MUSEUM+VIS

Process Book

PROJECT MILESTONE

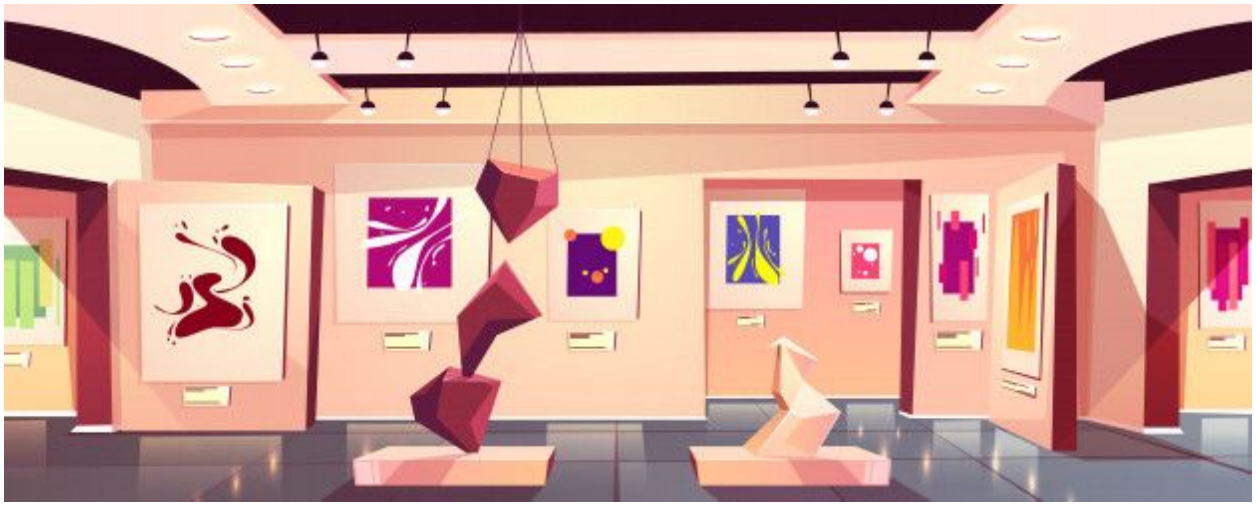
[WIP]

Table of Contents

Curated by:
Derya Akbaba
Madison Cooley
Cole Polychronis

Overview + Motivation

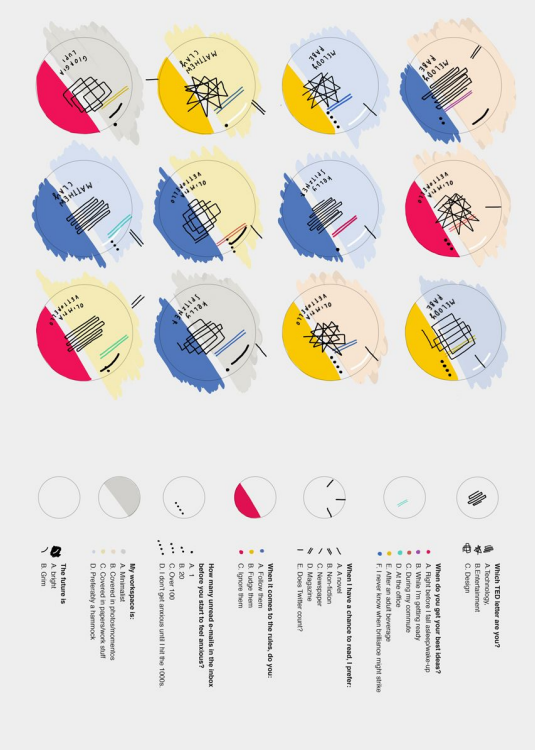
How prolific are museums in America? Are there trends in the artwork that they acquire? Is there anything interesting about the provenance of artwork? Our initial interest was sparked by the opacity of museums: the images we see at a museum only represent a small portion of the museums' collection. Our project is an attempt to unveil the artwork and paint a picture, pun intended.



Frame vector created by vectorpouch, obtained from Freepik.com

Related Work

Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.



Data Portraits

Commonly used by the self-proclaimed data humanist, Georgia Lupi, data portraits are an illustrative method of conveying differences across entities that share similar attributes. While not the most precise data visualization, it is an artistic and interesting way to portray differences and similarities within a population.

Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these "data stories" differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and messaging, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the interface) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 Introduction

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reporters use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent feature in The Economist [6] explores the proliferation of digital data and notes that visualization designers are *"melding the skills of computer science, statistics, artistic design and storytelling."*

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful "data stories" requires a diverse set of skills. Gershon and Page [12] note that effective story-telling *"require[s] skills like those familiar to movie directors, beyond a technical expert's knowledge of computer engineering and science."* While techniques from oration, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data similarly can be organized in a linear sequence, they can also be interactive, inviting verification, new questions, and alternative explanations.

Currently, most sophisticated visualization tools focus on data exploration and analysis. Applications such as spreadsheets and visualization tools support an array of analysis routines and visual encodings, but beyond exporting images for presentation typically provide scant support for crafting stories with analysis results. As such, they provide powerful vehicles for discovering "stories", but do little to aid narrative communication of these findings to others. As tools mature and more richly integrate with the web (e.g., Many Eyes [25], Tableau Public [22], GeoTime Stories [8]), they are enabling the publication of dynamic graphics with variably constrained levels of interactivity. It remains an open question how the design of such tools might be evolved to support richer and more diverse forms of storytelling.

- *The authors are with Stanford University, Stanford, CA 94305. E-mail: {segel, jheer}@stanford.edu.*

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

In this paper, we investigate the design of narrative visualizations and identify techniques for telling stories with data graphics. We take an empirical approach, analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. After reviewing related work, we share five selected case studies which highlight varied design strategies and illustrate our analytic approach. We then formulate a design space constructed from an analysis of 58 examples. Our analysis identifies salient dimensions of visual storytelling, including how graphical techniques and interactivity can enforce various levels of structure and narrative flow. We describe seven genres of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and video. These genres can be combined with interactivity and messaging to produce varying balances of author-driven and reader-driven experiences. Finally, we discuss the implications of our framework, noting recurring design strategies, promising yet under-utilized approaches to integrating visualization with other media, and the potential for improved user interfaces for crafting data stories. By focusing on the graphical and interactive elements of narrative visualization, our approach gives less attention to the cognitive and emotional experience of the reader. We recognize the importance of these elements, however, and describe directions for future reader-centric research in our conclusion.

2 Related Work

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as *"the world's second-oldest profession"* [12]. Without summarizing millennia of achievement, we describe a few of the key concepts informing narrative visualization.

2.1 Narrative Structure

The Oxford English Dictionary defines *narrative* as "an account of a series of events, facts, etc., given in order and with the establishing of connections between them." Central to this definition is the notion of a chain of causally related events. Stories of this form often have a beginning, middle, and end [3, 24]: an introduction to the situation, a series of events often involving tension or conflict, and a resolution.

Since ancient times, people have tried to understand and formalize the elements of storytelling. For example, writers (e.g., [5, 19, 21]) have developed typologies of dramatic situations and identified plot lines common to many narratives, such as the "hero's journey" [5]. This research typically distinguishes between the content of the story and the form in which it is told. While stories often concern interacting characters, they may also present a sequence of facts and observations linked together by a unifying theme or argument.

