Project 1: What is in a name?

Author: Gabriel Sanahuano

Date: 1/23/2021

**Professor: J. Hathaway**

Elevator pitch

This project consists of using Data Science to generate charts with statistical data about birth name counts by year for each US state. This method will mainly be used to compare the data, and in which years there was a notable increase in a specific name.

TECHNICAL DETAILS

- **GRAND QUESTION #1**
  How does your name at your birth year compare to its use historically?

As you can see in the graph below, the name "Gabriel" was not popular between the years 1910's and 1960's, not exceeding 1000 names per year. From the year 1970 to 1990, it began to have a popular trend almost bordering 3,000 names every year. It caught my attention, which is close to 1993, my year of birth, which has an impressive boom in the number of names per year. From 1990 to 2010, it has a non-stop increase from 2000 names per year to around 12,000. The year 1990 was decisive in the growth of the curve. My year of birth is marked by this boom of increase from the year 1990.
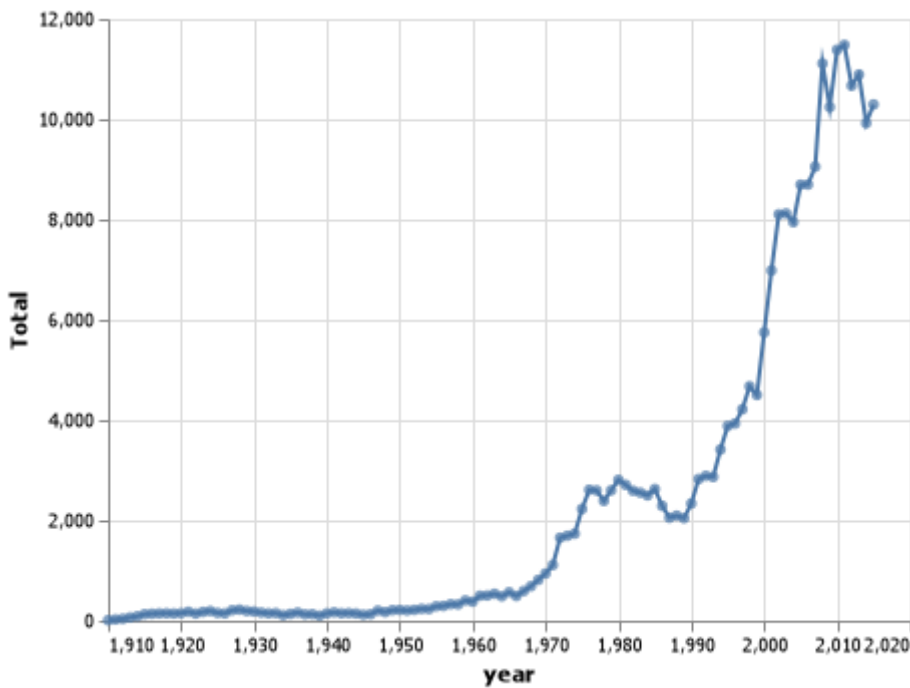
- **GRAND QUESTION #2**
  If you talked to someone named Brittany on the phone, what is your guess of their age?
 **What ages would you not guess?**

From the graph below, you would guess that she is between 35 and 25 years old. If I had to guess a specific age, I would say that it is 30 years old exactly. I would not say that she is between 40 and 55 years old. Nor would I say that she is under the age of 17. It's clear from the chart that there was a boom in people with the name "Brittany" during the years around 1985 and 1995.
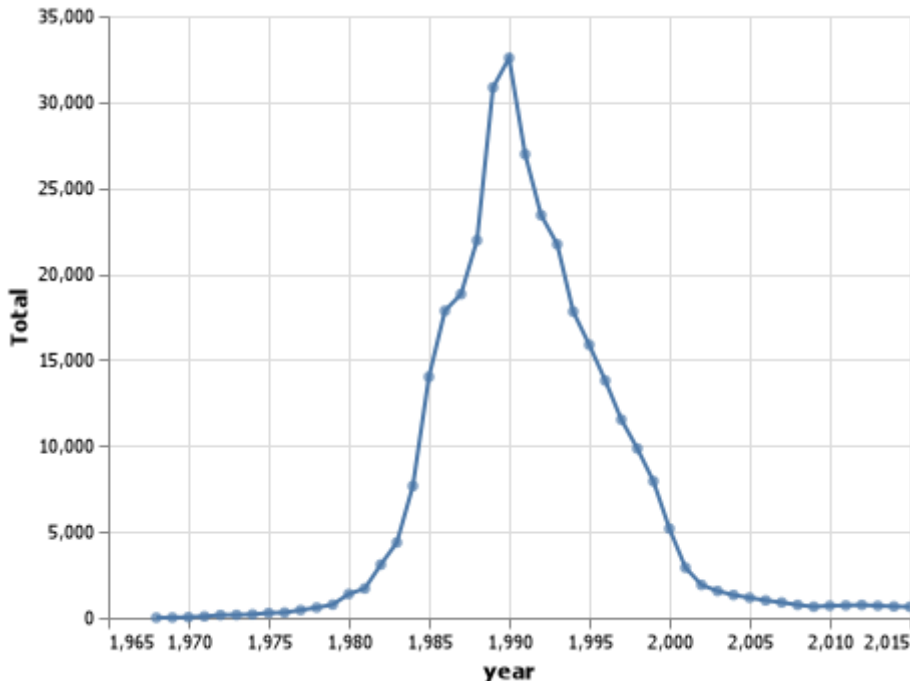
- **GRAND QUESTION #3**

