

Rapport de la quatrième partie du TP4 de statistique exploratoire

Rémy Gaudré

17/10/2019

Présentation des données

```
## donnees
##
## 14 Variables      27 Observations
## -----
## ORTH
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      18    0.997    10.02    5.553    4.50    4.80
##      .25      .50      .75      .90      .95
##      5.75    11.00    14.00    15.50    15.85
##
## Value      0.0  4.5  5.0  5.5  6.0  6.5  7.0  8.5  9.0  11.0
## Frequency      1   2   1   3   1   2   1   1   1   2
## Proportion 0.037 0.074 0.037 0.111 0.037 0.074 0.037 0.037 0.037 0.074
##
## Value      11.5 12.0 13.0 14.0 15.0 15.5 16.0 20.0
## Frequency      1   1   2   2   2   2   1   1
## Proportion 0.037 0.037 0.074 0.074 0.074 0.074 0.037 0.037
## -----
## GRAM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      15    0.989    7.556    3.142    2.95    4.00
##      .25      .50      .75      .90      .95
##      6.50    7.50    8.50    10.80    13.05
##
## Value      2.0  2.5  4.0  4.5  5.5  6.5  7.0  7.5  8.0  8.5
## Frequency      1   1   2   1   1   3   1   4   5   2
## Proportion 0.037 0.037 0.074 0.037 0.037 0.111 0.037 0.148 0.185 0.074
##
## Value      9.0 10.0 12.0 13.5 14.5
## Frequency      2   1   1   1   1
## Proportion 0.074 0.037 0.037 0.037 0.037
## -----
## EXPR
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      17    0.993    8.889    3.407    4.8    5.8
##      .25      .50      .75      .90      .95
##      7.5     9.0    10.5    12.2    13.9
##
## Value      2.0  4.5  5.5  6.0  6.5  7.5  8.0  8.5  9.0  10.0
## Frequency      1   1   1   2   1   4   2   1   4   2
## Proportion 0.037 0.037 0.037 0.074 0.037 0.148 0.074 0.037 0.148 0.074
##
## Value      10.5 11.0 11.5 12.0 12.5 14.5 16.5
```

```

## Frequency      2      1      1      1      1      1      1
## Proportion 0.074 0.037 0.037 0.037 0.037 0.037 0.037
## -----
## RECI
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      15    0.994    10.07    4.171     5.0     5.6
##      .25     .50     .75     .90     .95
##      6.5    10.0    12.5    14.7    15.7
##
## Value      4.0    5.0    6.0    6.5    8.0    9.5    10.0    11.0    11.5    12.0
## Frequency      1      2      3      2      1      2      4      2      1      2
## Proportion 0.037 0.074 0.111 0.074 0.037 0.074 0.148 0.074 0.037 0.074
##
## Value      13.0   14.0   14.5   15.0   16.0
## Frequency      1      2      1      1      2
## Proportion 0.037 0.074 0.037 0.037 0.074
## -----
## MATH
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      12    0.989    12.57    3.251     9.0     9.6
##      .25     .50     .75     .90     .95
##     10.0    12.0    14.5    16.4    17.7
##
## Value      8.0    9.0   10.0   11.0   12.0   12.5   13.0   14.0   15.0   16.0
## Frequency      1      2      5      3      3      1      3      2      2      2
## Proportion 0.037 0.074 0.185 0.111 0.111 0.037 0.111 0.074 0.074 0.074
##
## Value      17.0   18.0
## Frequency      1      2
## Proportion 0.037 0.074
## -----
## ANGL
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      19    0.995    10.93    3.803     6.65     7.30
##      .25     .50     .75     .90     .95
##      8.75   11.00   12.75   15.30   16.85
##
## Value      3.0    6.5    7.0    7.5    8.0    8.5    9.0    9.5   10.0   10.5
## Frequency      1      1      1      2      1      1      2      1      1      1
## Proportion 0.037 0.037 0.037 0.074 0.037 0.037 0.074 0.037 0.037 0.037
##
## Value      11.0   11.5   12.0   12.5   13.0   13.5   14.5   16.5   17.0
## Frequency      2      1      4      1      1      2      1      1      2
## Proportion 0.074 0.037 0.148 0.037 0.037 0.074 0.037 0.037 0.074
## -----
## HIST
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      11    0.972     9.204    2.373     7.00     7.00
##      .25     .50     .75     .90     .95
##      7.75     9.00   11.00   11.70   12.00
##
## Value      5.0    7.0    7.5    8.0    9.0    9.5   10.0   11.0   11.5   12.0
## Frequency      1      5      1      3      7      1      1      4      1      2
## Proportion 0.037 0.185 0.037 0.111 0.259 0.037 0.037 0.148 0.037 0.074

