

ANALYSE DE DONNÉES SUR LES COMMUNES DE FRANCE

Structure des données

La base de données initiales décrit 36689 communes de France à l'aide de 33 variables. Dans cette analyse, les communes sont d'abord regroupées par départements et de nouvelles variables sont créées : la densité moyenne de population par département en 2013 et la croissance de population entre 2008 et 2013. Le code R qui correspond à ces modification est décrit ci-dessous.

Le bloc suivant charge la base de données depuis le répertoire courant et supprime les colonnes des variables non pertinentes telles que celles qui donnent le code géographique, le nom de la commune ou encore la région. Enfin, la dernière instruction permet d'afficher la table dans l'environnement.

```
1 #Load and create the dataframes
2 table = read.csv("base_cc_resume_20161013_COM.csv")
3 table <- table[,!(colnames(table) %in% c("CODGEO", "LIBGEO", "REG"))]
4 View(table)
```

Dans la routine suivante, on crée la data frame *department*. Chaque individu est un département, et chaque département est décrit par les mêmes variables que dans la table précédente.

```
1 #Create and dimension department table
2 department = data.frame()
3 listdepartment = levels(table[,colnames(table) == "DEP"])
4
5 nb_departments = length(listdepartment)
6 department[1:nb_departments,1] <- NA
7 rownames(department) <- listdepartment
8 rm(nb_departments)
9
10 nb_variables = length(colnames(table))-1
11 department[,1:nb_variables] <- NA
12 colnames(department) <- colnames(table)[2:(nb_variables+1)]
13 rm(nb_variables)
14 rm(listdepartment)
15
16 View(department)
```

La routine suivante permet de remplir la data frame *department* en sommant, pour chaque département, les caractéristiques de toutes les communes dudit département. On s'occupe aussi individuellement des variables intensives pour lesquelles la somme n'aurait aucun sens. Pour ces variables, on considère plutôt la moyenne pondérée par le nombre d'habitant en 2013.

```
1 #Fill in the department table
2 variables = colnames(table)[2:length(colnames(table))]
3
4 for(i in rownames(department))
5 {
6   conditionL = (table[, "DEP"] == i)
7   for(j in colnames(department)){
8     department[i,j] = sum(na.omit(table[conditionL,j]))
9   }
10  department[i,c("PIMP13", "MED13", "TP6013")] = c(NA,NA,NA)
11
12  ponderation = table[conditionL, "P13_POP"]/sum(na.omit(table[conditionL, "P13_POP"]))
13
14  meanPIMP13 = ponderation * table[conditionL, "PIMP13"]
15  department[i, "PIMP13"] = sum(na.omit(meanPIMP13))
16 }
```

```

16
17 meanMED13 = ponderation * table[conditionL,"MED13"]
18 department[i,"MED13"] = sum(na.omit(meanMED13))
19
20 meanTP6013 = ponderation * table[conditionL,"TP6013"]
21 department[i,"TP6013"] = sum(na.omit(meanTP6013))
22
23 }
24 rm(conditionL, i, j, meanMED13, meanPIMP13, meanTP6013, ponderation, variables)
25 colnames(department)[3] <- "SUPERF_TOT"

```

Enfin, cette dernière routine de traitement ajoute trois variables à la data frame : la superficie moyenne des communes du département, la densité moyenne d'habitant par département et la croissance algébrique de population entre 2008 et 2013 dans chaque département.

```

1 #Add new variables to the department dataframe
2 department$SUPERF_MOY <- NA
3 department$DENSITY13 <- NA
4 department$POPO813 <- NA
5 for(i in rownames(department))
6 {
7   conditionL = (table[, "DEP"] == i)
8   department[i, "SUPERF_MOY"] = mean(na.omit(table[conditionL, "SUPERF"]))
9   department[i, "DENSITY13"] = department[i, "P13_POP"] / department[i, "SUPERF_TOT"]
10  department[i, "POPO813"] = department[i, "P13_POP"] - department[i, "P08_POP"]
11 }
12 rm(conditionL)

