Data Science in the 21st Century University

David Culler
University of California, Berkeley

Data Science Transforming Research

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

- Astronomy: LSST
- Physics: LHC
- Oceanography: OOI
- Sociology: The Web
- Biology: Sequencing
- Economics: POS terminals
- Neuroscience: EEG, fMRI
- Earthquake: Strong Shaking in 11 seconds

Reconstructing the movies in your mind

The Fourth Paradigm
1. Empirical = experimental
2. Theoretical
3. Computational
4. Data intensive

Berkeley

Data Science throughout campus

Berkeley
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’

[Graph showing enrollment trends in computing and statistics courses over years, with annotations for size of freshman class, computing intro, and stats intro.]
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.

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

Today's Majors

DS Minor

DS Major

Student Tracks

Connectors

Phys

CS

Data Science Core

Concentrations

Existing course

Thesis, Research

DS in field

DS focus

Rolling out to the students

Data Science in the Undergrad Experience

Undergrad Class

Data8: Foundations of Data Science

http://data8.org

Fundamental co-mingling of CS & Stat concepts

Learn computing concepts by doing interesting things on data

Learn statistical concepts by observing what's interesting

Codify understanding of concepts symbolically

Explorations

Foundations of Data Science

http://data8.org

Fundamental co-mingling of CS & Stat concepts

Learn computing concepts by doing interesting things on data

Learn statistical concepts by observing what's interesting

Codify understanding of concepts symbolically

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

Berkeley
UNIVERSITY OF CALIFORNIA
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.