Investigating the Gender and Nationality Distributions of the MoMA: While Modern, Not Diverse in Its Artists and Decision-making Staff*

An analysis of the MoMA's public datasets

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 in terms of gender and nationality. It was found that the MoMA is comprised largely of American men and particularly lacking in their representation of women, both as artists and as directors and department heads. However, it was found that women representation has a positive relationship with time, indicating that the MoMA's gender distribution has improved slightly over time. This finding gives quantitative proof of the need for more gender representation in the MoMA, both in the artists displayed and in the top decision-making staff hired.

Table of contents

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

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 (Manuel Charr 2020). 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.

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 0.08 percentage points per year, while the number of female directors and department heads has increased on average by 0.24 percentage points per year. It was also found that the percentage of female directors and department heads has a positive relationship with the percentage of female artists displayed, with an increase of an average of 0.10 percentage points in women artists with every 1% increase of women directors and department heads. 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 (Museum of Modern Art 2016) 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, specifically 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 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 1: First ten rows of the MoMA Exhibits dataset

Exhibition Title	Artist Name	Nationality	Gender	Exhibition Start	Exhibition End
C<e9>zanne, Gauguin, Seurat, Van Gogh	Paul C<e9>zanne	French	Male	1929	1929
C<e9>zanne, Gauguin, Seurat, Van Gogh	Paul Gauguin	French	Male	1929	1929
C<e9>zanne, Gauguin, Seurat, Van Gogh	Vincent van Gogh	Dutch	Male	1929	1929
C<e9>zanne, Gauguin, Seurat, Van Gogh	Georges-Pierre Seurat	French	Male	1929	1929
Paintings by 19 Living Americans	Charles Burchfield	American	Male	1929	1930
Paintings by 19 Living Americans	Charles Demuth	American	Male	1929	1930
Paintings by 19 Living Americans	Preston Dickinson	American	Male	1929	1930
Paintings by 19 Living Americans	Lyonel Feininger	American	Male	1929	1930
Paintings by 19 Living Americans	George Overbury ("Pop") Hart	American	Male	1929	1930
Paintings by 19 Living Americans	Edward Hopper	American	Male	1929	1930

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

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 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.

The large number of American artists may be explained in part because the MoMA itself is American.

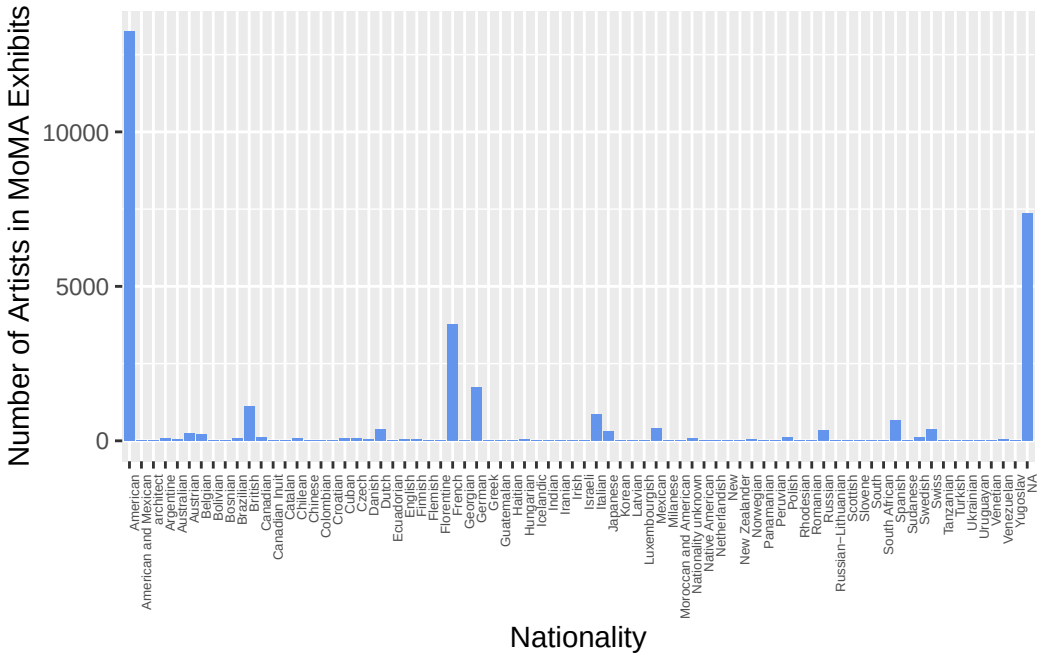

Figure 1: Nationality Distribution of Artists in MoMA Exhibits

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.

