Baby Names

Fun with Statistics

David R. Miller

{dave}@millermattson.com

October 11, 2025

Abstract

This is a report of various statistical measures of the U.S. Social Security "Baby Names" database. Analyses include trends and predictions about popular names, surges in specific name occurrences over time, trends in name complexity and pronounceability, palindromic names, sexually ambiguous names, "forgotten" names that have fallen out of use, and more. Some of the unexpected findings include a trend toward less pronounceable names, and lexical evidence that girls are not pink and boys are not blue.

1 Introduction

The U.S. Social Security Administration maintains a database of the first names given to newborns when parents apply for Social Security numbers at birth.[1] The data is organized by sex and by year for the 144-year span from 1880 through 2024. For security reasons, names that occur fewer than five times in any given year are excluded from the database. That leaves data for 371,807,666 births. Let's have some fun with the data.

---

[1]Social Security Administration, "Baby Names from Social Security Card Applications - National Data," Data.gov, metadata updated July 4, 2025, accessed October 8, 2025, https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-national-data.
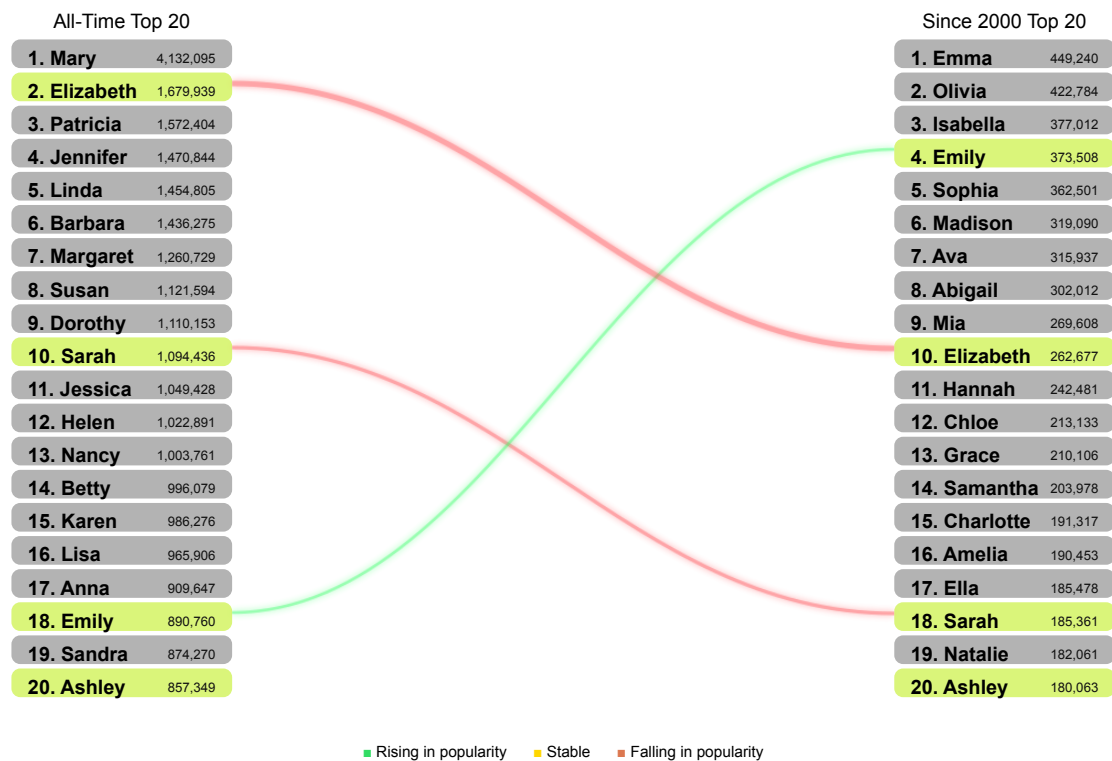
Girl Names: All-Time vs. Since 2000

Figure 1: Girl names

2 Common names

Figures 1 and 2 show the most common names for girls and boys. The left side shows the ranking over the entire 144-year span of data; the right side shows the ranking for the years 2000 through 2024.

The green highlighting shows which names occur both in the top 20 in all time and since 2000. Eleven of the 20 historically common boy names are still popular in recent history, while only four girl names remain popular. This suggests a more conservative practice in naming boys compared to naming girls. Perhaps this reflects cultural attitudes about gender roles and change, where traditional masculinity has historically been identified by stability and continuity, while women's roles have undergone more dramatic transformations over the past few decades.

