

Étude des goûts musicaux et cinématographiques chez les jeunes

Clément Weinreich - Angeline Rondeau
ENSC 1A Groupe 4

June 16, 2021

1 Introduction

Dans le cadre de notre enseignement à l'ENSC nous avons réalisé une analyse d'un jeu de données en utilisant les méthodes de statistiques inférentielles et d'analyse de données comme l'ACP. Le choix du jeu de données était libre. Nous avons donc décidé de nous intéresser aux goûts musicaux et cinématographiques d'adolescents et jeunes adultes (entre 15 et 30 ans).

Les données analysées sont issues d'une étude réalisée en Slovaquie en 2013 sur 1010 participants. Cette étude demandait à chaque participant de classer sur une échelle de 1 à 5 leur affinité pour chaque thématique proposée. Ici nous n'avons donc gardé que les genres musicaux et cinématographiques proposés. C'est cette valeur de 1 à 5 qui constitue nos données.

L'étude réalisée a questionné chaque participant sur un grand nombre de genres musicaux (19) et de genres cinématographiques (12). Ce grand nombre de variables rendait l'analyse de données complexe (5 valeurs propres > 1 pour l'ACP). Nous avons donc choisi de nous limiter à trois genres musicaux (Pop, Musique Classique et Metal/Hard-Rock) et trois genres cinématographiques (Romantique, Documentaire et Thriller). Nous avons choisi ces genres en fonction des stéréotypes que nous avons à l'esprit.

Nous avons décidé d'axer notre étude sur la problématique suivante: Peut-on déduire de ces données des groupes de genres musicaux et/ou cinématographiques types en fonction de l'appréciation des individus ? Si de tels profils existent correspondent-ils aux stéréotypes connus dans notre imaginaire commun (par exemple quelqu'un qui aime le Hard Rock et le Métal n'aime pas les films Romantiques) ?

Le code a été réalisé dans un notebook R collaboratif sur la plateforme Kaggle.

2 Description du jeu de données

Dans un premier temps, récupérons les données à analyser.

```
[1]: data <- read.csv("../input/young-people-survey/responses.csv") # Lecture des données
data <- data[complete.cases(data),] # On supprime toutes les valeurs NA
attach(data) # Pour pouvoir utiliser les noms de colonnes en tant que variable
```

```
[2]: dataMusic <- cbind(Metal.or.Hardrock, Pop, Classical.music) # On récupère les données
      ↳liées aux goûts musicaux
dataFilm <- cbind(Romantic,Thriller,Documentary) # On récupère les données liées aux
      ↳goûts cinématographiques
dataMusicFilm <- cbind(dataFilm,dataMusic) # On joint les données liées aux films et à
      ↳la musique dans un seul dataframe
```

```
[3]: head(dataMusicFilm,n=4) # On affiche une partie des données
```

	Romantic	Thriller	Documentary	Metal.or.Hardrock	Pop	Classical.music
4	2	3	1	5	2	
3	2	4	4	3	1	
2	4	2	3	3	4	
2	4	3	1	5	4	

On a donc pour chacun des genres, l'appréciation de l'individu avec une note allant de 1 à 5. Étudions maintenant la répartition de ces appréciations pour certaines variables.

```
[4]: summary(Romantic) # résumé de l'échantillon contenant l'appréciation des films
      ↳romantiques
sd(Romantic) # écart type de cet échantillon
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   3.000   4.000   3.472  4.750   5.000
1.203991998017
```

On observe ici, qu'en moyenne, les films romantiques sont appréciés à 3.475/5, avec une médiane de 4. Il y a donc autant d'individu qui ont noté leur appréciation des films romantiques, en dessous et au dessus de 4.

```
[5]: summary(Metal.or.Hardrock) # résumé de l'échantillon contenant l'appréciation du genre
      ↳musical Métal ou Hardrock
sd(Metal.or.Hardrock) # écart type de cet échantillon
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   1.000   2.000   2.356  3.000   5.000
1.39509716480719
```

Pour le genre musical "Métal ou Hardrock", la moyenne d'appréciation est bien plus basse car elle est à 2.356. La médiane est également en dessous avec une valeur de 2.356. On observe également que l'écart type est plus grande, l'appréciation de ce genre musical est donc plus dispersé que l'appréciation des films romantiques.

```
[6]: summary(Pop) # résumé de l'échantillon contenant l'appréciation du genre musical Pop
sd(Pop) # écart type de cet échantillon
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.00   3.00   4.00   3.44   4.00   5.00
1.16959632607128
```

