Réduction de dimensionalité-clustering

Fabien Haury

2022-10-05

1 Librairies

```
library(dplyr)
library(tidyverse)
library(FactoMineR)
library(factoextra)
library(scatterplot3d)
library(corrplot)
library(ggfortify)
library(flextable)
library(ggthemes)
library(parameters)
library(NbClust)
```

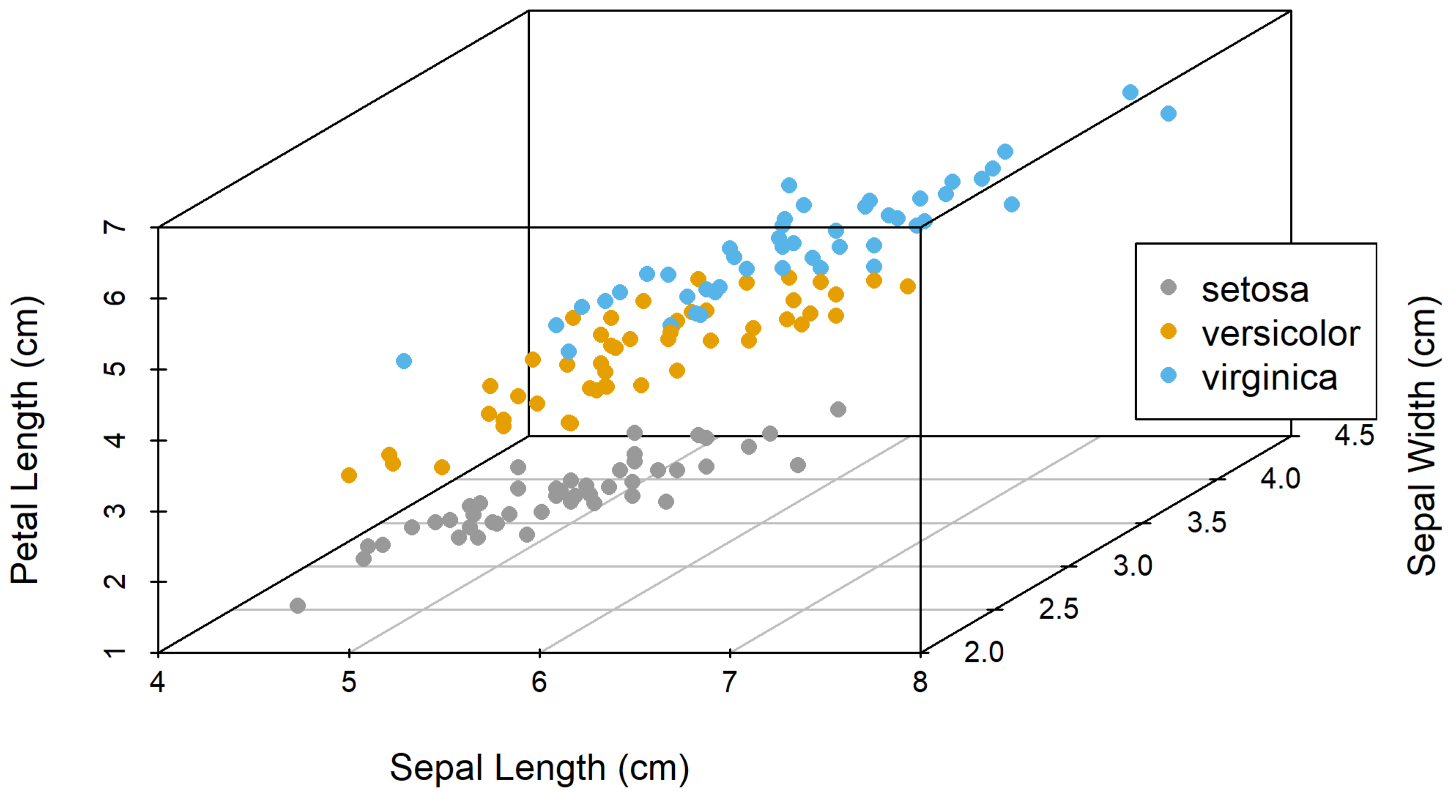
2 Import données

```
iris <- read_csv("Dataset/Iris.csv",
  locale = locale(encoding = "ISO-8859-2"),
  na = "NA", comment = "#"
)

iris$Species <- iris$Species %>%
  fct_recode(
    "setosa" = "Iris-setosa",
    "versicolor" = "Iris-versicolor",
    "virginica" = "Iris-virginica"
  )
```

3 Projet 1

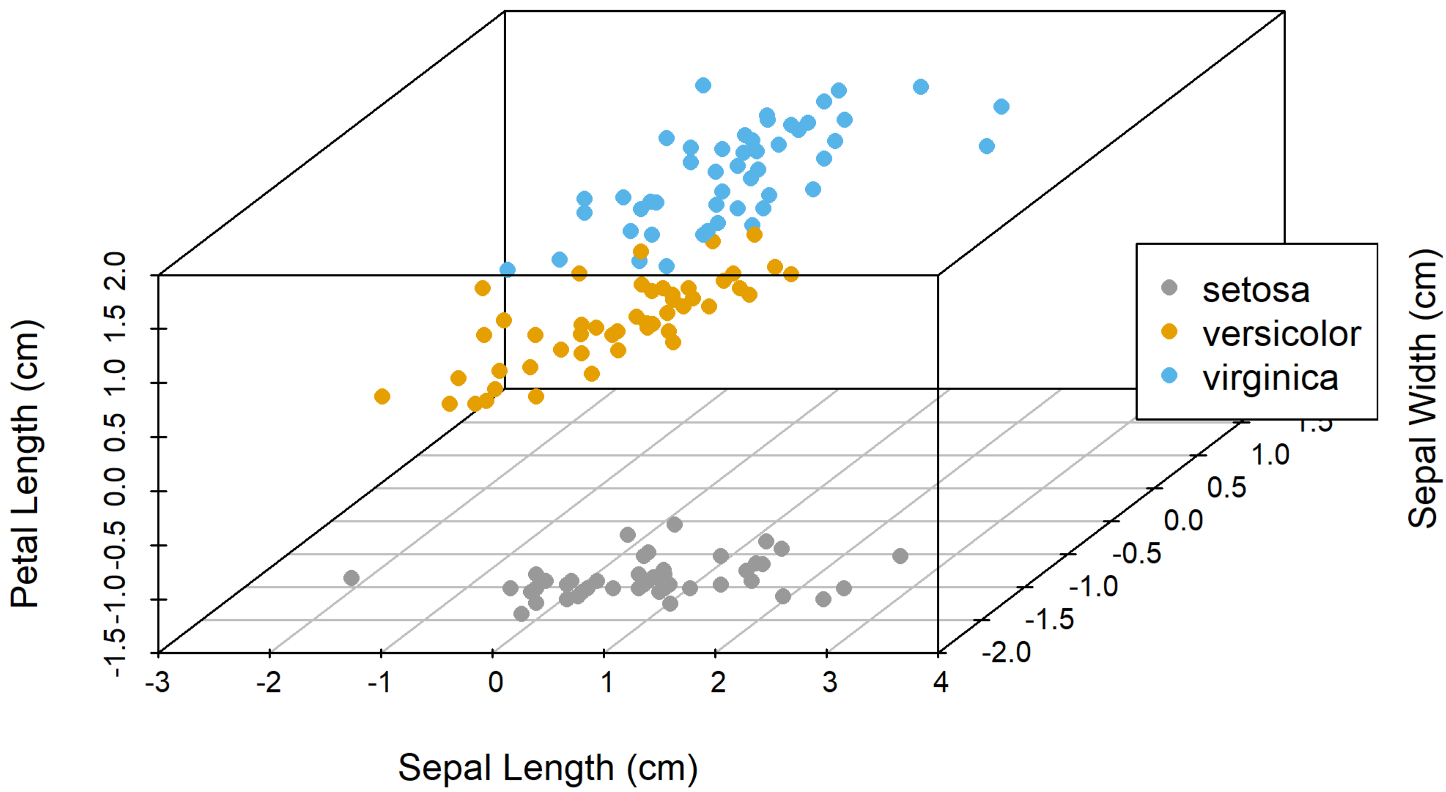
3.1 Question 1

```
colors <- c("#999999", "#E69F00", "#56B4E9")
colors <- colors[as.numeric(iris$Species)]

scatterplot3d(iris[, 2:4],
  xlab = "Sepal Length (cm)",
  ylab = "Sepal Width (cm)",
  zlab = "Petal Length (cm)",
  angle = 55,
  pch = 16,
  color = colors
)

legend("right",
  legend = levels(iris$Species),
  col = c("#999999", "#E69F00", "#56B4E9"), pch = 16
)
```



3.2 Question 2

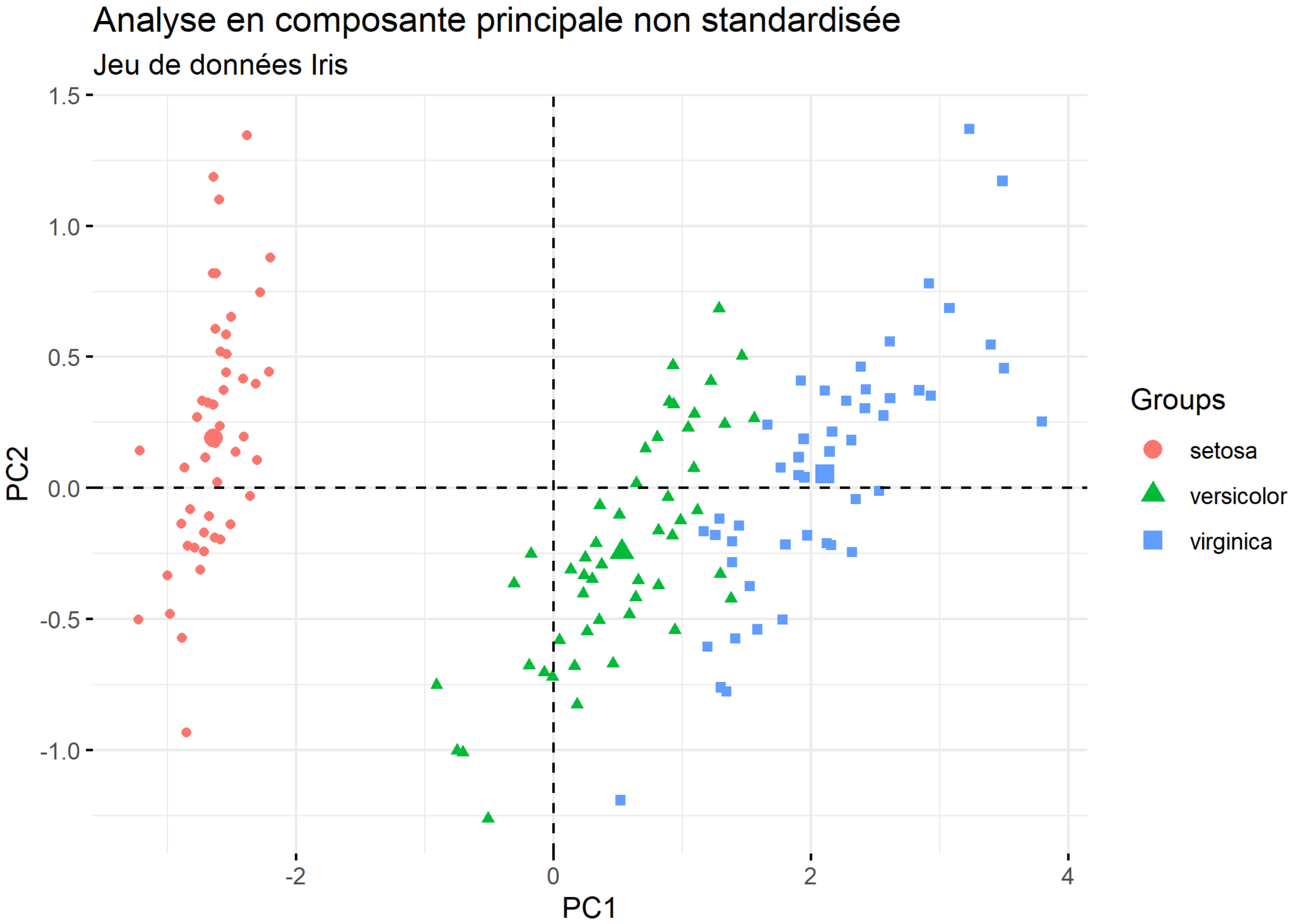
```
iris_scaled <- as_tibble(scale(iris[, 2:5], scale = TRUE, center = TRUE))

scatterplot3d(iris_scaled[, 2:4],
  xlab = "Sepal Length (cm)",
  ylab = "Sepal Width (cm)",
  zlab = "Petal Length (cm)",
  angle = 55,
  pch = 16,
  color = colors
)

legend("right",
  legend = levels(iris$Species),
  col = c("#999999", "#E69F00", "#56B4E9"), pch = 16
)
```



3.3 Question 3

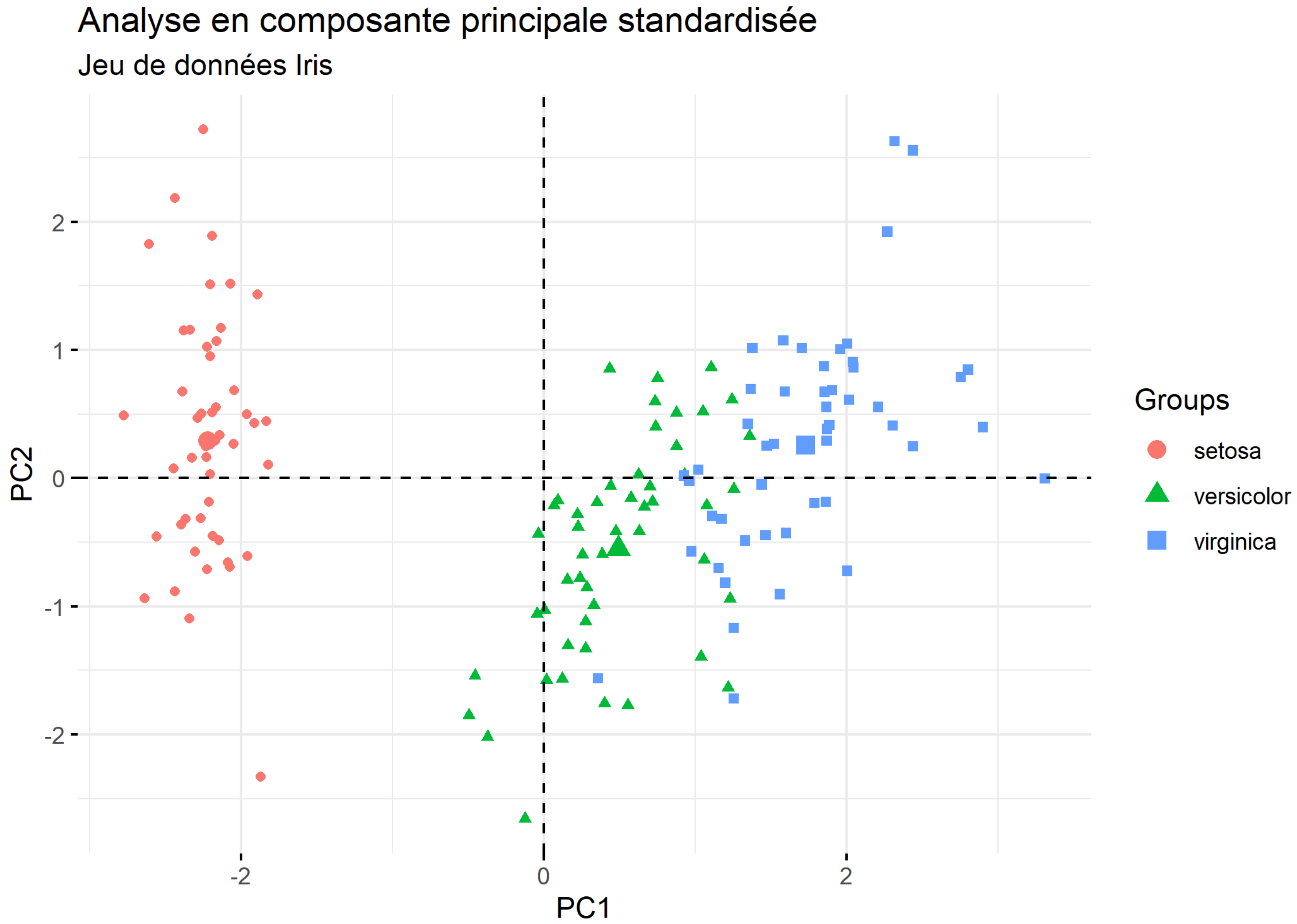
3.3.1 Données non standardisées

```r
iris_not_scaled <- PCA(iris[, 2:5], scale.unit = FALSE, graph = FALSE)

fviz_pca_ind(iris_not_scaled, label = "none", habillage = iris$Species) +
  labs(
    x = "PC1",
    y = "PC2",
    title = "Analyse en composante principale non standardisée",
    subtitle = "Jeu de données Iris"
  )
```



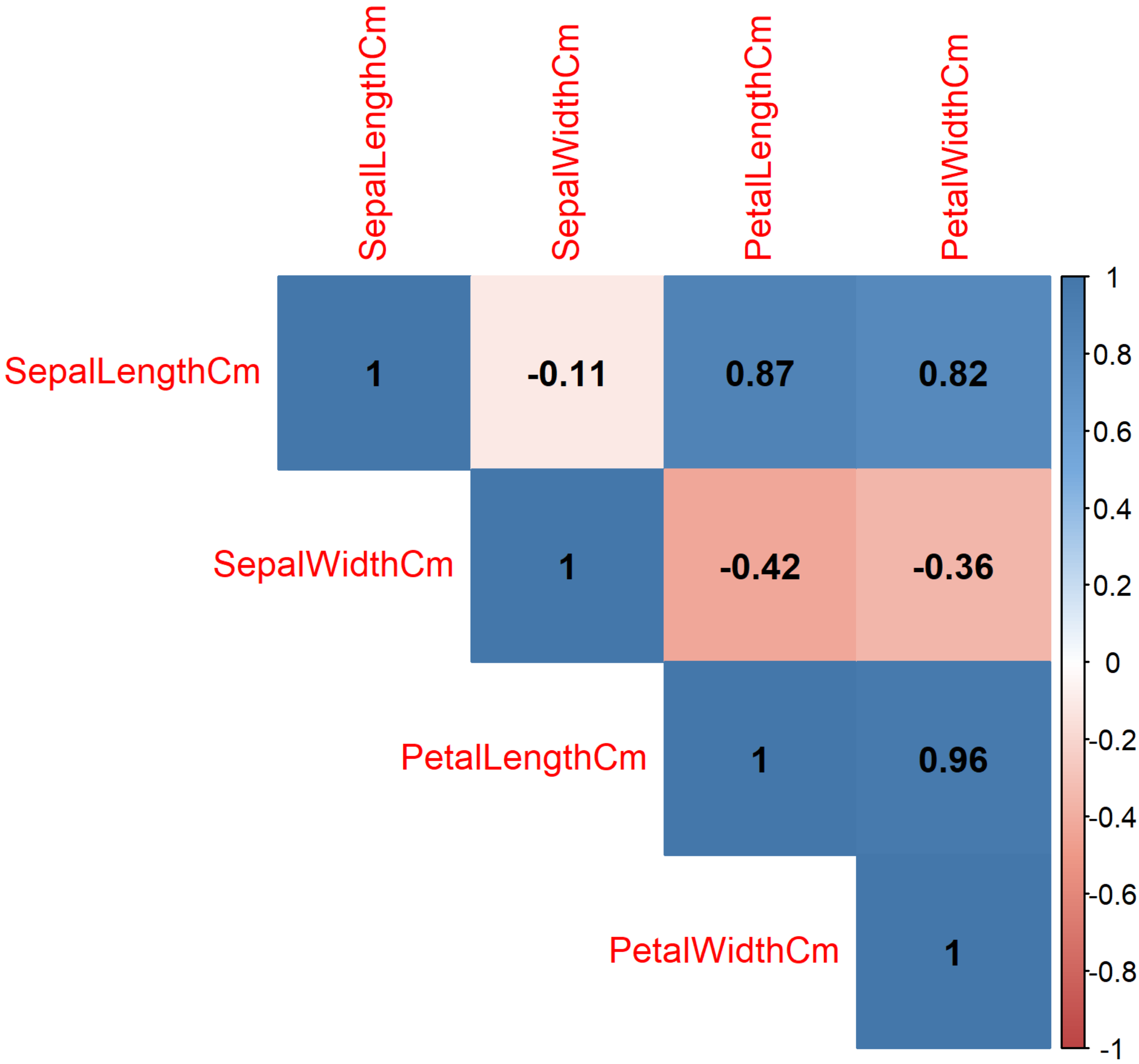
3.3.2 Données standardisées

```
pca_iris_scaled <- PCA(iris[, 2:5], scale.unit = TRUE, graph = FALSE)

fviz_pca_ind(pca_iris_scaled, label = "none", habillage = iris$Species) +
  labs(
    x = "PC1",
    y = "PC2",
    title = "Analyse en composante principale standardisée",
    subtitle = "Jeu de données Iris"
  )
```



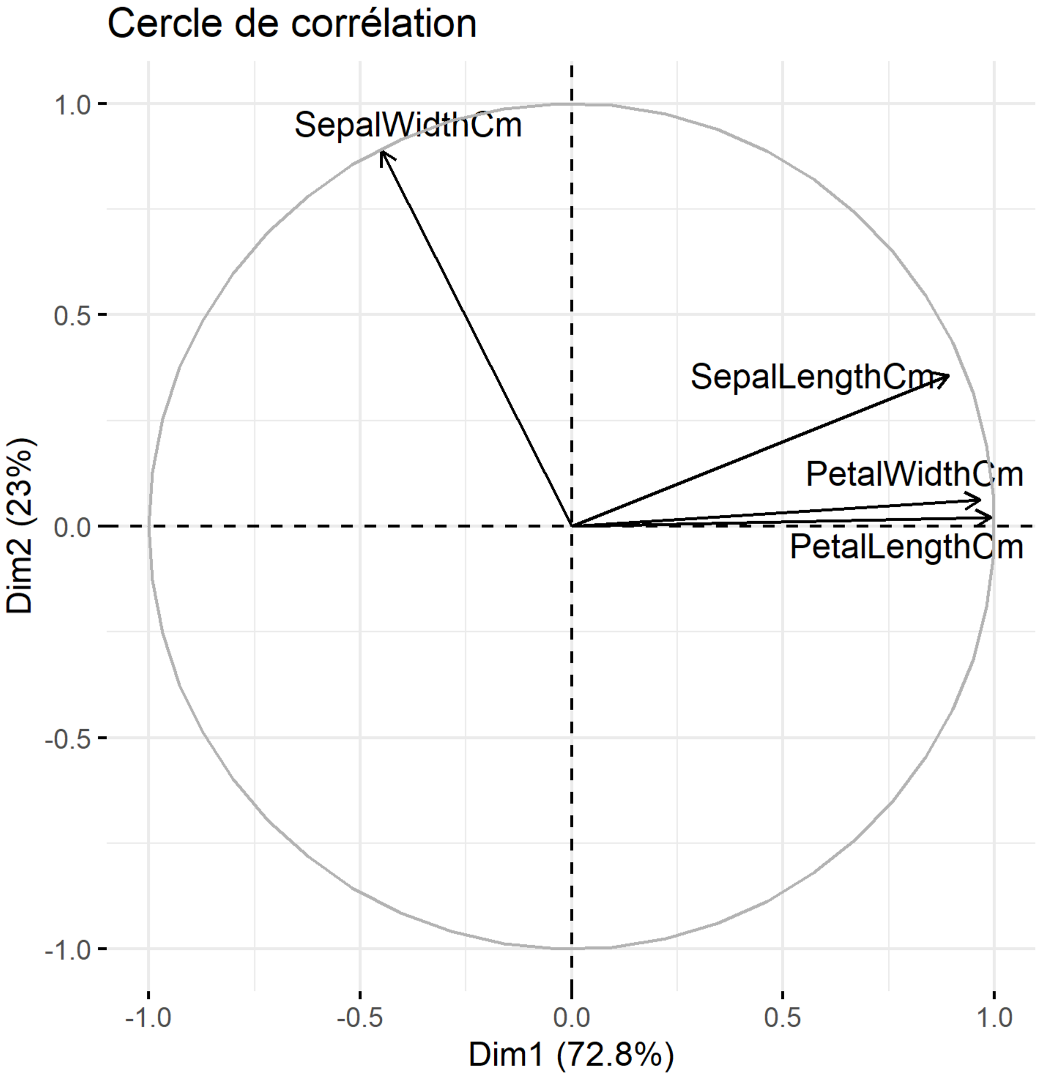
3.4 Question 4

```r
cor_iris_scaled <- cor(iris_scaled[, 1:4])

col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))

corrplot(cor_iris_scaled, method = "color", col = col(200),
         type = "upper",
         addCoef.col = "black")
```