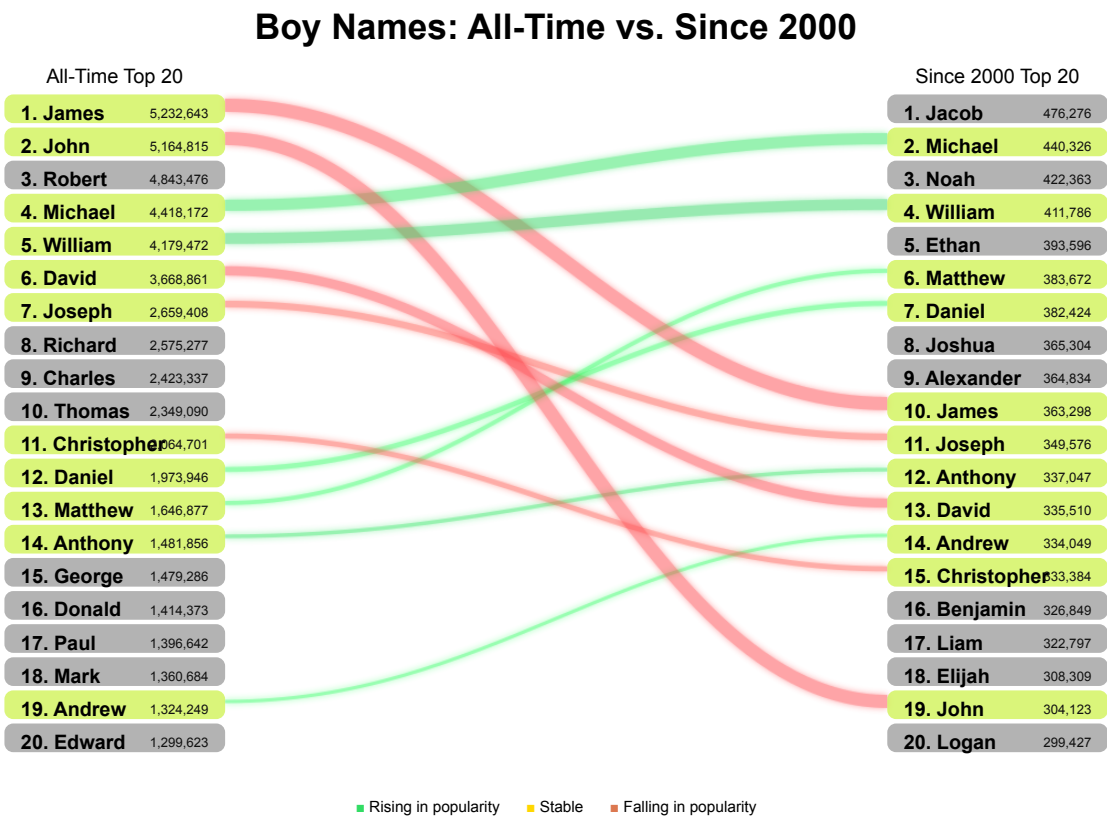I'll leave further speculation for the sociologists.

Boy Names: All-Time vs. Since 2000

All-Time Top 20

1. James	5,232,643
2. John	5,164,815
3. Robert	4,843,476
4. Michael	4,418,172
5. William	4,179,472
6. David	3,668,861
7. Joseph	2,659,408
8. Richard	2,575,277
9. Charles	2,423,337
10. Thomas	2,349,090
11. Christopher	2,064,701
12. Daniel	1,973,946
13. Matthew	1,646,877
14. Anthony	1,481,856
15. George	1,479,286
16. Donald	1,414,373
17. Paul	1,396,642
18. Mark	1,360,684
19. Andrew	1,324,249
20. Edward	1,299,623

Since 2000 Top 20

1. Jacob	476,276
2. Michael	440,326
3. Noah	422,363
4. William	411,786
5. Ethan	393,596
6. Matthew	383,672
7. Daniel	382,424
8. Joshua	365,304
9. Alexander	364,834
10. James	363,298
11. Joseph	349,576
12. Anthony	337,047
13. David	335,510
14. Andrew	334,049
15. Christopher	333,384
16. Benjamin	326,849
17. Liam	322,797
18. Elijah	308,309
19. John	304,123
20. Logan	299,427

■ Rising in popularity   ■ Stable   ■ Falling in popularity

Figure 2: Boy names

3 Name length

Figure 3 shows the average length of baby names weighted by count.

