

concordance=TRUE SinputVerbatimformatcom = SoutputVerbatimformatcom =

## L2S1 - UE4-2 - INITIATION AU LOGICIEL R 2

# TP4 - ACP - CORRIGÉ EXERCICE 1

### Exercice 1.

1. Importer le fichier `DonneesVoiture.csv` dans un `data.frame` intitulé `data`. Les séparateurs sont des tabulations.

```
data <- read.table("DonneesVoitures.txt", sep=";", header=T, row.names=1)
head(data)
```

	Car	Prix	Cyl	Puissance	Couple	Trans
C3 Picasso 1.6 HDI BMP 6	Diesel	22650	1.6	92	230	Avant
Citroen C5 II 2.2 HDI	Diesel	37450	2.2	204	450	Avant
Citroen C1 1.0i	Essence	9550	1.0	68	93	Avant
Citroen C8 HDI	Diesel	37050	2.0	160	340	Avant
Renault Clio Energy Tce 90	Essence	16100	0.9	90	135	Avant
Renault Grand Espace 4 2.0 dCi	Diesel	36900	2.0	130	320	Avant

	VitMax	X0a100	C1	C2	C3	C02	Niveau
C3 Picasso 1.6 HDI BMP 6	174	13.5	5.5	4.0	4.6	109	B
Citroen C5 II 2.2 HDI	230	8.3	8.0	4.7	5.9	155	D
Citroen C1 1.0i	160	12.3	3.8	5.1	4.3	99	A
Citroen C8 HDI	203	9.8	7.4	5.0	5.9	155	D
Renault Clio Energy Tce 90	182	12.2	6.2	4.2	4.9	114	B
Renault Grand Espace 4 2.0 dCi	184	12.0	7.0	5.0	5.7	150	D

	Autonomie	Poids	L	l	h	Reservoir
C3 Picasso 1.6 HDI BMP 6	902	1313	4.08	1.73	1.63	50
Citroen C5 II 2.2 HDI	985	1731	4.78	1.86	1.46	70
Citroen C1 1.0i	676	790	3.43	1.63	1.47	35
Citroen C8 HDI	1125	1770	4.73	1.86	1.86	80
Renault Clio Energy Tce 90	762	1009	4.06	1.73	1.45	45
Renault Grand Espace 4 2.0 dCi	1209	1901	4.86	1.86	1.75	83

	VolCoffre
C3 Picasso 1.6 HDI BMP 6	385
Citroen C5 II 2.2 HDI	467
Citroen C1 1.0i	130
Citroen C8 HDI	1160
Renault Clio Energy Tce 90	300
Renault Grand Espace 4 2.0 dCi	456

2. 

```
library("FactoMineR", lib.loc="C:/Program Files/R/R-3.2.0/library")
```

```
## Warning: package 'FactoMineR' was built under R version 3.2.3
```

3. On ne peut directement effectuer une ACP sur ce tableau de données car les variables ne sont pas toutes quantitatives.

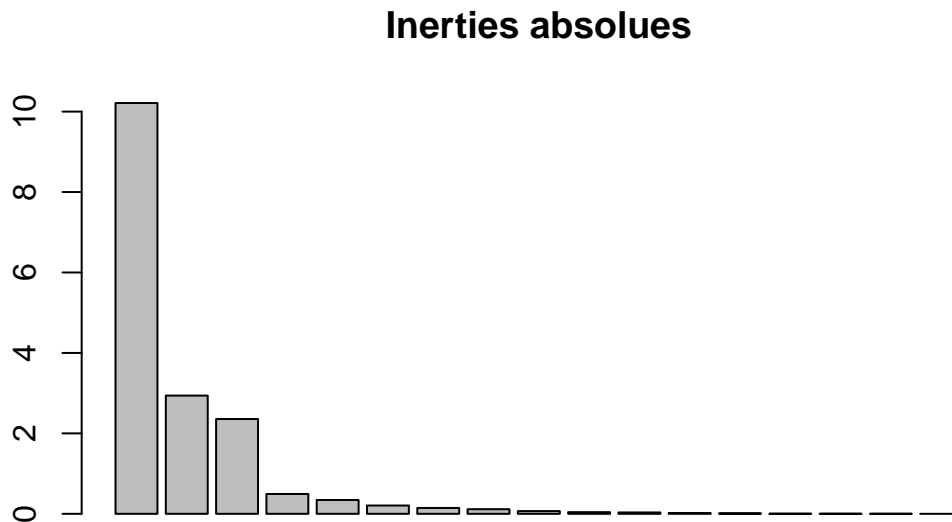
```
resacp<-PCA(data, quali.sup=c(1,6,13))
```

Dans cette commande, on précise que les colonnes 1, 6 et 13 sont des variables qualitatives, la variable 2 (Prix) est supplémentaire. Si on veut automatiser la procédure de recherche des variables qualitatives on peut faire ainsi avec une boucle :

```
nc<-ncol(data)
filtre<-NULL
for(i in 1:nc){
  filtre<-c(filtre, is.numeric(data[,i]))
}
resacp<-PCA(data, quali.sup=which(filtre==F))
```

4. Effectuer un graphique qui représente les inerties et qui permet de justifier le nombre de facteurs à conserver (commandes possibles `barplot` ou `plot`)

```
barplot(resacp$eig[,1],main="Inerties absolues",
        xlab="Valeurs propres ou Inerties absolues")
```



### Valeurs propres ou Inerties absolues

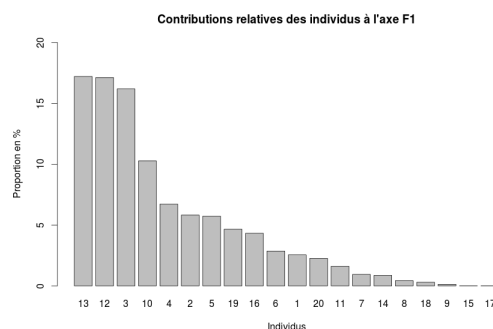
5. La commande suivante :

```
summary.PCA(resacp,file="resume.txt")
```

permet de créer un fichier texte contenant le résumé des caractéristiques des 10 premiers individus et des premières variables pour les 3 premiers axes avec 3 décimales. La commande suivante permet d'obtenir un résumé qui ne contienne que les données pour les deux premières composantes, mais pour tous les individus.

