

Rapport du TP ANAD

Thème :

Analyse des Composants Principales

Réalisé par :
MAHMAHI Anis

Groupe :
SID

Dataset choisie : “ PV 1CS S1 ”

Comme étape initiale, il était obligé d'éliminer les trois derniers individus ayant bloqué l'année pour ne pas fausser les résultats de l'ACP.

On peut toujours les rajouter comme des individus supplémentaires à la fin de l'ACP .

Individu : 35

	SYS1	RES1	ANUM	RO	ORG	IGL	THP	LANG1
35	12.82	15.87	18.25	11.20	13.94	12.95	9.50	14.10

```
# Installation et importation des packages nécessaires
install.packages(c("FactoMineR", "factoextra"))
library("FactoMineR")
library("factoextra")
```

```
res <- PCA(my_data, graph = FALSE)
```

```
eig.val <- get_eigenvalue(res)
print(eig.val)
```

Valeur Propres	5.6700	0.684	0.3878	0.3233	0.3215	0.2809	0.1830	0.1485
Inertie (%)	70.875	8.558	4.8476	4.0418	4.0195	3.5117	2.2885	1.8571
Inertie Cum (%)	70.875	79.43	84.281	88.323	92.342	95.854	98.142	100.00

```
fviz_eig(res, addlabels = TRUE, ylim = c(0, 80))
```

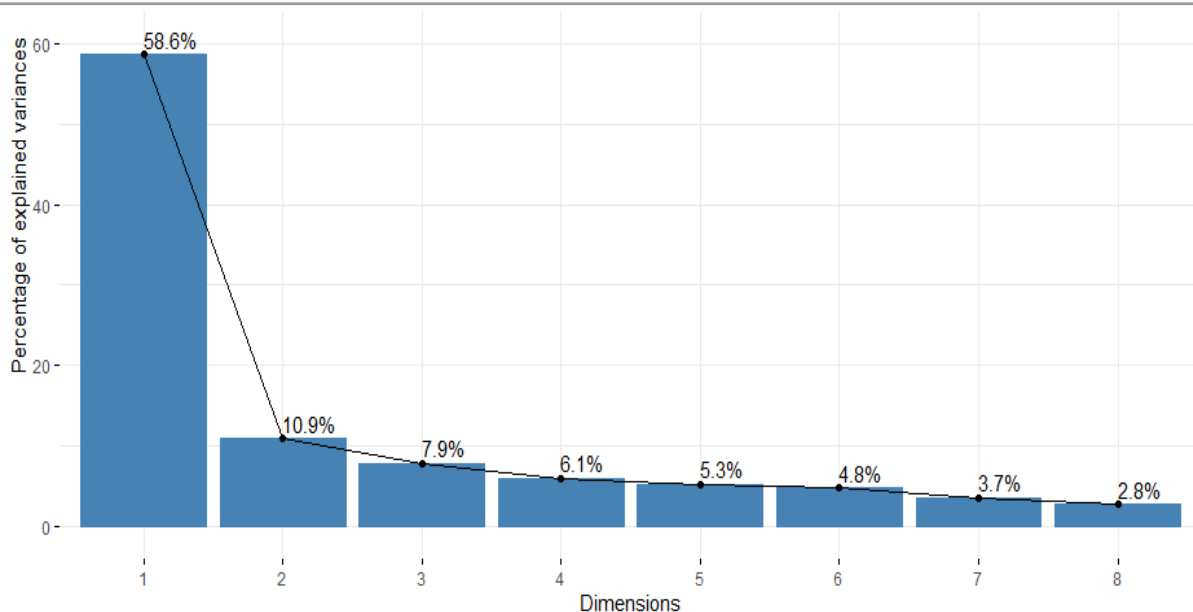


Figure 1: Inerties correspondantes aux valeurs propres

D'après la figure ci-dessus, on peut retenir seulement 4 composantes principales. 83.5% des informations contenues dans les données sont retenues par les 4 premières composantes principales.

```
var <- get_pca_var(res)
print(var$cos2); print(var$contrib); print(var$coord)
```

	V1	ctr	cos2	V2	ctr	cos2	V3	ctr	cos2
SYS1	0.804 4959	13.81 263	0.647 2137	-0.231 225	6.113 129	0.053 46504	0.227 13542	8.147 64069	0.051 59050
RES1	0.854 8709	15.59 659	0.730 8043	-0.113 9396	1.484 372	0.012 98223	-0.033 87739	0.181 25165	0.001 14767
ANUM	0.773 2448	12.76 035	0.597 9075	-0.182 5378	3.809 776	0.033 32006	-0.072 12628	0.821 57874	0.005 20220
RO	0.788 5822	13.27 158	0.621 8619	-0.258 2317	7.624 522	0.066 68360	0.013 65487	0.029 44673	0.000 18645
ORG	0.690 1018	10.16 377	0.476 2405	0.462 2083	24.42 6942	0.213 63651	-0.457 32921	33.03 08696	0.209 15000
IGL	0.809 8937	13.99 860	0.655 9278	0.160 7034	2.952 865	0.025 82557	-0.268 63092	11.39 65690	0.072 16257
THP	0.810 6594	14.02 508	0.657 1687	-0.216 7741	5.372 894	0.046 99100	0.178 62835	5.039 21497	0.031 90808
LANG1	0.546 3905	6.371 40	0.298 5426	0.649 3764	48.21 5500	0.421 68975	0.511 71094	41.35 34285	0.261 84808

