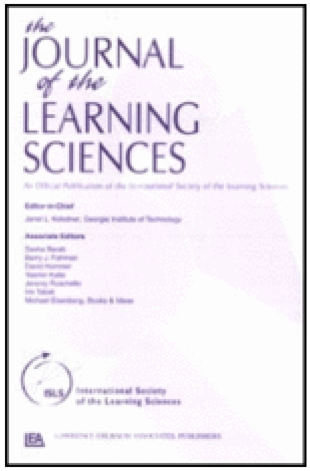

Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space

Journal:	*Journal of the Learning Sciences*
Manuscript ID	HLNS-2018-0082.R3
Manuscript Type:	Special Issue Article
Keywords:	Learning Technology < Topics, Instructional Design, Design Principles < Topics, Culture and Cultural Processes < Topics, Qualitative (includes ethnography, case study, small N) < Methodology

Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space

Sarah Van Wart
Northwestern University

Kathryn Lanouette
University of California Berkeley

Tapan S. Parikh
Cornell Tech

Abstract

Data increasingly mediates how we understand the world. As such, there is growing interest in designing initiatives to help young people learn about data — not only the techno-mathematical skills necessary to work with data, but also the dispositions needed to participate in data-centric ways of knowing and doing. In this article, we argue that as this educational goal is pursued, it is important to attend to the normative scripts that are often associated with data, and how they relate to learners' perspectives and prior experiences. We do this by examining two initiatives that aimed to help young people learn about data and its 'real-world' applications by engaging them in participatory mapping activities, directed towards the study of local community challenges. We argue that when there are mismatches between students' realities and how reality is described to work in data science, making the time and space to examine these contradictions can lead to a robust engagement with data science and its applications. These findings have implications for how we might better design tools and learning activities that connect data science with the broader contexts that frame young people's lives.

Data increasingly mediates modern life, with implications for how the world is shaped and understood across a range of human activities. As such, there is growing interest in helping young people learn *data science* (e.g., Berman et al., 2016) and how it might be applied to various contexts and domains central to their lives. In this article, we argue that as this goal is pursued, it is important to examine the various *scripts*[1] that surround data science, and how some of the values and expectations embedded in these scripts can shape data science practice.

A growing body of research has found that situated approaches to data science education can bring these scripts to the surface, creating both tensions and possibilities within data science education. On one hand, situating data science in familiar, disciplinary contexts can help learners connect their local and experiential knowledge to new disciplinary ideas, build technical, methodological and spatial reasoning skills, and develop a more critical perspective towards the places they inhabit (Elwood & Mitchell, 2013; e.g., Enyedy & Mukhopadhyay, 2007; Lanouette, Van Wart, & Parikh, 2016; Philip, Schuler-Brown, & Way, 2013; Rubel, Hall-Wieckert, & Lim, 2017; Taylor & Hall, 2013). On the other, a situated approach can also lead students to challenge some of the assumptions that surround data science — particularly when data-centric forms of participation do not account for students' existing knowledge, experiences, and positionalities. For instance, students may already be quite knowledgeable of phenomena, such as residential segregation, that data science activities are intended to illuminate (Enyedy & Mukhopadhyay, 2007). Students may also reject the idea of data-informed decision-making, citing decisions instead made on the grounds of existing authority and power. They may reject conclusions drawn from data that do not resonate with their experiences and values (Rubel et al., 2017). They may

[1] Common-sense descriptions of data that imbue values and assumptions of the world, elaborated in the Conceptual Framing.

also find that the assumptions embedded in data-centered activities neither reflect their notion of their local community nor their vision of how they might want to participate in helping their community (Philip, Way, Garcia, Schuler-Brown, & Navarro, 2013). In short, while situating data science in familiar contexts can be a productive way to learn about data, it can just as easily present challenges to data-centric ways of knowing and doing which are important to acknowledge and understand.

This article examines the role of scripts in data science education by considering a key site of situated data science practice: managing public resources, including air, water, schools, parks, community centers, transportation, and housing. This process is technically, politically, and logistically complex, shaped by formal and informal procedures, policies, and laws that are often mediated by data-intensive practices (Elwood, 2006). It also involves negotiating divergent priorities across diverse political constituencies, amid unequal power relations. This context, because it is data- and people-intensive and because it has such a profound impact on individual and community wellbeing, offers a situated way for young people to learn about data, and to examine *whether* and *how* data-centric ways of knowing figure into broader civic efforts.

In this paper, we consider two projects where young people participated in various data-intensive practices including participatory mapping (Chambers, 2006), participatory sensing (Burke et al., 2006), and community ethnography (Calabrese Barton & Tan, 2019) as a means to understand and advocate for important shared local public resources. Our analysis of these projects is guided by the following research questions:

1. What were the prevalent scripts around data in each project and how did they emerge?
2. How did the negotiations surrounding these scripts shape students' engagement with data science?

In the sections that follow, we expand upon our conceptual framework, and describe the varied project contexts, design activities, and supporting tools. We then examine the different scripts that surfaced during the two data science-focused projects and how emergent tensions were negotiated. We conclude by discussing implications, considering not only how data science is introduced to young people but also how student critiques can be productively integrated into shaping a more robust and critical data science.

Conceptual Framework

Scripts, Counterscripts and Third Spaces

To help us examine why scripts matter and how they might be productively navigated within young peoples' learning activities, we draw on Gutiérrez, Rymes, and Larson's (1995) notion of scripts, counterscripts, and third spaces. Gutiérrez et al. define *scripts* as familiar and comfortable ways of knowing and acting in the world, which vary across contexts, cultural communities, and power relations. They demonstrate this idea by examining an activity in a high school classroom, where the teacher (an actor with more positional authority) implicitly defines being 'knowledgeable about the world' as someone who knows the *Los Angeles Times* headlines. The authors ask us to consider how this definition is so easily taken as a given, and why it so neatly intersects with the teacher's own daily practice. They argue that certain cultural practices that constitute 'knowing about the world' have become dominant or 'transcendent' — defined by people who are in a position to make these determinations often without even noticing. With no way into the learning activity and no opportunities to negotiate what counts as knowledge (given the teacher's seemingly immutable script), students create their own counterscripts, using the teacher's words for jokes and side conversations.

However, these clashing scripts which each have something to say about the world from a particular perspective also have the potential to "refract" and bend one another into a "responsive/collaborative" script, or "*third space"* (p. 465, ibid). When this happens, new opportunities open up for genuine dialog across multiple knowledges, expectations, and values, thereby shifting what counts as knowledge within a social practice. While this third space is neither stable nor sustainable — more of an ideal than a steady state — these momentary shifts in perspective have the power to change how people think, thereby shifting the transcendent script, and culture itself.

Social Valences of Data

Scripts are also examined in the science and technology studies (STS) literature. As boyd and Crawford (2012) note, data science is often described in terms of its potential to advance human understanding and inform action. Fiore-Gartland and Neff (2015) elaborate on this idea by deconstructing these normative data science scripts into smaller units of analysis, which they call the *social valences of data* — common expectations and values that explain how and why people gather, interpret, and marshal data towards particular goals. The authors argue that these valences can be assembled in different ways as data practitioners make sense of their work. In this article, we examine three of these 'valences,' which we found to be prevalent across both projects:

1. *Discovery*: the idea that new forms of data analysis, applied to bigger, more diverse datasets, can lead to important insights and discoveries.
2. *Actionability*: the idea that data-driven discovery can subsequently be used to improve efficiency, service delivery, and lives based on facts.
3. *Truthiness*: the idea that data-backed ideas carry more authority over other ways of knowing, regardless of the validity of the analysis.

