

# Beyond Binary

Enhancing Data Analysis through  
Sex and Gender Disaggregation



Hailey Byrde and Tess Gompper

# Table of Contents

---

Introduction.....	page 1
1. Basics.....	page 4
2. How do we sex and gender disaggregate our data?.....	page 11
3. Being Inclusive.....	page 15
4. Think Critical.....	page 24
5. Taxonomy.....	page 27
6. Further Reading.....	page 31
7. About the Authors.....	page 34
Endnotes.....	page 35
Bibliography.....	page 38

# Introduction

**Do you work with human-centric data? Are you interested in making your data practices more inclusive, equitable and ethical? If so, this toolkit is for you!**

**We invite you to engage with our toolkit and integrate these feminist and queer practices into your work, life, and beyond!**

In this toolkit, we intend to demonstrate how **sex and gender data disaggregation**, or the process of categorizing and separating data according to sex or gender, is a vital practice in making **human-centric data more representative of our social world**. We will provide you with a concrete understanding of what sex and gender disaggregation is, what it entails, and how we can incorporate this practice into every step of the data process. We aim for our guide to **incorporate feminist and queer data approaches** to provide suggestions on how to inclusively represent transgender people as well as folks **beyond the binary**. Our guide encourages critical thinking and consideration of data, especially when it comes to questioning our dominant societal structures that “other” and marginalize women and queer people. We will supply you with the necessary inclusive vocabulary to talk about sex and gender-disaggregated data, as well as provide you with further resources that detail certain aspects of our discussion.

Any human-centric data collectors and analysts need to be aware of sex and gender disaggregation processes and need to understand why they are important. **Surveying of any kind about individual people must collect inclusive, responsible sex and/or gender data so that disaggregation can be effectively carried out and individuals can be represented to the fullest extent.**

We believe sex and gender disaggregation of data is fundamentally important for equitable and inclusive research, specifically research in the humanities or policy-making fields. As Caroline Criado Perez describes in her book Invisible Women, "seeing men as the human default is fundamental to the structure of human society"<sup>1</sup> and how in many cases the "**male default can be both a cause and a consequence of the gender data gap**"<sup>2</sup>. In order to look beyond this male default, we need to sex and gender disaggregate our data.

We intend for this toolkit to be used as a jumping-off point, not an end-all-be-all definitive scope of the topic. There is so much research around data feminism and queer data right now, and there is still so much more to be done. We provide a general overview and introduction of the topic and a guide on how it can be incorporated into data practices in different ways, but these ideas are just the start!

# 1. Basics

## What is it?

The sex and gender disaggregation of data is the separation and categorization of human-centric data by sex and/or gender which allows for gender-specific data analysis.

"Any data on individuals broken down by sex. Gender statistics rely on these sex-disaggregated data and reflect the realities of the lives of women and men and policy issues relating to gender."

- European Institute for Gender Equality<sup>3</sup>

"Gender-differentiated data and information must be available for policymakers to be able to assess the situation and develop appropriate, evidence-based responses and policies"

-Organization for Economic Co-operation and Development<sup>4</sup>

## When do we use it?

People of different genders experience and are affected by the world in different ways and data must reflect that. In an ideal world, all open, human-centric data will include sex and or gender data to provide the opportunity to be disaggregated.

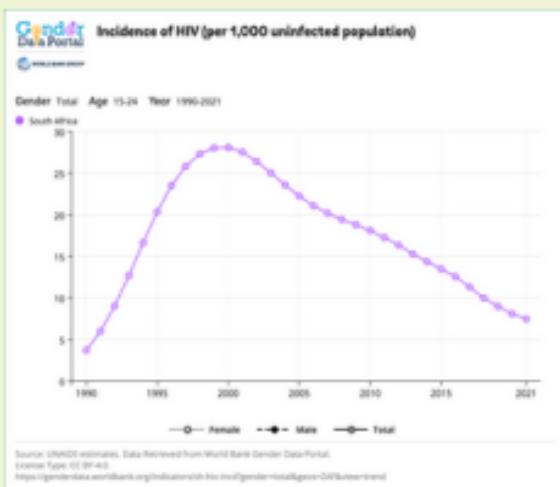
## Why do we use it?

