MoMa Dataset Visualization

https://valxwz.github.io/data-viz/gp/Final_Webpage%20Final/

SI 649 Section 1 Group 5

Linou Dai, Valyn Dall, Xinyue (Joy) Huang,
Chengyue (Amor) Qiu, Valorie Zhang

Executive Summary

We wanted to analyze the Museum of Modern Art (MoMa) artwork datasets to investigate disparities between gender representation and the artworks found in the collection. Specifically, we compared the gender distribution by analyzing the amount of artwork acquired over time; artist nationality; acquired artwork classification overtime; most productive age ranges; and how long it takes for artwork to be recognized.

These explorations were achieved by completing a series of visualizations that detailed the trends we were interested in after the MoMa dataset was cleaned and normalized. This included the following:

- World map examining artist representation by nationality and gender.
- Line graph displaying the imparity of artwork acquisition over time.
- Bar chart paired with a line graph visualizing the acquisition of artwork by classification over time and by gender
- Bar chart displaying the relationship between when the artist created their artworks and gender
- Bar chart visualizing when by gender comparing when artworks were created and later acquired

We found that in all visualizations, males dominated the artwork landscape of the MoMa collection, but we did see a few interesting trends. For example, female artists representation began to increase in line with the feminist movements of the 1970s. Additionally, some artwork classifications (e.g. sculpture and painting) have slightly more equal representation starting in the 2000s.

Questionnaire

1) What is your data?

Our data is from The Museum of Modern Art (MoMa) dataset. In its original form, it contains roughly 15,668 records. We cleaned and manipulated the data slightly in order to utilize it for our visualization (removed data not relevant to our research like artwork size and images). Additionally, the nationalities of artists were changed to their home country; data that included negative or irregular values for the date were removed; and gender values were normalized.

2) What are the domain tasks or communication goals you want to support? What should someone be able to understand after seeing / using your visualization?

Five domain tasks were chosen to communicate different insights about gender as well as artist representation within the MoMa data:

1. how the acquisition of artworks changed over time and if the woman's activities in history impact females' artworks acquisition.
2. Where do MoMA's artworks come from?
3. What is the relationship between different artwork classification acquisitions and gender?
4. Usually, at what age an artist would good artworks that were acquired by MoMA?
5. Usually, how many years would it take for artworks to be recognized by MoMA?

3) How are you encoding the data visually?

Graph 1 - Male and Female Distribution across the World

Tool: Tableau Public

Data entry used: Gender, Nationality/Country

We calculated the number of artists of different countries within each gender, and used map location to indicate geographical location and bubble size to indicate the number of artists. The bigger the bubble is, the more artists are in the country. We also used color

to encode the gender information to differentiate male distribution and female distribution.
X and Y: Nationality/Country
Color: Gender
Bubble size: Count of artists
Annotation: title = “Male Artists” and “Female Artists”

Graph 2 - Gender Imparity Throughout MoMA History
Tool**:** Tableau Public

Data entry used: Gender (Artist Gender), Date_acquired(the year when the artwork was acquired by MoMA)
We manipulated the data by creating a new column to indicate if a single artwork is created by “only female artists”, “only male artists” or “cross-gender collaboration” Then we used the count of each category by date_acuqired. In order to normalize the data, we used logarithm on the count of artworks.
Graph: Line chart.
X - Date_Acquired
Y - Count of artworks
Color - Gender & Cross-gender Collaboration
Annotation: Important feminist artworks & feminist art movements along the graph

Graph 3 - Artwork Classification Acquisition by Gender
Tool**:** Altair

Data entry used: Gender (Artist Gender), date_acquired (the year when the artwork was acquired by MoMA), Classification (the classification of the artwork)
Created a new column to indicate if artwork was created by female artists or male artists. Then created line graph to show the change overtime by adding date_acquired. A bar graph was also created to compare male and female acquisitions of different classifications of artwork over time.

Graph: Bar chart.
X - 'Gender:O',
Y - Count of artwork
Color - Gender
Column: Classification

Graph: line chart.
X - 'Date Acquired',
Y - Count of artwork
Color - Classification

Annotation: title= "Artwork Classification Acquisition by Gender"

Graph 4 - At what age was an artwork created?
Tool**:** Altair.
Data entry we used: Birthdate (when the artist of the artwork was born), Date_created(when the artwork was created), Gender (artist gender)
We used pandas to clean the data first. birthdate= date_created-birthdate and then group it by each year to calculate the number of artwork by each year.
Graph: bar chart.
X - age
Y - The count of artworks
Column - age
Color - Gender

Graph 5 -When Would an Artwork be Recognized

Tool**:** Altair.
Data entry we used: DateAcquired (when the artwork is acquired by MoMA), Date(when the artwork was created), Gender (artist gender)
We first cleaned the date to only contain year. Then we have: Years_an_artwork_got recognized = Year_Acquired - Year_Created.
We also need to clean gender to make it only contains male, female, or collaboration.
Graph: bar chart.
X - Gender
Y - The count of artworks
Column - Years_an_artwork_got recognized
Color - Gender

4) Why is your solution effective?