Storytelling strategies vary among media and genre. For instance, stories told through writing have access to a different set of formal mechanisms and narrative structures (e.g., stream of consciousness) than stories told through film (e.g., split-screen sequences [3]). Blundell [2] describes narrative devices for journalism such as the *anecdo-*

Narrative Visualization

The paper by Segel and Heer outlines different methods of storytelling in data-driven narratives. This paper has informed our choices on how to incorporate storytelling elements into our project.

Questions:

ARE MUSEUMS TRENDY?

With our exploration into museum data we are trying to understand the differences and similarities across museum art collections. We are interested in whether or not museums are trendy. More specifically, are there trends across museums (in the US) in artwork provenance? By looking at where artwork originates, is it possible to see trends in how trendy certain types of artwork may be across time? Do these trends change over time? For example, maybe museums pre-1950s were acquiring only European art, but post WWII with the surge of nationalism and rise in art movements across the US, museums began to focus on acquiring pieces made in America.

Do historical events, like depression or war, reflect in what artworks exist in a museum's collection? Can we use a museum's collection as a proxy for the trendy-ness of categories of artwork (eg sculptures, oil paints, wtc)?

Although our data is limited, we may even be able to speculate whether museums tend to act independently (we would see different trends in artwork provenance and acquisition) or do they act like a pack (we would see similar trends in provenance and acquisition).

Data:

Our data is comprised of seven different publically available datasets published by the following North American museums:

- Penn Museum (link: https://www.penn.museum/collections/objects/data.php)
- Museum of Modern Art (link: https://github.com/MuseumofModernArt/collection)
- Canada Science and Technology Museums (link: http://data.techno-science.ca/en/dataset/cstmc-smstc-artifacts-artefact)
- Cooper-Hewitt Smithsonian Design Museum (link: https://github.com/cooperhewitt/collection/tree/master/meta)
- Minneapolis Institute of Art (link: https://github.com/artsmia/collection)
- Metropolitan Museum of Art (link: https://github.com/metmuseum/openaccess)
- Cleveland Museum of Art (link: https://github.com/ClevelandMuseumArt/openaccess)

Initially, we had also planned to incorporate a dataset for the Williams College Museum of Art (link: https://github.com/wcmaart/collection). However, upon beginning our data processing, we realized that this dataset did not have acquisition date data, which is an attribute which is necessary for multiple parts of our proposed visualizations.

Given that our data came from seven unique sources, each of which had its own style that it used to format the data, there was a substantial amount of data cleaning required in order collate all of these datasets into a single dataset to feed into D3. As stated in our proposal, at a minimum, each dataset needed to describe an artifacts using a name, a country or region that it originated from, a date for when it was in use, and a date for when it was acquired by the museum. Each of the above datasets recorded these attributes, but in order to access them, we needed to create a different feature

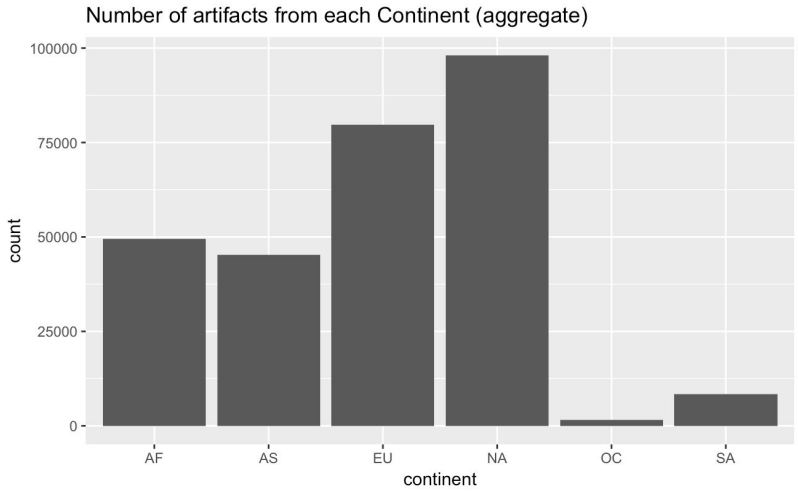
extractor for each dataset. To this end, we developed a Python script that processed each museum dataset separately, collated them into a single csv file, and then dropped any rows for which data was missing, which resulted in a cleaned dataset with over 280,000 artifacts. Contrary to what we initially proposed and at the suggestion of our TA, we did not subset any of the data and only dropped data from our final dataset to be used with D3 if there were missing values.

In order to make the data in our final dataset more easily usable in D3, we derived two attributes for all of our original data: continent of origin and the ISO three digit country code, which were both derived from the country of origin for a given artifact. This necessity also encouraged us to use Python to develop our processing script, as there is a Python package (PyCountry) which enabled us to derive these attributes from the country of origin.

Exploratory Data Analysis:

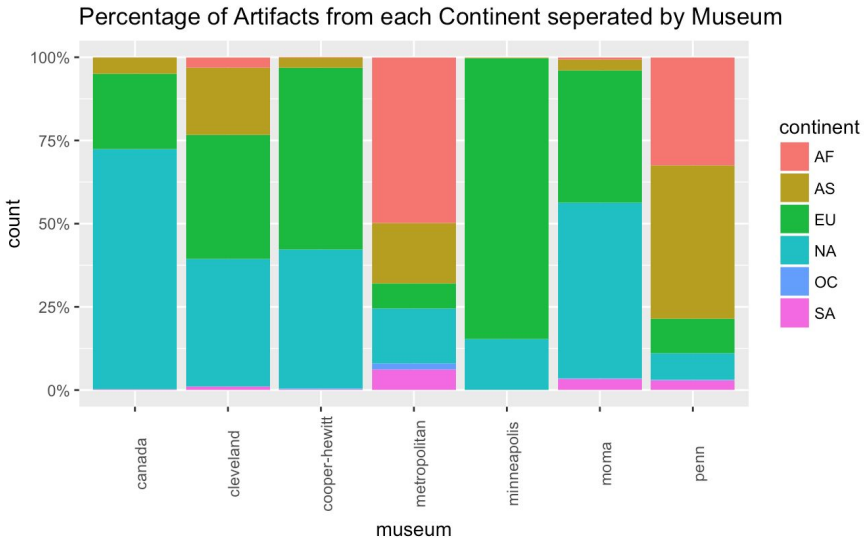
What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

One of the first visualizations that we used to explore our data was a simple bar chart to see which continents artifacts were predominantly coming from. This chart is shown below:



