Baby Name Trend Analysis in the United States

Marjella Ernst[a]

[a] *Stellenbosch University, Stellenbosch, South Africa*

Abstract

Abstract to be written here. The abstract should not be too long and should provide the reader with a good understanding what you are writing about. Academic papers are not like novels where you keep the reader in suspense. To be effective in getting others to read your paper, be as open and concise about your findings here as possible. Ideally, upon reading your abstract, the reader should feel he / she must read your paper in entirety.

1. Introduction

This report analyses trends in US baby names over the years 1910 until 2014. The analysis is supported by data on famous singers and actors names in the same time period. In the course of the analysis, persistence and volatility of name popularity are being explored over time, to analyse whether names that are popular in a given year tend to remain popular three years in the future.

To do this, I calculate and visualise the Spearman rank correlation for the 25 most popular names in each year, comparing them to the top names three years thereafter. This provides a measure of how stable or fast-changing baby name trends have been and whether these trends have become more volatile after 1990.

In addition, I investigate sudden surges in name popularity and cross-reference these with publications in music and film. Specifically, I cross-check, whether names have spiked in popularity following the success of certain songs, singers, actresses or characters. By combining name data with cultural references, the analysis aims to identify potential drivers of naming trends and provide guidance for selecting toy character names that have the potential to be liked by future baby generations.

Email address: `marjella.ernst@web.de` (Marjella Ernst)

2. Methodology

2.1. Data Overview

In a first step, data is loaded in from the data folder and stored in variables.

```
\begin{Highlighting}[]
\NormalTok{baby\_names }\OtherTok{\textless{}{-}} \FunctionTok{readRDS}\NormalTok{(}\StringTok{
\NormalTok{billboard }\OtherTok{\textless{}{-}} \FunctionTok{readRDS}\NormalTok{(}\StringTok{"d
\NormalTok{hbo\_titles }\OtherTok{\textless{}{-}} \FunctionTok{readRDS}\NormalTok{(}\StringTok{
\NormalTok{hbo\_credits }\OtherTok{\textless{}{-}} \FunctionTok{readRDS}\NormalTok{(}\StringTok
\end{Highlighting}
```

Consequently, data is being checked for NAs and duplicates.

The baby names dataset comprises US baby names from 1910 to 2014, with state-level observations differentiated by gender. In the next step, a graph is being created, showing the total number of babys in the data set over time, to get a feeling for the data. As shown in Figure @ref(fig:name-count), the total number of recorded names demonstrates significant variation over this 104-year period.

