# WHY DATA LITERACY MATTERS

**Kristin Fontichiaro**
font@umich.edu

**Jo Angela Oehrli**
jooehrli@umich.edu

## Introduction

From their first research project, students notice numbers. From how many toes a polar bear has to the population of Azerbaijan to the percentage of Americans who own iPhones, student researchers sense instinctively that numeric data is a powerful way of communicating information. Even though students often believe numbers convey an air of authority in their work, their lack of critical awareness is undermining their success.

As many states move forward with adoption or adaptation of the College, Career, and Civic Life (C3) Framework for Social Studies State Standards, Common Core State Standards, and/or Next Generation Science Standards, students are expected to be fluent with data: to collect and analyze it, create figures and tables, integrate quantitative information, and move fluidly between text and visually represented numerical information.

Despite these formal expectations, students receive little guidance on how to move nimbly between text and numbers beyond what examples they see in textbooks or instructions for in-class controlled lab experiments. There is a disconnect between classwork and the data and statistical literacy skills needed beyond the classroom. Whether researching cancer statistics or the best car to buy, students don't often have a strong sense of what those numbers *mean*. Students often believe that numbers are objective, though data in the real world is rarely so. In fact, visualized data—even from authoritative sources—can sometimes be anything but. School librarians increasingly recognize that students either make poor decisions about the quality of statistics, data, and visualizations, or that they lack the ability to comprehend these resources altogether.

School librarians can play a significant role in helping students gain understanding of real-world numbers, statistics, charts, graphs, and visualizations. Librarians are unique cross-disciplinary pollinators who can fill the gaps between subject areas and help students gain skill in comprehending and critically evaluating data at home, at school, and in life. We collectively refer to these skills as *data literacy* and define data as:

1. Information represented numerically via raw numbers, percentages, percentiles, averages (mean, median, mode), etc.

2. Information that can be used algorithmically to determine compatibility (OKCupid), fitness levels (Fitbit), personality (BuzzFeed quizzes), etc.

3. Numerical information rendered visually (charts, graphs, coded maps, tables, etc.) to aid in pattern-finding and comprehension. (Fontichiaro and Abilock 2015)

Michael Bowen and Anthony Bartley wrote, "Data literacy is important for your students [...] because data are used to argue and persuade people to, among other things, vote for political agendas [...] or lease a car. An improved understanding of data practices means that better questions can be asked" (2014, ix) and better decisions made.

To build students' capacity as thoughtful, active citizens in this brave new world, we must first build our own data literacy capacity. Given limited time and access to students, we must distill and prioritize data skills, building out our toolkits with mini-lessons and high-leverage data literacy rules of thumb that bridge the gap between awareness and action.

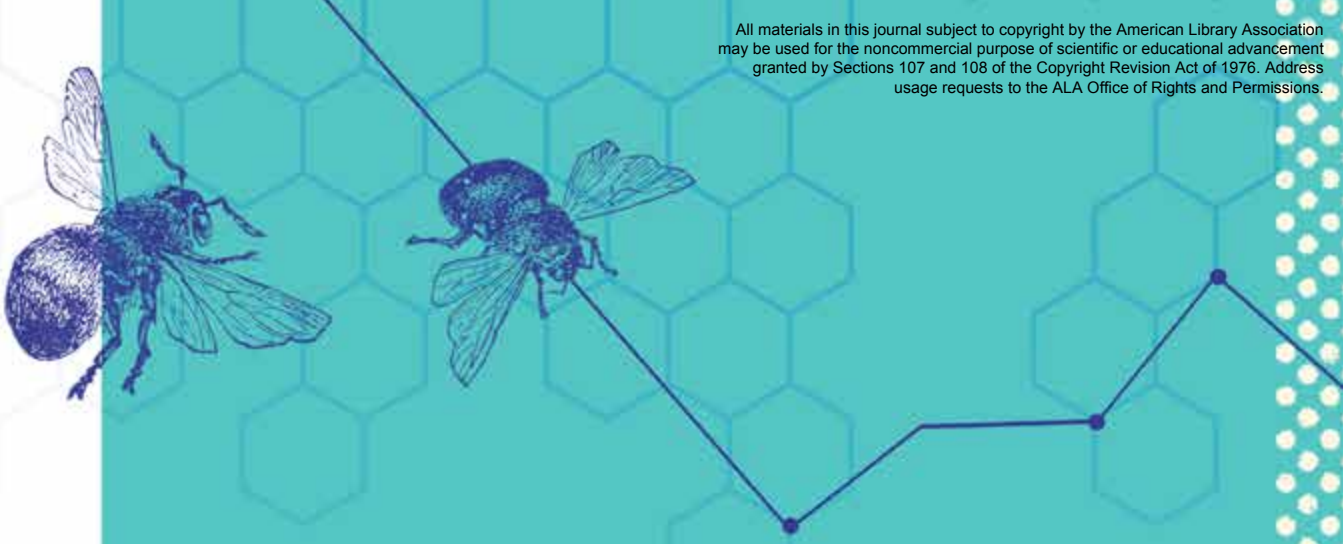
## How Do You Eat an Elephant? One Bite at a Time

