

One Hundred Years of US Baby Names

June 30, 2020

Group 4

Dhara Bhansali

Sarah Grant

Keke Li

Matt Debnar

Overview

Analysis of 5.6 MM rows of US baby name data by state and gender from 1910 thru 2014.

Data source: US Baby Names 1910 thru 2014

Project History

Found NYC Baby Name Data 5 years

Suggested move to Kaggle US Baby Name Data 100 years

Imported, checked and cleaned for nulls

Added Decade ($x//10*10$), Added Region (`np.where` nested)

Doc questions with tasks

Outline Questions

- **Q1 Are Names Getting Longer or Shorter? More or less prevalent by region?**
DB
Name Length vs Popularity over Time - per year, per gender then per decade, then per region
- **Q2 How many names are used by both genders? More or less over time? More or less prevalent by region in total? KL**
Child Gender vs Name Percent Match
- **Q3 How many new names are added each year? More or less prevalent by region? MD**

Datasets Used

- US Baby Names
<https://www.kaggle.com/kaggle/us-baby-names/data?select=StateNames.csv>
- Popular Baby Names in NYC: [Popular Baby Names](#)
Behind the Name API (Gender, Usage):
<https://www.behindthename.com/api/help.php>

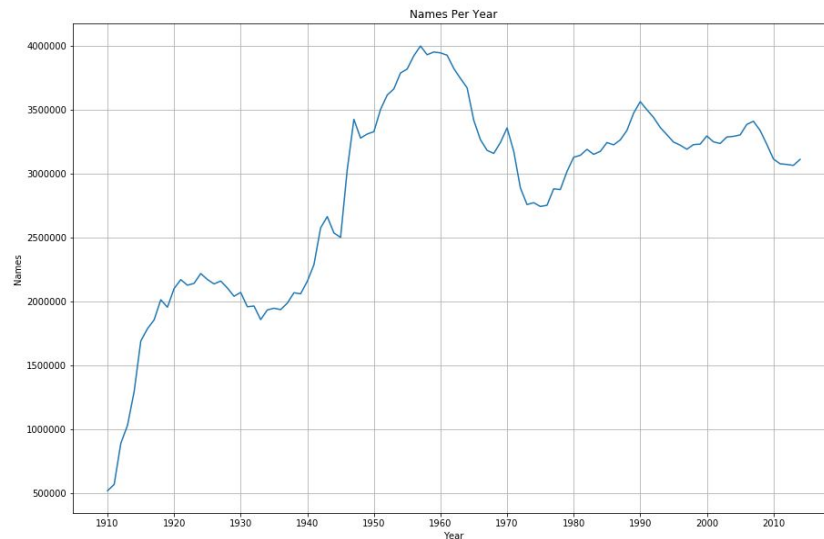
- To reproduce for “How many names are used by both genders?”
<https://babynames1000.com/gender.php?y=2016>

Github Link

- **Presentation**
<https://github.com/bbixby/BootcampProject/blob/master/US%20Baby%20Names%20Presentation%20Fixed.ipynb>
- **Repo:**
<https://github.com/bbixby/BootcampProject>

