



République tunisienne
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Carthage - École d'ingénieurs en statistiques et analyse de l'information



Projet en Analyse des Données

École Supérieure de la Statistique et de l'Analyse de l'Information



Présenté par

Nour Sfar

Explorer La Dépression

Février – May 2023

Encadrant :

Mr. Ghazi Bel Mufti

I. Introduction :

La dépression est un trouble de santé mentale fréquent qui affecte des millions de personnes à travers le monde. Cette maladie peut causer des symptômes tels que la tristesse persistante, la perte d'intérêt pour les activités quotidiennes, la fatigue, l'insomnie et l'anxiété. La dépression peut avoir des conséquences graves sur la vie des individus, notamment en affectant leurs relations interpersonnelles, leur capacité à travailler et leur bien-être généra... Dans ce cadre, 127 personnes, âgées de 17 ans au plus que 40 ans, ont été interrogés.

- Objectif : L'ACP (Analyse en composantes principales) et l'ACM (Analyse des Correspondances Multiples) sont des méthodes pertinentes pour l'analyse des données et la statistique multivariée. On va adopter ces méthodes pour analyser notre base de données. Puis faire une classification pour expliquer la carte des individus.

II. Découvrir la base de données :

La base de données "Explorer La Dépression" est construite à partir d'un questionnaire destiné à des individus ayant des différentes catégories socio-professionnelles et plusieurs classes d'âge. Les données contiennent principalement. Le questionnaire, entièrement anonymisé, était décomposé en différentes questions :

- Colonne1 : est une variable qualitative qui représente le sexe.
- Colonne2 : est une variable qualitative ordinale qui représente les classes d'âge.
- Colonne 3 à 5 : ce sont des variables qualitatives qui représentent la classe socio-professionnelle, type de lycée (pour les lycéens) et filière d'étude (pour les étudiants).

Le reste des colonnes sont partitionné en quatre blocs :

- Deux blocs de variables ont été obtenus à partir de questions à réponse binaire (oui/non).
- Deux blocs de variables comportant soit des symptômes, soit des habitudes ont été obtenus à partir d'une échelle de Likert graduée de 1 à 5.

Le tableau de données contient 127 observations et 45 colonnes.

1. Importer et préparer le jeu de données :

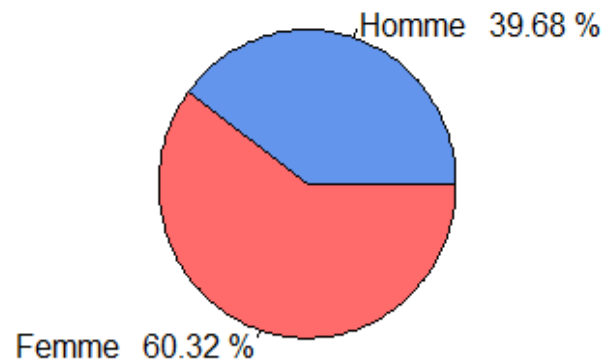
```
df<-read_csv("Explorer La Dépression.csv", show_col_types = FALSE)
```

2. Analyse Descriptive des Données :

2.1. Répartition par genre:

```
df$genre <- factor(df$genre, levels = c('Homme', 'Femme'))  
  
#Camembert Homme/Femme  
df['genre'] %>%  
  group_by(genre) %>%  
  summarize(count = n()) -> d  
pct = d$count / sum(d$count) * 100  
p2 <- pie(d$count, labels = paste(d$genre, ' ', round(pct, 2), '%'), col =  
c('cornflowerblue', 'indianred1'),  
  main = 'Distribution des Genres')
```

Distribution des Genres



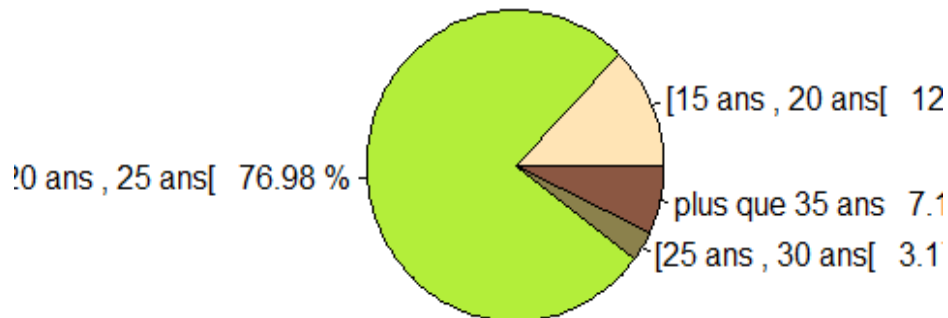
Interprétation :

- A l'issue de ce questionnaire, on a 60.32% des réponses sont des femmes (76 femmes), 39.68% sont des Hommes (50 hommes).

2.2. Répartition par âge :

```
#Camembert âges
df[ 'âge' ] %>%
  group_by(âge) %>%
  summarize(count = n()) -> d
pct = d$count / sum(d$count) * 100
p4 <- pie(d$count, labels = paste(d$âge, ' ', round(pct, 2), '%'), col =
c('moccasin', 'olivedrab2', 'lightgoldenrod4', 'lightsalmon4'),
  main = 'Distribution des âges')
```

Distribution des âges



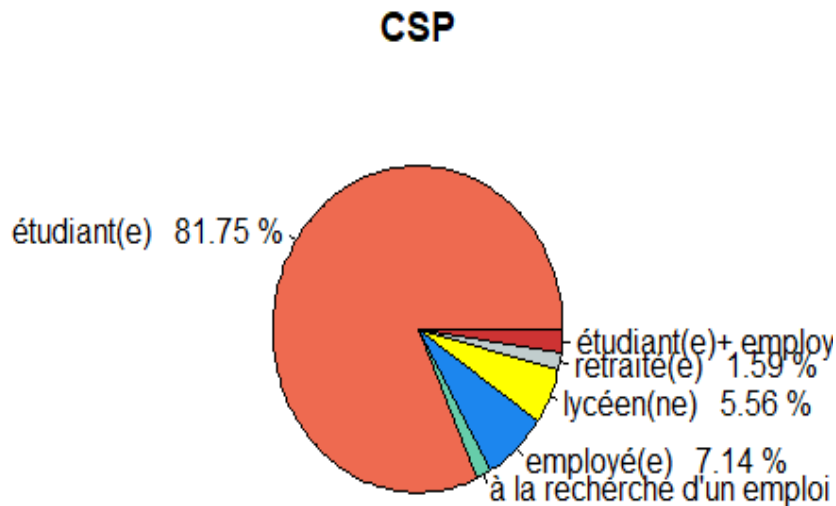
Interprétation :

- On remarque que la tranche d'âge majoritaire dans notre échantillon est du 20 ans à 25 ans.

