Lecture #1:
Welcome to CS88!
Goals today

• Introduce you to
  – the field
  – the course
  – the team

• Answer your questions

• Big Ideas:
  – Abstraction
  – Data Type
Data Science

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

Data Science growing organically everywhere

W I R E D
Spark: Open Source Superstar Rewrites Future of Big Data

AMP Lab
Ion Stoica, CS
Michael Franklin, CS

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web

The data deluge
AND HOW TO HANDLE IT: A 54-PAGE SPECIAL REPORT

Analytics in Healthcare

Emmanuel Saez, Economics

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

Astronomy: LSST
Physics: LHC
Oceanography: OOI

Biology: Sequencing
Economics: POS terminals

Sociology: The Web
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.
Greatest Artifact of Human Civilization …
A Connected World

1969

ARPANet

RFC 675 TCP/IP

1974

Internet

2.0 B 1/26/11

3,293,151,639

Internet Users in the world

2,652,887,737

Google searches today

5,835,884,253

Videos viewed today on YouTube

3.0 B 11/15

2010
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
How does CS88 relate to CS61A?

<table>
<thead>
<tr>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS61A</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>CS Concepts and Techniques</td>
</tr>
<tr>
<td>Intro Programming &amp; Tools</td>
</tr>
</tbody>
</table>

| Thinking w/ Data |
| Statistics Concepts in a Computational Approach |
| Intro Programming |

| Working w/ Data |
| CS Concepts and Techniques |
| & Tools |

01/19/18

UCB CS88 Sp18 L1
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

- Adjunct Assistant Professor, EECS UC Berkeley
- Principal Data Scientist, Lawrence Livermore National Labs
Projects you might want to check out:

- [http://mmcommons.org](http://mmcommons.org)
  - Work with 100M images, 1M videos in your own Amazon instance.

- [http://www.teachingprivacy.org](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
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”
WHERE ARE YOU FROM?
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!
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?

Real (or ideal) world

42

Computer representation
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