# Answers 2 - Base graphics

**Osama Mahmoud**

We are going to investigate a brief form of the IMDB (*movies*) data set. Movies were selected for inclusion if they had a known length, had been rated by at least one IMDB user and had an mpaa rating. This gives $4847$ films, where each film has $24$ associated variables. The data can be called and viewed using:

```
data(bmov, package = "BristolVis")
head(bmov)
```

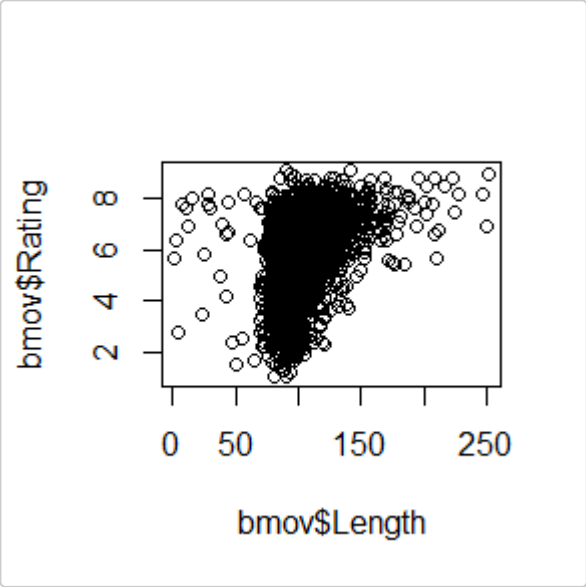
```
##                     Title Year Length   Budget Rating Votes   r1  r2  r3
## 1     A.k.a. Cassius Clay 1970     85       -1    5.7    43  4.5 0.0 4.5
## 2                     AKA 2002    123       -1    6.0   335 24.5 4.5 4.5
## 3 AVP: Alien Vs. Predator 2004    102 45000000    5.4 14651  4.5 4.5 4.5
## 4                 Abandon 2002     99 25000000    4.7  2364  4.5 4.5 4.5
## 5               Abendland 1999    146       -1    5.0    46 14.5 4.5 4.5
## 6              Aberration 1997     93       -1    4.8   149 14.5 4.5 4.5
##     r4   r5   r6   r7   r8  r9  r10  mpaa Action Animation Comedy Drama
## 1 14.5  4.5 24.5 14.5 14.5 4.5 14.5    PG      0         0      0     0
## 2  4.5  4.5  4.5 14.5 14.5 4.5 14.5     R      0         0      0     1
## 3  4.5 14.5 14.5 14.5  4.5 4.5  4.5 PG-13      1         0      0     0
## 4 14.5 14.5 14.5 14.5  4.5 4.5  4.5 PG-13      0         0      0     1
## 5  4.5  4.5  4.5 14.5 14.5 4.5 24.5     R      0         0      0     0
## 6 14.5 14.5 14.5 14.5  4.5 4.5  4.5     R      0         0      0     0
##   Documentary Romance Short
## 1           1       0     0
## 2           0       0     0
## 3           0       0     0
## 4           0       0     0
## 5           0       0     0
## 6           0       0     0
```

## Scatter plots (20 minutes)

Let's start with some simple scatter plots using the `bmov` data:

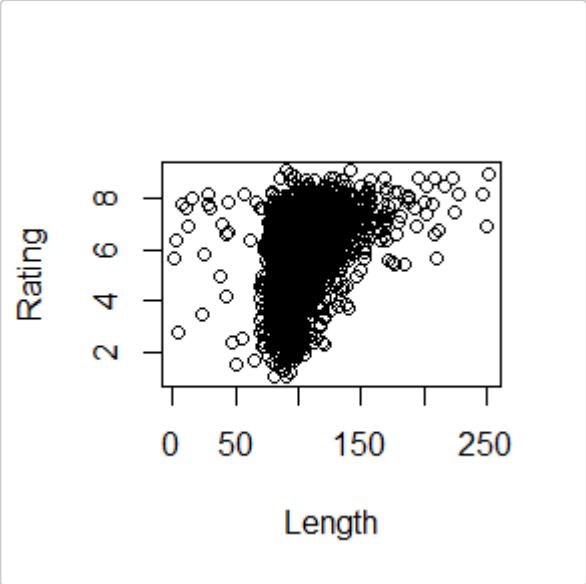
1. Plot length Vs. rating.

```
plot(bmov$Length, bmov$Rating)
```



