Exercice 1

```
1. distX = as.matrix(X)
    distX = distX ^2

2. Mthode 1
  XC = scale(X, scale=T)
  W = XC\%*\%t(XC)

  Mthode 2
  QN = diag(nrow(X)) - matrix(1, nrow(X), nrow(X))/nrow(X)
  W = -1/2*QN\%*\%distX\%*\%QN

3. Pour vrifier si elle est dfinie semi-positive, il suffit de vrifier que les valeurs propr
eigen(W)

4. L = eigen(W)$values
  L = diag(nrow(X))*L

  V = eigen(W)$vectors

5. C = V\%*\%sqrt(L)
  pas oublier de retirer les NaN

  plot(C)
  idem  biplot(princomp(X))
```

Exercice 2

```
m = as.vector(mutation)
b = cmdscale(mutation, 2, T)
c = as.vector(dist(b$points))
plot(b,c) problme mais on est pas loin
qualit  calculer avec les valeurs propres b[,1]$eigen etc... / sum

on refait de mme avec cmdscale(mutation, 3, T) jusqu' 5
```

Exercice 3

library(cluster) clusplot

¡¡iris, fig=TRUE¿¿= par(mfrow=c(1,3)) clusplot(iris, res2$cluster)res3 = kmeans(iris, 3)clusplot(iris, res3$cluster) res4 = kmeans(iris, 4) clusplot(iris, res4$cluster)$@

Figure 1: Visualisation de kmeans avec 2, 3 et 4 partitions

Iris

Question 1 - Diffrents nombres de partiton

Premirement, nous remarquons que les partitions n'ont pas toutes le mme nombre d'lments. Ensuite, elle varie suivant le nombre de partitions. En effet, nous pourrions penser qu'entre eux 3 et 4 partitions, l'ajout d'une partition subdiviserait une partition dj existante. Comme le montre les graphes ci dessous cela n'est pas le cas. En effet les 3 partitions de droit pour K=4 ne sont pas pas contenus dans entirement 2 partitions de K=3. Toutes les partitions sont redfinis chaque fois que nous augmontons le nombre.

Question 2 - Stabilit des partitions

De plus, mme pour un mme K, dans notre cas k=3, les partitions peuvent changer. Ici, nous avons deux cas diffrents avec des inerties de classes de 143 ou 78.9. Cela est du au choix alatoire des centres au dbut de l'algorithme.

Question 3 - Nombre de partitions optimales

```
$for(j in 2:10){
  for(i in 1:100){
    test[j, i] = kmeans(iris, j)$tot.withinss
  }
}$
apply(test, 2, min)
```

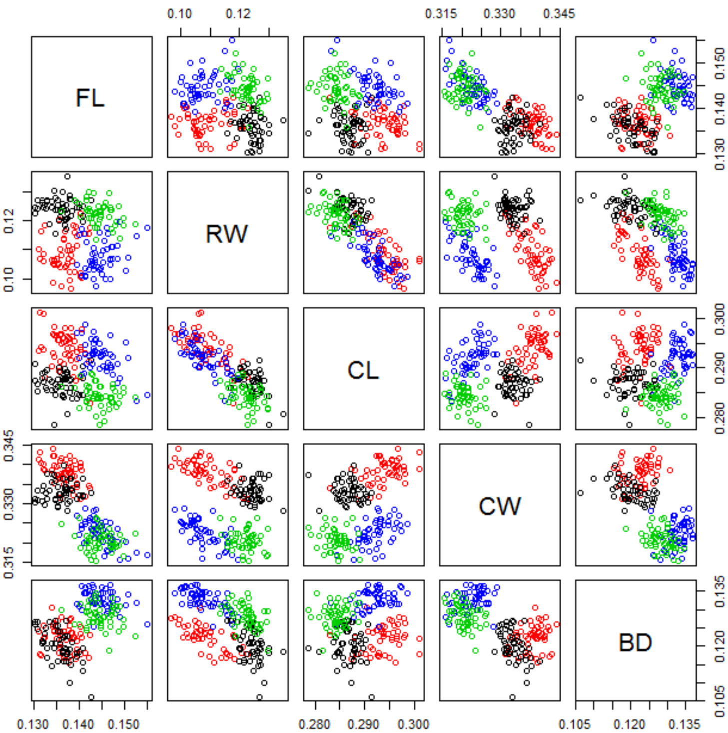
La solution optimale semble tre en 3 classes. Pourtant celle-ci n'est pas flagrante avec le tableau des minimums des inerties. La mthode du coude ne fonctionne pas trs bien, elle ne fait pas apparaitre de coude. Le minimum d'inertie de fait que diminuer en fonction du nombre de classes. Une solution serait de pnaliser un grand nombre de classes par le nombre d'individus prsents dans la classe.

Question 4 - Partitions relles

Crabs

```
library(MASS)
data(crabs)
crabsquant <- crabs[,4:8]
crabsquant <- crabsquant/matrix(rep(crabsquant[,4],dim(crabsquant)[2]),
nrow=dim(crabsquant)[1],byrow=F)
```

```
clusplot(crabsquant, kmeans(crabsquant, 4)$cluster)
plot(crabsquant, col =kmeans(crabsquant, 4)$cluster)
```



Mutations

```
res = kmeans(mutations2, 2)
plot(cmdscale(mutations), col=res$cluster)
```

Avec 3 vert au milieu des noirs

4 cluster seulement un point dans le dernier

tableau de contingence pour comparer les partitions table(res$cluster, res2$cluster)