```

Analyse en composantes principales

Une analyse en composantes principales est faite sur le tableau de données des départements. On pourrait envisager de pondérer les départements selon le nombre d'habitants, mais pour le moment, on les pondère identiquement. On ne considère pas toutes les variables dans l'ACP puisque certaines sont des combinaisons linéaires des autres. De telles contraintes linéaires réduiraient le rang de la matrice d'inertie du tableau de données. La routine suivante permet de réaliser l'ACP en utilisant la librairie FACTOMINEE :

```

1 library(FactoMineR)
2 variablesPCA = !(colnames(department) %in% c("ETTOT14", "P08_POP", "P13_LOG", "P08_EMPLT"))
3 departmentsPCA = !(rownames(department) %in% c("971", "972", "973", "974", "75"))
4 variablesSupPCA = which(colnames(department) == c())
5 departmentsSupPCA = which(rownames(department) == c("75"))
6 ncp = 5
7 res.pca = PCA(department[departmentsPCA, variablesPCA],
8               scale.unit=TRUE,
9               ncp,
10              quanti.sup=NULL,
11              quali.sup=NULL,
12              ind.sup=departmentsSupPCA,
13              graph=F
14              )
15 rm(departmentsPCA, variablesPCA)

```

Dans le bloc précédent, les vecteur *variablesPCA* et *departmentsPCA* permettent de sélectionner les variables et les départements sur lesquels porteront l'ACP, sachant qu'on élimine celles et ceux qui figurent explicitement dans les vecteurs qui leur servent de définition. On élimine ainsi les variables *ETTOT14* et *P13_LOG* car ce sont des combinaisons linéaires d'autres variables, et on élimine les variables donnant des statistiques de 2008 pour ne garder que celles postérieures à 2013. On élimine également les départements 971, 972, 973 et 974 car il y manque trop de données. Enfin, le département 75 sera placé en tant que supplémentaire car il déforme trop les axes principaux.

Analyse des valeurs propres de la matrice d'inertie

Les résultats concernant les valeurs propres et les axes principaux sont regroupés dans le tableau 1 :

	Eigenvalue	Percentage of variance	Cumulative percentage of variance
comp 1	2.279811e+01	7.354228e+01	73.54228
comp 2	2.730124e+00	8.806852e+00	82.34913
comp 3	1.503328e+00	4.849446e+00	87.19857
comp 4	1.288055e+00	4.155016e+00	91.35359
comp 5	6.109021e-01	1.970652e+00	93.32424
comp 6	5.587377e-01	1.802380e+00	95.12662
comp 7	4.859642e-01	1.567626e+00	96.69425
comp 8	3.343937e-01	1.078689e+00	97.77294
comp 9	2.050343e-01	6.614008e-01	98.43434
comp 10	1.867114e-01	6.022948e-01	99.03663
comp 11	9.035974e-02	2.914830e-01	99.32811
comp 12	6.138620e-02	1.980200e-01	99.52613
comp 13	5.032867e-02	1.623506e-01	99.68849
comp 14	3.021366e-02	9.746341e-02	99.78595
comp 15	2.224731e-02	7.176553e-02	99.85771
comp 16	1.687326e-02	5.442986e-02	99.91214
comp 17	8.432002e-03	2.720001e-02	99.93934
comp 18	5.949301e-03	1.919129e-02	99.95854
comp 19	3.739072e-03	1.206152e-02	99.97060
comp 20	3.562860e-03	1.149310e-02	99.98209
comp 21	1.478398e-03	4.769026e-03	99.98686
comp 22	1.404677e-03	4.531217e-03	99.99139
comp 23	1.078105e-03	3.477759e-03	99.99487
comp 24	7.184952e-04	2.317727e-03	99.99719
comp 25	3.541949e-04	1.142564e-03	99.99833
comp 26	2.833242e-04	9.139492e-04	99.99924
comp 27	1.368280e-04	4.413807e-04	99.99968
comp 28	5.273217e-05	1.701038e-04	99.99985
comp 29	3.213915e-05	1.036747e-04	99.99996
comp 30	1.319754e-05	4.257272e-05	100.00000
comp 31	2.138784e-31	6.899303e-31	100.00000

Table 1: Valeurs propres et inertie des axes principaux

On ne garde que les axes dont l'inertie est supérieure à 3,23%, étant donné que l'inertie totale est partagée entre 31 axes principaux. On doit donc garder les 4 premiers axes d'inertie. En cumulé, ces 4 axes expliquent 93,3% de l'inertie totale du nuage.