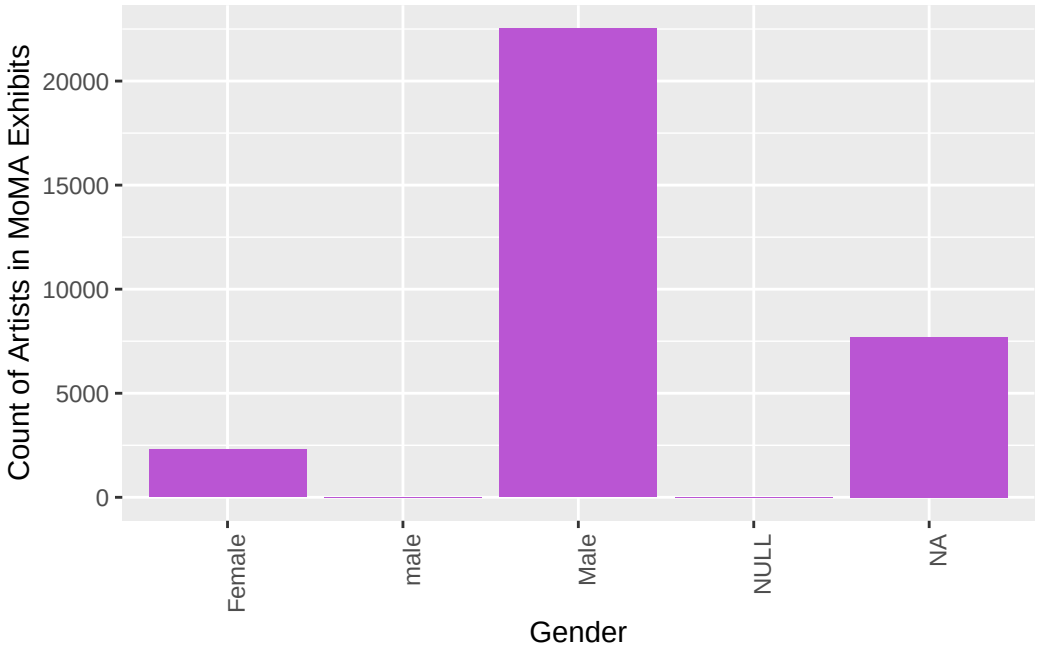

Figure 2: Gender Distribution of Artists in MoMA Exhibits

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 out of 60 Americans.

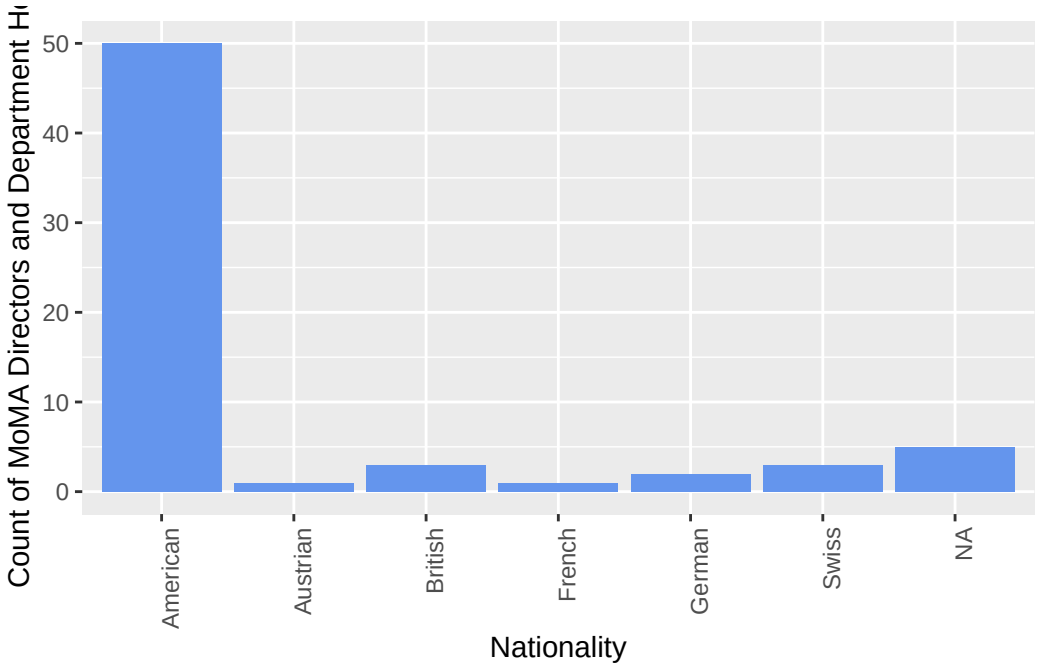

Figure 3: Nationality Distribution of Directors and Department Heads in MoMA

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.