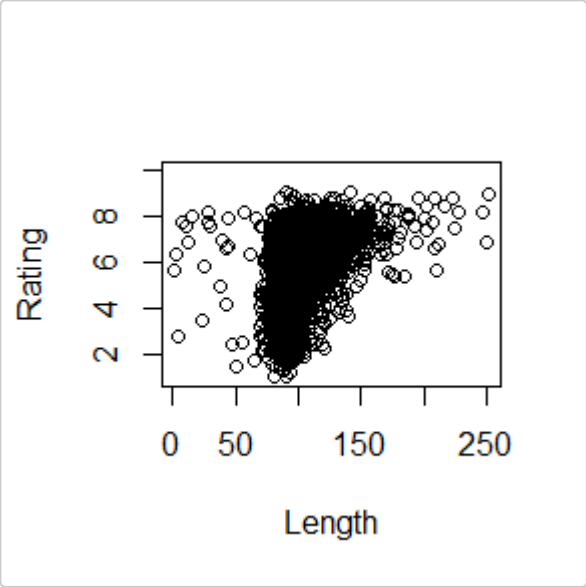

2. Use the `xlab` and `ylab` arguments to specify suitable axis labels.

```
plot(bmov$Length, bmov$Rating, xlab="Length", ylab="Rating")
```



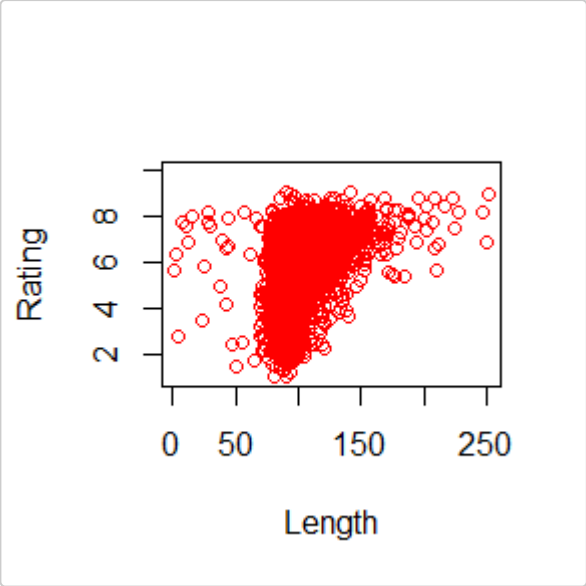

3. Use the `ylim` argument to specify a y-axis range from 1 to 10.

```
plot(bmov$Length, bmov$Rating, xlab="Length", ylab="Rating", ylim = c(1,10))
```



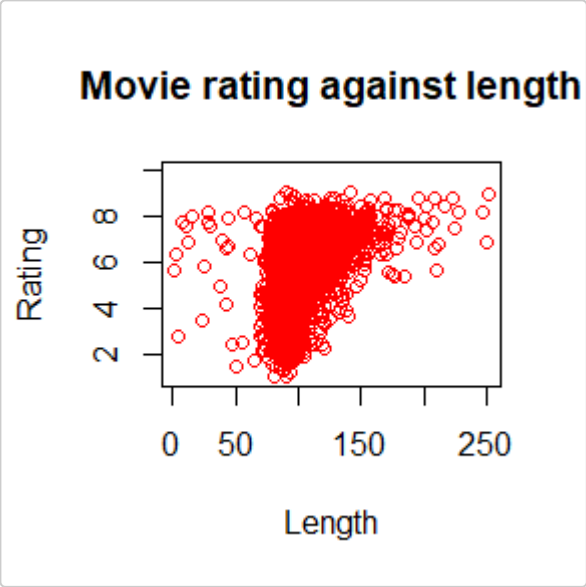

4. Use the `col` argument to change the colour of the points.

```
plot(bmov$Length, bmov$Rating, xlab="Length", ylab="Rating",
     ylim = c(1,10), col = 2)
```




5. Use the `main` argument to give the plot a suitable title.

```
plot(bmov$Length, bmov$Rating, xlab="Length", ylab="Rating",
     ylim = c(1,10), col = 2, main = "Movie rating against length")
```




6. If we altered the default plot parameters using:

```
op = par(mar=c(3,3,2,1), tck=-.01, las=1, cex.axis=0.4)
```

and generated our plot in (5) again, what will happen? can you figure out what *mar*, *tck*, *las* and *xec.axis* do?

