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?

```
problem1a_3_1 = babynames %>%
+    group_by(name) %>%
+    summarise(NamesTotal = sum(n))
> arrange(problem1a_3_1, desc(NamesTotal))
# A tibble: 97,310 x 2
   name      NamesTotal
   <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
```

```
problem1b_3_1 = babynames %>%
+    filter(year >= 1990) %>%
+    group_by(name) %>%
+    summarise(NamesTotal = sum(n))
> arrange(problem1b_3_1, desc(NamesTotal))
# A tibble: 77,092 x 2
   name         NamesTotal
   <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
```

2. 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!

```
problem2a_3_1 = babynames %>%
+   filter(year >= 1990, sex == "F") %>%
+   group_by(name) %>%
+   summarise(NamesTotal = sum(n))
> arrange(problem2a_3_1, desc(NamesTotal))
# A tibble: 51,183 x 2
   name        NamesTotal
   <chr>            <int>
 1 Emily           561299
 2 Ashley          467686
 3 Jessica         409833
 4 Samantha        407773
 5 Emma            397978
 6 Sarah           388002
 7 Elizabeth       383232
 8 Olivia          379362
 9 Madison         369832
10 Hannah          367423
```

```
problem2b_3_1 = babynames %>%
+   filter(year >= 1990, sex == "M") %>%
+   group_by(name) %>%
+   summarise(NamesTotal = sum(n))
> arrange(problem2b_3_1, desc(NamesTotal))
# A tibble: 33,454 x 2
   name         NamesTotal
   <chr>             <int>
 1 Michael          835096
 2 Jacob            712218
 3 Matthew          678528
 4 Joshua           652762
 5 Christopher      650925
 6 Daniel           587569
 7 Andrew           565863
 8 Joseph           551741
 9 William          543294
10 David            530892
```

3. 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?

```
problem3a_3_1 = babynames %>%
+   filter(name == "Jacob") %>%
+   group_by(name) %>%
+   summarise(NamesTotal = sum(n))
> problem3a_3_1
# A tibble: 1 x 2
  name   NamesTotal
  <chr>       <int>
1 Jacob      921726
```

```
problem3b_3_1 = babynames %>%
+    filter(name == "Jacob") %>%
+    group_by(sex) %>%
+    summarise(NamesTotal = sum(n))
> problem3b_3_1
# A tibble: 2 x 2
  sex    NamesTotal
  <chr>       <int>
1 F            2245
2 M          919481




problem3c_3_1 = babynames %>%
+    filter(name == "Jacob", year >= 2002) %>%
+    group_by(name) %>%
+    summarise(NamesTotal = sum(n))
> problem3c_3_1
# A tibble: 1 x 2
  name   NamesTotal
  <chr>       <int>
1 Jacob      347466
```

4. 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=problem3a_3_1)+
  geom_point(mapping = aes(x=year, y = n), color = "violetred1")+
  labs(title = "Popularity of the name Jacob")

This was working at one point then I changed something and it stopped working so idk

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

```
problem5_3_1 = babynames %>%
+    filter(name == "Abcde") %>%
+    group_by(name) %>%
+    summarise(NamesTotal = sum(n))
> problem5_3_1
# A tibble: 1 x 2
  name   NamesTotal
  <chr>       <int>
1 Abcde         373
```