

# ACP sur les ODDs

*NIANG & SAWADO*

2022-11-27

## 2ème PARTIE :

**Analyse en composantes principales du taux de pauvreté total par département et de 10 autres variables choisis dans les objectifs de développement durable (ODD).**

### Introduction :

Les dernières décennies ont vu un changement dans la définition et la compréhension du progrès social avec des avancées dans les domaines de l'économie et de la technologie.

Par conséquent, les indicateurs (nos variables ici) ont aussi évolué.

Le taux de pauvreté est la proportion d'individus appartenant à des ménages dont le niveau de vie, c'est-à-dire le revenu disponible (après transferts, impôts et prestations sociales) par unité de consommation (UC) est inférieur au seuil de 60 % de la médiane du revenu disponible par UC de l'ensemble de la population. En 2017, ce seuil est de 1 041 euros par mois en France.

Il met en relief en particulier le bien-être économique mais ignore les inégalités, la faible satisfaction sociale et la dégradation de l'environnement.

C'est en cela que les autres indicateurs sont précieux pour estimer la gain en bien-être humain, de justice sociale et de progrès écologique.

Dans le cadre de notre étude nous choisirons 10 variables au-delà du taux de pauvreté total et nous utiliserons l'analyse en composantes principales pour rechercher leur corrélation avec le taux de pauvreté total.

```
library(tidyverse)
```

```
## Warning: le package 'tidyverse' a été compilé avec la version R 4.1.3
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: le package 'ggplot2' a été compilé avec la version R 4.1.3
```

```
## Warning: le package 'tibble' a été compilé avec la version R 4.1.3
```

```
## Warning: le package 'tidyr' a été compilé avec la version R 4.1.3
```

```
## Warning: le package 'readr' a été compilé avec la version R 4.1.3

## Warning: le package 'purrr' a été compilé avec la version R 4.1.3

## Warning: le package 'dplyr' a été compilé avec la version R 4.1.3

## Warning: le package 'stringr' a été compilé avec la version R 4.1.3

## Warning: le package 'forcats' a été compilé avec la version R 4.1.3

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

## 1 . Etude statistique :

Nous devons choisir nos 11 variables en ayant en tete que les objectifs sont : Eradition de la pauvreté et Infrastructures résilientes.

Auparavant, sur les 101 départements retenus, les départements d'Outre-mer ( Guadeloupe, Martinique, La Réunion, Guyane, Mayotte) se caractérisent par une énorme quantité de données manquantes. Ce qui témoigne de la faiblesse générale du système de collecte de donnée dans ces zones en particulier à Mayotte ou les données sont inexistantes. Nous avons choisi de l'exclure de l'étude parce n'apporte aucune information.

- *Remarque* : On peut d'ores et déjà remarquer que la qualité de l'appareil statistique d'un département semble liée à son niveau de développement.

Nous travaillerons donc avec 100 departements dont 4 d'Outre-mer.

Pour les données manquantes dans les départements d'Outre-mer, nous avons fait le choix de remplacer par la moyenne ( toutes les données étant quantitatives ).

Ensuite dans notre choix de variables, nous avons privilégié le Taux de pauvreté total , chez les jeunes (<30 ans) et chez les seniors (60-74 ans) et  $\geq 75$  , l'esperance de vie des hommes/des femmes, la part de voitures criticaid essence, nombre de population hlm, part de foyers fiscaux imposables, apl\_medgen 65, revenu decl median.

```
data_fin = select(data,-c(4,5,6,9,13,14,15,17)) # Selection des 11 colonnes
```

```
data_fin = data_fin[1:100,] #selection de tous les départements sauf Mayotte.
head(data_fin)
```

```
##              X Taux.de.pauvrete.total Taux.de.pauvrete...30.
## 1              Ain                    10.3                  14.9
## 2              Aisne                   18.4                  29.7
## 3              Allier                   15.5                  25.3
## 4 Alpes-de-Haute-Provence              16.8                  24.9
## 5              Hautes-Alpes            13.9                  19.9
## 6              Alpes-Maritimes         15.8                  21.7
##  Taux.de.pauvrete.60..74 Taux.de.pauvrete...75 PDDT.tec
## 1                    7.5                    7.2      6.24
## 2                    11.6                    9.5      5.26
## 3                    10.6                   10.2      2.91
## 4                    12.3                   11.3      2.81
## 5                     9.6                   10.2      3.93
## 6                    13.2                   11.9     13.57
##  esperance.de.vie.homme X.part.des.vehicules.criticair.1.essence
## 1                    80.7                    60055
## 2                    76.9                    36821
## 3                    78.2                    25209
## 4                    79.4                    13438
## 5                    80.5                    11636
## 6                    80.5                    136217
##  nombre.de.personnes.en.logement.social
## 1                    89353.24
## 2                    78869.44
## 3                    32756.86
## 4                    14330.12
```

```
## 5 15079.84
## 6 118838.75
## accessibilité.potentielle.au.medecins.generalistes
## 1 2.722141
## 2 2.940012
## 3 2.889753
## 4 3.651232
## 5 4.380195
## 6 3.905107
## revenu.fiscal.median.declare.par.unite esperance.de.vie.des.femme
## 1 23300 85.9
## 2 19050 83.4
## 3 19690 84.8
## 4 20190 84.8
## 5 20640 86.4
## 6 22390 85.8
```

```
for (i in 1:ncol(data_fin)) {
  data_fin[is.na(data_fin[,i]), i] = mean(data_fin[,i], na.rm = T)
}
```

```
## Warning in mean.default(data_fin[, i], na.rm = T): l'argument n'est ni
## numérique, ni logique : renvoi de NA
```

```
kable(summary(data_fin))
```

X	Taux.de.fraude.pourcent	Taux.de.patients.pourcent	Taux.de.medecins.pourcent	Taux.de.medecins.generalistes.pourcent	esperance.de.vie.des.femme	revenu.fiscal.median.declare.par.unite	accessibilite.potentielle.au.medecins.generalistes
Length	Min.	Min.	Min.	Min.	Min.	Min.	Min.
:	:12.00	:	:	:	:76.10	4281	5123
8.90		6.30	5.30	1.220			
Class 1st	1st	1st	1st	1st	1st	1st Qu.:	1st Qu.:
:char- Qu.:	12.47	19.60	Qu.:	Qu.:	Qu.:	78.38	726
ac-		8.40	8.00	2.605			
ter							
Mode	Median	Median	Median	Median	Median	Median	Median
:char- :	14.50	:21.90	:	:	:79.40	42442	: 64715
ac-			9.85	9.70	4.710		
ter							
NA	Mean	Mean	Mean	Mean	Mean	Mean	Mean
:	:14.93	:22.92	:10.59	:10.66	:79.28	54584	:101730
NA	3rd	3rd	3rd	3rd	3rd	3rd Qu.:	3rd
Qu.:	15.88	24.93	11.65	11.95	Qu.:	80.00	1615
NA	Max.	Max.	Max.	Max.	Max.	Max.	Max.
:	:38.90	:54.30	:31.00	:39.90	:64.31	682.40	:196912

\* Pour aller plus vite, nous appellerons par moment TPT pour le taux de pauvreté total.

L'espérance de vie moyen d'un homme est plus faible que celle d'une femme.

Le revenu fiscal médian déclaré par unité de consommation est 200650 euros environ.

La part moyenne de voitures critcaira 1 ( moins polluantes) par département est de 54584.

