**Situating Data Science: Lenses on Learning and Meaning Making with Data**

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Journal of the Learning Sciences</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>HLNS-2017-0085.R1</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Special Issue Proposal</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Other &lt; Content Areas, Middle/High School &lt; Age of Learners, Informal Learning and Learning Environments &lt; Topics, Other &lt; Topics, Adult &lt; Age of Learners</td>
</tr>
</tbody>
</table>

URL: http://mc.manuscriptcentral.com/jls  Email: journaloflearningsciences@gmail.com
Situating Data Science:
Lenses on Learning and Meaning Making with Data

The emerging field of Data Science has had a large impact on science and society. This has led to over a decade of calls to establish a corresponding field of Data Science Education (Berman et al., 2016; Cleveland, 2001; Finzer, 2013). Data Science, the argument goes, prepares students for high-paying jobs, prepares society for scientific advancement, and provides communities with new tools for expression and empowerment. There is still a need, however, to more deeply conceptualize what a field of Data Science Education might entail. In particular, it is important to understand what makes learning Data Science sufficiently different that it requires a new field of study, and to explore the theoretical and practical implications of these differences for constructing an ethical and effective Data Science Education.

This special issue will explore one key feature of Data Science that we argue has serious implications for education and research. In contrast to data constructed to answer particular questions, Data Science is concerned with data collected in an incidental or automated manner—often from activities and contexts within which learners themselves are deeply situated. Emerging research suggests this can significantly impact how learners engage with and make sense of these data, at times in ways not recognized by educators or designers (Philip, Olivares-Pasillas, & Rocha, 2016; Rubel, Lim, Hall-Wieckert, & Sullivan, 2016). Our goal is to explore the variety of ways that learners’ situatedness—relative to data, and relative to the field of data science—can impact learning in ways that necessitate new lines of research, new theoretical and methodological development, and new approaches to educational design and practice.

Data Science and Data Science Education

Colloquially, the term data science refers to the use of computational tools and methods to collect, process, analyze, store, and visualize large quantities of data. It is broadly associated with the emergence of a growing number of data visualizations, open data repositories, and infographics intended for public consumption (McGhee, 2010), as well as shifts in professional disciplines, from the sciences to the arts, as they integrate computational data work into their everyday activities (Hey, Tansley, & Tolle, 2009). Data Science as a formal field of study, however, is still not well defined; the diversity of perspectives regarding its identity has led some to avoid defining it altogether (Cassel & Topi, 2015). Instead of attempting to create our own definition, we highlight some distinguishing characteristics that emerge as points of agreement in discussions of Data Science and Data Science Education (Baumer, 2015; Berman et al., 2016; De Veaux, et al., 2017; Donoho, 2015; Hardin et al., 2015) and that motivate our current work.

One of the most commonly cited characteristics of Data Science is that it is concerned with a new class of data that are not only “big” in the traditional sense of scale, but “…pervasive, tacit, and often collected without a specific intent” (Cassel & Topi, 2015, p. 10)—for instance, social and environmental information collected through passive, automated recording. It is the pervasive and connected nature of data, rather than merely its size, which is of note (Donoho, 2015). Common examples of such “big data” include social network and clickstream data from popular websites such as Facebook; large-scale civic data about legislation, policy, voting trends, and municipal demographics; or, weather and climate benchmark data collected multiple times per day from satellites, radars, ships, weather stations, tide gauges, and other automated devices. Such datasets are often public, encompassing in scope, and reflect tacit features of our shared world or details about ourselves and our behavior.
Such data can be exploited (Donoho, 2015) for a variety of purposes. Individual preferences can be inferred from website interaction data to target advertisements, and predictive climate models can be constructed from environmental data to aid in policy development and city planning. But to exploit data in this manner requires a great deal of exploratory work, post-hoc transformation, and automated analysis. And, in addition to computational and statistical knowledge, transforming large collections of data for a particular application or domain requires deep knowledge of both the target domain, and the original context in which the data were collected. As a result, other defining characteristics of Data Science include that it is intensely computational, and increasingly interdisciplinary. New tools, techniques, and statistical methods that incorporate and innovate on methods from Statistics, Computer Science, and other disciplines to visualize, store, organize, process, and interpret data are constantly in development.

