

"Bettering Data": The Role of Everyday Language and Visualization in Critical Novice Data Work

Ben Rydal Shapiro^{1*}, Amanda Meng², Annabel Rothschild², Sierra Gilliam¹, Cicely Garrett², Carl DiSalvo² and Betsy DiSalvo²

¹Georgia State University, USA // ²Georgia Institute of Technology, USA // ben@benrydal.com // am377@gatech.edu // arothschild@gatech.edu // sgilliam7@student.gsu.edu // cicelycmgarrett@gmail.com // cdisalvo@gatech.edu // bdisalvo@cc.gatech.edu

*Corresponding author

ABSTRACT: Informed by critical data literacy efforts to promote social justice, this paper uses qualitative methods and data collected during two years of workplace ethnography to characterize the notion of critical novice data work. Specifically, we analyze everyday language used by novice data workers at DataWorks, an organization that trains and employs historically excluded populations to work with community data sets. We also characterize challenges faced by these workers in both cleaning and being critical of data during a project focused on police-community relations. Finally, we highlight novel approaches to visualizing data the workers developed during this project, derived from data cleaning and everyday experience. Findings and discussion highlight the generative power of everyday language and visualization for critical novice data work, as well as challenges and opportunities to foster critical data literacy with novice data workers in the workplace.

Keywords: Data science education, Critical data literacy, Social justice, Data visualization, Workplace ethnography

1. Introduction

Data collection? ... that's the first time I've heard that term – Shannon

I realized that, hey one thing about data, you have to learn it, you have to get in with it, you have to love it for what it is, and you have to be able to love it to organize it. And I was like oh this is too much! – Nia

It's a trauma there too that keeps me like, "ok keep it easy." It's just a cautionary thing that I constantly think about...like do not insult the man in blue...don't give him a reason to think that we are against the man in blue...like we want to collaborate safely [with this data] and um, not overstep any type of boundaries and also protect them as well. – Leo

These reflections come from young adults at the beginning of their careers in data work, whom we characterize in this paper as "novice data workers." These people are beginning to learn — or aspire to learn — how to perform data work, but do not have formal training or education in computing, visualization, or statistics. The above quotes represent how these workers characterized early-stage data entry and cleaning work while employed at DataWorks, a workplace training program and data services provider that employs historically excluded populations to work with community data sets. The language they use contrasts with that of professionally trained data scientists, demonstrating how these novice data workers can hold different views of academic terms such as "data collection" and characterize challenges of critical data work differently. Each novice worker also carries personal histories and even trauma, which impacts how they work with data — particularly, in this case, during a project focused on critically engaging with arrest data to support police-community relations.

This paper is grounded in research that identifies everyday language and other sense-making practices as intellectual resources for learning and teaching (Warren et al., 2001; Nasir et al., 2006), and is aligned with critical data literacy efforts that promote justice-centered approaches in computer and data science education (see Benjamin, 2019; Brown, 2021; D'Ignazio & Klein, 2020; Matuk et al., 2020; Pargman et al., 2021; Vakil, 2018; Wilkerson & Polman, 2020; Yadav et al., 2022). We build on these efforts using qualitative methods and data collected during two years of ethnographic fieldwork at DataWorks to characterize the notion of critical novice data work.

We begin by describing our motivation and DataWorks. Then, we review relevant research concerning critical data literacy, situative studies of data science in the workplace, and alternative approaches to visualization as used in the learning sciences. Subsequently, we describe the empirical data and methods that inform our

subsequent two-part analysis. First, we present key findings from a thematic analysis of interviews with the first cohort of five novice data workers employed at a DataWorks program we launched at a large public technology university located in the US South. This cohort included three women and two men — all Black and/or African American people aged between 19–26 years old who possessed little previous experience working with data (e.g., typically only a high school diploma or GED) — named Shannon, Nia, Jalen, Leo, and Terrance (pseudonyms). Our thematic analysis identifies their simultaneous unfamiliarity with data science terminology at the start of their employment and ability to characterize and humanize such terminology using everyday language, including gesture.

Second, we use a case study approach to analyze a project in which three of these novice data workers critically engaged with arrest data from their local police department to support police-community relations. We focus on the challenges they faced when cleaning data and engaging critically with data. Likewise, we share the workers' novel use of participatory visualization with our research team to contextualize and humanize arrest data. Notably, the issue of police-community relations is a particularly salient social justice issue for Black Americans in the United States. Police violence remains a leading cause of death for Black men (Edwards et al., 2019); Black youth are stopped, questioned, and physically detained at rates far higher than white youth (Crutchfield et al., 2012; Epp et al., 2014); and tensions between police and the community following recent killings of Black men and women including Michael Brown (age 18), Breonna Taylor (age 26), Tamir Rice (age 12), George Floyd (age 46), Atatiana Jefferson (age 28), and Daunte Wright (age 20) have led to national protests about the racialized nature of police violence. Thus, this case study contributes to studies that have examined personal experiences and narratives of trauma from police violence in Black communities (see Smith Lee & Robinson, 2019).

Findings and discussion synthesize our two-part analysis to highlight the value of everyday language and visualization for critical novice data work, as well as challenges and opportunities to foster critical data literacy with novice data workers in the workplace. We conclude by outlining future research directions that expand on the notion of critical novice data work and address inherent limitations to this qualitative study.

2. Motivation and background

2.1. Broadening and diversifying participation in computing

For decades, United States researchers and practitioners have attempted to broaden participation in computing — particularly for Black Americans — by enlarging the pipeline of potential computer science students. Despite these efforts having been in place since the early 2000s, there has been little measurable impact on the percentage of Black Americans majoring in computing fields. Over the past two decades, for example, only 3.6% of undergraduate degrees in computing have been awarded to Black Americans (Computing Research Association, 2019).

More recently, researchers and practitioners are moving beyond quantitative inclusion of underrepresented groups. They now advocate for more expansive, participatory, and democratic approaches that foreground social justice to foster participation in computing and data science. Some demonstrate how computing education can apply critical traditions in education to interrogate the sociopolitical context of computing education (Vakil, 2018). Others critically explore the role of Black youth in computing and how they have experienced power imbalances (Rankin & Henderson, 2021). Still others use an abolitionist framework to explore Black empowered futures in computing (Benjamin, 2019).

These efforts are beginning to inform a paradigm shift in computer and data science education that motivates our work. This shift encourages researchers to explore race and identity in computer science settings; structure learning environments to address the power imbalance that permeates the tech sector; and conduct critical research on the ethical and fair use of data across school and workplace settings (see Crooks & Currie, 2021; DesPortes et al., 2022; Gray et al., 2022; Jayathirtha et al., 2020, Santo et al., 2019; Silvis et al., 2022; Yadav et al., 2022).

2.2. What is DataWorks?

DataWorks is a new model for providing data services to companies and non-profit organizations in ways that aim to foster diverse approaches to data science, support equitable labor practices, and develop just forms of

engagement between universities and communities. DataWorks employees are from historically excluded communities and learn entry-level data science skills such as cleaning, formatting, and labeling datasets (e.g., for use in machine learning algorithms) using real datasets submitted by local companies and non-profit organizations. This training aims to empower workers and support pathways to long-term, full-time employment in other kinds of organizations. Through DataWorks, we hope to outline one kind of model for members of historically excluded communities to learn entry-level data science skills. We also introduce DataWorks to provide an example of a sustained research context and community of practice that develops new knowledge and approaches to teaching data science that acknowledge and respect diverse subjectivities.

3. Literature review

3.1. Critical data literacy

Critical data literacy is vital to broadening and diversifying participation in computing, particularly as we face realities of how technologies perpetuate racism and systemic oppression. Scholars of critical data literacy identify technical and social components of data literacy that are complementary and indivisible (Tygel & Kirsch, 2016). For example, the technical skills required to analyze and work with datasets should be paired with efforts to understand how that data is embedded in specific local contexts, including efforts to work with communities and people who live in those contexts from which the data is about. Additionally, Wolff et al. (2016) posit we are more likely to gain competencies that allow us to learn from and solve problems with data when we participate in a full inquiry process. These scholars echo the work of Bhargava et al. (2015), who assert that data literacy must (1) foster adaptive capacities and resilience instead of teaching platforms and technical languages, and (2) empower people in meaningful and effective ways. Bhargava et al. (2015) also identify a critical challenge to data literacy concerning the need to better understand the importance of context.

