

# TP5

Ehoarn Surace

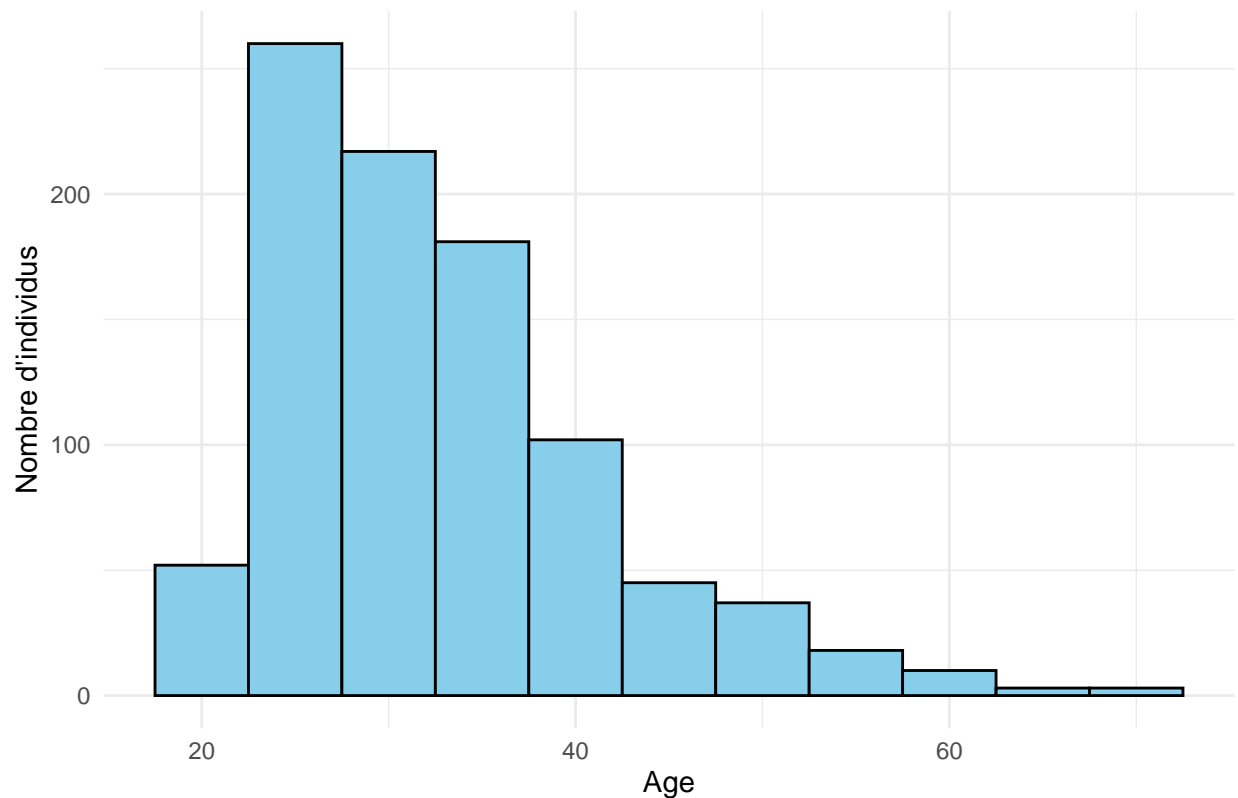
2025-09-29

## 1) Description des variables :

Ce jeu de données représente le choix de langage de programmation d'un individu classé selon plusieurs de ses critères personnels. Le jeu de données contient 932 lignes et 7 colonnes. Chaque ligne représente un individu qui a répondu au questionnaire. La première colonne représente le sexe de la personne. Il y a 3 possibilités. La seconde représente l'âge de la personne. La troisième colonne représente le statut de l'employé. La quatrième variable représente les études. La cinquième variable représente le secteur d'activité de sa profession. La sixième variable représente le choix de l'individu au questionnaire : Matlab, Python, R ou SAS. La dernière variable représente le domaine d'activité de la personne.

Des personnes ont 0 an. Nous décidons de ne pas conserver ces individus dans notre analyse car elle pourrait la fausser. Au vu du nombre de réponses à notre questionnaire, cela n'aura que très peu d'impact.

