

Project 1

Alice Pao

Project Summary

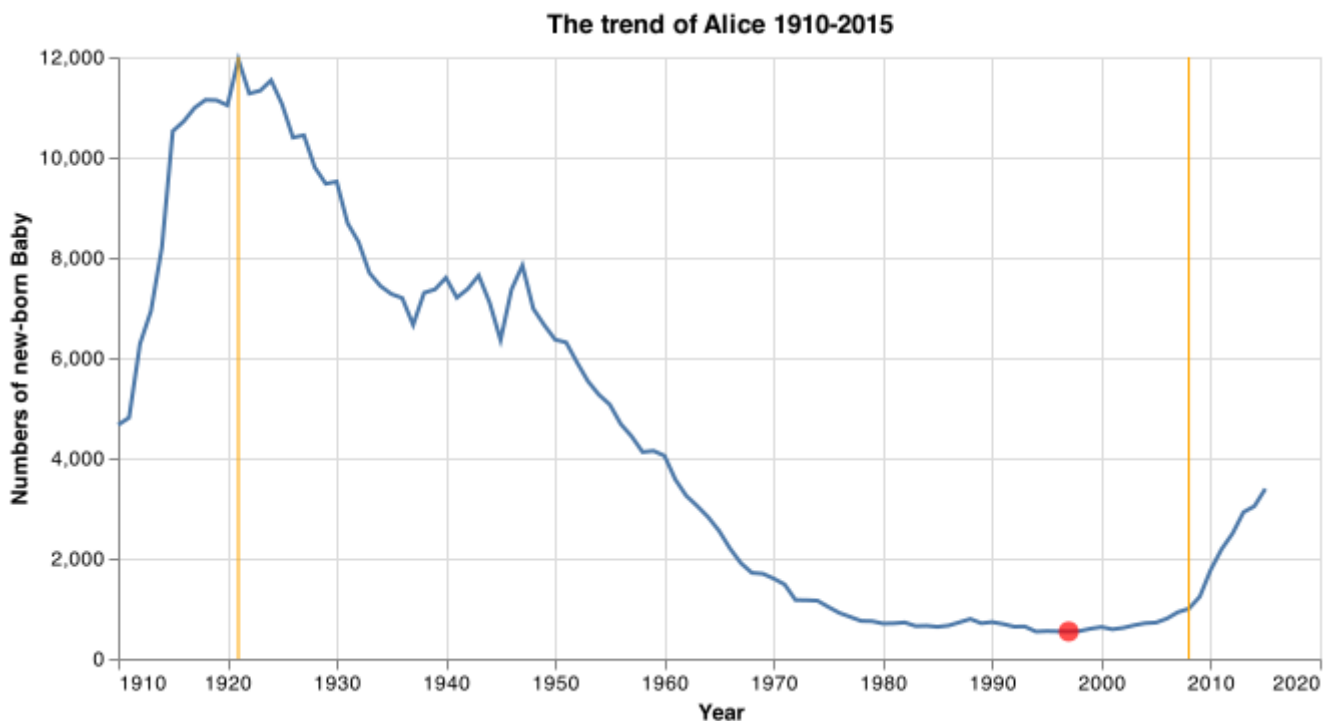
This project is to analyze name usage from an imported database. I used line chart and bar chart to support my analysis. I also created different data frame, marks, set colors to explain my points more clearly.

Technical Details

Grand Question 1

How does your name at your birth year compare to its use historically?