Nuage des Variables :

Cercle de Corrélation :

```
fviz_pca_var(res, col.var = "black")
```

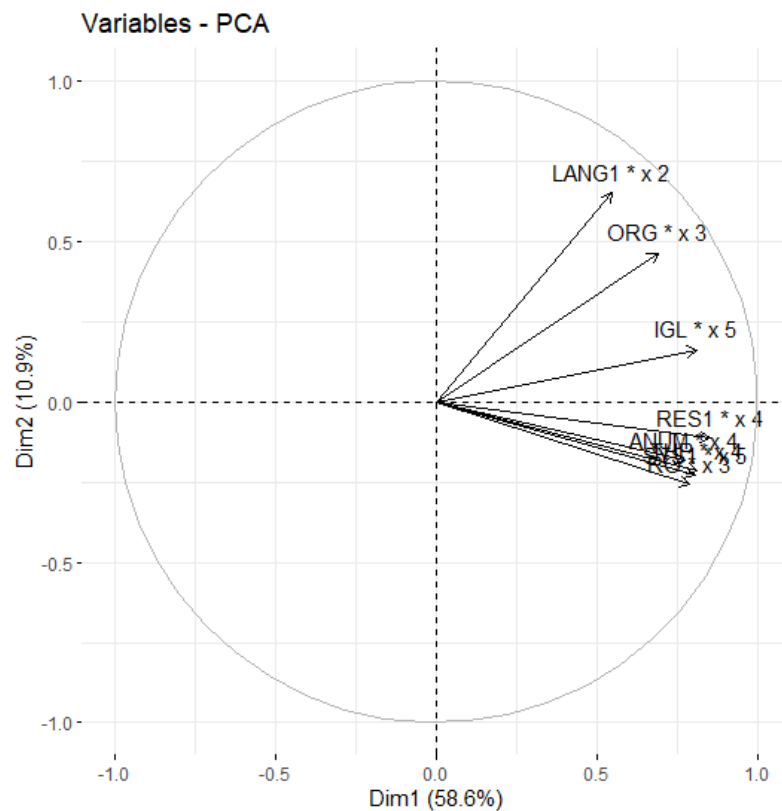


Figure 2: Cercle de Corrélation du premier plan factoriel

La longueur du vecteur mesure la qualité de représentation des variables, et on remarque que RES1, LANG1, THP, SYS1 sont les variables les mieux représentées par le premier plan factoriel.

Axe 1 : En considérant les variables dont : $C_{ab} \geq \frac{1}{P}$

```
fviz_contrib(res, choice = "var", axes = 1)
```

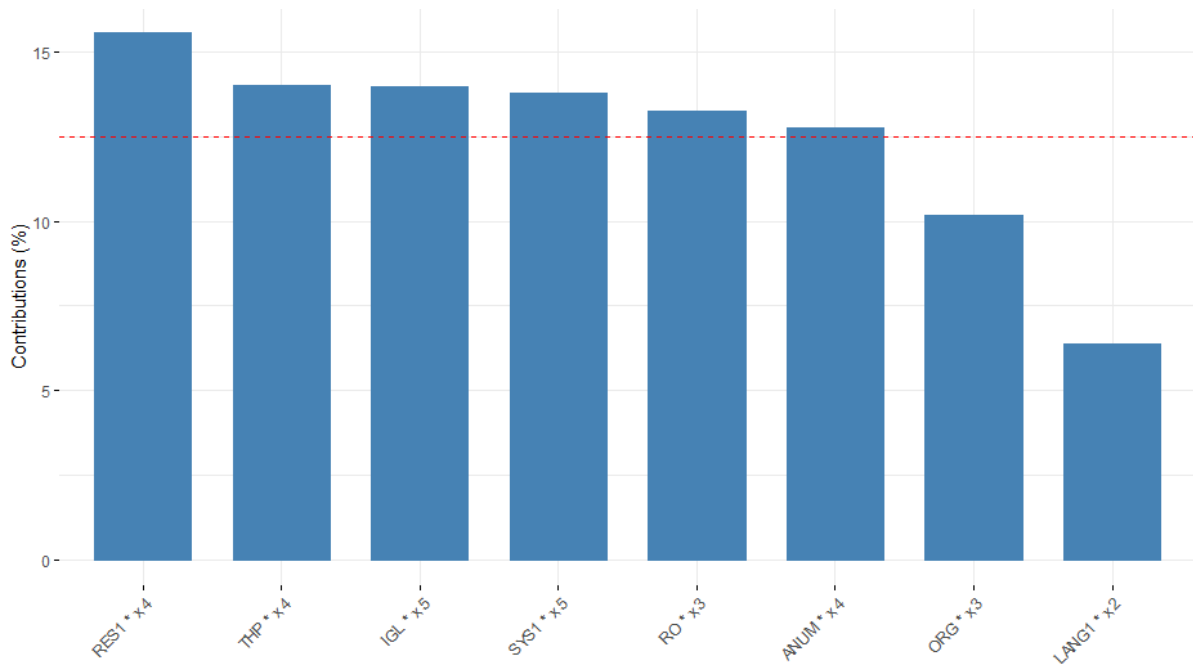


Figure 5: Contribution des variables sur l'axe 1

-	+
	RES1,THP,IGL,SYS1,RO,ANUM

Interprétation :

Axe 1, a un effet taille. Il mesure les notes des modules RES1, THP, IGL, SYS1, RO, ANUM.

