What is the role of data science in K-12 (pre-collegiate) teaching and learning?

A view from the education side of the fence

Joshua M. Rosenberg, Ph.D., Department of Theory and Practice in Teacher Education, University of Tennessee, Knoxville

Introduction

How I came to this work

Introduction

- 2011: Seeing my high school science students feel proud of an Excel graph they could tailor
- 2014: Learning R
- 2015: Being pointed to educational research on “data modeling”
- 2017: Seeing a peer in my graduate program in a quantitative/computational research methods class share with joy a word cloud they created
- 2017: Starting to ask, “What power do students have when it comes to analyzing data in education?”
- 2018: Learning about mixed effects linear regression models for the analysis of complex educational data—and their Bayesian interpretation
- 2018: Presenting on “Data science education dilemmas (and opportunities)”
- 2020: Publishing “Data science in education using R”

A bit more about me

Introduction

- I study teaching, learning, and educational systems—and engaging in design and development to support these
- A lot of the time, I focus on science education, but I also explore other sub-disciplines
- I do so using a range of methods—qualitative and *(especially)* quantitative and computational (more on this in a bit!)
- I draw extensively on my experience as a high school science teacher
- I'm in my fourth year as a faculty member at UTK, having completed my Ph.D. in 2018 from Michigan State University

A bit more about me

Introduction

Goals for today

Introduction

- Establish the context and problem
- Present an opinionated solution
- Provide two examples of research that reflect this opinionated stance
- Share some ways you can engage in educational data science research
- And, to try to share some ideas that interest you (or build on your interest) in doing educational research

But first

Introduction

- On a scale of 1-10 (use your hands), how confident are you in your knowledge and capabilities regarding:
  - Mathematics?
  - How about statistics?
  - Computer science?
  - Educational research?
  - Educational systems?

Context and Problem

The transformative potential of educational data science

Context and Problem

- “The main message of this article is that **the digital age is having a profound impact on statistics and the nature of data analysis**, and these changes necessitate reevaluation of the training and education practices in statistics.

- “In particular, **computing is an increasingly important and necessary aspect of a statistician’s work**, and needs to be incorporated more fully into statistics training.”

Nolan and Temple Lang (2012), *The American Statistician*

The transformative potential of educational data science

Context and Problem

- Not only having an extensive provenance, such calls have also grown over time:
  - “Through data science, students can learn to answer questions that are relevant to their lives and communities, to be critical consumers of the data that surround them every day, and to wield the power of data analysis.” (LaMar & Boaler, 2021)

See also: Bargagliotti et al. (2021), Hardin et al. (2015), Hardin et al. (2021), McFarland et al. (2021), Lee & Campbell (2020), Rosenberg et al. (2020), Wilkerson & Polman (2018)

Others have expressed some caution

Context and Problem

- Is data science really different from statistics education? (Rubin, 2019)
- How should we think about the myriad factors that influence whether and how students engage with data? (Lee et al., 2021)
- Who benefits from the abundance of available data? Who is empowered and who is disempowered? Who receives credit? Is it ethical? (D'Ignazio & Klein, 2020)

So, is there a role for data science in education?

