

# ANALYSE FACTORIELLE DES DONNEES

## Présentation des données

Dans le cadre de ce projet d'analyse factorielle de données, j'ai choisi un fichier Excel contenant une enquête sur la connectivité réalisée auprès de 2036 français représentatifs de la population française. Elle s'est déroulée du 17 novembre au 8 décembre 2016. Le fichier utilisé s'appelle Base enquête Capacity-WIP et est disponible à l'adresse : <https://www.data.gouv.fr/en/datasets/enquete-capacity-world-internet-project/>

1<sup>ère</sup> étape :

- \*lire le fichier Excel des données brutes.
- \*filtrer les départements de la région AURA.
- \*sélectionner les variables sans données manquantes.

Les départements d'AURA présents sont : AIN (1), DROME (26), ISERE (38), LOIRE (42), HAUTE-LOIRE (43), RHONE (69), SAVOIE (73), HAUTE-SAVOIE (74).

Ensuite, j'ai également supprimé toutes les questions de skills (compétences), elles sont au nombre de 20. Ce sont des questions ouvertes. Elles sont des questions du genre : Avez-vous un diplôme en informatique, lequel ? Qu'est-ce que vous savez faire avec internet ? car les réponses sont multiples et très différentes.

Après cela, je me suis retrouvé avec 262 individus et 35 variables.  
Pour finir cette étape, je vais faire une description rapide des variables.

Identifiant : chaque personne interrogée a un identifiant qui est numérique.  
version : les différentes versions de l'enquête : version 1, version 2, version 3.  
dep : le département de la personne interrogée.  
Agglo : le numéro d'agglomération.  
Sexe : le sexe de la personne interrogée.  
Age : l'âge de la personne interrogée.  
age\_rec : l'âge recodé :

- 1. 18 A 24 ANS**
- 2. 25 A 34 ANS**
- 3. 35 A 49 ANS**
- 4. 50 A 64 ANS**
- 5. 65 ANS ET PLUS**

sexe\_age : c'est une variable où il y a en même temps, le sexe et l'âge recodé.  
prof : la profession de la personne interrogée.

- 1 Agriculteurs exploitants**
- 2 Artisans, commerçants, chefs d'entreprise**
- 3 Cadres et professions intellectuelles supérieures, professions libérales**
- 4 Professions intermédiaires**
- 5 Employés**
- 6 Ouvriers**

**7 Retraité ou pré retraité**

**8 Elève, lycéen ou étudiant**

**9 Sans activité professionnelle (à la recherche d'un premier emploi, homme ou femme au foyer, invalide...)**

prof\_quota :

**1. CODES 1+2**

**2. CODE 3**

**3. CODE 4**

**4. CODE 5**

**5. CODE 6**

**6. CODE 7**

**7. CODES 8+9**

WIP.util.Internet : Avez-vous déjà utilisé Internet pour défendre une cause sur une question politique / sociale / syndicale / environnementale / etc. en

Internaute : les internautes sont classés en deux groupes en fonction de la réponse à la question : A quelle fin, utilisez-vous smartphone : « visite de sites Internet, envoi ou réception d'emails, utilisation d'applications comme WhatsApp, Viber, Skype, Facebook, jouer à des jeux vidéo en ligne

internaute2 : Dépend de la réponse à la question précédente.

WIP.debut.Internet\_mois : Depuis combien de mois utilisez-vous Internet ?

WIP.debut.Internet\_annee : Depuis combien d'années utilisez-vous Internet ?

WIP.impact.pol\_1 : En utilisant Internet, on peut mieux comprendre les questions politiques

WIP.impact.pol\_2 : En utilisant Internet, on peut avoir un plus grand impact politique

WIP.impact.pol\_3 : En utilisant Internet, on peut s'exprimer davantage sur les questions politiques

WIP.impact.pol\_4 : Les responsables politiques prennent en compte l'avis de ceux qui s'expriment sur Internet

Maintenant, je vais vous lire quelques phrases. J'aimerais que vous me disiez dans quelle mesure vous êtes d'accord ou non avec chacune de ces phrases. Utilisez une échelle de 1 à 5, où 1 signifie « pas du tout d'accord » et 5 signifie « tout à fait d'accord ». N'oubliez pas que vous pouvez choisir n'importe quel nombre compris entre 1 et 5.

