Proposal - MuseumVis

Basic Info. The project title, your names, email addresses, UIDs, a link to the project repository.

Madison Cooley, u1266040@utah.edu, u1266040
Derya Akbaba, deryakbaba22@gmail.com, u1265930
Cole Polychronis, u0963138@utah.edu, u0963138

Repo link: https://github.com/madicooley/museum_vis/invitations

Background and Motivation. Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

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.

Project Objectives. Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

Are museums trendy? Are there trends across museums (in the US) in artwork provenance?

- By looking at where artwork originates, can we see trends in how trendy certain types of artwork may be across time?
    - Does this trend 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. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

We will be collecting data from approximately 10 different museums across North America. These datasets are publically available, primarily through GitHub pages created by the museum. In particular we will be collecting .csv files for the following 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)
- Williams College Museum of Art (link: https://github.com/wcmaart/collection)
- Metropolitan Museum of Art (link: https://github.com/metmuseum/openaccess)
- Cleveland Museum of Art (link: https://github.com/ClevelandMuseumArt/openaccess)

Data Processing. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

While the overall formatting of the data is clean, there is some substantial data cleaning that we will have to do.

First, our visualization techniques require that a museum artifact has at least a name, a country or region that it originated from, a date for when it was in use, and ideally a date for when it was acquired by the museum. In all of the datasets listed above, these values are missing for some items. However, each dataset contains tens if not hundreds of thousands of items, so simply removing rows with missing values should be okay.

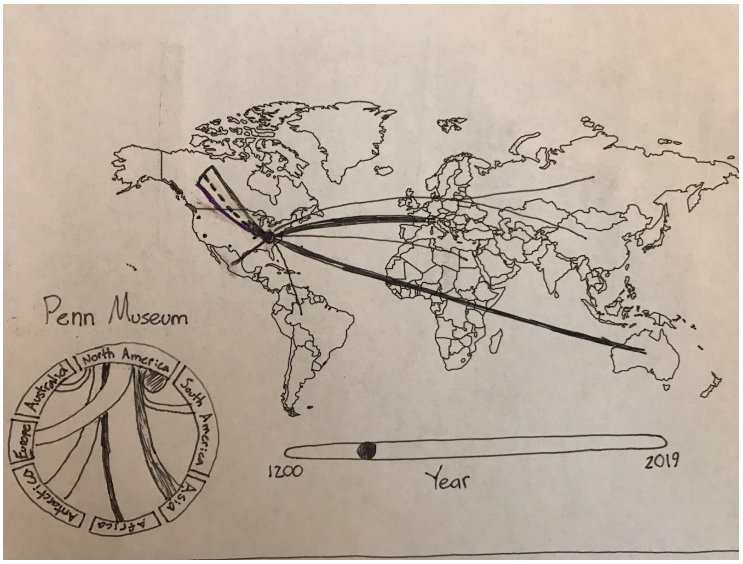
This brings us to the second data cleaning activity: each dataset is far too big to effectively visualize using D3. To address this, we will be taking random samples from each dataset. Third, since we are dealing with separate .csv files for each museum, each file is formatted slightly differently with different column names, which means we will need to write a script to unite all of this differently labeled data into one format to be read by our visualization system.

Our final data cleaning challenge is that for two datasets we will be using nationality of originator instead of the country/region of origin (for example: Japanese instead of Japan). This means that we can likely get country or region of origin for this data, but we will have to derive it using some string parsing.

In terms of how our actual data processing will be implemented from a project structure perspective, we will develop a python script (or a series of them), that will be used to pull all of the separate .csv files into one file that can be fed into our visualization system. While we will provide the generated

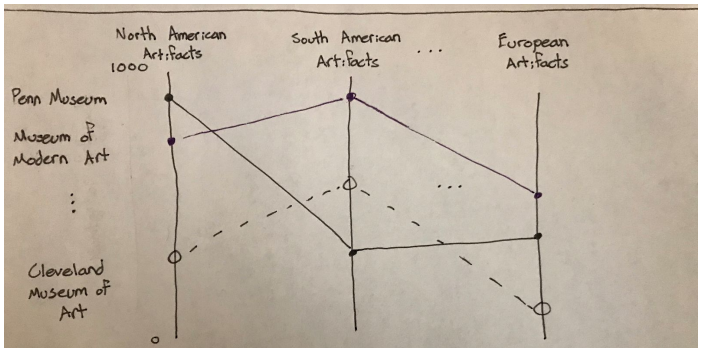
.csv file that we actually use to visualize the data, we will also include these scripts and the original .csv files for reproducibility.

Visualization Design. How will you display your data? Provide some general ideas that you have for the visualization design. Develop **three alternative prototype designs for your visualization.** Create **one final design that incorporates the best of your three designs.** Describe your designs and justify your choices of visual encodings. We recommend you use the Five Design Sheet Methodology.