Axe 2 : En considérant les variable dont : $C_{ab} \geq \frac{1}{p}$

```
fviz_contrib(res, choice = "var", axes = 2)
```

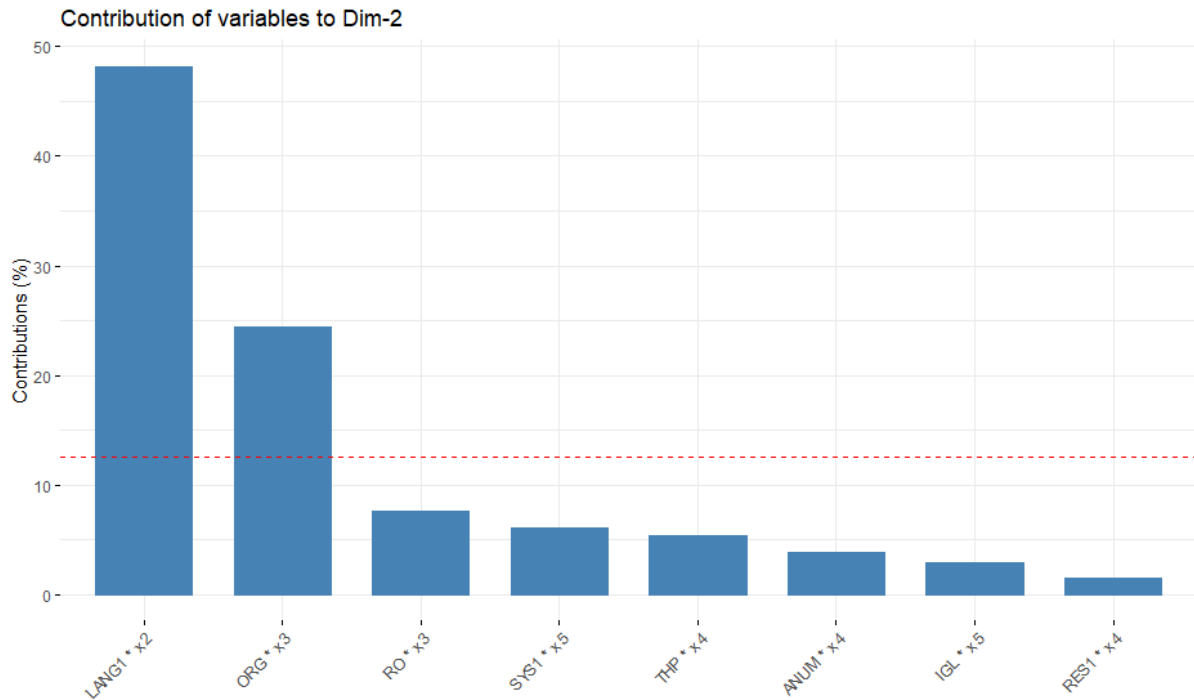


Figure 6: Contribution des variables sur l'axe 2

-	+
	ORG, LANG 1

Interprétation :

Axe 2, a un effet taille. Il mesure les notes des modules ORG, LANG 1.

Axe 3 : En considérant les variable dont : $C_{ab} \geq \frac{1}{P}$

```
fviz_contrib(res, choice = "var", axes = 3)
```