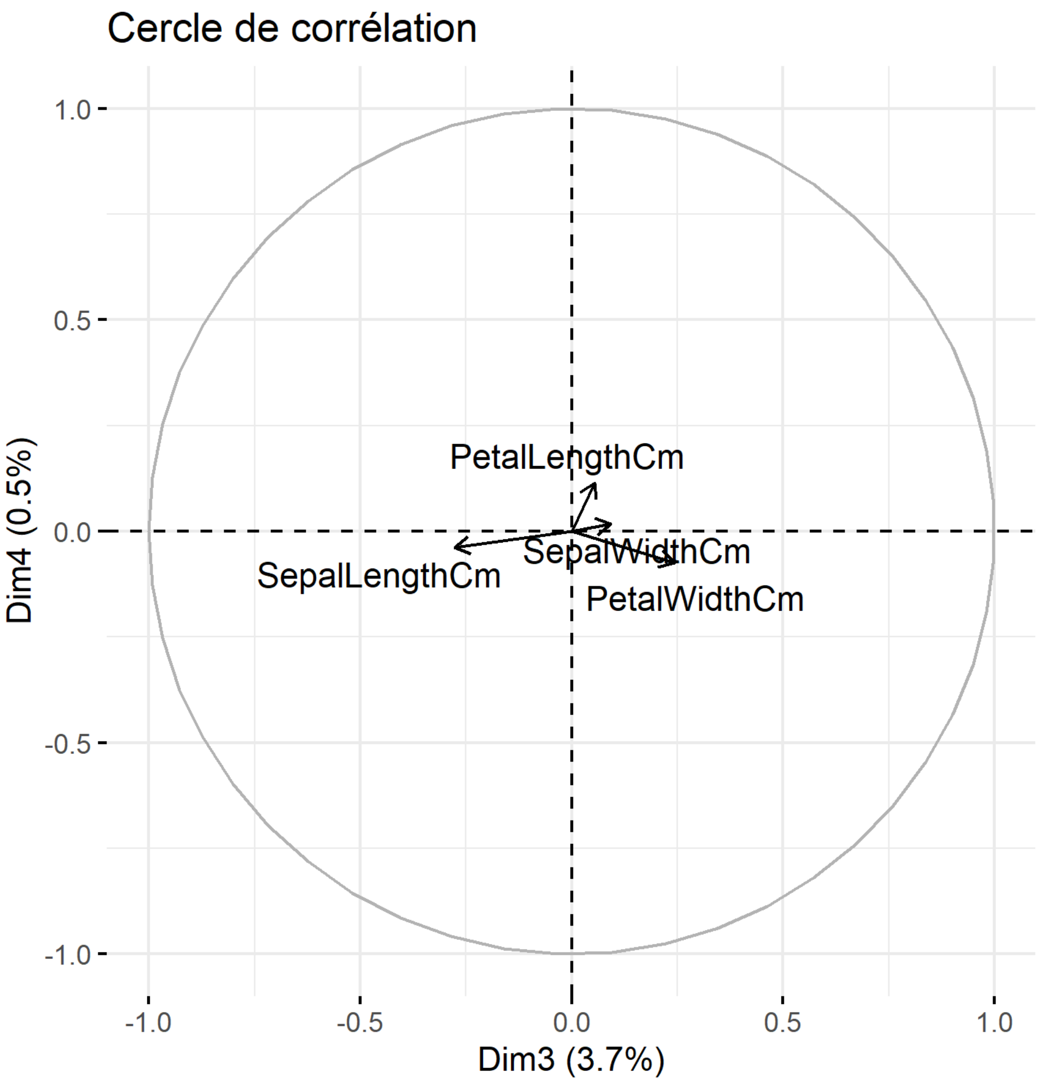
3.5 Question 5

```
fviz_pca_var(pca_iris_scaled, col.var = "black", repel = TRUE) +
  labs(title = "Cercle de corrélation")
```



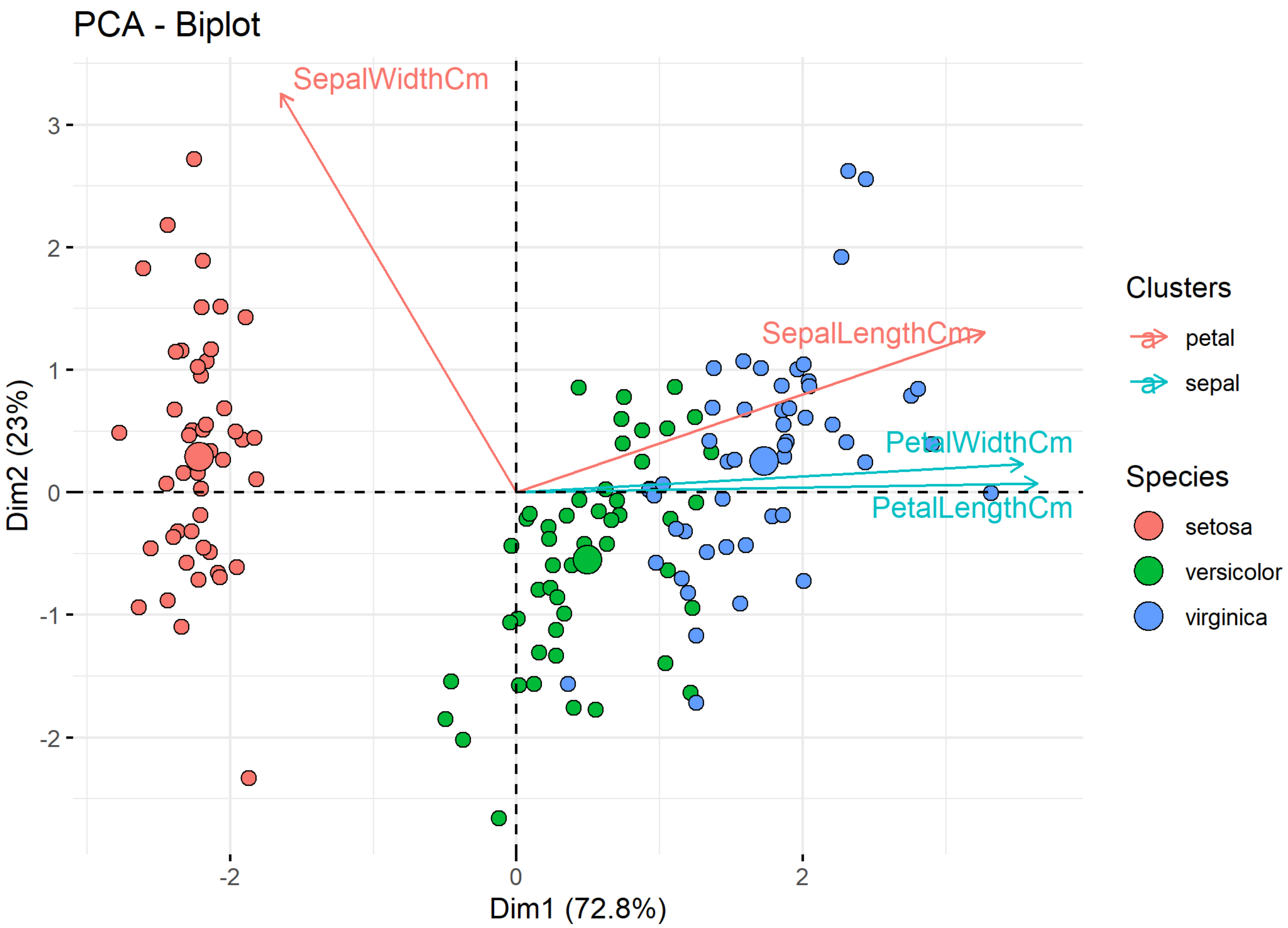
Les variables PetalWidthCm et PetalLengthCm avec un score de corrélation de 0.96 sont les deux variables les plus corrélées. Ce lien est montré dans le cercle de corrélation par là les deux flèches représentant les deux variables citées ci-dessus, celles-ci ont le même sens et même direction. De même, ces deux flèches possèdent une longueur quasi identique signifiant qu'elles contribuent au modèle de façon quasi identique.

```r
fviz_pca_var(pca_iris_scaled, col.var = "black", repel = TRUE, axes = c(3, 4)) +
  labs(title = "Cercle de corrélation")
```



3.6 Question 6

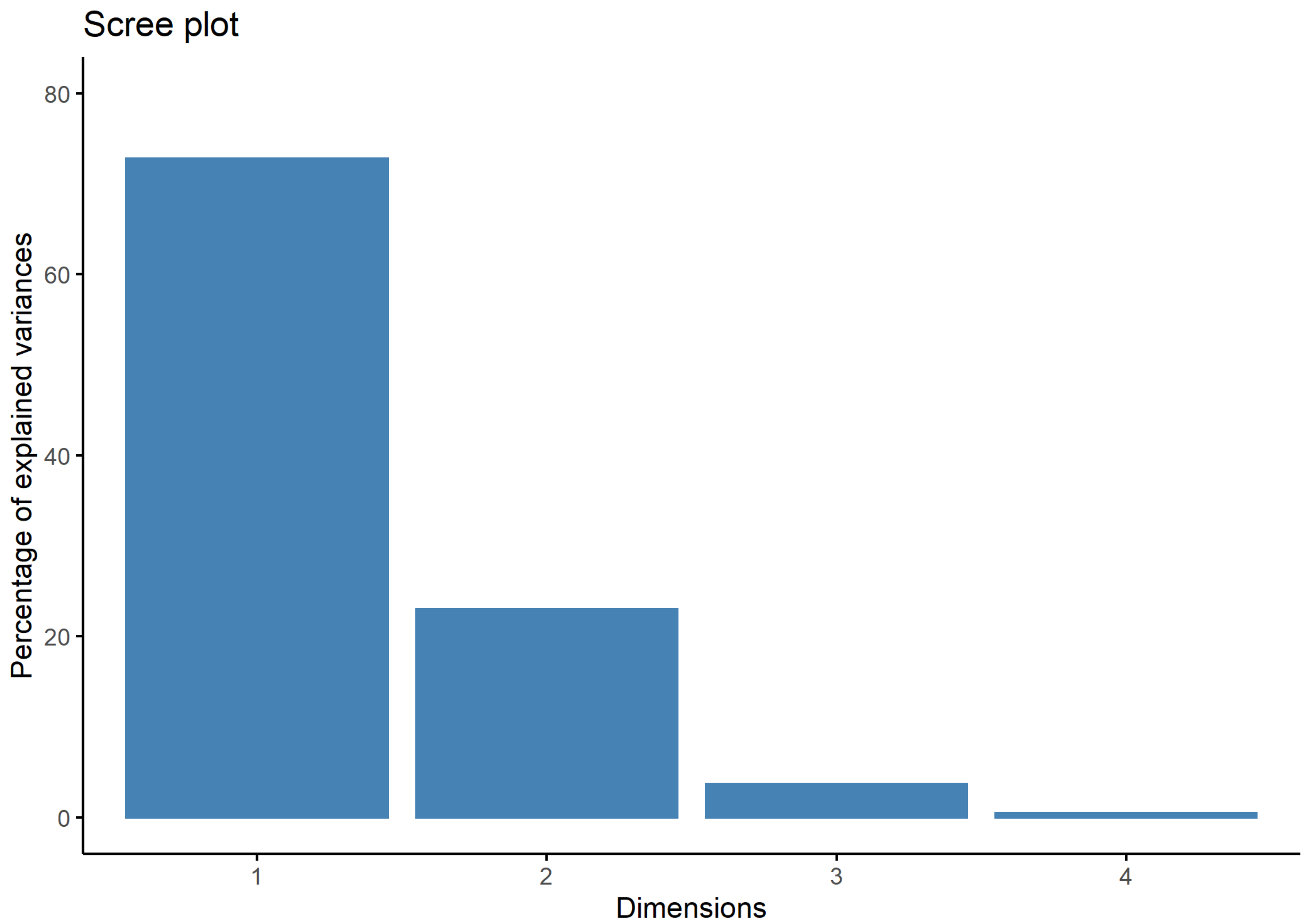
```
fviz_pca_biplot(pca_iris_scaled,
  geom.ind = "point",
  pointshape = 21
,
  pointsize = 2.5,
  fill.ind = iris$Species,
  col.ind = "black",
  col.var = factor(c("petal", "petal", "sepal", "sepal")),
  legend.title = list(fill = "Species", color = "Clusters"),
  repel = TRUE
)
```



La variance est représentée par les nombres donnés entre parenthèses des titres des axes. Ici, 72,8 % pour la dimension 1 et 23 % pour la dimension 2. Ces deux dimensions expliquent dont 95,8 % de la variance totale.

3.7 Question 7

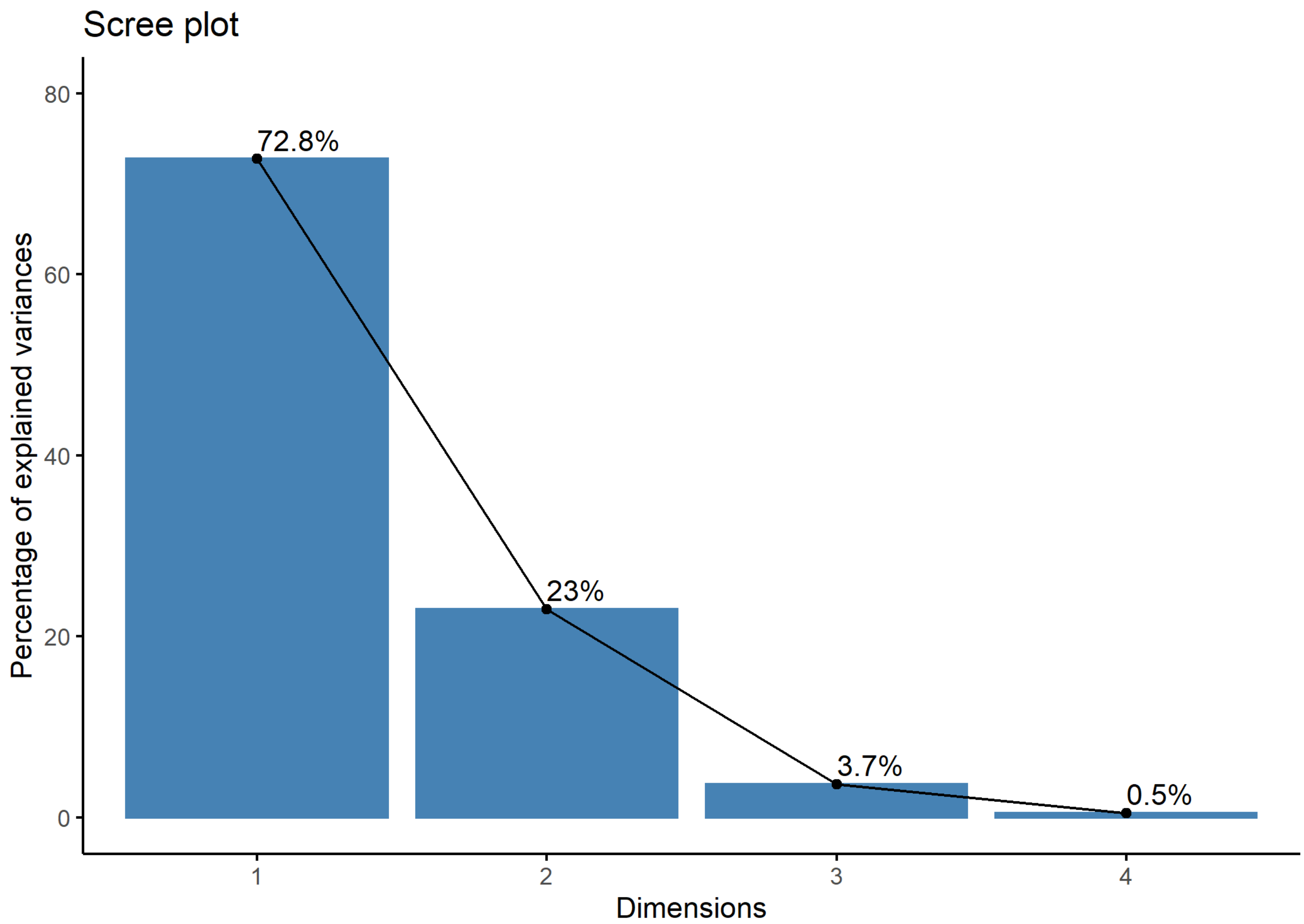
```
fviz_eig(pca_iris_scaled, geom = "bar", ggtheme = theme_classic(), ylim = c(0, 80)
)
```



Retenir deux dimensions semble un choix pertinent, car cela semble expliquer 95 % de l’ensemble de la variance total.

3.8 Question 8

```
fviz_eig(pca_iris_scaled, addlabels = TRUE, ggtheme = theme_classic(), ylim = c(0, 80)
)
```



3.9 Question 9

```
as_tibble(get_eig(pca_iris_scaled)) %>%
  add_column(Composant = 1:4) %>%
  select(Composant, 1:3) %>%
  rename(
    "% cumulé de la variance" = "cumulative.variance.percent",
    "% de la variance" = "variance.percent"
  )
```

Composant	eigenvalue	% de la variance	% cumulé de la variance
1	2.9108181	72.7704521	72.77045
2	0.9212209	23.0305233	95.80098
3	0.1473533	3.6838320	99.48481
4	0.0206077	0.5151927	100.00000

Quand le score de eigenvalue est grand, alors le score du pourcentage de la variance est grand lui aussi et inversement quand le score de eigenvalue est petit, alors le score du pourcentage de la variance est petit. Aussi, il existe un lien de proportion, ici d'environ 25.

3.10 Question 10

```
dimdesc(pca_iris_scaled)
```

```
## $Dim.1
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##               correlation       p.value
## PetalLengthCm   0.9916844 1.081470e-133
## PetalWidthCm    0.9649958  6.382632e-88
## SepalLengthCm   0.8912245  1.115929e-52
## SepalWidthCm   -0.4493130  8.048271e-09
##
## $Dim.2
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##               correlation      p.value
## SepalWidthCm    0.8883515 6.886044e-52
## SepalLengthCm   0.3573521 7.149886e-06
##
## $Dim.3
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##               correlation      p.value
## PetalWidthCm     0.243295 0.0026995537
## SepalLengthCm   -0.276774 0.0006065947
```

Pour la CP1, la variable PetalLengthCm est celle qui présente la saturation la plus forte (0.99). La table de saturation nous permet d'imaginer le cercle de corrélation sans le faire, car chaque valeur du tableau correspond au sens et à la direction de sa flèche de corrélation et de quel côté de l'axe de sa dimension, il se trouve. Seul les variables ayant une p value statistiquement significatives sont gardées.

3.11 Question 11

```
tibble(
  Composant = 1:4,
  Variable = c("PetalLengthCm", "SepalWidthCm", "SepalLengthCm", "PetalLengthCm")
)
```

Composant	Variable
1	PetalLengthCm
2	SepalWidthCm
3	SepalLengthCm
4	PetalLengthCm

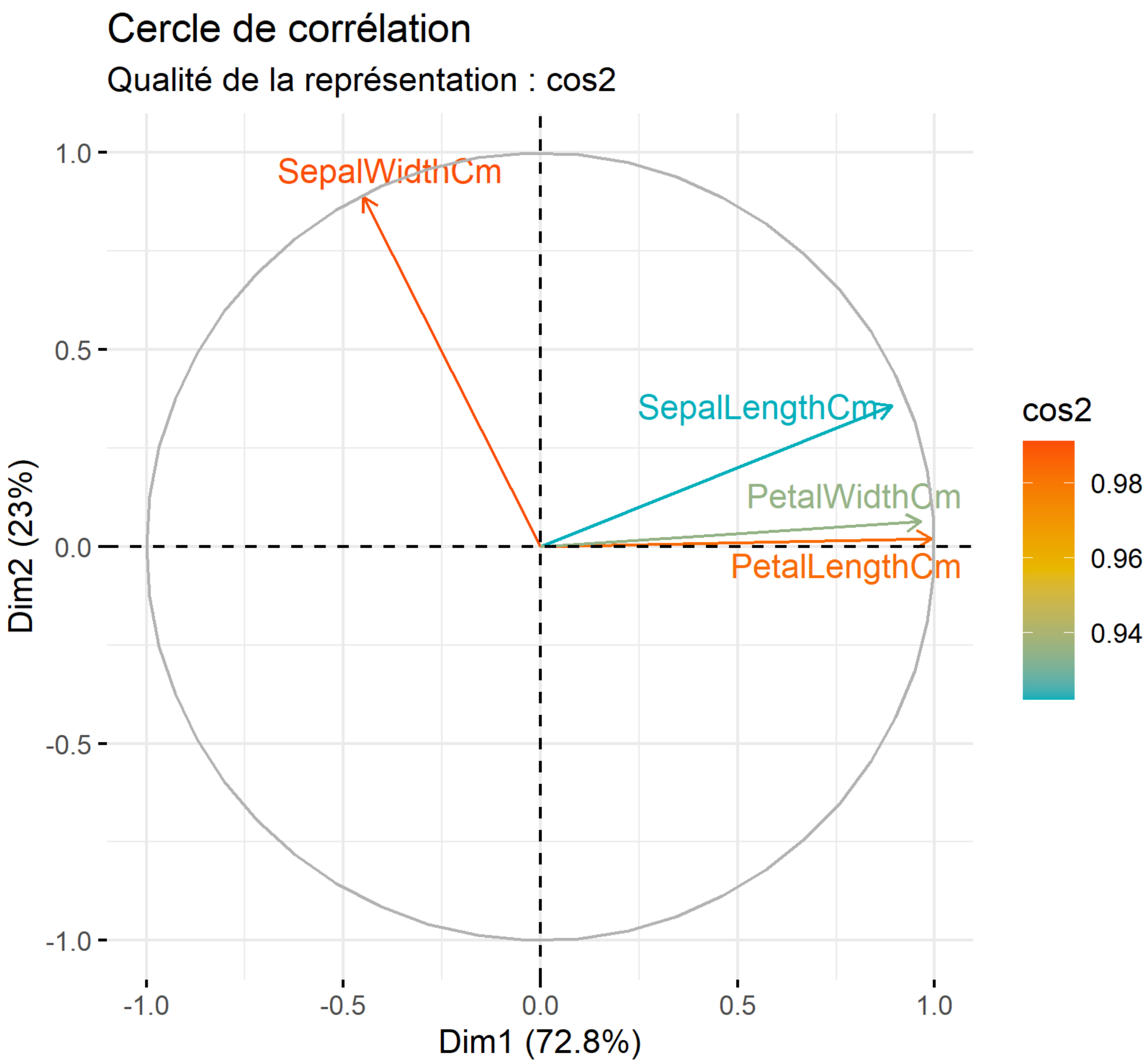
3.12 Question 12

Toutes les variables à droite de Dim1 sont positives et celles à gauche sont négatives, et plus elles tendent vers 1, plus elles se trouvent perpendiculaires à l'axe Dim1.

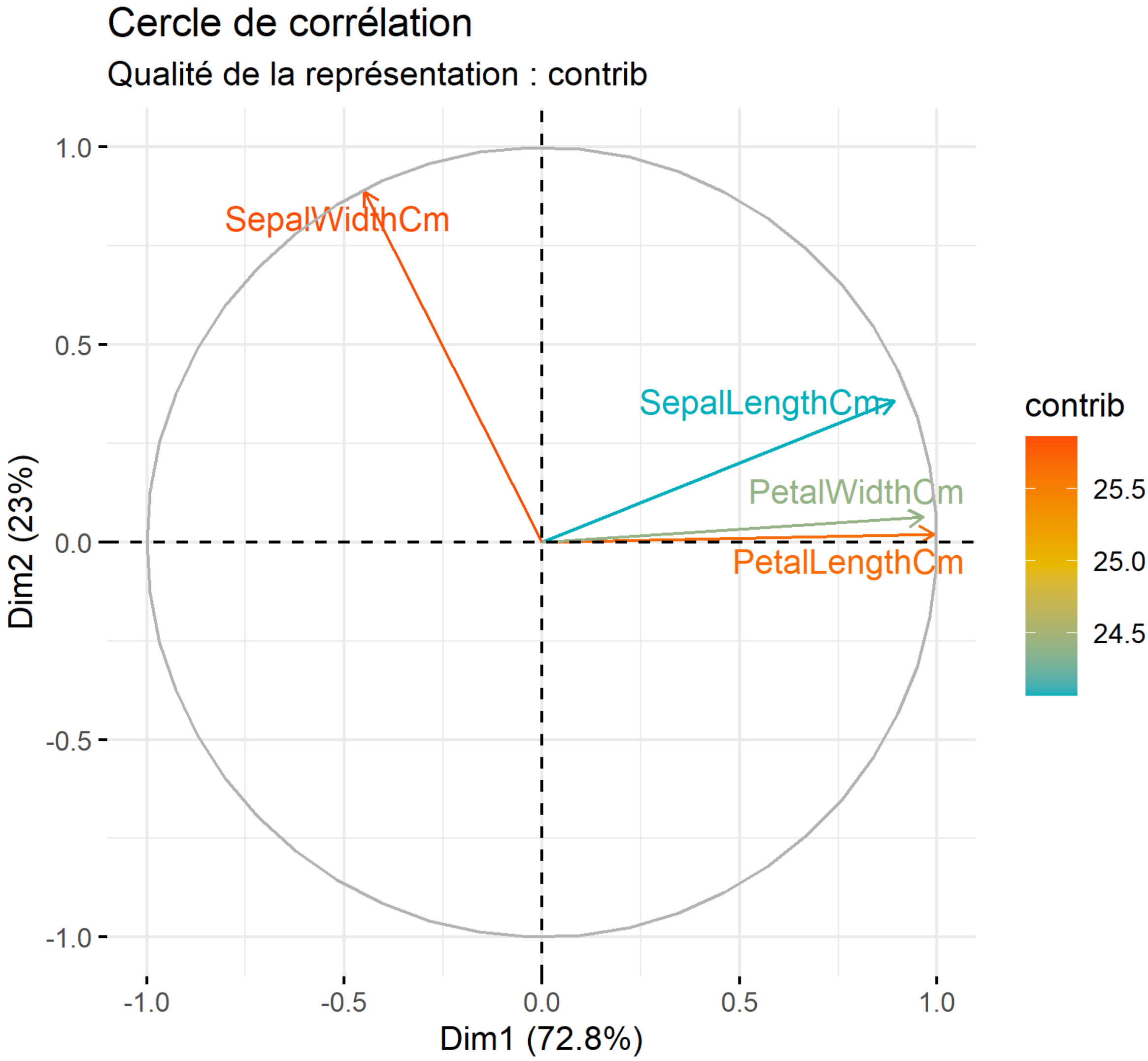
De même, toutes les variables en haut de l'axe Dim2 sont positives et celles en bas sont négatives.

