1261 HW 5

Sam Sadow

# 1. Youngest Male Names

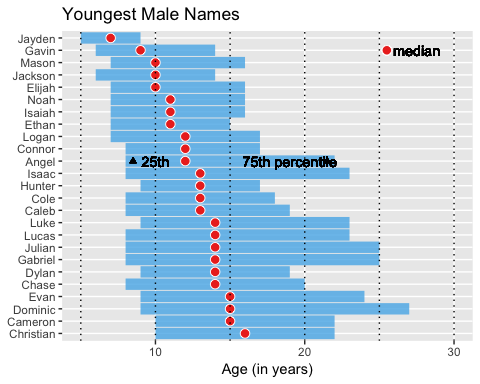
library(Hmisc)  
library(mdsr)  
library(babynames)  
BabynamesDist <- make\_babynames\_dist()

com\_men <-BabynamesDist %>%   
 filter(sex == "M") %>%   
 group\_by(name) %>%   
 mutate(N=n()) %>%   
 filter(N>=4) %>%   
 summarise(  
 est\_num\_alive = sum(est\_alive\_today),  
 q1\_age = wtd.quantile(age\_today, est\_alive\_today, probs = 0.25),  
 med\_age = wtd.quantile(age\_today, est\_alive\_today, probs = 0.5),  
 q3\_age = wtd.quantile(age\_today, est\_alive\_today, probs = 0.75)  
 ) %>%   
 filter(est\_num\_alive >= 100000) %>%   
 arrange(med\_age) %>%   
 head(25)  
com\_men[1:10,]

## # A tibble: 10 x 5  
## name est\_num\_alive q1\_age med\_age q3\_age  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Jayden 107980. 5 7 9  
## 2 Gavin 123936. 6 9 14  
## 3 Elijah 173815. 7 10 16  
## 4 Jackson 137370. 6 10 14  
## 5 Mason 146915. 7 10 16  
## 6 Ethan 301045. 7 11 15  
## 7 Isaiah 147352. 7 11 16  
## 8 Noah 235671. 7 11 16  
## 9 Angel 175908. 8 12 22  
## 10 Connor 153242. 8 12 17

m\_plot <- ggplot(  
 data = com\_men,  
 aes(x=reorder(name, -med\_age), y=med\_age)  
) +  
 ylim(5,30) +  
 xlab(NULL) +  
 ylab("Age (in years)") +  
 ggtitle("Youngest Male Names")

m\_plot +  
 geom\_linerange(  
 aes(ymin = q1\_age, ymax = q3\_age),  
 color = "#56B4E9",  
 size = 4,  
 alpha = 0.8  
 ) +  
 geom\_point(fill = "#ed3324", colour = "white", size = 3, shape = 21) +  
 geom\_point(aes(y=25.5,x=24), fill = "#ed3324", colour = "white", size = 3, shape = 21) +  
 geom\_text(aes(y=27.5,x=24),label="median",cex=3.8)+  
 geom\_text(aes(y=10,x=15),label="25th",cex=3.8)+  
 geom\_text(aes(y=19,x=15),label="75th percentile",cex=3.8)+  
 geom\_point(aes(y=8.5,x=15),shape=17)+  
 geom\_point(aes(y=21.5,x=15),shape=17)+  
 geom\_hline(data = com\_men, yintercept =c(5,10,15,20,25,30), linetype=3) +  
 coord\_flip()



# 2 Ten Closet Ratio Years

?mutate

## Help on topic 'mutate' was found in the following packages:  
##   
## Package Library  
## plyr /Library/Frameworks/R.framework/Versions/3.5/Resources/library  
## dplyr /Library/Frameworks/R.framework/Versions/3.5/Resources/library  
##   
##   
## Using the first match ...

library(dplyr)  
  
babynames2 <- babynames %>%   
 filter(name == "Jackie") %>%   
 group\_by(year) %>%   
 summarise(  
 N=n(),  
 total = sum(n),  
 boys = sum(ifelse(sex=="M", n, 0))  
 ) %>%   
 mutate(ratio = abs(0.5-(boys/total)))  
#head(babynames2 %>% arrange(ratio), 10)  
head(babynames2 %>% arrange(ratio), 10)$year

## [1] 2006 1997 1925 1999 2015 1956 1927 1926 2003 2002

# 3

babyname1<- babynames %>%   
 group\_by(year) %>%   
 summarise(  
 N=n(),  
 total = sum(n)  
 ) %>%   
 arrange(desc(total)) %>%   
 head(1)  
babyname1$year