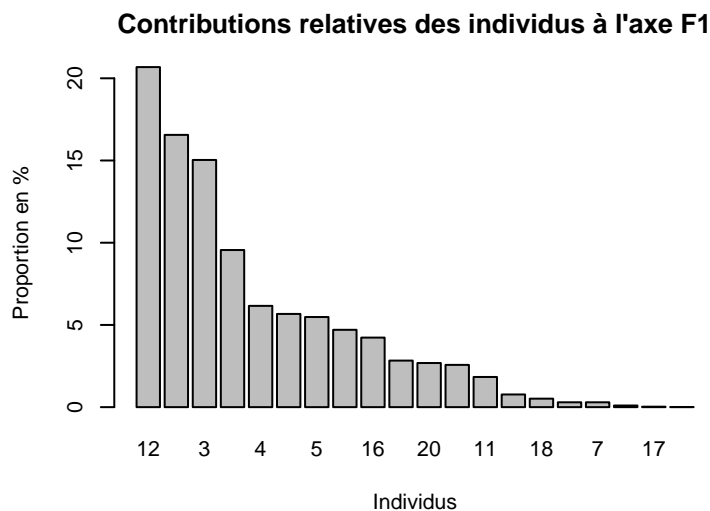
```
summary.PCA(resacp,file="resume.txt",ncp=2,nbelements=Inf,nbind=Inf)
```

6. On veut étudier finement les individus qui ont contribué à l'axe  $F_1$ .
- Créer un vecteur `ContriInd1` contenant les contributions des individus pour le premier axe.
  - Modifier le nom des valeurs de `ContriInd1` en les numérotant de 1 à 20 en utilisant la commande `names()`.
  - En utilisant la commande `sort` classer dans l'ordre décroissant `ContriInd1`.
  - Effectuer le graphique suivant :



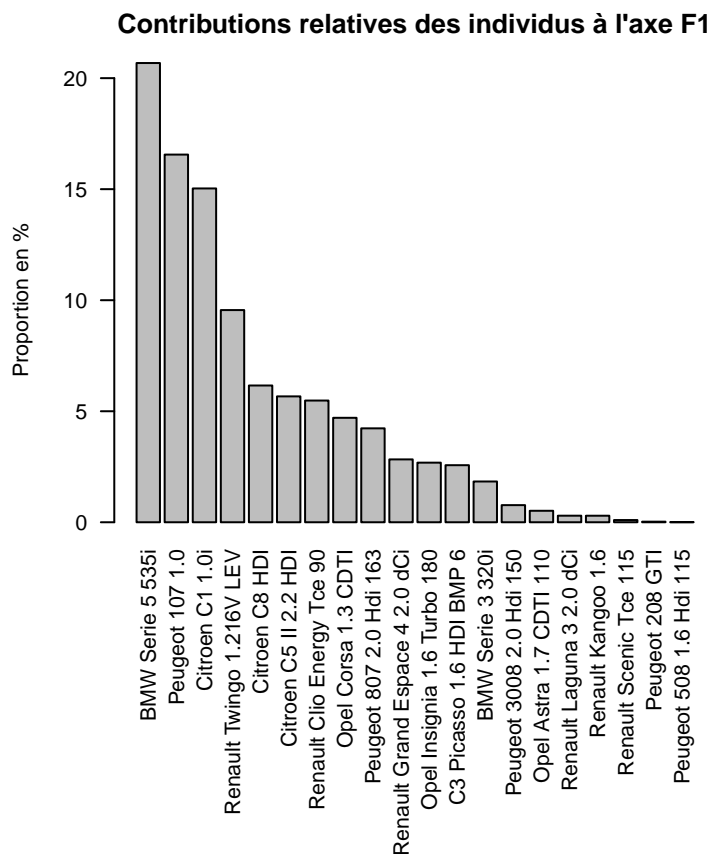
```
par(cex=0.7)
ContriInd1<-resacp$ind$contrib[,1]
names(ContriInd1)<-1:20
vec<-sort(ContriInd1,decreasing=T)
```

```
barplot(vec,main="Contributions relatives des individus à l'axe F1",
        xlab="Individus",ylab="Proportion en %",ylim=c(0,20))
```