		1. Pas du tout d'acco rd	2	3	4	5. Tout à fait d'accor d	DK	RF
1	En utilisant Internet, on peut mieux comprendre les questions politiques	1	2	3	4	5	7	9
2	En utilisant Internet, on peut avoir un plus grand impact politique	1	2	3	4	5	7	9
3	En utilisant Internet, on peut s'exprimer davantage sur les questions politiques	1	2	3	4	5	7	9
4	Les responsables politiques prennent en compte l'avis de ceux qui s'expriment sur Internet	1	2	3	4	5	7	9

WIP.expr.pol\_1 : En général (que ce soit en ligne ou hors ligne), vous vous sentez à l'aise pour dire ce que vous pensez sur la politique.

WIP.expr.pol\_2 : Sur Internet, on peut dire sans crainte ses opinions politiques.

WIP.expr.pol\_3 : Les gens devraient pouvoir critiquer librement leur gouvernement sur Internet.

WIP.expr.pol\_4 : Il n'y a PAS de problème à ce que les gens expriment leurs opinions sur Internet, même si celles-ci sont extrêmes.

WIP.expr.pol\_5 : Le gouvernement devrait réglementer davantage Internet qu'il ne le fait aujourd'hui.

		1. Pas du tout d'accord	2	3	4	5. Tout à fait d'accord	DK	RF
1	En général (que ce soit en ligne ou hors ligne), vous vous sentez à l'aise pour dire ce que vous pensez sur la politique.	1	2	3	4	5	7	9
2	Sur Internet, on peut dire sans crainte ses opinions politiques	1	2	3	4	5	7	9
3	Les gens devraient pouvoir critiquer librement leur gouvernement sur Internet.	1	2	3	4	5	7	9
4	Il n'y a PAS de problème à ce que les gens expriment leurs opinions sur Internet, même si celles-ci sont extrêmes.	1	2	3	4	5	7	9
5	Le gouvernement devrait réglementer davantage Internet qu'il ne le fait aujourd'hui	1	2	3	4	5	7	9

WIP.int.fiable : Quelle proportion des informations disponibles sur Internet vous semble fiable ?

**Utilisez une échelle de 1 à 5, où 1 signifie « aucune information n'est fiable » et 5 signifie « toute l'information est fiable ». Vous pouvez choisir n'importe quel nombre compris entre 1 et 5.**

niveau.vie : Quelle description s'approche le plus de l'opinion que vous avez du revenu actuel de votre ménage.

**Vous avez une vie très confortable avec le revenu actuel de votre ménage**

**Vous avez une vie confortable**

**Vous vous en sortez**

**Vous trouvez la vie difficile**

**Vous trouvez la vie très difficile avec le revenu actuel de votre ménage**

DK

Etude : Quel est votre niveau d'études ?

**Niveau collège ou élémentaire**

**CAP, BEP**

**Bac, Brevet professionnel ou équivalent**

**Supérieur court (Bac +1/+2)**

**Supérieur long (supérieur ou égal à Bac +3)**

DK

RF

WIP.statut.fam : Etes-vous

**Célibataire**

**Marié**

**En concubinage**

**Pacsé**

**Divorcé/séparé**

**Veuf**

DK

RF

nb.foyer : Combien de personnes composent votre foyer, y compris vous-même ?

nb.mineur.foyer : Combien y-a-t-il de personnes de moins de 18 ans ?

nb.ado.foyer : Si enfants de moins de 18 ans] Et parmi elles, combien ont de 11 à 18 ans ?

WIP.revenu :

Moins de 1400 euros	1
Entre 1400 et 2000 euros	2
Entre 2000 et 2900 euros	3
Entre 2900 et 4000 euros	4
Plus de 4000 euros	5
(DK)	8
(REF)	9

Actu : A quelle fréquence vous informez-vous sur l'actualité (tous médias confondus : tv, radio presse, Internet, etc.) ?

Ecrit : Dans le cadre de votre travail ou de votre vie personnelle, vous arrive-t-il de rédiger des textes, des comptes-rendus, des notes de synthèses, des courriers, des textes personnels, etc. (que ce soit sur papier ou sur ordinateur) ?

**Plusieurs fois par jour**

**Tous les jours ou presque**

**Toutes les semaines ou presque**

**Tous les mois ou presque**

**Moins souvent**

**Jamais**

**DK**

Poids : Le poids de la personne interrogée.

