

Projet IF36

Les Tontons Flingueurs

2024-06-24



Introduction

Le dataset que nous avons choisi provient de Kaggle (<https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results/data>) et vise à identifier les corrélations qui peuvent exister entre la musique et la santé mentale autodéclarée d'un individu, que ce soit à travers l'écoute de styles particuliers, par la pratique d'un instrument ou encore par la composition.

Nous avons choisi ce dataset en raison de sa pertinence pour les analyses que nous voulons mener, tout en offrant la possibilité d'explorer les liens potentiels entre la santé mentale et la musique, étant donné notre intérêt marqué pour celle-ci.

Données

Source

La collecte des données a été gérée via un formulaire Google. Les répondants n'étaient pas limités par l'âge ou le lieu.

Le formulaire a été publié sur divers forums Reddit, serveurs Discord et plateformes de médias sociaux. Des affiches ont également été utilisées pour annoncer le formulaire dans les bibliothèques, les parcs et autres lieux publics.

Le formulaire était relativement bref afin que les répondants soient plus susceptibles de terminer le sondage. Les questions « plus difficiles » (telles que le BPM) sont restées facultatives pour la même raison.

Format

Les données sont fournies dans un seul fichier au format CSV.

Description

Ainsi, nous avons un total de **736** observations pour **33** features.

5 d'entre-elles sont de type booléen :

- **While working** (écoute de la musique en travaillant / étudiant)
- **Instrumentalist** (régulièrement)
- **Composer**
- **Exploratory** (explore régulièrement de nouveaux artistes / styles de musique)
- **Foreign languages** (écoute régulièrement de la musique en langue étrangère)

6 sont de type entier :

- **Age**
- **BPM** (nombre de battements par minute du style de musique favori)
- **Anxiety** (échelle de 0 à 10)
- **Depression** (de même)
- **Insomnia** (de même)
- **OCD** (de même)

1 de type flottant :

- **Hours per day** (de 0 à 24)

1 de type date :

- **Timestamp** (date de soumission de la réponse au formulaire)

Et les 20 restantes sont de type string et sont ordonnables (et le nombre de réponses possibles était limité) :

- **Primary streaming** (plateforme d'écoute principale)
- **Fav genre**
- **Music effect** (choix entre improve, no effect et worsen)
- **Frequency (Classical)** (choix entre never, rarely, sometimes et very frequently)
- **Frequency (Country)**
- **Frequency (EDM)**
- **Frequency (Folk)**
- **Frequency (Gospel)**

- Frequency (Hip hop)
- Frequency (Jazz)
- Frequency (K pop)
- Frequency (Latin)
- Frequency (Lofi)
- Frequency (Metal)
- Frequency (Pop)
- Frequency (R&B)
- Frequency (Rap)
- Frequency (Rock)
- Frequency (Video game)
- Permission (autorisation de rendre publique la réponse, ainsi il n'y a qu'une seule valeur : "I understand.")

On retrouve réellement deux sous-groupes de features : les fréquences d'écoutes et les troubles mentaux.

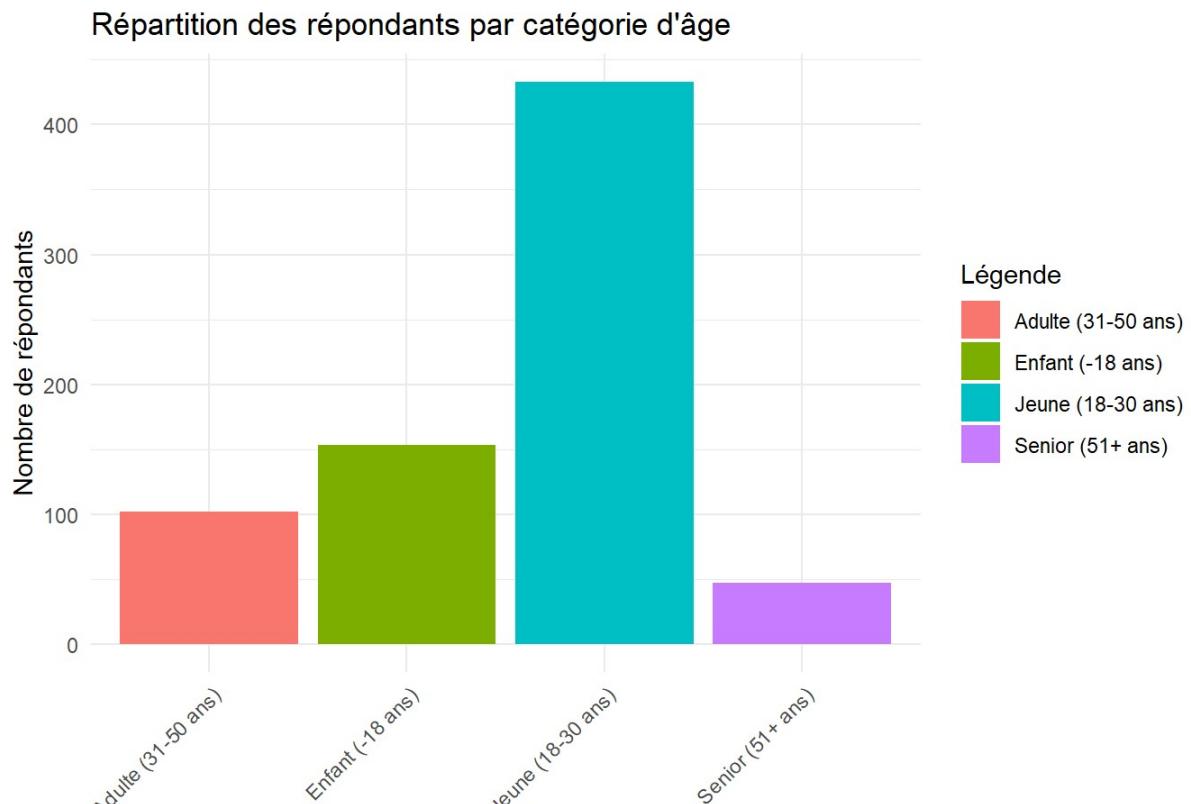
Analyse

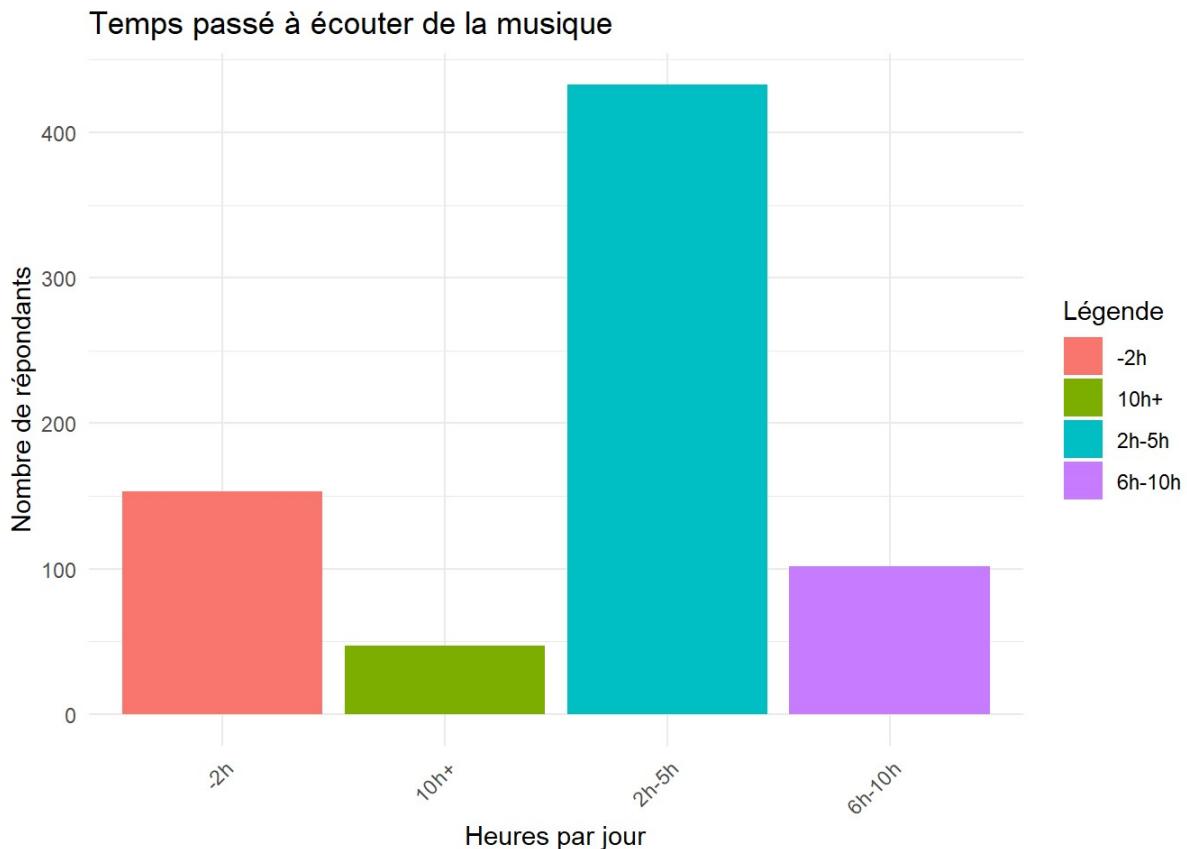
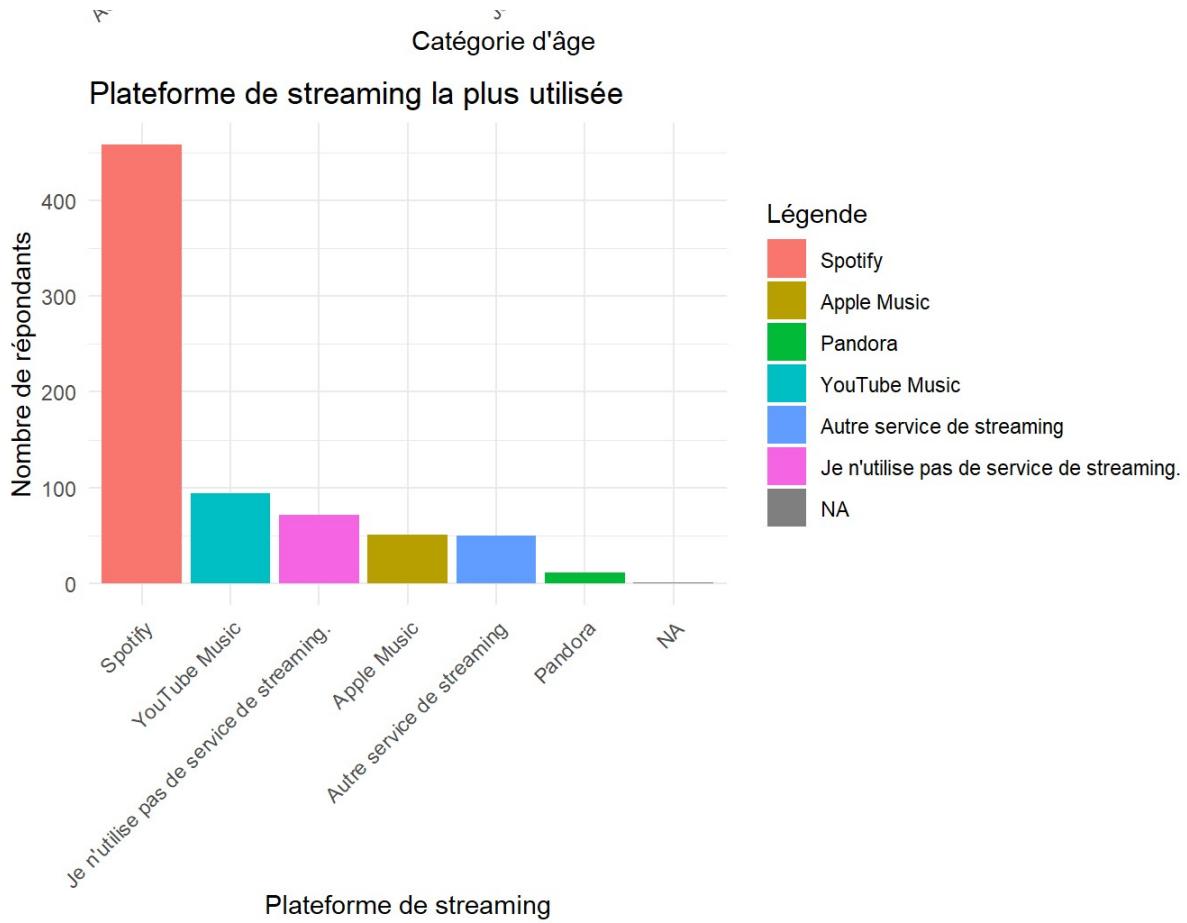
Il est important de noter que notre dataset est d'une part relativement petit, et d'autre part auto-déclaratif. Ainsi, les réponses peuvent être biaisées par la perception de la personne qui répond, et les conclusions que nous pourrons tirer ne seront pas nécessairement généralisables à l'ensemble de la population. De plus, nous ne connaissons pas la répartition des répondants par pays ou par culture, ce qui peut également influencer les résultats de l'analyse.