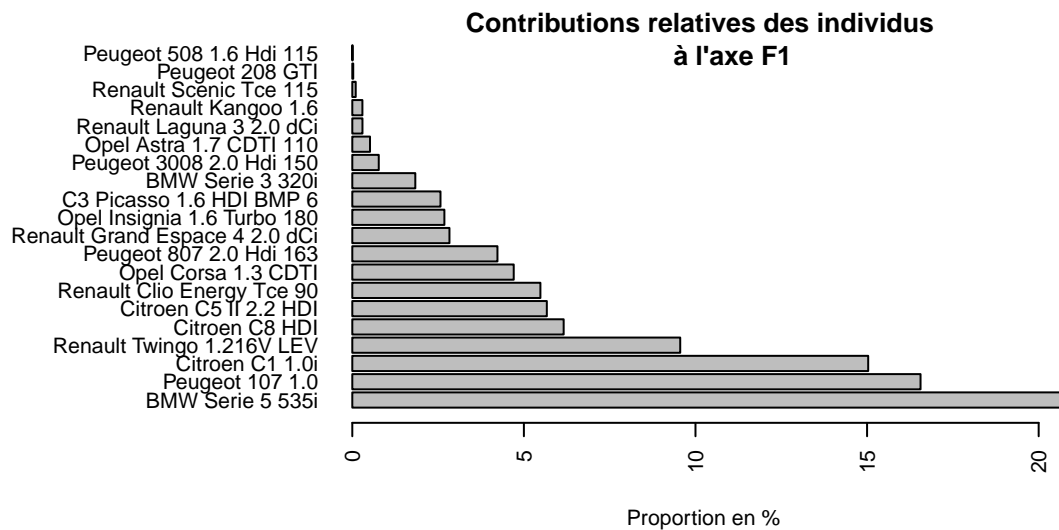
ou sans modifier les noms des variables mais les lire quand même :

```
ContriInd1b<-resacp$ind$contrib[,1]
vecb<-sort(ContriInd1b,decreasing=T)
par(mar=c(15.1,4.1,4.1,2.1))
par(cex=0.7)
barplot(vecb,main="Contributions relatives des individus à l'axe F1",
        xlab="",ylab="Proportion en %",ylim=c(0,20),las=2)
```



ou encore...

```
par(mar=c(5.1,14.1,4.1,2.1))
par(cex=0.7)
barplot(vecb,main="Contributions relatives des individus
à l'axe F1",
ylab="",xlab="Proportion en %",ylim=c(0,20),horiz=T,las=2)
```



```
cumsum(vecb)[3]
## Citroën C1 1.0i
## 52.27104
```

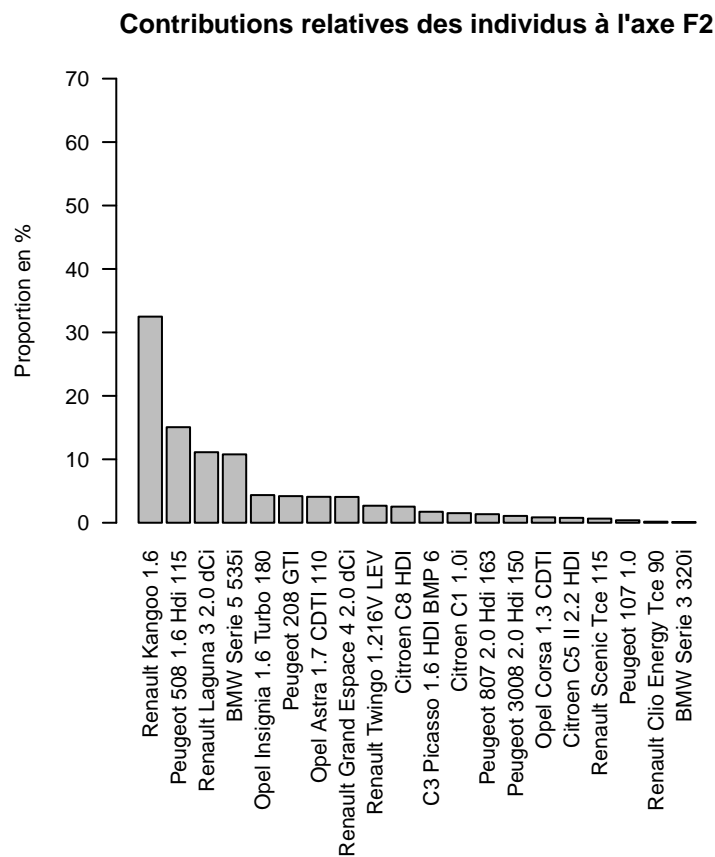
La commande `par` permet de modifier les paramètres graphiques : ici les marges (argument `mar`) de la fenêtre graphique ont été adaptées, la ainsi que la taille des polices avec l'argument `cex`.

- (e) Les voitures ont contribué à l'axe F1 sont la Peugeot 107, BMW série 5 et Citroën C1 avec une contribution cumulée de 50,5 %

```
cumsum(vec)[3]
## 3
## 52.27104
```

- (f) Faire de même avec l'axe F2. A partir d'un copier/coller dans le script on obtient rapidement :

```
ContriInd2<-resacp$ind$contrib[,2]
vec2<-sort(ContriInd2,decreasing=T)
par(mar=c(15.1,4.1,4.1,2.1))
par(cex=0.7)
barplot(vec2,main="Contributions relatives des individus à l'axe F2",
xlab="",ylab="Proportion en %",ylim=c(0,70),las=2)
```



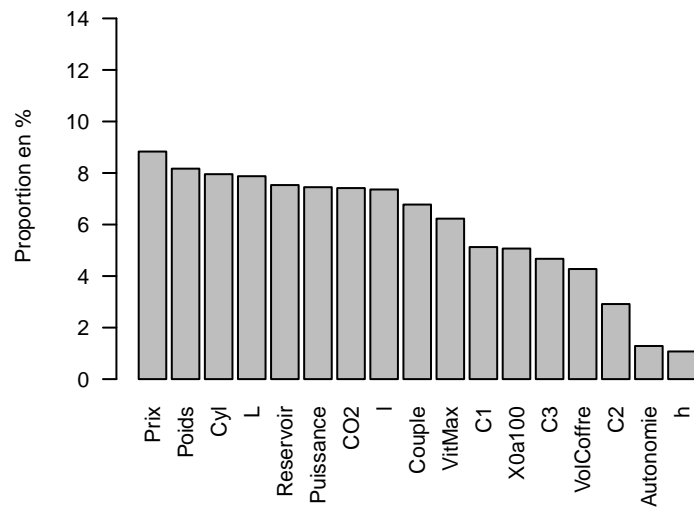
```
cumsum(vec2)[3]
## Renault Laguna 3 2.0 dCi
## 58.66619
```

Ici aussi on peut garder les 3 premières voitures, surtout pour montrer l'opposition que suggère F2.