The influence of these drivers—proliferation of pervasive, exploitable data, new computational tools and methods for working with such data, and the increasingly interdisciplinary nature of data work—are reflected in the calls to establish Data Science Education as a formal field of study (Berman et al., 2016; Cleveland, 2001; Finzer, 2013). New programs of study, primarily in higher education, focus on helping students learn to “think with data” (Baumer, 2015; Hardin et al., 2015), learn computational methods for working with and communicating about data (Nolan & Temple Lang, 2010), and develop data-related skills for science and industry. Others focus on developing data science competencies as part of the civic and information literacies needed to navigate a data-rich world (Bergstrom & West, 2017; Grawe, 2011).

Recommendations emerging from these efforts (e.g., De Veaux, et al. 2017) highlight two core pedagogical commitments required for effective Data Science Education. The first is that educators must balance technical skill in key areas (programming languages such as R or Python, data storage and manipulation techniques, statistical methods and machine learning) with flexibility so that students can learn and develop their own future tools and methods for working with data. The second is that Data Science must be grounded in specific, consequential investigations. To this end, the Data Science Educators advocate for an expanded view or full “cycle” of data analysis that includes posing questions, obtaining data, and communicating findings in meaningful disciplinary contexts. Some go so far as to suggest that Data Science courses should always be offered concurrent to, and connected with, content courses in relevant disciplines.

The Need for a Situated Perspective

Conceptualizations of Data Science Education such as those described above, while certainly productive, have focused more on identifying required curricular targets than identifying existing learner knowledge and experiences of data and data science. However, given the ubiquity of data and the popularization of data science, the potential personal and societal relevance of large datasets, and the increasing need for analysts to understand and coordinate multiple contexts in which data are collected and exploited, we expect learners’ own relationships to data to significantly shape how they engage in Data Science.

This has become particularly apparent in recent efforts to introduce Data Science Education at pre-collegiate levels (Gould, Machado, Ong, Johnson, & Molyneux, 2016).

---

1 See UC Berkeley’s “Data 8 Connector Program” for one example; http://data8.org/connector/
Whereas undergraduate programs situate data science within students’ *disciplinarily* relevant areas of study, precollegiate efforts have sought to do the same through the use of *socially* relevant datasets. For example, they may explore data about their local communities, pop culture, or social networks; or, they may collect and work with data from personal mobile devices or ambient environmental sensors. The assumption driving these designs is that using data relevant to, or collected from, students’ everyday experiences would make such explorations meaningful to students, allowing them to reason deeply about the contexts in which data were collected and what patterns in those data may reveal.

Thus far some such efforts have proved problematic, or have led to unintended consequences. In one case, the design and enactment of spatially-grounded examinations of the lottery and local alternative banking institutions originally developed to teach students statistics were found to reproduce deficit narratives about communities of color (Rubel, Hall-Wiekert, & Lim, 2016). In a study to explore how new features of the Scratch programming language could introduce young learners to Data Science, designers were surprised when children warned that allowing code that takes others’ user statistics as input could create exclusionary programs that only “popular” programmers could access (Hautea, Dasgupta, & Hill, 2017). In another case, high school students exploring Netflix data as part of a data science course attempted to reason about racial and economic dimensions of movie production and marketing, but this engagement was not supported by the teacher (Philip et al., 2016). At the same time, other activities that invite learners to explicitly situate themselves within, and tell their own stories through, available data have yielded rich engagement in analysis and manipulation of data, even at short time scales (Kahn, 2017; Wilkerson & Laina, 2015; Under Review).

These emerging findings illustrate just a few ways in which learners’ relationships to data influence their engagement with data, and are likely consequential to their learning. This is no surprise to Learning Scientists, who have long studied the ways that learning is shaped not only by curriculum and instruction by also by learners’ experiences, tools and materials, environments, communities, and position within social and political structures—all with important implications for educational practice and theory (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). It is for these reasons we dedicate this issue to situated perspectives toward learning Data Science. We use the term *situated* in its broadest sense, to refer to a collection of approaches toward learning, cognition, and participation (Roth & Jornet, 2013) that we expect can grant new insight into the complex territory marked by the emerging field of Data Science Education. Some progress has already been made, for example, in exploring the role of embodiment and mobility in learners’ reasoning with and about data, leading to the development of successful interventions that engage learners deeply with data analysis cycles in both formal and informal settings (Kahn, 2017; Lee, 2013; Taylor & Hall, 2013).

