Information Design Journal 23(1), **6–18**

DOI: 10.1075/idj.23.1.03dig

Catherine D'Ignazio

Creative data literacy

Bridging the gap between the data-haves and data-have nots

Keywords: data literacy, empowerment, data visualization, inequality

Working with data is an increasingly powerful way of making knowledge claims about the world. There is, however, a growing gap between those who can work effectively with data and those who cannot. Because it is state and corporate actors who possess the resources to collect, store and analyze data, individuals (e.g., citizens, community members, professionals) are more likely to be the subjects of data than to use data for civic purposes. There is a strong case to be made for cultivating data literacy for people in non-technical fields as one way of bridging this gap. Literacy, following the model of *popular education* proposed by Paulo Freire, requires not only the acquisition of technical skills but also the emancipation achieved through the literacy process. This article proposes the term *creative data literacy* to refer to the fact that non-technical learners may need pathways towards data which do not come from technical fields. Here I offer five tactics to cultivate creative data literacy for empowerment. They are grounded in my experience as a data literacy researcher, educator and software developer. Each tactic is explained and introduced with examples. I assert that working towards creative data literacy is not only the work of educators but also of data creators, data publishers, tool developers, tool and visualization designers, tutorial authors, government, community organizers and artists.

> "*The future is already here. It's just not very evenly distributed.*"
> – William Gibson

1. The problem: Data inequality

Despite the grand hype around "Big Data" and the knowledge revolution it will create (Schönberger & Cukier 2013), there is profound inequality between those who are benefitting from the storage, collection and analysis of data and those who are not (Andrejevic 2014; boyd & Crawford 2012; Tufekci 2014). Data has become a currency of power. Decisions of public import, ranging from which products to market, to which prisoners to parole and which city buildings to inspect, are increasingly being made by automated systems sifting through large amounts of data (Pasquale 2015). As a result, knowing how to collect, find, analyze, and communicate with data is of increasing importance in society.

Yet, ownership of data is largely centralized, mostly collected and stored by corporations and governments. Critically, the technical knowledge of how to work effectively with data is in the hands of a small class of specialists. People are far more likely to be discriminated against with data or surveilled with data than they are to use data for their own civic ends (O'Neil 2016). This has implications on how people do social science (Crawford et al. 2014; Sandvig et al. 2014; Welles 2014), practice law (Pasquale 2015), produce policy (Goldsmith & Crawford 2014), govern the city (Jacobs et al 2016) and create the news (Diakopoulos 2015; Kirchner 2016; D'Ignazio & Bhargava 2015), among other things.

The scholarship of Critical Data Studies (Dalton, Taylor & Thatcher 2016) has focused on algorithmic transparency, data discrimination and privacy concerns. There has been, however, comparatively less effort on issues of equity in terms of *who* has access to the computing power and know-how to be able to make sense of data and how they come to acquire and deploy that knowledge. Mark Andrejevic has termed this the "Big Data Divide" (Andrejevic 2014) and Boyd and Crawford have referred to data-haves and have-nots (Boyd & Crawford 2012). Crawford has written eloquently on "Artificial Intelligence's White Guy Problem" (Crawford 2016). Certainly, the fact that there are equity and inclusion issues in data science is not surprising given the persistence of digital inequality (DiMaggio & Hargittai 2001) and the lack of women and minorities in STEM fields (Neuhauser 2015). Cultivating data literacy in a more diverse population should therefore be part of any solution or mitigating strategy for data inequality.

2. Creative data literacy

Data literacy includes the ability to read, work with, analyze, and argue with data as part of a broader process of inquiry into the world (D'Ignazio & Bhargava 2016; Letouzé et al. 2015). The popular press has argued for broad data literacy education (Harris 2012; Maycotte 2014). Workshops for nonprofits and activists throughout the world are introducing tools and practices that can help use data to advocate for social change (Tygel & Kirsch 2015; Emerson & Tactical Tech 2013). However, there is a lack of consistent and appropriate approaches for helping novices learn to "speak data" (Bhargava 2014). Some approach the topic from a math—and statistics-centric point of view (Maine 2015). Some build custom tools to support intentionally designed activities based on strong pedagogical imperatives (Williams, Deahl, Rubel & Lim 2015). Still others have brought together diverse communities of interested parties to build documentation, trainings, and other shared resources in an effort to propagate the "open data movement" (Gray 2012). Regrettably, data literacy has been relegated to a set of technical skills, such as reading charts and making graphs, rather than connecting those skills to broader concepts of citizenship and empowerment. Drawing from Paulo Freire's popular education, literacy involves not just the acquisition of technical skills but also the emancipation achieved through the literacy process (Freire 1968; Tygel & Kirsch 2015). In other words, it is not enough to teach people how to read a chart, you must also teach them how to use that chart to make the world a fairer place. The practice of literacy is the practice of freedom, as conceived by Freire.

