Exercise2_StutzSascha

Sascha Stutz

25 September 2017

Exercise 2

Exploratory Data Analysis

Do an exploratory data analysis of a matrix of expression values. Load the data and display: * distributions: *boxplot*, *density*, *limma::plotDensities* * normalization: *limma::normalizeQuantiles* * clustering: *hclust* * heatmap: *heatmap.2* or *pheatmap* * correlation matrix: *cor* and *image* * reduced dimensionality representation: *cmdscale* and *prcomp*

import

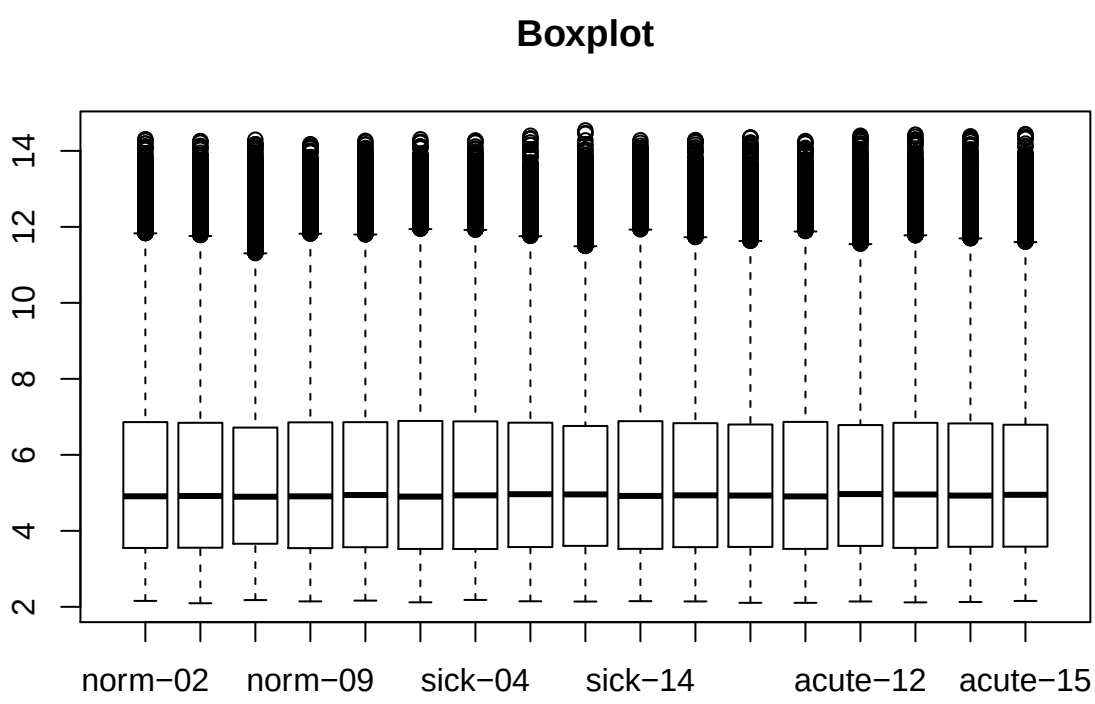
```
anno = read.table("Data/SampleAnnotation.txt", as.is=TRUE, sep="\t", quote="", row.names=1, header=TRUE)

x = read.table("Data/expressiondata.txt", as.is=TRUE, sep="\t", quote="", row.names=1, header=TRUE, che
x = as.matrix(x)
```

define smaples and colors and phenotype

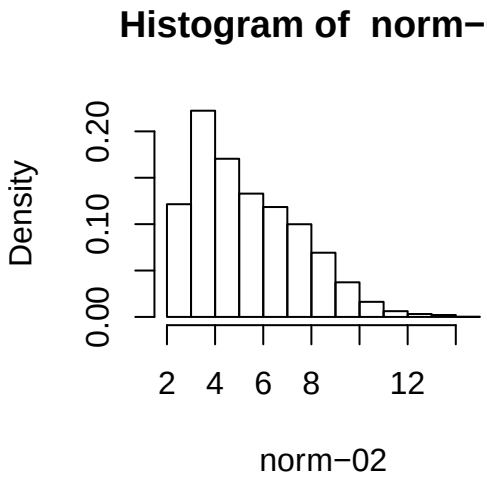
```
samples = rownames(anno)
colors = rainbow(nrow(anno))
isNorm = anno$TissueType == "norm"
isSick = anno$TissueType == "sick"
isAcute = anno$TissueType == "acute"
```

