Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019)

Examples of Chapter 2: Looking at Data

Kimmo Vehkalahti and Brian S. Everitt

8 October 2018

Contents

Examples

Table 2.1: Crime Rates for Drinkers and Abstainers

```
drink <- c(50, 88, 155, 379, 81)
abst <- c(43, 62, 110, 300, 158)
drink <- 100*drink/sum(drink)
abst <- 100*abst/sum(abst)
da <- cbind(drink, abst)
da
```

```
##          drink      abst
## [1,]  6.640106  6.389302
## [2,] 11.686587  9.212481
## [3,] 20.584329 16.344725
## [4,] 50.332005 44.576523
## [5,] 10.756972 23.476969
```

Figure 2.1

```
labels <- c("Arson", "Rape", "Violence", "Stealing", "Fraud")
par(mfrow = c(1,2))
pie(drink, density = -10, labels = labels, col = 0, lwd = 2)
title("Drinkers")
pie(abst, density = -10, labels = labels, col = 0, lwd = 2)
title("Abstainers")
```

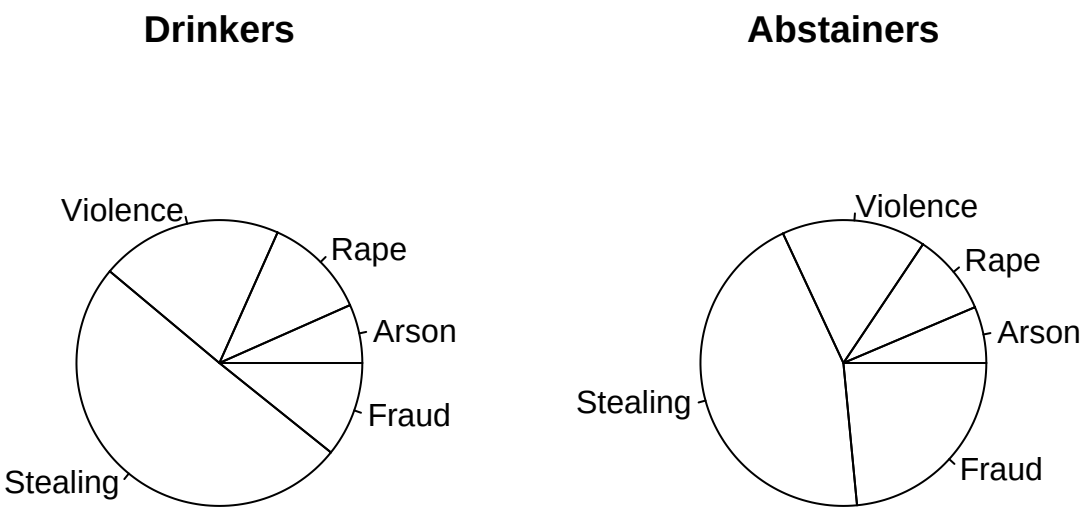


Figure 2.2

```
par(mfrow = c(1,2))
labs <- c("A", "R", "V", "S", "F")
barplot(drink, names.arg = labs, col = 1, lwd = 2, ylab = "Percent", ylim = c(0,55))
title("Drinkers")
barplot(abst, names.arg = labs, col = 1, lwd = 2, ylab = "Percent", ylim = c(0,55))
title("Abstainers")
```

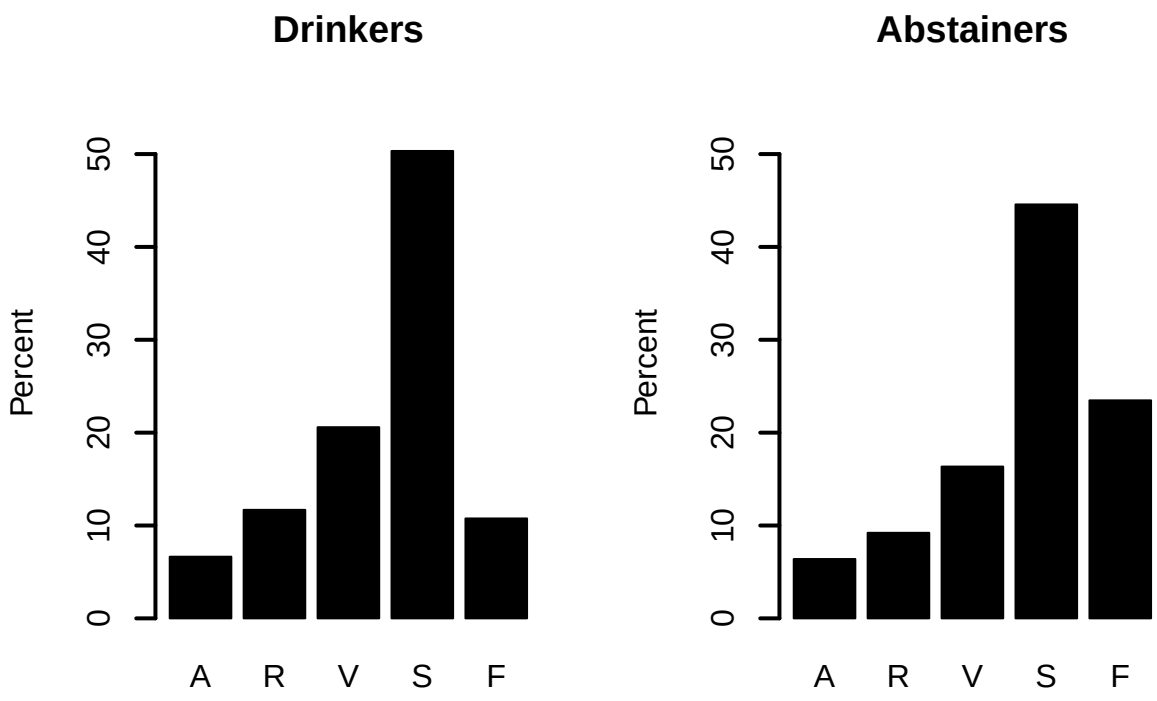


Figure 2.3

```
band <- c(7.0, 12.0, 8.0, 13.0, 7.2, 11.5, 7.4, 11.6, 8.2, 12.2)
labels1 <- paste("Band", 1:10)
par(mfrow = c(1,1))
pie(band, density = -10, labels = labels1, col = 0, lwd = 2)
```

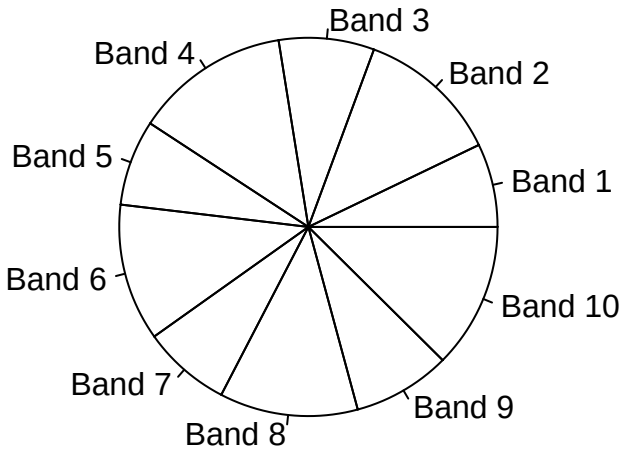


Figure 2.4

```
require("lattice")
```

```
## Loading required package: lattice
```

```
dotplot(1:10 ~ band, ylab = "Band", xlab = "Percentage", lwd = 3)
```

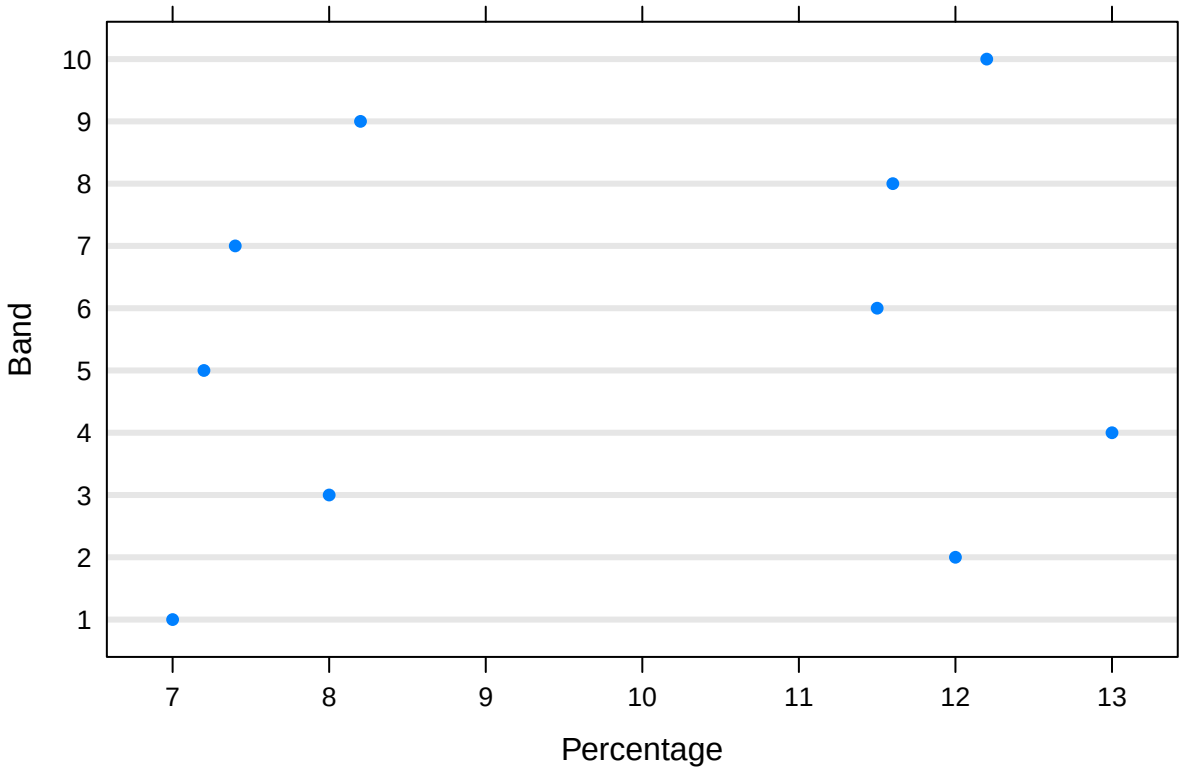


Figure 2.5

```
require("lattice")
dimnames(da) <- list(labels, c("Drinkers", "Abstainers"))
dotplot(da, groups = FALSE, xlab = "Percentage", ylab = "Crime")
```

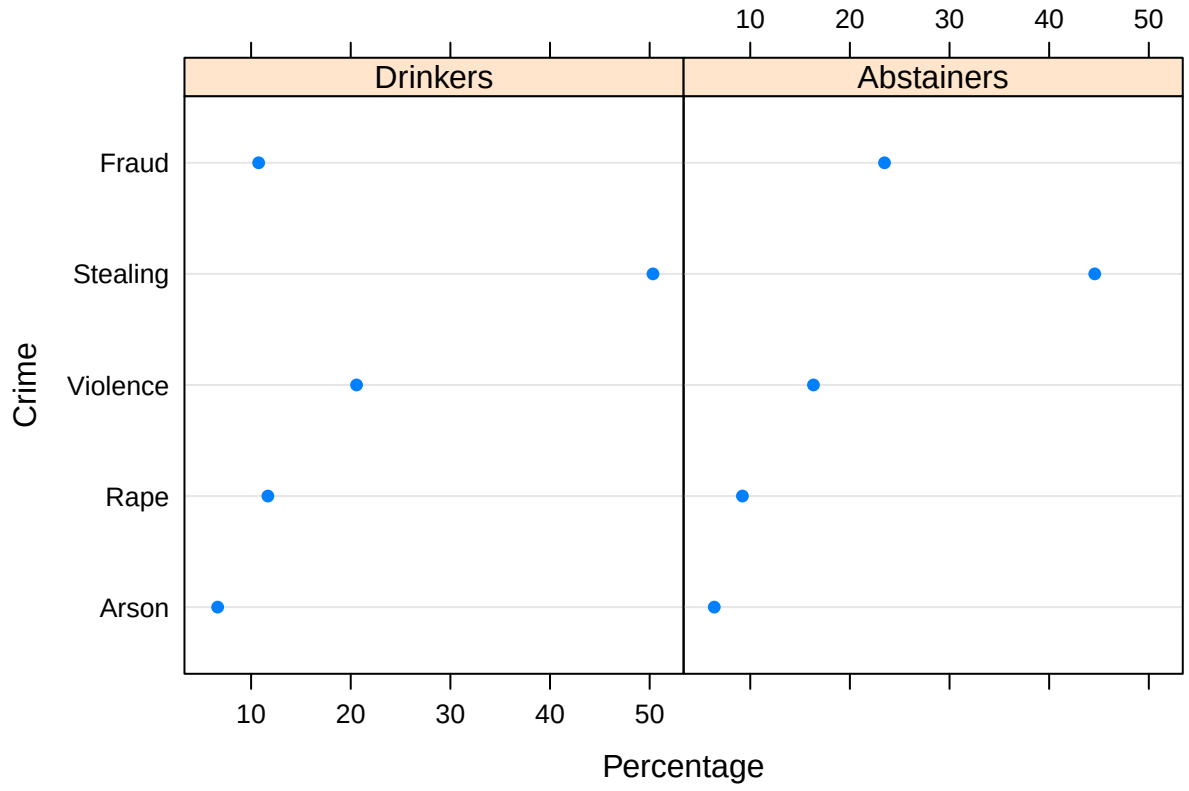


Figure 2.8

```
life <- c(76, 68, 77, 69, 76, 67, 77, 69, 75, 68, 77, 71, 75, 62, 78, 72, 76, 69, 77, 69)
life <- matrix(life, ncol = 2, byrow = T)
countries <- c("Austria", "Canada", "Finland", "France", "Germany",
               "Japan", "USSR", "Sweden", "UK", "USA")
dimnames(life) <- list(countries, c("Women", "Men"))
barplot(t(life), beside = T, space = c(0,3), horiz = F, names.arg = countries,
        legend.text = T, cex.names = 0.5, ylim = c(0,105),
        xlab = "Life expectancy at birth (years)")
```

