Exploring Data Science Across the Curriculum and Across Grade Levels

Session Co-chairs:
Joshua M. Rosenberg, University of Tennessee, Knoxville
Bodong Chen, University of Minnesota
Discussant:
Victor Lee, Stanford University

Abstract
This structured poster session brings together wide-ranging efforts to integrate data science in education across the curriculum and grade levels. Given the ever-growing impact of data and analytics in modern society and in education, it is imperative to engage all learners in the emerging field of data science. The objective of this session is to bring together ideas about what data science is and how data science can be taught in schools and beyond. Contributions explore a range of perspectives on data science, contexts for learning, and pedagogical approaches. These projects expand our views of data science and to make disciplinary learning in STEM learning environments more meaningful and relevant for learners, especially those who may be otherwise disinterested or disengaged.

Objectives
Data science is commonly characterized as a field that integrates disciplinary expertise, statistics, and advanced computational techniques to answer questions or solve problems. Contemporary perspectives on data science recognizes the necessity of going beyond understanding and applying specific statistical methodologies. Data science requires, among other things, collaboration, data processing, visualization, and communication of processes and results (National Academies of Sciences, Engineering, and Medicine, 2018; Wickham, Bryan, & Lazar, 2018).

The limited attention given to the role of data science in education calls for an emphasis on this distinctive field in teaching and learning across the disciplines. This session, then, aims to explore what data science may mean from an educational perspective. It explicitly intends to highlight a diverse set of ideas about what data science is and how it can be taught at the elementary, secondary, and tertiary levels of education and to open up a conversation about:

- What subject areas could data science be situated in?
- How not only researchers but also learners may come to do data science?
- What activities or practices data science might include?
- What challenges and opportunities may emerge from taking up data science?

Overview

The proposed structured poster session will meet the objectives described above through seven poster presentations on a variety of topics relevant to data science in K-12 and tertiary education. These studies use a variety of methodological approaches, from design-based research to the assessment of narrative patterns:

1. Bodong Chen and colleagues on elementary students' 'data expeditions' using real-world, open data
2. Joshua Rosenberg and colleagues on elementary students' art-inspired data visualizations
3. Kayla DesPortes and colleagues on co-designed data literacy-focused professional development for middle school math and art teachers
4. Michelle Wilkerson and colleagues on the middle school science students' 'data claims' as steps toward causal reasoning from data
5. Shiyan Jiang and Jennifer Kahn on middle school students' storytelling about mobility and migration with data visualizations
6. Seth Jones and colleagues on taking a data modeling approach to support statistical reasoning in undergraduate biology contexts
7. Daniel Anderson on teaching data science and computational social science to graduate-level educational researchers

After the poster presentations, Victor Lee will lead a discussion on the objectives of the session and future directions for data science in education.

Significance

While the field of data science has seen growth in industry, less attention has been made to what data science is or can be in education. Contributions explore a range of perspectives on data science, contexts for learning, and pedagogical approaches. These projects expand our views of data science and to make disciplinary learning in STEM learning environments more meaningful and relevant for learners, especially those who may be otherwise disinterested or disengaged.

Structure

1. Rationale and introduction by co-chairs (5 minutes)
2. Brief poster introductions by lead authors (15 minutes)
3. Concurrent poster interactions (40 minutes)
4. Remarks from the discussant (10 minutes)
5. Plenary discussion and Q&A (20 minutes)

Presentation Summaries

Presentation #1: Data Expedition in a Knowledge Building Community

Bodong Chen, Valerie Barbaro, University of Minnesota
Leanne Ma, OISE/University of Toronto
Ben Peebles, Jackman Institute of Child Study Lab School, University of Toronto

The ongoing "data revolution" is transforming knowledge production, governance, and civic engagement. This paper addresses how data science can be integrated into social studies at the elementary level through a pedagogical model named *Data Expedition*. Data Expedition integrates student inquiry with open data and is distinguished from existing initiatives by its foci on (a) harnessing real-world open data for student usage, (b) integrating data science in the school curriculum through computationally rich inquiry, and (c) scaffolding students' self-expression and computational participation within and across knowledge communities.

In our pilot study, 22 sixth-grade students were supported to conduct "data expeditions" in a knowledge-building community (Scardamalia & Bereiter, 2014) using the Common Online Data Analysis Platform (CODAP) and Knowledge Forum (KF). Throughout the semester, the teacher incorporated data expeditions in social studies. Students worked in flexible groups to examine world issues such as gender equality and climate change. Our research questions were: (a) To what extent were students able to analyze open data using CODAP? (b) How did data and analysis facilitate students' problem finding, theory building, and peer interaction?

