Client Report - W02 Project 1: What’s in a name?

Course DS 250

Scott LeFoll

## Elevator pitch

You can tell a lot about a civilization from it’s language, and one of the most important aspects of language are the words that we used to represent people. A name is the essence of a thing, and the trends in child names over the last century tell a story about the evolution of our culture. In this project I will show visualizations that compare the relative prominence of names over time, and I will also show how the popularity of names has changed over time.

# Include and execute your code here  
  
#read in url to df  
url = 'https://github.com/byuidatascience/data4names/raw/master/data-raw/names\_year/names\_year.csv'  
dat = pd.read\_csv(url)  
# Show some data  
dat.tail(-5)

|  | name | year | AK | AL | AR | AZ | CA | CO | CT | DC | ... | TN | TX | UT | VA | VT | WA | WI | WV | WY | Total |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | Aaden | 2011 | 0.0 | 10.0 | 0.0 | 0.0 | 39.0 | 7.0 | 0.0 | 0.0 | ... | 6.0 | 26.0 | 0.0 | 7.0 | 0.0 | 5.0 | 5.0 | 0.0 | 0.0 | 228.0 |
| 6 | Aaden | 2012 | 0.0 | 7.0 | 0.0 | 5.0 | 38.0 | 0.0 | 0.0 | 0.0 | ... | 5.0 | 17.0 | 0.0 | 0.0 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 | 167.0 |
| 7 | Aaden | 2013 | 0.0 | 5.0 | 0.0 | 0.0 | 29.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 154.0 |
| 8 | Aaden | 2014 | 0.0 | 0.0 | 0.0 | 5.0 | 31.0 | 6.0 | 0.0 | 0.0 | ... | 0.0 | 24.0 | 0.0 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 181.0 |
| 9 | Aaden | 2015 | 0.0 | 5.0 | 0.0 | 14.0 | 34.0 | 5.0 | 7.0 | 0.0 | ... | 0.0 | 35.0 | 0.0 | 9.0 | 0.0 | 0.0 | 6.0 | 0.0 | 0.0 | 247.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 393379 | Zyon | 2011 | 0.0 | 0.0 | 0.0 | 6.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 5.0 | 9.0 | 0.0 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 117.0 |
| 393380 | Zyon | 2012 | 0.0 | 0.0 | 0.0 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 7.5 | 0.0 | 12.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 118.5 |
| 393381 | Zyon | 2013 | 0.0 | 0.0 | 0.0 | 0.0 | 6.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 93.0 |
| 393382 | Zyon | 2014 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.0 | ... | 0.0 | 7.0 | 0.0 | 9.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 73.0 |
| 393383 | Zyon | 2015 | 0.0 | 0.0 | 0.0 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 10.0 | 0.0 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 87.5 |

## GRAND QUESTION 1

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

The name ‘Scott’ showed 26,177 occurrences in the year 1964, with the largest geographic areas of usage in the Mid-west and West. Historically, the name showed minimal usage throughout the 20th century until the 1950’s, when it began to appear in popular culture. The greatest spike in usage for ‘Scott’ appears in the mid-1960’s, with the largest years being 1963 - 1966, including my own birth year. There was at least one very popular ‘hit’ song in 1963 that featured the name prominently, and the name was also used in a popular television show in 1964. The name ‘Scott’ has been used in popular culture since the 1950’s, but it was not until the 1960’s that it became a popular name for children. Starting in 1972 - 1973 the name began to decline in popularity, and it has been on a steady decline since then. The name ‘Scott’ is currently ranked 1,000th in popularity.

GQ1 TABLE 1A - showing the occurrence of the name ‘Scott’ in each U.S. state for the year 1964:

# Include and execute your code here\  
  
# query the larger data file for name and year  
mydat = dat.query('name == "Scott" & year == 1964')  
# melt the data so that states become rows instead of columns  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
mydat\_melt.drop(columns=['name'], inplace=True)  
mydat\_melt.drop(columns=['year'], inplace=True)  
  
# rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
  
# Get indexe where variable column equals 'Total'  
indexState = mydat\_melt[mydat\_melt['State'] == 'Total'].index   
# Delete these row indexes from dataFrame  
mydat\_melt.drop(indexState , inplace=True)  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| State | Total |
| --- | --- |
| AK | 54 |
| AL | 197 |
| AR | 133 |
| AZ | 191 |
| CA | 2581 |
| CO | 383 |
| CT | 249 |
| DC | 104 |
| DE | 63 |
| FL | 652 |
| GA | 340 |
| HI | 100 |
| IA | 801 |
| ID | 116 |
| IL | 1773 |
| IN | 1002 |
| KS | 325 |
| KY | 234 |
| LA | 293 |
| MA | 1114 |
| MD | 384 |
| ME | 285 |
| MI | 1779 |
| MN | 1088 |
| MO | 642 |
| MS | 108 |
| MT | 154 |
| NC | 314 |
| ND | 186 |
| NE | 388 |
| NH | 184 |
| NJ | 947 |
| NM | 70 |
| NV | 63 |
| NY | 1179.5 |
| OH | 930.5 |
| OK | 257 |
| OR | 302 |
| PA | 1864 |
| RI | 152 |
| SC | 145 |
| SD | 217 |
| TN | 257 |
| TX | 819 |
| UT | 272 |
| VA | 320 |
| VT | 86 |
| WA | 605 |
| WI | 1248 |
| WV | 168 |
| WY | 58 |

GQ1 CHART 1A - showing the occurrence of the name ‘Scott’ in each U.S. state for the year 1964:

# Include and execute your code here  
  
# compute the total number of occurrences for 'Scott' in 1964 across all states  
  
alt.Chart(mydat\_melt, title="Occurrences of the name 'Scott' in U.S. States in 1964")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 10)\  
 .properties(  
 width=800,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 2A - showing the total occurrences of the name ‘Scott’ in all U.S. states for the year 1964:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Scott' in 1964 for each state, plus total  
mydat\_sum = mydat[["name","year","Total"]]  
Markdown(mydat\_sum.to\_markdown(index=False))  
  
# mydat\_sum = mydat\_melt['Total'].sum()  
# mydat\_sum  
# Markdown(mydat\_sum.to\_markdown(index=False))

| name | year | Total |
| --- | --- | --- |
| Scott | 1964 | 26177 |

GQ1 CHART 2A - showing the total occurrences of the name ‘Scott’ in all U.S. states for the year 1964:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Scott' in 1964 for each state, plus total  
alt.Chart(mydat\_sum, title="Total occurrences of the name 'Scott' in the U.S. in 1964")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 50)\  
 .properties(  
 width=200,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 3A - showing the 5 states with the least occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# compute the 5 min states for occurences of 'Scott' in 1964  
  
mydat\_min = mydat\_melt.nsmallest(5,'Total')  
# mydat\_min  
Markdown(mydat\_min.to\_markdown(index=False))

| State | Total |
| --- | --- |
| AK | 54 |
| WY | 58 |
| DE | 63 |
| NV | 63 |
| NM | 70 |

GQ1 CHART 3A - showing the 5 states with the least occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# Chart the 5 min states for occurences of 'Scott' in 1964  
  
alt.Chart(mydat\_min, title="Five U.S. States with the least occurrences of the name 'Scott' in the U.S. in 1964")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 20)\  
 .properties(  
 width=400,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 4A - Table showing the 5 states with the most occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# compute the max 5 states for occurences of 'Scott' in 1964  
  
mydat\_max = mydat\_melt.nlargest(5,'Total')  
mydat\_max  
Markdown(mydat\_max.to\_markdown(index=False))

| State | Total |
| --- | --- |
| CA | 2581 |
| PA | 1864 |
| MI | 1779 |
| IL | 1773 |
| WI | 1248 |

GQ1 CHART 4A - Table showing the 5 states with the most occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# Chart the max 5 states for occurences of 'Scott' in 1964  
  
alt.Chart(mydat\_max, title="Five U.S. States with the largest occurrences of the name 'Scott' in the U.S. in 1964")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 20)\  
 .properties(  
 width=400,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 1B - showing the 5 states with the least occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# Occurrences of 'Scott' for all U.S. States for each year from 1910 - 2015   
  
# # compute the total number of occurrences for 'Scott' for all years across all states  
mydat = dat.query('name == "Scott" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 7 |
| 1911 | 21 |
| 1912 | 49 |
| 1913 | 67 |
| 1914 | 82 |
| 1915 | 105 |
| 1916 | 105 |
| 1917 | 123 |
| 1918 | 108 |
| 1919 | 133 |
| 1920 | 150 |
| 1921 | 144 |
| 1922 | 152 |
| 1923 | 152 |
| 1924 | 175 |
| 1925 | 131 |
| 1926 | 103 |
| 1927 | 114 |
| 1928 | 118 |
| 1929 | 106 |
| 1930 | 89 |
| 1931 | 66 |
| 1932 | 115 |
| 1933 | 111 |
| 1934 | 120 |
| 1935 | 148 |
| 1936 | 170 |
| 1937 | 159 |
| 1938 | 192 |
| 1939 | 228 |
| 1940 | 269 |
| 1941 | 344 |
| 1942 | 474 |
| 1943 | 513 |
| 1944 | 569 |
| 1945 | 835 |
| 1946 | 1502 |
| 1947 | 2019 |
| 1948 | 2286 |
| 1949 | 3039 |
| 1950 | 4270 |
| 1951 | 5380 |
| 1952 | 6197 |
| 1953 | 7489 |
| 1954 | 9456 |
| 1955 | 10770 |
| 1956 | 13219 |
| 1957 | 15141.5 |
| 1958 | 19166 |
| 1959 | 19757 |
| 1960 | 21608 |
| 1961 | 21203 |
| 1962 | 27307.5 |
| 1963 | 26419.5 |
| 1964 | 26177 |
| 1965 | 21132.5 |
| 1966 | 22922.5 |
| 1967 | 22329 |
| 1968 | 21204.5 |
| 1969 | 21667 |
| 1970 | 21954.5 |
| 1971 | 23554.5 |
| 1972 | 18757.5 |
| 1973 | 17315 |
| 1974 | 15579.5 |
| 1975 | 14277.5 |
| 1976 | 14065.5 |
| 1977 | 12638.5 |
| 1978 | 12406.5 |
| 1979 | 11438.5 |
| 1980 | 11478 |
| 1981 | 11708.5 |
| 1982 | 11596 |
| 1983 | 10270.5 |
| 1984 | 8815.5 |
| 1985 | 9204.5 |
| 1986 | 9113.5 |
| 1987 | 9149 |
| 1988 | 9344 |
| 1989 | 8039.5 |
| 1990 | 7587 |
| 1991 | 7109 |
| 1992 | 6059 |
| 1993 | 5572 |
| 1994 | 4916 |
| 1995 | 4080 |
| 1996 | 3665 |
| 1997 | 3339 |
| 1998 | 3062 |
| 1999 | 2717 |
| 2000 | 2453 |
| 2001 | 2290 |
| 2002 | 2018 |
| 2003 | 1758 |
| 2004 | 1632 |
| 2005 | 1455 |
| 2006 | 1345 |
| 2007 | 1234 |
| 2008 | 1008 |
| 2009 | 888 |
| 2010 | 837 |
| 2011 | 776 |
| 2012 | 740 |
| 2013 | 653 |
| 2014 | 678 |
| 2015 | 608 |

