

# olympics

```
library(tidyverse)
library(skimr)

theme_set(theme_light())

data <- tidyuesdayR::tt_load('2021-07-27')

##
## Downloading file 1 of 2: `olympics.csv`
## Downloading file 2 of 2: `regions.csv`

olympics <- data$olympics

rm(data)

skim(olympics)
```

Table 1: Data summary

Name	olympics
Number of rows	271116
Number of columns	15
Column type frequency:	
character	10
numeric	5
Group variables	None

## Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
name	0	1.00	2	108	0	134731	0
sex	0	1.00	1	1	0	2	0
team	0	1.00	2	47	0	1184	0
noc	0	1.00	3	3	0	230	0
games	0	1.00	11	11	0	51	0
season	0	1.00	6	6	0	2	0
city	0	1.00	4	22	0	42	0
sport	0	1.00	4	25	0	66	0
event	0	1.00	15	85	0	765	0
medal	231333	0.15	4	6	0	3	0

## Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
id	0	1.00	68248.95	39022.29	1	34643	68205	102097.2	135571	
age	9474	0.97	25.56	6.39	10	21	24	28.0	97	
height	60171	0.78	175.34	10.52	127	168	175	183.0	226	
weight	62875	0.77	70.70	14.35	25	60	70	79.0	214	
year	0	1.00	1978.38	29.88	1896	1960	1988	2002.0	2016	

```
olympics %>%
  summarise(years = unique(year)) %>%
  arrange(years) %>%
  pull(years)
```

```
## [1] 1896 1900 1904 1906 1908 1912 1920 1924 1928 1932 1936 1948 1952 1956 1960
## [16] 1964 1968 1972 1976 1980 1984 1988 1992 1994 1996 1998 2000 2002 2004 2006
## [31] 2008 2010 2012 2014 2016
```

```
summer_olympics <- olympics %>%
  filter(season == "Summer")
```

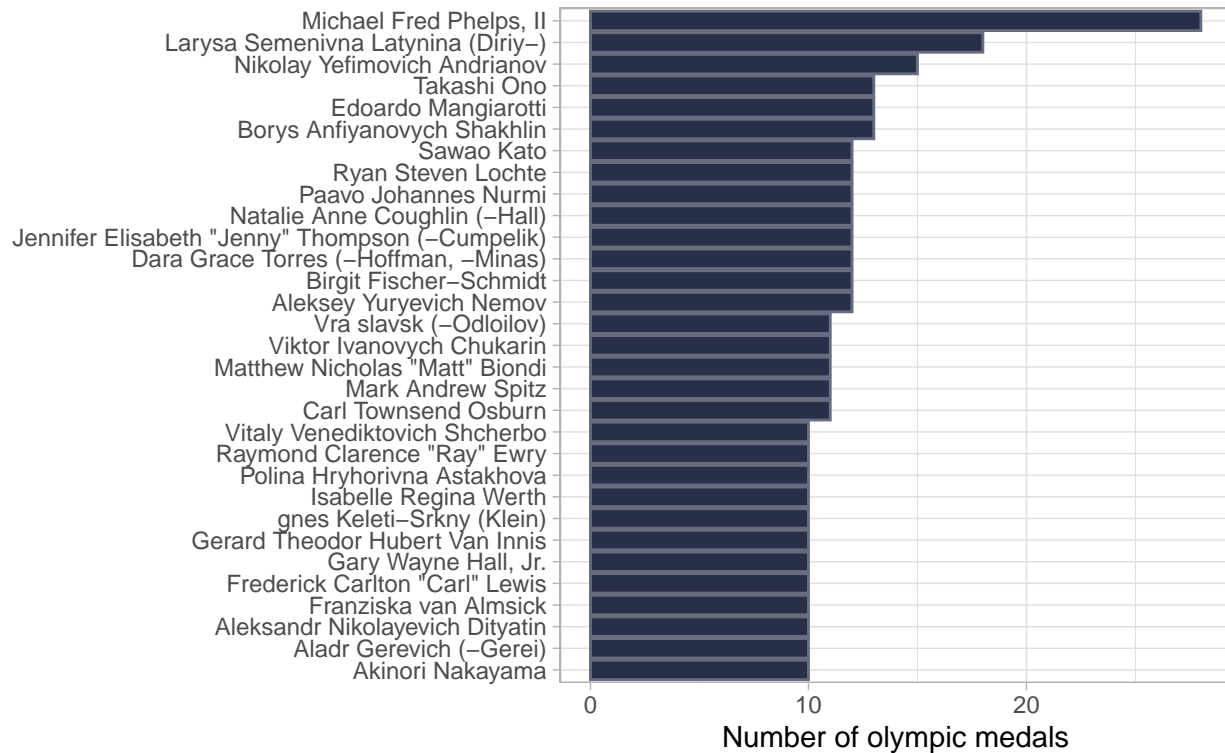
## Athletes with the greatest number of medals

```
most_medals <- summer_olympics %>%
  filter(!is.na(medal)) %>%
  count(name, sort = TRUE) %>%
  slice_max(n, prop = 0.001)

most_medals %>%
  ggplot(aes(reorder(name, n), n)) +
  geom_col(fill = "#272F49", color = "#676D7F") +
  coord_flip() +
  labs(x = NULL, y = "Number of olympic medals",
       title = "Top 0.1% of olympic athletes",
       subtitle = "Ordered by total number of medals")
```