7. Reprendre les questions précédentes avec les variables.

```
ContriVar1<-resacp$var$contrib[,1]
vec3<-sort(ContriVar1,decreasing=T)
par(mar=c(10.1,4.1,4.1,2.1))
par(cex=0.7)
barplot(vec3,main="Contributions relatives des variables à l'axe F1",
        xlab="",ylab="Proportion en %",ylim=c(0,15),las=2)
```

## Contributions relatives des variables à l'axe F1



```
cumsum(vec3)[8]
```

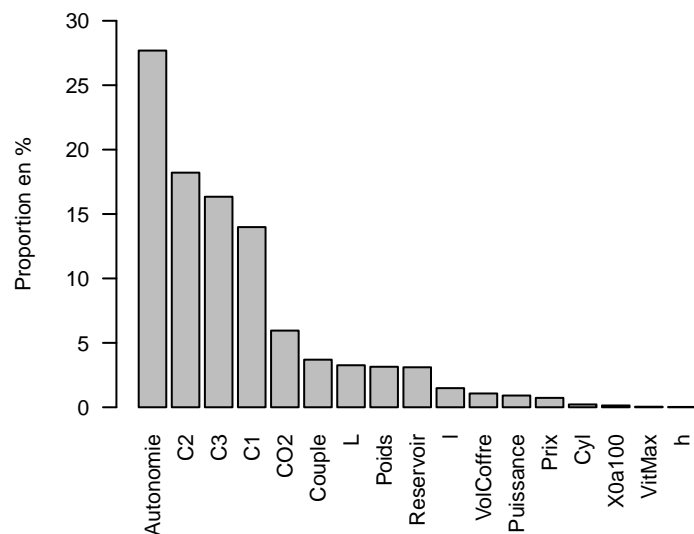
```
##      1
```

```
## 62.58682
```

On peut expliquer le premier axe par les 8 premières variables qui cumulent 70% des contributions

```
ContriVar2<-resacp$var$contrib[,2]
vec4<-sort(ContriVar2,decreasing=T)
par(mar=c(10.1,4.1,4.1,2.1))
par(cex=0.7)
barplot(vec4,main="Contributions relatives des variables à l'axe F2",
        xlab="",ylab="Proportion en %",ylim=c(0,30),las=2)
```

## Contributions relatives des variables à l'axe F2



```
cumsum(vec4)[4]

##          C1
## 76.22608
```

On conserve les 4 premiers variables qui contribue à 76%.

