

Lab 4 solutions

19-20th February 2018

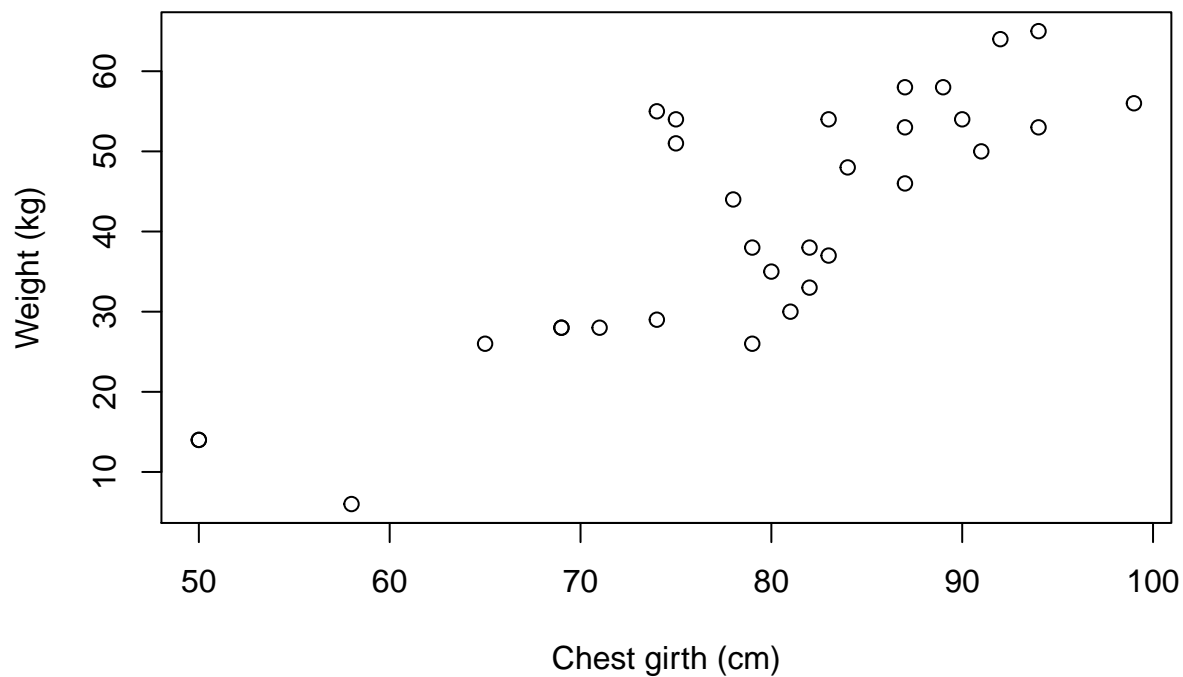
Exercise 1

Read in the goats data set

```
goats = read.csv("http://www.massey.ac.nz/~jcmarsha/rcourse/goatsdata1.csv")
```

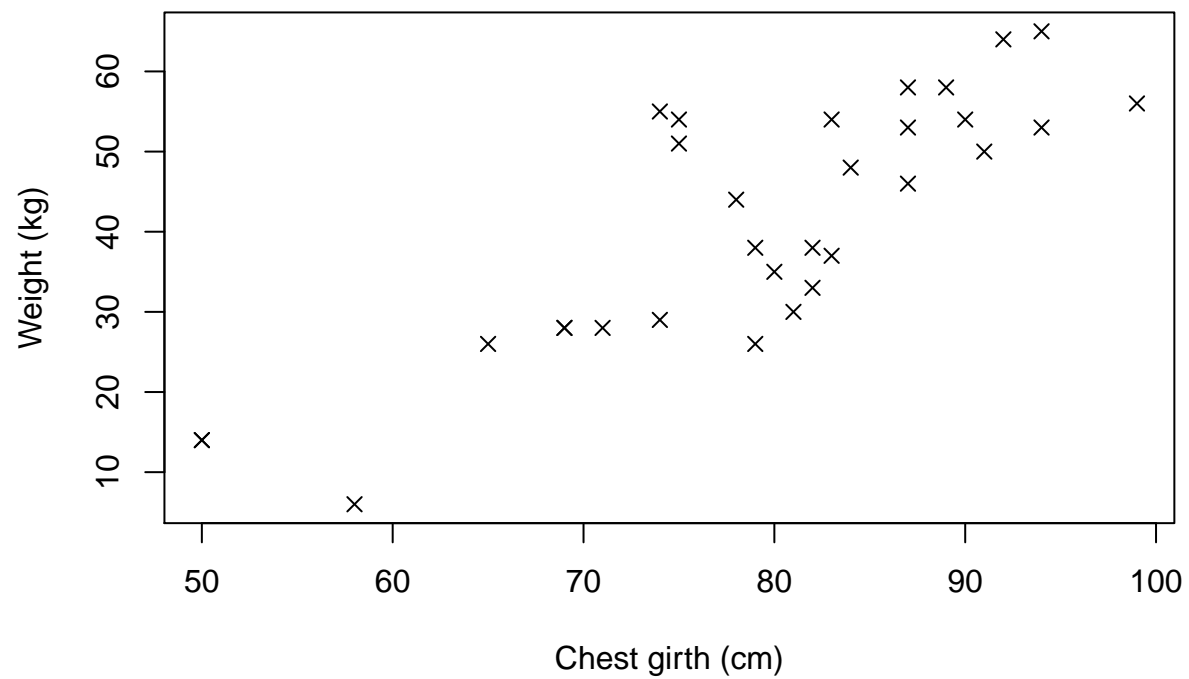
Produce a scatter plot of chest by weight

```
plot(goats$chest, goats$weight, xlab="Chest girth (cm)", ylab="Weight (kg)")
```



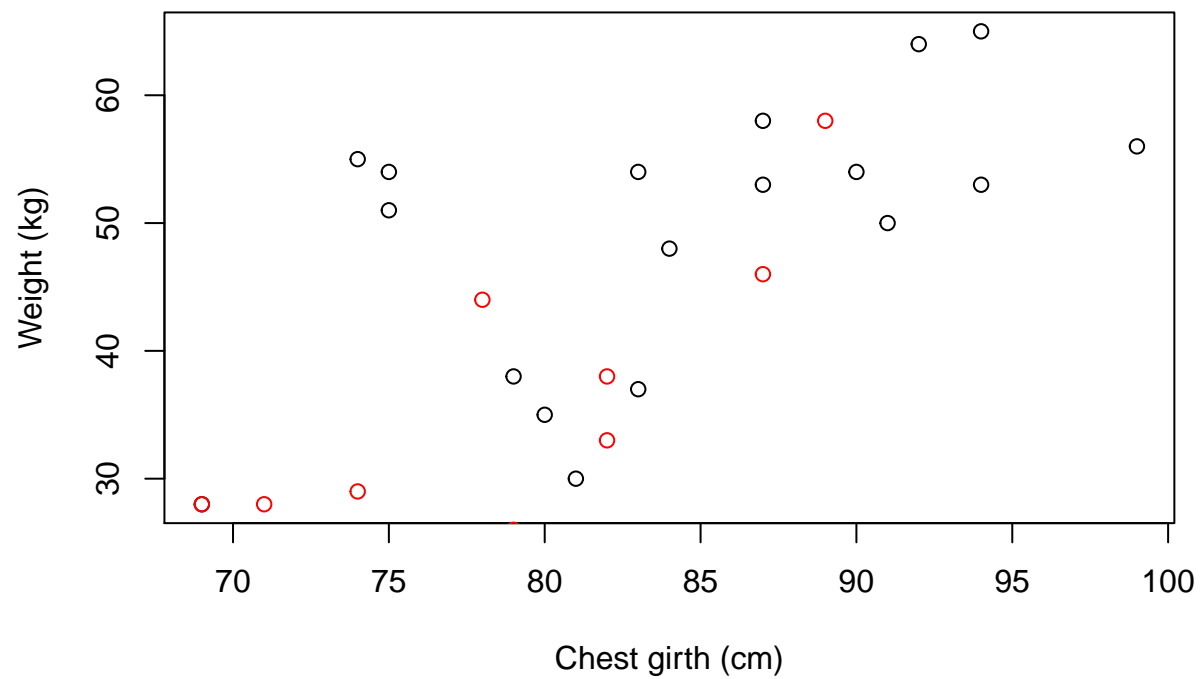
Alter the plotting symbol using pch

```
plot(goats$chest, goats$weight, xlab="Chest girth (cm)", ylab="Weight (kg)", pch=4)
```



