Evocative visualization of void and fluidity

by

Tomiko Karino

A master's capstone project submitted to the Graduate Faculty in Data Analysis and Visualization in partial fulfillment of the requirements for the degree of Master of Science, The City University of New York

Evocative visualization of void and fluidity

by

Tomiko Karino

This manuscript has been read and accepted for the Graduate Faculty in Data Analysis and Visualization in satisfaction of the capstone project requirement for the degree of Master of Science.

Date	Aucher Serr Capstone Project Advisor
Date	Matthew Gold Program Director

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Evocative visualization of void and fluidity

by

Tomiko Karino

Advisor: Aucher Serr

In the last few years alone, we have seen a significant increase in a demand for inclusion and diversity in every sector in our society. While there are numerous data visualization projects that point out and raise concerns about the lack of diversity, they often merely illustrate the disparity. By using the permanent collection data of the Museum of Modern Art, this capstone project explores ways to creatively visualize gender disparity among the artists in the museum's collection and questions what makes evocative visualizations that make an impression on the audience. It examines how to create visualizations that not only illustrate disparity but also visualize a void. It then further delves into the question of the accuracy of gender data as the definition of gender quickly evolves in the United States and contemplates challenges of visualizing such data. It also touches on the regional tension that may arise as the definition of gender primarily evolves in the Global North. It advocates for the need to continue collecting and analyzing gender data in a conventional manner in order to produce disaggregated data to identify gender inequality.

Link to the project: https://tomikokarino.github.io/Capstone/

ACKNOWLEDGEMENTS

I would like to thank professor Aucher Serr for her positive and gentle mentorship, especially when I was struggling to put my thoughts into a concrete project. Her encouragement and feedback undoubtedly uplifted me and gave me the confidence to proceed and complete the project.

I am grateful for Jason Nielsen, Filipa Calado and Rafael Davis Portela for their administrative support and knowledge of the CUNY system, as well as their dedication to helping students like myself to succeed in the program.

Among the faculty members that inspired me during the two and a half years I had studied at the Graduate Center, I would like to show my gratitude to two professors: Ellie Frymire and Katherine Baher.

Everything I know about d3.js, I learned in professor Frymire's classes. The practical coding lessons were definitely challenging, but reminded me how much I love to code. Her friendly demeanor and pragmatic approach made her classes some of the best that I have experienced in my entire academic career. Learning a new programming language and finding technical solutions gave me a sense of fulfillment during the pandemic when I often felt powerless.

While the assigned readings in professor Baher's class were arguably the most esoteric, they were also most intellectually stimulating and inspiring. The theoretical frameworks that I had learned in her class fundamentally changed the way I think about data and how it influences our society, and will stay with me for the rest of my life. The four months in the Spring 2021 semester when I was in her class, which was the first

winter months since the pandemic had begun, would have been much darker if it were not for the readings and discussions in her class.

I would also like to thank Katrin Bichler, to whom I was introduced by professor Serr in early 2023. As a visual designer, she has also been working with gender-related data, and has generously agreed to talk to me about my project. The conversation with her helped to foment my ideas and to realize my own subconscious bias caused by the binary gender system.

I would like to acknowledge that the class discussions that I had with my peers at the Graduate Center have been truly exciting and transformative. To be exposed to the fellow students' diverse backgrounds and points of view enriched my learning and vision for this project.

Lastly, I am grateful for my friends, who have supported me through not only the capstone process but also the entire program. Without their patiently listening to my babbling about the undeveloped ideas and kind words and encouragement that gave me confidence, I could not have completed the project.

DEDICATION

To my mother’s resilience, stoicism, and warmth.

TABLE OF CONTENTS

LIST OF FIGURES

DIGITAL MANIFEST

I. Capstone Project Whitepaper (PDF)

II. WARC File

A. Project Website

Archived version of https://tomikokarino.github.io/Capstone/index.html

III. Code and other deliverables

Zip file containing the contents of GitHub repository at the time of deposit.

(https://github.com/tomikokarino/Capstone)

A NOTE ON TECHNICAL SPECIFICATIONS

This capstone project is a web page, hosted on the GitHub Pages: https://tomikokarino.github.io/Capstone/index.html It is built with HTML, CSS and JavaScript. For the visualization, I used D3.js library and enter-view.js library to determine the cursor location to start the animation of the bubble chart. For the scrollytelling function, I used Scrollama.js library.

