Biehler, R., Kawakami, T., Weiland, T., & Zapata-Cardona, L. (2025). *Statistics and data science education as a vehicle for empowering citizens - A survey report for ICME 15*. Paderborn: Universitätsbibliothek.
https://doi.org/10.17619/UNIPB/1-2633

Statistics and Data Science Education as a Vehicle for Empowering Citizens—A Survey Report for ICME-15

ROLF BIEHLER

Paderborn University, Germany
biehler@math.upb.de

TAKASHI KAWAKAMI

Utsunomiya University, Japan
t-kawakami@a.utsunomiya-u.ac.jp

TRAVIS WEILAND

University of North Carolina, Charlotte, United States
tweilan1@charlotte.edu

LUCÍA ZAPATA-CARDONA

Universidad de Antioquia (Colombia)
lucia.zapata1@udea.edu.co

ABSTRACT

This survey reviews research in statistics and data science education with a focus on citizenship education. It examines the growing role of data in society, the rise of data science and AI, and their impact on science, society, and everyday life. We identify a spectrum of data-related literacies, from classical statistical literacy to AI, algorithm, and datafication literacy. The survey focuses on four themes: (1) Civic statistics and humanistic approaches to data literacy in the U.S. and Europe; (2) Critical data literacy perspectives from Latin America; (3) Connections between mathematical modeling and statistics/data science education; and (4) Contributions of mathematics/statistics education to AI and machine learning literacy.

Keywords: *Data science education, citizenship education, data literacies, statistical literacies, AI literacy*

1. DATA IN SOCIETY, DATA SCIENCE EDUCATION, AND CITIZEN EMPOWERMENT

1.1. THE NEW ROLE OF DATA IN SOCIETY AS A CHALLENGE FOR CITIZEN EDUCATION IN MATHEMATICS, STATISTICS AND OTHER SUBJECTS

Data is increasingly pervasive in daily life. School statistics has not kept pace with how citizens now engage with data—navigating X feeds, using AI to identify photos, or streaming GPS data into Google Maps for travel times. Data science and data-driven AI have transformed science and society. They can support social good, such as environmental protection and climate action, but can also serve narrow economic or political interests. Their use raises major concerns about privacy, data misuse, ethics, and surveillance. Growing awareness of the non-objective nature of data—including gender and racial biases in algorithm-training data—has become central to these discussions.

These developments require rethinking how education can empower citizens. Data, new visualizations, data-driven modeling, and algorithmic (predictive) modeling—including machine learning—are gaining importance not only in mathematics and statistics education, but also in science, social science, and computer science education. This interdisciplinary shift is reflected in theoretical discussions and a growing number of innovative projects and research studies. Theoretical approaches vary and appear under terms such as data and statistical literacy, critical data literacy, AI/data literacy, data-related mathematical modeling, and data-driven citizen science. We can also observe different approaches in different regions of the world.

Educator communities face the complex task of deciding which knowledge, competencies, and attitudes related to data, AI, and statistics are appropriate and feasible at different educational levels. This process, known as didactic transposition (Chevallard, 1985), begins by transforming scholarly knowledge into teachable classroom content. Because data science is an emerging, interdisciplinary, and rapidly evolving field, it poses greater challenges for didactic transposition than established single disciplines such as statistics or mathematics. Moreover, decisions about what to teach cannot rely solely on academic knowledge but must also consider the societal importance of data, which increasingly shapes the lives of individuals and communities.

Because school curricula are already crowded, two unresolved questions arise: What content must be removed to make space for data science, and what can different subjects—mathematics/statistics, computer science, natural and social sciences, and media education—contribute to data science education?

A reflective didactic transposition requires analysis on several levels:

- What roles do data, data science, and AI play in society and in citizens' lives that should be addressed in K–12 education?
- Which fundamental ideas of data science and AI should appear in K–12 curricula?
- How can diverse data literacy frameworks be related and synthesized, and what views of citizen education do they imply?
- What are key educational concepts and classroom examples aligned with these frameworks?
- What research evidence shows that interventions achieve their goals, especially for citizenship education?
- What support do teachers need to implement new data literacy approaches?

Addressing these issues requires broad, interdisciplinary scholarship. Our aim here is not to answer all questions but to survey relevant literature as a first step toward doing so.

1.2 The Review Focus of This Study

This paper is based on the survey report that was conducted for and presented at the International Congress on Mathematical Education (ICME 15) conference in Sydney in July 2024.[1] We examine recent trends in data science and statistics education (2020–2024) from the perspective of their contribution to citizen education. It is important to note that different stakeholders justify the inclusion of data science for different reasons. For example, data science and AI skills are often viewed as crucial for national competitiveness and economic growth (World Economic Forum, 2019; Green, 2024), which can lead to priorities different from those of citizen education.

---

[1] ICME is the largest international conference of mathematics education in the world and only occurs every four years. Survey reports are meant to provide a scholarly overview of the development of a domain to inform a general audience of mathematics educators, not just statistics educators. The focus of surveys is on the developments in the last four years but with enough background provided to understand the emergence of new trends and how they are situated in the field. The teams are appointed by the International Commission on Mathematical Instruction (ICMI) who is responsible also for organizing the ICME conferences. The title of the survey is suggested by the ICMI and the appointed team can then decide on the details of the survey content and procedure. We note this survey is the first commissioned by ICMI with a focus on data and statistics. The slides of the presentation with detailed examples can be found here in the appendix.

How mathematics education can and should contribute to citizen education is examined by Geiger et al. (2023) and related papers in the ZDM special issues. They argue that citizen education must be updated to reflect new uses of mathematics and technology in society, new media channels such as social media, and broader trends like the UN's sustainable development goals. Citizen education itself also needs reconceptualization: What kinds of citizenship—across democratic and more authoritarian contexts—should be addressed?

Geiger et al. draw on Westheimer and Kahne's (2004) distinction between personally responsible, participatory, and justice-oriented citizens. Personally responsible citizens obey laws and support community efforts; participatory citizens help organize such initiatives; justice-oriented citizens critically examine root causes of societal problems and act to challenge systemic injustices. These categories reflect different visions of democratic participation and highlight the need to align educational goals with broader civic and ethical aims.

Geiger et al. (2023) also note renewed traditions of critical citizenship education (Gibson, 2020) and global citizenship education (Oxley & Morris, 2013), which are increasingly reflected in mathematics education (see also Center for Universal Education at the Brookings Institution et al., 2017). What these citizen roles imply for mathematics, statistics, and data science education remains an open question and will evolve as disciplines and societal needs change. Although statistics and data science appear in broader discussions of mathematics and citizenship education, a detailed analysis specific to these domains is still lacking.

Such a review must consider the new role of data, algorithms, and data-based AI methods in science and society and critically examine related conceptions of citizen education—often framed as statistical literacy, data literacy, or AI literacy. Statistical literacy has long been discussed in relation to citizenship and democracy (Gal, 2002; Steen, 2001; Wallman, 2012; Watson, 2006), but these frameworks now require updating in light of today's expanded data practices and technologies.

For this survey, we concentrated on selected parts of the broader agenda outlined in Section 1.1. Specifically, we focused on:

- brief overview of the emerging field of data science and its role in science and society (1.3)
- A review of educational discourses and literacy conceptions related to data science for citizenship (1.4)
- A detailed analysis and exemplification of the most influential literacy frameworks
    - Civic Statistics and Humanistic Perspectives on Data Literacies Education in the U.S. and Europe (2)
    - Critical Perspectives on Data Literacy Emerging from Latin America (3)
    - Joint Discourse Between Mathematical Modeling and Statistics/Data Science Education Communities for Data Literacy (4)
    - Contributions of mathematics and statistics education on AI literacy (5)

The literature in Sections 1.3 and 1.4 provides the overall framework for the more detailed analyses in Sections 2–5. Civic Statistics, humanistic perspectives, and critical data literacy represent the most developed conceptions of data-related citizen education (Sections 2 and 3). The review of joint discourse between data-based mathematical modeling and data science education (Section 4) helps relate current trends in mathematical modeling to data literacy for citizenship. AI literacy—partly connected to data literacy—focuses on citizen education about AI and is often discussed within computer science education (Section 5). The respective section reviews and exemplifies how mathematics education may contribute to this facet. Bringing these four perspectives together and highlighting differences and commonalities is a unique contribution of this survey paper. These topics emerged from discussions within the survey team about how trends in statistics and data science education appear in different regions and areas of expertise. This approach was necessary because empowering citizens through statistics and data science draws on scholarship from many disciplines. We also observed that the topic is developing differently across the world, leading us to address these developments separately but unite them under the overarching theme of citizenship education.

We have chosen to take a non-traditional approach to our survey forgoing a systematic methodology, which we found did little to capture the diversity of scholarship around the theme. Instead, each section reviews respective theoretical frameworks and pedagogical examples. We see this as the special contribution of our survey, whereas most other surveys remain in their own disciplinary perspective, we try to present theory and practice across disciplines related to the core theme.

1.3 The Emergence of Data Science as a New Discipline And Data Science Education in the School Curriculum

Before discussing data science education, we must first discuss data science, which is an emerging field with many different definitions. Here is a compact definition from the National Library of Medicine:

> "Data Science is an interdisciplinary field which uses statistics, computer science, programming, and domain knowledge to collect, process, and analyze data for the purpose of acquiring knowledge or solving a problem. Data science also includes sharing acquired knowledge through storytelling, visualization, and other means of communication. Data science often employs methods such as machine learning, AI, natural language processing, algorithms, and other analytic tools to process and understand data." (https://www.nnlm.gov/guides/data-glossary/data-science)

Hazzan and Mike (2023) characterize data science as an interdisciplinary field:

> "Data science is emerging from an interdisciplinary integration of mathematics, statistics, computer science, and many other application domains such as business, biology, and education. On the surface, nothing about data science is new; any data science method or tool used today can be traced back to statistics, computer science, data mining, bioinformatics, and other data-intensive disciplines. The innovation, however, is in the integration—the holistic approach to data and methods to obtain knowledge and value, either financial, social, or educational." (Hazzan & Mike, 2023, p. 19)

A historical account of the emergence from within statistics, where John Tukey's (1962) paper and his Exploratory Data Analysis (1977) played a prominent role, is provided by Donoho (2017).

Data scientists ideally should have competencies in all these domains. Various Venn diagrams can be found that attempt to describe the structure of data science (e.g., http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram; https://upload.wikimedia.org/wikipedia/commons/4/44/DataScienceDisciplines.png).

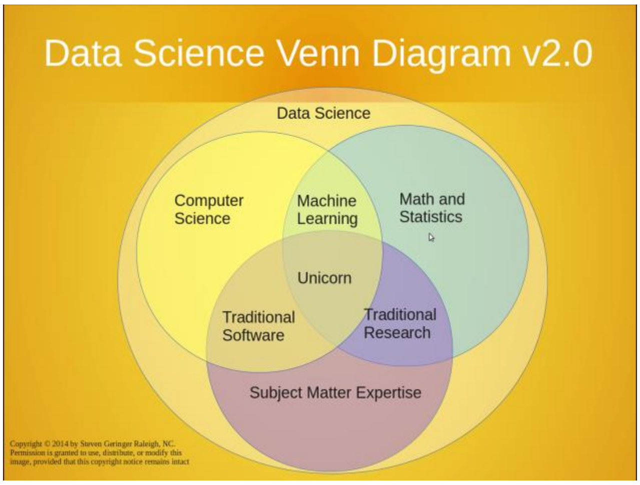


Figure 1: Data Science Venn diagram (http://www.anlytcs.com/2014/01/data-science-venn-diagram-v20.html). Copyright © 2014 by Steven Geringer Raleigh, NC. Permission is granted to use, distribute, or modify this image, provided that this copyright notice remains intact.

We like the version in Figure 1, which shows a unicorn at the intersection of all the disciplinary circles, demonstrating with humour that it is as easy to find a single person with all the competences as it is to find a unicorn. In essence, data science can be practiced with different sets of disciplinary expertise. The emergence of data science, with a growing number of programs at the tertiary level (Schwab-McCoy et al., 2021), is related to the new role of data in society and science.

In today's digital world, the concept of data has expanded far beyond numerical formats to include images, texts, and entire web pages. This broader view reflects major shifts in how information is created, shared, and interpreted. Data is now highly visible in media through interactive dashboards from data journalists (e.g., New York Times graphics), designed to make complex information accessible. At the same time, data-driven systems—search engines, streaming platforms, social media, wearables, and fitness trackers—collect and process large amounts of personal and behavioral data, while news portals use algorithms to curate content.

These processes, described as datafication, have reshaped scientific disciplines and societal practices, embedding algorithmic decision-making across many fields (Edmond et al., 2021; Gulson et al., 2022; Sander, 2024). As a result, citizens need strong data literacy to understand and critically engage with the systems influencing their daily lives. Increasingly, data also fuels AI and algorithmic decision systems that automate or support human decision-making.

These systems can be used for societal benefit—as seen in initiatives like "data science for social good" (https://www.datascienceforsocialgood.org) and "AI for social good" (https://www.mckinsey.com/capabilitiesquantumblack/our-insights/ai-for-social-good)—but they also raise serious concerns. Organizations such as the International Red Cross warn about the use of AI in warfare(https://www.icrc.org/en/document/what-you-need-know-about-artificial-intelligence-armed-conflict).. In several countries, scientific committees have examined ethical and political issues and reported them publicly. For example, the German Ethics Council (Deutscher Ethikrat, 2023) highlights ten cross-cutting concerns, including risks from statistical stratification, threats to privacy and autonomy through surveillance, bias and discrimination, lack of transparency and traceability, and questions of control and responsibility. These developments have led to calls for AI or algorithm literacy alongside data literacy.

We identified several initiatives that take an interdisciplinary view of school-level data science education. In April 2024, five major U.S. educational organizations—the National Council of Teachers of Mathematics (NCTM), the National Science Teaching Association (NSTA), the American Statistical Association (ASA), the National Council for the Social Studies (NCSS), and the Computer Science Teachers Association (CSTA)—released a joint position statement on the role of data science in K–12 education (NCTM et al., 2024). The statement argues that data science spans multiple disciplines and should not be confined to a single subject. Instead, it should be integrated across the curriculum to reflect the interdisciplinary nature of modern data work. Such integration is seen as essential for helping students think critically with data, conduct data-driven investigations, and understand ethical and societal implications.

This position aligns with other U.S. framework initiatives supporting didactic transposition, including the International Data Science in Schools Project (IDSSP Curriculum Team, 2019), the National Academies of Sciences, Engineering, and Medicine (NASEM) (2023), and work by Sukol (2024a, 2024b), who also reviews international projects. A recent effort to define coherent K–12 learning goals is the Data Science Learning Progressions (https://datasciencelearning.org).

1.4 Different conceptions of data-related literacies

The notion of statistical literacy has a long tradition in statistics education. A foundational paper is Gal (2002) and an early book is Watson (2006). The notion of data literacy is comparably new, and conceptions are formulated from other disciplines, not necessarily from statistics (e.g., Ridsdale et al. (2015). A very rich insight about the different current discourses can be gained from several special issues published recently. Special issues from *Teaching Statistics* (MacGillivray et al., 2021) and the *Statistics Education Research Journal* (Biehler et al., 2022) are close to our field. Additional special issues were published in interdisciplinary journals, such as the *Journal of the Learning Sciences* (Wilkerson & Polman, 2020), *Educational Technology and Society* (Matuk et al., 2022), *Information and Learning Sciences* (Acker et al., 2024), with three double issues; and *Computers* *and Education Open* (Chiu & Sanusi, 2024). Together, these publications represent a rich and diverse body of work shaping contemporary thinking and future directions in data science education. We found that quite a few publications frame their recommendations for secondary education as part of a broadly understood literacy education. We review these developments in this section and focus on some of them in depth in sections 2, 3, 4, and 5 of this paper.

Data Literacy and Statistical Literacy. Gould (2017) proposed that data literacy is effectively statistical literacy, enhanced by further components that reflect the new role of data in society including:

- "understanding who collects data about us, why they collect it, how they collect it;
- knowing how to analyze and interpret data from random and non-random samples;
- understanding issues of data privacy and ownership;
- knowing how to create basic descriptive representations of data to answer questions about real-life processes;
- understanding the importance of the provenance of data;
- understanding how data are stored;
- understanding how representations in computers can vary and why data must sometimes be altered before analysis; and
- understanding some aspects of predictive modeling" (Gould, 2017, p. 23).

There are some extensions and modifications of these notions, such as official statistical literacy introduced by Gal and Ograjensek (2018), as an important aspect that highlights the importance of understanding statistics produced by government agencies and other formal bodies. The contributions in Ridgway (2022) have advanced the concept of *civic statistical literacy*, which frames data understanding as essential to democratic participation and civic engagement. Complementing these views is the notion of *probabilistic literacy*, which focuses on reasoning under uncertainty and understanding probabilistic concepts in real-world contexts (Álvarez-Arroyo et al., 2024) and *risk literacy* referring "to one's ability to understand and evaluate risk, in order to support and make appropriate decisions." (Aven, 2024, p.1011).

The notions of *critical statistical* or *critical data literacy* have been put forth by Louie (2022), Weiland (2017), and Martínez-Castro et al. (2023), who emphasize critical data and statistical literacy as a necessary foundation for interrogating power, inequality, and the ethical use of data. These approaches will be studied in sections 2 and 3 of this paper.

Connecting data literacy and mathematical modeling (MM). It is important for citizens to be able to critically interpret model-based reports and products in the media and other public forums, as well as in AI systems, which can influence individuals' and society's decision-making (Gal & Geiger, 2022; Geiger, 2023; Ridgway et al., 2022). It is crucial for citizens to understand that predictions are based on models utilising assumptions and available data (Maass et al., 2022). Thus, a perspective of MM is essential for developing personally responsible, participatory, justice-oriented citizens or global citizenship for the data-driven society. In this context, collaborative efforts at the *boundary of MM and data science* (MM/DS) have recently emerged for data literacy for citizenship (Ärlebäck & Kawakami, 2023; English & Watson, 2018; Kazak et al., 2023; Makar et al., 2023). We will review these developments in section 4 of this paper.

Connecting data literacy and Artificial Intelligence. Modern AI, especially machine learning, heavily relies on data. It is no surprise that separate discourses have emerged within data literacy conceptions, such as AI literacy (Almatrafi et al., 2024; Casal-Otero et al., 2023), machine learning literacy (Chiang & Yin, 2022; The Royal Society, 2017) or algorithm literacy (Frau-Meigs, 2024; Gagrčin et al., 2024) . These concepts stem from computer science or media education, which have to be included in our survey as they are a part of our general view of data science education and are key for empowering citizens. Computer science education has a specific perspective on data-related practices in the context of AI (Grillenberger & Romeike, 2018; Olari & Romeike, 2024). We will discuss these approaches from the perspective of mathematics education in section 5 of this paper.

Positioning data literacy among broader literacies. Data literacy can be viewed as part of a broader constellation of literacies essential for navigating the digital age. UNESCO (2005) emphasized *literacy for life* as a foundational goal, and in this context, data literacy intersects with digital humanities (Abner, 2020), media and information literacy (Leaning, 2017), and digital literacy

(Mendoza-Chan & Pee, 2024). As technologies evolve, so too must literacies, leading to emergent forms such as AI literacy (Almatrafi et al., 2024) and machine learning algorithm literacy (Ma et al., 2023).

Disciplinary facets and versions of data literacy. Data literacies are not one-size-fits-all but are shaped by the epistemologies of specific disciplines, leading to a multiplicity of data literacies, each different in aspects based on the discipline they were created from and for. For example, Gulson et al. (2022) examined how datafication has reconfigured knowledge production. In the natural, earth, and life sciences, data practices underpin both professional research and public engagement through citizen science (Vohland et al., 2021). Socio-scientific problem-solving further illustrates how students must grapple with evidence and uncertainty (Mostacedo-Marasovic et al., 2024). In the social sciences, data literacy is also about understanding how data practices reflect and reproduce societal structures (Shreiner, 2020; Shreiner, 2023; Shreiner & Guzdial, 2024).

Expanding the concept to data-related competencies and attitudes. Emerging scholarship has also expanded data literacy to include affective and ethical dimensions. New terms and concepts reflect this broader understanding: *critical datafication literacy* (Sander, 2024), *personal data literacy* (Pangrazio & Selwyn, 2019), *data awareness* (Höper & Schulte, 2023), *data acumen* (Bargagliotti et al., 2020), *data conscience* (Marshall, 2022), and *data ethics* (Murillo et al., 2023). Activist and feminist approaches are also gaining ground, as seen in *data activism* (Ślosarski, 2023) and *data feminism* (D'Ignazio & Klein, 2020). These developments call for educational frameworks that do not merely teach students how to use data but also foster critical reflection, ethical reasoning, and social responsibility.

Last not least, we identified Fagerlund et al. (2025), as an important paper that situates citizen education in data science in a more general educational framework that distinguishes: "three interconnected domains—becoming equipped with skills (qualification), becoming a part of social orders (socialization), and becoming the subject of one's own life (subjectification)" (p. 1). Building on the work of Biesta (2020), they elaborate on the meaning of these functions in the context of data literacy education and base their review of potential K-12 goals for data science education on these concepts. Socialization is closely related to citizen education. Goals for citizen education set by educational administrations will differ, for instance, between democratic and autocratic societies.

2. CIVIC STATISTICS AND HUMANISTIC PERSPECTIVES ON DATA LITERACIES EDUCATION IN THE U.S. AND EUROPE

2.1. DATA LITERACY EDUCATION IN U.S. AND EUROPE

Situated within the frameworks presented in the previous section, we now zoom in to explore scholarship on emerging data literacy education in the U.S. and Europe, where scholarship has been prolific, often supported by federal funding initiatives from the National Science Foundation in the United States, and ERASMUS in the European Union. There has also been headway on adding data science or data literacy to the standards for K-12 curriculum in the U.S. (see Figure 2 for status of state implementations), pushed forward by recommendations from organizations such as the National Academies of Sciences Engineering Medicine (2023). In the 2023-2024 school year, over 193,000 students were enrolled in a dedicated data science course, and at least 277 schools delivered data science education (Sukol, 2024b). In this section, we focus on three different goals related to emerging work on civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe: (1) Present new emerging frameworks for considering data literacies for citizenship; (2) Overview the diversity of work around data literacies for citizenship coming out of the U.S. and Europe; (3) Provide some examples of relevant projects.

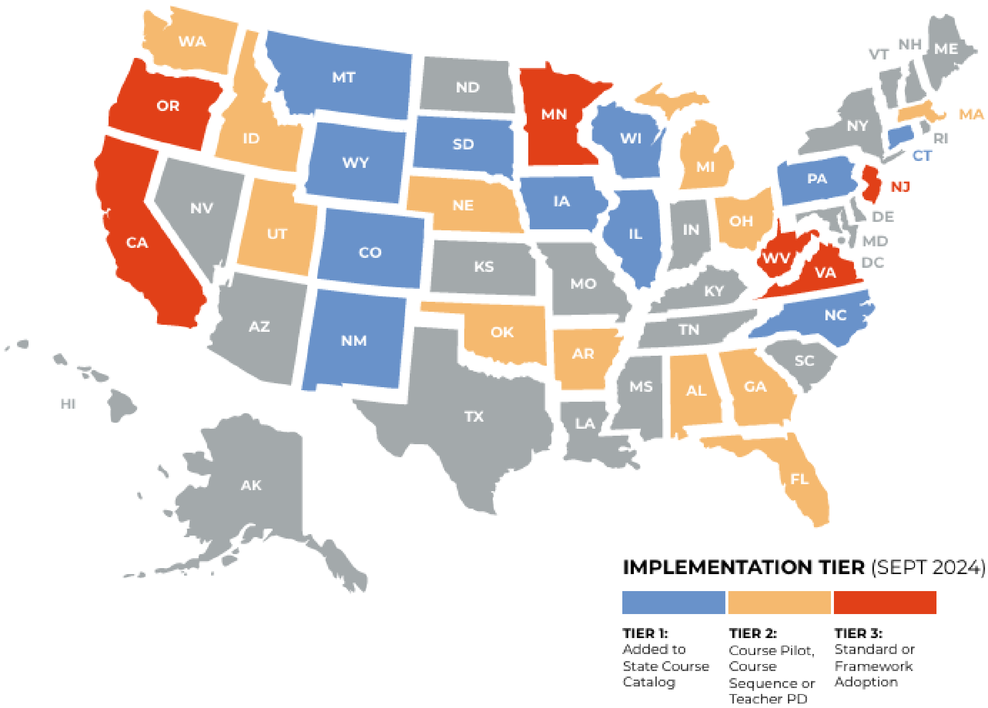


Figure 2: Spatial display of implementation levels of states for data science and data literacy from Sukol (2024b, p.7). Used with permission.

2.2. Goal 1: Emerging frameworks

In this section, we discuss three frameworks that are gaining importance in the emerging literature base of data literacies for citizenship. The first is the conceptual model of civics statistics from the ProCivicStats (PCS) project (Ridgway, 2022). Civic statistics is described as,

> statistics and quantitative evidence about key social phenomena that permeate civic life, such as migration, demographic change, crime, employment and poverty. PCS is predicated on the assertion that an understanding of Civic Statistics is essential for adults in democratic societies; however, the comprehension and critical interpretation of Civic Statistics often requires understanding of topics and issues that are different from or go beyond the knowledge gained from regular statistics curricula. (Nicholson el al., 2018, p.2)

The model includes 11 facets organized into three groups (see Figure 3) and has served as the conceptual basis for a growing body of literature. Furthermore, this is one of the only models that explicitly addresses the demands of citizenship.

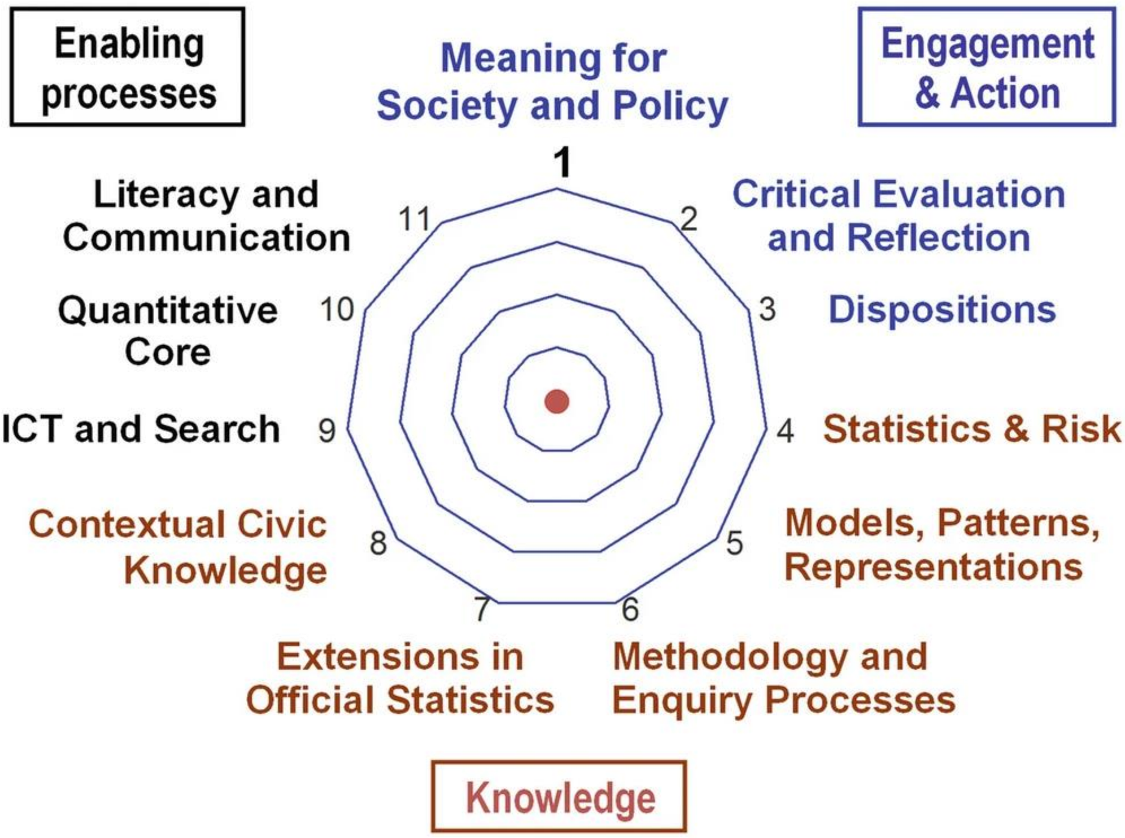


Figure 3. Conceptual framework for civic statistics (Nicholson el al., 2018). Used with permission.

A growing focus on humanistic approaches to STEM education broadly, and from the learning sciences specifically, has led scholars to call for a humanistic stance towards K-12 data science education (Lee et al., 2021). Lee et al. (2021) created a framework to describe core data practices from synthesizing decades of research as constellations of practices that extend across multiple layers. The layers they describe are,

- Personal: Students' personal and direct experiences with data, measurement, and the contexts in which data are collected.
- Cultural: The cultural and sociotechnical infrastructures and values enacted during a data set's collection and use (including but not limited to routines, technologies, and norms associated with various classroom, cultural, and disciplinary communities).
- Sociopolitical: The enduring political and social narratives that affect the purposes and methods by which data sets are constructed, interpreted, and used as social texts. (p. 665).

This call has sparked a growing literature base for humanistic data science research, which spans well beyond the boundaries of mathematics and statistics education. Though the demands of citizenship are not explicitly encompassed in Lee et al.'s (2021) framework, sociopolitical aspects of data science are, which do relate directly to demands of citizenship. Furthermore, the framework focuses on personal and cultural aspects of data practices that are a part of people's day-to-day lives and therefore part of being a citizen.

A third framework that does not come from education, but has been cited in a number of new scholarly works in statistics and data science education is that of data feminism (D'Ignazio & Klein, 2020). D'Ignazio & Klein (2020) draw upon feminist scholarship to look at data science and outline seven principles of data feminism that include: examine power, challenge power, elevate emotion and embodiment, rethink binaries and hierarchies, embrace pluralism, consider context, and make labor visible (see Figure 4). Projects working in this space were summarized by Lee et al. (2022) and continue to grow since this paper was published. A principle of data feminism that several studies have made inroads on is emotion and embodiment (Kahn et al., 2023; Lim et al., 2023; Radinsky & Tabak, 2022). Emotion is not an idea seen in the prior frameworks described and brings a new dimension to consider to data literacy. This framework does not explicitly address the demands of citizenship. However, there are lots of elements related to critical citizenship in this framework such as examining and challenging power, which are key aspects of critical citizenship, and are necessary for transformative action.

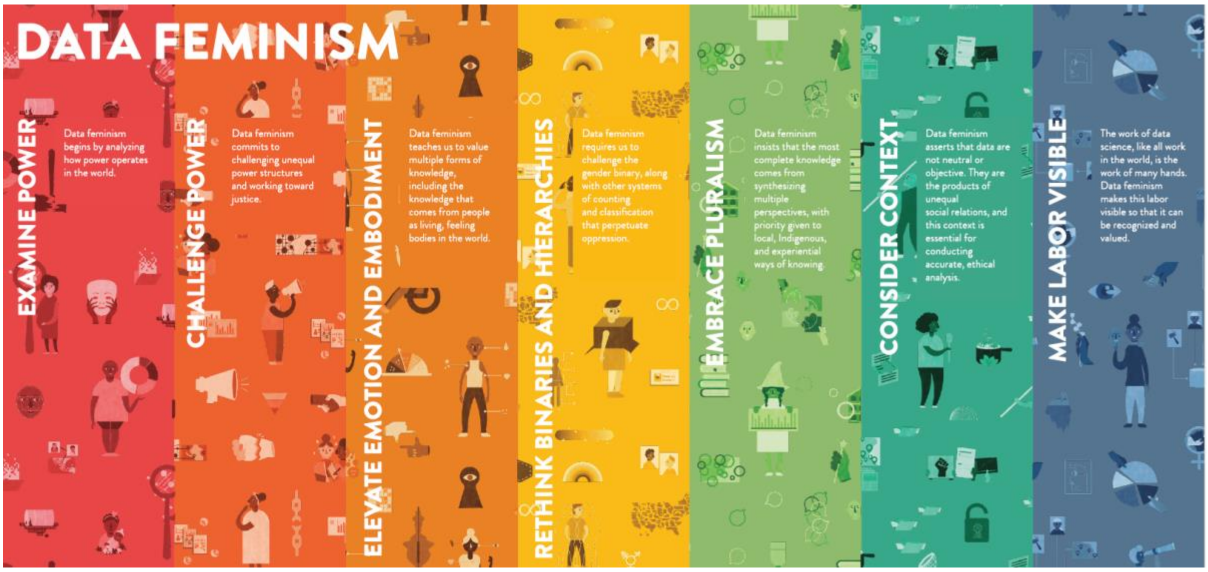


Figure 4. Data Feminism infographics by Catherine D'Ignazio, Lauren Klein and Marcia Diaz (2020). https://datafeminism.io/blog/book/data-feminism-infographic/. This graphic is licensed for re-use with the CC-BY-SA license.

There is overlap between these frameworks and strong connections. For example, the practices described in data feminism (D'Ignazio & Klein, 2020) constitute the type of data practices that Lee et al. (2021) talk about for humanizing data science, with constellations of data feminist practices existing through the personal, cultural, and sociopolitical layers of practices. There is also overlap of these practices with civic statistics, particularly the engagement and action dimensions. Civic statistics differs from the others in that it goes more specifically into technical skills and disciplinary practices of statistics, such as extensions in official statistics-. That said there is clear synergy between the frameworks which could be foregrounded or backgrounded in studies in different ways to highlight different issues and approaches in data science and statistics education research.

2.3. GOAL 2: OVERVIEW OF RELEVANT WORK

During a survey of literature related to the theme of this survey bound to the U.S. and Europe, several themes emerged that we describe in this section. To begin, much of the recent work is aligned with participatory and justice-oriented views of citizenship, looking to support citizens in becoming active and engaged. A wide range of approaches including conceptual, theoretical, and empirical were employed in the work reviewed with a large proportion being conceptual/theoretical (e.g., Louie, 2022; Rubel & Nichol, 2020). Empirical research has predominantly been qualitative in nature, consisting of small sample sizes, typically consisting of a single school class or a small group of children, often in an out-of-school setting (e.g., Van Wart et al., 2020; Wilkerson & Laina, 2018). Much of the work has been exploratory in nature, often occurring in informal education settings (e.g., Kahn 2020; Philip et al., 2013). Only a handful of studies were done in formal education settings (e.g., Louie et al., 2023). Interestingly this scholarship is being done in a diverse array of settings outside of mathematics education including science, social studies, computer science, and art/dance (e.g., Aridor et al., 2023; Rosenberg et al., 2022). This level of interdisciplinarity provides interesting insights and possibilities for the future. As this field is an emerging one, it makes sense to see a large portion of conceptual/theoretical work, but there is a need for this to shift more towards empirical work and larger-scale work.

There are several themes in the work we reviewed. The main themes we will describe are *reading the world with data*, and *writing the world with data*, along with descriptions of two emerging themes: data structures and handling, and technology. The first two themes are purposefully named reading and writing the world with data, drawing from Freire's (1970a) notion of critical literacy, because of the humanistic and sociopolitical framing of the scholarship emerging in relation to data literacy and citizenship. Much of this work was very focused on using statistics to bring sociopolitical issues such as climate change (e.g., Lanouette et al., 2024), inequitable access to resources (e.g., Rubel, Lim, et al., 2016), or sexism (e.g., Louie et al., 2023) into K-12 education to not only learn statistics, but also to support making sense of the world and one's lived experiences.

2.3.1 Reading the World with Data

The critical consumption and evaluation of data-based arguments have been a part of statistical literacy work for decades (e.g., Gal. 2002). However, in emerging scholarship we are seeing a more sociopolitical perspective on dealing with such arguments considering the subjectivity inherent in their creation and their impacts on the discourse in society. Scholarship on reading the world with data seems to have been heavily sparked by the pandemic with much of the literature being produced after 2020 and explicitly citing the pandemic as a rationale (e.g., Kahn et al., 2023). During the pandemic, citizens were bombarded with data visualizations about death rates, infection numbers, hospital beds, etc. Furthermore, the pandemic highlighted significant shifts in data visualization as it can now be dynamic, interactive, and connected in ways it never was previously, opening up new possibilities and new demands for citizens (da Silva et al., 2021; Engledowl & Weiland, 2021; Gundlach et al., 2015; Kahn et al., 2023; Lim et al., 2023; Philip et al., 2016; Rubel et al., 2021; Rubel & Nicol, 2020). The importance of spatial data and the ideas of spatial justice were also very prevalent in this growing literature base, highlighting the importance of considering such data visualizations in K-12 education (e.g., Lanouette et al., 2024; Poling & Weiland, 2020; Reigh et al., 2022; Rubel, Hall-Wieckert, et al., 2016; Rubel, Lim, et al., 2016; Rubel & Nicol, 2020).

2.3.2 Writing the World with Data

Advances in the consideration of critically consuming data-based arguments have similarly occurred in relation to new scholarship on data enquiry. Recent data enquiry work has also had a particular focus on making sense of real-world issues that are relevant to students' lived experiences. Students are provided opportunities to engage in investigating and telling stories about a wide range of contexts and issues including urban planning, participatory mapping, citizen science, alternative banking institutions, lottery sales, comparing pay based on sex, racial climate justice, COVID-19, socio scientific issues, social issues etc. (Aridor et al., 2023; Engel, 2017; Engledowl & Weiland, 2021; Kahn, 2020; Kahn et al., 2023; Lanouette et al., 2024; Lee, Drake, et al., 2021; Lee et al., 2022; Louie et al., 2023; Reigh et al., 2022; Rosenberg et al., 2022; Rubel, Lim, et al., 2016; Van Wart et al., 2020; Wilkerson & Laina, 2018). We therefore refer to this theme as writing the world through data, which includes scholarships that focus on students actively engaging in doing statistics through investigations and engaging in producing statistics. Recent ethnographic work that helps to capture the data enquiry process from Lee et al. (2022) builds from past work on statistical investigations (Bargagliotti et al., 2020; Franklin et al., 2015) and statistical enquiry (Wild & Pfannkuch, 1999) but extends into the modern practices of data scientists and provides insights into how to do that work with students and teachers. Modeling is an aspect of the data investigative process, and there are connections between this literature and that on data-rich mathematical modeling, which is discussed in detail in section 4. Another new aspect of literature in the writing the world theme is the notion of data storytelling (Kahn, 2020; Reigh et al., 2022; Wilkerson & Laina, 2018) or data narratives (Radinsky & Tabak, 2022). This literature is built from prior work on data arguments or interpreting results but takes a more humanistic perspective toward the work of data enquiry by acknowledging the personal and subjective nature of such data communications.

2.3.3 Emerging Themes: Technology and Data Handling

Two emerging themes in statistics and data science education are a focus on data structures and handling and on technology use. These two issues are deeply intertwined and have long held importance in the discipline of statistics. For example, consider Wickham's (2014) *Tidy Data*, which has become foundational for statistics and data science and serves as the basis for the popular R package ggplot2 using technology to structure and clean data. Kim et al. (2018) further describe the idea of taming data so that it is not too messy for students to make sense of, but it is not yet clean. One emerging area of promise that translates the ideas of tidy data and taming data into the K-12 educational setting is the idea of *data moves* (Erickson et al., 2019; Erickson & Chen, 2021). Such data moves include joining, grouping, calculating, summarizing, and filtering. Some of the work in this theme focuses on making sense of publicly available data which requires making sense of data structures and preparing the data for exploration, analysis, and modeling in different ways (Jiang et al., 2022; Poling & Weiland, 2020; Wilkerson et al., 2022; Wilkerson & Laina, 2018). However, these concepts rarely trickle down into the statistics curriculum most students experience, which is instead dominated with clean, structured datasets immediately ready for analysis (Weiland, 2019).