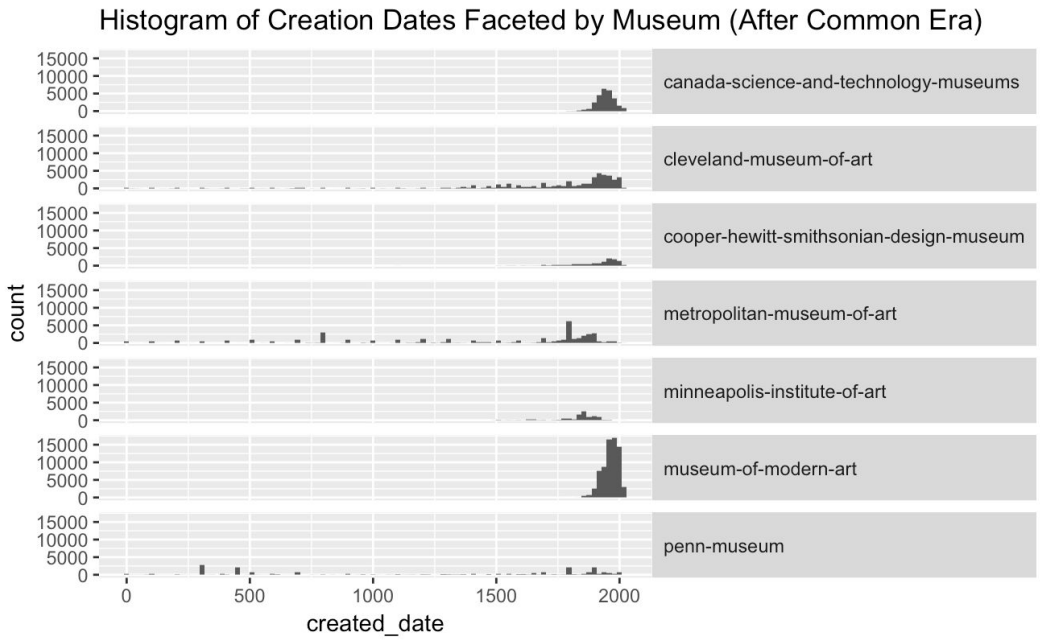
While the visualization itself is incredibly simple, the message it delivers is also readily clear, if not slightly predictable: there is a clear bias in North American museums to display artifacts from North America and Europe.

To explore this bias further, we created stacked bar charts, which shows the proportion of artifacts from each continent for each museum. This chart is shown below:



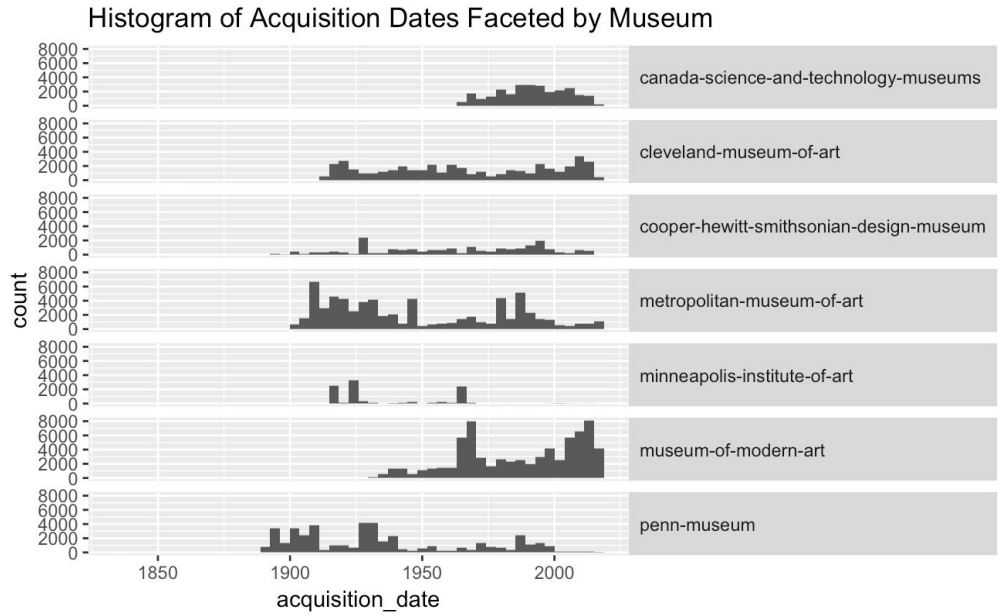
From this chart, the potential for bias that we saw above becomes even more readily apparent. For all but the Met and the Penn Museums, North American and European artifacts make up well over three-quarters of their collections and for several, it appears that proportion is close to one-hundred percent. Additionally, we can see that a proportionally large number of African artifacts can be found at the Met and a proportionally large amount of Asian artifacts can be found at the Penn Museum. This initial data exploration is what encouraged us initially to use parallel coordinates to display this data and as we approach the milestone, we have not yet figured out the best way to present this information to a user of our visualization, as we believe that this is a fairly impactful piece of the story that we want to tell.

Moving away from this particular exploration, we also analyzed the creation dates of the artifacts present in these museums to see if there was anything potentially interesting in the spread of creation dates for these artifacts. The histograms showing this spread for creation dates during the Common Era are shown below:



Again, while these results are not surprising, they do point us towards useful things to potentially highlight in our storytelling piece of the visualization: it is clear that the Canada Science and Tech museums and the Museum of Modern Art are clearly more inclined towards collecting artifacts from within the last 500 years, whereas the other museums have a considerably longer-tailed range of dates from which there artifacts were collected. On the opposite end of the spectrum, the Cleveland Art Museum, the Met, and the Penn Museum all pull a significant number of artifacts from a much earlier time period than their peers.

Moving on to the final piece of data analysis that is intimately tied with our question of trendiness of museums, we turn our attention towards acquisition dates of artifacts. Below, we have a histogram showing the spread of acquisition dates for each museum:



From these histograms of acquisition dates, we can see that there is a substantial amount of collection during the early 1900's, especially for the Met and the Penn Museum. Additionally for these two museums, we can see that there is specifically a large amount of collection in the decade leading up to World War II. In this same vein, we can see that the Cleveland Museum and the Minneapolis Institute has a substantial amount of collection during the 1920's, which could possibly be due, in part, to World War I and the economic success of America during the Roaring 20s.

Moving on, we see that the Cleveland Museum, the Minneapolis Institute, and especially the MoMA acquired a large amount of artifacts during the 1960's and the early 1970's. The Met collected a lot of artifacts during the 1970's and early 1980's. Finally, and perhaps unsurprisingly, the Canada Science and Tech Museums and the MoMA collected a substantial amount of artifacts after the 2000's.

These are definitely salient points to emphasize in both the storytelling portion of our visualization, as well as providing reasonable evidence that looking at the trendiness of museums is an interesting subject to consider.

Design Evolution:

The following section walks through individual components of our design and then presents the layout as a collection of these components. We considered many different visualizations that would enable us to tell a story and show aggregate comparisons of museums in North America as well as characteristics of individual museums.

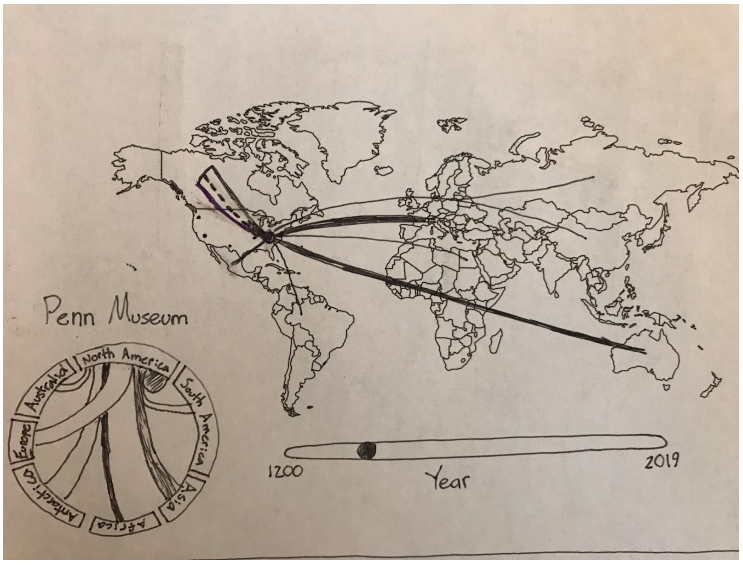
We received feedback from Kiran (our TA) and Lizzie Indra Kumar, Bond Denhalter, and John Lund. The notes from these feedback sessions can be found in the Appendix.