The GitHub repository can be found at https://github.com/tomikokarino/Capstone, and contains the following files and folders:

- Index.html - HTML file that contains the structure of the webpage and the narrative text and images.
- main.js - Loads the MoMAArtists data to render visualizations by using D3.js and enter-view libraries. Functions contained are predefined d3 and enter-view functions.
- Scroll.js - Creates scrollytelling behavior by using Scrollama library. Functions contained are predefined Scrollama functions.
- Style.css - Contains css styles of elements on the page
- /Data - Folder that contains MoMAArtists.csv file
- /FirstNames - Folder that contains firstNames.mp4 file
- /Images - Folder that contains images embedded in the HTML.

LIST OF VARIABLES

MoMAArtists.csv

ConstituentID	A unique number assigned to each artist, integer
FirstName	Artist's first name, string
DisplayName	Artist's last name, string
ArtistBio	Place of artist's birth and the years they were alive, encased in double quotes. Ex: "American, 1930-1992"
Nationality	Country of artist's nationality, string
Region	Geographical region, in which the artist's country of nationality exists, string
Continent	Continent, where the artist's country of nationality exists, string
Gender	Artist's gender, string

Introduction

What is gender? The question is at the core of this capstone project, *Evocative visualization of void and fluidity*. It not only questions its definitions but also how gender shapes our perceptions of ourselves and others, and how it is in turn constructed as a means of control to enforce social hierarchies. The project ponders these questions about gender while attempting to creatively visualize gender inequality to surpass solely illustrating the disparity and to push the boundary of data visualizations. It also brings attention to the regional contrast in terms of differently evolving discussions around gender, and underscores the importance of continuing to collect and monitor the conventional gender data.

The project is a narrative essay with data visualizations on my GitHub (https://tomikokarino.github.io/Capstone/). Its intention is not to pursue solutions nor answers to the questions. Instead, it hopes to instigate much needed conversations about gender, and how data visualizations should be employed to further gender equality.

Background and inspirations

This capstone project is a culmination of my course work in the Data Analysis and Visualization program at CUNY Graduate Center (GC). The technical knowledge I gained to visualize data, as well as the class readings and discussions during the five semesters I had attended the GC provided the foundation for this project. For the

technical implementation, the practical training in d3.js in the Interactive Data Visualization Studio course was crucial, while theoretical works that I had read in the Data, Culture and Society course and the Working with Data: Fundamentals course guided me, a heterosexual cis female that never gave much thought to gender identity, to critically think about the complexity of gender.

The course work that became the direct impetus of this project is the final project for the Advanced Interactive Data Visualization Studio course that I took in the Spring 2022 semester, in which I learned advanced visualization techniques with d3.js. While searching for the final project topic, my interest in art history as well as the criticisms about the lack of diversity at art institutions in New York City led me to the permanent collection data of the Museum of Modern Art (MoMA). Thus came a project that visualizes the lack of diversity in gender and nationality among the artists in MoMA's permanent collection.[1]

While the project was a success and achieved the original objective of illustrating the disparities, it felt lacking. The charts produced for the project clearly visualized the overwhelming disparities, but that was all they did. Moreover, the inequality in nationality and gender of the artists was so staggering that merely visualizing it felt unchallenging and uninspiring. The dissatisfaction towards the project led me to start thinking about how to visualize a void or a lack of presence more creatively.

Simultaneously, I began thinking about the quality of gender data. In MoMA's permanent collection data 'Male', 'Female' and 'Non-Binary' are the values of the gender variable, and there are only three artists categorized as 'Non-Binary' in the

[1] See the project on my GitHub: https://tomikokarino.github.io/Data78000/

entire collection. Since the concept of non-binary gender was not introduced until recent years, it is understandable that the majority of the artists are categorized either as 'Male' or 'Female.' However, this led to the question of accuracy. While it is important to identify and illustrate the gender disparity, what does it achieve if the gender data is inaccurately collected? As our understanding of gender and its definition are rapidly evolving, this issue is not unique to MoMA's permanent collection data. It is a universal challenge for those who work with data that are intended to represent individual identity but result in reducing their subjectivity.

These questions and the dissatisfaction I had about the final project for the Advanced Interactive Data Visualization Studio course propelled me to continue working with the gender data in MoMA's permanent collection dataset for the capstone.

Research of existing works

After the Spring 2022 semester, I began researching existing projects that investigate gender bias in the art world. The first project that came to mind was *A Sort of Joy (Thousands of Exhausted Things)*, performed at MoMA in spring of 2015 (Romeo, "Thousands"). I originally learned of this project in the Data, Society, and Culture class that significantly expanded my understanding of data visualization.