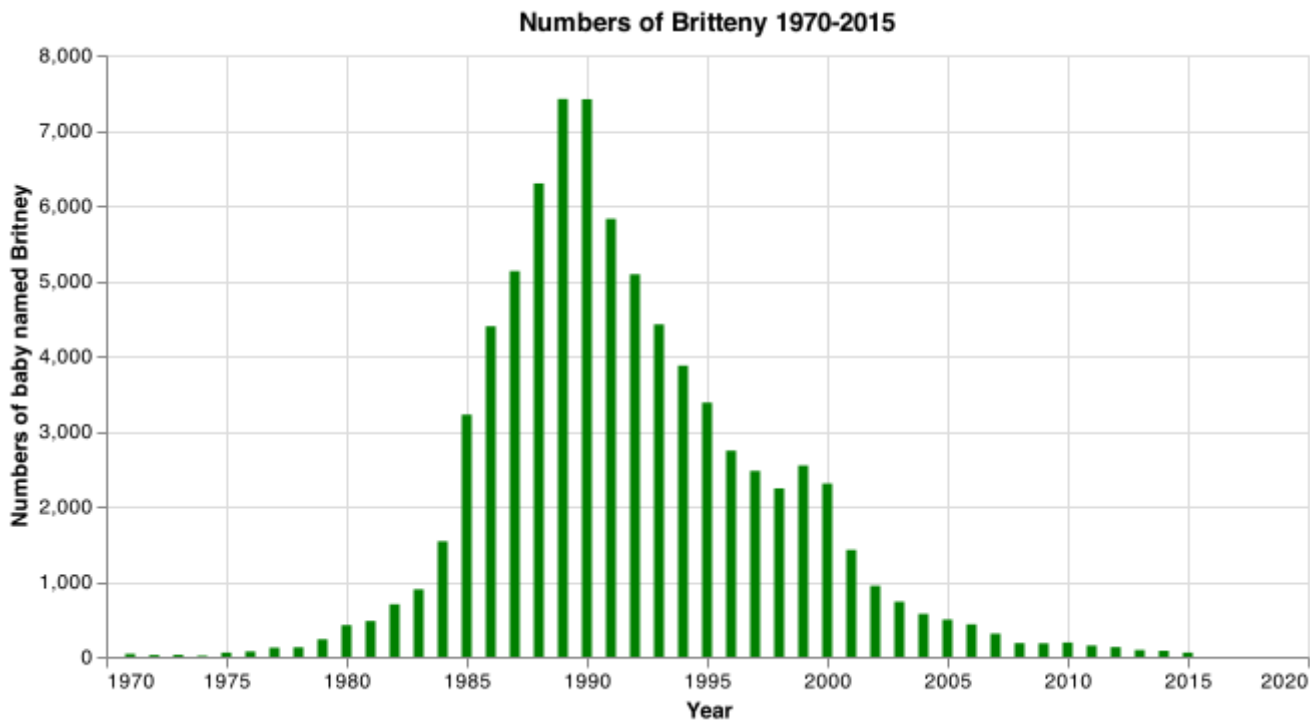
The name Alice has been around since 1910. It gained drastic popularity during 1910-1921. However, after 1921, the general trend shows a gradual decline. I was born in 1997. It is a part of the steady trend when the number stays the same for about 32 years. Around 2008, the name slowly earns back its popularity.



Grand Question 2

If you talked to someone named Brittany on the phone, what is your guess of his or her age? What ages would you not guess?

I'd guess that she is 33 or 34 years old because 1989 and 1990 have the most new born babies named Brittney. I wouldn't guess that she is 48.



Grand Question 3

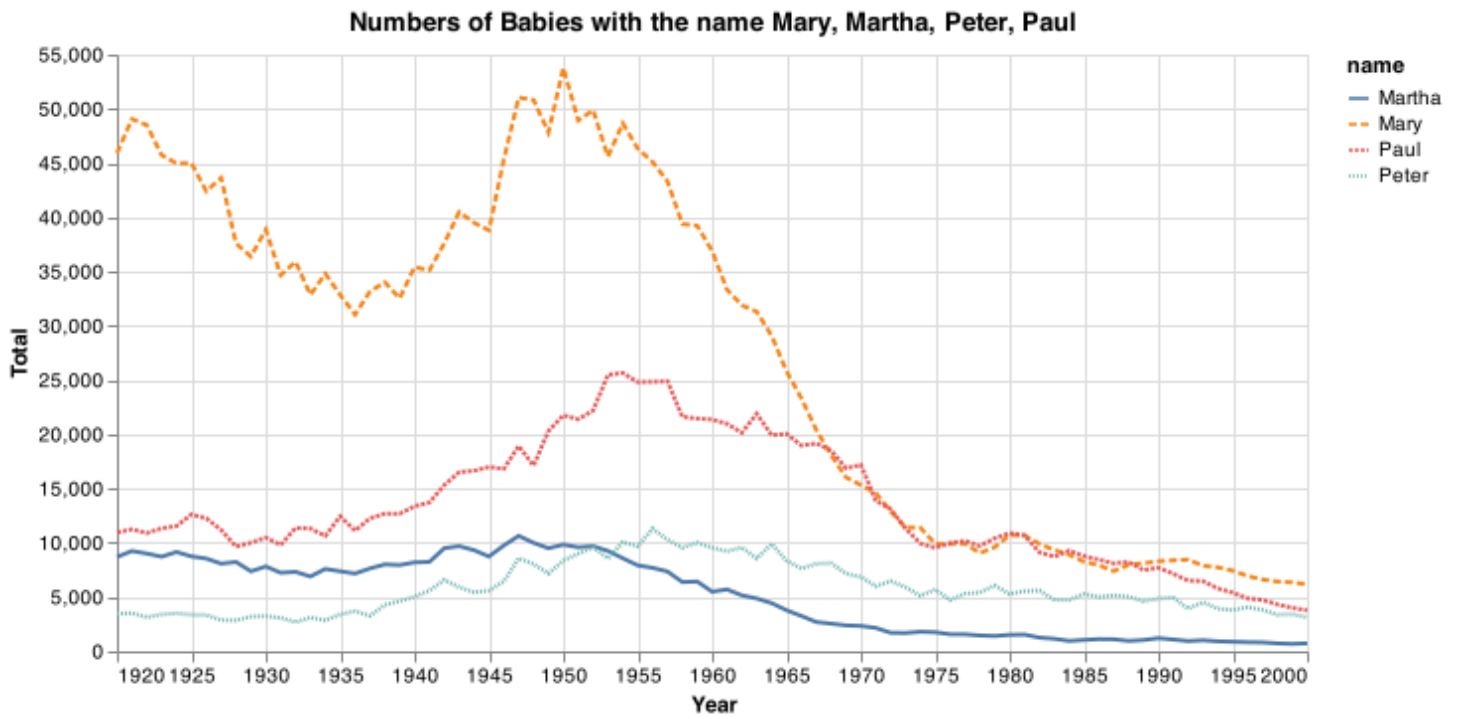
Mary, Martha, Peter, and Paul are all Christian names. From 1920 to 2000, compare the name usage of each of the four names.