Designs in order of appearance:

- Chord Diagram
- Parallel Coordinates
- Heat Map
- Data Portraits
- Layout

CHORD DIAGRAM

Original Design and Justification



Chord Diagram overlaid on Geographic map

Each ribbon represents the number of artifacts that a museum gets from a certain country (or continent if country would be too cluttered). Clicking on the ribbon breaks it up into multiple ribbons, which encode the type of artifact using color. A year slider is provided to allow users to explore changes over time. In addition, a chord diagram is shown in the bottom left for the selected museum to show the same data, but not geographically positioned.

Initial Feedback on Design

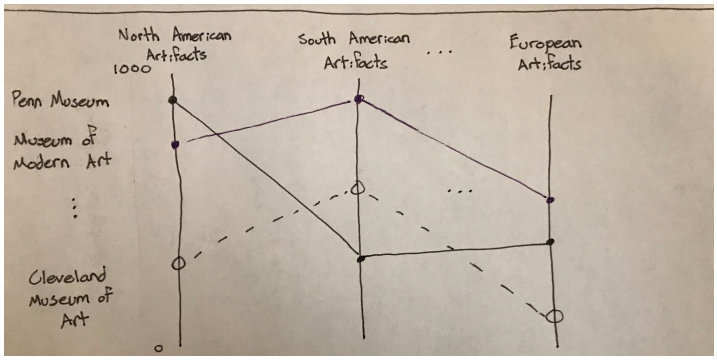
Chord Diagram
On review, we found that chord diagrams may not be necessary, since they will only be showing the connection between North America and other continents.

As a result we decided not to include the chord diagram in this view.

Map Connections
Again on review, we received multiple comments that ribbons would result in a visually cluttered map. As a result, we decided to experiment with other marks and channels to indicate data provenance based on country.

PARALLEL COORDINATES

Original Design and Justification



Parallel Coordinate of Artifact Continent of Origin

Each vertical axis represents a specific continent. The line chart connecting this axis represents how many artifacts from that continent each museum has. This allows for easier comparison than the geographical map does. Additionally, a slider could be implemented to see how this changed over time.

Initial Feedback on Design

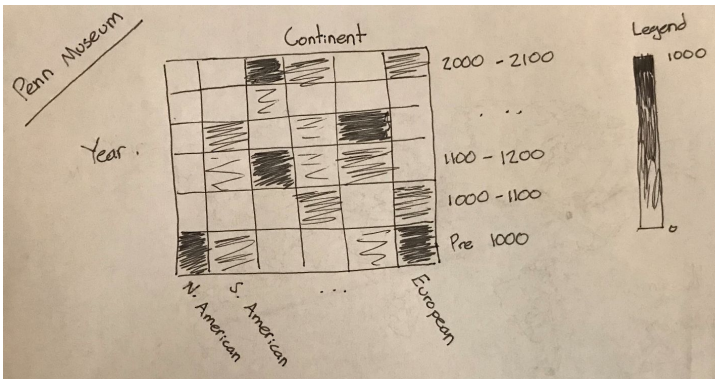
We received mixed feedback on the use of Parallel coordinates. While visually attractive, the flow of across the parallel coordinates is not representative of the data structure nor is it semantically meaningful.

An alternative design to show the distribution of artifact location origin would be to use a tree map.

At the time of the project milestone, we have not discussed an alternative visualization.

HEAT MAP

Original Design and Justification



Heat Map comparing continent of origin with time of use/creation of artifact

A heatmap like this would be generated for each museum. The vertical axis represents bins for when the artifact was created/used and the horizontal axis represents the continents of origin for the artifact. The heatmap allows for exploration of if museums tend to favor newer or older materials and from where.

Initial Feedback on Design

While the heat map is an appropriate way of showing the relative distribution of artifact provenance it may not be a super interesting visualization.

We decided to make this a nice to have feature after further data exploration.

DATA PORTRAITS

Original Design and Justification

Data Portraits

Using attributes of the collection, we will generate portraits for each museum. These visualizations are less for legibility and more for comparabing the general attributes of the museum.

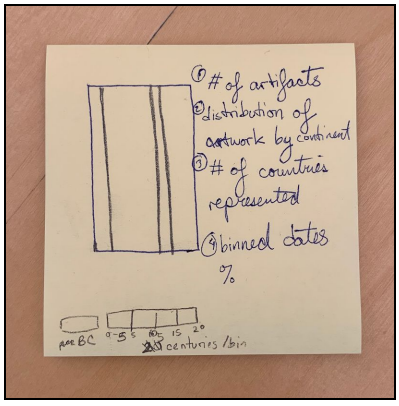
Initial Feedback on Design

The initial inspiration (*to the left*) is highly stylized, and instead we decided to opt for a more simplistic iteration of a data portrait.

Inspired by Georgia Lupi and the Abstract Expressionism movement of post-WWII artists, we opted for a more formulaic and simpler visual representation of the museums.

Mockups

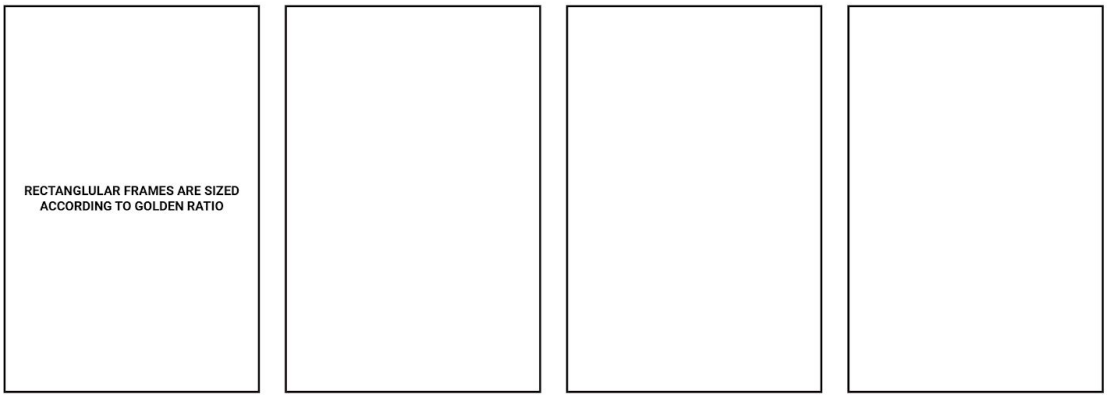
We began by looking at different Abstract Expressionism art, looking specifically for the use of bold colors and geometric shapes. Once we selected an artist for inspiration we created a rough mockup using pen + paper of what the portrait could look like.



Pleased with the rough sketch, we used Adobe Illustrator for cleaner mock-ups. After implementing the design in code using fake (but representative) data, we altered the design slightly to both make the coding easier and incorporate a more organic element into the design.

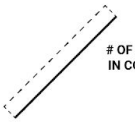
MUSEUM+VIS Gallery

OUTLINE



CREATED DATE OF ARTIFACT

ANYTHING BC | 0 - 5TH C. | 6 - 10TH C. | 11 - 15TH C. | 16 - 20TH C.