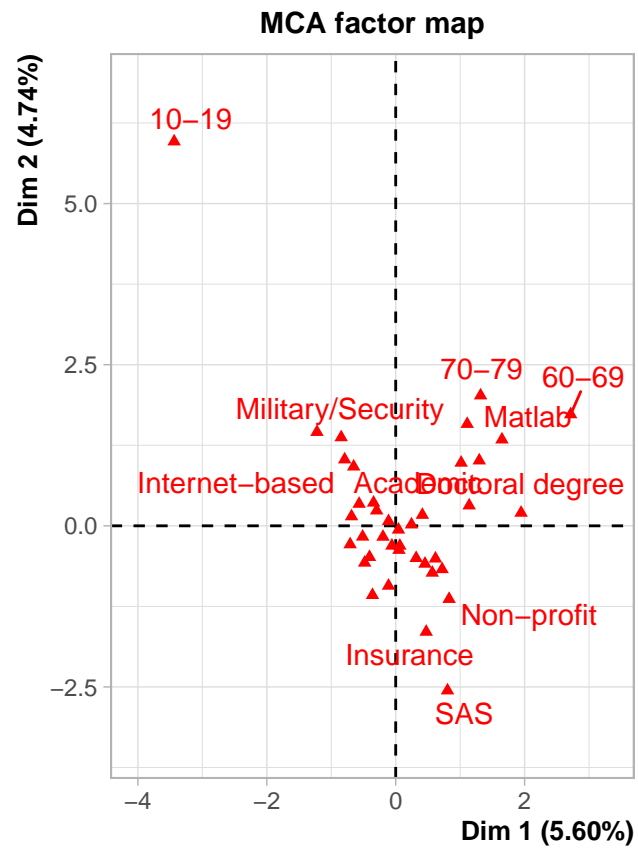
Histogramme des âges (>0 ans)

