

## Solution Sheet 1

The data we are using are from a vegetation study on an alp in the canton of Ticino. For each of 82 “plots”, the data set tells us the abundance of 64 plant species, values of 4 soil chemistry variables at each of 3 depths, 2 variables describing the intensity of grazing, as well as further variables. For a more detailed description of this data set, see

<http://stat.ethz.ch/~stahel/courses/multivariate/datasets/vegenv.html>

You can load these data into R by typing

```
d.vegenv ← read.table(
  "http://stat.ethz.ch/~stahel/courses/multivariate/datasets/vegenv.dat",
  header=TRUE) at the prompt.
```

**The full R code will be made available in a separate file on the course homepage.**

1. a) There is a clear linear relationship between C10 and N10.  
 b) To visualize the the distribution wrt. grazing in a monochrome picture, we vary the shape of the plotting symbols rather than their colour (in a slight change from the problem sheet):  

```
pairs(d.vegenv[c("pH10", "P10", "N10", "C10")], pch=
  ifelse(d.vegenv$Dungdensity>10, "o", "x"))
```

We can see that the group with intense grazing (“o”) tends towards higher values of P10.

 c) The plot displaying histograms and correlations is shown in Figure 1.
  
2. A coplot is shown in Figure 2. We can see that the variability “pH10” and “P10” is much smaller for low “Dungdensity” and high “Slope” than it is for high “Dungdensity” and low “Slope”.
  
3. Figure 3 displays a star plot. The advantage of this type of plot is that five different pieces of information can be displayed simultaneously (for each of the 82 “plots” in the study). Each symbol represents a single “plot”. The positions of the symbols in the graphic tells us what values the variables N (y-coordinate) and P (whose logarithm in base 10 is the x-coordinate here) have. The shape of the symbols can be interpreted as shown (in the legend) in the lower right corner of the plot: the length of the line section pointing to the top left shows us how many “Caluvulg” plants were found on that particular plot of land. The counts of “Nardstri” and “Festrubr” are displayed by the lines pointing to the right and to the bottom left, respectively. Thus if the symbol at a particular point also consists of a line from top left to bottom right, it means that only “Caluvulg” was found on this plot of land.

To facilitate the plotting of this legend, we have created an alternative function for R.