Our data sources included 21 student CODAP notebooks and students' written discussion in KF. Content analysis was applied on these data. First, we coded all CODAP notebooks in terms of (a) *graph type* (e.g., scatterplot) (Angra & Gardner, 2016); (b) *structural complexity* (Friel, Curcio, & Bright, 2001); and (c) *level of graph comprehension*—"Level 1–Reading the data," "Level 2–Reading between the data," "Level 3–Reading beyond the data" (Curcio, 1987; Friel et al., 2001). Second, we examined KF notes related to CODAP notebooks with a focus on the question–theory dynamics in the data cycle (Gould et al., 2016).

CODAP analysis. Students predominantly relied on scatterplots, the software's default graph choice. Students demonstrated a command of structural components such as regression lines and variable legends. Content analysis of student interpretations showed their capability in reading between and even beyond the data (M=2.05, SD=0.79; see *Figure 1*). These graphs also triggered various conceptual and emotional expressions, including explanations, questions, surprises, and discontent (see *Figure 2*).

KF analysis. Content analysis of student notes revealed that students primarily focused on sharing perspectives and seeking resources to raise awareness about a world issue, using data to support findings and to create a sense of urgency. For example, in one data cycle, a student used CODAP to explore a dataset on which countries pollute the most and was surprised to find that North America had the highest CO2 emission per capita (see Figure 3). This unexpected finding prompted some students to start a second data cycle on why people have not taken action on climate change, while others felt compelled to share this result on social media to inform the public.

This study demonstrates that with proper support, 12-year-olds are capable of engaging in productive data analysis, graph comprehension, and collaborative discourse around data. CODAP as a user-friendly data science tool mitigates technical challenges while collaboration tools like KF can further aid student expression, storytelling, and knowledge building.

Presentation #2: Art as a Context for Data Science: Exploring Fourth-Grade Students' Data Visualization Practices

Joshua Rosenberg, Lynn Hodge, Joy Bertling, and Shande King
University of Tennessee, Knoxville

Socializing, working, and even teaching and learning are increasingly impacted by data. For students, work with data can serve as a cross-curricular practice (Lee & Wilkerson, 2018) that is powerful in terms of students to reason about the world (Lehrer & Schauble, 2015). While research has defined approaches to support student work with data (Lehrer & Romberg, 1996), less attention has been paid to how activities such as recording and analyzing data can build upon the important assets and interests of students and their communities (Civil, 2007), especially at the elementary level.

In our study, we sought to support students to both create and analyze data and communicating findings through data visualizations that they create. The following research questions guided our work:

1. Within an art and statistics integrated pedagogical approach, what mathematics and statistical concepts do participants learn?
2. What aesthetic features do the data visualizations students create depict?

To answer these questions, we designed an art and statistics integrated pedagogical approach and explored its impacts with 14 10-11-year-old students in an elementary school in the Southeast United States. The project included eight sessions that took place from January – April of 2019. Early in the project, students created, together, a data representation that reflected their individual and collective interests. Students then used a survey to understand the interests of other 4th-grade students and analyzed the data from the survey to create data visualizations using visual methodologies (Grodoski, 2018) individually or in pairs (see *Figure 4* for an example of student visualizations). Work samples from students' data visualizations and reflective notes following each class session were examined to identify relevant themes across students (Saldaña, 2015). Students' data visualizations were analyzed through the use of a framework, the *grammar of graphics*, for the components of effective data visualizations (Wilkinson, 2005).

Findings reveal that students are able to explain the rationale behind their visualizations, and most of these explanations are grounded in their data or data summaries. Students' visualizations reflected important mathematical ideas of additive and multiplicative thinking. Students' visualizations regularly depicted data through a number of components of effective visualizations to represent amount and proportions through, particularly area, shape, colors, labels, and scale. Relationships between amounts as well as distributions of or variability in quantities were less-frequently represented.

Our findings suggest that students can consider the design and aesthetic properties of their data visualizations without the mathematical and statistical ideas becoming ignored, a key challenge for data analysts and data scientists at all levels (Wilke, 2019). The use of the *grammar of graphics* to highlight what appeared (and did not appear) in students' data visualizations may serve as a helpful analytic tool for scholars researching data visualization in other contexts, particularly as this frame has been used to develop a popular statistical graphics library for creating data visualizations (ggplot2; Wickham, 2016).

