FBI NATIONAL STOLEN ART FILE:

LOST ARTIFACTS AND MISSING DATA

by

ALLI ARNOLD

A master's capstone submitted to the Graduate Faculty in Data Analysis and Visualization in partial fulfillment of the requirements for the degree of Master of Sciences, The Graduate Center of the City University of New York.

2026

Approval

FBI National Stolen Art File:

Lost Artifacts and Missing Data

by

ALLI ARNOLD

This report has been read and accepted by the Graduate Faculty in Data Analysis and Visualization in satisfaction of the capstone requirement for the degree of Master of Sciences.

Approved: January 2026

Michelle A. McSweeney, PhD, Advisor

Matthew K. Gold, PhD, Executive Officer

THE GRADUATE CENTER - THE CITY UNIVERSITY OF NEW YORK

Abstract

FBI National Stolen Art File:

Lost Artifacts and Missing Data

by

ALLI ARNOLD

Advisor: Michelle A. McSweeney, PhD

This data exploration project studies the FBI's National Stolen Art File (NSAF). The NSAF is a repository for information on stolen artistic and cultural objects. By design, the objects in the NSAF should have distinguishing features and some monetary value, with a present minimum of $5,000 USD. As a public awareness tool, the objects in the NSAF should have a reasonable chance of being identified and potentially restituted in the future.

Using object data taken from the NSAF on August 25, 2024, my data exploration work includes a narrative of the data preparation process and a series of data visualizations. The most significant findings in this report are reflections on the overarching challenges and opportunities offered by the NSAF dataset.

The analysis and accompanying visualizations in this project provide examples for how further research and more consistent object data might be employed to identify physical conditions and social contexts associated with higher incidences of cultural object theft. The code repository is publicly available at: https://github.com/alliarnold/fbi_stolen_art_research.

Acknowledgments

To my advisor, Michelle A. McSweeney, PhD

> Your balance of enthusiasm, sincere curiosity, and pragmatism made it possible for me to move forward with this work when I didn't think I would. This project would not be done without you guiding me to do it.

To the faculty and staff at the Grad Center,

> Thank you for the opportunity to learn from you; for imbuing data ethics and critical thinking into every aspect of study; for your industry knowledge and historical insights, your time, and your patience.

Table of Contents

Images & Figures

Chapter 1: Introduction

This data exploration project began by extracting over 4,500 catalog entries for stolen artistic and cultural objects from the FBI's National Stolen Art File (NSAF).

When I embarked on this project, I intended to reconfigure and expand on the information in the NSAF dataset so that I could explore potential patterns and relationships between various object variables. I suspected then, and still do, that identifying overarching patterns in the physical and contextual attributes of the lost objects may help identify other objects that are most at risk for cultural loss and help to advance broader prevention and restitution efforts.

After becoming more acquainted with the NSAF dataset, I came to understand that I needed to adjust my intended project scope. The variety of object types included in the NSAF dataset requires niche areas of expertise to properly vet and to expand upon those object descriptions. Additionally, many object descriptions were missing important information and, therefore, will require more time to conduct supplementary research. Upon recognizing this challenge, I shifted from my original plan of analyzing a cross-category sample of the NSAF dataset and, instead, focused my data analysis efforts on segmented samples for the following three object categories: **Books and Ephemera**; **Drawings and Watercolors**; and **Paintings**.

The pattern analysis and accompanying visualizations created using the segmented sample data for those three object categories serve as snippet examples of what more robust object data and research could teach us about stolen artwork and cultural artifacts. However, the most interesting takeaways from my NSAF study relate to the importance of what is missing in the object descriptions and how the way that this dataset was compiled (and/or is published) is serving to extend the mysteries around these objects.

Chapter 2: The Art Crime Program & The NSAF

Since its inception in 2004, the FBI's Art Crime Program has recovered more than 20,000 objects, with a total value of over one billion dollars USD.[1] The program has three key elements: field investigations, a Rapid Deployment Art Crime Team, and the National Stolen Art File.[2]

Image 1.1: FBI's NSAF landing page, artcrimes.fbi.gov

The Rapid Deployment Art Crime team includes highly trained investigators, who work heavily with external subject matter experts domestically and abroad.[3] Art Crime Team agents are tasked with addressing art and cultural property crime cases within assigned

[1] FBI, "Art Crime," Federal Bureau of Investigations, December 26, 2025, https://www.fbi.gov/investigate/violent-crime/art-crime
[2] Ellen Ferrante, host, with Kristin Koch, guest, "The Art Crime Program," in *Inside the FBI Podcast*, produced by the FBI https://www.fbi.gov/news/podcasts/inside-the-fbi-podcast-the-art-crime-program
[3] Ellen Ferrante, host, with Kristin Koch, guest, "The Art Crime Program," in *Inside the FBI Podcast*

geographic regions in the US, as well as assisting with international art-related investigations in cooperation with foreign law enforcement and attaché offices.[4]

The Art Crime Team investigations include fraud and forgeries, antiquities trafficking, and violations of the Native American Graves Protection and Repatriation Act (NAGPR) and the Archeological Resources Protection Act (ARPA).[5] Money is a key driver for cases being brought to the FBI Art Crime Team. Art crime is a multi-billion-dollar global market, spanning theft and illicit sales.[6] The United States' robust art market, unfortunately, makes the US a prime location for the buying and selling of illegitimate art from other countries.[7] Despite significant progress being made in combating cultural property crime, it remains a growing industry, necessitating continued FBI resources.[8]

The National Stolen Art File dates further back than the Art Crime Team. It began in 1979 and has evolved from a paper filing system into a public website and, now, a mobile app.[9] The repository contains objects stolen in the United States and abroad, with each record containing various levels of information about the stolen objects.[10] For an item to be eligible to be included in the NSAF today, it must be valued at over $5,000,[11] though the previous limit was $2,000.[12] Just as important, any objects added to the registry must have some uniquely identifying characteristics, such as a distinguishing physical design with a photographic record, or a known artist's edition with the recorded edition number.[13]

Over 8,000 objects have been registered in the NSAF.[14] However, the public NSAF object count was at 4,561 at the time that I transferred the dataset from the NSAF website in August, 2024. I did not find a publicly stated reason for the difference between the number of objects purported to have been added (8,000+) and the number of objects listed (4,561).

---

[4] FBI, "Art Crime," Federal Bureau of Investigations, December 26, 2025
[5] Ibid
[6] Ibid
[7] FBI, February 9, 2015, *FBI Art Crime Team Marks 10-Year Anniversary*, FBI.gov, https://www.fbi.gov/video-repository/newss-fbi-art-crime-team-marks-10-year-anniversary/view
[8] Ellen Ferrante, host, with Kristin Koch, guest, "The Art Crime Program," in *Inside the FBI Podcast*
[9] Ibid
[10] Ibid
[11] Ibid
[12] FBI, March 13, 2013, *FBI Art Theft Program*, FBI.gov, https://www.fbi.gov/video-repository/newss-fbi-art-theft-program/view
[13] Ellen Ferrante, host, with Kristin Koch, guest, "The Art Crime Program," in *Inside the FBI Podcast*
[14] Ibid

One logical conjecture is that once objects are found or their criminal investigations are closed, the objects are removed from the public-facing NSAF site.

This theory is also supported by the presence of objects on the NSAF site that appear to predate current NSAF entry requirements. Many of the objects that appear to have older case files also have poorer quality catalogue entries, e.g., missing or poor-quality images, few identifying features, and limited artwork descriptors. The public NSAF dataset likely represents a mix of the oldest, hardest-to-close cases and the most recently opened investigations.

Given the likely disproportionate number of cold cases in the NSAF dataset, my approach to this project was destined to change from one focused on identifying patterns in the object attributes to a reflection on how this dataset was assembled and, by extension, how that assembly process impacts any intentions of studying it.

Chapter 3: Research Framework

Art history and material cultural are innately intersectional fields that require a deep appreciation for where social context meets materiality. With that in mind, my approach to working with the NSAF dataset was heavily influenced by Catherine D'Ignazio and Lauren F. Klein's *Data Feminism* and Johanna Drucker's data humanist concept of data as capta, not given, "something that is taken and constructed."[15]

In *Data Feminism*, D'Ignazio and Klein write, "The process of converting life experience into data always necessarily entails a reduction."[16] Each object description in the NSAF repository represents an abbreviated version of the long and twisty history of that object: who made it, who acquired it, how and why it was stolen and by whom, who was there to record the loss, and how did they capture that information. With so many layers of context at play, understanding the data assembly and vetting process for the information in the NSAF has been both a key challenge and requirement when cleaning and expanding upon the object descriptions in the NSAF.