A Sort of Joy was created by The Office of Creative Research (the OCR)[2] that participated in the 2013 - 2014 Artists Experiment initiative at MoMA.[3] The OCR used

[2] At this point, the OCR's website does not seem to exist. Their Twitter account is still publicly available, though the last tweet is from June 2017: https://twitter.com/The_O_C_R
[3] Artists Experiment initiative at MoMA: https://www.moma.org/calendar/programs/57#current

the museum's collection data, which only contains text data and no image, and pursued how a database can be performed (Thorp, "A Sort of Joy"). They developed a script, some of which were pre-written and the other were algorithmically generated on the fly during the performance. The script was then read out loud by Elevator Repair Service, a performance group based in New York City.[4]

According to the online article about this performance (Romeo, "Thousands"), for approximately 40 minutes, a group of six performers in a gallery at the museum spoke and sang seemingly random words and phrases. Every word they read was taken verbatim from the MoMA's database, including the title, artist's name, year, medium, and dimension. In one section of the show, four male performers, standing in a small circle with their backs facing each other in the middle of the gallery, read names of the artists. Each performer simultaneously read out a name that is displayed on the tablets in their hands. When this portion of the performance begins, they initially read out the full names of the artists: "John Timothy," "John Thompson," "John Sturgeon." Soon, three of the performers stop calling out the names as there are too many Johns and the names appearing on the tablets pass by too quickly for them to read. Instead of reading out the names, three performers flip the tablets to face the audience so that the audience can see the names whizzing by on the devices while one male performer continues to read just the first names of the artists. After they finish with John, they move onto Robert, then David, then Paul and so on. As one performer continues to read the first names of the artists, the pause between each name gets shorter and shorter. This continues for a few minutes while the audience quietly watches the

[4] See their web site for more information about the group and their work: https://www.elevator.org/

performance, and then suddenly two female voices yell “Mary.” I believe that this is where the audience realized the intention and meaning of this performance. This production directly influenced the video in this capstone project and gave me a completely new perspective on data visualization.

The other similar project that I found was the research conducted as part of the 2019 event at the National Gallery of Art (NGA) that revealed that the overwhelming majority of artworks displayed in the museum’s galleries since the 1980s is produced by white men artists (Greenwald, “What Can Data”). Many museums in the U.S., including the NGA, use The Museum System (TMS), the software to catalog and track their collections. The data collected in TMS contains the content, history, and changing display locations of the museum’s collection. Using the static data exported from TMS, the NGA hosted *Coding our Collection: The National Gallery of Art Datathon* to analyze the museum’s permanent collection in collaboration with researchers.[5] Five out of six analyses presented at this event produced findings that the museum holds artworks by white men artists, whose styles are predominantly Western, and less than half of their donations come from women donors.

Though the visualizations created for the NGA Datathon were insightful, I realized that, for my capstone, I wanted to move more towards what the OCR has done in terms of creativity and pushing the boundary of data visualization.

I have also read scholarships on gender, data and classification. Rereading *The Second Sex* by Simone de Beauvoir and *Invisible Women* by Caroline Criado Perez reminded me how much this world defaults and caters to men, their beliefs and needs.

[5] Read the press release about the event on the NGA website: https://www.nga.gov/press/2019/datathon.html

This automatic adherence to the male standards is not unintentional but treacherous. Lisa Gitelman, the editor of *"Raw Data" Is an Oxymoron*, unfolds in her book the history of data and its dependence on culture that is just as insidious. This subtle conditioning of our perceptions of the world was also a common theme in *Sorting Things Out: Classification and Its Consequences* by Geoffrey C. Bowker and Susan Leigh Star. Along with Michel Foucault's *The History of Sexuality*, the book helped me understand how the systematic categorizations of entities influence our behaviors towards standardizing the physical world. Together with the other scholarships that I will discuss in detail later in the white paper, these books guided me to understand that moving away from the male-centric way of seeing our world is a herculean task that requires time and concerted efforts by every sector in our society.

Process and theoretical frameworks

Data and data cleaning

Technical implementation of the visualizations was possibly the easiest part of this project, because I repurposed the data and two of the charts that I had created for the final project for the Advanced Data Visualization Studio course.

I originally downloaded Artists.csv file from MoMA's GitHub in spring 2022.[6] To clean the data, I used Microsoft Excel and BBEdit, a text editing software that offers a robust search and replace function using Regular Expressions.