Within our projects, these three valences worked together to constitute a version of the normative data science *script*, namely that data science can help produce and discover knowledge that can in turn be used to inform principled decision-making and action. However, just as the teacher's 'current events' script described above reflected a privileged vantage point, this normative script also imbues beliefs about the world. Specifically, while this data science script might hold true in certain situations, data-centric ways of knowing also have a long history of being used by more powerful actors to authorize, maintain, and amplify unequal power relations (versus ameliorating them) (see Scott, 1998, for an extended discussion). Moreover, even when 'everyday people' do engage in data-driven knowledge production efforts towards their own goals (e.g., farmworkers working to ban pesticides that are making them sick), their findings and their lived experiences are often dismissed as 'unscientific' or irrelevant (see Irwin, 1995, for a review). To summarize, scripts about data tend to be optimistic and ahistorical, and typically do not attend to the broader social, political and historical contexts that also underlie action and change. As such, these scripts, when invoked in the 'real world,' have the potential to reify deterministic beliefs about data, while at the same time dismissing people whose life experiences do not align with this depiction of the world.

Local Ground: Designing for a third space through digital participatory mapping

In designing for this third space ideal, Gutiérrez and Jurow (2016) argue that the activities and goals of learning, namely the means and the ends, must both be reconfigured so that students can make meaningful connections across contexts and between 'everyday' and 'scientific' forms of knowledge (Gutiérrez & Jurow, 2016; Vygotsky, 1978). We argue that participatory mapping offers one avenue to integrate these knowledge forms, allowing communities to collectively take stock of the places they live (Chambers, 2006) and locate themselves within a broader social and historical narrative. By allowing people to represent the many perspectives, critiques, and visions that they have, in relation to a place with a partially shared history and culture, the goal is to document and engage multiple knowledges and perspectives through data, towards a collectively negotiated end (Chambers, 2006). Digital mapping technologies enable a different set of place-based knowledge production activities, such as detecting patterns, sampling environmental features, and broadly disseminating findings. By bringing participatory mapping, data, and geospatial technologies together, we conjectured that we could create a context through which 'everyday' and 'scientific' knowledges could be brought into dialog, by supporting key data science activities (e.g., constructing data, conducting analysis, making inferences), and bringing the resulting analyses and representations to bear on students' local neighborhoods and communities (Enyedy & Mukhopadhyay, 2007; Rubel et al., 2017; Taylor & Hall, 2013).

To support this vision, we designed and implemented *Local Ground*, a participatory digital mapping tool (Van Wart & Parikh, 2013) that aimed to support youth bringing their own knowledge and histories of their local neighborhoods and communities into conversation with new data scientific techniques and disciplinary ideas. Central to the *Local Ground* design was supporting youths' engagement in the full range of data science activities, including defining a data protocol, collecting and analyzing data and visualizing and sharing findings. *Local Ground* was also designed to support multiple "everyday" and "scientific" forms of data, including photos, audio recording notes and drawings as well as more traditional data types such as spreadsheets and quantitative sensor data (Van Wart & Parikh, 2013). To date, several research groups have used *Local Ground* across multiple youth action research projects in the context of larger instructional designs with their own pedagogical and curricular goals.

Methods

In this article, we engage in a case study analysis (Yin, 2014) of two projects involving youth's data science activities. This analysis came out of a larger research project that involved designing mapping- and data-related software technologies to support project-based learning experiences. We began this research with a belief that data science could be an important tool for community empowerment, and that helping local community-based organizations to learn about data was a critical part of this agenda (a version of the normative script of data science). Both the first author (a white female) and third author (an Indian-American male) had experience designing and building data-related technologies for public, private, and non-profit institutions, and wanted to learn from youth educators and organizers working on the ground. The second author, a white female, had experience in design-based research and analysis. Our data sources include the first author's field notes (Lofland, Snow, Anderson, & Lofland, 2005), audio and video recordings of group activity, facilitators' digital and written instructional artifacts and student-created artifacts, such as their field notes, drawings, photographs, charts, graphs, maps and final presentations (see Table 1).

Case A: The Air Quality Project

In our first project, the first and third authors worked with a five-week summer science program to support high school students as they conducted an air quality study of the regional transit system using particulate matter sensor data (Van Wart, Lanouette, & Parikh, 2016). The program's mission was to design community-based science experiences for middle- and high-school youth living in underserved communities that could help them develop an understanding of important scientific concepts as well as an appreciation of how scientific research could contribute to addressing issues relevant to students' lives. The cohort we studied included 11 students between the ages of 15-18 years old, self-identifying as follows: Latino, Chicana, Mexican (7), African American (3) and Filipina (1) (see Visintainer, 2017, for additional analysis of this project). Students attended one of three large public urban high schools in the region, where 75-80% were eligible for free or reduced lunch. Several students enrolled in the program to make up science credits needed for high school graduation requirements. The lead instructor, Mark[2] who self-identified as a white male, had designed and taught environmental science programs for several years.

Case B: The Park Planning Project

In our second project, the first author worked with a workshop-based city planning course, taken by undergraduate and graduate students, at a large public university. The goal of the course was to engage high school and university students in a "real world" city revitalization effort, which encompassed a public housing development, the nearby park, school, and community center. Serving as mentors, university students (including the first author) worked with two 11th grade social studies classes, twice a week for 16 weeks, during their regularly scheduled class time. Approximately 60% of the high school students were identified as Latinx, 30% as Black or African American, 8% as Asian / Pacific Islander, and 1% as White / other. 80% of students were eligible for free and reduced lunch. The university students were a mixture of fifteen undergraduate and three graduate students, who self-identified as White (10), Latinx (4), and Asian (4).

Data Analysis

To analyze our data, we carried out two cycles of iterative coding (Saldaña, 2016), with the help of an undergraduate research assistant using MaxQDA, an audio and video coding software. In our initial round of coding, we attended primarily to data science activities across

[2] All names are pseudonyms.

the different phases of each project: (a) framing questions and opening activities, (b) gathering and analyzing data, and (c) sharing findings and making recommendations. We began our coding by attending to how each project was initially framed and how data was positioned within each project, using the first author's field notes and the curriculum documents. Next, we examined students' data collection and protocol design activities (e.g., developing categories, keen observation), which we derived from field notes and from students' own data and notes. Next, we coded data pertaining to students' data analysis, including student-generated artifacts (posters, notes, collages, models), transcripts of in-class presentations and discussions, students' written reflections, intermediate charts and graphs, and the first author's field notes. Across these multiple sources, we noted students' different types of data analysis (e.g., filtering, zooming, map making). Finally, we coded students' final project presentation transcripts, attending to how students communicated and marshaled data to garner support for their arguments and priorities.

