Museum Data in MET*

A data analysis about collection diversities 1900 to 2020

Rae Zhang

20 April 2023

This paper aims to explore the change in collections of the Metropolitan Museum of Art (MET) over the years, focusing on the change of nationality, culture, and medium of the artwork collections. The research utilizes data made available by the MET, which includes information on more than 470,000 artworks in its collection. The paper discusses the change of collections over years, strengths and weaknesses of the dataset, and also highlights how the collections reveals biases and limitations in cultural studies.

Table of contents

*Code and data are available at: https://github.com/JunweiZhang130/Museums_Data.git.

1 Introduction

Technology has all but taken over the modern world, and a big part of this digital revolution is data collection. Museums are no exception. As institutions preserve memories for our society, museums have tremendous amounts of data, which holds power and yields interesting ways to analyse artifacts, collections, objects and their creators in new ways.

This paper wants to explore the **change of collections in Metropolitan Museum of Art (MET)** over 150 years, and how the museum selects the highlights collection over years based on the artists, content, and the region of the artwork, and what kind of actions are taken as a result of this data. *We estimate that American artists would be the mainstream in the museum because of the museum's location, and the museums are accepting more artworks from artists from diverse background compare to the past.* By analyzing the dataset in the museum, we are able to determine the best way to serve the visitors and help museums better accomplish their own mission.

2 Data

2.1 Data Source and Methodology

The Metropolitan Museum of Art presents over 5,000 years of art from around the world for everyone to experience and enjoy. The Museum lives in two iconic sites in New York City—The Met Fifth Avenue and The Met Cloisters. Millions of people also take part in The Met experience online.

Since it was founded in 1870, The Met has always aspired to be more than a treasury of rare and beautiful objects. Every day, art comes alive in the Museum's galleries and through its exhibitions and events, revealing both new ideas and unexpected connections across time and across cultures(Ypsilantis et al. 2021).

The MET provides select datasets of information on more than 477,804 artworks in its Collection for unrestricted commercial and noncommercial use. To the extent possible under law, The Metropolitan Museum of Art has waived all copyright and related or neighbouring rights to this dataset using Creative Commons Zero. This work is published from: The United States Of America. The data is in CSV format.

The Museum dedicates select artworks in its collectionm, both works it believes to be in the public domain and those under copyright or other restrictions to the public domain.People can download, share, modify, and distribute the data for any purpose, including commercial and noncommercial use, free of charge and without requiring permission from the Museum.

2.2 Data Collection

The Metropolitan Museum of Art collected the data for their dataset from their internal collection management system, called the Museum System (TMS). TMS is a database that tracks information about each object in the museum's collection, including its title, artist, date, medium, dimensions, provenance, and exhibition history.

The MET's staff members enter and update this information in TMS as they catalog new acquisitions or revise existing records over years. The information in TMS is then used to create the museum's online collection database, which includes images and descriptions of each object.(Art 2021)

2.3 Data Cleaning

To begin the data cleaning process, we first reviewed the dataset and used a set of questions to guide our variable selection. Our research goal is to analyze the collection gathering in the MET, and to answer several key questions about the museum's artwork, including:

- The department with the largest number of artworks
- The cultures that are most represented in the collection
- The types of artists from different nationalities' artwork that are most prominent in the overall collection
- The representation of highlighted artifacts across all collections
- The trends in accession over time and potential factors that may impact accession

Based on these research questions, we identified several variables that we believed could help answer these questions, including **Object ID, Highlight, Timeline_Work, Department, AccessionYear, Object Name, Title, Culture, Artist Display Name**, and **Artist Nationality**.

After identifying these variables, we then proceeded to drop all missing values from the dataset that showed N/A. We also standardized the formatting of the **Country** and **Culture** variable, as the recording of this variable was done by different people and in different formats over the years. By cleaning and standardizing the data, we aimed to ensure that our subsequent data analyses would be more accurate and effective in answering our research questions.

2.4 Data Analysis

We obtained the dataset from the MET museum Github (Art 2021). Data can be downloaded from Met Github directly. We loaded in, cleaned and analyzed the data using `R` (R Core Team 2022), `dplyr` (Wickham et al. 2022), `tidyverse` (Wickham et al. 2019), `haven` (Wickham, Miller, and Smith 2022), `kableExtra` (Zhu 2021), `readr` (Wickham, Hester, and Bryan 2022), `ggplot2` (Wickham 2016), `stringr` (Wickham 2019), and `forcats` (Wickham 2021) packages.

