

# PROJET 8: Scooter

Reau Vincent - Buani

2023-04-05

```
library("readxl")
library(dplyr)

##
## Attachement du package : 'dplyr'
## Les objets suivants sont masqués depuis 'package:stats':
##
##   filter, lag
## Les objets suivants sont masqués depuis 'package:base':
##
##   intersect, setdiff, setequal, union
library(FactoMineR)
library(factoextra)

## Le chargement a nécessité le package : ggplot2
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
d = read_excel("/home/b/d/bdokieng/mod_av_projet_scooter/scooter.xls")
d$ident=NULL
d$`note-satisfaction` = as.numeric(d$`note-satisfaction`)
head(d)

## # A tibble: 6 x 25
##   sexe   âge      CSP   type--1 type--2 critè-3 note--4 imp-m-5 imp-m-6 imp-e-7
##   <chr> <chr>   <chr> <chr>   <chr>   <chr>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 femme 25-29 ans emplo~ 125   domici~ à la m~      8       5       3       5
## 2 homme 30-39 ans class~ 150-200 domici~ innova~    10      4       4       5
## 3 homme 25-29 ans emplo~ 150-200 domici~ sportif     8      2       4       5
## 4 homme 30-39 ans ouvri~ 125   domici~ innova~     8      4       4       4
## 5 homme 40-49 ans class~ 250-400 domici~ innova~     9      3       3       4
## 6 homme 30-39 ans class~ 250-400 domici~ discre~     9      5       3       4
## # ... with 15 more variables: `imp-prix` <dbl>, `imp-confort-pilote` <dbl>,
## #   `imp-confort-passager` <dbl>, `imp-dimensions` <dbl>, `imp-freinage` <dbl>,
## #   `imp-cylindrée` <dbl>, `imp-antivol` <dbl>, `imp-tableau-de-bord` <dbl>,
## #   `imp-accessoires` <dbl>, `imp-rangement` <dbl>, `imp-propulsion` <dbl>,
## #   `imp-refroidissement` <dbl>, `imp-tablier-avant` <dbl>, `imp-feux` <dbl>,
## #   `imp-fiabilité-moteur` <dbl>, and abbreviated variable names
## #   1: `type-cylindrée`, 2: `type-utilisation`, 3: `critère-esthétique`, ...

#d %>% mutate_all(as.factor)
#class(d)
```

```
#str(d) # affiche la structure (le type) des données
```

