



Computational Structures in Data Science



UC Berkeley EECS
Adj. Ass. Prof.
Dr. Gerald Friedland

Lecture #1: Welcome to CS88!



January 19, 2018

<http://inst.eecs.berkeley.edu/~cs88>



Goals today

- Introduce you to
 - the field
 - the course
 - the team
 - Answer your que

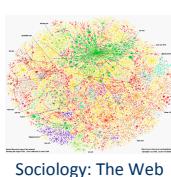
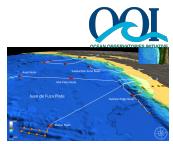
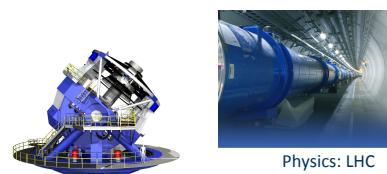


- **Big Ideas:**
 - Abstraction
 - Data Type



Data Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

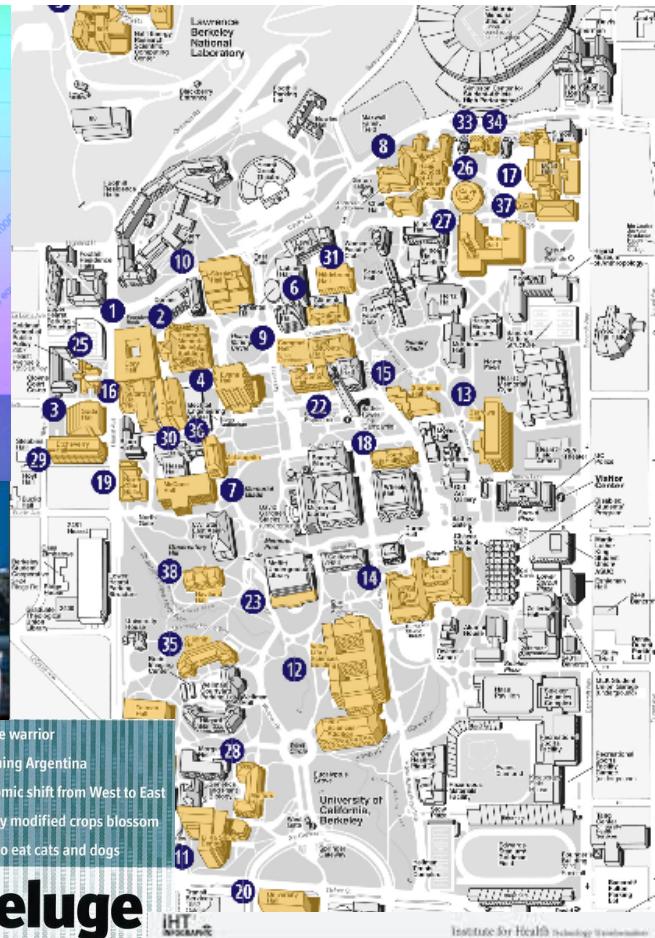


Demystifying Big Data in Government

A practical guide to transforming the business of government

The Economist

Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs



Berkeley
UNIVERSITY OF CALIFORNIA

Data Science growing organically everywhere

WIRED Spark: Open Source Superstar Rewrites Future of Big Data



Reconstructing the movies in your mind



AMP Lab
Ion Stoica, CS
Michael Franklin, CS

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Adam Arkin,
Bioengineering



Fernando Perez,
Brain Imaging Center
