

Paths through Data: Successes and Future Directions in Supporting Student Reasoning about Environmental Racism

Emily V. Reigh, Stanford University, evreigh@stanford.edu
Meg Escudé, Cherise McBride, Xinyu Wei, Michael Bakal, Edward Rivero, Collette Roberto, Michelle Hoda
Wilkerson, Kris Gutiérrez
{meg.escude; cherisem; weixinyu0820; edward_rivero; roberto_c; michaelbakal; mwilkers;
gutierrez}@berkeley.edu
University of California, Berkeley

Abstract: We report on a curriculum development project in which students explore environmental racism through data. Recognizing that quantitative data alone is insufficient to understand the sociohistorical contexts of racism, we draw from syncretic approaches to learning that put everyday experiences and qualitative evidence into direct conversation with quantitative datasets through storytelling. Through two focal cases, we demonstrate how one student leveraged personal experience to engage in deep integrative analysis of data, while another with fewer perceived personal connections to environmental racism focused more specifically on patterns, with less structural or racial analysis. Implications of the analysis include the need to carefully attend to the use of quantitative data related to race and to scaffold the integration of other sources of information with quantitative data sets.

Introduction

There are increasing calls for students in STEM classrooms to reimagine the use of data (Lee & Wilkerson, 2018), as well as to engage with issues of social and racial justice, like environmental inequities (e.g., Varelas et al., 2017). These two foci can be brought together through student exploration of socioscientific data. Recent studies have shown that race is a stronger indicator of exposure to pollution than income (Tessum et al., 2021). Yet, literature points to challenges in engaging students with issues of racial justice in STEM. Even when engaging with social phenomena, data literacy activities may fail to attend to the structural and ideological contexts in which data is situated and the embedded issues of power (Philip, Olivares-Pasillas & Rocha, 2016). Quantitative data alone is insufficient to understand racism; it must be supplemented with other information about local and historical contexts (Gilborn, Warmington and Demack, 2018).

The *Writing Data Stories* project seeks to address these challenges by engaging youth in expansive forms of learning that integrate analysis of data with students' own experiences and cultural repertoires of practice. We report on a youth workshop exploring environmental racism through data science inquiry in which students shared their own experiences, investigated and manipulated data sets, and engaged in journalistic research. We found that engaging students' personal experiences with environmental racism deepened their treatment of data as well as their analysis of historical and structural factors related to race. In some cases, however, students had more limited personal connections to the topic; in these cases, their investigations focused on correlative relationships among variables and included structural or racial analysis.

Theoretical Framework

Our work is grounded in a syncretic approach to development (Gutiérrez, 2008, 2014) that leverages the expansive forms of learning that occur when youth integrate everyday and school-based contexts and practices. Such approaches seek to reorganize the relationship between everyday and scientific practices while preserving tensions between them, maintaining the value of the everyday genre vis-a-vis the dominant form. In doing so, youth develop a deeper understanding of concepts and texts, consider the different spheres of life in which these texts function and are valued (or not), and have opportunities to use their knowledge to participate in disciplinary domains and civic life more broadly (Gutiérrez, 2014).

We build from lines of work that have established how students can leverage their personal connections to (and tensions with) data to make sense of it and how students can collect, explore, and manipulate data to create representations that matter to them (see Wilkerson & Polman, 2020). As they work with data, students can engage in data moves (Erickson et al., 2019), such as filtering to isolate cases of interest, grouping data in particular ways, or adding to and merging data sets to include new information. In our work, we are particularly interested in the ways that students use their everyday experiences as a resource to motivate or deepen their engagement with data sets. Students can transform data sets to reflect their own sociohistorical lives and engage in syncretic data reasoning, the ongoing interplay between everyday experiences and analysis

of data.

