

Spotify_Analyse_Factorielle

2024-01-03

Contents

Présentation de la base de données	2
Variables	2
Valeurs manquantes	3
Statistiques descriptives	3
Résumé des variables	3
Graphiques de densité	5
ACP	7
Matrice des corrélations	7
Inertie	8
Description des axes	9
AFC	13
Profil lignes	13
Profil colonnes	13
Contributions aux lignes et aux colonnes	14
Visualisation graphique	15
ACM	17
Dim 1 / Dim 2	18
Dim 3 / Dim 4	23
Classification non supervisée	29
Parangon	32
Distant	32
Description des clusters	33

Présentation de la base de données

Variables

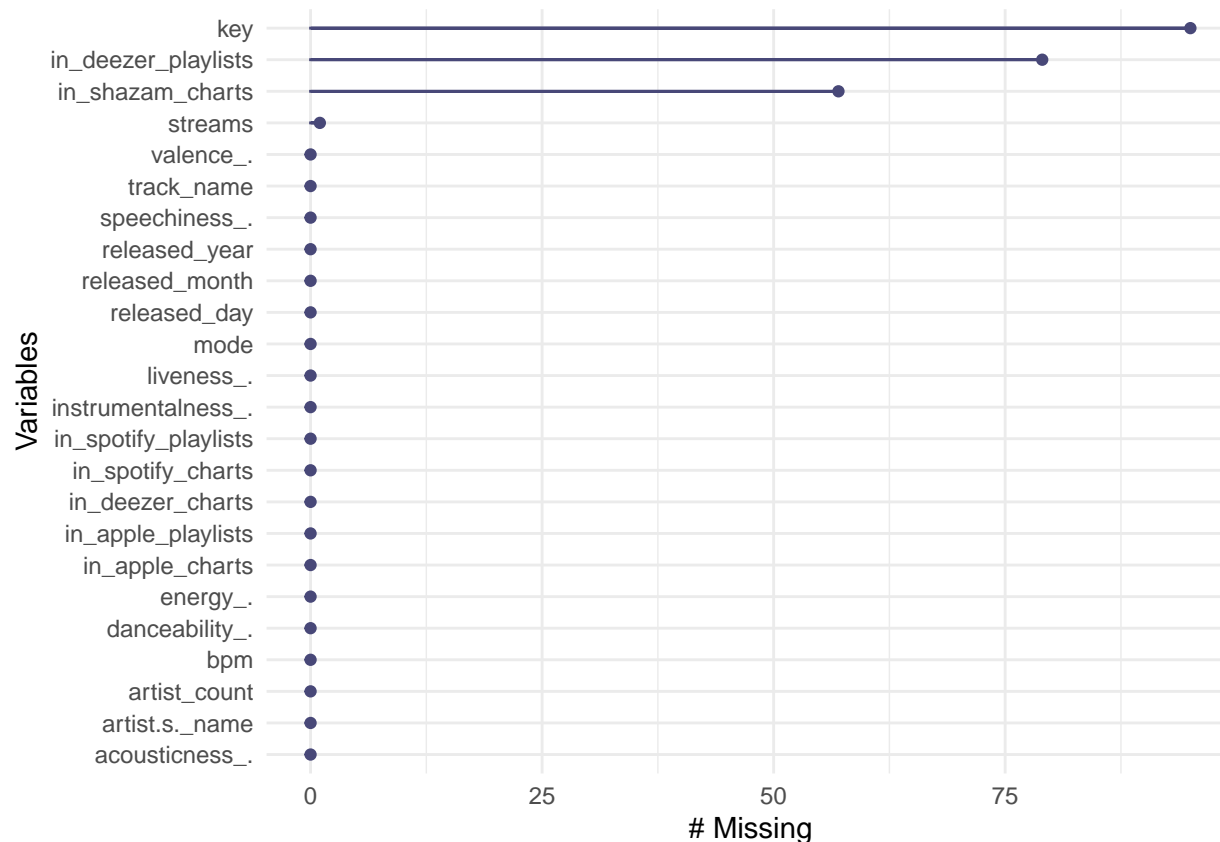
Nous avons travaillé sur la base de données spotify contenant 953 musiques et comportant 24 variables, dont 20 quantitatives et 4 qualitatives.

Les variables présentes dans la base de données sont : Nom de la musique et de l'artiste , nombre d'artistes présent sur la chanson, date de sortie , nombre de streams , nombre de playlist où la chanson est présente sur les différentes plateformes de streaming, classement dans les charts des différents services de streaming ainsi que certaines caractéristiques de la musique qui vont être l'étude principale de notre projet et qui sont :

- BPM (tempo de la musique)
- Key
- Mode (permet de distinguer les sons joyeux des sons mélancoliques , variable binaire)
- Danceability (permet de mesurer à quel point la musique est dansante , valeur comprise entre 0 et 100)
- Valence (permet de mesurer à quel point une musique est positive, valeur comprise entre 0 et 100)
- Energy (permet de mesurer le dynamisme/calme d'une musique, valeur comprise entre 0 et 100)
- Acousticness (permet de mesurer le degré d'instrument électronique/acoustique présent dans une musique, valeur comprise entre 0 et 100, si valeur faible alors la musique utilise principalement des instruments électroniques)
- Instrumentalness (permet de mesurer le degré de voix chantée dans la musique, valeur comprise entre 0 et 100, si valeur forte alors la musique n'est pas chantée)
- Liveness (permet de mesurer la probabilité que la musique soit enregistrée en studio ou bien plutôt en direct , valeur comprise entre 0 et 100, si valeur faible alors il y a de fortes chances que la musique soit enregistrée en studio)
- Speechiness (permet de mesurer la présence de voix comme des extraits de passages télévisées, de films , valeur comprise entre 0 et 100, si valeur faible alors il n'y pas d'éléments de discours)

Nous avons décidé de choisir ces variables pour notre étude afin de répondre à la problématique suivante : Existe-t-il des relations entre les caractéristiques d'une musique ?

Valeurs manquantes



On voit qu'il y a 95 valeurs manquantes dans `key`, 79 dans `deezer_playlist`, 57 dans `shazam_charts` et 1 dans `streams`, ce qui fait au total 232 valeurs manquantes.

Statistiques descriptives

Résumé des variables

Dans cette base de données, la musique avec le plus de streams est "Blinding Lights" de The Weekend avec 3.7 milliards d'écoutes. La base de données contient des musiques plutôt récentes avec 50% des musiques qui sont sorties après 2022. Le bpm moyen dans la base de donnée est de 122.5 ce qui correspond à un tempo rapide.

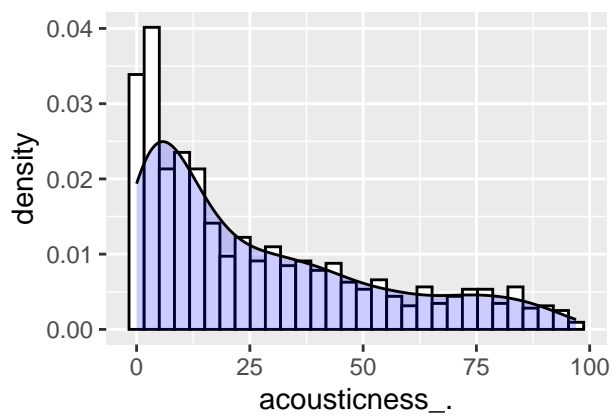
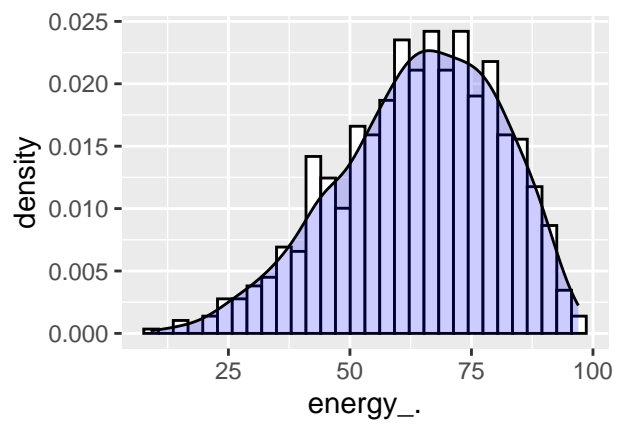
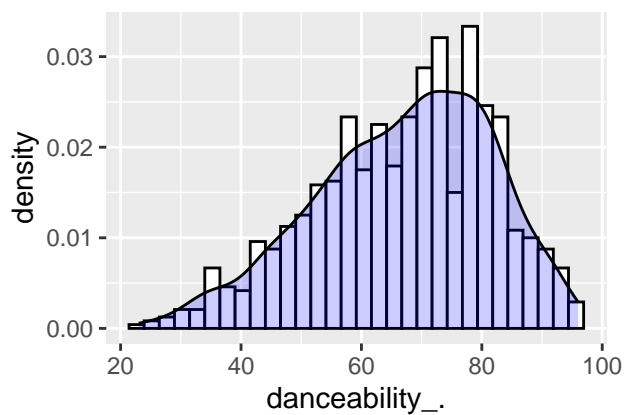
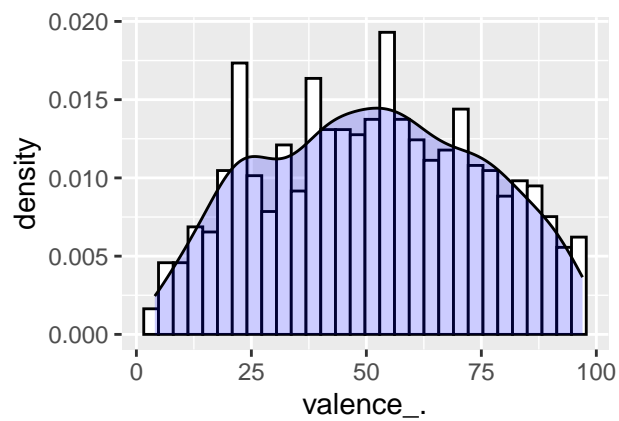
Ce tableau représente la fréquence d'apparition de chaque modalité pour les variables `clé` et `mode`. La clé la plus représentée est `C#` (c'est un do, tonalité neutre souvent perçue comme simple et directe) avec 14%. La moins représentée est `D#` (c'est un ré, peut avoir une sensation de légèreté et d'énergie) avec 3.5%. 58% des musiques ont une sonorité plutôt joyeuse et positive et 42% ont une sonorité plus sombre, introspective et mélancolique.

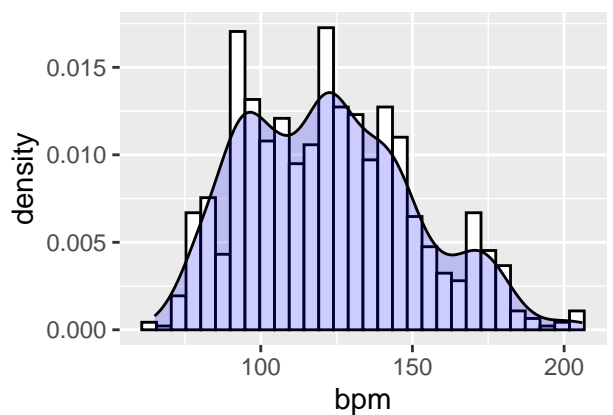
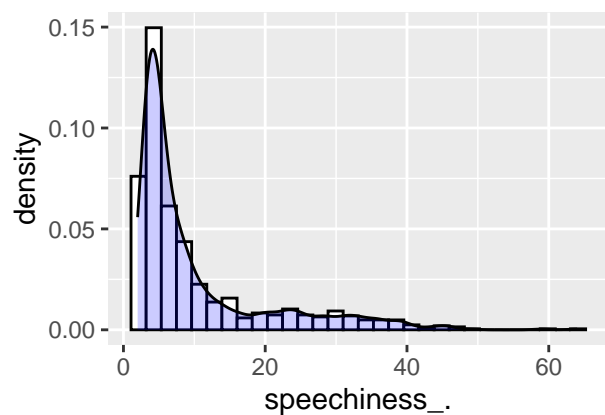
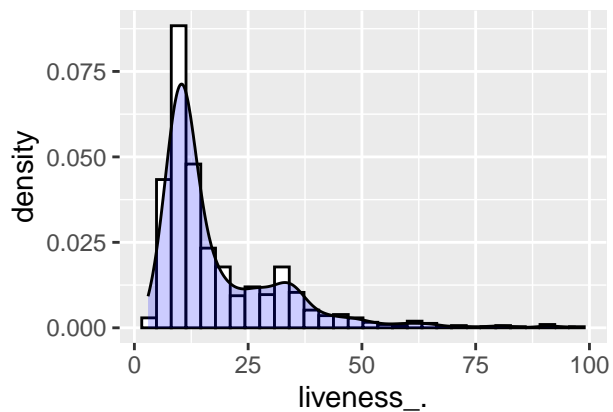
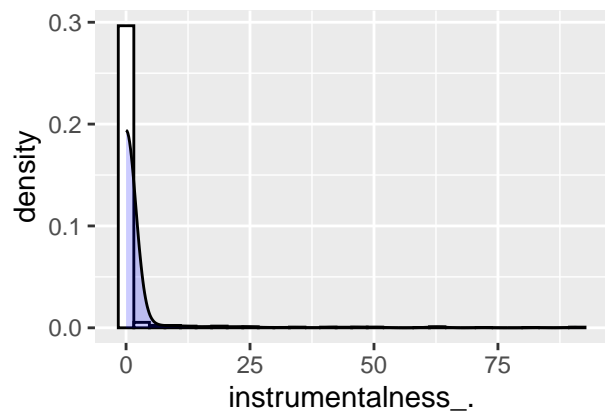
Table 1:

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
artist_count	8	0	1.6	0.9	1.0	1.0	8.0
released_year	50	0	2018.2	11.1	1930.0	2022.0	2023.0
released_month	12	0	6.0	3.6	1.0	6.0	12.0
released_day	31	0	13.9	9.2	1.0	13.0	31.0
in_spotify_playlists	879	0	5200.1	7897.6	31.0	2224.0	52 898.0
in_spotify_charts	82	0	12.0	19.6	0.0	3.0	147.0
streams	949	0	513 707 456.7	566 714 616.2	2762.0	290 228 626.0	3 703 895 074.0
in_apple_playlists	234	0	67.8	86.4	0.0	34.0	672.0
in_apple_charts	172	0	51.9	50.6	0.0	38.0	275.0
in_deezer_playlists	272	0	121.5	193.1	0.0	40.0	974.0
in_deezer_charts	34	0	2.7	6.0	0.0	0.0	58.0
in_shazam_charts	191	0	48.3	123.2	0.0	2.0	953.0
bpm	124	0	122.5	28.1	65.0	121.0	206.0
danceability_.	72	0	67.0	14.6	23.0	69.0	96.0
valence_.	94	0	51.4	23.5	4.0	51.0	97.0
energy_.	80	0	64.3	16.6	9.0	66.0	97.0
acousticness_.	98	0	27.1	26.0	0.0	18.0	97.0
instrumentalness_.	39	0	1.6	8.4	0.0	0.0	91.0
liveness_.	68	0	18.2	13.7	3.0	12.0	97.0
speechiness_.	48	0	10.1	9.9	2.0	6.0	64.0

		N	%
key	A	81	8.5
	A#	60	6.3
	B	85	8.9
	C#	137	14.4
	D	93	9.8
	D#	33	3.5
	E	64	6.7
	F	99	10.4
	F#	80	8.4
	G	117	12.3
	G#	104	10.9
mode	Major	550	57.7
	Minor	403	42.3

Graphiques de densité





Pour instrumentalness : la densité est concentrée vers 0, donc la plupart des musiques sont chantées, il y a peu de musiques spécifiquement instrumentales (sans paroles).

Pour liveness : la densité est concentrée entre 0 et 25, donc la plupart des sons ont été enregistrés en studio.

