Predicting Artwork Size to aid Museum Workers with Exhibition Curation*

The Influence of Gender, Nationality, and Year of Completion at the Museum of Modern Art

Sima Shmuylovich

April 18, 2024

This paper explores the determinants of artwork size within the Museum of Modern Art (MoMA) collection, focusing on the effects of artist gender, nationality, and year of completion. Utilizing the MoMA Artworks dataset, a linear regression model was employed to analyze how these factors influence the physical dimensions of artworks. The study reveals that artworks by male artists and those identified with multiple artists tend to be smaller in size, challenging conventional expectations about gender and artistic production. Notably, artworks from Algerian and Ghanaian artists are significantly larger, possibly reflecting unique cultural and artistic traditions that favor grand-scale works. The analysis also indicates a clear trend of increasing artwork size over time, suggesting an evolution in artistic practices and the spaces available for art display. These findings not only enhance our understanding of the factors influencing artistic expression but also have practical implications for museum curators and cultural historians in planning exhibitions and interpreting artistic trends. The paper recommends further research using non-linear models and a broader set of variables to more accurately reflect the complex variables that determine artwork size.

Table of contents

*Code and data are available at: https://github.com/sshmuylovich/moma-artworks.git

1 Introduction

Since its inception as America's first museum devoted exclusively to modern art in 1929, MoMA has played a crucial role in showcasing groundbreaking artworks, influencing art appreciation and research worldwide (Robot 2024). As art evolves, so do the characteristics of the artworks themselves, including their size, which often speaks volumes about the cultural, technological, and social contexts in which they were created. This study seeks to explore the determinants of artwork size within the MoMA collection, focusing on the roles of artist gender, nationality, and the year of completion. Understanding artwork size is pivotal not only for academic purposes but also for practical considerations in museum curation and exhibition design. Artwork dimensions can influence curatorial decisions, affecting how art is displayed, experienced, and interpreted by audiences. Despite its importance, the factors determining artwork size have not been extensively quantified within the context of modern art museums.

This paper employs a linear regression model to analyze the MoMA Artworks dataset, investigating how specific artist characteristics impact the physical dimensions of artworks. The estimand is as follows: this paper aims to estimate the size of an artwork as defined by its area based on its artist's gender, its artist's nationality, and its year of completion.

Preliminary findings reveal intriguing variations: artworks by male artists and those identified with multiple artists tend to be smaller, while significant increases in size are noted for artworks from certain nationalities such as Algerian and Ghanaian. Additionally, a clear trend shows that artworks have generally increased in size over the years. These patterns not only challenge conventional narratives about the representation of gender in art but also highlight the impact of cultural and national influences on artistic expression.

The subsequent sections follow a structured format. Section 2 outlines the source and variables central to our analysis. Section 3 details the construction and methodology of the statistical models used. Section 4 presents the key findings of our analysis, while Section 5 critically reviews the content, addresses the implications of the results, acknowledges model limitations, and suggests potential research directions.

2 Data

2.1 Methodology

The data used in this paper was gathered from the MoMA Artworks dataset on Github which is supplemented by the MoMA Collection dataset hosted on Zenodo (Robot 2024). The Artworks dataset was analyzed using the R programming language (R Core Team 2023) and essential packages from `tidyverse` (Wickham et al. 2019), specifically `dplyr` (Wickham et al. 2023) and `tibble` (Müller and Wickham 2023), for efficient data manipulation. The `janitor` package (Firke 2023) facilitated data cleaning by tidying variable names and removing duplicates. For visualizations, `ggplot2` (Wickham 2016) was used to create insightful graphics, and `kableExtra` (Zhu 2021) for stylized data tables. The `here` package (Müller 2020) was used for managing file paths and ensuring reproducibility. For modeling, `modelsummary` (Arel-Bundock 2022) and `rstanarm` (Goodrich et al. 2022) were used. Data testing was conducted with the help of `testthat` (Wickham 2011) and the pdf was rendering with the help of `knitr` (Xie 2023).

2.2 Dataset Source: The Museum of Modern Art Collection

The MoMA Collection Artworks dataset is a record of all artworks that have been featured in the museum since its first acquisition on November 19, 1929. It is updated automatically via their internal system and was last updated on March 7, 2024 following an acquisition of 123 new artworks. A total of 153,687 artworks are included in the MoMA Collection Artworks dataset, 99,389 of which have been cataloged by MoMA curators. The dataset includes 30 variables, many of which could have been included in the analysis but was narrowed down to 3 variables that could correlate with well with the area of an artwork (Robot 2024).

2.3 Variable Selection Justifications

Gender was picked as one of the predictor variables because gender of an artist can influence artwork size. For example, historically, access to resources, training, and spaces where large-scale works could be created and displayed were often restricted based on gender. Male artists, for instance, had greater access to larger studios or more patronage to undertake ambitious projects, which could lead to larger artwork sizes. Conversely, female artists, who were often marginalized in the art world, might have adapted their artistic output to smaller, more intimate scales, which were more feasible within their constraints. Analyzing the gender of artists in relation to the size of their artworks could provide insights into the social and economic factors that influenced artists' work and their artistic decisions. This kind of analysis can deepen our understanding of the systemic biases in the art world and highlight the evolution or persistence of these dynamics over time. A visualization of the proportion of artwork size by gender in the MoMA collection can be found here, Figure 1. Note that in the figure, Gender is shown as male, female, or other (all non-gender binary genders grouped together). While this was done to make the figure easier to digest visually the model uses the actual distribution of gender in the dataset which can be found in Table 3.