So the question to be asked is: How do we go about empowering new learners with data? Rather than proposing a systematic framework for data literacy at scale, this paper offers five tactics for creative data literacy for empowerment. I use the term *creative data literacy*, rather than simply "data literacy", to draw attention to the fact that these techniques are geared towards non-technical learners who may need an alternative to the traditional quantitative approach to working with data. Moreover, rather than presuming that creative data literacy is the educators' domain only, each of the five strategies outlined in this paper specifies which audiences it targets in the data pipeline. The assertion here is that different groups of professionals can contribute to data literacy and that data learning may take place in a variety of settings. The groups that may play a role in engendering and enhancing data literacy include educators as well as data creators, data publishers, tool developers, tool and visualization designers, tutorial authors, government, community organizers, and artists.

3. Five tactics for creative data literacy for empowerment

These tactics are not systemic answers to the problem of data inequality and literacy. They are, however, starting points for building an inclusive set of practices to introduce new learners to "speaking data" (Bhargava 2014) and develop a "data mindset" (Miller 2014). They also challenge the legitimacy of the current data status quo which is producing discriminatory technologies and centralizing data-based power in state and corporate actors. These tactics are derived from my own work as an educator and tool designer, and from that of some of my colleagues, such as Rahul Bhargava, with whom I have developed pedagogical materials and the data literacy platform DataBasic.io. I teach undergraduate and graduate students majoring in the fields of Journalism, the Arts and Communication. I also run data workshops for those in municipal government, journalism, the nonprofit sector, and the arts. Although the tactics I introduce are neither exhaustive nor appropriate for all cases of data literacy learning, they can assist professionals in these fields to improve data literacy learning. My hope is that we can draw from tactics such as these while we work on developing a more systemic design and research agenda to cultivate data literacy for empowerment across numerous sectors, including and especially "the accountability industries": Law, Government, Journalism, Education and the Arts.

3.1 Work with community-centered data

Who can do this: developers; data creators and publishers; tutorial authors; educators

This first and crucial tactic involves the careful sourcing and selection of data that are relevant to the community that is learning to work with data. Ideally, this is data that are about the learners themselves, their field of work, or related to an issue they are facing. In most cases, sample data provided for learning purposes is either highly generic (height and weight distributions of people, for example) or only relevant to a small number of learners. For example, many online tutorials in R feature the *mtcars* data set.[1] This data set is from the Motor Trends magazine in 1974 and consists of fuel consumption and performance metrics for cars based on parameters such as number of cylinders, horsepower, rear axle ratio, and weight. Although for learners who are car mechanics or car enthusiasts this is very relevant data, for those who are not, it is alienating to work with data about something that they do not know (or care) much about.

Working with community-centered sample data opens up possibilities for connecting context and lived experience to the data. It also makes it easier for learners to apply their learning to their everyday lives or work contexts more quickly and directly. For example, in the project Local Lotto[2]—a collaboration between the Center for Urban Pedagogy, Brooklyn School for Social Justice, MIT's Civic Data Design Lab, and CUNY Brooklyn College—urban high school students were charged with determining whether the lottery was a good or bad thing for their neighborhoods. They had to make a data-driven argument by collecting qualitative and quantitative information about their neighborhood and learning about probability. The students charted where lottery tickets were sold, interviewed shopkeepers and residents, and created digital maps and graphs to explain winners and losers. In this case, the students had a context for working with the data. They had collected the data, they had deep, ongoing everyday relationships with the people and the place and, most importantly, they had a stake in the outcome of the data analysis.

