

Facteurs influençants la prise des transports en commun pour l'agglomération grenobloise

RACHIDI Mustapha & SAUNIER Florent & SAADALLAH Malek

Janvier 2023

```
#J'ai tenté l'image  
knitr:::include_graphics("Image_BUS_TRAM.jpg")
```



Introduction

Ce projet se base sur des données récoltées en 2010 dans la région Grenobloise. L'étude a pour but de déterminer les facteurs influençant la prise des transports en commun . Pour cela nous nous sommes pris comme limites : le réseau Mtag qui comprend les bus qualifiés de “ville” (Nous n'avons pas pris en compte les bus régionaux comme par exemple le bus Grenoble - Chamrousse) et le réseau du tramway dont les lignes depuis 2010 ont été augmentées.

Articles de la littérature

Familiarisation avec la base de données

La base de données contient 30 702 lignes et 116 colonnes ce qui correspond à nos variables , on peut la qualifier de base de données “moyenne” mais qui saura nous occuper. Concernant le nombre de valeurs manquantes, toutes variables confondues nous avons 971 658 valeurs manquantes soit 27.3% de notre base de données. De plus, 0% des lignes ont toutes leurs valeurs et c'est 21% des colonnes qui n'ont pas de valeurs manquantes. Il peut être intéressant de voir où sont les valeurs manquantes.

L'échantillon comporte 5189 personnes

```
library(cluster)
library(factoextra)

## Le chargement a nécessité le package : ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(FactoMineR)
```

Visualisation valeurs manquantes titre à changer peut être

En annexe, quelques graphiques permettant de visualiser quelles variables ont le plus de valeurs manquantes. Ces graphiques nous permettront d'adopter un regard critique sur les variables que nous choisirons par la suite. Cependant, on peut établir quelques critères avec r : ration de valeurs manquantes dans la colonne.

Bon : r<=5% Moyen : 5%<r<=20% Mauvais 20%<r<=45% Très mauvais : r>45%

Plusieurs variables ont entre 80% 99% de valeurs manquantes JAI TROUVÉ PQ c'est jusque que beaucoup de gens n'ont tout simplement pas plus de 1 véhicule, ce qui fait que les variables correspondantes sont vides.
À CHANGER

```
sample <- sample(c(TRUE,FALSE),size=nrow(DB_projet_full),replace=TRUE,prob=c(0.7,0.3))
DB_BBB <-DB_projet_full[!sample,]
DB_BBB$freqtcu2<-DB_BBB$freqtcu

#DB_reduced$car_ownership_2<-c(1,nrow(DB_reduced))
DB_reduced$car_ownership_2<-ifelse(DB_reduced$dispovp==1 | DB_reduced$dispovp==2,"OUI","NON")

table(DB_reduced$car_ownership_2)

##
##  NON  OUI
##  866 3910
```

Variables du projet

Freqtcu : Variable d'intérêt (Y) catégorielle qui indique la fréquence d'utilisation des transports en communs chez une personne.

Elle prend les valeurs :

- 1 : Utilisation des transports en commun tous les jours
- 2 : Utilisation des transports en commun au moins deux fois par semaine
- 3 : Utilisation des transports en commun au moins deux fois par mois
- 4: Utilisation des transports en commun très rare
- 5 : Utilisation des transports en commun inexistante

Nous avons décidé de construire freqtcu de manière à ce qu'elle prenne la valeur 0 ou 1

```
DB_projet_full$freqtcu2<-DB_projet_full$freqtcu  
DB_projet_full<-DB_projet_full%>%mutate(freqtcu=ifelse(freqtcu<=3,1,0))  
DB_projet_full$freqtcu<-factor(DB_projet_full$freqtcu)
```

Pour toutes les personnes qui prennent les transports de manière : régulière/tous les jours , au moins deux fois par semaine et au moins deux fois par mois se sont vues attribuées la valeur 1 car le “au moins” présage une prise des transports en communs plus élevée.

Tailmng : Variable qui indique le nombre de personnes composant le ménage.

```
DB_projet_full<-rename(DB_projet_full, "tailmng"="NO_PERS")
```

On change simplement le nom de la variable “NO_PERS” qui indique le nombre de personne dans le ménage

Permis : Variable indiquant si la personne effectuant le trajet possède le permis ou pas.

```
DB_projet_full<-DB_projet_full%>%mutate(permis=ifelse(any(permis==1 | permis==3),"YES","NO"))  
DB_projet_full$permis<-factor(DB_projet_full$permis)
```

Car_ownership : Variable indiquant si la personne effectuant le trajet possède une voiture pour ses déplacements

```
#DB_projet_full<-DB_projet_full%>%mutate(car_ownership=ifelse(DB_projet_full$VP_DISPO>0 & (DB_projet_full$GENRE=="voiture" | DB_projet_full$GENRE=="camion" | DB_projet_full$GENRE=="autre"),1,0))  
  
#DB_projet_full$car_ownership<-factor(DB_projet_full$car_ownership)  
  
DB_projet_full<-DB_projet_full%>%mutate(car_ownership=ifelse(DB_projet_full$dispovp==1 | DB_projet_full$dispovp==2,1,0))  
DB_projet_full$car_ownership<-factor(DB_projet_full$car_ownership)  
  
#table(DB_reduced$car_ownership_2)
```

Cette variable dépend de trois variables qui sont VP_dispo qui doit être strictement supérieur à 0, puis GENRE (type de véhicule utilisé) , nous avons exclu les campings cars car notre sujet se prête au milieu urbain et de POSSE (Est ce que la voiture appartient à la personne). Nous nous sommes contentés de prendre exclusivement les véhicules possédés par la personne.

Création de la nouvelle base de données

Variables complémentaires

Grâce aux variables précédentes et aux articles que l'on a trouvé dans la littérature, nous allons construire notre base de données pour notre modèle.

Nous exploiterons un ensemble de caractéristiques socio-économiques puis certaines variables liées au “confort” du trajet.

Cette base de données va regrouper toutes les variables nécessaires à notre étude puis on la redévisera encore par la suite pour notamment l'utilisation de cluster

Restriction géographique

Définissons ce que l'on entend par “transports en communs”.

Pour notre étude nous nous concentrerons sur les transports en communs de la société MTag, c'est à dire les tram et bus du réseau.

Notre délimitation géographique sera simplement les terminaux des différentes lignes de tram/bus confondues.

Par la suite, quand on parlera de transports en communs, on se réfère à la définition au dessus.

Toutes les zones répertoriées dans le vecteur “Vec_zone” ont au moins un arrêt du réseau Mtag.

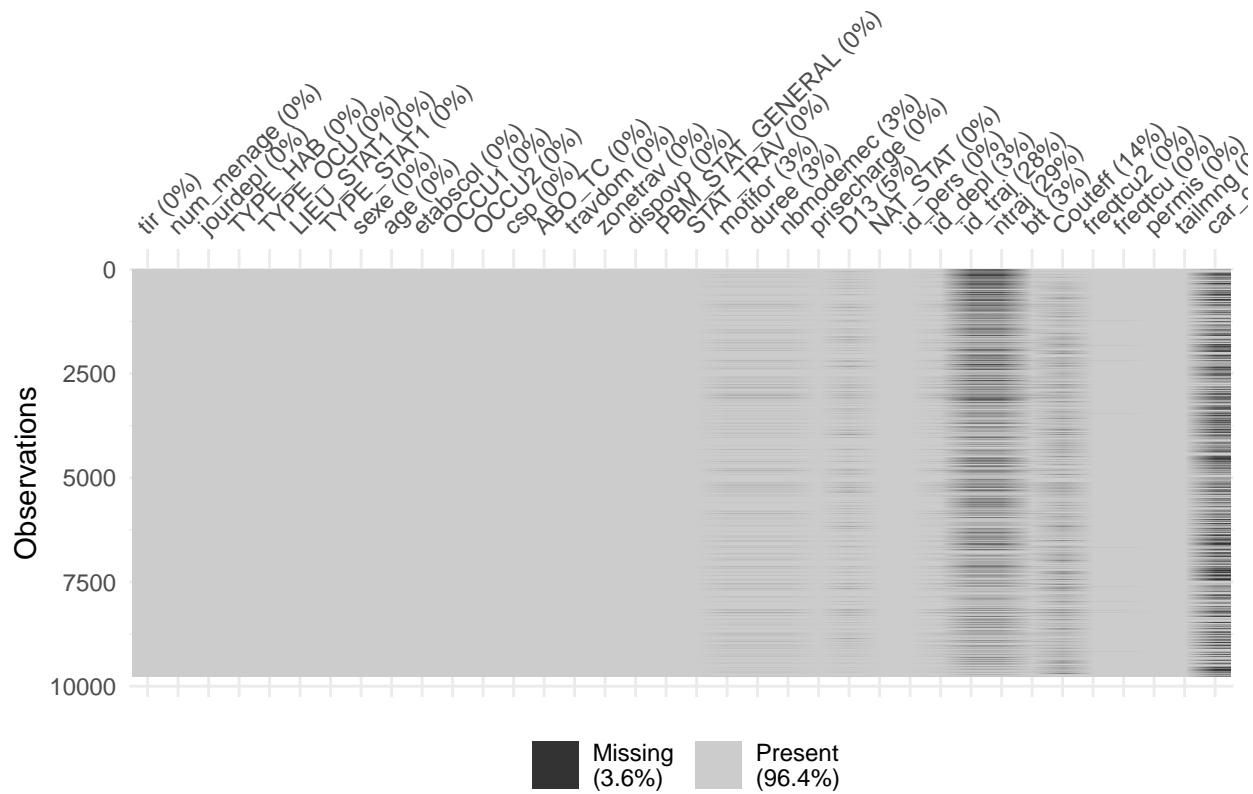
Restriction sur l'âge

Il est nécessaire de préciser que les mineurs se déplacent majoritairement via les transports en communs car ils n'ont tout simplement pas le choix...

Et aussi les étudiants car après plusieurs recherches, nous savons que les étudiants de l'étude utilisent majoritairement les transports en commun (plus de 80% d'entre eux), on détaillera cela un peu plus loin dans le rapport.

Revue de la base de données Il va falloir enlever les variables où il y a des valeurs manquantes

```
vis_miss(New_DB_filtered)
```



```
#ntraj 28% on va certainement l'enlever
```

```
##Clustering Les étudiants
```

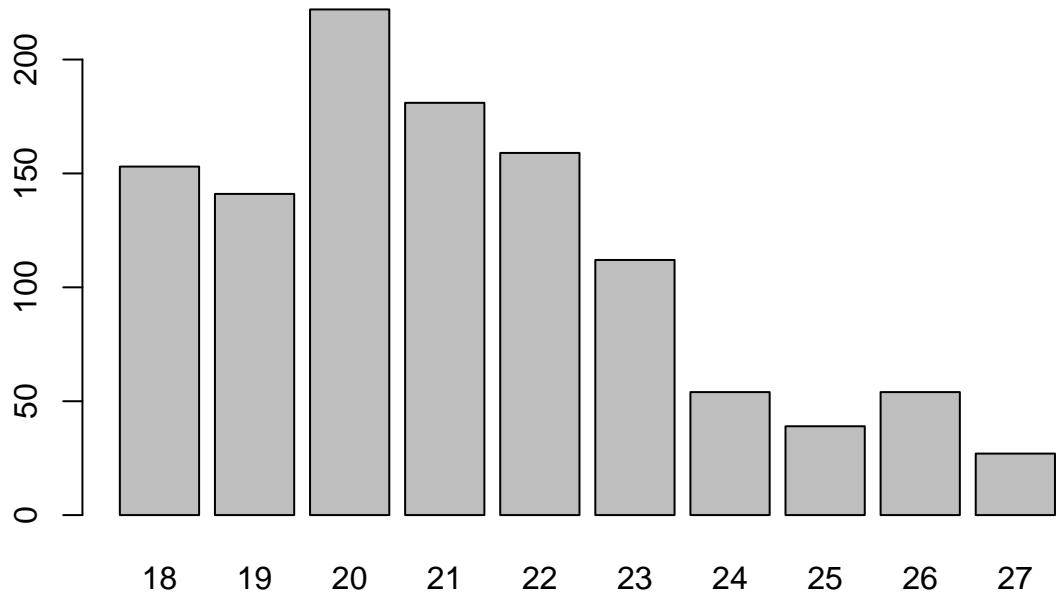
Test 0 Étudiants

```
New_DB_filtered_etu<-dplyr::filter(New_DB,csp=="ETUDES")
New_DB_filtered_etu<-dplyr::filter(New_DB,OCCU1=="Etudiant", age %in% c(18:27))
#New_DB_filtered_etu<-dplyr::filter(New_DB,OCCU2=="NON_Con")
```

```
DB_pers_socio_etu<-New_DB_filtered_etu[,c(1,8,9,10,11,13,14,15,16,17,26,33,34,35,36)]
DB_pers_socio_etu<-DB_pers_socio_etu[!duplicated(DB_pers_socio_etu$id_pers),]
DB_pers_socio_etu<-dplyr::filter(DB_pers_socio_etu,!is.na(DB_pers_socio_etu$car_ownership))
DB_pers_socio_etu<-dplyr::filter(DB_pers_socio_etu,!is.na(DB_pers_socio_etu$age))
```

```
#vis_miss(DB_pers_socio_etu)
```

```
#table_2<-na.omit(DB_pers_socio_etu)
barplot(table(New_DB_filtered_etu$age))
```

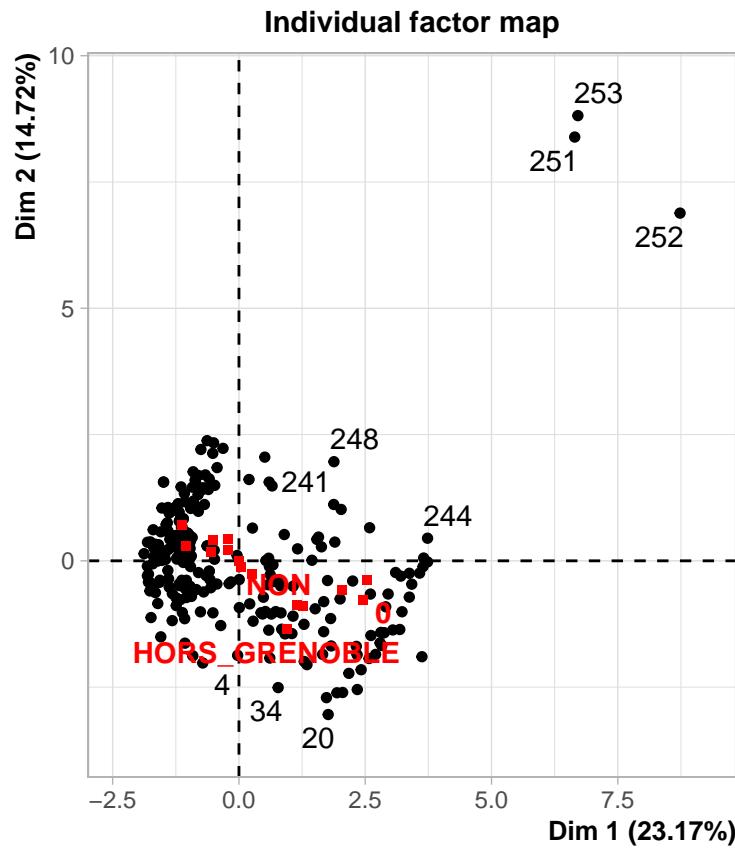


Les étudiants vont de 18 à 27 ans car au dessus on pouvait considérer certaines personnes comme des outliers , des personnes plus agés en reconversion certainement.

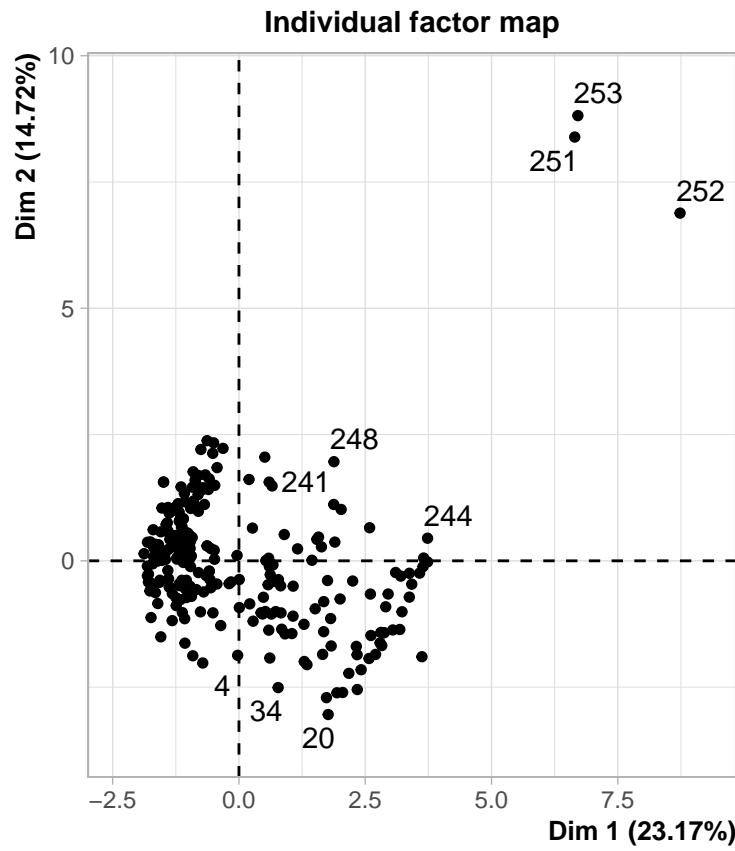
```
afm_etu<-FAMD(base=DB_pers_socio_etu,ncp=5)
```

```
## Warning: ggrepel: 244 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
## Warning: ggrepel: 17 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

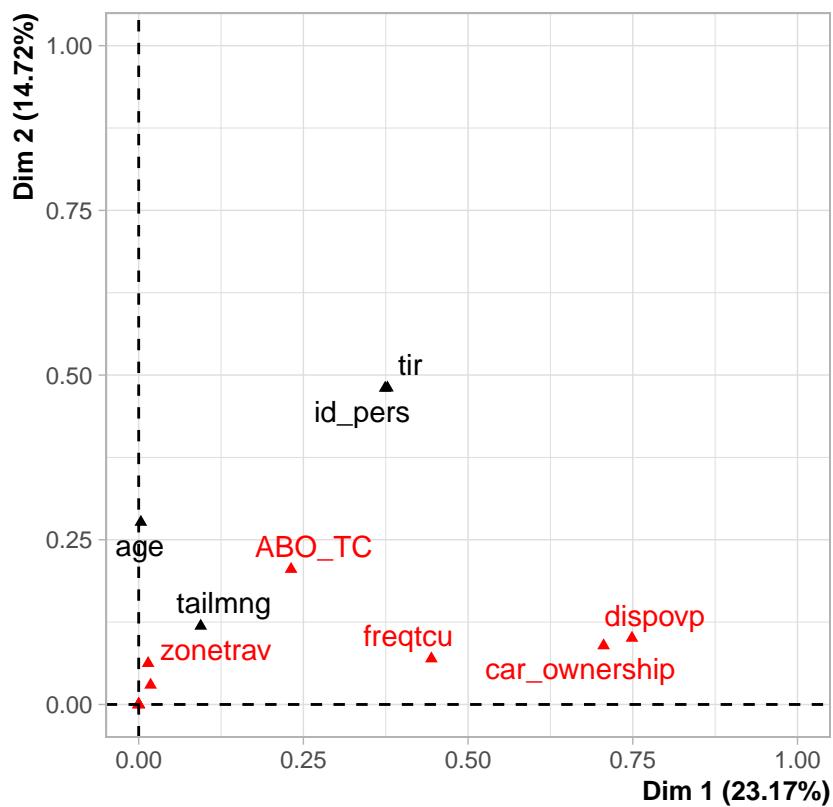


```
## Warning: ggrepel: 244 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