We envision a number of other ways in which situated perspectives can enrich the study and practice of Data Science Education. For instance, learners leave “digital traces” (Latour, 2007) everywhere to be found in data: recorded on transit passes, documented through clicks and keystrokes on social media sites, or logged by data-rich learning technologies (Pardo, 2017). These traces reflect different experiential trajectories through the same data context, and may be productively leveraged to support conversation about, and negotiation of, perspectival understandings (Greeno & Van De Sande, 2007) of a dataset that subsequently influence how the data are analyzed and interpreted. Sociopolitical perspectives (Philip, Jurow, Vossoughi, Bang, & Zavala, 2017) can shed light on how learners navigate the codification of personal, demographic, and political data, and the ways in which data reflect or conflict with lived experience. And, much can be learnt from attending to how learners frame and identify with
Data Science, both as a field and as a pursuit (Bell, Van Horne, & Cheng, 2017; Hammer, Elby, Scherr, & Redish, 2005; Hand, Penuel, & Gutiérrez, 2013; Polman & Miller, 2010), at a time when internet users are invited to become casual data scientists through interactive visualizations and data explorer tools, and graduate students are apprenticed as practitioners of a field that is ill-defined, well-funded, and ethically fraught.

Rather than constraining our focus further before having a good sense of the final papers to be included in the issue, we intend to keep our call for submissions relatively broad. We expect that the final, compiled issue will illuminate a variety of ways in which learners’ relationships to data influence their understanding of and engagement with both the data, and with data science as a field. Our review of those contributions may replace or supplement the examples provided above. This may include, for example, how learners’ spontaneous self-to-data perspective taking facilitated reasoning with data in museum contexts; how learners’ histories and the histories of their communities influence if data science is viewed as a means to construct evidence; whether the ways in which data resonate (or not) with students’ experiences inform how they analyze, represent, and question those data; or how teachers, researchers, or others expected to become data science practitioners understand and identify with the field.

Contributions of the Special Issue

The goal of Situating Data Science is to better conceptualize what a Data Science Education might be, given learners’ existing situatedness relative to the complex data with which they are expected to engage, and relative to a quickly growing set of communities, including citizens and scientists, who leverage data science practices in their work. Given the field’s central concern with large, pervasive datasets and contexts for which learners are likely to have experience, we specifically ask: In what ways are learners’ ways of engaging with and making sense of data affected by their situatedness: relative to data, data contexts, and data science as a field? And, How might learners’ prior experiences with and relationships to data equip them for formal and structured Data Science Education experiences?

The issue will begin with a historical review of Data Science Education as it relates to over a half-century of research on learning with data and learning in statistics (Rubin, expected contribution). This review will describe decades of research in statistics and mathematics education that has engaged students with rich, relevant datasets that can powerfully inform current Data Science Education efforts. It will also highlight several ways in which the emerging field of Data Science departs from, and thus requires reconceptualization of, some core principles underlying those early efforts. For example, Rubin describes how the statistics educators’ strong focus on sample representativeness and inference is challenged by an age when data about entire populations are readily available (see also Ainley, Gould, & Pratt, 2015). However, other features of data, such as the assumptions that underlie what is measured and how, become more critical.

The issue will then present a collection of articles examining how learners’ work with and experiences of data about their communities, families, and experiences across museum, after-school, school, and home contexts reveal a wealth of existing knowledge and emergent supports and practices for making sense of data that can powerfully inform the foundations of a more formal Data Science Education. Data Science is often described in emerging policy and consensus documents as a highly novel set of tools and practices, or as an eclectic combination of fields with which learners are likely to have had little experience with outside of formal, collegiate instruction. We expect this special issue to show, however, that learners in fact have extensive and rich knowledge of and experiences with data and data science practices.
Additionally, the special issue will offer examples that demonstrate these experiences can indeed be productively leveraged in formal and structured learning experiences to advance both education and career-oriented goals as well as ways of thinking relevant to everyday life and democratic participation.

