



Scatterplot

The MAA Journal of Data Science



ISSN: 2993-2955 (Online) Journal homepage: www.tandfonline.com/journals/usca20

Leveraging Mathematical Knowledge to Prepare Future Math Teachers to Teach Data Science

Eric Weber, Heather Gallivan, Lydia Butters & Stephen Nathan Mercil

To cite this article: Eric Weber, Heather Gallivan, Lydia Butters & Stephen Nathan Mercil (2026) Leveraging Mathematical Knowledge to Prepare Future Math Teachers to Teach Data Science, Scatterplot, 3:1, 2644686, DOI: [10.1080/29932955.2026.2644686](https://doi.org/10.1080/29932955.2026.2644686)

To link to this article: <https://doi.org/10.1080/29932955.2026.2644686>



[View supplementary material](#)



Published online: 08 Apr 2026.



[Submit your article to this journal](#)





[View related articles](#)



[View Crossmark data](#)



Leveraging Mathematical Knowledge to Prepare Future Math Teachers to Teach Data Science

Eric Weber^a , Heather Gallivan^b , Lydia Butters^b, and Stephen Nathan Mercil^a

^aDepartment of Mathematics, Iowa State University, Ames, IA, USA; ^bDepartment of Mathematics, University of Northern Iowa, Cedar Falls, IA, USA

ABSTRACT

Multiple professional societies – including mathematics, statistics, and mathematics educators – have recently released statements in support of teaching data science (DS) at the secondary level. Such statements implicitly (if not explicitly) seek preparation of teachers to deliver DS content. We present our experiences in developing, delivering, and refining curriculum which endeavors to prepare pre-service mathematics teachers to deliver this content. Our curriculum takes the form of a 5-week self-contained module that is delivered within coursework taken by pre-service mathematics teachers at the University of Northern Iowa and Iowa State University. We describe this curriculum, together with our inspiration and sources for this curriculum, as well as the design choices we made in its preparation. After offering the curriculum over 3 iterations – each of which were refinements of prior implementations – we have arrived at two themes that we think are foundational to preparing mathematics teachers to teach DS. These two themes are: 1) we can leverage future mathematics teachers' prior mathematical knowledge to master DS concepts, and 2) we can illustrate challenges to DS methodology through a mathematical framework. We conclude with a description and rationale for those two themes.

ARTICLE HISTORY

Received 18 September 2025
Revised 18 February 2026
Accepted 22 February 2026

KEYWORDS

Data science curriculum;
teacher preparation;
mathematics teachers;
knowledge for teaching

1. Introduction


Data science (DS) is ubiquitous, in many forms and in many contexts. While not necessarily new (we'll discuss a bit about the history of the subject), DS ideas, techniques, and consequences now permeate much of everyday life. The mathematics and statistics disciplines advocate for “data literacy” or “quantitative literacy” at all levels of academic preparation – and have done so for decades – yet these calls assume significantly more urgency in our nascent era of readily available artificial intelligence tools.

Widespread calls to teach DS content at the secondary levels appeared just in the last few years [1–3], though efforts to design DS curriculum at the secondary level have been ongoing for longer [4–7]. However, efforts to prepare secondary teachers to teach DS have lagged – we published such a call recently [8], focusing our position specifically on secondary mathematics teachers. Our contention is that current and future mathematics teachers are best positioned to teach DS at the secondary level.

This article discusses our efforts to prepare future (pre-service) mathematics teachers, or PSTs, to teach DS. Succinctly, we developed a 5-week curriculum module that exposes mathematics PSTs to elementary DS concepts. The curriculum was delivered to mathematics PSTs in content courses at both the University of Northern Iowa (UNI) and Iowa State University (ISU) over multiple years. The cohorts of students enrolled in the two universities have several differences in characteristics. All are mathematics/mathematics education majors; however, UNI students were typically in their first or second years, whereas ISU students were typically in their third. At UNI they were enrolled in a statistics and DS content course for teachers and at ISU they were enrolled in a topics course. While we assume no prerequisite knowledge for participating in the module, students that have had vector calculus and linear algebra are better prepared to engage with the content.

The collaboration for developing the curriculum consisted of the named authors on this article together with Jennifer Newman (Professor Emerita of

CONTACT Eric Weber  esweber@iastate.edu  Department of Mathematics, Iowa State University, Ames, IA, USA.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/29932955.2026.2644686>.

© 2026 The Mathematical Association of America

Mathematics at Iowa State University) and Sarah Diesburg (Associate Professor of Computer Science at the UNI). The collaboration for delivering the curriculum consisted of the named authors together with Alejandro Andreotti, and Heather Bolles. We also benefited from the input of several undergraduate research assistants: Meredith Tucker, Ryan Friedhoff, and Ian Tunicliff.

The article is organized as follows:

- In [Section 2](#), we first state for the record our answer to the question “What is Data Science?”. We then present the history and the sources of inspiration for the design goals and learning objectives ([Table 1](#)) of the curriculum. We also address choices that we made during the initial development of the curriculum.
- In [Section 3](#), we outline the current form of the curriculum, as well as highlight some of the changes that were made as successive iterations of the curriculum were delivered. We discuss briefly our usage of technology learning aids in the lessons. (The corresponding lesson plans and technology aids are available in [Appendices B and C](#).) We also briefly mention the outcome of a study conducted during our first iteration in which we administered a pre-post test to measure the impact of the module and discuss our efforts to conduct future research on the scholarship of teaching DS.
- In [Section 4](#), we describe the overarching *themes* that we have identified as a result of several iterations of the curriculum offering. These themes are: 1) we can leverage prior mathematics content knowledge by relating DS objects to mathematical objects; 2) we can illustrate grand societal challenges with prior mathematical exposure and terminology. For example, we can leverage prior understanding of: (i) functions to illustrate the characteristics of models, and (ii) optimization to contextualize trustworthiness of machine learning (ML) algorithms, respectively. These themes will inform our ultimate goal of creating the framework of preparing secondary teachers of DS.

2. Initial major design goals

2.1. What is data science?

Let us begin by dispensing with the question that often confronts us as we endeavor to prepare mathematics PSTs, present our findings, or discuss our project with various stakeholders. The phrase “Data Science” is relatively new, with initial references to the phrase appearing in the mid-1990s [9,10]. The phrase became commonplace in the late 2000s. We can conceptualize DS as an emerging academic discipline, with programs – and several departments – now existing at institutions of higher education across North America and Europe. The first lesson of our 5-week module consists of a lecture entitled “What is Data Science?” that Eric Weber has been presenting to audiences in various venues – typically undergraduate students in multiple majors – for several years. The slides from recent versions of this presentation are available in [Appendix A \(Supplementary appendix\)](#).

Many of the ideas in DS predate the initial usage of the phrase, and many scientific contributors advanced the methodology. One slide from the first lesson of our module is depicted in [Figure 1](#), in which we present a reverse chronological history of ideas that led to what we now often refer to as DS, highlighting the “prior art” of DS to its formal reference. The slide also depicts a Venn diagram of phrases that regularly appear in related contexts: DS, artificial intelligence (AI), and machine learning (ML). The diagram counteracts the conflation of these phrases, as they are frequently used interchangeably but in fact are distinct objects that are interrelated.

To provide broader context, we provide a few working definitions of DS as well as our own conception of DS, particularly within the context of mathematics. Johnstone and Robert (2014) state: “We define the processes that plan for, acquire, manage, analyze, and infer from data collectively as the data science lifecycle” [11] puts it more succinctly: “The would-be notion takes DS as the science of learning from data...” (p. 745). Others depict the notion of DS through figures, including a Venn diagram [12]

Table 1. Learning objectives for the data science module for pre-service teachers.