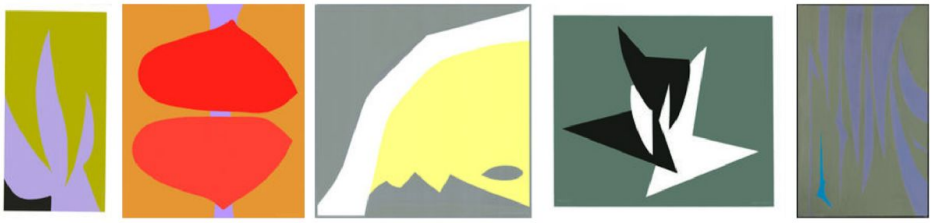


OF COUNTRIES

IN COLLECTION

MAUD MORGAN

1903-1999



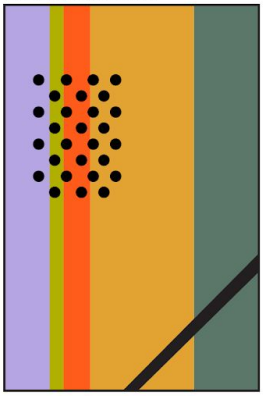
IMAGES COURTESY OF ARTSY.NET

COLOR PALLETTE

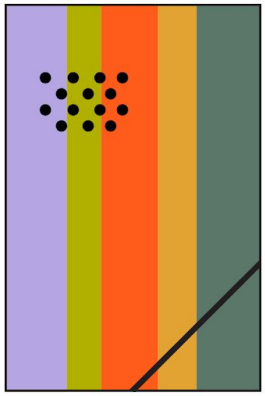
MUSEUM+VIS Gallery

PRE-CODING MOCK-UP

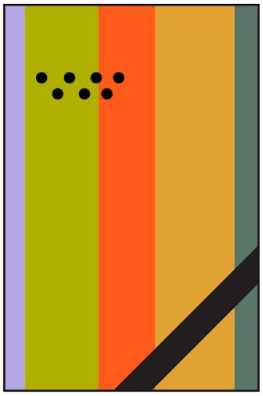
MUSEUM OF MODERN ART



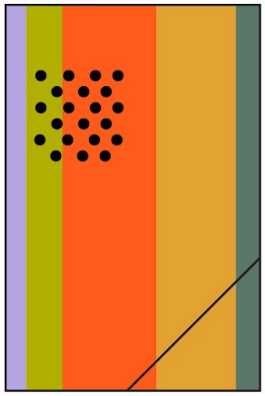
CLEVELAND MUSEUM OF ART



COOPER-HEWITT

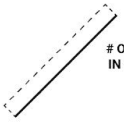


PENN MUSEUM



CREATED DATE OF ARTIFACT

ANYTHING BC | 0 - 5TH C. | 6 - 10TH C. | 11 - 15TH C. | 16 - 20TH C.



OF ARTIFACTS IN COLLECTION

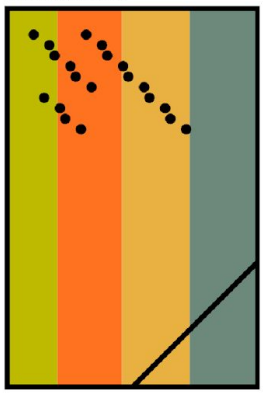
OF COUNTRIES

IN COLLECTION

MUSEUM+VIS Gallery

POST-CODING

FAKE MUSEUM DATA



DOTS EVOLVED INTO MORE ORGANIC STRUCTURE
EACH DOT REPRESENTS 5 COUNTRIES

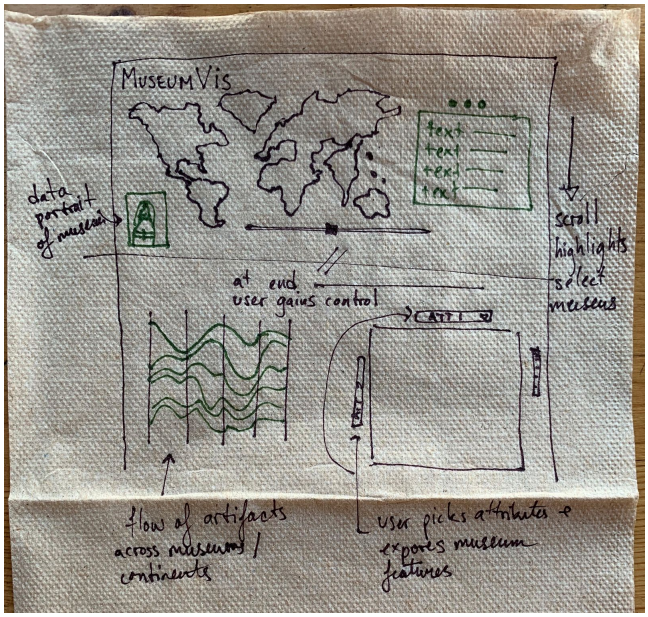
THE VARIATION OF THE DIAGONAL LINE WIDTH IS LESS DRAMATIC

IN ORDER TO AVOID CODING IT AS A <PATH> THE MAX WIDTH OF THE LINE WILL REMAIN THE SAME AS THE FRAME BORDER

COLORS AND MAPPING TO DATES REMAIN THE SAME

LAYOUT

Original Design and Justification



Layout

User scrolls through ~3 museums (there will be an option to click through the same examples). At the end of the examples the exploratory buttons appear and enable the user to see the flow of attributes across museums over continents and over time (by dragging the timeline). The heat map lets users explore the different attributes present in one museum at the given time and is also responsive to the timeline drag.