Interprétation des variables actives

Intéressons-nous aux variables actives de l'ACP et à leur interprétation sur les 4 premiers axes principaux d'inertie. On se place pour cela dans le nuage des variables et décompose spectralement sa matrice d'inertie pour trouver les mêmes vecteurs propres que ceux de la matrice d'inertie du nuage des départements. Les tableaux suivants récapitulent les coordonnées, les contributions et les cosinus carrés des 4 axes principaux. Le bloc de code suivant donne les instructions nécessaires pour construire les tableaux qui suivent car, par défaut dans FactoMineR, les tableaux de résultats sont donnés par grandeur et non par axes, ce qui ne facilite pas l'interprétation "visuelle".

```
1 axes = list()
2 for(i in 1:npc)
```

```

3 {
4   columns = c(res.pca$var$coord[,i],res.pca$var$contrib[,i],res.pca$var$cos2[,i])
5   axe = matrix(data = columns,nrow = dim(res.pca$var$coord)[1],ncol = 3)
6   rownames(axe) = rownames(res.pca$var$coord)
7   colnames(axe) = c("Coord","Contr","Cos2")
8   axes[[i]] = axe
9 }

```

Les figure suivantes donnent également le positionnement des variables dans différents plans principaux. On utilise le bout de code suivant pour plotter les graphes qui nous intéressent, représentés par la figure 1, 2 et 3 :

```

1 #Plot of the factor and individual maps
2 plot(res.pca, axes = c(1,2), choix = "ind")
3 plot(res.pca, axes = c(1,2), choix = "var")
4 plot(res.pca, axes = c(2,3), choix = "ind")
5 plot(res.pca, axes = c(2,3), choix = "var")
6 plot(res.pca, axes = c(3,4), choix = "ind")
7 plot(res.pca, axes = c(3,4), choix = "var")

```

Premier axe principal d'inertie

	Coord	Contr	Cos2
P13_POP	0.9913531	4.31080078	0.98278090
SUPERF_TOT	-0.2377461	0.24792935	0.05652320
NAIS0813	0.9687385	4.11636987	0.93845434
DECE0813	0.9387379	3.86536002	0.88122885
P13_MEN	0.9948958	4.34166600	0.98981759
NAISD15	0.9672305	4.10356374	0.93553478
DECESD15	0.9426134	3.89734127	0.88851997
P13_RP	0.9948958	4.34166600	0.98981759
P13_RSECOCC	0.2242953	0.22066911	0.05030838
P13_LOGVAC	0.9295734	3.79025707	0.86410680
P13_RP_PROP	0.9686142	4.11531356	0.93821352
NBMENFISC13	0.9952413	4.34468188	0.99050515
PIMP13	0.8165408	2.92453654	0.66673892
MED13	0.5097700	1.13985557	0.25986547
TP6013	0.5903659	1.52877584	0.34853193
P13_EMPLT	0.9807853	4.21938480	0.96193979
P13_EMPLT_SAL	0.9754006	4.17318119	0.95140624
P13_POP1564	0.9896083	4.29564006	0.97932454
P13_CHOM1564	0.9562516	4.01093471	0.91441712
P13_ACT1564	0.9886610	4.28742032	0.97745060
ETAZ14	0.1176312	0.06069406	0.01383710
ETBE14	0.9310414	3.80223736	0.86683808
ETFZ14	0.9264996	3.76523207	0.85840157
ETGU14	0.9724019	4.14756176	0.94556550
ETGZ14	0.9783586	4.19853110	0.95718554
ETOQ14	0.9638253	4.07472117	0.92895922
ETTEF114	0.9850496	4.25615519	0.97032274
ETTEFP1014	0.9869798	4.27285125	0.97412913
SUPERF_MOY	-0.1594272	0.11148749	0.02541704
DENSITY13	0.4184817	0.76816459	0.17512697
POP0813	0.7189136	2.26701627	0.51683676

Table 2: Axe principal d'inertie 1

Sur le premier axe d'inertie, on garde toutes les variables dont les contributions sont supérieures à 2,23%, c'est-à-dire : "P13_POP", "NAIS0813", "DECE0813", "P13_MEN", "NAISD15", "DECESD15", "P13_RP", "P13_LOGVAC", "P13_RP_PROP", "NBMEFISC13", "P13_EMPLT", "P13_EMPLT_SAL", "P13_POP1564", "P13_CHOM1564", "P13_ACT1564", "ETBE14", "ETFZ14", "ETGU14", "ETGZ14", "ETOQ14", "ETTEF114", "ETTEFP1014". On peut utiliser la commande suivante pour obtenir directement les labels des variables qui répondent au critère précédent :

```
1 labels(which(axes[[1]][,2]>=3,23))
```

