

QUESTION 1: Baby Names

Talyah Greyling^a

^a*Stellenbosch University, Cape Town, South Africa*

Abstract

This report is a concise analysis of baby naming trends in the US through the years utilising graphs, tables and bullet summaries to explain the findings to the client, a New York based kids' toy design agency.

1. Introduction

You are welcome to add some sort of an introduction here. Spend 5 minutes detailing the purpose here, e.g.

Look at the README.txt file and maybe give some background - especially also in understanding the definition of the variables.

Notice that for the below plots I created a parameters to play around with xaxis labels.

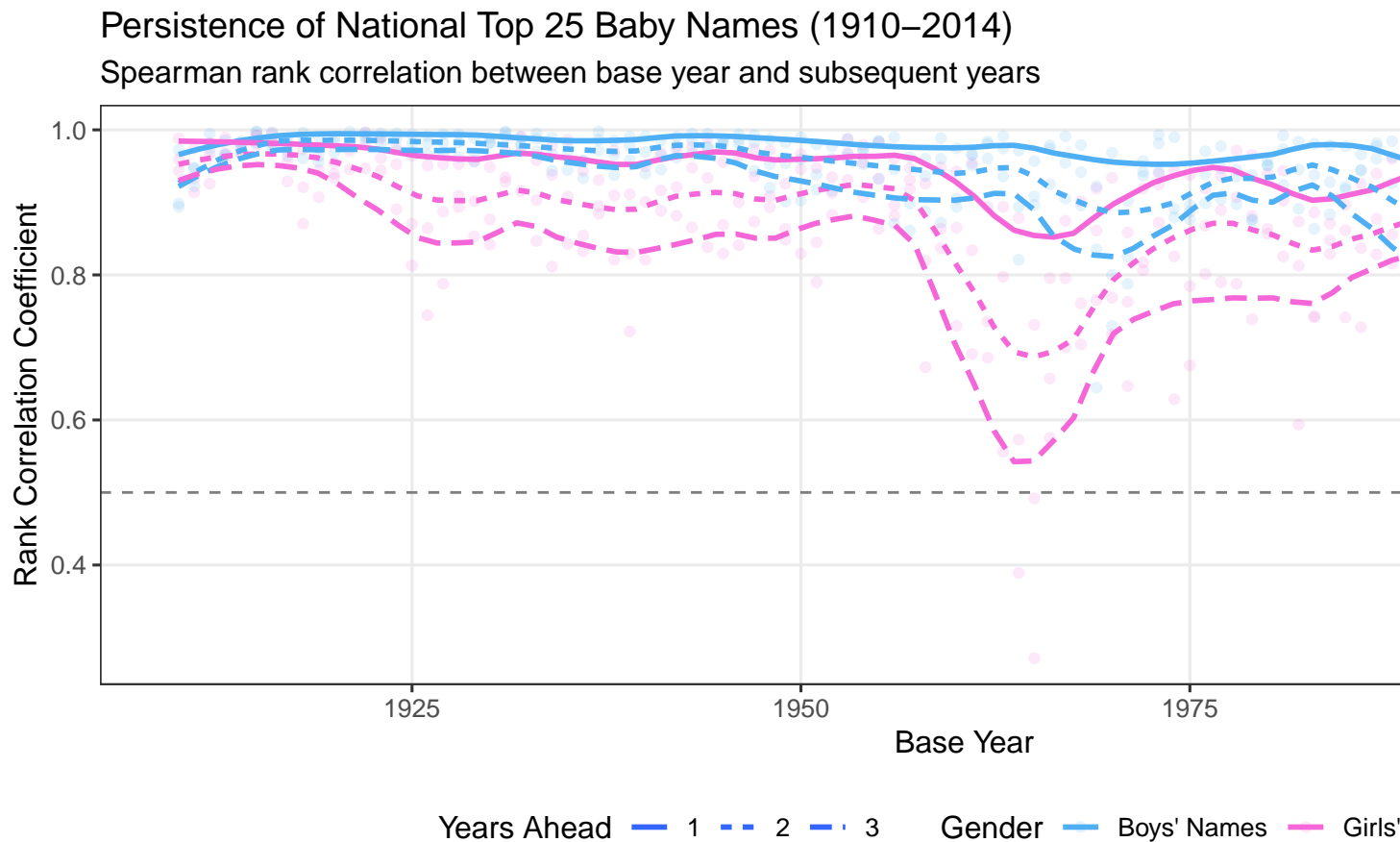
Corresponding author: Talyah Greyling

Email address: 23761067@sun.ac.za (Talyah Greyling)

Contributions:

The authors would like to thank the DNS server syncing for taking its time so that she could finish her morning cup of coffee before starting.