7. Reset your plot device using:

```
par(op)
```

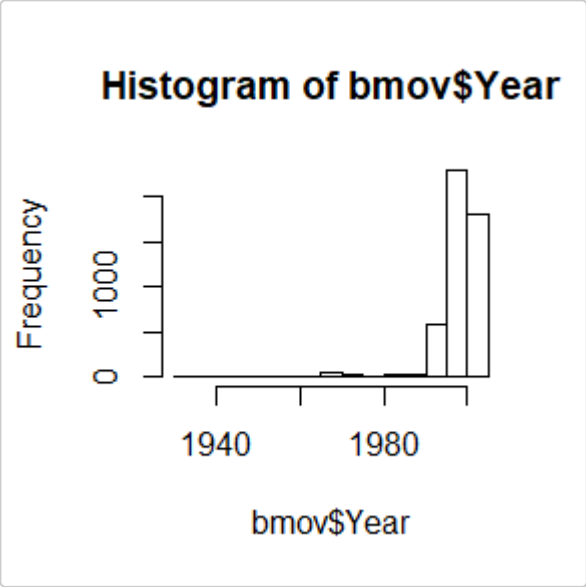
and generate the last plot again. Do you see the reset effect?

## Histograms (20 minutes)

We will now investigate the distribution of movie years using histograms.

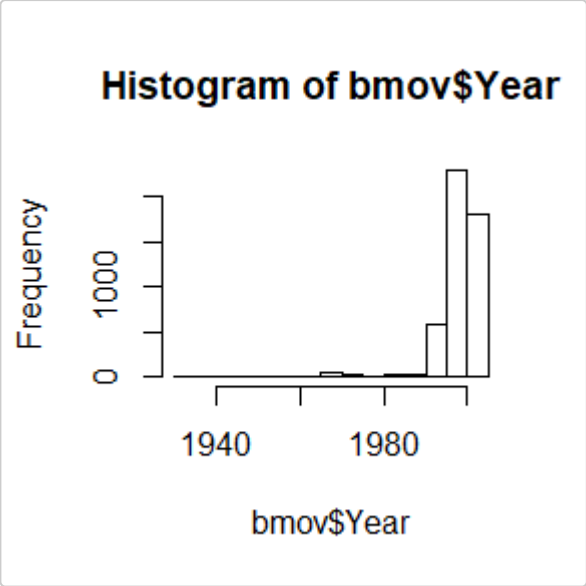
1. Use the `hist` function to plot a histogram of the movie years.

```
hist(bmov$Year)
```



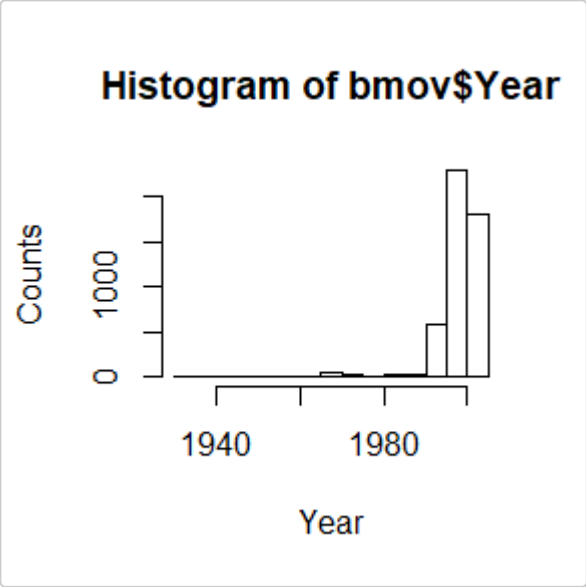

2. The default method for determining the number of bins isn't great. Use different rules to set breaks to `15`.

```
hist(bmov$Year, breaks = 15)
```



