



## Data and Social Worlds: How Data Science Education Supports Civic Participation and Social Discourse

Katherine M. Miller (Co-Chair), Concord Consortium, [kmiller@concord.org](mailto:kmiller@concord.org)  
Joseph L. Polman (Co-Chair), University of Colorado Boulder, [joseph.polman@colorado.edu](mailto:joseph.polman@colorado.edu)

Susan A. Yoon, Jooeun Shim, Vivian Y. Leung, Yen Nguyen  
[yoonsa@upenn.edu](mailto:yoonsa@upenn.edu), [jshim@upenn.edu](mailto:jshim@upenn.edu), [vyleung@upenn.edu](mailto:vyleung@upenn.edu), [nguyeny@upenn.edu](mailto:nguyeny@upenn.edu)  
University of Pennsylvania

Andee Rubin, Traci Higgins, Jessica M. Karch, Jim Hammerman  
[andee\\_rubin@terc.edu](mailto:andee_rubin@terc.edu), [traci\\_higgins@terc.edu](mailto:traci_higgins@terc.edu), [jessica\\_karch@terc.edu](mailto:jessica_karch@terc.edu), [jim\\_hammerman@terc.edu](mailto:jim_hammerman@terc.edu)  
TERC

Camillia Matuk, Kayla Desportes, Anna Amato, Suzanne Dikker, Xavier Ochoa, Esteban Romero  
[cmatuk@nyu.edu](mailto:cmatuk@nyu.edu), [kd90@nyu.edu](mailto:kd90@nyu.edu), [ada437@nyu.edu](mailto:ada437@nyu.edu), [suzanne.dikker@nyu.edu](mailto:suzanne.dikker@nyu.edu), [xavier.ochoa@nyu.edu](mailto:xavier.ochoa@nyu.edu),  
[a01639031@tec.mx](mailto:a01639031@tec.mx)  
New York University

Suzanne Podworny, Paderborn University, [podworny@math.upb.de](mailto:podworny@math.upb.de)  
Yannik Fleischer, Paderborn University, [yanflei@math.upb.de](mailto:yanflei@math.upb.de)  
Rolf Biehler, Paderborn University, [biehler@math.upb.de](mailto:biehler@math.upb.de)  
Justice T. Walker, University of Texas at El Paso, [jtwalker@utep.edu](mailto:jtwalker@utep.edu)  
Amanda Barany, University of Pennsylvania, [amanda.barany@gmail.com](mailto:amanda.barany@gmail.com)  
Alex Acquah, University of Texas at El Paso, [aacquah@miners.utep.edu](mailto:aacquah@miners.utep.edu)  
Andi Scarola, University of Texas at El Paso, [adscarola@miners.utep.edu](mailto:adscarola@miners.utep.edu)  
Sayed Mohsin Reza, Pennsylvania State University, Harrisburg, [skr6024@psu.edu](mailto:skr6024@psu.edu)  
Trang C. Tran, University of Colorado Boulder, [trang.tran@colorado.edu](mailto:trang.tran@colorado.edu)  
Ralph Vacca, Fordham University, [rvacca2@fordham.edu](mailto:rvacca2@fordham.edu)  
Megan Silander, Education Development Center, [msilander@edc.org](mailto:msilander@edc.org)  
Peter Woods, University of Nottingham, [peter.woods@nottingham.ac.uk](mailto:peter.woods@nottingham.ac.uk)  
Cassia Fernandez, University of São Paulo/Columbia University, [cassia.ofernandez@gmail.com](mailto:cassia.ofernandez@gmail.com)  
Adelmo Eloy, University of São Paulo/Columbia University, [adelmo@fablearn.net](mailto:adelmo@fablearn.net)  
Josh Radinsky, University of Illinois Chicago, [joshuar@uic.edu](mailto:joshuar@uic.edu)  
Iris Tabak, Ben Gurion University of the Negev, [itabak@bgu.ac.il](mailto:itabak@bgu.ac.il)  
Victor R. Lee, Stanford University, [vrlee@stanford.edu](mailto:vrlee@stanford.edu)  
Dora Demszky, Stanford University, [ddemszky@stanford.edu](mailto:ddemszky@stanford.edu),  
Sarah Levine, Stanford University, [srlevine@stanford.edu](mailto:srlevine@stanford.edu),  
Josephine Louie (Discussant), Education Development Center, [jlouie@edc.org](mailto:jlouie@edc.org)

