

Etude de la criminalité et des indicateurs sociaux des départements Français en 2020

LO Mouhamadou Mansour & HAMZA Assoumani Chissi

14 Avril 2021

Introduction

Depuis 2012, les crimes et délits enregistrés par les services de gendarmerie et de police nationale dans l'ensemble des départements français, y compris celles des DOM-TOM sont recensés. Les statistiques des crimes et délits enregistrés par les services de police et les unités de gendarmerie reposent sur des critères et des méthodes d'enregistrement communs et identiques autorisant la comparaison des données entre l'année de référence et la précédente ou la suivante. Elles sont établies à partir des plaintes reçues ou d'après les constatations diligentées par les différents services.

À ces données sur la criminalité des différents départements français, des données sur des indicateurs sociaux-economiques ont aussi été rajoutées. Ces indicateurs portent sur la classe d'âge, niveau de vie, indicateur d'activité, etc.

Nous ferons une étude sur les crimes et délits enregistrés durant l'année 2020, et sur les indicateurs sociaux selon les départements. Pour répondre à cette problématique, nous allons réaliser des analyses statistiques pour satisfaire aux questions suivantes :

- Est-ce que l'on commet le même délit ou crime d'un département à l'autre ?
- Les départements ont-ils les memes tendances criminelles ou delictuelles ?
- Est-ce que les départements ont le même niveau de développement ?
- Existe-t-il des indicateurs liés entre eux ?
- Existe-t-il un lien entre le niveau de développement d'un département et sa criminalité ?

Les réponses à ces différentes questions, seront le but de notre travail.

description de données

Nos jeu de données proviennent des sites gouvernementaux : <https://www.data.gouv.fr/fr/datasets/crimes-et-delits-enregistres-par-les-services-de-gendarmerie-et-de-police-depuis-2012/> , <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/les-indicateurs-sociaux-departementaux> et <https://www.data.gouv.fr/fr/datasets/departements-et-leurs-regions/>

Finalement, notre dataset final contient 99 observations qui sont les départements français et 23 variables. Selon la nature de nos variables et des individus, nous avons un pseudo table de contingence qui croise la variable département en ligne avec 99 modalités et en colonne d'une part la variable crime_delit avec 10 modalités et d'autre part les variables sur les indicateurs sociaux qui intègrent à la fois des indicateurs transversaux, reflétant le contexte territorial (parmi lesquels l'espérance de vie, le taux de pauvreté monétaire, les catégories sociales...) et des indicateurs propres à chaque thématique.

Prétraitement de données

Les jeux de données servant pour la construction de notre dataset final ont été téléchargé en OpenSource sur les sites data.gouv.fr et du drees.solidarites-sante.gouv.fr . Nous avons pratiqué plusieurs transformations telles que : le nettoyage, la fusion, l'agrégation, regroupement de certaine modalité etc. L'ensemble de toutes ces transformations sont disponibles sur le fichier **données initiaux/creation_dataset.R** joint avec ce document.

```
library("FactoMineR")
library("factoextra")
library("missMDA")
library("markdown")
library("knitr")
library("rmarkdown")
#install('magick')
library(magick)

## Warning: package 'magick' was built under R version 4.0.4

#setwd("C:/Users/Mansour Lo/Desktop/Dataset")
dataset <- read.csv2("dataset.csv", encoding = "UTF-8", row.names = 1)
dataset <- dataset[, -c(9)]
d <- dataset[, 1:10]
indicsociaux <- dataset[1:98, 11:23]
```

Analyse de données

Notre jeu de données est basé sur deux analyses. La première analyse consiste à étudier la criminalité et la seconde analyse consiste à étudier l'impact des indicateurs sociaux. Notre but est de tirer une conclusion, à savoir s'il existe des départements qui peuvent se ressembler au niveau de la vie sociale, mais aussi au niveau de la criminalité et délinquance.

Etude sur les crimes et délits

Nous étudions un pseudo table de contingence qui croise en ligne la variable **département** avec 99 modalités et en colonne la variable **Crimes et délits** avec 10 modalités à savoir *Homicide_et_tentative*, *Coups_et_blessures*, *Prisedotage_Sequestration_Attentat*, *violence_Menaces*, *viol_Harcelements*, *Ventes_usage_stupefiant*, *Falsification_escroquerie*, *Destruction_Degradation*, *Vols_Cambriolages*, *Port_et_Detention_Armes_prohibe*.

Vu la nature des variables, nous allons utilisé **Une Analyse Factorielle des Correspondances(AFC)**

Vérifions si les deux variables sont indépendants

```
contingence<-as.table(as.matrix (d))
chisq <- chisq.test(contingence)
chisq

##
## Pearson's Chi-squared test
##
## data:  contingence
## X-squared = 102999, df = 882, p-value < 2.2e-16
```

On rejette l'hypothèse d'indépendance. Les deux variables sont liées. Nous n'avons pas la même criminalité suivant les départements.

Réalisation d'une AFC pour expliquer le lien entre les deux variables

```
library ("FactoMineR")
res.ca <- CA(contingence, graph = FALSE)
```

Choix du nombre d'axes

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

```
library ("factoextra")
eig.val <- get_eigenvalue(res.ca)
eig.val
```

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	0.0197370662	56.5147829	56.51478
## Dim.2	0.0058783377	16.8319333	73.34672
## Dim.3	0.0042005049	12.0276550	85.37437
## Dim.4	0.0027157389	7.7762010	93.15057
## Dim.5	0.0009640792	2.7605282	95.91110
## Dim.6	0.0005963978	1.7077154	97.61882
## Dim.7	0.0005077111	1.4537714	99.07259
## Dim.8	0.0002619246	0.7499904	99.82258
## Dim.9	0.0000619625	0.1774224	100.00000

Les 2 premiers axes de l'analyse expriment **73.34%** de l'inertie totale du jeu de données ; cela signifie que 73.34% de la variabilité totale du nuage des lignes (ou des colonnes) est représentée dans ce plan. C'est un pourcentage élevé, et le premier plan représente donc bien la variabilité contenue dans une très large part du jeu de données actif.

```
fviz_eig(res.ca, addlabels = TRUE, ylim = c(0, 60), main = "Décomposition de l'inertie totale")
```

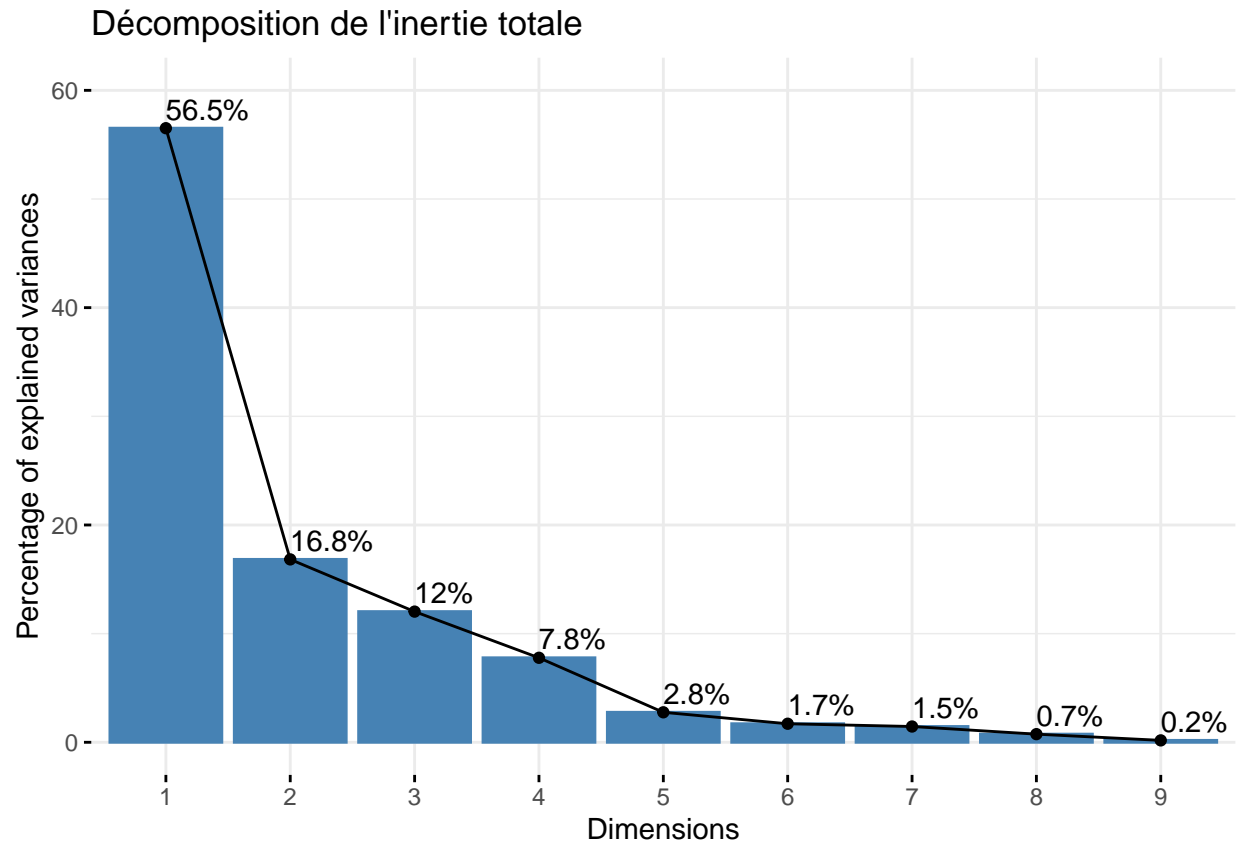


Figure 1 - Décomposition de l'inertie totale

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 2 premiers axes. Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

Interprétation sémantique des profils lignes

1. *Suivant l'axe 1*

```
row<-get_ca_row(res.ca)

coord<-row$coord[,1]
contrib<-row$contrib[,1]
cos2<-row$cos2[,1]
display<-cbind(coord,contrib,cos2)
#display
library(magick)
image<- image_read("image/display1.JPG")
plot(image, main="my image")
```

	coord	contrib	cos2
Paris	-0.2685698094	2.784490e+01	9.713134e-01
La Réunion	0.3906799273	6.362729e+00	6.533655e-01
Gironde	-0.2004206684	5.921421e+00	7.561267e-01
Loire-Atlantique	-0.2178425174	5.804384e+00	8.567343e-01
Pas-de-Calais	0.2243770271	4.805836e+00	8.606860e-01
Rhône	-0.1554431251	4.669098e+00	9.261434e-01
Haute-Garonne	-0.1586893846	3.254022e+00	9.055925e-01
Haut-Rhin	0.2661018469	2.894330e+00	8.746552e-01
Bouches-du-Rhône	-0.0991100681	2.164395e+00	5.532179e-01
Meurthe-et-Moselle	0.2041610112	1.955342e+00	7.834536e-01
Martinique	0.2566786396	1.609807e+00	4.929852e-01

Le coté négatif de l'axe 1 est caractérisé par le profil *Paris* qui s'opposent aux profils *La Réunion* et *Pas de Calais*. Cela veut dire que la distribution statistique de la criminalité n'est pas la même entre Paris et La Réunion et Pas de Calais.

2. suivant l'axe 2

```
row<-get_ca_row(res.ca)

coord<-row$coord[,2]
contrib<-row$contrib[,2]
cos2<-row$cos2[,2]
display<-cbind(coord,contrib,cos2)
#display
image<- image_read("image/display2.JPG")
plot(image)
```

	coord	contrib	cos2
Seine-Saint-Denis	1.797144e-01	2.279166e+01	7.768925e-01
Hauts-de-Seine	-1.199630e-01	7.921812e+00	5.424207e-01
Guadeloupe	2.109382e-01	4.865319e+00	3.592410e-01
Bouches-du-Rhône	7.720438e-02	4.409740e+00	3.356948e-01
Indre-et-Loire	-1.619935e-01	3.619751e+00	7.581784e-01
Puy-de-Dôme	-1.597886e-01	3.381091e+00	7.376226e-01
Val-d'Oise	8.432289e-02	2.591942e+00	5.486770e-01
Guyane	1.697667e-01	2.481301e+00	1.499989e-01
Martinique	1.705712e-01	2.386898e+00	2.177038e-01
Val-de-Marne	7.213221e-02	2.380042e+00	4.657431e-01

Sur l'axe 2, on remarque que le profil *Seine-Saint-Denis* est caractérisé sur le coté positif de l'axe, en opposition du profil *Hauts-de-seine*. Ce qui veut dire que la distribution caractérisant la criminalité et délit n'est pas la même entre Seine-Saint-Denis et Hauts-de-seine.