A. Visualization and exploration	1) Understand the problem of classification of data points in space 2) Utilize projections for visualization and classification 3) Identify gaps in data
B. Prediction	1) Design methods to predict classification 2) Evaluate performance of prediction 3) Identify assumptions and biases in methodology
C. Communication	1) Communicate solutions to prediction problems 2) Communicate effectively methodology and sources of errors
D. Optimization	1) Understand types and components of optimization problems 2) Translate prediction problems into optimization problems

What is Data Science?

- Data-Driven Discovery (2020)
- Data-Driven Decisions (2020)
- Data Science (2010's)
- Data Analytics (2000's)
- Big Data (1990's)
- Data Mining (1980's)
- Artificial Intelligence (1960's)
- Machine Learning (1940's)
- Statistics (1880's)
- Mathematics (5000 BC)

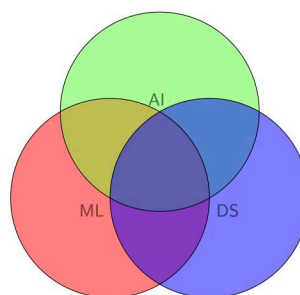


Figure 1. Slide from “What is Data Science?” Presentation introducing the module to Mathematics PSTs.

and a process cycle [13]. Our high-level view of DS is similar—we describe DS as the reversal of the scientific method to our students. By this, we mean that often, in a DS context, the process begins with the data and a hypothesis or research question is reverse-engineered. At a lower level, we scaffold DS over prior mathematical knowledge, which is reflected in our choice of learning objectives and lesson plans.

2.2. Timeline for curriculum design and implementation

Though our current working definition of DS as just stated informs our curriculum design and delivery, our original approach to DS for mathematics PSTs consisted of an amalgamation of several sources—sources that were significantly more algorithmic or technical in philosophy. We will highlight those sources in our description. Moreover, our curricula offerings occurred independently, prior to our ongoing collaboration. We describe the curricula of our two offerings that existed separately from each other as well as the output of our merged curriculum. After we relay the evolution of the curriculum over several iterations, we will then share the lessons we have learned in the form of identification of major “themes” which combine both rigorous mathematical content and our working definition of DS. These themes will be developed in Section 3.

At ISU, Eric Weber led a group consisting of Stephen Nathan Mercil, Jennifer Newman, and Alejandro Andreotti to design the curriculum in fall 2019 semester for a course that was offered during the spring 2020 semester. Dr. Andreotti was the instructor of the course; however, Stephen Nathan Mercil delivered the content to the class for the module. The module was not delivered in spring 2021. At the UNI, the mathematics department was in need of a new

course in probability and statistics for the mathematics teaching major. Heather Gallivan designed the course and implemented it for the first time during the spring 2021 semester, using materials designed by Heather Gallivan as well as those from the ESTEEM project [14]. During the spring 2022 semester, Heather Gallivan piloted the MODULE(S2) statistics curriculum [15] and has used this curriculum, in part, to varying degrees in each subsequent spring semester the course was offered.

During the fall 2021 semester, Heather Gallivan and Eric Weber were independently invited to join a large group of researchers in DS to submit a grant proposal to the National Science Foundation. While that proposal was not funded, Heather and Eric subsequently submitted a curriculum development proposal to the Iowa Space Grant Consortium (ISGC) during the summer of 2022. The ISGC proposal was funded during fall 2022 semester, at which point the curricula from ISU and UNI were integrated and delivered during the spring 2023 semester. We applied for continued funding during summer 2023 which was awarded, and the curricula was delivered again in spring 2024 and spring 2025 semesters in updated form. As of the writing of this article, the curriculum is continuing to be improved and offered in future spring semesters.

2.3. First iteration of our curriculum design

The curriculum initially written at ISU was informed and inspired by multiple sources in DS research and training at the graduate level. A team consisting of Dr. Alex Andreotti, Stephen Nathan Mercil, Dr. Jennifer Newman and Eric Weber met regularly over the course of a semester to conceptualize the module. Early in the process, the team’s discussions focused on learning objectives to be met by specific lesson plans that were drafted later in the process.

The first slate of learning objectives originated from the Midwest Big Data Summer School that transpired at ISU in the years 2016–2019 [16]. The main topics of the Summer School consisted of data acquisition and curation, visualization, generation of elementary ML models (*via* optimization), and mathematical concepts in support of those topics. The learning objectives that now inform our curriculum are presented in Table 1, which evolved from the earlier slate together with an emphasis on mathematical content as we will explain presently.

The first drafts of the lessons centered on synthetic datasets that simulated animal tracking (we provide one such lesson in Appendix B). We also incorporated real data as will be described shortly. The full data sets consisted of timestamps, 2-dimensional locations, and velocities given in the form of speed and heading. Lessons included multiple data sets which progressed by adding covariates—in earlier lessons, the data set contained fewer of the covariates; in later lessons all covariates were provided. The purpose of each lesson was to predict the trajectory of the animals (sometimes the number of animals was provided, and sometimes not). This synthetic dataset mimicked the real dataset provided to Stephen Nathan Mercil and Eric Weber as part of a funded research project [17,18]. More information about that dataset, including our analysis, can be found in the repository [19].

The animal tracking formulation provided us with the context to present material on visualization, projections, optimization, and algorithm design. Since the module appeared in a mathematics course for mathematics PSTs, our emphasis was on mathematical concepts which serve DS techniques. For example, we utilized one version of the animal tracking data set with timestamps and locations (so the data points were three-dimensional). Students were tasked with creating visualizations of the data under two-dimensional projections, then to make inferences informed by those visualizations. (A completed version of this lesson is included in Appendix B.) We chose to use the synthetic data to cleanly illustrate the mathematics we emphasized while also providing a more tangible context for the problem the data presents—one instructor artfully referred to the tracking data as “cowpies in a pasture”. Moreover, through game-play of “connect-the-dots”, students are motivated to think about how to get a computer algorithm to make associations through computations that humans make visually.

In addition to the pedagogical advantages to the animal tracking formulation just highlighted, object

tracking poses a significant challenge encountered in multiple contexts. Science, engineering, and military applications all encounter the challenge of object tracking. For example, our own inspiration for the animal tracking came from a challenge problem funded by the National Science Foundation and the National Geospatial Intelligence Agency to track substantial numbers of ocean vessels. One of our favorite examples of the premise of DS, and perhaps one of the earliest in history, is Johannes Kepler’s derivation of his laws of planetary motion. He observed the locations of planets (requiring him to track them), then derived (*learned*) the laws from the data.

The curriculum initially written at the UNI focused on probability and statistics content. Generally, this included exploratory data analysis, statistical association (including linear and multiple regression of quantitative and categorical variables), and statistical inference (including the Central Limit Theorem, confidence intervals, and hypothesis testing). The materials used have come from multiple sources including the MODULE(S2) statistics materials [20], the ESTEEM materials [14], and lessons and activities designed by Heather Gallivan. Incorporating DS content first occurred during the spring 2023 semester after collaboration started with the team at ISU.

Our merged curriculum included a heavily modified version of the statistical association material from the MODULE(S2) curriculum [20]. Our variations lead directly into multiple concepts that we think are crucial for mathematics PSTs. First, and foremost, we emphasize the concept of a model in DS—while simultaneously bridging the concept with that of a function from mathematics—which primes the students to associate DS objects as being founded upon familiar mathematical objects (e.g. a linear regression model is constructed from a linear function), even though we likely interact with those objects in different manners. We also incorporated multiple linear regression material from UNI, further developing the familiar notion of functions as DS objects. In later lessons on linear/multilinear regression, we had students model property valuations using (real) data obtained from public records in Waterloo and Cedar Falls, IA.

