dplyr_practice

Grace Becker

9/26/2019

```
library('tidyverse')
```

```
## -- Attaching packages ---------------------------------------------------------------------------------
## v ggplot2 3.2.1     v purrr   0.3.0
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.5.2

## Warning: package 'tibble' was built under R version 3.5.2

## Warning: package 'tidyr' was built under R version 3.5.2

## Warning: package 'purrr' was built under R version 3.5.2

## Warning: package 'dplyr' was built under R version 3.5.2

## Warning: package 'stringr' was built under R version 3.5.2

## Warning: package 'forcats' was built under R version 3.5.2

## -- Conflicts ------------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df.marathon = read.csv('https://raw.githubusercontent.com/chuckpr/BIOS512/master/in-class-exercises/data
df.marathon %>% head
```

```
##   year                        name       country       time race.time    MF
## 1 2019           Lawrence Cherono         Kenya  2H 7M 57S  2.132500 Men's
## 2 2018              Yuki Kawauchi         Japan 2H 15M 58S  2.266111 Men's
## 3 2017             Geoffrey Kirui         Kenya  2H 9M 37S  2.160278 Men's
## 4 2016               Lemi Berhanu      Ethiopia 2H 12M 45S  2.212500 Men's
## 5 2015              Lelisa Desisa      Ethiopia  2H 9M 17S  2.154722 Men's
## 6 2014 Mebrahtom "Meb" Keflezighi United States  2H 8M 37S  2.143611 Men's
##   country.simple
## 1          Kenya
## 2          Japan
## 3          Kenya
## 4       Ethiopia
## 5       Ethiopia
## 6  United States
```

Filter the parathon data to just the runners from Kenya.

```
df.marathon %>% filter(country == 'Kenya')
```

Select just the columns that represent runner names and country from the marathon data.

```
df.marathon %>% select(name, country)
```

Show best race time sfor US and Kenyan women.

```
df.marathon %>%
        filter(MF == "Women's", country %in% c("United States", "Kenya")) %>%
        group_by(country) %>%
        summarize(best.race.time = min(race.time))
```

```
## # A tibble: 2 x 2
##   country       best.race.time
##   <fct>                  <dbl>
## 1 Kenya                   2.35
## 2 United States           2.38
```

Show the best women's race time in the data and the country of the runners.

```
df.marathon %>%
        filter(MF == "Women's") %>%
        select(country, race.time) %>%
        arrange(race.time) %>%
        head()
```

```
##         country race.time
## 1      Ethiopia  2.333056
## 2         Kenya  2.345278
## 3       Germany  2.362500
## 4         Kenya  2.364444
## 5         Kenya  2.376667
## 6 United States  2.378611
```

What's the best race time ever for a US man?

```
df.marathon %>%
        filter(MF == "Men's", country == "United States") %>%
        select(name, race.time, time) %>%
        arrange(race.time) %>%
        head()
```

```
##                         name race.time       time
## 1 Mebrahtom "Meb" Keflezighi  2.143611  2H 8M 37S
## 2            Alberto Salazar  2.147778  2H 8M 52S
## 3              Greg A. Meyer  2.150000   2H 9M 0S
## 4               Bill Rodgers  2.157500  2H 9M 27S
## 5               Bill Rodgers  2.165278  2H 9M 55S
## 6               Bill Rodgers  2.170278 2H 10M 13S
```

How does this compare to the best race time in the data overall?

```
df.marathon %>%
        select(name, race.time, time) %>%
        arrange(race.time) %>%
        head()
```

```
##                         name race.time      time
## 1             Geoffrey Mutai  2.050556  2H 3M 2S
## 2  Robert Kiprono Cheruiyot  2.097778 2H 5M 52S
## 3 Robert Kipkoech Cheruiyot  2.120556 2H 7M 14S
## 4               Cosmas Ndeti  2.120833 2H 7M 15S
## 5                Moses Tanui  2.126111 2H 7M 34S
## 6 Robert Kipkoech Cheruiyot  2.129444 2H 7M 46S
```

The best race time for a US man is 2 hours 8 mins and 37 seconds by Meb Keflezighi. The best race time overall is 2 hours 3 mins and 2 seconds by Geoffrey Mutai.

For all countries with at least 5 winners in the men's category, which country has the best median race time in the men's category?

```
df.marathon %>%
        filter(MF == "Men's") %>%
        group_by(country) %>%
        mutate(N.winners = n()) %>%
        filter(N.winners >= 5) %>%
        summarize(median.race.time = median(race.time)) %>%
        arrange(median.race.time) %>%
        head()
```

```
## # A tibble: 6 x 2
##   country       median.race.time
##   <fct>                    <dbl>
## 1 Kenya                     2.16
## 2 Ethiopia                  2.16
## 3 Japan                     2.28
## 4 Finland                   2.35
## 5 United States             2.48
## 6 Canada                    2.50
```