Graph 1 - Male and Female Distribution across the World
One purpose of the map visualization is to identify which countries most artists came from. We used the bubble size to represent the number of artists which is straightforward and the biggest bubble is easy to pop out visually, completing our goal.
The bubble size and color encoding also help us realize our goal of telling the difference between the number of female and male artists.

Graph 2 - Gender Imparity Throughout MoMA History
In order to create an informative graph as well as an effective storytelling, we decided to put the year that the artworks are acquired by MoMA on Y axis to let viewers scroll through the MoMA history. In addition, with the line chart indicating the numbers and clear color encoding, users are able to compare the difference between the number of artwork acquired from male, female artists and collaboration artworks. It is easy for them

to tell that male artists keep the lead through the entire history of MoMA and how female artists and collaborative artworks rise through the process.

Graph 3 - Artwork Classification Acquisition by Gender
We used a bar chart and line graph because the bar chart displayed the difference of representation of male and female artists across classifications, but users can drag the year slider to see changes overtime. The line chart shows the change in acquisition of different classification over time overall.

Graph 4 - At what age was an artwork created?
We want to compare the difference of artworks between different ages. By binning them together and using a column, it's easier to see the change over time rather than plot them side by side. It's easier to make a comparison between different genders.

Graph 5 -When Would an Artwork be Recognized
We want to compare the number of artworks in terms of Years_an_artwork_got recognized, and we want to see the difference between male, female and collaboration work. Therefore, we made a bar chart that puts male,female and collaboration artworks side by side, and group them by years_an_artwork_got recognized. We used bin in grouping in order to shrink the horizontal view to a comfortable range.

5) How are you using text to support your visualization? Do you have any narrative structure?

Graph 1:
We used titles to specify the gender and text to show our conclusions.

Graph 2:
There is a clear legend to help users understand what each color represents. The annotation of important feminist art movements marked along the years also enhance the storytelling to help users better understand the history.

Graph 3:

The title communicates what the comparative visualization is going to achieve. Both charts have a corresponding legend and provide interactivity to display information when the user hovers over them. A time slider is also provided for the bar chart.

Graph 4 & Graph 5:
We used text annotation to show what x axis and y axis represent for. We put legend on the right hand side and made it represent gender in order to be consistent across different graphs.

6) How are you using interactivity (if at all)? Why does it support your task? (use the language from the 7 categories we described in the interaction lecture)

Hovering (Select):
We used hovering for tooltips for all of our graphs to allow users to know detailed information about a specific bar/point/bubble. For some graphs, like Graph 1 & 2, user can also select a specific point/bubble or category to focus. For graph 4&5, users can hover over the bar to see the changes of y axis.

Zooming in & out (Abstract/Elaborate):
There are some graphs with large data that we allow users to zoom in & out, for example Graph 1 with the geographical map visualization, it allows the user to drill down to specific countries to see more details. Graph 4 &5 use semantic zoom to allow users to zoom vertically as some collaboration y axis is too short to be selected and view the tooltip.

Time Slider (Filter):
In Graph 3, we allow users to use a time slider to filter down to a specific year to investigate more on the gender differences in the top 5 classifications.

7) What are the limitations of your solution?

Because of the large dataset and complex data format, our original data that we used to is not a hundred percent cleaned. So there might be some bias on this visualization. Our visualization requires users to manually interact with it - you have to select, drag, zoom in and out in order to see the change. A better way is to use animation to automatically display the change without requiring them to actively interact.

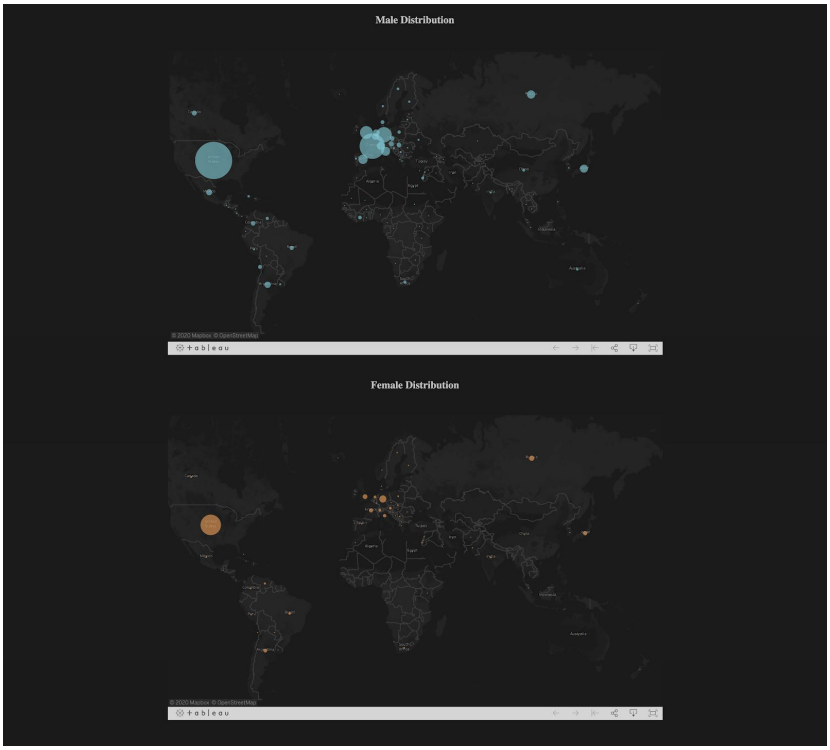
Also, using logarithm on the second graph does help normalize some extremely large values and makes the graph much readable. However, though we have informed users that we use logarithmic scale, there might still create some biases.