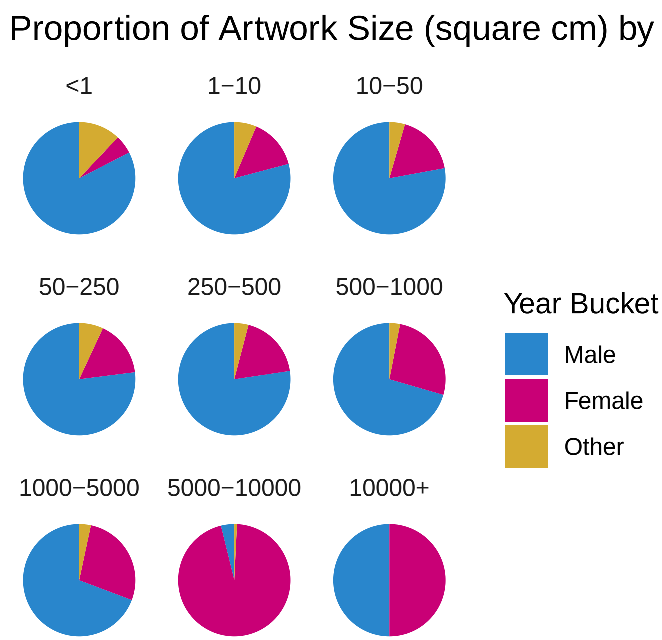

Figure 1: Proportion of Artwork Size by Gender in the MoMA Collection

Nationality was picked as one of the predictor variables because the nationality of an artist can reflect a variety of economic, cultural, and material conditions that differ across countries and influence the scale at which artists work. Artists from countries with robust funding for the arts, access to large studio spaces, and significant commercial or state support might produce larger artworks. These factors often allow for ambitious projects that require more space and resources. In contrast, artists from regions with limited access to materials and funding might focus on smaller-scale works, which are more practical under constrained conditions. Moreover, cultural factors linked to nationality can play a crucial role. Some cultures might emphasize grand, expansive artistic expressions as a reflection of national pride or cultural heritage, while others might prioritize intimacy, detail, and small-scale craftsmanship. Analyzing how nationality correlates with the size of artworks at MoMA can uncover patterns that tell stories of economic disparity, cultural emphasis, and resource availability. This form of analysis not only enhances our understanding of the art itself but also opens up discussions about global equity and the diversity of artistic expression across different cultural landscapes. A sample of 10 nationalities can be observed in Table 4.

Year of completetion was picked as one of the predictor variables because the year an artwork was completed can influence artwork size. The year of completion serves as a reflection of the artistic, cultural, and technological era in which artists were working, affecting the scale and ambitions of their projects. Over different periods, artistic movements and styles have emphasized varying scales and forms. For instance, monumental public art and large-scale abstract expressions were particularly prominent during the mid-20th century, reflecting post-war optimism and the availability of new spaces and materials. Conversely, earlier or later periods might show a preference for smaller, more detailed works due to economic conditions, material shortages, or shifts in consumer and public tastes. Technological advancements also play a significant role. The development of new materials and tools can enable larger or more intricate artworks. For example, the introduction of acrylic paints or metal welding techniques opened up new possibilities for size and form in sculpture and painting. This approach not only enhances our appreciation of individual artworks but also provides a broader understanding of art's evolution within its historical and cultural framework. A visualization of the proportion of artwork size by gender in the MoMA collection can be found here, Figure 2.

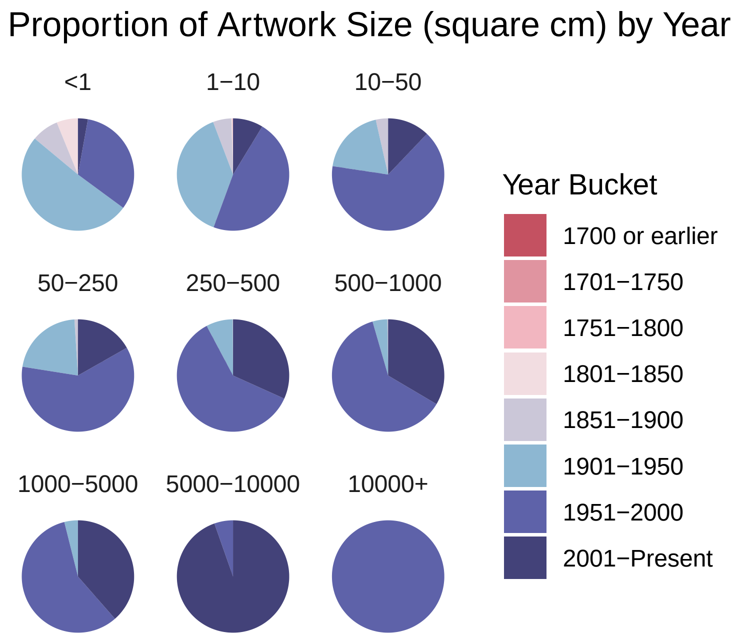

Figure 2: Proportion of Artwork Size by Year in the MoMA Collection

2.4 Analysis of Variance

The following is the results of having conducted an Analysis of Variance (ANOVA) on the dataset to examine the effects of Gender, Nationality, and Year on the variable Area.

Table 1: Anova Test for Gender, Nationality, and Year

term	df	sumsq	meansq	statistic	p.value
Gender	5	4.004497e+11	80089930627	93.352632	0
Nationality	111	4.173581e+11	3759983285	4.382628	0
Year	1	4.005067e+11	400506673726	466.829625	0
Residuals	101964	8.747787e+13	857929002	NA	NA

1. **F-statistic**: This value measures how much the means of the different groups vary compared to the variation within the groups themselves. A higher F-statistic suggests that the group means are very different in the context of the noise (variability) within each group, which implies a stronger effect of that factor.