GQ1 CHART 1B - showing the 5 states with the least occurrences of the name ‘Scott’ in the U.S. for the year 1964:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Scott' for all years for each state, plus total  
  
alt.Chart(mydat\_melt, title="Occurrences of the name 'Scott' accross all U.S. States from 1910 to 2015")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 2B - showing the total historical occurrences of the name ‘Scott’ for all U.S. states for the years 1910 - 2015:

# Include and execute your code here  
  
# # compute the total number of occurrences for 'Scott' for all years for each state  
mydat = dat.query('name == "Scott" & year > 1880')  
  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
  
# rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.groupby('State').Total.sum().reset\_index()  
  
# Get index where variable column equals 'Total'  
indexState = mydat\_melt[mydat\_melt['State'] == 'Total'].index   
# Delete these row indexes from dataFrame  
mydat\_melt.drop(indexState , inplace=True)  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| State | Total |
| --- | --- |
| AK | 1524 |
| AL | 4953 |
| AR | 3867 |
| AZ | 7554 |
| CA | 51340.5 |
| CO | 10876 |
| CT | 11707 |
| DC | 2668 |
| DE | 1773 |
| FL | 18252 |
| GA | 9042 |
| HI | 2999 |
| IA | 16363 |
| ID | 4358 |
| IL | 40272.5 |
| IN | 21790.5 |
| KS | 9231 |
| KY | 6785 |
| LA | 8614 |
| MA | 25722.5 |
| MD | 10309.5 |
| ME | 6776 |
| MI | 37879.5 |
| MN | 27106 |
| MO | 17201 |
| MS | 2924 |
| MT | 3735 |
| NC | 8884 |
| ND | 4379 |
| NE | 9555 |
| NH | 4968 |
| NJ | 23639.5 |
| NM | 2499 |
| NV | 2203 |
| NY | 47941 |
| OH | 41483.5 |
| OK | 7072 |
| OR | 10505 |
| PA | 38899.5 |
| RI | 3899 |
| SC | 4168 |
| SD | 4246 |
| TN | 6775 |
| TX | 23667.5 |
| UT | 10627 |
| VA | 10422 |
| VT | 2538 |
| WA | 17621 |
| WI | 29682.5 |
| WV | 4482 |
| WY | 1615 |

GQ1 CHART 2B - showing the total historical occurrences of the name ‘Scott’ for all U.S. states for the years 1910 - 2015:

# Include and execute your code here  
  
# # compute the total number of occurrences for 'Scott' for all years across all states  
alt.Chart(mydat\_melt, title="Total occurrence of the name 'Scott' accross each U.S. State from 1910 to 2015")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=800,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 3B - showing the states with the least historical occurrences of the name ‘Scott’ for the years 1910 - 2015:

# Include and execute your code here  
  
# compute the 5 min states for occurences of 'Scott' across all years  
  
mydat\_min = mydat\_melt.nsmallest(5,'Total')  
mydat\_min

|  | State | Total |
| --- | --- | --- |
| 0 | AK | 1524.0 |
| 51 | WY | 1615.0 |
| 8 | DE | 1773.0 |
| 33 | NV | 2203.0 |
| 32 | NM | 2499.0 |

GQ1 CHART 3B - showing the states with the least historical occurrences of the name ‘Scott’ for the years 1910 - 2015:

# Include and execute your code here  
  
# Chart the 5 min states for occurences of 'Scott' across all years  
  
alt.Chart(mydat\_min, title="Five U.S. States with the least occurrences of the name 'Scott' in the U.S. from 1910 - 1915")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 20)\  
 .properties(  
 width=400,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

GQ1 TABLE 4B - showing the states with the most historical occurrences of the name ‘Scott’ for the years 1910 - 2015:

# Include and execute your code here  
  
# compute the max 5 states for occurences of 'Scott' in 1964  
  
mydat\_max = mydat\_melt.nlargest(5,'Total')  
mydat\_max

|  | State | Total |
| --- | --- | --- |
| 4 | CA | 51340.5 |
| 34 | NY | 47941.0 |
| 35 | OH | 41483.5 |
| 14 | IL | 40272.5 |
| 38 | PA | 38899.5 |

GQ1 CHART 4B - showing the states with the most historical occurrences of the name ‘Scott’ for the years 1910 - 2015:

# Include and execute your code here  
  
# Chart the max 5 states for occurences of 'Scott' in 1964  
  
alt.Chart(mydat\_max, title="Five U.S. States with the largest occurrences of the name 'Scott' in the U.S. from 1910 - 1915")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 20)\  
 .properties(  
 width=400,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

*include figures in chunks and discuss your findings in the figure.*

## 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?

*type your results and analysis here*

Based upon the data my best guess would be 1985 to 1998 for a birth year. The name ‘Brittany’ showed very little occurrence until the rise of the pop music star “Brittany Spears,” and rose and fell with her popularity. The name shows dramatic increase in usage in 1985, and then a dramatic decrease starting in 1998.

