How Diverse is the MoMA: While Modern, Not Diverse in Its Artists and Directors*

An analysis of the MoMA's public datasets on their collections and exhibits

Sakura Ariga

19 April 2023

The MoMA's public GitHub data on their exhibits and collections was analyzed to determine how diverse the artists and directors at the MoMA are. It was found that the MoMA is comprised largely of white men, but that this relationship has changed with time. This finding gives quantitative proof of the need for more representation within not jus the MoMA, but all American museums as a whole.

1 Introduction

Diversity in cultural institutions is a topic that is increasingly being discussed and raised as a problem to be tackled, in museums in particular (**museumnext?**). While the main goal of museums is to serve as cultural institutions and a place of learning, oftentimes the objects on display or the decision-making staff who work behind the scenes do not meet this mission of inclusion.

This paper will investigate the level of diversity in one such museum: the Museum of Modern Art (also known as MoMA). Founded in 1929, the MoMA, as its name suggests, houses a collection of contemporary art pieces. It was originally created to go against traditional art museums by housing only modern art, and has since grown to become an influential cultural institution (Museum of Modern Art, n.d.b). Using data made publicly available on the MoMA's GitHub account, this paper aims to evaluate how diverse the artists of the artworks on display are, as well as how diverse the directors and department heads who chose to display these artworks are, in terms of gender and nationality. Thus, the estimand is: the percentage of artists, directors, and department heads in the MoMA's history who are / were not white men.

*Code and data are available at: https://github.com/sakura-ariga/MOMA_diversity_analysis.git.

It was found that the MoMA's artists and directors were predominantly men. Simple linear regression models show that time affects this percentage of female artists, directors, and department heads in that the number of female artists has increased on average by _____ percentage points per year, while the number of female directors and department heads has increased on average by 0.25 percentage points per year. These findings are important because they provide numerical proof of the lack of gender diversity in the MoMA, both in the artists displayed and in the directors and department heads who choose the artists to display, which could motivate the MoMA to take active steps to better the gender representation within their institution.

This paper's structure will begin by explaining the MoMA's publicly made available data regarding its artists and staff that is the subject of analysis. It will then explore this data to understand overall trends in nationality and gender distribution in the MoMA. Thirdly, linear models will be used to further analyze the MoMA's gender distribution and its changes over time. Finally, key findings, limitations, and next steps will be discussed.

2 Data

2.1 Data Source

In this report, the 32,537 observations from the MoMA's exhibition index dataset and 65 observations from the directors and department heads dataset are used to investigate the level of diversity among MoMA decision-making directors and displayed artists. This data was obtained from the MoMA's publicly available GitHub, which has several public repositories containing data regarding the museum (Museum of Modern Art, n.d.a). For this analysis, the MoMA's "exhibitions" GitHub repository (Manuel Charr 2020) was used. The exhibition index dataset was created by a MoMA Archives team and contains information regarding the exhibitions featured in the MoMA from 1929 to 1989. The team is currently working on adding information on exhibits from 1990 onwards to the dataset. The directors and department heads dataset contains information regarding all of the directors of the entire museum and the department heads of curatorial departments within the museum from 1929 to the present.

This report was created using the R statistical programming language (R Core Team 2020). For the results and analysis of this report, all figures were created using the tidyverse package (Wickham et al. 2019). Additionally, the tables were created using the packages knitr (Xie 2023) and kableExtra (Zhu 2021), the graphs were created using the packages tidyr (Wickham, Vaughan, and Girlich 2023) and ggplot2 (Wickham 2016), and the models were displayed using the modelsummary package (Arel-Bundock 2022).