iPython tools and community



Charles Marshall
Rosie Gillespie
Integrative Biology
Digitized Museum

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT





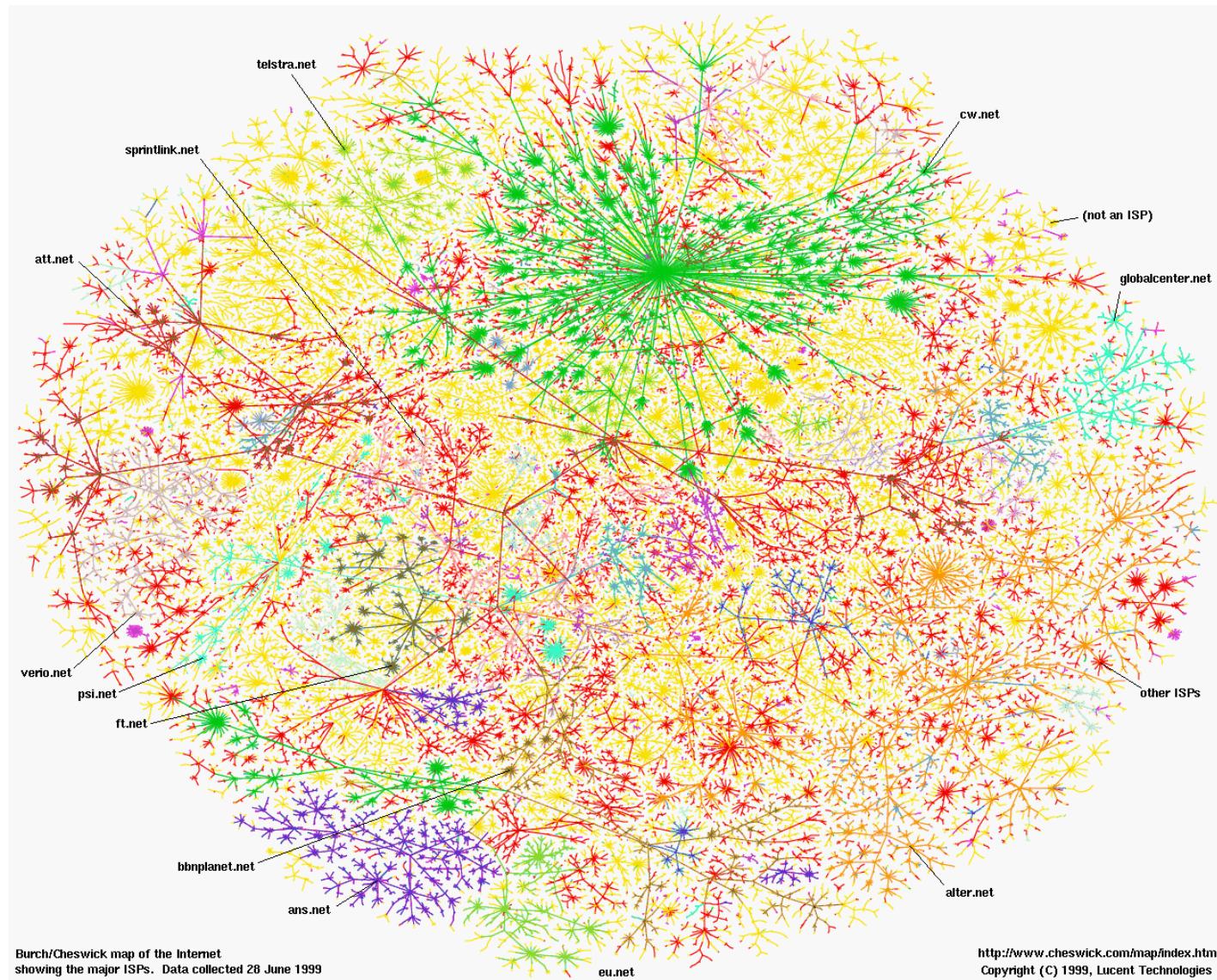
Data Science

In the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average size of a graduate class of data science students is 23 students. With approximately only 110 universities offering data science studies, the growing market will continue to pressure the supply in the US.

The screenshot shows the homepage of datanami. At the top is the logo "datanami" with three teal sunburst icons above it, followed by the text "• BIG DATA • BIG ANALYTICS • BIG INSIGHTS •". Below the logo is a navigation bar with links for Home, About, Whitepapers, Events, and Subscribe. A secondary navigation bar below features links for HOME, FEATURES ▾, SECTORS ▾, APPLICATIONS ▾, and TECHNOLOGIES. At the bottom left is the HPCw logo. To the right, the date "January 22, 2016" and the title "Data Scientists: The Myth and the Reality" by Seamus Breslin are displayed.