Though skilled investigators, the Art Crime Team have limited resources to work backwards on long cold cases, and much of the public data provided in the NSAF does not appear to have been compiled by subject matter experts. Moreover, few of the object descriptions provide details about the context of an object's theft or disappearance. The year for when an object's theft was reported is not one of the standard catalogue points in the NSAF, nor is the object's last known location. This information is likely withheld because NSAF objects are related to criminal investigations. However, those kinds of gaps in the data serve to hinder researchers' ability to study the missing objects. At the same time, that missing information limits the general public's ability to provide potentially useful information towards restitution efforts.

Due to the limited scope of this project and the limited public information available about the individual objects, I balanced the need to interrogate the reported information with

[15] Drucker, Johanna. "Humanities Approaches to Graphical Display." Digital Humanities Quarterly, vol. 5, no. 1, 2011.
[16] D'Ignazio, C., & Klein, L. (2020). "Introduction: Why Data Science Needs Feminism" in *Data Feminism*. Retrieved from https://data-feminism.mitpress.mit.edu/

targeted areas of acceptance and trust. While I did recategorize certain objects from paintings (something unique) to prints (a multiple, and generally less valuable) as part of my data cleaning process, I did not have the privilege of distrusting artist attributions or considering if any of the works were potential fakes or forgeries. The NSAF's limiting requirement that all objects have a certain minimum financial value and the connection between insurance claims and police reports do, in the broader context of American society and wealth, indicate that the fine art objects in the NSAF are more likely to come from public institutions or private collections with enough resources to have property insurance. Therefore, the stolen objects' most recent possessors had a greater likelihood of acquiring artwork through more vetted sources. Additionally, given the FBI Art Crime Program's work investigating the creation and sale of artistic forgeries, we can hope that any object added to the NSAF after the Art Crime Program began in 2004 has had some level of further authenticity vetting.

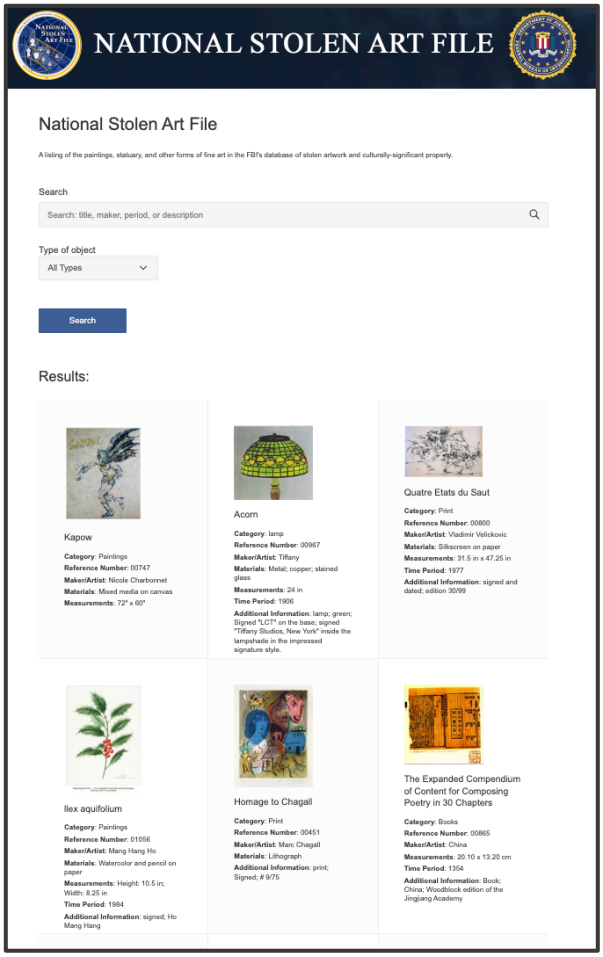
Chapter 4: Data Preparation Process

My first step in working with this NSAF data was to convert the object descriptions into the NSAF into a table format. Each object in the NSAF has a catalogue entry with one or more of the following elements:

- **Image –**an image of the artwork itself, a crime scene photo of the object in situ, or, more rarely, a drawing of the object
- **Title –**an official title or an abbreviated description of the object
- **Category –** 80+ categories were used in the NSAF site, with overlap and varying degrees of specificity, e.g., "Other – Jewelry," "Jewelry – Pendant," "Necklaces," "Doors," "Wine Cooler"
- **Reference Number –** no definition provided on the NSAF site, but context suggests this is a case number for the object's investigation; some reference numbers are associated with numerous objects
- **Maker/Artist –** use ranges from the name of the artist or manufacturer, to the artist's expected culture or the believed place of production, e.g., "Alexander Calder," "Asian," "Lalique," "France"
- **Materials –** use ranges from physical materials to production technique, e.g., "oil on canvas," "wool," "etching"
- **Measurements –** units ranged from inches, feet, pounds, ounces, and centimeters
- **Time Period –** ranged from specific years, to larger date ranges, to approximate centuries, to eras, e.g., "1979," "circa 1884," "1611-1650," "19th century," "Byzantine Empire," "Before 1400," "Contemporary," "Copyright: 1983"
- **Description –** only visible after clicking through to an object's unique page, ranged from detailed artwork descriptions, object provenance information, and further identifiers or keywords
- **Additional Information –** additional descriptive details, object identifiers, and information on object to provenance, or words

Most objects in the dataset were missing at least one of the above elements.

My first approach for extracting the object data from the NSAF was to use Python to conduct a simple web scraping exercise. After attempting several approaches in my code, I accepted that the FBI's website security would not allow the object data to be collected by any standard scraping approach. This meant that I had to navigate to the "View All" option on the NSAF search, and then copy and paste the object catalogue covers from each page of results into an Excel document. Having to use this approach resulted in a cascade of additional data cleaning steps.



	A	B	C	D	E	F
1	**Page 1**					
2						
3	Kapow					
4	Category: Paintings					
5	Reference Number: 00747					
6	Maker/Artist: Nicole Charbonnet					
7	Materials: Mixed media on canvas					
8	Measurements: 72" x 60"					
9						
10	Acorn					
11	Category: lamp					
12	Reference Number: 00967					
13	Maker/Artist: Tiffany					
14	Materials: Metal; copper; stained glass					
15	Measurements: 24 in					
16	Time Period: 1906					
17	Additional Information: lamp; green; Signed "LCT" on the base; signed "Tiffany Studi					
18						
19	Quatre Etats du Saut					
20	Category: Print					
21	Reference Number: 00800					
22	Maker/Artist: Vladimir Velickovic					
23	Materials: Silkscreen on paper					
24	Measurements: 31.5 in x 47.25 in					
25	Time Period: 1977					
26	Additional Information: signed and dated; edition 30/99					
27						
28	Ilex aquifolium					
29	Category: Paintings					
30	Reference Number: 01056					
31	Maker/Artist: Mang Hang Ho					
32	Materials: Watercolor and pencil on paper					
33	Measurements: Height: 10.5 in; Width: 8.25 in					
34	Time Period: 1984					
35	Additional Information: signed; Ho Mang Hang					
36						
37	Homage to Chagall					
38	Category: Print					
39	Reference Number: 00451					
40	Maker/Artist: Marc Chagall					
41	Materials: Lithograph					
42	Additional Information: print; Signed; # 9/75					

Images 4.1 & 4.2: Screenshots of the "view all" results on the NSAF search and the initial copy-paste data in Excel

The copy-and-paste version of the object data did not have one row per object, and the object attributes had their attribute name with the attribute content. For this reason, I used Excel's TEXTSPLIT function to separate the object variable labels (ex, "Materials:") from the subsequent information. I then transposed each object entry from columns into rows. Additionally, because each object did not contain the same number of elements, this step

required additional hours to redistribute each object datapoint into consistently laid out columns before the dataset could be trusted to be in an appropriate structure for sampling.

	A	B	C	D	E	F	G	H
1	**Title**	**Category**	**Ref_Num**	**Maker_Artist**	**Materials**	**Measurements**	**Time Period**	**Additional Information**
2	Kapow	Paintings	00747	Nicole Charbonnet	Mixed media on canvas	72" x 60"		
3	Acorn	lamp	00967	Tiffany	Metal; copper; stained gla	24 in	1906	lamp; green; Signed "LCT" o
4	Quatre Etats du Saut	Print	00800	Vladimir Velickovic	Silkscreen on paper	31.5 in x 47.25 in	1977	signed and dated; edition 30
5	Ilex aquifolium	Paintings	01056	Mang Hang Ho	Watercolor and pencil on	Height: 10.5 in; Width:	1984	signed; Ho Mang Hang
6	Homage to Chagall	Print	00451	Marc Chagall	Lithograph			print; Signed; # 9/75

Images 4.3: Screenshot of the restructured data in Excel

Once I had data in the necessary format, I did an initial check of the number of categories used. Given the uneven distribution of different types of objects within the NSAF, I wanted to create a stratified sample for my analysis. I used Excel's Pivot Table capacity to quickly count how many objects fell under each category and then began to consolidate the categories. I found over 80 categories being used, over 100 if counting categories that were technically the same but written differently (ex, "other,jewelry" vs "Other – Jewelry"). I began combining the categories that were essentially the same, and then considered which overlapping categories could be merged, such as "Dolls and Figurines" and "Figurines" or "Necklaces" and "Jewelry." Following this initial consolidation, there remained quite a few categories with fewer than ten objects in them, so I made the decision to further combine categories to reduce the risk of any one object greatly skewing category-level analysis later. Using more common art and antiques cataloguing subject groupings as my frame of reference, I was able to narrow the list of categories down to 27, allowing me to move forward with sampling.

