# 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

18 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 museums is a topic that is increasingly being discussed and raised as a problem in museums. 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. 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: how much of the artists and directors in the MoMA's history are/were not white men.

It was found that the MoMA's artists and directors were predominantly men. A simple linear regression model shows that time affects this proportion of men.

---

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

The paper's structure begins by explaining the data used, followed by a linear model, a discussion of the relationship between variables in the data and in the model, and a discussion on key findings.
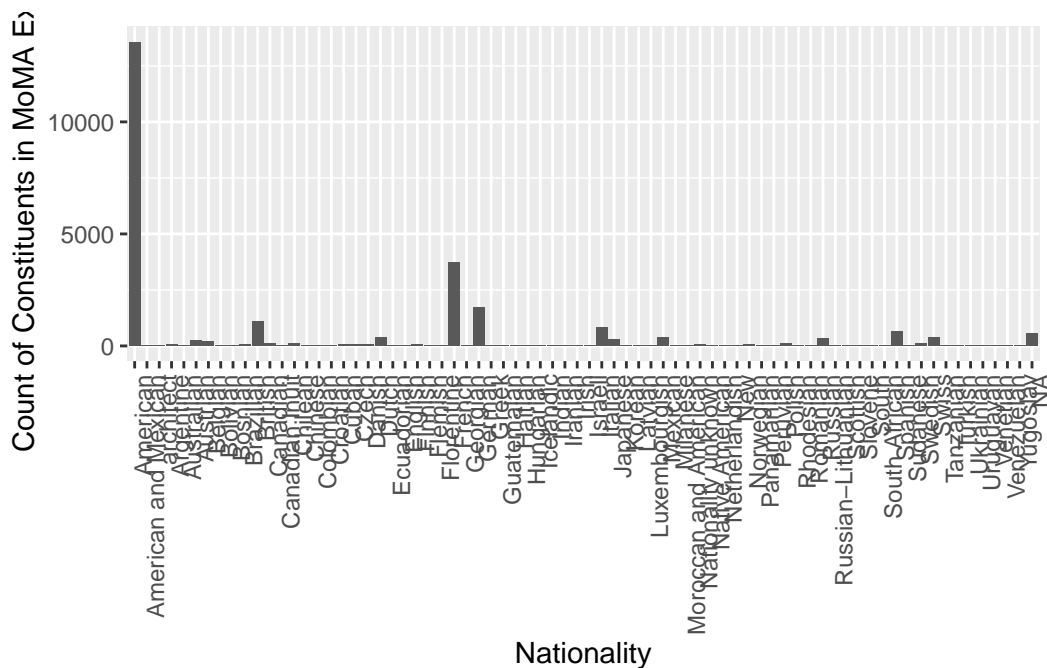
## 2 Data

Investigating artists:



Figure 1: Nationality Distribution of Consituents (Artists, Staff) in MoMA Exhibits
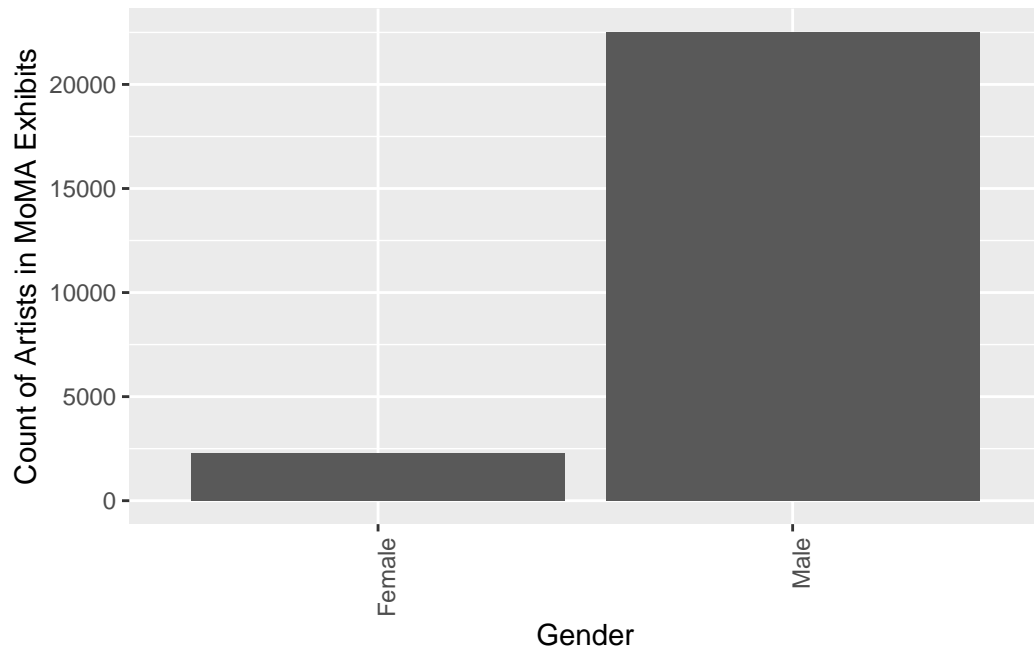
Figure 2: Gender Distribution of Artists in MoMA Exhibits