Technology was prevalent in all literature reviewed, which mirrors statistics and data science practice. CODAP (The Concord Consortium, 2014) has emerged in many research projects (e.g. Louie et al., 2023; Poling & Weiland, 2020; Wilkerson et al., 2022) and continues to grow in features and functionality. This dynamic action technology allows students to drag and drop attributes onto graphs to explore data primarily through visualization followed by adding measures. It is also designed to allow for hierarchical data structures (Konold et al., 2017) and data moves. Other technology tools that were present in the research we reviewed included professional tools like Excel, R and R studio, and Python. We also found Jupyter notebooks (Barba et al., 2019; Toomey, 2017) used to help scaffold the learning of programming languages like R and Python. This approach seems promising for scaffolding students learning to prepare them for professional practice (Biehler & Fleischer, 2021). Since reviewing the literature, it has become clear that generative AI will also begin to take up a role in this space supporting the writing of code and analysis for professional practices and has potential as a learning tool when used appropriately.

Technology use and data handling and structure are important themes to consider in relation to critical and humanizing perspectives in statistics and data science, as well as for data literacy and citizenship in that most publicly available data for citizens to investigate is large and often ill structured requiring data handling skills and technology to carry out such moves. We are also increasingly seeing the use of data dashboards that show live data streams which rely on technology to collect and display the data and also often make the data freely available though often in complex structures.

2.4. GOAL 3: EXAMPLES

In this section, we will describe three projects that are related to data literacy for citizenship that showcase the themes described in the previous section. Those projects include the *Introduction to Data Science* (https://www.idsucla.org/), *ProCivicStats* (https://iase-web.org/islp/pcs/), and *Writing Data Stories* (https://fi.ncsu.edu/projects/data-stories/). All these projects have substantial funding, allowing for large-scale efforts in new areas of study. They also all include aspects of reading and writing the world with data. Two of these projects also include substantial technology development.

Introduction to Data Science (IDS) is a long-term curriculum development project in the U.S. based out of the University of California, Los Angeles. The year-long curriculum is focused on introducing secondary students to data science in the mathematics curriculum and involves teaching students to reason with, and think critically about, data in all forms and is paired with professional development for teachers for adoption to occur at scale. At the time of writing (April 2025) the IDS curriculum was in use in 74 districts, 151 high schools and had reached 42,200 students (https://www.mobilizingcs.org/). The project leverages data relevant to the students to spark engagement and employs several types of technology to support student learning including participatory sensing technology, an interactive dashboard, and R. The use of participatory sensing "an approach to data collection and interpretation in which individuals, acting alone or in groups, use their personal mobile devices and web services to systematically explore interesting aspects of their worlds, ranging from health to culture" (IDS, 2022) is a personal form of data. To facilitate the handling and analysis of the data collected, the curriculum uses R/Rstudio to develop data science concepts and practices. The curriculum has four modules that cover: data and visualization; distributions, probability, and simulations; data collection methods: traditional and modern; and prediction and models. To date, this curriculum is the largest in scale related to data science in K-12 settings and could serve as a model to others. In an external evaluation of the impacts of the IDS curriculum on students, it was found that "evidence of significant student progress in their conceptual understanding of statistics. Perhaps just as important, students enrolled in IDS reported that they felt they had learned important skills over the course of the year, specifically in the area of coding." (National Center for Research on Evaluation, Standards and Student Testing, 2017, p.5). Furthermore, IDS classrooms have been the sites of various research projects (e.g., Philip et al., 2016; Philip et al., 2013).

The second example project is the *ProCivicStats* project (Ridgway, 2022; https://iase-web.org/islp/pcs/), which is a large-scale project across many countries in Europe and Israel building theoretical frameworks, empirical research, and producing curricular resources. The research project is primarily aimed at secondary and post-secondary audiences. Products of the project include: a conceptual framework, a database of teaching and learning material, sample lesson plans, datasets, sample syllabuses, and a review of dynamic visualization tools. The conceptual framework described earlier in this section serves as the basis for all the resources that were developed creating a coherent set of materials. A particularly impressive aspect of this project is that they have organized all their curricular resources into an Rshiny app that allows for searches based on language, statistics topic, tools, theme, level of difficulty, and material type. Furthermore, they provide an evaluation of statistical technology tools that is extensive and can support instructors in deciding which technology to incorporate into their classes. This was one of the only projects that was international, focusing not just on a single country, but issues of citizenship and data literacy globally.

The third example project is the *Writing Data Stories* (https://live-core-lab.pantheon.berkeley.edu/2022/11/18/writing-data-stories/) project that includes multiple partners in the U.S. (North Carolina State University, University of California, Berkeley, and the Concord Consortium). The project integrates computational data analysis into the middle school science curriculum in a longitudinal and interdisciplinary way (Lanouette, Rivero et al., 2024). It draws from the computer and data sciences, literacy studies, statistics, and science education. The project involves curriculum, resource, and tool development. The curriculum is centered around four elements *personal connections* - personal or community connection (with analysis of tensions or resonances with findings in data), *data* - analysis of a generally externally-sourced quantitative dataset, *interdisciplinary/multimodal* exploration of qualitative/contextual elements (journalism, interviews, observations) that help explain the tensions or resonances, and *call to action* or concluding argument. An important aspect of this work included using technology to make sense of publicly available datasets that also required the use of data moves to modify for exploration (Lanouette, Cortes, et al., 2024). This work directly addresses issues of critical literacy by having a focus on transformative action such as community planning and advocacy. Also making sense of publicly available datasets to investigate issues of personal and community importance leading to transformative action is an aspect of critical literacy that may be unique and emblematic of critical approaches in data literacy.

2.5. Section Summary

In this brief survey of the scholarship on data literacy in relation to citizenship in the U.S. and Europe, we discussed emerging theoretical frameworks, major themes in the literature, and provided some example projects that showcase the themes. There remain many questions to be answered in this area of study such as how does data literacy for citizenship get incorporated into the K-12 curriculum at scale? What concepts and practices need to be included? How should technology be created to support student learning? What are the changing data literacy needs for citizenship and how does this look differently for different types of government and citizenship? Something else important to note is that much of this work is being done with interdisciplinary teams including science, social studies, art, and dance. There are also aspects of critical theory that emerge at times in the literature, which is why the themes of reading and writing the world with data were used. However, we do not dig deeper into this aspect in this section but will address it in the next section in the context of Latin America where critical theory has long played a role in educational research.

3. CRITICAL PERSPECTIVES ON DATA LITERACY EMERGING FROM LATIN AMERICA

3.1. Introduction

As illustrated in the introduction, data science plays a crucial role in society and scientific development today. While every citizen is vulnerable to manipulation when using data-based digital artefacts (O'Neil, 2016), they can also benefit from the available data to generate useful information that enables them to make informed decisions (Ridgway, 2022). In a similar way, many disciplines are evolving more rapidly thanks to the availability of data in their fields, which was previously unimaginable. The previous section presented multiple conceptions of the citizen in educational research on data science, classified by Westheimer and Kahne (2004) as responsible, participatory and justice-oriented. However, the Latin American tradition is primarily rooted in critical educational approaches that aim to foster justice-oriented citizens who can recognise situations of injustice, challenge economic structures, address social issues and injustices, and act, which we delve into in this section.

Critical perspectives on data literacy in Latin America have flourished in response to a context of profound social, economic, and political inequalities that have fuelled resistance movements, popular struggles, and the urgent need for social transformation. Critical education in Latin America has been strongly influenced by liberation theories and social movements of indigenous peoples, peasants, workers, racialized peoples, feminists, youth, popular sectors, and vulnerable populations. However, critical education was essentially a response to the dictatorial contexts experienced in several Latin American countries since the second half of the 20th century: Colombia (General Gustavo Rojas Pinilla, 1953-1957), Brazil (successive military presidents, 1964-1985), Chile (General Augusto Pinochet, 1973-1990), Argentina (successive dictatorial presidents, 1976-1983). Dictatorships that implemented repressive policies to combat subversion, but which were accompanied by systematic violations of the fundamental rights of citizens. Philosophers and educators who lived through dictatorships identified the role of education in reproducing inequalities and proposed educational models that fostered critical awareness and emancipation. The critical educational perspective continues to influence contemporary educational studies, particularly those that promote education as a means of challenging social injustice and politically empowering marginalized communities. Critical data science seeks to theorize beyond data colonialism (Valente & Grohmann, 2024).

The aims of this section are to: (1) highlight critical perspectives on data literacy emerging from Latin America; (2) introduce some relevant scholars who have influenced work in the region; and (3) showcase initiatives involving critical data literacy.

3.2. Critical perspectives on data literacy emerging from Latin America

Critical data literacy is an educational approach closely related to critical thinking about data (Cabrera et al., 2023). In this perspective of data literacy, there is a strong link to the context in which data are produced (Tygel & Kirsch, 2016). Data are primarily used, in educational settings, to understand and transform social crises such as oppression (Valente & Grohmann, 2024), conflict (Ferreira da Silva & de Oliveira Souza, 2024), contradictions (Giordano et al., 2022), misery, injustice (Raffaghelli, 2022), ecological destruction (Zapata-Cardona, 2023) and exploitation. The type of citizen from a critical data literacy perspective would be one close to the classification of the justice-oriented citizen proposed by Westheimer and Kahne (2004) and the responsible citizenship suggested by Geiger et al. (2023).

Critical data literacy is the set of skills that enable people to critically use and produce data. It requires a combination of technical skills, such as data handling, computational skills, and mathematical methods, as well as the ability to critically analyze data, understand the context in which the data was produced, and understand the reality behind the data (Tygel & Kirsch, 2016).

Within this perspective, it is recognized that the use of data, algorithms, and AI is not neutral but, on the contrary, influenced by political choices, historical biases, and structural inequalities (de Sousa, 2022; Tironi & Valderrama, 2021). Moreover, citizens are easy prey for manipulation by big data algorithms (Mejía, 2020), while their privacy is violated and democracy is threatened (de Sousa, 2022), and many social and political decisions are based on data that is not accessible to communities (de Sousa, 2022). For example, recruitment systems discriminate based on gender (López Carreño et al., 2022), race (Villa Villa, 2020), or social class, and credit allocation algorithms perpetuate the exclusion of already marginalized sectors such as the elderly, women, and vulnerable populations (i.e., low income; de Olloqui et al., 2015). These social tensions should be studied empirically in educational settings.

3.3. THE DEBATE THAT HAS INFLUENCED CRITICAL DATA LITERACY IN LATIN AMERICA

The theoretical debate on critical data literacy from Latin America has been strongly influenced by critical pedagogy and the principles of popular education proposed by Paulo Freire (1970b). The foundation of Freire's popular education is the literacy method, which seeks to help people develop their ability to *read and write* the world. This is the ability to understand and transform a world in crisis. Popular education involves an emancipatory perspective that seeks to develop consciousness through transformative practices that are contextualized in the concrete experiences of the learners. Paulo Freire was influenced by the dialectical logic proposed by Marxist philosophers (Ilyenkov, 1977), in which contradictions and constant change are essential to understanding reality. In dialectical logic, as opposed to formal logic, individuals must consider the wholeness of contexts and the interconnectedness of subjects and contexts to understand social crises. Confronting crises is the essence of critical theory. Critical data literacy is a commitment to decolonizing practices that seek consciousness, autonomy, emancipation, problematization of crises, appropriation of historical reality, and transformation (Valente & Grohmann, 2024).

Following Freire, critical theory became a key part of the debate in Latin America. The work of Giroux (1993) and Skovsmose (2014) has been instrumental in shaping what Campos (2016) refers to as a *critical statistics education*. Although Giroux's most important contributions are in the field of critical pedagogy, and not necessarily in mathematics, his thinking has helped to translate critical theories into democratic contexts. Giroux (1993) suggests that students should be seen as active subjects who interpret and transform the world. In a complementary sense, Skovsmose's work has been fundamental to critical mathematics education because it has translated critical theories into the field of mathematics education to help understand the role of mathematics in society and in students' lives. Skovsmose (2014) proposes a critical, socially engaged, and reflective approach to mathematics education and how it can be used for social change. The critical statistics education proposed by Campos (2016) is the result of an articulation among Freire's ethical-political commitment, Giroux's ideological analysis and Skovsmose's critical methodology. This means teaching statistics from and for social reality. It also means providing an education that problematizes social issues and empowers students to become active agents of social transformation. Accordingly, students are expected to use data not only to describe the world, but also to change it. In essence, it is an effort to develop reflective citizenship, democratic attitudes, modeling and transformative action based on the study of data with some social interest. Against this background, critical data literacy has matured as an educational approach that goes beyond simply teaching how to read, interpret and use data: it seeks to develop the ability to critically analyze how, why, and for what purposes data is produced, communicated, and used in society. Critical data literacy involves asking questions such as: Who produced the data? For what purpose? What information is hidden or suppressed? How might the data be used to manipulate, exclude or empower? (Valente & Grohmann, 2024).

3.4. WHAT IS THE NEED FOR CRITICAL DATA LITERACY IN LATIN AMERICA?

Critical data literacy in a region with high levels of social, cultural, and economic inequalities could help people: make sense of the socially and economically important data that affects their lives, make informed decisions, participate in public life, recognize the harm that powerful interests can inflict with data (Mejía, 2020), recognize that data is not neutral (de Sousa, 2022), recognize that biased algorithms can exacerbate social and economic inequalities, and expose systematic social injustices. In Latin America, critical data literacy initiatives have emerged to address democracy (Giordano et al., 2022), social justice (Raffaghelli, 2022), visible and invisible complexities (Buehring & Grando, 2023), developing awareness of social issues awareness (Martínez-Castro et al., 2022), and co-liberation / contra culture (Raffaghelli, 2022).

We illustrate these conceptions by briefly discussing three examples from work in classrooms or with teachers, embedded in research studies. The following example is a classroom experience where data literacy was used to strengthen democracy. People cannot participate in decision-making if the information they receive from the authorities is not accurate. This was a research-based classroom experience involving two third-year high school classes (ages 17-21) from a public school in São Paulo, Brazil (Giordano et al., 2022). The aim was to support or refute statements with evidence. The experience began with the discussion of some statements made by the then-President Jair Bolsonaro in his speech at the 76th United Nations General Assembly. The statement that caused controversy was "In the Amazon, we had a 32% reduction in deforestation in the month of August, compared to August last year". To question the truth of the statement, students gathered information from multiple sources, used different ways of presenting data, integrated technical knowledge and language, improved evidence-based argumentation, and developed scepticism about public speech (Figure 5). Through this classroom experience, students progressed from being merely informed and engaged citizens, as described by Geiger et al. (2023) and Westheimer & Kahne (2004), to acting as responsible, justice-oriented citizens. They critically analysed public data, recognising that data is not neutral (de Sousa, 2022; Tironi & Valderrama, 2021) and that the way data is communicated can influence public opinion.

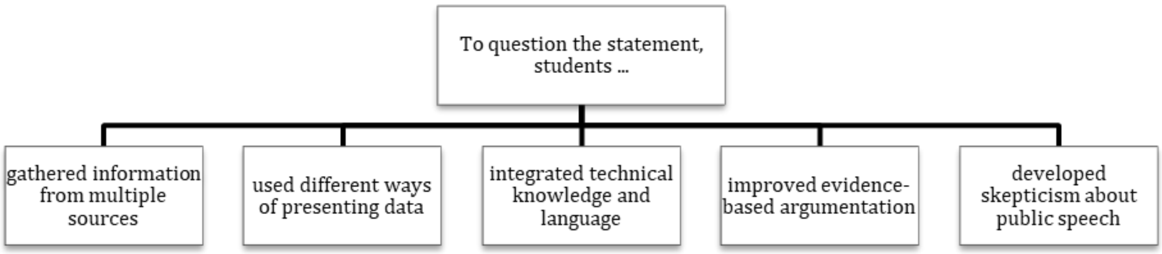


Figure 5: Actions taken by students to question the controversial statement. Diagram created by Lucía Zapata-Cardona.

A second example illustrates ways to develop awareness and visibility of social issues. The aim was to use open data to study inequalities between men and women in different dimensions of society (Martínez-Castro & Zapata-Cardona, 2022). The experience was carried out with 10 pre-service teachers in two two-hour sessions within a methods class for teaching statistics. The pre-service teachers started the experience by reading and discussing a news article about the "gender gap". They were then asked to find ways to answer a statistical question: "Is there a gender gap in your country? The participants collected and analysed information from public records; used some data skills such as data mining, data management, data visualization, and data cleaning; developed awareness of the gaps; proposed a list of possible solutions, such as speaking publicly about these gaps and offering tax incentives to companies that hire women. This classroom experience enabled students to recognise situations of injustice and act as justice-oriented citizens (Westheimer & Kahne, 2004), challenging injustice.

A third example comes from a workshop experience with in-service teachers and university instructors to illustrate co-liberation and social justice (Raffaghelli, 2022). The context was the *Mothers of the Plaza de Mayo*, an Argentine human rights organization formed in response to the military dictatorship of Jorge Rafael Videla. During this dictatorship, around 30,000 people disappeared. The *Mothers of the Plaza de Mayo* developed strategies to search for missing people using DNA; developed a 'grandparenthood index'; developed a memory map in several Argentinian cities; valued 'small data' to 'unravel' the complexity of narratives; and used open data as a form of activism. As an educational experience, the teachers and university instructors recreated and reflected on the tools used by the *Mothers of the Plaza de Mayo*. In addition, they asked questions such as: When is it possible to generate data co-liberation approaches? What are the necessary triggers? How can educational media serve the purpose of data co-liberation? The workshops were spaces where open data was constituted as a form of activism and as an open educational resource. Data collection and publication were discussed as mediators for working with local communities. This workshop demonstrates how the *Mothers of the Plaza de Mayo* used data for collective action. They used data to both *read* and *write* the world (Freire, 1970b), in an act of co-liberation. Participants in the workshop were able to engage in an emancipatory experience.

3.5. Section Summary

In Latin America, a critical data literacy perspective must help people identify who holds power over data, how biases can be reduced, and how information can be a tool for social transformation. These actions allow for the construction of more just, equitable models aligned with the needs of the most vulnerable communities. Looking at critical data literacy perspectives from Latin America provides a novel way of looking at critical perspectives. Today, technological and digital sovereignty in the United States, China, and the European Union resemble forms of capitalism (extractivist, patriarchal, and hegemonic); but, in Latin America, anti-colonial and community-based forms appear to be emerging, such as community data networks to protect Indigenous territories (Lehuedé, 2024).

4. JOINT DISCOURSE BETWEEN MATHEMATICAL MODELING AND STATISTICS/DATA SCIENCE EDUCATION COMMUNITIES FOR DATA LITERACY

4.1. Introduction

The development of critical thinking among citizens, as discussed in section 3, has been one of the key educational goals of mathematical modeling (MM) in mathematics education for over thirty years (Blum & Niss, 1991). Citizens need to critically understand that predictions are based on models that use assumptions and the best available data, and that they should be able to critically interpret reports and other public forms, as well as AI, that are based on data and models that influence decision-making in individuals and society (Geiger et al., 2023; Maass et al., 2022; Skovsmose, 2023). Such critical and interdisciplinary data literacy can be promoted through MM education as part of citizenship education, as discussed in sections 1 and 2. In this context, collaborative efforts at the boundary of MM and data science education communities have recently emerged for citizenship education (Ärlebäck & Kawakami, 2023; English & Watson, 2018; Kazak et al., 2023; Makar et al., 2023). A new integration of critical and interdisciplinary data literacy for citizenships with MM reflects an emerging shift in both communities towards data-rich MM—more authentic and interdisciplinary inquiry with data at its core. This section focuses on new trends in the joint discourse between the two scientific communities for data literacy for citizenships, focusing on three relevant discourses on data-rich MM since 2020: (1) data-rich MM cycle integrating statistics and mathematics, (2) interdisciplinary data-rich MM, and (3) societal data-rich MM. These three discourses are not complementary and independent but overlapping and interrelated.

4.2. Discourse 1: Data-rich MM cycle integrating statistics and mathematics

In contrast to the statistical inquiry cycle (Wild & Pfannkuch, 1999) referred to in statistics education, the typical MM cycles shown in Figure 6 (e.g., Blum, 2015, p. 76, originated from Blum & Leiss, 2007, p. 225) referred to in mathematics education often do not explicitly present the element of "data".

However, in recent years, a modeling process with statistics and mathematics at its core, in which the role of data is explicitly mentioned, has been proposed to develop statistical and/or mathematical literacy and/or learning (English, 2021; Kawakami & Saeki, 2022). In this section, we refer to this modeling process as the *data-rich MM cycle*. The discourse of the data-rich MM cycle also reflects the affective aspect of MM (Niss & Blum, 2020), which requires a proactive and autonomous disposition, as well as a critical disposition, to evaluate and improve the model in order to repeat the modeling cycle. These attitudes are essential for empowering responsible citizens.

For example, Figure 7 shows the data-rich MM cycle as suggested in English (2021). This cycle is similar to the MM cycle shown in Figure 6 (Blum & Leiss, 2007), in that it involves generating, evaluating, and modifying models. However, the data-rich MM cycle in Figure 7 includes data modeling (Lehrer & English, 2018), which is a type of statistical modeling (Pfannkuch et al., 2018). The cycle begins with *interpret and understand problem parameters* and consists of the phases: *pose questions*, *organize data*, *develop and refine models*, *draw inference*, and *communicate to peers*.

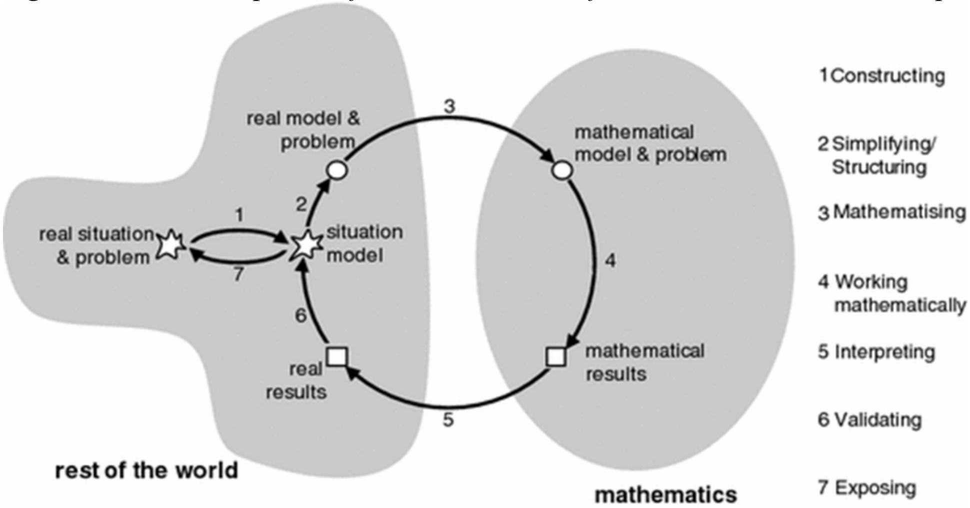


Figure 6: MM cycle, taken from Blum (2015, p. 76)

Data are clearly shown in the phases, *interpret and understand problem parameters* and *organize data*. At other phases, we can also identify the role of data in the modeling cycle in terms of developing, refining, and using models as a means of structuring and representing data. Additionally, the characteristics of statistics, namely uncertainty, and the characteristics of mathematics, such as generalize, are indicated in the phase *draw inference*. Thus, the data-rich MM cycle emphasises the importance of data in generating and validating models with mathematics and statistics, which distinguishes it from the typical MM cycle (e.g., Blum & Leiss, 2007), depicted in Figure 6. This highlight can foster citizens' understanding of the entire process and encourage their engagement in it, either partially or fully. English (2021) also points out that the data-rich MM cycle in Figure 7 can foster students' (future citizens') critical thinking about models constructed and the data and information presented. This cycle can enable their reflection on the model's purpose and intent, as well as critical examination of the justification for conclusions drawn from the model through constructing and refining models.

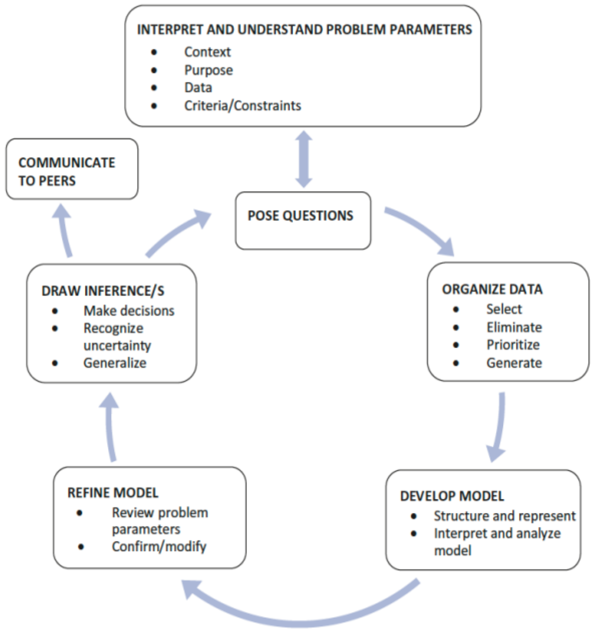


Figure 7: Data-rich MM cycle, taken from English (2021, p. 7), permission granted by Springer Nature

In this way, the discourse of data-rich MM cycles integrating statistics and mathematics emphasizes the modeling process as a cycle and explores the relationship between discipline-specific approaches to modeling and the role of data in them. This discourse also suggests that incorporating the data-rich MM process into data science education, including K-12 mathematics curricula, can help to develop citizens' data literacy and enable them to engage in data-driven investigations as modelling processes involving the critical iteration of model construction and refinement using statistics and mathematics.

4.3. Discourse 2: Interdisciplinary data rich-MM

Recent research on data-rich MM has been proposed not only with statistics and mathematics, but also with other disciplines/subjects in order to also promote STEM literacy (e.g., Bybee, 2018), which is essential for citizenships, also discussed in section 2 (Aridor et al., 2023; Fry et al., 2024; Kawakami & Saeki, 2024a; Lehrer et al., 2024) . In this section, we refer to this type of modeling as *interdisciplinary data-rich MM*. The background to this discourse is the harmony between the interdisciplinary nature of MM (English, 2016; Stillman et al., 2023) and the interdisciplinary nature of variation in data (Lehrer & Schauble, 2002; Makar et al., 2023; Watson et al., 2020). Modeling processes to explain and make predictions about phenomena are universal to STEM disciplines and subjects; data are not limited to mathematics but are also utilized in other STEM disciplines (English, 2016; Lehrer & Schauble, 2002).

Several empirical studies have tended to focus on developing a multifaceted perspective and knowledge through interdisciplinary data-rich MM (Aridor et al., 2023; Fry et al., 2024; Kawakami & Saeki, 2024a). These studies have pointed out that interdisciplinary data-rich MM promotes back-and-forth movement between deterministic views, stochastic views, and other STEM views, such as scientific views and design or engineering views. For example, Aridor et al (2023) explored middle school students' integration of deterministic, stochastic, and scientific reasoning, and understanding of the nature of science in a citizen science project on *radon* contamination. Fry et al. (2024) investigated a primary school teacher's communication supporting students' changing conceptions about data in a real-life STEM context with non-traditional forms of data (color photocopied leaves). Kawakami and Saeki (2024a) presented a case study of a primary school student's interdisciplinary data-rich MM process to design, test, and evaluate prototypes or design models of the tree seed dispersal, similar to the paper helicopters (Box, 1992), to maximize flight time using mathematical and statistical evidence. They identified model-based transitions the student made between mathematical (deterministic) reasoning, statistical (stochastic) reasoning, and other STEM reasoning, such as scientific and design-based reasoning, documenting evidence of the intricate connections the student made between the big ideas of these subjects. The importance of citizens' understanding of the role of uncertainty in the generation of data-based interdisciplinary knowledge has also been emphasized (Aridor et al., 2023; Lehrer et al., 2024).

Thus, the discourse on interdisciplinary data-rich MM suggests that interdisciplinary MM and data science, which is interdisciplinary in nature and cuts across different disciplines, are highly compatible. This discourse also suggests the possibility of implementing data science education and promoting opportunities to develop citizens' interdisciplinary data literacy throughout the K-12 curriculum, as discussed in section 1, by integrating data and modeling into the teaching of multiple school subjects.

4.4. DISCOURSE 3: SOCIETAL DATA-RICH MM

There are also recent studies on data-rich MM that focus on societal aspects to promote critical data literacy and citizenship, as also discussed in section 3. In this section, we refer to this type of modeling as *societal data-rich MM*. The discourse of societal data-rich MM reflects the social, critical, and prescriptive/performative aspects of MM (Barbosa, 2006; Davis & Hersh, 1986; Niss, 2015; Skovsmose, 2021). Davis and Hersh (1986), in discussing applied mathematics as a social instrument, indicated that "[by] the prescriptive function of mathematics I mean those situations where mathematics leads to human action or automatically to some sort of technological action" (p. 120). Niss (2015) called the prescriptive characteristics of MM *prescriptive modeling* and defined it as "a way for taking action based on decisions resulting from a certain kind of mathematical considerations, in other words to change the world" (p. 69). From the perspective of critical mathematics education, Skovsmose (2021), in discussing the formatting power of mathematics to project and reconstruct reality, stated that "[a mathematical model] formats the way we are acting in the situation. The model is not only descriptive; it is also performative" (p. 378). This means that mathematical models, on the one hand, serve a function of describing aspects of reality as *descriptive* models, and on the other hand, actually change or construct reality as *prescriptive/performative* models, carrying the risk that this function consciously or unconsciously shapes human perception and behavior and the world. The prescriptive/performative function of mathematical models would also apply to the context of data science and models based on big data and AI (O'Neil, 2016).

Empirical or theoretical studies on societal data-rich MM have addressed global, social, political, ethical, and daily life contexts to create authentic data science practices. For example, the following contexts have been used: COVID-19 or epidemics (Kawakami & Saeki, 2024b; Maass et al., 2023); climate change (Kazak et al., 2023; Steffensen & Kacerja, 2021; Zapata-Cardona & Martínez-Castro, 2023); pandemic-related media items (Gal & Geiger, 2022); reliability of public data sets (Wilkerson et al., 2022); mapping crime in the regions (Andersson & Register, 2023); social justice in fair distribution of school funding (Jung & Wickstrom, 2023). Wilkerson et al. (2022) explored pairs of secondary students' critical reflection on potential sources of variation in public socio-scientific datasets through inverse societal data-rich MM (model → data → real world). Gal and Geiger (2022) argued the importance of citizens' critical understanding of statistical and mathematical products (StaMPs) in media based on societal data-rich MM. The object of this understanding includes models, predictions, causality, risk, data quality, and strength of evidence.

4.5. AN EXAMPLE OF SOCIETAL DATA-RICH MM FROM TEACHER EDUCATION

Societal data-rich MM also serves as exemplary learning for developing tasks and lessons on the prescriptive/performative power of data-informed models in society. We illustrate this with an example from teacher education in Kawakami and Saeki (2024b). This research reported a teaching experiment for a group of mathematics pre-service teachers (PSTs) to experience basic societal data-rich MM and become aware of the societal benefits and risks of the MM reconstructing reality in society. The context used in this research was data-informed decision-making on the COVID pandemic in Japan. The PSTs addressed the task, "In which prefecture would you declare a state of emergency?" by visualizing and analyzing actual data using the tool CODAP mentioned in section

2. This research showed two ways in which pre-service teachers used models with statistics and mathematics (e.g., scatterplot, mean, quartiles, linear regression, correlation coefficient) through societal data-rich MM. One is that the PSTs used the models for *descriptive* purposes to visualize the trends and variability of data about the current world. For example, they used scatterplots of all prefectures to show the covariation between population density and the number of infected people in all prefectures using correlation coefficients (Figure 8). The other is that the PSTs used the models for *prescriptive/performative* purposes, to articulate data-driven societal decisions and guide human action for a preferred world. For example, they used scatterplots of 47 prefectures to demonstrate indicators of the spread of infection using the percentage of hospitalized and critically ill patients with COVID-19 (Figure 9).

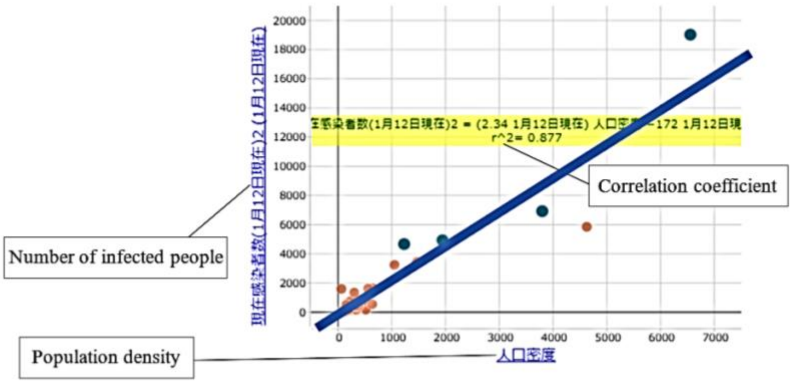


Figure 8: Model use for descriptive purpose (Adapted from Kawakami & Saeki, 2024b, p. 601)

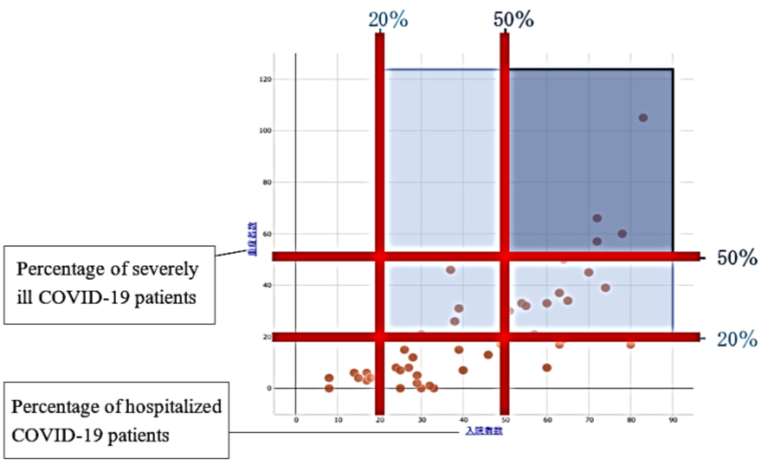


Figure 9: Model use for prescriptive/performative purpose (Adapted from Kawakami & Saeki, 2024b, p. 601)

Thus, the discourse on societal data-rich MM provides implications for empowering citizens by promoting awareness and critical reflection on the prescriptive power of data-based models to reconstruct reality in society. This power of models has been noted by O'Neil (2016) and Skovsmose (2021), and in an era where the role of models in media and AI is increasing, we cannot afford to overlook the prescriptive/performative nature of data-based models in order to empower responsible and reflective citizens equipped with critical data literacy.

4.6. Section Summary

This section reported on three recent discourses on data-rich MM that have been commonly discussed in the MM and statistics/data science education communities. All three discourses emphasized the role of models, the modeling process, and data in cultivating citizens equipped with critical and interdisciplinary data literacy. The findings suggest that the use of data-rich MM in data science education can be a vehicle for empowering citizens (1) to engage in the data-driven investigation iterating critically through model configurations and revisions, (2) to examine data from an interdisciplinary perspective, using deterministic and stochastic reasoning, as well as reasoning from other disciplines, and (3) to think critically about the usage of data-based models in society. Data-rich MM is a *boundary object* when mathematics educators and statistics educators collaboratively design, implement, and discuss data science education practices for citizens in schools.

5. MATHEMATICS AND STATISTICS EDUCATION'S CONTRIBUTION TO ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING LITERACY

5.1. INTRODUCTION

Frameworks of humanistic, civic, and critical data or statistical literacy address societal developments relevant to citizens. When these frameworks were created, data-based AI was still emerging. In principle, they can include AI-related data aspects, but the pedagogical examples we discussed did not. Meanwhile, AI literacy research (see section 1.4) highlights issues often absent from data science and statistics education—such as algorithmic fairness and ethical concerns when machine decisions occur in black-box systems without responsible human oversight.

AI also gives modeling a new role, especially predictive modeling, and relies more heavily on mathematical ideas such as matrices, partial derivatives, and high-dimensional optimization—topics not reflected in traditional modeling conceptions. Still, the discourses reviewed in section 4 offer an important basis for a broader understanding of data-related mathematical modeling for citizenship, which can be enriched with ideas from AI literacy. This section therefore reviews conceptions of AI literacy from a data science and mathematics education perspective and highlights pedagogical examples of accessible AI methods that can support data and AI literacy.

The rise of AI has created new demands for citizen education that go beyond current statistics and data science curricula. We can distinguish technical knowledge of AI from reflective knowledge about its societal use. An open question is what technical knowledge—and at what depth—is necessary to enable meaningful reflection. Another issue is which school subjects should teach which aspects. Traditionally, mathematics, statistics, and computer science address technical content, while social and political science focus on reflective dimensions. Yet to build reflective knowledge, students must also understand concepts such as machine learning, algorithms, classification and regression, types and magnitudes of classification errors, and the role of training data, including issues of bias.

What risks arise when decisions are made by opaque, black-box systems, and who is responsible for them? Understanding why and how such systems are created also requires technical knowledge. Ethical concerns—summarized in section 1 (Deutscher Ethikrat, 2023)—add further complexity. Citizens increasingly face AI-generated fake news, photos, and videos. These developments affect all levels of citizenship described by Westheimer and Kahne (2004): personally responsible, participatory, and justice-oriented. On a personal level, individuals must decide how to manage their data in social networks and on digital devices where AI operates in the background. Justice-oriented citizens may engage politically to prevent harmful AI uses or advocate for regulations that constrain companies developing AI systems.

5.2. DISCUSSIONS ABOUT AI LITERACY IN THE CONTEXT OF COMPUTER SCIENCE EDUCATION

The question of what secondary students should and could learn about AI is predominantly discussed in the community of computer science educators. Many publications on this topic are being published, even quite a few reviews of the literature (Almatrafi et al., 2024; Casal-Otero et al., 2023; Chiu & Sanusi, 2024; Floridi et al., 2018; Olari & Romeike, 2024; Sanusi, 2023; Tedre et al., 2021). What is most important from the perspective of mathematics and statistics education? A framework often quoted in this literature is the "Five Big Ideas in Artificial Intelligence" (AI4K12.org). The third big idea from the framework is *Learning*:

> "Computers can learn from data. Machine learning is a kind of statistical inference that finds patterns in data. Many areas of AI have progressed significantly in recent years thanks to learning

algorithms that create new representations. For the approach to succeed, tremendous amounts of data are required. (https://ai4k12.org/big-idea-3-overview/).”

These characteristics put machine learning in the neighborhood of statistical reasoning and statistical literacy. Machine learning is a specific type of predictive modeling, which Gould (2017), as we quoted in section 1, considers as one of the necessary extensions of traditional conceptions of statistical literacy.

We found the three dimensions of AI literacy that Casal-Otero et. al. (2003) identified in their systematic literature review very helpful for the perspectives from which machine learning can be taught in education.

> “AI literacy can be defined as a set of skills that enable a solid understanding of AI through three priority axes: learning about AI, learning about how AI works, and learning for life with AI (Long & Magerko, 2020; Miao et al., 2021).The first axis focuses on understanding AI concepts and techniques to enable the recognition of which artifacts/platforms use AI and which do not. The second axis addresses the understanding of how AI works, to effectively interact with it. The third axis seeks to understand how AI can affect our lives, allowing us to critically evaluate its technology.” (Casal-Otero et al., 2023, p. 2)

The second axis is related to technical knowledge of AI, which is necessary but not sufficient to address the third axis, which addresses reflective knowledge for citizen education.