3. Représentation graphique pour illustrer nos analyses

Graphique Colorer en fonction du cos2

```
fviz_ca_row (res.ca, col.row = "cos2",
             title = "Profils lignes selon leur cosinus²",
             gradient.cols = c ("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```

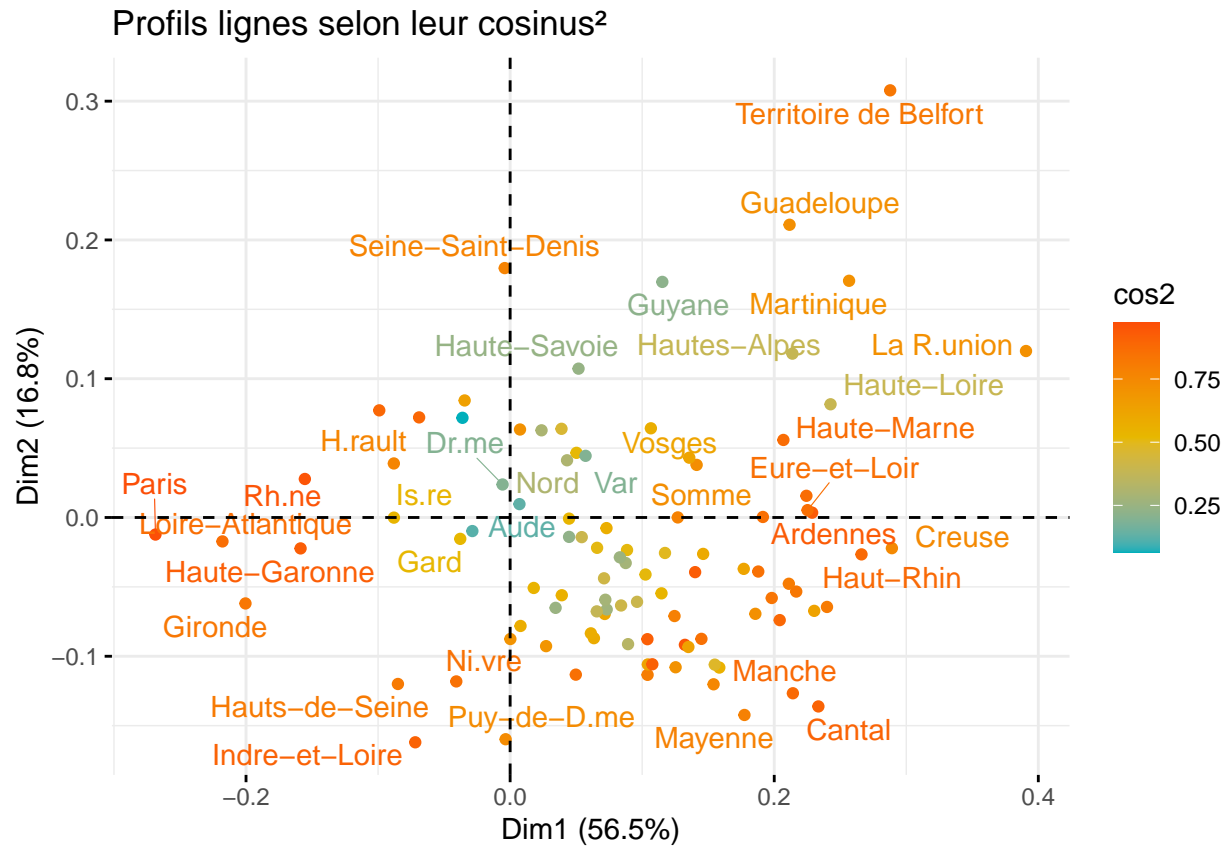


Figure 2: Grapique Colorer en fonction du cos2

Grapique Colorer en fonction de la contribution

```
fviz_ca_row(res.ca, pointsize = "contrib",
  title="Profils lignes selon leur contribution",
  repel = TRUE)
```

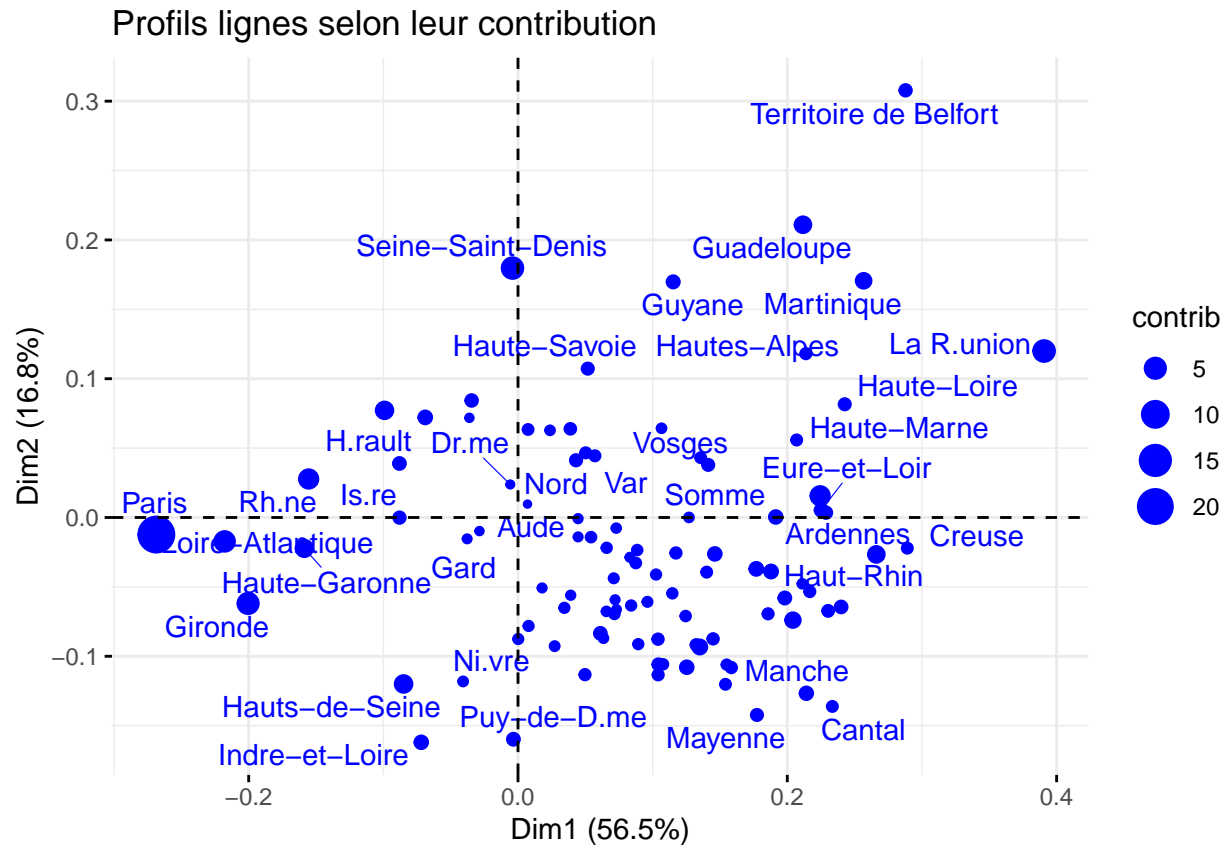


Figure 3: Profils lignes selon leur contribution

Interprétation sémantique des profils colonnes

1. Suivant l'axe 1

```
col<-get_ca_col(res.ca)
```

```
coord<-col$coord[,1]
```

```
contrib<-col$contrib[,1]
```

```
cos2<-col$cos2[,1]
```

```
display<-cbind(coord,contrib,cos2)
```

```
display
```

##	coord	contrib	cos2
## Homicide_et_tentative	0.13365748	0.12447653	0.03537915
## Coups_et_blessures	0.20948500	20.50313661	0.77050550
## Prisedotage_Sequestration_Attentat	0.09221998	0.06348087	0.06936789
## violences_Menaces	0.19858326	17.20380672	0.84222285
## viol_Harcelements	0.25247821	7.91138683	0.60776952
## Ventes_usage_stupefiant	0.13714737	5.81392585	0.24203603
## Falsification_escroquerie	0.04505479	1.45098129	0.06822667
## Destruction_Degradation	0.10006645	3.77512552	0.20701845
## Vols_Cambriolages	-0.12943760	43.02811709	0.97498091
## Port_et_Detention_Armes_prohibe	0.04757086	0.12556269	0.02614971

Le profil *Vols_Cambriolages* caractérise le côté négatif de l'axe 1 par opposition au profil *Coups_et_blessures* sur le côté positif de l'axe 1. Cela signifie que les départements se différencient sur les crimes et délits de types *Vols_Cambriolages* et du type *Coups_et_blessures*.

2. Suivant l'axe 2

```
col<-get_ca_col(res.ca)
```

```
coord<-col$coord[,2]
contrib<-col$contrib[,2]
cos2<-col$cos2[,2]
display<-cbind(coord,contrib,cos2)
display
```

##	coord	contrib	cos2
## Homicide_et_tentative	0.30138107	2.125010	0.17988412
## Coups_et_blessures	0.05414939	4.599701	0.05148218
## Prisedotage_Sequestration_Attentat	0.20782571	1.082479	0.35229542
## violences_Menaces	0.03573275	1.870252	0.02726933
## viol_Harcelements	-0.12089760	6.090715	0.13935626
## Ventes_usage_stupefiant	0.16019595	26.633340	0.33022359
## Falsification_escroquerie	-0.13536948	43.979417	0.61590503
## Destruction_Degradation	-0.07684228	7.474514	0.12207664
## Vols_Cambriolages	0.01492313	1.920345	0.01295970
## Port_et_Detention_Armes_prohibe	0.15058104	4.224228	0.26201458

Contrairement sur l'axe 2, on remarque une forte contribution du profil *Falsification_escroquerie* sur le côté négatif de l'axe, par opposition du profil *Vente_usage_stupefiant* dans le côté positif de l'axe 2. Ce qui signifie que la distribution statistique pour les départements sur la criminalité de type *Falsification_escroquerie* et *Vente_usage_stupefiant* n'est pas la même.

3- Représentation graphique

```
fviz_ca_col(res.ca, repel = TRUE)
```

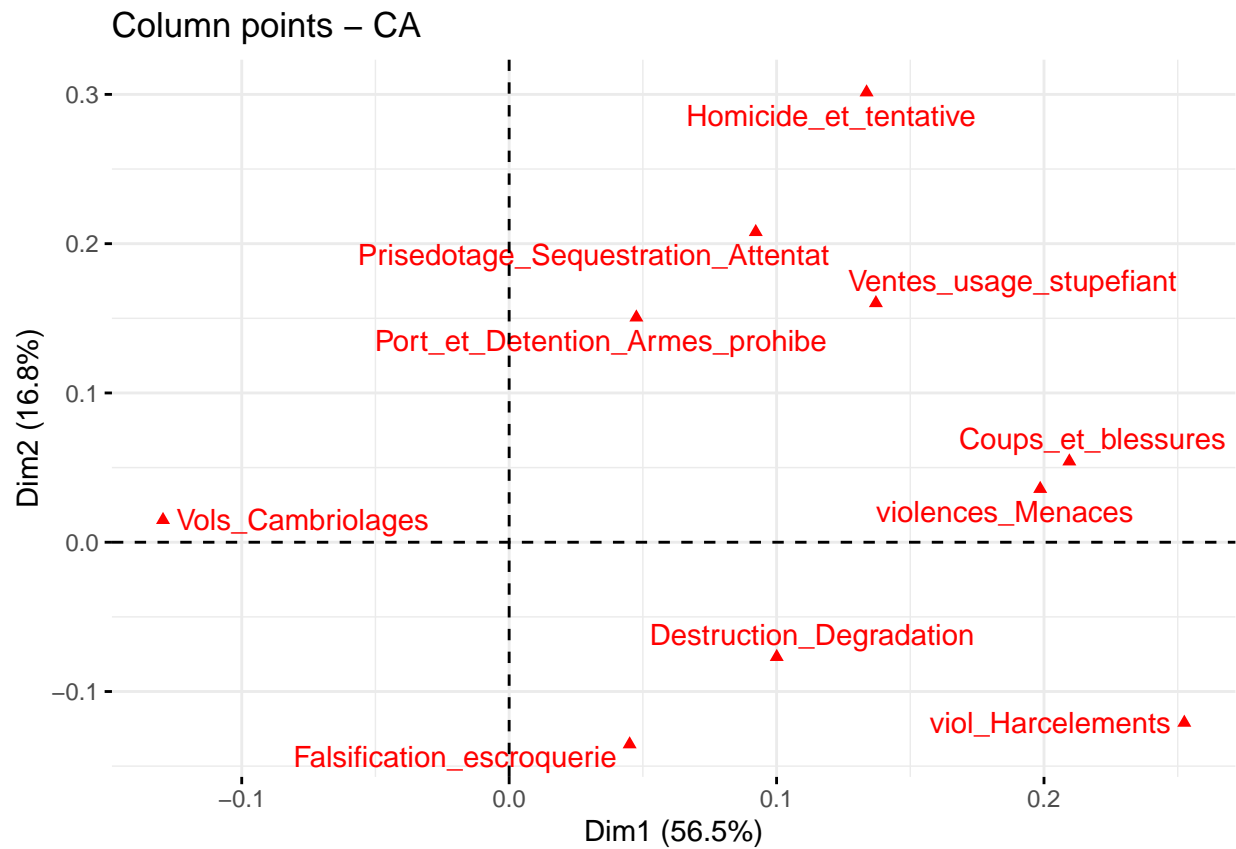


Figure 4: graphique des profils colonnes

Graphique colorer en fonction du Cos