Figure 2.10

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require("lattice")
# B&W version for the book:
#trellis.device(color = FALSE)
barchart(Class ~ Freq | Sex + Age, data = as.data.frame(Titanic),
         groups = Survived, stack = TRUE, layout = c(4, 1),
         auto.key = list(title = "Survived", columns = 2))
```

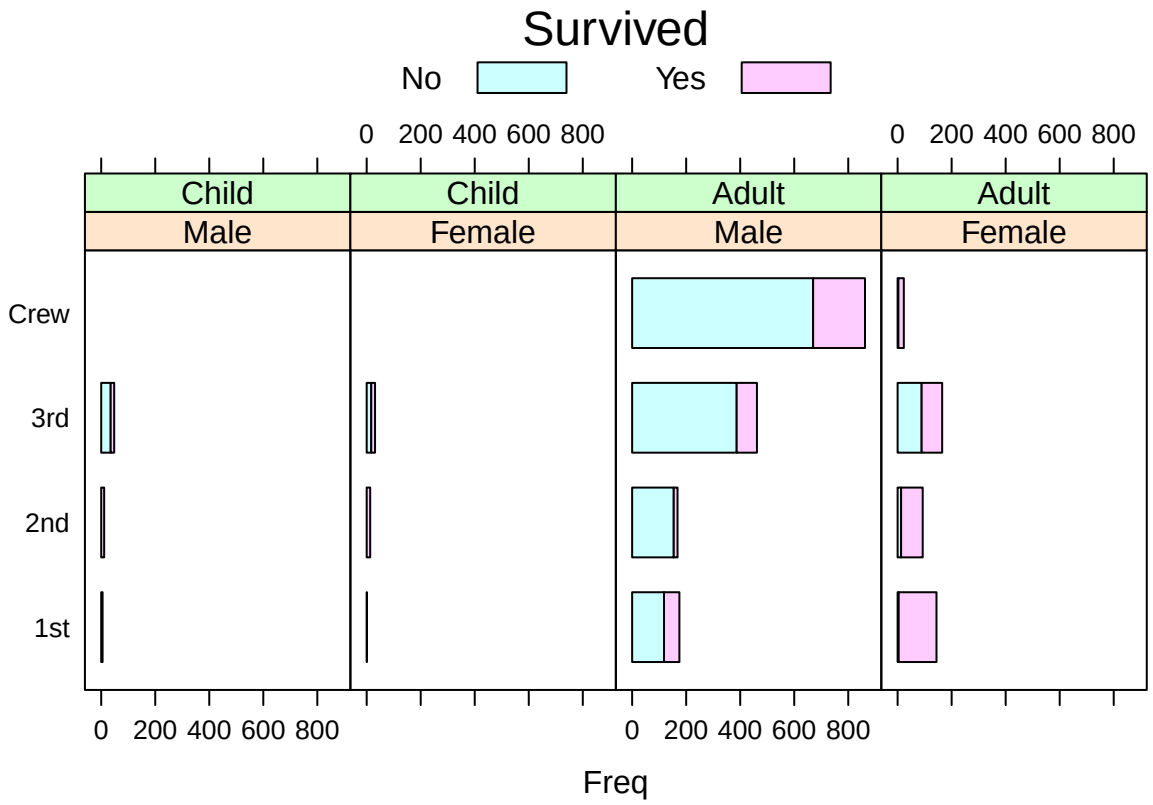


Figure 2.11

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require("lattice")
# B&W version for the book:
#trellis.device(color = FALSE)
barchart(Class ~ Freq | Sex + Age, data = as.data.frame(Titanic),
         groups = Survived, stack = TRUE, layout = c(4, 1),
         auto.key = list(title = "Survived", columns = 2),
         scales = list(x = "free"))
```

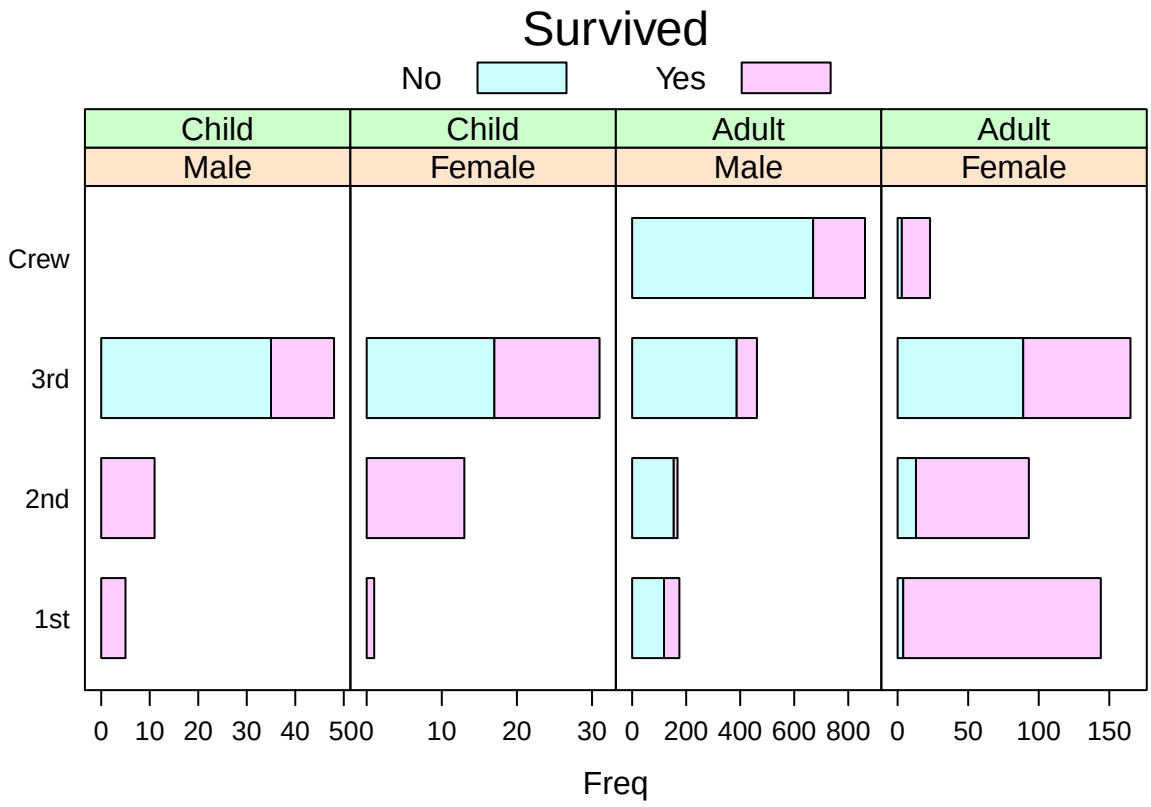


Table 2.2: Number of "and Then..." Statements Made by 50 Children

```
kids <- c(18, 15, 22, 19, 18, 17, 18, 20, 17, 12, 16, 16, 17, 21, 23, 18, 20,
          21, 20, 20, 15, 18, 17, 19, 20, 23, 22, 10, 17, 19, 19, 21, 20, 18,
          18, 24, 11, 19, 31, 16, 17, 15, 19, 20, 18, 18, 40, 18, 19, 16)
kids
```

```
##  [1] 18 15 22 19 18 17 18 20 17 12 16 16 17 21 23 18 20 21 20 20 15 18 17
## [24] 19 20 23 22 10 17 19 19 21 20 18 18 24 11 19 31 16 17 15 19 20 18 18
## [47] 40 18 19 16
```

Table 2.3: Number of "and Then..." Statements Made by 50 Adults

```
adults <- c(10, 12, 5, 8, 13, 10, 12, 8, 7, 11, 11, 10, 9, 9, 11, 15, 12,
            17, 14, 10, 9, 8, 15, 16, 10, 14, 7, 16, 9, 1, 4, 11, 12, 7,
            9, 10, 3, 11, 14, 8, 12, 5, 10, 9, 7, 11, 14, 10, 15, 9)
adults
```

```
##  [1] 10 12  5  8 13 10 12  8  7 11 11 10  9  9 11 15 12 17 14 10  9  8 15
## [24] 16 10 14  7 16  9  1  4 11 12  7  9 10  3 11 14  8 12  5 10  9  7 11
## [47] 14 10 15  9
```

Figure 2.12

```
hist(kids, ylab = "Frequency Count",
     xlab = "Number of 'and then...' statements", main = "")
```

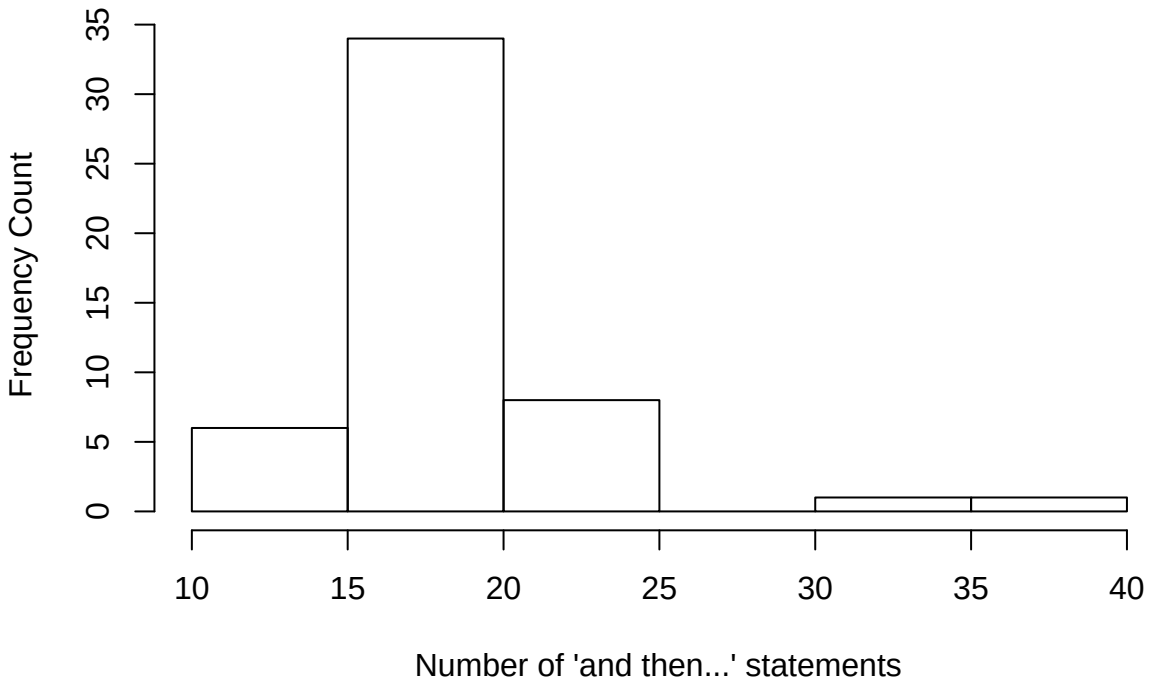


Figure 2.14

```
boxplot(kids, ylab = "Number of 'and then...' statements")
```

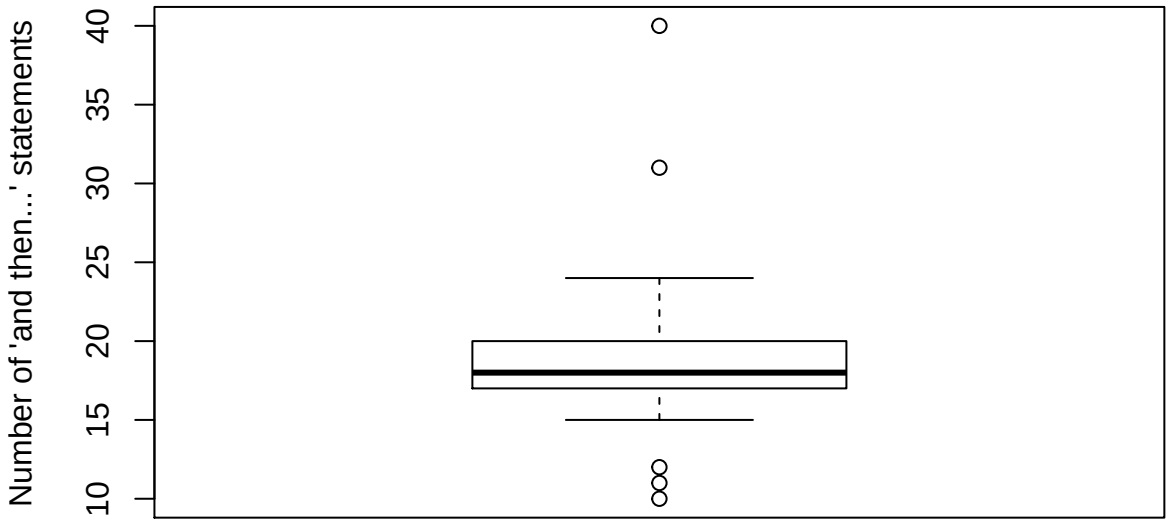


Figure 2.15

```
boxplot(kids, adults, names = c("Children", "Adults"),
        ylab = "Number of 'and then...' Statements")
```