Finally, we also struggled to match everything in each graph to be visually consistent.

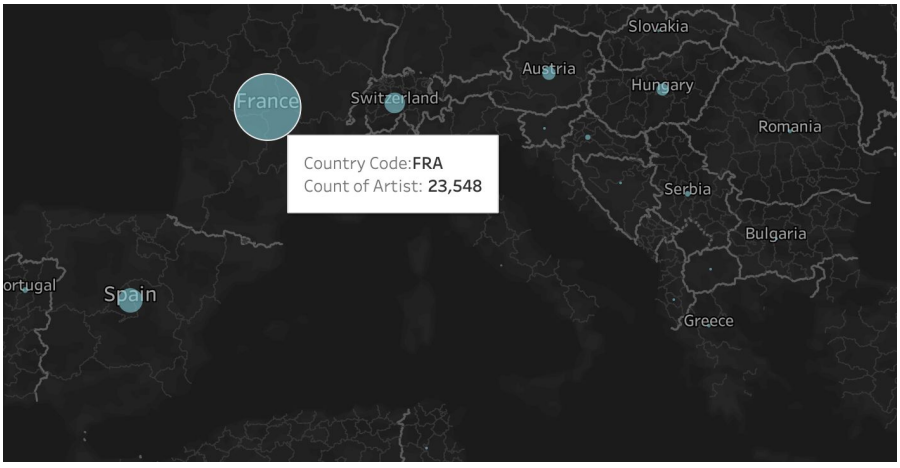
Implementation / Screenshots

Graph 1

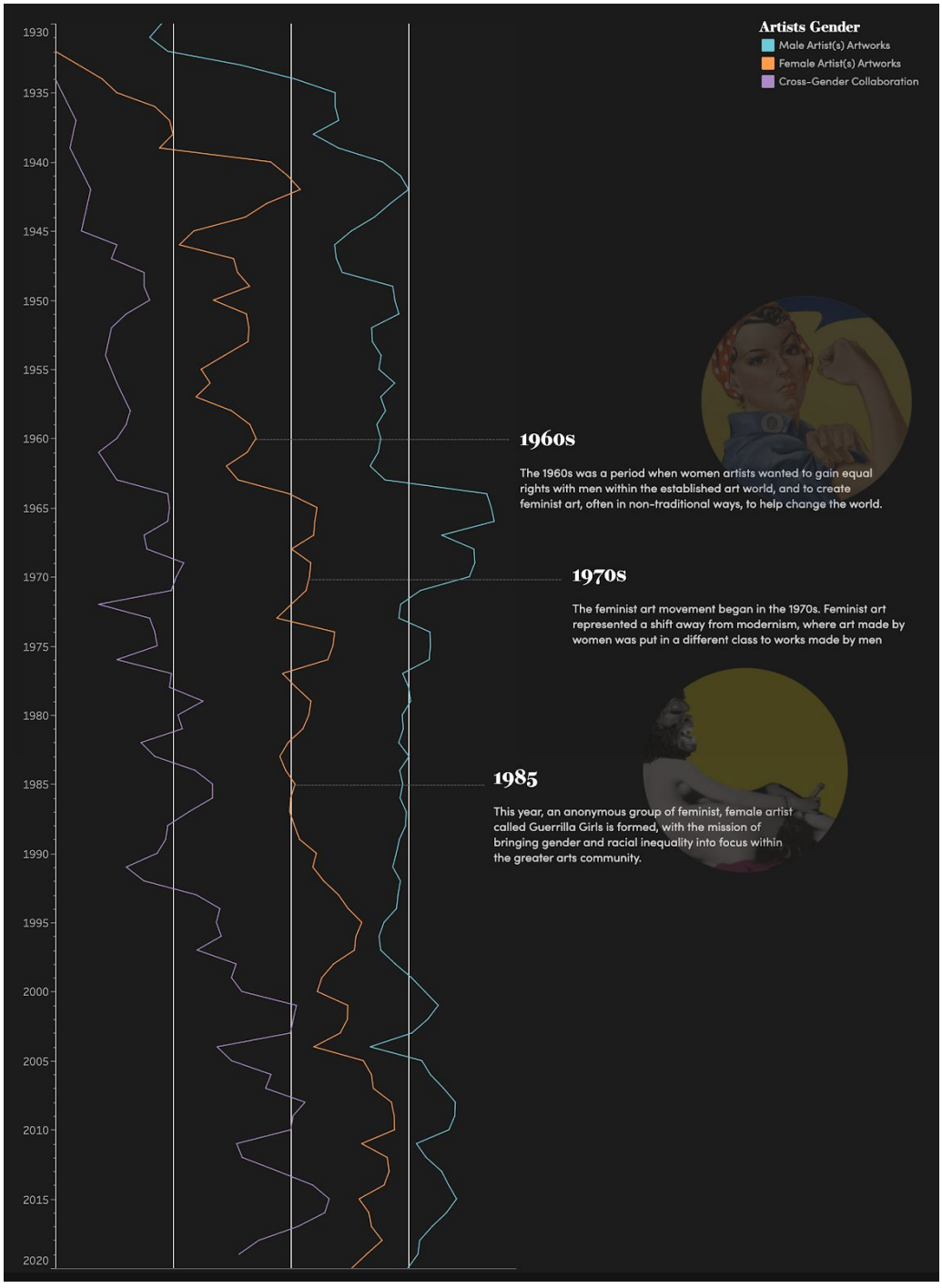
We used the map visualization to show where most artists came from and how the number of artists is different across gender.