To apply and expand on these theoretical foundations of data literacy, we highlight a context in which critique and critical inquiry are dissuaded: the workplace. We also address the challenge of understanding context through analysis of everyday language used by novice data workers and a case study of a project to clean and visualize arrest data, demonstrating how critical inquiry and data work interrelate. Our understanding of the critical data literacy process during this project—whereby data literacy and critical consciousness circulate in a positive feedback loop—comes from Paulo Freire’s theory of reflection and praxis in learning in *Pedagogy of the Oppressed* (Freire, 1968). Freire defines how critical consciousness is achieved through reflection in dialogue and practice that involves learning to perceive contradictions. Using the Freirean process of critical consciousness building, we investigate the context of novice data workers as they clean and visualize local arrest data.

3.2. Situative studies of data science in the workplace

In fields including statistics education and computer-supported cooperative work, ethnographic methods have long been employed to study work practices and reveal invisible labor (see Bakker et al., 2006; Noss et al., 2000; Star & Strauss, 1999; Suchman, 1995). Ethnography collects the formal knowledge people use in their work. That is, as a methodology, it examines the people who do the work, not just tools or technical systems for work.

Informed by this history, researchers are beginning to conduct situative studies of data science in the workplace. For example, Muller et al. (2019) use a grounded theory analysis of interviews with professional data scientists to characterize five ways humans influence data work practices and the outputs of data science systems: Through the discovery (e.g., finding a public dataset), capture (e.g., making selections or substitutions), curation (e.g., cleaning or converting), design (e.g., imputing missing or validating data), and creation (e.g., simulating) of data. Passi and Jackson (2018) use ethnographic fieldwork similarly, to reveal how trust is negotiated during the everyday work of corporate data science teams and how this negotiation shapes the development of data science systems. Furthermore, scholars of human-centered data science in the workplace are beginning to ask questions such as: “How might we mitigate the (environmental, social) harms imposed upon workers involved in industrial technology production? How might we design for “good” jobs? How might we as researchers inform policy initiatives that directly influence the conditions of digital labor?” (Fox et al., 2020; also see Aragon et al., 2022).

These situative accounts of data science continue to shape data science education and practice. Yet, they focus almost exclusively on the perspectives of expert data scientists or corporate data science teams (see Rothschild et al., 2022 for an exception). For example, they rarely have considered the experiences of Black Americans in the

workplace. In response, we center the perspectives of novice data workers from historically excluded communities in data science. Their perspectives expand on and challenge prevailing accounts and assumptions about data and data work. Further, we detail the types of challenges that can arise for people and organizations through efforts to foster critical data work in the workplace.

3.3. Alternative approaches to data visualization and the learning sciences

Our project case study also draws from a growing body of research in the learning sciences expanding on what we see as alternative approaches to data visualization. While information visualization has prioritized the exploration, analysis, and presentation of raw data (see Card et al., 1999), alternative approaches to visualization foreground alternative ways of knowing, non-binary approaches to data, and the contextual or human dimensions of data (see D'Ignazio & Klein, 2016; Dörk et al., 2013; Lupi, 2017).

The project described in our case study sought to empower workers by using visualization in ways informed by Matuk et al. (2022) who demonstrate how artistic practices can foster an accessible and personally relevant approach to critical data literacy. Specifically, their research draws from Data Humanism (Lupi, 2017) and shows how data-driven art provides ways to contextualize ideas about data science in the real world and pursue personal interests. In our work we provided opportunities for workers to integrate art and more traditional forms of data visualization to embrace ideas such as subjectivity and serendipity over objectivity and prediction and to encourage more personalized visualization designs and grammars. Similarly, our understanding of data cleaning and visualization as both technical and social processes is informed by Kahn (2020), who highlights social and familial dimensions of technical practices such as data wrangling. Notably, her work leverages new digital mapping and dynamic geovisualization tools that allow youth and families to link personal reflections about their own data with broader societal issues. These broad issues are represented in aggregate data through interactive, digital maps that represent personal, family migration stories or family geobiographies (also see Taylor & Hall, 2013; Shapiro et al., 2020). Finally, we draw from learning scientists demonstrating the need for data visualization pedagogies to better support teachers and students to negotiate racialized contexts of data as they emerge during discussion (Philip et al., 2016). These ideas begin to outline aspects of a critical visualization pedagogy that require an equal commitment to develop racial literacy through environments that interrogate processes of race, racism, and racialization.

We draw from and contribute to this body of work by exploring everyday perspectives novice data workers use to understand data visualizations and illustrating new roles visualization can play to support critical novice data work.

4. Methodology

We answer the following research questions:

- What sense-making resources (e.g., ways of using language, experience, social practices) do novice data workers bring to data work?
- What challenges do novice data workers face as they begin to critically engage with data for the first time?
- How does the process of cleaning data support novice data workers' critical questioning and visualization of data?

Our analysis is broken into two parts. First, we share a thematic analysis of semi-structured interviews with Shannon, Nia, Jalen, Leo, and Terrance as they began their employment at DataWorks. These in-person interviews lasted 1–2 hours and elicited workers' perceptions about data and data work through questions including: What does the term data mean to you? Where or do you encounter data in your daily life? Can you describe any representations of data you encounter in your daily life? We analyzed their responses using thematic analysis in the grounded theory tradition (Charmaz, 2006; Glaser & Strauss, 1967). All interviews were video recorded, and transcripts were produced by the research team. Afterwards, we conducted an open coding process, identifying themes relevant to how workers perceived data and data work. We then iteratively discussed and refined these themes over five months to produce the key themes presented in this paper.

Second, we use a case study approach (Yin, 2009) to analyze audio and video recordings, representations, and reflections collected during a real world, 10-week long project completed by three of these workers at DataWorks approximately one year into their employment. Our motivation for this project was multifaceted.

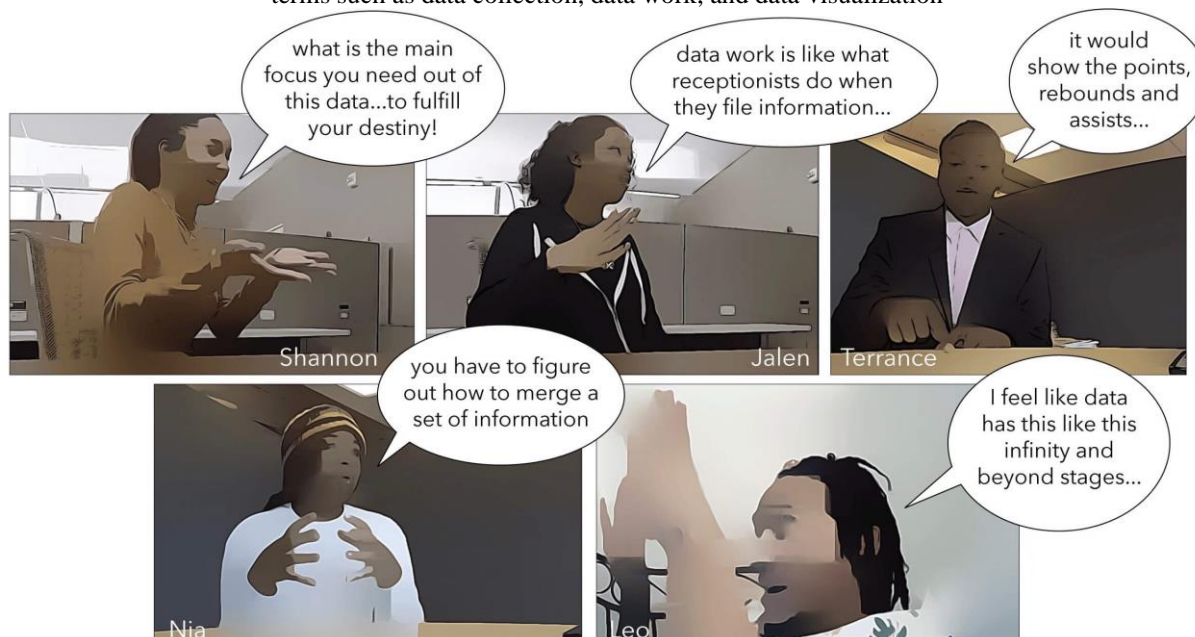
First, this project was the first time that workers developed their own data cleaning plan, as opposed to entering data, allowing us to see how novices approached such work for the first time. Second, most projects at DataWorks required highly specific deliverables for clients and provided little time for critical or creative dialogue and reflection. DataWork “clients” were members of a local criminology department who provided datasets and orienting tasks for the project, which empowered us to critically engage with and communicate arrest data in ways that foregrounded workers’ perspectives on police-community relations. Finally, as we opened this paper, the issue of police-community relations is a particularly salient social justice issue for Black Americans.