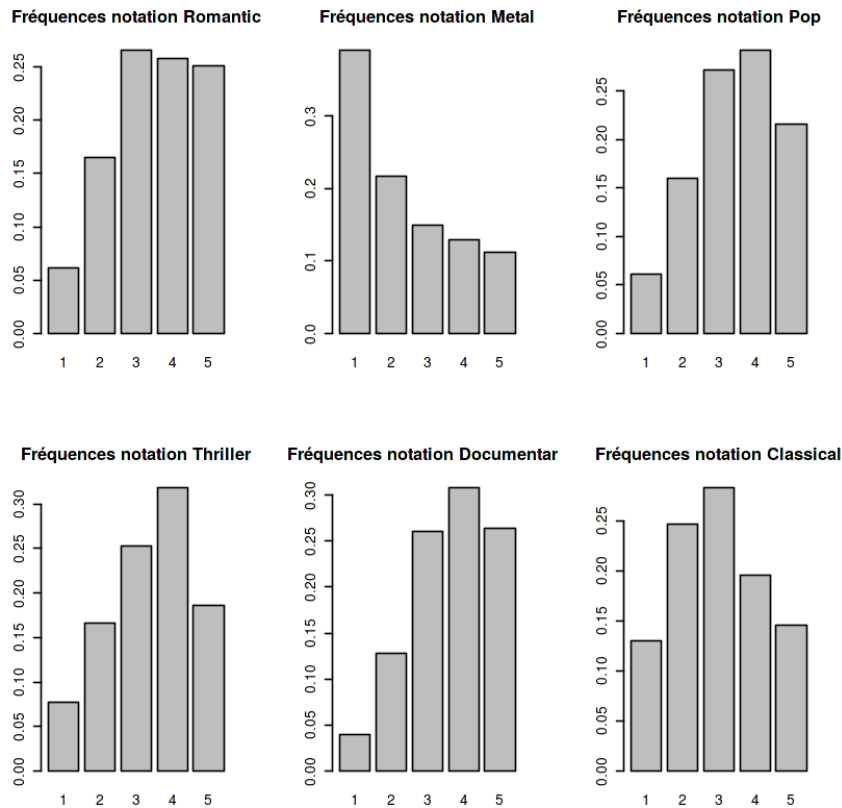
On observe que l'échantillon sur l'appréciation de la Pop et celui des films romantiques sont assez identiques avec la même médiane et des valeurs quasiment identiques pour la moyenne et l'écart type.

```
[7]: summary(Documentary) # résumé de l'échantillon contenant l'appréciation des
      ↳documentaires
sd(Documentary) # écart type de cet échantillon
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000   3.000   4.000   3.628  5.000   5.000
1.12061725057259
```

On observe pour l'échantillon correspondant à l'appréciation des documentaires que la médiane est identique à celle de la musique pop et des films romantiques. On peut également noter que la moyenne d'appréciation est plus haute que pour les autres variables évaluées précédemment avec 3.628/5. De plus, le premier quartile est assez haut (3), ce qui laisse penser que peu de personnes n'aiment pas les documentaires.

```
[8]: par(mfrow=c(2,3))
# Affichage de graphiques en bâtons pour représenter la fréquence des modalités
  →(1,2,3,4,5) pour chaque variable
barplot(table(Romantic)/sum(table(Romantic)));title("Fréquences notation Romantic")
barplot(table(Metal.or.Hardrock)/sum(table(Metal.or.Hardrock)));title("Fréquences
  →notation Metal")
barplot(table(Pop)/sum(table(Pop)));title("Fréquences notation Pop")
barplot(table(Thriller)/sum(table(Thriller)));title("Fréquences notation Thriller")
barplot(table(Documentary)/sum(table(Documentary)));title("Fréquences notation
  →Documentary")
barplot(table(Classical.music)/sum(table(Classical.music)));title("Fréquences notation
  →Classical")
```



Ces graphiques en bâtons permettent de montrer la répartition en fréquence de chacune des notes d'appréciations données aux 6 genres. On observe que certaines catégories sont très appréciées comme les films romantiques, la musique Pop et les documentaires. On peut noter que les films thrillers sont bien appréciés avec plus de 30% de 4/5. En revanche, l'appréciation du métal est très mauvaise parmi les sujets interrogés dans cette étude, on observe plus de 30% de note à 1/5.

Ces graphiques sont utiles pour visualiser la répartition des notes mais ne prennent pas en compte le fait

que les données ne sont pas indépendantes. En effet chaque individu note ces catégories, les données sont donc appariées. Nous allons maintenant prendre en compte cette information pour la suite de l'analyse.

