

## Projet personnel : CSP et la principale source d'information

Nous étudierons la relation entre la catégorie socio-professionnelle (CSP) et la principale source d'information sur les problèmes d'environnement. Sept CSP sont étudiées :

agriculteur (AGRI),  
cadre supérieur (CSUP),  
cadre moyen (CMOY),  
employé (EMPL),  
ouvrier (OUVR),  
retraité (RETR),  
chômeur (CHOM).

Les 1283 personnes interrogées devaient indiquer leur principale source d'information sur les problèmes d'environnement, parmi les six sources suivantes :

télévision (TEL),  
journaux (JOU),  
radio (RAD),  
livres (LIV),  
associations (ASS) ,  
mairie (MAI).

Table 1: Tableau de contingence

	TEL	JOU	RAD	LIV	ASS	MAI	Total
AGRI	26	18	9	5	4	6	68
CSUP	19	49	4	16	5	3	96
CMOY	44	87	4	39	14	3	191
EMPL	83	87	13	24	5	1	213
OUVR	181	107	16	31	7	7	349
RETR	167	95	29	15	7	7	320
CHOM	27	9	4	2	2	2	46
Total	547	452	79	132	44	29	1283

Voici le jeu de données à notre disposition pour mener notre analyse. Il s'agit d'un tableau des effectifs croisée entre les deux variables que nous étudions. On apprend par exemple que 181 ouvriers ont comme principale source d'information sur les problèmes d'environnement la télévision.

## Indépendance

Avant de commencer à analyser nos données on vérifie si il y a un lien entre nos deux variables. Pour cela on réalise un test d'indépendance du  $\chi^2$ .

```
##  
## Pearson's Chi-squared test  
##  
## data: media_sm  
## X-squared = 156.33, df = 30, p-value < 2.2e-16
```

On trouve un pvalue très proche de 0, on rejette donc l'hypothèse d'indépendance, il y a un lien entre la socio-professionnelle et la principale source d'information sur les problèmes d'environnement. Une AFC est donc légitime.

## Analyse Factorielle des Correspondances

Nous allons donc réaliser l'AFC sur nos données.

## Valeurs propres

On cherche d'abord combien de dimension nous allons garder pour représenter au mieux nos données.

Table 2: Valeurs propres

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.092	75.099	75.099
Dim.2	0.022	18.001	93.100
Dim.3	0.006	5.269	98.369
Dim.4	0.002	1.437	99.806
Dim.5	0.000	0.194	100.000