GQ2 TABLE 1A - showing the total historical occurrences of the name ‘Brittany’ for each year across all U.S. states for the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Brittany' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Brittany" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1968 | 5 |
| 1969 | 12 |
| 1970 | 32 |
| 1971 | 81 |
| 1972 | 158 |
| 1973 | 166 |
| 1974 | 198 |
| 1975 | 277 |
| 1976 | 304 |
| 1977 | 448 |
| 1978 | 592 |
| 1979 | 764 |
| 1980 | 1383 |
| 1981 | 1701 |
| 1982 | 3093 |
| 1983 | 4377 |
| 1984 | 7664 |
| 1985 | 14010 |
| 1986 | 17856.5 |
| 1987 | 18825.5 |
| 1988 | 21952 |
| 1989 | 30848 |
| 1990 | 32562.5 |
| 1991 | 26963.5 |
| 1992 | 23416.5 |
| 1993 | 21728 |
| 1994 | 17808.5 |
| 1995 | 15875.5 |
| 1996 | 13796 |
| 1997 | 11527 |
| 1998 | 9843 |
| 1999 | 7942 |
| 2000 | 5183 |
| 2001 | 2915 |
| 2002 | 1912 |
| 2003 | 1559 |
| 2004 | 1323.5 |
| 2005 | 1168 |
| 2006 | 1009 |
| 2007 | 891 |
| 2008 | 749 |
| 2009 | 644 |
| 2010 | 698 |
| 2011 | 717 |
| 2012 | 745 |
| 2013 | 699 |
| 2014 | 660 |
| 2015 | 636 |

*include figures in chunks and discuss your findings in the figure.*

GQ2 CHART 1A - showing the total historical occurrences of the name ‘Brittany’ for each year across all U.S. states for the years 1910 - 2015:

# Include and execute your code here  
alt.Chart(mydat\_melt, title="Annual occurrence of the name 'Brittany' in the U.S. across all states from 1910 - 1915")\  
 .encode(x = "year:N", y = "Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

My useless chart

GQ2 TABLE 2A - showing the total occurrences of the name ‘Brittany’ in each U.S. state for the years 1910 - 2015:

# Include and execute your code here  
  
# # compute the total number of occurrences for 'Brittany' for all years for each state  
mydat = dat.query('name == "Brittany" & year > 1880')  
  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
  
# rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.groupby('State').Total.sum().reset\_index()  
  
# Get index where variable column equals 'Total'  
indexState = mydat\_melt[mydat\_melt['State'] == 'Total'].index   
# Delete these row indexes from dataFrame  
mydat\_melt.drop(indexState , inplace=True)  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| State | Total |
| --- | --- |
| AK | 750 |
| AL | 7431.5 |
| AR | 4362 |
| AZ | 4963 |
| CA | 18956.5 |
| CO | 5165 |
| CT | 3169 |
| DC | 1122.5 |
| DE | 1043 |
| FL | 18438 |
| GA | 14179.5 |
| HI | 747 |
| IA | 3004 |
| ID | 1101 |
| IL | 15387 |
| IN | 10255 |
| KS | 2350 |
| KY | 9042.5 |
| LA | 8699.5 |
| MA | 5352 |
| MD | 7343 |
| ME | 1362 |
| MI | 11028 |
| MN | 5228 |
| MO | 7714 |
| MS | 5129 |
| MT | 823 |
| NC | 14090 |
| ND | 1072 |
| NE | 1807 |
| NH | 1238 |
| NJ | 7747 |
| NM | 1630 |
| NV | 1732 |
| NY | 15134.5 |
| OH | 17697.5 |
| OK | 4920 |
| OR | 3252 |
| PA | 12759 |
| RI | 1034 |
| SC | 7604 |
| SD | 905 |
| TN | 9325.5 |
| TX | 21272.5 |
| UT | 3199 |
| VA | 10986 |
| VT | 654 |
| WA | 4758 |
| WI | 6260 |
| WV | 3915 |
| WY | 581 |

GQ2 CHART 2A - showing the total occurrences of the name ‘Brittany’ in each U.S. state for the years 1910 - 2015:

# Include and execute your code here  
  
# # compute the total number of occurrences for 'Brittany' for all years across all states  
alt.Chart(mydat\_melt, title="Total occurrence of the name 'Brittany' accross each U.S. State from 1910 to 2015")\  
 .encode(x="State:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=800,  
 height=300  
  
 )

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.Chart(...)

## 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?

*type your results and analysis here*

Usage of the names ‘Mary’ and ‘Martha’ both show spikes during W.W. I and W.W. II, increasing their pre-war totals by 100%, from about 20,000 to 45,000 during each war and droping back to about 30,000 between the wars. Both names exhibit the same pattern of occurrence for the last 70 years, with sharp declines in usage right after W.W. II followed by a steady decline from the 1970’s to the present.

Usage of the names ‘Peter’ and ‘Paul’ both show spikes during W.W. I, increasing their pre-war totals by about 200%, with no drop between the wars and a further spike of another 150% during W.W.II. Both names exhibit the same pattern of occurrence for the last 70 years, with sharp spikes in the 1960’s with the rebirth of interest in spiritual concepts and folk music in the 1960’s follwed by a steady decline from the 1970’s to the present.

The rise for all four names during both world wars makes sense considering that it was a time of grave instability and danger for the world, for the nation and for individuals. People were reaching for stability, for peace, for tradition and for continuity and looking to the Bible to find it.