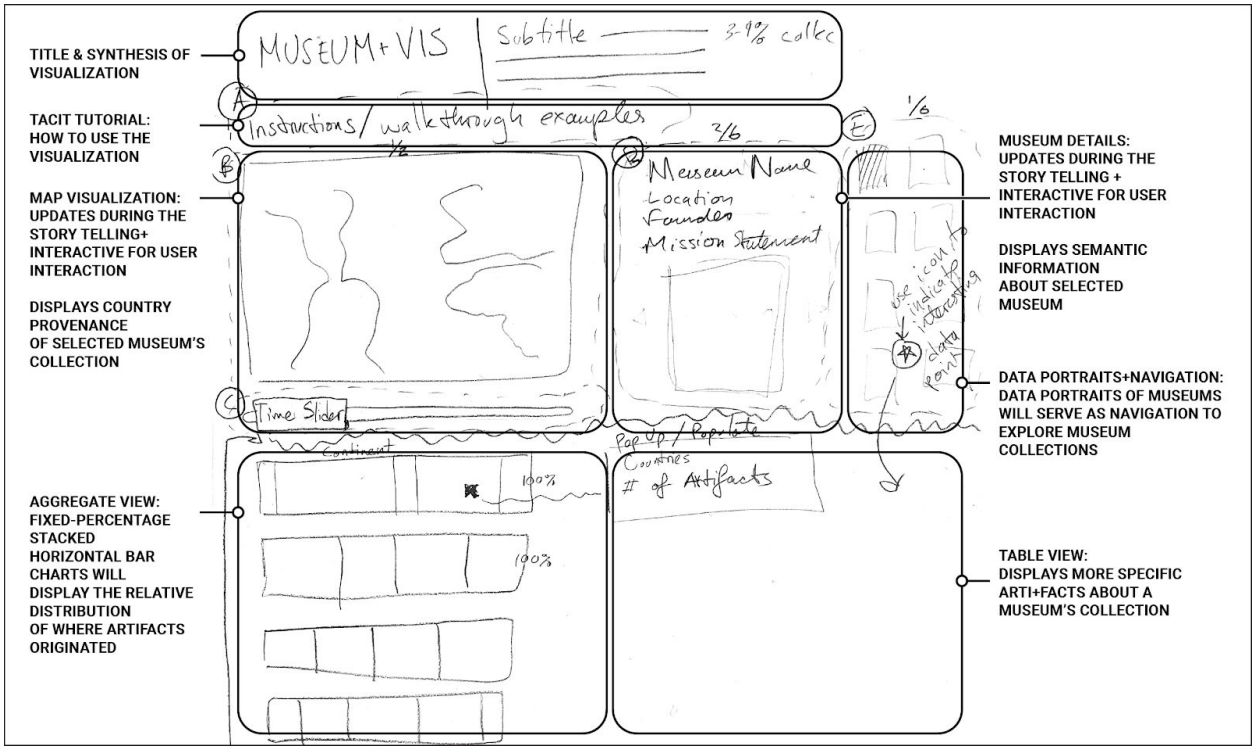
Initial Feedback on Design

Initially we had planned to incorporate scrollytelling into our design, but decided against it because it was unjustified interaction and we believed we could have more straightforward storytelling using buttons and the interactive slideshow method.

Layout Revisited

After receiving feedback from both other students and our TA, we decided to revisit the layout to clarify the different types of interaction and the storytelling components of the vis.

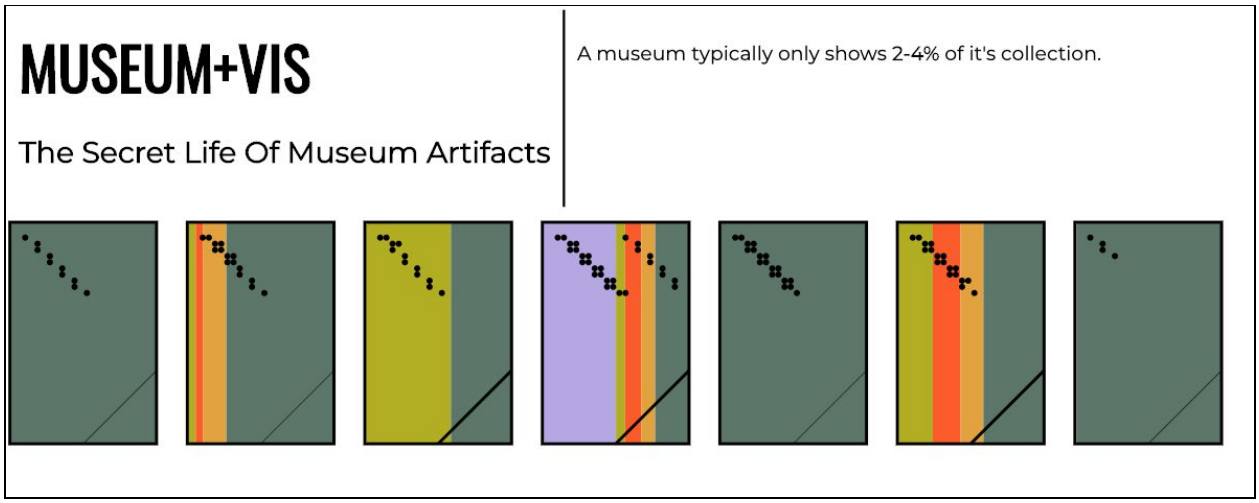
Mockups



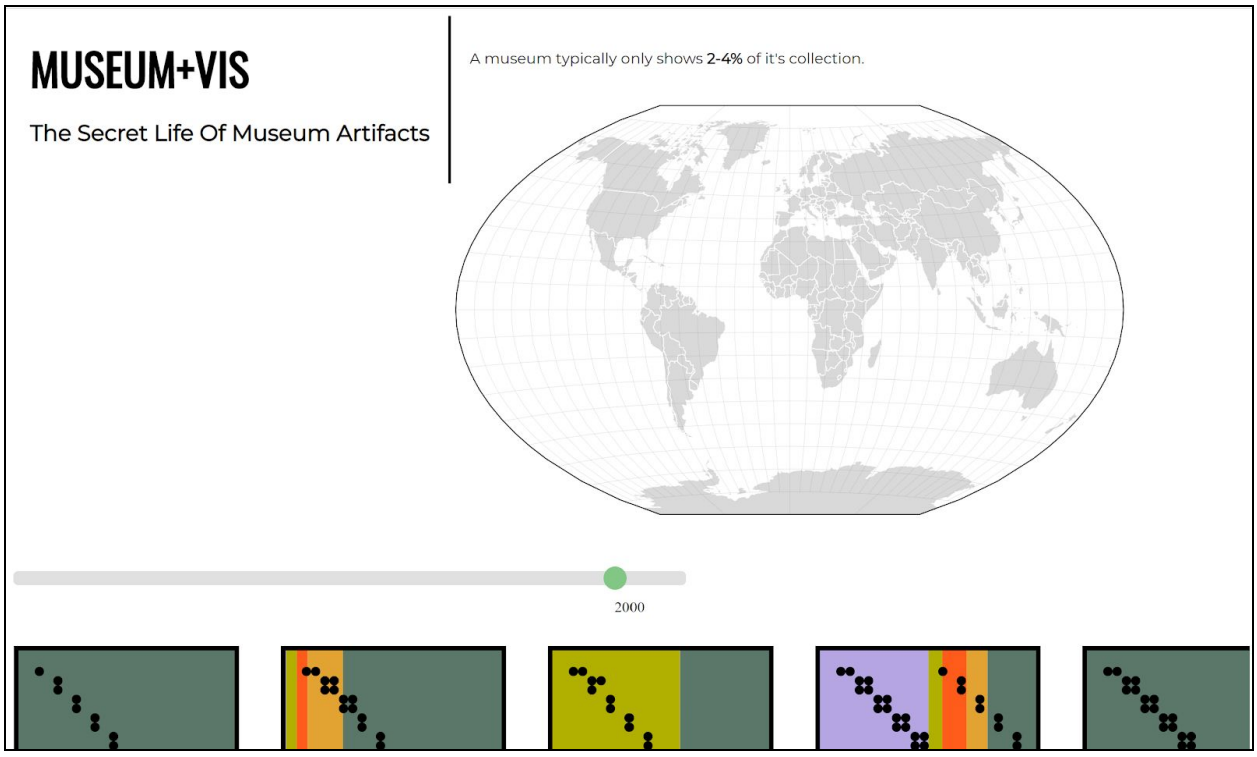
11.4.2019

This was our initial sketch where we outlined the sections and fleshed out the interactivity and storytelling elements.

