

# A Data-Driven Exploration of Baby Naming Trends in the United States

Linda Dube<sup>a</sup>

*<sup>a</sup>Stellenbosch University, Western Cape*

---

## Abstract

This study employs Spearman rank correlation analysis to examine longitudinal trends in U.S. baby naming practices from 1910 to 2014. By comparing annual rankings of the top 20 male and female names across consecutive years, we assess the temporal persistence of naming conventions. The results indicate significant transformations in naming patterns during the 1990s, reflecting evolving cultural influences on parental decision-making. These findings contribute to our understanding of how sociocultural factors shape intergenerational naming traditions in American society.

---

## 1. Introduction

This project investigates the dynamic evolution of baby naming conventions in the United States from the early 20th century to present day, analyzing how cultural, musical, and media influences have shaped naming trends. Using longitudinal data from Social Security records, Billboard charts, and entertainment media, I employ quantitative methods to identify patterns and disruptions in naming persistence. The study particularly examines the growing volatility in name popularity since the 1990s, contrasting traditional naming practices with modern, media-influenced choices. Through this analysis, I aim to illuminate the complex interplay between cultural phenomena and parental naming decisions across generations.

## Data and method

This study leverages four comprehensive datasets to analyze cultural influences on baby naming trends in the United States. The primary dataset consists of official Social Security Administration records documenting the annual popularity of baby names through 2015, providing a foundation for tracking naming patterns over time. To examine potential pop culture influences, we incorporate Billboard Hot 100 chart data spanning 1958-2015, which tracks weekly song and artist rankings based on sales, radio play, and streaming metrics. Additionally, we utilize movie and television data from HBO/TMDB,

---

*Email address:* 23084103@sun.ac.za (Linda Dube)

including titles, release years, and popularity scores, to identify correlations between character names and baby name trend

Table 1.1: Data summary

Name	baby_names
Number of rows	5647426
Number of columns	6
Column type frequency:	
character	3
numeric	3
Group variables	None

#### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Name	0	1	2	15	0	30274	0
Gender	0	1	1	1	0	2	0
State	0	1	2	2	0	51	0

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Id	0	1	2823713.501630271.61	1	1411857	2823714	4235570	5647426		
Year	0	1	1972.39	29.57	1910	1949	1977	1999	2014	
Count	0	1	52.92	180.81	5	7	13	34	10023	

The dataset contains over 5.6 million baby name records from all U.S. states between 1910 and 2014, spread across six variables—three character and three numeric. Names are consistently recorded, with no missing values and over 30,000 unique entries. The distribution of years is broad and fairly even, while name counts per entry vary widely, suggesting a skew toward rare names. The dataset is clean, complete, and well-suited for time-series and trend analysis of naming patterns across gender, time, and state.

