Homework

Hyeonho Lee

2018년 8월 17일

(homework) Re-create the data graphic, "Youngest Male Names" (again, your result will be different from the chart below). You can recycle some of the codes above. In particular, the youngest men names are given by the ascending order or median_age. Your chart should be restricted to birth names given to at least 100,000 male Americans since 1900. Use filter() to filter names with at least 100,000 est_num_alive. Can you make the color of the bars Carolina blue?

```
# library(Hmisc)
# library(mdsr)
# library(babynames)
# library(dplyr)
```

```
BabynamesDist <- make_babynames_dist()
head(BabynamesDist, 2)
```

```
## # A tibble: 2 x 9
##    year sex   name      n   prop alive_prob count_thousands age_today
##   <dbl> <chr> <chr> <int>  <dbl>      <dbl>           <dbl>     <dbl>
## 1  1900 F     Mary  16707 0.0526          0            16.7       114
## 2  1900 F     Helen  6343 0.0200          0             6.34      114
## # ... with 1 more variable: est_alive_today <dbl>
```

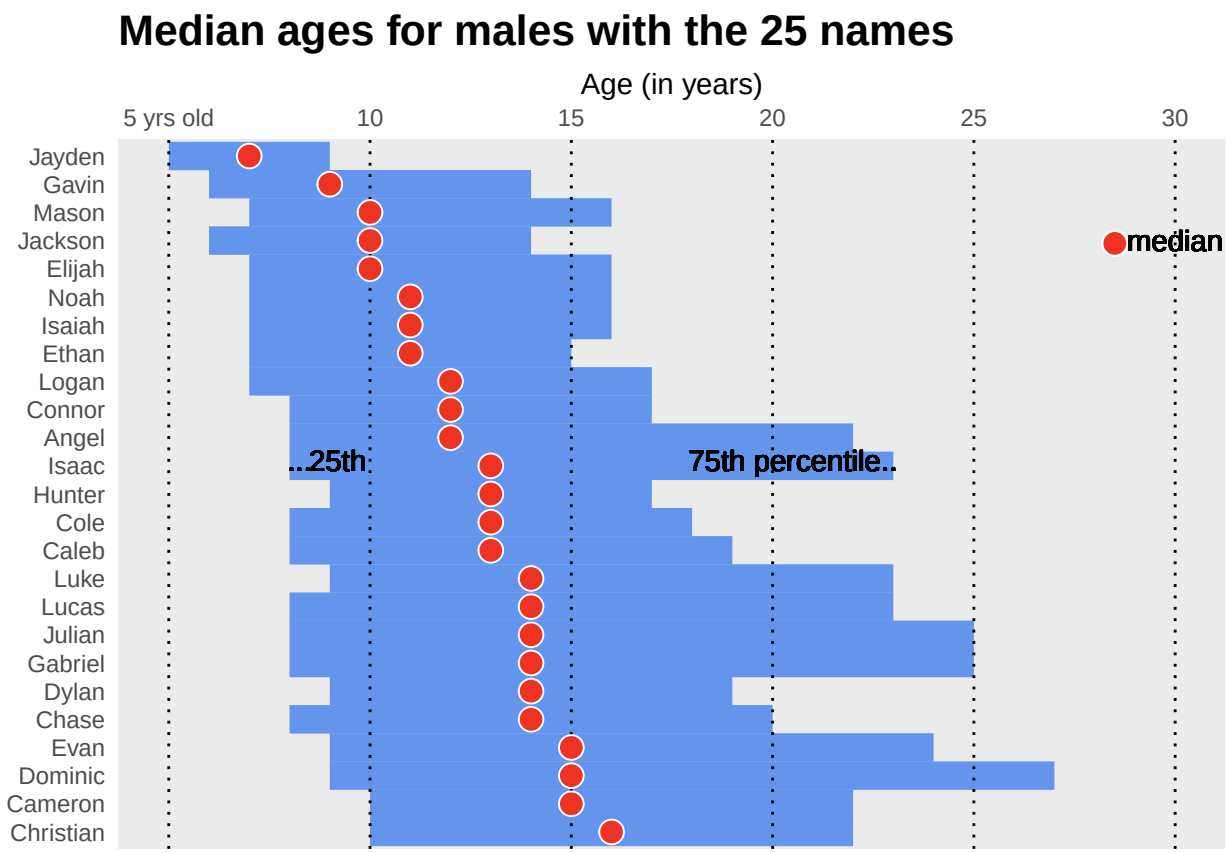
```
com_fem <- BabynamesDist %>%
  filter(sex == "M") %>% filter(est_alive_today!=0) %>%
  group_by(name) %>%
  summarise(
    N = n(), est_num_alive = sum(est_alive_today),
    q1_age = wtd.quantile(age_today, est_alive_today, probs = 0.25),
    median_age = wtd.quantile(age_today, est_alive_today, probs = 0.5),
    q3_age = wtd.quantile(age_today, est_alive_today, probs = 0.75)) %>%
  filter(est_num_alive >= 100000) %>%
  arrange(median_age) %>%
  head(25)
```

```
com_fem
```

```
## # A tibble: 25 x 6
##    name        N est_num_alive q1_age median_age q3_age
##    <chr>   <int>         <dbl>  <dbl>      <dbl>  <dbl>
##  1 Jayden     31       107980.      5          7      9
##  2 Gavin      96       123936.      6          9     14
##  3 Elijah    110       173815.      7         10     16
##  4 Jackson   110       137370.      6         10     14
##  5 Mason     110       146915.      7         10     16
##  6 Ethan     103       301045.      7         11     15
##  7 Isaiah    110       147352.      7         11     16
##  8 Noah      110       235671.      7         11     16
##  9 Angel     104       175908.      8         12     22
## 10 Connor     80       153242.      8         12     17
## # ... with 15 more rows
```

