

# Women In Art

Group Members: Téa Hassan, Nicholas Ngan

## Abstract / Summary

This project explores the underrepresentation of women in the art world and the mission of the Guerrilla Girls, a feminist art collective, to address gender and racial inequalities in the arts community. It discusses their use of culture jamming tactics to bring attention to disparities and discrimination within museums. It also highlights the Guerrilla Girls' findings regarding the low percentage of women artists exhibited compared to the prevalence of artworks featuring naked female bodies. Inspired by their work, the project outlines the aims of conducting a comprehensive analysis to examine the portrayal and representation of women in art history, the potential impact of societal changes on women artists' representation, and the broader underrepresentation of artists from various countries.

## I. Motivation

Women have been source of inspiration for artists throughout history. From the Venus de Milo to Andy Warhol's multi-coloured Marilyn, women (and especially naked women) have been plastered all over murals, exhibitions and museums. As two fellow art aficionados it struck us that all of our favourite pieces from all eras, were most, if not all, made by men. Why is it that women are at the centre of representation and yet not the ones behind the canvas? Why are there so little female artists? If art is supposed to depict an era, paint an accurate picture of the morals and ideals of that time, it appears as if that story has only been told by men, and thus leaves us with a biased recollection.



*Venus de Milo, Alexandros of Antioch, 150 to 125 BC*



*Girl with A Pearl Earring, Johannes Vermeer, 1665*

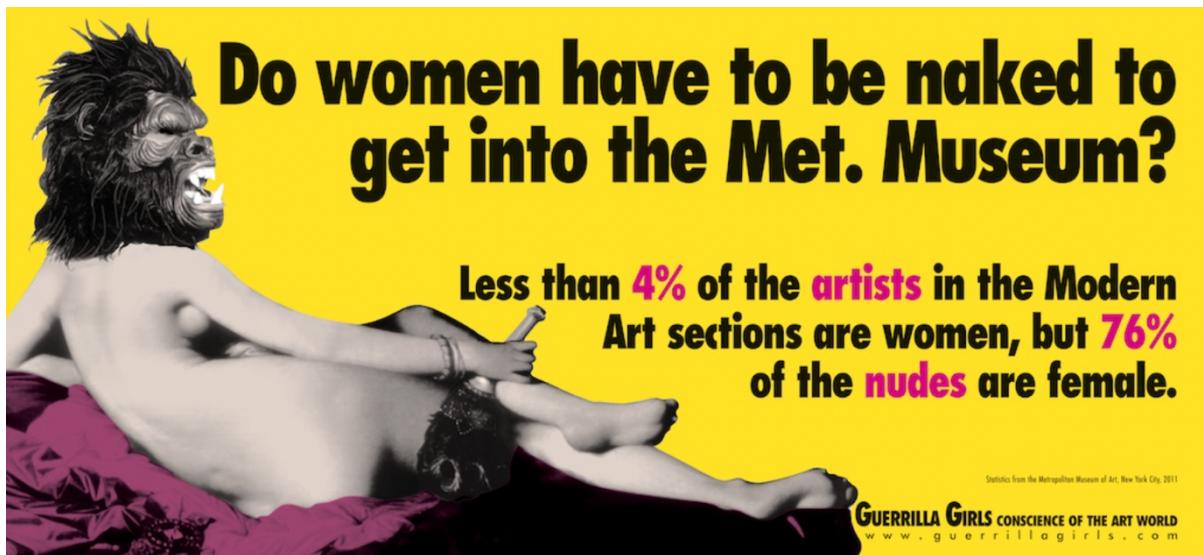


*Debbie Harry, Andy Warhol, 1980*

## The Guerrilla Girls

Our interest was sparked when visiting an exhibition by the Guerrilla Girls at the Saatchi Gallery back in March. The Guerrilla Girls are a group of feminist, female artists who formed in 1985 in New York City. Their mission is to combat sexism and racism within the art world. Through culture jamming tactics such as

posters, books, billboards, and public appearances, they bring attention to gender and racial inequalities in the arts community and society as a whole. The Guerrilla Girls use humor and anonymity, wearing gorilla masks and adopting pseudonyms, to shift the focus onto the issues they address rather than their individual identities or personal work. Their work exposes disparities, discrimination, corruption, and conflicts of interest within museums and aims to engage audiences through their thought-provoking messages.



Source: The Guerrilla Girls, 2012

When asked to create a billboard for the Public Art Fund, the Guerrilla Girls seized the opportunity to create something that would resonate with a wide audience. To gather data, they visited the Metropolitan Museum of Art and compared the number of women artists on display to the number of artworks featuring naked female bodies. The findings were eye-opening. However, the Public Art Fund rejected the billboard design, citing a lack of clarity. Undeterred, the Guerrilla Girls published their findings as an advertisement on New York City buses instead. Since then, they've done a few recounts and here are a few numbers:

1989	Less than 5% of artists in the Modern Art sections are women, but 85% of the nudes are female.
2005	Less than 3% of artists in the Modern Art sections are women, but 83% of the nudes are female.
2012	Less than 4% of artists in the Modern Art sections are women, but 76% of the nudes are female.

Interestingly, unlike what we would have thought, the number of female artists exposed at the Met Museum decreased over the years!

Here are a few more of their insightful projects:

# BUS COMPANIES ARE MORE ENLIGHTENED THAN NYC ART GALLERIES.



# Q. HOW MANY WORKS BY WOMEN ARTISTS WERE IN THE ANDY WARHOL AND TREMAINE AUCTIONS AT SOTHEBY'S?



\*The Contemporary Art Auction

Please send \$ and comments to:  
Box 1056 Cooper Sta. NY, NY 10276

**GUERRILLA GIRLS**

CONSCIENCE OF THE ART WORLD

With all these numbers under our eyes, we couldn't help but want to expand on the existing analysis. Indeed, while ground-breaking, The Guerrilla Girls' projects were often restricted to a single exhibition or museum. We believed that we had the right tools at hand to conduct a multi-level analysis, that did not limit itself to a country or museum.

## Initial Aims

- How has the portrayal of women changed throughout history?
- As women are given more independence in society, are women artists being more represented in museums?
- Why are there so little female artists? Or are they just underrepresented?

## II. Data Sources

### The Official MoMa Data Set

As we started looking for data, we came across the official MoMa Data set in the form of two CSV files, one with listing artists one listing artworks, published on GitHub.

The Museum of Modern Art (MoMA) is a highly influential institution, making it essential for understanding the role of women in art. As a renowned museum with a vast collection, studying MoMA allows us to gain insights into the representation and contributions of women artists throughout history. By examining its exhibition history, we can trace the evolving narrative of women's participation in the art world.

The Museum of Modern Art (MoMA) is a highly influential institution, making it essential for understanding the role of women in art. As a renowned museum with a vast collection, studying MoMA allows us to gain insights into the representation and contributions of women artists throughout history. By examining its exhibition history, we can trace the evolving narrative of women's participation in the art world.