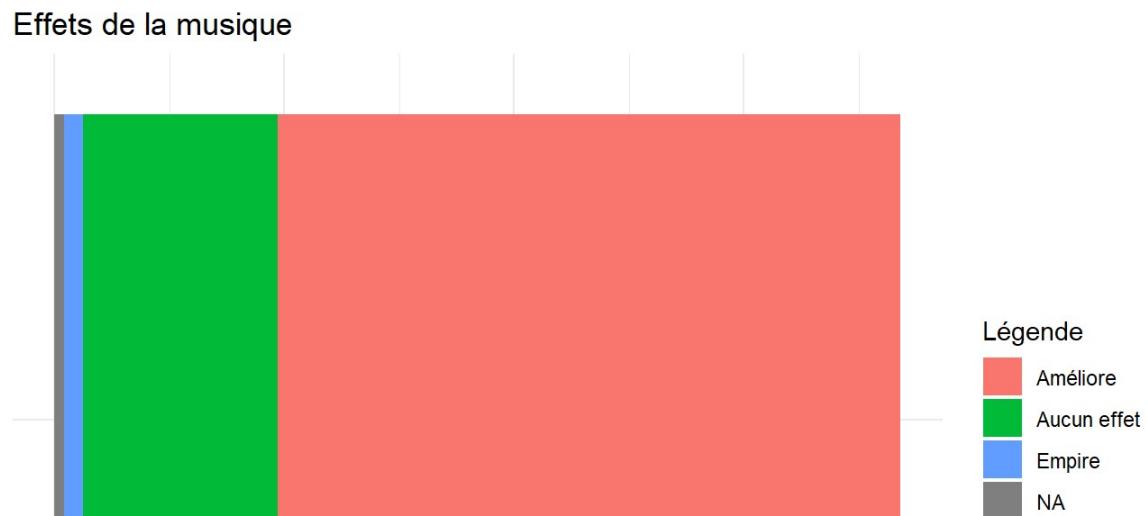
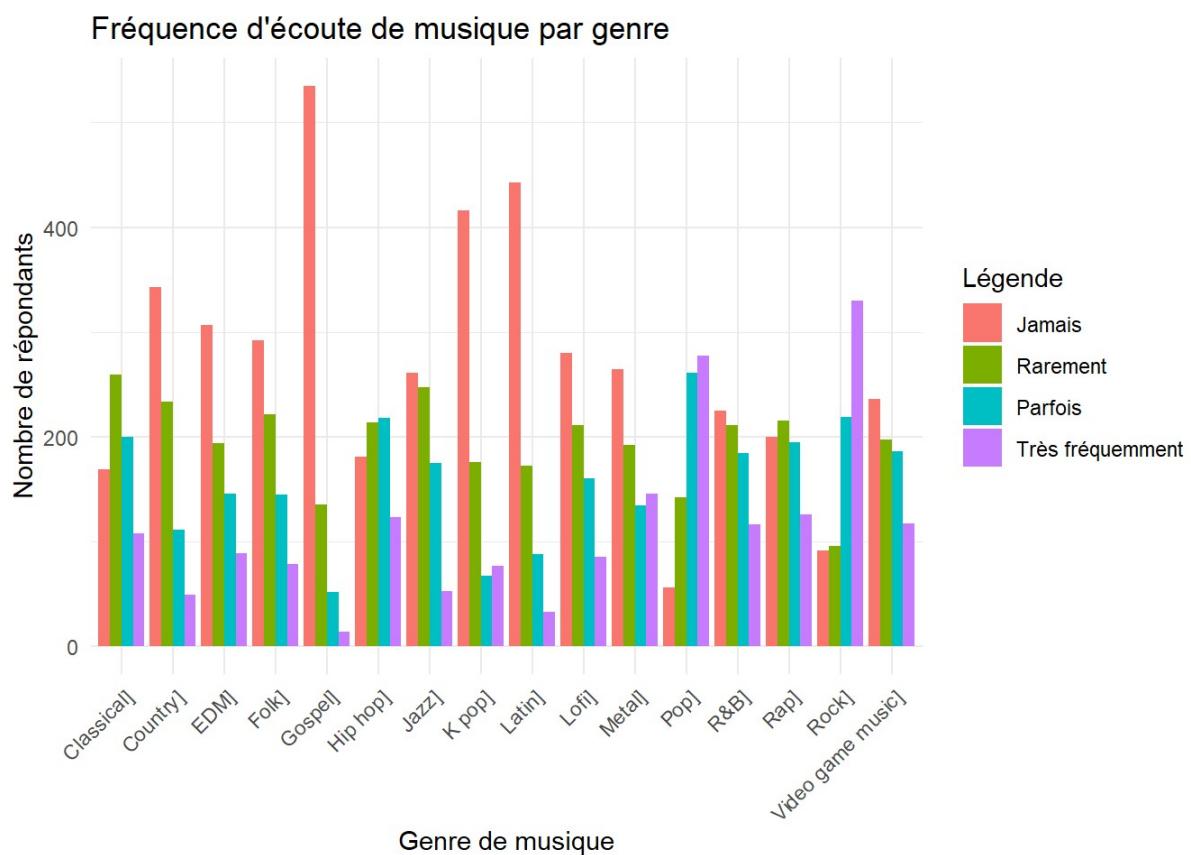
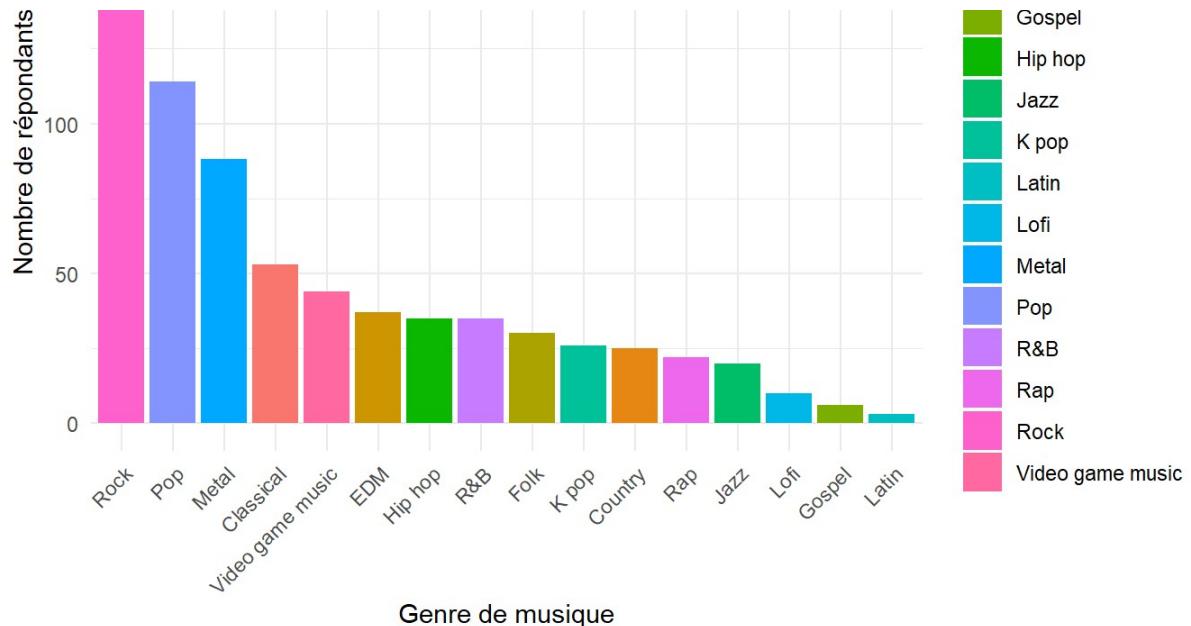
1) Quelles sont les données qui composent le dataset ?

