LUISS

Corso di laurea in Data Science and Management

Cattedra Data Visualization

Data Feminism and the Investigation of Gender Inequalities in the Italian Parliament

Prof.ssa Blerina Sinaimeri

RELATORE

Prof. Alessio Martino

CORRELATORE

Gaia Alberti (Matr. 762671)

CANDIDATO

Anno Accademico 2023/2024

TABLE OF CONTENTS

ABSTRACT

This study focuses on assessing the impact of applying feminist data principles to data analysis and visualization, particularly as a tool to avoid the biases and prejudices that are deeply embedded within traditional data practices, from data collection to data visualization. Based on the principles outlined in the book "Data Feminism" written by D'Ignazio and Klein, this research aims to assess if and how feminist data analysis can be a tool to uncover disparities, illuminate hidden narratives, and how feminist data visualization can provide a better understanding of the data presented. This research uses data about the dimensions of gender representation within the Italian Parliament in order to assess the impact of applying feminist data principles in both data analysis and visualization.

The research is divided into two sections, one dependent on the other. The initial part of the study applies feminist data principles to the analysis of parliamentary data, investigating female representation and participation in parliamentary activities with respect to male deputies. This investigation not only counts their presence but also focuses on analyzing their contributions in parliamentary activities to uncover women's active engagement with respect to their male counterparts. The analysis reveals a gradual increase in female representation over the years, though with recent declines, and a more active role of female deputies in certain parliamentary activities, especially when addressing social issues regarding gender equality, LGBTQ+ rights and disability. This section also highlights the limitations of data analysis due to how the data was collected, such as the use of the male/female binary classification to classify the gender of the deputies instead of taking into consideration the whole spectrum of genders.

The second part of the research uses the insights found in the preceding analysis to design specific data visualizations following both the more traditional approach and the feminist approach. Through a survey, the two visualization designs are compared to assess user preference for a particular visualization against specific attributes such as clarity, memorability, neutrality and emotional engagement. By contrasting traditional visualization, which often prioritize neutrality and objectivity, with feminist visualizations that aim to engage viewers on a more emotional and empathetic level, the study aims to assess the relative effectiveness of each approach in facilitating a deeper understanding of the data presented. In the survey, feminist data visualizations received higher scores in most of the attributes except for data neutrality. These results confirm the belief that feminist data visualization promote better data understanding and memorability since, by leveraging emotions and focusing on storytelling, the viewers feel more connected with the data.

This study enriches the framework of Data Feminism by showing that the application of feminist data principles in data analysis and data visualization can promote a more in depth understanding of the topic being analyzed. The findings of this research support the idea that data feminism promotes key analysis that helps to uncover specific insights, like the experiences of marginalized groups, that otherwise would remain hidden with more traditional approaches. For this reason, by adopting feminist principles, the research is able to shift the narrative from viewing women only as statistical figures to recognizing their fundamental and active roles in parliamentary activities. These concepts not only demonstrate the potential of Data Feminism to uncover hidden insights but also to advocate for societal change. These concepts challenge to rethink the relationship with data which is not only used to inform but also to advocate for a future where data serves as a tool for positive change, advancing social causes like gender equality.

1 INTRODUCTION

For many years, the movements striving for the achievement of gender equality have always been associated only with social and political activism. These activities had the goal to shed light on existing inequalities and address the social injustices and biases related to gender within society. The focus has mainly been on providing equal rights and creating equal opportunities across different areas of social and political life, including education, employment and political representation.

However, in recent years, the greater awareness that gender inequality was being reinforced in many other fields of society made these movements expand toward other disciplines, such as the discipline of data science. In fact, there was a recognition that gender biases were deeply embedded in the processes and technologies that shape our society, symbolizing that gender inequality was not just a social and political issue (D'Ignazio & Klein, 2020). For this reason, movements toward gender equality started focusing their attention on the field of data science, described as a pivotal area in which gender biases and inequalities are perpetuated and reinforced.

Data science involves a series of data practices, such as data collection, data analysis and the creation of algorithms through which insights and information are extracted and knowledge is produced. What data scientists do is to make discoveries while swimming in data (Harvard Business Review, 2012). Data science helps society by harnessing the power of data to make informed decisions, bring innovation, and address complex challenges (Vicario et al., 2019). However, the outcomes of these data practices can be deeply influenced by the biases of their creators, predominantly white heterosexual males. This can perpetuate social inequalities and obscure the perspectives of marginalized groups, excluding them from data narrative. Therefore, this causes an urgent need to incorporate gender equality at every stage of the data processes.

At this critical point, the concept of Data Feminism appears as a powerful framework for examining and reshaping the field of data science. Based on the work of Catherine D'Ignazio and Lauren F. Klein, Data Feminism and its guiding 7 principles call for the integration of feminist theories into data science practices to directly address and mitigate biases and disparities.

1.1 Problem Statement

Even though there has been significant effort in achieving gender equality into different areas, data science still remains a field in which specific consolidated methodologies continue to cause the perpetuation of biases and existing inequalities. For example, the methodologies used in the data collection process can affect the performance of machine learning models, with the risk of perpetuating specific social dynamics (Noble, 2018). This highlights the importance of taking into consideration that data is far from being neutral and the biases of AI algorithms when used in decision-making processes can significantly affect the outcome, distorting the output of processes like job recruitment and loan approvals (O'Neil, 2016).

The way in which data is collected, analyzed and visualized can create biases and therefore strongly affect the outcome of the research. If these processes are not critically analyzed, then data science can be used as a form of power, since its practices lead to the generation of knowledge and decisions that affect society (D'Ignazio & Klein, 2020). This point of view acknowledges that the mechanisms designed to collect, analyze and interpret data are inevitably pervaded by the positionalities, biases and values of the people who created them. It is important to highlight how the tech industry, including data science, has traditionally been and continues to be male dominated. This introduces an additional layer of gender bias into these processes since they inherently influence the outcomes of their data practices with their biases. This imbalance in the field can lead to the perpetuation of existing gender stereotypes since the perspectives and experiences of women and non-binary individuals are underrepresented in the creation and application of data systems. Moreover, according to Data Feminism, this should not only be limited to the gender dimension but should include all the other dimensionalities like ethnicity, disability and so on. Indeed, according to the Bureau of Statistics, in 2021, Black and Hispanic people made up respectively 9% and 15% of the U.S. STEM workforce population against 64% of white people.

Moreover, there is a growing emphasis not only on who is involved in the fields of Data Science but also on the disparities in data that may lead to the underrepresentation or misrepresentation of certain genders. This focus extends to how datasets are created, therefore which are the sources, the guiding question of data collection and the frameworks used in analysis. It highlights the need for a critical examination of data practices to ensure they do not make invisible or distort the experiences and realities of women, non-binary, and other marginalized groups (Eubanks, 2018).

Not only data collection but also data analysis and data visualization must be carefully examined. Traditional data analysis often relies on methodologies that do not take into account the possibility that the datasets can have biases and lack gendered perspectives. In fact, they often do not consider the social factors that might influence the data collection, leading to distorted conclusions. These methods typically rely on quantitative metrics that can be easily aggregated and compared which, however, may obscure specific narratives that show the presence of disparities (Criado, 2019). For example, aggregate data might show overall trends but fail to provide an intersectional viewpoint. In fact, by intersecting the gender dimension with other dimensions particularly those related to ethnic backgrounds, disabilities or socioeconomic statuses, analysis can reveal more complex patterns of discrimination. The non-adoption of an intersectional approach can result in analyses that overlook critical disparities, leading to conclusions that accidentally reinforce these inequalities.

On the visualization side, traditional techniques often rely on the principle of neutrality and visual minimalism. On the one hand, this approach is beneficial for quick and easy-to-understand visualization, on the other hand, these principles tend to oversimplify complex realities (D'Ignazio & Klein, 2020). Visualizations that do not reflect the complexity of data concerning social inequalities can result in misleading representations of the reality, which may not effectively communicate the sense of urgency to solve these disparities. By relying on the principle of visual minimalism, many traditional visualizations are often characterized by simple bar charts that show the numerical differences between genders but at the same time they fail to communicate the social contexts that contribute to these numbers. Additionally, a narrative element that would enable viewers to engage with the data on a deeper level is often absent from traditional visualizations. In fact, traditional data visualization often prioritizes the principle of neutrality at the cost of greater emotional engagement (Kennedy & Hill, 2016). The absence of storytelling lowers the users' empathy and engagement, elements that are critical in motivating societal change (Segel & Heer, 2010).

These shortcomings of traditional data practices also characterize data about political representation in the Italian Parliament. Traditional analysis may fail to capture the real women's contributions, contributing to the creation of a non-truthful belief in their political representation and participation. The risk of a misrepresentation highlights the importance of adopting data practices that actively challenge the current structures of power by uncovering and highlighting hidden narratives and disparities (D'Ignazio & Klein, 2020).

1.2 Framework & Research Questions

This research is based on the theoretical framework proposed by Catherine D'Ignazio and Lauren F. Klein in their work “Data Feminism”. Data Feminism focuses on proposing an approach that first examines the causes that perpetuate existing biases in the field of data science and then challenges these power dynamics to ensure that data practices promote fairness, inclusivity and equality. In their book, they propose seven guiding principles that should be adopted in order to challenge conventional data collection, analysis and visualization, ensuring that everyone’s perspective is included.

This study will adopt these 7 principles in data analysis and visualization, using as a case study data about gender representation in the Italian Parliament. The goal is to contribute to the Data Feminism discourse by assessing the impact of the adoption of feminist data principles to uncover and understand better gender disparities through data analysis and visualization,

In particular, the study aims to answer the following two research questions:

1. **How can feminist data analysis reveal hidden narratives within traditional datasets?** Therefore, what are the implications of applying feminist data principles to the analysis of data about gender inequality within the Italian Parliament? This question aims to examine the potential of feminist data principles to uncover hidden patterns of gender inequality that are often obscured by traditional data practices.
2. **In what ways does feminist data visualization offer a better understanding of gender dynamics compared to traditional visualization methods?** This research question aims to assess the level of memorability and engagement of feminist data visualizations and therefore, to investigate whether the viewers are able to better understand the data presented with respect to more traditional data visualization.

In order to answer these two research questions that are strictly connected to one another, this study is based on a two-step approach. This approach is used to assess the potential of applying feminist data principles by combining quantitative parliamentary data with quantitative and qualitative survey data, which will enable to provide a comprehensive understanding of the subject matter.

The first research question will be answered by running a quantitative analysis of gender representation within the Italian Parliament using data collected from the website of the Italian Chamber of Deputies through the SPARQL endpoint. The focus will not only be on numerical gender representation but will center on the deputies' actual involvement and contributions in parliamentary activities. This first part will be instrumental in identifying patterns of gender inequality and assessing progress toward gender parity in political representation, as well as highlighting women's contributions in politics.

After the quantitative data analysis, the study employs A/B testing to compare the effectiveness of traditional versus feminist data visualizations. Participants of the survey are asked to assess which approach better facilitates understanding and engagement with the data.

1.3 Contributions of the Study

This research contributes to the discourse of Data Feminism in three key areas:

1. Challenging traditional methodologies

The study wants to challenge traditional approaches to data practices in favor of feminist approaches by showing that traditional methodologies in data collection, analysis and visualization are often the cause of the perpetuation of existing inequalities. If on one hand, traditional data practices often provide partial perspectives by simplifying complex realities, on the contrary, data feminist principles highlights the complexities of the data, making data practices more inclusive and uncovering the experiences of marginalized groups. Indeed, this research aims to show how the adoption of data feminist principles can uncover hidden narratives and disparities and therefore, contributing to data science practices that are more sensitive to gender and other forms of inequality. The study examines the principles of data neutrality and visual minimalism applied to data visualization and challenges them by proposing feminist data visualizations as an alternative that puts emphasis on emotions and storytelling with the end goal of assessing whether they are more engaging and memorable than traditional data visualization.

2. Discovering hidden insights

This study not only contributes on a theoretical level to the framework of Data Feminism but it also provides a practical application of the principles of Data Feminism in a critical domain of public life: political representation. By analyzing the data about the Italian Parliament and through the adoption of specific feminist principles, the aim is to assess the challenges that still exist in achieving gender parity in political representation as well as to highlight the unseen contributions of women to parliamentary activities.

3. Achieving Social Change

The ultimate goal is to contribute to the development of data practices that are not only fairer and more inclusive but also capable of driving societal change. The insights gathered through data analysis and visualization can inform and promote change and, in this case, can be pivotal for contributing to the achievement of the Sustainable Development Goal (SDG) 5.5, which champions the cause of "ensuring women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life."

Thanks to the principles of Data Feminism and the two-step approach using quantitative and qualitative data, this thesis aims to assess the ability of Data Feminism to reveal gender disparities and enhance the comprehension and impact of data visualization, with the ultimate goal of fostering broader social goals.

2 LITERATURE

2.1 DATA FEMINISM

The concept of Data Feminism was introduced by Catherine D' Ignazio and Lauren F. Klein, as a powerful new approach that incorporates feminist theory into the field of Data science to rethink and challenge the methods through which data is collected, processed, interpreted and visualized. The core of this innovative and novel approach is the understanding that data is far from being impartial and objective since it is strongly influenced by power dynamics and social structures. This leads to the creation and perpetuation of societal prejudices and disparities. In fact, Data Feminism claims that Data Science is a form of power in which the creators, mostly males, influence the creation of knowledge. Catherine D' Ignazio and Lauren F. Klein introduced the notions of "Data Feminism" and its related seven guiding principles in their book "Data Feminism" released in 2020, which has been pivotal in defining and popularizing the concept.

In their book, Catherine D' Ignazio and Lauren F. Klein propose a series of steps and concepts that challenge the conventional techniques of data science drawing upon the long history of female activism. Their work highlights the importance of taking into consideration power dynamics when performing data tasks in order to include the voices and lived experiences of marginalized people in data narratives.

Data feminism emerged from the need to address and eliminate the biases and prejudices of the “dominant group”, composed of white cis males, that are frequently present in the processes of gathering, analyzing and visualizing data (D'Ignazio & Klein, 2020). This claim addresses the increasing recognition of the risks connected to the new technologies based on data processes, such as artificial intelligence algorithms which have the potential to perpetuate inequalities if they are not carefully conducted with an ethical approach. By examining data science with feminist principles, on one hand Data Feminism acknowledges that the output of data processes is always affected by the social context in which it is created, on the other hand it also aims to identify and address these biases by using an ethical approach to have more inclusive and fairer data practices.

The contributions of D'Ignazio, Klein, and other scholars are meant to challenge people to rethink the role of data in society with the associated risks and benefits.

2.1.1 Historical Context and Evolution

The origins of Data Feminism are connected to the movements that started criticizing traditional concepts of science and technology in the second half of the 20th century. During this time, many scholars began to challenge the principles of neutrality and objectivity that were characterizing scientific research. In fact, according to these scholars, scientific work was deeply affected by the perspectives of the creators that are never completely neutral and objective.

Two important scholars of these movements were Donna Haraway and Sandra Harding, who claimed that all the knowledge produced is influenced by the context in which the individual lives as well as by the relationships created. This implies that knowledge reflects the prejudices of its creators with regard to specific social themes such gender, race and class.

Haraway, in her work "A Cyborg Manifesto", combines elements of science fiction with critical theory to challenge the dichotomies of human/machine, natural/artificial, and, especially, male/female. In her work, she introduced the concept of the "cyborg" as a hybrid entity which is used as a metaphor to advocate for a more inclusive and flexible understanding of identity. Such a metaphor wants to go against the strict classifications that have traditionally marginalized women and other groups. The work of Haraway is meant to challenge the traditional classification of identity and rethink it in a way that is not bound by traditional boundaries. "Technology and science should be inclusive and acknowledge the interconnections between social, political, and biological factors." (Haraway, 1985)

Another important scholar that contributed to this discourse was Sandra Harding, a feminist philosopher who introduced the concept of "strong objectivity." According to Harding, the traditional notion of objectivity was actually "weak objectivity" since the outcome of scientific practices has been long influenced by power dynamics and societal hierarchies, leaving out from the production of knowledge the lived experiences of women and other marginalized groups. She proposed "strong objectivity" as an alternative, which requires the acknowledgment that the producer of knowledge can deeply influence research findings based on its own positionality, experiences and values. Research bias can never really be removed (Harding, 2005). Moreover, Harding (1986) and Hartsock (1983) claimed that taking into consideration the viewpoints of marginalized groups may offer stronger objectivity since they are more motivated to understand the views and perspectives of those in positions of power. These key statements have laid the groundwork for the advent of Data Feminism, highlighting the importance of encouraging researchers to consider how their own past experiences and biases may affect the outcome of their work.

As the digital age began and data became the new currency of knowledge and power, the questions raised by feminist theorists took on new dimensions. Feminist critiques were not just limited to the theoretical realm but also expanded in the concrete and practical aspects of life. Londa Schiebinger, professor of History of Science at Stanford University, and other scholars showed how gender biases in scientific research can lead not only to missed market opportunities but also to tangible harm. *"*Doing research wrong costs lives and money" (Schiebinger, 2008). With her work "gendered innovations", Schiebinger claims that failure to consider gender can lead to flawed research outcomes, such as medical treatments that do not work as effectively for women or safety car equipment that does not fit female bodies.

In fact, the creation of algorithms for decision-making further increased the urgency to establish a system of vigilance. If on the one hand, these advancements have the potential to improve the quality of humans' lives, on the other, they may pose significant risks connected to the perpetuation of specific discrimination through algorithm bias.

Examples of such algorithms demonstrate how biases either in data collection, design or implementation can lead to discriminatory outcomes. Amazon's recruitment algorithm is an example of algorithmic bias that led to discriminatory outcomes. In fact, this algorithm ended up favoring men over women when screening their resumes. This was due to the fact that the algorithm was trained mostly on a set of resumes of men since most of the people working in the tech industry were male, reinforcing social inequalities. The algorithm learned and replicated those biases, penalizing resumes containing the word "women". Other examples of algorithm bias are facial recognition technologies that exhibit lower accuracy rates for women and people of color. These biases are not just technical bugs, but they represent the outcome of deeper societal inequalities that have been transferred into the digital domain.