“Situating Data Science” is particularly well-suited for the Journal of Learning Sciences. The Aims and Scope of JLS highlight the value of “…interdisciplinary and methodological innovation; grounding research in real-world contexts; answering questions about learning process and mechanism, alongside outcomes; pursuing technological and pedagogical innovation; and maintaining a strong connection between research and practice.” Aligned with these aims, the special issue would offer a series of early, yet in-depth explorations of learners’ engagement with Data Science as an emerging, interdisciplinary field. It would do so through explorations situated in a diverse variety of contexts, in each case examining learners’ positions as already embedded within authentic, data-rich contexts and examining the effects of such positioning on learning and participation. In this way, we expect the issue to serve as a learner-experience centered contribution to the existing discourse about Data Science Education, which, until now, has focused primarily on how particular tools and procedures of Data Science should be taught.

Acknowledgements

Many of the ideas discussed in this introduction emerged from collective discussion at the 2016 Youth, Learning, and Data Science Summit. This small workshop was funded by National Science Foundation Cyberlearning capacity building grant IIS-1541676. We are grateful for the very helpful feedback on this introduction provided by several planned contributors, and several anonymous reviewers.
Call for Proposals

The emerging field of Data Science has had a large impact on science and society. This has led to over a decade of calls to establish a corresponding field of Data Science Education (Berman et al., 2016; Cleveland, 2001; Finzer, 2013). Initial efforts to do so, while productive, have focused primarily on curricular structures and materials rather than learner knowledge and experience. There is still a need to more deeply conceptualize what makes learning Data Science sufficiently different that it requires a new field of study, and to explore the theoretical and practical implications of these differences for constructing an ethical and effective Data Science Education.

This special issue will explore one key feature of Data Science that we argue has serious implications for education and research: learners’ relationships to data. In contrast to Statistics or Science Education, Data Science Education is typically concerned with data collected in an incidental or automated manner, often from activities and contexts within which learners themselves are deeply situated. Emerging research suggests this can significantly impact how learners engage with and make sense of these data—limiting students’ opportunities to learn when these relationships are not recognized by educators or designers (Philip et al., 2016; Rubel, Lim, et al., 2016) and enriching them when they are leveraged (Kahn, 2017; Lee, 2013; Taylor & Hall, 2013). The goal of Situating Data Science is to sketch the contours of what a Data Science Education might entail given these relationships. More specifically, we ask: In what ways are learners’ ways of engaging with and making sense of data affected by their situatedness: relative to the data themselves, the data contexts from which they are derived, and the tools and practices of data science that learners are likely to come into contact with? And, How might learners’ prior experiences with and relationships to data equip them for formal and structured Data Science Education experiences?

We invite contributions that explore how learners’ situatedness—relative to data, and relative to the field of data science—can impact learning in ways that necessitate new lines of research, new theoretical and methodological development, and new approaches to educational design and practice. We use the term situated (Brown et al., 1989; Lave & Wenger, 1991) in its broadest sense, to refer to a collection of approaches toward learning, cognition, and participation that we expect can grant new insight into the complex territory marked by the emerging fields of Data Science and Data Science Education. Papers may focus on questions like:

- How do learners’ different experiences of the same data context affect their engagement with data, and how might this diversity be leveraged pedagogically?
- How do different framings of data science activity goals (e.g., using data for commercial exploitation, predictive modeling, advocacy, self-monitoring, scientific inquiry) influence learners’ engagement with and treatment of data?
- How are current data scientists, such as practitioners of learning analytics or educational data mining, apprenticed into the discipline? In what ways do such apprenticeships leverage (or not) learners’ own data or learning experiences?
- What might be fruitful theoretical and methodological approaches for uncovering orientations toward or experiences with data that are likely to be especially powerful for supporting formal Data Science Education?
- How can insights about learners’ experiences with the types of datasets, tools, and methods characteristic of Data Science in informal (e.g., home, online, museum,
hobbyist, advocacy) contexts inform the design of formal Data Science Education experiences?