## [1] 1957

# 4

babynames4 <- babynames %>% #filter(name=="Jackie") %>%   
 group\_by(name) %>%   
 mutate(early = min(year), late = max(year))  
babynames4[c(1, 305:310, 900:950),]

## # A tibble: 58 x 7  
## # Groups: name [58]  
## year sex name n prop early late  
## <dbl> <chr> <chr> <int> <dbl> <dbl> <dbl>  
## 1 1880 F Mary 7065 0.0724 1880 2015  
## 2 1880 F Linda 27 0.000277 1880 2015  
## 3 1880 F Octavia 27 0.000277 1880 2015  
## 4 1880 F Sudie 27 0.000277 1880 1984  
## 5 1880 F Zula 27 0.000277 1880 2014  
## 6 1880 F Adella 26 0.000266 1880 2015  
## 7 1880 F Alpha 26 0.000266 1880 2015  
## 8 1880 F Manervia 5 0.0000512 1880 1948  
## 9 1880 F Manuela 5 0.0000512 1880 2015  
## 10 1880 F Margarett 5 0.0000512 1880 2015  
## # ... with 48 more rows

# 5

babyname <- head(babynames4 %>%   
 filter(prop >= 0.01 & name == "Olivia") %>%   
 arrange(desc(early)),10)#$name  
babyname

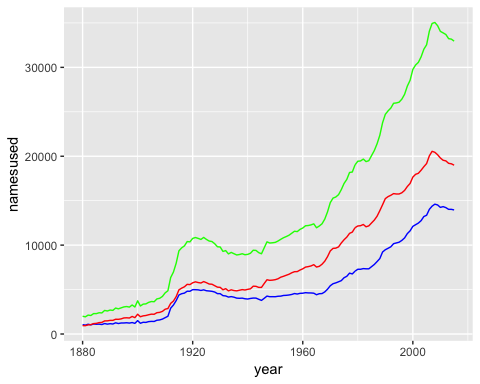
## # A tibble: 2 x 7  
## # Groups: name [1]  
## year sex name n prop early late  
## <dbl> <chr> <chr> <int> <dbl> <dbl> <dbl>  
## 1 2014 F Olivia 19761 0.0101 1880 2015  
## 2 2015 F Olivia 19553 0.0101 1880 2015

# 6

baybnames5 <- babynames %>%   
 group\_by(year) %>%   
 summarise(namesused = n(), namesusedM=sum(ifelse(sex=="M",1,0)), namesusedF=sum(ifelse(sex=="F",1,0)))  
  
baybnames5

## # A tibble: 136 x 4  
## year namesused namesusedM namesusedF  
## <dbl> <int> <dbl> <dbl>  
## 1 1880 2000 1058 942  
## 2 1881 1935 997 938  
## 3 1882 2127 1099 1028  
## 4 1883 2084 1030 1054  
## 5 1884 2297 1125 1172  
## 6 1885 2294 1097 1197  
## 7 1886 2392 1110 1282  
## 8 1887 2373 1067 1306  
## 9 1888 2651 1177 1474  
## 10 1889 2590 1111 1479  
## # ... with 126 more rows

tryit <- ggplot(baybnames5, aes(x= year)) +   
 geom\_line(color = "green", aes(y=namesused)) +  
 geom\_line(color = "blue", aes(y=namesusedM)) +  
 geom\_line(color = "red", aes(y=namesusedF))  
   
  
tryit

 Green is number of names used for all names.

Red is number of names used for female names.

Blue is number of names used for male names.

# 7

nineties <- babynames %>%   
 filter(year >= 1990 & year < 2000) %>%   
 group\_by(name) %>%   
 summarise(total = sum(n)) %>%   
 arrange(desc(total))   
  
nineties

## # A tibble: 45,921 x 2  
## name total  
## <chr> <int>  
## 1 Michael 464221  
## 2 Christopher 361234  
## 3 Matthew 352320  
## 4 Joshua 330029  
## 5 Jessica 303837  
## 6 Ashley 303113  
## 7 Jacob 298871  
## 8 Nicholas 275899  
## 9 Andrew 273475  
## 10 Daniel 273274  
## # ... with 45,911 more rows

The most common name in the nineties for males was Joshua and the most popular name for girls was Jessica.