In this work, we explore how students write “data stories.” Data stories are a type of syncretic text that uses narrative to integrate data, personal experiences, and other forms of social, cultural, and historical knowledge and research (Wilkerson et al., 2021). Although narrative is not often employed in STEM contexts, students can use storytelling to organize their understanding of data based on previous experiences, thereby making sense of the connections between descriptive and inferential statistics and understanding data analysis as complex endeavor guided by human decisions (Gillborn, Warmington & Demack, 2018). For the topic of environmental racism, students can use data stories to critically examine how race has structured communities’ exposure to environmental health hazards and to advocate for related social change.

Research Questions

Given our interest in the construction of texts that bring together disparate ways of reasoning, we ask:

1. What experiences do youth leverage to generate research questions about environmental racism?
2. How do youth use qualitative and quantitative evidence to reason about and advocate against environmental racism through the creation of a data story?

The *Writing Data Stories* project takes a syncretic approach to integrating narrative and data. This paper reports on a week-long, online summer workshop on the topic of environmental racism. Participants (n=12) ranged in age from 12 to 14, were from both dominant and non-dominant backgrounds, and were mostly from the greater Bay Area. In the workshop, students explored case studies of environmental racism around the world, interviewed family members, and listened to a presentation from a local African American environmental activist. They used the Common Online Data Analysis Platform (CODAP) data visualization tool (Finzer & Damelin, 2016) to engage in open-ended investigation and manipulation of public data sets on social indicators (e.g., income, race, asthma) and environmental indicators (e.g., air pollution, water pollution, pesticides) in addition to conducting internet searches and other forms of journalistic research. Many examples in the course and the initially provided data sets were grounded in the physical geography and history of the Bay Area. As a culminating project, participants developed a data story about a research question of their own choosing.

Data sources included video recordings of the class sessions and one-on-one conferences with students, student artifacts, and post-session student interviews. We analyzed these data for the subset of consented students who participated in all workshop activities (n=6). We identified the *questions* that these students posed and coded for the *experiences* that inspired the questions (e.g., personal experience, internet research, preliminary observation from provided data sets). We then identified the *claims* that students made in response to their question, coding for the type of *evidence* they leveraged (e.g., provided data set, external data set, personal experience, qualitative research) and the presence of any *data moves* (e.g., sorting, grouping, filtering, merging). Using the coding, we created a timeline of each student’s reasoning about environmental racism and constructed a narrative case describing how their thinking developed over the course of the workshop. We then selected two students who served as contrasting cases and reflected broader patterns in the larger data set.

Findings

We describe two cases that represent distinct approaches that emerged in our analysis. One student, Elena, drew from her own experience and the narrative of an activist to explore pollution in North Carolina. She engaged in a wide range of data moves, merged new information with data sets, and considered race in her reasoning. Another student, Narun, generated his question about the relationship between air pollution and poverty from his initial observations of patterns emerging from the data set. In contrast to Elena, he incorporated fewer data sources and did not mention race as a factor in his reasoning. We describe these two examples below.

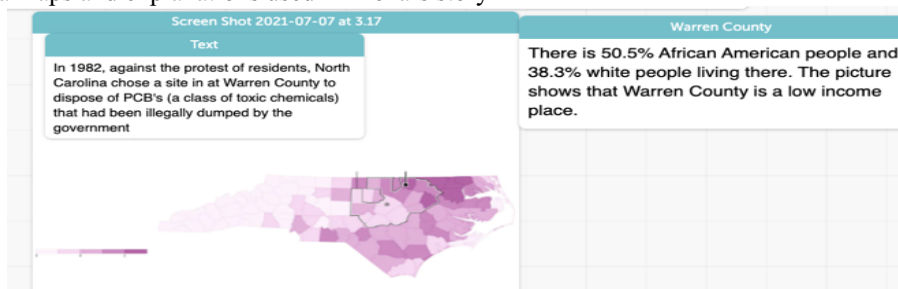
Case 1: Elena

Elena is fourteen years old, and her family immigrated to California from Central America. Early in the workshop, she reported on her own family’s exposure to air and water pollution as farm workers in California’s Central Valley. For her project, Elena chose to explore the history of Ben Chavis, an African American environmental activist from North Carolina she learned about through internet research (see Figure 1). She commented “I felt that I already had, like, covered the topic of the Latin community...and I wanted to also cover other kinds of races. [The activist of color who spoke to the group] talked about how they put factories next to an African American community and it inspired me.” By grounding her story in the actions of activists protesting against the dumping of toxic waste in their community (see Figure 1), she sought to relate her family’s own experience to those of other minoritized groups, revealing broader systemic injustice.