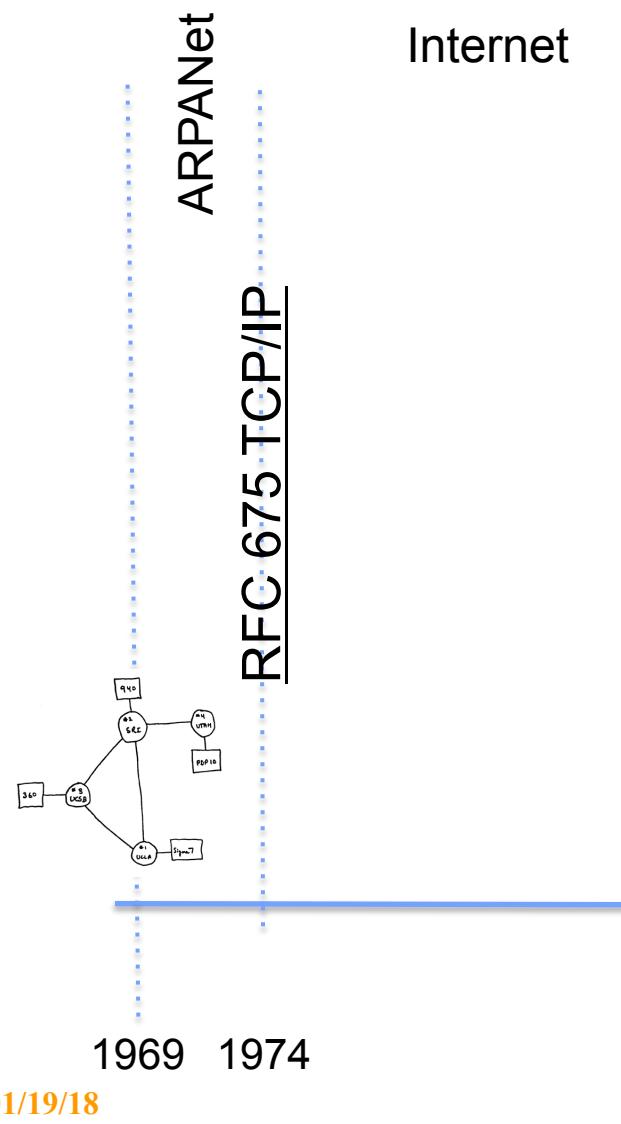


Greatest Artifact of Human Civilization ...