```
library(knitr)
R = cor(data_fin)
kable(R)
```

[illegible]

On a aussi une forte corrélation entre PDDT et nombre de personnes en logement social.

Les espérances de vie entre sexes sont très corrélées. Plus l'espérance de vie des hommes augmentent plus celle des femmes augmente aussi.

On peut remarquer dans une moindre mesure une corrélation négative entre le revenu.fiscal.median.declaré.par.unité et le taux de pauvreté dans les différentes tranches d'âge.

En d'autres termes plus le revenu median augmente, moins il y'a de pauvreté.

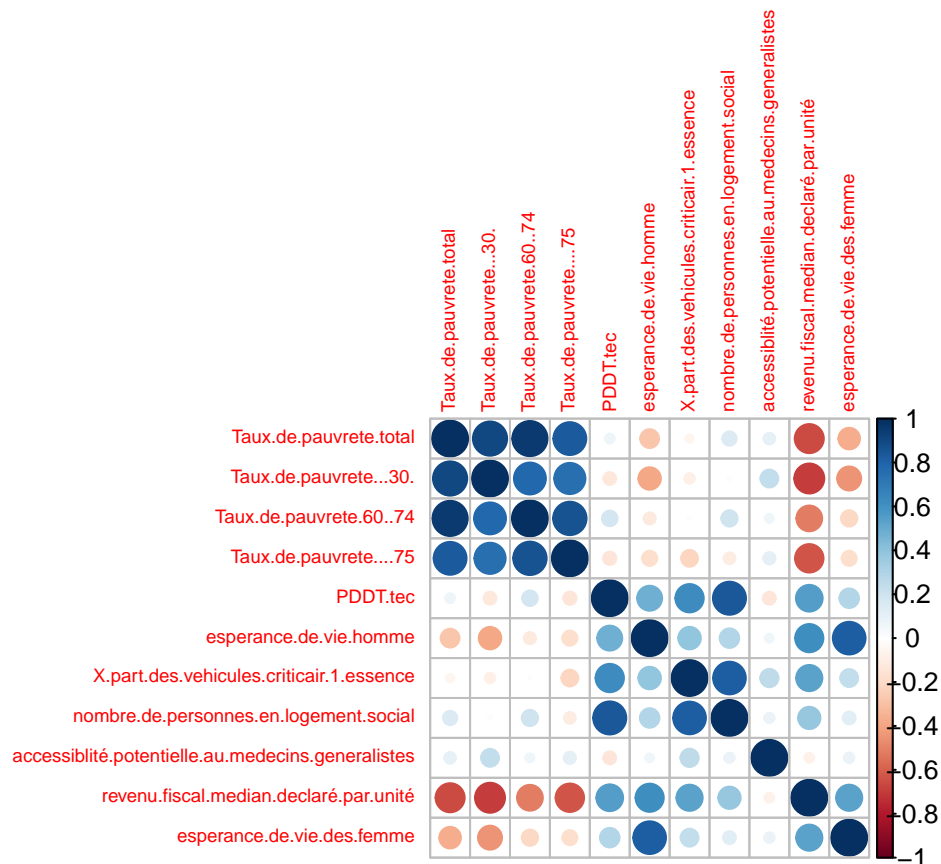
**Remarque** Nous allons dans ce dresser un corrplot pour mieux s'en rendre compte

```
#library(zoom)
library(corrplot)
```

```
## Warning: le package 'corrplot' a été compilé avec la version R 4.1.3
```

```
## corrplot 0.92 loaded
```

```
corrplot(R, tl.cex=0.6)
```



```
library("FactoMineR") #Importation de la library FactoMiner
```

2. Début de notre ACP :

```
## Warning: le package 'FactoMineR' a été compilé avec la version R 4.1.3
```

```
data.pca=PCA(data_fin, graph = FALSE)
data.pca
```

```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 100 individuals, described by 11 variables
## *The results are available in the following objects:
##
##   name                description
## 1  "$eig"              "eigenvalues"
## 2  "$var"              "results for the variables"
## 3  "$var$coord"        "coord. for the variables"
## 4  "$var$cor"          "correlations variables - dimensions"
## 5  "$var$cos2"         "cos2 for the variables"
## 6  "$var$contrib"      "contributions of the variables"
## 7  "$ind"              "results for the individuals"
## 8  "$ind$coord"        "coord. for the individuals"
## 9  "$ind$cos2"         "cos2 for the individuals"
## 10 "$ind$contrib"      "contributions of the individuals"
## 11 "$call"             "summary statistics"
## 12 "$call$centre"      "mean of the variables"
## 13 "$call$ecart.type"  "standard error of the variables"
## 14 "$call$row.w"       "weights for the individuals"
## 15 "$call$col.w"       "weights for the variables"
```

```
library("factoextra")
```

Valeurs propres :

```
## Warning: le package 'factoextra' a été compilé avec la version R 4.1.3
```

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

```
val_propres =get_eigenvalue(data.pca)
kable(val_propres)
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.6988220	42.7165641	42.71656
Dim.2	2.9888943	27.1717667	69.88833
Dim.3	1.2977371	11.7976102	81.68594
Dim.4	1.1177651	10.1615006	91.84744
Dim.5	0.2563509	2.3304630	94.17790
Dim.6	0.2202078	2.0018886	96.17979
Dim.7	0.1834476	1.6677056	97.84750
Dim.8	0.1299388	1.1812621	99.02876
Dim.9	0.0686902	0.6244562	99.65322
Dim.10	0.0330467	0.3004244	99.95364
Dim.11	0.0050994	0.0463586	100.00000

- `eigen$value` : valeurs propres ou Inertie
- `percentage of variance` : pourcentage d'inertie ou part d'inertie ou proportion de variance expliquée par chaque valeur propre
- `cumulative percentage of variance` : Inertie cumulée

