

Baby Name Trends - CS3300 Report

Arun Pidugu ap639, Junie Khang jk2235, Gillian Boehringer ggb39

Our data is from the Social Security Administration's website. We used two different datasets: national baby name data- where the data for each year is stored in separate .txt files, and state baby name data- where the data for each state is stored in a separate .txt. For the national data, in each .txt, we have the following variables: Name (name of the baby), Gender (female or male), and Count (the number of babies with the name we are looking at). For the state data, in each .txt, the variables were: Name, Gender, Year, and Count. The data was originally in the form of .txt files without any headers so we wrote a python script to generate the headers and write the new data file into .csv files. For the state level data we specifically had to filter out everything except for the most popular male and female name in 1915 and 2015 and wrote this data into the appropriate csv files. In order to "combine" all of the csv files into one for easier parsing, we used d3-queue. Additionally we used the us.json map file from the lecture notes in our map charts, as well as created another dataset by hand which mapped each state abbreviation to its corresponding ANSI code.

When mapping the data, we took different approaches based on what we were trying to show:

With the visualization of number of new baby names per year, a simple bar graph worked best in order to make our point about more variability in names over time. The x-axis was for the year, and the y-axis was the number of new names added in that year. The color of all the bars are the same because we were not trying to have each individual bar stand out and wanted the focus to be on the overall trend.

For the pop-culture visualization, we decided it would be best to do a line graph connecting different points. The line graph seemed like the best way to visualize a name's noticeable reaction to popular culture, so the viewer can follow along with the previous years and actually see the jump in popularity. We created a scale for the x-Axis for the years 1913-2015, and we scaled the y-axis as the percentage of births for the specific name being looked at over the total female births in that year. For example, each data point for the name Rachel is (year, number of babies named Rachel year / number of babies born in year). This was done in order to account for varying population as the number of births increases per year, particularly the baby boom.

For the biblical names, we created a bubble chart in which we linearly scaled values along our x-axis by year (from 1915 - 2015) and scaled our values along the y-axis by the specific names (John, Andrew, etc.). The size of our circles was determined using a square root scale. The larger the percentage, the greater the area of the circle, and the size of each circle represents the percentage of boys named that particular biblical name (like the percentages in the pop-culture line graphs). We thought that using a bubble chart was appropriate because of the three variables we wanted to show: names, year, and percentages.

For the maps we found the most popular baby names by state in 1915 and 2015 for both males and females. We then used the same color palette from our biblical names chart to shade in the states with the same name in the same color. This is different for every chart, for example the color used for “Mary” in the 1915 map is used for “Olivia” in the 2015 map, in order to maintain a consistent color scheme. After this the most popular name in each state was added to each of the respective maps using the centroid points of the different svg paths outlining each state. For smaller states the position of the text was modified manually for clarity.

In terms of color for all visualizations, we decided to stick with darker muted earth tones in order to allow the viewer to understand and have the ability to view them without all of the colors being too overwhelming. This was a natural decision due to the business of our visualizations- particularly in our Biblical names bubble chart. In each of our visualizations, we were not trying to get the viewer to look at a specific point or area- rather, to notice trends and patterns, so we avoided bright colors that would lead to an unintended focus on a particular aspects. Our graphs different aspects don't rely on one another but are for comparison, so we decided to ensure that each aspect of the graphs was a different color for ease of comparison.

Story:

Naming your child is the first thing that happens to them. It is the root of their identity for the rest of their lives. We decided to look into how people choose these names that carry so much weight. Our data set of the top 1000 male and female baby names for every year starting in 1913 until 2015 was the root of our exploration into these trends.

We began by comparing the baby name lists in general, noticing that there was a continuous increase in new names each year as we moved closer to the present. This is displayed in our first bar chart. We decided to add on to this demonstration with the four maps which demonstrate the difference in popular names by state during 1915 and 2015. This allows the viewer to begin understanding how large of a difference all of the new baby names shown in the bar chart make throughout the country. Following that, we decided to look at this data from the perspective of parents- they have to get these names from somewhere. That's when we thought that comparing the popularity of names of Popular Icons would give us something interesting, and it did. As displayed in our line graph, Pop-Culture names have a tendency to become more popular after incidents occur that bring attention to the Pop-Icon with that name.

Lastly, we looked away from current events and back at classic literature, particularly the Bible. We decided this through discussing popular boy names that have been fairly constant over time, realizing that a lot of them came from the Bible. This made sense to us, but we also learned as we developed the graph that the overall popularity of using Bible names to name a baby boy has become lesser as we move closer to present day.

Overall, our data demonstrates that the continuous influx of new baby names in the top 1000 is due to the desire of parents for their children to have unique names. Initially, we did not expect the theme to be centered around the uniqueness of the names in each year's data set.

As we began to see the trends fall into place, it became easier to continue to develop our idea around the developing uniqueness of the data sets for each year. Concluding in a cohesive and well rounded view of the trends of changing data and exploring possible contributors to those trends.