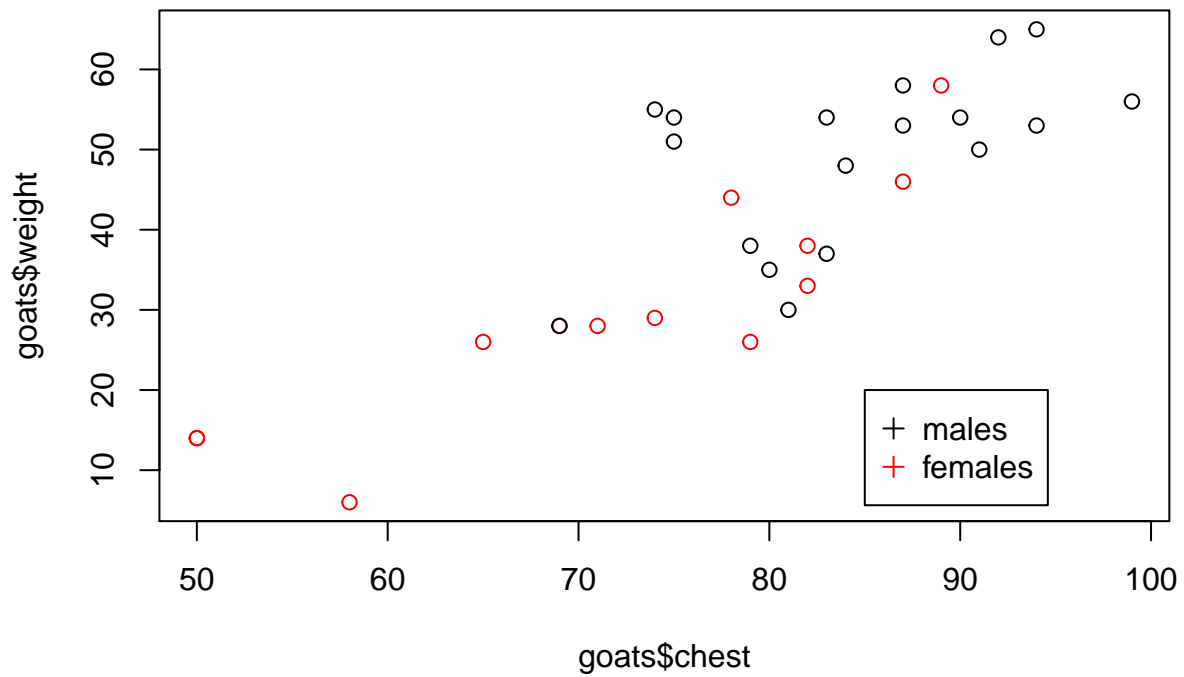
Subset the data to build a plot coloured differently for each gender

```
goatsM = goats[goats$gender==1,]  
goatsF = goats[goats$gender==2,]  
plot(goatsM$chest, goatsM$weight, xlab="Chest girth (cm)", ylab="Weight (kg)")  
points(goatsF$chest, goatsF$weight, col="red")
```



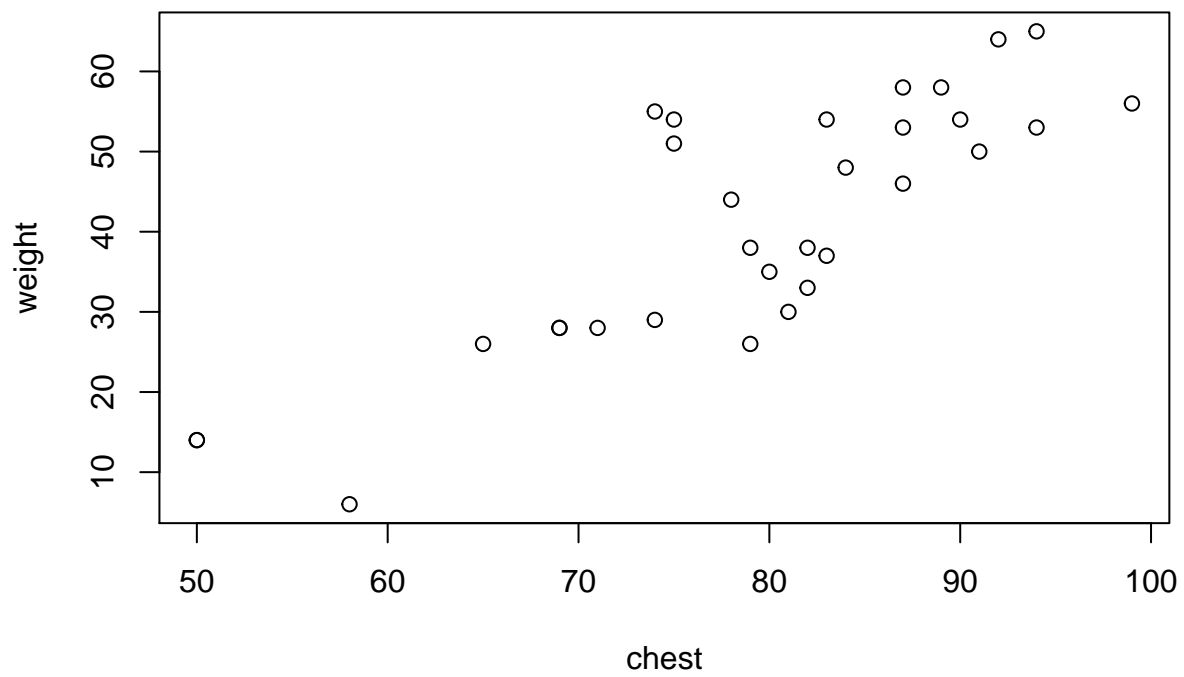
Alternatively colour each point with one plot statement

```
plot(goats$chest, goats$weight, col=goats$gender)
legend(85,20,legend=c("males", "females"), col=c(1,2),pch=3)
```

