA22 Week 1
Will McCarthy

Intros

Will McCarthy

Did Computer Science and Philosophy for UG

5th year PhD student in CogSci

I study cognitive mechanisms for *making things*, including perception, abstraction learning, and language

I also like to *make things*- art, music, food- and am using my time in California to surf and see cool nature

discord (willmccarthy) better than email

Come to office hours (Thurs 2pm on [zoom](#) pw: 185569)!



Discussion section logistics

Attendance: *optional*, but encouraged :)

Sections will **not** be recorded, but all materials will be available here:

<https://github.com/wpmccarthy/cogs9FA22>

Section philosophy: sections exist to make your learning experience easier and more thorough; different way of engaging with the material (i.e. not another lecture); concrete examples where possible; supplement the readings

Rough outline of every section: logistics, content, demos, touch base with group/ ask questions

Course logistics

Everything you will ever need (in theory) is on the [course website](#).

Quizzes, and exam are done individually (on gradescope).

Everything else is done in *groups*.

You will be assigned a group today, and that will remain your group for all assignments and the final.

Making things easy for yourself

1. Make a bookmarks folder, add all relevant course pages:

<https://kshannon.github.io/ucsd-cogs9/syllabus/>

<https://www.gradescope.com/courses/447181>

<https://github.com/wpmccarthy/cogs9FA22>

2. Make a cogs9 folder, subfolders for every week

3. Download all of the materials now (readings, reading guides)

4. Read through the syllabus

Assigning groups/ demo

[Note: there isn't any programming required for this course, however programming makes up a *large part* of data science and understanding a bit about how data is processed using code can help you to understand the limits and possibilities of data science. I will therefore strive to give you some concrete coding examples, but understanding them isn't required to do well in the course. If you want to learn how to actually *do* data science, take the cogs course: *Data Science In Practice*]

Making things easy for your group

1. Make a discord channel
2. Make a shared google drive folder or git repo
3. Decide when/ how you will meet (online? in-person? how regularly? I suggest revisiting this after you've completed the first assignment)

(Take 10 mins to introduce yourselves, and do these things now)

4. Read *Teamwork Expectations* in the Syllabus

Assignment 1

Prep for final project! (unlike assignments 2, 3, 4 which are standalone)

First attempt at: formulating a data science question, doing some background research, finding a dataset

More about documenting your efforts than having something polished

Ethical considerations > 50% of points

Think *very carefully* about this part

Even If you think everything “should be fine”, you need to justify why.

Tip: don't pick a project because you think ethical issues won't apply.
(Obviously don't pick something unethical.. but if you have something to discuss about ethics it's actually easier for you)

General tips for course readings

Read “Guide to Reading Academic Research Papers”

Use the reading guides on the syllabus!

Read *title*, *abstract*, and *year*

Skim first, deep read second

Ask yourself, why is this on the reading list?

Reading 1: 50 Years of Data Science, David Donoho

Discuss (5 mins)

As a group, come up with an answer to this question:

What was the purpose of this paper?

Provide a historical account of data science (in relation to statistics)?

Define data science?

Presenting a case for a new science of data?

Describe what data science could be?

Four major influences act on data analysis today* (Tukey) *1962

1. The formal theories of statistics
2. Accelerating developments in computers and display devices
3. The challenge, in many fields, of more and ever larger bodies of data
4. The emphasis on quantification in an ever wider variety of disciplines

Cleveland's 6 foci of activity (and allocations of effort)

Multidisciplinary investigations (25%)

Models and Methods for Data (20%)

Computing with Data (15%)

Pedagogy (15%)

Tool Evaluation (5%)

Theory (20%)

“An action plan to expand the technical areas of statistics focuses on the data analyst... . The value of technical work is judged by the extent to which it benefits the data analyst, either directly or indirectly.”

Generative vs. Predictive cultures

Generative: understand the giving rise to the data

Predictive: predict new data

Breiman says: stats should include both (not just generative)

Common Task Framework (“secret sauce” of predictive)

(a) A publicly available training dataset

(b) A set of enrolled competitors

(c) A scoring referee

“All the competitors share the *common task* of training a prediction rule which will receive a good score; hence the phase *common task framework*.”

Greater Data Science: 6 Divisions

Data Exploration and Preparation

Data Representation and Transformation

Computing with Data

Data Modeling

Data Visualization and Presentation

Science about Data Science

Similar to Cleveland's but more inclusive.

An exercise that may be useful: describe the differences between each of these.