```

```

##
## Value      15.0
## Frequency   1
## Proportion 0.037
## -----
## BIOL
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      16    0.992    9.593    4.262    4.65    5.00
##      .25     .50     .75     .90     .95
##      7.00    10.00    12.00    14.10    15.70
##
## Value      2.0  4.5  5.0  7.0  7.5  8.0  8.5  10.0  10.5  11.0
## Frequency   1   1   2   5   1   2   1   2   1   2
## Proportion 0.037 0.037 0.074 0.185 0.037 0.074 0.037 0.074 0.037 0.074
##
## Value      11.5 12.5 13.5 15.0 16.0 17.0
## Frequency   2   2   2   1   1   1
## Proportion 0.074 0.074 0.074 0.037 0.037 0.037
## -----
## EDMU
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      15    0.993    15.2    3.487    8.10    11.00
##      .25     .50     .75     .90     .95
##      13.75   16.00    17.50    18.00    18.35
##
## Value      7.0  7.5  9.5 12.0 13.5 14.0 15.0 15.5 16.0 16.5
## Frequency   1   1   1   2   2   2   1   1   3   2
## Proportion 0.037 0.037 0.037 0.074 0.074 0.074 0.037 0.037 0.111 0.074
##
## Value      17.0 17.5 18.0 18.5 19.0
## Frequency   2   3   4   1   1
## Proportion 0.074 0.111 0.148 0.037 0.037
## -----
## ARTS
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      17    0.996    9.019    4.561    2.10    3.80
##      .25     .50     .75     .90     .95
##      6.75    9.00    12.50    13.70    14.00
##
## Value      1.5  3.5  4.0  5.0  6.5  7.0  8.0  8.5  9.0  9.5
## Frequency   2   1   1   2   1   2   3   1   2   1
## Proportion 0.074 0.037 0.037 0.074 0.037 0.074 0.111 0.037 0.074 0.037
##
## Value      11.0 11.5 12.0 12.5 13.5 14.0 14.5
## Frequency   1   1   1   2   3   2   1
## Proportion 0.037 0.037 0.037 0.074 0.111 0.074 0.037
## -----
## TECH
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      11    0.987    12.63    5.128    1.8     6.0
##      .25     .50     .75     .90     .95
##      11.5    14.0    16.0    17.4    18.0
##
## Value      0   6   8   11   12   13   14   15   16   17