From this initial analysis, it became clear that the scripts that surrounded each of the projects, and students' responses to them (including students' counterscripts) were fundamentally shaping students' experiences with data. Therefore, we did a second round of coding that attended to these scripts – noting whether any data valences (Fiore-Gartland & Neff, 2015) showed up within the projects and any counterscripts that emerged, given actors varied values, experiences, and goals. The authors and research assistant coded in parallel, and later discussed codes and themes together until they reached agreement. Examples of these codes, including scripts, counterscripts, and data practices, are grouped into three data science phases of each project and are presented in Table 2 (Air Quality project) and Table 3 (Park Planning project).

Findings

To show some of the ways in which the project scripts and data practices were emergent and mutually constitutive, we have organized our findings chronologically by case. Within each case, we focus on three broad data practices: (a) framing questions and activities – where each project was introduced to students, (b) gathering and analyzing data – where the bulk of the data science activities happened, and (c) sharing findings and making recommendations – where students reflected on their findings and what they meant. As a convention, we have bolded the focal scripts and counterscripts, some of which are paraphrased based on field notes and curriculum documents, and italicized the valences within these scripts (e.g., *discovery*, *actionability*, *truthiness*).

Case A: Air Quality Project

We begin by examining the *Air Quality* project, which followed a fairly 'traditional,' quantitative data science process, but which nonetheless revealed several of the scripts and counter scripts that might surround knowledge production efforts that are directed towards effecting change.

Framing questions and activities

The *Air Quality* project was initially conceptualized as a way to give students from underserved Bay Area communities the opportunity to participate in rigorous, hands-on science experiences that could be brought to bear on their lives and communities. To do this, the sponsoring organization partnered with a local environmental justice organization made up of residents and volunteers living in West Oakland, a predominantly working class, African American neighborhood adjacent to the Port of Oakland. Volunteers from the organization met with students during the first week of the project and described their neighborhood health advocacy initiatives and the role of data within them, including a successful effort to pass a city ordinance that banned diesel-burning semi-trucks from idling in residential neighborhoods. During these meetings, they explained that ***data can be an activist tool for effecting change***, and invited students to join their ongoing air quality monitoring efforts, lending them several scientific-grade air quality sensors to conduct a study of their own. Through this process, the organizational volunteers established a pervasive script within the project that positioned data-driven knowledge production as a way to lend authority (*truthiness*) to civic advocacy (*actionability*).

Mark, the facilitator, also introduced an additional script during the first week, which positioned data science as a tool for *discovery* that could have significant consequences for community health and well-being. As he taught students to use the sensors, he invoked the discovery valence by saying (paraphrasing): *'**Any of you can make a scientific discovery with these sensors. You never know what you'll find, or when you'll find it**.'* Mark subsequently challenged students to see what they might discover for themselves by walking around the neighborhood with the sensors. Taken together, these mutually reinforcing scripts framed the purpose of the data science activities and the potential contributions each student could make.

The *discovery* valence appeared to resonate with students from the beginning, leading to inventive sampling methodologies and attention to extreme values. For example, during their first data collection activity, where students first piloted the sensors, students went out of their way to direct the sensors towards car exhaust fumes and dust from a leaf blower to find out how various emissions sources impacted air quality (in contrast to just walking around the block as Mark had instructed). These experiments turned into a spontaneous game that continued throughout the project: Who could capture the most extreme measurement? Given their curiosity about the sensors due to the invisibility of air quality, the creativity of their sampling strategies, and their mysterious air quality spikes visible in the *Local Ground* visualizations, students seemed invested in the idea of *discovery* and what could be learned from the sensors.

[Figure 1: Map and scatter-plot view of particulate densities gathered by students during pilot sensor activity]

Gathering and analyzing data

The *discovery* valence continued to shape students' process of formalizing a sampling strategy, building directly from students' own curiosities (e.g., the impact of temperature, cloud cover, underground versus above ground, proximity to freeways, trees, and water on air quality). Students split into teams and rode the subway trains to their respective destinations, sensors in hand. By the end of their very first day of sampling, they identified what would come to be their key finding — the air in some of the underground transit stations was noticeably unhealthier than anywhere else along the transit system (see Figure 2 below). Mark cautioned that in order for this discovery to be taken seriously, it was necessary to understand what was actually causing the anomaly, and to demonstrate that their discovery was a regular, systematic pattern (*truthiness*). The overarching valence therefore shifted from one of ***discovery*** to one that emphasized ***proof and verifiability*** (*truthiness*).

Students' emphasis on proof and verification (*truthiness*) in turn motivated varied data science practices including designing and revising their data protocol, repeat-sampling, and seeking out and documenting potential causal relationships. For example, each team assigned themselves a designated note taker and photographer to document potential factors that might influence air quality. One student, Maya, meticulously documented her observations at 30-

second intervals as she rode the train, noting multiple potentially relevant variables on her notepaper, including how stuffy each train car felt and the number of passengers per car.

Sharing findings and making recommendations

The *truthiness* valence remained dominant as students worked to communicate the importance and regularity their findings to various stakeholders. To do this, students used *Local Ground* to transcribe and link their field notes (Figure 2a) to particular measurements, so as to connect their sensor data with what was happening when the sample was being collected. With the help of Mark and the first author, students also used *Local Ground* and *Tableau Public*[3] to create summary statistics, histograms, heat maps, and scatter plots (Figure 2b) to be used in various presentations.

[Figure 2(a): Students' geo-referenced, transcribed notes of their field observations]

[Figure 2(b): Scatterplot of two transit stations, showing a regular pattern of unhealthy air]

As students shared their presentations with family and friends, the *actionability* valence became salient, namely the idea that ***data can be a tool for effecting change***. However, this script was also challenged as students and audience members discussed whether and how their findings might lead to action and change. For example, as Leonard shared:

> The reason for us doing this project…is because we want to inform the people who have asthma and explain to them exactly what they're breathing in — because it can kill you and help them fix it by talking to the [transit authority]. ...For me, personally, I want to follow up on this project by having someone come every few months or whatever, and take samples of the trains, and compare to other data that we have before…and you can actually tell if they listened to what you were saying. And if it decreased, you would know that they changed

[3] https://public.tableau.com

> something about the transit system, but if it increased, that means that you have to keep buggin' them until they change it.

Leonard, who figured the transit authority would probably have to be regularly monitored ("bugged") through continued data collection and analysis, therefore tempered the idea of *actionability* with a caveat: **data can be a tool for effecting change – with persistence and other tactics.**

Following Leonard's presentation, another student, Melanie, used scatterplots (Figure 2b) to show the regularity and extent of the discrepancy between above-ground and underground stations (truthiness). "As you can see," she said, "[the underground station] had more than 20 times the particulate matter concentration as [the above ground station]" (Figure 2b). This tactic had an immediate impact, prompting an audience member, Temina, to challenge the idea of *actionability*:

> I don't know why we're measuring air quality with what the results you guys have come up with in your research, there's nothing being done about it! When I hear in the mornings, at 5 in the morning, that it's a "Spare The Air" day, I don't know what that means. Because it use to be that you didn't have to pay a fee to get on the [transit line]. And so, what good is it doing us to know!? And I think that's one of the questions that maybe tomorrow, or when you guys get to D.C., you can lobby and ask.