## Top 0.1% of olympic athletes

Ordered by total number of medals



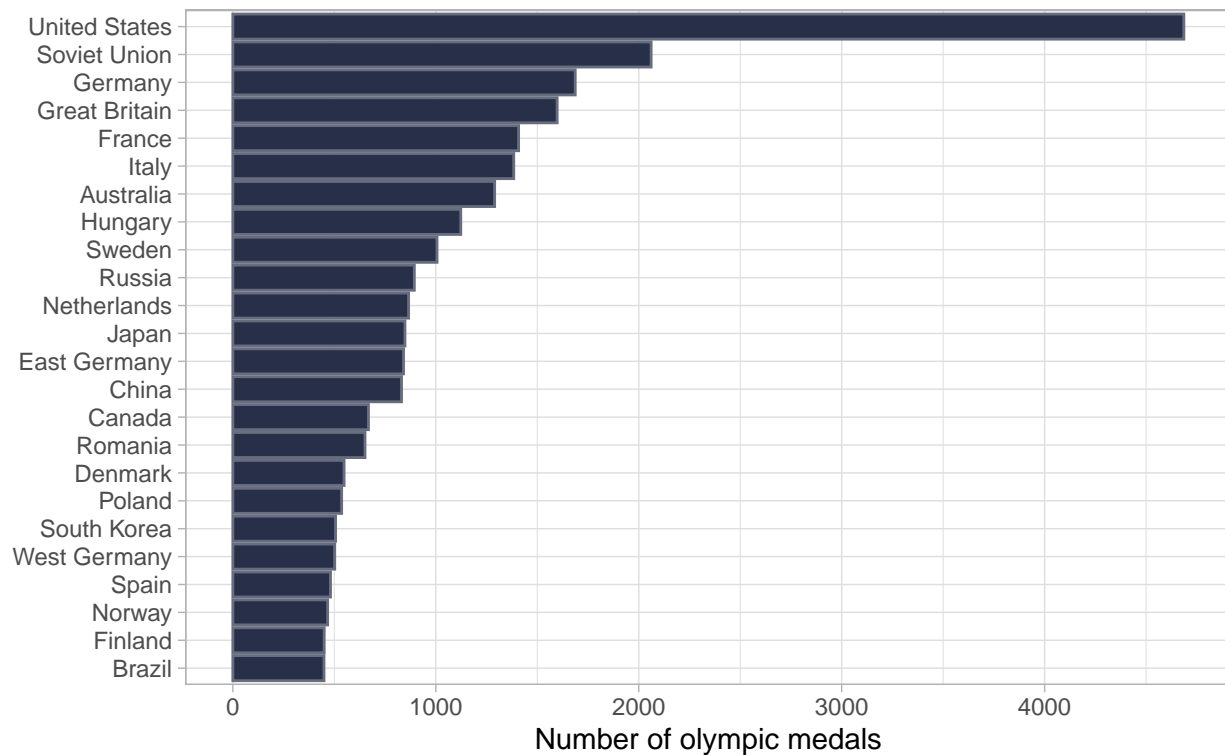
## Countries with the greatest number of medals

```
countries_most_medals <- summer_olympics %>%
  filter(!is.na(medal)) %>%
  count(team, sort = TRUE) %>%
  slice_max(n, prop = 0.05)

countries_most_medals %>%
  ggplot(aes(reorder(team, n), n)) +
  geom_col(fill = "#272F49", color = "#676D7F") +
  coord_flip() +
  labs(x = NULL, y = "Number of olympic medals",
       title = "Top 5% of olympic teams in history",
       subtitle = "Ordered by total number of medals")
```

