

Question 1

Vincent Reinshagen^a

^a*Stellenbosch University, Stellenbosch, South Africa*

Abstract

I will analyze baby naming trends from 1910 to 2014, exploring influences like popular culture and societal trends, assessing trend longevity, and using Spearman rank correlation to examine if popular name persistence has slowed since the 1990s, leveraging US population data for additional insights.

1. Introduction

A toy design agency has approached me to analyze baby naming trends in the US spanning from 1910 to 2014. They are particularly interested in understanding the influences behind naming trends, such as popular culture phenomena like movie characters, celebrities, and political figures. Additionally, they want insights into the longevity of these trends, distinguishing between enduring names and passing fads. Using Spearman rank correlation, I will assess whether popular names from the 1990s onwards have exhibited slower persistence compared to earlier decades. The analysis will be supplemented with US population data to provide contextual insights at the state level.

To find out if the characteristics of the top names given to new borns is the same for Males and Females, I conducted a regression which shows the correlation between the counts of the top names for each year between Male and Female. This will be done for the 25 most popular names for Males and Females in each year.

```
##
##
## |      | Correlation.Coefficient..rho. | p.value|
## |:---|-----:|-----:|
## |rho |                0.9472603|          0|
```

Corresponding author: Vincent Reinshagen
Email address: vreinshagen@outlook.de (Vincent Reinshagen)

Contributions:
The authors would like to thank no institution for money donated to this project. Thank you sincerely.

```
##
## Spearman's rank correlation rho
##
## data: splitted_gender$M$Count and splitted_gender$F$Count
## S = 158991761, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9472603
```

To design a new toy is expensive. It is common to calculate with three years of amortization until the new design pays itself. We know, that in the 1920s, this was no issue and the time until new names lost their relevance was well above three years. The word got around that the top name list got way more dynamic since the 1990s. To test this, a spearman correlation test will be setup which investigates this and provides a time series plot.

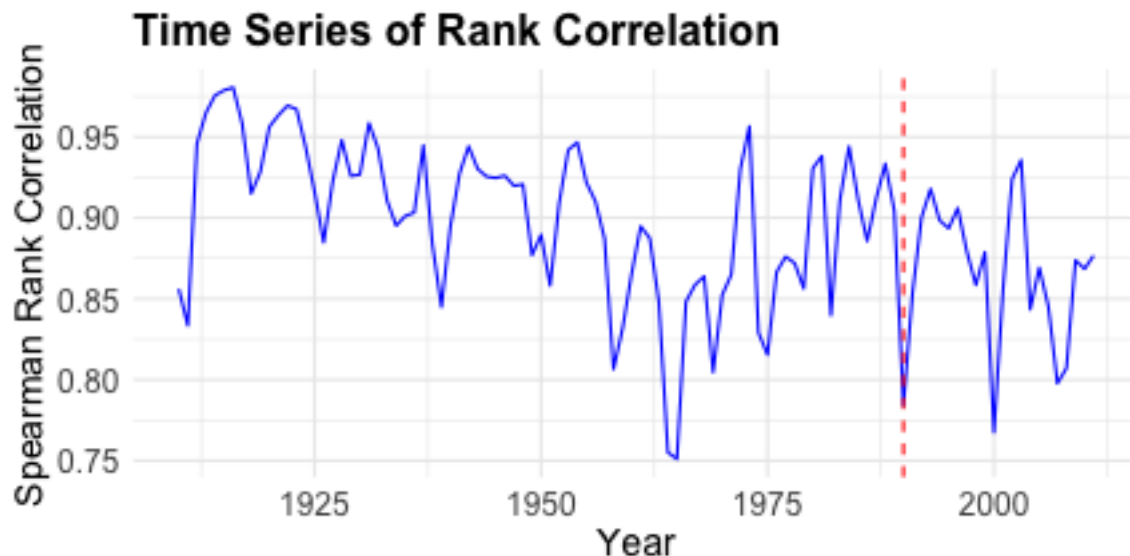


Figure 1.1: Caption Here

The plot shows that there is a declining trend of the persistence of popular names, which means that top names are usually shorter on the top names list. Nonetheless the persistence still appears high enough that we should not worry about tht for now.

2. Influence of other variables

We know that pop culture can have an influence on names for new borns. To be able to be the first to sell new nme personalized toys in the future we will investigate if the actors who will play in popular movies will have a strong influence on the names given to new borns.

We will start to invetigate the effect of movies on the names given to newborns

```
##
## Call:
## lm(formula = relative_change ~ in_movie, data = Popular_Names_with_following_year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.57  -5.50  -1.37   1.86  392.47
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3382     1.4831   0.228   0.820
## in_movie      1.0357     1.4921   0.694   0.488
##
## Residual standard error: 12.76 on 6195 degrees of freedom
## Multiple R-squared:  7.777e-05, Adjusted R-squared: -8.364e-05
## F-statistic: 0.4818 on 1 and 6195 DF, p-value: 0.4876
```

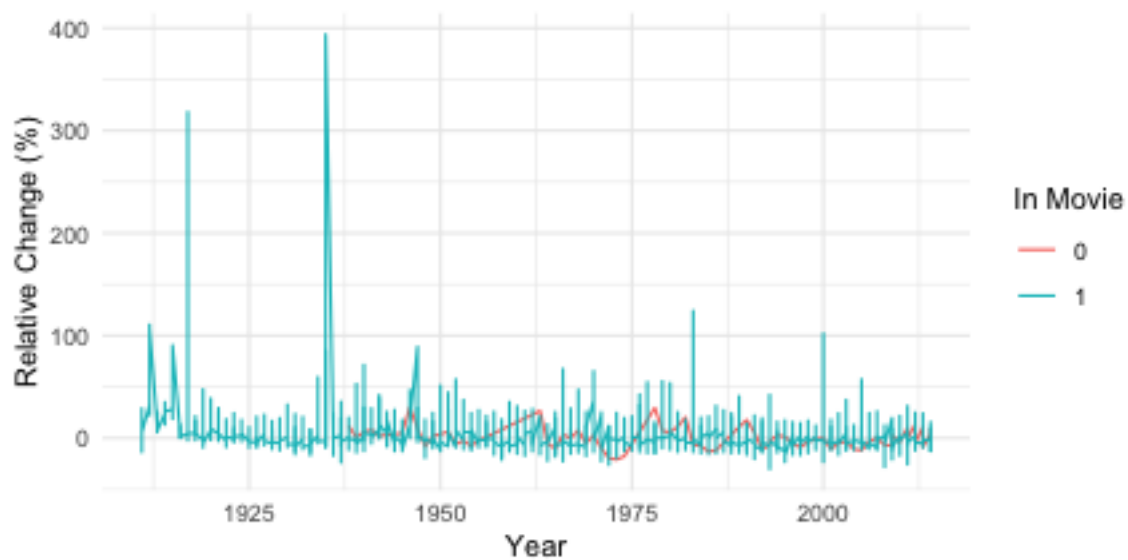


Figure 2.1: Caption Here