Data feminism was also shaped by the projects carried out in the field of digital humanities in which feminist principles were integrated in order to shed light on the contributions of marginalized groups, like the contribution of women. Digital humanity is an interdisciplinary field that combines computational methods and traditional humanities, exploring cultural artifacts, texts, and historical records through digital tools and platforms. An example of such an approach is the Orlando Project (2016), which digitally reported the contributions of women to literary history, shedding light on their work. There have been other projects like the Women Writers Project (2016) which have leveraged digital technologies to bring the work of female writers to a broader audience.

Data feminism was born thanks to the contributions of many scholars and artists that challenged the conventional principles of neutrality by adopting a feminist viewpoint. Data Feminism builds on the foundational concepts of feminist scholars like Haraway and Harding, as well as practical examples of gendered innovations and feminist digital humanities projects. This approach to data science prioritizes inclusivity, reflexivity and the removal of social disparities, with the end goal of producing more inclusive data practices.

2.2 THE 7 PRINCIPLES OF DATA FEMINISM

The formalization of Data Feminism as a distinct field owes much to the work of Catherine D'Ignazio and Lauren F. Klein. In their book "Data Feminism," published in 2020, they merged feminist theory with data science to propose an innovative feminist approach for thinking about and working with data. This framework is based on seven principles to make data practices more ethical and equitable. These principles go from considering power and challenging existing power structures to embracing multiple perspectives and making labor visible in data work.

2.2.1 Examine Power

Examining power is the first principle of Data Feminism and involves understanding how power operates in the world, investigating the power structures that characterize the data field: from data collection and analysis to data visualization. These power dynamics are deeply embedded in our society and take the form of “structural privilege” for certain groups and “structural oppression” for others (D'Ignazio & Klein, 2020). The people that suffer from systematic disadvantage are called “minoritized groups”, contrary to “minorities groups”, to symbolize the fact that they might not always represent a small percentage of the population. Women are an example of a minoritized group, even though they constitute most of the world population. Such disparities in power and privilege are not due to chance but are the outcome of systems created by dominant groups that do not take into consideration the lives and needs of marginalized groups.

Examining power in the field of data science starts by investigating the demographics of the workers involved in data processes: who is involved and who is not. Examining power means also stimulating people to think about these questions and always considering who is left out of the scene. People involved in this field should therefore question whose goals are being prioritized, and whose are not considered, who is going to benefit from the outcomes of data science, and who is at risk of being overlooked. These questions must be answered in order to understand the broader implications of data practices for society.

It is important to highlight that data science is the result of human labor and human decision making. The teams responsible for collecting and analyzing data often share similar characteristics and backgrounds, representing, therefore, a small portion of the global population. According to the Bureau of Statistics, in 2021, the percentage of female STEM workers was only 26%. The lack of diversity in terms of ethnic background may also cause certain perspectives to become prominent, leading to bigger consequences if scaled globally. The dominance of certain groups within data teams means that their own biases are more likely to be embedded in the datasets and algorithms that increasingly control aspects of our lives.

Kate Crawford, a social scientist, has commented on this problem, highlighting how AI systems could create potential dangers if not controlled by society. In fact, AI is neither artificial nor intelligent. It is made from natural resources, and it is people who are performing the tasks to make the systems appear autonomous (Crawford, 2021). According to Crawford, the main danger of the future of artificial intelligence does not involve a situation in which human intelligence will be surpassed by machines, instead, it will be one where our digital systems become pervaded with sexism, racism and other types of discrimination, This hard-coding of biases into AI systems will perpetuate and even exacerbate inequalities, making it critical to examine and address the power dynamics at play in data science.

It is important to highlight not only the risk of disposing of biased or unrepresentative datasets due to the dominance of certain groups but also the risk of never collecting certain types of data. Data Feminism works hard to shed light on the so called “Missing Datasets”. The “Library of Missing Datasets” is an artwork carried out by Mimi Onuoha in 2016, who decided to represent all the datasets about social themes that people were sure to exist but in reality, they had never been collected up to that point. This is strictly connected to the power dynamics that operate in our world. It is those who have power and those in a position of privilege who get to decide who gets counted and who is not counted in the statistics. Using a feminist approach means also collecting data about realities that have never been analyzed and for which data does not exist. The goal is to make these situations visible to the eyes of society since until data is not collected, it will never be addressed by policies.

The principle of examining power advocates for a structural change in how data processes are carried out by first examining how power dynamics operate in our society. It sheds light on the importance of diversifying the teams of people involved in data practices in order to include a broader range of perspectives, ensuring that the outcome of data practices is inclusive and fair. Moreover, examining power also means promoting transparency in every step of the data process, during which the people involved should acknowledge their positionality and the potential biases that they could apply to their work.

2.2.2 Challenge Power

The second principle of Data Feminism is challenging power which is a call to challenge and eliminate the unequal power structures that characterize our society with a focus on the inequalities and biases perpetuated in the data field. Therefore, after examining power, the next step is to challenge it. Data Feminism believes in challenging power by presenting counter data to quantify and visualize inequalities and oppression. Counter data represents data that is intentionally collected to shed light on the lives of those minoritized groups who are often excluded by the statistics.

Challenging power means involving end users in the data analysis and visualization process. In fact, the feminist framework is characterized by a participatory approach where all the stakeholders, especially marginalized groups and those directly involved in the themes under study are integrated into the whole design process. This approach makes sure that the analysis and visualization truly reflect the lived experiences of those affected, providing empowerment to the community by acknowledging their contribution and perspective. As suggested by the book Data Feminism, the community of interest can be involved through workshops, focus groups and feedback sessions, allowing individuals to contribute with their insights and points of view.

If on the one hand, Data feminism wants to uncover the experiences of marginalized people, on the other hand, the principle of challenging power wants to prevent the creation of deficit narratives that depict marginalized groups solely as victims or passive subjects in need of rescue. Data feminism advocates for presenting data in ways that highlight the actions and contributions of marginalized groups rather than just the systems of oppression. An example of work that tries to avoid deficit narratives is the one carried out by Sara Blecker who decided to focus on women's unseen contributions to the Uruguayan economy, rather than exclusively on the violences that Uruguayan women suffer from. This project demonstrates of how data can be used to tell empowering stories that challenge stereotypical narratives, highlighting the valuable roles women play in society.

According to this principle, eliminating the existing structural inequalities must be a shared goal between the dominant and minoritized groups representing a commitment toward the so called: co-liberation. This implies that a feminist approach to the design of data analysis and data visualization should provide benefits for both groups. The concept of co-liberation involves working with the community directly involved and learning from them making sure that the data projects are based on the needs and knowledge of these communities.

Data feminism challenges the existing structure of oppression through the key pillars previously described: inclusion of the community of interest, avoidance of deficit narratives and adopting the principle of co-liberation.

2.2.3 Elevate Emotions

The fourth principle of Data Feminism is elevating emotion and embodiment. Traditionally, data visualization scholars have always criticized the use of emotion in data visualization projects, symbolizing a lack of neutrality and therefore, reliability of the data presented. Introducing emotions would have meant introducing bias in the representation of the data. The underlining idea is that the more plain, the more neutral; the more neutral, the more objective; and the more objective, the more true (D'Ignazio & Klein, 2020). Technical visuals should be characterized by a plain style meant to create a neutral emotional field in which the viewers are not influenced by how to interpret the data represented (Amare, N. et al., 2016). Edward Tufte emphasized in his work "*The Visual Display of Quantitative Information*" how data visualization should rely on the principles of simplicity, integrity and objectivity to effectively communicate data without relying on unnecessary elements and decorations. According to the traditional view, emotional involvement might comprise the integrity of the study and therefore rationality should be the mode through which data is communicated. The concept of "distance" when performing statistical work was remarked by Karl Pearson who claimed that "Quantification is a technology of distance" in which there should be a detachment between individuals and the knowledge they produce, leaving behind their feelings and emotions. However, this belief was strongly criticized by Donna Haraway who introduced the concept of the "God Trick", claiming that "data visualization is the god trick of seeing everything from nowhere". This is because people believe that they are able to see everything with a glance from the visualization that they are analyzing. In reality, that visualization is displaying just a small part of the whole narrative, hiding the people, the methods, the experiences that lie behind a plain visualization. Therefore, this is a trick because what people believe to be everything is actually just a partial viewpoint and, most of the times, this partial perspective is the perspective of the creator of the visualization who is likely to be part of the dominant group. Building on Haraway's critiques, Data Feminism wants to question the actual neutrality of visual minimalism and shed light on the potentiality of leveraging emotion in data visualization to help people learn, remember and communicate data.

Most of the times, what people see and understand from a visualization is affected by the editorial choice of the creator. The creator, with his own intent and bias, decides what to show and what not to show. In order to show this concept, the New York Times used data about unemployment rates in September 2012 to create two types of data visualization, one using the perspective of the Democrats, Figure 1, and the other with the perspective of Republicans, Figure 2.

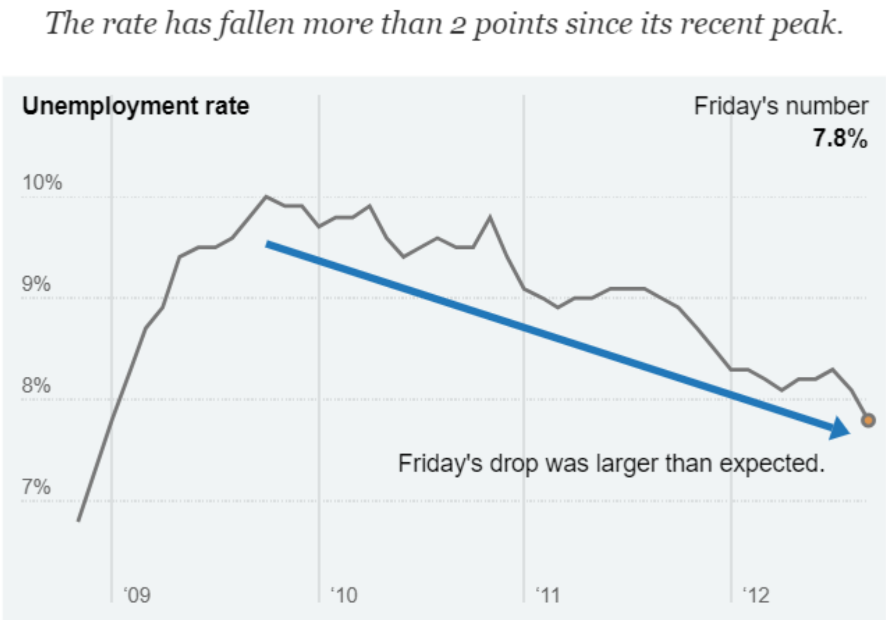

Figure 1 *Data visualization of the September 2012 jobs report from the perspective of Democrats. Images by Mike Bostock, Shan Carter, Amanda Cox, and Kevin Quealy, New York Times. (https://archive.nytimes.com/www.nytimes.com/interactive/2012/10/05/business/economy/one-report-diverging-perspectives.html)*

The rate was above 8 percent for 43 months.

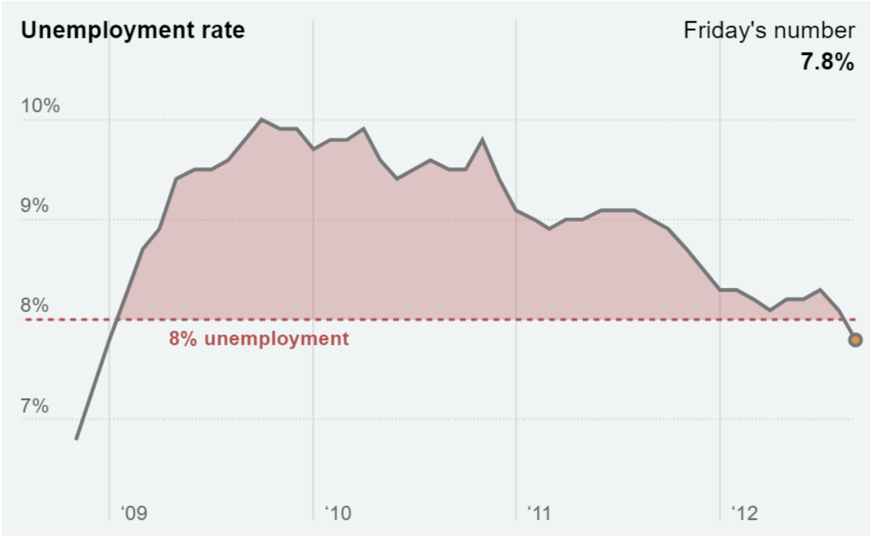

Figure 2 *Data visualization of the September 2012 jobs report from the perspective Republicans. Images by Mike Bostock, Shan Carter, Amanda Cox, and Kevin Quealy, New York Times. (https://archive.nytimes.com/www.nytimes.com/interactive/2012/10/05/business/economy/one-report-diverging-perspectives.html)*

By comparing the two visualizations and the message they carry, it is clear how the editorial perspective and bias will lead to different interpretations by the viewer. While the Democrats' graph focuses on the drop in the unemployment rate as shown by Figure 1, the Republicans' graph, Figure 2, focuses on the fact that the unemployment rate has never been below the 8% level in the past three years. Even though both visualizations, analyzed separately, seem to be quite simple, characterized by a high data-ink ratio and little decoration, these two different visualizations cannot be considered neutral, challenging the idea that minimalist visuals should be associated with the concept of neutrality.

Furthermore, the principle of elevating emotions recognizes that data often contains limitations and uncertainty. This acknowledgement challenges the traditional thinking that visual minimalism necessarily implies objectivity, therefore proving that simplicity rarely takes into consideration the complexities and uncertainties that characterize data. The explicit representation of uncertainty can be considered a means to foster transparency and trust in data visualization (Fleischmann & Wallace, 2005).

According to Data Feminism, including emotions in data visualization can enhance people's understanding of the data presented by acknowledging the emotional aspects of human experiences. Data Feminism wants to show data not just as a mere number or abstract quantification but as a representation of human experiences and human lives, which are inherently characterized by emotions (D'Ignazio & Klein, 2020). According to Data Feminism, including emotional elements in the visualization does not decrease the study's credibility but it enriches the narrative of the data by recognizing humans behind the datasets. This approach does not lead to the creation of bias, but it just helps the viewer to have a deeper understanding of the data and greater empathy toward the topic being analyzed. Using emotions in data visualization projects is meant not just to inform but to create a connection with the reader who is able to experience the data in a way that goes beyond simple reading. The work of Giorgia Lupi and Stefanie Posavec in *Dear Data* (Lupi & Posavec, 2016) shows the powerful role of emotional storytelling in data visualization. Through their project, they demonstrate how data narratives based on emotions can facilitate deeper learning, enhance memory retention and ensure more impactful communication, proving that emotion and rationality are not mutually exclusive but "complementary forces in the pursuit of knowledge" (Lupi & Posavec, 2016). By incorporating emotions like humor, anger, sadness and happiness, data visualization can become more accessible, engaging and memorable (Cairo, 2016). An example of the effectiveness of elevating emotion in data visualization is the Periscope Data project carried out in 2013 which visualizes the numbers of people killed by guns in the United States.

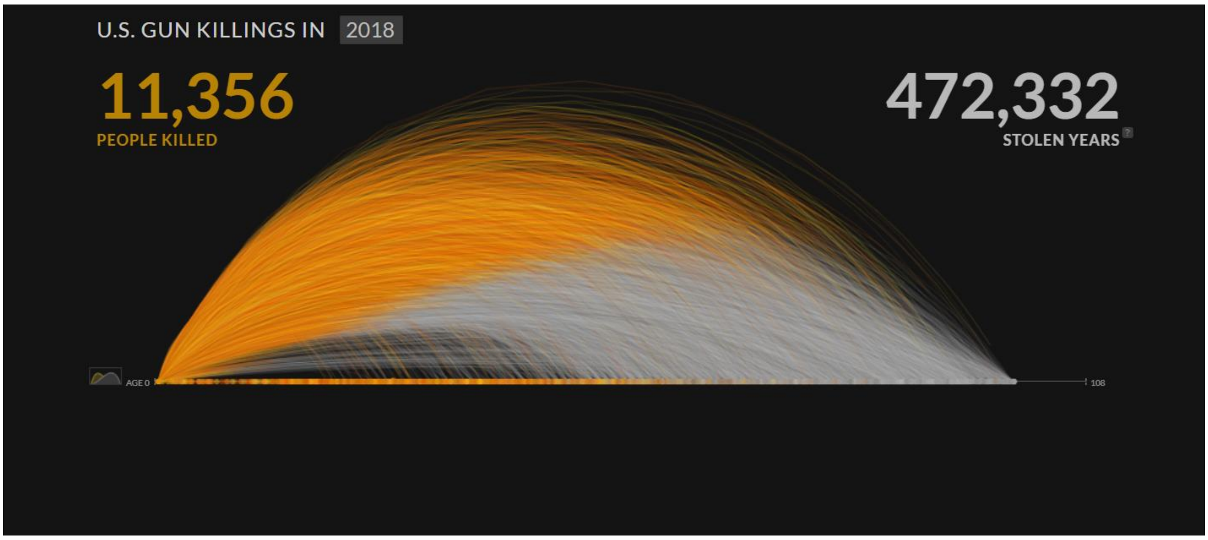

Figure 3 *An animated visualization of the "stolen years" of people killed by guns in the United States in 2013. Images by Periscopic. https://guns.periscopic.com/*

Figure 3 shows the stolen years of the people that were killed because of guns, each line represents a human life where the orange part stands for the years they have lived, while the white part represents the "stolen years". Each arc is labeled with the name and the age of the victim. It is the emotional aspect that sets this visualization apart from a more traditional visualization like the one published for the Washington Post "The era of "active shooters" in Figure 4.