2. **P-value**: This value indicates the probability of observing the results, or results more extreme, if the null hypothesis is true (the null hypothesis typically states that there is no effect). A very small p-value (typically less than 0.05) suggests that the observed effect is statistically significant, meaning it is unlikely to have occurred by chance.

The results of the ANOVA:

- Year has an F-statistic of 469 and a p-value of (1.01 ×10^{-103}) rounded to 0. This extremely high F-statistic, combined with an extremely low p-value, strongly suggests that the year has a very powerful effect on Area.
- Gender and Nationality also show significant effects, but the magnitude of their F-statistics (78.9 for Gender and 4.29 for Nationality) and their corresponding p-values ((8.14 ×10^{-99}) and (9.83 ×10^{-48}), respectively with both rounded to 0) indicate that while significant, their effects are less dramatic than that of Year.

2.5 Measurement

Table 2: Sample of Data

Gender	Nationality	Year	Area (cm^2)
male	Austrian	1896	8208.540
male	French	1987	1212.908
male	Austrian	1903	1090.740
male	Unknown	1980	2580.640
male	Austrian	1903	733.440

MoMA employs a well-thought out process for collecting and documenting information about the artworks in its collection. This process involves several stages, beginning with the acquisition of an artwork, during which detailed records are created. Information such as artist details, the title of the work, the year of creation, medium, dimensions, and provenance are systematically recorded. The museum's curatorial staff, alongside conservation experts, assesses and measures the artworks using standardized methods to ensure accuracy and consistency across the collection. These measurements include physical dimensions, which are particularly crucial for planning exhibitions and storage. Additionally, MoMA uses advanced digital tools to catalog and manage its collection data. This data is then maintained in a comprehensive, internally managed database that is periodically updated and reviewed to reflect new acquisitions, research, and changes in the conservation status of the artworks. This approach to data management not only supports the museum's operational needs but also contributes to scholarly research, allowing for detailed analysis and interpretation of trends and patterns within the collection.

Gender of the artist was taken directly from MoMA's records. In the case of missing data, the value "Unknown" was assigned. In the case of multiple artists working on the same artwork: if they were all of the same gender, then that gender was recorded, otherwise the walue "Multiple Artists" was recorded.

Nationality of the artist was taken directly from MoMA's records. In the case of missing data, the value "Unknown" was assigned. In the case of multiple nationalities working on the same artwork: if they were all the same, then that nationality was recorded, otherwise the walue "Multiple Nationalities" was recorded.

Year of completion was adapted from MoMA's recorded variable, Date, which represented the dates through which the artist(s) worked on the artwork. Year of completion is the maximum year within the Date interval.

Area was calculated from MoMA's records of Height and Width (both recorded in centimeters). It is important to note that this study is only interested in artwork that is for the most part 2-dimensional, therefore all artworks missing values for height or width or with values for depth were not included.

3 Model

While conducting the data analysis, a significant relationship between an artworks area the following three variables was observed: the artist's gender, the artist's nationality, and the year the artwork was completed. To explore this relationship further and predict future trends the size of artworks in the MoMA collection, a linear regression model was constructed.

The model was formalized as follows:

$$Y_i = \beta_0 + \beta_1 Gender_i + \beta_2 Nationality_i + \beta_3 Year_i \tag{1}$$

In the model:

- Y_i represents the area of an artwork created with the characteristics i.
- β_0 is the intercept, indicating the estimated area of an artwork without any known characteristics.
- β_1 is the coefficient for the gender, capturing the rate of change of an artwork's area across genders.
- β_2 is the coefficient for the nationality, capturing the rate of change of an artwork's area across nationalities.
- β_3 is the coefficient for the year, capturing the rate of change of an artwork's area as time passes.

The independent variables in the model are Gender, Nationality, and Year. The dependent variable is the Area, denoting area of the artwork in centimeters squared.

3.1 Model Justification

In this study, linear regression serves as the primary statistical tool to analyze the factors influencing artwork size within the Museum of Modern Art (MoMA) collection. The choice of linear regression is driven by its ability to provide clear, interpretable results regarding the relationship between multiple independent variables (artist gender, nationality, and year of completion) and a continuous dependent variable (artwork size). This model enables the quantification of how much each predictor variable affects the dependent variable when all other variables are held constant, offering precise insights into individual and collective impacts.

One of the most significant advantages of linear regression is its straightforward interpretability. The model's coefficients can be directly understood as the effect size of each independent variable on the dependent variable. This clarity is crucial for stakeholders at MoMA, including curators and educators, who rely on actionable insights to make informed decisions about exhibitions and acquisitions.

Linear regression not only identifies relationships between variables but also quantifies them. This quantification is essential for evaluating the scale of impact that factors like gender or nationality have on the size of artworks. Such quantitative assessments are invaluable in a museum setting, where space management and exhibit design are based on precise dimensions.

The model can adjust for multiple confounding variables simultaneously, allowing for a more accurate estimation of the effect of each variable. This aspect is particularly important in the complex environment of art collections, where many overlapping cultural, historical, and social factors can influence artistic outcomes.

The regression model's fit and predictive power are assessed through the R-squared and adjusted R-squared values, providing an estimate of how much of the variance in artwork sizes can be explained by the variables considered. While these values are typically lower in social sciences and humanities due to the complexity and multiplicity of influencing factors, they are still useful for gauging the effectiveness of the model. A significant F-statistic further validates that the model captures essential dynamics, even if the explanatory power is modest.

While linear regression offers numerous advantages, it also comes with limitations. The assumption of a linear relationship may not fully capture the complexities of art dimensions influenced by non-linear interactions among variables. Additionally, the potential for omitted variable bias exists if important influencers are not included in the model. These limitations necessitate a cautious interpretation of the results and suggest areas for further methodological enhancement in future studies.