Submission Instructions: We are currently soliciting abstracts for proposed papers for the special issue. Abstracts should be no longer than 500 words and be accompanied by up to six keywords.
[information on deadline will be added to the CfP upon acceptance of this proposal; proposed timeline is below]
Expected Audience

We have reason to believe there will be a broad audience for a special issue on Data Science Education. For the past several years, there have been public calls for a focused and intentional community-building effort around Data Science Education. Appeals to draw together interested parties were made at the inaugural Cyberlearning Summit Meeting (Finzer & Konold, 2012) and at the 2013 Cyberlearning Synthesis and Envisioning meeting (Dorsey, 2013). In the past year, researchers and educators have come together at two NSF-sponsored small conferences, both in Berkeley California: the Youth, Learning, and Data Science Summit in August 2016 (YDS16; https://www.ocf.berkeley.edu/~mwilkers/yds2016/), and the Data Science Education Technology Conference in February 2017 (http://codap.concord.org/dset/). The former drew 30 scholars and educators and engendered this proposed special issue. The latter drew over 100 scholars, educators, and software/curriculum developers and led to the establishment of a Data Science Education Webinar series (concord.org/meetup#webinar). Below, we list 12 notable Learning Scientists who are likely to be interested in the special issue we are proposing.

- Dani Ben Zvi, University of Haifa
- Cynthia Carter Ching, UC Davis
- Daniel Edelson, BSCS
- Noel Enyedy, UCLA
- Bill Finzer, Concord Consortium
- Rogers Hall, Vanderbilt University
- Cliff Konold, University of Massachusetts
- Thomas Philip, UCLA
- Catherine Cramer, New York Hall of Science
- Jim Hammerman, TERC
- Fred Martin, University of Massachusetts, Lowell
- Richard Lehrer, Vanderbilt University

We have also included with this proposal the attendee list for YDS16, which reflects broad interest from a number of related fields (e.g., Computer Science, Human-Computer Interaction, Math Education, Science Education, Youth Outreach and Engagement).
Overview of the Special Issue

We have consulted with several potential contributors to this Special Issue, and list all of our expected contributors below. All of these papers touch on themes central to Data Science Education, including youth’s relationships to data across formal, informal, and familial contexts, and the use of new technologies for collecting, exploring, and communicating about data. Each paper will contribute examples of ways in which youth are already working with and making meaning of data that is directly connected to their everyday lives. We have also included above proposed text for an open call for submissions. If this proposal is accepted, the special issue editors will solicit submissions from the authors noted below and the overall learning sciences community (through ISLS and AERA's SIG-Learning Sciences listserves, and through social media), then select a final set of six, and encourage those not selected to submit to JLS or another appropriate venue separately.

Data Science Education: How Did We Get Here, and What Do We Know?
Andee Rubin
While data science (and, therefore, data science education) is often presented as a new field, the field of statistics education research has been active for several decades. An international community of researchers has learned a considerable amount about how youth reason with data, how technology can support more sophisticated statistical reasoning and what core statistical ideas are particularly difficult for people to master. This article will describe a series of projects, beginning with a non-technological elementary curriculum development effort in the late 1980’s and proceeding through technology-rich projects of the recent past, identifying the lessons and insights gleaned from each. It will then consider both how the emerging field of Data Science education can build on this knowledge and what new challenges and issues it introduces to the learning sciences.

Engaging the Socio-Political Dimensions of Data Science Using Local Participatory Mapping
Sarah Van Wart, Kathryn Lanouette, and Tapan Parikh
While data science education has emphasized modeling, prediction, and pattern-finding, less attention has been given to its discursive and communicative aspects, including how data might be used (or not used) to justify, challenge, bolster, and/or reveal young people’s perspectives about people and places. Using examples from several design experiments, this article considers how high school students and teachers integrated empirical data methodologies into a variety of local research and advocacy efforts, in ways that honored students’ experience, memory, and imagination. Drawing on Gutiérrez’s idea of a sociocritical literacy (2008), we describe some of the ways in which students' histories of experience, and those of their communities, shaped their participation in these activities — from their analytical approaches, to their communication strategies, to their faith (or lack thereof) in the power of a well-reasoned, scientific argument (Enyedy & Mukhopadhyay, 2007). We argue that the normative ethos of data science (boyd & Crawford, 2012), in terms of its stated goal of informing more objective, accurate, evidence-based decision-making, may not always be shared by those whose truths have been historically discounted. By attending to some of the ways in which teachers and students negotiated these tensions, we can better connect, support, and situate data science — a powerful academic literacy and way of knowing — to local social issues and concerns that have a profound impact young people’s lives.