Summary

Total Names per Year

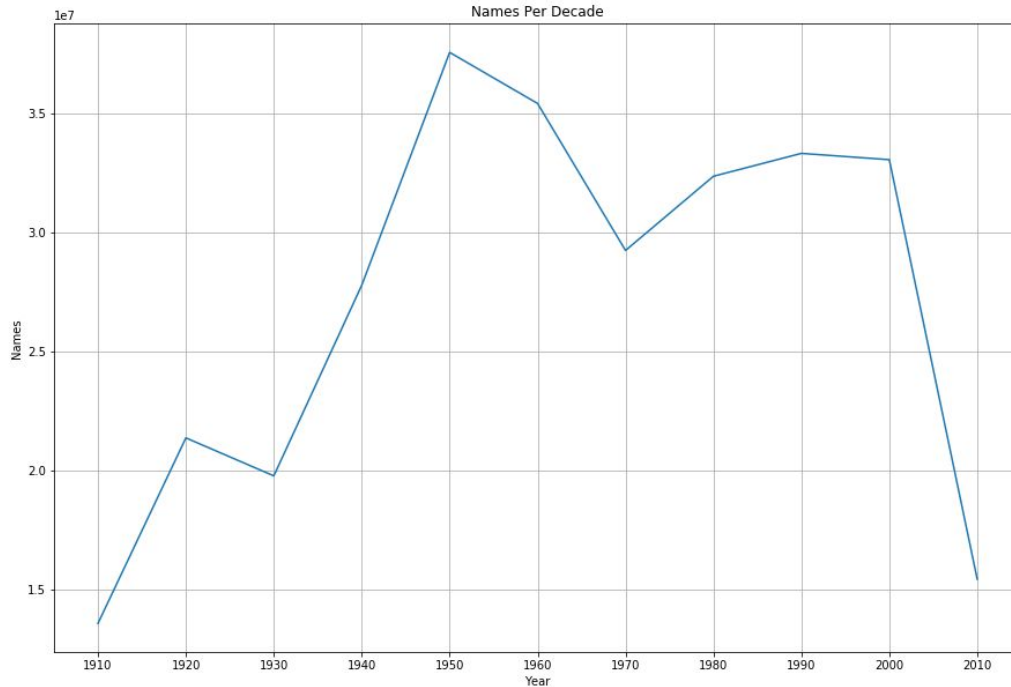


Observation: The baby boom is real! Count of names jumped at the end of WWII in 1945 and peaked just before 1960

Takeaways

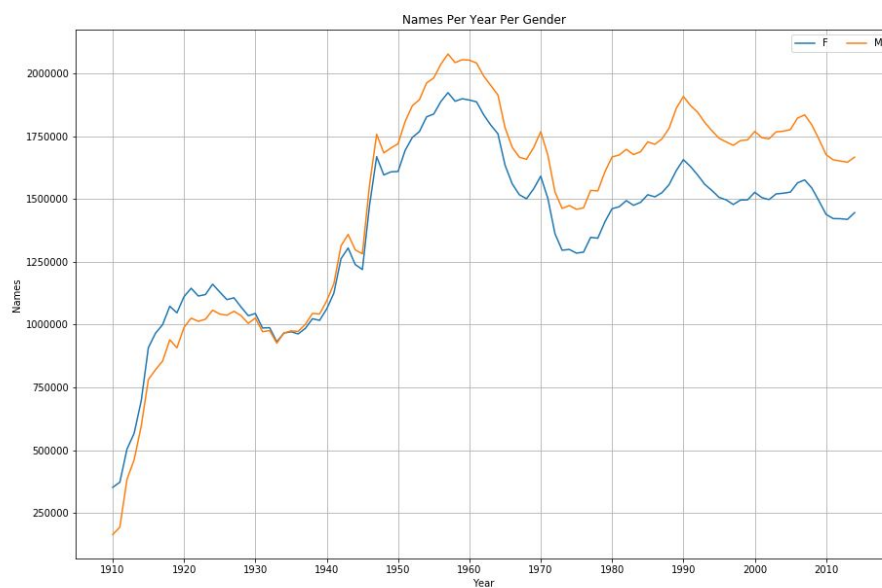
- Between 1910 and 2014, the number of names increased over time.
- The first sharp incline in baby names occurs in the mid-1940s (from 250,000 to just under 350,000), which correlates to WWII ending in 1945.
- These births will mark the beginning of the largest generation in the U.S. (until the millennials), the Baby Boomers.
- The sharp incline continues and peaks in the late-fifties and early sixties (circa 1957 - 1962), the period yielding the greatest volume of names to date -- these names signify the end of the Baby Boomers.

Names per Decade



Observation: baby boom confirmed in 1950s then slight bounce back until 2000s (note: 2010 is a partial decade, only half thru 2014)