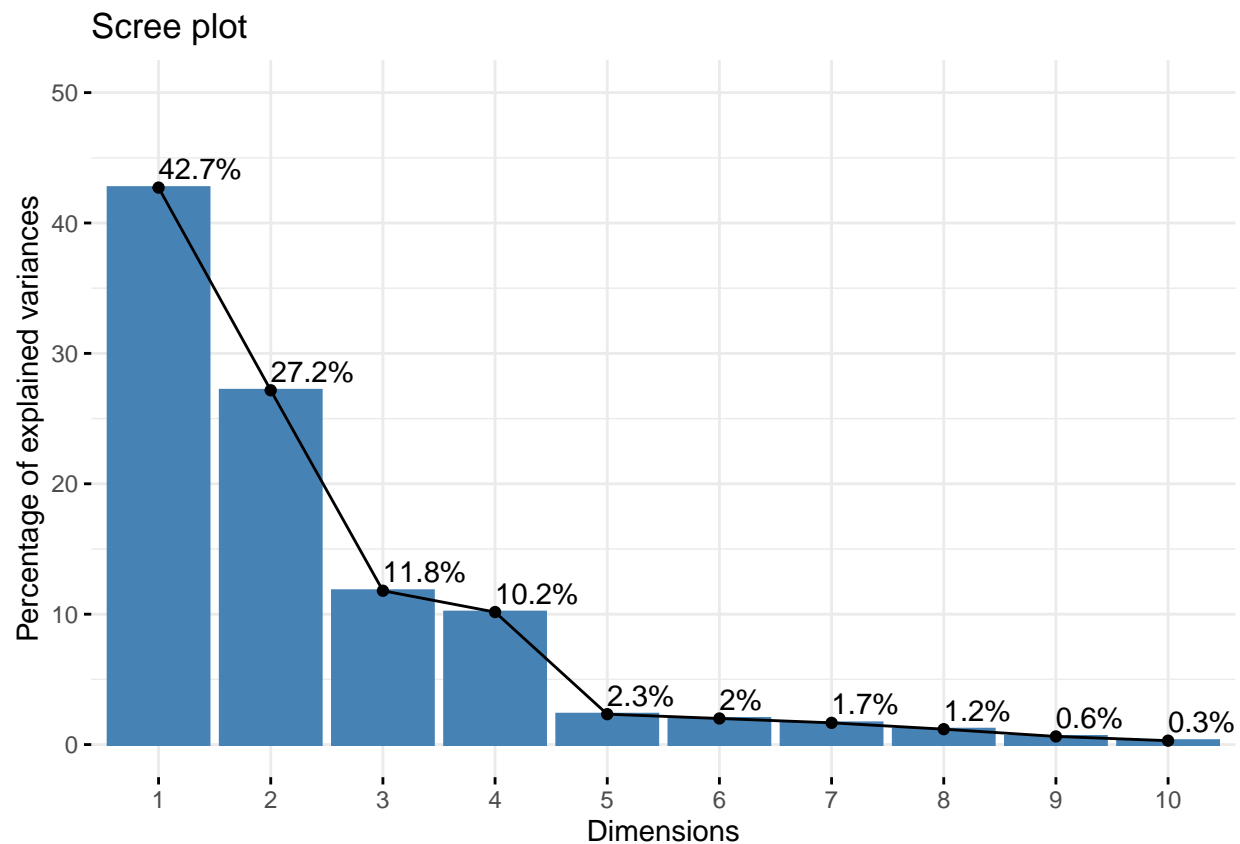
Les trois premières composantes principales cumulent  $\approx 82\%$  de l'information du tableau de données .

Mais dans le cadre de notre étude, nous choisirons deux axes par la règle du coude:

*En effet, compte tenu du nombre important de variables, environ 70 % de l'information nous semble un chiffre acceptable pour continuer notre étude*

**Eboulli des valeurs propres : ( règle du coude)**

```
fviz_eig(data.pca, addlabels = TRUE, ylim = c(0, 50))
```



La représentation des variables diffère de celle des individus: les individus sont représentées par leurs projections, mais les variables sont représentées par leurs corrélations.

On peut mieux le voir grace ci-dessous tri en ordre décroissant :

```
#Coordonnées des variables
```

```
kable(data.pca$var$coord[,1:2])
```



	Dim.1	Dim.2
Taux.de.pauvrete.total	0.8385959	0.5198570
Taux.de.pauvrete. . . 30.	0.8658662	0.3545420
Taux.de.pauvrete.60..74	0.7343342	0.6030038
Taux.de.pauvrete. . . 75	0.8047496	0.3324917
PDDT.tec	-0.4198895	0.7832763
esperance.de.vie.homme	-0.6272612	0.4086562
X.part.des.vehicules.criticair.1.essence	-0.4516081	0.7109027
nombre.de.personnes.en.logement.social	-0.3037669	0.8391960
accessibilité.potentielle.au.medecins.generalistes	0.0907374	0.2026389
revenu.fiscal.median.declaré.par.unité	-0.9028965	0.1927215
esperance.de.vie.des.femme	-0.6079003	0.2245451

- Nous remarquons que la variable revenu.fiscal.median.declaré.par.unité (cor ≈ -0.9) est la plus corrélée ( négativement) avec l'axe 1.
- L'esperance de vie chez les differents sexes dans une moindre mesure ( cor ≈ -0,6).
- Les variables taux de pauvreté totale, moins de 30 ans, 60-74 ans et plus de 75 ans sont tres corrélées avec l'axe 1.
- Les variables nombre.de.personnes.en.logement.social (0.8391960) , PDDT.tec (0.7832763) X.part.des.vehicules.criticair.1.essence (0.7109027) sont très corrélées avec l'axe 2.
- Remarque : Si on s'intéresse a la correlation des variables avec le 3ème axe, nous vérions que elles sont toutes très peu corrélées avec cet axe. C'est aussi une des raisons qui ont motivé, notre choix de 2 axes factoriels. (voir le ci-dessous)

```
kable(sort(data.pca$var$coord[,3],decreasing = T)) # correlation avec l'axe 3
```

	x
esperance.de.vie.des.femme	0.6974306
esperance.de.vie.homme	0.5781304
Taux.de.pauvrete. . . 75	0.3481762
accessibilité.potentielle.au.medecins.generalistes	0.2837926
Taux.de.pauvrete.60..74	0.1419217
Taux.de.pauvrete.total	0.0316177
Taux.de.pauvrete. . . 30.	0.0058214
revenu.fiscal.median.declaré.par.unité	-0.0303602
PDDT.tec	-0.2149284
X.part.des.vehicules.criticair.1.essence	-0.2454504
nombre.de.personnes.en.logement.social	-0.3831329

**Inertie totale :** On sait que

$$Inertie(Totale) = \sum I(projetées) = \sum_{i=1}^n \lambda_i$$

Nous allons sommer toutes les valeurs propres grace la fonction sum sur la colonne eigenvalue.