2.3. Répartition par catégorie socioprofessionnelle :

```
#Camembert CSP
df[ 'CSP' ] %>%
  group_by(CSP) %>%
  summarize(count = n()) -> d
pct = d$count / sum(d$count) * 100
```

```
p6 <- pie(d$count, labels = paste(d$CSP, ' ', round(pct, 2), '%'), col = c('coral2', 'aquamarine3', 'dodgerblue2', 'yellow', 'azure3', 'brown3', 'gray43'),
        main = 'CSP')
```

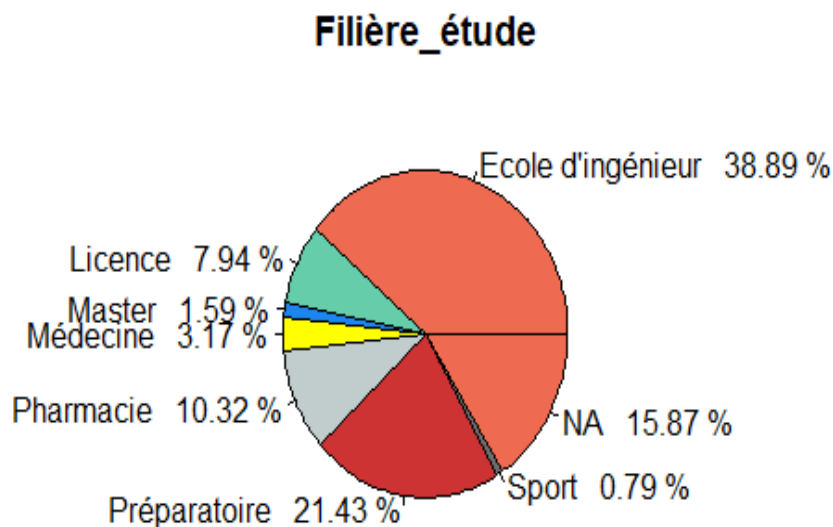


Interprétation :

- Les Etudiants sont les majoritaires dans notre échantillon avec un ratio qui dépasse les 80% des individus de l'échantillon.

2.4. Répartition des étudiants par filière d'étude :

```
#Camembert Filière_étude
df['Filière_étude'] %>%
  group_by(Filière_étude) %>%
  summarize(count = n()) -> d
pct = d$count / sum(d$count) * 100
p8 <- pie(d$count, labels = paste(d$Filière_étude, ' ', round(pct, 2), '%'), col =
c('coral2', 'aquamarine3', 'dodgerblue2', 'yellow', 'azure3', 'brown3', 'gray43'),
        main = 'Filière_étude')
```

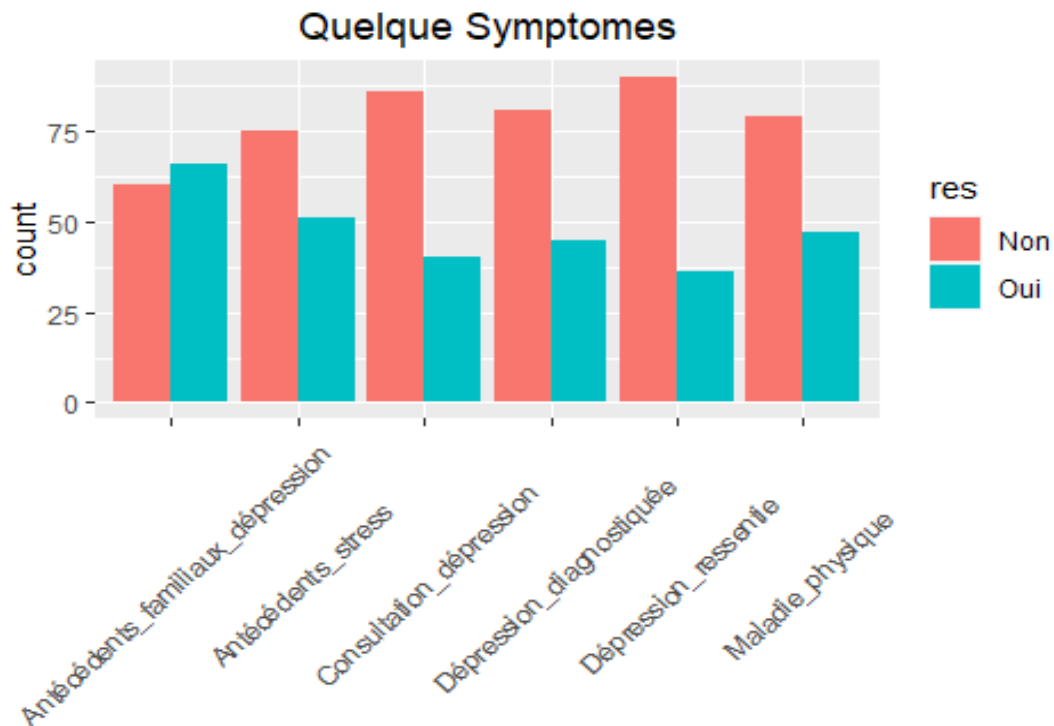


Interprétation :

- On remarque que la filière qui persiste dans notre échantillon c'est le cycle ingénieur (Préparatoire + Ecole d'ING) avec un ratio qui dépasse les 60% de l'échantillon.
- 15.87% des individus de notre échantillon ne sont pas des étudiants.

2.5. Quelques Comportement des individus :

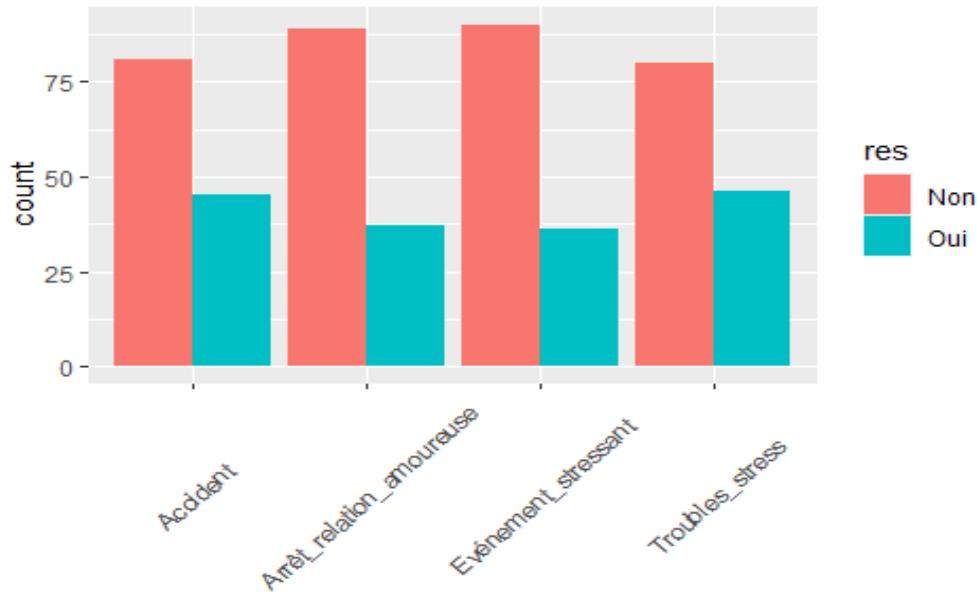
```
l1 <- list(medium = rep(c("Dépression_ressentie", "Dépression_diagnostiquée",
"Consultation_dépression", "Antécédents_familiaux_dépression",
"Maladie_physique", "Antécédents_stress"),
each = length(df[,1])), res = c(df$Dépression_ressentie,
df$Dépression_diagnostiquée, df$Consultation_dépression,
df$Antécédents_familiaux_dépression, df$Maladie_physique, df$Antécédents_stress))
data1 <- data.frame(l1)
ggplot(data1, aes(medium, fill = res)) +
  geom_bar(position = 'dodge') +
  labs(title = 'Quelque Symptomes', x = 'medium') +
  theme(plot.title = element_text(hjust = 0.5),
axis.text.x = element_text(angle = 45, vjust = 0.5)
) -> p9
```



2.6. Quelques événements peuvent arriver aux individus :

```
l2 <- list(medium = rep(c("Arrêt_relation_amoureuse", "Accident",
"Événement_stressant", "Troubles_stress"),
each = length(df[,1])), res = c(df$Arrêt_relation_amoureuse,
df$Accident, df$Événement_stressant, df$Troubles_stress))
data2 <- data.frame(l2)
ggplot(data2, aes(medium, fill = res)) +
  geom_bar(position = 'dodge') +
  labs(title = 'Quelques événements peuvent arriver aux individus', x = 'medium') +
  theme(plot.title = element_text(hjust = 0.5),
axis.text.x = element_text(angle = 45, vjust = 0.5)
) -> p10
p10
```

Quelques événements peuvent arriver aux individus



III. ACP et son interprétation :

- L'ACP est une technique d'analyse multivariée qui permet de réduire la dimensionnalité d'un ensemble de variables tout en préservant au maximum l'information qu'elles contiennent.
- Dans notre cas l'ACP pourrait être utilisée pour identifier les variables qui sont les plus importantes pour décrire les symptômes de la dépression et pour trouver des groupes de symptômes qui se regroupent souvent ensemble.