Names per Gender per Year

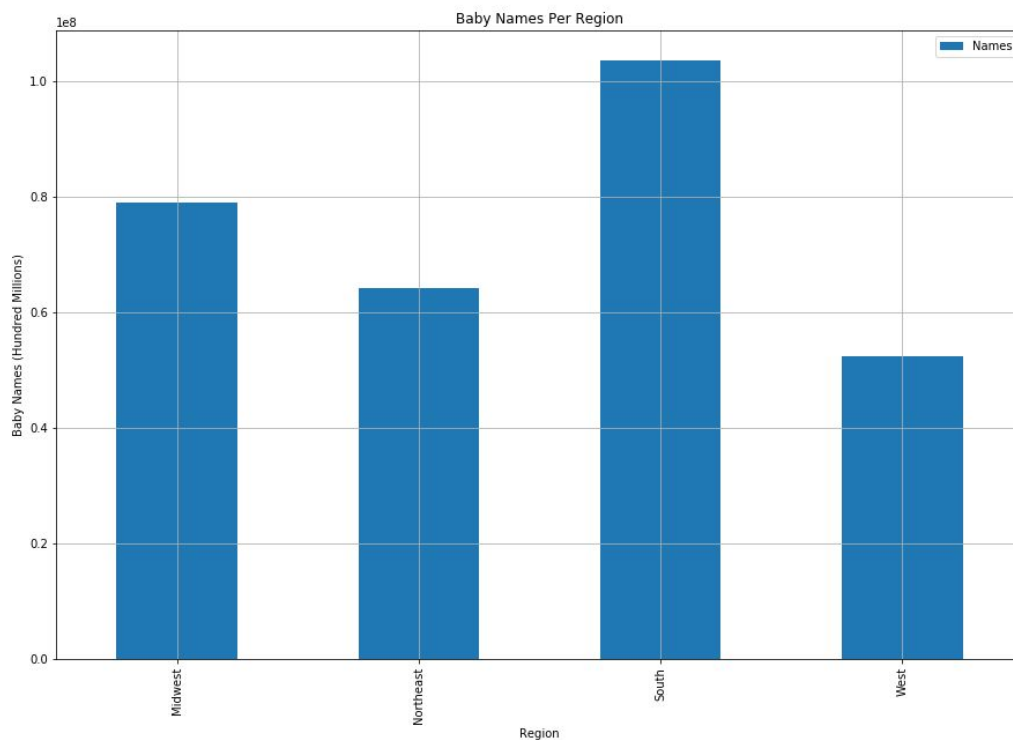


Observation: girls names outpaced boys until 1930; after 1930 the number of boys' names outpaced girls' consistently thru 2014

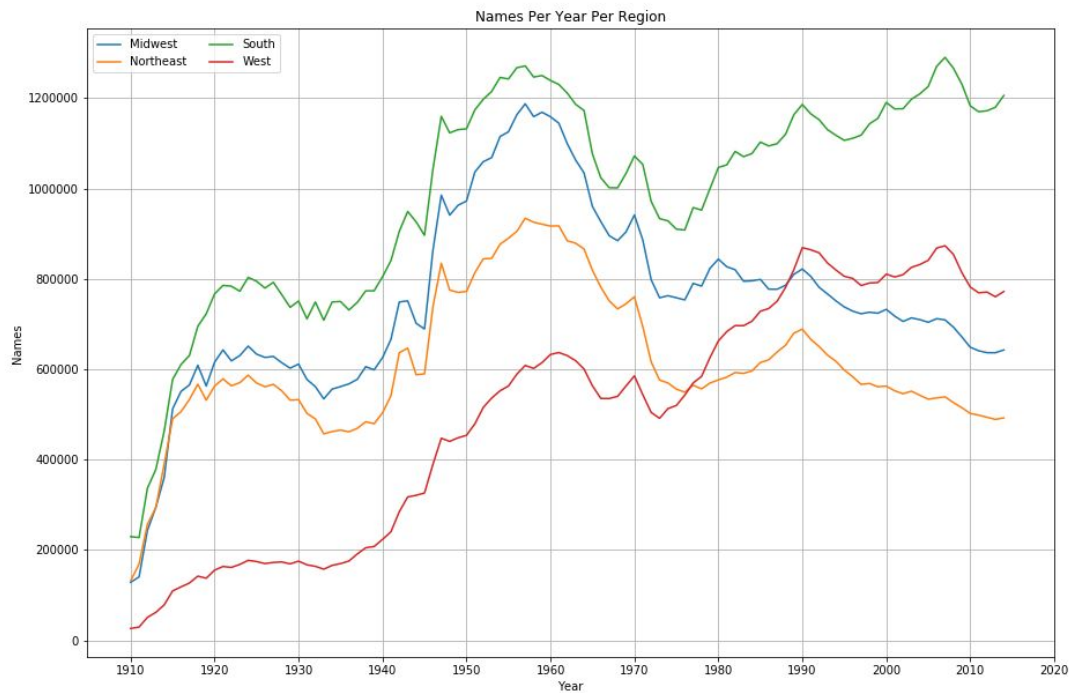
Takeaways

- The trends for male and female baby names are parallel to one another, but in the latter half of the century, they are not equal.
- There are significantly more male baby names beginning in the early 1970s -- a trend that remains the same through 2014.
- One reason for the increase in male baby names is the rising trend in unisex names, many of which are considered “male.”
- For example, the name “Madison” was historically considered a male name, but after the 1984 movie “Splash” (in which the female protagonist names herself Madison after the New York City street), there was a marked rise in U.S. females named Madison. “Madison” was ranked second for girls in the United States by 2001. This rise from obscurity to prominence in only 18 years represents an unprecedented 550,000 percent increase in usage.
- The name “Taylor” is another example of this trend of a traditionally male name that has been co-opted as feminine in modern times.

