miaR\_babynames

In this blog, I will be working with … INTRODUCE data

The first step is to load necessary packages and the data.

library(tidyverse)  
library(gridExtra)  
  
load("/Users/mia/Downloads/babynames.rda")  
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

Now for a little bit of data carpentry. First, I narrowed down the criteria that I wanted to work with by using filter(). I chose to work with babynames from the year 2002 — my birth year. After this, I am essentially creating a separate M and F column instead of having a sex column, where each name has an M and F row.

df <- babynames %>%   
 filter(year == 2002) %>%  
 group\_by(name, sex) %>%  
 summarize(total = sum(n)) %>%  
 pivot\_wider(  
 names\_from = sex, values\_from = total, values\_fill = 0)

## `summarise()` has grouped output by 'name'. You can override using the `.groups` argument.

I am changing M and F to nM and nF so R does not confuse F with FALSE. Next, I am adding to the dataset by creating a total column that includes the number of births (both M and F) for every name.

colnames(df) <- c("name", "nM", "nF")  
df <- df %>% mutate(ratio = nM/nF, total = nM + nF)

Creating a function to graph the sex distribution of names over time

gg\_name <- function(idname = NULL, color1 = NULL, color2 = NULL) {  
 p <- ggplot(babynames %>% filter(name==idname),  
 aes(x=year, y=n, color=sex)) +  
 geom\_line(color1="hotpink", color1="darkblue") +  
 ggtitle(idname) +  
 scale\_color\_manual(values=c(color1,color2))  
 return(p)  
}  
  
grid.arrange(  
 gg\_name("Mia", color1 = "hotpink", color2 = "darkblue"),  
 gg\_name("James", color1 = "hotpink", color2 = "darkblue"),  
 gg\_name("Jolie", color1 = "hotpink", color2 = "darkblue"),  
 gg\_name("Tyler", color1 = "hotpink", color2 = "darkblue"),  
 gg\_name("Maddie", color1 = "hotpink", color2 = "darkblue")  
 )

## Warning: Duplicated aesthetics after name standardisation: colour1

## Warning: Ignoring unknown parameters: colour1

## Warning: Duplicated aesthetics after name standardisation: colour1

## Warning: Ignoring unknown parameters: colour1

## Warning: Duplicated aesthetics after name standardisation: colour1

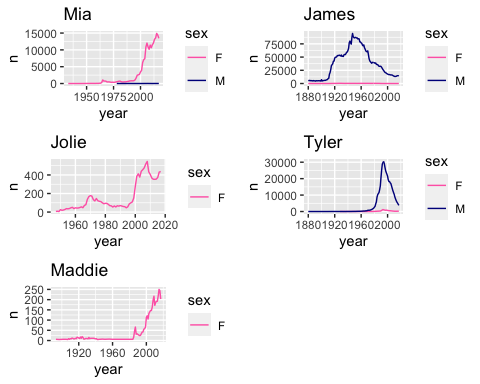
## Warning: Ignoring unknown parameters: colour1

## Warning: Duplicated aesthetics after name standardisation: colour1

## Warning: Ignoring unknown parameters: colour1

## Warning: Duplicated aesthetics after name standardisation: colour1

## Warning: Ignoring unknown parameters: colour1



Trying to graph percentages