8. Tableau caractéristique (complet) des individus :

```
TabInd<-data.frame(round(resacp$ind$coord[,c(1,2,3)],2),
                    round(resacp$ind$contrib[,c(1,2,3)],2),
                    round(resacp$ind$cos2[,c(1,2,3)],3),
                    round(resacp$ind$cos2[,1]+resacp$ind$cos2[,2],3),
                    round(resacp$ind$cos2[,3]+resacp$ind$cos2[,2],3))
colnames(TabInd)<-c("F1","F2","F3","ctr(F1)","ctr(F2)","ctr(F3)",
                    "qlt(F1)","qlt(F2)","qlt(F3)",
                    "qlt(F1,F2)","qlt(F2,F3)")
head(TabInd)

##          F1      F2      F3 ctr(F1) ctr(F2) ctr(F3)
## C3 Picasso 1.6 HDI BMP 6 -2.29  1.01  0.68    2.57    1.73    0.98
## Citroen C5 II 2.2 HDI    3.40  0.67 -1.56    5.67    0.77    5.17
## Citroen C1 1.0i         -5.54 -0.94 -0.33   15.03    1.51    0.23
## Citroen C8 HDI          3.55  1.22  2.51    6.16    2.54   13.33
## Renault Clio Energy Tce 90 -3.35 -0.31 -0.31    5.48    0.16    0.20
## Renault Grand Espace 4 2.0 dCi 2.40  1.55  1.90    2.83    4.07    7.63
##          qlt(F1) qlt(F2) qlt(F3) qlt(F1,F2)
## C3 Picasso 1.6 HDI BMP 6    0.651    0.126    0.058    0.777
## Citroen C5 II 2.2 HDI      0.744    0.029    0.157    0.774
## Citroen C1 1.0i            0.900    0.026    0.003    0.926
## Citroen C8 HDI             0.525    0.062    0.262    0.588
## Renault Clio Energy Tce 90  0.906    0.008    0.008    0.914
## Renault Grand Espace 4 2.0 dCi 0.409    0.169    0.255    0.579
##          qlt(F2,F3)
## C3 Picasso 1.6 HDI BMP 6    0.184
## Citroen C5 II 2.2 HDI      0.186
## Citroen C1 1.0i            0.029
## Citroen C8 HDI             0.325
## Renault Clio Energy Tce 90  0.015
## Renault Grand Espace 4 2.0 dCi 0.424

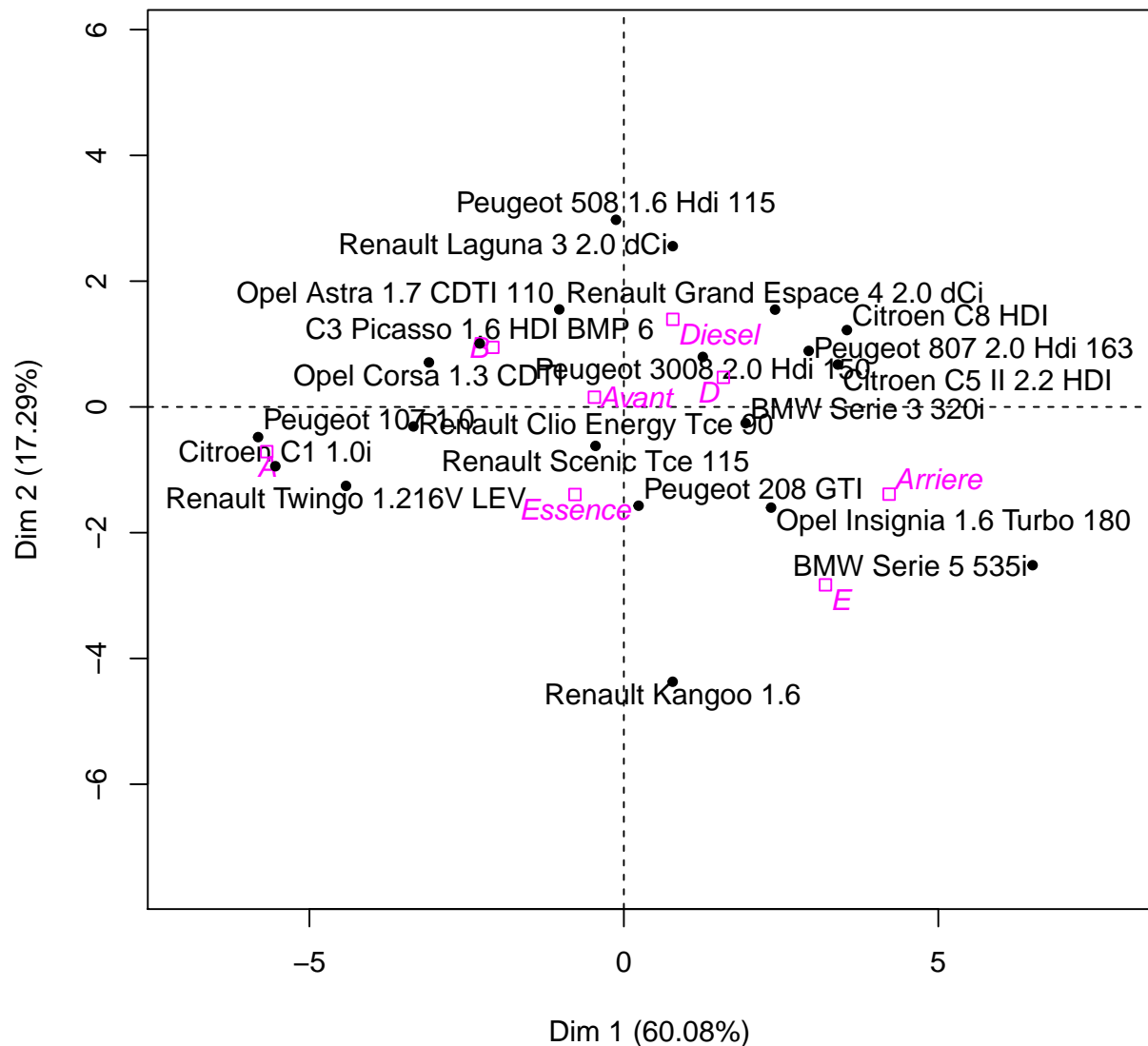
write.table(TabInd,file="TabInd.csv",sep=";")
```

le format csv permet qui utilise comme séparateur le point virgule est facilement lu par les tableurs et insérable dans les traitement de texte pour former des tableaux.