Context and Problem

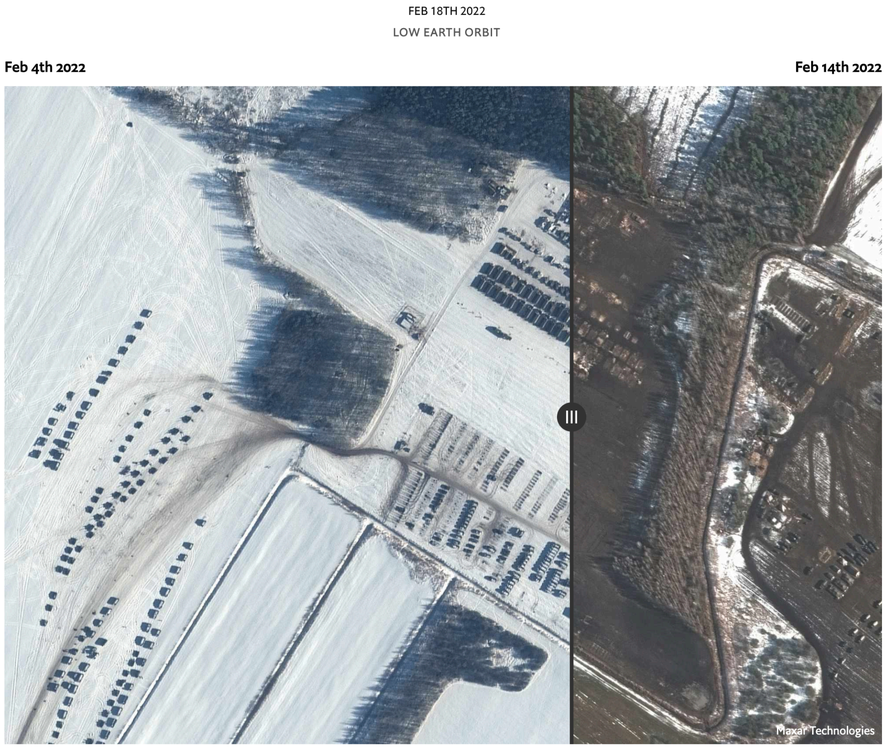
- Yes, there is in pre-collegiate (K-12) settings:
  - To develop cross-discipline capabilities
  - To support learners to deploy a critical data literacy
  - To ensure that the benefits of being able to work with data are widespread
  - To study teaching and learning in new, important ways
  - To build capacity within educational settings to ask and answer data-related questions (and to identify and work to solve problems)

So, is there a role for data science in education?

Context and Problem

The law, called the Digital Markets Act, would be the most sweeping piece of digital policy since the bloc put the world's toughest rules to protect people's online data into effect in 2018. The legislation is aimed at stopping the largest tech platforms from using their interlocking services and considerable resources to box in users and squash emerging rivals, creating room for new entrants and fostering more competition.

What that means practically is that companies like Google could no longer collect data from different services to offer targeted ads without users' consent and that Apple might have to allow alternatives to its App Store on iPhones and iPads. Violators of the law, which would most likely take effect early next year, could face significant fines.



Rechitsa, Belarus, was host to troops from Russia's 36th Combined Arms Army. By February 14th they had largely vanished.

Conjecture

Context and Problem

- Educational data science *does* have a role in education, but how we do this work matters
  - In terms of impact *and* sustainability over time

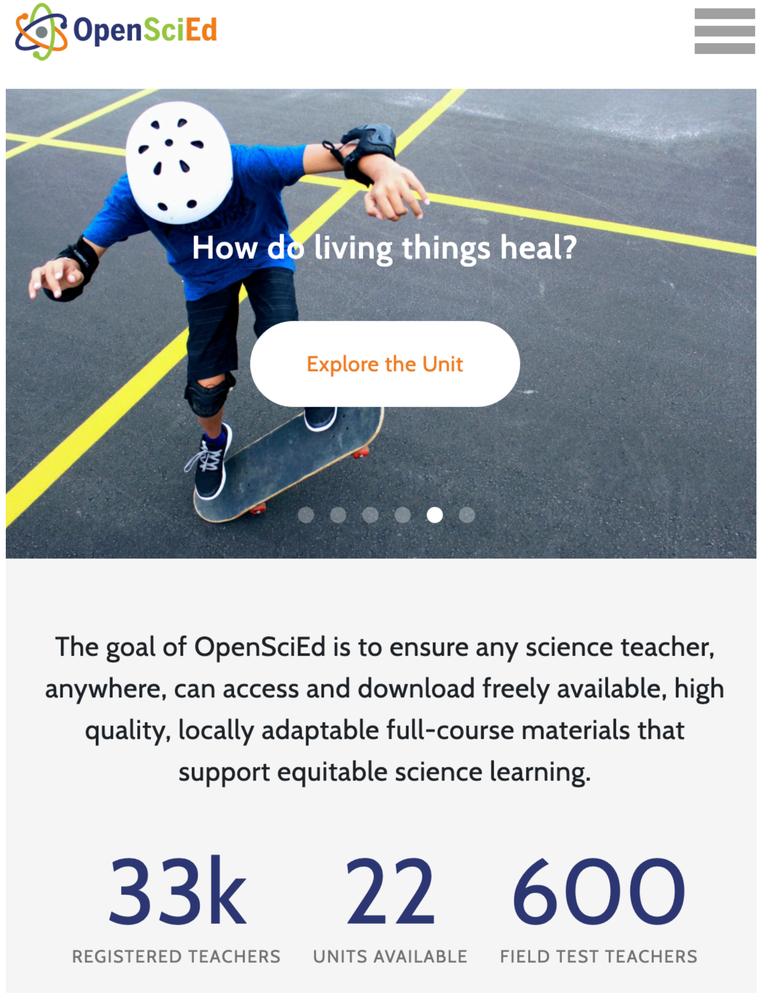
Reactions or questions?