3 Results

The Metropolitan Museum of Art (MET) has 19 departments dedicated to different collections of art objects from various cultures and time periods. Each department has a unique focus, ranging from Drawings and Prints to Islamic Art to Libraries.

We examined the total artwork amount in each department. We found that the Drawing and Painting department had the most number of objects with just over 16,7152, while the Libraries department had the smallest collection of 390 objects.

We observed that there was a significant variation in the distribution of objects across departments. For instance, the European Sculpture and Decorative Arts department had a collection of 42,934 objects, which was significantly larger than the number of objects in the Arts of Africa, Oceania, and the Americast, which had a collection of only 12,335 objects. The variation in collection size could be attributed to factors such as the popularity of different art forms, the historical significance of certain periods or cultures, and the availability of art objects for acquisition.

Our findings provide valuable insights into the distribution of artwork at the MET, shedding light on the representation of different art forms and cultures in the museum's holdings.

Table 1: MET Collections by Department

Department	Number of Collections
Drawings and Prints	167152
European Sculpture and Decorative Arts	42934
Photographs	37192
Asian Art	36817
Greek and Roman Art	33750
Costume Institute	31412
Egyptian Art	27962
The American Wing	18412
Islamic Art	15473
Modern and Contemporary Art	14241
Arms and Armor	13611
Arts of Africa, Oceania, and the Americas	12335
Medieval Art	7150
Ancient Near Eastern Art	6222
Musical Instruments	5210
European Paintings	2617
Robert Lehman Collection	2586
The Cloisters	2338
The Libraries	390

From Table 1 we can see MET has 19 departments dedicated to different collections of art objects from various cultures and time periods. Each department has a unique focus, ranging from Drawings and Prints to Islamic Art to Libraries.

We examined the total artwork amount in each department. We found that the Drawing and Painting department had the most number of objects with just over 16,7152, while the Libraries department had the smallest collection of 390 objects.

We observed that there was a significant variation in the distribution of objects across departments. For instance, the European Sculpture and Decorative Arts Department had a collection of 42,934 objects, which was significantly larger than the number of objects in the Arts of Africa, Oceania, and the Americas, which had a collection of only 12,335 objects. The variation in collection size could be attributed to factors such as the popularity of different art forms, the historical significance of certain periods or cultures, and the availability of art objects for acquisition.

Our findings provide valuable insights into the distribution of artwork at the MET, shedding light on the representation of different art forms and cultures in the museum's holdings.

In our analysis of the Metropolitan Museum of Art's (MET) dataset, we explored the representation of different artist nationalities in the museum's collections. Using the "Artist Nationality" variable, we identified the top 10 nationalities represented among the artists in the dataset(Figure 1).