Mary: This is the most liked name. 1920 to 1950 was its most popular years. After 1950, it decreased each year.

Martha: In 1920, Martha was more used than Peter; however, it became the least favorite after 80 years.

Peter: Peter started out to be the least used name among the four but it slightly received popularity and ended up becoming the second least favored name, defeating Martha by the end of 2000.

Paul: The peak of this name was around 1953 and 1954. The decrease started soon after.



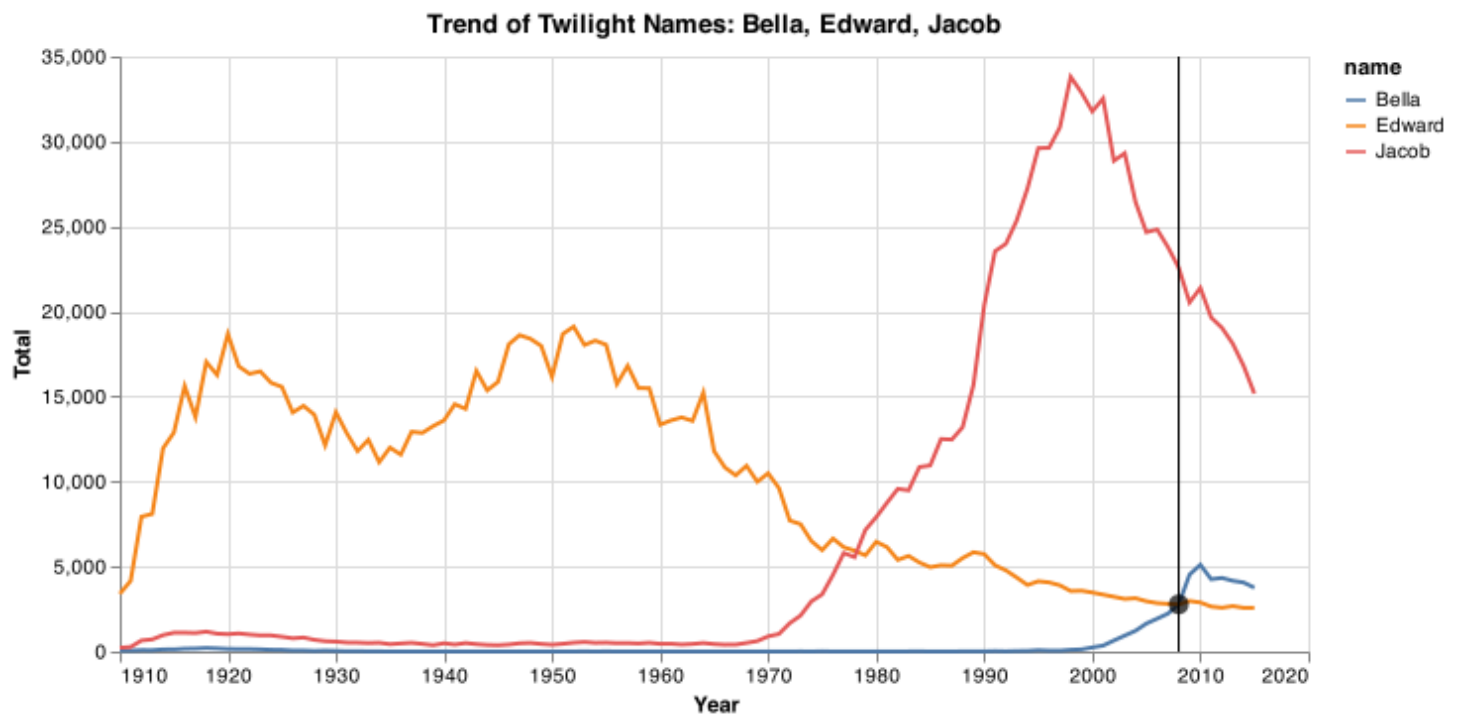
Grand Question 4

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

I picked the movie Twilight. It was released in 2008. Edward, Bella, and Jacob are the main three characters.