Sex and gender disaggregation of data “[allows] for the measurement of differences between women and men on various social and economic dimensions and are one of the requirements in obtaining gender statistics.”<sup>5</sup>

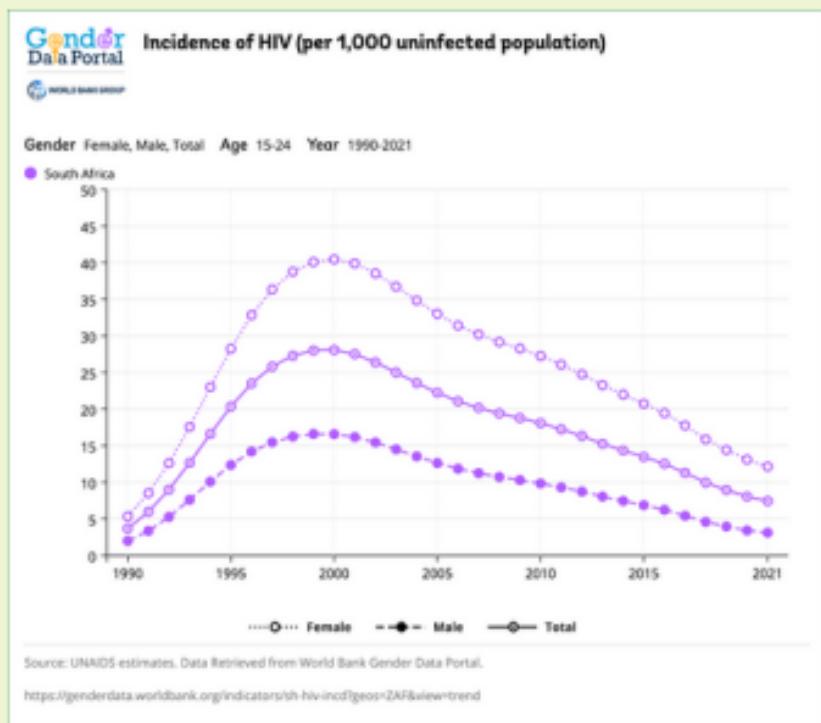
We need sex and gender-disaggregated data to close the **Gender Data Gap**. Living in a patriarchal society, data tends to be centred around the male subject, especially since it is men who dominate the data sector. This normalization of the cis-male in data frames them as the “default” human and effectively marginalizes women and gender non-conforming individuals when data is biased towards men<sup>6</sup>. This leads to an inevitable gender gap in our data where those who are not men are either left out of data completely, underrepresented, or not accurately represented. This has serious implications for policy, public life, and health outcomes when the experiences of different genders are not taken into account.

Sex and gender disaggregation in data is therefore a crucial process that can help close this gap and provide us with data that gives the full picture and accurately captures the experiences of women and gender non-conforming individuals. Sex and gender data disaggregation can thus work to minimize data bias towards men while providing a fuller and more accurate representation of our social world and gendered differences.

Sex or gender-disaggregated data is incredibly meaningful, and can actually make data look different. For example, in looking at the incidence of HIV in South Africa between 1990-2021 for the age group of 15-24<sup>7</sup>, the non-disaggregated data does not show the full picture.



However, once the data for the total population is sex-disaggregated, we can see a significant disparity between the number of women versus men who are affected by HIV. Women have historically had and currently have a notably higher incidence of HIV than men, and when this data is not sex-disaggregated, as demonstrated in the first graph, this difference is rendered invisible.



Thus, sex and gender disaggregated data is incredibly important for understanding different gendered impacts of experiences! In seeing this gendered difference through the utilization of sex-disaggregated data, researchers can then be prompted to consider why this gendered difference exists and which societal structures could be impacting this. In this example, researchers could consider and explore different drivers of higher incidences of HIV in young women, such as sexual violence, lack of sexual education, or inadequate healthcare, which could then be remedied in policy intervention. Disaggregating data according to sex and gender allows us to observe these significant gaps in experience.

# What if My Project Doesn't Need Sex and Gender Data?

There may be instances where you and your team feel your data project would in no way benefit from collecting respondents' sex and/or gender. If this is the case, **we encourage you to:**

1

Consider the social and political contexts surrounding the data you are collecting and whether or not there is any possibility at all that people may report different experiences due to their gender.

2

Consider who may want to use your dataset in the future. Particularly if you are creating an open-access dataset, you should keep in mind that researchers who use your data in the future may greatly benefit from having access to sex and gender data in relation to the dataset. There may be researchers looking specifically for this data, or researchers who may find that their research question is, in fact, affected by sex and gender.

