CSC 466 Lab 1 Report

Jack de la Motte (jdelamot@calpoly.edu) & Luke Watts (luwatts@calpoly.edu)

Abstract

In this report, we look at historical baby name data in the United States to try and identify social, cultural, or demographic shifts that can be identified in the ways that newborn babies were named. Although there are countless trends to try and identify, we focused our efforts on three. Firstly, how Irish Immigration loosely correlated to the use of the name Patrick in New England in the first half of the twentieth century. Secondly, how the widely influential film, "Breakfast at Tiffany's" which starred Audrey Hepburn, caused a massive increase in the use of the name Tiffany. Lastly, we analyzed how the presidency and assasination of former president John F. Kennedy affected the number of boys named John and girls named Kennedy.

Introduction

In this report we are given a dataset of baby names in the United States and attempt to identify three important sociological trends that are reflected in our data. By looking at individual names and how they increase and decrease in popularity throughout the time frame provided by our dataset, we attempt to find evidence of these trends impacting name popularity as well as uncover trends and events we had not expected to have impact by studying the data.

Dataset Description

For our assignment we were given two data files to work with. The first contains United States baby name data at a national level (no delineation between states) and spans from 1880 to 2014. In addition to each baby name we are given a unique identifier, the year, the gender, and count. The second file contains baby name data by state with the same fields otherwise; this data only spans from 1910 to 2014.

Research Questions

Before making a single query, we suspected that reflected in the baby name records were stories about American attitudes, preferences, prejudices, people groups, pop-culture trends, political figures, and more. After spending time researching and pondering some of the most influential cultural and sociopolitical demographic trends of the twentieth century, we landed on three questions for our dataset that we thought could potentially uncover some of these societal changes.

Firstly, we wondered about immigration. America has a rich history of bringing in foreigners and incorporating and assimilating them into the culture, which is best understood as a "melting pot" of innumerable heritages and cultures giving and taking to form something new and wonderfully diverse. In order to successfully identify a wave of immigration reflected in baby name data while also remaining in the scope of this assignment, we decided to concentrate our search down to a single population, the Irish.

We know that due to harsh farming conditions in conjunction with sociopolitical and religious factors, the Irish migrated to the United States in large quantities. Some estimates suggest that between 1841 and World War II, some 4.5 million Irish men, women, and children traveled to the United States to start new lives. However, this left us with a challenge: how were we going to identify a massive cultural shift caused by this immigration in… baby names? Of course, in a perfect analysis, we would have access to each baby's ethnic heritage so we could track exactly how many Irish children were born and when and where. But, our challenge was to accomplish the same task but with a more limited dataset. We decided to approach this difficulty by narrowing our query even further to a single name: Patrick. From the Latin name *Patricius* adopted by Saint Patrick in the 5th century (who is credited with Christianizing Ireland), this name is a wildly popular and traditional Irish name for boys. Although it doesn't provide a perfect method, we hoped to find evidence of the Irish Immigraiton to the United States through the tracking the use of the name Patrick. We decided to add on another layer of specificity by limiting the search to the seven states that make up New England: Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, and New York, since most late 19th century, early 20th century Irish immigrants came through Ellis Island, landing them in New York, and then spreading out to neighboring states. Lastly, since our state name data only extends back as far as 1910, we limited our scope to 1910 through 1950. This brings us to our first research question:

1. **"Did the use of the name 'Patrick' for baby boys increase in New England between 1910 and 1930?"**

For our second inquiry, we wanted to explore a different kind of baby name trend: one influenced by popular culture. Pop-culture is a complex phenomenon but some of it's biggest contributors are undoubtedly the musicians, authors, artists, producers, directors, actors, and actresses of a given period in time. Although the internet has made this more complicated by giving a platform to the masses, these societal characters were especially crucial in defining the culture in the mid twentieth century.

