

***Note: The program should run normally in VS Code (open html with Live Server)**

Domain & Techniques:

For the Assignment 3, I have a dataset about Forbes Billionaires in the past 20 years which I am interested in. I found this dataset when I explore potential topics for Assignment 2 but ended up using another topic for that assignment. I figured that this topic is interesting to explore on and its complexity makes it very difficult to be represented in detail using a single static visualization or even a few static visualizations. So, it is a good opportunity to apply interactive visualization techniques on this dataset. I want to try and implement interactive visualization with two major functionalities: brush and linking, and Direct manipulation (selection and filtering) for Assignment 3, since they are powerful techniques that at the same time easy to use from the users' perspective and they are good match for my dataset. To incorporate these interactive techniques, I would need to have multiple visualizations on the same page, each reflecting one aspect of the dataset (especially for brush and linking).

The dataset about Billionaires on Forbes includes more than 35000 observations, however, since there are a lot of missing data for the years before 2010, I am only using the portion of this dataset within the time range 2010 – 2023, which still contains 26717 observations. The reason why I think this dataset is a good match for the interactive techniques I planned to use is because this it contains a lot of variables. Besides the net worth and industry information, it contains a lot of other information about the Billionaires, from geographical information (country of citizenship) to demographic information (age, gender). To know a domain comprehensively, it is important to know multiple aspects of the data and all these information cannot be effectively represented on one visualization. So, brush and linking becomes a very suitable technique when user what to study how different aspects correlate or how Billionaires distribute across different aspects. Also, as a dataset that includes worldwide information, the variation within variables can be large. For example, it contains 89 different countries, 14 years information and in total 26 different industries. User will need ways to take a closer look at some subset of these different variables. Using selection by pointing and filtering to get dynamically updated visualization that are rapid and reversible can support detailed exploration of the aspect user interested in.

Description of the data domain:

Number of records (observation): 26717

Number of variables: 12

forbes_id, year, rank, net_worth, full_name, age, country_of_citizenship,
city_of_residence, business_category, company, self_made, wealth_status

Data Cleaning before used in D3:

- Filtered data before 2010.
- Unified how the same industry and country are called throughout the form to allow d3 recognize when grouping.
- Deleted not interested columns (e.g. forbes_id and city_of_residence).

Description of the visualization with Storyboards:

Aim of this interactive visualization: Help user to explore potential questions they have on Billionaires. It will be useful if people are curious about how these Billionaires are distributed across countries and industries as well as other demographic characteristics and how these trends changed over time in the past decade.

Sample questions:

1. How are the Billionaires distributed in different countries? How does this distribution change over time?
2. Are there some industries more popular in certain countries compared to others?
3. What are the industries that make more wealth? What characteristics do the Billionaires in these industries have? E.g., Age, citizenship...
4. Are Billionaires age influenced by whether their wealth is self-made or not?
5. What might be some industries that exist younger Billionaires?
6. Who have been ranked as Top 10 during the past 14 years? What about in a specific industry? Who got ranked as Top 10 for the longest period?

Plan to have 4 visualizations, connected with each other.

Visualization1: Proportional Symbols map where size of the circle represents the number of Billionaires in that country. Billionaires who are on the Forbes for multiple years are not double counted.

Visualization2: Line graph with year on the x-axis and sum of net worth of the industry mapped to the y-axis. Each industry is represented by a line with a distinct color.

Visualization3: A heatmap about minimum age among the Billionaires. Rows representing different industry and columns represent geographical regions. It uses a continuous color scale to represent difference in minimum age in each industry x region.