2.2 Variables of Interest

This report selects 7 variables for analysis from the MoMA exhibition index dataset and 6 variables from the directors and department heads. Variables regarding social background, specif-

Table 1: First ten rows of the MoMA Exhibits dataset

Exhibition Title	Artist Name	Nationality	Gender	Exhibit
C<e9>zanne, Gauguin, Seurat, Van Gogh	Paul C<e9>zanne	French	Male	
C<e9>zanne, Gauguin, Seurat, Van Gogh	Paul Gauguin	French	Male	
C<e9>zanne, Gauguin, Seurat, Van Gogh	Vincent van Gogh	Dutch	Male	
C<e9>zanne, Gauguin, Seurat, Van Gogh	Georges-Pierre Seurat	French	Male	
Paintings by 19 Living Americans	Charles Burchfield	American	Male	
Paintings by 19 Living Americans	Charles Demuth	American	Male	
Paintings by 19 Living Americans	Preston Dickinson	American	Male	
Paintings by 19 Living Americans	Lyonel Feininger	American	Male	
Paintings by 19 Living Americans	George Overbury ("Pop") Hart	American	Male	
Paintings by 19 Living Americans	Edward Hopper	American	Male	

ically gender and nationality, were chosen in order to measure diversity within the MoMA.

Table 1 shows the 7 variables of the MoMA exhibition index dataset and the first 10 rows of it, where each row represents a different individual constituent in the exhibit. The Exhibition Title variable refers to the exhibit name. The Name variable indicates the artist. Nationality indicates the individual's nationality (e.g. "American") and is investigated further in (_____**?**). Gender has two options of "Male" or "Female". Exhibition Start and Exhibition End indicate the time period that the exhibit was displayed at the MoMA for, and are the only numerical variables in this dataset.

Table 2 shows the 6 variables of the MoMA directors and department heads dataset and the first 10 rows of it, where each row represents a different individual. The Department variable refers to the specific curatorial department within the MoMA that the individual is in charge of. As with the exhibition dataset, the Name variable indicates the person, Nationality indicates the individual's nationality (e.g. "American") and is investigated further in (_____**?**), and Gender has two options of "Male" or "Female". Start and End indicate the time period that the individual held the position for, and are the only numerical variables in this dataset.

The data was cleaned and modified using the tidyverse package (Wickham et al. 2019).

2.3 Data Visualization

The data was visualized as bar graphs in order to best understand the distribution of the diversity indicators in the data (i.e. the Gender and Nationality variables).

Figure 1 displays the nationality distribution of the artists whose works have been displayed in exhibits in the MoMA. Aside from the large number of missing values (marked as NA at the rightmost bar), the most common nationality among is American, at 14,112 artists. The second highest nationality is French, at 3,747 artists - nearly a fifth of the American artists.

Table 2: First ten rows of the MoMA Directors and Department Heads dataset

Department	Name	Start	End	Nationality	Gender
The Museum of Modern Art	Alfred H. Barr, Jr.	1929	1943	American	Male
The Museum of Modern Art	René d'Harnoncourt	1949	1968	American	Male
The Museum of Modern Art	Bates Lowry	1968	1969	American	Male
The Museum of Modern Art	John B. Hightower	1970	1972	American	Male
The Museum of Modern Art	Richard E. Oldenburg	1972	1994	American	Male
The Museum of Modern Art	Glenn D. Lowry	1995	2023	American	Male
Department of Painting and Sculpture	Alfred H. Barr, Jr.	1929	1940	American	Male
Department of Painting and Sculpture	Alfred H. Barr, Jr.	1940	1943	American	Male
Department of Painting and Sculpture	James Thrall Soby	1943	1945	American	Male
Department of Painting and Sculpture	James Johnson Sweeney	1945	1946	American	Male

The large number of American artists may be explained in part because the MoMA itself is American. It is surprising though that while the MoMA has presented 2,500+ exhibits throughout its history, _____.