Table 1 presents the learning objectives of our curriculum in its merged form. The objectives and the chosen content align as a result of contemporaneous development, each informing the other as development transpired. We have since made several changes to the content, our implementation of the learning objectives, as well as our higher-order objectives—

themes—through multiple iterations of offerings. We will expand on that evolution in [Section 3](#).

2.4. Comments on design choices in the curriculum

We made numerous design choices that legitimately raise obvious questions or objections. We would like to address some of those here. Our choices reflect both constraints imposed by the environment within which the curriculum is (or will be) implemented as well as our primary motivations. Because of these factors, the curriculum is not comprehensive in presenting DS content.

Our curriculum is presented over a period of 5 weeks within the context of a (single) topics course in mathematics for PSTs. We could easily envision a full development of DS for mathematics PSTs requiring 3 semester-long courses (though this is neither research grounded nor a recommendation). We view these courses as distinct from any required or recommended courses in probability and statistics – the *Statistical Education of Teachers* (SET) report recommends 3 courses in probability and statistics for well-prepared teachers, for example [21]. The content of such a 3 semester sequence would cover content only; methods courses in DS (which are being developed by the ESTEEM II project, for example); [22] would add additional coursework.

As our target demographic is mathematics PSTs, our curriculum intentionally prioritizes mathematical content, concepts, and methods. We fully endorse calls to prepare secondary mathematics teachers to teach DS [23,24], yet we are sensitive to the needs of our PSTs as circumstances currently exist, namely, that they need to receive the content needed for licensure and—ultimately—success in the field. As we will demonstrate in [Sections 3](#) and [4](#), the mathematical content meets these needs *and* facilitates PSTs subsequent mastery in DS content.

Our curriculum further aligns with exposing PSTs to DS content with which they likely have prior familiarity through their mathematics coursework. Our PSTs explore overarching DS concepts without focusing on skills and procedures (e.g. coding, data management, etc.) but rather on the foundational mathematical content. We target larger DS concepts of prediction and classification through the emphasis that developing predictive models is mathematical in nature (i.e. a model is really a function as a mathematical object).

During the several iterations of our curriculum, we had discussions focused on whether we should present content and lessons that the PSTs could take directly with them into their own classrooms. Our choice reflects two rationales: 1) existing curricula is readily (and freely) available for their usage including: Introduction to DS [5]; International DS in Schools Project; (IDSSP, 2019); and Bootstrap: DS (Bootstrap Community, 2017); and 2) our content contains elements that we would identify as *horizon knowledge* [25], an important dimension of teacher preparation. Horizon knowledge refers to the content knowledge teachers should master beyond the content that they themselves will teach.

These choices leave much behind. At a disciplinary level, since DS can be conceptualized as a merging of mathematics, statistics, and computer science, our curriculum touches briefly on statistics and excludes computer science entirely. Thematically, we do not present the DS lifecycle [26,27]. Topically, we do not discuss data management (wrangling, cleaning, archiving, privacy), techniques (ML, neural networks, clustering), or technology (coding, scripting, databases, etc.) Most of these important topics in DS are included in the common curricula available online (e.g., Introduction to DS) [5].

Despite these difficult choices—and the other constraints within which our module was deployed—our curriculum is consistent with national standards in mathematics education. For example, the *Committee on the Undergraduate Program in Mathematics (CUPM) Guide to Majors in the Mathematical Sciences* [28] includes a list of “Mathematical Outcomes” for statistics education—the following are entries within that list which map to several of our learning objectives (as labeled in [Table 1](#)):

- *Experience using technology to explore statistical concepts and to analyze data graphically, numerically, and inferentially* A.1, A.2, A.3, B.1, B.2, C.2.
- *The knowledge of deciding which statistical methods to use in which situations and the ability to check necessary conditions for those methods to be valid* B.1, B.2, B.3, D.1, D.2.
- *Extensive experience with interpreting results of statistical analyses and communicating conclusions effectively, all in the context of the research question at hand* C.1, C.2.
- *An awareness of the power and scope of statistical thinking ... in everyday life* A.2, B.3 (p. 30).

In addition, the CUPM [28] itemizes recommendations from the GAISE report [29], all of which we

follow in our curriculum design. Our classroom practices also utilize recommendations from the *MAA Instructional Practices Guide* (IP Guide), including (CP.1.5) *collaborative learning strategies*, (AP.4) *assessments that promote student communication*, (XT.4.1) *technology and exploratory activities*, and (XT.4.3) *technology as a tool* [30]. We wish to particularly highlight (CP.2.9) *procedural fluency emerges from conceptual understanding*. Indeed, we will describe a major theme emerging from our development in Section 3 that reflects this understanding.

3. Current form of the curriculum

3.1. Overview of the curriculum

The 5-week module consists of 8 lessons covering the main topics of linear regression, classification in the form of animal tracking, and optimization. Here we outline the 8 lessons with very brief descriptions of content, after which we will provide substantially greater detail of the lessons by category.

1. What is Data Science: history of DS; definitions; ML and artificial intelligence.
2. Linear regression: modeling housing data—linear models; correlation coefficient; r^2 ; graphing with technology; model fit.
3. Data visualization & exploration: animal tracking data—visual inspection of data; ill-posed problems; initial algorithmic thinking; position data only.
4. Data visualization II: animal tracking data—position and time attributes; visualizing different coordinates; projections for dimensionality reduction.
5. Data visualization III & Prediction: animal tracking data—position, time, and heading attributes; algorithmic thinking; inferences/predictions from more covariates.
6. Data visualization IV & Prediction: animal tracking data—missing data; data gaps; barriers to inferences.
7. Optimization: traveling salesman problem—algorithmic thinking, greedy algorithms; run-time tradeoffs.
8. Optimization II: animal tracking data—predictions using optimization routines; choice of metric; use of projections to calculate distances.

We have altered the lessons over several iterations. One substantial change involved adding the traveling salesman problem to help scaffold the use of optimization in later animal tracking lessons. Another

substantial change involved incorporating pre-configured spreadsheets in several of the lessons so that our students would not be distracted by attempting to design appropriate calculations within the spreadsheets but would alternatively focus on the algorithmic thinking and the understanding of optimization routines.

3.2. Content on linear regression

The first few lessons after the introduction to DS lecture were focused on statistical association. These lessons were heavily modified versions of lessons contained in the MODULE(S2) statistics materials [20]. We chose to start the module content with linear regression specifically as finding the least squares regression line is a well-known optimization problem that is likely familiar to mathematics PSTs. We utilize linear regression models as a means to introduce more complex ideas of optimization later in the module. Thus, major modifications were made to the MODULE(S2) statistics materials [20] on statistical association to introduce bivariate linear regression using a data set collected by local high school students who were enrolled in a DS course offered by the high school. The high school students gathered local housing data from the Internet that included categorical and quantitative variables such as whether the home was East or West of the river, assessed value of the home, and property taxes. We then moved past the content covered in the MODULE(S2) materials to explore multiple regression using this data set. This included associations between two quantitative and one categorical variable; and association between three quantitative variables. We used Plotly to create three-dimensional scatterplots and compute least squares regression lines so the PSTs could explore what associations between three (or more) variables could look like visually. The mathematics PSTs also explored and analyzed projects that the high school students had completed using this data to support the PSTs' development of mathematical knowledge for teaching [25].