Toutes ces variables ont une coordonnée positive sur l'axe. Ainsi, du côté positif de l'axe se trouve les départements qui ont à la fois :

- Une population importante et donc beaucoup de ménages en 2013 ;
- Un taux de naissance important entre 2008 et 2013 ;
- Un bon nombre de naissances et de décès domiciliés en 2015 ;
- Beaucoup de résidences principales, de propriétaires, de logements vacants ;
- Une forte part de ménages imposables fiscalement ;
- Beaucoup d'emplois au lieu de travail en 2013, et beaucoup d'emplois salariés ;
- Beaucoup de chômeurs, d'actifs entre 15 et 64 ans en 2013 ;
- Beaucoup d'établissements actifs dans dans l'industrie, la construction, les commerces services, dans la réparation automobile et l'administration publique.

Toutes les variables ci-dessus sont très fortement corrélées entre elles.

Second axe principal d'inertie

Sur le premier axe d'inertie, on garde toutes les variables dont les contributions sont supérieures à 2,23%, c'est-à-dire (en utilisant une commande similaire à la précédente) : "SUPERF_TOT", "P13_RSECOCC", "TP6013", "ETAZ14", "ETBE14", "SUPERF_MOY", "DENSITY13". Les coordonnées de ces variables ne sont pas toutes positives : du côté positif, on trouve "SUPERF_TOT", "SUPERF_MOY", "ETAZ14", "ETBE14" et "P13_RSECOCC" ; du côté négatif : "TP6013", "DENSITY13". Cet axe discrimine donc les départements :

- **Du côté positif** : les départements avec une forte superficie urbaine et une forte superficie moyenne des communes, avec beaucoup d'établissements actifs dans l'agriculture et dans l'industrie et de résidences secondaires.
- **Du côté négatif** : les départements avec un fort taux de pauvreté moyen par commune et une forte densité de populations.

Troisième axe principal d'inertie

Quatrième axe principal d'inertie

La contribution d'une variable sur un axe représente la part d'inertie qu'elle explique le long de cet axe. Sur chaque axe, on s'intéresse aux variables qui explique plus que 2,23% de l'inertie de l'axe, et on regarde leur coordonnées sur l'axe.

	Coord	Contr	Cos2
P13_POP	-0.0140925763	7.274420e-03	1.986007e-04
SUPERF_TOT	0.8523197906	2.660864e+01	7.264490e-01
NAIS0813	-0.1631378583	9.748260e-01	2.661396e-02
DECE0813	0.2172288345	1.728433e+00	4.718837e-02
P13_MEN	0.0340942424	4.257746e-02	1.162417e-03
NAISD15	-0.1805238711	1.193677e+00	3.258887e-02
DECESD15	0.2323164627	1.976868e+00	5.397094e-02
P13_RP	0.0340942424	4.257746e-02	1.162417e-03
P13_RSECOCC	0.4424433363	7.170227e+00	1.957561e-01
P13_LOGVAC	0.2303390060	1.943357e+00	5.305606e-02
P13_RP_PROP	0.1488746771	8.118191e-01	2.216367e-02
NBMENFISC13	0.0349240089	4.467513e-02	1.219686e-03
PIMP13	-0.2180643071	1.741754e+00	4.755204e-02
MED13	-0.2396784399	2.104144e+00	5.744575e-02
TP6013	-0.2818567855	2.909877e+00	7.944325e-02
P13_EMPLT	-0.0519661642	9.891427e-02	2.700482e-03
P13_EMPLT_SAL	-0.0799449855	2.340993e-01	6.391201e-03
P13_POP1564	-0.0385789361	5.451526e-02	1.488334e-03
P13_CHOM1564	-0.0402995387	5.948641e-02	1.624053e-03
P13_ACT1564	-0.0668864047	1.638677e-01	4.473791e-03
ETAZ14	0.7593267977	2.111908e+01	5.765772e-01
ETBE14	0.2561821891	2.403895e+00	6.562931e-02
ETFZ14	0.0428971302	6.740220e-02	1.840164e-03
ETGU14	-0.0194508955	1.385788e-02	3.783373e-04
ETGZ14	0.0584982231	1.253438e-01	3.422042e-03
ETOQ14	0.1404474736	7.225127e-01	1.972549e-02
ETTEF114	0.0854413136	2.673951e-01	7.300218e-03
ETTEFP1014	-0.0003879175	5.511837e-06	1.504800e-07
SUPERF_MOY	0.4905596780	8.814574e+00	2.406488e-01
DENSITY13	-0.6701679927	1.645072e+01	4.491251e-01
POP0813	0.0531813167	1.035943e-01	2.828252e-03