Audrey Hepburn was one of the most influential actresses of the golden age of Hollywood who starred in numerous critically acclaimed films, one of which was "Breakfast at Tiffany's", released in 1961, which not only was met by numerous film awards but was recognized by the U.S. Library of congress as "culturally, historically, and aesthetically significant". While Audrey Hepburn's cultural impact in "Breakfast at Tiffany's" is undeniable, we wanted to dig a little bit deeper and see if we could correlate an increase in the usage of the name "Tiffany" to the popularity that followed the release of the film. To capture a helpful context, we decided to look at the usage of the name Tiffany for female babies between the years 1950 and 2000, to try and understand how popular the name was before the release of the movie, and how it's popularity is affected around the time of release as well as the cultural aftermath. Thus, we have our second question:

2. **"Did the use of the name 'Tiffany' increase between 1950 and 2000 due to the film 'Breakfast at Tiffany's'?"**
With our final inquiry, we chose to focus on the topic of politics and how politicians may influence the naming of newborns by looking at data related to a single term president who retained generally high approval ratings. For this, we decided to select John F. Kennedy as his short presidency during an eventful period of American history left a mark on the nation that would not be forgotten. In addition to his legacy, John F. Kennedy's name would provide two possible sources of influence on newborn names: firstly with the effect of JFK's presidency on male newborns being named John, and secondly with its effect on female newborns being named Kennedy.

John F. Kennedy is often regarded as one of the greatest presidents of modern American history with high approval ratings from retrospective polls. A 2013 poll from Gallup concluded with its findings that seventy four percent of Americans believed that President Kennedy was an outstanding or above average president and that only three percent disapproved of Kennedy's presidency. During his time as president, John F. Kennedy maintained a record high lowest point of approval rating of any president from 1937 to modern day with his approval rating never dipping below fifty-six percent. With high respect both during and after his term in office combined with his tragic assanation in 1963, it would be reasonable to think that the legacy of JFK would have a large impact on the naming of newborns in the United States. This brings us to our third and final question:

3. **"Did former President John F. Kennedy's presidency affect the popularity of the name John for baby boys and Kennedy for baby girls after his victory in the 1960 presidential election and his assination in 1963?"**

Methods

For each of our queries we utilized Python along with its Pandas, NumPy, and Matplotlib packages for filtering, organizing, and plotting our data.

For our first question, all we needed to do was create a dataframe of all the male instances of the name Patrick from 1910 to 1950 and use pandas to filter out any rows that were not from one of our seven New England states, and then sum the number of occurrences by year from each state. We Then used Matplotlib to plot the year against the count.
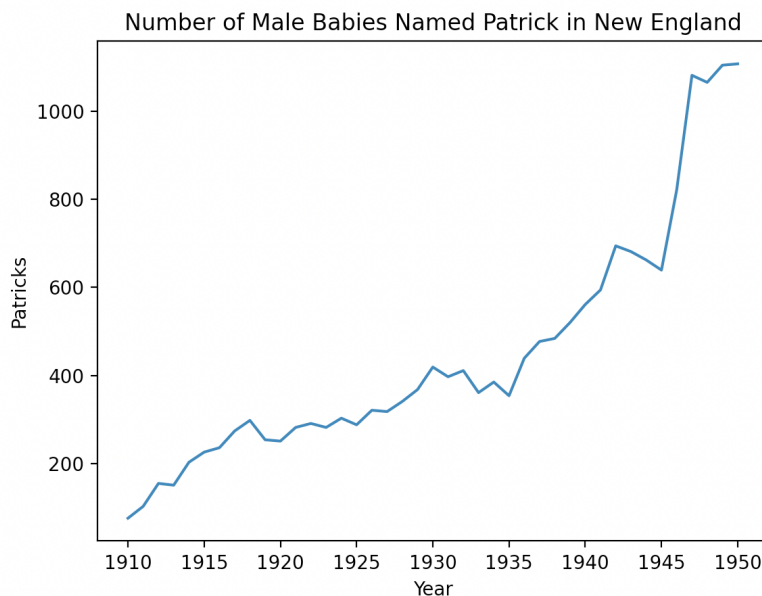
For our second question, we just created a dataframe for all the female instances of the name Tiffany from 1950 to 1990 and plotted the year against the count.