Few librarians received formal instruction in statistics. Anecdotal evidence points to a profession dominated by humanities and social science majors with little collegiate practice in data and statistics. On our campus, no school library candidates since 2010 have had STEM (science, technology, engineering, math) backgrounds, and we imagine the situation is similar elsewhere. Few library schools incorporate data or statistical literacy into information literacy courses, and few school library programs require coursework in research methods, where statistical literacy would be a core learning objective. If you are among the few who are well-versed in data and statistics, we salute you. Everyone else, keep reading—we see a terrific opportunity ahead for librarians.

Tackling self-study in data and statistical literacy can be a challenge. In our new IMLS-funded project to develop data literacy as a subset of information literacy skills, we have concentrated on how students read, comprehend, evaluate, and synthesize data and not on how they create and organize data via lab experiments. In doing so, we have identified six significant themes for school librarians to consider.

### *One: Statistical Literacy*

Statistics flood news articles, Facebook feeds, and scholarly journals. School librarians and their students must critically "read," contextualize, and interpret raw



School librarians can play a significant role in helping students gain understanding of real-world numbers, statistics, charts, graphs, and visualizations. **Librarians are unique cross-disciplinary pollinators** who can fill the gaps between subject areas and help students gain skill in comprehending and critically evaluating data at home, at school, and in life.

and synthesized data. Discerning correlation from causation; recognizing the difference in the meaning of mean, median, and mode; understanding what margin of error signifies in polling data; and recognizing potential biases in collected data, among other skills, are critical for reading scholarly research, understanding arguments in popular media, and interpreting government documents. For example, MyFitnessPal released a list of the ten healthiest and least healthy states (MyFitnessPal Staff 2015). A savvy librarian asks, "How did they gather the data?" and discovers that the list was determined based only on MyFitnessPal users. She then recognizes that those users might not be representative of all residents.

### *Two: Data Visualization*

Having skills to create and comprehend mapped data, graphs, pie charts, and emerging forms of visualizations will help students effectively navigate visually rich information sets. At a session at the 2015 Research Relevance Conference, librarians shared their concerns that the emotional overtones (e.g., color, icons) used in these visualizations have powerful influence over students. One critical question posed in the conversation that our project will address was, "Which comes first? Learning how to make graphics? Or how to interpret them?" Additionally, the need for data visualization skills across library types was evident at the 2015 American Library Association Annual Conference, where approximately 150 librarians gathered for a two-part data visualization session, the conference's only session on data literacy.

### *Three: Data in Argument*

Our students can assemble random bits of factual data. However, it takes far more skill to understand how data is used—both informationally and persuasively—to support arguments in resources students examine, and then for students to create viable arguments themselves. These arguments could take the form of statistics embedded as evidence in a research paper, shared charts and graphs with tweaked or non-standardized elements, advertising, or infographics.

Infographics have emerged in many schools as a novel way for students to represent what they have learned, yet many school librarians with whom we spoke expressed dissatisfaction with students' work, stating that it too often contained disconnected facts and lacked a cohesive argument. Similarly, a 2013 survey indicated that citizens in the UK overlooked statistics that would correct their misconceptions on topics like the rates of teen pregnancy and crime (Ipsos MORI 2013).