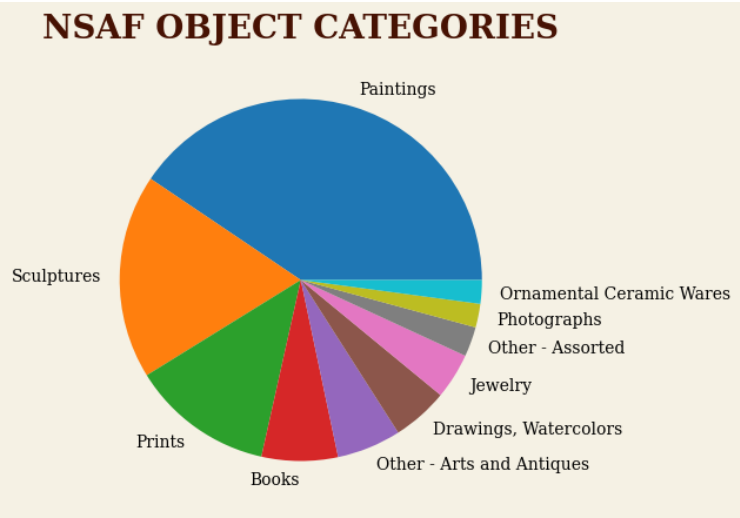


Figure 4.4: Pie chart showing the distribution of object categories in the NSAF dataset

Using Python with the Pandas library in a Jupyter Notebook, I loaded the data into a Pandas DataFrame, confirmed the distribution of the objects across categories, and then pulled a stratified sample of 10% of the NSAF data, with a matching distribution of objects by category.

The pie chart on the left shows the object categories for the full NSAF

dataset, with a unique color for each of the final 27 categories.

As pie charts lose readability when split into so many segments, I have also included the waffle chart below. The waffle shows the same distribution with a reduced color palette. The three object categories in my final sample are each represented with unique colors (green, yellow, and violet). The remaining object categories are shown in either light green (for categories I hope to study in the future) or light cream (for categories that would require outside expertise to properly research).

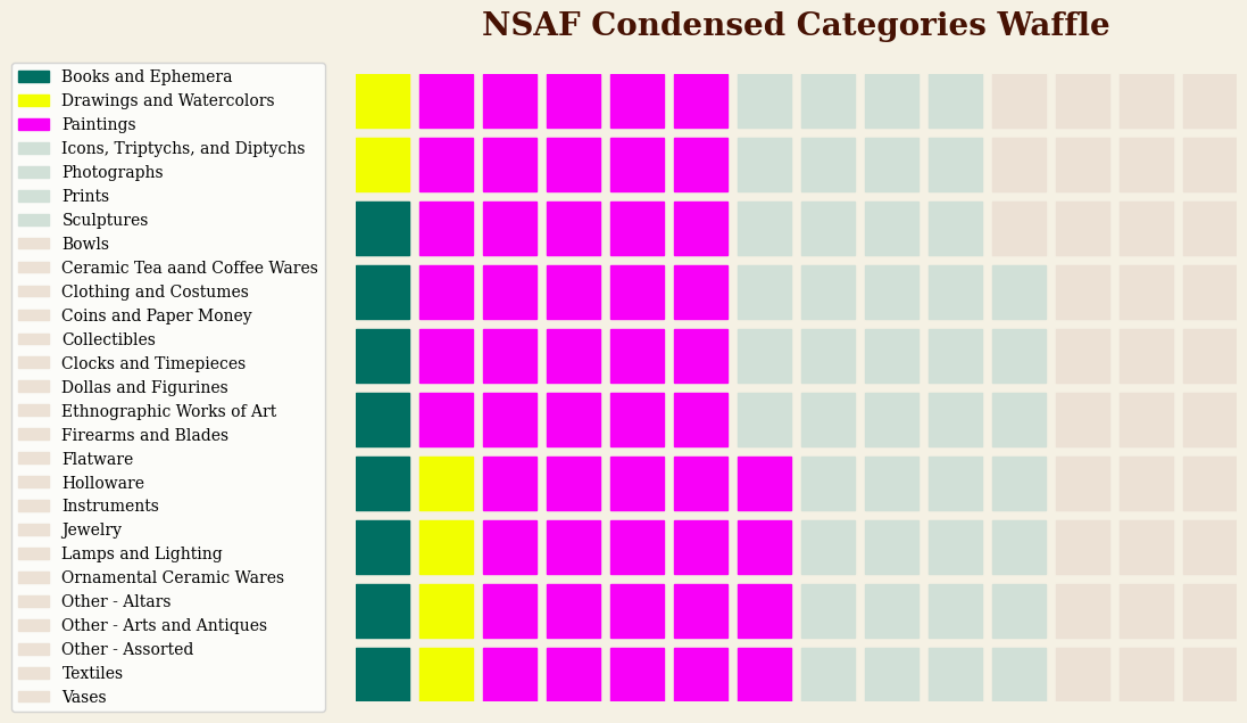


Figure 4.5: Waffle chart showing the distribution of object categories in the NSAF dataset, with condensed pallet.

As illustrated in the waffle, the three categories in my final sample, Books and Ephemera; Drawings and Watercolors, and Paintings, comprise 46.22% of the full NSAF dataset, or 6.01%, 4.41%, and 35.8% respectively.

While I was extracting the data from the NSAF website and restructuring it, I became more familiar with the scope of the objects. This allowed me to consider what variables would be more interesting to explore for prospective patterns in the objects. I set out to expand upon the available attributes for the 501 objects in my sample. I landed on the following list of variables for each object:

- **Title** – existing
- **Category** – existing
- **Unique-Multiple** – new, differentiating unique works or editions
- **Image** – new, note if an object record had an image (Y/N)
- **Reference Number** – existing
- **Maker_Artist** – refined to only list artist/maker name
- **Artist_2** – new, secondary artist or a manufacturer name, ex, books and artists' editions
- **Artist_Culture** – new, first artist's culture
- **Artist_2_Culture** – new, second artist's culture
- **Manufacture_Country** – new, country where an object was made
- **Materials** – existing
- **Material_Primary-Base** – new, object's primary material
- **Material_Secondary** – new, object's secondary material
- **Material_Regulated-Endangered** – new, if object included regulated material (Y/N)
- **Measurements** – existing
- **Height-Length** – new, split measurements by type
- **Width** – new, split measurements by type
- **Diameter_Depth** – new, split measurements by type
- **Largest Dimension** – new, split measurements by type
- **Weight** – new, split measurements by type
- **Time Period** – existing
- **TP_Start** – new, first year of creation/design time range
- **TP_End** – new, last year of creation/design time range
- **Manufacture_Start** – new, first year of manufacture time range
- **Manufacture_End** – new, last year of manufacture time range
- **Color-Primary** – new, predominant object color
- **Additional Information** – existing, combined text from "Description" and "Additional Information" variable labels

Initially, I had one more binary variable to flag if an object was created in a context of conflict, e.g., if there was a known war occurring in the country where it was made when it was made. With the numerous wide time frames and varying degrees of conflict

documentation, I had to let go of this variable. I accepted this in part because I believed that the more useful information would be if the object was first recognized as stolen during a conflict. Unfortunately, as already mentioned, the date and place of the reported thefts are not provided for the NSAF objects.

It was at this stage in the project that I reluctantly accepted that I would not be able to clean a comprehensive cross-category sample. I had begun my object research at the top of the sample, which was ordered alphabetically by category. This approach meant that I started my cleaning efforts with objects that required significant niche expertise in order to study them. I have a Master of Arts in the Connoisseurship of Fine & Decorative Art, and I have worked for several auction houses, most recently serving as a generalist specialist and head cataloguer. While I can research and identify a variety of fine art and antiques through in-person physical inspection, I do not have the subject matter expertise to identify and vet time periods for ancient coins without extensive supplemental research. I lost hours researching individual objects outside of my expertise before I accepted that I would not be able to complete this project within the required timeline if I continued on that path. I recognized enough inaccuracies in the object descriptions that I knew that including objects from categories I could not clean without extensive research would result in faulty findings. Ultimately, I shifted my focus to cleaning an initial set of three categories where I had a high level of existing knowledge.

