1. Save the following data in a text file called “DataGE1.txt” in the working directory of your R programming and calculate Euclidean distance and correlation matrix. You can use bioDist package.

|  |  |  |  |
| --- | --- | --- | --- |
| 1. name | 1. EX1 | 1. EX2 | 1. EX3 |
| 1. G1 | 1. 10 | 1. 8 | 1. 10 |
| 1. G2 | 1. 10 | 1. 0 | 1. 9 |
| 1. G3 | 1. 4 | 1. 8.5 | 1. 3 |
| 1. G4 | 1. 9.5 | 1. 0.5 | 1. 8.5 |
| 1. G5 | 1. 4.5 | 1. 8.5 | 1. 2.5 |
| 1. G6 | 1. 10.5 | 1. 9 | 1. 12 |
| 1. G7 | 1. 5 | 1. 8.5 | 1. 11 |
| 1. G8 | 1. 2.7 | 1. 8.7 | 1. 2.0 |
| 1. G9 | 1. 9.7 | 1. 2.0 | 1. 9.0 |
| 1. G10 | 1. 10.2 | 1. 1.0 | 1. 9.2 |

1. Run the following R commands and submit the results.

data <- read.table("DataGE1.txt", header=T, row.names="name")

dist <- dist(data, method = "euclidian")

plot(hclust(dist, method = "complete"))

heatmap(as.matrix(data))

1. Try to run the following commands on R and print and explain the results.

x <- rbind(matrix(rnorm(50, mean = 1, sd = 0.4), ncol = 2),

matrix(rnorm(50, mean = 5, sd = 0.3), ncol = 2),

matrix(rnorm(50, mean =10, sd = 0.5), ncol = 2))

colnames(x) <- c("A", "B")

cl <- kmeans(x, 3)

plot(x, col = cl$cluster)

points (cl$centers, col = 1:3, pch = 10, cex=2)

1. Calculate mahalanobis distances between genes/entities using the file "PCATutorial.txt" and plot the histogram of the distance values.