```

```

## Frequency      2      2      1      2      2      4      3      2      5      1
## Proportion 0.074 0.074 0.037 0.074 0.074 0.148 0.111 0.074 0.185 0.037
##
## Value          18
## Frequency      3
## Proportion 0.111
## -----
## EPS
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      17    0.991    13.07    3.365    9.30    10.30
##      .25     .50     .75     .90     .95
##     11.00    13.50    15.25    16.50    17.55
##
## Value          5.0   9.0  10.0  10.5  11.0  11.5  12.0  12.5  13.0  13.5
## Frequency      1     1     1     3     2     1     1     2     1     5
## Proportion 0.037 0.037 0.037 0.111 0.074 0.037 0.037 0.074 0.037 0.185
##
## Value          14.5  15.0  15.5  16.0  16.5  18.0  18.5
## Frequency      1     1     1     2     2     1     1
## Proportion 0.037 0.037 0.037 0.074 0.074 0.037 0.037
## -----
## GEO
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      27      0      12    0.968    13.8     2.875    11.10    12.80
##      .25     .50     .75     .90     .95
##     13.00    14.00    15.25    16.60    17.50
##
## Value          0.00 10.50 12.50 13.00 13.50 14.00 15.00 15.50 15.51 16.00
## Frequency      1     1     1     8     1     4     4     1     1     2
## Proportion 0.037 0.037 0.037 0.296 0.037 0.148 0.148 0.037 0.037 0.074
##
## Value          17.50 18.00
## Frequency      2     1
## Proportion 0.074 0.037
## -----
## EXP0
##      n missing distinct      Info      Mean      Gmd
##      27      0      9    0.976    14.48    3.963
##
## Value          0.0  13.0  14.0  15.0  16.0  16.5  17.0  17.5  18.0
## Frequency      2     5     1     6     3     1     4     1     4
## Proportion 0.074 0.185 0.037 0.222 0.111 0.037 0.148 0.037 0.148
## -----

```

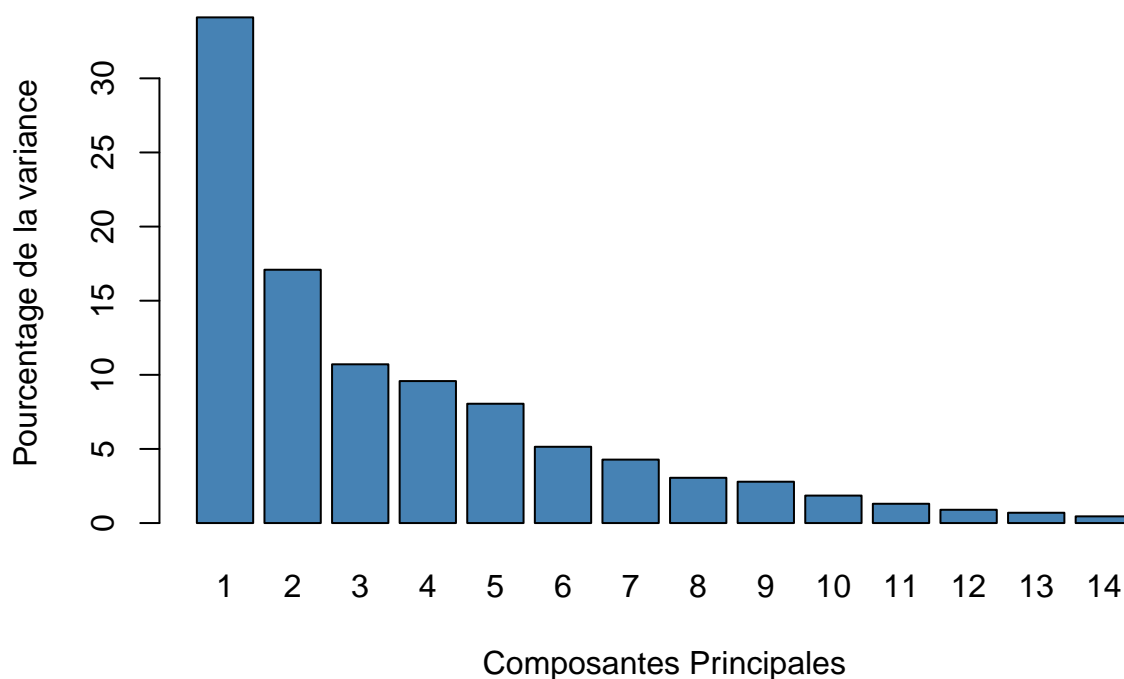
Les données sont composées de notes de 27 élèves dans 14 matières. Les variables sont toutes quantitatives continues. Les données sont toutes comprises entre 0 et 20. Les moyennes de la Grammaire, de l'Expression Ecrite, d'Histoire, de Biologie, d'Arts-Plastique sont en dessous de 10. Les autres matières ont une moyenne au dessus de 10. Les écarts types sont tous en dessous de 5. Les notes sont toute fois diversifié en regardant les minimum et les maximum. Les notes pouvant être très mauvaise comme très bonne. La matière la plus réussi en regardant les données brut étant l'éducation musicale avec une moyenne de 15.59 une note minimal de 7 et maximale de 19.