Before continuing our analysis, we acknowledge the positionality of our research team, all of whom are from the United States. The lead author, who interacted most with workers in the workshop, is a white male faculty member whose research draws from the learning sciences and critical visualization. The second author is a white female faculty member whose research focuses on data activism. Other authors include a Black female who is a leading racial equity consultant in the city where DataWorks is located, and two white faculty members, one male and one female whose research spans participatory design and learning. Additional team members include a white Ph.D. student and a Black Ph.D. student who have educational backgrounds in computer science. Team members have a long history of working with non-profit and civic organizations in the region and have consistently participated in outreach and community-based research with the goal of eliminating structural forms of oppression and to distribute the resources and opportunities we have access to for the benefit of minoritized communities. However, we acknowledge that we are not positioned to contribute to scholarship on the Black American experience. Rather, we see the context of DataWorks as an opportunity to better understand the development of a democratic data workplace that is inclusive of individuals from excluded groups and to better understand how and why our concepts of data need to be expanded to account for alternative perspectives in computer and data science. In this way, the experiences reported in this paper with lower income and Black novice data workers provide insights into our income privilege and whiteness and how that impacts data work.

5. Thematic analysis of intake interviews

Figure 1 shows excerpts from semi-structured interviews with each worker as they began their employment at DataWorks. In the following, we unpack these excerpts and share three primary themes that emerged from our analysis of these interviews.

Figure 1. Excerpts from interviews with each novice data worker at the start of their employment describing terms such as data collection, data work, and data visualization



5.1. Formal data science terminology is unfamiliar

Professional data scientists, researchers, and educators frequently assume that terms such as data, data collection, and data analysis are familiar to a general audience and students. An initial finding from our analysis of interviews with each novice data worker suggests this is not always the case, even for those who are aspiring to work with data in their professional careers. Two workers indicated the terms data and data collection were entirely new to them. For example, when asked about the term data collection, Shannon responded, *“Data collection? ... that’s the first time I’ve heard that term.”* While each of the other three workers shared that they had heard these terms before, they also indicated they were unsure of their meaning. For instance, Terrance defined data as *“The everyday use that people use in the world to like know what’s this and like what would be this and how would we be able to use it to make the world a better place.”* Further, all five workers we interviewed indicated that terms such as data analysis and data exploration were completely new to them.

5.2. Everyday language to humanize and contextualize data science terminology and practices

While these terms may have been new to workers at first, with some prompting by the interviewer to explain these terms in relation to their own lives, each novice data worker was able to characterize them using everyday language. Their everyday language highlighted what we interpreted as human and contextual dimensions.

For example, when asked to define the term data collection, Terrance shared *“collecting raw data and using it.”* Terrance was then prompted to describe the term in relation to basketball, a sport he strongly identified with, and expanded his definition through an example saying, *“First, I would show, like, it would be a name and it would show the points, rebounds and assists and, like, every point that he makes you write down, was it a two pointer or was it a three pointer or if it was a rebound, like, how many rebounds did he grab. You write, tally it down, and it would be a total. You sum that up, and you put that down on the sheet.”* While making this statement, Terrance gestured as if he was constructing a data table with each row characterizing a basketball player and columns indicating rebounds, points, and assists. The top right image in Figure 1 shows Terrance using a gesture to construct one row of this table.

Similarly, in characterizing the term data work Shannon shared, *“I guess what it means is just like when you’re working with that data it’s like you are trying to figure out the overall like what can you do to better that data I guess um I’m not cuz I don’t really work with data so once I do I feel like the overall when I do work with it it’s just to better that data.”* Subsequently, she posed different questions one might ask about data or data collection including *“What is the best way to collect data in the best way?”* These questions in turn led her to describe concepts such as the notion of *“fulfilling destiny.”* Fulfilling destiny characterizes the primary goal of data work in Shannon’s view: to make data useful to other people and organizations — that is, to fulfill the data’s destiny. The top left image of Figure 1 shows the moment Shannon used the term fulfilling destiny. It shows her gesturing as if she is handing a data set that she has made useful to another human being or organization, to describe fulfilling their destiny. Informed by Kahn (2020) as well as Van Wart et al. (2020), we suggest Shannon uses everyday language and gesture to characterize aspects of data wrangling as an inherently social and humanistic practice in ways that expand dominant conceptions of data wrangling (also see Rubel et al., 2017).

While Leo also at first struggled to talk about the term data, he eventually defined data as *“Yea, like music gives a lot of detail about you know where you are in your life. I feel like, you know, 24-25 [year-olds] most likely are going to listen to something a little bit more inappropriate...but then again you have to think about what year they were born in...and to me, that’s, like, data...to me, you gotta put the year they were born in to see what genre of music was out there, uh, collecting it from people you know and see what it is.”* Leo, who sang in his school’s chorus for many years, uses everyday language and music to characterize how age and genre of music must be understood relationally to figure out “where you are in your life.” We suggest his descriptions resemble the practice of data contextualization, or the notion that data must always be understood in context.

5.3. Using gesture to communicate data science terminology and practices

Gesture is a fundamental resource in human communication and has been studied in many disciplines including mathematics and science education (Alibali & Nathan, 2012; McNeil, 1992). Yet, computer and data science education research rarely studies gesture. Extant research primarily focuses on how teachers use gestures to communicate computer science concepts to students in classrooms (Solomon, 2021) rather than how gesture is culturally organized and improvised in situated activity (Rogoff, 2003; also see Davis et al., 2020).

Gesture was used by all data workers throughout their interviews to communicate data science terminology and practices in ways they could not do through talk alone. The previous analysis of Figure 1 highlighted Shannon's use of gesture to characterize the concept of fulfilling destiny (top left) as well as Terrance's use of gesture to construct a representation of one row of a data table (top right). Figure 1 also shows a moment when Jalen used a gesture to help her make an analogy between a receptionist who organizes files to a data worker who organizes data. Jalen's hand represents a file that she is slowly placing into a file cabinet as a receptionist would. Likewise, the figure shows Nia using a gesture to describe data analysis (bottom left). During the pictured moment, each of her hands represents a different set of information, and she brings her hands together to represent the merging of information sets necessitated by data analysis. Finally, the figure shows how Leo (bottom right) is using a gesture while he makes the following statement: *"So, I feel like data has like this infinity and beyond stages where you can do thousands of steps to complete whatever the data project that you are doing."* We suggest he uses gesture to characterize time and particularly, the significant amount of time he feels it can take to complete data-oriented projects.

These excerpts show the varying ways workers used gesture to amplify the concept of fulfilling destiny; make an analogy between a data worker and a receptionist; construct a representation of a data table; describe technical aspects of merging data sets; and characterize time through the notion of infinity and beyond.

In summary, we identified three themes that emerged from our analysis of interviews with workers as they began their employment: The unfamiliarity of data science terminology, use of everyday language to contextualize data science practices, and use of gesture to communicate data science terminology and practices. Our following case study analysis presents a different kind of analysis that expands on some of these themes and particularly, the role of everyday language in critical novice data work.