```
fviz_ca_col(res.ca, col.col = "cos2",
             title = "Profils colonnes selon leur cosinus2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```

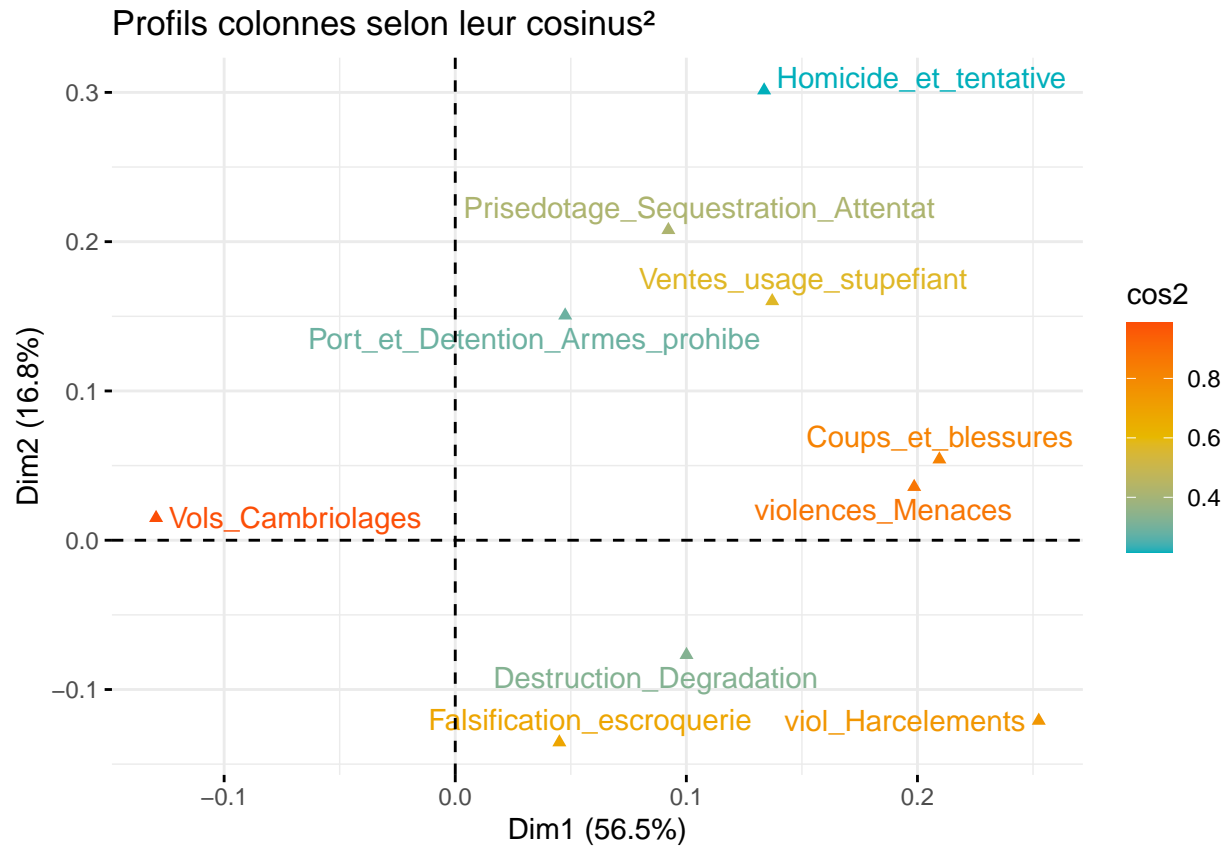


Figure 5: Graphique colorer en fonction du Cos

Graphique colorer selon la contribution

```
fviz_ca_col(res.ca, col.col = "contrib",
            title = "Profils colonnes selon leur contribution",
            gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
            repel = TRUE)
```

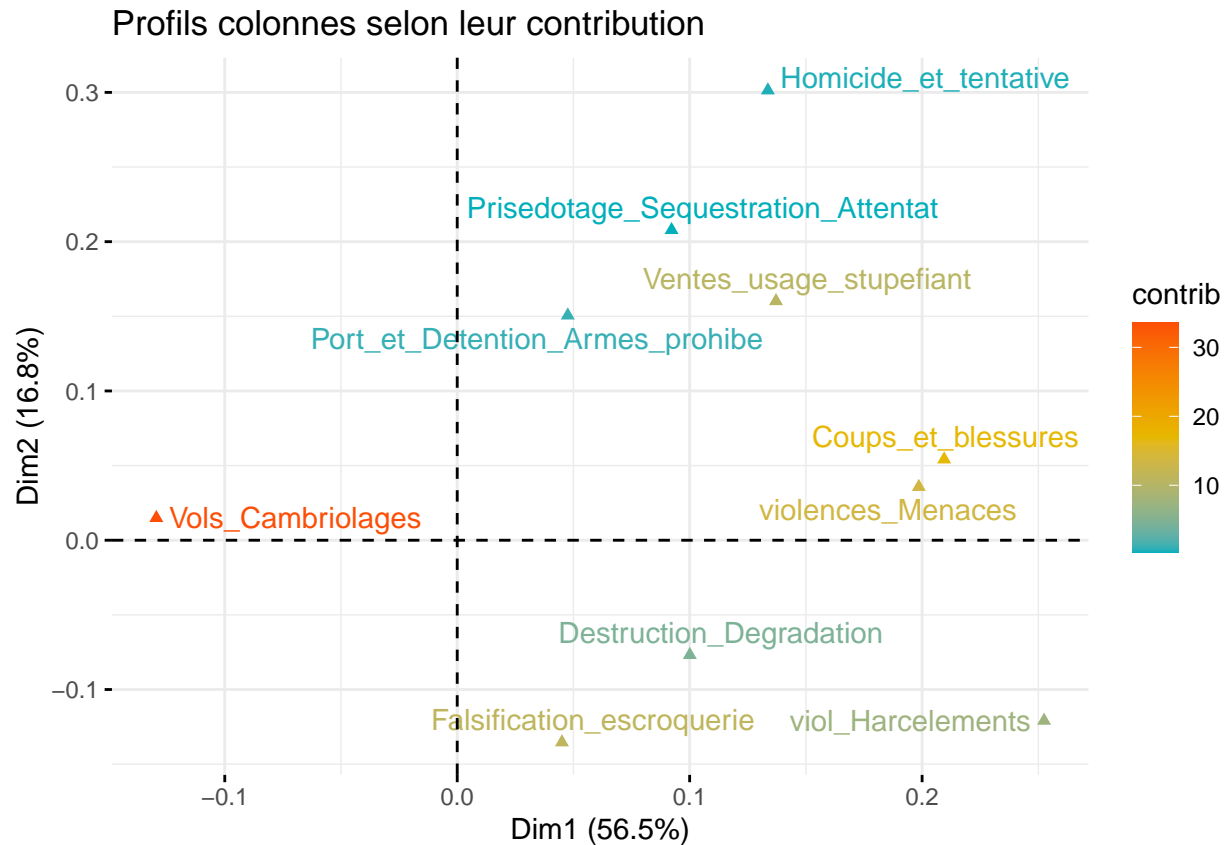


Figure 6: Graphique colorer selon la contribution

Classification

Dans cette partie, nous réalisons une classification pour explorer nos données. Les questions qui peuvent se poser et qui feront l'objet de l'analyse sont les suivantes :

1. Quels sont les profils lignes les plus caractéristiques pour une partition donnée.
2. Quelles sont les variables (les profils colonnes) qui caractérisent le mieux la partition de chaque classe.

Appliquons la classification hiérarchique sur le résultat de l'AFC

```
res.hcpc=HCPC(res.ca,graph = FALSE,nb.clust = -1)
```

Visualisons le dendrogramme généré par la classification

```
fviz_dend(res.hcpc, show_labels = FALSE)
```

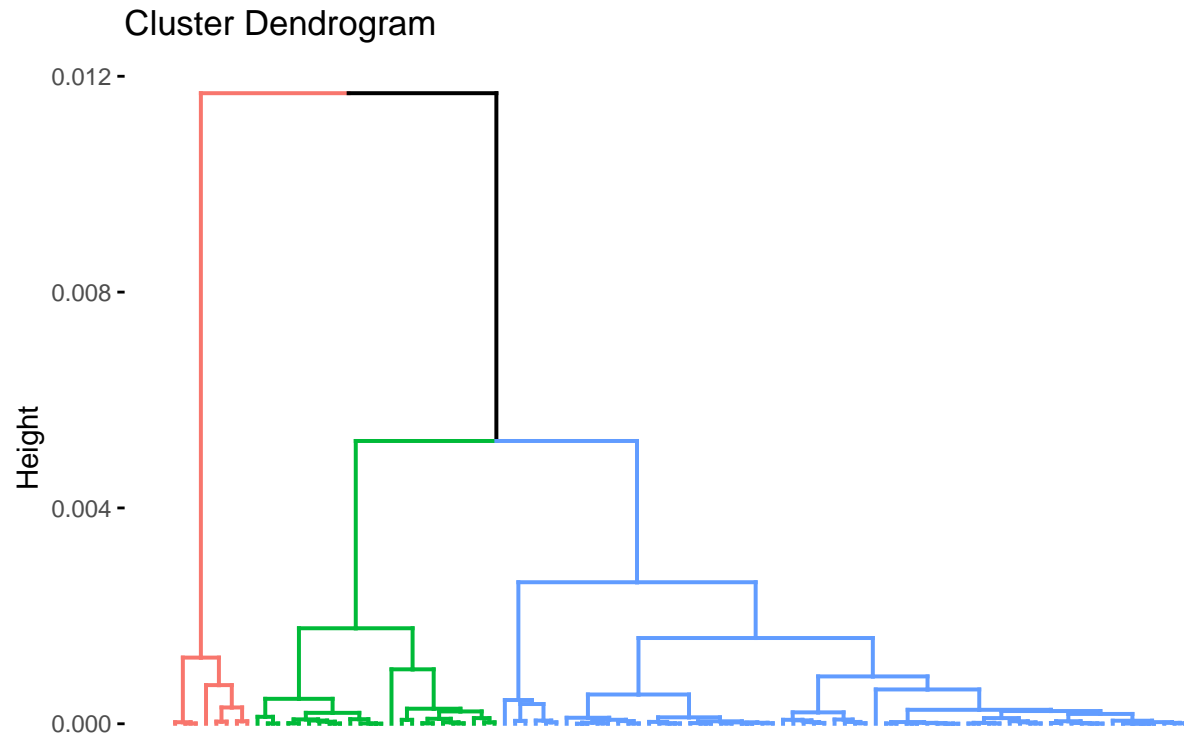


Figure 7: Dendrogramme des classes

Le dendrogramme suggère une solution à 3 groupes

Visualisons les profils et colorons par groupes sur le plan factoriel

```
fviz_cluster(res.hcpc,
  repel = TRUE,
  show.clust.cent = TRUE,
  palette = "jco",
  ggtheme = theme_minimal(),
  main = "Factor map"
)
```