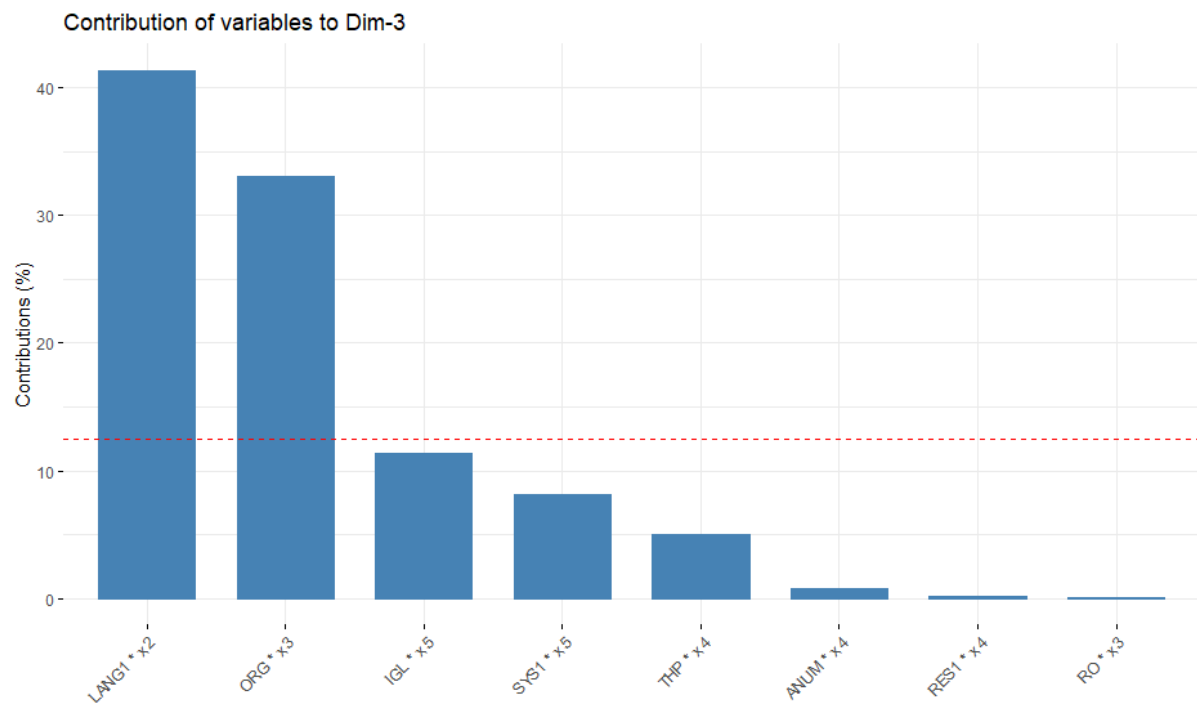


Figure 6: Contribution des variables sur l'axe 3

-	+
ORG	LANG 1

Interprétation :

Axe 3 est un axe d'opposition, il oppose les notes des modules ORG aux notes de LANG 1.

Axe 4 : En considérant les variable dont : $C_{ab} \geq \frac{1}{P}$

```
fviz_contrib(res, choice = "var", axes = 4)
```

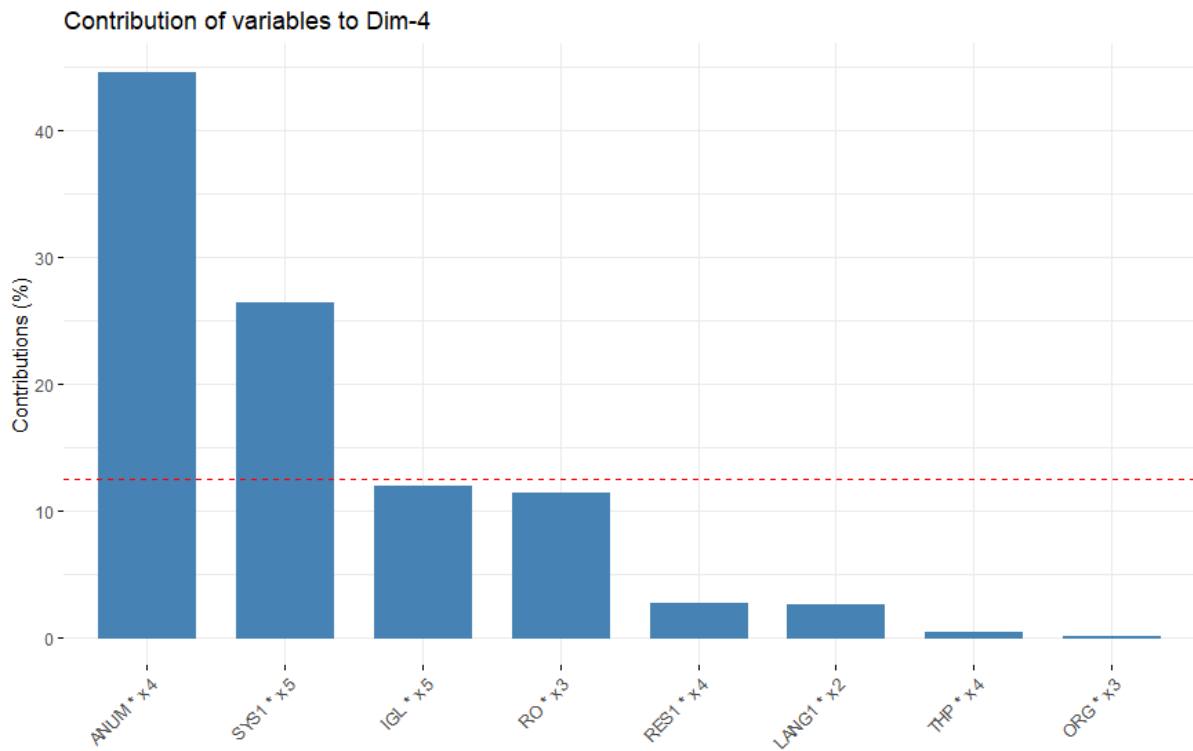


Figure 6: Contribution des variables sur l'axe 3

-	+
SYS1	ANUM

Interprétation :

Axe 4 est un axe d'opposition, il oppose les notes des modules ANUM aux notes de SYS 1.

