

Introduction to the **Heatplus** package

Alexander Ploner
Medical Epidemiology & Biostatistics
Karolinska Institutet, Stockholm
email: `alexander.ploner@ki.se`

May 12, 2008

The **Heatplus** offers an extended heatmap display, including covariates and color-coding of clusters.

We start with preparing a structured data matrix that looks somewhat like a gene expression matrix with 45 samples and 130 genes:

```
> mm = matrix(rnorm(1000, m = 1), 100, 10)
> mm = cbind(mm, matrix(rnorm(2000), 100, 20))
> mm = cbind(mm, matrix(rnorm(1500, m = -1), 100, 15))
> mm2 = matrix(rnorm(450), 30, 15)
> mm2 = cbind(mm2, matrix(rnorm(900, m = 1.5), 30,
+ 30))
> mm = rbind(mm, mm2)
> colnames(mm) = paste("Sample", 1:45)
> rownames(mm) = paste("Gene", 1:130)
```

Note that here the samples are the columns of the data matrix; this is common for gene expression data, whereas it is traditional in statistics to have the samples as rows of the data matrix.

For the annotated heatplots, we construct an artificial data frame of variables describing further properties of the samples:

```
> addvar = data.frame(Var1 = rep(c(0, 1, 0), c(10,
+ 20, 15)), Var2 = rep(c(1, 0, 0), c(10, 20, 15)),
+ Var3 = rep(c(1, 0), c(15, 30)), Var4 = rep(seq(0,
+ 1, length = 4), c(10, 5, 15, 15)) + rnorm(45,
+ sd = 0.5))
> addvar[3, 3] = addvar[17, 2] = addvar[34, 1] = NA
```

Note that this data frame is in the traditional form, i.e. samples as rows, as common for phenotypic information accompanying an expression matrix; the first three variables are binary, the fourth is quantitative, and we have thrown in a few missing values to make it more interesting.

```
> library(Heatplus)
> heatmap_2(mm)
```

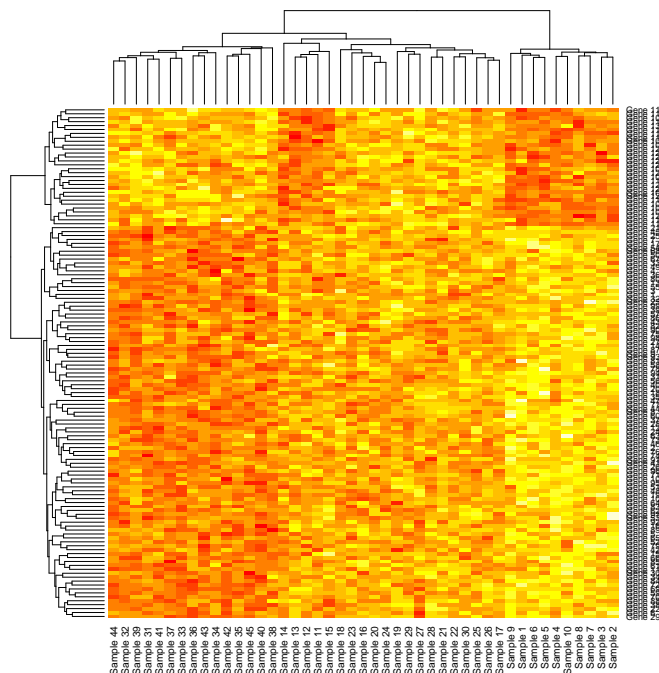


Figure 1: By default, the output of `heatmap_2` is very similar to the graph produced by `heatmap(mm)`.

The rest of this vignette consists of graphs produced by the functions `heatmap_2`, which produces a basic heat map similar to `heatmap`, and `heatmap_plus`, which produces annotated heat maps.

```
> heatmap_2(mm, do.dendro = c(FALSE, TRUE), legend = 1,
+   legfrac = 12)
```

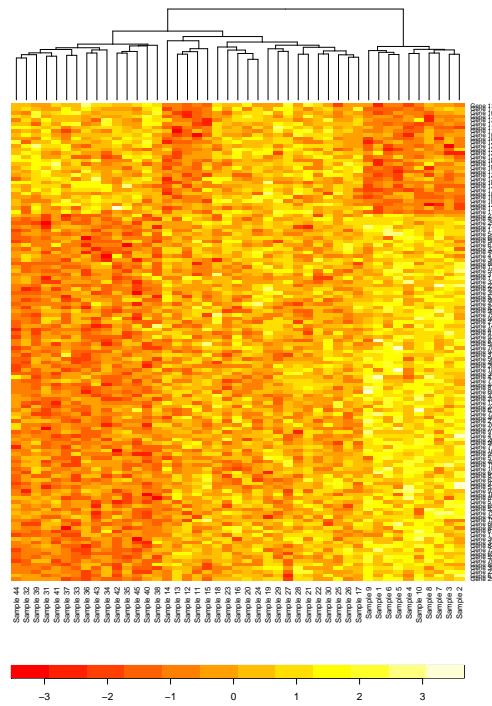


Figure 2: Here we suppress the plotting of the row dendrogram (though the rows are still sorted!), and we add a legend at the bottom, slightly smaller than the default.

```
> heatmap_plus(mm)
```