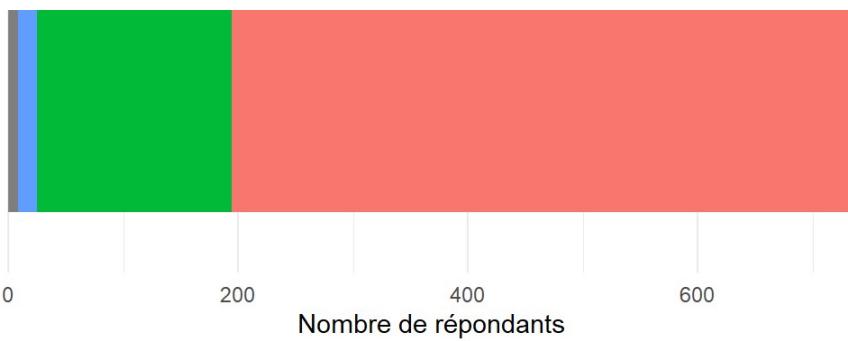
Analyse

Pour cela, nous allons réaliser quelques analyses de base et ainsi réaliser plusieurs graphiques pour visualiser les données.









Langues étrangères

Légende

- Oui
- Non
- NA

Pendant le travail

Légende

- Oui
- Non
- NA

Nombre de répondants

Conclusion

Nous avons 736 répondants dans ce dataset.

Nous avons volontairement exclu certaines des caractéristiques car nous nous y pencherons plus en détails dans les prochaines questions, notamment à propos de la santé mentale (dépression, anxiété, insomnie, OCD) ainsi que les BPM. En effet, ces caractéristiques sont plus complexes à analyser et nécessitent des analyses plus poussées.

En ce qui concerne les graphiques que vous avez pu voir, nous constatons une moyenne d'âge des répondants plutôt jeune, ce qui était prévisible. Nous constatons également que Spotify est la plateforme de streaming la plus utilisée. La majorité des répondants écoutent de la musique entre 2 et 5 heures par jour. Le genre de musique favori qui ressort le plus est le Rock. Enfin, nous constatons que la musique a un effet positif sur la majorité des répondants, qu'ils écoutent de la musique dans des langues étrangères et pendant le travail.

2) Est-ce que la pratique musicale (jouer d'un instrument ou composer de la musique) a une influence sur le temps passé à écouter de la musique ?

Analyse

Avant toute chose, puisque certaines personnes ont répondu qu'elles passaient plus de 18 heures par jour (ou moins de 30 minutes par jour) à écouter de la musique, nous avons décidé de gérer ces outliers à l'aide d'un écart interquartile. Pour cela, nous avons calculé le premier et le troisième quartile ainsi que l'écart interquartile pour la variable Hours per day . Nous avons ensuite filtré les valeurs qui étaient à plus de 1.5 fois l'écart interquartile du premier et du troisième quartile. Cela nous a permis de conserver 693 valeurs sur les 736 initiales.

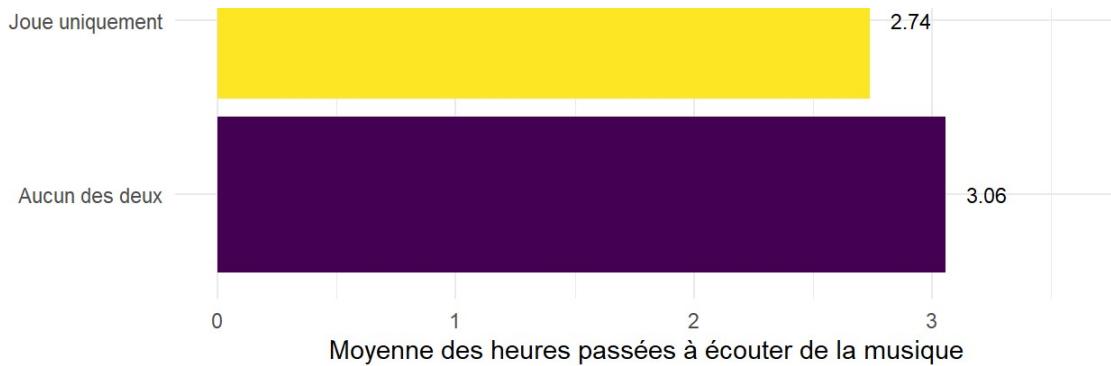
Pour répondre à la question, nous avons décidé de diviser les participants en 4 groupes qui regroupent les différentes pratiques musicales possibles :

1. Ceux qui ne jouent pas d'un instrument et ne composent pas de musique (441 personnes)
2. Ceux qui jouent d'un instrument mais ne composent pas de musique (141 personnes)
3. Ceux qui ne jouent pas d'un instrument mais composent de la musique (29 personnes)
4. Ceux qui jouent d'un instrument et composent de la musique (82 personnes)

Nous avons ensuite calculé la moyenne des heures passées à écouter de la musique pour chaque groupe avant de les comparer.

Moyenne des heures d'écoute en fonction de la pratique musicale





On remarque que les personnes qui composent de la musique passent en moyenne plus de temps à écouter de la musique que les autres groupes.

Étrangement, les personnes qui jouent d'un instrument mais ne composent pas de musique passent en moyenne moins de temps à écouter de la musique que les personnes qui ne font ni l'un ni l'autre. Cela peut s'expliquer par le fait que les personnes qui jouent d'un instrument passent déjà du temps à pratiquer leur instrument et ont donc moins le temps d'écouter de la musique.

Finalement, les personnes qui jouent et composent passent légèrement moins de temps à écouter de la musique que les personnes qui composent sans jouer d'un instrument.

Conclusion

Ainsi, on peut dire que jouer d'un instrument ou composer de la musique semble avoir une influence sur le temps passé à écouter de la musique. Il faudrait néanmoins un dataset avec plus d'échantillons pour confirmer cette hypothèse.

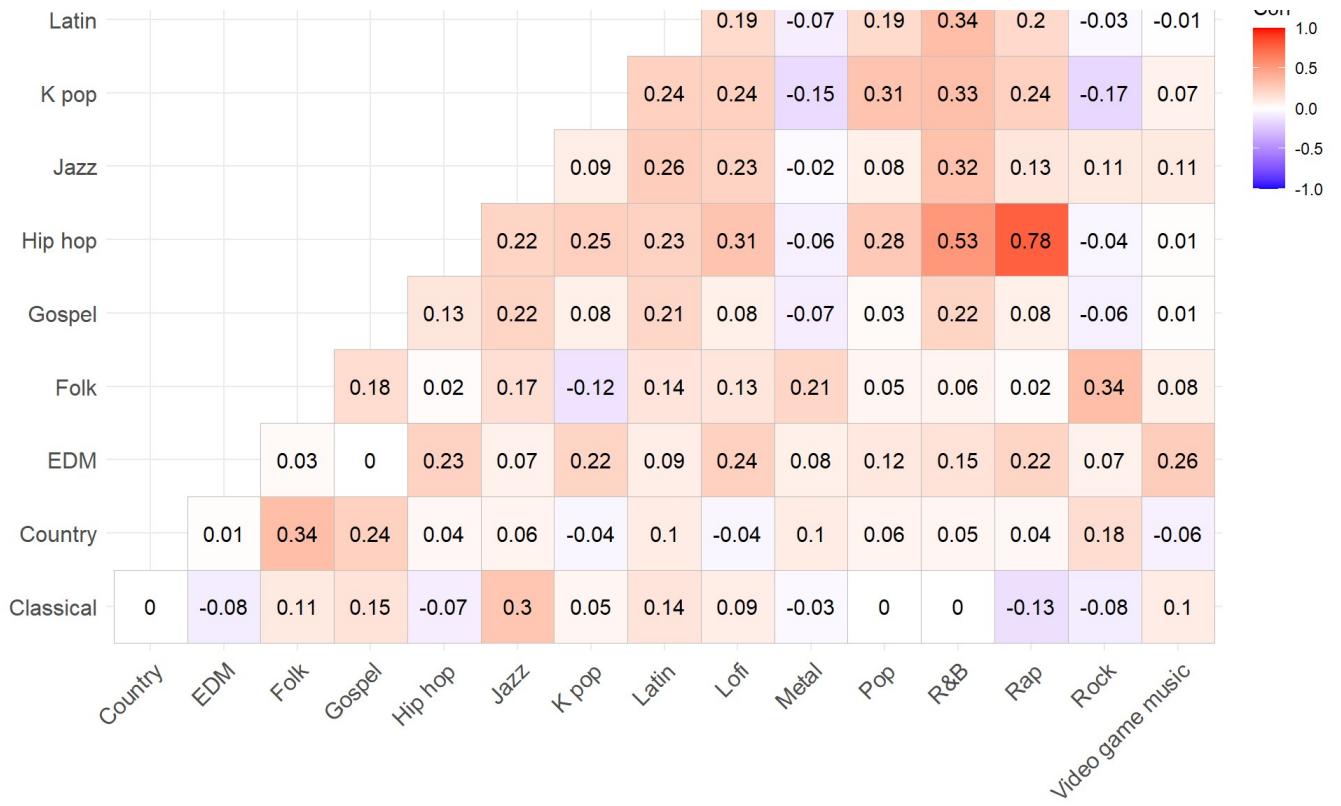
3) Est-ce que l'écoute d'un certain style de musique est corrélée à celle des autres styles de musique ?

Analyse

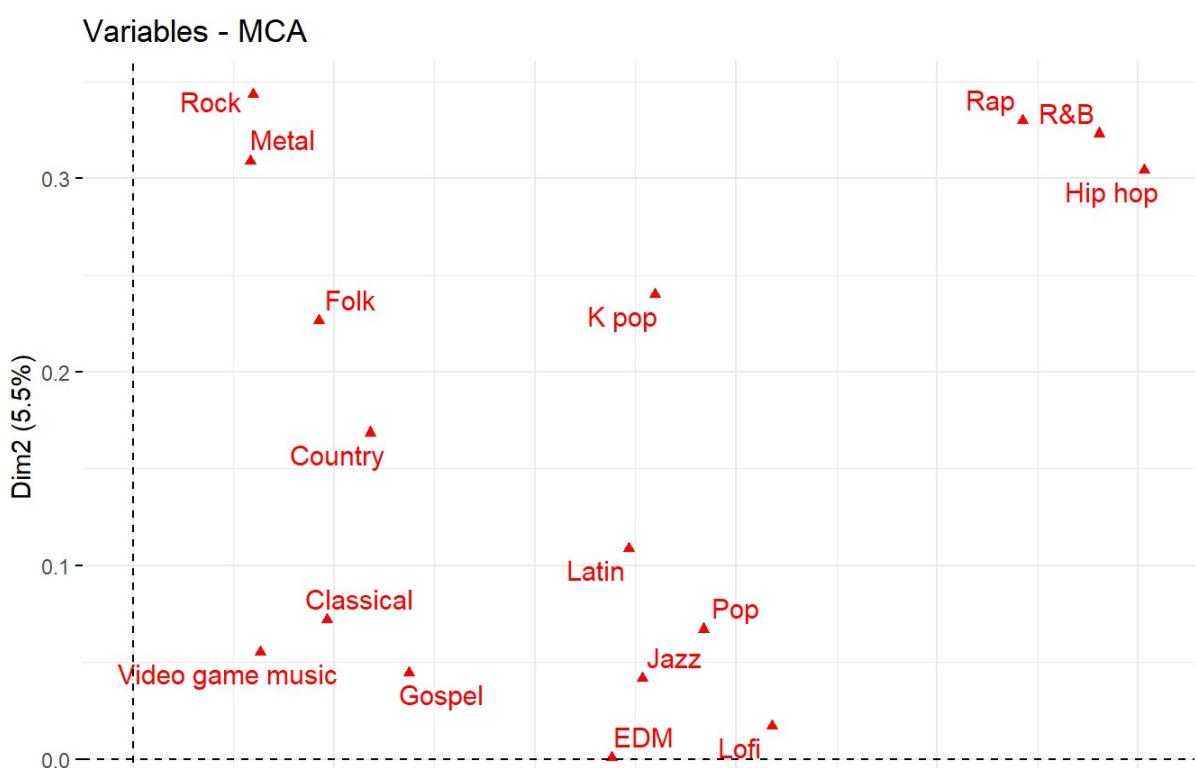
Ce dataset peut aussi nous permettre de nous demander si 2 styles de musiques sont compatibles entre-eux, c'est-à-dire si l'écoute de l'un est corrélée positivement (ou négativement) à l'écoute de l'autre. Pour cela, on peut représenter les coefficients de corrélation entre chaque style dans un tableau.