A Connected



3.0 B 11/15

3,293,151,639

Internet Users in the world

g

2,652,887,737

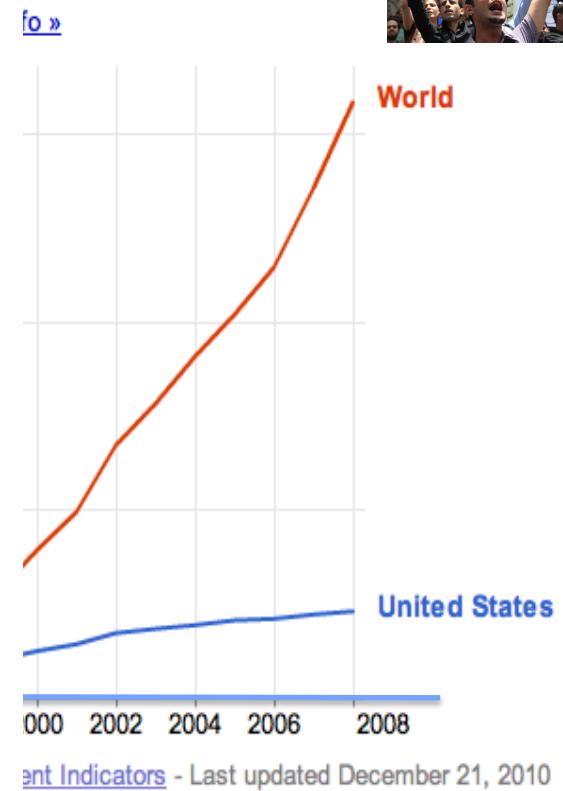
Google searches **today**



5,835,884,253

Videos viewed **today**
on YouTube

01/19/18



2010

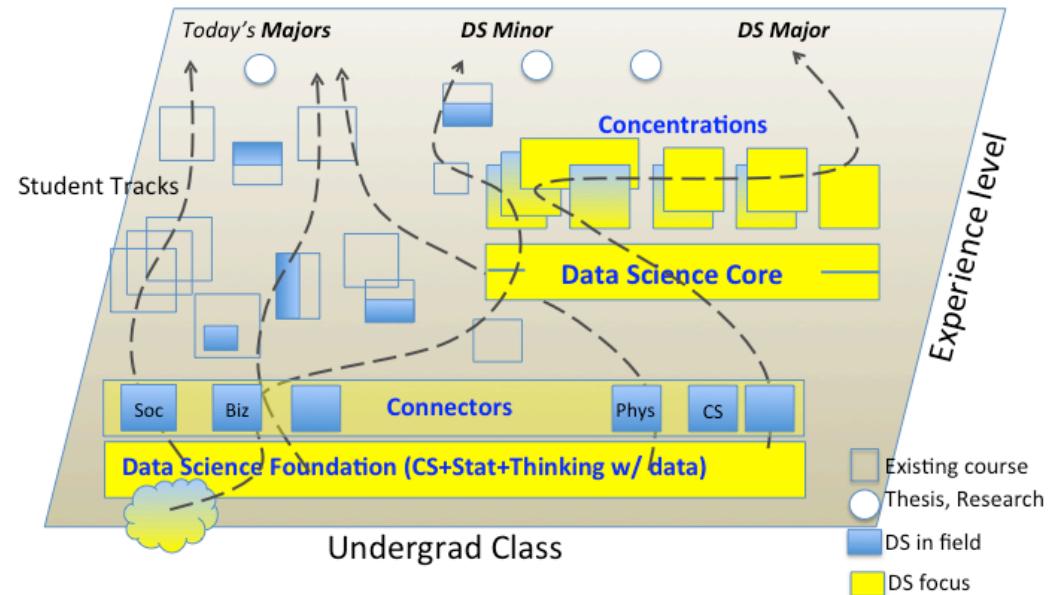
6

UCB CS88 Sp18 L1



Data 8 – Foundations of Data Science

- Computational Thinking + Inferential Thinking in the context of working with real world data
- Introduce you to several computational concepts in a simple data-centered setting
 - Authoring computational documents
 - Tables
 - Within Python3 and “SciPy”



CS88 – Computational Structures in Data Science



- **Deeper understanding of the computing concepts introduced in c8**
 - Hands-on experience => Foundational Concept
 - How would you create what you use in c8 ?
- **Extend your understanding of the structure of computation**
 - What is involved in interpreting the code you write ?
 - Deeper CS Concepts: Recursion, Objects, Classes, Higher-order Functions, Declarative programming, ...
 - Managing complexity in creating larger software systems through composition
- **Create complete (and fun) applications**
- **In a data-centric approach**