In Temina's view, public support for air quality protection was actually shrinking as knowledge of the issue continued to grow, building on Leonard's questioning remarks: ***Can data be a tool for effecting change?*** She also suggested that students take the transit authority to task by questioning them directly. Lelton, another student, responded to Temina by defending the *actionability* valence:

> You cannot make a change, if you don't know there's a problem, so therefore you have to recognize it, and allow others to acknowledge it, so therefore the process of coming up with a solution can be even possible.

While agreeing that data-driven evidence did not inevitably lead to change, Lelton also argued that "acknowledging" problems and spreading awareness was not insignificant. Their study was creating the preconditions necessary to begin the process of solving larger social environmental challenges.

Following the presentation to friends and family, students put the ideas of *actionability* and *truthiness* to the test by presenting to regional transit and port authority staff. According to Mark (we did not attend this presentation), students' ideas were well received by workers at the Port of Oakland. However, staff at the transit agency (who were implicated in the findings) dismissed students' findings as 'unscientific.' That said, students found other outlets for their work that were more receptive, including authoring and presenting a poster at the American Geophysical Union's (AGU) annual conference, and sharing of their findings with a national online news website.

Summary

In the Air Quality project, an early discovery of an air quality hazard (*discovery*) pushed students to examine the regularity and extent of the issue (*truthiness*) and the factors that might be causing it. These pursuits motivated numerous data science practices: revising the study design, sampling, data analysis (sorting, filtering, visualizing, and calculating summary statistics), and eventually presenting to various scientific, municipal and community audiences. Through this process, students ultimately constructed their own idea of how data might support action should the transit agency ignore their findings (*actionability*), conceptualizing it as a way to help others understand the problem, or as a vehicle to hold the transit agency accountable, should their findings be ignored.

Case B: The Park Planning Project

The Park Planners' engagement with data was primarily qualitative. While qualitative inquiries are not typically considered 'data science,' several of the project scripts and counterscripts were surprisingly similar to those that surfaced for the Air Surveyors. As such, this case offers a useful point of comparison regarding how different forms of data and different ways of positioning the value of data might influence students' data science practices.

Framing questions and activities

The organizers of the Park Planning project worked with the City Manager's Office and the Housing Authority to craft a question that young people might be uniquely positioned to answer, given the city's ongoing projects and priorities. The question focused on an ongoing revitalization effort, and asked: How could the public housing development, nearby park, school, and community center be made to feel more connected to one another? To introduce the project, Shawna, an employee of the City Manager's Office and long-time Richmond resident, met with students, explaining that the city wanted them to participate in the revitalization effort and make recommendations the city could implement (*actionability*). University mentors (including the first author) elaborated on this idea in subsequent in-class discussions, pointing out that the high school students had a wealth of local knowledge and ideas, as both young people and local residents, which were valuable but often overlooked within the planning process. Together, the mentors and the City Manager's office aimed to communicate that **students' ideas and perspectives mattered, and could help guide the revitalization effort** *(actionability*).

Some aspects of this script resonated with students, and particularly the idea that their own knowledge and experiences were a valuable asset that gave them an important perspective on the community. As one student, Carla, expressed in an early written reflection: "We have young, fresh ideas. We know the neighborhood better. We have connections there. We know what *we* want so we know what *we* need in the space."

However, many students were deeply skeptical of the projects' goals and the *actionability* valence that undergirded it, voicing several counterscripts that challenged the premise of the project. For instance, in a class discussion following the meeting with Shawna, students questioned whether anyone would care about, act on, or take their ideas and findings seriously. Students also criticized the relevance of the projects, questioning why *they* had to work on *these* problems, given the many other values and priorities they had. For example, one student wrote in an early reflection, "Why are we spending money on a random park? Why not our school?"

[Figure 3: Students question the city's priorities, given their own priorities for their community]

Gathering and analyzing data

During a subsequent field trip to the revitalization site, mentors introduced the role of data collection in the project, explaining to students that ***data could be used to systematically document and communicate needs to others (truthiness), which was necessary to effect change*** *(actionability)*. Mentors gave each student a paper map of the neighborhood and a worksheet to document the space, with prompts asking about which features of the environment made the space healthy and inviting as well as unhealthy and uninviting.

That said the site visit only gave students new material to critique the logic of the project. For instance, several students expressed in their written notes that they found it "random" or "boring" being asked to walk around and document the park and surrounding neighborhood, asking: **Why are we here?!** Moreover, students were confused about what they were supposed to document, given that nothing was particularly noteworthy about the space from their perspective: **What do we even write down?** To them, the story was clear. They were standing in an unmaintained park. What was the value of systematically documenting it (*truthiness*) when there was nothing else new to say about it?

Mentors responded to students' counterscripts by asking them to elaborate on their criticisms and record them as field notes (i.e. data). They asked students: 'What about this park makes it boring or random? What would you like to see here? What do you know about this place?' The mentors argued that there were specific, observable aspects of the space that made the park and surrounding neighborhood feel a certain way, which were important to document in order to ground their subsequent designs and to justify particular design choices to decision-makers (*truthiness*). Thus, a kind of re-mediated (Gutiérrez & Jurow, 2016) notion of truthiness emerged, which encouraged students to document their criticisms (i.e. counterscripts), interpretations, *and* observations about the surrounding natural and built environment. Both perspectives were positioned as important, valid forms of data. As a result, most students eventually participated in the activity, documented their feelings about the park (e.g., that it was "boring" and "abandoned") alongside empirical notes (e.g., "the fences are ugly" and "there are signs of alert everywhere") (see Figure 4).

[Figure 4: Planners document their qualitative data using Local Ground]

The re-mediated script, which now included a more expansive notion of *truthiness,* continued to guide students' data practices after returning from the field. During one activity, a poster-making session directed towards synthesizing and presenting their findings, students placed their ideas for the park and neighborhood (mostly drawings) alongside images (from their field notes and photographs) of things they wanted to change. For instance, one student wrote, next to a photograph of a partially intact chain linked fence: “We need places that have more trees and bushes that look nice. If we could have a place that we have to be free and not scared to go outside. And feel protected.” Another student drew a picture of a sign saying “Welcome to Richmond, City of Peace,” which he pasted next to a photo of a sign, taken by a student, that read: “NO Loitering, drinking, begging, soliciting. Subject to fine.”

Thus, students’ data analysis expanded to include additional ways of knowing – observations, personal experiences, feelings, perceptions, and so forth – to describe and symbolize their larger aspirations: feeling safe and free; being surrounded by beautiful and peaceful things; and having youth employment opportunities. Whereas data analysis had originally been positioned by mentors as a way to raise important issues to decision-makers, students’ directed their data analysis towards internal sensemaking – to collectively articulate a (partially) shared set of concerns, experiences, and hopes for the city.