Presentation #3: Co-designing to Support Middle School Data Literacy: Partnerships Among Researchers, Math and Art Teachers

Kayla DesPortes, kd90@nyu.edu, New York University
Camillia Matuk, cmatuk@nyu.edu, New York University
Megan Silander, msilander@edc.org, Education Development Center
Ralph Vacca, rvacca2@fordham.edu, Fordham University

A degree of data literacy can support decisions in diverse areas, from personal relationships to law enforcement (Provost & Fawcett, 2013); and prepare successful workers and citizens (Data-Pop Alliance, 2015; Zuiderwijk et al., 2012). In developing data literacy, learners benefit from integrating diverse perspectives. We explore how to support researchers and middle school teachers in integrating art and mathematics to create data literacy learning experiences.

Typical approaches to promoting data literacy fail to engage meaningful reasoning with data when they focus too narrowly on mathematics concepts with little relation to students' own experiences (Franklin et al., 2015). Recent initiatives incorporate different disciplinary perspectives including social justice, and physical activity, to increase the personal and societal relevance of data literacy (Deahl, 2014; Lee, Drake & Thayne, 2016; Wilkerson & Laina, 2018; Williams, 2015). Art offers an underexplored and approachable medium for working with data, and a perspective that engages students in critical reflection and expression of personally meaningful issues. There is little research on the potential of an art-based perspective on data science, and on how to equip teachers to develop such interdisciplinary pedagogical approaches to cultivate students' data literacy. Whereas co-design can ensure that diverse stakeholders' needs are addressed, and lead to more effective and sustained designs (Borko, 2004; DiSalvo et al., 2017); partners can lack sufficient design skills, and find it challenging to align different disciplinary perspectives toward a common goal (Dever & Lash, 2013; Hardré, Ge & Thomas, 2006).

(a) How do we support effective co-design of data literacy units among art teachers, mathematics teachers, and researchers? (b) What supports do teachers anticipate needing to successfully integrate art and mathematics approaches to build middle school students' data literacy?

Our team of researchers designed and led a 3-day professional development (PD) experience with 6 art and mathematics teachers from three different middle schools. The PD introduced data literacy concepts, activities, and technologies; and sought to elicit and integrate participants' ideas in the co-creation of classroom data science units aligned with both mathematics and art standards. Thematic analyses of focus group interviews with teacher participants, co-designed artifacts, and co-design observations, characterize co-designers' needs and experiences.

Findings identify specific supports that teachers and researchers need to engage in co-design of data-literacy curricular experiences, including supports for identifying authentic data, using technology to enable interactions with those data, and creating activities that are relevant to classroom contexts. We also identify specific disciplinary and pedagogical knowledge related to argumentation and visual representations that teachers drew upon during their co-design and the effectiveness of our approaches for drawing these out and building upon them. Finally, we describe the opportunities and challenges that teachers viewed, and their desire for specific curricular and technological supports in their instruction with art and math integrated data literacy materials.

Findings document the process of creating art-integrated data literacy curricula to broaden participation in data science among diverse youth. They also inform approaches to other co-design efforts among interdisciplinary educators and researchers.

Presentation #4: Making "Data Claims" as an (Inter)Disciplinary Practice in the Science Classroom

Michelle Wilkerson, Lisette Lopez, and Julio Jamarillo
University of California at Berkeley

We have been studying how to integrate analyses of large public datasets into science curricula. There are growing calls for such integration (NSF, 2018); but our primary motivation is to disrupt the assumed 'natural' and 'neutral' status of data in today's world (Gillborn, Warmington, & Deback, 2017). Thus we seek to position second-hand data as a *transformable* and *critiqueable* source of scientific evidence (McNeill & Berland, 2017). This involves not only asking students to analyze and argue from provided data but also to assess the utility and trustworthiness of data in relation to investigative needs (Authors, 2018).

Our work draws from theory about how disciplinary norms and practices emerge in classrooms (Lehrer, 2009), and how interdisciplinarity and sociocultural factors influence what is valued (Bing & Redish, 2009; Bang, et al. 2012). A lack of clarity around interdisciplinary expectations can create pedagogical confusion: Scientific norms emphasize causal and mechanistic relationships (Russ, et al. 2008), while statistical norms emphasize quantitative, low-inference reasoning (Makar & Rubin, 2009). Assessing the validity of data can bridge this gap, but requires disciplinary flexibility. Thus we ask: *How do students navigate the interdisciplinary expectations of data-intensive science investigations?*

We conducted a task analysis of science units enacted with multiple teachers in a large city in California. Each unit lasted 1-2 weeks and was driven by a causal question. This presentation will report on five seventh-grade classrooms in a suburban school with predominantly Latinx students who explored the question, "Did wolves change the rivers of Yellowstone?" Students used publicly available datasets, a visual analysis tool (CODAP; Finzer & Damelin, 2014), and prior content knowledge (predator-prey dynamics; energy webs) to address the question. Two analysts content logged screencast recordings of 36 student groups during the unit.