A Proposed Solution

Leverage key ideas about education

A Proposed Solution

- Draw on:
  - Theories of learning
  - Accounts of educational change
  - Models of professional learning
  - Design guidelines



Leverage key ideas about mathematics and statistics

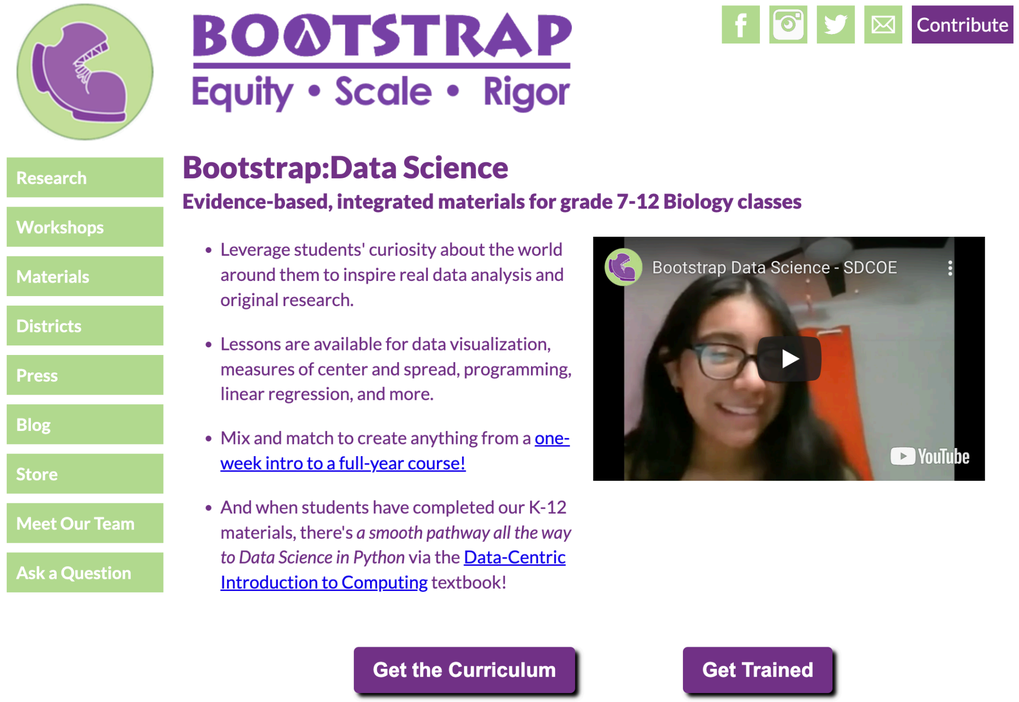
A Proposed Solution

- Draw on:
  - Formal methods of mathematical and statistical analysis
  - Approaches to exploratory data analysis
  - Key mathematical and statistical ideas
  - Mathematical ways of thinking

