Datasets and Variables:

We made use of 3 datasets.

a) Moma Museum Objects Infomation

https://www.kaggle.com/momanyc/museum-collection#artworks.csv

b) Met Museum Objects Infomation

https://github.com/metmuseum/openaccess/blob/master/MetObjects.csv

c) Well-Known Artworks Data

https://www.kaggle.com/ikarus777/best-artworks-of-all-time

For the first two datasets, since we are only interested in paintings and other two-dimensional artworks, and present some of them on-screen to make the Timeline visualization. We used an R program to only keep the variables we need, that is the "Titles", "Artist Name" and "Painting Time." And we filtered the data to keep the artworks classified as "Paintings", "Drawings" or other similar types. These two datasets are combined together and we added a "Museum" column to keep them trackable. We also added a "Unique Id" for programming purposes. The third dataset contains detailed data for a subset of paintings in the first 2 datasets. We used the third dataset to build the artist result information in our Quiz visualization. We manually combined country location data and artist photos from https://www.wikipedia.org and assembled a larger dataset. The original datasets are too large, we only included the processed ones.

Overview and Story: Have you ever wanted to explore the art world but you don't know where to start? Our visualization helps you get started by introducing you to an artist based on your style preferences. For this we used a quiz format, seeing as it was the best way for users to

compare images. Letting the user interact with actual images rather than just numeric or textual data might inspire more curiosity. We wanted the user to situate the artist in a particular time period and see when they may have been most active. The Moma and the Met both have large collections and much of the artwork from the two museums is from certain eras, such as the 16th century and the 19th century. If this is surprising to them, it would likely inspire them to further think about why there were more works in those time periods. Is it because the art world was blooming at those eras? Or the museums are getting more artworks from these eras for some other reasons? They can explore more with the brush and mouseover functions. We hope the detailed information will encourage the viewers to search more on the artworks they are interested in online.

Design Rationale

Part 1: Art Personality Quiz

Channels: Options, Font Size (d3 scale, In the word cloud component), Font Color (d3 scale, In the word cloud component), Map Region Color (In the map)

Other design details:This art personality quiz is a supplementary part that functionally serves as a filter for the Museum Timeline visualization. We chose the brown color as the primary color because it represents the classic wooden frame of an artwork. We added a secondary green color to make important action buttons such as the "Next" button standout, and to create a visual mapping between the buttons, selected options, and interested genres. We make use of shadows as a secondary visual cue to provide the user with a better visual hierarchy, indicating the different layers of the elements (as if they are actually objects in a art museum) and the states

(selected, disabled, etc) of these elements. Each painting and artist is associated with one or more genres. The word cloud shows the user's current art interest (genres) based on the decisions he/she made. In the end, the user's genres will be matched to an Artist's genre, the algorithm is simple, the most frequent genres the user liked will have a higher weight when matching. The word cloud makes use of Font Size and Font Color channels at the same time, making the major genres standing out.

Interactions: The user interacts with the quiz part by clicking on their preferred paintings. The page automatically scrolls up and down to centralize the user's focal point and minimize the mouse movement of the user. The word cloud changes while the user progress through the quiz, giving them feedback, indicating the current status of processed data.

Part 2: Museum Timeline

Channels: Dot bar length (height), dot color, dot x-axis positions

Interactions: The user interacts with the timeline through the brush. Deciding what kinds of interaction to use with this data set was difficult because of the sheer size and data cleaning issues. We ultimately decided to use a brush so the user could "pan" through the timeline but also see the detail in different eras and be able to point out which paintings from their artist were in which year. There were a few trade offs we had to make to design this interaction. One major problem we ran into was with the lack of even spread of the data points. More recent years have seen a lot more acquisitions than in the past, and because the MoMa is a modern art museum, the data was very uneven. To make sure that the majority of the painting dots could be seen on the

page at a zoomed in level, we had to place each year in rows of 4 dots rather than one vertical one. Even with doing this some of the dots had to get cut off on the top. We decided that these trade-offs were worth it overall so that the user could pan and see the points for individual painting.

Other design details: The timeline was designed so that the user could see the results from the quiz in the context of the collections of two of the larger art museums in the country. They can get a feel of how large the collections are and from when they pull most of their works, and can see the difference in time periods between the two museums. The integration with the quiz was the most important part of this design. Because the datasets from the Met and Moma were so large, we were worried about overwhelming the user with information. We were able to use the quiz to narrow the scope of what to present so the visualization was still meaningful to the user. It is versatile enough that the viewer gets a larger picture understanding of the data sets but also pulls a smaller, more personalized, experience from it.


Contribution:

Lily: Lily made the Museum Timeline part of this project and did designing, sketching for the visuals in this project.

Song: Song made the Quiz part of this project, and did data cleaning and processing.


Time breakdown by task:

Dataset discovery and design: 1.5 weeks

Data processing: 2 days (3 hrs)

Quiz: 3 days (10 hrs)

Timeline: 2 days (9 hrs)

Styling and debugging: throughout the development.