[6] Access MoMA's GitHub: https://github.com/MuseumofModernArt/collection

MoMA's artists dataset contains only the artists' information without the data of artworks in the collection. It consists of the following variables:

ConstituentID	A unique number assigned to each artist, integer
DisplayName	Artist's full name, string
ArtistBio	Place of artist's birth and the years they were alive, encased in double quotes. Ex: "American, 1930-1992"
Nationality	Country of artist's nationality, string
Gender	Artist's gender, string
BeginDate	Artist's birth year, integer
EndDate	Artist's death year, integer
Wiki QID	A unique identifier of a data item on Wikidata, Q followed by an integer. Ex: Q1063584
ULAN	Getty Union List of Artist Names Online ID, integer

I started by deleting the 'BeginDate', 'EndDate', 'Wiki QID', and 'ULAN' variables as I determined them irrelevant to the project. I proceeded to clean out the "artists" that are not individual human beings as the museum also includes in their collection artworks attributed to non-human entities such as the first generation iPod made by Apple Inc. or a mass-produced poster printed by a design company. I then separated the 'DisplayName' into first and last names. Finally, I used the United Nations' Definition of Regions and newly created the 'Region' and 'Continent' variables based on the 'Nationality' variable.[7]

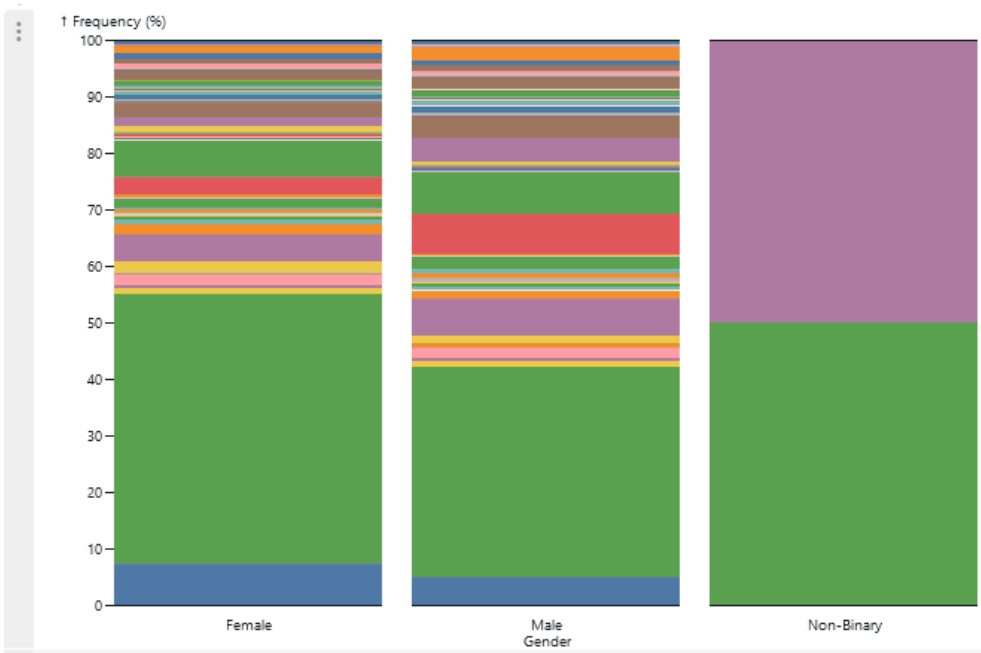
[7] You can read about the definition of regions on the UN website: https://population.un.org/wpp/DefinitionOfRegions

Technical choice and rationale

Visualization of the data was done in d3. This was a logical choice as the charts repurposed from the Advanced Data Visualization Studio course were coded in d3. For the scrolling effect, I used Scrollama and enter-view, both developed by Russell Samora.[8] The video was created using Camtasia, a video editing software by TechSmith.[9]

