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

```r
library(tidyverse)
```

Reading data into R

- Use read_tsv ("tab-separated values"), like read_csv.
- Data in ais.txt:

```
my_url <- "http://www.utsc.utoronto.ca/~butler/c32/ais.txt"
athletes <- read_tsv(my_url)
```

```
##
## -- Column specification ----------------------------
## cols(
##   Sex = col_character(),
##   Sport = col_character(),
##   RCC = col_double(),
##   WCC = col_double(),
##   Hc = col_double(),
##   Hg = col_double(),
##   Ferr = col_double(),
##   BMI = col_double(),
##   SSF = col_double(),
##   `%Bfat` = col_double(),
##   LBM = col_double(),
##   Ht = col_double(),
##   Wt = col_double()
## )
```

The data (some)

```
athletes
```

Sex	Sport	RCC	WCC	Hc	Hg	Ferr	BMI	SSF	%Bfat	LBM
female	Netball	4.56	13.30	42.2	13.6	20	19.16	49.0	11.29	53.14
female	Netball	4.15	6.00	38.0	12.7	59	21.15	110.2	25.26	47.09
female	Netball	4.16	7.60	37.5	12.3	22	21.40	89.0	19.39	53.44
female	Netball	4.32	6.40	37.7	12.3	30	21.03	98.3	19.63	48.78
female	Netball	4.06	5.80	38.7	12.8	78	21.77	122.1	23.11	56.05
female	Netball	4.12	6.10	36.6	11.8	21	21.38	90.4	16.86	56.45
female	Netball	4.17	5.00	37.4	12.7	109	21.47	106.9	21.32	53.11
female	Netball	3.80	6.60	36.5	12.4	102	24.45	156.6	26.57	54.41
female	Netball	3.96	5.50	36.3	12.4	71	22.63	101.1	17.93	55.97
female	Netball	4.44	9.70	41.4	14.1	64	22.80	126.4	24.97	51.62
female	Netball	4.27	10.60	37.7	12.5	68	23.58	114.0	22.62	58.27
female	Netball	3.90	6.30	35.9	12.1	78	20.06	70.0	15.01	57.28
female	Netball	4.02	9.10	37.7	12.7	107	23.01	77.0	18.14	57.30
female	Netball	4.39	9.60	38.3	12.5	39	24.64	148.9	26.78	54.18
female	Netball	4.52	5.10	38.8	13.1	58	18.26	80.1	17.22	42.96
female	Netball	4.25	10.70	39.5	13.2	127	24.47	156.6	26.50	54.46
female	Netball	4.46	10.90	39.7	13.7	102	23.99	115.9	23.01	57.20

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)

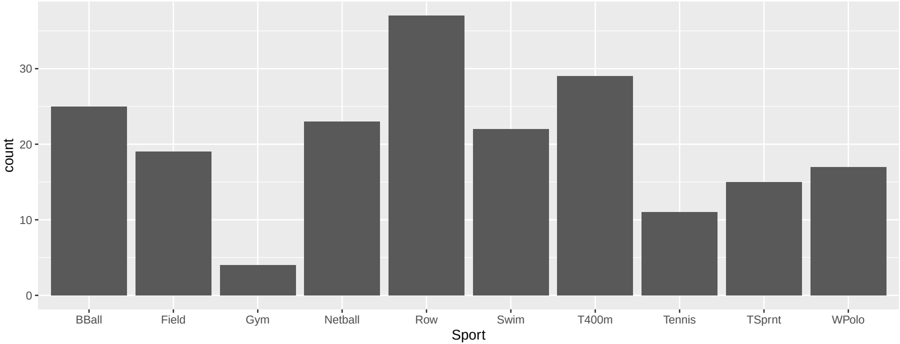
With more variables, might want *separate plots by groups*. This is called `facetting` in R.

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.