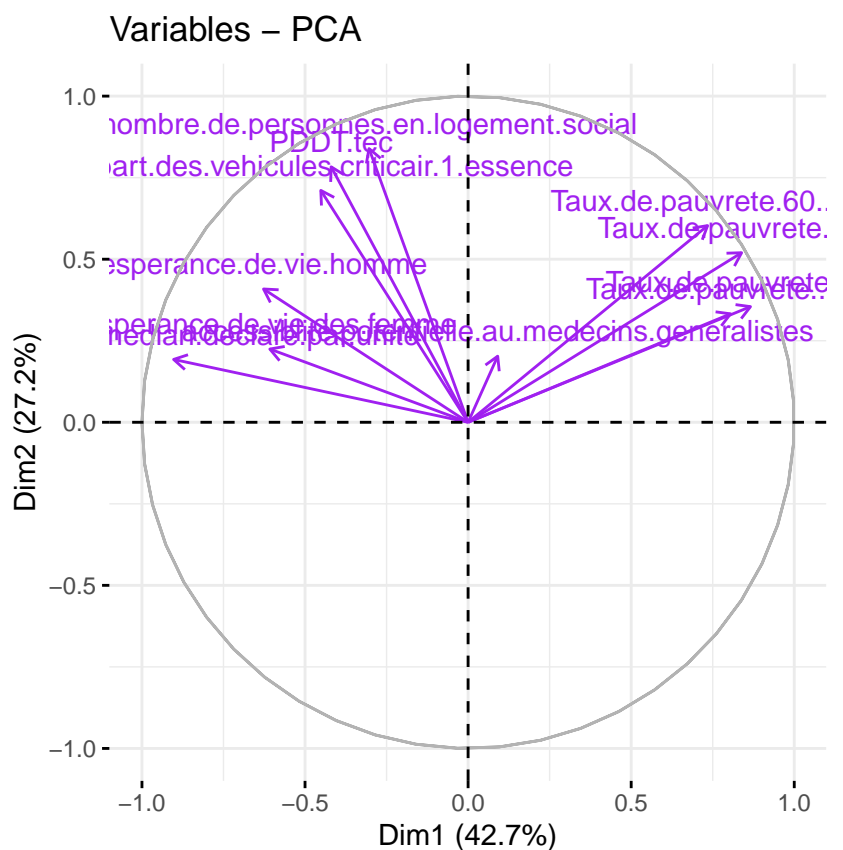
```
sum(val_propres[,1])
```

```
## [1] 11
```

```
#apply(val_propres,1,sum)
```

On pouvait s'y attendre, nous sommes dans une ACP normée

```
fviz_pca_var(data.pca, col.var = "purple")
```



### Cercle de corrélation :

Nous pouvons déjà remarquer deux catégories se distinguant :

- Du côté droit du graphe, les indicateurs de **l'objectif 1 : Eradication de la pauvreté**. L'indicateur 'accessibilité.potentielle.au.medecins.generalistes' de **l'objectif 11 : Villes et Communautés durables**, figure également dans cette partie, ce qui montre le lien étroit entre cet accès et le niveau de pauvreté dans un département, et donc l'importance de la fracture médicale entre les départements (certains pourvus de médecins généralistes et d'autres présentant des déserts médicaux).

Le revenu fiscal median par unité est la variable la plus corrélée avec l'axe 1 (négative).

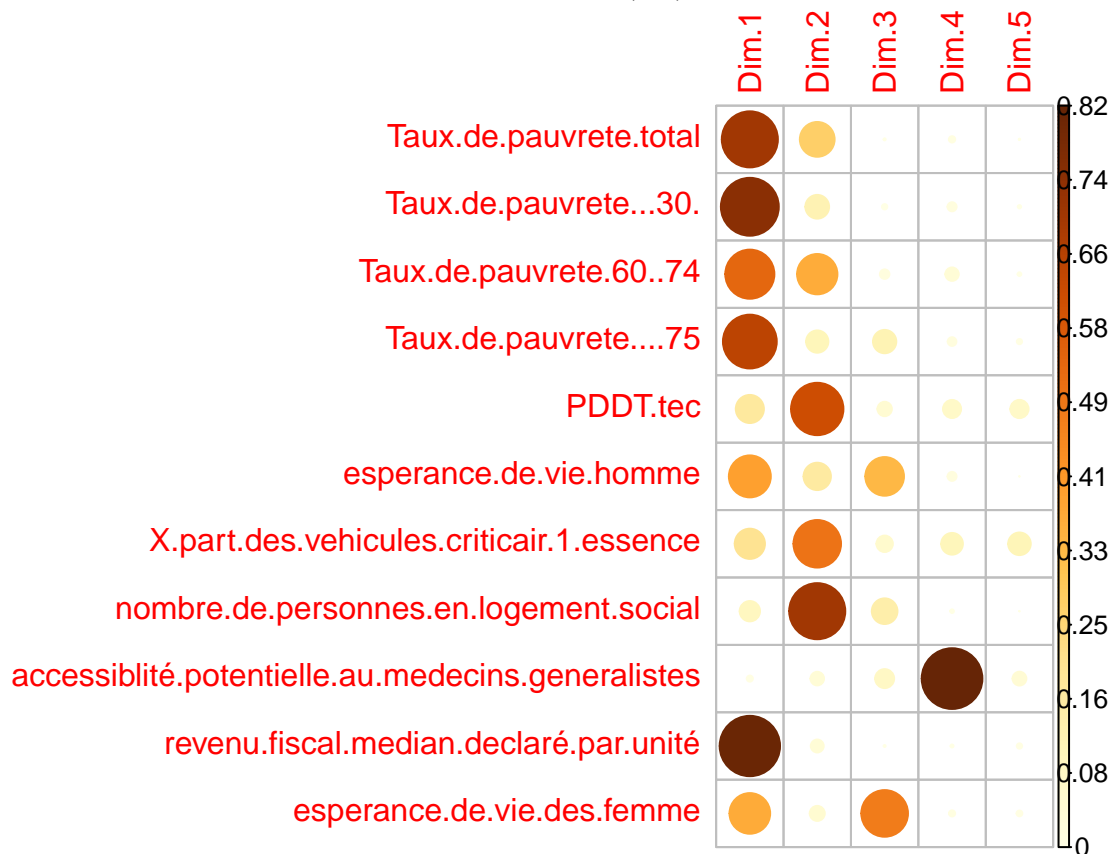
Etonnement, l'indicateur nombre de personnes en logement social de l'objectif 11 n'apparaît pas dans cette liste. On aurait pu s'attendre à un lien étroit entre le niveau de pauvreté et le nombre de personnes occupants

les logements sociaux. Du reste, il entretient une corrélation très forte avec le 2eme axe factoriel comme on a vu plus haut.