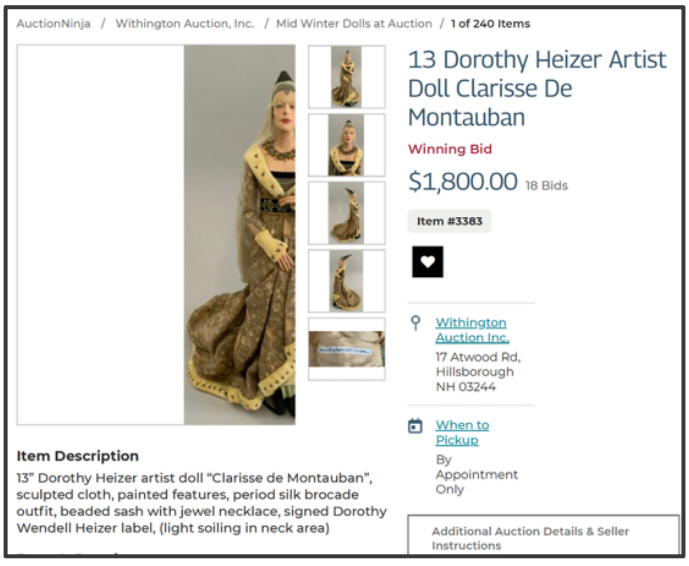
Once I shifted focus, I was able to move more quickly and more accurately, though there were certain variables that slowed me down more than others. Identifying the country where objects were made, for example, demanded more time to pin down when dealing with artists who moved regularly throughout their careers. I was more confident in recognizing mischaracterizations and time period errors. Many object time periods were listed with centuries misapplied. For example, many works that were associated with artists who lived in the 1800s had "18th century" listed as their time period. I also found many watercolors on paper erroneously categorized as paintings, when they would more traditionally fall under the Drawings and Watercolors category. I also found prints described as paintings, despite the artwork image being of a print and the object details listing information on the publisher. Recategorizing certain objects resulted in shifted sample sizes for the three categories I completed. However, I am hopeful that those shifting sample sizes are reflective of a more accurate breakdown of the objects in the NSAF. Objects in the Books and Ephemera category often lacked dimensions and often did not specify the lost volume's pressing.

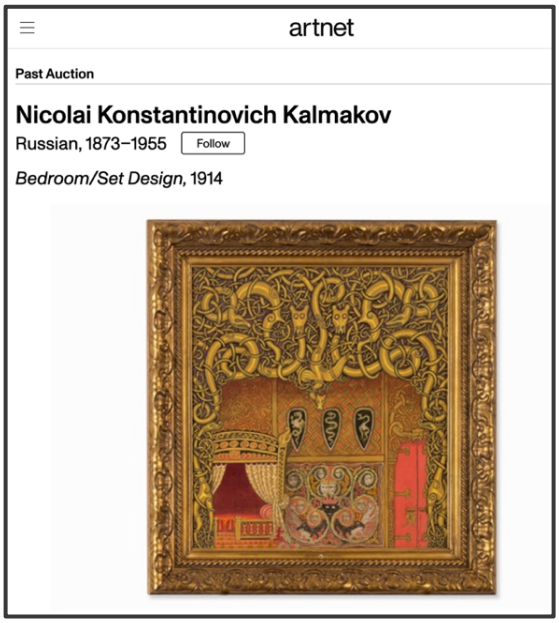
Fortunately, most books with modern printing records were produced in standardized sizes, so I was able to find their dimensions by finding information on the relevant edition through libraries and rare book sales listings.

While researching these objects, I found numerous echoes of their continued existence out in the world. Early on, I found what appeared to be the same Nisga'a carved mask, with matching provenance, on a webpage for the Smithsonian Museum of the American Indian.

Image 4.6: Nisga'a carved mask from the Smithsonian Museum of the American Indian (above)

I also came across an auction record for what appears to be the same, supposedly unique, handmade silk doll by Dorothy Heizer. Later on, I found an auction record for a set design by Nicolai Konstantinovich Kalmakov when I was researching the painting's time period. More surprisingly, the images from that artwork's auction record suggest that it was offered at auction two years in a row, in 2014 and again in November 2015. As the NSAF





Images 4.7 & 4.8: Screenshot from auction website with a record for the sale of Dorothy Heizer doll (left); screenshot of Nicolai Konstantinovich Kalmakov set design auction record (right)

does not divulge when the objects in the database were reported stolen, it is hard to say if any of these matches are clues to where the objects are today, or if what I found is just object provenance predating the theft.

If my research uncovered information that pre-dates these objects' theft, then the absence of these details from the NSAF is a telling example of how having the works submitted to the NSAF by local police without specialized training results in key gaps in object cataloguing. If my research did find clues about where these objects have gone, the only way for me to know is to try a tip to the FBI.

Chapter 5. Data Analysis, Visualizations, & Findings

I began my data analysis with a more targeted sample, focusing on only three object categories from the NSAF dataset. For this reason, several of the object attributes that I originally planned to study proved less useful for these specific categories. For example, regulated materials are found more often in the decorative arts (such as rosewood or ivory in furniture) or cultural artifacts (e.g., objects that are wrongfully removed from Indigenous communities). For my three categories, the most promising attributes to study were: object size, artist culture and country of manufacture, predominant object color, and object age/period of manufacture.

To analyze the sample, I used Python in Jupyter Notebooks with Numpy, Pandas, Matplotlib, and Seaborn libraries.

A. Size

I used the largest dimension available as the stand-in variable for object size. With this in place, the median size for objects in my targeted sample is 24 inches (60.96 centimeters) with a standard deviation of 17.4 inches.

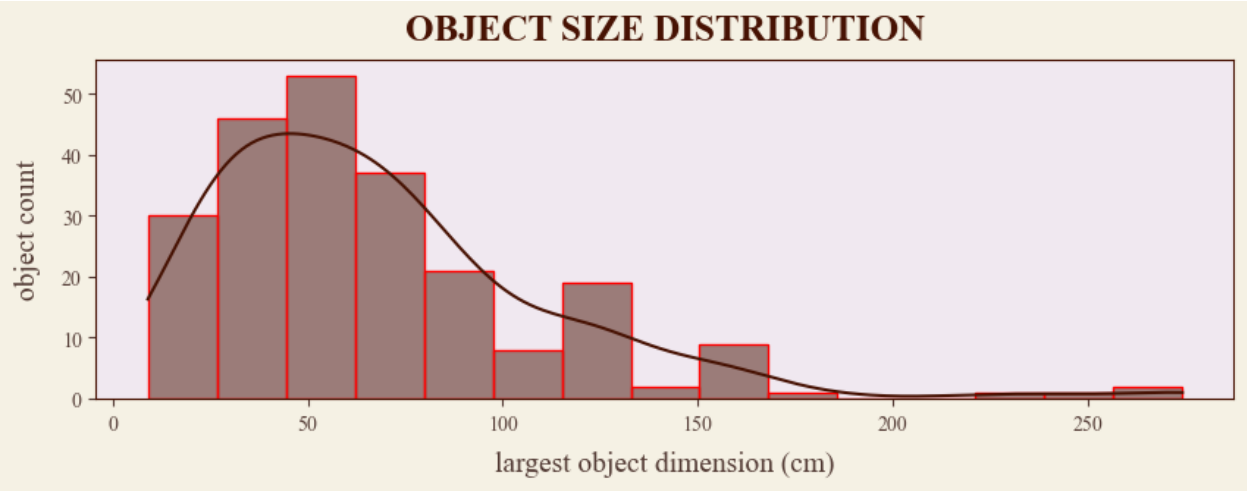


Figure 5.1: Histogram of object size distribution for all three sample categories (above)

As my sample only includes three categories, the size distributions are likely more informative on a category level.

- The Books and Ephemera sample has a right-skewed size distribution, with a mean of 10.3 inches and a median of 9.25 inches. The largest object in that category sample is 17.7 inches.
- The Drawing and Watercolors sample has a more normal size distribution, with a mean of 18.4 inches and a median of 18.7 inches. The largest object in that category sample is 30 inches.
- The Paintings sample also has a right-skewed size distribution, with a mean of 31.4 inches and a median of 28.5 inches. The largest artwork in that category sample is an outlier at 108 inches.

We can compare these three categories' largest object dimensions using a set of box and raincloud plots.