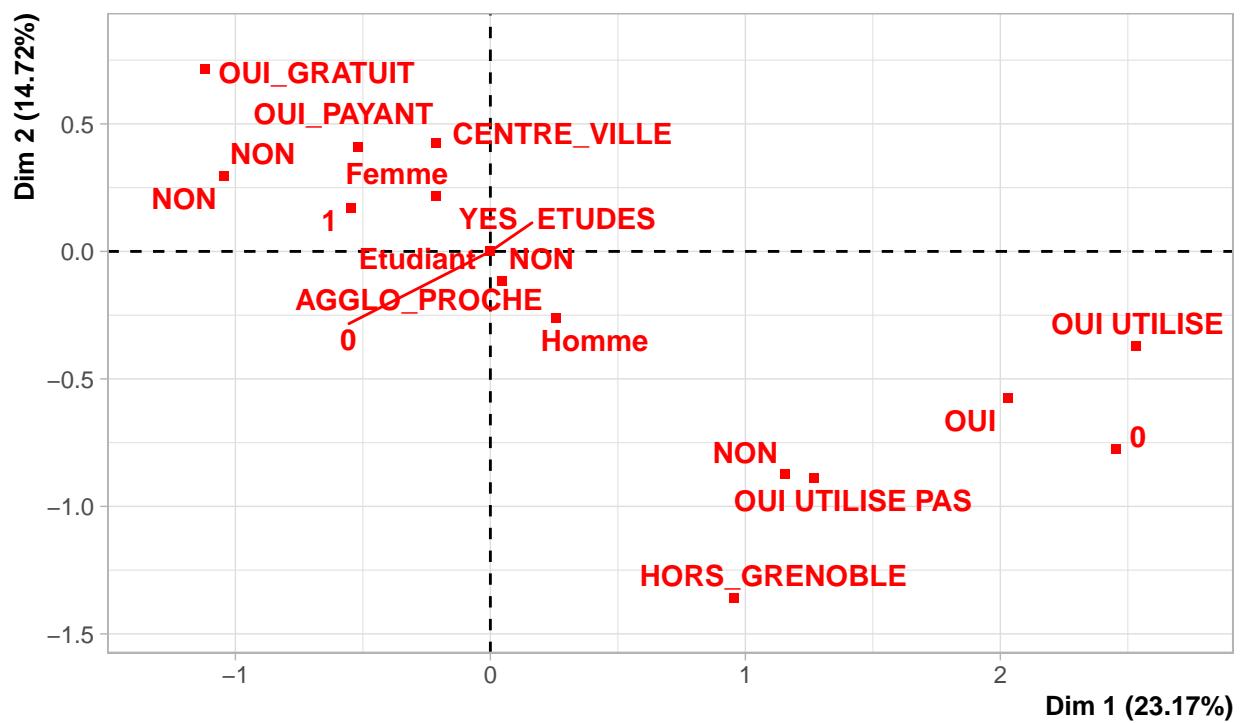


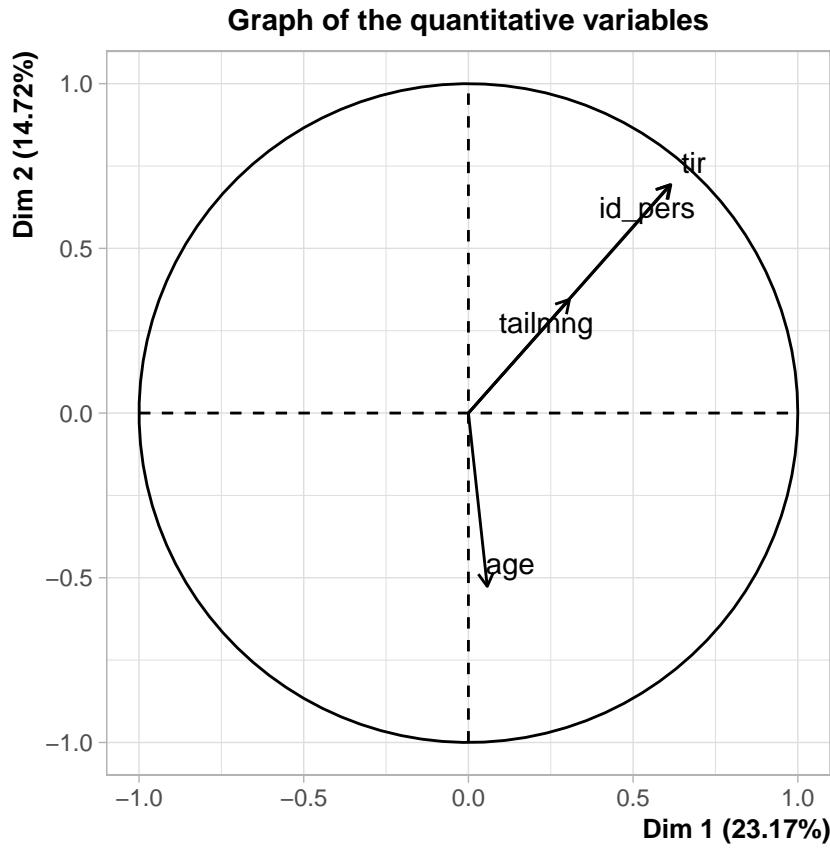
```
## Warning: ggrepel: 6 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

Graph of the variables



Graph of the categories





```
summary(afm_etu)
```

```
##
## Call:
## FAMD(base = DB_pers_socio_etu, ncp = 5)
##
##
## Eigenvalues
##              Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
## Variance      3.012   1.914   1.394   1.218   1.153
## % of var. 23.166 14.721 10.726  9.372   8.867
## Cumulative % of var. 23.166 37.887 48.613 57.985 66.852
##
## Individuals (the 10 first)
##          Dist    Dim.1     ctr    cos2    Dim.2     ctr    cos2    Dim.3
## 1       | 2.820 | -1.593  0.333  0.319 |  0.320  0.021  0.013 |  0.387
## 2       | 2.742 | -1.708  0.383  0.388 | -0.060  0.001  0.000 |  0.088
## 3       | 2.642 | -1.521  0.303  0.331 |  0.205  0.009  0.006 | -0.365
## 4       | 6.228 | -0.717  0.067  0.013 | -2.023  0.845  0.106 | -0.239
## 5       | 3.077 | -1.740  0.397  0.320 | -1.123  0.260  0.133 | -0.192
## 6       | 2.270 | -1.812  0.431  0.637 | -0.295  0.018  0.017 | -0.353
## 7       | 4.445 | -1.878  0.463  0.179 |  0.142  0.004  0.001 | -0.500
## 8       | 2.376 | -1.762  0.408  0.550 | -0.596  0.073  0.063 | -0.279
## 9       | 2.452 | -1.607  0.339  0.429 | -0.846  0.148  0.119 |  0.263
## 10      | 2.126 | -1.603  0.337  0.569 |  0.000  0.000  0.000 | -0.796
```

```

##          ctr   cos2
## 1      0.042 0.019 |
## 2      0.002 0.001 |
## 3      0.038 0.019 |
## 4      0.016 0.001 |
## 5      0.010 0.004 |
## 6      0.035 0.024 |
## 7      0.071 0.013 |
## 8      0.022 0.014 |
## 9      0.020 0.012 |
## 10     0.180 0.140 |

##
## Continuous variables
##           Dim.1    ctr   cos2   Dim.2    ctr   cos2   Dim.3    ctr
## tir        | 0.614 12.538 0.378 | 0.693 25.118 0.481 | 0.195 2.719
## age        | 0.057 0.109 0.003 | -0.526 14.463 0.277 | 0.088 0.549
## id_pers    | 0.612 12.422 0.374 | 0.693 25.094 0.480 | 0.199 2.831
## tailmng   | 0.307 3.127 0.094 | 0.345 6.222 0.119 | -0.504 18.232
##
##          cos2
## tir        0.038 |
## age        0.008 |
## id_pers    0.039 |
## tailmng   0.254 |
##
## Categories (the 10 first)
##           Dim.1    ctr   cos2 v.test   Dim.2    ctr   cos2 v.test
## Femme      | -0.213 0.273 0.051 -2.133 | 0.217 0.702 0.053 2.729 |
## Homme      | 0.255 0.327 0.051 2.133 | -0.261 0.843 0.053 -2.729 |
## 0          | 0.000 0.000 0.000 0.000 | 0.000 0.000 0.000 0.000 |
## Etudiant   | 0.000 0.000 0.000 0.000 | 0.000 0.000 0.000 0.000 |
## ETUDES     | 0.000 0.000 0.000 0.000 | 0.000 0.000 0.000 0.000 |
## NON         | 1.154 4.933 0.448 7.509 | -0.874 7.016 0.257 -7.138 |
## OUI_GRATUIT | -1.119 0.982 0.088 -2.832 | 0.717 0.998 0.036 2.277 |
## OUI_PAYANT  | -0.520 1.766 0.295 -5.737 | 0.410 2.715 0.183 5.671 |
## NON         | 0.000 0.000 0.000 0.000 | 0.000 0.000 0.000 0.000 |
## AGGLO_PROCHE | 0.045 0.015 0.004 0.603 | -0.117 0.254 0.027 -1.956 |
##
##           Dim.3    ctr   cos2 v.test
## Femme      -0.344 3.311 0.132 -5.059 |
## Homme      0.412 3.973 0.132 5.059 |
## 0          0.000 0.000 0.000 0.000 |
## Etudiant   0.000 0.000 0.000 0.000 |
## ETUDES     0.000 0.000 0.000 0.000 |
## NON         -0.266 1.222 0.024 -2.543 |
## OUI_GRATUIT -0.615 1.382 0.027 -2.287 |
## OUI_PAYANT  0.224 1.536 0.055 3.641 |
## NON         0.000 0.000 0.000 0.000 |
## AGGLO_PROCHE -0.325 3.690 0.208 -6.364 |

head(afm_etu$ind$coord,10)

```

```

##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## 1 -1.5931764 0.319504403 0.38675300 1.275095273 -0.4578470
## 2 -1.7083374 -0.060172320 0.08776164 -0.006859689 -0.3691743
## 3 -1.5206611 0.204561937 -0.36520258 0.310298831 -0.1026692

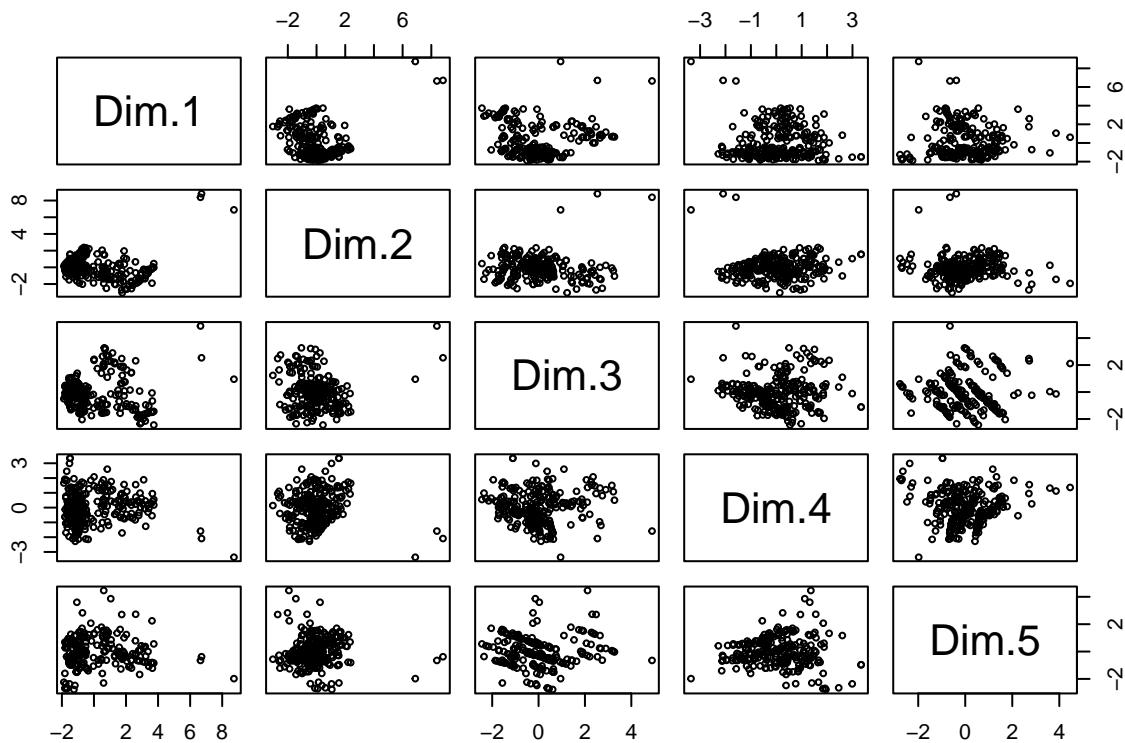
```

```

## 4 -0.7169158 -2.022966707 -0.23905604 0.253017747 2.8279816
## 5 -1.7402926 -1.123014890 -0.19192426 -1.692926272 0.5651522
## 6 -1.8120502 -0.294949841 -0.35328506 -0.622893664 0.6849146
## 7 -1.8781387 0.141669676 -0.50029122 1.661758106 -2.2401906
## 8 -1.7622138 -0.596231271 -0.27883465 -1.064898214 0.6323107
## 9 -1.6065736 -0.845681656 0.26326464 -1.096852898 -0.4956236
## 10 -1.6032158 -0.000235632 -0.79633131 -0.319743388 0.9467153

```

```
pairs(afm_etu$ind$coord, cex=0.6)
```

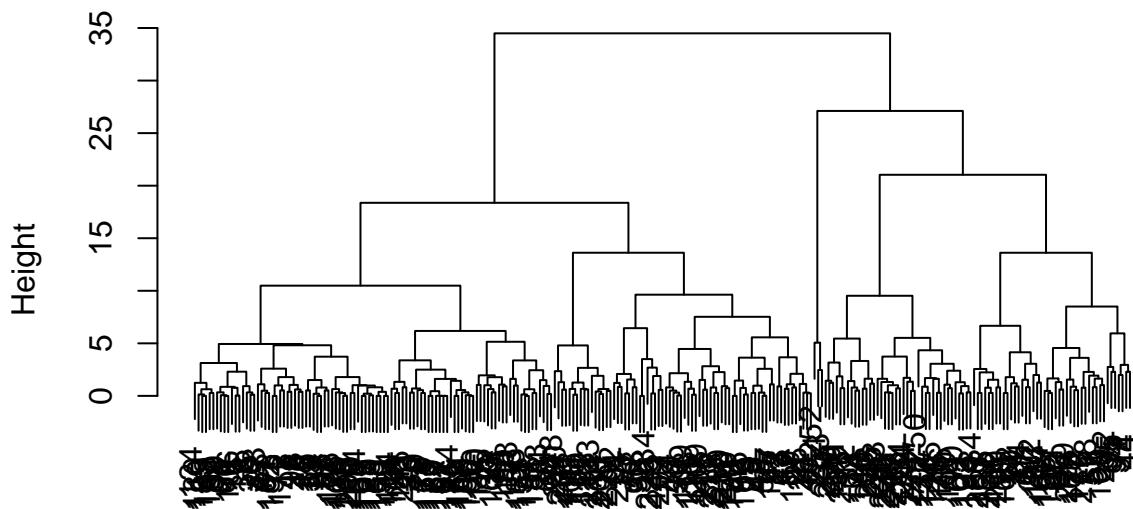


```

md_etu<-dist(afm_etu$ind$coord)
ch_etu<-hclust(md_etu,method='ward.D2')
plot(ch_etu)

```

Cluster Dendrogram



```
md_etu  
hclust (*, "ward.D2")
```

```
groups_etu<-cutree(ch_etu,k=2)
```

```
table(groups_etu)
```

```
## groups_etu  
##   1   2  
## 167  86
```

```
pairs(afm_etu$ind$coord,col=c('cornflowerblue','yellowgreen')[groups_etu],cew=0.75)
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
```

```
## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
```

```
## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un  
## paramètre graphique
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

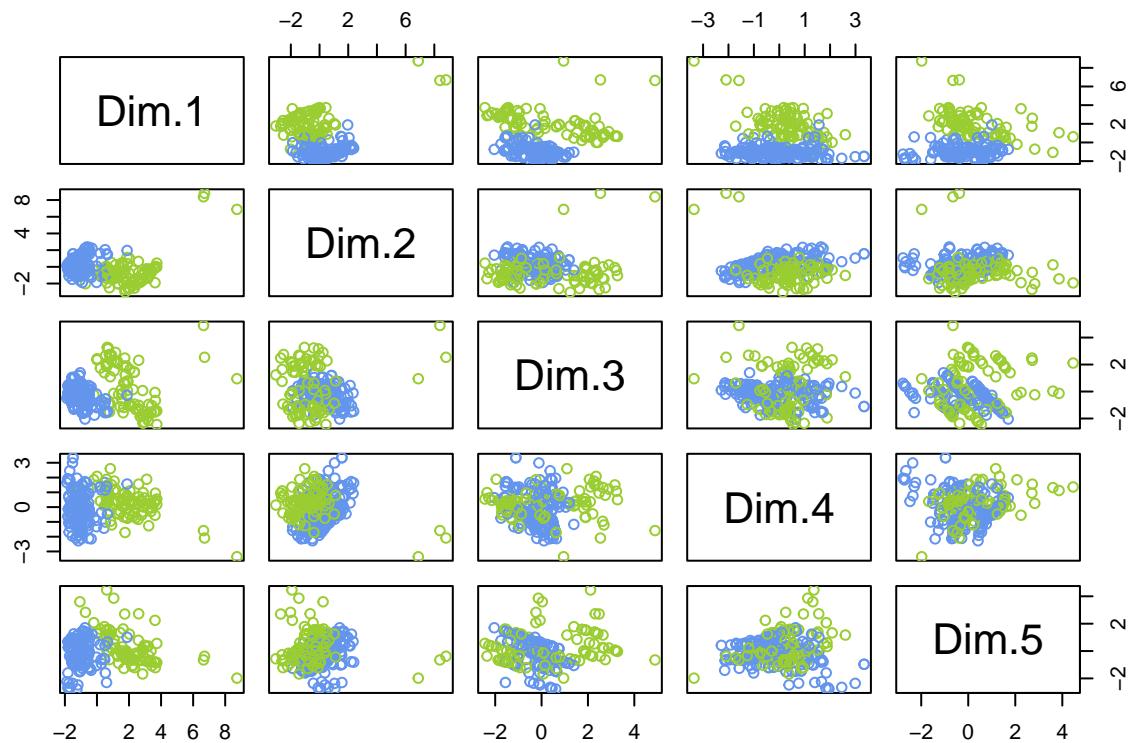
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

```



```

sapply(DB_pers_socio_etu[sapply(DB_pers_socio_etu,is.numeric)],mean)

##          tir         age      id_pers     tailmng
## 1.193794e+02 2.122530e+01 1.196606e+08 1.739130e+00

lapply(DB_pers_socio_etu[sapply(DB_pers_socio_etu,is.numeric)],function(x){tapply(x,groups_etu,mean)})

## $tir
##      1      2
## 115.4491 127.0116
##
## $age
##      1      2
## 21.08982 21.48837
##
## $id_pers
##      1      2
## 115771990 127211723
##
## $tailmng
##      1      2
## 1.646707 1.918605

sapply(DB_pers_socio_etu[sapply(DB_pers_socio_etu,is.factor)],function(x){prop.table(table(x))})

## $sexe
## x
##   Femme    Homme
## 0.5454545 0.4545455
##
## $etabscol
## x
##      0          8 APPRENTISSAGE          BAC      BAC >= 3
##      1          0          0          0          0
##    BAC+2        BEP          CAP    PAS DIPLOME
##      0          0          0          0
##
## $OCCU1
## x
## Appreti_Stage       Autre      Etudiant Reste_auFoyer      Retraite
##      0          0          1          0          0
##    RSA      Scolaire  TravailPartT  TravaiPleinT
##      0          0          0          0
##
## $csp
## x
## ETUDES HIGH_STATUT  LOW_STATUT
##      1          0          0
##
## $ABO_TC
## x

```