2<sup>ème</sup> étape : J'ai recodé les variables quantitatives en type factor vu que dans une ACM toutes les données doivent être qualitatives. Les variables quantitatives sont : identifiant, age, dep, aggro, nb.foyer, poids.

3<sup>ème</sup> étape : Problématique

J'essaye de voir les similitudes qu'il y a dans les différentes catégories de professions par rapport aux réponses sur la question de la fiabilité de l'internet et également s'il y a des similitudes entre le statut familial et le niveau de vie.

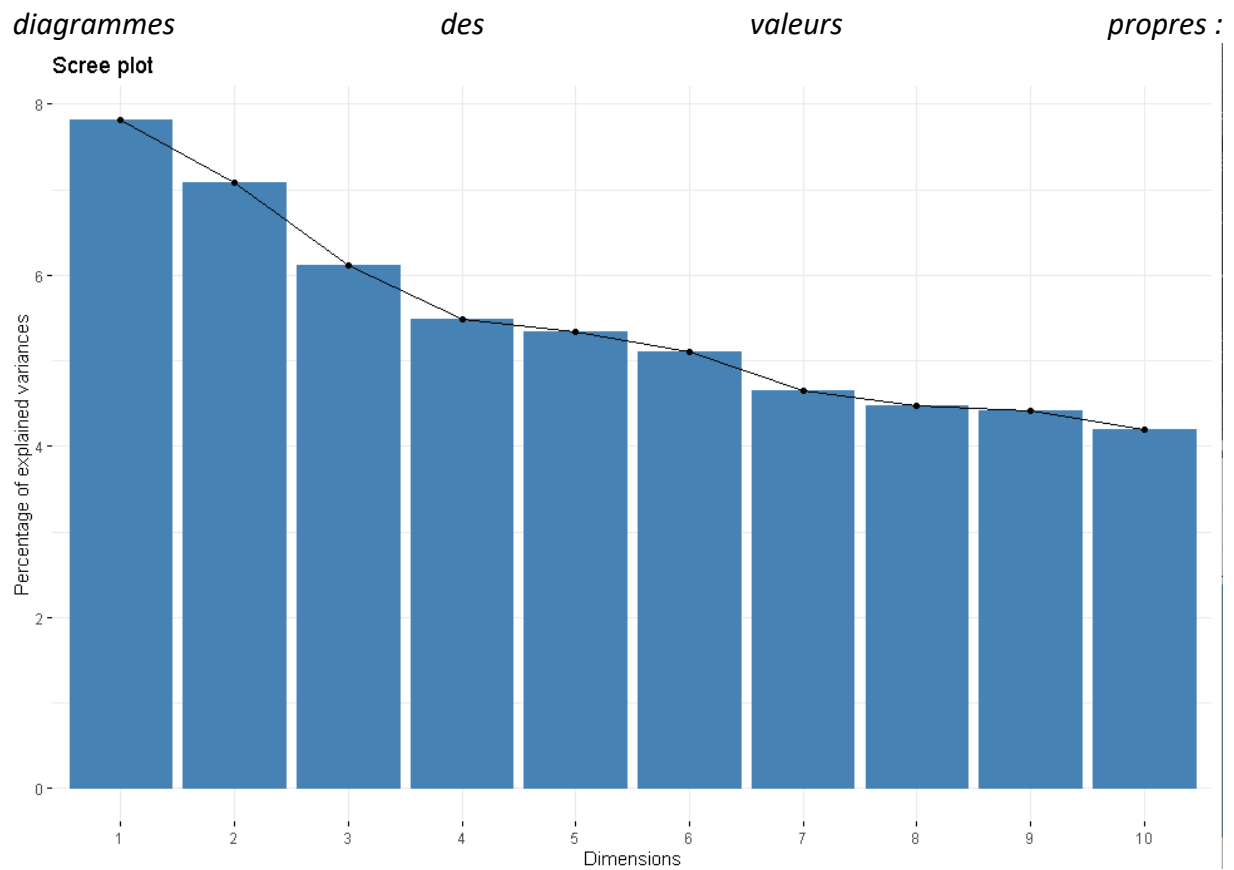
4<sup>ème</sup> étape : choix des variables actives

J'ai choisi des variables en rapport à ma problématique. J'ai pris les variables actives suivantes : Prof, Etude, WIP.int.fiable, WIP.statut.fam, niveau.vie.

