

Analysing the Persistence and Patterns of Baby Naming Conventions

Annika Pretorius^a

^a*Stellenbosch University*

Abstract

In this study, we analyze historical baby naming trends in the US from 1910 to 2014, focusing on the influence of cultural factors such as the media and celebrities. Our findings provide insights into the persistence of popular names and the impact of societal trends on naming conventions, aiding a toy design agency in selecting resonant character names for their products.

1. Analysing the Persistence and Patterns of Baby Naming Conventions

The naming trends of children have long been a subject of fascination and study. Recently, a New York-based kids' toy design agency has expressed interest in leveraging data analytics to understand these trends better. The primary goal is to identify the factors that influence baby names, such as popular movie characters, US presidential candidates, celebrities, and top-ranking songs or artists. By comprehending these patterns, the agency aims to predict which character names might resonate best with children, thereby enhancing the appeal of their toys.

To facilitate this analysis, we have been provided with an extensive dataset of baby names in the US, spanning from 1910 to 2014. This dataset includes annual data for all US states, offering a comprehensive view of naming patterns over more than a century. Additionally, we have access to population data by state and city, which will help us generate proportional insights at the state level. To supplement our primary dataset, we have also compiled data on music, movies, and TV series, including the Top 100 Billboard songs since 1958 and HBO movie and series titles with their audience popularity scores.

One of the key aspects of our analysis involves examining the longevity and persistence of baby names. Specifically, we aim to determine whether popular names tend to persist over time or if they fade away as fleeting trends. A useful method for this investigation is the Spearman rank correlation, which will help us assess the rank similarity of the top 25 most popular boys' and girls' names from one year to the next three years. This will provide insights into whether the trends of popular names have become more or less persistent over the decades, particularly since the 1990s.

Figure [1.1](#) below shows the total count of baby names per year by gender. The blue line represents male names, while the red line represents female names. This visualization highlights the fluctuations in the popularity of baby names over time, offering a foundation for our deeper analysis of naming trends.

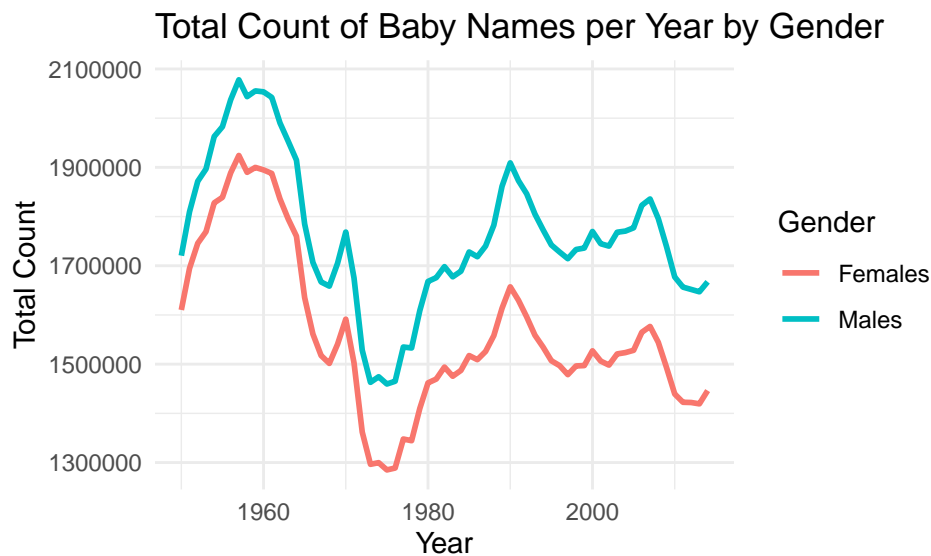


Figure 1.1: Total Count of Baby Names by Gender.

Figure 1.2 below illustrates the proportion of the top 25 baby names over time for both female and male names. The horizontal axis represents the years from 1950 to 2014. The vertical axis represents the proportion of babies given a specific name each year. This is calculated as the count of babies with that name divided by the total count of babies for that year. Each line represents a specific name that was in the top 25 for that year. The color of each line corresponds to a different name. The plot is divided into two facets: one for female names and one for male names. This separation helps to compare the trends for female and male names independently.