Si on s'intéresse plus précisément à deux catégories, les films romantiques et le Métal :

```
[9]: tabcontRomantiqueMetal <- table(Romantic,Metal.or.Hardrock) # Création de la table de
      ↳contingence entre l'appréciation des films romantiques, et du genre musical Métal
tabcontRomantiqueMetal
round(tabcontRomantiqueMetal/sum(tabcontRomantiqueMetal)*100,digits=2) # Table de
      ↳contingence en %
```

	Metal.or.Hardrock						Metal.or.Hardrock				
Romantic	1	2	3	4	5	Romantic	1	2	3	4	5
1	14	4	6	6	12	1	2.04	0.58	0.87	0.87	1.75
2	37	18	16	23	19	2	5.39	2.62	2.33	3.35	2.77
3	63	38	36	23	22	3	9.18	5.54	5.25	3.35	3.21
4	70	45	26	23	13	4	10.20	6.56	3.79	3.35	1.90
5	84	44	19	14	11	5	12.24	6.41	2.77	2.04	1.60

On peut faire la table de contingence entre l'appréciation des films romantiques, et du Métal. Cette table de contingence nous permet de remarquer que seulement 11 personnes ont noté leur appréciation des films romantique ET du métal à 5/5 (1.6% de l'échantillon), alors que 84 personnes (12.24% de l'échantillon) ont noté leur appréciation des films romantiques à 5/5 mais leur appréciation du métal à 1/5. Cela laisse donc penser que les gens qui aiment les films romantiques n'aiment pas le métal.

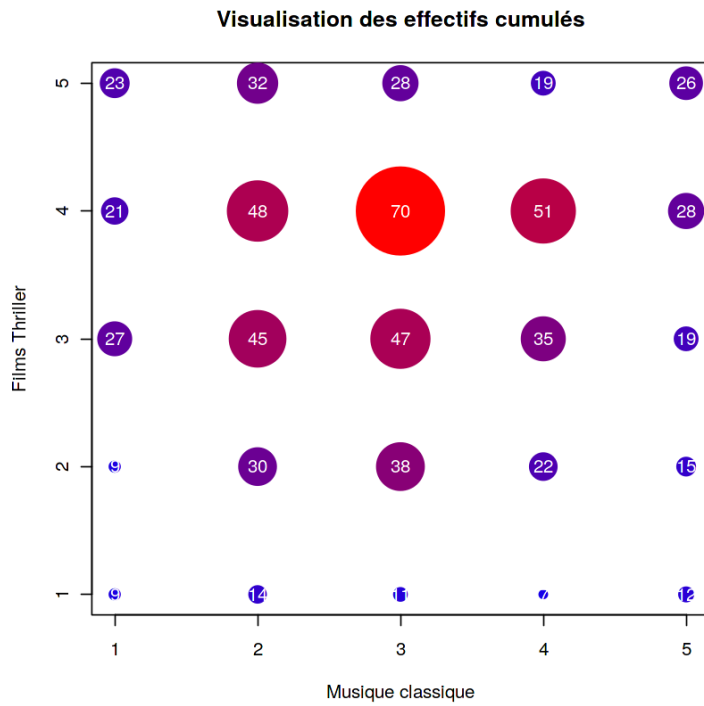
Les données étant appariées, il aurait été pertinent de faire un nuage de point. Or ici, il n'y a que 25 valeurs possibles pour 686 points. Cela n'a donc pas vraiment de sens. C'est pourquoi nous avons essayé de créer une autre forme de visualisation en nous basant sur la table de contingence.

```
[10]: # Fonction que nous avons créé pour visualiser plus facilement la table de contingence
visualiser_tabcont <- function(tabcont,labelX,labelY){
  colgradient <- colorRampPalette(c("blue","red"))(max(tabcont)) # Création d'un
  ↳gradient de couleur pour une meilleure intelligibilité du graphique

  # Création du graphique
  plot(1, xlab=labelX, ylab=labelY, xlim=c(1, 5), ylim=c(1, 5))
  for(i in 1:nrow(tabcont)){
    for(j in 1:ncol(tabcont)){
      nbPoints <- tabcont[i,j] # on récupère la valeur correspondante dans la
      ↳table de contingence
      points(i,j,cex=(nbPoints/
      ↳max(tabcont))*11,pch=16,col=colgradient[nbPoints]) # On créé un point de taille
      ↳correspondant à ce nombre, de même pour la couleur
      text(i,j,nbPoints,col="white",cex=1) # on ajoute le nombre sur le point
    }
  }
  title("Visualisation des effectifs cumulés")
}
```

