

## Drawing graphs

# Our data

- ▶ To illustrate making graphs, we need some data.
- ▶ Data on 202 male and female athletes at the Australian Institute of Sport.
- ▶ Variables:
  - ▶ categorical: Sex of athlete, sport they play
  - ▶ quantitative: height (cm), weight (kg), lean body mass, red and white blood cell counts, haematocrit and haemoglobin (blood), ferritin concentration, body mass index, percent body fat.
- ▶ Values separated by tabs (which impacts reading in).

## Packages for this section

```
library(tidyverse)
```

## Reading data into R

- ▶ Use `read_tsv` (“tab-separated values”), like `read_csv`.
- ▶ Data in `ais.txt`:

```
my_url <- "http://ritsokiguess.site/datafiles/ais.txt"  
athletes <- read_tsv(my_url)
```

## The data (some)

```
athletes
```

```
# A tibble: 202 x 13
```

|    | Sex    | Sport   | RCC   | WCC   | Hc    | Hg    | Ferr  | BMI   | SSB   |
|----|--------|---------|-------|-------|-------|-------|-------|-------|-------|
|    | <chr>  | <chr>   | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1  | female | Netball | 4.56  | 13.3  | 42.2  | 13.6  | 20    | 19.2  | 49    |
| 2  | female | Netball | 4.15  | 6     | 38    | 12.7  | 59    | 21.2  | 110.  |
| 3  | female | Netball | 4.16  | 7.6   | 37.5  | 12.3  | 22    | 21.4  | 89    |
| 4  | female | Netball | 4.32  | 6.4   | 37.7  | 12.3  | 30    | 21.0  | 98.3  |
| 5  | female | Netball | 4.06  | 5.8   | 38.7  | 12.8  | 78    | 21.8  | 122.  |
| 6  | female | Netball | 4.12  | 6.1   | 36.6  | 11.8  | 21    | 21.4  | 90.4  |
| 7  | female | Netball | 4.17  | 5     | 37.4  | 12.7  | 109   | 21.5  | 107.  |
| 8  | female | Netball | 3.8   | 6.6   | 36.5  | 12.4  | 102   | 24.4  | 157.  |
| 9  | female | Netball | 3.96  | 5.5   | 36.3  | 12.4  | 71    | 22.6  | 101.  |
| 10 | female | Netball | 4.44  | 9.7   | 41.4  | 14.1  | 64    | 22.8  | 126.  |

```
# i 192 more rows  
# i 1 more variable: Wt <dbl>
```

# Types of graph

Depends on number and type of variables:

| Categorical | Quantitative | Graph   |
|-------------|--------------|---|
| 1           | 0            | bar chart   |
| 0           | 1            | histogram   |
| 2           | 0            | grouped bar charts  |
| 1           | 1            | side-by-side boxplots                                       |
| 0           | 2            | scatterplot   |
| 2           | 1            | grouped boxplots  |
| 1           | 2            | scatterplot with points identified by group (eg. by colour) |

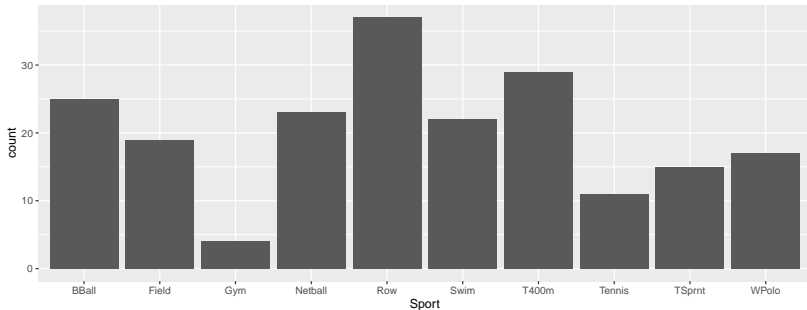
With more (categorical) variables, might want *separate plots by groups*. This is called *facetting* in R.

# ggplot

- ▶ R has a standard graphing procedure `ggplot`, that we use for all our graphs.
- ▶ Use in different ways to get precise graph we want.
- ▶ Let's start with bar chart of the sports played by the athletes.