GQ3 TABLE 1A - showing the total historical occurrences of the name ‘Mary’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Mary' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Mary" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 17568.5 |
| 1911 | 19058.5 |
| 1912 | 24685.5 |
| 1913 | 27013.5 |
| 1914 | 32734.5 |
| 1915 | 41425 |
| 1916 | 43642.5 |
| 1917 | 44589 |
| 1918 | 49738.5 |
| 1919 | 47296 |
| 1920 | 45928.5 |
| 1921 | 49070 |
| 1922 | 48514 |
| 1923 | 45749.5 |
| 1924 | 44981 |
| 1925 | 45006.5 |
| 1926 | 42416.5 |
| 1927 | 43645.5 |
| 1928 | 37597.5 |
| 1929 | 36396.5 |
| 1930 | 38872.5 |
| 1931 | 34680 |
| 1932 | 35923.5 |
| 1933 | 32867.5 |
| 1934 | 34809.5 |
| 1935 | 32865.5 |
| 1936 | 31045 |
| 1937 | 33171.5 |
| 1938 | 34026.5 |
| 1939 | 32546 |
| 1940 | 35422.5 |
| 1941 | 35127 |
| 1942 | 37619 |
| 1943 | 40505.5 |
| 1944 | 39509.5 |
| 1945 | 38813 |
| 1946 | 45337.5 |
| 1947 | 51053 |
| 1948 | 50828 |
| 1949 | 47835 |
| 1950 | 53791 |
| 1951 | 48928.5 |
| 1952 | 49898.5 |
| 1953 | 45603 |
| 1954 | 48678.5 |
| 1955 | 46351 |
| 1956 | 45111 |
| 1957 | 43334 |
| 1958 | 39392 |
| 1959 | 39242 |
| 1960 | 36910.5 |
| 1961 | 33353.5 |
| 1962 | 31908.5 |
| 1963 | 31314.5 |
| 1964 | 29094 |
| 1965 | 25823 |
| 1966 | 23328.5 |
| 1967 | 20413.5 |
| 1968 | 18096.5 |
| 1969 | 16045 |
| 1970 | 15333 |
| 1971 | 14606.5 |
| 1972 | 12914.5 |
| 1973 | 11429.5 |
| 1974 | 11398 |
| 1975 | 9916 |
| 1976 | 9941.5 |
| 1977 | 9936.5 |
| 1978 | 9057.5 |
| 1979 | 9605 |
| 1980 | 10730.5 |
| 1981 | 10692 |
| 1982 | 9912.5 |
| 1983 | 9345 |
| 1984 | 8916.5 |
| 1985 | 8269.5 |
| 1986 | 7938.5 |
| 1987 | 7402.5 |
| 1988 | 7992.5 |
| 1989 | 8127.5 |
| 1990 | 8308 |
| 1991 | 8400 |
| 1992 | 8454 |
| 1993 | 7891.5 |
| 1994 | 7746 |
| 1995 | 7438 |
| 1996 | 6946 |
| 1997 | 6627 |
| 1998 | 6434 |
| 1999 | 6360 |
| 2000 | 6184 |
| 2001 | 5725 |
| 2002 | 5452 |
| 2003 | 5004 |
| 2004 | 4770 |
| 2005 | 4445 |
| 2006 | 4078 |
| 2007 | 3666 |
| 2008 | 3488 |
| 2009 | 3149 |
| 2010 | 2856 |
| 2011 | 2684 |
| 2012 | 2554 |
| 2013 | 2626 |
| 2014 | 2621 |
| 2015 | 2588 |

GQ3 CHART 1A - showing the total historical occurrences of the name ‘Mary’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Mary' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Mary' accross all U.S. States from 1910 to 2015")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )  
  
lines\_df = pd.DataFrame({'big\_events': [1914, 1918, 1939, 1945, 1962, 1969]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "big\_events:N")  
 .properties(  
 width = 1100,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

GQ3 TABLE 2A - showing the total historical occurrences of the name ‘Martha’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Martha' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Martha" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 2837 |
| 1911 | 3008 |
| 1912 | 3902 |
| 1913 | 4527 |
| 1914 | 5430 |
| 1915 | 7217 |
| 1916 | 7606 |
| 1917 | 8259 |
| 1918 | 8339 |
| 1919 | 8318 |
| 1920 | 8705 |
| 1921 | 9254 |
| 1922 | 9018 |
| 1923 | 8731 |
| 1924 | 9163 |
| 1925 | 8757 |
| 1926 | 8557 |
| 1927 | 8088.5 |
| 1928 | 8245 |
| 1929 | 7381.5 |
| 1930 | 7830 |
| 1931 | 7248.5 |
| 1932 | 7329.5 |
| 1933 | 6905 |
| 1934 | 7603 |
| 1935 | 7383.5 |
| 1936 | 7176.5 |
| 1937 | 7645 |
| 1938 | 8017.5 |
| 1939 | 7956.5 |
| 1940 | 8204 |
| 1941 | 8250.5 |
| 1942 | 9514 |
| 1943 | 9700 |
| 1944 | 9329 |
| 1945 | 8744 |
| 1946 | 9763 |
| 1947 | 10651 |
| 1948 | 10006 |
| 1949 | 9508 |
| 1950 | 9840 |
| 1951 | 9598 |
| 1952 | 9699 |
| 1953 | 9262 |
| 1954 | 8585 |
| 1955 | 7930 |
| 1956 | 7703 |
| 1957 | 7365 |
| 1958 | 6393.5 |
| 1959 | 6439 |
| 1960 | 5504.5 |
| 1961 | 5712 |
| 1962 | 5164 |
| 1963 | 4879 |
| 1964 | 4448 |
| 1965 | 3793 |
| 1966 | 3267 |
| 1967 | 2708 |
| 1968 | 2557 |
| 1969 | 2399 |
| 1970 | 2346 |
| 1971 | 2167 |
| 1972 | 1718 |
| 1973 | 1698 |
| 1974 | 1811 |
| 1975 | 1768 |
| 1976 | 1580 |
| 1977 | 1572 |
| 1978 | 1469 |
| 1979 | 1420 |
| 1980 | 1525 |
| 1981 | 1549 |
| 1982 | 1270.5 |
| 1983 | 1166 |
| 1984 | 969.5 |
| 1985 | 1057 |
| 1986 | 1134 |
| 1987 | 1117 |
| 1988 | 969 |
| 1989 | 1049.5 |
| 1990 | 1222 |
| 1991 | 1096 |
| 1992 | 955.5 |
| 1993 | 1027 |
| 1994 | 928 |
| 1995 | 887 |
| 1996 | 849 |
| 1997 | 830 |
| 1998 | 737 |
| 1999 | 693 |
| 2000 | 741 |
| 2001 | 713 |
| 2002 | 670 |
| 2003 | 592 |
| 2004 | 548 |
| 2005 | 557 |
| 2006 | 551 |
| 2007 | 492 |
| 2008 | 452 |
| 2009 | 402 |
| 2010 | 353 |
| 2011 | 315 |
| 2012 | 340 |
| 2013 | 292 |
| 2014 | 337 |
| 2015 | 312 |