9. Effectuer et comparer les 3 commandes suivantes.

```
plot.PCA(resacp,choix="ind")
```

### Individuals factor map (PCA)

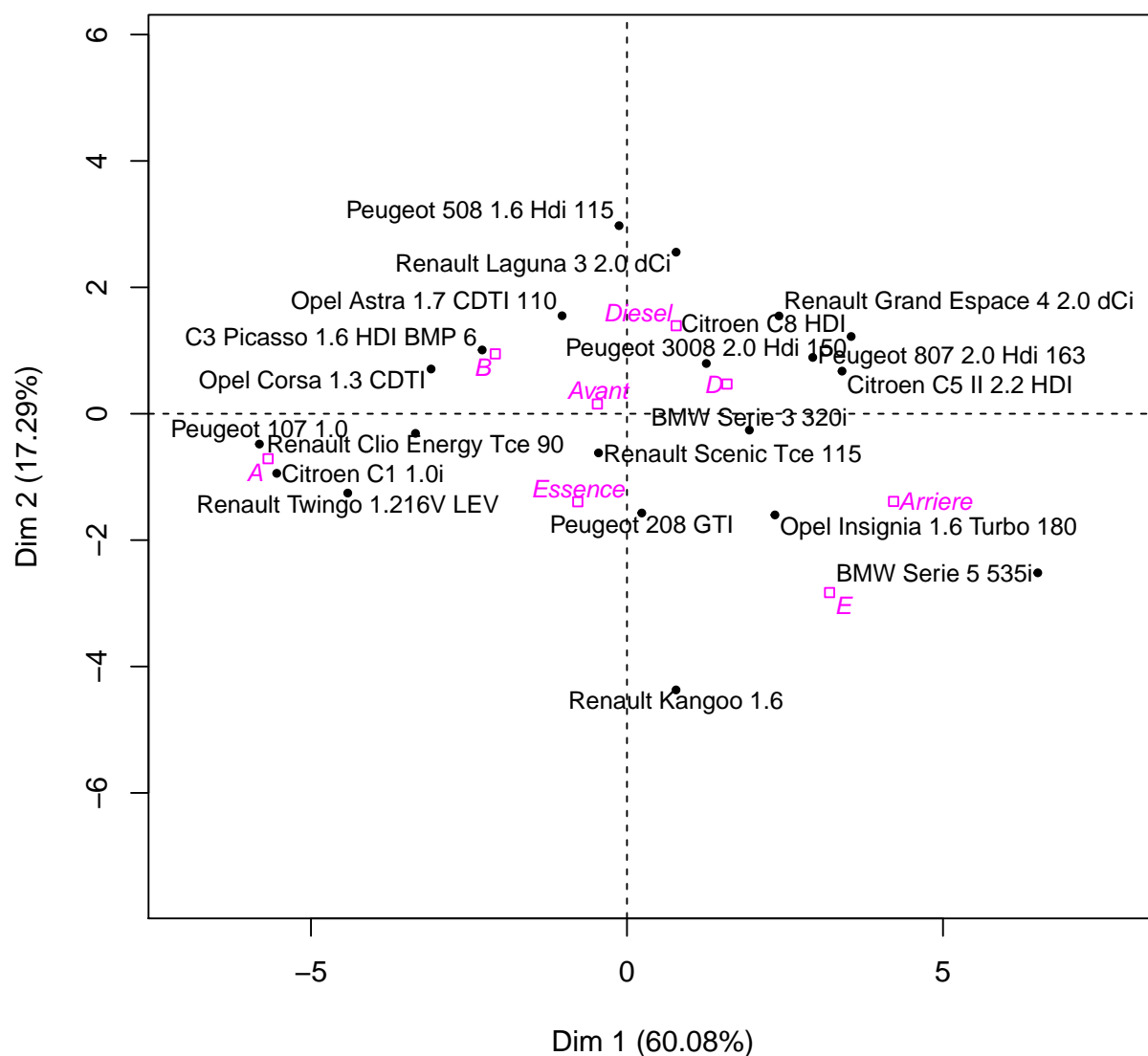


Représentation du nuage des individus avec titre par défaut, des caractères qui se superposent, et les variables qualitatives qui sont positionnées.

```
plot.PCA(resacp,choix="ind",
          title="Nuages des Individus sur le plan F1-F2",cex=0.8)
```



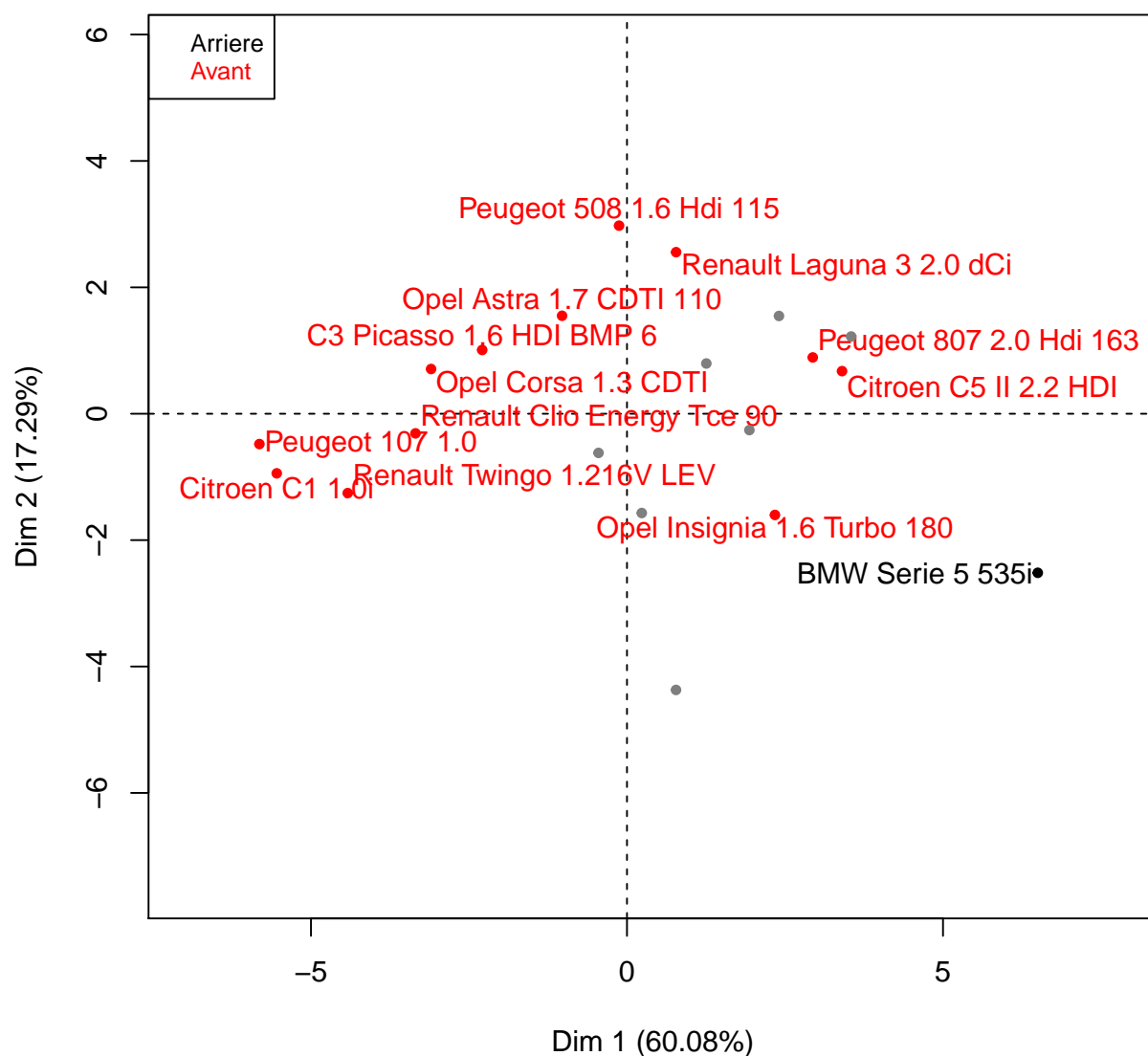
### Nuages des Individus sur le plan F1-F2



