```
################################## Clustering analysis ###############################


#### instaling and loading a package called "cluster"
>install.packages("cluster")
>library (cluster)


#### load a data
>data("animals")


### let's investigate more of the data
>?animals


#### to keep going with rest of the analysis we need to change the 1 and 2 values to 0 and 1
>animals.bin <- animals
>animals.bin[animals == 1] <- 0
>animals.bin[animals == 2] <- 1


#### let's calculate the paiwise distance between animals
## first let's figure out what code "daisy" does
>?daisy
>diss.animals <- daisy(animals.bin, type=list(symm=1:6))
>diss.animals


##### let's plot the dissimilarities
>par(mfrow=c(1,2))


#### using the code agnes to agglomerative tree algorithm, and specify a single linkage
### lat's first see what code "agnes" details are
>?agnes
>agnes.tree <- agnes(diss.animals, method="single")
```

```
>plot(agnes.tree, ask=T)
```

let's try a divisive method to make a tree

```
>diana.tree <- diana(diss.animals)
>plot(diana.tree)
```

how well do these two methods match?

```
>par(mfrow=c(1,2))
>plot(agnes.tree, ask=T)
>plot(diana.tree, ask=T)
```

let's remove the missing values

```
>animals.complete <- animals.bin[complete.cases(animals.bin),]
```

apply K-means and start it with K=2

```
>kmeans.animals <- kmeans(animals.complete,centers=2)
```

now let's see the variables

```
>names(kmeans.animals)
```

let's investigate variables into more details

```
>kmeans.animals$cluster
>kmeans.animals$centers
>kmeans.animals$withinss
>kmeans.animals$size
```

now let's plot clusters but we need to use the PCA to reduce the dimensionality

```
>pca.animals <- princomp(scale(animals.complete))
>plot(pca.animals$scores[,1:2],col=kmeans.animals$clust,
    pch=kmeans.animals$clust,main="K-Means clusters")
```

medoids are representative data points in each cluster whose average distance to all other objects in the cluster is minimal

```
>pam.animals <- pam(diss.animals, k=2)
```

let's plot it

```
>par(mfrow=c(1,2))
>clusplot(pam.animals)
>plot(pam.animals)
```

Let's explore the fitted model object

```
>names(pam.animals)
>pam.animals$medoids
>pam.animals$clustering
>pam.animals$clusinfo
```