

AI and ML and Data! Oh My!

**Supporting Teachers' and Learners' Work by
Considering the Human Sides of Data Science**

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

Presentation at the LEAD Graduate School and Research Network Retreat 2022-04-27
Slides: <https://joshuamrosenberg.com>

A Story

How I was *planning* to start this talk

A story

machine learning methods to address the constraints each has. Finally, I conclude with cautiously optimistic ways for people with expertise in education, mathematics and statistics, and computers and computation to collaborate to advance data science in K-12 settings.

What should I title a talk with the above abstract?

The Role of Data Science in K-12 Teaching and Learning

That is actually a pretty good title for a talk with the above abstract, but could you create a title that's a bit more interesting or even funny?

How to Use Data Science in K-12 Teaching and Learning (Without Losing Your Mind)

How I was *planning* to start this talk

A story

That's closer to what I had in mind. Could you suggest any alternate titles that are maybe even more funny or thought-provoking?

Data Science in K-12: A Recipe for Success (Or Disaster)

The Benefits (and Pitfalls) of Using Data Science in K-12 Teaching and Learning

How I was *planning* to start this talk

A story

These are definitely on the right track, but they're not quite there. Any last suggestions for me?

The Promise and Peril of Data Science in K-12 Education

Pretty good, once more. How about one with a tad bit more humor, please? Last attempt.

Data Science in K-12: To Infinity and Beyond!

That was missing something I wanted to say

A story

- Before coming, I was a bit anxious!
- At a bakery, I ordered *Dinkel Pur* (right) in German
- Having family nearby helped, too
- But, I have much more to learn



A little educational data science can go a long way

A story

- The thesis of this talk is that ML, AI, and data can make a sustained impact on education
- An implication of this is that **learning a bit** and **having some familiarity** with what these involve can help all of us

But first

A story

- On a scale of 1-10 (please use the digits on your hands), how confident are you in *your knowledge and capabilities* regarding:
 - Mathematics?
 - Science?
 - Statistics?
 - Programming?
 - Education:
 - Educational research?
 - Educational systems?

A brief aside

A story

- Many of you are already involved with these in exciting ways; as just a few examples from the program:
 - Katharina Fleig: *NLP-based learner assessment for feedback generation*
 - Lisa Minich: *The Development of a Teacher Dashboard for the FeedBook System*
 - Plus working groups on *Artificial Intelligence in Education* (Xiaobin Chen) and *analyzing log file data* (Ines Pieronczyk)

Defining some terms

A story

- By using the term, **Artificial Intelligence (AI)**, I refer to simulating human intelligence through the use of computers
- By, **Machine Learning (ML)**, I am referring to a subset of AI focused on how computers acquire new information/knowledge
- By **Data Science**, I am describing analytic approaches that involve statistics and computers/computation

Where is this work taking place?

A story

- The **Learning Analytics** (LA) community focuses primarily but not only on data from learning management systems, especially in higher education settings ([Wise, 2019](#))
- The **Educational Data Mining** (EDM) community focuses primarily but not only on data from educational technology tools, often through a technical lens ([Baker and Inventado, 2016](#))

Scholarship often cuts across these fast-developing fields (e.g., [Baker & Inventado, 2016](#); [Baker & Siemens, 2012](#); [Fischer et al. 2020](#))

Where is this work taking place?

A story

- The community exploring data science *in education* has two foci:
 1. **Educational Data Science** (EDS), focusing on using data science tools to support research ([McFarland et al., 2021](#))
 - “... it is an umbrella for a fleet of new computational techniques being used to identify new forms of data, measures, descriptives, predictions, and experiments in education.” ([McFarland et al. \(2021\)](#))
 2. **Data Science Education** (DSE), exploring data science as a context for learning ([Wilkerson & Polman, 2019](#))