1. But de l'ACP :

- Variables : Regrouper les variables corrélées et les remplacer par un petit nombre de nouvelles variables qui sont les composantes principales décrivant symptômes ou les causes de la dépression.
- Individus : Regrouper les individus qui ont les mêmes symptômes ou mêmes habitudes et indiquer les symptômes et les habitudes qui les caractérisent.

Remarque :

- On va ajouter les variables "sexe" et "âge" comme variables qualitatives supplémentaires. Les variables illustratives n'influencent pas la construction des composantes principales de l'analyse.

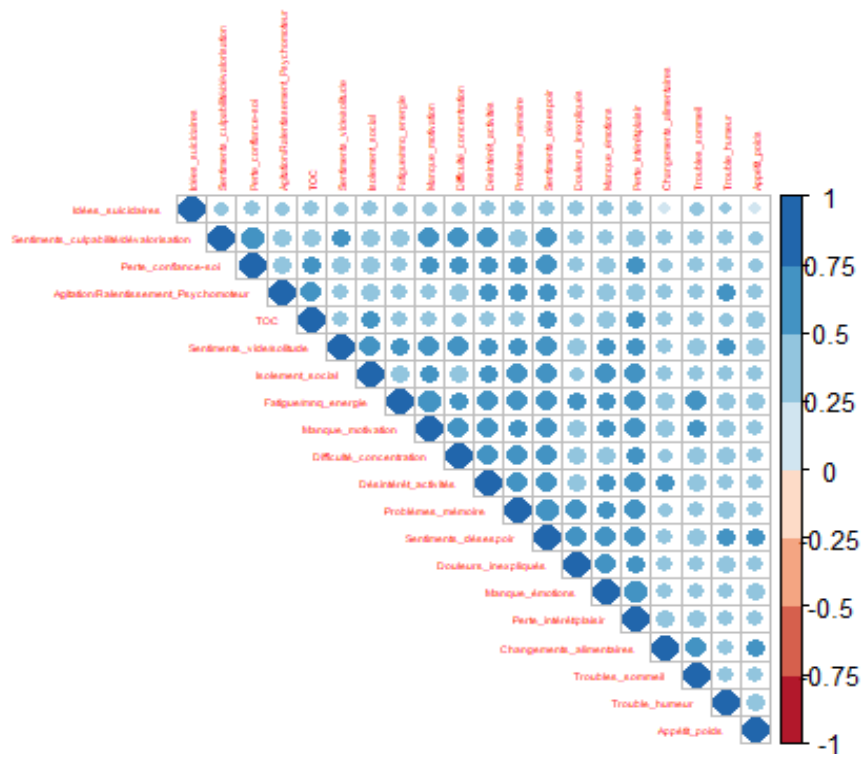
2. Pertinence de l'ACP :

2.1. Critère 1 : Le Corrplot :

- Le corrogramme donné ci-dessous permet d'étudier les corrélations entre les variables quantitatives.

```
m1<-cor(bloc1)
```

```
correlations <- corplot(m1, type="upper", order="hclust",  
                        col=brewer.pal(n=8, name="RdBu"),tl.cex = 0.3)
```



```
res.pca=PCA(bloc1, ncp = 10, graph = F)
```

Interprétation :

- Il est clair qu'il existe des corrélations importantes entre des groupes de variables ce qui suggère la pertinence de cette ACP.

2.2. Critère 2 : Les Valeurs Propres :

```
val_prop <- res.pca$eig
head(val_prop)
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
## comp 1 10.2116342          51.058171          51.05817
## comp 2  1.2022522           6.011261          57.06943
## comp 3  1.0945357           5.472678          62.54211
## comp 4  0.9624822           4.812411          67.35452
## comp 5  0.7663093           3.831547          71.18607
## comp 6  0.7298598           3.649299          74.83537
```

Interprétation :

- Une valeur très élevée a été obtenue pour la première valeur propre.

3. Choix du nombre d'axes à retenir :

3.1 Critère de Kaiser :

```
val_prop[,1]
```

```
##      comp 1      comp 2      comp 3      comp 4      comp 5      comp 6      comp 7
## 10.2116342  1.2022522  1.0945357  0.9624822  0.7663093  0.7298598  0.6805018
##      comp 8      comp 9      comp 10     comp 11     comp 12     comp 13     comp 14
##  0.5892378  0.5173245  0.4961391  0.4303420  0.3834948  0.3520654  0.3314169
##      comp 15     comp 16     comp 17     comp 18     comp 19     comp 20
##  0.2668980  0.2453968  0.2251123  0.1893941  0.1666464  0.1589565
```

Interprétation :

- On remarque qu'il y a 3 axes dont les valeurs propres sont supérieures à 1 donc 2 on retient 2 axes d'après le critère de Kaiser.

3.2 Critère du Taux d'Inertie Cumulée :

```
val_prop[,3]
```

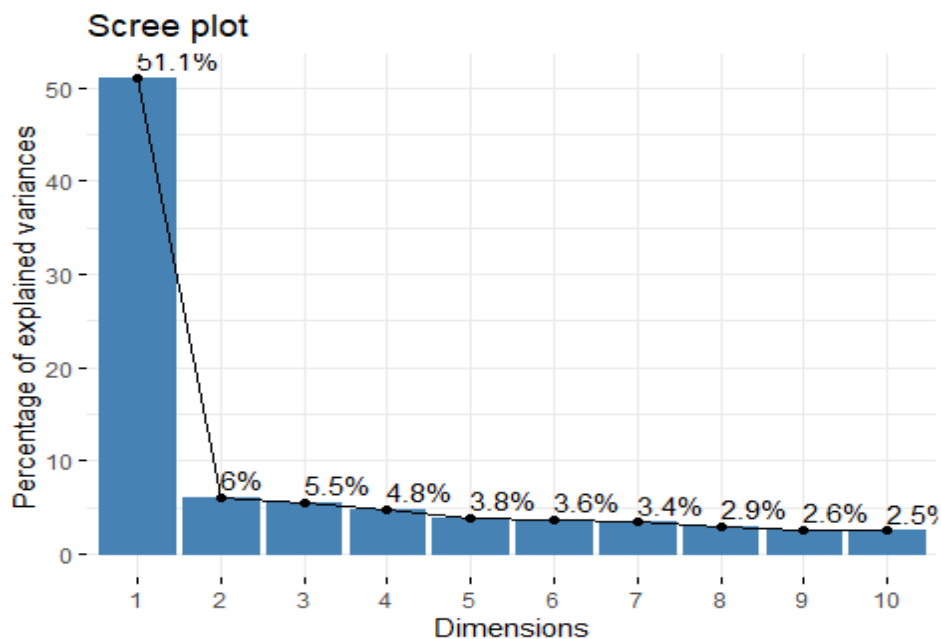
```
##   comp 1   comp 2   comp 3   comp 4   comp 5   comp 6   comp 7   comp 8
## 51.05817 57.06943 62.54211 67.35452 71.18607 74.83537 78.23788 81.18407
##   comp 9   comp 10   comp 11   comp 12   comp 13   comp 14   comp 15   comp 16
## 83.77069 86.25138 88.40309 90.32057 92.08089 93.73798 95.07247 96.29945
##   comp 17   comp 18   comp 19   comp 20
## 97.42501 98.37199 99.20522 100.00000
```