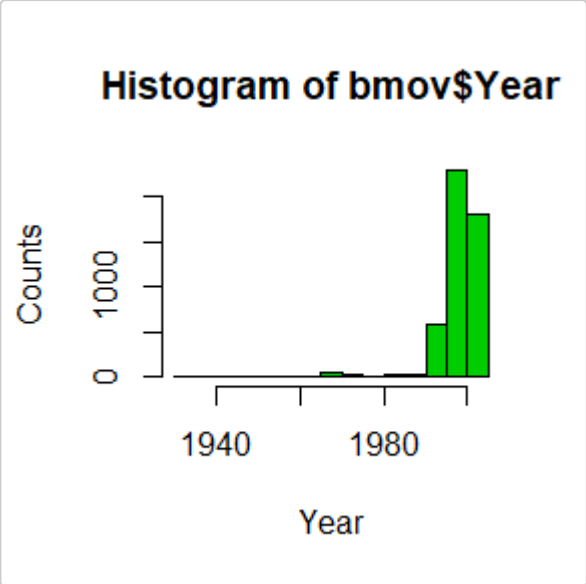

3. Use the `xlab` and `ylab` arguments to specify suitable axis labels.

```
hist(bmov$Year, breaks = 15, xlab = "Year", ylab = "Counts")
```



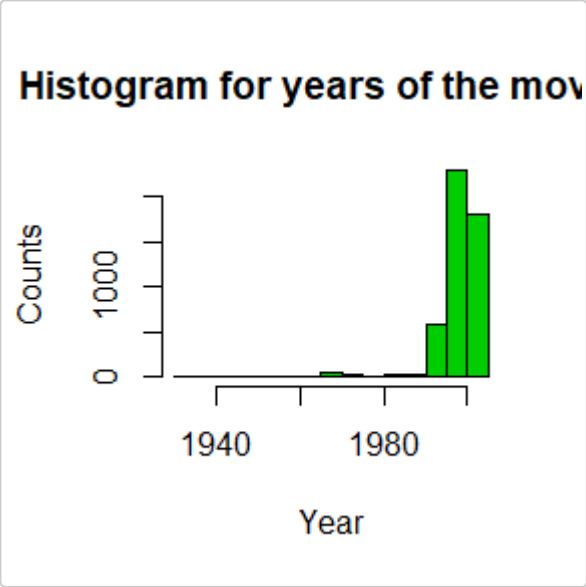

4. Use the `col` argument to change the colour of the histogram.

```
hist(bmov$Year, breaks = 15, xlab = "Year", ylab = "Counts", col = 3)
```




5. Use the `main` argument to give the plot a suitable title.

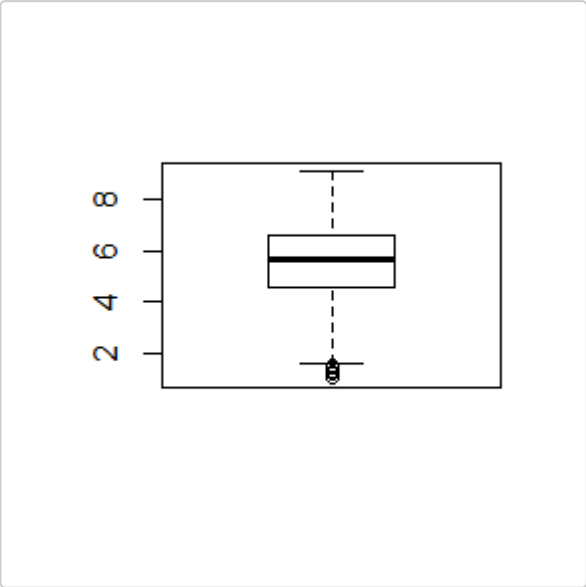
```
hist(bmov$Year, breaks = 15, xlab = "Year", ylab = "Counts", col = 3,
     main = "Histogram for years of the movies")
```