Both analysts independently identified *data claims* as one way students manage the interdisciplinary expectations of data-driven science investigations. These are "low inference" (Kerlin, McDonald, & Kelly, 2010) claims about what conclusions can be drawn directly from a dataset. Most groups struggled to make causal claims but frequently made valid data claims that reflected productive paths for investigation.

Instructors overlooked or dismissed these claims, which are more granular than the causal claims most valued in scientific argument, we suspect for both disciplinary and raciolinguistic reasons.

One example of such a dismissal is: "Your claim isn't what the graph says, that's evidence. I don't want to see any numbers in this claim box." [P1T1D4]. These dismissals represented missed opportunities to support student work, and interruption to students' progress. When data claims were supported by instructors, students tended to make progress and later link their data claims to causal inference: "There was 12 [beaver colonies] by 2009 and they might be making dams." [P3T5D4].

Data claims represent an intermediate step between the complex interdisciplinary expectations of analyzing data and making inferential claims. In this way, offer connective tissue between "statistical/data science" and "science" practices. They should be better understood and explicitly supported through instruction and curricular design.

Presentation #5: Storytelling with Data Visualizations: Narrative Patterns in Modeling Family Migration Narratives

Shiyan Jiang, North Carolina State University
Jennifer Kahn, University of Miami

Storytelling with data visualizations is a skill that is increasingly important for the next generation to be successful in a world of data and data-driven decision making (Pfannkuch et al., 2010; Wilkerson & Laina, 2018). However, the interdisciplinary practices that are needed for storytelling with data are not generally taught, as youth tend to learn storytelling with words and mathematical sensemaking with numbers (Knaflic, 2015), that is, as distinct disciplinary practices. With the goal of preparing youth to become better data storytellers, this study investigates youth's data narrative patterns and discusses the instructional support needed for youth to construct compelling data stories.

We collected 14 data stories created by 17 middle and high school youth, exploring the reasons for personal family mobility (*What moved my family?*) as well as national and global migration (*What moves families?*) in a free summer workshop at a city public library.

We adopted Bach et al. (2018)'s framework to assess narrative patterns. The results of our analysis illustrate the various ways in which youth tell family migration stories with data visualizations. The narrative patterns, in decreasing order of frequency, are:

- *Juxtaposition*. Participants set comparisons to show that the destination (where the family moved to) was better (e.g., equal job opportunities) than the origin (where the family moved from).

- *Silent data*. Participants emphasize or deemphasize data to draw the audiences' attention, such as adding tooltips for one location.
- *Multidimensional evidence*. Participants choose multiple indicators to strengthen the narrative, such as selecting three indicators to highlight why their family moved: high salary, large communities for African American, and large population attending colleges.
- *Gradual visual reveal*. Participants present multiple data visualizations to reveal narratives in a logical way, such as using one map to show people moving out of South Carolina during the great migration and another map to show people moving to the North in the same period of time.
- *Convention breaking*. Participants first establish a graphical convention and then disrupt it, such as presenting a model to show that the mean year in school was three in South Sudan and using texts to explain one's family member attended school every year.
- *Reflection*. Participants highlight key takeaways from the narrative. For instance, after showing that a family member migrated for better educational opportunities with data, a participant expressed appreciation of their own educational opportunities.
- *Familiar setting*. Participants illustrate ideas with different media to help the audience understand the meaning of data, such as using a picture of a crime scene to support the visualization of high crime rates in South Florida.

In addition to revealing the various ways youth use data in telling stories, our analysis highlights the following instructional supports that allow youth to construct compelling data stories: (a) identifying authentic audiences; (b) understanding data trends that are contradictory to the narrative; (c) connecting indicators in a meaningful and logical way; (d) focusing on one argument; (e) making fair comparisons; and (f) evaluating the consistency between different modes (e.g., texts and data visualization).