**Abstract:** As the world becomes increasingly awash in data, the ability to engage in civic participation and social discourse is becoming more dependent on the ability to engage with large real-world data sets. Data science education, as a growing focus across disciplines and age groups, strives to prepare learners to be active citizens by supporting them in engaging in inquiry with data that intersects with social and civic phenomena. The eleven projects represented in this symposium explore how learners across age groups and geographic locations are supported for civic engagement and the examination of social worlds through data based inquiry. The projects take place in an eclectic and diverse range of settings and disciplines, tied together through the convergence of data science education and civic and social engagement.

**Keywords:** Data Science Education, Social Discourse, Civic Participation, Agency

### Symposium overview



Engaging in civic participation and social discourse is becoming increasingly dependent on the ability to engage with large real-world data sets (e.g., Pangrazio & Selwyn, 2021). Data science education (DSE), as a growing focus across disciplines and age groups, strives to prepare learners to be active citizens by supporting them in engaging in inquiry with data that intersects with social and civic phenomena (Wise, 2020). While many efforts to engage students in inquiry around complex social concepts have been explored in the classroom (e.g., Hodson 2020) connecting inquiry to authentic action has been shown to be difficult in traditional classroom settings (e.g., Reichert & Print, 2018). In order to build civic participation, students need to develop their identities and agency as active participants, and see themselves as capable of engaging in authentic, meaningful action (e.g., Rap et al., 2022).

The National Academy of Education's 2021 report, *Educating for Civic Reasoning and Discourse* (C. D. Lee et al., 2021) concludes that civic engagement includes thinking through “a public issue using rigorous inquiry skills and methods to weigh different points of view and examine available evidence” (p. 1) as well as discussing the challenges of those issues as a way to grow collective and individual understanding. The report challenges educators to adopt a more ambitious stance toward teaching and learning that embraces this goal, with civic reasoning extending across all subjects, and inseparably connected to engaging in discourse with others. Hodson (2020) makes suggestions for how to support students in developing a civic identity and building towards authentic and public sociopolitical action. Hodson recognizes that students need to be explicitly taught how to take action in the world outlining three phases of “learning to act” which he labels learning *about*, learning *through* and learning *from* action. He posits that engaging in action is only one piece of the puzzle, supported by learning about what actions are available, how they can be pursued, and how previous actions and movements have met with success or failure. Translating knowledge into sociopolitical action is difficult and requires ownership of the issue and a sense of one’s own agency and potential for achieving impact (Rap et al., 2022; Vakil et al., 2018). In today’s world, such reasoning, discourse, and engaging in action inevitably involve engagement with data in some of its many forms.

DSE plays an important role in supporting civic participation by providing context and an environment for students to develop their voices and competencies for engaging in legitimate authentic action (Vakil et al., 2018). DSE’s grounding in authentic contexts and focus on critical thinking and engagement with the humanistic nature and influence of data collection and use provides a background for student agency and ownership over the inquiry process and the steps towards action (V. Lee et al., 2021). In this session, we aim to move beyond supporting learners in developing fluency with data practices, to examining the kind of social and civic contexts in which data and inquiry are embedded and how DSE can support students in building identity and agency towards civic participation within these contexts. The projects within this symposium represent a deliberate range of settings and disciplines in which students grapple with social issues through engagement with real-world data. Each project engages students with not only reading but also writing the world with data. This symposium will consist of three sections. First, the chairs, Katherine Miller and Joseph Polman, will introduce the topic, then the participants will present their research through two consecutive rounds of poster discussions. Finally, the discussant, Josephine Louie, will synthesize and reflect on findings, followed by a group discussion with the audience and presenters.

## 1. Telling data stories in middle school social studies classrooms

Katherine M. Miller, Joseph L. Polman, & Trang C. Tran

Understanding how data is used by those around us, as well as learning to interpret and use it in turn is a powerful tool, both for surviving in the fast-paced world of technology, and for thriving and wielding authority to help enact change in our communities through scientific action and civic participation (Hodson, 2020; Wise, 2020). Thus, fostering students’ development of data acumen and agency is increasingly important (V. Lee et al., 2021) and data storytelling within the context of interdisciplinary project-based learning (PBL) is a promising approach (Radinsky, 2020; Rubin & Mokros, 2018).