[Figure 5: Sample images taken from posters created immediately following fieldwork]

Sharing findings and recommendations

As students moved from analyzing the existing park and neighborhood to envisioning a future one, data science – broadly defined as a confluence of observations, experiences, memories, and hopes – made way for design. Students’ counterscripts fell off, and students became noticeably more invested in the subsequent activities leading up to the presentation – taking on many different self-appointed roles (e.g., crafting designs, poems, speeches, digital- and paper-based data displays, and spoken word performances), and attending several optional design sessions where mentors met with students after school. Six different student teams prototyped their visions for the park and neighborhood by creating a series of physical models and maps. In presenting her team's model to the class, one student, Cindy, shared: "Our park will create jobs, it will create artistic diversity, it will be all of these things...all wrapped up into one little park." Mario, another teammate, added: "and we'll have play structures for little kids, a skate park, and a welcoming gate — so people won't be like... 'that's only for white people.'" Using *Local Ground*, these future visions were overlaid on an aerial image of the existing park (Figure 6), offering a way to re-frame the "random" space with their own beautiful vision of what it might become.

[Figure 6: Local Ground allows images to be overlaid, annotated, and combined with other data. In this example, a representation of students' ideas for a park is overlaid on satellite imagery.]

On the day of the presentation, a student, Jonah introduced the students' message at City Hall, to the City Manager's office and the Housing Authority:

> The presentations you are about to see are the product of hard work, dedication, and a true desire for change. We have investigated, we have mapped, we have analyzed, and we have discussed what [the neighborhood] looks like today and what we think it can become. We want change. We want to be heard. We are young, but we are RICHMOND—Real Intelligent Community Helpers Modeling our Neighborhood's Diversity. We are Richmond.

Following Jonah, over a dozen students presented – describing their respective visions for the neighborhood, but using data only to the extent that it advanced a particular design recommendation. Through photographs and descriptions (i.e. data) of existing challenges (*truthiness*), students suggested that the city invest in community assets — a skate park, a community center, a "graffiti wall," beautiful winding pathways, and water features. They also asked for simple things, such as trashcans, someone to cut the grass, permanent bathrooms instead of porta-potties, streetlights, and moving the methadone clinic away from the children's playground. However, while the city officials praised the students for their energy, ideas, and dedication, when the park was revitalized a year later, many of students' most innovative ideas were not implemented.

Summary

For the Park Planners, students' rejection of scripts that involved *discovery*, *actionability*, and *truthiness* led them to leverage the representational and communicative aspects of data to systematically examine a world they already knew, and to creatively explore and present alternative visions. Moreover, rather than doubling down on a cheery *make a change with data and design!* script, facilitators including the first author became more aware of and therefore open to reconfiguring the data activities to be more geared towards supporting student communication and representation of youths' perspectives. Through these tacit negotiations made possible thorough a structured-but-flexible participatory mapping process, students and facilitators were able to learn from each other. As students provided facilitators with a reality check, facilitators helped structure students' analysis and design process, and helped them to understand how their ideas were still relevant to the broader context of the neighborhood revitalization effort, thereby creating a third space.

DISCUSSION

We have presented two situated accounts of data science learning, noting the rich data science activities that young people engaged in, and the negotiations that went into making the various data science scripts meaningful in light of three prominent data valences (*discovery*, *actionability*, and *truthiness*). In this section, we revisit the efficacy of these scripts, and consider how they might be re-mediated to better account for students' lived realities amid larger structural forces.

Re-mediating scripts that invoke truthiness and actionability

As we saw in the findings, the *actionability* and *truthiness* valences were tightly interleaved in both cases: data was positioned as a way to justify and lend credibility to students' ideas and recommendations when presenting to public agencies. However, public agencies were not compelled to act on students' "data-driven" findings/recommendations in either case. For the Park Planners, although students presented compelling evidence that the park was not being adequately maintained by the city and offered many wonderful suggestions for improving it, their favorite design ideas (e.g., a community "graffiti wall," an amphitheater, and a youth-run snack bar) were not carried forward (or even seriously considered).

It might be tempting to explain this outcome by invoking the *truthiness* valence: if students' had only backed up their recommendations with more and better (quantitative) data, then perhaps they would have been more convincing to decision-makers. However, the Air Quality case challenges this idea as well. Despite engaging in repeated, quantitative sampling involving the collection of over 100,000 air samples and producing a scientific poster that was accepted as valid science by the American Geophysical Union, the Air Surveyors' findings were still dismissed by the transit agency on the basis of not being sufficiently "rigorous."

These outcomes challenge a data science script that tightly couples data-driven knowledge production (*truthiness*) with change (*actionability*), absent larger shifts in power relations. Whereas students were able to control how their data were gathered, framed, and presented, they could not control whether their findings were viewed as authoritative by public agencies (*truthiness*), nor how local resources were allocated (*actionability*), at least in the short term. This finding is important, given the presumed importance of data science as a privileged epistemic form (e.g., Irwin, 1995; Philip, Schuler-Brown, et al., 2013). From a pedagogical perspective, these cases suggest that it is important to understand the limitations of data-driven argumentation, and to use this as an opportunity to consider the broader social, political and economic contexts within which these arguments are enacted.

These cases also provide us with insight into how data might play a role in meaningful action, where "action" is more broadly construed. In the case of the Park Planners, while students were initially skeptical of *any* change narrative, they eventually found meaning in their project by examining local injustices through data collection and analysis, formulating their own visions for their neighborhoods and communities, and collectively working to voice ideas and concerns to decision-makers (a form of action). Along the same lines, the Air Surveyors eventually formulated a notion of meaningful action that entailed generating awareness (a precursor to change) through data and holding the transit agency accountable through regular monitoring. While convincing public agencies to act was not the sole purpose of students' efforts, we argue that students' interactions with the agencies motivated them to produce rigorous, compelling findings (*truthiness*), focus their message, formulate their overarching goals (*actionability*), and decide how they wanted to go about achieving them. One implication of these findings is that taking a first step towards a broader goal can also count as meaningful action. Moreover, determining what meaningful action means is something that can be discovered through tension and struggle, and can even be formulated in opposition to an initial idea that does not resonate.

Re-mediating scripts that invoke discovery

Another prominent script in both cases was the idea of data-driven *discovery*: that through data science, new insights could be generated. For the Air Surveyors, the discovery valence aligned with students ' experiences of the project. With the help of sensors and data analysis software (*Tableau* and *Local Ground*), students were able to discover a genuine air quality hazard, share this finding with local agencies and the AGU, and participate in every aspect of a very data-intensive knowledge production process. However in the Park Planning project, the project script needed significant re-mediation: students viewed the status of the revitalization site as self-evident and obvious. Whereas surprise and mystery undergirded the Air Quality project, many of the Park Planners perceived that they were being asked to discover things they already knew from their everyday experiences. However when the discovery frame transitioned to one that involved taking stock of the site and connecting observations to broader themes in students' lives (a form of *truthiness*), it appeared to become more meaningful.

Implications for data science education

In this article, we have examined how young people engaged in data science, in the context of civic advocacy, with particular attention to the role of scripts and counterscripts in shaping data science practice. We argue that while some version of a normative script may always exist in data science activities, the resonance and relevance of this script depends on a number of situated factors. Providing avenues for students to critically examine these scripts, and reorient them as needed, is an important aspect in the design of data science experiences for young people. While more research is needed to examine how these negotiations might be better supported, we highlight three aspects of our digital participatory mapping process that we believe were particularly generative:

Engaging in data collection. Our findings illustrate that many of the most consequential (Gutiérrez & Jurow, 2016) data activities – including collectively analyzing the environment, thinking systematically about gathering information, articulating hypotheses, and even negotiating the very purpose of the project – happened as students gathered and analyzed data in the field. While data collection is often conceptualized as a rote activity that precedes the "real thinking," these case studies demonstrate that being in the field supported a collaborative, embodied way to develop a collective data protocol and a common analytic framework and also hash out the larger goals of the projects.