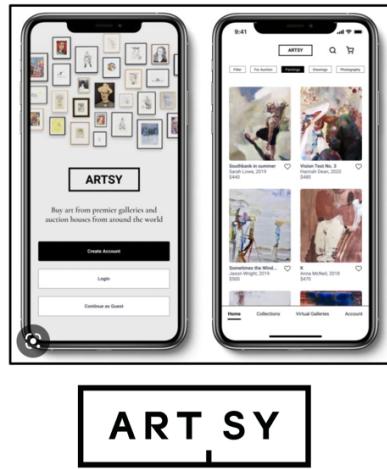
### *Dataset Characteristics*

- Data Volume: 138,185 artworks by 15,244 artists
- Dataset includes:
  - Basic metadata for each artist
  - Name
  - Nationality
  - Gender
  - Birth and death year
  - Wiki QID
  - Getty ULAN ID

### **The Artsy API**

To enhance our analysis on women in art and not limit ourselves to one museum, we incorporated data from the Artsy API. By leveraging the extensive resources provided by Artsy, including information on artworks, artists, exhibitions, and galleries, we were able to enrich our understanding of the representation and experiences of women in the art world. The integration of Artsy API data allowed us to gather comprehensive insights and broaden the scope of our project, providing a more holistic view of women's contributions and challenges within the art community.

Artsy is an online platform that provides an extensive array of art-related services, encompassing art commerce, education, and a robust database. Established in 2009, Artsy boasts an exceptional database featuring over 1 million artworks and 100,000 artists, making it a highly comprehensive and valuable online resource for art information. Its offerings contribute to a thriving ecosystem for art enthusiasts, collectors, and professionals seeking unparalleled access to the art world. The Artsy API is an interface that allows developers to access and retrieve art-related data from the Artsy platform. It provides us with a structured way to interact with the vast collection of artworks, artists, exhibitions, and galleries stored in the Artsy database.



### *Dataset Characteristics*

- Data Volume: over 1 million artworks by more than 100,000 artists
- Dataset includes:
  - Artist info: name, dates, nationality, bio
  - Artwork info: title, medium, date, dimensions, provenance and exhibition history
  - Gallery and museum info: name, location and exhibitions hosted
  - Auction data: sale date, auction house, sale price and lot number
  - Art market analysis: pricing trends, artist popularity and market outlook
  - Collection and curation data

## **III. Methodology**

### **1. Data Collection and Data Wrangling**

#### Data Collection MoMa Dataset

We collected the two CSV files from the Official MoMa GitHub Page:  
<https://github.com/MuseumofModernArt/collection>

To extract the relevant data, we wanted from the two CSV files we had to go through a few manipulations, to make the data more readable:

<p><b>Lowercasing the gender column using '<u>lower()</u>' method.</b></p>	<pre>gender = row[4].lower() if len(gender) &lt; 2:     gender = 'undefined'</pre>
<p><b>Handling missing or undefined values:</b> In case the gender column is empty or contains only one character, the value is set to 'undefined'</p> <p><b>Removing header row:</b> The header row is skipped during the loop through the CSV file</p>	
<p><b>Removing header row:</b> The header row is skipped during the loop through the CSV file</p>	<pre># Skip header row next(reader)</pre>
<p><b>Filtering data:</b> The data is filtered based on a condition to include only records where the value in column 6 is less than 1955.</p>	<pre>if int(row[6]) &lt; 1955:     if gender in pre_cutoff_counts:         pre_cutoff_counts[gender] += 1     else:         pre_cutoff_counts[gender] = 1</pre>

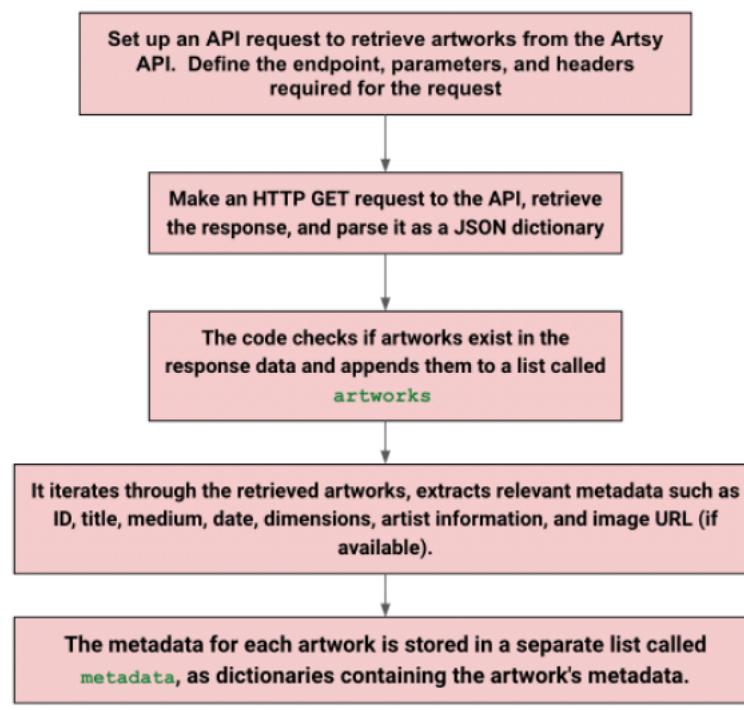
Finally, we merged the two files ‘artworks MoMa’ and ‘artists MoMa’ to have them into a single csv file called ‘moma’

Here is a sample of the data:

## Data Collection Artsy API

Working with the Artsy API was a more challenging endeavour. Between our final presentation and the project some time had passed, and the API had been completely updated, which meant the code we had so far did not work and we had to start all over.

The steps we followed are detailed in the chart below:



Here is a sample of the data:

ID	Title	Medium	Date	Width (cm)	Height (cm)	Depth (cm)	Diameter (cm)	Artist	URL	Name
44093844d4f81b2000000000	Caravaggio (The Kiss)	Oil and gold leaf on canvas	1567-1568	180.1	181.1			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000000	Caravaggio (The Kiss)	
44093844d4f81b2000000001	The Third of May	Oil on canvas	1804	230.2	345.2			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000001	Pierre Michel Alce after Jean-François Gérard	
44093844d4f81b2000000010	The Company of Frans Banning Coop and Willem van Heythuysen (The Night Watch)	Oil on canvas	1642	360.2	437.4			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000010	John Peter Paul Rubens after Frans Hals	
44093844d4f81b2000000012	Mona Lisa	Oil on poplar	ca. 1503-1506	77	55.1			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000012	Théodore Félix Eugène after Louis Rothschild	
44093844d4f81b2000000013	Portrait of the Painter Jan de Dijzeren a/c (Honthorst)	Oil on panel	1620	280.5	284.5			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000013	Gerard van Honthorst after Jan de Dijzeren	
44093844d4f81b2000000014	The Birth of Christ	Tempera on canvas	ca. 1480	127.2	278.6			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000014	Philipp Runggaldier	
44093844d4f81b2000000016	The Birth of Venus	Oil on canvas	1563	200.0	203.5			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000016	Hans Holbein, F. and Ulrich von Hutten after Jean-Jacques Henner	
44093844d4f81b2000000017	Portrait of Sir Matthew Paston	Oil on canvas	1499-1502	230.0	260.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000017	Wouter van den Bosch after Sir Matthew Paston	
44093844d4f81b2000000018	The Birth of the Virgin Mary	Oil on canvas	1499-1500	211.2	460.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000018	Philippe de Champaigne after Georges Ménageot	
44093844d4f81b2000000019	The Birth of the Virgin Mary	Oil on panel	1499-1500	211.2	460.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000019	Wouter van den Bosch after Georges Ménageot	
44093844d4f81b2000000020	The Last Supper	Tempera on panel, pitch and mastic	1495-1498	460.2	880.6			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000020	Gaspero Milesi	
44093844d4f81b2000000021	Vision of the Rocks	Wood, transferred to canvas in 1898 by Munch	ca. 1483-1486	198	122.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000021	Edvard Munch after Gustav Klimt	
44093844d4f81b2000000022	Death of Sardanapal	Oil on panel	1601	232.0	463.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000022	Antoine Caron after Titian	
44093844d4f81b2000000023	Le Libéreté guidant le Peuple ( Liberty Leading the People)	Oil on canvas	1830	360.1	305.4			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000023	Philippe Sauvanet	
44093844d4f81b2000000024	David	Bronze	1424-1432	158.2	158.2			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000024	Peter Paul Rubens	
44093844d4f81b2000000025	Self-Portrait with a Medallion	Oil on panel	1591-1594	489.5	719.5			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000025	Diego Velázquez	
44093844d4f81b2000000026	Portrait of Alessandro Farnese	Oil on panel	1598-1612	260.0	260.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000026	Andrea Mantegna after Andrea Mantegna	
44093844d4f81b2000000027	Le Titre à Dieu	Oil on canvas	1743	210.1	91.2			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000027	Joseph Barthélémy Kost after John-Jacques Henner	
44093844d4f81b2000000028	Le Grand Odalisque	Oil on canvas	1614	89.8	160.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000028	Gerard van Honthorst after Jean-Auguste-Dominique Ingres	
44093844d4f81b2000000029	Le Baiser (The Kiss)	Oil on canvas	1772	30.1	46.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000029	James Hobson after Francis Cotes	
44093844d4f81b2000000030	The Anatomy Lesson of Dr. Nicolaes Tulp	Oil on canvas	1632	216.7	198.7			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000030	Master of 1515	
44093844d4f81b2000000031	Dance at the Moulin de la Galette (Ball at the Moulin de la Galette)	Oil on canvas	1870	111.1	175.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000031	Isidor van Meckenem after Martin Schor	
44093844d4f81b2000000032	Allegory of the Four Elements	Oil on panel	1593	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000032	Workshop of Sava Henning	
44093844d4f81b2000000033	Le déambulation des Morts au port de Marseille le 2 Novembre 1600 (Martyr Medicis) arrives in Marseilles Nov. 2 1600	Oil on canvas	ca. 1602-1605	384.0	295.1			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000033	Quentin Perret after Jean-Baptiste Goujon	
44093844d4f81b2000000034	Self-Portrait with a Medallion	Oil on panel	ca. 1603	60.5	50.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000034	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000035	Portrait of a Frenchman	Oil on panel	ca. 1606	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000035	Nicolas Beaurain	
44093844d4f81b2000000036	L'Embarquement pour Cythère (The Embarkation for Cythera)	Oil on canvas	1717	129.1	194.1			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000036	James Fagan	
44093844d4f81b2000000037	A Buddha	Copper alloy with crystal and obsidian	1590	80.7	91.8	23.0		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000037	Giovanni Battista Troja	
44093844d4f81b2000000038	A Harmer Milkman of Sicily	Oil on cardboard	1617	90	90			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000038	Domenico Tristano	
44093844d4f81b2000000039	Portrait of a Frenchman, Father from a Monastier d'Asfeld (Portrait of Jean)	Oil on panel	1620	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000039	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000040	A Scène from the Life of St. Martin	Oil on panel	ca. 1737	38.7	26.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000040	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000041	Allegory of Navigation with a Cross-Staff Averell	Oil on panel	1537	245.4	116.8			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000041	Suzer C. Webster	
44093844d4f81b2000000042	Portrait of a Frenchman with an Armchair	Oil on panel	1597	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000042	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000043	Aux Antiochies	Etching	1679-1680	24.0	30.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000043	Lucas Verman	
44093844d4f81b2000000044	Styrian Orpheus	Stop-symmedys	200 BCE - CE 500	24.1	30.3	30.5		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000044	Jean-Baptiste Huet	
44093844d4f81b2000000045	Antennae	Carved wood with pigments	1470-1510	45.7	19.7	17.8		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000045	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000046	Antennae	Carved wood	1470-1510	13.7	13.7			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000046	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000047	Buddha Shakyamuni	Bronze	2nd-3rd century	100.0	40.0	49.0		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000047	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000048	Checkpoint from a Home Bit	Bronze	1000-1050 B.C.	15.2	21.6			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000048	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000049	Ornat in Limur	Oil on canvas	ca. 1653	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000049	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000050	Portrait of a Frenchman, Crown of Thorns	Portrait on panel	ca. 1658	73.7	52.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000050	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000051	Diana Resting	Oil on canvas	ca. 1685	40.8	33.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000051	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000052	Eve	Oil on panel	first model ca. 1681; this cast 1686	127.2	44.3	63.5		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000052	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000053	Bronze oil and varnished silver object with glass	Lead	late 16th century	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000053	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000054	Album leaf, ink, and color on silk	Ancient	24.1	20.1	24.1	20.3		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000054	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000055	Portrait with Landscape	Seaweed wave, wheel-thrown stoneware with blue-tinted glaze and overglaze enamel painted decoration	Lead Ming dynasty, 1520-1564	40.3	37.2	14.0		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000055	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000056	Grande Arquée	Porcelain	1600	30.0	30.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000056	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000057	Head of Buddha Shakyamuni	Terracotta	375-400	23.0	12.7	14.0		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000057	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000058	Henry IV Conferring the Regency upon Marie de' Medici (Medici before Rubens)	Oil on canvas	made before 1654	88.3	115.9			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000058	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000059	Je with Dragon and Clouds	Wheel-thrown porcelain with blue painted decoration under glaze	Johnson dynasty (1390-1510), 15th century	44.0	34.3			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000059	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000060	Landscape with Dunes	Oil on panel	ca. 1650	100.0	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000060	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000061	Manus (Dioris or the Falling Man) (2)	Bronze	first model ca. 1682; this cast 1670	102.4	73.7	45.7		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000061	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000062	Minotaur or Faun and Nymph	Painter	ca. 1688	34.3	27.9	22.2		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000062	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000063	Portrait of a Frenchman in the Desert	Oil on wood	ca. 1690	100.0	94.8	55.3		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000063	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000064	Movement to Honourable Nation	Bronze	first model 1687; this cast 1687	207.2	109.3	109.0		https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000064	Portrait of a Frenchman after Jean-Baptiste Goujon	
44093844d4f81b2000000065	Portrait of Peter I	Oil on canvas	ca. 1655-1658	88.2	99.1	88.1	11.1	https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000065	Portrait of a Frenchman after Peter I	
44093844d4f81b2000000066	Portrait of a Young Woman with a White Cloth	Oil and tempera on panel	ca. 1654	111.1	100.0			https://artsy.net/api/v1/artwork?hash=44093844d4f81b2000000066	Portrait of a Frenchman after Peter I	

## 2. Merging The Data

Once we were done extracting all the relevant data from the Artsy API we had to merge all our different csv files to create a final csv file that contained all the relevant information.