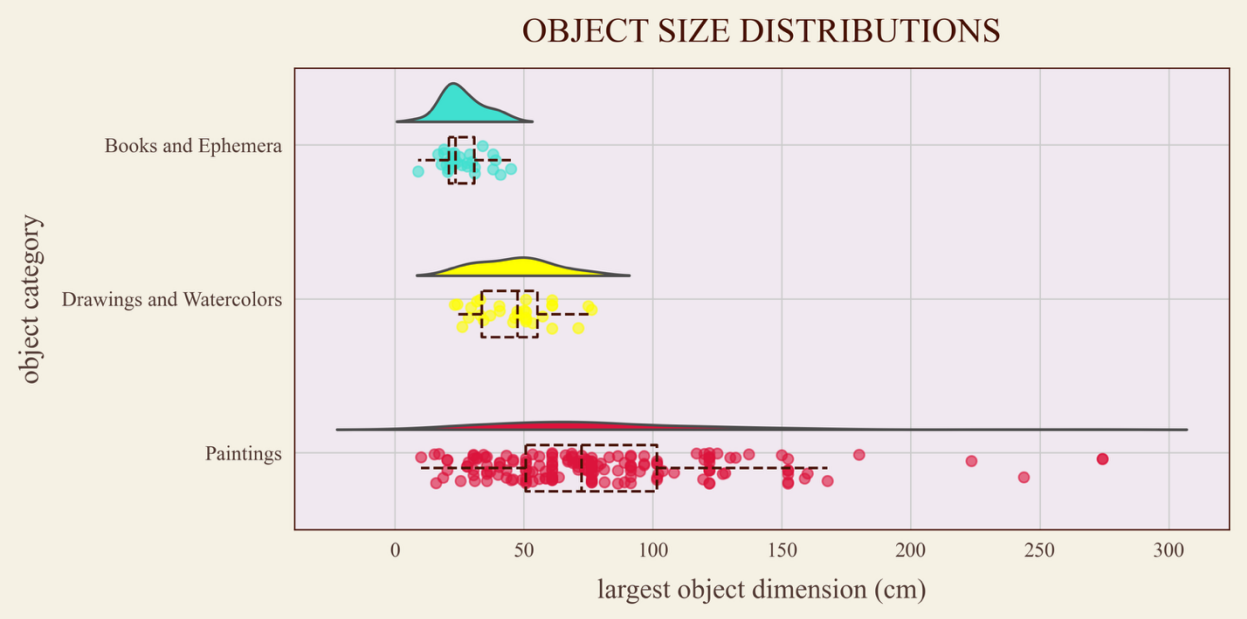


Figure 5.2: Box and raincloud plots of object size distribution, split by object category (above)

With the sample objects split by category, we can better see how the size distribution changes by category. Additionally, the number of large outliers in the Paintings category becomes more obvious.

Despite that, if we look at a histogram with the object counts overlaid with one another by category, it is clear that all objects veer towards smaller sizes, regardless of category, with a

significant drop in object counts after 100 centimeters (37.4 inches). Studying the object counts with stacked bars or by density also highlights how the object size distribution veers smaller.

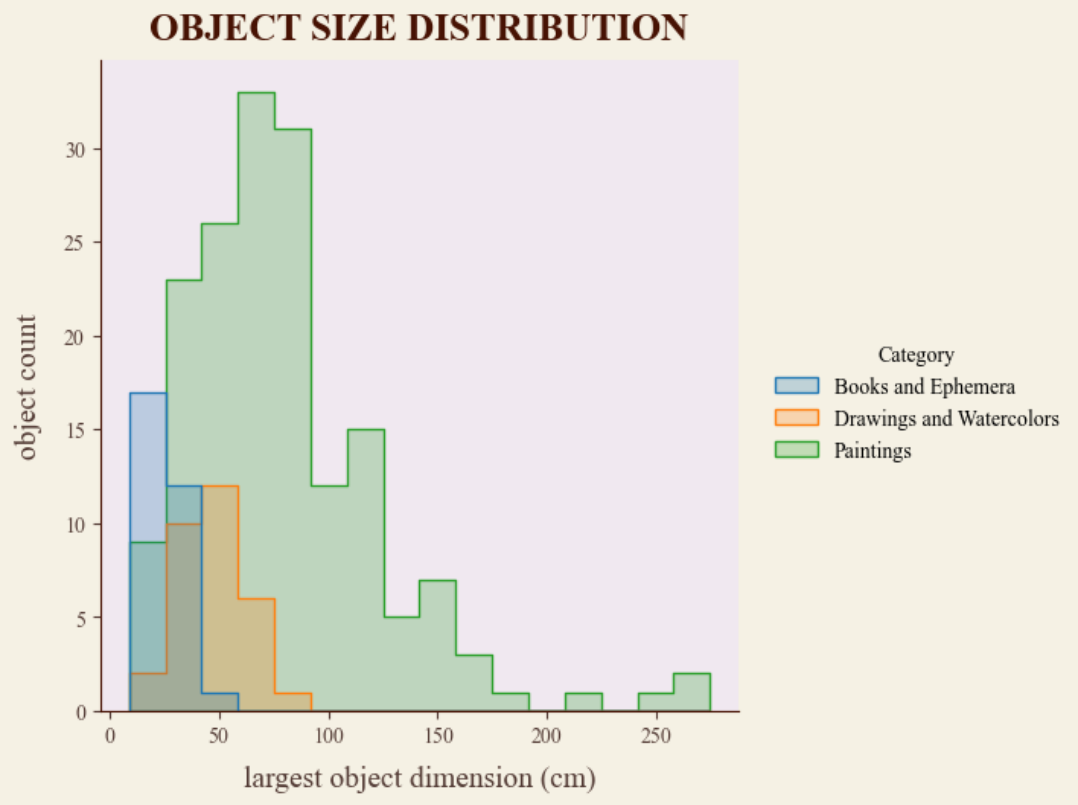
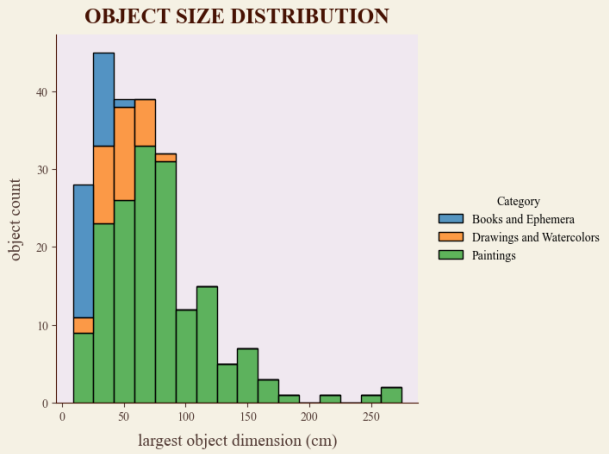
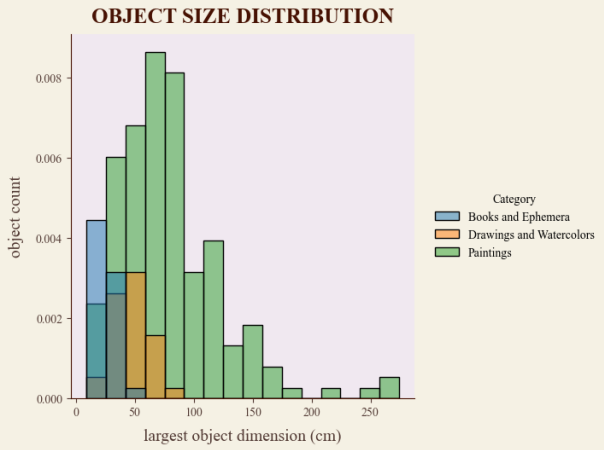


Figure 5.3: Histogram chart of object size distribution, split and overlaid by category (above)