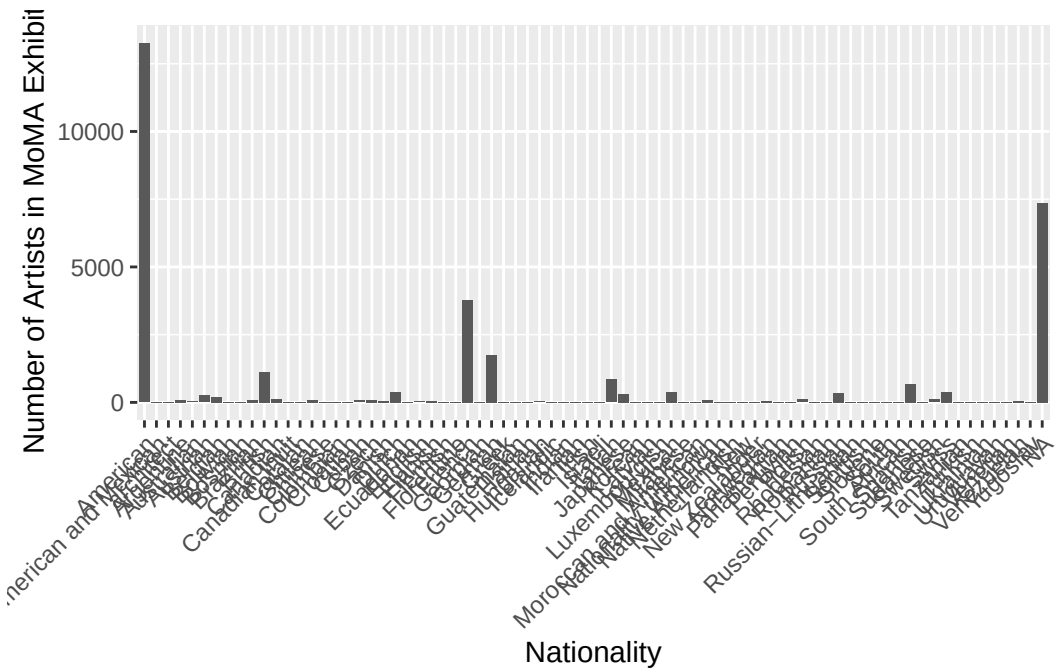

Figure 1: Nationality Distribution of Artists in MoMA Exhibits

Figure 2 displays the gender distribution of the artists whose works have been displayed in exhibits in the MoMA. As the graph demonstrates, there is a much larger number of male artists who have been featured in the MoMA exhibits than female artists. With 22,518 male artists and 2,301 female artists, the male to female ratio is approximately 10:1, showing significant inequality in gender representation in MoMA artists. This may be partly explained by _____.