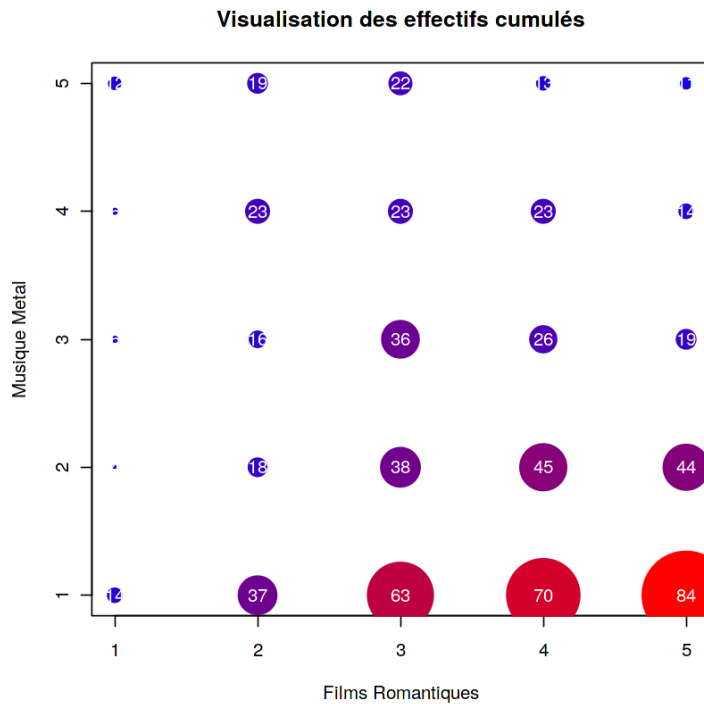
Cette fonction permet de créer des points de taille plus ou moins grande en fonction des effectifs conjoints à une certaine position. Un gradient de couleur est également utilisé pour aider à la visualisation.

```
[11]: visualiser_tabcont(table(Classical.music,Thriller),"Musique classique", "Films
      ↳Thriller") # Visualisation de la table de contingence entre musique classique et les
      ↳films Thriller
```



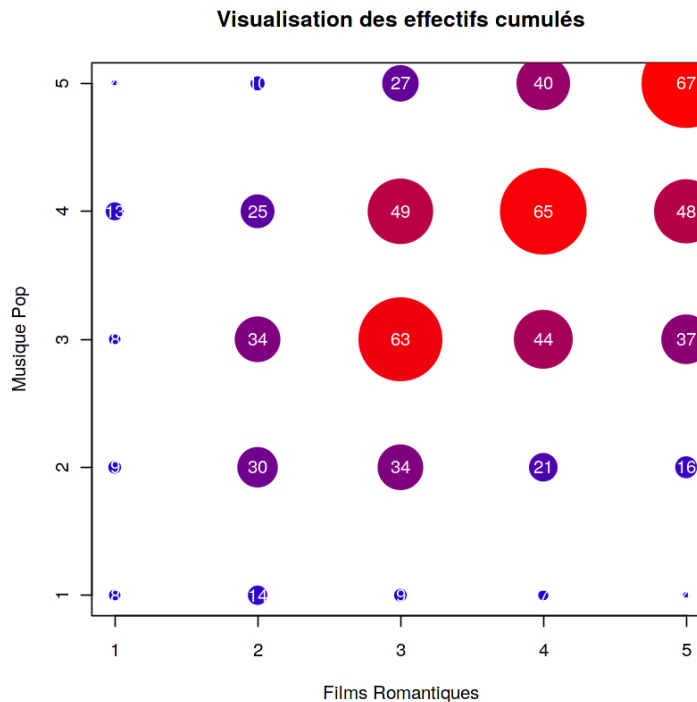
Ici par exemple, on observe qu'il n'y a pas de forte tendance dans l'appréciation des Thrillers et de la musique classique.

```
[12]: visualiser_tabcont(tabcontRomantiqueMetal, "Films Romantiques", "Musique Metal") #_
      → Visualisation de la table de contingence entre films romantiques et la musique Métal
```



Cependant, entre les films romantiques et le Metal, on remarque très facilement qu'il y a une forte tendance. Les individus qui aiment les films romantiques n'aiment pas le Metal.

```
[13]: visualiser_tabcont(table(Romantic,Pop), "Films Romantiques", "Musique Pop") #  
      → Visualisation de la table de contingence entre films romantiques et la musique pop
```



Concernant les films romantiques et la musique Pop, on peut observer une corrélation positive entre les deux. En effet, ce graphique laisse paraître qu'en général, les gens qui aiment la Pop aiment également les films romantiques. Cela reste encore à confirmer par la suite, ce n'est qu'une hypothèse.

3 Analyse et exploitation du jeu de données

3.1 Analyse en composantes principales

Nous allons maintenant réaliser une ACP sur notre jeu de données.

```
[ ]: require(PCAmixdata) # permet de charger le package "PCAmixdata"
```

```
[15]: # Mise en oeuvre de l'ACP et Choix du nombre d'axes à retenir  
ACPMusicFilm <- PCAmix(dataMusicFilm,graph=FALSE) # Stocke les calculs de l'ACP dans  
      → l'objet " ACPMusicFilm  
round(ACPMusicFilm$eig,digit=2) # Valeurs propres et pourcentage de variance expliqué  
      → pour chaque axe
```

	Eigenvalue	Proportion	Cumulative
dim 1	1.73	28.78	28.78
dim 2	1.18	19.72	48.50
dim 3	1.02	17.02	65.52
dim 4	0.84	13.98	79.50
dim 5	0.63	10.55	90.05
dim 6	0.60	9.95	100.00

Cette ACP est centrée réduite et donne le même poids à tous les individus de l'étude. L'espace des individus est R^6 car, pour chaque individu, on mesure quantitativement leur affinité pour 6 genres musicaux ou cinématographiques. La métrique associée est $M=D_{[1/s^2]}$ pour une ACP centrée.

