

Age Prediction from Naming Trends: A Demographic Study

Akshat Porwal¹

¹Department of Statistics, University of Michigan

January 24, 2024

Abstract

This report explores the relationship between first names, decades, sex, and the predictability of an individual's age within the United States. Leveraging a comprehensive dataset of U.S. baby names from the Social Security Administration, alongside life expectancy data, this study employs statistical models to analyze naming patterns across different decades and genders. The analysis focuses on identifying names that significantly indicate age, examining the influence of celebrities on naming conventions, and evaluating the predictability of age by name over time. Ethical considerations and potential biases in predictive modeling are discussed to underscore the implications of this research. Findings reveal that certain names, influenced by cultural and societal trends, can serve as reliable indicators of an individual's age cohort. This study contributes to the broader understanding of demographic trends through the lens of naming practices, offering insights into the intersection of sociolinguistics and data science.

1 Introduction

This study interrogates the extent to which temporal and presidential influences permeate naming conventions and their subsequent utility in deducing demographic age. Utilizing a composite dataset amalgamating United States baby names, life expectancy figures, and presidential tenures, the investigation quantifies the correlation between nomenclature frequency and historical epochs. It further employs U.S. census data to extract age distribution probabilities associated with gender-specific names, emphasizing those exceeding a threshold of fifteen thousand occurrences. Statistical measures, includ-

ing percentile ranks and interquartile ranges, facilitate the stratification of names by informativeness regarding age. This research delineates the predictive capacity of names as chronological indicators and elucidates the cultural imprints left by presidential incumbencies on naming trends. The implications of such findings extend to the domains of sociological research and demographic analytics.

2 Methodology

The initial phase of our methodology hinges on the meticulous selection and acquisition of datasets, setting the stage for a detailed examination of naming trends and their correlation with age predictability.

2.1 Data Acquisition

The analysis commenced with the acquisition of a comprehensive dataset from the U.S. Social Security Administration, detailing names, birth years, and gender-specific frequencies from 1880 to 2017. Additionally, life expectancy at birth data from the World Bank for the same period were procured to estimate the living population in 2017. Presidential term data were sourced from historical records to assess the influence of presidents on naming trends. U.S. Census data provided age distribution probabilities by gender, which were instrumental in understanding demographic name dispersion.

2.2 Data Processing

The raw data underwent rigorous preprocessing to ensure accuracy and relevance. A 'decade' column was calculated from birth years to facilitate trend analysis over time. The life expectancy figures were employed to deduce the

estimated death year, with an assumption of uniform mortality within each birth year cohort. Names shared by both genders were filtered to include only those with a total occurrence exceeding fifteen thousand to maintain statistical significance. Age percentiles were computed using census data to assess the informativeness of names concerning age.

2.3 Predictive Approach

The study adopted a probabilistic model that combined the historical prevalence of names with current age distributions. The model identified peaks in name counts corresponding to presidential terms, revealing culturally influenced naming spikes. Age percentiles and interquartile ranges derived from census data were used to rank names by their predictive informativeness. The upper quartile age identified the 'oldest' names, while a narrow interquartile range indicated names most indicative of an individual's age. Contrarily, a wide interquartile range pointed to names with the least predictive power.

3 Analysis

The analysis revealed distinct patterns in naming trends influenced by historical periods and gender. The examination of the name 'Elizabeth' demonstrated a cyclical pattern of popularity, with peaks corresponding to specific decades. The most substantial peak in the 1940s suggests a median age of 34 years in 2017, supported by a high frequency of births during that period.

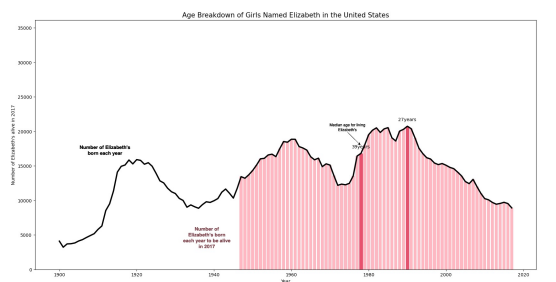


Figure 1: Age Breakdown for Elizabeth

Conversely, the analysis of the name 'Joseph' identified a pronounced peak around the 1960s, indicating a median age of 61 years. These patterns were visually represented in a combined bar and line graph, delineating the counts of individuals born with these names against the backdrop of their survival status in 2017.