- Du côté gauche du graphe, nous avons un bloc constitué par la l'esperance de vie homme/femme , le revenu médian par unité de consommation, la part de déplacement en tec et la part de voiture criticaire1(essence). On peut remarquer une forte corrélation entre l'esperance de vie et le nombre de voitures moins polluantes et dans une moindre mesure la part de déplacement en transport en commun.
- Un deuxième enseignement ressort de l'analyse de cet axe. Comme attendu, il existe une corrélation négative entre les différents indicateurs de pauvreté, les indicateurs de revenu median par unité et d'esperance de vie (bien représentés sur l'axe 1 ). Cela signifie que l'amélioration de la situation pour le premier groupe ( indicateurs pauvreté) d'indicateurs dans un département va de pair avec l'amélioration pour le second groupe ( dont quelques indicateurs de l'objectif 9 Infrastructures résilientes et innovation) Autrement dit, lorsqu'on cherche à se rapprocher de l'objectif 1, on va aussi se rapprocher de facto de l'ODD 9.

```
library("corrplot")
corrplot(data.pca$var$cos2, is.corr=FALSE)
```

Qualités de représentation des variables sur l'axe(1,2) par corrplot et lecture de données



Les indicateurs de pauvreté sont bien représentés sur l'axe 1 de meme que le revenu fiscal médian.

Le nombre de personnes en logement social, le PDDT.tec et la part des vehicules criticaire1 sont bien représentés sur l'axe 2 .

Seule l'accessibilité potentielle au medecin generaliste a une faible qualité de représenté dans notre plan (axe1,axe2).

- **Qualités de représentation des variables sur l'axe(1,2) par lecture des données :**

```
kable(data.pca$var$cos2[,1:2])
```

	Dim.1	Dim.2
Taux.de.pauvrete.total	0.7032431	0.2702513
Taux.de.pauvrete. . . 30.	0.7497243	0.1257000
Taux.de.pauvrete.60..74	0.5392467	0.3636136
Taux.de.pauvrete. . . 75	0.6476219	0.1105507
PDDT.tec	0.1763072	0.6135217
esperance.de.vie.homme	0.3934566	0.1669999
X.part.des.vehicules.criticair.1.essence	0.2039498	0.5053826
nombre.de.personnes.en.logement.social	0.0922743	0.7042499
accessibilité.potentielle.au.medecins.generalistes	0.0082333	0.0410625
revenu.fiscal.median.declaré.par.unité	0.8152221	0.0371416
esperance.de.vie.des.femme	0.3695428	0.0504205

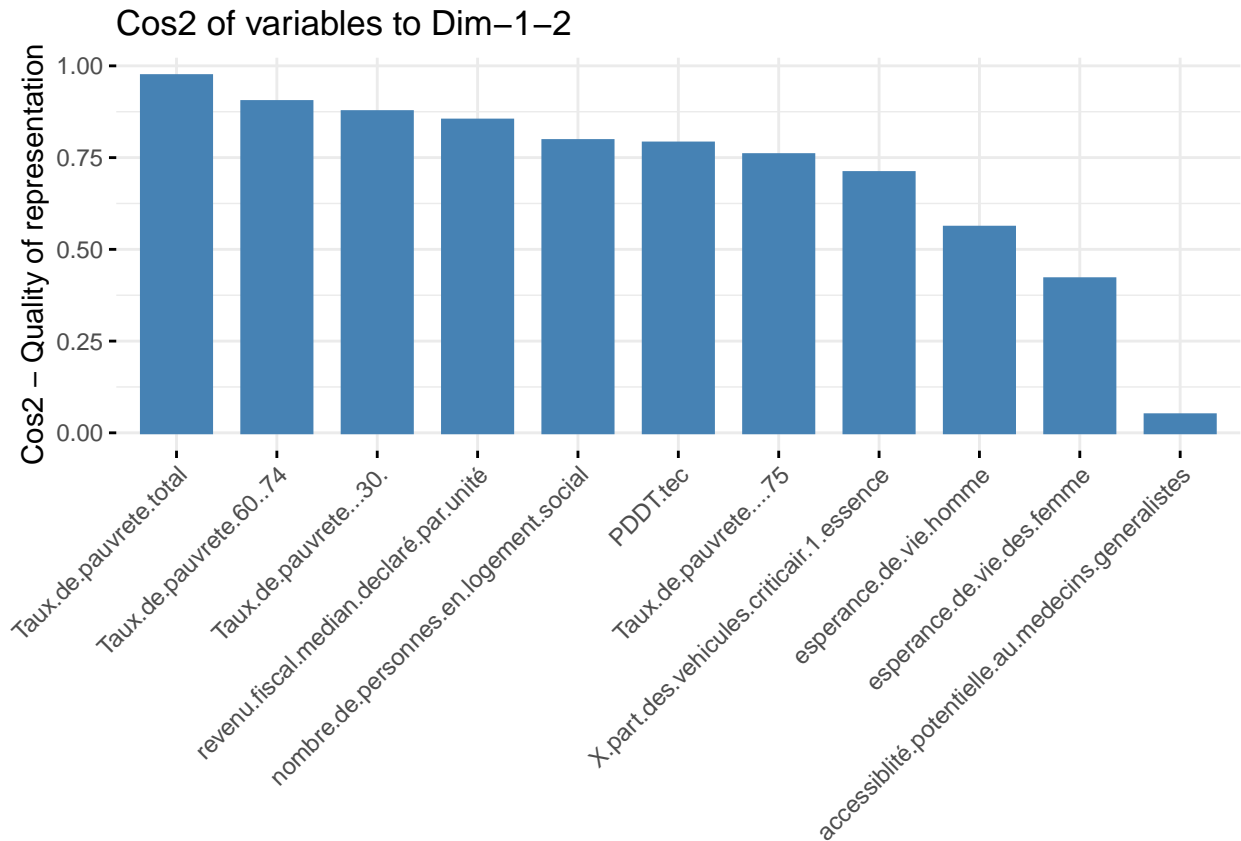
```
kable(data.pca$var$contrib[,1:2])
```

