

## Project 2 Baby Names and Births in NYC(2011-2017)

### Final Report

Team Members: Yuanhao Zhu(yz2969), Yuhan Hu(yh959), Ziqu Wang(zw522)

#### A. Description of the data

Data Source: [Popular Baby Names by Sex and Ethnic Group for New York City](#)

Description of the variables:

Variable Name	Description	Variable Type
Year of Birth	Year of Birth	Number
Gender	Sex of the baby	Plain Text
Ethnicity	Mother's Ethnicity	Plain Text
Child's First Name	Child's First Name	Plain Text
Count	Number of babies with this name	Number
Rank	Frequency of baby names in descending order	Number

Data Source: [Percent Live Births by Infant Sex and Mother's Race/Ethnicity for New York City, 2007-2017](#) (Nativity)

Description of the variables:

Variable Name	Description	Variable Type
Birth Year	The year of birth	Number
Sex Of Infant	The infant's sex	Plain Text
Ethnicity of Mother	Mother's Ethnicity/Race	Plain Text
Births	The number of infants born	Number
Percentage	Percent of male and female births within the year, and within the race/ethnic group	Number

#### Data Processing:

- We used two datasets to explore the name trends and the number of births of newborn babies, each dataset is loaded in an async function.
- The original datasets contained varied ranges of data, we kept the data from 2011 to 2017 in both datasets to improve usability.

- We capitalized the columns in both datasets and deleted duplicating rows to maintain consistency.
- In the Natality dataset, the categories of Ethnicity are in slightly different wording, we modified the wording to be the same as that of the Names dataset and combined the “Other” and “Unknown” categories as “Other/unknown” to keep consistency.

## **B. An overview of your visual design rationale**

We have created three types of charts for our visualization: a bubble chart, a bar chart, and a line chart. We used different color hues for different ethnicities, and implemented pink/blue color hues for the gender.

A bubble chart is selected to display the most popular names by ethnicity and gender, as it allows the name texts to be displayed within the bubbles. However, as we need to implement a fixed size for the bubbles to keep consistency and the length of the names varies, long names would escape the bounds of the bubbles. Additionally, we employed the force simulation for the bubble chart because the animation makes the chart more interesting and organized. The X axis represents the four different ethnicities, and the Y axis represents the rank of the names.

A bar chart is used to visualize the number of births each year by ethnicity. We used the same color scale for the ethnicities to keep consistency, the numbers are displayed on top of the bars for better readability.

When we are designing the bubble chart, we wanted to provide more interaction and information for the users in case they are interested in further exploring the data points, we then decided to implement a line chart to display the yearly trend of a particular name within that ethnicity from 2011 to 2017. Users can have an overview of the number of changes of a given name. The names can be selected from the bubble chart. Additionally, we have an input box that allows the user to type in the names they want to explore and we intentionally put the interactive elements close to the graph they interact with.

## **C. An overview of your interactive elements and their design rationale.**

The interactive elements include: a year filter, a rank filter, an user input box, and the interaction between the bubble plot and the line plot to show specific name’s popularity trends. Since we need both charts to be responsive to the year filter, and also to make the interaction more intuitive, we made the “Year” filter into a global slider and put it above the two charts to make it more accessible. We also made the “Rank” filter to a dropdown selector providing a selective range from 1 to 10. Because we think a dropdown can reduce clutter and can be placed closer to the bubble chart to indicate its interaction with the bubble chart.

The interactive element on the bubbles is that the users could click on any of the bubbles or texts to see the count trends of that name over the 7 years in the line chart below. To provide additional customization for users to explore the data on their terms, we implemented an input box for users to type in any name they want to see the trend of that name among all ethnicities.

#### **D. The story**

The visualizations are designed to explore the popular baby names and the number of births in New York City. Specifically, we want to enable the exploration of the naming trends(in NYC) based on year, ethnicity, and gender. In addition, to provide a little more context, we provided the births data by ethnicity which could enable us to see the trend of the number of births over the 2011-2017 time period.

From the data visualization, we discovered that different ethnicities have some unique preferences on choosing baby names. Some due to language differences, such as “Isabella” which is a Spanish name, is the most popular female baby name among hispanic populations (It was top 1 popular from 2011 to 2016, and top 2 in 2017). Male name “Moshe” is only popular among white non hispanic population, and not any single baby from other ethnicities used this name, likely because it’s a Hebrew name.

We also find there are some names that are favored across different ethnicities, and will stand the test of time. Such as “Ethan” and “Jayden” for male babies, “Mia” and “Emma” for female babies, those names are always on the top rank list from 2011 to 2017 across different ethnicities. While we type in those names in the dialog box and observe the line chart, we could see the absolute count of those names seems still slightly skew to certain populations due to higher counts. Combining the bar plot of Number of Newborn Babies above, we could see the new-borns have bigger fractions of white and hispanic babies, and less of asian/pacific islander and black babies. For “Ethan”, we could see asian/pacific islander people have outstanding interest in this name, even though it’s popular among all the populations.

Some general trend over time:

Names with consistently high popularity such as Michael, David for baby boys and Olivia, Emma, Sophia, and Madison for baby girls.

Names with increasing popularity such as: Liam, Kayla,

Names with decreasing popularity: Elijah, London, Emily

#### **E. Outline of team contributions to the project**

Yuanhao Zhu:

- Made graphical illustrations of babies, UI elements. Made eye-follow animation for the baby illustrations.
- Mainly made bubble chart and line chart. Made the bubble chart and the line chart interactive.
- Wrote the final report (partial)

Yuhan Hu:

- Implemented data cleaning and processing
- Developed the bar chart, the year filter and the rank filter
- Designed the layout of the visual elements including the graphs and filters
- Wrote the final report

Ziqu Wang:

- Contributed to the skeleton of bubble chart, line chart.
- Helped to adjust CSS style
- Wrote the final report (partial)