This project examines two 8th grade humanities teachers as they co-design and implement data enhanced project based learning modules within their social studies curriculum. Both teachers were in urban school districts teaching historically under-resourced minority students. The teachers both taught a unit focused on Japanese American internment and its relationship to internment, representation and U.S. citizenship today. We found that students achieved all four modes Radinsky (2020) identified for narrating with and around data: “(1) telling a story about oneself working with data; (2) animating a data representation; (3) incorporating data into extant narratives; and (4) narrating oneself into a data-represented world” (p. 374). Furthermore, we found that through engaging with data storytelling, students developed both data identity and increased agency for civic engagement. In post-implementation interviews, students were asked how their perception of data changed over the course of the PBL module. One student highlighted mode 3, saying, “I learned about how data can really affect our lives and show things that affect our lives. I learned that data can tell a story.” Additionally, in a post-implementation retrospective



likert-scale survey, there was significant growth in measurements of students' data identity and agency. This work reinforces interdisciplinary project-based learning as preparing learners to utilize data in meaningful and personally relevant sensemaking and to connect to extant narratives as well as supporting the use of data storytelling as a way to build students' identity and agency on the path towards increased civic engagement.

## 2. “OMG that’s crazy!” Examining the role of high school biology students’ epistemic emotions in data investigations of neighborhood air quality

Susan A. Yoon, Jooeun Shim, Vivian Y Leung, & Yen Nguyen

Arguably the need to address and arrest the debilitating impact of socioscientific issues on society such as the spread of disease, air pollution, and climate change has never been more important (Reiss, 2020). Consequently, researchers have challenged school science curricula to level up in raising awareness, emphasizing epistemic practices, and taking real-world action to stem their tide (Chinn et al., 2021; Hodson, 2020). Such investigations however, as Hodson (2020) points out can evoke profound emotions due to stressful implications for planetary well-being. Conversely, Chinn et al. (2021) make the case that promoting caring about such issues can develop interest in coordinating complex information. Similarly, Muis et al. (2021) highlight evidence illustrating positive links between epistemic emotions and critical thinking. Following this line of research, we examine the emergence of students' epistemic emotions while participating in a bioinformatics unit on asthma and air quality. We ask, what epistemic emotions were triggered, to what extent, and how did they influence student engagement?

We define epistemic emotions as surprise, curiosity, confusion, enjoyment, frustration, anxiety, and boredom arising from information-oriented appraisals during knowledge-generating activities (Muis et al., 2021; Pekrun et al., 2017). We analyzed video recordings of students as they collected air quality data (carbon monoxide and particulate matter) in their neighborhood with mobile sensors (76 videos of 84 students across 14 high school biology classrooms). Analysis revealed that curiosity and surprise were the most frequent categories, followed by enjoyment, anxiety, confusion, and frustration, with no evidence of boredom. Curiosity mainly arose while making scientific predictions about air pollution quality, e.g., “I think the air is more polluted because of the traffic, definitely...so I think it might be higher.” This, in turn, often spurred further hypothesis construction and testing. Surprise frequently occurred when students encountered unexpected data, e.g., “OMG, that’s crazy!” or “what? 95%!” Surprise sometimes shifted to anxiety, especially when data suggested environmental risks, e.g., “It feels so dangerous for children; that’s wrong...and kids die!” This prompted further discussions about solutions and actions. Additional findings are detailed in the poster. We also further discuss epistemic emotions’ generative influence on students' use of epistemic practices such as taking additional data points, comparing and contrasting hypotheses, and using data to support student scientific claims.