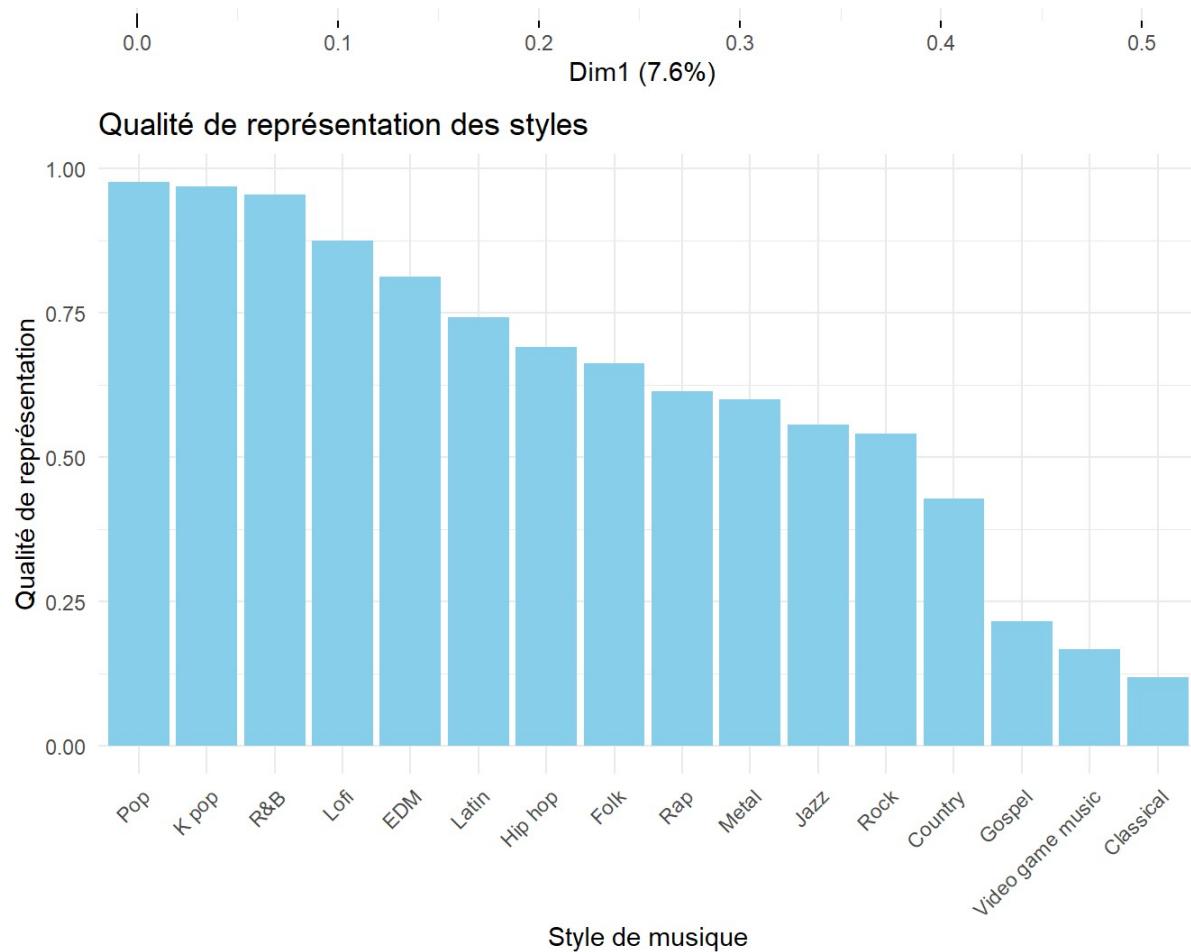
Corrélation entre les fréquences d'écoute des différents styles de musique

	Rock	Rap	R&B	Pop	Metal	Lofi	Corr
Rock							0.12
Rap							0.02 0.04
R&B							0.5 -0.02 -0.08
Pop							0.38 0.28 0.08 0.04
Metal							-0.14 -0.16 0.03 0.51 0.2
Lofi							-0.04 0.25 0.26 0.26 -0.02 0.33



La lecture de ce graphe est assez difficile car les informations sont assez peu visuelles. Nous allons donc faire une ACM. L'Analyse des Correspondances Multiples (ACM) est une méthode statistique utilisée pour explorer les relations entre les catégories de plusieurs variables qualitatives. Elle réduit la dimensionnalité des données et les représente dans un espace de dimension plus faible, facilitant ainsi leur interprétation visuelle. C'est une technique utile pour analyser des tableaux de contingence et découvrir des associations entre les différentes catégories des variables étudiées. Plus 2 variables sont proches sur le graphe, plus elles ont de chance d'avoir la même valeur pour 2 individus distincts. Plus elles sont éloignées, moins il y a d'individus pour lesquels la valeur de ces variables est la même. Autrement dit, ce graphe permet de montrer la corrélation (signée) entre les différents styles de musique.



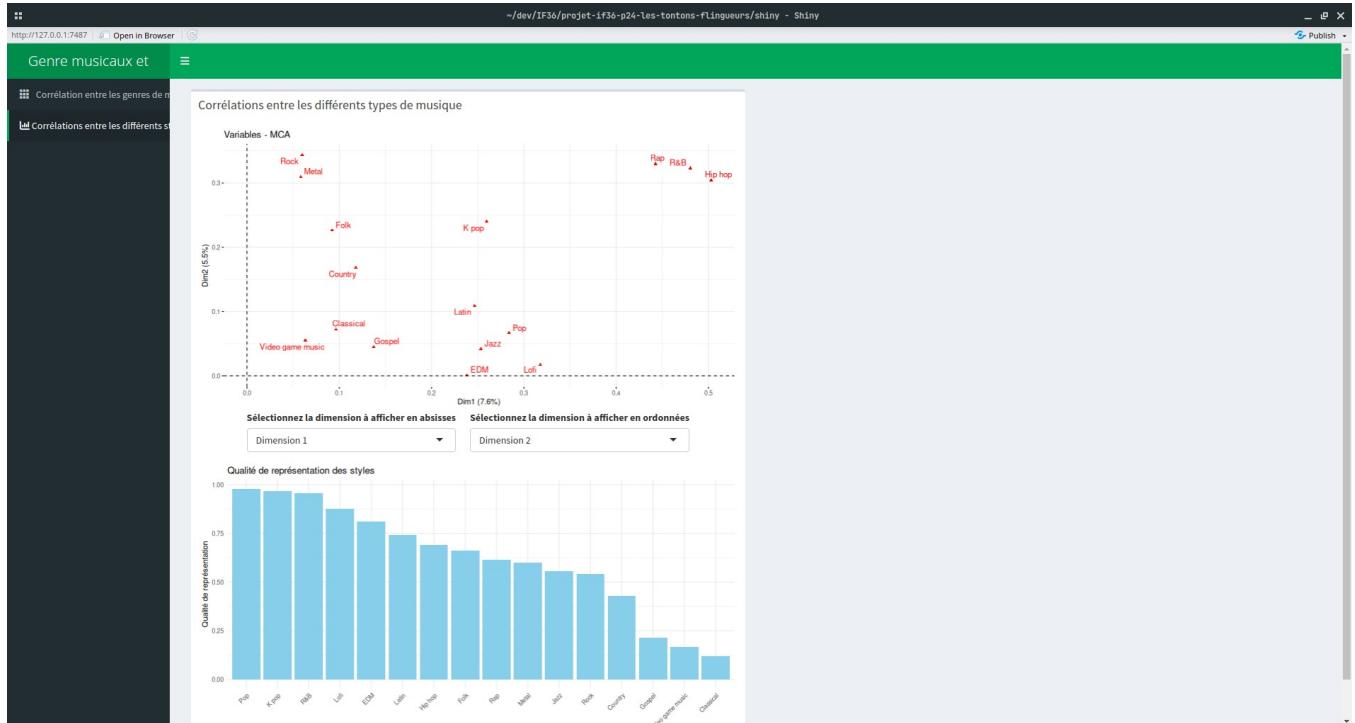


Conclusion

On voit donc bien qu'on peut grouper certains styles. Par exemple, les personnes écoutant du rap écoutent souvent du R&B et du hip hop, et les personnes n'écoutant pas de rock n'écouteront pas de métal. Certains styles sont plutôt à part, comme la K pop, dont la fréquence d'écoute n'a que peu de corrélation avec celle des autres styles. Cela se vérifie dans le tableau précédent.

Ce graphique ne doit cependant que nous donner des pistes de recherche, il ne faut pas tirer de conclusions avec : il n'explique que 13.1% de la variance, il n'est donc pas du tout exact. Certaines variables sont moins bien représentées que d'autres. En effet, pour avoir un graphe parfait, il faudrait le représenter dans un nombre de dimensions élevées. C'est évidemment impossible (sans compromettre la qualité de la lecture), on réduit donc l'espace à deux dimensions. Mais cela nous fait perdre des informations, et pour certaines variables (Gospel, Classical, Video game music), beaucoup d'informations seront perdues, tandis que pour d'autres (Pop, K pop, R&B), la majorité seront conservées. Le deuxième graphique montre donc cela. Il n'est donc pas vraiment possible de tirer de conclusions avec uniquement ce graphique, mais il est tout de même très efficace pour nous donner une idée de la corrélation entre chaque variable.