Another project that has made community-centered sample data a priority is the Rap Research Lab, an afterschool program geared towards youth of color and immigrant and transgender high school youth in New York City. Rap Research Lab is led by creative technologist Tahir Hemphill and uses his dataset of song lyrics from more than 100,000 hip-hop songs dating back to 1979 as a starting point for teaching data literacy to youth. Learners attend the free after school program and design their own research questions and visualization projects while learning techniques of data analysis and design. Project outputs have included work that has charted the incidence of crime-related lyrics and actual crimes, explorations of the semantic nuances of "the N-word" and sentiment analysis that compares negativity in East Coast versus West Coast rap lyrics. For Hemphill, hip hop is a cultural indicator and also constitutes the musical backdrop to many of these learners' lives. Students "feel the power" of data analysis and start to ask ethical questions about data by being able to apply it to something that they already have deep contextual knowledge about (Creative Capital 2015).

Finally, working with community-centered sample data does not necessarily mean founding an entirely new project-based program such as those described above. It could be as simple as carefully considering what sample data you provide in your application, visualization or workshop; thinking about whether it is culturally appropriate, and whether your community has any connection to it. My colleague Rahul Bhargava and I have been building a platform called DataBasic.io that introduces concepts of data analysis to new learners. DataBasic is geared towards journalists, government, non-profit staff, artists, educators and students. Because our audience is wide-ranging we needed sample data with broad appeal. Taking inspiration from Rap Research Lab, we include song lyrics for introducing quantitative text analysis, as well as presidential speeches (since we launched DataBasic during a campaign year). We also include data on UFO sightings, the crash of the Titanic, and the social network of Paul Revere. These data sets are fun and most English language learners bring some context to them. However, DataBasic is also offered in Spanish and Portuguese and these same data sets do not work for learners of these languages. Thus, we have sourced song lyrics in Spanish and Portuguese, as well as data about Brazilian soccer teams, speeches by Fidel Castro and baby names in Portugal to provide starting points. These light and fun data sets, however, can sometimes fail to connect to learner's more pressing professional contexts. For example, in a workshop for municipal government officials, I was asked the following question: "This is great that we can analyze Prince's lyrics but how does this relate to the work we are trying to do?" When running DataBasic workshops for more specific audiences, we often work to source a custom data set to use as sample data that will be relevant to the group's questions. For the municipal government officials, that consisted of text analysis of citizen ideas for the future of transportation collected by the City of Boston. Once I showed the questioner the sample data, her face lit up and she immediately connected it to community surveys that her group had been putting out to collect citizen feedback.

3.2 Write data biographies

Who can do this: data creators and publishers; government; educators, students and learners

Many people working with data—including journalists, researchers, entrepreneurs, citizens and artists—are increasingly encountering data sets "in the wild". Thanks to the open data movement there are now APIs and government portals. There are test data sets for network analysis,[3] machine learning,[4] social media, and image recognition.[5] There are compilations of fun data sets,[6] curious data sets,[7] newsletters of data sets[8] and so on. This may seem to be a "good thing" in that I can download a spreadsheet of stop and frisk incidents in Boston[9] rather than making a public records request, for example.

The downside of discovering data through a quick Google search is that the data arrives at our doorstep completely decontextualized, without explanation of why it was collected, who collected it, in what way, and what its known limitations are (data creators are usually very aware of the limitations of the data they collect and maintain). The very best open data sets provide data dictionaries, user guides (Gradeck 2016), playbooks (Jacobs 2016) and other metadata to help introduce the answers to these questions. However, for most publishers, just getting a spreadsheet posted online is a struggle, and they do not have resources to post detailed metadata.

This situation poses a challenge for new learners. This is because new learners tend to see information organized systematically in a spreadsheet as "true" and complete, especially if the data is not noticeably missing values. One way of narrowing the rift between the data owners and the data users is to ask learners to write "data biographies". These are stories of how the data set came to be in the world. Instead of following a typical data analysis process (Figure 1) where you acquire a data set and work forward to see what meaning there might be in the data, creating a data biography requires learners to go backwards in time before engaging in analysis, and describe how a data set came to be in the world (Figure 2). Understanding how the data was collected can be a very important step in estimating whether patterns in the data are an artifact of the collection process or a signal in themselves. For example, are there more parking tickets issued in Boston in September because there are more parking enforcement officers temporarily deployed at that time, or because there are actually more people violating parking regulations?