#### Contribution à la création des axes :

	Dim.1	Dim.2
Taux.de.pauvrete.total	14.9663702	9.041848
Taux.de.pauvrete. . . 30.	15.9555790	4.205569
Taux.de.pauvrete.60..74	11.4762096	12.165487
Taux.de.pauvrete. . . 75	13.7826447	3.698717
PDDT.tec	3.7521580	20.526711
esperance.de.vie.homme	8.3735154	5.587347
X.part.des.vehicules.criticair.1.essence	4.3404462	16.908682
nombre.de.personnes.en.logement.social	1.9637753	23.562220
accessibilité.potentielle.au.medecins.generalistes	0.1752201	1.373836
revenu.fiscal.median.declaré.par.unité	17.3494986	1.242653
esperance.de.vie.des.femme	7.8645829	1.686928

Les indicateurs de pauvreté ainsi que le revenu median par unité participent le plus à la création de l'axe 1 alors que PDDT.tec, esperance de vie homme, la part des vehicules criticair.1 essence et le nombre de personnes en logement social participent plus à la création de l'axe 2.

```
fviz_cos2(data.pca, choice='var', axes = 1:2)
```



Il est immédiat que la variable 'Taux de pauvreté total' est la mieux représentée sur le plan factoriel (1,2) et la variable 'accessibilité ptentielle aux médecins généralistes' est la moins bien représentée sur le plan factoriel (1,2).

#### Contributions des individus à la création des axes :

```
data.pca$ind$contrib[,1:2] #contribution des individus a la création des deux axes
```

##	Dim.1	Dim.2
## 1	1.390742877	0.203247726
## 2	1.292602277	0.213854420
## 3	0.256094322	0.421742652
## 4	0.307673608	0.126011692
## 5	0.078974104	0.132236395
## 6	0.157911045	1.179106679
## 7	0.004017406	0.377678328
## 8	1.606442366	0.110915294
## 9	0.645755647	0.055102800
## 10	0.258321205	0.296699060
## 11	1.978977347	0.013957228
## 12	0.001017504	0.193780447
## 13	0.020889131	4.378227637
## 14	0.095298259	0.177294699
## 15	0.212978993	0.899250039

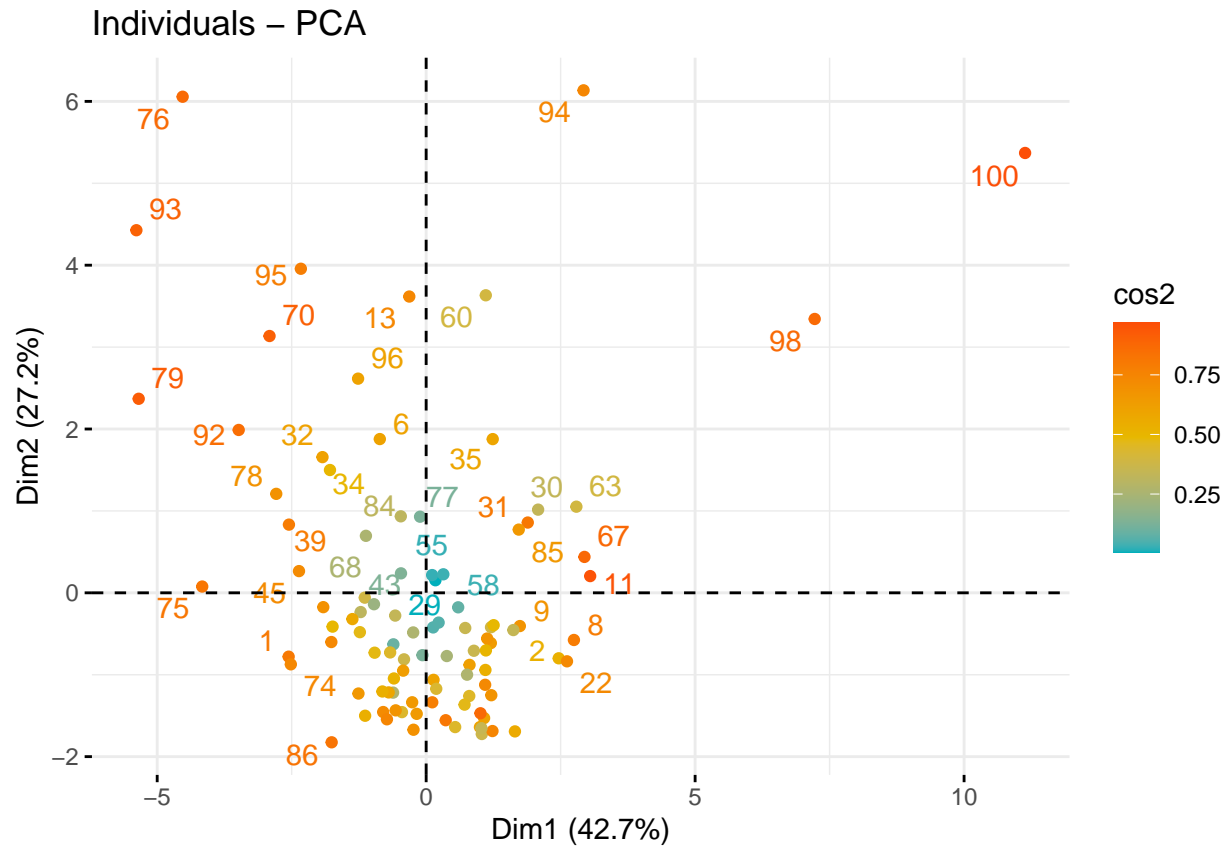
## 16	0.138372046	0.258915132
## 17	0.036046607	0.220631338
## 18	0.228047475	0.994141859
## 19	0.002731144	0.597745111
## 20	0.660927640	0.120721943
## 21	0.011771640	0.934869942
## 22	1.459271134	0.233888577
## 23	0.274062804	0.104080482
## 24	0.195211396	0.178762539
## 25	0.012348047	0.078292636
## 26	0.043005315	0.709712349
## 27	0.336888829	0.505838976
## 28	0.080418289	0.496786303
## 29	0.006307504	0.007887065
## 30	0.921857814	0.345218181
## 31	0.759369545	0.246336367
## 32	0.790518155	0.918352966
## 33	0.167722385	0.167308098
## 34	0.680231468	0.752362748
## 35	0.326039850	1.178358909
## 36	0.778188419	0.010257336
## 37	0.246902452	0.782298067
## 38	0.276886433	0.001166834
## 39	1.384604220	0.232052788
## 40	0.135820540	0.708693072
## 41	0.103320512	0.494648076
## 42	0.113112297	0.797496563
## 43	0.046550300	0.018928783
## 44	0.006554745	0.730349892
## 45	1.189310735	0.023524840
## 46	0.322986317	0.077208133
## 47	0.031379747	0.199202481
## 48	0.309570708	0.058894838
## 49	0.136717167	0.530973735
## 50	0.643913436	0.056971123
## 51	0.068197662	0.688805566
## 52	0.003578272	0.059659715
## 53	0.323357077	0.952775495
## 54	0.273861100	0.753546628
## 55	0.002635481	0.015712479
## 56	0.216912570	0.723249031
## 57	0.140646069	0.485390039
## 58	0.021738553	0.017096181
## 59	0.579985068	0.957689013
## 60	0.261622847	4.415190544
## 61	0.200093232	0.006482670
## 62	0.108483743	0.623334450
## 63	1.659727905	0.370118697
## 64	0.070447271	0.025840219
## 65	0.315235274	0.018473060
## 66	0.262514456	0.165484676
## 67	1.843195355	0.064295126
## 68	0.266642520	0.162015030
## 69	0.399866122	0.034368135

```
## 70 1.804107888 3.289212105
## 71 0.028143210 0.809381839
## 72 0.038805764 0.302133135
## 73 0.076391479 0.365854471
## 74 1.346927923 0.255064184
## 75 3.688063544 0.001970556
## 76 4.360885620 12.273295683
## 77 0.002727833 0.289017205
## 78 1.652888413 0.488622400
## 79 6.077627310 1.877907419
## 80 0.014350794 0.597428215
## 81 0.559049801 0.069123802
## 82 0.112377007 0.061648002
## 83 0.334954304 0.052251627
## 84 0.047264218 0.291854923
## 85 0.630747572 0.199380444
## 86 0.656613369 1.114305193
## 87 0.011450942 0.044166746
## 88 0.075528707 0.010473691
## 89 0.312117039 0.522832124
## 90 0.061932658 0.898841082
## 91 0.007504547 0.459300917
## 92 2.584056577 1.322870633
## 93 6.175928706 6.557038453
## 94 1.822473767 12.592349981
## 95 1.154295851 5.237390621
## 96 0.340717186 2.287635282
## 97 0.123117613 0.334128763
## 98 11.104696663 3.740345499
## 99 0.225151569 0.914339864
## 100 26.385698344 9.648325208
```

```
fviz_pca_ind (data.pca, col.ind = "cos2",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE # Évite le chevauchement de texte
              )
```