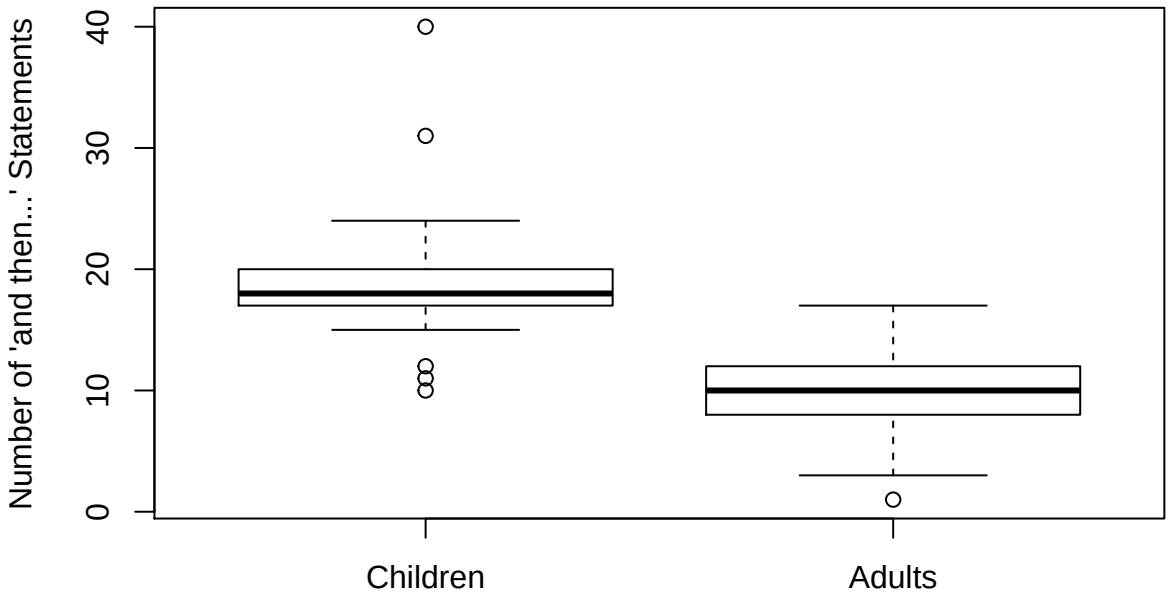


Figure 2.16

```
par(mfrow = c(1,2))
qqnorm(kids, main = "Children")
qqnorm(adults, main = "Adults")
```

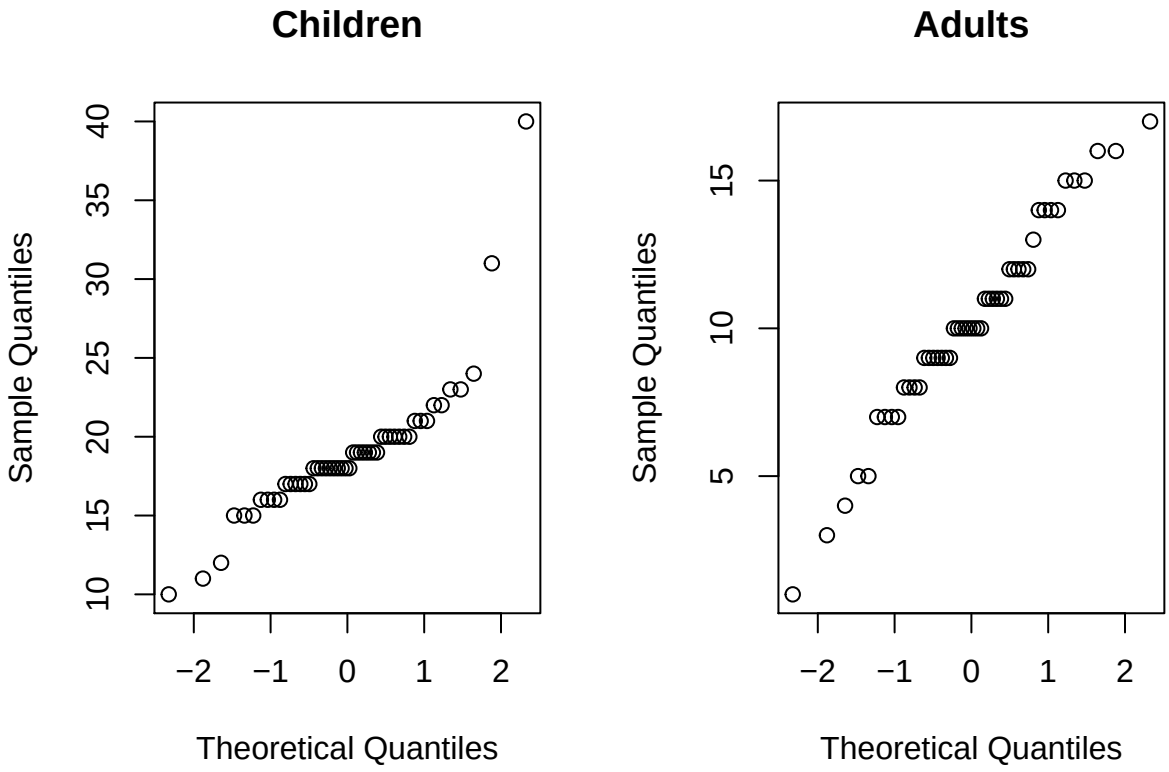


Table 2.4: Field Dependence Measure and Time to Complete a Task

```
#time to complete test; row and corner group data
timer <- c(317, 464, 525, 298, 491, 196, 268, 372, 370, 739, 430, 410)
eftr <- c(59, 33, 49, 69, 65, 26, 29, 62, 31, 139, 74, 31)
#
timec <- c(342, 222, 219, 513, 295, 285, 408, 543, 298, 494, 317, 407)
eftc <- c(43, 23, 9, 128, 44, 49, 87, 43, 55, 58, 113, 7)
te2 <- cbind(timer, eftr, timec, eftc)
te2
```

```
##       timer eftr timec eftc
##  [1,]   317   59   342   43
##  [2,]   464   33   222   23
##  [3,]   525   49   219    9
##  [4,]   298   69   513  128
##  [5,]   491   65   295   44
##  [6,]   196   26   285   49
##  [7,]   268   29   408   87
##  [8,]   372   62   543   43
##  [9,]   370   31   298   55
## [10,]   739  139   494   58
## [11,]   430   74   317  113
## [12,]   410   31   407    7
```

Figure 2.17

```
par(mfrow = c(2,1), mar = c(2, 4, 2, 0.1))
plot(eftr, timer, xlim = range(c(eftr, eftc)), ylim = range(c(timer, timec)),
     xlab = "EFT", ylab = "Completion time (secs)")
title(main = "Row group")
#
plot(eftc, timec, xlim = range(c(eftr, eftc)), ylim = range(c(timer, timec)),
     xlab = "EFT", ylab = "Completion time (secs)")
title(main = "Corner group")
```

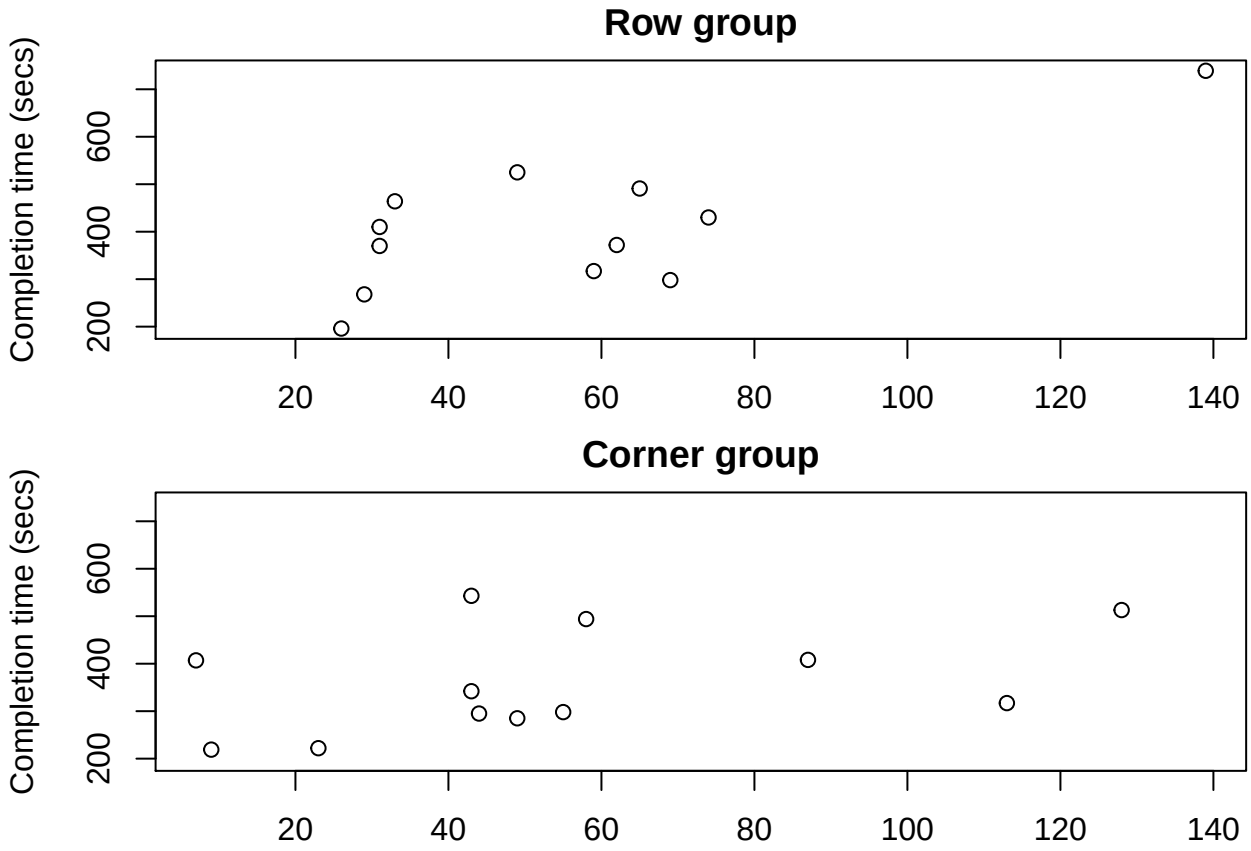


Figure 2.18

```
par(mfrow = c(2,1), mar = c(2, 4, 2, 0.1))
plot(eftr, timer, xlim = range(c(eftr, eftc)), ylim = range(c(timer, timec)),
     xlab = "EFT", ylab = "Completion time (secs)")
title(main = "Row group")
abline(lm(timer ~ eftr))
#
plot(eftc, timec, xlim = range(c(eftr, eftc)), ylim = range(c(timer, timec)),
     xlab = "EFT", ylab = "Completion time (secs)")
title(main = "Corner group")
abline(lm(timec ~ eftc))
```