Some logistics

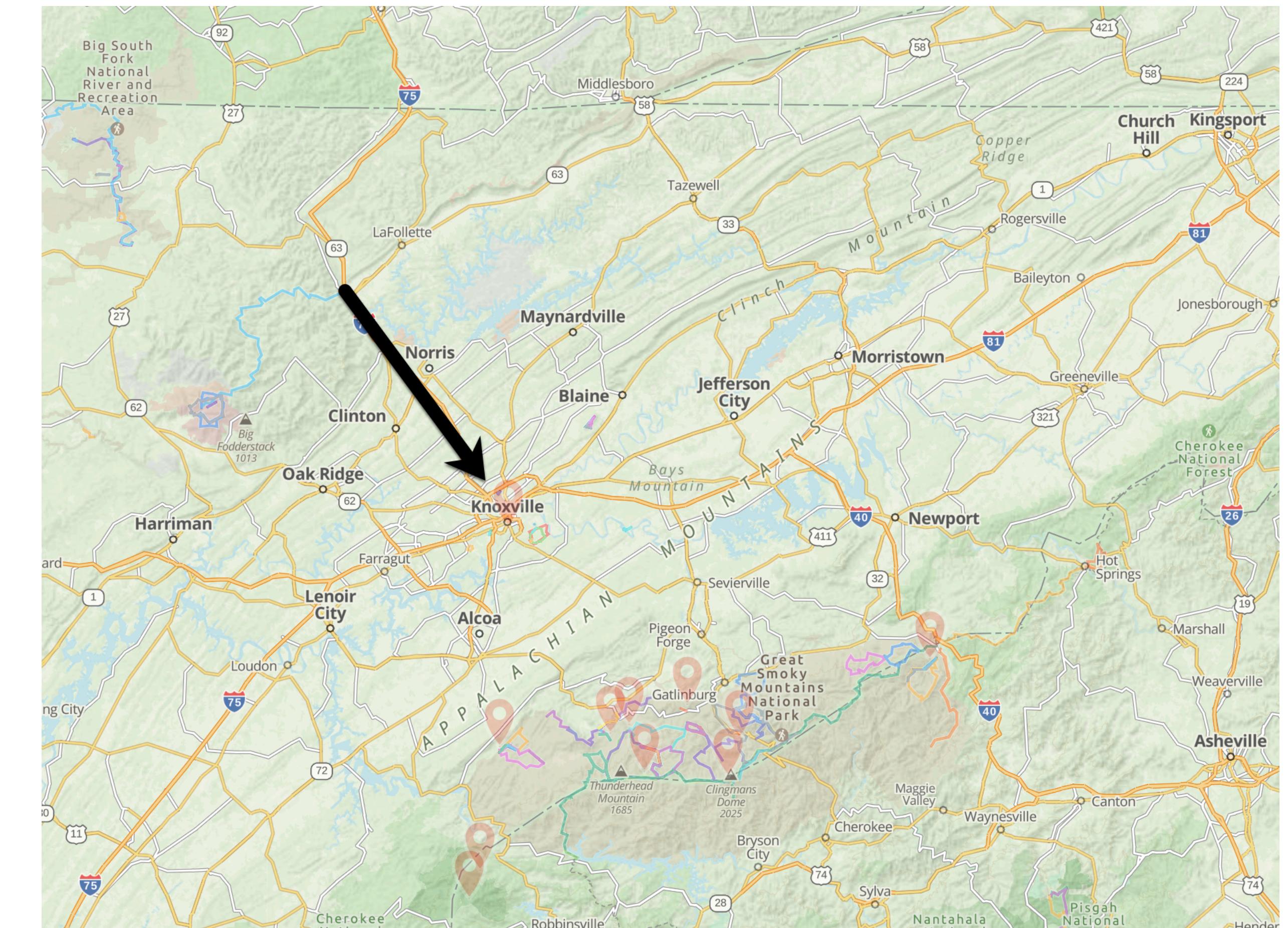
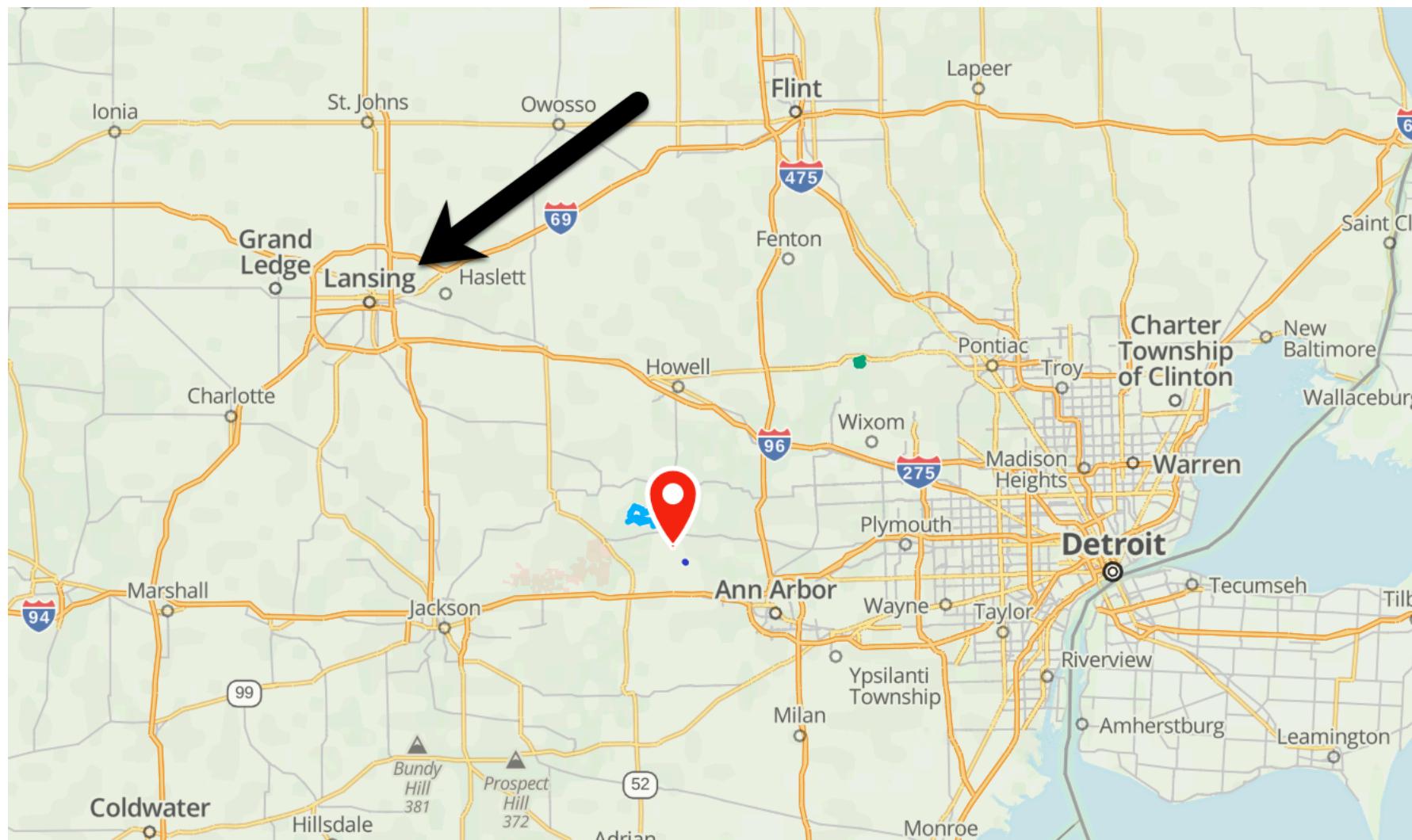
A bit about me

Some logistics



A bit about from where I come

Some logistics



A bit more about me

Some logistics

- I study teaching, learning, and educational systems
- I focus on science education, but I also explore STEM disciplines
- I do so using a range of methods, but my strengths are in educational data science



THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

And how I became interested in these topics

Some logistics

- **2011:** Saw high school students become proud of a graph they created
- **2014:** Learned R in a class on multi-level modeling
- **2015:** Was exposed to educational research on students' *data modeling*
- **2015-Present:** Engaged in cycles of learning, applying, and learning more around the topic of *data science in education*

Goals for today

Some logistics

1. **Articulate a thesis** about the nature of data science in education
2. **Describe some prior research** (mine and others') that speaks to this thesis
3. **Work toward a model** of how humans can use machines for research in a cooperative mode

Thesis

The transformative potential of data science in education

Thesis

- “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*

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

Others have expressed some caution

Thesis

- Is data science really that different from statistics education? ([Rubin, 2019](#))
- How should we think about the factors influencing how students engage with data? ([Lee et al., 2021](#))
- Who benefits from abundant data and computing power ([D'Ignazio & Klein, 2020](#))

An initial attempt at a thesis

Thesis

- Educational data science has a role in education, but **how we do this work matters** in terms of impact sustainability over time
- But, I want to use this retreat to rethink this work and progress toward a model of what this “how” can look like

Why does this matter?

Thesis

- To step back from this thesis, I hope working toward this model can inform the work of the LEAD Graduate School and Research Network (and this retreat)
- I also hope to complement the talks of other scholars—external and members of the network

Prior Research

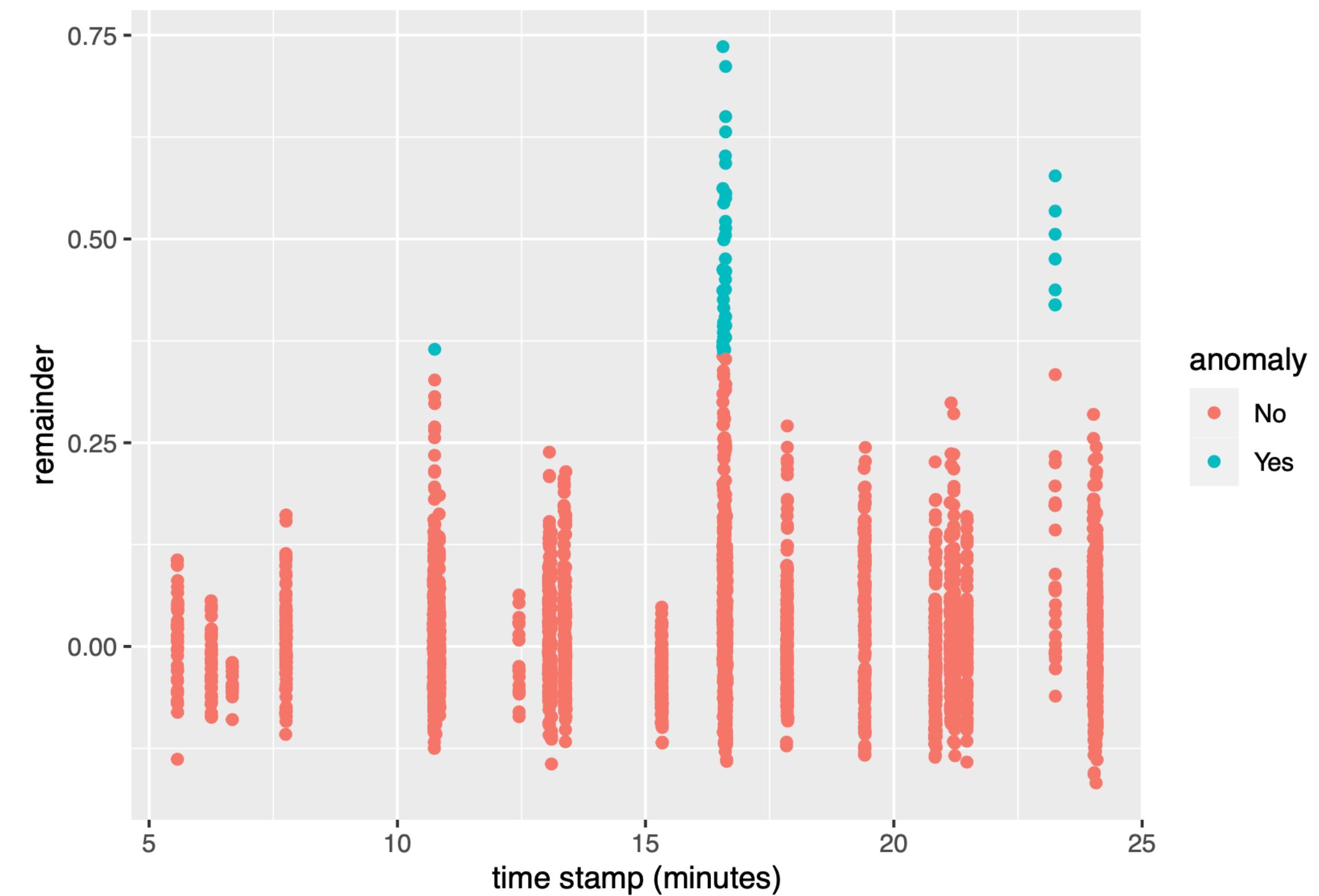
Combining qualitative and machine learning methods

Prior Research

n	Classification algorithm	Percentage agreement	Quadratic Weighted Kappa
169	Naïve Bayes	.60	.47
	Support vector machine	.59	.56
	Sequential neural network	.56	.53
1885	Naïve Bayes	.66	.62
	Support vector machine	.70	.66
	Sequential neural network	.70	.65

Rosenberg & Krist (2021), Combining Machine Learning and Qualitative Methods to Elaborate Students' Ideas About the Generality of their Model-Based Explanations