Interprétation :

- On remarque que le taux d'inertie cumulé des deux premiers axes est de 57.06943% qui est un taux important compte tenu du fait que nous avons 20 variables : on va donc, d'après le critère du taux d'inertie cumulé, retenir les 2 premiers axes.

3.3 Critère du Coude :

```
fviz_screepLOT(res.pca,addlabels=TRUE)
```



Interprétation :

- On remarque que le coude se trouve au niveau du deuxième axe, d'après le Critère du Coude, on devrait retenir les 2 premiers axes.

3.4. Axes à retenir :

- On peut conclure par vote des 3 critères qu'on devrait retenir les 2 premiers axes.

4. Interprétation de la carte des variables :

4.1. Coordonnées des variables :

- L'objet var de res.pca contient les 4 objets : coord, cor, cos2 et contrib.

- A noter que vu que notre ACP est normée, cor (i.e. la corrélations d'une variable avec la composante principale d'un axe) est identique à coord (i.e. la coordonnée de cette variable sur cet axe).

```
names(res.pca$var)
```

```
## [1] "coord" "cor" "cos2" "contrib"
```

- L'objet coord dans var contient les coordonnées des variables.

```
res.pca$var$coord[,1:3]
```

##	Dim.1	Dim.2	Dim.3
## Changements_alimentaires	0.6192530	0.458416825	0.25613281
## Troubles_sommeil	0.6756491	0.344732949	0.11026914
## Fatigue/mnq_energie	0.7780426	0.247502409	-0.14275928
## Manque_motivation	0.7842850	0.057647175	-0.29590996
## Difficulté_concentration	0.7475262	-0.049331685	-0.38829589
## Sentiments_culpabilité/dévalorisation	0.7199256	-0.282408932	-0.10301719
## Perte_confiance-soi	0.7100421	-0.416248810	0.01697140
## Idées_suicidaires	0.5000128	-0.400099580	0.22122254
## Trouble_humeur	0.6376272	0.316998646	-0.01576050
## Agitation/Ralentissement_Psychomoteur	0.6812892	-0.057290846	0.34509613
## Sentiments_vide/solitude	0.7515305	0.064088155	-0.15469836
## Désintérêt_activités	0.7982838	-0.041299397	-0.13795917
## Douleurs_inexpliqués	0.6687030	0.127905321	-0.20229106
## Problèmes_mémoire	0.7730520	-0.165983066	-0.20834087
## Sentiments_désespoir	0.8491699	-0.100774088	-0.07671791
## Manque_émotions	0.7468186	-0.003043481	0.07381884
## Appétit_poids	0.6172891	0.383297714	0.28642046
## Isolement_social	0.7137570	-0.176784159	0.12093041
## TOC	0.6355906	-0.222024513	0.58536257
## Perte_intérêt/plaisir	0.7931549	-0.045064495	0.03064487

- L'objet cos2 dans var est une matrice dont les lignes représentent le cos carrés de la variable (soit le carrée des coordonnées puisque l'ACP est normée).

```
res.pca$var$cos2[,1:3]
```

##	Dim.1	Dim.2	Dim.3
## Changements_alimentaires	0.3834742	2.101460e-01	0.0656040159
## Troubles_sommeil	0.4565017	1.188408e-01	0.0121592838
## Fatigue/mnq_energie	0.6053504	6.125744e-02	0.0203802127
## Manque_motivation	0.6151029	3.323197e-03	0.0875627032
## Difficulté_concentration	0.5587955	2.433615e-03	0.1507736964
## Sentiments_culpabilité/dévalorisation	0.5182929	7.975480e-02	0.0106125422
## Perte_confiance-soi	0.5041598	1.732631e-01	0.0002880285
## Idées_suicidaires	0.2500128	1.600797e-01	0.0489394120
## Trouble_humeur	0.4065684	1.004881e-01	0.0002483932
## Agitation/Ralentissement_Psychomoteur	0.4641550	3.282241e-03	0.1190913402
## Sentiments_vide/solitude	0.5647980	4.107292e-03	0.0239315837
## Désintérêt_activités	0.6372571	1.705640e-03	0.0190327337
## Douleurs_inexpliqués	0.4471637	1.635977e-02	0.0409216724
## Problèmes_mémoire	0.5976094	2.755038e-02	0.0434059175
## Sentiments_désespoir	0.7210894	1.015542e-02	0.0058856374
## Manque_émotions	0.5577380	9.262775e-06	0.0054492215
## Appétit_poids	0.3810459	1.469171e-01	0.0820366790
## Isolement_social	0.5094490	3.125264e-02	0.0146241643

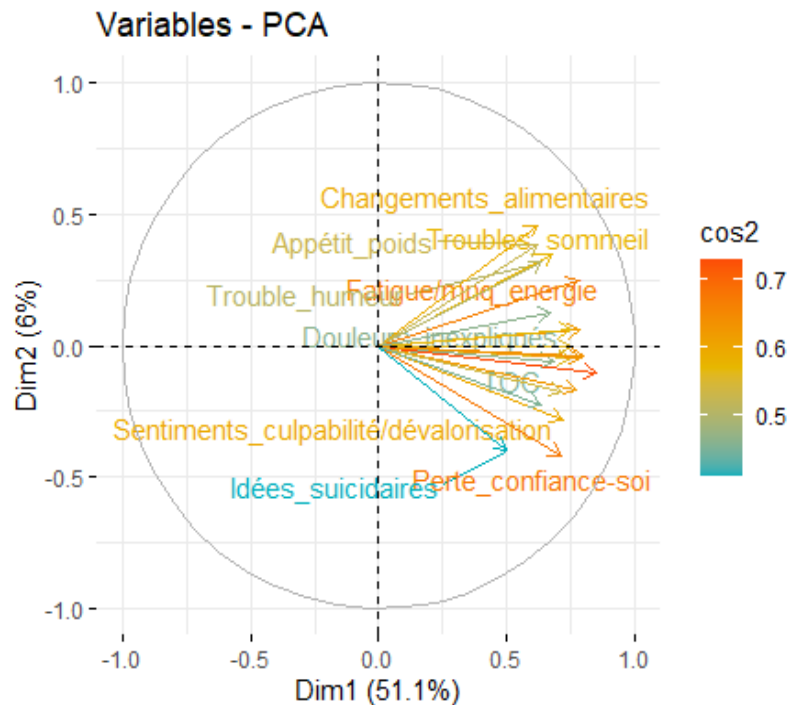
```
## TOC 0.4039754 4.929488e-02 0.3426493347
## Perte_intérêt/plaisir 0.6290947 2.030809e-03 0.0009391080
```