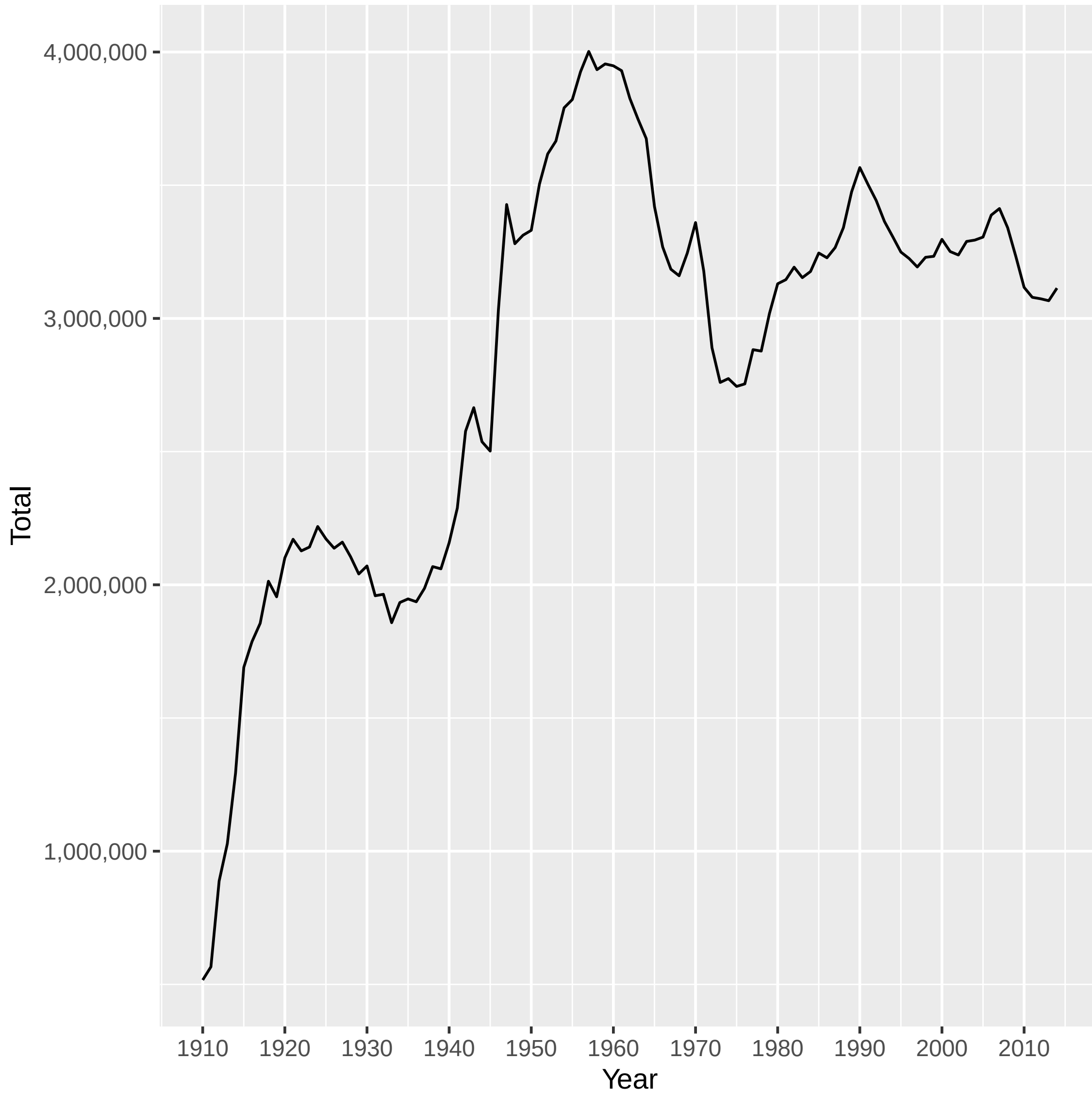

Figure 2.1: Annual Count of Unique Baby Names in the U.S., 1910-2014

2.2. Analysis Steps

The following chapter summarizes the results of the data analyses.

2.3. Mainstream Analysis

Now I carry out a "mainstream analysis" meaning I check how many children had the most popular name of the year, differentiated between male and female names.