The map could be zoomed in and country name and artist number information could be shown when hovered on.



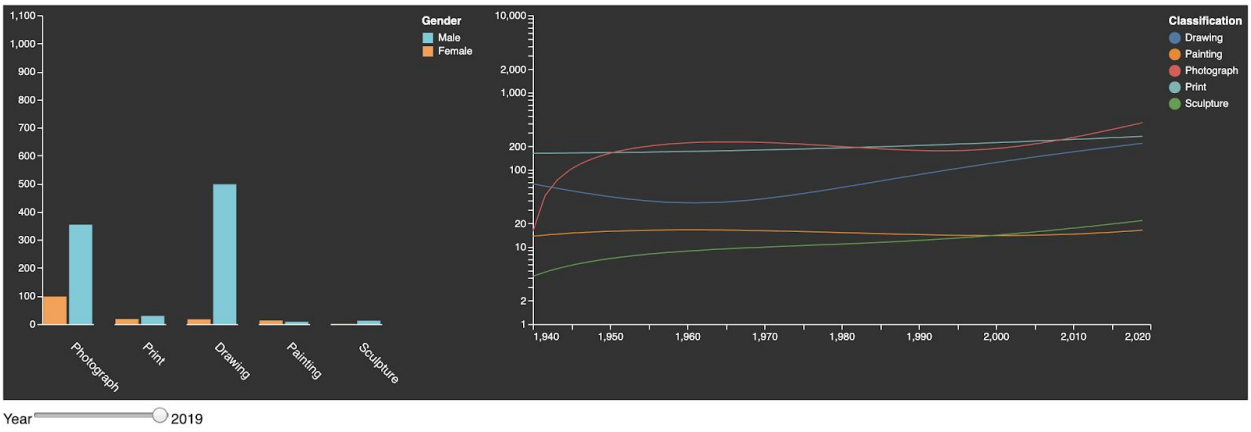
Graph2



We used Tableau Public to implement this graph. The users will be able to scroll through the history of MoMA to see the difference between the amount of artwork acquisition of only male artists, only female artists, and cross-gender collaboration artworks, along with the important feminist movements on the right. They can also hover over a specific point to see the details like the exact number of artworks.

