Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019) **Solutions to Exercises of Chapter 2: Looking at Data**

Kimmo Vehkalahti and Brian S. Everitt

8 October 2018

Solutions

Exercise 2.1

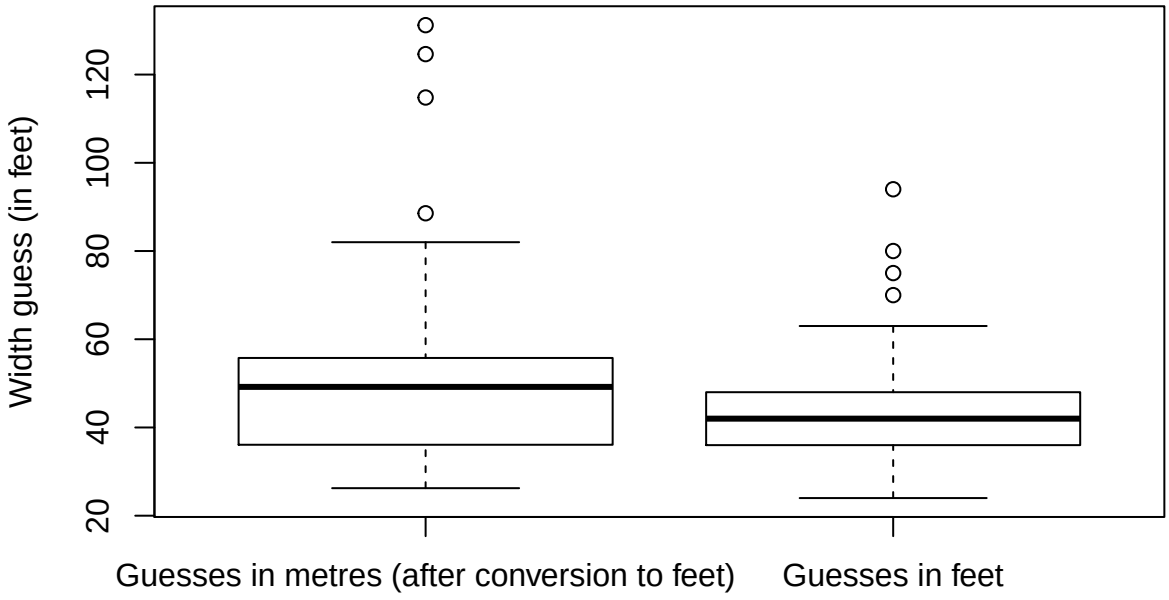
Exercise 2.2

Use the lengths data (see below).