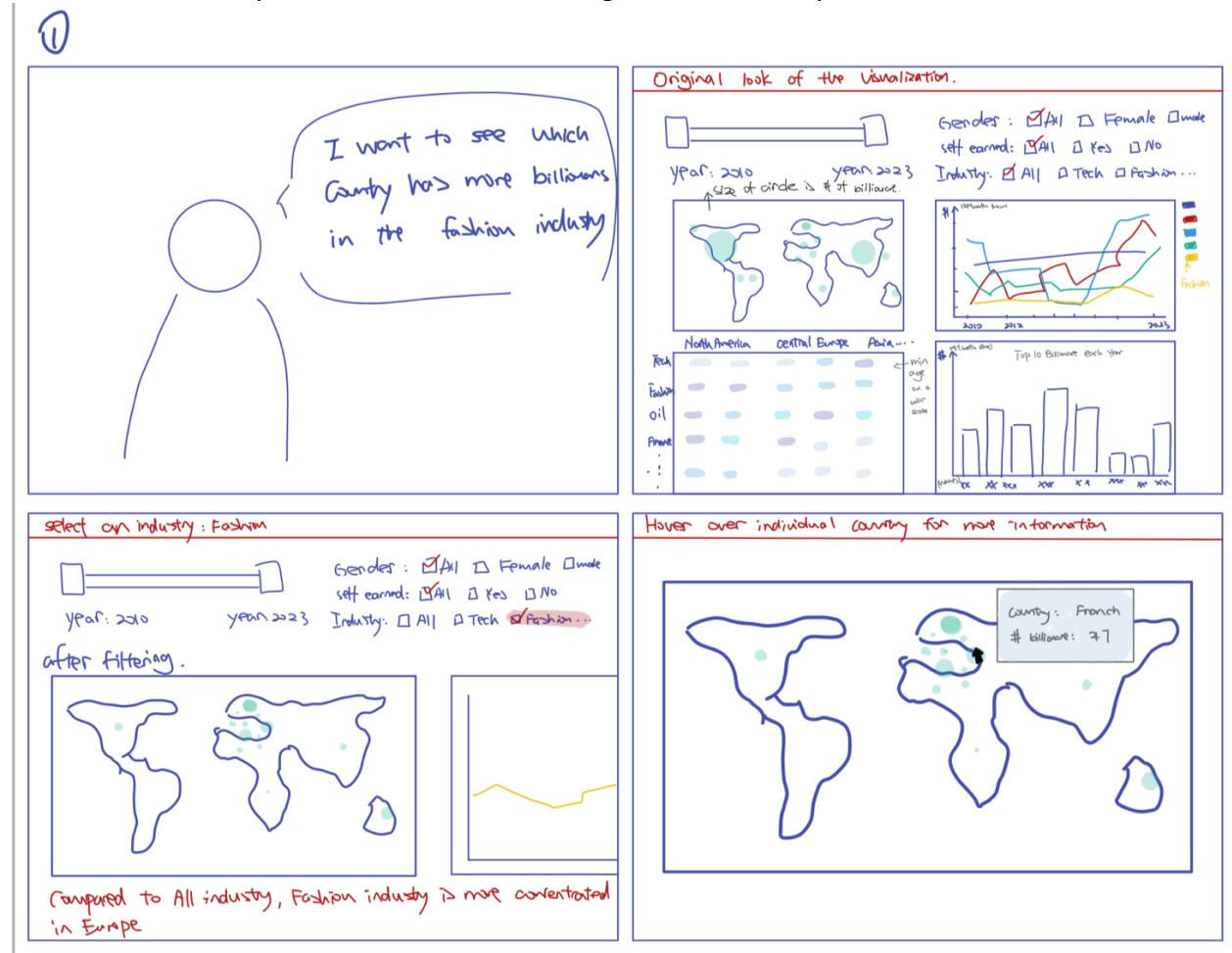
Visualization4: Bar chart with Top 10 Billionaires on the board each year. The color of the bar is mapped to the number of years that person gets in the Top 10. When selecting individual industries, this chart will reflect the Top 10 within that industry and draw less than 10 if there is in total less than 10 Billionaires in the selected industry.

Interactions:

1. An overall brushing slider (2 ends) to control the year range, applied to all 4 visualizations. For the line graph where x-axis is year, the slider acts as “zoom in” as the graph’s x-axis and y-axis range will update accordingly.
2. A series of checkbox for different industries, allows user to select all, select one or select multiple ones. This acts as a filter for Visualization 1, 2, and 4, only showing data on selected industries. However, Visualization 3 will always contain all industries information, but the industries selected will be highlighted.
3. Checkbox for filtering by Male vs. Female for Visualization 2.
4. Checkbox for filtering Self-Made vs. Not Self-Made.
5. Hovering over to view more details in each graph.
6. Click on individual people in Visualization 4 to filter data corresponded in Visualization 1 and 2, highlight in Visualization 3.
7. Hover over the country in Visualization 1 to highlight correspondent region (column) in Visualization 3.