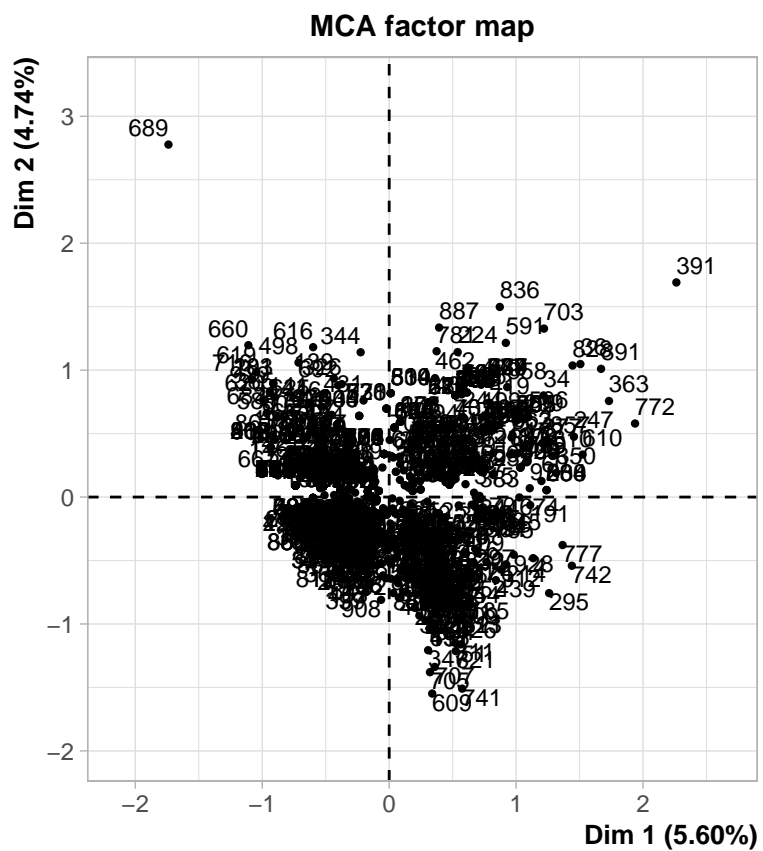


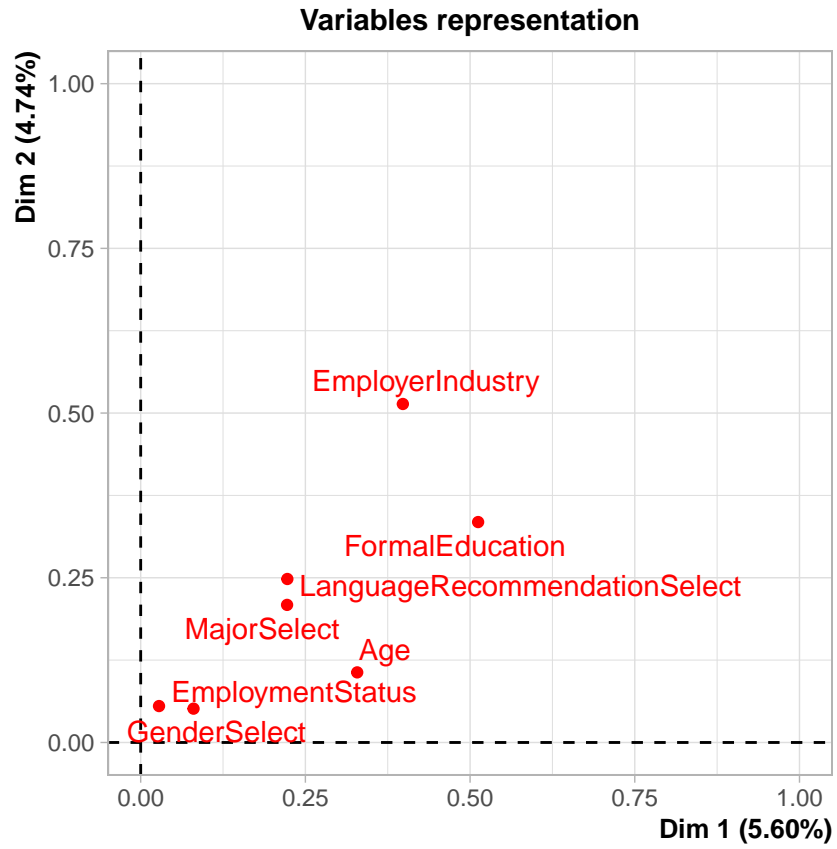
## Réalisation de l'ACM

On crée des classes d'âges pour la variable âge

```
## Warning: ggrepel: 28 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

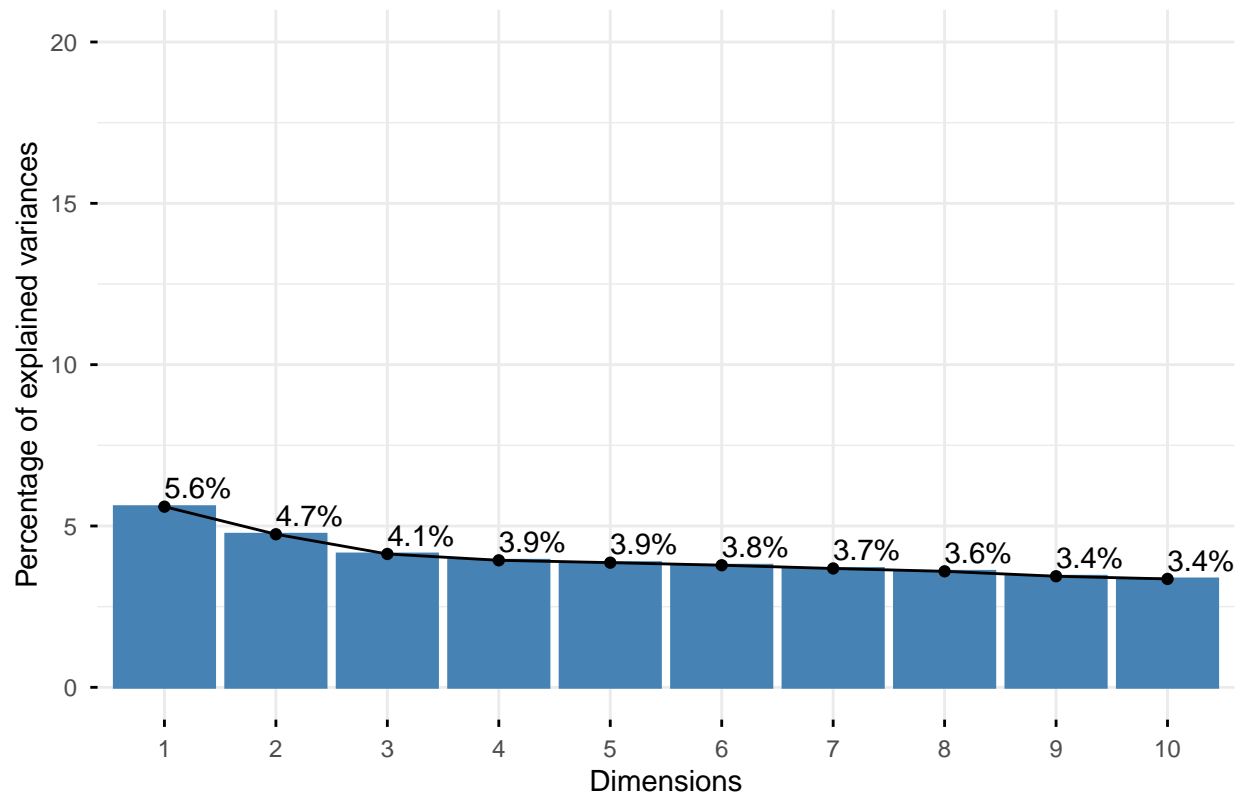






## Représentation des valeurs propres :

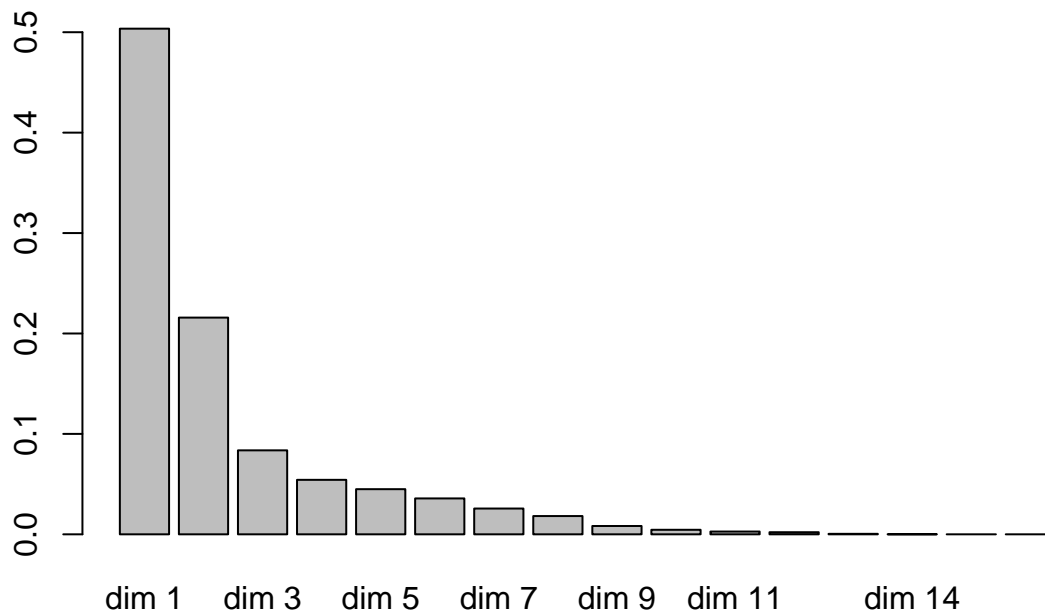
## Représentation des valeurs propres



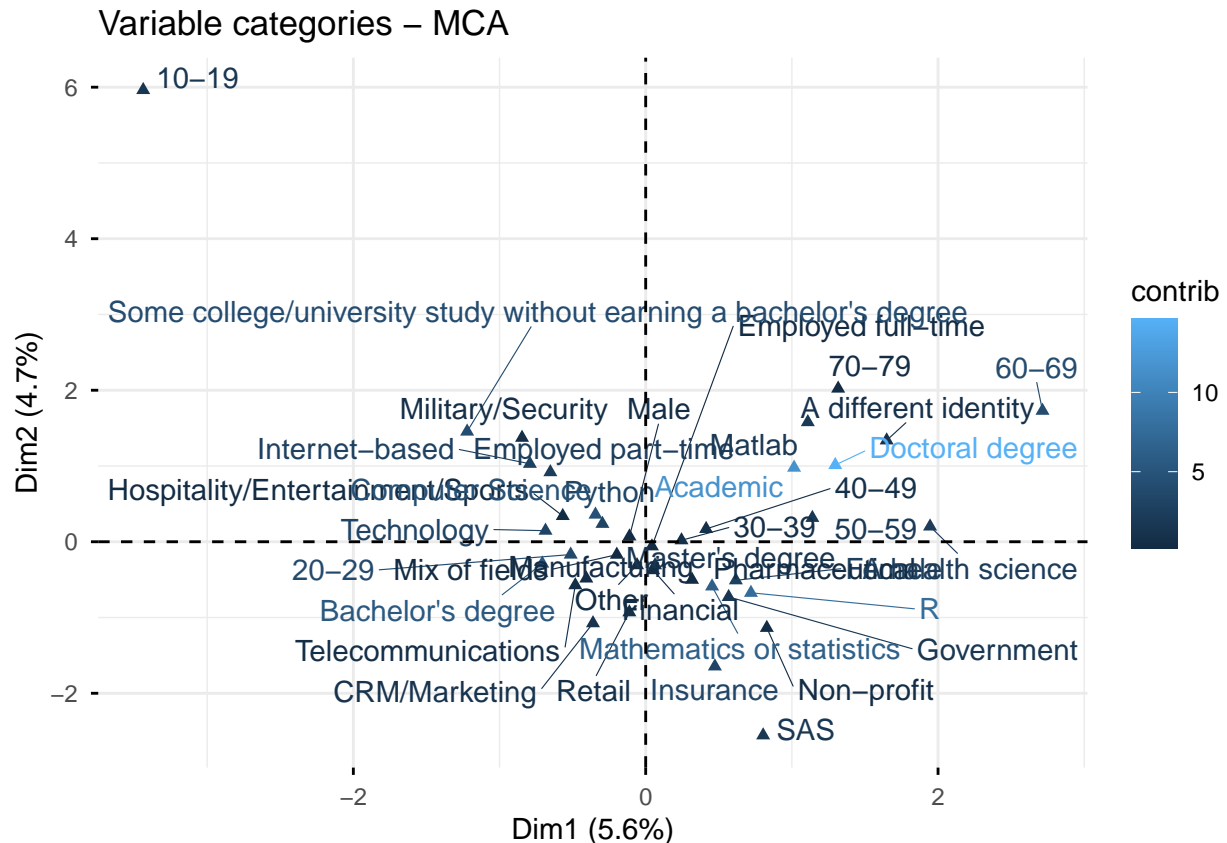
##	eigenvalue	percentage of variance	cumulative percentage of variance
## dim 1	0.25595357	5.598984	5.598984
## dim 2	0.21689568	4.744593	10.343577
## dim 3	0.18894292	4.133126	14.476704
## dim 4	0.17997743	3.937006	18.413710
## dim 5	0.17665477	3.864323	22.278033
## dim 6	0.17298884	