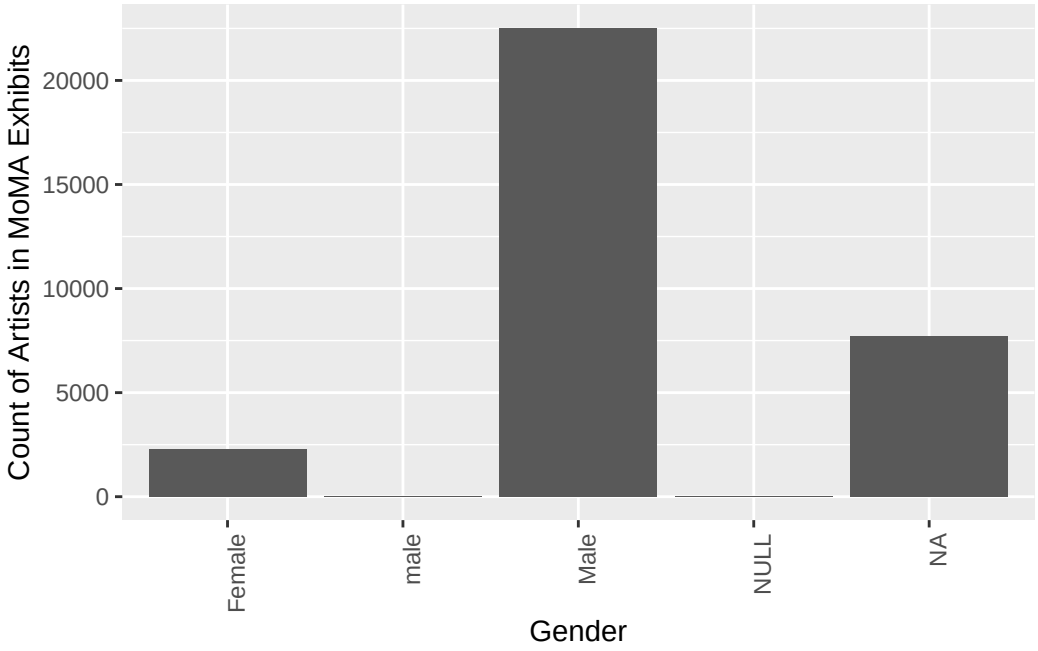

Figure 2: Gender Distribution of Artists in MoMA Exhibits

Figure 3 displays the nationality distribution of all of the directors and department heads at the MoMA. Here too, like in the nationality distribution of artists, there is a majority of American nationality among the directors and department heads, at 50 Americans. Compared to the 3 Swiss, 3 British, 2 German, 1 French, and 1 Austrian directors and department heads, _____.

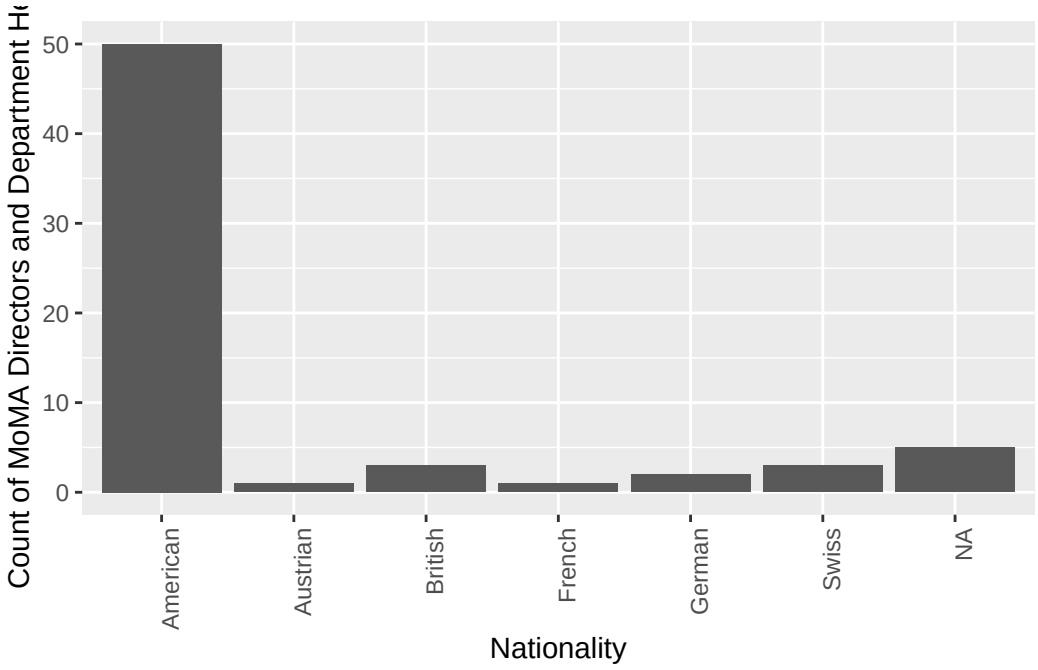

Figure 3: Nationality Distribution of Directors and Department Heads in MoMA

Figure 4 displays the gender distribution of all of the directors and department heads at the MoMA. Here too, there are many more male directors and department heads compared to female department heads. With 53 male directors and department heads and 12 female directors and department heads, the male to female ratio is approximately 4:1, showing inequality in gender representation in MoMA leadership. This may be partly explained by _____.