**So collect the data, you never know when you, or others, will need it!**

Collecting sex and gender data is not only for projects that are asking gender-specific questions, it's about collecting the data knowing you are **creating the potential to reveal gender disparities in our world.**

Remember! It is beneficial to record sex and/or gender when collecting data since **you can always choose not to include it** in your analysis! If you don't record sex or gender in the collection process, you may be missing the only opportunity to collect this data in this context.

# ★ 2. How do we Sex and Gender Disaggregate our Data?

## The Data Science Process

We can integrate sex and gender disaggregation principles, as well as feminist and queer thought at every step of the data science process:

### 1. Question Formulation

The first part of the data science process is always the formulation of a research question as it shapes the rest of the process. Here is where it is important for your question to consider that people of different sexes and genders experience the world in different ways and your data project needs to reflect this consideration. Where possible, formulate your research questions in a way that accounts for gendered differences in social experiences.

A key point to remember in the question formulation stage is: “Chances are you’re not going to be the (only) one to use the information”. Here it is important to remember that even if you absolutely do not feel it necessary to collect sex and gender data, there are likely to be others who come along after you who would greatly benefit from it.

## 2. Data Collection

The data collection stage involves the “obtaining and arranging data so it can be worked with”, in other words; it is the collection and preparation of raw data for analysis. It consists of deciding what data to collect, obtaining the raw data, cleaning the data, and organizing the data<sup>10</sup>.

In the collection process, we encourage data collectors to collect sex and gender data from respondents in an inclusive, sensitive, and respectful manner (discussed further in section 3: Inclusivity) in addition to the rest of your desired data.

In the cleaning phase, it is important that collectors do not delete or modify data based on respondents' self-declared identity. We instead suggest making note of potential question areas or typos to be further considered in the data annotation phase.

### REMEMBER!

\* Researchers are not all knowing and may not be aware of all gender identities so it is important they don't delete anything based on personal conceptions. Consult with others! Specifically women, gender-nonconforming, or queer people. \*

### 3. Data Annotation

The goal of annotation is typically to “accurately and reproducibly assign labels to a set of data objects”<sup>11</sup>. Here, data experts, along with members of the affected communities, should consider typos and different reported gender identities to create an inclusive set of categories.

### 4. Data Analysis

Because of the work done in the previous 3 steps, the data can now be effectively disaggregated by sex and/or gender. Analyze your data to come to separate conclusions for each gender or sex group so that ...

### 5. Data Analysis

... you can assess how the research topic affects each group differently!

## 6. Communication

Finally, in presenting the results of your study, you can now effectively communicate potential sex or gender-based inequalities and observations to stakeholders and the community.

### REMEMBER!

In all stages be gender sensitive and inclusive! Always offer opportunities for respondents to opt-out of sharing their sex and gender data.

\* In all stages of the process it is important to include women and genderqueer people in order to bring to light certain issues that men may not even think about or to offer contextual know-how.<sup>8</sup>\*

# 3. Being Inclusive

## Sex & Gender

In order to benefit and meet the needs of those whom you collect or analyze data about, it is vital to “**know who the people are and what their needs are**”<sup>12</sup>. Here it is key to remember that there are more people than just those who fit within the gender binary and that “collection methods can ensure that particular identities, or ways of thinking about identity, are designed out of the process in ways that further compound the invisibility of the most marginalized among LGBTQ people”<sup>13</sup>. In order to make marginalized people visible, data collectors must **challenge “the cis-heteronormative assumptions built into data collection methods”**<sup>14</sup> and ask inclusive questions about sex and gender.

The definitions of sex and gender are not concrete or universal! The concepts of sex and gender rely on one another and shape what each word means<sup>15</sup>. **Gender describes the more social and personal aspects of identity**, and is often based on sex characteristics. Gender has to do with processes of socialization and performativity, it is enacted on an everyday basis! Sex is often conceptualized to refer specifically to biological characteristics such as reproductive functions, but it is also intertwined with self-identity, and **just as gender does, it can exceed the binary categorizations of “male” and “female”.**

It is important to ask yourself whether you need to collect data on sex, gender, or both. New Zealand's Stats NZ tells us, "in most cases, a person's gender-- their social and personal identity-- is most relevant for policy making and research rather than their sex at birth, with gender-based analysis 'used in a range of areas, from income inequality to health and education.'"<sup>16</sup> So, often collecting data on a person's expressed gender is enough for representation of that individual.