Elena explored a data set on pollution in North Carolina, engaging in multiple data moves that were motivated by Ben Chavis’s story. She first *filtered* the data, selecting Chavis’s home county and comparing it to

a nearby county with higher income, marking both on the map with a specific color. She then *merged* the data, looking for another source that reported the racial composition of the two counties (see Figure 1). Her argument thus coordinated data about income, race, and toxic waste to support Chavis' argument that the dumping incident constituted an act of racism. She also investigated the chemicals that were dumped at the site (PCBs) and searched internet articles for evidence of racial strife in the area, integrating quantitative and qualitative data to explain the historical event.

Figure 1. Data maps and explanations used in Elena's story

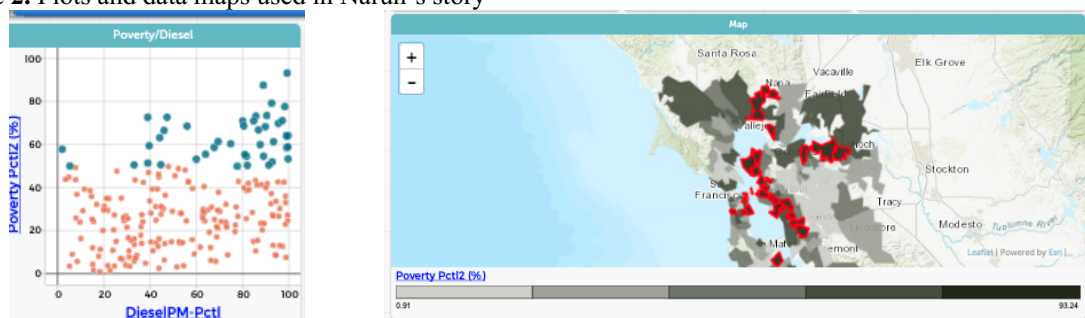


Case 2: Narun

Narun is thirteen years old, is of South Asian descent, and lives in a wealthy suburb of the Bay Area. In his interview, he reported that he did not have much personal experience with pollution. Early in the workshop, he reported researching Flint, Michigan. He commented “I think it was a racism case. It was basically a town in Michigan where they cut off clean water for the citizens of the town.” However, in his own project, Narun chose to investigate the relationship between income and pollution in the Bay Area. He reported that he chose this topic because “whenever I’m in San Francisco or the Bay Area, I see lots of homeless people outside and it doesn’t feel too good. It’s important because being homeless is very difficult.” Thus, he grounded his investigation in a situation that observed that invoked his empathy and concern.

Narun reported in the interview that he “looked at a bunch of ...factors, but chose to do poverty and air pollution.” He looked for which type of air pollution had the strongest relationship to poverty, and settled on diesel emissions. In his final presentation, he filtered to show the top quintile of poverty and demonstrated it on a map and a graph (Figure 2). He then drew relationships, referencing the map to claim that “poverty is high along the coast...and you also see that air pollution is high there too.” However, Narun did not report on the racial composition of these counties, even though it was provided in his data set, nor did he discuss the history of residential segregation and industrial development in the area, even though it was presented in the workshop. Narun’s analysis of poverty and income was well-reasoned and supported by the data; yet, it was incomplete in that it did not examine the relationships between race and income. While we have no direct evidence of why he did not incorporate race, we propose that we may not have adequately scaffolded the use of quantitative data and other sources of information related to race. We draw implications from this case in our discussion.