Nuage des individus:

```
fviz_pca_ind(res)
```

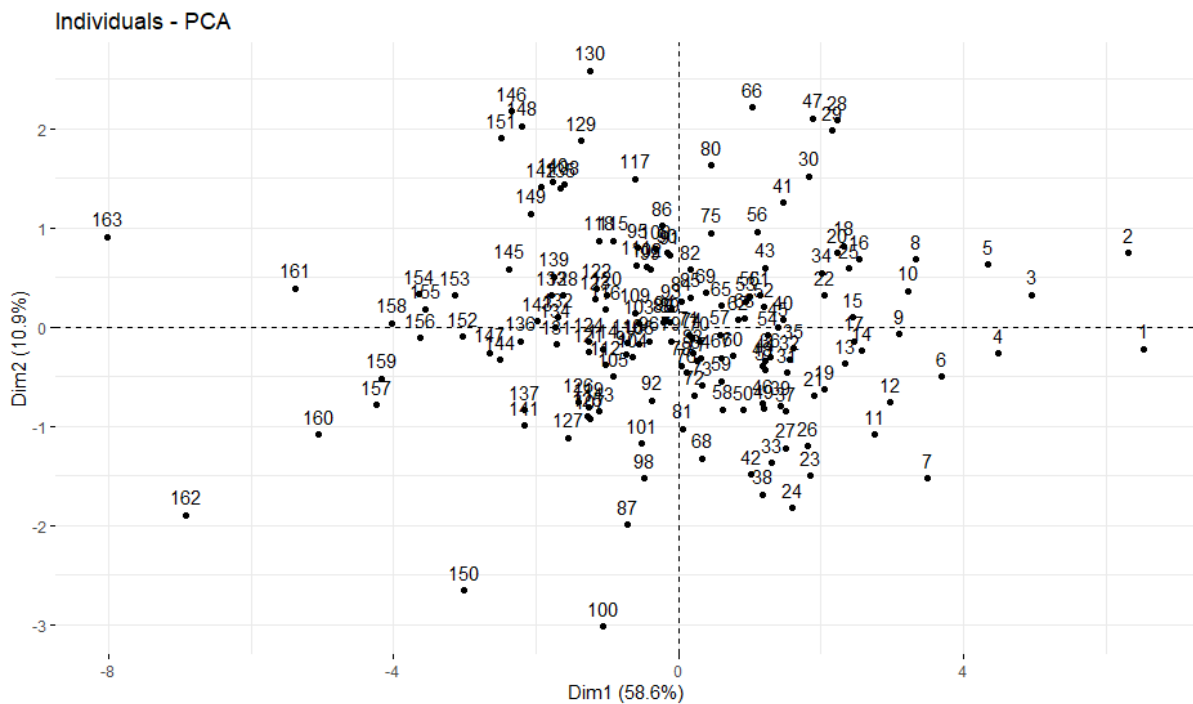


Figure 3: Nuage des Individus

```
fviz_pca_ind(res, col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = FALSE, )
```

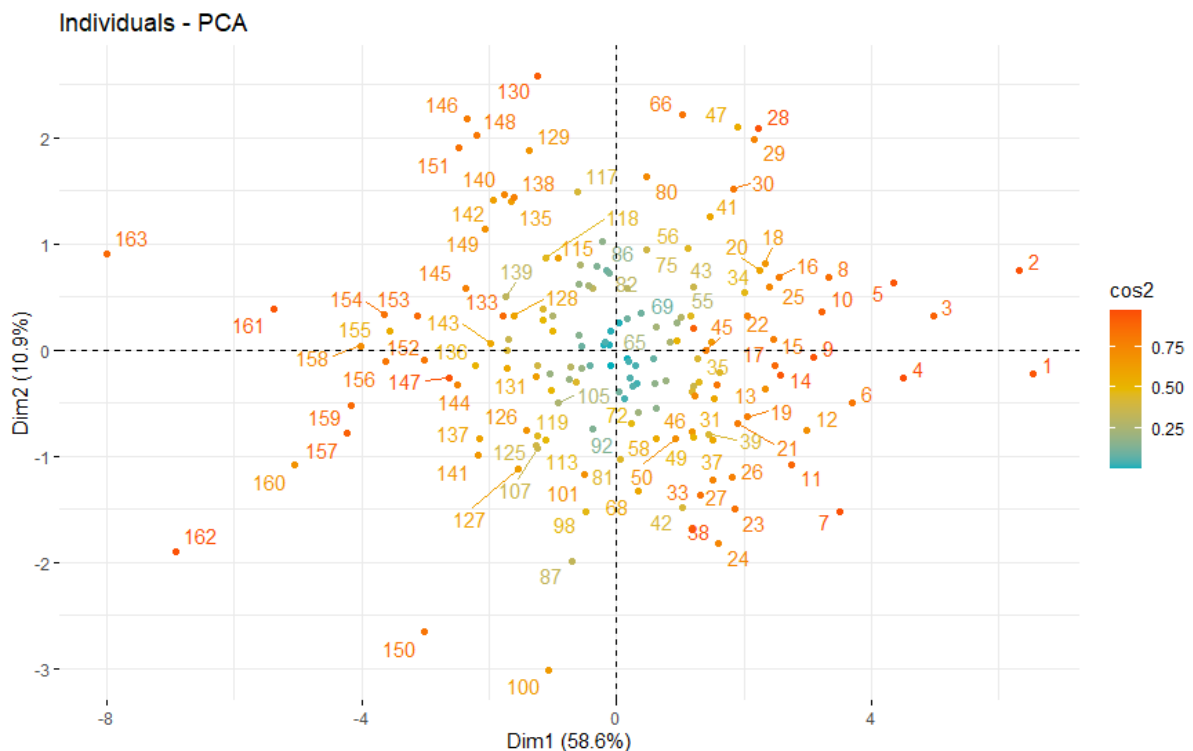


Figure 4: Qualité de représentation de chaque individu

*Les individus centrés (proches de l'origine) sont mal-représentés .
On note que les individus similaires sont proches entre eux.*