3.2 Model Prediction Example

Following the construction and validation of the linear regression model, which estimates area of an artwork as a function of artist's gender, artist's nationality, and year of completion, it is possible to forecast the amount of space a curator needs for a future acquisition. Figure 3, juxtaposed against observed data, offers a visualization of the expected area of artwork created by Russian female artists over different time periods. Figure 4, juxtaposed against observed data, offers a visualization of the expected area of artwork created by Swedish male artists over different time periods.

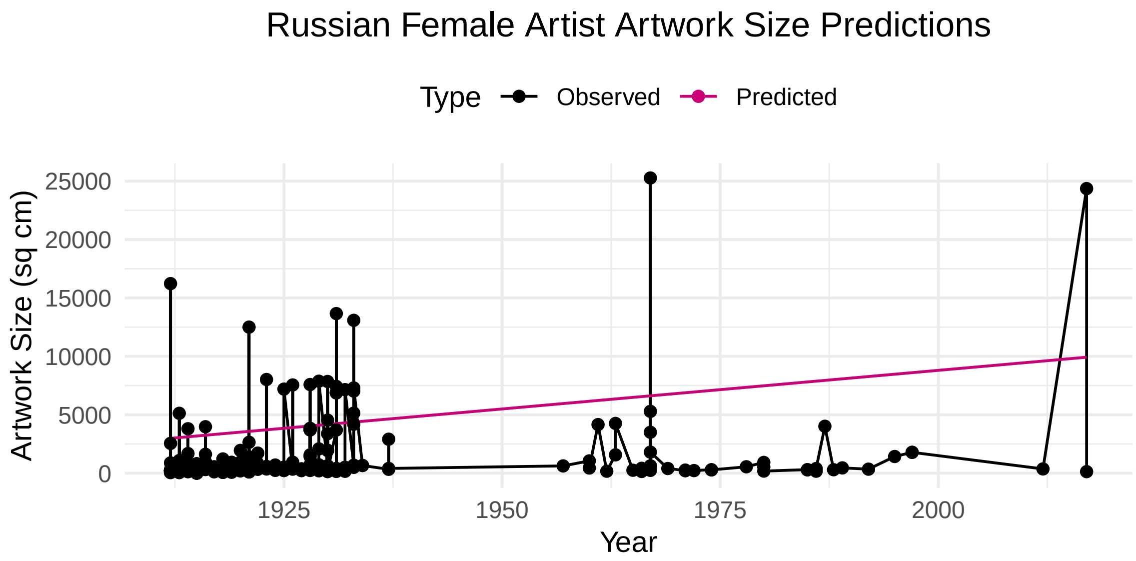

Figure 3: Russian Female Artist Predictions

Figure 3 predicts that for Russian female artists, the area of their artwork has increased over time, which is visible in the graph, especially in recent artworks.

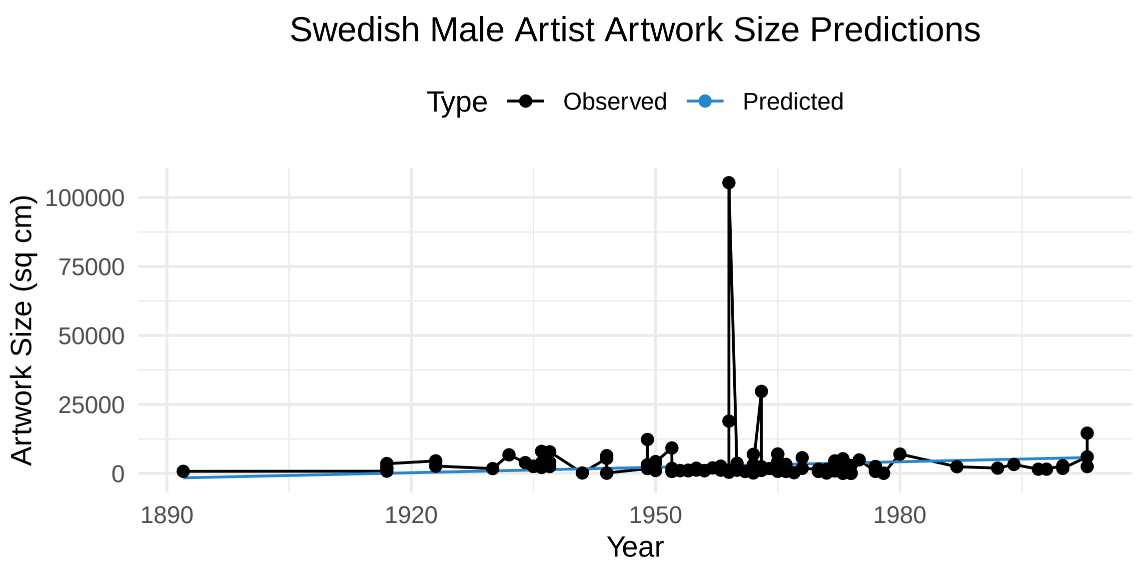

Figure 4: Swedish Male Artist Predictions

Figure 4 predicts that for Swedish male artists, the area of their artwork has increased over time, which is visible in the graph, especially in recent artworks. Notice, that this line fits more accurately to the observed trend than that of Russian female artists in Figure 3, prompting us to inquire as to what might have caused Russian female artists to have so many artworks of over 250 square meters in size which act as outliers.

3.2.1 Shiny Web App: MoMA Artwork Size Model Prediction Explorer

In our Shiny Web App, you have the flexibility to choose specific Gender and Nationality variables for detailed exploration. Examples of the types of analyses you can perform are illustrated in Figure 3 and Figure 4. Each graph displayed on the app is complemented by a comprehensive data table, which includes all graphed values clearly marked as either observed or predicted. This feature allows for an in-depth understanding of the data trends and model predictions based on your selections.