## 5<sup>ème</sup> étape : Construction des axes factoriels

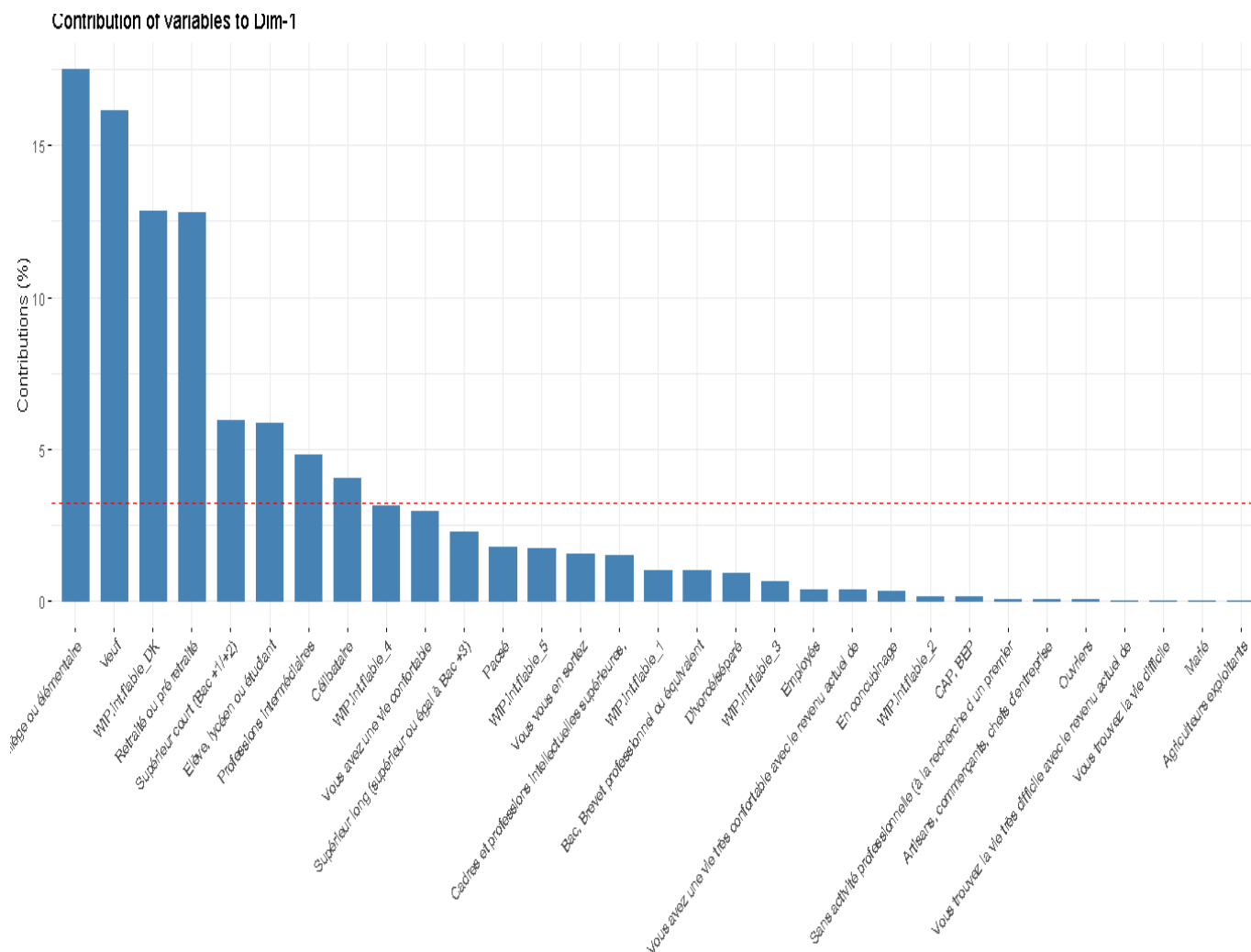
Valeurs propres :