## 3. Exploring data through the lens of the United Nations Sustainable Development goals: What actions can youth take?

Andee Rubin & Traci Higgins

The Data Explorers program, funded by NetApp as part of its Corporate Social Responsibility program, engages youth around the world in exploring data related to the United Nations Sustainable Development Goals, which include aspirations such as “no poverty,” and “zero hunger.” The program has an explicit goal of inspiring participants to take action related to such goals. It involves middle school-aged students from the United States, India, Ireland and the Netherlands in a 12-session out-of-school experience introducing participants to basics of data science and the potential of data to uncover economic, social and health inequities around the world. In creating the program, we made several design decisions to maximize chances participants would connect patterns they saw in data with potential actions. One strategy was moving from an initial focus on international, country-level data to local data. For the United States, local data comprised county-level data within the state where participants lived; in other countries, local data was at an appropriate governmental level that allowed comparisons across several dozen entities. Our design hypothesis was that youth would be more likely to feel a connection with inequities at a local level versus international level and be better able to imagine actions that could have real impact.

Another design decision was to suggest that participants create final projects in the form of Public Service Announcements (PSAs). To support connection to action, students choose between four courses of action to highlight in their PSA: Inquire (dig deeper into data and uncover more detailed findings); Invest (suggest allocation of resources to an under-resourced community); Inform (spread the word about the issue) and Invent (a creative

solution or technology to alleviate the issue). Our poster will describe these designs in more detail and, using participants' PSAs, reflect on how the designs provide authentic opportunities for youth to connect data and action.

#### **4. Exploring multivariate survey data: Predicting who is playing online games**

Susanne Podworny, Yannik Fleischer, & Rolf Biehler

In the age of information and technology, media has become an integral part of our daily lives. Understanding the dynamics of media usage profiles has become essential in today's digitally driven society. Given the growing societal relevance, there is a rising demand to incorporate data-driven methods into the educational curriculum (Biehler & Fleischer 2021). A data science project at school (Rubin & Mokros 2018) can teach different skills for data science by using student-related data and the right tool. In the Project Data Science and Big Data at School (ProDaBi, [www.prodabi.de/en](http://www.prodabi.de/en)), we offer a multivariate dataset on a survey of over 1200 young participants who completed a comprehensive online survey with over 150 questions concerning leisure time use and media behavior (Podworny et al. 2021) and the CODAP tool to introduce students to data science and machine learning methods.

Through a teaching unit for grades 8-10 (ages 13-16) developed in the ProDaBi-Project, we explore ways to support students in developing data science competencies. The digital tool CODAP, with its intuitive drag-and-drop functionality, provides an easy entry point to explore relationships between variables as well as providing an introduction to the machine learning method of decision trees. For example, students can explore the question of who plays games online by creating a decision tree. Decision trees turn out to be a highly transparent machine learning approach that allows students to understand the construction algorithm (Biehler & Fleischer 2021). Using the CODAP decision tree plugin, students create their own decision trees based on the multivariate data to predict who plays online games, thus engaging with socially relevant data. CODAP supports the intuitive and manual creation of decision trees and the successive use of different visualization and "semi-automatic" strategies that allow to "play the machine". This is used as a didactic concept for the teaching unit, which aims at understanding the decision tree algorithm without the need for programming (Futschek & Moschitz 2010). In the poster, we present how "playing the machine" supports students' understanding of a decision tree algorithm. Examples of decision trees created by students enrich the discussion.

#### **5. Creative data advocacy: Engaging youth in data remixing practices**

Ralph Vacca, Kayla DesPortes, Camilllia Matuk, Megan Silander, Anna Amato, & Peter Woods

Outside school, youth engage in rich storytelling practices through image- and video-based social media platforms, which go beyond consuming, to also producing media. These can be avenues for data advocacy, that is, offering data to support arguments made in civic and political discourse (e.g., Abbas et al., 2022). Yet, few data analysis tools prioritize data storytelling in ways that leverage youth's existing practices. Youth often "remix;" they take existing cultural artifacts and recombine them to create new products, as has been observed in fan fiction and game modding communities (e.g., Vourletsis & Politis, 2022). Remixing has re-emerged as a "new media" literacy practice as novel remixing technologies afford new ways of meaning making. We ask: What remixing affordances can support students' critical data reasoning practices?