## Top 5% of olympic teams in history

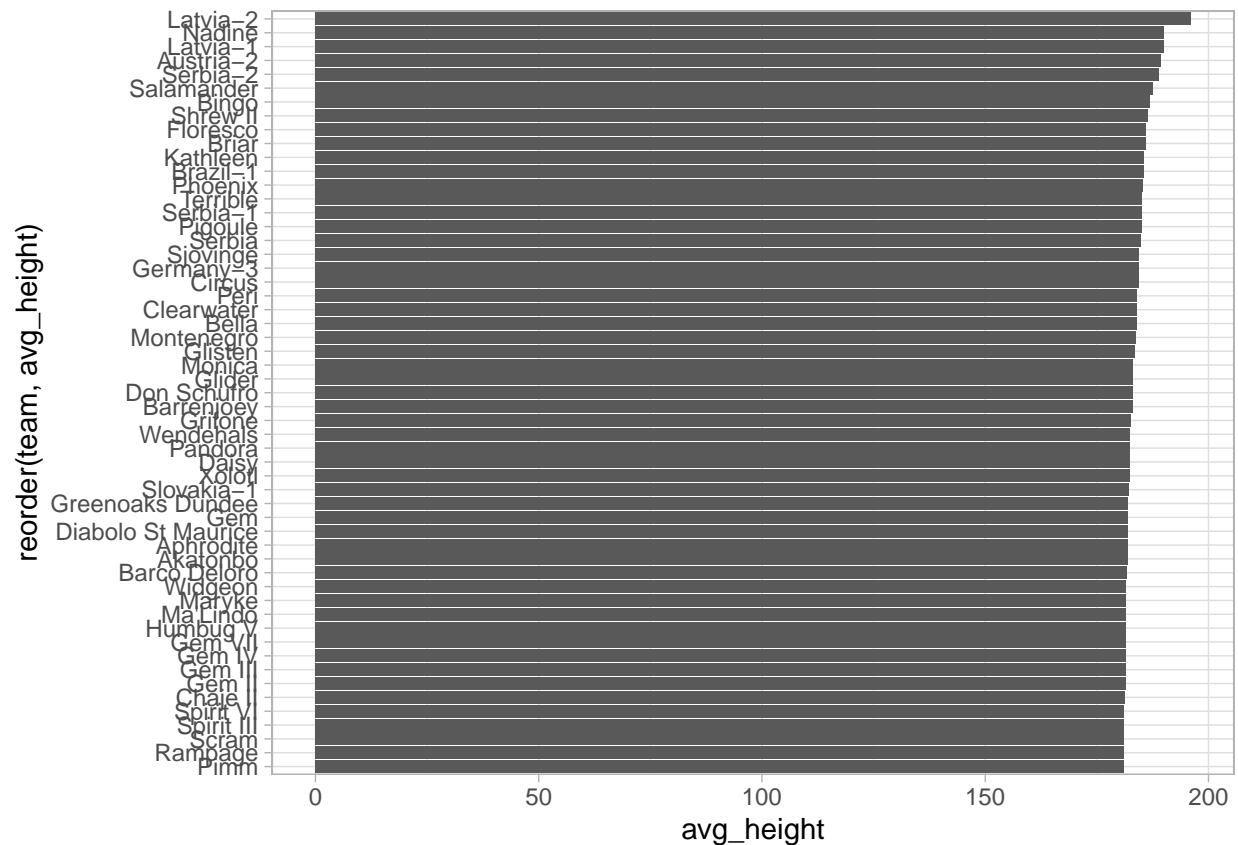
Ordered by total number of medals



## Average height by team

```
height_team <- summer_olympics %>%
  group_by(team) %>%
  summarise(avg_height = mean(height)) %>%
  drop_na(avg_height) %>%
  arrange(desc(avg_height)) %>%
  slice_max(avg_height, prop = 0.2)

height_team %>%
  ggplot(aes(reorder(team, avg_height), avg_height)) +
  geom_col() +
  coord_flip()
```