### *Four: Big Data and Citizen Science*

Recent media reports lament society's "Big Data Problem" (Kopytoff 2014; Pena Gangadharan 2014; and Salmon 2014). More and more data is being collected, often without citizens' knowledge, via frequent-shopper cards, step counters, social media, and more. Some data is life-saving, such as DIY systems that help parents monitor their children's Type 1 diabetes by transferring insulin data temporarily and anonymously online (Nightscout Project 2015). Careful *human interpretation* of big data is required for positive outcomes to be achieved.

One student-friendly entry point for interacting with big data is citizen science. Students should recognize that projects like eBird and the emerging PhotosynQ invite the crowd-sourcing of data for the greater good (monitoring birds' locations and migration, and tracking photosynthesis levels in leaves, respectively) but must incorporate mechanisms to address data authenticity and accuracy. However, students must also realize that some data-collection projects start out with good intentions—such as the failed inBloom initiative to passively monitor identifiable student data and achievement for decades—but backfire by making too much personal data vulnerable to outside access. Microsoft researcher danah boyd has called for more understanding of data and statistical literacy as tools to navigate these issues (Pearle 2015; Hardy 2012).

### *Five: Personal Data Management*

From Google's personalized search results to Facebook's custom ads, students have daily experiences—often unbeknownst to them—captured as their clicks and likes are converted into actionable data. While students might like seeing relevant ads or music recommendations that match their favorites, few know it is because of the breadcrumb trail they leave behind. Students may think the website CNN.com is serving up the news to them, but they are usually unaware that as many as fourteen bots are following their actions and converting their clicks into data, according to a recent experiment conducted with Google Chrome extension Ghostery, which monitors webpages for beacons, advertising, and click-counting tools.

School librarians have long taught digital citizenship and the importance of being cautious about personal information shared online. As students use the open Web for research, they need to be aware that even if they do not enter their names or phone numbers online, information about them—their search habits, their choices of which webpages to visit—may be attaching themselves to their Internet cookies or Google account. Librarians must extend their credibility lessons to help students recognize that today's online content creators and social networks are engaged in a balancing act between maximizing advertising revenues and delivering quality content. This nuanced understanding goes beyond more-obvious credibility markers such as relevance, authorship, or currency into a recognition that information online is a market-driven economy—and editorial choices may be influenced accordingly.

### *Six: Ethical Data Use*

Data is not inherently good or bad, but it can be framed, edited, manipulated, or otherwise modified for unethical purposes (such as swaying voters by citing a small or outdated study)—or simply to confuse (Is "take an additional 25% off our half-off prices" the same as 75% off?). Just as students need practice in rhetoric and information literacy, students also need practice learning to create and evaluate data-related arguments and information. This may seem like a minor extension of

Librarians must extend their credibility lessons to help students recognize that **today's online content creators and social networks are engaged in a balancing act between maximizing advertising revenues and delivering quality content.** This nuanced understanding goes beyond more-obvious credibility markers such as relevance, authorship, or currency into a recognition that information online is a market-driven economy—and editorial choices may be influenced accordingly.

existing practice. We already tell our students to use information accurately and to cite sources. Is using data accurately and citing it any different? We believe that there *is* a difference because of students' inherent belief that numbers are infallible. We must train our brains—and theirs—to remember to stop and analyze numerical arguments, not just text-based ones. Ethical informational use is more than merely citing sources. In fact, we can leverage discussions about ethical use of personal data—something of deep personal value—to ground discussions of citation, an abstract concept the value of which can be more difficult for teens to grasp.