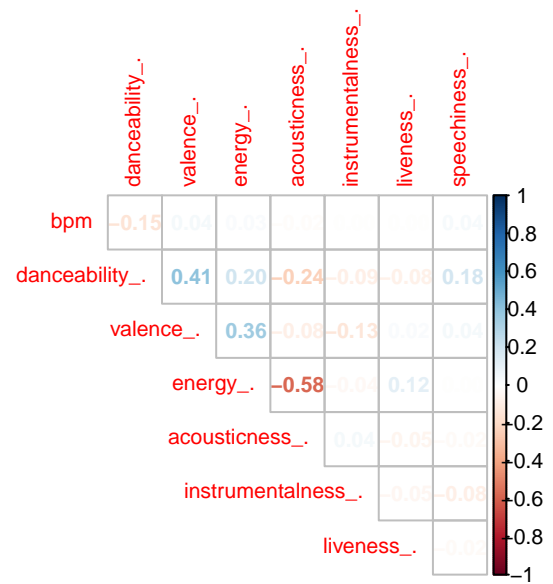
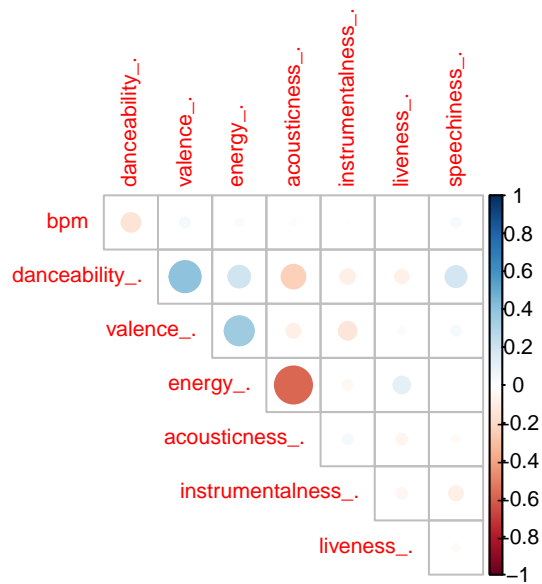
ACP

Nous avons décidé d'effectuer notre ACP sur les variables quantitatives qui caractérisent la musique, à savoir :

- BPM
- Danceability
- Valence
- Energy
- Acousticness
- Instrumentalness
- Liveness
- Speechiness

De plus, nous avons rajouté en tant que variables supplémentaires les variables liées aux streams , aux nombres de playlists dans lesquelles les musiques apparaissent et leurs places dans les charts dans les différents services de streaming, ainsi que les variables key et mode.

Matrice des corrélations



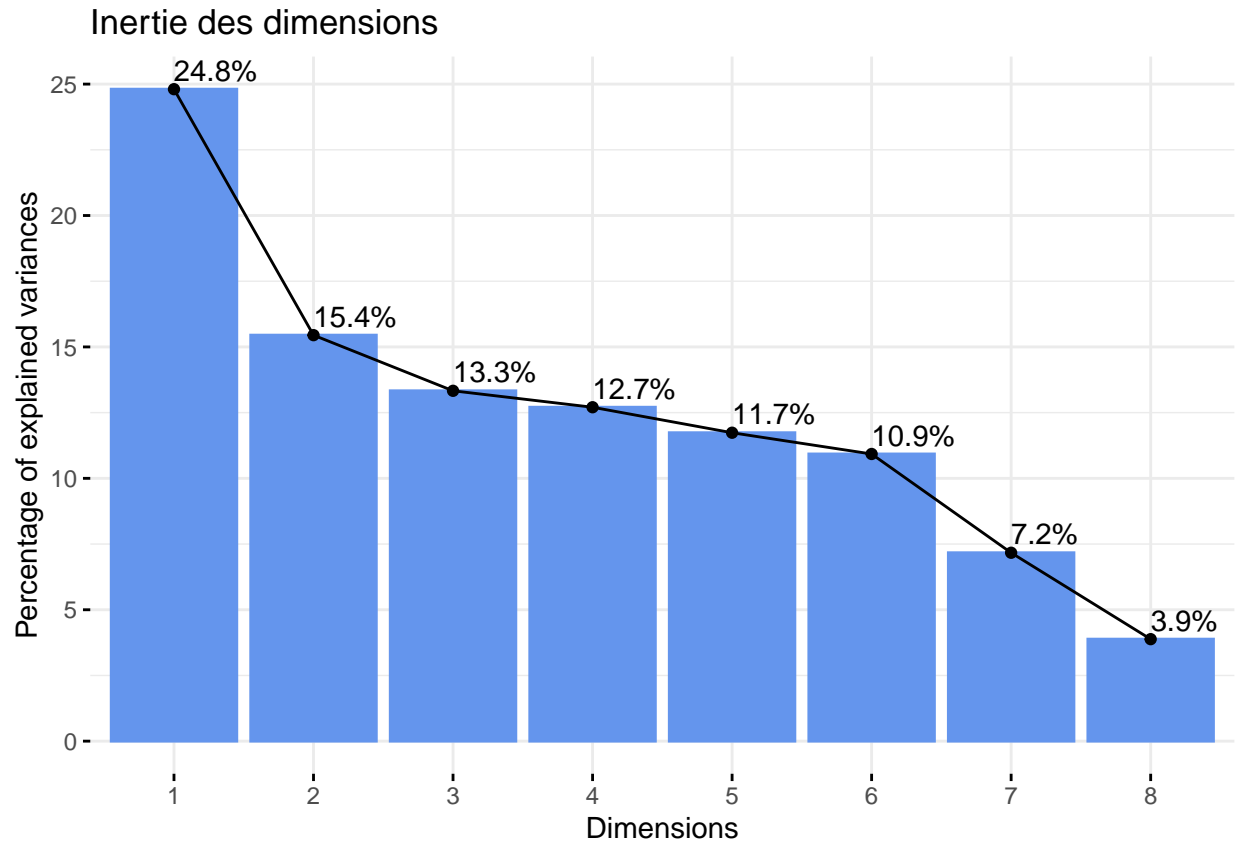
Voici la matrice des corrélations.

On peut voir notamment qu'il y a :

- Une corrélation positive (0.41) entre danceability et valence
- Une corrélation négative (-0.6) entre energy et acousticness
- Une corrélation positive (0.36) entre valence et energy

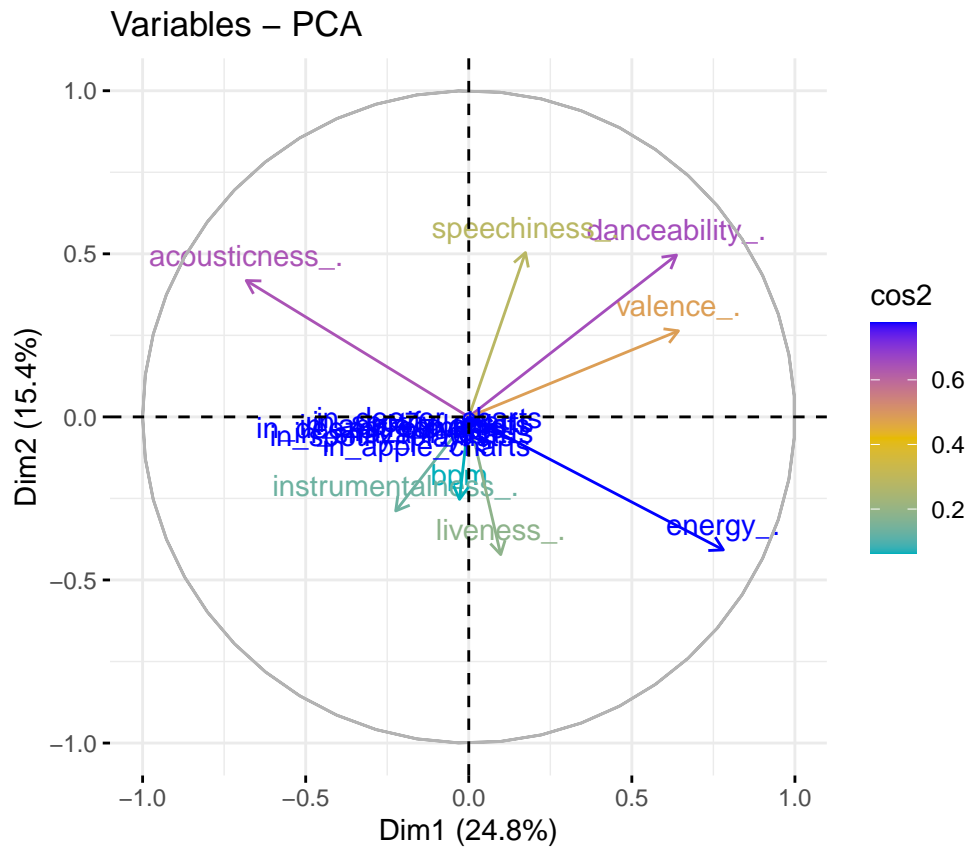
Globalement il existe quelques corrélations mais qui ne sont pas très fortes.

Inertie



On garde les 2 premiers axes qui représentent 40.2% de l'inertie.

Description des axes



Les variables qui expliquent l'axe 1 sont, à priori :

- Energy
- Acousticness
- Valence
- Danceability

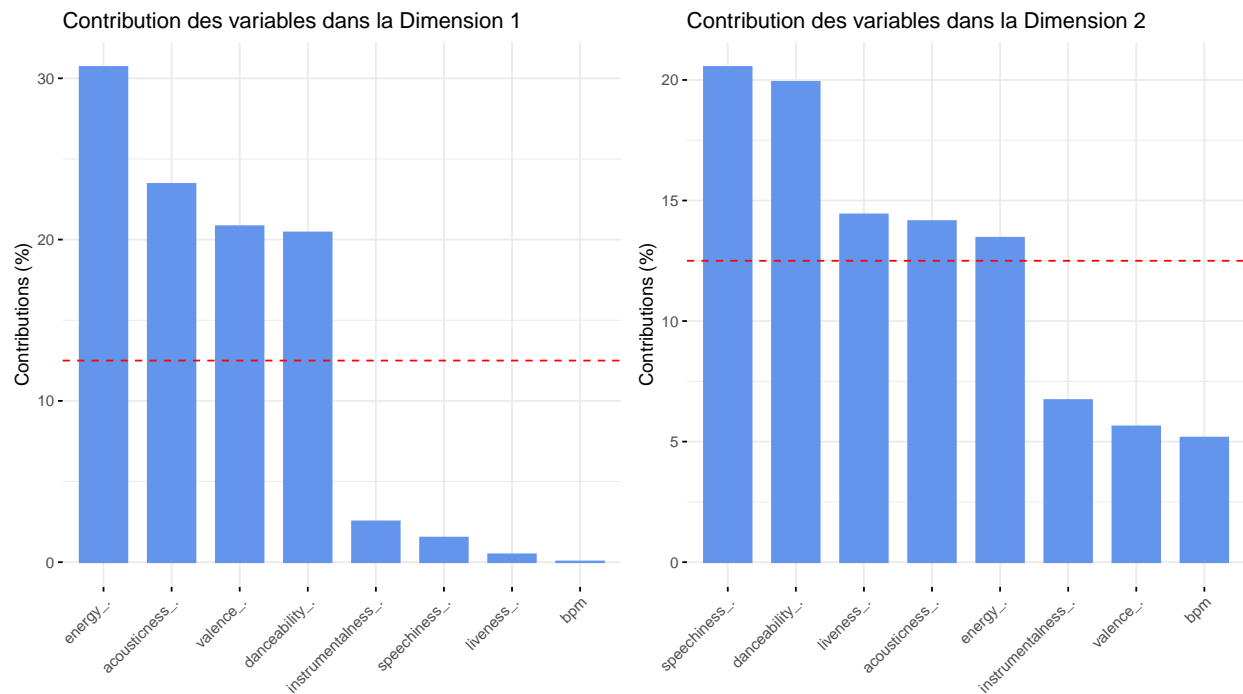
De plus, on peut voir que energy et acousticness ont l'air de s'opposer.

Les variables qui expliquent l'axe 2 sont, à priori :

- Speechiness
- Liveness
- Danceability
- Acousticness
- Energy

Les variables BPM et instrumentalness sont quant à elles mal représentées dans ces 2 premières dimensions donc on ne peut pas en dire davantage sur elles.

De plus, on ne peut pas tirer grand chose des variables supplémentaires qui ne nous donne pas d'informations supplémentaires.



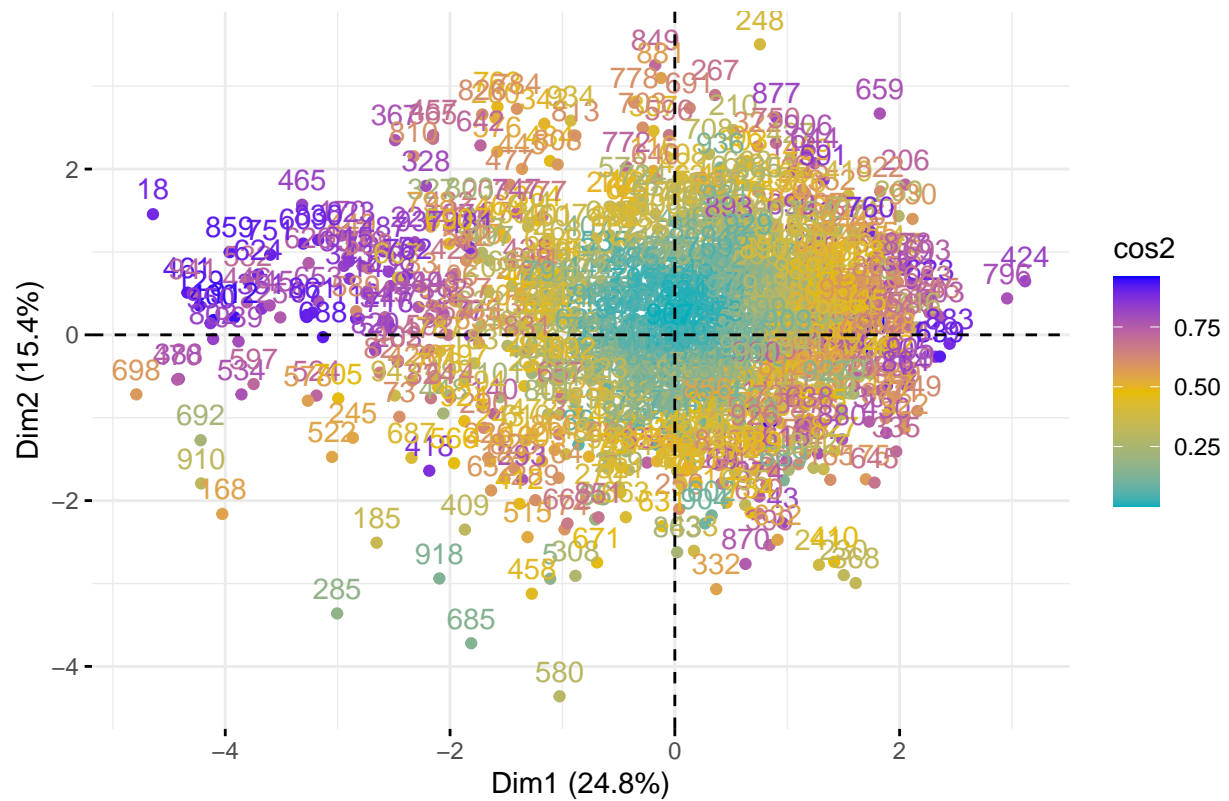
Les variables qui contribuent le plus à l'axe 1 sont energy, acousticness, valence et danceability, ce qui vient bien confirmer ce qu'on pouvait voir sur le graphique des variables.

Sur la partie positive de l'axe 1, on va retrouver des musiques avec une forte energy, une forte valence et danceability mais avec une faible acousticness (car il y a opposition entre energy et acousticness).