Figure 1. A typical data analysis process.

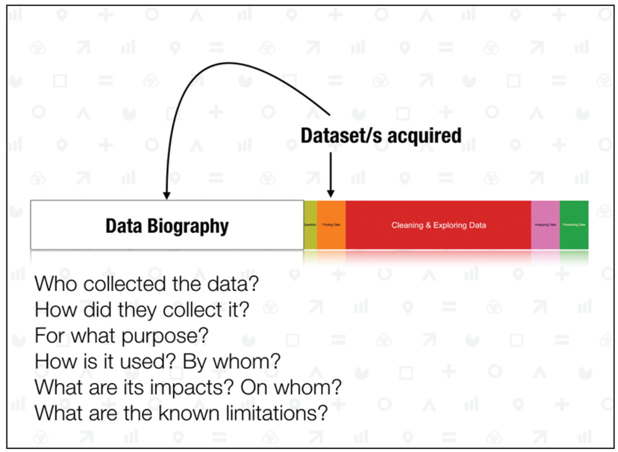

Figure 2. Going back in time to write a biography of a data set

Creating a data biography might be as simple as inviting the owner of a data set to present about the process of collection, or encouraging learners to interview one of the creators or maintainers of a data set. Occasionally for this assignment, learners have found that their data biography has turned into the whole story. For example, students in my data visualization course at Emerson College were investigating national data collected by the Clery Act Report comparing sexual assault incidents on college campuses. While doing their data biography the students learned that Clery Act data is self-reported on college campuses. They also learned that some of the campuses with the lowest rates of sexual assault were actually the ones with the fewest resources devoted to helping survivors, the least supportive environments and (possibly) places where institutions were turning a blind eye to sexual assault. This paradoxical finding became the subject of the students' excellent final story for the class (Torphy, Galnon & Meehan 2016). This would not have been possible if they had simply taken the data at face value.

3.3 Make data messy

Who can do this: community and event organizers; educators; government seeking public participation

Building on the idea of writing data biographies, this third tactic also advocates that we should *not* start the learning process with a data set. Rather, *making data messy* refers to introducing learners to the messy process of creating and categorizing data in the face of uncertainty and complexity. This tactic may seem counter-intuitive since the received wisdom is that people working with data spend up to 80% of their time cleaning their data (Lohr 2014), and there are standards around what constitutes "tidy data" (Wickham 2014).

As described in section 2 above, *write data biographies*, new learners often have the impression that data is "true", particularly quantitative data and data about the physical world. Cultivating skepticism of "raw" data, therefore, should be seen as one of the primary goals of any data literacy program that seeks to empower the learners. Taking the learners through the process of data collection, categorization and standard-creation helps them understand how inquiry goals, interests and politics contribute to the creation of data sets. Furthermore, it helps learners engage the critical thinking skills they will need to ask questions of other data sets in the future.

There are many possible ways to introduce learners to the data collection process. An extremely simple activity that I do when introducing the basic concept of data is to draw a spreadsheet on the whiteboard with two columns: "Name" and "Shirt Color". I then go around the room and ask individuals to give me their name and shirt color. Inevitably somebody in the group has a shirt with stripes or patterns. This opens up the conversation about which color bucket we apply to their shirt, whether we need to allow for multiple colors per shirt, whether we need to introduce the idea of adding a "pattern" column to make our data describe more of the world, and what we might need to use the data for in the future. I close the activity by asking people to name some of the things about the classroom and people that we are *not* including in our spreadsheet, so as to illustrate the idea that data is always an intentional simplification of a more complex and rich reality.

A more quantitative example is an activity that I do with the students on the module on sensor journalism. Students taking this module are asked to build extremely simple DIY water conductivity sensors, use them to test water samples and then reflect on what kind of data they would have to collect, and with what kind of rigor, in order to tell a story about urban water quality. In the process, students encounter complex problems, such as how to properly calibrate a sensor; the fact that measuring one point at one time in a water system is not meaningful unless you have more comprehensive data; the distance between using one simple measurement—water conductivity—when you really want to tell a story about water quality. The purpose of this assignment is to cultivate skepticism of raw data, especially data collected about the physical world by technological instruments.