eigenvalue	variance	percent	cumulative	variance	percent
0.406597592838976	7.81918447767261	7.81918447767261			
0.368630653673634	7.08905103218527	14.9082355098579			
0.31818742935225	6.11898902600481	21.0272245358627			
0.28534020865852	5.48731170497155	26.5145362408342			
0.27798122339103	5.34579275751981	31.8603289983541			
0.26533850890497	5.10266363278788	36.9629926311419			
0.242016628411312	4.65416593098676	41.6171585621287			
0.232804936389953	4.47701800749909	46.0941765696278			
0.229990166414425	4.42288781566202	50.5170643852898			
0.217896140620277	4.19031039654379	54.7073747818336			
0.206710679814123	3.97520538104083	58.6825801628744			
0.191940665552823	3.69116664524659	62.373746808121			
0.189859219352826	3.65113883370819	66.0248856418292			
0.184659657607499	3.55114726168267	69.5760329035119			
0.174195505184759	3.34991356124536	72.9259464647572			
0.169425654328698	3.25818566016727	76.1841321249245			
0.156041056339126	3.00078954498319	79.1849216699077			
0.153626642792806	2.95435851524626	82.139280185154			
0.141219292985136	2.71575563432954	84.8550358194835			
0.13925487216383	2.67797831084288	87.5330141303264			
0.138315712003998	2.65991753853843	90.1929316688648			
0.126565600285631	2.43395385164676	92.6268855205116			
0.109381970737048	2.10349943725092	94.7303849577625			
0.10261258764648	1.97331899320155	96.703703950964			
0.0907703490895148	1.74558363633682	98.4492875873009			
0.0806370454603554	1.55071241269914	100			

