Book Review: Data Feminism by Catherine D'Ignazio and Lauren Klein

By: Nick Zufelt

Use case: Data Feminism is a book focused on the ways in which the theories and practices of intersectional feminism can and ought to be employed in the technical world, and how the powerful impacts of data practices can be harnessed to create more equitable outcomes for the world (and how they traditionally haven't). A perfect read for anyone or any courses studying, critiquing, or leveraging these data practices.

Suggested age range: 9th grade-adult

Special topics: Objectivity in Data Science (p.73–81), Reductionary data practices such as classification ("Questioning Classification Systems", p.105–111), and the goals of surveillance and other data systems ("Data Science with Whose Interests and Goals", p.39–47)

As data practices — the collection, organization, shaping, visualizing, and employment of data to impact people's understandings and decisions — become ever more ubiquitous, many people have come to recognize the inherent harms in established techniques and procedures. Examples abound, from election interference caused by mis/disinformation campaigns on social media to bias in AI training processes resulting in racist facial recognition algorithms. As a teacher, I hear from my students that they see this and feel this, particularly students of color, and yet our education system continues to churn out data practitioners who largely learn status quo techniques and create technology that continues to exploit and other. Are there other ways to think about the technology sector in our schools? How can we empower thinkers of the next generation to challenge and rethink oppressive technological systems?

Data Feminism offers us a possible way forward. This book, written by professors Catherine D'Ignazio and Lauren Klein, takes readers on a journey that builds historical context and shares concrete case studies to contrast data practices "in the wild" that either empower or disempower historically marginalized (or *minoritized*, as the authors suggest, p. 26) peoples. These case studies build into powerful principles that can be applied to any number of situations in which data practices are employed.

The book offers seven chapters, each focusing on a different principle in their newly coined field. For example, Chapter 3 ("Elevate emotion and embodiment") takes on the notion of objectivity in data science, while Chapter 6 ("Consider context") shows what can go wrong when the origins of the data are ignored and offers an alternative data storytelling experience. Each chapter can easily be read in any data-themed course, or even just a 5–7 page case study within those chapters. Below, I offer a selection of these case studies and how they could be deployed in a classroom to challenge traditional thinking about data practices. *(Importantly for teachers, it should be noted that the entire text of the book is available online through MIT Open Access.)*

The goals of surveillance and other data systems
"Data Science with Whose Interests and Goals", p.39–47, within Chapter 1

All technologists ought to learn to question the goals of the data systems they create and all people ought to analyze the interests of those choosing to create them. This is a concrete skill that empowers technologists to help answer the question, "should I help create this?"

This case study tells the stories of a Pennsylvania county who inequitably surveilled poorer communities to identify children at risk of child abuse, of a statistician at Target who built a pregnancy detection system (because pregnancy is a key moment in a shopper's lifetime to commit to lifetime brand loyalty), and of the notion that "data is the new oil" and how vast sums of money (i.e. power) is spent to secure access to it.

> *"This extractive system creates a profound asymmetry between who is collecting, storing, and analyzing data, and whose data is collected, stored, and analyzed."*
> *(p. 45)*

I think this pairs well with students in statistics and computer science classes who are taking on the task of selecting open-ended projects or analyzing the field of data science more generally. Why is some work done and not others? Who benefits (and who, instead, could benefit from other applications of data projects)? I also could see this case study empowering conversations around current events in the technology sector that happen regularly in the STEM classroom, which could default to feeling like powerless complaining sessions.

Objectivity in Data Science
Introduction to Chapter 3, p.73–81

One of the strengths of data practices is how they are received by others. We often want to teach our students to be critical readers of graphs and other data visualizations, because the natural inclination is to trust them blindly. You've doubtless heard people say "The numbers never lie", that data practices are powerful because they are objective and all-seeing. This "God Trick", as the authors attribute to feminist philosopher Donna Haraway, is employed to lull readers into a perhaps undeserved attribution of authority over the graphed situation. In contrast, teachers want students to learn to be critical consumers of data visualizations, a key skill in media literacy.

> *"[The god trick is] also the most ethically complicated to navigate for the ways in which it masks the people, the methods, the questions, and the messiness that lies behind clean lines and geometric shapes."*
> *(p. 76)*

I would consider using this case study in any class where students are expected to generate simple data visualizations: natural science or math classes, for instance, or even statistics classes. We teachers often force students to "tell the story of the data" or even as simple as "label the axes" without attributing much reasoning behind why this is important. This case study can help shed some light on why context matters.

Reductionary data practices
"Questioning Classification Systems", p.105–111

Think of any form you've recently filled out. Are you male or female? No, you have to select one answer, there are no other options. There is power inherent in the reductions made by those designing data systems and employing data practices. One example is discussed in this case study: a TSA agent pushes a button to select your gender before you enter an airport scanner, which determines if you're a threat (because the pre-trained model requires this information to compare your body to other bodies with "your" gender). It's a powerful look at the ways the choices that technologists make can have huge impacts on the users of their work.

> *"Because flawed classification systems—like the one that underlies the airport scanner's risk-detection algorithm…—are not only significant problems in themselves, but also symptoms of a more global condition of inequity."*
> *(p. 106)*

The pairing here with app development classes, or realistically any computer science or related course, is clear. This case study will get students thinking about the impacts of their code well beyond the text editor. One simple "if" statement can make a world of difference, for better or worse, in people's lives.