Stat 344 – RMarkdown Challenge

Luca Seazzu

January 19, 2023

Mathematical Notation

The probability of both event A and event B happening is $P(A \cap B)$, and if A and B are independent, this quantity is equivalent to $P(A)P(B)$. One way of writing the equation for a simple linear regression of y as a function of x is:

$$y = \beta_0 + \beta_x + \epsilon,$$

$$\epsilon \sim N(0, \sigma)$$

A Dataset

A dataset on participants in the Olympic Games is available at https://sldr.netlify.com/data/athlete_events.csv

A glimpse() of the dataset is below:

```
glimpse(olympics)
```

```
## Rows: 271,116
## Columns: 15
## $ ID     <int> 1, 2, 3, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, ~
## $ Name   <chr> "A Dijiang", "A Lamusi", "Gunnar Nielsen Aaby", "Edgar Lindenau~
## $ Sex    <chr> "M", "M", "M", "M", "F", "F", "F", "F", "F", "F", "M", "M", "M"~
## $ Age    <int> 24, 23, 24, 34, 21, 21, 25, 25, 27, 27, 31, 31, 31, 31, 33, 33,~
## $ Height <int> 180, 170, NA, NA, 185, 185, 185, 185, 185, 185, 188, 188, 188, ~
## $ Weight <dbl> 80, 60, NA, NA, 82, 82, 82, 82, 82, 82, 75, 75, 75, 75, 75, 75,~
## $ Team   <chr> "China", "China", "Denmark", "Denmark/Sweden", "Netherlands", "~
## $ NOC    <chr> "CHN", "CHN", "DEN", "DEN", "NED", "NED", "NED", "NED", "NED", ~
## $ Games  <chr> "1992 Summer", "2012 Summer", "1920 Summer", "1900 Summer", "19~
## $ Year   <int> 1992, 2012, 1920, 1900, 1988, 1988, 1992, 1992, 1994, 1994, 199~
## $ Season <chr> "Summer", "Summer", "Summer", "Summer", "Winter", "Winter", "Wi~
## $ City   <chr> "Barcelona", "London", "Antwerpen", "Paris", "Calgary", "Calgar~
## $ Sport  <chr> "Basketball", "Judo", "Football", "Tug-Of-War", "Speed Skating"~
## $ Event  <chr> "Basketball Men's Basketball", "Judo Men's Extra-Lightweight", ~
## $ Medal  <chr> NA, NA, NA, "Gold", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

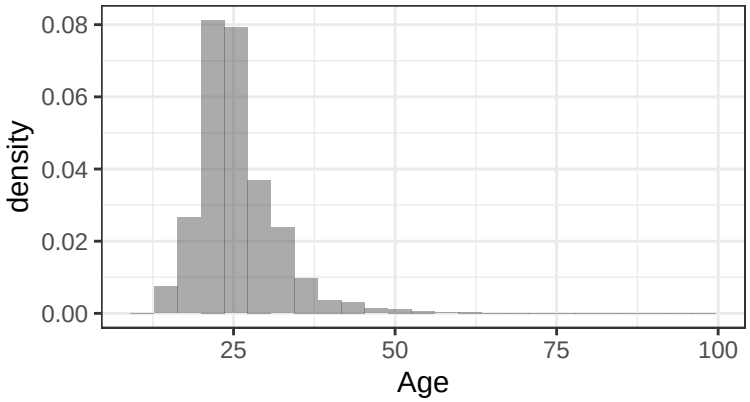
Graphics

#Age distribution of Olympians

```
gf_dhistogram(data = olympics, ~Age)
```

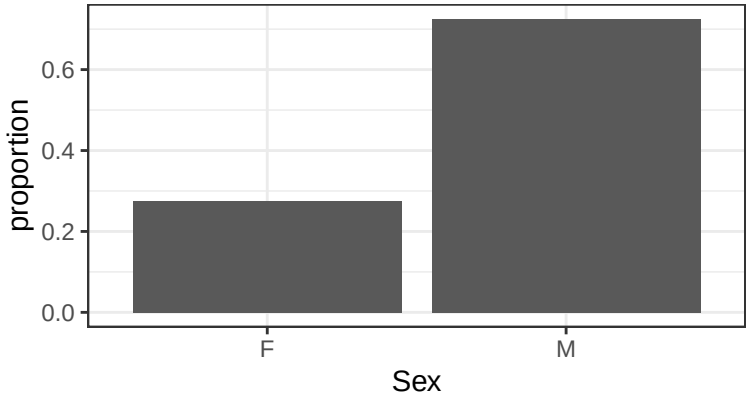
```
## Warning: `stat(density)` was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
```

```
## Warning: Removed 9474 rows containing non-finite values (`stat_bin()`).
```



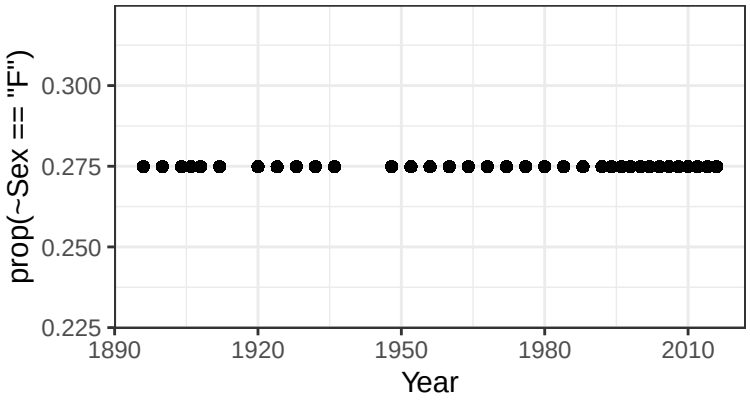
#Numbers of Male and Female Olympians

```
gf_props(~Sex, data = olympics)
```



#Extra Credit: Proportion Female by Year

```
gf_point(prop(~Sex == 'F') ~Year, data = olympics)
```



Couldn't quite remember how to do this off the top of my head.

#Extra Credit: Total US Medals, 1896-2016

Also couldn't quite remember how to do this but it is something like tally I think.