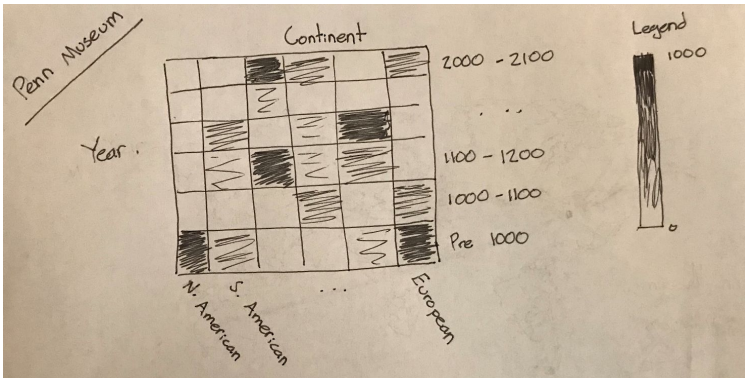
Chord Diagram overlaid on top of Geographic map

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.



Parallel Coordinate plot 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.



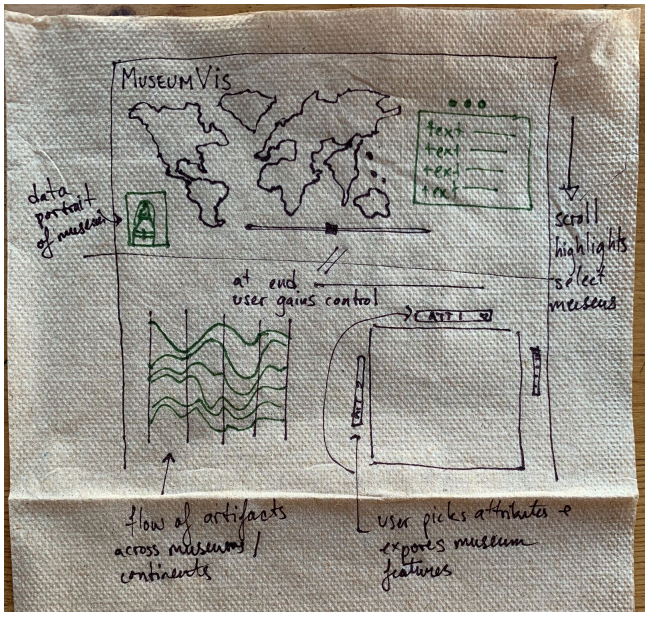
Heat map comparing continent of origin with time of use/creation of artifact

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.

While the inspiration above is highly stylized, we may find it a little ambitious and end up using radial area charts as our data portraits for museums.

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.



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.

Must-Have Features. List the features without which you would consider your project to be a failure.

1. Globe with ribbons to connect artifacts at a specific museum to the origin of the piece. This will help determine if museums follow certain trends in the types of pieces they acquire.
2. Some sort of “artistic” visualization such as a heat map to compare new and old pieces. This piece is essential to not only visualize the data, but to encode it in an appealing way to fit the theme of the project.
3. Year slider. This is essential to view trends over the years.
4. Chord diagram to also encode the ribbons which surround the globe. This will serve as another way to visualize the data to allow for additional insights.
5. Scrollytelling. We want to tell an exciting story of the data, without which the visualization will be less interesting.
6. Parallel coordinate plot. This will allow for easy visualization of trends.

Optional Features. List the features which you consider to be nice to have, but not critical.

1. Expand our python scripts to pull from several publicly available museum collection APIs so that we can include more museums in our visualization.
2. Sorted Stream graph as a either a replacement for the parallel coordinate chart or an additional visualization (https://datavizproject.com/data-type/sorted-stream-graph).

Project Schedule. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

1. **Oct. 25 - 29**
    a. Obtain Data - *All*
    b. Read in Data into a standard usable form - *Cole*
    c. Get data structures in place - *Cole*
2. **Oct. 30, 31, Nov. 1 - 5**
    a. Start View 1 & 2 - *Derya & Madi*
    b. Have functional prototype - *All*
3. **Nov. 6 - 12**
    a. Finalize process book for project milestone - *All*
    b. Clean up code to report for project milestone - *All*
    c. Schedule a time for project feedback w/ instructor (Nov. 11 - 15) - *All*
4. **Nov. 13 - 19**

a. Have all views finalized - *Derya & Madi*
b. Schedule a time for project feedback w/ instructor (Nov. 11 - 15) - *All*

5. **Nov. 20 - 27**
    a. Create a project website - *Cole*
    b. Finalize Process Book - *Madi*
    c. Finalize Code - *Derya*
    d. Create screencast of project
        i. Record (2 minutes) - *Cole*
        ii. Voiceover - *Madi*
        iii. Edit Video - *Derya*
    e. Submit!

October

Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25. **Project Proposal Due**	26.
27	28. Obtain Data	29. Read Data in	30. Start on Views 1 & 2	31		

November

Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
					1	2
3	4	5. Have functional prototype	6. Finalize process book	7. Clean up code	8. **Final Project Milestone**	9
10	11. Project Feedback->	12	13	14	15. Project <-Feedback	16

17	18	19. Have views finalized	20	21. Finalize Code	22. Create Video	23. Do Voiceover
24. Edit Video	25. Finalize Process Book	26.	27. **Final Project Due!**	28	29	30