

Rapport final - Sujet 8

Kenzée Aboustait, Buani Doki Engoué, Emma Le Franc, Vincent Reau, Tom Roquand, Maëlys Tieha

2023-03-31

Contents

Introduction	1
Présentation des données	1
Présentation du sujet	1
Les données	2
Import	2
Refactorisation	2
Observation des données manquantes	2
Analyse descriptive univariée	3
Analyse descriptive bivariée	4
Profils type	5
Le premier plan factoriel	10
CSP	13
Description des dimensions	13
Classification hiérarchique ascendante (CAH) avec l’acp	15
Description des clusters selon les variables	18
Explication des variables selon les clusters	19
Conclusion	21

Introduction

Présentation des données

Dans le cadre de ce projet, les données sont fournies dans un fichier *scooter.xls*. Dans cette table de données sont représentés 416 individus autour de 25 variables. Ces variables concernent les individus, leur scooter et l’importance portée à certaines caractéristiques de celui-ci. Les 6 premières variables (qualitatives) portent sur les caractéristiques de l’individu et son scooter, les suivantes (quantitatives) sont des notes attribuées par l’individu à l’importance de certaines caractéristiques de leur véhicule.

Présentation du sujet

En partant de ce jeu de données, il faut retrouver les différents profils “type” d’utilisateurs qui peuvent se dégager. Il faudra ensuite relever les différences entre les hommes et les femmes. Enfin nous déduirons quel type de scooter conviendrait le mieux à un homme, à une femme.

Les données

Import

```
##      sexe      âge      CSP      type-cylindrée
## Length:416 Length:416 Length:416 Length:416
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## type-utilisation critère-esthétique note-satisfaction
## Length:416 Length:416 Length:416
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
```

Refactorisation

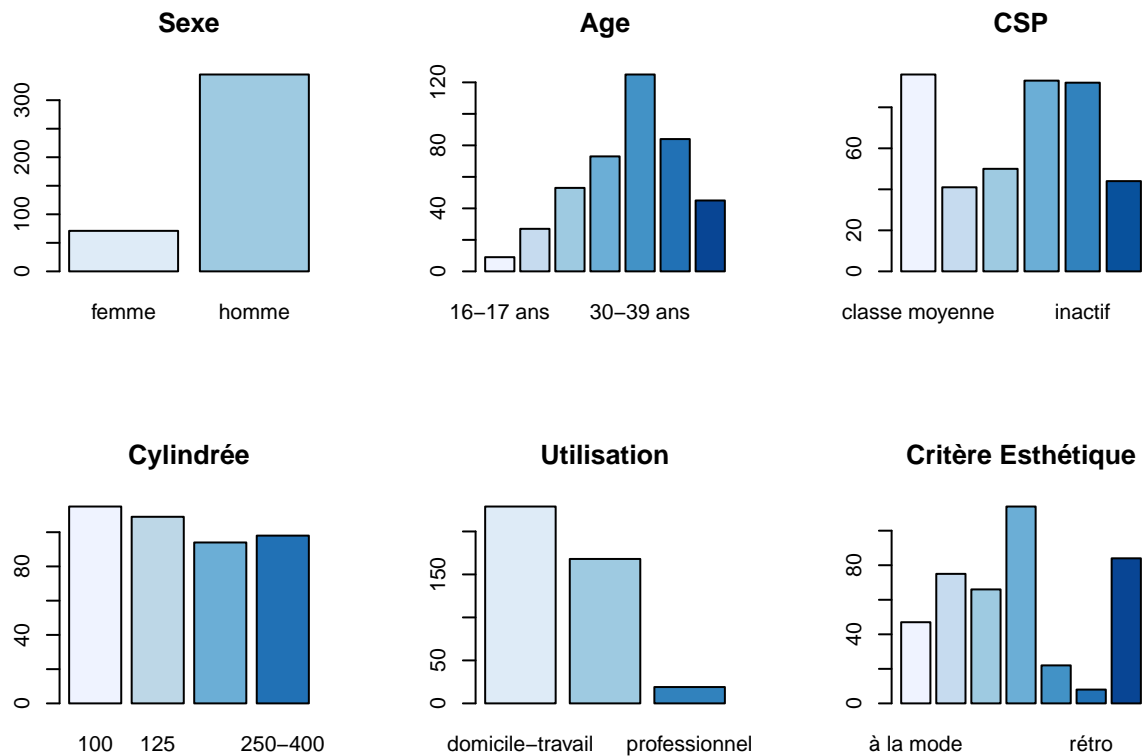
Comme expliqué en introduction, les 6 premières colonnes représentent des variables qualitatives. On remarque ici qu'elles sont lues comme des caractères, il faut donc utiliser la méthode **as.factor()** pour les étudier. La note de satisfaction qui est quantitative est également lue comme des caractères, on utilisera donc la méthode **as.numeric()**

Observation des données manquantes

```
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

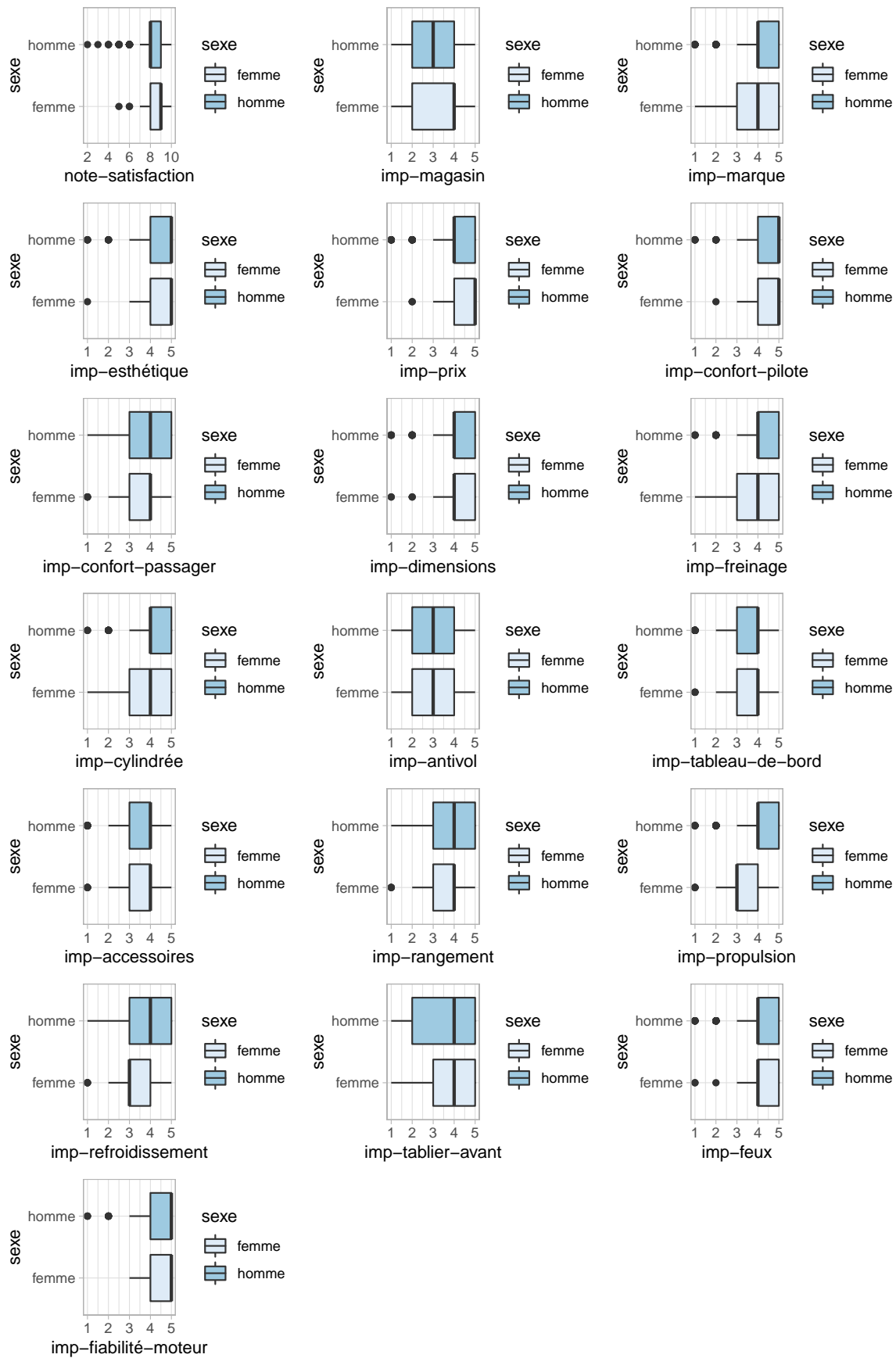
Cette table de donnée ne contient aucune donnée manquante. Il ne sera donc pas nécessaire d'imputer de valeurs dans notre dataframe.

Analyse descriptive univariée



On remarque que seulement 71 femmes sont représentées dans cette table de données (pour 345 hommes). Une grande majorité des individus se situent entre 25 et 50 ans. Les classes les plus représentées sont la classe moyenne, les employés et les inactifs. Une majorité utilise son scooter de manière quotidienne pour se rendre au travail. Aussi l'âge semble normalement distribué.

Analyse descriptive bivariable



On remarque avec seulement ces boîtes à moustache des différences de critères entre les hommes et les femmes. Effectivement la marque, le freinage, la propulsion, le refroidissement semblent plus importants pour les hommes, là où les femmes n'ont pas de critères qui ressortent. On peut relever dans ces statistiques le stéréotype des hommes qui accordent beaucoup d'importance aux caractéristiques techniques de leur véhicule et des femmes qui ne s'y intéressent pas.

Profils type

ANOVA sur les données qualitatives

Nous décidons de séparer les données qualitatives des données quantitatives afin de les étudier plus précisément. On regroupera ensuite les résultats obtenus entre l'ACP réalisée par les autres membres et l'ANOVA.

Récupérations des données qualitatives Pour réaliser une ANOVA, les données doivent respecter certains critères, tout d'abord la normalité des résidus (Shapiro test), ensuite l'homosédasticité (bartlett test), et l'indépendance qui est déjà vérifiée sur la matrice de corrélation.

Création du modèle

Réalisation des tests préliminaires à l'ANOVA