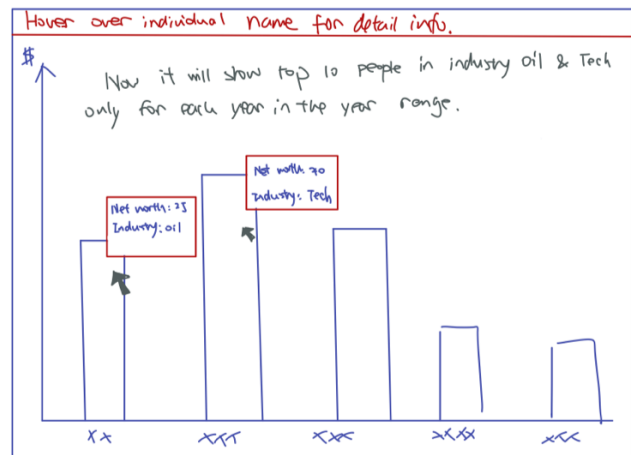
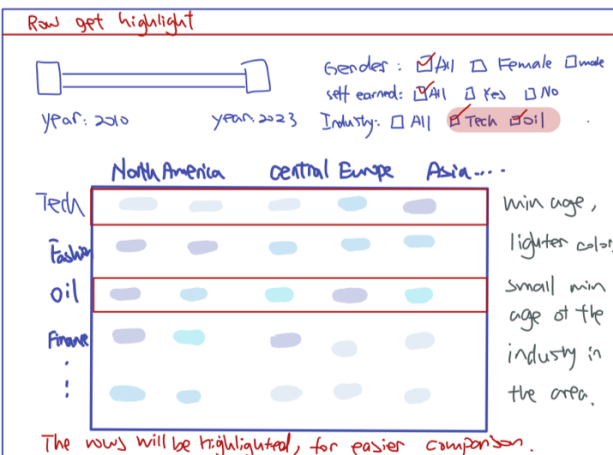
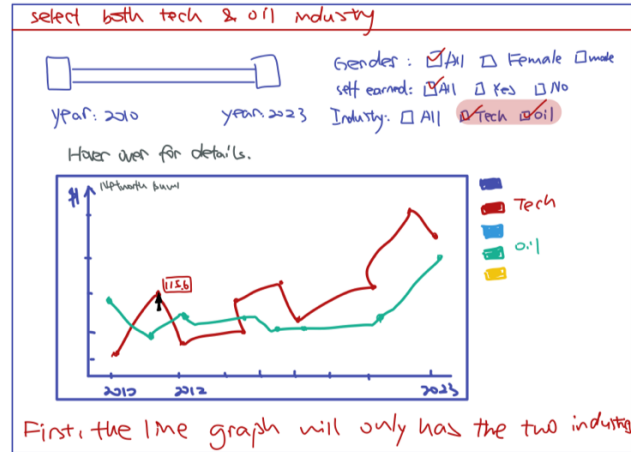
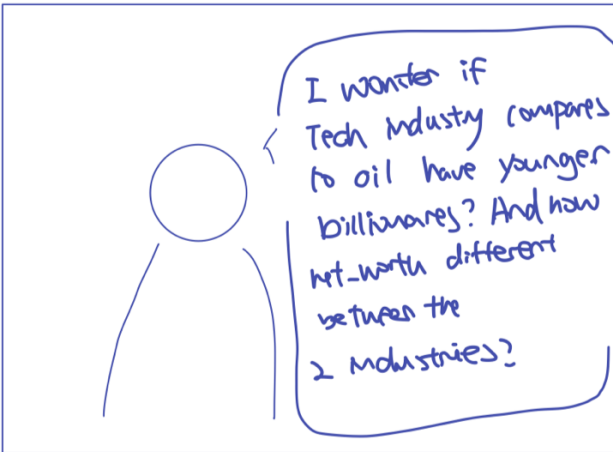
Storyboards:

Here are a few story boards showcase the design mentioned in previous sections.