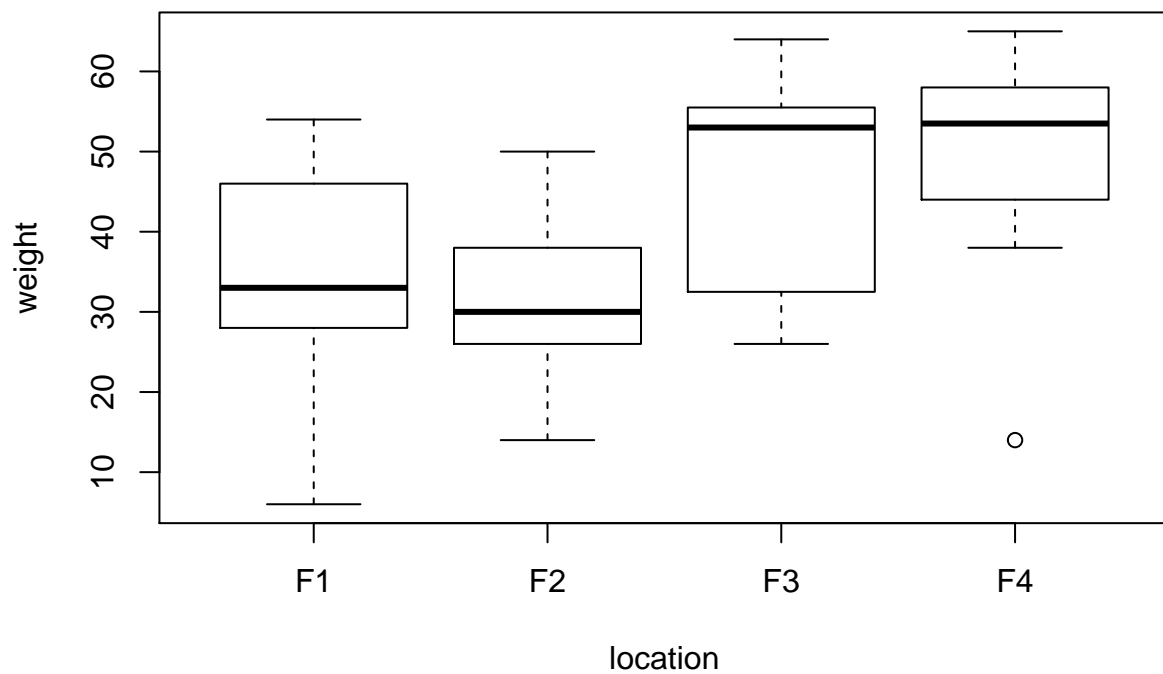


We can also use the formula notation, where we notice that R adapts the plot type based on the variable types.

```
plot(weight ~ chest, data=goats)
```



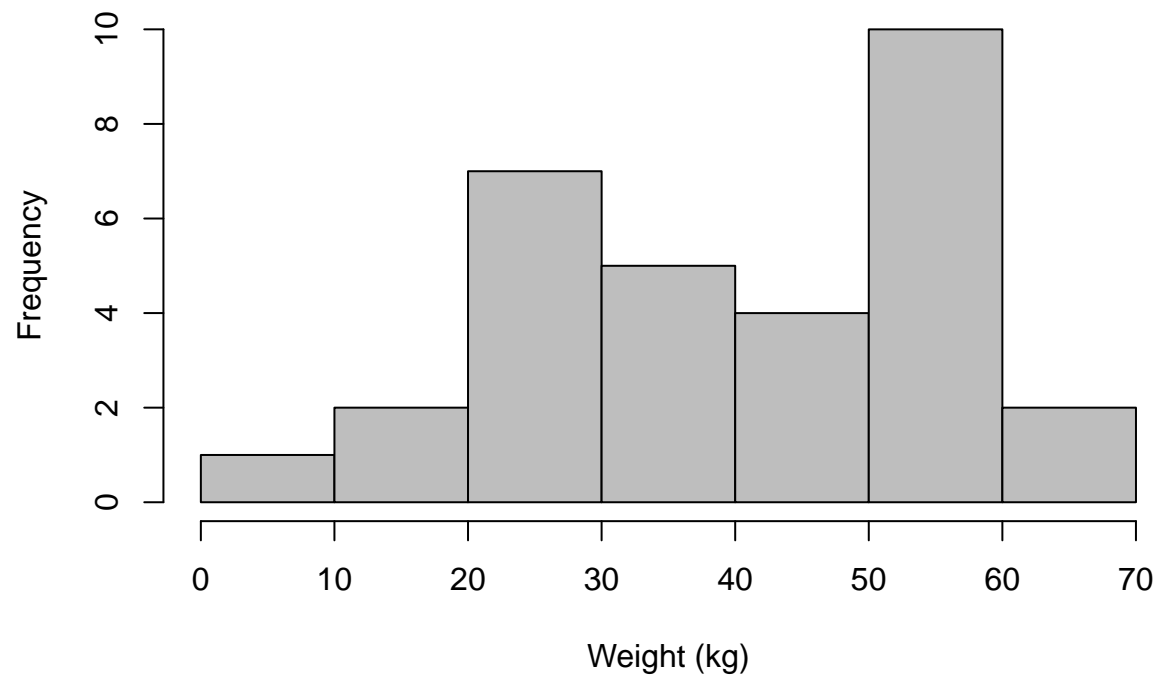
```
plot(weight ~ location, data=goats)
```



and a histogram

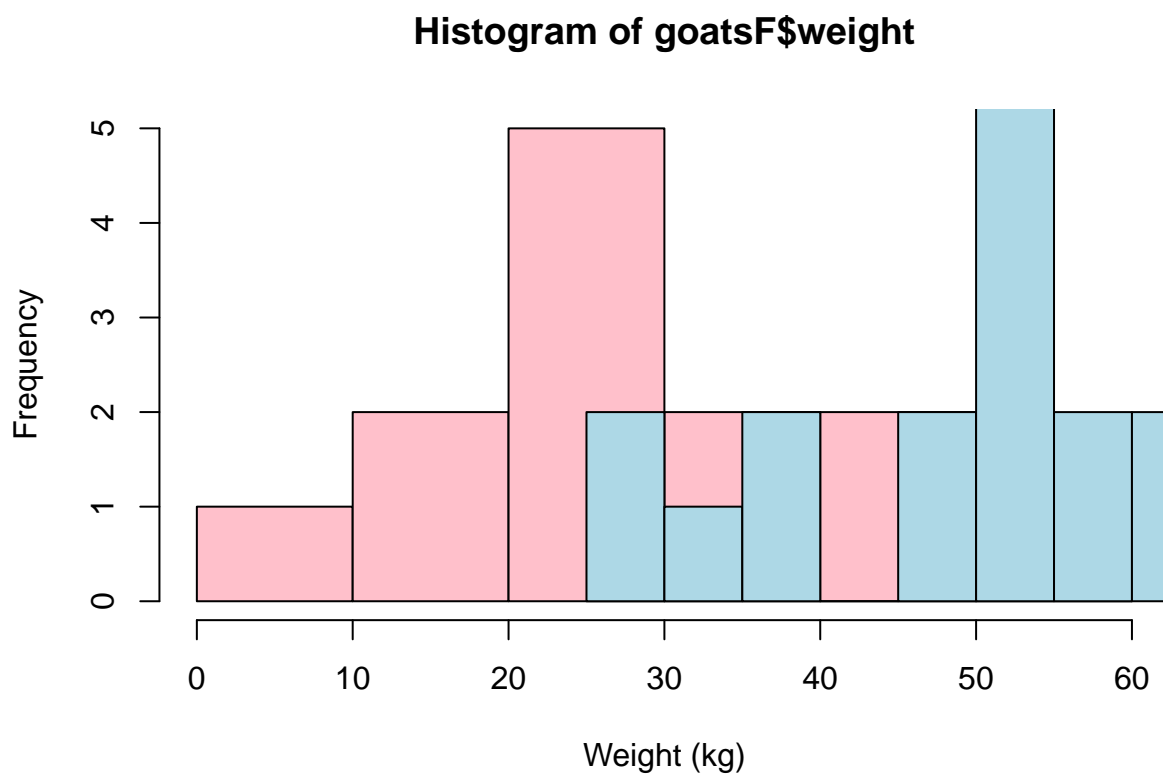
```
hist(goats$weight, xlab="Weight (kg)", col="grey")
```