```

##           NON OUI_GRATUIT   OUI_PAYANT      PAS_INFO
##  0.33596838  0.07114625  0.59288538  0.00000000
##
## $travdom
## x
##           NON      OUI PAS_INFO
##           1          0          0
##
## $zonetrav
## x
## AGGLO_PROCHE CENTRE_VILLE HORS_GRENOBLE      PAS_INFO
## 0.67984190    0.28853755    0.03162055  0.00000000
##
## $dispovp
## x
##           NON      OUI UTILISE OUI UTILISE PAS      PAS_INFO
##  0.6600791     0.2055336     0.1343874     0.0000000
##
## $freqtcu
## x
##           0          1
## 0.1818182  0.8181818
##
## $permis
## x
## YES
## 1
##
## $car_ownership
## x
##           NON      OUI
##  0.6600791  0.3399209

sapply(DB_pers_socio_etu[sapply(DB_pers_socio_etu,is.factor)],function(x){prop.table(table(groups_etu,x))})

## $sexe
## x
## groups_etu      Femme      Homme
##           1 0.5868263 0.4131737
##           2 0.4651163 0.5348837
##
## $etabscol
## x
## groups_etu 0 8 APPRENTISSAGE BAC BAC >= 3 BAC+2 BEP CAP PAS DIPLOME
##           1 1 0          0  0          0  0  0  0          0
##           2 1 0          0  0          0  0  0  0          0
##
## $OCCU1
## x
## groups_etu Appreti_Stage Autre Etudiant Reste_auFoyer Retraite RSA Scolaire
##           1          0  0          1          0  0  0  0          0
##           2          0  0          1          0  0  0  0          0
##
## x
## groups_etu TravailPartT TravaiPleinT

```

```

##          1          0          0
##          2          0          0
##
## $csp
##          x
## groups_etu ETUDES HIGH_STATUT LOW_STATUT
##          1          1          0          0
##          2          1          0          0
##
## $ABO_TC
##          x
## groups_etu      NON OUI_GRATUIT OUI_PAYANT PAS_INFO
##          1 0.2095808  0.1077844  0.6826347 0.0000000
##          2 0.5813953  0.0000000  0.4186047 0.0000000
##
## $travdom
##          x
## groups_etu NON OUI PAS_INFO
##          1  1  0          0
##          2  1  0          0
##
## $zonetrav
##          x
## groups_etu AGGLO_PROCHE CENTRE_VILLE HORS_GRENOBLE PAS_INFO
##          1 0.69461078  0.30538922  0.00000000 0.00000000
##          2 0.65116279  0.25581395  0.09302326 0.00000000
##
## $dispovp
##          x
## groups_etu      NON OUI UTILISE OUI UTILISE PAS PAS INFO
##          1 0.98203593  0.01796407  0.00000000 0.00000000
##          2 0.03488372  0.56976744  0.39534884 0.00000000
##
## $freqtcu
##          x
## groups_etu          0          1
##          1 0.05389222 0.94610778
##          2 0.43023256 0.56976744
##
## $permis
##          x
## groups_etu YES
##          1  1
##          2  1
##
## $car_ownership
##          x
## groups_etu      NON      OUI
##          1 0.98203593 0.01796407
##          2 0.03488372 0.96511628

```

On remarque deux groupes au sein des étudiants (392), il y a celui qui possède une voiture, le deuxième groupe (97% ont une voiture disponible pour aller au travail) et le premier groupe qui ne possède simplement pas de voiture (99%).

Dans le premier groupe qui n'a pas de voiture à disposition plus de 83% utilisent les transports en communs de manière régulière et dans le deuxième groupe, on utilise toujours majoritairement les transports bien que le pourcentage de personne n'utilisant pas régulièrement les transports en communs soit de 46%.

Parmi ceux qui ont à disposition une voiture, c'est 60% qui l'utilise pour aller étudier ou retourner au domicile

Les lieux études ne diffèrent pas et l'abonnement gratuit aux TC est possédé par 20% des étudiants du premier groupe. On remarque que le fait d'avoir une voiture disponible facilite, dans un premier temps, la non prise des transports en communs. Le fait d'avoir une deuxième occupation incite les étudiants à avoir recours à une voiture.

Test 1: Reste de la population

```
#Base donnée catégorie socio prof personnes

DB_pers_socio_1<-New_DB_filtered[,c(1,8,9,10,11,12,13,14,15,16,26,32,33,34,35,36)]
DB_pers_socio_1<-DB_pers_socio_1[!duplicated(DB_pers_socio_1$id_pers),]
DB_pers_socio_1<-dplyr::filter(DB_pers_socio_1,!is.na(DB_pers_socio_1$freqtcu))
DB_pers_socio_1<-dplyr::filter(DB_pers_socio_1,!is.na(DB_pers_socio_1$car_ownership))

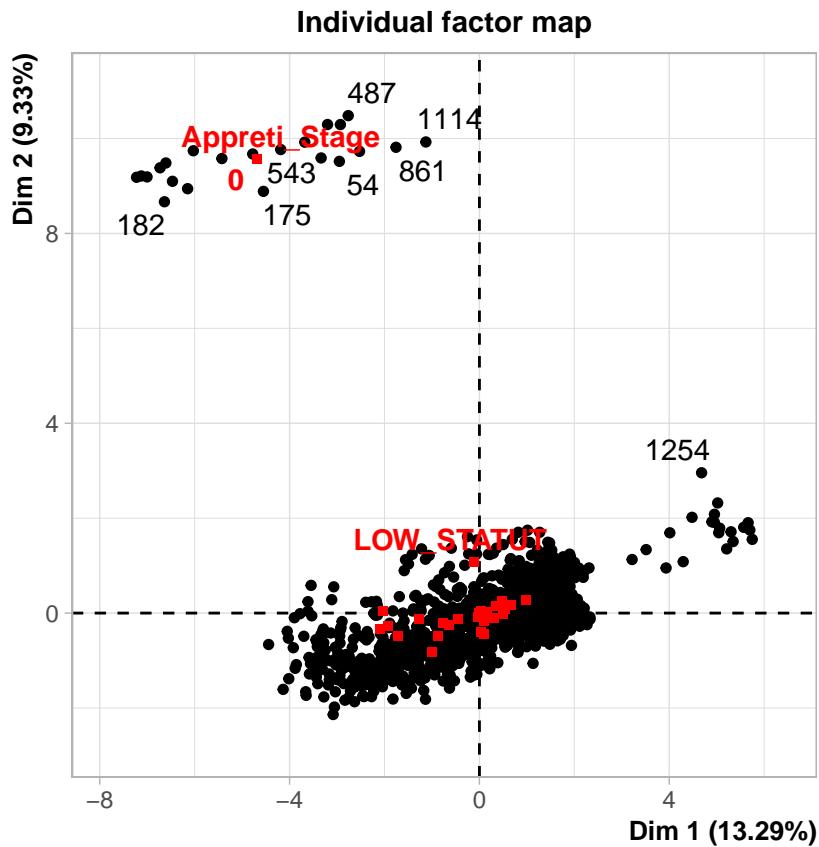
DB_pers_socio_1<-dplyr::filter(DB_pers_socio_1,DB_pers_socio_1$zonetrav!="PAS_INFO")
#vis_miss(DB_pers_socio_1) # un peu freqtcu manquant 1% on peut l'enlever

#DB_cluster_2<-DB_cluster_1[sample(c(TRUE, FALSE), size=nrow(DB_cluster_1), replace=TRUE, prob=c(0.05, 0.95))
afm_1<-FAMD(base=DB_pers_socio_1,ncp=5)

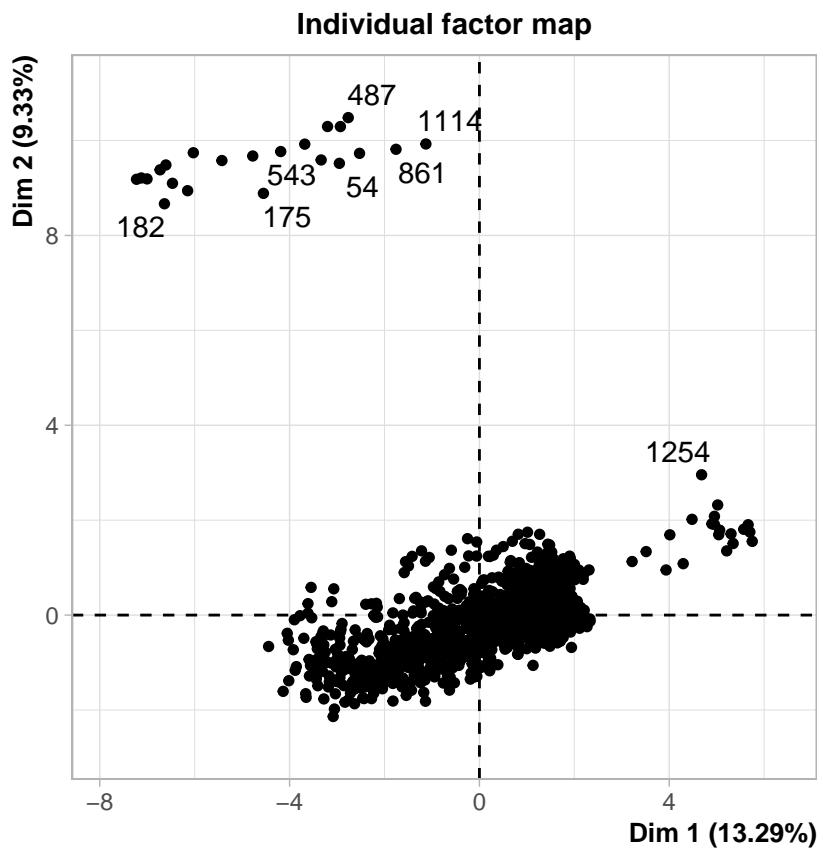
## Warning in sqrt(nombre): Production de NaN

## Warning: ggrepel: 1256 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps

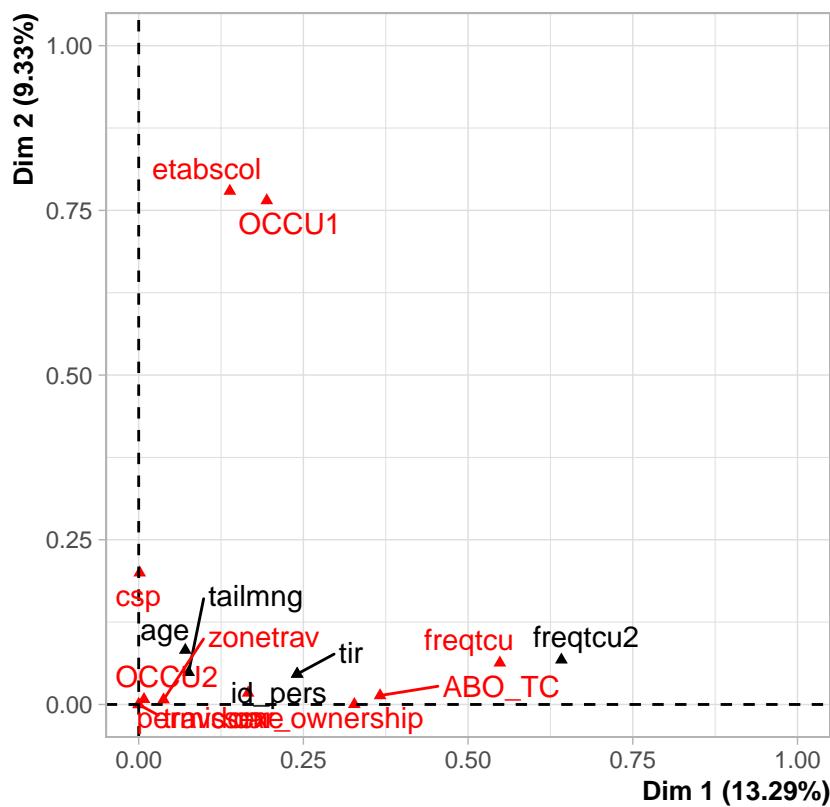
## Warning: ggrepel: 26 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
## Warning: ggrepel: 1256 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

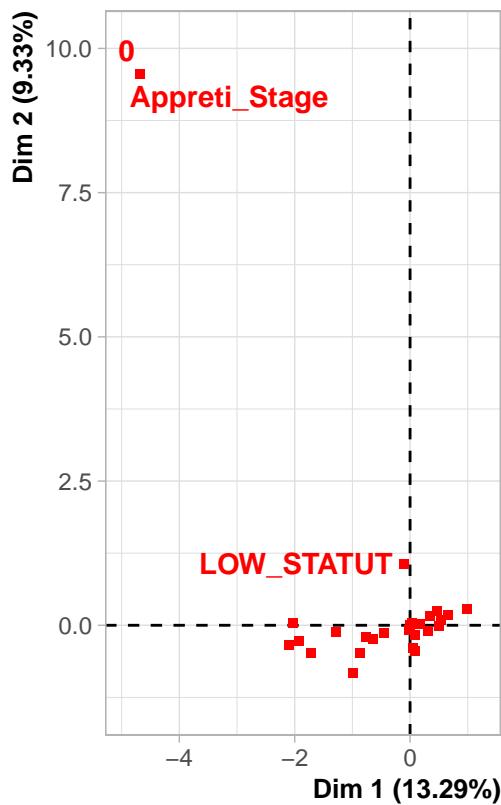


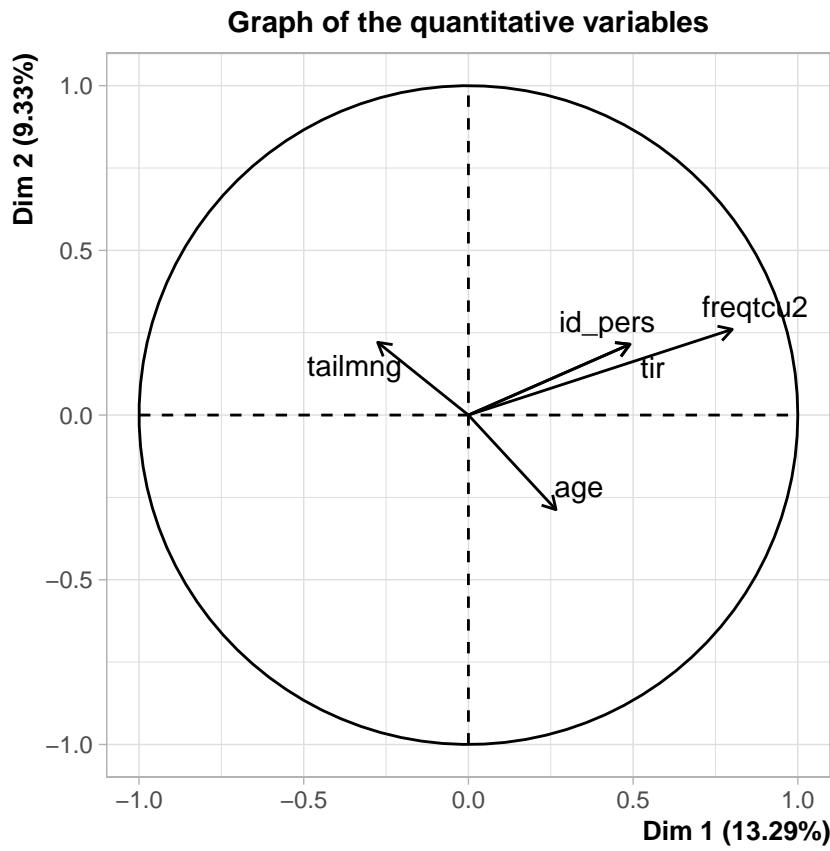
Graph of the variables



```
## Warning: ggrepel: 26 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

Graph of the categories





```
summary(afm_1)
```

```
##
## Call:
## FAMD(base = DB_pers_socio_1, ncp = 5)
##
##
## Eigenvalues
##              Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
## Variance      3.057   2.145   1.903   1.504   1.323
## % of var. 13.293  9.326   8.272   6.539   5.753
## Cumulative % of var. 13.293 22.619 30.891 37.430 43.183
##
## Individuals (the 10 first)
##          Dist    Dim.1     ctr    cos2    Dim.2     ctr    cos2    Dim.3
## 1 | 4.265 | 1.247  0.040  0.085 | -0.054  0.000  0.000 | -2.201
## 2 | 4.024 | -1.519  0.060  0.142 | -0.547  0.011  0.018 | -0.571
## 3 | 4.773 | -2.286  0.135  0.229 | -1.032  0.039  0.047 | -0.364
## 4 | 6.942 | -3.857  0.385  0.309 | -1.086  0.043  0.024 |  0.601
## 5 | 4.385 | -0.872  0.020  0.040 |  0.135  0.001  0.001 | -0.978
## 6 | 5.439 | -2.364  0.145  0.189 | -0.389  0.006  0.005 |  0.193
## 7 | 3.998 | -0.086  0.000  0.000 | -1.150  0.049  0.083 | -1.365
## 8 | 4.410 | -2.717  0.191  0.380 | -1.643  0.100  0.139 |  0.342
## 9 | 5.345 | -1.957  0.099  0.134 | -1.372  0.069  0.066 |  0.187
## 10 | 5.427 | -3.022  0.236  0.310 | -1.091  0.044  0.040 |  0.328
```

```

##          ctr   cos2
## 1      0.201  0.266 |
## 2      0.014  0.020 |
## 3      0.006  0.006 |
## 4      0.015  0.008 |
## 5      0.040  0.050 |
## 6      0.002  0.001 |
## 7      0.077  0.117 |
## 8      0.005  0.006 |
## 9      0.001  0.001 |
## 10     0.004  0.004 |

##
## Continuous variables
##          Dim.1    ctr   cos2   Dim.2    ctr   cos2   Dim.3    ctr
## tir        | 0.491  7.871  0.241 | 0.215  2.154  0.046 | 0.804 33.943
## age       | 0.266  2.313  0.071 | -0.287 3.845  0.082 | 0.188  1.855
## id_pers    | 0.490  7.864  0.240 | 0.215  2.156  0.046 | 0.804 33.946
## freqtcu2  | 0.801 20.986  0.642 | 0.260  3.159  0.068 | -0.226 2.693
## tailmng   | -0.275 2.481  0.076 | 0.221  2.273  0.049 | 0.235  2.896
##
##          cos2
## tir        0.646 |
## age       0.035 |
## id_pers   0.646 |
## freqtcu2 0.051 |
## tailmng   0.055 |

##
## Categories (the 10 first)
##          Dim.1    ctr   cos2 v.test   Dim.2    ctr   cos2
## Femme     | -0.764  2.905  0.346 -14.483 | -0.208  0.436  0.026
## Homme     | 0.665   2.527  0.346  14.483 | 0.181   0.379  0.026
## 0         | -4.690  4.096  0.182 -12.687 | 9.563   34.596  0.755
## APPRENTISSAGE | -1.281  0.083  0.008 -1.798 | -0.114  0.001  0.000
## BAC        | 0.093   0.011  0.001  0.710 | -0.167  0.075  0.004
## BAC >= 3  | 0.086   0.028  0.003  1.299 | -0.443  1.517  0.091
## BAC+2     | -0.020  0.001  0.000 -0.187 | -0.079  0.023  0.001
## BEP        | 0.025   0.001  0.000  0.205 | 0.034   0.003  0.000
## CAP        | 0.346   0.216  0.022  3.169 | 0.155   0.088  0.004
## PAS DIPLOME | -0.646  0.088  0.008 -1.864 | -0.246  0.026  0.001
##
##          v.test   Dim.3    ctr   cos2 v.test
## Femme     -4.698 | 0.627   5.047  0.233 15.059 |
## Homme     4.698 | -0.545  4.390  0.233 -15.059 |
## 0         30.884 | -0.132  0.008  0.000 -0.452 |
## APPRENTISSAGE -0.192 | 1.441   0.272  0.010  2.565 |
## BAC        -1.527 | 0.359   0.441  0.018  3.479 |
## BAC >= 3  -7.985 | 0.016   0.002  0.000  0.301 |
## BAC+2     -0.876 | -0.049  0.011  0.000 -0.570 |
## BEP        0.330 | -0.171  0.112  0.005 -1.766 |
## CAP        1.694 | -0.244  0.278  0.011 -2.833 |
## PAS DIPLOME -0.849 | 0.936   0.478  0.017  3.424 |

```

#Affichage des coordonnées 5 variables latentes deux à deux orthogonales
`head(afm_1indcoord,10)`

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
##					