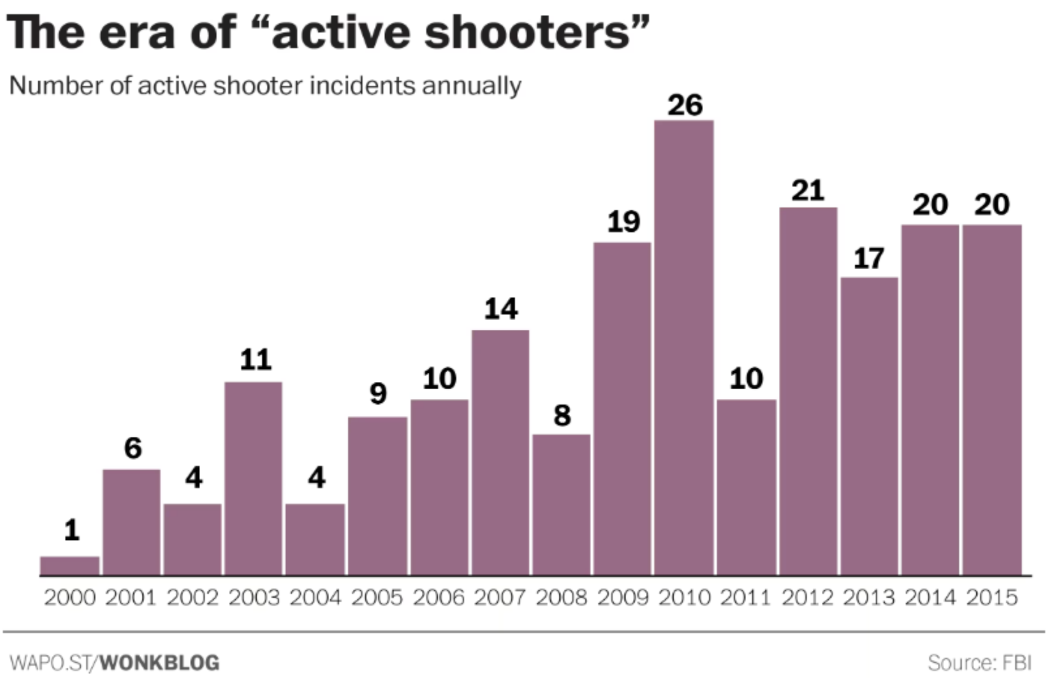

Figure 4 *A bar chart of the number of "active shooter" incidents in the United States between 2000 and 2015. Images by Christopher Ingraham for the Washington Post.*

The main difference is that the feminist visualization is able to transform numbers into people by giving to each death a name and the years stolen, while the second visualization only shows a simple bar chart with a very plain style that does not evoke any emotion.

By including emotion and embodiment, Data Feminism offers a different view of traditional data visualization practices, advocating for the recognition of the human dimension of data. This principle encourages the development of data visualizations that are not only informative but also empathetic and engaging, ultimately contributing to greater viewer involvement and, therefore, understanding of the data.

2.2.4 Rethinking Binaries and Hierarchies

The principle of rethinking binaries and hierarchies represents the 4th principle of Data Feminism which is meant to challenge and reevaluate the traditional classifications that perpetuate oppression. The most known classification is the gender binary classification, however, data feminism does not just limit itself to contrasting the binary male/female classification but also wants to include all the other types of traditional counting. These traditional classifications are the foundation of data gathering and affect how data is analyzed and therefore visualized meaning that “who gets counted counts” (Sager, 2018), while who is not included in the classifications becomes invisible.

Data feminism fights for the creation of a new framework of classification and counting that is as inclusive as possible to make visible those people and groups, the so called “minoritized people”, that have been invisible in the statistics so far (D'Ignazio & Klein, 2020). These structures are deeply embedded in our institutions, and they are part of scientific practices. They simplify the variety of identities that exist in the real world perpetuating the system of oppression that makes certain groups invisible. These restrictive categories do not take into consideration the experiences of those who most of the time suffer the most. It is, therefore, important to recognize the limitations and drawbacks of classifying people into these restrictive classes, making sure that this awareness leads to new ways of classification that have the potential to affect both data collection and analysis (Landstrom, 2007).

The binary male/female classification is the most evident example of how a certain group gets left out of statistics: non-binary people. Traditional methods of data collection regarding gender often offer only two options, erasing the identities of non-binary, genderqueer or transgender individuals. A Data Feminism approach might involve designing surveys and databases that allow respondents to self-identify their gender in open-ended formats or include a wider number of gender options. This would allow not only a more accurate reflection of gender diversity making data and therefore analysis more inclusive. Gender should be seen as a spectrum rather than a binary (D'Ignazio & Klein, 2020). In 2015, Facebook switched from the traditional “Gender” dropdown menu to a blank text field when registering on the platform, resolving the limitations imposed by the options in the dropdown menu. However, Facebook continued to classify the gender of its users with the binary classification of female/male when it comes to its advertising services, showing how it is who is in power that really controls how the data available gets utilized.

An example of data work challenging traditional classifications was the interactive visualization “Does the New Congress Reflect You?” created by Morris, Adolphe, and Salam for The Guardian after the election of 2018. From the visualization shown in Figure 5, the users can choose certain features and then they can visualize the number of members that share that characteristic. What is special about this visualization is that it is not limited to the binary classification of sex but provides an option for trans and non-binary people. The interesting fact is that when clicking that option, it shows that there is no member of the Congress belonging to that group.

Figure 5 *"Does the New Congress Reflect You?" Sam Morris, Juweek Adolphe, and Erum Salam for The Guardian (2018)*

This visualization is not meant to make complex data easy to understand as traditional visualizations try to aim, instead it operates in the opposite way: converting simple ideas into more complex and nuanced concepts. The goal is to highlight the complexities of the world in which we live and uncover the experiences of the most marginalized (Montañez, 2017).

2.2.5 Embrace Pluralism

Data work has originally been characterized by the belief that it is a solitary undertaking, since the people who gather, clean, analyze and in general work with data, most often, work alone. Data feminism advocates for the inclusion of multiple perspectives in the data lifecycle through the principle of embrace pluralism which relies on the idea that the most complete form of knowledge can be achieved by incorporating a diversity of perspectives throughout the data lifecycle. However, embracing pluralism first relies on the foundational idea that first of all people involved in data projects should acknowledge their own positionality and inherent bias, by making public their methods, choices and strategies. As David Weinberger (2009) claimed "Transparency is the new objectivity". Therefore, people in data projects should be open about their identities and the methods applied in their jobs.

Transparency should be coupled with the inclusion of multiple perspectives from various groups of society, especially those from marginalized groups, in all processes of data work including data analysis and storytelling processes. This includes involving the people directly affected by the topic under consideration. Data feminism believes that the combination of the technical skills of data people and the perspective and lived experiences of marginalized groups can provide a greater understanding and a more truthful representation of the phenomena under consideration. This makes sure that the outcomes of the data project are beneficial to the community of interest. Therefore, from a gender perspective, it involves including women and non-binary people.

Embracing pluralism means shifting from the concept of "doing good with data" to a model of co-liberation in which people from both the dominant and the marginalized groups collaborate to dismantle oppressive systems. "Doing good with data" model does not consider two positive outcomes that emerge with a co-liberation approach: knowledge transfer and building social infrastructure. Knowledge transfer entails a bidirectional benefit for both groups where the marginalized groups have the chance to gain technical skills to manage and enhance data projects independently, while technical collaborators are able to have insights on the community's lived experiences, broadening their knowledge and their understanding of the data under study through not only quantitative data but also qualitative data. The second element of co-liberation approach is the strengthening of community solidarity through the project, allocating resources not just to technical tasks but also to fostering community relations and support.

By following this principle of embracing multiple perspectives, the goal of data feminism is to make sure that people involved in data projects are not complete strangers to the datasets that are being handled. The local context of the data is fundamental to understand, and for this reason, the person behind those projects must collaborate with the people directly affected by the phenomenon under study. The concepts of co-liberation and participatory processes ensure that data projects are not only inclusive but also empowering, leading to outcomes that benefit all involved parties.

2.2.6 Consider Context

Data feminism challenges the notion that data is neutral and objective since data is created by humans and influenced by human decisions, methodologies and contexts, making data highly subjective. This means that data should not be considered a “raw input” but a cooked product shaped by structural and societal bias even before it is quantified or categorized (D'Ignazio & Klein, 2020). For this reason, it becomes fundamental to consider context when dealing with data and posing questions on how this data was collected, by whom, which methodologies were used and who is left out of the grand picture. Consider context is the sixth principle of Data Feminism that tries to emphasize the importance that numbers should never speak for themselves. Acknowledging the context in which data is produced should be the primary step in carrying out an accurate and ethical data project. Nowadays, there are an immense number of sources from which we can download data and it has become imperative now to understand the context in which that data was produced since data is the outcome of the context and the identities that collected it. While the open data movement advocates for the democratization of data, it often overlooks the necessity of providing context. Therefore, data should not be seen as an input that must be used directly in our data projects, but first people must acknowledge the context in which it was produced as well as its limitations.

The main problem with all the data that can be found online is that it comes without metadata or information about the context. Using data without context is very dangerous, as it can lead to misinterpretations and provide only a partial perspective of reality. For this reason, a feminist approach to data science emphasizes the significance of identifying what is absent from datasets to uncover biases in the "cooking" process of data. For example, health data collected primarily from urban hospitals might miss the nuances of rural healthcare challenges, skewing public health strategies that rely on such datasets. These gaps can reveal as much about societal inequalities and priorities as the data itself. Nowadays, it is becoming a common practice to provide documentation and user guides for datasets in order to increase transparency and fully consider the limitations associated with the data under consideration.

Considering context should not only be a step during the stage of data collection but it should also extend to how data is visualized and communicated. When visualizing data, not including context could lead to scenarios where the data is misunderstood by the viewers. By looking at visualization 6.1, the viewer may infer that white people are more likely to be diagnosed with a mental health disease. However, the truth is that white people appear to have a higher probability because people of other races are less likely to be visited and receive a diagnosis due to racial discrimination. The difference between Figure 6(a) and Figure 6(b) is that in the first one, context is not included and therefore, the results of the study undertaken are not captured by the reader, while the second visualization, by providing context, leads the viewer to truly capture the research findings.

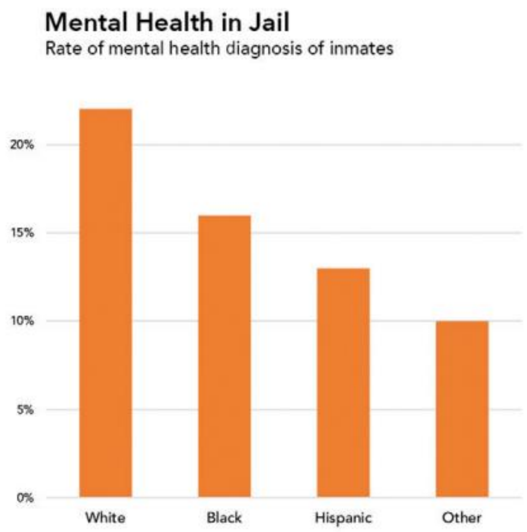

Figure 6(a) *Kaba F. et al., "Disparities in Mental Health Referral and Diagnosis in the New York City Jail Mental Health Service." Graphics by Catherine D'Ignazio.*

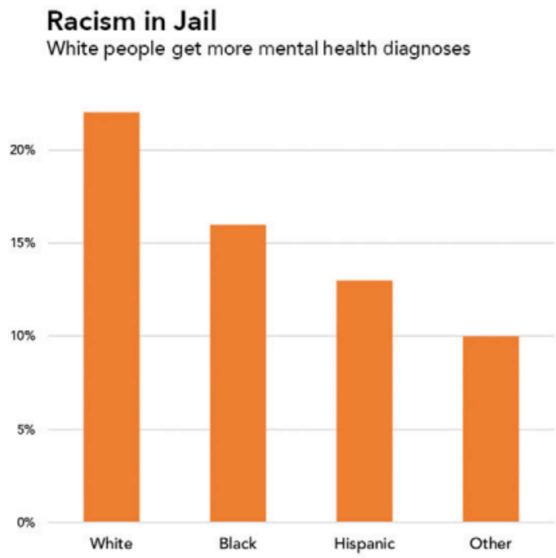

Figure 6(b) *Kaba F. et al., "Disparities in Mental Health Referral and Diagnosis in the New York City Jail Mental Health Service." Graphics by Catherine D'Ignazio.*

A feminist approach to data visualization might involve layering additional information or using interactive elements that allow users to explore how changing contexts affect the data, thereby offering a more nuanced and truthful representation.

2.2.7 Make Labour Visible

The last but not least principle of Data Feminism is making labor visible. This principle wants to give credit to all the people who contribute to data science projects. It highlights how it is important to also acknowledge the work that remains invisible, but that characterizes part of the process of the data project from data collection to data visualization. This is a critical step toward fostering a more inclusive data science practice. This principle allows people to consider every contribution as a critical step vital for the success of the project, therefore, also considering the contributions more peripheral, like data cleaning, as important as the others.

Data Feminism suggests some practices to make labor more visible depending on the project's context (D'Ignazio & Klein, 2020) such as project documentation, visualization annotations and public acknowledgments.

By making labor visible, Data Feminism advocates for data science practices that not only highlight but also value the diverse contributions necessary to bring data projects to life. This practice not only challenges the often hierarchical and obscured structures of labor within data science but also aligns with broader feminist principles of recognition and respect for all forms of work. Power imbalances characterize every process of data work, for this reason it is important to make labor visible, so that undervalued and invisible work receives the credit it deserves.

2.3 FEMINIST DATA VISUALIZATION

A major focus of Data feminism is Data Visualization. Data visualization can be defined as the graphical representation of information in order to communicate and make people understand complex data. Data visualization serves two scopes. On one hand, data visualization can be helpful to simplify data and uncover insights, on the other hand it is also a powerful tool to communicate findings and for storytelling, making these data more understandable and engaging to the audience. In fact, in today’s world, data visualization has become a vital tool for organizations that need to stay relevant and competitive in a continuously evolving landscape. Complex datasets are transformed into easy-to-understand visualizations that can guide actionability by capturing key information.

Data visualization has evolved along the advancements of society. The shift from traditional data visualization to feminist data visualization represents the recognition by the society that our social systems cause the perpetuation of inequalities. The growing awareness of the importance of adopting an intersectional approach, therefore considering race, gender and other factors, when analyzing data is influencing how data visualization will evolve.

2.3.1 Traditional Data Visualization: History and Principles

Traditional data visualization traces its roots back to the 17th century. Prior to this period, data visualization was only in the form of maps displaying cities, roads and lands used to help in navigation and exploration. Michael Florent Van Langren, a Flemish astronomer, is believed to be the first one to visually represent statistical data with a one-dimensional line graph in 1644 (Tufte, 1997). In the 1660s, the systematic collection and study of social data began in various European countries with the goal of informing the state about matters related to wealth, population, agricultural land, taxes and so on. The 18th century was characterized by the contributions of William Playfair, a Scottish engineer and political economist, who introduced the first statistical graphs like line graphs, bar charts, and pie charts. He argued that charts communicate better than tables of data. In fact, in his book "Lineal Arithmetic" he wrote: "As the knowledge of mankind increases and transactions multiply, it becomes more and more desirable to abbreviate and facilitate the modes of conveying information." However, at that time, data visualization remained rare because of a lack of publicly available data and experts in this area. Playfair's pioneering work showed how data visualization could have the potential to make data understandable and accessible, setting the basis for the principles that would later define traditional data visualization: clarity, accuracy, efficiency, and objectivity.

The 19th century is called the Golden Age of statical graphics by Friendly. This was due to the larger availability of datasets, the creation of government statistical offices and a growing recognition of the importance of having at disposal numerical data for better planning in all areas of public life. In this period, the contributions from different figures refined and expanded the principles of clarity and objectivity that would then characterize the evolution of traditional data visualization. A famous example of data visualization from that period is the map of cholera outbreaks made by John Snow during the London epidemic of 1854, in which the principles of clarity and efficiency were applied. Snow's map not only communicated vital public health information but also demonstrated how visual data representation could be used to solve real-world issues.

These principles became relevant in the second half of the 20th century, a period that was characterized by the work of Edward Tufte. He introduced six principles of graphical integrity in his book "The Visual Display of Quantitative Information", promoting for data-ink maximization and the elimination of "useless decorations" called chart junk. Data Ink maximization represents a core concept of Edward Tufte thinking, supporting the elimination of all ink that does not have a descriptive purpose. Maximizing the ink ratio would imply making the visualization clearer and more direct, focusing, therefore, the attention of the viewer just on the essential key information. Moreover, Tufte criticizes the use of decorations and design elements that may partially obscure key data insights and he argues that these elements have the potential to generate confusion and misinterpretation. The elements of simplicity and minimalism in visual representation strongly characterize all the work undertaken by Tufte. In fact, he emphasized the importance of creating visualizations that are informative and at the same time aesthetically pleasing, implying the need to integrate minimalism in the data work. According to Tufte, this type of visualization design would facilitate a deeper engagement with the data. The principles of Edward Tufte have influenced the evolution of data visualization encouraging practitioners to prioritize clarity, simplicity and integrity in their visual presentations.

Throughout its evolution, traditional data visualization has continuously focused on making complex information understandable and accessible. Practitioners have worked to make data visualizations free from any type of bias or subjectivity by applying the traditional principles of simplicity, clarity and objectivity. However, even though this approach is powerful in making data accessible, it is based on the strong assumption that using a traditional approach will always guarantee neutrality, therefore suggesting that data can be presented and interpreted without bias.

2.3.2 Evolution of Gender Data Visualization

During the second half of the 20th century, the need to use gender as a data dimension in visualizations became predominant. In fact, so far, existing approaches and methodologies had never revealed any contributions of women as well as the existing disparities between male and females. It is in this timeframe that gender data visualization started to spread as part of a broader movement that was advocating for gender equality and representation. During the 1970s and 1980s, as feminist movements gained popularity, there was an increasing need to start visualizing data disaggregated by gender in order to shed light on the discrimination that women were suffering in all areas of life. The feminist movement wanted to make their own experiences visible through data and the advancement in digital technologies allowed to carry out more sophisticated analyses and visualizations of gender data. Gender data visualization began to challenge traditional approaches by not only presenting data disaggregated by gender but also by starting to analyze and question how the biases and societal dynamics were influencing the collection, analysis and visualization of data. However, if on one hand, introducing gender as a data dimension represented a big advancement by providing more visibility to the lives of women, on the other hand, data visualization was still operating within certain boundaries of the traditional principles of data visualization that were causing the perpetuation of historical inequalities. Lupton (2016) and Criado-Perez (2019) highlighted the limitations of data practices in capturing the full spectrum of gender experiences. In fact, with gendered data visualization, gender classification is still binary, therefore, not considering the full range of gender identities.