Interprétation de l'individu n°35 :

Premier Plan :

Pour interpréter l'individu n°35, il est bien de l'isoler pour voir son abscisse et sa coordonnée.

D'après la figure 5, on remarque qu'il est mal-représenté dans le premier plan factoriel car il est très proche de l'origine et il a un cos2 très petit.

```
fviz_pca_ind(res, col.ind = "cos2",  
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
             repel = FALSE,  
             select.ind=list(name = c("1", "42", "35","163"))  
))
```

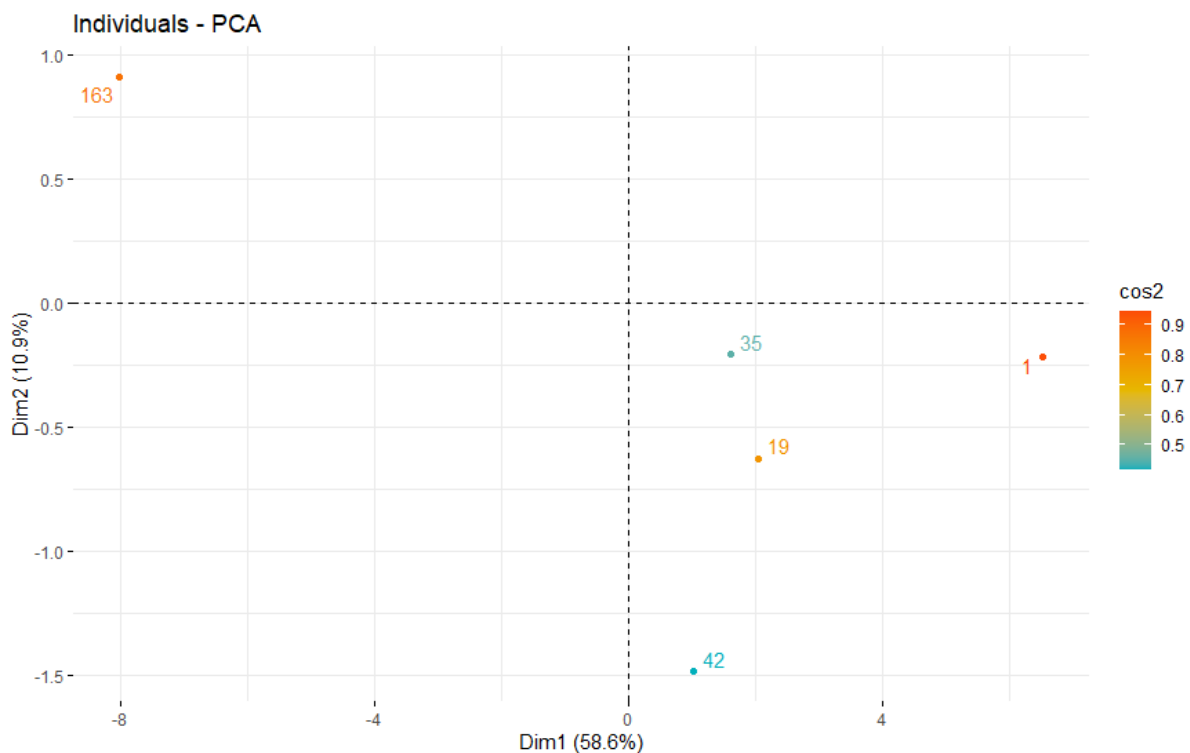


Figure 5: Qualité de représentation d'individu n° 35 dans le premier plan.

```
plot(res, choix = "ind", cex=1, select = 35, axes = c(1,3), xlim = c(-7, 7), ylim = c(-2, 2),)
```