Total Names per Region



Names per Region per Year



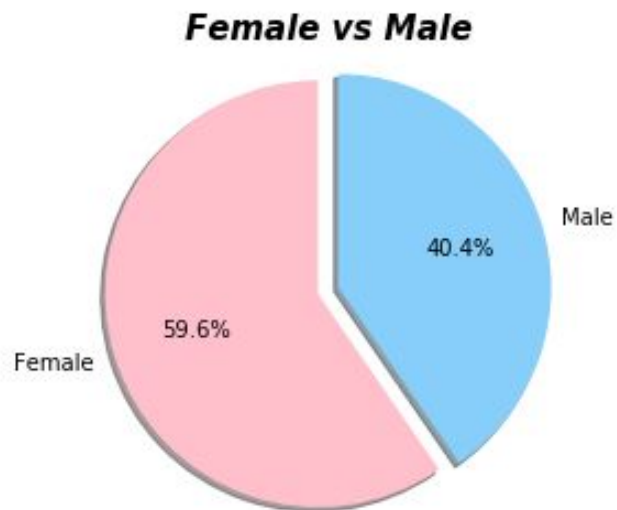
Observation: the South consistently registers the most names. The Midwest and Northeast counts peaked just before 1960 then fell off. The West gained in name counts starting in 1940 then increased again in 1975 to gain second place

Top Name Overall per Gender

	Name	Popularity
Name		
James	James	4957166.0
John	John	4845414.0
Robert	Robert	4725713.0
Michael	Michael	4312975.0
William	William	3839236.0
Mary	Mary	3740495.0
David	David	3562278.0
Richard	Richard	2534949.0
Joseph	Joseph	2485220.0
Charles	Charles	2252146.0

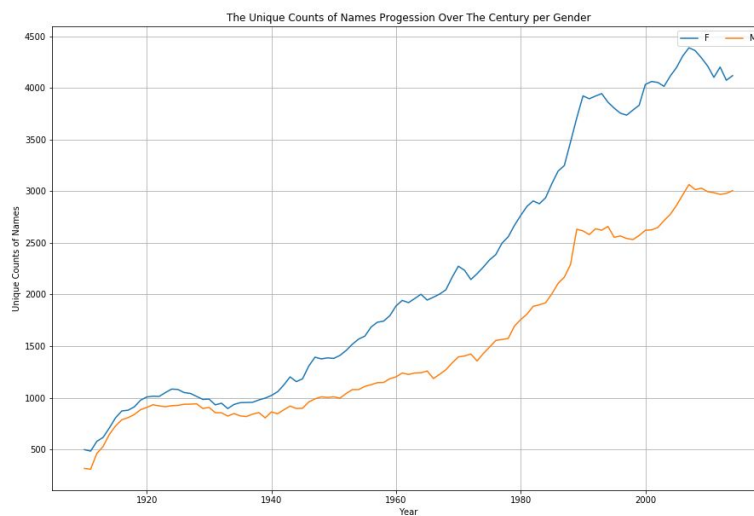
Observation: most of the top names are Male. Given more Female names registered than Male suggests more variability in Female names than Male