A key aspect of AI and machine learning literacy is addressing the multiple opacity problem—the difficulty of understanding processes that are often not transparent (Burrell, 2016). Opacity arises from the learning algorithm, the resulting model, and the training data, all contributing to the “black box” character of many AI systems. Developing AI/ML literacy therefore requires helping learners critically interpret how algorithms work, how data becomes a model, and how the model operates.

In machine learning, predictive modeling is central. Compared with the mathematical modeling approach in section 4, some differences stand out. Machine learning often prioritizes prediction accuracy over explanation, and relies heavily on training and test data—elements less emphasized in traditional modeling education. Overfitting is a key concept: a model may fit training data extremely well but fail to generalize, performing poorly on test data. This trade-off between fitting known data and broader applicability is less explicitly addressed in classical mathematical modeling.

To support AI literacy, students should work with at least one machine learning algorithm in a way that reveals—rather than hides—how machines learn from data. This calls for white- or grey-box models whose internal workings are understandable. Such an approach makes the shift from manual to automated model construction visible, enabling students to grasp core ML principles, follow algorithmic decisions, and understand real-world uses of these systems.

Ideal candidates include:

- *(Linear) regression*, revisited with a mindset focused on prediction and automation, test and training data, accuracy of prediction (Fergusson, 2024)
- *k-nearest neighbors (KNN)*, which is intuitive, visual, and grounded in similarity-based reasoning (Bata et al., 2022; Mike & Hazzan, 2022)
- *Decision trees (DT)*, which offer a rule-based structure that mirrors human decision-making and can be easily traced and interpreted (Biehler & Fleischer, 2021; Erickson & Engel, 2023; Fergusson & Pfannkuch, 2024; Fleischer, Biehler, et al., 2022; Fleischer et al., 2024; Podworny et al., 2021; Zieffler et al., 2021)

These models provide a concrete foundation for exploring core ML workflows, from data input to model output, while helping students critically understand how algorithms interact with data to produce predictions.

In the next section, we zoom into decision trees as our example.

5.3. Decision trees as an elementary machine learning method for secondary students

Decision trees can be used for classification and regression problems. Many applications of ML in the real world are concerned with classification problems, which have not had a place in the statistics curriculum so far. Recent guidelines, such as the latest GAISE guidelines for K-12 statistics education, recommend decision trees as part of new curricula (Bargagliotti et al., 2020), although leaving the automatic machine learning creation of trees as optional.

For some years, the ProDaBi-project has explored how decision trees (DT) can be taught as an ML method to students from different grade levels (5/6, 9/10, 12/13) in the ProDaBi project (www.prodabi.de/en; Biehler & Fleischer, 2021; Fleischer et al., 2022; Fleischer et al., 2024; Fleischer & Biehler, 2025; Podworny et al., 2022; Podworny et al., 2021). We use this project as a case study and summarize and generalize our approach and experiences so that they can be helpful for future research and development.

We distinguish several types of design challenges.

Theoretical challenge 1: DT within machine learning. A key task is situating decision trees in the broader ML landscape. In our case study, DT serve as an accessible entry point to concepts such as predictive modeling, classifiers, and data-based model-building algorithms—overlapping but distinct perspectives.

Theoretical challenge 2: Authentic elementarization. DT must be simplified so that students can understand them while still retaining the essential principles of their professional counterparts.

Practical challenge 1: Problems and datasets. Tasks and datasets must be motivating, relatable to students, and still capture core aspects of real decision-making.

Practical challenge 2: Selection of tools. Classroom tools—digital or unplugged—must be accessible without extensive coding skills yet reflect the logic of professional tools. K-12 data science tools are reviewed in Biehler et al. (2013), Pimentel et al. (2022), and Israel-Fishelson et al. (2023).

Theoretical challenge 3: Supporting the shift from technical to reflective knowledge as part of citizen education.

5.4. EXAMPLE OF TEACHING AND LEARNING DECISION TREES WITH A MEDIA USE DATA SET

We illustrate possible elementarizations with a classroom example. Students worked with a dataset of roughly 160 variables and 1200 cases, based on a questionnaire on classical and social media use by secondary students—the YOU-PB (Youth–Paderborn) dataset. The questionnaire mirrors that of the biannual German youth media study (Rathgeb & Schmid, 2020), although only aggregated data are published there. Our project collected its own online sample (not representative) from the Paderborn region so students could explore the microdata in preparation for building decision trees.

The YOU-PB dataset (https://www.prodabi.de/en/materialien/datendetektiv_innen/) covers demographics, digital device ownership, use of online platforms and social media, traditional and digital media consumption, and gaming behaviors. Variables include grade, age, gender; access to devices such as computers, consoles, tablets, and smartphones; frequency and type of use of services like Instagram, TikTok, YouTube, Twitch, Snapchat, and WhatsApp; engagement with newspapers, TV, radio, and magazines; and gaming frequency, platforms, genres, solo or group play, and typical duration..

The problem introduced was: How can TikTok infer your real age from how you use the app? A few years ago, TikTok tried to identify users under 13—who are not permitted—by making hidden inferences from usage data. Understanding such inferences is part of citizen education. As a classroom proxy, students worked with the YOU-PB dataset to address: How can age be predicted from other variables such as media use and device ownership?

The idea is to use data where age is known and build a decision tree to predict it from the other variables, then apply the tree to cases with unknown age. The dataset was simplified by transforming all variables into binary form. Media use was recoded as “rarely” vs. “frequently,” and age as under13 vs. 13_and_above. Students used CODAP (www.codap.concord.com) with the Arbor plug-in to semi-automatically construct decision trees (see Erickson & Engel, 2023). In Figure 10, under13 is treated as the positive class—an arbitrary choice that later clarifies false-positive and false-negative errors.

The tree uses *Use_Instagram*, *Read_Online_Newspapers*, and *Use_Twitch* to predict *Age*. On the first level, the data split into students who use Instagram frequently (left) versus rarely (right). In the left branch, 79 are under_13 and 760 are 13_and_above. By the majority-vote rule, all are predicted as 13_and_above, producing 79 misclassifications (9.4%). Tree growth stops here. In the right branch, two additional levels are added, again using majority votes. For example, the rightmost leaf contains those students with rare Instagram use, rare online newspaper reading, and rare Twitch use. The majority vote predicts under_13 for all, with 205 correct and 118 incorrect.

Overall misclassifications across all terminal nodes are (line below the tree)

- 118 false positives (falsely predicted under_13)
- 120 false negatives (falsely predicted 13_and_above)

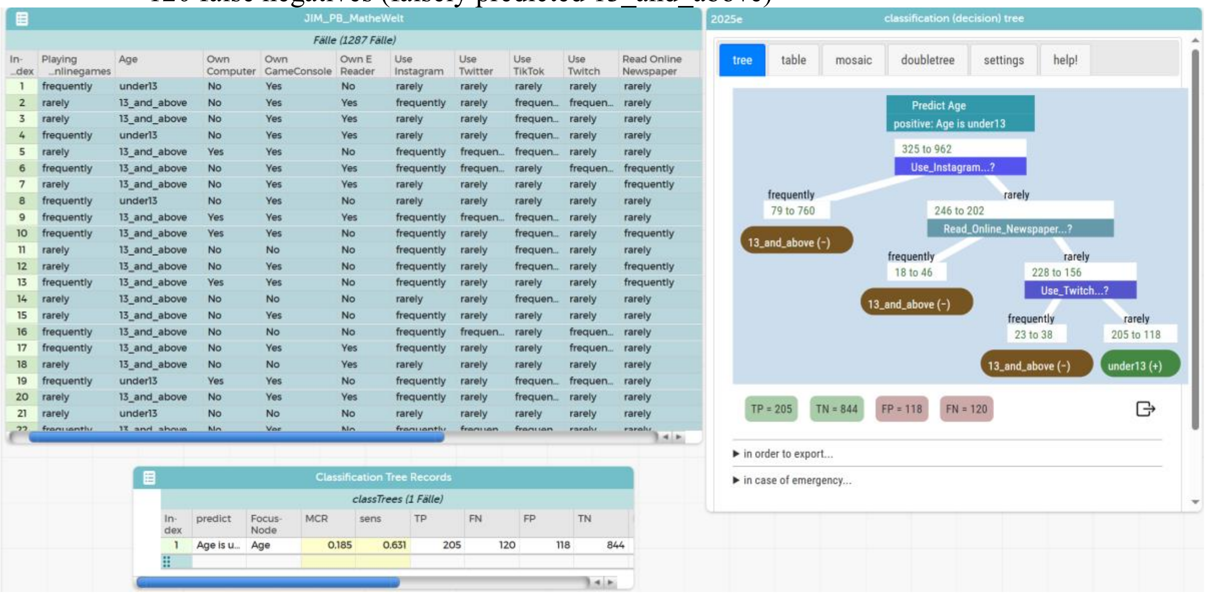


Figure 10: Screenshot from CODAP with Arbor plugin

In this context, false positives would be banned from TikTok but could continue if they provide proof of age, while the 120 false negatives—wrongly predicted as 13_and_above—could keep using the platform. If the goal is to block all under-13 users, this second error type is more serious. Distinguishing false positives from false negatives is therefore important for citizen education. The overall misclassification rate is (118+120)/(118+120+205+844) = 18.5% (also shown in Figure 10).

Students can drag and drop variables in CODAP to grow or modify a tree and may override choices to reduce misclassification. CODAP automatically reports correct and incorrect predictions, letting students focus on variable selection. Students choose variables they consider "informative" (context-based) or those expected to yield fewer misclassifications (data-based). The YOU-PB dataset is rich enough for other tasks as well, such as predicting interest in online games—an example of how platforms might target advertisements based on media-use data.

Overall, the classroom implementation followed a structured progression of tools and practices. Students first worked with a small subset of about a dozen cases represented on data cards to construct a DT manually. Next, they built trees in CODAP, deciding on splits based on data patterns or contextual reasoning. In a third step, CODAP was used to simulate how a machine constructs a tree purely from data, allowing students to enact the algorithm step by step.

Students then visualized the machine learning process with flowcharts before moving on to menu-based ProDaBi Jupyter Notebooks—developed by Yannik Fleischer—which implement the algorithm while hiding the code (Biehler & Fleischer, 2021; Fleischer, Biehler, et al., 2022; Fleischer, Hüsing, et al., 2022). These notebooks let students interact with the algorithm and experiment with hyperparameters.

For grade 12/13 students with Python experience, the final step involves coding: implementing parts of the algorithm or modifying worked-example notebooks (Toomey, 2017; Atkinson et al., 2000). This progression aligns with recommendations from Pimentel et al. (2022; National Academies, 2023) and draws on the documented educational value of Jupyter Notebooks (Barba et al., 2019) and worked examples.

5.5. Reflection on the Example from the Perspective of Citizen Education and Research Challenges for Implementing Decision Trees in Practice

Data set. The students worked with a data set collected from peers of the same age group. This context engages them, as it lets them draw on everyday knowledge to form hypotheses about relationships between variables. However, the data set is not an authentic one in which decision trees were used in real practice. Authentic data sets typically come from less familiar domains, making them harder for students to interpret or relate to.

Decision tree as a method. Students can construct simple trees by hand and value learning more systematic methods. This gives them a sense of agency in developing machine learning techniques. Comparing manual trees with automatically generated ones—considering prediction accuracy, explanatory power, and performance on training and test data—offers opportunities for reflection. Students also see that a decision tree depends heavily on the data used and that results can be misleading if the data are not representative. These insights are essential for reflective knowledge about AI.

Decision tree as a classifier. Students learn that classifiers always involve some misclassification and that the type of error can be crucial for evaluating a model's usefulness. They also encounter, often for the first time, a truly multivariate method that combines several variables for prediction—offering a concrete introduction to techniques common in modern AI.

Problems addressed with decision trees. Examples like personalized advertising or age prediction on TikTok illustrate authentic machine-learning applications. The used classroom data sets and variables, however, are only proxies for such real scenarios, and in practice other ML methods may be used. Still, these examples enable students to critically engage with the ethical, social, and technical implications of AI-based decision-making.

Computational tools. Authentic AI applications require coding and complex libraries, which can be a major barrier for many secondary students. A progression—from manual to semi-automatic construction and then to conceptualizing an algorithm for automatic construction—supported by pre-prepared Jupyter notebooks with minimal coding, can help. This scaffolding allows students to retain a sense of agency while gaining initial insights into professional AI practices beyond the school context..

F*rom technical to reflective knowledge*. Shifting the focus from purely technical aspects to reflective and ethical considerations is a major challenge, especially for mathematics and computer science teachers. Existing classroom research often centers on technical knowledge. Further design-based research is needed to improve learning materials, teacher guidance, and instructional strategies. This includes studying students' initial conceptions, learning difficulties, and learning outcomes, as well as teachers' challenges in facilitating decision-tree learning. To date, several qualitative classroom studies have been conducted (Fergusson & Pfannkuch, 2024; Fleischer & Biehler, 2025; Fleischer, Biehler et al., 2022; Fleischer et al., 2024; Podworny et al., 2025; Zieffler et al., 2021). However, systematic efforts to scale up teaching units and research are still lacking. Strengthening classroom implementations with reflective components is an important next step. While early educational designs naturally emphasize technical understanding of new concepts for both students and teachers, the future challenge is to integrate critical and reflective dimensions that foster AI literacy for citizenship.

5.6. Section Summary

This section reviewed the growing literature on AI literacy and AI education at the secondary level in the context of computer science education to identify possible contributions to the educational goals of mathematics and statistics education. Machine learning (ML) as part of AI is one of the topics that is relevant for mathematics and statistics education. We presented decision trees as an exemplary ML method and categorized and generalized the challenges for the didactic transposition of this topic from the research and development experiences in the ProDaBi project (www.prodabi.de/en). The theoretical and practical challenges, from the adequate elementarization of algorithms to the selection of data sets, problems, and tools that are accessible to students on the one hand, and that resemble authentic ML practices outside of school on the other hand, are discussed and exemplified. A future challenge is still how to combine technical knowledge with reflective critical knowledge about the use of AI in society.

6. CONCLUDING REMARKS ON FUTURE DIRECTIONS IN DATA SCIENCE EDUCATION

Our survey of work on statistics and data science education with a focus on citizenship education does not answer specific research questions. Across the four sections, different perspectives on data, modeling, and AI literacies for citizenship emerge that together outline a multi-layered research agenda for the coming years. Each regional or disciplinary perspective adds a distinctive focus—humanistic, critical, epistemological, or technological—that collectively advances our understanding of how citizens can be empowered to navigate datafied societies. It is hopefully clear to the reader that the theme of our survey, statistics and data science education as a vehicle for empowering citizens, requires considering a broad range of scholarship across many different disciplines and fields which are not well connected in the literature. For example, a central insight across all sections is the need for integrative frameworks that connect *humanistic and critical orientations* (sections 2–3) with *epistemic and methodological orientations* from modeling and AI education (sections 4–5). Such integration requires joint conceptual work between mathematics, statistics, computer science, and social studies educators. It also requires design-based research to test how tasks, tools, and pedagogies can jointly foster both technical proficiency and critical awareness. The idea of "data-rich modeling as a boundary object" exemplifies how interdisciplinary collaboration can succeed.

Currently the research landscape could be described as fragmented. Part of the problem is that technology development and use in this area is progressing rapidly, which in turn is leading to rapid changes in the literacy needs of citizens in today's societies. Educational research, on the other hand, is not known for developing quickly. This has led to many current initiatives for data science education, AI literacy, and data literacy broadly to be created and carried out with little research to rely on. What we are seeing as a result of all this is disparate evidence bases developing over similar topics, whereas what we need is more cross disciplinary collaborations to make sense of what is really a transdisciplinary space of data education, broadly defined. Such a shared space could be likened to the unicorn at the center of the Venn diagram in Figure 1 and requires visioning.

As we envision the evolution of data science and data literacy education for empowering citizens, several critical tasks emerge on the horizon. These are not merely incremental improvements, but foundational steps toward building a scalable, inclusive, and conceptually robust educational ecosystem.

- *Conceptual and Theoretical synthesis:* The concepts, terminology, and constructs relevant to data literacy and data education are developing fragmented across multiple disciplines and fields making it difficult to create a coherent foundation of scholarship to build from requiring synthesis and demarcation for clarity.
- *In-Depth Research:* Rigorous qualitative and quantitative studies are essential to unpack how teaching and learning unfold across various settings, curriculum, pedagogical approaches, tools used, and across time. This also includes developing meaningful assessments of key constructs with validity and reliability arguments for research.
- *Scaling Systemic Change:* Sustainable integration of data science into education requires organized collaboration with schools, curriculum developers, teacher preparation programs, administrators, policy makers and other stakeholders. This systemic change must be both top-down and bottom-up to be effective.
- *Tool, curriculum, and learning environment development:* To support scaling systemic change research-based resources and learning environments are needed including curriculum, data exploration tools, instructional routines, activities, etc.

Ultimately, the path forward lies in forging deeper collaboration across disciplines and educational systems to ensure all learners are empowered with the skills to thrive in a data-driven world. The ProDaBi project is a clear example of such collaboration, but we need significantly more collaborations that include different intersections of expertise across a broader spectrum of disciplines to continue moving forward. A perennial challenge to such work is the structural silos in academia, higher education and school educational systems. A major challenge being where do all of these next concepts and practices go in the curriculum? We do not have clear answers but finding them will take a transdisciplinary perspective viewing AI literacy or data literacy as the responsibility of education broadly. The central goal is not to produce mathematicians, computer scientists, data scientists, AI engineers, etc. but instead to empower citizens to make sense of their world and shape it for the future where data will only continue to play an important role in every aspect of life.

ACKNOWLEDGEMENTS

The survey presentation at ICME-15 was developed and delivered by the authors of this paper. We extend our sincere thanks to Erna Lampen (Stellenbosch University, South Africa), a valued member of the team, whose insightful feedback during the preparation of the survey and earlier versions of the slides greatly enriched the final presentation.

THE PRESENTATION OF THE TALK IS INCLUDED IN THE APPENDIX AFTER THE REFERENCES

REFERENCES

Abner, K. (2020). Data literacy as digital humanities literacy: Exploration of threshold concepts. In M. Hubbard, J. Perkins, & J. Russell (Eds.), *Literacies in a digital humanities context: A dh+lib special issue* (pp. 19–22). Dh+lib.

Acker, A., Bowler, L., & Pangrazio, L. (2024). Guest editorial: Special issue – perspectives on data literacies. *Information and Learning Sciences*, *125*(3/4), 157–162. https://doi.org/10.1108/ILS-03-2024-266

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, *6*. https://doi.org/10.1016/j.caeo.2024.100173

Álvarez-Arroyo, R., Batanero, C., & Gea, M. M. (2024). Probabilistic literacy and reasoning of prospective secondary school teachers when interpreting media news. *ZDM – Mathematics Education*. https://doi.org/10.1007/s11858-024-01586-8

Andersson, C. H., & Register, J. T. (2023). An examination of pre-service mathematics teachers' ethical reasoning in big data with considerations of access to data. *The Journal of Mathematical Behavior*, *70*, 1–23. https://doi.org/10.1016/j.jmathb.2022.101029

Aridor, K., Dvir, M., Tsybulsky, D., & Ben-Zvi, D. (2023). Living the DReaM: The interrelations between statistical, scientific and nature of science uncertainty articulations through citizen science. *Instructional Science*, *51*(5), 729–762. https://doi.org/10.1007/s11251-023-09626-8

Ärlebäck, B. J., Frejd, P., & Doerr, H. M. (2021). Exploring students' models of sampling and inference with nominal variables. *Quadrante*, *30*(1), 158–177. https://doi.org/10.48489/quadrante.23655

Ärlebäck, J. B., & Kawakami, T. (2023). The Relationships Between Statistics, Statistical Modelling and Mathematical Modelling. In G. Greefrath, S. Carreira, & G. A. Stillman (Eds.), *Advancing and Consolidating Mathematical Modelling: Research from ICME-14* (pp. 293–309). Springer International Publishing. https://doi.org/10.1007/978-3-031-27115-1_18

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Aven, T. (2024). Risk literacy: Foundational issues and its connection to risk science. *Risk Analysis*, *44*(5), 1011–1020. https://doi.org/10.1111/risa.14223

Barba, L. A., Barker, L. J., Blank, D. S., Brown, J., Downey, A. B., George, T., Heagy, L. J., Mandli, K. T., Moore, J. K., Lippert, D., Niemeyer, K. E., Watkins, R. R., West, R. H., Wickes, E., Willing, C., & Zingale, M. (2019). *Teaching and learning with Jupyter* https://github.com/jupyter4edu/jupyter-edu-book

Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *ZDM Mathematics Education*, *38*(3), 293–301. https://doi.org/10.1007/BF02652812

Bargagliotti, A., Binder, W., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., Huchting, K., Larson, S., Miric, N., Rovetti, R., Seal, K., & Zachariah, T. (2020). Undergraduate Learning Outcomes for Achieving Data Acumen. *Journal of Statistics Education*, *28*(2), 197–211. https://doi.org/10.1080/10691898.2020.1776653

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) - A Framework for Statistics and Data Science Education*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Bata, K., Schmitz, A., & Eichler, A. (2022). Insights into the design of an introductory course for data science and machine learning for engineering students. *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*, Feb. 2022 Bozen-Bolzano, Italy. hal-03751807,

Biehler, R. (2022). Revisiting Fundamental Ideas for Statistics Education From the Perspective of Machine Learning and Its Applications. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. https://doi.org/10.52041/iase.icots11.T1A2

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), Third International Handbook of Mathematics Education (pp. 643–689). Springer. https://doi.org/10.1007/978-1-4614-4684-2_21

Biehler, R., De Veaux, R., Engel, J., Kazak, S., & Frischemeier, D. (2022). Editorial: Research on Data Science Education. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.606

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*(S1), S133-S142. https://doi.org/10.1111/test.12279

Biesta, G. (2020). Risking Ourselves in Education: Qualification, Socialization, and Subjectification Revisited. *Educational Theory*, *70*(1), 89-104. https://doi.org/10.1111/edth.12411

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do? In J. S. Cho (Ed.), *Proceedings of the 12th international congress on mathematical education* (pp. 73–96). Springer.

Blum, W., & Leiß, D. (2007). How do students and teachers deal with modelling problems? In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modeling: Education, engineering, and economics* (pp. 222–231). Horwood. https://doi.org/10.1533/9780857099419.5.221

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects—State, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, *22*, 37–68. https://doi.org/10.1007/BF00302716

Box, G. E. P. (1992). Teaching engineers experimental design with a paper helicopter. *Quality Engineering*, *4*(3), 453–459.

Buehring, R. S., & Grando, R. C. (2023). Reading and writing the world with children: statistical thinking and multivariate data. *Statistics Education Research Journal, 22*(2), Special Issue: Early statistical and probabilistic thinking. https://doi.org/10.52041/serj.v22i2.446

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Bybee, R. (2018). *STEM education now more than ever*. NSTA Press.

Cabrera, G. P., Tauber, L. M., & Pochulu, M. D. (2023). Diseño e implementación de rúbricas para caracterizar y evaluar la alfabetización estadística crítica en contextos de lectura [Design and implementation of rubrics to characterize and assess critical statistical literacy in reading contexts]. *Revista de Educación Estadística, 2*(1). https://doi.org/10.29035/redes.2.1.4

Campos, C. R. (2016). *Towards Critical Statistics Education: Theory and practice*. Lambert Academic Publishing.

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00418-7
Center for Universal Education at the Brookings Institution, UNESCO, & GEFI-YAG (Eds.). (2017). *Measuring Global Citizenship Education - A Collection of Practices and Tools*. The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/04/global_20170411_measuring-global-citizenship.pdf.
Chevallard, Y. (1985). *La Transposition Didactique: Du Savoir Savant au Savoir Enseigneé*. La Penseé Sauvage.
Chiang, C.-W., & Yin, M. (2022). Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning models. *Proceedings of the 27th International Conference on Intelligent User Interfaces*, 148–161.
Chiu, T. K. F., & Sanusi, I. T. (2024). Define, foster, and assess student and teacher AI literacy and competency for all: Current status and future research direction. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100182
Davis, P. J., & Hersh, R. (1986). *Descartes' dream: The world according to mathematics*. Harcourt Brace Jovanovich.
de Olloqui, F., Andrade, G., & Herrera, D. (2015). *Inclusión financiera en América Latina y el Caribe: Coyuntura actual y desafíos para los próximos años* [Financial inclusion in Latin America and the Caribbean: The current situation and the challenges for the years to come]. Banco Interamericano de Desarrollo.
da Silva, A. S., Barbosa, M. T. S., de Souza Velasque, L., da Silveira Barroso Alves, D., & Magalhães, M. N. (2021). The COVID-19 epidemic in Brazil: How statistics education may contribute to unravel the reality behind the charts. *Educational Studies in Mathematics*, *108*(1–2), 269–289. https://doi.org/10.1007/s10649-021-10112-6
de Sousa, L. M. R. (2022). O lado obscuro dos algoritmos: como a promessa de liberdade da internet resultou em invasão de privacidade e ameaças à democracia [The Dark Side of Algorithms: How the Internet's promise of freedom has led to invasions of privacy and threats to democracy]. *Conexão ComCiência*, *2*(1), 1–8. https://revistas.uece.br/index.php/conexaocomciencia/article/view/5515
D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press. https://data-feminism.mitpress.mit.edu
Deutscher Ethikrat. (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz - Stellungnahme* [Man and machine – challenges of artificial intelligence – statement] https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine
Dvir, M., & Ben-Zvi, D. (2023). Informal statistical models and modeling. *Mathematical Thinking and Learning*, *25*(1), 79–99. https://doi.org/10.1080/10986065.2021.1925842
Donoho, D. (2017). 50 Years of Data Science. Journal of Computational and Graphical Statistics, 26(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734
Edmond, J., Horsley, N., Lehmann, J., & Priddy, M. (2021). *The Trouble With Big Data: How Datafication Displaces Cultural Practices*. Bloomsbury Academic.
Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. https://doi.org/10.52041/serj.v16i1.213
Engledowl, C., & Weiland, T. (2021). Data (mis)representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, *29*(2), 160–164. https://doi.org/10.1080/26939169.2021.1915215
English, L. D. (2016). Advancing mathematics education research within a STEM environment. In K. Fry, S. Dole, M. Goos, K. Maker, A. Bennison, & J. Visnovska (Eds.), *Research in mathematics education in Australasia 2012–2015* (pp. 353–371). Springer. https://doi.org/10.1007/978-981-10-1419-2_17
English, L. D. (2021). Mathematical and interdisciplinary modeling in optimizing young children's learning. In J. M. Suh, M. H. Wickstrom, & L. D. English (Eds.), *Exploring mathematical modeling with young learners* (pp. 3–24). Springer. https://doi.org/10.1007/978-3-030-63900-6_1

English, L. D., & Watson, J. (2018). Modelling with authentic data in sixth grade. *ZDM*, *50*(1), 103–115. https://doi.org/10.1007/s11858-017-0896-y
Erickson, T., & Chen, E. (2021). Introducing data science with data moves and CODAP. *Teaching Statistics*, *43*(S1). https://doi.org/10.1111/test.12240
Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education*, *12*(1). https://doi.org/10.5070/T5121038001
Erickson, T., & Engel, J. (2023). What goes before the CART? Introducing classification trees with Arbor and CODAP. *Teaching Statistics*, *45*(S1), S104–S113. https://doi.org/https://doi.org/10.1111/test.12347
Fagerlund, J., Palsa, L., & Mertala, P. (2025). Exploration of domains of educational purpose in K-12 data literacy education research. *Educational Research Review*, *46*. https://doi.org/10.1016/j.edurev.2024.100663
Fergusson, A. (2024). *Introducing a data science perspective on predictive modelling within a large introductory statistics course: Connecting research with practice* Presentation Paderborn Colloquium on Data Science and Artificial Intelligence at School. December 11, 2024, https://www.prodabi.de/colloquium/archiv
Fergusson, A., & Pfannkuch, M. (2024). Using Grayscale Photos to Introduce High School Statistics Teachers to Reasoning with Digital Image Data. *Journal of Statistics and Data Science Education*, *32*(4), 345–360. https://doi.org/10.1080/26939169.2024.2351570
Ferreira da Silva, A. F., & de Oliveira Souza, L. (2024). La interseccionalidad como forma de (re)pensar la educación estadística [Intersectionality as a way of (re)thinking statistical education]. *Revista De Educación Estadística*, *3*, 1–23. https://doi.org/10.29035/redes.3.1.8
Fleischer, Y., & Biehler, R. (2025). Exploring students' constructions of data-based decision trees after an introductory teaching unit on machine learning *ZDM Mathematics Education*, *57*(1). https://doi.org/10.1007/s11858-025-01663-6
Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and Learning Data-Driven Machine Learning with Educationally Designed Jupyter Notebooks. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.61
Fleischer, Y., Hüsing, S., Biehler, R., Podworny, S., & Schulte, C. (2022). Jupyter Notebooks for Teaching, Learning, and Doing Data Science. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. https://doi.org/10.52041/iase.icots11.T10E3
Fleischer, Y., Podworny, S., & Biehler, R. (2024). Teaching and Learning to Construct Data-Based Decision Trees Using Data Cards as the First Introduction to Machine Learning in Middle School. *Statistics Education Research Journal*, *23*(1). https://doi.org/10.52041/serj.v23i1.450
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach (Dordr)*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
Franklin, C., Bargagliotti, A. E., Case, C. A., Kader, G. D., Schaeffer, R. L., & Spangler, D. A. (2015). The statistical education of teachers. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/edu-set.pdf
Frau-Meigs, D. (2024). Algorithm Literacy as a Subset of Media and Information Literacy: Competences and Design Considerations. *Digital*, *4*(2), 512–528. https://doi.org/10.3390/digital4020026
Freire, P. (1970a). *Pedagogy of the oppressed*. Continuum.
Freire, P. (1970b). *Pedagogía del oprimido* [Pedagogy of the Oppressed]. Translator Jorge Mellado. Siglo veintiuno editores.
Fry, K., English, L., & Makar, K. (2024). Cognitive tuning in the STEM classroom: Communication processes supporting children's changing conceptions about data. *Mathematics Education Research Journal*, *36*, 67–89. https://doi.org/10.1007/s13394-023-00465-x
Gagrčin, E., Naab, T. K., & Grub, M. F. (2024). Algorithmic media use and algorithm literacy: An integrative literature review. *New Media & Society*, 14614448241291137.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, *70*(1), 1–25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics*, *111*(1), 5–28. https://doi.org/10.1007/s10649-022-10151-7

Gal, I., Nicholson, J., & Ridgway, J. (2022). A Conceptual Framework for Civic Statistics and Its Educational Applications. In J. Ridgway (Ed.), *Statistics for Empowerment and Social Engagement* (pp. 37–66). https://doi.org/10.1007/978-3-031-20748-8_3

Gal, I., & Ograjensek, I. (2018). Developing official statistics literacy: A proposed model and implications. In P. Kovács (Ed.), *Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat*. University of Szeged.

Geiger, V., Gal, I., & Graven, M. (2023). The connections between citizenship education and mathematics education. *ZDM – Mathematics Education*, *55*(5), 923–940. https://doi.org/10.1007/s11858-023-01521-3

Gibson, M. (2020). From deliberation to counter-narration: Toward a critical pedagogy for democratic citizenship. Theory & Research in Social Education, 48(3), 431–454. https://doi.org/10.1080/00933104.2020.1747034

Giordano, C. C., Pereira, F. A., & Souza, F. dos S. (2022). El desarrollo de la alfabetización estadística de los estudiantes de secundaria brasileña: un enfoque a través de las Estadísticas Cívicas [Developing statistical literacy among Brazilian high school students: A Civic Statistics approach]. *Revista Baiana De Educação Matemática*, *3*(01), e202212. https://doi.org/10.47207/rbem.v3i01.15516

Giroux, H. (1993). *La escuela y la lucha por la ciudadanía. La Pedagogía crítica de la época moderna* [School and the struggle for citizenship: Critical Pedagogy in the Modern Age]. Siglo veintiuno editores.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Green, A. (2024). *Artificial intelligence and the changing demand for skills in the labour market*. https://www.oecd.org/en/publications/artificial-intelligence-and-the-changing-demand-for-skills-in-the-labour-market_88684e36-en.html

Grillenberger, A., & Romeike, R. (2018). Developing a Theoretically Founded Data Literacy Competency Model. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education (WiPSCE '18), October 4–6, 2018, Potsdam, Germany*. ACM. https://doi.org/10.1145/3265757.3265766

Gulson, K. N., Sellar, S., & Webb, P. T. (2022). *Algorithms of education: How datafication and artificial intelligence shape policy*. University of Minnesota Press.

Gundlach, E., Maybee, C., & O'Shea, K. (2015). Statistical literacy social media project for the masses. *The Journal of Faculty Development*, *29*(2), 71–80.

Hazzan, O., & Mike, K. (2023). *Guide to Teaching Data Science: An Interdisciplinary Approach*. Springer. https://doi.org/https://doi.org/10.1007/978-3-031-24758-3

Höper, L., & Schulte, C. (2023). The data awareness framework as part of data literacies in K-12 education. *Information and Learning Sciences*. https://doi.org/10.1108/ils-06-2023-0075

IDSSP Curriculum Team. (2019). *Curriculum Frameworks for Introductory Data Science*. The International Data Science in Schools Project. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Introduction to Data Science (2022). *IDS Technology Suite*. https://www.mobilizingcs.org/introduction-to-data-science-curriculum/the-technology-suite-for-ids

Ilyenkov, Ė. V. (1977). *Dialectical logic: essays on its history and theory*. Moscow: Progress Publishers

Israel-Fishelson, R., Moon, P. F., Tabak, R., & Weintrop, D. (2023). Preparing students to meet their data: an evaluation of K-12 data science tools. *Behaviour & Information Technology*, 1–20. https://doi.org/10.1080/0144929x.2023.2295956

Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., Finzer, W., & Rosé, C. P. (2022). An empirical analysis of high school students' practices of modelling with unstructured

data. *British Journal of Educational Technology*, *53*(5), 1114–1133. https://doi.org/10.1111/bjet.13253
Jung, H., & Wickstrom, M. H. (2023). Teachers creating mathematical models to fairly distribute school funding. *The Journal of Mathematical Behavior*, *70*. https://doi.org/10.1016/j.jmathb.2023.101041
Kahn, J. (2020). Learning at the intersection of Self and Society: The Family Geobiography as a Context for Data Science Education. *Journal of the Learning Sciences*, *29*(1), 57–80. https://doi.org/10.1080/10508406.2019.1693377
Kahn, J. B., Peralta, L. M., Rubel, L. H., Lim, V. Y., & Herbel-Eisenmann, B. (2023). Notice, wonder, feel, act, and reimagine as a path toward social justice in data science education. *Educational Technology & Society*, 25 (4), 80–92. https://doi.org/10.30191/ETS.202210_25(4).0007
Kawakami, T., & Saeki, A. (2022). A framework for describing and analysing data-driven modelling activities in school mathematics: From the perspectives of mathematical and statistical models. *Journal of Science Education in Japan*, *46*(4), 421–437 [in Japanese]. https://doi.org/10.14935/jssej.46.421
Kawakami, T., & Saeki, A. (2024a). Extending data-driven modelling from school mathematics to school STEM education. In J. Anderson, & K. Makar (Eds.), *The contribution of mathematics to school STEM education: Current understandings* (pp. 221–239). Springer. https://doi.org/10.1007/978-981-97-2728-5_13
Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595–605). Springer. https://doi.org/10.1007/978-3-031-53322-8_49
Kazak, S., Fujita, T., & Turmo, M. P. (2023). Students' informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning*, *25*(1), 23–43. https://doi.org/10.1080/10986065.2021.1922857
Kim, A. Y., Ismay, C., & Chunn, J. (2018). The fivethirtyeight R Package: "Tame Data" Principles for Introductory Statistics and Data Science Courses. *Technology Innovations in Statistics Education*, *11*(1). https://doi.org/10.5070/T5111035892
Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, *16*(2), 191–212. https://doi.org/10.52041/serj.v16i2.190
Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305–325. https://doi.org/10.1007/s10649-013-9529-8
Lanouette, K., Cortes, K. L., Lopez, L., Bakal, M., & Wilkerson, M. H. (2024). Exploring climate change through students' place connections and public data sets. *Science Scope*, *47*(3), 18–25. https://doi.org/10.1080/08872376.2024.2340444
Lanouette, K. and Rivero, E. and Barton, J. and Bulalacao, N. and Lopez, M. L. and Cortes, K. and Roberto, C. and Gutiérrez, K. and Wilkerson, M. H. and Lee, H. and Stokes, D. and Finzer, W. and Erickson, T. and Petrosino, T. and Haldar, L. (2020). Writing data stories: Reauthoring scientific data through syncretic computational investigations in middle school science. In C. Matuk & S. Yoon (Orgs.) and J. Polman (Disc.), Data literacy for social justice. In *Proceedings of the 14th International Conference for the Learning Sciences (ICLS 2020).* Nashville, TN, USA: ISLS.
Leaning, M. (2017). *Media and information literacy: An integrated approach for the 21st century*. Chandos Publishing.
Lee, V. R., Drake, J., Cain, R., & Thayne, J. (2021). Remembering what produced the data: Individual and social reconstruction in the context of a "quantified self" elementary data and statistics unit. *Cognition and Instruction*, *39*(4), 367–408. https://doi.org/10.1080/07370008.2021.1936529
Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664–672. https://doi.org/10.3102/0013189X211048810
Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.41