Partie A1 : réalisation de l'ACP

Combien de composantes principales peut-on retenir ?

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.7757005	34.1121461	34.11215
comp 2	2.3923845	17.0884607	51.20061
comp 3	1.4993422	10.7095870	61.91019
comp 4	1.3408417	9.5774409	71.48763
comp 5	1.1267019	8.0478709	79.53551
comp 6	0.7203848	5.1456057	84.68111
comp 7	0.5991907	4.2799337	88.96104
comp 8	0.4274849	3.0534636	92.01451
comp 9	0.3904079	2.7886277	94.80314
comp 10	0.2593168	1.8522627	96.65540
comp 11	0.1821002	1.3007158	97.95611
comp 12	0.1255255	0.8966105	98.85273
comp 13	0.0972880	0.6949142	99.54764
comp 14	0.0633305	0.4523606	100.00000

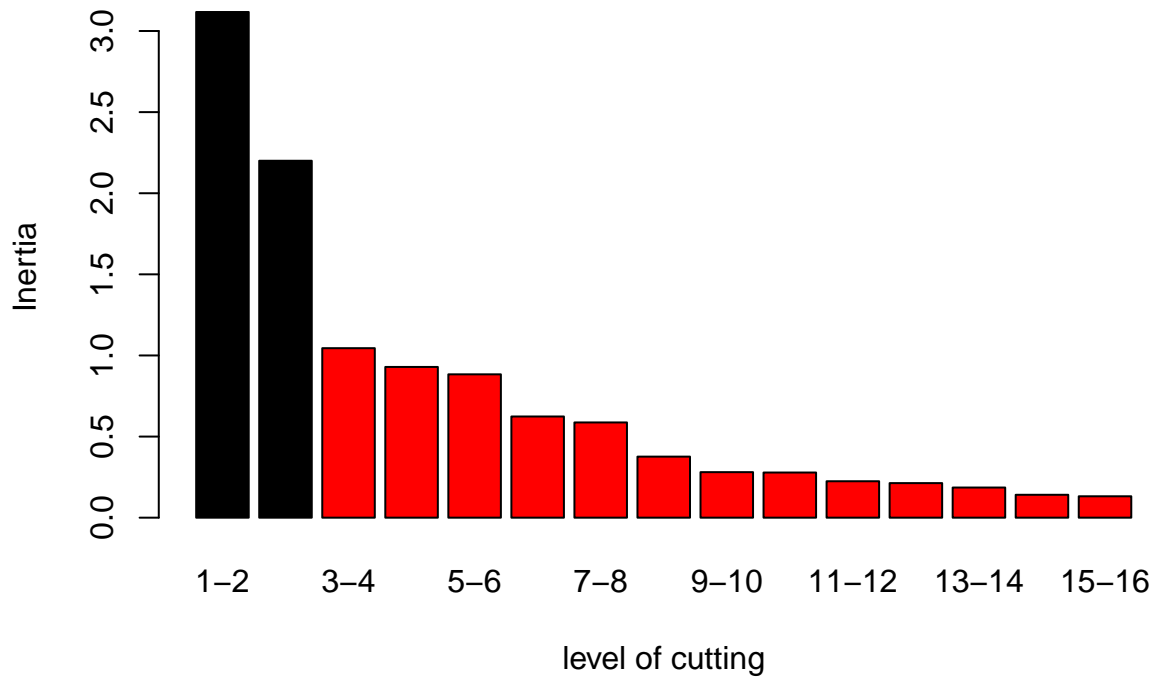
La variances expliquée pas les composantes principales (%)