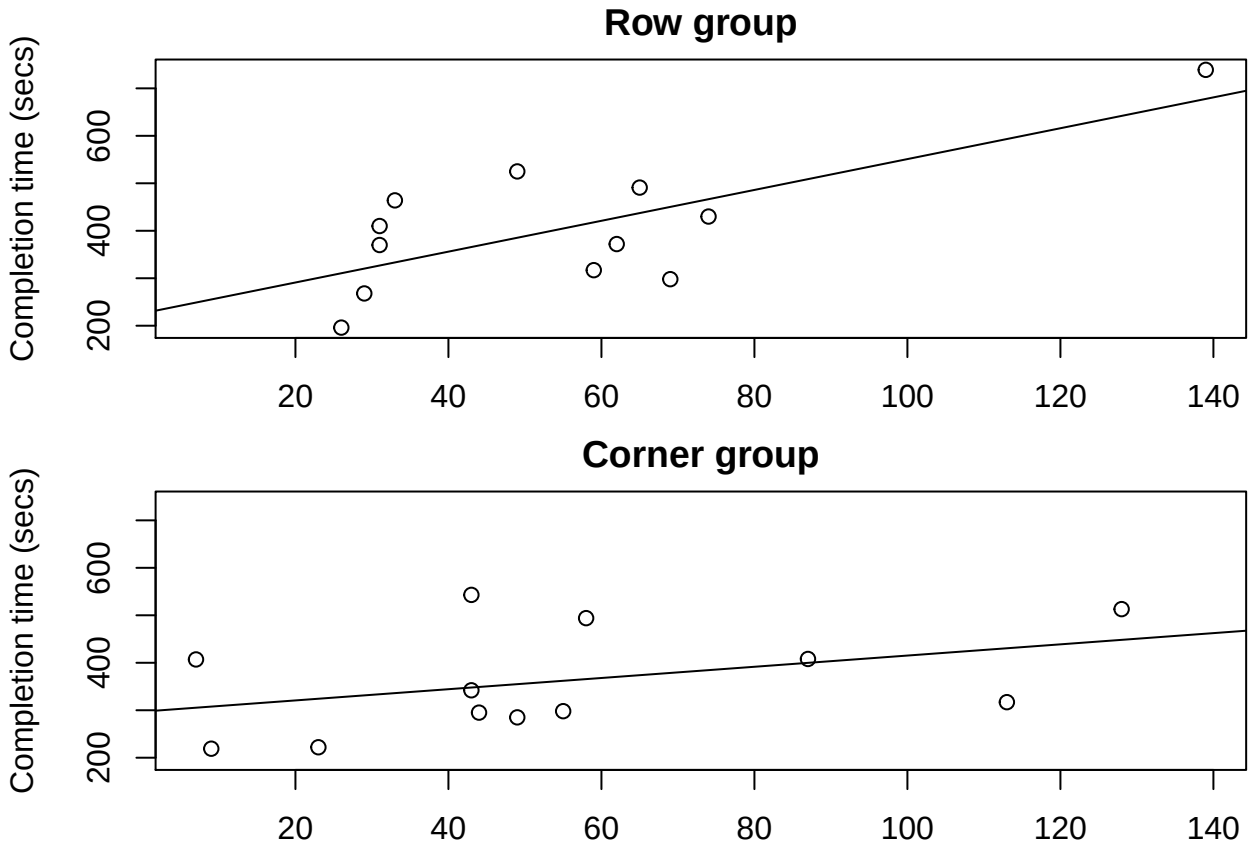


Figure 2.19

```
par(mfrow = c(1,1), mar = c(4, 4, 1, 0.1))
plot(eftr, timer, xlim = range(c(eftr, eftc)), ylim = range(c(timer, timec)),
     xlab = "EFT", ylab = "Completion time (secs)", pch = "R")
abline(lm(timer ~ eftr), lty = 1)
points(eftc, timec, pch = "C")
abline(lm(timec ~ eftc), lty = 2)
legend("bottomright", legend = c("Row group", "Column group"), pch = c("R", "C"))
legend("topleft", legend = c("Row group fit", "Column group fit"), lty = 1:2)
```

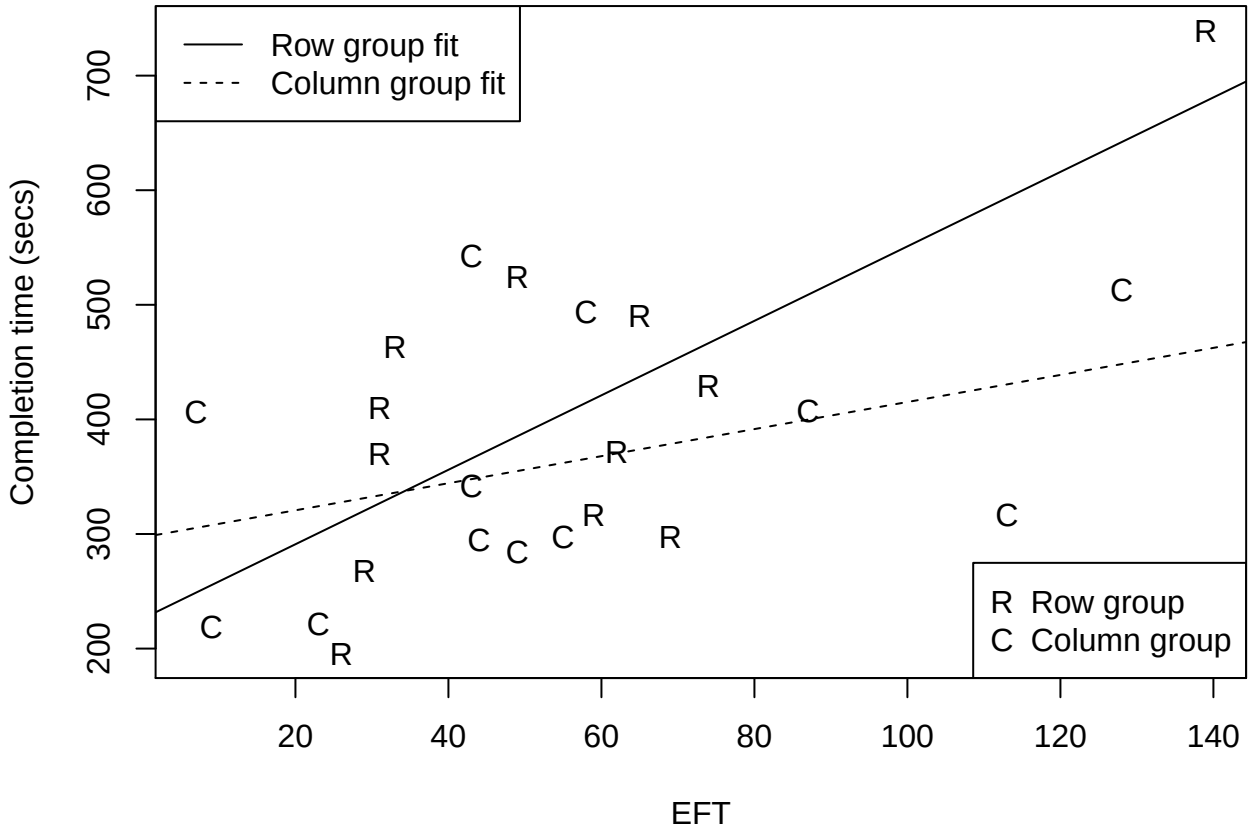


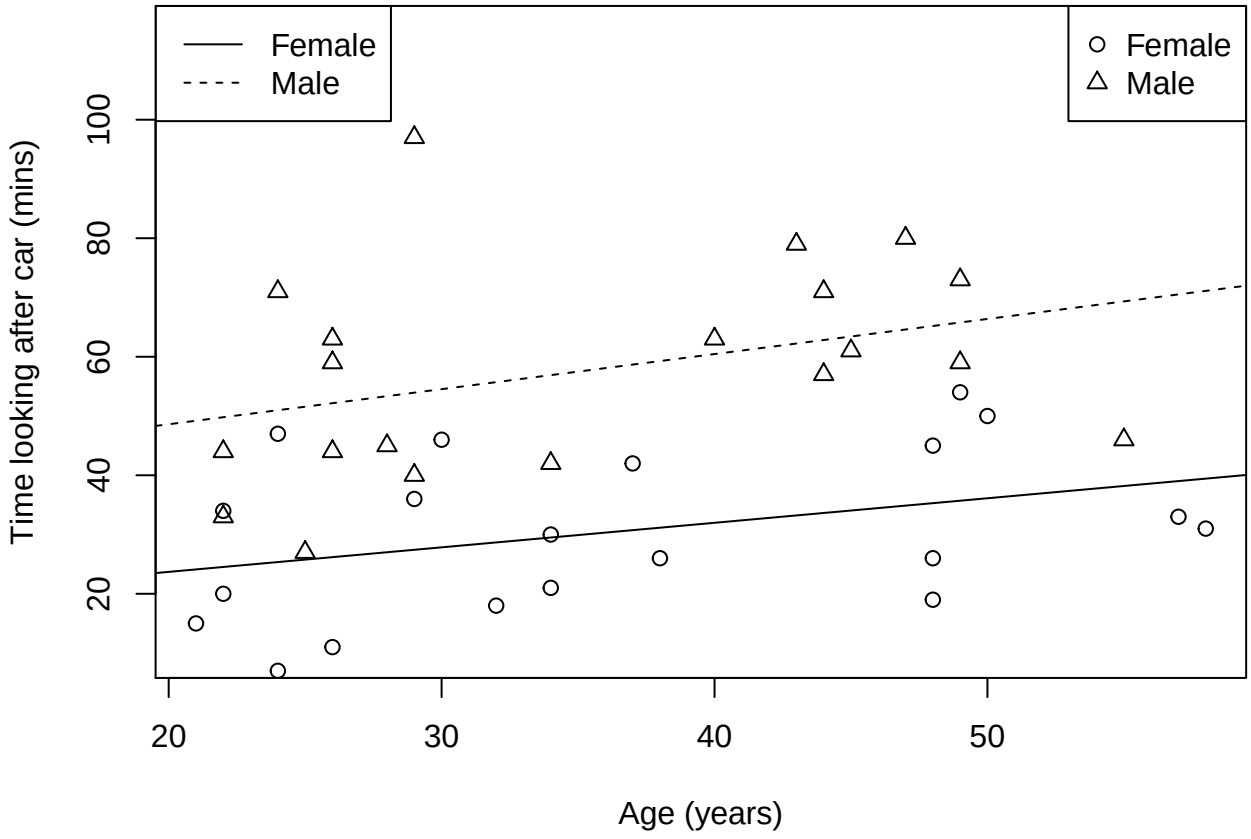
Table 2.5: Time Spent Looking After Car

```
sex <- c(1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0,
         0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1)
age <- c(55, 43, 57, 26, 22, 32, 26, 29, 40, 30, 34, 44, 49, 22, 34, 47,
         48, 48, 22, 24, 50, 49, 49, 48, 29, 58, 24, 21, 29, 45, 28, 37,
         44, 22, 38, 24, 34, 26, 26, 25)
extro <- c(40, 45, 52, 62, 31, 28, 2, 83, 55, 32, 47, 45, 60, 13, 7, 85,
           38, 61, 26, 3, 29, 60, 47, 18, 16, 36, 24, 12, 32, 46, 26, 40,
           46, 44, 3, 25, 43, 41, 42, 36)
time <- c(46, 79, 33, 63, 20, 18, 11, 97, 63, 46, 21, 71, 59, 44, 30, 80,
          45, 26, 33, 7, 50, 54, 73, 19, 36, 31, 71, 15, 40, 61, 45, 42,
          57, 34, 26, 47, 42, 44, 59, 27)
#
sex <- factor(sex, levels = c(0, 1), labels = c("Female", "Male"))
carclean <- as.data.frame(cbind(sex, age, extro, time))
head(carclean)
```

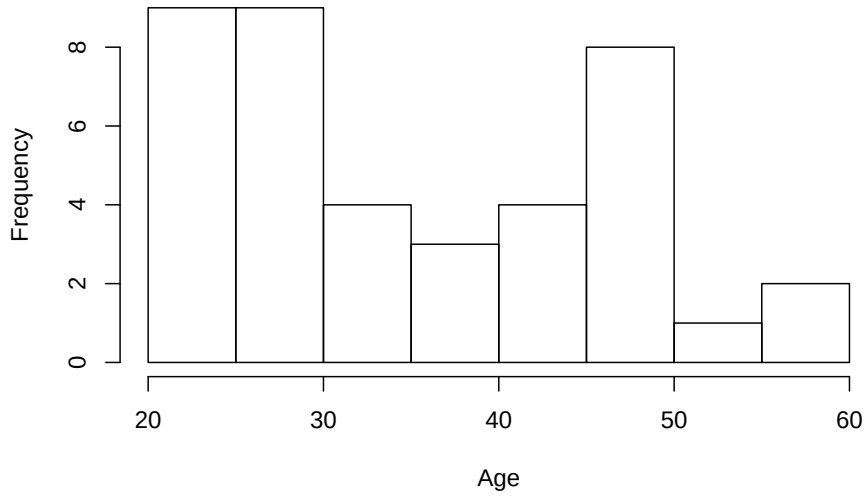
```
##   sex age extro time
## 1   2  55    40   46
## 2   2  43    45   79
## 3   1  57    52   33
## 4   2  26    62   63
## 5   1  22    31   20
## 6   1  32    28   18
```