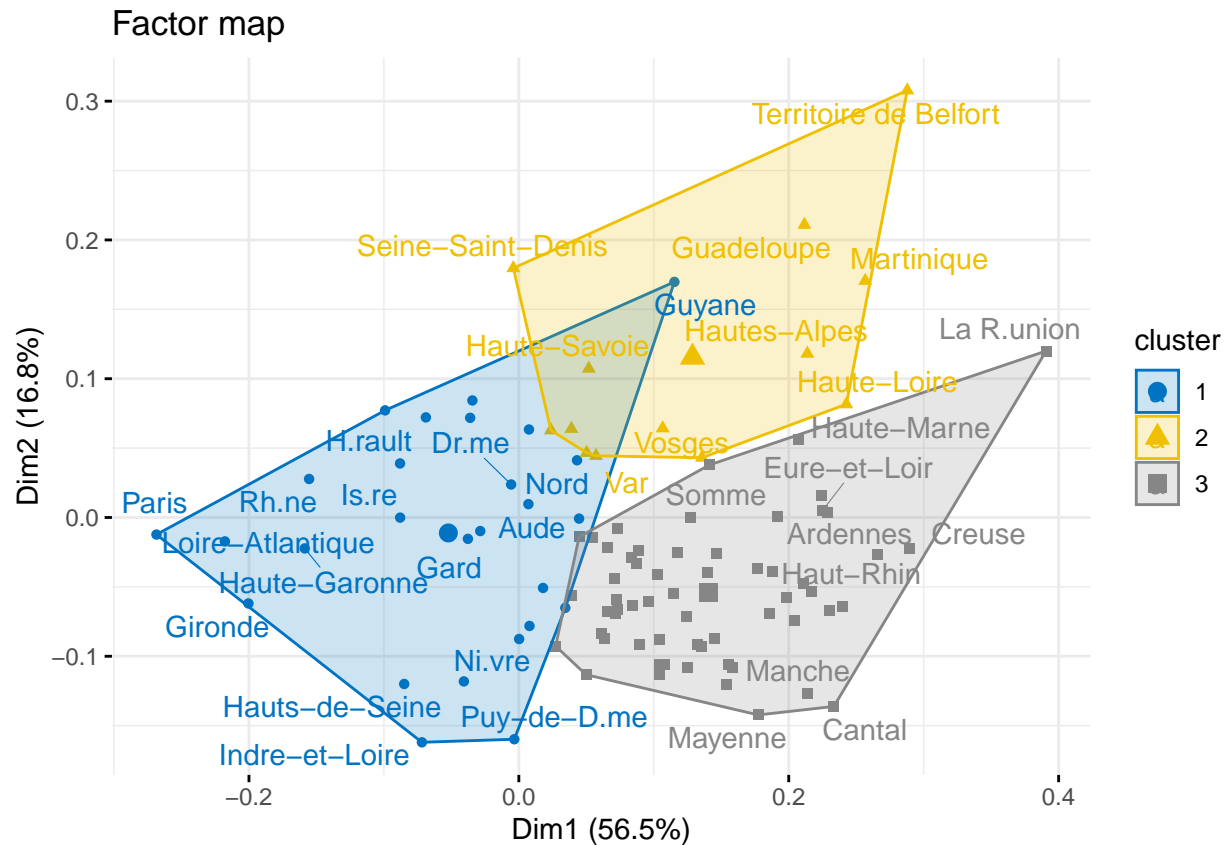


Figure 8: Repartition des classes sur plan

Description et sémantique de chaque groupe

Suivant les Profils lignes

res.hcpc\$desc.ind

```
## $para
## Cluster: 1
##           Gard           Aude Pyr<e9>n<e9>es-Orientales
##           0.02782297       0.06787928       0.07132780
##           Is<e8>re         H<e9>rault
##           0.07297453       0.07432051
## -----
## Cluster: 2
## Alpes-de-Haute-Provence   Vosges           Var
##           0.06028139       0.07893434       0.11797715
##           Seine-et-Marne   Vaucluse
##           0.11854779       0.12149858
## -----
## Cluster: 3
##           Tarn           Charente C<f4>tes-d'Armor   Loir-et-Cher
##           0.02998569     0.04266072       0.05184473   0.05915194
##           Tarn-et-Garonne
##           0.06517996
## $dist
```

```
## Cluster: 1
##      Guyane      Paris      Mayotte Loire-Atlantique
##      0.4352276    0.4121191    0.3746896      0.3741922
##      Gironde
##      0.3589288
## -----
## Cluster: 2
## Territoire de Belfort      Haute-Loire      Hautes-Alpes
##      0.4073727      0.3446192      0.3423180
##      Martinique      Guadeloupe
##      0.3135206      0.3123299
## -----
## Cluster: 3
## La R<e9>union      Mayenne      Cantal      Haut-Rhin      Manche
##      0.4253024      0.2790701      0.2778838      0.2678952      0.2645030
```

1. Pour la classe 1, les paragon, c'est à dire les départements représentatifs sont *Gard, Aude, pyrénées-orientales, Isère et Hérault* (ils ressemblent le plus à la moyenne de la classe 1). Les départements de la classe 1 peuvent être résumés par ces 5 départements. Pour la classe 1, les départements les plus caractéristiques c-à-d ceux qui s'éloignent les plus des autres classes sont *Guyane, Paris, Mayotte Loire-Athlantique et Gironde*.
2. Pour la classe 2, les paragon sont *Alpes-de-Haute-Provence, Vosges, Var, Seine-et-Marne, Vaucluse* (ils ressemblent le plus à la moyenne de la classe 2). Les départements de la classe 2 peuvent être résumés par ces 5 départements. Pour la classe 2, les départements les plus caractéristiques c-à-d ceux qui se distinguent les plus des autres classes sont *Territoire de Belfort, Haute-Loire, Hautes-Alpes*.
3. Même chose pour la classe 3, les paragon sont *Tarn, Charente, Côtes-d'Armor, Loir-et-Cher Tarn-et-Garonne* et les plus caractérisant sont *La Réunion, Mayenne, Cantal, Haut-Rhin et Manche*

Description des classes suivant les Profils colonnes(les variables)

```
res.hcpc$desc.var
```

```
## $`1`
##      Intern %      glob % Intern freq Glob freq
## Vols_Cambriolages      55.572249 50.6890351      855183      1494956
## Prisedotage_Sequestration_Attentat 0.140363 0.1473246      2160      4345
## Port_et_Detention_Armes_prohibe      1.025105 1.0951188      15775      32298
## Destruction_Degradation      6.937442 7.4410981      106758      219458
## Falsification_escroquerie      13.407786 14.1079027      206328      416080
## Ventes_usage_stupefiant      5.232291 6.1006643      80518      179925
## viol_Harcelements      2.023957 2.4495561      31146      72244
## violences_Menaces      7.477254 8.6103709      115065      253943
## Coups_et_blessures      8.046439 9.2214037      123824      271964
##      p.value      v.test
## Vols_Cambriolages      0.000000e+00      Inf
## Prisedotage_Sequestration_Attentat 1.198358e-03 -3.239271
## Port_et_Detention_Armes_prohibe      1.820617e-33 -12.055213
## Destruction_Degradation      2.116234e-259 -34.405089
## Falsification_escroquerie      9.231179e-285 -36.061126
## Ventes_usage_stupefiant      0.000000e+00      -Inf
## viol_Harcelements      0.000000e+00      -Inf
## violences_Menaces      0.000000e+00      -Inf
## Coups_et_blessures      0.000000e+00      -Inf
##
## $`2`
```

```
## Intern % glob % Intern freq Glob freq
## Ventes_usage_stupefiant 8.9128936 6.1006643 36709 179925
## Port_et_Detention_Armes_prohibe 1.4208574 1.0951188 5852 32298
## Coups_et_blessures 10.0263679 9.2214037 41295 271964
## violences_Menaces 9.3890216 8.6103709 38670 253943
## Prisedotage_Sequestration_Attentat 0.1847697 0.1473246 761 4345
## Homicide_et_tentative 0.1728726 0.1375256 712 4056
## viol_Harcelements 2.0562613 2.4495561 8469 72244
## Destruction_Degradation 6.5876114 7.4410981 27132 219458
## Falsification_escroquerie 12.8627411 14.1079027 52977 416080
## Vols_Cambriolages 48.3866033 50.6890351 199287 1494956
## p.value v.test
## Ventes_usage_stupefiant 0.000000e+00 Inf
## Port_et_Detention_Armes_prohibe 1.178278e-96 20.862315
## Coups_et_blessures 7.260460e-81 19.044770
## violences_Menaces 1.984652e-80 18.992040
## Prisedotage_Sequestration_Attentat 6.856415e-11 6.523765
## Homicide_et_tentative 1.884235e-10 6.370491
## viol_Harcelements 1.579666e-72 -18.011610
## Destruction_Degradation 1.558068e-115 -22.847202
## Falsification_escroquerie 4.508267e-138 -25.012160
## Vols_Cambriolages 9.176277e-223 -31.861315
```

```
## $`3`
## Intern % glob % Intern freq Glob freq
## Destruction_Degradation 8.5693284 7.4410981 85568 219458
## Falsification_escroquerie 15.7004541 14.1079027 156775 416080
## viol_Harcelements 3.2676773 2.4495561 32629 72244
## violences_Menaces 10.0354719 8.6103709 100208 253943
## Coups_et_blessures 10.7001436 9.2214037 106845 271964
## Ventes_usage_stupefiant 6.2789799 6.1006643 62698 179925
## Port_et_Detention_Armes_prohibe 1.0686624 1.0951188 10671 32298
## Homicide_et_tentative 0.1235807 0.1375256 1234 4056
## Vols_Cambriolages 44.1130933 50.6890351 440486 1494956
## p.value v.test
## Destruction_Degradation 0.000000e+00 Inf
## Falsification_escroquerie 0.000000e+00 Inf
## viol_Harcelements 0.000000e+00 Inf
## violences_Menaces 0.000000e+00 Inf
## Coups_et_blessures 0.000000e+00 Inf
## Ventes_usage_stupefiant 6.740401e-20 9.131745
## Port_et_Detention_Armes_prohibe 1.789880e-03 -3.123049
## Homicide_et_tentative 3.389907e-06 -4.645660
## Vols_Cambriolages 0.000000e+00 -Inf
```

```
## attr("class")
## [1] "descfreq" "list"
```

```
departement <- row.names(res.hcpc$data.clust)
clust_crim = cbind(departement,subset(res.hcpc$data.clust,select = "clust"))

names(clust_crim) <- c("departement","Type_criminalite")
clust_crim <- data.frame(lapply(clust_crim, function(x) {
  gsub("1", "type_vols", x)
```



```

    )))
  clust_crim <- data.frame(lapply(clust_crim, function(x) {
    gsub("2", "type_VUS", x)
  })))
  clust_crim <- data.frame(lapply(clust_crim, function(x) {
    gsub("3", "type_detruct&degra", x)
  })))

```

On peut dire que :

La classe 1 qui est composée des profils lignes telles que Paris, Loire-Atlantique, Gironde etc., est caractérisé par :

- une forte fréquence de la modalité *Vols_Cambriolages*
- une faible fréquence des modalités *Coups_et_blessures*, *violences_Menaces*, *viol_Harcelements* (du plus rare au plus commun).

La classe 2 qui est composée des profils lignes telles que *Seine-Saint-Denis*, *Guadeloupe*, *Martinique* etc., est caractérisé par :

- une forte fréquence des modalités *Ventes_usage_stupefiant*, *Port_et_Detention_Armes_prohibe*, *Coups_et_blessures*, *violences_Menaces*, *Prisedotage_Sequestration_Attentat* et *Homicide_et_tentative* (du plus commun au plus rare).
- une faible fréquence des modalités *Vols_Cambriolages*, *Falsification_escroquerie*, *Destruction_Degradation* et *viol_Harcelements* (du plus rare au plus commun).

La classe 3 qui est composée des profils lignes telles que *Pas-de-Calais*, *Haut-Rhin*, *La Réunion* etc., est caractérisé par :

- une forte fréquence des modalités *Destruction_Degradation*, *Falsification_escroquerie*, *viol_Harcelements*, *violences_Menaces*, *Coups_et_blessures* et *Ventes_usage_stupefiant* (du plus commun au plus rare).
- une faible fréquence des modalités *Vols_Cambriolages*, *Homicide_et_tentative* et *Port_et_Detention_Armes_prohibe* (du plus rare au plus commun).

Etude sur les indicateurs sociaux

Analyse de l'ACP

Nous procédons ici à une *Analyse en Composantes Principales (ACP) normée*, car les indicateurs sociaux sont des variables quantitatives .

