### Introduction

*Data Feminism* is a book that seeks to integrate existing work on feminism, data ethics, and data science. It provides information and it is also a call to action. *Data Feminism* is about considering systems of power in the world (particularly in terms of injustice), how technology can reflect and perpetuate those systems, and how we can do data science to change those systems. Rather than focusing strictly on gender, the authors define feminism as “shorthand for the diverse and wide-ranging projects that name and challenge sexism and other forces of oppression, as well as those which seek to create more just, equitable, and livable futures” (6-7). As the authors state their purpose: “A central aim of this book is to describe a form of intersectional feminism that takes the inequalities of the present moment as its starting point and begins its own work by asking: How can we use data to remake the world?” (5) There are two authors. Catherine D’Ignazio has experience as a freelance software developer and experimental artist and as a professor. She chose to write the book because “the hype around big data and AI is deafeningly male and white and technoheroic and the times is now to reframe that world with a feminist lens” and “my recent experience running a large, equity-focused hackathon taught me just how much people like me—basically, well-meaning liberal white people—are part of the problem in struggling for social justice” (9). Lauren F. Klein is a “professional nerd” who worked in software development, studied English in grad school “with a particular focus on early American literature and culture”, and works as a professor on research projects that “translate this history into contemporary concepts”. She chose to write the book to “show how the issues of racism and sexism that we see in data science today are by no means new” and to “help translate humanistic thinking into practice” (9).

### Summary

There are 7 principles listed in the book and one chapter for each principle. Examine power (chapter 1), challenge power (chapter 2), elevate emotion and embodiment (chapter 3), rethink binaries and hierarchies (chapter 4), embrace pluralism (chapter 5), consider context (chapter 6), and make labor visible (chapter 7). Each chapter includes a description of the principle, examples of the inequality that make the principle necessary, and examples of how the principle has been (and can be) applied.

### Response

Note that I have focused my response on the things that are new and/or interesting to me.

From chapter 2, I learned that it’s important to include (and defer to) people who experience/are familiar with oppression. I wonder, how can we engage with the community as part of our computing classes (data science, CS, statistics, etc.)? Would it be possible to partner with the community (local organizations, Calvin-affiliated organizations, etc.) so that the work we do provides a meaningful service to the community? Professor Arnold pointed out that work like that requires a relationship, and relationships can be difficult to develop. So perhaps as a simpler task, would it be possible to work with other departments in Calvin (e.g., Biology, English (for NLP)), giving us students a more interdisciplinary experience?

From chapter 3, I learned some creative examples of ways to show data and that the binary between emotion and reason can be harmful. I’d like to get more experience with showing data and really going into depth as to what methods of presentation are useful for conveying the message I want to convey.

From chapter 4, I learned that whether to count gender, when to count gender, and how to count gender is ethically complex. Think about context and power to determine “whose interests are being served by being counted, and who runs the risk of being harmed” (111). This applies to other classifications as well. I now have a better understanding for why it can be hard to get data with the detail of labeling that I might like and that “get more labeled data” is not a neutral idea.

Comments on chapter 5. I’ve learned a lot of technical skills about how to do data cleaning, but I don’t recall being taught much about the potential consequences of data cleaning. I think there should be more discussion of the consequences of data cleaning in our Statistics/Data Science education at Calvin. Another comment is that I hate the “lone genius” concept in general, and I’m glad they touched on why it is so harmful. (No one person can understand the situation perfectly, and particularly if you’re an outsider and go in assuming you know everything then you’re disregarding the knowledge and experience of the community.) I also appreciated the discussion of transparency and reflexivity. That helped me understand why it’s important for professors to share their background and good for students and professors to get to know each other so that we can understand the context that leads us to do what we do. The chapter also talked about co-liberation. I think the CS interim I did in the Dominican Republic had co-liberation in mind (even if Professor Schuurman wasn’t explicitly thinking that). I would say it met all the qualifications the authors listed in “Data for co-liberation” (just replace “data” with “computing” or “computer science”). (As a reminder, the qualifications are “Leadership by members of minoritized groups working in community, Money and resources managed by members of minoritized groups, Data owned and governed by the community, Quantitative data analysis “ground truthed” through a participatory, community-centered data analysis process, Data scientists are not rock stars and wizards, but rather facilitators and guides, Data education and knowledge transfer are part of the project design, Building social infrastructure--community solidarity and shared understanding--is part of the project design”.) In that interim, we went through the book “When Helping Hurts” and that book had a lot of the same concepts that Data Feminism does (consider context, work with the local community, etc.), so it’s also good to remember that the ideas expressed in Data Feminism are not new. Data Feminism collects, explains, and exemplifies existing feminist (using feminist here in the broad sense that authors use) work and applies it to a data science context.

From chapter 6, I learned the importance of context. Providing context is important, and I feel it often gets glossed over in class, by which I mean professors will point students to the documentation of a dataset and assume the students will read it and understand it independently. (That being said, providing students with the documentation is an important step in providing context and I do appreciate it.) This doesn’t happen every time or in every class, but it has happened many times in my experience at Calvin. When it comes to communicating context, I sometimes feel underprepared. I’ve gotten better at showing my work in a notebook format where I go through and have code and comments explaining my thought process so a reader can see all the work that went into my solutions. (This is what I do for my statistics classes.) But when it comes to something like producing a final report, I find that really hard. How do I reduce the work I’ve done and the considerations and exploration that went into it into a document that accurately represents the main points of my work (because even if I provide resources for more context, most people probably won’t read them)?

From chapter 7, I learned about invisible labor. I think there is invisible labor that goes into classes at Calvin. Sometimes the work the professors have to do of preparing the curriculum, doing grading, etc. is not always apparent to students (especially students new to college) and can lead to students having unfair expectations of professors. The authors also mentioned invisible labor of project management. In this class Professor Pruim thanked me for my project management work, but in other group projects I’ve done at Calvin I’ve done project management work that is not recognized (or not thanked) by the professor and/or the students. In part is because there is not an explicit “project management” role so there may not be a way built in to recognize that role, even though in many cases somebody needs to step up to fulfill that role.