6. Case study analysis

In the second part of our analysis, we describe Nia, Leo, and Jalen's experiences during a 10-week project to explore using data to support police-community relations. The project came to DataWorks from academic colleagues who work in a criminology department. In considering the project, our motivation to involve these data workers was three-fold: Empower workers to critically discuss policing through data collected in their local communities while simultaneously challenging and expanding our own worldviews and value systems about policing; engage workers in the process of creating their own data cleaning plan; and create data visualizations informed by workers interactions with police and oppression experienced by the communities they lived.

We first describe an initial workshop that built on our existing relationships with these data workers and aimed to empower workers to decide if they wanted to take on the project. Subsequently, we analyze their progression over six weeks to develop and carry out a data cleaning plan for an arrest dataset. Finally, we illustrate how workers visualized and critically reinterpreted this data in collaboration with our research team.

6.1. Introductory workshop

The workshop opened with an hour-long conversation about workers' alignments to arrest data and policing. Workers indicated they had never seen arrest data before and were deeply interested in this project because of personal experiences with policing. Leo and Nia willingly shared they had multiple encounters with the police and had been arrested multiple times. One described the dehumanizing experience of being *"locked up for 33 days"* and unable *"to shower for 5 days."* They also shared positive relationships they had with police; one worker described enjoying spending time with an officer in their neighborhood regularly. Among all the workers, police had a significant presence in their lives and communities.

After the discussion covering workers' personal, racialized, and political experiences with policing, we visually explored a sample tabular arrest dataset from the client that encoded arrests, along with officer activity from a local police department from 2014–2019. Arrests in the dataset were encoded with spatial features such as address, longitude, and latitude; temporal features such as the date and time; classifications based on gender, race, and age; and standardized measures for reporting crime used by police departments through systems such as the Unified Crime Reporting (UCR) Program including police beat, officer shift, and arrest type (e.g., homicide, burglary, theft).

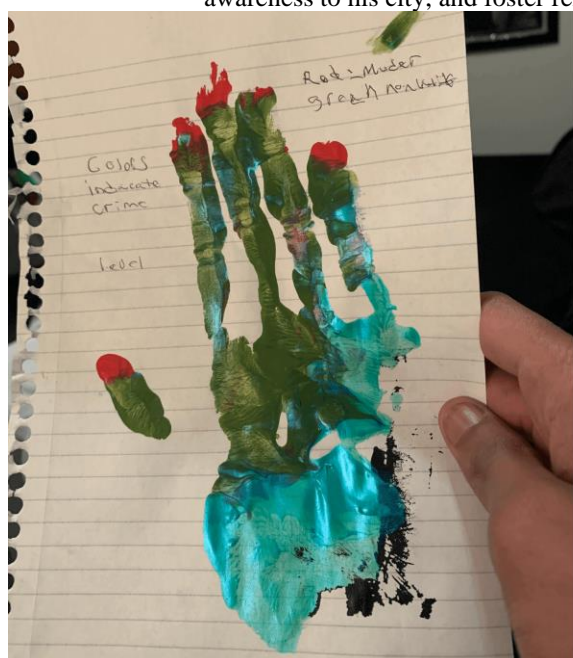
We then employed visualization tools to explore this data. While workers found such tools “cool,” they struggled to see their value and felt viewing the data as a table was more meaningful because it was a familiar representation and yielded more details of the data. Later in the analysis we return to their evolving perceptions about visualization.

The table view allowed each worker to recognize the enormous racial disparity in reported arrests as well as specific places encoded in the dataset. As Leo described, *“We know all... I know every place on this right here.”* The first hundred rows of the dataset predominantly displayed arrests of people encoded as Black from neighborhoods in which the workers lived and worked. Their lived experiences with police also led them to ask and answer questions about what constituted an arrest, what information went into a police report, and how other people and officers interpreted this dataset. Likewise, they were able to see how data become encoded from their lived experiences in ways that were new and previously invisible to them. The workers also suggested their own names for column headers such as UCR_Literal (a standard way to characterize arrest type), which signaled the type of arrest made by an officer. In this case they suggested *“gotcha”* for the header name.

As we collectively viewed this dataset, the research team introduced and discussed the notion of unreported arrests, inherent bias in arrest data collected by police departments; how the use of predictive policing systems based on such data reinforced the marginalization of communities where workers lived; and other critical questions about arrest data. These topics were quite new to workers and became areas we would revisit throughout the project.

At the workshop’s conclusion, workers emphasized they wanted to take on this project. To further empower them to shape the project in ways that were uniquely encouraged by the client, we asked them to reflect on the following prompt: What ways would you like to use these data with your community, family, or others as part of this project? Figure 2 illustrates part of Leo’s response that became central to later stages of the project. Leo used paint to create a handprint of his hand where color encodes different arrest types and the amount of space represented by each color on the handprint corresponds to Leo’s rough interpretation of the number of arrests for different arrest types. For example, red encodes homicide and covers only the fingertips of each finger indicating fewer homicides that occur in comparison to other arrest types in the dataset. In the figure text, Leo explains his view of the handprint as a method to humanize arrest data and speak both to police and the community.

Figure 2. Leo’s initial sketch and written reflection excerpt of a handprint to represent arrest data, bring awareness to his city, and foster relations between police and the community



“I feel this will bring the most awareness to our city. I chose these because of the reactions I received from family friends and a few strangers yesterday from examples. Far as being able to connect too on a human to human Level when looked at and understood. The handprint Is Definitely Something That shifts a person perception and allows them to bond with our crime data.”

6.2. Data cleaning

Subsequently, workers developed and carried out their own plan to clean a publicly available 2020 arrest dataset from a local police department. We selected this dataset in collaboration with our client because it was formatted

using two different crime reporting systems, coinciding with the start of a national effort to update ways arrest data were encoded. The data merging project opened space for workers to engage in common and important cleaning tasks including merging two differently formatted data tables, standardizing column formats (e.g., standardizing date stamps), and addressing missing data and encoding errors.

For each worker, developing an initial set of steps to begin cleaning this dataset was an imposing new task, one they viewed as extremely challenging and highly creative. For example, Nia and Leo shared the following exchange after attempting to develop an initial set of steps:

Nia: I'm going to be honest I have no structure with this data...I don't know what I'm looking for, I need some instructions...I feel like I'm looking for it myself and I don't know...I'm so used to us running the play

Leo: If I'm a doctor but you just gave me a quick little prep sheet and I'm running in here and then you got his whole body showing up and you just showing me his foot and you tell me to untie something...I'm not going to know the full extent of what is going on up all the way up under

Nia: Yeah there's no hand walking...if I don't get this step right...he's dead! Laughing...that's how I felt with this...If I don't get this one thing right it's not machine readable

Leo: Yeah but you see we have the freedom we got the freedom that's what they telling us...that's what we got to keep in mind we have the freedom to go about it like we're free right now that's why it's hard

As Nia describes, in prior projects the workers focused on “running the play” by following clear, predefined instructions about how to input and format data. Leo extends Nia’s statement, analogizing the challenges of developing initial cleaning steps to a doctor being able to see the entirety of a patient’s needs, but only having information about that patient’s foot. As Leo summarizes, having the “freedom” to develop their own set of cleaning steps was new and challenging.