3.13 Question 13

```r
fviz_pca_var(pca_iris_scaled,
  col.var = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE
) +
  labs(
    title = "Cercle de corrélation",
    subtitle = "Qualité de la représentation : cos2"
  )
```




```r
fviz_pca_var(pca_iris_scaled,
  col.var = "contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE
) +
  labs(
    title = "Cercle de corrélation",
    subtitle = "Qualité de la représentation : contrib"
  )
```



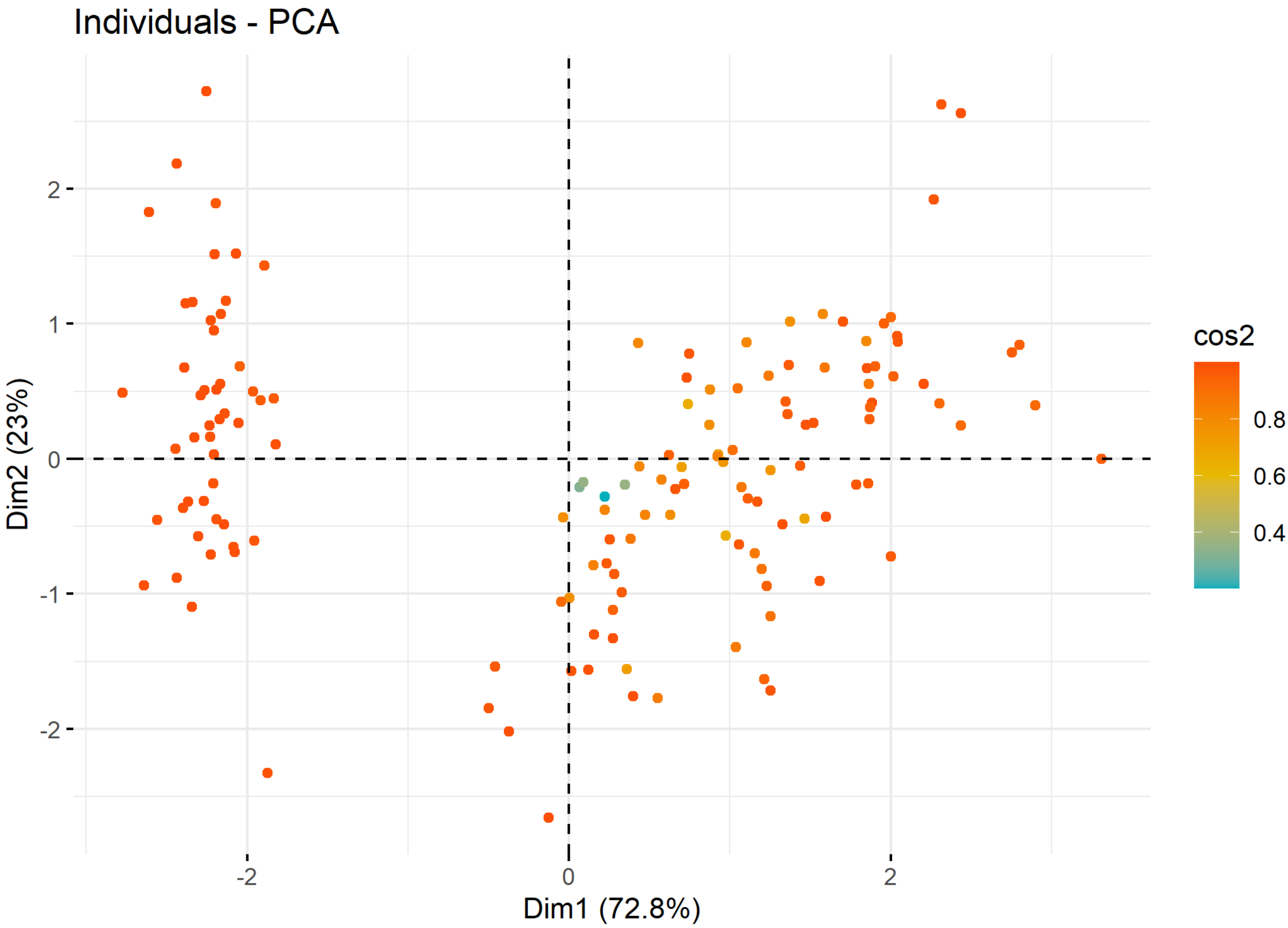
La corrélation de chaque point sur un axe exprime la qualité de représentation du point sur l'axe. Elle prend des valeurs entre 0 (pas corrélé du tout) et 1 (fortement corrélées). Si cette valeur est proche de 1, alors le point est bien représenté sur l'axe. Les points situés près du centre sont donc généralement mal représentés par le plan factoriel. Leur interprétation ne peut donc pas être effectuée avec confiance

cos2 : représente la qualité de la représentation des variables sur la carte factorielle.

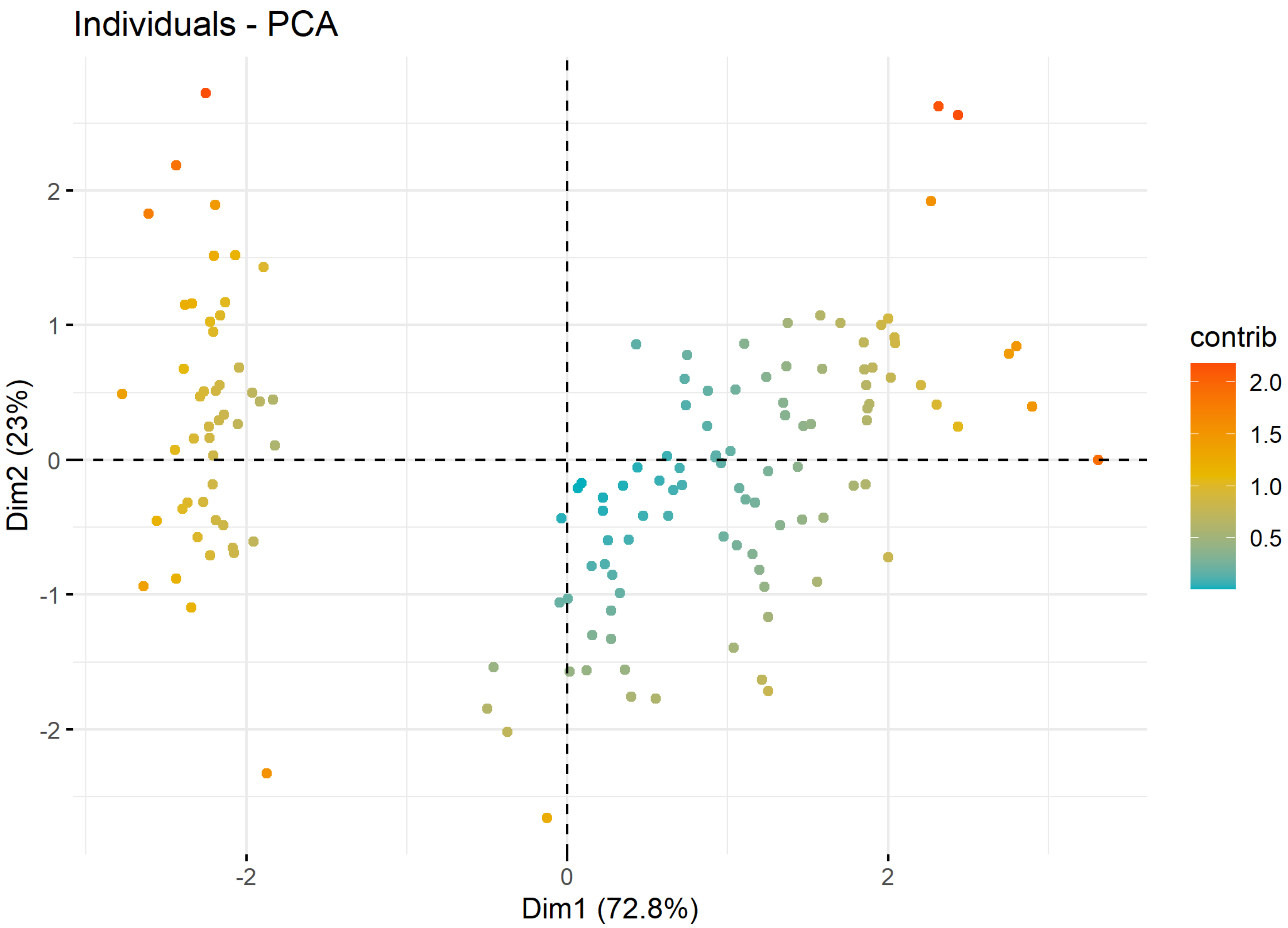
contrib : contient les contributions (en pourcentage) des variables aux composantes principales.

3.14 Question 14

```
fviz_pca_ind(pca_iris_scaled,
  col.ind = "cos2", label = "none",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE
)
```




```r
fviz_pca_ind(pca_iris_scaled,
  col.ind = "contrib", label = "none",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE
)
```



3.15 Question 15

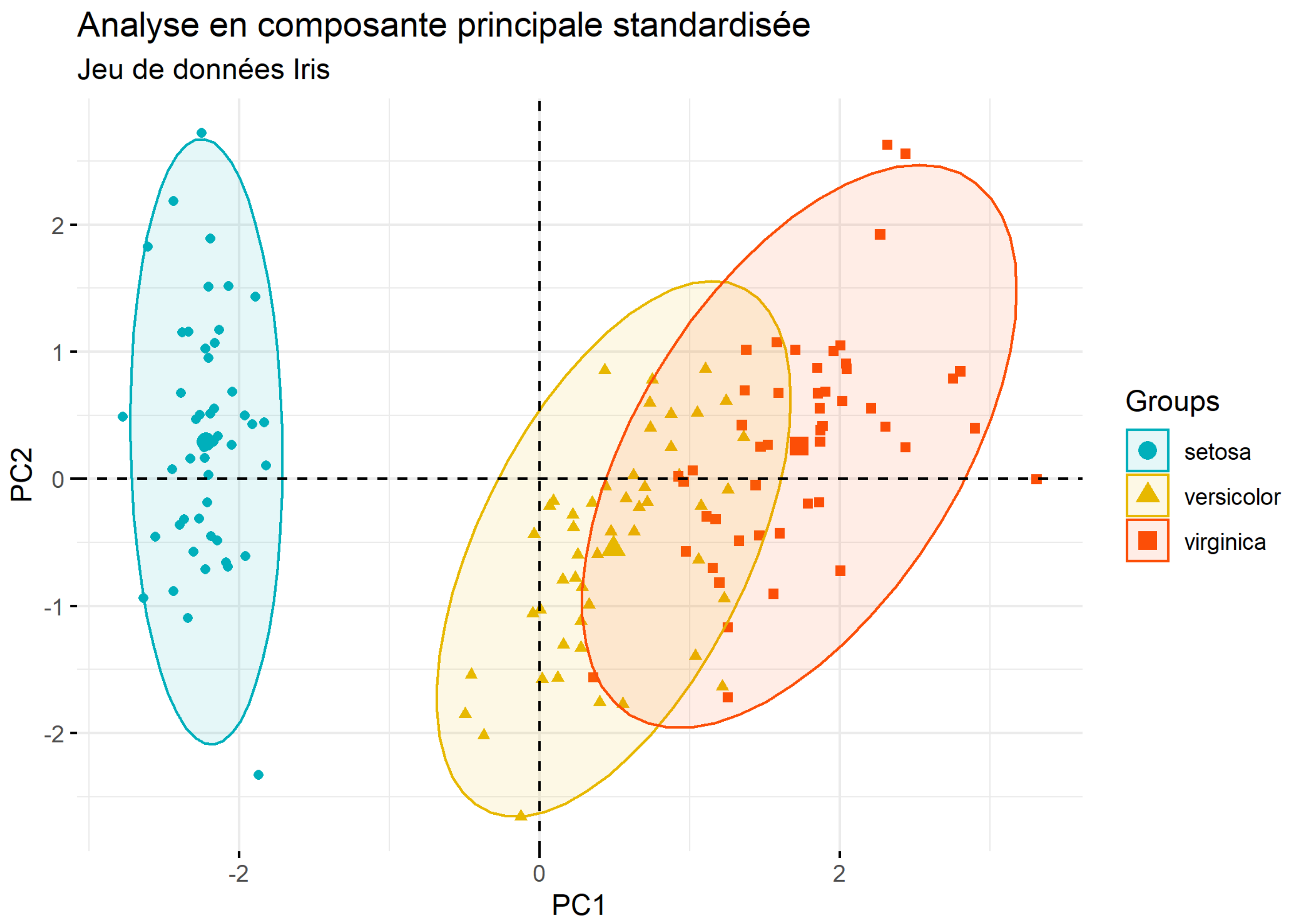
```
as_tibble(pca_iris_scaled$ind$contrib) %>%
  add_column(ID = 1:150) %>%
  select(ID, 1:2) %>%
  mutate(across(where(is.numeric), round, 2)) %>%
  head(10)
```

ID	Dim.1	Dim.2
1	1.17	0.19
2	1.00	0.31
3	1.28	0.07
4	1.22	0.24
5	1.31	0.33
6	0.98	1.67
7	1.37	0.00
8	1.14	0.04
9	1.26	0.87
10	1.10	0.15

Plus la valeur de la contribution est importante, plus l'observation contribue à la dimension en question.

3.16 Question 16

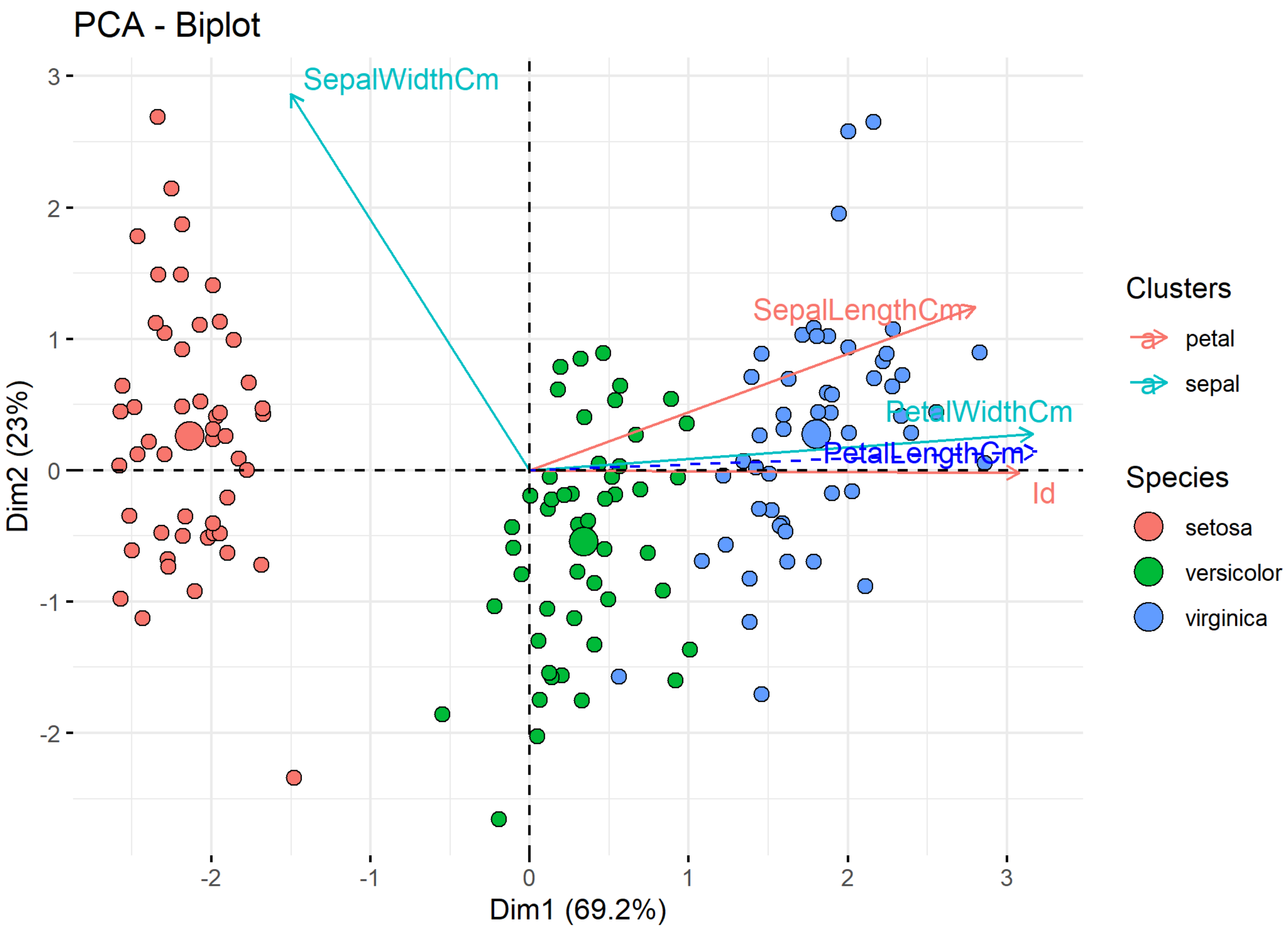
```
fviz_pca_ind(pca_iris_scaled,
  geom.ind = "point", col.ind = iris$Species,
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),
  addEllipses = TRUE
,
  legend.title = "Groups"
) +
  labs(
    x = "PC1",
    y = "PC2",
    title = "Analyse en composante principale standardisée",
    subtitle = "Jeu de données Iris"
  )
```



Ces ellipses se distinguent d'un algorithme de classification par plusieurs éléments. Premiers, tous les points ne sont pas inclus dans une ellipse. Second, les ellipses se chevauchent énormément. Trois, elles ne possèdent pas de centroïdes tel qu'ils sont dans un algorithme kmeans.

3.17 Question 17

```r
pca_iris_var_sup <- PCA(iris, quali.sup = c(6), quanti.sup = c(4), graph = FALSE)

fviz_pca_biplot(pca_iris_var_sup,
  geom.ind = "point",
  pointshape = 21
,
  pointsize = 2.5,
  fill.ind = iris$Species,
  col.ind = "black",
  col.var = factor(c("petal", "petal", "sepal", "sepal")),
  legend.title = list(fill = "Species", color = "Clusters"),
  repel = TRUE
)
```



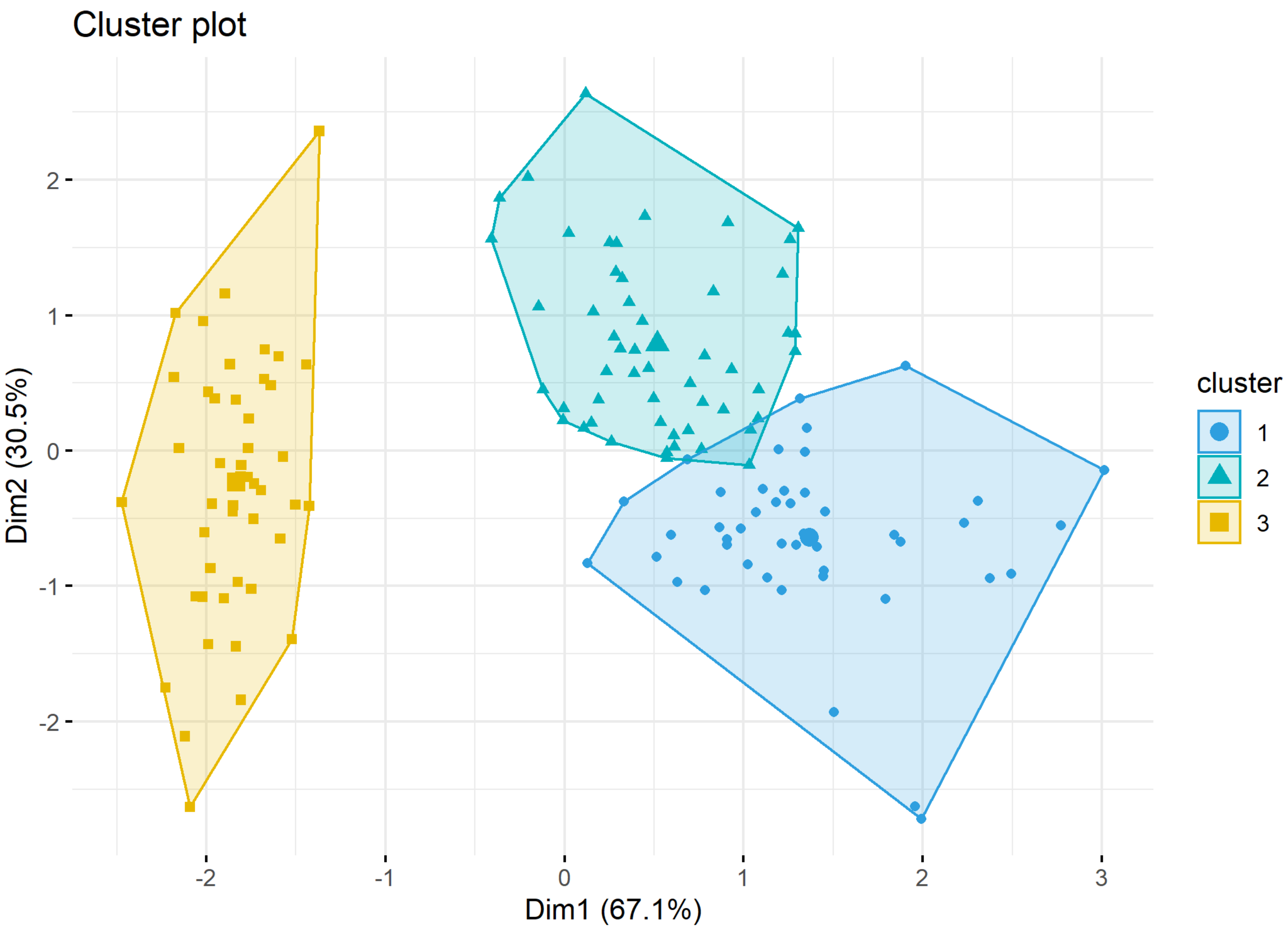
3.18 Question 18

Un algorithme k-means fonctionne comme suit :

- Première itération :
  - Disposition aléatoire des centroïdes dont le nombre est défini à l'avance (k).
  - k cluster sont générés en associant chaque observation à la moyenne la plus proche.
  - Le centroïde de chaque nouveau cluster devient la nouvelle moyenne.
- Seconde itération :
  - De nouveaux clusters sont calculés en fonction des nouveaux centroïdes.
  - Le centroïde de chaque nouveau cluster devient la nouvelle moyenne.
- Répétition jusqu'à stabilisation des centroïdes, c'est-à-dire qu'aucun nouveau centroïdes apparait.