  Mary, Martha, Peter, and Paul are all Christian names. From 1920 - 2000, compare the name usage of each of the four names.

Between 1920 and 1970, Maria was the most prominent name in use, and the second was Paul. Peter and Martha had very similar numbers, not exceeding 10,000 names per year. According to the graph, we can see that since 1970, the popularity of the four names was on the decline and they had very close numbers. Based on this information, it can be deduced that, by the year 2000, the four names were not popular at all.
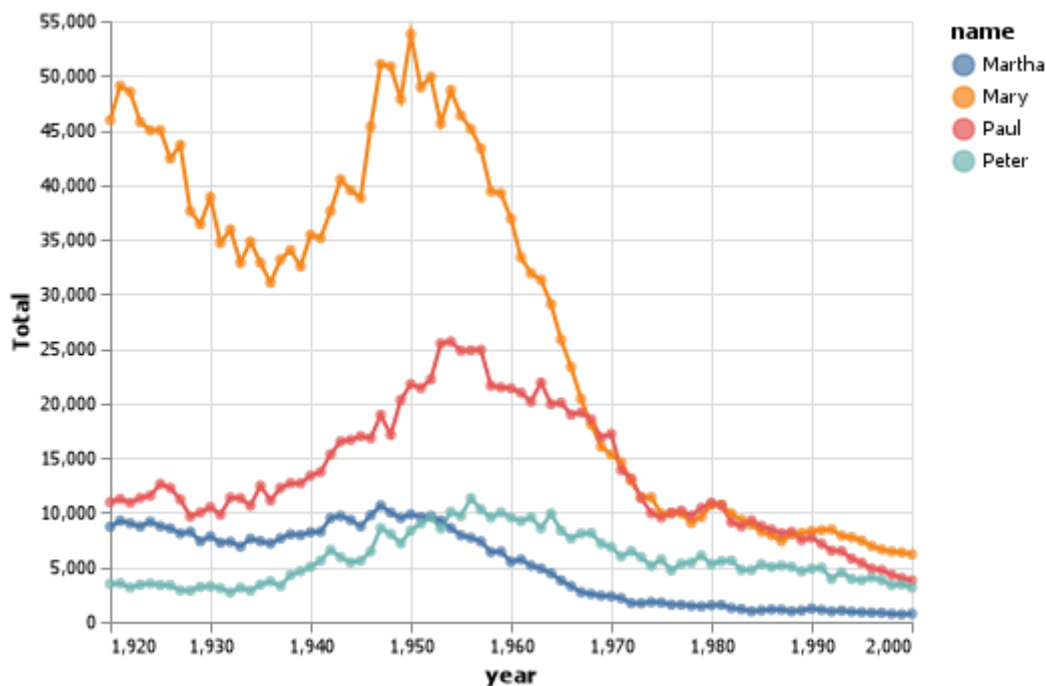
- **GRAND QUESTION #4**

  Think of a unique name from a famous movie. Plot that name and see how increases line up with the movie release.