3.784131	26.062164
## dim 7	0.16840151	3.683783	29.745947
## dim 8	0.16435516	3.595269	33.341216
## dim 9	0.15736147	3.442282	36.783498
## dim 10	0.15358085	3.359581	40.143079
## dim 11	0.15124129	3.308403	43.451482
## dim 12	0.15013568	3.284218	46.735700
## dim 13	0.14638624	3.202199	49.937900
## dim 14	0.14426227	3.155737	53.093637
## dim 15	0.14330736	3.134848	56.228485
## dim 16	0.14287816	3.125460	59.353945
## dim 17	0.14208051	3.108011	62.461956
## dim 18	0.14067949	3.077364	65.539320
## dim 19	0.13656371	2.987331	68.526651
## dim 20	0.13269223	2.902643	71.429294
## dim 21	0.13129897	2.872165	74.301459
## dim 22	0.12801106	2.800242	77.101701
## dim 23	0.12186158	2.665722	79.767423
## dim 24	0.11836020	2.589129	82.356552
## dim 25	0.11788074	2.578641	84.935193

## dim 26	0.11602565	2.538061	87.473254
## dim 27	0.11134444	2.435660	89.908914
## dim 28	0.10546356	2.307015	92.215929
## dim 29	0.10260479	2.244480	94.460409
## dim 30	0.10112621	2.212136	96.672545
## dim 31	0.08943398	1.956368	98.628913
## dim 32	0.06267825	1.371087	100.000000

Application de Benzécri :



Il faut sélectionner 3 dimensions.



Dimension 1 (5.6% de variance) Cet axe oppose deux profils professionnels et éducatifs : Côté gauche (valeurs négatives) :

Secteurs techniques et spécialisés : Technology, Telecommunications, Hospitality/Entertainment Formation : Bachelor's degree Domaines : Computer Science, Internet-based

Côté droit (valeurs positives) :