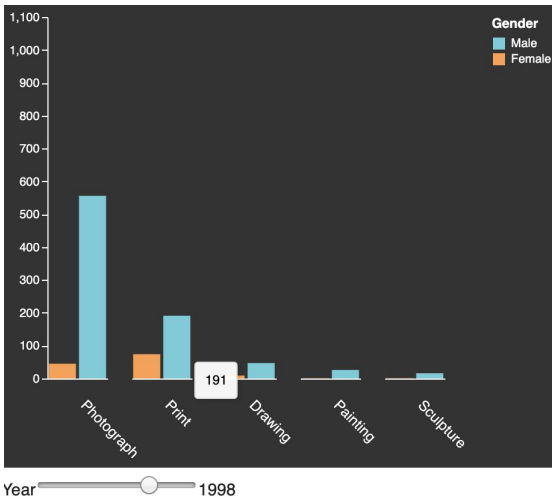
Graph 3

Artwork Classification Acquisition by Gender

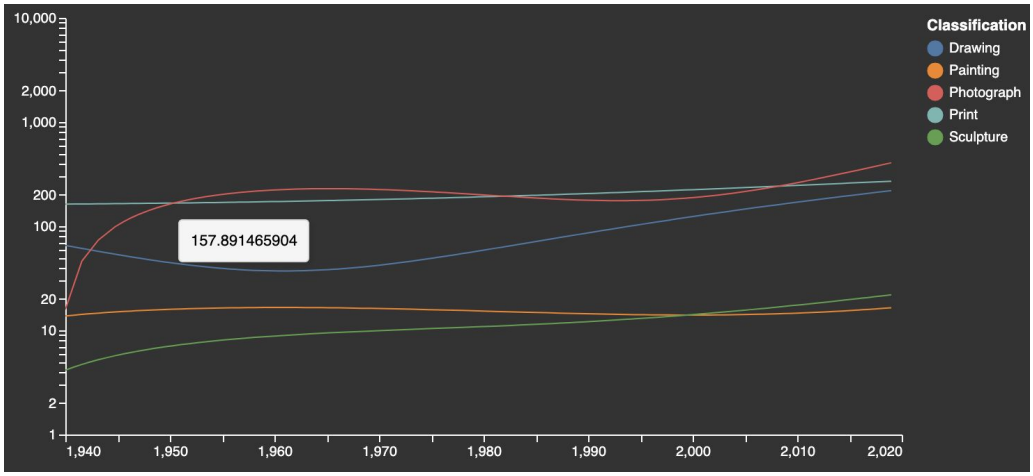


On the left is a bar chart that displays the acquisition rate (y-axis) of the top five kinds of art classifications by gender (x-axis).

The line graph on the right is a line chart that gives a sense of the acquisition rate (y-axis) of these five artwork types over time (x-axis).

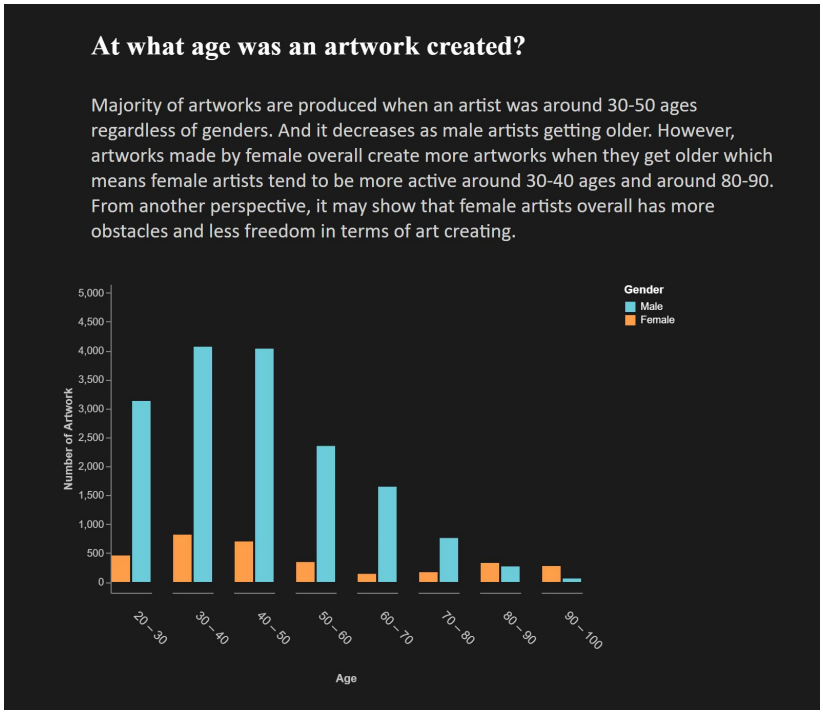


The bar chart has two types of interactivity. The first is a slider that allows the user to investigate male/female comparisons for the acquisitions of each year. The second allows the user to hover over the bar to display how many pieces were acquired for that year; classification; and gender.



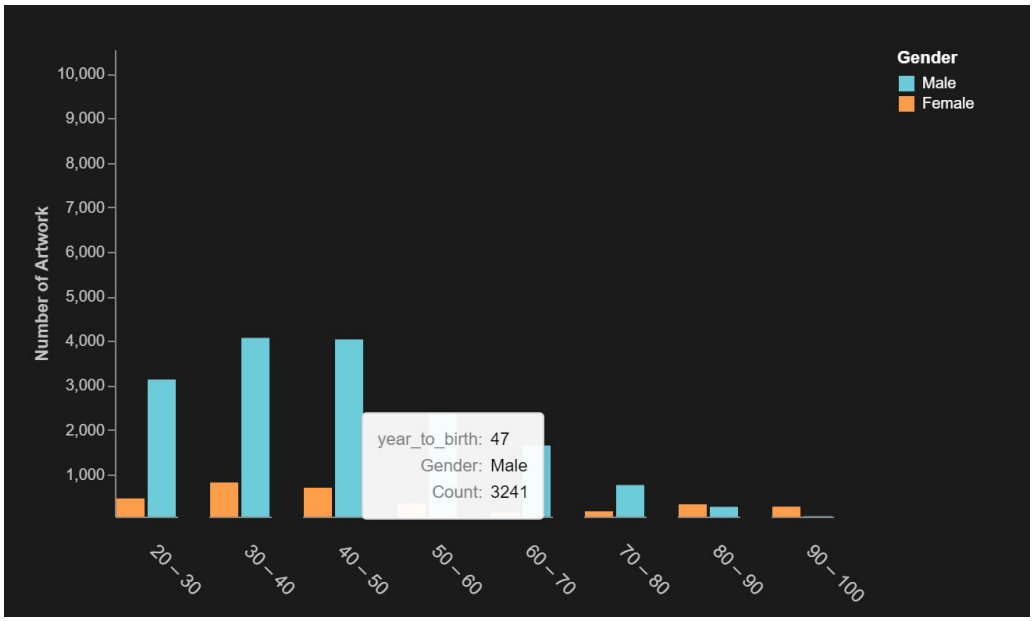
The line chart shows a regression line created from the original scatterplot showing the acquisition of the top five classifications over time. It was originally very noisy, so the visualization was better curated by using a log transform and implementing the regression line. There are two types of interactivity. The first allows the user to hover over the line to follow the acquisition trends. The second allows the user to “zoom in” if they desire to get a closer look at the lines.