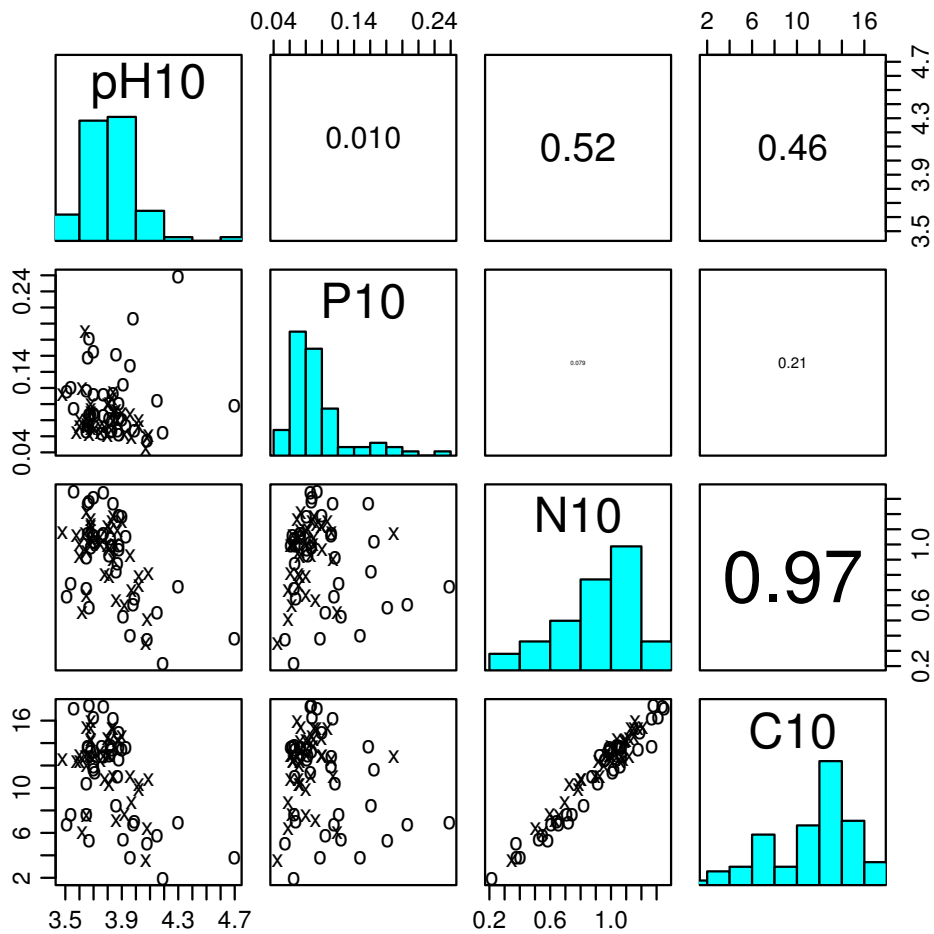


Figure 1: Pairs plot with histograms on the diagonal and correlations above it. The strong correlation between C10 and N10 is evident. Moreover, high grazing intensity seems to be associated to increased values of P10.

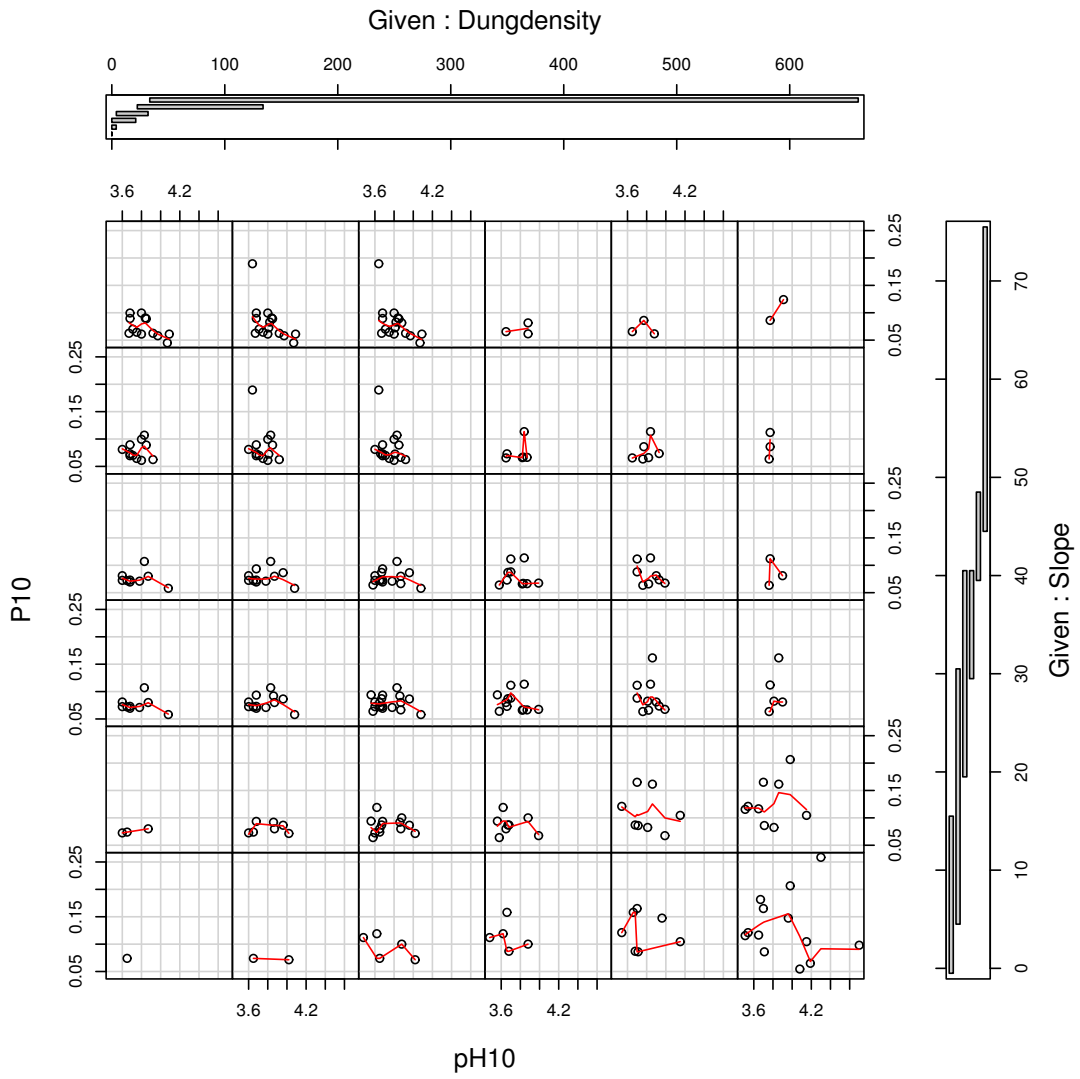


Figure 2: Coplot

```
stars(as.matrix(d.vegenv[,c("Nardstri","Caluvulg","Festrubr")]),
      location=cbind(log10(d.vegenv[, "P10"]),d.vegenv[, "N10"]),
      axes=TRUE, key.loc=c(-0.5,0.3),len=0.07,full=TRUE,
      xlab="log10(P)", ylab="N", labels=NULL)
```