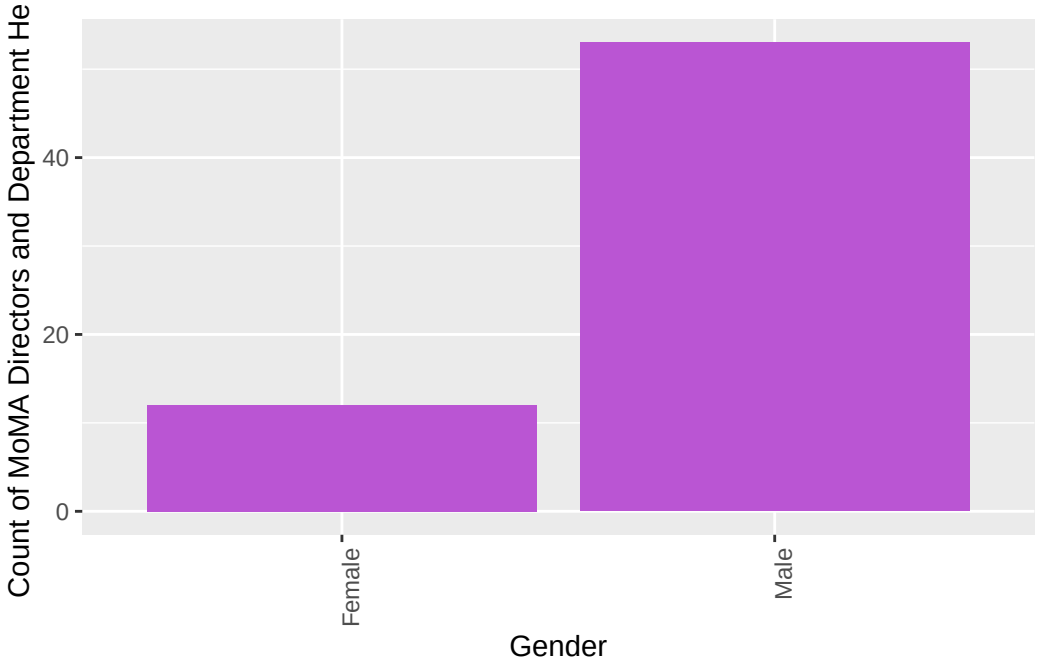

Figure 4: Gender Distribution of Directors and Department Heads in MoMA

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:

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

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).

3.2 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

The results of the first linear regression model of the effect of time on the percentage of female MoMA directors is summarized in Table 3. The prediction of a positive relationship between time in years and the percentage of female MoMA directors proved correct, with the percentage of female directors increasing every year by 0.25 percentage points since the MoMA's founding in 1929. We know that this finding is significant because the p-value of this model is less than the threshold of 0.05. Figure 5 demonstrates this relationship as a scatterplot, better visualizing the positive linear relationship between the amount of women leaders in the MoMA and time.

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

	First model
(Intercept)	−448.72
	(78.82)
Year	0.24
	(0.04)
Num.Obs.	95
R2	0.272
R2 Adj.	0.264
AIC	723.2
BIC	730.9
Log.Lik.	−358.612
RMSE	10.55