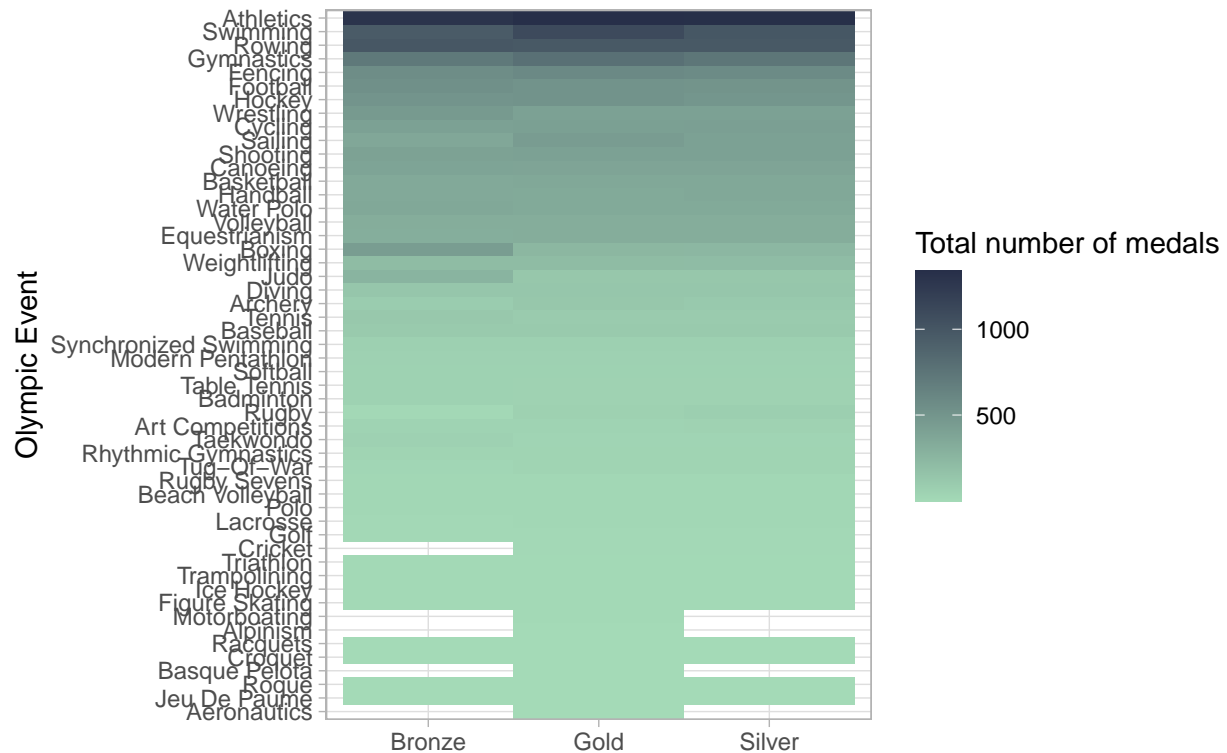


## Top 5% sports by number of delivered medals, in history

```
events_most_medals <- summer_olympics %>%
  drop_na(sport, medal) %>%
  group_by(sport) %>%
  count(medal, name = "total_medals")

events_most_medals %>%
  ggplot(aes(medal, reorder(sport, total_medals), fill = total_medals)) +
  geom_tile() +
  scale_fill_gradient(low = "#a4dab7", high = "#272f49") +
  labs(title = "Athletics is the sport that delivered the most medals",
       subtitle = "Considering all events in history", x = NULL,
       y = "Olympic Event", fill = "Total number of medals")
```

Athletics is the sport that delivered the most medals  
Considering all events in history