Findings # Spearman Rank Correlation of Baby Names

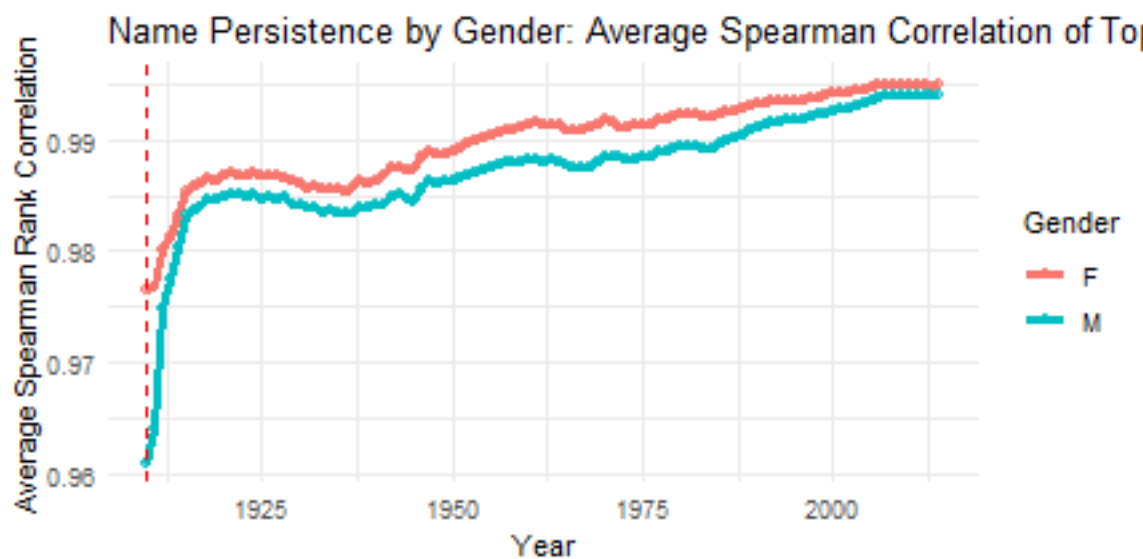


Figure 1.1: Caption Here

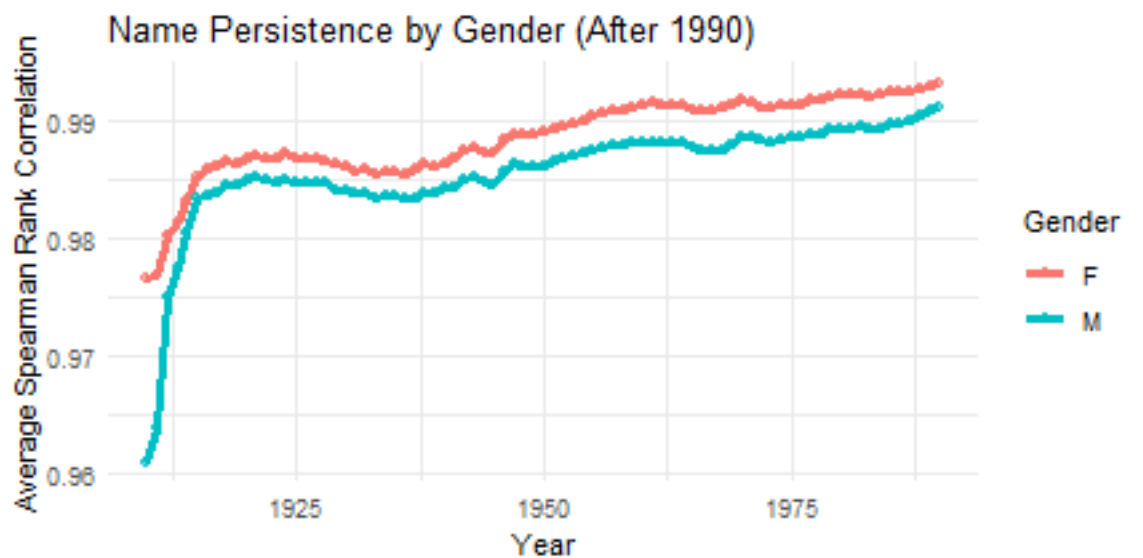


Figure 1.2: Caption Here

The Spearman rank correlation plot below illustrates the year-on-year persistence of baby name popularity before 1990, separated by gender. The correlation for female' names was consistently high, typically above 0.9, indicating that the most popular male names changed very little from one year to the next. In contrast, the correlation for boys' names was generally lower and more volatile, reflecting

greater fluctuations in naming trends. This suggests that female names were more susceptible to changing cultural influences, while male names remained relatively traditional and stable over time. Despite the fluctuations, both genders show a moderate-to-high level of persistence overall.