Figure 2.20

```
# nf <- layout(matrix(c(2,0,1,3), 2, 2, byrow = TRUE), c(2,1), c(1,2), TRUE)
par(mfrow = c(1,1), mar = c(4, 4, 1, 0.1))
psymb <- as.numeric(sex)
plot(time ~ age, pch = psymb, ylim = c(10, 115),
     xlab = "Age (years)", ylab = "Time looking after car (mins)")
abline(lm(time[sex == "Female"] ~ age[sex == "Female"]), lty = 1)
abline(lm(time[sex == "Male"] ~ age[sex == "Male"]), lty = 2)
legend("topright", legend = levels(sex), pch = c(1,2))
legend("topleft", legend = levels(sex), lty = 1:2)
```



```r
hist(age, ylab = "Frequency", xlab = "Age", main = "")
```




```r
boxplot(time, ylab = "Age")
```

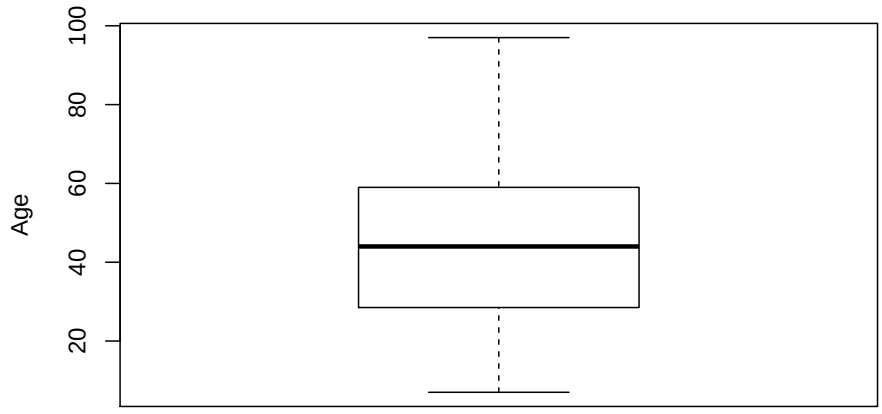
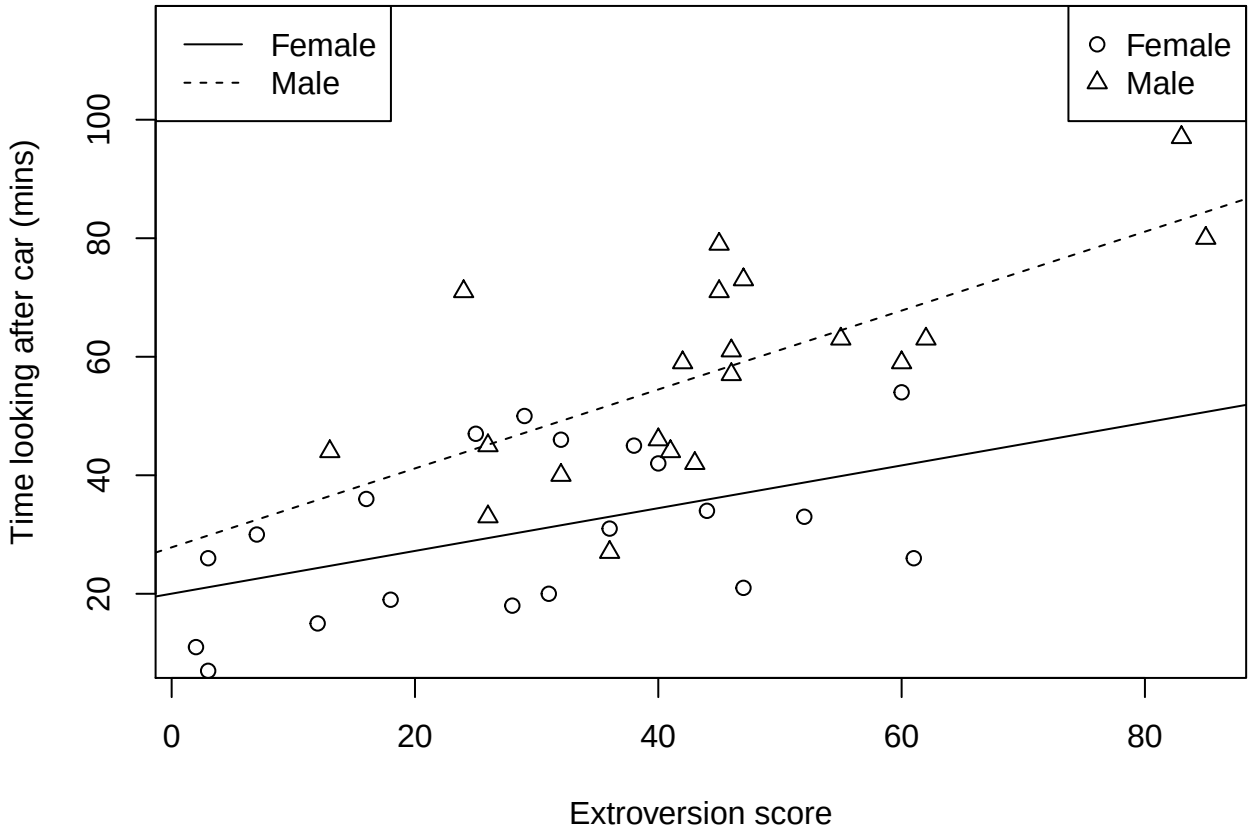
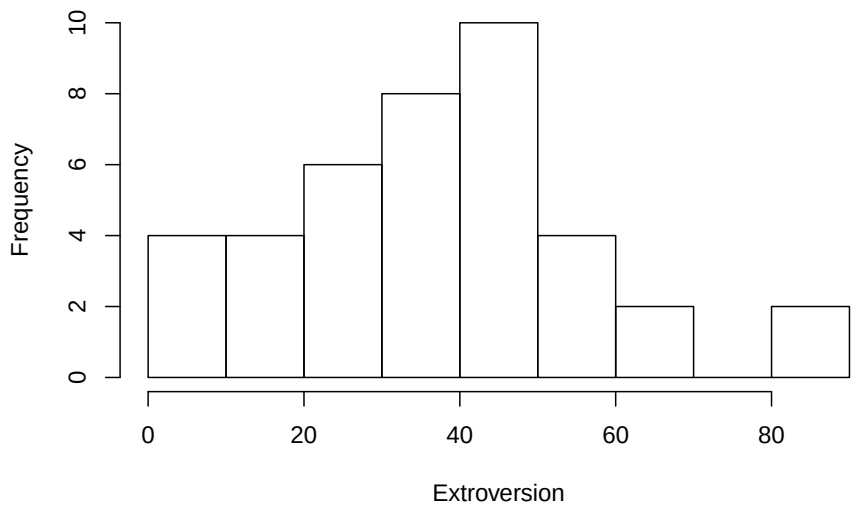


Figure 2.21

```
# nf <- layout(matrix(c(2,0,1,3), 2, 2, byrow = TRUE), c(2,1), c(1,2), TRUE)
par(mfrow = c(1,1), mar = c(4, 4, 1, 0.1))
psymb <- as.numeric(sex)
plot(time ~ extro, pch = psymb, ylim = c(10, 115),
     xlab="Extroversion score", ylab = "Time looking after car (mins)")
abline(lm(time[sex == "Female"] ~ extro[sex == "Female"]), lty = 1)
abline(lm(time[sex == "Male"] ~ extro[sex == "Male"]), lty = 2)
legend("topright", legend = levels(sex), pch = c(1,2))
legend("topleft", legend = levels(sex), lty = 1:2)
```



```r
hist(extro, ylab = "Frequency", xlab = "Extroversion", main = "")
```




```r
boxplot(time, ylab = "Extroversion")
```

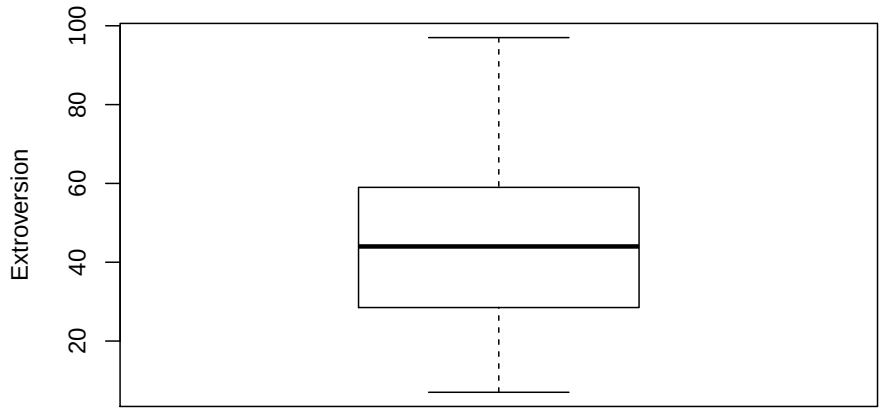


Figure 2.22

```
par(mfrow = c(1,1), mar = c(4, 4, 1, 0.1))
psymb <- as.numeric(sex)
plot(time ~ age, pch = psymb, xlim = c(20, 65), ylim = c(10, 115),
     xlab = "Age (years)", ylab = "Time looking after car (mins)")
legend("topright", legend = levels(sex), pch = c(1,2))
symbols(age, time, circles = extro, inches = 0.4, add = TRUE)
```

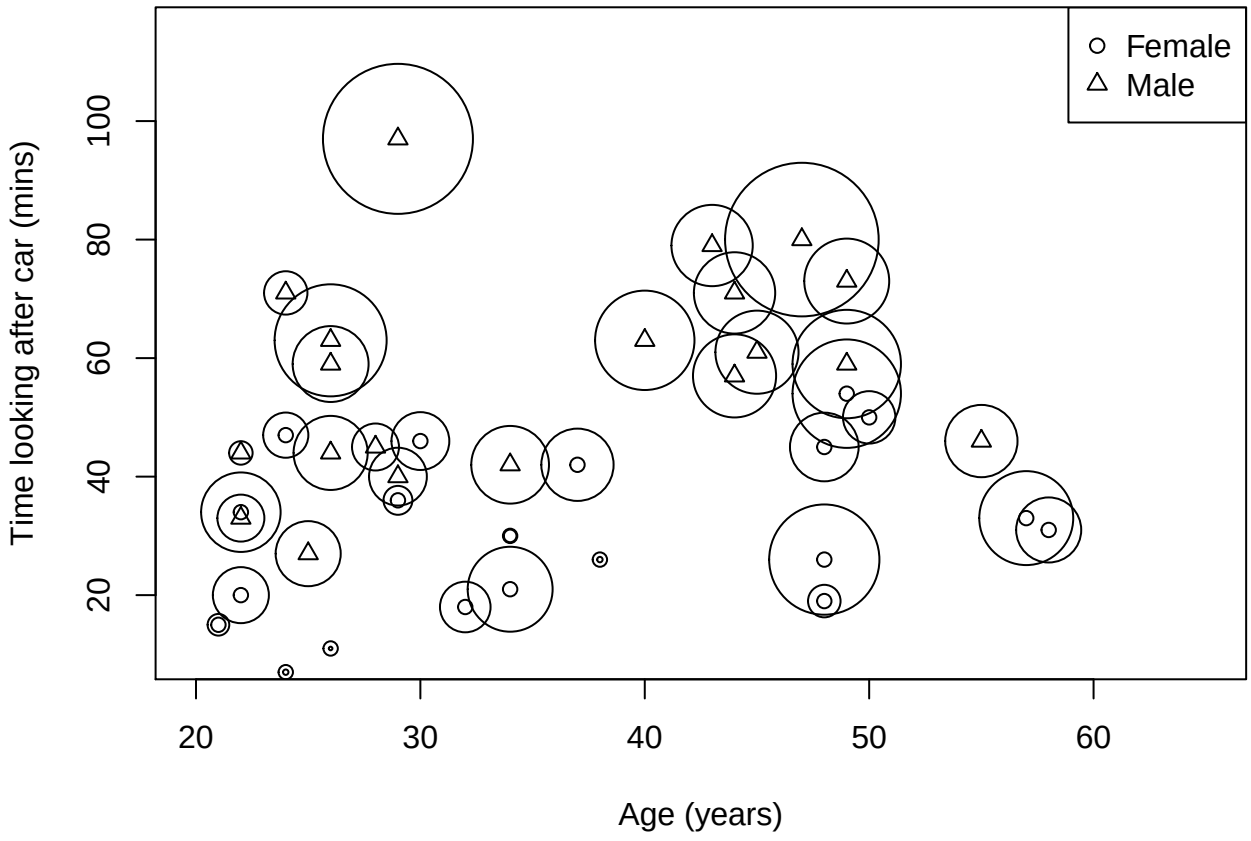


Figure 2.24

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

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

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

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

```
bvbox(cbind(age, time), mtitle = "", xlab = "Age (years)",
      ylab = "Time spent looking after car (mins)")
```