Studying a context you already know. Our findings also suggest that digital participatory mapping and sensing can be a productive way to bring students' own local knowledge and critical perspectives to bear on a number of key data science practices and concepts (Elwood & Mitchell, 2013; Rubel et al., 2017), and wrestle with some of the value propositions and contradictions inherent in data science (Philip, Schuler-Brown, et al., 2013) — namely the efficacy of evidence-based argumentation and the idea of data-driven discovery. In doing so, we must remain cognizant that local, situated contexts are not simply sources of intuition to leverage for learning with and about data, but also complex, personal, and sometimes painful settings with long histories.

Reframing 'discovery' as systematic inquiry. These cases also show us that the data-science-as-discovery valence can certainly lead to a productive learning experience with data, as we saw with the Air Surveyors. However, systematically and collectively confirming ones' previously held knowledge and beliefs is also a powerful way to affirm one's own knowledge and perspectives, which might otherwise be dismissed (Irwin, 1995). While the Park Planners may not have discovered anything they didn't already know, similar to Enyedy & Mukhopadhyay's (2007) findings, collectively revisiting and analyzing the familiar helped them to reorient the project, develop confidence in their ideas, build solidarity, and formulate and share a collective message.

The task of situating data science often becomes one of finding generative intersections between domain-specific knowledge (e.g., science, mathematics, social science) and data-related methodologies, so that data science can be both an end and a means of learning. However, as critical data studies and learning scholars also remind us (e.g., Irwin, 1995; Philip, Schuler-Brown, et al., 2013), the value propositions inherent in data science do not always hold for people who occupy non-dominant subject positions, nor does the world solely operate on the basis of a well-reasoned argument. Therefore, we contend that providing opportunities for students and instructors to examine these scripts – through activities that are open-ended and invite re-mediation (Gutiérrez & Jurow, 2016) and where each party has the agency to shape the nature and direction of the activities – can be a powerful way to make data science learning relevant and meaningful to a diverse cross-section of learners, with the potential of generating a third space.

Acknowledgements

We are grateful for the thoughtful, incisive and generous feedback from Joseph Polman, Kris Gutiérrez, Paul Duguid, Coye Cheshire, the doctoral research and theory workshop at the UC Berkeley I School, and the three anonymous reviewers. We would also like to thank Jazmin Marie Garcia for her detailed support with data analysis. And finally, we would like to thank the young people and project organizers who allowed us to learn with them.

Funding

This work was supported by the National Science Foundation (grants IIS-01054332 and IIS-1319849), and by fellowships from the Alfred P. Sloan Foundation and the Hellman Faculty Fund.

References

Berman, F. (co-chair), Rutenbar, R. (co-chair), Christensen, H., Davidson, S., Estrin, D., Franklin, M., … Szalay, A. (2016). Realizing the potential of data science. *Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group.*

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Burke, J., Estrin, D., Hansen, M., Parker, N., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. *Presented at World Sensor Web Workshop, ACM SenSys '06, Boulder, CO*. Retrieved from https://escholarship.org/uc/item/19h777qd

Calabrese Barton, A., & Tan, E. (2019). Designing for rightful presence in STEM: The role of making present practices. *Journal of the Learning Sciences*, 1–43. https://doi.org/10.1080/10508406.2019.1591411

Chambers, R. (2006). Participatory mapping and Geographic Information Systems: Whose map? Who is empowered and who disempowered? Who gains and who loses? *The Electronic Journal of Information Systems in Developing Countries*, *25*(1), 1–11. https://doi.org/10.1002/j.1681-4835.2006.tb00163.x

Elwood, S. (2006). Beyond cooptation or resistance: Urban spatial politics, community organizations, and GIS-based spatial marratives. *Annals of the Association of American Geographers*, *96*(2), 323–341. https://doi.org/10.1111/j.1467-8306.2006.00480.x

Elwood, S., & Mitchell, K. (2013). Another politics is possible: Neogeographies, visual spatial tactics, and political formation. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *48*(4), 275–292. https://doi.org/10.3138/carto.48.4.1729

Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *Journal of the Learning Sciences*, *16*(2), 139–174. https://doi.org/10.1080/10508400701193671

Fiore-Gartland, B., & Neff, G. (2015). Communication, mediation, and the expectations of data: Data valences across health and wellness communities. *International Journal of Communication*, *9*(1), 1466–1484.

Gutiérrez, K. D., & Jurow, A. S. (2016). Social design experiments: Toward equity by design. *Journal of the Learning Sciences*, *25*(4), 565–598. https://doi.org/10.1080/10508406.2016.1204548

Gutierrez, K., Rymes, B., & Larson, J. (1995). Script, counterscript, and underlife in the classroom: James Brown versus Brown v. Board of Education. *Harvard Educational Review*, *65*(3), 445–472. https://doi.org/0017-8055/95/0800-455

Irwin, A. (1995). *Citizen science: A study of people, expertise and sustainable development*. New

York: Routledge. https://doi.org/10.5334/opt.101109

Lanouette, K., Van Wart, S., & Parikh, T. S. (2016). Supporting elementary students' learning about ecological systems through data modeling and interactive mapping in local contexts. In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Proceedings of the International Conference of the Learning Sciences (ICLS)* (Vol. 1, pp. 570–577). Singapore.

Lofland, J., Snow, D. A., Anderson, L., & Lofland, L. H. (2005). *Analyzing social settings: A guide to qualitative observation and anal* (Fourth). Belmont, CA: Wadsworth.

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

Philip, T. M., Way, W., Garcia, A. D., Schuler-Brown, S., & Navarro, O. (2013). When educators attempt to make "community" a part of classroom learning: The dangers of (mis)appropriating students' communities into schools. *Teaching and Teacher Education*, *34*, 174–183. https://doi.org/10.1016/j.tate.2013.04.011

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

Saldaña, J. (2016). *The coding manual for qualitative researchers* (Third). Thousand Oaks, CA: Sage Publications.

Scott, J. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven: Yale University Press.

Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology, Knowledge and Learning*, *18*, 65–93. https://doi.org/10.1007/s10758-013-9201-5

Van Wart, S., Lanouette, K., & Parikh, T. (2016). Local Ground: Supporting local, data-driven inquiry with youth. *Paper Presented at the Annual Meeting of the American Educational Research Association, Washington, DC.*

Van Wart, S., & Parikh, T. S. (2013). Increasing Youth and Community Agency in GIS. *Paper Presented at the Geographic Human-Computer Interaction (GeoHCI) Workshop at SIGCHI*, 1–3. Retrieved from https://www-users.cs.umn.edu/~bhecht/geohci2013ps/GeoHCI2013_Proceedings.pdf

Visintainer, T. (2017). "Scientists do what we do": Empowering Youth of Color as Learners and Doers of Community-Based Scientific Research. In *Reframing Science Teaching and Learning* (pp. 135–152). New York and London: Routledge.