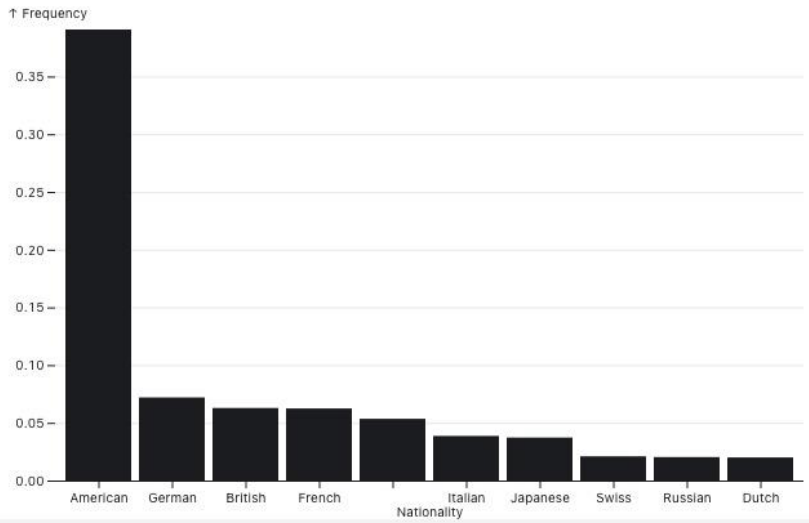
When I initially started working on producing the charts for the Spring 2022 course, I used Observable to create a proof of concept of my ideas.[10] Observable is a great platform for quickly creating a visualization to test if your idea is technically feasible or to find the best type of chart for the data you have. Over the course of two weeks in spring 2022, I created multiple visualizations using the MoMA data to see what would be viable for the final project (see below).

[8] Thank you Russell. His GitHub: https://github.com/russellsamora
[9] Camtasia website: https://www.techsmith.com/video-editor.html
[10] The Observable note I have created for the Spring 2022 course: https://observablehq.com/@45438b76c7484205/tomiko78000-1



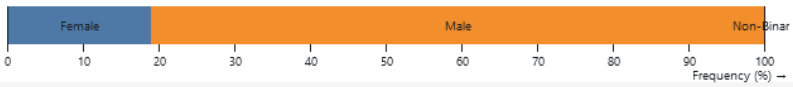

```
Plot.plot({
  y: {
    grid: true,
    percent: true
  },
  facet: {
    data: MoMAartistsGender,
    x: "Gender"
  },
  marks: [
    Plot.barY(MoMAartistsGender, Plot.groupZ({y: "proportion-facet"}, {fill: "Nationality"})),
    Plot.ruleY([0, 1])
  ]
})
```

Figure1: Screenshot of Observable Notebook, Stacked barplot of nationality distribution by gender.




```
Plot.plot({
  y: {
    grid: true
  },
  marks: [
    Plot.barY(MoMAartistsGender, Plot.groupX({y: "proportion"}, {x: "Nationality", sort: {x: "y", reverse:
true, limit: 10}})),
    Plot.ruleY([0])
  ]
})
```

Figure 2: Screenshot of Observable Notebook, Barplot of top 10 nationalities of artists in MoMA's collection data.




```
Plot.plot({
  x: {
    percent: true
  },
  marks: [
    Plot.barX(MoMAartistsGender, Plot.stackX(Plot.groupZ({x: "proportion"}, {fill: "Gender"}))),
    Plot.text(MoMAartistsGender, Plot.stackX(Plot.groupZ({x: "proportion", text: "first"}, {z: "Gender", text:
"Gender"}))),
    Plot.ruleX([0, 1])
  ]
})
```

Figure 3: Screenshot of Observable Notebook, Horizontal bar chart of gender distribution of artists in MoMA's collection data.

From the simple horizontal bar plot of the gender disparity I had created in Observable (Figure 3), I decided to create a treemap as it seemed visually more interesting. The gender treemap chart in the capstone project is essentially unchanged from the spring 2022 project. I have only adjusted the color scale.

For the first name bubble chart, I tweaked the nationality bubble chart from the spring 2022 project. When I was originally experimenting with various charts to display the rankings of nationalities of artists in MoMA's collection, I had created a few additional Observable notebooks to test different types of charts and consulted some online resources that explain and categorize various chart types. To create visualizations that illustrate rankings, I contemplated between the bubble chart and the circular barplot. While the latter can be visually interesting, I believed that it was not suited to display 118 different nationalities. Additionally, I preferred the bubble chart as the third visualization created for the Spring 2022 project was a circle pack of the artists' first names and nationalities, hence offering a visual unity.

Modifying the visualization from a nationality based bubble chart to a first name based one was straightforward. I simply used the 'FirstName' and 'Gender' variables to group the data by using d3.flatRollup, which calls a function on all the groups. Adding a transition and adjusting its timing required more time and care.