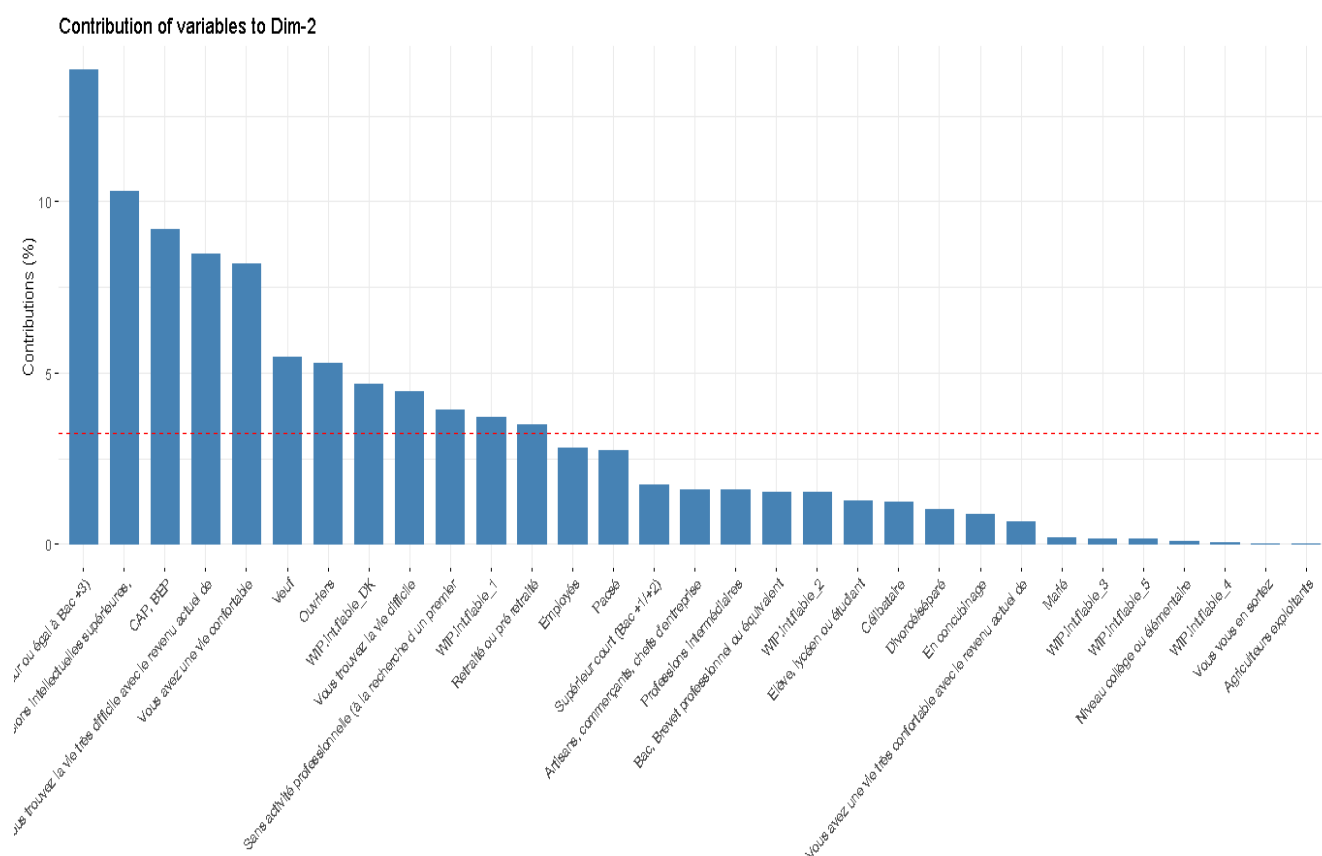


Les valeurs propres mesurent la quantité de variance expliquée par chaque axe principal. Elles sont grandes pour les premiers axes et petites pour les axes suivants. La proportion de variance expliquée par chaque valeur propre est donnée dans la deuxième colonne, par exemple 7,81 % de la variation est expliquée par la première valeur propre.

## contributions des variables initiales :



les catégories collège ou élémentaire, veuf, WIP\_int. fiable\_DK et Retraité ou Pré ou pré retraits sont les plus importantes dans la définition de la première dimension.

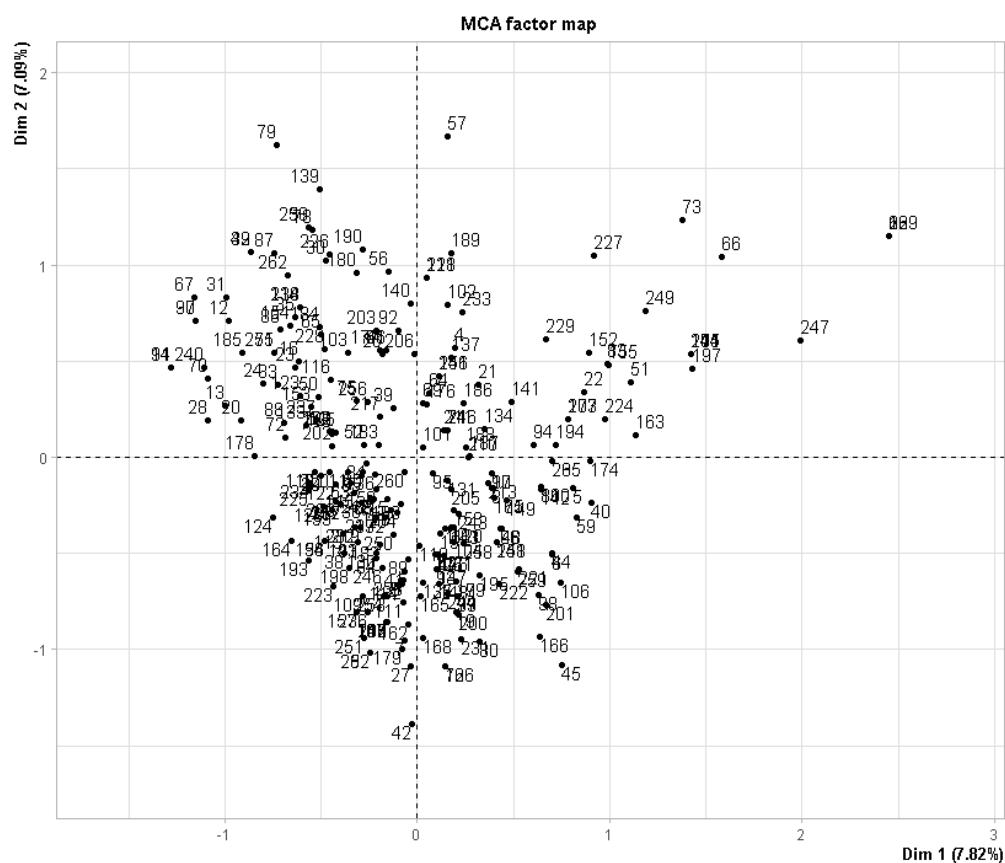


*Les catégories Supérieur long (supérieur ou égal à Bac +3), Cadres et professions intellectuelles supérieures, CAP, BEP et Vous trouvez la vie très difficile avec le revenu actuel sont les plus importantes dans la définition de la deuxième dimension.*



