Question 2: Relationships between the variables

a) **Compute the covariance and correlation matrices for the 7 variables. Is there any apparent structure in them? Save these matrices for future use.**

For computing both matrices asked, it is enough to call the two functions associated with covariance and correlation, which are:

```
cov_matrix <- cov(dataset[, 2:8])
cor_matrix <- cor(dataset[, 2.8])
```

being its output:

```
> cov_matrix
           X2         X3         X4          X5         X6          X7         X8
X2 0.15531572  0.3445608  0.8912960 0.027703564 0.08389119  0.23388281   4.334178
X3 0.34456080  0.8630883  2.1928363 0.066165898 0.20276331  0.55435017  10.384988
X4 0.89129602  2.1928363  6.7454576 0.181807932 0.50917683  1.42681579  28.903731
X5 0.02770356  0.0661659  0.1818079 0.007546925 0.02141457  0.06137932   1.219655
X6 0.08389119  0.2027633  0.5091768 0.021414570 0.07418270  0.21615514   3.539837
X7 0.23388281  0.5543502  1.4268158 0.061379315 0.21615514  0.66475793  10.706091
X8 4.33417757 10.3849876 28.9037314 1.219654647 3.53983732 10.70609113 270.270150
> cor_matrix
          X2        X3        X4        X5        X6        X7        X8
X2 1.0000000 0.9410886 0.8707802 0.8091758 0.7815510 0.7278784 0.6689597
X3 0.9410886 1.0000000 0.9088096 0.8198258 0.8013282 0.7318546 0.6799537
X4 0.8707802 0.9088096 1.0000000 0.8057904 0.7197996 0.6737991 0.6769384
X5 0.8091758 0.8198258 0.8057904 1.0000000 0.9050509 0.8665732 0.8539900
X6 0.7815510 0.8013282 0.7197996 0.9050509 1.0000000 0.9733801 0.7905565
X7 0.7278784 0.7318546 0.6737991 0.8665732 0.9733801 1.0000000 0.7987302
X8 0.6689597 0.6799537 0.6769384 0.8539900 0.7905565 0.7987302 1.0000000
```

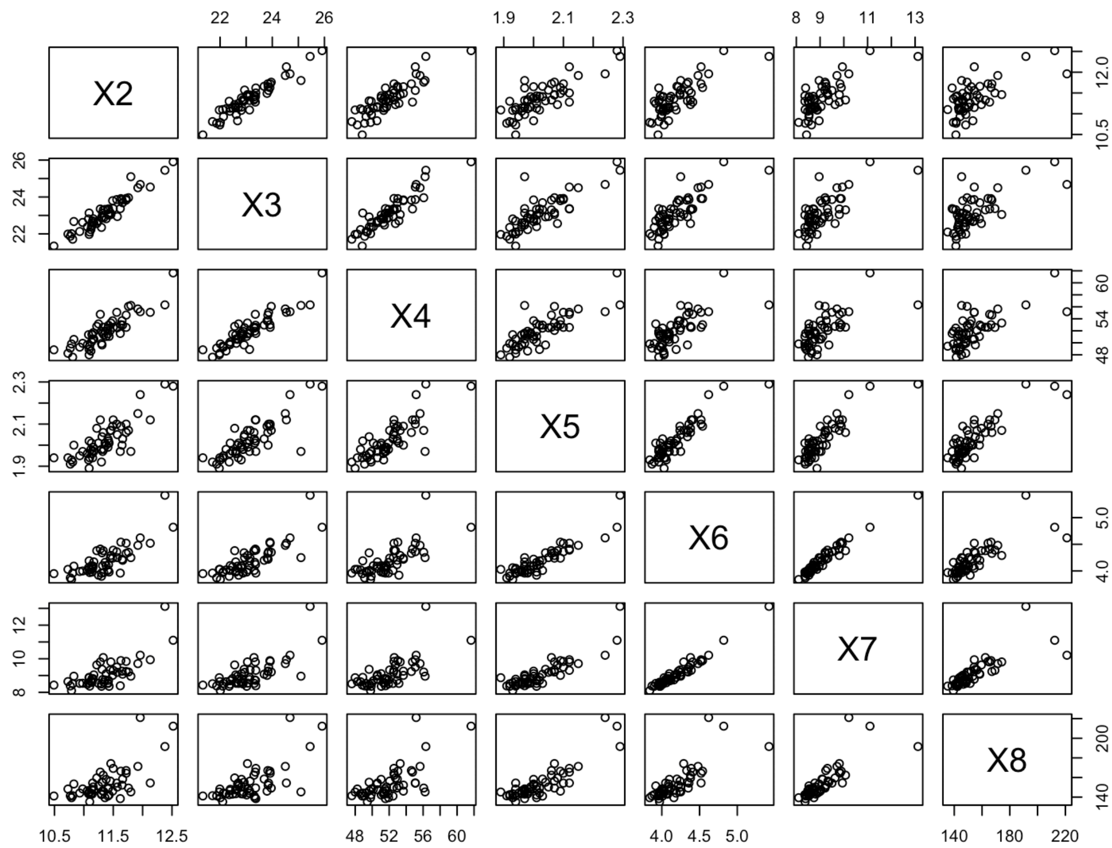
Those results highlight the symmetry of both matrices. Furthermore, from this particular dataset shows how only looking to the covariance matrix a strong relationship between X8 and the rest of the variables could be inferred, but then the correlation matrix turns down the hypothesis.