Leverage key ideas about computation

A Proposed Solution

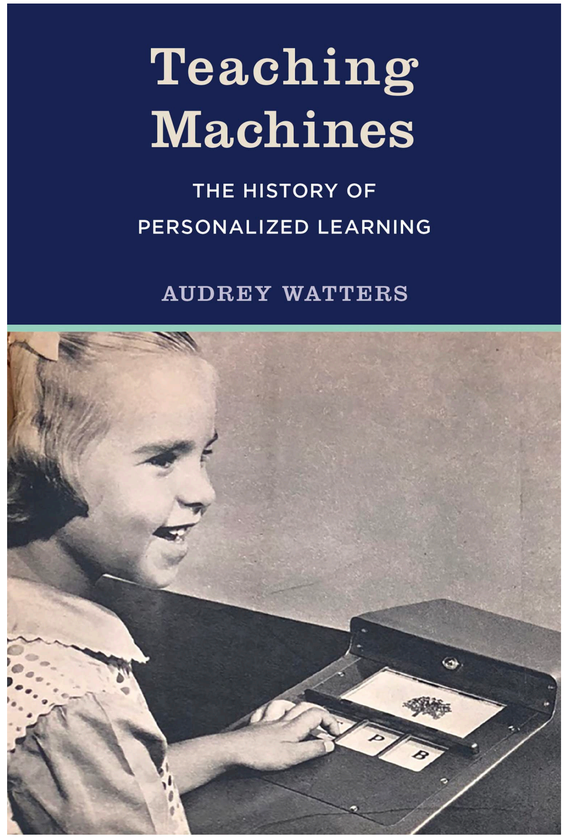
- Draw on:
  - Computationally-intensive methods
  - Digitally-collected data
  - Computational ways of thinking
  - Statistical software



Consider the context

A Proposed Solution

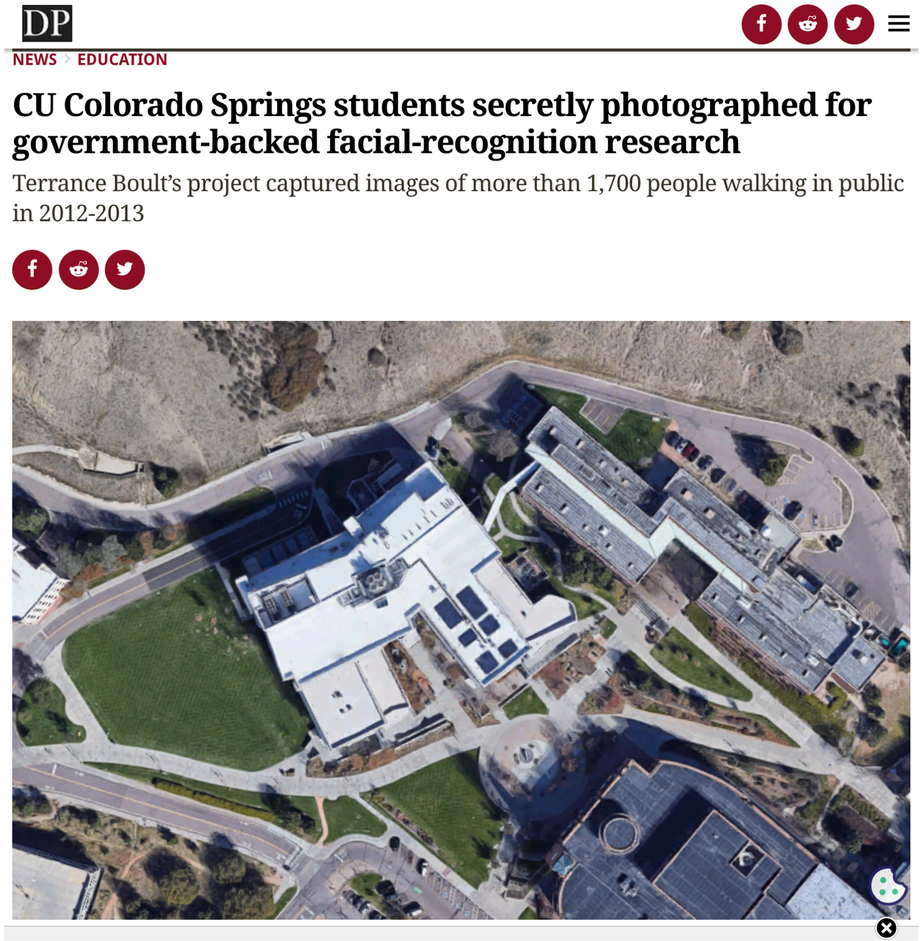
- Consider:
  - The need for a variety of expertise
  - The open nature of public education
  - The historical and structural factors involved in educational systems
  - The importance of policy, governance, and law