## Detection des corrélations les plus fortes

```
## corrplot 0.92 loaded
```

## L'ACP

```
## Le chargement a nécessité le package : shiny
```

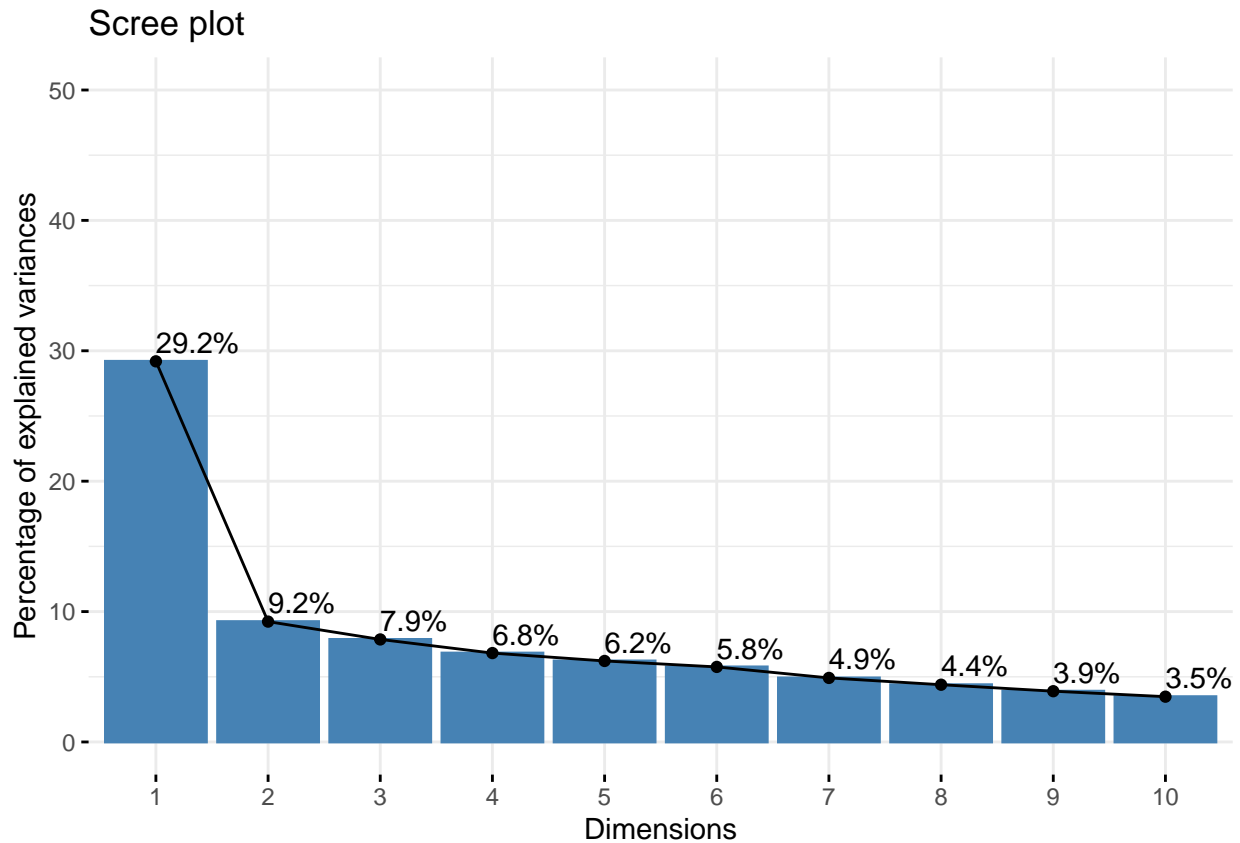
```
## Le chargement a nécessité le package : FactoInvestigate
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## comp 1    4.9624603             29.190943             29.19094
## comp 2    1.5692621             9.230954             38.42190
## comp 3    1.3368234             7.863667             46.28556
## comp 4    1.1589770             6.817512             53.10308
## comp 5    1.0563379             6.213752             59.31683
## comp 6    0.9786703             5.756884             65.07371
## comp 7    0.8352410             4.913182             69.98689
## comp 8    0.7465451             4.391442             74.37834
## comp 9    0.6621163             3.894802             78.27314
## comp 10   0.5907611             3.475065             81.74820
## comp 11   0.5629822             3.311660             85.05986
## comp 12   0.5298541             3.116789             88.17665
## comp 13   0.5005213             2.944243             91.12089
## comp 14   0.4225907             2.485828             93.60672
## comp 15   0.4102899             2.413470             96.02019
## comp 16   0.3776856             2.221680             98.24187
## comp 17   0.2988818             1.758128             100.00000
## [1] 0.4749425
```