Pathways

c8

c8 cs88

c8 cs88 cs47a cs61b

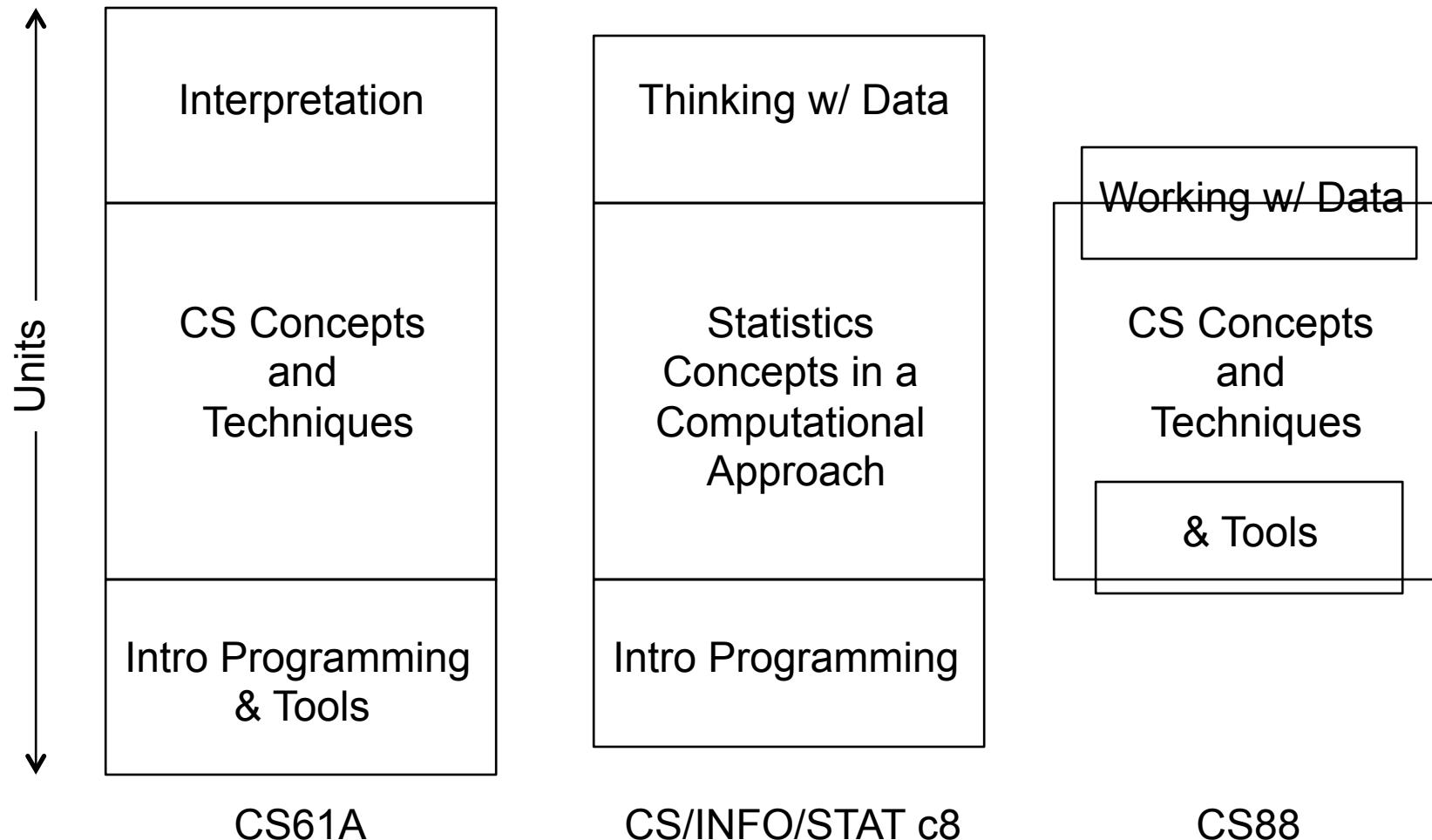
*** CS major

c8 cs61a

cs61a



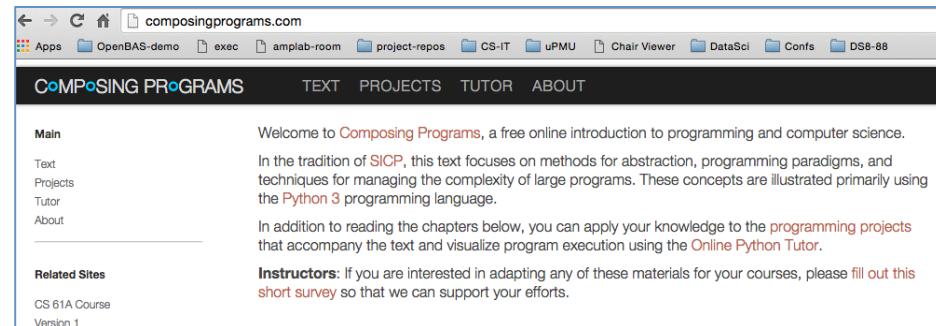
How does CS88 relate to CS61A ?





Course Structure

- **1 Lecture + 1 Lab/Discussion on Wednesday (!!!)**
- **Lecture introduces concepts (quickly)**
- **Lab provides concrete detail hands-on**
- **Homework (10) cements your understanding**
 - Out Monday, Due Sunday
- **Projects (3) put your understanding to work in building complete applications**
 - Maps
 - Hangman
 - Open Projects!



- **Readings:** <http://composingprograms.com>
 - Same as cs61a



CS88 Team



Prof. Gerald Friedland
fractor@eecs.berkeley.edu



Meghna Dasgupta
Meghna.dasgupta@berkeley.edu



Andrew Tan

Email: andrewtan@berkeley.edu



Jessica Gao

Email: gaojessicaping@berkeley.edu



Jobel Vecino

Email: jkpvecino@berkeley.edu



CS88 Team - me

- **Dr. Gerald Friedland (fractor@berkeley.edu)**
 - 424 Saturdai Daj Hall (CITRIS)
 - <http://www.gerald-friedland.org>
 - Office hours: Mo 1-2 @ 424 SDH
 - Before/after class



Berkeley
UNIVERSITY OF CALIFORNIA



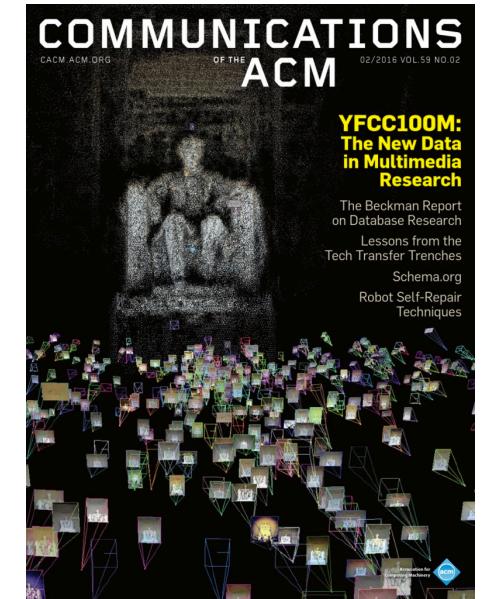
- **Adjunct Assistant Professor, EECS UC Berkeley**
- **Principal Data Scientist, Lawrence Livermore National Labs**



CS88 Team - me

Projects you might want to check out:

- <http://mmcommons.org>
 - Work with 100M images, 1M videos in your own Amazon instance.
- <http://www.teachingprivacy.org>
 - Creating teaching materials informing about data over sharing.