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 229–259). Springer. https://doi.org/10.1007/978-3-319-66195-7
Lehrer, R., & Schauble, L. (Eds.). (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. Teachers College Press.
Lehrer, R., Wisittanawat, P., & Schauble, L. (2024). Designing for epistemic development. In Y. Li, Z. Zeng, & N. Song (Eds.), *Disciplinary and interdisciplinary education in STEM* (pp. 121–148). Springer. https://doi.org/10.1007/978-3-031-52924-5_7
Lehuedé, S. (2024). An alternative planetary future? Digital sovereignty frameworks and the decolonial option. *Big Data & Society*, *11*(1). https://doi.org/10.1177/20539517231221778
Lim, V. Y., Peralta, L. M. M., Rubel, L. H., Jiang, S., Kahn, J. B., & Herbel-Eisenmann, B. (2023). Keeping pace with innovations in data visualizations: A commentary for mathematics education in times of crisis. *ZDM – Mathematics Education*, *55*(1), 109–118. https://doi.org/10.1007/s11858-022-01449-0
Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.337672
López Carreño, M. D., López González, L. C., & Peña Suárez, M. A. (2022). *Sesgos y prejuicios de género en los procesos de selección de personal de línea media en empresas del Valle de Aburrá* [Gender bias and prejudice in mid-level selection processes in Aburrá Valley companies] (Master's thesis, Universidad EAFIT). https://hdl.handle.net/10784/30915
Louie, J. (2022). *Critical data literacy: Creating a more just world with data* National Academy of Sciences' Workshop on Foundations of Data Science for Students in Grades K–12, https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D16254F310D01BBDA873920E4EFB8151F2D8334181AA
Louie, J., Stiles, J., Fagan, E., Chance, B., & Roy, S. (2023). Building toward critical data literacy with investigations of income inequality. *Educational Technology & Society*, *25*(4), 142–163. https://doi.org/10.30191/ETS.202210_25(4).0011
Ma, R., Sanusi, I. T., Mahipal, V., Gonzales, J. E., & Martin, F. G. (2023). Developing Machine Learning Algorithm Literacy with Novel Plugged and Unplugged Approaches *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (pp. 298–304). https://doi.org/10.1145/3545945.3569772
Maass, K., Zehetmeier, S., Weihberger, A., & Flößer, K. (2023). Analysing mathematical modelling tasks in light of citizenship education using the COVID-19 pandemic as a case study. *ZDM Mathematics Education*, *55*(1), 133–145. https://doi.org/10.1007/s11858-022-01440-9
MacGillivray, H., Ridgway, J., & Gould, R. (2021). Preface. *Teaching Statistics*, *43*(S1). https://doi.org/10.1111/test.12282
Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. *ZDM – Mathematics Education*, *55*(5), 967–979. https://doi.org/https://doi.org/10.1007/s11858-022-01450-7
Marshall, B. H. (2022). *Data Conscience: Algorithmic Siege on Our Humanity*. John Wiley & Sons.
Martínez-Castro, C. A., & Zapata-Cardona, L. (2022). Formación inicial de profesores de estadística en una perspectiva crítica [Initial training of statistics teachers from a critical perspective]. In A. Salcedo & D. Diaz-Levicoy (Eds.) *Formación del profesorado para enseñar estadística: retos y oportunidades* (pp. 367–382). Universidad Católica de Maule.
Martínez-Castro, C. A., Zapata-Cardona, L., & Jones, G. L. (2023). Critical Citizenship in Statistics Teacher Education. In *Research on Reasoning with Data and Statistical Thinking: International Perspectives* (pp. 213–226). https://doi.org/10.1007/978-3-031-29459-4_17
Matuk, C., Knight, S., & DesPortes, K. (2022). Educational Technology & Society: Learning at the intersection of data literacy and social justice - Call for papers. *Educational Technology & Society*, *25*(4).
McLean, J. A., Brady, C., Jung, H., Dominguez, A., & Glancy, A.W. (2023). Seeing the forest for the trees: Investigating students' data moves in a citizen science based model-eliciting activity. In G. Greefrath, S. Carreira, G. A. Stillman (Eds.), *Advancing and consolidating mathematical*

modelling: Research from ICME-14 (pp. 193–204). Springer. https://doi.org/10.1007/978-3-031-27115-1_12
Mejía, A. F. (2020). La libertad de expresión en jaque. El panóptico del siglo XXI. Big data como amenaza para la democracia. A propósito del caso Cambridge Analytica [Freedom of expression under threat. The 21st century panopticon. Big data as a threat to democracy. On the Cambridge Analytica case]. *Universitas, 32,* 79–105. https://doi.org/10.20318/universitas.2020.5512
Mendoza-Chan, J., & Pee, L. G. (2024). Digital skilling of working adults: A systematic review. *Computers & Education*, *218*. https://doi.org/10.1016/j.compedu.2024.105076
Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). AI and education: A guidance for policymakers. UNESCO Publishing.
Mike, K., & Hazzan, O. (2022). Machine Learning for Non-Majors: A White Box Approach. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.45
Mostacedo-Marasovic, S.-J., Olsen, A. A., & Forbes, C. T. (2024). Supporting Secondary Students' Understanding of Earth's Climate System and Global Climate Change Using EzGCM: A Cross-Sectional Study. *Journal of Science Education and Technology*, *33*(2), 178–194. https://doi.org/10.1007/s10956-023-10067-8
Murillo, L. F. R., Wylie, C., & Bourne, P. (2023). Critical data ethics pedagogies: Three (non-rival) approaches. *Big Data & Society*, *10*(2). https://doi.org/10.1177/20539517231203666
National Academies of Sciences Engineering Medicine. (2023). *Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop*. https://doi.org/https://doi.org/10.17226/26852
National Center for Research on Evaluation, Standards and Student Testing (2017). *Introduction to data science: Crest findings, student impacts*. https://drive.google.com/file/d/0B1Nho2Ty0ZnYbE5wT1FCU2RtZExSN290WVJsM0htc0NqbzFn/view?resourcekey=0-VL7HSox62czWW_XB9gGXYw
NCTM, NSTA, ASA, NCSS, & CSTA. (2024). *Data Science A Joint Position of NCTM, NSTA, ASA, NCSS, and CSTA*. https://www.nctm.org/Standards-and-Positions/Position-Statements/Data-Science
Nicholson, J., Ridgway, J., & Gal, I. (2018). Understanding civic statistics: A conceptual framework and its educational applications. *A Product of the ProCivicStat Project*. http://IASE-web.org/ISLP/PCS
Niss, M. (2015). Prescriptive modelling: Challenges and opportunities. In G. Stillman, W. Blum, & M. Salett Biembengut (Eds.), *Mathematical modelling in education research and practice* (pp. 67–79). Springer. https://doi.org/10.1007/978-3-319-18272-8_5
Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling*. Routledge
Olari, V., & Romeike, R. (2024). Data-related concepts for artificial Intelligence education in K-12. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100196
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
Oxley, L., & Morris, P. (2013). Global citizenship: A typology for distinguishing its multiple conceptions. *British Journal of Educational Studies*, *61*(3), 301–325. https://doi.org/10.1080/00071005.2013.798393
Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419–437. https://doi.org/10.1177/1461444818799523
Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*, *34*(4), 361–388. https://doi.org/10.1080/07370008.2016.1210418
Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

Pimentel, D. R., Horton, N. J., & Wilkerson, M. H. (2022). Tools to support data analysis and data science in K-12 education. *National Academies of Science*. https://nicholasjhorton.github.io/K12-Data-Tools/static/data_tools.pdf

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM – Mathematics Education*, *50*(7), 1113–1123. https://doi.org/10.1007/s11858-018-0989-2

Podworny, S., Biehler, R., & Fleischer, Y. (2025). Young students' engagement with data to create decision trees. *ZDM – Mathematics Education*, *57*(1). https://doi.org/10.1007/s11858-024-01649-w

Podworny, S., Fleischer, Y., & Hüsing, S. (2022). Grade 6 Students' Perception and Use of Data-Based Decision Trees. In *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. https://doi.org/10.52041/iase.icots11.T2H3

Podworny, S., Fleischer, Y., Hüsing, S., Biehler, R., Frischemeier, D., Höper, L., & Schulte, C. (2021). Using data cards for teaching data based decision trees in middle school. In *21st Koli Calling International Conference on Computing Education Research (Koli Calling '21), November 18-21, 2021, Joensuu, Finland*. ACM. https://doi.org/10.1145/3488042.3489966

Podworny, S., Frischemeier, D., Dvir, M., & Ben-Zvi, D. (Eds.). (2024). *Minerva School 2022: Reasoning with data models and modeling in the big data era*. https://doi.org/10.17619/UNIPB/1-1815

Poling, L. L., & Weiland, T. (2020). Using an interactive platform to recognize the intersection of social and spatial inequalities. *Teaching Statistics*, *42*(3), 108–116. https://doi.org/10.1111/test.12234

Radinsky, J., & Tabak, I. (2022). Data practices during COVID: Everyday sensemaking in a high-stakes information. *British Journal of Educational Technology*, *53*(5), 1221–1243. https://doi.org/10.1111/bjet.13252

Raffaghelli, J. (2022). Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía [Data Literacy and Social Justice: An oxímoron? Responses from the counter-hegemony]. *Revista Izquierdas, 51,* 1-18. https://www.izquierdas.cl/images/pdf/2022/51/art17.pdf

Rathgeb, T., & Schmid, T. (2020). *JIM-Studie 2020 - Jugend, Information, (Multi-)Media, Basisstudie zum Medienumgang 12- bis 19-jähriger in Deutschland*. Medienpädagogischer Forschungsverbund Südwest.

Reigh, E., Escudé, M., Bakal, M., Rivero, E., Wei, X., Roberto, C., Hernández, D., Yada, A., Gutiérrez, K., & Wilkerson, M. H. (2022). Mapping racespace: Data stories as a tool for environmental and spatial justice. *Occasional Paper Series*, (48), 79-95. https://doi.org/10.58295/2375-3668.1452

Ridgway, J. (Ed.). (2022). *Statistics for empowerment and Social Engagement – Teaching Civic Statistics to Develop Informed Citizens*. Springer. https://doi.org/10.1007/978-3-031-20748-8

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: knowledge synthesis report.* http://hdl.handle.net/10222/64578

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology*, *53*(5), 1179–1201. https://doi.org/10.1111/bjet.13245

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2016). Teaching mathematics for spatial justice: Beyond a victory narrative. *Harvard Educational Review*, *86*(4), 556–579.

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016). Teaching mathematics for spatial justice: An investigation of the lottery. *Cognition and Instruction*, *34*(1), 1–26. https://doi.org/10.1080/07370008.2015.1118691

Rubel, L. H., & Nicol, C. (2020). The power of place: Spatializing critical mathematics education. *Mathematical Thinking and Learning*, *22*(3), 173–194. https://doi.org/10.1080/10986065.2020.1709938

Rubel, L. H., Nicol, C., & Chronaki, A. (2021). A critical mathematics perspective on reading data visualizations: Reimagining through reformatting, reframing, and renarrating. *Educational Studies in Mathematics*, *108*(1–2), 249–268. https://doi.org/10.1007/s10649-021-10087-4
Sander, I. (2024). *Critical Datafication Literacy*. transcript.
Sanusi, I. T. (2023). *Machine Learning Education in the K–12 Context.* Ph.D. thesis. Itä-Suomen yliopisto.
Schiller, A., & Engel, J. (2018). The importance of statistical literacy for democracy - Civic-Education through statistics. In P. Kovács (Ed.), *Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat*. University of Szeged.
Schwab-McCoy, A., Baker, C. M., & Gasper, R. E. (2021). Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years. *Journal of Statistics and Data Science Education*, *29*(sup1), S40–S50. https://doi.org/10.1080/10691898.2020.1851159
Shreiner, T. L. (2020). Data-literate citizenry: how US state standards address data and data visualizations in social studies. *Information and Learning Sciences*, *121*(11/12), 909–931. https://doi.org/10.1108/ils-03-2020-0054
Shreiner, T. L. (2023). Teaching Data Literacy for Civic Competence: The Social Studies Teacher's Crucial Role. *Social Education*, *87*(4), 262–269.
Shreiner, T. L., & Guzdial, M. (2024). Everyday Equitable Data Literacy is Best in Social Studies. In *Improving Equity in Data Science* (pp. 133–150). https://doi.org/10.4324/9781003364634-7
Siller, H.-S., Geiger, V., & Kaiser, G. (2024a). *Researching mathematical modelling education in disruptive times*. Springer.
Siller, H-S., Geiger, V., & Kaiser, G. (2024b). Researching mathematical modelling education in disruptive times—An introduction. In Siller, H-S., Geiger, V., Kaiser, G. (Eds.), *Researching Mathematical Modelling Education in Disruptive Times* (pp. 3–11). Springer. https://doi.org/10.1007/978-3-031-53322-8_1
Skovsmose, O. (2014). *Critique as Uncertainty.* The Montana Mathematics Enthusiast: Monograph Series in Mathematics Education. Information Age Publishing
Skovsmose, O. (2021). Mathematics and crises. *Educational Studies in Mathematics*, *108*, 369–383. https://doi.org/10.1007/s10649-021-10037-0
Ślosarski, B. (2023). Data arenas: The relational dynamics of data activism. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517231177617
Steen, L. A. (2001). *Mathematics and democracy: the case for quantitative literacy*. Woodrow Wilson National Fellowship Foundation.
Steffensen, L., Kacerja, S. (2021). Carbon Footprints Calculators and Climate Change. In F. K. S. Leung, G. A. Stillman, G. Kaiser, & K. L. Wong (Eds.), *Mathematical modelling education in East and West* (pp. 513–523). Springer. https://doi.org/10.1007/978-3-030-66996-6_43
Stillman, G. A., & Brown, J. (2023). Modeling the phenomenon versus modeling the data set. *Mathematical Thinking and Learning*, *25*(3), 270–295. https://doi.org/10.1080/10986065.2021.2013144
Stillman, G. A., Ikeda, T., Schukajlow, S., de Loiola Araújo, J., & Ärlebäck, J. B. (2023). Survey of interdisciplinary aspects of the teaching and learning of mathematical modelling in mathematics education. In G. Greefrath, S. Carreira, G. A. Stillman (Eds.), *Advancing and consolidating mathematical modelling: Research from ICME-14* (pp. 21–41). Springer. https://doi.org/10.1007/978-3-031-27115-1_2
Sukol, S. (2024a). *State of The Field: Data Science and Data Literacy Education in US K-12*. DataScience 4Everyone. https://www.datascience4everyone.org/_files/ugd/0a9d2b_f183f7139980484a9816319a99393bc9.pdf
Sukol, S. (2024b). *Beyond Borders 2024: Primary and Secondary Data Science Education Around the World*. Data Science 4 Everyone.
Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching Machine Learning in K–12 Classroom: Pedagogical and Technological Trajectories for Artificial Intelligence Education. *IEEE Access*, *9*, 110558–110572. https://doi.org/10.1109/access.2021.3097962

The Royal Society. (2017). *Machine learning: the power and promise of computers that learn by example*. The Royal Society.

Tironi, M. & Valderrama, M. (2021). Descolonizando los sistemas algorítmicos: diseño crítico para la problematización de algoritmos y datos digitales desde el Sur [Decolonizing algorithmic systems: Critical design for problematizing algorithms and digital data from the South]. *Palabra Clave*, *24*(3), e2432. https://doi.org/10.5294/pacla.2021.24.3.2

Toomey, D. (2017). *Jupyter for Data Science - Exploratory analysis, statistical modeling, machine learning, and data visualization with Jupyter*. Packt Publishing.

Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, 33, 1–67.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley

Tygel, A.F., & Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics, 12*(3), 108–121. https://doi.org/10.15353/joci.v12i3.3279

UNESCO. (2005). *Education for all: Literacy for Life*. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54676/HFRH4626

Van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2021). Statistical modeling processes through the lens of instrumental genesis. *Educational Studies in Mathematics*, *107*(2), 235–260. https://doi.org/10.1007/s10649-020-10023-y

Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space. *Journal of the Learning Sciences*, *29*(1), 127–153. https://doi.org/10.1080/10508406.2019.1693378

Valente, J. & Grohmann, R. (2024). Critical data studies with Latin America: Theorizing beyond data colonialism. *Big Data & Society 11*(1), 1-13. https://doi.org/10.1177/20539517241227875

Villa Villa, J. (2020). *Discriminación racial en Latinoamérica: Análisis de los efectos de la raza sobre el empleo* [Racial discrimination in Latin America: Analysis of the effects of race on employment] (Bachelor's thesis, Universidad EAFIT). Escuela de Economía y Finanzas. https://repository.eafit.edu.co/bitstreams/99ac45ea-c53d-4b76-8190-222ec80274ea/download

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., & Wagenknecht, K. (Eds.). (2021). *The science of citizen science*. Springer. https://doi.org/10.1007/978-3-030-58278-4

Wallman, K. K. (1993). Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association*, *88*(421), 1–8. https://doi.org/10.1080/01621459.1993.10594283

Watson, J. M. (2006). *Statistical Literacy at School*. Lawrence Erlbaum.

Watson, J., Fitzallen, N., & Chick, H. (2020). What is the role of statistics in integrating STEM education? In J. Anderson, & Y. Li (Eds.), *Integrated approaches to STEM education* (pp. 91–115). Springer. https://doi.org/10.1007/978-3-030-52229-2_6

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, *96*(1), 33–47. https://doi.org/10.1007/s10649-017-9764-5

Weiland, T. (2019). The contextualized situations constructed for the use of statistics by school mathematics textbooks. *Statistics Education Research Journal*, *18*(2), 18–38. https://doi.org/10.52041/serj.v18i2

Westheimer, J., & Kahne, J. (2004). What kind of citizen? The politics of educating for democracy. *American educational research journal*, *41*(2), 237–269.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Wilkerson, M. H., & Laina, V. (2018). Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM*, *50*(7), 1223–1235. https://doi.org/10.1007/s11858-018-0974-9

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2022). Exploring variability during data preparation: A way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, *24*(4), 312–330. https://doi.org/10.1080/10986065.2021.1922838

Wilkerson, M. H., & Polman, J. L. (2020). Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the Learning Sciences*, *29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

World Ecomomic Forum. (2019). *Data science in the new economy: A new race for talent in the Fourth Industrial Revolution* https://www3.weforum.org/docs/WEF_Data_Science_In_the_New_Economy.pdf

Zapata-Cardona, L. (2023). The possibilities of exploring nontraditional datasets with young children. *Teaching Statistics*, *45*(S1), S22–S29. https://doi.org/10.1111/test.12349

Zapata-Cardona, L., & Martínez-Castro, C. A. (2023). Statistical modeling in teacher education. *Mathematical Thinking and Learning*, *25*(1), 64–78. https://doi.org/10.1080/10986065.2021.1922859

Zieffler, A., Justice, N., delMas, R., & Huberty, M. D. (2021). The Use of Algorithmic Models to Develop Secondary Teachers' Understanding of the Statistical Modeling Process. *Journal of Statistics and Data Science Education*, *29*(1), 131–147. https://doi.org/10.1080/26939169.2021

Statistics and data science education as a vehicle for empowering citizens

(Survey 3, presented at ICME 15, Sydney, July 10, 2024)

Rolf Biehler (Chair)

Takashi Kawakami, Erna Lampen, Travis Weiland, Lucía Zapata-Cardona

Illustration of our title by DALL-E

Team members

Rolf Biehler
Paderborn University
Germany

Takashi Kawakami,
Utsunomiya University
Japan

. Erna Lampen,
Stellenbosch University
South Africa

Travis Weiland
University of Houston
United States

Lucía Zapata-Cardona
Universidad De Antioquia
Colombia

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)

2. **Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe** (Travis Weiland)

3. **Critical perspectives on data literacy emerging from Latin America** (Lucía Zapata-Cardona)

4. **Joint discourse between mathematical modeling and statistics/data science communities** (Takashi Kawakami)

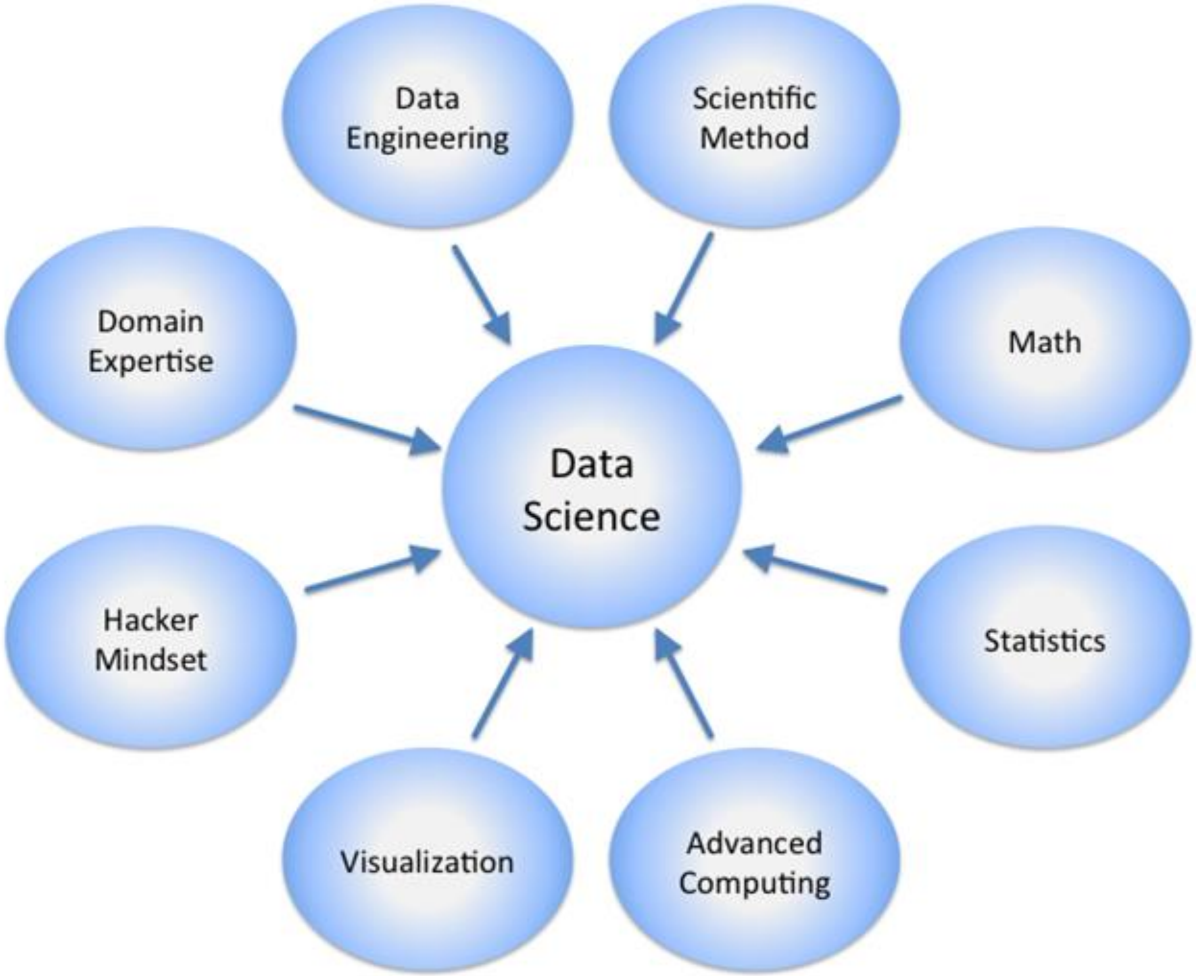
5. **What can mathematics/statistics education contribute to Artificial/Machine Learning literacy** (Rolf Biehler)

6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

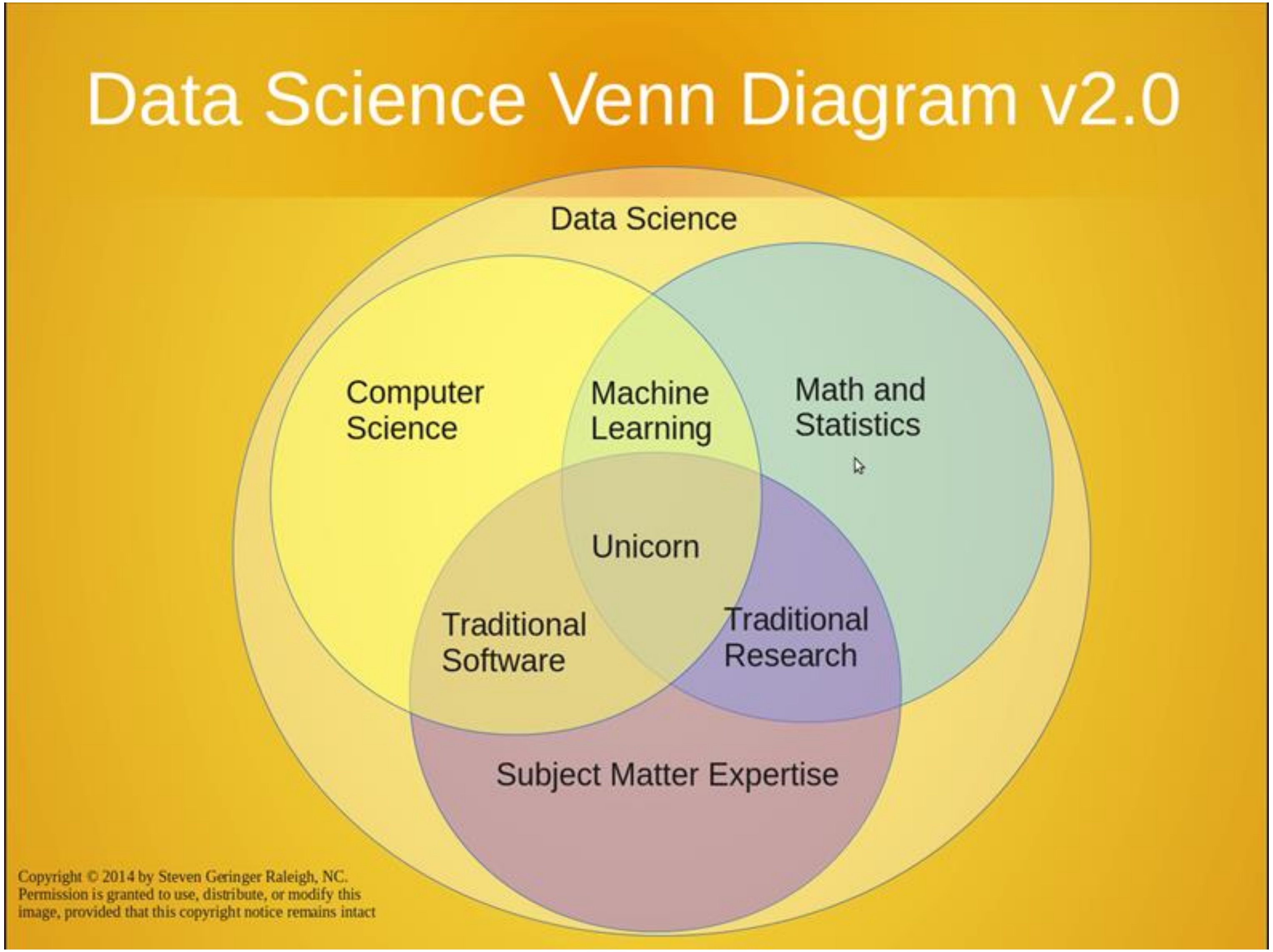
Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)
2. **Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe** (Travis Weiland)
3. **Critical perspectives on data literacy emerging from Latin America** (Lucía Zapata-Cardona)
4. **Joint discourse between mathematical modeling and statistics/data science communities** (Takashi Kawakami)
5. **What can mathematics/statistics education contribute to Artificial/Machine Learning literacy** (Rolf Biehler)
6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

What is data science?



https://en.wikipedia.org/wiki/File:DataScienceDisciplines.png



Data Science Venn diagram (http://www.anlytcs.com/2014/01/data-science-venn-diagram-v20.html). Copyright © 2014 by Steven Geringer Raleigh, NC. Permission is granted to use, distribute, or modify this image, provided that this copyright notice remains intact.

What should be taught in data science education?

Data Science

A Joint Position of NCTM, NSTA, ASA, NCSS, and CSTA

(April 2024)

- data science bridges disciplines
- should be taught across the curriculum in K-12 schools

Declarations or Guiding Principles

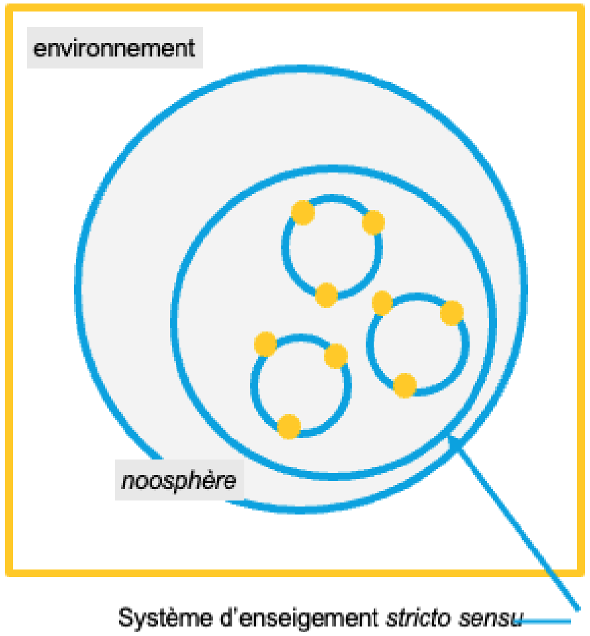
1. Data science is contextual and interdisciplinary
2. Data science is an investigative process
3. Data science understandings and experiences are for everyone
4. Data science educators must develop and practice ethical uses of data

Data Science

A Joint Position of NCTM, NSTA, ASA, NCSS, and CSTA

A fundamental question

New challenge for the didactic transposition of knowledge into the educational system



Chevallard, Y. (1985). *La Transposition Didactique: Du Savoir Savant au Savoir Enseigneé*. La Penseé Sauvage, p.24 (figure reproduced similar to origin)

The dilemma of noospherians (Didacticians of mathematics)

- scholarly knowledge is interdisciplinary, dynamic and unstable
- use and discourse in society is also relevant, not only knowledge in the disciplines

Data in society: New trends

- Changes in what counts as data: images, text, webpages …
- More data in the media (interactive dashboards), data journalism
- Data-driven digital artefacts
    - search engines
    - streaming services
    - social media
    - wearables
    - fitness trackers
    - news portals with content curation
- Datafication of scientific disciplines and societal practices
- Data-based algorithmic decision making
    - data-driven student monitoring in education
    - surveillance

Sage Journals Big Data & Society

New Data (interactive) visualizations in the media

- More data and new visualizations in classical media
- Websites with embedded interactive visualizations and dashboards
- Emergence of data journalism
- Incorporated within argumentative texts

AI and Data Science for Social Good

https://www.datascienceforsocialgood.org

https://www.mckinsey.com/capabilities/quantumblack/our-insights/ai-for-social-good

AI and Data Science: Risk and Dangers

International Red Cross is concerned:

“What you need to know about artificial intelligence in armed conflict”

“The advance of artificial intelligence (AI) for military purposes raises profoundly worrying questions for humanity. We take a look at some of the key questions and concerns surrounding the use of AI, especially machine learning, in armed conflict.”

https://www.icrc.org/en/document/what-you-need-know-about-artificial-intelligence-armed-conflict

Dangers and Risks of (data driven) AI and Data Technology

- Danger to the individual through statistical stratification
- Protection of privacy and autonomy versus threats from surveillance
- Bias and discrimination
- Transparency and traceability - control and responsibility

Selection from the 10 cross-cutting issues of the report

Deutscher Ethikrat (Ed.). (2023). *Stellungnahme: Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz.*

German Ethics Council: *Man and machine - Challenges posed by Artificial intelligence*. Position Paper (403 (!) pages)

(Conflicting) Goals of education

Who determines/constraints the goals and content of education? politicians? private companies? humanistic educators?

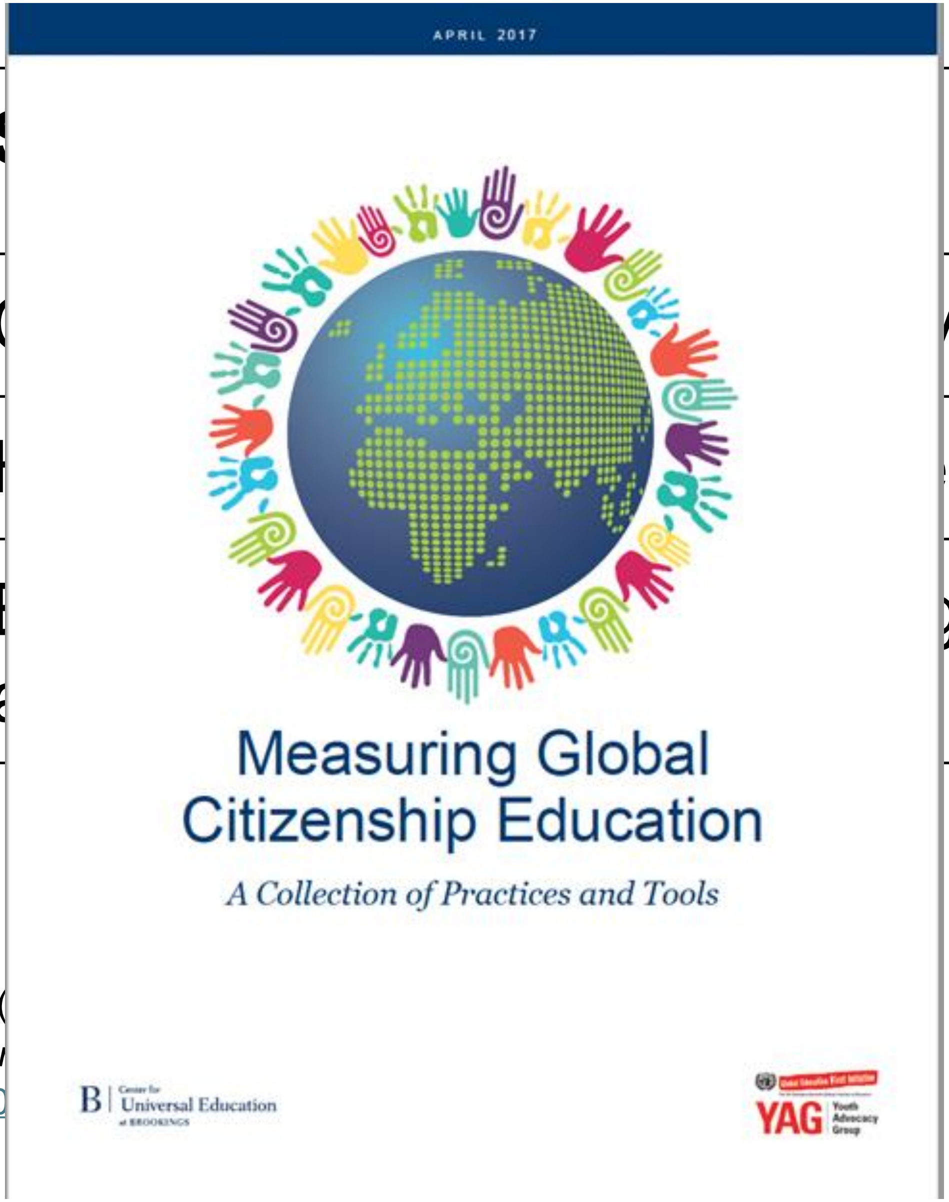
Distinctions in citizenship education

Types of citizen	**Sample action**
Personally responsible	Contributes food to a food drive
Participatory	Helps to organize a food drive
Justice oriented	Explores why people are hungry and acts to solve root causes

Westheimer, J., & Kahne, J. (2004). What Kind of Citizen? The Politics of Educating for Democracy. *American Educational Research Journal*, *41*(2), 237-269. https://doi.org/10.3102/00028312041002237

Distinctions in citizenship education

Types of citizen	S	
Personally responsible	C	ve
Participatory	H	e
Justice oriented	E a	gry and



Westheimer, J., & Kahne, J. (
for Democracy. *Amer*
https://doi.org/10.310

Reviewing the debates on data literacies

Data Literacy as closely related to statistical literacy

- Statistical Literacy (Gal 2002, 2018)
- Official statistical literacy (Gal 2018)
- Data literacy is statistical literacy (Gould 2017)
- Civic Statistical literacy (Ridgway (ed.) 2022, Gal et al 2022)
- Critical Data /Statistical literacy (Louie 2022, Weiland 2017, 2019, Martínez-Castro et al., 2023)

- Probabilistic literacy (Álvarez-Arroyo, Batanero, and Gea 2024)

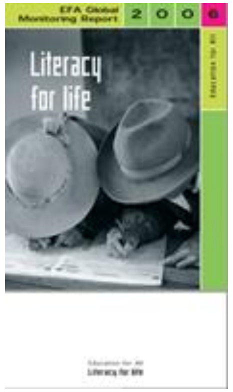
Data Literacy and mathematical modelling (MM) education

- Data: central in creating and validating a mathematical model
- Integrated approaches using MM and data
  - Required by global and disruptive events such as the COVID pandemic, climate change, sustainability etc. (e.g., Siller, Geiger, & Kaiser, 2024)
  - MM as related to citizenship education (Geiger, Gal and Graven 2023)
- MM more important in data science now due to
  - machine learning (predictive modeling, often also prescriptive modeling) (e.g., Biehler, 2022; Podworny, Frischemeier, Dvir, & Ben-Zvi, 2024)

Joint work at the boundary MM/DS

(e.g., Ärlebäck & Kawakami, 2023; Kazak, Fujita, & Turmo, 2023; Lehrer & English, 2018; Makar, Fry, & English, 2023; Gal & Geiger, 2022, Podworny, Frischemeier, Dvir, Ben-Zvi 2024)

Data literacy as part of



- General literacies (UNESCO 2005)
- “Digital Humanities Literacy” (Abner 2020)
- Media, news, and information literacy (Schield 2005, Leaning 2017)
- Digital literacy (Mendoza-Chan & Pee 2024).
- AI literacy (Almatrafi 2024), machine learning algorithm literacy (Ma et al 2023)

Data literacies as shaped by subject-specific epistemologies

New datafication in many disciplines (Gulson, Sellar, and Webb 2022)

- Data practices in the natural, earth and life sciences
  - Citizen Science (Vohland et al, 2021)
  - Socio-scientific problem solving (Mostacedo-Marasovic et al 2024)
- Data literacy as related to social sciences (Shreiner 2020, 2023. Shreiner & Guzdial 2024)
  - and their data practices
  - and their reflection on data practices in society

Further concepts of data literacy and data-related competencies and attitudes

- Critical datafication literacy (Sander 2024)
- Personal data literacy (Pangrazio & Selwyn 2018)
- Data awareness (Höper & Schulte 2023)
- Data acumen (Bargagliotti et al. 2020)
- Data conscience (Marshall 2022)
- Data ethics (Murillo, Wylie, & Bourne 2023)
- Data activism (Ślosarski 2023)
- Data feminism (D'Ignazio & Klein, 2020)

UNESCO. (2005). *Education for all: Literacy for Life*. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54676/HFRH4626

Data science education: Projects and frameworks

Overview of projects

- National Academies of Sciences Engineering Medicine. (2023). Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop. https://doi.org/https://doi.org/10.17226/26852
- Sukol, S. (2024a). *Beyond Borders 2024: Primary and Secondary Data Science Education Around the World*. Data Science 4 Everyone.
- Sukol, S. (2024b). State of The Field: Data Science and Data Literacy Education in US K-12. DataScience 4Everyone.

Frameworks

- IDSSP Curriculum Team. (2019). *Curriculum Frameworks for Introductory Data Science*. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Journal Special Issues on Data (AI) literacy and Data Science Education

1. Journal of the Learning Sciences 2020
2. Teaching Statistics 2021
3. Statistics Education Research Journal 2022
4. Educational Technology and Society 2022
5. Information and Learning Sciences 2024
6. Computers and Education Open 2024

(1) eds. Wilkerson, Polman, (2) eds. MacGillivray, Ridgway, Gould,(3) eds. Biehler, deVeaux, Engel, Kazak, Frischemeier, (4) eds. Matuk, Knight, Desportes (5) eds. Acker, Bowler, Pangrazio: 3 double issues (6) eds. Chiu & Sanusi

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)

2. **Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe** (Travis Weiland)

3. **Critical perspectives on data literacy emerging from Latin America** (Lucía Zapata-Cardona)

4. **Joint discourse between mathematical modeling and statistics/data science communities** (Takashi Kawakami)

5. **What can mathematics/statistics education contribute to Artificial Intelligence/Machine Learning literacy** (Rolf Biehler)

6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

2. **Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe** (Travis Weiland)

Goals

1. Present new emerging frameworks for considering data literacies for citizenship and humanizing education
2. Overview the diversity of work around data literacies for citizenship coming out of the U.S. and Europe
3. Provide a some examples of different types of projects coming out of the U.S. and Europe related to themes and frameworks in the field

Statistics, Data Science, Data Literacies and Citizenship Education

Statistical Literacy scholarship has used the demands of citizenship as a rationale for decades (Gal, 2002; Wallman, 1993)

Recent scholarship in data science education, data literacy, critical data literacy draw on similar rationales (Bargagliotti et al., 2020; Gould, 2017; Lee, Wilkerson, et al., 2021; Louie, 2022; Philip et al., 2013)

What kind of citizenship are we talking about? Personally responsible, participatory, justice oriented, global, spectator, referee, player, or some blending of these?

Conceptual Model for Civic Statistics

Civic Statistics, i.e. statistics and quantitative evidence about key social phenomena that permeate civic life, such as migration, demographic change, crime, employment and poverty.