Through an ongoing co-design project with six middle and high school Math, ELA, and Social Studies teachers in the United States, we are exploring ways to support disciplinary-based instructional goals related to students forming data-driven arguments for social justice advocacy. Here, we focus on the design of a tool to support students in drawing connections across different but thematically connected datasets. The tool will allow students to remix multiple data sources, while prompting reflection on the relationships among those sources, and on how data's context and analysis contribute to its use in an argument. We tested a prototype of the tool in one of our participating teacher's classrooms, in which students used it to explore ELA and Social Studies topics. Findings from a thematic analysis of interviews and group discussions with teachers as they reflected on student artifacts from this implementation suggest several learning opportunities with regard to the design of this tool. Specifically, in remixing data, students identify, connect and layer data in ways that can facilitate individual meaning making and argumentation. These processes can support critical evaluation of data sources when students build connections between data to their identities, cultures, and lived experiences. Remixing also centers the subjectivity of data sources, analyses and interpretations, as learners must situate their arguments within broader sociopolitical systems. We will illustrate our findings with examples of student artifacts, and teachers' ideas for scaffolds necessary to build students' critical data literacy through similar tools that engage learners in advocacy with data and narrative.

## 6. Leveraging multimodal data from the body to include non-scientists in human brain and behavioral inquiry

Suzanne Dikker, Esteban Romero, Xavier Ochoa & Camillia Matuk

Human data have become a billion-dollar industry, fueling innovations in technologies for tracking brain/body signals (Horst et al., 2022) and permeating diverse realms of society (Kim et al., 2019). These developments have brought challenges pertaining to the validity of the inferences that are drawn from the data, data privacy, and public trust. It is thus critical that learners develop data literacy skills that will prepare them to navigate a world where data from their brains and bodies have become a commodity (Drew, 2023). We describe a series of cross-disciplinary projects that illustrate how non-scientists can become critical agents through science participation (Raffaghelli, 2020) by making brain and body data more accessible to them through Brain/Body Computer Interface (BCI) technologies.

In the art/science installation *The Mutual Wave Machine* (Dikker et al., 2021) museum visitors sit in a dome surrounded by light patterns reflecting the degree to which their brainwaves are ‘in sync.’ Engaging with visualizations of their own brain synchrony prompts audiences to test different hypotheses about being ‘on the same wavelength’ (eye contact, conversation, holding hands). *You: Quantified* is a web application where learners use wearable biosensing devices to explore data from their own brains and bodies. For example, students can use the javascript-based tool P5.js to create visuals they can manipulate with their ‘minds’ by alternating their brain states (e.g., a “sun” that rises or sets as a function of fluctuations in alpha brainwaves, which have been linked to attention and sleepiness). Learners can toggle between these visualizations, raw data traces, and graphical representations of their brain data, providing hands-on exposure to the subjective decisions that underlie analyzing and representing human data. These projects demonstrate how creative ways to engage the public with brain data can promote curiosity, encourage experimentation, and spur new questions that may inform scientific research, build data literacy, and invite critical reflections on the use of data in society (Drew, 2023).

## 7. Revisiting the regionalization of the American continent through constructing data visualizations

Cassia Fernandez, Adelmo Eloy, Roseli Lopes & Paulo Blikstein

Data visualizations can be powerful inscriptions to support learning across a wide range of disciplinary contexts, including social studies (Philip et al., 2016). By inventing their own visualizations, students can also engage with data science activities in more personal ways, strengthening their relationships to data (V. R. Lee et al., 2021). We present initial results from an initiative to integrate computer and data science into 8th-grade social studies in Brazilian public schools by prompting students to program data visualizations. Eight 50-minute lessons were co-designed by researchers and geography and technology teachers with the goal of exploring cultural and socioeconomic similarities and differences among America’s nations. Research data include recordings of classroom implementations, co-design sessions, and teacher interviews.

Our findings underscore the value of collaborative design and how the infusion of data science into social studies opens avenues for students to engage in the unit’s central ideas personally and critically. During co-design, teachers emphasized regionalization, which they defined as “the categorization of geographical spaces based on specific criteria determined by someone”. To explore this, students used data sources that covered several aspects of all American countries (e.g., HDI, favorite sport) to devise personalized groupings. The visualizations were created using PlayData, a block-based programming environment for data visualization. The activities primarily granted students autonomy to generate personal visualizations and establish their own criteria, which increased participation. Teachers described how geography, “often viewed as a ‘boring’, theoretical subject, was made more accessible and engaging to students.” Teachers also highlighted how interaction with up-to-date data differed from students’ interactions with the typically outdated data found in their textbooks. Lastly, constructing visualizations helped students grasp the idea of regionalization from different perspectives, motivating them to transcend its conception as something given by an authority, to discuss alternative possibilities through critical viewpoints while supporting their understanding of the continent’s diversity. In this context, the unit was an initial step towards promoting students’ sociopolitical reasoning, where personal engagement with data fostered students’ critical perspectives.