GQ3 CHART 2A - showing the total historical occurrences of the name ‘Martha’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Martha' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Martha' accross all U.S. States from 1910 to 2015, with W.W.I, W.W.II & 60's Revolution plotted")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )  
  
lines\_df = pd.DataFrame({'big\_events': [1914, 1918, 1939, 1945, 1962, 1969]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "big\_events:N")  
 .properties(  
 width = 1100,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

GQ3 TABLE 3A - showing the total historical occurrences of the name ‘Peter’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Peter' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Peter" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 578 |
| 1911 | 793 |
| 1912 | 1647 |
| 1913 | 1903 |
| 1914 | 2604 |
| 1915 | 2993.5 |
| 1916 | 3359 |
| 1917 | 3452 |
| 1918 | 3698 |
| 1919 | 3358 |
| 1920 | 3479 |
| 1921 | 3530 |
| 1922 | 3143.5 |
| 1923 | 3403 |
| 1924 | 3510 |
| 1925 | 3369 |
| 1926 | 3355 |
| 1927 | 2883 |
| 1928 | 2858.5 |
| 1929 | 3201 |
| 1930 | 3251 |
| 1931 | 3093 |
| 1932 | 2694.5 |
| 1933 | 3120 |
| 1934 | 2865 |
| 1935 | 3413 |
| 1936 | 3706 |
| 1937 | 3283.5 |
| 1938 | 4289 |
| 1939 | 4633 |
| 1940 | 5039 |
| 1941 | 5608 |
| 1942 | 6582 |
| 1943 | 5920 |
| 1944 | 5450 |
| 1945 | 5604 |
| 1946 | 6449 |
| 1947 | 8591 |
| 1948 | 8050 |
| 1949 | 7180 |
| 1950 | 8366 |
| 1951 | 9004 |
| 1952 | 9581 |
| 1953 | 8561 |
| 1954 | 10073 |
| 1955 | 9689.5 |
| 1956 | 11321 |
| 1957 | 10284.5 |
| 1958 | 9555 |
| 1959 | 10028.5 |
| 1960 | 9545 |
| 1961 | 9225.5 |
| 1962 | 9577 |
| 1963 | 8571.5 |
| 1964 | 9907 |
| 1965 | 8364.5 |
| 1966 | 7626.5 |
| 1967 | 8087 |
| 1968 | 8133 |
| 1969 | 7170 |
| 1970 | 6854 |
| 1971 | 5995.5 |
| 1972 | 6479 |
| 1973 | 5918 |
| 1974 | 5128.5 |
| 1975 | 5707 |
| 1976 | 4721 |
| 1977 | 5323 |
| 1978 | 5425.5 |
| 1979 | 6076 |
| 1980 | 5296 |
| 1981 | 5551 |
| 1982 | 5611 |
| 1983 | 4760.5 |
| 1984 | 4743.5 |
| 1985 | 5285.5 |
| 1986 | 5018 |
| 1987 | 5161 |
| 1988 | 5042.5 |
| 1989 | 4620.5 |
| 1990 | 4866.5 |
| 1991 | 4956 |
| 1992 | 3934.5 |
| 1993 | 4532 |
| 1994 | 3928.5 |
| 1995 | 3817.5 |
| 1996 | 4069 |
| 1997 | 3821 |
| 1998 | 3377 |
| 1999 | 3430 |
| 2000 | 3137 |
| 2001 | 2944 |
| 2002 | 2851 |
| 2003 | 2780 |
| 2004 | 2787 |
| 2005 | 2540 |
| 2006 | 2464 |
| 2007 | 2331 |
| 2008 | 2167 |
| 2009 | 1992 |
| 2010 | 1914 |
| 2011 | 1814 |
| 2012 | 1841 |
| 2013 | 1835 |
| 2014 | 1896 |
| 2015 | 1902 |

GQ3 CHART 3A - showing the total historical occurrences of the name ‘Peter’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Peter' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Peter' accross all U.S. States from 1910 to 2015")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )  
  
lines\_df = pd.DataFrame({'big\_events': [1914, 1918, 1939, 1945, 1962, 1969]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "big\_events:N")  
 .properties(  
 width = 1100,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

