Client Report - What's in a name?

See code

Course DS 250

AUTHOR
Bracken Sant

▼ Show the code

```
import pandas as pd
import altair as alt
from altair_saver import save
from tabulate import tabulate
```

Elevator pitch

I used a name pool from 1920-2015 to look at the distributions of given names using pandas and altair. There are many different factors that go into the name pool and how much a given name is used in a certain year. One of the factors that I looked at was a popular movie franchise, Star Wars, and what it did to change the usage of names like Leia and Luke.

▼ Read and format project data

```
data = pd.read_csv('names_year.csv')
```

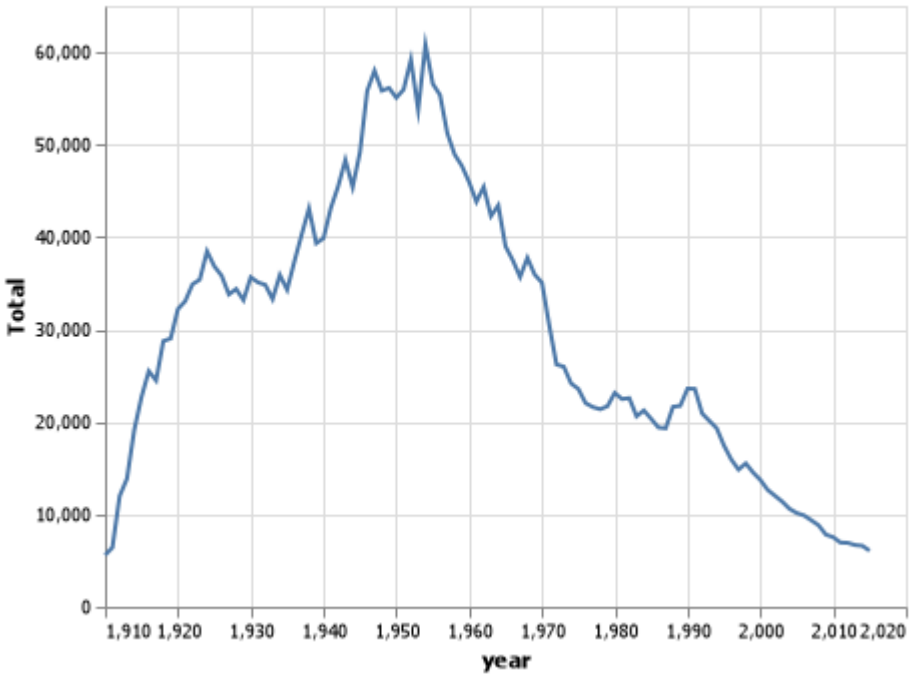
GRAND QUESTION 1

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

According to the chart my name, Robert, steadily became more popular until in the 1950's it hit a local maximum and then slowly declined until present date.

▼ Read and format data

```
myname_data = data.query("name == 'Robert'")
chart = alt.Chart(myname_data).encode(x = "year", y = "Total").mark_line()
save(chart, "mynamedata.png")
```



Robert Data

As you can see in the chart the name Robert has a local maximum in the 1950's and then slowly declines until 2015