Nous rajoutons comme variable supplémentaire quantitative **le taux de criminalité**

```

nb <- estim_ncpPCA(indicsociaux,ncp.max=5)
res.comp <- imputePCA(indicsociaux,ncp=2)
res.pca <- PCA(res.comp$completeObs, quanti.sup = 13, scale.unit = TRUE, ncp = 5,graph= FALSE)

```

Nous utilisons le package *missMDA* pour estimer les valeurs manquantes , la PCA classique renvoie un warning

```

eig.val <- get_eigenvalue(res.pca)
eig.val

```

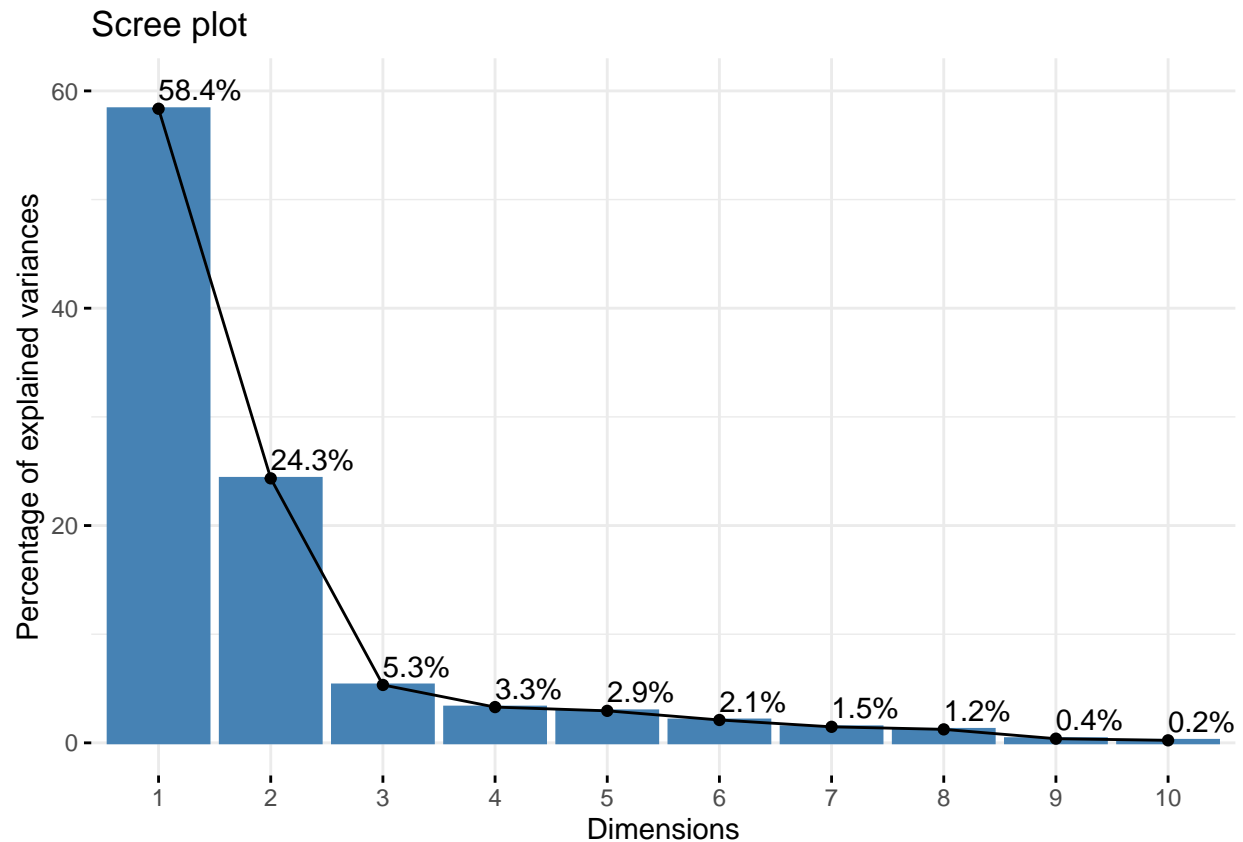
```

##          eigenvalue variance.percent cumulative.variance.percent
## Dim.1    7.00298516         58.3582097          58.35821
## Dim.2    2.92107033         24.3422527          82.70046
## Dim.3    0.63975628          5.3313023          88.03176
## Dim.4    0.39517215          3.2931012          91.32487

```

```
## Dim.5  0.35373882      2.9478235      94.27269
## Dim.6  0.25278055      2.1065046      96.37919
## Dim.7  0.17724900      1.4770750      97.85627
## Dim.8  0.14902721      1.2418934      99.09816
## Dim.9  0.04618412      0.3848677      99.48303
## Dim.10 0.02796480      0.2330400      99.71607
## Dim.11 0.01985790      0.1654825      99.88155
## Dim.12 0.01421368      0.1184473      100.00000
```

```
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 60))
```



Nous cherchons a trouver le nombre optimal d'axe a retenir . Ici c'est les deux premiers axes en utilisant le critere de Kaiser.

```
ind <- get_pca_ind(res.pca)
coord<-ind$coord[,1]
contrib<-ind$contrib[,1]
cos2<-ind$cos2[,1]
display<-cbind(coord,contrib,cos2)
#display
image<- image_read("image/image01.png")
plot(image, main="my image")
```

	coord	contrib	cos2
Guyane	15.09177276	3.318725e+01	0.8419101620
La Réunion	10.49445908	1.604763e+01	0.9323179661
Guadeloupe	9.46656633	1.305797e+01	0.9106329906
Martinique	6.54521433	6.242211e+00	0.8086613311
Haute-Savoie	-3.88739406	2.201952e+00	0.5704556562
Hauts-de-Seine	-3.61850675	1.907873e+00	0.3396325180
Pyrénées-Orientales	3.47837507	1.762964e+00	0.7145298203
Seine-Saint-Denis	3.45906633	1.743446e+00	0.3839530786
Yvelines	-3.32349076	1.609458e+00	0.4574139645
Paris	-3.12572783	1.423617e+00	0.2034400204
Pas-de-Calais	2.85983095	1.191712e+00	0.5366613802
Savoie	-2.72796561	1.084348e+00	0.8413648921

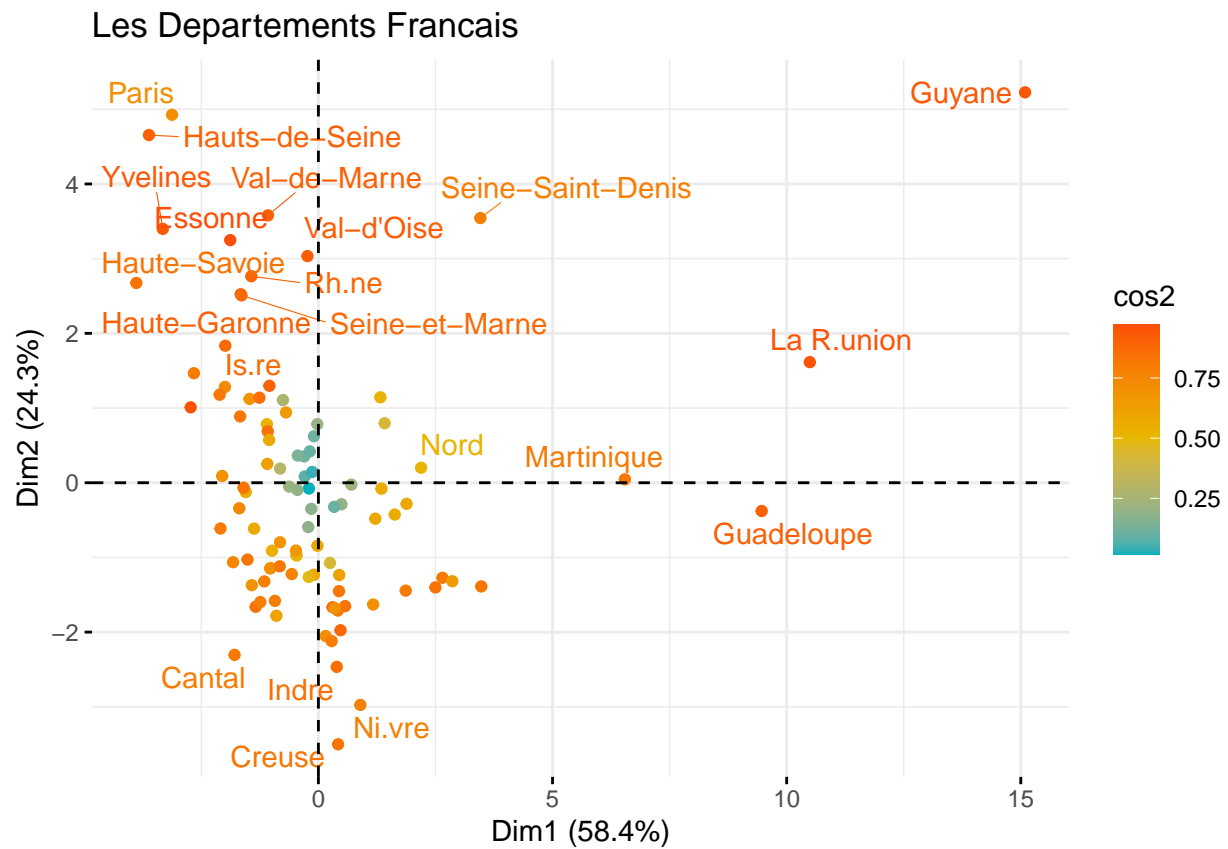
Le coté positif de l'axe 1 est représenté par Guyane, Reunion, Guadeloupe, Martinique. Ces départements ont des valeurs similaires sur les critères sociaux en opposition à la Haute-Savoie, Savoie, les Hauts-de-Seine et Yvelines qui représentent le coté négatif de l'axe 1.

```
ind <- get_pca_ind(res.pca)
coord<-ind$coord[,2]
contrib<-ind$contrib[,2]
cos2<-ind$cos2[,2]
display<-cbind(coord,contrib,cos2)
#display
image<- image_read("image/image02.png")
plot(image, main="my image")
```

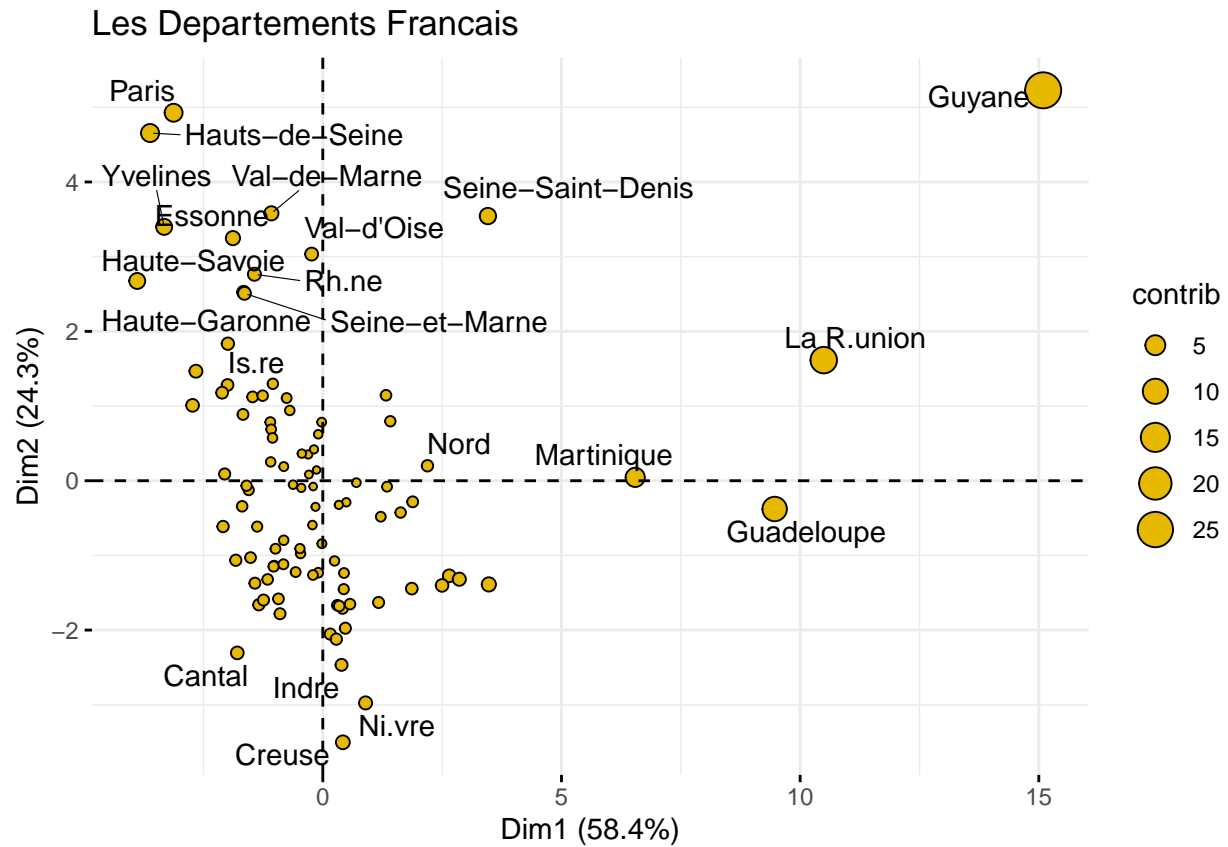
	coord	contrib	cos2
Guyane	5.22670871	9.5430786761	1.009815e-01
Paris	4.92660324	8.4786574097	5.053930e-01
Hauts-de-Seine	4.65498953	7.5695372203	5.620668e-01
Val-de-Marne	3.57990674	4.4768788013	8.352911e-01
Seine-Saint-Denis	3.54353132	4.3863619290	4.029331e-01
Creuse	-3.50198506	4.2841087611	8.155666e-01
Yvelines	3.39899622	4.0358338036	4.784338e-01
Essonne	3.24804384	3.6853240131	7.273936e-01
Val-d'Oise	3.03418390	3.2159975608	9.198465e-01
Nièvre	-2.97469269	3.0911218453	7.141261e-01
Rhône	2.76473789	2.6701756962	7.031776e-01
Haute-Savoie	2.67485533	2.4993812457	2.700880e-01
Haute-Garonne	2.52348988	2.2245135085	6.371850e-01

Et sur le coté positif de l'axe 2 est représenté par Paris et Val de Marne par opposition au coté négatif qui est représenté par Nièvre et la Creuse

Représentation graphique des individus sur les deux premiers axes.



