

Dana Center
Mathematics
PATHWAYS

**Everything a President/Chancellor/Provost/Dean
Needs to Know About Data Science Education**

Nicholas Horton, Amherst College

Roxy Peck, Cal Poly San Luis Obispo

November 25, 2019



 www.dcmathpathways.org

1

Plan for the webinar

1. Review recent data science curricular reports (National Academies initiatives and the Two Year College Data Science Summit)
2. Consider statistical and mathematical foundations for data science
3. Consider ways to engage academic leaders
4. Discussion and Q&A

Dana Center
Mathematics
PATHWAYS

2

DATA SCIENCE FOR UNDERGRADUATES

Opportunities and Options

Published in 2018

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/EnvisioningDS

5

Key Insights: Data Science

- We are in the infancy of data science
- Data science is a unique field that borrows heavily from multiple other fields
 - A major/minor/certificate/etc. should not be the same as, e.g., a degree in statistics or in computer science
 - There will need to be educational opportunities to expose faculty to the breadth of the field
 - There will need to be ways to share educational resources (e.g., course materials, etc.)
- Evaluation and assessment is critical

Dana Center
Mathematics
PATHWAYS

6

Knowledge for Data Scientists

Recommendation 2.1 Academic institutions **should embrace data science as a vital new field** that requires specifically tailored instruction delivered through majors and minors in data science as well as the development of a cadre of faculty equipped to teach in this new field.

Recommendation 2.2 Academic institutions should provide and evolve a range of **educational pathways** to prepare students for an array of data science roles in the workplace.

Recommendation 2.3 To prepare their graduates for this new data-driven era, academic institutions should encourage the development of a **basic understanding of data science in all undergraduates**.

Recommendation 2.4 Academic institutions should **ensure that ethics is woven into** the data science curriculum

NASEM 2018: Data acumen

Finding 2.3 A critical task in the education of future data scientists is to instill **data acumen**. This requires exposure to key concepts in data science, real-world data and problems that can reinforce the limitations of tools, and ethical considerations. Key concepts include the following:

- **Mathematical foundations**
- Computational foundations
- **Statistical foundations**
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- Communication and teamwork
- Domain-specific considerations
- Ethical problem solving.

Data Science as way to broaden participation

Finding 4.1: The nature of data science is such that it offers **multiple pathways** for students of different backgrounds to engage at levels ranging from basic to expert.

Finding 4.2: Data science would particularly **benefit from broad participation** by underrepresented minorities because of the many applications to problems of interest to diverse populations.

Recommendation 4.1: As data science programs develop, they should focus on attracting students with **varied backgrounds and degrees of preparation** and preparing them for success in a variety of careers.

Data acumen: Mathematical concepts

Key **mathematical** concepts/skills that would be important for all students in their data science programs and critical for their success in the workforce are the following:

- Set theory and basic logic,
- Multivariate thinking via functions and graphical displays,
- Basic probability theory and randomness,
- Matrices and basic linear algebra,
- Networks and graph theory, and
- Optimization.

Data acumen: Statistical concepts

Important **statistical foundations** might include the following:

- Variability, uncertainty, sampling error, and inference;
- Multivariate thinking;
- Nonsampling error, design, experiments (e.g., A/B testing), biases, confounding, and causal inference;
- Exploratory data analysis;
- Statistical modeling and model assessment; and
- Simulations and experiments

Two-Year College Data Science Summit:

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

- First programs in data science were primarily at the graduate and bachelors degree level.
- NAS report recognizes that two-year colleges have an important role to play in data science education.
- Report from the summit should be of value to both four-year colleges with transfer and articulation with two-year colleges as well as two-year colleges that are developing data science programs.
- Learning outcomes identified appropriate for two-year programs as well as first two years of college at four-year institutions

Two-Year College Data Science Summit:

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

Identified Three Potential Types of Data Science Programs at Two-Year Colleges

1. **Two-year degree designed for those who plan to enter the workforce**
2. **Two-year degree designed for those who plan to transfer to a four-year data science program**
3. **Certificate programs designed for those already in the workforce or for those with degrees in other areas that want to move in a data science direction**