## 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.

```
fviz_eig(res.PCA, addlabels = TRUE, ylim = c(0, 50))
```



## Etude des variables

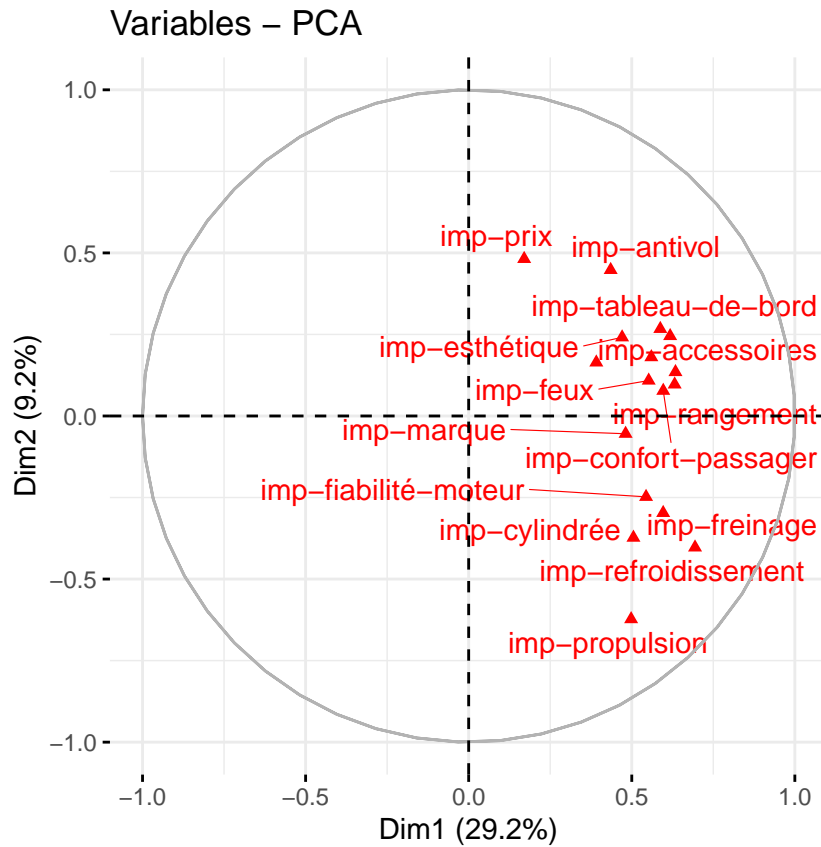
### Le cercle de corrélation des variables quantitatives

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 .

```
# création d'un dataframe avec uniquement les variables quantitatives
fviz_famd_var (res.PCA, repel = TRUE)
```