```
##
##  Shapiro-Wilk normality test
##
## data:  scooter_quali$`note-satisfaction`
## W = 0.89216, p-value < 2.2e-16

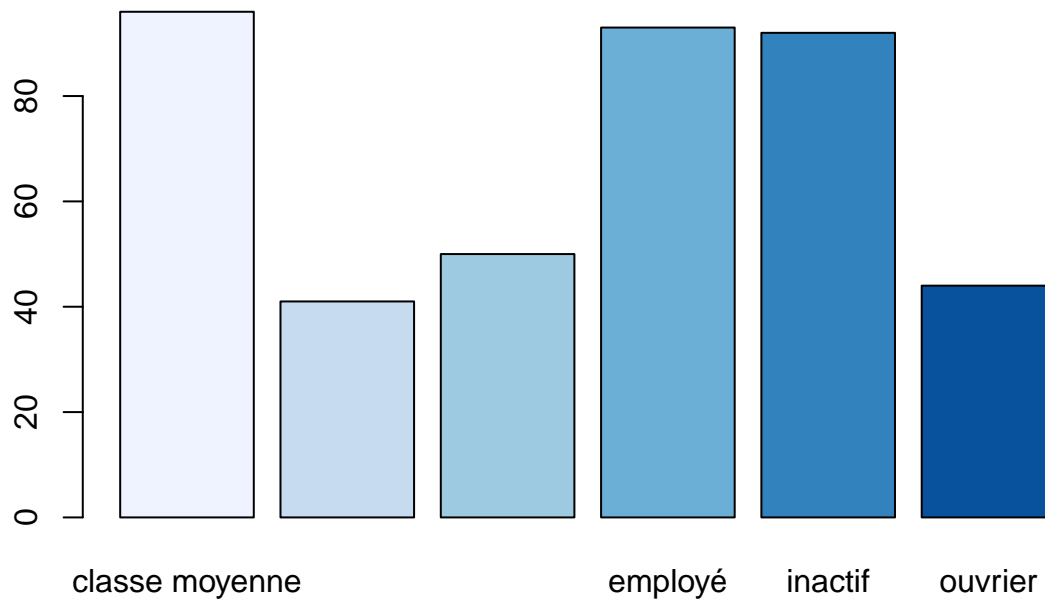
##
##  Shapiro-Wilk normality test
##
## data:  residuals(modele_note)
## W = 0.96174, p-value = 6.123e-09
```

On obtient ici une p-value largement inférieure à 0.05. La distribution n'est donc pas normale. L'hypothèse de la normalité de la distribution et des résidus est rejetée, on devra donc réaliser un test de Kruskal-Wallis non-paramétrique. Nous n'avons donc plus besoin de vérifier l'homogénéité des variances.

```
## [1] 0.0001596721
## [1] 0.3270871
## [1] 9.176094e-06
## [1] 5.306715e-05
## [1] 4.993413e-05
## [1] 0.0004221408
```

On remarque ici que les variables sont toutes représentatives sauf la CSP qui ne semble pas jouer tant que ça sur la note de satisfaction. Concernant les autres variables cela semble en effet logique.

On peut étudier de plus près la variable des catégories socio-professionnelles :



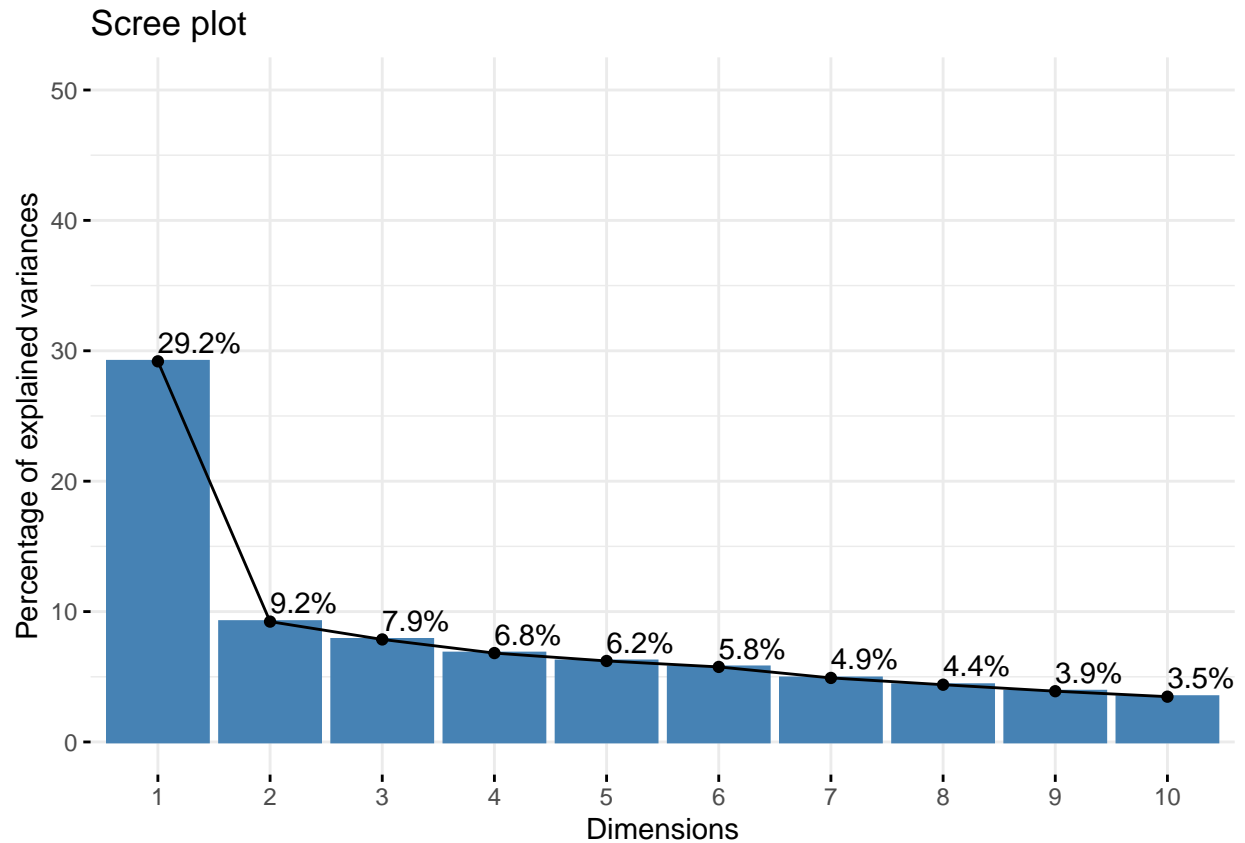
Il existe des différences de représentation en termes de quantités mais elles ne suffisent pas à expliquer les résultats obtenus. On pourrait vérifier les liens entre ces catégories par des tests d'égalité des moyennes comme t.test ou Wilcox test.

L'ACP des 416 évaluations de scooter

