

# Teaching data science

**Eric B. Laber**

Department of Statistics, North Carolina State University

ENAR<sup>®</sup>

March 6, 2016

# Acknowledgments

- ▶ Joel Dubin and Sujit Ghosh
- ▶ Cisco Systems Inc.
- ▶ NCSU Corporate Education
- ▶ NCSU and UNC AI precision medicine lab
- ▶ NCSU personalized medicine cluster



# Data science

*"... Data Science is the extraction of knowledge from large volumes of data that aren't structured..."*

—Google

- ▶ Teaching data science is not easy
  - ▶ Broadly defined
  - ▶ Heterogeneity across 'data scientist' responsibilities
  - ▶ Heterogeneity in background/training


**<BIG DATA /  >**

# Data science cont'd

- ▶ My experience teaching data science
  - ▶ Short workshops for industry (2 days - 2 weeks)
  - ▶ Semester long course for industry
  - ▶ Masters/PhD statistics course on big data
- ▶ Teaching philosophy for data science
  - ▶ Ideal: teach ideas not tools
    - ▶ Sampling, design of experiments, causal inference, theory of estimation and inference, computing, optimization, etc.
    - ▶ Case studies/extended illustrative examples use tools (software) to illustrate underlying principals
    - ▶ Possible with PhD students
  - ▶ Reality: risk 'master of none'

# Material delivery

- ▶ Mix of traditional 75 minute lectures and lab sessions
  - ▶ Background material including pre-recorded lectures, tutorials, readings, and comprehension quizzes provided online
  - ▶ Homework primarily open-ended, data-heavy, comp. intensive
  - ▶ Require students to read APIs, message boards, etc.
  - ▶ Communication of results weighted heavily in grading
- ▶ Launching online program at NCSU this year!



Thank you.  
laber@stat.ncsu.edu