```

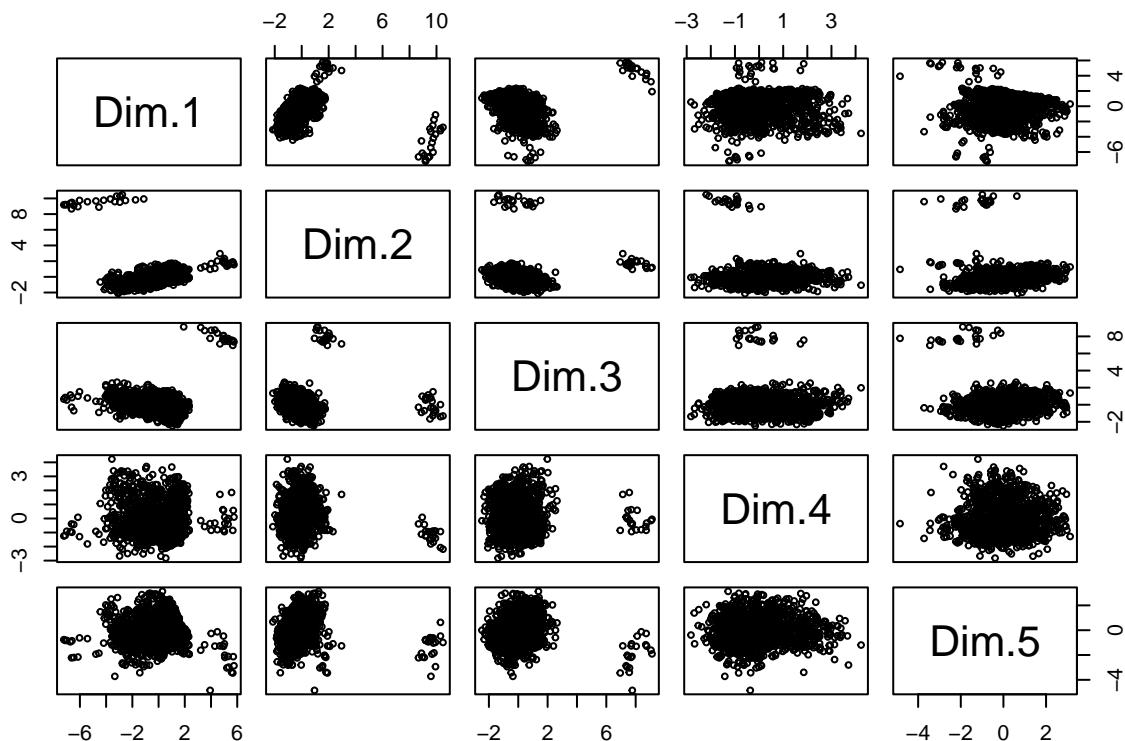
## 1   1.24682056 -0.05354719 -2.2010530 -0.18572532 -0.4191742
## 2  -1.51882938 -0.54702276 -0.5714611 -0.74142083  1.1665302
## 3  -2.28582424 -1.03215777 -0.3641212  0.07554820  0.4901664
## 4  -3.85716908 -1.08587593  0.6013141  0.03230997  1.4417965
## 5  -0.87230482  0.13457854 -0.9779062 -1.41993061  1.2672749
## 6  -2.36436231 -0.38892586  0.1927531 -2.49864866  1.6591910
## 7  -0.08606466 -1.15018401 -1.3648967 -0.82871910 -2.4990619
## 8  -2.71705401 -1.64254070  0.3416640 -0.86106357 -0.8739262
## 9  -1.95747362 -1.37193713  0.1871396 -0.33531089 -0.3741600
## 10 -3.02200097 -1.09136702  0.3279014 -0.19928567  1.0641681

```

```

#Mettre les individus deux à deux si on arrive à distinguer des patterns des formes
pairs(afm_1$ind$coord, cex=0.6)

```



```
#Trouver les groupes pas évident

```

```

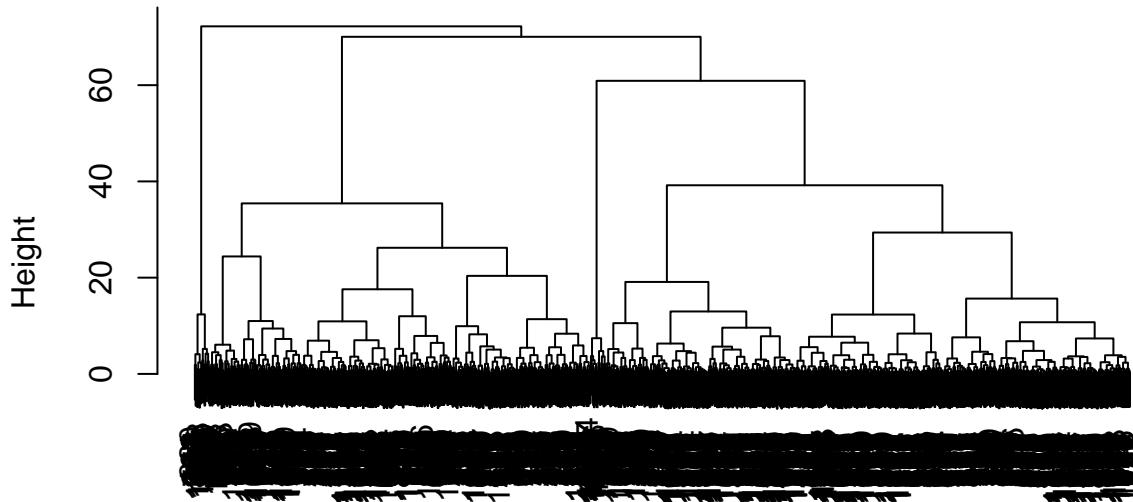
#Classification ascendante hiérarchique
md_1<-dist(afm_1$ind$coord)

ch_1<-hclust(md_1,method='ward.D2')

plot(ch_1)

```

Cluster Dendrogram



```
md_1  
hclust (*, "ward.D2")
```

```
#Le bazar mais c'est normal au vu de tous les individus , le dendrograme  
  
#2carts de hauteur entre les paliers , ça donne une idée de la distance entre les clusters , ici deux g  
  
groups_1<-cutree(ch_1,k=3)  
  
table(groups_1)  
  
## groups_1  
##   1   2   3  
## 729 513  22  
  
pairs(afm_1$ind$coord,col=c('cornflowerblue','yellowgreen','red')[groups_1],cex=0.75)  
  
## Warning in plot.window(...): "cex" n'est pas un paramètre graphique  
  
## Warning in plot.xy(xy, type, ...): "cex" n'est pas un paramètre graphique  
  
## Warning in title(...): "cex" n'est pas un paramètre graphique  
  
## Warning in plot.window(...): "cex" n'est pas un paramètre graphique  
  
## Warning in plot.xy(xy, type, ...): "cex" n'est pas un paramètre graphique
```

```
## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique
```

```

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

```

```

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

```

```

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

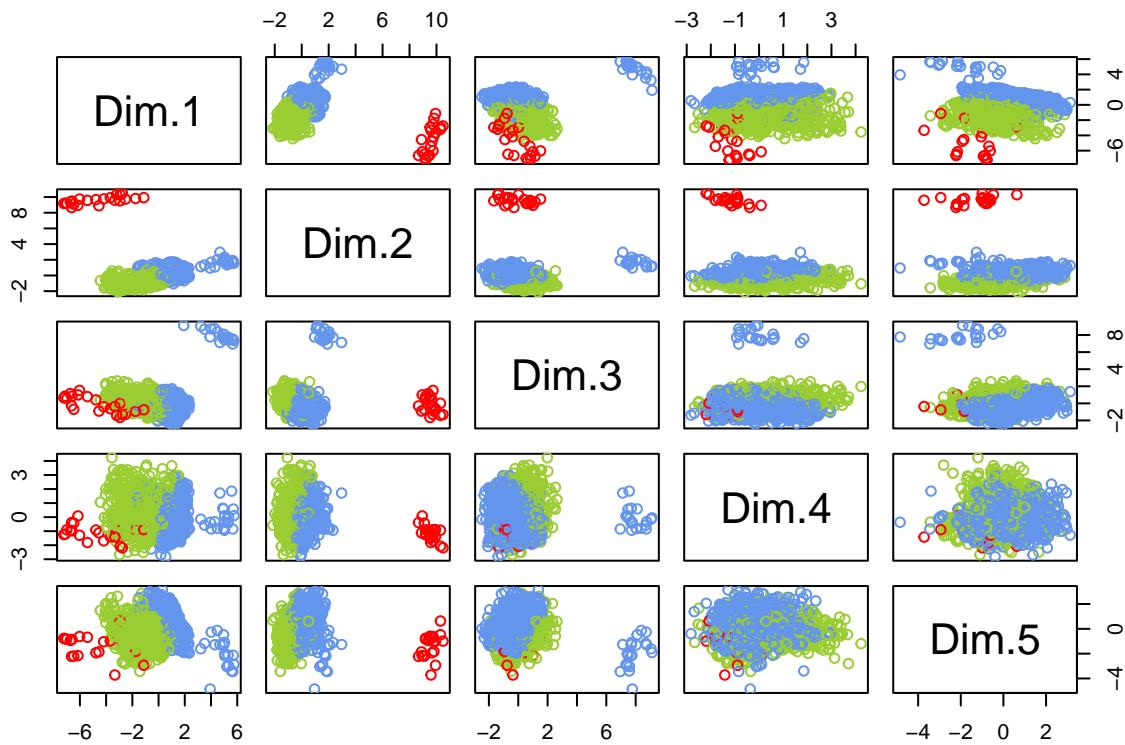
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique
```



```

sapply(DB_pers_socio_1[sapply(DB_pers_socio_1,is.numeric)],mean)

##          tir           age      id_pers     freqtcu2      tailmng
## 1.247002e+02 4.130142e+01 1.247743e+08 3.609968e+00 1.472310e+00

lapply(DB_pers_socio_1[sapply(DB_pers_socio_1,is.numeric)],function(x){tapply(x,groups_1,mean)})

## $tir
##      1       2       3
## 129.6900 117.7856 120.5909
##
## $age
##      1       2       3
## 41.50892 41.74854 24.00000
##
## $id_pers
##      1       2       3
## 129759392 117865105 120698638
##
## $freqtcu2
##      1       2       3
## 4.530864 2.335283 2.818182
##
## $tailmng

```

```

##      1      2      3
## 1.443073 1.485380 2.136364

sapply(DB_pers_socio_1[sapply(DB_pers_socio_1,is.factor)],function(x){prop.table(table(x))}

## $sexe
## x
##   Femme      Homme
## 0.4651899 0.5348101
##
## $etabscol
## x
##      0          8 APPRENTISSAGE          BAC      BAC >= 3
## 0.017405063 0.000000000 0.004746835 0.124208861 0.355221519
##      BAC+2        BEP         CAP PAS DIPLOME
## 0.171677215 0.138449367 0.168512658 0.019778481
##
## $OCCU1
## x
## Appreti_Stage       Autre      Etudiant Reste_auFoyer      Retraite
## 0.01740506 0.00000000 0.00000000 0.00000000 0.00000000
##      RSA      Scolaire TravailPartT TravaiPleinT
## 0.00000000 0.00000000 0.18433544 0.79825949
##
## $OCCU2
## x
##   Etudes    NON_Con     Travail
## 0.02452532 0.97547468 0.00000000
##
## $csp
## x
##      ETUDES HIGH_STATUT  LOW_STATUT
## 0.0000000 0.7262658 0.2737342
##
## $ABO_TC
## x
##      NON OUI_GRATUIT  OUI_PAYANT    PAS_INFO
## 0.79272152 0.03876582 0.16851266 0.00000000
##
## $travdom
## x
##      NON      OUI PAS_INFO
## 1      0      0
##
## $zonetrav
## x
## AGGLO_PROCHE CENTRE_VILLE HORS_GRENOBLE      PAS_INFO
## 0.5205696 0.3378165 0.1416139 0.0000000
##
## $freqtcu
## x
##      0      1
## 0.6352848 0.3647152
##

```

```

## $permis
## x
## YES
## 1
##
## $car_ownership
## x
##      NON      OUI
## 0.1954114 0.8045886

sapply(DB_pers_socio_1[sapply(DB_pers_socio_1,is.factor)],function(x){prop.table(table(groups_1,x),1)})

## $sexe
## x
## groups_1   Femme     Homme
## 1 0.3497942 0.6502058
## 2 0.6218324 0.3781676
## 3 0.6363636 0.3636364
##
## $etabscol
## x
## groups_1      0          8 APPRENTISSAGE          BAC      BAC >= 3
## 1 0.000000000 0.000000000 0.000000000 0.133058985 0.336076818
## 2 0.000000000 0.000000000 0.011695906 0.116959064 0.397660819
## 3 1.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## x
## groups_1      BAC+2      BEP      CAP PAS DIPLOME
## 1 0.186556927 0.146776406 0.189300412 0.008230453
## 2 0.157894737 0.132553606 0.146198830 0.037037037
## 3 0.000000000 0.000000000 0.000000000 0.000000000
##
## $OCCU1
## x
## groups_1 Appreti_Stage      Autre  Etudiant Reste_auFoyer Retraite      RSA
## 1 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## 2 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## 3 1.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x
## groups_1 Scolaire TravailPartT TravaiPleinT
## 1 0.0000000 0.1495199 0.8504801
## 2 0.0000000 0.2417154 0.7582846
## 3 0.0000000 0.0000000 0.0000000
##
## $OCCU2
## x
## groups_1      Etudes      NON_Con      Travail
## 1 0.01783265 0.98216735 0.00000000
## 2 0.03508772 0.96491228 0.00000000
## 3 0.00000000 1.00000000 0.00000000
##
## $csp
## x
## groups_1      ETUDES HIGH_STATUT LOW_STATUT
## 1 0.0000000 0.7146776 0.2853224

```

```

##      2 0.0000000  0.7738791  0.2261209
##      3 0.0000000  0.0000000  1.0000000
##
## $ABO_TC
##      x
## groups_1      NON OUI_GRATUIT  OUI_PAYANT    PAS_INFO
##      1 0.971193416 0.001371742 0.027434842 0.000000000
##      2 0.549707602 0.089668616 0.360623782 0.000000000
##      3 0.545454545 0.090909091 0.363636364 0.000000000
##
## $travdom
##      x
## groups_1 NON OUI PAS_INFO
##      1   1   0       0
##      2   1   0       0
##      3   1   0       0
##
## $zonetrav
##      x
## groups_1 AGGLO_PROCHE CENTRE_VILLE HORS_GRENOBLE  PAS_INFO
##      1 0.5445816 0.2976680 0.1577503 0.0000000
##      2 0.4931774 0.3898635 0.1169591 0.0000000
##      3 0.3636364 0.4545455 0.1818182 0.0000000
##
## $freqtcu
##      x
## groups_1      0       1
##      1 0.990397805 0.009602195
##      2 0.142300195 0.857699805
##      3 0.363636364 0.636363636
##
## $permis
##      x
## groups_1 YES
##      1   1
##      2   1
##      3   1
##
## $car_ownership
##      x
## groups_1      NON      OUI
##      1 0.06721536 0.93278464
##      2 0.36257310 0.63742690
##      3 0.545454545 0.45454545

```

La base de données compte 1246 observations pour 15 variables, on a filtré les individus de cette base pour que les caractéristiques d'une personne apparaissent une seule fois, les variables sont toutes issues de la fiche Personne

On distingue 3 groupes respectivement de 914 , 305 et 27 personnes avec le groupe 2 qui comporte pour notre variable Y freqtcu une répartition de 24% Y=0 et 76% Y=1.

On remarque que pour ce groupe, 76% des gens possèdent une voiture au sens propre , 67% n'ont pas de voiture disponible pour se rendre au travail 47% se rendent travailler dans le centre ville de Grenoble contre 43% se rendant dans l'agglomération proche. 66% sont des travailleurs à plein temps et 33 % à temps partiel. 73% ont un statut social dit "élév6" avec 30% de BAC+>=3

60% ont un abonnement TC payant tandis que pour le groupe 1 avec Y=0 77% et Y=1 23% , 96% des personnes n'ont pas d'abonnement aux transports en communs.

Ce que l'on peut exprimer c'est une tendance à prendre les transports en communs fréquemment quand on ne dispose pas de voiture pour se rendre au travail. Avoir un abonnement gratuit aux TC n'a à priori pas d'impact sur la prise des transports en communs. Le statut social contribue aussi bien à la prise fréquente de TC qu'à celle de la voiture

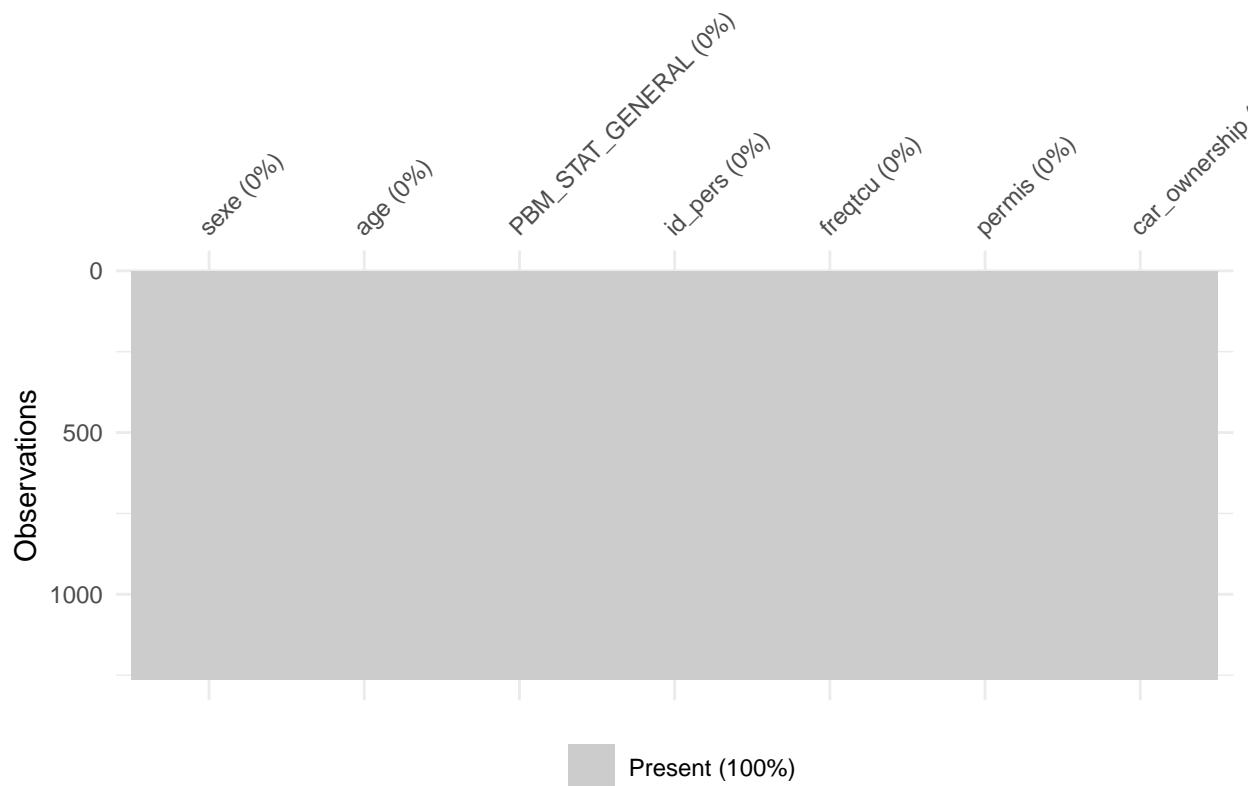
Le fait de travailler en ville contribue à la prise fréquente des TC.

Test 2 : Cluster avec Car_ownership : ça n'a pas été concluant Ici on va restreindre notre base de données avec les déplacements des personnes et l'influence des variables dites de confort reliées à la voiture , car_ownership n'est plus simplement le fait de posséder une voiture c'est plus global que cela