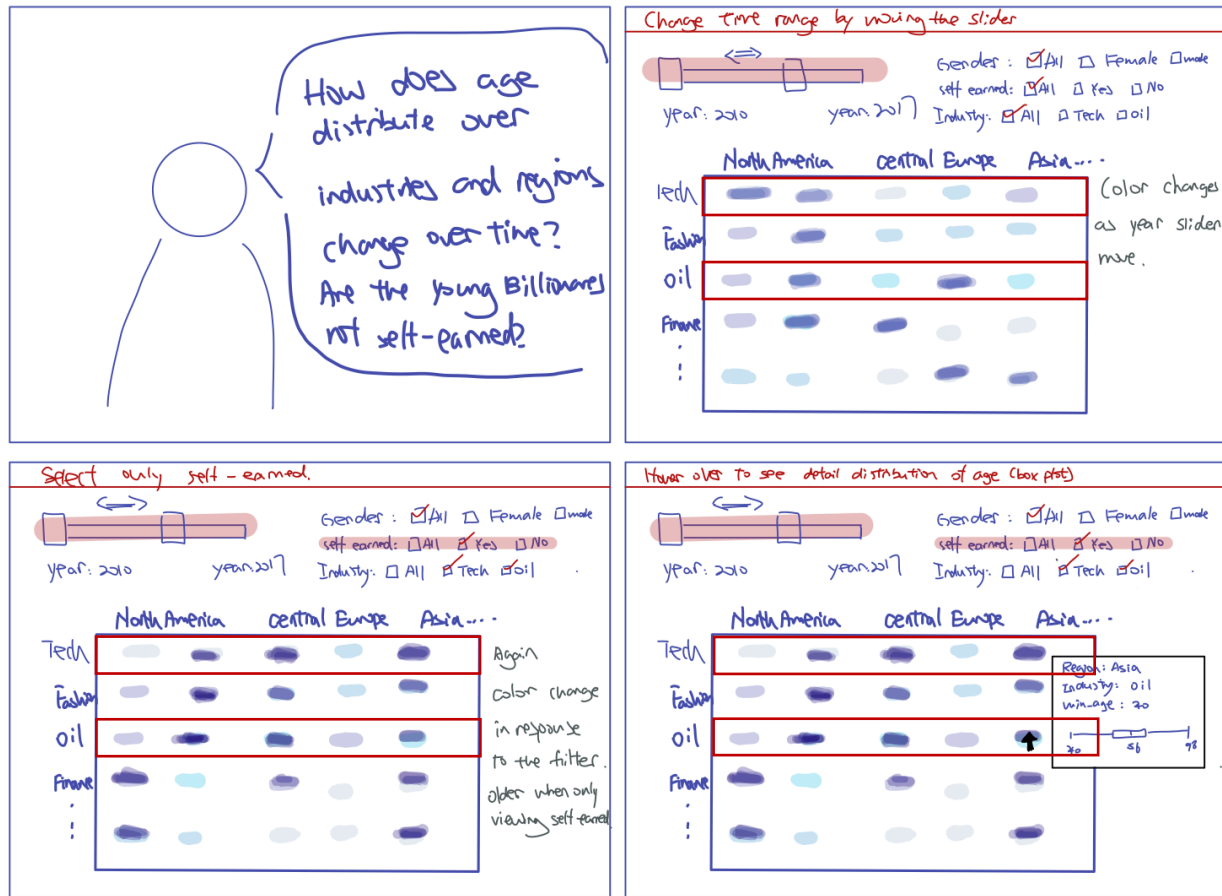
Storyboard 1 demonstrated the use of this visualization to explore industry specific information. In the complete design, selecting specific industry impacts all 4 visualizations, while the storyboard focused on showing how geographical distribution differs between industries. User can see the dramatic change of size of the circles on the map. This change helps user to notice the difference, and through hovering over, they can get more detailed information.

