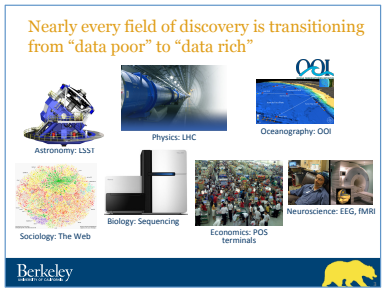
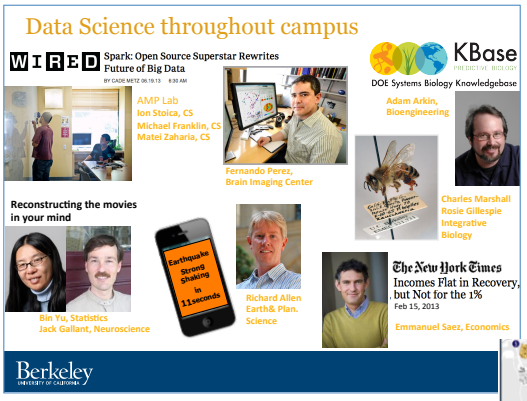
# Data Science in the 21st Century University

David Culler
University of California, Berkeley

Berkeley
UNIVERSITY OF CALIFORNIA

## Data Science Transforming Research



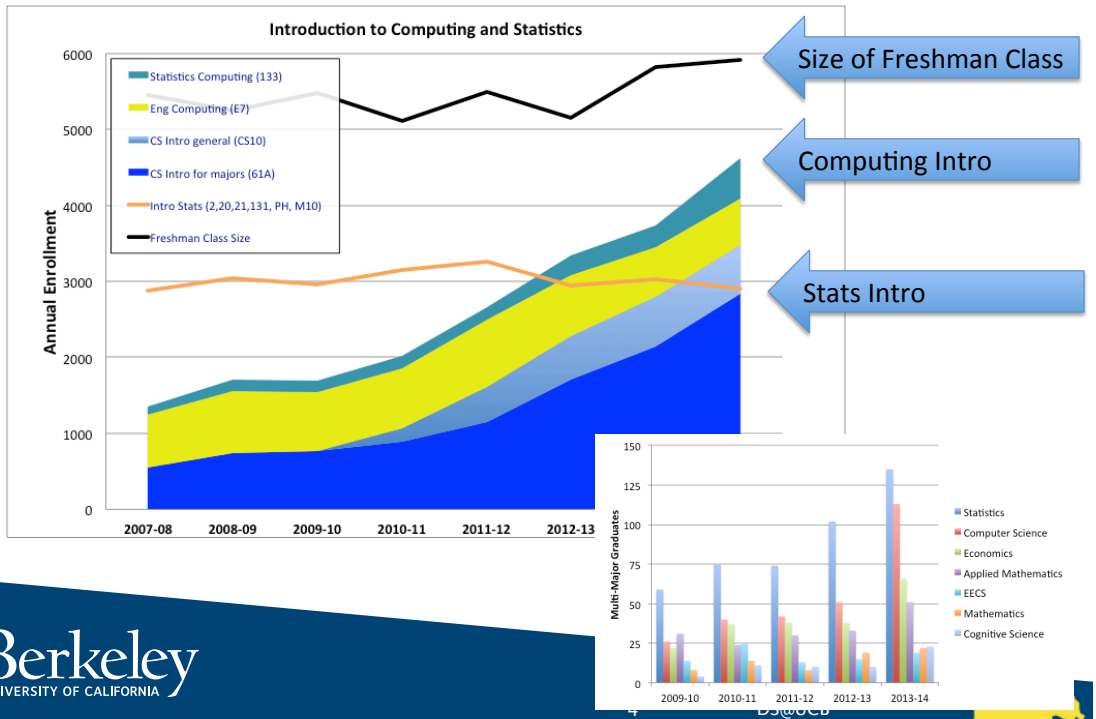
Berkeley
UNIVERSITY OF CALIFORNIA

# What is it?

- Principled use of computational methods and statistical techniques to draw robust insights and conclusions from complex, multi-dimensional, typically multi-sourced, often content-rich data
  - often in natural, evolving, time-sensitive, incompletely controlled, settings
- Development of theory and systems foundations of those processes,
- Informed by scientific, sociological, economic, or political context of questions, sources, and methodologies.