Pour pouvoir explorer les autres dimensions, nous avons fait un graphe sur Shiny :



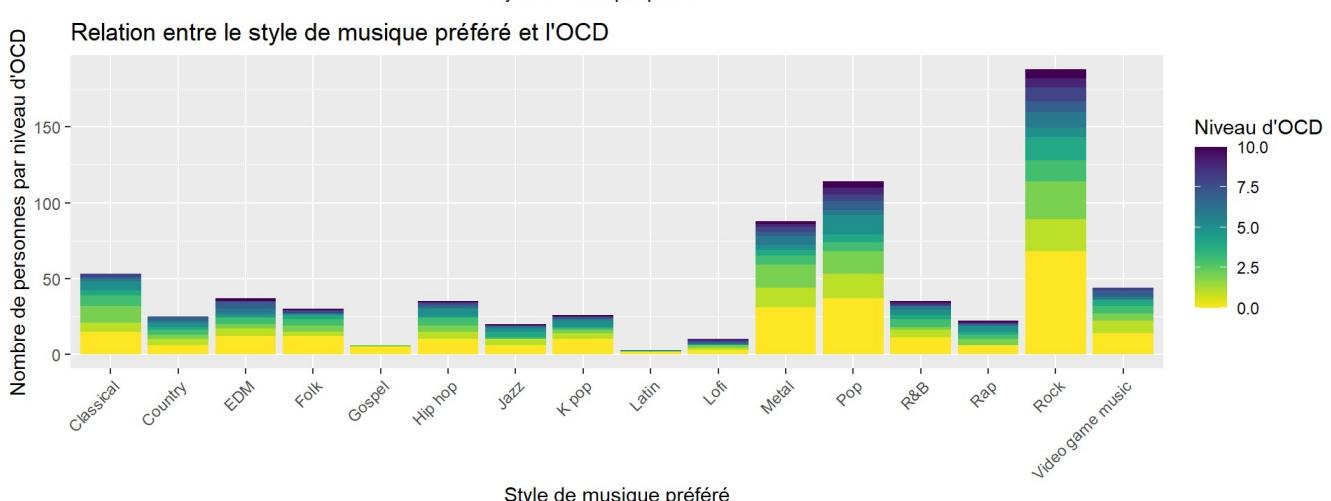
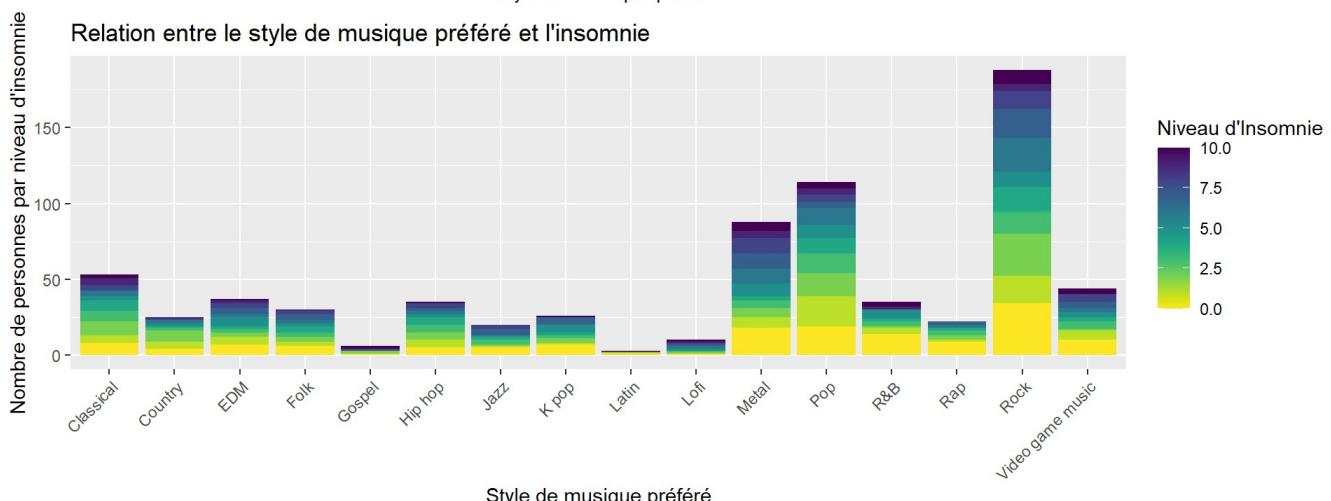
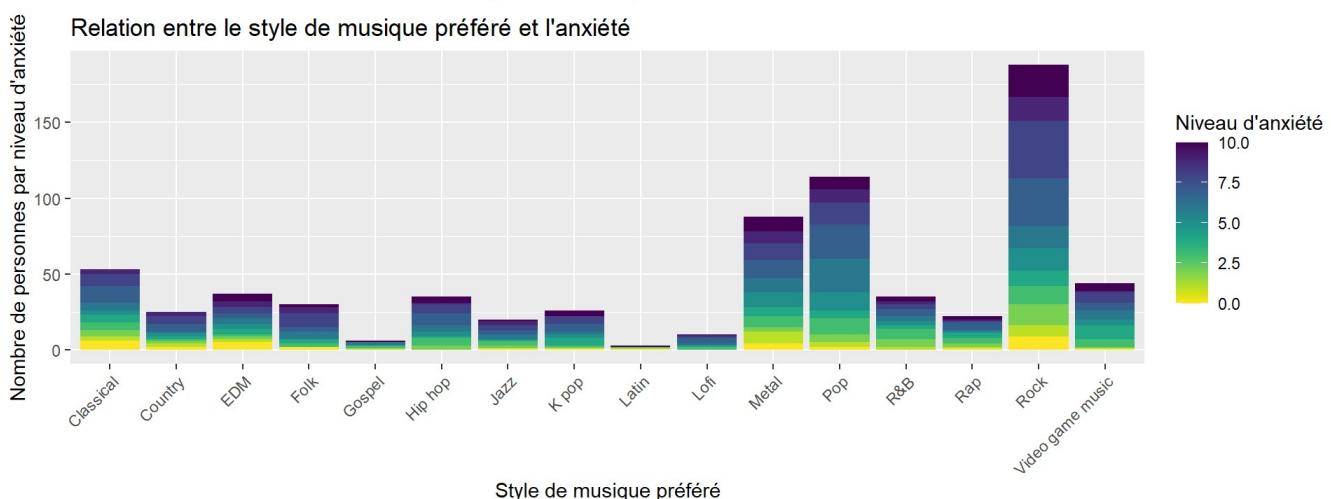
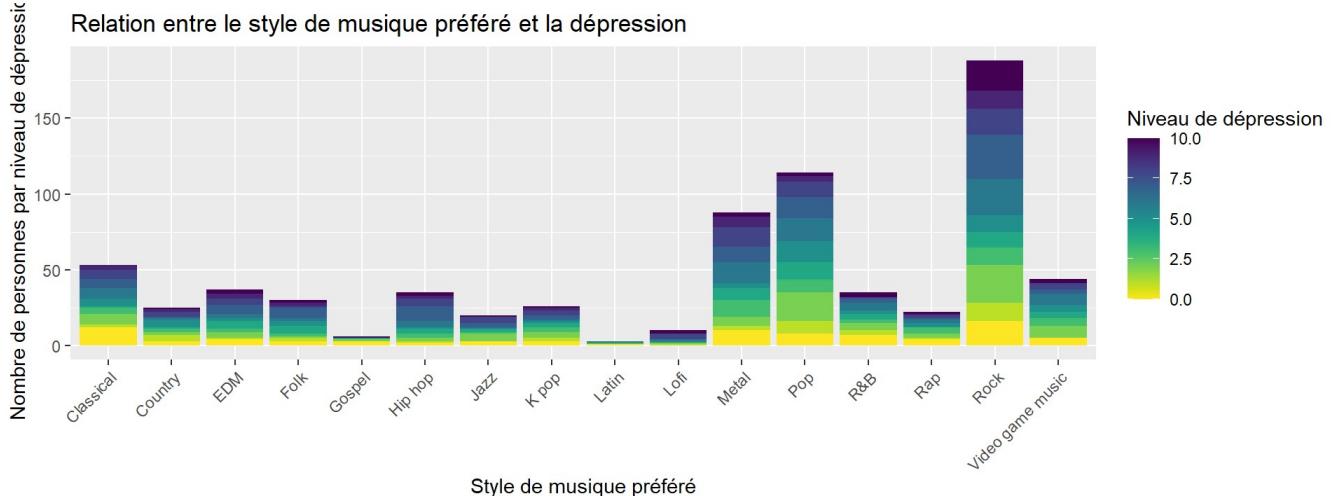
Comme on pouvait s'y attendre, il existe donc bien des corrélations entre les fréquence d'écoute des différents styles de musique. Le meilleur exemple est celle du rap et celle du hip hop, avec un coefficient de corrélation de 0.78.

4) Existe-t-il une corrélation entre le style favori d'une personne et sa santé mentale ?