The Shiny Web App can be found here: https://www.shinyapps.io/admin/#/application/11803830:~:text=https

4 Results

4.1 Model Results and Statistical Significance

Table 5 examines the effects of gender, nationality, and year on artwork size.

The linear regression model implemented to investigate the determinants of artwork size within the Museum of Modern Art (MoMA) collection highlighted significant variations attributable to the artist's gender, nationality, and the year of completion. The results provide insights into how these factors collectively and individually influence the physical dimensions of artworks.

The model identified notable differences in artwork size associated with the gender of the artist. Male artists and artists identified as part of multiple artists groups tend to create smaller artworks compared to the baseline category (female artists), with coefficients of -4090.29 and -1582.77 respectively, both statistically significant with p-values near zero. This finding challenges traditional perceptions that associate male artists with larger canvas sizes. In contrast, gender non-conforming and unknown gender categories did not show a statistically significant impact, which may be reflective of their smaller representation in the dataset or other sociocultural factors not captured by the model.

The coefficient for the year of completion was positive (66.13), indicating that artwork sizes have generally increased over the years. This trend is statistically significant with a p-value near zero, suggesting a steady evolution in the scale of artworks possibly due to changes in artistic styles, available technologies, and evolving cultural preferences that favor more immersive art experiences.

Figure 6 visually represents the influence of an artist's nationality on the size of their artwork. Each bar in the graph corresponds to a different nationality and illustrates the magnitude and direction (positive or negative) of the estimated coefficient. Positive values indicate a positive association between the nationality and the dependent variable, suggesting that artists from these nationalities tend to have higher values in the measured variable. Conversely, negative values suggest a negative association. The use of colors in the graph helps distinguish between positive and negative estimates at a glance, aiding in quick identification of which nationalities exhibit the strongest positive or negative influences.

Significant variances were also observed across artists' nationalities. Algerian and Ghanaian artists are associated with significantly larger artworks, with coefficients of 109109.33 and 95722.19, respectively, both with p-values indicating strong statistical significance. This may suggest a cultural or regional preference for larger-scale artworks in these countries. Conversely, most other nationalities, including American, did not demonstrate a significant deviation from the baseline, which could indicate a more varied approach to artwork size that transcends national boundaries.

Figure 5 helps assess the distribution of p-values across tested nationalities or other categorical variables. This histogram plots the frequency of p-values obtained from the linear regression analysis,showing many of the tests resulted in significant or non-significant outcomes based on a conventional significance level (set at 0.05). The almost uniform distribution suggests that few nationalities have a statistically significant effect.

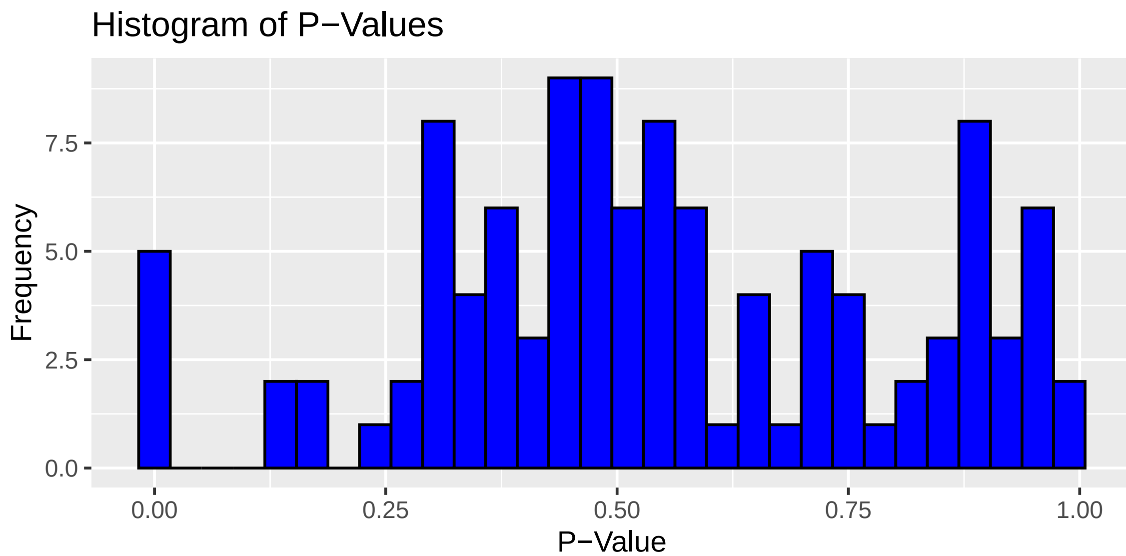

Figure 5: Histogram of P-values for Nationalities

4.2 Model Fit and Statistical Significance

- **Residual Standard Error (RSE)**: 29,290 (sq cm).
- **Degrees of Freedom**: 101,964.
- **Multiple R-squared (R^2)**: 0.01374.
- **Adjusted R-squared**: Adjusted to 0.0126.
- **F-statistic and its p-value**: The F-statistic is 12.14, with a p-value less than 2.2e-16.

The overall fit of the model, as indicated by the R-squared value (0.01374), suggests that while the model accounts for some variability in artwork size, a large portion remains unexplained by the variables included. This underscores the complexity of artistic dimensions and the potential influence of unaccounted factors. The Adjusted R-squared value (0.0126) slightly lower than the R-squared, reflects the adjustment for the number of predictors used, providing a more accurate estimate of the model's explanatory power.

The Residual Standard Error (RSE) was 29,290, suggesting that the typical prediction by the model could be off by approximately 29,290 square centimeters. Considering the large scale of some artworks, this error margin, while significant, is understandable given the vast diversity in artwork sizes.

