

Final Report

The Dataset:

For our visualization, we were drawn to the efforts of MIT's "Pantheon" project, which attempted to measure the cultural impact of historical figures across the planet utilizing pageview data from Wikipedia. However, the raw dataset of MIT's project had numerous limitations that made it difficult to use for this assignment. Specifically, data regarding the birthplace and birth year of many of the highlighted figures were incomplete, which would force us to either exclude the figures, or estimate the missing data points independently. Due to numerous issues regarding accuracy and time constraints, neither of these choices seemed viable for the project.

However, the Pantheon Project utilizes an additional data source, the 2003 book *Human Accomplishment*, which is described by the researchers as "a data compilation effort containing information on 3,896 eminent individuals from the arts and sciences who made a significant contribution prior to 1950" (<http://pantheon.media.mit.edu/methods>). This dataset more specifically utilizes the practice of histometry in 15 encyclopedia collections to measure the "significance" of historical figures in the arts, sciences, and humanities. Based on their prominence in these sources, the figures were given index score on a scale of 1 to 100 relative to their field (area of expertise), with 100 being the most referenced figure. By scoring each field separately, the dataset inherently highlights the "big names" of each of the area of expertise in a clean and concise manner. Furthermore, the dataset was far more complete and in-depth than the Pantheon Project in regards key fields such as birth year, birth country, work country, ethnicity, and gender. While not all of the fields were ultimately implemented into the final visualization, they dataset allowed us a degree of flexibility that the Pantheon Project didn't afford.

Of course, any dataset that attempts to measure the "significance" will surely contain limitations and biases that impact the accuracy and "fairness" of the results. Specifically, by utilizing encyclopedia data, the resulting index scores are likely to have biases towards specific people, time periods and cultures, all of which are acknowledged by the researchers themselves. Our visualization focuses solely on European History to minimize difficulties comparing achievement across cultures.

Ultimately the primary domain of our visualization consisted of the figure's significance (index score), their "floriut" (year of flourishing, defined as either the year the figure turned 40 or died, whichever first), and their area of expertise. The attributes in particular will allow the user to under where, when, and what arts were flourishing at different points throughout European history.

Storyboard Sketches:

The Design

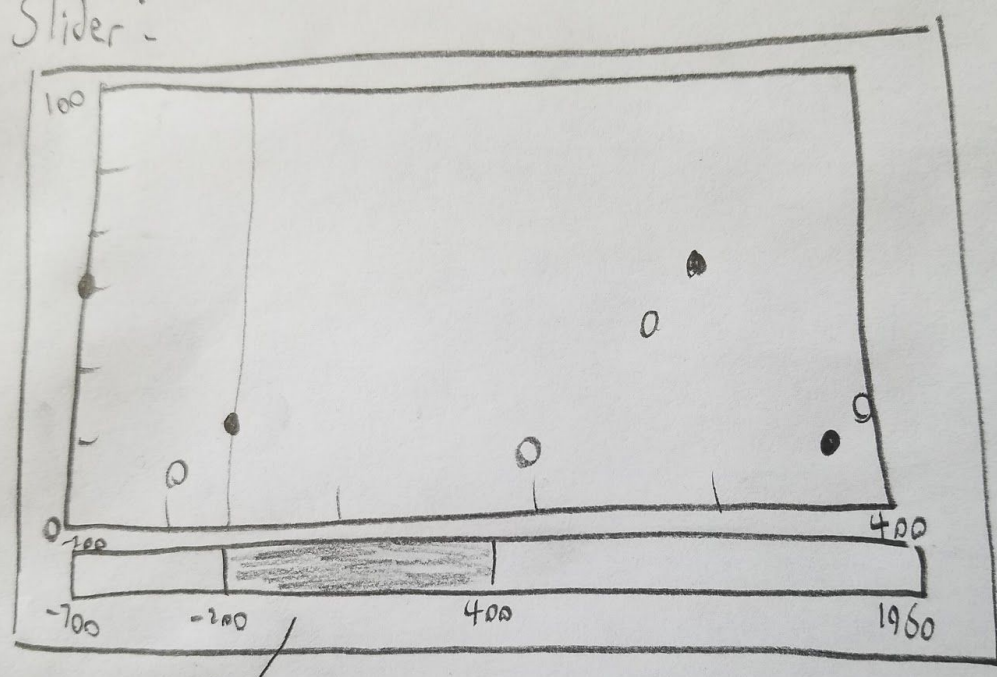
Plot Divided into
different eras or history