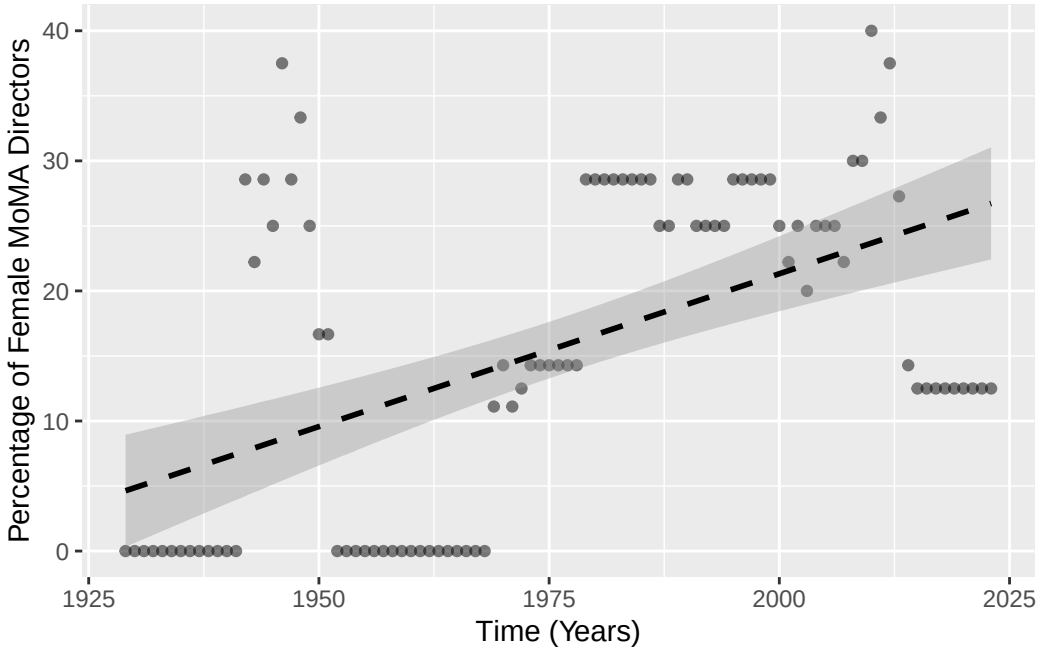

Figure 5: Linear model of percentage of female MoMA directors by year

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

	Second model
(Intercept)	−142.14
	(37.77)
Year	0.08
	(0.02)
Num.Obs.	1378
R2	0.011
R2 Adj.	0.010
AIC	10 692.3
BIC	10 708.0
Log.Lik.	−5343.137
RMSE	11.69

The results of the second linear regression model of the effect of time on the percentage of female MoMA artists is summarized in Table 4. The prediction of a positive relationship between time in years and the percentage of female MoMA artists proved correct, with the percentage of female artists increasing every year by 0.08 percentage points since the MoMA's founding in 1929. This finding is significant because the p-value of this model is less than the threshold of 0.05. Figure 6 demonstrates this relationship as a scatterplot, better visualizing the positive linear relationship between the amount of women artists who have been displayed in MoMA exhibits and time.