2.3.3 Feminist Data Visualization

Feminist data visualization builds on the concept of gendered data visualization but extends it to be as inclusive as possible, taking into consideration and focusing attention on the lived experiences of marginalized groups. Feminist Data Visualization introduces the critical perspective that data and therefore also the visualization of the data is never neutral but is the outcome of the power dynamics in which they are produced. Feminist thinking acknowledges that data is always situated within the social, cultural and political contexts in which it is created. This new thinking goes against the traditional view of data visualization as a mere technical field that reflects what data holds. The outcome of data visualization is never objective, but it is deeply influenced by the perspectives of the people that create it, holding the power to shape the narratives. For this reason, data visualizations can perpetuate stereotypes and biases if they are not carried out following specific steps. Such biases can be perpetuated by continuing to represent gender with a binary classification, making invisible the experiences of non-binary people. Also, the choice of the data sources as well as the color scheme can reinforce biases and the structure of oppression in place, perpetuating the invisibility of marginalized groups. Feminist Data Visualization proposes a framework in order to contrast these ongoing stereotypes by providing certain steps and principles to follow when visualizing data. Feminist data visualizations challenge the traditional principles by valuing different ideas like the acknowledgment of the researcher's positionality and the understanding that all data visualizations are influenced by the creator’s perspectives.

Moreover, feminist data visualization is used as a tool to advocate social change since it is based on the belief that the complexities of the world and human experiences, particularly those related to gender, cannot be summarized by a simple and clean chart. Feminist data visualization also focuses on the application of the principle of intersectionality, considering the intersection of gender with other categories of identity like race and class to shed light also on those individuals that are marginalized twice, suffering from more complex discrimination.

Feminist data visualization challenges the principles of traditional data visualization in order to support approaches that are more inclusive and impactful.

2.4 COMPARATIVE ANALYSIS: Feminist vs. Traditional Data Visualization

By comparing feminist and traditional data visualizations, the aim is to highlight the different principles that characterize each way of visualizing data. Traditional data visualizations often value simplicity, minimalism and objectivity, sometimes at the cost of oversimplifying complex social realities and provide only a partial perspective. Feminist data visualizations, in contrast, emphasize inclusivity, emotional engagement and the ethical consideration of design choices. They aim to make visible marginalized groups and to challenge rather than reinforce existing power structures.

The comparison between feminist and traditional data visualization reveals significant differences in key principles and graphical design decisions.

2.4.1 Objectivity vs. Situated Knowledge

Traditional data visualization values the principle of objectivity which is based on the idea that if data is collected and analyzed in a careful manner then it is possible to visualize data in a neutral way. The importance of the principle of objectivity, as emphasized by Cairo (2012) in *The Functional Art: An Introduction to Information Graphics and Visualization*, is reflected in several key graphical design decisions.

- **Quantitative data:** Traditional data visualization prioritizes quantitative data: numbers, metrics and statistical results with respect to qualitative data. Quantitative data is associated with the concept of objectivity, in line with a scientific approach to understanding and solving problems.
- **Standardized Visual Graphics**: In order to promote a clear and straightforward representation of data, traditional data visualization encourages the use of standard charts like bar charts and line graphs. Because these charts are widely recognized by the audience, individuals can interpret the data presented quicker. Moreover, the traditional approach highlights the value of standardization in data representation since standard graphs, which are also used in academic and scientific papers, provide an

element of trust and credibility. The viewers are therefore more likely to trust the accuracy of the data.

- **Minimalist Aesthetic**: This principle prioritizes simplicity and minimalism when designing a visualization. The minimalist design, advocated by Tufte (2001), is based on the belief that data should be communicated as clearly as possible and the viewer should be able to capture the key message as quickly as possible. For this reason, this principle advocates for the avoidance of any element that could distract the viewers, with the goal of allowing the data to "speak for itself." The design should comprise clear labels, legible text as well as straightforward graphical elements, minimizing the use of symbols or icons that do not have a clear communication goal. Stephen Few, in his work *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (Few, 2009), emphasizes the importance of simplicity in design to support clear and effective data communication. Moreover, minimalism is associated with the concept of professionalism and therefore, greater reliance on and credibility of the data presented.
- **Restrained Color Palettes**: Linked to the concept of Minimalist Aesthetic, the use of limited color palettes reflects the belief that color in data visualization should be limited and should be primarily used to differentiate data categories (Tufte, 2001). Color should be used not as a decorative element but as a way to enhance the interpretability of the data (Few, 2009). One important element of the traditional approach is the absence of emotional connection with the use of colors. Colors can be emotional and may have cultural meanings, these elements may affect and influence the viewer perception and interpretation of the data presented (Wong, 2010). Dona M. Wong advocates for neutral color palettes that do not evoke specific emotions, making sure that the visualization remains as neutral as possible.

These principles are used to minimize the designer's influence on the interpretation of the viewer, trying to visualize data as neutral as possible.

On the contrary, feminist data visualization relies on the concept of situated knowledge, acknowledging that all work reflects the biases and perspectives of the creators. The biases of the design choices will then shape and influence data interpretation. This perspective, based on Haraway's (1988) notion of *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, is then translated into specific graphical design decisions:

- **Quantitative and Qualitative Data:** Data feminism believes that the reliance on just quantitative data oversimplifies complex issues, removing the context and depth that give data its full meaning. For this reason, it recognizes the importance of using both quantitative and qualitative data when making a visualization. By integrating numbers with the data narratives, data is not seen as just a mere number but as a reflection of human lived experiences. As a result of that, the viewers are more engaged on an empathetic level, making the visualization more memorable. Edward Segel and Jeffrey Heer in their study on narrative visualization highlight the effectiveness of combining quantitative and qualitative data in storytelling, noting how this blend can guide viewers through complex information, enhancing comprehension and retention (Segel & Heer, 2010)

- **Unconventional Chart Types**: Feminist Data Visualization wants to move beyond traditional charts to better represent the complexities that characterize the real world and therefore, the datasets. D'Ignazio and Klein in *Data Feminism* highlight the need to introduce unconventional chart types in order to promote greater viewer engagement and deeper understanding. Unconventional charts can help uncover hidden insights and patterns that a traditional chart may not capture as well as provide a greater amount of information (Yau, 2013). Contrary to the belief in the importance of standardization for the traditional approach, data feminism believes that if people are confronted with an unfamiliar chart, then they are more likely to analyze and reflect more on the data (Berinato, 2016). Unconventional charts help also include the context of the data which is considered of paramount importance by the feminist movement. This implies the inclusion of a more complex layout that includes different types of visual elements like text, video and the statements of the people involved in the study.
- **Interactive Elements**: One characteristic of feminist data visualization is the use of interactive elements which encourage the audience to better focus on the data presented and to explore the visualization by playing with the filters, slicers and other elements. These interactive elements enable the audience to increase their level of engagement making data more accessible and meaningful. Ben Schneiderman introduced the concept of "direct manipulation," where users interact with visual representations of data to uncover insights (Shneiderman, 1983). By enabling users to filter, sort and manipulate data, interactive elements empower individuals to design their data exploration process, answering their own questions and interests (Ward et al., 2015). Murray Turoff and Starr Roxanne Hiltz highlighted the educational potential of interactive data visualizations which facilitate learning and discovery.

- **Expressive Color Schemes**: Data Feminism uses color differently from the traditional approach. Colors are used to convey emotions and to make the narrative more memorable and not only as a tool to differentiate data points (D'Ignazio & Klein, 2020). The communication of the data is strictly connected to the use of colors according to *Visual Insights: A Practical Guide to Making Sense of Data* (2014) by Katy Börner and David E. Polley who emphasize the capacity of color to evoke emotions and suggest that the careful selection of color palettes can make data visualizations more impactful and memorable. Moreover, data feminism highlights the importance of being cautious when selecting specific colors since certain color palettes can perpetuate stereotypes or biases such as "pink for girls and blue for boys" (D'Ignazio & Klein, 2020).

By adopting these design principles, feminist data visualization enriches the data presentation of elements that promotes engagement of the audience on a more ethical and social level.

2.4.2 Neutral Presentation vs. Advocacy

One difference between traditional data visualization and feminist data visualization is the principle of neutrality. Traditional data visualization focuses on presenting data in a way that does not influence viewer's interpretation. The audience should not be influenced by explicit elements in the visualization and therefore, the creator should be careful when designing the visualization as to not insert his perspective in the narrative. This principle of impartiality is not at the core of feminist data visualizations which, on the contrary, often aim to advocate for social change through storytelling and data narrative. The principles of data feminism such as elevating emotions are based on the belief that data visualization should not be just a tool to inform but should be used to highlight existing inequalities and injustices to inspire action and change. Since Data feminism is based on the belief that all types of visualization, even those created using a traditional approach, carry the biases and viewpoints of the creator, it chooses to leverage this concept to highlight injustices, uncover the experiences of marginalized groups that are often invisible and mobilize action towards inclusivity and justice. If data is presented without context and without specific comments, then the key messages the visualization carries might not be understood by the viewer and might remain in the dark. The advocacy approach of feminist data visualization is important for several reasons. First, it challenges the belief of neutral data, highlighting how choices in data collection, analysis and visualization are influenced by the creators and can perpetuate existing power dynamics. Second, it makes data accessible to broader audiences by connecting numbers to real world social issues and by highlighting realities that have never been investigated before.

This approach helps increase public engagement but it also provides power to the communities of interests by making available evidence through which they can support their needs and ask for more rights.

3 METHODOLOGY

The research is based on a two-step approach that contributes to answering the research questions of assessing the potential impact of feminist data principles to uncover hidden narratives and provide a greater understanding of the data through feminist data analysis and visualization techniques. This two-step approach enables a comprehensive exploration of the topic analyzed: gender representation in the Italian parliament. In fact, the research aims first to uncover insights typically kept obscured by applying feminist principles in the data analysis phase. The insights identified in the first step are then used to design specific visualizations to assess the effectiveness of feminist data visualization techniques in facilitating a deeper understanding of the data presented.

The initial phase of the quantitative analysis of parliamentary data aims to investigate gender dynamics within the Italian Parliament over time. By examining not just the mere female representation but investigating on their contributions and involvement in parliamentary activities, this phase aims to uncover hidden narratives and highlight the progress made and the challenges that still remain.

The second phase involves designing specific visualizations and adopting both a traditional and feminist approach to compare them with the end goal of assessing the effectiveness of feminist data visualization techniques in promoting a deeper understanding of gender disparities compared to traditional data visualization methods.

The collection, analysis and visualization of the data are conducted following feminist data techniques in order to offer insights that bridge the feminist theoretical framework with a practical application in data analysis and visualization, therefore, contributing to the field of Data Feminism.

3.1 PARLIAMENTARY DATA COLLECTION AND ANALYSIS

3.1.1 Data Collection

During the data collection phase of this research, all the parliamentary data was retrieved from the official website of the Italian Chamber of Deputies. The official website provides every kind of information about deputies and their parliamentary activities. The data was retrieved from the website's SPARQL endpoint (https://dati.camera.it/sparql). SPARQL, which stands for SPARQL Protocol and RDF Query Language, enables the end user to retrieve and manipulate data that is stored in Resource Description Framework (RDF) format. Utilizing this language, the necessary databases fundamental to run the analysis, were then extracted in CSV (Comma-Separated Values) format, focusing on the period from the 10th legislature (1987- 1992) to the 19th legislature (2022 – up to now). The timeframe was chosen to provide the possibility of running trend analysis of gender representation in the Italian Chamber of Deputies over time which allows to observe the evolution of female representation in politics and therefore understand when and why there has been progress or draw backs in this matter. The first dataset to be retrieved concerned the demographics of all the deputies that have been elected throughout the ten legislations taken into consideration. For each deputy, the following information was retrieved: deputy ID, name, surname, gender, year of birth, affiliated political group, starting date of affiliation, end date of political affiliation and number of mandates. It is important to highlight that certain deputies were present multiple times in the dataset because of changes in their political affiliation. For this reason, the variables about the starting and end date of political affiliation were included. This dataset was used to analyze the trend of female representation over time. However, since the goal of the study is to apply feminist data principles, the research does not only investigate the numerical representation of deputies but collects other datasets to further advance the analysis. These other datasets are meant to allow for investigating how much women are really involved in political matters. The objective is to study not only the numerical "superficial" side of gender representation but also to investigate the "facts": the difference across gender in their involvement and contributions in parliamentary activities. These datasets were collected to analyze specific matters of interest regarding gender dynamics in the Italian Parliament, including their participation in parliamentary activities such as the law proposals made, the oral interventions, the voting pattern, and the relative number of absences.

Regarding the dataset about law proposals, a critical limitation emerged in the process. Detailed data about the status of law proposals, i.e. the success rate of law proposals, was only available for the most recent three legislatures, the 17th ,18th and 19th . This limited the analysis of this topic by focusing on just 3 legislatures rather than 10.

Not all the information was retrieved from the SPARQL endpoint. In order to analyze the oral intervention and therefore the text of the debates, the collection phase required an additional step other than downloading the dataset in csv format from the SPARQL endpoint. In fact, from the SPARQL endpoint, it was possible to retrieve only the following information about oral intervention: the deputy ID, session title, topic, the date and the link of the text that brings the viewer to another window. However, the entire text of the oral interventions was not possible to include through the SPARQL endpoint. This information was collected through Python programming language using Beautiful Soup Library, a method of web scraping which parses HTML content from the links associated to each deputy intervention. The function initiates by sending an HTTP GET request to the provided URL using the requests.get method. If the request does not encounter any error, then the response content is parsed using BeautifulSoup. The parsed HTML is then searched for paragraphs with a class attribute value of 'intervento', indicating the sections of the text that contain deputies' interventions. For each paragraph found, the function checks if it contains the deputy's name. If the deputy's name is found within the text, it means that the paragraph is part of that specific deputy's intervention. The text following the deputy's name is extracted, cleaned (removing the name and whitespace) and returned as the result.

All the datasets, obtained through SPARQL queries and web scraping, were instrumental in conducting a comprehensive analysis of the gender dynamics characterizing the Italian parliament. The timeframe spanning different legislatures ensures that the analysis is made by comparing the trend across different terms. This methodological approach facilitates a meaningful exploration of the impact of feminist data principles on data analysis and visualization within the context of the Italian Parliament.

3.1.2 Data Analysis

After the data collection process, the analysis phase represents the core part of the research, representing the essential element needed to uncover insights about gender dynamics within the Italian Parliament. These insights will be later used in the second part of the research to make specific data visualizations following feminist principles and run the survey. All these steps are done following the guiding principles of data feminism in order to uncover the whole data narrative behind each dataset. The datasets were cleaned and subsequently analyzed using Python as programming language. Python is a high-level, object-oriented programming language which is very popular and easy to use especially for data analysis and manipulation. The analysis was carried out by relying on the Pandas library which is used for handling and analyzing the CSV files obtained from the SPARQL queries, enabling efficient data processing, cleaning and transformation tasks.

The data analysis was structured in two parts: the first part concerned the analysis of the numerical composition in terms of gender of the Italian Chamber of Deputies. In this first part, the trend in the Italian Chamber was also compared to the trend of gender representation both for the government’s ministers and in parliamentary commissions, in order to understand whether the trend was similar or not. The first part was characterized by an intersectionality approach since data was analyzed by intersecting the gender dimension with other dimensions like age, the political group of affiliations and the geographical area of election.

The second part was focused on the analysis of the gender differences in the involvement of parliamentary activities. Therefore, going beyond the analysis of the mere composition of the Chamber of Deputies, the research wants to investigate how much female deputies are involved in these activities and whether there is a difference between male and female participation levels. This second part is characterized by the analysis of three macro topics concerning parliamentary activities: law proposals, oral interventions and voting patterns.

Before delving into each topic, an overall analysis of the average number of total parliamentary acts by gender was carried out. Not only was the average computed, but also the median was used in order to compare the different results and verify whether there were outliers that were skewing the result. After that, an outliers' analysis was conducted to identify the outliers as the ones that had values above or below the threshold of the upper and lower whiskers. The number of outliers across genders was identified and then a comparison between the mean with the outliers and the mean without the outliers was carried out. In the outlier's analysis, the values of the skewness and of the kurtosis were also computed. The skewness measures the symmetry of the distribution of the data, a skewness equal to zero stands for perfectly symmetric data while a positive skewness signals a distribution skewed with a right-end tail towards the positive values. The kurtosis measures how heavy or light the tails are compared to a normal distribution. The computation of the median and the outliers' analysis was carried out for each analysis involving averages.

The first macro area to be investigated regards the general trend in the different gender contributions in law proposals across legislation. After that, the analysis was focused just on the deputies that contributed to the law proposals as “First Signatory” since it represents a more active involvement in this parliament activity than the other type of classification available: “Other Signatory”. Subsequently, in scope with the research goals, this phase involved the analysis of the topic of the law proposals with a focus on topics about gender inequality, LGBTQ + inclusion and disability. These topics align with the principles of Data Feminism to investigate themes about marginalized groups, therefore putting emphasis on them. The research analyzes how many law proposals regarding these topics were made and whether there has been an increase or decrease across legislatures. Then, the analysis was focused on how many of these law proposals were proposed by women and by men. This analysis was also conducted using an intersectional approach to the political party dimension. For the last part of this topic, the research focused on the three last legislations, due to the limited availability of data highlighted in the data collection methodology, to analyze the status of the law proposals: how many proposals have been approved, not approved or are still going through the approval steps.

The second macro area analyzed involves oral interventions. The analysis started by computing the average number of oral interventions and the average length of these interventions by gender across the last three legislations. As stated previously, since averages were computed, the median and the outlier's analysis were also included in the analysis. Moreover, when analyzing the average number of oral interventions, only the debates long more than one hundred words were taken into consideration since the web scraping technique captured as separate interventions short claims or greetings that the deputies were saying to the president of the Chamber of Deputies before starting their oral discourse. After this analysis, the same structure about topic analysis was re-applied for the debates. This analysis is meant to investigate how much these themes about marginalized groups are discussed in the debates. A greater number of debates about these topics implies an increased awareness and attention to these social issues. An additional element of the topic analysis in this section was that the research investigated also the difference between the average length of interventions on these sensitive topics compared to the average length of all interventions.