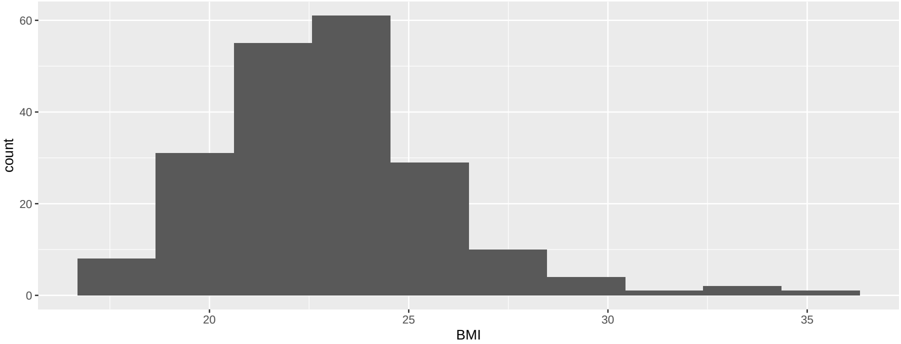
Bar chart

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



Histogram of body mass index

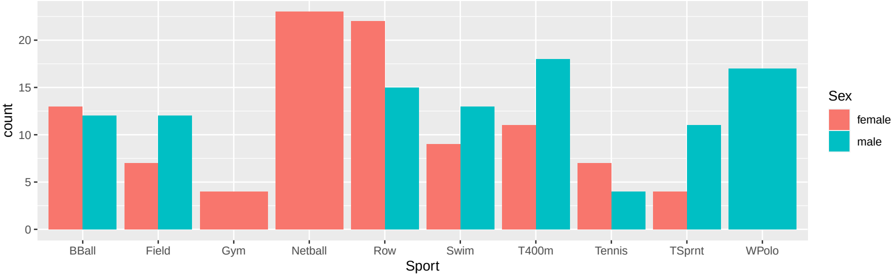
```
ggplot(athletes, aes(x = BMI)) + geom_histogram(bins = 10)
```



Which sports are played by males and females?

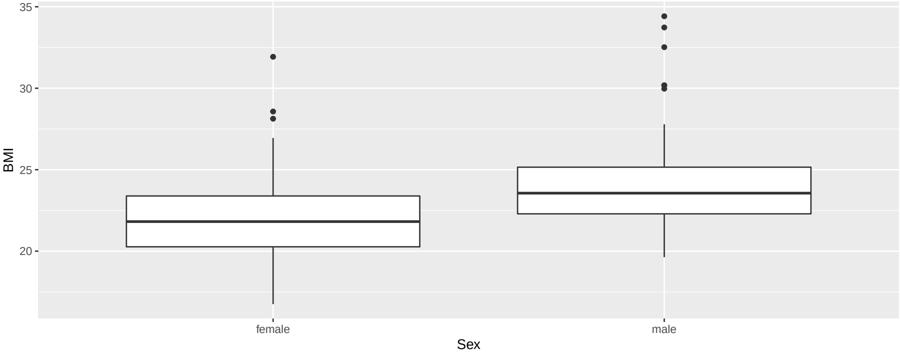
Grouped bar chart:

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



BMI by gender

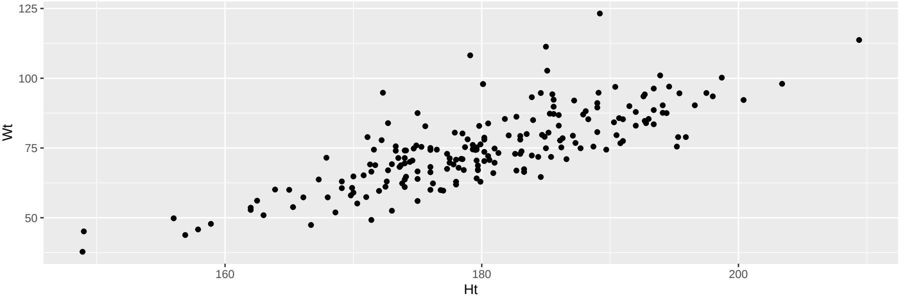
```
ggplot(athletes, aes(x = Sex, y = BMI)) + geom_boxplot()
```



Height vs. weight

Scatterplot:

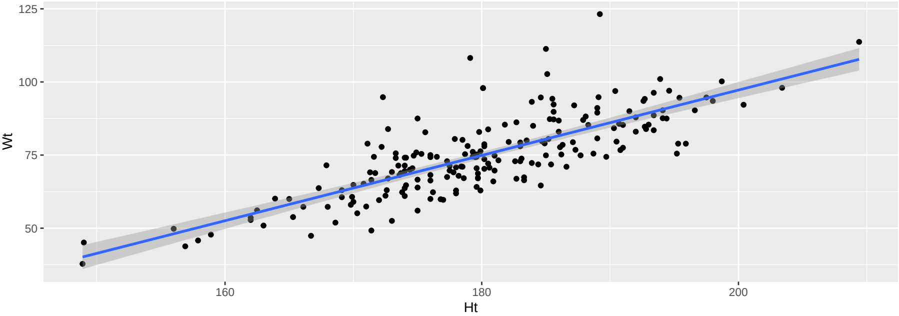
```
ggplot(athletes, aes(x = Ht, y = Wt)) + geom_point()
```