Graph 4



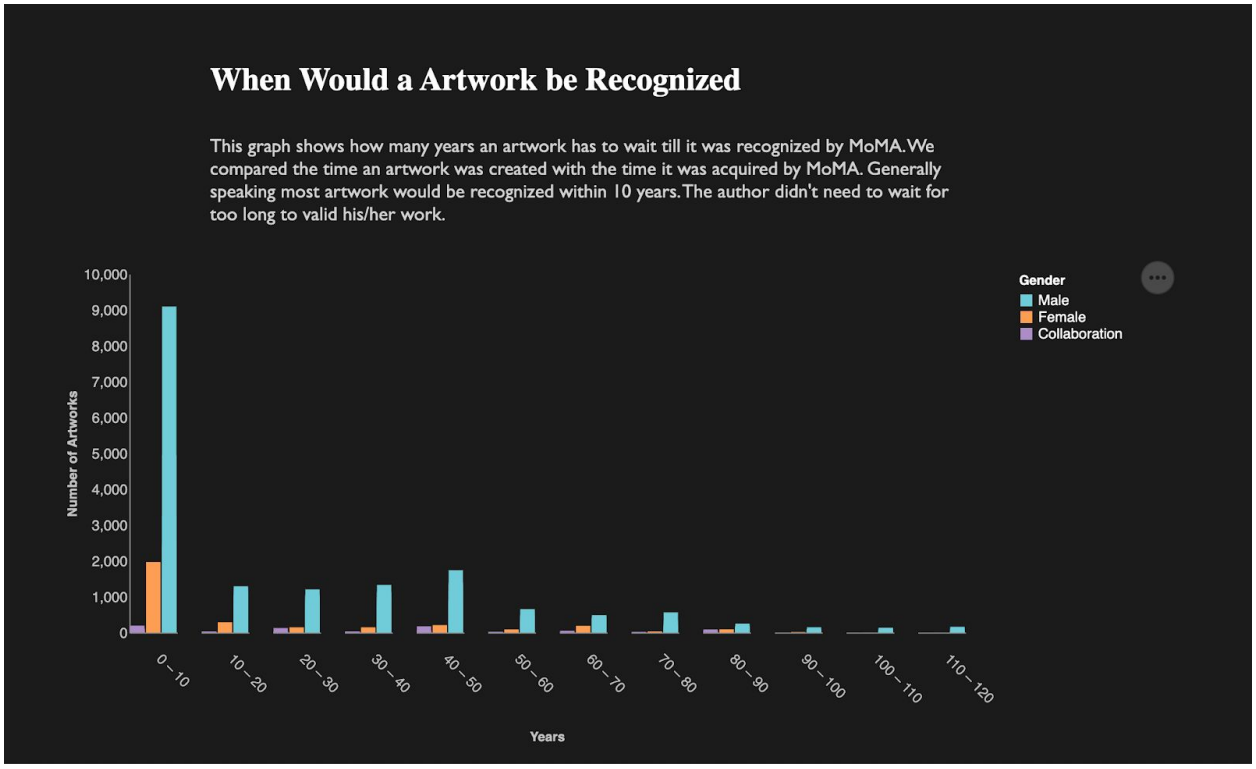
This is a bar chart. The x axis shows the age of the artist when the artwork was created. To calculate the x axis, we used date_created column minus birth date to get the age of the artist. And ages are binned by 10 years intervals.
Y axis shows the number of artworks created in this age. And genders are encoded into different colors.

This graph shows one level of interactivity. First one is a tooltip, hovering over a different bar chart will give you a detail of the number of artworks produced in each age and it's gender. As you hover along the bar,the year_to_birth will grow so you can see more details for different years.