L'espace des variables est R^{686} car chaque variable a été mesurée sur $n=686$ individus de manière complète (certains individus n'ont pas répondu à toutes les questions, nous les avons donc supprimés de l'analyse). La métrique associée est $D=(1/686)*I_{686}$ car tous les individus ont le même poids.

Pour choisir les axes retenus, on utilise le critère de Kaiser : on conserve seulement les axes factoriels associés à une valeur propre (eigenvalue) plus grande que 1. Ici on retient donc les 3 premiers axes: 1er axe factoriel explique 28,78% de l'inertie (ou de la variance) Second axe explique 19,72% d'inertie supplémentaire Troisième axe explique 17,02% d'inertie supplémentaire Ainsi, en considérant l'espace 1-2-3, on récupère 65,52% de l'information. Sur le plan 1-2 on a 48.5% de l'information et sur le plan 1-3 on a 45.8% de l'information.

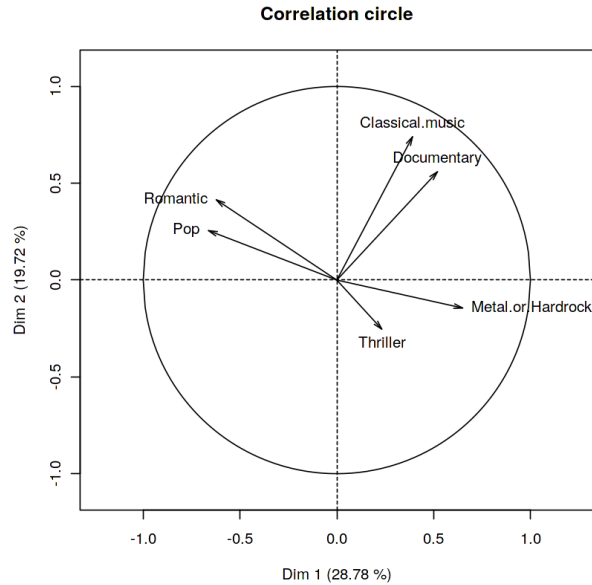
```
[16]: # Sorties numériques pour les variables
round(ACPMusicFilm$quant1$cos2,digit=3) # Cosinus carrés associés aux variables
```

	dim 1	dim 2	dim 3	dim 4	dim 5
Romantic	0.388	0.172	0.001	0.183	0.195
Thriller	0.054	0.064	0.806	0.001	0.012
Documentary	0.270	0.313	0.006	0.203	0.178
Metal.or.Hardrock	0.422	0.021	0.000	0.398	0.026
Pop	0.440	0.065	0.197	0.000	0.023
Classical.music	0.153	0.548	0.012	0.053	0.199

La qualité de représentation des variables sur les axes est donnée par les cosinus carrés. On a donc : * Sur l'axe 1, l'appréciation de Pop et Metal/Hardrock sont bien représentés. L'appréciation des autres genres sont moyennement représentés par l'axe sauf le genre Thriller qui est très mal représenté (seulement 5%) * L'axe 2 représente très bien l'appréciation de la musique Classique. Il complète bien l'axe 1 avec une bonne ou moyenne représentation des genres qui sont moyennement représentés sur l'axe 1 (Musique classique, Romantique et Documentaire) sauf Thriller qui est toujours très mal représenté. * L'axe 3 représente extrêmement bien l'appréciation des films Thriller (80%) mais ne représente pas quasiment pas les autres genres (autour de 0% pour tous les autres sauf Pop qui est à 20%)

Finalement, en étudiant ces trois axes nous pouvons raisonnablement discuter de l'appréciation de tous les genres car ils sont tous représentés à au moins 50%. Nous ferons tout de même attention à ne pas tirer de conclusion sur l'appréciation des films Thriller en observant le plan 1-2 où il est très mal représenté. Nous nuancerons aussi nos conclusions quand à l'appréciation du Metal/Hardrock qui est le genre le moins bien représenté sur ces 3 plans (44% sur le plan 1-2)

```
[17]: plot(ACPMusicFilm,axes=c(1,2),choice="cor") # Affichage du cercle des corrélations des
→variables (plan 1-2)
```

