If You Build It, They Will Come
Perspectives on an Undergraduate Statistics/Data Science Program

Rebecca Nugent

Department of Statistics
Carnegie Mellon University

December 12, 2016

Carnegie Mellon Undergraduate Statistics

- Housed in Dietrich College of Humanities & Social Sciences; students apply directly to college
- Statistics (with defined concentration), Economics-Statistics, Statistics & Machine Learning, Mathematical Statistics (track), Statistics and Neuroscience (track)
- over 300 majors; Statistics and Machine Learning rapidly growing
- intro level required for all Dietrich students
- teach about 1500 undergraduates (out of 6500) across the campus each semester
- class sizes at *all* levels are in the 100s

Statistics/Data Science Program Components

- Math/Intro level pre-requisites
- Solid probability/inference theory background
- Statistical computing (multiple languages, databases, etc)
- Modern regression/advanced data analysis methods
- wide breadth and depth of methodology courses
- research opportunities at all levels
- interdisciplinary projects in all classes
- no “textbook” data sets; all real problems
- reproducibility; well-structured, commented code
- communication, written and oral
- seeing an analysis through from start to end
- two-tiered advising system: Academic, Faculty Advisors

Lessons Learned/Learning?

Our viewpoint is that a modern, interdisciplinary undergraduate statistics program incorporating breadth of computing topics is doing data science; industry, grad schools seem to concur

- If you build it, they will come
- students love graphics/visualization, computing, data mining, upper level methods
- very wide computing/programming background; requires thoughtful revamp of related courses
- still need to work on adding languages/computing topics
- feedback is incredibly time-consuming
- faculty background and incentives
- hard to know what kinds of students to admit
- how to efficiently and effectively manage large programs