```
#c(6,7,8,9,17,18,19,20,21,23,24,28,31)
DB_car_ownership<-New_DB_filtered[,c(8,9,18,26,33,34,36)]
DB_car_ownership<-DB_car_ownership[!duplicated(DB_car_ownership$id_pers),]
#DB_car_ownership<-dplyr::filter(DB_car_ownership,!is.na(DB_car_ownership$id_depl))
#DB_car_ownership<-dplyr::filter(DB_car_ownership,!is.na(DB_car_ownership$duree))
DB_car_ownership<-dplyr::filter(DB_car_ownership,!is.na(DB_car_ownership$car_ownership))
DB_car_ownership<-dplyr::filter(DB_car_ownership,!is.na(DB_car_ownership$freqtcu))

#DB_car_ownership<-dplyr::filter(DB_car_ownership,DB_car_ownership$PBM_STAT_GENERAL!="PAS_INFO")
#DB_car_ownership<-dplyr::filter(DB_car_ownership,DB_car_ownership$STAT_TRAV!="PAS_INFO")

vis_miss(DB_car_ownership)
```



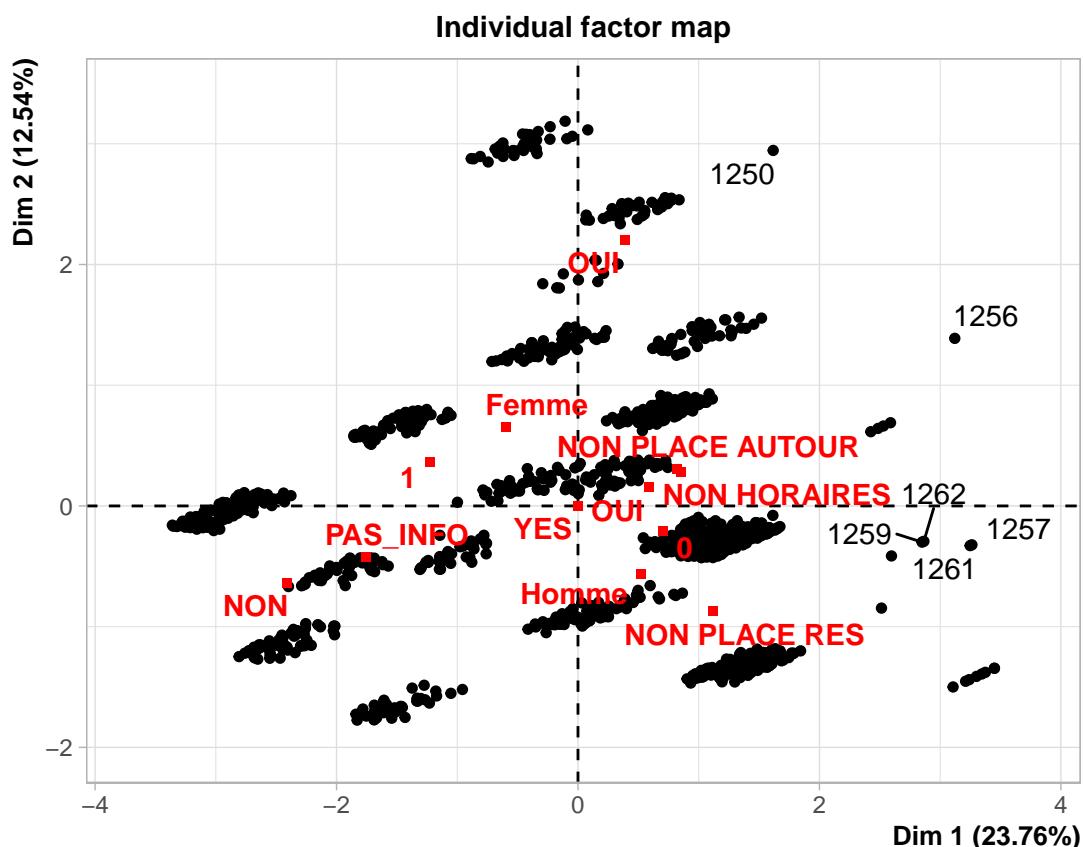
```
table(DB_car_ownership$car_ownership)
```

```
##  
## NON OUI  
## 247 1017
```

```
afm_3<-FAMD(DB_car_ownership, ncp=5)
```

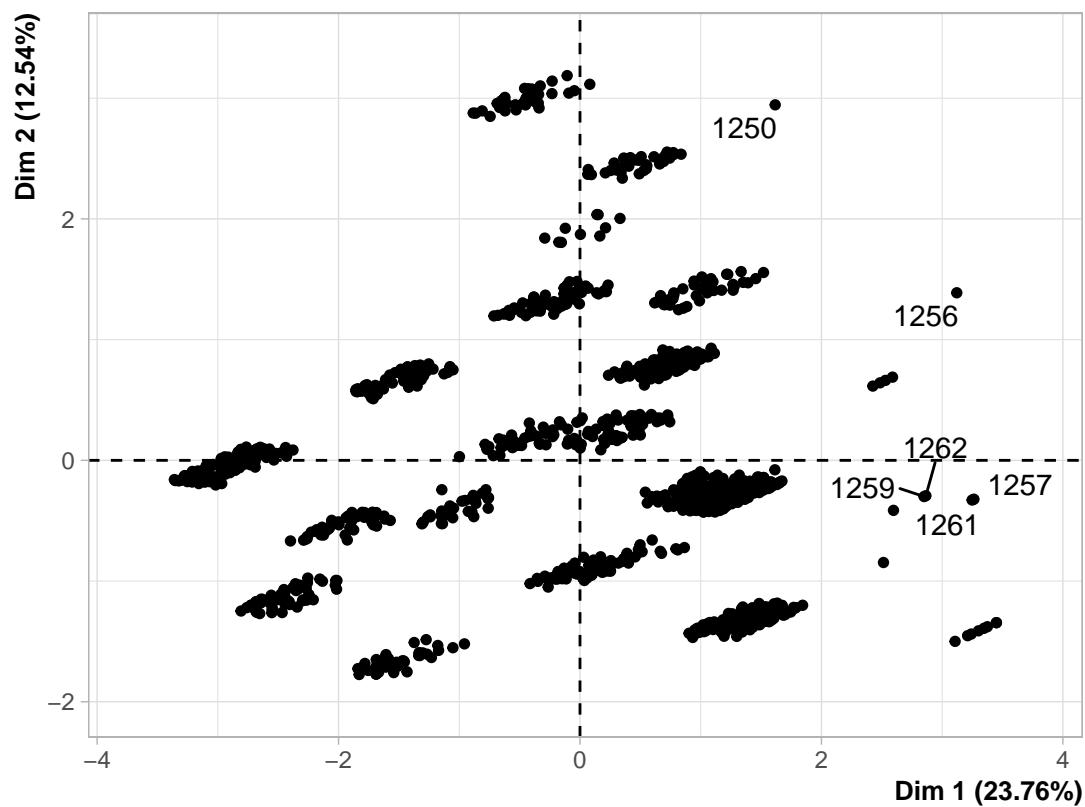
```
## Warning in sqrt(nombre): Production de NaN
```

```
## Warning: ggrepel: 1258 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

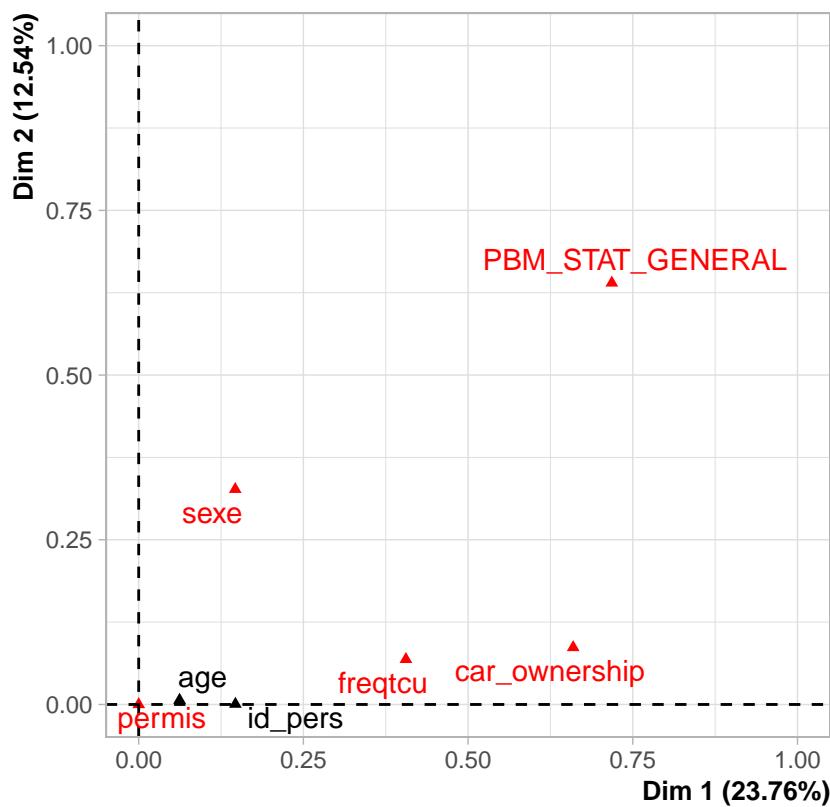


```
## Warning: ggrepel: 1258 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

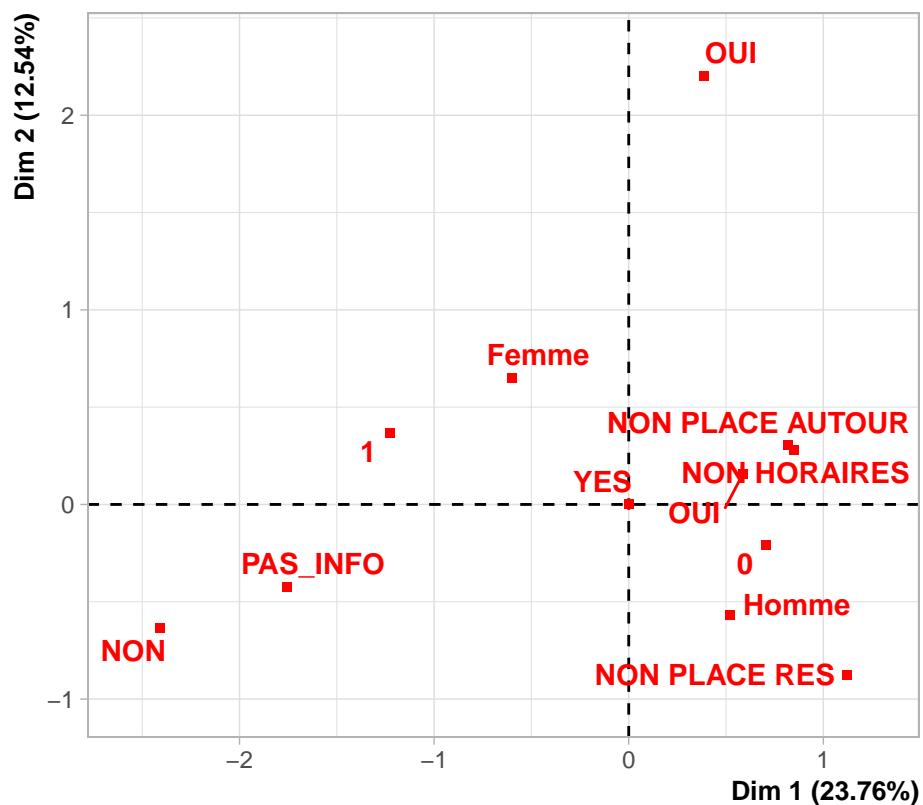
Individual factor map



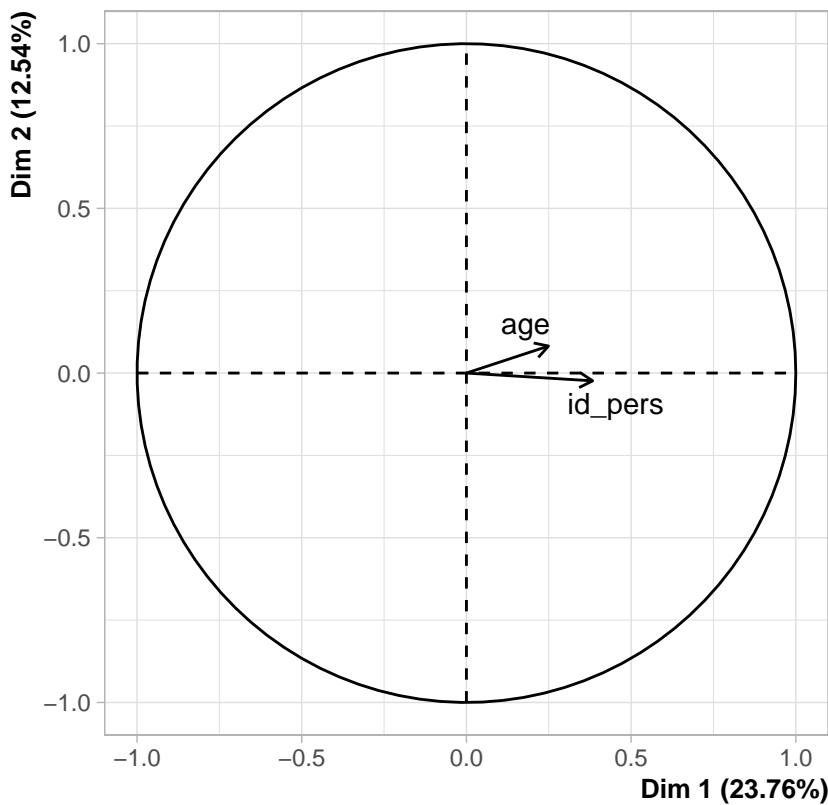
Graph of the variables



Graph of the categories



Graph of the quantitative variables



```
summary(afm_3)
```

```
##
## Call:
## FAMD(base = DB_car_ownership, ncp = 5)
##
##
## Eigenvalues
##              Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
## Variance      2.139   1.129   1.080   1.009   0.999
## % of var. 23.764 12.544 11.996 11.215 11.100
## Cumulative % of var. 23.764 36.307 48.304 59.519 70.619
##
## Individuals (the 10 first)
##          Dist   Dim.1     ctr    cos2   Dim.2     ctr    cos2
## 1 | 2.089 | 1.002 | 0.037 | 0.230 | -0.258 | 0.005 | 0.015 |
## 2 | 2.716 | -1.840 | 0.125 | 0.459 | 0.593 | 0.025 | 0.048 |
## 3 | 3.185 | -2.642 | 0.258 | 0.688 | -1.150 | 0.093 | 0.130 |
## 4 | 3.298 | -3.211 | 0.381 | 0.948 | -0.092 | 0.001 | 0.001 |
## 5 | 3.070 | -1.782 | 0.117 | 0.337 | -1.682 | 0.198 | 0.300 |
## 6 | 2.754 | -1.855 | 0.127 | 0.454 | 0.586 | 0.024 | 0.045 |
## 7 | 2.761 | -0.884 | 0.029 | 0.103 | -0.290 | 0.006 | 0.011 |
## 8 | 2.627 | -1.483 | 0.081 | 0.319 | 0.755 | 0.040 | 0.083 |
## 9 | 2.431 | -0.383 | 0.005 | 0.025 | 1.352 | 0.128 | 0.309 |
## 10 | 3.655 | -3.360 | 0.418 | 0.845 | -0.159 | 0.002 | 0.002 |
```

```

##          Dim.3    ctr   cos2
## 1      -1.210  0.107  0.336 |
## 2      -0.717  0.038  0.070 |
## 3      -0.649  0.031  0.042 |
## 4      -0.226  0.004  0.005 |
## 5      -1.085  0.086  0.125 |
## 6      -0.776  0.044  0.079 |
## 7       0.379  0.011  0.019 |
## 8       0.685  0.034  0.068 |
## 9      -0.176  0.002  0.005 |
## 10     -0.810  0.048  0.049 |

##
## Continuous variables
##          Dim.1    ctr   cos2   Dim.2    ctr   cos2   Dim.3    ctr
## age        |  0.249  2.901  0.062 |  0.082  0.595  0.007 |  0.695 44.761
## id_pers    |  0.383  6.864  0.147 | -0.024  0.049  0.001 |  0.608 34.277
##          cos2
## age        0.483 |
## id_pers    0.370 |
##
## Categories (the 10 first)
##          Dim.1    ctr   cos2 v.test   Dim.2    ctr   cos2
## Femme      | -0.600  3.664  0.278 -13.604 |  0.651 15.476  0.328
## Homme      |  0.522  3.187  0.278  13.604 | -0.566 13.461  0.328
## NON HORAIRES |  0.851  0.213  0.010  2.414 |  0.282  0.084  0.001
## NON PLACE AUTOUR |  0.819  5.164  0.311 14.672 |  0.303  2.543  0.043
## NON PLACE RES |  1.122  5.814  0.294 14.110 | -0.874 12.657  0.179
## OUI         |  0.389  0.319  0.016  3.087 |  2.204 36.775  0.503
## PAS_INFO    | -1.758 22.071  0.886 -29.757 | -0.425  4.621  0.052
## 0           |  0.706  6.917  0.657 22.634 | -0.211  2.214  0.059
## 1           | -1.229 12.048  0.657 -22.634 |  0.367  3.857  0.059
## YES         |  0.000  0.000  0.000    NaN |  0.000  0.000  0.000
##          v.test   Dim.3    ctr   cos2 v.test
## Femme      20.312 |  0.346  4.767  0.092 11.025 |
## Homme      -20.312 | -0.301  4.147  0.092 -11.025 |
## NON HORAIRES |  1.102  0.117  0.016  0.000  0.469 |
## NON PLACE AUTOUR |  7.481 -0.292  2.579  0.040 -7.367 |
## NON PLACE RES | -15.126  0.391  2.769  0.036  6.919 |
## OUI         |  24.091 -0.471  1.834  0.023 -5.262 |
## PAS_INFO    | -9.892  0.196  1.081  0.011  4.678 |
## 0           | -9.304 -0.102  0.571  0.014 -4.622 |
## 1           |  9.304  0.178  0.995  0.014  4.622 |
## YES         |    NaN  0.000  0.000  0.000    NaN |

```

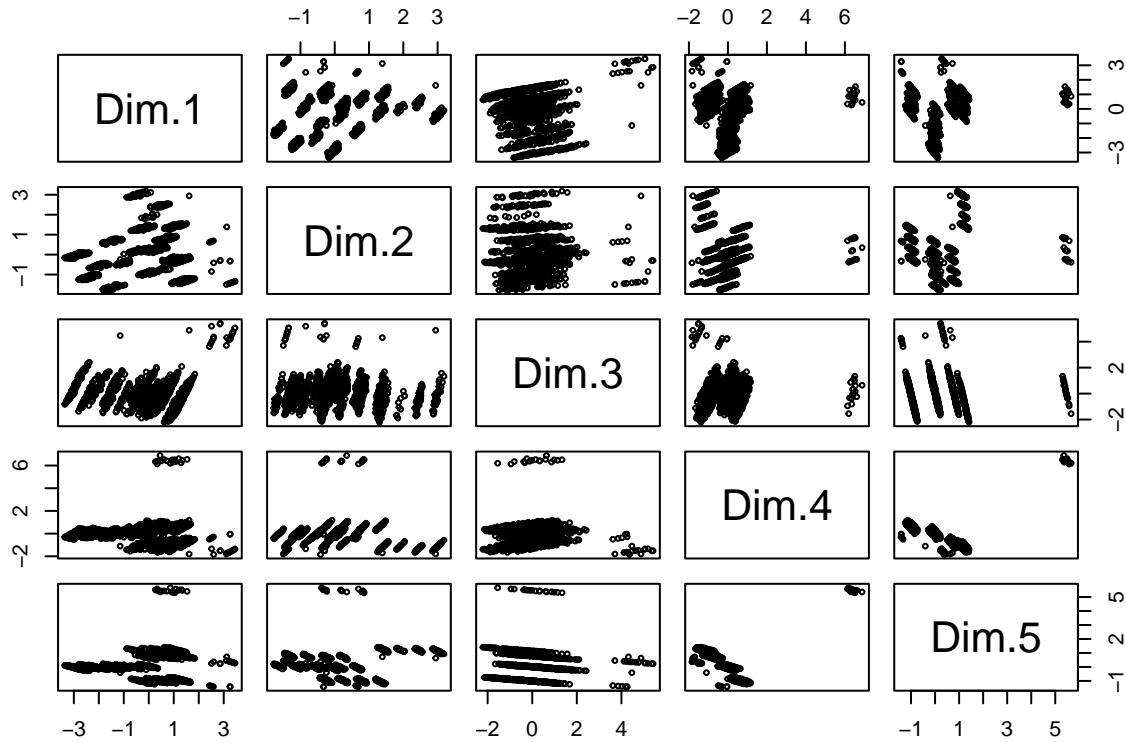
```
head(afm_3$ind$coord)
```