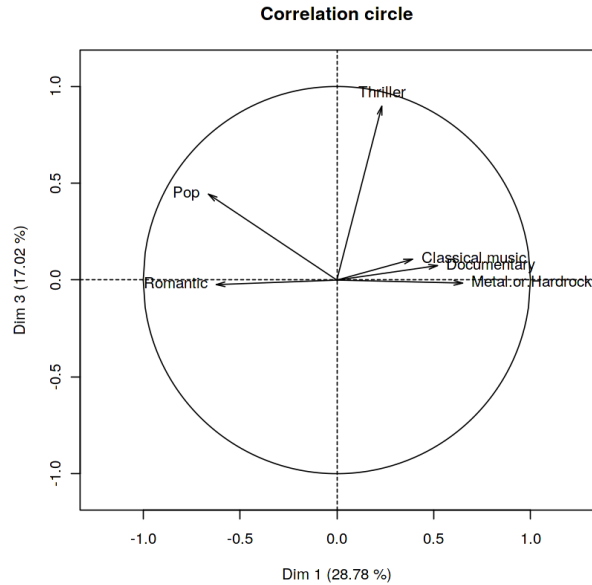


Comme nous l'avons discuté précédemment, les variables sont plus ou moins bien représentées dans ce plan: la variable musique Classique est très bien représentée (vecteur de projection proche de la circonférence du cercle) tandis que Thriller ne l'est pas du tout. Nous ne discuterons pas de la variable Thriller avec ce graphe.

On voit que les variables Romantique, Pop et Metal/ Hardrock semblent corrélées: Romantique et Pop sont positivement corrélées (vecteurs de projection dans le même sens). Ce qui signifie que plus la note d'appréciation de la Pop est élevée, plus la note d'appréciation du genre Romantique le sera et vice-versa. A contrario, la projection de Metal/Hardrock(MH) est opposée au deux précédentes: la variable MH est négativement corrélée à la Pop (et au genre Romantique). Cela signifie que plus l'individu appréciera la Pop (et le genre Romantique) moins il appréciera le MH et vice-versa.

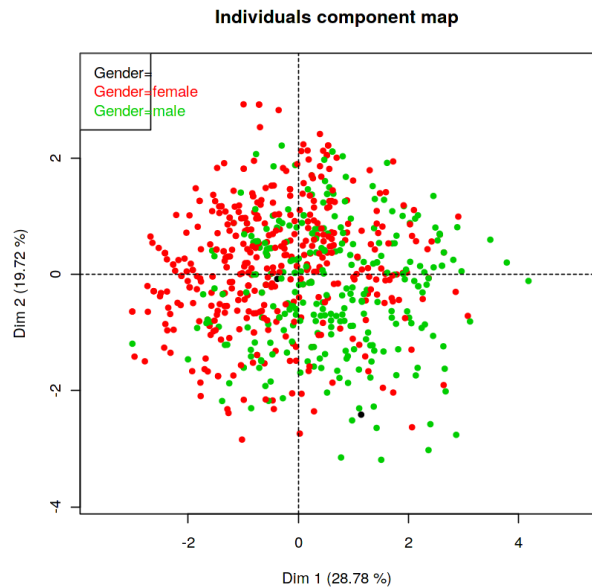
La musique Classique et les Documentaire sont positivement corrélées (pour les même raisons que la Pop et le genre Romantique). Ce groupe de 2 variables est quasiment orthogonal aux trois autres variables discutées plus tôt. Cela signifie, qu'à priori, le fait d'aimer la musique classique et les documentaires n'a pas d'impact sur le fait d'aimer le MH, la Pop ou les films Romantiques.

```
[18]: plot(ACPMusicFilm,axes=c(1,3),choice="cor") # Affichage du cercle des corrélations des
      →variables (plan 1-3)
```

Ici les variables Thriller et Pop sont bien représentées. Les documentaires et la musique classique sont très mal représentés. On retrouve l'opposition entre MH et Romantique. Ce graphe nous indique surtout l'orthogonalité (et donc la non corrélation à priori) entre le Thriller et la Pop, Le Romantique et le MH. Ceci contredit donc le stéréotype que nous avons imaginé : une personne écoutant du Métal aime les films Thriller. On ne peut rien dire de la corrélation ou non corrélation entre le Thriller et le Documentaire ou la Musique Classique.

```
[19]: plot(ACPMusicFilm, axes=c(1,2), choice="ind", coloring.ind=Gender, label=TRUE) # Affichage
      → du graphique des individus (plan 1-2)
```



Ce nuage de points montrant la répartition des individus nous apprend deux choses : * Le plan est assez équitablement occupé par les individus ce qui signifie qu'il n'existe pas de genre musical ou cinématographique délaissé par la cohorte. On peut remarquer tout de même une légère sur-densité au centre du graphique : la plupart des personnes apprécient moyennement tous les genres musicaux et ciné-

matographiques considérés. * En regardant la répartition selon les sexes des individus, on remarque que les fans de pop et de films romantiques ont tendance à être des femmes tandis que les fans de MH ont tendance à être des hommes.