## 8. Coding like a data miner: Computational data mining as a tool for exercising sociopolitical inquiry in teaching and learning

Justice T. Walker, Amanda Barany, Alex Acquah, Andi Scarola, & Sayed M. Reza



While data science has successfully been taken up in a number of academic disciplines, computer science has shown remarkable promise as a context to support computational literacies (Kafai et al., 2019) and subject area domain knowledge mastery (Rosenberg et al., 2020) situated in contexts congruent with learners, their cultural histories, and everyday life. Such efforts are motivated by growing need to support data science learning that attends to its increased ubiquity and societal impact. One such area is “big data” fields, where datasets can be unwieldy but effectively harnessed using computational techniques. Current pre-college efforts often involve datasets and investigations that are curated by others, rather than learners themselves. Thus, learning experiences, while sometimes valuable, often employ replication, limiting how much learners pursue authentic practice or meaningful lines of inquiry. This undermines educational efforts that seek to foster learners’ knowledge and gain mastery, while supporting enactment of literacies with agency that pursue personal interests and/or engage sociopolitical concerns.

One promising approach in addressing these challenges includes open inquiry, or sandboxing (Walker et al., 2023). In data-science-based computing, this can be characterized as learning environments that create opportunities for learners to curate and harness data for themselves using computation techniques. Coding Like a Data Miner (Barany et al., 2023) instantiates this approach, in which learners access social media platforms using publicly available APIs to access and analyze massive primary datasets and use them to construct knowledge about a variety of academic domains. In this study, we report on an exploratory study set in an urban city in the mid-atlantic region of the United States where we implemented an in-person pilot workshop with high school youth from demographic groups persistently underrepresented in computing education and industry. We then examined the various ways learners leveraged computational data mining to conduct inquiries into personal and often sociopolitical concerns. Our observations suggest sandboxing is a meaningful frame for supporting learning, empowerment, and agency.

## 9. Humanizing data and recruiting historical empathy as sociopolitical action

Josh Radinsky & Iris Tabak

Following Vakil (2018), we conceptualize DSE for sociopolitical action as incorporating attention to learners’ development of civic and political identities, attending to “the social and economic welfare of historically nondominant students and their communities” (Vakil, 2018, p. 37). Education for sociopolitical action involves not only modeling and practicing overtly political acts like protesting or lobbying (e.g., Hodson, 2020), but also cultivating data practices that contribute to, and do not damage, agentic civic and political identities, in the context of the data ecologies within which learners make sense of their social worlds (Philip et al., 2016). This includes developing habits and practices for narrating a represented world with data, in ways that attribute agency to nondominant and oppressed communities; identifying ways that data can erase or obscure those identities; and contesting extant social narratives that diminish those identities (Philip et al, 2016; Radinsky, 2020).

In this poster we illustrate three data practices: *problematising absences*, *humanizing data*, and *recruiting historical empathy*. We illustrate a classroom activity design for pre-service, elementary social studies teachers, which affords opportunities to model, practice, and critique these data practices. Analysis of vignettes from two classrooms shows how preservice teachers engaged in sensemaking about waves of U.S. immigration using mapped historical census data. Learners negotiate absences and irregularities in census data; characterize and compare complex data distributions over time and across regions; and interrogate how their data interpretations relate to extant societal narratives, historical sociopolitical events, and their own lives. We show how students’ negotiations of data narratives about immigration and immigrants created opportunities for ideological contestation (Philip et al, 2016), uptake of alternative framings, and shared attention to processes of “making up” “kinds of people” with data. A design analysis of the learning activity is presented, mapping pedagogical designs to the focal data practices.

