

Carly Hu (ch862) and Myna Lim (ml2326)

Professor Rzeszotarski

INFO 4310

22 February 2024

INFO 4310 Homework 2

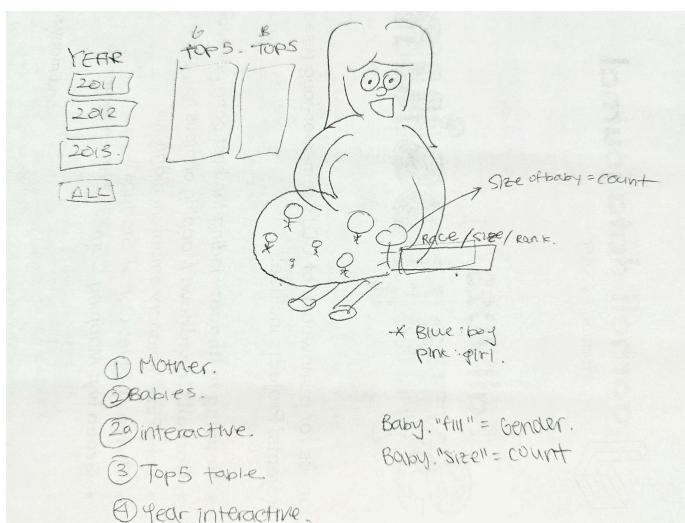
Dataset

For this assignment, we selected a dataset of [Popular Baby Names](#) from the City of New York's OpenData website. This data was collected via civil birth registration, where each row in the set corresponds to a baby name, along with its year of birth, gender, mother's ethnicity, count, and rank. We selected this specific dataset because the size of the data and number of features both fell within a reasonable range that we felt would create an interesting and diverse visualization, without requiring much pre-processing. In terms of specific insights and use cases for our dataset, we wanted to take advantage of essentially every feature in the dataset, allowing a user to see a specific name and instantly know its popularity in a given year.

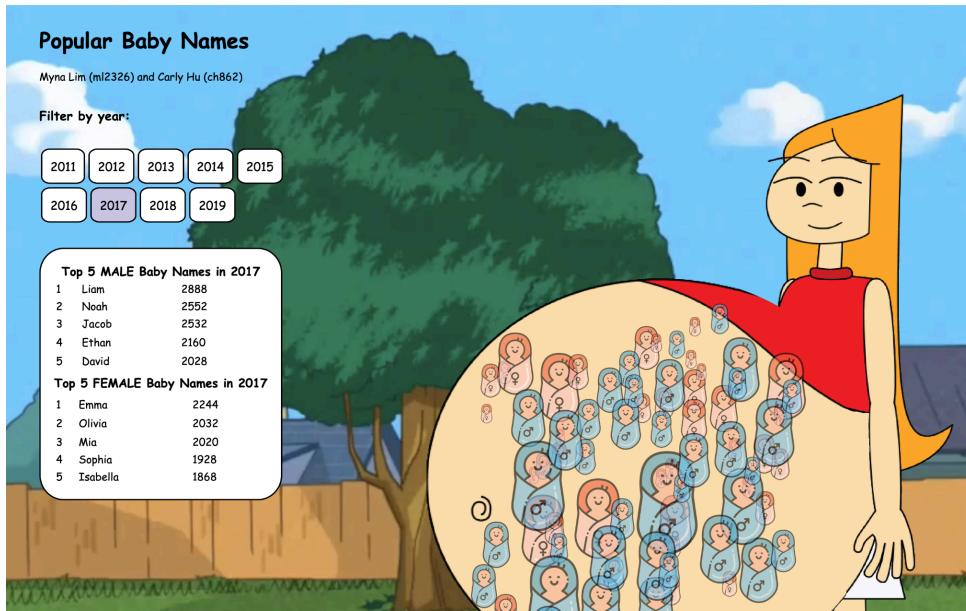
The dataset didn't require a lot of preprocessing because the data didn't have missing values or unmatching formats. However, in the 'Ethnicity' column, there were some inconsistencies such 'WHITE NON HISP' and 'WHITE NON HISPANI,' which we replaced to have the same values. We repeated this process with 'BLACK NON HISP' and 'ASIAN AND PACI.' Because the dataset had both columns 'Child's First Name' and 'Ethnicity,' there were numerous repeated names that are listed in multiple ethnicities. For simplicity, we merged the dataset with repeated names with different 'Year of Birth' and 'Gender.' Then we created columns with different ethnicities and stored their 'count' values for the corresponding ethnicity. We also created an 'overall_count' column that adds up all the count the counting for all ethnicities. The 'rank' column was changed to ranking the names for the year.