The F-statistic (12.14) with a p-value less than 2.2e-16 strongly indicates that the model is statistically significant. This implies that the predictors, as a group, have a meaningful impact on artwork size, despite the relatively low explanatory power of individual factors.

5 Discussion

This study provides a systematic exploration of the factors influencing artwork size within the Museum of Modern Art (MoMA) collection, with significant findings regarding the impact of artist gender, nationality, and year of completion. The implications of these results are profound, offering insights into the interplay between artistic creation and broader socio-cultural contexts.

5.1 Findings

Using a linear regression model, this paper analyzed the MoMA collection dataset with the intent of revealing the differences in artwork size attributable to artist's gender, artist's nationality, and the year of completion, hoping that the results would highlight significant patterns that may reflect broader socio-cultural dynamics within the art world.

The regression results indicate significant variations in artwork size based on the gender of the artist. Notably, artworks by male artists are, on average, smaller in area compared to the baseline, with a statistically significant negative coefficient. This finding is intriguing as it challenges some conventional expectations about the representation of gender in art. This could reflect a shift where male artists are more engaged in mediums or styles that tend to produce smaller works, or it could reflect curatorial choices within institutions that collect and catalog such data. The non-significant results for gender non-conforming and non-binary artists, while not statistically enlightening, open a conversation about the representation and visibility of non-traditional gender identities in art collections. Their smaller sample sizes might contribute to the lack of statistical significance, highlighting a potential area for further focused research and acquisition efforts. Perhaps, MoMA and similar institutions could benefit from diversifying their acquisitions to include more varied gender representations, which could alter future analyses.

Nationality proved to be a less important predictor of artwork size, with some variances observed across different national origins. Algerian and Ghanaian artists, for example, tend to create significantly larger artworks. This might be reflective of specific cultural traditions or artistic practices prevalent in these countries that favor large-scale works. Perhaps large mural traditions or monumental sculptures prominent in these cultures influenced the size of contemporary artworks. On the other hand, American artists, despite their prominence in global art narratives, did not show a significant difference from the baseline, suggesting a possible diversity in styles and sizes that neutralizes any clear national tendencies. The substantial sizes of artworks by Cuban artists, significantly larger than many other nationalities, could be interpreted within the context of the nation's rich history in large-scale public art and muralism. These findings suggest that national trends in art size can be reflective of deeper cultural, historical, and perhaps political influences that shape artistic expression.

One of the most striking findings is the positive coefficient for the year, indicating that artworks have progressively increased in size over the years covered by the dataset. The increase in artwork size over the years could be linked to technological advancements that allow for larger installations, increased availability of large exhibition spaces, and perhaps a cultural trend towards more immersive art experiences. This trend is significant as it underscores the need to adapt space and logistics planning to accommodate larger artworks.

The study's results can help museum curators and art historians in planning exhibitions and expanding collections with a more nuanced understanding of trends in artwork size. By recognizing the factors influencing artwork dimensions, museums can better tailor their spaces to the evolving nature of contemporary art. Additionally, these findings can enrich discussions about representation in art collections, pushing for more inclusive practices that reflect the true diversity of artistic expression across different genders and nationalities.

5.2 Study Weaknesses

The examination of artwork size through regression analysis provides valuable insights into the influence of gender, nationality, and time on artistic output. However, this study, like any empirical investigation, encounters several limitations that may affect the interpretation and generalizability of its findings.

The primary limitations of this study stem from its reliance on linear regression, which assumes linear relationships between variables. Artistic outcomes are often the result of complex, non-linear interactions that this model may not fully capture. The model's low R-squared value indicates that other unaccounted-for variables, such as the medium of the artwork(e.g., sculpture, painting, digital), the intended installation space, or even the economic resources available to the artist significantly impact the size of the artwork but were not considered in the model. The exclusion of these factors can lead to omitted variable bias, where the effects of included variables are inaccurately estimated because they may absorb the influence of omitted factors.

Additionally, the study is constrained by the data provided by MoMA, which may not uniformly cover all artists or artworks equally, leading to potential biases in representation or emphasis on certain types or styles of art. The absence of significant findings for some categories could also be due to smaller sample sizes rather than a true lack of effect.

5.3 Next Steps

The results of this analysis not only shed light on the practical aspects of art creation and collection but also open various avenues for further research. Understanding why certain genders or nationalities produce art of different sizes could lead to deeper insights into the socio-economic and cultural contexts that influence artistic expression. Furthermore, examining these trends over time could help predict future developments in the art world.

The implications of these findings are crucial for curators, historians, and cultural theorists who seek to understand the dynamics of artistic representation and its evolution. By recognizing the patterns in artwork size related to gender, nationality, and time, the art community can strive towards more inclusive and representative collection and exhibition practices. Next steps include looking at how this paper's limitations can be responded to in future research. The use of non-linear models or machine learning techniques that can handle complex interactions and non-linear relationships could be employed could greatly improve this study. Future studies should also include additional variables that might influence artwork size, such as art medium, the socio-economic background of the artist, and the context of the artwork's creation. By addressing the limitations of this study, future research can enhance our understanding of the dynamics that influence artistic expression and further enrich the dialogue within art history and cultural studies. This approach will also bolster the robustness of the findings, ensuring that they reflect a more comprehensive view of the artistic landscape.