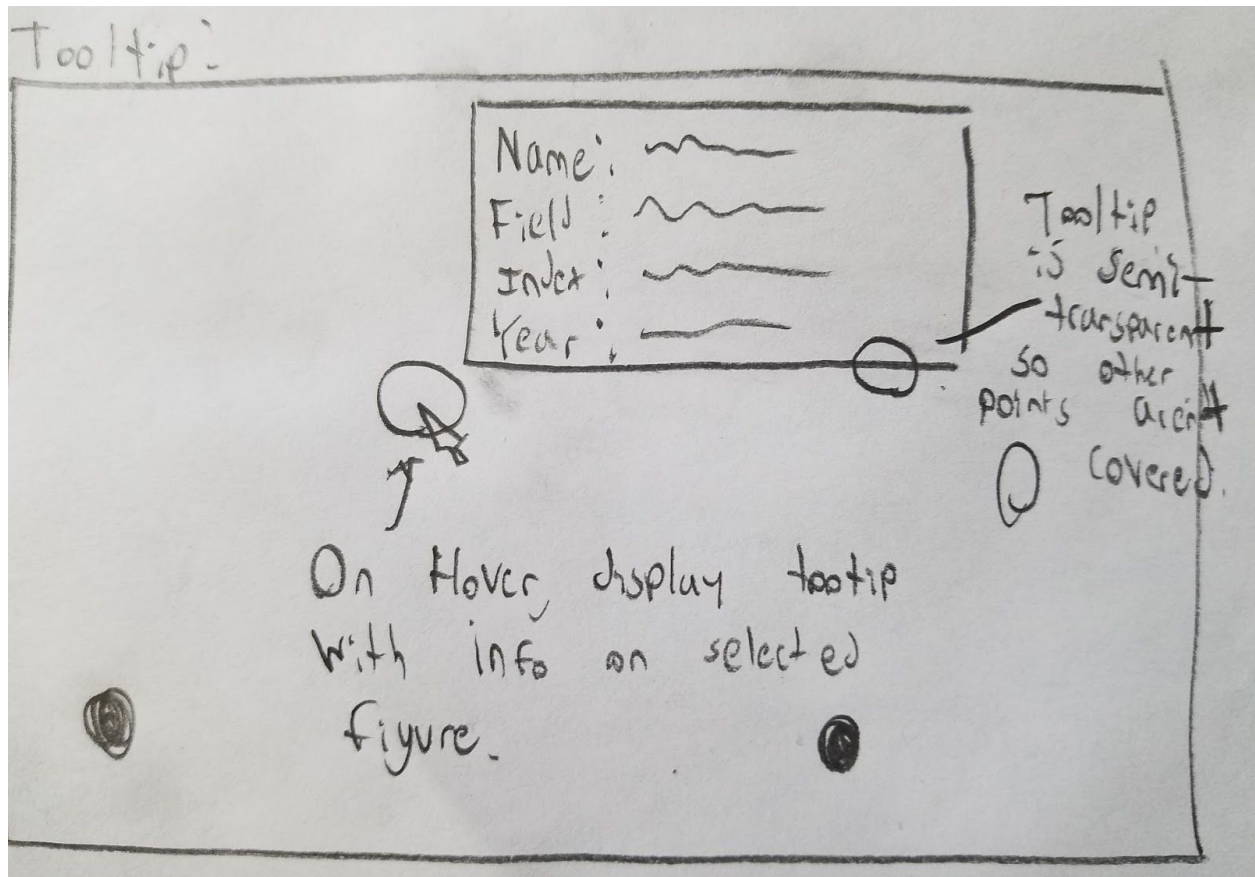
Filter hides
figures from
specific field
from scatterplot

Brush slider

Slider:



Brush slider can be used to select/zoom into a particular area.