In the 1800s, girl names tended to be shorter. The top 10 in the 1880s included names such as Mary, Anna, Emma, and Ida. One of my own grandmothers born in 1887 was named Oe. The baby names database contains 155 different two-letter first names.

The maximum average name length occurred around 1990 as a result of longer names such as Brittany, Samantha, Stephanie, and Elizabeth appearing in the top 10.

After the length of names peaked around 1990, there has been a clear trend toward shorter names. As shown in the figure, if we fit a 4th-order polynomial to the data and project into the future, then in another few decades, my grandmother's name Oe might come back into fashion.
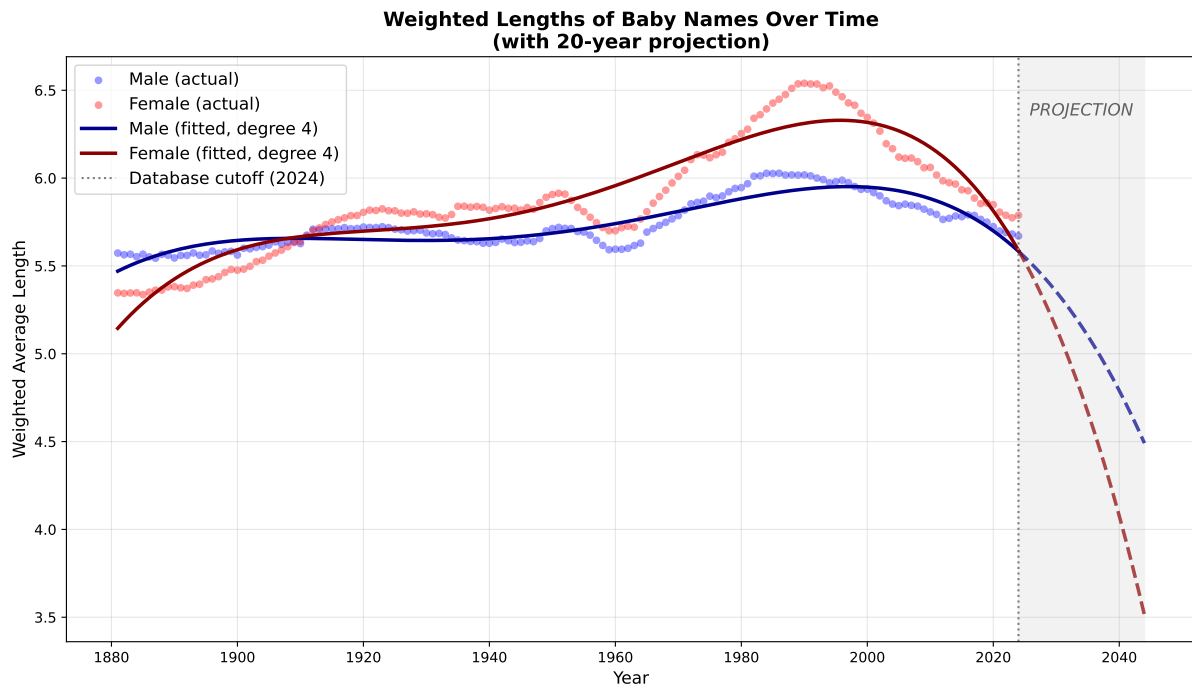
Figure 3: Name length

4 Name endings

This is an analysis of morphological similarities in name endings where only the first letter of the name is different. Table 6 shows the largest groups in order of popularity.

This analysis was limited to the 2000 most popular male and female names in each year over the period 1880 through 2024.