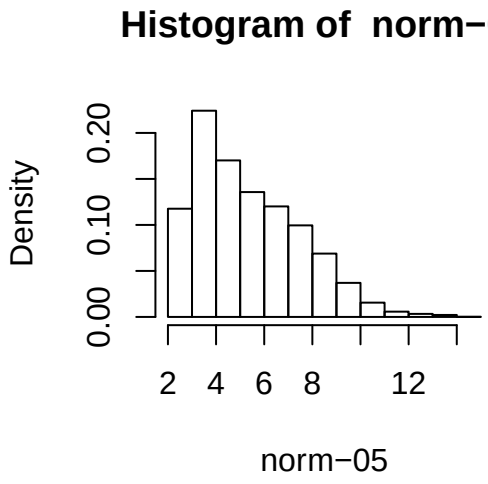
plot distributions

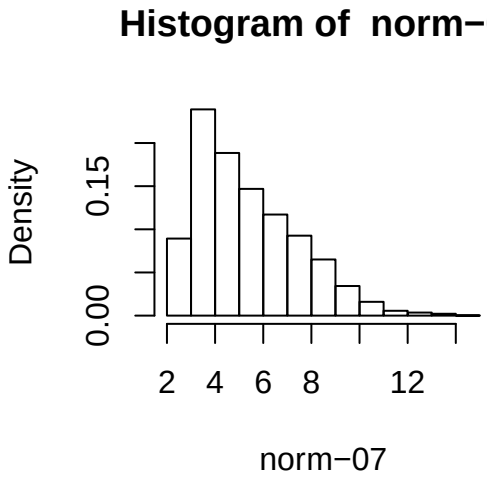
```
logx <- log2(x)
boxplot(logx, main = "Boxplot")
```

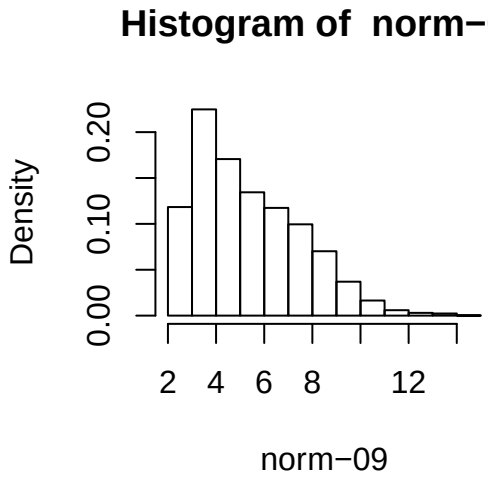


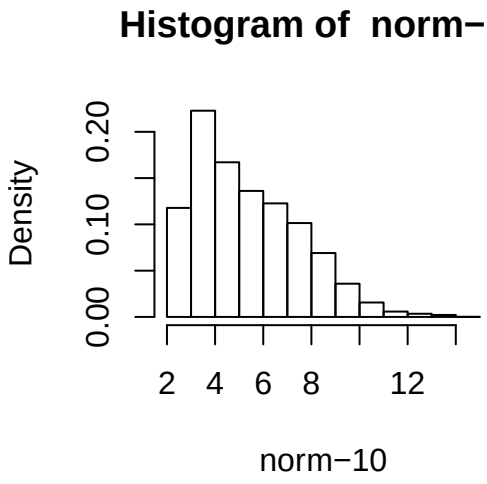

```
for(i in 1:ncol(logx)){
  hist(logx[,i], freq = FALSE, xlab = colnames(logx)[i], main = paste("Histogram of ", colnames(logx)[i]
}
```