Consider the underlying vision

A Proposed Solution

- Ask:
  - What are our underlying goals when using data?
  - Who is benefitting and who is being harmed by how we are using data?
  - How does our work respect the dignity and privacy of people?
  - What counts as data science?



DP

NEWS › EDUCATION

CU Colorado Springs students secretly photographed for government-backed facial-recognition research

Terrance Boult's project captured images of more than 1,700 people walking in public in 2012-2013

Let’s consider two ways of approaching this work

A Proposed Solution

- In this presentation, I claim that data science has a role as both a context for learning and as a research method
  - Perspective 1—*Data Science in Education*: Applying data science methods
  - Perspective 2—*Data Science Education*: Supporting data science learning

Doing educational data science work in a way that has an impact on teaching, learning, and educational systems and is sustained over time requires . . .

a **vision** and an understanding of three distinct domains: education, mathematics and statistics, and computers and computation and an attention to the educational context

a vision and an understanding of three distinct domains: **education**, mathematics and statistics, and computers and computation and an attention to the educational context

a vision and an understanding of three distinct domains: education, **mathematics and statistics**, and computers and computation and an attention to the educational context

a vision and an understanding of three distinct domains: education, mathematics and statistics, and **computers and computation** and an attention to the educational context

a vision and an understanding of three distinct domains: education, mathematics and statistics, and computers and computation and an **attention to the educational context**

Reactions or questions?

Example 1: Data Science in Education

Null Hypothesis Statistical Testing is hard

Example 1: Data Science Education

- It can be challenging for learners at all ages to grasp the logic of Null Hypothesis Statistical Testing (Makar & Rubin, 2018)
  - At the same time, in most K-12 classroom settings, formal inference is not the goal
  - But, making informal inferences in light of variation or uncertainty is
- There are also some known issues with Null Hypothesis Statistical Testing (Wagenmakers, 2007)
- Some have argued for the role of Bayesian methods as a (surprisingly) intuitive and accessible alternate, but such methods have mostly been out of reach

Exploring Bayesian methods

Example 1: Data Science Education

- *Education*: Design guidelines
- *Mathematics and statistics*: Approaches from the “data modeling” literature
- *Computers and computation*: Tools that make Bayesian estimation and inference more accessible
- *A vision*: Learners can do surprising things with the right support
- *Attention to context*: Designing for particular curricular standards and goals