Final Visualization Application:

Our final visualization is an interactive scatterplot that displays a collection of significant historical figures in the European Arts & Humanities. These figures are plotted by their "floruit" on the x-axis, and their index score (Significance) on the y-axis. The figures themselves are color coded by their area of expertise and can be filtered to highlight a specific field. Additionally, the scatter plot itself is divided into numerous sections highlighting different eras of European history, such as the Dark Ages and the Industrial Revolution.

The user can further explore the data through two features, the hoverable tooltip and the brush scroll. Upon hovering any point on the scatterplot, a tooltip appears that displays more detailed info about that hover figure, including their name, country of origin, and area of expertise. Additionally, the user can search through and zoom into specific areas of the timeline utilizing the brush scroll placed below the x-axis. As the user selects an area in the scrollbar, the scatterplot dynamically scales itself to that area. The brush scroll also displays a small area chart measuring the increase in significant figures over the whole timeline.

Ideally, the average user will be able to utilize this visualization to learn about points of cultural flourishing (and lack thereof) in Europe as well as highlight the biggest drivers of such accomplishment. For example, the user can utilize the filter tool to highlight trends in European art, specifically how the vast majority of significant artists flourished during the Renaissance Period.

Issues/Trade-offs:

1. Limiting the visualization to solely European History presents an inherently limited perspective on history as a whole. As mentioned before, this decision was made to circumvent difficulties in comparing the accomplishments of different cultures and societies. For example, eras of history are often designated to specific areas of the world (https://en.wikipedia.org/wiki/List_of_time_periods), meaning that the highlighted eras in the visualization aren't particularly relevant to other civilizations, since they had time periods and eras of their own. For example, while the European world was going through the Dark Ages, the Islamic world was experience a golden age.
2. Because of the sheer number of figures, many of the points on the scatter plot have ended up stacked on top of each other. While most figures are still hoverable with some zooming, it presents difficulties for the user if they're looking for a specific individual (even if the user knows the time period that individual was relevant).

Developmental Process:

Once the visualization was formally planned out, visualization generally remained fairly in line with what we originally planned. There were a number of issues however:

1. When implementing the Scatter plot, we played around with the idea of including significant figures in the sciences in the chart. However, when we tried to do this, numerous issues arose. Specifically, the increase in points made the chart look far more clustered, causing numerous figures to be hidden under the masses, even when zoomed in. Additionally, the overload of figures caused the brush scroll to lag significantly, especially when selecting any area post-1600. While some lag still exists in the current chart, it's minimized to the point of being fairly noticeable. Beyond the tradeoffs noted previously, there weren't any major overhauls to our design between the original storyboard and the final product.
2. Due to unforeseen coding difficulties, the filter feature only allows for a single field to be selected at any given time. Initially, this prevented the user from seeing the full chart again once they began filtering figures out. To circumvent this issue, we added a "Show All" button to restore visibility to all the figures. However, there still unfortunately is no way to select any more than one field with the filter box, meaning the user isn't able to only have two fields visible at once, it's either all of them, or one of them. This isn't a major issue, but still worth noting.

Additionally, there were a handful of features we ended up adding as the visualization began to flesh out. Specifically, we decided to include a legend so that the user can easily identify what color corresponds to what field. Additionally, the area chart present in the brush scroll wasn't in our original design either. This chart was added to serve a quick and simple function of showing the user the rise in significant figures that has occurred over the course of European History.

Work Breakdown:

Liam Delaney

Storyboard + description of dataset - 45 minutes

Formatting datasets - 30 minutes

Scatter plot (points, categories, era areas) - 2 hours

Implementing Brush scroll (+ area chart) - 2 hours

Final report - 1.5 hours

Sam Hamburger

Storyboard + description of dataset - 45 minutes

Tooltip - 2 hours

Era labels (+ compatibility with brush scroll) - 1 hour

Filters - 2 hours

Legend - 30 minutes