Histogram of goats\$weight



and one that splits into male and female

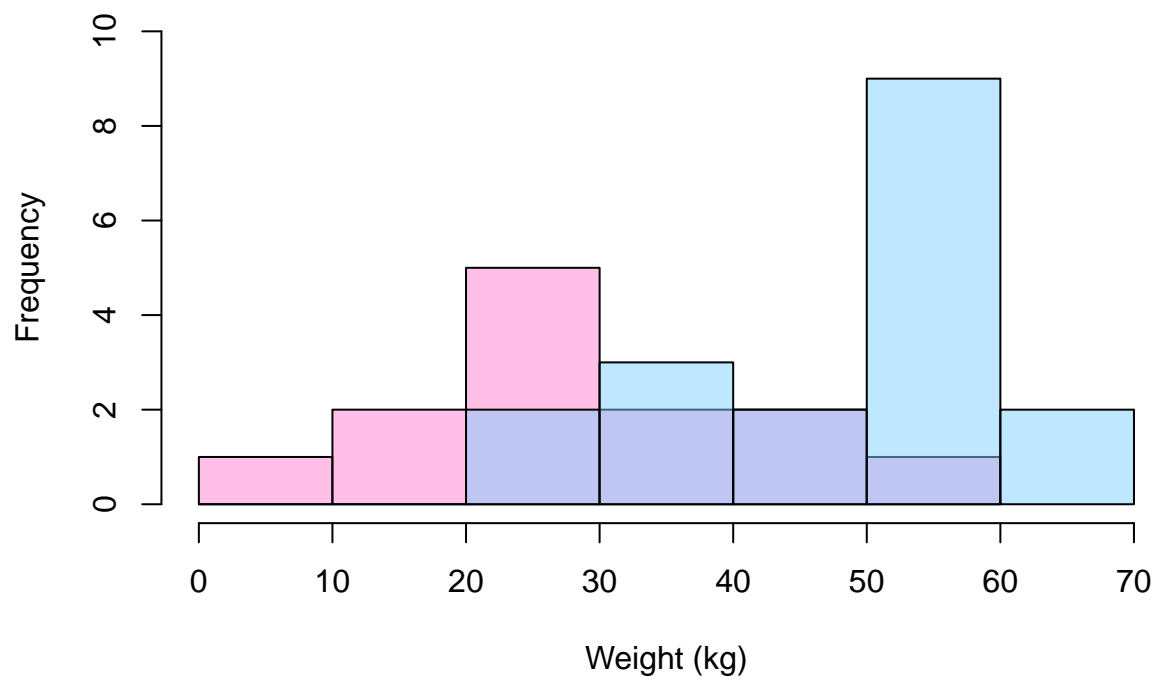
```
hist(goatsF$weight, xlab="Weight (kg)", col="pink")  
hist(goatsM$weight, col="lightblue", add=TRUE)
```



To do this better we'd really need to make sure that the scale covers both males and females. We can do this using the `xlim` and `ylim` arguments. we could also define the breaks, and maybe make the colours semi-transparent

```
hist(goatsF$weight, xlab="Weight (kg)", xlim=c(0,70), ylim=c(0,10), col="#ff7fcf7f")  
hist(goatsM$weight, col="#7fcfff7f", add=TRUE, breaks=seq(0,70,by=10))
```