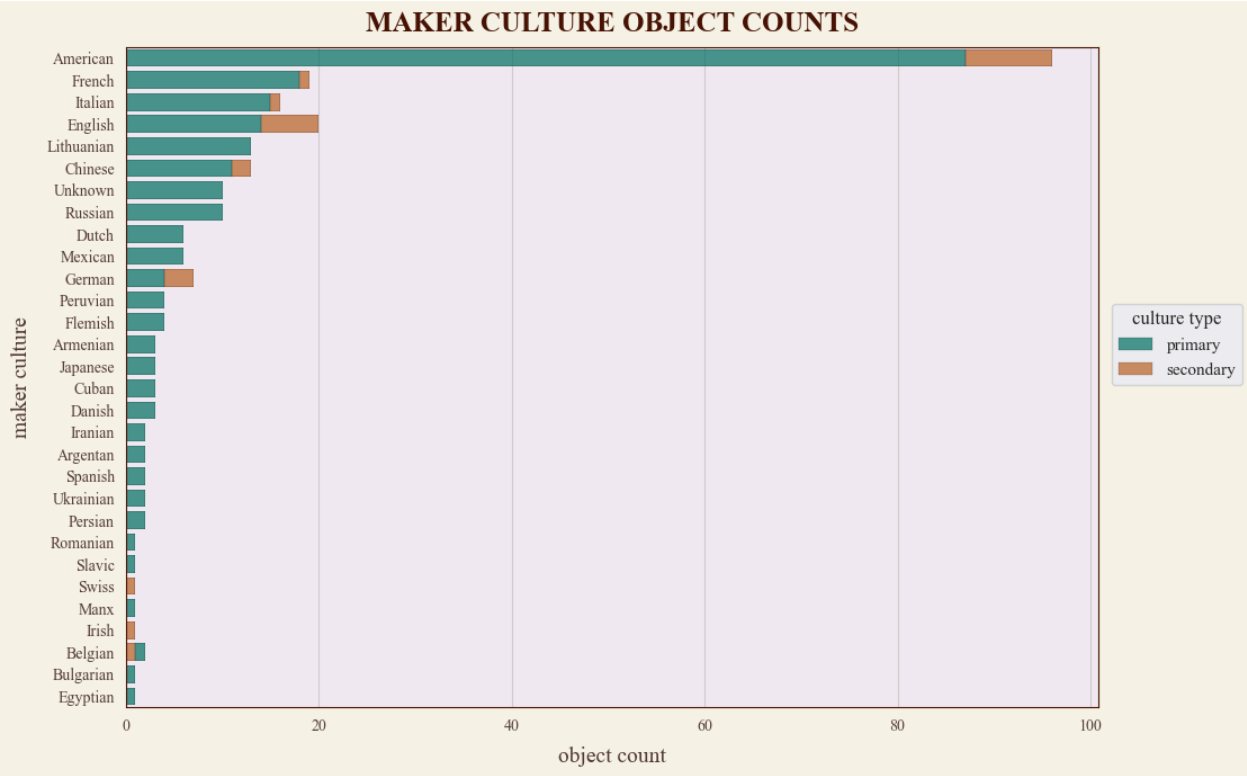




Figures 5.4 & 5.5: Histogram charts of object size distributions, stacked by category (left) and shown as object densities stacked by category (right)

As the paintings category jumps out as having larger objects with more outliers, I wanted to check to see if there might be a compounding variable related to that category. Paintings on canvas can be removed from their stretchers or cut out of frames. In this sense, paintings on canvas can be made smaller (more portable) without significantly altering their market value. When I sliced the largest dimension variable by material, looking at paintings on canvas vs. paintings on wood, I did find that canvas paintings were responsible for many of the larger outliers in the category. The median largest dimension for objects in the paintings category is 72.4 centimeters, which is similar to the median largest dimension for paintings on canvas (74.3 centimeters), but 11.4 centimeters larger than the median largest dimension for paintings on wood panels (60.98 centimeters).

B. Artist/Maker Culture



Figures 5.6: Horizontal stacked bar chart with object maker cultures, stacked by primary and secondary culture (above)

When analyzed across all three categories, American is the most frequently represented culture among this sample's artists, makers, and authors, with 87 objects or 37.8% of the sample. This makes sense given that this is, first and foremost, an American database.

American was followed by French, Italian, English, and Lithuanian. It is expected that we see a high number of French artists given the migration of artists and artworks between France and the US between and directly after the two world wars. Italian artists are highly represented in collections of Old Masters artworks. It is also reasonable that we see quite a few works by English artists and authors, given the relationship between our two countries and the importance of British publishing houses for consumers of English-language books.

The high frequency of objects made by Lithuanian artists is due to a 30-year-old crime skewing the data. In 1994, a storage unit belonging to the Lithuanian-American artist Arbit Blatas was robbed, with over 200 paintings taken.[17] Thirteen of those lost works ended up in my sample.

Most of the objects in these three sampled categories only have one associated artist, maker, or author. The objects with secondary makers are primarily in the books and ephemera category, which is why we see secondary cultures most frequently associated with cultures important to early and modern English-language and European publishing. If I move into preparing the sample of the objects in the prints category, I would also expect to see more objects with a secondary artist/maker in the form of the printers and publishers.

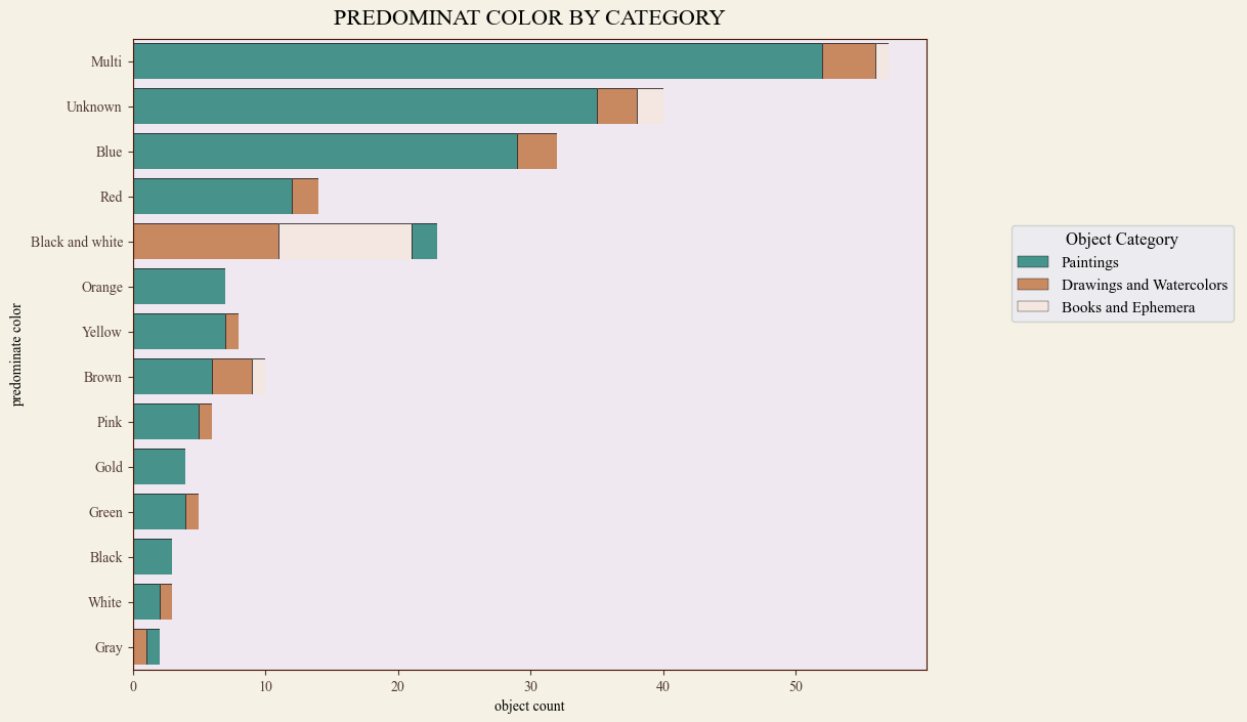
C. Predominant Color

I was curious about how aesthetic desirability related to incidents of theft. There is no universal scale for aesthetic desirability. However, studies have found that blue and red paintings sell for more than other colors at auction.[18] Anecdotally, gallerists and publishers will tell you that blue works are particularly strong.

---

[17] New York Times. *'Three Penny Opera' Paintings Born Anew*. New York Times, January 2001. Retrieved from https://www.nytimes.com/2001/01/20/arts/threepenny-opera-paintings-born-anew.html
[18] Nate Freeman. *Blue and red paintings make the most money at auction according to a study.* Artnet News, March 2019. Retrieved from https://www.artsy.net/article/artsy-editorial-blue-red-paintings-money-auction-study

With this in mind, I labeled the objects in the samples with their predominant color. Works without images or other relevant information available were coded as "Unknown" for their predominant color. Works that did not easily jump out as having a predominant color by the human eye were coded as "Multi."