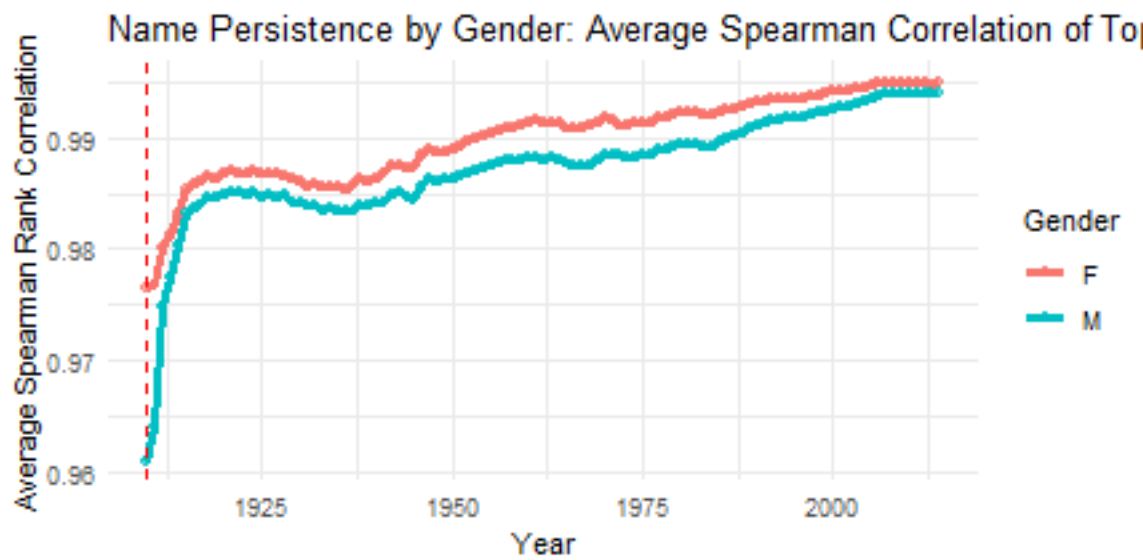


Figure 1.3: Caption Here

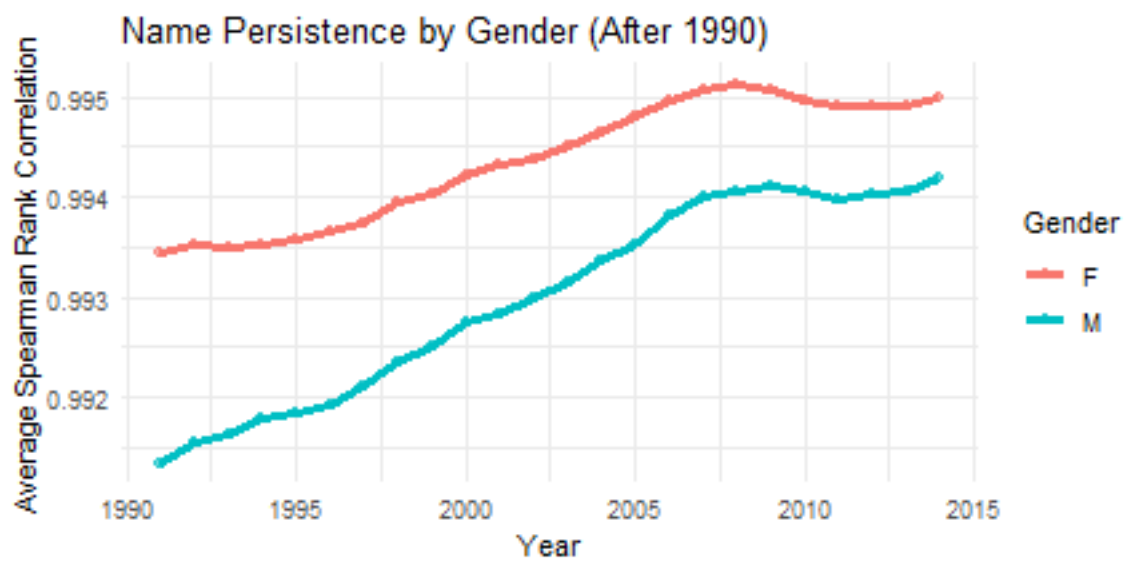


Figure 1.4: Caption Here

The Spearman rank correlation after 1990 shows slightly more fluctuation than in earlier decades, particularly for female names, which dip as low as 0.75. While male names remain relatively stable,

---

they also show occasional drops below 0.85. This suggests that name trends have become less persistent and more influenced by short-term cultural factors. Thus, I then confirm the suspicion that name popularity has become more volatile since the 1990s.