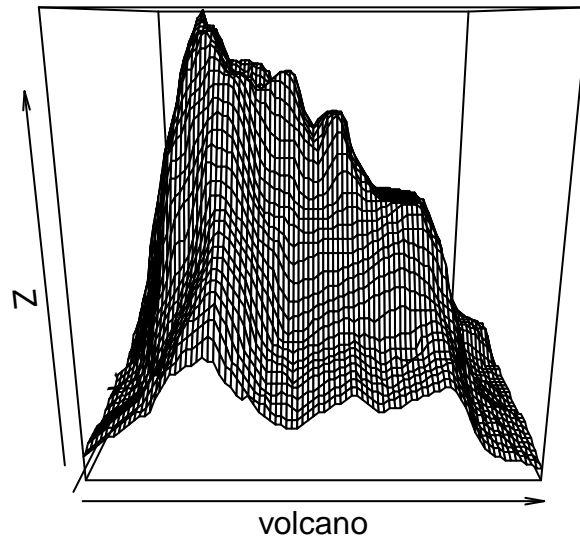

Histogram of goatsF\$weight



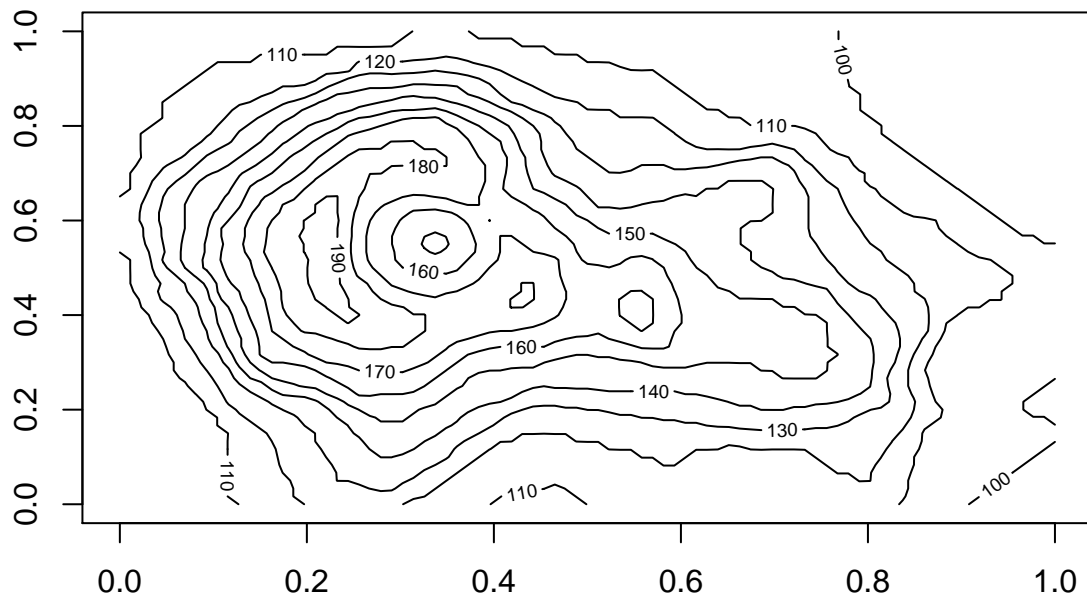
Exercise 2

The volcano data is included with R.

```
persp(volcano)
```

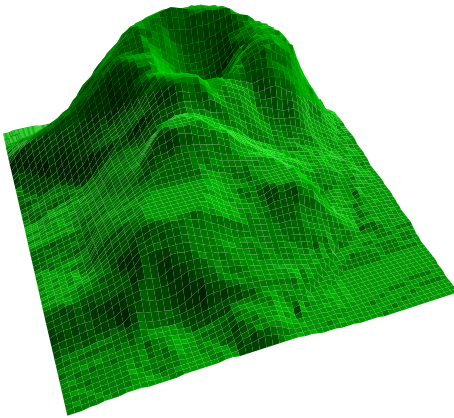


```
contour(volcano)
```



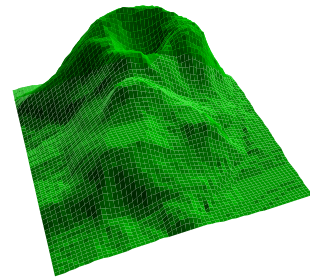
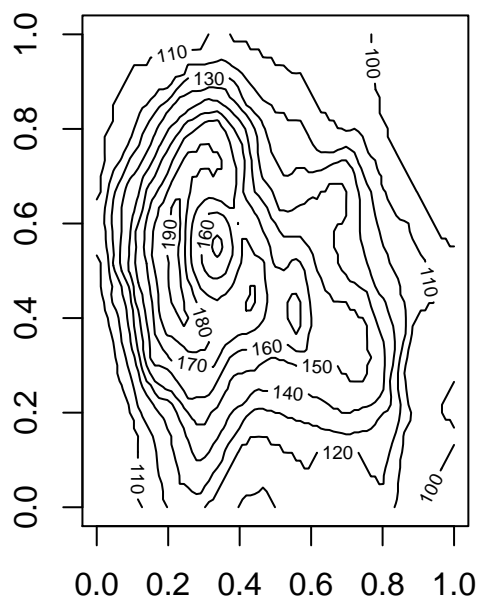
We can get a better perspective on the volcano data by changing the angles that it is viewed at.

```
persp(volcano,theta=70,phi=40,col ="green",box=FALSE,expand=0.5,border=NA,shade=0.75)
```



We can use `mfrow` in the `par` command to place plots next to each other

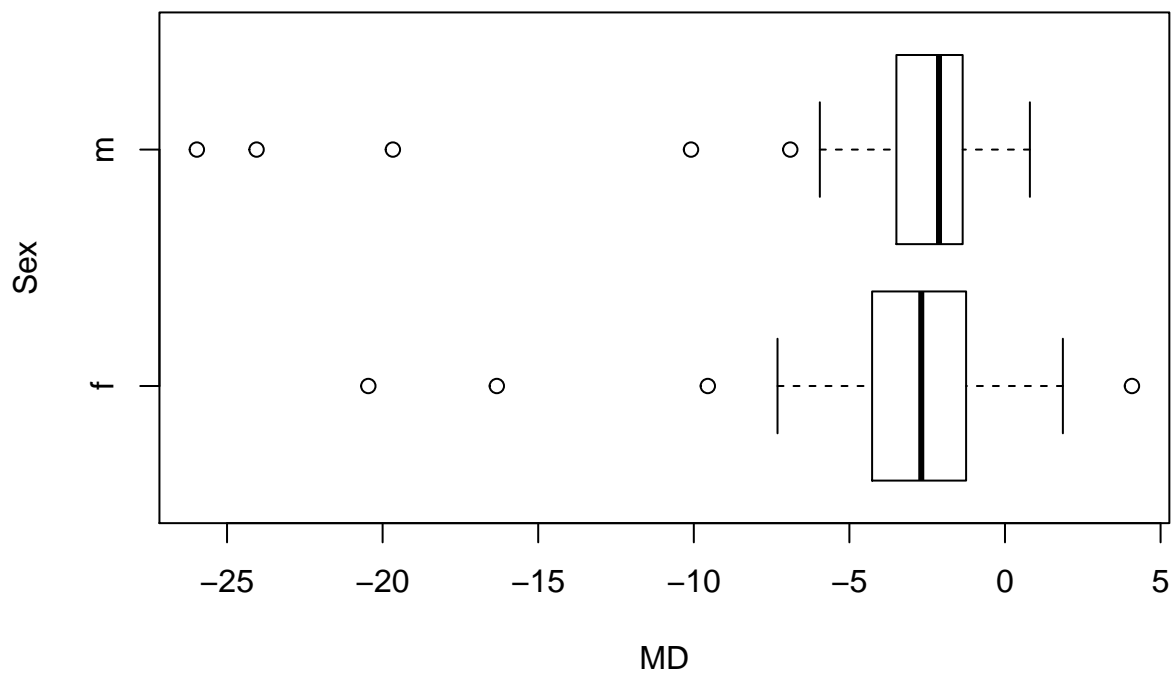
```
par(mfrow=c(1,2))  
contour(volcano)  
persp(volcano,theta=70,phi=40,col ="green",box=FALSE,expand=0.5,border=NA,shade=0.75)
```



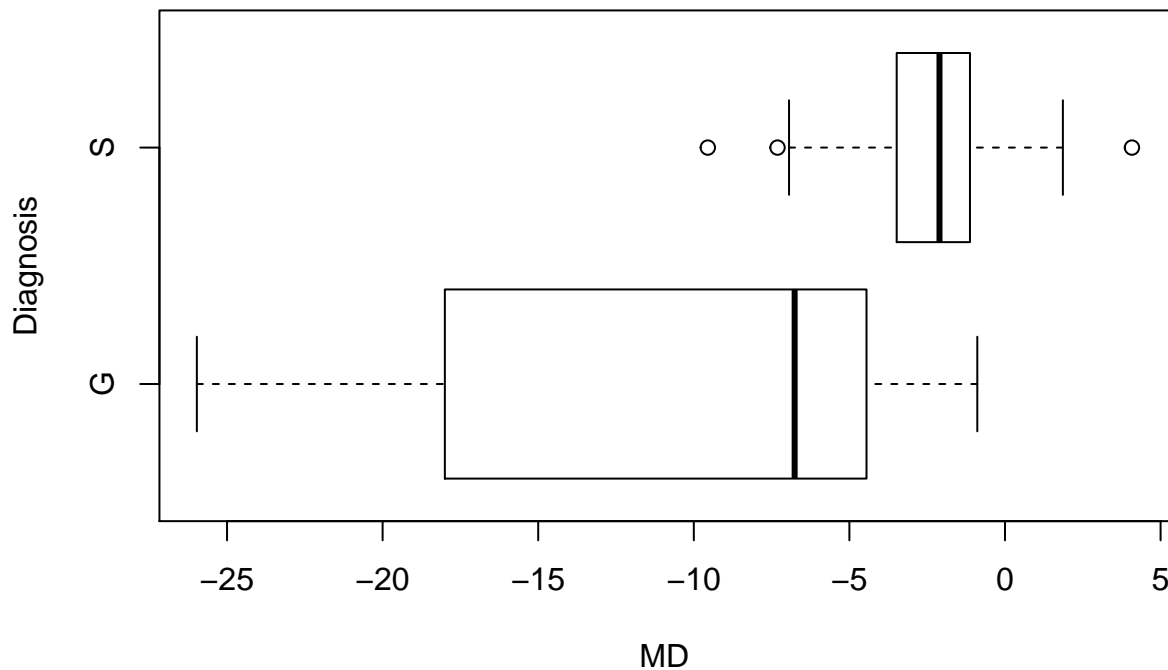
Exercise 3

Boxplots of the glaucoma data may be done using

```
glaucoma = read.csv("http://www.massey.ac.nz/~jcmarsha/rcourse/glaucoma.csv")
plot(MD ~ Sex, data=glaucoma, horizontal=TRUE)
```



```
plot(MD ~ Diagnosis, data=glaucoma, horizontal=TRUE)
```



By the looks the variance in the G diagnosis is larger than the S diagnosis, suggesting more variance in the lower MD scores.