2. Time time-series representation of the rank-correlation calculations
3. Time time-series representation of the rank-correlation visualisation



Note: Correlations computed between each year's top 25 names and names in subsequent years. Only includes names that were in the top 25 in both the base year and the subsequent year.

Figure 3.1: Persistence of National Top 25 Baby Names (1910-2014).

As you can see from Figure 3.1, name persistence is much stronger among names for boys than names for girls. Both display trends that appear to be generally downward sloping over time. Persistence is also seen to decrease the further the future year is from the target year. Considering the client's suspicion that popular name trends have been slower to persist since the 1990s than in earlier decades their suspicion is correct for boys names, but appears to be the opposite for girls names which display with an upwards trend since the 1990s.

4. Popular names per decade calculations

5. Popular names per decade visualisation

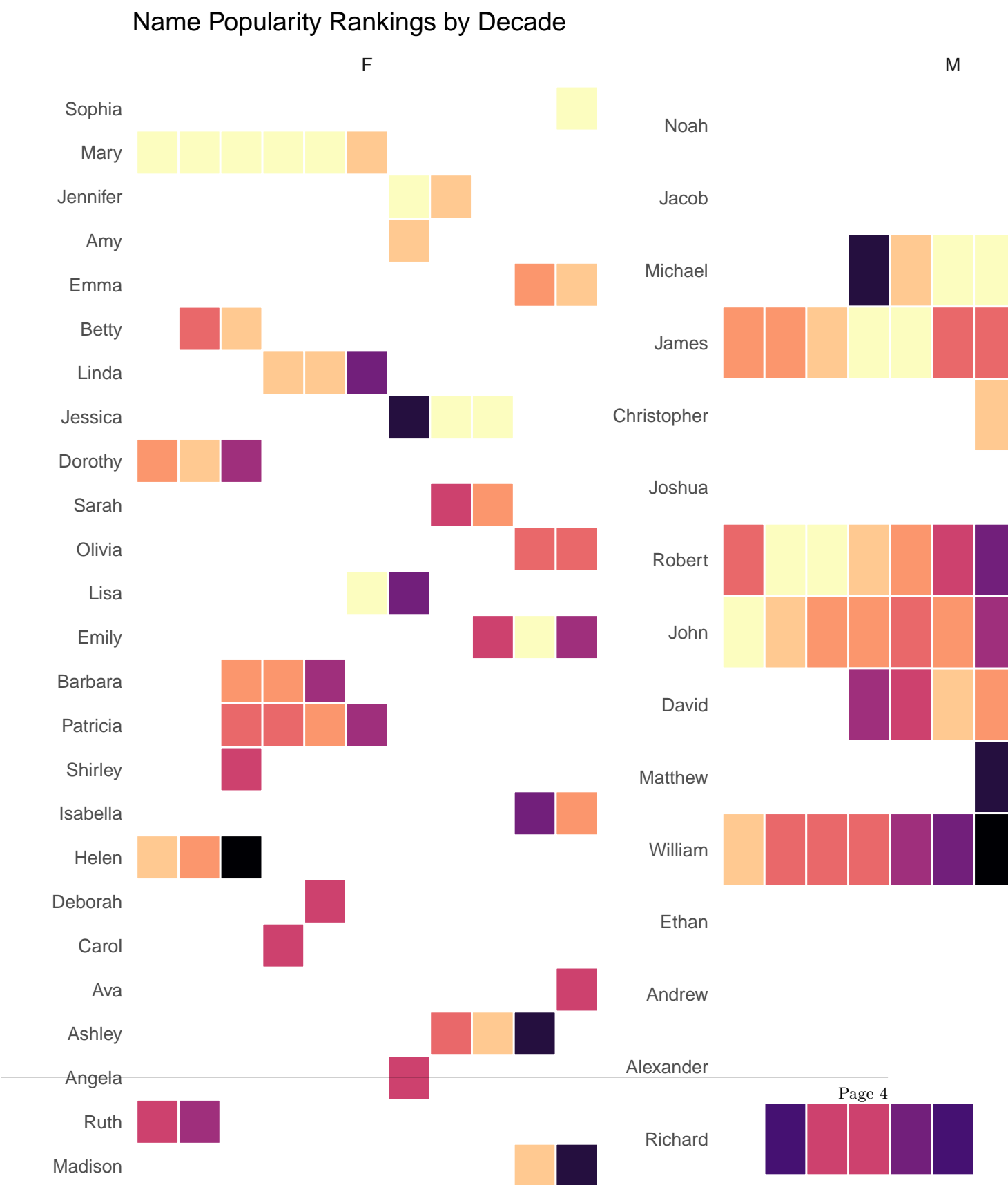


Figure 5.1 once again shows that name persistence is much stronger among males. By displaying the information in a heatmap we can see that popularity is concentrated among less names for males than for females, contributing to their higher persistence levels.