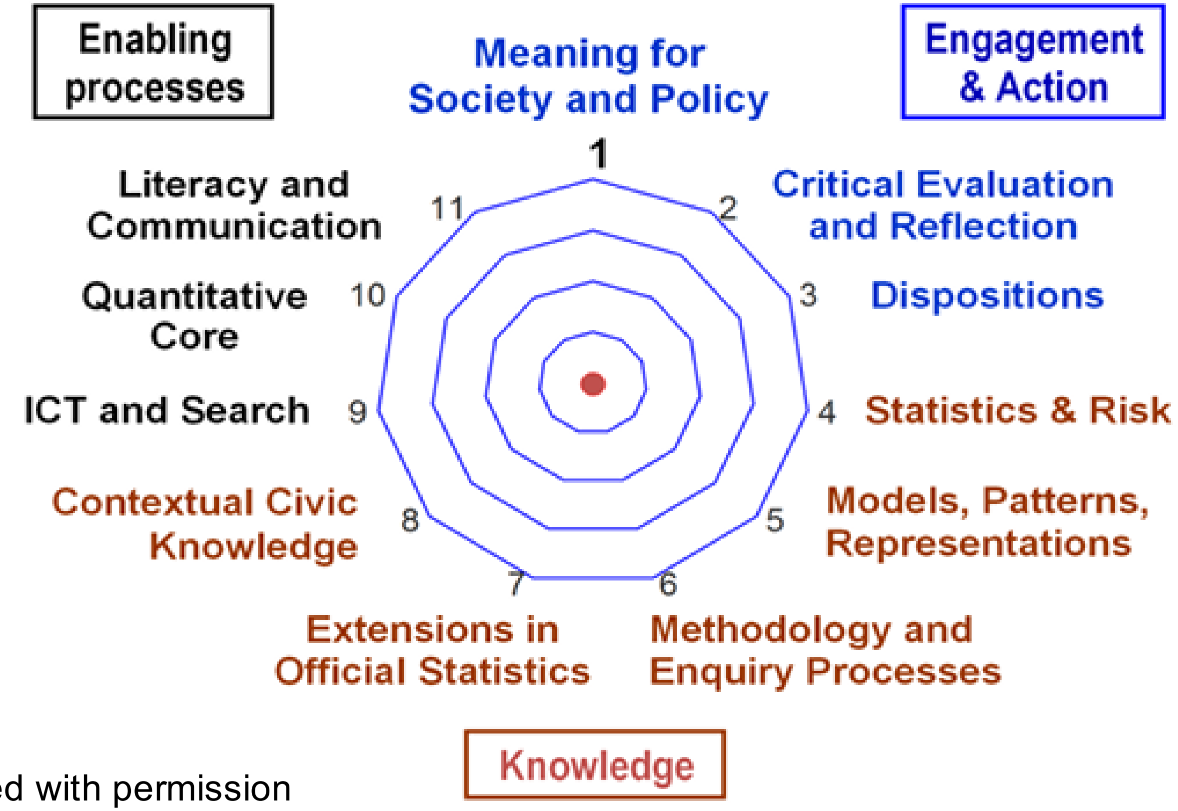


Diagram used with permission

Figure 4: A Conceptual Model for Civic Statistics

Suggested Citation: Nicholson, J., Gal, I., & Ridgway, J. (2018). Understanding Civic Statistics: A Conceptual Framework and its Educational Applications. A product of the ProCivicStat Project. Retrieved (Date) from: http://IASE-web.org/ISLP/PCS

Humanistic Stance Toward K–12 Data Science Education

A growing focus on humanistic approaches to STEM education broadly coming in particular from the learning sciences

A key driving frame comes from Lee, Wilkerson, and Lanouette who look at data practices through layers

A Call for a Humanistic Stance Toward K–12 Data Science Education

Victor R. Lee, Michelle Hoda Wilkerson, and Kathryn Lanouette View all authors and affiliations

Volume 50, Issue 9 | https://doi.org/10.3102/0013189X211048810

Framing: Data Feminism (D'Ignazio & Klein, 2020)

- New theoretical frontiers drawing upon feminism
- The seven principles of data feminism are:
  - **Examine power**
  - **Challenge power**
  - **Elevate emotion and embodiment**
  - **Rethink binaries and hierarchies**
  - **Embrace Pluralism**
  - **Consider Context**
  - **Make Labor Visible**
- Several projects working in this space as summarized by Lee et al. (2022).
- A key area several have made inroads that is especially unique is around the consideration of emotion and embodiment (J. B. Kahn et al., 2023; Lim et al., 2023; Radinsky & Tabak, 2022)

Survey of Scholarship

- Wide range of approaches including conceptual, theoretical, and empirical
- Approaches and Scale
  - Most work is **qualitative** or rhetorical in nature
  - Most projects are **small**, with a class or small group of kids for maybe a unit of instruction
  - Much of this work is taking a more critical or humanistic approaches
- Settings
  - More work occurring in **informal education** settings than formal education settings
  - Work is occurring in many settings outside of mathematics education, perhaps even more so than in mathematics education settings
- Citizenship
  - much of the recent work seems to align more with **participatory** and **justice-oriented views** of citizenship looking to support citizens in becoming **players**

Themes in the Literature

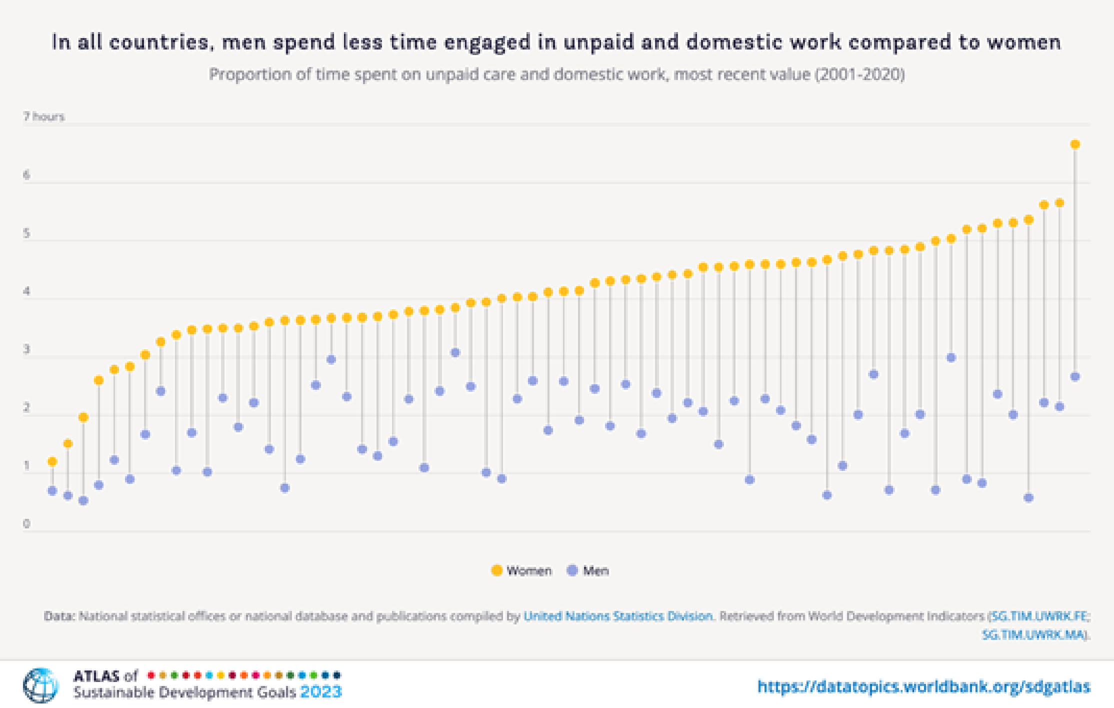
- Reading the world with data (i.e. making sense of the data based communication of others)
  - data viz
  - data journalism
- Writing the world with data (i.e. using data practices to investigate the world around us)
  - data investigations
  - data stories
- Data Structures and Handling (i.e. everything we do to the data to move from raw data to analysis)
  - data moves
  - tidy data
- Technology (i.e. including the development, interaction with, and learning from)
  - new professional tools are driving new demands for learning
  - new dynamic learning tools are allowing students to interact with and learn about data in new ways

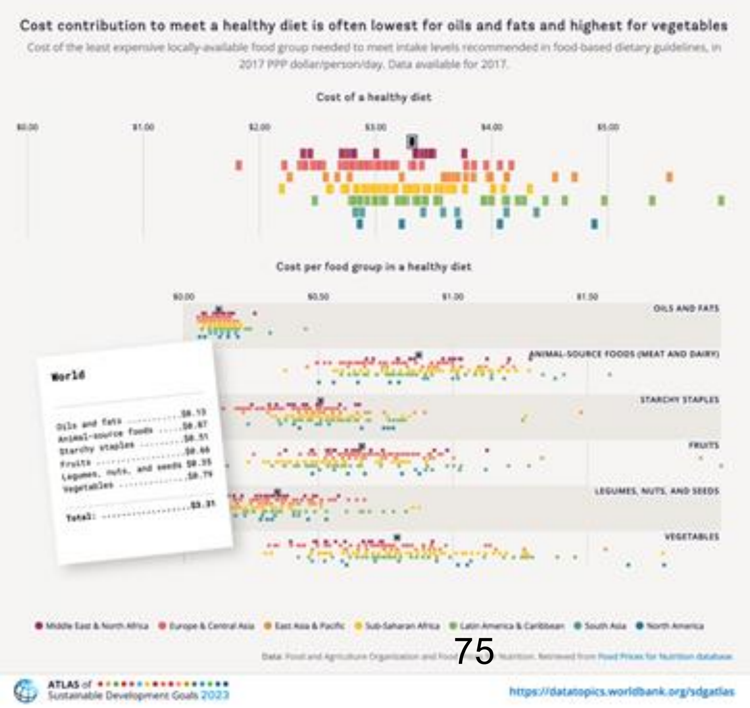
Reading the World through Data

Data Viz is now dynamic, interactive, and connected in ways it never was before opening up new possibilities and new demands for citizens

Reading new types of data visualization and with new lens on media and

(da Silva et al., 2021; Engledowl & Weiland, 2021; Gundlach et al., 2015; J. B. Kahn et al., 2023; Lim et al., 2023; Philip et al., 2016; Rubel et al., 2021b, 2021a; Shapiro et al., 2023)



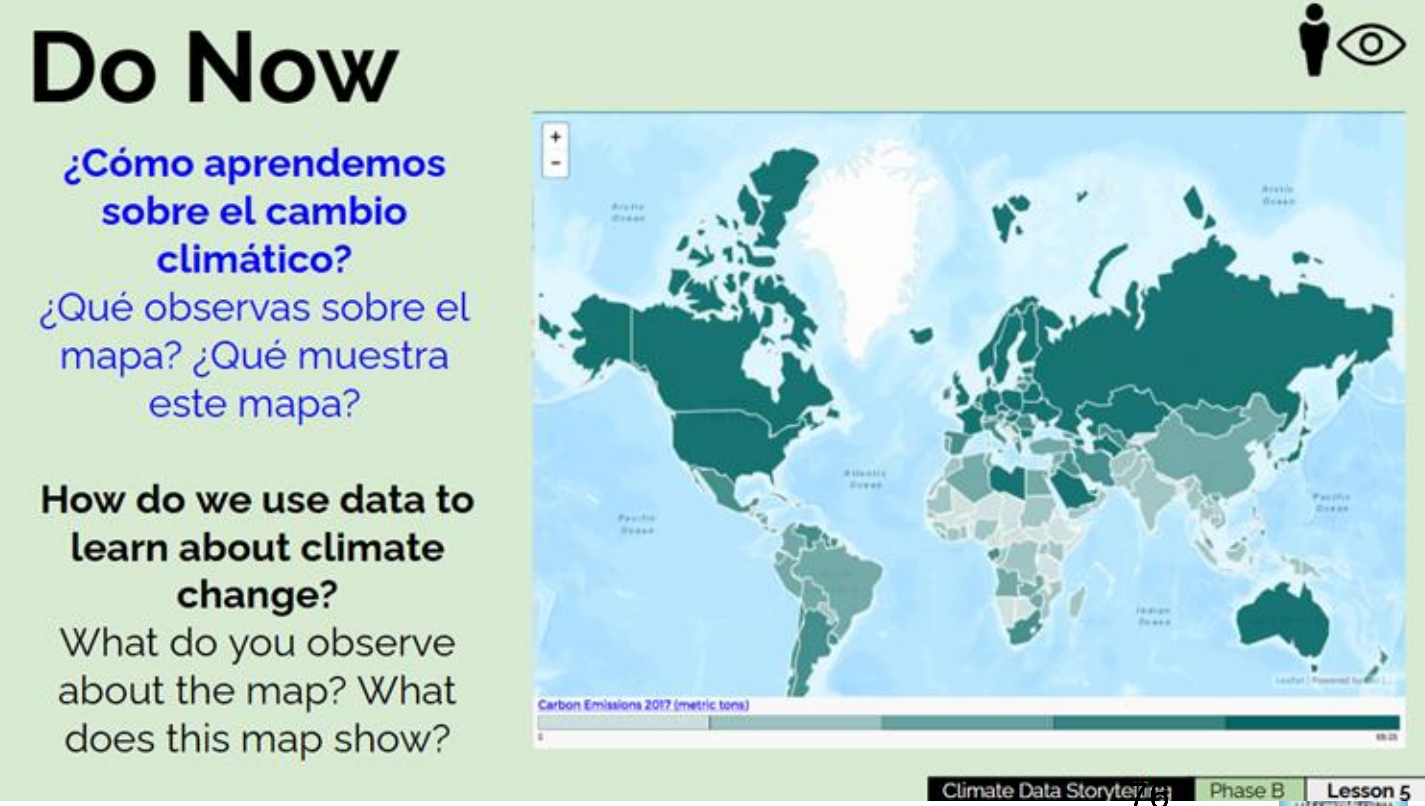


Data visualizations from World bank report
https://datatopics.worldbank.org/sdgatlas?lang=en

Reading the World through Data

In particular, spatial data and ideas of spatial justice are being seen frequently in the literature considering relationships to space

(Lanouette, et al., 2024; Poling & Weiland, 2020; Reigh et al., 2022; Rubel et al., 2017; Rubel, Hall-Wieckert, et al., 2016; Rubel, Lim, et al., 2016; Rubel & Nicol, 2020)

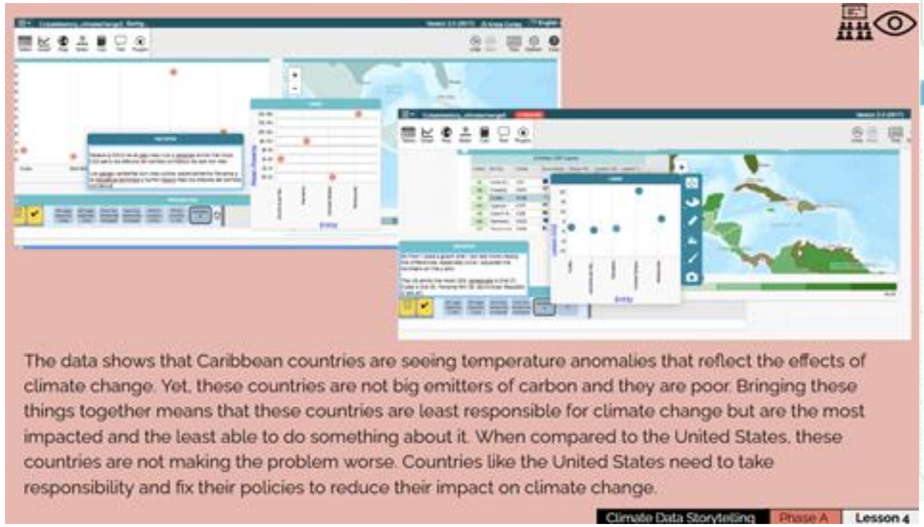


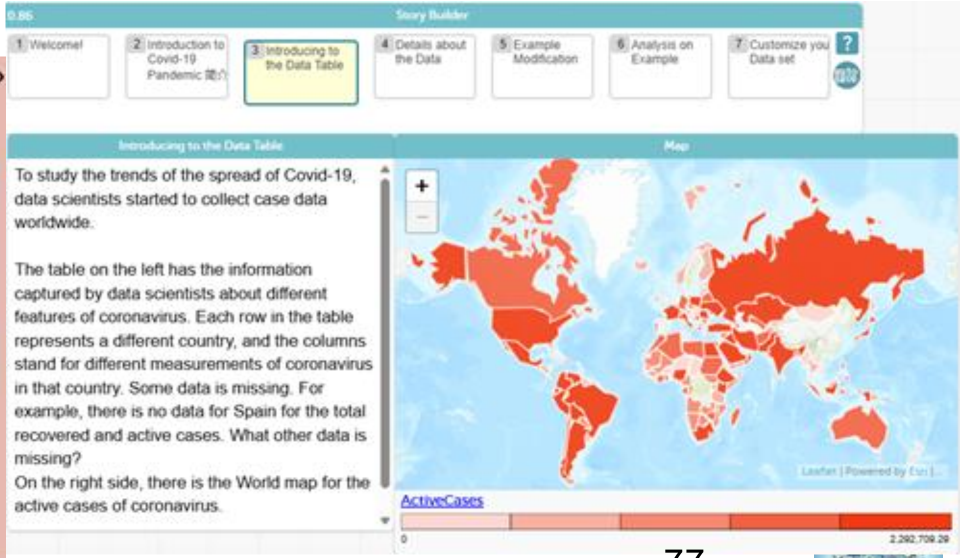
Screenshot from Writing Data Stories https://docs.google.com/presentation/d/1I6iCM8W_LC8ZCh2mBB70cdbEGPlgOXopSz0SHqmRDxA/edit?usp=sharing

Writing the World through Data

Data investigations (Lee et al., 2022), statistical investigations (Bargagliotti et al., 2020; Franklin et al., 2015), or statistical enquiry (Wild & Pfannkuch, 1999) are at the core of learning experiences

Data storytelling (Kahn, 2020; Reigh et al., 2022; Wilkerson & Laina, 2018) or data narratives (Radinsky, 2020)





Writing Data Stories https://docs.google.com/presentation/d/1I6iCM8W_LC8ZCh2mBB70cdbEGPlgOXopSz0SHqmRDxA/edit?usp=sharing

Writing the World through Data

Students are provided opportunities to engage in investigating and telling stories about a **wide range of contexts and issues** including:

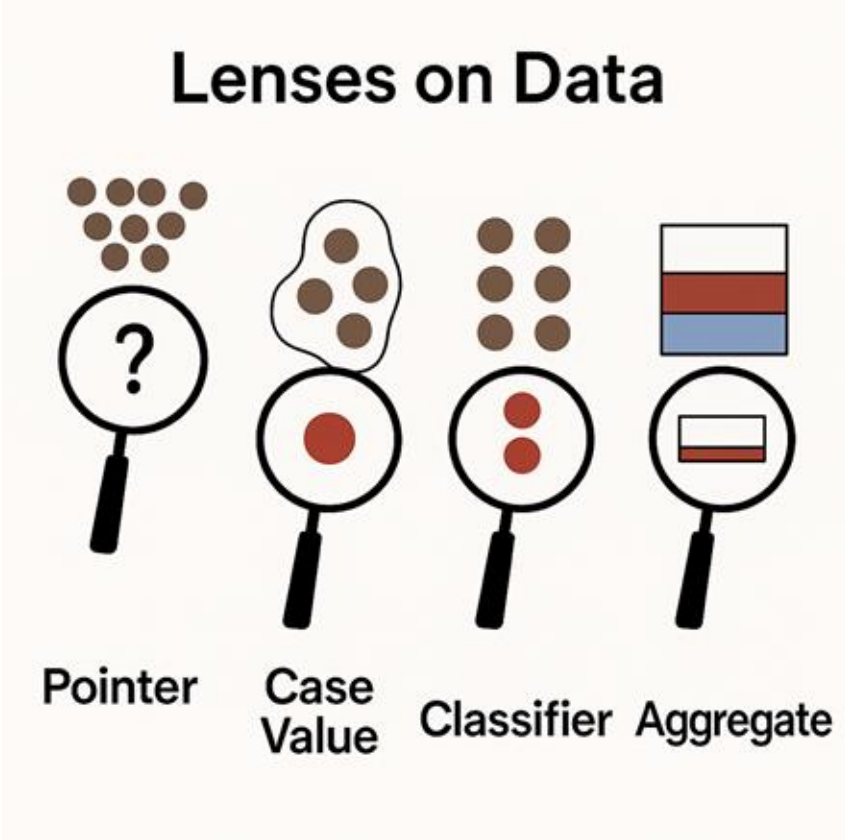
- urban planning
- participatory mapping
- citizen science,
- alternative banking institutions
- lottery sales
- comparing pay based on sex
- racial climate justice
- COVID-19
- socio scientific issues
- social issues, etc.

(Aridor et al., 2023; Engel, 2017; Engledowl & Weiland, 2021; Kahn et al., 2023; Kahn, 2020; Lanouette et al., 2024; Lee et al., 2022; Lim et al., 2023; Louie et al., 2023; Nicholson et al., 2018; Reigh et al., 2022; Rubel et al., 2021; Rosenberg et al., 2022; Rubel, Hall-Wieckert, et al., 2016; Rubel, Lim, et al., 2016; Shreiner & Guzdial, 2022; Van Wart et al., 2020; Wilkerson & Laina, 2018)

Created using ChatGTP

Data Structures and Handling

- Data lenses (Konold et al., 2015; Konold, Finzer, et al. 2017)
- Making sense of publicly available data requires making sense of data structures and preparing the data for exploration, analysis, and modeling in different ways (Poling & Weiland, 2020; Wilkerson et al., 2022; Wilkerson & Laina, 2018)



U.S. Government from Data.gov

Data Structures and Handling

- Dealing with unstructured data (Jiang et al., 2022)
- Hierarchical data structures (Erickson et al., 2017; Konold, et al., 2017)
- Tame data (Kim et al., 2018)
- Tidy Data (Wickham, 2014)
- Data moves (Erickson et al., 2019; Erickson & Chen, 2021)
  - Joining
  - Grouping
  - Calculating
  - Summarizing
  - Filtering

Technology

- Historically *Fathom* and *TinkerPlots* were groundbreaking for supporting reasoning about data and chance
- More recently CODAP has emerged in many research projects and continues to grow in features and functionality
- Other tools including:
    - Professional tools like: Excel, R/R studio, Python
        - Educational adaptations using Jupyter notebooks
    - Tools designed for learning like: TUVA, Desmos, curriculum specific web based apps

Example: Introduction to Data Science Project

- Long term curriculum development project in the U.S.
- How to introduce secondary students to Data Science in the mathematics curriculum
- Teaching students to reason with, and think critically about, data in all forms
- Curriculum is paired with professional development for teachers
- Adoption is being scaled

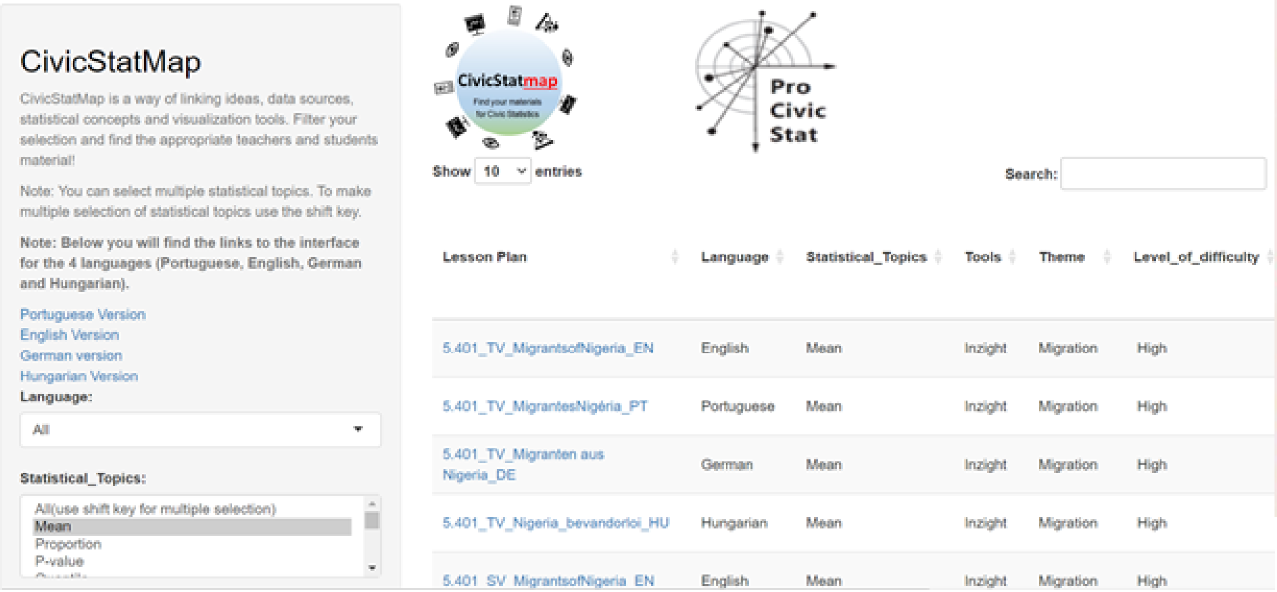
Curriculum IDS

- Participatory Sensing: An approach to data collection and interpretation in which individuals, acting alone or in groups, use their personal mobile devices and web services to systematically explore interesting aspects of their worlds, ranging from health to culture.
- Year long curriculum combines mathematics with computer science through the use of R/Rstudio
- Interactive Dashboard

Example: ProCivicStat Project

Large scale project across many countries in Europe and Israel building both theoretical frameworks, empirical research, and producing curricular resources

Pro Civic Stat



Durham University

אוניברסיטת חיפה University of Haifa

PH Ludwigsburg University of Education

PADERBORN UNIVERSITY

U.PORTO FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

SZTE UNIVERSITY OF SZEGED

ProCivicStats Project https://iase-web.org/islp/pcs/

Promoting civic engagement via explorations of evidence

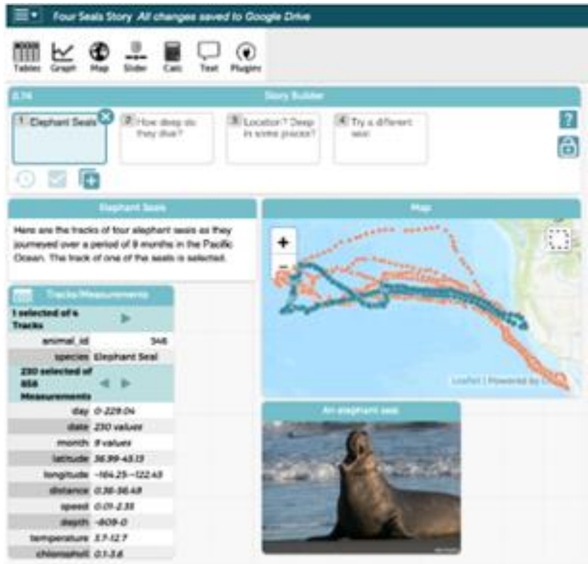
Wide range of resources primarily aimed at secondary and post-secondary audiences that have also been the basis of a wide range of research projects

- Conceptual framework
- Database of teaching and learning material
- Sample lesson plans
- Datasets
- Sample Syllabuses
- Review of dynamic visualisation tools

Example: Writing Data Stories

Data Writing Stories

- Project across multiple partners in U.S.
- Integrating computational data analysis into the middle school science curriculum in a longitudinal, interdisciplinary way
- Draws from the computer and data sciences, literacy studies, statistics, and science education
- Project involves, curriculum, resource, and tool development

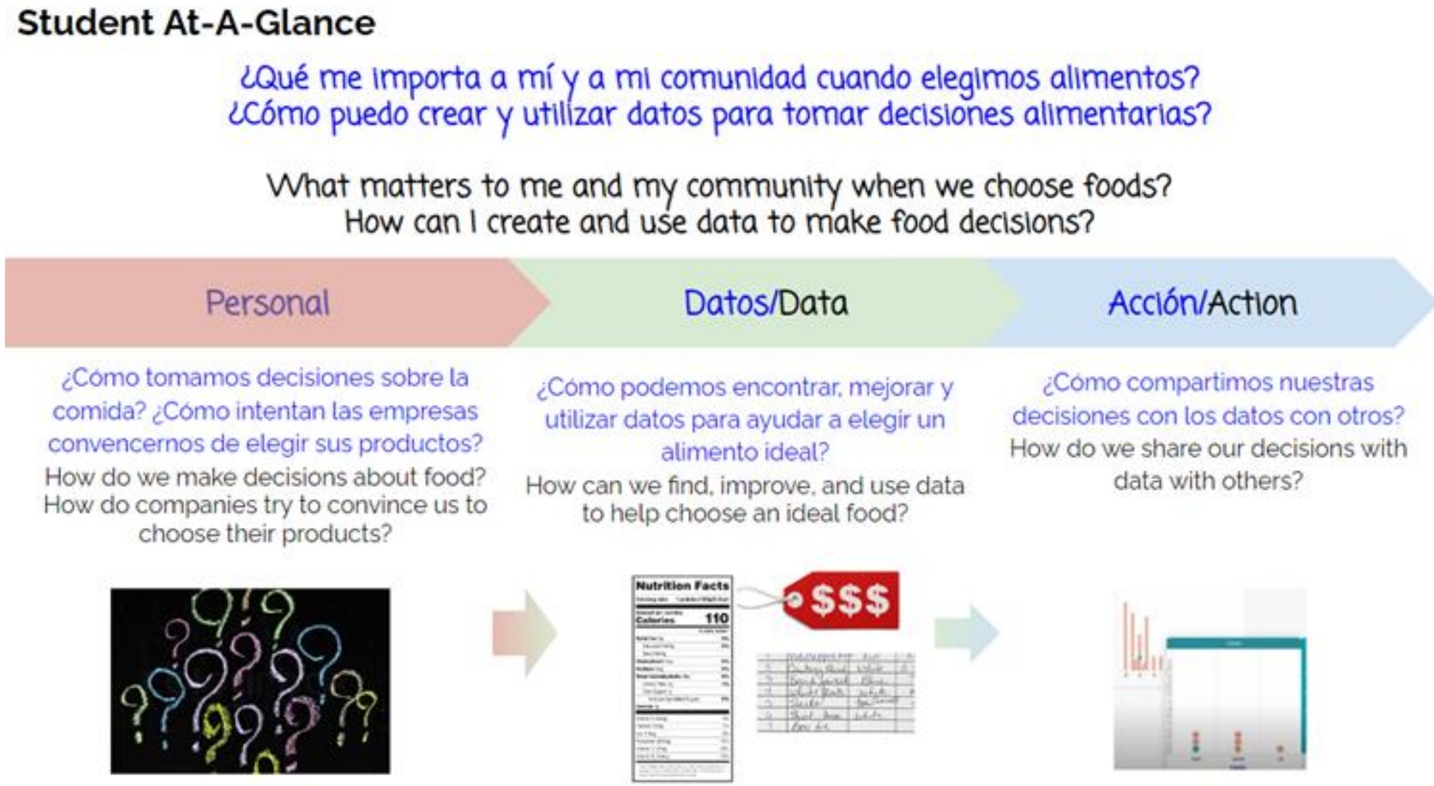


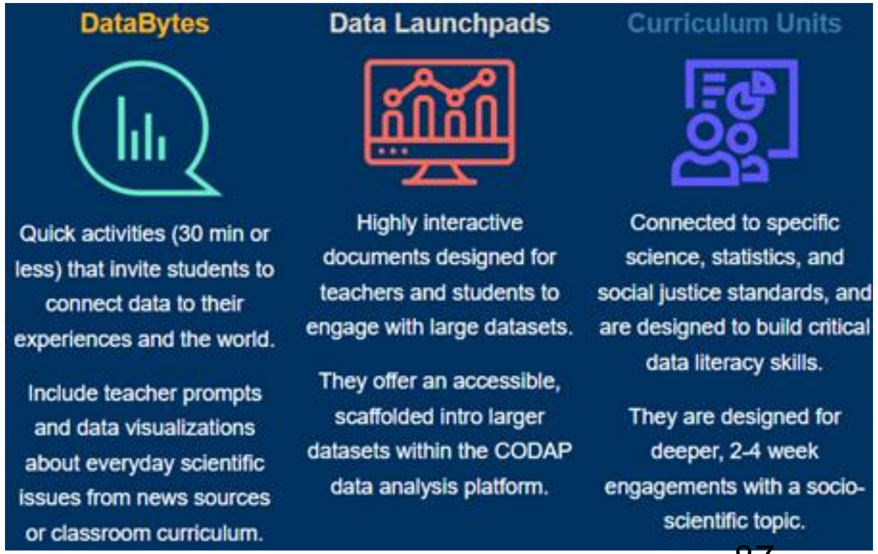
The Concord Consortium NC STATE College of Education Berkeley UNIVERSITY OF CALIFORNIA Graduate School of Education

Writing Data Stories https://live-core-lab.pantheon.berkeley.edu/2022/11/18/writing-data-stories/ Creative Commons

Curriculum materials

- PERSONAL CONNECTIONS - personal or community connection (with analysis of tensions or resonances with findings in data)
- DATA - analysis of a generally externally-sourced quantitative dataset
- INTERDISCIPLINARY/MULTIMODAL exploration of qualitative/contextual elements (journalism, interviews, observations) that help explain the tensions or resonances
- CALL TO ACTION or concluding argument





Writing Data Stories https://live-core-lab.pantheon.berkeley.edu/2022/11/18/writing-data-stories/ Creative Commons

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)

2. Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe (Travis Weiland)

3. **Critical perspectives on data literacy emerging from Latin America** (Lucía Zapata-Cardona)

4. **Joint discourse between mathematical modeling and statistics/data science communities** (Takashi Kawakami)

5. **What can mathematics/statistics education contribute to Artificial Intelligence/Machine Learning literacy** (Rolf Biehler)

6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

Goals

- Make evident critical perspectives on data literacy and data literacy work emerging from Latin America
- Present some relevant scholars that have influenced the work in the region
- Show some purposes behind critical data literacy
- Present some examples that illustrate those purposes

Contributions to the theoretical debate on Critical data literacy from Latin America

Critical data literacy → Critically think about data

↓

Use data to understand **crises** and work to transform them

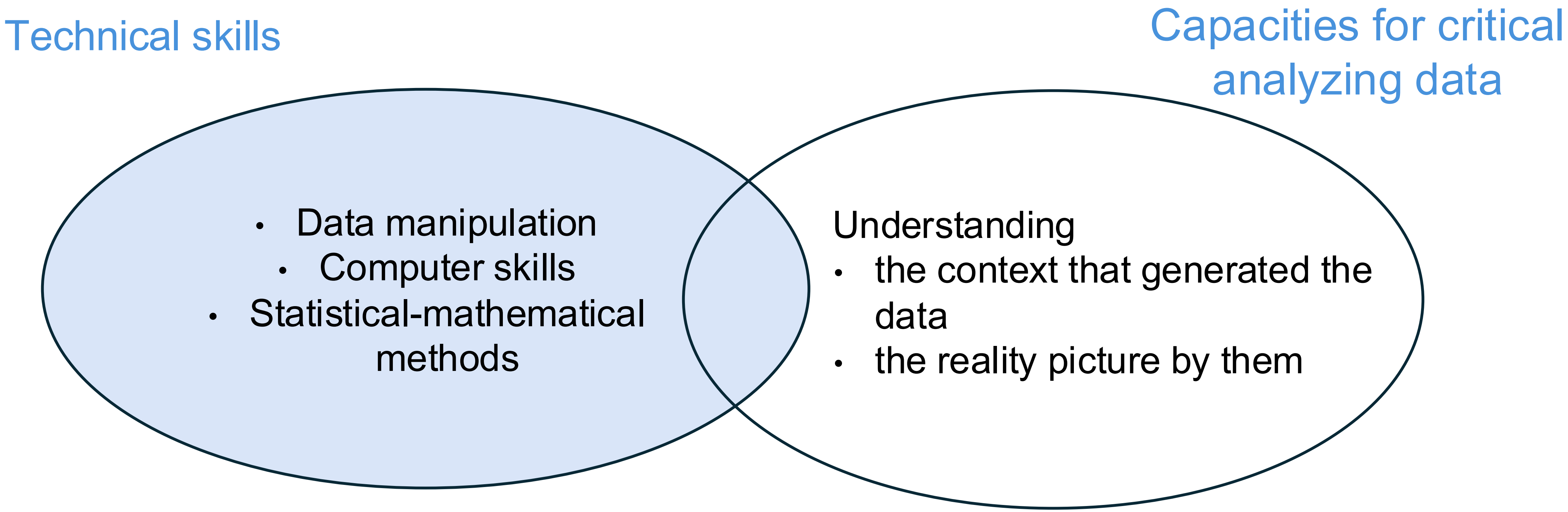
repression, conflict, contradiction, misery, inequity, ecological devastation, exploitation

(Cabrera et al., 2023, Raffaghelli, 2022, Tygel & Kirsch, 2016, Valente & Grohmann, 2024)

Tygel, A.F., & Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics, 12*(3), 108—121. https://doi.org/10.15353/joci.v12i3.3279

Contributions to the theoretical debate on Critical data literacy from Latin America

“Critical Data Literacy is the set of abilities which allows one to use and produce data in a critical way” (Tygel & Kirsch, 2016).



Tygel, A.F., Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. The Journal of Community Informatics, 12(3), 108—121.

Contributions to the theoretical debate on Critical data literacy from Latin America

Critical Theory

Marxist philosophy

Dialectical logic
is an approach that
emphasizes the importance of
✓ contradictions,
✓ continuous change,
✓ wholeness, and
✓ interconnection
in understanding reality.
Facing the **crises** is the essence of critical theory

There is a **decolonizing** purpose:

- Awareness
- Autonomy
- Emancipation
- Problematization of the crises
- Appropriation of historical reality
- Transformation

Critical Pedagogy

Paulo Freire

Popular Education

✓ Read and write the word
✓ Emancipatory perspective – liberation
✓ Consciousness
✓ Own practices – Contextualized
✓ Dialectical interaction
✓ Transformative praxis

Tironi, M. & Valderrama, M. (2021). Descolonizando los sistemas algorítmicos: diseño crítico para la problematización de algoritmos y datos digitales desde el Sur. *Palabra Clave*, *24*(3), e2432.
Tygel, A.F., Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. The Journal of Community Informatics, 12(3), 108—121.

Contributions to the theoretical debate on Critical data literacy from Latin America

Freire (1983)

- ✓ Liberation
- ✓ Democratic society

Giroux (1988)

- ✓ Reflection and critical action to overcome injustices

Skovsmose (2014)

- ✓ Critical reflection
- ✓ Empowerment
- ✓ Social issues

Campos (2016)

- ✓ Reflective citizenship
- ✓ Democratic attitudes
- ✓ Modeling
- ✓ Transform reflections into actions

Campos, C. R. (2016). La educación estadística y la educación crítica [Statistical education and critical education]. Proceedings of Encuentro Colombiano de Educación Estocástica. Bogotá, Colombia.

What is the need for critical data literacy in Latin America ?

Image taken from Google Maps

Critical data literacy in a region with high levels of inequality (social, cultural, economic) could help people to

- make sense of the socially and economically important data that affects their lives
- make informed decisions
- participate in public life
- recognize the harm that powerful interests can inflict with data
- recognize that data is not neutral
- recognize that biased algorithms can exacerbate social and economic inequalities
- expose systematic social injustices

Tironi, M. & Valderrama, M. (2021). Descolonizando los sistemas algorítmicos: diseño crítico para la problematización de algoritmos y datos digitales desde el Sur. *Palabra Clave*, *24*(3), e2432.