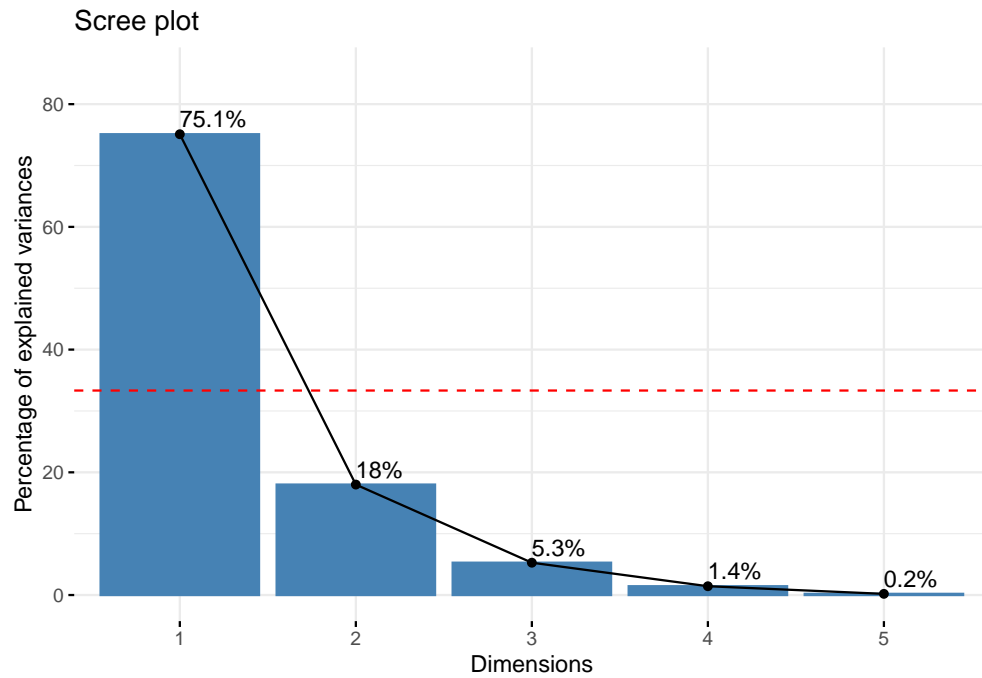


Figure 1: Visulation des valeurs propres

Dans le tableau on voit que les deux premiers axes expliquent 93.1% de la variance totale. C'est un pourcentage très acceptable. Avec le graphique on voit que les dimensions 1 et 2 expliquent environ 75.% et 18% de l'inertie totale, respectivement. On décide donc de retenir ces deux axes.

## Catégorie socio-professionnelle

On commence par étudier la variable CSP.

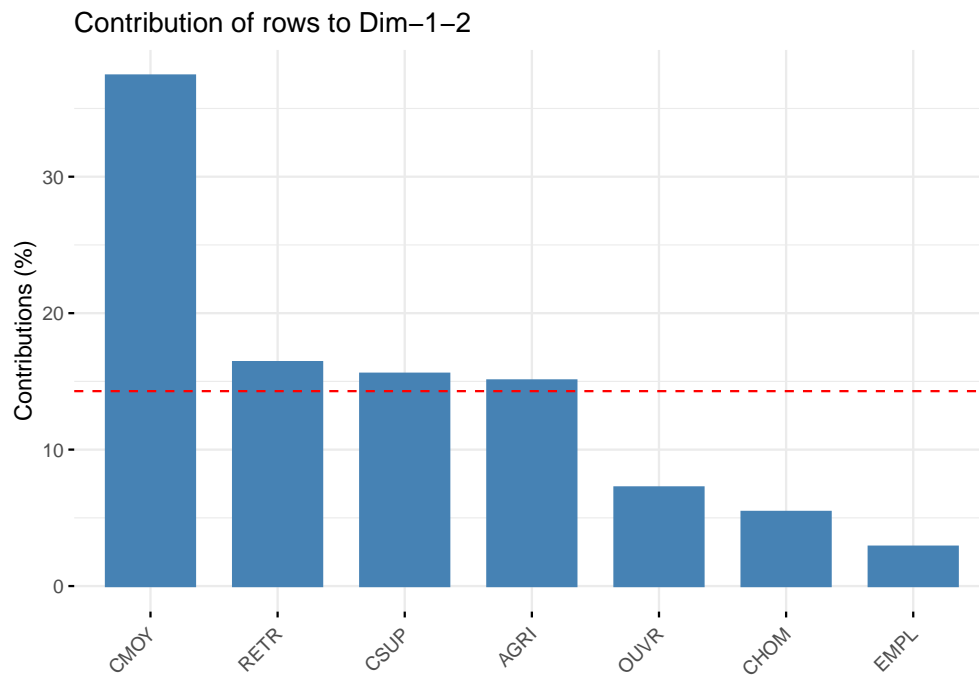


Figure 2: Visulation des contrivutions des CSP sur le premier plan

On cherche à savoir quelle catégorie socio-professionnelle contribue le plus au premier plan. Avec le diagramme en bar ci-dessus, on voit les différentes contributions pour chaque catégorie. La ligne poitillée rouge correspond à la contribution moyenne, toutes les variables au dessus de cette ligne peuvent être considérer comme fortement contribuant au premier plan. On retient donc les agriculteurs, la classe supérieur, les retraiters, et la catégorie la plus contribuant, la classe moyenne qui à la meilleur contribution au premier plan.

Table 3: Coordonnées des CSP sur le premier plan

	Dim 1	Dim 2
AGRI	-0.150	0.548
CSUP	0.477	0.093
CMOY	0.534	0.015
EMPL	0.085	-0.112
OUVR	-0.145	-0.096
RETR	-0.273	0.005
CHOM	-0.388	0.147

Avant de représenter le nuage des CSP, on regarde les coordonnées de chaque catégories sur le premier plan. Tous les individus avec une coordonnée négative sur la dimension 1, seront dans la partie gauche du graphique, et inversement pour les coordonnées positive. Et Tous les individus avec une coordonnée négative sur la dimension 2, seront dans la partie basse du graphique, et inversement pour les coordonnées positive. Par exemple les agriculteurs seront dans la partie haute gauche du graphique.