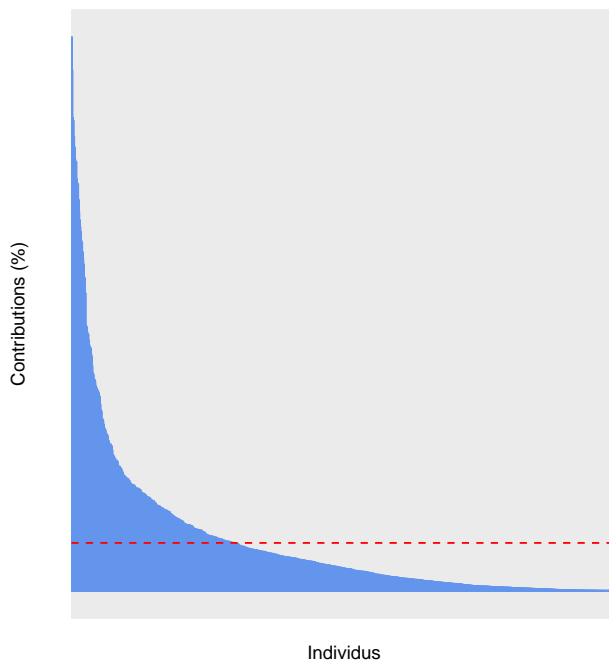
Les variables qui contribuent le plus à l'axe 2 sont surtout speechiness et danceability, et dans une moindre mesure acousticness, energy et liveness qui s'oppose légèrement à speechiness.

Sur la partie positive de l'axe 2, on retrouve les musiques avec une forte speechiness et danceability mais une faible liveness.

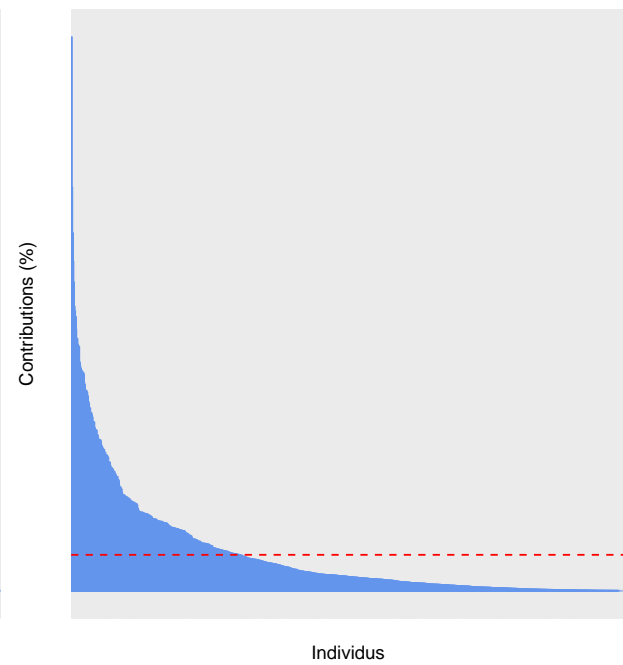
Individuals – PCA



Contribution des individus dans la Dimension 1



Contribution des individus dans la Dimension 2



	danceability__.	valence__.	energy__.	acousticness__.	instrumentalness__.	liveness__.	speechiness__.
424	95	91	89	6	0	31	24
796	91	91	96	3	0	3	9
18	44	14	9	96	0	10	3
859	42	32	20	94	0	11	4

	danceability__.	valence__.	energy__.	acousticness__.	instrumentalness__.	liveness__.	speechiness__.
659	90	79	71	26	0	7	37
870	44	41	92	0	0	44	7

Voici 2 extraits de la base de données afin d'appuyer nos résultats.

Le 1er tableau correspond aux individus situés aux extrémités de l'axe 1 avec d'un côté les individus 424 et 796 du côté positif de l'axe 1 et les individus 18 et 859 du côté négatif. On voit bien que les individus 659 et 870 ont de fortes valeurs pour les variables danceability, valence et energy et une faible valeur pour acousticness.

De même pour les individus 18 et 859 qui eux ont une forte acousticness et donc une faible energy, valence et danceability.

Pour le 2ème extrait, l'individu 659 se situe du côté positif de l'axe 2 a une forte speechiness, une forte danceability et une faible liveness. Tandis que pour l'individu 870, c'est l'inverse, c'est à dire qu'il a une faible speechiness et une faible danceability et une forte liveness.

AFC

On a ensuite réalisé une AFC, on a d'abord essayé de faire l'analyse sur key et mode car c'était les 2 seules variables qualitatives mais l'analyse ne donnait rien.

On avait peu de variables qualitatives donc a transformé des quantitatives utilisées pour l'ACP en qualitatives en les découpant chacune en 4.

Chaque variable est découpée en 4 modalités : - quand la valeur est inférieure à 25, - quand la valeur est entre 25 et 50, + quand la valeur est entre 50 et 75, ++ quand la valeur est supérieure à 75. Sauf pour le BPM : - inférieur à 60, - entre 60 et 120, + entre 120 et 167, ++ supérieur à 167.

L'analyse de l'AFC ne donnait rien non plus lorsqu'elle était réalisée avec key et une nouvelle qualitative, de même avec mode, le test du chi-2 ne nous permettait pas de conclure à une corrélation entre les 2 variables testées.

```
##
## Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 369.35, df = 16, p-value < 2.2e-16
```

Au final, on a fait l'analyse sur les variables energy et acousticness car elles étaient corrélées (pvalue < 2.2e-16).

```
##
## Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 369.35, df = 16, p-value < 2.2e-16
```

```
##
##          acoustique (--) acoustique (-) acoustique(+) acoustique(++) Sum
## energy (--)          9.435467      3.391396      1.897167      1.275971  16
## energy (-)         104.379853     37.517314     20.987408     14.115425 177
## energy (+)         287.192025    103.225603     57.745016     38.837356 487
## energy (++)         160.992655     57.865687     32.370409     21.771249 273
## Sum                562.000000    202.000000    113.000000     76.000000 953
```

Profil lignes

	acoustique (-)	acoustique (-)	acoustique(+)	acoustique(++)	Sum
energy(-)	6.25	0.00	12.50	81.25	100
energy(-)	25.42	19.21	26.55	28.81	100
energy(+)	60.78	25.05	11.70	2.46	100
energy(++)	80.59	16.85	2.56	0.00	100

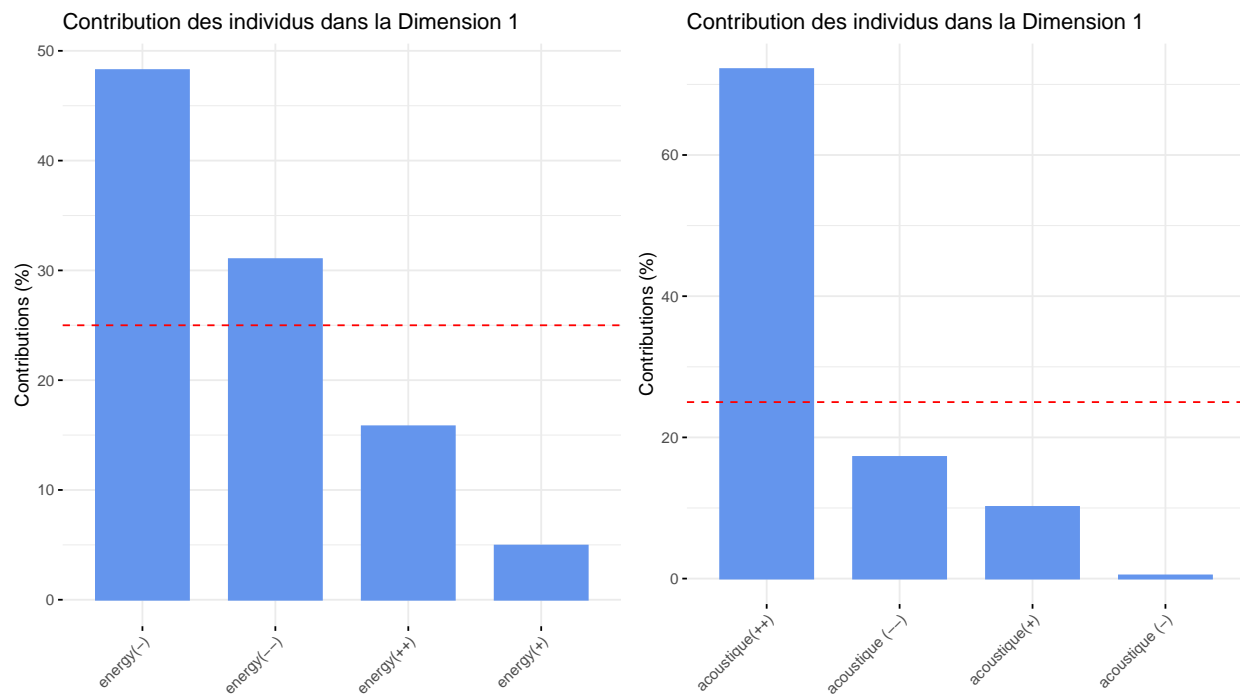
Les musiques qui ont une énergie très faible sont principalement regroupées avec celles qui ont une acousticness très élevée et inversement.

Profil colonnes

	acoustique (-)	acoustique (-)	acoustique(+)	acoustique(++)
energy(-)	0.18	0.00	1.77	17.11
energy(-)	8.01	16.83	41.59	67.11
energy(+)	52.67	60.40	50.44	15.79
energy(++)	39.15	22.77	6.19	0.00
Sum	100.00	100.00	100.00	100.00

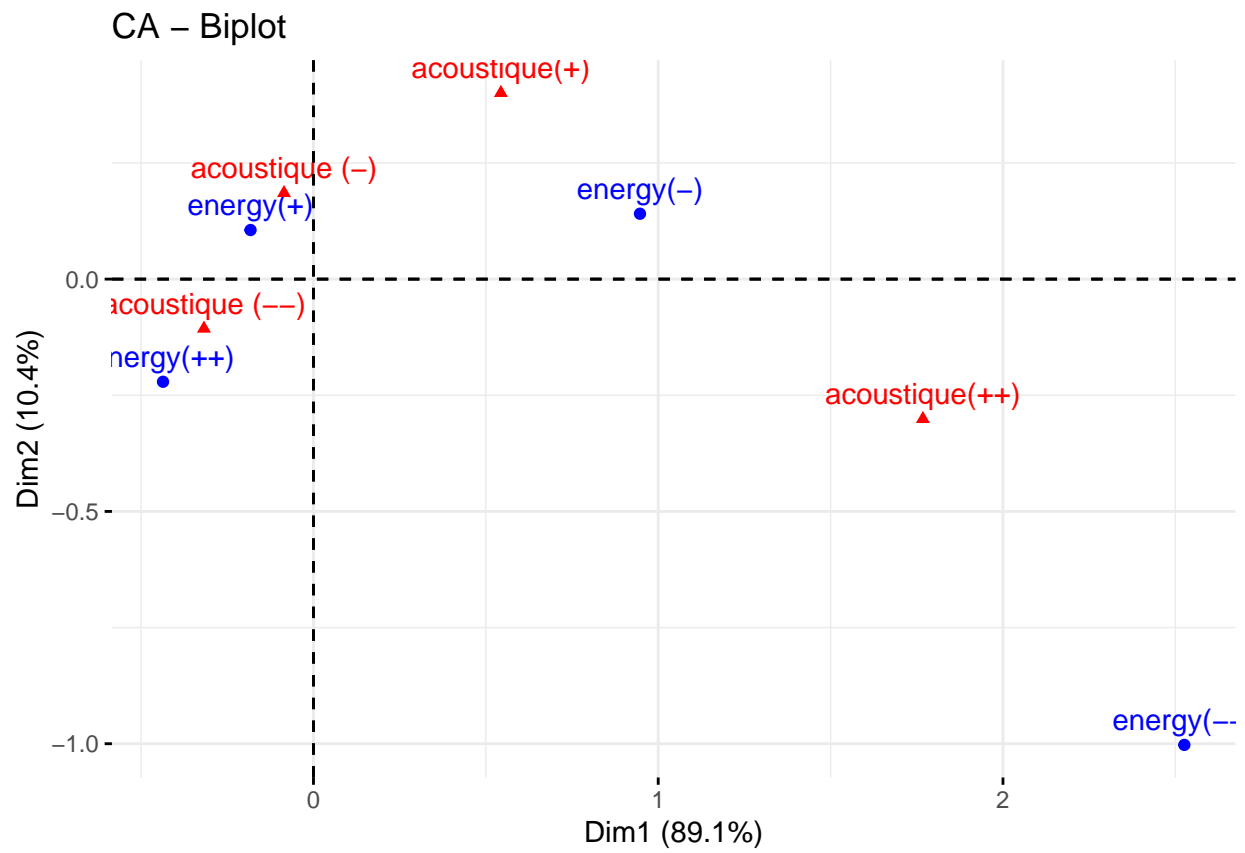
On voit que c'est principalement les musiques avec energy + et acoustictness - qui s'associent.

Contributions aux lignes et aux colonnes

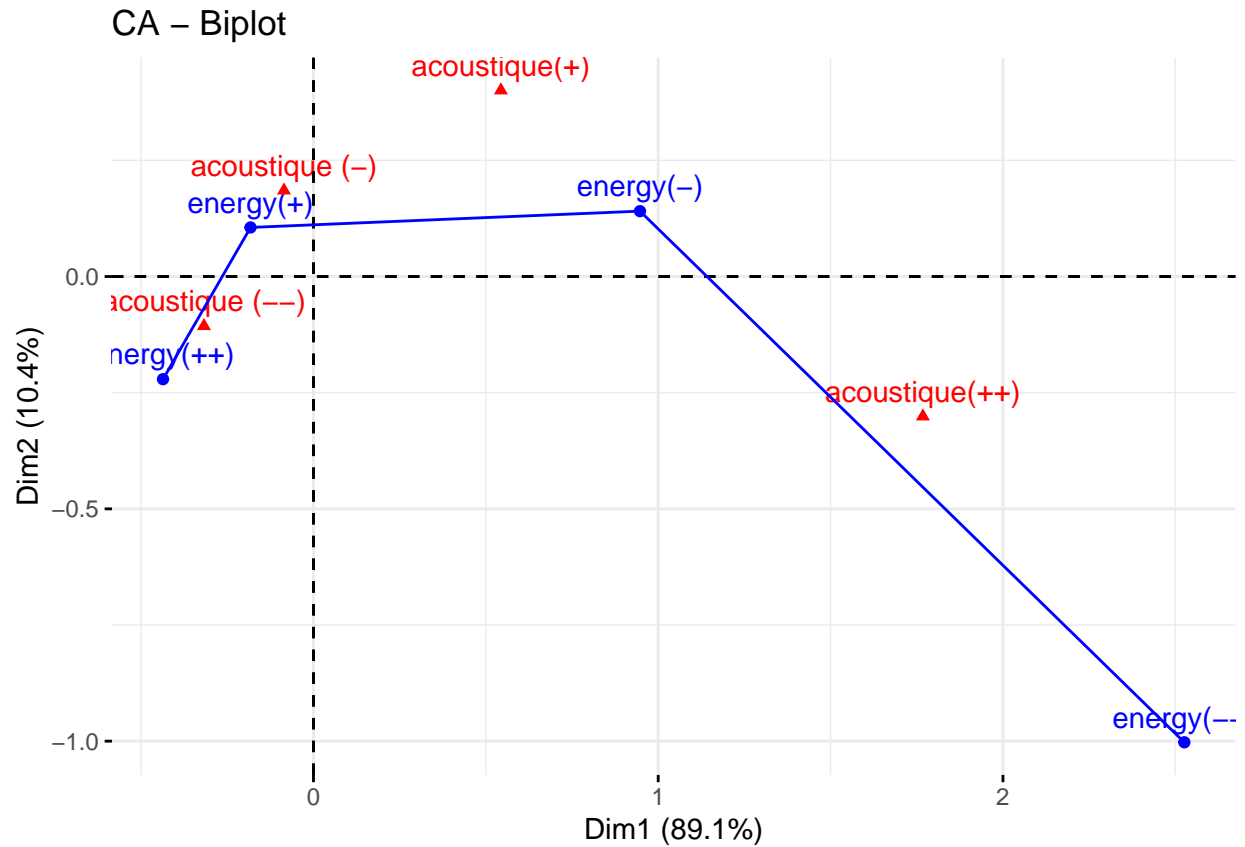


Pour la contribution des lignes, c'est surtout energy - et energy - qui contribuent le plus tandis que pour les colonnes c'est surtout acoustictness ++.

