Aligning expectations for the emergent discipline of data science education

NZSA Visiting Lecturer Seminar Massey University

28 April, 2025

Matthew Beckman

Assoc Research Professor, Penn State Director, CAUSE | www.causeweb.org

https://mdbeckman.github.io/



Penn State University | Department of Statistics









University of Minnesota

Driven to Discover™





Image credits (clockwise from left):

- Kennedy, P., (6 Feb 2023). Ecolab now selling products at Home Depot-the first time available at retail stores, Star Tribune.
- Medtronic Operational Headquarters, https://asiapac.medtronic.com/xp-en/about.html
- Rogers, G., (9 Apr 2018). World's smallest battery-powered implant treating back pain. 9 News.
- Lee, E., (4 Oct 2022), The Best Pulse Oximeter for Home Use, New York Times Wirecutter.

Consortium for the Advancement of Undergraduate Statistics Education



Connect. Share. Educate www.causeweb.org/

Community

- ~90 member institutions;
- thousands of individual educators;

Events & conferences

- eCOTS; USCOTS;
- eUSR; VOICES

Contests & Engagement

- o captions, a-mu-sing, SPARKS,
- USPROC, USCLAP, etc

Resources

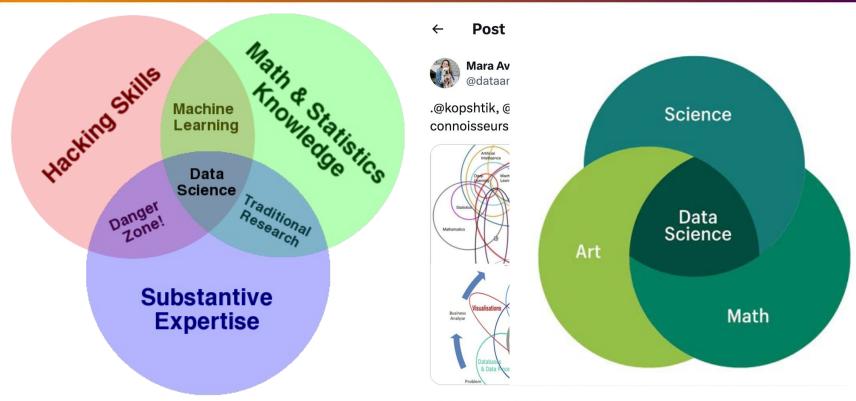
- webinars, webpages,
- professional development

Mission: support and advance undergraduate statistics education, in four target areas: resources, professional development, outreach, and research.

How might CAUSE embrace Data Science Education?

- Rebrand? (CAUDSE)
- Fully Integrate? (CAUSDSE)
- o Just let it blow over?

What IS Data Science anyway? Why is it even a thing?



10:51 PM · Apr 3, 2017

Image credits:

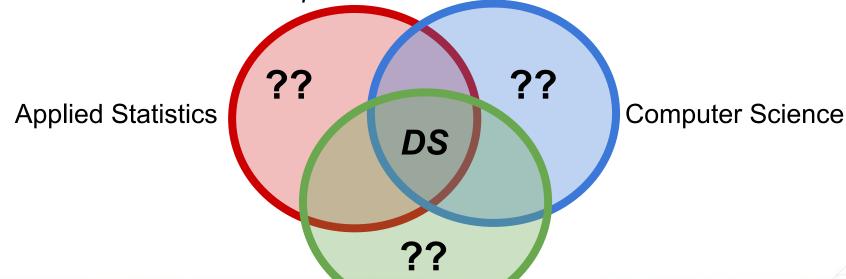
- Drew Conway; http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram
- https://shopify.engineering/data-storytelling-shopify

What ISN'T Data Science...

...but IS part of Applied Statistics?