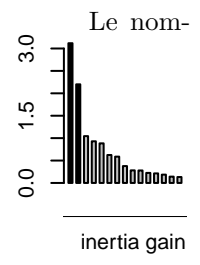
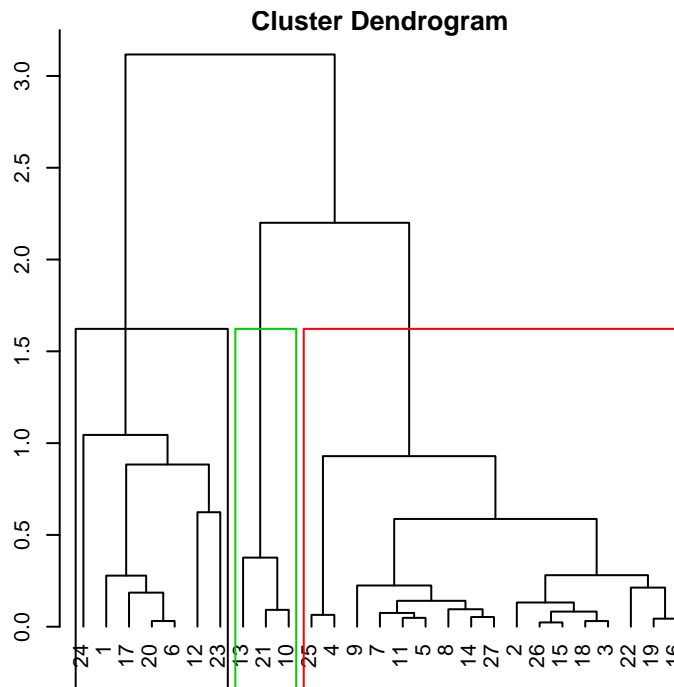
D'après le critère des 80% de variance expliquée : nous avons 6 composantes principales à retenir.
Le critère du coude nous indique un premier palier à partir de la troisième composante principale et une chute au sixième axe. D'après le critère de Kaiser, nous avons 5 composantes principales à retenir.

