

# Chapter 1

## Basic graphics

### 1.1 Introduction to graphics with R

There are a lot of nice functions for graphics in R. In this chapter we use some of the most common functions in base R. There are more advanced possibilities in the `ggplot2` package which we will see later on. However, also the plot functions in this chapter have a lot of options. It may be good to know about both worlds - sometimes you need quick graphics to illustrate your data and other times you need a nice layout or more complex illustration and then it is worth spending somewhat more time on how to do it.

### 1.2 Histogram

Let us import the data file `norsjo86` and do a histogram for body height. The setting `par(mfrow=c(2,2))` is used to divide the plot area in 4 cells (2 rows and 2 columns).

agegrp:	Age group	(30, 40, 50 ,60 years)
health:	Health status	(0=good, 1=not quite good/bad)
sex:	Sex	(1=man, 2=woman)
height:	Body height	(cm)
weight:	Body weight	(kg)
sbp:	Systolic blood pressure	
dbp:	Diastolic blood pressure	
cholesterol:	Cholesterol	
smoker:	Smoking status	(0=non-smoker, 1=smoker)
bmi:	Body mass index	( $kg/m^2$ )

```
# import spss file norsjo86
library(haven)
nors <- read_sav("../data/norsjo86.sav")
nors <- as.data.frame(nors)
head(nors)
```

	agegrp	health	sex	height	weight	sbp	dbp	cholesterol	smoker	bmi
1	60	0	2	157	61	110	70	6.7	0	24.74745
2	60	1	2	157	97	150	100	6.6	0	39.35251
3	60	0	1	170	74	136	96	8.2	0	25.60554
4	60	0	2	163	66	156	76	7.5	0	24.84098

```

5      60      0      2      166      66 110      70      10.2      0 23.95123
6      60      0      2      168      61 130      78      7.3      0 21.61281

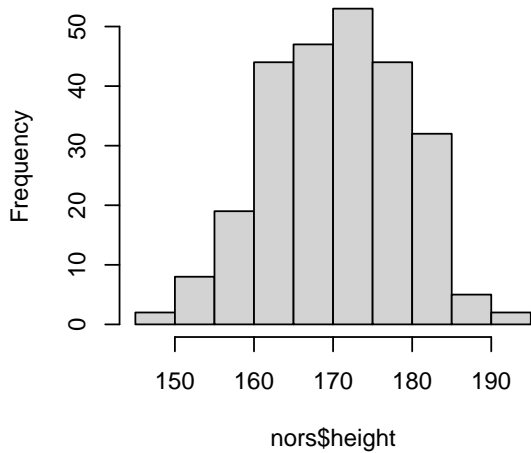
par(mfrow=c(2,2))
hist(nors$height)
# hist(nors$height)$breaks
# str(hist(nors$height))
range(nors$height,na.rm=T)

[1] 145 191

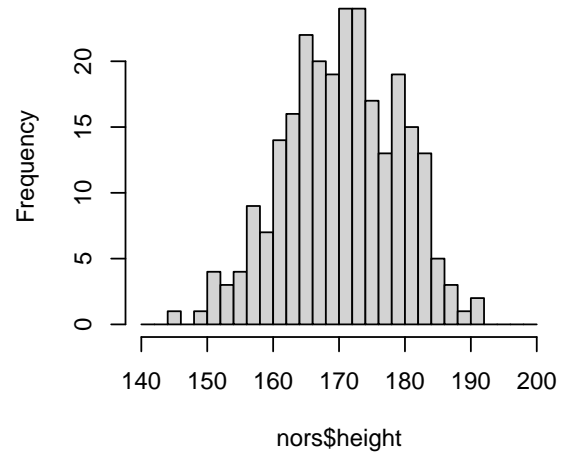
hist(nors$height,breaks=seq(140,200,by=2))
hist(nors$height,breaks=seq(140,200,by=10),col=3,density=30)
hist(nors$height,xlim=c(100,200),freq=F,xlab="Body height",main="Histogram")