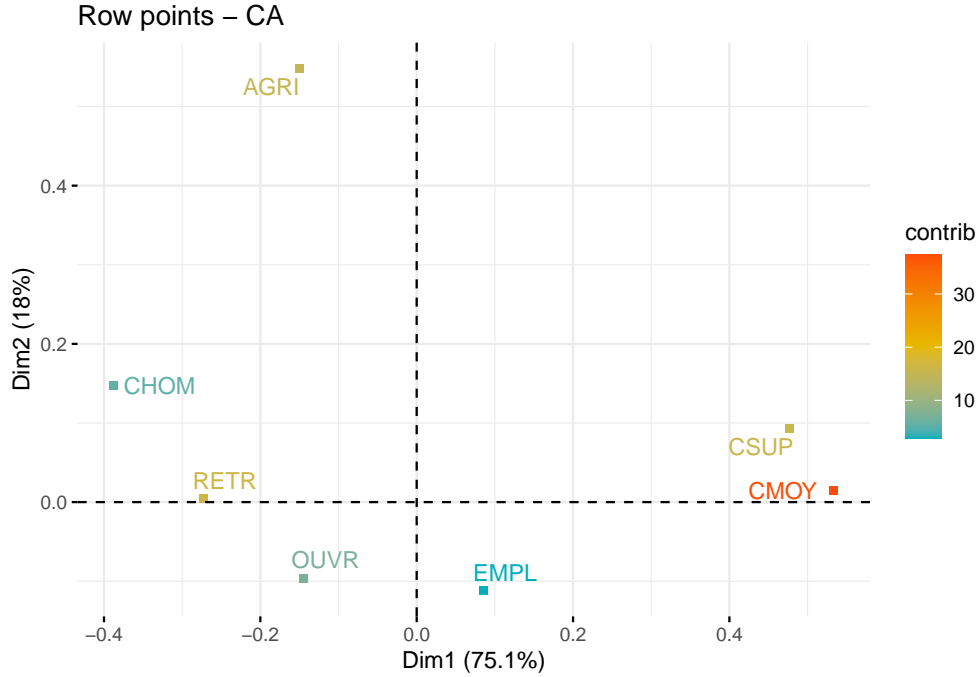


Figure 3: Nuage des CSP sur le premier plan

On peut ensuite tracer le nuage des catégories socio-professionnelle. Ici il y a une couleur selon la contribution, plus la contribution au premier plan est grande, plus la couleur sera chaude. On retrouve donc la classe moyenne avec la couleur la plus chaude car comme nous venons de le voir c'est la catégorie qui contribue le plus à ce plan.

ici on remarque que la classe moyenne et la classe supérieure sont regroupées, cela signifie qu'ils ont un profil similaire. À l'inverse les catégories qui s'opposent vont être corrélées négativement. Par exemple les employés et les agriculteurs ont ici des profils qui s'opposent mais cela reste léger.

## Source d'information

On s'intéresse maintenant à la deuxième variable de nos données.