Ces dernières considérations vont dans le sens des stéréotypes.

3.2 Test d'hypothèse

Nous allons maintenant réaliser quelques tests d'hypothèses quant à l'indépendance de certaines variables.

Nous allons commencer par étudier l'indépendance entre l'appréciation des films romantiques et du métal. Pour cela nous allons utiliser le test du chi-deux d'indépendance. Pour ce test on a H_0 : indépendance entre les deux variables contre H_1 : non H_0 .

```
[20]: chisq.test(tabcontRomantiqueMetal)$expected # Calcul des effectifs théoriques
```

	1	2	3	4	5
1	16.40816	9.122449	6.306122	5.44898	4.714286
2	44.14577	24.543732	16.966472	14.66035	12.683673
3	71.10204	39.530612	27.326531	23.61224	20.428571
4	69.14869	38.444606	26.575802	22.96356	19.867347
5	67.19534	37.358601	25.825073	22.31487	19.306122

Dans un premier temps, on récupère les effectifs théoriques calculés sous H_0 . Cela permet de repérer les cases où l'effectif théorique est inférieur à 5. En effet pour avoir une bonne approximation de la loi de la statistique de test par une loi du chi-deux, il faut avoir un effectif supérieur ou égal à 5 dans chacune des cases. Or ici on voit que dans la case (1,5) on a 4.71. Par conséquent, on peut travailler avec des simulations de la p-value pour palier ce manque.

```
[21]: chisq.test(tabcontRomantiqueMetal, simulate.p.value = TRUE) # Test du chi-deux
      → d'indépendance
      # H0 : indépendance entre l'appréciation du métal et "l'appréciation des films"
      → romantique versus H1 : non H0.
```

```
      Pearson's Chi-squared test with simulated p-value (based on 2000
      replicates)
data: tabcontRomantiqueMetal
X-squared = 46.652, df = NA, p-value = 0.0004998
```

On observe que la p-value est très proche de 0, et donc bien inférieure au risque de première espèce $\alpha = 5\%$, on rejette donc très fortement H_0 . Il existe donc un lien entre l'appréciation du métal et des films romantiques, comme nous le supposions précédemment. Pour confirmer cela, on peut calculer le coefficient de corrélation linéaire entre les deux variables :

```
[22]: cor(Romantic, Metal.or.Hardrock, method="pearson")
```

```
-0.212276387547616
```

Le coefficient de corrélation linéaire selon la méthode de pearson est $-0.21 < 0$, on peut donc dire qu'il y a une faible liaison linéaire négative entre les deux variables. Plus un individu aime les films romantiques, moins il aime le métal.

Passons maintenant au test d'indépendance entre la musique classique et les Thrillers.

```
[23]: chisq.test(table(Classical.music, Thriller))$expected # Calcul des effectifs théoriques
```

	1	2	3	4	5
1	6.876093	14.79009	22.44461	28.28280	16.60641
2	13.056851	28.08455	42.61953	53.70554	31.53353
3	14.988338	32.23907	48.92420	61.65015	36.19825
4	10.352770	22.26822	33.79300	42.58309	25.00292
5	7.725948	16.61808	25.21866	31.77843	18.65889

Ici, toutes les valeurs sont > 5, on peut donc continuer sans problème.

```
[24]: chisq.test(table(Classical.music,Thriller)) # Test du chi-deux d'indépendance
# H0 : indépendance entre l'appréciation de la musique classique et l'appréciation des
  →films Thriller versus H1 : non H0.
```

```
Pearson's Chi-squared test
data: table(Classical.music, Thriller)
X-squared = 25.913, df = 16, p-value = 0.05526
```

On obtient une p-value égale à 5.526% > 5 %, on ne peut donc pas rejeter H0, il y a indépendance entre l'appréciation de la musique classique et l'appréciation des Thrillers.

Pour afficher le coefficient de corrélation linéaire entre toutes les variables, on peut faire :

```
[25]: cor(dataMusic,dataFilm,method="pearson") # Correlation entre les goûts musicaux et
  →cinématographiques
```

	Romantic	Thriller	Documentary
Metal.or.Hardrock	-0.212276388	0.117138612	0.1024361
Pop	0.315532233	0.031788633	-0.1456218
Classical.music	0.008914091	0.005760452	0.3009092

Cela confirme bien le fait qu'il n'y a pas de liaison linéaire entre l'appréciation des Thrillers et la musique classique. on remarque également des liaisons linéaires positives entre l'appréciation de la pop et des films romantiques, ainsi qu'entre l'appréciation de la musique classique et des documentaires (environ 0.3 pour les deux).