```
##      imp-prix imp-antivol
##           3           9
```

selection du nombre d'axes

Nous avons décidé de garder 10 axes pour obtenir plus de 80% de l'information délivré par nos données. Nous n'interpréterons que les deux premiers axes car les autres ne révèlent pas des structures très pertinentes dans nos données.

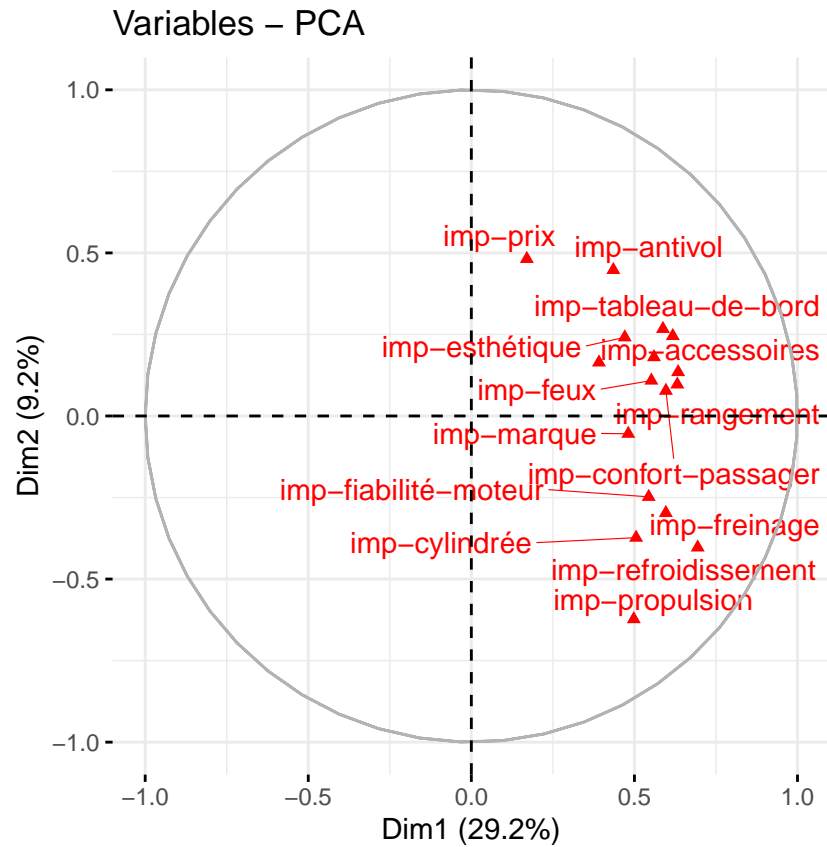


Le cercle de corrélation des variables quantitatives

On remarque toutes les variables sont positivement corrélés sur la première dimension 1.

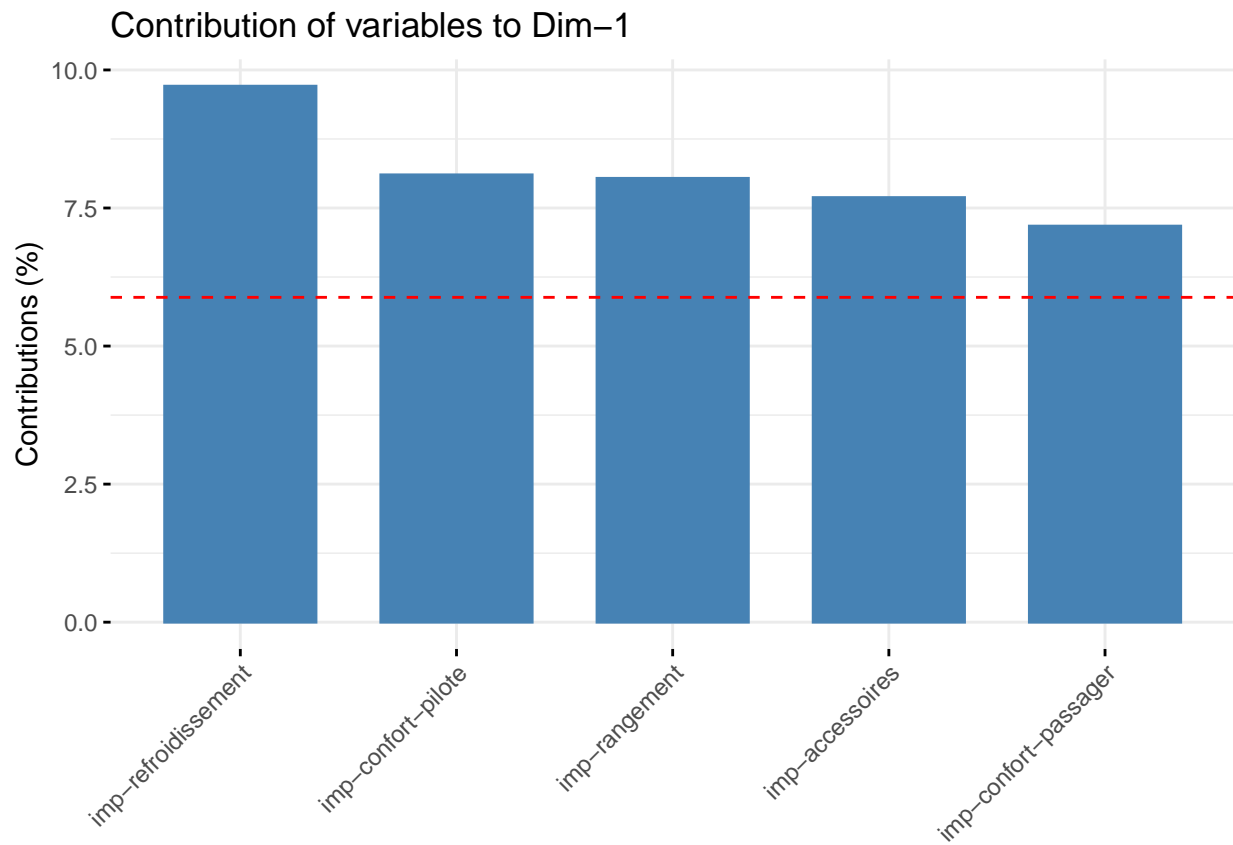
Ce qui n'est pas le cas sur la deuxième. Les variables imp-freinage imp-cylindrée , et imp-propulsion sont respectivement opposée aux variables imp-magasin, imp-marque et imp-esthétique .