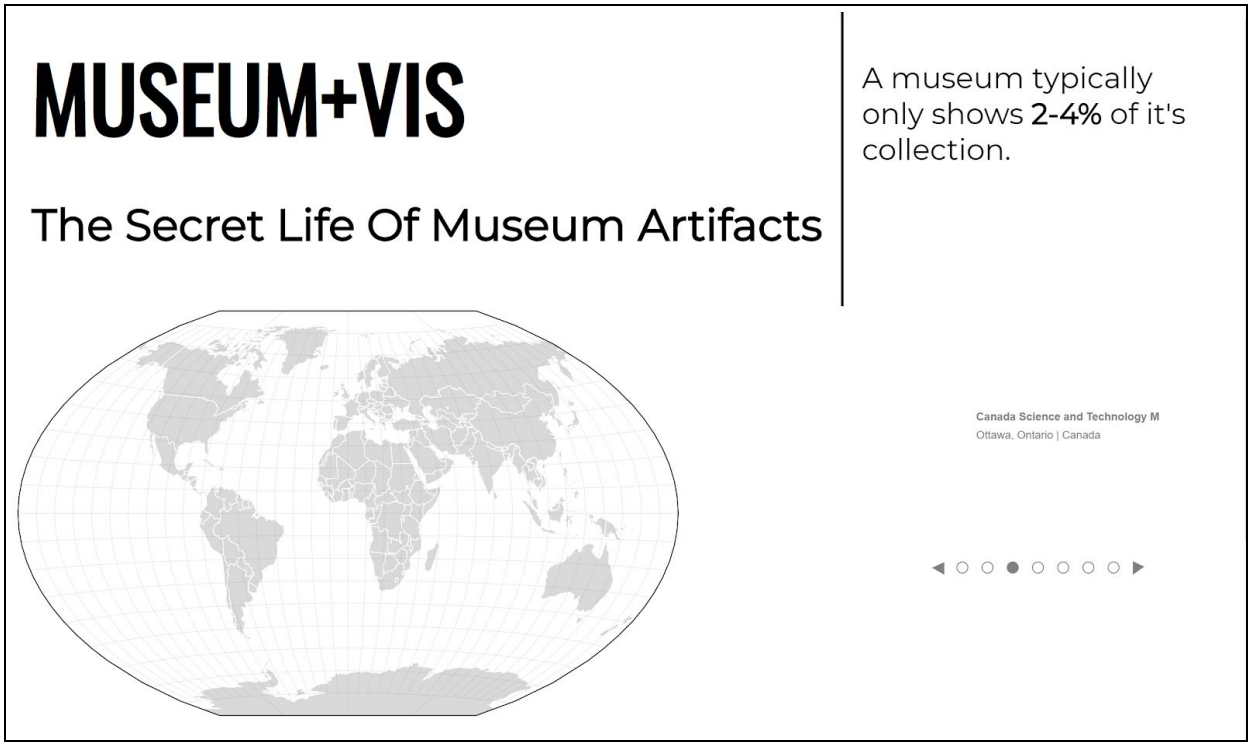
In-progress Screenshots



11.4.2019
Data portraits created on separate branch. No map vis.



11.5.2019
Map and data portrait visualizations merged. Formatting of containers leads to wonky views.



11.6.2019

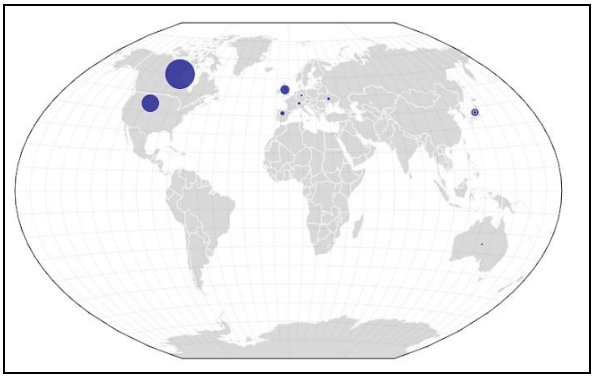
Implementation of grid, header elements, and storytelling buttons. The sizing of the font is too big.

11.6.2019

Data portraits scaled and adjusted. North American data added to dataset.

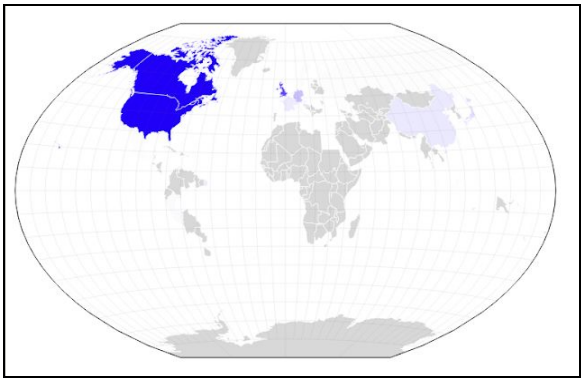
11.7.2019

We decided to explore different methods of visualizing the encoded collection data on the map. The two design spaces we explored were the bubble map and choropleth.



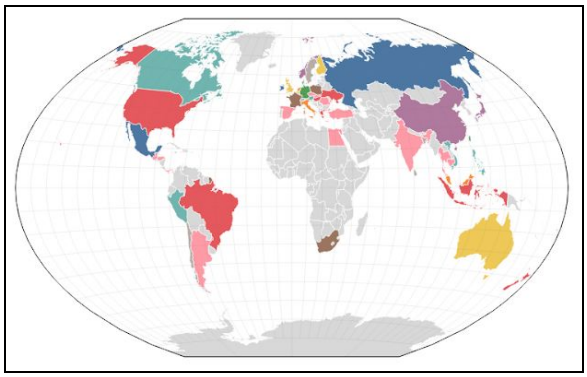
11.7.2019

Bubble map initial view.



11.8.2019

Opacity scale to visualize encoding the data on the map. Added the museum descriptions to the museum tabs along with museum website.

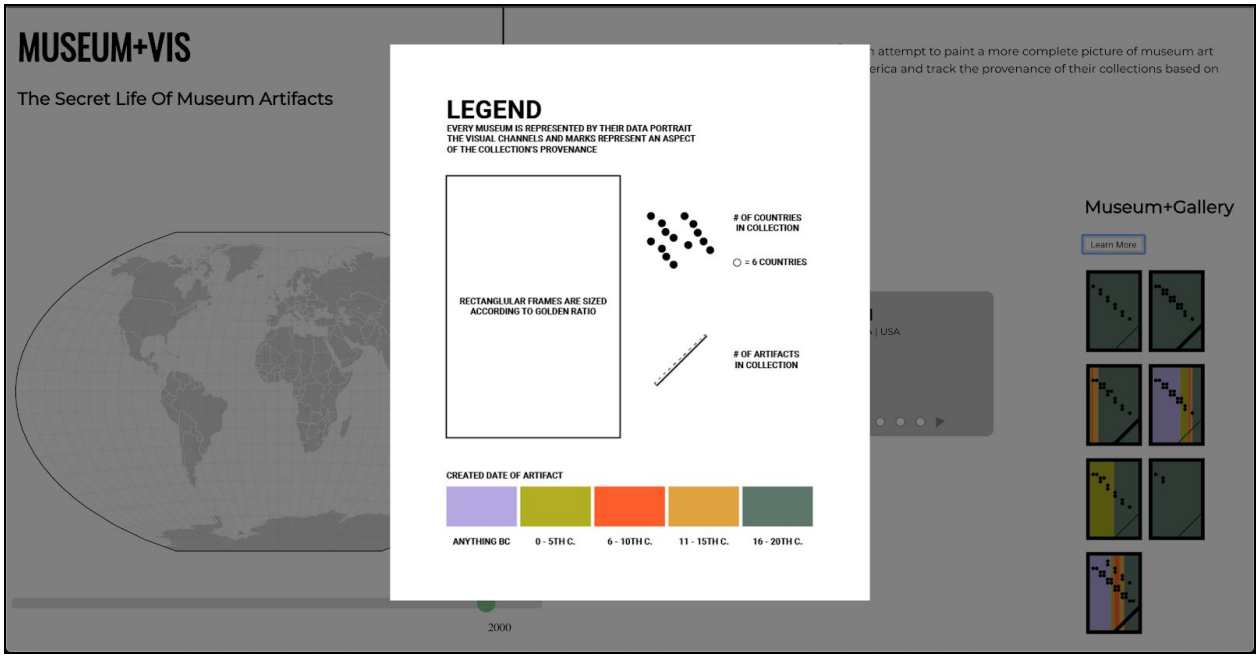


11.8.2019

Different color scale to visualize encoding the data on the map. Color scale needs some work, but it shows the countries where the artifacts originated at least.