Figure 2. Plots and data maps used in Narun's story



Discussion

Through this analysis, we add to lines of work that have demonstrated the value of syncretic approaches to science learning. Specifically, we have shown how Elena’s family experience with racism promoted the use of data moves (Erickson et al., 2019) that deepened her work with data, as well as expanded her understanding of environmental racism as a form of systemic injustice across minoritized communities. This case demonstrates

how a data story can act as a syncretic text (Gutiérrez, 2008) in which students integrate personal experience, data, and multiple types of evidence to forward claims about environmental racism. Such a text reorganizes disciplinary learning by elevating personal experience, narrative, and historical understandings as valued and connected resources for supplementing quantitative data sets in STEM contexts. Syncretic learning was supported by the open-ended nature of the inquiry, which allowed students to choose a topic that related to their own experiences, suggesting the value of integrating more student choice into explorations of data.

Yet, our analysis showed that some students explored income without considering race, which aligns with previous work showing race is often avoided in STEM contexts (Philip, Olivares-Pasillas & Rocha, 2016). A key implication of this analysis is that teaching about racial justice does not necessarily prepare students to adopt this lens in their own data reasoning and research. In considering Narun's case, and others like it, having less personal experience with structural forms of injustice like environmental racism may have made these students less likely to conduct an analysis with a racial lens. As a result, we are dedicating more time to discussing, modeling, and critiquing the use of racial categories and integrating more related data visualizations in our curriculum, as well as explicitly discussing the relationships between race and income.

Additionally, we are ever more aware of the limitations of quantitative data in exploring issues related to race (Gilborn, Warmington and Demack, 2018). This analysis surfaced a clear need to give more support to students to understand the social, political, and historical factors that could help explain the quantitative patterns they observed. As we develop the curriculum, we are carefully considering how to support students in linking historical policies and social structures to present-day environmental inequities. Contextualizing data in this way requires an interdisciplinary approach that draws tools from history/social studies to help students to understand how structural racism has been organized in the past and continues to be organized in the present, as well as tools from ethnic studies to help students to understand their own identities and relationships to systems of power. Overall, these findings underscore the importance of drawing from diverse pedagogical and disciplinary approaches that support students in considering race from multiple perspectives while analyzing quantitative data sets. Understanding the structural causes of current environmental inequities is critical to helping students to imagine more just futures and engage in related social action.

References

- Gutiérrez, K. D. (2008). Developing a sociocritical literacy in the third space. *Reading Research Quarterly*, 43(2), 148-164.
- Gutiérrez, K. D. (2014). Syncretic approaches to literacy learning: Leveraging horizontal knowledge and expertise. In P. Dunston, L. Gambrell, K. Hadley, S. Fullerton, & P. Stecker (Eds.), *63rd literacy research association yearbook* (pp. 48-61). Literacy Research Association.
- Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, 12(1).
- Finzer, W., & Damelin, D. (2016). Design perspective on the common online data analysis platform (CODAP). Paper presented at American Educational Research Association conference, Washington, D.C.
- 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.
- Lee, V. R., & Wilkerson, M. H. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. Commissioned paper for the National Academy of Sciences, Engineering, and Medicine.
- 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.
- Tessum, C. W., Paoletta, D. A., Chambliss, S. E., Apte, J. S., Hill, J. D., & Marshall, J. D. (2021). PM2.5 pollutants disproportionately and systemically affect people of color in the United States. *Science Advances*, 7(18).
- Varelas, M., Morales-Doyle, D., Raza, S., Segura, D., Canales, K., & Mitchener, C. (2018). Community organizations' programming and the development of community science teachers. *Science Education*, 102(1), 60-84.
- Wilkerson, M., Finzer, W., Erickson, T., & Hernandez, D. (2021). Reflective data storytelling for youth: The CODAP Story Builder. In *Proceedings of the 20th ACM SIGCHI Interaction Design and Children Conference (IDC '21)*. Worldwide online, 24-30 June.
- Wilkerson, M. H. & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning [Special Issue]. *Journal of the Learning Sciences*, 29(1), 1-10.