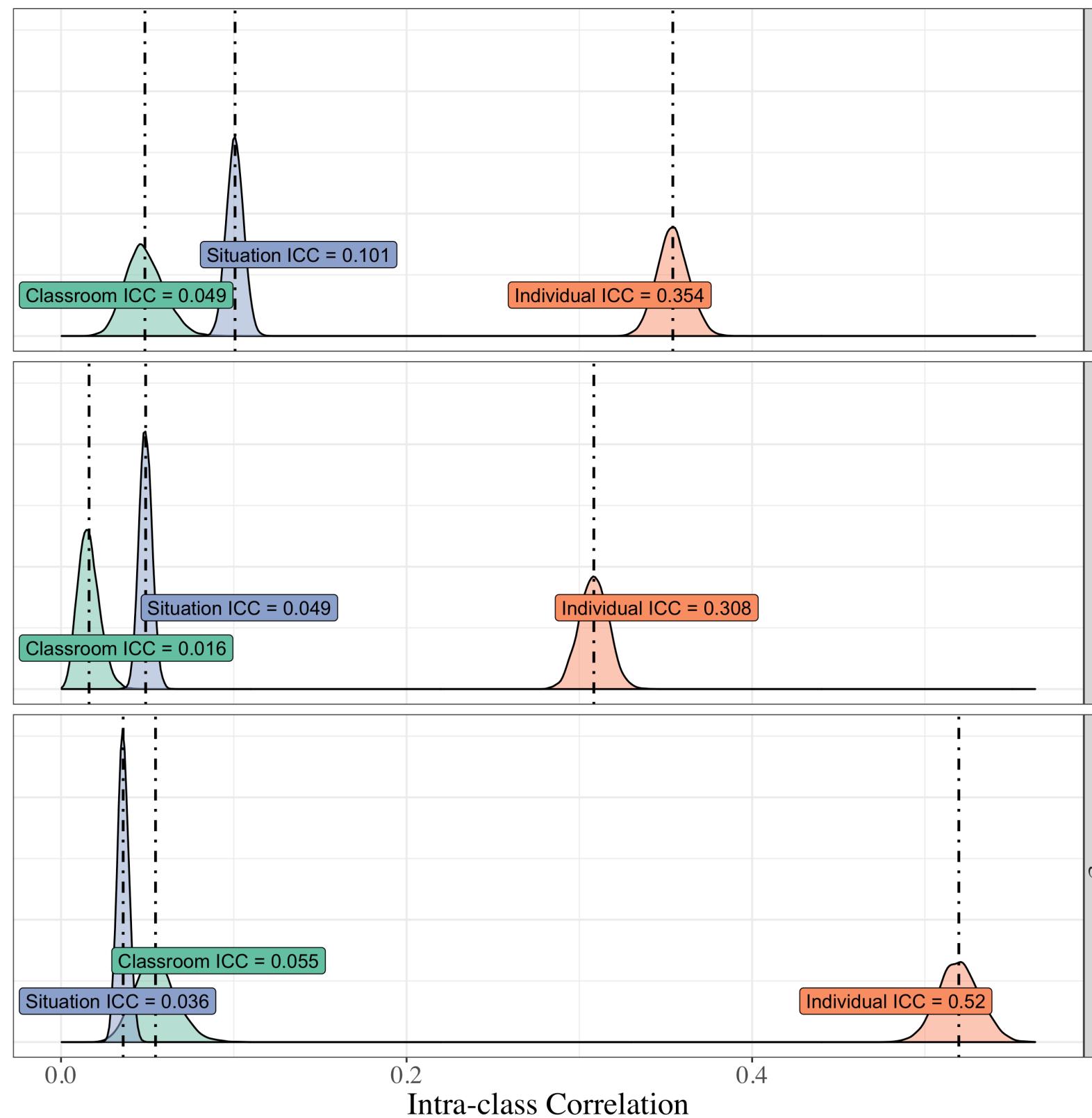
<https://faast.shinyapps.io/generality-shiny/>



Theory-based Computational Analysis of Classroom Audiovisual Data (NSF-supported)

Supporting learners' to use Bayesian methods

Prior Research



Rosenberg et al. (2020), How does situational engagement vary between learners, situations, and classrooms?

Confidence Updater

What I know Estimated confidence

What is your hypothesis?

I think the substance is a carbohydrate.

How sure are you that your hypothesis is true? Use the slider to select a percentage value that best fits with what you already know!

0% 55% 100%

How compatible is the evidence with your hypothesis relative to an alternative hypothesis? Choose the best fitting option!

- the evidence strongly favors my hypothesis
- the evidence favors my hypothesis
- the evidence somewhat favors my hypothesis
- the evidence not conclusive
- the evidence somewhat favors an alternative hypothesis
- the evidence favors an alternative hypothesis
- the evidence strongly favors an alternative hypothesis

Show numeric confidence level.