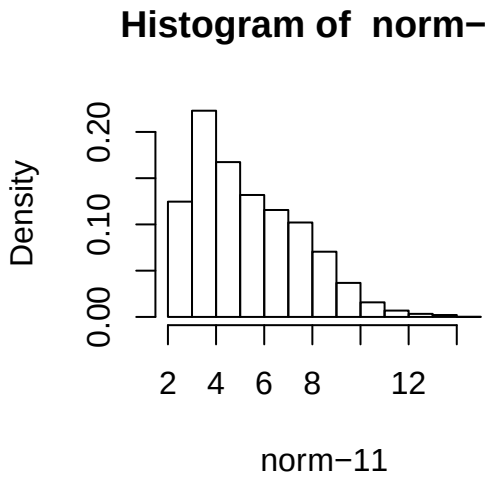

Histogram of norm−02
Density
0.00
0.10
0.20
2
4
6
8
12
norm−02

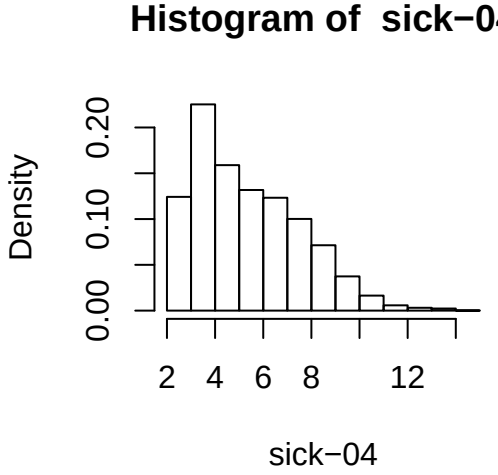

Histogram of norm−05
Density
0.00
0.10
0.20
2
4
6
8
12
norm−05

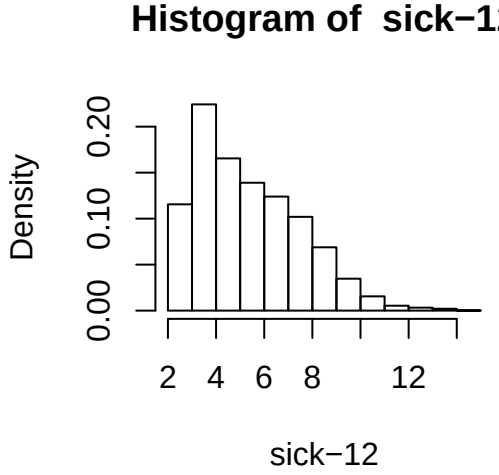

Histogram of norm−07
Density
0.00
0.15
2
4
6
8
12
norm−07

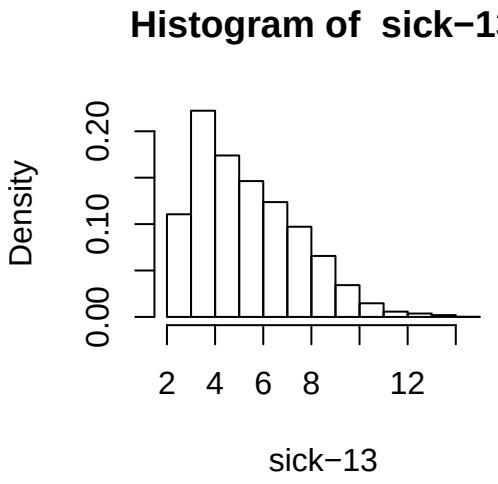

Histogram of norm−09
Density
0.00
0.10
0.20
2
4
6
8
12
norm−09

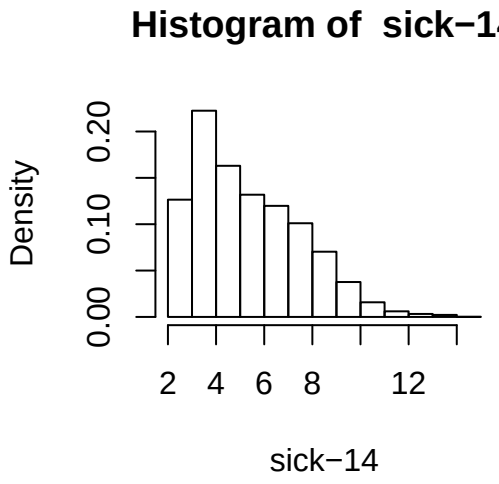

Histogram of norm−10
Density
0.00
0.10
0.20
2
4
6