Représentation graphiques des individus selon leurs Cos^2



Représentation graphique des individus selon leurs Contributions

```
var <- get_pca_var(res.pca)
coord<-var$coord[,1]
contrib<-var$contrib[,1]
cos2<-var$cos2[,1]
display<-cbind(coord,contrib,cos2)
#display
image<- image_read("image/image03.png")
plot(image, main="my image")
```

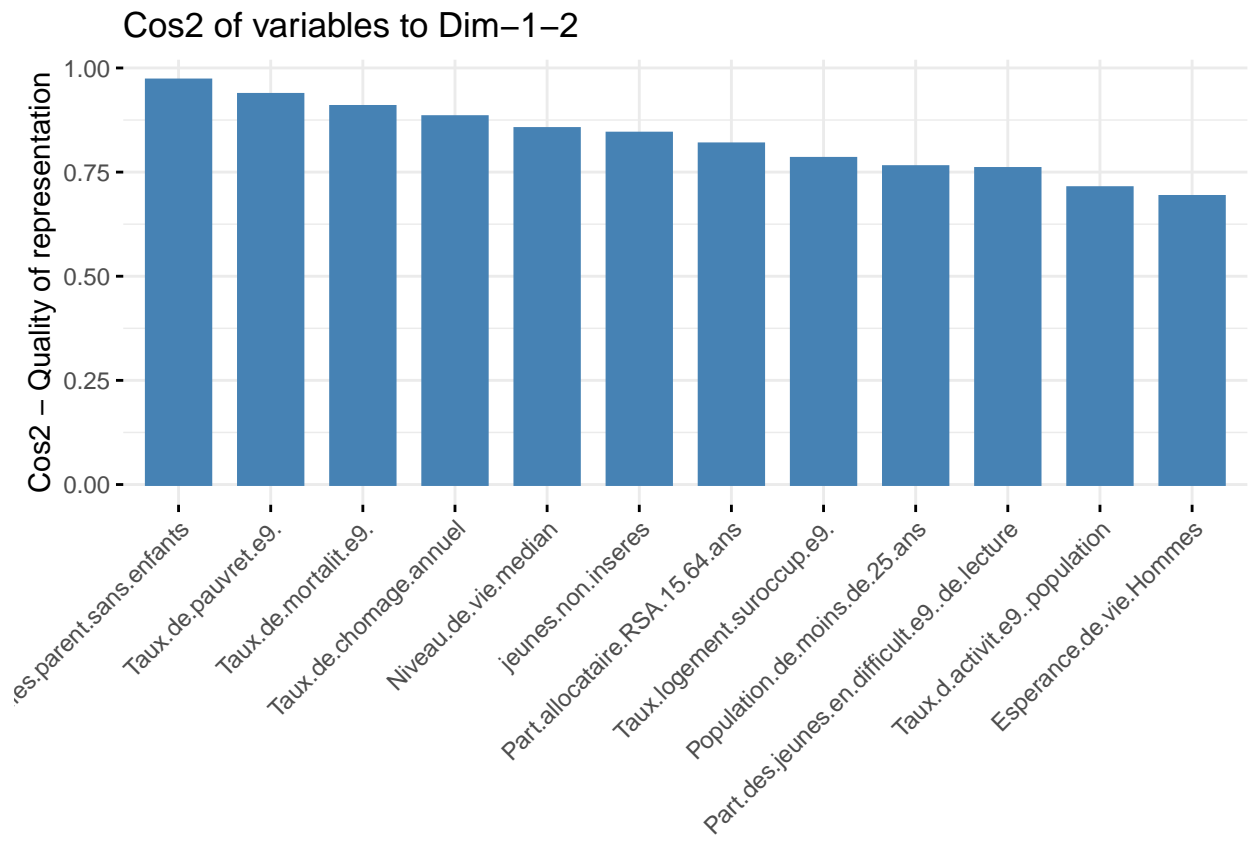
	coord	contrib	cos2
Part.d.enfant.dont.les.parent.sans.enfants	0.9778431	13.6538518	0.95617722
Taux.de.pauvreté	0.9591039	13.1355464	0.91988037
Taux.de.chomage.annuel	0.9367552	12.5305180	0.87751031
Part.allocataire.RSA.15.64.ans	0.9023068	11.6258636	0.81415750
jeunes.non.inseres	0.8850009	11.1841810	0.78322654
Part.des.jeunes.en.difficulté.de.lecture	0.8695835	10.7979029	0.75617554
Taux.d.activité.population	-0.8190368	9.5790752	0.67082122
Niveau.de.vie.median	-0.8050414	9.2545062	0.64809170
Esperance.de.vie.Hommes	-0.5643428	4.5478149	0.31848280
Population.de.moins.de.25.ans	0.3793430	2.0548537	0.14390110
Taux.logement.suroccupé	0.3108512	1.3798186	0.09662849
Taux.de.mortalité	-0.1339118	0.2560677	0.01793238

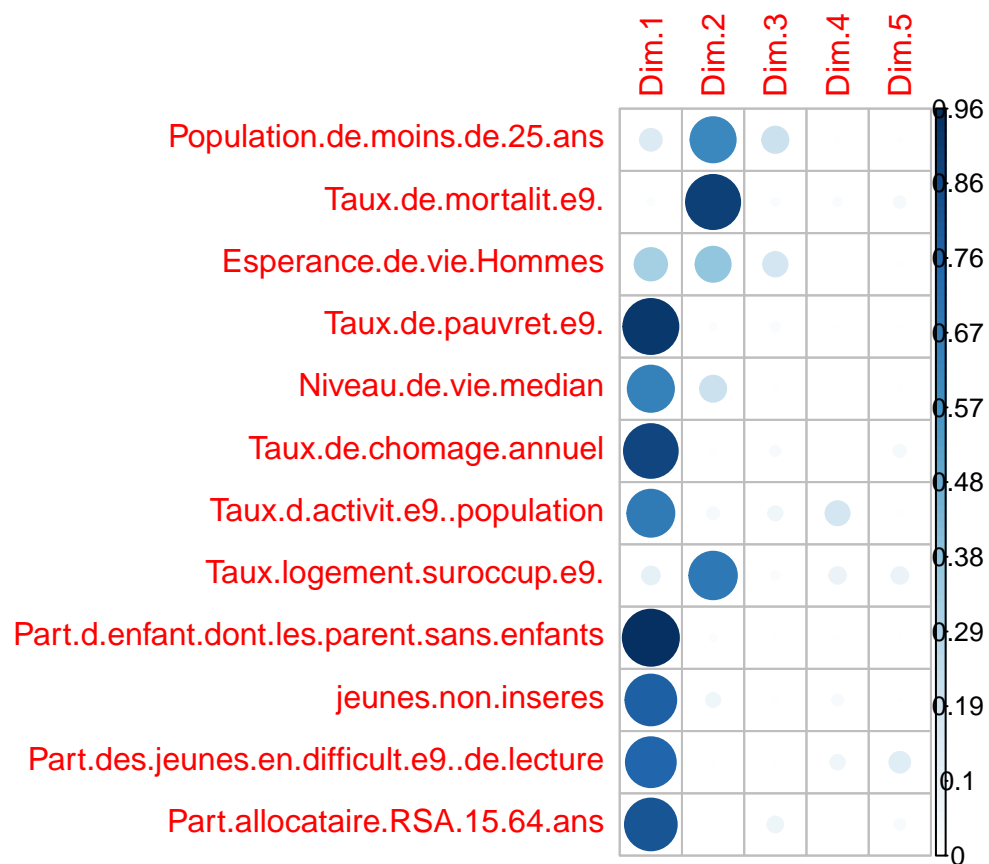
Le coté positif de l'axe 1 est représenté par les variables Part d'enfant dont les parents sont sans enfants, Taux de pauvreté, Taux de chômage annuel, Part allocataire RSA 15-64.ans. Les départements qui ont été mal notés dans ces critères ont été projetés dans ce coté. Le Taux d'activité population, Niveau de vie median, qui représentent le coté négatif de l'axe 1. Les départements qui ont été bien notés dans ces critères ont été projetés dans ce coté.

```
var <- get_pca_var(res.pca)
coord<-var$coord[,2]
contrib<-var$contrib[,2]
cos2<-var$cos2[,2]
display<-cbind(coord,contrib,cos2)
#display
image<- image_read("image/image04.png")
plot(image, main="my image")
```

	coord	contrib	cos2
Taux.de.mortalité	-0.94325900	30.4592989	0.889737542
Taux.logement.suroccupé	0.82851088	23.4992725	0.686430275
Population.de.moins.de.25.ans	0.78696778	21.2017588	0.619318286
Esperance.de.vie.Hommes	0.61086726	12.7747288	0.373158812
Niveau.de.vie.median	0.45449813	7.0716734	0.206568552
jeunes.non.inseres	-0.24565546	2.0659073	0.060346604
Taux.d.activité.population	0.20457759	1.4327621	0.041851989
Taux.de.pauvreté	0.12939937	0.5732213	0.016744197
Part.d.enfant.dont.les.parent.sans.enfants	0.12248847	0.5136276	0.015003424
Taux.de.chomage.annuel	0.07524312	0.1938169	0.005661527
Part.allocataire.RSA.15.64.ans	0.06049077	0.1252669	0.003659133
Part.des.jeunes.en.difficulté.de.lecture	0.05089192	0.0886657	0.002589987

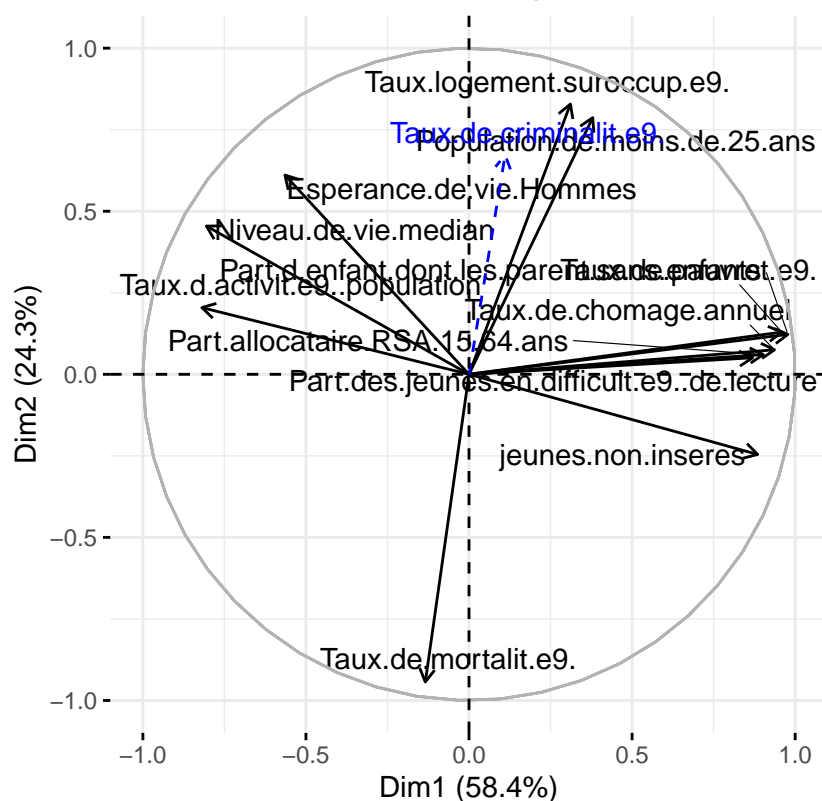
Le coté positif de l'axe 2 est représenté par les variables Taux logement suroccupé, Population de moins de 25 ans. Les départements qui ont été mal noté dans ces critères ont été projetés dans ce coté. Le Taux de mortalité représente le coté négatif de l'axe 2. Les départements qui ont été mal noté dans ce critère ont été projetés dans ce coté.