Working to make these methods accessible

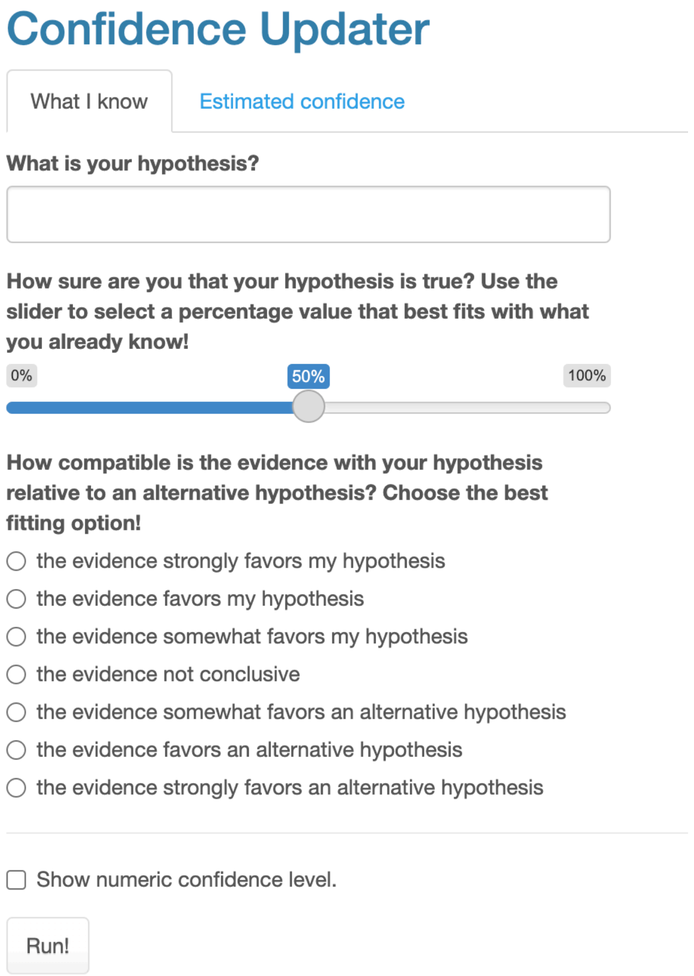
Example 1: Data Science Education

- Research in statistics, child development, and science education supported this approach
- We argued for three ways K-12 teachers and students could use Bayesian methods:
  - Epistemic principles
  - An accessible (Shiny) application
  - Designing games or activities for very young learners

Rosenberg, J. M., Kubsch, M., Wagenmakers, E.-J., & Dogucu, M. (in press). Making sense of uncertainty in the science classroom: A Bayesian approach. Science & Education. https://osf.io/aznyq/