Table 1: Popular endings

Ending	Examples
-arry	Larry, Harry, Barry, Garry, Darry, Marry, Karry, Carry, Tarry, Parry
-ara	Sara, Tara, Kara, Cara, Lara, Mara, Zara, Dara, Yara, Ara, Vara, Fara, Para, Eara
-ina	Tina, Gina, Nina, Ina, Dina, Lina, Mina, Vina, Zina, Rina, Jina, Sina, Bina, Kina, Pina
-ellie	Nellie, Ellie, Kellie, Mellie, Dellie, Cellie, Zellie, Tellie, Lellie, Pellie
-ennie	Jennie, Bennie, Pennie, Dennie, Tennie, Gennie, Vennie, Kennie, Rennie, Mennie, Hennie, Zennie, Sennie, Ennie
-ena	Lena, Rena, Dena, Jena, Gena, Tena, Nena, Zena, Xena, Vena, Ena, Sena, Kena, Bena, Cena
-allie	Callie, Allie, Sallie, Hallie, Kallie, Vallie, Mallie, Dallie, Lallie, Pallie, Wallie, Tallie
-ollie	Ollie, Mollie, Dollie, Hollie, Rollie, Pollie, Collie, Lollie, Zollie, Tollie, Vollie, Sollie, Nollie

5 Sexually ambiguous names

Table 2 shows names that are used in nearly equal number for girls and boys, ordered by the percentage of overlap.

Table 2: Sexually ambiguous names

Rank	Name	Male	Female	Ratio	Overlap	Total
1	Eslie	60	60	1.00:1 F	50.0%	120
2	Meco	43	43	1.00:1 F	50.0%	86
3	Najay	26	26	1.00:1 F	50.0%	52
4	Zyian	49	49	1.00:1 F	50.0%	98
5	Swae	38	38	1.00:1 F	50.0%	76
6	Wisdom	983	982	1.00:1 M	50.0%	1,965
7	Emanuelle	196	195	1.01:1 M	49.9%	391
8	Seneca	1,741	1,732	1.01:1 M	49.9%	3,473
9	Tahjae	121	122	1.01:1 F	49.8%	243
10	Mycah	827	820	1.01:1 M	49.8%	1,647
11	Tristyn	2,222	2,202	1.01:1 M	49.8%	4,424
12	Emerson	32,113	31,797	1.01:1 M	49.8%	63,910
13	Jayel	86	87	1.01:1 F	49.7%	173
14	Shine	255	252	1.01:1 M	49.7%	507
15	Hillery	501	495	1.01:1 M	49.7%	996
16	Cree	1,089	1,103	1.01:1 F	49.7%	2,192
17	Ronne	67	68	1.01:1 F	49.6%	135
18	Landry	5,647	5,562	1.02:1 M	49.6%	11,209
19	Mikele	66	65	1.02:1 M	49.6%	131
20	Sarang	66	65	1.02:1 M	49.6%	131

6 Girls are not pink and boys are not blue

We associate pink with girls and blue with boys. Does this convention affect how we name our baby girls and boys, perhaps at a subconscious level? Let's find out.

In this [dubious] analysis, we will look at every unique boy and girl name and see if the name contains the name of a color. For example, "Tristan" contains the word "tan" and "Goldie" contains "gold."

The algorithm works like this:

1. Construct a list of 95 different color names and their RGB values extracted from various online color charts.
2. Construct a list of all the unique baby names, their sex, and the number of babies with each name.
3. Examine each name to see which color(s), if any, appear as part of the baby name.
4. Accumulate all the RGB values for all the colors found in all the baby names, weighted by the number of babies with each name.

Table 3: Top names containing colors

Name	Sex	Count	Color
Mildred	F	452,763	red
Brittany	F	361,465	tan
Fred	F	344,027	red
Violet	F	178,215	violet
Constance	F	140,236	tan
Tanya	F	117,217	tan
Olive	F	65,475	olive
Winifred	F	45,823	red
Freda	F	42,499	red
Goldie	F	38,501	gold
Stanley	M	302,514	tan
Frederick	M	265,946	red
Oliver	M	255,373	olive
Alfred	M	244,381	red
Jared	M	206,307	red
Tristan	M	129,641	tan
Grayson	M	127,278	gray
Tanner	M	99,478	tan
Freddie	M	83,492	red
Meredith	M	81,382	red

Table 3 lists the top 10 girl and top 10 boy names (by count) that contain the name of a color.

As readily evident in the table, cool colors are not well represented in baby names, and both girls and boys are predominantly composed of reds and tans. The accumulated average color over all names is shown in Figure 4.

#A53830 Girls
Rusty Red

#AE3118 Boys
Tuscan Red

Figure 4: The colors of girls and boys

7 Palindromes

Table 4 shows names that are palindromes. This analysis was conducted over all names over all years.