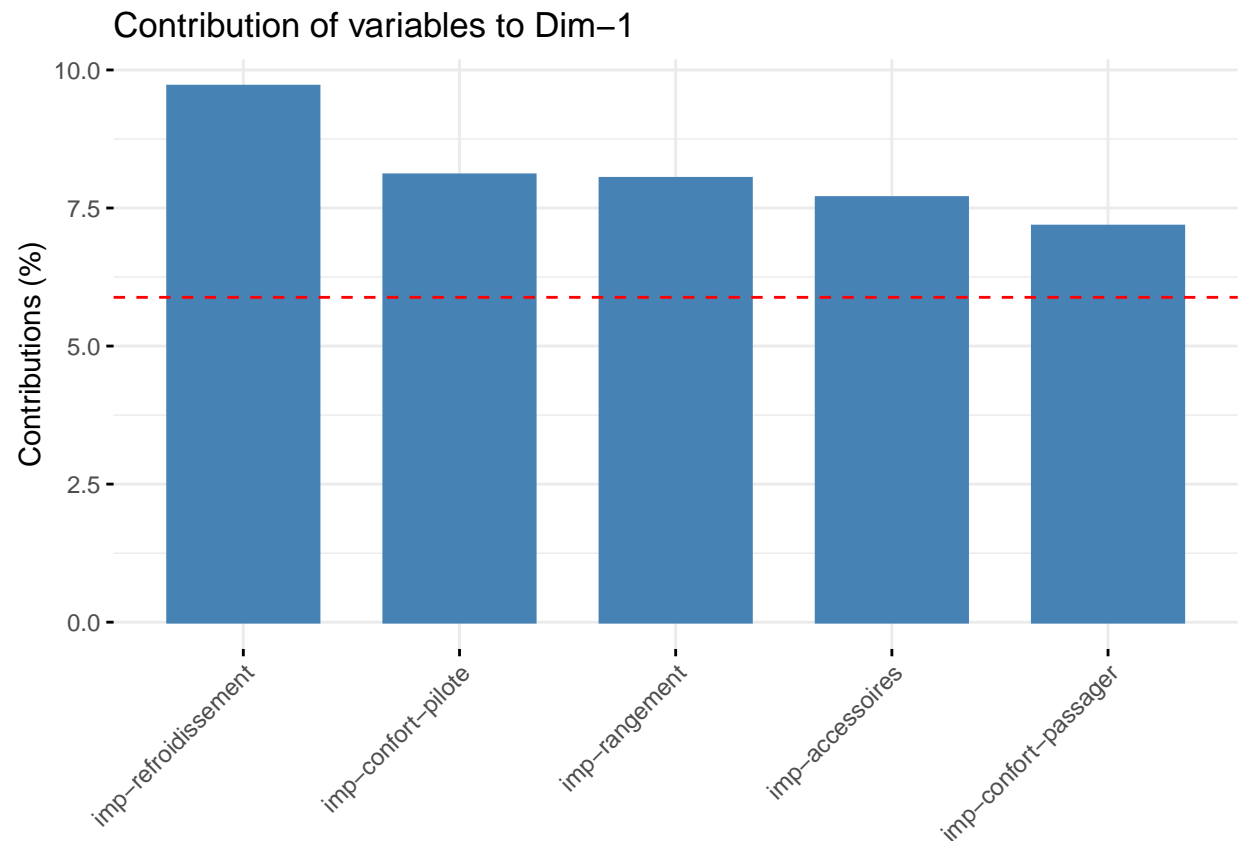
```
## Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



### 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.

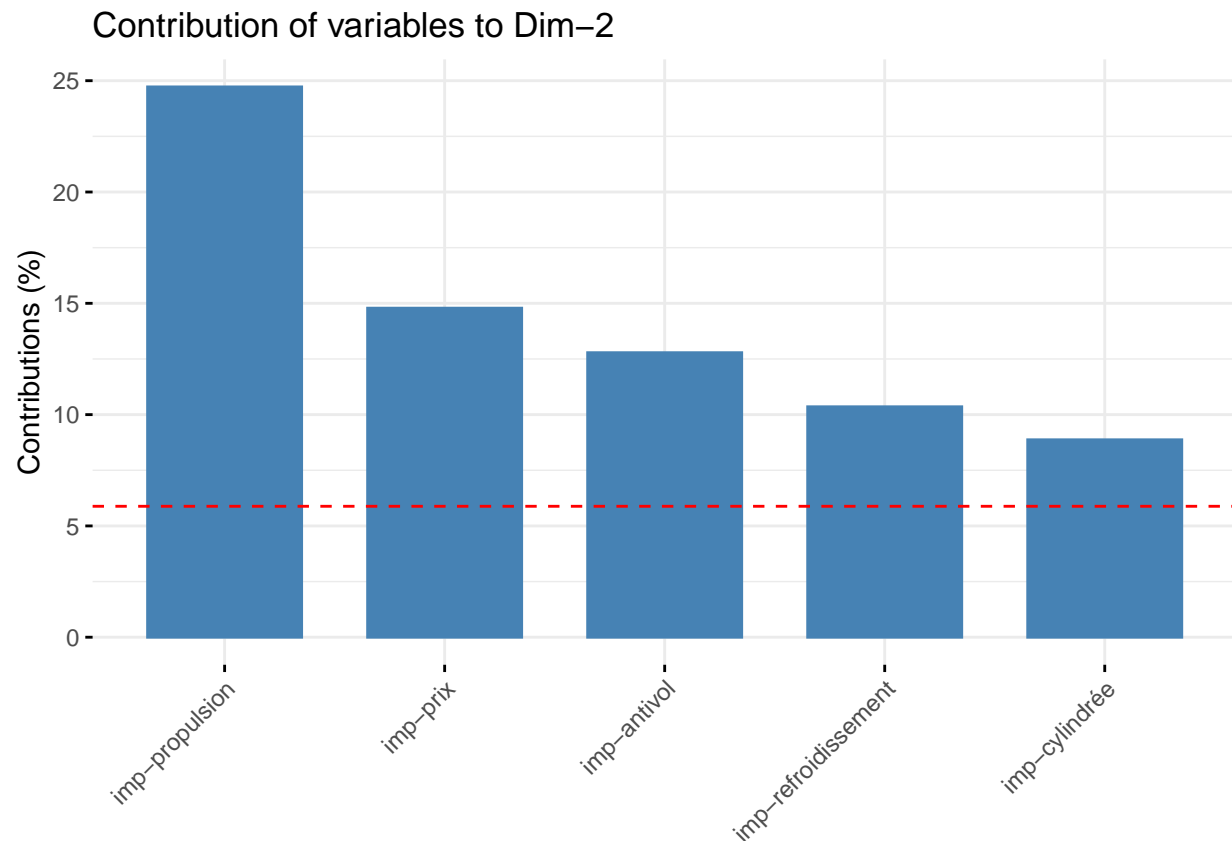
```
fviz_contrib(res.PCA, choice = "var", axes = 1, top = 5)
```



#### *# Contributions des variables à PC2*

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.