Q2: The Total Unique Names in Both Gender

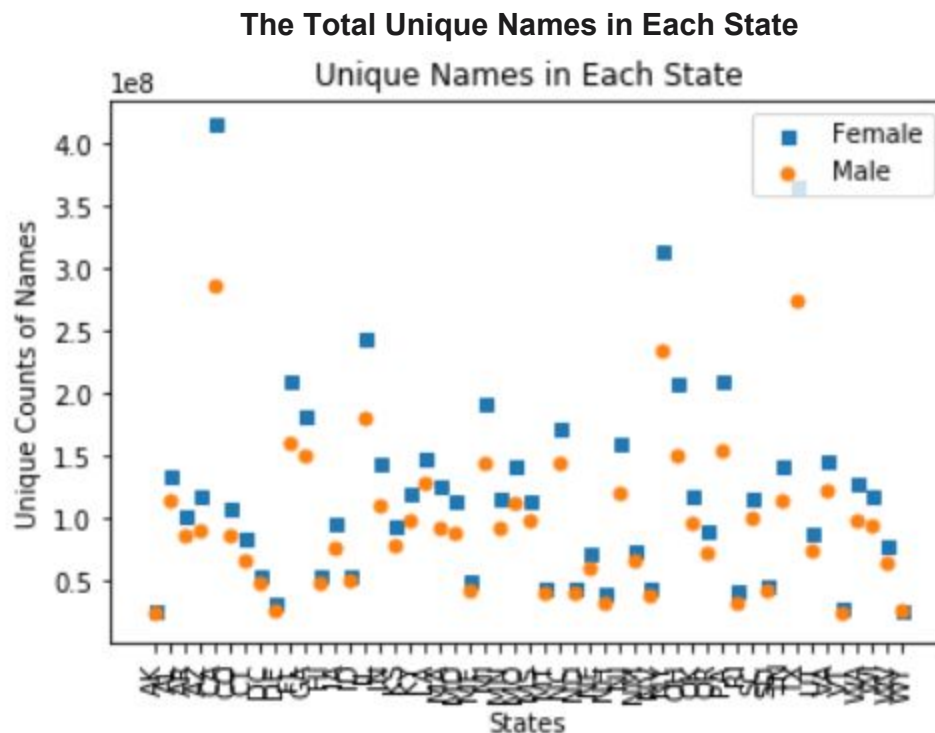


Proof! Female names have more variety than Male names

The Unique Counts of Names Progression Over The Century per Gender



Observation: The name uniqueness has a significant increase with a new influx of immigrants from the 1990s onward.



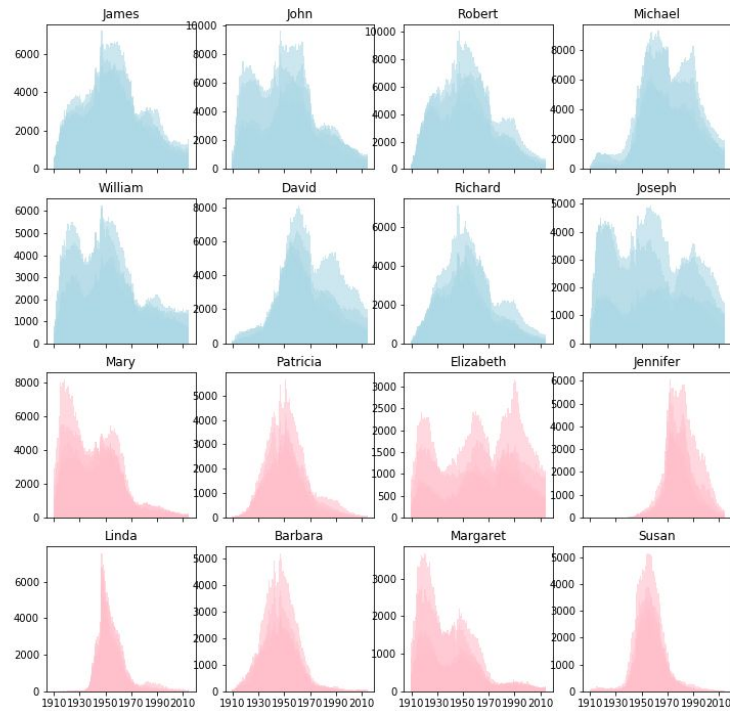
Observation: states CA and NY have the highest number of unique names across all states