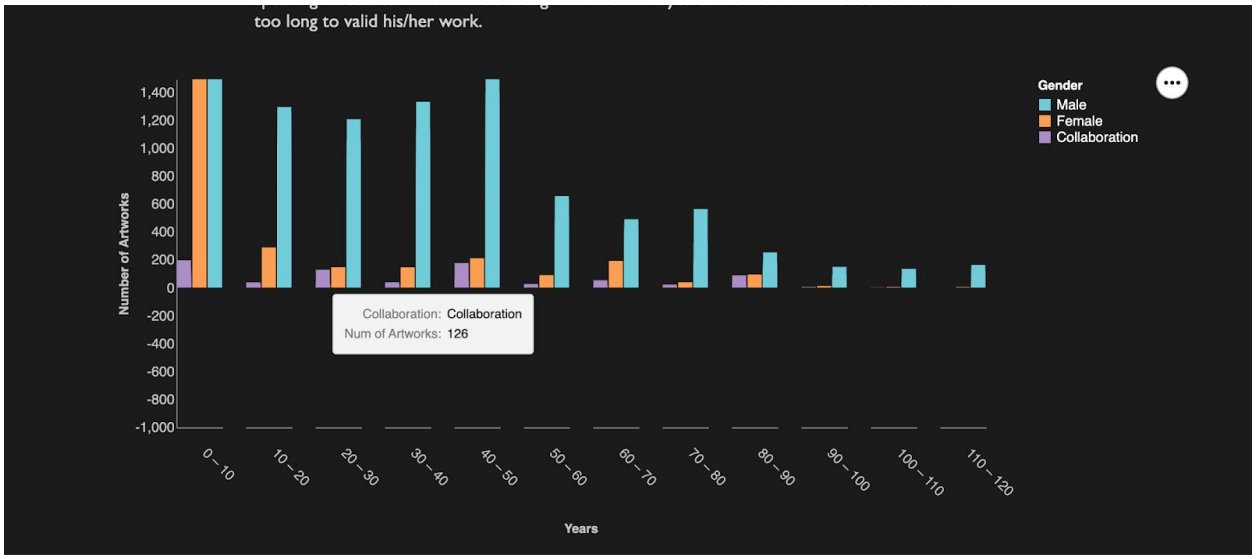


Graph 5

This graph shows how many years an artwork has to wait till it was recognized by MoMA. We compared the time an artwork was created with the time it was acquired by MoMA. Generally speaking most artwork would be recognized within 10 years. The author didn't need to wait for too long to validate his/her work.

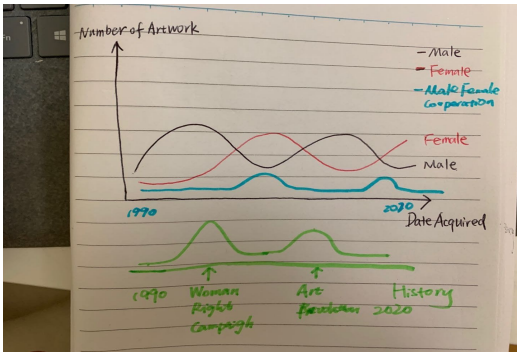


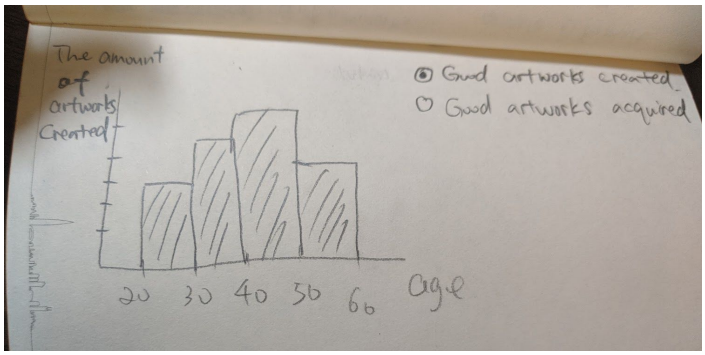
Users are able to zoom in and out vertically. By hovering over a bar, it's going to show detailed tooltip.