Planning & Final Visualization

Storyboarding:



Final Interactive Visualization:



Our final visualization ultimately stayed very loyal to our initial design plan:

- We thought that an engaging and creative way to visualize baby names (as opposed to a more typical word cloud or line chart), would be to plot each name as a baby icon on a comically large pregnant woman's stomach.
- This would draw users' attention and encourage them to interact with the baby icons, which would vary in color and size based on gender and count respectively, with a hover label that includes name, rank, and count.
- On the left side of our design, we included a simple table that displays the top 5 names by year and gender, which allows users to easily access and digest a subset of the data before diving into the main interactive baby visualization.
- Finally, since the data is organized by year, we also wanted to add a level of interactivity that allowed users to filter the data (both on the table and woman's belly) by year.

Development Process

Throughout our development process, we made several design decisions to enhance our user's experience and make our tool exciting to interact with. While searching for images of pregnant mothers (for the background of our interactive baby visualization), we stumbled upon an image of Candace from the children's show *Phineas and Ferb*, and decided to fully commit to this cartoon as the theme for our visualization. This involved adding the show's backyard as the background, using Comic Sans as the visualization's font, and using round corners and white backgrounds for our buttons, table, and labels.

For the Top 5 Names table, we kept the design relatively simple: users can see the top 5 names ranked by gender, along with the corresponding counts for each name. These two tables were cleanly combined into one small panel to clearly distinguish it from the main interactive visualization. For our buttons, we set the button bar the same width as the ranking table for visual alignment, and chose a light purple color to highlight the selected button.

The main interactive visualization features the 30 most common baby names for the selected year—this quantity was chosen because it resulted in the optimal amount of baby icons appearing on the visualization so that users could interact and browse the data, but not be overwhelmed by the number of interactive elements. As stated before, we implemented the icon size to vary based on the count of babies under a given name, and gave all of the icons an opacity of 0.6 to allow users to see through any overlap. Using `d3.forceSimulation()`, the babies within the visualization were given an animated entrance to enhance the overall visual experience and encourage user interaction and exploration.

In terms of actual user interactivity, we implemented a one-line label that appears when users hover over the baby visualization, to quickly provide all essential information (name, ranking, and count). The positioning of these labels was customized to respond to the icon location to prevent the label being cut off from the SVG. Moreover, users are able to interact directly with the baby icons, allowing them to freely drag any icons across the display—though it is not an essential level of interactivity, this permits the users to expose any icons that are potentially obscured by others.

Limitations

One of the limitations in the visualization was due to the setup of our SVG elements, which we were unable to make dynamic. This means that users may encounter inconsistencies in the display's layout across different browser widths, potentially impacting the visual experience. Another limitation worth mentioning pertained to the lack of a comprehensive “Total” data option for the table and interactive visualization. Instead, we only provide data segmented by year. Lastly, we noticed later into our development process that there were comparatively fewer data points available for later years in the dataset, which could potentially skew interpretations or analyses that heavily rely on longitudinal trends.

These limitations underscore areas that would require further refinement and enhancement in future iterations of the project to optimize the visualization’s usability.

Contributions

Our approach to this project was very collaborative—we very frequently updated each other on our progress and often passed a visualization back and forth, making incremental improvements on both ends until we were satisfied with it.

- **Carly** added the hover label to the interactive visualization and implemented the ranking table. The most time consuming tasks were properly positioning the hover label, as well determining how to format the table properly.
- **Myna** created the baby visualization and also implemented the entire year filter button functionality, including making the baby data points interactive such as forceSimulation and drag/move. She also photoshopped Candace's belly to enlarge it.