# Students voting 'with their seats'



# A National Challenge

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.

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

Report | McKinsey Global Institute
Big data: The next frontier for innovation, competition, and productivity
May 2011 | by James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers

CRUNCH NETWORK
How To Stem The Global Shortage Of Data Scientists
Posted Dec 31, 2015 by Amy Gershkoff (@amygershkoff)

Berkeley
UNIVERSITY OF CALIFORNIA

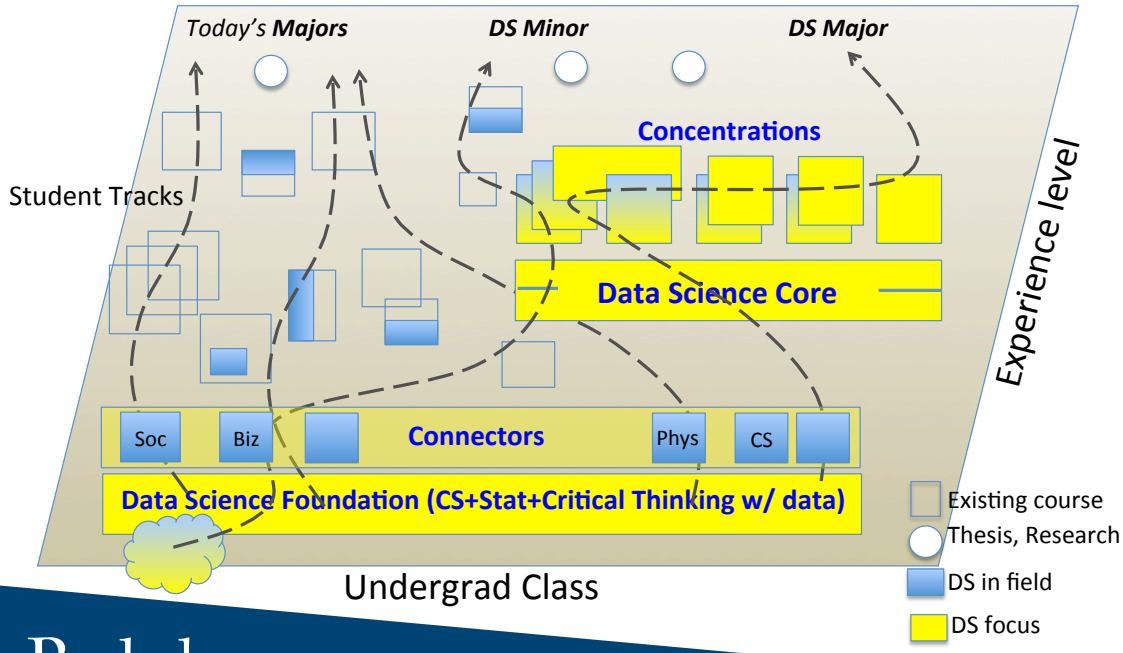


# Foundations of Data Science @ UCB

… rethinking at a fundamental level what every educated person must know about quantitative reasoning: how to effectively understand, process and interpret information, to inform decisions in their professional and personal lives and as citizens of the world in the 21st century.

-- from the Chancellor charge to the Data Science Education "Rapid Action Team," June 2014