- **Female Names:** The left facet shows the proportion trends for the top 25 female names over time. The plot indicates that the popularity of the top names was relatively higher in the past (1950s to 1980s) and has become more evenly distributed in recent years.
- **Male Names:** The right facet shows the proportion trends for the top 25 male names over time. Similar to female names, the plot indicates that the popularity of the top names was higher in the past and has become more evenly distributed in recent years.

The peaks in the plot indicate years when specific names were particularly popular. The valleys indicate years when the popularity of those names decreased. **Declining Trends:** For both female and male names, the plot shows a general declining trend in the proportion values, indicating that the popularity of top names has become less concentrated over time. This suggests a diversification in naming trends, with a wider variety of names being used in recent years.

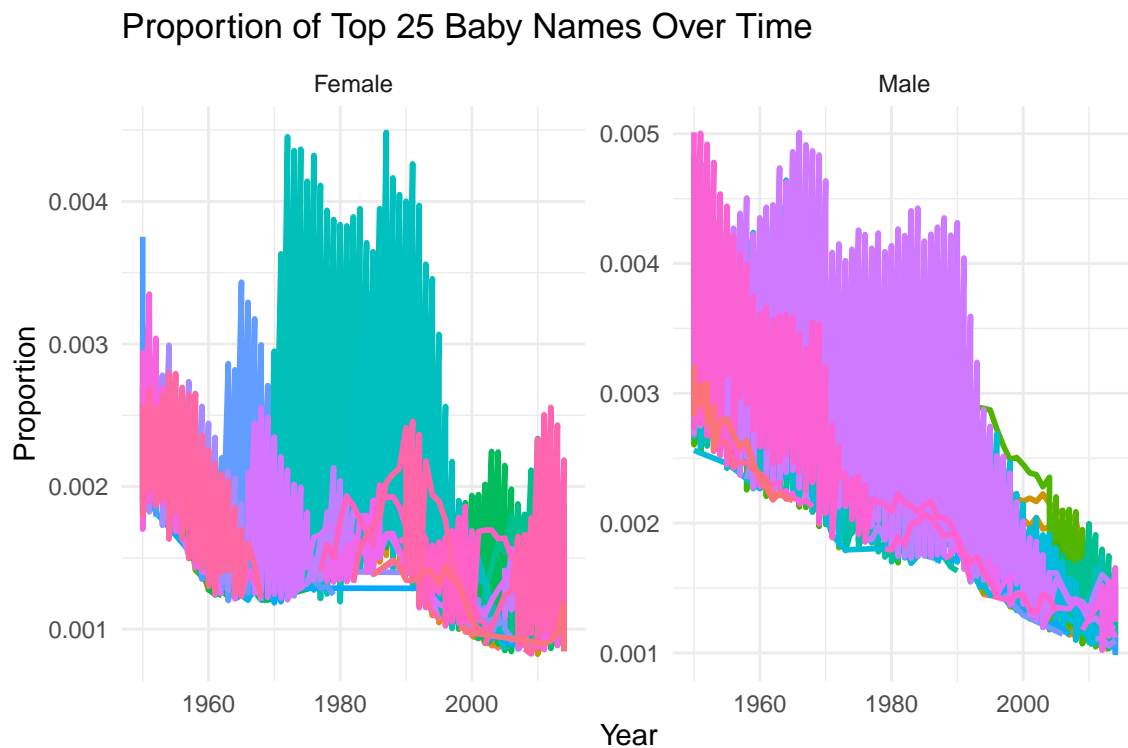


Figure 1.2: The Proportion of the Top 25 Baby Names and their Persistence Over Time

2. Analysing the Persistence of Names over Time: A Time Series Analysis

Figure 2.1 below showcases the persistence of the most popular baby names from a given dataset. The color gradient from white to blue represents the presence of a name in the top 25 baby names for each year, with darker shades of blue indicating higher persistence. Notable names such as “Jessica,” “Jennifer,” and “Emily” show strong persistence over the years, particularly in the 1980s and 1990s. Names like “Susan” and “Mary” were more popular in earlier decades, while “Ashley” and “Amanda” gained popularity later on. Names like “Robert,” “Michael,” and “John” demonstrate strong persistence, especially from the 1960s to the 2000s. Other names such as “Matthew,” “Joshua,” and “Joseph” show high presence in the top 25 names in the 1980s and 1990s. Male names show more consistent presence over a longer period compared to female names, which exhibit more variation and changes over the years.

```
## [1] "Data before pivoting:"
## # A tibble: 6 x 5
##   Year Name      Proportion Rank Gender
##   <dbl> <chr>          <dbl> <int> <chr>
```