From the following line chart, it looks like there wasn't an increase for the name Edward and Jacob. After 2008, the number of baby Edward stayed relatively the same while the popularity of Jacob dropped dramatically.

On the other hand, the number of Bella went up from 2778 to 5000 in two years after the movie release. Another interesting thing I noticed is that Edward and Bella had it's first crossing point in 2008.



Appendix A

```

# %%
import pandas as pd
import altair as alt

# import data
names = pd.read_csv("https://github.com/byuidatascience/data4names/raw/master/data-raw/names_year/names")

# filter Alice from the list of names
question_one = names.query("name == 'Alice'")

# create a line chart to analyze how Alice has been used historically
my_line_chart = (alt.Chart(question_one, title = 'The trend of Alice 1910-2015')
    .properties(width = 600)
    .mark_line()
    .encode(
        x = alt.X('year', axis=alt.Axis(format = 'd', title = 'Year')),
        y = alt.Y('Total', axis=alt.Axis(title = 'Numbers of new-born Baby'))))

# create mark charts to highlight the year 1921, 2008, and the year I was born
my_bd_data = names.query('name == "Alice" & year == 1997')
my_dot_chart = (alt.Chart(my_bd_data)
    .encode(x = 'year', y = 'Total')
    .mark_circle(color = "red", size = 100)
    )

line_1921_df = pd.DataFrame({'year': [1921]})
line_1921_df
line = alt.Chart(line_1921_df).mark_rule(color = "orange").encode(x = "year")

line_2008_df = pd.DataFrame({'year': [2008]})
line_2008_df
line_2008 = alt.Chart(line_2008_df).mark_rule(color = "orange").encode(x = "year")

# combine line & mark charts together
chart_1 = my_line_chart + my_dot_chart + line + line_2008

# save the chart to .png
chart_1.save('question_1.png')
# %%
# create the chart
brittney = names.query("name == 'Brittney'")
chart_2 = (alt.Chart(brittney, title = 'Numbers of Britteny 1970-2015').properties(width = 600)
    # black bar chart
    .mark_bar(color = 'green')
    # name the x & y axes
    .encode(
        x = alt.X('year', axis = alt.Axis(format = 'd', title = 'Year')),
        y = alt.Y('Total', axis = alt.Axis(title = 'Numbers of baby named Britney'))
    )
)

```

```

chart_2
# save the chart to .png
chart_2.save('question_2.png')
# %%
# create a list to find multiple names: Mary, Martha, Peter, Paul
christian_names = ['Mary', 'Martha', 'Peter', 'Paul']
# create a data frame with the names and years
question3_data = names.query("name == @christian_names & year >= 1920 & year <= 2000")

# create a chart with all 3 names
chart_3 = (alt.Chart(question3_data, title = 'Numbers of Babies with the name Mary, Martha, Peter, Pau
# line chart
.mark_line()
.encode(
    x = alt.X('year', axis = alt.Axis(format = 'd', title = 'Year')),
    y = alt.Y('Total'),
    color = 'name',
    strokeDash = 'name'
)
)
chart_3
chart_3.save('question_3.png')

# %%
# create a list to find multiple names: Edward, Bella, Jacob
twilight_names = ['Edward', 'Bella', 'Jacob']
# create a data frame with the names
question4_data = names.query("name == @twilight_names")
# create a stacked bar chart
chart_4_base = alt.Chart(question4_data, title = 'Trend of Twilight Names: Bella, Edward, Jacob').prope

chart_4_line = (chart_4_base.mark_line()
    .encode(
        x = alt.X('year', axis = alt.Axis(format = 'd', title = 'Year')),
        y = 'Total',
        color = 'name'
    )
)
# create a mark line for year 2008
chart_4_line_2008_df = pd.DataFrame({'year': [2008]})
chart_4_line_2008_df
chart_4_mark_line = alt.Chart(chart_4_line_2008_df).mark_rule(color = "black").encode(x = "year")

# create a dot mark in to point out the crossing in 2008
bella_2008 = names.query('name == "Bella" & year == 2008')
bella_dot_chart = (alt.Chart(bella_2008)
    .encode(x = 'year', y = 'Total')
    .mark_circle(color = "black", size = 100)
)

chart_4 = chart_4_line + chart_4_mark_line + bella_dot_chart

```

```
chart_4
chart_4.save('question_4.png')
```