```
## Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

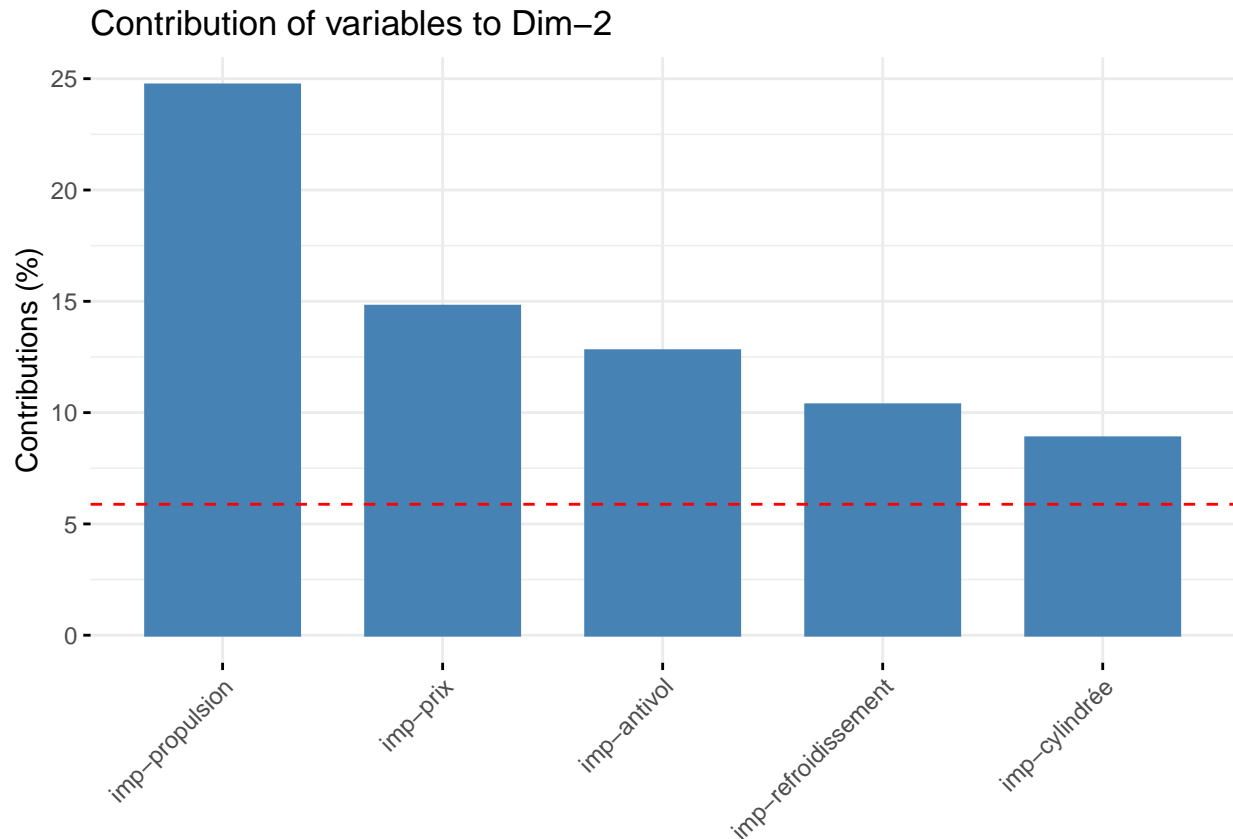


Les variables les plus contributives

Les variables qui contribuent le plus à la formation de l'axe 1 sont les notes sur le refroidissement, le confort du pilote, le rangement, les accessoires et le confort passager.



Les variables qui contribuent le plus à la formation de la deuxième dimension sont les notes sur la propulsion, le prix, le refroidissement, le cylindrée et l'antivol.



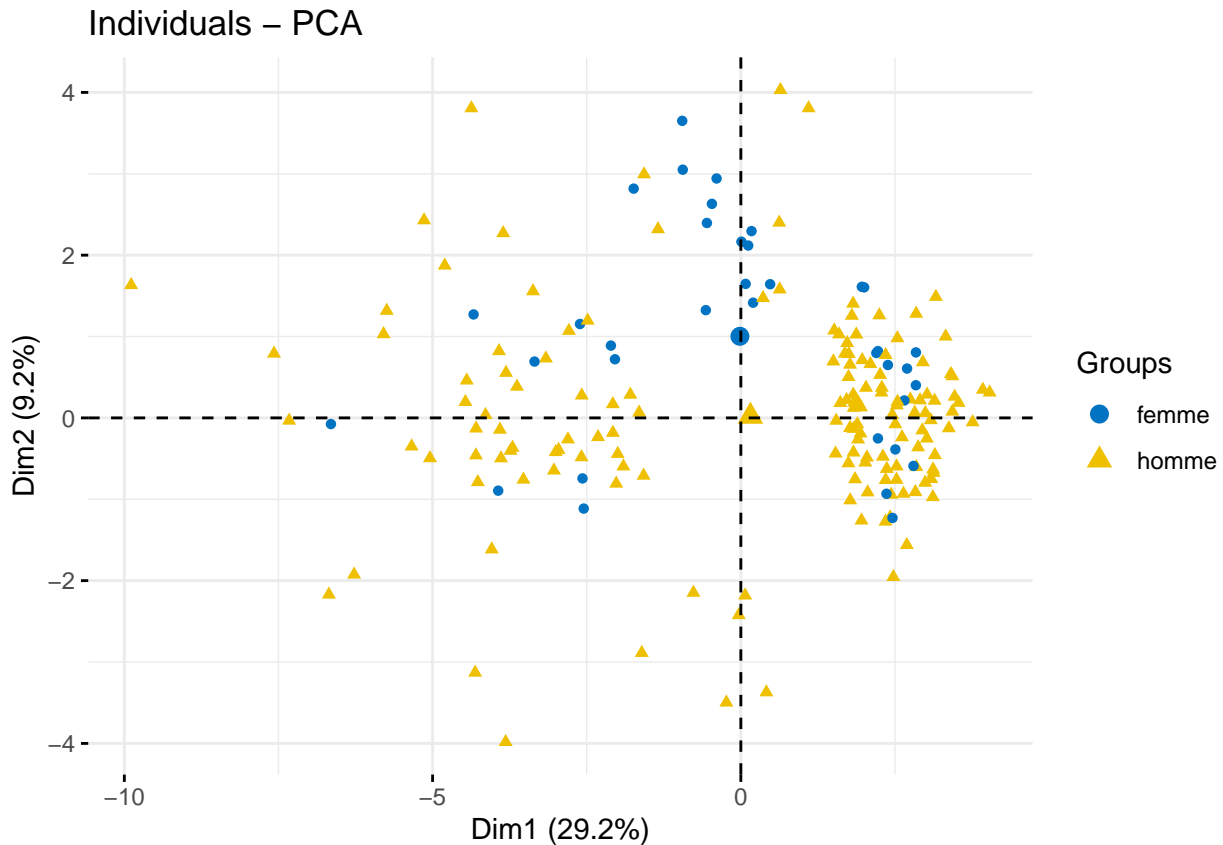
Les individus

Le but est de trouver des profils types des clients . Nous essayerons d'illustrer les choix des utilisateurs selon plusieurs caractéristiques . Comme l'âge , la CSP et le sexe.

Le premier plan factoriel

Critère du sexe

Ce graphique du premier plan factoriel présente les individus qui ont un cosinus carré supérieur à 0.3. Les individus qui ont de fortes valeurs positives sur la première dimension sont des hommes à l'opposé de la deuxième ou ceux qui ont une forte valeur positive sont des femmes. Ce sont hypothétiquement des hommes qui ont donné des notes élevées en termes de imp-marque, mp-esthétique, imp-confort-pilote .



Notre hypothèse est confirmée car lorsque l'on regarde de plus dans les données ces individus en moyenne donnent des notes supérieures à 4 sur 6 pour la marque, l'esthétique et le confort du pilote.

Tableau des moyennes des notes pour la marque , l'esthétique et le confort du pilote

##	imp-marque	imp-esthétique	imp-confort-pilote
##	4.004808	4.391827	4.341346

Tableau des moyennes des notes pour la marques , l'esthétique , le confort pour les idnividus ayant des coordonnées positives sur la dimension 1

##	imp-marque	imp-esthétique	imp-confort-pilote
##	4.514563	4.747573	4.844660

Inversement ceux possédant une forte valeurs positives donnent des notes inférieures à la moyennes générale pour ces mêmes critères.

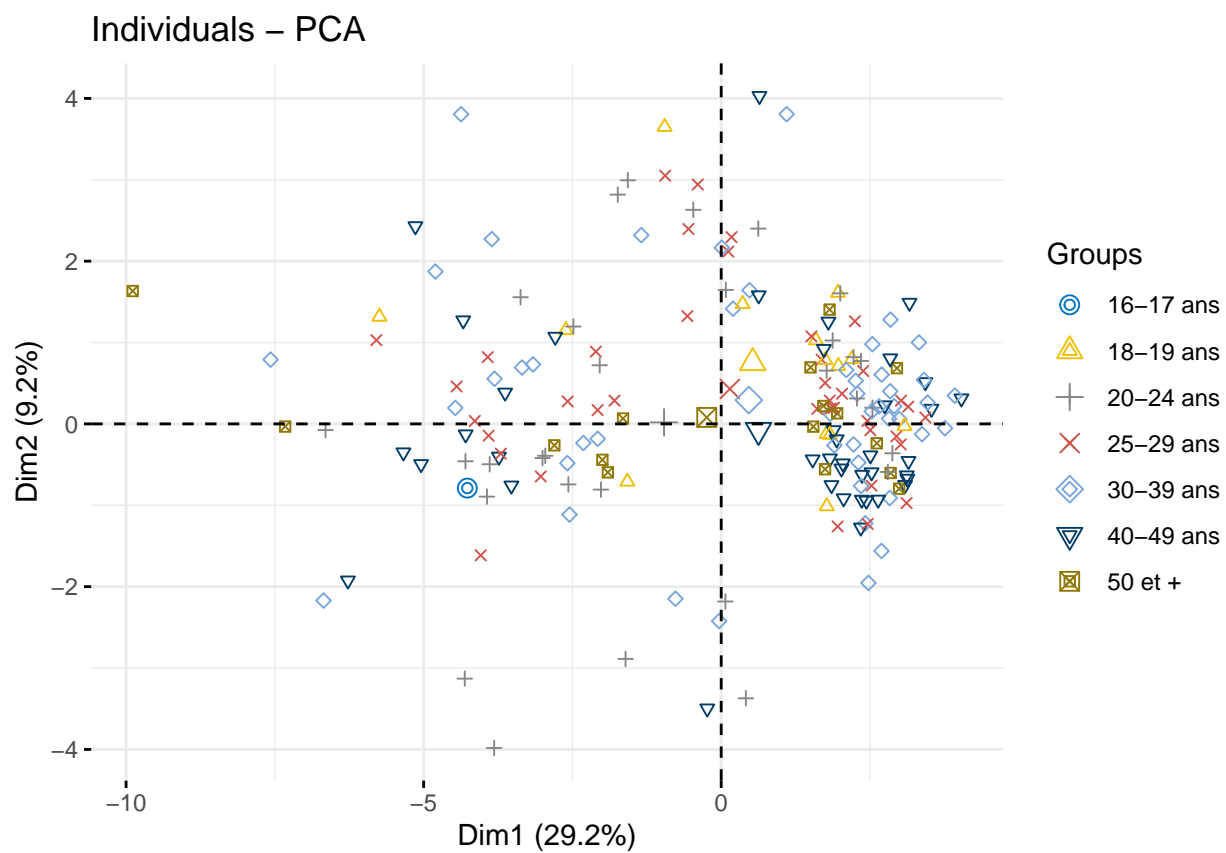
Tableau des moyennes des notes pour la marques , l'esthétique , le confort pour les idnividus ayant des coordonnées négatives sur la dimension 2