Analyse

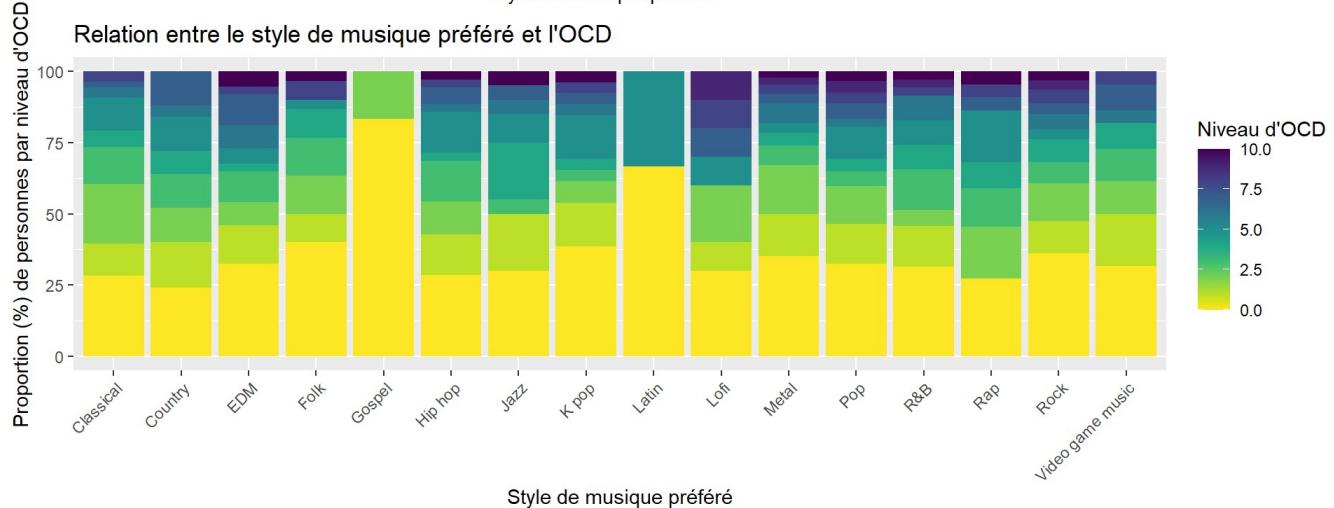
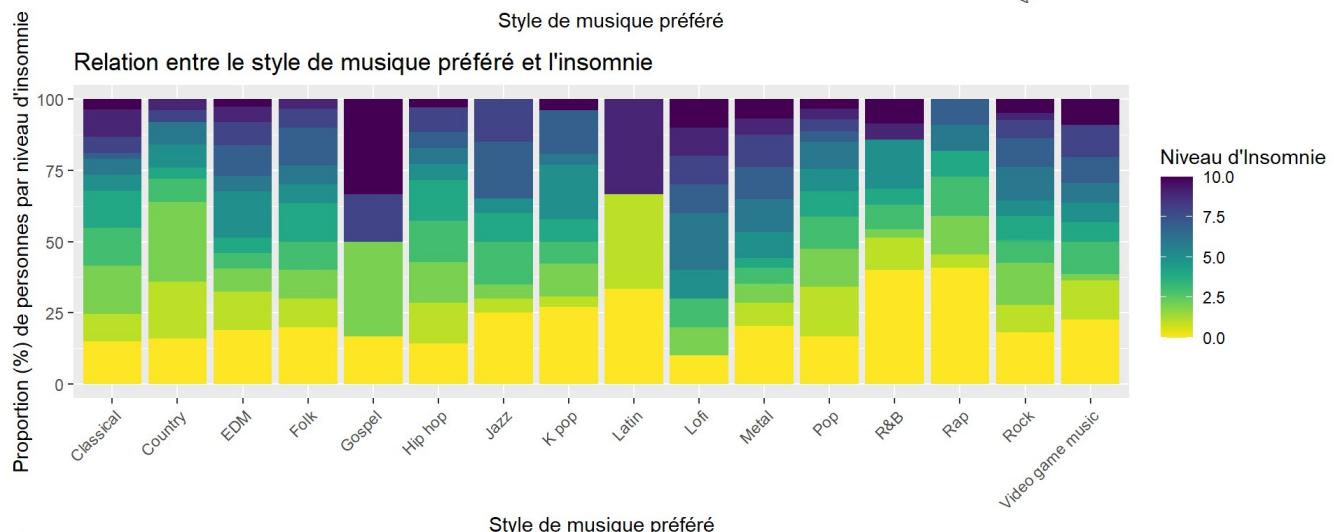
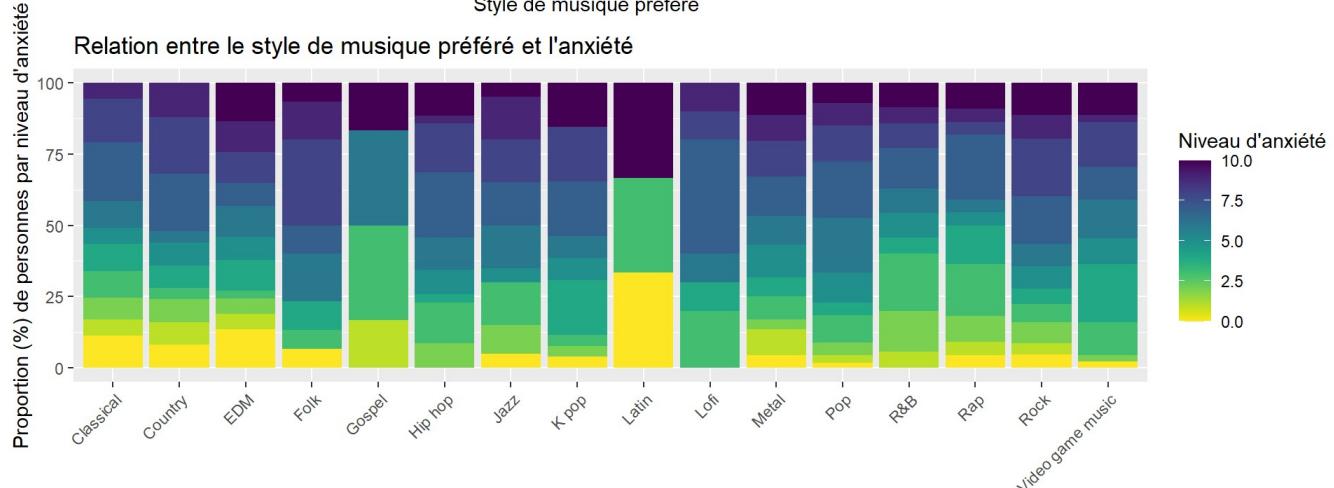
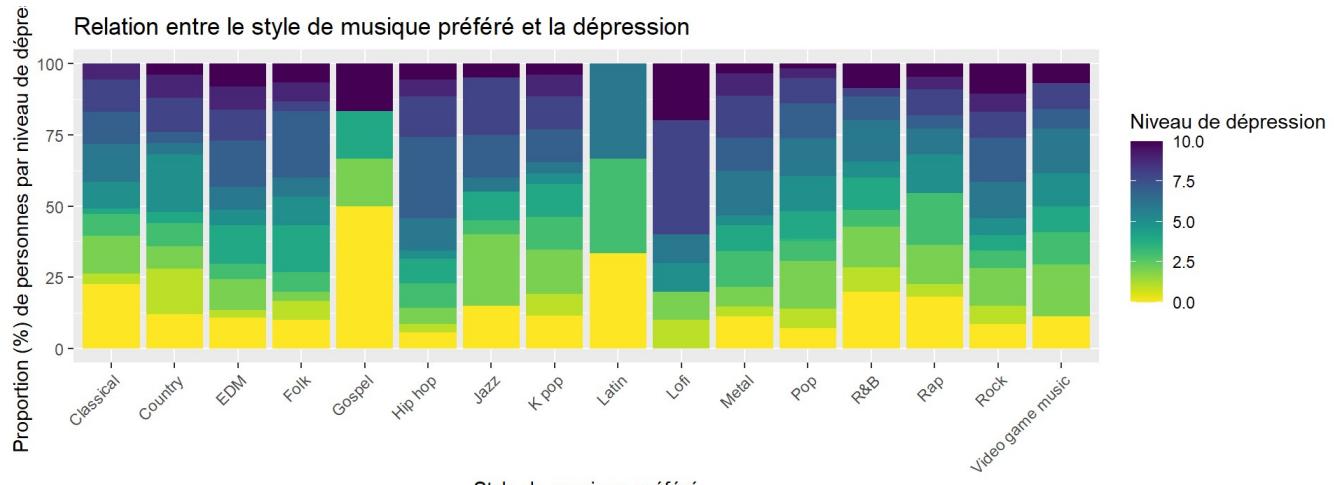
La réponse qui nous viendrait comme cela serait oui, car il semble logique de faire le lien entre “avoir comme style favori un style plutôt lent/calme” et “avoir des problèmes de santé mentale”.

Pour vérifier cela, observons d'abord pour chaque style de musique le nombre de personnes qui l'ont en favori ainsi que leur niveau auto-évalué de dépression, d'anxiété, d'insomnie et d'OCD. À savoir que les échelles sont continues car certains se sont évalués l'insomnie et la dépression à 3.5, alors que d'autres se sont évalués l'OCD à 5.5 ou 8.5 et l'anxiété à 7.5. Garder une échelle discrète rendrait ainsi bien moins simple la comparaison entre les 4 graphiques, l'échelle variant d'un graphique à un autre.



La première chose que nous remarquons est que certains styles de musiques sont bien moins populaires que d'autres, et cela peut avoir un impact sur la fiabilité des conclusions qu'on pourra tirés, des styles comme le Latino, le Lofi ou encore le Gospel n'ayant que des échantillons très réduit de personnes les représentant. La deuxième chose notable est la prévalence des différents problèmes mentaux, qui du plus prévalent au moins prévalent semblent être l'anxiété, la dépression, l'insomnie et enfin l'OCD. Mais difficile de conclure avec assurance ici.

Il nous est au final complexe de conclure quoique ce soit d'autre avec de tels graphiques, les grandes différences de taille d'échantillon rendant les différents styles difficilement comparable. Il nous faut donc aller plus en loin en regardant plutôt les proportions par rapport à chaque échantillon global de chaque style au lieu des nombres bruts.