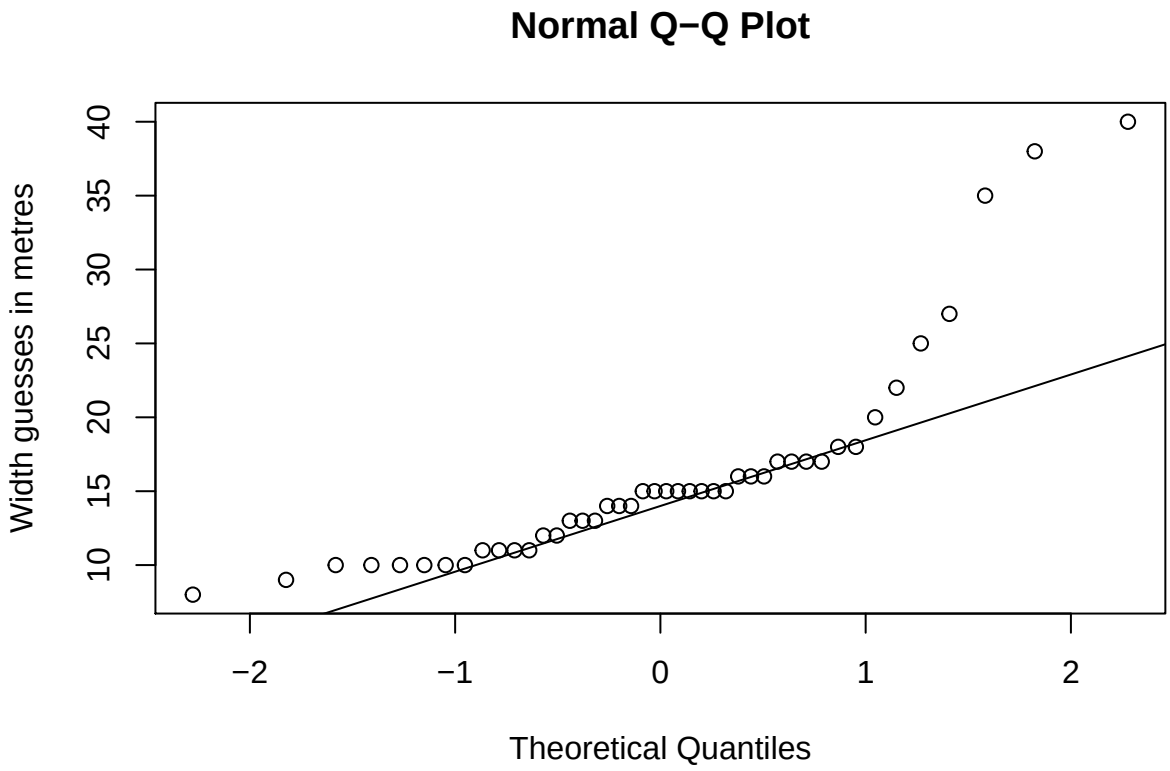
```
lengths <- structure(list(group = structure(

c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L),

.Label = c("Metres", "Feet"), class = "factor"),

guesses = c(8, 9, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 12, 12, 13,
13, 13, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 17, 17,
17, 17, 18, 18, 20, 22, 25, 27, 35, 38, 40, 24, 25, 27, 30, 30, 30, 30,
30, 30, 32, 32, 33, 34, 34, 34, 35, 35, 36, 36, 36, 37, 37, 40, 40, 40,
40, 40, 40, 40, 40, 40, 41, 41, 42, 42, 42, 42, 43, 43, 44, 44, 44, 45,
45, 45, 45, 45, 45, 46, 46, 47, 48, 48, 50, 50, 50, 51, 54, 54, 54, 55,
55, 60, 60, 63, 70, 75, 80, 94)),

.Names = c("group", "guesses"), row.names = c(NA, -113L), class = "data.frame")

head(lengths)
```

```
##    group guesses
## 1 Metres       8
## 2 Metres       9
## 3 Metres      10
## 4 Metres      10
## 5 Metres      10
## 6 Metres      10
```

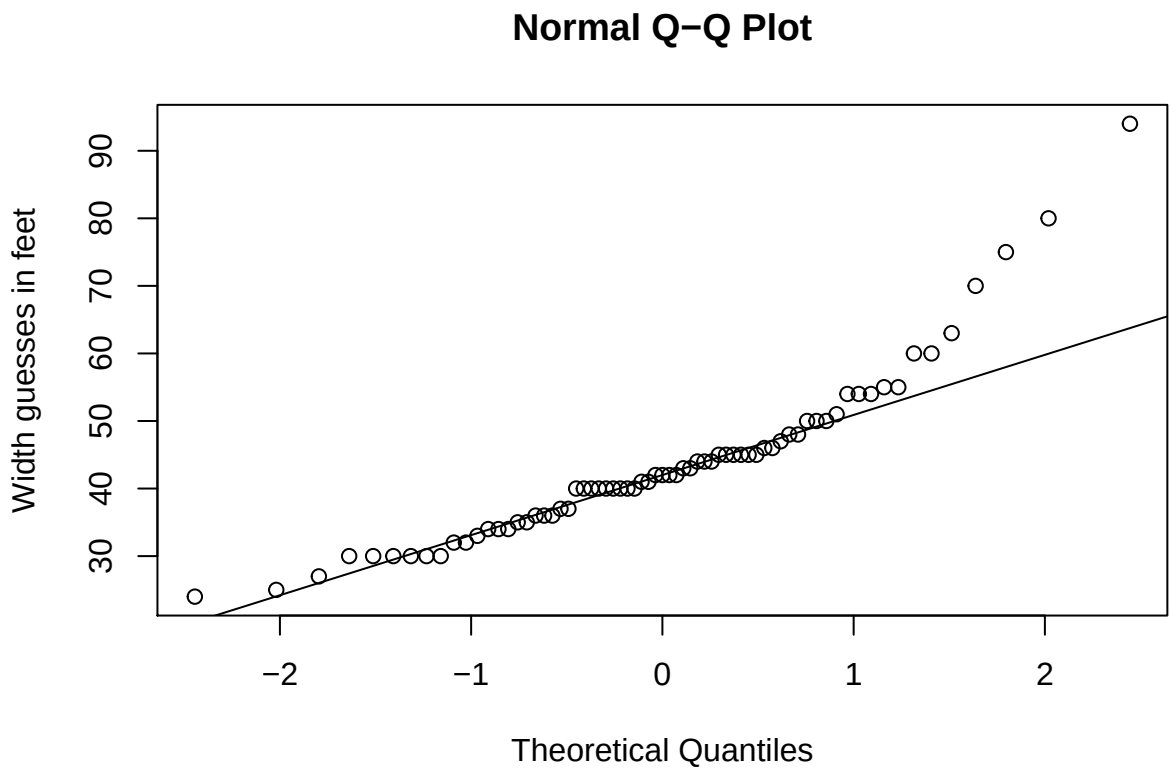
```r
# produce side-by-side boxplots and normal probability plots
# after converting guess in meters to feet
#
attach(lengths)
feet <- group == "Feet"
convert <- ifelse(feet, 1, 3.28)
# layout(matrix(c(1,2,1,3), nrow = 2, ncol = 2, byrow = FALSE))
boxplot(I(guesses * convert) ~ group, ylab = "Width guess (in feet)", var.width = TRUE,
        names = c("Guesses in metres (after conversion to feet)", "Guesses in feet"))
```



```r
qqnorm(guesses[!feet], ylab = "Width guesses in metres")
qqline(guesses[!feet])
```



```r
qqnorm(guesses[feet], ylab = "Width guesses in feet")
qqline(guesses[feet])
```




```r
detach(lengths)
```