Uniqueness

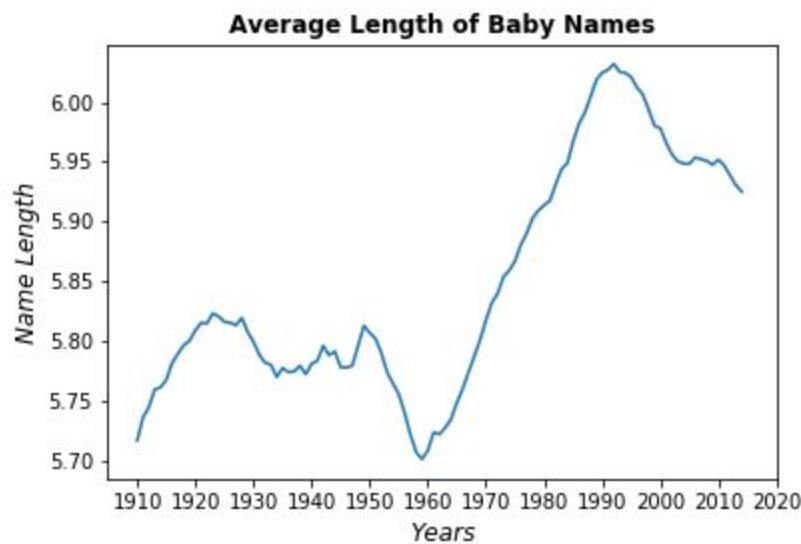
- Overall, there is an increase in “new” baby names overall, and particularly from 1970 onward.
 - One reason for this is U.S. population growth in general (Baby Boomers... now millennials). There are more names, therefore, more opportunities for newness.
 - Another explanation for more “new” names is cultural: the data suggest that modern parents are less likely than their predecessors to name their child after a family (“junior”). These new parents value individualism more than tradition.
 - There is a correlation between the volume of new names and increased immigration during peacetime.

Q3. To find the Name length from the year 1990-2014

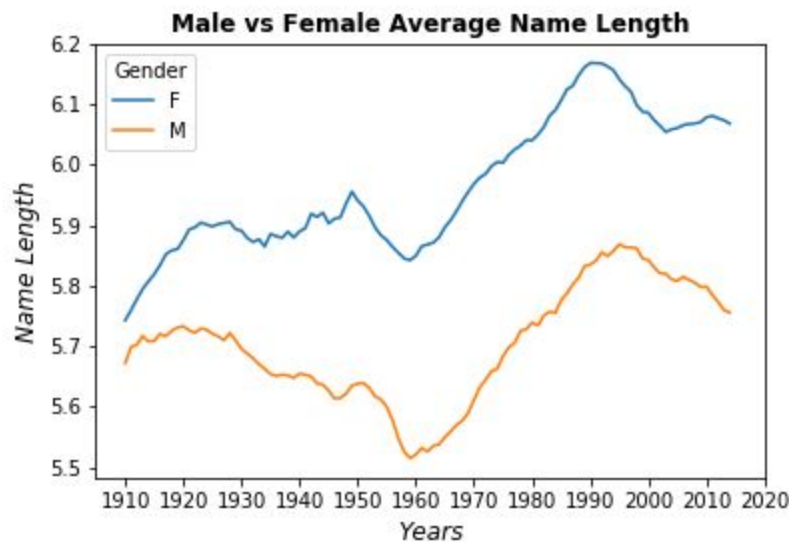
Trends for the most popular names



Observation: most Boys' names peak around 1960; Joseph and Michael remain popular. Most Girls names also peak in 1950 or 1960, Elizabeth and Jennifer remain popular



Observation: The baby names length started increasing from the year 1960 and reached a peak in the year 1990.



Observation: Looking at the pattern, average female names length are longer than males names length with a peak at 1990.

We also removed the Mean and standard Deviation on the Name Length and we could find out that there was no much variation in the Mean and the Standard Deviation.

Select national events that impacted “new names”:

- **Turn-of-the-century immigration boom:** Circa 1910, one in four American workers were foreign-born (according to the U.S. Bureau of Immigration).
- **The Great Depression and quota system hurt immigration** (thus slowing “new names”). The National Origins Act’s quota system, which took effect in 1929, diminished the wave of immigration from Southern and Eastern Europe. **The Great Depression** and **WWII** further impeded immigration into the U.S., so “new names” slumped during this time.
- The United States’ percentage of foreign-born citizens reached a **low of 4.7% in 1970**.
- **After 1970**, when Mexican, Chinese, Filipino and Indian immigration saw a massive rise, the incidence of “new names” also increased, per the data.
- “New names” in the U.S. are likely coming from the country’s newest immigrant groups, from China or India. **Those countries surpassed Mexico as countries of origin for immigrants arriving in the U.S. in 2013.**

Conclusions

Baby names are a window into looking at cultural trends in the U.S.

- *The diversity of baby names has increased over the last 100 years.*
- *There is a greater variety of female names than male names. However, male names have been consistently more popular over the last 50 years.*
- *The proliferation of baby names may be attributable to expanded immigration policies in the U.S.*