

Lisa Gitelman's *Raw Data Is an Oxymoron* Revisited For a Pandemic

Review by Lee Boot
Director, Imaging Research Center, UMBC

Gitelman, Lisa. Raw Data Is an Oxymoron. The MIT Press, 2013.

Raw Data is an Oxymoron is not a new book, but it is newly relevant. The use of, and reverence for, data have continued to increase hyperbolically in the years since the book was published, and data as an ambient phenomenon is now everywhere, all the time, for everyone. Data is now becoming a proxy for reality itself—beyond everywhere, it is becoming everything. Nonetheless, critical consideration of the relationship between human beings and our data may, for many in our technocracy, echo the cliché of fish pausing to consider water. But the need to be critical is now urgent. As I write this reflection on the book, a global pandemic is crashing into humans and our ways. Storytellers know that events both reveal and shape who we are; thus, as COVID-19 ravages our species, we are exposed—not only to it, but also by it.

A key insight that emerges from *Raw Data is an Oxymoron* is that the way we've imagined data into existence projects our hope that it can be all we need—that the data will be the whole story. The book disabuses us of that notion through essays that paint pictures of the contexts in which data are embedded, tamping down our data-topian tendencies but, at the same time, creating an appreciation for data used well. Unfortunately, events surrounding the spread of COVID-19 make obvious that we've imagined the virus as reductively as we tend to imagine data. The two are hopelessly intertwined, supporting one another symbiotically, and also mirroring one another analogically. While *Data journalists* show us the impact of the virus in charts, the *digitalmediadatasphere* reverberates with the voices of policymakers and politicians “interpreting” them for us, preferring us to see the virus alone as the cause of its lethal impact. We listen, comfortable in the extraction of infection and death numbers if not the horror itself, as data become a more and more refined fuel for the calculation enterprises we hope will save us. The mythology of objectivity that miscast data also associate intelligent competence with being dispassionate, or “cool and collected,” when faced with a crisis; suffering and stupidity are the wrong kind of data to make the news. If masks and gloves are the protective devices for our bodies, data is the prophylactic that protects us from having to confront the limitations of our most hallowed ways of thinking. Like data, a virus isn't really alive in the normal sense but seems to possess agency and intention. Both the virus and the data about it appear to have lives of their own.

The essays in *Raw Data is an Oxymoron* that trouble our vision of, and relationship to, data are worth reading again now. Matthew Stanley's chapter on how scientists, in order to derive data that gave us the calendar for predicting solar eclipses, cast themselves in the role of interpreters of human behavior, despite the mismatch of expertise. David Ribes and Steven Jackson chronicle the near impossibility of maintaining a viable long-term study of environmental measures because of the simple, practical challenges that emerge when people seek to precisely

repeat the ways they probe their physical world over an extended period of years. The COVID charts make invisible what we can only imagine to be Kafkaesque data collection practices around the globe. While such essays demonstrate that data can only be truthfully understood in light of their immediate context, others suggest that the distinction itself is a weak dichotomy. Ellen Gruber Garvey, in her chapter about Northern abolitionists' use of Southern newspaper ads about runaway slaves and slave markets to bring injustice to light, reminds us how shifting the context surrounding data is a rhetorical power move that can produce dramatic political effects. The chapter by Kevin Brine and Mary Poovey describes how using theory as a context for data creates a synergy that in turn becomes a powerful engine for understanding macroeconomics. Daily evening briefings on the coronavirus reveal that data have become a political football untethered by a viable theory that people can understand let alone agree upon.

Gitelman, along with Rita Bailey in her chapter on *dataveillance*, appropriately, if predictably, explore the nefarious potential of data to become intentionally weaponized or used in other, more banal, forms of evil. This is a story that has finally found ample representation in the mainstream press, and that is progress. But the most important contribution of the book is in the way it grills the notion of data as an *objective given* by pointing out that the value of data can only be realized if we can locate the context immediately surrounding and pertaining to the data: their provenance, circumstances and support structures.

The shortcoming of the book, in terms of its capacity to help us navigate our current crisis, is that it doesn't go nearly far enough. To suggest what going further might mean, I would begin with the book's introduction, in which Gitelman and Virginia Jackson touch on photography as both an example and a culturally reactive symbol of the apparent objectivity of data. I would suggest that by adding a dimension, so that photography becomes film, we would have a metaphor to help us understand more of the unsettling limits of our relationship with our data and why we need to dimensionalize it much further.

The beloved and maligned screenwriting sage, Robert McKee, in his book, *Story*, writes "If the scene is about what the scene is about, you're in deep shit." He cites the traumatizing scene in the brilliant film, *Chinatown*, in which a cascade of incest related plot-points are revealed to viewers like the slaps in the face happening on screen. Suddenly, the whole story makes sense. In and of itself, a sensational scene has no meaning without its relationship to the film as a whole. Creating parts in relation to the whole is fundamental to storytelling, the arts more generally, and across mature domains. Paintings are created one stroke at a time, but the art of painting is the construction of a gestalt experience that evokes affective cognition or to some other end, out of the individual strokes. This illustrates the overarching theme of *Raw Data is an Oxymoron*, that context is needed to reliably make meaning of data. A global pandemic, however, requires that connections be made that go far beyond the relatively internal logic of a scene in relation to a film, and yet storytelling is once again a useful example. Screenwriters discuss the "throughline" of the film as the structure of what happens, materially, to advance a film's plot. Dorothy, knocked unconscious during a tornado, dreams her way to a dangerous and wondrous fantasy world, and then back to the farm. If that were all there is to *The Wizard of Oz*, we might never have heard of the film. The same would be true if *Moby Dick* were just a story of adventure and revenge amongst interesting characters. We might enjoy these stories, but we miss most of their value if we don't look at the social and political landscape beyond them to see how, like the coronavirus, they collide with our culture and society and shed new light on our individual and collective decisions. Unless juxtaposed with evidence of the xenophobia in other promoted American narratives, we can't navigate a landscape of nationalism, racism and sexual discrimination. The larger human meaning of great works emerges from larger and more remote circles of context than the immediate (scene to film)