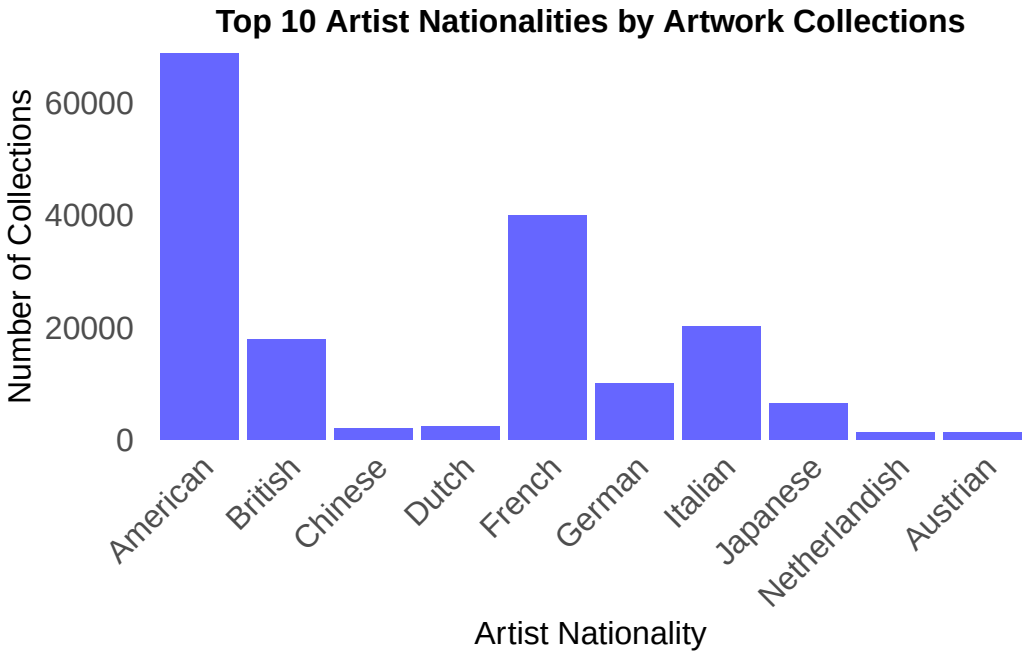


Figure 1: Top 10 Artist Nationalities by Artwork Collections

The top 10 artist nationalities in the MET's collections, ranked by the total number of artwork objects, are as follows: American - 47,797; French - 42,748; Italian - 29,910; British - 22,491; German - 18,293; Japanese - 15,097; Spanish - 12,599; Chinese - 10,322; Dutch - 9,876; Flemish - 7,303

Our analysis also revealed that the majority of the artwork in the MET's collections were created by artists from the United States, with over 60,000 objects. This was followed by French artists with over 42,000 objects, and Italian artists with over 29,000 objects.

In Figure 2, we examined the top 10 artist nationalities whose artwork was highlighted in the MET museum. We defined highlighted artwork as objects that were labelled with a "Highlight" tag in the dataset. Our analysis showed that American artists had the highest number of highlighted artwork with 387 objects, followed by French artists with 244 objects, and Italian artists with 116 objects. The other nationalities in the top 10 were German, British, Dutch, Netherlandish, Japanese, Spanish, Chinese.

Furthermore, we analyzed that even spanish artworks are not one of the top 10 large collections in the dataset, they become one of the 10 nationalities owns most highlights artworks.

Figure 3 showed that the most popular object medium in the MET's collection was prints. This was followed by photograph and drawings. Other popular mediums included book, fragment, and piece. Meanwhile, we can see Print has the largest collections among all other object mediums.