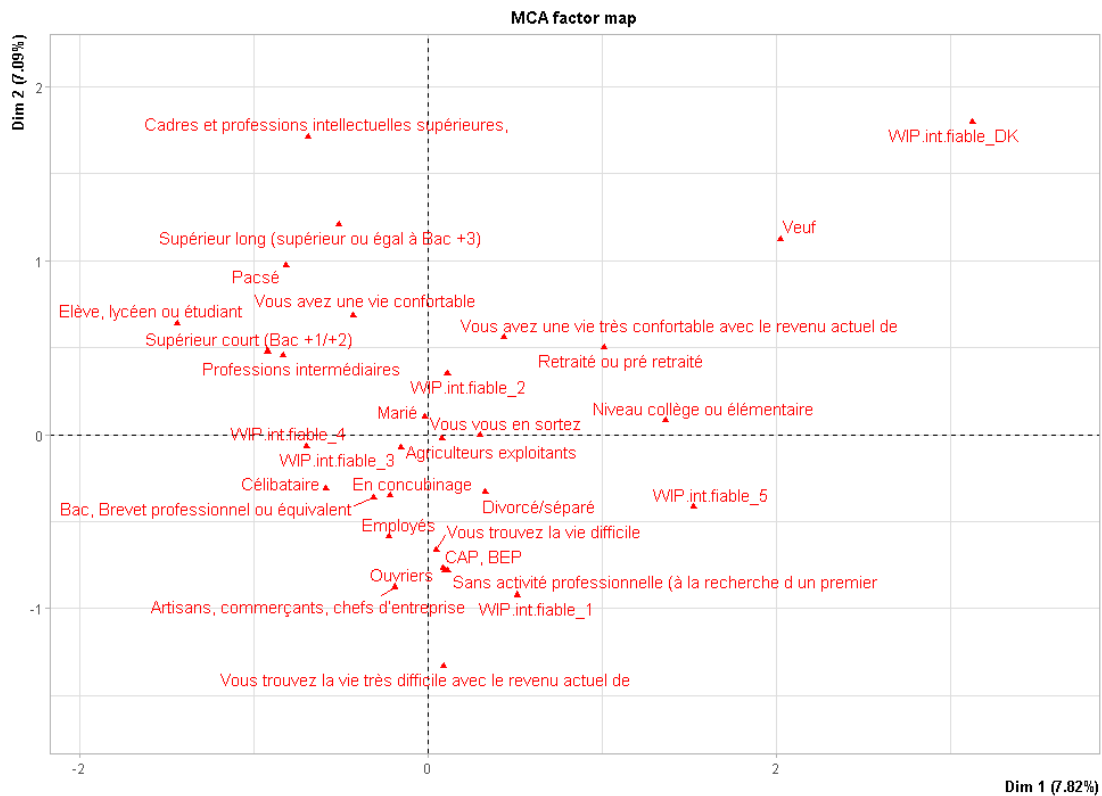
## 6ème étape : Représentation graphique, cartographie

### Représentation des individus :



Le graphe n'est pas très lisible vu que c'est un traitement d'enquête de 262 individus. Mais tous les individus proches ont les mêmes profils de réponses aux questions.

Représentation des variables :



Ici, on peut voir :

Les variables positivement corrélées sont regroupées.

Les variables négativement corrélées sont positionnées sur les côtés opposés de l'origine du graphique (quadrants opposés).

On peut dire par exemple dans cette enquête, que les personnes divorcées ou séparées trouvent la vie difficile avec leur revenu actuel.

Les cadres et professionnelles intellectuelles supérieures et les professeurs intermédiaires ne trouvent pas toute l'information sur internet fiable.

Les individus sans activité professionnelle, ont en général le CAP ou le BEP...

On peut dire plein d'autres choses à partir de ce graphe.

7<sup>ème</sup> étape : Typologie des individus par méthode de classification (automatique) ; caractérisation des classes

Dans cette partie, j'ai choisi de classer les individus en fonction du poids, âge, aggro, identifiant, nb. foyer et dep.

Le code utilisé est dans le script R.