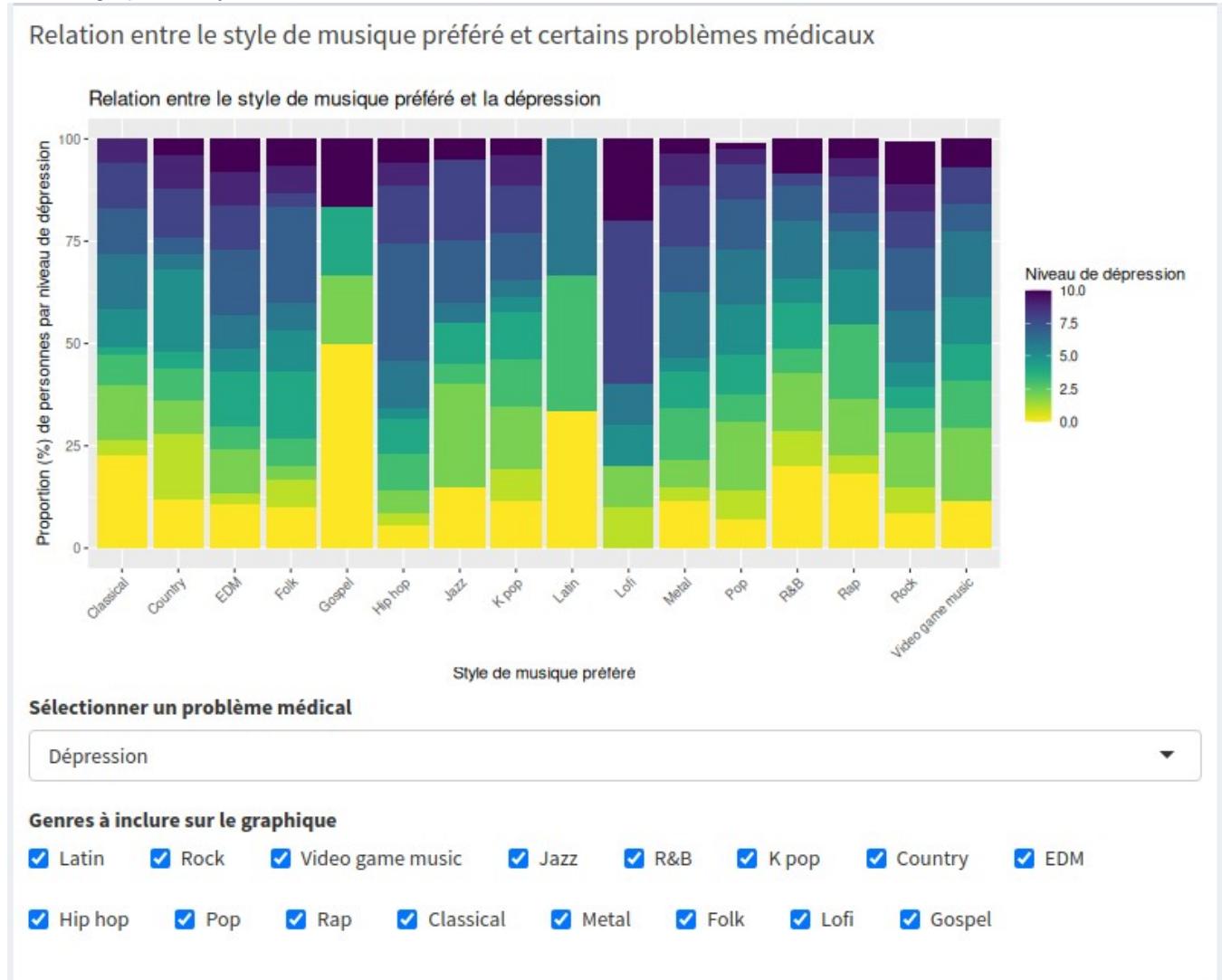


Les styles qui se démarquent le mieux sont comme attendu ceux qui possèdent les plus petits échantillons, on a d'abords le latino qui semble être le style le plus corrélé positivement aux différents problèmes mentaux, et à l'opposé on a le Lofi qui semblent être le plus corrélé négativement. Ce qui correspondrait à l'hypothèse de départ, le Lofi étant un style qui se caractérise principalement par sa lenteur et son calme, et le Latino étant un style très dansant. De plus, ces 2 styles sont possiblement les styles présents les plus codifiés dans leur ambiance, pouvant transmettre au final une palette assez restreinte d'émotions. Ceci pourrait expliquer ce que nous voyons, là où les autres styles vont posséder des morceaux ne transmettant pas du tout les même émotions l'un par rapport à l'autre. Mais il est important de garder en tête que l'échantillon Latino n'est composé que de 3 personnes, et Lofi d'une dizaine de personnes, rendant tout ceci très hypothétique.

Ensuite, le Gospel est le style dont les résultats varient le plus d'un problème mental à un autre, étant parfois très corrélé positivement comme avec l'OCD, et parfois très négativement comme avec l'insomnie, sans doute une conséquence de son échantillon de 6 personnes.

Les autres styles, possédant de plus importants échantillons, sont plus similaires les uns aux autres. Il n'y a pas vraiment de styles qui se démarquent franchement, puisque souvent par exemple si un style possède deux fois plus de niveau 1 que de niveau 2 qu'un autre, ils auront au final la même proportion de personnes avec un niveau inférieur à 2, comme le classique et le country pour la dépression. Ceci pourrait simplement s'expliquer par le fait qu'une division d'un problème mental en 11 niveaux différents est peu pertinent pour une auto-évaluation, l'Homme ayant du mal à voir une différence entre deux niveaux côté-à-côte.

Nous pouvons réduire rendre les graphiques moins spacieux et rendre les informations qui nous intéressent plus lisibles avec un graphe Shiny :



Conclusion

Ainsi, il semble difficile de répondre oui à notre question, surtout que le peu de corrélation que l'on a pu percevoir semble plus être une implication de l'ambiance des morceaux plutôt que des styles à proprement parler. Mais une chose importante que nous n'avons pas pris en compte est le fait qu'avoir un style favori ne signifie pas forcément que l'on écoute beaucoup plus ce style que d'autres. Ce pourrait même parfois être l'inverse. De plus regarder la corrélation entre la musique l'intensité des problèmes mentaux ne nous dit rien de si la musique a un impact positif ou négatif, cela peut au mieux juste nous dire si certains styles sont meilleurs pour la santé que d'autres. Il semble donc être plus intéressant de se poser la question "Les styles de musiques peuvent-ils avoir un impact négatifs ou positifs sur les problèmes mentaux ?".

5) La musique peut-elle avoir un impact négatif ou positif sur les problèmes mentaux en fonction du genre musical ?

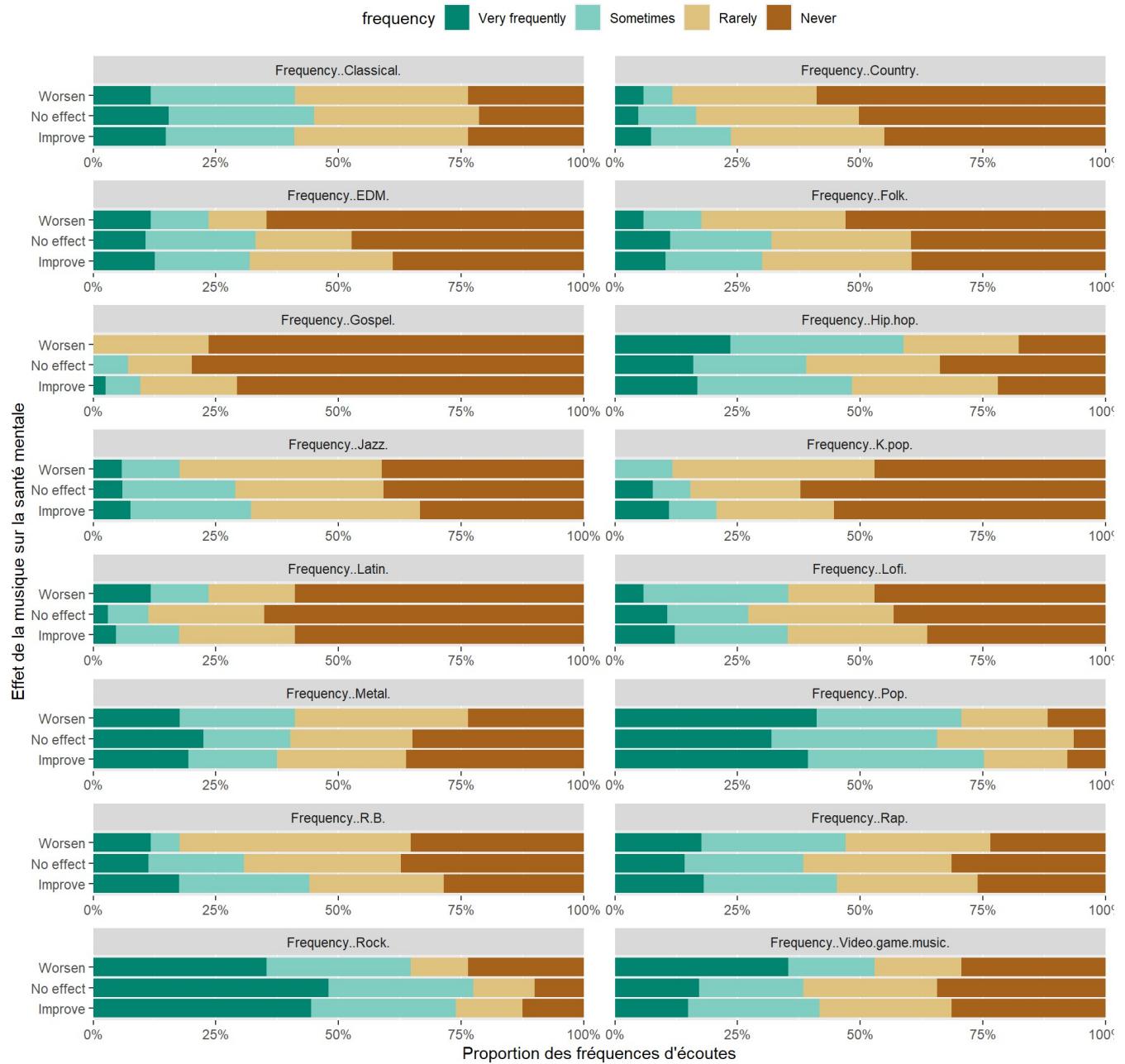
Analyse

Personnellement, les résultats de la question précédente me donnent peu d'espoir pour trouver une quelconque corrélation.

Pour répondre à cette question, étudions la possible corrélation entre la fréquence d'écoute d'un genre avec l'effet de la musique sur la santé mentale, au moyen d'un barplot qui présente le dit effet en ordonnée ainsi que la proportion des fréquences d'écoutes en abscisse.

Mais avant toute chose, deux choses sont à noter. La première est que 8 personnes n'ont rien indiqué quant à l'impact de la musique sur leur santé mentale, et qu'il a donc fallu retirer ces lignes du dataset pour ce graphique. La seconde est que le nombre de personnes ayant répondu "Worsen" à l'impact sur la santé mentale s'élève à 17, ce qui est donc un petit échantillon.

Etude de la corrélation entre la fréquence d'écoute et l'effet de la musique sur la santé mentale



Conclusion

En premier lieu, on remarque que la plupart des genres musicaux possèdent des proportions relativement équivalentes entre "Worsen", "No effet" et "Improve", ce qui tendrait à montrer qu'il n'y a pas de corrélation, au moins pour ces styles là. On a notamment le Classique, le Jazz, le Lofi, le Rap, la Pop, le R&B, le Gospel et le Métal. D'autres semblent avoir un impact positif minime, comme le Rock, la K-pop ou le Folk. D'autres au contraire semblent avoir un impact négatif minime, comme le Latino, le Hip-Hop, et surtout les musiques de jeux-vidéos.

Mais il n'y a pas l'air d'y avoir de lien entre les styles qui auraient un impact négatif, et de même pour ceux semblant avoir un impact positif. Ainsi, il est plus probable que ce soit des erreurs statistiques, sachant que la plupart des styles ne semblent pas avoir d'impact du tout. Et au final, le problème avec cette méthode est que l'effet de la musique sur la santé mentale est une variable globale, et non spécialisée à chaque genre musical. Ainsi, sa valeur est nécessairement multifactorielle et il est peu probable que par exemple, écouter des musiques de jeux vidéos suffisent seul à créer un impact négatif de la musique. De plus, ayant seulement un petit échantillon de "Worsen", ces résultats pourraient être invalidés.