# Data Science in the Undergrad Experience

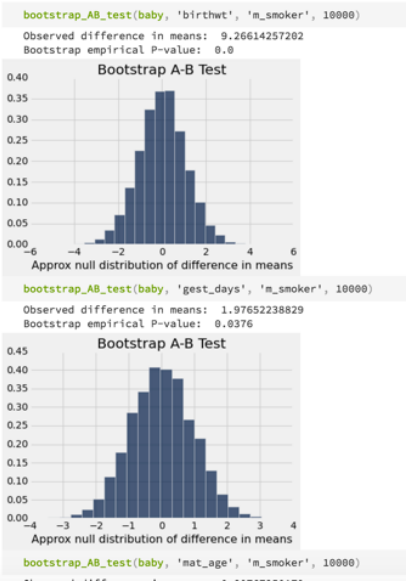


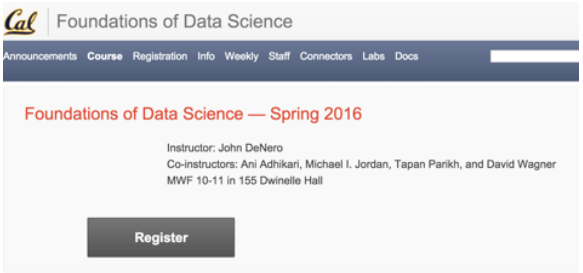

Berkeley
UNIVERSITY OF CALIFORNIA

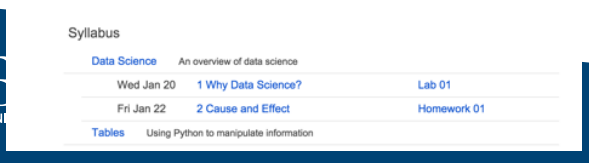


# Data8: Foundations of Data Science

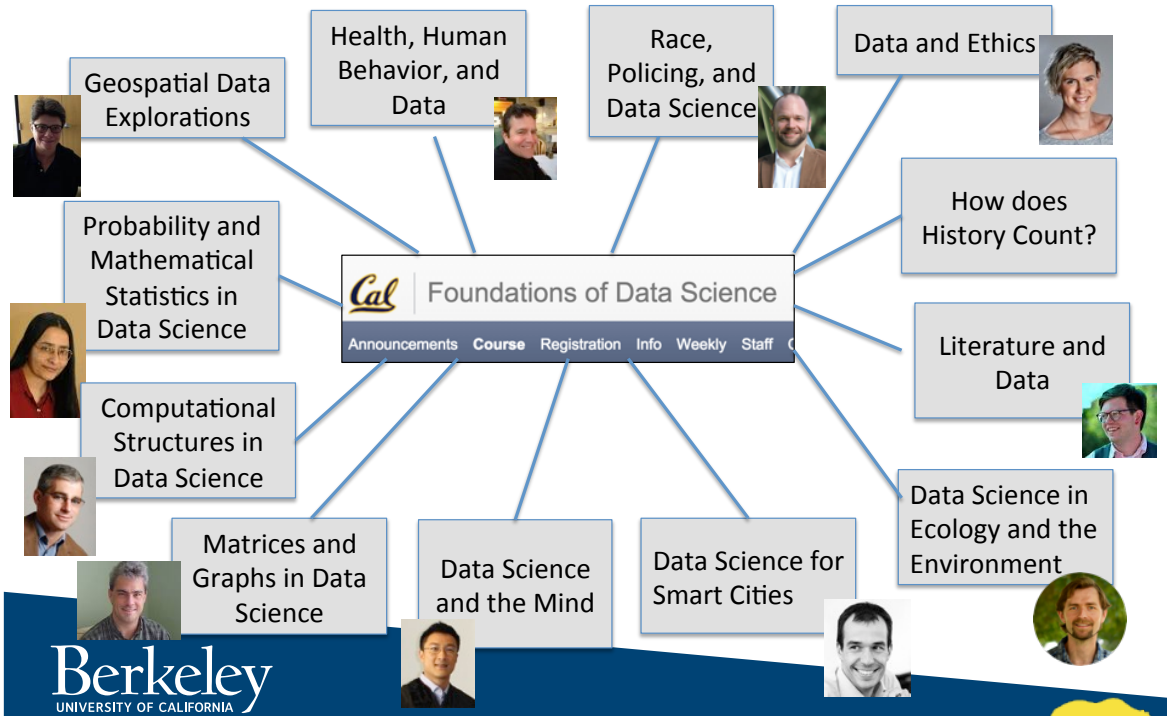
- http://data8.org
- Fundamental co-mingling of CS & Stat con
- Learn computing concepts by doing interes
- Learn statistical concepts by observing wha
- Codify understanding of concepts symbolic
- Explorations



# Data Science Connections






# DS UD Core – Centered on Life Cycle

research design → collection → preparation → analysis → utilization

*Perspectives*

Data Management: Acquisition | Organization | Storage | Retrieval | Curation

Data Analytics: Filtering | Fusing | Modeling | Prediction | Causal Inference | Diagnostics

Data Engineering: Pipeline Tools & Tech | Parallelization | Testing | Performance | Reproducibility Transparency

Decision Making: Visualization | Presentation | Interaction with Stakeholders

social, ethical, political, organizational, and privacy implications

# Class of 2020 …

Every college student should be prepared to understand and develop points of view based on the analysis of data as well as evaluate arguments made by others