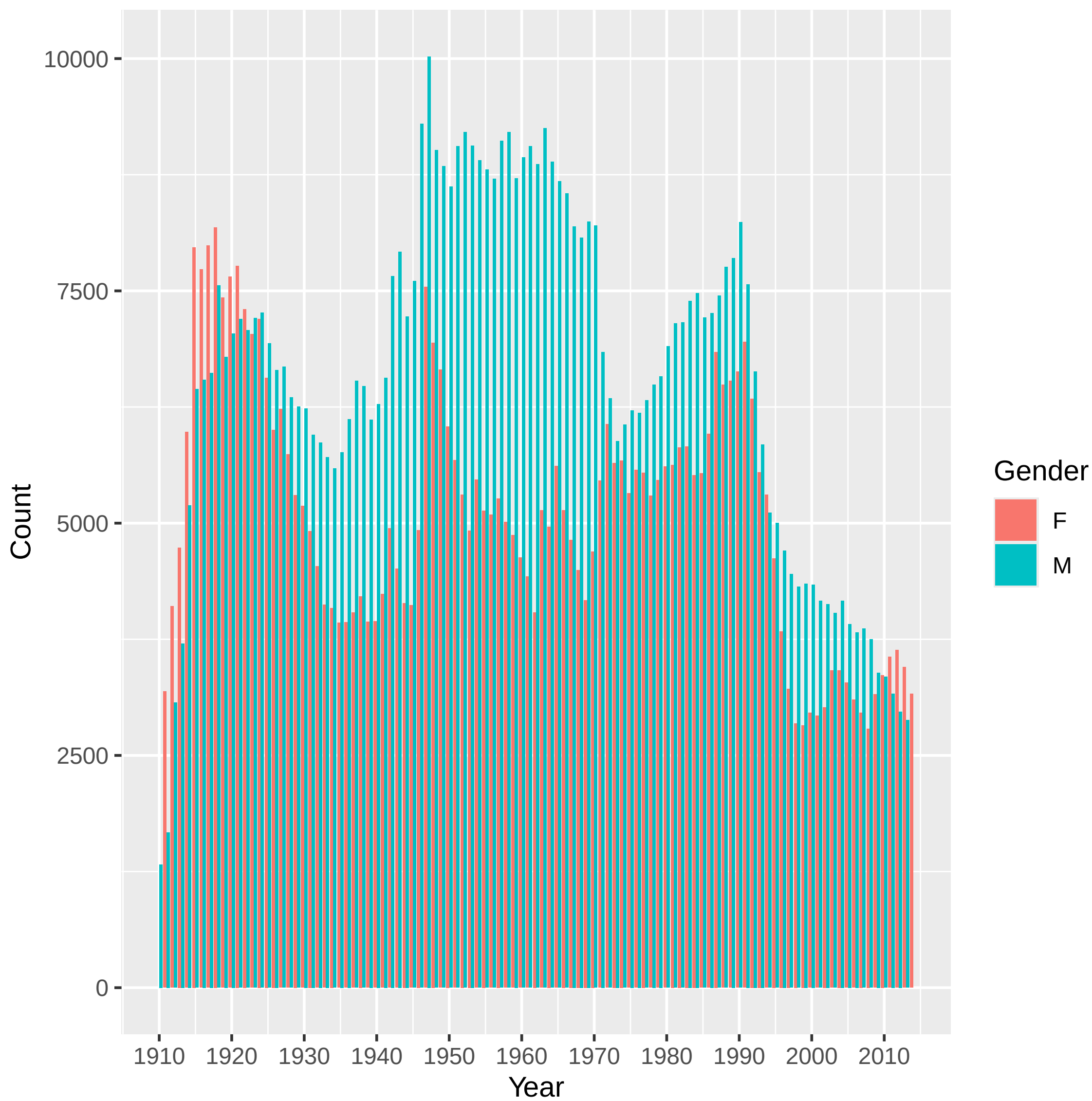

Figure 2.2: Mainstream Analysis

2.4. Persistence Analysis

To analyse whether the persistency of names has changed over the time horizon 1910 until 2011, a Spearmen Rank Correlation is being calculated between top 25 names on the country-level in year t0 and t3. The period had to be limited to 2011 as this is the last year for which the +3 years time lag can be calculated. The results of the persistency analysis (Figure @ref(fig:correlation)) suggests that persistence of names follows a decreasing trend overall, indicating that name preferences change faster today than they did in the mid 19th century. Comparing the means of Spearmen rank correlation coefficients before and after 1990 has confirms the hypothesis that popular names persisted slower after

1990. The mean of correlation coefficients before 1990 amounts to 0.857, while the mean thereafter reduces to 0.766, indicating faster-paced name trends.

Figure 2.3: Correlation Trend Over Time (1910–2011)

However, after correlation reduced considerably between the 1990s and early years of the present century, this trend has reversed after 2005. In the year 2011, Spearmen correlation is only slightly lower that in the beginning of the time series in 1910, indicating a recent shift towards more higher persistence of names.

