Class Activity 1

Your name here

2024-03-20

The R package **babynames** provides data about the popularity of individual baby names from the US Social Security Administration. Data includes all names used at least 5 times in a year beginning in 1880.

```
#install.packages("babynames")  # uncomment to install
library(babynames)
```

Below is the list for first few cases of baby names.

```
head(babynames)
```

```
# A tibble: 6 x 5
   year sex   name          n   prop
  <dbl> <chr> <chr>     <int>  <dbl>
1  1880 F     Mary       7065 0.0724
2  1880 F     Anna       2604 0.0267
3  1880 F     Emma       2003 0.0205
4  1880 F     Elizabeth  1939 0.0199
5  1880 F     Minnie     1746 0.0179
6  1880 F     Margaret   1578 0.0162
```

1. How many cases and variables are in the dataset **babynames**?

Answer:

```
dim(babynames)
```

```
[1] 1924665       5
```

There are 1924665 cases and 5 variables in the dataset `babynames`.

Let's use the package tidyverse to do some exploratory data analysis. `<-` is the assignment operator and `%>%` is the pipe operator. We will slowly learn these in detail later.

```
#install.packages("tidyverse")   # uncomment to install
library(tidyverse)
babynames %>% filter(name=='Aimee')
```

```
# A tibble: 150 x 5
    year sex   name      n      prop
   <dbl> <chr> <chr> <int>     <dbl>
 1  1880 F     Aimee    13 0.000133
 2  1881 F     Aimee    11 0.000111
 3  1882 F     Aimee    13 0.000112
 4  1883 F     Aimee    11 0.0000916
 5  1884 F     Aimee    15 0.000109
 6  1885 F     Aimee    17 0.000120
 7  1886 F     Aimee    17 0.000111
 8  1887 F     Aimee    18 0.000116
 9  1888 F     Aimee    12 0.0000633
10  1889 F     Aimee    16 0.0000846
# i 140 more rows
```

```
filtered_names <- babynames %>% filter(name=='Aimee')
```

```
#install.packages("ggplot2")   # uncomment to install
library(ggplot2)
```

```
ggplot(data=filtered_names, aes(x=year, y=prop)) +
  geom_line(aes(colour=sex)) +
  xlab('Year') +
  ylab('Prop. of Babies Named Aimee')
```

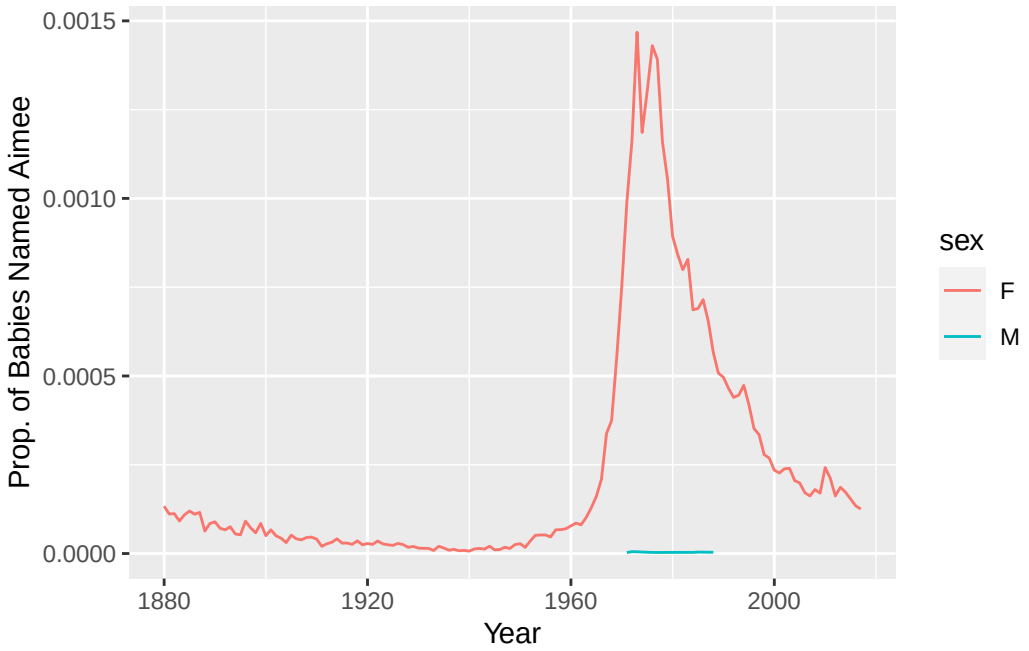

Figure 1: A trend chart

2. What do you see in the Figure 1? Explain in a few sentences.

Answer:

3. Repeat question 2 to infer how does the proportion of babies with your first name trend over time. Examine the generated plot and describe the trend of your name's popularity over time. Consider the following points:

Has the popularity of your name increased, decreased, or remained stable over the years? Is there a noticeable difference in popularity between sexes? Are there any interesting patterns or trends, such as sudden increases or decreases in popularity?

Answer:

```
# Replace 'YourName' with your first name
your_name <- "Dee"

your_name_data <- babynames %>% filter(name == your_name)
```

```
ggplot(data=your_name_data, aes(x=year, y=prop)) +
  geom_line(aes(colour=sex)) +
  xlab('Year') +
  ylab(paste('Prop. of Babies Named', your_name))
```