```

##          Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## 1  1.002134 -0.25816092 -1.2103568  0.71366595 -0.86862920
## 2 -1.840461  0.59307427 -0.7171669 -0.10280720  0.07161904
## 3 -2.641795 -1.15009798 -0.6493386  0.01612881  0.09395245
## 4 -3.211206 -0.09182596 -0.2260225 -0.10232344  0.03631924
## 5 -1.781693 -1.68216300 -1.0846177 -0.10203012  0.16482557
## 6 -1.855323  0.58634015 -0.7755663 -0.12167927  0.07927992

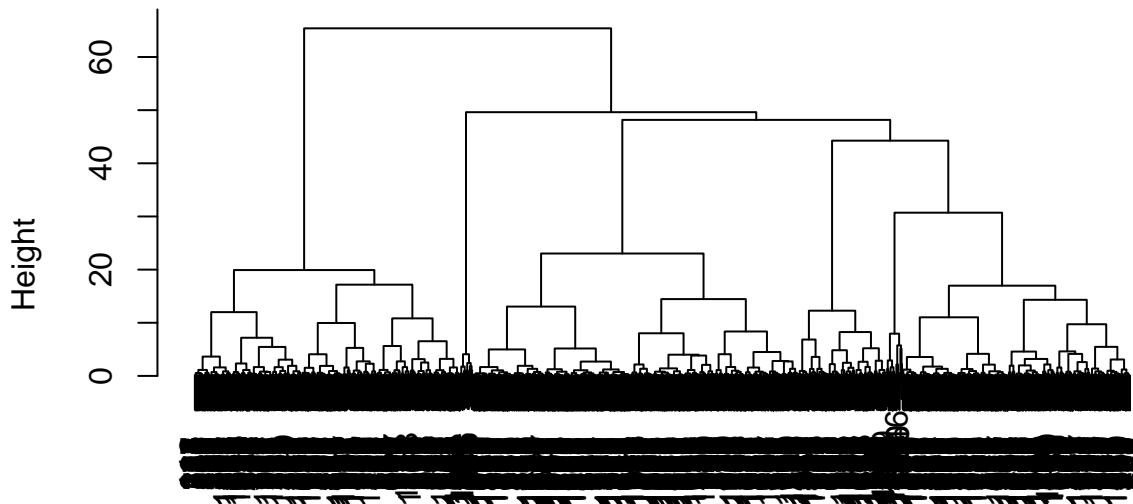
```

```
pairs(afm_3$ind$coord, cex=0.5)
```



```
md_3<-dist(afm_3$ind$coord)
ch_3<-hclust(md_3,method="ward.D2")
plot(ch_3)
```

Cluster Dendrogram



```
md_3  
hclust (*, "ward.D2")
```

```
groups_3<-cutree(ch_3,k=3)  
table(groups_3)
```

```
## groups_3  
##   1   2   3  
## 888 359  17
```

```
pairs(afm_3$ind$coord,col=c('cornflowerblue','yellowgreen','red')[groups_3],cew=0.75)
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
```

```
## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
```

```
## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un  
## paramètre graphique
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
```

```
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
```

```
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique
## Warning in plot.window(...): "cew" n'est pas un paramètre graphique
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique
## Warning in title(...): "cew" n'est pas un paramètre graphique
```

```

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "cew" n'est pas un
## paramètre graphique

## Warning in plot.window(...): "cew" n'est pas un paramètre graphique

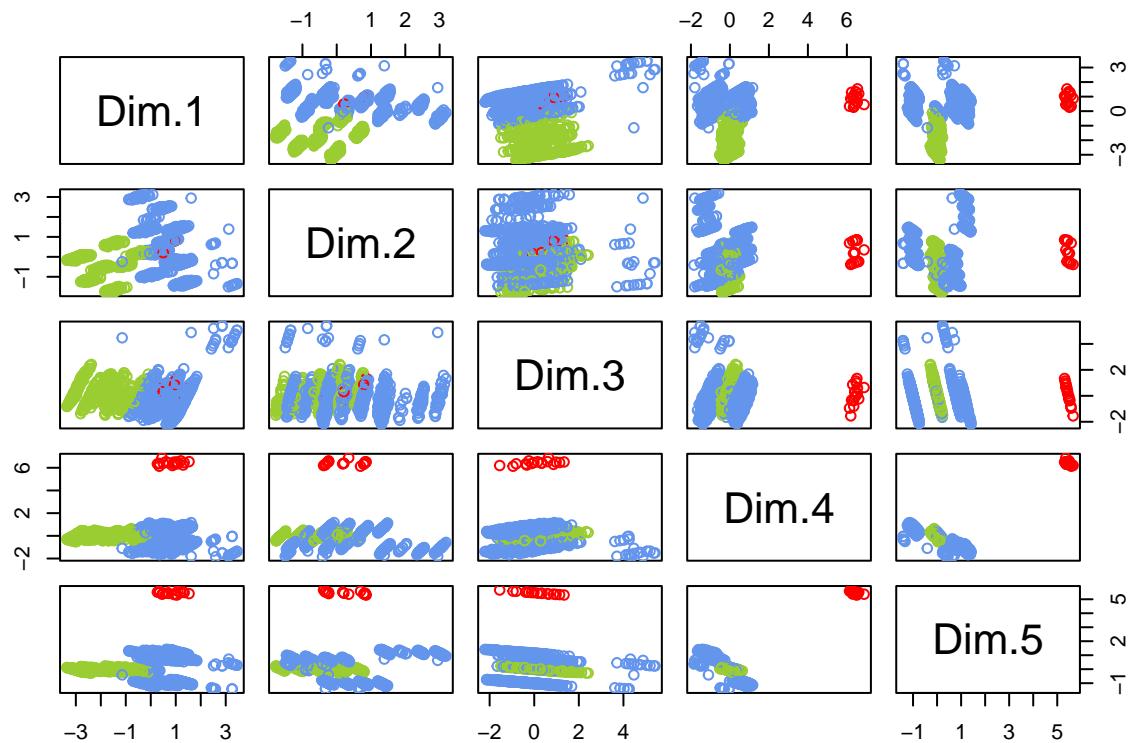
## Warning in plot.xy(xy, type, ...): "cew" n'est pas un paramètre graphique

## Warning in title(...): "cew" n'est pas un paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

## Warning in axis(side = side, at = at, labels = labels, ...): "cew" n'est pas un
## paramètre graphique

```



```

#lapply(DB_car_ownership[sapply(DB_car_ownership, is.numeric)], function(x){tapply(x, groups_3, sd)})
lapply(DB_car_ownership[sapply(DB_car_ownership, is.numeric)], function(x){tapply(x, groups_3, mean)})

## $age
##      1      2      3
## 42.08896 39.22006 44.11765
##
## $id_pers
##      1      2      3
## 128099344 116585829 124013884

#lapply(DB_car_ownership[sapply(DB_car_ownership, is.numeric)], function(x){tapply(x, groups_3, var)})

sapply(DB_car_ownership[sapply(DB_car_ownership, is.factor)], function(x){prop.table(table(groups_3, x), 1)})

## $sexe
##      x
## groups_3   Femme   Homme
##      1 0.3896396 0.6103604
##      2 0.6545961 0.3454039
##      3 0.4117647 0.5882353
##
## $PBM_STAT_GENERAL
##      x
## groups_3 NON HORAIRES NON PLACE AUTOUR NON PLACE RES          OUI    PAS_INFO
##      1 0.00000000 0.50112613 0.30067568 0.13738739 0.06081081
##      2 0.00000000 0.00000000 0.00000000 0.00000000 1.00000000
##      3 1.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##
## $freqtcu
##      x
## groups_3      0      1
##      1 0.7657658 0.2342342
##      2 0.3064067 0.6935933
##      3 0.7647059 0.2352941
##
## $permis
##      x
## groups_3 YES
##      1  1
##      2  1
##      3  1
##
## $car_ownership
##      x
## groups_3      NON      OUI
##      1 0.001126126 0.998873874
##      2 0.685236769 0.314763231
##      3 0.000000000 1.000000000

```

Le confort lié à la voiture se traduit par une place de parking disponible ou non une fois arrivé sur son lieu de travail ou d'études, il semblerait que le fait d'avoir une place de parking sur son lieu de travail encourage

les gens a prendre leur voiture plutôt que les transports en communs. 66% des personnes du groupe 2 (75% des personnes n'ont pas une utilisation régulière des transports en communs) n'auraient pas de problèmes de stationnement au travail. De plus, le simple fait d'avoir une voiture n'a pas un rôle significatif dans l'utilisation ou non des transports en communs. Il est important de noter qu'il y a beaucoup de valeurs manquantes en ce qui concerne les variables PBM STAT et STAT TRAV

Test 3 : Trajet : concluant

```
DB_trip<-New_DB_filtered[,c(20,21,22,23,26,27,28,29,30,31,33,36)]
DB_trip<-dplyr::filter(DB_trip,!is.na(DB_trip$id_traj))
DB_trip<-dplyr::filter(DB_trip,!is.na(DB_trip$ntraj))
DB_trip<-dplyr::filter(DB_trip,!is.na(DB_trip$Couteff))
DB_trip<-dplyr::filter(DB_trip,!is.na(DB_trip$car_ownership))
#DB_trip<-dplyr::filter(DB_trip,!is.na(DB_trip$D13))

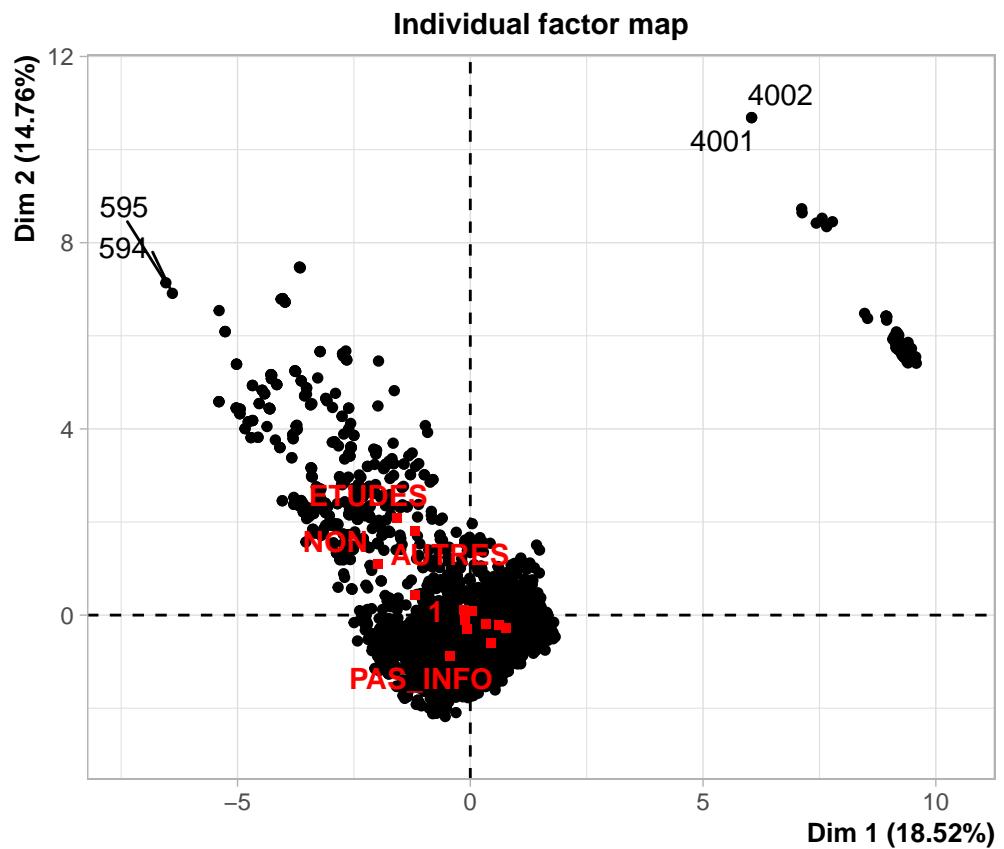
#vis_miss(DB_trip)
```

```
DB_trip_2<-DB_trip[,-c(5,6,9)]
```

```
afm_4<-FAMD(DB_trip, ncp=5)
```

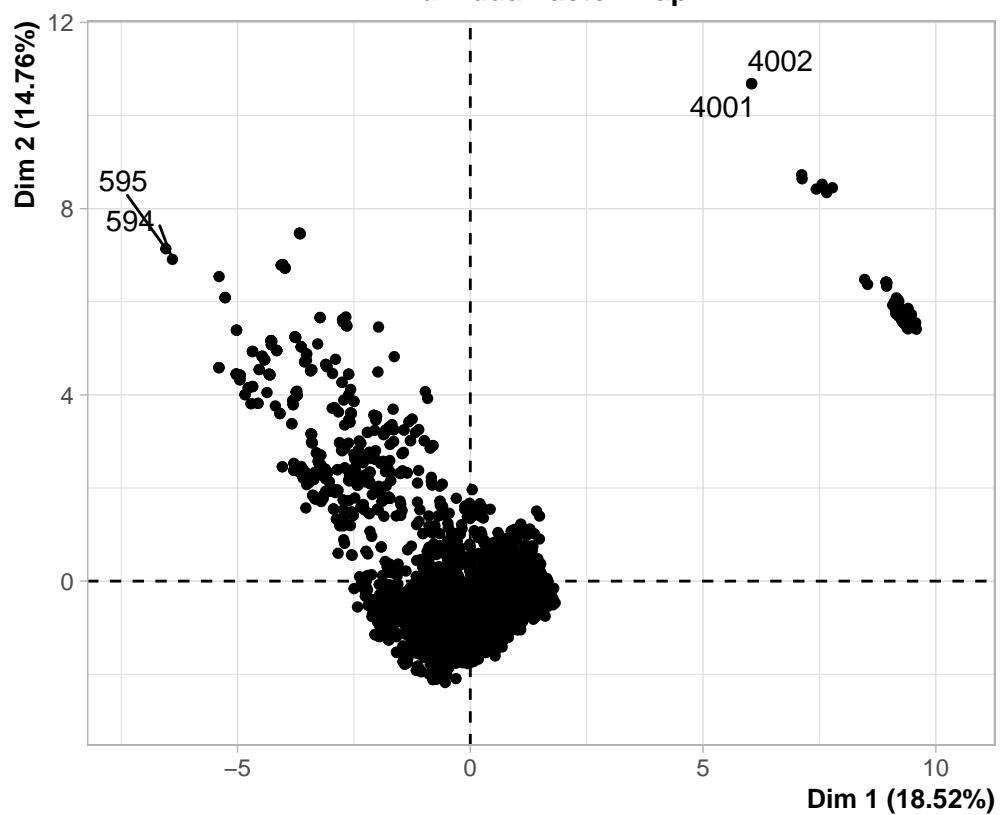
```
## Warning: ggrepel: 4039 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

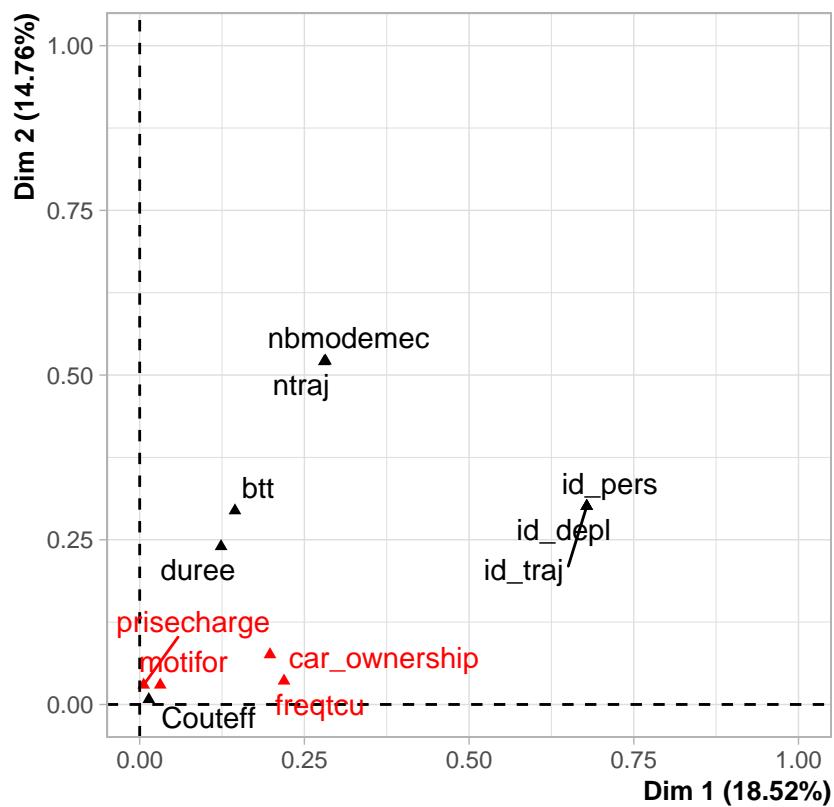


```
## Warning: ggrepel: 4039 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```

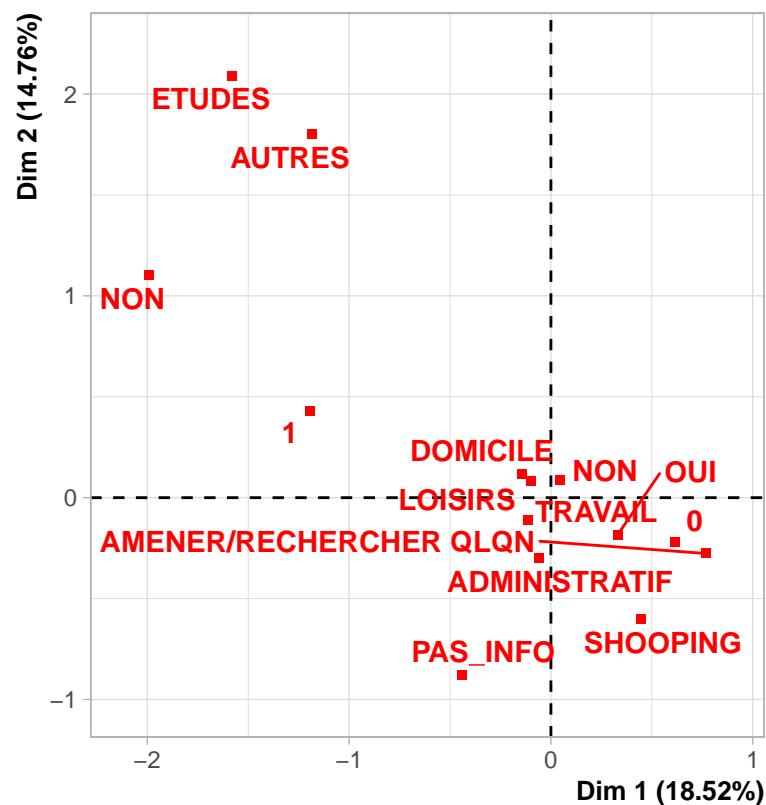
Individual factor map

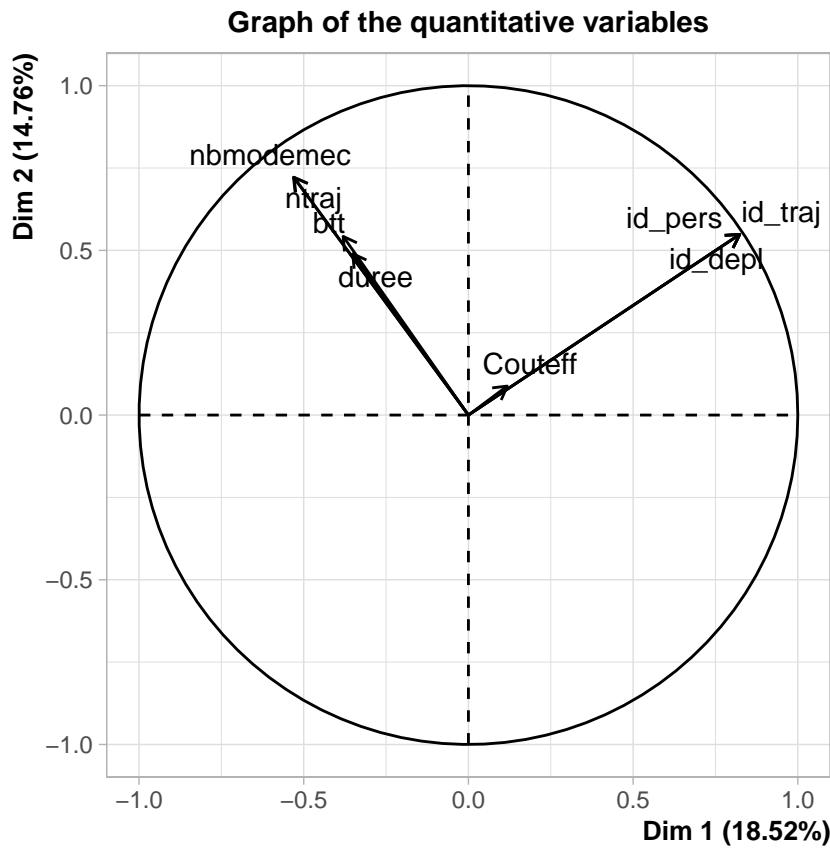


Graph of the variables



Graph of the categories