For the video, I used the free version of Voicemaker, an online text to voice conversion tool that offers various AI-generated voices with different accents, and produced mp3 files that read out the first names. I imported the first name audio files into Camtasia and sliced them into individual names so that I could time their start and end points. Determining the duration of each name was done by dividing the total

number of first names by the total number of artists in the bubble chart. For example, there are 236 Johns among the 2,841 artists in the bubble chart. 236 divided by 2841 is approximately 0.08. Therefore the name John takes up 10 seconds in the almost 2-minute video. For the background music, I downloaded *Moonshine*, created by Pridiga from Upbeat, which offers license-free music to be used in videos.

Design decisions

The font and color scheme were the primary design elements in this project. I knew that I should make the color scheme consistent in all three visualizations for visual cohesion. I also knew that I would not use a conventional color scheme for male and female, such as blue and pink respectively. The idea of using green for male and purple for female came from an example visualization included in *Data Feminism* by Catherine D'Ignazio and Lauren F. Klein. In their book, the authors discuss the interactive visualization, published in *the Telegraph* in 2018, that depicts the gender gap in the United Kingdom in terms of education, politics, business, and culture (D'Ignazio and Klein 112). The journalists at *the Telegraph* took inspiration from the Votes for Women campaign of early twentieth-century England, in which purple was used to represent freedom and dignity and green to represent hope (D'Ignazio and Klein 112). Because purple registers with greater contrast than green when placed against white, they decided to use purple for female and green for male (D'Ignazio and Klein 112). Where men significantly outnumber women in many visualizations, it is a simple method of bringing women back into focus (D'Ignazio and Klein 112).

While I am certain about the choice of color scheme, I am still unsure of the fonts. I selected from Google Fonts Andada Pro for the page title and section headings and Ysabeau for the rest of the project. As fonts and how they are received are extremely subjective, I realize that there is no one right font for the project. I selected Andada for the title and headings as the slightly serif typeface struck the right balance of legibility and non-utilitarian quality. Sans-serif Ysabeau, on the other hand, emits a clean, modern feel without being too impersonal. I also decided to use a thinner than the average weight of the font because I believed that it emanates sensitivity and creates nuanced ambiance.

Narrative development

Parallel to producing the visualizations, I worked on drafting the narrative, which proved to be much more difficult. Writing of any kind for me requires great effort. Succinctly putting into words the nebulous thoughts about gender that have accumulated in my mind over the course of three years was almost painful. To be completely frank, I am not entirely confident of my essay as I write this white paper. For one, I struggled to distill the scholarships on gender that I had read over the years into a narrative that drives my points in the capstone.

When thinking about more conscientious data collection and visualization, *Data Feminism* offered theoretical foundations and excellent examples. Divided into seven chapters, each of which focuses on a different principle, the book examines both the use and the limitation of data to question and challenge the existing systems of power, ultimately aiming to change the power distribution. In chapter 4, where the authors focus

on gender binary and hierarchy, they discuss the importance of being counted “for the basis of policymaking or resource allocation, as well as risks associated with being visible as a person of non-conventional gender” (D’Ignazio and Klein 97). D’Ignazio and Klein argue that gender is not just a physical trait but also a social construct that is instituted by systemic and systematic categorization and is understood by repeated performances that reinforce the categories of gender (108-109). The authors maintain that while all data need to be put into a category to be operable, once a system is in place, it becomes naturalized as “the way things are” and is rarely questioned how our classification systems are constructed, what values or judgments might be encoded into them, or why they were thought up in the first place (D’Ignazio and Klein 104).

To move away from the data practice that reinforces the binary gender categorization, D’Ignazio and Klein introduce examples of more conscientious data collection practices that are designed to accommodate non-binary genders. Through these examples they also point out the risks of disclosing one’s gender, especially when it is non-binary, and further discuss and question the need for collecting gender data in non-health research (110-111). Their arguments helped me to rethink the rationale for and methods of data collection, and provided a sobering view of how blindingly our society collects data without clear purpose or knowledge of its use.

Data Feminism also includes prime inspirations for creative and evocative visualizations.

Figure 4: Screenshot of the interactive visualization in the article, *How diverse is the 2018 US Congress?*, published on the Guardian's website.

In addition to the U.S. senators' gender identity visualization (Figure 4) that is mentioned in this capstone project, the authors introduce a diagram that visualizes the complexity of sex and gender, originally published in the September 2017 issue of *Scientific American* (Figure 5).