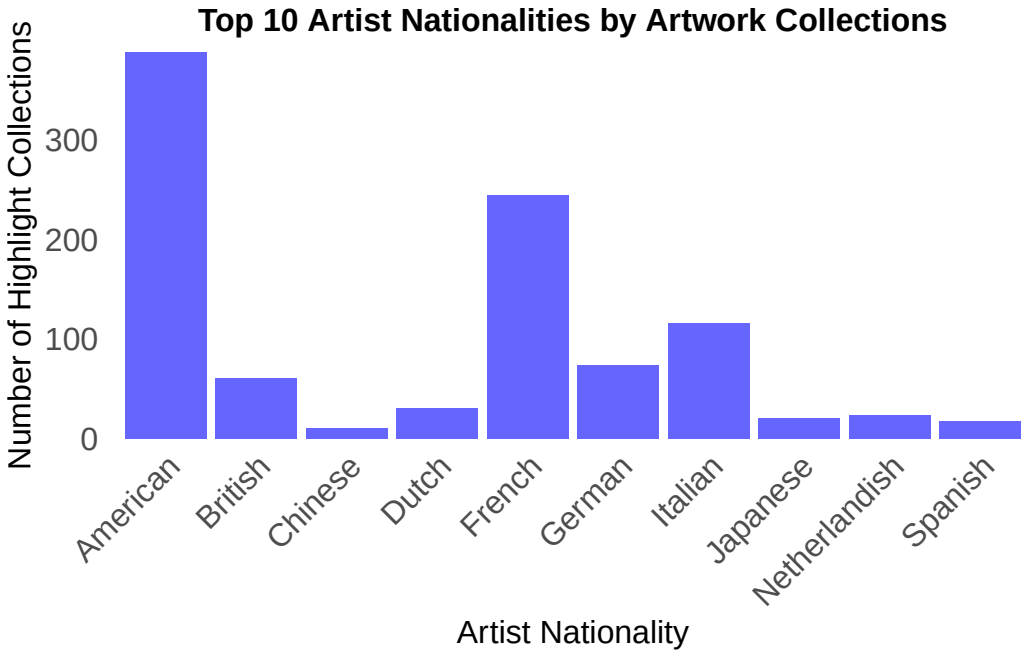


Figure 2: Top 10 Artist Nationalities by Highlight Artwork Collections

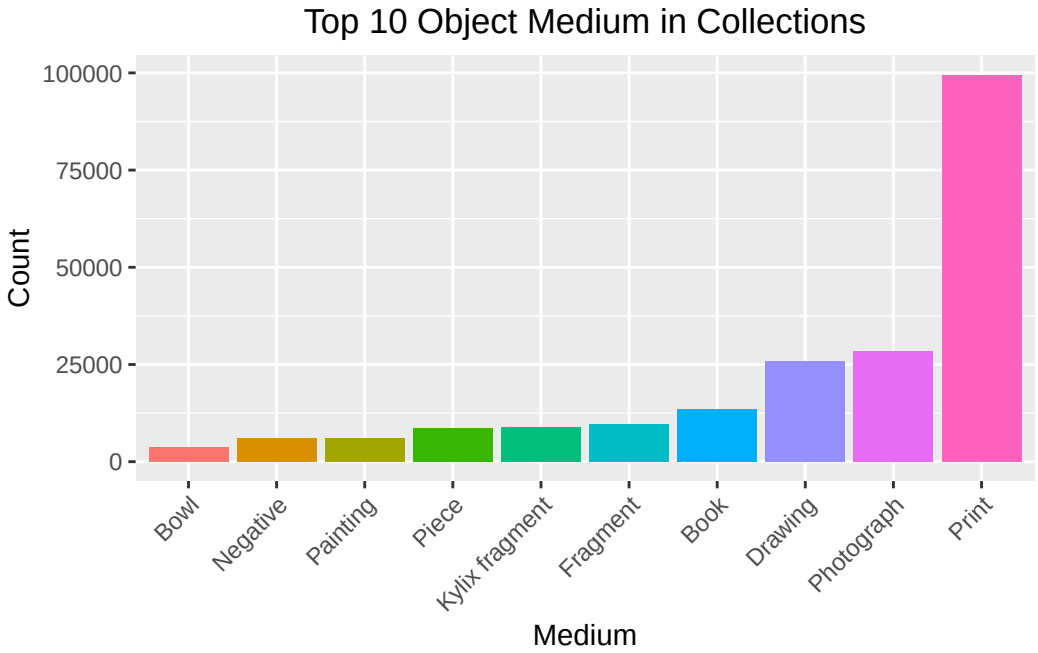


Figure 3: Top 10 Objects Mediums in Artwork Collections

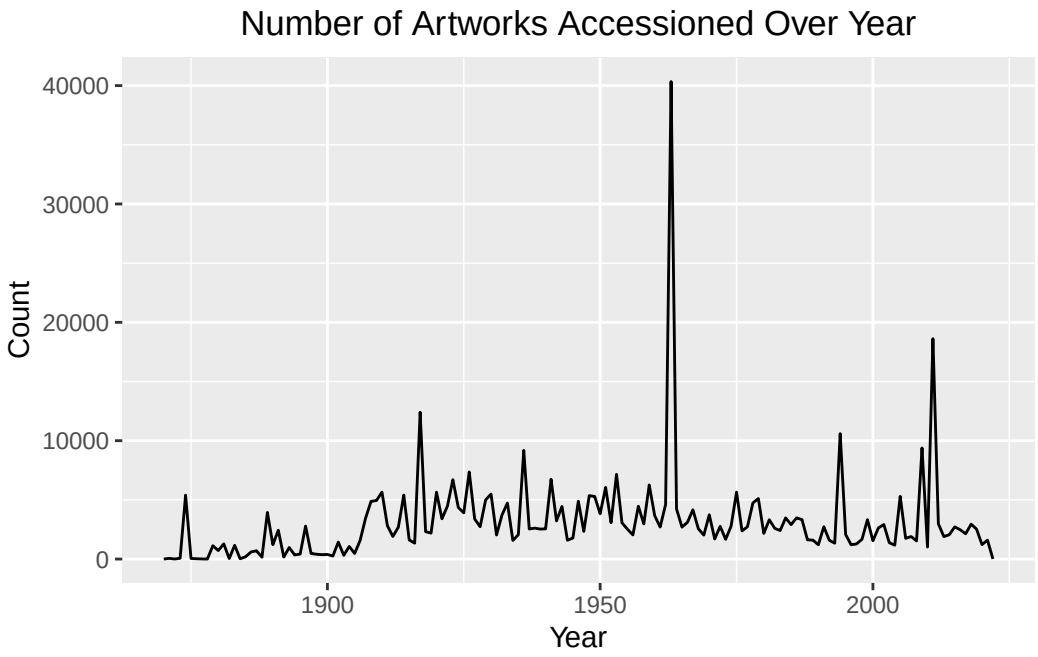


Figure 4: Number of Artworks Accessioned Over Year

The accession year of artwork in the Metropolitan Museum of Art (MET) provides valuable insights into the museum's acquisition history and patterns of collecting. We analyzed the number of artwork accessions over the years to understand the trends in the museum's collection over time in Figure 4. Our analysis revealed that the number of accessions per year has varied significantly over the course of the museum's history. For instance, the year with the highest number of accessions was 1960s. In contrast, the year with the lowest number of accessions was 2020. Furthermore, we examined the distribution of artwork accessions by decade. We found that the number of artwork access reached a peak in every 50 years. The 2010s had the second highest number of accessions, followed by the 1920s.

The Met defined "highlight artwork" as those that have been prominently featured in the museum's exhibitions or publications. Figure 5 revealed that the number of highlight artwork accessions varied from year to year. Since 1900, every year the MET has acquired a new highlight artworks. We also looked at the total number of artwork accessions per year and found that there was a general trend of increasing accessions over time. In the early 1900s and before, the number of accessions was relatively low, with fewer than 500 accessions per year. However, in the latter half of the 20th century, the number of accessions increased significantly.