#### # Year-on-Year Surges in Baby Name Popularity

## # A tibble: 20 x 8

##		Id	Name	Year	Gender	State	Count	Count_lag	Pct_Change
##		<dbl>	<chr>	<dbl>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
##	1	4350321	John	1912	M	PA	3067	70	42.8
##	2	4243423	Mary	1912	F	PA	4106	131	30.3
##	3	4351681	John	1915	M	PA	6443	207	30.1
##	4	3750309	Joseph	1912	M	NY	2250	77	28.2
##	5	4245041	Mary	1915	F	PA	7970	281	27.4
##	6	3750310	William	1912	M	NY	1922	70	26.5
##	7	4351166	John	1914	M	PA	5192	191	26.2
##	8	4244447	Mary	1914	F	PA	5981	235	24.5
##	9	4350022	John	1911	M	PA	1672	66	24.3
##	10	4352252	John	1916	M	PA	6544	284	22.0
##	11	4243917	Mary	1913	F	PA	4738	207	21.9
##	12	4242992	Mary	1911	F	PA	3188	141	21.6
##	13	4246415	Mary	1917	F	PA	7987	364	20.9
##	14	4350722	John	1913	M	PA	3706	172	20.5
##	15	4354658	John	1920	M	PA	7041	327	20.5
##	16	4353426	John	1918	M	PA	7558	353	20.4
##	17	4352830	John	1917	M	PA	6618	313	20.1
##	18	4245735	Mary	1916	F	PA	7730	368	20.0
##	19	4247118	Mary	1918	F	PA	8184	408	19.1
##	20	3751268	Joseph	1914	M	NY	3327	173	18.2

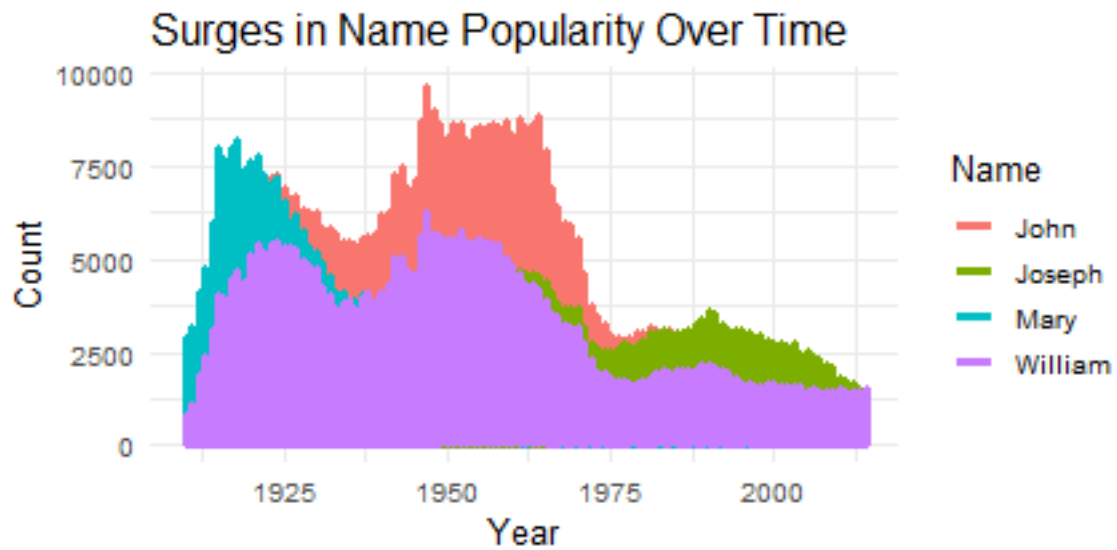


Figure 1.5: Caption Here

The graph highlights four names—John, Mary, William, and Joseph—that showed distinct surges in popularity over time. Mary and John peaked early, especially between the 1910s and 1950s, reflecting their strong biblical and traditional appeal. William remained relatively stable, while Joseph saw renewed interest around the 1990s–2000s, possibly tied to cultural or religious revivals. Their early “pop” suggests these names were historically dominant and shaped broader naming trends across decades.

2. Checking for baby name spikes after major events

Jeniro	Johni	Judas	Nami	Neini	Naria	Laury
Jewel	Johnnie	Jude	Kandi	Kelis	Kiki	Lainey
Jewell	Johnny	Judson	Kane	Kellee	Kiley	Lake
Jhene	Jon	Judy	Kansas	Kellie	Kim	Lalo
Jill	Jonas	Juelz	Kanye	Kelly	Kimberley	Lamo
Jim	Jonathan	Jules	Karen	Kelsea	Kimberly	Lana
Jimi	Jonell	Julia	Karl	Kem	King	Lando
Jimmie	Joni	Julian	Karla	Kendrick	Kip	Lani
Jimmy	Jordan	Juliana	Karmin	Kenny	Kirby	Lanier
Jinny	Jordin	Julianne	Karol	Kent	Kirk	Lara
Jo	Jordy	Julie	Karyn	Keri	Kitty	Lari
Joan	Jorgen	Juliet	Katalina	Kermit	Koryn	Larry
Joanie	Jose	Julio	Kate	Kerry	Kris	Laura