Secteurs tertiaires diversifiés : Healthcare, Finance, Government, Insurance, Retail Niveaux d'éducation variés : Master's degree, Doctoral degree Domaines académiques divers : Mathematics/Statistics, autres disciplines

Dimension 2 (4.7% de variance) - Axe vertical Cet axe structure l'échelle d'éducation et de qualification : Partie supérieure (valeurs positives) :

Éducation incomplète ou partielle : "Some college/university study without earning a bachelor's degree" Âge jeune : 10-19 ans Statut : Employed part-time

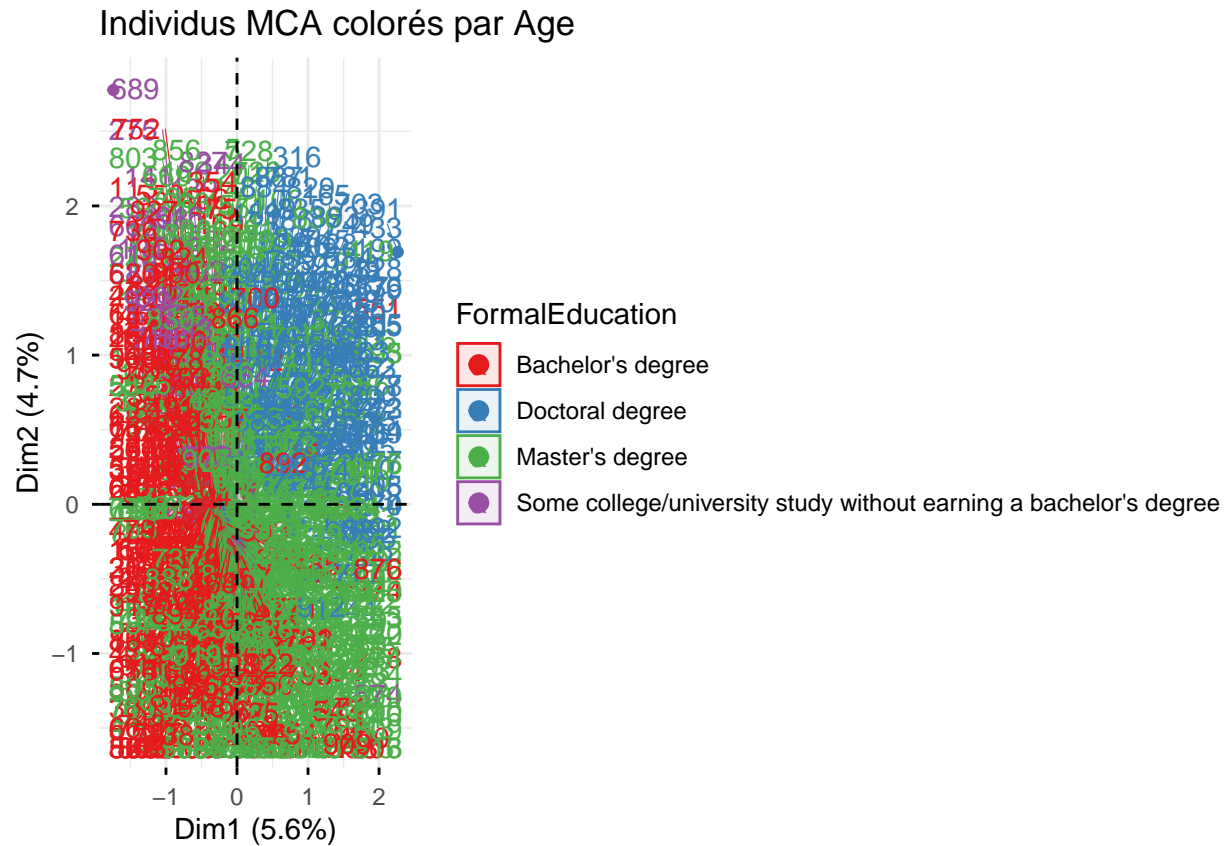
Partie inférieure (valeurs négatives) :

Secteurs spécialisés : Pharmaceutical, SAS, Insurance, CRM/Marketing Disciplines avancées : Mathematics/Statistics

Groupes de profils identifiables

