

TP : Classification de données qualitatives

On considère les données `canines` étudiées dans le chapitre sur l'ACM. On voudrait déterminer la meilleure partition des 27 races canines, visualiser les classes de cette partition puis les décrire.

1. Calcul de la matrice de distance sur variables binaires

<https://www.rdocumentation.org/packages/ade4/versions/1.7-17/topics/dist.binary>

```
library(ade4)
dj_canines<-tab.disjonctif(canines)
dist<-dist.binary(dj_canines, method = 2)
```

2. Détermination du nombre de classes optimal pour la méthode des k-modes selon l'indice Silhouette

https://www.rdocumentation.org/packages/factoextra/versions/1.0.7/topics/fviz_nbclust

```
library(klaR)
library(factoextra)
fviz_kmodes = fviz_nbclust(canines, kmodes, method = "silhouette",dist,
k.max = 10, print.summary = TRUE)+
labs(subtitle = "K-modes silhouette method")
fviz_kmodes
print(data.frame(fviz_kmodes$data$clusters, fviz_kmodes$data$y))

# fviz_nbclust est applicable pour les indices suivants :
wss , gap_stat, Silhouette.
```

3. Visualisation des classes sur la première carte factoriel de l'ACM

4. Description des classes