Figure 3. Diagrams and README file excerpts created by workers during data cleaning

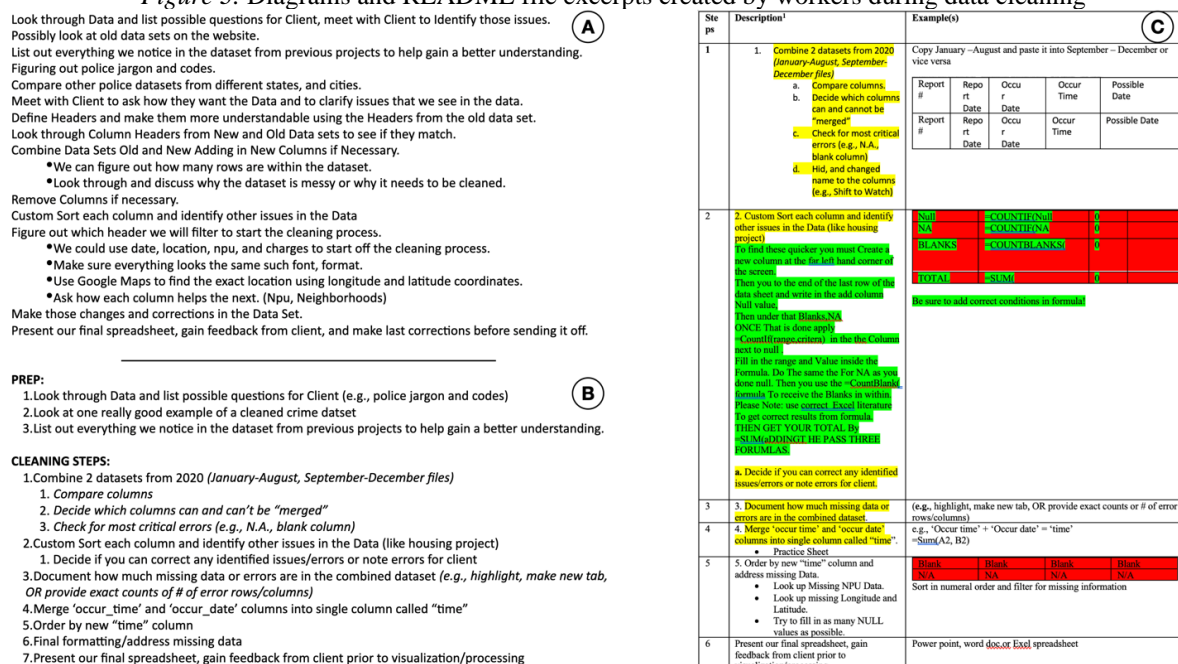


Figure 3 shows representations workers created at the beginning and end of their cleaning process. Figure 3A is a list created by workers outlining initial steps they created to clean this dataset. On one hand, it shows how workers were able to develop several clear steps important to data cleaning for this project, including recognizing the need to merge columns from different data tables and drawing from past projects to inform tasks for this project. On the other hand, it highlights a key challenge workers faced regarding how to determine what level of detail should be captured in each step. For example, lines such as “We could use date, location, npu [Neighborhood Planning Units], and charges to start off the cleaning process” are abstract. As workers would find, such steps were not detailed enough to characterize specific tasks required for the dataset. Figure 3B illustrates a subsequent list workers created a few days later that refines cleaning steps into “prep work” and

“cleaning tasks.” This list is more descriptive in its overarching organization and ordering of tasks, highlighting workers’ ability to refine their initial steps. For example, they recognized that they needed to check data formats for each column to merge columns from different data tables.

Figure 3C was produced at the end of their cleaning process. It is an excerpt of workers’ attempts to translate their steps into a README template file provided by our client. README files are used to describe a dataset’s contents, summarize data cleaning tasks, and provide other contextual documentation for future self and outsider comprehension. Such documentation is essential to conducting efficient, useful, and ethical data cleaning work. We introduced workers to practices of creating README files as they began cleaning data and workers iteratively updated a collaborative README file each week as they cleaned data.

The figure highlights an innovation workers developed to support their collaborative development of a README file. Specifically, they used color-coded text to indicate which worker performed certain tasks, to better support their collaborative work. At the same time, the figure helps reveal several challenges the workers faced when translating more informal descriptions of cleaning steps as represented by lists into a more formal README file representation. For example, providing a specific example of how they performed tasks for each step proved extremely challenging. Likewise, while indicating which worker performed each task was important for their own data cleaning work, this information would be less useful to others who would read and possibly use their README file in the future. Moreover, workers focused on carrying out their initial steps in the README file, even if they discovered that adding or removing other steps would enhance the data cleaning process. In other words, changing their initial steps was seen as too challenging. Thus, the same initial six steps from Figure 3B are reflected in Figure 3C.

In summary, this figure illustrates these data workers could develop a data cleaning plan for the first time, but also documents challenges they faced. We discuss later how many of these challenges highlight opportunities to expand the design of traditional data cleaning tools and practices to better support novice data work.

6.3. Participatory visualization

The final stage of the project began with a two-hour workshop where workers imported and explored their cleaned dataset with different exploratory visualization tools including a geovisualization tool we developed in prior work called the Interaction Geography Slicer or IGS (see Shapiro & Pearman, 2017). This workshop also served to assess the quality of their cleaned dataset. During this activity, workers engaged with interactive visualizations in ways quite different to the start of this project when they favored exploring data through tabular representations.

Notably, workers realized and focused on the value of interactive visualization to reveal what they identified as errors in the data and errors in their data cleaning process. For example, workers particularly engaged with geovisualization tools that provided ways to selectively highlight and understand how many geocoded arrest types with null values were clustered around locations such as an airport. Likewise, they were able to identify dates that had been incorrectly merged during their cleaning process. These experiences led Leo to state, *“This is actually a great way of like once you form like a good visualization like inputting your data and working between the spreadsheet and that would be so much easier especially visualizing it and not having to like do extreme amounts of research. ...that’s pretty cool...that’s pretty dope.”* In this statement, Leo identifies there can be “good” visualizations better suited for particular datasets, and he highlights the value of going back and forth between cleaning and visualizing data to identify errors in a dataset to better inform a data cleaning process. He contrasts this process with doing “extreme amounts of research” to try to identify errors from a tabular representation as he and the workers had done during their data cleaning process.

Workers also recognized the value of color and interaction techniques such as sorting or linking different views in a visualization to see trends in the quantity and type of arrests across different neighborhoods that were challenging to see in a tabular representation. Many of workers’ questions focused on comparing the type and number of arrests across different neighborhoods and neighborhood planning units. Likewise, while doing such comparative analysis, workers repeatedly made statements such as “this makes it real.” We interpreted these statements as further evidence that they found visualization more meaningful than before because they gained a sense of ownership of the dataset from developing and carrying out their own data cleaning plan.

To conclude this workshop, workers sketched different visualization ideas on post-it notes as shown in Figure 4A. The color of the post-it indicates the worker who developed the idea. These ideas included traditional bar

charts and scatterplots, virtual reality experiences, and representations integrating digital maps and handprints through tools the workers had been exposed to including the IGS.

Figure 4. Initial exploratory visualizations of arrest data created by workers



Figure 4B-D shows more refined visualizations that workers developed on their own time during the week following the workshop. For example, Figure 4C shows a series of drawings by Jalen, reflecting approaches to visualizing data she had learned previously in school. These drawings include bar charts grouped by a unit of time (e.g., a year or month) with color encoding type of arrest to compare how arrests varied over different time periods. These drawings also included the beginning of a histogram (middle drawing), which Jalen described as “more specific” than bar charts. Notably, her histogram only begins to sketch out potential ranges to group data. Leo developed quite different visualizations including a unique badging system for police officers, as well as a three-dimensional hand created by Leo that furthered his initial notion of a handprint. These visualizations by Leo became meaningful ways to see and talk about workers’ critical perspectives on this data at a subsequent project discussion during which Leo made the following statement to explain his officer badging system.

I’ve been around police on the wrong end and presenting something that tracks their badges doesn’t really sound like...it sounds kinda risky to approach with their own data...You know how we think about police you know don’t give them a reason to think that we’re trying to get them individually...I thought collectively as a county as a whole when they work together that wouldn’t be as irritating and would be more acceptable instead of getting them like physically badge by badge.

Leo’s statement reveals two things. First, it highlights a personal approach to developing police officer badges. This approach displays arrest data on a badge at a collective or county level as opposed to an individual or officer level, to protect officers’ privacy and foster relations between the police and community. Second, his statement shows how the personal experiences and trauma he experienced through encounters with police shaped his critical perspectives. Put simply, exploratory visualization allowed Leo and other workers to express that they were afraid of criticizing police officers, due to fear of police retaliation but also due to what we interpreted as a genuine value for collective action with and not against police officers.