The movie I chose was "Harry Potter and the Sorcerer's stone." It was released on November 14, 2001, and was a hit at the time. The most famous name of that movie was "Harry." But unfortunately, the film did not affect the number of names "Harry" from 2001 to 2015. No increase was seen, and more along the line, it seems that it has continued in a slow decline. But it does not stop being a great movie.
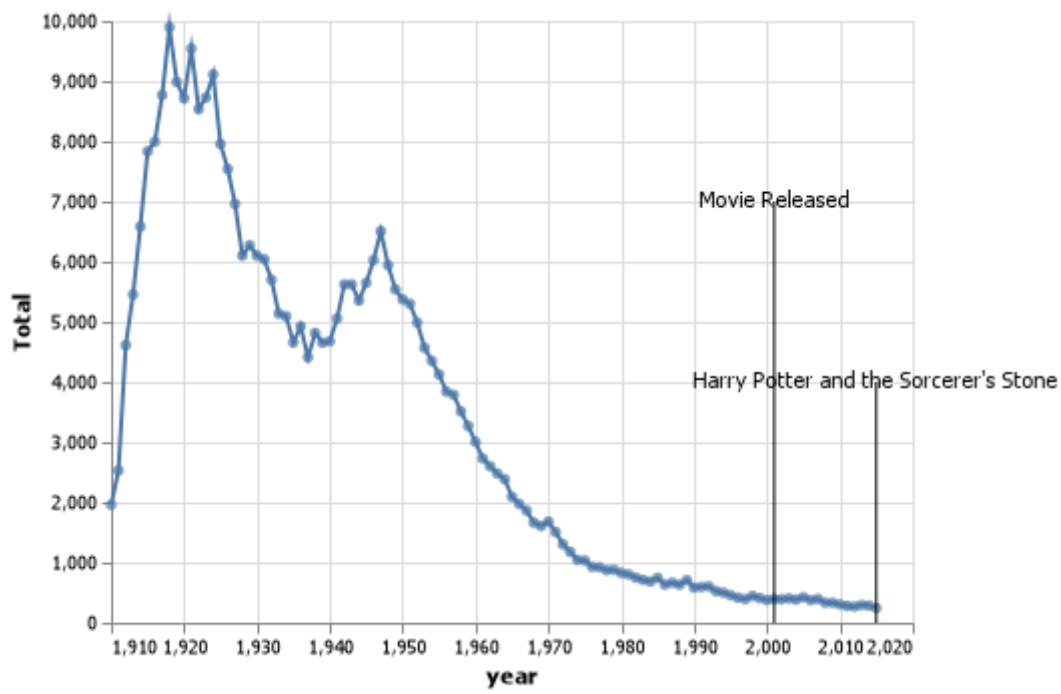
PYTHON SCRIPT

```python
#%%
import pandas as pd
import altair as alt

#%%
namesData = pd.read_csv('https://raw.githubusercontent.com/byuidatascience/data4names/master/data-raw/names_year/

# %%
gabrielNames = namesData[namesData['name'] == 'Gabriel']

chart = (alt.Chart(gabrielNames).encode(x = 'year', y = 'Total'))


graphic1 = chart.mark_line() + chart.mark_circle()

graphic1

# %%
graphic1.save("name_Gabriel.png")
# %%

brittanyNames = namesData[namesData['name'] == 'Brittany']
chart2 = (alt.Chart(brittanyNames).encode(x = 'year', y = 'Total'))


graphic2 = chart2.mark_line() + chart2.mark_circle()

graphic2

# %%
graphic2.save('name_Brittany.png')
# %%
start = 1920
end = 2000
startDate = namesData["year"] >= start
endDate = namesData["year"] <= end
between1920And2000 = startDate & endDate
In1920_2000 = namesData.loc[between1920And2000]

namesFiltered = (In1920_2000.query('name in ["Mary", "Martha", "Peter", "Paul"]').filter(items = ['name', 'year',

chart3 = (alt.Chart(namesFiltered).encode(x = alt.X('year'), y = alt.Y('Total'), color = 'name'))

graphic3 = chart3.mark_line() + chart3.mark_circle()

graphic3


# %%
graphic3.save('comparison.png')
# %%
```

```python
harryNames = namesData[namesData['name'] == 'Harry']

chart4 = (alt.Chart(harryNames).encode(x = 'year', y = 'Total'))

labels = pd.DataFrame({
    'year': [2001, 2015],
    'Total': [7000, 4000],
    'label': ['Movie Released', 'Harry Potter and the Sorcerer\'s Stone']

})

chartWithLabels = (alt.Chart(labels).encode(x = alt.X('year'), y = alt.Y('Total'), text = 'label')).mark_text()
chartWithRules = (alt.Chart(labels).encode(x = alt.X('year'), y = alt.Y('Total'), text = 'label')).mark_rule()

graphic4 = chart4.mark_line() + chart4.mark_circle() + chartWithLabels + chartWithRules

graphic4
# %%
graphic4.save('harry.png')
```