Exploring Trends in Baby Names Across Decades in the United States

Group 20: Tathagata Banerjee, Ketan Saini, Sujash Krishna Basak & Anjali Gond

16 November 2023

1. Introduction

Our project delves into the captivating world of baby names in the United States spanning from 1920 to 2010. Envision it as a journey through the evolving panorama of names over the years.

Motivation

Names transcend mere labels; they function as time capsules, offering insights into the societal trends of various decades. Our objective is to delve into these capsules, unraveling the stories they hold and deciphering how people's preferences for baby names have transformed. At the core of our exploration is the mission to unveil trends in US baby names across decades, distinguishing between timeless classics and passing fancies.

2. Data Set

2.1 Rank Table

We have gathered the top 100 names for each decade, culminating in a comprehensive dataset of 1,000 observations, encompassing both male and female babies.

Show [5 ∨] entries   Search: [   ]

	Decade	Decadewise.Rank	Male.Name	No..of.births.M.	Female.Name	No..of.births.F.
1	1920	1	Robert	576373	Mary	701755
2	1920	2	John	564054	Dorothy	368874
3	1920	3	James	515301	Helen	290403
4	1920	4	William	512403	Betty	283095
5	1920	5	Charles	298031	Margaret	245010

Showing 1 to 5 of 1,000 entries   Previous [1] 2   3   4   5   …   200   Next

In the provided dataset, the variables are outlined as follows:

Variable	Description
Decade	Represents the decade under consideration, taking values from 1920 to 2010.
Decadewise Rank	Denotes the rank of names within the respective decade, ranging from 1 to 100.
Male Name	Signifies the male name corresponding to the given rank in that decade.
Number of Babies (Male)	Specifies the number of babies with the corresponding male name.

Variable	Description
Female Name	Signifies the female name corresponding to the given rank in that decade.
Number of Babies (Female)	Specifies the number of babies with the corresponding female name.

2.2 History and Origin

In addition to charting the trajectory of baby names in the United States from 1920 to 2010, our project extends its reach to compile a rich tapestry of information. We meticulously gather data on the history and origin of every name that has graced the top 100 rank list throughout these decades. This multifaceted approach allows us not only to track the ebb and flow of naming preferences but also to unravel the fascinating stories behind each name.

Instead of static databases, we opt for real-time data scraping directly from the source. To bring this dynamic data to life, we showcase it seamlessly on our custom-designed RShiny app, specifically crafted for detailed Name-specific Analysis. The app serves as a vibrant portal, allowing users to explore the intricacies of baby names in the United States.

The coming sections our report will give more details on why this data was not stored statically and used in visualisation

3. Obtaining the data

Our dataset, on Rank Table, has been acquired through web scraping from the official Social Security website of the United States, https://www.ssa.gov (https://www.ssa.gov).
The supplementary layer has been acquired through diligent web scraping from https://www.behindthename.com/ (https://www.behindthename.com/), providing a comprehensive perspective on the historical roots and origin stories associated with each name.

3.1 Source and Reliability

The Social Security website serves as a trusted repository of demographic data, including comprehensive information on baby names in the United States. By obtaining our dataset directly from this official source, we enhance the reliability of our findings, instilling confidence in the accuracy and authenticity of the information presented. To enhance the trustworthiness of the historical and origin data obtained from https://www.behindthename.com (https://www.behindthename.com), we conducted a thorough validation process. Importantly, Behindthename.com provided additional links to support their information, reinforcing the validity of the presented details.

3.2 Data Scrapping and Processing

The process of data scraping from the ssa.gov website proved to be manageable. Acquiring information on the top-ranked names for each decade involved extracting data from various webpages. To streamline this, we created a deticated function. The baby counts contained commas as separators, which were removed for convenience during data processing.

Acquiring data from behindthename.com posed specific challenges due to the variable number of fields for various pieces of information. Due to this variability, scraping proved challenging, and as a result, certain data were left unscraped. The complexity further increased as each name necessitated visiting a new webpage. However, we successfully collected data from each webpage, and subsequently, the relevant information was meticulously cleaned

3.3 Potential Sources of Bias or Misinformation:

- Early Decades' Data Collection Inefficiencies: In the early decades, the data collection system may not have been as efficient, potentially leading to incomplete registrations of names.Changes in policies, administrative practices, or

data collection methods can affect the consistency and accuracy of the data.