We proceeded in the following way:

- we merged the artists moma.csv and artworks moma.csv into a final moma.csv file
- we merged the artists artsy.csv and artworks artsy.csv into a final artsy.csv

Then, we have to reshape the datasets so that they had the same column names. We saved these new datasets as moma\_simple.csv and artsy\_simple.csv. Here is a sample:

Title	ID	Name	Nationality	Gender
Der Kuss (The Kiss)	4d8b92eb4eb68a1b2c000968			
The Third of May	4d8b92ee4eb68a1b2c0009ab	Pierre-Michel Alix after Jean-François Garnery after Louis Michel Van Loo		male
The Company of Frans Banning Cocq and Willem van Ruytenburch (The Night Watch)	4d8b93394eb68a1b2c0010fa	John Jones after Thomas Beach		male
Mona Lisa	4d8b937c4eb68a1b2c001722	Thérèse-Élönore Lingée after Louis Rolland Trinquesse		
Luncheon on the Grass (Le Déjeuner sur l'herbe)	4d8b93b304eb68a1b2c001b9d	Hans Baldung	German	male
The Birth of Venus	4ea0ec3d2a58200001006e93	Géraud Vidar after P. Davesne		male
The Birth of Venus	4ea0eedd66899c800001006eac	Hans Weiditz, II and Ulrich von Hutten (author)		male
The Calling of St Matthew	4eaef38fec5cf6000100a3f9	Johann Christoph Dietzsch		male
The Burial At Ornans	4eaefaf6c76e78f0001009d40	Willem van der Leuuw after Sir Peter Paul Rubens		male
The Last Supper	4eaefc4976e78f0001009e66	Gasparo Mola		male
Virgin of the Rocks	4eaefdd86899c8000010081e1	Israël van Meckenem after Master W with the Key		male
La mort de Sardanapale (Death of Sardanapalus)	4ea03a1476e78f000100badb	Nicòlo Ricciolini		male
La Liberté guidant le peuple (Liberty Leading the People)	4ea0fb24bddef7000100b6ef	Philippe Rumpf		male
David	4ea0fc0176e78f000100bcd0	Pieter van Sompel		male
The Raft of the Medusa	4ea151bd2a5820000100ba46	Christoph Nathe		male
Isenheim Altarpiece	4ea1f1910bddaf7000100d117	Ezra Ames		male
The Tête à Tête	4ea1fd2876e78f00010014f	Joseph Bartholomew Kidd after John James Audubon		male
La Grande Odalisque	4ea1fe4435118f000100d32a			
Olympia	4ea209c35118f000100dd6	Domenico Tintoretto	Italian	male
Le Berceau (The Cradle)	4eb02597e3c380000100121	James Watson after Francis Cotes		male
The Anatomy Lesson of Dr. Nicolaes Tulp	4eb063a5b197640001007110	Master of 1515		male
Dance at Le Moulin de la Galette (Bal du moulin de la galette)	4eb0654269a04000010096da	Israël van Meckenem after Martin Schongauer		male
The Gates of Hell (Detail: The Thinker)	4eb068115fb4150000100a26a	Robert Polhill Bevan		male
Le débarquement de Marie de Médicis au port de Marseille le 3 November 1600 (Maria Medici arrives in Marseille, Nov. 3 1600)	4eb06c8c5fb4150000100b92d	Workshop of Severo da Ravenna		male
Self-Portrait with Bandaged Ear	4eb077c25fb4150000100dd91	Quentin-Pierre Chedel after Jean-Baptiste Oudry		male

Finally, we merged these two data sets and saved the final dataset into a csv file **merged.csv**.

## Cleaning The Data Frames

To clean up the data, we created a code that defined a function called 'extract\_year' that extracts the year from a given date. It removes whitespace from the 'Date' column in the 'artsy' dataframe and converts it to numeric values. It also removes brackets from the 'Gender' and 'Nationality' columns in the 'moma' dataframe and applies the 'extract\_year' function to the 'Date' column to extract the year.

To ensure consistency between the two dataframes, the code renames columns in both 'artsy' and 'moma' to match each other. It then merges the two dataframes into a new dataframe called 'df'.

We performed further data cleaning by removing rows with missing values in the 'Date' column of the 'df' dataframe and converting the 'Date' column to integer values. The 'Gender' column is cleaned up using the 'decide\_gender' function, which converts gender strings to lowercase, splits them into words, and determines the dominant gender based on word counts. To provide additional information, the code defines a function called 'year\_to\_century' that converts the year to a century. It adds a new column 'Century' to the 'df' dataframe by applying the 'year\_to\_century' function to the 'Date' column. Rows with 'Century' values greater than 21 are removed from the 'df' dataframe.

Finally, the code summarizes the data by grouping the 'df' dataframe based on 'Century' and 'Gender' columns and calculates the count of each group. The summarized data is stored in a new dataframe called 'summary\_df', which includes columns 'Century', 'Gender', and 'Count'.

This is useful for preparing and merging art data from two different sources, allowing for comprehensive analysis and visualization. By cleaning and transforming the data, it ensures data quality and enables us to gain insights into the gender distribution and artistic trends across different centuries.

#### IV. Exploratory Data Analysis

Now equipped with our clean and tidy data frames, we can move onto the analysis of the data. We reiterate our initial goals:

- How has the portrayal of women changed throughout history?
- As women are given more independence in society, are women artists being more represented in museums?
- Why are there so little female artists? Or are they just underrepresented?

#### Summary Statistics

##### MoMa Dataset

```
Title      ID      Artist Nationality  Gender
count    140811  139633  139633  139634
unique   99087   13736   13697   1062   414
top     Untitled  229   Eugène Atget  (American) (Male)
freq     8525    5050   5050    58388  105646
```

These summary statistics for the MoMa Data provide a snapshot of a dataset containing information about art pieces. The dataset includes over 140,000 observations, with almost 100,000 unique titles and more than 13,000 unique artists. The most common title in the dataset is "Untitled," appearing over 8,500 times, while the artist "Eugène Atget" is the most frequent, with over 5,000 occurrences. The majority of the artists in the dataset are identified as American, and the gender distribution is heavily skewed towards male artists, with over 105,000 instances.

##### Artsy Dataset

```
Title      ID      Name Nationality  Gender
count    140811  139645  78159   46525  72287
unique   99886   17536   986     27     4
top     Untitled  229   French   male
freq     8522    5050   6082    19218  65806
```

These summary statistics for the Artsy Data provide an overview of a dataset containing information about art pieces. The dataset consists of approximately 140,000 observations. There are nearly 100,000 unique titles and over 17,000 unique IDs. The most common title in the dataset is "Untitled," appearing around 8,500 times. The majority of the artists in the dataset have French nationality, with a frequency of approximately 19,000. The dataset includes four unique gender categories, with "male" being the most common, occurring over 65,000 times. However, it's worth noting that the "Name" attribute has missing or unspecified top and frequency values, which implies some missing data or variability in the dataset.

#### Proportion of Men vs Women Artists

The first natural question that comes to mind is how many women artists there are compared to men artists. In order to provide an answer, we first counted how many male and female artists were in each dataset and then added them up to create a final number. We tried to directly analyse the **merged.csv** file but the file was too big and we encountered many crashes. It was easier for us to break down the process.