Course Culture

- Learning
- Community
- Respect
- Collaboration
- Peer Instruction





Piazza for {ask,answer}ing questions

PIAZZA CS 10 Questions Statistics 35 | Search or ask a question... Add Question/Note | Dan Garcia Piazza Help

Popular tags: #instructor-question #admin #logistics #welcome

QUESTION FEED FILTERS

▼ This week

When are TA / professor office hours? Sun
When can I meet up with a GSI or professor to get help with the course material? #admin
#instructor-question #admin

▼ Last week

So, I'm here... now how exactly does Pia Mon
(No question details) Sun
#logistics #welcome

question. 3 Views, 1 Follows Actions

When are TA / professor office hours?
When can I meet up with a GSI or professor to get help with the course material? #admin
Last updated by Luke Segars 2 days ago

Good Question!

Instructors' response.

We haven't established our office hours yet, but we'll make that information available as soon as possible. Check back here for an update by the second week of classes.
Last updated by Luke Segars 2 days ago

Actions Good Answer! Ask a Followup »

Start off a Students' Response

followup discussions.

Still Confused? Ask New Followup

AVERAGE RESPONSE TIME SPECIAL MENTIONS USERS ONLINE THIS WEEK

N/A Luke Segars answered When are TA / ... in 1.1 hr. 2 days ago 3
Online Now: 1

About Piazza Privacy Policy Copyright Policy Terms of Use Report a Bug!



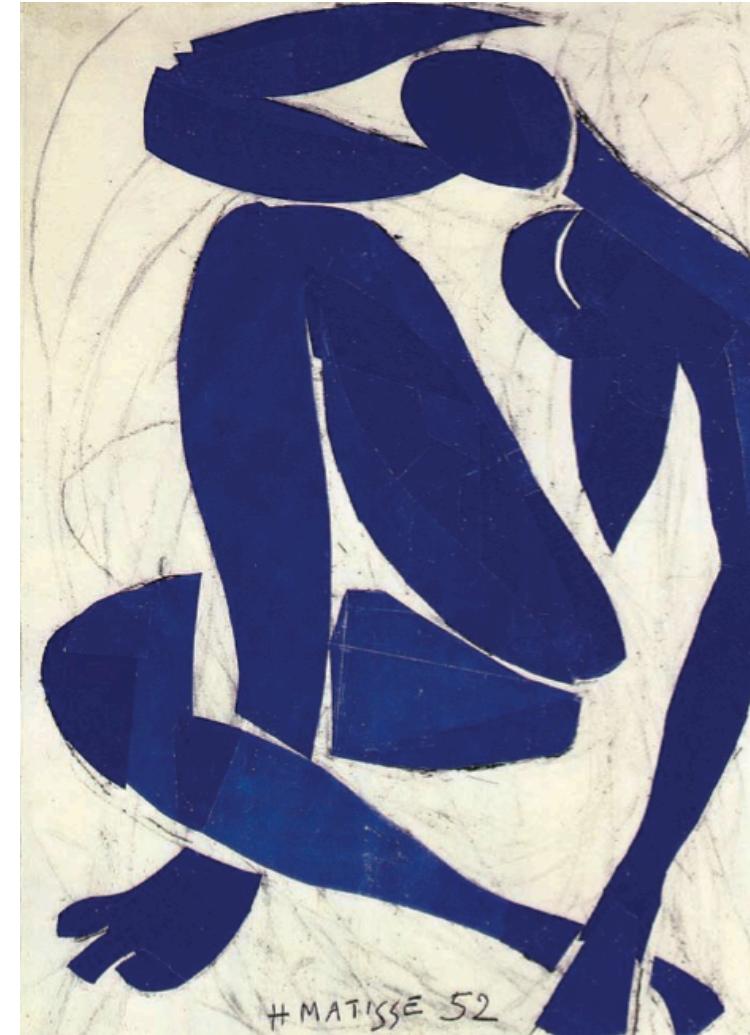
Pro-student Grading Policies

- EPA
 - Rewards good behavior
 - Effort
 - » E.g., Office hours, doing every single lab, hw, reading Piazza pages
 - Participation
 - » E.g., Raising hand in lec or discussion, asking questions on Piazza
 - Altruism
 - » E.g., helping other students in lab, answering questions on Piazza
- You have 2 “Slip Days”
 - You use them to extend due date, 1 slip day for 1 day extension
 - You can use them one at a time or all at once or in any combination
 - They follow you around when you pair up (you are counted individually)
 - » E.g., A has 2, B has 0. Project is late by 1 day. A uses 1, B is 1 day late



Abstraction

- **Detail removal**
 - “The act or process of leaving out of consideration one or more properties of a complex object so as to attend to others.”
- **Generalization**
 - “The process of formulating general concepts by abstracting common properties of instances”



Henri Matisse "Naked Blue IV"



Experiment

Standard Time Zones of the World





Where are you from?