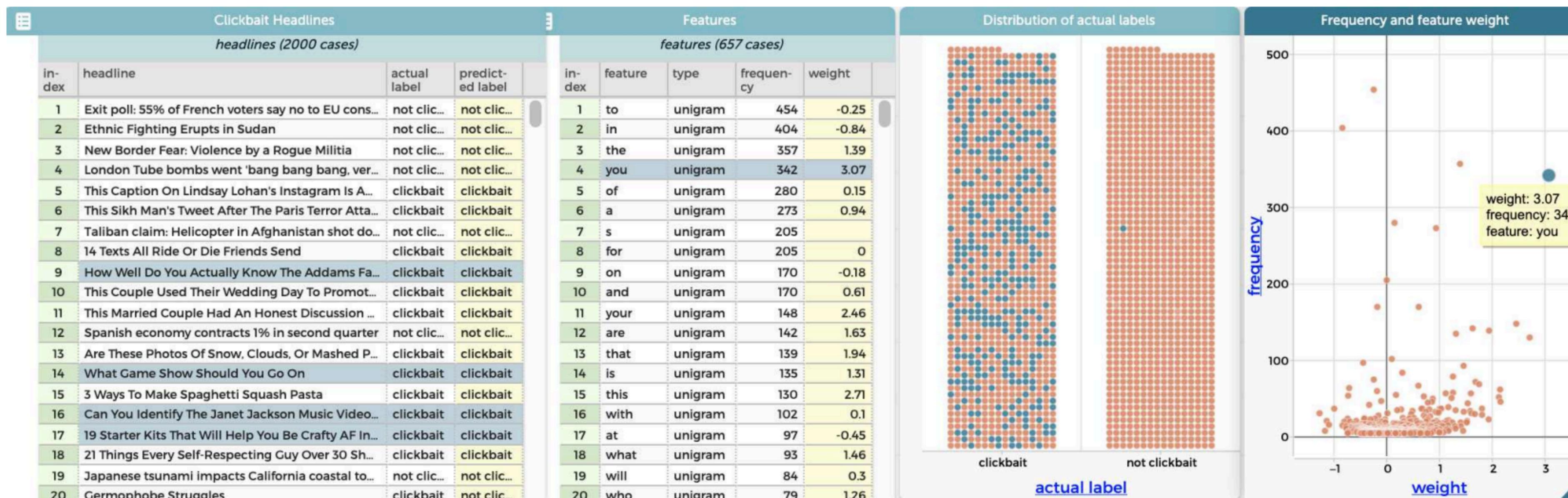
Run!

Rosenberg et al. (2022), Making Sense of Uncertainty in the Science Classroom: A Bayesian Approach

ML and AI Education

Prior Research

- Understanding how students make informal predictions (Konold, 1989, Informal Conceptions of Probability)



Tatar et al. (2021),
Modeling Unstructured
Data: Teachers as
Learners and Designers
of Technology-
enhanced Artificial
Intelligence Curriculum

Data Science and Justice

Prior Research



Rubel et al. (2016),
Teaching Mathematics
for Spatial Justice: An
Investigation of the
Lottery

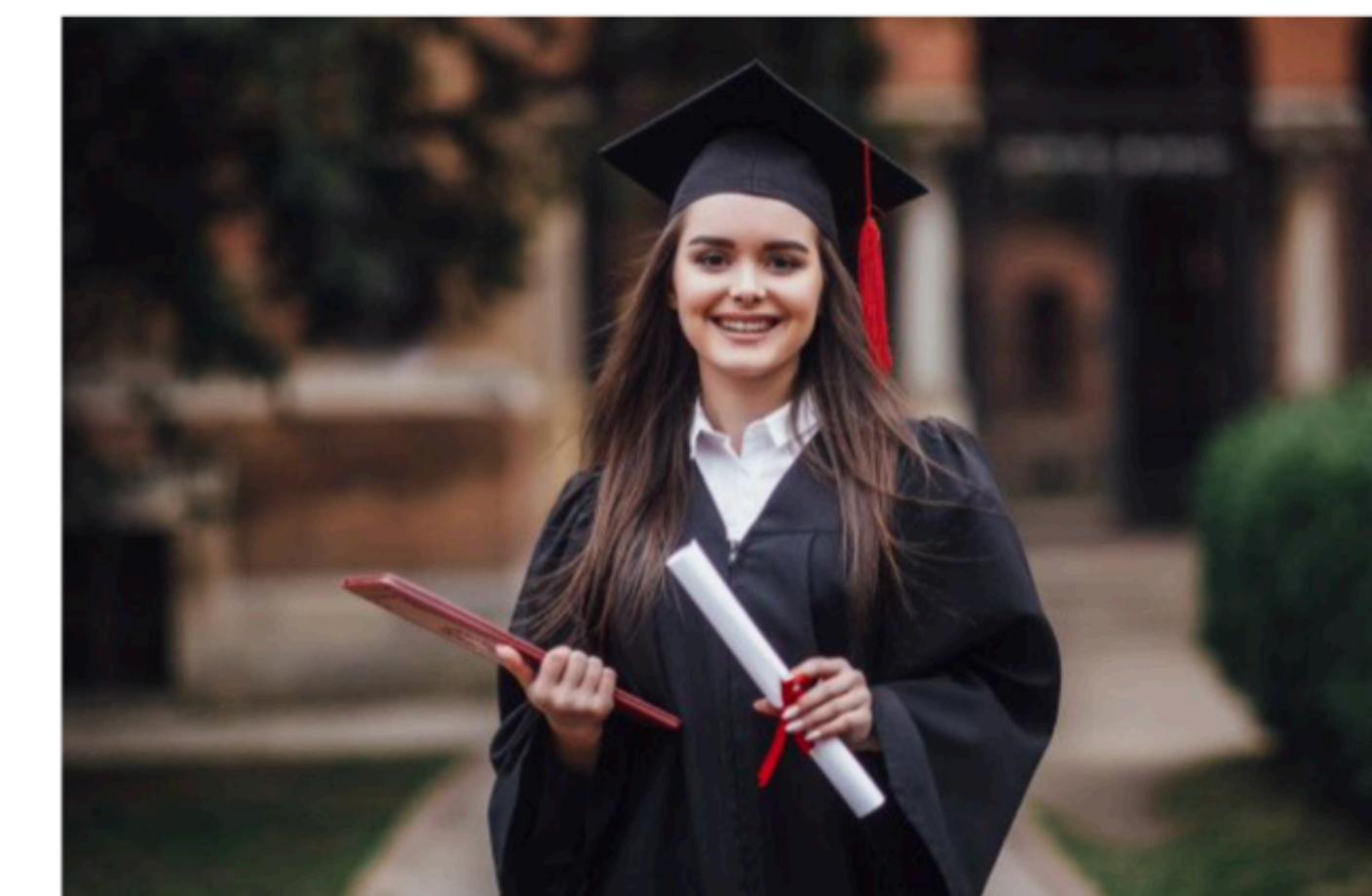
"And notice **it's ranked number one in certain areas, right.** Uh, ok and let's look—so it's orange around Inglewood and here this area between Compton and Carson."