8
12
norm−10

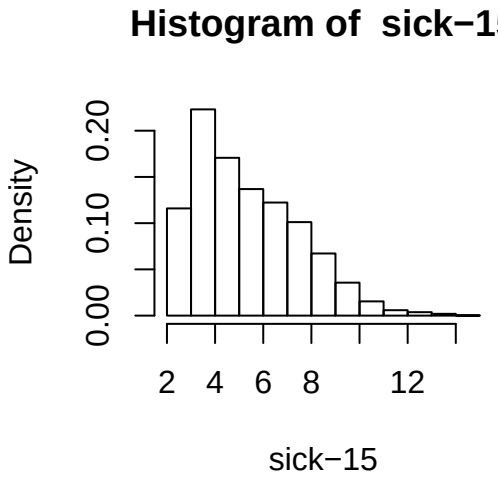

Histogram of norm−11
Density
0.00
0.10
0.20
2
4
6
8
12
norm−11

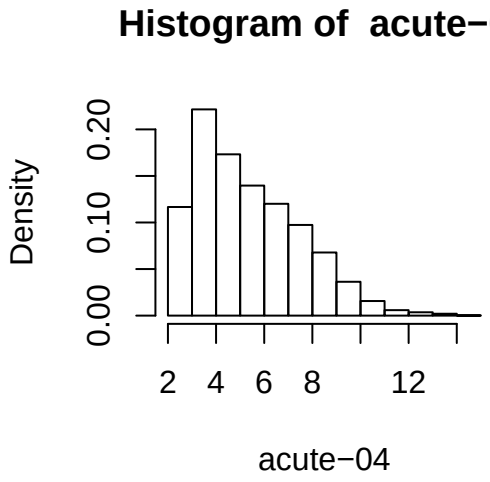

Histogram of sick−04
Density
0.00
0.10
0.20
2
4
6
8
12
sick−04

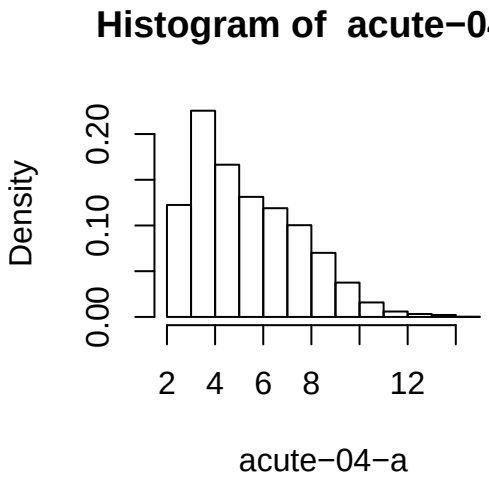

Histogram of sick−12
Density
0.00
0.10
0.20
2
4
6
8
12
sick−12

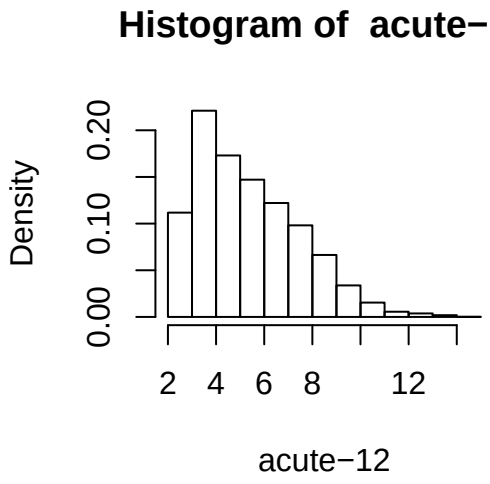

Histogram of sick−13
Density
0.00
0.10
0.20
2
4
6
8
12
sick−13

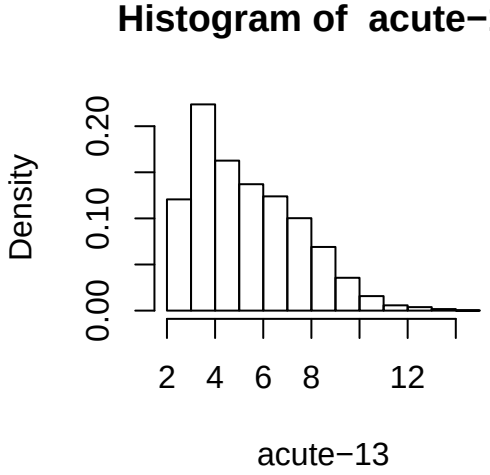

Histogram of sick−14
Density