Possible Answers:

- China
- California
- The Bay Area
- San Mateo
- 1947 Center Street,
Berkeley, CA
- 37.8693° N, 122.2696° W



All correct but different levels of abstraction!



Abstraction gone wrong!



I Can Stalk U

Raising awareness about inadvertent information sharing

Home How Why About Us Contact Us

What are people *really* saying in their tweets?

 [denisluque](#): I am currently nearby <http://maps.google.com/?q=-23.6193333333,-46.5506666667>
1 minute ago • [Map Location](#) • [View Tweet](#) • [View Picture](#) • [Reply to denisluque](#)

 [nikosofficiel](#): I am currently nearby <http://maps.google.com/?q=48.8699833333,2.3282833333>
5 minutes ago • [Map Location](#) • [View Tweet](#) • [View Picture](#) • [Reply to nikosofficiel](#)

 [dilmanarede](#): I am currently nearby <http://maps.google.com/?q=-15.7878333333,-47.8291666667>
7 minutes ago • [Map Location](#) • [View Tweet](#) • [View Picture](#) • [Reply to dilmanarede](#)

 [downtownvan](#): I am currently nearby <http://maps.google.com/?q=49.2833333333,-123.119833333>
10 minutes ago • [Map Location](#) • [View Tweet](#) • [View Picture](#) • [Reply to downtownvan](#)

 [MommaGooseBC](#): I am currently nearby 15745 Weaver Lake Rd
Maple Grove MN

Links

- Mayhemic Labs
- PaulDotCom
- SANS ISC
- Electronic Frontier Foundation
- Center for Democracy & Technology

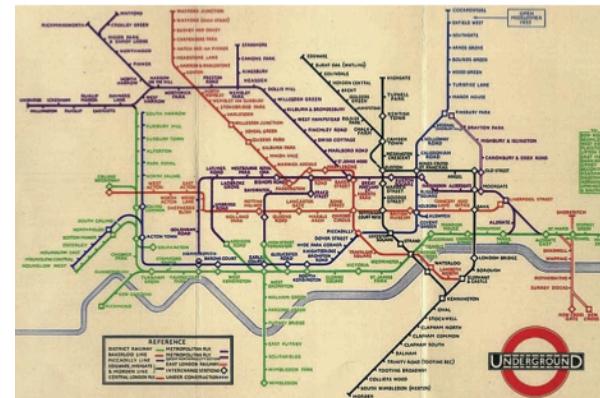
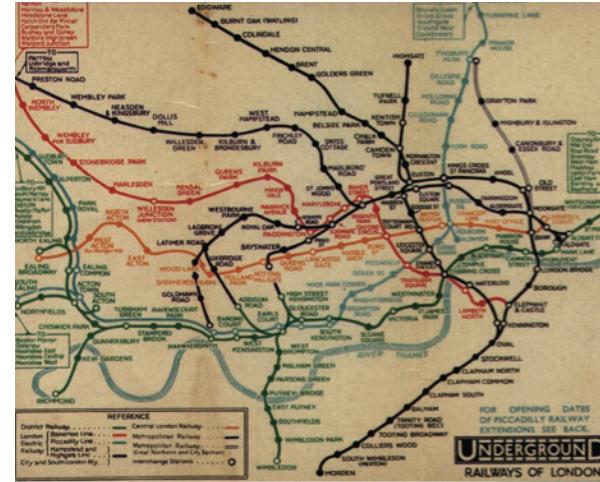
How did you find me?

Did you know that a lot of smart phones encode the location of where pictures are taken? Anyone who has a copy can access this information.



Detail Removal (in Data Science)

- You'll want to look at only the interesting data, leave out the details, zoom in/out...
- Abstraction is the idea that you focus on the essence, the cleanest way to map the messy real world to one you can build
- Experts are often brought in to know what to remove and what to keep!



The London Underground 1928 Map & the 1933 map by Harry Beck.



The Power of Abstraction, Everywhere!

- **Examples:**
 - Functions (e.g., $\sin x$)
 - Hiring contractors
 - Application Programming Interfaces (APIs)
 - Technology (e.g., cars)
- **Amazing things are built when these layer**
 - And the abstraction layers are getting deeper by the day!

