# Analysis of U.S. Baby Names

ERIN WITMER, CS440

January 2019

## 1 Introduction

This report analyzes trends in baby names in the United States between 1880 and 2017. This data is made available by the U.S Social Security office. The analysis includes the growth in unique male and female names, basic statistical analysis and a deeper look into two interesting conclusions.

## 2 Conclusions

The first trend analyzed was the growth in unique names in the U.S. between the years 1880 and 2017. In 1880, there were 2000 unique names (male and female, unisex names are counted twice). 9,655 boys were given the most popular name that year, John. In the same year, 7,065 girls were given the name Mary.

| | total | | | | | | | |
| | count | mean | std | min | 25% | 50% | 75% | max |
| year | | | | | | | | |
| **1880** | 2000.0 | 100.742000 | 466.108304 | 5.0 | 7.0 | 13.0 | 41.25 | 9655.0 |
| **1881** | 1935.0 | 99.584496 | 438.355601 | 5.0 | 7.0 | 13.0 | 42.00 | 8769.0 |
| **1882** | 2127.0 | 104.152797 | 468.469089 | 5.0 | 7.0 | 13.0 | 42.50 | 9557.0 |
| **1883** | 2084.0 | 104.100768 | 448.153015 | 5.0 | 7.0 | 13.0 | 43.00 | 8894.0 |
| **1884** | 2297.0 | 105.991293 | 466.053710 | 5.0 | 7.0 | 13.0 | 40.00 | 9388.0 |

Summary 1880-1884

By 2017, the number of unique names had grown to 33,469.

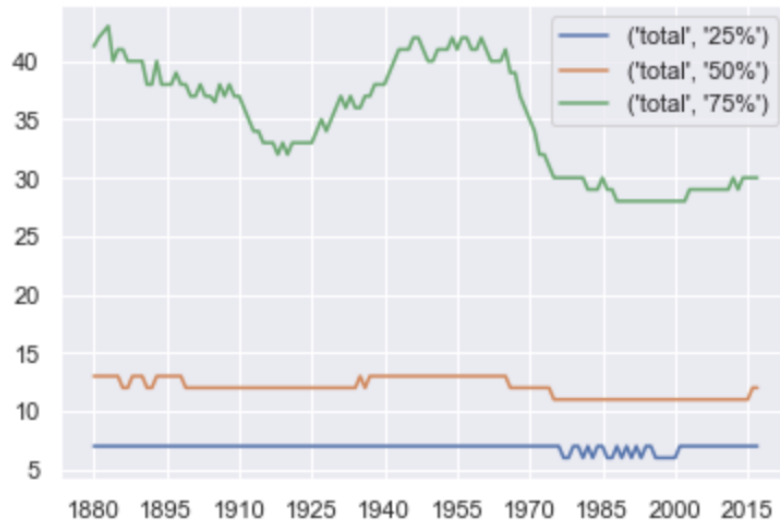|  | total | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | count | mean | std | min | 25% | 50% | 75% | max |
| year | | | | | | | | |
| **2013** | 33269.0 | 109.330307 | 678.812266 | 5.0 | 7.0 | 11.0 | 29.0 | 21213.0 |
| **2014** | 33228.0 | 111.240851 | 681.177437 | 5.0 | 7.0 | 11.0 | 30.0 | 20924.0 |
| **2015** | 33098.0 | 111.447429 | 674.338786 | 5.0 | 7.0 | 11.0 | 30.0 | 20435.0 |
| **2016** | 32979.0 | 110.766488 | 659.810676 | 5.0 | 7.0 | 12.0 | 30.0 | 19471.0 |
| **2017** | 32469.0 | 109.221134 | 639.965484 | 5.0 | 7.0 | 12.0 | 30.0 | 19738.0 |