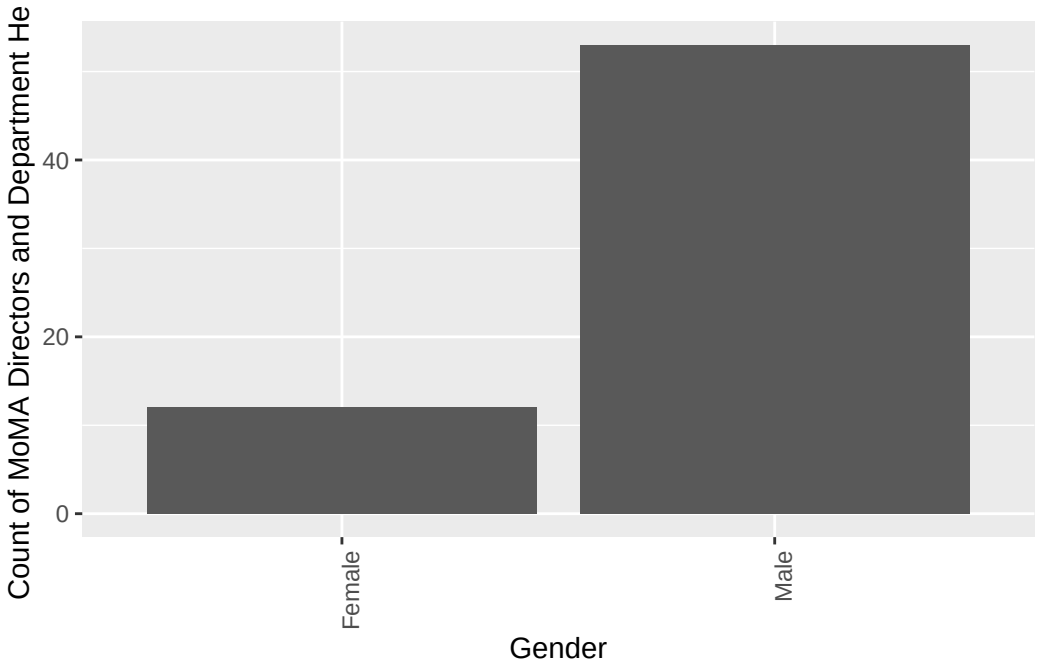

Figure 4: Gender Distribution of Directors and Department Heads in MoMA

3 Model

From the exploratory data analysis above, it is clear that both the number of female MoMA directors and the number of female artists displayed in the MoMA is much less than the number of male counterparts in these positions. This gender relationship is worth further investigation using linear models.

The goal of this modelling strategy is threefold. Firstly, the relationship between the percentage of female directors at the MoMA and years will be investigated to determine whether the gender distribution of those who have the highest decision-making power at the MoMA has improved over time. Secondly, another model will look at the relationship between the percentage of female artists in exhibits at the MoMA and years to determine whether the gender distribution of the artists displayed in MoMA exhibits has improved over time. Finally, the third model will investigate the effect that female director percentage in the MoMA has on female artist percentage in MoMA exhibits to determine whether the gender distribution of top MoMA decision-makers has an effect on the gender distribution of artists selected for exhibits.

Here we briefly describe the three linear regression models used to investigate this relationship between gender and time.

3.1 Model Set-Up

The first linear regression model regarding the effect of time on the percentage of female MoMA directors is as follows:

$$\hat{y} = \beta_0 + \beta_1 x_{year}$$

The output of this linear regression model gives the percentage of female MoMA directors. Thus, the response in the model is the percentage of female MoMA directors based on the year explanatory variable. A linear regression model was used because both the response variable and the explanatory variable are quantitative.

This model focuses on the following two aspects of MoMA directors:

- Percentage of female MoMA directors, the response variable, is used as a quantitative measure of gender diversity that is between 0 and 100
- Year, the explanatory variable, is a numeric variable between 1929 (the founding year of the MoMA) and 2023 (the year that this report was published) indicating a specific year in the history of the MoMA

The second linear regression model regarding the effect of time on the percentage of female MoMA artists in the exhibits is similar to the first and is as follows:

$$\hat{y} = \beta_0 + \beta_1 x_{year}$$

The output of this linear regression model gives the percentage of female MoMA artists whose works have been displayed in MoMA exhibits over time. Thus, the response in the model is the percentage of female MoMA artists based on the year explanatory variable. A linear regression model was used because both the response variable and the explanatory variable are quantitative.