```

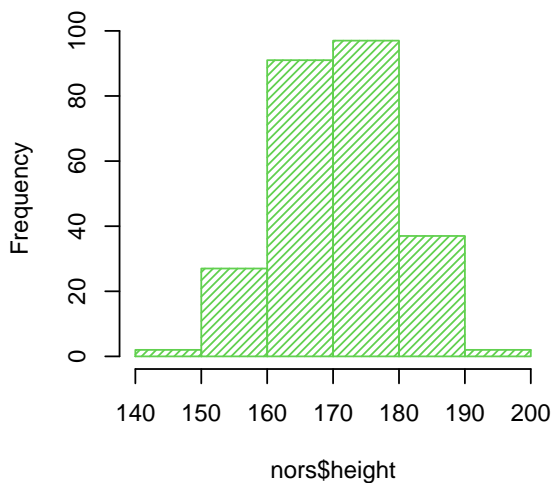
Histogram of nors\$height



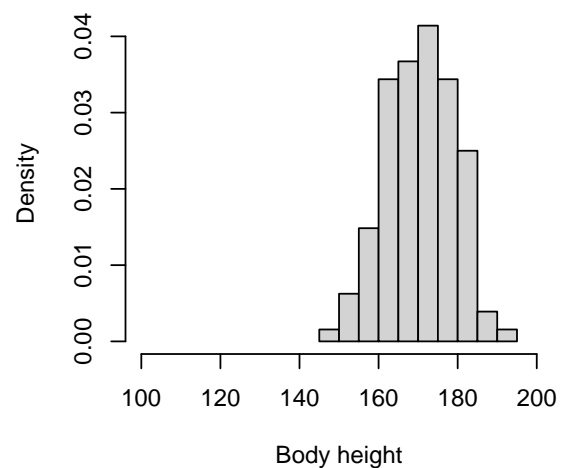
Histogram of nors\$height



Histogram of nors\$height



Histogram



```
# freq=F the total area is equal to one
```

## 1.3 Boxplot

Now let us look at boxplots for cholesterol by sex and age group. The plots look better if you zoom the result (plots tab).

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

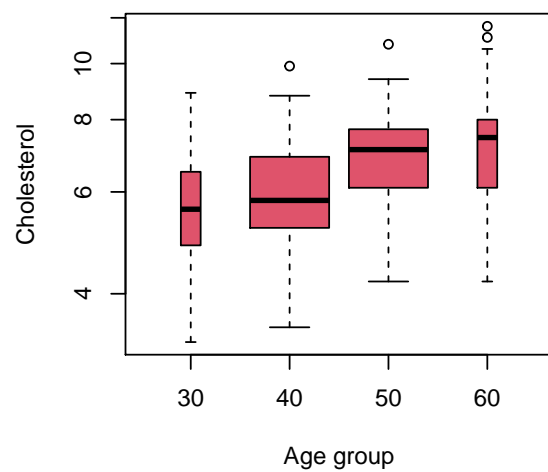
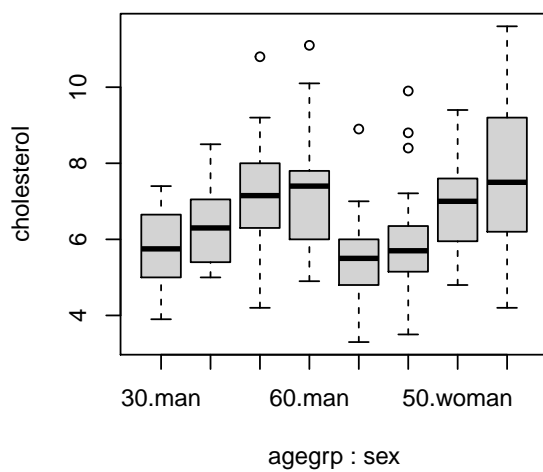
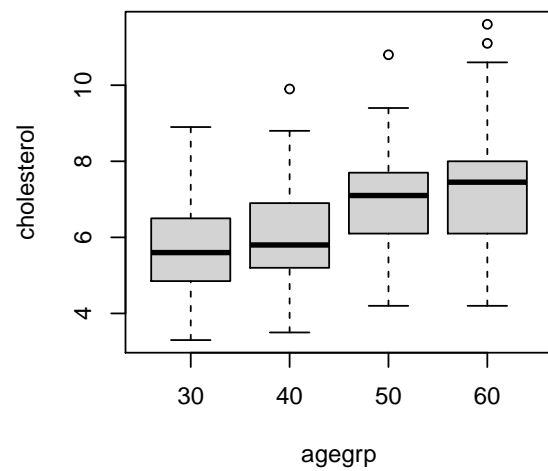
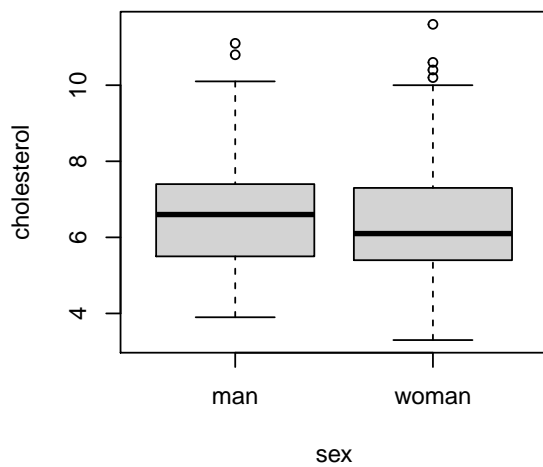
nors$sex<-factor(nors$sex,labels=c("man","woman"))
# make a factor and set labels