Sports ordered by average height, from low to high

```
avg_height <- summer_olympics %>%
  drop_na(height) %>%
  group_by(sport, sex) %>%
  mutate(sex_sport_avg_height = mean(height))

avg_height <- avg_height %>%
  ungroup() %>%
  group_by(sport) %>%
  mutate(sport_avg_height = mean(height)) %>%
  select(sex, sport, sex_sport_avg_height,
         sport_avg_height) %>%
  distinct()

both_sexes_avg_height <- avg_height %>%
  select(sport, sex_sport_avg_height = sport_avg_height) %>%
  mutate(sex = "Both") %>%
  distinct()

avg_height <- avg_height %>%
  select(-sport_avg_height) %>%
  bind_rows(both_sexes_avg_height) %>%
```

```

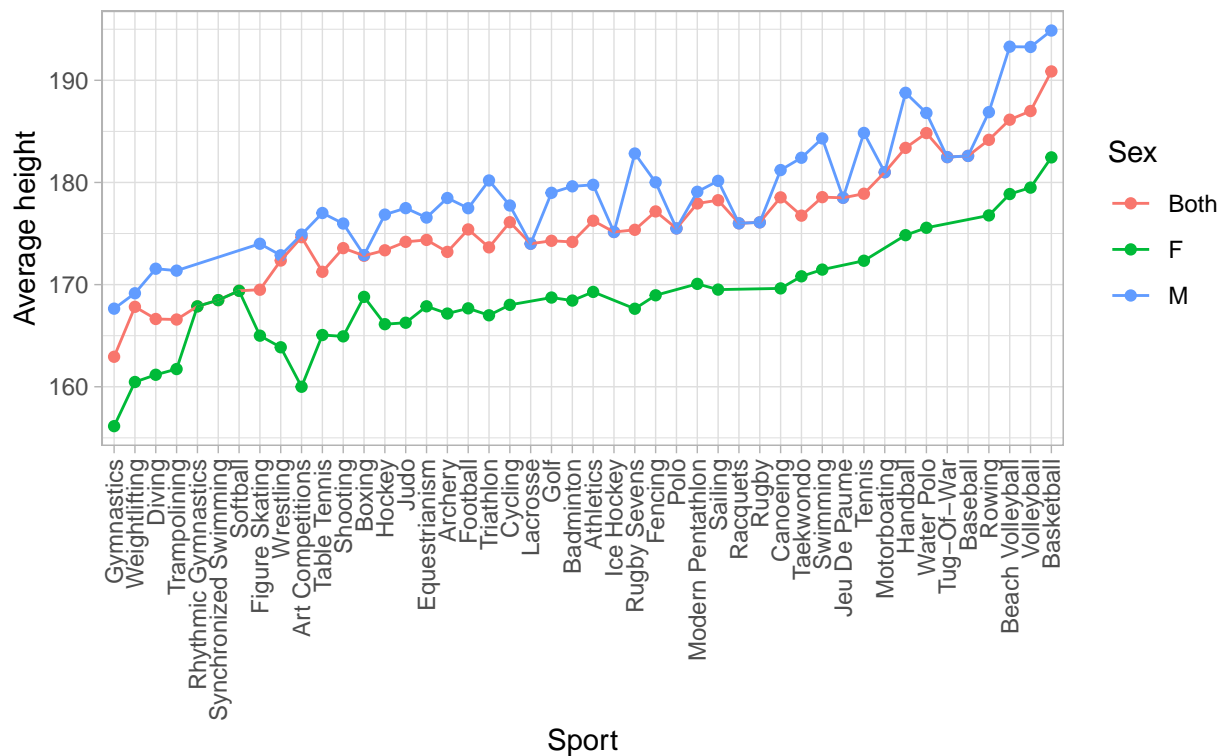
arrange(sport, sex)

height_plot <- avg_height %>%
  ggplot() +
  geom_point(aes(reorder(sport, sex_sport_avg_height), sex_sport_avg_height,
    color = sex, group = sex)) +
  geom_line(aes(reorder(sport, sex_sport_avg_height), sex_sport_avg_height,
    color = sex, group = sex))

height_plot +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  labs(x = "Sport", y = "Average height", color = "Sex",
    title = "Gymnastics and Basketball are on the extremes of average athlete height",
    subtitle = "Considering all summer olympic sports")

```

## Gymnastics and Basketball are on the extremes of average athlete height Considering all summer olympic sports



## Sports ordered by average weight, from low to high

```

avg_weight <- summer_olympics %>%
  drop_na(weight) %>%
  group_by(sport, sex) %>%
  mutate(sex_sport_avg_weight = mean(weight))