- The data available on ssa.gov is derived exclusively from individuals who registered for a Social Security Number (SSN). Consequently, individuals who did not register for an SSN are not represented in this dataset.
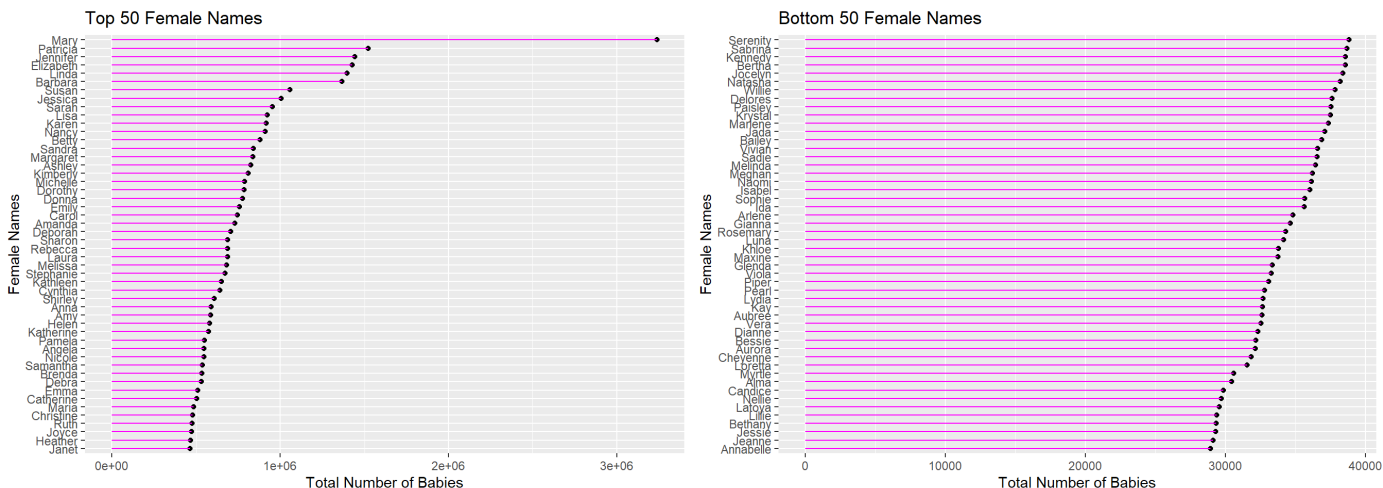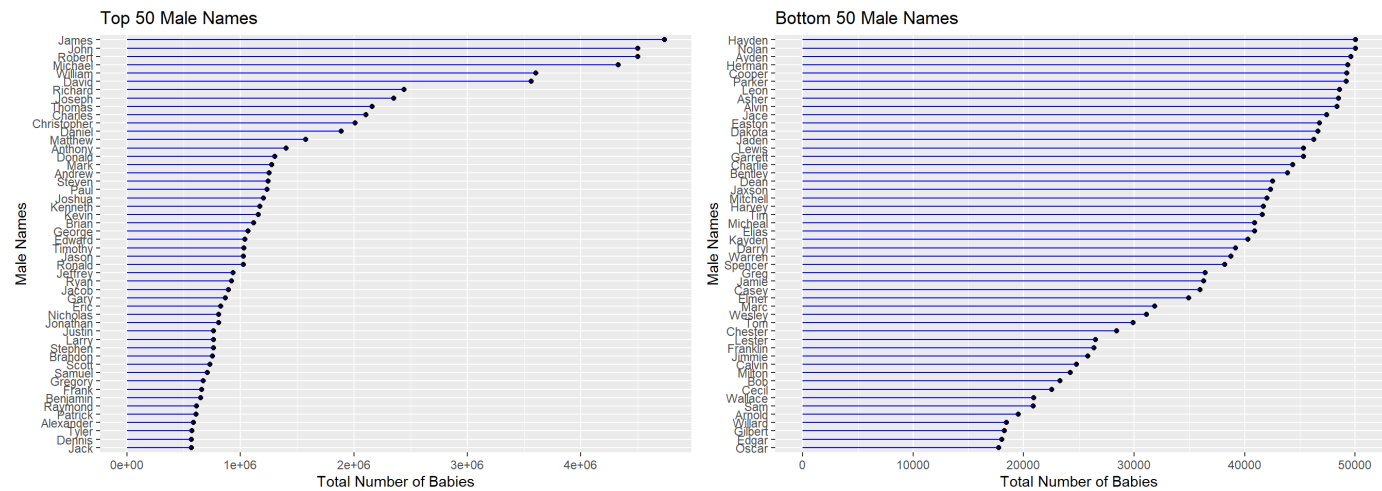
4. Interesting questions to ask from the data:

Our analysis will address several intriguing questions based on the dataset:

1. Evolution of Popularity: How has the popularity—measured by rank and the number of babies with the top-ranked names—changed over the decades?

2. Temporal Popularity Patterns: Which names have experienced fleeting popularity over just one or two decades, and which names have consistently maintained their popularity?

3. Gender-Based Ending Letters: How does the presence or absence of a vowel sound in the ending letter vary between male and female names?

4. Temporal Trends in Initial Letters & Last Letter: How much the popularity of a name depended on the initial and last letter of the name?

5. Name Length and Popularity: How does the popularity of names differ based on the length of the name, considering both male and female names?

6. Cross-Gender Popularity: Are there any names that have achieved top rankings in both male and female categories?

5. Normal Insights Through Data Analysis and Visualization

5.1 Top Name vs Total Number of babies

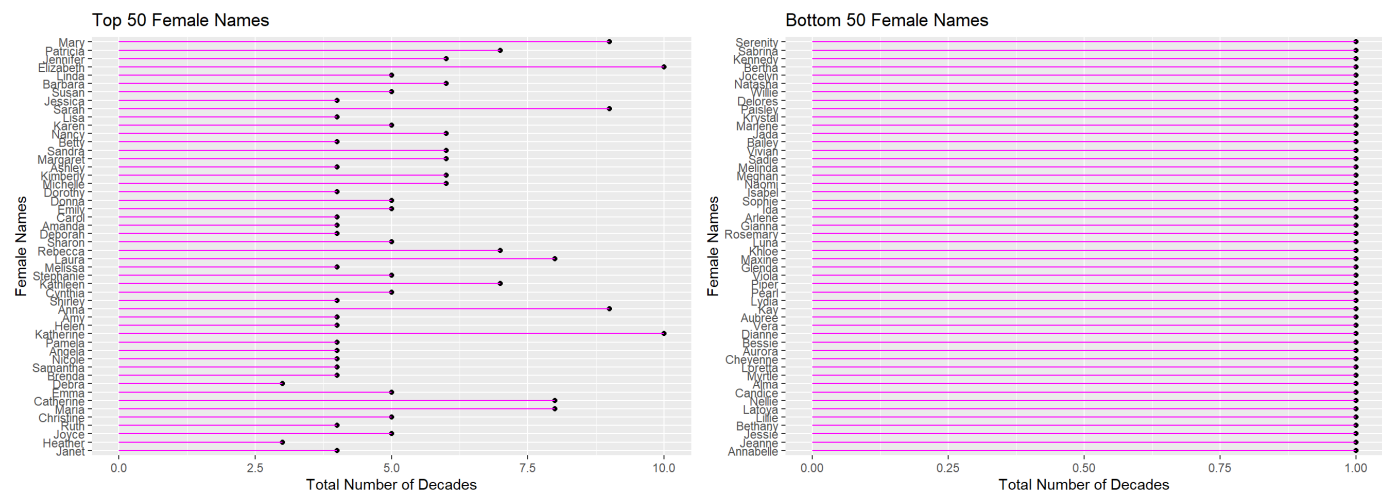Top 50 Male Names   Bottom 50 Male Names