Presentation #6: Engaging Students with Uncertainty Through Repeated Measure of Biological Qualities

Ryan Seth Jones, Anna Grinath, Zhigang Jia, and Angela Google
Middle Tennessee State University

Scientific practice creates models to represent and study the world, and these models are how scientists specify the "conditions of seeing" for others critiquing scientific claims (Bazerman, 1998; Lehrer & Shauble, 2010). These conditions often take the form of a *data model*, where sampling and measurement procedures convert qualities of the natural world into quantities for analysis. Science education, then, should not only support students to develop understandings of *what* scientists know, but also the data modeling *practices* scientists deploy to generate, critique and revise knowledge (AAAS, 2011; NRC, 2012; Duschl, Schweingruber, & Shouse, 2007).

In this presentation we describe our efforts to design learning environments that engage undergraduate students with opportunities to construct variable data within the context of measuring sea sponge abundance. Our approach to instruction was motivated by a body of research in statistics education that has shown the importance of a deep understanding of variability and distribution in reasoning with data (Franklin et al., 2007; Konold & Pollatsek, 2002), and of engaging students in the construction of variable data as they model the natural world (e.g. Lehrer & Romberg, 1996; Lehrer, Kim, & Schauble, 2007). Students in these classes independently measured the volume of identical 3D printed sponge fragments from a digital scan of a sea sponge specimen. Students then aggregated their independent measures to consider the distributional shape formed from their measurements and discussed the aspects of the distribution that informed our guiding question about sponge abundance. Students also created mathematical procedures to quantify the distributional characteristics they described in class.

We carried out this project using design-based research methodology (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) in order to create detailed accounts of how students reasoned about measurement, distribution, statistics, and the biology of sea sponges. We video-recorded 4 class meetings in two sections of a non-major biology course at a large public university in the southeastern United States. We also collected written artifacts from 218 students enrolled in these two classes. We open coded the video records and artifacts by cataloging the different ideas students made use of in the tasks and then created categories of similar ideas across the students' comments during class and the artifacts.

We found that students imagined many different measurement strategies to account for sponge abundance, but also came to value shared measurement procedures as critical to biological research. Their measurements of the identical sponge fragments created opportunities for students to come to see the center of the distribution as a signal about true measure, and the variability as an indication of error. Students developed strategies for quantifying distributional characteristics, such as center and variability, that have strong correspondence to disciplinary conventions of mean and standard deviation. In our presentation, we will emphasize the opportunities students' ideas provide for building correspondence between their thinking and disciplinary norms within biology research. This work contributes to a growing body of knowledge about how to support students in undergraduate biology courses to authentically engage in data modeling of biological phenomena.

Presentation #7: A Pedagogical Framework for Developing Computational Social Scientists in Educational Research

Daniel Anderson, University of Oregon

During the 2018-19 school year, a series of three pilot courses were run at the University of Oregon introducing PhD students in educational research programs to data science and computational approaches to educational research. These courses operated in a cohort model, with essentially all students enrolled in the first course taking all three courses. The first course introduced students to the R programming language, git/GitHub, and the principles of reproducible research, with an emphasis on R Markdown. The final project was a group project housed on GitHub, with each member making commits and pulling/pushing to the repository to develop an APA manuscript written entirely in R Markdown, with analysis code weaved with text, making all tables, figures, and in-text references to results dynamic to the code. The second course emphasized the communication of data through a variety of mediums, with a specific emphasis on data visualization, and working with relational data. The final project included a website or online dashboard displaying the fundamental findings of a study for more than one audience, including lay audiences. Finally, the third course focused on functional programming, writing custom functions to extend the functionality of R, and developing interactive web-based data applications. The final project included either an interactive data application deployed on the web or an online tutorial, describing more efficient programming techniques.

This presentation will cover lessons learned from the delivery of these three courses, described above, including areas of successes and areas for improvement. Permission has been obtained from students to share samples of their work, including the progression of both their code written and the products produced across the three courses. Sample text from course evaluations related to the content will also be shared. We share this information within a pedagogical framework for training graduate students on computational methods in social science research, with a specific emphasis on educational researchers. Key takeaways include (a) the need to focus on programming skills separate from statistics or domain-specific courses, including direct, explicit instruction of code; (b) starting with the "low hanging fruit" of skills most directly related to students' applied work, including data visualization; (c) placing greater emphasis on collaborative workflows with git/GitHub; (d) balancing package workflows to ease usability while not overly burdening users or their scripts with excess dependencies; and (e) placing an explicit emphasis on reproducible and open workflows, enabling the verification of not only scientific findings, but also procedures. In sum, we discuss the development of these courses, share students work and evidence of their growth, data from students' perceptions of the courses, and our developed pedagogical framework emerging from these experiences. We view this framework as a starting point for future research on best practices in teaching t to educational researchers.