3.3. Content on tracking data

Most of our lessons utilized a data set—or variations there-of—representing animal tracking information. This data set provided the framework for students to explore and engage with multiple mathematical and DS concepts. Major concepts included optimization routines, algorithmic thinking, projections, and visualization. Students also engaged with fundamental

questions within a DS context, such as “What do we mean by classification?”; “Is there a correct answer (or ground-truth)?”; “Are there explanations for the data that reveal ambiguity?”; “How do we engage data as human beings with senses?” and “How might that be different for a computer?”; “How do we interpret our work, our predictions?”

PSTs explored several different animal tracking data sets throughout these lessons. The data sets consisted of animal tracking information. For the earlier lessons, the PSTs were provided with a data set that had few attributes (such as spatial attributes as x - y coordinates only or x - y coordinates and a time stamp, t). For later lessons, PSTs were given data sets that consisted of x - y coordinates, time stamps, and two heading attributes (speed coordinate s ; heading coordinate θ). Then, they were given a data set where there were gaps in the timestamps. Further, in earlier lessons, the number of animals was indicated; for later lessons, the number was not specified and students needed to estimate or infer that number from the data given.

Visualizations of the data constituted a substantial fraction of the material. For several of the lessons, visualizations were provided using data sets with a subset of the attributes or all of them; then students were expected to construct their own visualizations using a different data set. Students were instructed on how to use CODAP [31]; a free, web-based data analysis software for K-16 students) and a spreadsheet utility for generating images. Figure 2 below shows an image of a visualization of a data set in the x - y coordinates only (the provided data included the time attribute).

This image prompted students to envision connecting the dots together with a methodology to explain

their choices of how to connect the dots. Students explored the ambiguity of multiple different ways in which to connect the dots to create plausible trajectories of the animals. Students were also challenged on how they might conceptualize the manner in which a computer program would connect the dots.

With this data set, we emphasize the distinctions in visual cues given the different coordinate systems. We discuss how the data set with 3 attributes (x - y - t coordinates) would appear in 3 dimensions, and then how the two-dimensional views (see Figure 3) are artifacts of projections. We can illustrate a significant concept that when data inhabits much higher-dimensional space, visualization is not possible, but we can still explore the data through projections in this manner. Sometimes those projections even reveal significant structure!

Ultimately, the lessons scaffold student thinking and content mastery—in mathematics and DS—to conceptualize a complete solution to the animal tracking problem. We share this lesson in its entirety in Appendix B. Later lessons incorporated content emphasizing optimization routines as methodology for classifying points in space. Those lessons also include some student led algorithm design together with prior developed spreadsheet calculation to overcome students’ potential computational barriers. We discuss additional details of the optimization content in the next subsection.

3.4. Content on optimization

Our emphasis on optimization reflects two significant realities of DS as it is currently practiced: 1) modern ML techniques are massive optimization routines; 2)

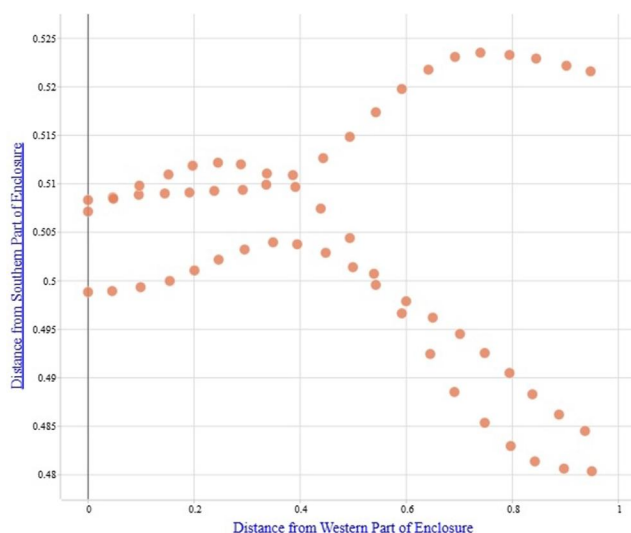


Figure 2. CODAP generated example plot of three-animal data set as presented to PSTs.

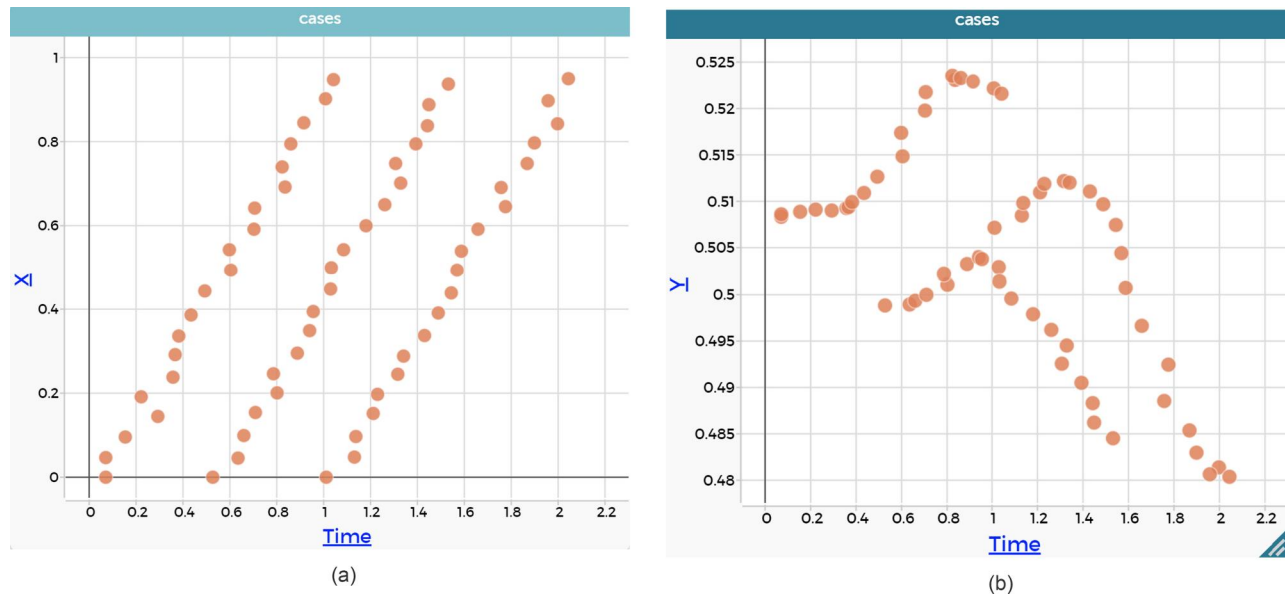


Figure 3. CODAP generated graphs of X vs. T and Y vs. T attributes.

those routines will always either: a) converge or b) terminate subject to a stopping condition. These two assertions form the foundation of multiple quantitative and ethical discussions with our PSTs.

ML forms much of the foundation of DS, particularly in the forms of artificial intelligence and large-language models. As Goodfellow et al. state:

A subfield of artificial intelligence, most modern machine learning, as used in production, can essentially be understood as applied function approximation; when there is some mapping from an input x to an output y that is difficult for a programmer to describe through explicit code, a machine learning algorithm can learn an approximation of the mapping by analyzing a dataset containing several examples of inputs and their corresponding outputs. The learning proceeds by defining a “model,” a parametric function describing the mapping from inputs to outputs (p. 56).

Similarly Hardin [32], notes “that the mathematics [of machine learning], e.g., gradient descent, is almost always hidden within the computational software, making it even more important that our students understand the underlying mathematics.”

We present elementary and illustrative examples of optimization problems in order to open the “black box” of ML. We emphasize that, while complex in scale and details, the essence of ML techniques is conceptually understood through a mathematical formulation. Our students encounter this formulation in prior coursework, including calculus.

