



Computational Structures in Data Science

Lecture #1: Welcome to CS88!

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

data word cloud

August 26, 2016 <http://inst.eecs.berkeley.edu/~cs88>

Goals today

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

BIG DATA word cloud

data science word cloud

8/26/16 UCB CS88 Fa16 L1 2

Data Science

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



Berkeley

UCB CS88 Sp16 L1 3

Data Science

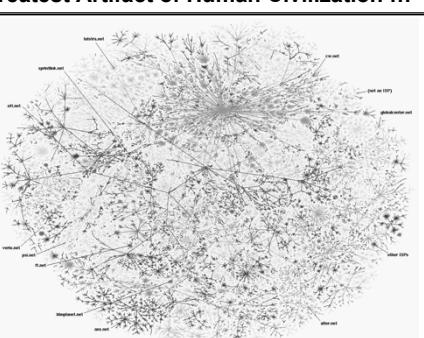
In the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average size of a graduate class of data science students is 23 students. With approximately only 110 universities offering data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

datanami
BIG DATA • BIG ANALYTICS • BIG INSIGHTS

HPC January 22, 2016 Data Scientists: The Myth and the Reality

8/26/16 UCB CS88 Sp16 L1 4

Greatest Artifact of Human Civilization ...



Internet ARPA/Net RFC 675-TCP/IP 1969 1974 3.0 B 11/15 3,293,151,639 Internet Users in the world 2.0 B 1/26/11 2,652,887,737 Google searches today 5,835,884,253 Videos viewed today on YouTube

8/26/16 UCB CS88 Sp16 L1 5

A Connected World

Internet ARPA/Net RFC 675-TCP/IP 1969 1974 3.0 B 11/15 3,293,151,639 Internet Users in the world 2.0 B 1/26/11 2,652,887,737 Google searches today 5,835,884,253 Videos viewed today on YouTube

8/26/16 00 2002 2004 2006 2008 2010 IT Indicators - Last updated December 21, 2010

World United States

1

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”

8/26/16 UCB CS88 Sp16 L1 7

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

1/25/16 UCB CS88 Sp16 L1 8

Pathways

8/26/16 UCB CS88 Sp16 L1 9

How does CS88 relate to CS61A ?

8/26/16 UCB CS88 Sp16 L1 10

Course Structure

- 1 Lecture + 1 Lab/Discussion on Monday (!!)
- 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

8/26/16 UCB CS88 Sp16 L1 11

CS88 Team - uGSIs

Dr. Gerald Friedland
fractor@eecs.berkeley.edu

Gunjan Baid
cunjan_baid@berkeley.edu

Lab Assistants (hopefully):

Anthony Xian, Rana Zee Maneri, Dashiell Brennan Stander, Pransu Dash, Niharika Jain, David Sang-chul Nahm, Minsu Kim, Caleb Casimir Chuck, Daniel Bernard Ricciardelli, Rena Chen, Kenneth Kao, Andrew Tan, Peter Yuan, Arman Madani, Calvin Dong, Erik Sanders Cheng

8/26/16 UCB CS88 Sp16 L1 12

CS88 Team - me

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



Berkeley
UNIVERSITY OF CALIFORNIA



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

8/26/16

UCB CS88 Sp16 L1

13

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.



1/25/16

UCB CS88 Sp16 L1

14

Course Culture

- Learning
- Community
- Respect
- Collaboration
- Peer Instruction



1/25/16

UCB CS88 Sp16 L1

15

Piazza for {ask,answer}ing questions

A screenshot of the Piazza platform. At the top, there's a navigation bar with 'piazza' and various tabs like 'Q & A', 'Questions', 'Answers', etc. Below the navigation, there's a search bar and a 'New Question/Note' button. The main area shows a question from a user asking about TA/professor office hours. Other users have responded with links to external resources and their own questions. At the bottom, there are statistics: 'AVERAGE RESPONSE TIME N/A', 'SPECIAL MENTIONS Luke Segars answered when are TA... in 1.1 sec, 2 days ago', and 'USERS ONLINE THIS WEEK 3 Online Now 1'.

8/26/16

UCB CS88 Fa16 L1

16

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

UCB CS88 Fa16 L1

17

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"

8/26/16

UCB CS88 Fa16 L1

18

Experiment



8/26/16

UCB CS88 Fa16 L1

19

Where are you from?

Possible Answers:

- China
- California
- The Bay Area
- San Mateo
- 1947 Center Street, Berkeley, CA
- $37.8693^\circ \text{N}, 122.2696^\circ \text{W}$



All correct but different levels of abstraction!

8/26/16

UCB CS88 Fa16 L1

20

Abstraction gone wrong!



I Can Stalk U

Raising awareness about inadvertent information sharing

What are people *really* saying in their tweets?

deniseguest: I am currently nearby http://maps.google.com
[Profile] 1 minute ago · [View Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to deniseguest](#)

nikosofficial: I am currently nearby http://maps.google.com
[Profile] 5 minutes ago · [View Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to nikosofficial](#)

dimmerude: I am currently nearby http://maps.google.com
[Profile] 7 minutes ago · [View Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to dimmerude](#)

downtowner: I am currently nearby http://maps.google.com
[Profile] 10 minutes ago · [View Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to downtowner](#)

MommaGosseBC: I am currently nearby 1574 Weaver Lake Rd
Maple Grove MN

Links

- Mayhem 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.

[Home](#) [How](#) [Why](#) [About Us](#) [Contact Us](#)

8/26/16

UCB CS88 Fa16 L1

21

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.

8/26/16

UCB CS88 Fa16 L1

22

The Power of Abstraction, Everywhere!



Examples:

- Functions (e.g., $\sin x$)
- Hiring contractors
- Application Programming Interfaces (APIs)
- Technology (e.g., cars)

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.

8/26/16

UCB CS88 Fa16 L1

23

Abstraction in CS: Data Type

• What's this?



42



Computer representation

8/26/16

UCB CS88 Fa16 L1

24

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

8/26/16

UCB CS88 Fa16 L1

25

Questions



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

8/26/16

UCB CS88 Fa16 L1

26

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

8/26/16

UCB CS88 Fa16 L1

27