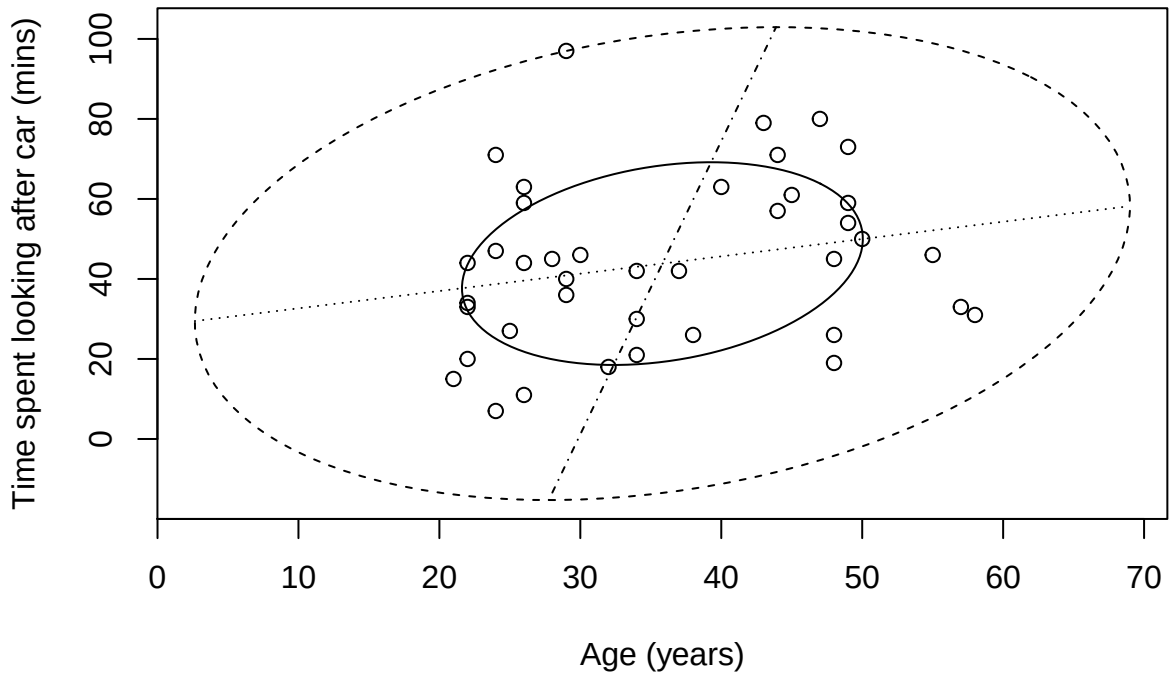


Figure 2.25

```
#install.packages("MVA")
require("MVA")
bvbox(cbind(extro, time), mtitle = "", xlab = "Extroversion score",
      ylab = "Time spent looking after car (mins)")
```

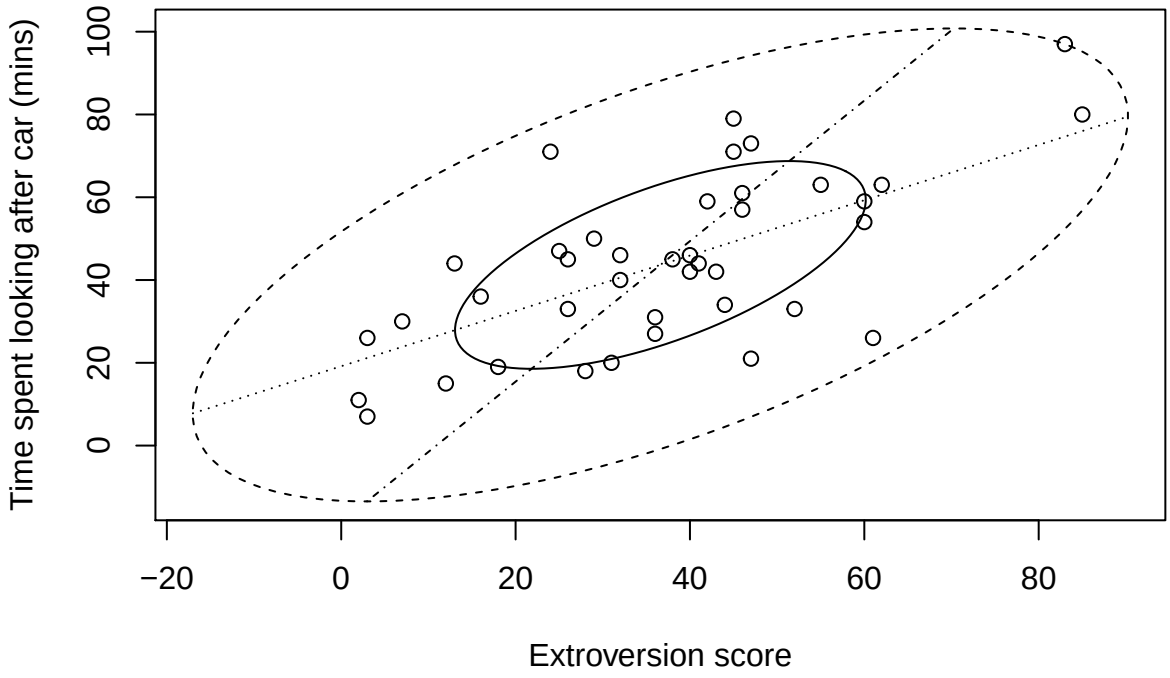


Table 2.6: Measure of Resistance Made on Five Different Types of Electrode

```
resis <- c(500, 400, 98, 200, 250, 660, 600, 600, 75, 310, 250, 370,
           220, 250, 220, 72, 140, 240, 33, 54, 135, 300, 450, 430, 70,
           27, 84, 135, 190, 180, 100, 50, 82, 73, 78, 105, 180, 32,
           58, 32, 90, 180, 220, 34, 64, 200, 290, 320, 280, 135, 15,
           45, 75, 88, 80, 160, 200, 300, 300, 330, 250, 400, 50, 50,
           92, 170, 310, 230, 20, 150, 66, 1000, 1050, 280, 220, 107,
           48, 26, 45, 51)
resis <- matrix(resis, nrow = 16, byrow = T)
dimnames(resis) <- list(NULL, c("E1", "E2", "E3", "E4", "E5"))
resis
```

```
##        E1   E2   E3  E4  E5
##  [1,] 500  400   98 200 250
##  [2,] 660  600  600  75 310
##  [3,] 250  370  220 250 220
##  [4,]  72  140  240  33  54
##  [5,] 135  300  450 430  70
##  [6,]  27   84  135 190 180
##  [7,] 100   50   82  73  78
##  [8,] 105  180   32  58  32
##  [9,]  90  180  220  34  64
## [10,] 200  290  320 280 135
## [11,]  15   45   75  88  80
## [12,] 160  200  300 300 330
## [13,] 250  400   50  50  92
## [14,] 170  310  230  20 150
## [15,]  66 1000 1050 280 220
## [16,] 107   48   26  45  51
```

Figure 2.26

```
pairs(resis)
```

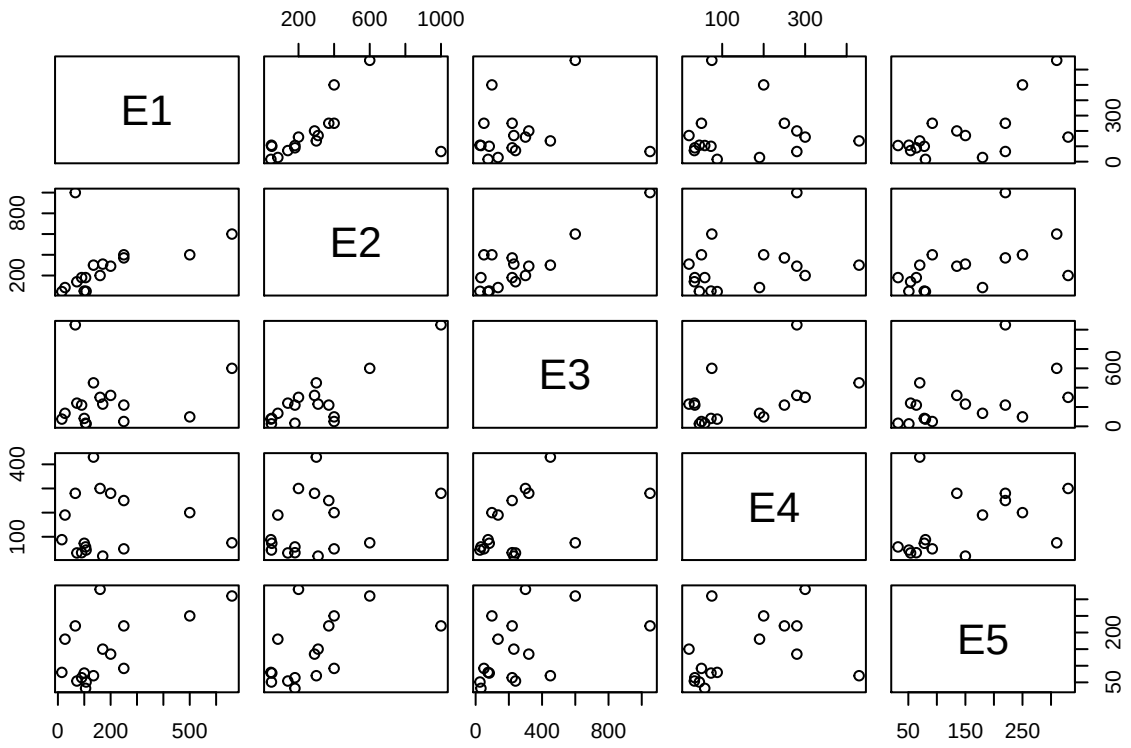
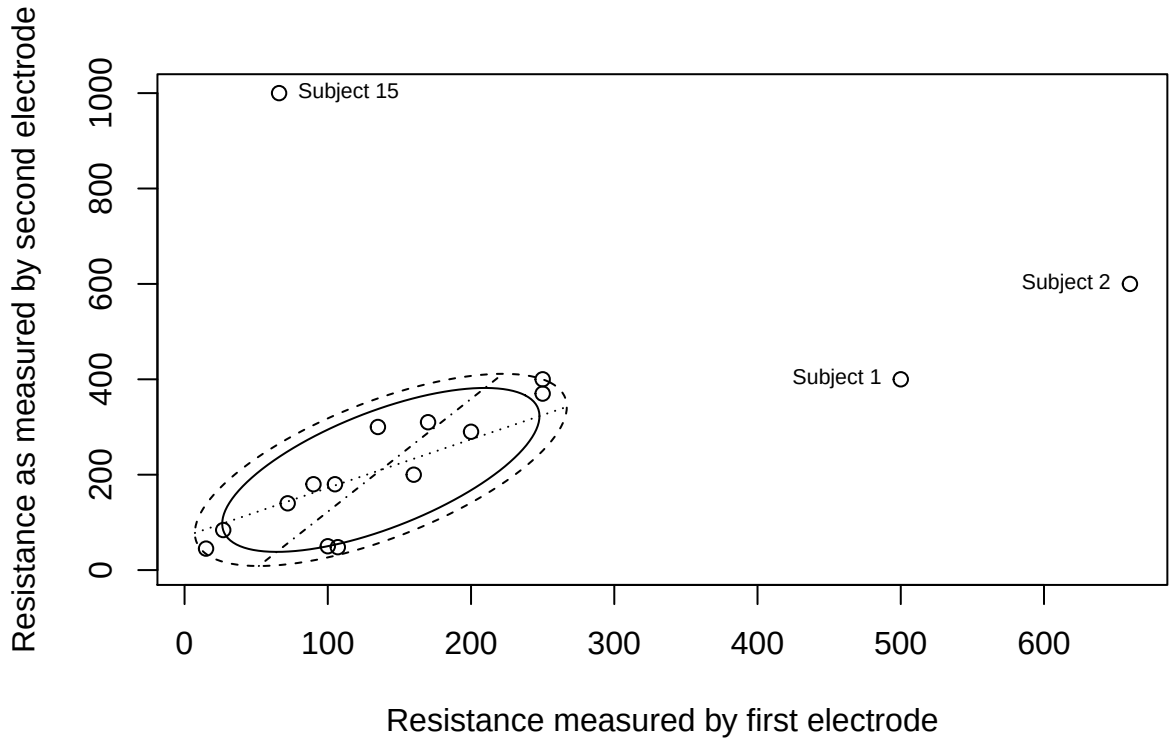


Figure 2.27

```
#install.packages("MVA")
require("MVA")
bvbox(resis[, c(1, 2)], mtitle = "",
      xlab = "Resistance measured by first electrode",
      ylab = "Resistance as measured by second electrode")
text(resis[c(1, 2, 15), c(1, 2)], labels = c("Subject 1", "Subject 2", "Subject 15"),
     cex = 0.7, pos = c(2, 2, 4))
```




```
# Compare the correlations with/without the outliers:
#
cor(resis[, c(1, 2)])
```

```
##           E1        E2
## E1 1.0000000 0.4103945
## E2 0.4103945 1.0000000
```

```
#
cor(resis[-c(1, 2, 15), c(1, 2)])
```

```
##           E1        E2
## E1 1.0000000 0.8773492
## E2 0.8773492 1.0000000
```

Figure 2.28

```
coplot(time ~ extro | sex, panel = function(x, y, col, pch) {
                                       points(x, y)
                                       abline(lm(y ~ x))
                                   })
```