##	1	1991	Jessica	0.00426	1	F
##	2	1987	Jessica	0.00448	1	F
##	3	1990	Jessica	0.00400	1	F
##	4	1989	Jessica	0.00405	1	F
##	5	1988	Jessica	0.00417	1	F
##	6	1992	Jessica	0.00397	1	F

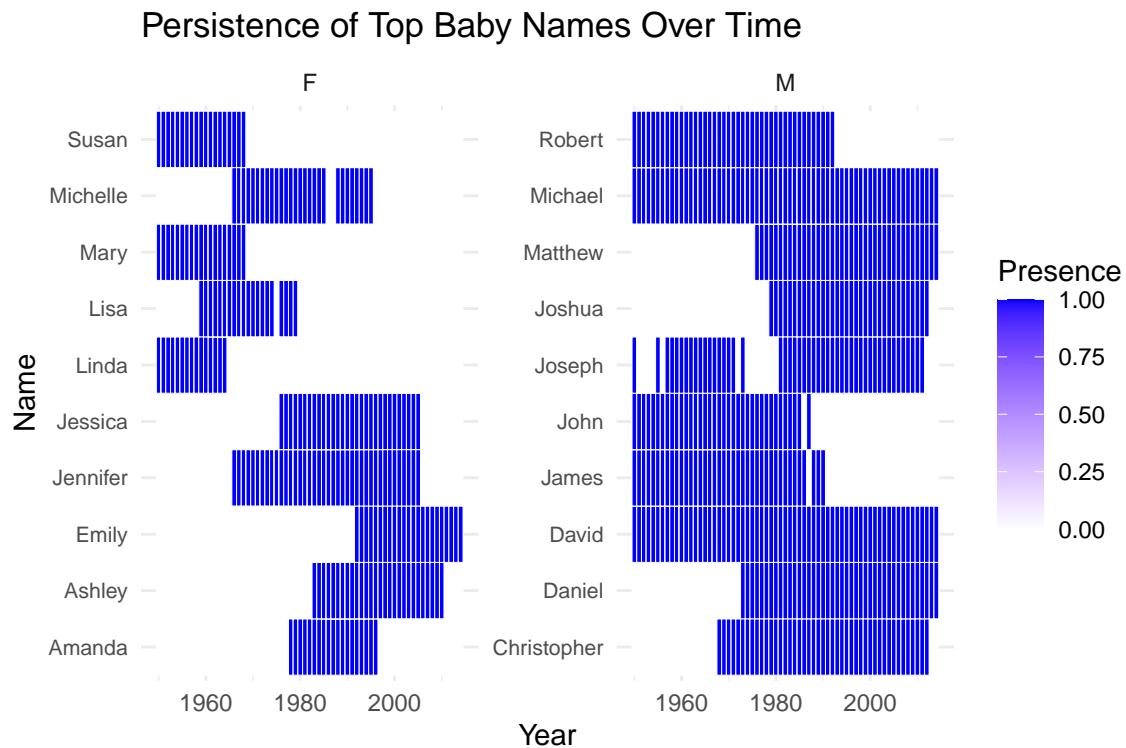


Figure 2.1: The Proportion of the Top 25 Baby Names and their Persistence Over Time for both Genders