Table 4: Palindromic names

Name	Total	Male	Female
Anna	912,391	2,744	909,647
Hannah	462,147	721	461,426
Ava	336,595	336	336,259
Ana	110,656	601	110,055
Ada	105,101	278	104,823
Bob	93,254	92,932	322
Otto	38,498	38,398	100
Eve	28,919	0	28,919
Asa	22,299	20,474	1,825
Elle	18,574	0	18,574
Nan	10,258	0	10,258
Aja	8,359	90	8,269
Aya	7,534	0	7,534
Ara	4,802	1,291	3,511
Aviva	3,552	0	3,552
Emme	3,533	0	3,533
Hanah	2,047	0	2,047
Aziza	1,918	0	1,918
Natan	1,562	1,562	0
Alla	1,191	10	1,181
Ama	1,035	0	1,035
Adda	957	0	957
Ala	921	142	779
Maram	881	0	881
Awa	842	0	842
Nosson	819	819	0
Layal	768	0	768
Sabas	739	739	0
Jj	728	728	0
Nitin	639	639	0

8 Forgotten names

Table 5 shows names that previously appeared more than 1000 times in a year but have not been seen for the past 15 years.

This analysis was conducted over all names over all years.

Table 5: Forgotten names

Name	Peak Year	Peak Count	Last Seen	Total Babies
Laverne	1924	1,383	2003	42,869
Pat	1941	1,900	1991	40,122
Pam	1959	3,677	1996	33,125
Dick	1934	1,131	2005	29,248
Vonda	1965	1,093	2008	10,312
Deneen	1964	1,604	2007	3,603
Christop	1989	1,082	1989	1,082

9 Pronounceability

It is difficult to algorithmically estimate the difficulty for a native English speaker to pronounce a name. We know that even a baby can easily pronounce a name like "Moe," but skilled orators might stumble over "Agastyareddy" (a real name). How can we quantify pronounceability?
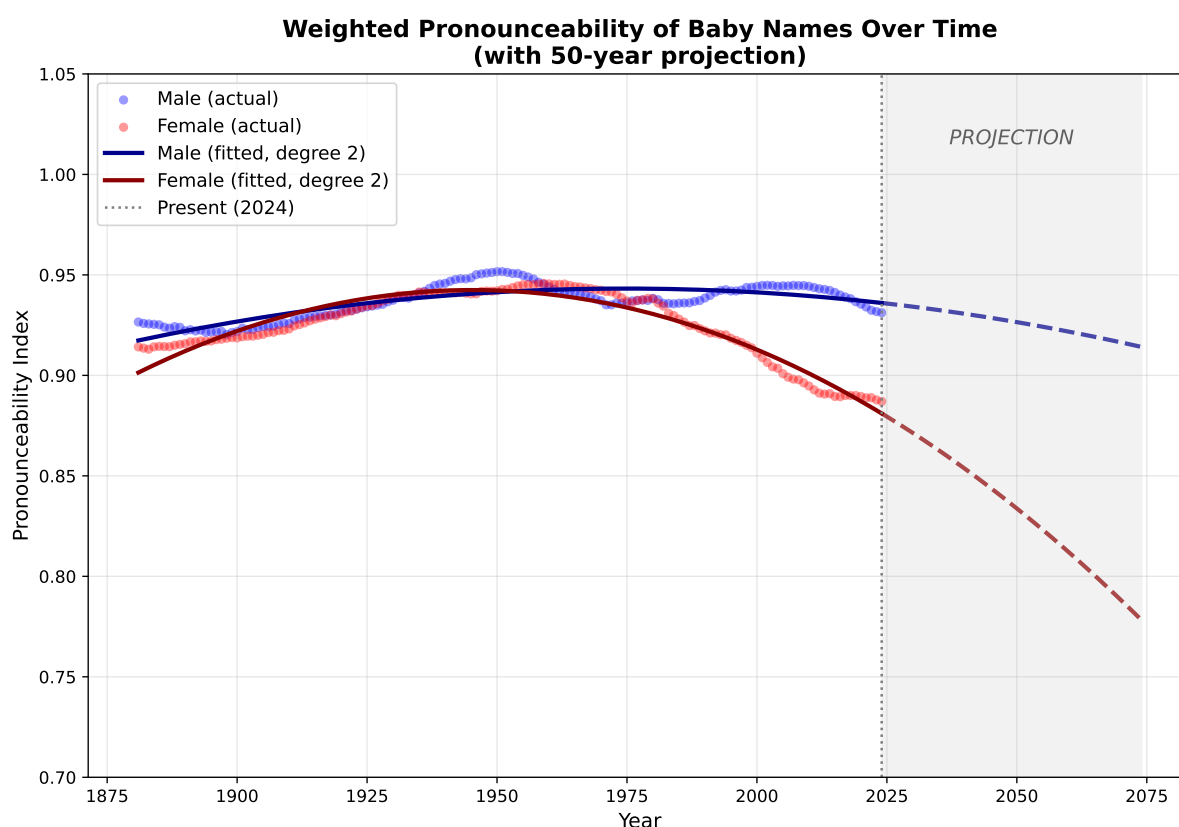