**Représentation graphique des \cos^2 variables suivant les dimensions*

Critères de notation des Departements avec variable illustrati



**Représentation graphique des variables*

Nous remarquons que les départements qui ont un *Taux de criminalité élevé*, ont aussi un *taux de logement sureoccupé* et *population de moins de 25 ans* élevés, par contre leur *taux de mortalité* est très faible**.

Les variables *Part d'enfant dont les parents sont sans enfants*, *Taux de pauvreté*, *Taux de chômage annuel*, *Part allocataire RSA 15-64ans* sont fortement liées elles.

•

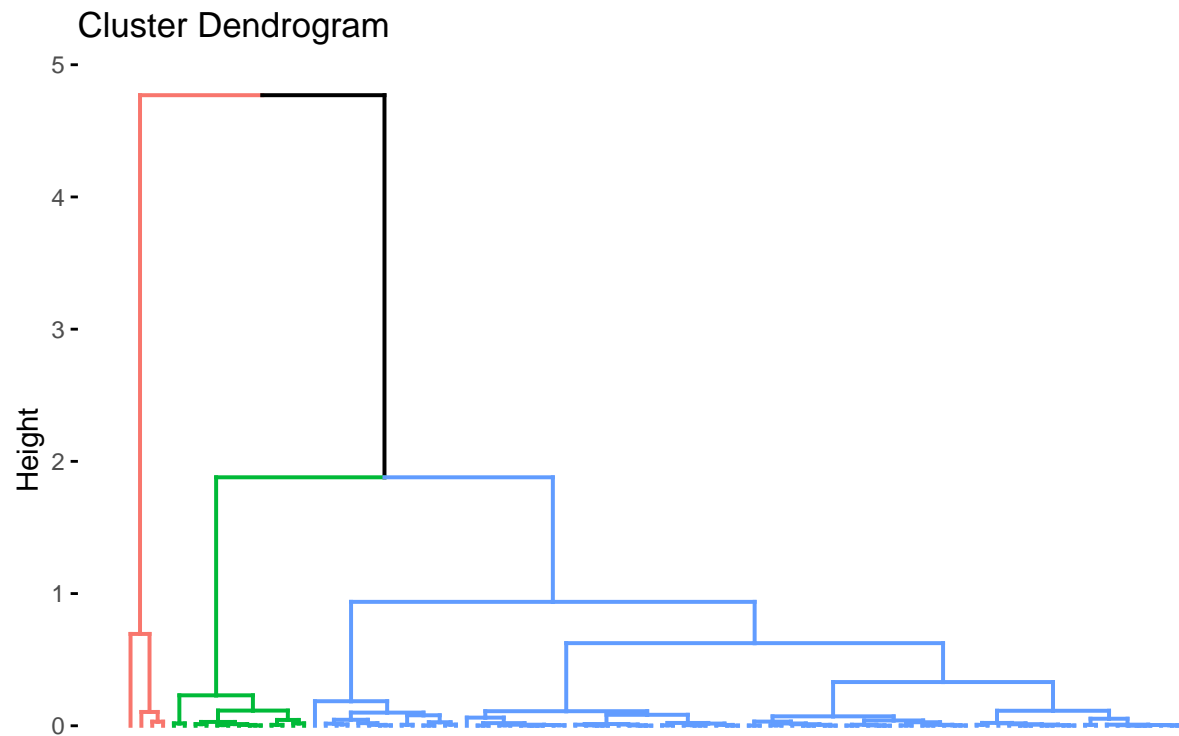
Classification

Dans cette partie, nous réalisons une classification pour explorer nos données. Les questions qui peuvent se poser et qui feront l'objet de l'analyse sont les suivantes :

1. Quels sont les individus les plus caractéristiques pour une partition donnée.
2. Quelles sont les variables qui caractérisent le mieux la partition de chaque classe.

Appliquons la classification hiérarchique sur le résultat de l'ACP

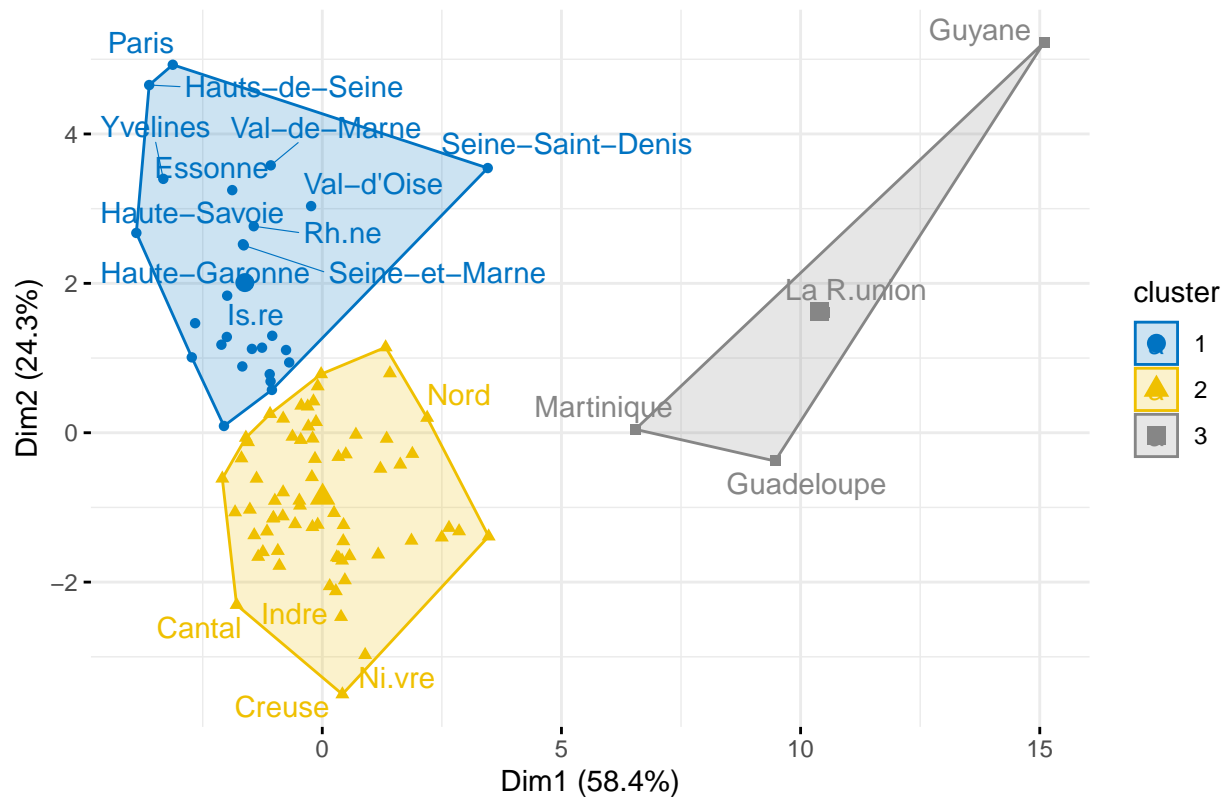
```
res.hcpc <- HCPC(res.pca, nb.clust = 3, graph=FALSE)
test1 = res.hcpc$data.clust[,10:11]
fviz_dend(res.hcpc, show_labels = FALSE)
```



Visualisons le dendrogramme généré par la classification

```
fviz_cluster(res.hcpc,  
  repel = TRUE,  
  show.clust.cent = TRUE,  
  palette = "jco",  
  ggtheme = theme_minimal(),  
  main = "Factor map"  
)
```

Factor map



Description des classes suivant Suivant les individus

res.hcpc\$desc.ind

```
## $para
## Cluster: 1
##      Is<e8>re  Haute-Garonne      Gironde Seine-et-Marne      Bas-Rhin
##      0.6135234    0.8399275    0.9621344    1.0218196    1.0311255
## -----
## Cluster: 2
##      Tarn      Moselle      Haute-Sa<f4>ne      Ard<e8>che
##      0.5473832    0.7469531    0.8237439    0.9239390
## Charente-Maritime
##      0.9263512
## -----
## Cluster: 3
## La R<e9>union      Guadeloupe      Martinique      Guyane
##      1.827388    2.600827    4.493562    7.145107
##
## $dist
## Cluster: 1
##      Paris      Hauts-de-Seine      Seine-Saint-Denis      Haute-Savoie
##      7.306464    6.806355    5.662404    5.592068
##      Yvelines
##      5.427930
## -----
## Cluster: 2
```

```
## Pyr<e9>n<e9>es-Orientales          Creuse          Pas-de-Calais
##          6.281839          6.053965          5.931137
##          Ni<e8>vre          Aisne
##          5.783691          5.487238
## -----
## Cluster: 3
##          Guyane La R<e9>union    Guadeloupe    Martinique
##          16.650833    11.005256    9.734225    7.148728
```

1. Pour la classe 1, les paragon, c'est à dire les départements représentatifs sont *Isère, Haute-Garonne, Gironde, Seine-et-Marne, Bas-Rhin* (ils ressemblent le plus à la moyenne de la classe 1). Les départements de la classe 1 peuvent être résumés par ces 5 départements. Pour la classe 1, les départements les plus caractéristiques c-à-d ceux qui s'éloignent les plus des autres classes sont *Paris, Hauts-de-Seine, Seine-Saint-Denis, Haute-Savoie, Yvelines*.
2. Pour la classe 2, les paragon sont *Tarn, Moselle, Haute-Saône, Ardèche, Bas-Rhin, Charente-Maritime*. Pour la classe 2, les départements les plus caractéristiques sont *Pyrénées-Orientales, Creuse, Pas-de-Calais, Nièvre, Aisne*.
3. Pour la classe 3, le paragon est *reunion*. Pour la classe 2, le département le plus caractéristique est *Guyane*.

Description des classes suivant les variables

```
res.hcpc$desc.var
```

```
##
## Link between the cluster variable and the quantitative variables
## =====
##          Eta2          P-value
## Part.allocataire.RSA.15.64.ans          0.7713140 3.663396e-31
## Part.des.jeunes.en.difficult.e9..de.lecture 0.7485038 3.351547e-29
## Part.d.enfant.dont.les.parent.sans.enfants 0.7392723 1.857340e-28
## Taux.de.pauvret.e9.          0.6803087 2.983405e-24
## Taux.de.chomage.annuel          0.6637944 3.263903e-23
## jeunes.non.inseres          0.5896992 4.191448e-19
## Niveau.de.vie.median          0.5707155 3.592488e-18
## Taux.de.mortalit.e9.          0.5318111 2.213104e-16
## Esperance.de.vie.Hommes          0.4351247 1.651008e-12
## Taux.d.activit.e9..population          0.4234278 4.370689e-12
## Taux.logement.suroccup.e9.          0.3833797 1.061321e-10
## Population.de.moins.de.25.ans          0.3592642 6.565237e-10
## Taux.de.criminalit.e9.          0.2130491 1.141598e-05
##
## Description of each cluster by quantitative variables
## =====
## $`1`
##          v.test Mean in category
## Esperance.de.vie.Hommes          6.243745          80.553846
## Niveau.de.vie.median          5.784359          22449.230769
## Taux.d.activit.e9..population          4.980923          75.653846
## Taux.logement.suroccup.e9.          4.463854          11.726923
## Taux.de.criminalit.e9.          4.346452          46.651154
## Population.de.moins.de.25.ans          4.173190          30.707692
## Part.allocataire.RSA.15.64.ans          -2.480081          3.430769
## Taux.de.pauvret.e9.          -2.554976          12.907692
## Part.des.jeunes.en.difficult.e9..de.lecture -2.712982          9.634615
```

```

## Part.d.enfant.dont.les.parent.sans.enfants -2.780448 9.184615
## Taux.de.chomage.annuel -2.816600 7.288462
## jeunes.non.inseres -5.065482 19.334615
## Taux.de.mortalit.e9. -6.163337 7.815385
## Overall mean sd in category
## Esperance.de.vie.Hommes 79.252041 0.8621221
## Niveau.de.vie.median 20581.131333 2099.2744591
## Taux.d.activit.e9..population 73.822449 1.5665324
## Taux.logement.suroccup.e9. 7.365306 6.9865562
## Taux.de.criminalit.e9. 37.964490 16.2203444
## Population.de.moins.de.25.ans 28.100000 2.2163593
## Part.allocataire.RSA.15.64.ans 4.591837 1.1808632
## Taux.de.pauvret.e9. 15.236510 3.5788828
## Part.des.jeunes.en.difficult.e9..de.lecture 12.350000 2.1375655
## Part.d.enfant.dont.les.parent.sans.enfants 12.229592 2.3054605
## Taux.de.chomage.annuel 8.595918 1.0839128
## jeunes.non.inseres 24.424490 3.0605357
## Taux.de.mortalit.e9. 10.300000 1.4311621
## Overall sd p.value
## Esperance.de.vie.Hommes 1.233977 4.272155e-10
## Niveau.de.vie.median 1911.397547 7.278924e-09
## Taux.d.activit.e9..population 2.176101 6.328168e-07
## Taux.logement.suroccup.e9. 5.782876 8.049857e-06
## Taux.de.criminalit.e9. 11.828357 1.383573e-05
## Population.de.moins.de.25.ans 3.698235 3.003639e-05
## Part.allocataire.RSA.15.64.ans 2.770754 1.313526e-02
## Taux.de.pauvret.e9. 5.394547 1.061950e-02
## Part.des.jeunes.en.difficult.e9..de.lecture 5.923668 6.668073e-03
## Part.d.enfant.dont.les.parent.sans.enfants 6.481500 5.428401e-03
## Taux.de.chomage.annuel 2.747315 4.853488e-03
## jeunes.non.inseres 5.946926 4.073674e-07
## Taux.de.mortalit.e9. 2.385885 7.122786e-10
##
## $`2`
## v.test Mean in category Overall mean
## Taux.de.mortalit.e9. 7.147069 11.45000 10.300000
## jeunes.non.inseres 2.190306 25.30294 24.424490
## Taux.d.activit.e9..population -2.788373 73.41324 73.822449
## Niveau.de.vie.median -3.239598 20163.52941 20581.131333
## Taux.de.criminalit.e9. -4.499746 34.37500 37.964490
## Esperance.de.vie.Hommes -4.884076 78.84559 79.252041
## Population.de.moins.de.25.ans -5.554273 26.71471 28.100000
## Taux.logement.suroccup.e9. -5.808483 5.10000 7.365306
## sd in category Overall sd p.value
## Taux.de.mortalit.e9. 1.6422277 2.385885 8.865045e-13
## jeunes.non.inseres 3.8757910 5.946926 2.850202e-02
## Taux.d.activit.e9..population 1.5484247 2.176101 5.297353e-03
## Niveau.de.vie.median 697.8322837 1911.397547 1.196982e-03
## Taux.de.criminalit.e9. 7.2681397 11.828357 6.803473e-06
## Esperance.de.vie.Hommes 0.9513819 1.233977 1.039150e-06
## Population.de.moins.de.25.ans 2.4999567 3.698235 2.787693e-08
## Taux.logement.suroccup.e9. 2.0542639 5.782876 6.304142e-09
##
## $`3`