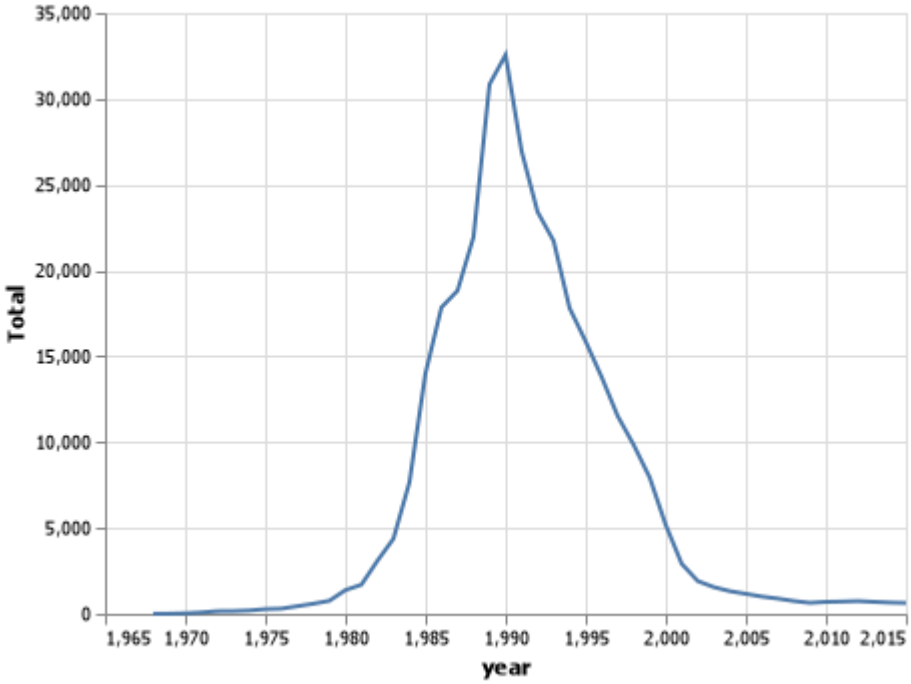
GRAND QUESTION 2

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?

The chart is almost a bell curve so about 68% of the people named Brittany will be between 24 and 38 years of age with the biggest spike being around 1990 meaning the most likely age would be 32. So I would not guess 0-18 and older than 40 years of age.

▼ Read and format data

```
brittany_data = data.query("name == 'Brittany'")
chart2 = alt.Chart(brittany_data).encode(x = "year", y = "Total").mark_line()
save(chart2, "brittany_data.png")
```



Brittany Data

As you can see on the graph it represents something like a bell curve so using standard deviations you can assert that about 68% of the people named Brittany will be between 24-38.

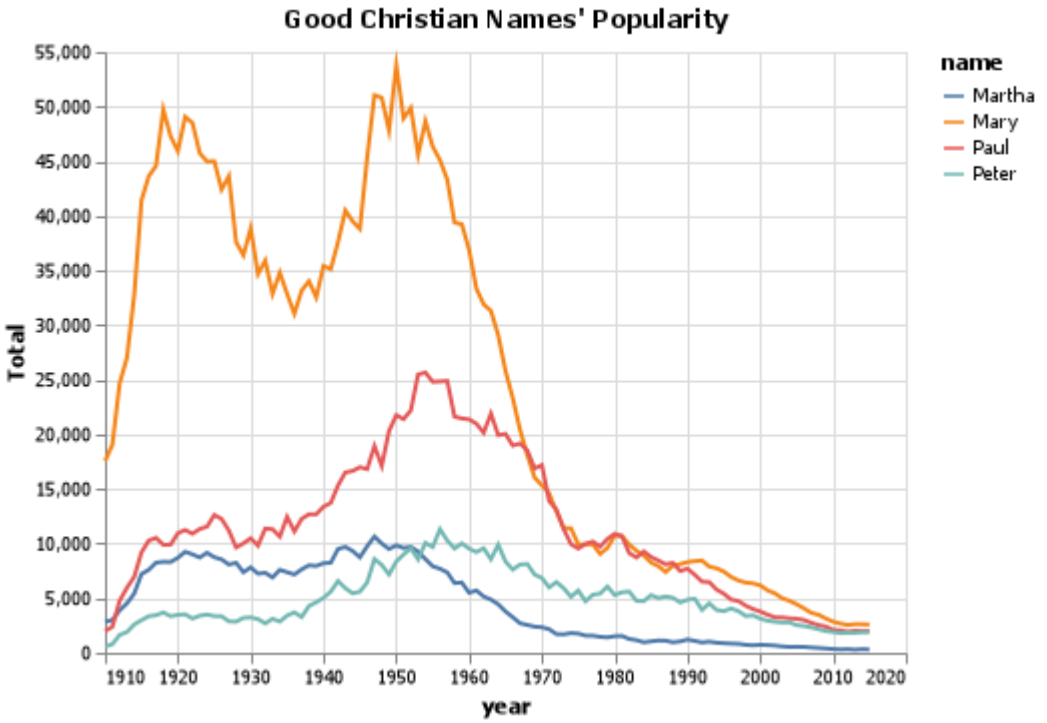
GRAND QUESTION 3

Mary, Martha, Peter, and Paul are all Christian names. From 1920 - 2000, compare the name usage of each of the four names. What trends do you notice?