## Librarian Action Steps

For us, data literacy has been an acquired taste, and we suspect the same may be true for you. Kristin's interest grew from Debbie Abilock's gentle but persistent prodding that the big data movement was something to watch. At the University of Michigan Library, Angie currently chairs the data information literacy task force, which focuses on exploring existing data literacy teaching strategies as well as creating new teaching practices in this area. She has been working for several years on initiatives to develop her colleagues' understanding of data literacy. We have found these books and resources to be helpful:

- Abilock, Debbie, and Connie Williams. 2014. "Recipe for an Infographic." *Knowledge Quest* 43 (2): 46–55.
  *In-depth guidance on constructing argument-rich infographics.*
- Best, Joel. 2013. *Stat-Spotting: A Field Guide to Identifying Dubious Data*, updated and expanded ed. Berkeley, CA: University of California Press.
  *Uses friendly, easy-to-read language to provide rules of thumb to keep in mind while reading for pleasure or scholarship.*
- Fontichiaro, Kristin. 2014. *Creating and Understanding Infographics.* Ann Arbor, MI: Cherry Lake.
  *A child-friendly overview of infographic comprehension and construction.*

**Kristin Fontichiaro** *is coprincipal investigator for the Supporting Librarians in Adding Data Literacy Skills to Information Literacy Instruction project, funded by the Institute of Museum and Library Services, and a clinical assistant professor at the University of Michigan School of Information in Ann Arbor. A member of AASL, she serves on ALA's* Book Links *advisory board. She was awarded the 2015 Joan Durrance Community Engagement Award and the 2014 Excellence in Instruction Award from the University of Michigan School of Information.* Booklist *named her* Makers as Innovators *series one of the 2014 Top Ten Series Nonfiction (author and series editor). She recently edited the e-book essay collection* Information Literacy, or How I Learned to Stop Worrying and Love Library Instruction <www.smashwords.com/books/view/601117>. *She coauthored the chapter "Digital Badges: Purposeful Design in Professional Learning Outcomes for K–12 Educators" in* Foundations of Digital Badges and Micro-Credentials: Demonstrating and Recognizing Knowledge and Competencies *(Springer in press). She also authored the children's books* Hacking Fashion: Denim, Hacking Fashion: Fleece, Hacking Fashion: T-Shirts, Design Thinking, Watch It! Researching with Videos, *and* Review It! Helping Peers Create Their Best Work *(Cherry Lake Publishing). She blogs at <http://bit.ly/fontblog>.*

**Jo Angela (Angie) Oehrli** *is coprincipal investigator for the Supporting Librarians in Adding Data Literacy Skills to Information Literacy Instruction project. A former teacher in an alternative high school and in middle schools, Angie works as an instructional librarian and has coauthored a chapter on the University of Michigan (U-M) librarians' data education professional development project for* The New Information Literacy Instruction *(Rowman and Littlefield 2015). At U-M she collaborates on instructional strategies with diverse groups, including the campus Women in Science and Engineering Residence Program. As the former chair of the University of Michigan Library Instructor College, she led the professional development efforts for U-M librarians in the area of instruction. She co-created the Michigan Instruction Exchange, a low-cost, statewide conference for instruction librarians. She is an adjunct lecturer in the University of Michigan School of Information, focusing on instructional practices for librarians and information professionals, and teaches both basic and advanced digital research methods courses for the U-M College of Literature, Sciences, and the Arts. She has also chaired the Top Twenty Committee for ALA's Library Instruction Round Table and currently serves on the LOEX Advisory Council.*

- Hoelter, Lynette. 2014. "Data, Data Everywhere and Not a Number to Teach." <www.youtube.com/watch?v=N1RKVP9O9_w> (accessed December 15, 2015). *A welcoming overview of data and statistics by an advisor to our project.*
- Wheelan, Charles J. 2013. *Naked Statistics: Stripping the Dread from the Data.* New York: Norton. *Provides user-friendly context for understanding statistics in the real world. Definitely not a stats class textbook!*
- Yau, Nathan. 2013. *Data Points: Visualization That Means Something*. Indianapolis: Wiley. *A superb introduction to data visualization analysis and construction. Remarkably user-friendly. Yau also publishes the flowingdata.com blog.*

## Conclusion

Data is more than charts, graphs, and spreadsheets. It is being used in powerful, sometimes nearly invisible, ways to shape how we view the world and our role within it. In a world where everyone is an author online, data use is at the crux of teens' daily lives. Never has there been a more critical time to declare data as an essential literacy for students. Is it ethical for someone's social media profile to be used as evidence against them? Should someone's Fitbit data be used to contradict her testimony? It already has been (Hambright 2015). Should algorithms predicting future crimes be used to sentence someone now? They already are (Barry-Jester, Casselman, and Goldstein 2015). Should neighborhood data be used to determine mortgage eligibility and rates, despite the legal ban on "redlining"? It already is (Knott 2015). Do schools and libraries have a responsibility to help future generations figure this out? Our answer is that we unequivocally do.