URL: http://mc.manuscriptcentral.com/jls  Email: journaloflearningsciences@gmail.com
Engagement with Data through Fluid and Flexible Self-to-Data Relationships in Spontaneous Dialogue
Jessica Roberts and Leilah Lyons

One of the known challenges in teaching Data Science is getting learners to engage with a complex and unfamiliar dataset. In particular, we know very little about how engagement with data manifests in out-of-school settings like museums. One promising marker of engagement with data can be found in the perspective a learner takes during a collaborative exploratory activity. Prior work on perspective taking has shown that taking on a first-person “Actor” perspective (e.g., Lindgren, 2012) or blending perspectives between first-person and third-person (Enyedy, Danish, & DeLiem, 2013; Ochs, Gonzales, & Jacoby, 1996) can support learners in a variety of reasoning and learning tasks. This paper extends that work by examining the spontaneous dialogue of visitors at an interactive museum exhibit, finding that not only do peer groups productively employ Actor perspective taking (APT) as a tool to engage in sensemaking, they do so through the construction of a variety of self-to-data relationships we identified and codified from their naturalistic talk. Moreover, the groups engaging in flexible usage of APT—that is, fluidly utilizing multiple self-to-data relationship blends—showed overall more productive learning talk during their exhibit interactions. Deepening our understanding of how novice learners naturally relate to data in informal settings can provide insight for how to support youths’ comprehension of data “in the wild”, and provide suggestions for how to capitalize on these relationships for teaching Data Science in multiple contexts.

Youth Type One Diabetes Management as a Context for Data Engagements
Victor Lee

Drawing from Distributed Cognition (Hutchins, 2014), this paper considers chronic disease management associated with Type 1 diabetes as a setting in which youth and families must continuously learn and teach one another about data through social coordination and use of various digital and analog artifacts. The paper considers three groups affected by diabetes data management: the child with the condition, the immediate family, and extended community such as schools, peers, and neighbors. In a sense, this could be seen as different systems in Bronfenbrenner’s ecological systems model and interactions between a child, a microsystem (family), and a mesosystem (the child’s caretaking community). For affected youth, the learning involves continuous numeracy encounters with parents creating new artifacts or engaging in spontaneous computational conversations with children to encourage them to practice recognizing quantities that will affect their health. For parents, there are encounters with data streams that can lead to different visual appraisals or methods of analysis. They learn to think about anomalies in data visualizations in new ways and create plausible stories for those anomalies to help them determine next actions. In one exceptional case, a family developed a computational model to try to understand how different foods affected their child. Finally, for those outside the family, the actions taken often involve actively educating school personnel and peers, which can take the form of science fair projects that a youth had pursued and made visible to his community or more formal invocation of special education laws and ongoing daily communication with teachers to make sure that various quantities are discussed, considered, and acted upon when a parent was not around. The contribution of this work to learning sciences is its illustration of learning that is initiated by biological circumstance and how that impacts and creates learning opportunities for multiple parties in larger encompassing systems. It also, similar
to work on informal mathematics and informal science knowledge research previously published in *JLS* (e.g., Azevedo, 2013; Nasir, 2005; E. V. Taylor, 2009), adds to our understanding of knowing-in-practice and the distinct character of data use in home settings.

**Stories of Our City: Student Engagement With Data through Visualization Design**

*Michelle Wilkerson and Vasiliki Laina*

“Telling stories with data” (Segel & Heer, 2010) is celebrated as a key element, if not the core, of Data Science. Research has shown that youth have a diversity of (counter)stories about data they encounter in the classroom (Enyedy & Mukhopadhyay, 2007; Philip et al., 2016; Rubel, Hall-Wiekert, et al., 2016). This study builds on such work to explore the tensions and opportunities that arise when students are invited to tell their own stories about publically available data through visualization design. Over two weeks, seventh grade students in a mathematics classroom analyzed and created visualizations of demographic data about their city. We focus on two student groups that explored a dataset about the city’s racial composition, but whose analyses and productions told very different stories. One group created a line graph that highlighted a consistent decrease in the proportion of the population identified as White, with relative increases in the proportions of the population identified as Black, Asian, and Other racial categories. This group celebrated that the city was becoming more racially diverse, and noted similarities to their own classroom. Another group created a pictogram with stick-figures of different colors to represent the relative proportion of the population identified as each racial category. They argued that though the city was often described as tolerant and diverse, an overwhelming majority of the population was White. We leverage microgenetic learning analysis (Parnafes & diSessa, 2013) to trace students’ construction of these data stories as they work to coordinate resources from statistics, representation, and their own experiences of the city. We also examine how putting these students’ stories into conversation deepened students’ engagement in statistical reasoning, and led to classwide inquiry about the design and purpose of the “race” category on the U. S. Census.