Purposes behind critical data literacy proposals in Latin-American tradition

- democracy (Giordano et al., 2022) - Example 1
- social justice (Raffaghelli, 2022) - Example 3
- providing visibility of under covered complexities (Martínez-Castro et al., 2022) - Example 2
- developing awareness about social issues (Martínez-Castro et al., 2022) - Example 2
- co-liberation / contra culture (Raffaghelli, 2022) - Example 3

Example 1: Data literacy for democracy

A classroom experience

- **Context:** A research-based experience
- São Paulo, Brazil
- Third grade of a public high school (17-21 years old students)
- Two classes
- **Goal:** Corroborate or refute statements with evidence
- No part of the curriculum
- In the Paulist curriculum, statistics is covered in two months

Giordano, C. C., Pereira, F. A., & Souza, F. dos S. (2022). El desarrollo de la alfabetización estadística de los estudiantes de secundaria brasileña: un enfoque a través de las Estadísticas Cívicas. *Revista Baiana De Educação Matemática*, *3*(01), e202212

Example 1: Data literacy for democracy

A classroom experience

- Jair Bolsonaro's speech at the 76th UN assembly

  "In the Amazon, we had a 32% reduction in deforestation in the month of August, compared to August of the previous year" (Sep. 21st , 2021)

Image taken from Pixabay (copyright free)

When truth dies, democracy dies

Giordano, C. C., Pereira, F. A., & Souza, F. dos S. (2022). El desarrollo de la alfabetización estadística de los estudiantes de secundaria brasileña: un enfoque a través de las Estadísticas Cívicas. *Revista Baiana De Educação Matemática*, *3*(01), e202212

Example 1: Data literacy for democracy
A classroom experience

- **Results:** To contrast the truthfulness of the statement, students
  - gathered information from multiple sources
  - used different ways to represent data
  - integrated technical knowledge and language
  - improved evidence-based argumentation
  - developed skepticism about public speech

Giordano, C. C., Pereira, F. A., & Souza, F. dos S. (2022). El desarrollo de la alfabetización estadística de los estudiantes de secundaria brasileña: un enfoque a través de las Estadísticas Cívicas. *Revista Baiana De Educação Matemática*, *3*(01), e202212

Example 2: Data literacy - providing visibility / developing awareness

- **Goal:** Use open data to study inequalities between men and women in different dimensions of society
- **Context:** 10 pre-service teachers, two two-hour sessions
- Methods class for teaching statistics in a mathematics education program
- Reading a news article “gender gap”
- Is there a gender gap in our country?
- Participants
  - ✓ gathered and analyzed information from public records
  - ✓ developed awareness of disparities
  - ✓ proposed a list of possible solutions
    - Speak publicly about these disparities
    - Offer tax incentives to companies that hire women

Martínez-Castro, C. A. & Zapata-Cardona, L. (2022). Formación inicial de profesores de estadística en una perspectiva crítica. En A. Salcedo & D. Diaz-Levicoy (Eds.) Formación del profesorado para enseñar estadística: retos y oportunidades (pp. 367-382). Universidad Católica de Maule.

Example 2: Data literacy - providing visibility / developing awareness

Data skills

- Data mining
- Data management
- Data visualization
- Data cleansing

Data tools

- Free data software
- Spreadsheets

Empowerment:

- Visibility
- Awareness
- Empathizing

Martínez-Castro, C. A. & Zapata-Cardona, L. (2022). Formación inicial de profesores de estadística en una perspectiva crítica. En A. Salcedo & D. Diaz-Levicoy (Eds.) Formación del profesorado para enseñar estadística: retos y oportunidades (pp. 367-382). Universidad Católica de Maule.

Example 3: Data literacy for co-liberation / social justice
A workshop with in-service teachers

Context: The Mothers of the Plaza de Mayo is an Argentine human rights association formed in response to the National Reorganization Process, the military dictatorship by Jorge Rafael Videla (1976 to 1983)

Image taken from Pixabay (copyright free)

- 30.000 missing people
- Developed strategies to look for missing people using DNA
- Developed “grandparenthood index”
- Developed memory map in several Argentinian cities
- “Small data” helps “unravel” the complexity of narratives
- Open data as a form of activism

Raffaghelli, J. (2022). Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía [Data Literacy and Social Justice: An oxímoron? Responses from the counter-hegemony]. Revista Izquierdas, 51

Example 3: Data literacy for co-liberation / social justice

A workshop with in-service teachers

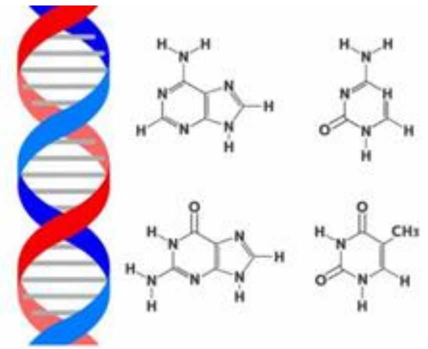


Image taken from illust AC (copyright free)

Co-liberation: Create alternative forms of power from civil society to obtain their rights.

Grandparenthood index based on the genetic material, determines with accuracy the probability of relationship between a grandparent and a grandchild

- 12 Workshops from 2018-2020 in academic events
- 80 of 298 participants from Latin America
- **Goal:** explore the educational potential of open data as an engine of interdisciplinary dialogue in learning design and pedagogical practices
- **Method:** data exploration, data production, reflection on results

Raffaghelli, J. (2022). Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía [Data Literacy and Social Justice: An oxímoron? Responses from the counter-hegemony]. Revista Izquierdas, 51

Example 3: Data literacy for co-liberation / social justice
A workshop with in-service teachers

Results:

Participants

- appreciated the importance of small data
- emphasized the need to **study contexts**, **work on educational designs** and **deepen technical knowledge** to understand how data can be used to develop solutions.

Memory map

Image taken from Pixabay (copyright free)

Experience with Open Data as educator

Raffaghelli, J. (2022). Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía [Data Literacy and Social Justice: An oxímoron? Responses from the counter-hegemony]. Revista Izquierdas, 51

Future challenges for the community in Latin America

- Incorporation of quantitative research designs to assess impact on teaching and learning
- There are several isolated research experiences reported as case studies, but larger projects require joining forces
- Develop new capacity for initial teacher education and professional development
- Strengthen networks and support groups for educators
- Enrich the limited data science curriculum in compulsory education

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)

2. Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe (Travis Weiland)

3. Critical perspectives on data literacy emerging from Latin America (Lucía Zapata-Cardona)

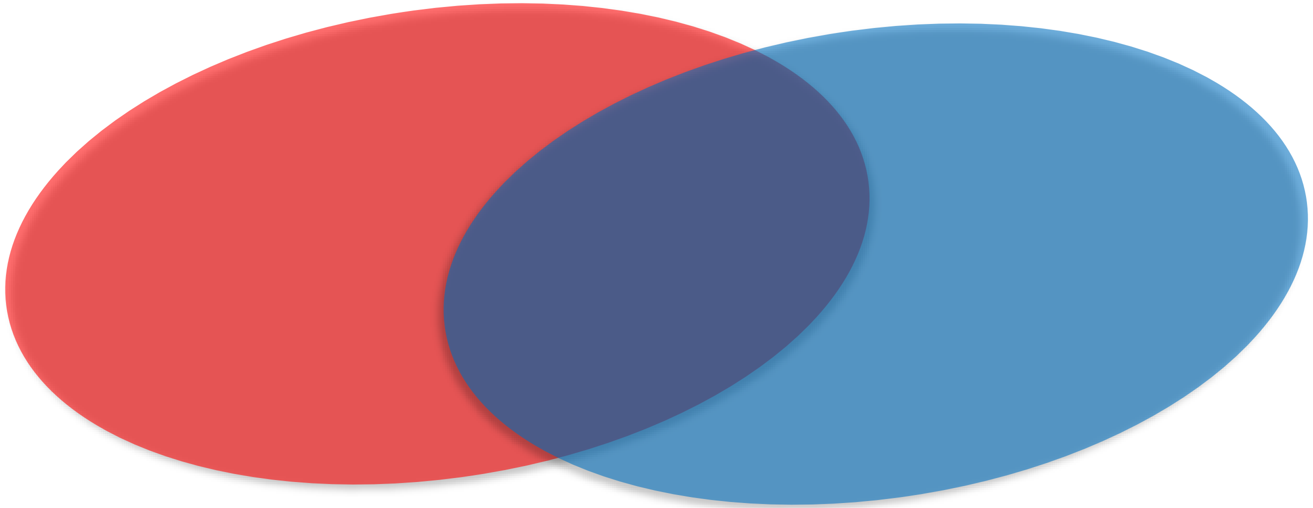
4. **Joint discourse between mathematical modeling and statistics/data science communities** (Takashi Kawakami)

5. **What can mathematics/statistics education contribute to Artificial Intelligence/Machine Learning literacy** (Rolf Biehler)

6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

Goal

The purpose of this section is to report on new trends in the joint discourse between the mathematical modeling (**MM**) and statistics/data science (**DS**) communities.



Plan: Analyzing three relevant discourses on data-rich MM since 2020

Discourse 1

Data-rich MM cycle integrating statistics & mathematics

Discourse 2

Interdisciplinary data-rich MM

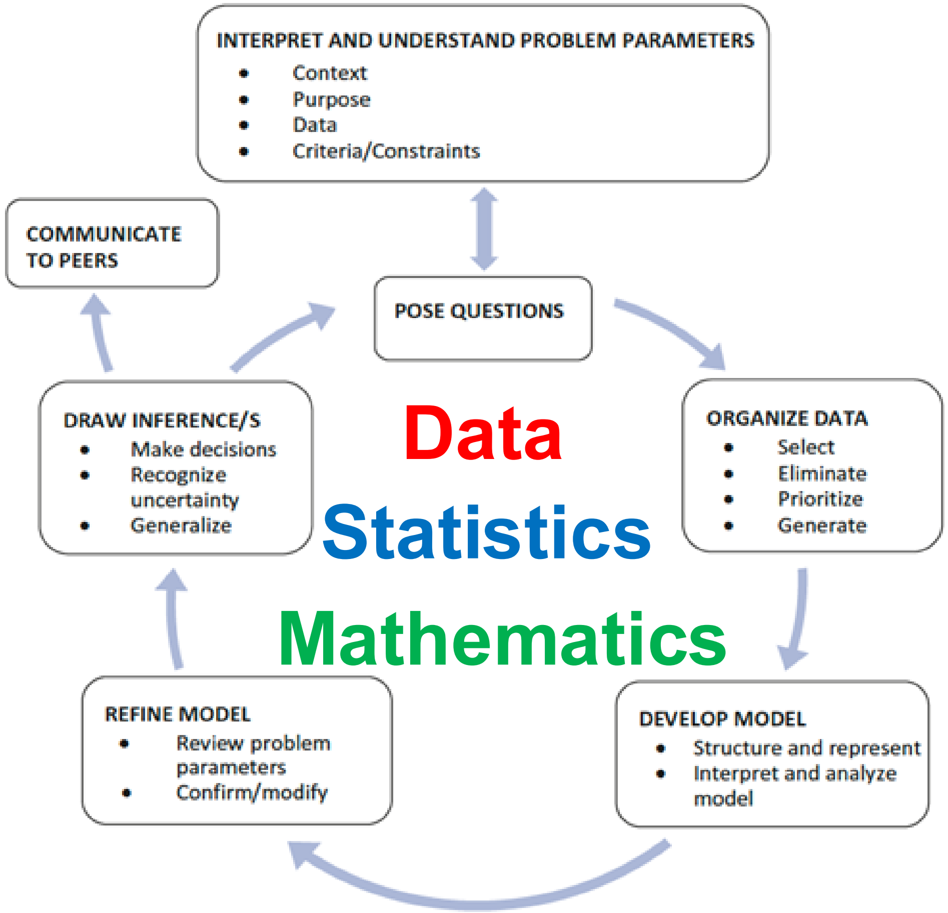
Discourse 3

Societal data-rich MM

Discourse 1: Data-rich MM cycle integrating statistics & mathematics

Data-rich MM cycle integrating statistics & mathematics

- **Data-rich MM processes** with **statistics** and **mathematics at their core** to develop **statistical/mathematical literacy/learning** (e.g., English, 2021; Kawakami & Saeki, 2022)
    - **Proactive, autonomous, and critical disposition as the affective aspects** of **MM** (e.g., Niss & Blum, 2020; Burkhardt 2020; Frejd & Vos, 2024; Galbraith, 2015; Maass et al., 2022; Stillman, 1998)



English (2021, p.7), permission granted by Springer Nature

Data-rich MM cycle integrating statistics & mathematics

Role of data highlighted by empirical studies of data-rich MM in empowering citizens

- **Provide contextual knowledge, stories, and patterns (trends and variation) behind or in the data**
  (e.g., Kazak et al., 2023; McLean et al., 2023; Stillman & Brown, 2023; Van Dijke-Droogers et al., 2021)

Source: Illust AC (copyright free)

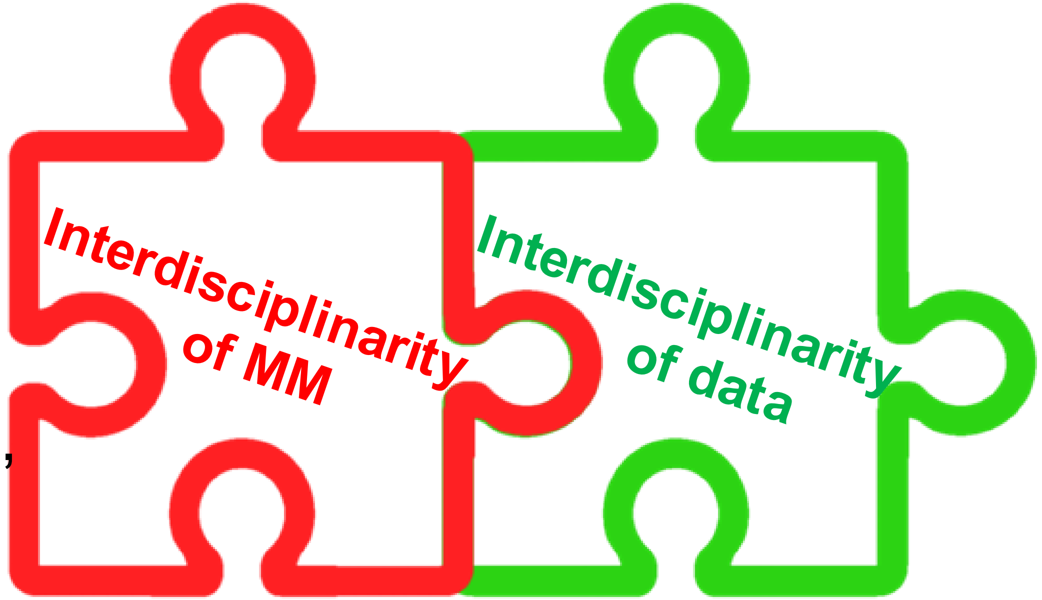
- **Generate a hypothetical model for prediction with one data set and then test it on another one**
  (e.g., Ärlebäck, Frejd, & Doerr, 2021; Dvir & Ben-Zvi, 2023; Kawakami & Mineno, 2021)

Source: Illust AC (copyright free)

Interdiclinary data-rich MM

Discourse 2: Expanding data-rich MM into interdisciplinary aspects

- **Interdisciplinary data-rich MM** with not only statistics and mathematics, **but also other disciplines/subjects** to promote **also "STEM literacy"** (e.g., Bybee, 2018) essential for citizens
  - **The inherent interdisciplinarity** of **MM** (e.g., English, 2016; Maass et al., 2019; Pollak, 1977; Stillman et al., 2023)
  - **The inherent interdisciplinarity** of **variation in data** (e.g., Lehrer & Schauble, 2002; Makar et al., 2023; Watson et al., 2020)



Developing multifaceted views and knowledge through interdisciplinary data-rich MM

- **Interdisciplinary data-rich MM** promotes back-and-forth movement between **deterministic views**, **stochastic views**, and **other STEM views** such as scientific views, design/engineering views (e.g., Aridor et al., 2023; Fry et al., 2024; Kawakami & Saeki, 2024a)
  - DS-oriented citizen science project (Aridor et al., 2023)
  - The importance of citizens' understanding of the role of uncertainty in generating data-based interdisciplinary knowledge (e.g., Lehrer, Wisittanawat, & Schauble, 2024)

Source: Illust AC (copyright free)

Source: Illust AC (copyright free)

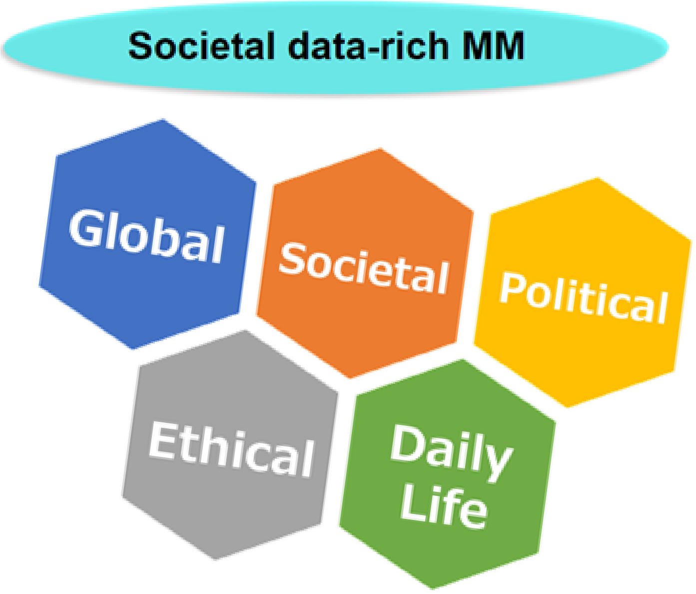
Societal data-rich MM

Discourse 3: Expanding data-rich MM into societal aspects

Source: Illust AC (copyright free)

- **Societal data-rich MM** with statistics and mathematics to promote **critical thinking and citizenship**
  - **The social, critical, and prescriptive aspects** of **MM** (e.g., Barbosa, 2006; Davis & Hersh, 1986; Niss, 2015; Skovsmose, 1994)
    - **Prescriptive modeling** (Niss, 2015): 'pave the way for taking action based on decisions resulting from a certain kind of mathematical considerations, in other words to *change the world*' (p.69)
  - **Need for global, societal, political, ethical, and daily life contexts** to create authentic DS practices

Context example of societal data-rich MM found in studies



- *COVID-19 or epidemics* (e.g., Kawakami & Saeki, 2024b; Maass et al., 2023)
- *Climate change* (e.g., Kazak et al., 2023; Steffensen & Kacerja, 2021; Zapata-Cardona & Martínez-Castro, 2023)
- *Pandemic-related media items* (Gal & Geiger, 2022)
- *Reliability of public data sets* (e.g., Wilkerson et al., 2022)
- *Mapping crime in the regions* (Andersson & Register, 2023)
- *Social justice in fair distribution of school funding* (Jung & Wickstrom, 2023)
- *Trash production* (Rosa, Orey, de Sousa Mesquita, 2023)

An example from teacher education of societal data-rich MM (Kawakami & Saeki, 2024b)

Societal data-rich MM

- **Setting**: A teaching experiment for a group of math pre-service teachers in Japan to experience **basic societal data-rich MM** and **become aware of the societal benefits and risks of the MM reconstructing reality in society**

- **Task/context**: Data-informed decision-making on the COVID-19 pandemic in Japan,

'In which prefecture would you declare a state of emergency?'

Source: Illust AC (copyright free)

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling A prescriptive modelling perspective. In H.S. Siller, V. Geiger, V., & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

An example from teacher education of societal data-rich MM (Kawakami & Saeki, 2024b)

Societal data-rich MM

- **Data** such as *population*, *number of COVID-19 cases and admissions*, *number of beds for COVID-19 patients and critical cases* in Japan’s 47 prefectures

都道府県のCOVID-19データ（1月12日現在）

都道府県データ (47データ)

索引	No	県	境界	海	人口	面積 (km²)	現在感染者数(3月31日現在)2	現在感染者数(4 …4月30日現在)	現在感染者数(5 …(5月31日現在)2	現在感染者数(7 …月31日現在)2	現在感染者数 …(9月30現在)	現在感染者数 …(11/29現在)	現在感染者 …(1/12現在)	感染者数感染 …(3月31日現
1	1	北海道		有り	5286000	83424	38	485	195	82	138	2323	1588	
1	2	青森県		有り	1263000	9646	8	11	1	2	1	16	70	
2	3	岩手県		有り	1241000	15275	0	0	0	4	0	95	65	
3	4	宮城県		有り	2316000	6859	6	36	0	19	45	167	497	
4	5	秋田県		有り	981000	11638	4	5	0	2	0	17	28	
5	6	山形県		有り	1090000	6652	1	20	3	1	2	31	60	
6	7	福島県		有り	1864000	13784	4	49	6	6	37	54	355	
1	8	茨城県		有り	2877000	6097	24	119	7	68	37	406	749	
2	9	栃木県		無し	1946000	6408	12	33	14	56	49	107	1335	
3	10	群馬県		無し	1952000	6362	18	100	12	30	55	200	589	
4	11	埼玉県		無し	7330000	3768	75	572	66	504	264	1386	4925	
5	12	千葉県		有り	6255000	5083	143	632	55	370	272	966	4667	
6	13	東京都		有り	13822000	2109	473	2730	416	2742	2177	4407	19029	
7	14	神奈川県		有り	9177000	2416	111	474	222	342	570	1928	6923	
1	15	新潟県		有り	2246000	10364	22	36	5	22	8	103	159	
2	16	富山県		有り	1050000	2046	2	157	16	6	10	25	176	
3	17	石川県		有り	1143000	4186	9	185	54	19	43	17	164	
4	18	福井県		有り	774000	4191	20	44	4	14	4	22	40	
5	19	山梨県		無し	817000	4254	5	16	4	17	12	48	175	
6	20	長野県		無し	2063000	13104	5	48	7	23	8	135	530	
7	21	岐阜県		無し	1997000	9769	24	66	3	123	31	184	703	

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.S. Siller, V. Geiger, V., & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

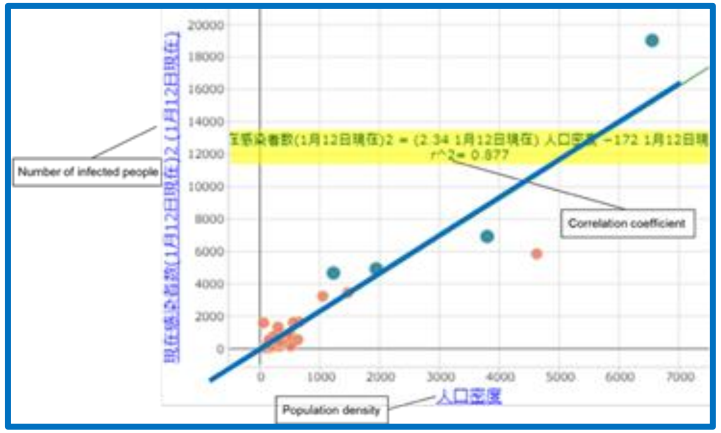
An example from teacher education of societal data-rich MM (Kawakami & Saeki, 2024b)

Societal data-rich MM

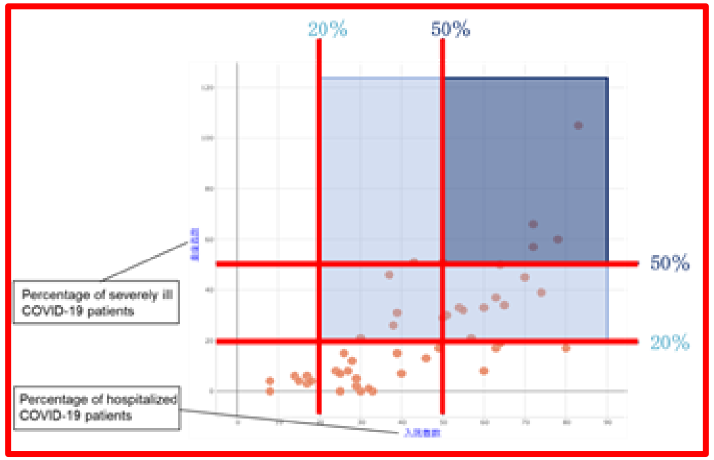
- **Focus on**:

  Teachers' model use with statistics and mathematics (e.g., scatterplot, mean, quartiles, linear regression, correlation coefficient) for

  - descriptive purposes to visualize the trends and variability of data on the current world
  - **prescriptive purposes** to articulate data-informed societal decision-making and **lead human action for a preferred world**

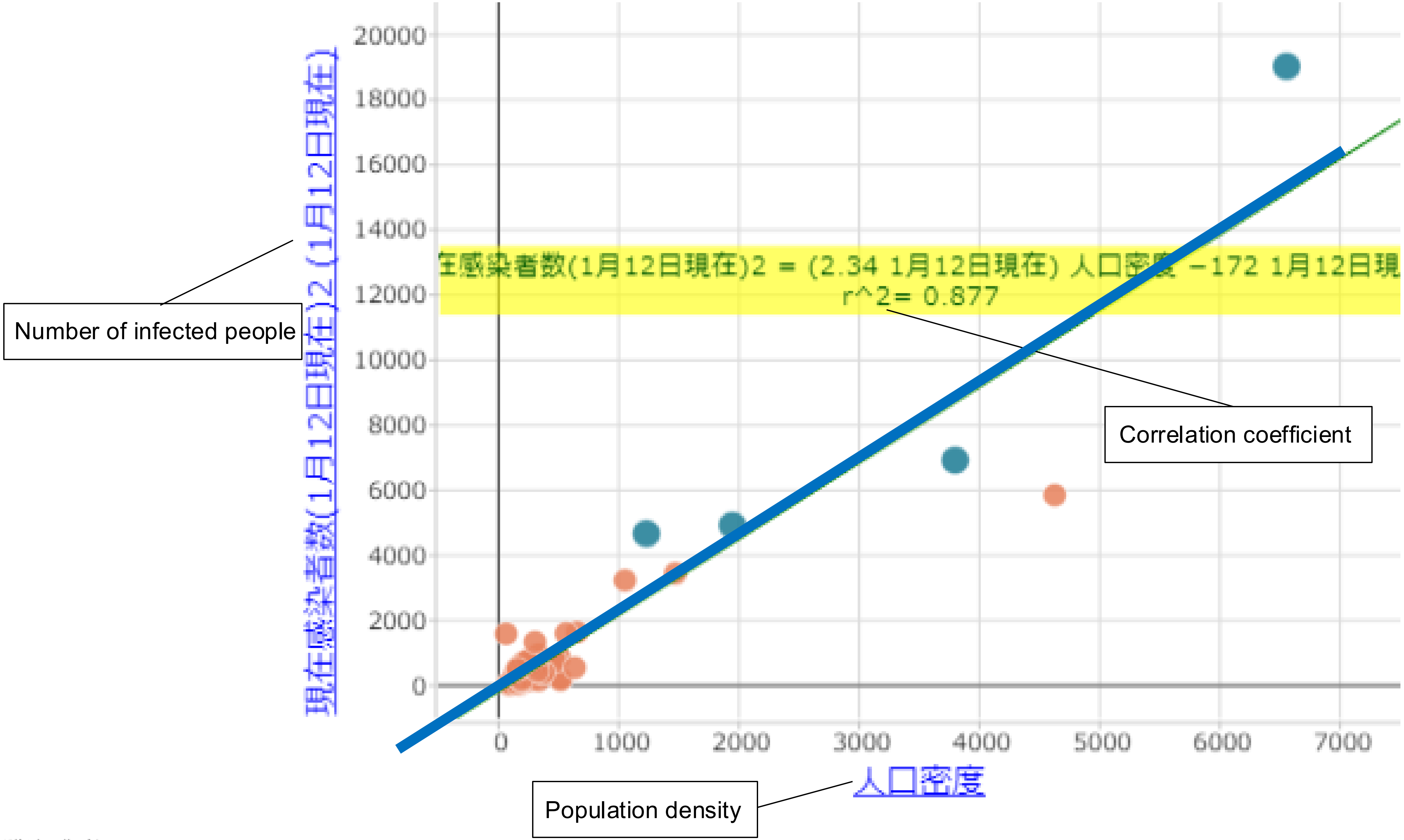


Adapted from Kawakami & Saeki (2024b, p. 601)



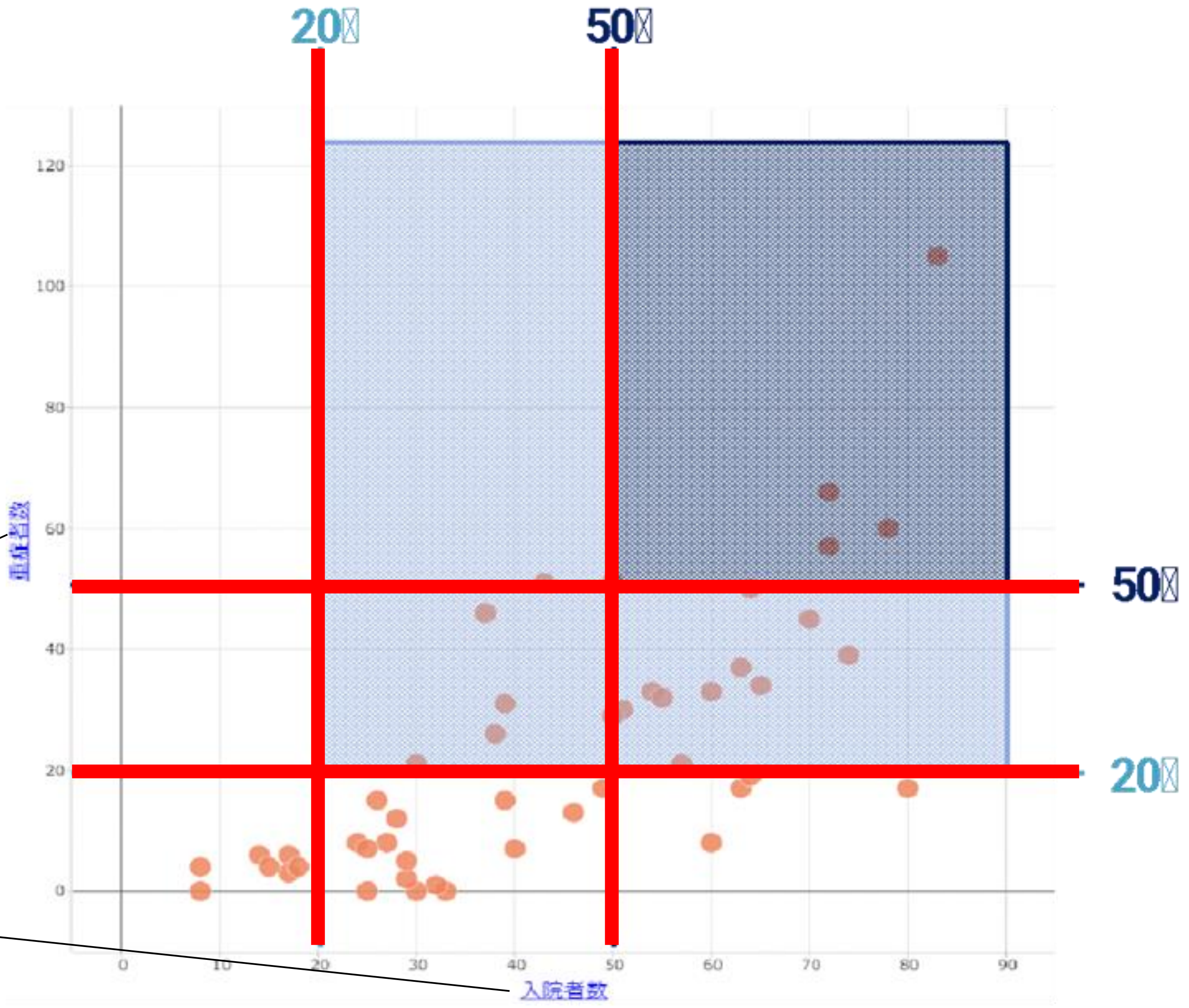
Adapted from Kawakami & Saeki (2024b, p. 601)

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.S. Siller, V. Geiger, V., & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.



Adapted from Kawakami & Saeki (2024b, p. 601)

Percentage of severely ill COVID-19 patients



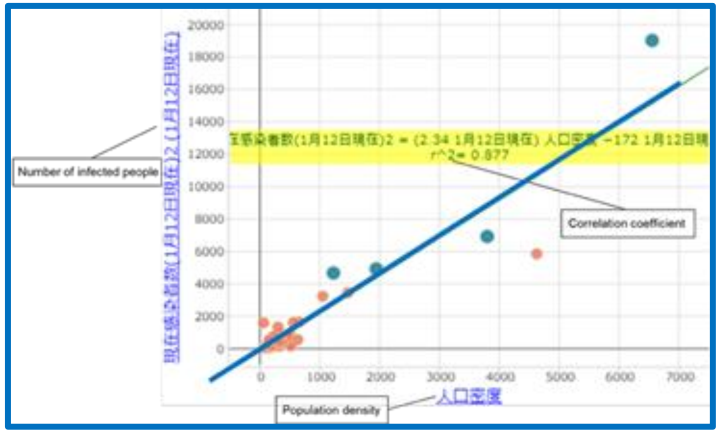
Percentage of hospitalized COVID-19 patients

Adapted from Kawakami & Saeki (2024b, p. 601)

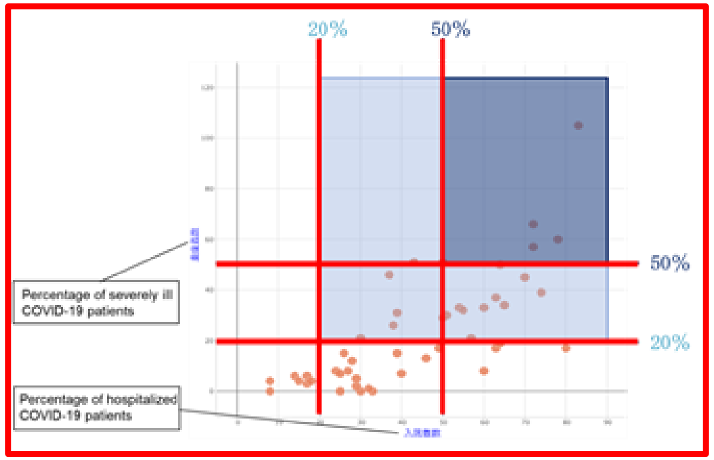
An example from teacher education of societal data-rich MM (Kawakami & Saeki, 2024b)

Societal data-rich MM

- **Focus on** (cont.):
  - 2) Suggestions for **responsible citizens**
    - **Societal data-rich MM** promotes awareness and critical reflection on **the prescriptive power of data-based models** to reconstruct reality in society (i.e., O'Neil, 2016; Skovsmose, 2023)



Adapted from Kawakami & Saeki (2024b, p. 601)



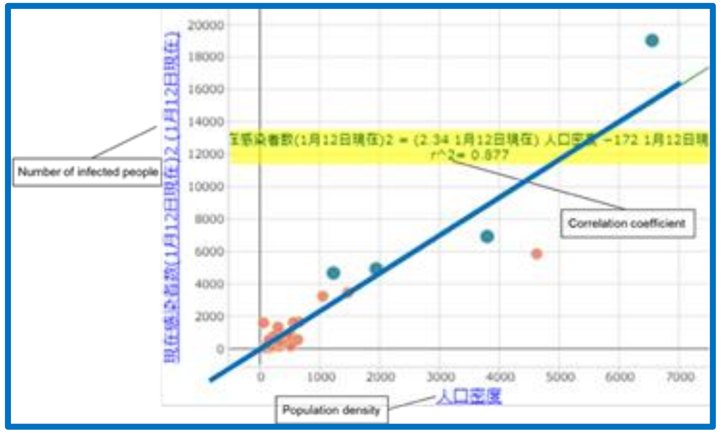
Adapted from Kawakami & Saeki (2024b, p. 601)

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.S. Siller, V. Geiger, V., & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

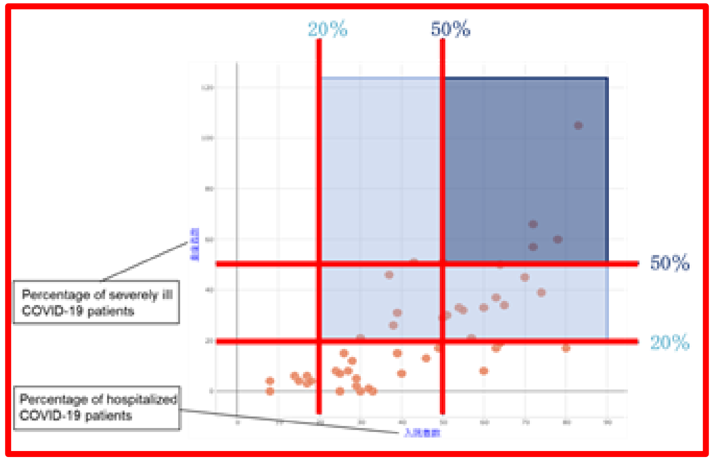
An example from teacher education of societal data-rich MM (Kawakami & Saeki, 2024b)

Societal data-rich MM

- **Focus on** (cont.):
  - 3) Suggestions for **teacher education on DS education**
    - **Societal data-rich MM** serves as **an exemplary learning for developing tasks and lessons on the prescriptive power of data-informed models in society** for empowering responsible citizens



Adapted from Kawakami & Saeki (2024b, p. 601)



Adapted from Kawakami & Saeki (2024b, p. 601)

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.S. Siller, V. Geiger, V., & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595-605). Springer.