Partie A2 : Réalisation de la classification Ascendante Hiérarchique

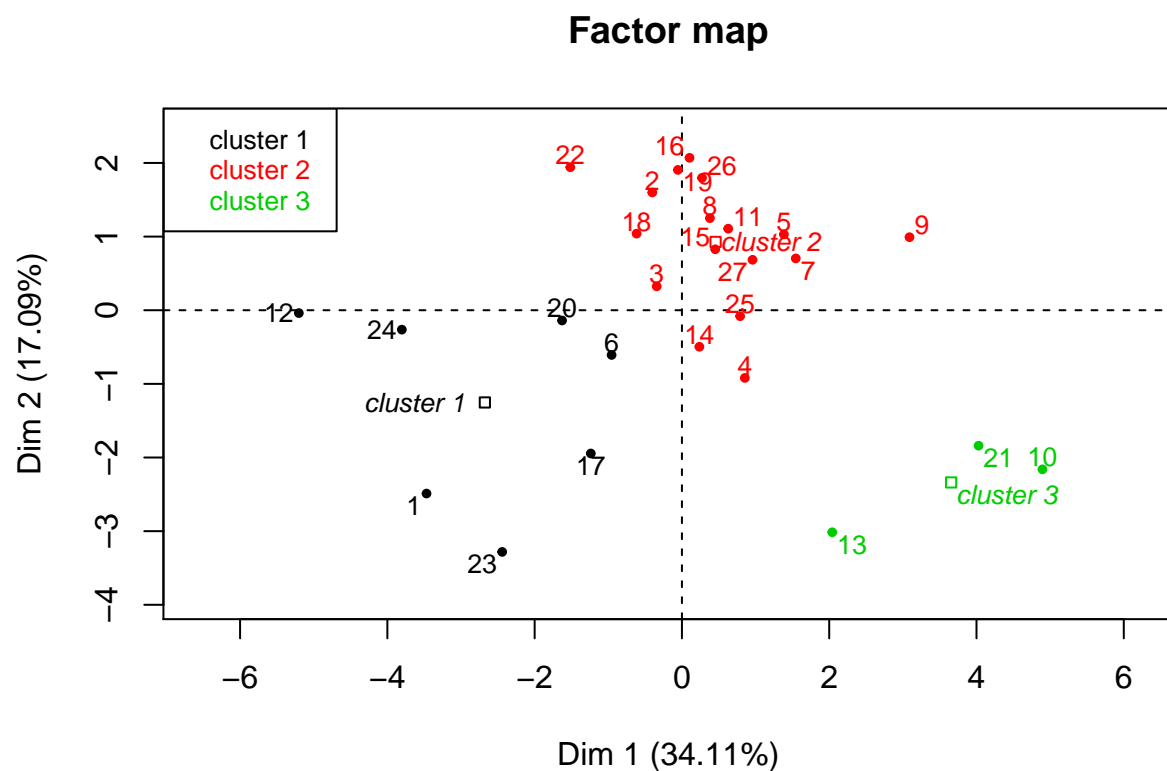
Inter-cluster inertia gains



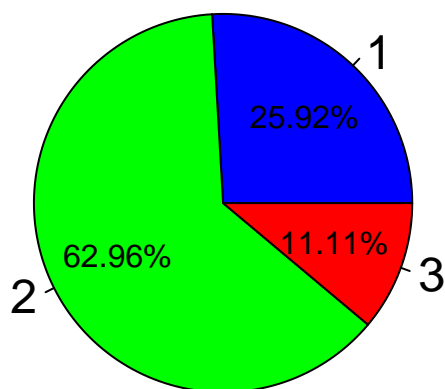
Hierarchical clustering



bre de classe optimale est 3.



Répartition des élèves selon les classes



La classe la plus représentée par les élèves est la classe numéro 2 avec plus de la moitié des élèves y appartenant (62.96%). La classe numéro 1 contient 25,92% des élèves et la classe numéro 3 est en minorité avec 11,11% des élèves.

Etude des élèves

Table 2: Les matières significatives de la classe 1

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
EXPO	-2.620437	10.642857	14.481481	6.890544	4.418986	0.0087817
GEO	-2.645103	11.000000	13.796667	4.605897	3.189472	0.0081666
BIOL	-3.226408	5.714286	9.592593	1.749635	3.626132	0.0012535

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
EXPR	-3.420592	5.500000	8.888889	1.732051	2.988662	0.0006249
ARTS	-3.587296	4.428571	9.018519	2.258770	3.859767	0.0003341
EDMU	-4.261928	10.714286	15.203704	2.519313	3.177640	0.0000203

Les élèves de la classe 1 ont en général des notes inférieures au reste de la classe. Ils ont des notes en dessous de la moyenne en Exposé, en Géographie, en Biologie, en Expression écrite, en Art plastique et en Education musicale de façon significative. Les matières étant les plus significatives étant l'Education musicale, l'Art plastique, et l'Expression écrite.

Table 3: Les matières significatives de la classe 2

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
BIOL	3.065725	11.264706	9.592593	2.926138	3.626132	0.0021714
EDMU	2.958275	16.617647	15.203704	1.334270	3.177640	0.0030937
GEO	2.386904	14.941765	13.796667	1.570920	3.189472	0.0169909
ARTS	2.349155	10.382353	9.018519	2.943235	3.859767	0.0188161
EPS	2.238990	14.058824	13.074074	2.078478	2.924050	0.0251566
RECI	2.030183	11.147059	10.074074	3.328861	3.513739	0.0423380
EXPO	2.019094	15.823529	14.481481	1.790015	4.418986	0.0434774
GRAM	-2.278116	6.588235	7.555556	2.109051	2.822966	0.0227197

Les élèves de la classe 2 sont généralement des élèves ayant des bonnes notes en Biologie, en Education musicale, en Géographie, en Art plastique, en EPS, en Récitation, et en Exposé de façon significative. Ils ont cependant des lacunes en Grammaire de façon significative. Les matières étant les plus significative de cette classe étant la Biologie et l'Education musicale. Ces élèves sont donc bons dans 7 matières et mauvais dans 1 seule. Les autres matières s'approchant plus de la moyenne de classe.