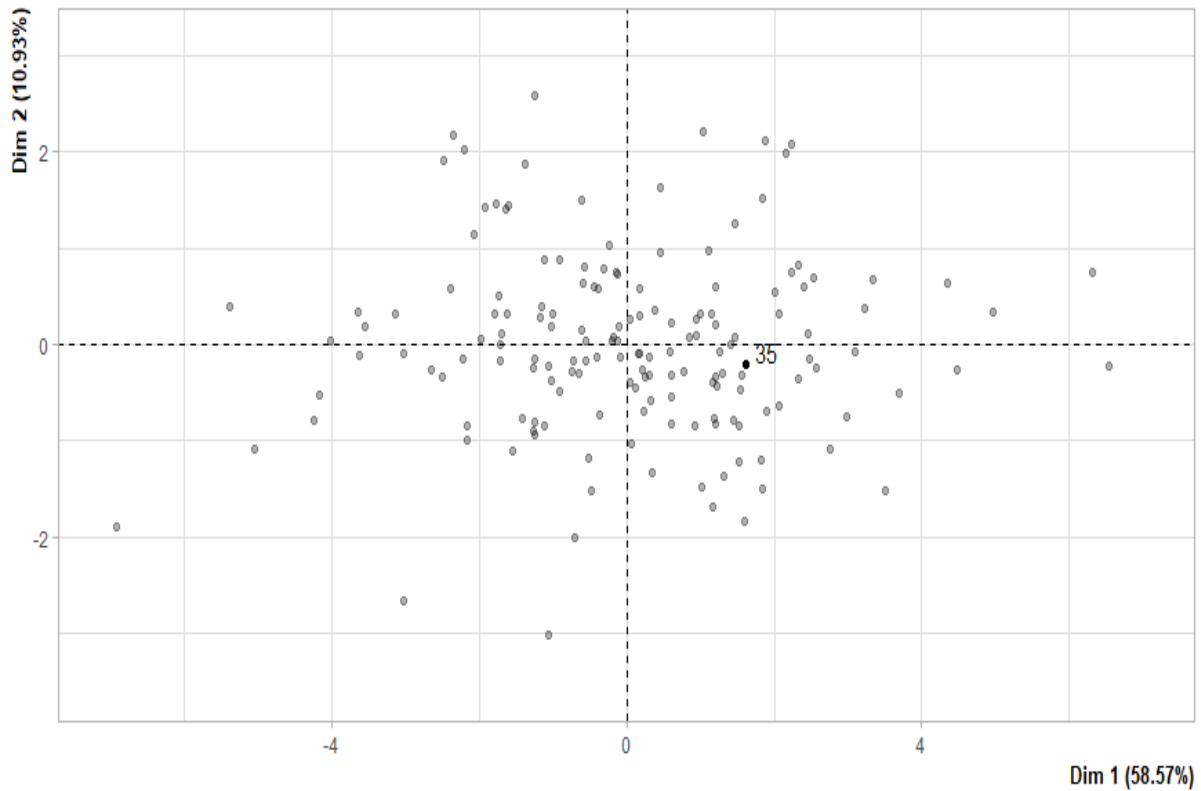


Figure 6: Biplot de l'individu n°35 dans le 1er plan factoriel

Interprétation :

On remarque que l'individu n°35 est mal-représenté dans le 2eme axe (ordonné très proche de l'origine).

Sur l'axe 1, on peut dire qu'il se situe pas très proche ni trop loin de l'origine donc il est moyen à bon dans RES1, THP, IGL, SYS1, RO, ANUM.

	SYS1	RES1	ANUM	RO	IGL	THP
35	12.82	15.87	18.25	11.20	12.95	9.50

Deuxième plan factoriel : constitué de (Dim 1, Dim 3).

D'après la figure 7, on remarque que l'individu $n^{\circ}35$ est très bien représenté sur l'axe 3.