```
fviz_contrib(res.PCA, choice = "var", axes = 2, top = 5)
```



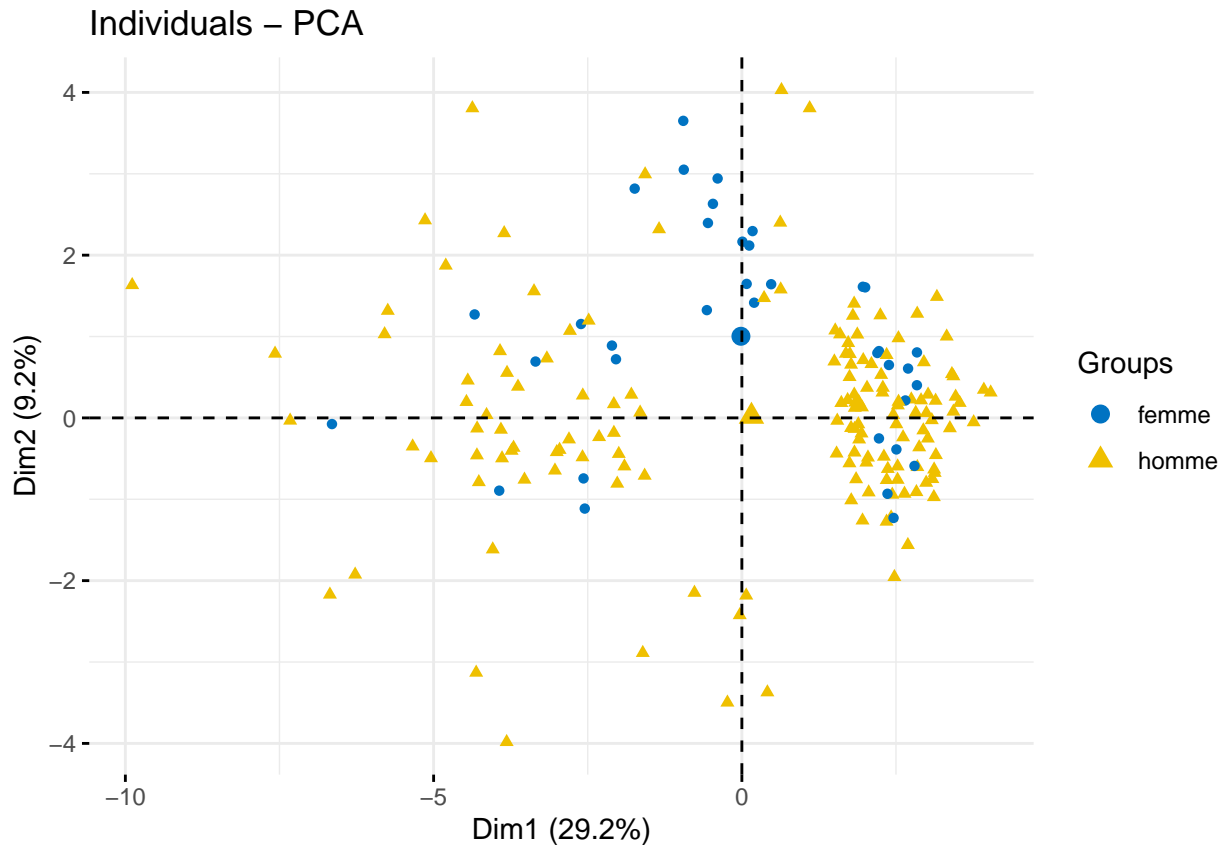
### Les individus

Le but 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.

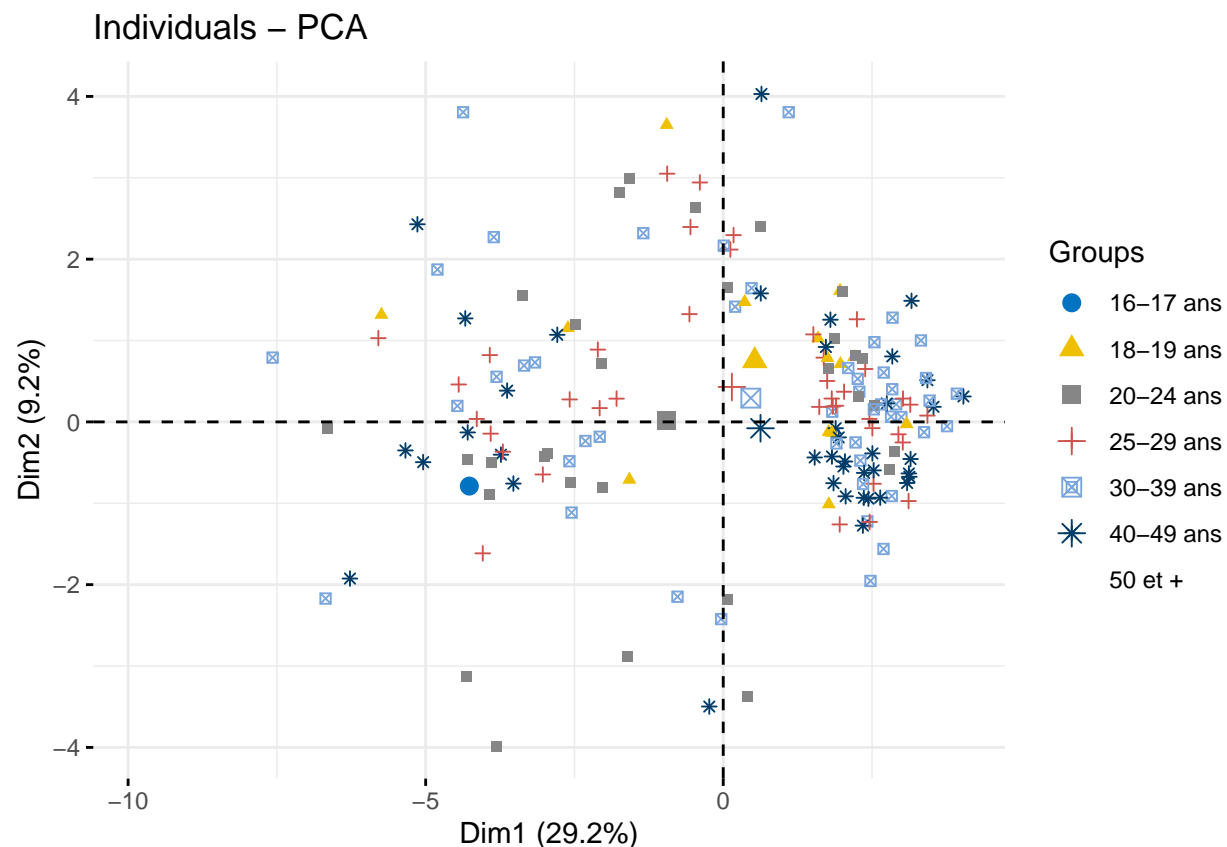
*Tableau des moyennes des notes pour la marques , l'esthétique , le confort pour les idnividus ayant des coordonnées négatives sur la dimension 2*

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