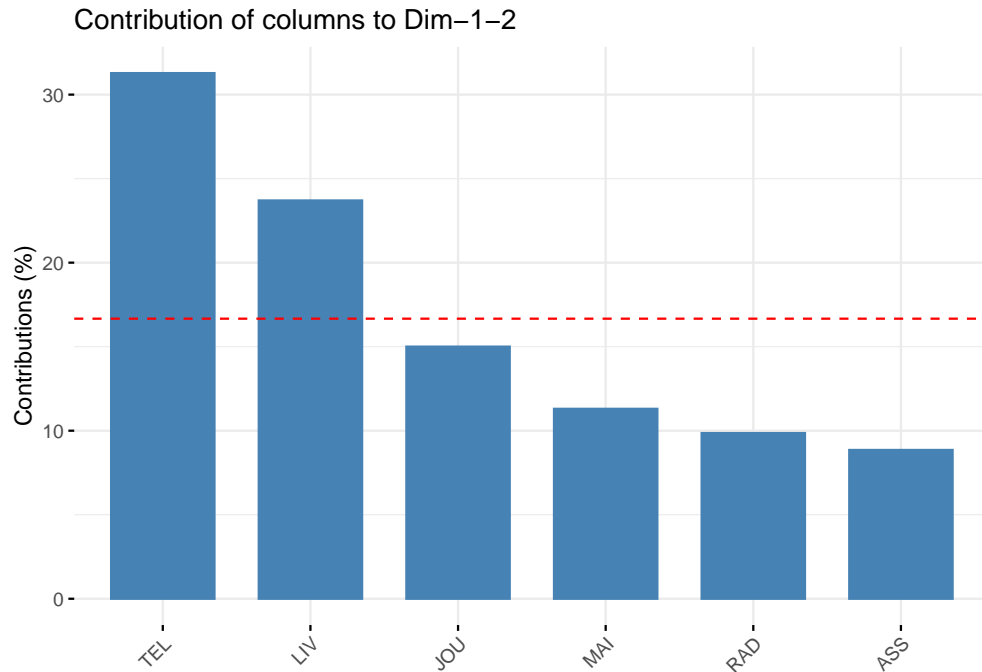


Figure 4: Visulation des contributions des sources d'information

On commence avec les contributions sur le premier plan. On voit avec le graphique que deux sources d'information ont une contribution supérieure à la contribution moyenne. Ce sont la télévision et les livres, ce seront donc ces deux modalités qui contribueront le plus au premier plan.

Table 4: Coordonnées des sources d'information sur le premier plan

	Dim 1	Dim 2
TEL	-0.281	-0.065
JOU	0.217	-0.037
RAD	-0.334	0.265
LIV	0.511	-0.025
ASS	0.450	0.301
MAI	-0.169	0.734

Avant de représenter le nuage des sources d'information, on regarde les coordonnées de chaque source sur le premier plan. Tous les individus avec une coordonnée négative sur la dimension 1, seront dans la partie gauche du graphique, et inversement pour les coordonnées positives. Et tous les individus avec une coordonnée négative sur la dimension 2, seront dans la partie basse du graphique, et inversement pour les coordonnées positives. Par exemple la télévision sera dans la partie basse gauche du nuage.