Graphe des individus :

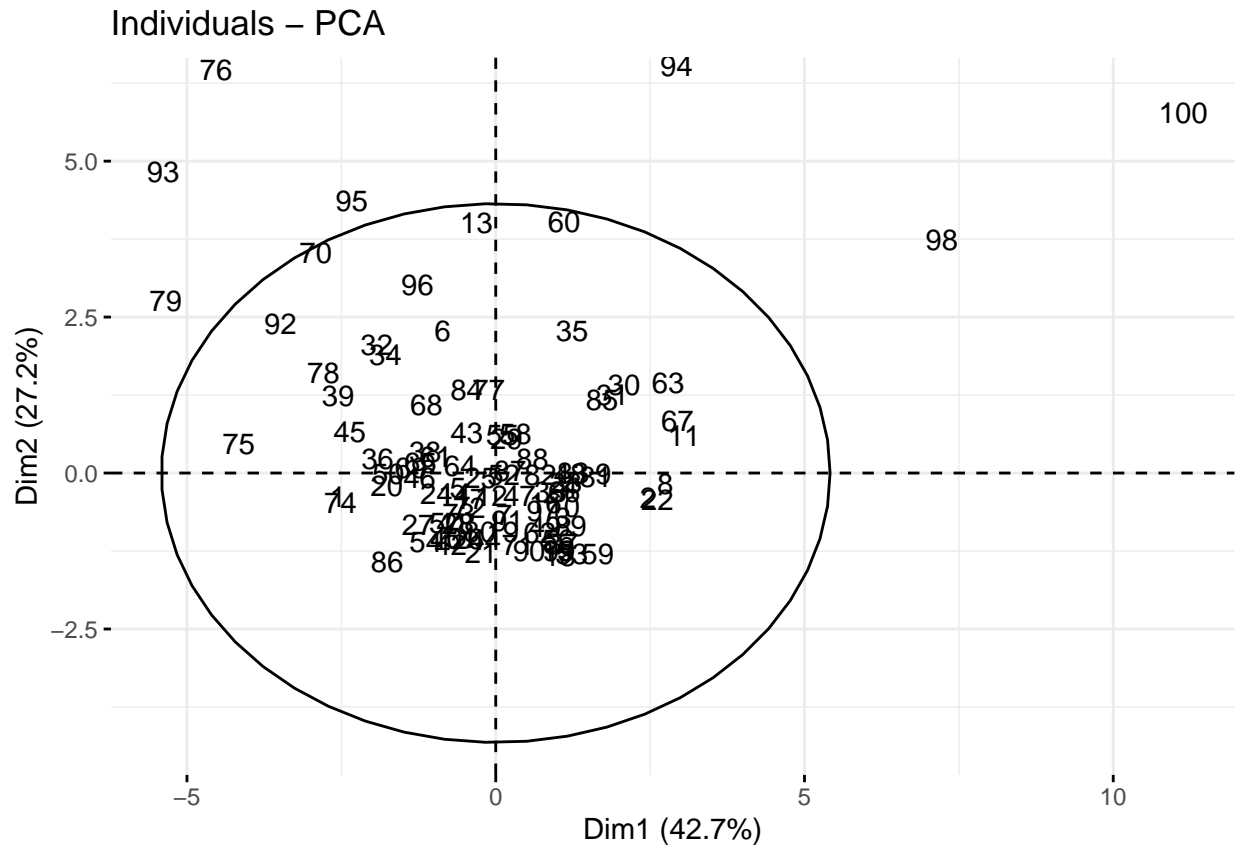
```
## Warning: ggrepel: 60 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



#### Création d'un cluster