We'll ignore this for now and fit a two-way analysis of variance.

```
mod = lm(MD ~ Sex*Diagnosis, data=glaucoma)
anova(mod)
```

```
## Analysis of Variance Table
##
## Response: MD
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Sex      1   18.75    18.75   1.3498  0.24822
## Diagnosis 1  827.54   827.54  59.5886 1.17e-11 ***
## Sex:Diagnosis 1   64.16    64.16   4.6200  0.03414 *
## Residuals 95 1319.31    13.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the anova table we see that diagnosis and the interaction of diagnosis and sex are significant thus sex must be important. The summary table will give us the effects

```
summary(mod)
```

```
##
## Call:
## lm(formula = MD ~ Sex * Diagnosis, data = glaucoma)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -13.1771 -1.2944  0.4516  1.4856  11.1229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -8.472      1.318  -6.430 5.09e-09 ***
## Sexm             -4.320      1.929  -2.240  0.0274 *
## DiagnosisS        6.032      1.415   4.262 4.77e-05 ***
## Sexm:DiagnosisS   4.519      2.103   2.149  0.0341 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.727 on 95 degrees of freedom
## Multiple R-squared:  0.4083, Adjusted R-squared:  0.3896
## F-statistic: 21.85 on 3 and 95 DF,  p-value: 7.592e-11
```

Here we see that males tend to have a lower MD in the G diagnosis compared to females (4.32 units lower) while in the S diagnosis they are similar to females (as the 4.5 increase from the interaction balances the 4.3 unit decrease from earlier). Females have MD 6 units higher under S diagnosis than G.

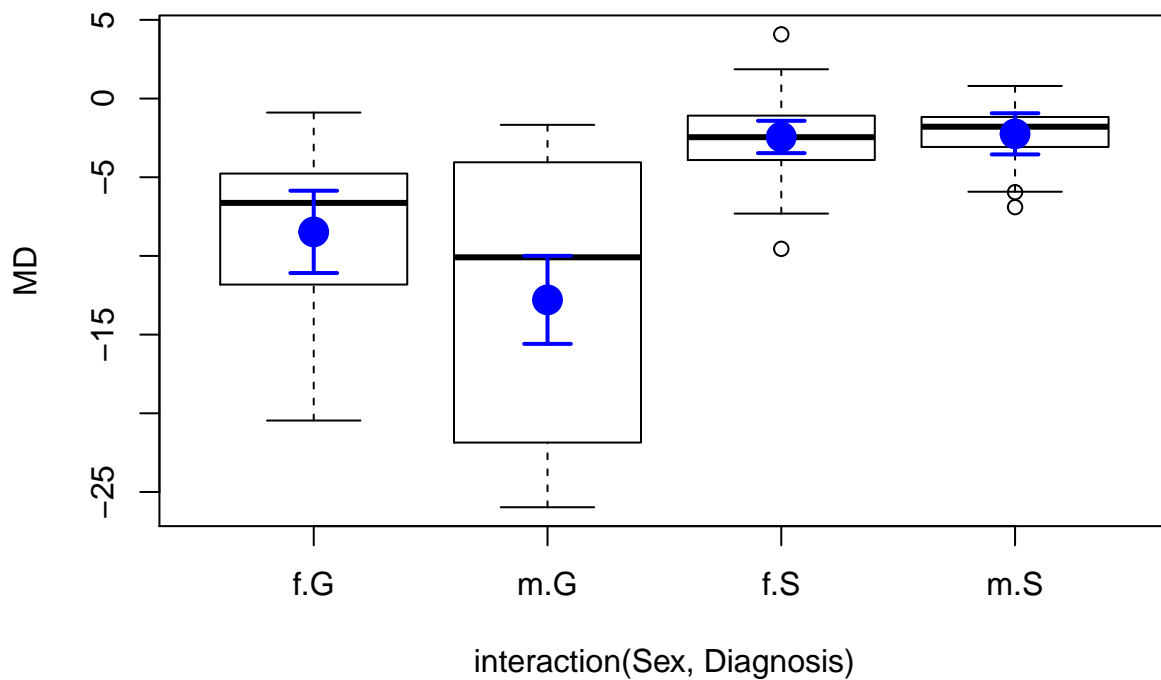
We'll now try and plot the model fit on top of the data. We begin by plotting MD by both Sex and Diagnosis. We do this using the `interaction` command to combine them, and then computing fits and confidence intervals

```
new_data = expand.grid(Sex=c("f","m"),Diagnosis=c("G","S"))
pred = data.frame(predict(mod, new_data, interval="confidence"))
pred
```

```
##      fit      lwr      upr
## 1  -8.472500 -11.088166 -5.8568340
## 2 -12.792857 -15.589122 -9.9965926
## 3  -2.440577  -3.466526 -1.4146283
## 4  -2.241562  -3.549395 -0.9337295
```

We can plot these using

```
plot(MD ~ interaction(Sex, Diagnosis), data=glaucoma)
points(1:4, pred$fit, col="blue", pch=19, cex=2)
segments(1:4, y0=pred$lwr, y1=pred$upr, col="blue", lwd=2)
segments(1:4+0.1, pred$lwr, x1=1:4+0.1, col="blue", lwd=2)
segments(1:4+0.1, pred$upr, x1=1:4+0.1, col="blue", lwd=2)
```

Alternatively we could use the visreg package

```
library(visreg)
```

```
## Warning: package 'visreg' was built under R version 3.4.3
```

```
par(mfrow=c(1,2))
```

```
visreg(mod)
```

```
## Warning: Note that you are attempting to plot a 'main effect' in a model that contains an
## interaction. This is potentially misleading; you may wish to consider using the 'by'
## argument.
```

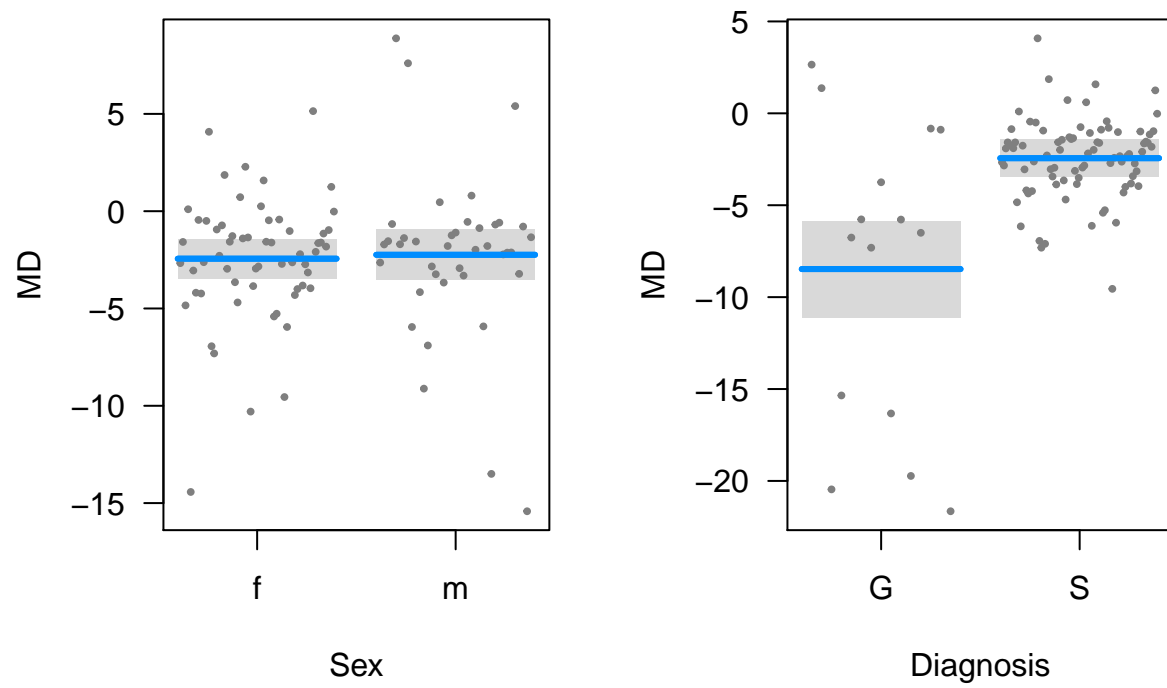
```
## Conditions used in construction of plot
```

