Three Models for Undergraduate Data Science Curricula
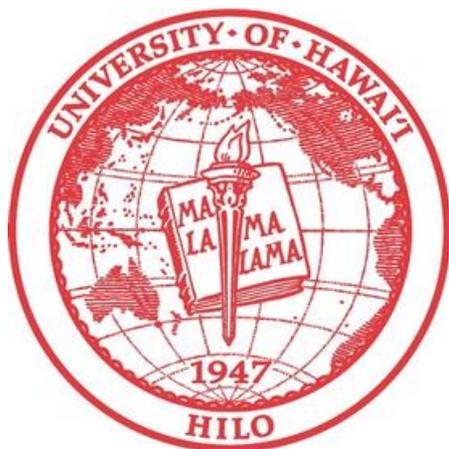
Bob Pelayo

University of California, Irvine

CSU MC^2 Series on Data Science Education

A Tale of Three Programs

A Tale of Three Programs

Underline: University of Hawai`i at Hilo

- Certificate of Data Science
- Planning for BS in Data Science (housed in CS Dept)

University of California, Irvine – Math

- Concentration within the Mathematics BS

University of California, Irvine – Stats

- BS in Data Science

The Birth of Three Programs – UH Hilo
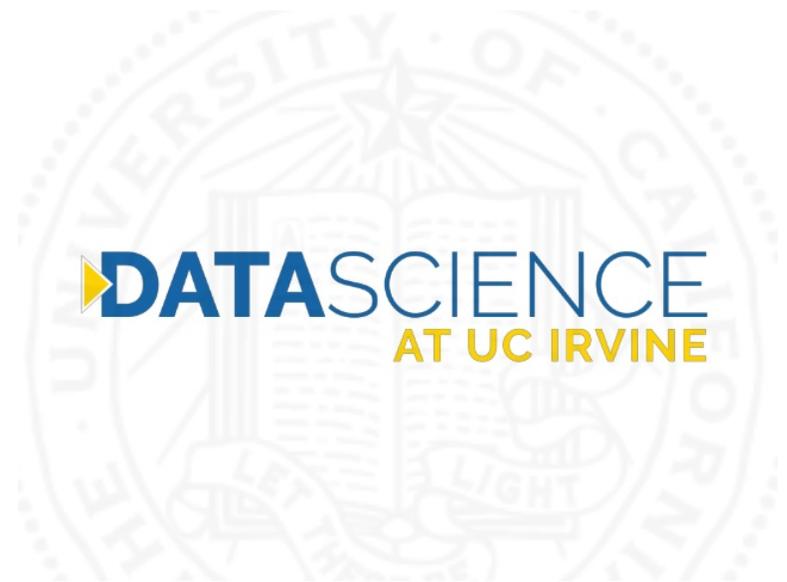
Certificate Program at UH Hilo

- Funded by a $20M NSF EPSCoR grant
- Initiated at the University level
- Hired 4 DS Faculty:
  - 1 in Math (Stats)
  - 1 in CS (ML)
  - 1 in Marine Science (Computer Vision)
  - 1 in Bus/Econ (NLP)
- Faculty in 2 Colleges & 4 Depts

'IKE WAI

The Birth of Three Programs – UCI Stats
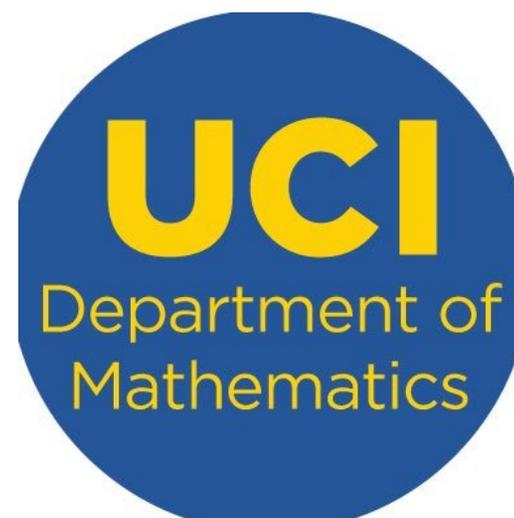
BS in Data Science

- Housed in the School (College) of Information & Computer Science
  - Stats and CS Depts!
- ICS School started in 2002
- BS in Data Science started in 2015

DATASCIENCE
AT UC IRVINE

The Birth of Three Programs – UCI Math

Data Science Concentration in Math

- Started in 2017 as 'overflow' for financial math concentration
- Parallel (independent) program to UCI Stats BS in DS
- Very few overlapping courses between Math & Stats