Our first specific instantiation of an optimization problem is, in fact, linear regression. We introduce linear regression – as often practiced in secondary

and post-secondary coursework – as the “line of best fit”. We do so initially through visual motivation; using CODAP, students can visualize a line with the squares of the residuals depicted in real time as the parameters of the line vary. Students are then challenged to visually determine when the sum of the areas of the squares is minimized (see Figure 4). CODAP can then graphically display the actual line which minimizes the area – i.e. the squares of the residuals. This example has all of the essential features of an optimization problem – an objective function (the sum of the squares of the residuals) and a search space (all linear functions). As an optimization problem, however, it presents differently than prior exposure in calculus and so motivates deeper conceptual thinking about optimization routines. Moreover, the minimizer has a closed form expression and has relevant properties to the question of whether a linear model is appropriate for the given data. Therefore, we can use linear regression as an example of when the minimizer can be properly interpreted, so that the output of the optimization routine can be evaluated for trustworthiness [27].

Our second instantiation of an optimization problem is the traveling salesman problem. We present our students with a list of 5 points in the x - y plane together with the goal of finding the path of shortest distance that includes the 5 points. Using a spreadsheet that has been prepared for them, they can test various paths among the 5 points to see how different selections affect the total path length. We then introduce them to a “greedy algorithm”; after selecting an initial point, students are instructed to find the closest

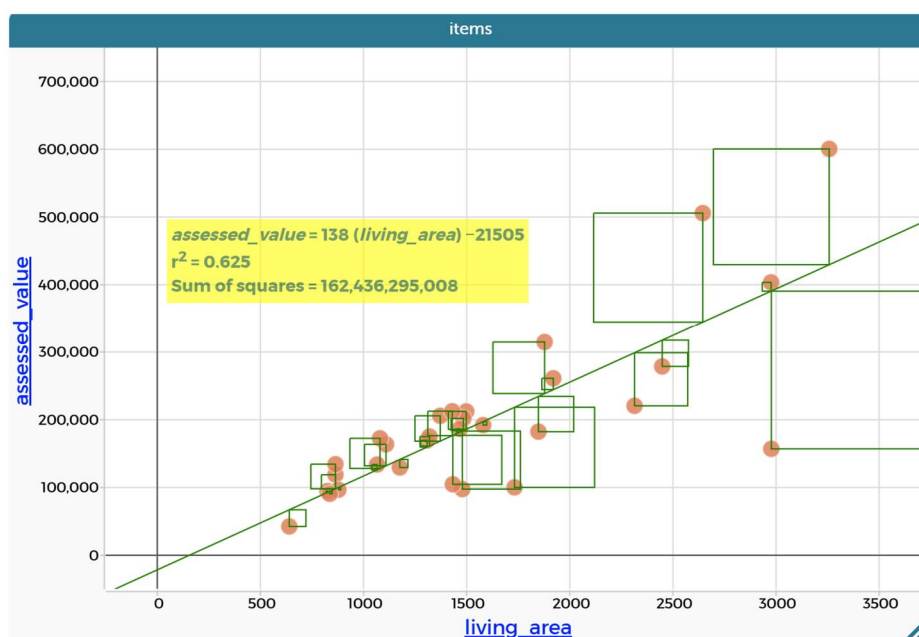


Figure 4. CODAP least squares regression line sum of square errors representation with housing data.

point of the remaining 4. That point will be the second destination of the traveling salesman; so students are instructed to next find the closest point of the remaining three. Students iterate the calculations until all 5 points are selected. We prompt students to try several different initialization points within the greedy algorithm, again to see how selections alter the total path length.

After their exploration with the greedy algorithm, we inform them of the shortest path together with the observation that the greedy algorithm would not always find that path (we design the locations of the 5 points so that this property holds). We thus illustrate that the greedy algorithm, as an optimization routine, will provide an estimate of the minimizer but will not necessarily provide the actual minimizer. Students are also challenged to consider how an algorithm would attempt to find the actual minimizer and conceptualize the difficulty level of the problem – i.e. it is a combinatorial search problem. We share this lesson and spreadsheet in its entirety in [Appendix C](#).

Using our animal tracking framework, we are able to then introduce several interrelated applications of optimization. These applications further develop students’ understanding of the main elements of optimization as well as provide exposure to see algorithms within the context of making predictions/classifications. Utilizing the animal tracking data set that contains the speed and heading attributes, we prompt students to consider how those attributes of a given data point (i.e. current time, location, *and* velocity) could be used to estimate the animal’s location at a

later time. See [Figures 5](#) and [6](#) for how we visualize the velocity vectors with speed and heading attributes. Students are able to define a metric using this intuition—assuming that the animal travels in a straight line over short time intervals – to then consider future location data relative to that straight line. Doing so motivates using projections – the distance from a point to a line can be computed using the scalar projection, a topic from advanced calculus. Students are able to design a methodology in which future data points are classified according to which are closest to the line estimated from a previous (in time), already classified data point.

An additional variation of this theme is for students to consider pairs of future data points (assuming two animals in this scenario), then minimizing the sum of the distances between pairs of points and the two lines associated with the two animals. Through this scenario, students immediately experience the ramifications of altering the objective function.

The use of projections was chosen for several reasons: 1) to connect to PSTs’ horizon content knowledge [25] and the content they were learning in their upper level mathematics coursework; 2) to support PSTs’ learning algorithmic thinking to find a prediction model using an optimization algorithm; and 3) to support PSTs’ understanding that optimization algorithms are what we need to code into a computer for it to generate a prediction model.

PSTs were asked to evaluate the “solutions” each of these algorithms provided to determine which solutions made sense in context as well as through visual

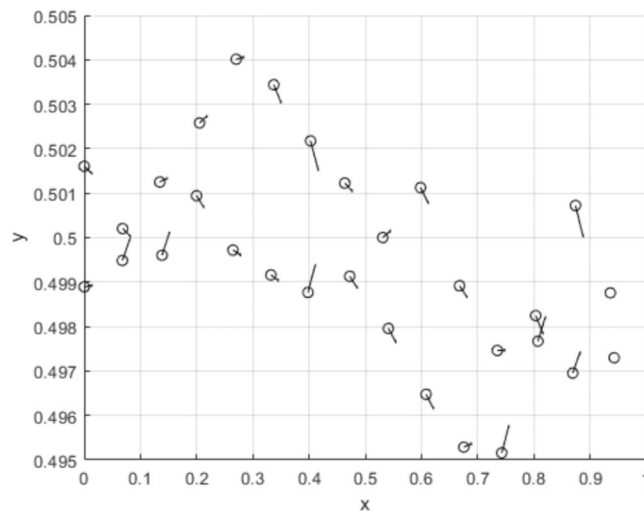


Figure 5. Graph representing X and Y coordinates, speed, and heading.

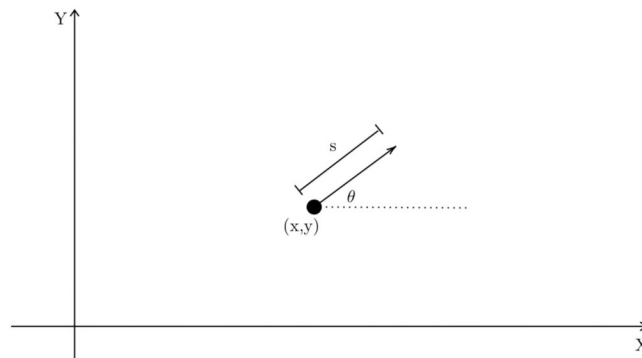


Figure 6. Illustration of how speed (s) and heading (θ) are represented graphically.

inspection. Students were then asked to communicate their reasoning. PSTs also discussed the idea that programming a computer to use a specific algorithm will generate a solution, just not necessarily the “correct” one. The role of humans is to evaluate the computer output within the context of the study, the data itself, and any ethical or societal concerns. We share this lesson and spreadsheet in its entirety in [Appendix D](#).