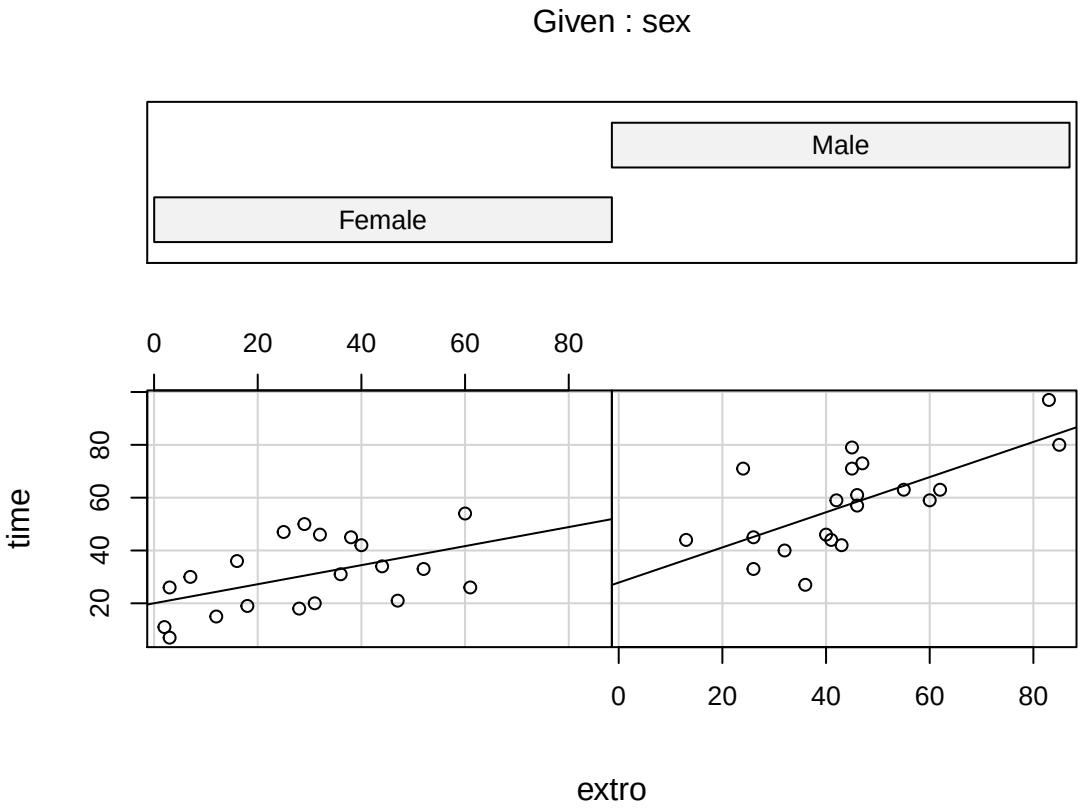


Figure 2.29

```
coplot(time ~ extro | age, panel = function(x, y, col, pch) {
                                      points(x, y)
                                      abline(lm(y ~ x))
                                  })
```

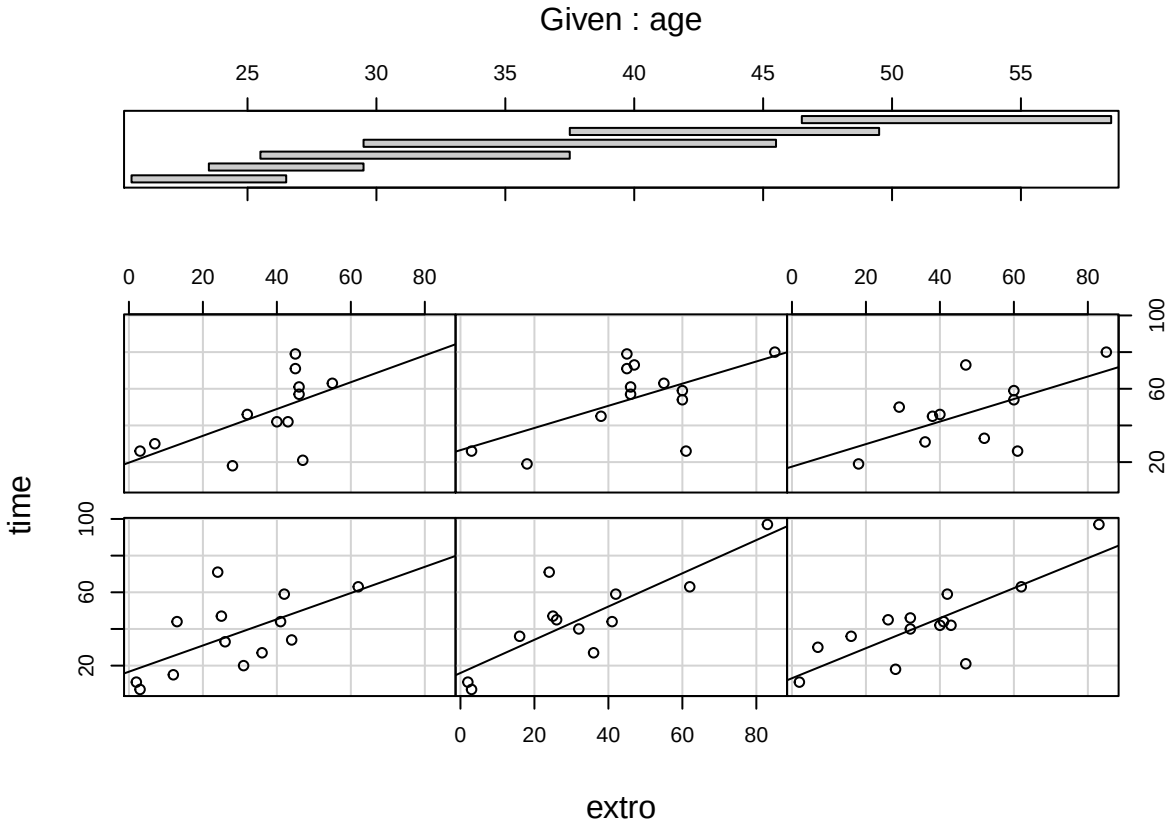


Figure 2.31

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require(lattice)
data(postdoc, package = "latticeExtra")
barchart(prop.table(postdoc, margin = 1), xlab = "Proportion",
         auto.key = list(adj = 1))
```

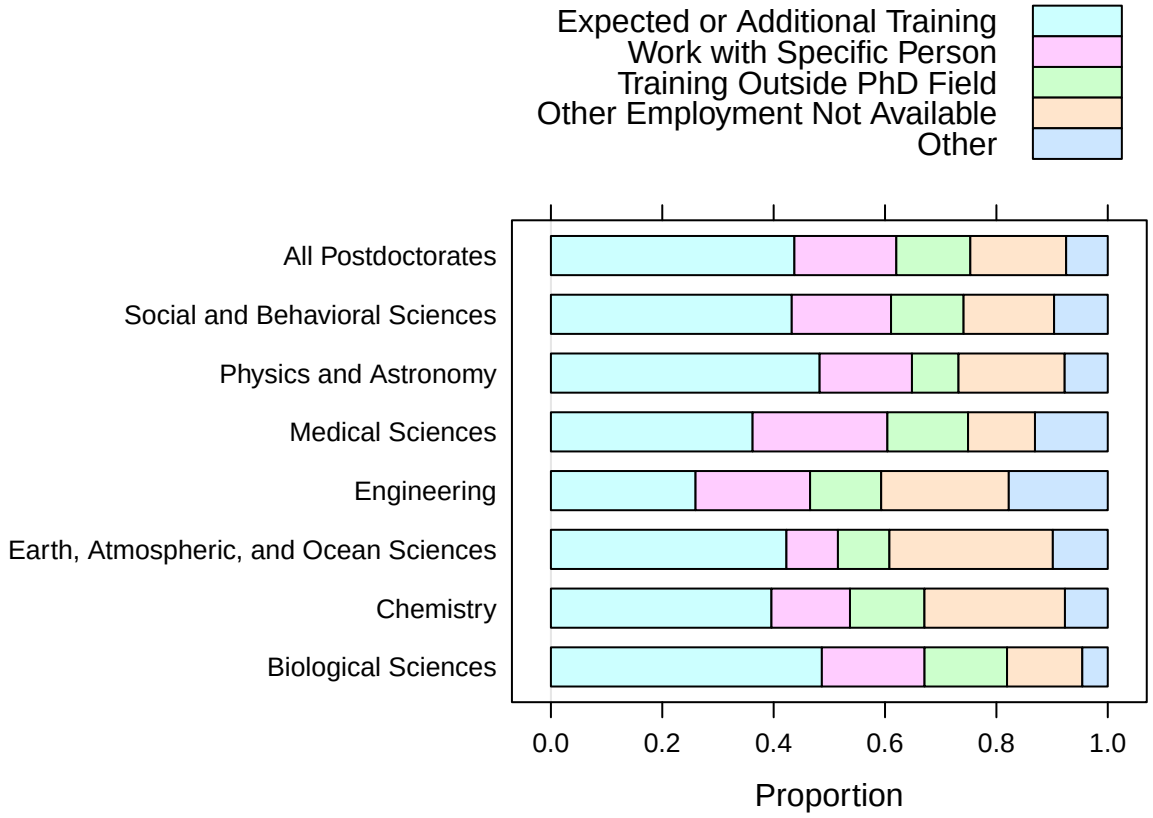


Figure 2.32

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require(lattice)
dotplot(prop.table(postdoc, margin = 1), groups = FALSE,
        xlab = "Proportion", par.strip.text = list(abbreviate = TRUE, minlength = 10))
```