##	imp-marque	imp-esthétique	imp-confort-pilote
##	3.280702	3.666667	3.421053

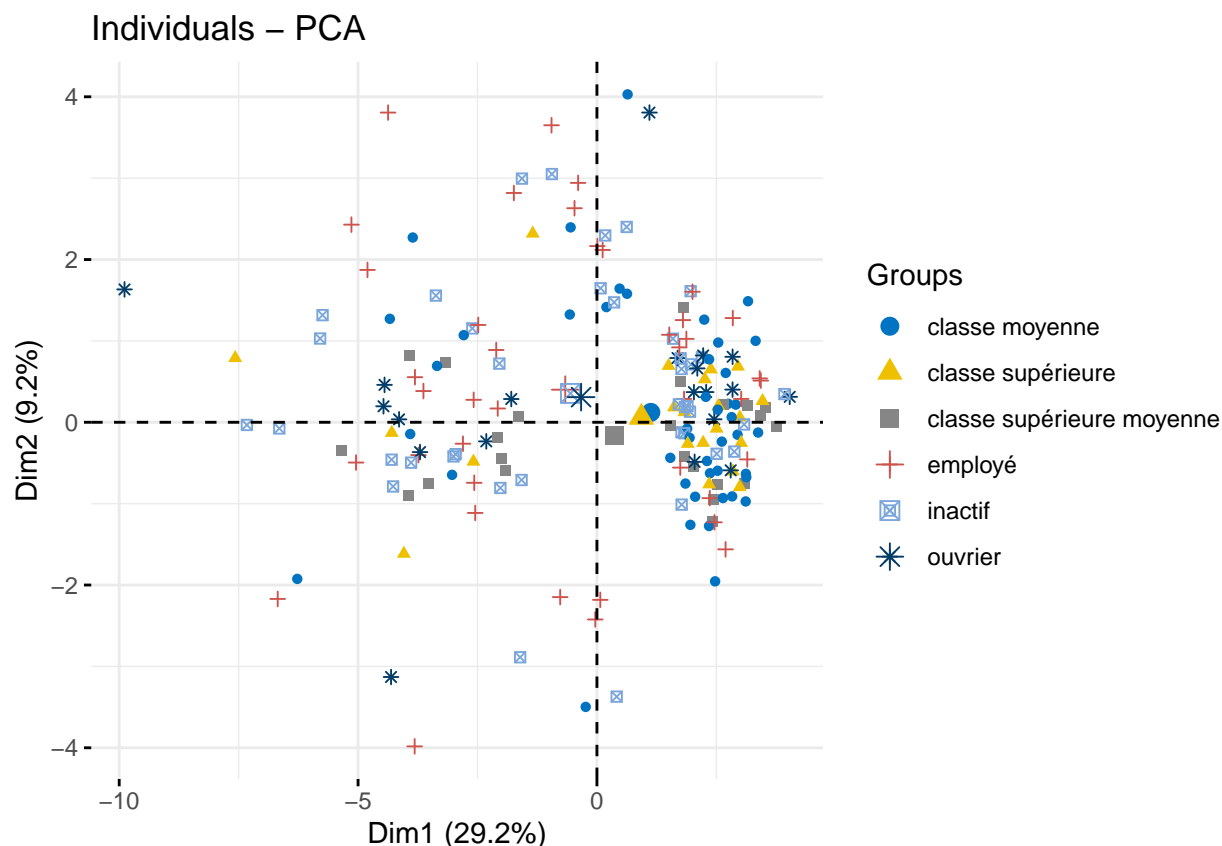
Critère de l'âge

Le chargement a nécessité le package : shiny

Le chargement a nécessité le package : FactoInvestigate



CSP



Description des dimensions

Le plan du cylindrée

Le cylindrée est la variable qui permet le mieux de distinguer les individus sur le plan composé de la dimension 1 et 2. **La dimension 1**

La dimension 1 oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) est caractérisé par des individus satisfaits.

Le groupe 2 (caractérisés par une coordonnée négative sur l'axe) est caractérisé par des individus qui ne sont pas satisfaits par le confort des passagers, le rangement, le confort du pilote, le refroidissement, les feux, le freinage, la fiabilité du moteur et les dimensions du scooter (de la plus extrême à la moins extrême). **Ce sont majoritairement des femmes qui ont entre 16 et 17 ans et 20 et 24 ans qui aiment les scooters sobres et rétro.**

Le groupe 3 (caractérisés par certaines positives sur l'axe) est caractérisé par des individus satisfaits par les qualités de tableau de bord et le confort du pilote pour le scooter. Ce sont majoritairement des hommes entre 20-24ans, 25-29ans, 30-39ans et qui sont inactifs. **Ce sont majoritairement des femmes qui ont entre 16 et 17 ans et 20 et 24 ans qui aiment les scooters sobres et rétro.**

La dimension 2

La dimension 2 oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

de faibles valeurs pour des variables telles que imp-refroidissement, imp-accessoires, imp-confort-pilote, imp-freinage, imp-rangement, imp-confort-passager, imp-fiabilité-moteur, imp-tableau-de-bord, imp-cylindrée et imp-propulsion (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage :

de fortes valeurs pour les variables imp-magasin, imp-esthétique, imp-accessoires et imp-tableau-de-bord (de la plus extrême à la moins extrême)

de faibles valeurs pour les variables imp-propulsion, imp-refroidissement, imp-freinage et imp-cylindrée (de la plus extrême à la moins extrême).

Les individus de ces 2 groupes sont des femmes qui préfèrent les scooter à la mode avec des cylindrée de type 100 et qui ont entre 18 et 19 ans ou entre 25 et 29 ans.

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

de faibles valeurs pour des variables telles que imp-prix, imp-antivol, imp-magasin, imp-tableau-de-bord, imp-confort-passager, imp-accessoires, imp-tablier-avant, imp-rangement, imp-feux et imp-esthétique (de la plus extrême à la moins extrême).

Le groupe 4 (caractérisés par une coordonnées négative sur l'axe) partage :

de fortes valeurs pour des variables telles que imp-refroidissement, imp-freinage, imp-confort-passager, imp-accessoires, imp-propulsion, imp-rangement, imp-tableau-de-bord, imp-fiabilité-moteur, imp-cylindrée et imp-confort-pilote (de la plus extrême à la moins extrême).

Les individus de ces 2 groupes sont des hommes qui préfèrent les scooter innovants, qui ressemblent à des motos et qui ont un gros cylindrée mais qui sont sobre et discret . Ce sont des hommes qui sont de classe supérieur ou moyenne .

La dimension 3

C'est l'esthétisme qui permet de mieux séparer nos variables sur cet axes.

La dimension 3 oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

de fortes valeurs pour les variables imp-tablier-avant, imp-accessoires, imp-antivol, imp-feux, imp-rangement et imp-refroidissement (de la plus extrême à la moins extrême). de faibles valeurs pour les variables imp-esthétique, imp-marque, imp-cylindrée et imp-magasin (de la plus extrême à la moins extrême).

Ce sont des individus qui préfèrent les scooter sobre, classique et qui ont entre 49 et 50 ans.