6. Movie star names calculations

```
## # A tibble: 29,118 x 7
```

```
##   name          Name    Surname character  title release_year tmdb_score
##   <chr>         <chr>    <chr>    <chr>      <chr>      <dbl>      <dbl>
## 1 Judy Garland   Judy    Garland Dorothy Ga~ The ~      1939       7.6
## 2 Ray Bolger     Ray     Bolger  Hunk / Sca~ The ~      1939       7.6
## 3 Jack Haley     Jack    Haley   Hickory / ~ The ~      1939       7.6
## 4 Bert Lahr      Bert    Lahr    Zeke / Cow~ The ~      1939       7.6
## 5 Margaret Hamilton Margaret Hamilton Elmira Gul~ The ~      1939       7.6
## 6 Frank Morgan   Frank   Morgan  Professor ~ The ~      1939       7.6
## 7 Clara Blandick Clara   Blandick Auntie Em  The ~      1939       7.6
## 8 Charley Grapewin Charley Grapewin Uncle Henry The ~      1939       7.6
## 9 Billie Burke   Billie  Burke   Glinda     The ~      1939       7.6
## 10 Pat Walshe     Pat     Walshe   Nikko      The ~      1939       7.6
## # i 29,108 more rows
```

```
## # A tibble: 10 x 11
```

```
##   name      Name  Surname character title release_year tmdb_score Decade Gender
##   <chr>    <chr> <chr>    <chr>    <chr>      <dbl>      <dbl> <dbl> <chr>
## 1 Mary Ast~ Mary  Astor   Brigid O~ The ~      1941       7.8  1940 F
## 2 Mary And~ Mary  Anders~ Maybelle~ Gone~      1939       8    1940 F
## 3 Mary Nash Mary  Nash    Margaret~ The ~      1940       7.7  1940 F
## 4 Mary Cla~ Mary  Clare   Baroness~ The ~      1938       7.5  1940 F
## 5 Mary Wic~ Mary  Wickes  Dora Pic~ Now,~      1942       7.4  1940 F
## 6 Mary Mor~ Mary  Morris  Halima    The ~      1940       7.2  1940 F
## 7 Mary Cla~ Mary  Clare   Mrs. Cor~ Oliv~      1948       7.5  1940 F
## 8 Mary Bol~ Mary  Boland  Mrs. Ben~ Prid~      1940       6.9  1940 F
## 9 Mary Jo ~ Mary  Jo Ell~ Clementi~ The ~      1949       7    1940 F
## 10 Mary Ast~ Mary  Astor   Brigid O~ The ~      1941       7.8  1950 F
## # i 2 more variables: Total_Dec_Count <dbl>, Rank <int>
```

```
## # A tibble: 10 x 11
```

```
##   name      Name  Surname character title release_year tmdb_score Decade Gender
##   <chr>    <chr> <chr>    <chr>    <chr>      <dbl>      <dbl> <dbl> <chr>
## 1 Mary Ast~ Mary  Astor   Brigid O~ The ~      1941       7.8  1940 F
## 2 Mary And~ Mary  Anders~ Maybelle~ Gone~      1939       8    1940 F
## 3 Mary Nash Mary  Nash    Margaret~ The ~      1940       7.7  1940 F
```

```
## 4 Mary Cla~ Mary Clare Baroness~ The ~ 1938 7.5 1940 F
## 5 Mary Wic~ Mary Wickes Dora Pic~ Now,~ 1942 7.4 1940 F
## 6 Mary Mor~ Mary Morris Halima The ~ 1940 7.2 1940 F
## 7 Mary Cla~ Mary Clare Mrs. Cor~ Oliv~ 1948 7.5 1940 F
## 8 Mary Bol~ Mary Boland Mrs. Ben~ Prid~ 1940 6.9 1940 F
## 9 Mary Jo ~ Mary Jo Ell~ Clementi~ The ~ 1949 7 1940 F
## 10 Mary Ast~ Mary Astor Brigid O~ The ~ 1941 7.8 1950 F
## # i 2 more variables: Total_Dec_Count <dbl>, Rank <int>
```