Dana Center
Mathematics
PATHWAYS

13

Two Year College Data Science Summit Statistics Learning Outcomes

- Determine if conclusions are appropriate for a study based on study design (observational or controlled experiment), including identifying potential confounding factors and appropriate controls.
- Produce and interpret data visualizations, including dashboards, graphs and charts to describe and explore data and communicate findings.
- Produce and interpret numerical summaries to describe and explore data
- Produce and interpret confidence intervals
- Formulate statistical claims in terms of null and alternative hypotheses, carry out and interpret basic hypothesis tests
- Investigate and explore relationships between more than two variables.
- Use statistical models to describe relationships between variables.
- Fit, interpret and evaluate basic statistical models (e.g. linear, logistic, exponential).
- Explain the purpose of cross-validation

Dana Center
Mathematics
PATHWAYS

14

Statistics Outcomes from Two-Year College Data Science Summit...

Consistent with the recommendations in Data Science for Undergraduates: Opportunities and Options, The National Academies of Science (2018)

Statistical Foundations

- Variability, uncertainty, sampling error and inference
- Multivariate thinking
- Nonsampling error, experimental design, bias, confounding
- Exploratory data analysis
- Statistical modeling and model assessment
- Simulations and experiments

Statistics for Data Science

- **New course?**
- **Re-think of existing introductory statistics course?**
 - Incorporation of multivariate thinking
 - Confounding and causal inference
 - Appropriate technology (not Excel or calculators)
 - Build on foundation of statistics in high school

Two-Year College Data Science Summit

- **Recommendations related to statistics foundations include:**
 - Reduce barriers to entry
 - Create courses with modern and compelling introduction to statistics
 - Ensure opportunities to engage with realistic problems with real data
 - Integrate ethical considerations throughout
 - Foster active learning

Mathematics Foundations for Data Science

- How do we ensure that the mathematics foundations are included?
- Current standard option: Pre-calc/calc 1/calc 2/calc 3, linear algebra
- New pathways?

Data acumen: Mathematical concepts

Key **mathematical** concepts/skills that would be important for all students in their data science programs and critical for their success in the workforce are the following:

- Set theory and basic logic,
- Multivariate thinking via functions and graphical displays,
- Basic probability theory and randomness,
- Matrices and basic linear algebra,
- Networks and graph theory, and
- Optimization.

More on mathematical foundations

- Hardin and Horton, "Ensuring that Mathematics is Relevant in a World of Data Science", AMS Notices, 2017)
- Proposed new course in **continuous** mathematics that focused on modeling and optimization for problem solving
- Proposed a course in **discrete** mathematics organized around linear algebra, counting, graph theory
- Both tightly integrated with computation and modeling
- Consistent with NASEM and Two Year College Summit recommendations

Making this happen at your institution

- **Having a conversation with your academic leader**
 - Perspectives
 - Faculty member
 - Academic leader
- **What are the questions that might be relevant?**
- **How to set up those conversations for success?**

The big meeting (role playing)

- Having a (productive) conversation with your academic leader
- What are the questions that might be relevant?
- How to set up those conversations for success?

Preparing for discussion with academic leaders

- **Some questions that administrator might ask:**
 - What are the opportunities provided by new program?
 - We already have a stat major/minor. How is this different?
 - Sounds expensive...what is it going to cost?
 - Do we have faculty that can do this?
 - How does this fit with the mission of the university?
 - Is there evidence of a demand—students/employers?
 - Won't this take away enrollment in STEM disciplines?
 - Can you get the math faculty, the statistics faculty and the CS faculty to work together effectively?

Talking with academic leaders: key points

- Don't ask for too little
- Don't ask for too much
- Tie back to institutional mission and goals
- Do your homework!
- Build across departments and programs

Questions and Discussion

Please submit a question or comment through the “chat” window and we will respond

Dana Center
Mathematics
PATHWAYS

25

Dana Center
Mathematics
PATHWAYS

Everything a President/Chancellor/Provost/Dean Needs to Know About Data Science Education

Nicholas Horton, Amherst College

Roxy Peck, Cal Poly San Luis Obispo

November 25, 2019



The University of Texas at Austin
Charles A. Dana Center

www.dcmathpathways.org

26