3.19 Question 19

```r
kmean_iris_scaled <- kmeans(pca_iris_scaled$ind$coord, 3, nstart = 50)

fviz_cluster(kmean_iris_scaled,
  data = iris[, 2:4],
  palette = c("#2E9FDF", "#00AFBB", "#E7B800"),
  geom = "point",
  ellipse.type = "convex",
  show.clust.cent = TRUE,
  ggtheme = theme_minimal(
)
)
```




```r
as_tibble(kmean_iris_scaled$centers[, 1:2]) %>%
  add_column(Cluster = 1:3) %>%
  select(Cluster, 1:3)
```

Cluster	Dim.1	Dim.2
1	1.722369	0.5999051
2	0.567138	-0.8076751
3	-2.220193	0.2922248

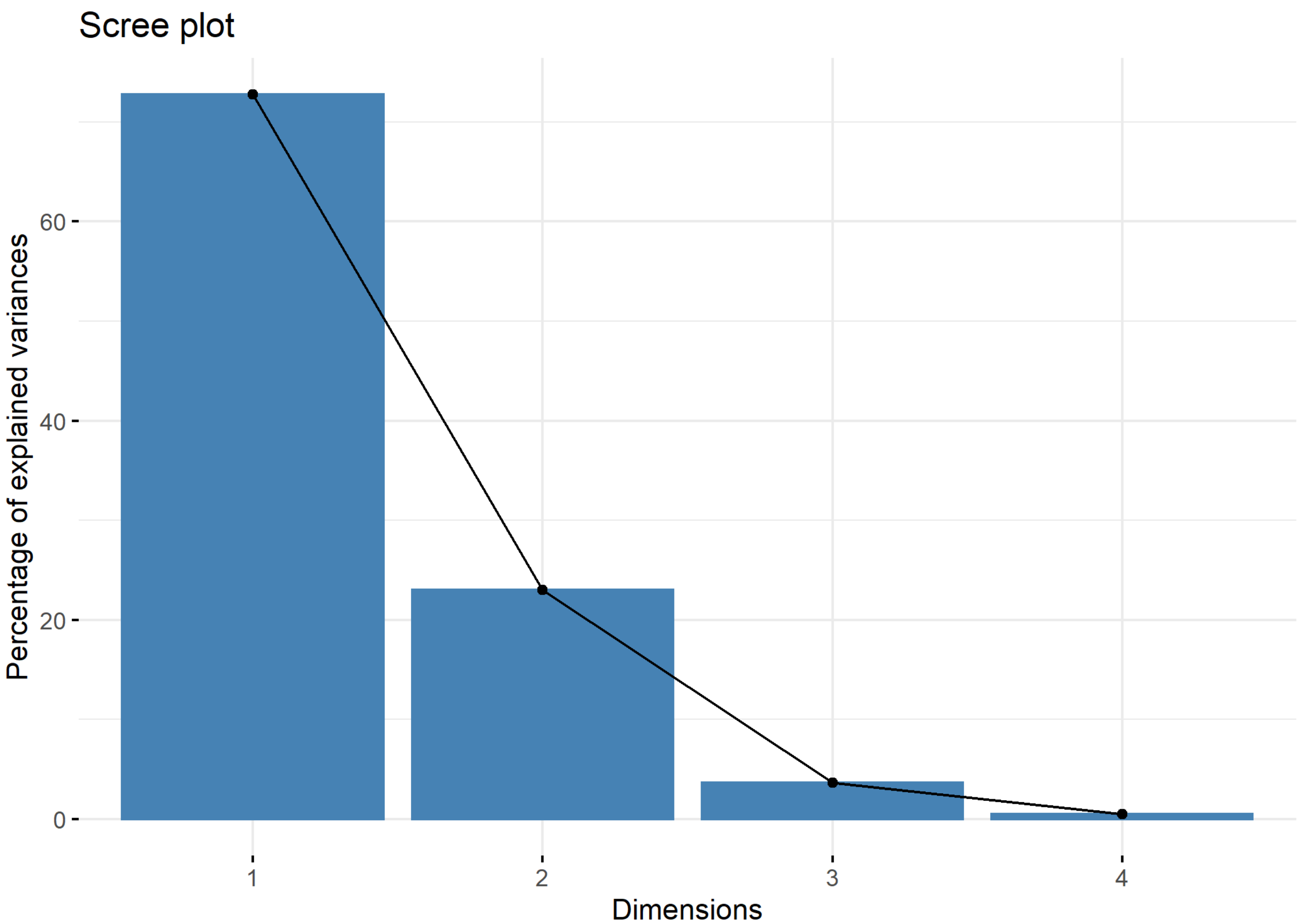
```r
as_tibble(kmean_iris_scaled$size) %>%
  add_column(Cluster = 1:3) %>%
  rename("Nb d'éléments" = value) %>%
  select(Cluster, "Nb d'éléments")
```

Cluster	Nb d'éléments
1	47
2	53
3	50

3.20 Question 20

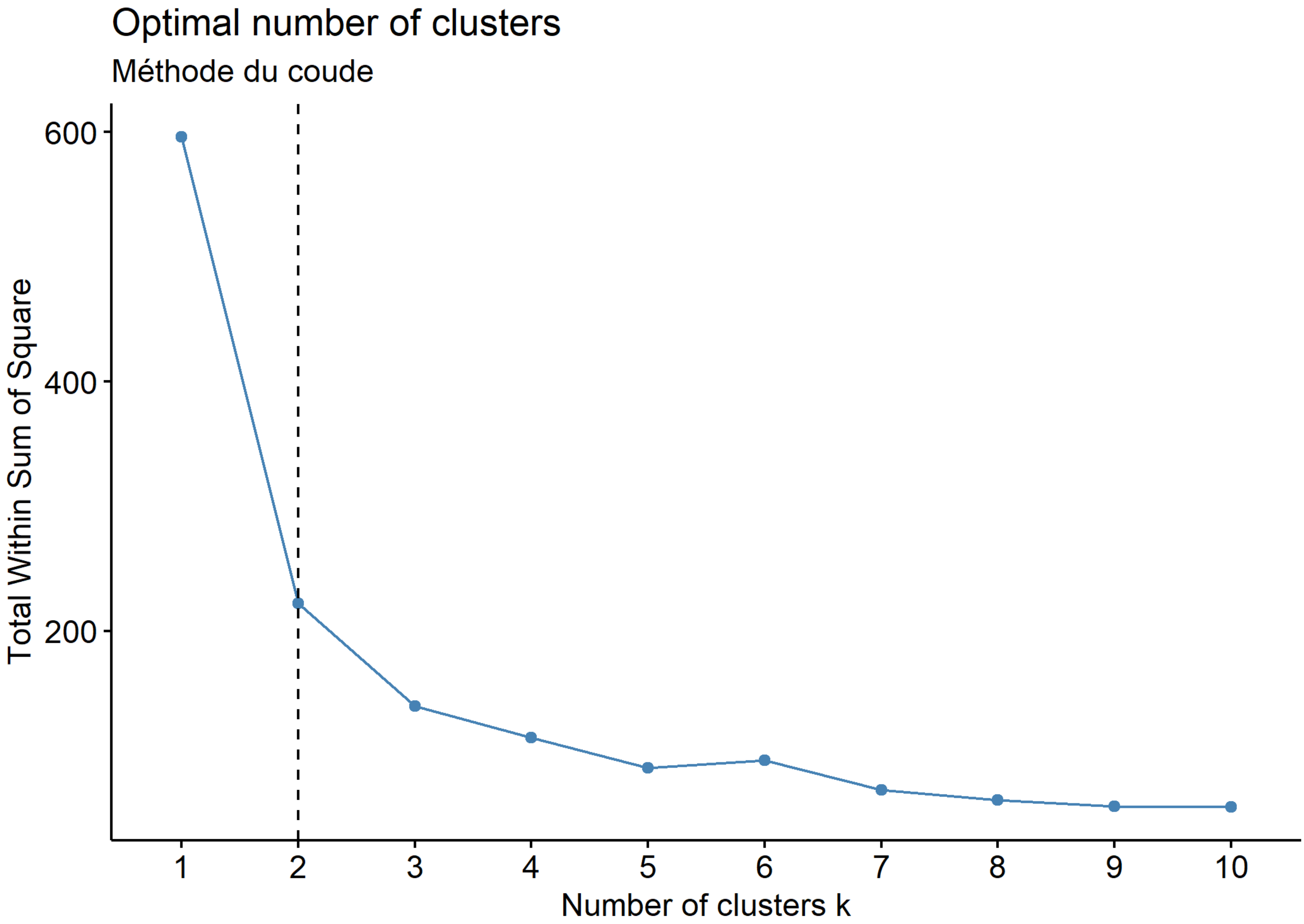
3.20.1 Scree plot de la variance intracluster

```r
fviz_screeplot(pca_iris_scaled, choice = c("variance"))
```



3.20.2 Méthode du coude

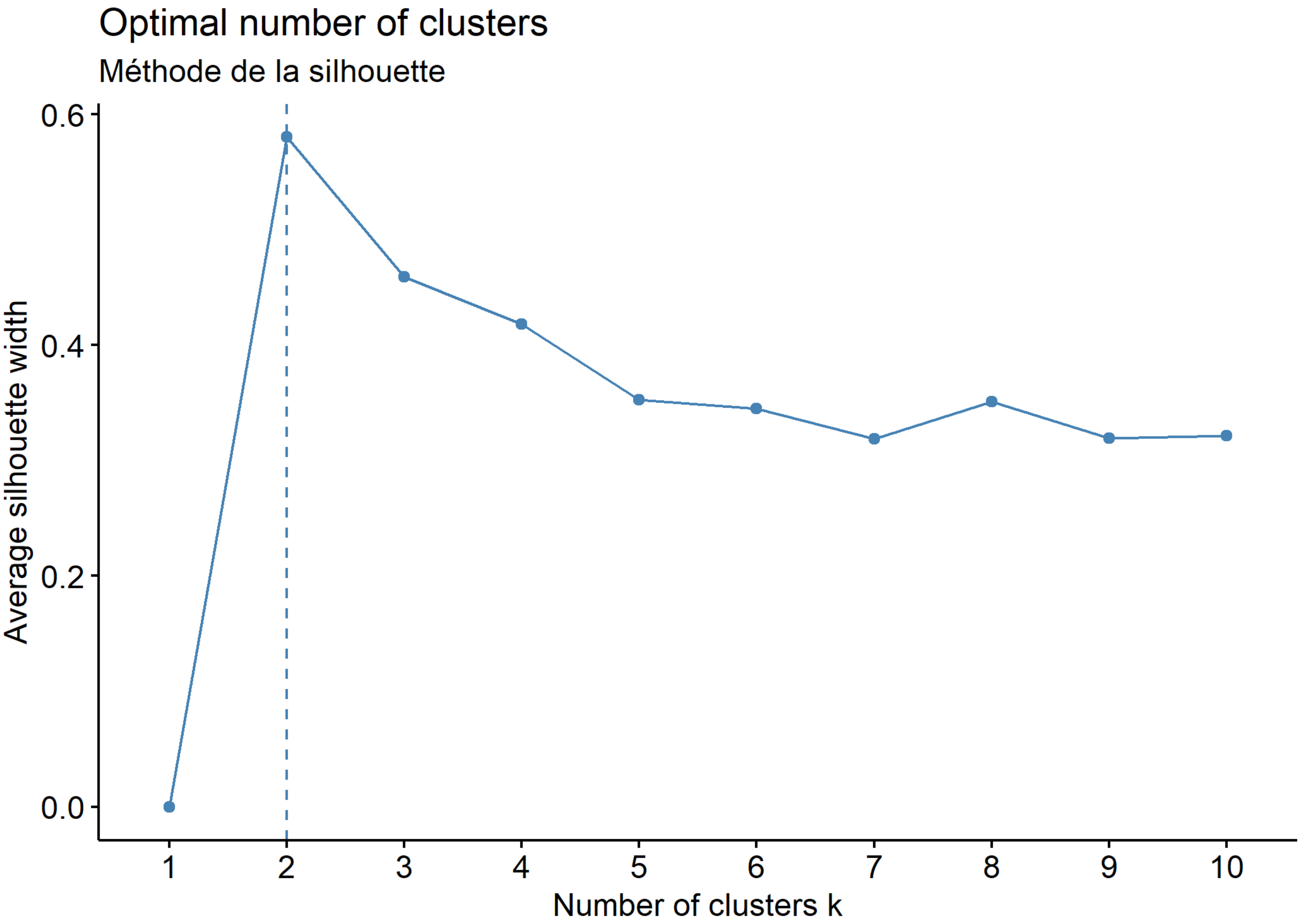
```r
fviz_nbclust(iris_scaled, kmeans, method = "wss") +
  geom_vline(xintercept = 2, linetype = 2) +
  labs(subtitle = "Méthode du coude")
```



Dans la méthode du coude, nous faisons varier le nombre de clusters (K) de 1 à K. Pour chaque valeur de K, nous calculons le WCSS (Within-Cluster Sum of Square). Le WCSS est la somme des carrés de la distance entre chaque point et le centroïde d'un cluster. Lorsque nous traçons le WCSS avec la valeur K, le tracé ressemble à un coude. Lorsque le nombre de clusters augmente, la valeur WCSS commence à diminuer. La valeur WCSS est la plus grande lorsque K = 1.

3.20.3 Méthode de la silhouette

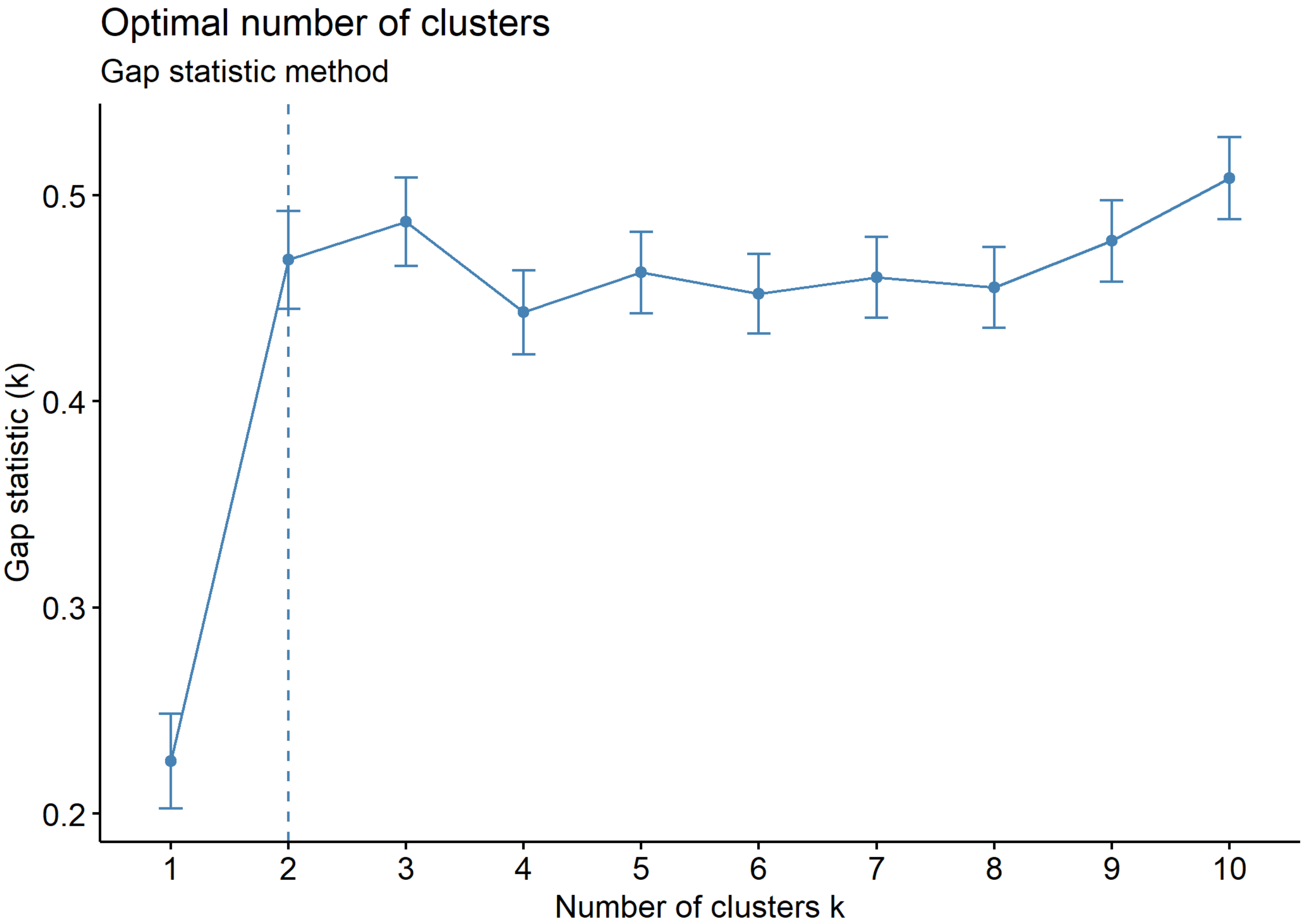
```
fviz_nbclust(iris_scaled, kmeans, method = "silhouette") +
  labs(subtitle = "Méthode de la silhouette")
```



La méthode de la silhouette calcule les coefficients de silhouette de chaque point qui mesurent à quel point un point est similaire à son propre cluster par rapport aux autres clusters. en fournissant une représentation graphique succincte de la façon dont chaque objet a été classé.

3.20.4 Méthode de la statistique de l'écart

```
fviz_nbclust(iris_scaled, kmeans,
  nstart = 25,
  method = "gap_stat",
  nboot = 500
) +
  labs(subtitle = "Gap statistic method")
```

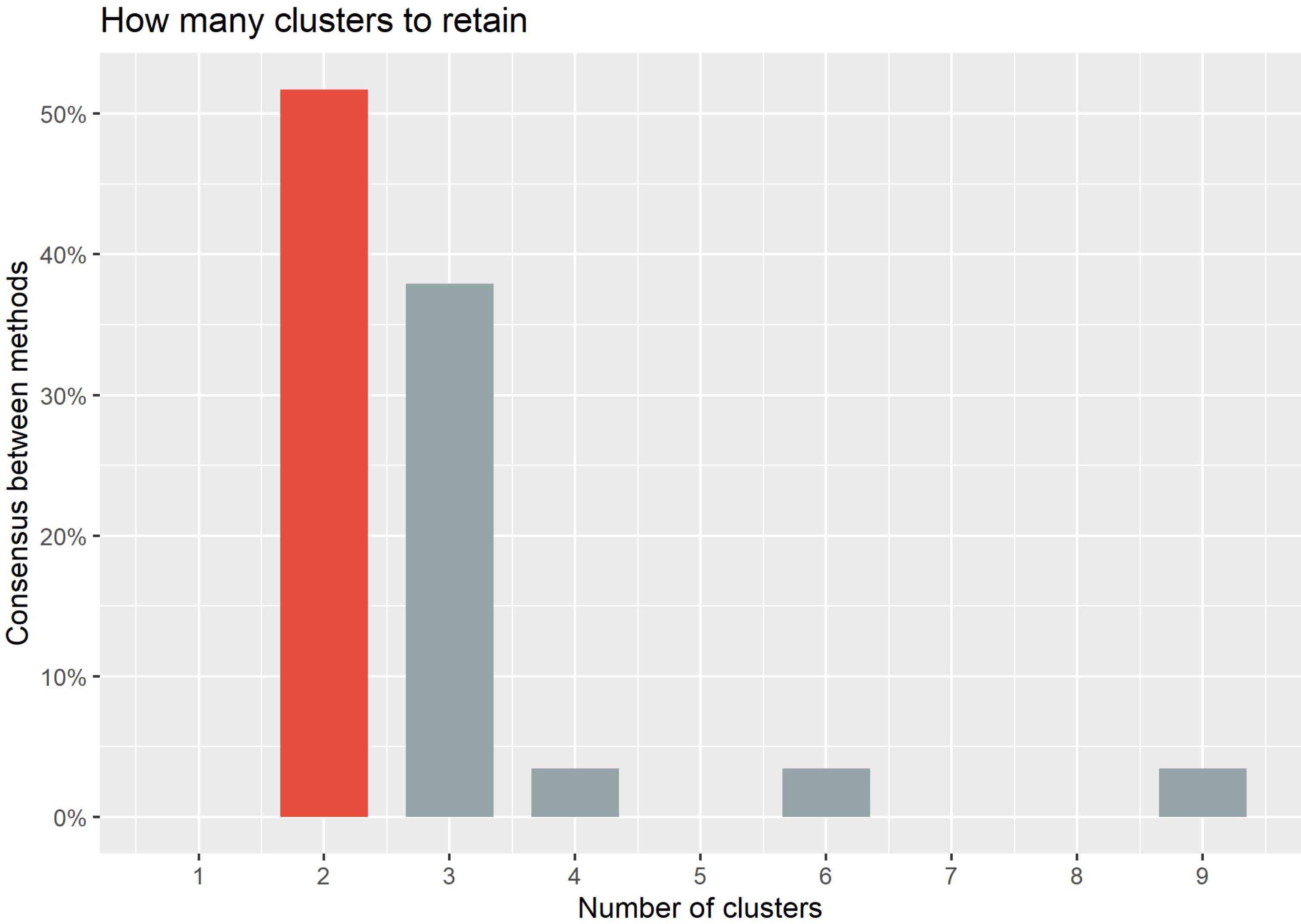


La statistique d'écart mesure à quel point la variation totale au sein d'un cluster peut être différente entre les données observées et les données de référence avec une distribution uniforme aléatoire. Une statistique d'écart importante signifie que la structure de clustering est très éloignée de la distribution uniforme aléatoire des points.

Le nombre de clusters peut être choisi comme la plus petite valeur de k telle que la statistique d'écart soit dans un écart-type de l'écart à k+1.

3.20.5 Algorithme basé sur le consensus

```
n_clust <- n_clusters(iris_scaled,
  package = c("easystats", "NbClust", "mclust"),
  standardize = FALSE
)

plot(n_clust)
```



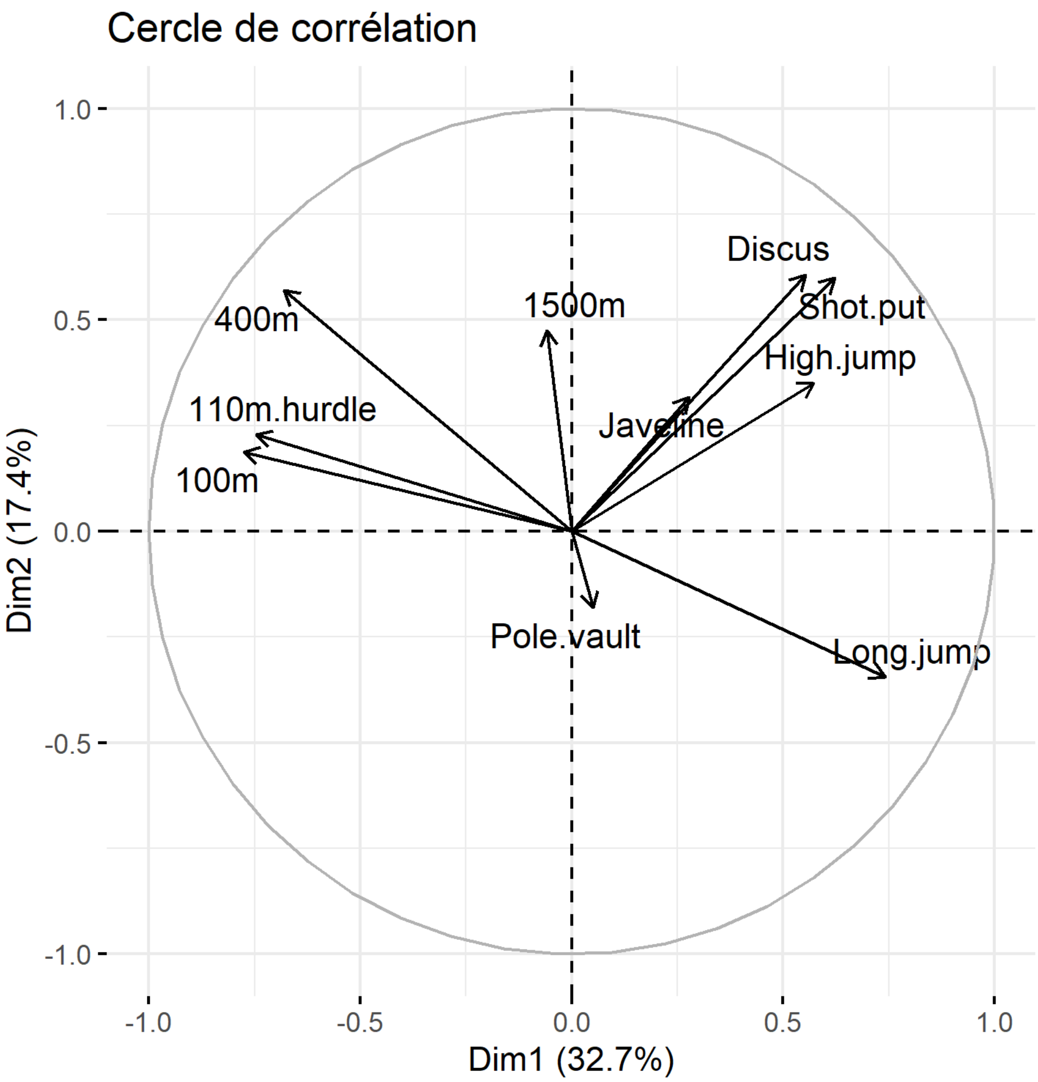
Un algorithme de consensus est un processus qui permet de trouver un accord sur une valeur unique de données entre des processus ou des systèmes distribués.