Lastly, for our third question, we created two dataframes for the female name Kennedy, one for the national dataset and one for the state dataset filtered down to states that John F. Kennedy won the majority of electoral college votes for in the 1960 election. In addition, we created one dataframe for the name John and one dataframe for the name Jon to compare whether events in the presidency of President Kennedy influenced the popularity of all spellings of the name

John/Jon, just the gospel spelling as the president has, or none of the variations of the name at all. Graphs with these dataframes would be created with Matplotlib with the year against the count of the names' appearance.
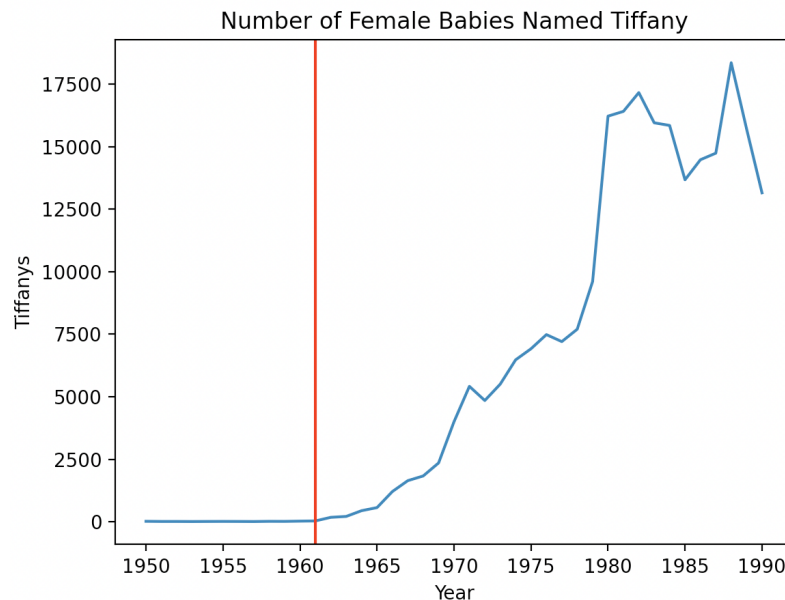
Results

The results of our first query are documented in the following chart:

As this data demonstrates, the usage of the name Patrick appears to increase steadily until 1946 when it sees a significant increase. The spike at 1946 is very likely to be correlated to the start of the post World War II 'baby boom' so it is not particularly relevant to the question we are hoping to answer since many names would have seen a large increase as well around that time. What is noteworthy is the increase in the number of babies given the name Patrick in New England between 1910 and 1945. Although, again, the measurement of a single name is an inadequate way to properly track the immigration of an entire population, it is not altogether inadequate. A hypothesis we would be interested in studying in further research would be if other 'Irish' names follow a similar trend to Patrick at this time. Another limitation on this analysis is the fact that much of the Irish immigration to America took place before 1910. However, 1910 is the year that our state data starts so we have to work with what we can. Furthermore, research that we did using outside sources showed that in fact, Irish immigration to the United States was actually on the decline by the 1940's and 50's so we hypothesize that perhaps the

continued increase in number of Patricks by that time is attributable to first generation Irish immigrants naming their baby boys Patrick.

