Visualization-HW

Seoyun Yang

2018년 8월 21일

load packages and data

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

BabynamesDist <- make_babynames_dist()
dim(BabynamesDist)
```

```
## [1] 1639490       9
```

```
names(BabynamesDist)
```

```
## [1] "year"            "sex"             "name"            "n"
## [5] "prop"            "alive_prob"      "count_thousands" "age_today"
## [9] "est_alive_today"
```

group of male names whose est_num_alive is under 100,000

```
not_mybaby <- BabynamesDist %>% filter(sex=="M") %>% group_by(name) %>%
  summarise(est_num_alive = sum(est_alive_today)) %>%
  filter(est_num_alive<100000) %>% select(name)
```

select data of male names except not_mybaby

```
mybaby <- BabynamesDist %>%
  anti_join(not_mybaby, by="name") %>%
  filter(sex == "M") %>%
  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)) %>%
  arrange(median_age) %>%
  head(26)
```

draw a plot

```
b.plot <- ggplot(data=mybaby, aes(x = reorder(name, -median_age), y=median_age)
                 ) + xlab(NULL) + ylab("Age (in years)") +
```

```
  ggtitle("Youngest Male Names",
          subtitle = "By estimated median age for Americans since 1900")

b.plot + geom_linerange(
  aes(ymin = q1_age, ymax = q3_age),
  color = "#76bde0",
  size=5, alpha=0.7) +
  geom_point(fill = "#ed3324", colour = "white", size = 4, shape = 21) +
  annotate(geom = "text", x=15.1, y=9.5 , label=" 25th", colour="black", size=3.8) +
  annotate(geom = "text", x=15.1, y=19.5, label="75th percentile  ", colour="black",
           size=3.8) + theme(legend.position = "top") + coord_flip()
```