From these many ideas, the workers decided to focus on integrating Leo’s notion of what workers began calling “the crime hand” with digital maps and arrest data visually animated in the IGS. Jalen described how this approach provided a way to “humanize” the data; Nia and Leo described this approach provided ways to communicate to the police and community; and Leo further emphasized this approach reflected ideas such as “a hand holding the city’s problems” and experiences from encounters with police such as having their fingerprints taken from them when arrested.

To advance this idea, Jalen and Leo developed many sketches of their own hands, some of which are shown in Figure 5. These sketches also explored coloring schemes workers felt were most meaningful to the communities they lived and different hand positions that could interact with data displayed in the IGS.

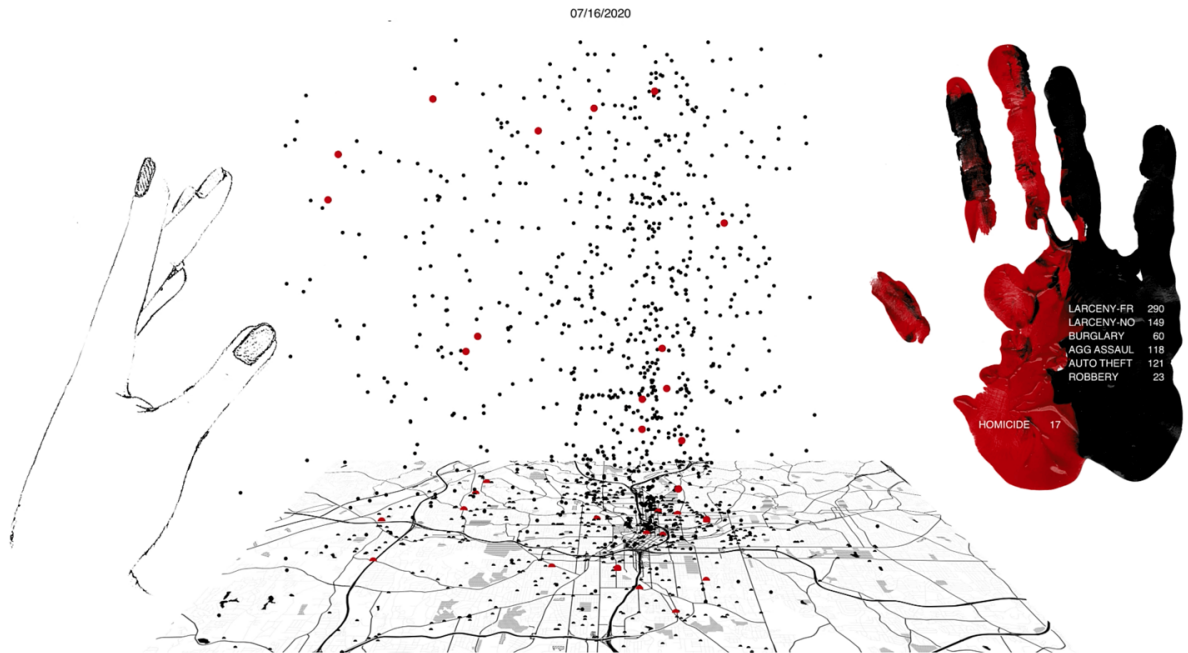
In a final workshop together, we integrated these hands with interactive IGS visualizations. This necessitated demonstrating to workers how to use tools such as Adobe Photoshop to cut out their crime hands and a video

editor to place them as shape files over interactive IGS visualizations we recorded together. This process generated several decisions and critical conversations.

Figure 5. Crime hand sketches by Jalen and Leo



Figure 6. Screenshot from final video produced by workers and the research team integrating crime hands with IGS visualizations of arrest data (Video available at <https://youtu.be/RROp5xg59gw>)



For example, workers first created data visualizations with the IGS that used several colors to encode arrests based on neighborhood planning units. This allowed comparisons of the amount and type of arrests made by police across different neighborhoods and how these evolved across 2020. Yet workers felt the use of multiple colors did not convey the dark and more serious tone they wanted to communicate. Thus, they decided to use only two colors in their final visualization: Red to encode homicides and black to encode all other arrest types. Likewise, we had critical discussions about workers' and our own stances regarding arrest data and policing. These discussions revisited previous comments and assumptions shared by workers and informed decisions about data encodings, views, and the goals of a final collaborative visualization. Notably, our research team expanded upon issues such as bias in arrest data collected by police and predictive policing software to share how such work inspired us to pursue a final visualization critical of certain policing tactics and such uses of

arrest data. Yet, workers continued to share they did not want to make something that could be interpreted as being critical of police officers. For example, as a potential title for the final visualization workers suggested: *“There is a thin line between crime and mistakes, the choices go hand in hand.”* Through such titles workers suggested orientations where they and their communities assumed blame for past encounters with officers.

Throughout this project, but especially during these conversations, the workers and our research team experienced contradictions, tensions, and risk that were deeply meaningful and expanded both our worldviews and understandings of data. For example, members of our research team who had not experienced trauma and fear of an arrest at the hands of police, were able to experience white privilege and how our relationships with the data were distinctly different from the data workers. Likewise, workers began to appreciate bias and power dynamics encoded in arrest data collected by police departments such as the notion of unreported arrests. Along with a desire to protect the privacy of individuals encoded in the data, an awareness of power inspired the workers to create a visualization that presented an aggregate depiction of arrests across their entire city as opposed to focusing on a few neighborhoods where they lived.

Figure 6 is a screenshot from a final video-based visualization that reflects these experiences and negotiations. It aims to encourage more contextual data-driven conversations about arrest data to support police-community relations. This video can be viewed at the following link: <https://youtu.be/RR0p5xg59gw>

7. Discussion

7.1. The generative power of everyday language and visualization for data science education and practice

Data science is an emerging discipline in need of more expansive language and terminology to expound, teach about, and communicate the contextual and human dimensions of data. This paper underscores the generative power of novices’ everyday language to contextualize and humanize data science concepts and terminology. For example, terms such as “bettering data” and “fulfilling destiny” coined by workers in this paper meaningfully characterize humanistic dimensions of technical concepts such as data wrangling. Our thematic analysis also illustrates how everyday language encompasses non-verbal dimensions, and that recognizing the multimodal nature of everyday language in data science is important. In particular, gesture provided a way for novices to draw from their own cultural and historical backgrounds in communicating more formal and unfamiliar data science terminology and practices. Likewise, across the project described in this paper workers drew from their personal and cultural experience with policing to describe arrest data and challenges associated with cleaning and being critical with data. We suggest their language has a descriptive power that is missing from current language used in data science. It would be valuable, for example, to develop more varied and diverse concepts to structure and work with data from the perspectives of the places and communities a dataset describes (see Eglash et al., 2021).

This paper also illustrates the value of incorporating Black novice data workers in data visualization. Their decision to include the hand and handprint to communicate and creatively reinterpret data was done to humanize the data. It prompted the viewer to connect, even identify with and embody the data pouring down onto the map in the IGS. Workers’ invention and development of the crime hand reflected their connection to crime formed by their direct experience with policing. The crime hand served as a novel and meaningful approach to communicate more than data to diverse audiences. The hand, and the activity of designing the data visualization served as another form of dialogue and communication between the data workers and our research team.

The final visualization presents an artifact that we suggest is valuable to emerging discussions and perspectives in data science education and practice. We find Black feminist theory and particularly, the theories of Lorde (2007) and hooks (1988), useful to understand and interpret this data visualization. Notably, Lorde and hooks make it clear that embodied emotion and intuition must not be ignored to gain knowledge and ultimately achieve liberation. With this theoretical lens we can begin to understand the data visualization as a product of a deep knowing and of an ongoing articulation of arrest data and the experience of being Black in a police state. Data workers had to both understand and articulate the arrest data and the experience of arrests to produce a visualization that embodied both. Likewise, in our analysis we highlighted how some of the workers’ decisions in creating this final visualization illustrated forms of critical data literacy. They chose to communicate emotion through more aggregated depictions of data due to potential biases encoded in arrest data and to protect the privacy of individuals represented in the data.