The third macro area focused on the investigation of the different voting pattern between male and female deputies and the number of absences from voting. Moreover, the research also investigated the gender differences in the likelihood of voting differently from the voting of the people belonging to the same political party.

To better analyze the results of this phase, the findings were visually presented using Plotly, a popular open-source visualization library that allows the creation of interactive visualizations.

3.1.3 Feminist data principles

The phases of data collection and analysis were conducted applying feminist data principles, keeping always in mind the restrictions inherent to the initial datasets available. In fact, the first step involved recognizing the limitations of both phases to ensure that the analysis would not replicate existing biases but rather seeks to uncover and address them. In order to achieve this, several key principles were applied to these processes:

- **Challenge Power**: Both the data collection and the data analysis phases were designed to challenge traditional power structures by analyzing and highlighting the engagement and contributions of female deputies, an area which often is not explored in traditional political analyses. These analyses aim to recognize the role that women play in parliamentary activities and are meant to empower their contributions. In fact, one of the core feminist data principles applied was the commitment to making visible their contributions and their works instead of portraying only the negative experience and the situation of discrimination that women often suffer from.
- **Intersectionality**: This research applies the principle of intersectionality in the data analysis phase, a principle that is central to the framework of Data Feminism. In fact, the research analyzes not only the general participation of female deputies in parliamentary activities but also investigates in depth their contribution in law proposals that address social issues like disability inclusion, LGBTQ+ rights and gender equality. In this study, the gender dimension was intersected with topics that are strictly connected with marginalized groups, enabling to uncover insights into the advocacy roles that female deputies might occupy.
- **Consider Context and Rethink Binaries**: Data Feminism is based on the idea that data is never neutral but is the outcome of the social and cultural contexts in which it is created. Consider context is a principle applied in the data collection phase by recognizing the limitations of the dataset collected about deputies' demographics in which sex is classified as a binary option and does not give information about the existence of non-binary or transgender people. This binary classification does not allow to carry out an intersectional analysis on the other identities of gender, limiting the scope of the research only to female and male division. This research highlights this limitation of the data available and recognizes that this classification perpetuates existing biases and excludes specific groups from the narrative. Being aware of this is essential to provide context for the analysis made and interpret the findings in the right way.
- **Consider uncertainty in the data:** During the analysis phases, the research focused specific attention and acknowledgment on the uncertainties characterizing data. Among the uncertainties identified, the research reveals the limitations of available data regarding both the limited data available for specific topics, such as law status and the full text of debates, and the limitation of the binary classification regarding sex.

Moreover, the research aims to assess the variability in data by computing specific measures and detecting outliers that could skew the averages computed. For this reason, outliers' analyses were carried out. This approach contributed to a more transparent interpretation of the insights found, acknowledging that data are not always complete and do not always reflect all shades of reality.

By following the principles proposed by Data Feminism, the research aims to avoid perpetuating existing biases and contribute to a more comprehensive understanding of gender dynamics within the Italian Parliament. Both phases of data collection and data analysis were carefully conducted by considering all the limitations of the data and the impact of these limitations on data analysis. This research tries to show how the adoption of feminist data principles can transform our approach to understanding gender disparities in political representation.

3.2 VISUALIZATION AND SURVEY DESIGN

After carrying out the analysis phase, the research focuses on data visualization. It does so by employing some of the findings from the previous phase to investigate the comparative effectiveness of feminist versus traditional data visualizations in conveying information about gender dynamics within the Italian Parliament. The survey is designed to collect feedback on participants' perceptions and emotional responses to these two different visualization methods. This will contribute to a deeper knowledge of how different techniques of data visualization influence viewer engagement and understanding.

The survey displays two pairs of visualizations. Each pair focuses on the comparison of two visualizations, in which one adopts traditional data visualization principles and the second one adopts feminist data visualization principles. The survey is divided into two sections: while the first pair of visualizations is aimed at investigating the differences in viewers' perceptions in terms of clarity, ease of understanding, memorability, neutrality and emotional involvement, the second pair is meant to investigate only the difference in the ability to convey emotions.

3.2.1 Visualization Design Part 1

The first part of the survey is characterized by the first pair of visualizations. These visualizations display the same data but use different visualization methods: the traditional visualization approach and the feminist data visualization approach. Both visualizations show the evolution of gender representation within the Italian Chamber from legislature X (1987 - 1992) to legislature XIX (2022 - Present). Using Virtuoso SPARQL Query Editor, a list of deputies elected for each legislation was extracted with variables regarding gender, date of birth, geographical origin and mandate. Then, specific manipulations were made to the csv files using Python programming language and specifically the Pandas module in order to end up with a dataframe displaying for each legislation the percentages of female deputies and the percentages of male deputies. These percentages were used to design both visualizations, applying both traditional and feminist data visualization principles respectively. The aim is to compare these two different visualization methodologies. Each visualization is strategically placed within the survey to gather truthful and detailed feedback on their comparative effectiveness in conveying complex information.

Figure 7 shows the traditional visualization which is based on a simple static bar chart to depict the percentage of female deputies across different legislations.

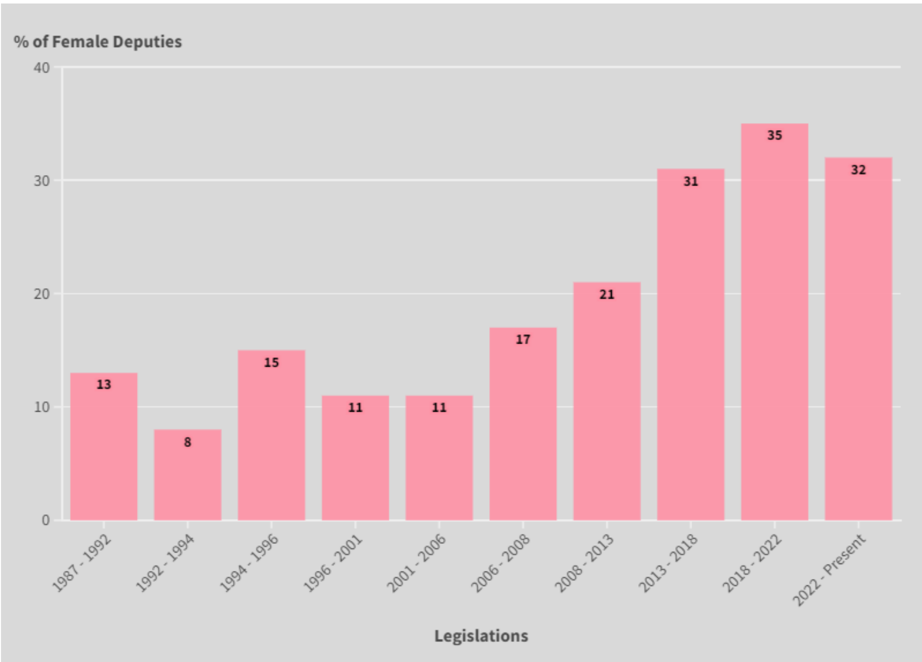

Figure 7 *Visualization 1 adopting traditional visualization principles representing % of female representation in the Chamber of Deputies across legislatures.*

Figure 7 follows important principles that are rooted in established theories of data visualization. These principles highlight the importance of delivering data in a clear way and in a manner that is immediately understandable. Following these principles, the data should then speak for itself without the influence of external narratives or biases. In order to make this first visualization, the following principles were adopted:

1. **Clarity:** The visualization was made with the aim to present the data in a clear and straightforward manner. This principle is central to traditional data visualization, according to which visualizations should be made as transparent as possible to the viewer, even though there is no familiarity with the topic shown. In order to apply this principle, visualizations must rely on the use of clean and simple shapes as well as a restricted color palette that should be only used to distinguish different data points. The visualization created for the survey follows this principle by presenting data in an uncomplicated manner. This is achieved by using a bar chart as graphic element which allows the viewer to easily read and interpret the data at one glance. Each bar corresponds to a legislative period, and the length of the bar represents the percentage of female deputies. This representation ensures that the viewer can quickly grasp the key data points without confusion. The visualization's design is made in a way to minimize distraction, in fact no other graphical elements are included that could make the viewer lose its focus on the trend shown. The use of the bar charts was meant also to allow a direct comparison of the percentage of female deputies across different legislative periods. The visualization clearly shows the trends over time of female representation also thanks to the consistent scale and spacing between the bars.
2. **Objectivity:** The visualization is designed to be as neutral as possible without introducing any extra elements that could introduce bias or a specific viewpoint in the reader's interpretation of the data. The visualization uses just a single color across all bars, which makes sure that no bar stands out over the others. The objective principle is strengthened by the choice of the title which is " % of Female Representation in the Chamber of Deputies" which only has a descriptive purpose of the data and does not provide any hints of the data and trend depicted. This allows the viewers to have their own opinions based on what they see from the visualization and without external influences.
3. **Traditional Gender Color Scheme:** The bar chart of the visualization is pink signaling the percentage of female deputies. The color pink was chosen because it is traditionally associated with women in most of the world's cultures. The use of this specific color is not only meant to make the visualization more appealing but also to enable a more rapid identification and association of the information presented. In this case, the bars in pink

are intended to be associated with the concept of female representation, ensuring that the message is well understood without the need to provide additional information.

The second visualization was designed using a feminist approach. It is a dynamic infographic which shows also here the evolution of the gender composition in the Italian Chamber. It does so by using a parliamentary chart composed of dots and each dot represents a deputy. Grey dots represent male members, while orange dots represent female members.

The dynamic visualization starts with a statement: "The road toward Gender Equality in the Italian Parliament is still a long one." The statement shown in Figure 8 is meant to engage the viewer emotionally and to provide context from the beginning, symbolizing a journey through time and drawing from the beginning attention to the challenge of achieving gender equality.

Figure 8 *Visualization 2, initial statement: "The road toward Gender Equality in the Italian Parliament is still a long one."*

The visualization takes the viewer through successive legislative periods, beginning with legislature X (1987 – 1992). By illustrating these periods dynamically, the viewer can observe changes over time in the gender composition of the Italian Parliament, offering a temporal narrative of progress and regression.

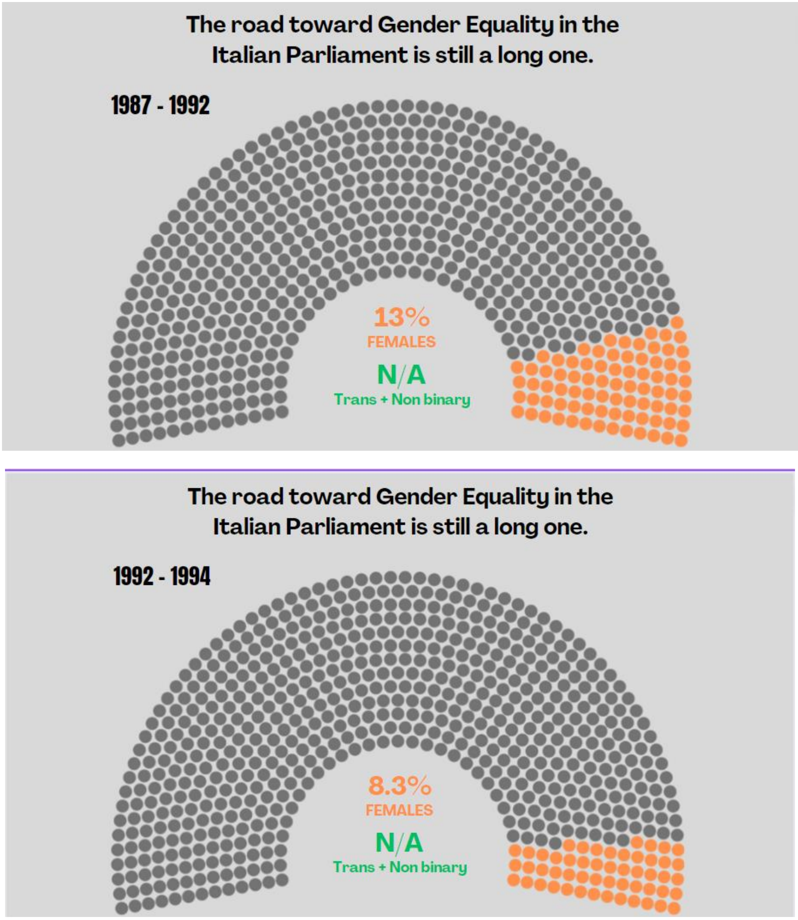

Figure 9 *Visualization 2, shots of the dynamic visualization showing changes over time in the gender composition.*

This visualization displays, directly below the parliamentary chart, the percentage of female deputies. This element was designed to provide the viewer an immediate understanding of the number of deputies. In fact, it allows for a straightforward translation of the visual dots into quantifiable numbers. Below the percentage of female deputies, the visualization includes the representation of "Trans - Non-binary". However, this label is always characterized by the word N/A in every legislation, indicating that the data is not available since the classification of the deputies is done using the conventional binary classification of gender.

The last shot focuses on the current visualization and reinforces the message that there is still a long road to achieve gender equality in the Italian parliament. This visualization is designed with the objective to incorporate specific feminist data visualization principles to differentiate a traditional visualization from a feminist one.

1. **Examine Power:** The visualization is based on the principle "Examine Power" by stating at the beginning "The road toward Gender Equality in the Italian Parliament is

still a long one". The visualization wants to focus the attention of the audience on the issue of gender inequality and how power is currently distributed in politics.

2. **Elevate emotions:** The principle "Elevate Emotion" is included in every frame of the visualization: starting from the initial statement that evokes negative emotions since the road toward gender parity is still a long one. The use of dots to represent each individual deputy is a way to humanize data, a core element of feminist data visualization. The visualization makes the audience more connected with the data by humanizing the dots and by not reducing deputies to mere aggregate numbers. This approach increases the likelihood in the creation of empathy between the data and the reader. As the visualization unfolds showing the evolution of female representation across legislations, the slow progression of the numbers and, in some cases, the stagnation across different legislative periods may evoke strong emotions for the readers. Moreover, the visualization highlights the limitations of the datasets about the gender classification by showing the absence of data for trans and non-binary individuals. This element further emphasizes inclusivity and acknowledges those typically excluded. The final frame ends with the current percentage of female representation to remind the audience of the progress made and the strides still need to be done, evoking a sense of encouragement or a call to action. This approach focuses on the narrative and storytelling to engage on a deeper level the viewer.
3. **Rethink Binaries and Hierarchies:** The visualization highlights how the data collected are limited to the conventional binary classification of gender, excluding from the narrative trans and non-binary. Their inclusion in the visualization challenges the viewer to think beyond the traditional standards of classification and consider the invisibility that this group of people is suffering from. By emphasizing the lack of non-binary and trans data (N/A Trans - Non-binary), the visualization acknowledges the existence and erasure of non-binary individuals in political data. Moreover, to distinguish between male and female deputies, the visualization relies on specific color coding that do not perpetuate the existing stereotypes of pink for girls and blue for boys. It challenged them by using grey for males and orange for females. These colors were chosen to highlight the numerical representation of female deputies so that the viewer is focused more on the increase or decrease of the number of orange dots.
4. **Consider context:** In the feminist visualization, the context is included firstly with the initial statement "The road toward Gender Equality in the Italian Parliament is still a long one," which invites the viewer to consider the historical journey and ongoing

challenges of gender equality. Legislative periods and related data change as the video goes on, resulting in a final current level of female representation at 32%. By highlighting this change across time, the visualization urges viewers to take into account both the historical and current context. By demonstrating how the Italian Parliament's composition has changed over time, viewers are able to observe patterns and the advancement (or lack) of gender representation, emphasizing the dynamic nature of social change.

5. **Make Labour Visible:** at the end of the visualization (Figure 10), contributions are made visible highlighting where the data come from, who made the visualization and the people behind the theoretical framework that was followed to make such visualizations.

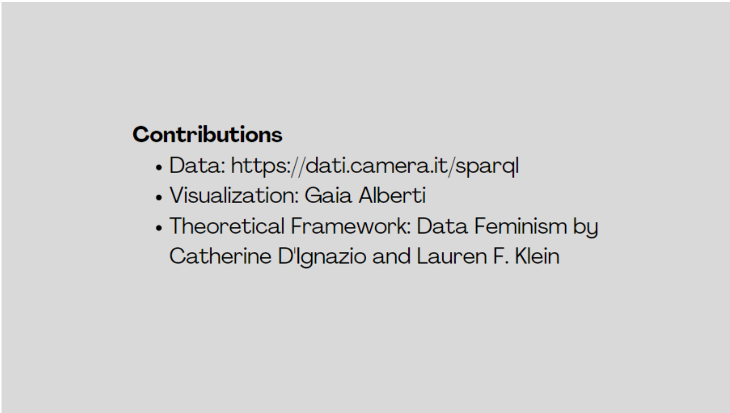

Figure 10 *Visualization 2, contributions at the end of the visualization.*

This visualization is designed to engage participants on an emotional level and to encourage them to acknowledge the progress and challenges of gender equality in political representation. The dynamic visualization is designed to apply the feminist data visualization principles of storytelling, emotional engagement, and inclusivity.

3.2.2 Visualization Design PART 2

The second part of the survey is also characterized by a pair of visualizations similar to the previous pair. However, this part is intended to investigate just the emotional aspect of the two visualization approaches. The second pair of visualizations display data about law proposals on gender equality and shows how many have been converted into official laws and how many have not been passed. This second pair of visualizations adopt the same principles outlined before for both the traditional visualization and the feminist visualization.

The first visualization adopts traditional data visualization principles by applying the principle of clarity, ease of comparison and objectivity. The visualization is designed using a stacked bar chart and characterized by the use of pink to highlight the law proposals that have been passed. The visualization is clear and provides a straightforward message. Moreover, the title of the visualization is neutral by not providing any viewpoint to avoid influencing the viewer's interpretation of the data.