b) **Generate and study the scatterplots between each pair of variables. Any extreme values?**

These scatterplots can be generated by the following command:

```
pairs(dataset[, 2:8])
```

which returns the following figure:

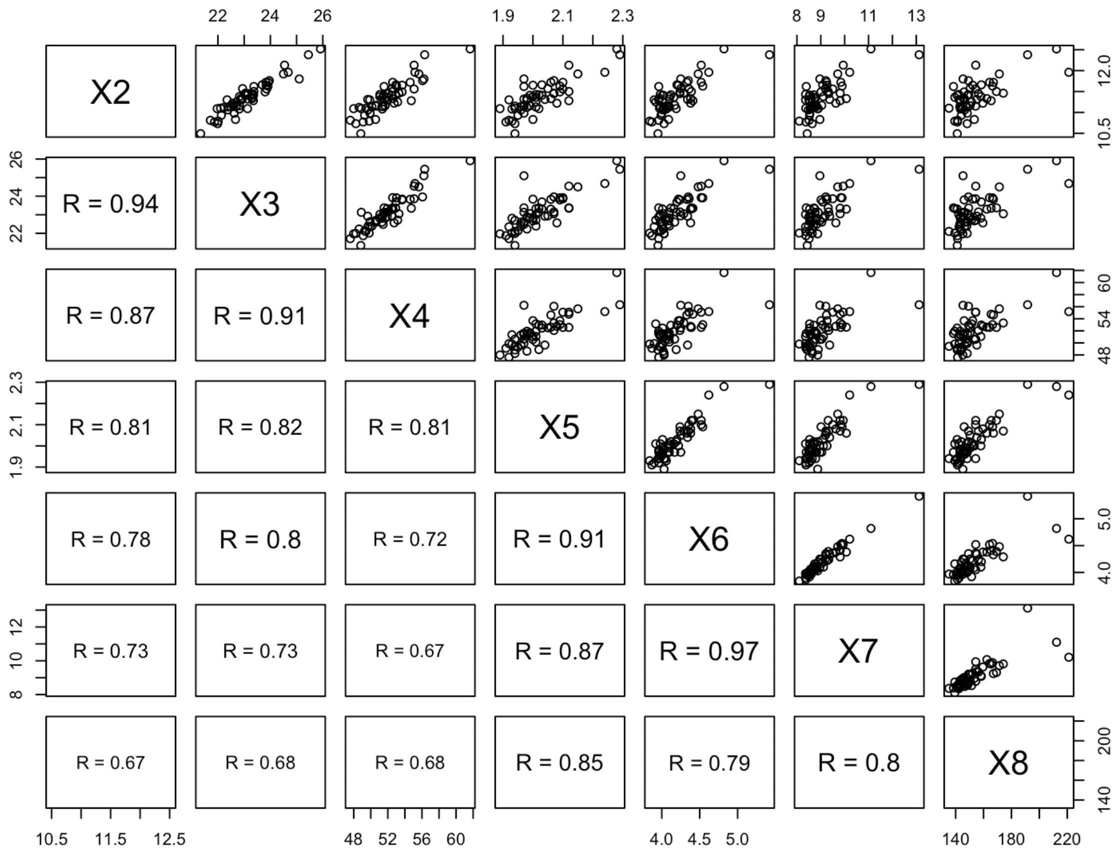


Deducing the strongest and weakest relationship from this plot could mean a hard job, but the arguments of the previous function can be slightly modified in order to facilitate the work:

```
panel.cor <- function(x, y){
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- round(cor(x, y), digits=2)
  txt <- paste0("R = ", r)
  cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}

pairs(dataset[, 2:8], lower.panel = panel.cor)
```

The *panel.cor* function calculates the correlation between two certain variables and associates a font size to it. The new figure looks like the following:



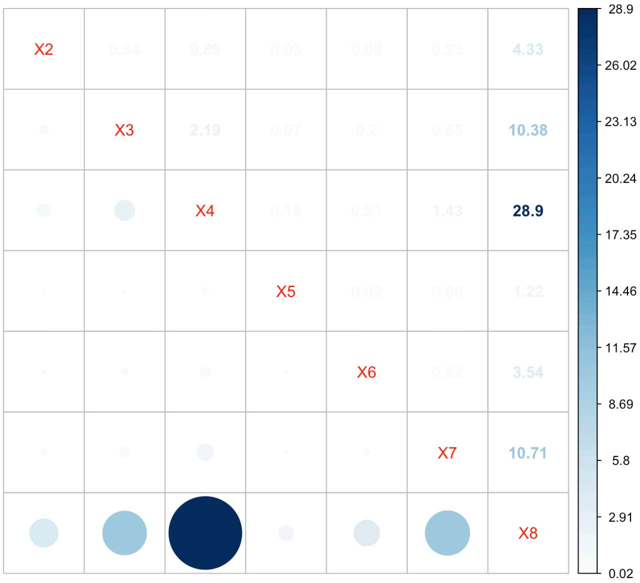
c) **Explore what other plotting possibilities R offers for multivariate data. Present other (at least two) graphs that you find interesting with respect to this data set.**

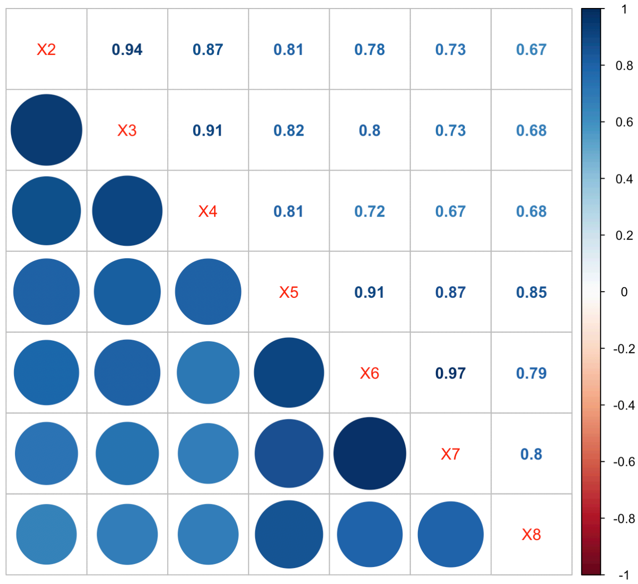
R offers interesting option for plotting a multivariate dataset. Including the needed libraries, three attractive figures are shown. If the following code is typed:

```
library(corrplot)
library(ellipse)

corrplot.mixed(cov_matrix, is.corr = FALSE, upper = "number", lower = "circle")
corrplot.mixed(cor_matrix, upper = "number", lower = "circle")
plotcorr(cor_matrix, type="lower", diag=FALSE, main="Bivariate correlations")
```

These three functions respectively return the next plots:





Bivariate correlations