Our interpretation suggests the final visualization contributes to scholarship highlighting how alternative visualization approaches that foreground more personal and emotional aspects of data support critical data literacy efforts and pedagogies (Matuk et al., 2022). However, one could also argue that the final visualization is evidence that the emotional impact of the data on this project overwhelmed workers' and our own interest in analysis, resulting in a visualization process and final visualization that focused too strongly on communicating emotions as opposed to insights from the data. Alternatively, one could also argue that the representational system of the space-time cube on which the IGS is based provides powerful ways to communicate emotion and tell stories with data but backgrounds more traditional forms of exploratory data analysis that are important for novice data workers to engage with as they begin to work with data for the first time.

7.2. Fostering critical data literacy with novice data workers in the workplace

Throughout the project described in this paper we drew on existing research to introduce and scaffold critical perspectives of policing and arrest data. We hoped such an approach would empower novice data workers and challenge our own worldviews and value systems. Reflection and dialogue comprised our primary methods of engaging novice data workers in critical conscious building during group meetings. While forms of oppression and trauma that workers experienced through past encounters with police fostered deep interest in this project, they also caused triggered conflicting feelings. They continually expressed that they did not want to be critical of police for fear of retaliation, even as they held a genuine respect for officers. The participatory visualization stage of the project rendered this tension especially visible. Whereas we supported critique of policing and arrest data, what emerged was data workers highlighting the humans in the data, the concept of intentionality, and an invitation to bond with what the data represented. But their work was not completely absent of critique. Inserting their own hand and handprint into a data visualization critiqued how to consume the data, or at the very least, was an invitation to embody the data. Likewise, their choice to create more aggregated data visualizations of homicide and all other arrest types served to obfuscate and thereby protect the humans the data represented. Such choices also align with contemporary social movements (e.g., Data for Black Lives, Abolish Big Data) that advocate for abolishing "big data" collection of Black bodies. Without training in forms of data protection and data privacy, the workers intuitively used such a method to protect and to reduce harm of visualizing arrest data.

These findings highlight two important realities specific to our workplace case study with novice workers. First, when we conceptualize critical data literacy in the workplace, we must account for how dominant conditions of work can obscure the very histories and perspectives a critical data literacy is meant to foreground (see Johnson et al., 2021). Second, Freire highlights the importance of trusting your ability to reason, and that whoever lacks this trust will fail to initiate the dialogue, reflection, and communication that leads to critical conscious building and liberatory action. Freire would explain that this first step towards self-affirmation or self-efficacy precedes the choice of critique. Those who lack self-affirmation or self-efficacy will choose the security of fear over the risk of freedom. With this theoretical grounding, we recognize that being critical with data is more accessible to data workers with a sense of self-efficacy. However, with this project we challenged both workers' capacity to critique and their capacity to design and execute a data cleaning process, which they struggled with. After the conclusion of the project, however, we did see the workers begin to self-affirm. As Nia explained, *"The crime data project ALONE has definitely pushed us and our mindset in a different space and aspect because this is a project that we had to run, like, completely from the beginning to the end... this was something that actually made us who we are today and communicate our...you know what we do today."* It is not that critiquing data is a privilege, but that critiquing data comes from a place of power and a place of knowing. In this context, self-efficacy to clean and visualize the data was as important for generating critique as the situated knowledge the data workers brought from their lived experience.

8. Conclusion

In this paper we characterized the notion of critical novice data work, highlighting the generative power of everyday language and visualization for that work. We also presented challenges and opportunities to foster critical data literacy with novice data workers in the workplace. We conclude by outlining several research areas to expand on this work and build on inherent limitations of this qualitative study. First, this paper highlights the need to study novice data work in different kinds of workplace settings. Such research offers opportunities to contribute more descriptive language about the human and contextual dimensions of data and data work to inform data science and data science education. Second, our work highlights challenges associated with supporting critical data work in the workplace. Put simply, our work raises questions to orient future critical studies in the workplace such as how do we teach critical approaches to data and data work while also sustaining

a business practice? Third, this paper begins to detail the types of challenges novices face when using computer and data science tools such as README files for the first time. We believe there is a rich design space to explore how such tools can be expanded to scaffold novice learning in and out of schools. Finally, this work highlights how data cleaning, a process typically relegated to the background of discussions in data science education, can be an empowering activity that merits further research.

Acknowledgement

We would like to thank all the workers at DataWorks who entrusted us with their lived experiences and who generously shared their insights and time with us in ways that have resulted in deep friendships and a shared commitment to bettering data in our communities. We would also like to thank each of the reviewers of this paper who provided critically important feedback at all stages in the review process. This work was supported by the National Science Foundation under Grant No. 1951818. Any opinions, findings, and conclusions or recommendations expressed in this material are our own and do not necessarily reflect the views of the National Science Foundation.

References

- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, 21(2), 247-286. <https://doi.org/10.1080/10508406.2011.611446>
- Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-centered data science*. MIT Press.
- Bakker, A., Hoyles, C., Kent, P., & Noss, R. (2006). Improving work processes by making the invisible visible. *Journal of Education and Work*, 19, 343-361. <https://doi.org/10.1080/13639080600867133>
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
- Bhargava, R., Deahl, E., Letouze, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond data literacy: reinventing community engagement and empowerment in the age of data. *Data-Pop Alliance White Paper Series*. Harvard Humanitarian Initiative, MIT Lad and Overseas Development Institute Internews. <http://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf>
- Brown, Q., Grandison, T., Jenkins, O. C., Burge, J. D., Dillahun, T., Thomas, J. O., Erete, S., & Rankin, Y. A. (2021). Reflecting on being black in computing. *Interactions*, 28(5), 34-37. <https://doi.org/10.1145/3480237>
- Crooks, R., & Currie, M. (2021). Numbers will not save us: Agonistic data practices. *The Information Society*, 37(4), 201-213. <https://doi.org/10.1080/01972243.2021.1920081>
- Crutchfield, R. D., Skinner, M. L., Haggerty, K. P., McGlynn, A., & Catalano, R. F. (2012). Racial disparity in police contacts. *Race Justice*, 2, 1-21. <https://doi.org/10.1177/2153368712448063>
- Card, S. K., Mackinlay, J. D., & Schneiderman, B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.
- Charmaz, K. (2006). *Constructing grounded theory: A Practical guide through qualitative analysis*. Sage.
- Computing Research Association. (2019). *CRA Taulbee survey*. <https://cra.org/resources/taulbee-survey/>
- Davis, N. R., Vossoughi, S., & Smith, J. F. (2020). Learning from below: A Micro-ethnographic account of children's self-determination as sociopolitical and intellectual action. *Learning, Culture and Social Interaction*, 24. <https://doi.org/10.1016/j.lcsi.2019.100373>
- DesPortes, K., McDermott, K., Bergner, Y., & Payne, W. (2022). "Go [ing] hard... as a woman of color": A case study examining identity work within a performative dance and computing learning environment. *ACM Transactions on Computing Education*, 22(4), 1-29. <https://doi.org/10.1145/3531000>
- D'Ignazio, C., & Klein, L. (2020). *Data feminism*. The MIT Press.
- Dörk, M., Feng, P., Collins, C., & Carpendale, S. (2013). Critical InfoVis: Exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)* (pp. 2189-2198). <https://doi.org/10.1145/2468356.2468739>
- Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race-ethnicity, and sex. *Proceedings of the National Academy of Sciences*, 116(34), 16793-16798. <https://doi.org/10.1073/pnas.1821204116>