Mary and Paul are the most popular of the names but the others hold up well, until they reach the 1970's then they all decline rapidly, until they are all under the 5000 mark on the y-axis which represents the Total from all the states that named their child those names.

▼ Read and format data

```
christian_names = ["Mary", "Martha", "Peter", "Paul"]
christian_name_data = data.query("name in @christian_names")
chart3 = alt.Chart(christian_name_data).encode(x = alt.X("year", axis = alt.Axis(format =
         "d")), y = "Total", color = "name").mark_line().properties(title = "Good Christian
         Names' Popularity")
save(chart3, "christian_names_data.png")
```



Christian Names Data

As you can see in on the chart Mary and Paul are the most popular out of the Christian names given and all of them were most popular from the 1920's to the 1970's where they began to decline rapidly.

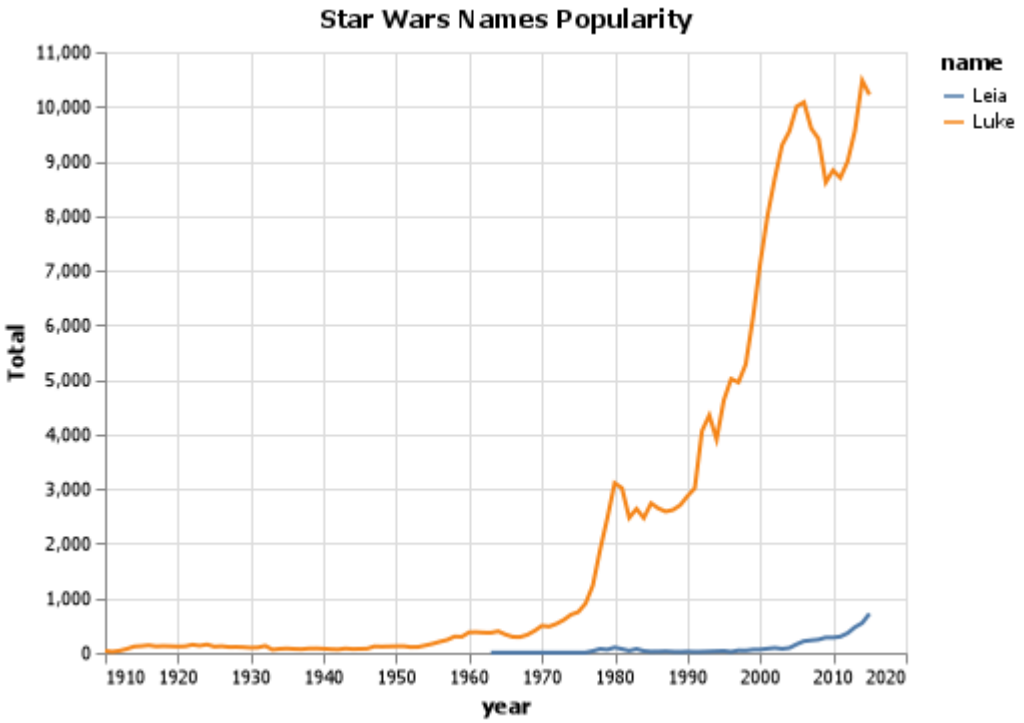
GRAND QUESTION 4

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. Does it look like the movie had an effect on usage?

I tested out the name Leia from the popular franchise Star Wars, there wasn't a significant spike after the movies came out but there was a rise in the name popularity the real spike came between 2005 and 2015. I believe the reasoning behind it is that all the fans who grew up watching Star Wars decided to name their children after some of the characters, one of the bigger examples is Luke which is also on the graph.

▼ Read and format data

```
starwars_names = ['Leia','Luke']
starwars_data = data.query("name == @starwars_names")
chart4 = alt.Chart(starwars_data).encode(x = alt.X("year", axis = alt.Axis(format = "d")), y =
         "Total", color = "name").mark_line().properties(title = "Star Wars Names Popularity")
save(chart4, "starwarsdata.png")
```



Star Wars Data

As you can see in the chart Leia and Luke get a small spike a couple years after the first movie comes out in the 1970's. However, their real spike is the decade inbetween 2005 and 2015.