In some cases, however, you may find, for whatever reason, that you only need to collect data on a person's sex. This could happen, for example, when we ask questions about medical care, medical funding, or biological characteristics. In these cases, **it remains important to be respectful and inclusive of individuals' sex identities such as cis, trans or intersex identities** as we will see in the following section.

# How do We Collect Data Inclusively?

You may be wondering how you can inclusively collect gender data while still having viable, relatively clean raw data, given the large number of potential gender expressions respondents may have. Let's discuss.

gen|*derfluid*

gen*derfluid*

gen*der nonconforming*

gen*derqueer*

Here we recommend using **autofill** text boxes where possible, as detailed in Queer Data<sup>17</sup> (See Chapter 6: Further Reading). The autofill suggestion method allows respondents to submit whatever their gender expression may be while mitigating typos and encouraging relatively standardized responses.

Non-exhaustive (you should never limit responses!) response options may include the following:

- Cis woman
- Cis Man
- Genderfluid
- Genderqueer
- Gender nonconforming
- Agender
- Woman
- Man
- Non-binary
- trans woman
- trans man
- \*

If you are collecting data on sex, you may consider utilizing a similar format with the following response options:

- Female
- Male
- Intersex
- Cis female
- Cis male
- Trans female
- Trans male
- \*

\*There are many more sexes and gender identities and you should **always consult educated members of the communities** of which you are collecting data to provide as many relevant options as possible, while also leaving open-ended response options.

This technique can create a large number of categories upon sex or gender disaggregation, likely with some categories containing very few data points. It is acceptable, and often necessary from a data perspective, to combine certain categories in order to provide substantive and relevant insights for these groups of people. However, in combining sex or gender categories, you must consult community members so as not to aggregate groups in inappropriate ways.

**For example:** You must not categorize all respondents who do not identify as “male” or “female” into a third, “other” group.

# Trans Visibility

Transgender-related sex and gender data offers more inclusivity questions for us to consider. In some cases, whether someone identifies as transgender is irrelevant to the data being collected, however in other contexts, it may be relevant. In some contexts, both women and trans women experience the world in the same way, such as in experiencing oppression from men. However in other contexts, such as medical funding, trans status may be relevant.

We recommend offering respondents an opportunity to identify their sex or gender **as they wish** and offering a second question to determine if they have a **trans history** if it is pertinent to the project. For example:

Do you identify as transgender or as someone with a trans history?

- Yes
- No
- Prefer not to say

We advise against a “two-question” approach that first asks respondents to report their gender identity and then asks for their sex assigned at birth, as it inappropriately asks the respondent to report their sex at birth and can position the respondent as “not really” male or female<sup>10</sup>. Asking participants if they currently or have previously identified as trans is a less invasive way of deducing if a respondent is trans by allowing them the opportunity to self-identify as such. While this means of asking for this information is still not perfect, in the sense that it still assumes cis as the “default” and trans as the “other”, it is less assumptive and gives the respondent more agency than the “two-question” approach.

We would also advise against asking about sex and gender in this format:

What is your gender?

- Male
- Female
- Other (please specify): \_\_\_\_\_

This approach similarly enforces the gender binary by normalizing “male” and “female” and marginalizes gender non-conforming individuals by framing them as “others”.

## Remember!

Identities are often **fluid** and can change in different ways over time, both for individuals and societies as a whole. On the individual level, one way we can account for the fluidity of identity in data is to allow respondents opportunities to **update or modify their information** where possible.<sup>19</sup> This practice can also provide researchers and analysts with the opportunity to explore how identities and conceptions around them can change over time!

It is important to understand that data is reflective of our current conditions and **can always be reinterpreted!** Our current conceptions of gender and transness may not be the same in the future, this data is merely a representation of our current world, our perceptions of it, and our biases. (More about this in section 4: Think Critically!)