```r
w_plot <- ggplot(
  data = com_fem,
  aes(x = reorder(name, -median_age),  y = median_age)) +
  xlab(NULL) +
  ylab("Age (in years)") +
  ggtitle("Median ages for males with the 25 names")
```

```r
w_plot +
  geom_linerange(
    aes(ymin = q1_age, ymax = q3_age),
    color = "#6495ED",
    size = 5,
    alpha = 1) +
  geom_hline(aes(yintercept = 5), linetype = 'dotted') +
  geom_hline(aes(yintercept = 10), linetype = 'dotted') +
  geom_hline(aes(yintercept = 15), linetype = 'dotted') +
  geom_hline(aes(yintercept = 20), linetype = 'dotted') +
  geom_hline(aes(yintercept = 25), linetype = 'dotted') +
  geom_hline(aes(yintercept = 30), linetype = 'dotted') +
  geom_point(fill = "#ed3324", colour = "white", size = 4, shape = 21) +
  geom_point(aes(y = 28.5, x = 21.9), fill = "#ed3324", colour = "white", size = 4, shape = 21) +
  #theme(panel.background = element_rect(fill='gray')) +
  geom_text(aes(y = 8.25, x = 14.2, label = ' ')) +
  geom_text(aes(y = 9.2, x = 14.2, label = '25th')) +
  geom_text(aes(y = 22.8, x = 14.2, label = ' ')) +
  geom_text(aes(y = 20.3, x = 14.2, label = '75th percentile')) +
  geom_text(aes(y = 30, x = 22, label = 'median')) +
  scale_y_continuous(position = "bottom", breaks = c(5, 10, 15, 20, 25, 30),
                     labels = c("5 yrs old", "10", "15", "20", "25", "30")) +
  theme(plot.title = element_text(face = 'bold', size = 16),
        panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        axis.ticks = element_blank(), plot.caption = element_text(size = 10)) +
  coord_flip()
```


Median ages for males with the 25 names
Age (in years)
5 yrs old
10
15
20
25
30
Jayden
Gavin
Mason
Jackson
Elijah
Noah
Isaiah
Ethan
Logan
Connor
Angel
Isaac
Hunter
Cole
Caleb
Luke
Lucas
Julian
Gabriel
Dylan
Chase
Evan
Dominic
Cameron
Christian
median
...25th
75th percentile..