6 Appendix

Table 3: Sample of Gender Variable Values How Many Instances of Each Are in the Dataset

Gender	n
female	15494
gender non-conforming	2
male	80264
multiple artists	1708
non-binary	1
unknown	4613

Table 4: Sample of Nationality Variable Values

Nationality
Austrian
French
American
German
Dutch
Italian
Swedish
Multiple Nationalities
British
Japanese

Table 5: Model of Artwork Size: Coefficients of Gender, Nationality, and Year Completed

term	estimate	std.error	statistic	p.value
(Intercept)	-128316.60	9277.76	-13.83	0.00
Gendergender non-conforming	-8862.24	20713.07	-0.43	0.67
Gendermale	-4090.29	269.05	-15.20	0.00
Gendermultiple artists	-3718.32	812.61	-4.58	0.00
Gendernon-binary	16704.67	29291.67	0.57	0.57
Genderunknown	-3194.65	768.87	-4.15	0.00
NationalityAlgerian	109109.33	30093.06	3.63	0.00
NationalityAmerican	5959.26	6906.75	0.86	0.39
NationalityArgentine	4835.45	7000.11	0.69	0.49
NationalityAustralian	4287.04	7204.14	0.60	0.55

term	estimate	std.error	statistic	p.value
NationalityAustrian	5339.84	7032.55	0.76	0.45
NationalityAzerbaijani	2206.24	30093.11	0.07	0.94
NationalityBahamian	10987.94	21835.42	0.50	0.61
NationalityBelgian	4637.10	6958.05	0.67	0.51
NationalityBeninese	1124.11	18265.83	0.06	0.95
NationalityBolivian	6169.29	18268.17	0.34	0.74
NationalityBosnian	12929.32	13047.35	0.99	0.32
NationalityBrazilian	4099.87	6994.84	0.59	0.56
NationalityBritish	7198.18	6920.56	1.04	0.30
NationalityBulgarian	5206.23	11552.81	0.45	0.65
NationalityBurkinabé	1356.01	8981.36	0.15	0.88
NationalityCambodian	2225.09	12446.35	0.18	0.86
NationalityCameroonian	-247.57	10064.09	-0.02	0.98
NationalityCanadian	7474.10	7008.78	1.07	0.29
NationalityCanadian Inuit	3691.05	13047.81	0.28	0.78
NationalityCatalan	10522.96	30102.70	0.35	0.73
NationalityChilean	3899.72	7055.49	0.55	0.58
NationalityChinese	6650.01	7163.79	0.93	0.35
NationalityColombian	2129.31	7011.15	0.30	0.76
NationalityCongolese	9793.67	9408.77	1.04	0.30
NationalityCosta Rican	4430.44	7831.25	0.57	0.57
NationalityCroatian	7595.33	7219.27	1.05	0.29
NationalityCuban	40294.75	7231.37	5.57	0.00
NationalityCzech	4887.43	7022.45	0.70	0.49
NationalityCzechoslovakian	7131.83	21845.11	0.33	0.74
NationalityDanish	4783.76	7085.43	0.68	0.50
NationalityDutch	4276.95	6950.49	0.62	0.54
NationalityEcuadorian	4313.80	11553.96	0.37	0.71
NationalityEgyptian	384.37	7815.62	0.05	0.96
NationalityEmirati	-194.47	13810.19	-0.01	0.99
NationalityEstonian	4120.14	30093.20	0.14	0.89
NationalityEthiopian	4932.47	21834.93	0.23	0.82
NationalityFinnish	4525.39	7688.70	0.59	0.56
NationalityFrench	5292.48	6911.44	0.77	0.44
NationalityGeorgian	7043.70	8033.85	0.88	0.38
NationalityGerman	5385.15	6915.52	0.78	0.44
NationalityGhanaian	57844.44	16191.03	3.57	0.00
NationalityGreek	8928.83	8144.43	1.10	0.27
NationalityGuatemalan	5645.99	7742.28	0.73	0.47
NationalityHaitian	5575.50	9409.71	0.59	0.55
NationalityHungarian	4438.18	7423.01	0.60	0.55

term	estimate	std.error	statistic	p.value
NationalityIcelandic	10971.66	10437.99	1.05	0.29
NationalityIndian	9935.44	7364.47	1.35	0.18
NationalityIranian	9254.74	13808.04	0.67	0.50
NationalityIraqi	10690.41	10661.21	1.00	0.32
NationalityIrish	22151.81	8631.50	2.57	0.01
NationalityIsraeli	6946.74	7135.60	0.97	0.33
NationalityItalian	5531.34	6933.50	0.80	0.43
NationalityIvorian	903.74	7033.06	0.13	0.90
NationalityJapanese	4665.53	6938.76	0.67	0.50
NationalityKenyan	6334.22	10918.68	0.58	0.56
NationalityKorean	9399.54	7815.78	1.20	0.23
NationalityKuwaiti	2614.50	30094.80	0.09	0.93
NationalityLatvian	6541.40	7875.36	0.83	0.41
NationalityLebanese	16721.45	10915.99	1.53	0.13
NationalityLithuanian	1588.62	12446.35	0.13	0.90
NationalityLuxembourger	3728.65	8913.12	0.42	0.68
NationalityMacedonian	11204.21	30093.14	0.37	0.71
NationalityMalaysian	1428.38	21833.47	0.07	0.95
NationalityMalian	10595.08	10661.90	0.99	0.32
NationalityMexican	4879.58	6962.82	0.70	0.48
NationalityMoroccan	7137.50	10439.91	0.68	0.49
NationalityMozambican	12194.49	16191.37	0.75	0.45
NationalityMultiple Nationalities	3880.85	6946.04	0.56	0.58
NationalityNamibian	3174.81	21831.89	0.15	0.88
NationalityNative American	875.93	8225.35	0.11	0.92
NationalityNepali	9725.51	21844.23	0.45	0.66
NationalityNew Zealander	37345.65	13046.99	2.86	0.00
NationalityNicaraguan	5806.38	16195.57	0.36	0.72
NationalityNigerian	7332.65	8251.75	0.89	0.37
NationalityNorwegian	6519.93	7480.65	0.87	0.38
NationalityPakistani	-1205.50	8099.68	-0.15	0.88
NationalityPanamanian	4325.82	18272.60	0.24	0.81
NationalityPeruvian	6168.56	7453.50	0.83	0.41
NationalityPolish	5290.45	7044.30	0.75	0.45
NationalityPortuguese	2804.32	7356.58	0.38	0.70
NationalityPuerto Rican	11868.19	8349.39	1.42	0.16
NationalityRomanian	3632.48	8052.90	0.45	0.65
NationalityRussian	4866.09	6940.41	0.70	0.48
NationalitySalvadoran	19234.97	21831.83	0.88	0.38
NationalityScottish	2555.36	8251.93	0.31	0.76
NationalitySerbian	-356.95	7974.93	-0.04	0.96