4 Projet 2

```
data(decathlon)

decathlon_scaled <- as_tibble(scale(decathlon[, 1:10], scale = TRUE, center = TRUE))

pca_decathlon <- PCA(decathlon[, 1:10], graph = FALSE, scale.unit = TRUE)
```

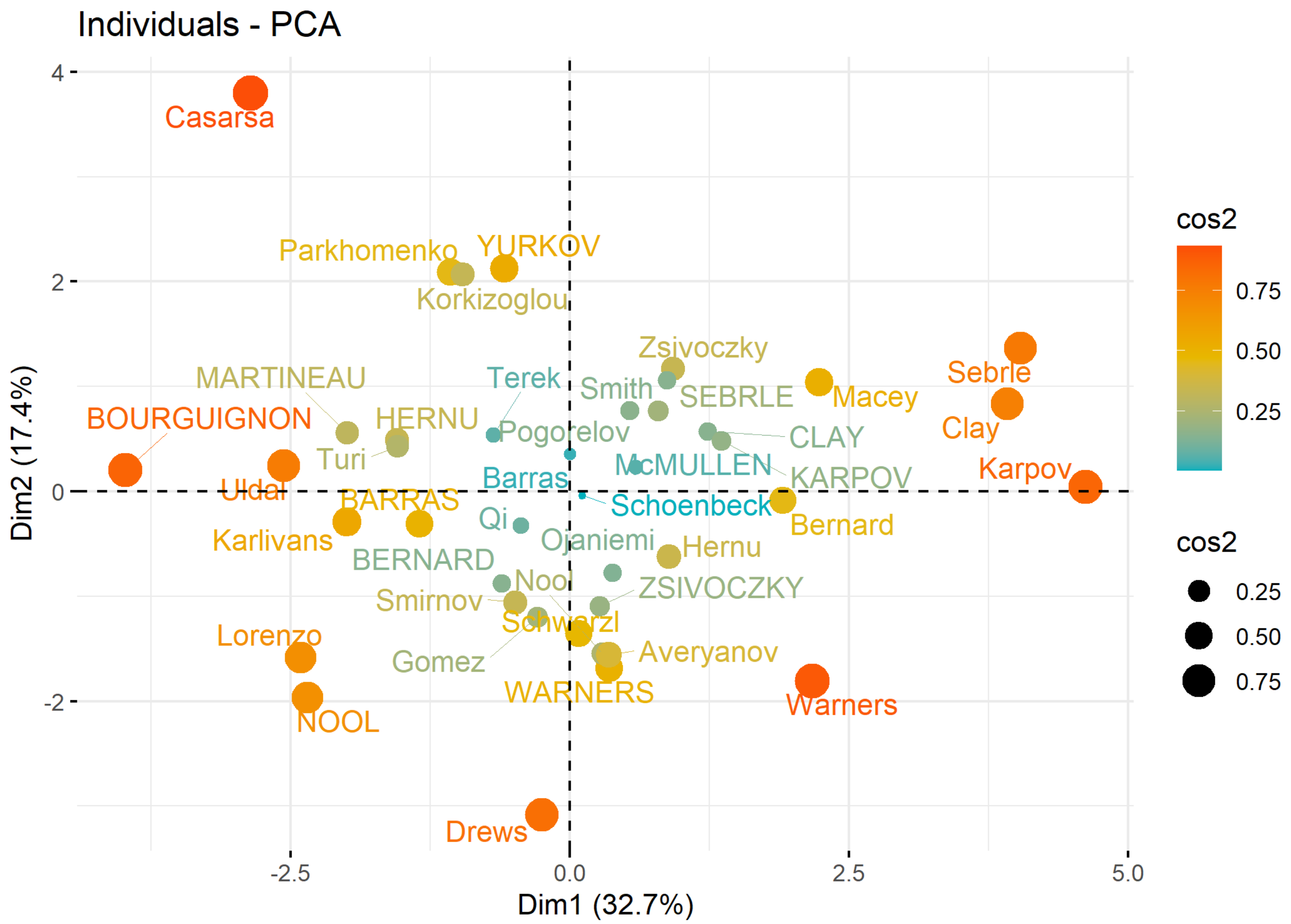
4.1 Question 1

```
fviz_pca_var(pca_decathlon, col.var = "black", repel = TRUE) +
  labs(title = "Cercle de corrélation")
```



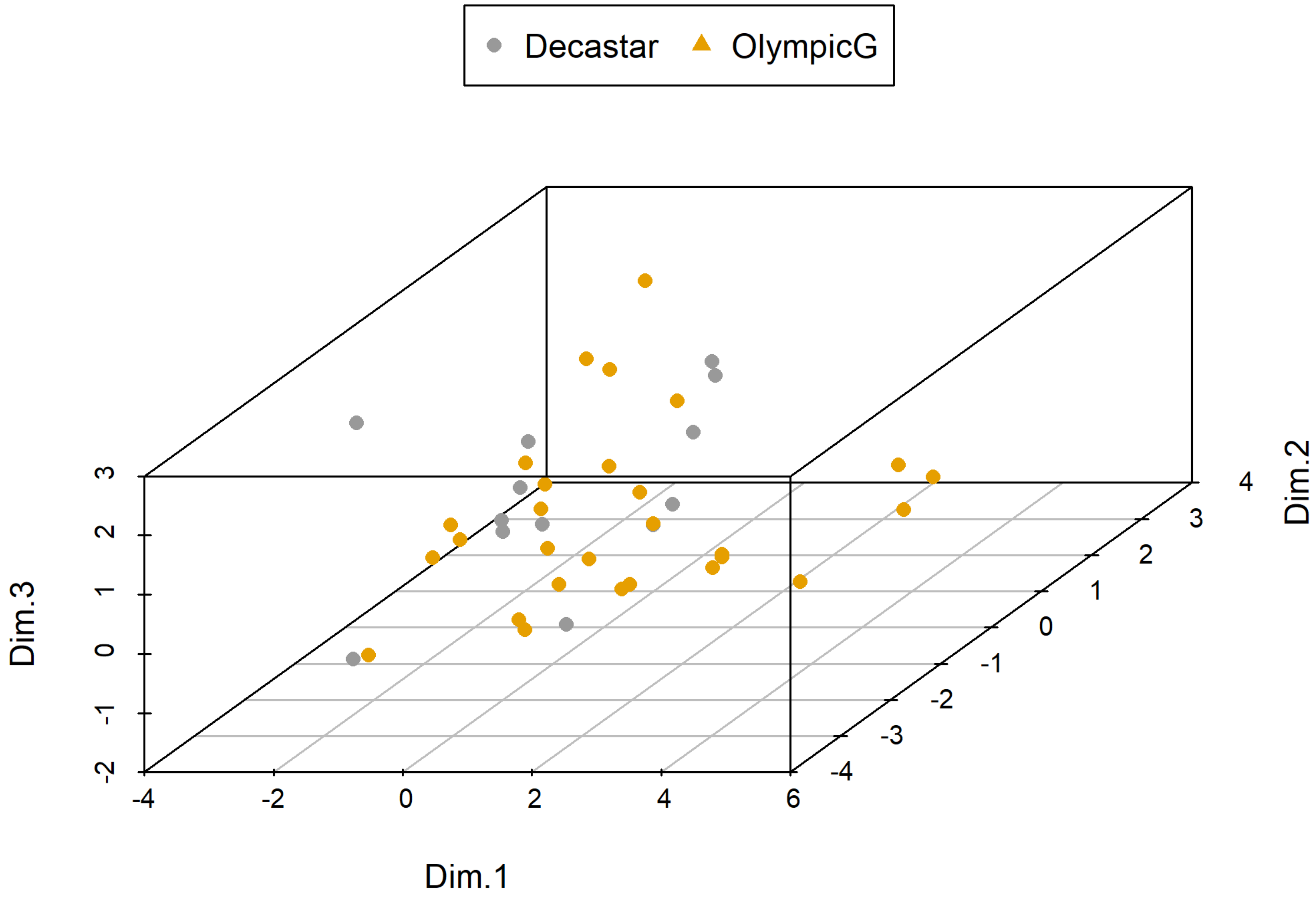
4.2 Question 2

```r
fviz_pca_ind(pca_decathlon,
  col.ind = "cos2", pointsize = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE
)
```



4.3 Question 3

```
colors <- c("#999999", "#E69F00", "#56B4E9")
colors <- colors[as.numeric(decathlon$Competition)]

s3d <- scatterplot3d(pca_decathlon$ind$coord[, 1:3],
  angle = 55,
  pch = 16,
  color = colors
)
legend("top",
  legend = levels(decathlon$Competition),
  col = c("#999999", "#E69F00", "#56B4E9"), pch = c(16, 17, 18),
  inset = -0.25, xpd = TRUE, horiz = TRUE
)
```



4.4 Question 4

```
dimdesc(pca_decathlon, axes = 1:5)
```

```
## $Dim.1
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##             correlation      p.value
## Long.jump     0.7418997 2.849886e-08
## Shot.put      0.6225026 1.388321e-05
## High.jump     0.5719453 9.362285e-05
## Discus        0.5524665 1.802220e-04
## 400m         -0.6796099 1.028175e-06
## 110m.hurdle  -0.7462453 2.136962e-08
## 100m         -0.7747198 2.778467e-09
##
## $Dim.2
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##           correlation      p.value
## Discus      0.6063134 2.650745e-05
## Shot.put    0.5983033 3.603567e-05
## 400m        0.5694378 1.020941e-04
## 1500m       0.4742238 1.734405e-03
## High.jump   0.3502936 2.475025e-02
## Javeline    0.3169891 4.344974e-02
## Long.jump  -0.3454213 2.696969e-02
##
## $Dim.3
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##            correlation      p.value
## 1500m        0.7821428 1.554450e-09
## Pole.vault   0.6917567 5.480172e-07
## Javeline    -0.3896554 1.179331e-02
##
## $Dim.4
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##            correlation      p.value
## Javeline     0.7122773 1.761578e-07
## Pole.vault   0.5515340 1.857748e-04
##
## $Dim.5
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##            correlation      p.value
## High.jump    0.5554396 0.0001635051
## Pole.vault   0.3299593 0.0351316637
```

Dim 1 : Saut/Lancé

Dim 2 : Lancé

Dim3 : ??

Dim 4 : Utilisation d'un baton

Dim 5 : Hauteur

4.5 Question 5

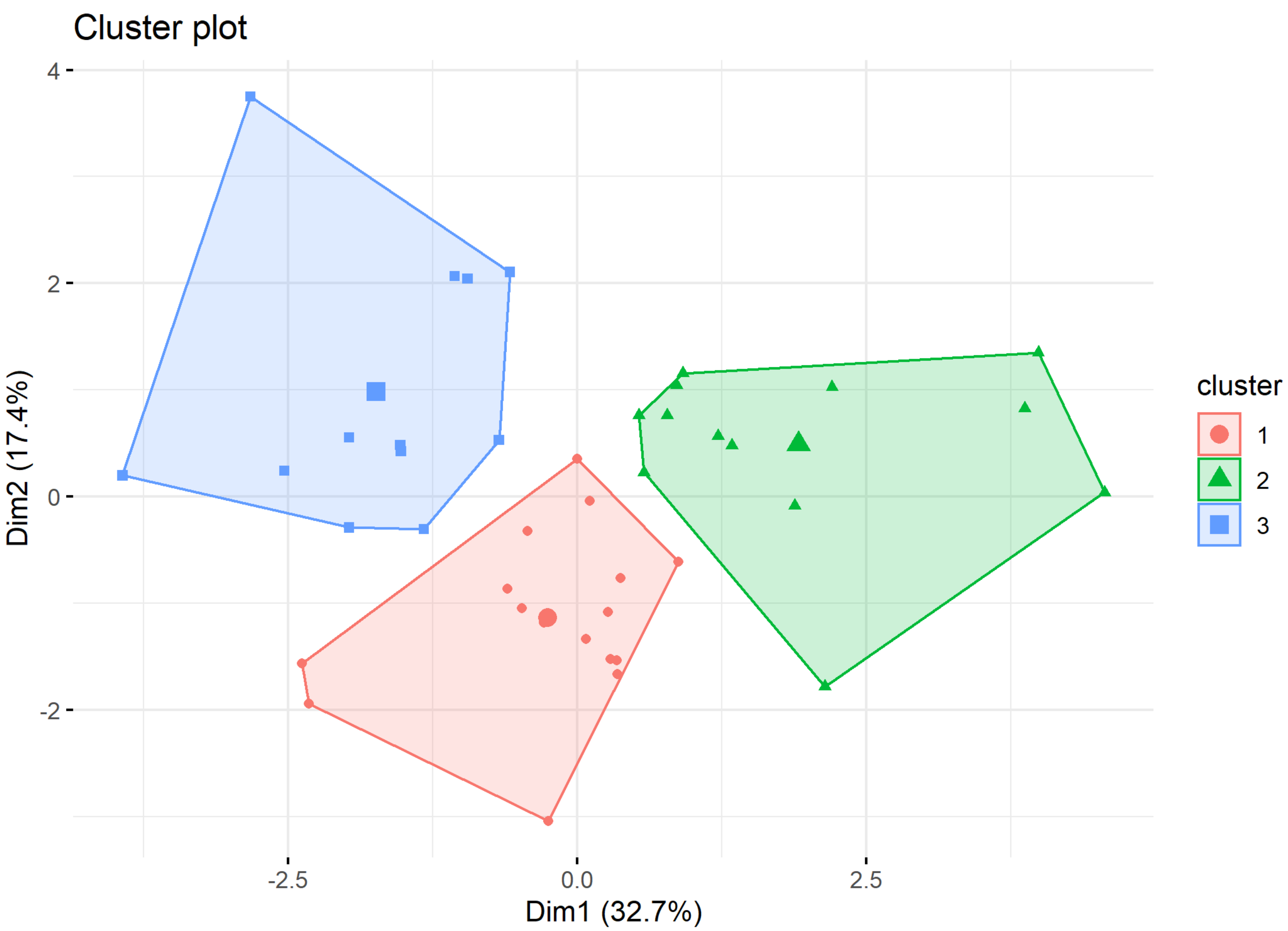
- K-mean :
  - Simple, flexible, faible coût de calcul,.
  - Fonctionne sur des grands jeu de données.

- Non reproductible
- CAH :
    - Choix du nombre de clusters.
    - Reproductible.
    - Permet d'obtenir une hiérarchie de partitions et ainsi de choisir le nombre de classes optimal.
    - N'est pas adaptée à des tableaux de données volumineux.

4.6 Question 6

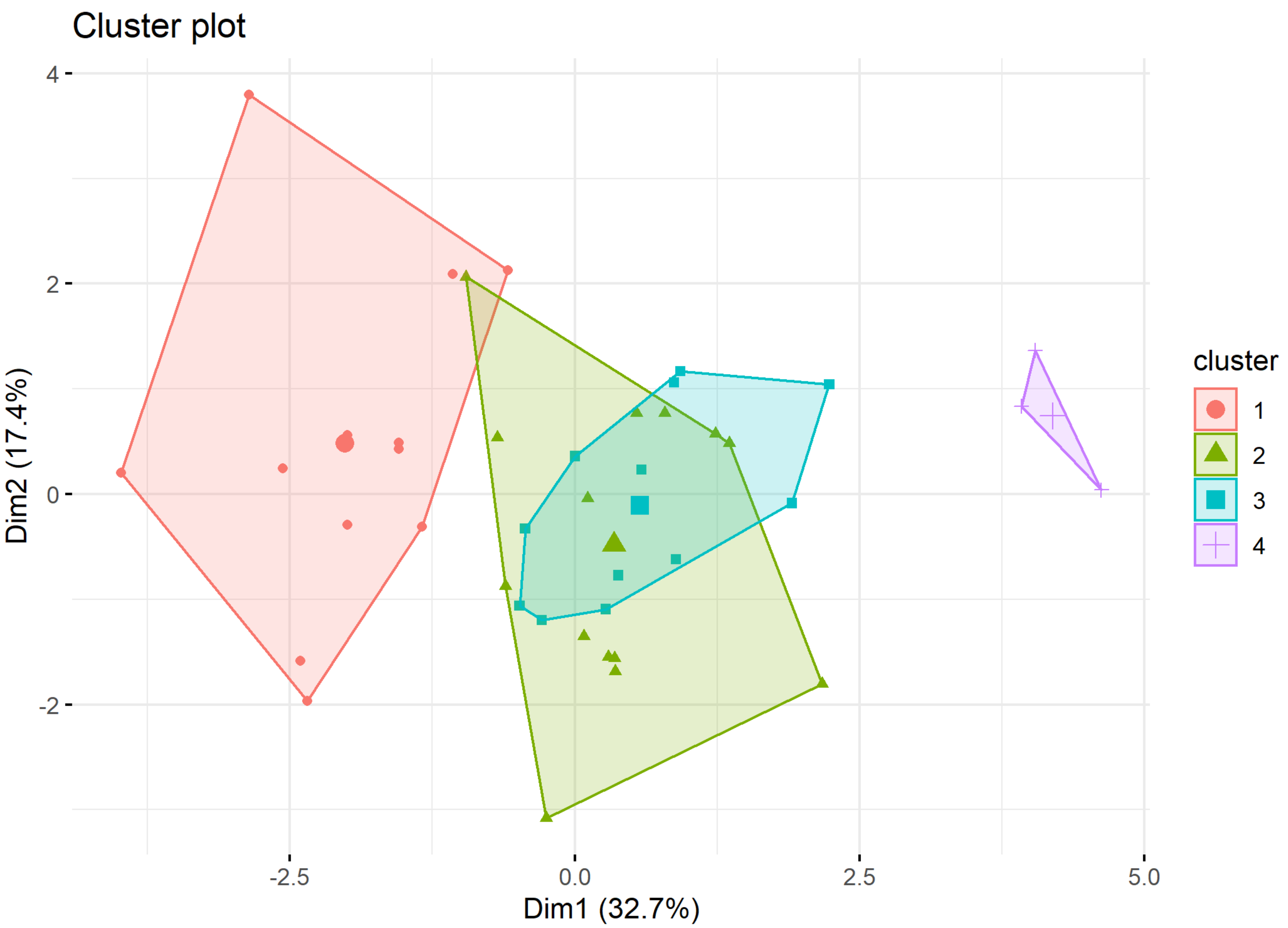
4.6.1 K-mean

```
kmean_decathlon_scaled <- kmeans(pca_decathlon$ind$coord, 3, nstart = 50, iter.max = 500)

fviz_cluster(kmean_decathlon_scaled,
          data = decathlon[, 1:10],
          geom = "point",
          ellipse.type = "convex",
          ggtheme = theme_minimal()
)
```

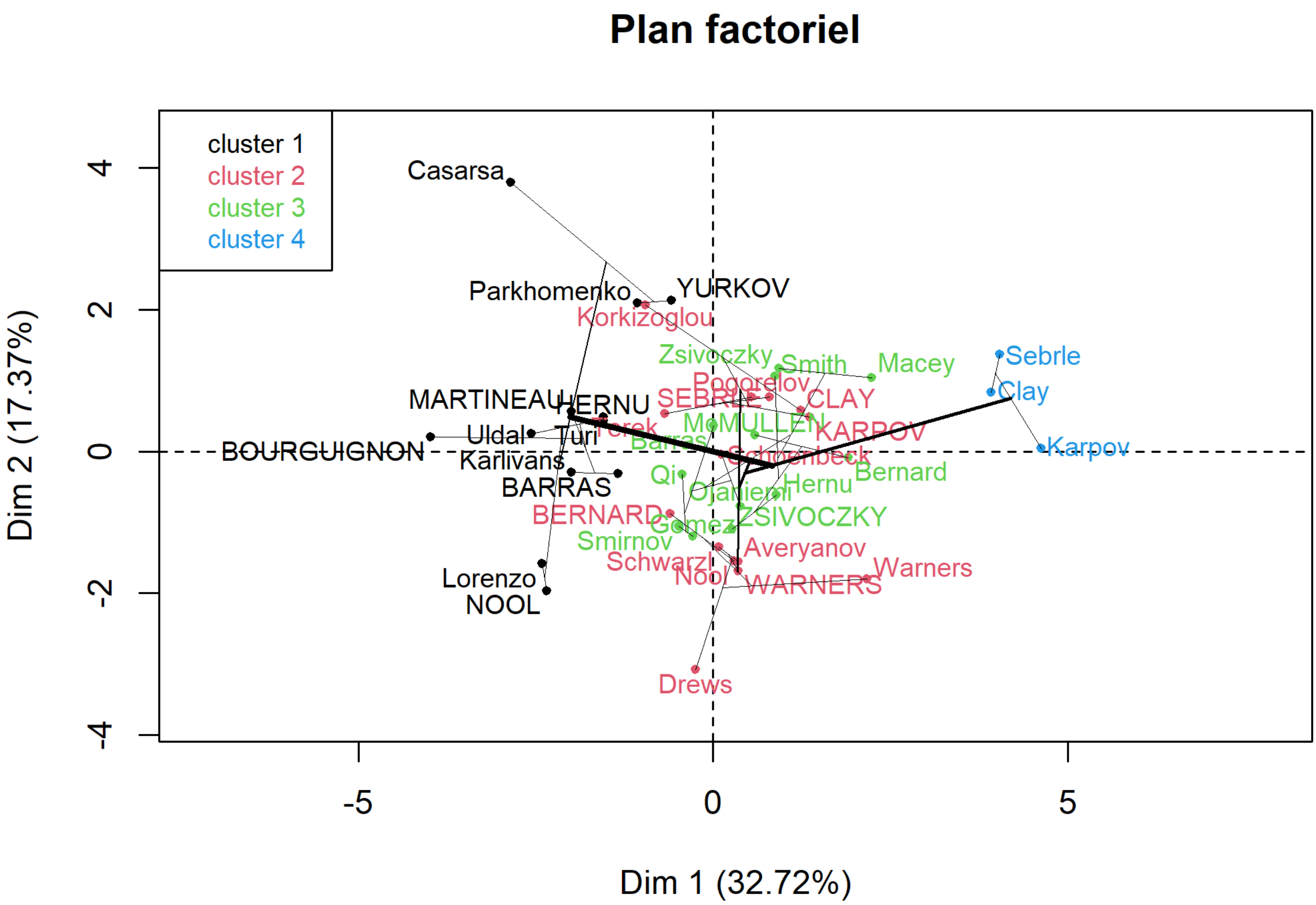


4.6.2 CAH

```r
cah_hcpc_decathlon <- HCPC(pca_decathlon, graph = FALSE)

fviz_cluster(cah_hcpc_decathlon, geom = "point", ellipse = TRUE,
          ggtheme = theme_minimal())
```

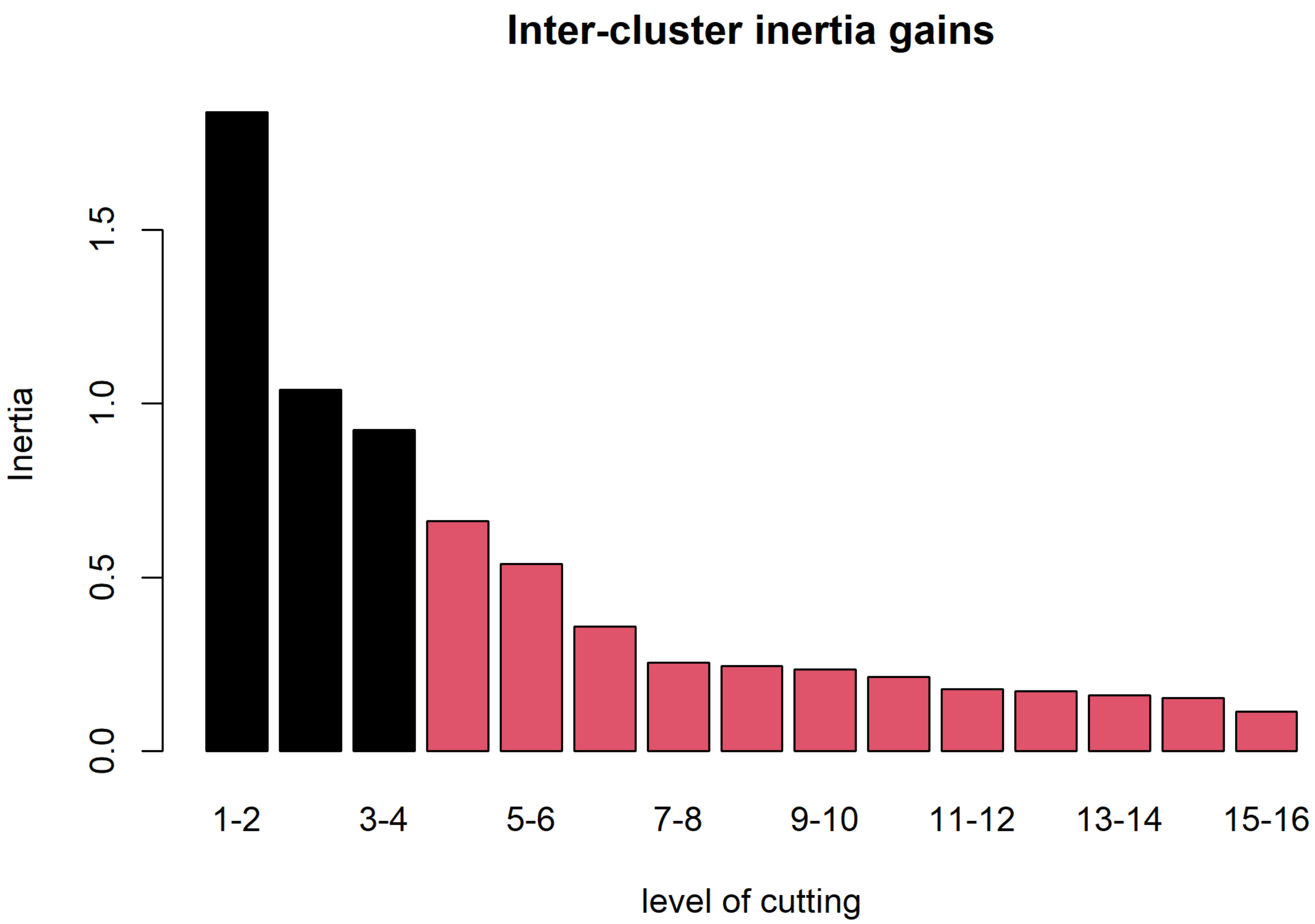



```r
plot.HCPC(cah_hcpc_decathlon, choice = 'map', draw.tree = TRUE, title = 'Plan factoriel')
```