Making data messy can also be employed by institutions working with diverse stakeholders and constituents as a way to engage in collaborative analysis and meaning-making from data. For example, from 2014 to 2016 the City of Boston ran a participatory planning process around the creation of a transportation master plan for the year 2030. GoBoston2030 employed creative methods to collect citizen ideas about the future of transportation, such as painted trucks and bicycles. Such methods allowed them to collected more than 5000 ideas as qualitative, unstructured text. While a data analyst could have been hired to produce an analysis in the shadows of city hall, the project leadership instead treated the analysis as an opportunity for further community engagement. Several events were staged. Each one of these events had over 75 local community leaders, citizens and policymakers, and participants worked in groups to sift through the data, prioritize the best ideas and highlight new categories for consideration.

Ultimately *making data messy* does not mean it has to stay messy. New learners are introduced to the challenges of collecting and categorizing data so as that they can conduct an inquiry into the world. The challenges that result are epistemological (i.e., how can we gather data to make a claim about the world?) and editorial (i.e., what do we need to include/exclude?). A key learning goal for creative data literacy is understanding the potential and limitations of what aspects of the world data does and does not represent.

3.4 Build learner-centered tools

Who can do this: developers and designers; tutorial authors

The growing popularity of data has led to a proliferation of tools to collect, analyze and visualize data. In separate works in progress, Rahul Bhargava, Dalia Othman and I are logging the more than 500 free or freemium tools designed for non-experts to collect and analyze data or create data visualizations. This large number of tools causes tremendous complexity for learners as there is little guidance on when to use which tool. For example, should a person with geographic data make a map with CARTO, Google Maps, ESRI StoryMap, Knight StoryMaps, ZeeMaps, OpenStreetMap, plot.ly, Tableau, D3 or R? Additionally, many tools prioritize the creation of quick, flashy visualizations rather than scaffolding a learning process that helps the learner through each stage of the data processing pipeline.

So the question is: What do learner-centered tools look like? The first fact to keep in mind is that the provision of more meta-information about the tool space and how to decide on an appropriate tool is crucial. Rahul, Dalia and I have a small-scale online experiment in progress called NetStories[10] where we ask students once or twice a semester to learn a tool, write a review of it, and assess what kinds of tasks it is good for and whether it is worth learning. Although these reviews work well for peer learning, the students have reviewed less than a quarter of the tools on our master list. Thus we need to scale up such an effort so as to provide a useful and comprehensive resource.

In prior work Rahul and I have outlined design principles for learner-focused rather than output-focused data tools (D'Ignazio & Bhargava 2016) and tried to design DataBasic with these principles in mind. DataBasic consists of four online, free, digital tools, with accompanying participatory activities, for data literacy learners in academic and workshop settings. In the list that follows, I enumerate our learner-centered principles and how we tried to enact them for DataBasic.

A learner-centered data literacy tool is:

1. Focused: Strives to do one thing well. Provides a low barrier to entry for the data literacy learner. Each tool in the DataBasic suite is focused on taking in one type of input and producing a single web page report or visualization as output.
2. Guided: Introduced with strong activities and sample data to get the learner started so they do not have to imagine use cases. The input screen of each DataBasic tool starts with sample data so the learner can quickly run it to see what kind of output it generates.
3. Inviting: Appeals to the learner—either because of direct relevance to the learner themselves or through the use of play, humor and visual design. DataBasic's visual design uses bright colors and simple layouts. The sample data draw from pop culture, something the user may be familiar with.
4. Expandable: Helps the learner take the next step (possibly to another, more advanced tool). Each tool in the DataBasic suite recommends two other tools that learners can use once they are ready to take the next step.

Learner-centered tools are focused and simple for beginners to use. Therefore, they will probably not be the tools that advanced users and professionals will choose to use in the long-term to produce flexible and complex outputs. This is why learner-centered tools must be expandable and help the learner "graduate" to more open-ended tools. The purpose of a learner-centered data tool is to introduce new vocabulary to beginners, to introduce them to data-centric thinking, and to help build their confidence and identity as someone who can work with data. This perception of "self-efficacy" has been shown to be extremely important to the learning process (Zimmerman 2000).