```
## Diagnosis: S
```

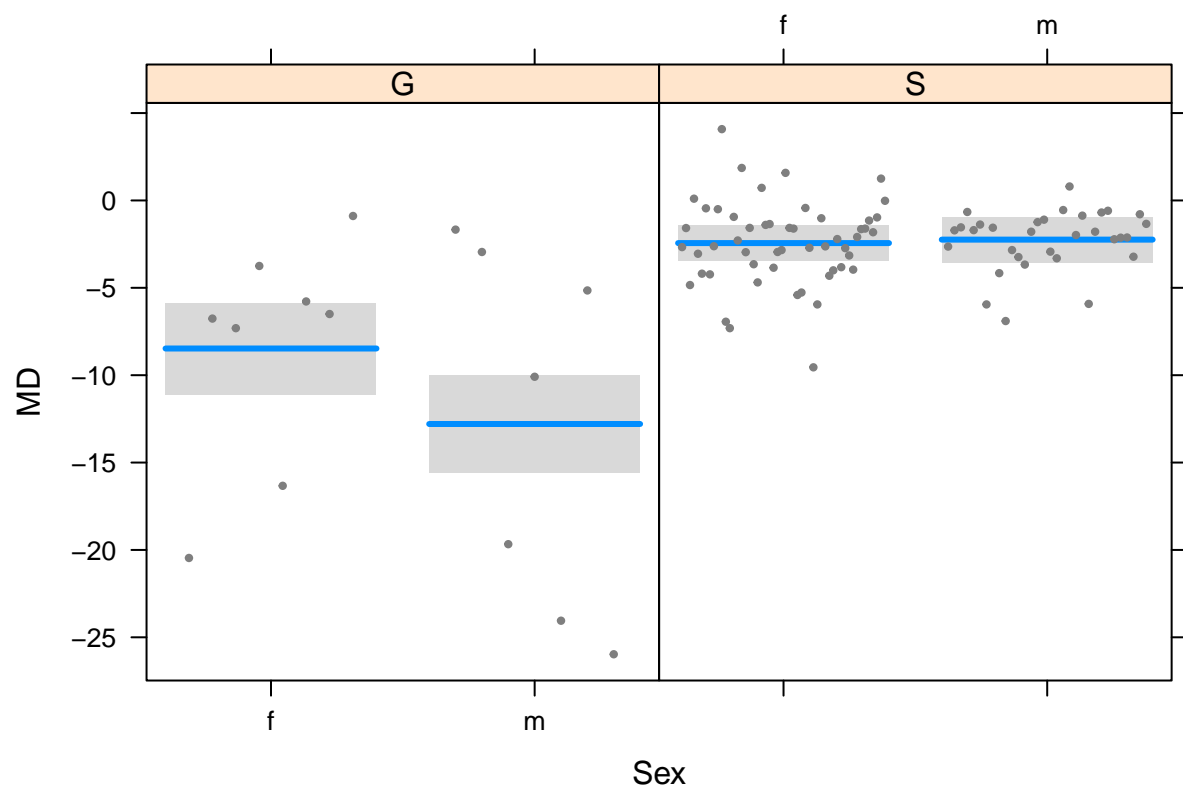
```
## Warning: Note that you are attempting to plot a 'main effect' in a model that contains an
## interaction. This is potentially misleading; you may wish to consider using the 'by'
## argument.
```

```
## Conditions used in construction of plot
```

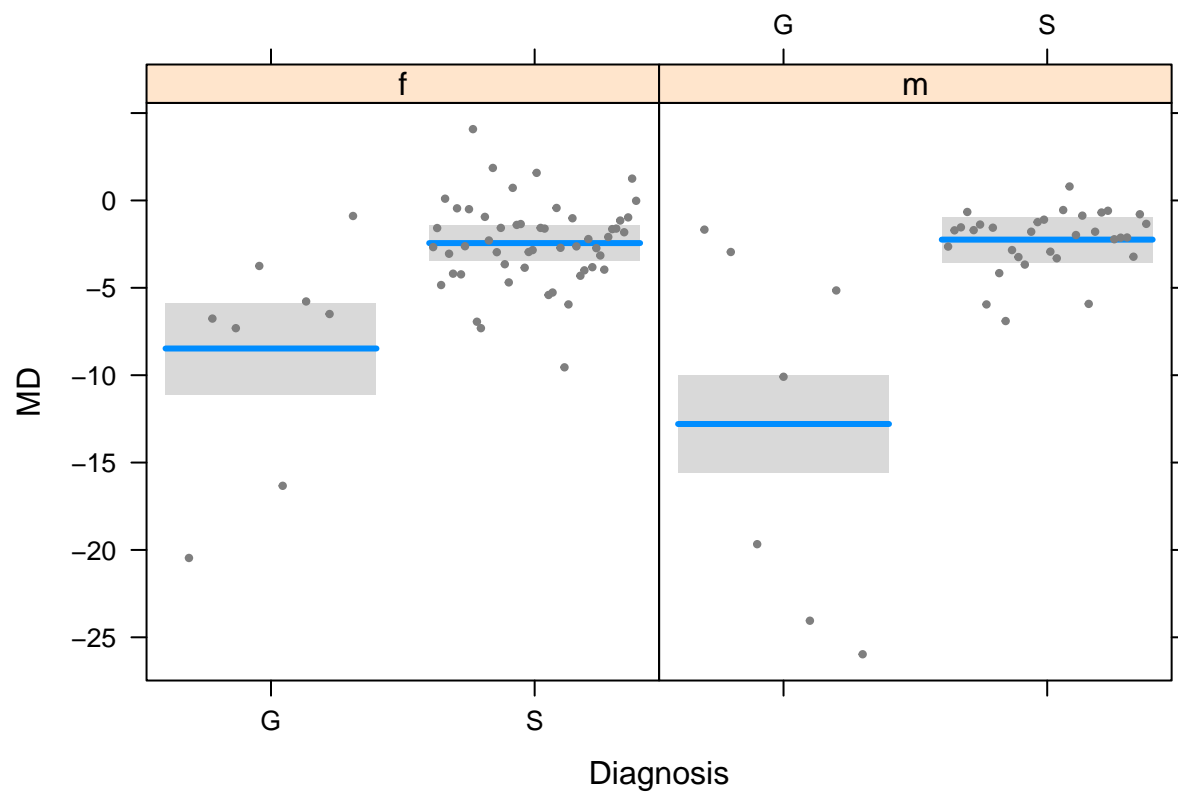
```
## Sex: f
```