Le groupe 2 (caractérisés par une coordonnées négative sur l'axe) partage :

de fortes valeurs pour la variable imp-tableau-de-bord. de faibles valeurs pour les variables imp-prix, imp-feux, imp-magasin, imp-fiabilité-moteur, imp-dimensions et imp-tablier-avant (de la plus extrême à la moins extrême). Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

de fortes valeurs pour les variables imp-magasin, imp-prix, imp-dimensions, imp-esthétique, imp-marque, imp-fiabilité-moteur, imp-cylindrée et imp-feux (de la plus extrême à la moins extrême). de faibles valeurs pour les variables imp-accessoires, imp-tablier-avant, imp-antivol et imp-tableau-de-bord (de la plus extrême à la moins extrême).

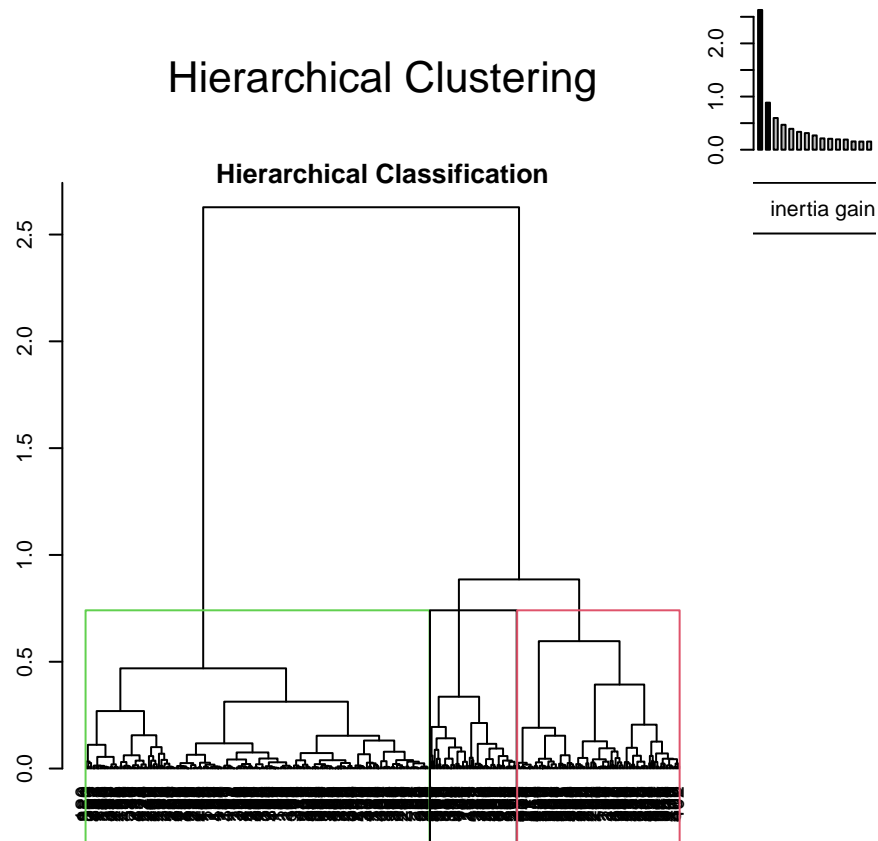
Ce sont des individus qui préfèrent les scooter rétro , qui ressemblent à des moto, ils ont entre 20 et 24 ans.

Synthèses et établissement de profils clients

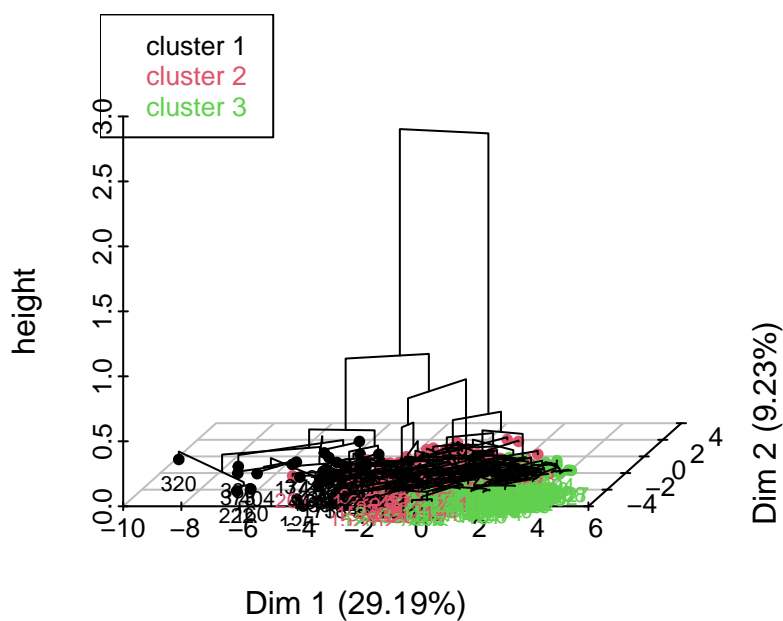
Ainsi l'acp nous a permis d'identifier plusieurs profils clients:

- Des hommes qui ont entre 18-19 ans 20-25 ans, qui aiment les cylindrées de type 250-400 , qui appartiennent à la classe moyenne.
- Des femmes qui ont entre 16-17 ans ou 20-24 ans qui aiment les scooters sobres et rétro.
- Des femmes qui préfèrent les scooter à la mode avec des cylindrée de type 100 et qui ont entre 18 et 19 ans ou entre 25 et 29 ans.
- Des hommes qui préfèrent les scooter innovants, qui ressemblent à des motos , qui ont un gros cylindre mais qui sont sobre et discret . Ce sont des hommes qui sont de classe supérieur ou moyenne .
- Des personnes qui aiment les scooter sobres, classiques et qui ont entre 49 et 50 ans.
- Des individus qui préfèrent les scooter rétro , qui ressemblent à des moto, qui ont entre 20 et 24 ans.**

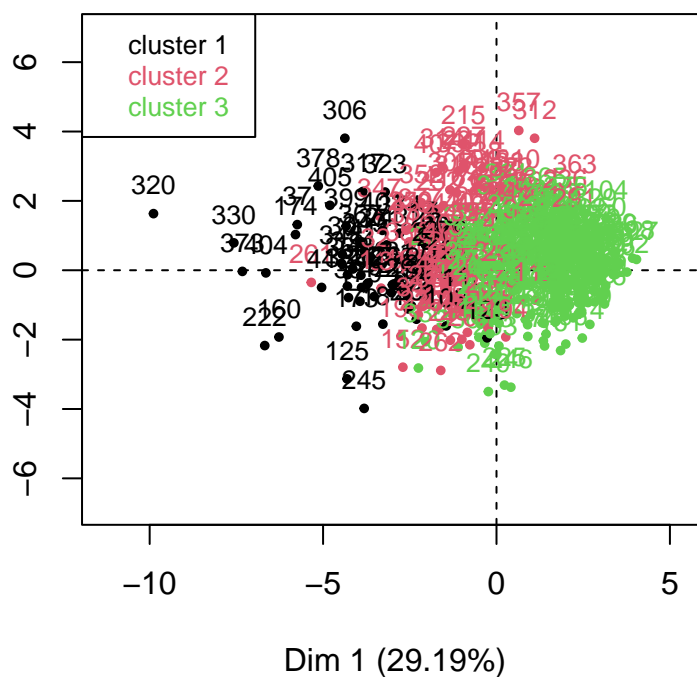
Classification hiérarchique ascendante (CAH) avec l'acp