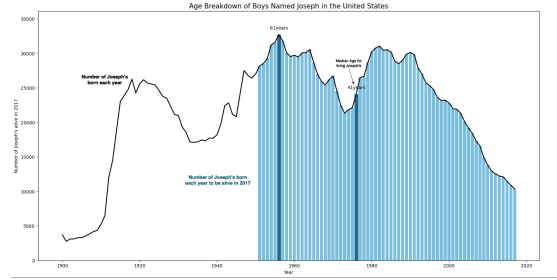


Figure 2: Age Breakdown for Joseph

Presidential influences were markedly evident in the naming trends for 'Woodrow' and 'Lyndon'. During Woodrow Wilson's presidency, there was a significant spike in the name 'Woodrow,' while 'Lyndon' saw a similar surge during Lyndon B. Johnson's term.

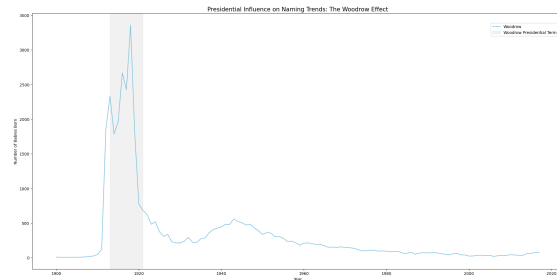


Figure 3: Presidential Influence on Naming Trends: The Woodrow Effect

These correlations were depicted through line graphs overlaid with shaded areas representing the presidential terms, providing a clear visual link between political leadership and naming choices.

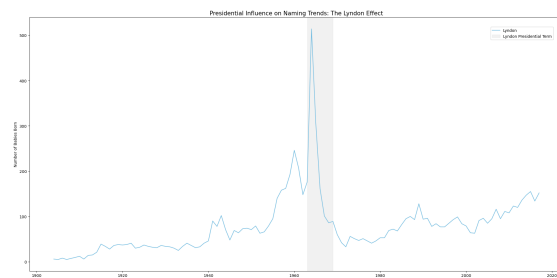


Figure 4: Presidential Influence on Naming Trends: The Lyndon Effect

The use of age percentiles unveiled names like 'Harper' and 'Aria' with a notably narrow interquartile range (IQR), suggesting these names are highly informative for predicting age due to their concentrated use in recent times.

Table 1: Most Informative Names about ages

Name	Gender	Inter-Quartile Range
Harper	F	3
Jase	M	3
Bentley	M	3
Aria	F	3
Adalynn	F	3
Paisley	F	4
Zayden	M	4
Isla	F	4
Sheena	F	4
Brantley	M	4

In contrast, names such as 'Olive' and 'Hazel' showed a wide IQR, reflecting a broader age distribution and thus less predictability.

Table 2: Least Informative Names about ages

Name	Gender	Inter-Quartile Range
Olive	F	74
Hazel	F	74
Eloise	F	72
Eleanor	F	70
Willa	F	68
Violet	F	66
Lola	F	66
Clara	F	65
Cora	F	64
Evelyn	F	64

Tables of 'oldest' names, based on the upper quartile age, listed 'Lucile' and 'Gertrude' as the most senior, suggesting that these names are predominantly carried by older individuals. This statistical evidence, derived from the U.S. census data, underpins the notion that certain names can indeed serve as chronological markers.

Table 3: Oldest Names

Name	Gender	Upper Quartile Age
Lucile	F	89
Gertrude	F	89
Imogene	F	88
Muriel	F	87
Opal	F	87

Conclusion

This analysis substantiates the premise that nomenclatural frequencies are not only reflective of societal inclinations but also responsive to temporal leadership figures, as evidenced by the nomenclature spikes during the presidencies of Woodrow Wilson and Lyndon B. Johnson. The

quantification of name-age distributions, characterized by measures such as interquartile ranges, yielded a spectrum of predictability with respect to age, thereby underscoring the potential for names to serve as demographic indicators.

The study acknowledges the constraints imposed by its methodology, particularly the uniform mortality assumption and the exclusive reliance on American datasets, which may not account for variances in global naming conventions or longevity. Such factors delineate the scope and applicability of the findings and suggest a cautious interpretation. The implications of this research invite further inquiry into the broader applicability of its conclusions, advocating for the integration of diverse datasets to enhance the robustness and cultural representativeness of predictive nomenclatural models.

References

- [1] <https://www.ssa.gov/oact/babynames/>
- [2] https://data.cityofnewyork.us/Health/Popular-Baby-Names/25th-nujf/about_data
- [3] <https://fivethirtyeight.com/features/how-to-tell-someones-age-when-all-you-know-is-her-name/#fn-6>
- [4] <https://kiwidamien.github.io/name-to-age.html>