La taille des caractères des individus a diminué, et le titre est maintenant précisé.

```
plot.PCA(resacp, choix="ind", title="Nuages des Individus sur le plan F1,F2",
  select="cos2 0.6", unselect="grey50", invisible="quali", habillage=6)
```

### Nuages des Individus sur le plan F1,F2

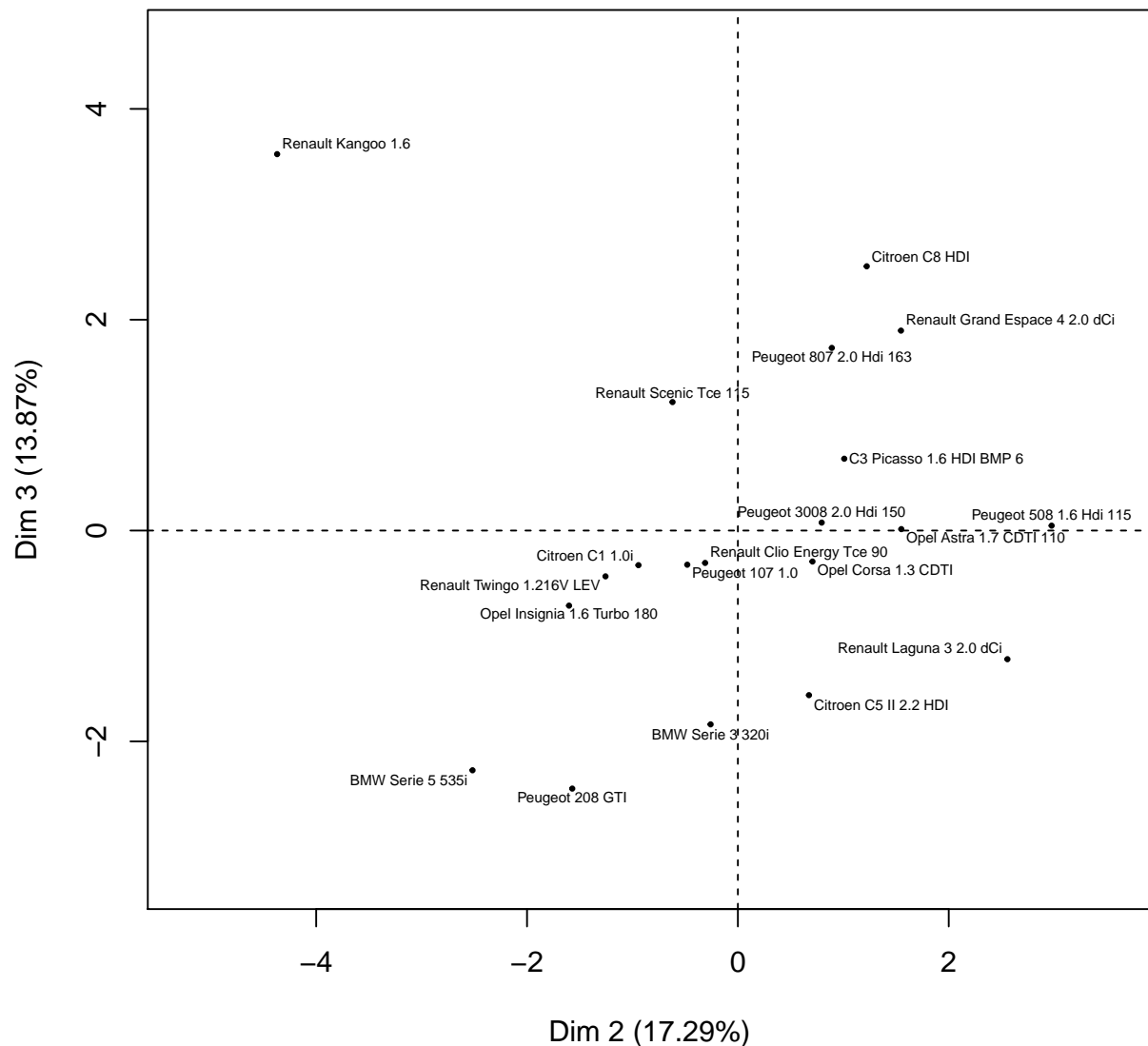


Seul les individus ayant une qualité de représentation supérieure à 0.6 sont représentés avec leur désignation, les autres sont signalés par un point gris clair. Les variables qualitative ne sont plus représentés. Les individus sont colorés (habillés) en fonction de leur valeur qualitative de la colonne 6 : la transmission.

10. Effectuer la représentation graphique du nuage des individus sur le plan F2-F3.

```
plot.PCA(resacp,choix="ind",axes=c(2,3),cex=0.5,invisible="quali",
         title="Nuages des Individus sur le plan F2,F3")
```

## Nuages des Individus sur le plan F2,F3



11. Produire le tableau des caractéristiques des variables dans le même format que celui des individus. Effectuer la représentation graphique du nuage des variables sur F1-F2 et sur F2-F3. Soigner la forme.