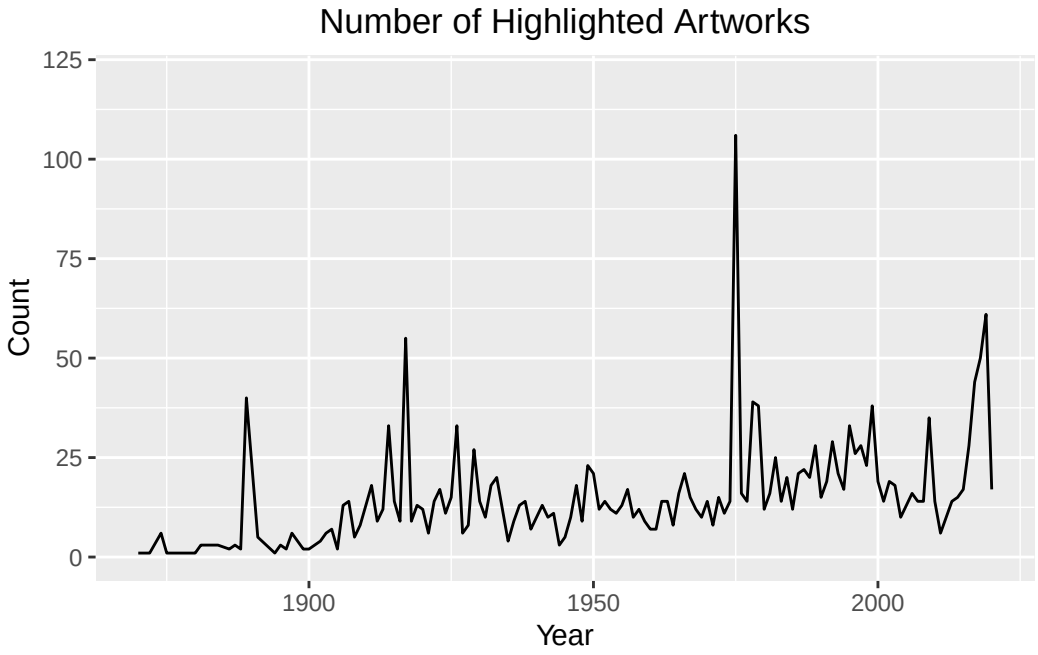


Figure 5: Number of Highlight Artworks Accessioned Over Year

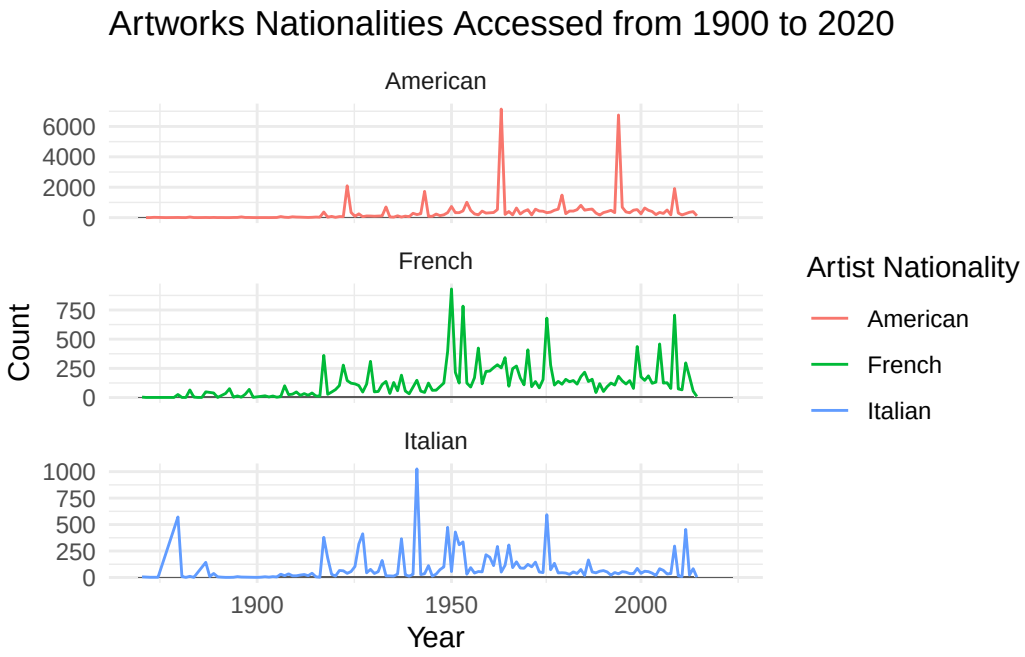
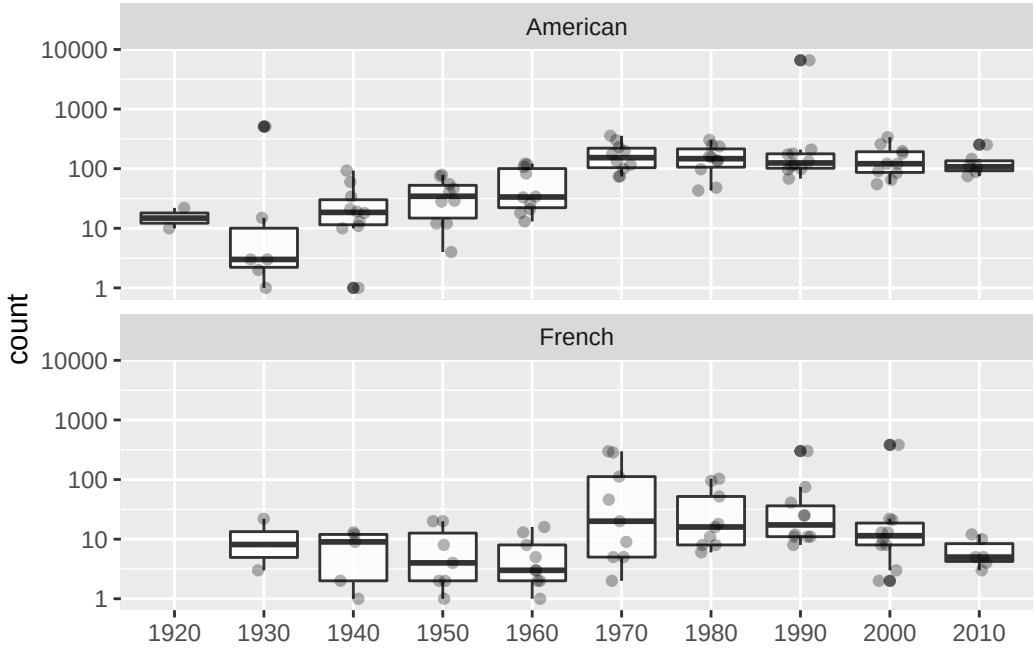


Figure 6: Artworks Nationalities Accessed from 1900 to 2020

4 Discussion

4.1 Bias and Concerns about MET Collections



5 Discussion

5.1 First discussion point

Our analysis revealed that the majority of artworks in the dataset were created by male artists, which is consistent with previous studies on gender representation in the art world.However, we also found there is a heavy load of american and Europe art represent in the gallery, and as time goes by, the variety of collections is not changing.