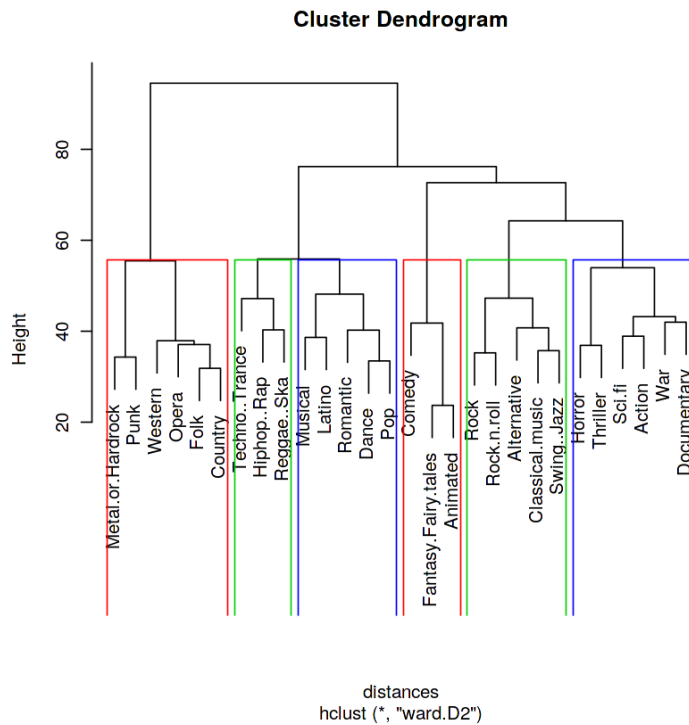
3.3 Classification hiérarchique

Maintenant que nous avons observé que l'appréciation de certaines catégories étaient corrélées avec l'appréciations d'autres catégories, il est pertinent d'essayer de dégager des groupes de catégories en fonction de l'appréciation des individus. Pour cela, nous n'allons pas nous cantonner à 3 genres cinématographiques et musicaux mais à tous les genres cinématographiques (12) et tous les genres musicaux (19).

```
[26]: tousFilmsMusiques <- cbind(data[3:19], data[21:31]) # Récupération de tous les genres
  →musicaux et cinématographiques d'après leur indice
```

Réalisation de la classification hiérarchique grâce à la méthode d'aggregation de Ward :

```
[27]: distances<-dist(t(tousFilmsMusiques)) # calcul des distances entre l'appréciation des
  →genres musicaux et cinématographiques
arbre <- hclust(distances,method="ward.D2") # construction de la classification
  →hiérarchique avec la methode d'aggregation de Ward
plot(arbre) # representation graphique de la hierarchie : dendogramme
rect.hclust(arbre,k=6,border=2:4) # ajout de k=6 rectangles de couleurs pour faire
  →ressortir visuellement 6 groupes de genres de films et musiques
# en fonction de l'appréciation des jeunes
```



On obtient finalement un dendrogramme avec 6 groupes de genres musicaux et cinématographiques (les deux catégories sont ici confondues). On observe que certains groupes ne sont pas surprenants et correspondent aux stéréotypes communs comme le groupe contenant les comédies musicales, la musique latine, les films romantiques, la musique Dance et la musique Pop. Cela correspond bien à nos analyses précédentes, les individus qui apprécient la musique pop apprécient également les films romantiques. Cependant d'autres corrélations que nous avons trouvées n'ont pas été représentées comme celle entre les documentaires et la musique classique.

4 Conclusion

Pour rappel, notre problématique était de nous demander si on pouvait déduire de ces données des groupes de genres musicaux et/ou cinématographiques types en fonction de l'appréciation des individus. De plus, si tel était le cas, nous nous demandions si ces groupes et ces corrélations correspondaient aux stéréotypes connus.

Cette étude nous a permis de dégager des corrélations entre l'appréciation de certains genres musicaux et cinématographiques ce qui nous permet de répondre positivement à la première partie de notre problématique : il semble exister des profils types d'appréciation de musiques et de films. Nous pouvons nous aider du dendrogramme pour visualiser de tels profils même s'il serait pertinent de pousser cette analyse car celui-ci montre pas certaines corrélation (négatives ou positives) trouvées par nos analyse précédentes.

Concernant la deuxième partie de notre problématique, les résultats de notre étude vont dans le sens de certains stéréotypes : - les personnes aimant la Pop aiment les films romantiques et ce groupe de personne n'apprécie pas le Métal/Hard Rock. Dans ce groupe les femmes sont largement présentes. - les personnes aimant les documentaires aiment aussi la musique classique. Cependant certains stéréotypes ont été infirmés par nos résultats : les personnes aimant le Métal n'aime pas forcément particulièrement les Thriller. Nous avons beaucoup aimé travailler sur ce sujet et il serait très intéressant de continuer à essayer de dégager des corrélations entre l'appréciation des différents genre artistiques. Malheureusement le nombre limité de page ne nous a pas permis de nous étendre sur le sujet comme nous l'aurions souhaité.