2



This storyboard 2 shows how user can easily use the interactive component to enable comparison by viewing different categories at the same time. After selecting multiple industries, only the selected industries will be shown in the line graph together, easier for comparison. For the heat map on min_age, selected industries will be highlighted, which again supports comparison. The border around the entire row is used for highlighting, avoiding interruption to the original color coding. The bar chart will also be updated, to pick Top 10 for each year within the selected industries. Sometimes all the top ones could be all from one industry if the two industries differ in size/net wealth largely and I think it is better to remain a closer connection with the original ranking feature. User can still get a sense of the net_worth difference from the line graph. And the absence of one industry itself also carries meaningful information.

③



Besides industry, another important variable in this dataset is time. The two-end slider enables user to select the time frame they are interested in. At the same time, by filtering with selections like “self-made”, user can use multiple interactive methods together to get an answer for their specific query. I decided to elaborate minimum age in visualization 3 instead of mean of age is because after exploring the data, I figured that minimum age is a more interesting feature which differs larger among regions and industries. However, to avoid losing the distribution, I planned to add a box plot for each cell to be visible when hovers over it. This provides additional information to support user’s exploration.

Final interactive visualization application & changes:

1. Change on how gender is represented:

In the original design, gender was designed as a filter object that will only impact visualization 2. Originally, I want that it is separately controlled for visualization 2, if user want to see how the wealth is distributed between gender, a new line for male or a new line for female will be drawn. Also, it could draw line together for comparison. However, I later figured that this only makes sense when one industry is selected. Also, this means beside the 26 color already exist for different industry, I need another color coding for

gender if I want to support this comparison. And this design interfered with other interactions for the industry comparison, which is the aspect I want to focus on. The line graph will become hard to read if I want to show comparison between gender and industry at the same time. And if I simply change this to a filter that applies to all 4 visualizations, it is not that useful. Finally, I decided to drop this checkbox, filter design for gender. Since gender ratio is the aspect of gender information that most frequently wanted, I add it as a waffle chart when user hover over the country in visualization 1, which allows them to see the number of Billionaires in each country as well as gender distribution.

2. Change on highlight representation in heatmap:

The heatmap has industries as rows and regions as columns. In the previous design, both row and column can be highlighted in response to user actions (red boarder of the entire row or column). However, when row and column are highlighted at the same time, highlights on both directions can interfere with each other. So, I changed how the industries are highlighted. Now, the selected industry will be moved to the top of the rows and other not selected industries' height will be reduced, making the selection more obvious. By moving the selected industries to the top, selected ones can be closer to each other can making comparison easier.

3. Change on color scale for heatmap:

In the original design, a continuous color scale is used to represent minimum age in heatmap. However, a divergent scale is more application as the real question this heatmap trying to answer is "which industry or region has younger Billionaires". So relative comparison to the mean is meaningful visual information and mean of minimum age can be identified and set as the midpoint.

4. Change on selection based on individual Billionaire (visualization 4):

In the original design, there is a selection interaction based on individual billionaire in visualization 4, in which clicking individual bar in visualization 4 will results in filtering of the dataset. However, I later find that this interaction overlaps with other brush and linking techniques and decided to drop this interaction for consistency in other operations. Also, investigating individual billionaire is not the major focus of this interactive visualization. So, instead of this interaction, I extended the original brush & linking when hover over the country, highlight the Billionaires in bar chart together with the region column in the heat map.

Development process:

This project took me quite long to complete and the start of making different visualizations in d3 is the part that is most difficult. Finding data domain & design visualizations (including brainstorming and drawing drafts of storyboards) took me around 5 – 6 hours. Start to get used to d3 and construct the rough drafts for the 4 visualizations took me more than 15 hours. I had limited experience with JavaScript and HTML before, so it is quite hard at the beginning. Later refine and adding interaction was not that difficult, partially because I am much more familiar

with d3 at that point. Overall, this project took me around 30 hours to complete. I am not sure if the exact hours I estimated are correct, but the ratio between different parts should be correct.

I feel that at the beginning of this project I was really stressed, because d3 seems difficult and I am doing this project alone. However, as I get familiar with d3, the tasks become much more straight forward and predictable, debug took shorter than I expected, and the entire project becomes more manageable.