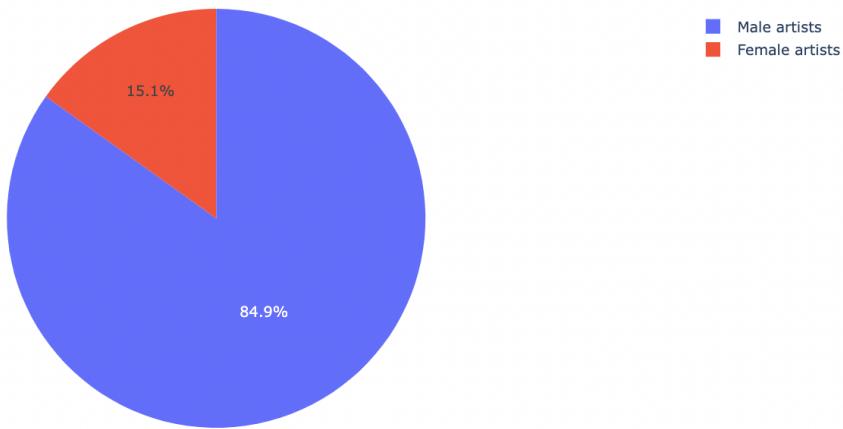
We first counted the number of male and female artist present in the MoMa dataset:

Total artworks: 140851

Female artists: 18830  
Male artists: 105689

Then, we plotted the results in a pie chart using plotly:

Gender Distribution of Artists from MoMa Data

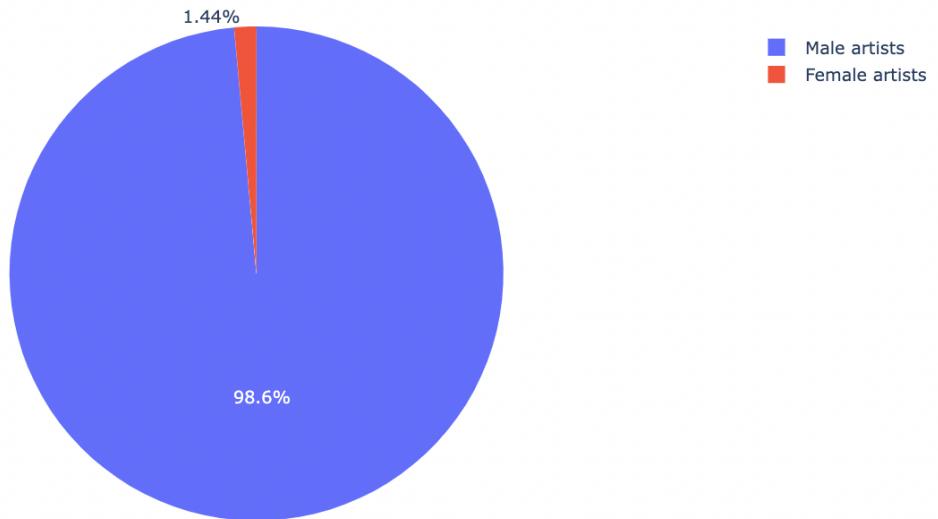


Unsurprisingly, the proportion of male artists compared to female artists in the MoMa Data set is much higher.

We then did the same for the Artsy Dataset:

Total artworks: 140851  
Female artists: 963  
Male artists: 65806

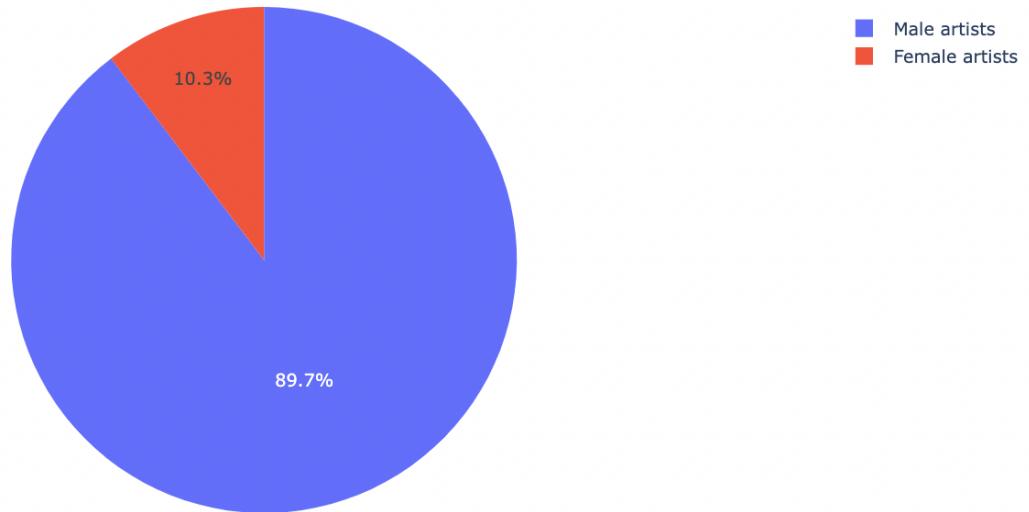
Gender Distribution of Artists from Arsty Data



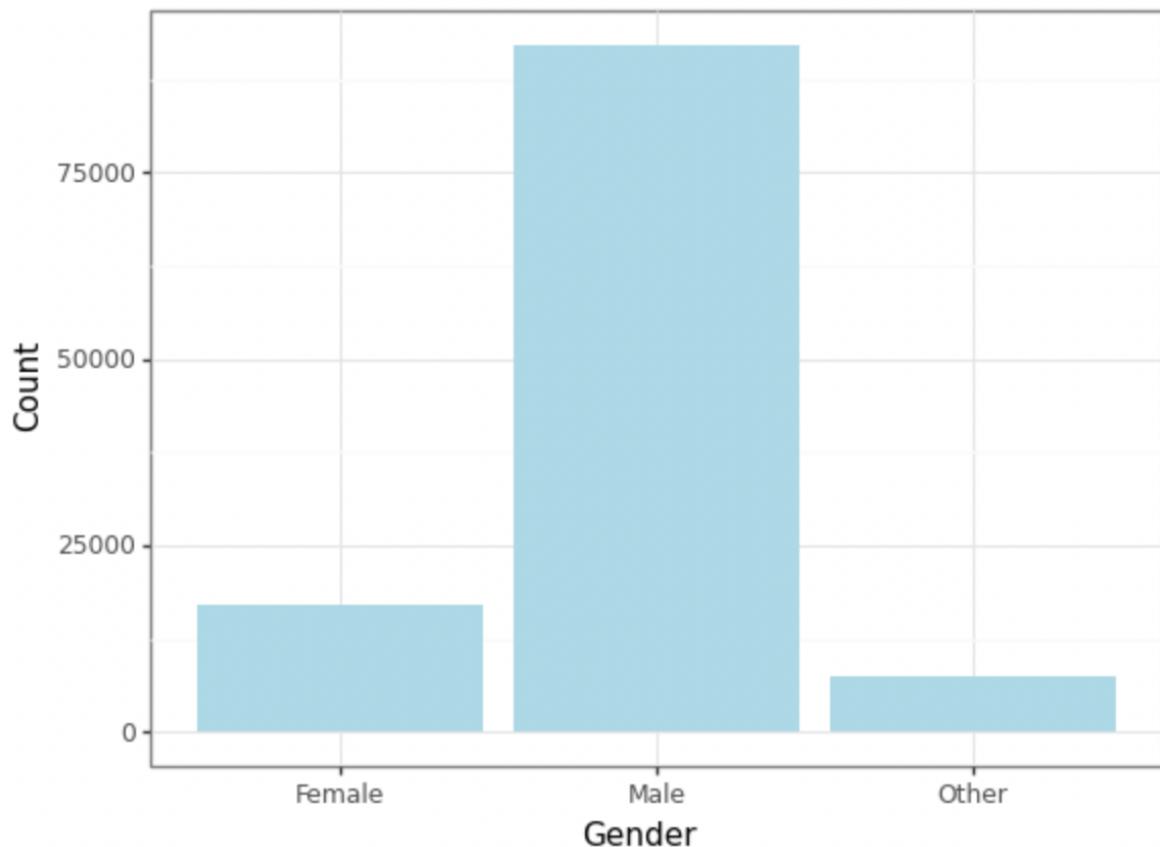
For the Artsy Data, the proportion of women artists is even smaller. This is probably due to the fact that the Artsy API extract data from all time periods, starting in around the 1500s, where there were relatively no female artists as women had a very limited role in society. Since MoMa exposes Modern and Contemporary Art, which starts around the 1950s, it makes sense that they expose a lot more female artists.