```

	v.test	Mean in category
## Part.allocataire.RSA.15.64.ans	8.530055	16.22500
## Part.des.jeunes.en.difficult.e9..de.lecture	8.351417	36.70000
## Part.d.enfant.dont.les.parent.sans.enfants	8.281632	38.65000
## Taux.de.pauvret.e9.	7.968406	36.39451
## Taux.de.chomage.annuel	7.804839	19.15000
## jeunes.non.inseres	6.200807	42.57500
## Population.de.moins.de.25.ans	3.625780	34.70000
## Taux.logement.suroccup.e9.	3.569346	17.52500
## Esperance.de.vie.Hommes	-2.555335	77.70000
## Taux.de.mortalit.e9.	-2.895218	6.90000
## Taux.d.activit.e9..population	-4.619067	68.87500
## Niveau.de.vie.median	-5.360749	15537.71767
	Overall mean	sd in category
## Part.allocataire.RSA.15.64.ans	4.591837	2.2264041
## Part.des.jeunes.en.difficult.e9..de.lecture	12.350000	10.8625503
## Part.d.enfant.dont.les.parent.sans.enfants	12.229592	8.7514285
## Taux.de.pauvret.e9.	15.236510	6.2839876
## Taux.de.chomage.annuel	8.595918	2.5034976
## jeunes.non.inseres	24.424490	6.2655307
## Population.de.moins.de.25.ans	28.100000	8.7478569
## Taux.logement.suroccup.e9.	7.365306	10.7792335
## Esperance.de.vie.Hommes	79.252041	0.9246621
## Taux.de.mortalit.e9.	10.300000	2.4768932
## Taux.d.activit.e9..population	73.822449	3.1909050
## Niveau.de.vie.median	20581.131333	1225.2609901
	Overall sd	p.value
## Part.allocataire.RSA.15.64.ans	2.770754	1.462781e-17
## Part.des.jeunes.en.difficult.e9..de.lecture	5.923668	6.744889e-17
## Part.d.enfant.dont.les.parent.sans.enfants	6.481500	1.214981e-16
## Taux.de.pauvret.e9.	5.394547	1.607340e-15
## Taux.de.chomage.annuel	2.747315	5.957742e-15
## jeunes.non.inseres	5.946926	5.617430e-10
## Population.de.moins.de.25.ans	3.698235	2.880908e-04
## Taux.logement.suroccup.e9.	5.782876	3.578731e-04
## Esperance.de.vie.Hommes	1.233977	1.060855e-02
## Taux.de.mortalit.e9.	2.385885	3.788952e-03
## Taux.d.activit.e9..population	2.176101	3.854693e-06
## Niveau.de.vie.median	1911.397547	8.287744e-08

On peut dire que :

La classe 1 qui est composée des individus telles que Paris,Hauts-de-Seine, Seine-Saint-Denis etc., est caractérisé par :

- des valeurs supérieures a la moyenne *Esperance de vie Hommes,Niveau de vie median...*
- des valeurs inférieures a la moyenne *Taux de mortalité,jeunes non inserés...*

La classe 2 qui est composée des individus telles que Creuse,Pas-de-Calais,Nièvre etc., est caractérisé par :

- des valeurs supérieures a la moyenne *Taux de mortalité,jeunes non inserés...*
- des valeurs inférieures a la moyenne *Taux de logements suroccupé,population de moins de 25 ans...*

La classe 3 qui est composée des individus telles que reunion,Guyane,Martinique etc., est caractérisé par :

- des valeurs superieures a la moyenne *Part allocataire RSA 15-64 ans,Part des jeunes en difficulté de*

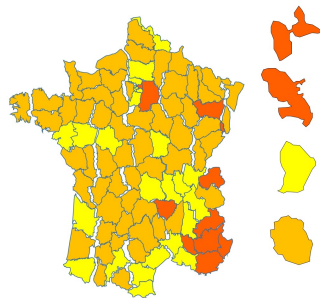
lecture...

- des valeurs inferieures a la moyenne *Niveau de vie median, Taux d'activité population.*

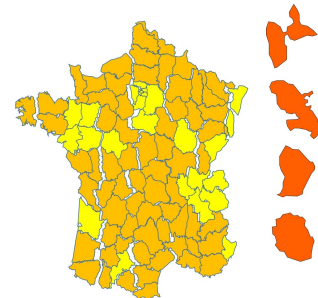
Conclusion

```
image1<- image_read("image/carte criminalite.JPG")
image2<- image_read("image/carte indicateurs sociaux.JPG")
par(mfrow=c(1,2))
plot(image1)
title(main = "carte classification criminalite",font.main= 1)
plot(image2)
title(main = "carte classification indicateurs",font.main= 1)
```

carte classification criminalite



carte classification indicateurs



Pour chacune de nos analyses ci-dessus, nous avons eu à remarquer qu'il existe des différences entre les départements peu importe l'axe d'analyse. En ce qui concerne la criminalité, nous avons remarqué qu'il existe 3 profils ou types de département qui sont :

- **Groupe 1(jaune)** : les départements où il y a une plus de *vols et cambriolages* que la moyenne, mais moins de *crimes et délits violents* Ex : Paris, Gironde, Loire-Atlantique
- **Groupe 2(rouge)** : les départements où il y a plus de *ventes de stupéfiants et l'usage ou détention d'armes prohibées* que la moyenne, mais moins de *vols et cambriolages* Ex : Seine-Saint-Denis, Guadeloupe, Martinique
- **Groupe 3(orange)** : puis les départements où il y a plus de *destruction, dégradation de biens, falsification, escroquerie* que la moyenne, mais moins de *Vols et cambriolage ou détention d'armes*

prohibé Ex : Pas-de-Calais, Haut-Rhin, La Réunion

La répartition de ces 3 différents profils sont visibles sur la carte de criminalité ci-dessus.

En ce qui concerne les indicateurs sociaux, nous avons remarqué qu'il existe aussi 3 profils ou type de département qui sont :

- **Groupe 1(jaune)** : les départements où il y a une plus grande *espérance de vie Hommes* et un *niveau de vie médian supérieur* a la moyenne Ex : Paris, Hauts-de-Seine, Seine-Saint-Denis
- **Groupe 2(orange)** : les départements où il y a un plus grand *taux de mortalité* et de *jeunes non insérés* que la moyenne, mais *moins de logements suroccupés* ou de *population jeune* Ex : Creuse, Pas-de-Calais, Nièvre
- **Groupe 3(rouge)** : puis les départements où il y a plus de *Part allocataire RSA 15-64 ans*, *Part des jeunes en difficulté de lecture* que la moyenne Ex : Réunion, Guyane, Martinique

La répartition de ces 3 différents profils sont visibles sur la carte d'indicateur sociaux ci-dessus.

Nous avons remarqué la plupart des départements qui ont généralement les mêmes types de criminalités sont également regroupés suivant les mêmes critères sociaux Cf : la coloration des cartes.

Cela pourrait laisser à penser que le niveau de développement d'un département a un impact sur le type de crimes ou délits commis même si bien évidemment une conclusion ne peut pas encore être tirée.

Les départements les plus développés comme Paris connaissent plus de vols et cambriolages que la moyenne, les départements avec des scores dans la moyenne auront plus de destructions et dégradation des biens et les départements les moins développés comme la Martinique, nous avons plus de ventes de stupéfiants ou détention ou usage d'arme prohibés.

```
departement <- row.names(res.hcpc$data.clust)
clust_indc = cbind(departement,subset(res.hcpc$data.clust,select = "clust"))

names(clust_indc) <- c("departement","indic_sociaux")
clust_indc <- data.frame(lapply(clust_indc, function(x) {
  gsub("1", "indic_bon", x)
}))
clust_indc <- data.frame(lapply(clust_indc, function(x) {
  gsub("2", "indic_moyen", x)
}))
clust_indc <- data.frame(lapply(clust_indc, function(x) {
  gsub("3", "indic_faible", x)
}))

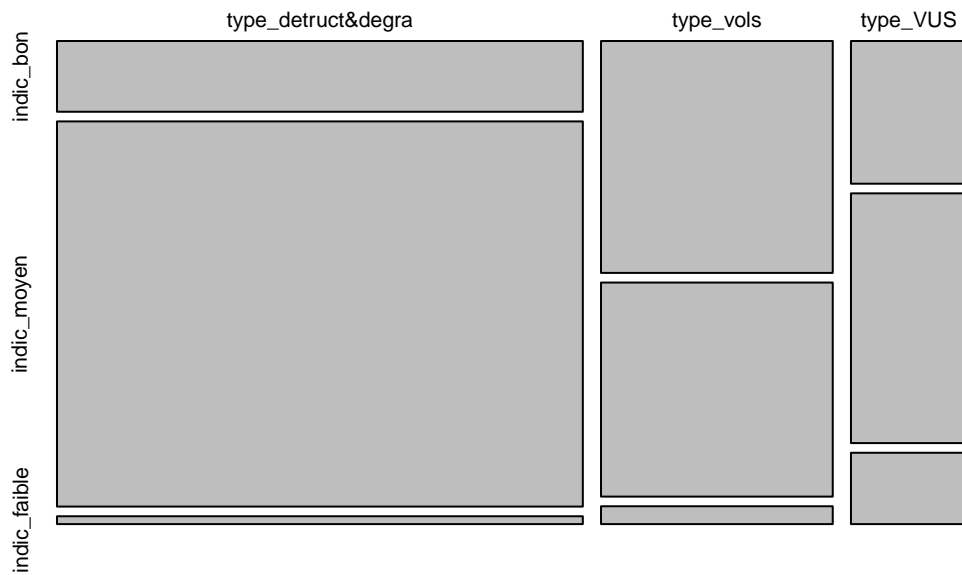
matrice <- merge(clust_indc,clust_crim,by = "departement")

contingence<-table(matrice$Type_criminalite,factor(matrice$indic_sociaux, levels = c("indic_bon", "indic_moyen", "indic_faible")))
contingence

##
##               indic_bon indic_moyen indic_faible
## type_detruct&degra      9         49          1
## type_vols              13         12          1
## type_VUS                4          7          2

plot(contingence)
```

contingence



```
fisher <- fisher.test(contingence)
fisher
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  contingence
## p-value = 0.001086
## alternative hypothesis: two.sided
```

Nous utilisons un test de Fisher, car nous avons un effectif trop faible pour le test du χ^2 , mais la conclusion est la même avec une p value de 0.001086, nous rejetons l’hypothèse d’indépendance des variables : **Le type de crimes ou délits commis dans un département est lié à son “niveau de développement”**