Table 3: Axe principal d'inertie 2

	Coord	Contr	Cos2
P13_POP	-0.09511364	0.60177178	0.0090466054
SUPERF_TOT	-0.22029080	3.22803979	0.0485280355
NAIS0813	-0.11092386	0.81845753	0.0123041036
DECE0813	-0.12252645	0.99863294	0.0150127315
P13_MEN	-0.07396723	0.36393591	0.0054711515
NAISD15	-0.08362054	0.46512758	0.0069923945
DECESD15	-0.10403257	0.71992101	0.0108227763
P13_RP	-0.07396723	0.36393591	0.0054711515
P13_RSECOCC	0.70629524	33.18323489	0.4988529593
P13_LOGVAC	-0.02390789	0.03802144	0.0005715871
P13_RP_PROP	-0.09222891	0.56582264	0.0085061718
NBMENFISC13	-0.05996994	0.23922873	0.0035963932
PIMP13	0.34856233	8.08178061	0.1214956946
MED13	0.16424625	1.79447365	0.0269768301
TP6013	0.11903939	0.94260029	0.0141703769
P13_EMPLT	-0.06854568	0.31254057	0.0046985109
P13_EMPLT_SAL	-0.08905560	0.52755606	0.0079308996
P13_POP1564	-0.10381249	0.71687817	0.0107770324
P13_CHOM1564	-0.14765521	1.45025276	0.0218020601
P13_ACT1564	-0.09033389	0.54280967	0.0081602114
ETAZ14	-0.34877148	8.09148241	0.1216415445
ETBE14	0.06418266	0.27401956	0.0041194137
ETFZ14	0.23999996	3.83149717	0.0575999811
ETGU14	0.15935437	1.68917303	0.0253938162
ETGZ14	0.09364392	0.58331792	0.0087691833
ETOQ14	0.08313260	0.45971528	0.0069110299
ETTEF114	0.06470107	0.27846402	0.0041862284
ETTEFP1014	-0.05984859	0.23826160	0.0035818541
SUPERF_MOY	0.61901466	25.48872051	0.3831791474
DENSITY13	0.08304477	0.45874440	0.0068964344
POP0813	0.19965466	2.65158214	0.0398619846

Table 4: Axe principal d'inertie 3

	Coord	Contr	Cos2
P13_POP	-0.0073254584	4.166153e-03	5.366234e-05
SUPERF_TOT	0.1204175679	1.125759e+00	1.450039e-02
NAIS0813	-0.0335548800	8.741320e-02	1.125930e-03
DECE0813	-0.1281775209	1.275526e+00	1.642948e-02
P13_MEN	0.0020843946	3.373071e-04	4.344701e-06
NAISD15	-0.0365793423	1.038813e-01	1.338048e-03
DECESD15	-0.1154520456	1.034830e+00	1.332917e-02
P13_RP	0.0020843946	3.373071e-04	4.344701e-06
P13_RSECOCC	-0.0639098580	3.171037e-01	4.084470e-03
P13_LOGVAC	-0.0727041852	4.103784e-01	5.285899e-03
P13_RP_PROP	0.0565227731	2.480348e-01	3.194824e-03
NBMENFISC13	0.0001345995	1.406542e-06	1.811704e-08
PIMP13	0.2005831755	3.123594e+00	4.023361e-02
MED13	0.7596473996	4.480121e+01	5.770642e-01
TP6013	-0.6504520739	3.284704e+01	4.230879e-01
P13_EMPLT	0.0903126228	6.332316e-01	8.156370e-03
P13_EMPLT_SAL	0.0937305208	6.820680e-01	8.785411e-03
P13_POP1564	0.0018863425	2.762528e-04	3.558288e-06
P13_CHOM1564	-0.2095500332	3.409111e+00	4.391122e-02
P13_ACT1564	0.0434165362	1.463444e-01	1.884996e-03
ETAZ14	0.0974455763	7.372078e-01	9.495640e-03
ETBE14	0.0173184701	2.328545e-02	2.999294e-04
ETFZ14	-0.1253619355	1.220104e+00	1.571561e-02
ETGU14	0.0335623629	8.745219e-02	1.126432e-03
ETGZ14	-0.0937251889	6.819904e-01	8.784411e-03
ETOQ14	0.0241914273	4.543480e-02	5.852252e-04
ETTEF114	0.0088601645	6.094656e-03	7.850252e-05
ETTEFP1014	0.0709252602	3.905418e-01	5.030393e-03
SUPERF_MOY	-0.1198493588	1.115160e+00	1.436387e-02
DENSITY13	0.0012287000	1.172080e-04	1.509704e-06
POP0813	0.2647555620	5.441966e+00	7.009551e-02