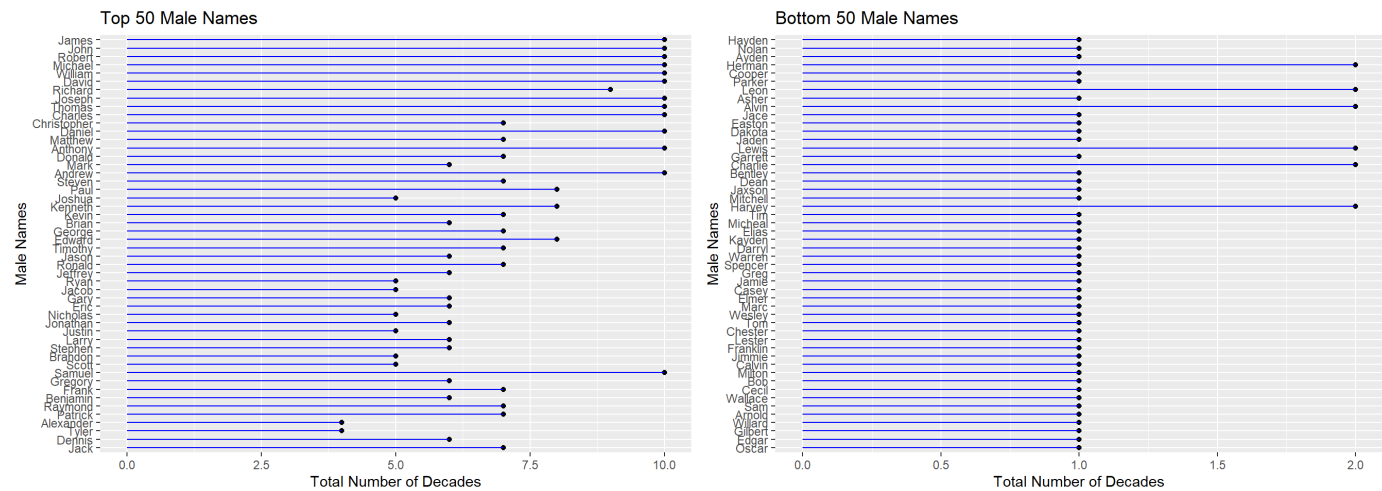### 5.1.a Top Female Name vs Total Number of babies:-

- Mary holds the top spot as the most popular female name, with Patricia and Jennifer following closely. Notably, there is a significant decline in popularity from Mary to Patricia.
- Annabelle emerges as the least favored female name on the list, with a total count of fewer than 30,000 babies.
- The names at the lower end of the popularity scale have markedly fewer instances compared to the top names. This discrepancy may be attributed to their sporadic appearance across decades. In instances where these names are absent in certain decades, they are recorded as having zero occurrences.

5.1.b Top Male Name vs Total Number of babies:-

- James claims the title of the most popular male name, closely trailed by John. Notably, both names share the same initial, 'J.'
- The overall count of names experiences intermittent drops, even within the top ranks.
- Names such as Oscar, Edgar, and Gilbert emerge as the least favored, each having fewer than 20,000 occurrences in the list.

5.2 Top Name vs Total Number of Decades

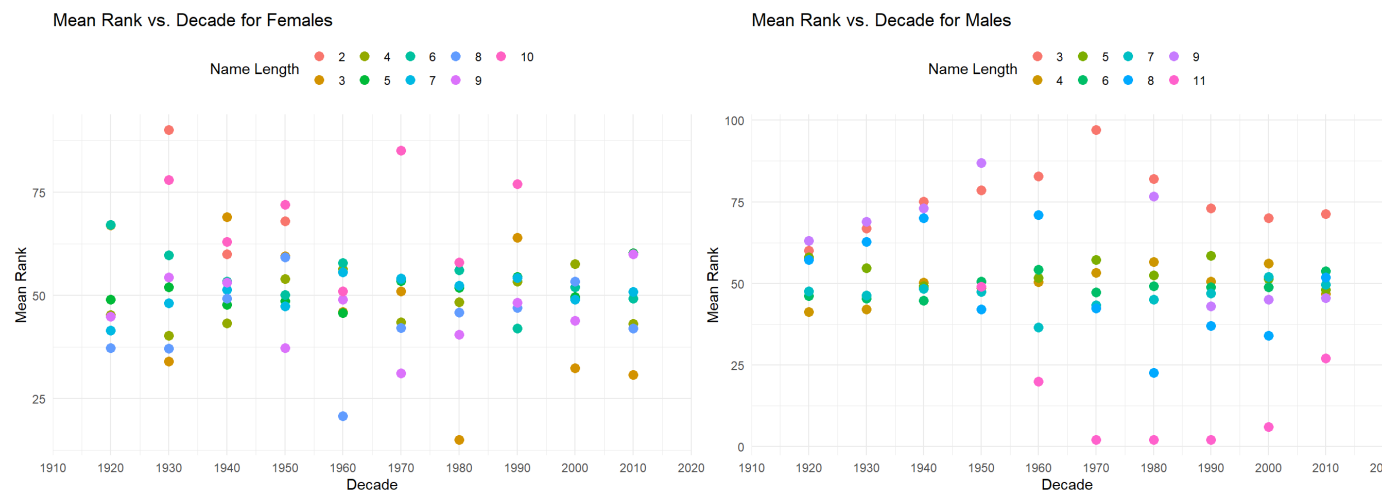Top 50 Female Names   Bottom 50 Female Names