The Confidence Updater App

Example 1: Data Science Education

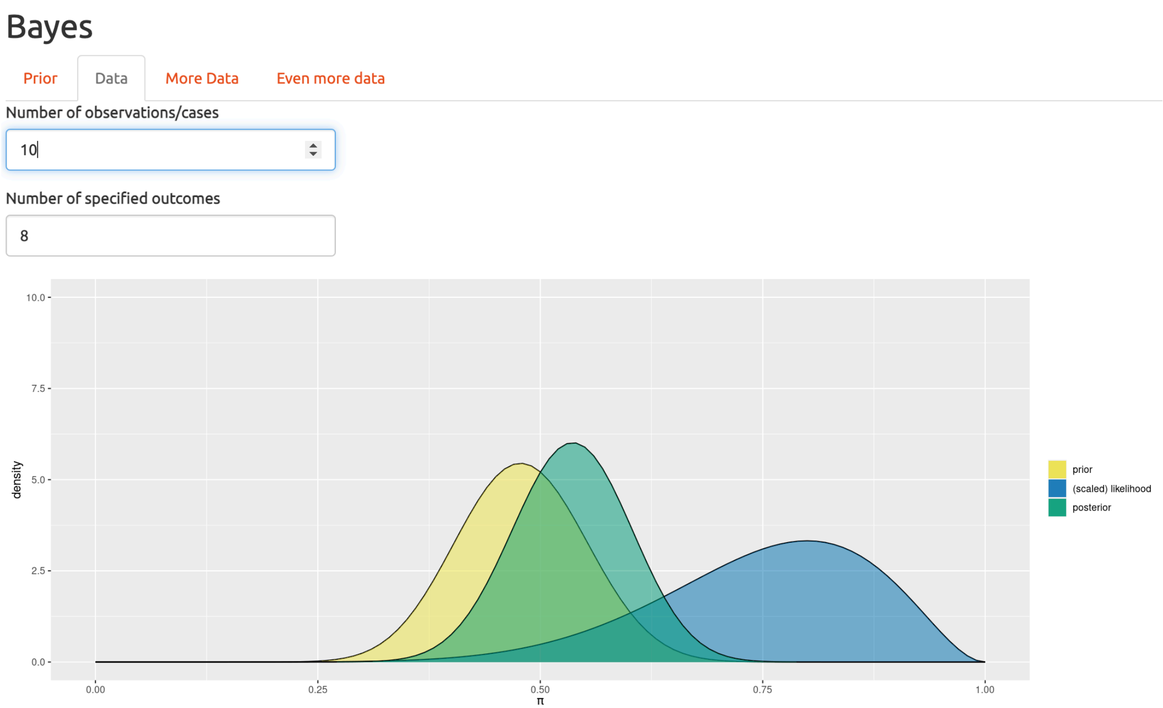


https://kubsch.shinyapps.io/Confidence_Updater/

Next steps around Bayesian methods

Example 1: Data Science Education

- Adapting Data Nuggets for use with the Confidence Updater app
- Developing modules for JASP
- Working pre-service mathematics and science teachers around using Bayesian methods with their students (and working with Mine Dogucu and Sibel Kazak and an app they have developed)

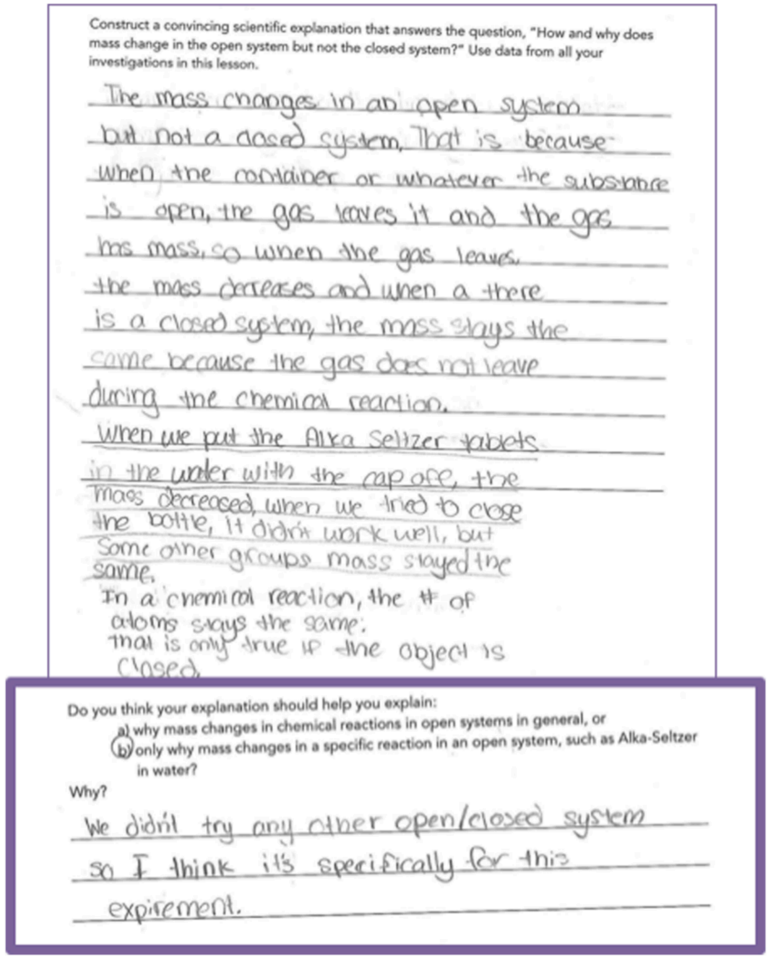


Reactions or questions?