# Be Sensitive with Sensitive Data

For some individuals, sex and gender information is sensitive. This could be for many reasons, in some cases a person may not be “out” as their preferred identity and may fear exposure. In many places in the world, identity expressions outside of the male/female binary are not accepted and a person may rightfully fear societal or governmental persecution. Thus it is critical to:

1

**Explain to respondents the reason for and importance of collecting their sex and gender data.** Ensure respondents that their data will be used anonymously and responsibly in ways that will not hurt them (this should be the case!). Explain to respondents that by collecting this sex and gender information, the data from your project will be better able to serve them and their communities, for example via appropriate resource allocation or serving of needs based on this comprehensive view.

2

**Always offer an opt-out option** for respondents with regard to sex and gender questions. This could be in the form of a “Prefer not to say” response option.

# Inclusive Teams Create Inclusive Data

**"When data teams are primarily composed of people from dominant groups, those perspectives come to exert outsized influence on the decisions being made—to the exclusion of other identities and perspectives"** -Data Feminism<sup>20</sup>

Data Feminism tells us that **"data scientists must proceed with awareness of context...and analysis of power in the collection environment ... to determine whose interests are being served by being counted, and who runs the risk of being harmed"**<sup>21</sup>. Those who most know the context, interests to be served and the risks of harm are those who are in these minority groups themselves. Thus **it is essential to include women and gender non-conforming people in the data collection and analysis processes in order to avoid falling prey to privilege hazards.**

**Privilege hazard** describes the impacts of a decision-making privileged person who may lack the ability to notice when they are committing harm, whether it be due to lack of expertise or lack of lived experience within a marginalized group<sup>22</sup>.

# 4. Think Critical

Data can and should be critically engaged with as interpretations of our social world, **these numbers are representative of individuals, but they may not tell the whole story!** We must take opportunities to reinterpret data and reconsider what data aims to accomplish<sup>23</sup>. Data should intend to “**benefit those about whom data relates**”<sup>24</sup>, so we must approach data processes with this in mind, and critically assess certain aspects of the data process.

In creating more inclusive and equitable data, it is important to think critically about the data collected and your role in analyzing, collecting or cleaning data. There are many decisions made in every step of the data process, and these **decisions are not neutral!** Decisions made reflect our dominant social structures, as well as our specific positioning regarding privilege and identity. Thus, these decisions are biased in the sense that they are always reflective of a certain time and place, and this is important to remember. And certain judgements can shape the results of data and how it is interpreted<sup>25</sup>.

Providing transparency in all steps of the data process can work to combat the “naturalization” of data’s journey from collection to use. If users are given insight into what decisions were made and why, they can better understand what may have influenced the decisions and how they shape the data and conclusions drawn from it.

"Queer Data encourages researchers, practitioners, and activists to think about data differently, and ask critical questions such as '**Why do we collect data this way?**', '**Whose interests does data serve?**' and '**Why do we collect data at all?**'"

- Queer Data<sup>26</sup>

"Who benefits from data science (and who is either overlooked or actively harmed)? ... there are groups of people who are disproportionately benefitting from data science and there are groups of people who are disproportionately harmed"

- Data Feminism<sup>27</sup>

It is necessary to consider bias at every step in the data process, as **biases that influence the collection of data "continue as data progresses through analysis."**<sup>28</sup> Assess your privileges and unintended biases. In thinking about **privilege hazard**, it can be helpful to **work with or seek the insight of those who are different from you**, as they may be able to point out gaps and shortcomings where you may not notice them!

Finally, it is important to remember that data collectors are not entitled to individuals' data! Information isn't owed to anyone, thus, it should always be an option to say no or withhold responses to certain questions.<sup>29</sup> **Data practices should always be clear and consensual on what data is being extracted.**

To engage in a critical examination of data processes, here is a (non-exhaustive) list of questions to consider when collecting, analyzing, annotating, or engaging with someone else's data.

**Overall:**

- Does this data project create more good than harm?
- Will this data accurately represent those from whom the data was taken?
- Is the process behind this data transparent? If not, why not?
- Is this project exclusionary or deliberately harmful to any certain group?

**Question Formulation:**

- Why are we researching this?
- What is the aim of this project?
- What dominant structures or biases do these questions reflect?
- Whose interests will this data serve?

**Data Collection:**

- Why are we collecting this data?
- Who benefits from the collection of this data?
- Will the collection of this data actively help the population whom the data was taken from?

**Annotation & Analysis:**

- Who is making decisions about who and what gets included in data?
- Why were these decisions made, are they reflective of bias?
- Is anyone being overlooked or harmed by this data?