connections. One dataset must be seen in light of another, and another. Like Ahab, we must scan the horizon in all directions—far from the center of our data and even its origins to the myriad scaffolding contexts that reach way, way over the rainbow. A culturally influential film or book is often far more than its plot.

To protect ourselves from a global pandemic such as COVID-19 it has to become more normal to connect the full landscape of factors that have led to people dying now, today, in (best case) an emergency room near you. The data in the coronavirus narrative function as something screenwriters might call a *macguffin*: an object chosen only for its ability to make the mechanics of the plot work. Robert McKee cites the Maltese Falcon statuette in the film of the same name as an example. Nobody cares about the bronze bird, and data are the macguffin in the story of COVID-19.

In her article, “It Wasn't Only Trump Who Got it Wrong,” social scientist and technology journalist Zeynep Tufekci points out that what has made COVID-19 so deadly, particularly in the US, is that we did not scan the horizon and thus did not connect the dots. We did not look for negative synergies among various pools of data—even data regarding trends in global travel, much less long-term political trends in the US and beyond to try to run governments like businesses with their just-in-time supply practices, meager spending on preparedness, or, using an appropriately wide lens, the way intentionally reducing trust in government or mixing public messaging may be the most impactful data of all: right there in front of us but difficult to use. Fortunately, Tufekci's article pulls no punches in this regard. In a refreshingly accusatory tone, she highlights the vast crimes of omission, calling the thinking of politicians and policymakers (including those in public health): *asystemic* to the point of inexcusable negligence. To those who think about how we think about our civilization's most overwhelming challenges and what would be required to meet them, the article is a clarion call to advocate against Modern era reductive thinking alone as a cure anymore. It's a call that cuts across the ruts of partisanship as liberals and conservatives are both more than capable of thinking that is far too narrow.

Beyond the current bio-siege, other, so called “wicked” challenges where we fail decade after decade to make progress, are, logically, those that have developed “herd immunity” to asystemic thinking. How else could it be?

Read in light of the coronavirus crisis, the book usefully conjures a ghost. Our vision of a few essential, properly cleaned, and isolated data as a proxy for truth is haunted by ontological spirits whose efficacy to deal with the problems we face has died without our noticing. Our vestigial impulses are still narrow. Despite plenty of discourse to the contrary (occurring in more remote datasets), we still intuitively jump to the conclusion that poor communities are the source of their poverty rather than the results of the designs of systems they had no power to shape. To use a data engineering term, we have not *joined* those datasets, and we rarely do. According to Dr. Tufekci and others, this is now killing us.

Slicing and dicing reality the way the past guides us to do, leads to us solving problems that are nothing like the ones we have. We dream of solving the COVID-19 crisis with a vaccine while leaving in place all the systems, beliefs and attitudes that made the virus so much deadlier than it needed to be. We want to believe we can draw tight circles around problems, leading us to fantasize about narrow technological solutions to global warming, underachievement in education, transportation.

Beyond the scope of this reflection is the discussion we need to have about the cultural change that will be necessary for us to be more wise about how we use the data we have. To move beyond slogans such as the

ubiquitous “think outside the box” or Apple’s 1997 “think different” will require new tools and practices that allow us to rework how we’ve been doing our most celebrated technocratic thinking for at least two and a half centuries. No biggy.

Of course, *Raw Data is an Oxymoron* as a project made no claim that it was aiming to look far beyond discrete datasets in specific use cases. Reality makes that case. Instead the book makes powerful and useful points about context, society, and the “runaway train” of data-utopianism. In his Afterword, *Data Flakes*, Geoffrey Bowker draws an astute analogy, pointing out that we see the rawness of data similarly to the ways societies have viewed the rawness of nature versus the social world of humans. He discusses the ways in which the elevation of quantitative data, exacerbated by how well fit digital computing, threatens to distort our understanding of ourselves and the world. All this reveals Gitelman’s laudable intention to find us the insights we need regarding this gargantuan subject. Perhaps another edition, or even a new volume, will take us on the next leg of the journey. As I sit here, self-isolating in my home for the sixth week, it can’t arrive soon enough.

References

Gitelman, L. (2013). *Raw Data is an Oxymoron*. The MIT Press.

McKee, R. (1997) *Story: Substance, Structure, Style and The Principles of Screenwriting* (1st ed.). Regan Books: p.253.

Tufekci, Z. (2020, March 24). “It Wasn’t Just Trump Who Got It Wrong.” *The Atlantic*. <https://www.theatlantic.com/technology/archive/2020/03/what-really-doomed-americas-coronavirus-response/608596/>

Cite this Review

Boot, Lee. “Lisa Gitelman’s *Raw Data Is an Oxymoron* Revisited For a Pandemic.” *Rhizomes: Cultural Studies in Emerging Knowledge*, no. 36, 2020, doi:10.20415/rhiz/036.r01