GQ3 TABLE 4A - showing the total historical occurrences of the name ‘Paul’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Paul' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Paul" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 1998 |
| 1911 | 2384 |
| 1912 | 4777 |
| 1913 | 5962 |
| 1914 | 6966.5 |
| 1915 | 9209 |
| 1916 | 10284.5 |
| 1917 | 10534 |
| 1918 | 9890 |
| 1919 | 9913.5 |
| 1920 | 10958.5 |
| 1921 | 11251.5 |
| 1922 | 10907 |
| 1923 | 11341.5 |
| 1924 | 11553.5 |
| 1925 | 12626.5 |
| 1926 | 12262 |
| 1927 | 11203 |
| 1928 | 9650.5 |
| 1929 | 10022.5 |
| 1930 | 10493 |
| 1931 | 9812 |
| 1932 | 11379 |
| 1933 | 11324.5 |
| 1934 | 10642 |
| 1935 | 12450.5 |
| 1936 | 11121.5 |
| 1937 | 12269 |
| 1938 | 12690.5 |
| 1939 | 12685.5 |
| 1940 | 13382 |
| 1941 | 13725.5 |
| 1942 | 15335 |
| 1943 | 16520.5 |
| 1944 | 16654 |
| 1945 | 16997 |
| 1946 | 16823.5 |
| 1947 | 18934 |
| 1948 | 17120.5 |
| 1949 | 20309 |
| 1950 | 21763 |
| 1951 | 21399 |
| 1952 | 22197.5 |
| 1953 | 25497 |
| 1954 | 25662.5 |
| 1955 | 24818.5 |
| 1956 | 24852 |
| 1957 | 24897.5 |
| 1958 | 21623.5 |
| 1959 | 21464.5 |
| 1960 | 21377 |
| 1961 | 20980.5 |
| 1962 | 20155 |
| 1963 | 21904.5 |
| 1964 | 19928.5 |
| 1965 | 20051.5 |
| 1966 | 18978.5 |
| 1967 | 19168.5 |
| 1968 | 18531 |
| 1969 | 16884 |
| 1970 | 17193 |
| 1971 | 13902 |
| 1972 | 13123.5 |
| 1973 | 11327.5 |
| 1974 | 9951.5 |
| 1975 | 9554 |
| 1976 | 9980 |
| 1977 | 10191 |
| 1978 | 9725.5 |
| 1979 | 10428.5 |
| 1980 | 10898.5 |
| 1981 | 10694.5 |
| 1982 | 9129.5 |
| 1983 | 8733 |
| 1984 | 9284.5 |
| 1985 | 8774.5 |
| 1986 | 8454 |
| 1987 | 8112 |
| 1988 | 8238 |
| 1989 | 7494.5 |
| 1990 | 7705.5 |
| 1991 | 7150.5 |
| 1992 | 6522 |
| 1993 | 6477 |
| 1994 | 5796 |
| 1995 | 5413 |
| 1996 | 4879 |
| 1997 | 4748 |
| 1998 | 4319 |
| 1999 | 4015 |
| 2000 | 3786 |
| 2001 | 3500 |
| 2002 | 3257 |
| 2003 | 3248 |
| 2004 | 3161.5 |
| 2005 | 3129 |
| 2006 | 3012 |
| 2007 | 2753 |
| 2008 | 2545 |
| 2009 | 2409 |
| 2010 | 2111 |
| 2011 | 2055 |
| 2012 | 1933 |
| 2013 | 2023 |
| 2014 | 1988 |
| 2015 | 2006 |

GQ3 CHART 4A - showing the total historical occurrences of the name ‘Paul’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Paul' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Paul' accross all U.S. States from 1910 to 2015")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
  
 )  
  
lines\_df = pd.DataFrame({'big\_events': [1914, 1918, 1939, 1945, 1962, 1969]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "big\_events:N")  
 .properties(  
 width = 1100,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

## 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?

* Leia and Luke from Star Wars

*type your results and analysis here*

‘Luke’ and ‘Leia’ are the names of the two principal characters in the Star Wars series of movies. The popularity of both names rose significantly on the release of the first movie in 1977, and showed additional bumps with each subsequent release (1980, 1983, 1999, 2002, 2005, 2015, 2017, 2019). The popularity of the names has remained high since the release of the first movie, and has not declined since the release of the last movie in 2019. The last series of movies (2015, 2017, 2019) created a new bump in popularity for the names, far exceeding previous spikes (500%).

GQ4 TABLE 1A - showing the total historical occurrences of the name ‘Leia’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Leia' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Leia" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1963 | 6 |
| 1965 | 5 |
| 1967 | 6 |
| 1976 | 6 |
| 1977 | 34 |
| 1978 | 73 |
| 1979 | 60 |
| 1980 | 99 |
| 1981 | 71 |
| 1982 | 39 |
| 1983 | 73 |
| 1984 | 35 |
| 1985 | 23 |
| 1986 | 24 |
| 1987 | 30 |
| 1988 | 21 |
| 1989 | 18 |
| 1990 | 24 |
| 1991 | 19 |
| 1992 | 21 |
| 1993 | 26 |
| 1995 | 36 |
| 1996 | 19 |
| 1997 | 46 |
| 1998 | 43 |
| 1999 | 60 |
| 2000 | 62 |
| 2001 | 77 |
| 2002 | 91 |
| 2003 | 74 |
| 2004 | 90 |
| 2005 | 158 |
| 2006 | 215 |
| 2007 | 230 |
| 2008 | 243 |
| 2009 | 282 |
| 2010 | 280 |
| 2011 | 298 |
| 2012 | 361 |
| 2013 | 470 |
| 2014 | 547 |
| 2015 | 718 |