5.2 Collection acquiring

Furthermore, we examined the accession year data and found that the majority of artworks in the dataset were acquired by the Met in the 20th century. This could be attributed to several factors, such as the Met's focus on contemporary art in the 20th century or the increase in the number of art collectors and donors during this time period. However, it is worth noting that the dataset includes a significant number of artworks that were acquired in the 19th and 21st centuries as well, indicating that the Met has been actively acquiring artworks from different time periods.

5.3 Misrepresenting the World Culture

Misrepresentation of culture can occur when metadata highlight and accession year data are not considered carefully. Metadata highlight data can lead to a biased view of a culture if the selected artworks only represent a narrow range of themes, styles, or time periods within that culture. For example, if an art museum only highlights ancient Egyptian art that depicts pharaohs and monumental architecture, it may present a skewed and incomplete view of the complex and diverse culture of ancient Egypt. Similarly, accession year data can perpetuate stereotypes or misconceptions about a culture if the selected artworks only reflect a limited period of its history or a particular moment of cultural exchange.

Therefore, it is important for art museums to consider the context and significance of each artwork when selecting pieces for metadata highlight and accession year data. By choosing a diverse range of artworks that represent different themes, styles, and time periods, museums can provide a more accurate and nuanced representation of a culture. Additionally, museums should make an effort to include works by lesser-known or underrepresented artists, as well as artworks that challenge or subvert dominant narratives about a culture. This can help to promote a more inclusive and equitable representation of cultures in art museums.

5.4 Weaknesses and next steps

The Metropolitan Museum of Art's open dataset is a valuable resource for data scientists and researchers looking to study art, history, and culture. However, like any dataset, it has its strengths and weaknesses. The dataset from The Metropolitan Museum of Art (The Met) offers several strengths that make it a valuable resource for analysis. First, the dataset is large and diverse, containing over 470,000 records of art objects from around the world and across time periods. This diversity provides a vast collection for analysis, allowing researchers to explore various topics from different angles. Another strength is that the dataset is available to anyone for free, making it easily accessible to researchers and art enthusiasts alike. This open access encourages more people to use and explore the dataset, promoting broader participation and facilitating collaborative research. In addition, the dataset contains rich metadata about each object, such as its title, artist, date, medium, and dimensions. This information provides a wealth of data for analysis, offering researchers insights into the cultural and historical contexts of the objects and the artists who created them.

Despite these strengths, the dataset also has some limitations that must be considered. One limitation is that the data is collected and uploaded in different formats over time, which can make it challenging to compare and analyze data across different periods. Another limitation is that the dataset primarily includes text-based information, such as titles and descriptions, and images of the art objects. Other types of data, such as audio or video recordings, are not included, which may limit certain types of analysis. Moreover, missing data is a potential issue in the dataset, as some objects may lack detailed descriptions or may be missing important metadata. This limitation can limit the reliability and validity of analyses, and it is important to consider the completeness and accuracy of the data when conducting research using this dataset. Additionally, the gender of artists is significantly missing from the dataset, which may impact the analysis of gender representation in art.

Overall, while the Met's open dataset is a valuable resource for data science analysis, researchers should be aware of its limitations and potential biases when interpreting the results of their analyses.

References

Art, The Metropolitan Museum of. 2021. *The Met Collection API [Data Set]*. https://metmuseum.github.io/.

R Core Team. 2022. *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.

———. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

———. 2021. *Forcats: Tools for Working with Categorical Variables (Factors)*. https://CRAN.R-project.org/package=forcats.

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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Evan Miller, and Danny Smith. 2022. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Ypsilantis, Nikolaos-Antonios, Noa Garcia, Guangxing Han, Sarah Ibrahimi, Nanne Van Noord, and Giorgos Tolias. 2021. "The Met Dataset: Instance-Level Recognition for Artworks." In *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. https://openreview.net/forum?id=fnuAjFL7MXy.

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