Practical 1

Jumping Rivers

We will continue to investigate the movies data set.

- Make sure you've downloaded the GitHub repository
- Link: `bit.ly/nhs-ggplot2`.
- Click on the .Rproj file to open the R project for today. This ensures that our working directory is correct!
- Open a new R script (`Ctrl+Shift+N` in RStudio.)

```
library("ggplot2")
movies = readRDS(file = "data/movies.rds")
```

When loading in data, it's always a good idea to carry out a sanity check. I tend to use the commands

```
head(movies)
names(movies)
dim(movies)
View(movies)
```

Scatter plots

(1) Make a basic scatter plot with votes on the x axis and rating on the y axis. Remember, we use the `ggplot()` function to create a basic plot and them `geom_point()` to add points.

(2) Use the `labs()` function to change the axis labels & and title to something better

(3) The range of possible ratings is between 0 and 10, however because the maximum rating is below 10 the y-axis stops before 10. We can change the axis range using the `ylim` function. For instance, if our graph was stored in an object `g`, we could change the y-axis limit to (0,10) like so `+ ylim(0, 10)`.

Do this to your graph.

Histograms

(1) Make a basic histogram of the year of releases.

(2) Change the binwidth to one year using the `binwidth` argument in `geom_histogram()`

(3) Try changing `geom_histogram()` to `geom_density()`. You won't need a binwidth argument for the density plot.

Bonus questions - an introduction to colours and fills

Let's go back to the classification bar chart we had in the slides.

```
ggplot(movies, aes(x = classification)) + geom_bar()
```

(1) Inside the `geom_bar()` function, try adding the argument `colour = "blue"`. What happens?

```
ggplot(movies, aes(x = classification)) + geom_bar(colour = "blue")
```

(1) Change `colour = "blue"` to `fill = "blue"`

(1) Experiment with other colours, you can find a list of colours that R takes using the `colours()` function i.e. run

```
colours()
```

(1) Remove `fill = "blue"` from `geom_bar()`. Now try adding `fill = classification` to the `aes()` function. What happens?

```
ggplot(movies, aes(x = classification)) + geom_bar(aes(fill = classification))
```