```
summary(afm_4)
```

```
##
## Call:
## FAMD(base = DB_trip, ncp = 5)
##
##
## Eigenvalues
##              Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
## Variance      3.334   2.658   1.497   1.237   1.035
## % of var. 18.522 14.764   8.316   6.870   5.753
## Cumulative % of var. 18.522 33.286 41.603 48.472 54.225
##
## Individuals (the 10 first)
##          Dist    Dim.1     ctr    cos2    Dim.2     ctr    cos2
## 1 | 3.518 | -0.480  0.002  0.019 | -1.400  0.018  0.158 |
## 2 | 3.625 | -0.537  0.002  0.022 | -1.287  0.015  0.126 |
## 3 | 2.749 | -1.102  0.009  0.161 | -1.416  0.019  0.265 |
## 4 | 6.228 | -1.077  0.009  0.030 | -1.573  0.023  0.064 |
## 5 | 6.735 | -2.254  0.038  0.112 | -0.323  0.001  0.002 |
## 6 | 3.789 | -2.343  0.041  0.382 | -0.058  0.000  0.000 |
## 7 | 4.521 | -1.855  0.026  0.168 | -0.704  0.005  0.024 |
## 8 | 4.853 | -2.197  0.036  0.205 | -0.521  0.003  0.012 |
## 9 | 4.851 | -2.081  0.032  0.184 | -0.702  0.005  0.021 |
## 10 | 4.174 | -1.395  0.014  0.112 | -1.692  0.027  0.164 |
```

```

## Dim.3    ctr   cos2
## 1       1.674  0.046  0.226 |
## 2       2.072  0.071  0.327 |
## 3      -0.299  0.001  0.012 |
## 4      -0.162  0.000  0.001 |
## 5      -0.326  0.002  0.002 |
## 6      -0.114  0.000  0.001 |
## 7      -1.239  0.025  0.075 |
## 8      -0.789  0.010  0.026 |
## 9      -1.052  0.018  0.047 |
## 10     -1.570  0.041  0.141 |

##
## Continuous variables
##           Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3
## duree      | -0.351  3.705  0.124 |  0.490  9.027  0.240 |  0.534
## nbmodemec  | -0.531  8.458  0.282 |  0.722 19.626  0.522 | -0.188
## id_pers    |  0.824 20.354  0.679 |  0.549 11.322  0.301 | -0.107
## id_depl    |  0.824 20.354  0.679 |  0.549 11.322  0.301 | -0.107
## id_traj    |  0.824 20.354  0.679 |  0.549 11.322  0.301 | -0.107
## ntraj      | -0.529  8.405  0.280 |  0.722 19.591  0.521 | -0.188
## btt        | -0.380  4.337  0.145 |  0.542 11.072  0.294 |  0.377
## Couteff    |  0.118  0.415  0.014 |  0.088  0.291  0.008 |  0.603
##           ctr   cos2
## duree      19.071  0.285 |
## nbmodemec  2.353  0.035 |
## id_pers    0.767  0.011 |
## id_depl    0.767  0.011 |
## id_traj    0.767  0.011 |
## ntraj      2.372  0.035 |
## btt        9.488  0.142 |
## Couteff    24.320  0.364 |

##
## Categories (the 10 first)
##           Dim.1    ctr   cos2  v.test    Dim.2    ctr
## ADMINISTRATIF | -0.061  0.001  0.000 -0.371 | -0.301  0.038
## AMENER/RECHERCHER QLQN |  0.768  0.541  0.064  9.012 | -0.277  0.110
## AUTRES        | -1.186  0.066  0.007 -2.984 |  1.801  0.239
## DOMICILE       | -0.145  0.069  0.012 -3.821 |  0.117  0.071
## ETUDES         | -1.579  0.089  0.010 -3.464 |  2.091  0.245
## LOISIRS        | -0.113  0.009  0.001 -1.143 | -0.113  0.014
## SHOOPING       |  0.448  0.129  0.015  4.331 | -0.602  0.367
## TRAVAIL        | -0.100  0.031  0.005 -2.528 |  0.083  0.034
## NON            |  0.045  0.017  0.017  4.918 |  0.090  0.104
## PAS_INFO       | -0.441  0.163  0.017 -4.918 | -0.876  1.011
##           cos2  v.test    Dim.3    ctr   cos2  v.test
## ADMINISTRATIF 0.003 -2.042 |  0.050  0.003  0.000  0.455 |
## AMENER/RECHERCHER QLQN 0.008 -3.634 | -0.146  0.096  0.002 -2.547 |
## AUTRES         0.016  5.077 |  7.758 13.951  0.290 29.129 |
## DOMICILE       0.008  3.452 | -0.155  0.390  0.014 -6.089 |
## ETUDES          0.017  5.141 |  0.331  0.019  0.000  1.085 |
## LOISIRS         0.001 -1.280 |  0.266  0.246  0.006  4.019 |
## SHOOPING        0.027 -6.519 | -0.403  0.517  0.012 -5.805 |
## TRAVAIL         0.004  2.351 |  0.105  0.169  0.006  3.953 |
## NON             0.069 10.943 |  0.125  0.632  0.133 20.284 |

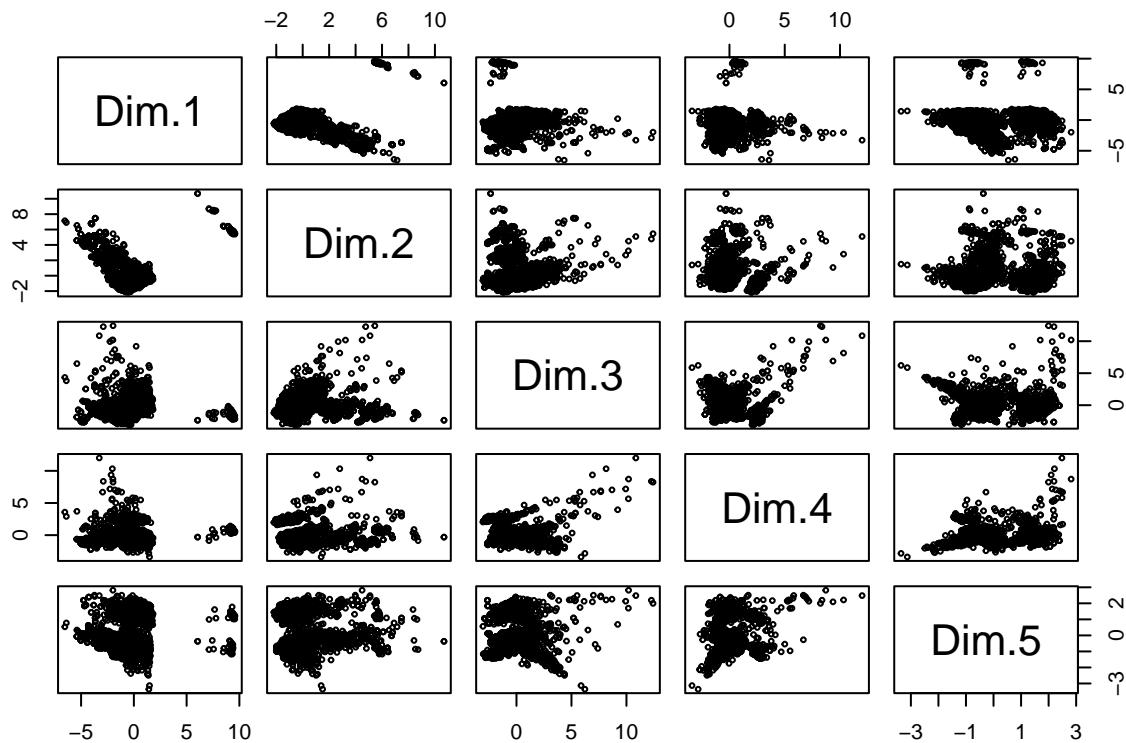
```

```
## PAS_INFO          0.069 -10.943 | -1.219  6.168  0.133 -20.284 |
```

```
head(afm_4$ind$coord)
```

```
##      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## 1 -0.4796943 -1.39975149  1.6735107 -1.62272467 -1.2187872
## 2 -0.5369225 -1.28657975  2.0716064 -1.10339134 -1.4743002
## 3 -1.1017878 -1.41583724 -0.2993740 -0.91838073 -0.5018004
## 4 -1.0766628 -1.57296275 -0.1623657 -0.86571795  1.3930771
## 5 -2.2543323 -0.32291463 -0.3256734 -0.01336904  1.3354352
## 6 -2.3431621 -0.05803611 -0.1141418  0.49146565 -0.7870002
```

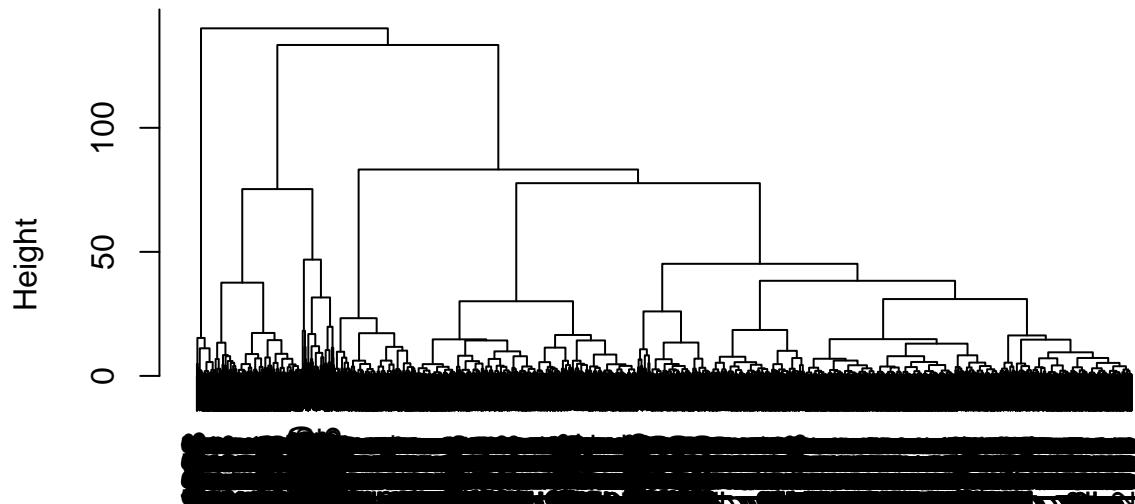
```
pairs(afm_4$ind$coord, cex=0.5)
```



```
md_4<-dist(afm_4$ind$coord)
ch_4<-hclust(md_4,method="ward.D2")
```

```
plot(ch_4)
```

Cluster Dendrogram



```
md_4  
hclust (*, "ward.D2")
```

```
groups_4<-cutree(ch_4,k=3)  
table(groups_4)
```

```
## groups_4
##      1     2     3
## 3450  514  79
```

```
#pairs(afm[4$ind$coord, col=c('cornflowerblue', 'yellowgreen')]) [groups 47, cew=0.75]
```

```
#lapply(DB_trip[sapply(DB_trip,is.numeric)],function(x){tapply(x,groups_4,sd)})  
lapply(DB_trip[sapply(DB_trip,is.numeric)],function(x){tapply(x,groups_4,mean)})
```

```

## $duree
##           1          2          3
## 17.22319 57.10895 14.97468
##
## $nbmodemec
##           1          2          3
## 1.000000 1.986381 1.126582
##
## $id_pers
##           1          2          3
## 123333045 119863025 302069419
##

```

```

## $id_depl
##           1          2          3
## 1233330452 1198630252 3020694196
##
## $id_traj
##           1          2          3
## 123333045223 119863025159 302069419566
##
## $ntraj
##           1          2          3
## 1.000290 1.994163 1.126582
##
## $btt
##           1          2          3
## 96.21913 268.46887 84.72152
##
## $Couteff
##           1          2          3
## 1.304822 1.093664 1.376748

#lapply(DB_trip[sapply(DB_trip, is.numeric)], function(x){tapply(x, groups_4, median)})

#lapply(DB_trip[sapply(DB_trip, is.numeric)], function(x){tapply(x, groups_4, var)})

sapply(DB_trip[sapply(DB_trip, is.factor)], function(x){prop.table(table(groups_4, x), 1)})

## $motifor
##           x
## groups_4 ADMINISTRATIF AMENER/RECHERCHER QLQN      AUTRES      CRECHE
##           1 0.030144928                  0.110724638 0.000000000 0.000000000
##           2 0.025291829                  0.031128405 0.040856031 0.000000000
##           3 0.025316456                  0.177215190 0.000000000 0.000000000
##           x
## groups_4     DOMICILE      ETUDES      LOISIRS      SHOOPING      TRAVAIL
##           1 0.357391304 0.002318841 0.079130435 0.080289855 0.340000000
##           2 0.412451362 0.015564202 0.073929961 0.017509728 0.383268482
##           3 0.354430380 0.000000000 0.050632911 0.037974684 0.354430380
##
## $prisecharge
##           x
## groups_4        NON    PAS_INFO
##           1 0.89710145 0.10289855
##           2 0.95914397 0.04085603
##           3 1.00000000 0.00000000
##
## $freqtcu
##           x
## groups_4          0          1
##           1 0.70753623 0.29246377
##           2 0.30155642 0.69844358
##           3 0.94936709 0.05063291
##
## $car_ownership

```

```

##      x
## groups_4    NON     OUI
##      1 0.1000000 0.9000000
##      2 0.4494163 0.5505837
##      3 0.0000000 1.0000000

```

Notre clustering mix s'est porté sur un échantillon après filtrage de 4043 trajets, deux groupes ont pu être mis en évidence, dans le premier groupe (3599 trajets) 72% des trajets ont été faits par des personnes n'utilisant pas fréquemment les transports en communs et dans le deuxième group (444 trajets) 76% des trajets ont été faits par des personnes utilisant fréquemment les transports en communs.

En moyenne les temps de trajets sont plus longs avec l'utilisation des transports en communs (btt_mean c'est à dire le temps passé dans un trajet quotidiennement qui vaut 268 minutes, dans le deuxième groupe contre 96 minutes dans le premier) ou bien duree qui est en moyene de 57 minutes contre 17 minutes dans le premier groupe. Il faudrait néanmmoins voir ce que nous disent le groupe travaillant sur les heures de pointes car il y a peut être un rapport différent durant ces horraires là.

Le motif de déplacement est assez similaire lorsqu'il s'agit de se rendre au travail, il n'y a pas de différence notoire, cependant, lorsqu'il s'agit de déposer une personne , l'usage des transports en communs n'est pas privilégié.

Toujours en lien avec notre variable (W) car_ownership, le fait d'avoir à disposition un véhicule favorise grandement son utilisation au détriment des TC , les trajets du premier groupe sont effectués avec 90% des personnes qui ont à disposition un véhicule contre 1 personne sur deux pour le deuxième groupe.

Analyse Univariée

Analyse Univariée : freqtcu

```

summary(DB_pers_socio_1$freqtcu)

##      0      1
## 803 461

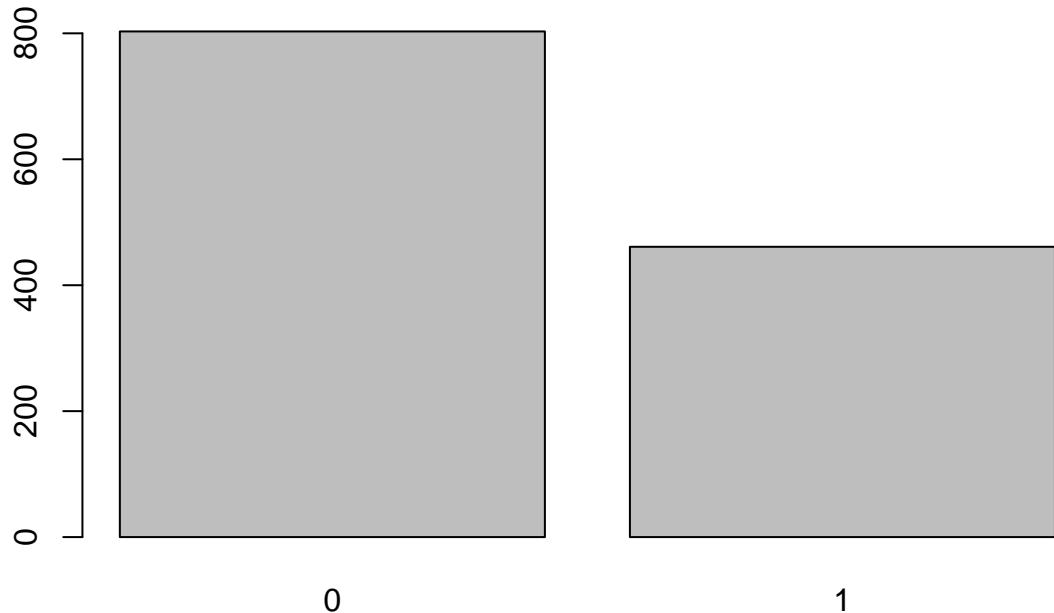
counts<-table(DB_pers_socio_1$freqtcu)
(counts/nrow(DB_pers_socio_1))*100

## 
##      0      1
## 63.52848 36.47152

barplot(counts)
title("Fréquence d'utilisation des transports en communs")

```

Fréquence d'utilisation des transports en communs



63% des personnes de notre base de données ne prennent pas régulièrement les transports en communs et 37% de manière régulière.

Analyse Univariée : permis

Toutes les personnes de notre échantillonnage possède le permis de conduire.

Analyse Univariée : tailmng

Pour ce qu'il en est de tailmng, la moyenne étant plus élevée que la médiane nous avons une asymétrie du côté droit , c'est à dire qu'il y a une concentration plus importante de valeurs à gauche de la moyenne.

58% des gens de notre restriction géographique vivent seul, 37% vivent à deux , 5% vivent à trois et moins de 1% vit tout seul

Analyse Univariée : car_ownership

80% des gens auraient à disposition une voiture pour se rendre au travail ou chez eux.

Analyse Bivariée

Pour cette partie, nous allons faire appel à plusieurs tests statistiques pour tenter de comprendre les relations qu'il peut y avoir entre nos variables.

Le test statistique du Chi^2 est utile pour établir ou non une relation entre deux variables qualitatives.