3. Analyzing how long each name stayed in top 25

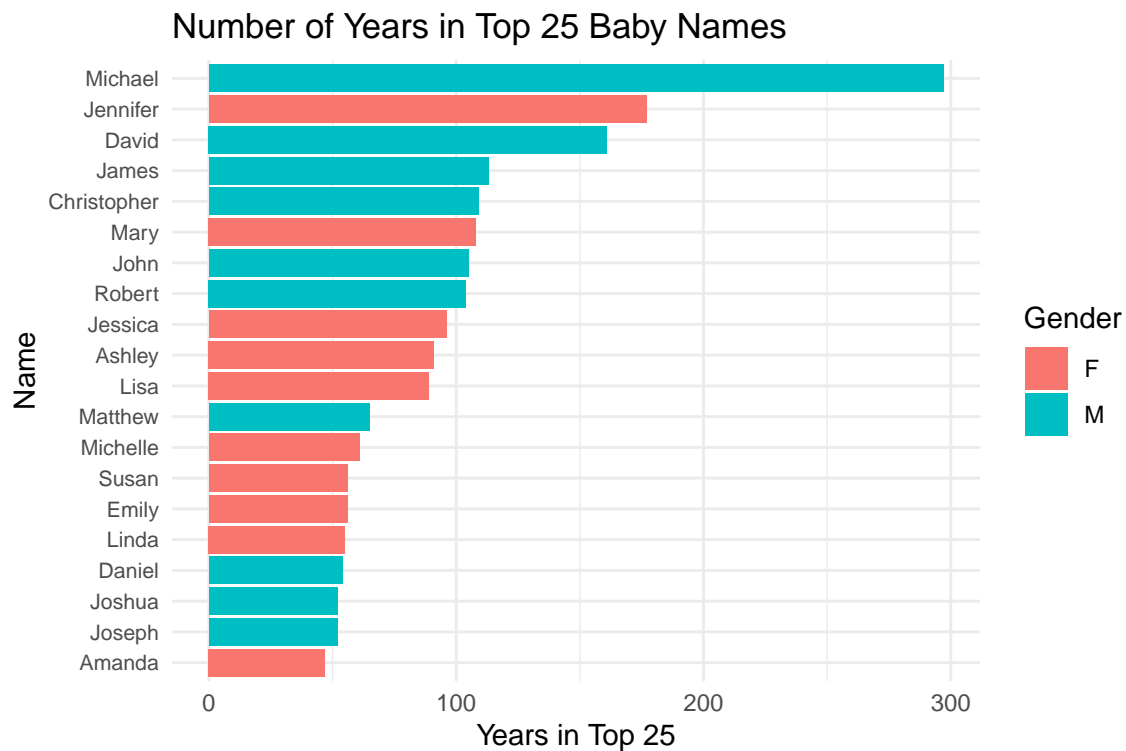


Figure 3.1: The Number of Years the Top 25 Baby Names Persisted Over Time for both Genders

We can therefore see that the top boy names over the last few decades illustrate significant persistence in comparison to that of girl names, therefore, we will further analyse factors influencing girl names as they display significantly higher variability which could be linked to alternate factors.