5.2.a Top Female Name vs Total Number of Decades:-

- The top 50 female names show fluctuating patterns when we consider the number of decades they have been popular for. The top name,Mary,has been popular for only 8 decades. Elizabeth and Katherine(which is not in the top 25),are the only names in the top 50 that have appeared in all decades.
- All the bottom 50 names have appeared in only 1 decade,and hence the total number of babies for those names is very less compared to the top ones.

5.2.b Top Male Name vs Total Number of Decades:-

- It can be observed that the top 5 most popular names have all been popular throughout 10 decades, and most of the top names are so.
- "Christopher" is the topmost name having appeared in less than 8 decades,but it's recent popularity has pushed it up the ranks. Samuel has been a name consistenly moderately popular,ranking close to 50 even after being in top 100 in all decades.
- Similar to the female names,most of the bottom male names have been popular for just one or two decades.
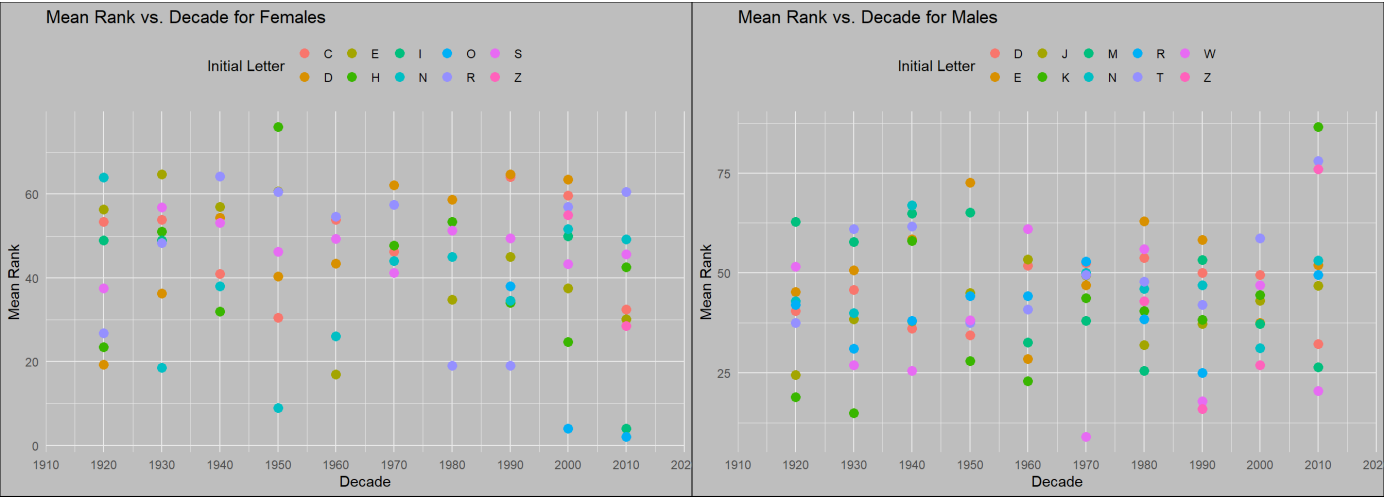
5.3 Mean Rank vs Decades (Name Length)