```
## Warning: The shape palette can deal with a maximum of 6 discrete values because
## more than 6 becomes difficult to discriminate; you have 7. Consider
## specifying shapes manually if you must have them.
```

```
## Warning: Removed 16 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



### Critère de l'âge

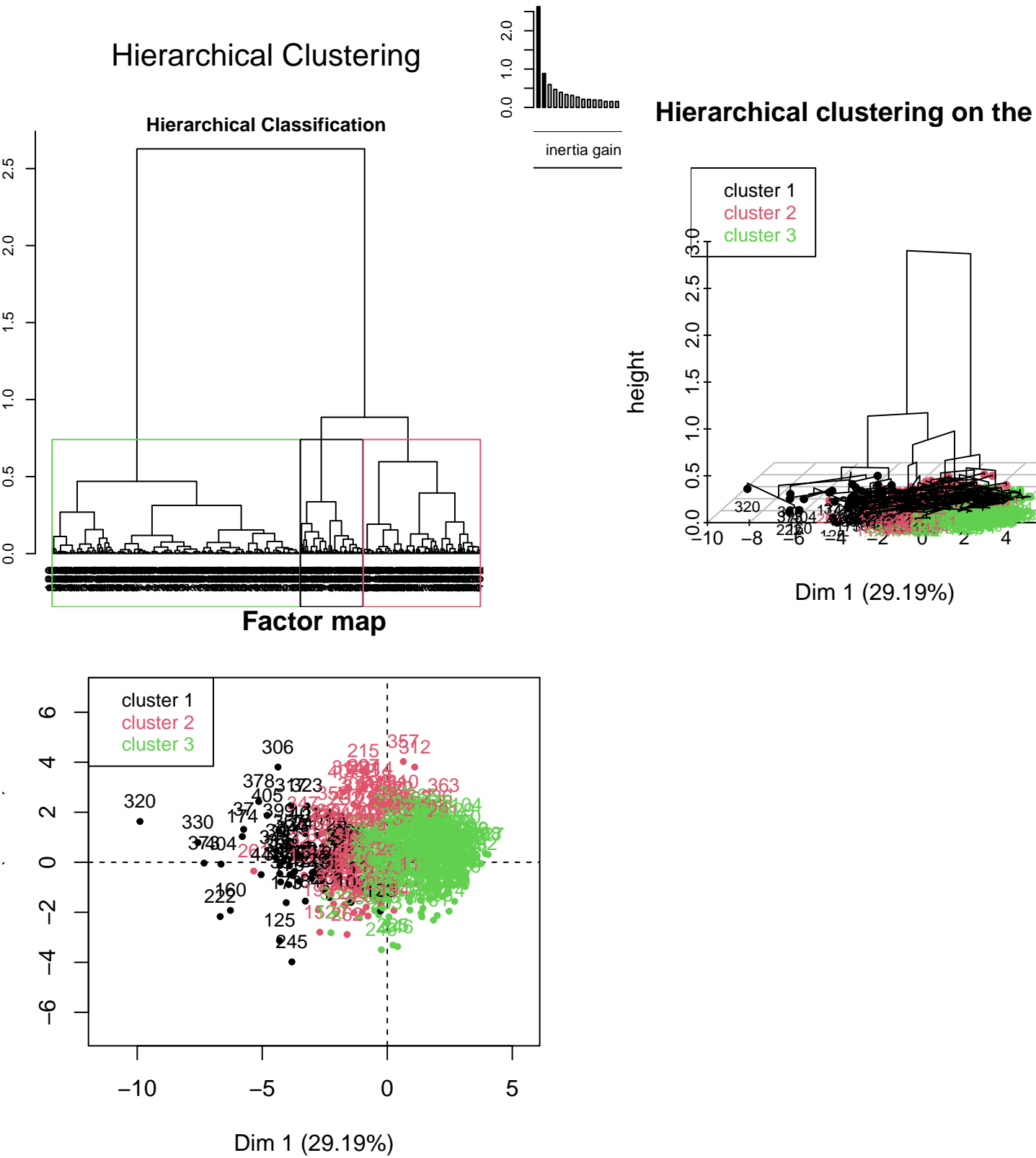
## Description des dimensions

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 par les qualités de refroidissement, de freinage, de propulsion, de cylindrée, de fiabilité du moteur, du confort pour le passager, la qualité des accessoires, du rangement, du tableau de bord et du confort du pilote pour le scooter. Ce sont majoritairement des hommes ont entre 30-39 ans 40-49 ans et qui appartiennent à classe moyenne.