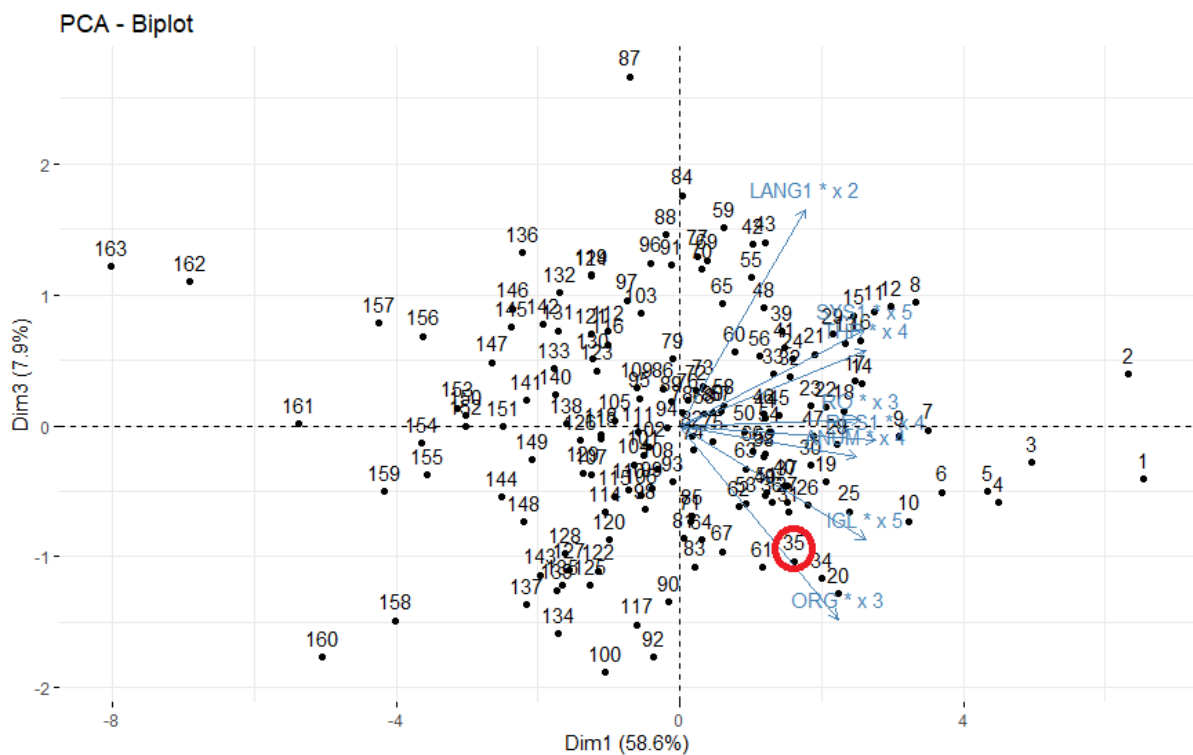


Figure 7: Biplot de l'individu $n^{\circ}35$ dans le deuxième plan factoriel

Interprétation :

L'individu $n^{\circ}35$ a une ordonnée importante donc il est bon en ORG mais pas grande que celle de $n^{\circ}100$, 92 par exemple, et cela ne veut dire pas que ces individus sont mieux que $n^{\circ}35$ en ORG, mais ils sont plus mauvais en LANG 1 que $n^{\circ}35$ parce que cet axe oppose simultanément ORG au LANG 1, et c'est pour ça 35 s'est situé un peu plus haut que ces individus.

	ORG	LANG 1
35	13.94	14.10
100	10.70	7.63

92	12.84	11.37
----	-------	-------

Troisième plan factoriel : constitué de (Dim 1, Dim 4).

D'après la figure 8, on remarque que l'individu $n^{\circ}35$ est très bien représenté sur l'axe 4.

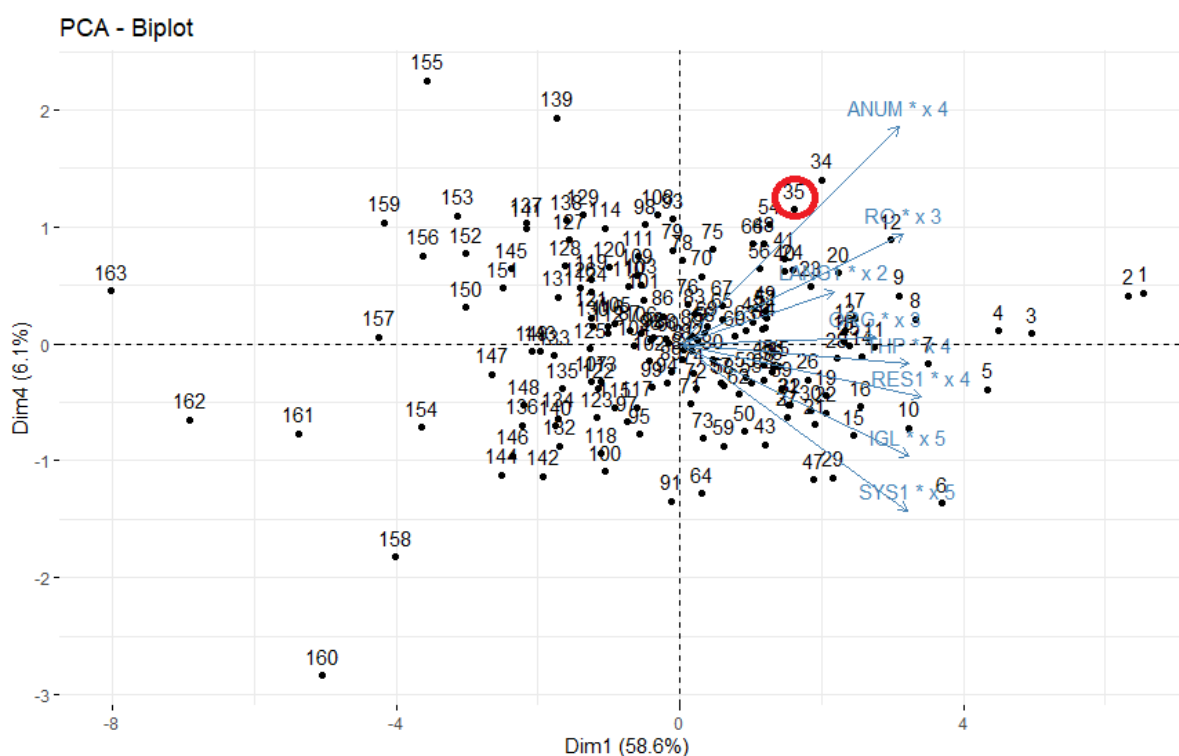


Figure 8: Biplot de l'individu $n^{\circ}35$ dans le troisième plan factoriel

Interprétation :

L'individu $n^{\circ}35$ a une ordonnée importante donc il est bon en ANUM mais pas grande que celle de $n^{\circ}34$ par exemple, et cela ne veut dire pas qu'il mieux que $n^{\circ}35$ en ANUM, mais il est plus mauvais en SYS 1 que $n^{\circ}35$ parce que cet axe oppose simultanément ANUM au SYS1 , et c'est pour ca 35 s'est situé un peu plus bas.

	ANUM	SYS1
35	18.25	12.28
34	17.25	10.68