Finally, we combined the results to get a final count:

### Gender Distribution of Artists Total



### Count of Artworks by Gender

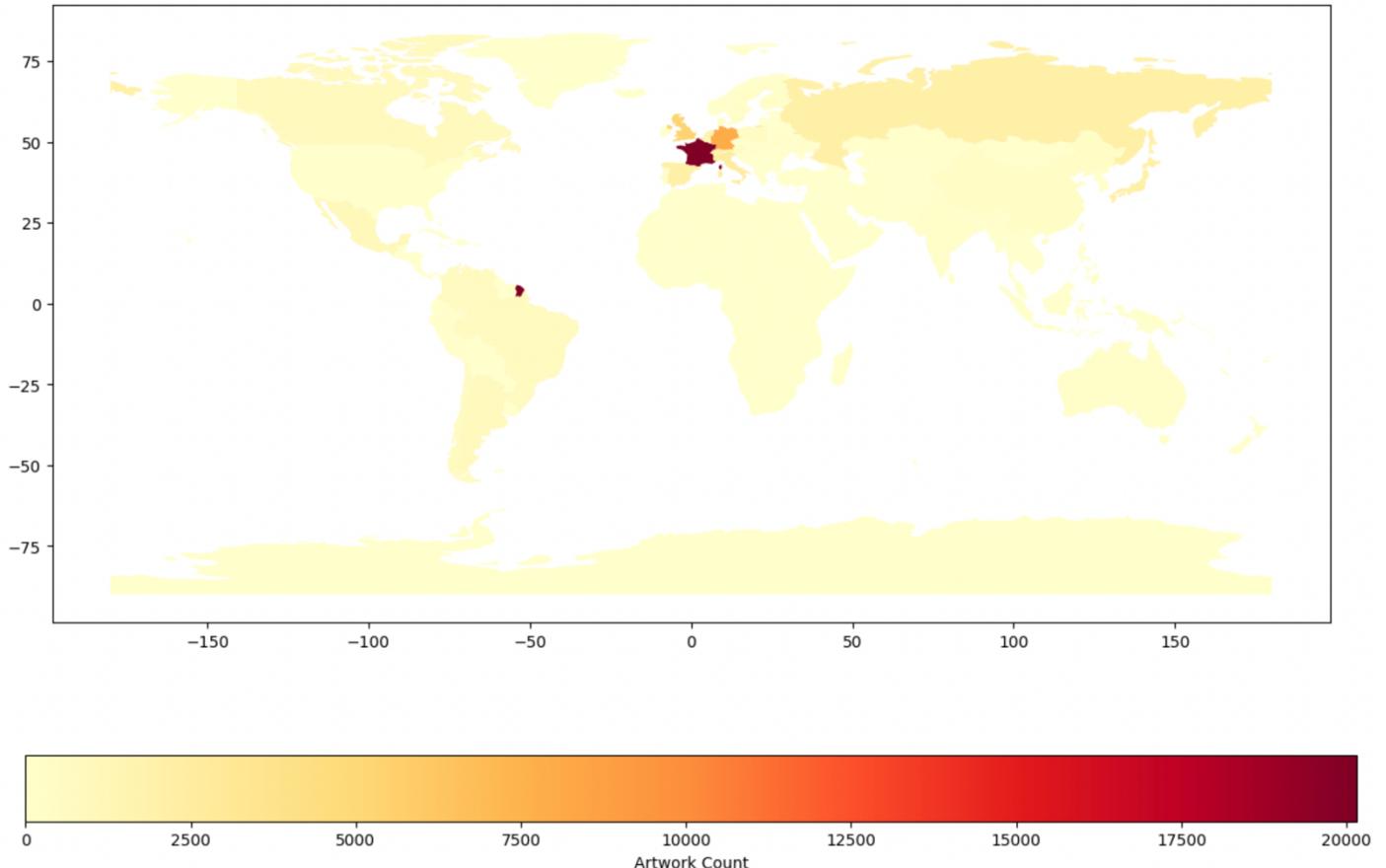


*note:* the other factor accounts for rows with missing data where the gender was not specified, it also encompasses gender neutral artists.

From these final plots, it is clear that male artists are heavily overrepresented. What is interesting is that from the MoMA data we understand that male artists are more exposed in museums, as the MoMA provides us with data about what is in the MoMA collection but also in private collections and small galleries as the Arsty Database provides us with data on artworks from all sorts of collections, from very prestigious museums and galleries to more local and niche ones. Through the use of these two datasets we clearly see that the underrepresentation of women in art is a global phenomenon that needs to be addressed.

## Country Analysis

We then performed an analysis to see if the phenomenon identified in the previous section was global or located in certain countries. We came up with the following results:



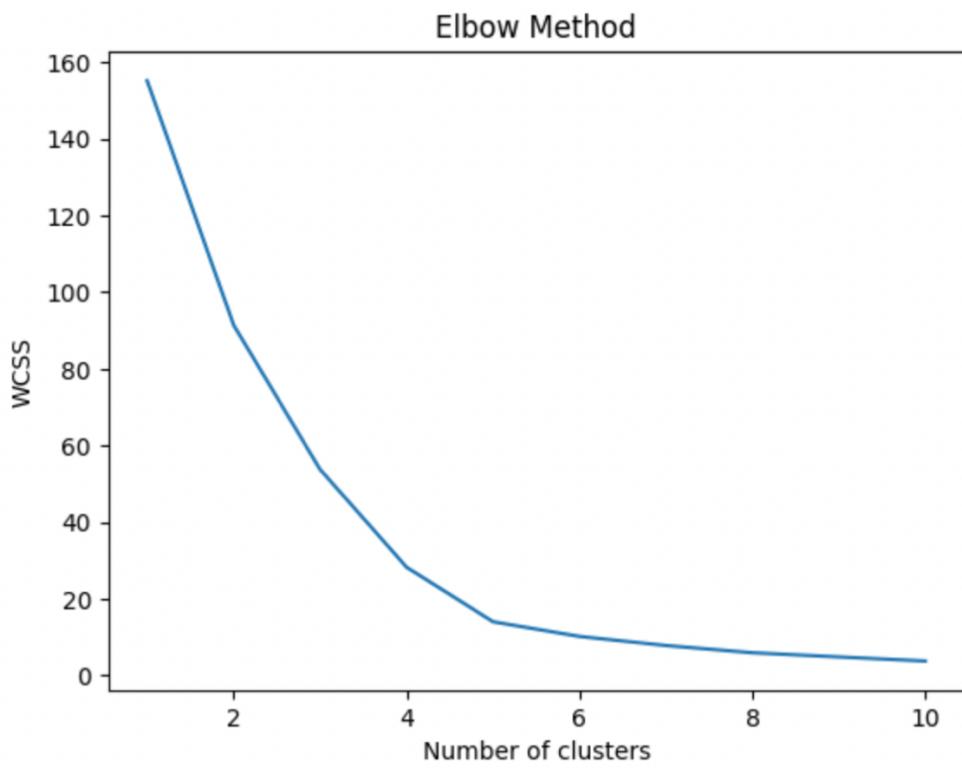
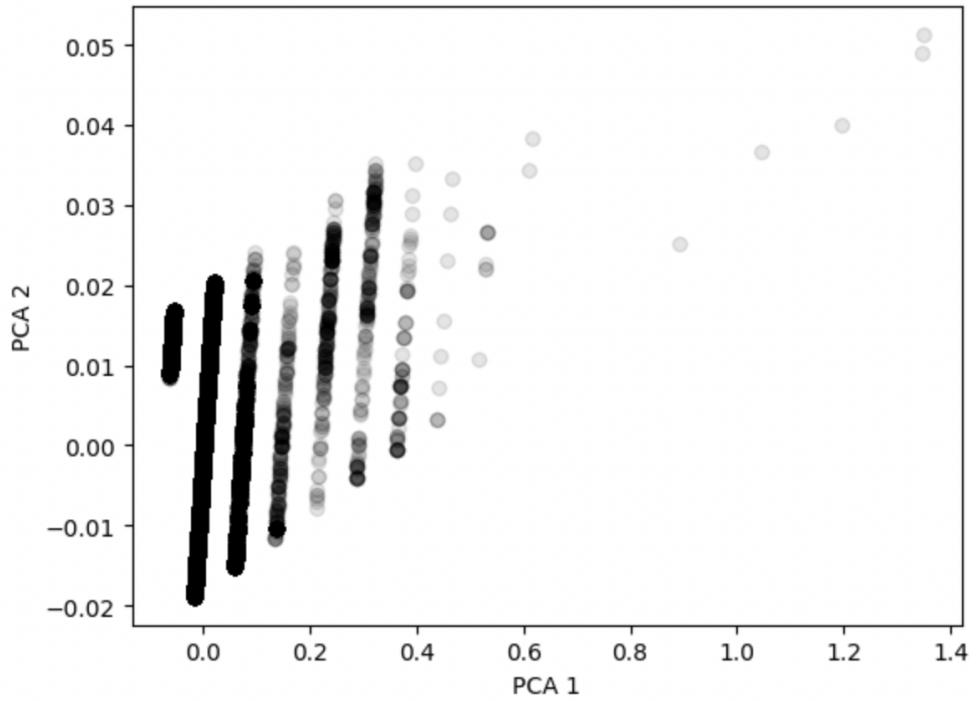
This plot shows a world map visualization of artwork counts by country intensity. Interestingly, it shows that the most artworks are generated by France and some other European countries. This aligns with our predictions that not only women are underrepresented but different nationalities as well.

## K Means Clustering Analysis

**Definition:** K-means clustering is a machine learning algorithm that groups similar data points together based on their feature similarities. It aims to partition the data into a predetermined number of clusters, where each data point belongs to the cluster with the nearest mean (centroid). The algorithm iteratively adjusts the cluster centroids until the clustering solution converges, resulting in distinct groups or clusters of data points.

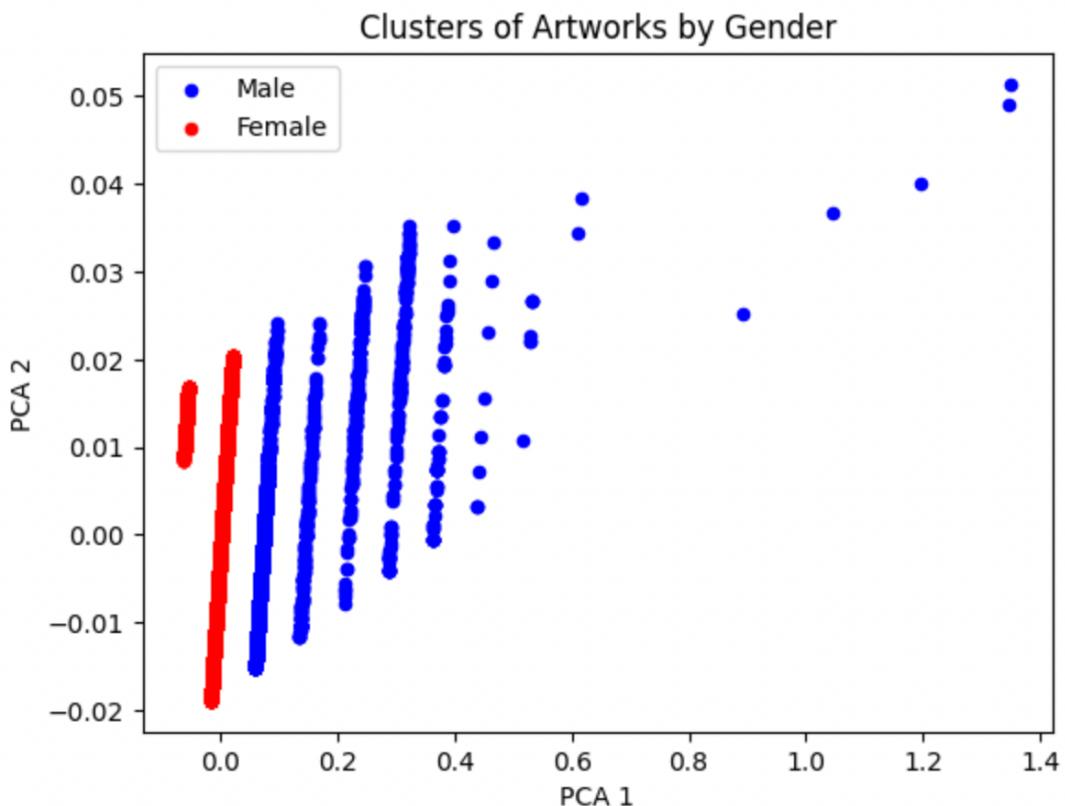
We performed a k-means clustering analysis to see what factors affect the gender of art. It selects numeric columns from the dataset, normalizes the data, applies principal component analysis (PCA) to reduce dimensionality, determines the optimal number of clusters using the elbow method, assigns data points to clusters based on the chosen number of clusters, and finally visualizes the clusters on a scatter plot. The analysis helps identify patterns and relationships between gender and the selected features in the dataset.

Creating the clusters and principal components of the analysis:



Based on the elbow method applied to the clustering results, it becomes evident that selecting 2 or 3 clusters is appropriate for distinguishing gender in the dataset. The elbow point in the plot indicates a significant drop in the within-cluster sum of squares (WCSS) as the number of clusters increases, implying that the most substantial gains in clustering accuracy are achieved with 2 or 3 clusters. This finding aligns with the expectation that the majority of the data points predominantly belong to the male or female clusters, while only a small percentage is categorized as "other," suggesting a clear gender differentiation in the dataset.