Figure 5: Weighted average name pronounceability, higher is easier to pronounce

This analysis uses a heuristic that estimates pronounceability based on a combination of measures:

- Length of consonant clusters.
- Consonant-to-vowel ratio.
- Consonant density.
- Extra penalty for known difficult consonant sequences.
- Estimated syllable complexity.
- Bigrams and trigrams weighted for known N-gram frequency in English.[2]

---

[2]Lyons, James. "English Letter Frequencies," Practical Cryptography, http://www.practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/english-letter-frequencies/ (accessed Oct 8, 2025).

These measures are weighted and combined to form a pronounceability estimate from 0.0 (impossible to pronounce) to 1.0 (easy to pronounce).

All the names in the baby names database were evaluated by this formula for pronounceabilty and weighted by their counts. The results are plotted in Figure 5.

In the Figure, the dotted lines are a second-order polynomial regression that predicts the trend for the next 50 years. At that rate, in two more centuries, the average baby girl name will be unpronounceable.[3]

10 Spikes

Figure 6 shows names that suddenly increased in popularity.

In 1883, the name Grover was given to 127 babies, but in the next year, 802 babies were given that name, more than a five-fold increase. One of my grandfathers was born in 1885 and given the name Grover. We can assume that the election of Grover Cleveland to the U.S. presidency in 1884 was the principal influence for that surge.

More recently, 1,849 babies were given the name Jayceon in 2013, up from only 182 in the preceding year. Most likely, rapper Jayceon Terrell Taylor was responsible for that surge.

The causes behind the other spikes are left an exercise for the reader.

Table 6 contains the detailed data ranked by $score$, which is somewhat arbitrarily defined as

$$score = \frac{rate_{after}}{rate_{before}} \sqrt{rate_{after}}$$

where $rate_{before}$ and $rate_{after}$ are the rates per 10,000 births before and after the spike.

---

[3]In case the reader has any doubts, this is indeed a dubious application of statistics.
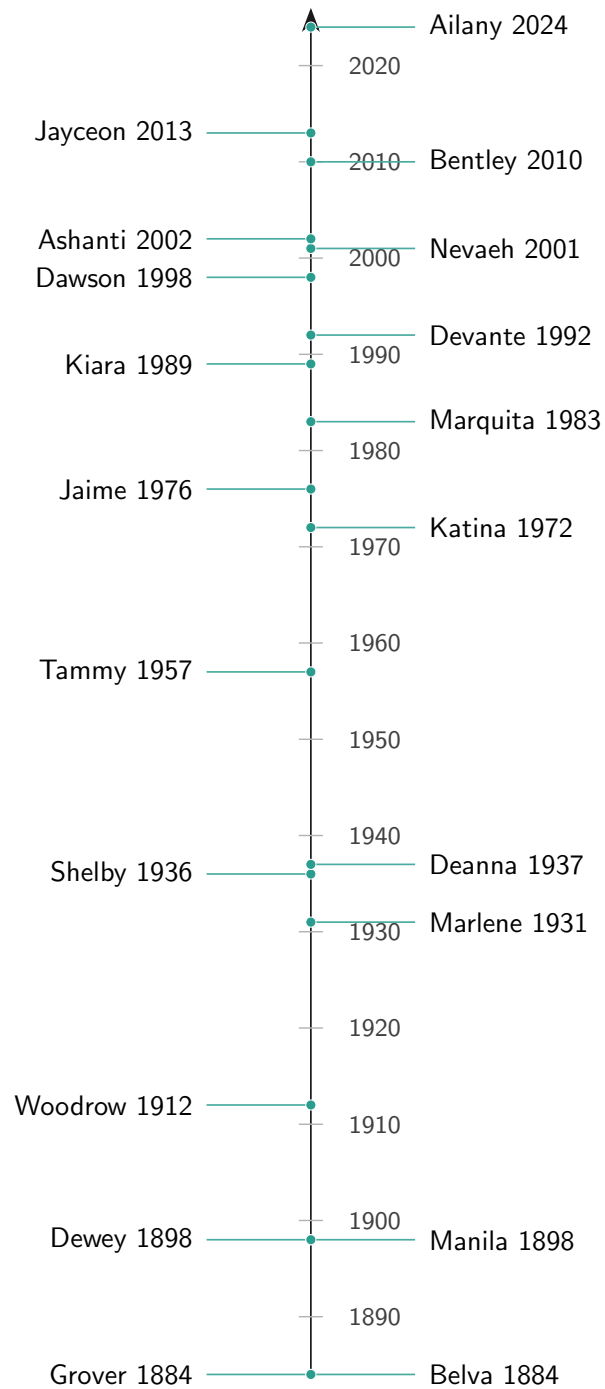
Figure 6: Spikes in name usage

