CSC 466 Lab 1 Report: Modern Sociocultural Baby Name Trends in the United States

Martin Solomon Hsu, mshsu@calpoly.edu
Lana Mai Huynh, lmhuynh@calpoly.edu

Abstract

In this exploratory data analysis, we examined datasets of newborn baby names in the U.S. covering the period from 1880 to 2014. We guided our analysis through the context of three distinct social and cultural phenomena that we hypothesize may have influenced modern newborn naming conventions: 1) the effect of the television show *Game of Thrones*, 2) regional differences in spelling "Ashley," and 3) changes in gender distribution by name. After reshaping the data to answer our queries, we identified three significant findings: an effect of Game of Thrones on the popularity of certain names present in the show, an organized regional divide on the spellings "Ashley" versus "Ashly" between the northern and southern US, and ten names that significantly changed in proportion from male to female or female to male. We further analyzed patterns and trends in these findings and explored possible sociocultural explanations.

I. Introduction

The act of naming a baby is one of the most important decisions that a parent can make. Thus, they probably put a great amount of thought into such selection. Where do parents get inspiration to name their child? Do they want to name it after their role model? Perhaps they want to name them after their favorite singer or artist? Do they want to use a classic or unique spelling of a name? How are these decisions influenced by the current culture of the time? In this report, we investigate these baby name trends to get a better understanding of how external societal and cultural influences play a role in baby naming.

II. Dataset Description

We examined two variations of a tabular dataset containing information regarding baby names in the united states, spanning the period from 1880 to 2014:

- *NationalNames.csv*. This file contains information about the total number of babies with a given name born each year from 1880 to 2014 in the United States. The variables in this dataset include *Id*, *Name*, *Year*, *Gender*, and *Count*.
- *StateNames.csv.* This file contains state-by-state information about the frequency of baby names in the years 1910 to 2014. The variables in this dataset include *Id*, *Name*, *Year*, *Gender*, *State*, and *Count*.

III. Research Questions

Within the scope of the baby names data, we asked three distinct questions regarding sociocultural naming trends in the United States:

1. ***Game of Thrones*:** In which years did certain names unique to the popular television show *Game of Thrones* peak in their popularity during the period between 2010 to 2014, inclusive?
2. **Variations of "Ashley":** For each state, which phonetically identical spelling variation of the name "Ashley" was most popular during the period between 2010 to 2014, inclusive?
3. **Changing Gender Conventions:** Throughout the period between the years 1900 to 2000, inclusive, which names changed the most in proportion from female to male, and male to female?

We additionally aimed to explore sociocultural explanations behind trends identified in investigating the history behind these questions.

IV. Methods

Data transformation and computations were performed in Python 3.10, using the Pandas package. Visualizations were created primarily using the Matplotlib package, with choropleth maps created with the assistance of the Geopandas package.

i) *Game of Thrones*

For this question, we wanted to transform our data to identify the most popular year for unique newborn names that were based on the *Game of Thrones* franchise from 2010-2014. We wanted each row to be identified by the selected name and contain the year between 2010-2014 that the name was most popular among newborns.
Firstly, we identified ten names from the *NationalNames.csv* dataset that corresponded to major characters in the *Game of Thrones* television show: Arya, Bran, Brienne, Daenerys, Sandor, Talisa, Theon, Tyrion, Tyrell, and Tywin. We then filtered the data to observations containing those names with a *Year* variable falling in between 2010 and 2014, inclusive. Taking into account that there may be both male and female people with these names, we grouped by name and year and took the total within these groups. We then reshaped the data from long to wide so that each row was a unique name and each column was a unique year, with each value being the count of newborns corresponding to the appropriate name and year.

ii) Variations of "Ashley"

To answer this question, we wanted to transform our data to identify the most popular spelling of "Ashley" for each state during the years 2010 to 2014. We wanted a table where each row was identified by a single state, and contained the most popular spelling variation of "Ashley" for that state.

Similarly to the *Game of Thrones* data restructuring methods, we identified 13 common variations of the name “Ashley” from online sources. We filtered the data in *StateNames.csv* to these variations, and then further filtered to names with a *Year* variable falling in between 2010 and 2014, inclusive. We then grouped by state and name and took the total within these groups, and then identified the variation of “Ashley” with the maximum count within each state. This table was joined to a Geopandas table containing boundary geometry from the US Census Bureau in order to plot the data on a choropleth map.

iii) Changing Gender Conventions

For this question, we wanted a list of the five names that changed the most from male to female, and the five that changed the most from female to male, between 1900 to 2000. We additionally wanted to see the magnitude of the change for each name, and the proportions across each year.