3.5. Outcomes of our study to evaluate curriculum effectiveness

We assessed the PSTs’ DS content knowledge through a pre-post test. This pre-post test consisted of three open response tasks that addressed each of the learning objectives that we have discussed for the module above. The pretest was given two weeks prior to the start of the DS module and the post-test was given directly following the end of the module. A rubric was created for each task to score the participants’ tests. The authors independently scored tasks 1 and 3

on all of the pre-post tests and resolved any disagreements collaboratively. Paired samples t -tests were utilized to determine whether there was a significant change in the PSTs’ DS content knowledge from before the DS module was implemented and after. The second task did not get scored with a rubric since we felt that the responses (and lack thereof) from the PSTs were better analyzed qualitatively as there were several potential possible solution pathways. For those few PSTs who did provide a response to the task, we felt that the responses from the PSTs were better analyzed using the constant comparative method of qualitative analysis [33] as such results will better illuminate what PSTs learned.

The first iteration of our course delivery was conducted under prior ISGC Curriculum Development Grant support. Our delivery was successful in that on the pre-post, mathematics PSTs showed a significant gain in understanding of DS concepts ($t=9.21$; $df=18$, $p<0.001$). However, a more qualitative exploration of what the PSTs learned shows there is

room for further development of the materials – specifically the content delivery and assessments of PST learning. A complete presentation of the study results can be found in Gallivan et al.

3.6. Future directions of the curriculum design

For our first goal, we intend to refine the course activities to incorporate more use of technological tools to better facilitate the learning of certain concepts (e.g. CODAP, R, and/or Jupyter Notebooks). For example, we intend to incorporate more scaffolds for PSTs to develop their understanding of classification models using CODAP add-ons (see <https://codap.xyz/>) to introduce the idea of decision trees – a topic that is accessible to their future high school students and is content included in already existing high school curricula [5]. Additionally, we are developing more user-friendly spreadsheet templates so that PSTs can interact more deeply with our animal tracking content in support of learning objectives *B. Prediction* and *D. Optimization*.

While our curriculum emphasizes conceptual understanding of broader DS ideas, our second goal will be to begin creating a bridge toward procedural fluency [30,34], e.g. learning a coding language. We assert that algorithmic thinking can be that bridge. We intend to pursue research that supports that assertion by incorporating lessons and activities on writing pseudocode for the traveling salesman problem activity as well as the animal tracking data. The purpose of including pseudocode activities will be to: 1) focus on mathematical and algorithmic thinking; 2) avoid having to teach a particular coding language in a 5-week module; 3) start thinking about how one might “tell” a computer to classify data points; and 4) and how to communicate the chosen methodology to others.

After data analysis during the first iteration of the project, we realized that the pre-post test we utilized to measure the PSTs’ learning of DS content needs some revision to align better with our learning outcomes (including assessing their use of technological tools) and to better measure student learning. We will refine this instrument as well as develop other measures of student learning that clearly align with our intended learning objectives.

Finally, we also would like to incorporate more opportunities for the mathematics PSTs to consider how to teach DS concepts, including the analysis of K-12 student work and completing and analyzing tasks they can use with high school students. Thus, we argue that existing research on mathematics and

statistics knowledge for teaching (MKT; SKT); [25,35,36] can help inform what we will call *data science knowledge for teaching* (DSKT). Through analysis of data (see Gallivan et al.), discussions of the overall goals for the module, and conversations with participants at conference presentations (e.g.,) [37], we developed two themes as an articulation of what we believe are important foundations for a theory of DS knowledge for teaching as we describe below.

4. Transition toward major themes

Based on our implementation and assessment of our DS module, we as Mathematics teacher educators have carefully reflected on and discussed what revisions we need to make to the module to better support PST learning of our learning outcomes (Table 1). Through this reflection and discussion, the authors determined that there are two overarching themes that we feel drive our decisions for what our module includes. We share them in the hope of being transparent with the progression of our work as well as share a glimpse of where our module revisions are going for the future.

These themes are: 1) we can leverage prior mathematics content knowledge by relating DS objects (e.g., models, predictions, and learning) to mathematical objects (functions, probabilities, and optimization); 2) we can illustrate grand societal challenges (e.g., trustworthiness of AI) with prior mathematical exposure and terminology (e.g., optimization algorithms always provide an output from a given input, though we do not always understand the outputs of such an algorithm).

We emphasize these two themes that pervade the content and inform the learning objectives for the module. The first theme recognizes that key DS concepts are predicated on prior mathematical knowledge. Indeed, our content demonstrates that our students actually engage multiple DS concepts in prior mathematics coursework under different nomenclature (e.g., linear models are functions). Consequently, our DS content actually provides PSTs with horizon knowledge in both mathematics and DS. The second theme prompts students to engage with ethical dimensions of DS (and therefore mathematics!). Because optimization algorithms are foundational techniques in DS applications [32], student exploration of these algorithms expose them to a critical lens for understanding the outputs of these techniques. Through the content, students consider the role and

trustworthiness of technologies deployed in processing and modeling data [27].

Thus, our goal for module revisions is to explicitly weave these themes throughout the content. Our intended results are two-fold. First, in a very short timeframe, we intend to provide PSTs with sufficient foundational knowledge through bridge building that they can pursue further study of DS independently. Second, we aim to instill in PSTs an interest in the significance of DS to themselves and their future students through an ethical lens.

4.1. Theme 1: Building bridges between data science and mathematics

The mathematics curriculum for PSTs is well developed—as PSTs progress through their programs, they are exposed to significantly more and deeper mathematical content. This content focuses both on mathematical objects and mathematical techniques. While PSTs’ level of mastery and self-efficacy varies, they encounter mathematical concepts repeatedly across multiple courses, providing them with a foundation of mathematics understanding. However, PSTs often receive very little exposure to DS concepts, potentially creating a barrier to their level of comfort with teaching DS in the field.

To overcome this barrier, we want to emphasize in our module that there is substantial overlap between the concepts in a DS context and those in mathematics. We endeavor to bridge the gaps that exist between DS and mathematics caused by distinctions in disciplinary nomenclature, practices, and ontology. A specific example of such a gap is the concept of a function. Functions as mathematical objects appear prominently in many mathematics courses where students, including PSTs, learn their formal mathematical definition as well as intuitive conceptualization (e.g., black box; vertical line test). In a DS context, the word function appears rarely, and if it does, it usually refers to a routine in a programming language. Rather, the object that mathematics refers to as “functions”, DS refers to as “models”. At a base level, a model is a function that is defined in some way by data, which is a collection of input–output pairs [38].

We assert that making the connection–bridge building – between DS and mathematics concepts more explicit can support PST learning. Specifically, if PSTs can recognize that models are really functions, they can develop a deeper understanding of DS content. The understanding that models (likely unfamiliar objects to PSTs) are functions (familiar objects) is our bridge’s first pillar – and perhaps the most significant,

though not the only instance of translating DS concepts into familiar mathematics.

A second pillar for our bridge utilizes optimization as a unifying concept. We call upon their prior experience with optimization in calculus, referring to common elements of an optimization problem—a target or objective function, and a search space (often described through constraints). Our goal here is to build intuition into algorithmic thinking—how can we conceptualize a computer program doing what humans do (visually or through other mechanisms)?

For example, we utilize our animal tracking problem to instantiate the transition from human interaction with data to algorithmic modes of interactions with data. After we prompt PSTs to consider visual interpretations of the animal tracking data, we (re-)introduce the Euclidean distance formula as a means of developing initial consideration of optimization algorithms. In our animal pathway examples, this means encouraging students to find the next closest point using Euclidean distance (and not simply through visual inspection) to determine which point belongs to which animal.