```
visreg(mod, "Sex", by="Diagnosis")
```

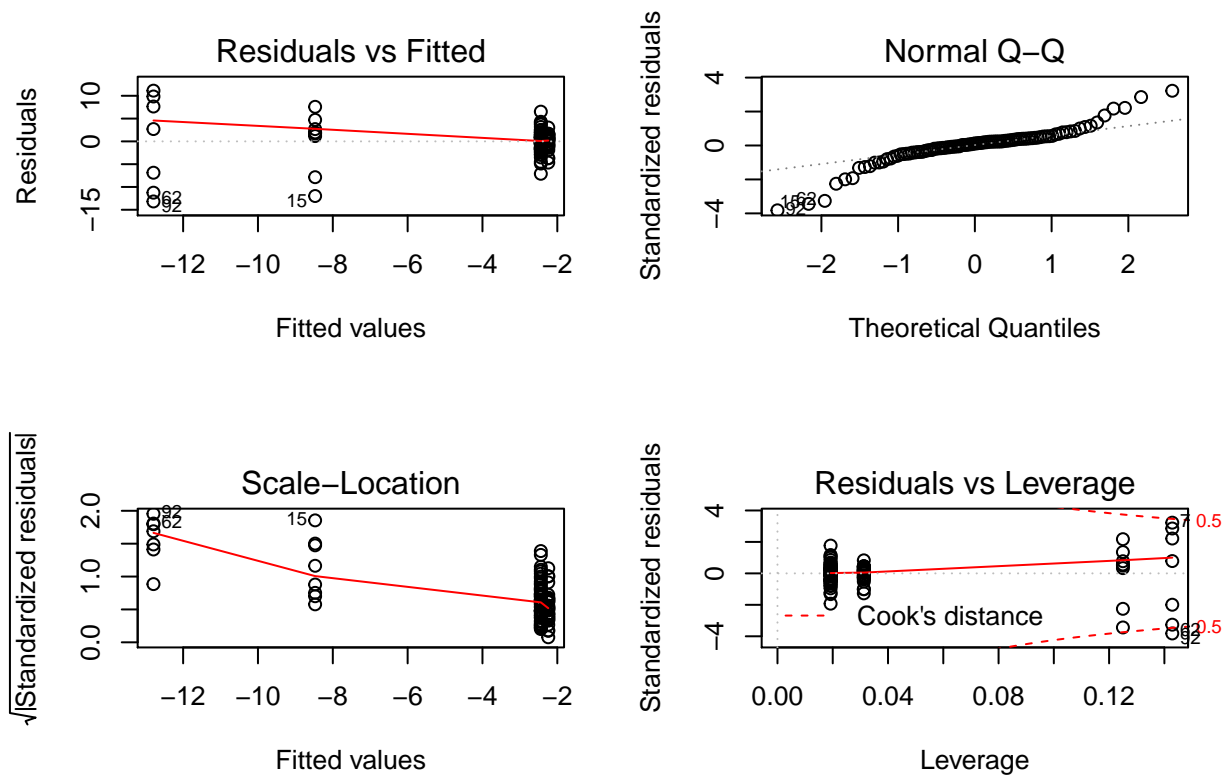


```
visreg(mod, "Diagnosis", by="Sex")
```



In the first plot we have just the sex effects plotted for the S diagnosis group. In the second we have just the diagnosis effect plotted for the females. In the last two we have diagnosis by sex and sex by diagnosis. The diagnostics suggest that the model isn't a good fit as there is heteroskedasticity

```
par(mfrow=c(2,2))
plot(mod)
```



Repeating with signed square-root transformed data

```
glaucoma$MDtr = sign(glaucoma$MD)*sqrt(abs(glaucoma$MD))
mod2 = lm(MDtr ~ Sex*Diagnosis, data=glaucoma)
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: MDtr
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Sex         1  1.217    1.217   1.2259    0.2710
## Diagnosis   1 33.668   33.668  33.9217 7.789e-08 ***
## Sex:Diagnosis 1  0.731    0.731   0.7360    0.3931
## Residuals   95 94.290    0.993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With this model the interaction isn't important, and nor is Sex Let's simplify it

```
mod4 = lm(MDtr ~ Diagnosis, data=glaucoma)
anova(mod4)
```

```
## Analysis of Variance Table
##
## Response: MDtr
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Diagnosis   1 34.343   34.343   34.86 5.222e-08 ***
## Residuals  97 95.562    0.985
```

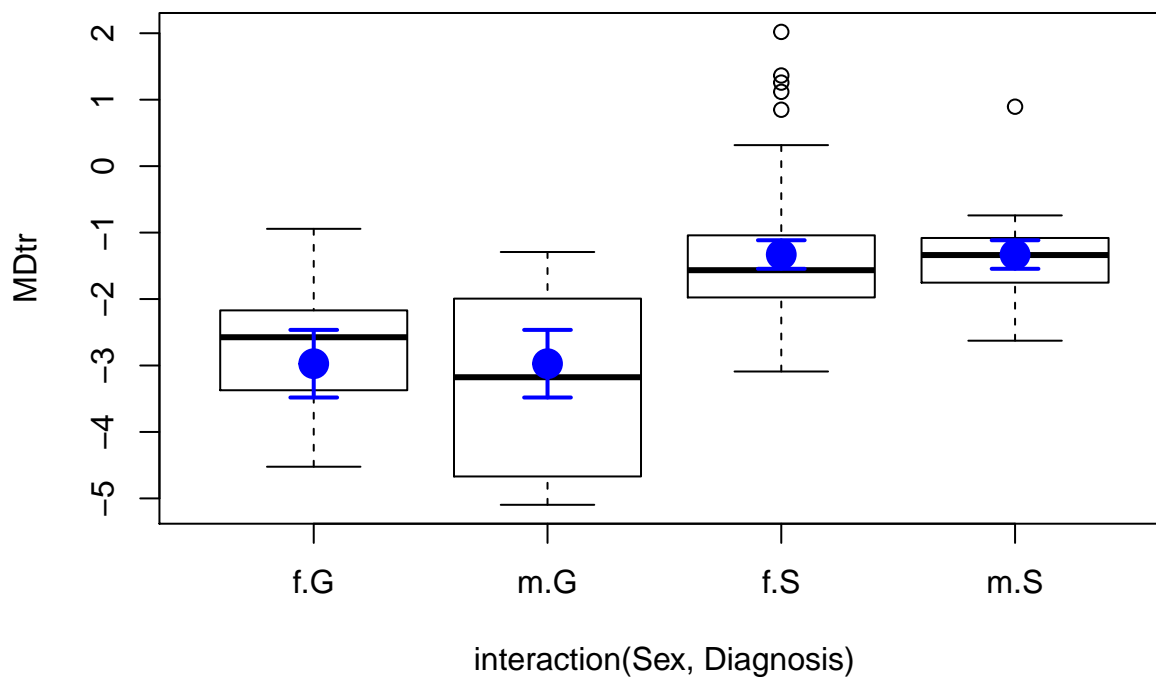
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To plot the model fit then we do it in the same way as before

```
new_data = expand.grid(Sex=c("f","m"),Diagnosis=c("G","S"))
pred = data.frame(predict(mod4, new_data, interval="confidence"))
```

We can plot these using

```
plot(MDtr ~ interaction(Sex, Diagnosis), data=glaucoma)
points(1:4, pred$fit, col="blue", pch=19, cex=2)
segments(1:4, y0=pred$lwr, y1=pred$upr, col="blue", lwd=2)
segments(1:4-0.1, pred$lwr, x1=1:4+0.1, col="blue", lwd=2)
segments(1:4-0.1, pred$upr, x1=1:4+0.1, col="blue", lwd=2)
```



This plot suggests little difference between the sexes.