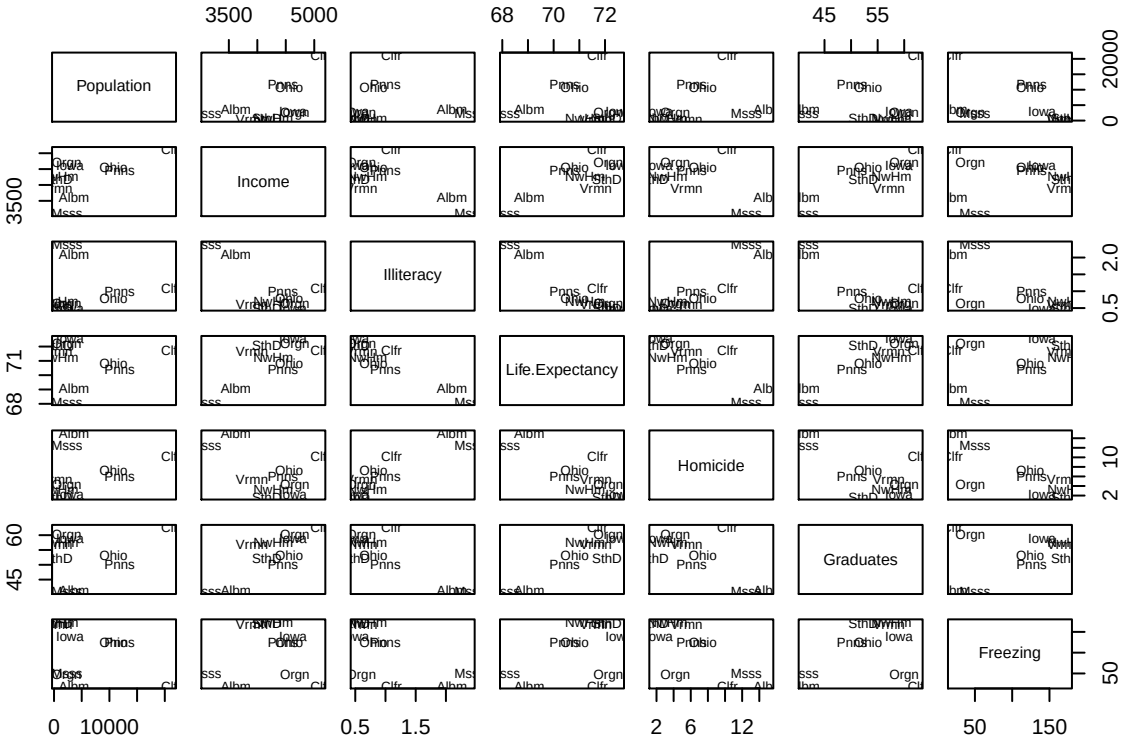
Exercise 2.3

Use the states data (see below).

```
states <- structure(list(

Population = c(3615, 21198, 2861, 2341, 812, 10735, 2284, 11860, 681, 472),

Income = c(3624, 5114, 4628, 3098, 4281, 4561, 4660, 4449, 4167, 3907),

Illiteracy = c(2.1, 1.1, 0.5, 2.4, 0.7, 0.8, 0.6, 1, 0.5, 0.6),

Life.Expectancy = c(69.05, 71.71, 72.56, 68.09, 71.23, 70.82, 72.13, 70.43, 72.08, 71.64),

Homicide = c(15.1, 10.3, 2.3, 12.5, 3.3, 7.4, 4.2, 6.1, 1.7, 5.5),

Graduates = c(41.3, 62.6, 59, 41, 57.6, 53.2, 60, 50.2, 52.3, 57.1),

Freezing = c(20, 20, 140, 50, 174, 124, 44, 126, 172, 168)),

.Names = c("Population", "Income", "Illiteracy", "Life.Expectancy", "Homicide",
           "Graduates", "Freezing"),

row.names = c("Alabama", "California", "Iowa", "Mississippi", "New Hampshire",
              "Ohio", "Oregon", "Pennsylvania", "South Dakota", "Vermont"),
             class = "data.frame")
states
```

```
##               Population Income Illiteracy Life.Expectancy Homicide
## Alabama             3615   3624        2.1           69.05     15.1
## California         21198   5114        1.1           71.71     10.3
## Iowa                2861   4628        0.5           