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)
2. Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe (Travis Weiland)
3. Critical perspectives on data literacy emerging from Latin America (Lucía Zapata-Cardona)
4. Joint discourse between mathematical modeling and statistics/data science communities (Takashi Kawakami)
5. **What can mathematics/statistics education contribute to Artificial Intelligence/Machine Learning literacy** (Rolf Biehler)
6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

Reviews on AI Literacy (from computer science education perspective)

Almatrafi, O., Johri, A., & Lee, H. (2024). **A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023)**. *Computers and Education Open*, *6*. https://doi.org/10.1016/j.caeo.2024.100173

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). **AI literacy in K-12: a systematic literature review.** *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00418-7

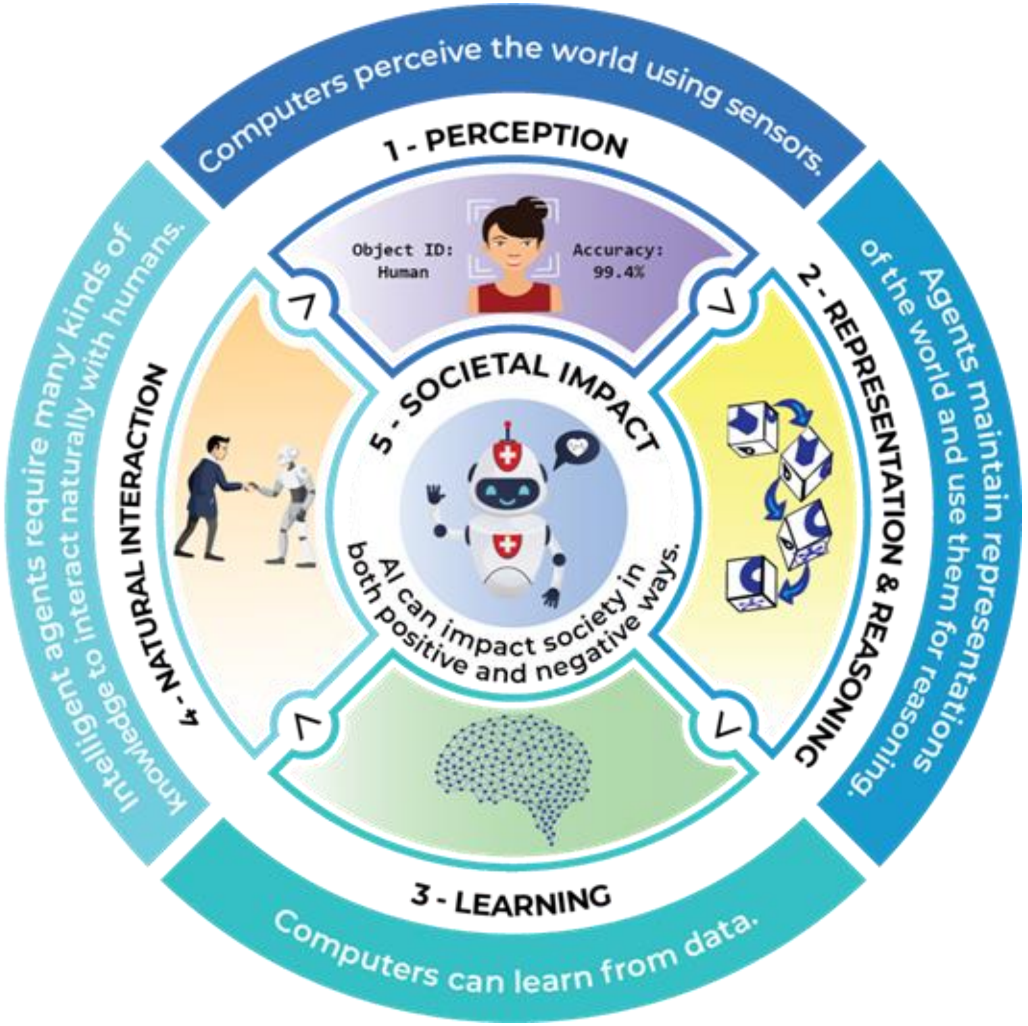
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). **AI4People-An Ethical Framework for a Good AI Society:** Opportunities, Risks, Principles, and Recommendations. *Minds Mach (Dordr)*, *28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Olari, V., & Romeike, R. (2024). Data-related Concepts for Artificial Intelligence Education in K-12. Computers and Education Open. https://doi.org/10.1016/j.caeo.2024.100196

Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). **Teaching Machine Learning in K–12 Classroom:** Pedagogical and Technological Trajectories for Artificial Intelligence Education. IEEE Access, 9, 110558-110572. https://doi.org/10.1109/access.2021.3097962

Sanusi, I. T. (2023). ***Machine Learning Education in the K–12 Context*** Itä-Suomen yliopisto (Ph.D.thesis)

Five Big Ideas in Artificial Intelligence



AI4K12.org

jointly sponsored by AAAI and CSTA.

https://ai4k12.org/resources/big-ideas-poster

3. Learning

"Computers can learn from data. Machine learning is a kind of statistical inference that finds patterns in data. Many areas of AI have progressed significantly in recent years thanks to learning algorithms that create new representations. For the approach to succeed, tremendous amounts of data are required.

This "training data" must usually be supplied by people, but is sometimes acquired by the machine itself."

https://ai4k12.org/resources/big-ideas-poster/#jp-carousel-996

A minimal requirement for data science education's contribution to AI literacy?

Students should understand **at least one machine learning algorithm** that "learns from data"

Requirements: should be teachable emphasizing

- white or grey box approach to models and algorithms
- highlight the transition from manual construction (human-driven model building) to automation (machine-driven model building).
- illustrating central concepts and workflows of predictive modeling and its use

Candidates

- (linear) regression, however with a new mindset
- k nearest neighbors (KNN) (Mike & Hazzan 2022, Bata 2024)
- **Decision trees (DT)**

Decision Trees (DT) as an Exemplary ML method

Design challenges

- **elementarizing** the algorithms and the concepts
- accessible and motivating **authentic examples, questions and data**
- **digital tools** for learning, creating and using DT

Research challenges

- Design-based research
- Studies of teaching and learning processes
- Assessment of learning outcomes

DT publications: Zieffler et al. 2021, Ferguson & Pfannkuch 2024, 5; Erickson & Engel 2023, Martignon, Engel & Erickson 2022; Various publications from Biehler/Fleischer/Podworny 2021, 2022, 2024, 2025

Classification task: Predicting the type of habitat lizards have come from **(target variable)** by several **predictor variables**

Training data

Marnocha's Lizards

cases (160 cases)

index	Island	Habitat	Mass (g)	SVL (mm)	Hindlim... ngth (mm	Hindspan (mm)	Forelimb (mm)	Forespan (mm)	Gape width (mm)	Head depth (mm)	Toe pad width (mm)	Tail length (mm)
1	New Pr...	Natural	2	49	11.2	23.3	7.1	20.5	7	4	0.9	
2	New Pr...	Natural	2.2	45	10.4	25.4	7	20.8	7.6	5.1	1.2	78
3	New Pr...	Natural	2.4	50	11.6	27.1	7.1	22.5	7.2	4.6	1	
4	New Pr...	Natural	2.6	48	11.2	25.8	7.2	21.4	7.8	4.5	1.1	83
5	Exuma	Natural	2.7	53	11.4	27	7.3	22.9	7	5.1	0.8	
6	New Pr...	Natural	2.7	48	11.3	25.1	8.1	21.9	7.2	4.5	1.2	
7	Harbor I...	Natural	2.7	47	10.8	24.2	7	19.9	8.5	4.7	0.9	83
8	Eleuthe...	Natural	2.7	46	10.9	25	7.3	21	7.5	4.6	1.1	
9	Exuma	Natural	2.8	49	12.2	25.8	7.2	22.6	7.4	4.4	1.1	94
10	New Pr...	Natural	2.8	49	11.9	25.7	8.7	23.2	8	4.5	0.9	89
11	New Pr...	Natural	2.8	51	11.4	25.3	7.6	22.7	8.1	4.8	1.1	
12	New Pr...	Natural	2.8	50	12.1	26.4	7.6	22.5	7.4	4.4	1.2	
13	New Pr...	Natural	2.9	51	11.8	27	8.8	23.3	8.7	5.2	1.1	90
14	Eleuthe...	Natural	2.9	48	11.5	25.8	7	21.6	8.1	4.6	1	
15	New Pr...	Natural	3	51	10.4	26.2	8.1	22.5	7.3	4.3	1.1	86
16	New Pr...	Disturb...	3	52	11.6	25	7.5	21	7.7	4.1	0.9	

GAISE II Guidelines for Statistics Education)

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) - A Framework for Statistics and Data Science Education.* American Statistical Association.

Decision tree for predicting the type of habitat lizards have come from

Misclassification rate:
10% for these **training data**

How to find a good decision tree?

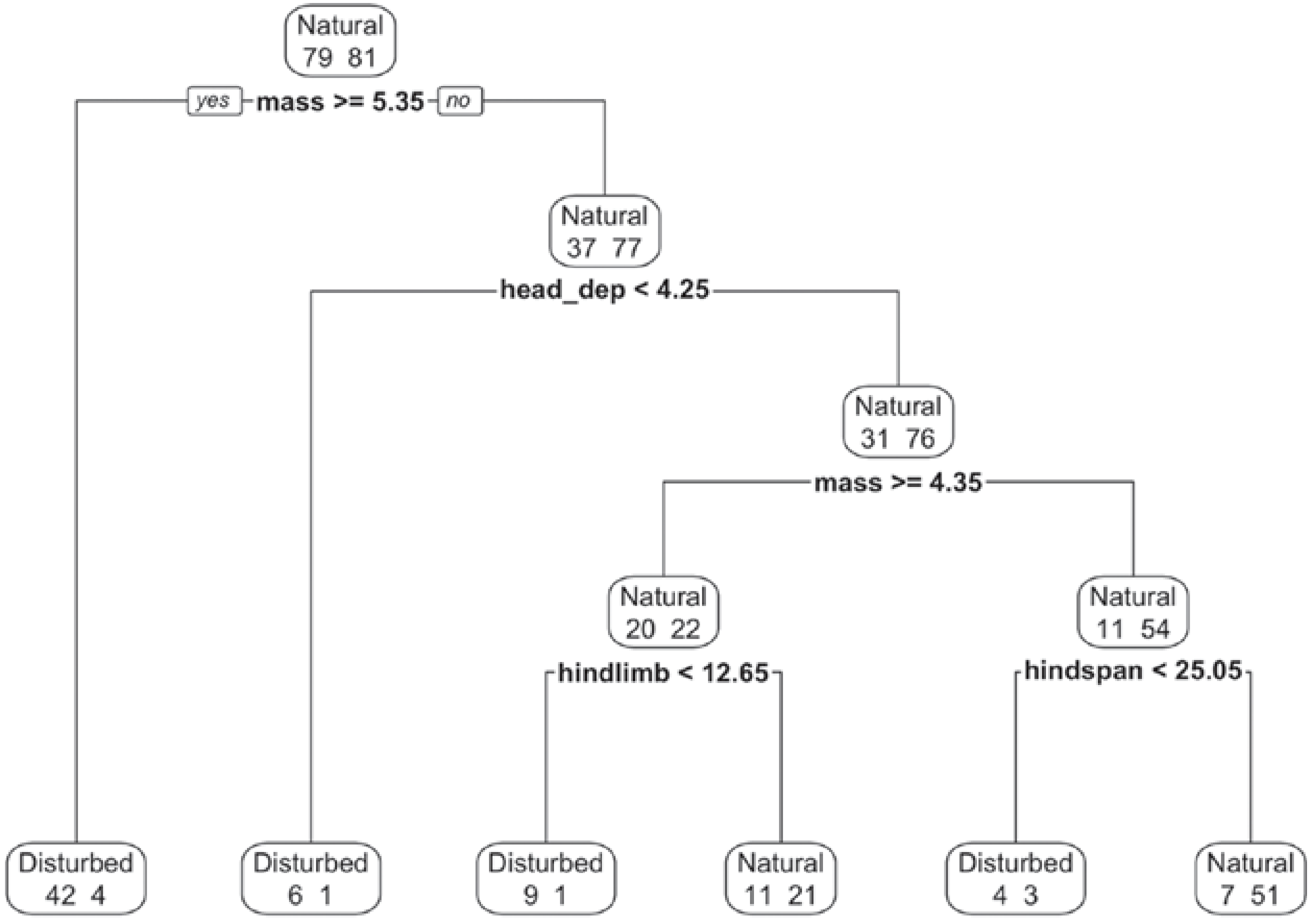


Figure 20: CART Tree for Lizard data where the left branch is taken when rule applies.

https://en.wikipedia.org/wiki/Lizard#/media/File:Lizard_Collage.jpg

GAISE II, p. 94

Basic ideas needed I

- successive splitting of the data set into two subsets
- classification by “majority vote” (with misclassification errors)
- selecting variables and splits by
  - data-based criteria (lowest misclassification rate)
  - context-related reasoning: which predictor variables are informative for or correlated with the target variable
- machines only use data-based criteria, testing all possible variables and splits
- machine vs. humans: pros and cons
  - lower misclassification rates
  - discovering influential variables
  - correlation not causation

Basic ideas needed II

Interface or **boundary concepts** for discussing ethical aspects

- Quality criteria of DT:
  - misclassification rates
  - misclassification rates of different types
  - explainability
- DT is to be used for predicting future cases
  - This has to be evaluated with test data
  - Problem of overfitting to training data (not generalizable)
  - predictive quality has to be monitored
- Prediction can turn into prescription (decision rather than predictions)
- Bias: if training data are biased / not representative, so is the DT

Two examples on DT from the ProDaBi project

1. How TikTok can find out your true age?
2. How can we use nutrition information to decide whether a food is rather recommendable or rather not recommendable?

Joint work RB with Yannik Fleischer und Susanne Podworny

ProDaBi
Project Data Science and Big Data at School

PADERBORN UNIVERSITY

Deutsche Telekom Stiftung

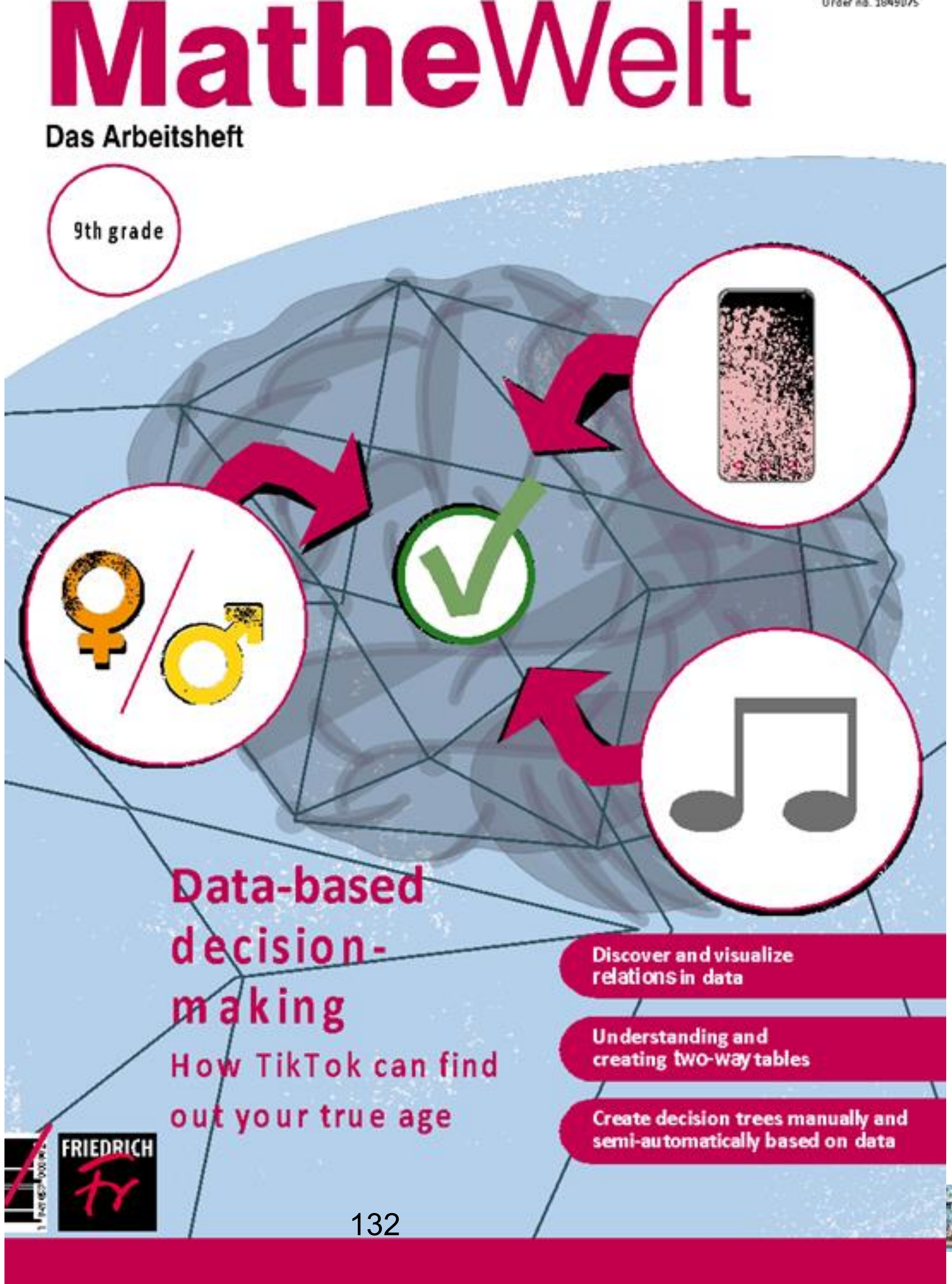
www.prodabi.de/en

Example 1 (Grade 9): How TikTok can find out your true age

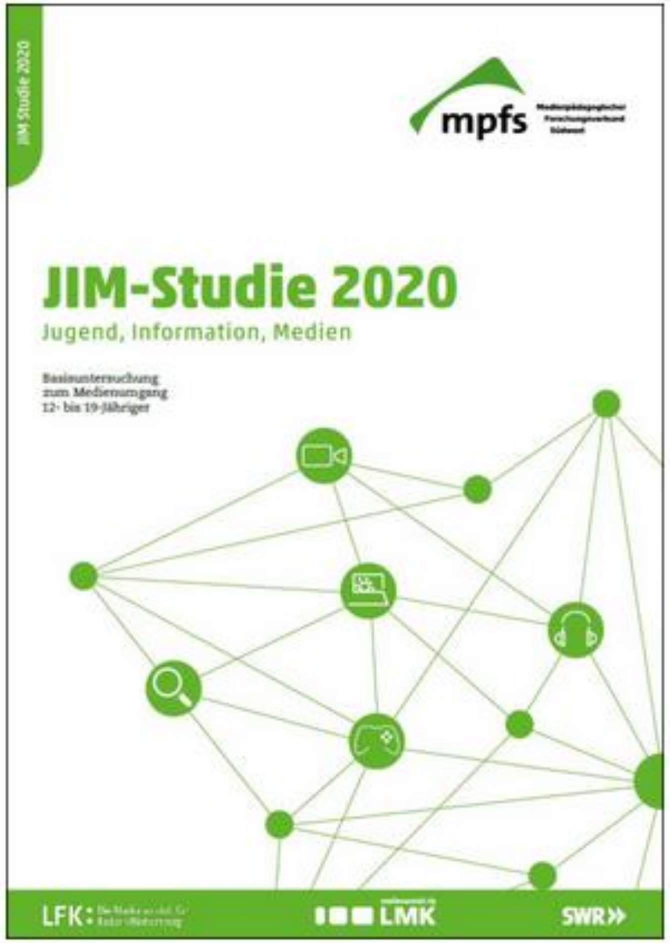
Yannik Fleischer,
Susanne
Podworny,
Rolf Biehler (2024)

Teacher Journal
„mathematik lehren“ 244 (June 2024)

AI in mathematics teaching at school
(edited by RB & Sarah Schönbrodt)



Example 1: How can TikTok find out your true age?



Background

- We collected own micro data of ~1200 teenagers
- “YOU-PB” data
- Based on questionnaire of an official representative German survey on media use of 12-19 year olds
- ~150 Questions about media use and device ownership

(Official Report Rathgeb & Schmid (2020) shows only aggregated data)

Variables (~150) of the YOU-PB data set

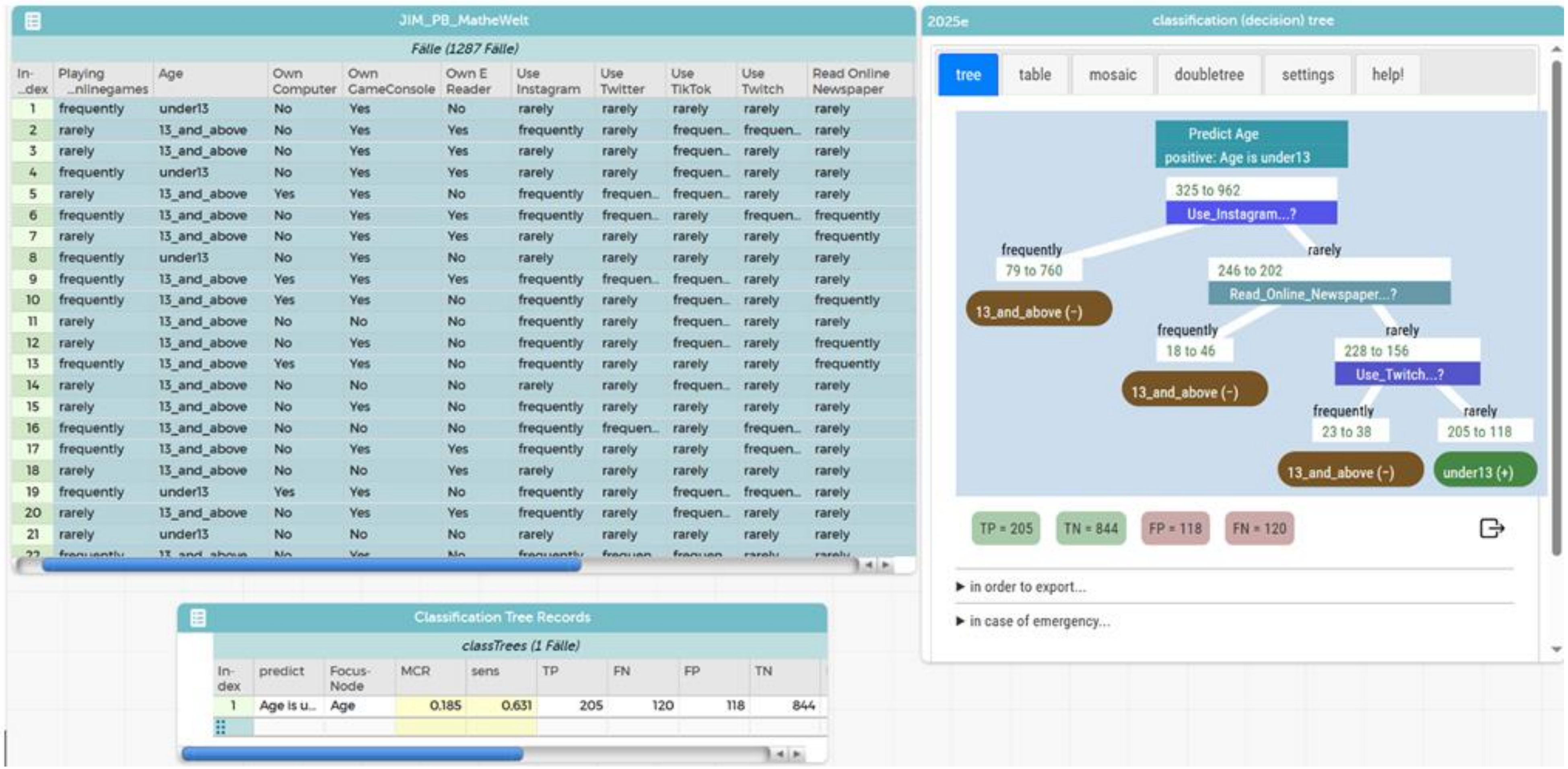
- Grade, Age, Gender, …
- Owning digital devices
    - Computer, Game Console, Tablet, …
- Use of online platforms, social media
    - Instagram, Facebook, TikTok, Youtube, Twitch…
- Use of classical media: Online-Newspapers. …
- Gaming
- ….

Variables (~150) of the YOU-PB data set

- Grade, Age, Gender, ..
- Owning digital devices
  - Computer, Game Console, Tablet, …
- Use of online platforms, social media
  - **Instagram**, Facebook, TikTok, Youtube, **Twitch…**
- Use of classical media: **Online-Newspapers. …**
- Gaming
- ….

Age

How can TikTok find out your true age?



JIM_PB_MatheWelt

Fälle (1287 Fälle)

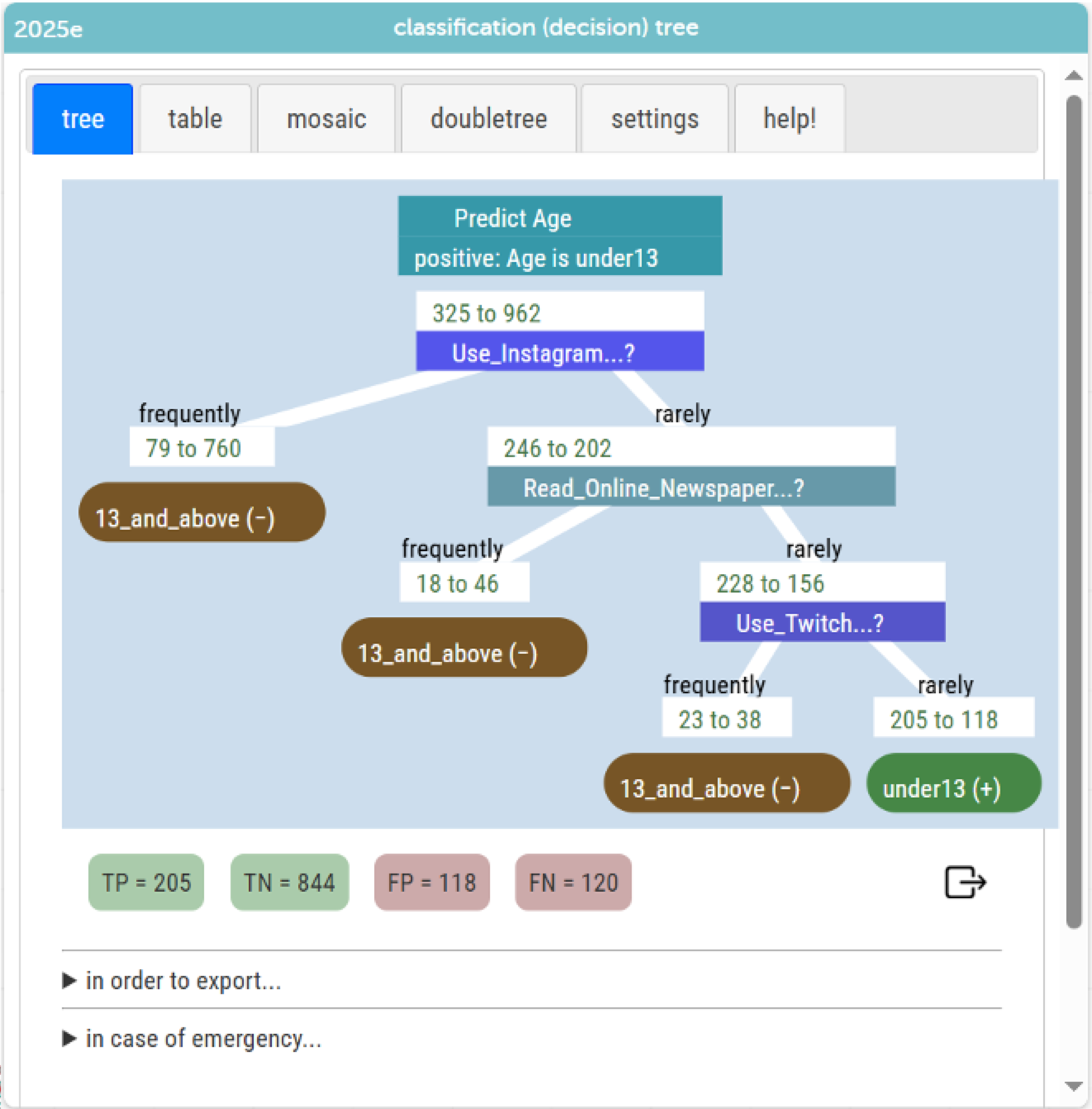
In-..dex	Playing ...nlinegames	Age	Own Computer	Own GameConsole	Own E Reader	Use Instagram	Use Twitter	Use TikTok	Use Twitch	Read Online Newspaper
1	frequently	under13	No	Yes	No	rarely	rarely	rarely	rarely	rarely
2	rarely	13_and_above	No	Yes	Yes	frequently	rarely	frequen...	frequen...	rarely
3	rarely	13_and_above	No	Yes	Yes	rarely	rarely	frequen...	rarely	rarely
4	frequently	under13	No	Yes	Yes	rarely	rarely	frequen...	rarely	rarely
5	rarely	13_and_above	Yes	Yes	No	frequently	frequen...	frequen...	rarely	rarely
6	frequently	13_and_above	No	Yes	Yes	frequently	frequen...	rarely	frequen...	frequently
7	rarely	13_and_above	No	Yes	Yes	rarely	rarely	rarely	rarely	frequently
8	frequently	under13	No	Yes	No	rarely	rarely	rarely	rarely	rarely
9	frequently	13_and_above	Yes	Yes	Yes	frequently	frequen...	frequen...	rarely	rarely
10	frequently	13_and_above	Yes	Yes	No	frequently	rarely	frequen...	rarely	frequently
11	rarely	13_and_above	No	No	No	frequently	rarely	frequen...	rarely	rarely
12	rarely	13_and_above	No	Yes	No	frequently	rarely	frequen...	rarely	frequently
13	frequently	13_and_above	Yes	Yes	No	frequently	rarely	rarely	rarely	frequently
14	rarely	13_and_above	No	No	No	rarely	rarely	frequen...	rarely	rarely
15	rarely	13_and_above	No	Yes	No	frequently	rarely	rarely	rarely	rarely
16	frequently	13_and_above	No	No	No	frequently	frequen...	rarely	frequen...	rarely
17	frequently	13_and_above	No	Yes	Yes	frequently	rarely	rarely	frequen...	rarely
18	rarely	13_and_above	No	No	Yes	rarely	rarely	rarely	rarely	rarely
19	frequently	under13	Yes	Yes	No	frequently	rarely	frequen...	frequen...	rarely
20	rarely	13_and_above	No	Yes	Yes	frequently	rarely	frequen...	rarely	rarely
21	rarely	under13	No	No	No	rarely	rarely	rarely	rarely	rarely

Classification Tree Records

classTrees (1 Fälle)

In-dex	predict	Focus-Node	MCR	sens	TP	FN	FP	TN
1	Age is u...	Age	0.185	0.631	205	120	118	844

Screenshot from CODAP (ARBOR plugin)



Misclassification rate 18,5 %

Evaluation of the tree: confusion matrix

Real age	Predicted age ≥ 13	Predicted age < 13
≥ 13	844	118 (FP)
< 13	120 (FN)	205

Misclassification rate 18,5 %

The two error types have different consequences
- 120 can continue to use TikTok, although < 13
- 118 are temporarily banned, but can be asked for an age proof for re-approval

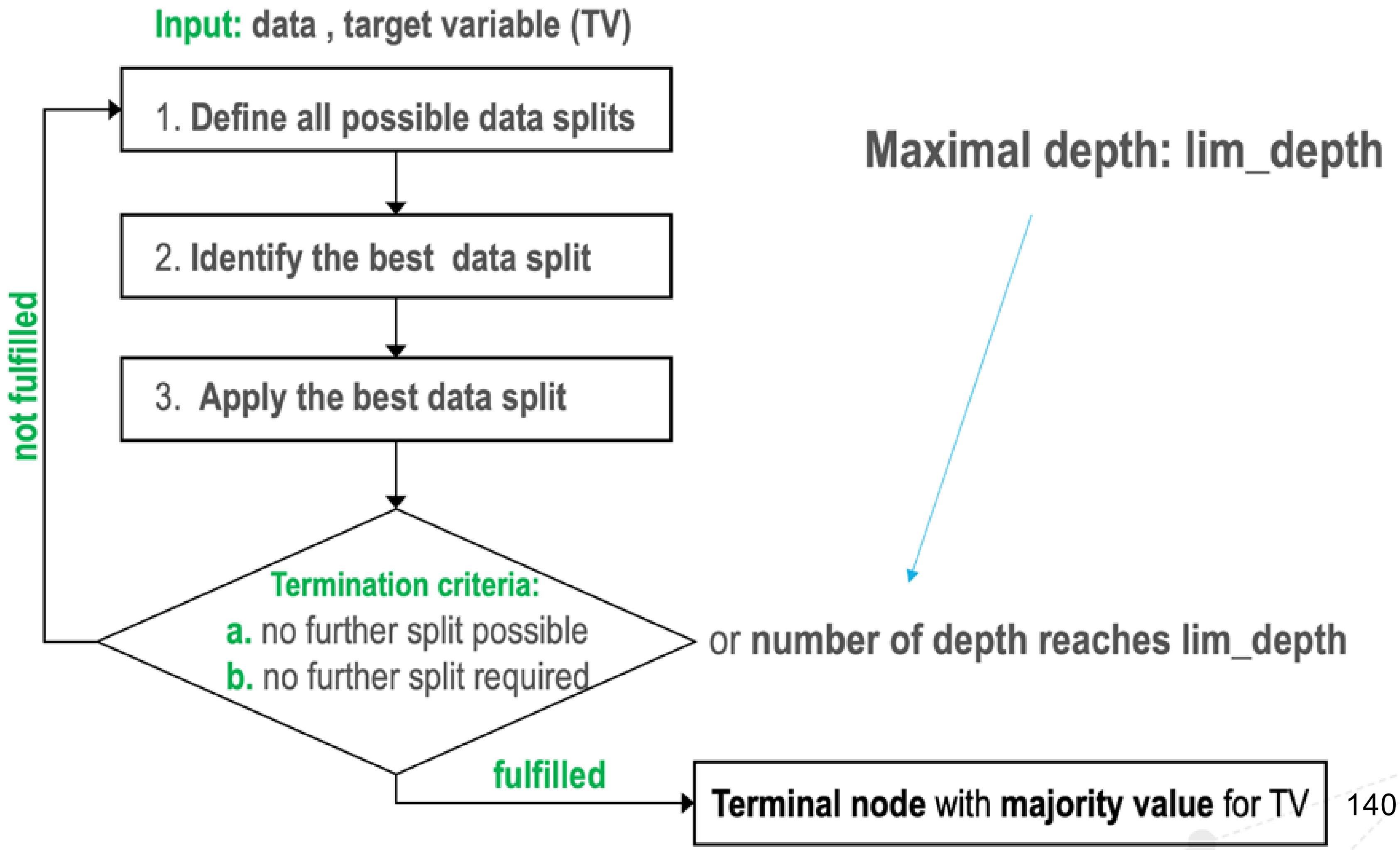
Construction of trees in the classroom

- Students can choose variables themselves
    - context based: A variable X1 may (probabilistically) predict age
    - data-based: A variable X2 is tentatively used for splitting, and the number of misclassifications can be assessed in CODAP
    - several variables can be compared by number of misclassifications
- Pragmatic termination criteria can be used
    - maximal depth
    - pure subset
    - no (relevant) variable left

Representation of the learning algorithm for building a decision tree: algorithmic thinking

Implementation in Python-based Jupyter Notebooks
(Yannik Fleischer)

- menu-based
- different types with different coding skills required

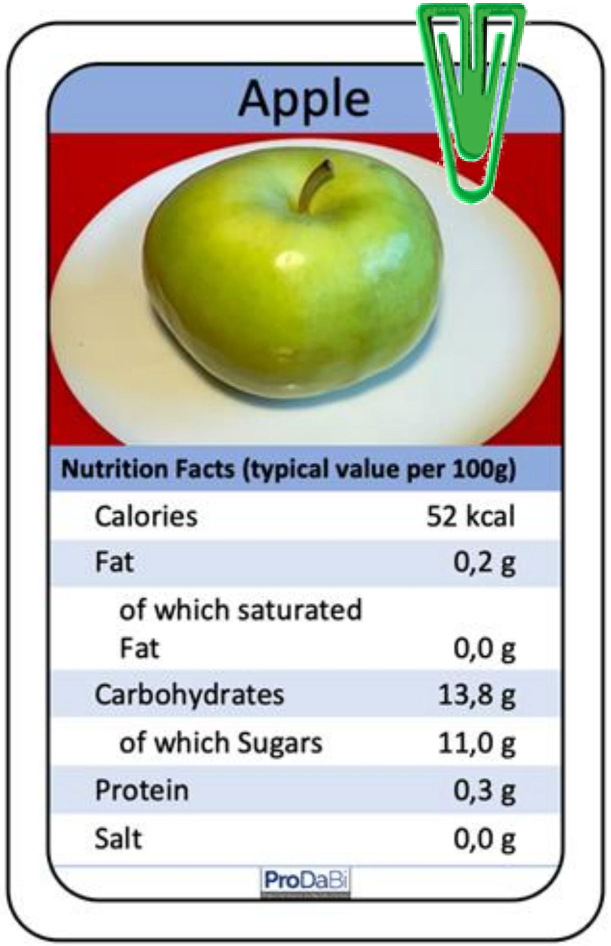


Example 2 Grade 5/6: Decision Trees unplugged with data cards

Guiding Questions:

- How can we use **nutrition facts** to decide whether a food is rather **recommendable** or rather **not recommendable**?

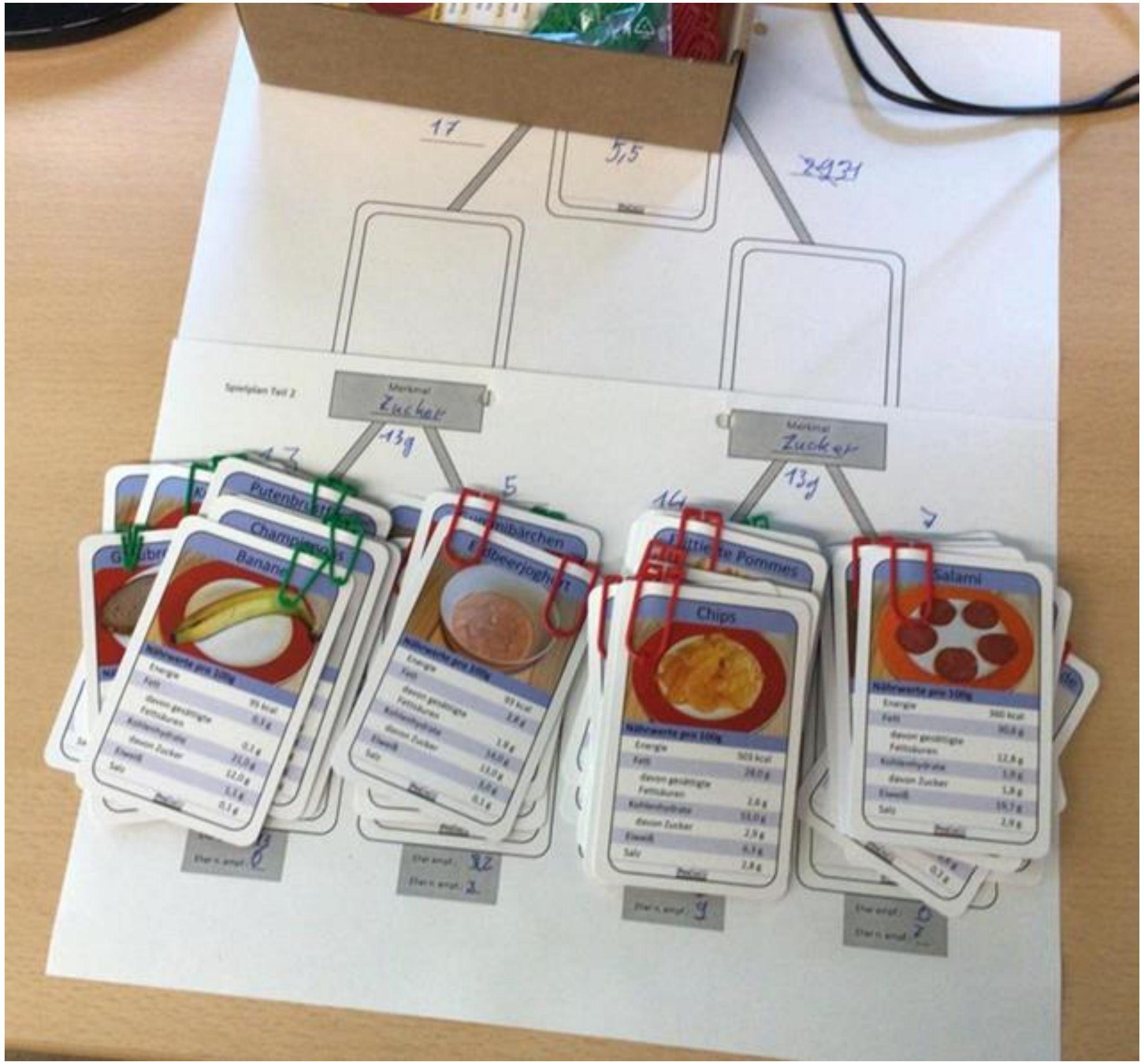
The material



- 55 data cards about food items
  - nutrition facts (typical value per 100g)
- green and red paper clips to label the cards
- worksheets and slides

Also joint work RB with Yannik Fleischer und Susanne Podworny (Podworny, Biehler, R., & Fleischer, Y. 2025)

Impressions of students work in class



Photos: Susanne Podworny

AI meets theatre education

Theatre project „Apple vs. Popcorn – Why not ask an AI“ – Gymnasium Theodorianum Paderborn

Photo Universität Paderborn, Johanna Pietsch https://www.uni-paderborn.de/nachricht/99642, Used with permission

Structure of presentation: our perspectives

1. Introduction: **Data in society, data science education and citizen empowerment** (Rolf Biehler on behalf of the team)

2. Civic statistics and humanistic perspectives on data literacies education in the U.S. and Europe (Travis Weiland)

3. Critical perspectives on data literacy emerging from Latin America (Lucía Zapata-Cardona)

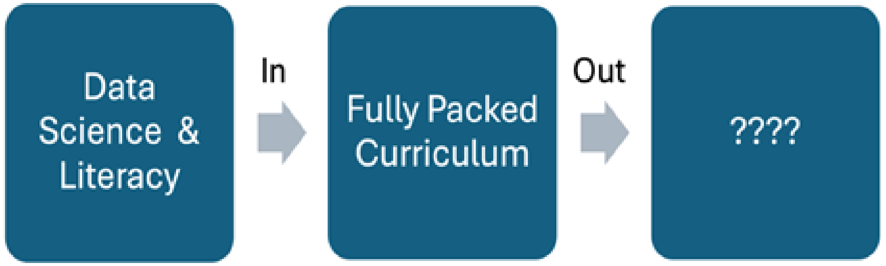
4. Joint discourse between mathematical modeling and statistics/data science communities (Takashi Kawakami)

5. What can mathematics/statistics education contribute to Artificial Intelligence/Machine Learning literacy (Rolf Biehler)

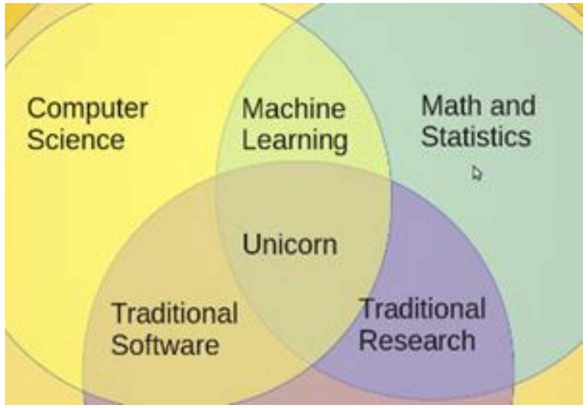
6. Conclusion: **Challenges for future development** (Rolf Biehler and Erna Lampen)

Future tasks

- Conceptual reviews and clarifications of the literacies debate
- Identifying priorities of topics, tools and contexts
- In-depth qualitative and quantitative studies of teaching and learning processes
- Designing assessment tools in particular for higher order skills
- Scaling up: Organizing systemic change including collaboration with schools, teacher education, school administration
- Work on the following problems:
- The curriculum overload problem



The data science **education** unicorn problem



An AI-liking unicorn (according to DALL-E)

Conceptual challenges as imminent task

"It has become a truism, at least among statisticians, that while statistics is a mathematical science, it is not a subfield of mathematics." Moore & Cobb (2000)

Calls for and increased cooperation between professional societies (ASA, MAA, AMS, IMS in USA) has yet to yield consensus about curriculum options between mathematics and statistics.