```

```

avg_weight <- avg_weight %>%
  ungroup() %>%
  group_by(sport) %>%
  mutate(sport_avg_weight = mean(weight)) %>%
  select(sex, sport, sex_sport_avg_weight,
         sport_avg_weight) %>%
  distinct()

both_sexes_avg_weight <- avg_weight %>%
  select(sport, sex_sport_avg_weight = sport_avg_weight) %>%
  mutate(sex = "Both") %>%
  distinct()

avg_weight <- avg_weight %>%
  select(-sport_avg_weight) %>%
  bind_rows(both_sexes_avg_weight) %>%
  arrange(sport, sex)

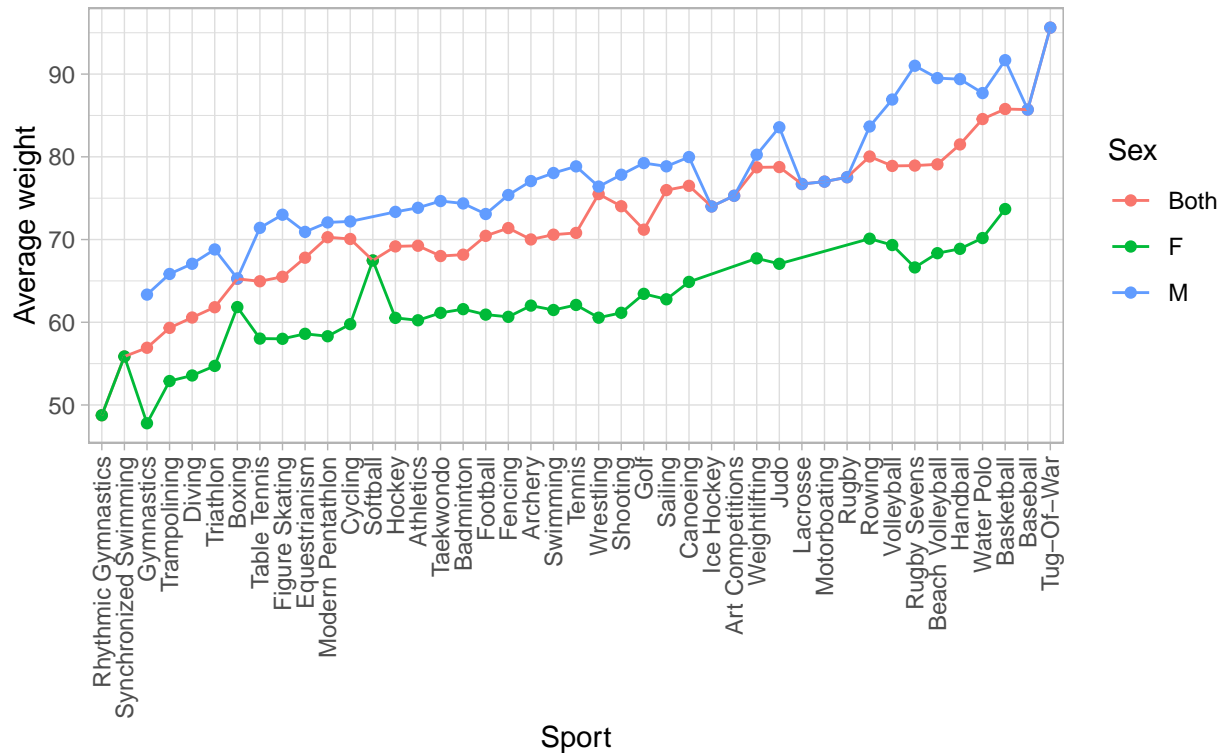
weight_plot <- avg_weight %>%
  ggplot() +
  geom_point(aes(reorder(sport, sex_sport_avg_weight), sex_sport_avg_weight,
                      color = sex, group = sex)) +
  geom_line(aes(reorder(sport, sex_sport_avg_weight), sex_sport_avg_weight,
                      color = sex, group = sex))

weight_plot +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  labs(x = "Sport", y = "Average weight", color = "Sex",
       title = "Tug-of-war is the absolute leader in average athlete weight",
       subtitle = "Considering all summer olympic sports")

```



Tug-of-war is the absolute leader in average athlete weight  
Considering all summer olympic sports



Sports ordered by average age, from low to high

```
avg_age <- summer_olympics %>%
  drop_na(age) %>%
  group_by(sport, sex) %>%
  mutate(sex_sport_avg_age = mean(age))

avg_age <- avg_age %>%
  ungroup() %>%
  group_by(sport) %>%
  mutate(sport_avg_age = mean(age)) %>%
  select(sex, sport, sex_sport_avg_age,
         sport_avg_age) %>%
  distinct()

both_sexes_avg_age <- avg_age %>%
  select(sport, sex_sport_avg_age = sport_avg_age) %>%
  mutate(sex = "Both") %>%
  distinct()

avg_age <- avg_age %>%
  select(-sport_avg_age) %>%
  bind_rows(both_sexes_avg_age) %>%
```

```

arrange(sport, sex)

age_plot <- avg_age %>%
  ggplot() +
  geom_point(aes(reorder(sport, sex_sport_avg_age), sex_sport_avg_age,
    color = sex, group = sex)) +
  geom_line(aes(reorder(sport, sex_sport_avg_age), sex_sport_avg_age,
    color = sex, group = sex))

age_plot +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  labs(x = "Sport", y = "Average age", color = "Sex",
    title = "A few sports have an average age greater than 30",
    subtitle = "Considering all summer olympic sports")

```

A few sports have an average age greater than 30  
Considering all summer olympic sports