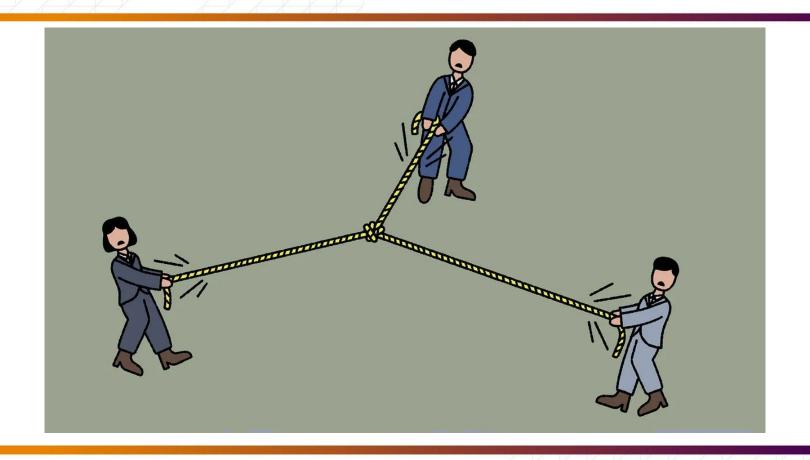
...but IS part of Computer Science?

...but IS part of Information Science?



Domain Knowledge

Where does Data Science Education belong?

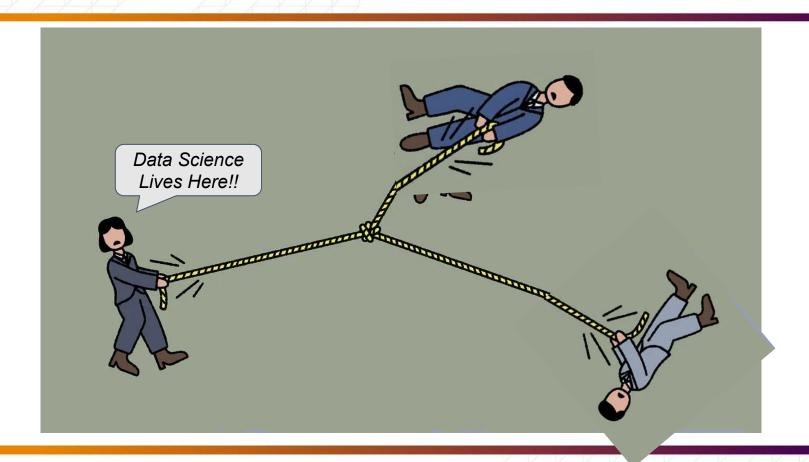


Tension within and among the disciplines

We want [our department] to be the home for DS at our institution so we can...

- Stay relevant in a changing world
- Hire additional faculty with complementary expertise
- Attract students
- Bolster industry connections
- Generate revenue for the department

After all, DS needs to have a home *somewhere*, so who better than us?



Alignment of Expectations

- What makes DS education different from Statistics, IS, or CS Education?
- What can we learn from Statistics, IS, & CS Education?
- Where does / should Data Science "live" within academic institutions?
- Is the whole truly greater than the sum of it's parts?

So now [your department] is home of Data Science

- Faculty / administration experience?
- Student experience?
- Industry experience?

Alignment: Faculty / Administration Experience

- Curricular turf wars & "copy-cat" programs
- Accreditation confusion
- Disciplinary silos vs true collaboration
- Pressure to teach fast-moving content (AI, LLMs, etc.)
- Concern for fundamentals

Alignment: Student Experience

- Statistics vs statistics
 - "statistic"--collective property of some data (e.g., calculated from a sample)
 - the field of <u>S</u>tatistics offers the world is the means to appropriately accommodate, characterize, and even quantify variability and uncertainty
- Perception vs potential of exploratory data analysis (EDA)
 - Especially true of "advanced novices"
 - In a rush to use fancy tools and "get the answer"

Alignment: Student Experience

- Statistics vs statistics
 - "statistic"--collective property of some data (e.g., calculated from a sample)
 - the field of <u>S</u>tatistics offers the world is the means to appropriately accommodate, characterize, and even quantify variability and uncertainty
- Perception vs potential of exploratory data analysis (EDA)
 - Examine the data source(s): provenance, missingness, data viz, etc
 - Discover features that affect modelling decisions: independence, outliers, missingness
 - Address research questions: intuition for results, build up from minimalism, refine RQ's

