

# Feminist Text Analysis, Spring 2023

Can there be a feminist text analysis? Feminism, text, and analysis in a computational world

## Book Review – “Raw Data” Is an Oxymoron

Data, in the modern-day world, became such a phenomenon that people began (and continue) claiming that it provides an objective set of facts (the fundamental stuff of truth itself) and consequently, methods and tools for its analytics would truly revolutionize everything and allow us an understanding of the “raw” and “unfiltered” sense of the world around us.

The book “Raw Data is an Oxymoron” stands as an answer to that claim, critiquing and analyzing the fundamental element of the process itself – data. While much has changed ever since the book was released in 2013 (like the advent of large language models and other AI-powered automated tools released daily), it can be taken as a rather early work on the subject to uncover what data really is.

The book is actually a compilation of essays providing different perspectives on data through different lenses of observation. While it is divided into chapters each being essay by a different author(s), every two sets of chapters can be taken to be different “topics” or “sections” in the book (as outlined in the introduction).

The first section looks rather at a historical and cultural account of some of its early beginnings (including how the word came to be, its meanings, and development over time) and critique of early modern arithmetic (I believe the inclusion of this chapter is because data is tied to math but it felt a little out of touch because it began to delve more into a critique of arithmetic than data). The second section looks at the formulation, creation, and collection of data in different disciplines (one being economics, the other astronomy). The third section relates to essays involving ideas of databases, data aggregation, and information management exploring the ethical implications of data and how it can be used to influence and manipulate. The fourth and final section relates to the use of data in society, a sort of exploration of the future of data, how it comes to impact people in the (then) current day world, and the challenges that large (and ever-growing) sets of data possess as data itself seems to be becoming an organism on its own.

As the title suggests what the book discusses through its essays and chapters is the idea that data is never “raw”

and is rather “cooked”. When considered alone, data is still not raw in the sense that it’s a reflection of the world which itself has its biases and issues and so they are transferred over. Data is also “conceptualized” and “collected” and in that process contains biases of the humans or human-made tools collecting them. Also, data has to be manipulated, cleaned, structured, processed, etc. before any analysis can be performed and observations can be made. These end up “cooking” up the data we first begin with. Now when an analysis is performed through the usage of particular models run through tools, both the models and tools have embedded biases in them and contexts that they do not consider and that apply to any field whether economics or astronomy or others. Finally, even after the analysis is done, the results have to be interpreted which brings in biases from the humans doing it.

So, what we essentially understand is that data can never be raw or objective, that at every stage, it contains, grows, and evolves on its limitations and biases and so it is important to consider that since data or rather the analysis and conclusions we draw from data has wide implications whether in public policy, politics, education, business, etc.

The book connects to our class conversation in all the stages we have discussed. By simply replacing “data” with “text” – which is the element of critique in our class, we can still apply the facts that text itself, the processes performed to collect, process, clean, analyze and interpret it all have hidden biases which give rise for the need to approaching the problem from a feminist perspective.

Two quotes really showed the duality that data (text in our class discussion case) possesses. “Data require our participation. Data need us.” and “Yet if data are somehow subject to us, we are also subject to data.” This connects to the discussions we have that data requires labor and participation of humans and does not exist without a form of human intervention and, quite equally, we are subject to data itself. And being subject to data, we are ourselves being shaped by it. This is further discussed in an essay with the writing “The work of producing, preserving, and sharing data reshapes the organizational, technological, and cultural worlds around them.”

Another interesting idea I found can be summarized by the quote “When phenomena are variously reduced to data, they are divided and classified, processes that work to obscure – or as if to obscure – ambiguity, conflict, and contradiction.” And that is a point of concern to us, especially as feminist scholars critiquing the topic that when we reduce real-world phenomena to data, we are in many cases just oversimplifying things and lose out on the depth and complexity of the topic (for example classification algorithms can fit real-world phenomena into a binary while these can be on a spectrum rather than a strict binary).

The chapter on economic theory and data also tied to our conversations about how “all models are wrong, but some are useful” showing that economists’ use of data and the modeling done are actually just approximations of reality only useful “for the formulation in which they appear.”

Something that I did not think about earlier but now completely agree with can be summarized through the following sentence from the book: “Raw data is the material for informational patterns still to come, its value unknown or uncertain until it is converted into the currency of information.” It made me think how data is only valuable and useful to us when it can be converted into the “currency of information” or in other words when it provides us with information that is useful or profitable (whichever sense the word is applied to whether purely business or even academic) to us. This brings to the sense that data has in most cases become a utilitarian as well as a capitalist tool.

Also, the book does touch on the topic of how power dynamics influence the production, analysis, and interpretation of data and how that can be used to influence and manipulate as well as exert power – which is an important factor to consider when building a feminist critique.

While the book is now outdated since many more technological leaps (and dangers) have grown within the last decade, it does provide a good introduction to the topic especially given that it is a relatively early work in the field of data studies. Also, it provides a rather introductory, theoretical understanding of data but does not provide a practical framework that can be used to develop strategies to deconstruct and counteract the issues being brought up (something that later works like D'Ignazio and Klein's book “Data Feminism” achieves).

However, it does produce early thoughts and perspectives that can be applied to the landscape of computational text analysis.

For example, the central idea of the book that data is “cooked” or in other words constructed, is a core topic discussed in computational text analysis (as well as by feminist scholars) and prompts all CTA work to be well documented and thought out of the ways that biases and assumptions influence it.

The book discusses the importance of context when it comes to understanding data which is very crucial and a hotly debated topic in the field.

It also touches not just on how humans influence data, but how data evolves at a point to almost mimic a live organism itself that beings to make us its subjects and influence and control us, which is important to consider in modern computational text analysis, especially in a feminist perspective, when algorithms and AI dominate the field to influence society in every manner. However, the book itself does not consider the current day situation of the topic.

Finally, the book draws on ideas from a diverse set of disciplines be It an anthropological account, an economic analysis, a humanities critique, scientific development, etc. That is important when applied to computational text analysis which itself is a subject that is highly interdisciplinary and can benefit from the incorporation of insights from different disciplines to provide a more nuanced understanding. Also, building a feminist text analysis itself requires us to bring diversity in the perspectives of participants.



This entry is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.

This entry was posted in Posts on May 14, 2023 [<https://femethods2023.commonsgc.cuny.edu/book-review-raw-data-is-an-oxymoron/>] by Muhammad Rakibul Islam (Rakib).

**CUNY ACADEMIC COMMONS**

**People** **Groups** **Sites** **Courses** **Events** **Activity** **About**

Help Privacy Terms of Service Creative Commons (CC) license unless otherwise noted

Built with WordPress Protected by Akismet

Powered by CUNY