Philip et al. (2016),
Becoming Racially
Literate About Data and
Data-Literate About Race:
Data Visualizations in the
Classroom as a Site of
Racial-Ideological Micro-
Contestations

Science Education and New Forms of Media

Prior Research

- Most participants in a Twitter-based network were involved in substantive conversations (Rosenberg et al., 2020, Idle chatter or compelling conversation?)
- Posts on Twitter about the NGSS were very positive—much more than posts about the CCSS (Rosenberg et al., 2020, The Next Generation Science Standards on Twitter)



Rosenberg et al. (2022), Schools' Posts on Facebook Systematically Risk Students' Privacy

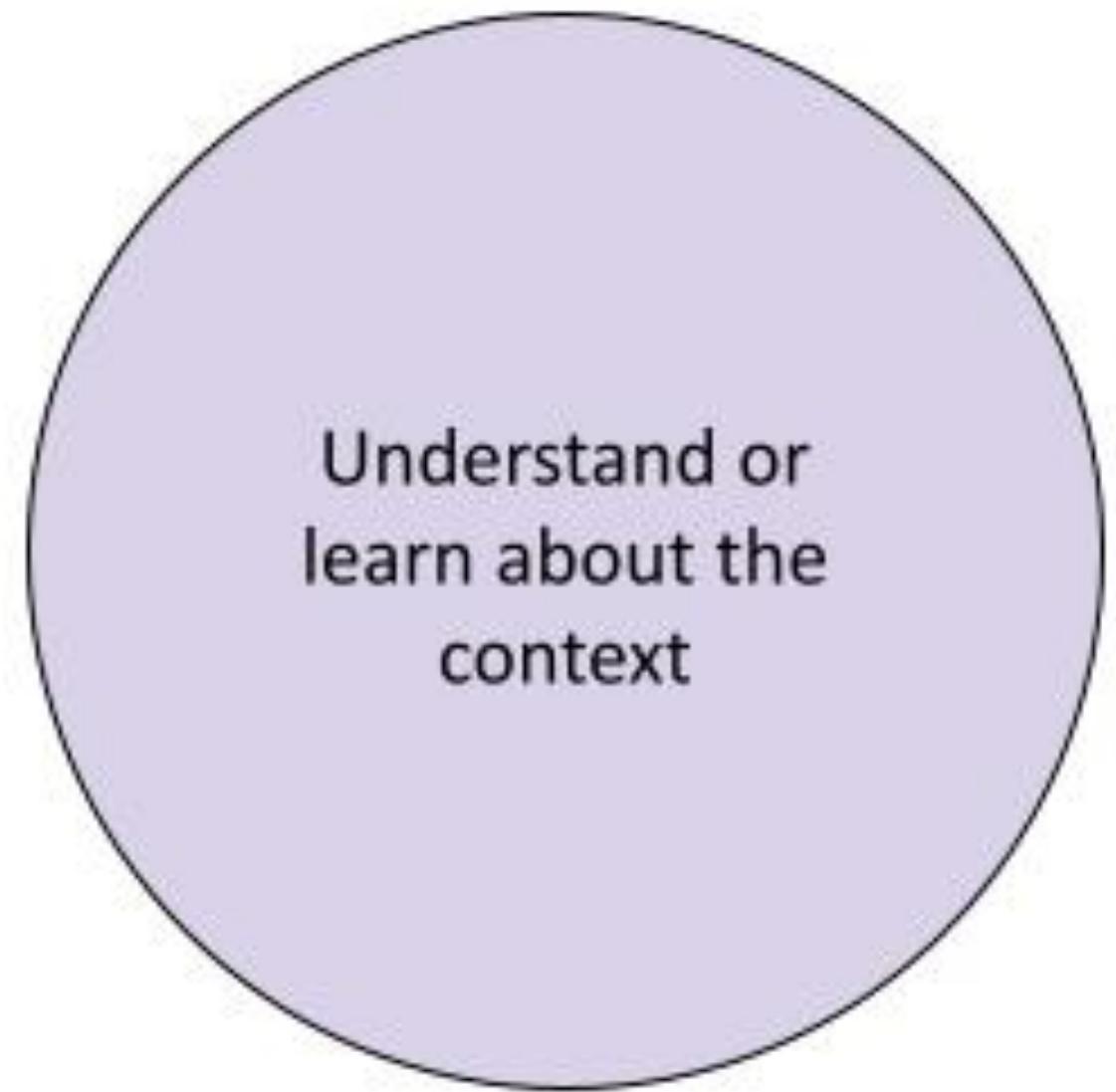
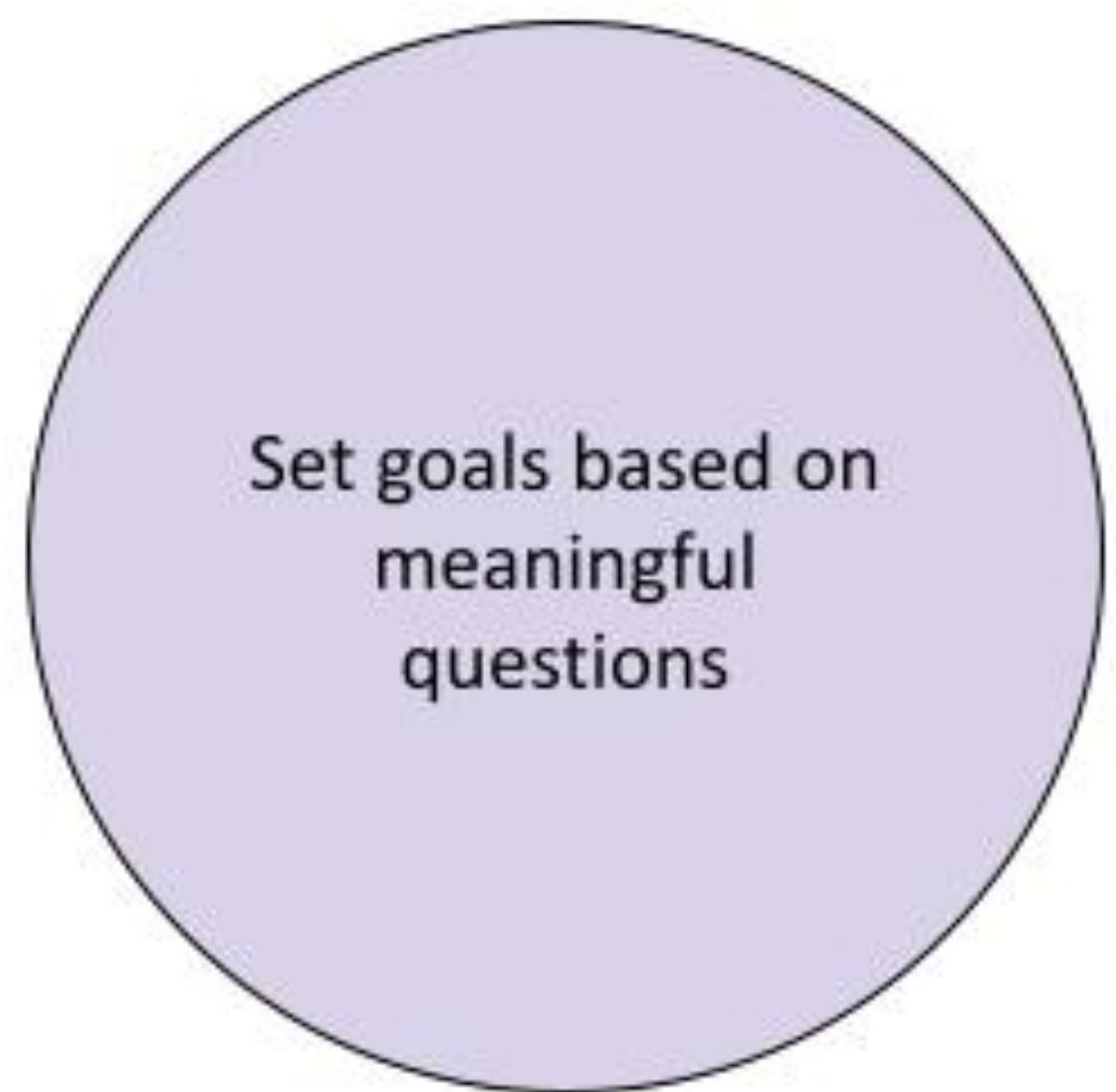
Toward a Model

