

```
#####
# 9. PCA and CLUSTER ANALYSIS
#####
## clear memory
rm(list = ls())

##-----
## 9.1. Iris data
##-----

## show the data
iris
str(iris)

## plot iris data
x11()
plot(iris[,-5])
## more beautiful
plot(iris[,-5],col = iris[,5],pch=19)

##-----
## 9.2 PCA
##-----

## Let's transform data frame into numerical matrix
## data - numerical data
## classes - type of iris, 1=setosa, 2=versicolor, 3=virginica
Data = as.matrix(iris[,-5])
row.names(Data) = as.character(iris[,5])
classes = as.integer(iris[,5])

PC = prcomp(Data)
str(PC)
x11()
plot(PC$x[,1],PC$x[,2],col = classes,pch=19)

## plot in 3D
library(rgl)
plot3d(PC$x[,1],PC$x[,2],PC$x[,3],col = classes,type="s")

##-----
## 9.3 k-Means clustering
##-----

## try k-means clustering
clusters = kmeans(x=Data,centers=3,nstart=10)$cluster

## show clusters on PCA
x11()
plot(PC$x[,1],PC$x[,2],col = classes,pch=clusters)
legend(2,1.4,levels(iris$Species),col=c(1,2,3),pch=19)
legend(-2.5,1.4,c("c1","c2","c3"),col=4,pch=c(1,2,3))

##-----
## 9.4 Hierarchical clustering
##-----

## use heatmap
heatmap(Data)

## use heatmap with colors
color = character(length(classes))
```

```
color[classes == 1] = "black"
color[classes == 2] = "red"
color[classes == 3] = "green"
heatmap(Data,RowSideColors=color)

##-----
## 9.5 Solution of task 9a
##-----
Data = read.table("http://edu.sablab.net/data/txt/all_data.txt",
                  as.is=T,header=T,sep="\t")
str(Data) ## see here - experiments are in columns!!!

## create read.table te classes: 1 - normal, 2 - cancer
classes = integer(ncol(Data[, -c(1,2)]))
idx.norm = grep("normal", names(Data[, -c(1,2)]))
idx.all = grep("ALL", names(Data[, -c(1,2)]));
classes[idx.norm] = 1
classes[idx.all] = 2

## PCA
PC = prcomp(t(Data[, -c(1,2)]))
str(PC)
x11()
plot(PC$x[,1], PC$x[,2], col = classes, pch=19)

## 3D
library(rgl)
plot3d(PC$x[,1], PC$x[,2], PC$x[,3], col = classes, type="s")
```