0.00
0.10
0.20
2
4
6
8
12
sick−14

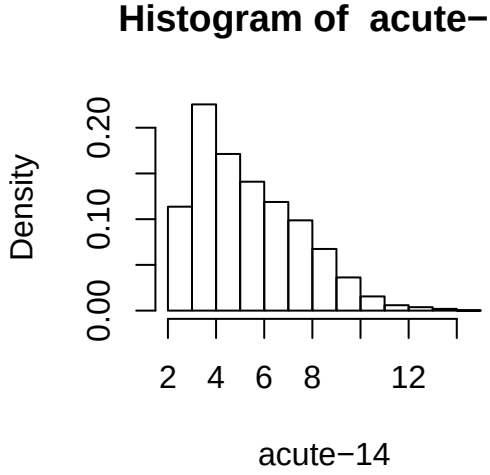

Histogram of sick−15
Density
0.00
0.10
0.20
2
4
6
8
12
sick−15

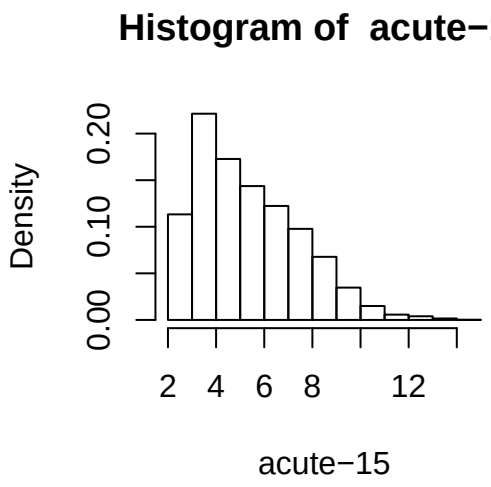

Histogram of acute−04
Density
0.00
0.10
0.20
2
4
6
8
12
acute−04


Histogram of acute−04−a
Density
0.00
0.10
0.20
2
4
6
8
12
acute−04−a

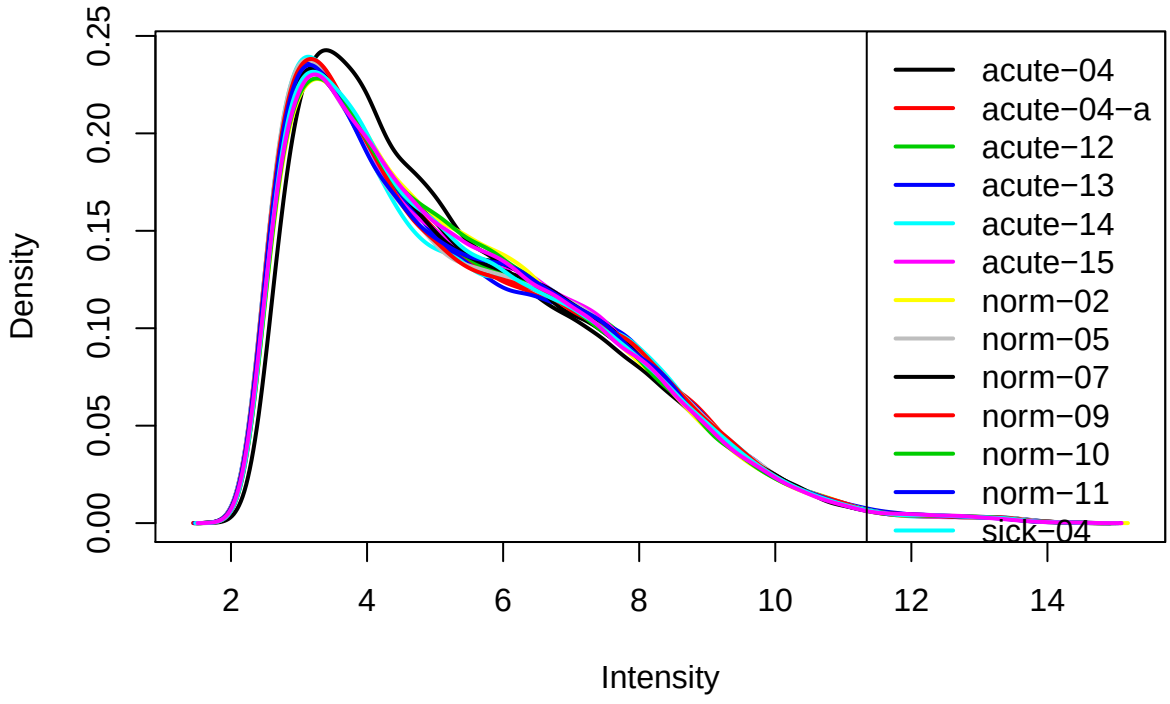

Histogram of acute−12
Density
0.00
0.10
0.20
2
4
6
8
12
acute−12


Histogram of acute−13
Density
0.00
0.10
0.20
2
4
6
8
12
acute−13


Histogram of acute−14