**Data Science for The Rest of Us: Data to Help Our Community**

*Clifford Lee & Elisabeth Soep*

Mood Ring is a mobile app developed by the Youth Radio Interactive (YRI) team of youth and adult staff. Terrell and Nancy*, two of the original teenage developers, recognized an exigent need amongst their peers to reach out to trusted friends and family members when they felt heightened and sustained emotions of anger, frustration, anxiety, and depression. They noticed this growing phenomenon affecting many in their social circles in significant ways. As youth of color growing up in the flatlands of Oakland, many have to contend with the effects of institutional racism, intergenerational poverty, underfunded schools, and other systems of oppression. Internalizing these structural inequalities with few positive outlets, some turned to activities that led to a further downward spiral.

Through extensive brainstorms, user and market research, as well as interviews with professional mental health clinicians, Terrell and Nancy, along with the YRI team developed a mobile app that allowed users to keep track and share data of their emotions and feelings through a set of nine emojis and written reflections to a pre-selected group of people. Their research led them to many of the final functionalities of their app. What is particularly unique and different from traditional Data Science approaches is that these teens created a platform to collect, process, and represent quantitative (frequency of emojis) and qualitative data (narrative discussion of feelings/ emotions) for a select audience to take action when necessary. By creating
an app to address deeply personal and often private issues in their lives, these teen developers provided a real-world solution in their immediate communities. Both the instructional approach and the skills shown by the creators of this app demonstrated Critical Computational Literacy or a macro sociopolitical awareness of their audience and society while utilizing the computational thinking skills necessary to create the mobile app.

This case study demonstrates the capacity for a more expansive notion of what constitutes data and what the field of Data Science should include, especially when we involve youth as creators and developers of tools to serve needs unsupported by current infrastructures. From our work at Youth Radio, we argue that teaching data science must begin with an informed and critical examination of the world (Friere, 1970). This provides the foundation for developing tools for change that does not simply reify the inequitable conditions many historically marginalized groups face.

*Pseudonyms

**Invited Commentary.** We will invite commentary from individuals involved in large-scale data science efforts at the high school and undergraduate levels, such as Dani Ben-Zvi (The University of Haifa, Israel; long-time researcher of Exploratory Data Analysis in K-12 education); Alyssa Wise (leader in Learning Analytics techniques); Rob Gould (UCLA; Mobilize project for Data Science in high schools), Deborah Nolan (UC Berkeley; at the forefront of Data Science Education at the undergraduate level), David Donoho (Stanford University; author of “50 years of data science”) and David Custer (Decatur High School/Editorial Board, *The Mathematics Teacher*).
Proposed Timeline

We have attracted contributors to this Special Issue from the pool of individuals that participated in the initial NSF-funded Youth, Learning, and Data Science Summit held at UC Berkeley. Therefore, we expect initial invitations to proceed quickly; indeed, as soon as possible. We intend to invite abstracts several months before first drafts of manuscripts are due, with hopes that this will streamline the identification and recruitment of potential reviewers. The current proposed timeline allows for a full year of review and revision; this should be sufficient given the average recent review times at *JLS*.

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts Due</td>
<td>Feb 15, 2018</td>
</tr>
<tr>
<td>Invitations to Submit Drafts Sent</td>
<td>Feb 22, 2018</td>
</tr>
<tr>
<td>First Drafts Due</td>
<td>March 15, 2018</td>
</tr>
<tr>
<td>Reviews of First Drafts</td>
<td>June 15, 2018</td>
</tr>
<tr>
<td>Ongoing Revisions</td>
<td>July-December 2018</td>
</tr>
<tr>
<td>Final Manuscript Submitted</td>
<td>January 31, 2019</td>
</tr>
<tr>
<td>Publication</td>
<td>Spring 2019</td>
</tr>
</tbody>
</table>
References


Berman, F., Rutenbar, R., 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 Informatin Science and Engineering Advisory Committee Data Science Working Group.*