4.2. Représentation de la Carte des variables :

- La qualité de la représentation dépend de la grandeur cosinus carré, plus il est élevé plus on dit qu'un axe est bien représenté.

```
fviz_pca_var(res.pca, col.var = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = T)
```

```
## Warning: ggrepel: 10 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



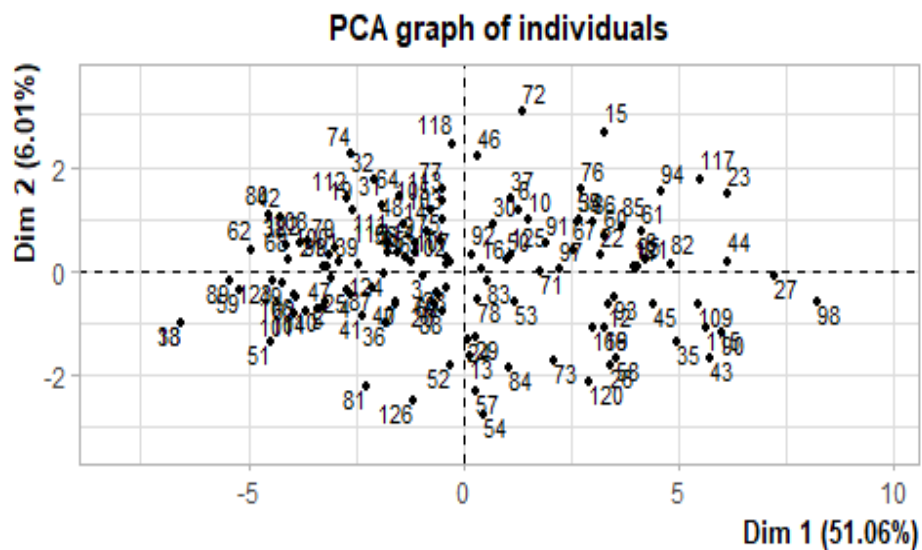
Interprétation :

- Les variables à faible valeur de cos2 seront colorées en "blanc".
- Les variables avec les valeurs moyennes de cos2 seront colorées en "Blue".
- Les variables avec des valeurs élevées de cos2 seront colorées en "rouge".
- Exemple : Les \cos^2 de « **Sentiments_désespoir** » et « **Perte_confiance-Soi** » sont élevés ce qui indique une bonne représentation des deux variables sur les axes principaux. Dans ce cas, ces variables sont positionnées à proximité de la circonférence du cercle de corrélation tandis que le \cos^2 de « **Idées_Suicidaires** » est faible par rapport aux autres variables, d'où la variable est un peu proche du centre du cercle.
- La variable **{Trouble_humeur}** est positivement corrélée à la variable **{Appétit_poids}**. L'individu qui des troubles d'humeur est susceptible de d'augmenter ou diminuer son appétit et son poids.
- La variable **{Douleurs_inexpliqués}** est positivement corrélée a la variable **{TOC}**. L'individu qui a des troubles obsessionnels-compulsifs susceptible d'avoir des douleurs inexpliquées.

- Axe 1 : symptômes de la dépression.
- Axe 2 : Facteurs de la dépression.

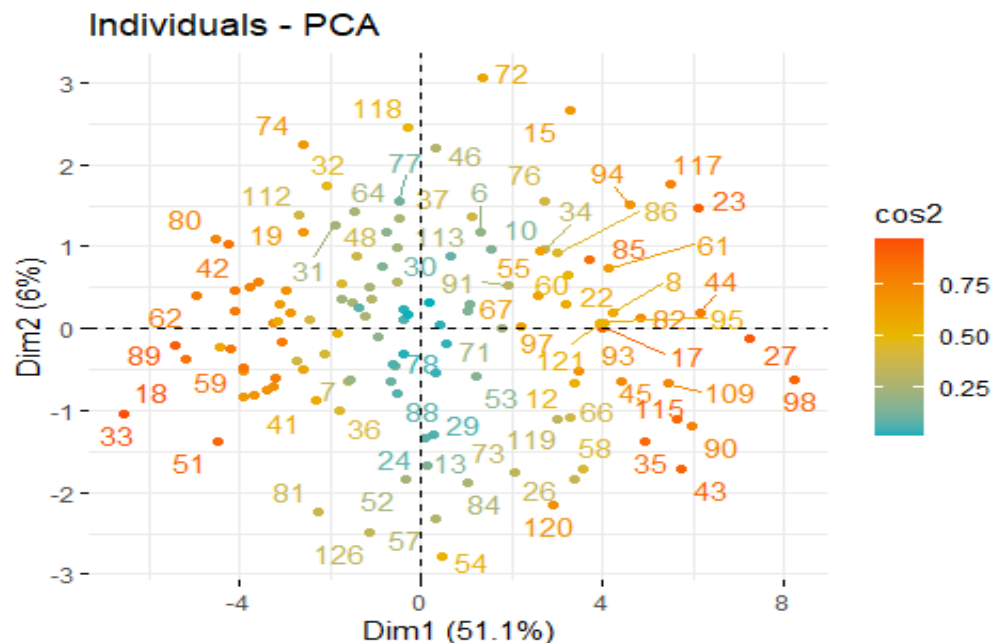
- ## 5. Interprétation de la carte des individus :

- ```
res.pca = PCA(bloc1, graph=F)
plot(res.pca, choix="ind", cex=0.7)
```



```
quali.sup <- list(sexe = df$genre)
fviz_pca_ind(res.pca, col.ind = "cos2",
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
 repel = T, quali.sup = quali.sup)

Warning: ggrepel: 49 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



#### IV. ACM et son interprétation :

- L'ACM, ou Analyse en Composantes Multiples, est également une technique d'analyse multivariée utilisée pour explorer les relations entre plusieurs variables. Contrairement à l'ACP, l'ACM est utilisée lorsque les variables sont catégorielles et non continues.
- Dans notre contexte, l'objectif de l'Analyse en Composantes Multiples (ACM) pourrait être d'explorer les relations entre nos plusieurs variables catégorielles liées à la dépression, telles que l'âge, le sexe, le niveau d'éducation, la CSP, etc. En appliquant l'ACM, nous pouvons identifier les combinaisons de variables qui sont les plus fréquentes et voir comment elles sont associées à la dépression. Cela peut nous aider à mieux comprendre les facteurs de risque et les symptômes de la dépression chez différents groupes de personnes, ainsi qu'à identifier les tendances et les schémas généraux liés à la dépression.

##### 1. But de l'ACM :

Résumer et visualiser le tableau de données contenant 6 tranches d'âge, 2 sexes et 6 types de CSP. 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.

```
bloc3 <- df[, 7:12]
bloc3
```

```
A tibble: 126 × 6
Dépression_ressentie Dépression_diagnostiquée Consu...1 Antéc...2 Malad...3 Antéc...4
<chr> <chr> <chr> <chr> <chr> <chr>
1 Oui Non Non Non Non Non
2 Oui Non Non Non Non Non
3 Oui Non Non Non Non Oui
4 Non Non Non Non Non Oui
```

```
5 Non Non Non Non Non Non
6 Oui Oui Oui Non Oui Oui
7 Non Non Non Non Non Non
8 Oui Non Non Oui Non Oui
9 Oui Non Non Non Non Oui
10 Oui Oui Non Oui Non Oui
... with 116 more rows, and abbreviated variable names
^1Consultation_dépression, ^2Antécédents_familiaux_dépression,
^3Maladie_physique, ^4Antécédents_stress

res.mca <- MCA (bloc3, graph = FALSE)
```

## 2. Visualisation et interprétation :

```
res.mca$eig

eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.32479709 32.479709 32.47971
dim 2 0.20289368 20.289368 52.76908
dim 3 0.15939515 15.939515 68.70859
dim 4 0.11206776 11.206776 79.91537
dim 5 0.10499360 10.499360 90.41473
dim 6 0.09585272 9.585272 100.00000
```

- $m = 6$  est le nombre de variables et  $p = 2$  est le nombre de modalités alors l'inertie totale de cette ACM est égale à 0.33.