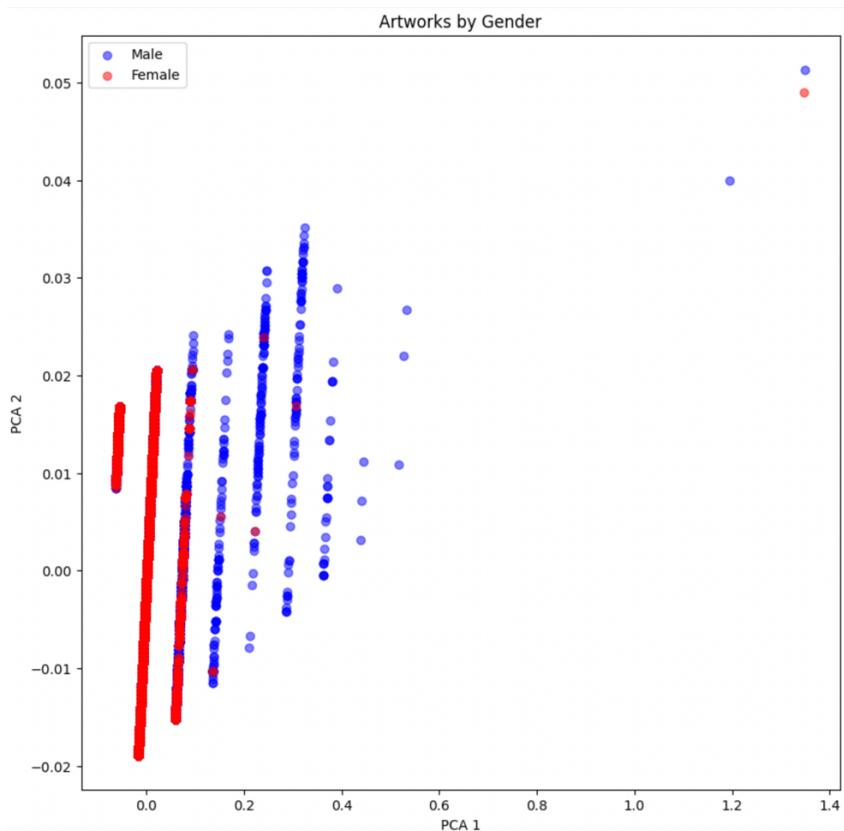
Results:



We generated a scatter plot to visualize the clusters obtained from the k-means clustering analysis. It plots the data points in the two-dimensional space defined by the principal components 'PC1' and 'PC2'. The code first plots the data points that belong to cluster 0, representing the 'Male' cluster, as blue dots. It then plots the data points belonging to cluster 1, representing the 'Female' cluster, as red dots. The 's' parameter determines the size of the dots, while the 'c' parameter assigns colors to the dots based on the cluster label. The plot is given a title of 'Clusters of Artworks by Gender' and labels the x-axis as 'PCA 1' and the y-axis as 'PCA 2'. The legend indicates the colors associated with each cluster label ('Male' and 'Female'). The plot provides a clear representation of how the data points are grouped based on the identified gender categories.

Unsurprisingly, we find a lot more data points associated with the male cluster (blue dots).

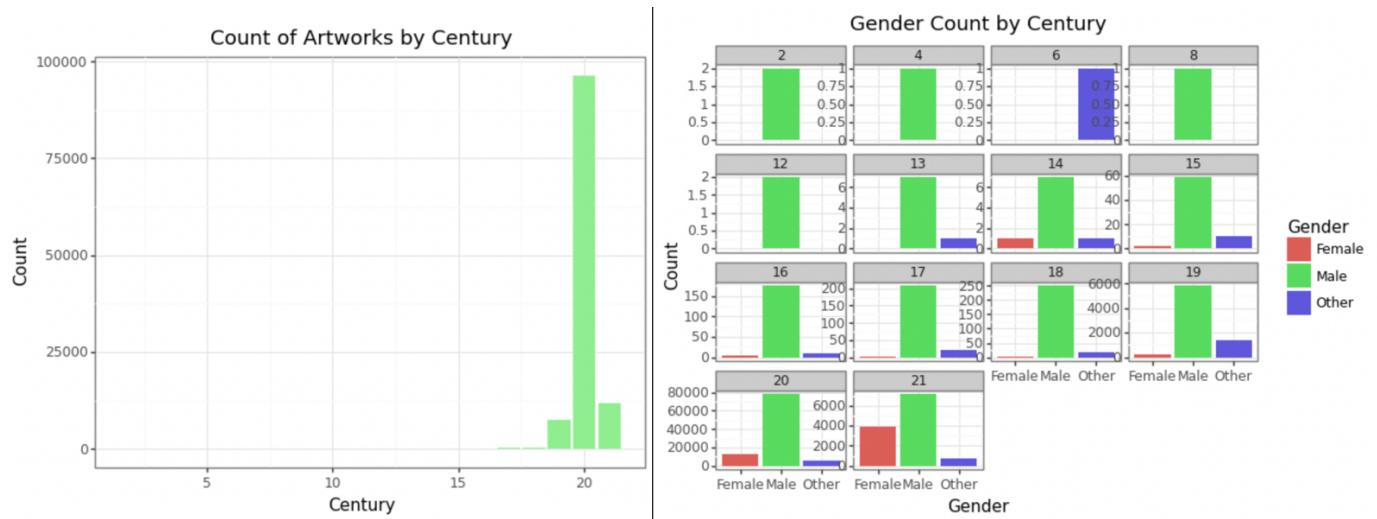
The analysis demonstrates that the data has been effectively mapped back after performing PCA and k-means clustering. By examining the PCA components used in the analysis, it becomes apparent that they mainly capture features related to the year of artwork release. Consequently, it can be inferred that artworks created by women are relatively more recent compared to those created by men, as indicated by the clustering results and the distribution of data points along the principal components.



## Analysis over Time

From this, we performed an analysis of over time to see when most artworks were created and how the proportion of female artists evolved over centuries.

Here are our results:



Again, this is coherent with our hypothesis that there are more female artists as recently and that the proportion grows over time.

## Final Results

1. There are significantly more male than female artists
2. The proportion of female artists has grown over time
3. There is also underrepresentation of many other countries

## **Problems Faced**

During the data analysis process, several challenges were encountered. One significant issue was the presence of missing values in the dataset, which required extensive data cleaning to ensure the reliability and completeness of the results. Additionally, the files containing the data were large and difficult to handle, leading to slower processing and making it challenging to read and manipulate the data efficiently. The data processing also required a substantial amount of time due to the extensive amount of information to go through, and the program would occasionally crash, causing further delays and interruptions. Another obstacle was the inconsistent or missing gender information for artworks, which often posed difficulties when attempting to categorize and analyse the data based on gender.

## **V. Conclusion**

In conclusion, the underrepresentation of female artists in the art world is a significant issue that calls for attention and action. Despite the central role women have played as subjects in art, their presence as creators has been limited throughout history. The work of the Guerrilla Girls, a group of feminist artists, highlights the gender and racial inequalities that persist in the arts community. Through their thought-provoking projects, they have exposed the disparities in representation, revealing the low percentage of women artists compared to the abundance of artworks featuring female bodies.

The analysis conducted aimed to expand on the Guerrilla Girls' efforts by providing a broader and more comprehensive examination of gender representation in art across different countries and museums. The findings confirmed the existing gender imbalance, with a significantly higher number of male artists compared to female artists. However, there were indications of progress over time, with a growing proportion of female artists being represented in museums. The analysis also revealed that the underrepresentation issue extends beyond gender, as many countries are not adequately represented in the art world. This lack of diversity calls for greater efforts to promote inclusivity and provide platforms for artists from diverse backgrounds to showcase their work.

Finally, the art world must address the biases and inequalities that persist in order to create a more inclusive and representative environment. It is crucial to recognize and celebrate the contributions of female artists and artists from underrepresented groups, ensuring that their voices and perspectives are heard and appreciated. Only by embracing diversity and challenging the existing norms can the art world truly reflect the rich and varied experiences of humanity.