5.3.a Mean Rank vs Decade:-

- Among females, moderately long names with 5,6,7 and 9 letters have been consistently moderately popular.3 letter names look like they have been very popular recently , possibly due to a specific name("Ava"). 10 letter names have been consistently less popular,possibly because parents prefer shorter names for their daughters.
- Among males,moderately long names with 5,6 and 7 letters have been moderately popular.3 letter names saw a drop in popularity but regained it to an extent after 1970. 11 letter names look like they have been very popular recently, due to the name "Christopher" which has been the most popular name in 1970,1980 and 1990.

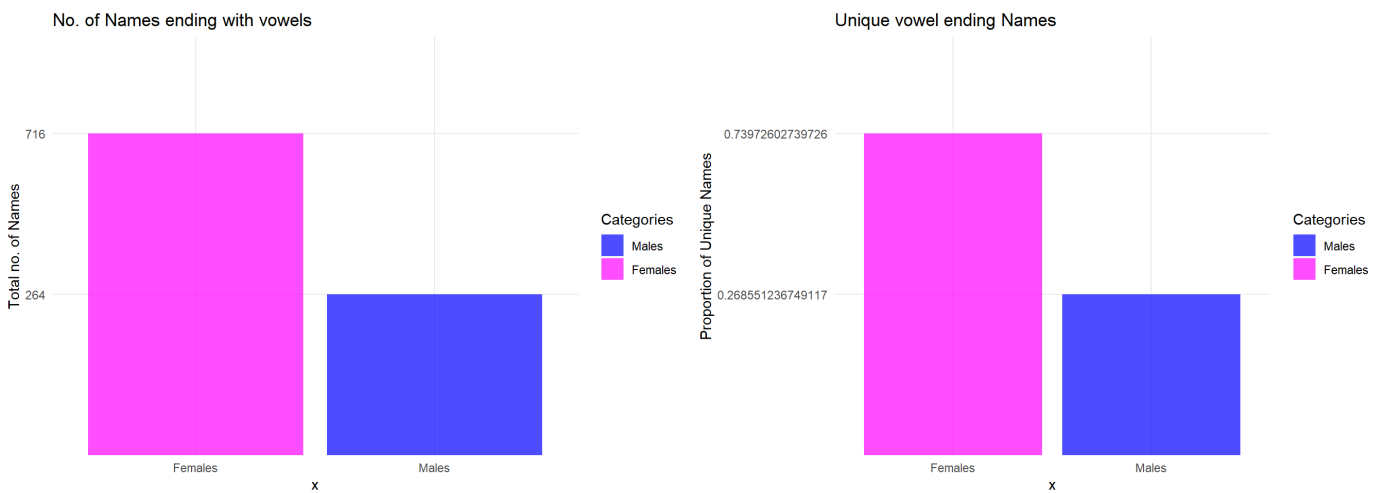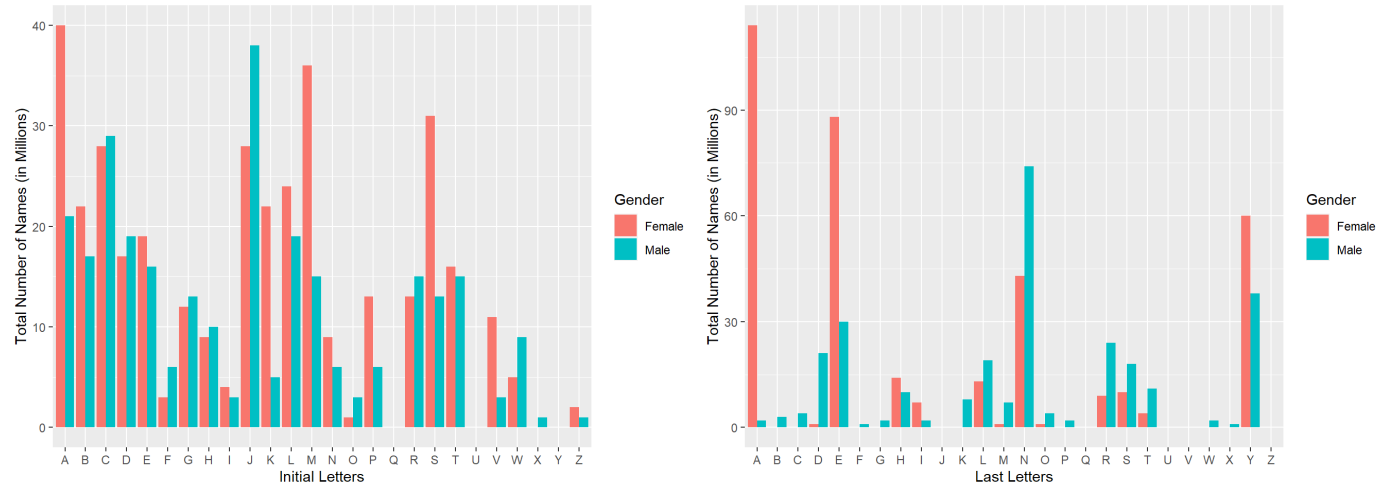## 5.4 Mean Rank vs Decades (Initial Letter)