# 5. Taxonomy

This taxonomy is intended to equip you with the relevant vocabulary to inclusively talk about sex and gender disaggregated data and the issues it can address!

## Cis

Describes someone who identifies with or expresses the gender they were assigned at birth

## Data feminism

"A way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought"; "the work of data feminism is first to fine-tune into how standard practices in data science serve to reinforce [existing inequalities] and second to use data science to challenge and change the distribution of power"<sup>30</sup>.

## Gender

"Gender is used to describe a person's social and personal identity as a man, woman, something between or beyond these concepts."<sup>31</sup>

*For more: see Data Feminism  
by Catherine D'Ignazio and  
Lauren F. Klein*

## Gender Binary

The construction of sex and gender as two, distinct and opposite, poles, categorizing everyone as either male or female.

## Gender Data Gap

Describes a gap in our knowledge surrounding women and gender-non-conforming individuals due to the cultural dominance of men that constructs them as the “default” of humanity.

The “absent-presence” of women in our culture; “both a cause and a consequence of the type of unthinking that conceives of humanity as almost exclusively male”; “a gap in our knowledge that is at the root of perpetual, systemic discrimination against women, and that has created a pervasive but invisible bias with a profound effect on women’s lives”.<sup>32</sup>

For more... See *Invisible Women* by Caroline Criado-Perez

## Gender non-conforming

“A person who doesn’t adhere to societal pressures to conform to gender norms and roles.”<sup>33</sup>

## Heteronormative

The presumption, normalization, and privileging of heterosexuality.

## Male Default

When the male experience is seen as universal, and as the “default” of humanity. It is the result of male dominance in our patriarchal society and is both a cause and consequence of the gender data gap.



The Introduction of  
*Invisible Women* has  
a great explanation of  
this!

## Privilege Hazard

Those in privileged positions are less likely to notice that they are in such a position, and are thus less likely and less equipped to notice instances of oppression, marginalization, and harm.

## Queer

“An umbrella term used to refer to lesbian, gay, bisexual and transgender people. Some use as an alternative to “LGBT” in an effort to be more inclusive. Depending on the user, the term has either a derogatory or an affirming connotation, as many within the LGBT community have sought to reclaim the term that was once widely used in a negative way.”<sup>34</sup>

## **Queer data**

"Queer data is more than using data to tell stories about the lives and experiences of LGBTQ individuals: the presentation of data is also an opportunity for LGBTQ people to see themselves reflected, although this mirror image is a never truly accurate representation"; data capturing "sexual orientation, gender identity, and trans status/history". Also refers to the process of queering data processes, changing the way that things are done and questioning the norms and "natural" processes of data<sup>35</sup>.

*for more: see Queer Data by Kevin Guyan*

## **Sex**

"Sex is an identity based on primary and secondary sex characteristics, such as genitalia, reproductive functions, hormones, breasts and facial hair"; "Consideration of the sexed body in isolation is never enough"; Kevin Guyan's "definition of sex... accommodates the binary poles of female/male as well as space between and beyond those poles"<sup>36</sup>.

## **Sex and Gender Data Disaggregation**

The process of separating and categorizing human-centric data by sex and/or gender. This practice then allows for gender-specific data analysis.

# 6. Further Reading

## Invisible Women

London based writer, broadcaster and feminist activist Caroline Criado-Perez brings us "a provocative examination of how a gender gap in data perpetuates bias and disadvantages women"<sup>37</sup>. Invisible Women exposes the "default male" in our society's data through a wide variety of examples and shows us the appalling effects this default has on women.

We recommend Invisible Women to those who are looking for real-life examples of the gender data gap and scenarios where sex disaggregation was not used when it should have been. Criado-Perez provides a deep insight into how the construction of the "default male" is pervasive across cultures, time periods and most of all, data.

## Queer Data

Kevin Guyan, a Research Fellow in the School of Culture and Creative Arts at the University of Glasgow, dives into the importance of better collecting, analyzing and understanding queer data. Through examples of current data practices, such as UK Censuses, Guyan shows readers the harms these practices perpetuate for LGBTQ people and what we can do about it. Queer Data "shows how greater knowledge about queer identities is instrumental in informing decisions about resource allocation, changes to legislation, access to services, representation and visibility"<sup>38</sup>.

We recommend Queer Data to those looking to learn more about inclusivity within the collection of trans data, sexuality data, and gender spectrum data, particularly with respect to policy making.

