Big Picture for a Big Data Science Education Network: Data Science Education Initiatives

Nicholas Horton

Amherst College

nhorton@amherst.edu

ASA Guidelines for Statistics Major and Minor Programs (2014)

- Students need integration of computational skills, statistical theory, statistical practice, and communication

- Need to restructure curriculum to help them "think with data"

- Stresses key role of data science

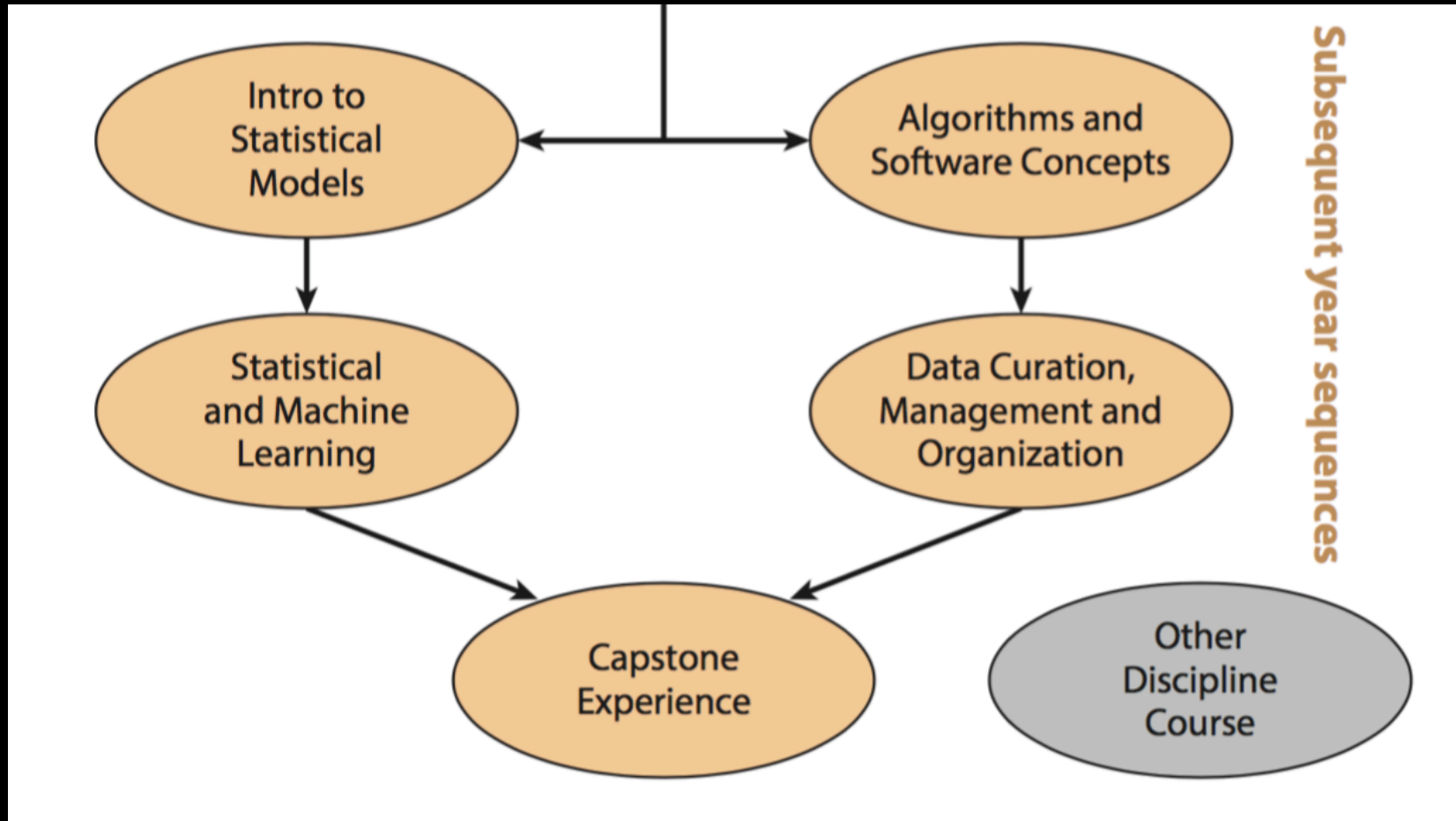http://amstat.tandfonline.com/doi/full/10.1080/00031305.2015.1094283