Visualisation graphique



Sur le graphique, on voit que l'axe 1 représente l'opposition entre les modalités - et +, et l'axe 2 représente plutôt l'opposition entre les modalités ++ / + et les modalités - / -, par exemple energy - opposée à energy -



En reliant les modalités d'une variable (ici energy), on peut observer un effet Guttman, les modalités sont ordonnées, de energy ++ à gauche à energy - à droite, inversement pour la variable acoustique.

On peut dessiner des groupes par exemple energy ++ avec acoustique -, energy + avec acoustique -, energy - avec acoustique + et energy - avec acoustique ++.

ACM

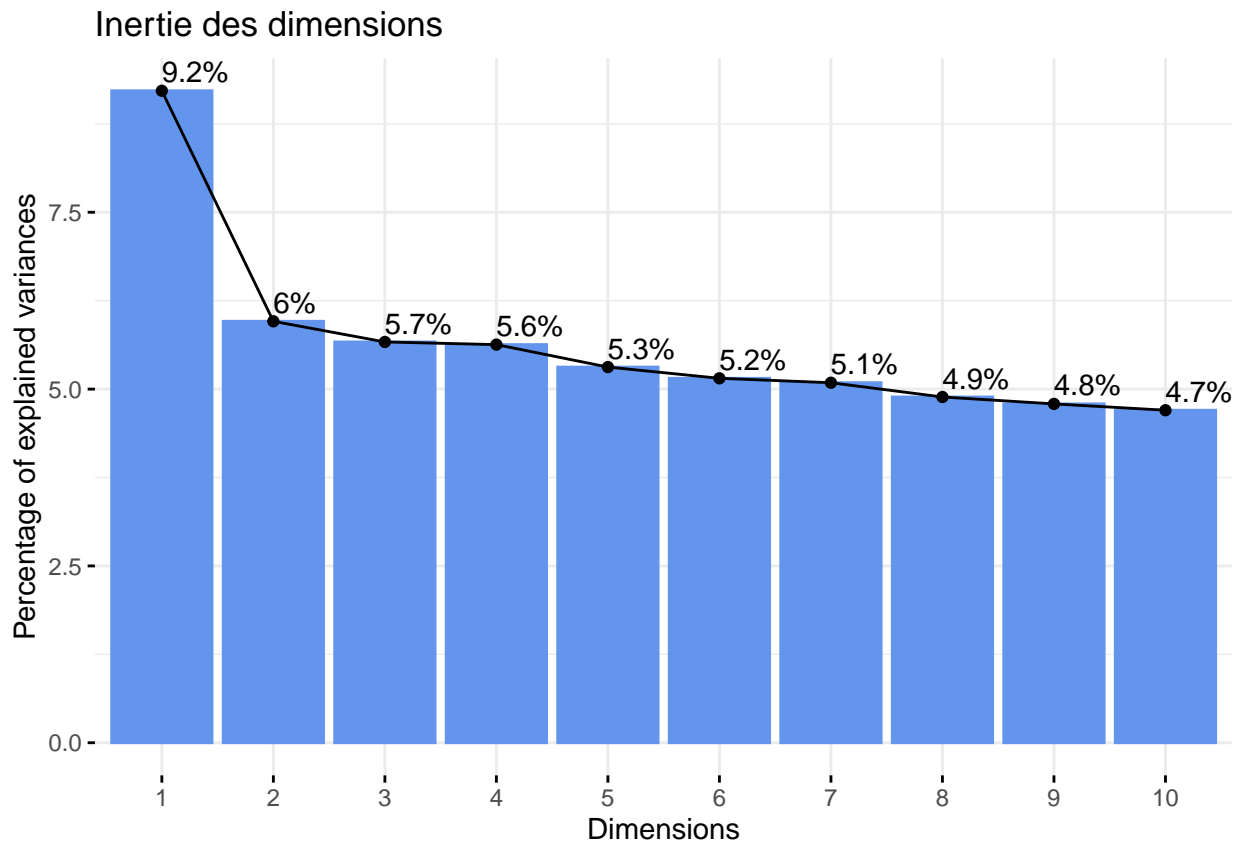
Pour réaliser notre ACM, on a choisi d'utiliser les variables qui caractérisent la musique bpm, danceability, valence, energy, acousticness, instrumentality, liveness et speechiness, que l'on a coupée en 4.

L'axe 1 permet de distinguer les individus qui prennent la modalité rare dansant – et energy –, et d'autre part permet de distinguer un groupe qui se détache des autres en prenant acousticness ++.

L'axe 2 met en avant les individus qui prennent la modalité rare speechiness + (individu 248, 936) et dansant – (individu 470, 448). Un groupe se détache légèrement et prend la modalité acousticness +

Sur l'axe 3, on voit un groupe se détacher et prendre la modalité instrumentality –.

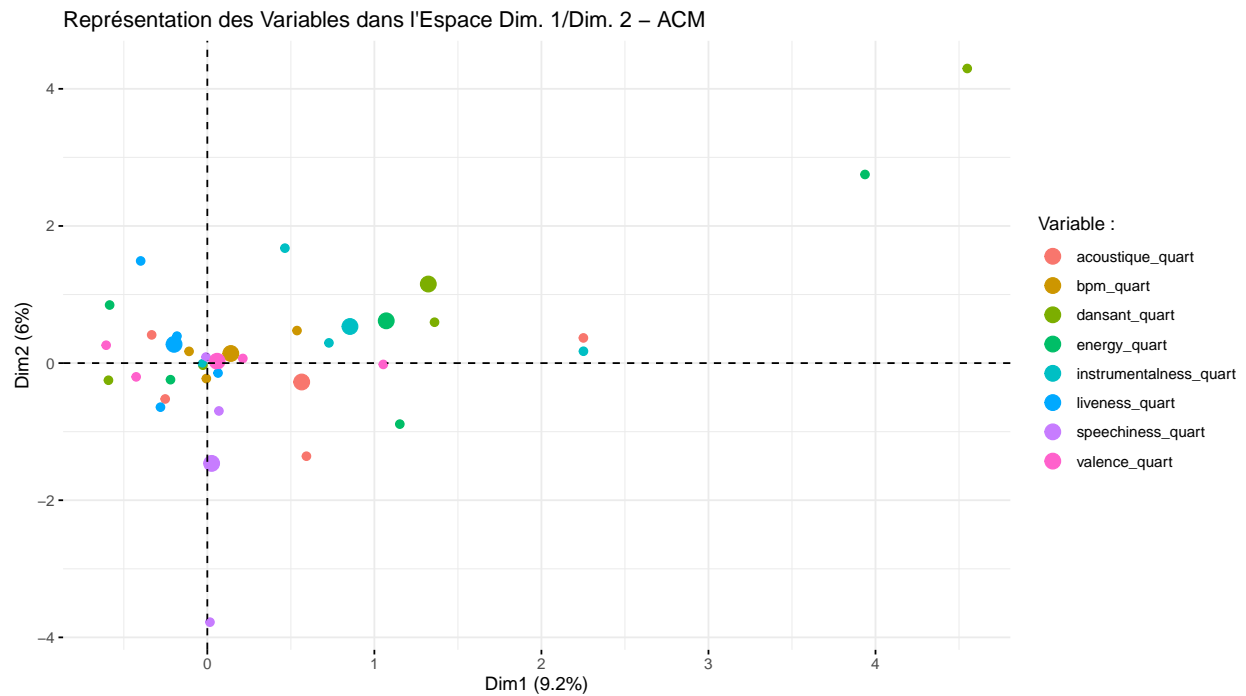
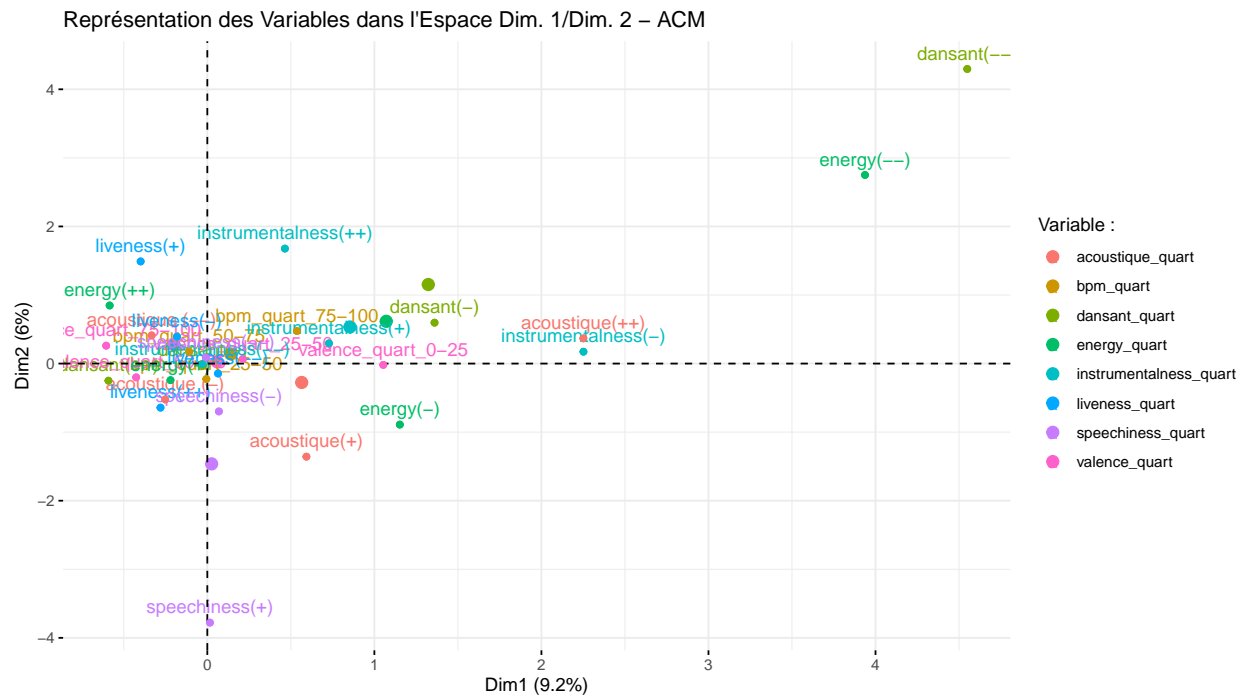
L'axe 4, on distingue un groupe avec instrumentality ++ en bas, le groupe dansant – en haut



Nous avons choisi les 4 premiers axes ce qui nous permet de garder 26.47% d'inertie

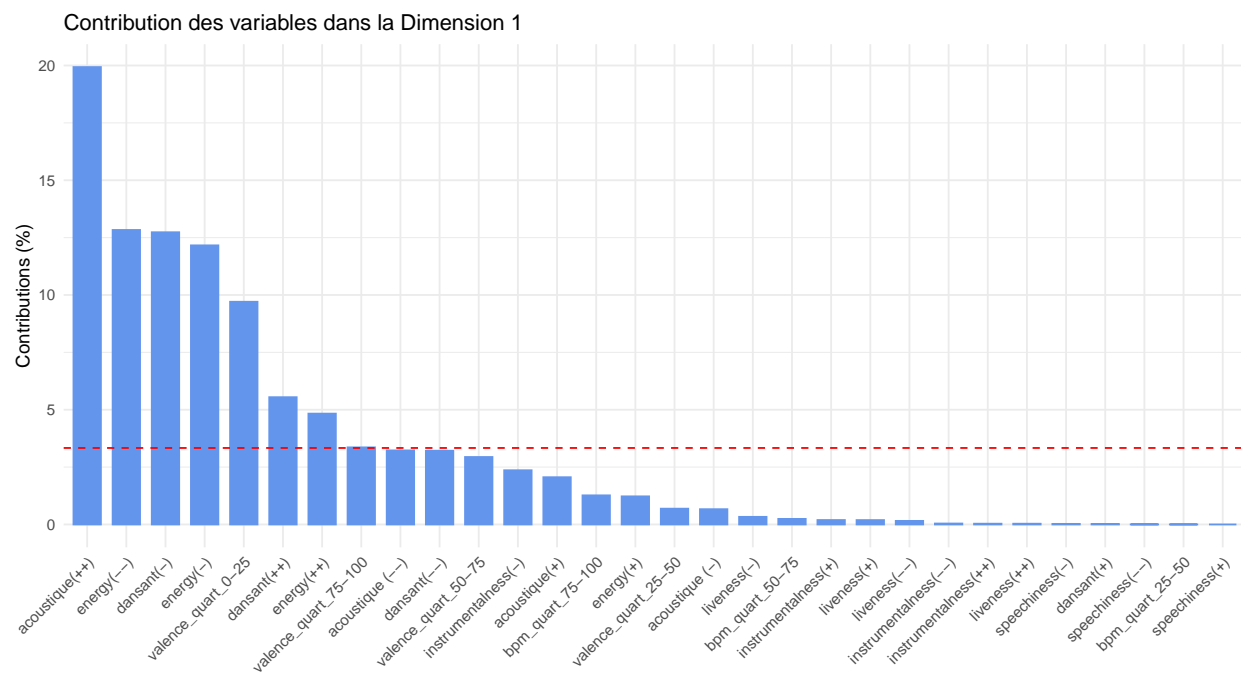
Dim 1 / Dim 2

Variables

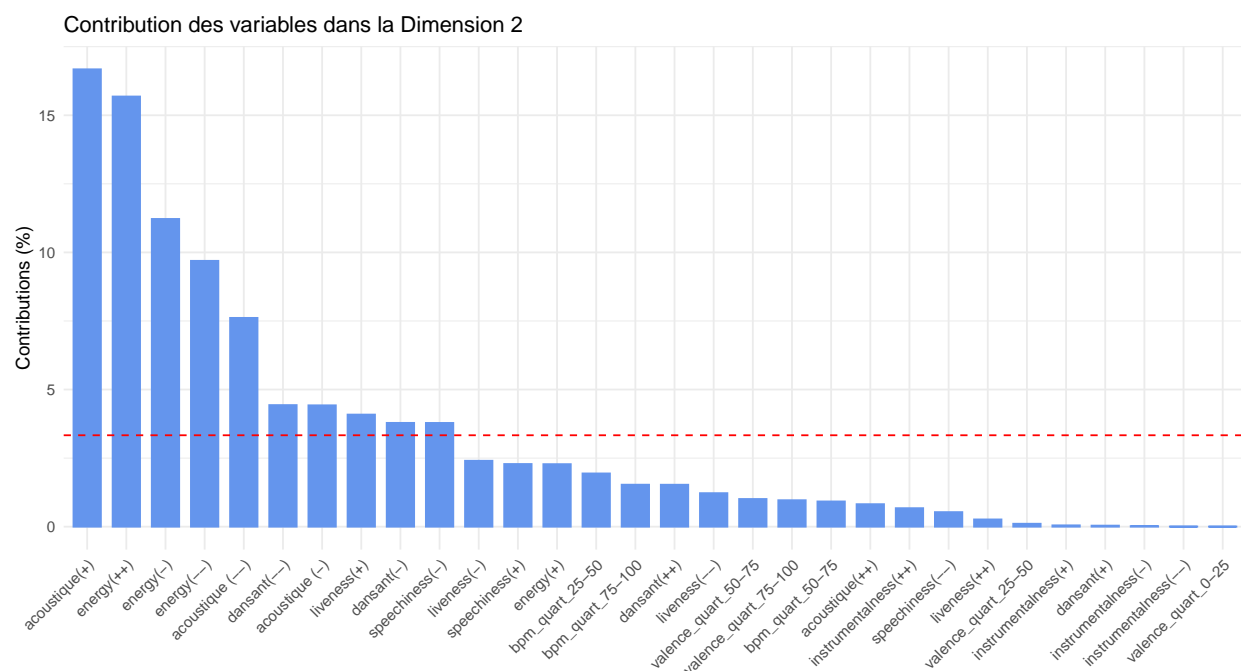


Nous pouvons apercevoir deux groupes de modalités se détacher :