4. Analysing the HBO Dataset

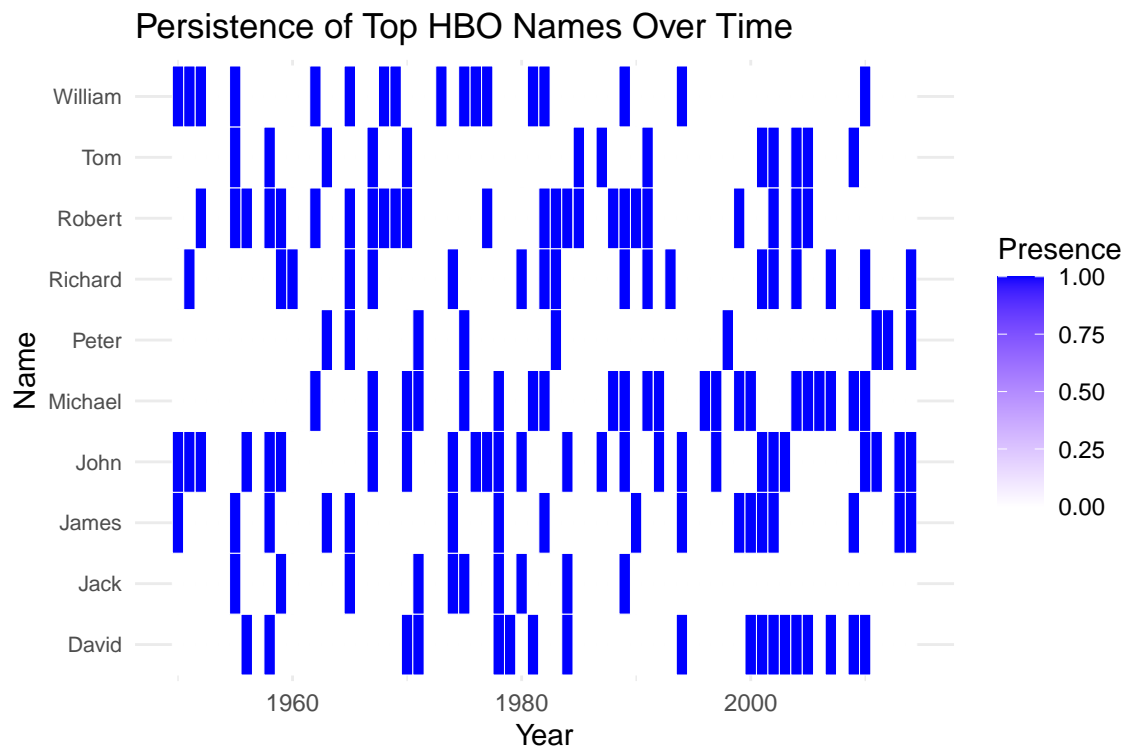


Figure 4.1: The Persistence of HBO names in the top rankings over time

We can see the above persistence is all male names within data regarding movies on HBO over the last few decades.

5. The Persistence of HBO names in the Top Rankings over Time.

In the Figure 5.1 below, the name “John” stands out as the most persistent name, having stayed in the top rankings for 40 years. This indicates that “John” was a popular name in HBO shows over a prolonged period. Names like “Robert,” “Michael,” “David,” and “Richard” also exhibit high persistence, each maintaining their presence in the top rankings for over 20 years. The presence of these names over many years suggests a trend in character naming conventions in HBO shows. The recurrence of these names might reflect cultural preferences or character archetypes frequently used in HBO’s storytelling.

The distribution of the years in top rankings shows that while some names had a very long presence, there is a gradual decline as we move down the list. Names like “Tom,” “Peter,” and “Jack,” though

still significant, have a lesser persistence compared to the top names.

The bar chart effectively highlights the relative persistence of each name, with the length of the bars corresponding to the number of years each name stayed in the top rankings. The use of horizontal bars makes it easy to compare the persistence across different names. This analysis helps in understanding the longevity and popularity of certain names within HBO shows, providing insights into naming trends and cultural influences over the years.

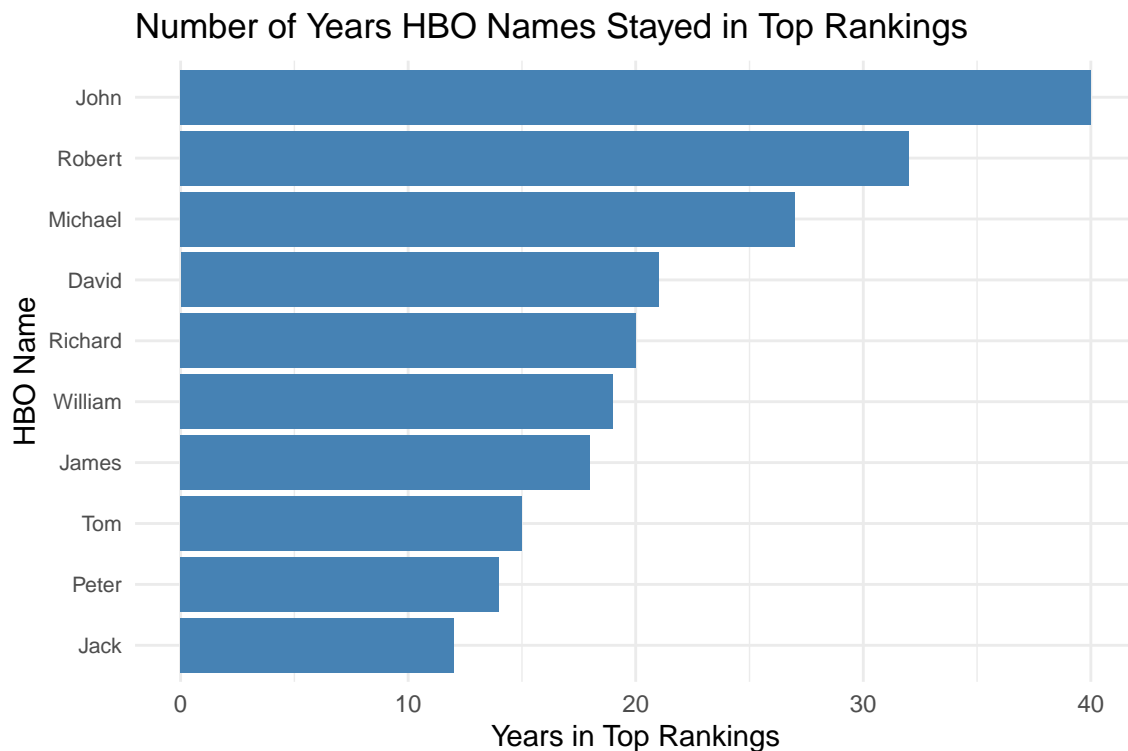


Figure 5.1: The Persistence of HBO names in the top rankings over time.