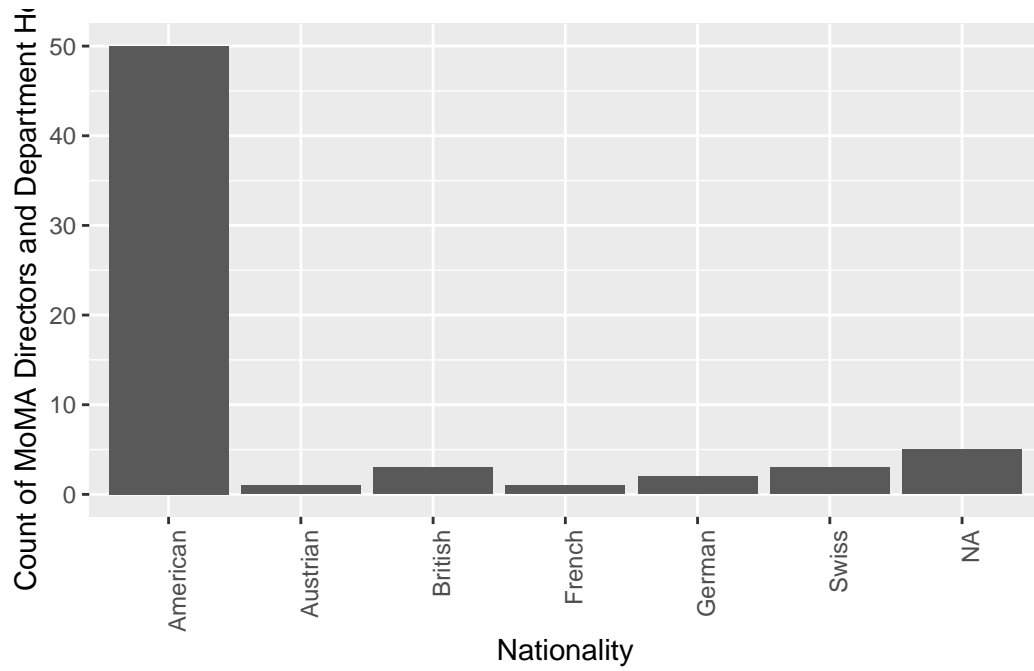Investigating directors and department heads:



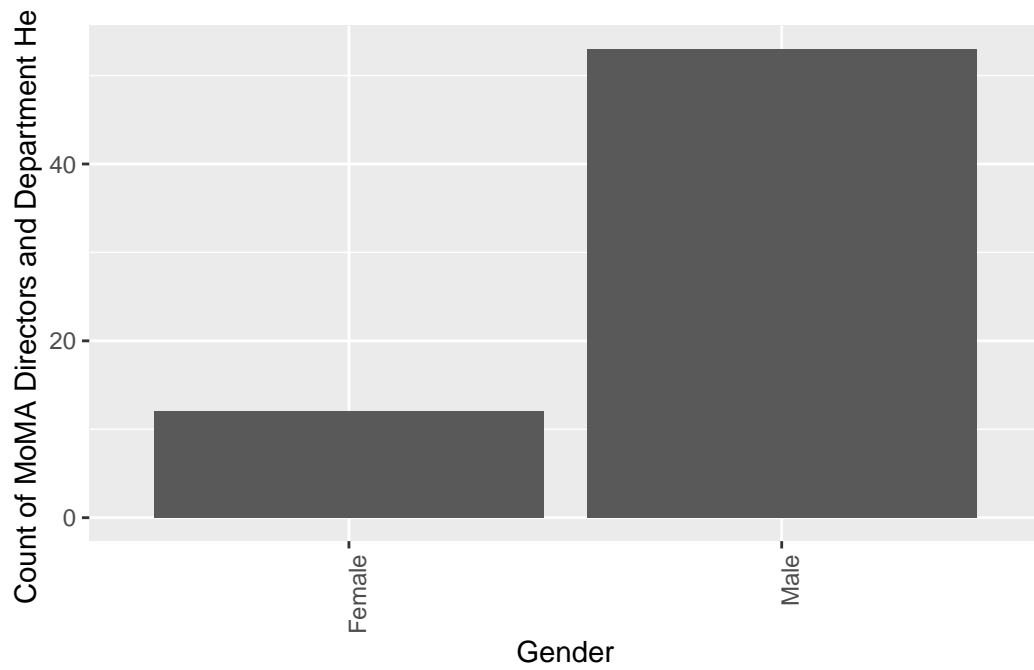Figure 3: Nationality Distribution of Directors and Department Heads in MoMA

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. Background details and diagnostics are included in Appendix B.

## 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 (**citeR?**) using the `stats` package of (**stats?**).

### 3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance $\theta$.

Table 1: 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 |

# 4 Results

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

# 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

Table 2: Explanatory model of percentage of female MoMA artists based on exhibit year

|  | Second model |
| --- | --- |
| (Intercept) | −96.67 |
|  | (33.06) |
| Year | 0.05 |
|  | (0.02) |
| Num.Obs. | 1420 |
| R2 | 0.007 |
| R2 Adj. | 0.006 |
| AIC | 11 025.6 |
| BIC | 11 041.4 |
| Log.Lik. | −5509.815 |
| RMSE | 11.72 |

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

|  | Third model |
| --- | --- |
| (Intercept) | −47.28 |
|  | (34.54) |
| PercentageFemale.y | 0.13 |
|  | (0.03) |
| Year | 0.03 |
|  | (0.02) |
| Num.Obs. | 1420 |
| R2 | 0.021 |
| R2 Adj. | 0.020 |
| AIC | 11 006.6 |
| BIC | 11 027.6 |
| Log.Lik. | −5499.287 |
| RMSE | 11.63 |

# Appendix

# A Additional data details

# B Model details

## B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected
by, the data

Figure 5: **?(caption)**

## B.2 Diagnostics

**?@fig-stanareyouokay-1** is a trace plot. It shows... This suggests...

**?@fig-stanareyouokay-2** is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC
algorithm

Figure 6: **?(caption)**

# C  References