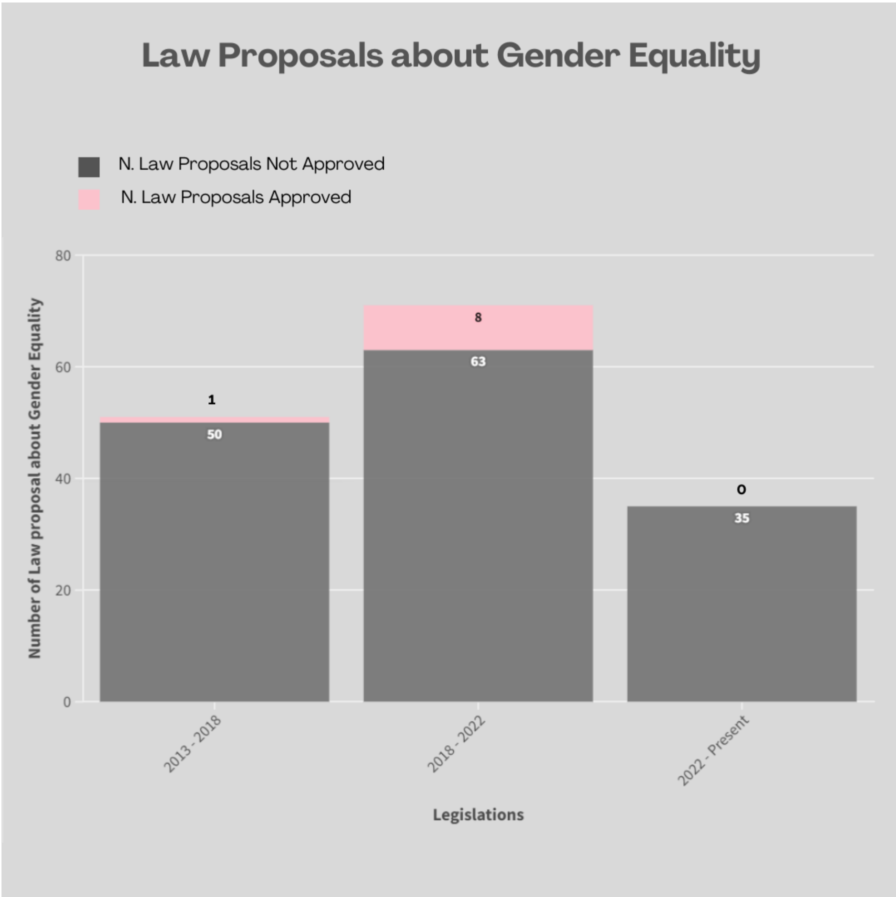

Figure 11: *Visualization 1 of PART 2, number of law proposals on gender equality that have been converted into official laws.*

The second visualization adopts a feminist approach by including the principles of elevating emotions, consider context and examine power. As in the first set of visualization, this feminist visualization is dynamic. For the last three legislations, it shows the count of law proposals not approved and the count of law proposals approved. However, the data are displayed differently from the traditional visualization. In fact, the counts dynamically increase as the visualization unfolds showing the general trend of the outcome of these law proposals.

These shots (Figure 12(a) – Figure 12(b)) incrementally update to reflect the growing number of law proposals approved and not approved. For each update also the title of the law proposal associated changes. making it clear that these numbers represent concrete initiatives. By providing context, the visualization engages the viewer on a more personal level, moving beyond abstract numbers to highlight the tangible contributions and efforts aimed at tackling issues of gender inequality. This visualization is, therefore, intended to engage the viewer on an emotional level by showing how there are real human efforts behind each number. Moreover, the dynamic increment of the count of laws not approved is meant to motivate the participants to advocate for changes.

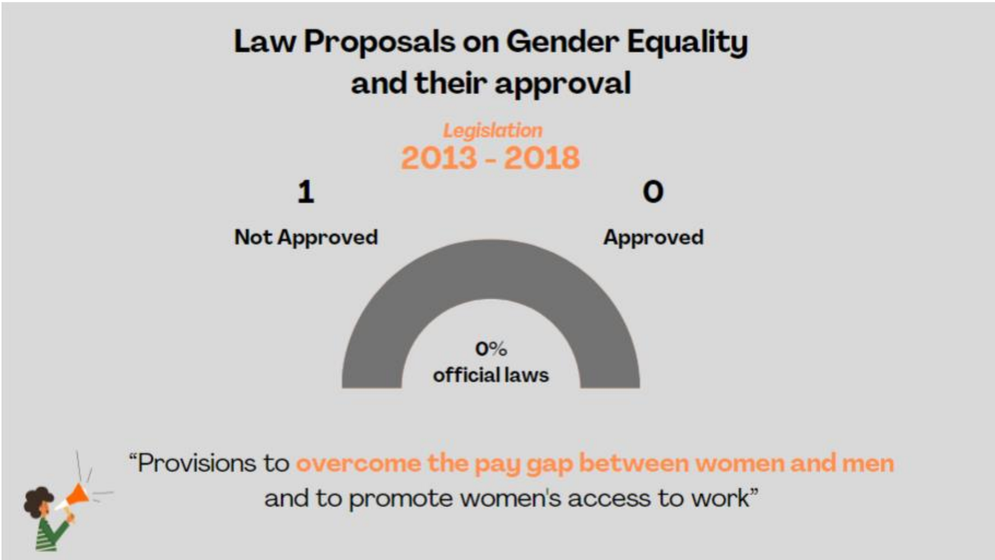

Figure 12(a) Visualization 2 of PART 2, dynamically presenting the number of law proposals on gender equality that have been converted into official laws.

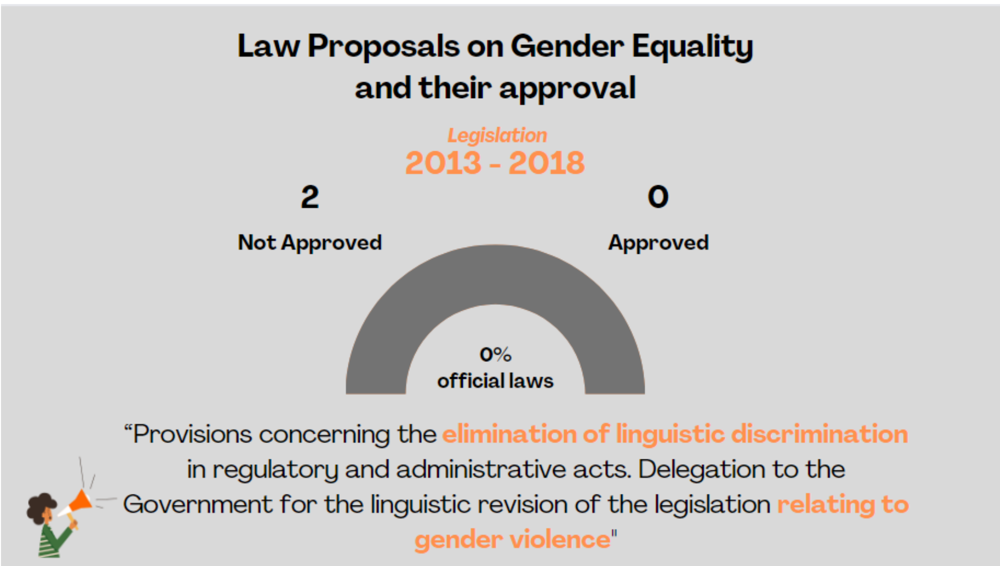

Figure 12(b) Visualization 2 of PART 2, dynamically presenting the number of law proposals on gender equality that have been converted into official laws

3.2.3 Survey Structure

1. **Introduction**

The survey starts with a short written paragraph meant to explain the goals of the study as well as how the data gathered will be used for the purpose of the study. This initial section is divided into three sections.

The first lines explain the objective of the research by informing the participants on the structure of the survey and on the subject matter. Knowing the topic of the survey and the sections of the survey will make participants understand the significance of their contribution and the context of the study.

The second section states that the responses will be anonymous and that the email used to fill out the form will not be gathered. Moreover, this section also seeks informed consent from participants by explaining how the data they provide will be used for research purposes. These statements aim to build trust with participants, encouraging them to contribute to the research.

The last section highlights the importance of the contribution of each individual in filling out the survey to carry out a research of quality.

2. Demographic Information:

In order to analyze data by groups of people sharing different characteristics, it is instrumental for the research objective to insert, at the beginning, a section to investigate the demographics of the participants. Data about age, gender and educational background were gathered through compulsory multiple-choice questions. The question about gender goes beyond the traditional binary classification of gender and provides multiple options, including trans and non-binary, as well as the possibility to write a different answer if it is not available among the options given. This type of question has been formatted in a way to be as inclusive as possible and challenge the traditional analysis of male/female classification. The scope of collecting these demographic data is to understand whether different groups of participants perceive and interpret visualization differently.

3.Visualization Sections:

Part 1 of the survey displays the first pair of visualizations previously described, first separately and then side by side. This section begins with the traditional visualization which is then followed by different types of questions. The same questions are then displayed also for the feminist data visualization. The questions of this section are the following:

1. **Open-ended questions:** the survey aims to investigate through two open-ended questions what are the key information that the participants identify as well as the graphical elements that stand out. These questions are meant to understand which type of insights and elements participants have spotted. The goal is then to compare the answers based on the traditional visualization with the answers based on the feminist visualization to highlight any differences that are caused by the different technique designs.
2. **Likert Scale questions:** participants are asked to express their level of agreement using a scale from 1 (strongly disagree) to 5 (strongly agree) assessing 5 different characteristics of the visualization: clarity, memorability, neutrality, curiosity, and emotional impact.

The traditional visualization serves as a benchmark for comparing how different data visualization techniques can impact the participant's understanding and engagement with the data. After this first section, the dynamic feminist visualization is displayed which is then followed by the same series of open-ended and Likert scale questions, mirroring the structure used for the traditional visualization. This allows the direct comparison of the answers between the two visualization techniques.

In the final section of part 1, both visualizations are presented side by side, and participants are asked through an open-ended question to point out the main differences between the two visualizations. After that, participants must answer 5 closed-ended questions in which they must claim which visualization is better with respect to the following 5 areas: clarity, ease of understanding, neutrality, emotional engagement, and overall preference.

The placement of these visualizations within the survey, first separately, then side-by-side, was meant to facilitate a layered understanding of participant responses. This setup aims to assess not just the immediate impressions of each visualization but also allows for a deeper comparative analysis of their effectiveness in communication, emotional impact and overall preference.

Part 2 of the survey has a similar structure: it displays first the traditional data visualization, then the feminist one and at the end it shows them side by side. However, the questions in this case focus just on emotional involvement. After each visualization, the survey investigates if the visualization has impacted participants emotionally and if the answer is yes, they are asked to explain the type of emotion. In the final section, the survey asks through a closed-ended question which visualization is more emotionally engaging.

Through this structure, the survey aims to test the hypothesis that traditional visualizations, while clear and straightforward, may not engage viewers as much as the feminist visualizations do in terms of storytelling and emotional elements. Therefore, the survey seeks to assess through quantitative and qualitative data how feminist principles can enrich data visualization practices (Appendix A).

4 RESULTS

The results of the research are divided into two sections: the results of the data analysis and the survey results. In the first part, the focus is on highlighting the insights found during the analysis of parliamentary data about gender representation using feminist data principles. In the second part, the aim is to show how different feminist and traditional data visualization, designed with the data previously collected, perform against specific attributes: clarity, memorability, neutrality, curiosity and emotional engagement.

4.1 Analysis Results

The aim of the data analysis phase is to better understand gender representation in the Chamber of Deputies, uncovering gender inequalities and hidden insights on the role of female deputies through the application of feminist data principles. In fact, the aim of the research is to assess the ability of the Data Feminism framework to uncover hidden narratives. By applying feminist data principles, the analysis does not just investigate the numerical representation of female and male deputies but goes beyond it by highlighting their engagement and the quality of their participation in parliamentary activities. Therefore, through the application of these principles, the analysis is not merely based on traditional numerical metrics, since they often fail to highlight the real contributions of individuals and fail to provide a detailed picture of the topic taken into consideration. In this case, the application of feminist data principles aims to challenge traditional data analysis by highlighting the roles and contributions of female deputies which are often neglected or underrepresented in traditional political analysis. These principles are based on the work "Data Feminism" by Catherine D'Ignazio and Lauren F. Klein, which allows to carry out a comprehensive analysis of data. In fact, Data Feminism highlights the importance of taking into consideration power dynamics, intersectionality in data practices as well as focusing on the unseen experiences of marginalized people. In analyzing gender representation within the Italian Parliament, the adoption of the Data feminism framework has been instrumental in revealing the progress and the systemic issues that affect gender representation in Italian politics.

The numerical analysis of gender representation reveals a steady increase in female representation from legislature 14th, with 11,4 % of female representation, to legislature 18th with 35.7%. From term 10th to term 14th, the trend has been characterized by a lot of fluctuations. Despite the overall gains, there has been a decrease to 32% in female representation for the current legislation. This trend depicted in Figure 13 emphasizes how fast progress can be achieved but also how quick female representation can decrease, highlighting the volatility in these gains.

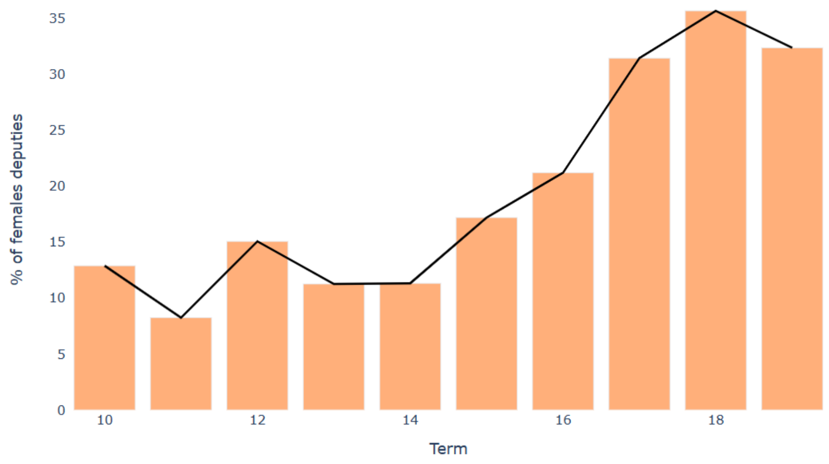

Figure 13 *Visualization about the % of female deputies in the Chamber of Deputies from term 10th to term 19th*

These insights highlight how important it is to always analyze data by taking into consideration the context of the data under study, a key principle of Data feminism. Moreover, additional analysis of female representation in parliamentary commissions as well as in the government ministries revealed a similar trend in the percentage of female deputies as in the Chamber of Deputies. Relating to parliamentary commissions, despite an initial increase, the percentage of female deputies have significantly decreased in the current legislature, particularly in leadership roles. The decrease represents a major drawback: starting from 32,09% of women with presidential roles in the 18th legislature to 22,1% in the current legislature. Such variability and fluctuation in female representation is also visible by looking at the composition of government ministries. While with the previous government under Prime minister Draghi there has been a huge increase in female representation peaking to 42,21%, with the current government of Giorgia Meloni, the positive trend suffered a drawback, declining to 25,5%. Even though Giorgia Meloni is the first female prime minister in the Italian history, there has been a decline in female representation across all other parliamentary bodies, indicating that a progress in one area does not imply an advancement in all the others. This variability prompts critical questions about the strive to gender parity as well as the influence that the political context and the type of government may have on gender dynamics. Once again, this highlights how the principle "Consider Context" of Data Feminism is fundamental in understanding data within its specific socio-political framework.

A more in-depth investigation in the political participation of Italian deputies revealed a different mean duration of the parliamentary role between genders, where male deputies are elected for 2.1 legislatures compared to 1.71 for females. Even when adjusting for outliers, therefore by taking into consideration the median, the gap increases becoming 2 for males and 1 for female, suggesting that male deputies are more likely to serve multiple terms. These findings highlight systemic biases and by uncovering it, the research challenges these existing power structures as the framework of Data Feminism advocates.

Transitioning from the quantitative analysis of gender representation to the quantitative and qualitative analysis of the engagement level in parliamentary activities, it is essential to highlight the importance, following data feminism principles, of not merely accounting for the presence of women but also illuminating their involvement and contributions. With respect to the level of involvement in parliamentary acts of female deputies as "First Signatory" and as "Other Signatory", the analysis shows how female deputies contribute substantially to the parliamentary activities, demonstrating a more active engagement with respect their male counterpart as shown in Figure 14.

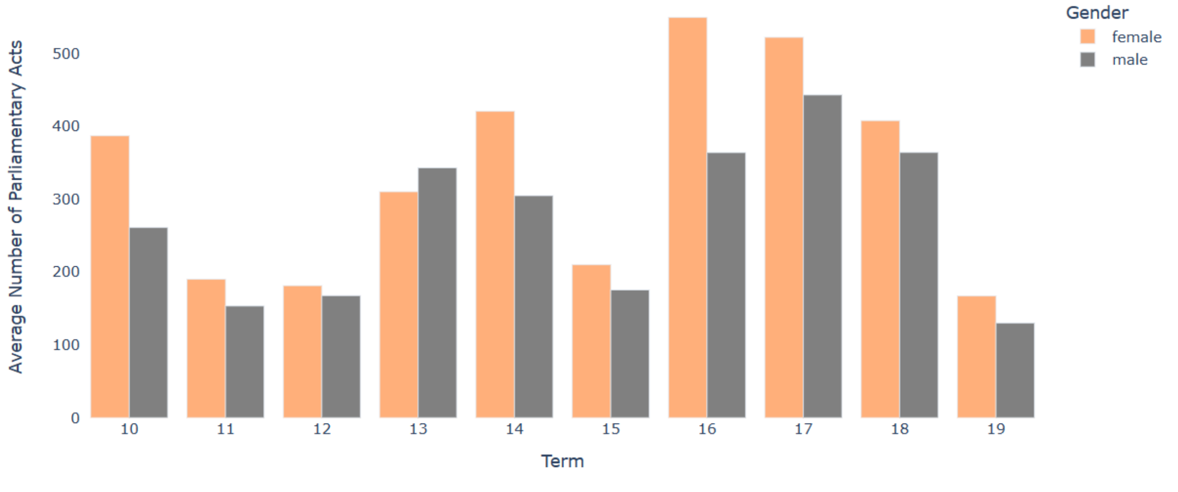

Figure 14 *Average Number of Parliamentary Acts as both "First Signatory" and as "Other Signatory" by Gender Across Terms.*

When analyzing the data using the median, the difference between males and females remains roughly the same. By carrying out an outliers' analysis for the current term, the findings highlight the presence of 17 outliers for males (6% of total male deputies) and 6 outliers for females (4% of total female deputies) with a greater standard deviation among the female group compared to males. Nonetheless, since the distributions for both groups are skewed to the right, the median shows very similar results across terms, suggesting a greater female involvement. This information shows how women' contribution is essential for parliamentary legislative efforts and therefore, it should not be undervalued.

