**Practice with dplyr package: Baby Names**

Install the R package babynames: *install.packages("babynames")*

For this example, you will use the *babynames* data set in the babynames package. It contains the United States Social Security Administration’s tabulation of names given to babies between 1880 and 2017. This R data set includes all names with at least 5 uses.

The data set is in a tibble (data frame), and R only shows the top rows of a tibble when you try to view it in the console window. You can change how many rows are displayed using options(). For example: *options(tibble.print\_min = 20)*

Use the dplyr and ggplot2 packages to answer the following questions:

1. Overall, what are the top 10 most popular baby names? How about since 1990?

names.MostPop<-babynames %>%

group\_by(name) %>%

summarize(total\_n=sum(n)) %>%

arrange(-total\_n)

names.MostPop

name total\_n

<chr> <int>

1 James 5173828

2 John 5137142

3 Robert 4834915

4 Michael 4372536

5 Mary 4138360

6 William 4118553

7 David 3624225

8 Joseph 2614083

9 Richard 2572613

10 Charles 2398453

How about since 1990?

names.MostPop\_1990<-babynames %>%

filter(year>=1990) %>%

group\_by(name) %>%

summarize(total\_n=sum(n)) %>%

arrange(-total\_n)

names.MostPop\_1990

name total\_n

<chr> <int>

1 Michael 838131

2 Jacob 713501

3 Matthew 679781

4 Joshua 654300

5 Christopher 652509

6 Daniel 589381

7 Andrew 567048

8 Emily 562187

9 Joseph 552864

10 William 544247

1. Now differentiate between genders. What are the top 10 names given to girls since 1990? Top 10 for boys? Hint: You can group by more than one variable!

Top 10 Female names since 1990

names.MostPop\_1990<-babynames %>%

filter(year>=1990,sex=='F') %>%

group\_by(name,sex) %>%

summarize(total\_n=sum(n)) %>%

arrange(-total\_n)

names.MostPop\_1990

name sex total\_n

<chr> <chr> <int>

1 Emily F 561299

2 Ashley F 467686

3 Jessica F 409833

4 Samantha F 407773

5 Emma F 397978

6 Sarah F 388002

7 Elizabeth F 383232

8 Olivia F 379362

9 Madison F 369832

10 Hannah F 367423

Top 10 male names since 1990

names.MostPop\_1990<-babynames %>%

filter(year>=1990,sex=='M') %>%

group\_by(name,sex) %>%

summarize(total\_n=sum(n)) %>%

arrange(-total\_n)

names.MostPop\_1990

name sex total\_n

<chr> <chr> <int>

1 Michael M 835096

2 Jacob M 712218

3 Matthew M 678528

4 Joshua M 652762

5 Christopher M 650925

6 Daniel M 587569

7 Andrew M 565863

8 Joseph M 551741

9 William M 543294

10 David M 530892

1. How many babies overall have been given *your* name since 1880? How many babies of your same gender have been given your name? How about since the year you were born?

My name since 1880, Craig

names.craig<-babynames %>%

filter(name=="Craig") %>%

summarize(total\_n=sum(n))

names.craig

total\_n

<int>

1 299583

Same name and gender, Craig Male

names.craig<-babynames %>%

filter(name=="Craig",sex=='M') %>%

summarize(total\_n=sum(n))

names.craig

total\_n

<int>

1 298571

Since the year I was born, Number of craigs since 2001

names.craig<-babynames %>%

filter(name=="Craig",sex=='M',year>=2001) %>%

summarize(total\_n=sum(n))

names.craig

total\_n

<int>

1 6826

1. Make a graph showing how the popularity of your name has changed over time since 1880. During what year was your name most popular?

ggplot(data=names.craig)+

geom\_point(mapping=aes(x=year,y=n,color=sex))+

labs(x="Year",y="Number of Babies Named Craig",

title = "Number of Babies Named Craig since 1880",

subtitle = "(Only includes years where more than 5 babies are named Craig)")

Chart

Description automatically generated with medium confidence

Craig was most popular in 1960

1. How many babies have been named Abcde (no joke)?

names.Abcde<-babynames %>%

filter(name=="Abcde") %>%

summarize(total\_n=sum(n))

names.Abcde

total\_n

<int>

1 373