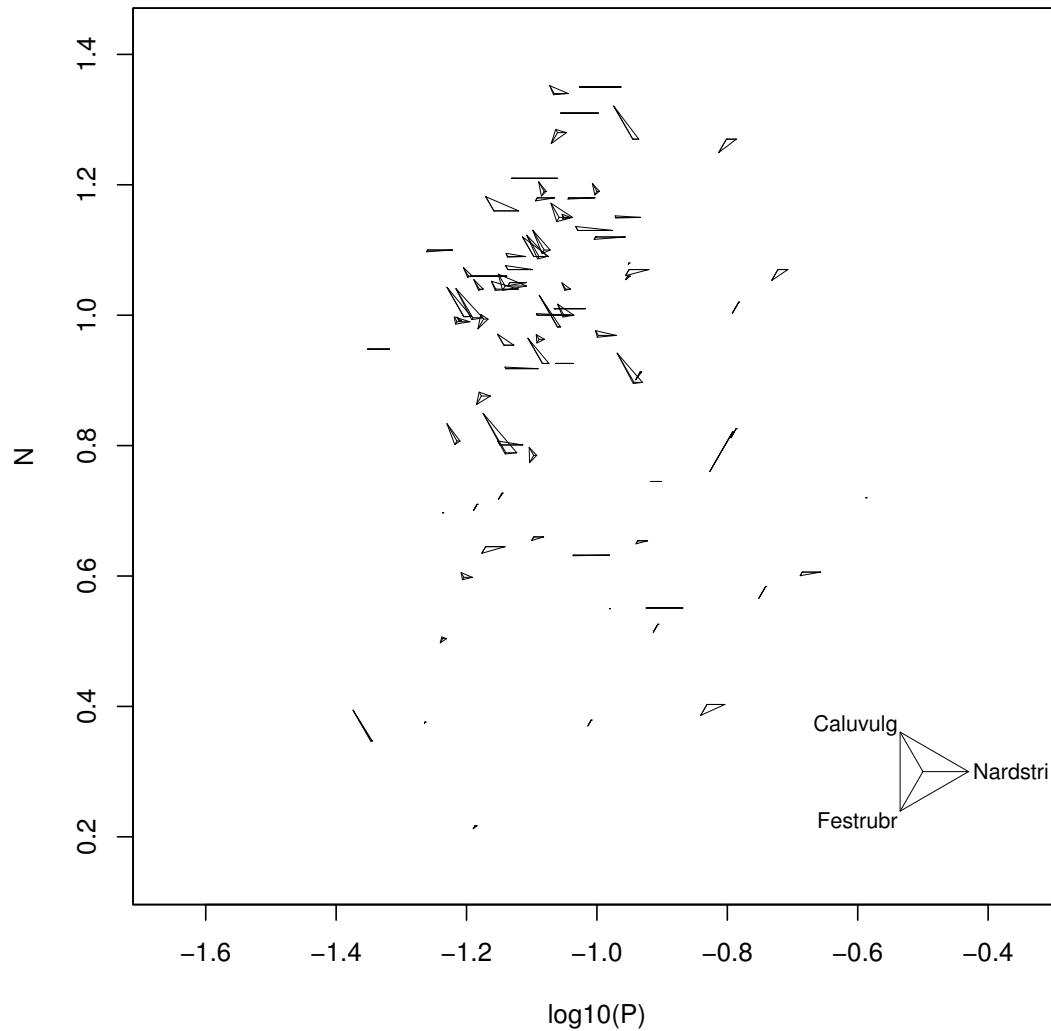


Figure 3: Star plot: We see that “Caluvulg” plants are found mainly in areas with low “P” and high “N”.

We see that “Caluvulg” plants are to be found mainly in areas with low “P” and high “N”. The other plants seem to grow in other circumstances, too (such as high “P” and low “N”). However, this is not a very clear effect.