Subsequently, the research focuses on a specific parliamentary act: law proposals. The average number of law proposals when considering only "First Signatory", is greater for female deputies in most legislatures. However, when computing the median, the difference diminishes in most legislatures and parity is reached in the last two terms. The standard deviation for females results to be bigger than the standard deviation for males indicating that there are few highly active females that inflate the mean.

The next step involves investigating on the topics of these law proposals particularly those regarding themes sensitive to minoritized groups such as gender inequality which encompasses topics like discrimination, sexism, domestic violence, and equal rights as well as LGBTQ+ inclusivity and disability. Figure 15 highlights how the number of law proposals regarding the topic of gender equality has steadily increased throughout the terms, therefore, symbolizing a greater effort and a greater attention in politics to reach gender equality. These laws have been proposed by a greater number of females than males, a trend that is persistent in almost all legislative terms.

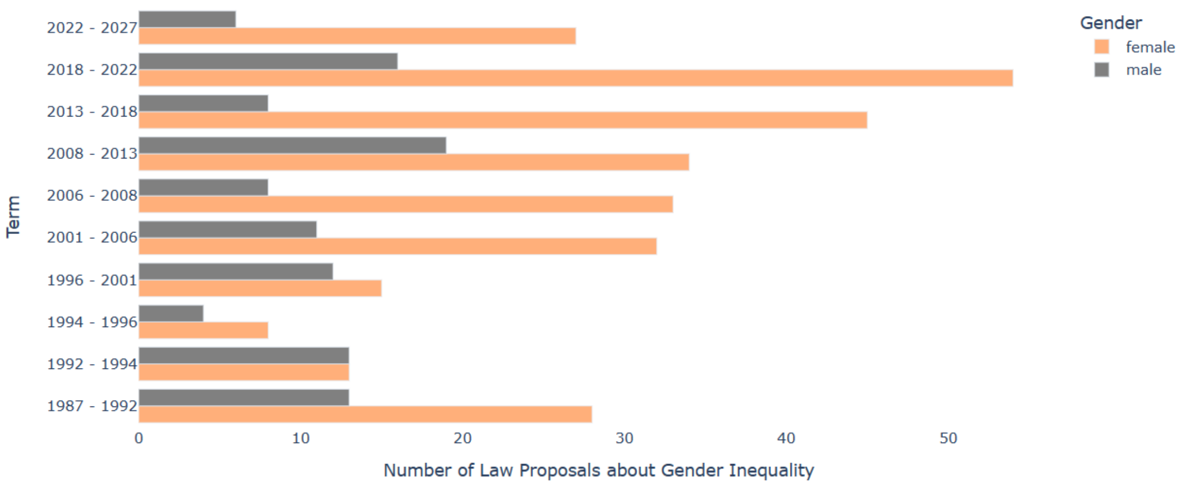

Figure 15 *Visualization about the Number of Law Proposals about Gender Equality by Gender and across Terms*

This greater involvement by female deputies highlights the importance of the role of female deputies in advancing and addressing social issues. However, even though there has been an increase in the number of law proposals regarding gender equality, only a very small proportion has become an official law passing all the steps of the legislative path. The highest approval rate occurred in the 18th legislatures, where 8 law proposals about gender equality out of 67 became official law.

By taking into consideration only the laws regarding LGBTQ+ inclusivity and rights, the number of laws proposed in the current legislation are so far just three. Moreover, in the previous legislature the law proposals were five. These small numbers highlight the little interest toward this topic which is often advanced by deputies that are themself part of the LGBTQ+ group. Moreover, none of these law proposals was passed and therefore none became official laws. The low number of proposals and the lack of any successful approvals highlight the challenges in advancing LGBTQ+ rights during the parliamentary activities.

Lastly, law proposals regarding disability, another sensitive topic for a specific minoritized group, were analyzed to further investigate the difference in gender involvement in such laws. This analysis indicates that there has been an increased number of law proposals focusing on advancing these rights with a higher number of proposals initiated and proposed by female deputies. However, with regard to the status of these proposals in their legislative path and their enactment, only one law passed in the 17^{th} term and none in the 18^{th} and 19^{th} term.

The analysis of women contributions in parliamentary activities incorporates feminist data principles by emphasizing their engagement in proposing laws that are beneficial to face social issues and advance the rights of minoritized groups. Moreover, this analysis highlights how the increasing efforts of deputies for advancing theses topic do not lead to substantial legislative outcomes, indicating how much important it is to go beyond the mere number of law proposal to really assess the concrete effectiveness of these actions. The information collected highlights how there are still obstacles in the advancement of the rights of minoritized groups which is strongly influenced by the current political and social context.

Proceeding to other analysis about the contributions of Italian deputies, data about oral interventions in the Italian Parliament reveals important insights that demonstrate how feminist data principles have the potential to uncover hidden narratives. During the 18^{th} legislature, the average number of oral interventions was roughly equal between genders, with males being slightly more active on average. However, by computing the median, the difference between genders diminishes, indicating that there is a certain number of male deputies that inflates the mean by making a very high number of oral interventions. The computation of the median and the outliers' analysis highlight the importance of using different statistical measures to have a more accurate picture of participation. This is in line with feminist principles that advocate for a comprehensive and inclusive analysis. As with law proposals, the topics of the oral interventions were examined, focusing in particular on the social issues highlighted above. The topic of gender inequality was discussed 224 times, however the number of deputies involved in this discussion was only 133. This symbolizes that there are certain deputies that tend to bring this topic up more than once, therefore suggesting the presence of a subgroup of parliamentarians that are highly involved with these topics. It is important to highlight how female deputies discussed gender inequality 165 times compared to the 59 times of male deputies and that women's intervention on average tend to be longer. Moreover, the analysis revealed that overall, the discussion regarding these social issues was longer with an average of 1514 words compared to the general average discussion length of 974 words. This big difference highlights how much these social themes are complex and significant.

The oral interventions were also examined with respect the LGBTQ+ social themes by examining both the support and opposition within the Parliament. Among the deputies that intervene in this topic, 62% of female deputies made a speech in favor of the LGBTQ+ rights compared to 44.44% of male deputies. While with regard opposition, 33% of males was against these topics compared to only 9.5% from females. The rest of the deputies' interventions were classified as "neutral" since they did not express explicitly their positioning. The length of discussions on LGBTQ+ issues is also greater than the average general discussion length, further indicating the heightened engagement and both significance and complexity of these debates. Similarly, the topic of disability rights shows a more frequent and involved participation of female deputies with 156 interventions compared to 126 oral interventions by male deputies.

Concluding the analysis with the exploration of the voting patter for the 18th term, the analysis highlights a very similar pattern for both female and male deputies. By looking at the average voting rate of deputies, it can be noticed how male deputies did not vote 30.6% of the times compared to around 28.5% for female deputies. Moreover, the standard deviation for the number of absences among the male group is much higher than the female group even though both distributions are highly skewed to the right, indicating the presence of a minority group which is significantly less likely to vote either for official missions' reasons or types of absences that in the dataset are not explained. Moreover, female deputies exhibit less dissent in their voting behavior compared to males, suggesting that when they do vote, they may be more aligned with their party's stance.

By applying feminist data principles, the analysis provides a comprehensive understanding of gender dynamics in the Italian parliament. Specifically, it shed lights on the often unseen contributions and engagement of women in parliamentary activities. This analysis aims to shift the narrative from viewing women only as statistical figures or victims to recognizing their active and impactful roles. Moreover, the analysis demonstrates not only the active engagement of women but also the key role that they have in advancing social issues both with regard law proposals and oral debates. This type of analysis contributes in revealing hidden narratives and disparities that might remain obscured without such an in-depth analysis.

4.2 Survey Results

After analyzing parliamentary data, key insights were used to design different types of visualizations following the feminist and traditional approach. The visualizations were used in the survey to assess the different impact of these visualization methodologies with respect to the attributes of clarity, memorability, neutrality, curiosity and emotional engagement, The survey showed relevant insights about the application of feminist data visualization principles and their impact on viewer engagement and understanding.

The survey is based on 104 answers characterized by 52 female and 52 male respondents (Figure 16).

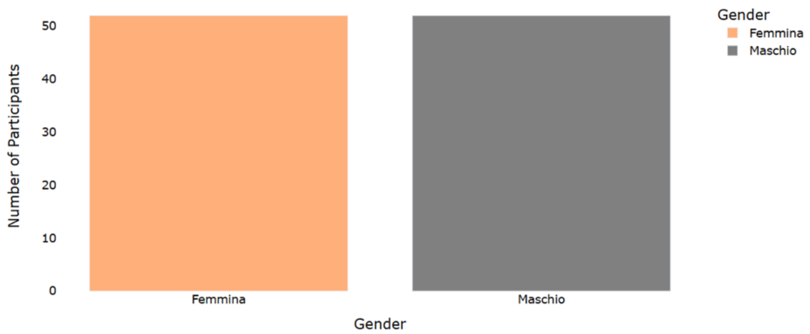

Figure 16 *% of Gender representation of the participants of the survey*

47% of the respondents belong to the age group 18-24 and 33% are between 25 and 34. All the other age groups are minor. 37% of the respondents have finished a Master's degree, another

37% have finished a Bachelor's degree and 19% of the respondents have finished High School. With regard the area of study or work, Business & Management represents the most popular choice with 42%, followed by Architecture and Engineering (16%) and Science & Tech (14%)

By comparing the answers to the multiple-choice questions for both types of visualization, key insights emerged in support to the effectiveness of feminist data visualization. By computing the mean scores (Table 1) for each of the multiple-choice questions about clarity, memorability, neutrality, curiosity, and emotional engagement, Visualization 2, the feminist visualization, consistently outperforms Visualization 1, the more traditional visualization, in all measured aspects. This suggests that the design or data presentation methods used in the feminist visualization might be more effective in engaging the audience and communicating the message.

Attribute	**Visualization 1**	**Visualization 2**
Clarity	4.0096	4.3846
Memorability	3.3942	4.2692
Objectivity & Neutrality	3.7885	4.0096
Curiosity	3.3942	4.0769
Emotional Engagement	2.5192	3.5962

Table 1 *Mean scores for Visualization 1 and Visualization 2 for the 5 attributes under study*

However, since the sample is not large enough to make conclusions only on averages, it is fundamental to analyze whether the differences between the scores of Visualization 1 and the scores of Visualization 2 are statistically significant. The statistical test was performed using Wilcoxon signed rank, a non-parametric test used for data that may not follow a normal distribution. Moreover, it is also used for categorical and ordinal variables as in this case. The following *p*-value for each question were calculated:

Attribute	***p*-value**
Clarity	0.0013
Memorability	0
Objectivity & Neutrality	0.0521
Curiosity	0
Emotional Engagement	0

Table 2: *Assessing statistical significance for the differences in the mean score between the two visualizations. If p-value < 0.05 then the difference is statistically significant.*

The first and second visualization differ significantly in how clearly, they communicate their message, how memorable they are, how much curiosity they generate, and their emotional impact showing a significant difference with a *p*-value below 0.05. The difference in their perceived objectivity and neutrality is not statistically significant since the difference between the two mean scores is very slight.

Moreover, additional analysis was carried out with other variables. The gender variable was used to compute the mean score across gender groups in order to understand whether there was a different pattern in perception between males and females (Table 3). The findings highlight a difference voting pattern in assigning the scores for visualization 2 with respect to the curiosity attribute and emotional engagement. The difference in voting for these two aspects is statistically significant as the *p*-value computed is below the threshold level of 0.05.

Attribute	**Visualization 1 *p*-value**	**Visualization 2 *p*-value**
Clarity	0.1050	0.4842
Memorability	0.5726	0.4564
Objectivity & Neutrality	0.0786	0.8113
Curiosity	0.6635	**0.0001**
Emotional Engagement	0.9571	**0.0013**

***Table 3** Assessing statical significance for the difference in the voting pattern between the female group and the male group for each Visualization.*

Moreover, by analyzing the voting pattern with the age variable, the findings suggest that age does not significantly influence how participants perceived the visualizations with respect the attributes of clarity, memorability, objectivity, curiosity or emotional impact. This could imply that the visualizations were perceived equally across all age groups. All attributes have high *p*-values, which indicates that there are no statistically significant differences in the perceptions of the two visualizations across the different age groups for most attributes. Therefore, Visualization 2, which adopts feminist principles, is perceived to be clearer, more memorable and more likely to generate curiosity and elevate emotions. This is in line with the belief of Data feminism that feminist data visualization facilitates viewer engagement and therefore understanding, making the visualization memorable and inspiring action. Moreover, female participants provided a higher score with respect their male counterparts with regard the capacity of Visualization 2 to spark curiosity and elevate emotion, implying that they may be more impacted by the visualization design and topic.

With respect the overall preference, the answers demonstrate a clear overall preference for the feminist data visualization (Visualization 2) in most categories, especially in being more memorable, highlighting key information better and being more emotionally engaging (Table 4).

Attribute	**Visualization 1**	**Visualization 2**
Clarity	33	71
Memorability	24	80
Key info	30	74
Objectivity & Neutrality	51	53
Emotional Engagement	10	94
Overall Preference	24	80

Table 4 *Number of people that expressed their preference for either Visualization 1 or Visualization 2 with respect to each attribute.*

The statistical analysis shows how this difference in voting between the two visualizations is statistically significant as all the *p*-values are below the threshold level of 0.05. Moreover, by examining the differences in the voting pattern between female and males, it is interesting to notice how they differ especially in the first question about which visualization is the clearest (Table 5). 80.77% of the female respondents believe that Visualization 2 is clearer in contrast with only 55.77% of males. This difference in voting is statistically significant since the *p*-value is 0.0115, below the threshold level. This difference highlights how males can have a different perception of the two visualizations compared to females, even though for all the other attributes the distribution is roughly the same. Still, the percentage of female participants supporting visualization 2 is always higher for all attributes.

Attribute	**Gender Group**	**Visualization 1**	**Visualization 2**
Clarity	Females	**19.23%**	**44.23%**
	Males	**80.77%**	**55.77%**
Memorability	Females	17.31%	28.85%
	Males	82.69%	71.15%
Key Info	Females	19.23%	38.46%
	Males	80.77%	61.54%
Objectivity & Neutrality	Females	48.08%	50%
	Males	51.92%	50%
Emotional Engagement	Females	9.62%	9.62%
	Males	90.38%	90.38%
Overall	Females	11.54%	34.62%

Preference	Males	88.46%	65.38%

Table 5 *% of people divided by gender group that expressed their preference for either Visualziation1 or Visualization 2 with respect to each attribute,*

The analysis of the multiple-choice questions for part 1 shows how on average people tend to prefer Visualization 2 with respect to all the attributes given, with only an exception about the attribute of neutrality and objectivity. For this attribute, the analysis does not provide statistical significance to the difference score between the two visualizations meaning that the respondents do not have a preference.

These quantitative data are then enriched with qualitative insights extracted by the participants' answer to the open-ended questions. Through the open-ended questions, it is possible to understand the reasons behind the participants' answer and draw more meaningful conclusions. Qualitative data helps to understand the why behind the numbers.

The first open-ended question aims to understand what are the key information that the participants captured. The goal is to compare the answers of both visualizations to understand whether each visualization design has influenced differently people's interpretation of the data presented, which is the same for both. By analyzing these answers, a significant difference emerged in participants' answers. While for the first visualization, most of the participants (78%) highlighted the increasing trend in female representation with only 13% of the participants remarking the slight decrease in representation in the last legislation, for the second visualization most of the participants' answer have a different connotation. In fact, while they still highlight how the female representation has increased throughout the legislations, they also remarked how much disparity there is still between male and female and how there is still a long road to achieve gender equality. This suggest that the interpretation has been influenced partially by the initial statement of Visualization 2 which states that there is still a long way to go to achieve gender parity in the Italian Parliament, demonstrating how the way a visualization is designed affects the interpretation of the data presented. This is one feature of feminist data visualization which is based on the belief that a visualization should highlight the findings of the research by inserting claims and elements that elevate emotion, making the visualization more memorable. Data feminism advocates for this approach, arguing that striving for neutrality in data visualization often leads to a diluted perspective that fails to convey the full implications of the data. In contrast, by explicitly acknowledging that "there is still a long way to go," Visualization 2 provides a more comprehensive and impactful narrative. This message is absent in Visualization 1, whose title merely describes the data without providing the broader context of gender disparity, limiting its interpretative depth.

The second open-ended question about the main graphical elements that were noticed reveals that the color pink and use of bar charts were the key elements that captured the viewer’s attention.

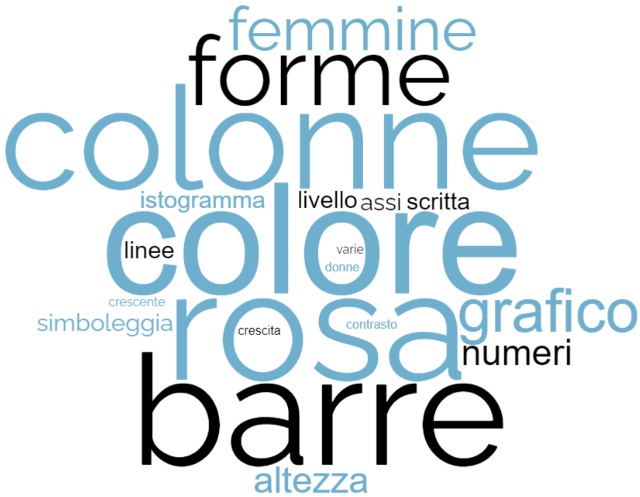

Figure 17 *Word Cloud highlighting the words most used by participants in answering the second open-ended question for Visualization 1*

While for the second visualization, participants noticed mainly the parliamentary chart with points representing each single deputies as well as the use of contrasting colors (grey vs orange) that make female deputies emerge among all the data points.