*We only need to worry about the interface, or specification, or contract
NOT how (or by whom) it's built*

Above the abstraction line

Abstraction Barrier (Interface)
(the interface, or specification, or contract)

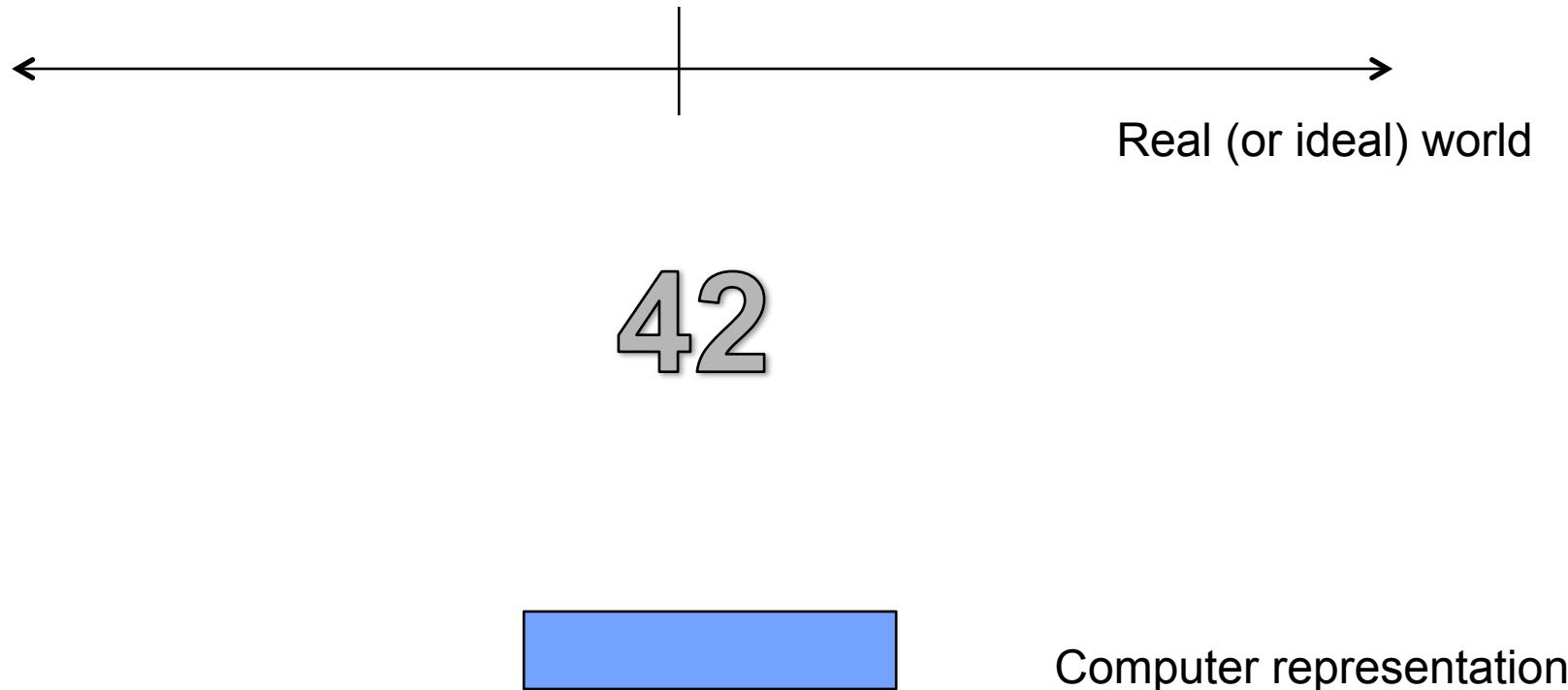
Below the abstraction line

This is where / how / when / by whom it is actually built, which is done according to the interface, specification, or contract.



Abstraction in CS: Data Type

- What's this?





Data Types and Operations

- **Set of elements**
 - with some internal representation
 - E.g. Integers, Floats, Booleans, Strings, ...
- **Set of operations on elements of the type**
 - e.g. +, *, -, /, %, //, **
 - ==, <, >, <=, >=
- **Properties**
 - Commutative, Associative, ... , Closure (???)
- **Expressions are valid well-defined sets of operations on elements that produce a value of a type**



Questions

- What's the difference between '==' and '=' ?



Lab and HW this week

- Lab will get you to where you have a *program development environment*
 - Even on your computer
- HW will give practice and explain subtleties of types, operators, and expressions
 - In a program development environment