PCMI Guidelines for Data Science Programs
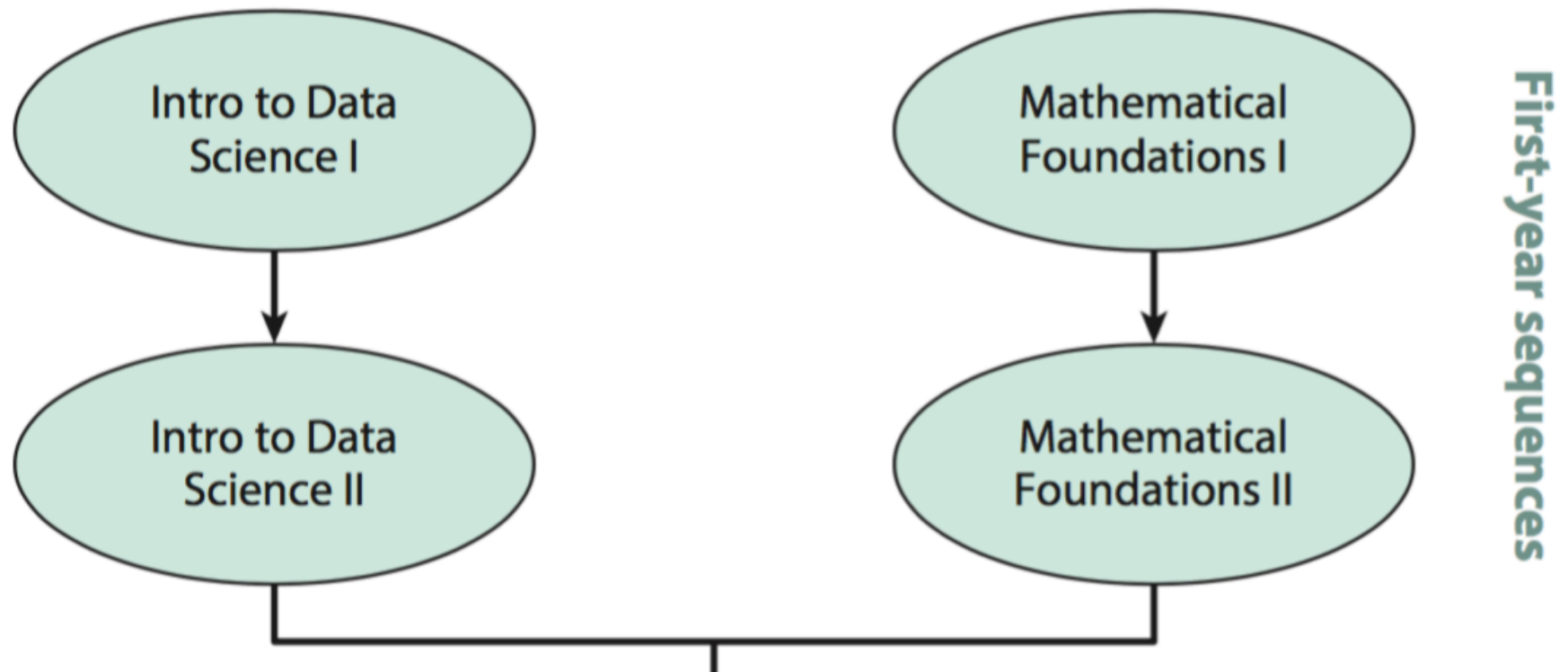
PCMI Guidelines for Data Science Programs

Envisioning the Data Science Discipline: The Undergraduate Perspective

Carry out a **consensus study** to

- Identify core underlying principles, intellectual content, and pedagogical issues specific to data science, including core concepts that distinguish it from neighboring disciplines.

http://nas.edu/EnvisioningDS

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Envisioning the Data Science Discipline: The Undergraduate Perspective

Carry out a **consensus study** to
- Focus on undergraduate level, but address topics related to middle and high schools, community colleges, and draw on experiences of Master's-level programs.
- Consider opportunities to engage underrepresented student populations.

http://nas.edu/EnvisioningDS

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

The Data Science Education Roundtable (DSERT)

nas.edu/DSERT

ROUNDTABLE ON DATA SCIENCE POST-SECONDARY EDUCATION

GORDON AND BETTY **MOORE** FOUNDATION

NIH National Institutes of Health

acm

ASA AMERICAN STATISTICAL ASSOCIATION
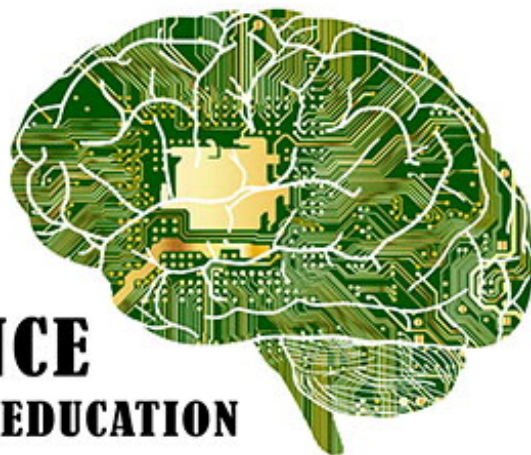
Bring together CS, stat, math, and domain fields

Strengthen ties between industry and academia

Highlight new and innovative programs

http://nas.edu/DSERT

Every Meeting Advertised, Live Streamed, Archived & Summarized

Roundtable on Data Science Postsecondary Education Meeting #2:

Examining the Intersection of Domain Expertise and Data Science

March 20, 2017

Beckman Center, Huntington Room
100 Academy Way
Irvine, CA 92617

ROUNDTABLE ON
DATA SCIENCE
POSTSECONDARY EDUCATION

The National Academies of Sciences, Engineering, and Medicine held a one-day meeting and webcast on data science postsecondary education on March 20, 2017. This meeting brought together data scientists and educators to discuss how to define and strengthen data science education in data intensive domains such as digital humanities and astronomy, and to discuss several case studies of domain-focused data science education ongoing at several universities.