Curricular Structure – UH Hilo

<u>Certificate Program</u>

- 4 required lower-division courses:
  - CS 171: *Data Science Fundamentals in R*
  - CS 172: *Python for Data Analysis*
  - CS 272: *Machine Learning for Data Science*
  - Math 271: *Applied Statistics in R*
- 2 upper-division electives
  - Upper-div courses from various departments
  - Includes *Multivariate Modeling in R*
- Continue Building Towards BS Degree in Data Science

Curricular Structure – UCI Stats

BS in Data Science

- Several lower-division ICS, Math, and Stats courses
- Upper-division Stats
  - *Statistical Methods*, *Bayesian Stats*, & *Probability*
- Upper-division CS
  - *Machine Learning*, *Data Management*, *Design of Algorithms*
- Capstone
  - Two-quarter team-based projects

Curricular Structure – UCI Math

DS Concentration in Math Major

- All standard lower-division Math courses
- Math 10 – *Intro to Programming for Data Science*
  - Intro to Python and Machine Learning techniques
- Upper-division requirements
  - *Optimization*
  - *Numerical Analysis*
  - *Probability & Stochastic Processes*
- Many unrelated upper-division Math courses
  - *Group Theory, Ring Theory, Complex Analysis*

Strengths of Programs

UH Hilo Certificate in DS

- Recruit from different departments
- Courses relevant to domain science

UCI Stats BS in DS

- Nice blend of Stats & CS expertise

UCI Math Concentration in DS

- Strong mathematics foundation makes grad school admission easier

Curricular Challenges

UH Hilo Certificate in DS

- Interdisciplinary faculty have different expectations/priorities
- Difficult to control pre-requisites for upper-div courses

UCI Math Concentration in DS

- Lack of sufficient programming experience
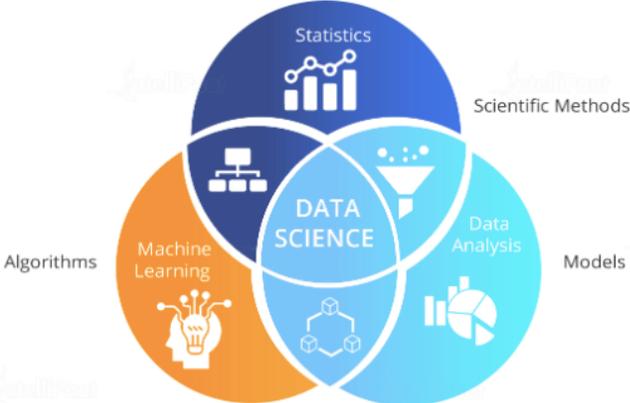- Difficulty getting Stats or CS course electives
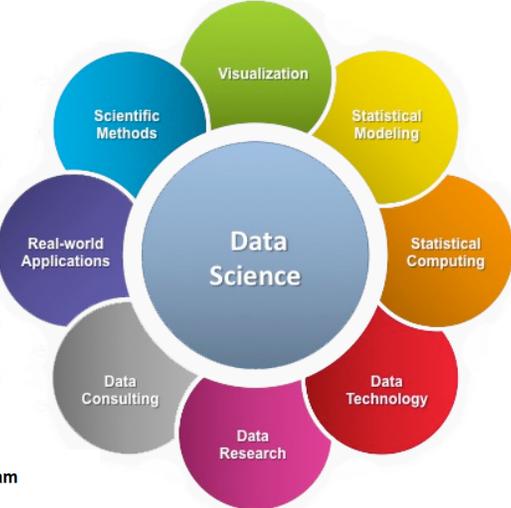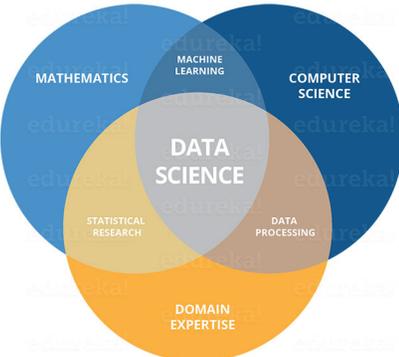
Concluding Thoughts

Our Data Science curricula are largely results of top-down decisions.

We need to leverage our strengths and find collaborations when we can.

Even if we had complete freedom, *could we even decide on what the ideal Data Science curricula is?*

Thank you!

What is Data Science?