To obtain the male and female proportions of names between 1900 to 2000, we first filtered the *NationalNames.csv* data into two separate tables: one table of female newborn names from 1900 to 2000 inclusive and another table for male newborn names from 1900 to 2000 inclusive. We immediately joined the two tables together so that each row was a unique name and year, and contained information on the count of female newborns and the count of male newborns with that name in that year. From this information, we calculated two new fields: one for the proportion of newborns with the respective name that were female in the respective year, and an equivalent field for the male proportion. We pivoted the female proportion values wider so that each row was a unique name, each column was a unique year, and each element or value was the proportion of newborns with the respective name that were female in the respective year.

Finally, for each name, we took the average of the female newborn proportions in the first five years (1900-1904) and an average of the proportions in the last five years (1996-2000), and then took the difference between the two. We took the top 5 names with positive differences, which represents the newborn names that changed the most in proportion from male to female between 1900 to 2000. We additionally took the top 5 names with negative differences, which represents the newborn names that changed the most in proportion from female to male between 1900 to 2000.

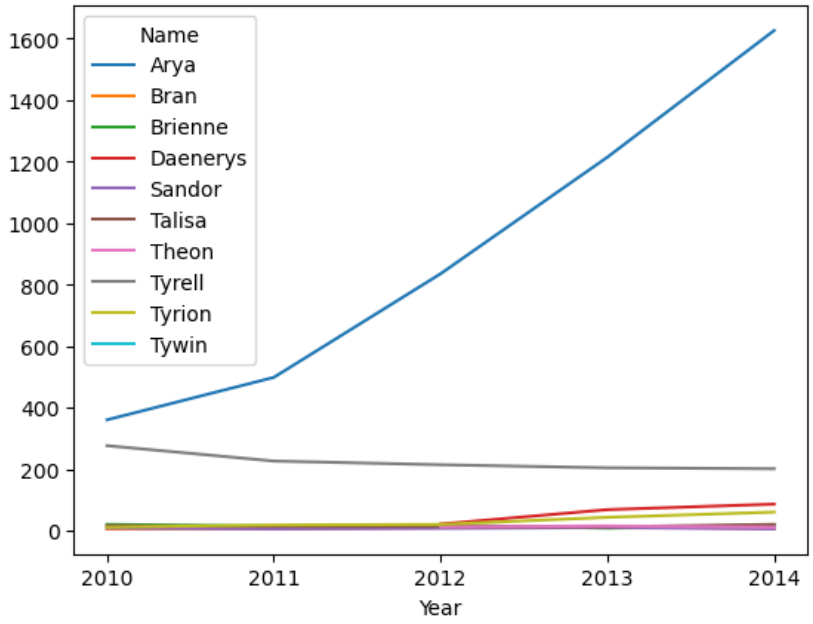
V. Results

i) *Game of Thrones*

Table 1. Name of Game of Thrones character and their most popular year in newborn names

Name	*Year*
Arya	2014
Bran	2013
Brienne	2010
Daenerys	2014
Sandor	2013
Talisa	2014
Theon	2013
Tyrell	2010
Tyrion	2014
Tywin	2014

Figure 1. Popularity of Game of Thrones character baby names over years 2010-2014

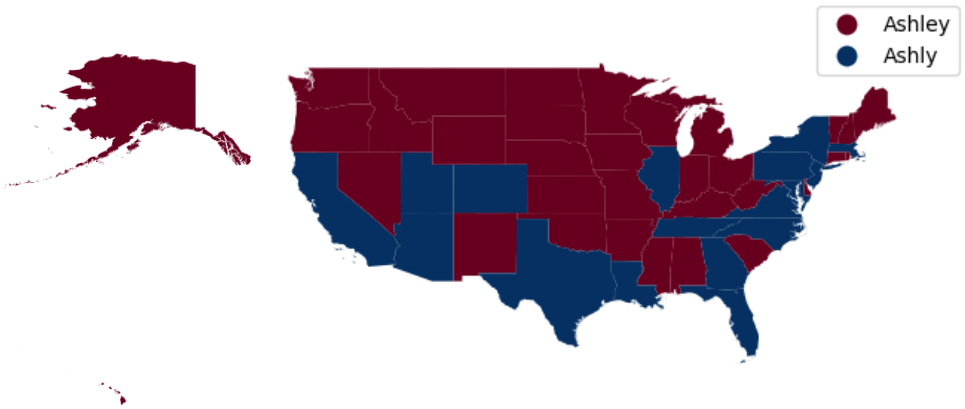


As we can see from figure 1, the baby names of *Game of Thrones* characters stayed relatively consistent with slight increases except for one. The name “Arya” steadily increases from the