Hierarchical clustering on the factor map



Factor map



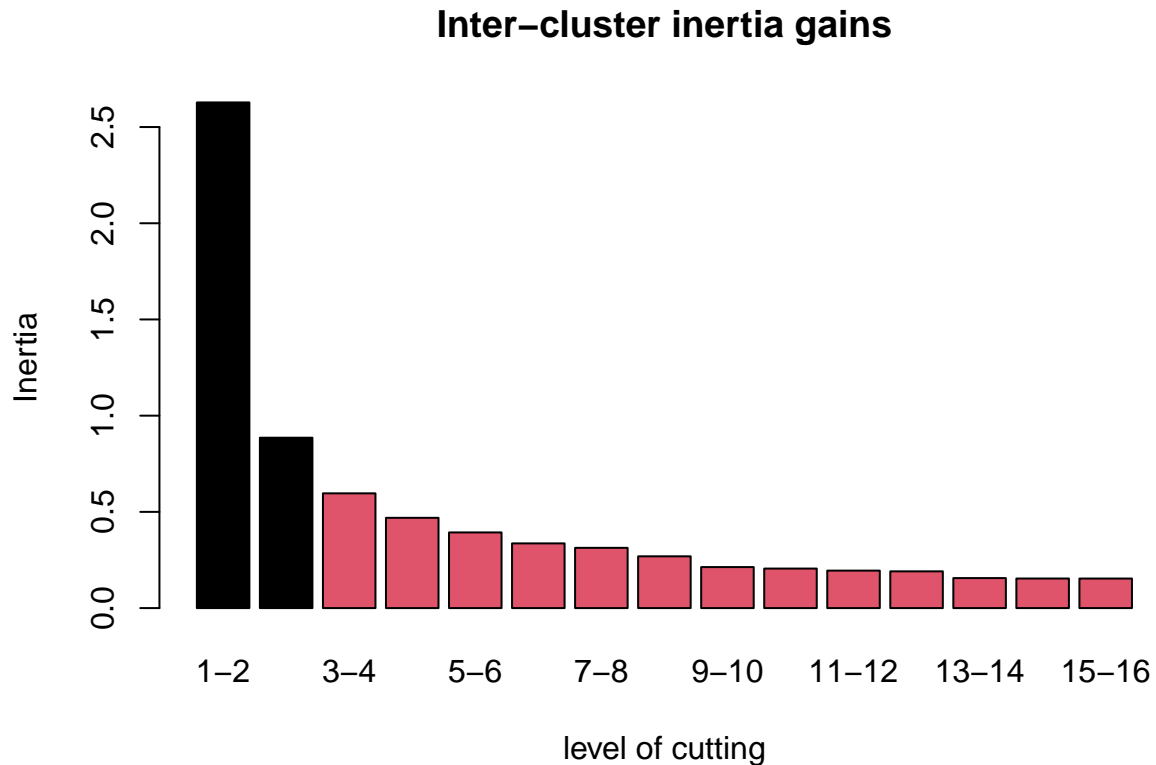
```
## $`1`
##
## Cla/Mod   Mod/Cla   Global   p.value
## CSP=employé      22.580645  34.426230  22.35577  0.019331931
## type-cylindrée=type-cylindrée_100  20.869565  39.344262  27.64423  0.032413995
## type-cylindrée=type-cylindrée_250-400  6.122449   9.836066  23.55769  0.003984272
##
## v.test
## CSP=employé      2.339068
```



```

## type-cylindrée=type-cylindrée_100      2.139268
## type-cylindrée=type-cylindrée_250-400 -2.879404
##
## $`2`
##
## Cla/Mod  Mod/Cla  Global    p.value
## type-cylindrée=type-cylindrée_100 38.26087 38.59649 27.64423 0.002739337
## critère-esthétique=à la mode 44.68085 18.42105 11.29808 0.007243528
## sexe=femme 39.43662 24.56140 17.06731 0.015755576
## sexe=homme 24.92754 75.43860 82.93269 0.015755576
## type-cylindrée=type-cylindrée_150-200 17.02128 14.03509 22.59615 0.008807407
##
## v.test
## type-cylindrée=type-cylindrée_100 2.995569
## critère-esthétique=à la mode 2.685436
## sexe=femme 2.414529
## sexe=homme -2.414529
## type-cylindrée=type-cylindrée_150-200 -2.619441
##
## $`3`
##
## Cla/Mod  Mod/Cla  Global    p.value
## type-cylindrée=type-cylindrée_250-400 73.46939 29.875519 23.55769 3.151920e-04
## sexe=homme 61.15942 87.551867 82.93269 3.757729e-03
## type-cylindrée=type-cylindrée_150-200 70.21277 27.385892 22.59615 5.865350e-03
## âge=50 et + 73.33333 13.692946 10.81731 2.592162e-02
## critère-esthétique=innovant nouveau 66.66667 31.535270 27.40385 2.658795e-02
## critère-esthétique=à la mode 42.55319 8.298755 11.29808 2.588047e-02
## CSP=employé 47.31183 18.257261 22.35577 1.987286e-02
## âge=20-24 ans 41.50943 9.128631 12.74038 1.077441e-02
## sexe=femme 42.25352 12.448133 17.06731 3.757729e-03
## type-cylindrée=type-cylindrée_100 40.86957 19.502075 27.64423 1.583558e-05
##
## v.test
## type-cylindrée=type-cylindrée_250-400 3.602483
## sexe=homme 2.897814
## type-cylindrée=type-cylindrée_150-200 2.755215
## âge=50 et + 2.227384
## critère-esthétique=innovant nouveau 2.217514
## critère-esthétique=à la mode -2.228001
## CSP=employé -2.328740
## âge=20-24 ans -2.549932
## sexe=femme -2.897814
## type-cylindrée=type-cylindrée_100 -4.316733

```



On cherche à minimiser l'inertie inter-classe. Sur cet histogramme, on observe 1 coude significatif entre 1 et 2. Le choix le plus judicieux est donc de faire répartir nos données en 2 classes.

Description des clusters selon les variables

Description du cluster 1

D'après ces résultats, la classe n°1 est plutôt caractérisée par une population de femmes, jeunes et des scooters à type de cylindre petit. On pourrait aussi ajouter une tendance au niveau du critère esthétique qui serait plutôt classique et à la mode. De plus, dans cette classe, il y a très peu d'hommes, de scooters à type de cylindre gros, et dont le critère esthétique est innovant.

Expliquons plus en détail ces résultats :

Dans la classe n°1, il y a une surreprésentation des individus dont le type de cylindre choisi est 100, puisque Mod/Cla (34.46%) > Global (27.64%). De la même façon, il y a une surreprésentation des individus de sexe féminin, puisque Mod/Cla (22.59%) > Global (17.06%). Pour la modalité age = 20-24 ans, cela est aussi le cas puisque Mod/Cla (16.94%) > Global (12.74%). A l'inverse, il y a une sous représentation des individus dont le type de cylindre choisi est 250-400, puisque Mod/Cla (18.64%) < Global (23.55%). Idem pour les modalités sexe masculine et critère esthétique est innovant avec respectivement Mod/Cla (77.40%) < Global (82.93) et Mod/Cla (22.03%) < Global (27.40%).

Logiquement, la classe 2 serait caractérisée par les variables qui sont très faiblement représentées dans la classe n°1 soit la population d'hommes, de scooters à type de cylindre gros, et de le critère esthétique est innovant.

Description du cluster 2

Il y a une surreprésentation des hommes dans le cluster 2, il sont de 82,93% dans la population et contre 87,03% dans le cluster. Il y a une surreprésentation de la modalité "innovant nouveau" 27.40% contre 31,38% dans notre cluster. Ensuite on remarque une surreprésentation des cylindres entre 250 et 400 (23,56 % dans la population globale contre 27,19% dans notre cluster). Toutefois, on constate une sous représentation des individus qui ont entre 20 et 24 ans, des femmes et des cylindrée 100 respectivement (12,74% dans la

population globale contre 9,62% dans notre cluster).17,06% dans la population globale contre 12,97% dans notre cluster ,27,64% dans la population globale contre 22,59% dans notre cluster . Ainsi les individus du cluster 2 sont majoritairement des hommes qui préfèrent les scooters modernes et des cylindres assez puissants entre 250 et 400.

Les 2 profils types qui ressortent donc ici sont les suivants : population de femmes, jeune et des scooters à type de cylindre petit population d'hommes, à type de cylindre gros et au critère esthétique nouveau.

Il serait alors intéressants de proposer des scooters avec de petits cylindres à des femmes plutôt jeunes et de puissants scooter dernier cris aux hommes .

Explication des variables selon les clusters

Nous avons réalisé une CAH avec seulement deux clusters et nous nous sommes aperçu que toutes les variables quantitatives sont significatives mais également que celles ayant un fort impact positif dans l'un des deux clusters ont un fort impact négatif dans l'autre cluster. Le v-test était inversement proportionnel entre les deux clusters. Nous avons alors décidé d'étudier le cas des 3 clusters pour avoir des cas plus intéressants.