These are the questions that affect all of our students. As part of our IMLS-funded grant, we will be publishing professional development materials for school librarians and other educators, as well as conducting virtual conferences in summer 2016 and 2017, focusing first on statistical literacy, data in argument, and data visualization, with the remaining themes scheduled for development in the final year. We invite you to view our progress and join our e-mail notification list at <http://dataliteracy.si.umich.edu>.

## Works Cited:


Barry-Jester, Anna Maria, Ben Casselman, and Dana Goldstein. 2015. "Should Prison Sentences Be Based on Crimes That Haven't Been Committed Yet?" *FiveThirtyEight* (August 4). <http://fivethirtyeight.com/features/prison-reform-risk-assessment> (accessed December 15, 2015).

Bowen, Michael, and Anthony Bartley. 2014. *The Basics of Data Literacy: Helping Your Students (And You!) Make Sense of Data*. Arlington, VA: NSTA.

Fontichiaro, Kristin, and Debbie Abilock. 2015. "Understanding Dressed-Up Data: Helping Students Unpack Data Visualizations." Concurrent session presented at the American Association of School Librarians 17th National Conference, November 6. Columbus, OH. Slides available at <www.fontichiaro.com/activelearning/2015/11/06/aasl-data-visualization-presentation> (accessed January 30, 2016).

Hambright, Brett. 2015. "Woman Staged 'Rape' Scene with Knife, Vodka, Called 9-1-1, Police Say." *Lancaster Online* (June 19) <http://lancasteronline.com/news/local/woman-staged-rape-scene-with-knife-vodka-called--/article_9295bdbe-167c-11e5-b6eb-07d1288cc937.html> (accessed December 21, 2015).

Hardy, Quentin. 2012. "Rethinking Privacy in an Era of Big Data." *New York Times* (June 4). <http://bits.blogs.nytimes.com/2012/06/04/rethinking-privacy-in-an-era-of-big-data> (accessed December 21, 2015).

Ipsos MORI. 2013. "Perceptions Are Not Reality." 2013. <www.ipsos-mori.com/researchpublications/researcharchive/3188/Perceptions-arenot-reality-the-top-10-we-get-wrong.aspx> (accessed December 21, 2015).

Knott, Jeff. 2015. "How to Use Big Data to Underwrite Safer Mortgages." *National Mortgage News.* <www.nationalmortgagenews.com/news/technology/how-to-use-big-data-to-underwrite-safer-mortgages-1058989-1.html> (accessed December 21, 2015).

Kopytoff, Verne. 2014. "Big Data's Dirty Problem." *Fortune Data* (June 30). <http://fortune.com/2014/06/30/big-data-dirty-problem> (accessed December 21, 2015).

MyFitnessPal Staff. 2015. "The Healthy Habit Index: The Top (And Bottom) 10 Healthiest States." *Hello Healthy* (July 6). <https://blog.myfitnesspal.com/the-healthy-habit-index-the-top-and-bottom-10-healthiest-states> (accessed December 21, 2015).

Nightscout Project. 2015. [homepage]. <www.nightscout.info> (accessed December 21, 2015).

Pearle, Laura. 2015. "Big Data and danah boyd at #alaac15." *Venn Librarian* (July 8). <https://lpearle.wordpress.com/2015/07/08/big-data-and-danah-boyd-at-alaac15> (accessed December 21, 2015).

Pena Gangadharan, Seeta. 2014. "The Dangers of High-Tech Profiling, Using Big Data." *Room for Debate* (August 7). <www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/the-dangers-of-high-tech-profiling-using-big-data> (accessed December 21, 2015).

Salmon, Felix. 2014. "Why Quants Don't Know Everything." *Wired* (January 7). <www.wired.com/2014/01/quants-dont-know-everything> (accessed December 21, 2015).