## 3. Nombre d'axes à retenir :

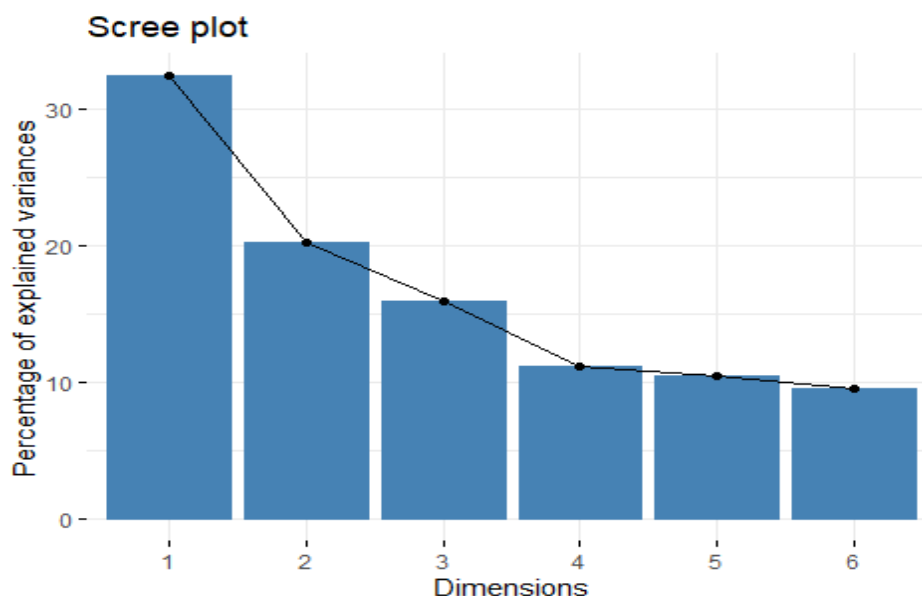
### 3.1. Critère 1 :

- Retenir les axes dont les valeurs propres sont supérieures à la moyenne soit  $1/m = 0.166$  donc les 2 premiers axes.

### 3.2. Critère 2 :

- Tracer le graphique des valeurs propres et noter que le coude se trouve au niveau du 2ème axe.

```
fviz_screepplot(res.mca)
```



### 3.2. Critère 3 :

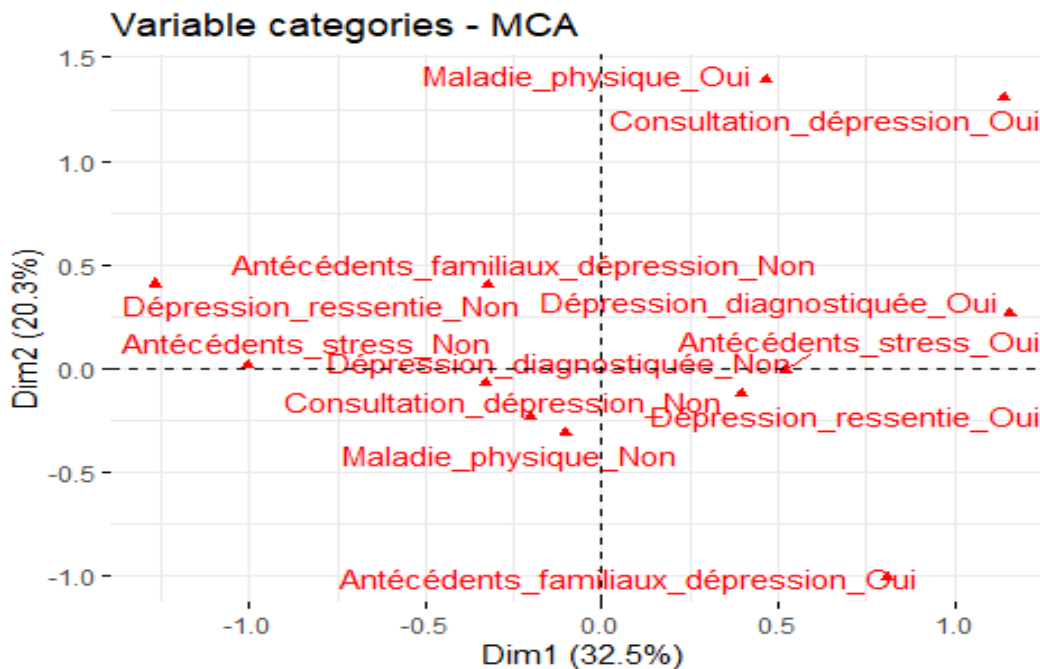
- Raisonner sur le taux d'inertie cumulé et retenir les 2 premiers axes qui, sur les 9 axes de l'ACM, accumulent quasi la moitié de l'inertie.

**En conclusion, le choix des 2 premiers axes serait justifié.**

### 4. Interprétation de la première carte des modalités :

*#Représentation de la première carte des modalités*

```
fviz_mca_var (res.mca,
 repel = TRUE,
 ggtheme = theme_minimal ())
```



#### Interprétation :

- Il suffit d'identifier les groupes de modalités qui sont proches sur la carte.
- On a un premier groupe de modalités à gauche de la carte, constitué des modalités suivantes: Dépression\_ressentie\_Non et Antécédents\_stress\_Non. Ces modalités peuvent être interprétées comme indiquant les individus qui ne souffrent pas de dépression.
- Un deuxième groupe de modalités se trouve à droite de la carte et est constitué des modalités suivantes : Dépression\_diagnostiquée\_oui, antécédent\_Stress\_oui, dépression\_ressentie\_oui. Ces modalités peuvent être interprétées comme indiquant les individus qui souffrent de dépression.
- Un troisième groupe se trouve en haut de la carte et est constitué des modalités suivantes : Maladie\_physique et Consultation\_dépression. Ces modalités peuvent être interprétées comme indiquant les individus qui souffrent d'une maladie physique et ont tendance à consulter un psychiatre pour évaluer leur situation psychique, peut-être en raison de l'impact de la maladie physique sur leur santé mentale ou de la prise de médicaments associés à cette maladie.