- le groupe dansant(-) et energy(-) en haut à droite
- speechiness(+) en bas sur l'axe 2

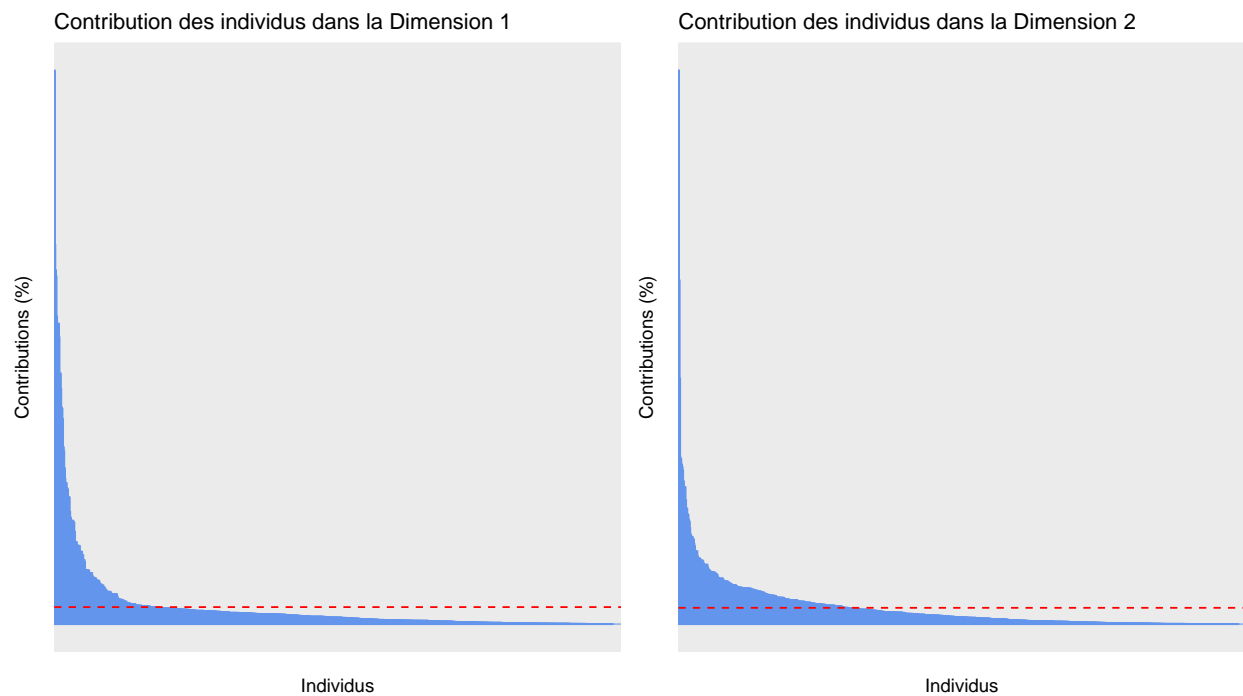
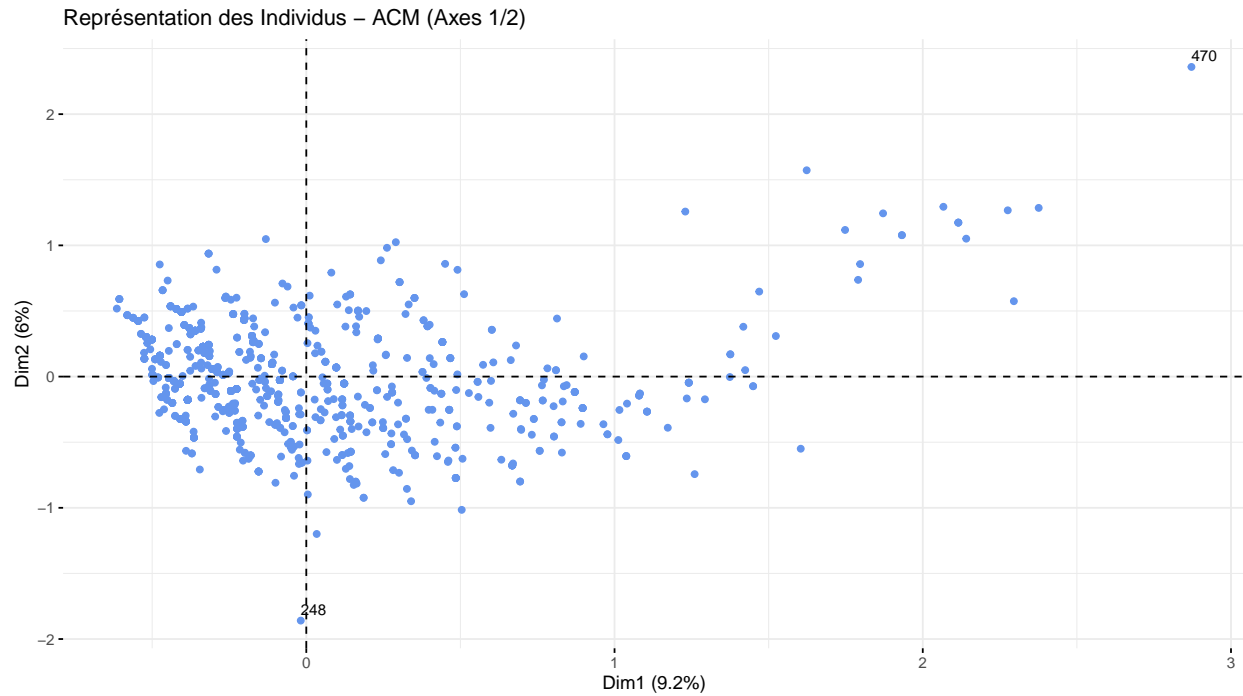


Les modalités qui contribuent le plus dans la dimension 1 sont acoustique(++) (19.93%), energy(-) (12.83%), dansant(-) (12.73%), energy(-) (12.16%) et valence(-) (9.7%).



Les modalités qui contribuent le plus dans la dimension 1 sont acoustique(+) (16.68%), energy(++) (15.68%), energy(-) (11.22%), energy(-) (9.69%) et acoustique(-) (7.61%)

Individus



Pour la dimension 1 nous pouvons voir une tendance particulière, il n'y a pas une grosse proportion qui ont une contribution au dessus de la moyenne. Pour la dimension 2 nous avons une tendance qui se généralise avec près de un tiers de la proportion d'individus qui ont une contribution au dessus de la contribution moyenne.

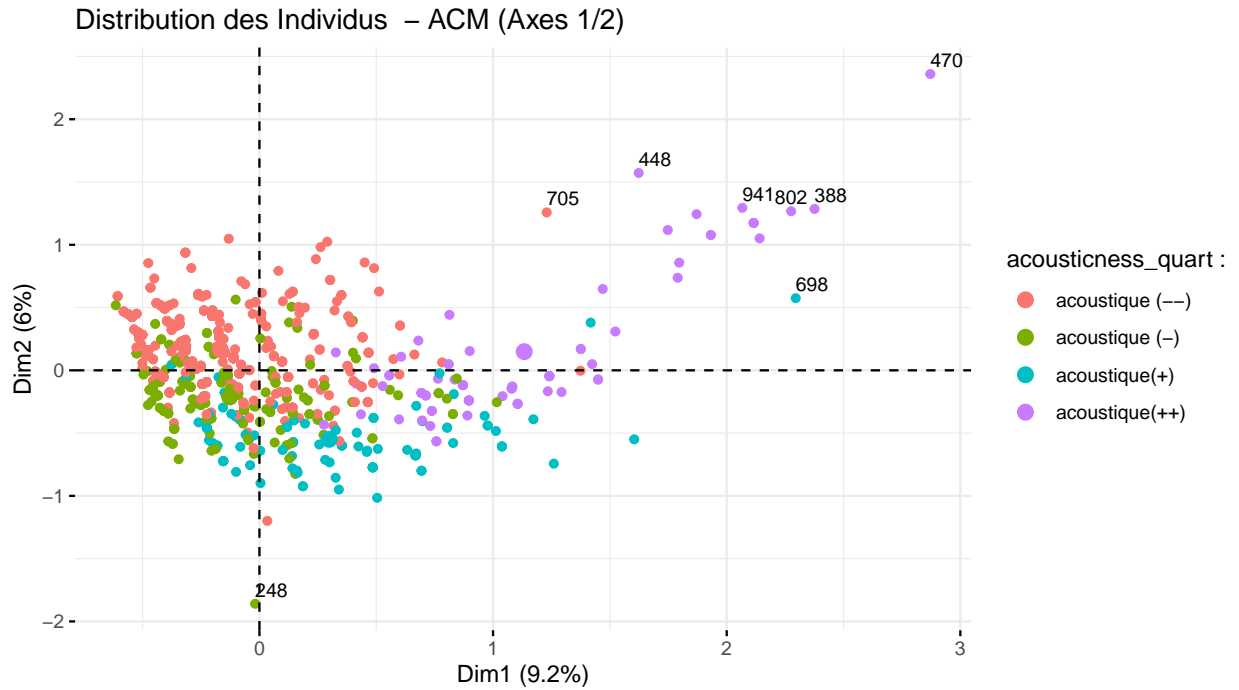
Les individus qui contribuent le plus à la dimension 1 sont le 470, le 388, le 698 et le 802; et les individus qui contribuent le plus à la dimension 2 sont le 470, le 248, le 448 et le 941.

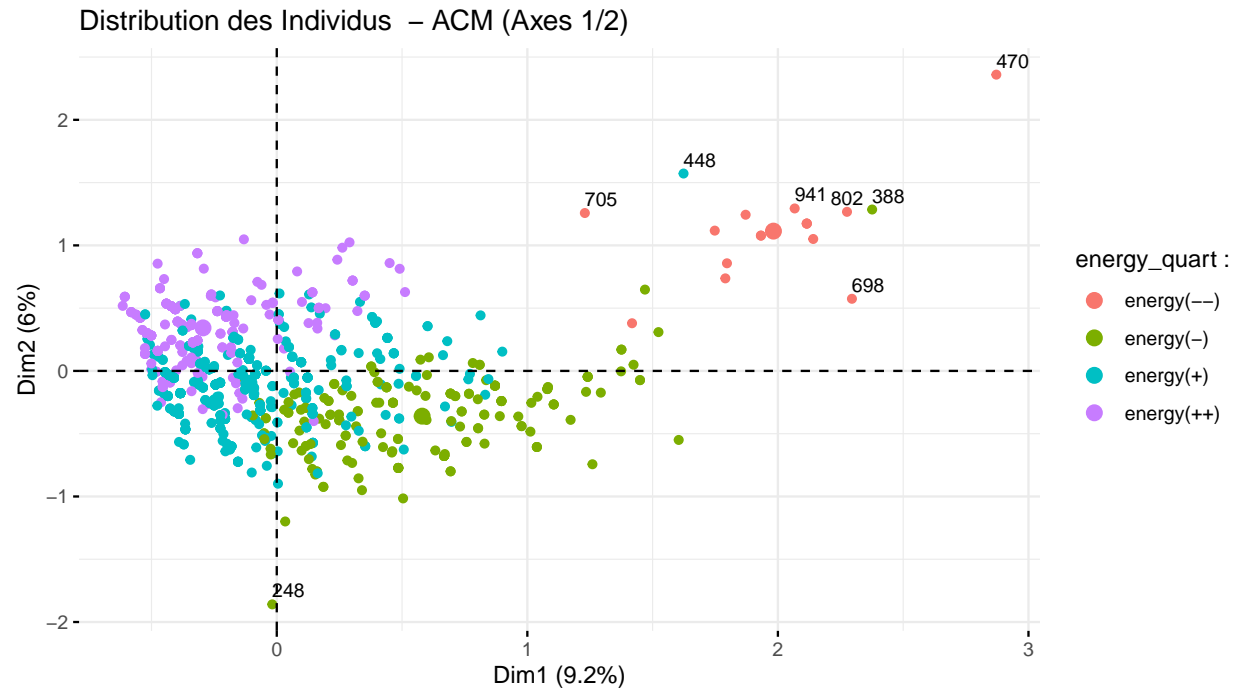
Table 2: Tableau des contributions des individus dans la Dim1

Individu	Contribution en %
470	3.413799
388	2.337171
698	2.181264
802	2.143863
18	1.897510
400	1.852263
461	1.852263
950	1.852263
941	1.768598
532	1.545532
859	1.545532
446	1.449474

Table 3: Tableau des contributions des individus dans la Dim2

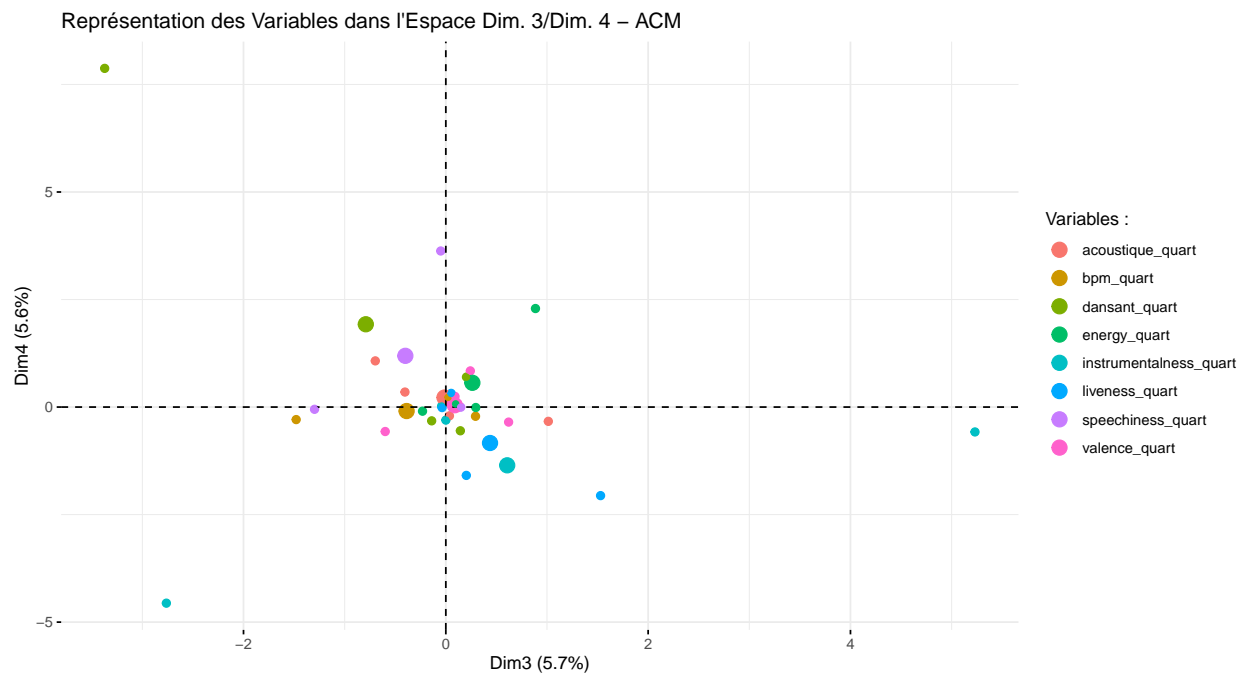
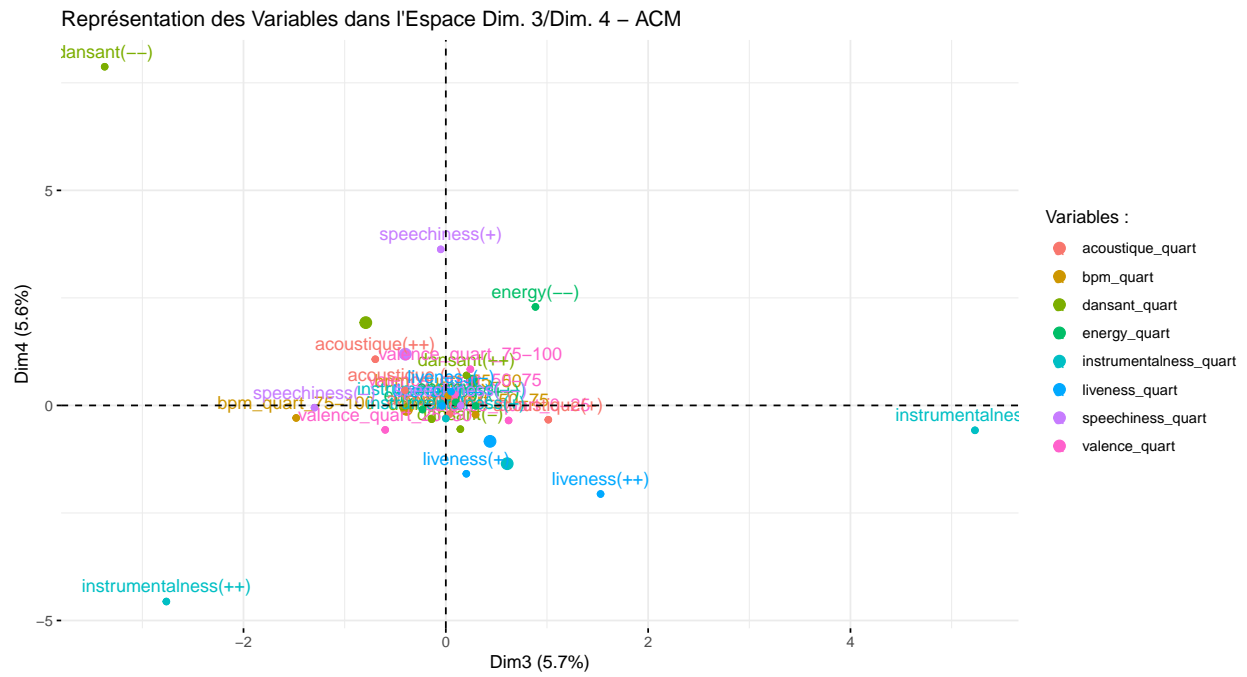
Individu	Contribution en %
470	3.5663593
248	2.2141231
448	1.5829516
941	1.0722986
388	1.0584009
802	1.0282988
705	1.0128825
446	0.9910731
936	0.9209655
400	0.8819739
461	0.8819739
950	0.8819739