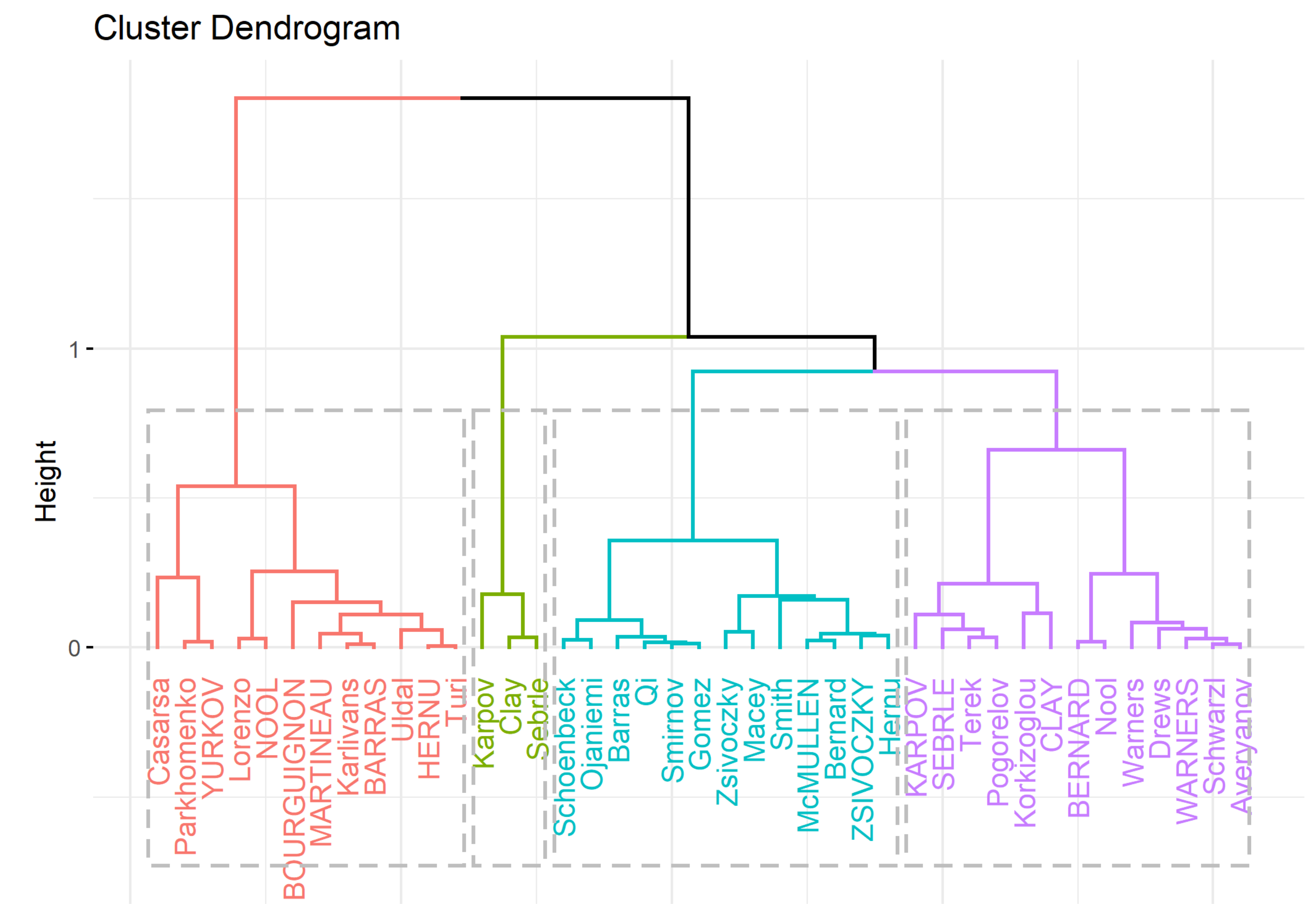


4.7 Question 7

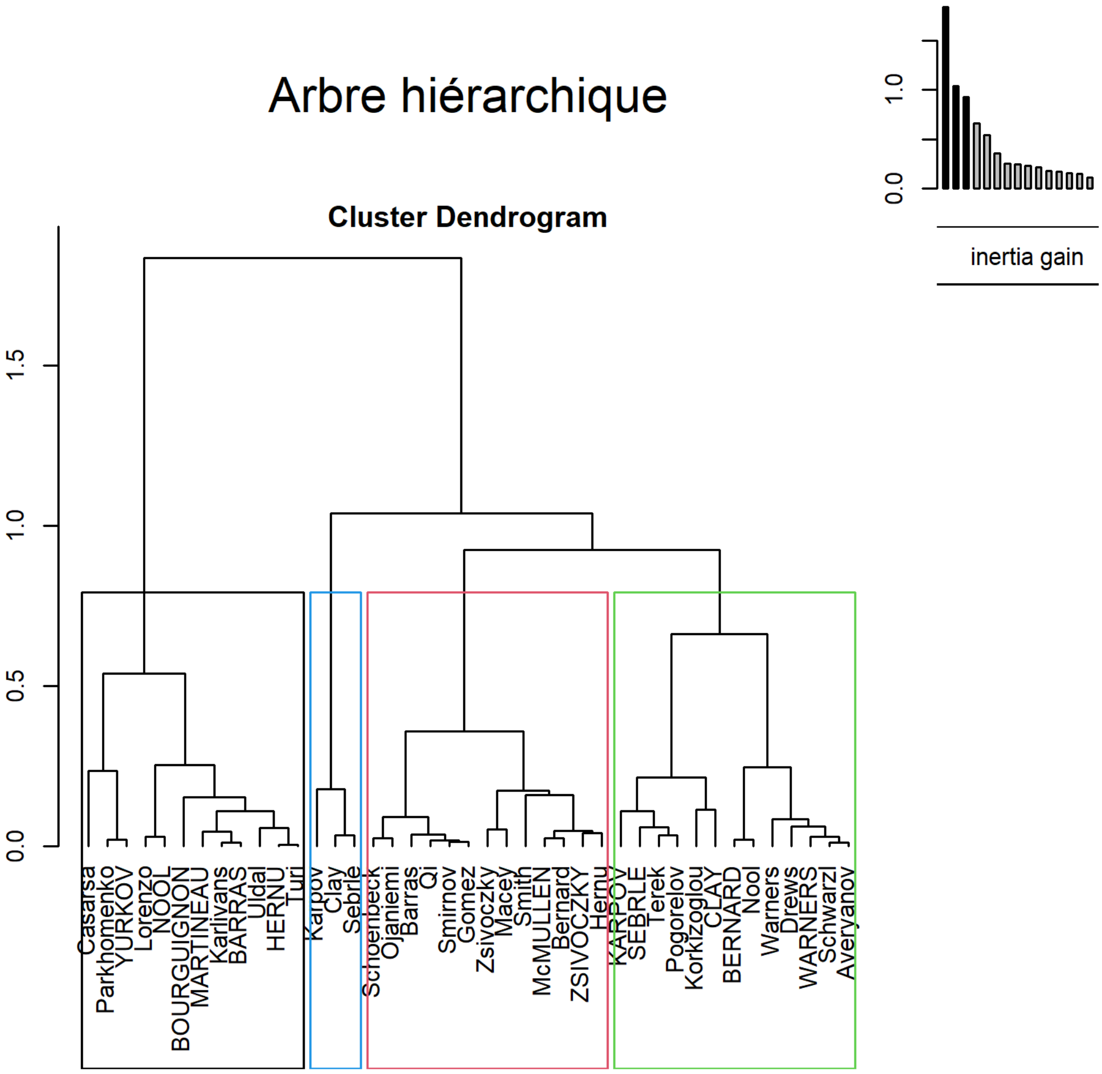
```r
# Calcul du nombre de cut.
plot(cah_hcpc_decathlon, choice = "bar")
```




```
fviz_dend(cah_hcpc_decathlon, rect = TRUE, ggtheme = theme_minimal()
)
```



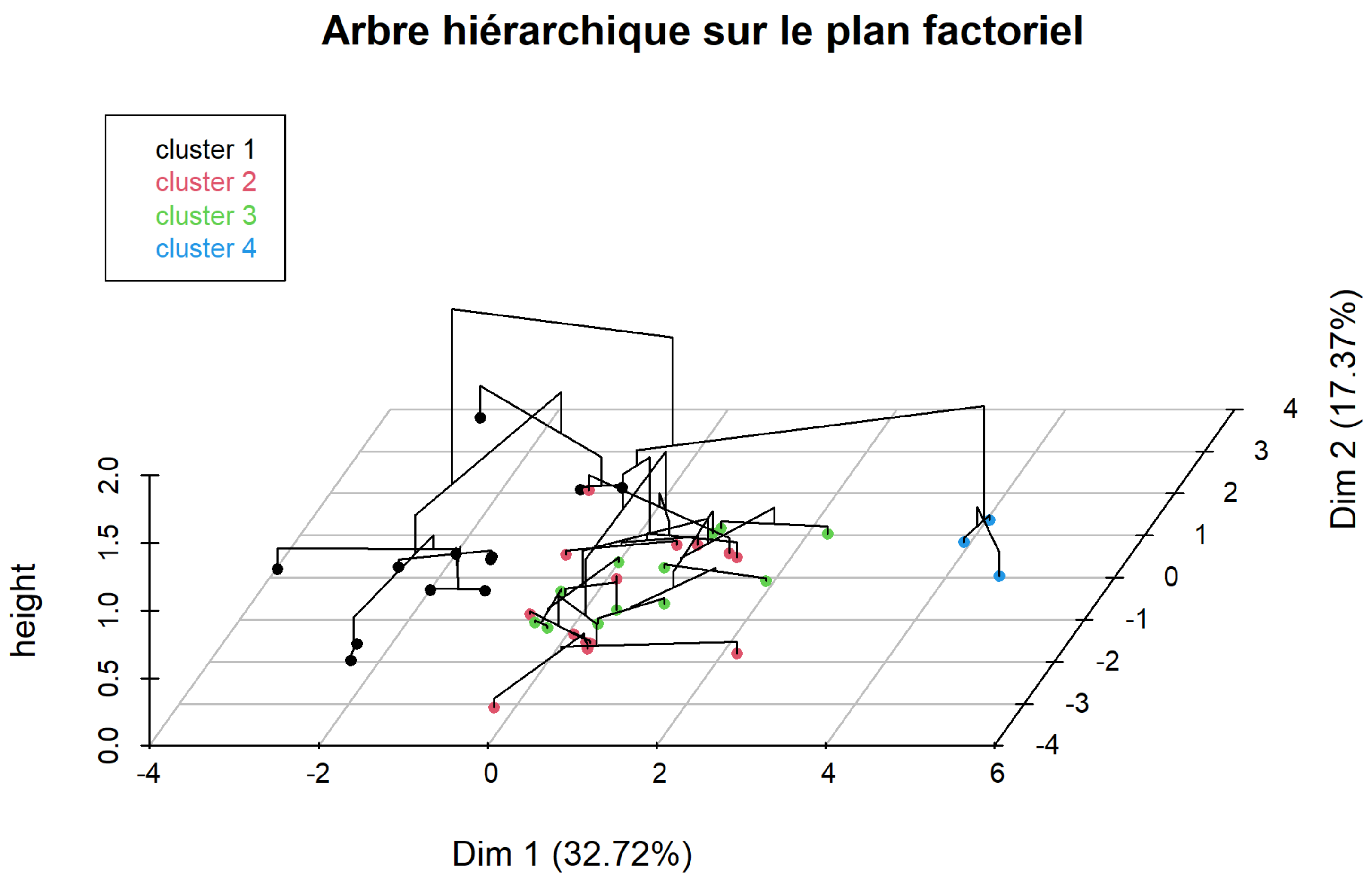

```
plot.HCPC(cah_hcpc_decathlon, choice = 'tree', title = 'Arbre hiérarchique')
```



Le graphique des gains d'inerties inter-cluster nous permet de voir que 3-4 groupes semble le nombre optimal.

4.8 Question 8

```
plot(cah_hcpc_decathlon, choice = "3D.map", ind.names=FALSE, centers.plot = FALSE, angle = 70,
    title = 'Arbre hiérarchique sur le plan factoriel')
```



5 Projet 3

```
users_db <- read_csv("Dataset/users.db.csv")
users_db <- subset(users_db, select = -c(1, 9))
names <- c(3,4,5,8,11,12,13,14)
users_db[,names] <- lapply(users_db[,names] , factor)
```

```
res_MCA <- MCA(users_db,
  ncp = 5,
  quali.sup = c(1,7,8),
  quanti.sup = c(2,9, 10),
  graph = FALSE
)
```

5.1 Question 1

L'analyse des correspondances multiples (ACM) est une technique d'analyse des variables catégorielles. Il s'agit essentiellement d'une forme d'analyse factorielle pour les données catégorielles.

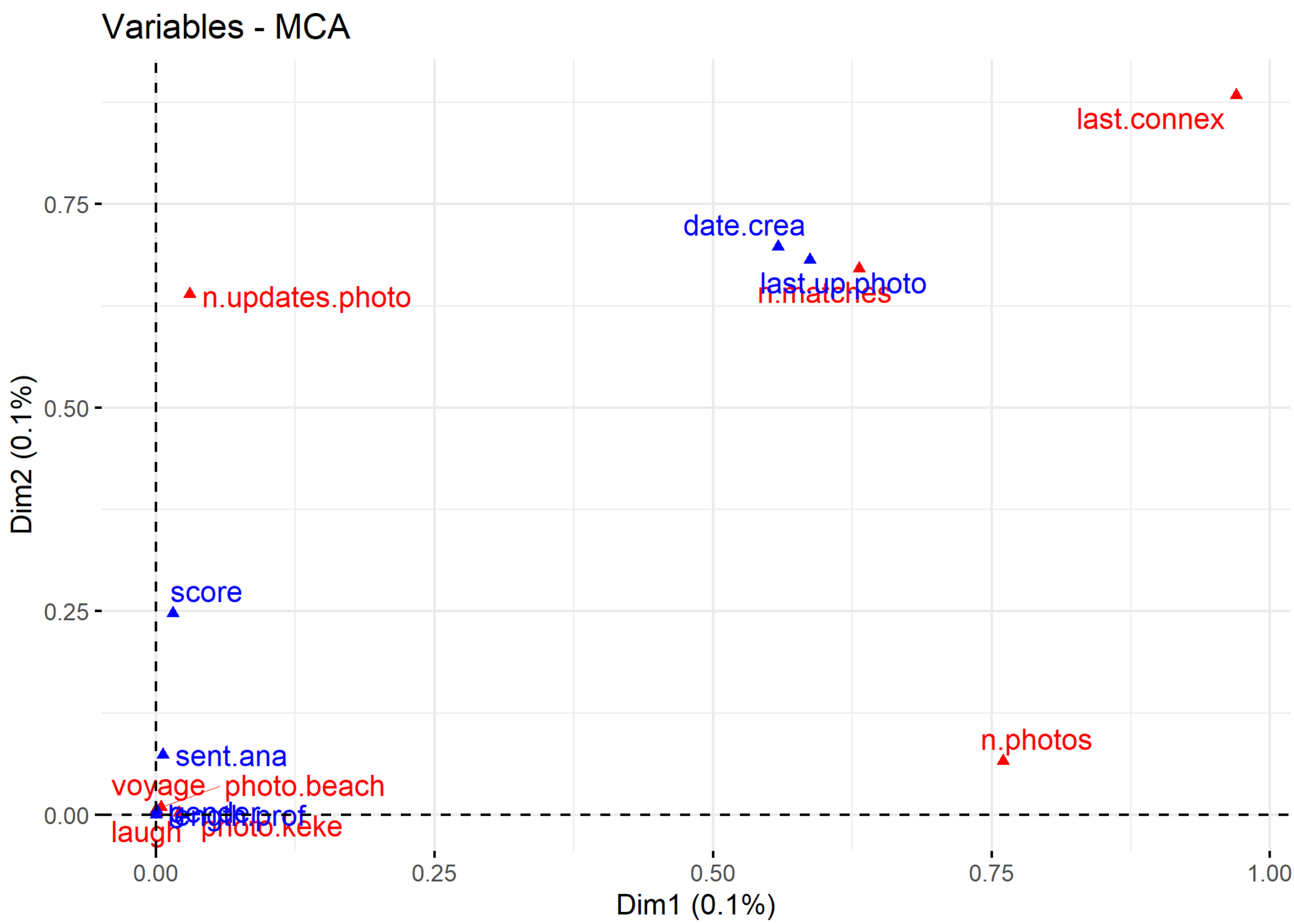
L'ACM est généralement utilisée pour analyser des données d'enquête ou de sondage.

L'objectif est d'identifier:

- Un groupe de personnes ayant un profil similaire dans leurs réponses aux questions
- Les associations entre les catégories des variables

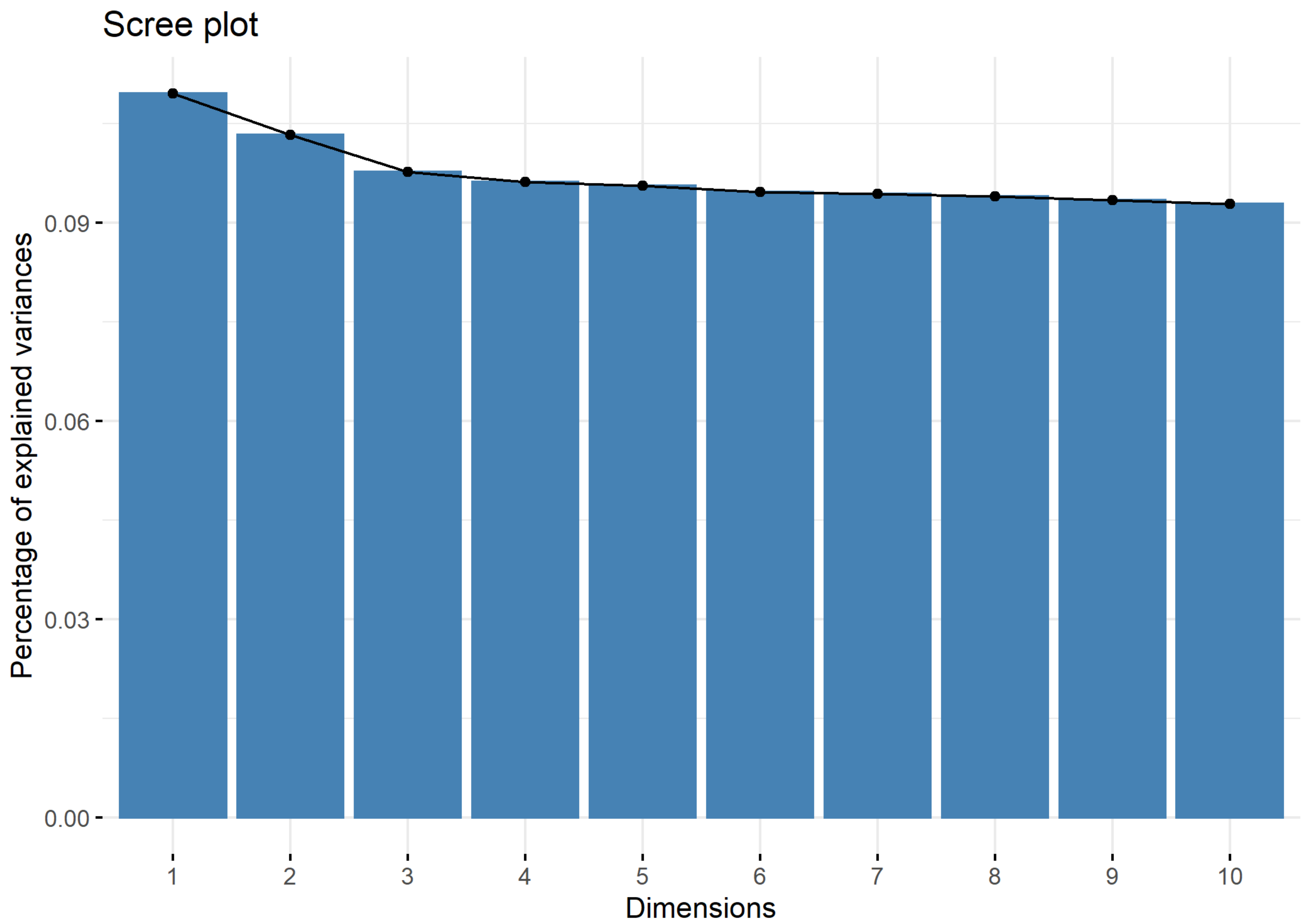
5.2 Question 2

```r
fviz_mca_var(res_MCA, choice = c("var"), repel = TRUE)
```



5.3 Question 3

```r
fviz_eig(res_MCA)
```



Nous pouvons voir que la variance représentée pour Dim1 et Dim2 est faible (< 0,5%), et que celle-ci est de moins en moins représentée pour chaques dimensions.