Figure 2.1: Caption Here

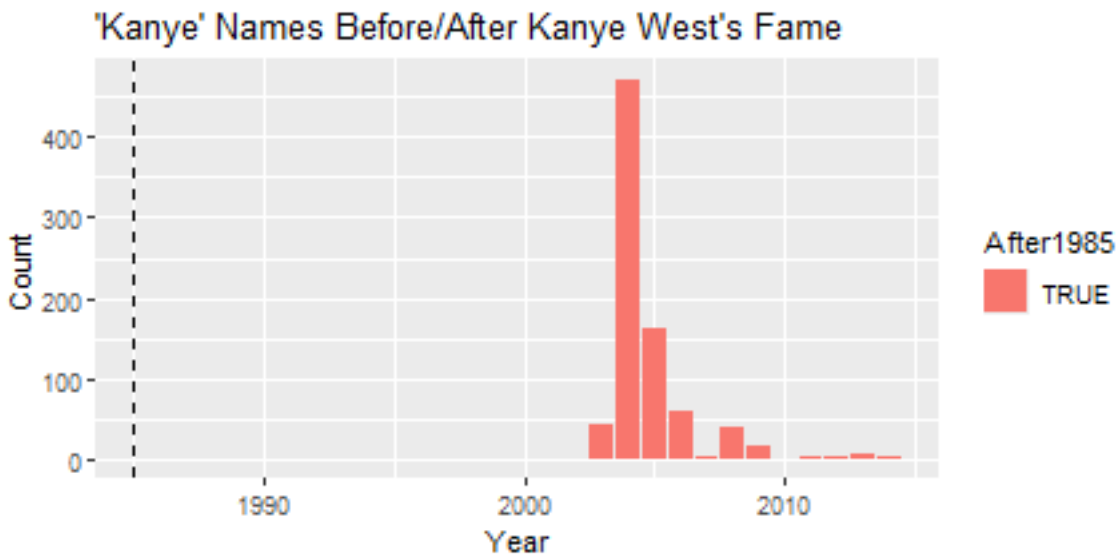


Figure 2.2: Caption Here

The graph shows a dramatic rise in babies named “Kanye” starting in 2001, coinciding exactly with Kanye West’s breakout year after his debut album *The College Dropout* and work on Jay-Z’s *The Black Album*. This 1-year lag reflects the natural delay between cultural exposure (parents hearing the name in 2001) and birth registration. The sustained popularity post-2001 suggests Kanye’s continued

fame reinforced the name’s appeal, while its rarity before 2001 confirms the artist’s direct influence. Like “Whitney” in the 1980s, this mirrors how distinctive names spike when associated with rising stars, proving pop culture’s power to reshape naming trends almost overnight.