## Data Feminism

American professors, Catherine D'Ignazio and Lauren F. Klein, bring us Data Feminism, a comprehensive lesson in data science and data ethics informed by intersectional feminist thought. D'Ignazio and Klein challenge the male/female binary in data to explore power: "who has it, who doesn't," and "how those differentials of power can be challenged and changed" and to change the white, male, "techno-heroic" narrative around big data and data science<sup>39</sup>.

We recommend Data Feminism to anyone who works with data and to those who wish to explore making data a more inclusive space; for those who work in it as well as those affected by it. Data Feminism is a useful guide on gender spectrum inclusivity, and the importance of more equitable data practices. We recommend the book to help anyone looking to challenge the binary classification of gendered data and to critically examine who data serves.

## The Feminist Data Manifest-No

Authors Marika Cifor, Patricia Garcia, along with TL Cowan, Jasmine Rault, Tonia Sutherland, Anita Say Chan, Jennifer Rode, Anna Laura Hoffmann, Niloufar Salehi, and Lisa Nakamura document a manifesto of refusals and commitments in current and future data practices. The 32 declarations of the manifesto condemn current harmful and reductionist data practices and commit to a more partial and relational understanding of data, one that understands that data is an interpretation of our social world at a specific point in time and can and should be re-interpreted.

We recommend The Feminist Data Manifest-No to anyone who is interested in data ethics and is looking for a (non-exhaustive) list of ways to commit to data feminism in several different stages of data collection and dispersal.

## The World Bank's Gender Data Portal

The World Bank's Gender Data Portal provides a database of (majority) sex-disaggregated data and data on gendered impacts on a global scale. The database is highly interactive, offering data on a variety of indicators relating to global development, and includes filters and customizable visualizations. The portal offers narratives as well to contextualize the gender data and the indicators offered. It also provides a resources repository that consists of guidelines for collecting, using and disseminating gender related data and links other datasets and portals where gender related data can be found.

We recommend the Gender Data Portal to anyone who is interested in current gender-related data in relation to global development and the UN's Sustainable Development Goals. This tool is especially helpful in understanding how collecting and disseminating sex-disaggregated data is incredibly important in shaping policy!

## Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning

This article by Susan Leavy, Assistant professor at University College Dublin School of Information and Communication, explores how the dominance of men in machine learning design perpetuates and reproduces gender bias in artificial intelligence. Leavy provides a critical examination of machine learning practices and the embeddedness of gender bias in our language, and calls for the presence of women in the processes of machine learning, as they are better suited to identify and remedy these biases.

We recommend Gender bias in artificial intelligence to anyone interested in critical examinations of our current data practices and how they are influenced and biased by our dominant social structures.

# 7. About the Authors

## Tess Gompper (she/her)

Tess is a Computer Science and Communication Studies student at McGill University, graduating in Spring 2024. Post-grad, she will be working as a data and software engineer and hopes to further spread the feminist, queer, and inclusive data practices detailed in this guide.

## Hailey Byrde (she/her)

Hailey is an International Development, Communication Studies, and Art History student at McGill University, graduating in Spring 2024. Post-grad she is planning on continuing her education by exploring the intersections of Queer-centered approaches, organizing, and global development.

# Notes

1. Caroline Criado-Perez, *Invisible Women: Data Bias in a World Designed for Men* ( New York, NY: Abrams Press, 2019), 6.
2. Criado-Perez, 2.
3. "Sex-Disaggregated Data," European Institute for Gender Equality, June 15, 2017, [https://eige.europa.eu/gender-mainstreaming/tools-methods/sex-disaggregated-data?language\\_content\\_entity=en](https://eige.europa.eu/gender-mainstreaming/tools-methods/sex-disaggregated-data?language_content_entity=en).
4. "Gender-Disaggregated Data - Organisation for Economic Co-Operation and Development," Toolkit for Mainstreaming and Implementing Gender Equality, accessed April 7, 2024, <https://www.oecd.org/gender/governance/toolkit/government/assessment-of-gender-impact/disaggregated-data/>.
5. "Sex-Disaggregated Data." European Institute for Gender Equality.
6. Criado-Perez, *Invisible Women*, 1.
7. "Incidence of HIV (per 1,000 Uninfected Population)," The World Bank Gender Data Portal, accessed April 7, 2024, <https://genderdata.worldbank.org/indicators/sh-hiv-incd/?gender=total&geos=ZAF&view=trend>.
8. Criado-Perez, *Invisible Women*.
9. Derek Ruths, "What does a good question look like?" COMP 370: Introduction to Data Science (class lecture, McGill University, Montréal, QC, October 2023).
10. Derek Ruths, "Data collection and annotation," COMP 370: Introduction to Data Science (class lecture, McGill University, Montréal, QC, September 2023).
11. Derek Ruths, "Annotation & Typology," COMP 370: Introduction to Data Science (class lecture, McGill University, Montréal, QC, November 2023).
12. Kevin Guyan, *Queer Data: Using gender, sex and sexuality data for action* (London: Bloomsbury Academic, 2022) 154.