5.4.a Mean Rank vs Decade:-

- Among females, N has been a popular initial recently. S,D and R have been consistently moderately popular. Z has been an uncommon initial,popular for only 1 decade.
- Among males, W has been very popular recently, probably due to "William". K was popular in the past,but has fallen in ranks now.J, D and M have been consistently moderately popular.

6. Intresting Insights Through Data Analysis and Visualization

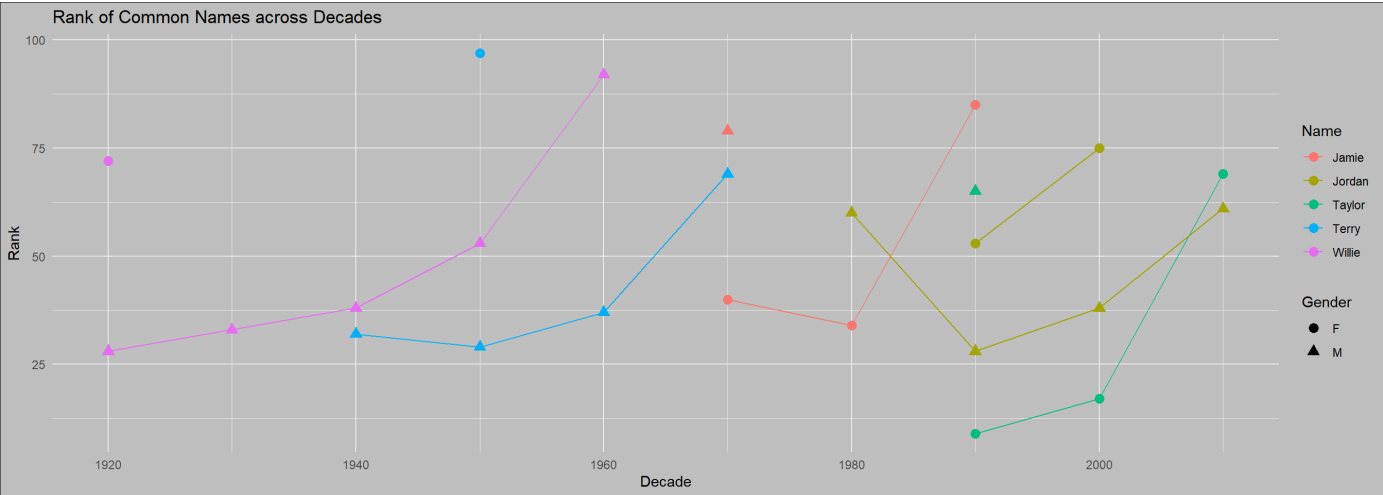## 6.1 Intial and Last Letter of Names

6.1.a Observations:-

- From the plots, it is evident that female names exhibit a preference for ending with vowel sounds. Across decades, the total number of female names within the top 100 ranks is more than twice that of male names.

- When considering the proportion of unique names, the disparity becomes more pronounced. Approximately 74% of female names conclude with vowel sounds, whereas only 27% of male names share this characteristic.