72.56      2.3
## Mississippi         2341   3098        2.4           68.09     12.5
## New Hampshire        812   4281        0.7           71.23      3.3
## Ohio               10735   4561        0.8           70.82      7.4
## Oregon              2284   4660        0.6           72.13      4.2
## Pennsylvania       11860   4449        1.0           70.43      6.1
## South Dakota         681   4167        0.5           72.08      1.7
## Vermont              472   3907        0.6           71.64      5.5
##               Graduates Freezing
## Alabama            41.3       20
## California         62.6       20
## Iowa               59.0      140
## Mississippi        41.0       50
## New Hampshire      57.6      174
## Ohio               53.2      124
## Oregon             60.0       44
## Pennsylvania       50.2      126
## South Dakota       52.3      172
## Vermont            57.1      168
```

```
# pairs plot with abbreviated state names
pairs(states, panel = function(x,y) text(x, y,
                        abbreviate(row.names(states)), cex = 0.6))
```



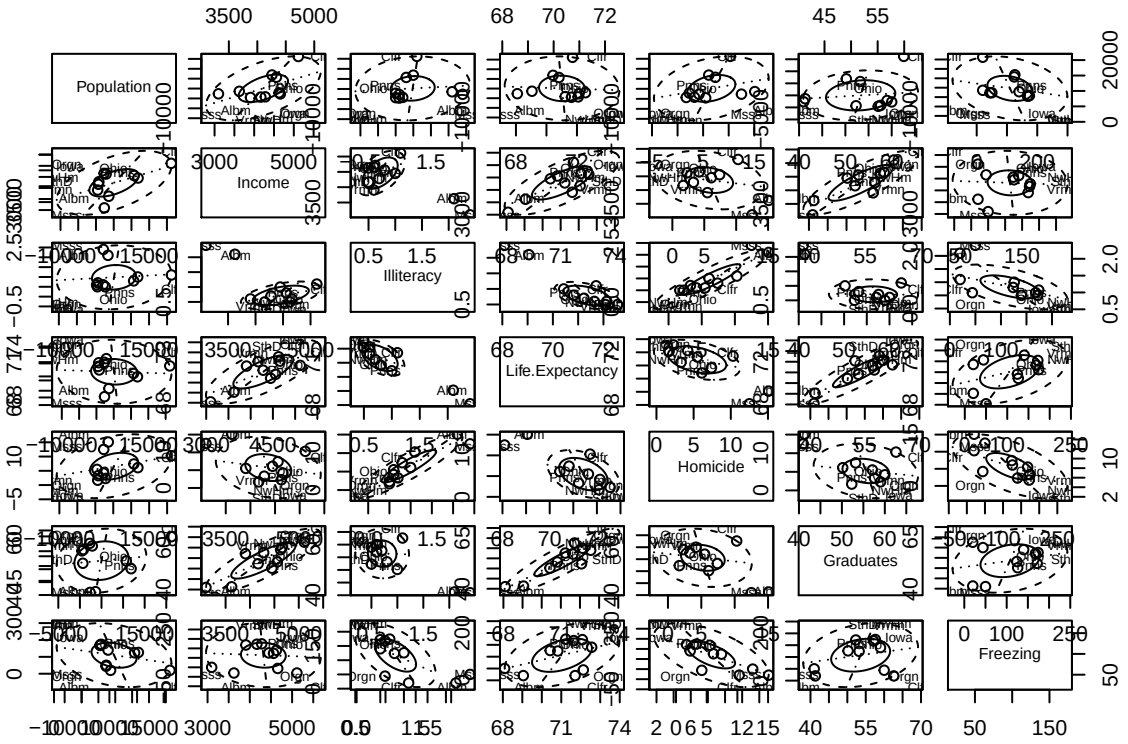
```
#install.packages("MVA")
require("MVA")
```

```
## Loading required package: MVA
```