5.4 Question 4

5.5 Question 5

```
dimdesc(res_MCA, axes = 1:2)
```

```
## $`Dim 1`
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##          correlation      p.value
## score      0.12337049 1.198604e-11
## sent.ana  0.07912544 1.432128e-05
##
## Link between the variable and the categorical variable (1-way anova)
## =============================================
##                        R2        p.value
## n.matches       0.631175167  0.000000e+00
## n.photos        0.760345527  0.000000e+00
## last.connex     0.969542455 2.100670e-286
## photo.keke      0.020725834  2.313046e-15
## n.updates.photo 0.030108503  2.700181e-15
## photo.beach     0.004688925  1.744430e-04
##
## Link between variable abd the categories of the categorical variables
## ================================================================
##                                      Estimate       p.value
## last.connex=last.connex_2017-11-11     26.12922084  0.000000e+00
```

```
## n.photos=n.photos_12                 23.93521746  0.000000e+00
## n.matches=n.matches_45               10.36903901  0.000000e+00
## date.crea=date.crea_2017-11-01       13.08228154 3.547294e-311
## last.up.photo=last.up.photo_2017-11-05 13.00135167 2.057581e-306
## last.up.photo=last.up.photo_2011-07-31  7.52570334  5.652975e-44
## last.connex=last.connex_2011-08-24      7.52147480  5.652975e-44
## last.up.photo=last.up.photo_2009-08-24  7.48500987  1.669088e-43
## last.connex=last.connex_2009-08-31      7.48078133  1.669088e-43
## date.crea=date.crea_2009-08-03          7.48661586  1.669088e-43
## last.up.photo=last.up.photo_2012-05-22  5.51626933  4.594902e-24
## date.crea=date.crea_2012-05-02          5.51787531  4.594902e-24
## last.connex=last.connex_2014-07-07      3.88364467  6.356889e-24
## date.crea=date.crea_2011-07-18          3.80476836  6.033425e-23
## last.connex=last.connex_2012-07-06      2.58604271  2.247154e-16
## last.up.photo=last.up.photo_2014-06-07  3.16609397  2.609095e-16
## photo.keke=photo.keke_1                 0.10784518  2.313046e-15
## date.crea=date.crea_2014-05-14          2.95991650  2.054299e-14
## n.updates.photo=n.updates.photo_3       0.03136509  1.421820e-12
## photo.beach=photo.beach_0               0.04991156  1.744430e-04
## last.up.photo=last.up.photo_2014-06-09  1.65526248  2.589439e-03
## date.crea=date.crea_2014-06-04          1.65686846  2.589439e-03
## date.crea=date.crea_2012-06-11          1.16112494  3.489244e-02
## last.up.photo=last.up.photo_2012-07-09  1.09502550  4.635029e-02
## n.updates.photo=n.updates.photo_4      -0.05006507  3.605912e-02
## n.matches=n.matches_4                  -0.33355235  4.066466e-03
## n.matches=n.matches_6                  -0.33049705  2.362962e-03
## n.matches=n.matches_5                  -0.31529663  2.356929e-03
## n.photos=n.photos_3                    -2.14452007  2.097287e-03
## n.photos=n.photos_4                    -2.26641209  9.510528e-04
## n.updates.photo=n.updates.photo_-1     -0.33066165  8.377592e-04
## n.updates.photo=n.updates.photo_1      -0.17749679  5.359699e-04
## photo.beach=photo.beach_1              -0.04991156  1.744430e-04
## n.updates.photo=n.updates.photo_0      -0.24774881  3.925452e-07
## photo.keke=photo.keke_0                -0.10784518  2.313046e-15
##
## $`Dim 2`
##
## Link between the variable and the continuous variables (R-square)
## =================================================================================
##          correlation       p.value
## score      0.4975133 2.151094e-187
## sent.ana   0.2711594  1.024427e-51
##
## Link between the variable and the categorical variable (1-way anova)
## =============================================
##                          R2       p.value
## n.matches       0.670732983 0.000000e+00
## n.updates.photo 0.639460667 0.000000e+00
## last.connex     0.883684183 1.041497e-74
## n.photos        0.066214493 6.783679e-38
## photo.beach     0.009109332 1.628550e-07
## voyage          0.006110353 1.817112e-05
## gender          0.005420945 2.900715e-04
## photo.keke      0.002741828 4.120542e-03
## laugh           0.001739712 2.233719e-02
##
## Link between variable abd the categories of the categorical variables
## ================================================================
##                                          Estimate       p.value
## n.updates.photo=n.updates.photo_6       0.25714922  3.790845e-34
## n.matches=n.matches_28                  0.81697194  8.622866e-33
## n.matches=n.matches_26                  0.50213685  1.375791e-22
## n.matches=n.matches_27                  0.44509283  9.190268e-20
## last.connex=last.connex_2009-06-09      4.52939172  1.282079e-17
## n.updates.photo=n.updates.photo_9       3.90008773  1.282079e-17
## date.crea=date.crea_2009-04-30          4.53634208  1.282079e-17
## n.photos=n.photos_2                     0.37899043  1.531581e-17
## n.matches=n.matches_31                  0.39746876  2.557503e-13
## n.photos=n.photos_4                     0.07355372  3.065610e-13
## n.matches=n.matches_25                  0.23198509  5.361721e-11
```

```
## n.matches=n.matches_24               0.23635067  6.250489e-11
## n.matches=n.matches_32               0.33101755  1.151762e-09
## n.matches=n.matches_29               0.30887815  2.519244e-09
## n.matches=n.matches_33               0.51543918  4.056848e-08
## n.photos=n.photos_6                  0.01918924  9.132640e-08
## photo.beach=photo.beach_0            0.06755702  1.628550e-07
## n.updates.photo=n.updates.photo_7    0.03815851  9.349258e-07
## n.matches=n.matches_30               0.22409657  9.397045e-07
## last.up.photo=last.up.photo_2009-05-29  1.48423083  1.477809e-06
## n.matches=n.matches_38               0.53137174  4.716010e-06
## voyage=voyage_0                      0.05023113  1.817112e-05
## last.up.photo=last.up.photo_2009-11-19  2.25399162  2.431544e-05
## last.connex=last.connex_2009-12-30   2.24471364  2.431544e-05
## n.updates.photo=n.updates.photo_8    1.61540965  2.431544e-05
## n.matches=n.matches_35               0.29897310  3.263289e-05
## n.matches=n.matches_19               0.05184366  5.438359e-05
## n.matches=n.matches_37               0.33234335  9.351394e-05
## n.matches=n.matches_34               0.23179326  1.073581e-04
## n.matches=n.matches_44               0.41524204  1.340400e-04
## gender=gender_0                      0.01451947  1.923470e-04
## n.matches=n.matches_41               0.39758336  2.124474e-04
## n.matches=n.matches_40               0.40884437  2.998702e-04
## last.up.photo=last.up.photo_2016-04-23  1.36082144  3.187960e-04
## n.matches=n.matches_39               0.29666960  4.308029e-04
## last.connex=last.connex_2018-12-12   1.62335941  2.261849e-03
## last.up.photo=last.up.photo_2015-04-08  1.12993839  2.824124e-03
## last.connex=last.connex_2015-04-13   1.57340251  3.079043e-03
## n.photos=n.photos_11                 1.79830354  3.079043e-03
## last.connex=last.connex_2016-05-16   1.56311461  3.277619e-03
## last.connex=last.connex_2018-07-29   1.10145281  3.332596e-03
## photo.keke=photo.keke_1              0.03809145  4.120542e-03
## last.connex=last.connex_2018-12-25   0.84141574  5.942684e-03
## n.matches=n.matches_42               0.25574456  7.802463e-03
## last.up.photo=last.up.photo_2018-02-01  1.38663281  9.545609e-03
## last.connex=last.connex_2018-02-28   1.37735482  9.545609e-03
## date.crea=date.crea_2018-01-23       1.38430518  9.545609e-03
## n.matches=n.matches_43               0.42125559  1.292953e-02
## last.up.photo=last.up.photo_2018-12-11  0.66390612  1.327476e-02
## last.connex=last.connex_2011-04-21   0.92721238  1.340351e-02
## date.crea=date.crea_2017-07-08       0.75609897  1.424494e-02
## last.connex=last.connex_2016-12-31   1.29885602  1.450033e-02
## date.crea=date.crea_2016-12-02       1.30580638  1.450033e-02
## last.up.photo=last.up.photo_2010-12-08  1.26356117  1.822437e-02
## last.connex=last.connex_2010-12-09   1.25428319  1.822437e-02
## date.crea=date.crea_2009-04-10       0.86950143  2.137275e-02
## last.up.photo=last.up.photo_2009-07-20  1.23034281  2.152015e-02
## last.connex=last.connex_2009-08-25   1.22106483  2.152015e-02
## last.up.photo=last.up.photo_2011-02-06  1.22846167  2.172140e-02
## last.connex=last.connex_2011-02-08   1.21918368  2.172140e-02
## date.crea=date.crea_2011-01-21       1.22613404  2.172140e-02
## laugh=laugh_0                        0.02806305  2.233719e-02
## date.crea=date.crea_2015-09-23       0.85110020  2.428900e-02
## date.crea=date.crea_2009-10-19       0.84667502  2.503969e-02
## date.crea=date.crea_2016-04-10       0.84385577  2.552835e-02
## last.connex=last.connex_2015-07-06   0.58677837  2.630383e-02
## last.connex=last.connex_2018-12-16   1.16626471  2.809384e-02
## last.connex=last.connex_2012-01-26   1.16010081  2.893170e-02
## date.crea=date.crea_2011-11-10       1.16705117  2.893170e-02
## last.up.photo=last.up.photo_2018-07-05  1.16224541  2.992809e-02
## date.crea=date.crea_2018-06-22       1.15991778  2.992809e-02
## last.connex=last.connex_2010-06-17   1.15053813  3.027408e-02
## n.matches=n.matches_55               0.98475288  3.027408e-02
## last.connex=last.connex_2015-11-14   1.14908802  3.048225e-02
## last.up.photo=last.up.photo_2009-07-17  1.15444344  3.105150e-02
## last.connex=last.connex_2009-07-23   1.14516545  3.105150e-02
## date.crea=date.crea_2009-06-08       1.15211581  3.105150e-02
## last.up.photo=last.up.photo_2019-06-01  1.15444344  3.105150e-02
## last.connex=last.connex_2019-06-19   1.14516545  3.105150e-02
## last.up.photo=last.up.photo_2019-05-12  1.15121855  3.152631e-02
## last.connex=last.connex_2019-06-30   1.14194056  3.152631e-02
```

```
## date.crea=date.crea_2019-04-27          1.14889092  3.152631e-02
## last.connex=last.connex_2016-05-15      1.13997230  3.181915e-02
## last.up.photo=last.up.photo_2014-06-26  1.14797188  3.201060e-02
## last.connex=last.connex_2014-07-30      1.13869389  3.201060e-02
## last.connex=last.connex_2016-10-02      0.79910882  3.291890e-02
## last.connex=last.connex_2014-02-24      1.13156394  3.309655e-02
## date.crea=date.crea_2014-02-08          1.13851430  3.309655e-02
## last.up.photo=last.up.photo_2017-07-16  1.13429098  3.412194e-02
## last.connex=last.connex_2017-07-22      1.12501299  3.412194e-02
## last.up.photo=last.up.photo_2015-06-11  1.13388904  3.418573e-02
## date.crea=date.crea_2015-05-28          1.13156141  3.418573e-02
## last.connex=last.connex_2018-01-05      0.55639774  3.505583e-02
## last.up.photo=last.up.photo_2015-09-08  1.11451041  3.738433e-02
## last.connex=last.connex_2015-09-10      1.10523242  3.738433e-02
## date.crea=date.crea_2015-07-25          1.11218278  3.738433e-02
## last.up.photo=last.up.photo_2015-03-04  0.64144241  3.836402e-02
## last.up.photo=last.up.photo_2015-09-20  1.10045506  3.986055e-02
## last.connex=last.connex_2015-10-02      1.09117707  3.986055e-02
## last.connex=last.connex_2014-10-27      0.75824174  4.289833e-02
## last.up.photo=last.up.photo_2016-04-29  0.76638780  4.320767e-02
## last.up.photo=last.up.photo_2014-04-01  1.07989952  4.373153e-02
## last.connex=last.connex_2014-04-10      1.07062153  4.373153e-02
## date.crea=date.crea_2014-03-13          1.07757189  4.373153e-02
## last.connex=last.connex_2015-08-31      0.75433517  4.397390e-02
## date.crea=date.crea_2011-02-05          0.76066560  4.414665e-02
## date.crea=date.crea_2014-05-19          0.75957337  4.445241e-02
## last.up.photo=last.up.photo_2018-12-06  0.76090863  4.473177e-02
## last.up.photo=last.up.photo_2019-04-25  1.07031631  4.564195e-02
## last.connex=last.connex_2019-04-28      1.06103833  4.564195e-02
## date.crea=date.crea_2019-04-08          1.06798869  4.564195e-02
## last.connex=last.connex_2011-03-13      1.05164196  4.758289e-02
## last.up.photo=last.up.photo_2016-12-26  0.74948872  4.805439e-02
## date.crea=date.crea_2018-04-17          0.60611427  4.955870e-02
## date.crea=date.crea_2016-02-08         -1.04517251  4.998330e-02
## date.crea=date.crea_2011-12-18         -0.73991710  4.957596e-02
## last.connex=last.connex_2014-06-28     -0.74944379  4.878844e-02
## date.crea=date.crea_2019-06-03         -0.74542865  4.790404e-02
## last.connex=last.connex_2010-03-09     -0.75267825  4.781464e-02
## last.up.photo=last.up.photo_2009-03-04 -1.05301473  4.779672e-02
## last.connex=last.connex_2009-03-06     -1.06229272  4.779672e-02
## date.crea=date.crea_2009-02-17         -1.05534235  4.779672e-02
## last.up.photo=last.up.photo_2009-08-24 -1.05460900  4.746134e-02
## last.connex=last.connex_2009-08-31     -1.06388699  4.746134e-02
## date.crea=date.crea_2009-08-03         -1.05693663  4.746134e-02
## last.up.photo=last.up.photo_2016-04-01 -0.74532898  4.724177e-02
## last.connex=last.connex_2017-01-28     -0.61833349  4.679788e-02
## n.matches=n.matches_7                  -0.08249924  4.632841e-02
## last.connex=last.connex_2012-08-13     -1.07274383  4.563395e-02
## date.crea=date.crea_2012-10-25         -1.06658924  4.547271e-02
## date.crea=date.crea_2009-11-15         -1.06958049  4.487088e-02
## last.connex=last.connex_2011-04-04     -1.07850864  4.447666e-02
## date.crea=date.crea_2011-03-03         -1.07155828  4.447666e-02
## last.connex=last.connex_2018-01-02     -0.76432358  4.444273e-02
## last.connex=last.connex_2016-09-13     -0.62583716  4.416752e-02
## last.up.photo=last.up.photo_2012-03-06 -0.75828990  4.353987e-02
## date.crea=date.crea_2015-06-19         -1.07793074  4.322628e-02
## date.crea=date.crea_2013-03-04         -1.08062309  4.270699e-02
## date.crea=date.crea_2011-12-16         -0.76438303  4.251142e-02
## date.crea=date.crea_2010-10-30         -0.54107583  4.209880e-02
## last.connex=last.connex_2012-01-19     -1.09106158  4.204206e-02
## date.crea=date.crea_2011-12-26         -1.08411121  4.204206e-02
## date.crea=date.crea_2019-03-27         -0.76708104  4.178723e-02
## n.matches=n.matches_23                 -0.03701606  4.146203e-02
## last.connex=last.connex_2016-11-17     -0.63572006  4.089441e-02
## last.connex=last.connex_2013-11-06     -0.63626360  4.072051e-02
## last.connex=last.connex_2012-02-16     -1.09842194  4.066767e-02
## date.crea=date.crea_2012-01-04         -1.09147158  4.066767e-02
## last.up.photo=last.up.photo_2019-02-11 -1.08914396  4.066767e-02
## last.connex=last.connex_2019-03-14     -1.09842194  4.066767e-02
## date.crea=date.crea_2019-01-20         -1.09147158  4.066767e-02
## date.crea=date.crea_2016-07-19         -1.09187031  4.059432e-02
```

```
## date.crea=date.crea_2016-07-19         -1.09187031  4.059432e-02
## last.up.photo=last.up.photo_2014-01-17 -0.62757503  4.053266e-02
## last.connex=last.connex_2018-10-31     -1.09968099  4.043643e-02
## last.up.photo=last.up.photo_2012-12-27 -0.77040419  4.030423e-02
## last.up.photo=last.up.photo_2012-08-30 -1.09509735  3.958403e-02
## last.connex=last.connex_2012-10-03     -1.10437534  3.958403e-02
## date.crea=date.crea_2012-08-15         -1.09742498  3.958403e-02
## last.connex=last.connex_2011-02-04     -1.10437534  3.958403e-02
## date.crea=date.crea_2011-01-25         -1.09742498  3.958403e-02
## last.connex=last.connex_2010-05-11     -1.10784110  3.896455e-02
## date.crea=date.crea_2010-04-19         -1.10089074  3.896455e-02
## last.connex=last.connex_2018-05-11     -0.78521384  3.889595e-02
## last.connex=last.connex_2014-07-29     -1.11139608  3.833768e-02
## last.up.photo=last.up.photo_2011-06-04 -1.10251639  3.826799e-02
## date.crea=date.crea_2011-05-06         -1.10484402  3.826799e-02
## date.crea=date.crea_2012-08-09         -0.78171454  3.803869e-02
## last.connex=last.connex_2016-07-26     -0.64560592  3.782822e-02
## last.connex=last.connex_2013-11-22     -1.12330623  3.629961e-02
## date.crea=date.crea_2013-10-31         -1.11635587  3.629961e-02
## last.up.photo=last.up.photo_2015-11-20 -1.11478791  3.617280e-02
## last.up.photo=last.up.photo_2017-12-27 -0.64268757  3.596086e-02
## last.connex=last.connex_2011-12-10     -1.12538849  3.595294e-02
## last.up.photo=last.up.photo_2013-04-05 -0.78848492  3.585535e-02
## last.connex=last.connex_2017-12-11     -0.80108454  3.508528e-02
## last.connex=last.connex_2010-07-11     -0.80214741  3.484188e-02
## last.connex=last.connex_2012-08-31     -0.65757346  3.437997e-02
## last.up.photo=last.up.photo_2013-08-01 -1.13133689  3.350268e-02
## n.matches=n.matches_13                 -0.07530942  3.263071e-02
## last.up.photo=last.up.photo_2018-03-29 -1.13757645  3.254065e-02
## last.connex=last.connex_2018-05-06     -1.14685444  3.254065e-02
## last.up.photo=last.up.photo_2010-09-25 -1.13757645  3.254065e-02
## last.connex=last.connex_2010-09-23     -1.14685444  3.254065e-02
## date.crea=date.crea_2010-08-21         -1.13990407  3.254065e-02
## last.up.photo=last.up.photo_2019-05-14 -1.14117412  3.199677e-02
## date.crea=date.crea_2019-04-18         -1.14350175  3.199677e-02
## last.up.photo=last.up.photo_2016-09-23 -0.80645636  3.185515e-02
## last.up.photo=last.up.photo_2019-08-26 -1.14286008  3.174459e-02
## last.connex=last.connex_2019-09-02     -1.15213806  3.174459e-02
## last.connex=last.connex_2016-08-31     -0.66949015  3.121675e-02
## last.up.photo=last.up.photo_2015-02-06 -1.14695610  3.113900e-02
## last.connex=last.connex_2015-03-03     -1.15623409  3.113900e-02
## date.crea=date.crea_2015-01-24         -1.14928373  3.113900e-02
## last.connex=last.connex_2014-06-11     -0.82850916  2.924982e-02
## last.connex=last.connex_2017-03-04     -0.82853863  2.924402e-02
## last.connex=last.connex_2016-04-04     -0.82866213  2.921977e-02
## date.crea=date.crea_2011-03-13         -0.82830226  2.794998e-02
## date.crea=date.crea_2011-10-03         -0.82938764  2.774549e-02
## last.connex=last.connex_2014-12-20     -0.68657328  2.711926e-02
## date.crea=date.crea_2014-05-14         -0.83336112  2.700782e-02
## last.up.photo=last.up.photo_2015-08-04 -1.18979930  2.537634e-02
## date.crea=date.crea_2014-11-07         -0.68929175  2.501250e-02
## date.crea=date.crea_2017-05-10         -0.84686514  2.462591e-02
## last.up.photo=last.up.photo_2010-06-10 -1.20067222  2.406979e-02
## last.connex=last.connex_2010-06-12     -1.20995021  2.406979e-02
## date.crea=date.crea_2018-02-06         -0.69563423  2.370876e-02
## last.up.photo=last.up.photo_2010-10-31 -0.85034929  2.365815e-02
## last.up.photo=last.up.photo_2018-03-03 -0.60016375  2.363816e-02
## n.matches=n.matches_14                 -0.28213551  2.357668e-02
## last.up.photo=last.up.photo_2011-09-05 -1.20648359  2.339567e-02
## last.connex=last.connex_2011-10-03     -1.21576158  2.339567e-02
## last.up.photo=last.up.photo_2017-09-05 -1.20648359  2.339567e-02
## last.connex=last.connex_2017-10-07     -1.21576158  2.339567e-02
## date.crea=date.crea_2017-07-28         -1.20881122  2.339567e-02
## last.connex=last.connex_2014-10-20     -1.21576158  2.339567e-02
## date.crea=date.crea_2014-08-31         -1.20881122  2.339567e-02
## last.up.photo=last.up.photo_2019-11-12 -0.85225276  2.334845e-02
## date.crea=date.crea_2012-08-02         -0.85549600  2.320073e-02
## last.up.photo=last.up.photo_2019-01-30 -1.21072808  2.291372e-02
## last.connex=last.connex_2019-02-24     -1.22000606  2.291372e-02
## last.up.photo=last.up.photo_2015-12-14 -0.85653379  2.266472e-02
## last.up.photo=last.up.photo_2014-03-26 -1.21445893  2.249724e-02
```

```
## last.connex=last.connex_2014-04-05     -1.22373692  2.249724e-02
## last.connex=last.connex_2010-04-22     -1.22373692  2.249724e-02
## laugh=laugh_1                         -0.02806305  2.233719e-02
## date.crea=date.crea_2015-11-04         -0.86159020  2.223804e-02
## date.crea=date.crea_2012-11-24         -0.70364401  2.214682e-02
## last.up.photo=last.up.photo_2009-05-08 -0.70527310  2.140881e-02
## last.connex=last.connex_2018-12-18     -0.71490650  2.134358e-02
## last.connex=last.connex_2016-12-03     -1.24758810  1.998689e-02
## date.crea=date.crea_2016-10-26         -1.24063774  1.998689e-02
## last.up.photo=last.up.photo_2019-10-11 -1.23831012  1.998689e-02
## last.connex=last.connex_2019-10-19     -1.24758810  1.998689e-02
## date.crea=date.crea_2019-09-29         -1.24063774  1.998689e-02
## last.connex=last.connex_2018-11-02     -0.88992149  1.912987e-02
## last.connex=last.connex_2016-08-15     -1.26501041  1.831108e-02
## last.connex=last.connex_2019-07-04     -1.26501041  1.831108e-02
## last.connex=last.connex_2016-05-30     -0.89965333  1.784562e-02
## n.matches=n.matches_18                -0.30189798  1.742153e-02
## last.up.photo=last.up.photo_2017-02-22 -0.89800228  1.689256e-02
## date.crea=date.crea_2017-12-15         -0.91168401  1.555648e-02
## last.up.photo=last.up.photo_2017-12-06 -1.29281829  1.514802e-02
## last.connex=last.connex_2017-12-02     -1.30209627  1.514802e-02
## date.crea=date.crea_2017-12-06         -1.29514591  1.514802e-02
## last.up.photo=last.up.photo_2011-11-02 -1.29958802  1.462567e-02
## last.connex=last.connex_2011-11-10     -1.30886601  1.462567e-02
## date.crea=date.crea_2019-07-31         -0.75398916  1.422889e-02
## last.connex=last.connex_2014-05-19     -0.93138010  1.416817e-02
## last.up.photo=last.up.photo_2016-10-20 -1.30993403  1.385802e-02
## date.crea=date.crea_2016-10-10         -1.31226166  1.385802e-02
## last.connex=last.connex_2009-05-17     -1.31938010  1.384585e-02
## date.crea=date.crea_2009-04-04         -1.31242973  1.384585e-02
## date.crea=date.crea_2015-10-08         -0.92944958  1.365213e-02
## last.up.photo=last.up.photo_2016-07-23 -1.31427032  1.354700e-02
## last.up.photo=last.up.photo_2018-05-23 -1.31972337  1.316464e-02
## date.crea=date.crea_2011-08-31         -1.32619613  1.288040e-02
## last.up.photo=last.up.photo_2013-07-28 -0.93601733  1.277832e-02
## last.connex=last.connex_2015-11-10     -1.33811597  1.254680e-02
## last.up.photo=last.up.photo_2010-07-23 -0.77009092  1.202789e-02
## date.crea=date.crea_2010-07-09         -0.94645212  1.202542e-02
## date.crea=date.crea_2014-05-17         -1.33995804  1.197526e-02
## last.connex=last.connex_2019-12-04     -0.95588908  1.180250e-02
## date.crea=date.crea_2019-11-10         -0.94893871  1.180250e-02
## last.up.photo=last.up.photo_2012-06-27 -1.34388499  1.158285e-02
## last.connex=last.connex_2012-07-01     -1.35316297  1.158285e-02
## last.connex=last.connex_2012-06-22     -0.96369850  1.112604e-02
## last.up.photo=last.up.photo_2013-02-02 -0.95706427  1.090497e-02
## last.connex=last.connex_2015-08-11     -0.96677199  1.086940e-02
## last.up.photo=last.up.photo_2012-09-04 -1.35591406  1.086009e-02
## last.connex=last.connex_2012-09-27     -1.36519205  1.086009e-02
## date.crea=date.crea_2012-08-19         -1.35824168  1.086009e-02
## last.connex=last.connex_2012-11-04     -0.79106259  1.079301e-02
## date.crea=date.crea_2013-07-17         -0.96659139  1.032258e-02
## n.matches=n.matches_20                -0.02758319  9.754650e-03
## last.up.photo=last.up.photo_2015-11-08 -1.37583043  9.751196e-03
## last.connex=last.connex_2015-11-07     -1.38510842  9.751196e-03
## last.up.photo=last.up.photo_2014-06-15 -1.37790968  9.641454e-03
## last.up.photo=last.up.photo_2010-10-05 -0.97503542  9.502854e-03
## last.up.photo=last.up.photo_2012-06-23 -1.39126905  8.962227e-03
## date.crea=date.crea_2016-07-16         -1.39366540  8.958846e-03
## last.up.photo=last.up.photo_2015-10-25 -0.98415786  8.854554e-03
## last.up.photo=last.up.photo_2017-11-30 -1.39977226  8.552467e-03
## date.crea=date.crea_2017-10-22         -1.40209989  8.552467e-03
## last.up.photo=last.up.photo_2014-05-15 -1.40171197  8.461373e-03
## date.crea=date.crea_2014-04-24         -1.40403959  8.461373e-03
## last.up.photo=last.up.photo_2017-05-31 -0.99319196  8.251747e-03
## date.crea=date.crea_2017-09-04         -0.99640742  8.194533e-03
## last.connex=last.connex_2019-08-04     -0.82548751  7.785722e-03
## last.up.photo=last.up.photo_2018-12-12 -1.41710016  7.768892e-03
## last.connex=last.connex_2018-12-24     -1.42637814  7.768892e-03
## last.connex=last.connex_2014-05-18     -1.42637814  7.768892e-03
## last.connex=last.connex_2019-03-24     -1.42775464  7.709493e-03
```