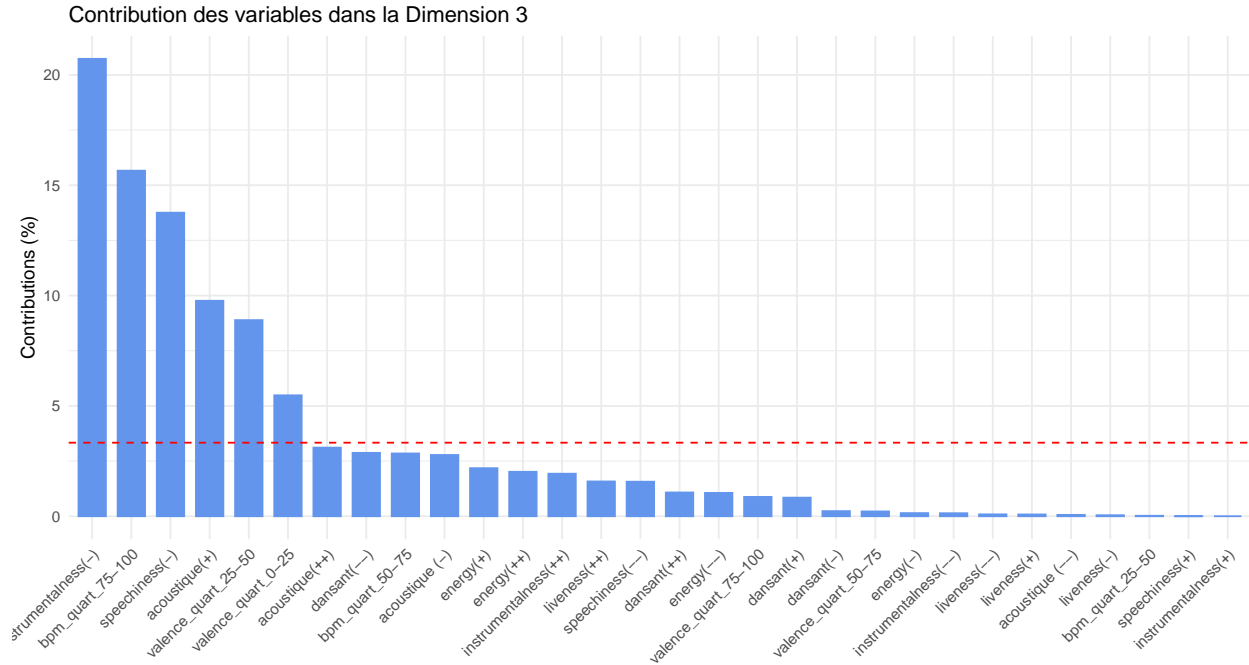




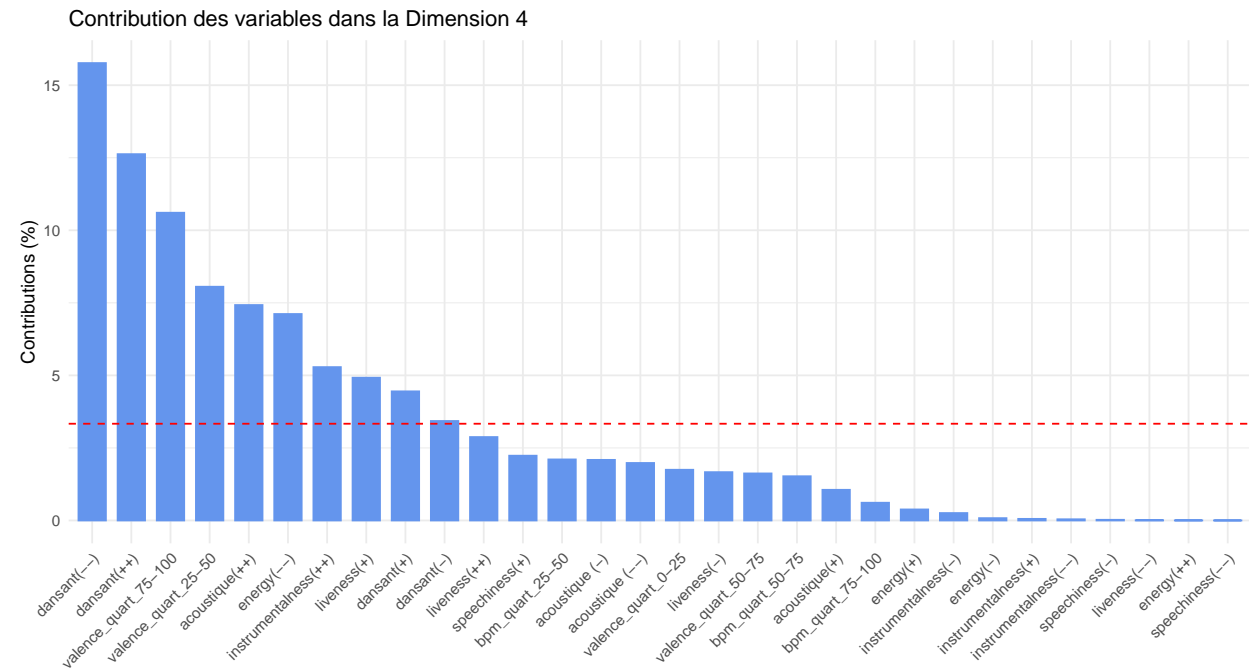
Dim 3 / Dim 4

Variables



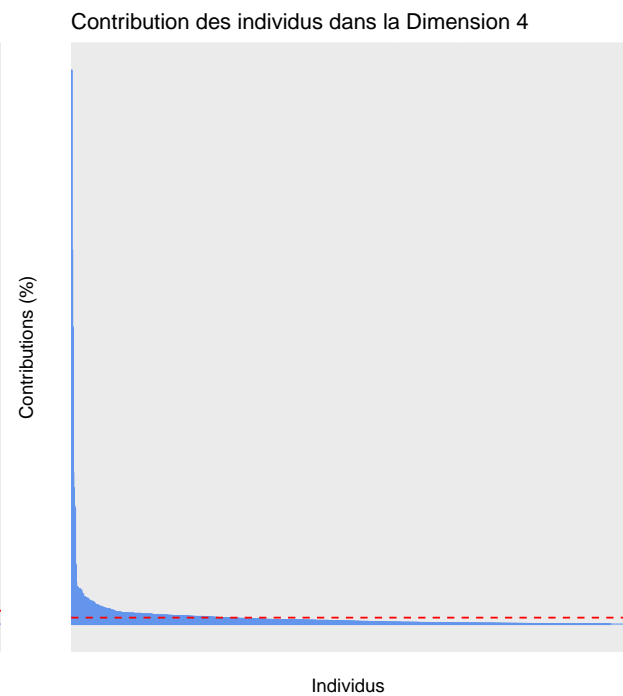
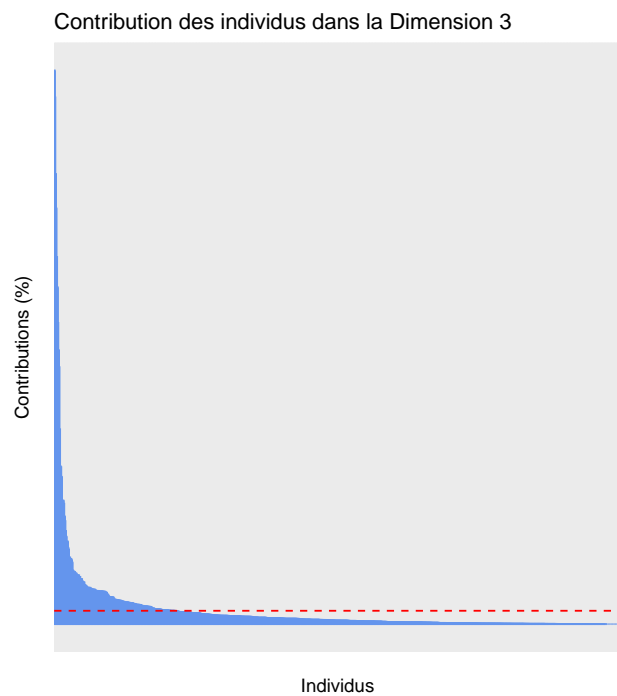
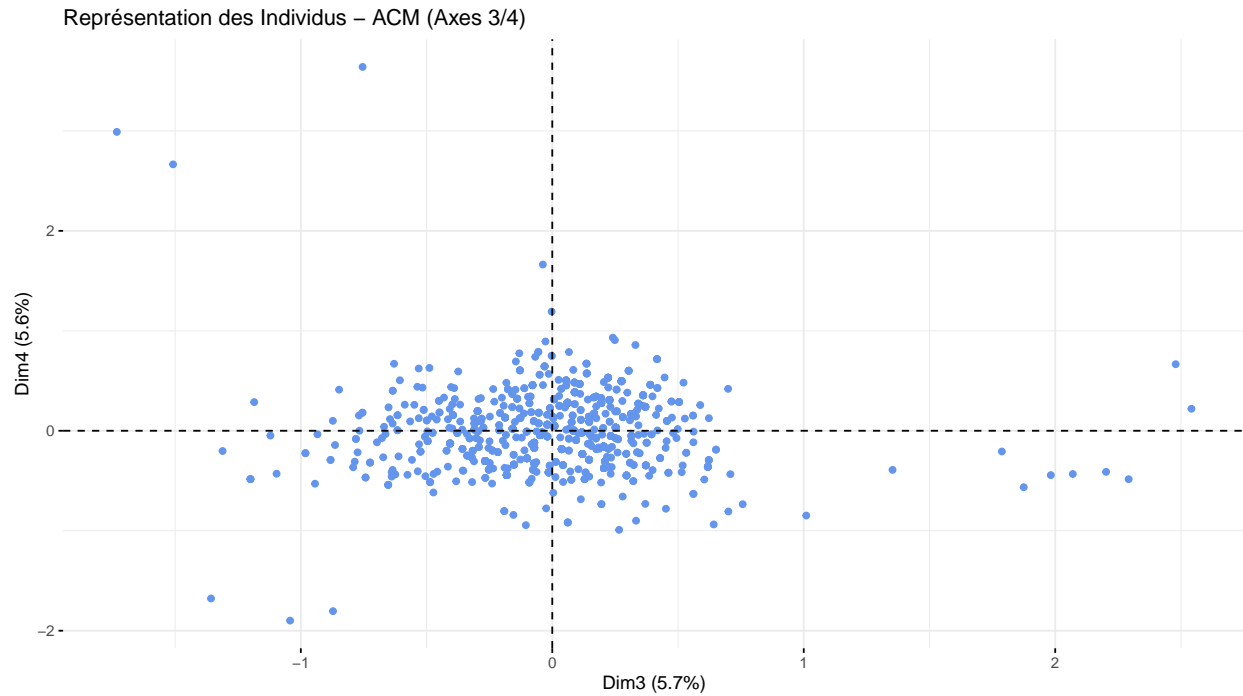


Les modalités qui contribuent le plus dans la dimension 1 sont instrumentalness(-) (20.73%), bpm(++) (15.66%), speechiness(-) (13.76%), acoustique(+) (9.76%) et valence(-) (8.89%)



Les modalités qui contribuent le plus dans la dimension 1 sont dansant(-) (19.93%), dansant(++) (12.83%), valence(++) (12.73%), valence(-) (12.16%) et acoustique(++) (9.7%)

Individus



Pour la dimension 3 et 4, nous pouvons voir une tendance particulière, il n'y a pas une grosse proportion qui ont une contribution au dessus de la moyenne.

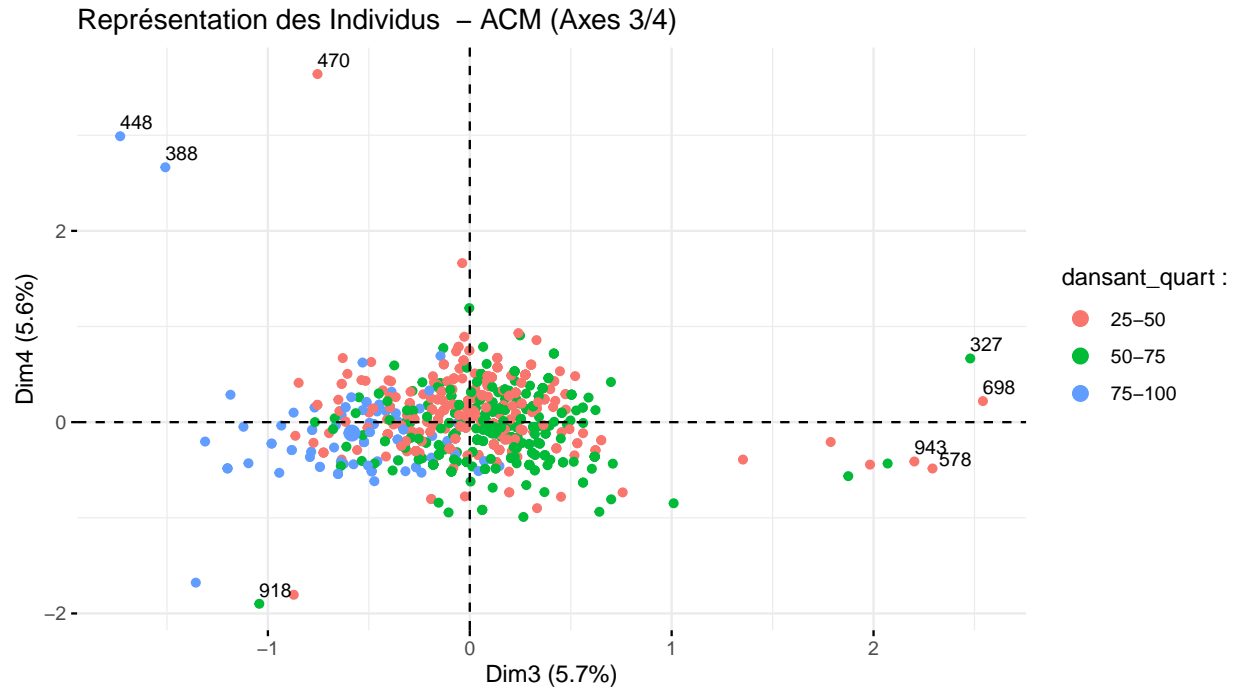
Les individus qui contribuent le plus à la dimension 3 sont le 698, le 327, le 578 et le 943; et les individus qui contribuent le plus à la dimension 4 sont le 470, le 448, le 388 et le 918.