We are then able to further their consideration of optimization algorithms through varying the parameters of the problem formulation. For example, we can use the next closest point to develop algorithmic intuition, while also contrasting the greedy, iterative nature of this method *vs.* a search over all possible paths. Another variation we discuss is whether we would want to minimize angles between velocity vectors (as in the data) and the vector determined by the current point and the next potential point. Without becoming distracted by the details (due to the complexity of the dataset itself), we can illustrate how changing the algorithm or changing the search space would alter the outcome of the optimization routine. We can also challenge the PSTs to attempt to determine how their visual inspections may correspond to one of these variations on optimization.

To encapsulate this first theme, we endeavor to provide PSTs with sufficient foundational knowledge of DS through our mathematical bridge that they can pursue further study of DS independently, all within a very short timespan. We emphasize again the significant design choice of focusing on developing conceptual understanding to facilitate their future learning of DS concepts [30,34].

We assert that focusing on developing their conceptual understanding will best position them to teach DS content at the high school level. We note the variety of topics that are covered in existing high school

DS curricula – which are evolving as the field progresses. Therefore, we cannot possibly teach PSTs all of the DS content they may need, now or in the future. Thus, our goal is to expose PSTs to the underlying concepts of DS that are grounded in mathematical ideas to support their future exploration of other DS content.

4.2. Theme 2: Role and trustworthiness of technology

We acknowledge throughout the module that with the advent of everything from companies to social media collecting large amounts of data from people, that having the technology to assist in the analysis of that data is important. Our discussion on optimization provides the initial foundation upon which PSTs consider that technology. The goal of our module is not to have PSTs learn to write and implement computer code, but instead to encourage students to think algorithmically. Specifically, we have PSTs think informally about ways that a computer might be coded to classify points (e.g., through the use of projections) to highlight both the complexity and the utility of technology for analyzing large amounts of data that humans could not do by hand.

Our emphasis on optimization (as part of Theme 1) leads naturally into Theme 2. Without using the formalism and the language of ML, PSTs' engagement with optimization provides us with the opportunity to explore ethical and philosophical dimensions of DS, ML, and artificial intelligence. We are able to explain, in broad strokes, how AI (particularly large language models) are trained through optimization. We motivate PST thinking regarding the scales of magnitudes between our datasets (dozens of datapoints) in comparison to datasets within industrial applications. We are then able to prompt students to ponder important questions that result from those drastic differences in magnitude.

Through this prompting, we are able to briefly introduce PSTs to the concepts of “Explainable” or “Interpretable” AI [39,40]. We are also able to touch upon the question of trustworthiness of models. While we could cite a plethora of research publications on these issues [27,41–44], our emphasis is to have our students consider these issues without exposure to numerous research results.

Indeed, by building upon their understanding of optimization methods, we note that if one develops a model for a computer to use to analyze a data set, it will provide a “solution” based on that model. For example, if we want to classify points by connecting

them on a coordinate plane, we could have a computer find the shortest possible pathways using Euclidean distance. We can use this model for any set of x- and y-coordinate data and the computer will generate a solution. We problematize this idea for PSTs and ask them to be critical of it as the “accuracy” of that solution needs to be evaluated. In other words, we encourage the PSTs to consider what responsibility they have as humans in interpreting the model output and evaluating the solutions of the models that the computer is generating (i.e. the trustworthiness of models more broadly [27]). Thus, our intent is to revise the DS module to incorporate more in-depth discussions of how to interpret models generated by technology and the role humans have in evaluating that output.

To encapsulate our second theme, we introduce PSTs to the ethical dimensions of DS sufficient to pique their interest in understanding the relevance to their personal lives and those of their future students. Relative to DS content for the PSTs, the theme on role and trustworthiness emerged recently in our refinement of the intervention, as we have reflected on our several prior implementations. Therefore, the theme requires further development in future iterations, which will be a primary focus of ours looking forward.

5. Conclusion

There is an increasing need to build the data literacy of our students graduating from K-12 schools (cite GAISE II, SET). This call has been taken up through state initiatives to write K-12 DS standards [45], creation of curriculum materials [5], and DS frameworks for K-12 [46]. Thus, teacher preparation programs have also started to increase their capacity to prepare PSTs in DS. We reiterate our stance that mathematics teachers are in the best position to teach DS [8]. The future of our project is to continue to develop materials mathematics PSTs need to successfully teach DS and to further develop a DSKT framework. Acknowledging our environment in higher education and in teacher preparation specifically, we will continue to focus our efforts on developing DS content knowledge in tandem with their learning of other mathematics content (e.g., learning about optimization from a calculus and DS perspective). Our themes, curriculum, and burgeoning DSKT framework serves our ultimate goal as mathematics teacher educators, which is to ensure the preparation and development of

mathematics teachers to teach DS at the secondary level.

Acknowledgments

We emphasize our appreciation for the many collaborators that have contributed to the overall project of designing, delivering, and evaluating our curriculum for mathematics PSTs. We also appreciate the multiple funding sources that have facilitated these efforts. Eric Weber and Nathan Harding Mercil have been supported by the “Algorithms for Threat Detection” (ATD) program at the National Science Foundation through awards #1803254 and #2219959. The ATD program issues challenge problems for researchers supported by ATD funding; one of those challenge problems served as inspiration for our lessons on tracking data. ATD funding is also supported by the National Geospatial Intelligence Agency. Eric Weber has been supported by the Transdisciplinary Research in Principles of Data Science (TRIPODS) program at the National Science Foundation through award #1934884. Sarah Diesburg, Heather Gallivan, and Eric Weber have been supported by the Iowa Space Grant Consortium under NASA Award No. 80NSSC20M0107.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Iowa Space Grant Consortium, 80NSSC20M0107; National Science Foundation, 1830254, 2219959, and 1934884.