- Consequently, if a name concludes with a vowel sound, it is significantly more likely to be associated with a female name than a male name. Conversely, for names concluding with a consonant sound, the likelihood of it being a male name is higher.

- Even though girls constitute a smaller proportion of the overall population, the prevalence of names ending with the letter "A" among them is quite astonishing.

6.2 Common Names (Gender Neutral)

6.2.a Observations:-

- Jamie,Jordan,Terry,Taylor and Willie are the five names common between males and females.
- The common names have been popular at similar times for either gender.
- Among common names, only Taylor has reached the very top ranks among females.The others have been moderately or less popular among top ranks.

7. Limitation

They rise in popularity, enjoy a period of dominance, and then fall. "Emma" and "Liam" will be hot for a while, until suddenly … they're not. But why? What makes a name suddenly pop—and then die?

The driving force behind these fluctuations is often linked to pop culture. Parents draw inspiration from a myriad of sources, including their favorite celebrities and characters from best-selling books.

It's evident that the popularity of names is influenced by a multitude of factors. While analyzing the history and origin of every name is beyond our scope, we have conducted research on some prominent ones.

Anomalies :-

- The female name "Aria" has experienced a surge in popularity, likely influenced by the well-known character from the web series "Game of Thrones."

- When Kool and the Gang's song "Joanna" hit the Billboard Hot 100 List in 1984, the name Joanna shot up in popularity. The same thing happened to "Rosanna" after Toto's song of that name in 1982.

- Political events can also wield influence on naming trends. Following Ronald Reagan's presidency in the 80s, the name "Reagan"—once relatively uncommon—experienced a notable surge in popularity, eventually securing a place in the top 100 names by 2010.

8. Conclusion

In this project, we delved into the popularity trends of the most favored male and female baby names in the U.S. spanning the past ten decades. Our exploration encompassed factors such as diverse name choices, variations in initial and ending letters, and differences in name lengths.

Methodology:

For our analysis, we employed metrics such as the total number of babies and the mean rank corresponding to various factors, providing a comprehensive view of name popularity.

Key Insights:

- Biblical Origins Dominating: Mary emerges as the foremost female name, while James and John reign as the leading male names, all rooted in biblical origins.

- Length Matters: Names with moderate lengths (5, 6, 7 characters) consistently maintain moderate popularity. Distinct short (e.g., Ava) and long (e.g., Christopher) names exhibit heightened popularity among females and males, respectively.

- Initial Letter Influence: The initial "J" stands out as the most popular for males and exhibits considerable popularity among females as well.

- Variety in Female Initials: Females exhibit a broader spectrum of popular initials compared to males.

- Temporal Dynamics of Initial Choices: The preference for initial letters has significantly fluctuated over time for both genders, with only a select few maintaining consistent popularity.

- Gendered Ending Sounds: Notably, female names predominantly conclude with vowel sounds, while male names typically end with consonants.

- Gender-Neutral Names: While names like Jamie, Jordan, Taylor, Terry, and Willie are popular across genders, their moderate popularity within each gender suggests that parents may not commonly opt for gender-neutral names for their children.

- Increased Variety in Names: Back in 1900, for example, 91% of all children of any gender were given a name from the top 1,000 most popular names. But a century later in 2000, only 75% of girls were given a name from the top 1,000 most-popular girl names, and that percentage had dropped for boys too, to 86%. In other words, more kids were getting names that would have been considered unusual or new. (This data was also collected from ssa.gov)

9. Acknowledgement

We want to express our special thanks of gratitude to our professor Dr Dootika Vats, who gave us this fantastic opportunity to do this great project of Data Science on "Exploring Trends in Baby Names" and who also helped us out at various stages and eventually helped us in completing the project with our best efforts and in time. We learnt a lot of new things

throughout this project's journey, and it was an amazing experience for us as a team to understand and implement things collaboratively.

10. References:-

- https://www.ssa.gov/oact/babynames/decades/names2010s.html (https://www.ssa.gov/oact/babynames/decades/names2010s.html)

- https://www.behindthename.com/ (https://www.behindthename.com/)

```