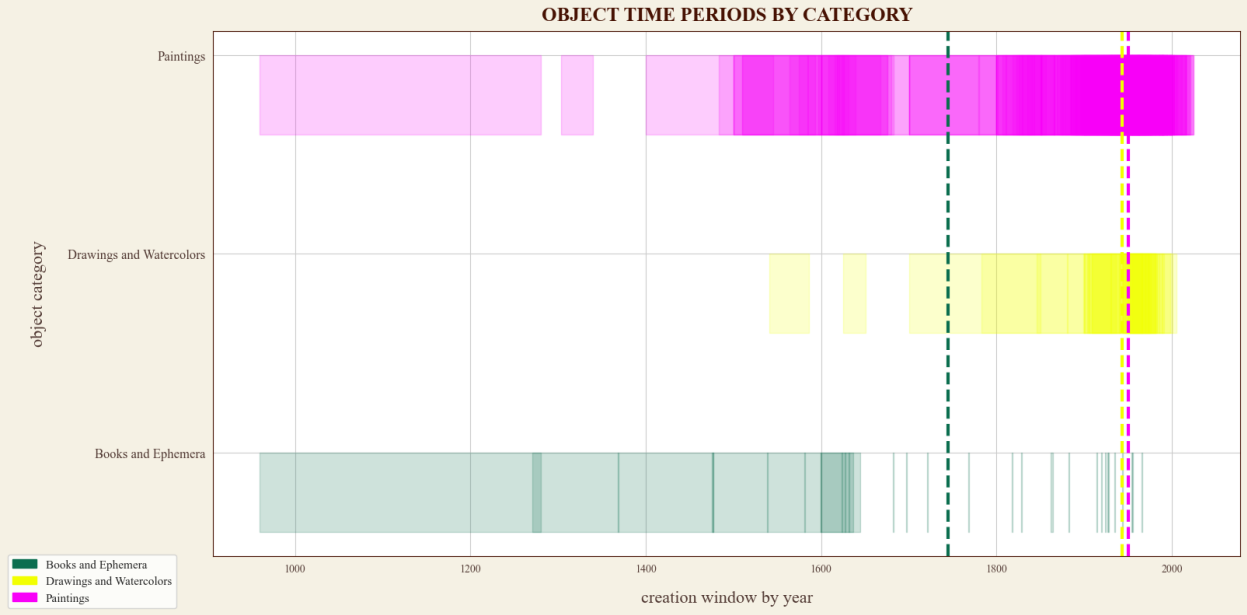
Figures 5.7: Horizontal bar chart with objects' predominant colors, stacked by object category (above)

Black and white was predominant for the books and ephemera category, as well as in the drawings and watercolors category. Just as blue and red are key colors for auction sales, blue and red were the most popular predominant colors for paintings other than those that fell under "multi". This suggests that just as there is a correlation between artwork color and achieved sales, there may be a correlation between artwork color and theft.

D. Time Period

I was also interested to see if the period when an object was made had any relationship to theft. Before making any extrapolations on this variable, however, it is important to understand that there are innate flaws with this attribute in the context of the NSAF dataset.

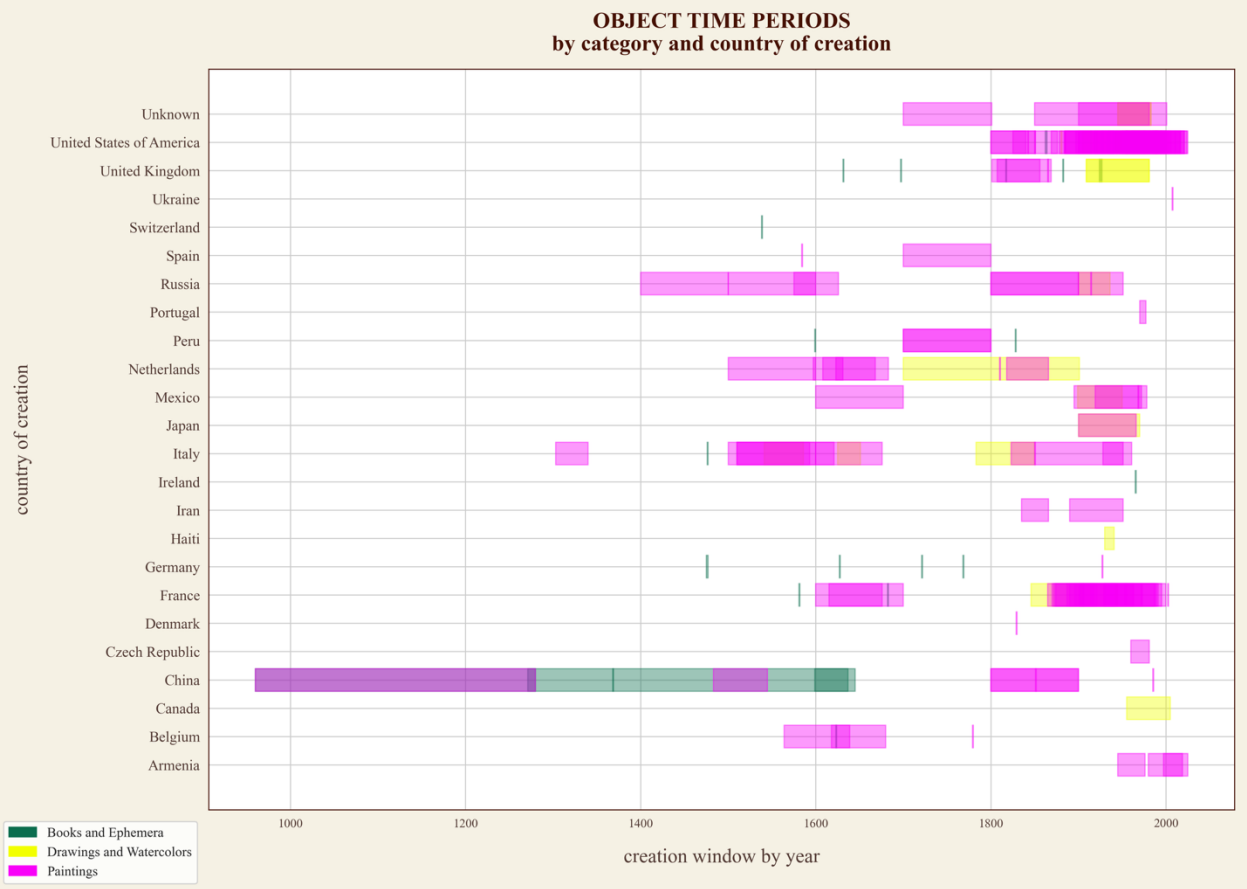
If this is a dataset of open cases, the more time that passes from a case being opened, the more likely it is to be solved, and thereby have any of the case's related objects removed from the NSAF. You could hypothesize that the longer an object exists, the more chances it has had of being stolen and ultimately entered on this list. However, because the NSAF appears to be predominantly comprised of new cases and long-cold cases, the dates of the thefts also impact the potential ages of the objects featured in the NSAF. Nonetheless, I still felt that this was a variable worth exploring.



Figures 5.8: Gantt chart of object time periods, split by category (above)

This first time period chart looks at the estimated creation windows for the objects in the NSAF sample by category. Each object in the chart is represented by a semitransparent bar, with start and end points on the x-axis representing the open and close of the nth object's potential creation period. The chart shows years with high object counts through areas of higher color saturation due to the overlapping bars. The median year for each object category is also shown with a dotted vertical line in corresponding colors.

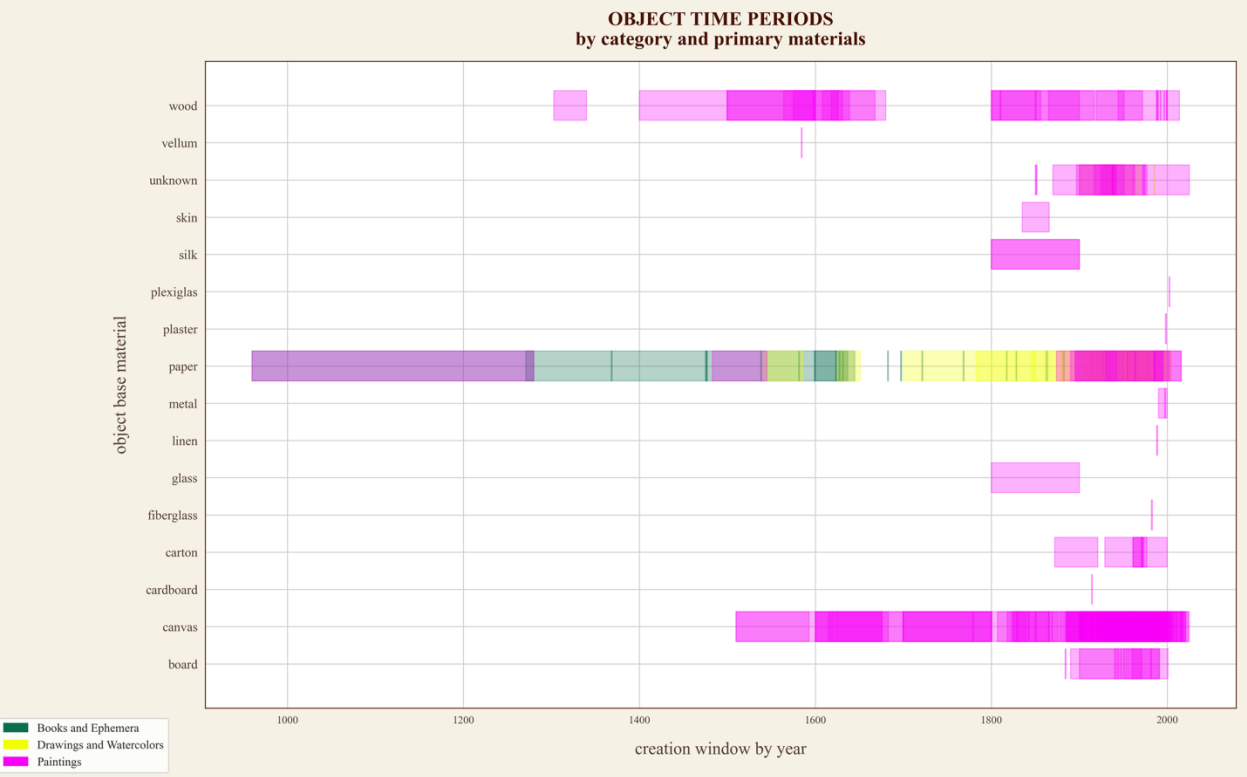
The next chart is sorted by country of creation, with object categories shown once again by color.