boxplot(cholesterol~sex,data=nors)

boxplot(cholesterol~agegrp,data=nors)

boxplot(cholesterol~agegrp+sex,data=nors) # combining two categorical variables

# Example of some extra options
boxplot(cholesterol~agegrp,data=nors,col=2,ylab="Cholesterol",
        xlab="Age group",width=c(0.1,0.4,0.4,0.1),log="y")
```



## 1.4 Scatterplot

Scatterplots are very useful for illustrating relationships between two continuous variables. They can also be used to illustrate trends e.g. over time. There is a standard alternative how to give the x and y variables but it is usually more convenient to use the formula object alternative.

```
# two similar alternatives with different symbols
par(mfrow=c(2,2))
plot(nors$bmi,nors$cholesterol)
plot(cholesterol~bmi,data=nors) # using formula object

# Example of some extra options
plot(cholesterol~bmi,data=nors,col=2,pch=2,cex=0.7,ylab="Cholesterol",
      xlab="Body mass index (BMI)",ylim=c(0,15),xlim=c(10,50))

# two groups in the same plot
plot(height~bmi,data=subset(nors,sex=="man"),pch=16,main="Figure 4")
points(height~bmi,data=subset(nors,sex=="woman"),pch=2,col=2)
legend(x=33,y=185,legend=c("Men","Women"),col=c(1,2),pch=c(16,2),cex=0.7)