3.5 Favor creative, community-centered outputs over Tuftean purity

Who can do this: designers and artists; educators; community and event organizers

Edward Tufte has done prescient and important work to define the field of data visualization. However, as more people start working with data, particularly those outside the fields of graphic design, statistics and computer science, the range of outputs produced should rightly expand to accommodate people's increasingly diverse perspectives, goals and situations. While economy of visual elements, two-dimensional manicured charts and precise comparison may be goals if one's audience consists of scientists or designers, such graphical language might not be appropriate for community gardeners, policymakers or children. Journalists, artists, citizen scientists and makers are experimenting with ways of visualizing, physicalizing or even "visceralizing"[11] data in order to more effectively communicate their data-driven ideas.

For example, Rahul and Emily Bhargava, collectively known as DataTherapy,[12] work with community-based organizations to build their capacity to work with data. They focus on the organization's own data and work to collaboratively analyze it, so as to draw out a data-driven story that the organization will want to tell to a wider public. The result of this process is a "Data Mural"—a codesigned, large scale public painting (Bhargava, Kadouaki, Castro, Bhargava & D'Ignazio 2016). Some data murals, such as the one in Figure 3, have been painted on outdoor walls and others have been painted on banners, designed to be rolled out at community events.

Likewise, in a recent arts-based project called *Boston Coastline Future Past*, commissioned by the DeCordova Museum & Sculpture Park, I worked with artist Andi Sutton to create a "walking data visualization". While comparing old maps of the Boston coastline from 1630 and the project models of the coastline for 2100 based on sea-level rise, we saw a striking similarity. Thus Andi and I decided to host the walking data visualization event as a way of having a public conversation about this potential return to the past in the face of Climate Change. Around 35 people walked the past/future coastline of Boston (nowhere near the current coastline) and stopped along the way to listen to short speeches on climate adaptation from the Mayor's office, local scientists and media scholars. At the end of the walk, participants stenciled a temporary message about the future onto the Boston Common. The idea behind the project was to feel the future coastline with our bodies rather than to see it on a map.

Experimental outputs by others have included stitching photographs together to make aerial maps,[13] physicalizing data via 3D printing (Huron, Carpendale, Thudt, Tang & Mauerer 2014) and jewelry (Dwyer 2016), and using hand-drawn personal data as a starting point for intimate conversations (Lupi, Posavec & Popova 2016). While not all of these projects have a pedagogical

Figure 3. Creating a Data Mural. Datatherapy with the Somerville Food Security Coalition, 2014.

Figure 4. Boston Coastline Future Past. A walking data visualization by Catherine D'Ignazio and Andi Sutton.

focus, they represent a developing visual, tactile and experiential language that brings data back into the world, as an object of discussion, contention and communal production. These creative outputs often use vernacular visual forms (such as sketching, stick figures and balsa wood) intentionally, so as to convey the participatory form of production or the idea that the output and meaning of the data process do not end with a pristine expert analysis but continues to unfold through community discussion.

4. Conclusion

The collection, storage and processing of large amounts of data create a situation of asymmetry and inequality. The actors who collect, store and process the data are very different from those whose data are collected, stored and processed. This is a complex problem and part of the solution is in cultivating data literacy for empowerment in non-technical audiences. By doing so we may increase the number and type of people who can "speak data", frame problems that can be solved with data, and use data to transform (not just reproduce) the status quo. In this paper, I have offered five tactics for *creative data literacy* for empowerment, recognizing that non-technical learners may need pathways towards speaking data other than those coming from technical fields. While these tactics do not replace a more comprehensive research agenda around data literacy, they may be starting points towards this goal.

Submission date: 9 October, 2016
Accepted date: 8 February, 2017

Notes

1. https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

2. http://citydigits.mit.edu/locallotto

3. https://snap.stanford.edu/data/

4. http://archive.ics.uci.edu/ml/

5. http://www.cs.utexas.edu/~grauman/courses/spring2008/datasets.htm

6. http://koaning.io/fun-datasets.html

7. http://blog.yhat.com/posts/7-funny-datasets.html

8. https://tinyletter.com/data-is-plural

9. https://data.cityofboston.gov/Public-Safety/Boston-Police-Department-FIO/xmmk-i78r

10. http://netstories.org/tools

11. *Data visceralization* is a term coined by Kelly Dobson, formerly the head of the Digital + Media Program at the Rhode Island School of Design. It has to do with making data *felt* using various sensory and experiential techniques rather than only *seen* with the eyes.