GQ4 CHART 1A - showing the total historical occurrences of the name ‘Leia’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Leia' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Luke' accross all U.S. States from 1910 to 2015, with Star Wars movie releases plotted")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 10)\  
 .properties(  
 width=800,  
 height=300  
 )  
  
lines\_df = pd.DataFrame({'movie\_releases': [1977, 1980, 1983, 1999, 2002, 2005, 2015]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "movie\_releases:N")  
 .properties(  
 width = 800,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

GQ4 TABLE 2A - showing the total historical occurrences of the name ‘Luke’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Occurrences of 'Luke' for all U.S. States for each year from 1910 - 2015   
  
mydat = dat.query('name == "Luke" & year > 1880')  
# melt the data so that the states become rows  
mydat\_melt = mydat.melt(id\_vars=['name', 'year'])  
# # rename columns  
mydat\_melt.rename(columns = {'variable':'State'}, inplace = True)  
mydat\_melt.rename(columns = {'value':'Total'}, inplace = True)  
mydat\_melt = mydat\_melt.query('State != "Total"')  
mydat\_melt = mydat\_melt.groupby('year').Total.sum().reset\_index()  
  
Markdown(mydat\_melt.to\_markdown(index=False))

| year | Total |
| --- | --- |
| 1910 | 37 |
| 1911 | 21 |
| 1912 | 39 |
| 1913 | 73 |
| 1914 | 117 |
| 1915 | 125 |
| 1916 | 140 |
| 1917 | 119 |
| 1918 | 125 |
| 1919 | 121 |
| 1920 | 116 |
| 1921 | 120 |
| 1922 | 150 |
| 1923 | 131 |
| 1924 | 154 |
| 1925 | 114 |
| 1926 | 122 |
| 1927 | 108 |
| 1928 | 109 |
| 1929 | 104 |
| 1930 | 95 |
| 1931 | 99 |
| 1932 | 129 |
| 1933 | 59 |
| 1934 | 74 |
| 1935 | 80 |
| 1936 | 72 |
| 1937 | 68 |
| 1938 | 80 |
| 1939 | 82 |
| 1940 | 76 |
| 1941 | 68 |
| 1942 | 63 |
| 1943 | 80 |
| 1944 | 71 |
| 1945 | 73 |
| 1946 | 77 |
| 1947 | 119 |
| 1948 | 115 |
| 1949 | 118 |
| 1950 | 121 |
| 1951 | 123 |
| 1952 | 106 |
| 1953 | 108 |
| 1954 | 137 |
| 1955 | 167 |
| 1956 | 205 |
| 1957 | 238 |
| 1958 | 299 |
| 1959 | 295 |
| 1960 | 376 |
| 1961 | 379 |
| 1962 | 371 |
| 1963 | 368 |
| 1964 | 396 |
| 1965 | 336 |
| 1966 | 293 |
| 1967 | 290 |
| 1968 | 331 |
| 1969 | 404 |
| 1970 | 496 |
| 1971 | 482 |
| 1972 | 532 |
| 1973 | 603 |
| 1974 | 700 |
| 1975 | 748 |
| 1976 | 904 |
| 1977 | 1235 |
| 1978 | 1888 |
| 1979 | 2471 |
| 1980 | 3108 |
| 1981 | 3017 |
| 1982 | 2473 |
| 1983 | 2638 |
| 1984 | 2469 |
| 1985 | 2739 |
| 1986 | 2646 |
| 1987 | 2588 |
| 1988 | 2617 |
| 1989 | 2704 |
| 1990 | 2870 |
| 1991 | 3015 |
| 1992 | 4064 |
| 1993 | 4351 |
| 1994 | 3910 |
| 1995 | 4628 |
| 1996 | 5015 |
| 1997 | 4952 |
| 1998 | 5282 |
| 1999 | 6133 |
| 2000 | 7141 |
| 2001 | 8004 |
| 2002 | 8692 |
| 2003 | 9297 |
| 2004 | 9549 |
| 2005 | 10003 |
| 2006 | 10082 |
| 2007 | 9602 |
| 2008 | 9412 |
| 2009 | 8614 |
| 2010 | 8835 |
| 2011 | 8697 |
| 2012 | 8998 |
| 2013 | 9558 |
| 2014 | 10480 |
| 2015 | 10219 |

GQ4 CHART 2A - showing the total historical occurrences of the name ‘Luke’ in all U.S. states for each of the years 1910 - 2015:

# Include and execute your code here  
  
# Display the total number of occurrences for 'Luke' for all years for each state, plus total  
  
chart = alt.Chart(mydat\_melt, title="Occurrences of the name 'Luke' accross all U.S. States from 1910 to 2015, with Star Wars movie releases plotted")\  
 .encode(x="year:N", y="Total:Q")\  
 .mark\_bar(width = 5)\  
 .properties(  
 width=1100,  
 height=300  
 )  
  
lines\_df = pd.DataFrame({'movie\_releases': [1977, 1980, 1983, 1999, 2002, 2005, 2015]})  
  
rules = (  
 alt.Chart(lines\_df)  
 .mark\_rule(  
 color = c\_cp[4],  
 opacity= 0.35,  
 size = 1  
 )  
 .encode(  
 # x = alt.X("x:Q", scale = alt.Scale(domain = (1910, 2015)))  
 x = "movie\_releases:N")  
 .properties(  
 width = 1100,  
 height = 400  
 # scale = alt.Scale(domain = (1910, 2015))  
 )  
)  
  
  
chart + rules

C:\Python310\lib\site-packages\altair\utils\core.py:317: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.  
 for col\_name, dtype in df.dtypes.iteritems():

alt.LayerChart(...)

## APPENDIX A (Additional Python Code)

#paste other your code from your python file (.py) here