Density
0.00
0.10
0.20
2
4
6
8
12
acute−14

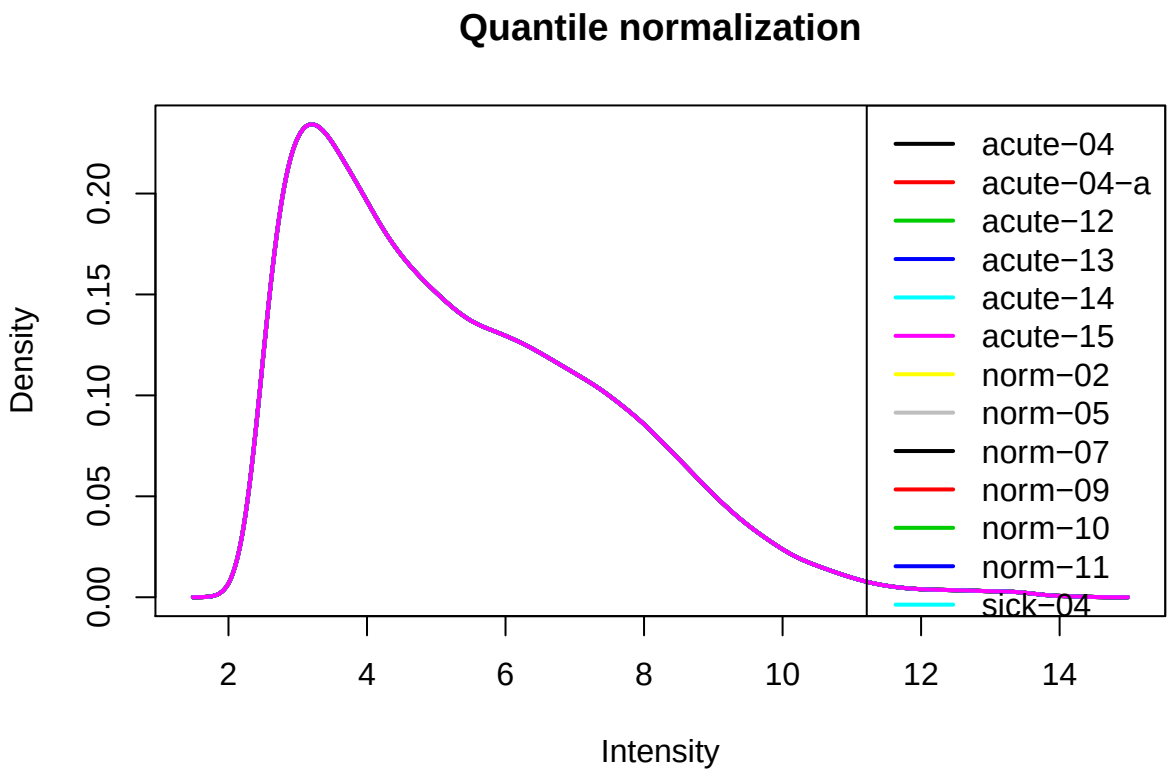



```
limma::plotDensities(logx, legend = "topright")
```



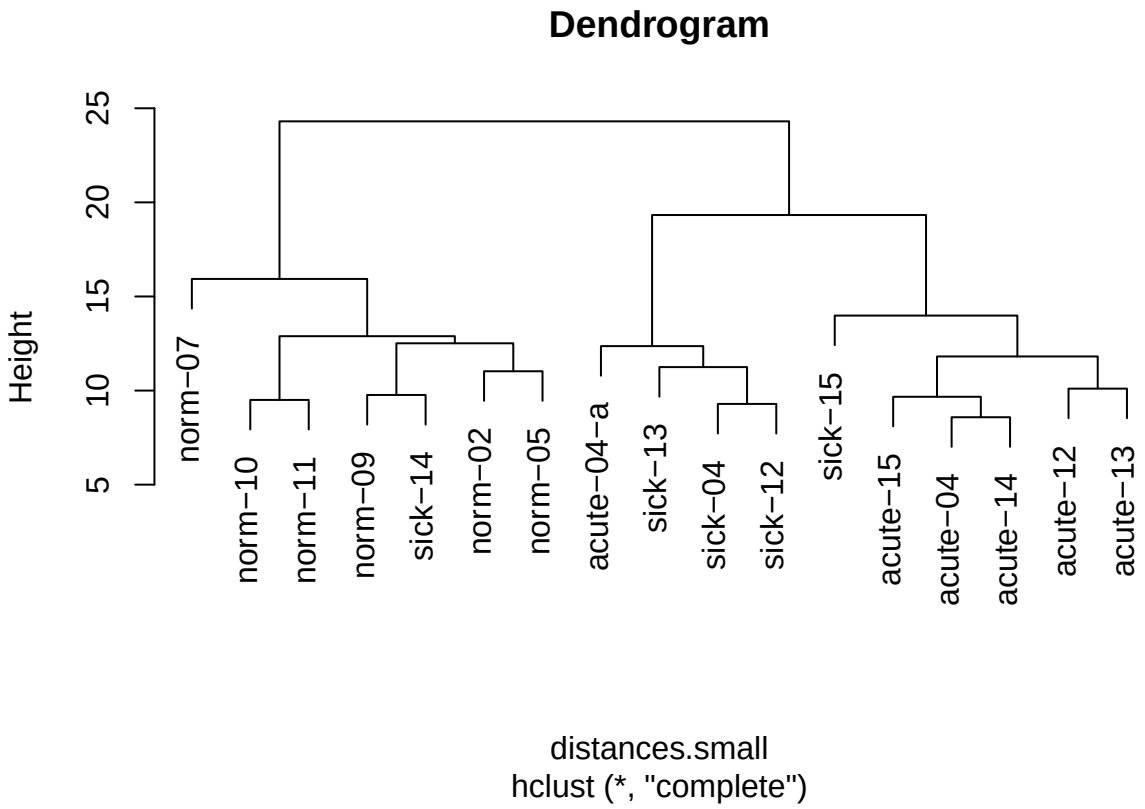
normalization

```
norm.logx = limma::normalizeQuantiles(logx)
limma::plotDensities(norm.logx,main = 'Quantile normalization',legend="topright")
```



hclust

```
distances.small = dist(t(as.matrix(logx[sample(1:nrow(logx), size = 1000),])))
clusters = hclust(distances.small)
plot(clusters, main="Dendrogram")
```