References

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC. Retrieved from http://visionandchange.org/finalreport

Angra, A., & Gardner, S. M. (2016). Development of a framework for graph choice and construction. *Advances in Physiology Education, 40*(1), 123–128. https://doi.org/10.1152/advan.00152.2015

Bach, B., Stefaner, M., Boy, J., Drucker, S., Bartram, L., Wood, J., ... & Tversky, B. (2018). Narrative design patterns for data-driven storytelling. In *Data-driven storytelling* (pp. 125-152). AK Peters/CRC Press.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press

Bing, T. J., & Redish, E. F. (2009). Analyzing problem solving using math in physics: Epistemological framing via warrants. *Physical Review Special Topics-Physics Education Research, 5*(2), 020108.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher, 32*(1), 9-13.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393. https://doi.org/10.2307/749086

Data-Pop Alliance. (2015). *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*. White Paper Series.

Deahl, E. (2014). Better the data you know: Developing youth data literacy in schools and informal learning environments. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2445621

Dever, R., & Lash, M. J. (2013). Using common planning time to foster professional learning: Researchers examine how a team of middle school teachers use common planning time to cultivate professional learning opportunities. *Middle School Journal*, *45*(1), 12–17. https://doi.org/10.1080/00940771.2013.11461877

DiSalvo, B., Yip, J., Bonsignore, E., & DiSalvo, C. (Eds.). (2017). *Participatory design for learning: Perspectives from practice and research*. Taylor & Francis.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press.

Finzer, W. (2014). *Common Online Data Analysis Platform (CODAP)* [Computer Software]. Concord, MA: The Concord Consortium. https://codap.concord.org/

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *The statistical education of teachers*. Washington, DC: American Statistical Association. Retrieved from http://www.amstat.org/education/SET/SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Alexandria: American Statistical Association.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 124–158.

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race Ethnicity and Education*, *21*(2), 158–179. https://doi.org/10.1080/13613324.2017.1377417

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., … Zanontian, L. (2016). *Teaching data science to secondary students: The mobilize introduction to data science curriculum*. IASE 2016 Roundtable Conference, 1–12.

Grodoski, C. (2018). Visualizing data in art education: Research, practice, and advocacy. *Art Education, 71*(2), 36-45. https://doi.org/10.1080/00043125.2018.1414537

Hardré, P. L., Ge, X., & Thomas, M. K. (2006). An investigation of development toward instructional design expertise. *Performance Improvement Quarterly, 19*(4), 63-90.

Kerlin, S. C., McDonald, S. P., & Kelly, G. J. (2010). Complexity of secondary scientific data sources and students' argumentative discourse. *International Journal of Science Education, 32*(9), 1207–1225. https://doi.org/10.1080/09500690902995632

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Konold, C., & Pollatsek, A. (2002). Data Analysis as the Search for Signals in Noisy Processes. *Journal for Research in Mathematics Education*, *33*(4), 259–289. https://doi.org/10.2307/749741

Ladson-Billings, G. (2017). "Makes me wanna holler": Refuting the "culture of poverty" discourse in urban schooling. *The Annals of the American Academy of Political and Social Science, 673*(1), 80-90.

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects*. Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

Lee, V. R., Drake, J. R., & Thayne, J. L. (2016). Appropriating quantified self technologies to support elementary statistical teaching and learning. *IEEE Transactions on Learning Technologies, 9*(4), 354-365.

Lehrer, R. (2009). Designing to develop disciplinary dispositions: Modeling natural systems. *American Psychologist, 64*(8), 759-771.

Lehrer, R. & Schauble, L. (2015). Developing scientific thinking. In L. S. Liben & U. Müller (Eds.), *Cognitive processes. Handbook of child psychology and developmental science* (Vol. 2, 7th ed., pp. 671-174). Hoboken, NJ: Wiley.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69-108.

Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model?. In *Instructional explanations in the disciplines* (pp. 9-22). Springer, Boston, MA.

Lehrer, R., Kim, M. J., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195-216.

Makar, K., & Rubin, A. (2009). A Framework for Thinking about Informal Statistical Inference. *Statistics Education Research Journal, 8*(1), 82–105.