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



Classification hiérarchique ascendante (CAH) avec l'acp



##

```

## Link between the cluster variable and the categorical variables (chi-square test)
## =====
##                p.value df
## type-cylindrée 1.155782e-05 6
## sexe          1.149047e-02 2
##
## Description of each cluster by the categories
## =====
## $`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
##
##
## Link between the cluster variable and the quantitative variables
## =====
##              Eta2      P-value
## imp-refroidissement  0.38346917 4.223435e-44
## imp-freinage         0.36157668 5.687943e-41
## imp-propulsion       0.27412658 1.852568e-29
## imp-rangement        0.24959075 1.774489e-26
## imp-confort-pilote   0.22841176 5.560653e-24
## imp-tableau-de-bord  0.22304584 2.326212e-23
## imp-fiabilité-moteur 0.20199000 5.817288e-21
## imp-cylindrée        0.20003760 9.635338e-21
## imp-accessoires      0.18991878 1.291662e-19
## imp-feux             0.18730528 2.511979e-19
## imp-dimensions       0.18011896 1.547211e-18
## imp-antivol          0.17657634 3.768945e-18
## imp-esthétique       0.17162293 1.300459e-17
## imp-confort-passager 0.17040004 1.763562e-17
## imp-tablier-avant    0.15870847 3.172893e-16
## imp-marque           0.15563751 6.733249e-16
## imp-magasin          0.05113919 1.960282e-05
## note-satisfaction    0.04561466 6.500642e-05
##
## Description of each cluster by quantitative variables
## =====
## $`1`
##              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

```

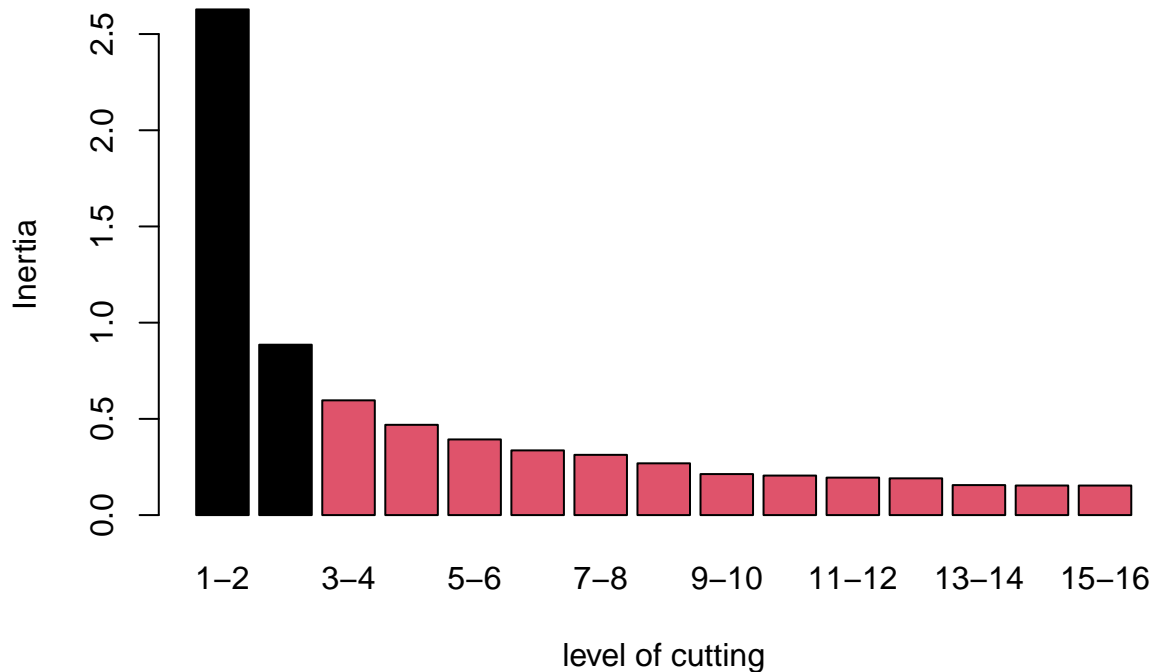
```

## 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
##
## $`2`
##               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
##
## $`3`
##               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		

## Inter-cluster inertia gains



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, ils 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 cylindrées 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 cooters avec de petits cylindrés à des femmes plutôt jeunes et de puissants scooter dernier cris aux hommes .