year 2010 to 2011, but it increases at an even faster rate from 2011 to 2014. The name “Tyrell” steadily decreased from 2010 to 2014. The other names mostly plateaued from 2010 to 2012 and somewhat increased from 2012 to 2014.

ii) Variations of “Ashley”

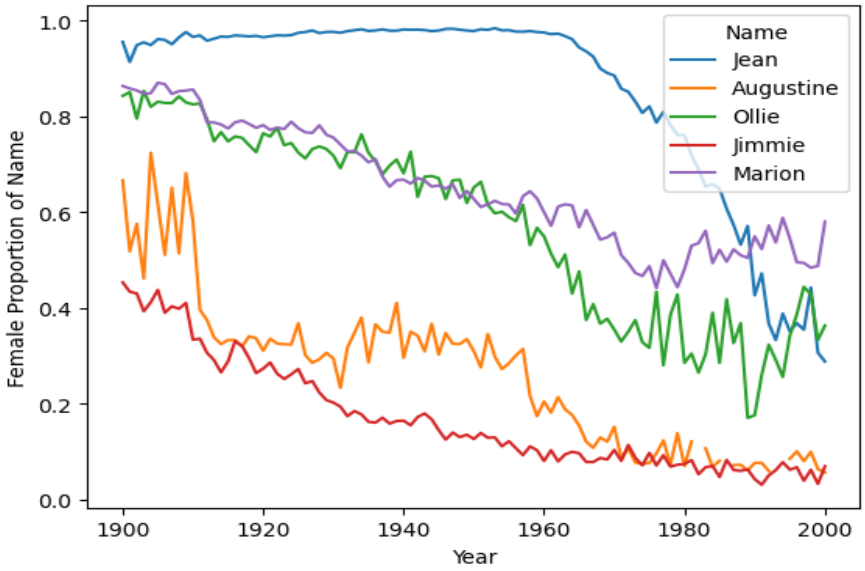
Figure 2. Map of most popular spelling variation of “Ashley” by state, 2010-2014



From the choropleth map, it appears that the name “Ashley” is the most popular variant in Northern states, while “Ashly” is the most popular in Southern states with exceptions in the Northeast states.

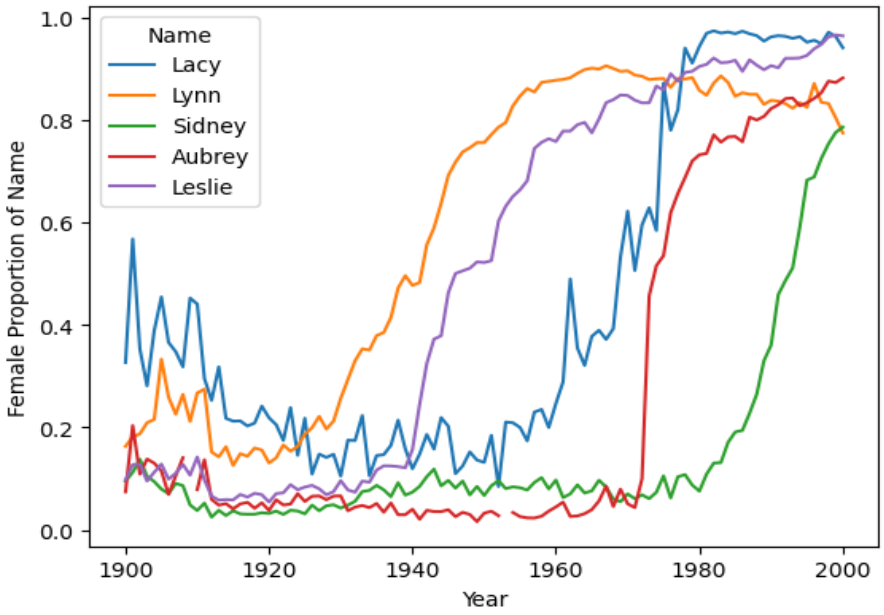
iii) Changing Gender Conventions

Figure 3. Top 5 US baby names with greatest change in proportion from female to male over years 1900-2000