# Bar chart

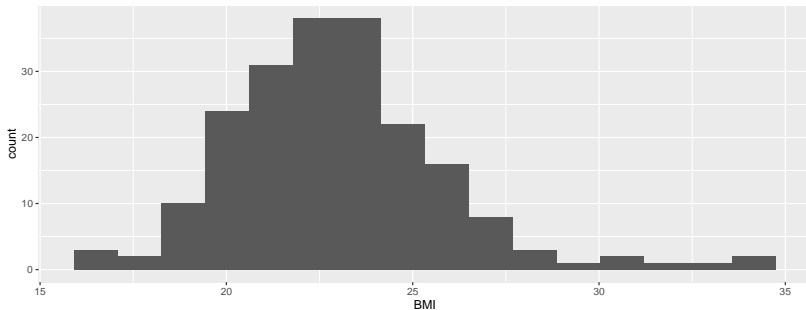
```
ggplot(athletes, aes(x = Sport)) + geom_bar()
```





# Histogram of body mass index

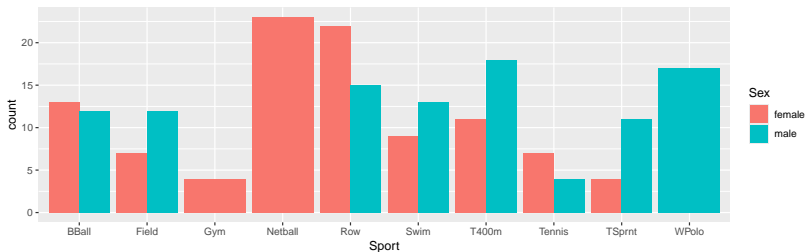
```
ggplot(athletes, aes(x = BMI)) + geom_histogram(bins = 16)
```



# Which sports are played by males and females?

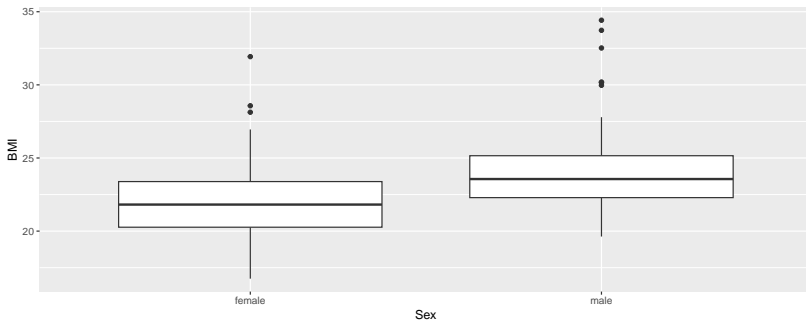
Grouped bar chart:

```
ggplot(athletes, aes(x = Sport, fill = Sex)) +  
  geom_bar(position = "dodge")
```



# BMI by gender

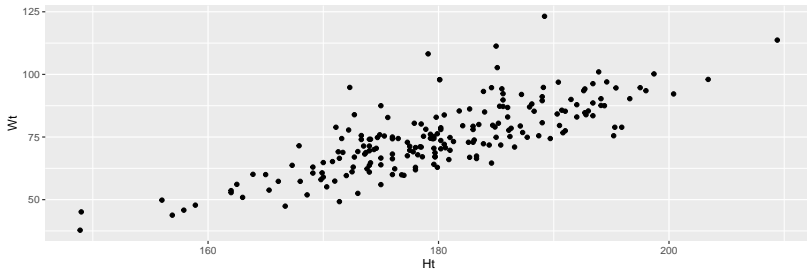
```
ggplot(athletes, aes(x = Sex, y = BMI)) + geom_boxplot()
```



# Height vs. weight

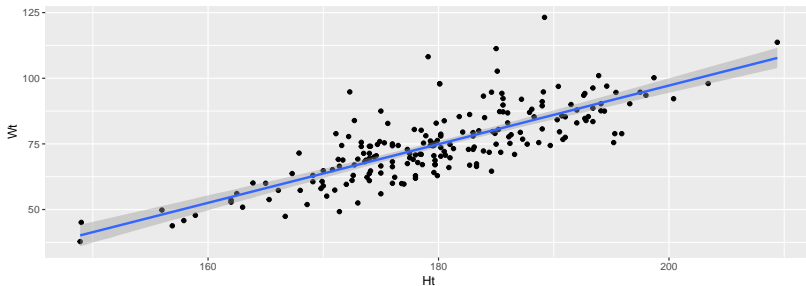
Scatterplot:

```
ggplot(athletes, aes(x = Ht, y = Wt)) + geom_point()
```



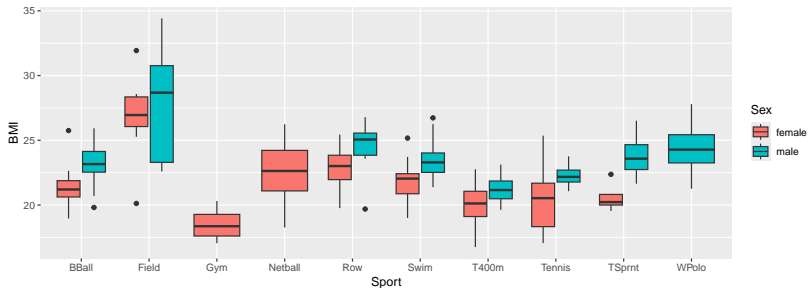
## With regression line

```
ggplot(athletes, aes(x = Ht, y = Wt)) +  
  geom_point() + geom_smooth(method = "lm")
```