```

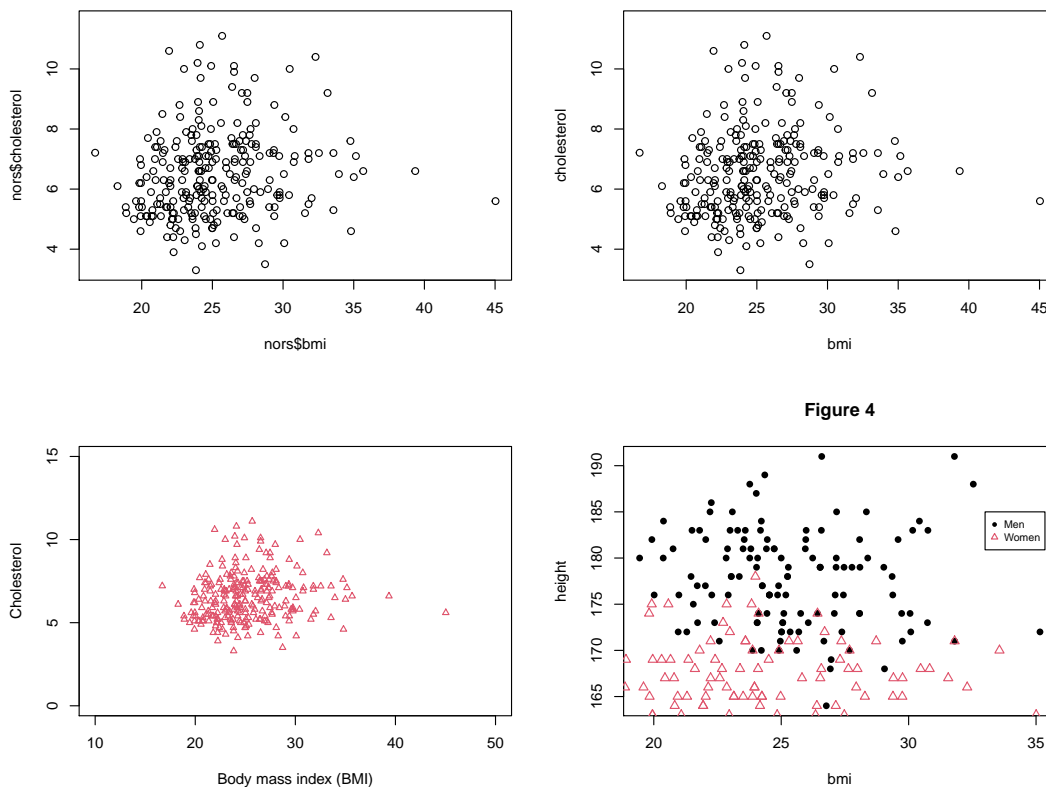


Figure 1.1: Scatterplots

We can also use scatterplot to do plots based on mathematical functions.

```
x<-seq(-20,20,by=0.001)

y1<-10*x^2-100
plot(x,y1,type="l",ylim=c(-5000,5000),ylab="y")

y2<-5+-x-3*x^2+x^3
lines(x,y2,col=2,lty=2,lwd=2)

y3<-sin(x)*1000
lines(x,y3,col=3,lwd=3)

abline(h=0,lty=2) # adding a horizontal line
```

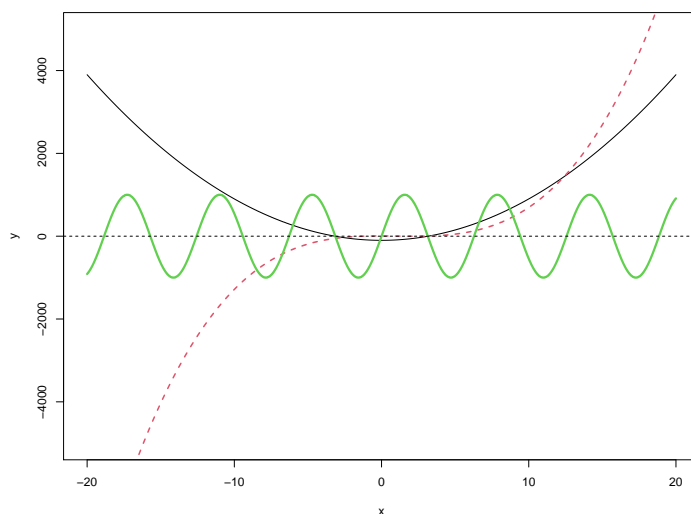


Figure 1.2: Plot of mathematical functions

### Own experimentation

The plot with heading Figure 4 does not show all points - why? Also the legend need to be adjusted (it may look ok in this layout but not if it is plotted using the script.R file). Can you fix it? Try to plot some own mathematical functions.

## 1.4.1 Graphical parameters

You have seen that for many functions there are a varying number of arguments, some important and some more optional with default values we may not be aware of. Plot functions usually have a large number of arguments. The argument parameters can be set globally by `par()` but usually they are set in the respective plot function. However you can find a list of these options in the help for `par` under Graphical Parameters.