Now i will analyse the correlation between the persistent boy names within the Baby names dataframe in comparison to that of the most persistent names within the HBO baby frame, which all happen to be male.

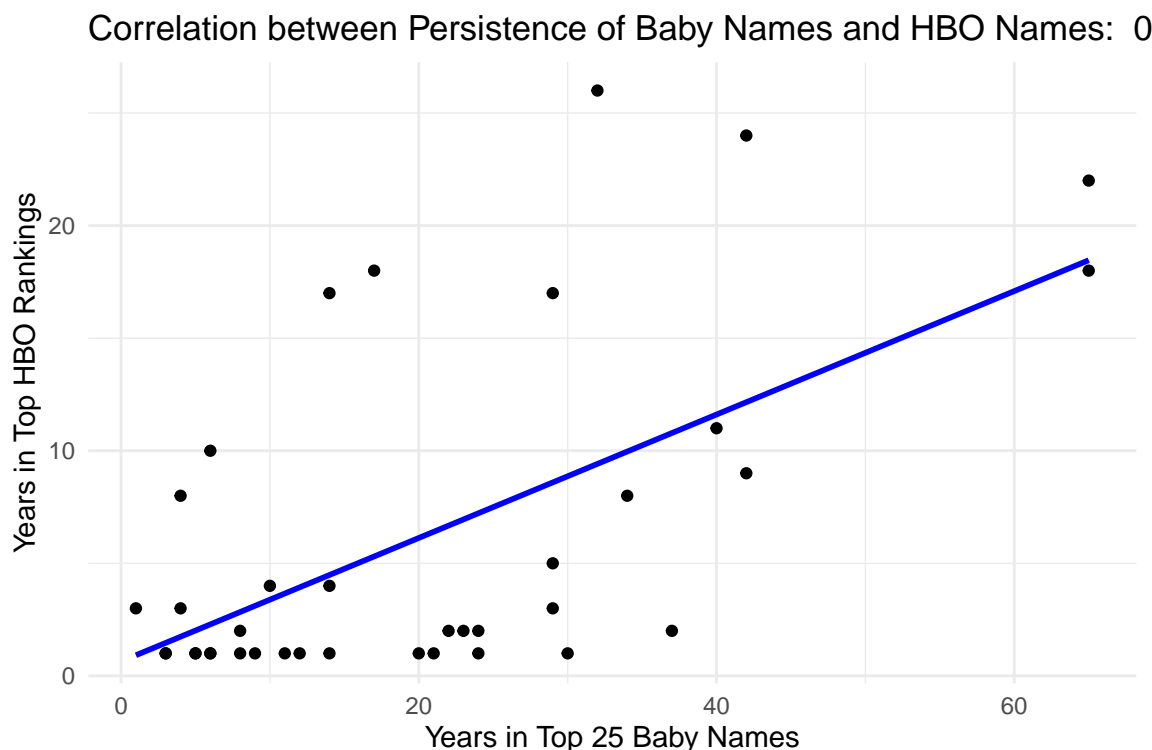


Figure 5.2: The Correlation Between the Persistence of HBO names and Baby Names.

```
## [1] "Spearman correlation value: 0.51"
```

In Figure 5.2 above the correlation between the persistence of baby names in the top 25 rankings and HBO names in the top 25 rankings over time are analysed. Each point on the scatter plot represents a name that appears in each of the datasets. The x-axis shows the number of years a name remained in the top 25 baby names, while the y-axis indicates the number of years the same name stayed in the top 25 HBO rankings.

A positive Spearman correlation value of 0.509. was calculated, which indicates a moderate positive correlation between the persistence of baby names and HBO names. This suggests that names that tend to be popular for longer periods as baby names are also likely to remain popular in HBO rankings for a significant duration. The blue line represents a linear fit, illustrating the trend of this correlation. While there is some spread in the data, with a number of outliers, the general trend suggests that there is a relationship between the duration of name popularity in both contexts.