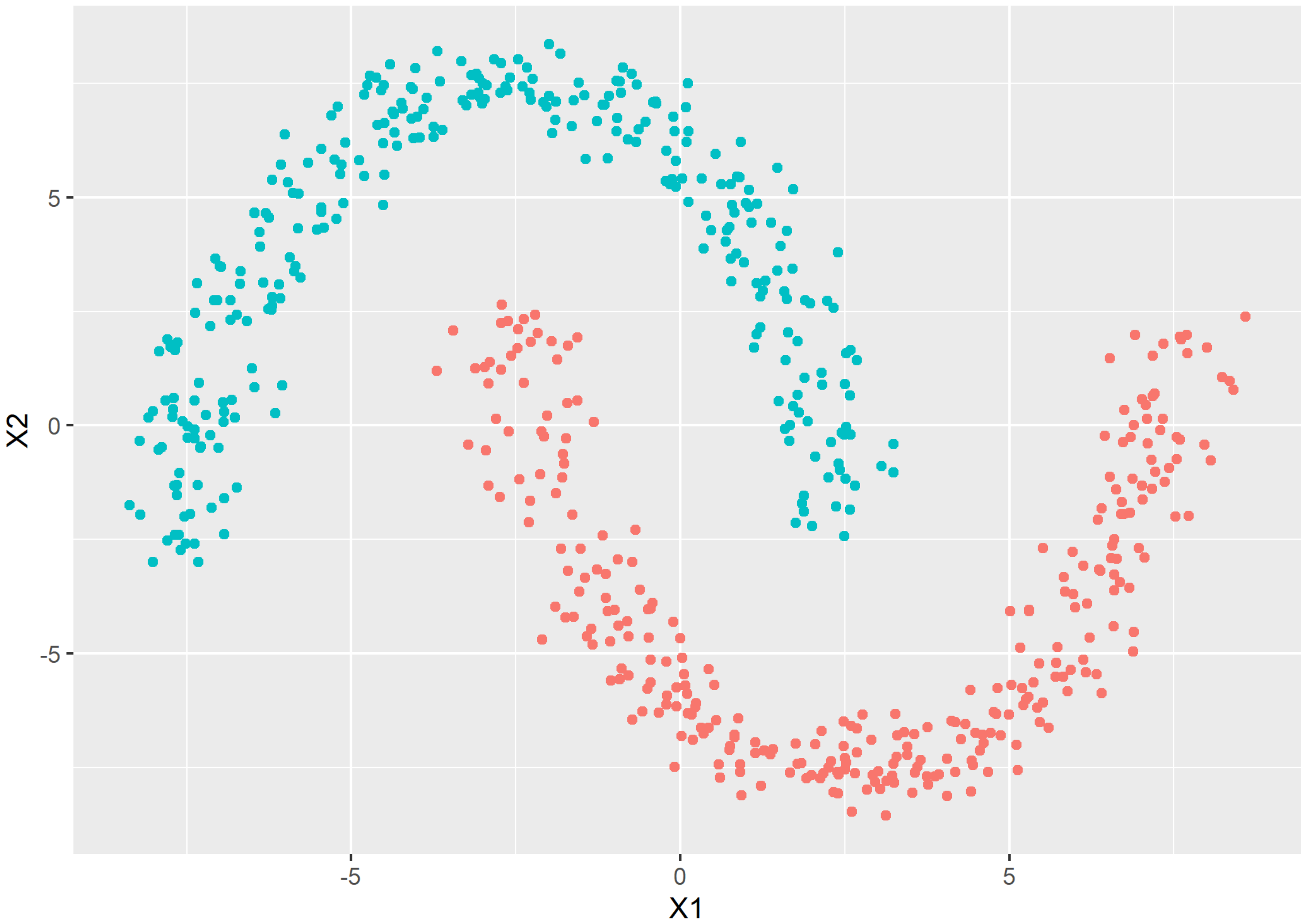
```
## last.up.photo=last.up.photo_2012-08-11 -1.00327397  7.622578e-03
## last.up.photo=last.up.photo_2012-09-01 -1.42239760  7.542519e-03
## last.connex=last.connex_2012-09-18     -1.43167559  7.542519e-03
## date.crea=date.crea_2019-10-06         -1.00769353  7.497546e-03
## last.up.photo=last.up.photo_2014-03-11 -1.42404546  7.473317e-03
## date.crea=date.crea_2014-02-20         -1.42637308  7.473317e-03
## last.connex=last.connex_2017-03-01     -1.01560799  7.440544e-03
## date.crea=date.crea_2017-08-08         -1.42741957  7.429665e-03
## last.up.photo=last.up.photo_2012-03-04 -1.00848800  7.314360e-03
## date.crea=date.crea_2018-09-27         -1.43030060  7.310670e-03
## last.connex=last.connex_2016-08-26     -1.01902088  7.241868e-03
## last.connex=last.connex_2010-05-14     -1.02229569  7.055718e-03
## date.crea=date.crea_2019-09-08         -1.02499675  6.531822e-03
## last.up.photo=last.up.photo_2019-07-30 -1.45243881  6.367049e-03
## date.crea=date.crea_2019-07-02         -1.45476644  6.367049e-03
## last.connex=last.connex_2009-05-15     -1.03581982  6.331628e-03
## date.crea=date.crea_2010-04-24         -1.03191866  6.177907e-03
## last.connex=last.connex_2014-08-16     -0.85454941  5.856698e-03
## date.crea=date.crea_2014-03-31         -1.47944385  5.527685e-03
## date.crea=date.crea_2018-10-16         -1.48564066  5.333237e-03
## n.matches=n.matches_15                 -0.02475416  4.903613e-03
## last.up.photo=last.up.photo_2018-04-28 -1.50948302  4.578365e-03
## date.crea=date.crea_2018-04-10         -1.51181064  4.578365e-03
## date.crea=date.crea_2014-04-10         -1.51233972  4.564155e-03
## date.crea=date.crea_2018-11-07         -1.51940737  4.378126e-03
## last.up.photo=last.up.photo_2010-04-12 -1.51753092  4.366487e-03
## last.up.photo=last.up.photo_2013-04-17 -1.52181049  4.257474e-03
## last.connex=last.connex_2014-05-07     -0.88707967  4.216913e-03
## photo.keke=photo.keke_0                -0.03809145  4.120542e-03
## last.up.photo=last.up.photo_2011-04-10 -1.52797247  4.104840e-03
## last.connex=last.connex_2011-04-14     -1.53725046  4.104840e-03
## last.up.photo=last.up.photo_2015-08-25 -1.08033028  4.067773e-03
## n.matches=n.matches_22                 -0.34770700  4.053140e-03
## last.up.photo=last.up.photo_2019-09-23 -1.54337550  3.744776e-03
## last.connex=last.connex_2019-09-27     -1.55265348  3.744776e-03
## date.crea=date.crea_2019-09-03         -1.54570312  3.744776e-03
## last.up.photo=last.up.photo_2017-12-02 -1.56051086  3.378113e-03
## last.up.photo=last.up.photo_2014-05-01 -1.10396655  3.329015e-03
## last.up.photo=last.up.photo_2014-08-09 -1.56519395  3.283756e-03
## date.crea=date.crea_2014-07-23         -1.56752158  3.283756e-03
## last.up.photo=last.up.photo_2014-04-30 -1.11570776  3.009445e-03
## date.crea=date.crea_2010-09-14         -1.12968948  2.720150e-03
## last.connex=last.connex_2019-08-22     -1.14004694  2.640505e-03
## date.crea=date.crea_2015-08-04         -1.13339859  2.633549e-03
## date.crea=date.crea_2017-12-03         -1.60946964  2.539620e-03
## last.up.photo=last.up.photo_2010-04-27 -1.13965360  2.442737e-03
## last.connex=last.connex_2010-05-02     -1.14893159  2.442737e-03
## last.connex=last.connex_2015-11-21     -1.14929201  2.435007e-03
## last.up.photo=last.up.photo_2016-10-02 -0.93154797  2.395478e-03
## date.crea=date.crea_2015-11-25         -1.15803574  2.119376e-03
## date.crea=date.crea_2016-09-17         -1.63869285  2.116087e-03
## last.connex=last.connex_2012-08-12     -1.65822835  1.954490e-03
## date.crea=date.crea_2012-08-06         -1.65127799  1.954490e-03
## date.crea=date.crea_2017-05-21         -1.65196561  1.945996e-03
## last.up.photo=last.up.photo_2017-02-26 -1.64969315  1.945316e-03
## date.crea=date.crea_2017-01-31         -1.65693624  1.885594e-03
## last.up.photo=last.up.photo_2014-12-17 -1.66393428  1.776892e-03
## last.connex=last.connex_2010-11-08     -0.83968180  1.763860e-03
## date.crea=date.crea_2015-09-28         -1.68057714  1.621189e-03
## last.up.photo=last.up.photo_2011-02-09 -1.72804740  1.172169e-03
## date.crea=date.crea_2011-01-31         -1.73037503  1.172169e-03
## last.connex=last.connex_2015-12-10     -1.00671435  1.151746e-03
## last.up.photo=last.up.photo_2019-12-24 -1.22827166  1.092034e-03
## last.connex=last.connex_2017-06-20     -1.24728963  9.963846e-04
## last.up.photo=last.up.photo_2011-03-10 -1.77543451  8.543630e-04
## last.connex=last.connex_2011-04-03     -1.78471250  8.543630e-04
## date.crea=date.crea_2011-03-06         -1.77776214  8.543630e-04
## last.connex=last.connex_2013-05-11     -1.29025804  6.599774e-04
## last.connex=last.connex_2012-03-09     -1.06023027  6.154696e-04
## last.up.photo=last.up.photo_2015-09-07 -1.28933142  6.083268e-04
```

```
## last.up.photo=last.up.photo_2018-01-14 -1.84399826  5.335313e-04
## last.connex=last.connex_2018-01-25     -1.85327625  5.335313e-04
## date.crea=date.crea_2018-01-05         -1.84632588  5.335313e-04
## last.up.photo=last.up.photo_2013-05-08 -1.84924620  5.143128e-04
## date.crea=date.crea_2013-04-25         -1.85157383  5.143128e-04
## last.connex=last.connex_2010-04-17     -1.31819432  5.015836e-04
## date.crea=date.crea_2011-01-10         -1.85817789  4.910445e-04
## last.connex=last.connex_2017-12-26     -1.32485747  4.694409e-04
## last.up.photo=last.up.photo_2019-10-01 -1.87218791  4.376300e-04
## last.connex=last.connex_2017-05-26     -1.90341159  3.743811e-04
## last.up.photo=last.up.photo_2014-01-05 -1.89623991  3.687825e-04
## last.connex=last.connex_2014-01-13     -1.90551789  3.687825e-04
## date.crea=date.crea_2013-11-30         -1.89856753  3.687825e-04
## last.connex=last.connex_2012-09-23     -1.95562623  2.565613e-04
## last.connex=last.connex_2013-07-31     -1.39222131  2.363322e-04
## last.connex=last.connex_2019-10-09     -1.39537713  2.286833e-04
## last.connex=last.connex_2016-10-07     -1.15809933  1.816212e-04
## last.up.photo=last.up.photo_2009-04-30 -2.01213573  1.572997e-04
## last.connex=last.connex_2009-05-04     -2.02141371  1.572997e-04
## date.crea=date.crea_2009-04-11         -2.01446335  1.572997e-04
## last.connex=last.connex_2020-02-08     -2.03412502  1.428712e-04
## date.crea=date.crea_2019-12-10         -2.02717466  1.428712e-04
## last.connex=last.connex_2010-10-13     -2.05904973  1.181207e-04
## last.connex=last.connex_2010-08-26     -2.06919025  1.092575e-04
## last.up.photo=last.up.photo_2017-09-13 -2.05991227  1.092575e-04
## last.connex=last.connex_2017-09-22     -2.06919025  1.092575e-04
## last.up.photo=last.up.photo_2010-04-20 -1.46215496  1.014010e-04
## last.up.photo=last.up.photo_2012-01-09 -2.07238053  9.921867e-05
## last.connex=last.connex_2012-02-02     -2.08165852  9.921867e-05
## date.crea=date.crea_2011-12-10         -2.07470815  9.921867e-05
## last.connex=last.connex_2009-03-07     -2.08165852  9.921867e-05
## last.up.photo=last.up.photo_2018-07-27 -2.08265919  9.160397e-05
## last.connex=last.connex_2018-07-26     -2.09193717  9.160397e-05
## date.crea=date.crea_2018-06-20         -2.08498681  9.160397e-05
## last.connex=last.connex_2017-06-27     -2.09193717  9.160397e-05
## last.connex=last.connex_2015-09-25     -2.11705689  7.525021e-05
## last.connex=last.connex_2015-09-21     -2.12247900  7.210254e-05
## date.crea=date.crea_2015-08-08         -2.11552864  7.210254e-05
## last.connex=last.connex_2019-12-01     -2.12376369  7.137520e-05
## last.up.photo=last.up.photo_2015-10-28 -1.50354375  6.407462e-05
## last.up.photo=last.up.photo_2013-03-17 -2.13449012  6.090724e-05
## last.connex=last.connex_2013-03-24     -2.14376811  6.090724e-05
## date.crea=date.crea_2013-02-20         -2.13681775  6.090724e-05
## gender=gender_1                        -0.06240029  6.090321e-05
## last.connex=last.connex_2018-03-22     -2.14768905  5.903359e-05
## last.up.photo=last.up.photo_2012-12-01 -2.14643116  5.536972e-05
## last.connex=last.connex_2012-12-07     -2.15570914  5.536972e-05
## last.connex=last.connex_2013-03-07     -2.15570914  5.536972e-05
## date.crea=date.crea_2012-12-25         -2.14875878  5.536972e-05
## last.up.photo=last.up.photo_2018-01-06 -2.23684010  2.648499e-05
## last.connex=last.connex_2018-01-14     -2.24611809  2.648499e-05
## last.up.photo=last.up.photo_2014-09-10 -2.24076105  2.563514e-05
## last.connex=last.connex_2014-10-17     -2.25003903  2.563514e-05
## date.crea=date.crea_2014-08-19         -2.24308867  2.563514e-05
## voyage=voyage_1                        -0.05023113  1.817112e-05
## last.connex=last.connex_2015-12-11     -2.30612202  1.598603e-05
## last.connex=last.connex_2014-01-19     -2.39418587  7.451406e-06
## date.crea=date.crea_2013-12-11         -2.38723551  7.451406e-06
## date.crea=date.crea_2017-11-01         -1.76576785  2.745429e-06
## last.connex=last.connex_2011-03-26     -1.51183132  9.755483e-07
## photo.beach=photo.beach_1              -0.06755702  1.628550e-07
## last.up.photo=last.up.photo_2017-11-05 -2.19574613  5.152552e-09
## n.matches=n.matches_45                 -1.70865757  1.023883e-10
## n.updates.photo=n.updates.photo_2      -0.52535504  1.296183e-11
## last.connex=last.connex_2017-11-11     -4.16258202  5.097033e-15
## n.photos=n.photos_12                   -3.93768099  5.097033e-15
## n.updates.photo=n.updates.photo_1      -0.78221734  1.194556e-16
## n.updates.photo=n.updates.photo_5      -0.22385348  5.767436e-20
## n.updates.photo=n.updates.photo_4      -0.28016906  5.889870e-34
## n.matches=n.matches_3                  -1.28208469  4.505729e-37
```

```
## n.updates.photo=n.updates.photo_3      -0.38154328  8.233998e-38
## n.matches=n.matches_5                 -0.86320795  3.619015e-67
## n.matches=n.matches_6                 -1.00575934  2.191916e-78
## n.matches=n.matches_4                 -1.22094058 7.851369e-108
## n.updates.photo=n.updates.photo_0      -1.24817152 1.768097e-139
## n.updates.photo=n.updates.photo_-1     -2.36949539 1.941374e-191
```

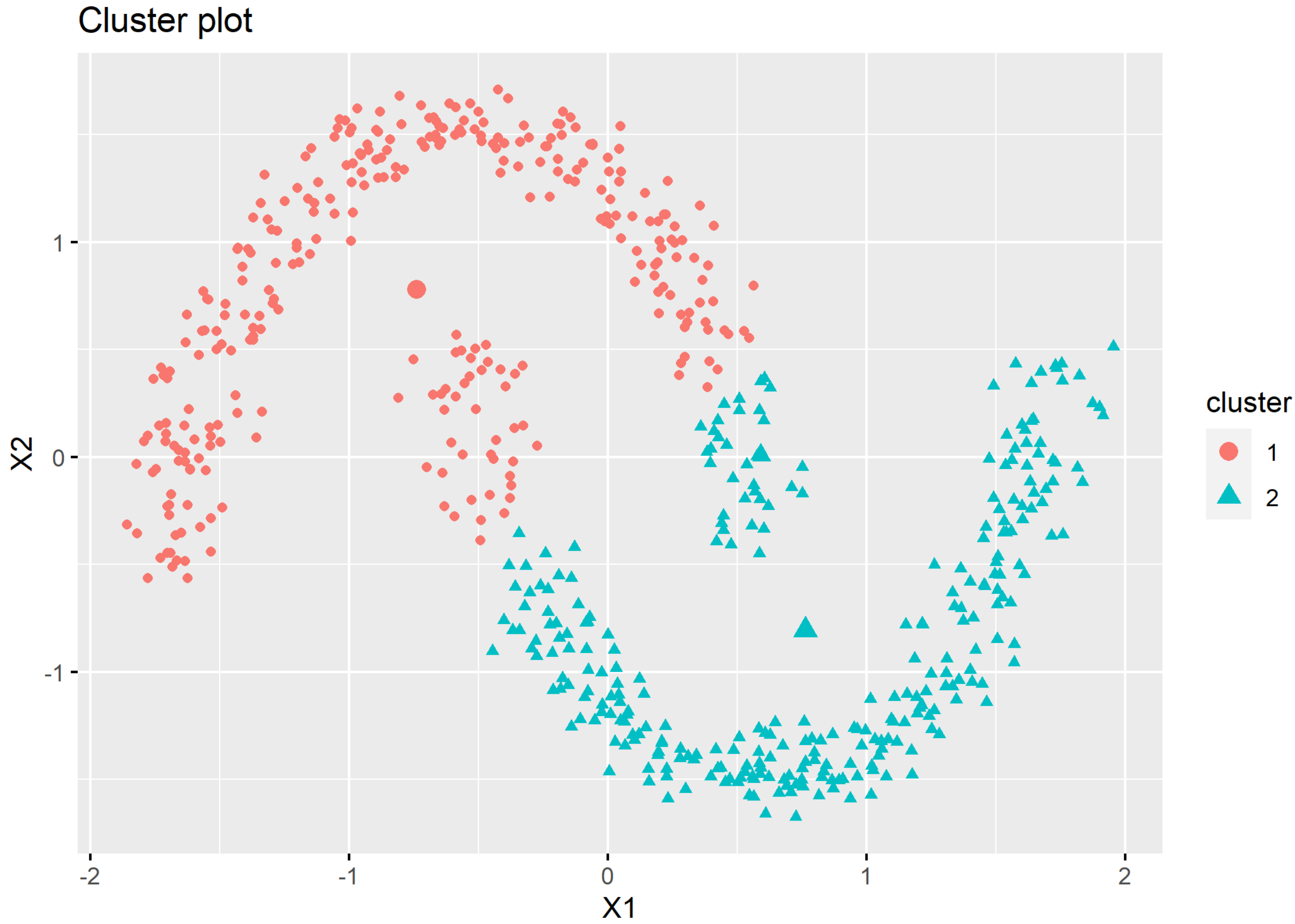
6 Projet 4

```
data <- RSSL::generateCrescentMoon(300, 2, 0.5)

df <- data[, 2:3]

ggplot(df, aes(X1, X2, col = data$Class)) +
  geom_point() +
  theme(legend.position = "none")
```



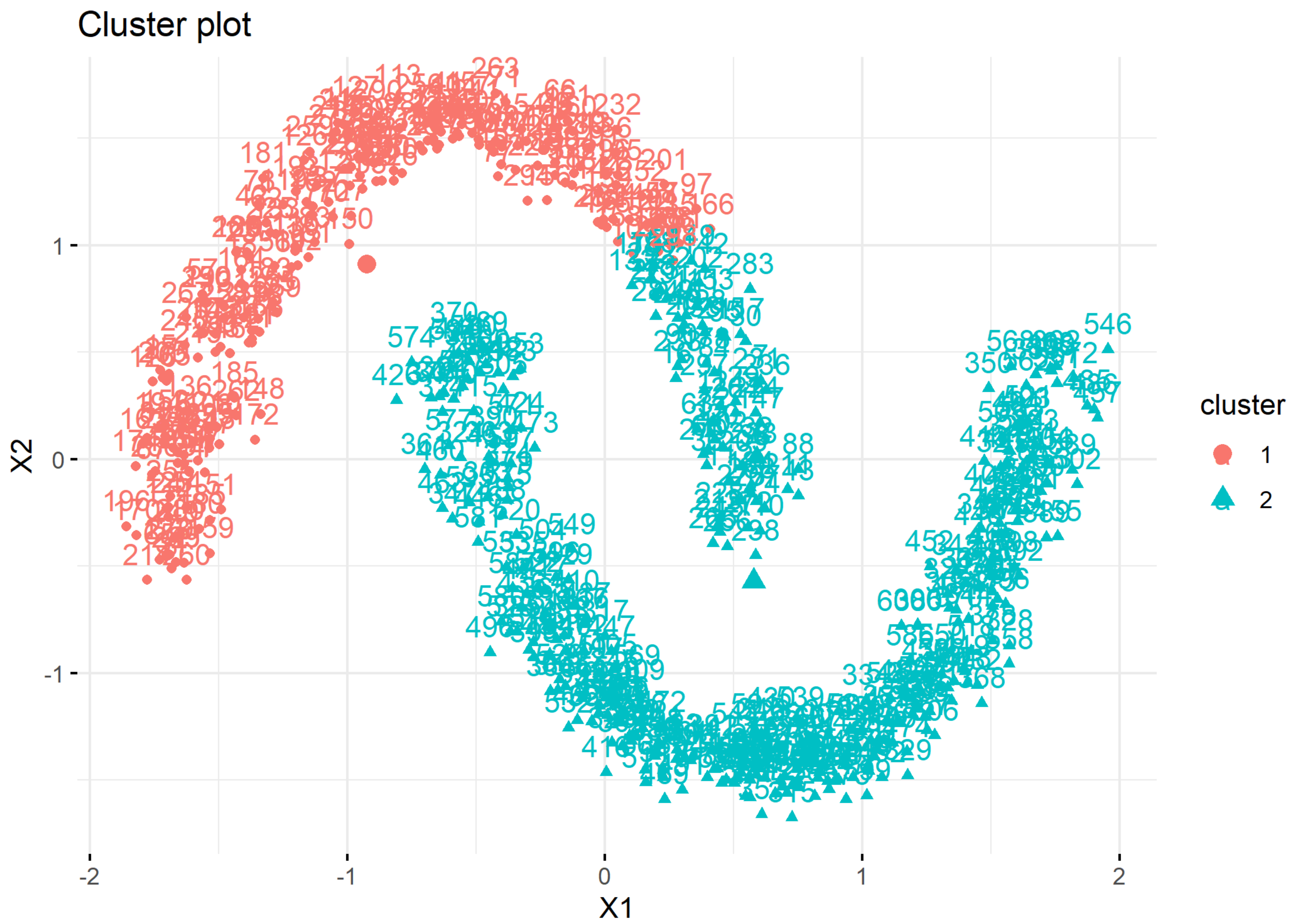
6.1 K-means

```
km.res <- kmeans(df, 2, nstart = 25)

fviz_cluster(km.res, df, ellipse = FALSE, geom = "point")
```



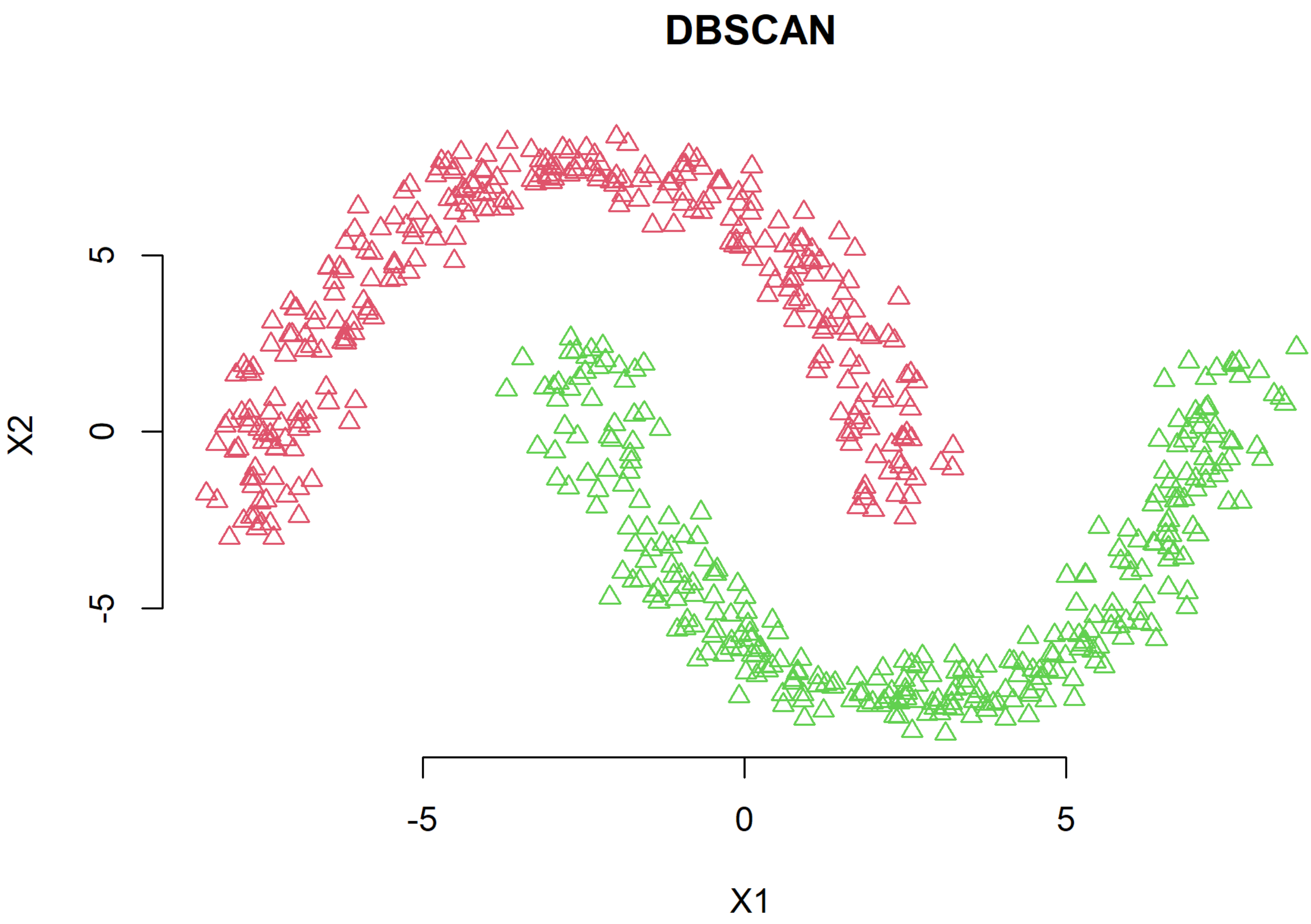
6.2 CAH

```
cah_hcut_data <- hcut(df)

fviz_cluster(cah_hcut_data, ggtheme = theme_minimal(), ellipse = FALSE
)
```



6.3 DBSCAN

```
db <- fpc::dbscan(df, eps = 1, MinPts = 2)

plot(db, df, main = "DBSCAN", frame = FALSE)
```


DBSCAN
X2
5
0
-5
X1
-5
0
5