```
fviz_pca_ind(data.pca,
  geom.ind = "text", # Montre les points seulement (mais pas le "text")
  palette = c("#00AFBB", "#E7B800", "#FC4E07"),
  addEllipses = TRUE, # Ellipses de concentration
  legend.title = "Groups"
)
```



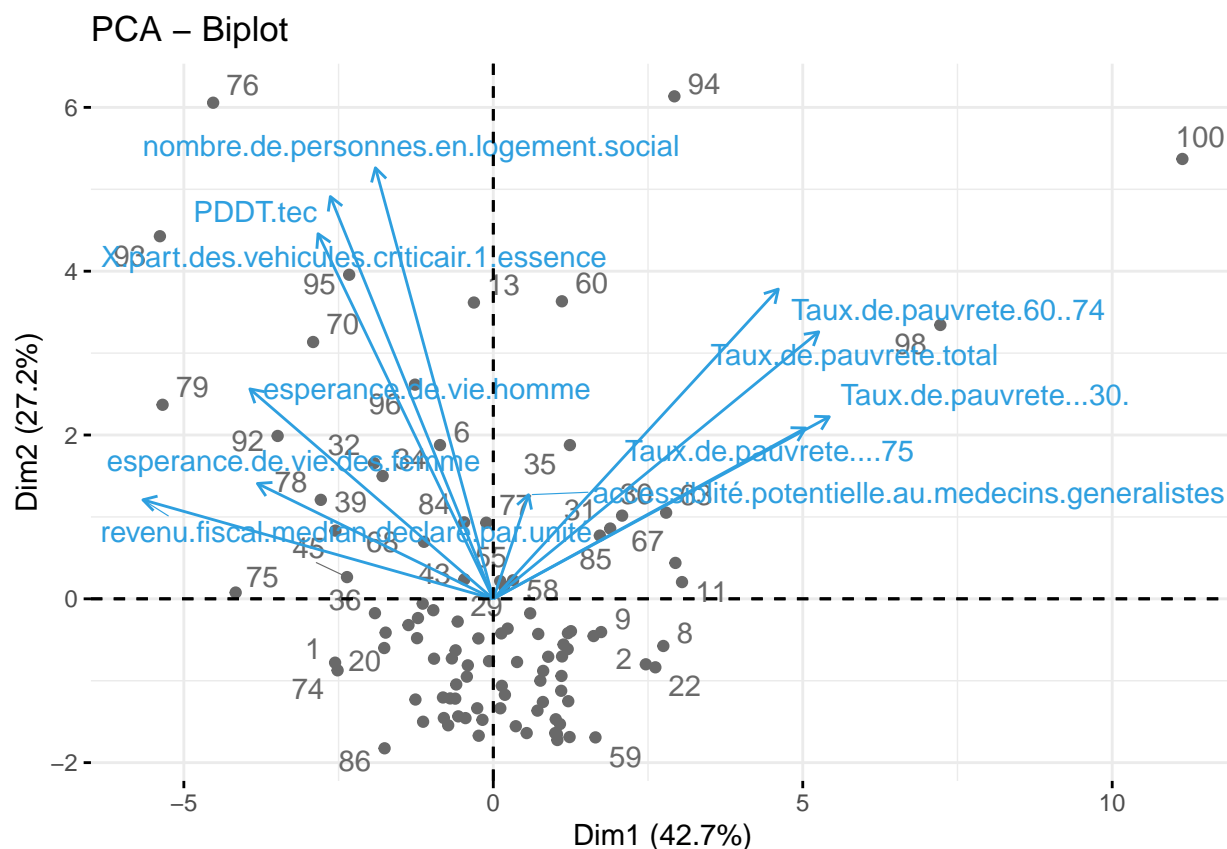


Nous remarquons que la plupart des départements de France ont des comportement similaires en moyenne mais 6 départements se detachent : 2 départements d’Outre-mer ( *La Guyane et la Guadeloupe*), 4 départements d’île-de-France ( *Hauts-de-Seine, Essonne, Seine-Saint-Denis, Seine-et-Marne*) et la *Haute-Savoie*.

Pour analyser leur comportement, nous allons faire un biplot :

```
fviz_pca_biplot(data.pca, repel = TRUE,
  col.var = "#2E9FDF", # Couleur des variables
  col.ind = "#696969" # Couleur des individus
)
```

```
## Warning: ggrepel: 57 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



#### #### Interprétation des axes :

Nous voyons ainsi à travers ce graphe, que par exemple que les individus *Guyane* et la *Martinique* ( 100 et 98 ) sont très proches : Ca veut dire qu'ils ont des niveaux de pauvreté très proches en moyenne avec de fortes valeurs en *Guyane*.

Les départements a droite de l'axe 1 sont des départements qui ont de fortes valeurs pour tous les indicateurs de pauvreté.

Le département 94 (*Hauts-de-Seine*) concentre beaucoup de personnes en logement HLM. Ce dont on pouvait s'attendre en raison du prix élevé du metre carré dans ces zones.

Quelques départements d'*Ile-de-France* (*Seine-Saint-Denis*(95),*Essonne*(93),*Val-de-Marne*(96)) prennent de fortes valeurs pour PDDT.tec. Donc la part de déplacement de domicile travail en transport en commun en *Ile-de-France* est très élevée en raison du forts taux d'alternatives de transport (RER, Metro, Train, bus, Tram).

A l'opposé le département du *Nord* (59) est très peu desservi en Transport en commun d'ou de faibles valeurs en PDDT.tec. *Le Vaucluse* (86) aussi dans une moindre mesure.

Le revenu fiscal médian par unité de consommation le plus élevé de France est celui de *Seine-et-Marne* (79).

La part de véhicules criticair1 est très élevée dans plusieurs départements d'*Ile-de-France*. En partie due au mesures de circulations alternées pair/impair et les arrêtés municipaux d'interdiction de circulation dans certaines zones sans vignette criticair1.

L'espérance de vie homme/femme est très élevée dans les *Territoires de Belfort* (92) et en *Seine-et-Marne*(79). Il est intrinsèquement lié au revenu median par unité puisque ces departements prennent de fortes valeurs sur cette variable.

Nous pouvons dire que de maniere générale, l'axe 1 sépare :

- d'un coté (droit) les départements les plus pauvres de France, peu équipées en transport en commun, d'espérance de vie faible par sexe et de revenu median par unité faible
- et de l'autre (coté gauche) les départements ou le revenu median est très élevé, bien desservis avec une espérance de vie élevée et un revenu median élevé.

Par rapport a l'axe 2, les corrélations sont moins forte sur cet axe (représente 27% de l'information).

Nous remarquons aussi que la corrélation entre revenu median par unité et l'axe 1  $< 0$ . Donc les départements à gauche du graphe ont tendance à avoir un revenu fiscal médian par habitant très élevé.

Inversement les départements à droite du graphe ont des revenus fiscal median par unité très faibles.

**Conclusion :** Nous remarquons que les indicateurs de l'objectif 1 (Eradication pauvreté) sont très corrélées avec l'axe 1.

Ainsi les indicateurs de l'objectif 1 (pauvreté totale et par catégories + revenu médian qui est aussi dans objectif 10) sont toutes corrélées avec l'axe 1.

Et les indicateurs de l'objectif 11 (Villes et communautés durables) dont le nombre de personnes en HLM, et de l'objectif 9 ( Infrastructures et Innovation) dont le PDDT.tec et les voitures criticairl sont très corrélées avec l'axe 2.

Nous pouvons en conclusion dire que ***l'axe 1 mesure le niveau de pauvreté par département.***

Alors que ***l'axe 2 mesure la qualité de la vie par département.***