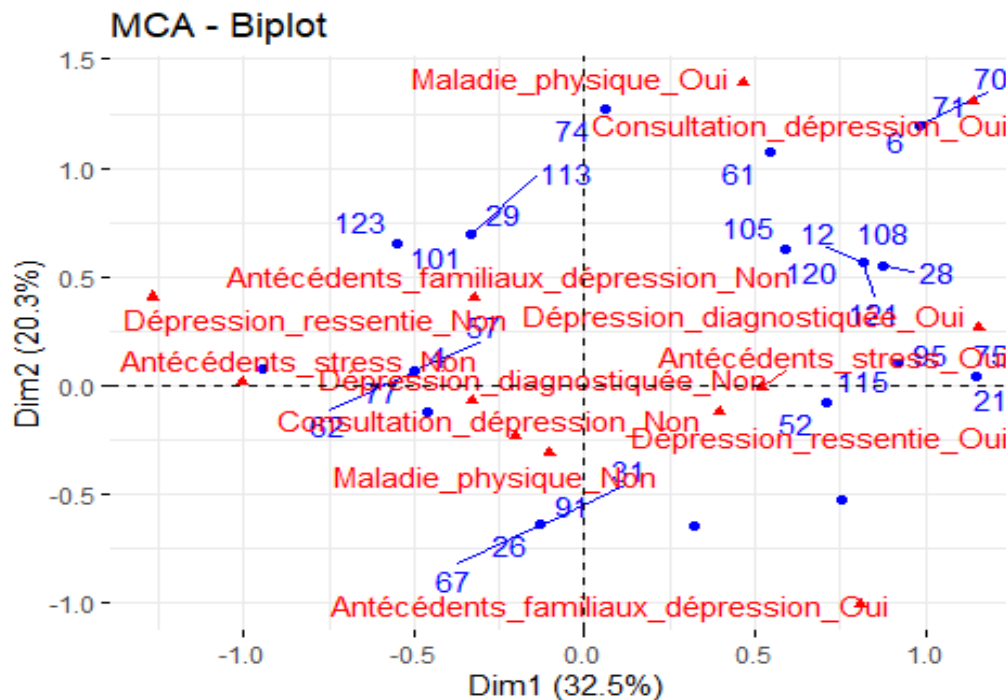
```
dimdesc(res.mca, axes=1:2, proba=0.05)$`Dim 1`
```

## 5. Interprétation de la première carte des individus :

```
#Représentation de la première carte des individus
```

```
fviz_mca_biplot (res.mca,select.ind = list(cos2 = 0.3),
 repel = TRUE,
 ggtheme = theme_minimal ())
```

```
Warning: ggrepel: 53 unlabeled data points (too many overlaps). Consider
increasing max.overlaps
```



## V. Classification :

## 1. Méthode Hiérarchique :

- Les méthodes non hiérarchiques permettent de traiter des populations importantes (même 1000 et plus) à des coûts raisonnables. Ces méthodes visent à constituer directement  $k$  types à partir de  $n$  objets en essayant d'optimiser un indice global mesurant la qualité de la classification.

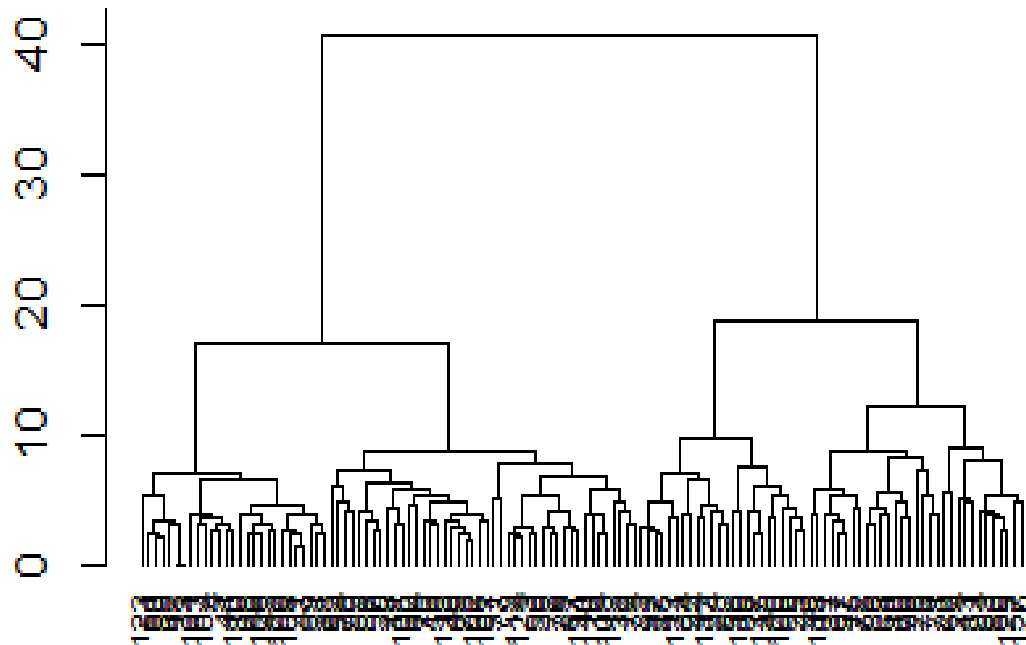
### 1.1. Arbre Hiérarchique :

- L'arbre hiérarchique est un outil d'analyse de données qui permet de regrouper des variables ou des individus en fonction de leurs similarités ou de leurs différences. Sur une carte factorielle, qui représente les résultats d'une analyse en composantes principales (ACP).

### 1.1.1. Afficher l'arbre :

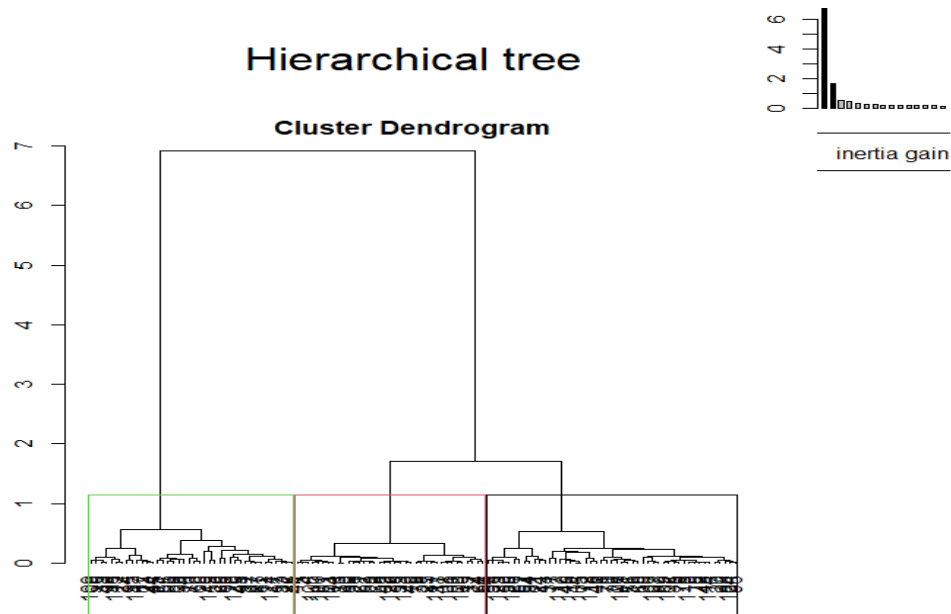
```
c<-dist(scale(bloc1),method="euclidean")
h<- hclust(c, method="ward.D2")
plot(h, hang = -1, cex =0.6)
```

# Cluster Dendrogram



## 1.1.2. Choisir Les Classes :

```
res.HCPC<-HCPC(res.pca, consol=TRUE, graph=F)
plot.HCPC(res.HCPC,choice='tree', title ='Hierarchical tree')
```