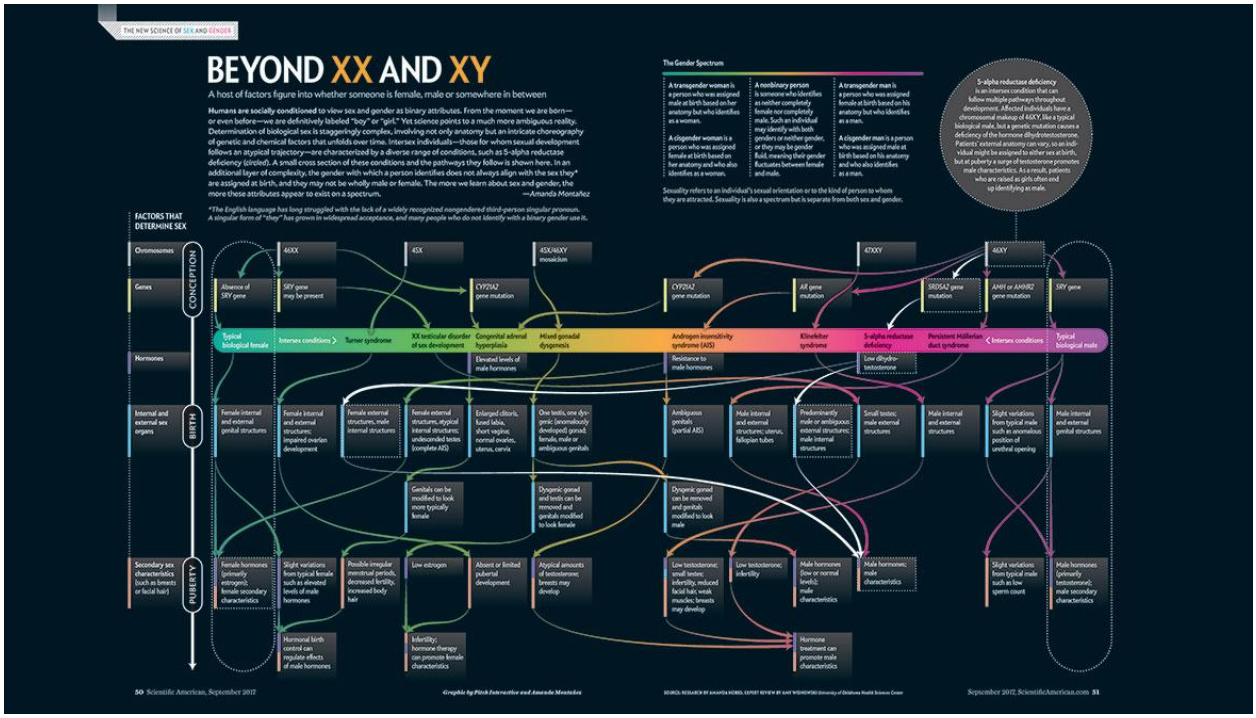


Figure 5: *Beyond XX and XY* diagram, published in the September 2017 issue of *Scientific American*.

To represent sex and gender in a scientifically accurate manner, the designer of the diagram upended the usual simplification that occurs in data visualization, and pushed simple, oppressive ideas to be more complex, nuanced, and just (D'Ignazio and Kelin 115). The idea that a data visualization can be intentionally complex and demand a greater thought process to be understood gave me much inspiration.

Race and/as Technology; or, How to Do Things to Race by Wendy Hui Kyong Chun, written as the introduction to the May 2009 issue of *Camera Obscura*, the journal that provides a forum for scholarship and debate on feminism, culture, and media studies, also enormously influenced my understanding of gender.[11]

[11] You can find out more about the online journal on their website: https://read.dukeupress.edu/camera-obscura

In the article, Chun introduces “race and/as technology,” a strange formulation that forces the reader to reread for accuracy and full comprehension. The author argues that:

> Race as technology reveals how race functions as the “as,” how it facilitates comparisons between entities classed as similar or dissimilar. This comparison of race and technology also displaces claims of race as either purely biological or purely cultural because technological mediation, which has been used to define humankind as such (“man” as a “tool-using” animal), is always already a mix of science, art, and culture. [...] Race, it therefore follows, has never been simply biological or cultural; rather, it has been crucial to negotiating and establishing historically variable definitions of biology and culture. (Chun 8)

Chun contends that race is a fundamental organizing principle of social relationships (14). She further maintains that race, like media, is also a heuristic, a way to understand, to reveal, the world around us (Chun 14).

