Yes or no. Motivate you answer:

1. Performing hierarchical clustering with a correlation based similarity metric applied on the non-rescaled data gives exactly the same results as when you would use hierarchical clustering with a Euclidian distance applied on the non-rescaled data.
2. Increasing the number of clusters as a parameter reduces the cluster size
3. The is the output of a PCA analysis on the following dataset

Explain in words what PCA is doing.

What is the meaning/interpretation of loadings and scores (see graph below) .





3)

a) What would this code be doing? Golub is a gene-patient dataset. The dataset contains x patients and 3051 genes. Explain the code line per line. Explain why you need to perform the different steps. Explain what you expect in the output.

golub\_m = golub – rowMeans(golub) [code on the whole dataset]

SD = apply(golub, 1, sd, na.rm=TRUE)

golub\_r = golub\_m/SD

d.euclidian <- dist(golub\_r, method="euclidian")

m <- as.matrix(dist(golub))

dim(m)

#(3051,3051)

kmeangolub100= kmeans(d.euclidian\_r,100)

b) what will happen if you change the following parameter setting

kmeangolub10= kmeans(d.euclidian\_r,10)