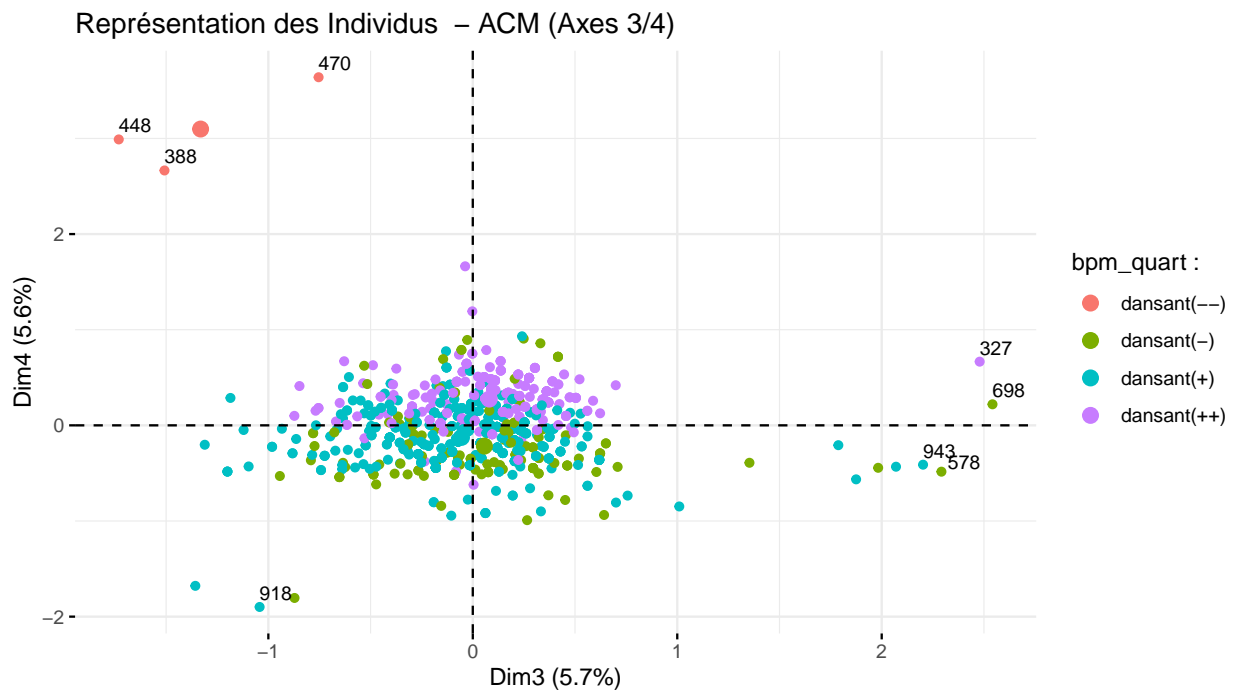
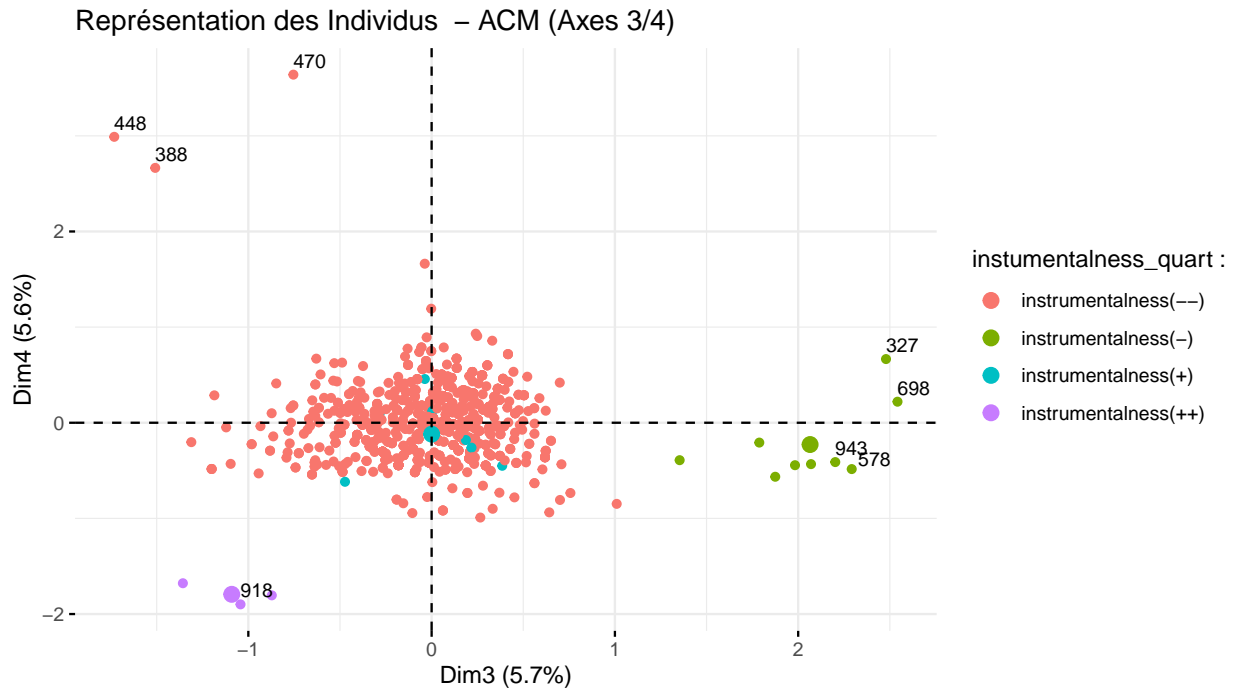
Table 4: Tableau des contributions des individus dans la Dim3

Individu	Contribution en %
698	4.350690
327	4.138687
578	3.537761
943	3.265878
308	2.885278
185	2.646879
409	2.366866
863	2.152434
448	2.019084
388	1.531259
285	1.240092
687	1.233090

Table 5: Tableau des contributions des individus dans la Dim4

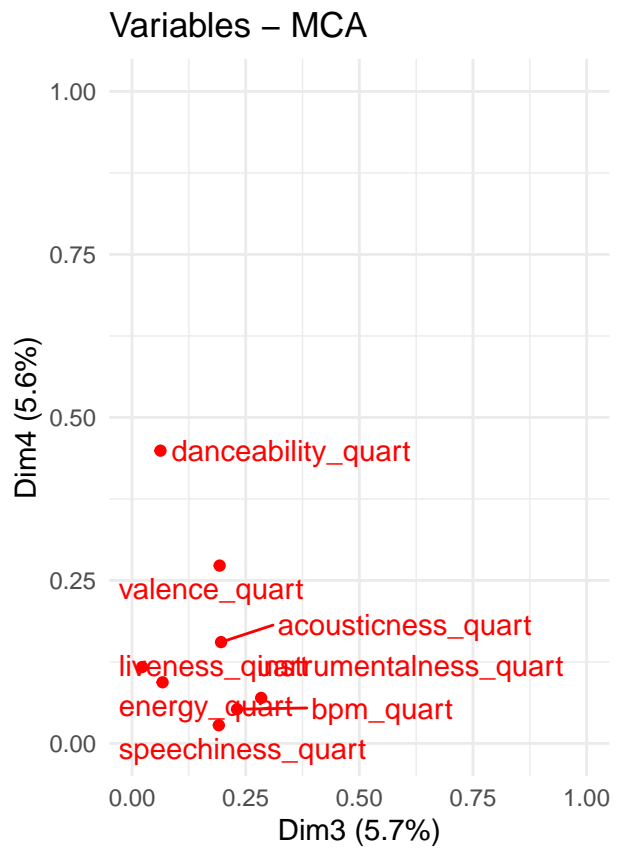
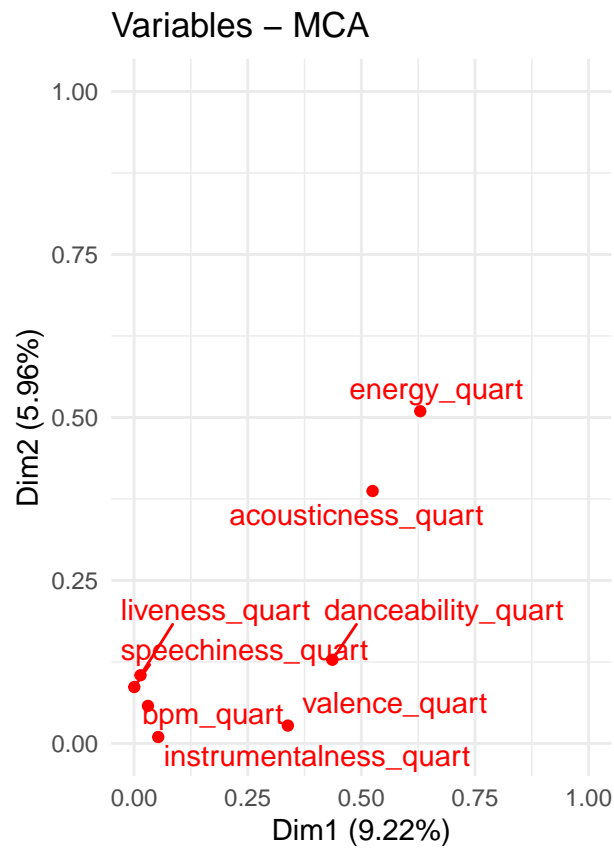
Individu	Contribution en %
470	8.9801841
448	6.0588789
388	4.8146048
918	2.4443025
685	2.2075102
285	1.9084854
248	1.8746562
936	0.9637995
458	0.6660190
695	0.6043319
602	0.5951912
465	0.5877881





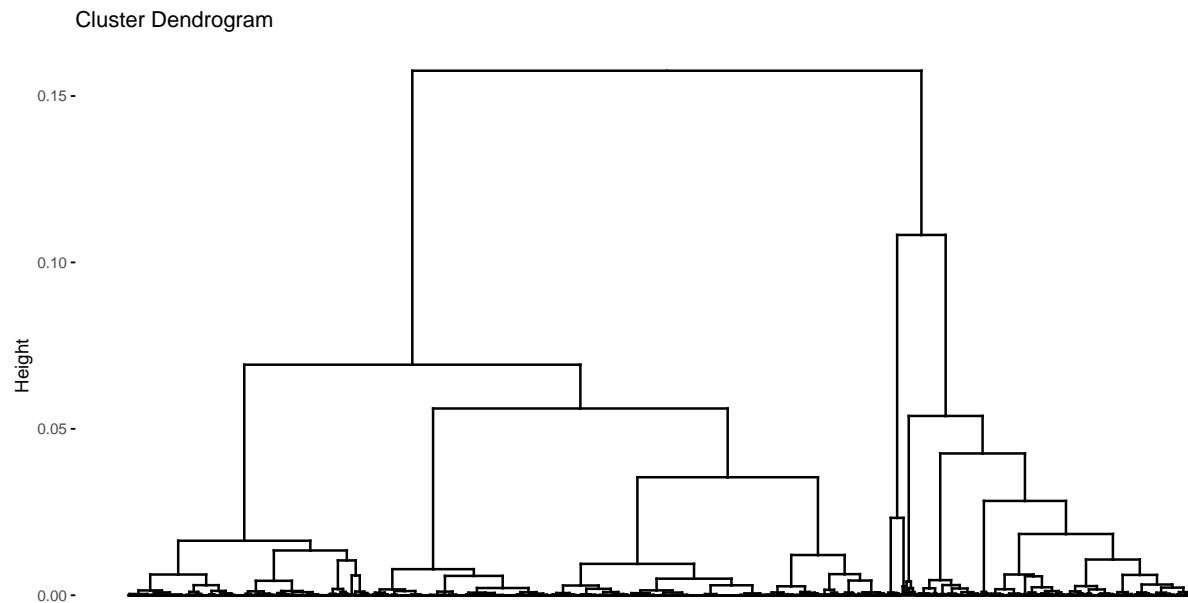
Eta

	dim1	dim2	dim3	dim4
energy_quart	0.63	0.51	0.07	0.09
acousticness_quart	0.52	0.39	0.20	0.16
danceability_quart	0.44	0.13	0.06	0.45
valence_quart	0.34	0.03	0.19	0.27
instrumentalness_quart	0.05	0.01	0.28	0.07
bpm_quart	0.03	0.06	0.23	0.05
liveness_quart	0.01	0.10	0.02	0.12
speechiness_quart	0.00	0.09	0.19	0.03

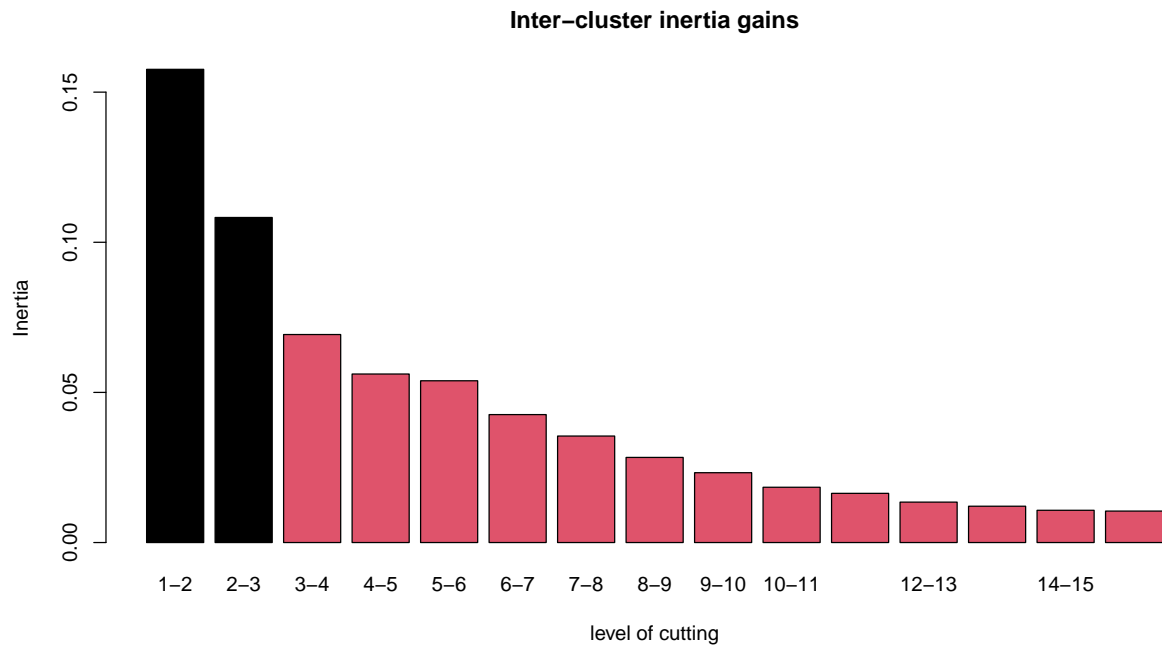


Classification non supervisée

On a ensuite réalisé une classification hiérarchique.