Summary 2013-2017

However, the number of unique names has been trending downward over the past 10 years.
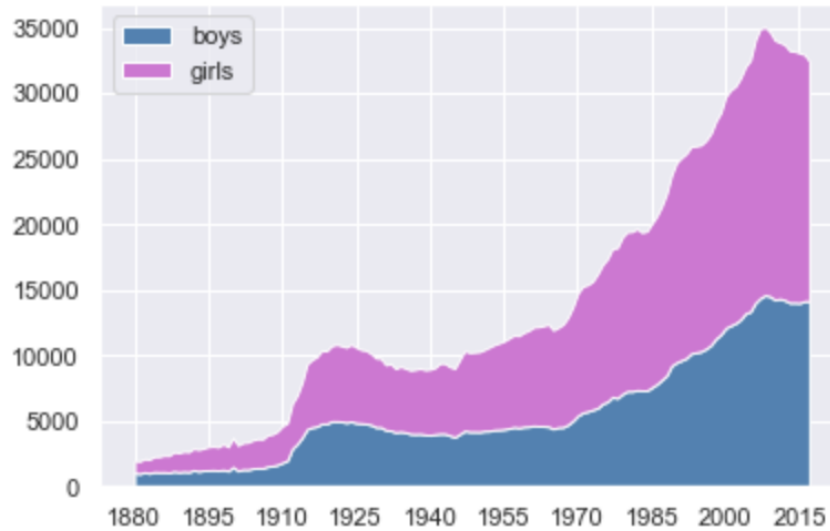


Number of Unique Names

One might assume that most of the growth in unique names comes from the tail, a proliferation of names with a small number of children given the name per year. An analysis of the year by year frequency distribution of names shows that is not the case. There is little change in the average number of children born with a name over time. In sum, there is a very long tail of names that are not very popular, however this is not a new phenomenon. It has always been the case.

2

Statistics on Frequency Distribution by Year

Turning the analysis to top quartile, or popular names, there is a noteworthy trend. This analysis shows the concentration of popular names by looking at the percentage of total names that the ten most popular female names accounted for each year. In 1880, approximately 25 percent of girls born were given one of the top 10 most popular names. By 2017, only approximately 8 percent of girls born were given one of the 10 most popular names, and the top 60 names accounted for 25 percent of births. In conclusion, the most noteworthy trend is that now there are many more "popular" names. The proliferation has been in the top quartile, not the bottom quartile.
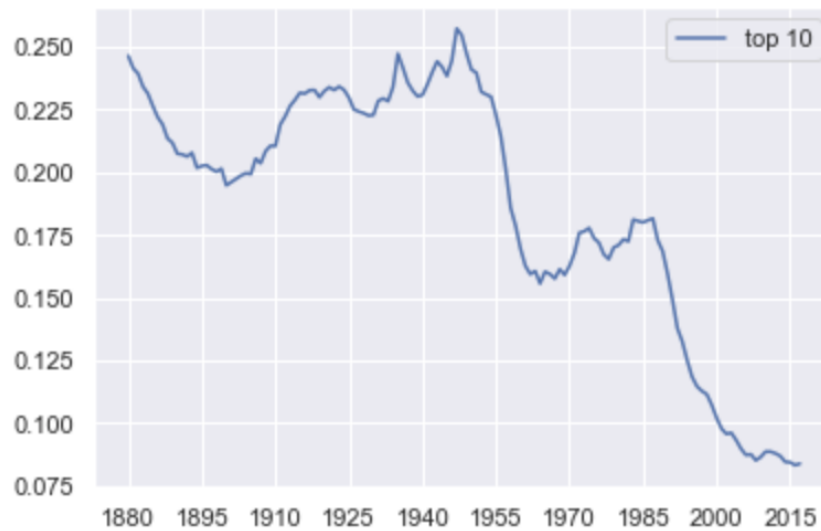
10 Most Popular Names as a Percentage of Total

## References

Pandas documentation https://pandas.pydata.org and Matplotlib documentation https://matplotlib.org