13. Guyan, 71.
14. Guyan, 46.
15. Guyan, 9.
16. Guyan, 107.
17. Guyan, 124.
18. Guyan, 108.
19. Guyan, 55.
20. Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: The MIT Press, 2020).
21. D'Ignazio and Klein, 111.
22. Guyan, *Queer Data*, 153-154
23. M. Cifor, P. Garcia, T.L. Cowan, J. Rault, T. Sutherland, A. Chan, J. Rode, A.L. Hoffmann, N. Salehi, and L. Nakamura, "Feminist Data Manifest-No," 2019, accessed April 7, 2024, <https://www.manifestno.com/>.
24. Guyan, *Queer Data*, 180.
25. Guyan. 120.
26. Guyan, 5.
27. D'Ignazio and Klein, *Data Feminism*, 26.
28. Guyan, *Queer Data*, 145.
29. Guyan. 179.
30. D'Ignazio and Klein, *Data Feminism*, 8-9.
31. Guyan, *Queer Data*, 8.
32. Criado-Perex, Invisible Women, XIX.

33. "Terminology Related to Transgender and Gender Non-Conforming Identity," Anti-Defamation League, 2018, <https://www.adl.org/sites/default/files/terminology-related-to-transgender-and-gender-non-conforming-identity.pdf>.

34. Ibid.

35. Guyan, *Queer Data*, 7.

36. Guyan, 9.

37. Criado-Perez, *Invisible Women*, inside book cover.

38. Guyan, *Queer Data*, back book cover.

39. D'Ignazio and Klein, *Data Feminism*, inside book cover.

# Bibliography

Criado-Perez, Caroline. *Invisible Women: Data bias in a world designed for men*. New York, NY: Abrams Press, 2019.

D'Ignazio, Catherine, and Lauren F. Klein. *Data feminism*. Cambridge, MA: The MIT Press, 2020.

"Gender-Disaggregated Data - Organisation for Economic Co-Operation and Development." Toolkit for Mainstreaming and Implementing Gender Equality. Accessed April 7, 2024.

<https://www.oecd.org/gender/governance/toolkit/government/assessment-of-gender-impact/disaggregated-data/>.

Guyan, Kevin. *Queer Data: Using gender, sex and sexuality data for action*. London: Bloomsbury Academic, 2022.

"Incidence of HIV (per 1,000 Uninfected Population)." The World Bank Gender Data Portal. Accessed April 7, 2024.

<https://genderdata.worldbank.org/indicators/sh-hiv-incd/?gender=total&geos=ZAF&view=trend>.

Ruths, Derek. "Annotation and Typology." COMP 370: Introduction to Data Science. Class lecture at McGill University, Montréal, QC, September, 2023.

Ruths, Derek. "Data Collection and Annotation." COMP 370: Introduction to Data Science. Class lecture at McGill University, Montréal, QC, September, 2023.

Ruths, Derek. "What does a good question look like?" COMP 370: Introduction to Data Science. Class lecture at McGill University, Montréal, QC, October, 2023.

"Sex-Disaggregated Data." European Institute for Gender Equality, June 15, 2017. [https://eige.europa.eu/gender-mainstreaming/tools-methods/sex-disaggregated-data?language\\_content\\_entity=en](https://eige.europa.eu/gender-mainstreaming/tools-methods/sex-disaggregated-data?language_content_entity=en).

"Terminology Related to Transgender and Gender Non-Conforming Identity." Anti-Defamation League, 2018.

<https://www.adl.org/sites/default/files/terminology-related-to-transgender-and-gender-non-conforming-identity.pdf>.