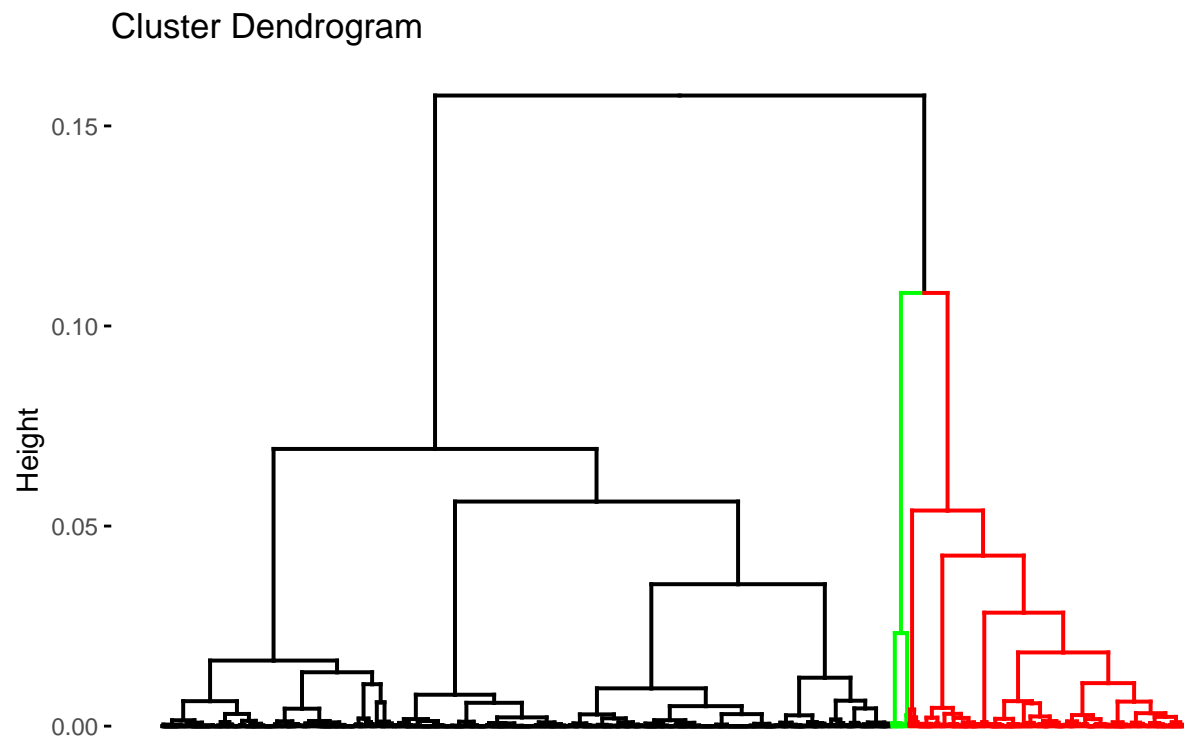


Sur le dendrogramme, on peut voir qu'il serait cohérent de découper en 3 ou 4 groupes.



Le graphique du gain d'inertie inter groupe nous montre qu'on peut choisir de découper 3 ou 4 groupes, au delà, le gain d'inertie est assez faible.

A partir de ces informations, on a décidé de choisir un découpage en 3 groupes.



Parangon

Les parangons : ce sont les individus qui sont le plus au centre de leur cluster

```
## Cluster: 1
##      178      196      479      100      441
## 0.1849608 0.1849608 0.1849608 0.2069922 0.2069922
## -----
## Cluster: 2
##      238      312      139      183      488
## 0.2164301 0.2164301 0.2421538 0.2421538 0.2421538
## -----
## Cluster: 3
##      449      446      465      18      532
## 0.3414631 0.3505304 0.4151873 0.4616657 0.4619336
```

Les parangons et leur distance au centre du cluster.

Distant

```
## Cluster: 1
##      685      918      285      863      390
## 2.563160 2.513485 2.377944 2.047715 1.738405
## -----
## Cluster: 2
##      578      943      185      248      689
## 3.155402 2.898287 2.772377 2.598000 2.517611
## -----
## Cluster: 3
##      470      448      388      698      327
## 5.265890 4.256552 4.107942 3.610850 3.103744
```

Les individus les plus éloignés du centre du cluster.

Description des clusters

Cluster 1

	Cla/Mod	Mod/Cla	Global	p.value	v.test
acousticness_quart=acoustique (-)	91.1	75.6	59.0	6.274e-62	16.61
energy_quart=energy(++)	96.7	39.0	28.6	1.440e-35	12.45
energy_quart=energy(+)	82.3	59.2	51.1	2.553e-15	7.91
danceability_quart=dansant(++)	86.3	39.0	32.1	1.177e-13	7.42
valence_quart=valence_quart_75-100	83.6	21.9	18.6	2.222e-05	4.24
valence_quart=valence_quart_50-75	79.2	36.6	32.8	7.633e-05	3.96
liveness_quart=liveness(+)	95.7	3.2	2.4	4.021e-03	2.88
liveness_quart=liveness(-)	78.4	22.5	20.4	1.071e-02	2.55
key=C#	79.6	16.1	14.4	1.548e-02	2.42
acousticness_quart=acoustique (-)	77.2	23.0	21.2	2.734e-02	2.21
instrumentalness_quart=instrumentalness(-)	71.5	98.7	98.0	3.174e-02	2.15
key=A	60.5	7.2	8.5	3.348e-02	-2.13
danceability_quart=dansant(-)	0.0	0.0	0.3	2.410e-02	-2.26
liveness_quart=liveness(-)	69.1	74.3	76.4	1.609e-02	-2.41
instrumentalness_quart=instrumentalness(-)	11.1	0.1	0.9	3.228e-04	-3.60
liveness_quart=liveness(++)	0.0	0.0	0.8	4.601e-05	-4.08
energy_quart=energy(-)	0.0	0.0	1.7	1.785e-09	-6.02
danceability_quart=dansant(-)	41.4	8.1	14.0	1.096e-14	-7.73
valence_quart=valence_quart_0-25	39.6	9.9	17.7	3.079e-21	-9.46
acousticness_quart=acoustique(++)	2.6	0.3	8.0	9.674e-41	-13.37
acousticness_quart=acoustique(+)	6.2	1.0	11.9	2.256e-55	-15.67
energy_quart=energy(-)	6.8	1.8	18.6	3.519e-93	-20.48

91.1% des musiques prenant acoustique – sont dans ce cluster, ce qui représente 75% de la population de ce cluster.

96.7% des musiques prenant energy ++ et 82% des musiques prenant energy + se retrouvent dans ce cluster, ce qui représente à eux deux quasiment 100% des individus de ce cluster.

Valence + et valence ++ représentent à eux deux quasiment 60% des individus du cluster.

86% des musiques prenant dansant ++ sont dans ce cluster, ce qui représente 39% des individus de ce cluster.

98.7% des ind de ce cluster prennent la modalité instrumentalness –

On peut en conclure que ce cluster comporte globalement des musiques plus énergiques, peu acoustiques et plutôt dansantes.

Cluster 2

	Cla/Mod	Mod/Cla	Global	p.value	v.test
energy_quart=energy(-)	92.7	63.6	18.6	1.383e-98	21.07
acousticness_quart=acoustique(+)	92.0	40.3	11.9	9.547e-56	15.73
acousticness_quart=acoustique(++)	77.6	22.9	8.0	9.220e-22	9.59
valence_quart=valence_quart_0-25	54.4	35.7	17.7	5.292e-17	8.38
danceability_quart=dansant(-)	49.6	25.6	14.0	2.062e-09	5.99
liveness_quart=liveness(++)	100.0	3.1	0.8	2.663e-05	4.20
instrumentalness_quart=instrumentalness(-)	66.7	2.3	0.9	1.678e-02	2.39
liveness_quart=liveness(-)	28.8	81.4	76.4	2.513e-02	2.24
liveness_quart=liveness(-)	20.1	15.1	20.4	1.293e-02	-2.49
liveness_quart=liveness(+)	4.3	0.4	2.4	6.871e-03	-2.70
energy_quart=energy(-)	0.0	0.0	1.7	6.105e-03	-2.74
key=C#	17.5	9.3	14.4	5.204e-03	-2.79
valence_quart=valence_quart_50-75	20.1	24.4	32.8	6.347e-04	-3.42
valence_quart=valence_quart_75-100	15.8	10.9	18.6	1.091e-04	-3.87
danceability_quart=dansant(++)	13.4	15.9	32.1	1.125e-11	-6.79
energy_quart=energy(+)	17.5	32.9	51.1	6.872e-12	-6.86
energy_quart=energy(++)	3.3	3.5	28.6	4.909e-32	-11.78
acousticness_quart=acoustique (-)	8.7	19.0	59.0	6.777e-54	-15.46

Toutes les musiques prenant la modalité liveness ++ sont dans le cluster (il y a peu d'individus qui ont cette modalité, ils ne représentent que 3% du cluster).

63.6% des individus du cluster prennent la modalité energy - ce qui représente 92.7% des individus prenant cette modalité dans la base de données.

92% des individus prenant la modalité acoustique + se retrouvent dans ce cluster et 77.6% des individus prenant la modalité acoustique ++ se retrouvent dans ce cluster, ces modalités concernent en tout 60% des individus du cluster.

50% des sons qui prennent dansant - sont dans ce cluster et cela représente 25.6% des individus du cluster 2.

Ce cluster contient des musiques qui sont globalement assez acoustiques et moins énergiques.

Cluster 3

	Cla/Mod	Mod/Cla	Global	p.value	v.test
energy_quart=energy(-)	100.0	88.9	1.7	7.849e-33	11.93
acousticness_quart=acoustique(++)	19.7	83.3	8.0	5.847e-15	7.81
danceability_quart=dansant(-)	9.0	66.7	14.0	3.258e-07	5.11
danceability_quart=dansant(-)	100.0	16.7	0.3	5.675e-06	4.54
valence_quart=valence_quart_0-25	5.9	55.6	17.7	3.262e-04	3.59
instrumentalness_quart=instrumentalness(-)	22.2	11.1	0.9	1.167e-02	2.52
valence_quart=valence_quart_50-75	0.6	11.1	32.8	4.171e-02	-2.04
acousticness_quart=acoustique (-)	0.0	0.0	21.2	1.315e-02	-2.48
danceability_quart=dansant(++)	0.3	5.6	32.1	9.338e-03	-2.60
energy_quart=energy(++)	0.0	0.0	28.6	2.153e-03	-3.07
danceability_quart=dansant(+)	0.4	11.1	53.6	2.167e-04	-3.70
energy_quart=energy(+)	0.2	5.6	51.1	4.634e-05	-4.07
acousticness_quart=acoustique (-)	0.2	5.6	59.0	2.490e-06	-4.71

88.9% prennent la modalité energy – et tout les individus qui ont la modalité energy – sont dans ce cluster.

Il y a 83% des individus de ce cluster qui prennent la modalité acoustique ++.

Tous les individus qui prennent la modalité dansant – sont dans ce cluster.

55% des individus dans ce cluster prennent la modalité valence –.

Ce cluster possède moins d'individus, il rassemble les musiques qui sont plus calmes, plus acoustiques, et moins dansantes.

Exemple

Pour montrer nos résultats vis-à-vis des clusters , voici le tableau des individus des parangons avec leurs caractéristiques musicales.

	bpm_quart	danceability_quart	valence_quart	energy_quart	acousticness_quart
Cluster 1					
178	25-50	dansant(+)	75-100	energy(+)	acoustique (-)
196	25-50	dansant(+)	75-100	energy(+)	acoustique (-)
479	25-50	dansant(+)	75-100	energy(+)	acoustique (-)
100	25-50	dansant(+)	50-75	energy(+)	acoustique (-)
441	25-50	dansant(+)	50-75	energy(+)	acoustique (-)
Cluster 2					
238	25-50	dansant(+)	0-25	energy(-)	acoustique (-)
312	25-50	dansant(+)	0-25	energy(-)	acoustique (-)
139	25-50	dansant(+)	0-25	energy(-)	acoustique (-)
183	25-50	dansant(+)	0-25	energy(-)	acoustique (-)
488	25-50	dansant(+)	0-25	energy(-)	acoustique (-)
Cluster 3					
449	50-75	dansant(-)	50-75	energy(-)	acoustique(++)
446	25-50	dansant(-)	25-50	energy(-)	acoustique(++)
465	25-50	dansant(+)	0-25	energy(-)	acoustique(++)
18	25-50	dansant(-)	0-25	energy(-)	acoustique(++)
532	25-50	dansant(-)	25-50	energy(-)	acoustique(++)

	instrumentalness_quart	liveness_quart	speechiness_quart
Cluster 1			
178	instrumentalness(-)	liveness(-)	speechiness(-)
196	instrumentalness(-)	liveness(-)	speechiness(-)
479	instrumentalness(-)	liveness(-)	speechiness(-)
100	instrumentalness(-)	liveness(-)	speechiness(-)
441	instrumentalness(-)	liveness(-)	speechiness(-)
Cluster 2			
238	instrumentalness(-)	liveness(-)	speechiness(-)
312	instrumentalness(-)	liveness(-)	speechiness(-)
139	instrumentalness(-)	liveness(-)	speechiness(-)
183	instrumentalness(-)	liveness(-)	speechiness(-)
488	instrumentalness(-)	liveness(-)	speechiness(-)
Cluster 3			
449	instrumentalness(-)	liveness(-)	speechiness(-)
446	instrumentalness(-)	liveness(-)	speechiness(-)
465	instrumentalness(-)	liveness(-)	speechiness(-)
18	instrumentalness(-)	liveness(-)	speechiness(-)
532	instrumentalness(-)	liveness(-)	speechiness(-)

Et, voici le tableau des individus des distants avec leurs caractéristiques musicales.

	bpm_quart	danceability_quart	valence_quart	energy_quart	acousticness_quart
Cluster 1					
685	25-50	dansant(-)	25-50	energy(++)	acoustique (-)
918	50-75	dansant(+)	25-50	energy(+)	acoustique (-)
285	75-100	dansant(+)	0-25	energy(+)	acoustique (-)
863	25-50	dansant(+)	50-75	energy(++)	acoustique (-)
390	50-75	dansant(+)	25-50	energy(++)	acoustique (-)
Cluster 2					
578	25-50	dansant(-)	0-25	energy(-)	acoustique(+)
943	25-50	dansant(+)	0-25	energy(-)	acoustique(+)
185	25-50	dansant(-)	0-25	energy(-)	acoustique (-)
248	25-50	dansant(++)	50-75	energy(-)	acoustique (-)
689	50-75	dansant(+)	50-75	energy(++)	acoustique(+)
Cluster 3					
470	25-50	dansant(-)	0-25	energy(-)	acoustique(++)
448	75-100	dansant(-)	75-100	energy(+)	acoustique(++)
388	75-100	dansant(-)	0-25	energy(-)	acoustique(++)
698	25-50	dansant(-)	0-25	energy(-)	acoustique(+)
327	50-75	dansant(++)	50-75	energy(-)	acoustique(+)

	instrumentalness_quart	liveness_quart	speechiness_quart
Cluster 1			
685	instrumentalness(++)	liveness(-)	speechiness(-)
918	instrumentalness(++)	liveness(-)	speechiness(-)
285	instrumentalness(++)	liveness(-)	speechiness(-)
863	instrumentalness(-)	liveness(-)	speechiness(-)
390	instrumentalness(-)	liveness(+)	speechiness(-)
Cluster 2			
578	instrumentalness(-)	liveness(-)	speechiness(-)
943	instrumentalness(-)	liveness(-)	speechiness(-)
185	instrumentalness(-)	liveness(-)	speechiness(-)
248	instrumentalness(-)	liveness(-)	speechiness(+)
689	instrumentalness(-)	liveness(++)	speechiness(-)
Cluster 3			
470	instrumentalness(-)	liveness(-)	speechiness(-)
448	instrumentalness(-)	liveness(-)	speechiness(-)
388	instrumentalness(-)	liveness(-)	speechiness(-)
698	instrumentalness(-)	liveness(-)	speechiness(-)
327	instrumentalness(-)	liveness(-)	speechiness(-)