## 10. Investigating large, authentic, complex datasets at DataFest: How the goal of communicating actionable insights shapes the process

Traci Higgins, Jessica M. Karch, & James K.L. Hammerman

Our research explores how teams of undergraduate students with multidisciplinary knowledge, different skill sets, and diverse experiences draw on these in an interdisciplinary way as they navigate being awash in large, authentic, complex (LAC) data at DataFest (DF), a co-curricular, hackathon-style competition. DF takes place at over 50 sites, with over 2000 participating students. Each year the American Statistical Association (ASA) seeks industry or public sector sponsors to provide LAC data for students to explore within an open-ended problem space. Students sign up as teams or are placed in teams of up to five students. On Friday evening, datasets are revealed. Teams have

until Sunday midday to extract meaning and insight from the LAC dataset to address open-ended real-world problems (Gould, 2014). On Sunday afternoon, teams present via short presentations and vie for awards determined by a judge panel, such as Best Use of Outside Data and Best Data Visualization. In Spring 2023, teams were tasked with finding insights to improve the effectiveness of the American Bar Association's (ABA) online platform providing *pro bono* answers to low-income users' questions about civil law. The purposefully vague problem space forced teams to define the problem and generate questions. The data sets contained different information types (e.g., demographic, time data) including an especially messy and complex dataset with text from posts to the platform.

A core research question examines how teams leverage interdisciplinary thinking when using DataFest's complex real-world data. An emerging finding from our mixed methods research is that teams leverage a broad range of knowledge assets, both computational and non-computational, to frame their investigation, identify actionable insights, and communicate findings. Moreover, the problem framing and communicating & proposing action phases of data investigation (Lee, H. et al., 2022) shape work progress in many ways. We conclude that DSE needs to attend more deeply to how interdisciplinary assets and reasoning processes support data investigation.

## 11. Data by the book: New lenses on social worlds depicted in middle school texts

Victor R. Lee, Dora Demszky, & Sarah Levine

Texts in US-based English/Language Arts classes invite readers to engage with described worlds and situations that are windows into and mirrors reflecting back our understandings of our social worlds. Through reading, interpreting, and discussing texts, students are offered new perspectives on events and see how text can be structured to achieve communicative or evocative goals. For instance, a unit with a middle school text can both explore self-discovery and also how sentence structure and specific word choices deliberately shape readers' experiences.

Through an ongoing co-design project with middle school English/Language Arts teachers, we are exploring ways to support comprehending texts with relevant social commentary and to understand authorial style as exemplified by text and literary devices, and to do so using data visualizations derived from those texts. For this session, we focused on the design and enactment of an activity exploring how suspense is developed in a short dystopian story, "The Lottery" by Shirley Jackson. The story tells of a village where residents gather to participate in a drawing that has been done for as long as the residents can remember. The winner of this lottery is then immediately stoned to death by the rest of the town, a tragic end for one of the main characters that features throughout the story. This story is a cautionary tale of conformity and tradition, and invites questions about choice and human nature. It is also a story that our partner teacher has used successfully for years as she covers the techniques where an author creates suspense. The class participated in an activity focusing on the visualization of sentence length in the Lottery and the extent to which sentence length was used to generate suspense. We present a summary informed by our video recordings of the discussion that took place across two English classes as they examined and discussed this dystopian text with data visualizations providing an additional semiotic resource and explore what opportunities this may provide for participating in social discourses in the future.

## References

Abbas, L., Fahmy, S. S., Ayad, S., Ibrahim, M., & Ali, A. H. (2022). TikTok intifada: Analyzing social media activism among youth. *Online Media and Global Communication*, 1(2), 287-314.

Barany, A., Reza, S., Johnson, M., Barrera, A., Badreddin, O., Fuentes, C., & Walker, J. T. (2023). Towards the Design of a Culturally Relevant Curriculum for Equitable, Data Mining-Based CS Education. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, pp. 1498-1501.

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43(S1), 133–142.

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2021). Education for a “post-truth” world: New directions for research and practice. *Educational Researcher*, 50(1), 51-60.

Dikker, S., Michalareas, G., Oostrik, M., Serafimaki, A., Kahraman, H. M., Struiksma, M. E., & Poeppel, D. (2021). Crowdsourcing neuroscience: Inter-brain coupling during face-to-face interactions outside the laboratory. *NeuroImage*, 227, 117436.