- Eglash, R., Bennett, A., Cooke, L., Babbitt, W., & Lachney, M. (2021). Counter-hegemonic computing: Toward computer science education for value generation and emancipation. In *ACM Transactions on Computing Education (TOCE)*, 21(4), 1-30. <https://doi.org/10.1145/3449024>
- Epp, E. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race and citizenship*. University of Chicago Press.
- Fox, S. E., Khovanskaya, V., Crivellaro, C., Salehi, N., Dombrowski, L., Kulkarni, C., Irani, L., & Forlizzi, J. (2020). Worker-centered design: Expanding HCI methods for supporting labor. In *Proceedings of the 2020 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1-8). <https://doi.org/10.1145/3334480.3375157>
- Freire, P. (1968). *Pedagogy of the oppressed*. Routledge.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of grounded theory: Strategies for qualitative research*. Aldine.
- Gray, C. M., Obi, I., Chivukula, S. S., Li, Z., Carlock, T., Will, M., Pivonka, A. C., Johns, J., Rigsbee, B., Menon, A. R., & Bharadwaj, A. (2022). *Practitioner trajectories of engagement with ethics-focused method creation*. <https://doi.org/10.48550/arXiv.2210.03002>
- hooks, b. (1988). *Talking back: Thinking feminist, thinking black*. Between the Lines.
- Jayathirtha, G., Shaw, M., Kafai, Y. B., & Lui, D. (2020). When a glove becomes a gun: From personally meaningful to socially critical restorying in maker activities. In *Proceedings of the FabLearn 9th Annual Conference on Maker Education* (pp. 94-97). <https://doi.org/10.1145/3386201.3386205>
- Johnson, B., Shapiro, B. R., DiSalvo, B., Rothschild, A., & DiSalvo, C. (2021). Exploring Approaches to Data Literacy through a Critical Race Theory Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI'21) (pp. 1-15). ACM. <https://doi.org/10.1145/3411764.3445141>
- Kahn, J. (2020). Learning at the intersection of self and society: The Family geobiography as a context for data science education. *Journal of the Learning Sciences*, 29(1), 57–80. <https://doi.org/10.1080/10508406.2019.1693377>
- Lorde, A. (2007). *Sister outsider essays and speeches by Audre Lorde*. Ten Speed Press.
- Lupi, G. (2017). Data humanism: The Revolutionary future of data visualization. *Print Magazine*, 30(3).
- Matuk, C., DesPortes, K., Amato, A., Vacca, R., Silander, M., Woods, P. J., & Tes, M. (2022). Tensions and synergies in arts-integrated data literacy instruction: Reflections on four classroom implementations. *British Journal of Educational Technology*, 53, 1159-1178. <https://doi.org/10.1111/bjet.13257>
- Matuk, C., Yoon, S., Polman, J., Amato, A., Barton, J., Bulalacao, N. M., Cafaro, F., Haldar, L. C., Cottone, A., Cortes, K., DesPortes, K., Erickson, T., Finzer, W., Taylor, K. H., Herbel-Eisenmann, B., Graville, C., Gutiérrez, K., Higgins, T., Himes, B., Lanouette, K., Lee, H., Lim, V., Lopez, M. L., Lyons, L., Milz, D., Olivares, M. C., Osche, E., Parikh, T. S., Philip, T., Rubel, L., Shelley, J., Rivero, E., Roberts, J., Roberto, C., Petrosino, T., Rubin, A., Shim, J., Silander, M., Sommer, S., Stokes, D., Tes, M., Trajkova, M., Urbanowicz, R., Vacca, R., Van Wart, S., Vasudevan, V., Wilkerson, M., & Woods, P. J. (2020). Data literacy for social justice. In *Proceedings of the 14th International Conference of the Learning Sciences* (Vol. 1, pp. 343-349). International Society of the Learning Sciences.
- McNeil, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., & Erickson, T. (2019). How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). <https://doi.org/10.1145/3290605.3300356>
- Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2006). Learning as a cultural process: Achieving equity through diversity. In R. K. Sawyer (Ed.), *The Cambridge Handbook of: The Learning Sciences* (pp. 489-504). Cambridge University Press.
- Noss, R., Hoyles, C., & Pozzi, S. (2000). Working knowledge: Mathematics in use. In A. Bessot, & J. Ridgway (Eds.), *Education for mathematics in the workplace* (pp. 17-35). Kluwer Academic.
- Pargman, C. T., McGrath, C., Viberg, O., Kitto, K., Knight, S., & Ferguson, R. (2021). Responsible learning analytics: creating just, ethical, and caring. In *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21)*. The Open University.
- Passi, S., & Jackson, S. (2018). Trust in data science: Collaboration, translation, and accountability in corporate data science projects. In *Proceedings of the ACM on Human-Computer Interaction* (Article 136, pp. 1-28). <https://doi.org/10.1145/3274405>
- Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*, 34(4), 361-388. <https://doi.org/10.1080/07370008.2016.1210418>

- Rankin, Y. A., & Henderson, K. K. (2021). Resisting racism in tech design: Centering the experiences of black youth. In *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (pp. 1-32). <https://doi.org/10.1145/3449291>
- Rogoff, B. (2003). *The Cultural nature of human development*. Oxford University Press.
- Rothschild, A., Meng, A., DiSalvo, C., Johnson, B., Shapiro, B. R., & DiSalvo, B. (2022). Interrogating data work as a community of practice. In *Proceedings of the ACM on Human-Computer Interaction*, CSCW2 (Vol. 6, Article 307). <https://doi.org/10.1145/3555198>
- Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making space for place: Mapping tools and practices to teach for spatial justice. *Journal of the Learning Sciences*, 26(4), 643-687. <https://doi.org/10.1080/10508406.2017.1336440>
- Shapiro, B. R., & Pearman II, F. A. (2017). Using the interaction geography slicer to visualize New York City stop & frisk. In *Proceedings of the IEEE VIS 2017 Arts Program, VISAP'17* (pp. 1-8). <https://doi.org/10.1109/VISAP.2017.8282370>
- Shapiro, B. R., Meng, A., O'Donnell, C., Lou, C., Zhao, E., Dankwa, B., & Hostetler, A. (2020). Re-shape: A Method to teach data ethics for data science education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI'20) (pp. 1-13). ACM. <https://doi.org/10.1145/3313831.3376251>
- Silvis, D., Clarke-Midura, J., Shumway, J. F., Lee, V. R., & Mullen, S. (2022). Children caring for robots: Expanding computational thinking frameworks to include a technological ethic of care. *International Journal of Child-Computer Interaction*, 33, 100491. <https://doi.org/10.1016/j.ijcci.2022.100491>
- Smith Lee, J. R., & Robinson, M. A. (2019). "That's my number one fear in life. It's the police": Examining young black men's exposures to trauma and loss resulting from police violence and police killings. *Journal of Black Psychology*, 45(3), 143-184. <https://doi.org/10.1177/0095798419865152>
- Solomon, A. S. (2021). *Embodiment in computer science learning: How space, metaphor, gesture, and sketching support student learning* (Unpublished doctoral dissertation). Georgia Institute of Technology, Atlanta, GA. <http://hdl.handle.net/1853/64777>
- Star, S. L., & Strauss, A. L. (1999). Layers of silence, arenas of voice: The Ecology of visible and invisible work. *Computer-Supported Cooperative Work (CSCW): The Journal of Collaborative Computing*, 8, 9-30. <https://doi.org/10.1023/A:1008651105359>
- Suchman, L. (1995) Making work visible. *Communications of the ACM*, 38(9), 56-64. <https://doi.org/10.1145/223248.223263>
- Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology, Knowledge and Learning*, 18(1-2), 65-93. <https://doi.org/10.1007/s10758-013-9201-5>
- Tygel, A.F., Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics*, 12(3), 108-121. <https://doi.org/10.15353/joci.v12i3.3279>
- 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. <https://doi.org/10.17763/1943-5045-88.1.26>
- Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and counterscripts in community-based data science: Participatory digital mapping and the pursuit of a third space. *Journal of the Learning Sciences*, 29(1), 127-153. <https://doi.org/10.1080/10508406.2019.1693378>
- Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The Logic of everyday sense-making. *Journal of Research in Science Teaching*, 38, 529-552. <https://doi.org/10.1002/tea.1017>
- Wilkerson, M., & Polman, J. (2020) Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, 29(1), 1-10. <https://doi.org/10.1080/10508406.2019.1705664>
- Wolff, A., Gooch, D., Caverio Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3), 9-26. <https://doi.org/10.15353/joci.v12i3.3275>
- Yadav, A., Heath, M., & Hu, A. D. (2022). Toward justice in computer science through community, criticality and citizenship. *Communications of the ACM*, 65(5), 42-44. <https://doi.org/10.1145/3527203>
- Yin, R. (2009). *Case study research. Design and methods* (4th ed.). Sage.