This event is the second of an ongoing series of roundtable meetings that will take place approximately four times per year. This roundtable is organized by the Committee on Applied and Theoretical Statistics in conjunction with the Board on Mathematical Sciences and Their Applications, the Computer Science and Telecommunications Board, and the Board on Science Education.

This roundtable is sponsored by the Gordon and Betty Moore Foundation, the National Institutes of Health, the National Academy of Sciences W. K. Kellogg Foundation Fund, the Association for Computing Machinery, and the American Statistical Association.

Meeting #2 Highlights

Roundtable Members

Meeting Agenda

9:00 a.m. Welcome and overview of meeting
Eric Koalczyk, Boston University and Kathy McKeown, Columbia University
Video

Emerging Needs and Opportunities in Data-Intensive Domains
9:15 a.m. English
Ted Underwood, University of Illinois at Urbana-Champaign
Video - Presentation

9:45 a.m. Astronomy

August 2017

Roundtable
HIGHLIGHTS

Roundtable on Data Science Postsecondary Education
Meeting #3 - May 1, 2017

The third Roundtable on Data Science Postsecondary Education met on May 1, 2017, at the Pew Research Center in Washington, D.C. Stakeholders from data science education programs, funding organizations, government agencies, professional societies, foundations, and industry convened to discuss data science training in the workplace. This Roundtable Highlights summarizes the presentations and discussions that took place during the meeting. The opinions presented are those of the individual participants and do not necessarily reflect the views of the National Academies or the sponsors. Watch meeting videos or download presentations at nas.edu/DSERT.

DATA SCIENCE TRAINING IN THE WORKPLACE: GOVERNMENT

Practicing Data Science in the Government
Ron Prevost, U.S. Census Bureau

Prevost explained that data produced by the U.S. Census Bureau are expected to be unbiased, statistically accurate, delivered quickly at low cost, useful to determine causality, reproducible, transparent, and protected. While striving to meet these expectations, statistical agencies confront many challenges, including greater than expected costs and lower than expected response rates for surveys, complex information requests, competition among data products and questions of product validity, new data sources and methodologies, and policy requirements.

The Census Bureau hopes to supplement survey data with data that have been assessed from other sources. However, this data integration needs to be transparent and reliable, utilize quality measures, and ideally incorporate model-based estimation and data source acquisition and integration processes. In his view, the Census Bureau could advance this paradigm shift by taking several actions: solutions, (1) consolidating business processes and systems and optimizing solutions, (2) supplementing current business processes with new processes, (3) developing new products, (4) building new capabilities, and (5) optimizing current business processes. Prevost noted that institutional, bud-

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

More than 1000 viewers across the webcasts

http://nas.edu/DSERT

8

Past Meetings of the DSERT

- December 14, 2016: Foundations of Data Science from Statistics, Computer Science, Mathematics, and Engineering

- March 20, 2017: Examining the Intersection of Domain Expertise and Data Science

- May 1, 2017: Data Science Education in the Workplace

- October 20, 2017: Alternative Mechanisms for Data Science Education

- December 8, 2017: Integrating Ethics and Privacy Concerns into Data Science Education

Looking forward, new topics and continuing the discussion of this group

- **March 5, Berkeley CA:** Teaching Data Science as a Scientific Process

- **June 13, 2018:** 2 Year Colleges, Washington, DC (tentative)

- **September 17, 2018:** Improving Female and Underrepresented Minority Representation in Data Science, Atlanta, GA (tentative)

- **December 10, 2018:** Balancing Education in Employable Skills and General Knowledge, Washington, DC (tentative)

Two-Year College Data Science Summit

- NSF funded workshop on two-year college data science programs

- Organized by the American Statistical Association (Steve Pierson)

- Will take place in May, 2018 in Washington, DC

https://www.amstat.org/ASA/Education/Two-Year-College-Data-Science-Summit.aspx

Other related projects

- Practical Data Science (PeerJ collection): Bryan and Wickham

https://peerj.com/collections/50-practicaldatascistats/

- 50 years of Data Science (with discussion and focus on educational implications)

http://www.tandfonline.com/toc/ucgs20/26/4

Big Picture for a Big Data Science Education Network: Data Science Education Initiatives

Nicholas Horton

Amherst College

nhorton@amherst.edu