12. www.datatherapy.org

13. https://publiclab.org/wiki/mapknitter

References

Andrejevic, M. (2014). Big Data, Big Questions: The big data divide. *International Journal of Communication*, 8, 1673–1689.

Bardzell, S. (2010). Feminist HCI: Taking Stock and Outlining an Agenda for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1301–1310). New York, NY, USA: ACM. doi: 10.1145/1753326.1753521

Bhargava, R. (2014). Speaking Data – Data Therapy. Retrieved October 7, 2016, from https://datatherapy.org/2014/07/09/speaking-data/

Bhargava, R., Kadouaki, R., Castro, G., Bhargava, E. & D'Ignazio, C. (2016). Data Murals: Using the Arts to Build Data Literacy. *Journal of Community Informatics*, 12. Retrieved from http://ci-journal.net/index.php/ciej

Boyd, Danah & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. doi: 10.1080/1369118X.2012.678878

Crawford, K. (2016, June 25). Artificial Intelligence's White Guy Problem. *The New York Times*. Retrieved from http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Crawford, K., Gray, M.L. & Miltner, K. (2014). Big Data| Critiquing Big Data: Politics, Ethics, Epistemology| Special Section Introduction. *International Journal of Communication*, 8, 10.

Dalton, C.M., Taylor, L. & Thatcher, J. (2016). *Critical Data Studies: A Dialog on Data and Space* (SSRN Scholarly Paper No. ID 2761166). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2761166

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 1–18. doi: 10.1080/21670811.2014.976411

D'Ignazio, C. & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Bloomberg Data for Social Good*. Retrieved from http://www.kanarinka.com/wp-content/uploads/2015/07/Big_Data_Literacy.pdf

D'Ignazio, C. & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal Of Community Informatics*, 12(3). Retrieved from http://www.ci-journal.net/index.php/ciej/article/view/1294

DiMaggio, P., Hargittai, E. & others. (2001). From the "digital divide" to "digital inequality": Studying Internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, 4(1), 4–2.

Blackburn-Dwyer, B. (2016). This Necklace Shows Just How Clean (or Dirty) Your Air Is. Retrieved December 30, 2016, from https://www.globalcitizen.org/en/content/pollution-necklace-jewelry-wearable-tech/

Emerson, JTactical Technology Collective. & . (2013). *Visualizing information for advocacy*. Bangalore, India: Tactical Technology Collective.

Freire, P. (1968). *Pedagogy of the oppressed*. New York: Continuum.

Gray, J., Bounegru, L., Chambers, LEuropean Journalism Centre & Open Knowledge Foundation., . (2012). *The data journalism handbook: how journalists can use data to improve news*. Sebastopol, CA: O'Reilly Media.

Goldsmith, S. & Crawford, S. (2014). *The Responsive City: Engaging communities through data-smart governance*. John Wiley & Sons.

Harris, J. (2012, September 13). Data Is Useless Without the Skills to Analyze It. Retrieved September 12, 2016, from https://hbr.org/2012/09/data-is-useless-without-the-skills

Huron, S., Carpendale, S., Thudt, A., Tang, A. & Mauerer, M. (2014). Constructive visualization. In *Proceedings of the 2014 conference on Designing interactive systems—DIS '14* (pp. 433–442). New York, New York, USA: ACM Press. doi: 10.1145/2598510.2598566

Jacob, N. (2016). Boston Smart City Playbook — from the Mayor's Office of New Urban Mechanics. Retrieved September 15, 2016, from https://monum.github.io/playbook/

Keller, E.F. (1996). *Reflections on Gender and Science: Tenth Anniversary Paperback Edition (Anniversary)*. New Haven: Yale University Press.

Julia Angwin, Surya Mattu, Jeff Larson, & Lauren Kirchner. (2016, May 23). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. Retrieved September 12, 2016, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Landström, C. (2007). Queering feminist technology studies. *Feminist Theory*, 8(1), 7–26. doi: 10.1177/1464700107074193

Letouzé, E. (2015). *"Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data."* Retrieved from http://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/

Lohr, S. (2014, August 18). For Big-Data Scientists, "Janitor Work" Is Key Hurdle to Insights. *New York Times*.