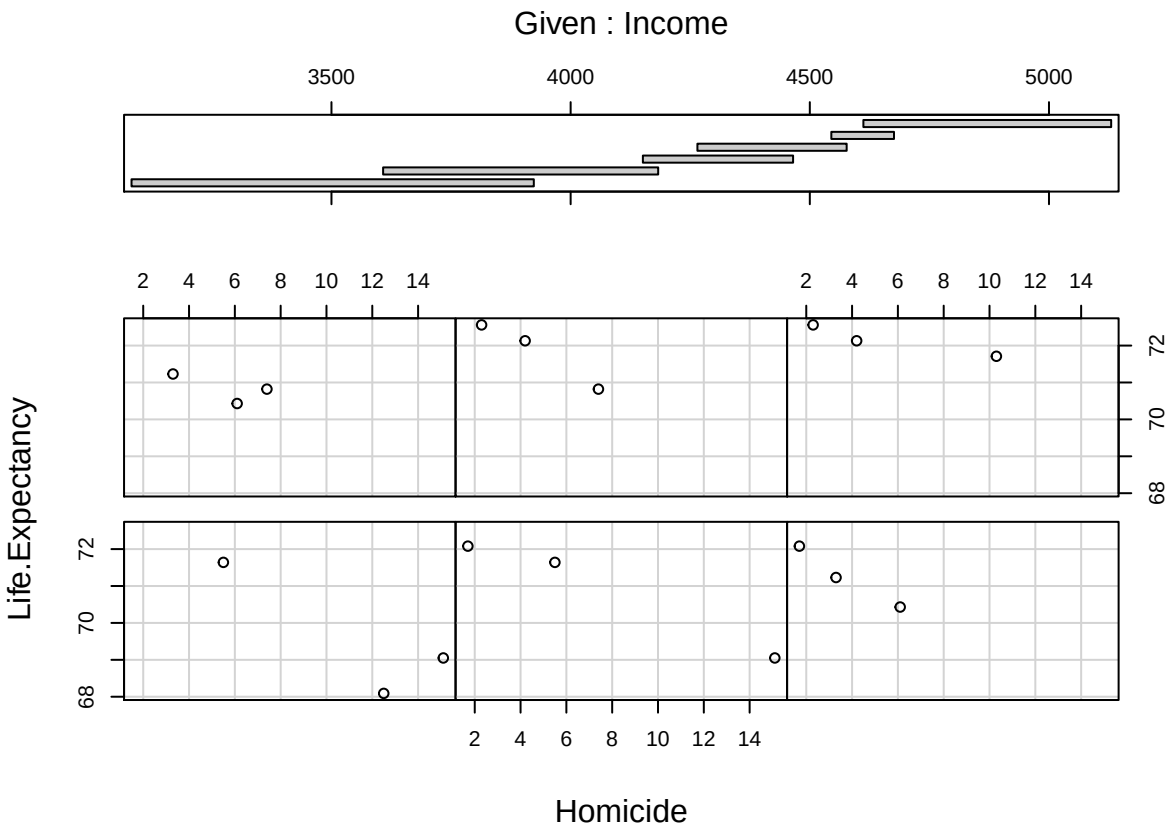
```
## Loading required package: HSAUR2
```

```
## Loading required package: tools
```

```
pairs(states, panel = function(x,y) {
   text(x, y, abbreviate(row.names(states)), cex = 0.6)
   par(new = T)
   bvbox(cbind(x, y), mtitle = "")
   })
```



```
attach(states)
coplot(Life.Expectancy ~ Homicide | Income)
```




```
detach(states)
```

Exercise 2.4

The graph commits the cardinal sin of quoting data out of context; remember that graphics often lie by omission, leaving out data sufficient for comparisons. Here, a few more data points for other years in the area would be helpful, as would similar data for other areas in which stricter enforcement of speeding had not been implemented.

