

## Book Reviews – “Calling Bullshit” and “The Data Detective”

Lydia Dawe

The University of Notre Dame Australia

Lydia.dawe@nd.edu.au | <https://orcid.org/0000-0002-7275-0260>

Bergstrom, C. T., & West, J. D. (2020). *Calling bullshit: The art of skepticism in a data-driven world*. Random House.

Harford, T. (2021). *The data detective: Ten easy rules to make sense of statistics*. Riverhead Books.

While we as information professionals may feel confident in the world of words, it is easy to be duped by numerical nonsense. The words “Big Data” and “algorithm” are often accompanied by grandiose promises and daunting predictions. Dubious statistics and de-contextualised numbers flood social media and fast-news outlets. Two recent books offer an antidote - good old-fashioned quantitative reasoning.

In *The Data Detective*, journalist and economist Tim Harford lists the habits of mind that are essential when encountering statistics. He stresses that his book is not about “debunking” statistics, but confidently scrutinising them. Academics Carl T. Bergstrom and Jevin D. West adapted *Calling Bullshit* from their undergraduate course of the same name. The syllabus and course materials are freely available online and are well worth exploring (Bergstrom & West, 2019). For Bergstrom and West, “bullshit” is about manipulation, superfluous detail, and “statistical snake oil [...] with a blatant disregard for truth, logical coherence, or what information is actually being conveyed” (p.40). The two books have complementary approaches: *Calling Bullshit* covers statistical analysis methods in more detail, while *The Data Detective* illustrates its “rules” with fascinating detours into history, science and politics. Both books are entertaining reads that challenge our perspective on the power of numbers. Here’s some of what I learned from them.

### Slow down and ask questions

Bergstrom and West lament that the internet has resulted in not only more bullshit, but bullshit spread at lightning speed, while “the truth comes limping after” (p.15). The solution? What librarians already do in a reference interview: slow down, ask questions, and gather information. The crowdfunding website *Kickstarter* boasts its success stories - someone once raised \$55,492 to fund the making of a potato salad - but how many other campaigns failed to raise a cent? A popular *TED* talk proclaims that “power posing” positively affects your hormones, but have the results been replicated in other studies? Admittedly, the answers to these questions often throw a

wet blanket on the excitement, but when the stakes are high, we must probe the headlines. Therefore, don't be embarrassed to ask simple questions such as "how do you know that?" or "what does that mean?" If the answer sounds like bullshit, keep investigating.

### Beware the infographic

Both books devote an entire chapter to data visualisation. As Harford puts it, "misinformation can be beautiful, too". Infographics are not new - Florence Nightingale used them to argue the case for sanitation in the Crimean War - but they are more pervasive than ever. Data visualisation is storytelling, and by making a few adjustments in scale, shading and labelling, the story can be easily manipulated. Bergstrom and West go through diverse examples, from the bar graph with axes starting at arbitrary numbers (e.g. 35 instead of 0) to line graphs with dubious comparisons (e.g. MMR vaccination and autism rates). Both books reserve particular derision for the "duck" infographic - those that favour form over function and that are shaped like the thing they are explaining. We've all seen these - literacy rates shown by bar graphs in the form of book spines, transportation pie charts shown in the wheels of a bike. Harford suggests you check your emotional responses when reading an infographic: do you feel angry, defensive, triumphant, satisfied? Evaluate that feeling. Before you get swept away by the beauty of a graph, ask yourself: do you understand what is being measured or counted? The pause we take to understand can make all the difference.

### Call bullshit on Big Data

We have never generated so much data just by living our lives. Our every transaction, search, and movement becomes a data point that is swept up and aggregated by a tech giant. Silicon Valley is guilty of falling victim to its own hubris, convinced that an abundance of data means all the answers are there to be found, and proving causation is irrelevant. Both books tell the story of Google's *Flu Trends*, a project big on promises, absent on theory. Google claimed that people's search terms (data they refused to reveal) correlated with flu outbreaks - and for a while, *Flu Trends* seemed to be on the right track. Soon after, however, the predictions fell apart. Google failed to account for the other reasons people might be searching for flu information, such as news headlines, predictive text prompts, or self-diagnosis (people often mistake a cold for the flu). If Google were more transparent in its process, and worked more collaboratively with others, it may have developed a successful tool. The bottom line is, we don't need to understand the algorithm, but we do need to question what went into it: what is the data? Where did it come from? How representative is it? The golden rule of machine learning is "garbage in, garbage out". As big tech companies have an incentive to avoid scrutiny, we should be all the more suspicious of their claims.

### Put the numbers in context

Because of publication timing, Covid-19 and the subsequent explosion of misinformation are not covered by these titles. If written a year later, it is likely that case studies about Covid-19 statistics would have taken centre stage in both books. The pandemic is a complex topic at the crossroads of epidemiology, virology, statistics and sociology that poses a challenge for anyone with basic scientific literacy (Braund, 2021). Throw in an abundance of alien-sounding metrics such as mortality risk, case fatality rate, and death numbers vs. case numbers (Roser et al., 2020), and you've got a recipe for mass confusion. The result is people questioning why they should take a vaccine that is 90% effective if they have a 99% chance of surviving the virus (Reuters, 2020). Due to the 24-hour news cycle and resulting click-bait headlines, numbers are often ill-defined and falsely compared to each other. Harford stresses that journalists should not only define a measure, but put numbers in a context people will relate to. Americans may think that a 1% mortality rate sounds good, without realising it means the loss of between 700,000 and 1.5 million people - or the entire population of Hawaii (Coltrain, 2020).

### Be skeptical, not cynical

Bergstrom and West assert that the chief source of bullshit to contend with is yourself: "what you see depends on where you look" (p.106). With the Internet, you will find what you are looking for eventually. We fail to fact-check sources we "feel" are true because the information confirms our existing beliefs. We can never be rid of cognitive bias but we can be mindful of it in our evaluation process. One of those biases is cynicism. Harford dislikes the notion that all statistics are smoke and mirrors, inherently misleading. Taking that view, he explains, is intellectually lazy - it gives us permission to dismiss a claim without investigating it. Nor should we become skeptics whose sole aim is to "shoot down falsehoods" (p.11). Instead, quantitative reasoning should be about curiosity: having an open mind and seeking the truth.

Information literacy is a work in progress, and quantitative literacy is part of that. These books provide an accessible starting point to the latter. I encourage all information professionals to read them and take their lessons to heart: keep asking questions, and don't be afraid to call bullshit when you see it.

### References

Bergstrom, C. T., & West, J. D. (2019). *Syllabus for Calling Bullshit*.

<https://www.callingbullshit.org/syllabus.html#Introduction>

Braund, M. (2021). Critical STEM literacy and the COVID-19 pandemic. *Canadian Journal of Science, Mathematics and Technology Education*, 1-18.

<https://doi.org/10.1007/s42330-021-00150-w>

Coltrain, N. (2020, May 5). *Fact check: Does COVID-19 have a mortality rate of 1%-2%?* USA Today. <https://www.usatoday.com/story/news/factcheck/2020/05/05/covid-19-fact-check-coronavirus-mortality-rate-misleading/3019503001/>

Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). *Mortality risk of Covid-19*. OurWorldInData. <https://ourworldindata.org/mortality-risk-covid>

Reuters. (2020, November 21). *Fact check: Misleading comparison between COVID-19 mortality rate and vaccine efficiency*. <https://www.reuters.com/article/uk-factcheck-vaccine-comparison/fact-check-misleading-comparison-between-covid-19-mortality-rate-and-vaccine-efficiency-idUSKBN2802H4>