## Boxplots (10 minutes)

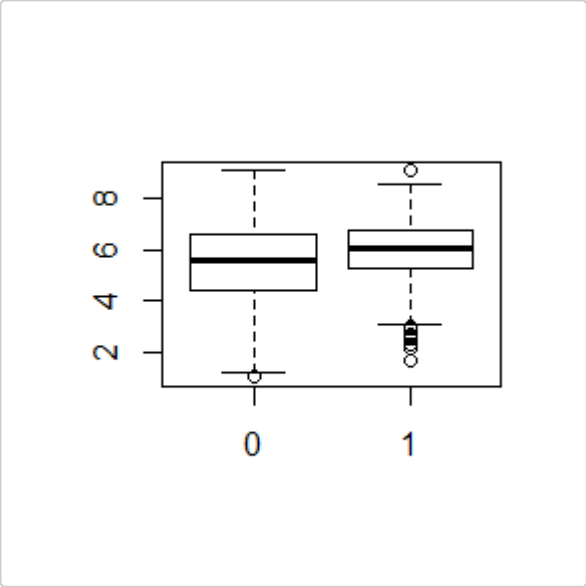
1. Generate a boxplot for the ratings data.

```
boxplot(bmov$Rating)
```



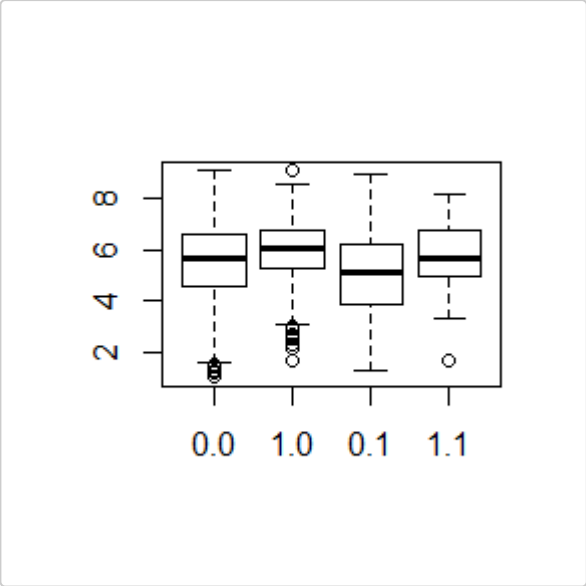

2. Separate the data by whether the movie is a romantic movie.

```
boxplot(bmov$Rating ~ bmov$Romance)
```




3. Try generating a similar boxplot, but for other variables. What happens when you condition on more than one variable? e.g. `Romance` and `Action`.

```
boxplot(bmov$Rating ~ bmov$Romance + bmov$Action)
```

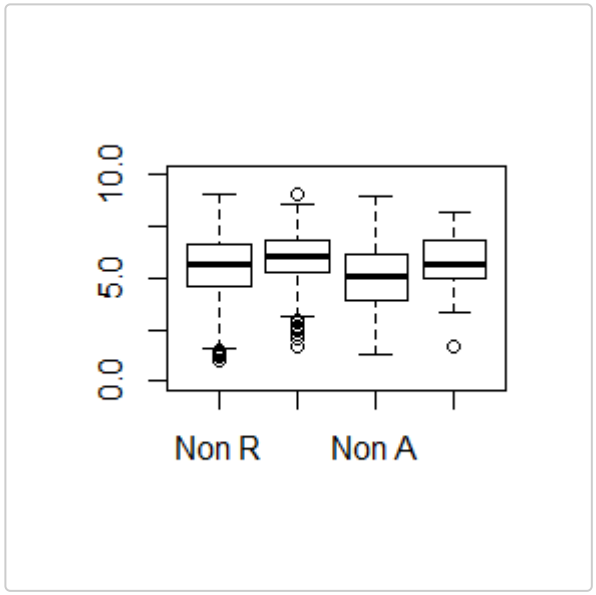



We can also change axis labels by:

```
# Plot a boxplot but skip the labels
boxplot(bmov$Rating ~ bmov$Romance + bmov$Action, axes=FALSE, frame.plot=TRUE, ylim=c(0,10))

# Y-axis: 0 to 10 in steps of 2.5
axis(2, at=seq(0,10,2.5))

## X-Axis, at points x=1 : x=4,
## but with labels "non R" and "Romantic"
axis(1,at=1:4, labels=c("Non R", "Romantic", "Non A", "Action"))
```


10.0
5.0
0.0
Non R
Non A