Gender as a tool of control then is analogous to race. Both gender and race are understood as biological and physical traits, yet they are not merely indexical; they can be extremely forcible apparatuses in organizing social relationships (Chun 14). In this sense, gender too can be considered as technology as Chun argues race to be. Gender, just like race, is not simply an object of representation and portrayal but rather a technique, “a carefully crafted, historically inflected system of tools, mediation, or enframing that build history and identity” (Chun 7).

Chun's arguments dovetail with Michel Foucault's writings in *The History of Sexuality*. Reading the chronological dissection of classification of sexuality as a way to monitor and govern "abnormal" sexualities was such a revelation. It also helped me see the risks of being classified, as being labeled and made visible causes the Other sexualities to be compared against the default sexuality (hetrosexuality).

Using binary gender as a means of control is still very much the dominant practice in most regions in the world. As we push for more inclusive gender discussions in the Global North, we must keep in mind that such conversations may never be allowed in many parts of the globe. Gender inequalities caused by the stubborn systems that believe in and enforce the binary gender are abound. Girls in most parts of the world are twice less likely than boys of the same age to have access to the internet or digital skills (UNICEF, "Bridging" 18-23). The HIV epidemic is influenced by gender inequalities as girls are less educated about protection against sexually transmitted diseases (UNICEF, "HIV/AIDS"). As climate change exacerbates and access to clean water becomes scarcer, more girls will likely miss out on school because water fetching duty disproportionately falls on women and girls (UNICEF, "Climate change"). Gender inequality is a cross-cutting injustice that disadvantages half of the world's population. This is why it is imperative to continue to collect conventional gender data so that we can have disaggregated data analysis to identify where gender biases occur and how to mitigate them.

Reflections

After completing the capstone project, I am still unable to find answers to the question “What is gender?” The more I think about gender, the more I have trouble answering it. One thing that is clear is that we are at a critical point of rethinking data collection and usage in our society as our personal identity information and data of our behavioral patterns become valuable commodities.

Since the days of Web 1.0, we slowly became conditioned to give our personal information in order to access content on the web. Combined with this conditioned behavior, the classification systems that enforce social hierarchies subtly acclimate us to perpetuate the power structure that oppresses people of non-binary gender. Instead of continuing with the existing data practices that operate on the flawed assumptions of binary gender, it is time for more inclusive data collection processes and the development of new technologies.

Through my research on the aforementioned theoretical frameworks, when I think about this capstone project’s contribution to the data visualization and digital humanities communities, I hope that it brings about serious discussions on the current data collection and analysis methods and calls for these changes that enable truly equitable and just data practices.

BIBLIOGRAPHY

Bowker, Geoffrey C. and Susan Leigh Star. Sorting Things Out: Classification and Its Consequences. The MIT Press, 2000.

Bridging the Gender Digital Divide: Challenges and an Urgent Call for Action for Equitable Digital Skills Development. United Nations Children's Fund (UNICEF), 2023.

Criado Perez, Caroline. Invisible Women: Data Bias in a World Designed for 33 Men. Abrams Press, 2019.

De Beauvoire, Simone. The Second Sex. Everyman's Library, 1993.

D'Ignazion, Catherine and Lauren F. Klein. Data Feminism. The MIT Press, 2020.

Foucalut, Michel. *The History of Sexuality*. Vintage Books, 1990.

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

Greenwald, Diana. "What Can Data Teach Us About Museum Collections?" *aam-us.org*, 27 April 2020, www.aam-us.org/2020/04/27/what-can-data-teach-us-about-museum-collections/

Ọnụọha, Mimi. "The Library of Missing Datasets (2016)." mimionuoha.com, mimionuoha.com/the-library-of-missing-datasets

Romeo, Fiona. "Thousands of Exhausted Things, or why we dedicated MoMA's collection data to the public domain." *Medium*, 27 July 2015, medium.com/digital-moma/thousands-of-exhausted-things-or-why-we-dedicated-moma-s-collection-data-to-the-public-domain-7e0a7165e99

Thorp, Jer. "A Sort of Joy." *Medium*, 27 July 2015, medium.com/memo-random/a-sort-of-joy-1d9d5ff02ac9

UNICEF. "HIV/AIDS." data.unicef.org, July 2023, data.unicef.org/topic/gender/gender-and-hiv-aids/

UNICEF. "Climate change." data.unicef.org, May 2023, data.unicef.org/topic/climate-change/overview/

Wendy Hui Kyong Chun; Introduction: Race and/as Technology; or, How to Do Things to Race. Camera Obscura 1 May 2009; 24 (1 (70)): 7–35. doi: https://doi.org/10.1215/02705346-2008-013