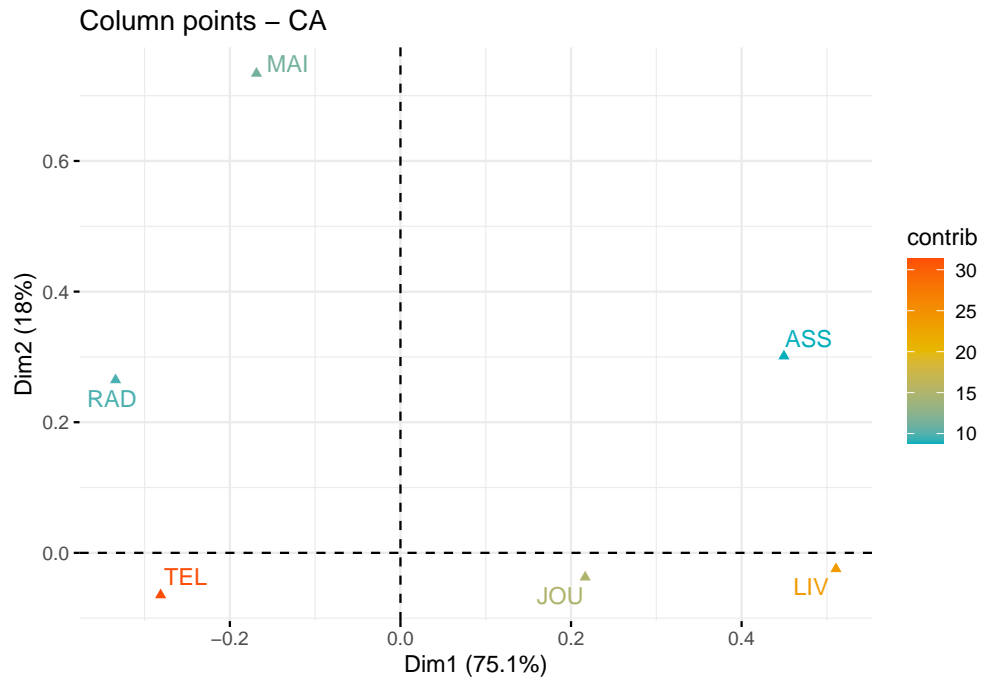


Figure 5: Nuage des sources d'informations

On peut ensuite tracer le nuage des sources d'informations. Comme pour le nuage précédent, il y a une couleur selon la contribution, plus la contribution au premier plan est grande, plus la couleur sera chaude. On retrouve donc la télévision et les livres avec les contributions les plus élevées.

Ici il n'y a pas de groupe qui se forme, donc aucune source d'informations a un profil similaire. Il y a quelque opposition, comme la télévision et les associations, donc les profils de ces deux sources d'information s'opposent, mais cela n'est pas net.

## CSP et source d'information

On peut maintenant croiser les deux variables pour voir les liens et les dissimilarités.

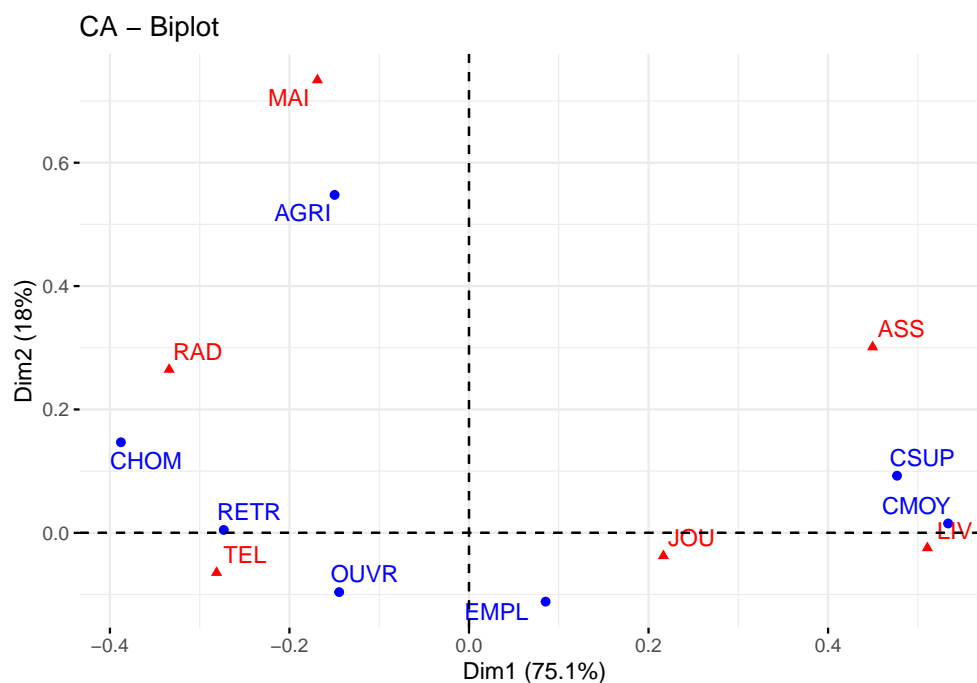


Figure 6: Graphe superposé

Pour cela on superpose les deux nuages. Les CSP sont représentées par des points bleus et des sources d'information par des triangles rouges.

On remarque que les modalités sont disposées en arc de cercle. Ce phénomène est connu sous le nom d'effet Guttman. Il y a donc un ordre sous-tend les modalités. On voit que la classe moyenne et la classe supérieure sont fortement liées au livre, les employés aux journaux, les ouvriers et retraités à la télévision, les chômeurs aux radios, et les agriculteurs à la mairie.

Pour aller plus loin, on imagine avec ces résultats que le coût d'une source d'information a un impact. En effet on remarque que les sources d'information gratuites comme la télévision, la radio et la mairie, vont être associées à la catégorie sociale la plus modeste. Tandis que les livres et les journaux, des sources d'informations payantes vont plutôt être liées à la catégorie la plus aisée.