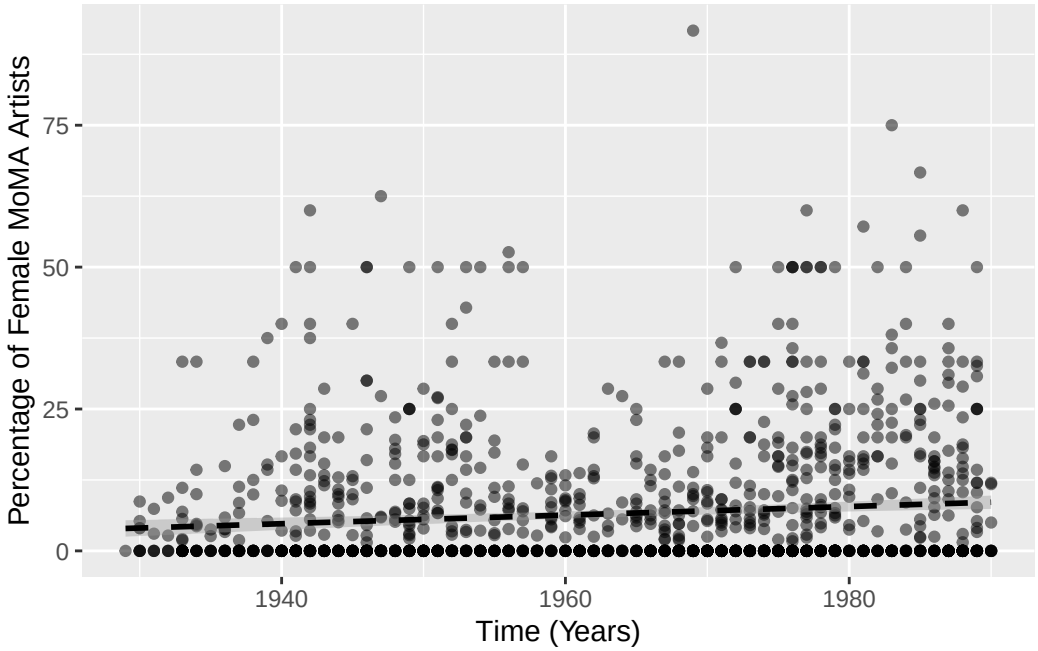

Figure 6: Linear model of percentage of female MoMA artists by year

The results of the third linear regression model of the effect of time and the percentage of female MoMA directors on the percentage of female MoMA artists is summarized in Table 5. The prediction of a positive relationship between time in years and the percentage of female directors and the percentage of female MoMA artists proved correct, with the percentage of female artists increasing by 0.10 percentage points with every 1% increase in the percentage of female directors.

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

	Third model
(Intercept)	−92.88
	(40.33)
PercentageFemale.y	0.10
	(0.03)
Year	0.05
	(0.02)
Num.Obs.	1378
R2	0.019
R2 Adj.	0.018
AIC	10 682.8
BIC	10 703.7
Log.Lik.	−5337.402
RMSE	11.64

5 Discussion

5.1 Gender Distribition

From Figure 2 and Figure 4, we see that there is a significant gap between representation of men versus women in the MoMA, both for the artists displayed and for the museum staff with the most power. Comparing Figure 2 and Figure 4, the men:women ratio is more unequal in the artists, at a ratio of 10:1, than in the directors and department heads, at a ratio of 4:1.

The models found that _____

This distribution is surprising, especially given that the MoMA was founded due to the works and ideas of women: Lillie P. Bliss, Mary Quinn Sullivan, and Abby Aldrich Rockefeller - three influential women who were patrons of the arts in the 1920s (Museum of Modern Art, n.d.b). One potential reason for this could be American society's historical attitude towards women versus men, tending to elevate and encourage men to pursue careers as artists or museum curators and discouraging women to. Another reason could be related to the findings of the third model found in Table 5: if women directors and department heads are more likely to choose women artists' work to exhibit, the same may go for men directors and department heads with men artists. These two reasons could be related to each other as well, with male directors and department heads more likely to choose male artists due to social conditioning from society's expectations of gender roles.

While the above models (found in Figure 5, Figure 6, Table 5) show that change has occurred and that the percentage of women represented in the MoMA's displayed artists and decision-making staff has increased over time, there is still much more room for improvement. As an institution with "modern" in its name, the MoMA should act as a role model for other cultural institutions and be an active leader in incorporating more diversity in the museum, both in showcasing more female artists and in hiring more female directors and department heads. This holds true especially now that it has been nearly a century since the MoMA's opening and thus much more modern art has been created and can be displayed.

5.2 Nationality Distribition

Figure 1 and Figure 3 show us that the museum is predominantly made up of American representation, both in its artists and in its decision-making staff. Most interesting is that the nationalities most represented in the artists is also represented well in the directors and department heads: American, French, British. One possible explanation of this is that these three countries place a heavy emphasis on the modern arts, and so produce many artists and museum staff. Another explanation could be that the MoMA places a eurocentric, Western-heavy emphasis on their exhibits, and thus is biased towards selecting American, French, and British artists and staff. Because nationality is not the best indicator for social inclusion practices, analysis on racial discrimination cannot be made here from this data.

5.3 Weaknesses and Next Steps

One weakness of this analysis is the binary understanding of gender as only male or female. This excludes artists, directors, and department heads who do not identify within this gender binary and thus this analysis was not able to completely measure gender diversity in the MoMA. Another bias present in this analysis is the removal of missing values that occurred during the data cleaning step of this paper, which made up a large number of responses, particularly in the gender of artists. This inevitably led to the results of the gender distribution analysis being biased and not entirely accurate.

Thus, an appropriate next step would be to conduct more research on the demographics of the artists, directors, and department heads of the MoMA so that a more accurate and meaningful analysis can be conducted. Not only would this account for some of the missing values that were excluded in this paper, this could result in analysis on racial equality in the MoMA. While the nationality variable that was analyzed in this paper was helpful in understanding some aspect of inclusion in the MoMA, in a museum that is so dominated by Americans, both as artists and as decision-making museum staff, an analysis on the racial representation in the MoMA would yield more actionable insights.

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. 2016. *The Museum of Modern Art (MoMA) Exhibition and Staff Histories*. New York, United States of America: Museum of Modern Art. https://github.com/MuseumofModernArt/exhibitions.

———. 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.