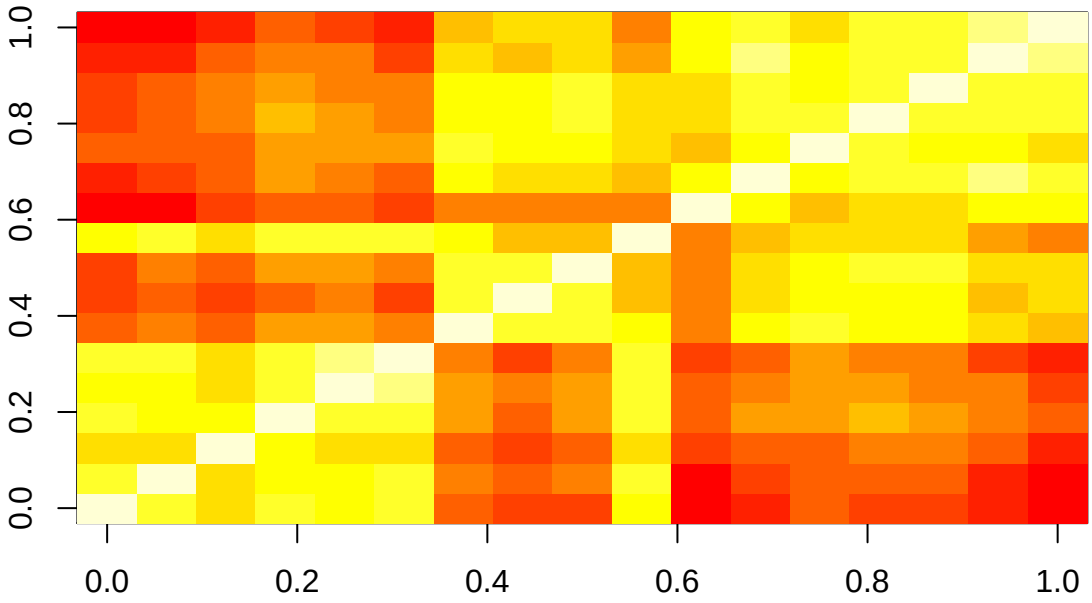


heatmap.2

```
library(pheatmap)
# pheatmap(distances.small,main="Heatmap",legend = TRUE,annotation_legend = TRUE, annotation_names_row =
# data way too big, can't run code
```

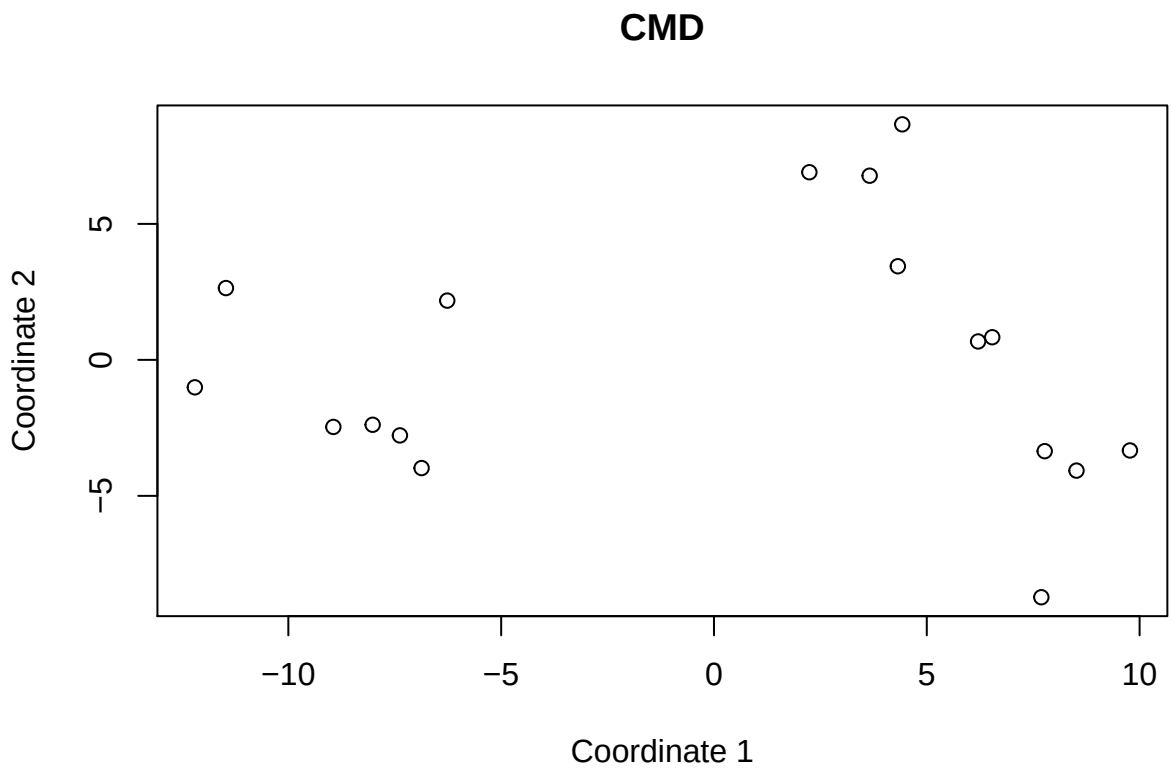
correlation matrix

```
image(cor(logx))
```