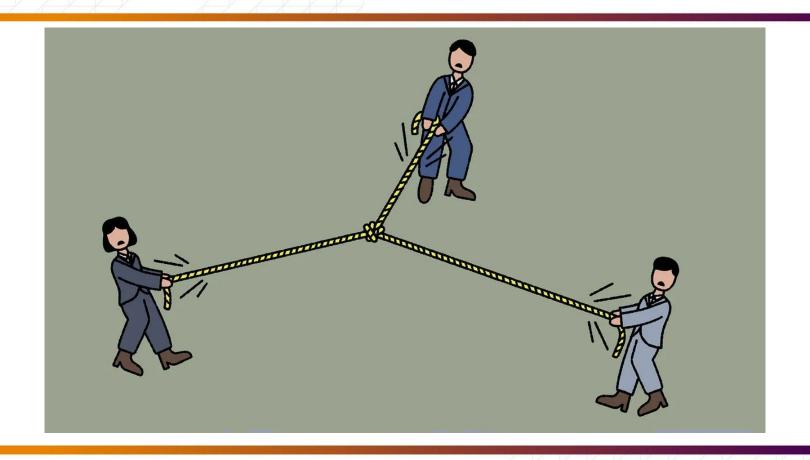
Alignment: Student Experience

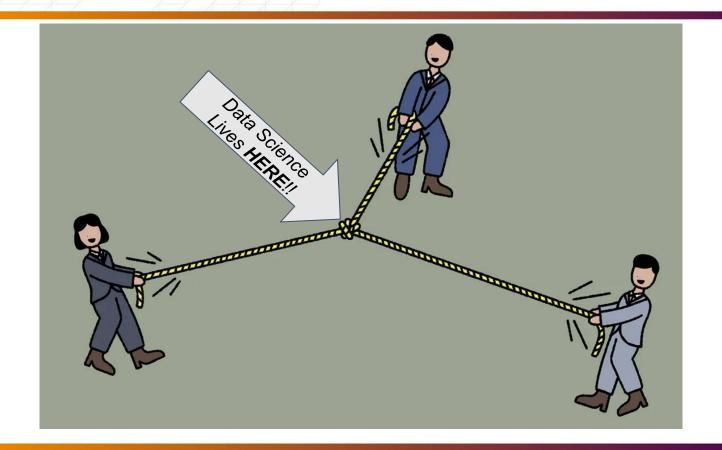
- Demand for fast-moving content (AI, LLMs, etc.)
- Expecting specific tech skills?
- "learning everything" vs "specializing"
- Implementation vs understanding of tech & tools
- Patience with fundamentals
- Expectations of entry level positions

Alignment: Industry Expectations

- "You're a data scientist?! I don't know what that is, but I know we need them!"
 - -Anonymous recruiter at University Career Fair (true story)
 - Shifting expectations about what a "data science degree" actually means
 - Often preoccupied with specific tools & contexts
 - Entry-level expectations
- On the job, Statisticians & Data Scientists are regularly requested to engage with myriad issues that require no statistical or data analysis at all. . .
 - @ Ecolab, Medtronic, Nonin, and elsewhere
 - o **Bad news**—maybe management is unclear how to utilize the skill set
 - Good reason—we don't just "crunch numbers" & make apps... expertise includes a
 disciplined approach to problem solving & critical thinking; due consideration for
 uncertainty, alternative explanations, and practical implications

Where does Data Science Education belong?





•

What Is Data Science? Who "Owns" It?

- Not just tools, but essential ways of thinking
 - Statistical thinking
 - Computational thinking
 - Domain expertise
 - Ethics & communication
- Interdisciplinarity to the core
- Why this conversation matters now

Regarded as a distinct discipline...

- Data Science Curriculum is a negotiation
- Curriculum ownership / responsibility is collaborative
- The whole is greater than the sum of its parts
- Norms of teaching, learning, and assessment for Data Science Education are questioned, challenged, & refined

Case Study: Data Science Education at Penn State

Opportunities & Optimism

- Data Science Curriculum Guidelines (Park City Mathematics Institute)
- GAISE College Report
 - Ongoing revision addresses both Statistics & Data Science
 - Assessment in the age of Al
- Organizations Embracing Data Science Education
 - Data Science 4 Everyone
 - CAUSE
 - ACM SIGCSE (Special Interest Group in CS Education)
 - ASA Section on Statistics and Data Sciences Education
- Journals that regularly publish DS Education
 - Harvard Data Science Review (Stepping Stones)
 - Journal of Statistics and Data Science Education

Aligning expectations for the emergent discipline of data science education

NZSA Visiting Lecturer Seminar University of Otago

14 April, 2025

Matthew Beckman

Assoc Research Professor, Penn State Director, CAUSE | www.causeweb.org

https://mdbeckman.github.io/