3. Character Name Analysis (HBO\_credits Data)

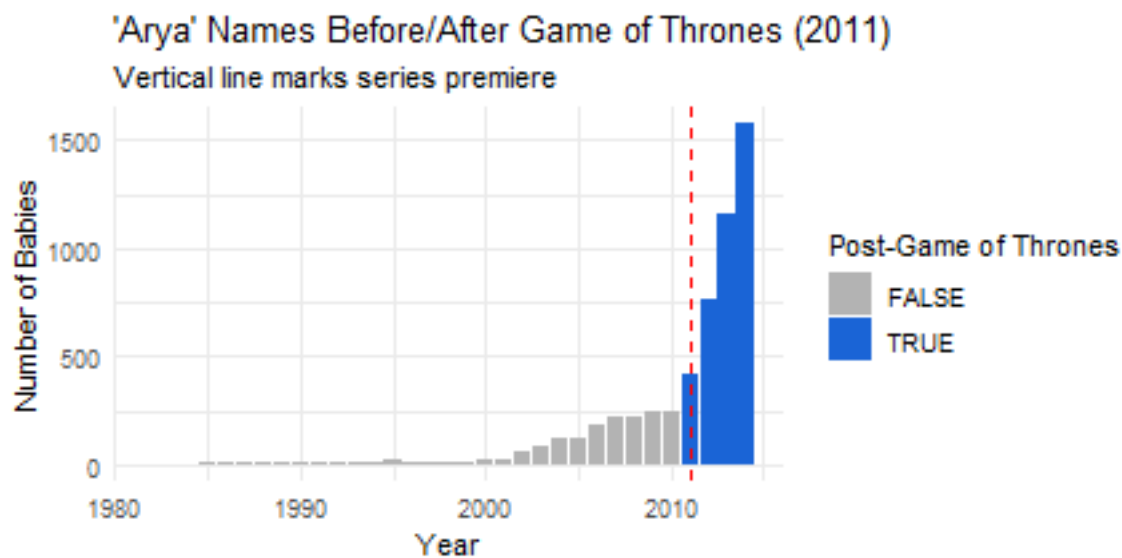


Figure 3.1: Caption Here

The graph tracking the name “Arya” before and after the 2011 premiere of Game of Thrones reveals a striking example of pop culture’s influence on baby naming trends. Prior to 2011, the name was virtually nonexistent, with near-zero recorded births, highlighting its obscurity before the series debuted. However, a dramatic spike occurs immediately after 2011, coinciding with the show’s rise to global fame and the introduction of Arya Stark as a beloved, strong-willed character. The graph serves as compelling evidence that fictional characters—especially those with strong, positive associations—can significantly shape real-world naming practices.