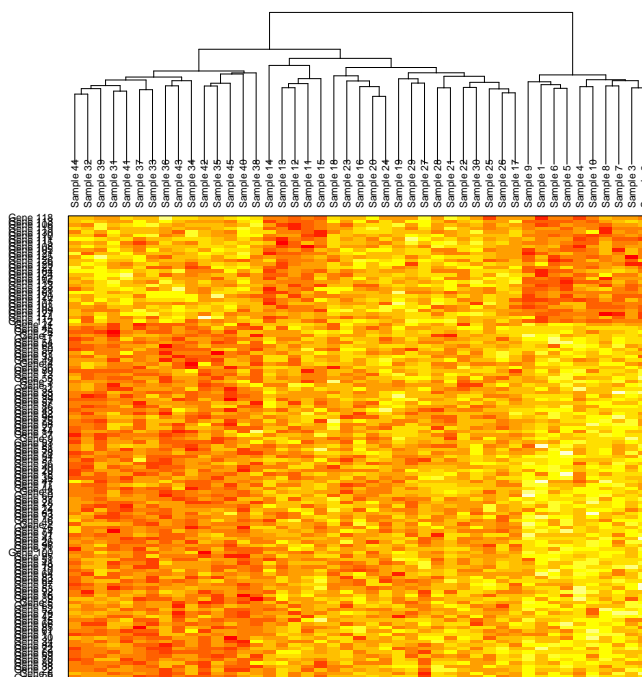


Figure 3: If we do not specify extra variables, `heatmap_plus` just produces a heatmap without drawing the row dendrogram (though as in Figure 2, the rows are sorted according to the dendrogram). Note however, how the sample labels are now plotted under the column dendrogram; this does not look nice, but the need for this will become clear in Figure 4.

```
> colnames(mm) = NULL
> heatmap_plus(mm, addvar = addvar, cov = 4)
```

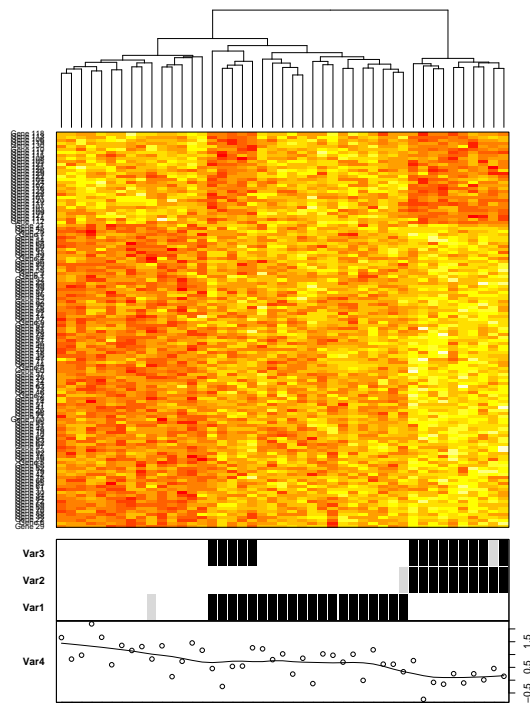


Figure 4: Here we set column names to `NULL` in order to get rid of the ugly labels seen in Figure 3. Then we produce an annotated heat map, specifying that we do have a covariate in column 4 of `addvar`.

```
> heatmap_plus(mm, addvar = addvar, cov = 4, h = 20,
+   col = RGBColVec(64))
```

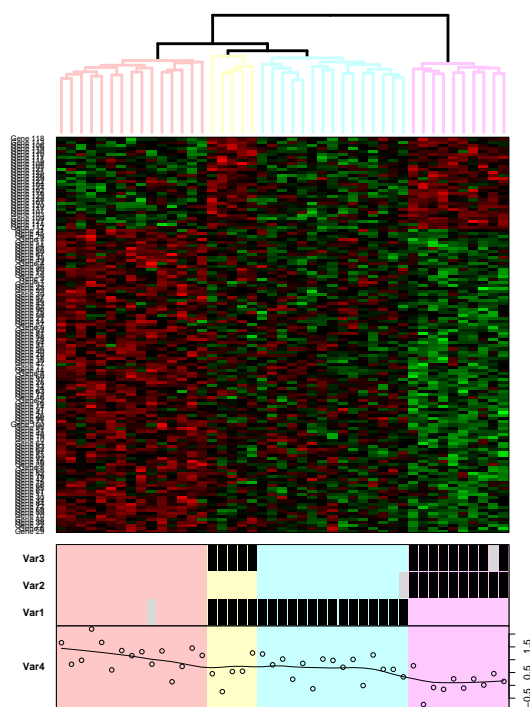


Figure 5: Here we produce an annotated heat map as in Figure 4, but we also specify that the column dendrogram is to be cut at height $h=20$. Note the different color scheme.