# BMI by sport and gender

```
ggplot(athletes, aes(x = Sport, y = BMI, fill = Sex)) +  
  geom_boxplot()
```



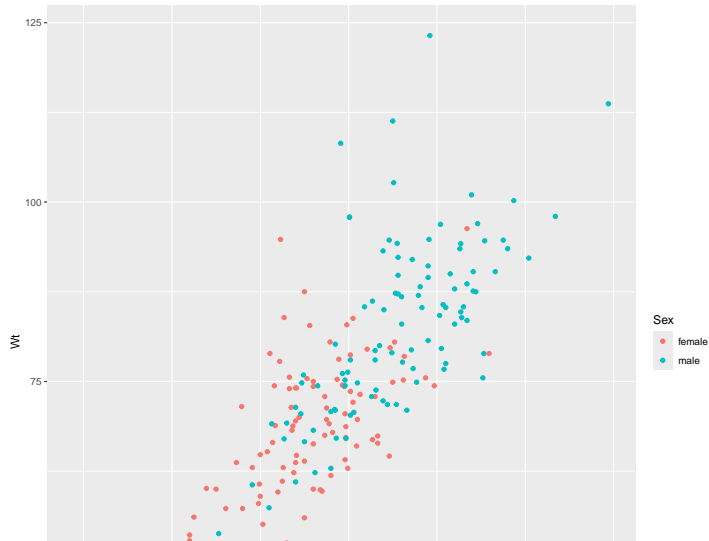
A variation that uses colour instead of fill:

```
ggplot(athletes, aes(x = Sport, y = BMI, colour = Sex)) +  
  geom_boxplot()
```



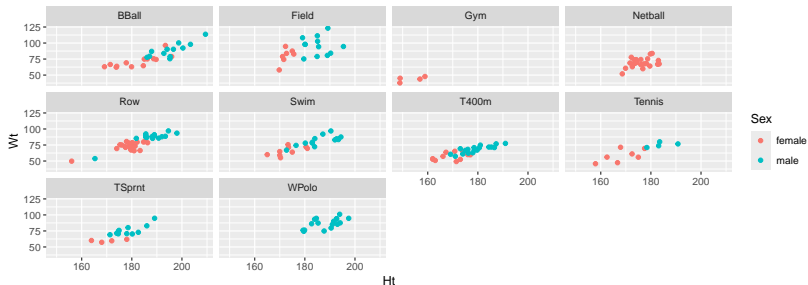
# Height and weight by gender

```
ggplot(athletes, aes(x = Ht, y = Wt, colour = Sex)) +  
  geom_point()
```



# Height by weight by gender for each sport, with facets

```
ggplot(athletes, aes(x = Ht, y = Wt, colour = Sex)) +  
  geom_point() + facet_wrap(~Sport)
```

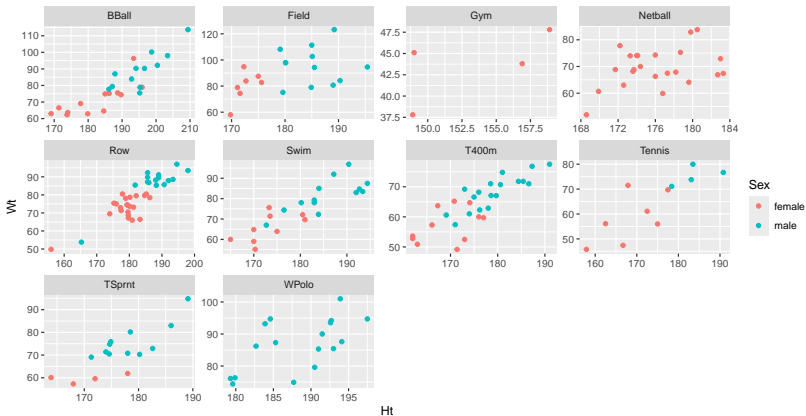




## Filling each facet

Default uses same scale for each facet. To use different scales for each facet, this:

```
ggplot(athletes, aes(x = Ht, y = Wt, colour = Sex)) +  
  geom_point() + facet_wrap(~Sport, scales = "free")
```



## Another view of height vs weight

```
ggplot(athletes, aes(x = Ht, y = Wt)) +  
  geom_point() + facet_wrap(~ Sex)
```