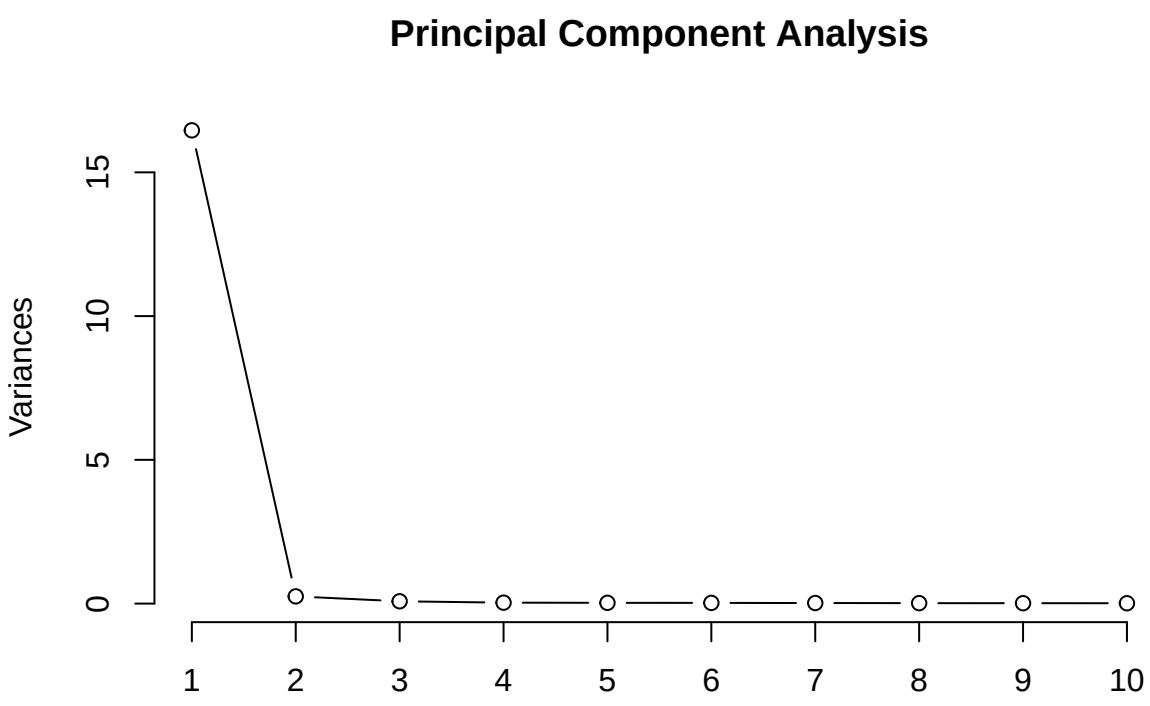


reduced dimensionality representation

```
fit.cmd <- cmdscale(distances.small, eig=TRUE, k=2)
x <- fit.cmd$points[,1]
y <- fit.cmd$points[,2]
plot(x, y, xlab="Coordinate 1", ylab="Coordinate 2",
  main="CMD")
```




```
fit.pca <- prcomp(logx, center = TRUE, scale. = TRUE)
plot(fit.pca, type = "l", main = "Principal Component Analysis")
```


Principal Component Analysis
Variances
0
5
10
15
1
2
3
4
5
6
7
8
9
10