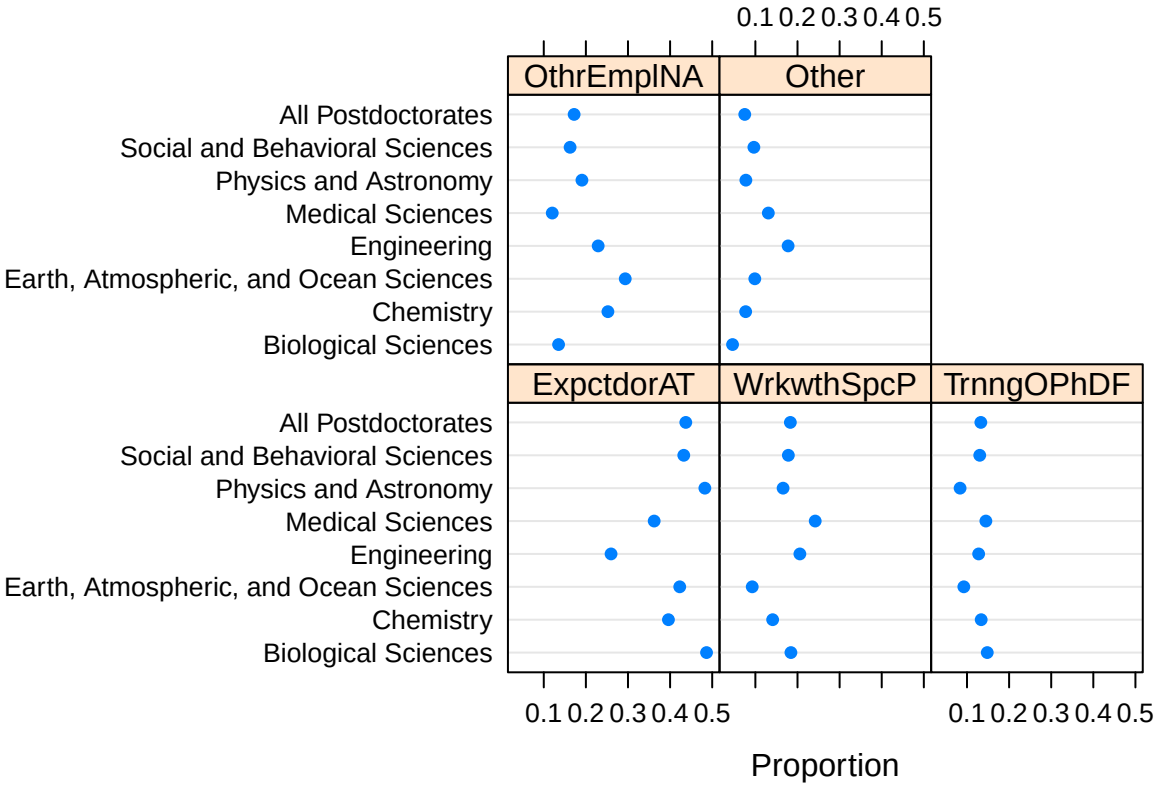


Figure 2.33

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require(lattice)
dotplot(prop.table(postdoc, margin = 1), groups = FALSE,
        index.cond = function(x, y) median(x),
        xlab = "Proportion", layout = c(1, 5), aspect = 0.15,
        scales = list(y = list(relation = "free", rot = 0)),
        prepanel = function(x, y) {
                       list(ylim = levels(reorder(y, x)))
                   },
        panel = function(x, y, ...) {
                     panel.dotplot(x, reorder(y, x), ...)
                 })
```

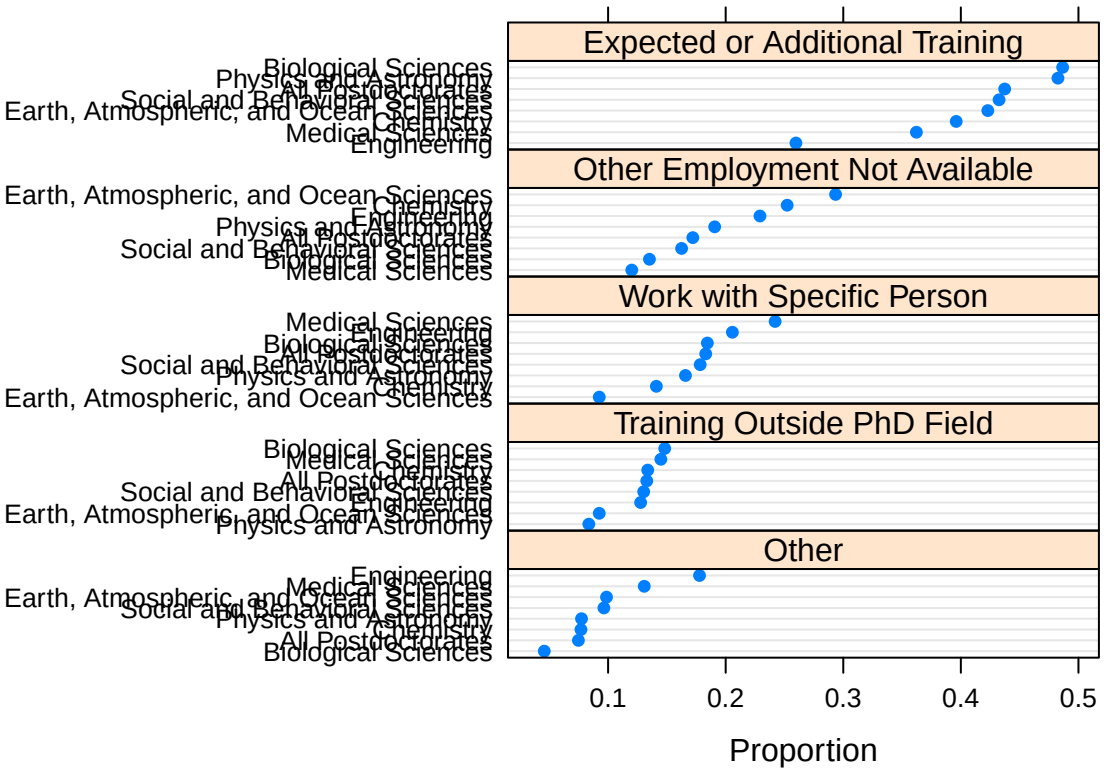


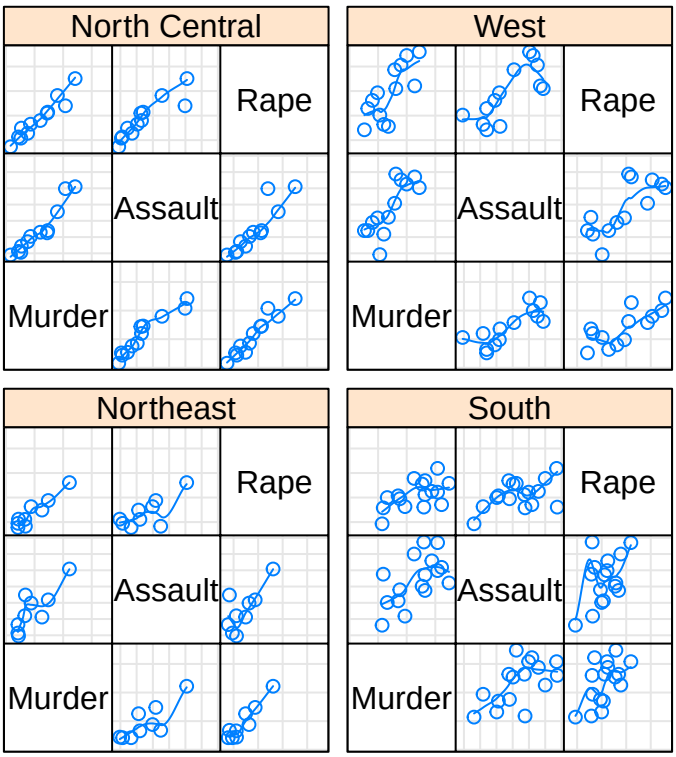
Figure 2.34

Example taken with permission from:

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*, Springer.

http://lmdvr.r-forge.r-project.org/

```
require(lattice)
splom( ~ USArrests[c(1, 2, 4)] | state.region, pscales = 0,
       type = c("g", "p", "smooth"))
```



Scatter Plot Matrix

Figure 2.35

```
RATES <- read.table("data/deathrates.txt", sep = ' ' , header = TRUE)
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
p1 <- ggplot(RATES, aes(x = year, y = rate))
p2 <- p1 + geom_line() + geom_point(shape=22, size=2, fill = "black")
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(),
                              panel.grid.minor = element_blank())
p4a <- p3 + scale_x_continuous(name = "(a)")
p5a <- p4a + scale_y_continuous(name = "Death Rate by Million")

p4b <- p3 + scale_x_continuous(name = "(b)")
p5b <- p4b + scale_y_continuous(name = "Death Rate by Million",
                                breaks = seq(0, 500, 100), limits = c(0, 500))
p5a; p5b
```

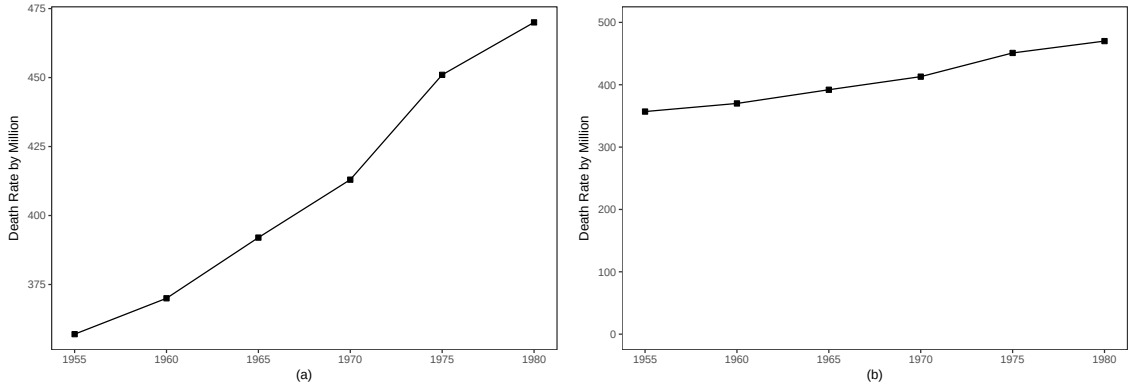


Figure 2.38

Source of the data (with the jittered NumIncJ variable manually added):

Dalal, S. R., Fowlkes, E. B. and Hoadley, B. (1988). Risk analysis of the space shuttle: Pre-Challenger prediction of failure. *Journal of the American Statistical Association*, 84, 945–957.

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
orings <- read.table("data/orings.txt", header = TRUE)
orings
```

```
##    Flight     Date TempF NumInc NumIncJ
## 1       1  4/12/81    66      0    0.00
## 2       2 11/12/81    70      1    0.95
## 3       3  3/22/82    69      0    0.00
## 4       5 11/11/82    68      0    0.00
## 5       6  4/04/83    67      0   -0.10
## 6       7  6/18/83    72      0    0.00
## 7       8  8/30/83    73      0    0.00
## 8       9 11/28/83    70      0   -0.10
## 9    41-8  2/03/84    57      1    1.00
## 10   41-C  4/06/84    63      1    1.00
## 11   41-D  8/30/84    70      1    1.05
## 12   41-G 10/05/84    78      0    0.00
## 13   51-A 11/08/84    67      0    0.00
## 14   51-C  1/24/85    53      3    3.00
## 15   51-D  4/12/85    67      0    0.10
## 16   51-8  4/29/85    75      0    0.00
## 17   51-G  6/17/85    70      0    0.10
## 18   51-F  7/29/85    81      0    0.00
## 19   51-1  8/27/85    76      0   -0.05
## 20   51-J 10/03/85    79      0    0.00
## 21   61-A 10/30/85    75      2    2.00
## 22   61-8 11/26/85    76      0    0.05
## 23   61-C  1/12/86    58      1    1.00
```

```
orings123 <- orings %>% filter(NumInc > 0.5)
```

```
require(ggplot2)

xlab = expression("Calculated Joint Temperature " ( degree*F))
ylab = "Number of Incidences"

p1 <- ggplot(orings123, aes(x = TempF, y = NumIncJ))
p2 <- p1 + geom_point(shape = 1, size = 3) + labs(x = xlab,
                y = ylab) + ylim(-0.1,3) + xlim(53,81)
p3a <- p2 + theme_bw() + theme(panel.grid.major = element_blank(),
                               panel.grid.minor = element_blank())
p3a
```

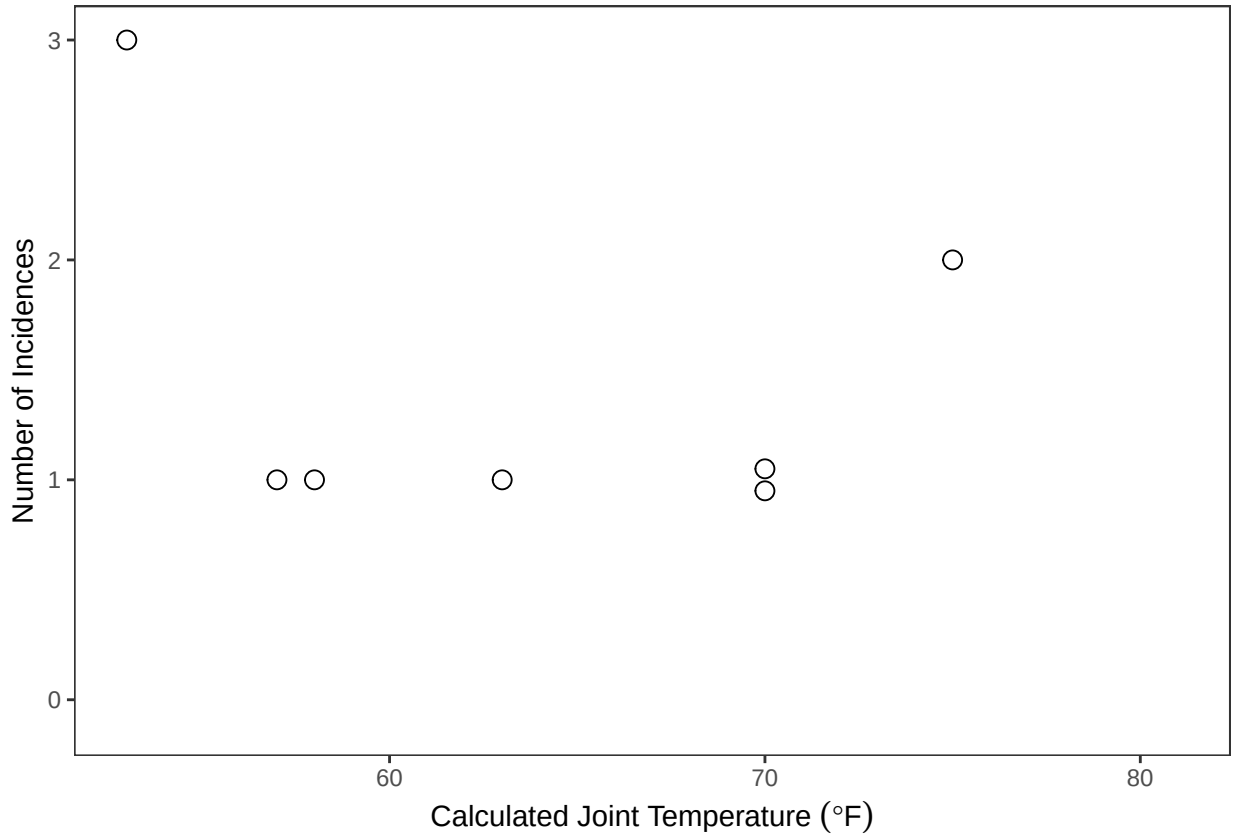


Figure 2.39

```
require(ggplot2)
# using the complete data (compare with Fig. 2.38)
p1 <- ggplot(orings, aes(x = TempF, y = NumIncJ))
p2 <- p1 + geom_point(shape = 1, size = 3) + labs(x = xlab,
                y = ylab) + ylim(-0.1,3) + xlim(53,81)
p3b <- p2 + theme_bw() + theme(panel.grid.major = element_blank(),
                               panel.grid.minor = element_blank())
p3b
```