McNeill, K. L., & Berland, L. (2017). What is (or should be) scientific evidence use in k‐12 classrooms?. *Journal of Research in Science Teaching, 54*(5), 672-689.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

National Science Foundation (2018). *Harnessing the Data Revolution @ NSF*. Accessed from https://www.nsf.gov/cise/harnessingdata/.

Pfannkuch, M., Regan, M., Wild, C., & Horton, N. J. (2010). Telling Data Stories: Essential Dialogues for Comparative Reasoning. *Journal of Statistics Education*, *18*(1), 9. https://doi.org/10.1080/10691898.2010.11889479

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51-59.

Researchers examine how a team of middle school teachers use common planning time to cultivate professional learning opportunities. *Middle School Journal, 45*(1), 12-17.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92*(3), 499-525.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. New York, NY: Sage.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 397–417). New York, NY, USA: Cambridge University Press.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.

Wilke, C. O. (2019). *Fundamentals of data visualization: A primer on making informative and compelling figures*. O'Reilly Media.

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.

Williams, S. (2015). More than data: working with big data for civics. *ISJLP, 11*, 181.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of e-Government, 10*(2).

Figures

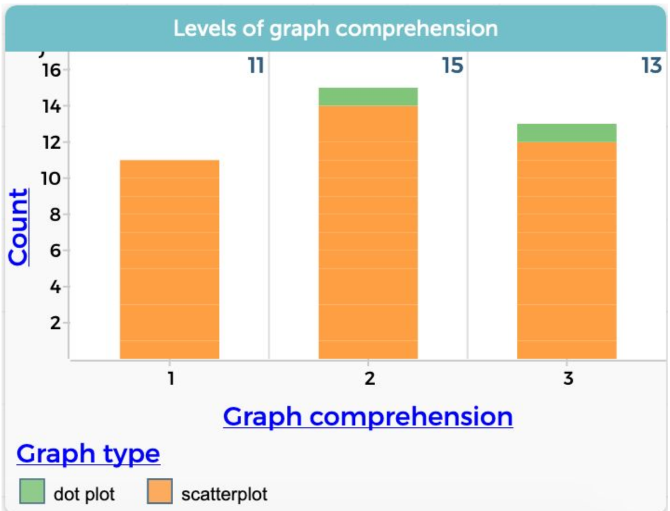

Figure 1. Distribution of the level of graph comprehension by graph type.