Vygotsky, L. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Yin, R. K. (2014). *Case study research: Design and methods* (Fifth Edit). Thousand Oaks, CA: Sage Publications.

Tables

Table 1: Data collected for each project group

Case	Duration	Data Analyzed
Air Surveyors	5 weeks (5 days/week for 3 hours)	- Field notes - Students' photos and notes - Screenshots of data visualizations using *Local Ground* - Transcript of final presentation - Instructional materials and curriculum
Park Planners	16 weeks (2 days/week for 1 hour)	- Field notes - Participants' posters, photos, notes using *Local Ground* - Transcripts of 3 final presentations

Table 2: Air Quality Project: Scripts And Data Science Practices

	Example Scripts	Example Counterscripts	Data Science Practices
Phase 1: Set-up	Facilitator(s): 'Science is a process of gathering and analyzing data to understand human and environmental health.' **[D]** Community Partner(s): 'Data can compel decision-makers to listen **[T]** and to act.' **[A]**		Piloting the sensors, doing sampling, visualizing results, brainstorming causal relationships
Phase 2: Fieldwork	Facilitator(s): 'Look out for potential causes of particulate matter.' **[D]**		Designing and adjusting the sampling strategy, sampling, taking notes/photos
Phase 3: Data Analysis	Facilitator(s): 'Making charts and calculating summary statistics is a way to formally present scientific findings.' **[T]**		Filtering data, authoring data visualizations (histograms, scatterplots, and heat maps), calculating summary statistics, researching explanatory relationships
Phase 4: Presenting	Decision-makers: Your findings are not scientific **[T]** Student: "You cannot make a change **[A]** if you don't know there's a problem." **[D]**	Student: "Keep buggin 'em 'til they change it." **[A]**	Pamphlets, presentations, scientific posters

The valences (embedded in these scripts) are coded as follows: T = Truthiness, D = Discovery, A = Actionability

Table 3: Park Planning Project: Scripts And Data Practices

	Examples of Scripts	Examples of Counterscripts	Data Practices
Phase 1: Introducing the project	Facilitator(s): 'Civic participation involves doing research **[T]** and working across constituencies to effect change.' **[A]** Student: 'We're young, energetic, know the area, and have fresh ideas.' **[A]**	Student(s): 'Why do we have to work on a random park?' Student(s): 'They're not going to listen to us **[T]** or use our ideas' **[A]**	Students writing about their own experiences with public space (personal experience as data)
Phase 2: Doing fieldwork	Facilitator(s): 'Data helps you to understand the opportunities and challenges **[D]**, and to communicate them to others.' **[T]**	Student(s): 'We already know about this park. Why are we here?' **[D]**	Taking notes (drawing on maps, written notes), taking photos, discussing observations and what they meant
Phase 3: Data Analysis	Student: "Our park will create jobs, it will create artistic diversity, it will be all of these things all wrapped up into one little park;" **[A]**	Student: "We need a welcoming gate so people won't be like...*that's only for white people*." **[A]** Student: "We need places to feel free...and protected." **[A]**	Content analysis (via interpretive posters), class discussions of neighborhood investment vis-à-vis resource constraints; constructing new visions (informed by data and experience)
Phase 4: Presenting	Decision-makers: 'We want to hear your ideas' Decision-makers: 'we want to know that you went through a process' **[T]**	Student: "Man, who's cutting the grass?" **[A]**	Data (mostly photos) only used when arguing for a particular public investment (benches, trees, community programs, etc.)

The valences (embedded in these scripts) are coded as follows: T = Truthiness, D = Discovery, A = Actionability

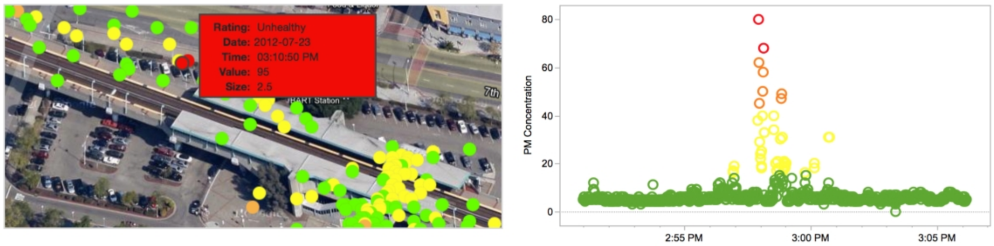

Figure 1: Map and scatter-plot view of particulate densities gathered by students

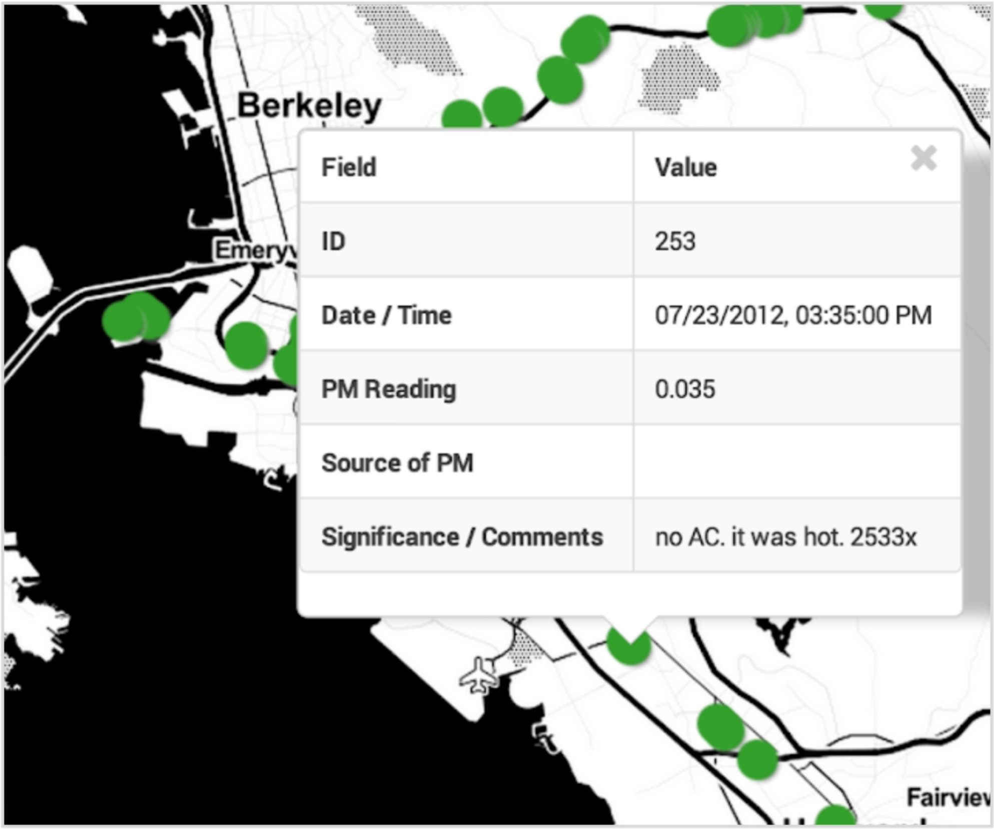

Figure 2a: Students' geo-referenced, transcribed notes of their field observations