With regression line

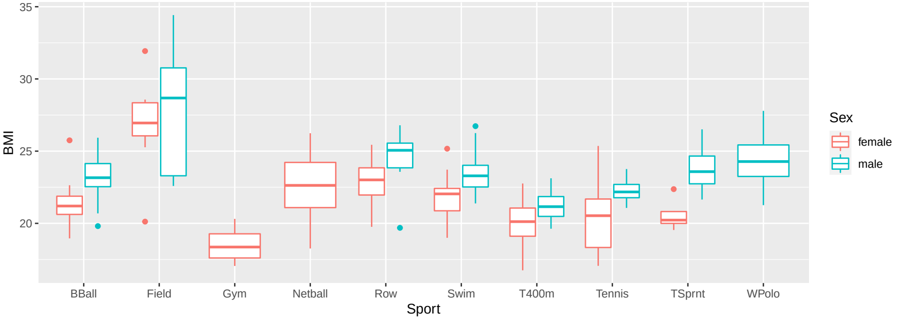
```r
ggplot(athletes, aes(x = Ht, y = Wt)) +
  geom_point() + geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



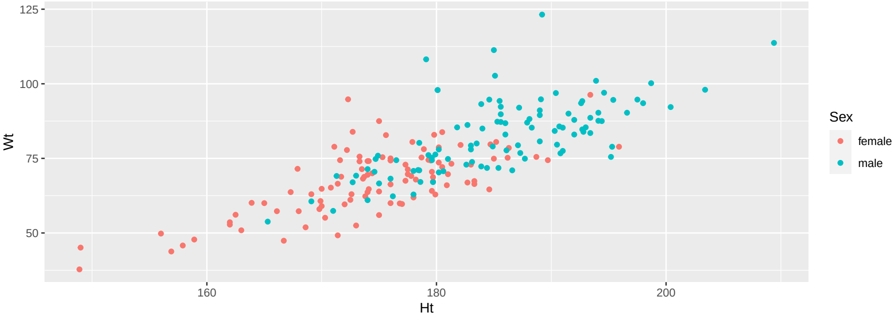
BMI by sport and gender

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



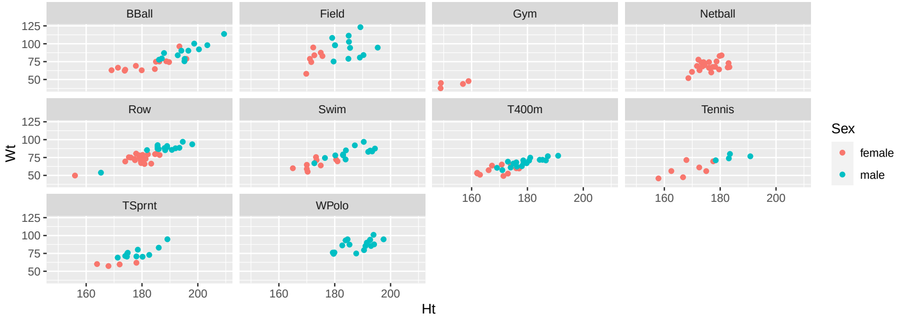
Height and weight by gender

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



Height by weight for each sport, with facets

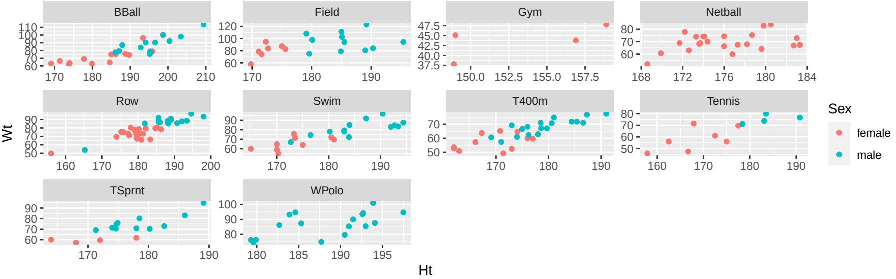
```
ggplot(athletes, aes(x = Ht, y = Wt, colour = Sex)) +
  geom_point() + facet_wrap(~Sport)
```



Filling each facet

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

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


BBall
Field
Gym
Netball
Row
Swim
T400m
Tennis
TSprnt
WPolo
Wt
Ht
Sex
female
male