2.5. Spike Analysis

The spike analysis was carried out to test which names have experienced the largest hypes in the period under analysis. @ref(tab:top-spikes) displays the 15 mostly hyped names, as indicated by their percentage increase in usage between two years. Interestingly, 12 out of the 15 names are female names, indicating names for baby girls are experiencing more extreme hypes. It has also been analysed, whether these names appear in movie or music productions. Surprisingly, out of these top 15 most spiked names, only the name Aja shows a relation to the film industry. The next subchapter analyses the impact of film and music publications on names overall.

Table 2.1: Top 15 Baby Name Spikes by Percentage Increase

Year	Name	Gender	% Increase
1964	Deneen	F	31340.0
1994	Aaliyah	F	28140.0
1983	Mallory	F	13060.0
2010	Tenley	F	12900.0
1957	Tammie	F	11580.0
1992	Jalen	M	9616.7
1978	Aja	F	8620.0
2012	Cataleya	F	7600.0
1928	Jeannine	F	7440.0
1981	Krystle	F	6816.7
1968	Dustin	M	6550.0
1995	Lyric	F	6320.0
1966	Tabitha	F	5960.0
2000	Elian	M	5922.2
2005	Mikalah	F	5680.0

2.6. Pop-Culture Analysis

Thereafter, billboard data is being used to check whether famous baby names relate to music pop-culture. The music_check function organizes the billboard data by top 20 songs per year and checks whether artist or song names relate to baby names. It then assigns two dummy columns to the spikes data frame, telling us whether names appear in top 20 singer or song names within the time frame 1910 until 2014.

A similar analysis is being done for the movie data stored in hbo_titles and hbo_credits. In a first step, both data frames are being merged based on the id column. Therafter, unnecessary columns are deselected and only films with rating higher than 8.5 and published before 2014 selected. Then, str_detect() is being used to check whether names occur in movie titles, character or actor names, or the movie description. In a last step, a dummy column is being added to provide information if the name is related to a popular movie.

In a last step, it is being analysed, how many names per gender are related to famous singers, songs or movies. This analysis is insightful to provide recommendations for the toy company, which elements of pop culture to get toy name inspiration from.

Table 2.2: Name Representation in Pop Culture

Gender	pct_artist	pct_song	pct_film
F	8.4%	2.6%	16.1%
M	15.1%	3.5%	30.5%

3. Recommendations

In a last step, names from actors or actresses of recent (as of 2014) and highly rated movies (> viewer rating 8.5) are being isolated to serve as source of inspiration for new toy names. Movie-related names are being chosen as they were found to influence baby name choice more than singer or song names.

Table 3.1: Top Name Predictions

Names
Andy
Asami
Cailee
Chukwudi
Guy
Juan
Junya
Kabir
Konomi
Lochlyn
Lydia
Manaka
Mitsuki
Olly

Peter
Ramon
Reiji
Rikako
Robert
Yuko

4. Conclusion

In conclusion, the analysis has provided evidence for the fact that baby name trends have become more fast-paced over the course of the recent decades. This is supported by the finding that mean correlation coefficients before 1990 are higher than those after 1990. However, data visualization in @ref(fig:correlation)) suggests a recent increase in correlation coefficients, indicating a slow-down of the previous fast-paced name trend development.

To analyse the sources of spikes in names over the years, billboard data on famous singers and songs, as well as Hbo data on movies, movie characters and actors has been analysed for their effect on name popularity. The findings of this analysis suggest that film-related names have had the largest effect on both male (30%) and female names (16%), followed by artist names (5.8% on female and 10.5% on male names). The lowest effect was found by song names, only contributing to 2.5% of male and 1.7% of female names.

These findings suggest that toy names should be inspired by movie characters or actresses, as these are the major sources of inspiration for baby names. Looking at recent film trends, possible candidates could for instance be Mahershala, Ivana or Yuki.