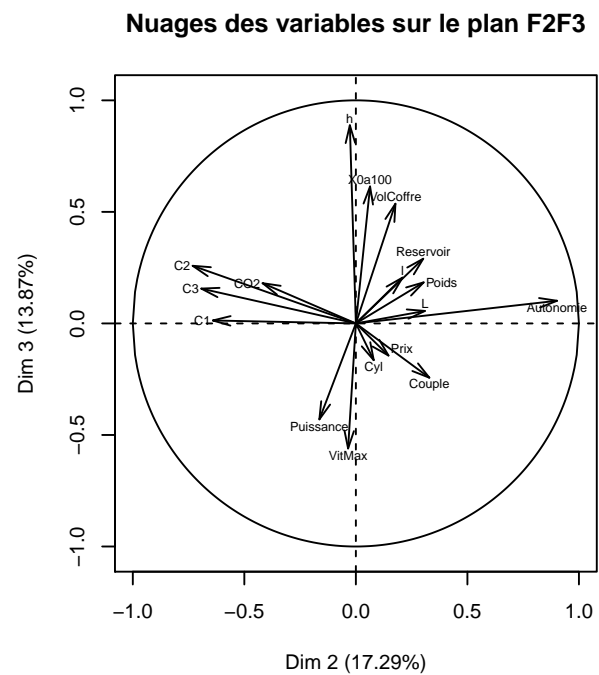
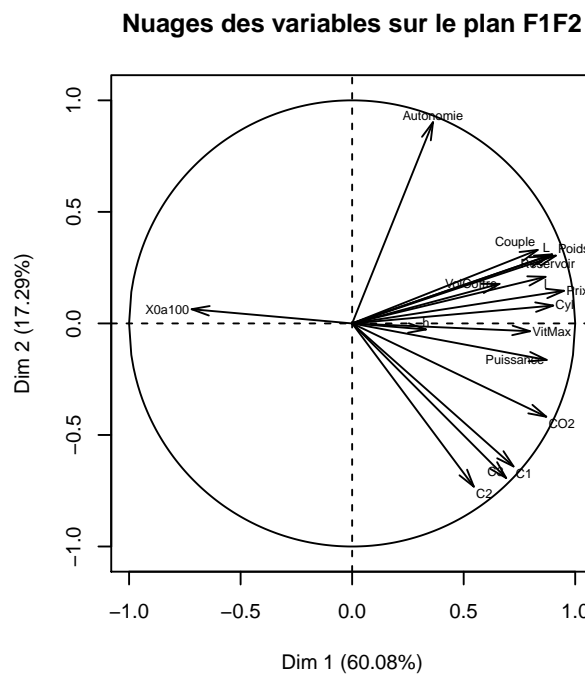
```
TabCarVar<-data.frame(round(resacp$var$coord[,c(1,2,3)],2),
  round(resacp$var$contrib[,c(1,2,3)],2),
  round(resacp$var$cos2[,c(1,2,3)],3),
  round(resacp$var$cos2[,1]+resacp$var$cos2[,2],3),
  round(resacp$var$cos2[,3]+resacp$var$cos2[,2],3))
colnames(TabCarVar)<-c("F1", "F2", "F3", "ctr(F1)", "ctr(F2)", "ctr(F3)",
  "qlt(F1)", "qlt(F2)", "qlt(F3)",
  "qlt(F1,F2)", "qlt(F2,F3)")
head(TabCarVar)
```

	F1	F2	F3	ctr(F1)	ctr(F2)	ctr(F3)	qlt(F1)	qlt(F2)
## Prix	0.95	0.15	-0.14	8.83	0.73	0.88	0.902	0.021
## Cyl	0.90	0.08	-0.16	7.95	0.22	1.15	0.812	0.006
## Puissance	0.87	-0.16	-0.43	7.45	0.91	7.85	0.761	0.027
## Couple	0.83	0.33	-0.24	6.77	3.69	2.51	0.692	0.109
## VitMax	0.80	-0.03	-0.56	6.23	0.04	13.32	0.636	0.001
## X0a100	-0.72	0.06	0.61	5.07	0.14	15.98	0.518	0.004
##	qlt(F3)	qlt(F1,F2)	qlt(F2,F3)					

```
## Prix      0.021    0.924    0.042
## Cyl       0.027    0.819    0.034
## Puissance 0.185    0.788    0.212
## Couple    0.059    0.801    0.168
## VitMax    0.314    0.637    0.315
## X0a100    0.377    0.522    0.381
```

```
write.table(TabCarVar,file="TabCarVar.csv",sep=";")
```

```
par(mfrow=c(1,2))
par(cex=0.7)
plot.PCA(resacp,choix="var",title="Nuages des variables sur le plan F1F2",cex=0.6)
plot.PCA(resacp,choix="var",axes=c(2,3),title="Nuages des variables sur le plan F2F3",cex=0.6)
```



12. Commenter les résultats obtenus, signification des axes, association de véhicules... On peut remarquer à partir des contributions et du nuages des variables que  $F_1$  correspond aux caractéristiques moteurs en lien avec le prix. L'axe  $F_2$  explique plutôt les différences liés à l'autonomie de la voiture et la consommation. Le nuage des individus confirme la signification de l'axe  $F_1$ , on observe que les petites voitures citadines sont situées sur la partie gauche du graphique, les grandes routières et les grands monospace sont situés sur la partie droite. De plus on remarque que les véhicules sportifs ou fortement motorisés sont en bas à droite, "consommateur et cher", tandis qu'en haut se situe les véhicules avec une forte autonomie. *On pouvait à partir de là effectuer des regroupements...*