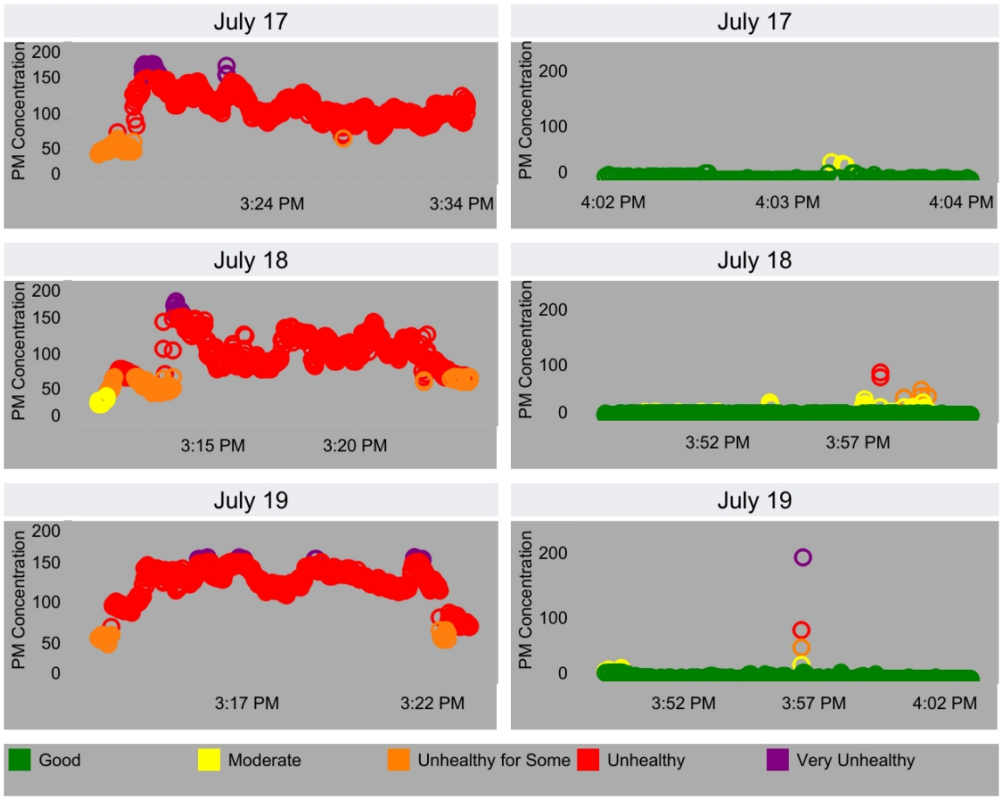

Figure 2b: Scatterplot of two transit stations, showing a regular pattern of unhealthy air

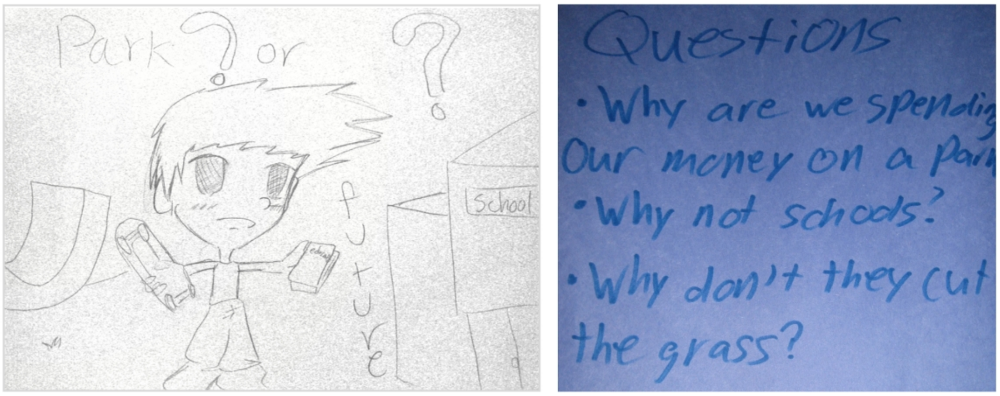

Figure 3: Students question the city's priorities, given their own priorities for their community

Figure 4: Planners document their qualitative data using MapTool

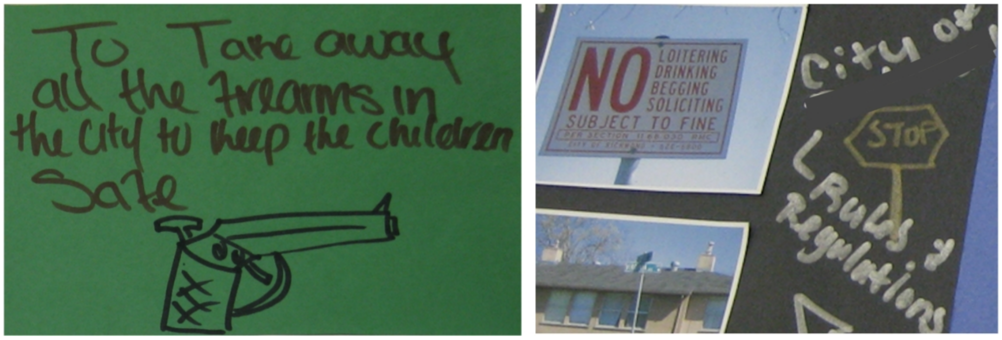

Figure 5: Sample images taken from posters created immediately following fieldwork

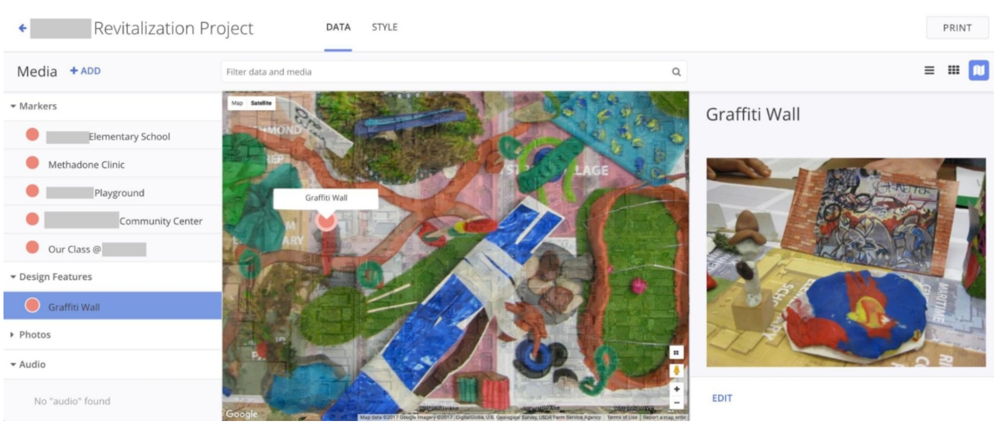

Figure 6: MapTool allows images to be overlaid, annotated, and combined with other data. In this example, a representation of students' ideas for a park is overlaid on satellite imagery.