Figures 5.9: Gantt chart of object time periods, split by country of creation (above)

- The oldest objects in these three categories were made in China, followed by Italy, and then Russia.
- The highest saturation of objects from France occurs in the 1900s, which again likely reflects the migration of artists and collectors from Paris, a major art center of the time, to the United States around the period of the two world wars.
- With the US being a younger country, it makes sense that all of the objects from the US in this sample were produced sometime after 1800. However, one major caveat to that finding is that cultural objects made by Indigenous peoples in what is now the US generally fell under other object categories than the three analyzed here. If more objects from Indigenous creators were represented in this chart, the United States timeline would reach much further back.

The next chart shows object time periods by primary material, with categories again differentiated by color.



Figures 5.10: Gantt chart of object time periods, split by primary material (above)

Paper objects appear in all object categories and have the longest timeline of any primary material in this sample. This is an incredibly hopeful data point to come across. Compared to canvas and wood, paper is a much less hearty material. More practically, paper is also one of the historically more accessible materials, both by price and by fabrication. Industrialization has made wood panels and canvas far more accessible in recent centuries than they were for artists long ago, making those materials more popular choices now than they would have been in 960, the earliest date in this sample.

Chapter 6. Key Takeaways

In the end, my research into the NSAF was 90% data preparation and 10% data analysis. While this is not unusual when dealing with historical data, it changed the focus of my project from pattern analysis to an assessment of how the NSAF object information was compiled and shared with the public. The following were the primary challenges of working with the NSAF dataset:

- Evidence of an inconsistent data collection process;
- Inconsistent data formatting;
- Lack of backward-looking revisions;
- Influence of external market pressures;
- Unclear level of review by subject-matter experts;
- And, a lack of information on the circumstances of the thefts.

After conducting a detailed review of the various object descriptions, there are clear signs of inconsistency in the way object descriptions are collected from one case to another. The object descriptions show varying levels of subject matter expertise. That is reasonable given that most of these objects are first reported as stolen by non-specialized local police departments before being entered into the FBI NSAF. If NSAF object descriptions were, as a rule, passed by subject matter experts for review before being made public, that could result in more consistent data entries and data quality. That may be a step that the FBI has already begun doing, but objects associated with older cases do not appear to be updated as new requirements are established. More recent public statements from the FBI include requirements for objects to have more distinguishing features, quality images, and a minimum market value of $5,000. Despite that, there are objects in the database without any images or distinguishing features. In a 2023 episode of the Inside the FBI podcast, the Art Crime Program Manager goes as far as to explain why ancient gold coins are tricky and unlikely to be found. Nonetheless, many such coins are in the NSAF.

In terms of the value limitations, many of the objects I reviewed have fair market values well below $5,000. I blame that partly on the number of older cases in the NSAF and partly on the link between insurance filings and police reports. Insurance agencies typically require that losses be reported to the police before a claim is paid, and subsequently, many police

officers are given inflated values on objects, which are then reported to the FBI as meeting the NSAF threshold. In the field of art appraisal, it is understood that insurance estimates are often inflated. In cases when artwork may have been stolen from a gallery or a private residence, a gallery is likely to provide its insurance with the maximum retail price of the lost work, even if it is highly unlikely that said artwork would have been sold for that amount, with artworks often priced with space for discounts to preferred collectors. Similarly, I saw objects that were marked as original paintings, which by description were clearly prints, suggesting that some prior owners may have misunderstood what they had, and by extension, the value of what they had was misreported to insurance and the police. Again, the challenge of having misinformed victims who misdescribe their property to the police could be addressed by adding an initial object review by a subject-matter expert. Of course, that added step would come with its own administrative burden and expenses.

Despite these challenges, the data analysis portion of this project did provide a glance into the potential of studying lost and stolen art datasets, like the NSAF. An analysis of sample data from the books and ephemera, drawings and watercolors, and paintings categories in the NSAF was enough to find patterns in object sizes pointing to higher frequencies of smaller (or collapsible) objects, developed an increased understanding of the represented artist cultures, and find a suggestion of a relationship between object aesthetic appeal and theft (via object color).

Chapter 7. Future Potential

Although this exploratory exercise highlighted challenges to working with the NSAF, the patterns that my analysis found were promising enough to encourage further research into conducting a more cross-category approach. I would like to expand my sample by adding objects from the following categories: Icons, Triptychs, and Diptychs; Photographs; Prints, and Sculptures. With those four additional categories, my sample would then be representative of 75.9% of the NSAF dataset. When a more robust cross-category sample is available, I would hope to be able to share a more in-depth set of trends in the lost objects. By sharing and discussing those findings with other arts and cultural researchers, together we be able to apply that knowledge to prevent future loss.

Should other researchers find this approach to studying lost and stolen objects useful, I may move forward with this line of study by working with a subset of the Interpol Stolen Works of Art Database for a specific country, comparing patterns in objects and object record keeping.

Lastly, as I move forward, I should also begin submitting tips to the FBI Art Crimes Team with the provenance details I come across for specific objects (such as the sales record for the set design by Nicolai Konstantinovich Kalmakov). Regardless of whether or not this is new information to the FBI team, the exercise might lead to a deeper understanding of how their process and public record keeping works.

Bibliography

D'Ignazio, C., & Klein, L. (2020). *Data Feminism*. Retrieved from https://data-feminism.mitpress.mit.edu/

Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly*, vol. 5, no. 1, 2011.

SMU DataArts. *Arts Vibrancy Map*. Received from https://dataarts.smu.edu/ArtsVibrancyMap/

Burns, Charlotte, & Halperin, Julia. *The 2022 Burns Halperin Report*. Artnet, December 2022. Retrieved from https://news.artnet.com/art-world/burns-halperin-report

Thornes, Robin, with contributions by Dorrell, Peter and Lie, Henry. *Introduction to Object ID: Guidelines for Making Records that Describe Art, Antiques, and Antiquities*, 1999. Retrieved from https://www.getty.edu/publications/virtuallibrary/0892365722.html

International Council of Museums and J. Paul Getty Trust. *Object ID*. ICOM, 1999. Retrieved from https://icom.museum/en/resources/standards-guidelines/objectid/

FBI, "National Stolen Art File, FBI Art Crime Program, August 25, 2024. Retrieved from https://artcrimes.fbi.gov/

FBI, "Art Crime," Federal Bureau of Investigations, December 26, 2025. Retrieved from https://www.fbi.gov/investigate/violent-crime/art-crime

Ellen Ferrante, host, with Kristin Koch, guest, "The Art Crime Program," in *Inside the FBI Podcast*, produced by the FBI, November 1, 2023. Retrieved from https://www.fbi.gov/news/podcasts/inside-the-fbi-podcast-the-art-crime-program

FBI, *FBI Art Crime Team Marks 10-Year Anniversary*, FBI.gov, February 9, 2015. Retrieved from https://www.fbi.gov/video-repository/newss-fbi-art-crime-team-marks-10-year-anniversary/view

FBI, *FBI Art Crime Team Training*, FBI.gov, February 5, 2010. Retrieved from https://www.fbi.gov/video-repository/newss-mp4-art_crime_training.mp4/view

Nate Freeman. *Blue and red paintings make the most money at auction according to a study*. Artnet News, March 2019. Retrieved from

https://www.artsy.net/article/artsy-editorial-blue-red-paintings-money-auction-study

New York Times. 'Three Penny Opera' Paintings Born Anew. New York Times, January 2001. Retrieved from https://www.nytimes.com/2001/01/20/arts/threepenny-opera-paintings-born-anew.html