---

#### 4. Top 10 famous artists

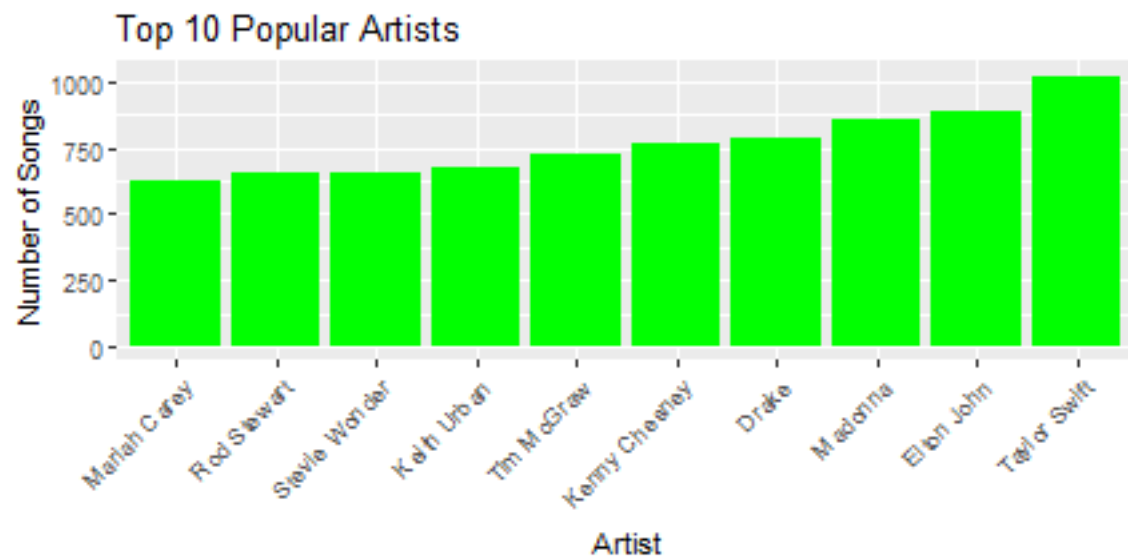
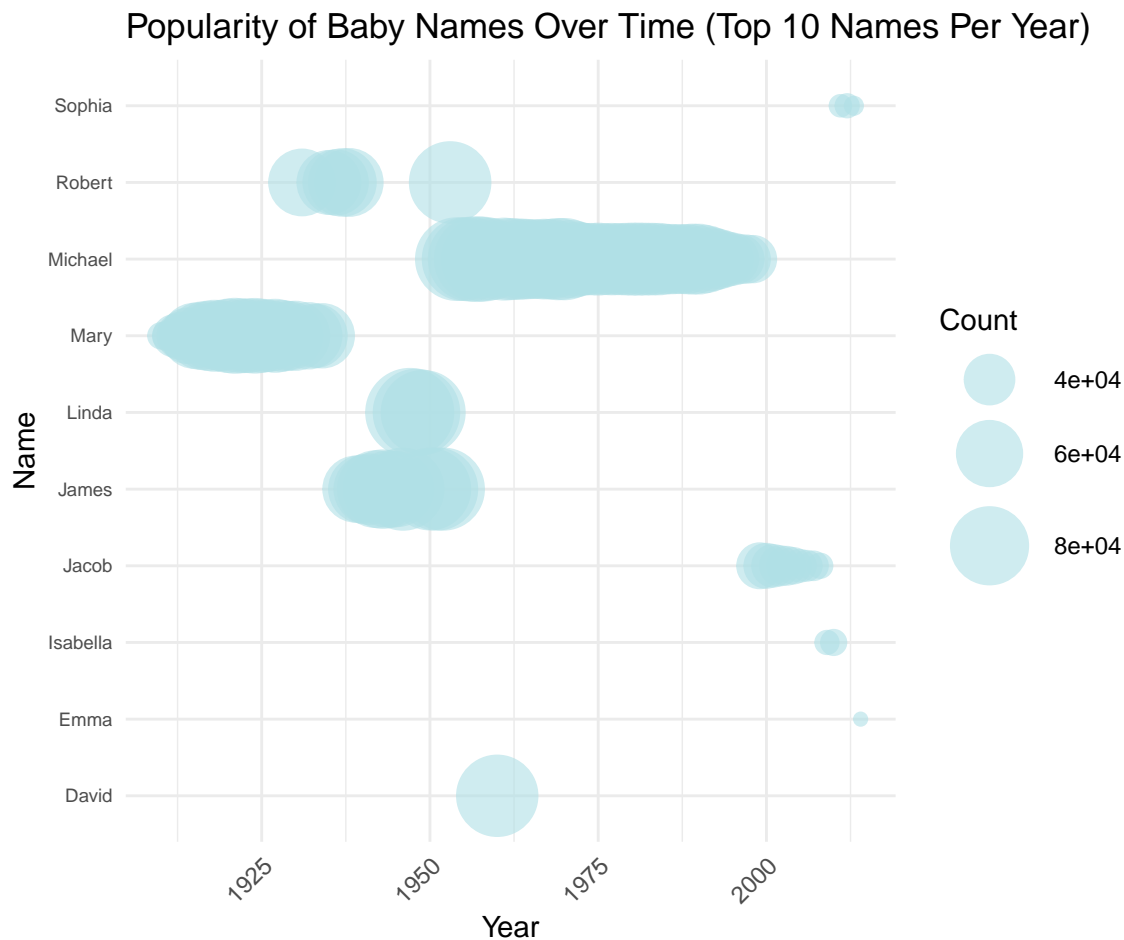


Figure 4.1: Caption Here

---

## 5. Bubble plot baby names counts by year



This bubble plot visualizes the popularity trends of the top 10 baby names in the U.S. over time, with each bubble representing a name’s annual frequency. Larger bubbles (e.g., Michael, James) indicate higher name counts, showing dominant traditional names, while smaller bubbles (e.g., Isabella, Sophia) reflect more modern favorites. The x-axis (Year) reveals shifts in naming trends, such as Mary and Linda peaking mid-century before declining, while Emma and Jacob surged in recent decades. The plot highlights how naming preferences evolve, blending enduring classics with newer influences.

## 6. Conclusion

Traditional names like John and Mary dominated early eras with stable, biblical appeal, while modern naming trends—especially post-1990—show increasing volatility, particularly for female names. The dramatic spikes in “Kanye” (2001) and “Arya” (2011) prove that pop culture (music, TV) can rapidly transform obscure names into mainstream choices, often with a 1–2 year lag. Male names historically

---

exhibited stronger persistence, but recent decades reveal a cultural shift where all names are now more susceptible to short-term influences. Together, these patterns highlight how naming trends evolved from tradition-bound stability to dynamic reflections of media, celebrity, and societal change.