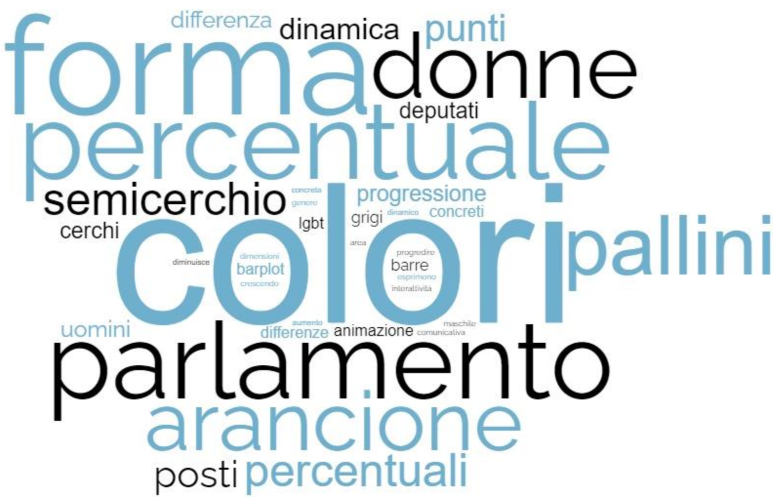

Figure 18 *Word Cloud highlighting the words most used by participants in answering the second open-ended question for Visualization 2*

With respect to this question, participants on average wrote a longer answer for Visualization 2 compared to Visualization 1. The median length for the traditional data visualization is 14 characters while for the feminist visualization is 34.5 characters.

All these insights are key elements that add to the third open ended question asking participants to describe the key differences between the two visualizations. This question is key to understanding the reasons behind the user's preference toward Visualization 2. In fact, most of the participants state that while the first visualization is very clear and enables them to capture the trend across legislations with one glance, the second visualization is more nuanced and engaging.

"Both graphical representations are effective. The first allows a direct comparison and detailed observation of data relating to different periods, proving particularly incisive in the communication of its message. The second, however, evokes greater emotions and tends to remain more imprinted in the memory"

The qualitative data suggests a general preference for Visualization 2, which is often described as more impactful, both emotionally and in terms of data communication. This visualization seems to have utilized comparative data more effectively, making the problem more tangible and highlighting the still huge disparity between male and female representation. The use of the parliamentary chart, as highlighted in the second open question for graphical elements, makes the data more concrete.

"Visualization 2 is more impactful as it involves comparison between groups, which adds value"

"The second visualization had a significant emotional impact on me, the presentation of the data in direct comparison evoked an emotional response that solved the problem much more tangible."

"The second one stays in your head better because it is easier to read and more figurative"

Moreover, when highlighting the key differences and when commenting the graphical elements of Visualization 2, some of the participants noted that the second visualization was giving attention to the missing data about trans and non-binary people, a feature that was not present in the first dataset. These responses highlight a critical observation about Visualization 2's lack of inclusive data representation, emphasizing the need for more comprehensive data that includes all gender identities to ensure a complete and accurate analysis.

There are no people of "undefined" gender present in parliament, as if their thoughts are worth less than "man" or "woman"

The second part of the survey reveals a preference for the feminist visualization when discussing emotional engagement. 77% of the participants expressed their preference toward the feminist visualization when looking at the level of emotions and engagement. In fact, while for the traditional visualization, 41 of the respondents answered “No emotion” for the open-ended questions asking the type of emotions that the visualization has evoked, for the feminist visualization, the number of people with no emotional involvement decreased to 21.

The insight of the survey reveals the power of feminist data visualization in facilitating data understanding and memorability. By leveraging emotions, the viewers are more connected with the data, making the visualization more memorable. In fact, the analysis of the survey showed how the feminist visualization outperformed Visualization 1 in clarity, memorability, and emotional engagement with statistically significant *p*-values. This emphasizes the effectiveness of the adoption of feminist data principles in data visualization. The connection with the viewer has been further advanced by the use of dots to represent individual deputies which were used to humanize data. In fact, participants felt that the data were more concrete, fostering greater empathy toward this subject and making the data more relatable. This emotional connection characterizing feminist data visualization is highlighted by the response of participants who reported a stronger emotional response to Visualization 2. In fact, participants felt more involved by looking at the dynamic Visualization 2 which illustrates the slow progression and sometimes stagnation in gender representation, term by term. This approach not only engages viewers more deeply but also prompts a call to action, since one of the key principle of data feminist is to inspire change. Moreover, by including trans and non-binary individuals in the visualization and by acknowledging the missing data about them, Visualization 2 aims to challenge the traditional binary classification and prompts participants to reflect on the reasons why these categories are often overlooked when collecting data. This type of reflection is absent in Visualization 1 which has the mere goal of presenting data without the aim to challenge traditional methods of classification and to inspire change. This approach, as revealed by the survey, not only broadens the viewer's perspective but also educates them about the limitations of traditional data categorization. This call to action for a more inclusive data collection aligns with the principles “Rethink Binaries” and “Challenge Power” of Data Feminism.

Additionally, Visualization 2 contextualizes the data by providing an historical overview throughout the terms of gender representation and ending with the present status. In this way, participants better capture the progress made as well as the drawbacks. Moreover, by contrasting the orange points representing female deputies with the grey points representing male deputies, participants have a better perspective of the proportion of gender representation as highlighted in the comments of the survey for Visualization 2. In fact, while in Visualization 1, participants only perceive the increasing trend in female participation, in Visualization 2, this increase is complemented with the awareness that the progress made is not yet enough and there is still a lot of work to do.

In conclusion, Visualization 2 is not only used as a tool for data presentation, but it acts as a tool for storytelling, education and advocacy through the application of feminist data principles. By leveraging on the emotions, the visualization becomes more engaging and memorable. Moreover, it demonstrates how data visualization can influence perceptions, provoke thoughts and inspire change.

5 CONCLUSION

The objective of this research was to assess the impact of integrating feminist data principles into data analysis and data visualization to uncover and better understand gender disparities within the Italian Parliament. This research was based on the principles laid out by the framework of Data Feminism which aims to challenge the existing traditional data practices. This framework does so by highlighting the need to consider the existing power structures operating in the filed of data science and challenge them by relying on an intersectional approach, elevating emotions and illuminating the overlooked experiences of marginalized groups. With the adoption of the framework of Data Feminism in the context of women's political representation, the research aimed to contribute to fairer data science practices by proving how feminist data principles can uncover deeper insights into gender representation and participation and how feminist data visualization can inform, educate and drive change. Every section of the research was conducted keeping in mind and recognizing the limits of the datasets available.

The first part of the study, the analysis of data about gender representation in the Italian Parliament, shows not only the progress made and the challenges that still persist in achieving gender parity in terms of numerical representation, but it also highlights the unseen contributions of female deputies in parliamentary activities. These analyses demonstrate the impact of applying feminist approaches to uncover hidden insights in political data analysis. By applying feminist data principles, in particular the principle "Challenge Power", the study moved beyond simple numerical counts of female versus male deputies to assess the quantity and quality of their engagement in parliamentary activities. This feminist approach highlighted not only the fluctuations in female representation across legislative terms but also the significant contributions of female deputies in parliamentary activities and law proposals, particularly concerning gender equality and other social issues. These findings challenge the traditional numerical analyses of political representation and emphasize the importance of using engagement metrics to analyze the real contribution of individuals. This approach allows data scientists to gather more in depth and detailed insights about complex social phenomena, helping in uncovering the experiences of marginalized groups that are often overlooked with traditional data analysis approaches.

The research also investigated how different visualization techniques may differently affect viewer understanding and engagement through a survey in which a more traditional visualization is compared with a feminist visualization. The responses showed how the feminist visualization significantly outperformed the traditional visualization across different areas including clarity, memorability, curiosity and emotional engagement suggesting that the principles of Data Feminism enable a more effectiveness communication of data through emotional engagement with the content. In fact, one characteristic of feminist data visualization is that it focuses on storytelling and on placing the data into the appropriate context by applying the principle of “Elevating Emotion” and “Consider Context”. This approach allows to engage more deeply the users since feminist data visualization has not only the purpose of presenting data but also to engage and inspire action among the audience. By gathering feedback on the effectiveness of feminist data visualization on data understanding and memorability, the research provides evidence to consider more carefully how data is presented to the public. In fact, the greater preference for feminist visualization for the dimensions of emotional engagement, data understanding and engagement suggests that the principles of Data Feminism may be applied in data visualization to lead to more effective advocacy and education campaigns. This visualization technique may better inform the public on social and political themes and inspire change. Lastly, both the data analysis and the data visualization phase shed light on the limitation of the datasets regarding the classification of gender that still does not take into consideration non-binary and trans besides the traditional male and female binary classification. By putting emphasis on this lack of data, the research applies the principle “Rethink Binaries and Hierarchies”.

The reliance on the framework of Data Feminism also implies the recognition of the limitations that characterize the research. This is meant to allow the reader to contextualize in the right manner the findings and their applicability. The first inherent limitation of the study is that the research is focused on a single case study regarding the theme of gender representation in the Italian Parliament. These findings regarding the dynamics of gender representation and the effectiveness of feminist principles in data analysis, are therefore linked to the Italian political and cultural context and may not be directly transferable to other countries without considering cultural and political differences. Moreover, the analysis was limited to the data available from the Italian Chamber of Deputies and only specific terms were used to analyze oral discussions and voting pattern, not capturing the trend across legislatures. The binary gender classification and the absences of nationality information limited the application of intersectionality in the analysis. Regarding the assessment of the impact of feminist data visualization through survey, limitations are linked to the limited number of participants whose responses may not be representative of the entire Italian population, especially regarding the age, since most of the respondents are between the age of 18 and 34. Moreover, the study analyzed the effectiveness of feminist data visualizations based on immediate participant's perceptions. Such responses may not fully capture information about long-term understanding and retention of key insights.

Keeping in mind this limitations, future works based on Data Feminism may expand on the findings highlighted in this research by applying feminist data principles across different political systems and cultural contexts to understand how different cultures respond to feminist methodologies in data science. The differences in the response to the application of feminist data principles should also be assessed based on the gender. In the research, female respondents seemed to be more deeply engaged when looking at the feminist visualization with respect to male respondents. Future research should investigate better in how different genders perceive and react to feminist data visualizations. This could involve more psychological studies that investigate the emotional responses to different techniques of data visualization. Lastly, it would be important to conduct a study that focuses on assessing the long-term impact of feminist data visualization against traditional data visualization to assess the audience retention of the data showed. These types of studies would be fundamental to uncover the memorability and effectiveness of different visualization techniques over time and provide deeper insights into how data visualization can affect policy making and public opinions.

This research shows how the framework of Data feminism is not just theoretical but is characterized by actionable principles that can be applied in concrete case studies to promote fairness and inclusivity in data practices. The research proves how feminist data principles, applied in data analysis and data visualization, can uncover hidden insights, highlight the contributions of marginalized groups as well as focus on the emotions to enhance understanding and memorability of the data presented.

The research highlights the positive impact of feminist principles in data analysis and visualization therefore aiming to contribute to a greater application of these principles in data science and a greater recognition of the biases that still pervade this field.

To conclude, this research supports the idea that feminist data analysis and visualization can be used as a tool to advocate for greater inclusion and greater visibility to specific social issues, inspiring change of the current power structures through more inclusive data practices.

BIBLIOGRAPHY

Amare, N., & Manning, A. (2016). A Unified Theory of Information Design. In Routledge eBooks. Informa. https://doi.org/10.4324/978131523273

Berinato, S. (2016). *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations*. Harvard Business Review Press.

Börner, K., & Polley, D. E. (2014). *Visual Insights: A Practical Guide to Making Sense of Data*. MIT Press.

Cairo, A. (2012). *The Functional Art: An Introduction to Information Graphics and Visualization.* New Riders.

Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication.* New Riders.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press

Criado Perez, C. (2019). *Invisible Women: Data Bias in a World Designed for Men.* Abrams.

Davenport , T. H., & Patil, D. (2012). *Data scientist: The sexiest job of the 21st Century*. Harvard Business Review. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press. doi: 10.7551/mitpress/11805.001.0001 1, 2

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Few, S. (2009). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.

Fleischmann, K. R., & Wallace, W. A. (2005). *A Covenant with Transparency: Opening the Black Box of Models*. Communications of the ACM, 48(5), 93-97.

Haraway, D. (1985). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century. The International Handbook of Virtual Learning Environments, 117–158. https://doi.org/10.1007/978-1-4020-3803-7_4

Haraway, D. (1989). *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*. Feminist Studies, 14(3), 575-599. doi: 10.2307/3178066

Harding, S. (2004). "Rethinking Standpoint Epistemology: What is "Strong Objectivity?" 10.1093/oso/9780198751458.003.0016.

Harding, S. (1991). *Whose Science? Whose Knowledge? Thinking from Women's Lives.* Cornell University Press.

Hartsock, N.C.M. (1983). The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism. In: Harding, S., Hintikka, M.B. (eds) Discovering Reality. Synthese Library, vol 161. Springer, Dordrecht. https://doi.org/10.1007/0-306-48017-4_15

Kennedy, H., & Hill, R. L. (2016). The Pleasure and Pain of Visualizing Data in Times of Data Power. Television & New Media, 18(8), 769–782. 10.1177/1527476416667823

Landstrom. C. (2007). Queering feminist technology studies. ¨ Feminist Theory, 8(1):7–26, doi: 10.1177/1464700107074193

Lupi, G., & Posavec, S. (2016). *Dear Data.* Princeton Architectural Press.

Montañez, A. (2017). *Visualizing Sex as a Spectrum*. Scientific American. https://blogs.scientificamerican.com/sa-visual/visualizing-sex -as-a-spectrum/

Morris, S., Adolphe, J., & Salam, E. (2019) "Does the New Congress Reflect You?," Guardian, https://www.theguardian.com/us-news/ ng-interactive/2018/nov/15/new-congress-us-house-of-representatives-senate.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown.

Schiebinger, L. (2008). *Gendered Innovations in Science and Engineering*. Stanford University Press.

Seager, J. (2018). The Women's Atlas, 5th ed. Penguin Books, 2018

Segel, E., & Heer, J. (2010). *"Narrative Visualization: Telling Stories with Data."* IEEE Transactions on Visualization and Computer Graphics, 16(6), 1139-1148. 10.1109/TVCG.2010.179

Shneiderman, B. (1983). "*Direct Manipulation: A Step Beyond Programming Languages."* IEEE Computer, 16(8), 57-69.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information, 2nd Edition*. Graphics Press.

Vicario, Grazia & Coleman, Shirley. (2019). A review of data science in business and industry and a future view. Applied Stochastic Models in Business and Industry.36. 10.1002/asmb.2488

United Nations. (2015). Goal 5 | Department of Economic and Social Affairs. Sdgs.un.org; United Nations. https://sdgs.un.org/goals/goal5#targets_and_indicators

Ward, M. O., Grinstein, G., & Keim, D. (2015). *Interactive Data Visualization: Foundations, Techniques, and Applications,* 2nd Edition. A K Peters/CRC Press.

Wong, D. M. (2010). *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figure*s. W. W. Norton & Company.

Yau, N. (2013). *Data Points: Visualization That Means Something*. Wiley.

APPENDIX

Appendix A

Link for Survey in Italian: https://forms.gle/c1bHeG5nPr2H8XH96

Link for Survey in English: https://forms.gle/bfq69pNtej651PKu9

Demographic Information

1.1. Age
1.2. Gender (*Please select the option that best describes you. We strive to be inclusive and recognize that these options might not fully encompass all identities.)*
1.3. Educational Background *(Please select the highest level of education you have completed.)*
1.4. Please indicate your area of study / work

PART 1

1. Visualization 1

- What is the key information from the visualization above?
  (Open-ended question)
- What graphic element(s) of the visualization (colors, writing, shapes, etc.) caught your attention the most?
  (Open-ended question)
- Visualization 1 is a visualization that conveys a clear message.
  (Likert Scale 1-5)
- Visualization 1 is a memorable visualization.
  (Likert Scale 1-5)
- Visualization 1 is an objective and neutral visualization.
  (Likert Scale 1-5)
- Visualization 1 is a visualization that arouses curiosity in me.
  (Likert Scale 1-5)
- I'm impacted emotionally by Visualization 1.
  (Likert Scale 1-5)

2. Visualization 2

- What is the key information from the visualization above?
  (Open-ended question)
- What graphic element(s) of the visualization (colors, writing, shapes, etc.) caught your attention the most?
  (Open-ended question)
- Visualization 1 is a visualization that conveys a clear message.

(Likert Scale 1-5)

- Visualization 1 is a memorable visualization.
  (Likert Scale 1-5)
- Visualization 1 is an objective and neutral visualization.
  (Likert Scale 1-5)
- Visualization 1 is a visualization that arouses curiosity in me.
  (Likert Scale 1-5)
- I'm impacted emotionally by Visualization 1
  (Likert Scale 1-5)

3. Visualization 1 against Visualization 2

- Having seen both visualizations, what are the main differences you notice? Please comment on their clarity, effectiveness in communicating the data, and any other observations."
  (Open-ended question)
- Which visualization do you believe presents the data more clearly?
  (Visualization 1 or Visualization 2)
- Which visualization is easier to remember and is more memorable?
  (Visualization 1 or Visualization 2)
- Considering how each visualization highlights key information, which one do you prefer?
  (Visualization 1 or Visualization 2)
- Which visualization do you consider to be more neutral and objective?
  (Visualization 1 or Visualization 2)
- Which visualization makes you feel more engaged or emotionally impacted by the data presented?
  (Visualization 1 or Visualization 2)
- Based on all aspects considered (clarity, ease of understanding, effectiveness in highlighting key information, and emotional impact), which visualization do you overall prefer?
  (Visualization 1 or Visualization 2)

PART 2 – FOCUS ON EMOTIONAL ASPECT

1. Visualization 1 (Traditional)
    - I'm impacted emotionally by this kind of visualization.
      (Likert Scale 1-5)
    - If you are impacted emotionally, what emotions did it convey to you? *(if you do not perceive any emotion, just write "NO")*
      (Open-ended question)
2. Visualization 2 (Feminist)
    - I'm impacted emotionally by this kind of visualization.

(Likert Scale 1-5)

- If you are impacted emotionally, what emotions did it convey to you? *(if you do not perceive any emotion, just write "NO")*
  (Open-ended question)

3. Visualization 1 vs Visualization 2
   - Which visualization emotionally impacts you the most?
     (Visualization 1 or Visualization 2)