Table 5: Axe principal d'inertie 4

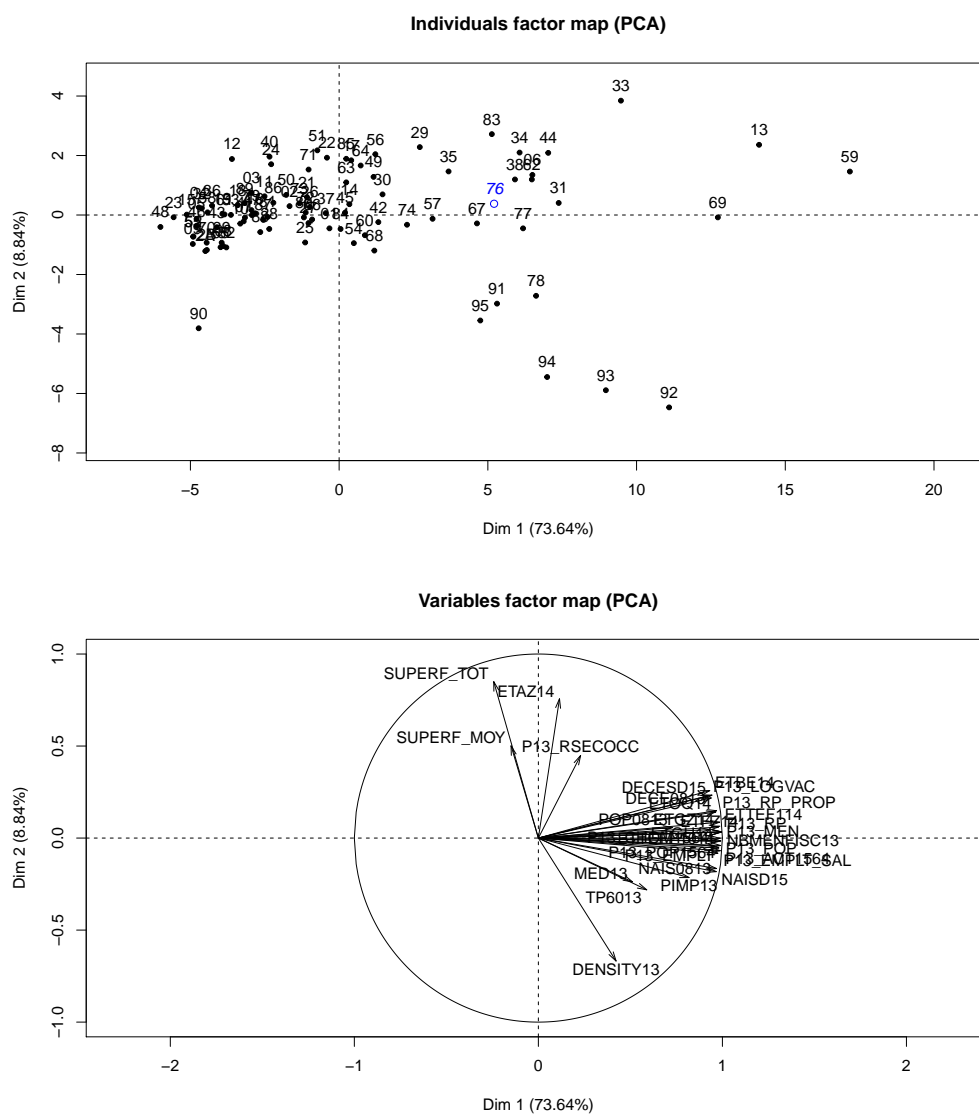


Figure 1: Plans factoriels principaux