Drew, L. (2023, November 8). *The rise of brain-reading technology: what you need to know*. Nature Publishing.

Futschek, G. & Moschitz, J. (2010). Developing algorithmic thinking by inventing and playing algorithms. *Proceedings of the 2010 constructionist approaches to creative learning, thinking and education: Lessons for the 21st century (constructionism 2010)*, 1-10.

Gould, R. (2014). Datafest: Celebrating data in the data deluge. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics* (pp. 1-4). International Association for Statistical Education (IASE).

Hodson, D. (2020). Going beyond STS education: Building a curriculum for sociopolitical activism. *Canadian Journal of Science, Mathematics and Technology Education*, 20, 592-622.

Horst, R. T., ter Horst, R., & Dresler, M. (2022). The Quantified Scientist: Citizen Neuroscience and Neurotechnology. *Neuroscience*, 13(1), 63–65.

Kim, J., Campbell, A. S., de Ávila, B. E.-F., & Wang, J. (2019). Wearable biosensors for healthcare monitoring. *Nature Biotechnology*, 37(4), 389–406.

Lee, C. D., White, G., & Dong, D. (Eds.). (2021). *Educating for Civic Reasoning and Discourse*. Committee on Civic Reasoning and Discourse. Washington, DC: National Academy of Education.

Lee, H., Mojica, G., Thrasher, E., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal*, 21(2), 1-23

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.

Muis, K. R., Chevrier, M., Denton, C. A., & Losenno, K. M. (2021). Epistemic emotions and epistemic cognition predict critical thinking about socio-scientific issues. *Frontiers in Education*, 6, 669908.

Pangrazio, L., & Selwyn, N. (2021). Towards a school-based ‘critical data education’. *Pedagogy, Culture & Society*, 29(3), 431-448.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion*, 31(6), 1268-1276.

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.

Podworny, S., Fleischer, Y., Hüsing, S., Biehler, R., Frischmeier, D., Höper, L., & Schulte, C. (2021). Using data cards for teaching data based decision trees in middle school. *21st Koli Calling International Conference on Computing Education Research*, 1–3.

Radinsky, J. (2020). Mobilities of Data Narratives. *Cognition and Instruction*, 38(3), 374-406.

Raffaghelli, J. E. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education. *Education Sciences*, 10(9), 233.

Rap, S., Blonder, R., Sindiani-Bsoul, A., & Rosenfeld, S. (2022). Curriculum development for student agency on sustainability issues: An exploratory study. *Frontiers in Education*, 7, 871102.

Reichert, F., & Print, M. (2018). Civic participation of high school students: The effect of civic learning in school. *Educational Review*, 70(3), 318-341.

Reiss, M. J. (2020). Science education in the light of COVID-19: The contribution of history, philosophy and sociology of science. *Science & Education*, 29(4), 1079-1092.

Rosenberg, J. M., Lawson, M., Anderson, D. J., Rutherford, T., & Jones, R. S. (2020) Making Data Science “Count”: Data Science and Learning, Design, and Technology Research.

Rubin, A., & Mokros, J. (2018). Data clubs for middle school youth: Engaging young people in data science. In *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)*, Kyoto, Japan. Voorburg, The Netherlands: International Statistical Institute.

Vakil, S. (2018). Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review*, 88(1), 26-52.

Vourletsis, I., & Politis, P. (2022). Exploring the effect of remixing stories and games on the development of students' computational thinking. *Computers and Education Open*, 3, 100069.

Walker, J.T., Barany, A., Barrera, A., Reza, S., Guzman Del Rio, K., Acquah, A., Badreddin, O., Johnson, M. (October, 2023). Sandbox Data Science: Culturally Relevant K-12 Computing. FIE. In *2023 IEEE Frontiers in Education Conference (FIE)*. IEEE.

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, 29(1), 165-181.

## Acknowledgements

We acknowledge support of the National Science Foundation through the DRK12 program (2200887, 2009803, 1908482), ITEST (2241751, 2241483), CISE (2137708), IIS (2302657, 2302658, 2302659), IUSE (2216023), of NetApp, Inc. through their Corporate Social Responsibility program, and of the Deutsche Telekom Stiftung.