term	estimate	std.error	statistic	p.value
NationalitySierra Leonean	30507.47	30093.05	1.01	0.31
NationalitySlovenian	6049.14	8734.16	0.69	0.49
NationalitySouth African	3305.65	7087.86	0.47	0.64
NationalitySouth Korean	8282.31	10661.04	0.78	0.44
NationalitySpanish	5098.03	6938.72	0.73	0.46
NationalitySri Lankan	6997.93	9309.75	0.75	0.45
NationalitySudanese	1234.90	7857.22	0.16	0.88
NationalitySwedish	5688.96	7172.72	0.79	0.43
NationalitySwiss	6092.79	6947.56	0.88	0.38
NationalitySyrian	19049.74	30093.32	0.63	0.53
NationalityTaiwanese	11793.33	18266.83	0.65	0.52
NationalityTanzanian	5141.51	30093.47	0.17	0.86
NationalityThai	13442.08	9132.90	1.47	0.14
NationalityTrinidad and Tobagonian	-1260.72	7456.51	-0.17	0.87
NationalityTunisian	1614.56	10682.57	0.15	0.88
NationalityTurkish	4972.28	8630.31	0.58	0.56
NationalityUgandan	1694.23	30093.06	0.06	0.96
NationalityUkrainian	7383.59	7816.83	0.94	0.34
NationalityUnknown	5067.82	6952.57	0.73	0.47
NationalityUruguayan	6430.56	8382.26	0.77	0.44
NationalityVenezuelan	3889.88	7043.16	0.55	0.58
NationalityVietnamese	17445.64	30102.20	0.58	0.56
NationalityWelsh	2881.68	30093.31	0.10	0.92
NationalityWest African	7869.95	14825.37	0.53	0.60
NationalityZimbabwean	10388.98	11209.88	0.93	0.35
Year	66.13	3.06	21.61	0.00

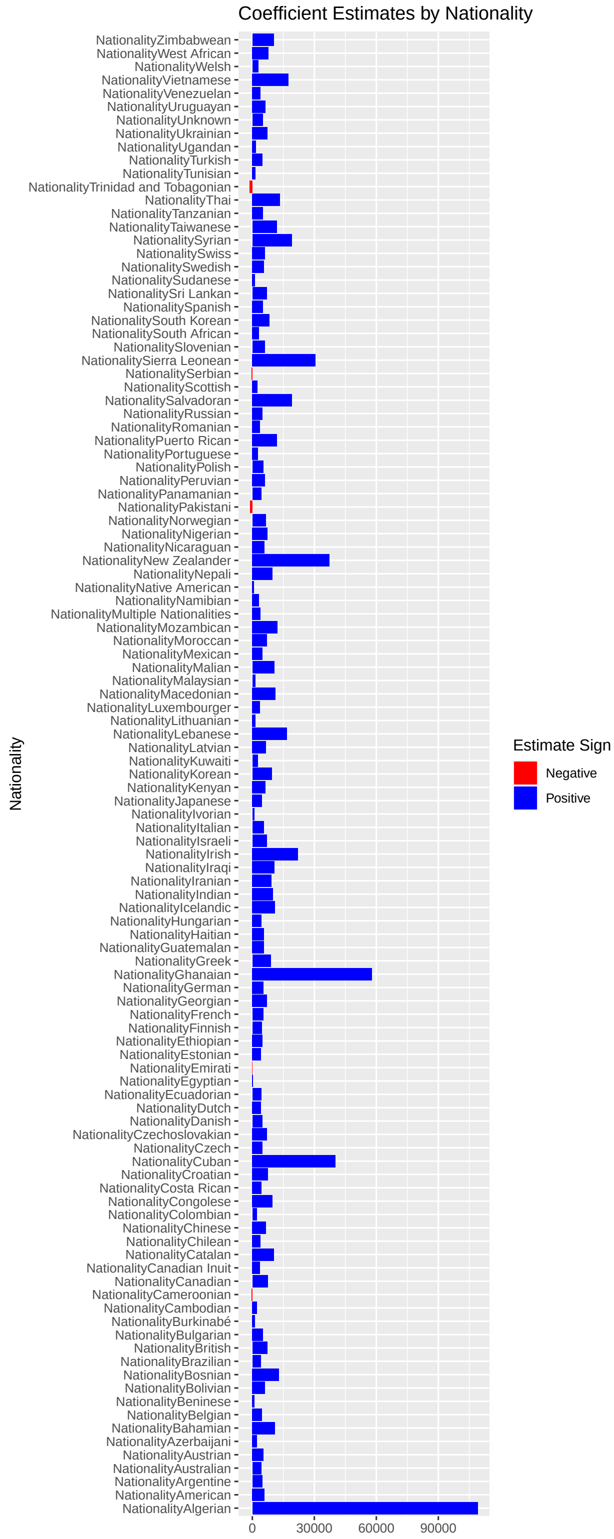

Figure 6: Coefficient Estimates by Nationality

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.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robot, O. D. 2024. *Moma Collection - Automatic Update. Zenodo*. doi:10.5281/zenodo.10977334.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 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, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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.