From the visualization in Figure 3, we can identify that the five names that changed the most from proportionally female to proportionally male were Jean, Augustine, Ollie, Jimmie, and Marion. With the exception of “Jean,” the change appears to be a steady, relatively linear decrease. The name “Jean” appeared to contain a clear majority of female newborns in 1900, which decreased to a more even split between male and female in 2000, with the decline starting between 1960 and 1970. The names Augustine, Ollie, and Marion were also majority female newborn names in 1900, and slowly decreased to a more even split by 2000. In contrast, the name “Jimmie” was never majority female, but rather was majority male with a more even split between genders. However, it also decreased gradually in female proportion until it reached a clear male majority by 2000.

Figure 4. Top 5 US baby names with greatest change in proportion from male to female over years 1900-2000



As seen in Figure 4, the five names that changed the most from male to female in the time period given are Lacy, Lynn, Sidney, Aubrey, and Leslie. In comparison to names that changed in proportion from female to male, the names that changed in proportion the most from male to female did not do so gradually, but rather saw drastic change from majority male to majority female, often within one or two decades. The most prevalent change can be found in the name “Aubrey,” where the name held a clear male newborn majority until the 1970s, where it suddenly held a clear female newborn majority by 1980.

VI. Discussion

i) *Game of Thrones*

The name that stands out the most in our visualization is clearly “Arya.” After the debut of the first season of *Game of Thrones* in 2011, this name became significantly more popular. In the show, “Arya Stark” is one of the main protagonists, and she is greatly loved by the fandom. In fact, she consistently tops popularity rankings for characters from the franchise. Thus, the reality of the name “Arya” growing in popularity could be attributed to her adoration.
There are 8 seasons of the show *Game of Thrones*, and consequently, there are a myriad of names. We narrowed the list of names to use in our analysis down to those that were uniquely from the show and had a decent popularity. Thus, our analysis is not representative of every single name in the show, but the most important names in this context.

ii) Variations of “Ashley”

Because we were looking at the variations of “Ashley” by state, counts of the special spellings of the name were much more limited. This made it more difficult to see the differences between the spellings of the name.
The variant “Ashly”, from English origin, is a unisex baby name that was at peak popularity in the 1980s and 90s, meaning “dweller near the ash tree clearing.” The British mostly settled in the West, South and Northeast, and those regions are where the variant “Ashly” was most popular. It is difficult to tell whether or not these facts are related to one another.
The two most common ways to spell “Ashley” are “Ashley” and “Ashly.” Because the other variants of the name were not as popular, it is even more difficult to see the minute differences between the variants at the state level. In the future, we may want to consider excluding the original spelling of “Ashley” from our analysis in order to hone in on the other variants.

iii) Changing Gender Conventions

The most notable difference between our two visualizations is that the top male to female names decreased steadily, while the top female to male names drastically increased at a certain point in time. In the case of the name “Aubrey,” underwent a gender switch because of the song “Aubrey” (1973) by the group Bread. This event influenced societal norms and shifted the gender of the name “Aubrey.” It is our hypothesis that a single event influenced the switch from male to female names, while gradual societal changes created the shift from female to male names.

VII. Conclusion

Through the utilization of different data manipulation techniques, we were able to identify multiple sociocultural trends regarding baby naming. The effect of the television show *Game of Thrones* seemed to increase the naming of “Arya.” Regional differences appear to exist for the spelling of “Ashley” and “Ashly.” The top male to female names decreased steadily, while the top female to male names drastically increased at a certain point in time. These trends from all three of our research questions uncover what the types of societal changes that exist throughout history.

Bibliography

[1] Alternate name speller - used this to see what all the different spellings of “Ashley” are

- https://www.bestlittlebaby.com/alternate-name-speller.aspx?name=Ashley

[2] Game of Thrones character list - used this to base what names we should include in our analysis

- https://en.wikipedia.org/wiki/List_of_Game_of_Thrones_characters

[3] Significance of variant “Ashly”

- https://www.momswhothink.com/baby-name/ashly/

[4] Explanations behind “Aubrey” name gender change

- https://www.nancy.cc/2013/02/25/baby-name-aubrey/

[5] Most popular Game of Thrones characters audience poll

- https://www.ign.com/faceoffs/game-of-thrones-face-off-who-is-the-best-character-in-westeros