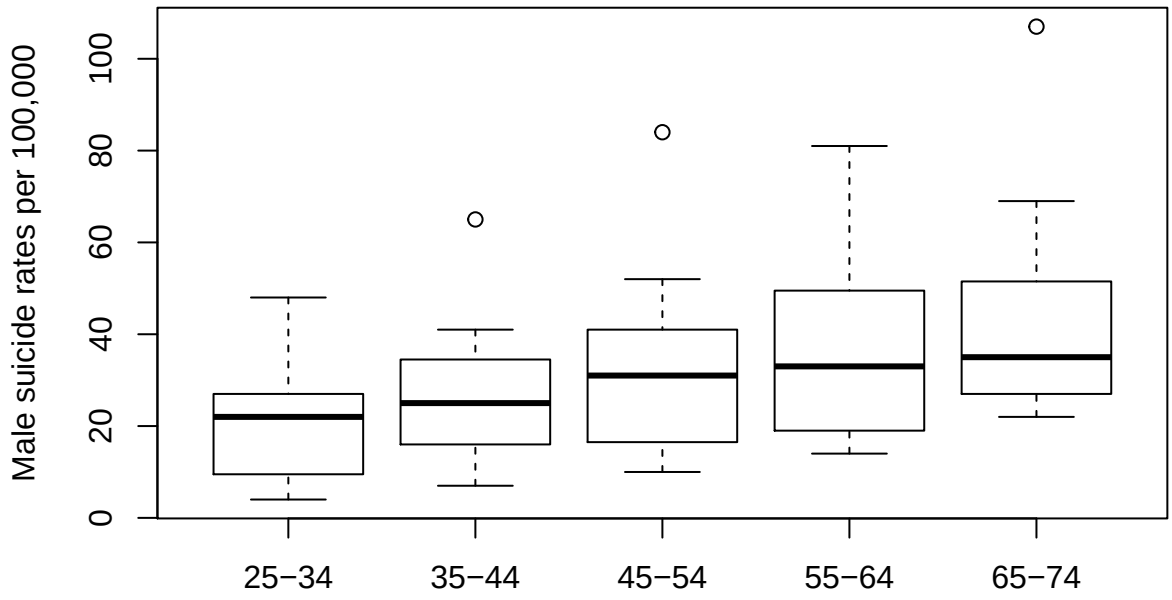
Exercise 2.5

Use the suicides data (see below).

```
suicides <- structure(list(
A25.34 = c(22, 9, 22, 29, 16, 28, 48, 7, 8, 26, 4, 28, 22, 10, 20),
A35.44 = c(27, 19, 19, 40, 25, 35, 65, 8, 11, 29, 7, 41, 34, 13, 22),
A45.54 = c(31, 10, 21, 52, 36, 41, 84, 11, 18, 36, 10, 46, 41, 15, 28),
A55.64 = c(34, 14, 31, 53, 47, 49, 81, 18, 20, 32, 16, 51, 50, 17, 33),
A65.74 = c(24, 27, 49, 69, 56, 52, 107, 27, 28, 28, 22, 35, 51, 22, 37)),

.Names = c("A25.34", "A35.44", "A45.54", "A55.64", "A65.74"),

row.names = c("Canada", "Israel", "Japan", "Austria", "France", "Germany",
              "Hungary","Italy", "Netherlands", "Poland", "Spain", "Sweden",
              "Switzerland", "UK", "USA"),
            class = "data.frame")
suicides
```

```
##             A25.34 A35.44 A45.54 A55.64 A65.74
## Canada          22     27     31     34     24
## Israel           9     19     10     14     27
## Japan           22     19     21     31     49
## Austria         29     40     52     53     69
## France          16     25     36     47     56
## Germany         28     35     41     49     52
## Hungary         48     65     84     81    107
## Italy            7      8     11     18     27
## Netherlands      8     11     18     20     28
## Poland          26     29     36     32     28
## Spain            4      7     10     16     22
## Sweden          28     41     46     51     35
## Switzerland     22     34     41     50     51
## UK              10     13     15     17     22
## USA             20     22     28     33     37
```

```
boxplot(suicides[, 1], suicides[, 2], suicides[, 3], suicides[, 4], suicides[, 5],
        names = c("25-34", "35-44", "45-54", "55-64", "65-74"),
        ylab = "Male suicide rates per 100,000")
```



A set of boxplots for the different countries might also be interesting.