The results of our second query are documented in the following chart:
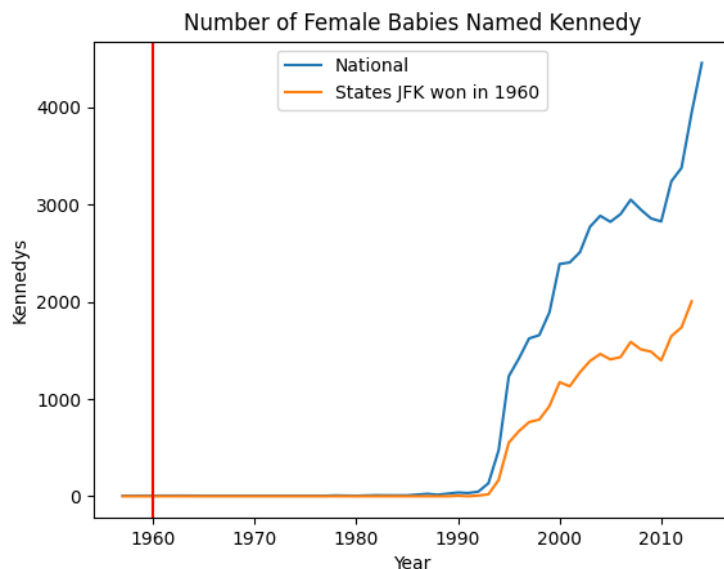

Number of Female Babies Named Tiffany

As is especially evident in this plot, the name Tiffany, for reasons that will be discussed, exploded in popularity in the latter half of the twentieth century. There are a few curious facts about this data, some of which have obvious answers and others that we haven't been able to fully explain yet. What is important to note first is that the name Tiffany was relatively unheard of in the United States until 1961 when "Breakfast at Tiffany's was released", the data strongly suggests therefore that the release of the film had *significant* impact on the number of usages of the name Tiffany for babies born thereafter. Two idiosyncrasies of this data that were not immediately explainable for us were the two spikes: the first being the most dramatic in 1979 and the second in 1987. Although we can't prove exactly why these exist, and a number of factors could have influenced them, we do have two compelling events that could shed some light on these strange increases.

Firstly, 1979, the year of the first spike, is also the year the Breakfast at Tiffany's was released on VHS. What we have to keep in mind about movies in this period of time is that there was no Netflix or Amazon Prime Video to make movies completely accessible in most homes in America. However, when movies started coming out on VHS, and especially when Audrey

Hepburn's first movie was released on VHS (Breakfast at Tiffany's), these films once only accessible in theaters were now playing and replaying on every family television in America.

As for the second, lesser spike, there are two factors that could have been at play. One line of thinking we followed was that the teenage girls who had fallen in love with "Breakfast at Tiffany's" when it came out on VHS in 1979 were now grown up and having children of their own. Alternatively, 1987 is also the year of the smash hit pop-single "I Think We're Alone Now", sung by none other than the widely influential singer, "Tiffany". Whether one of these had more effect on the second spike than the other or the conjunction of the two of them is what created the spike at all is unclear but those are our best attempts at explaining the causes of the two increases.
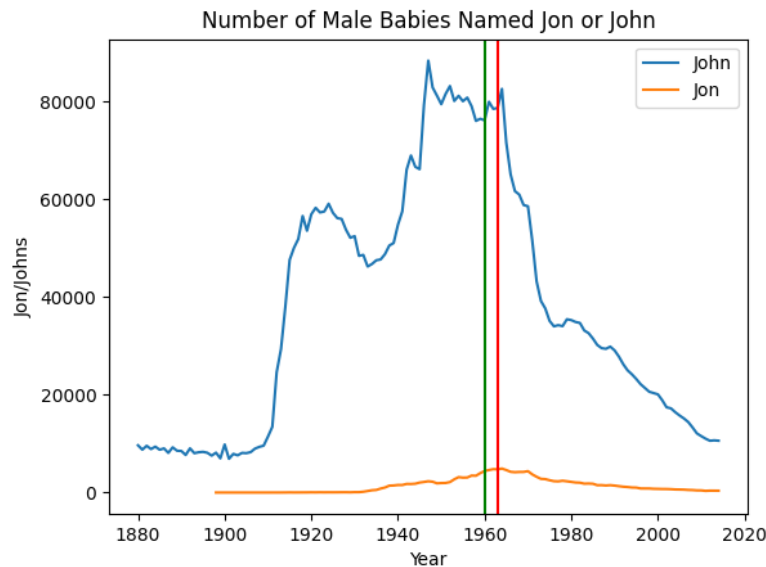
The results for our third query are documented in the two following charts:

Number of Female Babies Named Kennedy

* Note for black and white printing: the upper trendline correlates to the National data and the lower to the states JFK won in 1960.