Back to the thesis

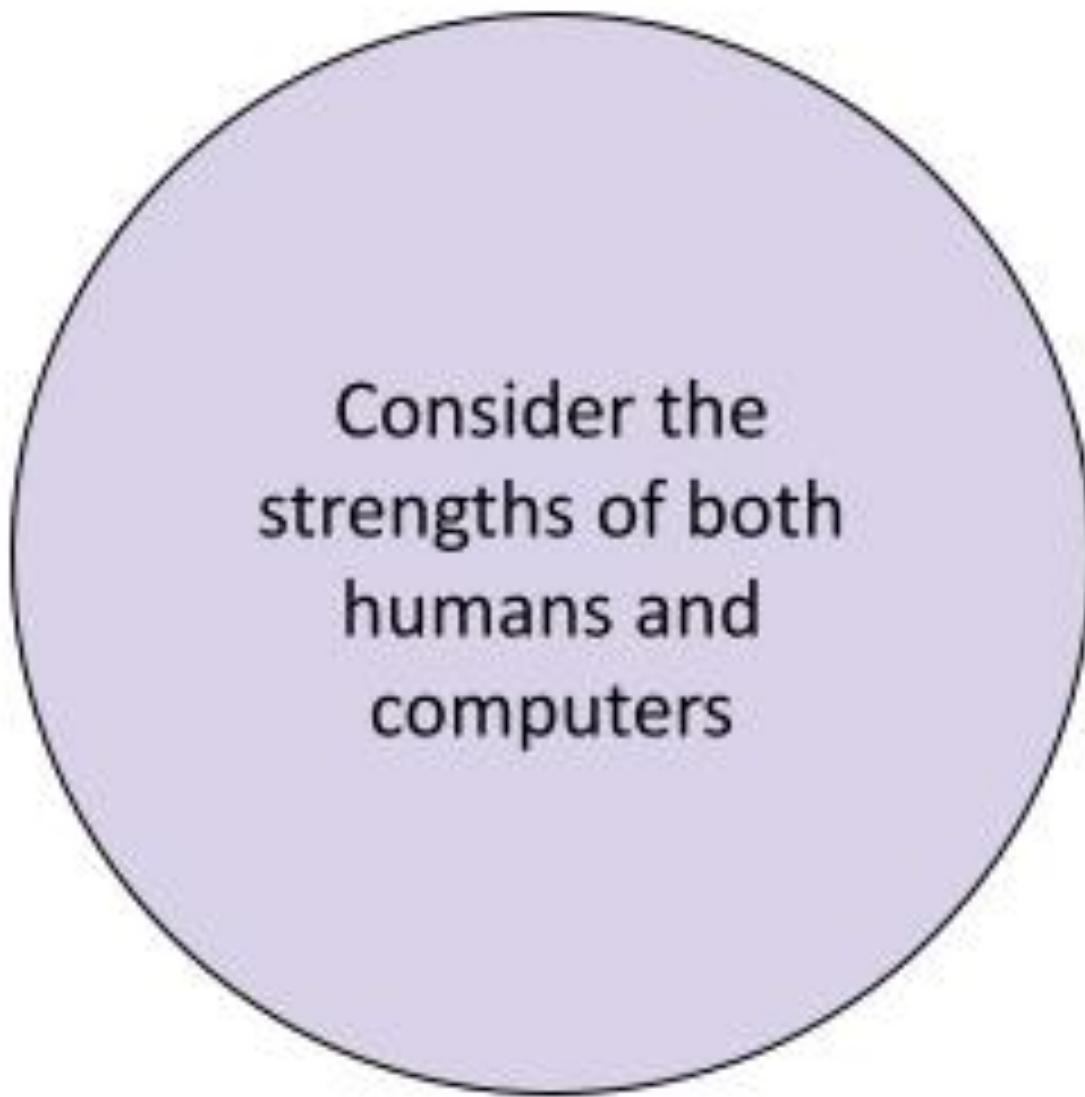
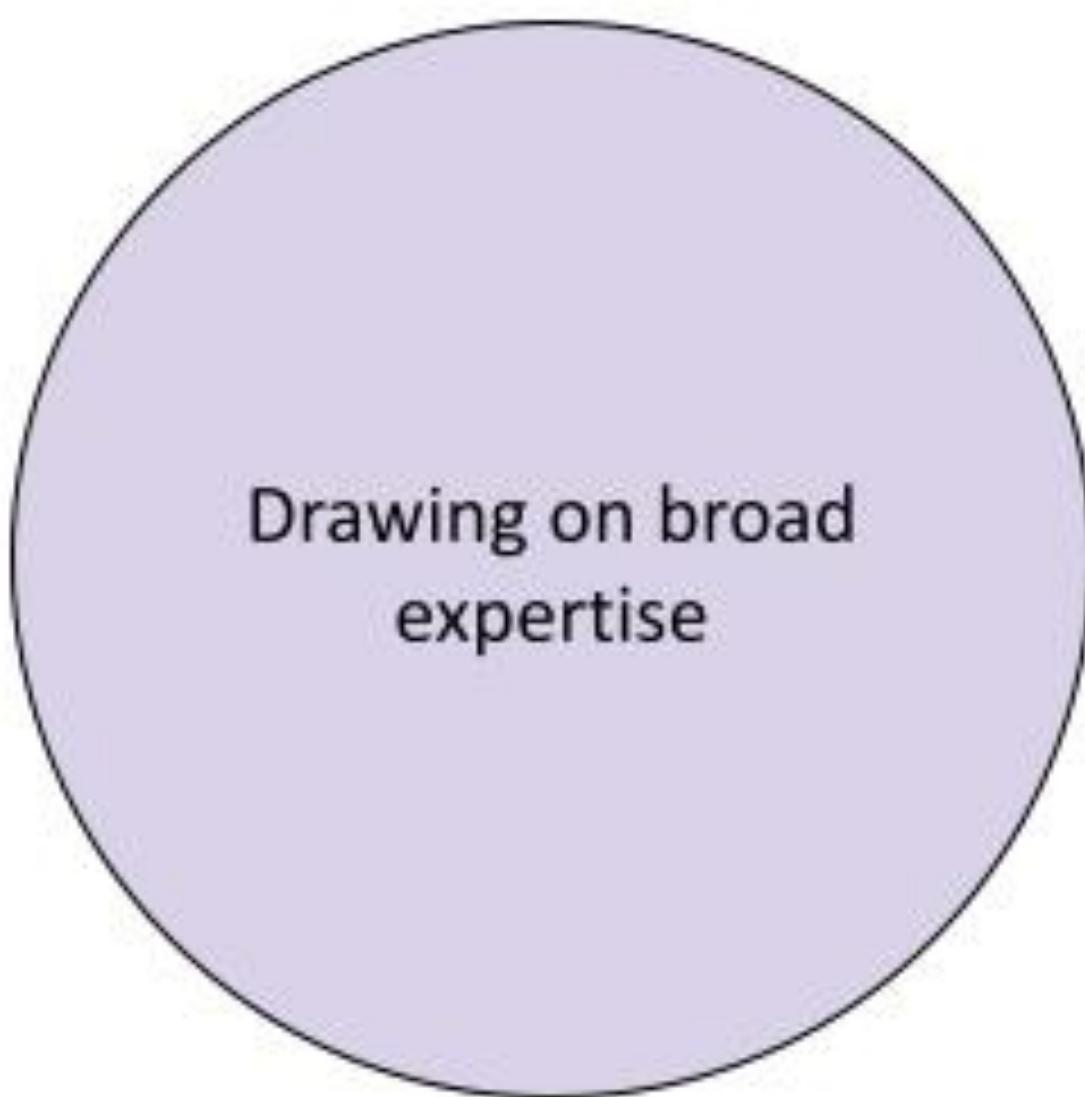
Toward a Model

- Educational data science has a role in education, but how we do this work matters in terms of impact and sustainability over time
- How should research related to data science in education progress?
- Let's try to *model* this (cf. Schwarz et al., 2009)

Understanding of education



Understanding of AI, ML, or data science



Why familiarity can matter

Conclusion

- There is room for humans and computers in our work, but we need to know about both to be able to answer hard questions (Kubsch et al., 2022)
 - What are the tasks?
 - Who or what (human or computer) is doing which tasks?
 - Who or what should be doing which tasks?

Conclusion

Why familiarity can matter

Conclusion

- More of us in education can have an active role in shaping how AI, ML, and data are used
- We can ignore, embrace, or work out *for which tasks under what conditions* we should use these advanced technologies (Kissinger et al., 2022)
- I look forward to learning with and from you about this topic and others over the upcoming days

Please share the questions or remarks you have!

Conclusion

- Thank you kindly to:
 - Alisa Schmid and the LEAD Scientific Coordination Team
 - LEAD members (special thanks to Christian Fischer)
 - LEAD Post-docs, Interns, and Students
 - Those hosting us in Untermarchtal
 - And the baker of the Dinkel Pur
- Contact information:
 - Joshua Rosenberg, Ph.D.
 - <https://joshuamrosenberg.com> (slides posted here)
 - jmrosenberg@utk.edu