##	v.test	Mean in category	Overall mean	sd in category
## imp-tablier-avant	-2.343777	3.032787	3.411058	1.3666666
## imp-magasin	-2.839133	2.786885	3.235577	1.3564026
## note-satisfaction	-4.053331	7.590164	8.199519	1.4069738
## imp-propulsion	-5.477924	3.196721	3.901442	1.1848729
## imp-marque	-5.545803	3.327869	4.004808	1.1695776
## imp-confort-passager	-7.065641	2.672131	3.634615	1.1554761
## imp-feux	-7.104308	3.196721	4.052885	1.1281738
## imp-accessoires	-7.485085	2.377049	3.456731	0.9778520
## imp-fiabilité-moteur	-7.619074	3.786885	4.468750	0.9601017
## imp-cylindrée	-7.689747	3.426230	4.218750	0.9993279
## imp-dimensions	-8.156671	3.098361	4.045673	1.1970581
## imp-antivol	-8.418620	1.721311	3.045673	0.9256127
## imp-esthétique	-8.433343	3.573770	4.391827	1.2474106
## imp-refroidissement	-8.513220	2.459016	3.600962	1.0491803
## imp-confort-pilote	-8.522082	3.508197	4.341346	1.0499485
## imp-rangement	-8.597680	2.672131	3.740385	1.1120991
## imp-tableau-de-bord	-9.438336	2.213115	3.514423	1.0261300
## imp-freinage	-10.308642	2.934426	4.151442	0.9895992
##	Overall sd	p.value		
## imp-tablier-avant	1.3628917	1.908959e-02		
## imp-magasin	1.3345568	4.523631e-03		
## note-satisfaction	1.2695016	5.049339e-05		
## imp-propulsion	1.0863656	4.303452e-08		
## imp-marque	1.0307652	2.926073e-08		
## imp-confort-passager	1.1503151	1.598758e-12		
## imp-feux	1.0176762	1.209266e-12		
## imp-accessoires	1.2180742	7.150073e-14		
## imp-fiabilité-moteur	0.7557373	2.555020e-14		
## imp-cylindrée	0.8703085	1.474261e-14		
## imp-dimensions	0.9807427	3.443857e-16		
## imp-antivol	1.3284355	3.809456e-17		
## imp-esthétique	0.8191403	3.359296e-17		
## imp-refroidissement	1.1327294	1.691691e-17		
## imp-confort-pilote	0.8255663	1.567097e-17		
## imp-rangement	1.0492231	8.134393e-18		
## imp-tableau-de-bord	1.1642857	3.787464e-21		
## imp-freinage	0.9969414	6.440465e-25		

Dans le cluster 1, nous avons toutes les variables qui sont significatives. Les plus représentatives sont les 5 suivantes : imp-freinage, imp-tableau-de-bord, imp-rangement, imp-confort-pilote et imp-refroidissement. Ce cluster est représenté par des notes faibles dans ces variables.

##	v.test	Mean in category	Overall mean	sd in category
## imp-esthétique	2.457046	4.552632	4.391827	0.6225599
## imp-confort-pilote	-2.382296	4.184211	4.341346	0.8009401
## imp-confort-passager	-2.610043	3.394737	3.634615	1.1593890
## imp-accessoires	-2.710123	3.192982	3.456731	1.2764891
## imp-cylindrée	-2.767464	4.026316	4.218750	0.8831813
## imp-magasin	-2.785247	2.938596	3.235577	1.3395802
## imp-fiabilité-moteur	-2.969102	4.289474	4.468750	0.7098762
## imp-rangement	-3.076837	3.482456	3.740385	1.0107532
## imp-feux	-3.239643	3.789474	4.052885	1.1506953
## imp-freinage	-3.773479	3.850877	4.151442	1.0448911
## imp-marque	-4.212440	3.657895	4.004808	1.1981634
## imp-refroidissement	-6.834241	2.982456	3.600962	0.9998461
## imp-tablier-avant	-6.916726	2.657895	3.411058	1.3561358
## imp-propulsion	-7.454853	3.254386	3.901442	1.2125904
##	Overall sd	p.value		
## imp-esthétique	0.8191403	1.400847e-02		
## imp-confort-pilote	0.8255663	1.720507e-02		
## imp-confort-passager	1.1503151	9.053097e-03		
## imp-accessoires	1.2180742	6.725832e-03		
## imp-cylindrée	0.8703085	5.649420e-03		
## imp-magasin	1.3345568	5.348692e-03		
## imp-fiabilité-moteur	0.7557373	2.986717e-03		
## imp-rangement	1.0492231	2.092098e-03		
## imp-feux	1.0176762	1.196793e-03		
## imp-freinage	0.9969414	1.609867e-04		
## imp-marque	1.0307652	2.526263e-05		
## imp-refroidissement	1.1327294	8.244018e-12		
## imp-tablier-avant	1.3628917	4.621996e-12		
## imp-propulsion	1.0863656	8.996797e-14		

Dans le cluster 2, nous avons également toutes les variables qui sont significatives. Les 3 plus représentatives sont les suivantes : imp-propulsion, imp-tablier-avant et imp-refroidissement. Toutes les autres ont une p-value bien plus élevée mais inférieure à 5%. Ce cluster est représenté par des notes faibles dans ces 3 variables.

##	v.test	Mean in category	Overall mean	sd in category
## imp-refroidissement	12.274958	4.182573	3.600962	0.7835625
## imp-freinage	10.796083	4.601660	4.151442	0.5752839
## imp-propulsion	10.660715	4.385892	3.901442	0.6541030
## imp-rangement	8.940662	4.132780	3.740385	0.7935190
## imp-confort-pilote	8.258974	4.626556	4.341346	0.5703244
## imp-fiabilité-moteur	8.142096	4.726141	4.468750	0.5612867
## imp-feux	8.017670	4.394191	4.052885	0.7101863
## imp-cylindrée	8.010558	4.510373	4.218750	0.6512805
## imp-tablier-avant	7.928718	3.863071	3.411058	1.1639134
## imp-accessoires	7.812097	3.854772	3.456731	1.0304835
## imp-marque	7.779833	4.340249	4.004808	0.7233497
## imp-confort-passager	7.421119	3.991701	3.634615	0.9597359
## imp-dimensions	6.472506	4.311203	4.045673	0.7391252
## imp-tableau-de-bord	6.221610	3.817427	3.514423	0.9853003

## imp-antivol	5.450495	3.348548	3.045673	1.2369183
## imp-magasin	4.550876	3.489627	3.235577	1.2659342
## imp-esthétique	3.823040	4.522822	4.391827	0.6182712
## note-satisfaction	3.353367	8.377593	8.199519	1.1056982
##	Overall sd	p.value		
## imp-refroidissement	1.1327294	1.234555e-34		
## imp-freinage	0.9969414	3.592007e-27		
## imp-propulsion	1.0863656	1.553984e-26		
## imp-rangement	1.0492231	3.868449e-19		
## imp-confort-pilote	0.8255663	1.469293e-16		
## imp-fiabilité-moteur	0.7557373	3.884945e-16		
## imp-feux	1.0176762	1.077696e-15		
## imp-cylindrée	0.8703085	1.141895e-15		
## imp-tablier-avant	1.3628917	2.214199e-15		
## imp-accessoires	1.2180742	5.624419e-15		
## imp-marque	1.0307652	7.262015e-15		
## imp-confort-passager	1.1503151	1.161352e-13		
## imp-dimensions	0.9807427	9.639085e-11		
## imp-tableau-de-bord	1.1642857	4.920790e-10		
## imp-antivol	1.3284355	5.022968e-08		
## imp-magasin	1.3345568	5.342293e-06		
## imp-esthétique	0.8191403	1.318166e-04		
## note-satisfaction	1.2695016	7.983471e-04		

Dans le cluster 3, nous avons également toutes les variables qui sont significatives. Les 5 plus représentatives sont les suivantes : imp-refroidissement, imp-freinage, imp-propulsion, imp-rangement et imp-confort-pilote. Ce cluster est représenté par des notes élevées dans toutes ces variables mais surtout dans les 5 citées précédemment.

On retient que les individus du groupe 1 et 2 vont mettre des notes faibles pour les variables significatives citées plus haut telles que imp-freinage ou encore imp-propulsion et que les individus du groupe 3 mettront eux des notes élevées pour ces mêmes variables.

Conclusion

Pour résumer, suite à nos analyses, nous avons identifié 4 groupes associés aux clients de la base. Nous avons trouvé des similarités entre les 2 premiers groupes à savoir les individus de ces 2 groupes sont des hommes qui préfèrent les scooter innovants, qui ressemblent à des motos et qui ont un gros cylindre mais qui sont sobre et discret . Ce sont généralement des hommes qui sont de classe supérieur ou moyenne. Concernant les deux derniers groupes, ils sont caractérisés par des hommes qui préfèrent les scooters innovants, qui ressemblent à des motos et qui ont un gros cylindre mais qui sont sobre et discret . Ce sont des hommes qui sont de classe supérieur ou moyenne. Concernant les analyses en fonction du sexe, nous avons identifié les scooters à proposer en fonction de ceux-ci. A une population de femmes, nous pouvons proposer des scooters à type de cylindre petit et au critère esthétique classique et à la mode. Cependant, pour les hommes, le type de scooter à proposer serait plutôt du style gros cylindre et au critère esthétique nouveau. La principale différence se trouve au niveau de type de cylindre mais aussi au niveau du critère esthétique.