11.8.2019

We ultimately decided to use the bubble map because it was the better encoding for the differences in artifact acquisition amounts.

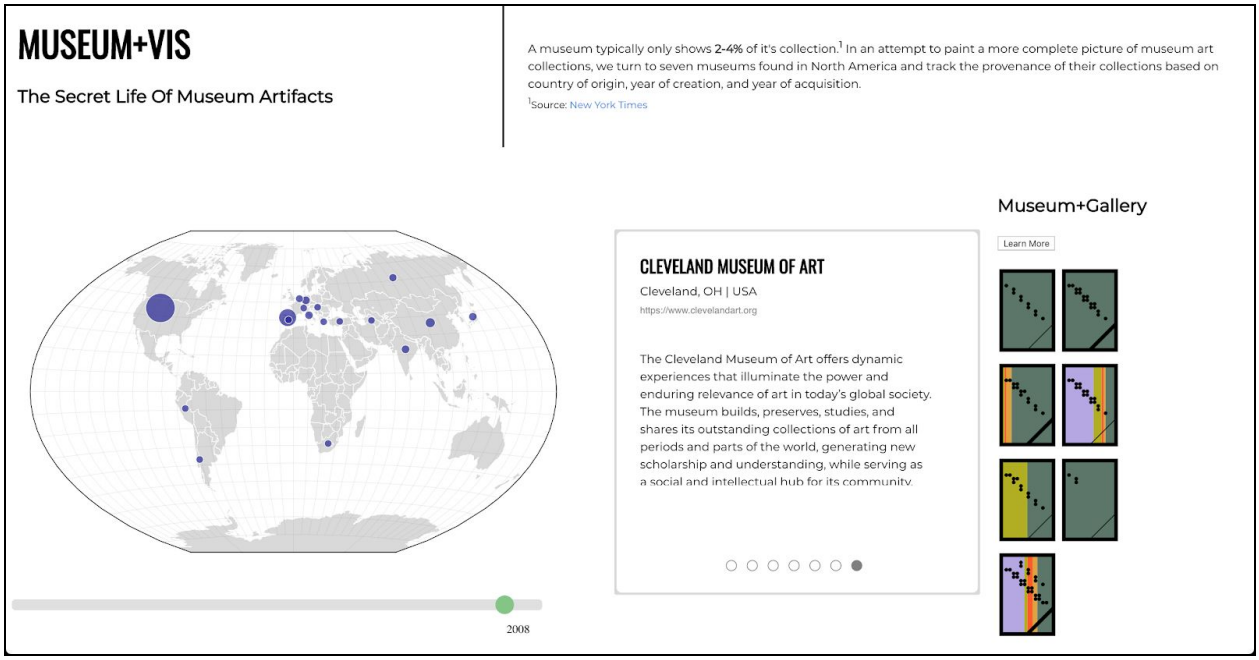


11.8.2019

We added a legend that users could click to better understand how the data portraits were created.

11.8.2019

Fixed the formatting of the middle portion (the description boxes). The description boxes now change on each click of the data portraits on the right.



11.8.2019 - Project Milestone Release

Link: https://github.com/madicooley/museum_vis/releases/latest

Implementation:

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

In this first iteration, we have three separate elements, which can be seen in the screenshot on the previous page. The first is the map, which shows the countries from where the selected museum acquires their artifacts. The circles encode the number of artifacts acquired from that specific country, relative to the area. The museum tabs (middle element) show the name of the selected museum along with a description, location, and website. The third element is the data portraits on the right. Each data portrait represents a museum, which on click, changes the map view and the tab view to the selected museum. Under the map, there is a year slider, which changes the data seen on the map to show the artifacts acquired by the chosen museum at the specific year.

Evaluation:

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

From our visualizations, we were able to see that most museums collected a large number of artifacts from both Canada and the United States each year. Many of the artifacts also came from Europe. We did notice some exciting trends related to the artifacts collected in specific years, as in an interesting number of artifacts collected during historical events. These are outlined in the Exploratory Data Analysis section. We have not been able to visualize these trends specifically yet, so that is something we would like to continue improving upon before the final deadline. To do this, we hope to point out a few significant historical events to emphasize these trends in artifact acquisition. We also hope to improve upon the museum specific data visualizations. We are planning on adding a stacked horizontal bar chart to show the continents in which each museum collected artifacts from, to easily compare all museums at once. We also will add a chart to visualize specific artifacts from each museum. So far though, our visualization works very well. The implementation is simple, conveys information about the trends in artifacts, and is visually appealing.

APPENDIX

Feedback (in class)

10.29.2019

Group: Lizzie Kumar, Bond Denhalter, John Lund

- Chord diagram may not be necessary, since they will always be going to NA
    - As it stands, what we have right now is not showing interesting data
    - Parallel coordinates might be a suitable replacement - see how they look
- Before we finalize everything, look at the numbers from the datasets to make sure it actually fills it out well
- Shift chord diagram to a nice to have feature, see if data is missing which chord diagram show
- Nice to have feature: overlay which shows important historical periods or civilization culture
- Parallel coordinates are misleading in our visualization, bar charts may be better, since there is no relation or correlation between different continents
    - Proportions should be shown, think about having proportion on parallel coordinate viz (if a museum had 20% from each category, it would look bad)
- Tree map could be an interesting way to show country and continent data, so this may be a suitable replacement for parallel coordinate
- Maybe move heatmap to a nice to have
- Overall, maybe spend more time on scrolly-telling aspect and make it stand out
- Might need a way to drill down to individual artifacts - table that populates from clicking on a country

Feedback (TA)

11.02.2019

1. Chord Diagram + Map: The ribbons will add visual clutter which is not required, you can color code the countries as heatmap for number of artifacts with a legend. On hovering a country, a popup shows the exact number of artifacts + break down of artifact types. This is also easier to implement, less cluttered and encodes all the information from ribbons. You can use continents or geographic areas instead of countries. Maybe user decides the granularity?

2. PCP: This is perfect!

3. Heatmap: This is good. Maybe you should consider the number in the cells itself, possibly in shortened format when applicable. E.g. 1000 as 1k, this is easier to understand than referring a legend.

4. Data Portraits: This is a good idea in theory, however I agree radial area charts are better than actual portraits. Except for the visual appeal of using actual portraits, it's still hard to compare general attributes without spending time with the visualization, a comprehensive legend and memorizing.

5. I like the layout and your choice of scrolling through three examples before exploration! Add a button to directly skip to exploration so that user doesn't have to go through examples every time.