ORCID

Eric Weber  <http://orcid.org/0000-0002-9153-8089>

Heather Gallivan  <http://orcid.org/0009-0002-4940-8659>

References

1. Dykema K. The importance of data science—national council of teachers of mathematics. 2024. Available from: <https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Kevin-Dykema/The-Importance-of-Data-Science/>.
2. National Council of Teachers of Mathematics. A joint position of NCTM, NSTA, ASA, NCSS, and CSTA. 2024. Available from: <https://www.nctm.org/Standards-and-Positions/Position-Statements/Data-Science/>.
3. National Council of Teachers of Mathematics & American Statistical Association. Preparing pre-K-12 teachers of statistics. 2022. Available from: <https://www.nctm.org/Standards-and-Positions/Position-Statements/Preparing-Pre-K-12-Teachers-of-Statistics/>.
4. CourseKata. CourseKata statistics & data science for high school. 2024. Available from: <https://www.coursekata.org/highschool/textbooks>.
5. Gould R, Machado S, Johnson TA, et al. Introduction to data science. 2022. Available from: <https://www.ucladatascienceed.org/introduction-to-data-science-curriculum>.
6. IDSSP Curriculum Team. Curriculum frameworks for introductory data science. International Data Science in Schools Project. 2019. Available from: http://idssp.org/files/IDSSP_Frameworks_1.0.pdf.
7. YouCubed. YouCubed. 2013. Available from: <https://www.youcubed.org/>.
8. Gallivan H, Weber E. Artificial intelligence, data literacy, and preservice mathematics teacher training. *Iowa Council of Teachers of Mathematics Journal*. 2023. Available from: <https://www.iowamath.org/Articles/13285066>.
9. Hayashi C, With Yajima K, Bock HH, et al. Data science, classification, and related methods. *Proceedings of the fifth conference of the international federation of classification societies (IFCS-96)*; March 27–30, 1996; Kobe, Japan. Japan: Springer; 1998.
10. Ohsumi N. 2000). From data analysis to data science. In: Kiers HAL, Rasson JP, Groenen PJF, Schader M, editors. *Data analysis, classification, and related methods*. Berlin, Germany: Springer. p. 329–334. doi: 10.1007/978-3-642-59789-3_52.
11. Donoho D. 50 Years of data science. *J Comput Graph Stat*. 2017;26(4):745–766. doi: 10.1080/10618600.2017.1384734
12. Chizari M. The core of data science: a venn diagram perspective. 2025. Available from: <https://www.linkedin.com/pulse/core-data-science-venn-diagram-perspective-mohamed-chizari-zlqye/>.
13. Lee H, Mojica G, Thrasher E, et al. The data investigation process. *Invigorating statistics teacher education through professional online learning, friday institute for educational innovation: NC state university*. 2020. Available from: <http://cdn.instepwithdata.org/DataInvestigationProcess.pdf>.
14. Lee H, Mojica G, Casey S, et al. Enhancing statistics teacher education with e-modules. 2016. Available from: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1625713&HistoricalAwards=false.
15. Casey S, Bondurant L, Ross A. Engaging teachers in the combination of statistical investigation and social justice: Fairness in school funding. *Math Teach Educ*. 2023;11(2):139–142. doi: 10.5951/MTE.2022.0016
16. Iowa State University Department of Computer Science. Midwest big data summer school. 2021. Available from: <https://digital.ag.iastate.edu/event/2022/midwest-big-data-summer-school>.
17. National Science Foundation. Algorithms for threat detection. 2024. Available from: <https://www.nsf.gov/funding/opportunities/atd-algorithms-threat-detection/503427/nsf24-526>.
18. Weber E, Keinert F, Hegde C, et al. ATD: efficient and stable algorithms for non-euclidean regression in discrete geometries. 2018. Award #1830254. Available from: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1830254&HistoricalAwards=false.

19. Weber E, Harding SN, Przybylski L. AIS challenge problem dataset. n.d. Retrieved September 9, 2025, from <https://github.com/esweber/DS-for-PST-Challenge-Problem/tree/main>.
20. Casey S, Ross A, Maddox S, et al. MODULE(S2): statistical knowledge for teaching. 2022. Available from: <https://modules2dotcom.wordpress.com/statistics/>.
21. Franklin CA, Bargagliotti AE, Case CA, et al. Statistical education of teachers. American Statistical Association. 2015. Available from: <https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf>.
22. Lee H, Mojica G, Casey S, et al. Enhancing statistics teacher education with e-modules II. 2022. Available from: <https://fi.ncsu.edu/projects/enhancing-data-science-and-statistics-teacher-education/>.
23. Bargagliotti A. Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II). 2nd ed. Alexandria (VA): American Statistical Association; 2020.
24. Rosenberg J, Jones RS. A secret agent? K-12 data science learning through the lens of agency. Workshop on foundations of data science for students in grades K-12. 2022. Washington (DC): The National Academies Press. doi: [10.35542/osf.io/eyzkv](https://doi.org/10.35542/osf.io/eyzkv)
25. Ball DL, Thames MH, Phelps G. Content knowledge for teaching: what makes it special? *J Teach Educ.* 2008;59(5):389–407. doi: [10.1177/0022487108324554](https://doi.org/10.1177/0022487108324554)
26. Biswas S, Wardat M, Rajan H. The art and practice of data science pipelines: a comprehensive study of data science pipelines in theory, in-the-small, and in-the-large. Proceedings of the 44th International Conference on Software Engineering (ICSE '22); 2022 May 21–29; Pittsburgh, Pennsylvania. New York (NY): Association for Computing Machinery; 2022. p. 2091–2103. doi: [10.1145/3510003.3510057](https://doi.org/10.1145/3510003.3510057)
27. Varshney KR. Trustworthy machine learning. 2022. Available from: <http://trustworthymachinelearning.com/trustworthymachinelearning.pdf>.
28. Schumacher C, Siegel M. 2015 CUPM guide to majors in the mathematical sciences. In: Zorn P, editor. Washington (DC): Mathematical Association of America; 2015. Available from: <https://maa.org/resource/cupm-guide/>.
29. Franklin CA, Kader G, Mewborn D, et al. Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-K-12 curriculum framework. Alexandria (VA): American Statistical Association; 2007.
30. Abell M, Braddy L, Ensley D, et al. MAA instructional practices guide. Washington (DC): MAA Press; 2018. Available from: <https://maa.org/resource/instructional-practices-guide/>.
31. Common online data analysis platform [computer software]. Concord (MA): The Concord Consortium; 2014. Available from: <https://concord.org>.
32. Hardin J. A mathematical lens for teaching data science. *Notices Amer Math Soc.* 2025;72(11):1276–1284. doi: [10.1090/noti3253](https://doi.org/10.1090/noti3253)
33. Merriam SB, Tisdell EJ. Qualitative research: a guide to design and implementation. Hoboken (NJ): John Wiley & Sons; 2016.
34. Leinwand S, Brahier D, Huinker D. Principles to actions: ensuring mathematical success for all. Reston (VI): National Council of Teachers of Mathematics; 2014.
35. Groth RE. Towards a conceptualization of statistical knowledge for teaching. *J Res Math Educ.* 2007;38(5): 427–437.
36. Groth RE. Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Math Think Learn.* 2013;15(2):121–145. doi: [10.1080/10986065.2013.770718](https://doi.org/10.1080/10986065.2013.770718)
37. Gallivan H, Weber E, Butters L, et al. Data science curriculum for preservice secondary mathematics teachers. [Presentation]. Orlando (FL): Annual Meeting of the Association of Mathematics Teacher Educators; 2024.
38. Goodfellow I, McDaniel P, Papernot N. Making machine learning robust against adversarial inputs. *Commun ACM.* 2018;61(7):56–66. doi: [10.1145/3134599](https://doi.org/10.1145/3134599)
39. Mittelstadt B, Russell C, Wachter S. Explaining explanations in AI. Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019. p. 279–288. doi: [10.1145/3287560.3287574](https://doi.org/10.1145/3287560.3287574).
40. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?": explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). New York (NY): Association for Computing Machinery; 2016. p. 1135–1144. doi: [10.1145/2939672.2939778](https://doi.org/10.1145/2939672.2939778)
41. Harper SB, Weber ES. Fiduciary responsibility: facilitating public trust in automated decision making. *J Soc Comput.* 2022;3(4):345–362. doi: [10.23919/JSC.2022.0017](https://doi.org/10.23919/JSC.2022.0017)
42. IBM Research AI. Trustworthy AI. 2020. Available from: <https://research.ibm.com/topics/trustworthy-ai>
43. Lum K, Isaac W. To predict and serve? Significance. 2016;13(5):14–19. doi: [10.1111/j.1740-9713.2016.00960.x](https://doi.org/10.1111/j.1740-9713.2016.00960.x)
44. Toreini E, Aitken M, Coopamootoo K, et al. The relationship between trust in AI and trustworthy machine learning technologies. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). New York (NY): Association for Computing Machinery; 2020. p. 272–283. doi: [10.1145/3351095.3372834](https://doi.org/10.1145/3351095.3372834).
45. California Department of Education. Mathematics framework for California public schools: Kindergarten through grade twelve (mathematics framework). Sacramento (CA): California Department of Education 2023.
46. Charles A. Dana center at the university of Texas at austin. Data science course framework. New York (NY): Author2021.