Table 6: Spikes in name occurrences

Rank	Name	Sex	Years	Rate Before	Count Before	Rate After	Count After	Ratio	Score
1	Katina	F	1971→1972	0.28	95	8.73	2745	31.6	93.3
2	Deanna	F	1936→1937	0.37	77	7.62	1624	20.6	56.8
3	Marquita	F	1982→1983	0.36	128	7.34	2543	20.1	54.5
4	Tammy	F	1956→1957	0.62	255	10.39	4365	16.8	54.1
5	Jaime	F	1975→1976	3.03	914	25.83	7838	8.5	43.4
6	Woodrow	M	1911→1912	1.88	121	18.65	1843	9.9	42.9
7	Shelby	F	1935→1936	0.32	67	5.16	1071	16.1	36.5
8	Dewey	M	1897→1898	4.55	158	29.23	1115	6.4	34.7
9	Grover	M	1883→1884	5.85	127	32.94	802	5.6	32.3
10	Marlene	F	1930→1931	1.38	306	12.28	2585	8.9	31.3
11	Kiara	F	1988→1989	0.60	222	6.83	2624	11.4	29.7
12	Ashanti	F	2001→2002	0.75	281	7.88	2946	10.5	29.5
13	Devante	M	1991→1992	0.34	131	4.07	1565	12.1	24.5
14	Bentley	M	2009→2010	1.37	522	10.22	3776	7.5	23.9
15	Ailany	F	2023→2024	0.94	311	7.80	2596	8.3	23.2
16	Jayceon	M	2012→2013	0.50	182	5.07	1849	10.2	23.0
17	Manila	F	1897→1898	0.20	7	2.73	104	13.5	22.3
18	Nevaeh	F	2000→2001	0.26	99	3.20	1199	12.2	21.9
19	Dawson	M	1997→1998	0.58	209	5.15	1895	8.9	20.3
20	Belva	F	1883→1884	0.23	5	2.71	66	11.8	19.4
21	Tristan	M	1994→1995	1.32	492	8.43	3088	6.4	18.5
22	Mandy	F	1974→1975	1.47	446	8.98	2710	6.1	18.3
23	Kayla	F	1981→1982	0.92	317	6.49	2276	7.1	18.0
24	Brianne	F	1978→1979	0.64	202	5.00	1662	7.8	17.5
25	Farrah	F	1975→1976	0.27	82	2.76	836	10.1	16.8
26	Audra	F	1965→1966	0.25	90	2.57	892	10.3	16.6
27	Hobart	M	1895→1896	0.46	16	3.78	135	8.3	16.1
28	Bret	M	1957→1958	0.46	194	3.66	1514	7.9	15.2
29	Iesha	F	1990→1991	0.72	285	4.87	1896	6.8	14.9
30	Alondra	F	1994→1995	0.40	149	3.29	1205	8.2	14.9

11 Conclusion

In this report, all of the statistical analysis is mathematically precise, but statistics have to be interpreted in context. Remember that figures don't lie, but liars can figure.