STATISTICS IS, ALAS, DIFFERENT. Should statisticians gloat?

"Not at all. …The picture that emerges when all the pieces come together is one of **organizational weakness**. The advantages of statistics over mathematics in our current environment are cultural, and cultural strength rarely outweighs organizational weakness. God is on the side of the big battalions. Mathematics has the big battalions, and statistics has a few guerrillas scattered about the academic jungles. Mathematics is far likelier than statistics to have an extended, if not prosperous, future."

Moore & Cobb 2000, p.620

Weak organizational players

- Statisticians active in reform have reached consensus about teaching, but they **lack institutional power**. (see Moore & Cobb 2000, p. 620)
- The lack of institutional power at undergraduate level feeds math/stats wars at school level
- Statistics have links to many fields, as **a methodological discipline rather than a core substantive area** - is this why hardly any calls for statistics as a full, separate high school subject are heard?

Weak organizational players

- Does the field have a core?
  " Statisticians become more identified with the area of their applied work than with statistics as a free-standing discipline. A pessimistic (or perhaps simply realistic) vision of the future sees statistics dissipating back into other fields."

- The threat from information technology.
  "Although statistics has been re-energized and redirected by computing, it now risks being **engulfed by information technology.** New areas in which statistical ideas offer promise are being pursued more vigorously by non-statisticians than by statisticians. **Information technology is now the most important methodology for most scientific fields, displacing both statistics and mathematics from their traditional roles**. (Moore & Cobb 2000, p.621)

References

Section 1 References

Abner, K. (2020). Data literacy as digital humanities literacy: Exploration of threshold concepts. *Literacies in a digital humanities context: A dh+ lib special issue*, 19-22.

Acker, A., Bowler, L., & Pangrazio, L. (2024). Guest editorial: Special issue – perspectives on data literacies. *Information and Learning Sciences*, *125*(3/4), 157-162. https://doi.org/10.1108/ILS-03-2024-266

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, *6*. https://doi.org/10.1016/j.caeo.2024.100173

Álvarez-Arroyo, R., Batanero, C., & Gea, M. M. (2024). Probabilistic literacy and reasoning of prospective secondary school teachers when interpreting media news. *ZDM – Mathematics Education*. https://doi.org/10.1007/s11858-024-01586-8

Ärlebäck, J. B., & Kawakami, T. (2023). The Relationships Between Statistics, Statistical Modelling and Mathematical Modelling. In G. Greefrath, S. Carreira, & G. A. Stillman (Eds.), *Advancing and Consolidating Mathematical Modelling: Research from ICME-14* (pp. 293-309). Springer International Publishing. https://doi.org/10.1007/978-3-031-27115-1_18

Bargagliotti, A., Binder, W., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., Huchting, K., Larson, S., Miric, N., Rovetti, R., Seal, K., & Zachariah, T. (2020). Undergraduate Learning Outcomes for Achieving Data Acumen. *Journal of Statistics Education*, *28*(2), 197-211. https://doi.org/10.1080/10691898.2020.1776653

Biehler, R. (2022). Revisiting Fundamental Ideas for Statistics Education From the Perspective of Machine Learning and Its Applications. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. https://doi.org/10.52041/iase.icots11.T1A2

Biehler, R., De Veaux, R., Engel, J., Kazak, S., & Frischemeier, D. (2022). Editorial: Research on Data Science Education. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.606

Center for Universal Education at the Brookings Institution, UNESCO, & GEFI-YAG (Eds.). (2017). *Measuring Global Citizenship Education - A Collection of Practices and Tools*. The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/04/global_20170411_measuring-global-citizenship.pdf.

Chevallard, Y. (1985). *La Transposition Didactique: Du Savoir Savant au Savoir Enseigneé*. La Penseé Sauvage.

Chiu, T. K. F., & Sanusi, I. T. (2024). Define, foster, and assess student and teacher AI literacy and competency for all: Current status and future research direction. *Computers and Education Open*. https://doi.org/10.1016/j.caeo.2024.100182

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press.

Section 1 References

Deutscher Ethikrat. (2023). Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz - Stellungnahme https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/

Fielding, J., Makar, K., & Ben-Zvi, D. (2025). Developing students' reasoning with data and data-ing. ZDM – Mathematics Education, 57(1), 1-18. https://doi.org/10.1007/s11858-025-01671-6

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. International Statistical Review, 70(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. Educational Studies in Mathematics, 111(1), 5–28. https://doi.org/10.1007/s10649-022-10151-7

Gal, I., Nicholson, J., & Ridgway, J. (2022). A Conceptual Framework for Civic Statistics and Its Educational Applications. In J. Ridgway (Ed.), Statistics for Empowerment and Social Engagement (pp. 37-66). https://doi.org/10.1007/978-3-031-20748-8_3

Gal, I., & Ograjensek, I. (2018). Developing official statistics literacy: A proposed model and implications. In P. Kovács (Ed.), Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat. University of Szeged.

Geiger, V., Gal, I., & Graven, M. (2023). The connections between citizenship education and mathematics education. ZDM – Mathematics Education, 55(5), 923-940. https://doi.org/10.1007/s11858-023-01521-3

Gould, R. (2017). Data literacy is statistical literacy. Statistics Education Research Journal, 16(1), 22-25.

Gulson, K. N., Sellar, S., & Webb, P. T. (2022). Algorithms of education : how datafication and artificial intelligence shape policy. University of Minnesota Press.

Höper, L., & Schulte, C. (2023). The data awareness framework as part of data literacies in K-12 education. Information and Learning Sciences. https://doi.org/10.1108/ils-06-2023-0075

IDSSP Curriculum Team. (2019). Curriculum Frameworks for Introductory Data Science. The International Data Science in Schools Project. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Kazak, S., Fujita, T., & Turmo, M. P. (2021). Students' informal statistical inferences through data modeling with a large multivariate dataset. Mathematical Thinking and Learning, 25(1), 23-43. https://doi.org/10.1080/10986065.2021.1922857

Leaning, M. (2017). Media and information literacy: An integrated approach for the 21st century. Chandos Publishing.

Section 1 References

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), International handbook of research in statistics education (pp. 229–259). Springer. https://doi.org/10.1007/978-3-319-66195-7

Louie, J. (2022). Critical data literacy: Creating a more just world with data National Academy of Sciences' Workshop on Foundations of Data Science for Students in Grades K–12, https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D16254F310D01BBDA873920E4EFB8151F2D8334181AA

Ma, R., Sanusi, I. T., Mahipal, V., Gonzales, J. E., & Martin, F. G. (2023). Developing Machine Learning Algorithm Literacy with Novel Plugged and Unplugged Approaches Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1,

MacGillivray, H., Ridgway, J., & Gould, R. (2021). Preface. Teaching Statistics, 43(S1). https://doi.org/10.1111/test.12282

Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. ZDM – Mathematics Education, 55(5), 967-979. https://doi.org/https://doi.org/10.1007/s11858-022-01450-7

Marshall, B. H. (2022). Data Conscience: Algorithmic Siege on Our Humanity. John Wiley & Sons.

Martínez-Castro, C. A., Zapata-Cardona, L., & Jones, G. L. (2023). Critical Citizenship in Statistics Teacher Education. In Research on Reasoning with Data and Statistical Thinking: International Perspectives (pp. 213-226). https://doi.org/10.1007/978-3-031-29459-4_17

Matuk, C., Knight, S., & DesPortes, K. (2022). Educational Technology & Society: Learning at the intersection of data literacy and social justice - Call for papers. Educational Technology & Society, 25(4).

Mendoza-Chan, J., & Pee, L. G. (2024). Digital skilling of working adults: A systematic review. Computers & Education, 218. https://doi.org/10.1016/j.compedu.2024.105076

Mostacedo-Marasovic, S.-J., Olsen, A. A., & Forbes, C. T. (2024). Supporting Secondary Students' Understanding of Earth's Climate System and Global Climate Change Using EzGCM: A Cross-Sectional Study. Journal of Science Education and Technology, 33(2), 178-194. https://doi.org/10.1007/s10956-023-10067-8

Murillo, L. F. R., Wylie, C., & Bourne, P. (2023). Critical data ethics pedagogies: Three (non-rival) approaches. Big Data & Society, 10(2). https://doi.org/10.1177/20539517231203666

National Academies of Sciences Engineering Medicine. (2023). Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop. https://doi.org/https://doi.org/10.17226/26852

NCTM, NSTA, ASA, NCSS, & CSTA. (2024). Data Science A Joint Position of NCTM, NSTA, ASA, NCSS, and CSTA. https://www.nctm.org/Standards-and-Positions/Position-Statements/Data-Science/

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. New Media & Society, 21(2), 419-437. https://doi.org/10.1177/1461444818799523

Section 1 References

Podworny, S., Frischemeier, D., Dvir, M., & Ben-Zvi, D. (Eds.). (2024). Minerva School 2022: Reasoning with data models and modeling in the big data era. https://doi.org/10.17619/UNIPB/1-1815.
Ridgway, J. (Ed.). (2022). Statistics for empowerment and Social Engagement – Teaching Civic Statistics to Develop Informed Citizens. Springer. https://doi.org/10.1007/978-3-031-20748-8.
Sander, I. (2024). Critical Datafication Literacy. transcript.
Schield, M. (2005). Statistical Literacy Curriculum Design. In G. Burril & M. Camden (Eds.), Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable (pp. 54-74). International Statistical Institute.
Shreiner, T. L. (2020). Data-literate citizenry: how US state standards address data and data visualizations in social studies. Information and Learning Sciences, 121(11/12), 909-931. https://doi.org/10.1108/ils-03-2020-0054
Shreiner, T. L. (2023). Teaching Data Literacy for Civic Competence: The Social Studies Teacher's Crucial Role. Social Education, 87(4), 262-269.
Shreiner, T. L., & Guzdial, M. (2024). Everyday Equitable Data Literacy is Best in Social Studies. In Improving Equity in Data Science (pp. 133-150). https://doi.org/10.4324/9781003364634-7
Siller, H.-S., Geiger, V., & Kaiser, G. (2024). Researching mathematical modelling education in disruptive times. Springer.
Ślosarski, B. (2023). Data arenas: The relational dynamics of data activism. Big Data & Society, 10(1). https://doi.org/10.1177/20539517231177617
Sukol, S. (2024a). Beyond Borders 2024: Primary and Secondary Data Science Education Around the World. Data Science 4 Everyone.
Sukol, S. (2024b). State of The Field: Data Science and Data Literacy Education in US K-12. DataScience 4Everyone.
UNESCO. (2005). Education for all: Literacy for Life. United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54676/HFRH4626
Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., & Wagenknecht, K. (Eds.). (2021). The science of citizen science. Springer. https://doi.org/10.1007/978-3-030-58278-4.
Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. Educational Studies in Mathematics, 96(1), 33-47. https://doi.org/10.1007/s10649-017-9764-5
Westheimer, J., & Kahne, J. (2004). What kind of citizen? The politics of educating for democracy. American Educational Research Journal, 41(2), 237-269.
Wilkerson, M. H., & Polman, J. L. (2020). Situating Data Science: Exploring How Relationships to Data Shape Learning. Journal of the Learning Sciences, 29(1), 1-10. https://doi.org/10.1080/10508406.2019.1705664

Section 2 References

Aridor, K., Dvir, M., Tsybulsky, D., & Ben-Zvi, D. (2023). Living the DReaM: The interrelations between statistical, scientific and nature of science uncertainty articulations through citizen science. *Instructional Science*, *51*(5), 729–762. https://doi.org/10.1007/s11251-023-09626-8

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II)* (Second edition). American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf

da Silva, A. S., Barbosa, M. T. S., de Souza Velasque, L., da Silveira Barroso Alves, D., & Magalhães, M. N. (2021). The COVID-19 epidemic in Brazil: How statistics education may contribute to unravel the reality behind the charts. *Educational Studies in Mathematics*, *108*(1–2), 269–289. https://doi.org/10.1007/s10649-021-10112-6

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. https://doi.org/10.52041/serj.v16i1.213

Engledowl, C., & Weiland, T. (2021). Data (mis)representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, *29*(2), 160–164. https://doi.org/10.1080/26939169.2021.1915215

Erickson, T., & Chen, E. (2021). Introducing data science with data moves and CODAP. *Teaching Statistics*, *43*(S1). https://doi.org/10.1111/test.12240

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education*, *12*(1). https://doi.org/10.5070/T5121038001

Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Gundlach, E., Maybee, C., & O'Shea, K. (2015). Statistical literacy social media project for the masses. *The Journal of Faculty Development*, *29*(2), 71–80.

Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., Finzer, W., & Rosé, C. P. (2022). An empirical analysis of high school students' practices of modelling with unstructured data. *British Journal of Educational Technology*, *53*(5), 1114–1133. https://doi.org/10.1111/bjet.13253

Kahn, J. (2020). Learning at the intersection of self and society: The family geobiography as a context for data science education. *Journal of the Learning Sciences*, *29*(1), 57–80. https://doi.org/10.1080/10508406.2019.1693377

Kahn, J. B., Peralta, L. M., Rubel, L. H., Lim, V. Y., & Herbel-Eisenmann, B. (2023). Notice, wonder, feel, act, and reimagine as a path toward social justice in data science education. *Educational Technology & Society*, *25*(4), 80–92. https://doi.org/10.30191/ETS.202210_25(4).0007

Section 2 References

Kim, A. Y., Ismay, C., & Chunn, J. (2018). The fivethirtyeight R Package: “Tame Data” Principles for Introductory Statistics and Data Science Courses. *Technology Innovations in Statistics Education*, *11*(1). https://doi.org/10.5070/T5111035892

Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, *16*(2), 191–212. https://doi.org/10.52041/serj.v16i2.190

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305–325. https://doi.org/10.1007/s10649-013-9529-8

Lanouette, K., Cortes, K. L., Lopez, L., Bakal, M., & Wilkerson, M. H. (2024). Exploring climate change through students’ place connections and public data sets. *Science Scope*, *47*(3), 18–25. https://doi.org/10.1080/08872376.2024.2340444

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664–672. https://doi.org/10.3102/0013189X211048810

Lim, V. Y., Peralta, L. M. M., Rubel, L. H., Jiang, S., Kahn, J. B., & Herbel-Eisenmann, B. (2023). Keeping pace with innovations in data visualizations: A commentary for mathematics education in times of crisis. *ZDM – Mathematics Education*, *55*(1), 109–118. https://doi.org/10.1007/s11858-022-01449-0

Louie, J. (2022). *Critical data literacy: Creating a more just world with data* (Workshop on Foundations of Data Science for Students in Grades K-12). National Academy of Sciences Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D16254F310D01BBDA873920E4EFB8151F2D8334181AA?noSaveAs=1

Louie, J., Stiles, J., Fagan, E., Chance, B., & Roy, S. (2023). Building toward Critical Data Literacy with Investigations of Income Inequality. *Educational Technology & Society*, *25*(4), 142–163. https://doi.org/10.30191/ETS.202210_25(4).0011

Nicholson, J., Ridgway, J., & Gal, I. (2018). Understanding civic statistics: A conceptual framework and its educational applications. *A Product of the ProCivicStat Project*. http://IASE-web.org/ISLP/PCS

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming Racially Literate About Data and Data-Literate About Race: Data Visualizations in the Classroom as a Site of Racial-Ideological Micro-Contestations. *Cognition and Instruction*, *34*(4), 361–388. https://doi.org/10.1080/07370008.2016.1210418

Section 2 References

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

Poling, L. L., & Weiland, T. (2020). Using an interactive platform to recognize the intersection of social and spatial inequalities. *Teaching Statistics*, *42*(3), 108–116. https://doi.org/10.1111/test.12234

Radinsky, J., & Tabak, I. (2022). Data practices during COVID Everyday sensemaking in a high-stakes information. *British Journal of Educational Technology*, *53*(5), 1221–1243. https://doi.org/10.1111/bjet.13252

Reigh, E., Escudé, M., Bakal, M., Rivero, E., Wei, X., Roberto, C., Hernández, D., Yada, A., Gutiérrez, K., & Wilkerson, M. H. (2022). Mapping racespace: Data stories as a tool for environmental and spatial justice. *Occasional Paper Series*, *2022*(48), 79–95. https://doi.org/10.58295/2375-3668.1452

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology*, *53*(5), 1179–1201. https://doi.org/10.1111/bjet.13245

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2016). Teaching mathematics for spatial justice: Beyond a victory narrative. *Harvard Educational Review*, *86*(4), 556–579.

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences*, *26*(4), 643–687. https://doi.org/10.1080/10508406.2017.1336440

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016). Teaching mathematics for spatial justice: An investigation of the lottery. *Cognition and Instruction*, *34*(1), 1–26. https://doi.org/10.1080/07370008.2015.1118691

Rubel, L. H., & Nicol, C. (2020). The power of place: Spatializing critical mathematics education. *Mathematical Thinking and Learning*, *22*(3), 173–194. https://doi.org/10.1080/10986065.2020.1709938

Rubel, L. H., Nicol, C., & Chronaki, A. (2021a). A critical mathematics perspective on reading data visualizations: Reimagining through reformatting, reframing, and renarrating. *Educational Studies in Mathematics*, *108*(1–2), 249–268. https://doi.org/10.1007/s10649-021-10087-4

Rubel, L. H., Nicol, C., & Chronaki, A. (2021b). A critical mathematics perspective on reading datavisualizations: Reimagining through reformatting,reframing, and renarrating. *Educational Studies in Mathematics*, *108*, 249–268. https://doi.org/10.1007/s10649-021-10087-4

Shapiro, B. R., Meng, A., Rothschild, A., Gilliam, S., DiSalvo, C., & DiSalvo, B. (2023). *"Bettering Data": The Role of Everyday Language and Visualization in Critical Novice Data Work*.

Section 2 References

Shreiner, T. L., & Guzdial, M. (2022). The information won't just sink in: Helping teachers provide technology-assisted data literacy instruction in social studies. *British Journal of Educational Technology*, *53*(5), 1134–1158. https://doi.org/10.1111/bjet.13255

Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space. *Journal of the Learning Sciences*, *29*(1), 127–153. https://doi.org/10.1080/10508406.2019.1693378

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1–8.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23.

Wilkerson, M. H., & Laina, V. (2018). Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM*, *50*(7), 1223–1235. https://doi.org/10.1007/s11858-018-0974-9

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2022). Exploring variability during data preparation: A way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, *24*(4), 312–330. https://doi.org/10.1080/10986065.2021.1922838

Section 3 References

Cabrera, G. P., Tauber, L. M., & Pochulu, M. D. (2023). Diseño e implementación de rúbricas para caracterizar y evaluar la alfabetización estadística crítica en contextos de lectura [Design and implementation of rubrics to characterize and assess critical statistical literacy in reading contexts]. *Revista de Educación Estadística, 2* (1). https://doi.org/10.29035/redes.2.1.4

Campos, C. R. (2016). *Towards Critical Statistics Education: Theory and practice*. Lambert Academic Publishing.

de Sousa, L. M. R. (2022). O lado obscuro dos algoritmos: como a promessa de liberdade da internet resultou em invasão de privacidade e ameaças à democracia [The Dark Side of Algorithms: How the Internet's promise of freedom has led to invasions of privacy and threats to democracy]. *Conexão ComCiência*, *2*(1), 1-8. https://revistas.uece.br/index.php/conexaocomciencia/article/view/5515

Ferreira da Silva, A. F., & de Oliveira Souza, L. (2024). La interseccionalidad como forma de (re)pensar la educación estadística [Intersectionality as a way of (re)thinking statistical education]. *Revista De Educación Estadística*, *3*, 1-23. https://doi.org/10.29035/redes.3.1.8

Freire, P. (1970). *Pedagogía del oprimido* [Pedagogy of the Oppressed]. Translator Jorge Mellado. Siglo veintiuno editores.

Giordano, C. C., Pereira, F. A., & Souza, F. dos S. (2022). El desarrollo de la alfabetización estadística de los estudiantes de secundaria brasileña: un enfoque a través de las Estadísticas Cívicas [Developing statistical literacy among Brazilian high school students: A Civic Statistics approach]. *Revista Baiana De Educação Matemática*, *3*(01), e202212. https://doi.org/10.47207/rbem.v3i01.15516

Section 3 References

Giroux, H. (1993). *La escuela y la lucha por la ciudadanía*. *La Pedagogía crítica de la época moderna* [School and the struggle for citizenship: Critical Pedagogy in the Modern Age]. Siglo veintiuno editores.

Martínez-Castro, C. A., & Zapata-Cardona, L. (2022). Formación inicial de profesores de estadística en una perspectiva crítica [Initial training of statistics teachers from a critical perspective]. In A. Salcedo & D. Diaz-Levicoy (Eds.) *Formación del profesorado para enseñar estadística: retos y oportunidades* (pp. 367-382). Universidad Católica de Maule.

Raffaghelli, J. (2022). Alfabetización en datos y justicia social ¿Un oxímoron? Respuestas desde la contra-hegemonía [Data Literacy and Social Justice: An oxímoron? Responses from the counter-hegemony]. *Revista Izquierdas, 51,* 1-18. https://www.izquierdas.cl/images/pdf/2022/51/art17.pdf

Skovsmose, O. (2014). *Critique as Uncertainty.* The Montana Mathematics Enthusiast: Monograph Series in Mathematics Education. Information Age Publishing.

Tygel, A.F., & Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics, 12*(3), 108—121. https://doi.org/10.15353/joci.v12i3.3279

Section 3 References

Tygel, A.F., & Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics, 12*(3), 108—121. https://doi.org/10.15353/joci.v12i3.3279

Tironi, M. & Valderrama, M. (2021). Descolonizando los sistemas algorítmicos: diseño crítico para la problematización de algoritmos y datos digitales desde el Sur [Decolonizing algorithmic systems: Critical design for problematizing algorithms and digital data from the South]. *Palabra Clave*, *24*(3), e2432. https://doi.org/10.5294/pacla.2021.24.3.2

Valente, J. & Grohmann, R. (2024). Critical data studies with Latin America: Theorizing beyond data colonialism. *Big Data & Society 11*(1), 1-13. https://doi.org/10.1177/20539517241227875

Zapata-Cardona, L. (2023). The possibilities of exploring nontraditional datasets with young children. *Teaching Statistics*, *45*(S1), S22–S29. https://doi.org/10.1111/test.12349

Section 4 References

Andersson, C. H., & Register, J. T. (2023). An examination of pre-service mathematics teachers' ethical reasoning in big data with considerations of access to data. *The Journal of Mathematical Behavior*, *70*. https://doi.org/10.1016/j.jmathb.2022.101029

Aridor, K., Dvir, M., Tsybulsky, D., & Ben-Zvi, D. (2023). Living the DReaM: The interrelations between statistical, scientific and nature of science uncertainty articulations through citizen science. *Instructional Science*, *51*(5), 729–762. https://doi.org/10.1007/s11251-023-09626-8

Ärlebäck, B. J., Frejd, P., & Doerr, H. M. (2021). Exploring students' models of sampling and inference with nominal variables. *Quadrante*, *30*(1), 158–177. https://doi.org/10.48489/quadrante.2365

Ärlebäck, J. B., & Kawakami, T. (2023). The relationships between statistics, statistical modelling and mathematical modelling. In G. Greefrath, S. Carreira, & G. Stillman (Eds.), *Advancing and consolidating mathematical modelling: Research from ICME-14* (pp. 293–309). Springer. https://doi.org/10.1007/978-3-031-27115-1_18

Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *ZDM Mathematics Education*, *38*(3), 293–301. https://doi.org/10.1007/BF02652812

Burkhardt, H. (2021). Modelling in school mathematics: Past achievements—current challenges. In F. K. S. Leung, G. A. Stillman, G. Kaiser, & K. L. Wong (Eds.), *Mathematical modelling education in East and West* (pp. 341–361). Springer. https://doi.org/10.1007/978-3-030-66996-6_29

Bybee, R. (2018). *STEM education now more than ever*. NSTA press.

Davis, P. J., & Hersh, R. (1986). *Descartes' dream: The world according to Mathematics*. Harcourt Brace Jovanovich.

Section 4 References

Dvir, M., & Ben-Zvi, D. (2023). Informal statistical models and modeling. *Mathematical Thinking and Learning*, *25*(1), 79–99. https://doi.org/10.1080/10986065.2021.1925842

English, L. D. (2016). Advancing mathematics education research within a STEM environment. In K. Fry, S. Dole, M. Goos, K. Maker, A. Bennison, & J. Visnovska (Eds.), *Research in mathematics education in Australasia 2012–2015* (pp. 353–371). Springer. https://doi.org/10.1007/978-981-10-1419-2_17

English, L. D. (2021). Mathematical and interdisciplinary modeling in optimizing young children's learning. In J. M. Suh, M. H. Wickstrom, & L. D. English (Eds.), *Exploring mathematical modeling with young learners* (pp. 3–24). Springer. https://doi.org/10.1007/978-3-030-63900-6_1

Frejd, P., & Vos, P. (2024) The spirit of mathematical modeling – a philosophical study on the occasion of 50 years of mathematical modeling education. *The Mathematics Enthusiast*, *21*(1), 269–300. https://doi.org/10.54870/1551-3440.1626

Fry, K., English, L., & Makar, K. (2024). Cognitive tuning in the STEM classroom: Communication processes supporting children's changing conceptions about data. *Mathematics Education Research Journal*, *36*, 67–89. https://doi.org/10.1007/s13394-023-00465-x

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics*, *111*(1), 5–28. https://doi.org/10.1007/s10649-022-10151-7

Kawakami, T., & Saeki, A. (2022). A framework for describing and analysing data-driven modelling activities in school mathematics: From the perspectives of mathematical and statistical models. *Journal of Science Education in Japan*, *46*(4), 421–437. https://doi.org/10.14935/jssej.46.421

Section 4 References

Kawakami, T., & Saeki, A. (2024a). Extending data-driven modelling from school mathematics to school STEM education. In J. Anderson, & K. Makar (Eds.), *The contribution of mathematics to school STEM education: Current understandings* (pp. 221–239). Springer. https://doi.org/10.1007/978-981-97-2728-5_13

Kawakami, T., & Saeki, A. (2024b). Roles of mathematical and statistical models in data-driven modelling: A prescriptive modelling perspective. In H.-S. Siller, V. Geiger, & G. Kaiser (Eds.), *Researching mathematical modelling education in disruptive times* (pp. 595–605). Springer. https://doi.org/10.1007/978-3-031-53322-8_49

Kazak, S., Fujita, T., & Turmo, M. P. (2023). Students' informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning*, *25*(1), 23–43. https://doi.org/10.1080/10986065.2021.1922857

Lehrer, R., & Schauble, L. (Eds.). (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. Teachers College Press.

Lehrer, R., Wisittanawat, P., & Schauble, L. (2024). Designing for epistemic development. In Y. Li, Z. Zeng, & N. Song (Eds.), *Disciplinary and interdisciplinary education in STEM* (pp. 121–148). Springer. https://doi.org/10.1007/978-3-031-52924-5_7

Maass, K., Artigue, M., Burkhardt, H., Doorman, M., English, L. D., Geiger, V., & Schoenfeld, A. (2022). Mathematical modelling–A key to citizenship education. In N. Buckholtz, B. Schwarz, & K. Vorhöter (Eds.), *Initiationen mathematikdidaktischer Forschung* (pp. 31–50). Springer. https://doi.org/10.1007/978-3-658-36766-4_2

Maass, K., Geiger, V., Romero Ariza, M., & Goos, M. (2019). The role of mathematics in interdisciplinary STEM education. *ZDM Mathematics Education*, *51*(6), 869–884. https://doi.org/10.1007/s11858-019-01100-5

Section 4 References

Maass, K., Zehetmeier, S., Weihberger, A., & Flößer, K. (2023). Analysing mathematical modelling tasks in light of citizenship education using the COVID-19 pandemic as a case study. *ZDM Mathematics Education*, *55*(1), 133–145. https://doi.org/10.1007/s11858-022-01440-9

Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. *ZDM Mathematics Education*, *55*(5), 967–979. https://doi.org/10.1007/s11858-022-01450-7

McLean, J. A., Brady, C., Jung, H., Dominguez, A., & Glancy, A.W. (2023). Seeing the forest for the trees: Investigating students' data moves in a citizen science based model-eliciting activity. In G. Greefrath, S. Carreira, G. A. Stillman (Eds.), *Advancing and consolidating mathematical modelling: Research from ICME-14* (pp. 193–204). Springer. https://doi.org/10.1007/978-3-031-27115-1_12

Niss, M. (2015). Prescriptive modelling: Challenges and opportunities. In G. Stillman, W. Blum, & M. Salett Biembengut (Eds.), *Mathematical modelling in education research and practice* (pp. 67–79). Springer. https://doi.org/10.1007/978-3-319-18272-8_5

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling*. Routledge.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Pollak, H. O. (1977). The interaction between mathematics and other school subjects (Including integrated courses). In H. Athen, & H. Kunle (Eds.), *Proceedings of the Third International Congress on Mathematical Education* (pp. 255-264). Zentralblatt für Didaktik der Mathematik.

Rosa, M., Orey, D.C., & de Sousa Mesquita, A. P. S. (2023). An ethnomodelling perspective for the development of a citizenship education. *ZDM Mathematics Education*, *55*(5), 953–965. https://doi.org/10.1007/s11858-023-01472-9

Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. Springer. https://doi.org/10.1007/978-94-017-3556-8

Skovsmose, O. (2023). A Performative Interpretation of Mathematics. In M. A. V. Bicudo, B. Czarnocha, M. Rosa, & M. Marciniak (Eds.), *Ongoing advancements in philosophy of mathematics education* (pp. 269–292). Springer. https://doi.org/10.1007/978-3-031-35209-6_14

Section 4 References

Steffensen, L., Kacerja, S. (2021). Carbon Footprints Calculators and Climate Change. In. F. K. S. Leung, G. A. Stillman, G. Kaiser, & K. L. Wong (Eds.), *Mathematical modelling education in East and West* (pp. 513–523). Springer. https://doi.org/10.1007/978-3-030-66996-6_43

Stillman, G., & Brown, J. (2023). Modeling the phenomenon versus modeling the data set. *Mathematical Thinking and Learning*, *25*(3), 270–295. https://doi.org/10.1080/10986065.2021.2013144

Stillman, G.A., Ikeda, T., Schukajlow, S., de Loiola Araújo, J., & Ärlebäck, J. B. (2023). Survey of interdisciplinary aspects of the teaching and learning of mathematical modelling in mathematics education. In G. Greefrath, S. Carreira, G. A. Stillman (Eds.), *Advancing and consolidating mathematical modelling: Research from ICME-14* (pp. 21–41). Springer. https://doi.org/10.1007/978-3-031-27115-1_2

Van Dijke-Droogers, M., Drijvers, P., & Bakker, A. (2021). Statistical modeling processes through the lens of instrumental genesis. *Educational Studies in Mathematics*, *107*(2), 235–260. https://doi.org/10.1007/s10649-020-10023-y

Watson, J., Fitzallen, N., & Chick, H. (2020). What is the role of statistics in integrating STEM education? In J. Anderson, & Y. Li (Eds.), *Integrated approaches to STEM education* (pp. 91–115). Springer. https://doi.org/10.1007/978-3-030-52229-2_6

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2022). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, *24*(4), 312–330. https://doi.org/10.1080/10986065.2021.1922838

Zapata-Cardona, L., & Martínez-Castro, C. A. (2023). Statistical modeling in teacher education. *Mathematical Thinking and Learning*, *25*(1), 64–78. https://doi.org/10.1080/10986065.2021.1922859

Section 5 References

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, *6*. https://doi.org/10.1016/j.caeo.2024.100173

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) - A Framework for Statistics and Data Science Education*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Bata, K., Schmitz, A., & Eichler, A. (2022). Insights into the design of an introductory course for data science and machine learning for engineering students. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12), Feb. 2022 Bozen-Bolzano, Italy. hal-03751807,

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*(S1), S133-S142. https://doi.org/10.1111/test.12279

Biehler, R., & Schönbrodt, S. (Eds.). (2024). *KI verstehen: Wie Maschinen lernen [Understanding AI - How machines learn].mathematik lehren 244*. Friedrich Verlag.

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-023-00418-7

Erickson, T., & Engel, J. (2023). What goes before the CART? Introducing classification trees with Arbor and CODAP. *Teaching Statistics*, *45*(S1), S104-S113. https://doi.org/https://doi.org/10.1111/test.12347

Fergusson, A., & Pfannkuch, M. (2024). Using Grayscale Photos to Introduce High School Statistics Teachers to Reasoning with Digital Image Data. *Journal of Statistics and Data Science Education*, 1-16. https://doi.org/10.1080/26939169.2024.2351570

Fleischer, Y., Podworny, S., & Biehler, R. (2024). Datenbasiertes Entscheiden - Wie TikTok dein wahres Alter herausfinden kann [Data-based decision making: How TikTok can find out your true age]. MatheWelt. *mathematik lehren 244.*

Fleischer, Y., & Biehler, R. (2025). Exploring students' constructions of data-based decision trees after an introductory teaching unit on machine learning *ZDM Mathematics Education*, *57*(1).

Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and Learning Data-Driven Machine Learning with Educationally Designed Jupyter Notebooks. *Statistics Education Research Journal*, *21*(2). https://doi.org/10.52041/serj.v21i2.61

Fleischer, Y., Hüsing, S., Biehler, R., Podworny, S., & Schulte, C. (2022). Jupyter Notebooks for Teaching, Learning, and Doing Data Science. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics. https://doi.org/10.52041/iase.icots11.T10E3

Section 5 References

Fleischer, Y., Podworny, S., & Biehler, R. (2024). Teaching and Learning to Construct Data-Based Decision Trees Using Data Cards as the First Introduction to Machine Learning in Middle School. Statistics Education Research Journal, 23(1). https://doi.org/10.52041/serj.v23i1.450

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds Mach (Dordr), 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Martignon, L., Engel, J., & Erickson, T. (2022). A Transparent, Simple AI Tool for Constructing Efficient and Robust Fast and Frugal Trees for Classification Under Risk Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics

Mike, K., & Hazzan, O. (2022). Machine Learning for Non-Majors: A White Box Approach. Statistics Education Research Journal, 21(2). https://doi.org/10.52041/serj.v21i2.45

Olari, V., & Romeike, R. (2024). Data-related concepts for artificial Intelligence education in K-12. Computers and Education Open. https://doi.org/10.1016/j.caeo.2024.100196

Podworny, S., Biehler, R., & Fleischer, Y. (2025). Young students' engagement with data to create decision trees. ZDM – Mathematics Education, 57(1). https://doi.org/10.1007/s11858-024-01649-w

Podworny, S., Fleischer, Y., & Hüsing, S. (2022). Grade 6 Students' Perception and Use of Data-Based Decision Trees. In Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics. https://doi.org/10.52041/iase.icots11.T2H3

Podworny, S., Fleischer, Y., Hüsing, S., Biehler, R., Frischemeier, D., Höper, L., & Schulte, C. (2021). Using data cards for teaching data based decision trees in middle school. In 21st Koli Calling International Conference on Computing Education Research (Koli Calling '21), November 18-21, 2021, Joensuu, Finland. ACM. https://doi.org/10.1145/3488042.3489966

Rathgeb, T., & Schmid, T. (2020). JIM-Studie 2020 - Jugend, Information, (Multi-)Media, Basisstudie zum Medienumgang 12- bis 19-jähriger in Deutschland. Medienpädagogischer Forschungsverbund Südwest.

Sanusi, I. T. (2023). Machine Learning Education in the K–12 Context Itä-Suomen yliopisto].

Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching Machine Learning in K–12 Classroom: Pedagogical and Technological Trajectories for Artificial Intelligence Education. IEEE Access, 9, 110558-110572. https://doi.org/10.1109/access.2021.3097962

Zieffler, A., Justice, N., delMas, R., & Huberty, M. D. (2021). The Use of Algorithmic Models to Develop Secondary Teachers' Understanding of the Statistical Modeling Process. Journal of Statistics and Data Science Education, 29(1), 131-147. https://doi.org/10.1080/26939169.2021.1900759

Section 6 references

Moore, D. S., & Cobb, G. W. (2000). Statistics and Mathematics: Tension and Cooperation. *The American Mathematical Monthly*, *107*(7), 615-630. http://www.jstor.org/stable/2589117