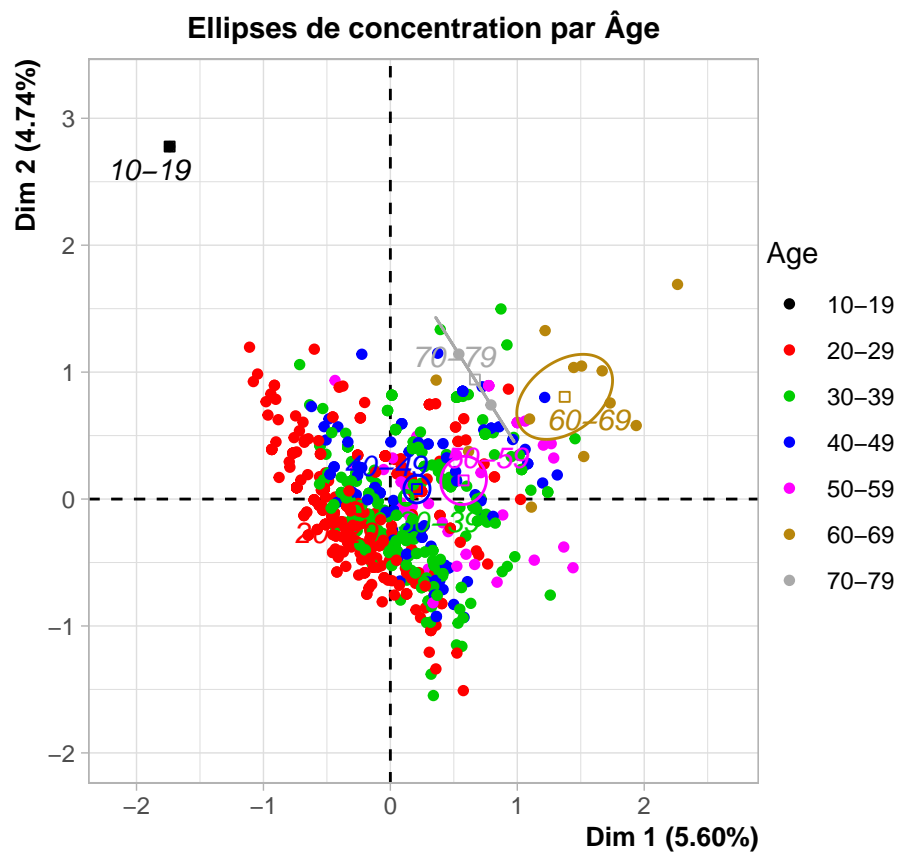
Profil technique-technologique (gauche) : jeunes professionnels en tech avec Bachelor's degree Profil académique avancé (droite-haut) : hautes qualifications (Master's/Doctoral) dans des secteurs variés Profil en formation (haut-centre) : étudiants en cours d'études, emploi partiel Profil spécialisé (bas) : secteurs nécessitant expertise pointue (pharma, finance, statistiques)

## Analyse des corrélations :



```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## i The deprecated feature was likely used in the FactoMineR package.
## Please report the issue at <https://github.com/husson/FactoMineR/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```





## NULL

## NULL