Lupi, G., Posavec, S. & Popova, M. (2016). *Dear data*. New York: Princeton Architectural Press.

Maine Data Literacy Project. (n.d.). Retrieved September 12, 2016, from http://participatoryscience.org/project/maine-data-literacy-project

Maycotte, H.O. (n.d.). Data Literacy—What It Is And Why None of Us Have It. Retrieved September 12, 2016, from http://www.forbes.com/sites/homaycotte/2014/10/28/data-literacy-what-it-is-and-why-none-of-us-have-it/

Mayer-Schönberger, V. & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.

Merriam, S.B. (1998). *Qualitative Research and Case Study Applications in Education. Revised and Expanded from "Case Study Research in Education"*. Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104; phone: 415-433-1740; fax: 800-605-2665; World Wide Web: www.josseybass.com ($21.95). Retrieved from http://eric.ed.gov/?id=ed415771

Miller, S. (2014). Collaborative Approaches Needed to Close the Big Data Skills Gap. *Journal of Organization Design*, 3(1), 26–30. doi: 10.7146/jod.3.1.9823

Neuhauser, A. (2015, June 29). 2015 STEM Index Shows Gender, Racial Gaps Widen. Retrieved September 12, 2016, from http://www.usnews.com/news/stem-index/articles/2015/06/29/gender-racial-gaps-widen-in-stem-fields

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press. doi: 10.4159/harvard.9780674736061

Phillips, A. (1991). *Engendering democracy*. University Park, Pa.: Pennsylvania State University Press.

Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*. Retrieved from https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf

Tufekci, Z. (2014a). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/4901 DOI: 10.5210/fm.v19i7.4901

Tufekci, Z. (2014b). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/4901 DOI: 10.5210/fm.v19i7.4901

Tygel, A. & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. Retrieved from http://www.researchgate.net/profile/Alan_Tygel/publication/278524333_Contributions_of_Paulo_Freire_for_a_critical_data_literacy/links/5581469d08aea3d7096e6a31.pdf

Walters, S. & Manicom, L. (1996). *Gender in Popular Education. Methods for Empowerment*. ERIC. Retrieved from http://eric.ed.gov/?id=ED398449

Welles, B.F. (2014). On minorities and outliers: The case for making Big Data small. *Big Data & Society*, 1(1), 2053951714540613. doi: 10.1177/2053951714540613

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10). doi: 10.18637/jss.v059.i10

Williams, S., Deahl, E., Rubel, L. & Lim, V. (2014). City Digits: Local Lotto: Developing Youth Data Literacy by Investigating the Lottery. *Journal of Digital and Media Literacy*, 2(2). Retrieved from http://www.jodml.org/2014/12/15/city-digits-local-lotto-developing-youth-data-literacy-by-investigating-the-lottery/

Yin, R.K. (2013). *Case study research: Design and methods*. Sage publications. Retrieved from http://scholar.google.com/scholar?cluster=17071144043372907427&hl=en&oi=scholarr

Zimmerman, B.J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, 25(1), 82–91. doi: 10.1006/ceps.1999.1016

About the author

Catherine D'Ignazio is an Assistant Professor of Civic Media and Data Visualization at Emerson College, a Principal Investigator at the Engagement Lab and a Research Affiliate at the MIT Center for Civic Media. Her work focuses on data literacy, feminist technology and civic art. She is the Faculty Chair of the Boston Civic Media Consortium, a network of 250+ researchers and community organizations committed to the ethical study and design of technology in/with communities. D'Ignazio has co-developed a suite of tools for data literacy (DataBasic.io), developed custom software to geolocate news articles and designed an application, "Terra Incognita", to promote global news discovery. She is currently working with the Public Laboratory for Technology and Science to explore the possibilities for journalistic storytelling with DIY environmental sensors. Her art and design projects have won awards from the Tanne Foundation, Turbulence.org, the LEF Foundation, and Dream It, Code It, Win It. Her work has been exhibited at the Eyebeam Center for Art & Technology, Museo d'Antiochia of Medellin, and the Venice Biennial.

Email: catherine_dignazio@emerson.edu