```
DB_var_zg<-New_DB_filtered[!duplicated(New_DB_filtered$id_pers),]  
#DB_var_zg<-allgreI faire avec allegreI  
#DB_var_zg<-New_DB_filtered%>%group_by(tir) faire un group_by tir et trouver des variables relatives au
```

Création des variables pour l'analyse bivariée

```

#DB_ABV<-
DB_ABV<-DB_pers_socio_1
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_permis=sum(permis=="YES"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_proprietaire_car=sum(car_ownership=="OUI"))

DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_actif=sum(OCCU1=="TravailPleinT" | OCCU1=="TravailPartT"))
#DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_inactif=sum(OCCU1=="RSA" | OCCU1=="Reste_auFoyer" | OCCU1==)
#DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_retraites=sum(OCCU1=="Retraite")) #pas de retraité à vérifi
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_high_profil=sum(csp=="HIGH_STATUT"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_low_profil=sum(csp=="LOW_STATUT"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_BAC_PLUS=sum(etabscol=="BAC+2" | etabscol=="BAC>=3"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_DIP_PRO=sum(etabscol=="BEP" | etabscol=="CAP"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_ABO_TC GRATUIT=sum(ABO_TC=="OUI_GRATUIT"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_ABO_TC_PAYANT=sum(ABO_TC=="OUI_PAYANT"))
#DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_TT=sum(travdom=="OUI"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_trav_ville=sum(zonet trav=="CENTRE_VILLE"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_trav_agglo=sum(zonet trav=="AGGLO_PROCHE"))
DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_trav_loin=sum(zonet trav=="HORS_GRENOBLE"))
#DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_voiture_dispo_utilisee=sum(dispo vop=="OUI UTILISE"))
#DB_ABV<-DB_ABV%>%group_by(tir)%>%mutate(NB_voiture_dispo_pas_utilisee=sum(dispo vop=="OUI UTILISE PAS"))

```

```

#permis


```

```

##
## Pearson's Chi-squared test
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_permis)
## X-squared = 103.93, df = 20, p-value = 2.48e-13

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_permis,data=DB_ABV))

##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_permis, data = DB_ABV)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -2.6175 -0.6121  0.3892  1.3874  1.3978
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.624243   0.174310  20.792   <2e-16 ***
## DB_ABV$NB_permis -0.000449   0.005348  -0.084    0.933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.364 on 1262 degrees of freedom
## Multiple R-squared:  5.585e-06, Adjusted R-squared:  -0.0007868
## F-statistic: 0.007048 on 1 and 1262 DF,  p-value: 0.9331

```

```

#chisq p<2.2e-16
#lm 1.472 10-6

#tailmng
fisher.test(table(DB_ABV$freqtcu,DB_ABV$tailmng))

## 
## Fisher's Exact Test for Count Data
##
## data: table(DB_ABV$freqtcu, DB_ABV$tailmng)
## p-value = 0.07788
## alternative hypothesis: two.sided

summary(lm(DB_ABV$freqtcu2~DB_ABV$tailmng,data = DB_ABV))

## 
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$tailmng, data = DB_ABV)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6903 -0.6903  0.3097  1.3097  1.8201
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.86050   0.09985 38.662 <2e-16 ***
## DB_ABV$tailmng -0.17016   0.06265 -2.716  0.0067 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.36 on 1262 degrees of freedom
## Multiple R-squared:  0.005812, Adjusted R-squared:  0.005024
## F-statistic: 7.377 on 1 and 1262 DF,  p-value: 0.006695

# p-value 0.07685
#R^2 0.005996

#nb_actif
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_actif),simulate.p.value = TRUE)

## 
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
## 
## data: table(DB_ABV$freqtcu, DB_ABV$NB_actif)
## p-value = 0.0004998
## alternative hypothesis: two.sided

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_actif,data=DB_ABV))

```

```

## 
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_actif, data = DB_ABV)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.7207 -0.6382  0.4031  1.3618  1.4855 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.49383   0.10447 33.442 <2e-16 ***
## DB_ABV$NB_actif 0.02062   0.01726  1.195   0.232  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.363 on 1262 degrees of freedom
## Multiple R-squared:  0.00113, Adjusted R-squared:  0.0003389 
## F-statistic: 1.428 on 1 and 1262 DF, p-value: 0.2323

```

```
#p-value 0.0004998
#R^2 0.0007595
```

```
#high_profil
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_high_profil),simulate.p.value = TRUE)
```

```

## 
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
## 
## data: table(DB_ABV$freqtcu, DB_ABV$NB_high_profil)
## p-value = 0.0004998
## alternative hypothesis: two.sided

```

```
summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_high_profil,data=DB_ABV))
```

```

## 
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_high_profil, data = DB_ABV)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.7146 -0.6522  0.3826  1.3617  1.4867 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.770154   0.150253 25.092 <2e-16 ***
## DB_ABV$NB_high_profil -0.006941   0.006295 -1.103   0.27  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.363 on 1262 degrees of freedom
## Multiple R-squared:  0.0009624, Adjusted R-squared:  0.0001708 
## F-statistic: 1.216 on 1 and 1262 DF, p-value: 0.2704

```

```

# pvalue = 0.0004998
#R^2 = 0.001598

#low_profil
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_low_profil),simulate.p.value = TRUE)

## 
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_low_profil)
## p-value = 0.0004998
## alternative hypothesis: two.sided

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_low_profil,data=DB_ABV))

## 
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_low_profil, data = DB_ABV)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7295 -0.6803  0.3690  1.3361  1.5166
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.46704   0.09629 36.008 <2e-16 ***
## DB_ABV$NB_low_profil 0.01640   0.01014  1.618   0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.362 on 1262 degrees of freedom
## Multiple R-squared:  0.00207,    Adjusted R-squared:  0.001279
## F-statistic: 2.618 on 1 and 1262 DF,  p-value: 0.1059

#p value = 0.0004998
#R^2 0.003519

#trav_ville
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_trav_ville),simulate.p.value = TRUE)

## 
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_trav_ville)
## p-value = 0.0004998
## alternative hypothesis: two.sided

```

```

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_trav_ville))

##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_trav_ville)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.7920 -0.7222  0.3707  1.3010  1.6729 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             3.861692   0.094226 40.983 < 2e-16 ***
## DB_ABV$NB_trav_ville -0.023243   0.007952 -2.923  0.00353 ** 
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 1.359 on 1262 degrees of freedom
## Multiple R-squared:  0.006724, Adjusted R-squared:  0.005937 
## F-statistic: 8.543 on 1 and 1262 DF, p-value: 0.00353

#p value = 0.0004998
#R^2 = 0.005998

```

```

#trav_aggro
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_trav_aggro),simulate.p.value = TRUE)

```

```

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_trav_aggro)
## p-value = 0.0004998
## alternative hypothesis: two.sided

```

```

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_trav_aggro))

```

```

##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_trav_aggro)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.9859 -0.8797  0.4034  1.1911  1.7927 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             3.03030   0.12187 24.864 < 2e-16 ***
## DB_ABV$NB_trav_aggro  0.03539   0.00707  5.005 6.36e-07 *** 
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

```

```

## 
## Residual standard error: 1.35 on 1262 degrees of freedom
## Multiple R-squared:  0.01947,   Adjusted R-squared:  0.01869
## F-statistic: 25.06 on 1 and 1262 DF,  p-value: 6.361e-07

#R^2 0.01874

#trav_loin
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_trav_loin),simulate.p.value = TRUE)

## 
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_trav_loin)
## p-value = 0.0004998
## alternative hypothesis: two.sided

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_trav_loin))

## 
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_trav_loin)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8924 -0.7692  0.3541  1.2308  1.9701
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.89237   0.07168  54.305 < 2e-16 ***
## DB_ABV$NB_trav_loin -0.06161   0.01325 -4.648  3.7e-06 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.352 on 1262 degrees of freedom
## Multiple R-squared:  0.01683,   Adjusted R-squared:  0.01605
## F-statistic: 21.61 on 1 and 1262 DF,  p-value: 3.7e-06

#R^2 0.01611

#voiture_dispo_utilisee

#fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_voiture_dispo_utilisee),simulate.p.value = TRUE)
#summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_voiture_dispo_utilisee))

#R^2 0.04952

#voiture_pas_utilisee

#fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_voiture_dispo_pas_utilisee),simulate.p.value = TRUE)
#summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_voiture_dispo_pas_utilisee))
#R^2 0.04983

```

```
#ABO_TC_GRATUIT
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_ABO_TC_GRATUIT),simulate.p.value = TRUE)
```

```
##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_ABO_TC_GRATUIT)
## p-value = 0.0004998
## alternative hypothesis: two.sided
```

```
summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_ABO_TC_GRATUIT))
```

```
##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_ABO_TC_GRATUIT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8899 -0.8899  0.3368  1.1101  2.0167
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.88988   0.05406  71.961 < 2e-16 ***
## DB_ABV$NB_ABO_TC_GRATUIT -0.22665   0.03145 -7.206 9.9e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.336 on 1262 degrees of freedom
## Multiple R-squared:  0.03952,    Adjusted R-squared:  0.03876
## F-statistic: 51.92 on 1 and 1262 DF,  p-value: 9.898e-13
```

```
#R^2 0.03367
```

```
#ABO_TC_PAYANT
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_ABO_TC_PAYANT),simulate.p.value = TRUE)
```

```
##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_ABO_TC_PAYANT)
## p-value = 0.0004998
## alternative hypothesis: two.sided
```

```
summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_ABO_TC_PAYANT))
```

```
##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_ABO_TC_PAYANT)
##
```

```

## Residuals:
##      Min     1Q Median     3Q    Max
## -3.2709 -0.9139  0.3681  0.9847  2.5182
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.27088   0.06936  61.58 <2e-16 ***
## DB_ABV$NB_ABO_TC_PAYANT -0.12779   0.01139 -11.21 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.3 on 1262 degrees of freedom
## Multiple R-squared:  0.09064, Adjusted R-squared:  0.08992
## F-statistic: 125.8 on 1 and 1262 DF, p-value: < 2.2e-16

```

#R^2 0.08841

```

#BAC_PLUS
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_BAC_PLUS),simulate.p.value = TRUE)

```

```

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_BAC_PLUS)
## p-value = 0.0004998
## alternative hypothesis: two.sided

```

```
summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_BAC_PLUS))
```

```

##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_BAC_PLUS)
##
## Residuals:
##      Min     1Q Median     3Q    Max
## -2.8387 -0.7139  0.3694  1.2861  1.6191
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.38088   0.08974  37.673 < 2e-16 ***
## DB_ABV$NB_BAC_PLUS    0.04162   0.01475   2.822  0.00485 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.359 on 1262 degrees of freedom
## Multiple R-squared:  0.006269, Adjusted R-squared:  0.005481
## F-statistic: 7.961 on 1 and 1262 DF, p-value: 0.004854

```

#R^2 0.005318

```

#DIP_PRO
fisher.test(table(DB_ABV$freqtcu,DB_ABV$NB_DIP_PRO),simulate.p.value = TRUE)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 2000 replicates)
##
## data: table(DB_ABV$freqtcu, DB_ABV$NB_DIP_PRO)
## p-value = 0.0004998
## alternative hypothesis: two.sided

summary(lm(DB_ABV$freqtcu2~DB_ABV$NB_DIP_PRO))

##
## Call:
## lm(formula = DB_ABV$freqtcu2 ~ DB_ABV$NB_DIP_PRO)
##
## Residuals:
##      Min      1Q      Median      3Q      Max 
## -2.8390 -0.8073   0.3829   1.2561   1.6365 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.300105  0.102971  32.05 < 2e-16 ***
## DB_ABV$NB_DIP_PRO 0.031701  0.009783   3.24  0.00122 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.358 on 1262 degrees of freedom
## Multiple R-squared:  0.008252, Adjusted R-squared:  0.007466 
## F-statistic: 10.5 on 1 and 1262 DF,  p-value: 0.001225

```

Matrice de corrélation

D=data.frame(DB ABV\$NB DIP PRO.DB ABV\$NB BAC PLUS.DB ABV\$NB ABO TC GRATUIT.DB ABV\$NB ABO TC PAYANT.DB A)

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
MC<-cor(D)
#MC<-cor(DB_ABV[,-c(21,22,29)])
symnum(MC, abbr.colnames=FALSE)
```

```
## DB_ABV.NB_DIP_PRO DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_DIP_PRO 1
## DB_ABV.NB_BAC_PLUS 1
## DB_ABV.NB_ABO_TC GRATUIT
```

```

## DB_ABV.NB_ABO_TC_PAYANT
## DB_ABV.NB_trav_loin
## DB_ABV.NB_permis
## DB_ABV.NB_trav_loin.1
## DB_ABV.NB_trav_agglo
## DB_ABV.NB_trav_ville
## DB_ABV.NB_low_profil
## DB_ABV.NB_high_profil
## DB_ABV.NB_actif
## DB_ABV.NB_DIP_PRO
## DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_ABO_TC_GRATUIT 1
## DB_ABV.NB_ABO_TC_PAYANT . 1
## DB_ABV.NB_trav_loin .
## DB_ABV.NB_permis
## DB_ABV.NB_trav_loin.1 .
## DB_ABV.NB_trav_agglo .
## DB_ABV.NB_trav_ville .
## DB_ABV.NB_low_profil
## DB_ABV.NB_high_profil
## DB_ABV.NB_actif
## DB_ABV.NB_trav_loin DB_ABV.NB_permis
## DB_ABV.NB_DIP_PRO
## DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_ABO_TC_GRATUIT
## DB_ABV.NB_ABO_TC_PAYANT
## DB_ABV.NB_trav_loin 1
## DB_ABV.NB_permis . 1
## DB_ABV.NB_trav_loin.1 1 .
## DB_ABV.NB_trav_agglo .
## DB_ABV.NB_trav_ville .
## DB_ABV.NB_low_profil .
## DB_ABV.NB_high_profil .
## DB_ABV.NB_actif
## DB_ABV.NB_trav_loin.1 DB_ABV.NB_trav_agglo
## DB_ABV.NB_DIP_PRO
## DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_ABO_TC_GRATUIT
## DB_ABV.NB_ABO_TC_PAYANT
## DB_ABV.NB_trav_loin
## DB_ABV.NB_permis
## DB_ABV.NB_trav_loin.1 1
## DB_ABV.NB_trav_agglo . 1
## DB_ABV.NB_trav_ville .
## DB_ABV.NB_low_profil .
## DB_ABV.NB_high_profil .
## DB_ABV.NB_actif
## DB_ABV.NB_trav_ville DB_ABV.NB_low_profil
## DB_ABV.NB_DIP_PRO
## DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_ABO_TC_GRATUIT
## DB_ABV.NB_ABO_TC_PAYANT
## DB_ABV.NB_trav_loin

```

```

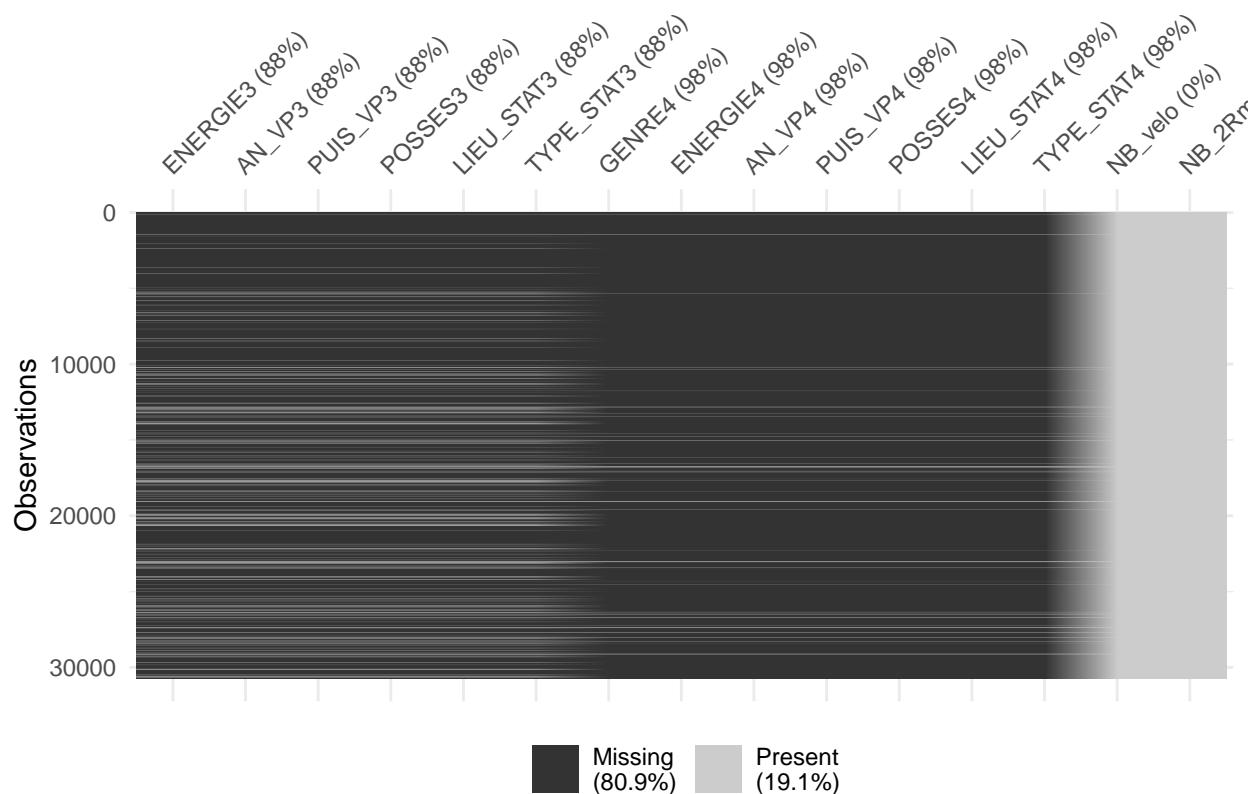
## DB_ABV.NB_permis
## DB_ABV.NB_trav_loin.1
## DB_ABV.NB_trav_agglo
## DB_ABV.NB_trav_ville      1
## DB_ABV.NB_low_profil           1
## DB_ABV.NB_high_profil       .
## DB_ABV.NB_actif
##                               DB_ABV.NB_high_profil DB_ABV.NB_actif
## DB_ABV.NB_DIP_PRO
## DB_ABV.NB_BAC_PLUS
## DB_ABV.NB_ABO_TC_GRATUIT
## DB_ABV.NB_ABO_TC_PAYANT
## DB_ABV.NB_trav_loin
## DB_ABV.NB_permis
## DB_ABV.NB_trav_loin.1
## DB_ABV.NB_trav_agglo
## DB_ABV.NB_trav_ville
## DB_ABV.NB_low_profil
## DB_ABV.NB_high_profil      1
## DB_ABV.NB_actif           1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1

```

```
#corrplot(MC, type="upper", order="hclust", tl.col="black", tl.srt=50)
```

Annexes

```
data_2<-DB_projet_full[,c(30:44)]  
vis_miss(  
  data_2,  
  cluster = FALSE,  
  sort_miss = FALSE,  
  show_perc = TRUE,  
  show_perc_col = TRUE,  
  large_data_size = 9e+06,  
  warn_large_data = TRUE  
)
```



Listes variables à plus de 80% de valeurs manquantes

-motoracc -situveil -STAT_TRAV -TYPE_STAT4 -LIEU_STAT4 -POSSES4 -PUIS_VP4 -AN_VP4 -ENERGIE4 -GENRE4 -TYPE_STAT3 -LIEU_STAT3 -POSSES3 -PUIS_VP3 -AN_VP3 -ENERGIE3 -motdeacc -nbarret -abonpeage