This model focuses on the following two aspects of MoMA artists:

- Percentage of female MoMA artists, the response variable, is used as a quantitative measure of gender diversity that is between 0 and 100
- Year, the explanatory variable, is a numeric variable between 1929 (the founding year of the MoMA) and 1990 (the year that the MoMA public GitHub has data until) indicating a specific year in the history of the MoMA

The third and final linear regression model differs slightly from the previous two and is as follows:

$$\hat{y} = \beta_0 + \beta_1 x_{femaledirectorpercentage} + \beta_2 x_{year}$$

The output of this linear regression model gives the percentage of female MoMA artists whose works have been displayed in MoMA exhibits over time. Thus, the response in the model is

the percentage of female MoMA artists based on the percentage of female MoMA directors and year explanatory variables. A linear regression model was used because both the response variable and the explanatory variables are quantitative.

This model focuses on the following three aspects of MoMA artists:

- Percentage of female MoMA artists, the response variable, is used as a quantitative measure of gender diversity that is between 0 and 100
- Percentage of female MoMA directors, an explanatory variable, is a numeric variable between 0 and 100 indicating the gender distribution of the directors who had the decision-making power to choose the artists to be displayed
- Year, an explanatory variable, is a numeric variable between 1929 (the founding year of the MoMA) and 2023 (the year that this report was published) indicating a specific year in the history of the MoMA

All of the models are run in R (R Core Team 2020) and displayed using modelsummary (Arel-Bundock 2022).

3.1.1 Model Justification

For the first two models regarding the effect of time on the percentage of female MoMA directors and on artists, time was chosen to be evaluated because there has been an increased push for gender equality in workplaces across the United States over the years. Because of this, a positive relationship between time in years and the percentage of female directors and female artists is expected for both the first linear regression model (regarding the effect of time on the percentage of female MoMA directors) and the second linear regression model (regarding the effect of time on the percentage of female MoMA artists).

Given these two models, it is worth investigating whether there is a relationship between the percentage of female directors and the percentage of female artists, to determine whether female directors are more likely to be aware of gender inclusion and thus display more female artists than their male counterparts. Because women are subjects of workplace gender discrimination, they are more likely to be aware of the need for inclusive gender representation, so a positive relationship between time in years and the percentage of female directors and female artists is expected.

4 Results

Our results are summarized in **?@tbl-modelresults**.

Table 3: Explanatory model of percentage of female MoMA directors based on year

	First model
(Intercept)	−483.62
	(80.40)
Year	0.25
	(0.04)
Num.Obs.	95
R2	0.294
R2 Adj.	0.286
AIC	727.0
BIC	734.7
Log.Lik.	−360.497
RMSE	10.76

Table 4: Explanatory model of percentage of female MoMA artists based on exhibit year

	Second model
(Intercept)	379.57
	(76.49)
Year	−0.19
	(0.04)
Num.Obs.	1387
R2	0.016
R2 Adj.	0.015
AIC	12 729.2
BIC	12 744.9
Log.Lik.	−6361.598
RMSE	23.75

Table 5: Explanatory model of percentage of female MoMA artists based on exhibit year and percentage of female MoMA directors

	Third model
(Intercept)	461.54
	(82.14)
PercentageFemale.y	0.16
	(0.06)
Year	−0.23
	(0.04)
Num.Obs.	1387
R2	0.021
R2 Adj.	0.020
AIC	12 723.9
BIC	12 744.8
Log.Lik.	−6357.954
RMSE	23.69

5 Discussion

5.1 Gender distribition

5.2 Nationality distribition

5.3 Analyzing the model: is time a factor?

5.4 Comparing artists vs directors

5.5 Weaknesses and next steps

- a LOT of NAs in the exhibit dataset, especially regarding artist gender, so results are skewed bc of that
- binary understanding of gender that is exclusive

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Manuel Charr. 2020. *What Can Museums Teach Us about Diversity?* Museum Next. https://www.museumnext.com/article/what-can-museums-teach-us-about-diversity/.

Museum of Modern Art. n.d.a. *MoMA Public GitHub*. New York, United States of America: Museum of Modern Art. https://github.com/MuseumofModernArt.

———. n.d.b. *The Museum of Modern Art History*. New York, United States of America: Museum of Modern Art. https://www.moma.org/about/who-we-are/moma-history.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.