Example 2: Data Science Education

We had a lot of text data for a research project

Example 2: Data Science in Education



Construct a convincing scientific explanation that answers the question, "How and why does mass change in the open system but not the closed system?" Use data from all your investigations in this lesson.

The mass changes in an open system but not a closed system. That is because when the container or whatever the substance is open, the gas leaves it and the gas has mass, so when the gas leaves, the mass decreases and when a there is a closed system, the mass stays the same because the gas does not leave during the chemical reaction. When we put the Alka Seltzer tablets in the water with the cap off, the mass decreased, when we tried to close the bottle, it didn't work well, but some other groups mass stayed the same. In a chemical reaction, the # of atoms stays the same. That is only true if the object is closed.

Do you think your explanation should help you explain:

a) why mass changes in chemical reactions in open systems in general, or

(b) only why mass changes in a specific reaction in an open system, such as Alka-Seltzer in water?

Why?

We didn't try any other open/closed system so I think it's specifically for this experiment.

Combining Machine Learning and Qualitative Methods

Example 1: Data Science Education

- *Education*: Theories of learning
- *Mathematics and statistics*: Latent Semantic Analysis
- *Computers and computation*: Supervised Machine Learning (SVM, SNN)
- *A vision*: Humans and machines have complementary strengths
- *Attention to context*: Knowledge of the data collection

Computational Grounded Theory

Example 2: Data Science in Education

- Adapted the Computational Grounded Theory approach (Nelson, 2020)
- Started with a construct map that could be used to code students' written responses and carried out three steps
    1. Exploratory analysis using latent semantic analysis
    2. Close reading of the written responses with the output from the first step
    3. Validation using supervised machine learning methods
- This approach led to a construct map that was aligned with theory and was better grounded in the data

Rosenberg, J. M., & Krist, C. (2021). Combining machine learning and qualitative methods to elaborate students' ideas about the generality of their model-based explanations. Journal of Science Education and Technology, 30(2), 255-267. https://link.springer.com/article/10.1007%2Fs10956-020-09862-4

Next steps

Example 2: Data Science in Education

- Exploring the usefulness of the approach for audiovisual data — T(CA)^2
- Developing a framework for combining machine learning and qualitative analysis (in science education research; Kubsch et al., 2021)
- Combining qualitative and machine learning methods to explore questions around student privacy (Rosenberg et al., 2022)

Reactions or questions?

Discussion

Summary

Conclusion

- Educational data science has a transformative potential
- But, some are cautious or critical
- Educational data science can make an impact as a research method and a context for learning
- Approaching educational data science from a single perspective may lead to limited impact and sustainability

A conjecture

Conclusion

To have an impact and be sustained over time, this work requires:

- A vision
- An understanding of education
- An understanding of mathematics and statistics
- An understanding of and computers and computation
- An attention to the educational context

What can we do?

Conclusion

- Educational researchers
  - Partner with mathematicians and statisticians with deep subject area expertise and become computationally-literate
- Mathematicians and statisticians
  - Advance mathematical ideas that may be useful in some form, even for long learners
- Computer scientists
  - Consult mathematicians and statisticians
- Others
  - Have a voice in how data science as a discipline develops

Conclusion

(Humbly offered) epistemic virtues

Conclusion

- Knowing a little about the different domains involved with data science
- Being willing to learn
- Avoiding credentialism
- Taking time
- Considering the underlying vision and context

Implications of these virtues

Conclusion

- We should try out a range of approaches to educational data science
- We need some time and support to do this work well
- We should involve a range of stakeholders, including students
- What else?

Thank you and contact

Conclusion

- Joshua Rosenberg, Ph.D.
- https://joshuamrosenberg.com
- https://twitter.com/jrosenberg6432
- Your questions, reactions, and ideas are welcome!