With this first chart, it can be seen that the name Kennedy for females did not exist in the databases prior to 1957 and was not a popular name in the United States until around the mid 1990's. During this time, the name had an explosive growth with a second spike occurring within the 2010's. The trends between the blue and yellow lines of national data vs state specific data

from states that Kennedy won a majority of electoral college votes from in 1960 shows that states that favored Kennedy at the time do not variate heavily from the national perception of the name. The lack of female Kennedys for thirty years after his death indicates that President Kennedy may not be the main influence for the growing popularity of the name. However, the rapid growth is in line with the year of the death of President Kennedy's wife, Jacqueline Kennedy Onassis, who remarried after the president died and passed away in 1994.

This second chart shows the popularity of the names John and Jon over time while highlighting the years of President Kennedy's victory in 1960 and death in 1963. As seen from the visual, the name John has long been more popular than its variant spelling which omits the letter H. In fact, the name Jon never reached more than a few thousand at its peak and did not record five or more births with the name prior to 1998. The name John, on the other hand, experienced explosive growth from the 1910's to the 1920's and once again from the 1930's to the 1940's, having over 80,000 newborns with the name at its peak. Just before 1960, the name was on a slow decline, although still very popular, until Kennedy's election to the position of Commander in Chief. Just after that year, a strong bounce in the data can be seen and this occurred once again in 1964, the year after the president was assassinated. After 1964, there is a harsh decline that persists with few exceptions until the most recent data in 2014 where the name rests at nearly the same amount of newborns with the name as before the 20th century despite the nation's total population growing by over two hundred million during the one hundred years that passed. Given the timing of the decline, it is likely that the sorrow surrounding the late president's death contributed to the decline, although other factors may have been involved in

this shift in popularity. A comparison of other gospel names over time may highlight whether a greater trend in a move away from biblical names was occurring in America during this period or if the assassination of John F. Kennedy led to the slow death of the name John all together. The decline in John may also have contributed to an increase in the many variant spellings of the name such as Jonathan, Johnny, etc. Further investigation into the data sets would be required to find if these reasons held greater influence than the legacy of former President Kennedy in the popularity and spelling of John related names.

Discussion and Conclusion

In conclusion, our hypotheses both did and did not get supported by evidence in the ways we had predicted. For our first query of the name Patrick, we were challenged by the limited time scope of our dataset; we would have liked to look further back in time, but based on the time period we did look at, it does appear that the name Patrick was steadily increasing in popularity. A primary limitation of our method in this case was that the correlation between the usage of the name Patrick could ultimately be loosely correlated with Irish immigration patterns. For this reason our method could have been improved if we had individual ethnicity data. Furthermore, for our last question, a difficulty faced was understanding why the popularity of the name Kennedy for females exploded years after JFK's death. We were able to give our best speculation as to why this occurred but with a limited dataset and the limitations of our own historical knowledge, we can't say with certainty what caused the phenomenon or if the source was a singular point of influence.

Bibliography

https://iowaculture.gov/history/education/educator-resources/primary-source-sets/irish-immigration-beyond-potato-famine

https://www.behindthename.com/name/patrick

https://en.wikipedia.org/wiki/Audrey_Hepburn

https://en.wikipedia.org/wiki/Breakfast_at_Tiffany%27s_%28film%29

https://news.gallup.com/poll/165902/americans-rate-jfk-top-modern-president.aspx

https://www.presidency.ucsb.edu/statistics/data/presidential-job-approval

https://www.cantonrep.com/article/20140112/NEWS/140119826