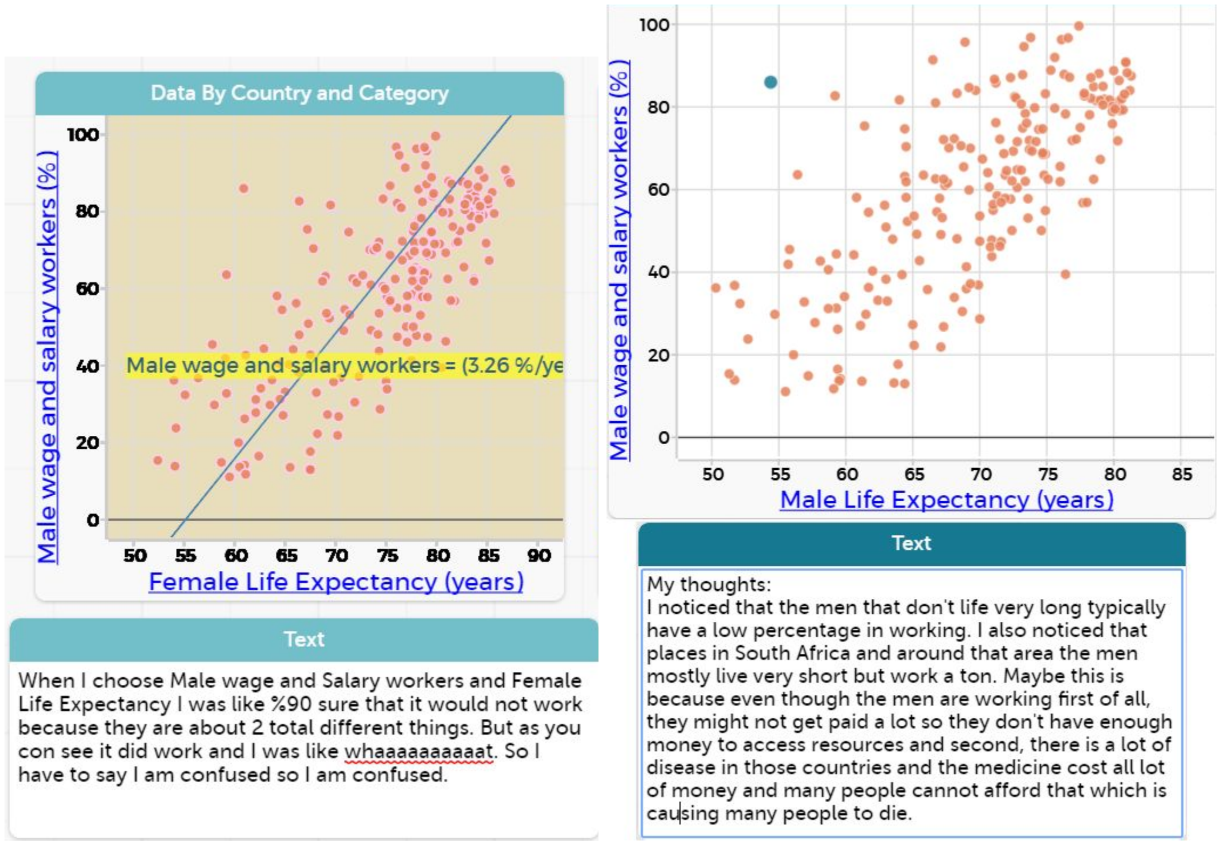

Figure 2. Examples of student CODAP notebooks. The first example is structurally more complex but demonstrates less sophisticated graph comprehension (*Level 1*), whereas the second example is structurally simpler but shows sophisticated reasoning with the graph (*Level 3*).

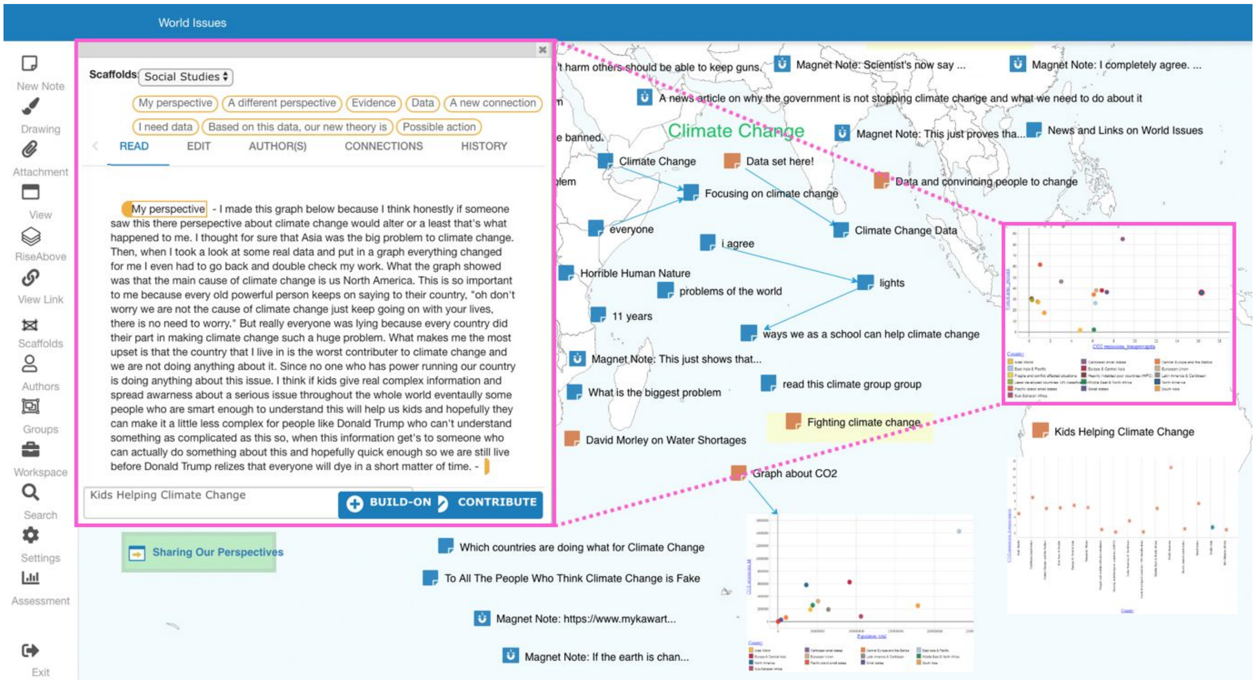

Figure 3. Screenshots of KF discussion about climate change datasets.

Figure 4. *Examples of art-inspired, student-created visualizations of survey data that students collected.*