Table 4: Les matières significatives de la classe 3

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
GRAM	3.689750	13.33333	7.555556	1.0274023	2.822966	0.0002245
EXPR	3.284116	14.33333	8.888889	1.8408935	2.988662	0.0010230
ANGL	3.255989	16.83333	10.925926	0.2357023	3.270813	0.0011300
MATH	2.886181	17.00000	12.574074	0.8164966	2.764536	0.0038995
HIST	2.847026	12.50000	9.203704	1.7795130	2.087260	0.0044130
ORTH	2.725159	17.16667	10.018519	2.0138410	4.728717	0.0064271

Les élèves de la classe 3 sont généralement des élèves littéraires. Ils sont bons en Grammaire, en Expression, en Anglais, en Math, en Histoire et en Orthographe. Les matières les plus significatives de la classe étant la Grammaire, l'Expression écrite et l'Anglais.

Etude des axes de l'ACP

Table 5: Les axes significatifs de la classe 1

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Dim.2	-2.442027	-1.252118	0	1.209146	1.546733	0.0146050
Dim.1	-3.694421	-2.676358	0	1.437537	2.185338	0.0002204

Les élèves de la première classe ont en moyennes des valeurs faibles dans les 2 premières composantes principales de l'ACP de façon significative. La première composante principale est la plus significative.

Table 6: Les axes significatifs de la classe 2

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Dim.2	3.989715	0.9282099	0	0.8259869	1.546733	6.62e-05

Les élèves de la seconde classe ont en moyenne des valeurs élevées dans la seconde dimension de l'ACP de façon significative.

Table 7: Les axes significatifs de la classe 3

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Dim.1	3.015181	3.655024	0	1.1948009	2.185338	0.0025683
Dim.2	-2.725314	-2.338247	0	0.4960709	1.546733	0.0064240

Les élèves de la troisième classe ont des valeurs élevées dans la première dimension et faible dans la seconde dimension de l'ACP de façon significative. La première composante principale est la plus significative.

Conclusion de la classification ascendante hiérarchique

Pour rappel : on a conclu l'ACP en déterminant que le premier axe de l'ACP déterminait caractériser le niveau de l'élève et le second axe le profil des matières que l'élève réussit le mieux. Les élèves sont répartis dans 3 classes. La première classe contient un quart des élèves. Les élèves de cette classe ont en moyenne des notes assez faibles en Exposé, en Géographie, en Biologie, en Expression écrite en Art Plastique et en Education musicale. Ils sont aussi caractérisés par des valeurs faibles sur les 2 premiers axes de l'ACP. Les élèves de la première classe sont en général en difficulté.

Les élèves de la seconde classe sont en moyenne bons en biologie, en éducation musicale, en géographie en art plastique, en EPS, en récitation et en exposé. Ils sont cependant mauvais en grammaire en moyenne. Ils sont aussi caractérisés par une forte valeur sur l'axe 2 de l'ACP. C'est-à-dire que les élèves sont bons dans 7 des 14 matières et mauvais dans 1 seule. Cette classe représente la majorité des élèves (63% de la classe). Les élèves de la troisième classe ont en moyenne de bonnes notes en grammaire, en expression, en anglais, en mathématique, en histoire et en orthographe. Ils ont donc un profil plus "littéraire". Ce qui est confirmé par la bonne valeur moyenne sur l'axe 1 et la faible valeur moyenne sur l'axe 2 de l'ACP. Ils sont plus atypiques car seulement 11,11% des élèves sont dans cette classe.