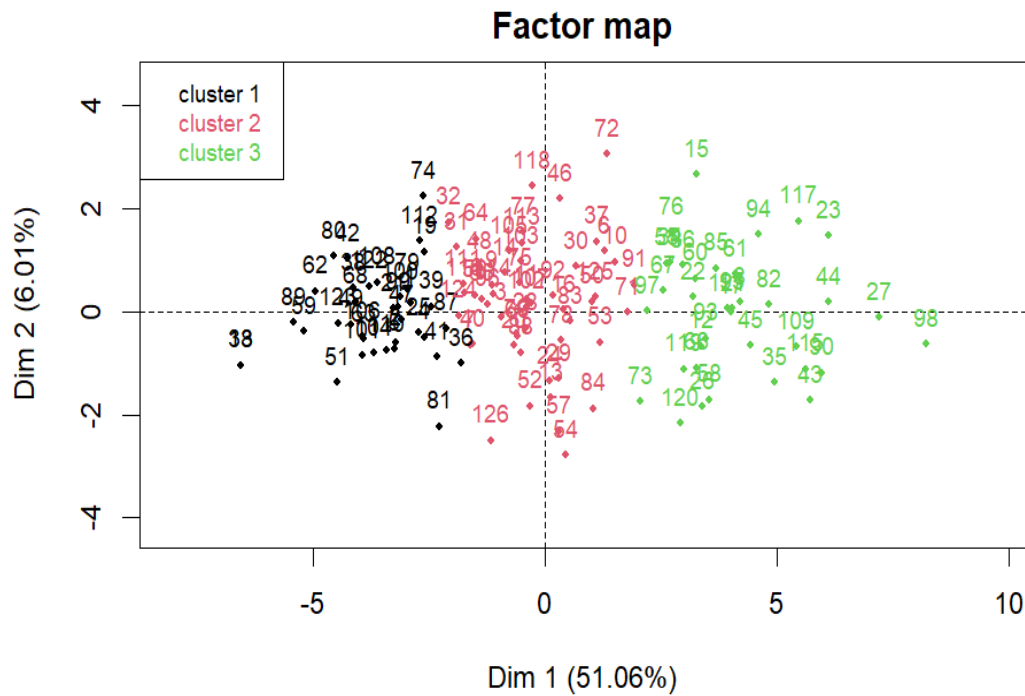


**Conclusion :** on peut choisir 3 classes.

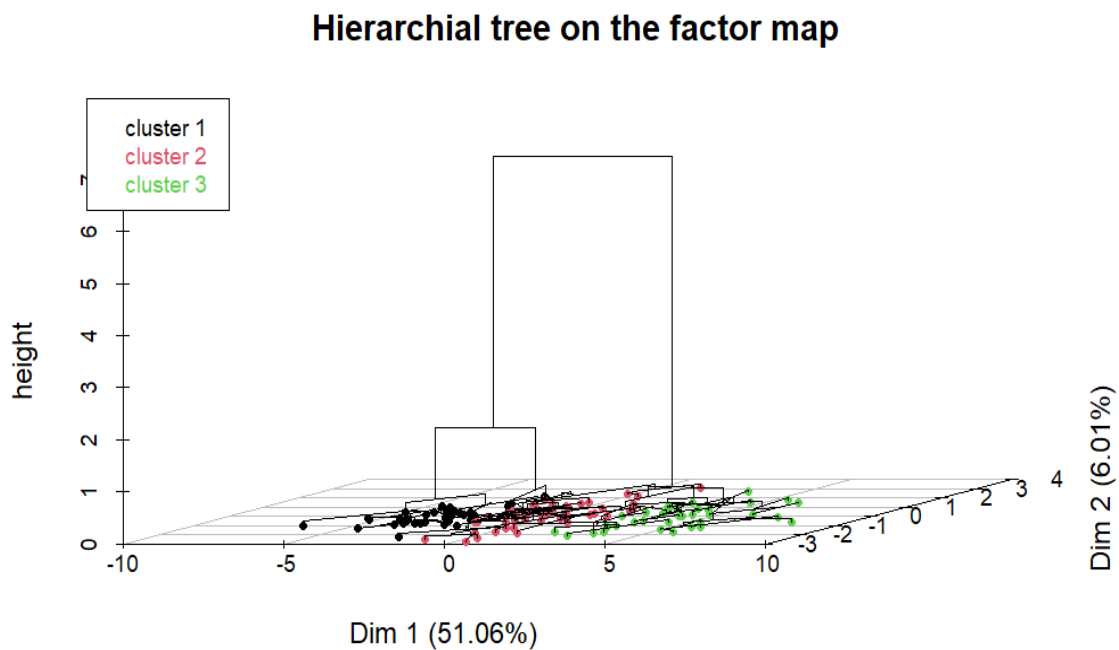
## 1.2. Arbre hiérarchique sur la carte factorielle :

```
plot.HCPC(res.HCPC, choice = 'map', draw.tree=F, title ='Factor map')
```





```
plot.HCPC(res.HCPC,choice = '3D.map', ind.names = FALSE, title='Hierarchial tree on the factor map')
```

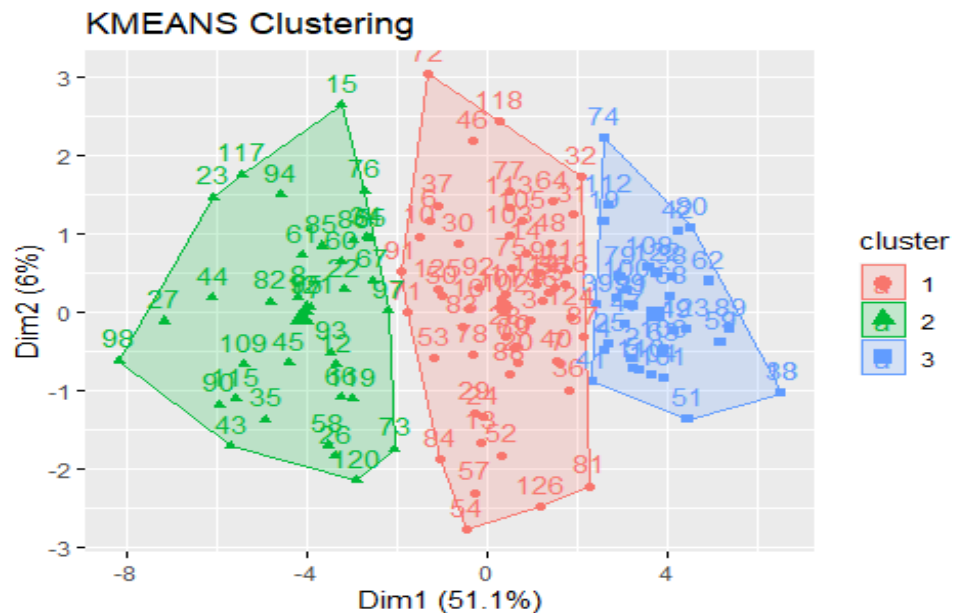


## 2. Méthode non Hiérarchique :

### 2.1. K-means :

Ci-dessous, nous exposons comment effectuer une coloration par groupes d'individus. Nous démontrons également comment intégrer des ellipses de concentration et des ellipses de confiance pour chaque groupe.

```
res.km <- eclust(bloc1, "kmeans",nstart = 25)
```



### 3. Description des classes par les variables :

- Cluster **“Bleu”** : Ce sont les individus qui ne souffrent presque pas des symptômes de dépression, on peut nommer la classe Bonne santé mentale.
- Cluster **“Vert”** : Ce sont les individus qui souffrent presque toujours des symptômes de dépression, on peut nommer la classe Situation Dépressive.
- Cluster **“Rouge”** : Ce sont les individus qui souffrent moyennement des symptômes de dépression, on peut nommer la classe Pré-dépression.

### 4. Classification Automatique :

#### 4.1. Silhouette :

`fviz_silhouette(res.km)`

```
cluster size ave.sil.width
1 1 56 0.06
2 2 36 0.24
3 3 34 0.32
```