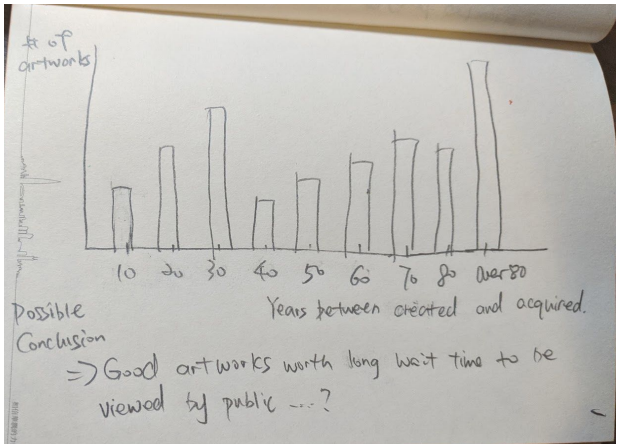


Sketches / Experiments

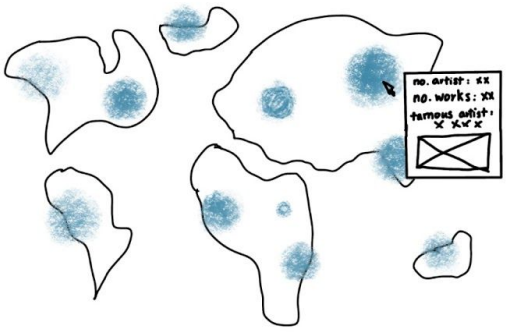
Sketch - Earliest Version

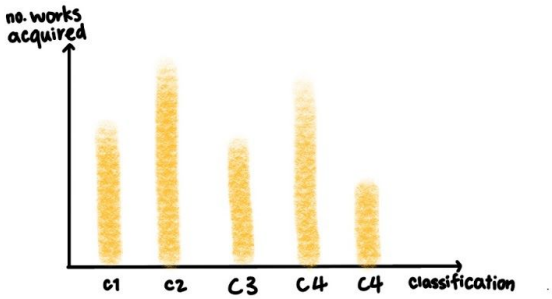




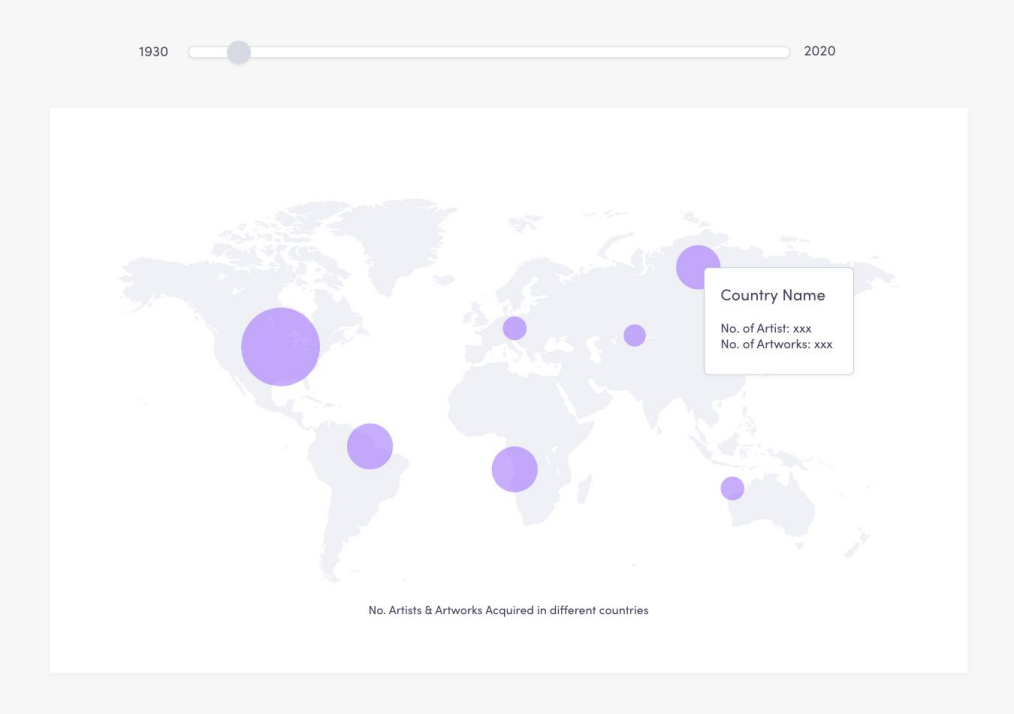


Prototype - Later Versions





Changes of the graph above:
We decided to separate these two charts in our final design, because it was not very clear the exact connection between these two graphs. The geographic graph we tried to focus more on the distribution of the artist nationalities as well as the genders. The second bar chart would be more focused on the diversity of artworks types.

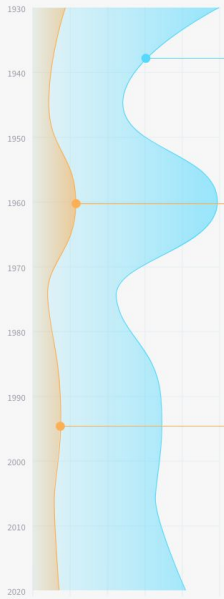


Changes of the graph above:
We did not add the time slider in the end because we realized the change of the number of artists did not show any patterns. The chart without time slider had already completed our domain task and there was no use to complicate the chart by adding a meaningless slider.

Changes of the graph above:
Original idea for displaying artwork based on classification and gender. This idea was pursued and paired with a line chart displaying artwork acquisition by classification over time. One design decision we made after implementing this graph was to limit the number of classifications to five. Originally there were about 14 and many of them did not show any trends of interest either temporally or when comparing genders.

Changes of the graph above:
We decided to get rid of the radio button but put these two bar charts side by side. Since we are not comparing these two graphs, it doesn't make sense to us radio button to view these two graphs.

History of Gender Distribution in MoMA



1938
Xxx Movement: some simple one sentence explaination

1960
Xxx Movement: some simple one sentence explaination

1994
Xxx Movement: some simple one sentence explaination

Male
Female