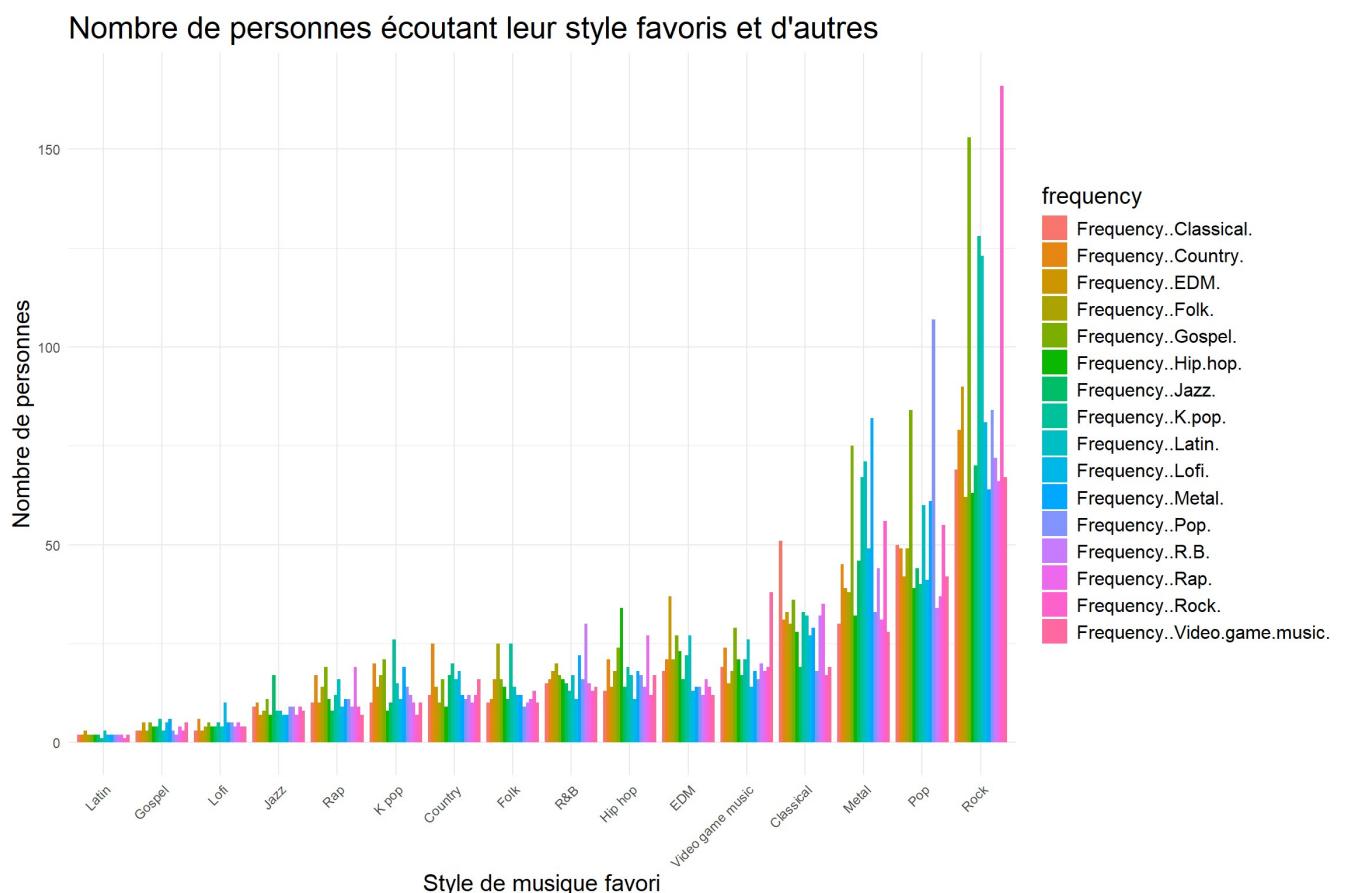
À travers tout ça, la réponse à la question posée semble donc plus être non.

6) Le style de musique favori influence-t-il l'écoute d'autres styles de musique ?

Analyse

Au début de la question précédente, nous avons fait l'hypothèse qu'une personne n'écoute pas seulement son style favori, voir même qu'il pouvait en écouter d'autres davantage. Ainsi, le style de musique favori influence-t-il l'écoute d'autres styles de musique ? La réponse devrait être positif, mais difficile d'estimer un ordre de grandeur. Pour cela, nous allons regarder la répartition des préférences musicales des répondants en fonction de leur style de musique favori.

Pour répondre à cette question, nous avons analysé la répartition des préférences musicales des répondants en fonction de leur style de musique favori. Le graphique ci-dessus représente le nombre de personnes écoutant leur style favori ainsi que d'autres styles, ventilé par fréquence d'écoute.

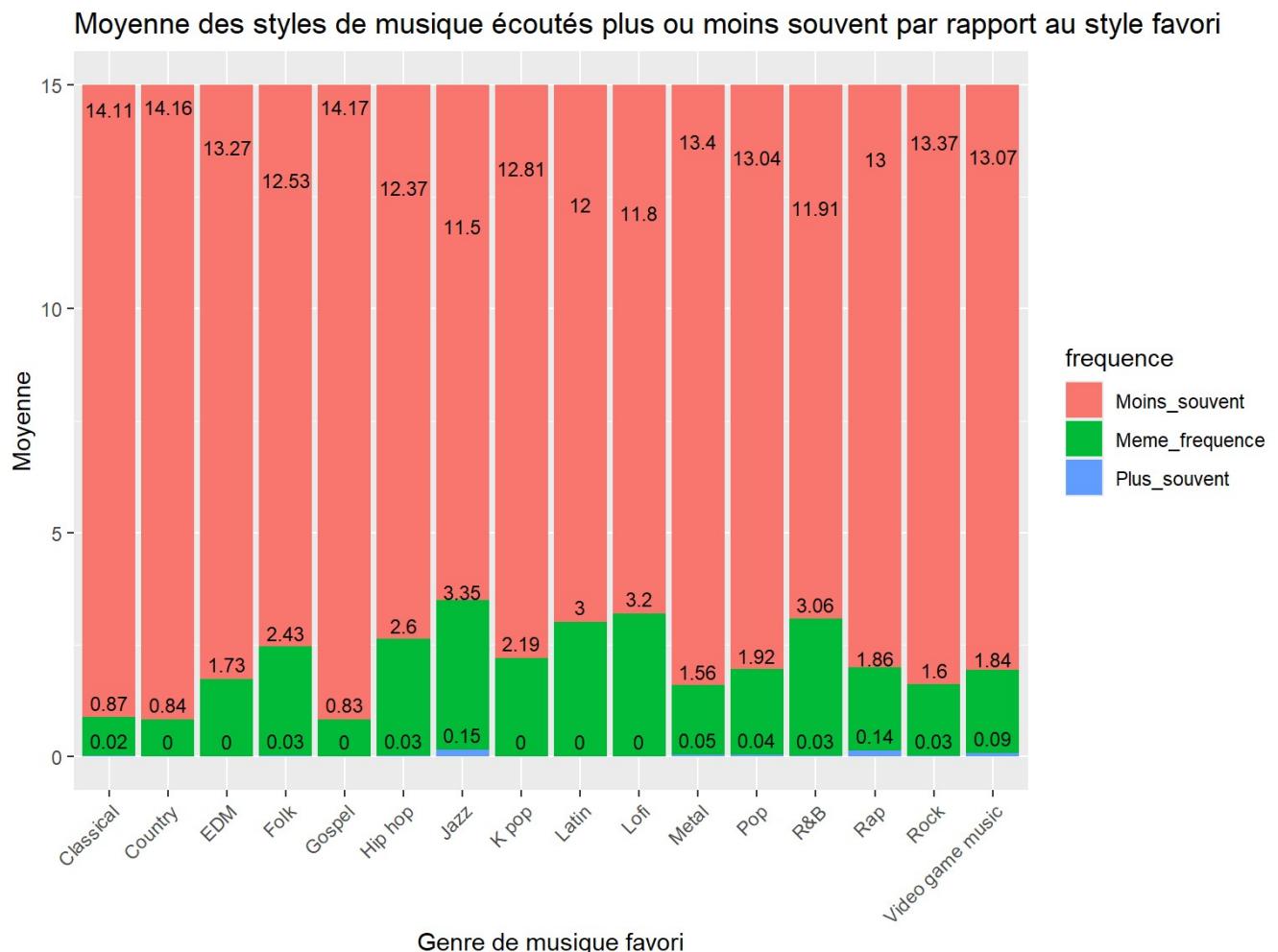


Nous constatons que les préférences musicales varient considérablement en fonction du style favori de chaque individu. Par exemple, les amateurs de musique pop ont tendance à écouter d'autres genres musicaux de manière plus sporadique, tandis que les fans de rock et de metal ont une propension plus élevée à explorer une variété de styles de musique.

En allant davantage dans les détails, on peut voir que pour le style rock, beaucoup écoutent également du gospel, de la kpop ou du latin, et un peu plus de moitié moins les autres styles. Pour le style pop, on peut voir que beaucoup écoutent également du gospel, et quasiment moitié moins du metal ou du latin. En ce qui concerne le style classique, la répartition est plutôt homogène, bien que beaucoup restent à n'écouter que du classique. Voilà pour ce qui est des trois plus gros styles de musique du dataset en terme de nombre de personnes les ayant comme favoris. En ce qui concerne

les autres, les répartitions sont plutôt homogènes, bien que beaucoup restent à n'écouter que leur style favori. Il faut néanmoins rester vigilant quant à cette conclusion car, et c'est la raison pour laquelle je ne fais qu'une rapide conclusion sur ceux-là, le nombre de répondants n'est pas suffisant pour en tirer des conclusions fiables.

Mais le graphique étant relativement complexe, essayons d'y voir plus claire en retirant des informations peu utiles. Pour cela, observons donc par style de musique favori, la moyenne de styles écoutés plus souvent, à la même fréquence et moins souvent que le style favori.



On remarque tout de suite qu'en moyenne, il y a très peu de gens qui écoutent davantage un style autre que leur style favori, variant entre 0 et 0.15 styles écoutés plus souvent. Mais ce chiffre grimpe tout de suite lorsqu'il s'agit de la même fréquence, ce qui correspond à ce que nous attendions. On remarque aussi qu'il existe une certaine disparité entre les différents, avec 4 styles favori à plus de 3, et 3 autres style en dessous de 1, alors qu'en plus ces deux groupes possèdent un nombre de personnes relativement important.

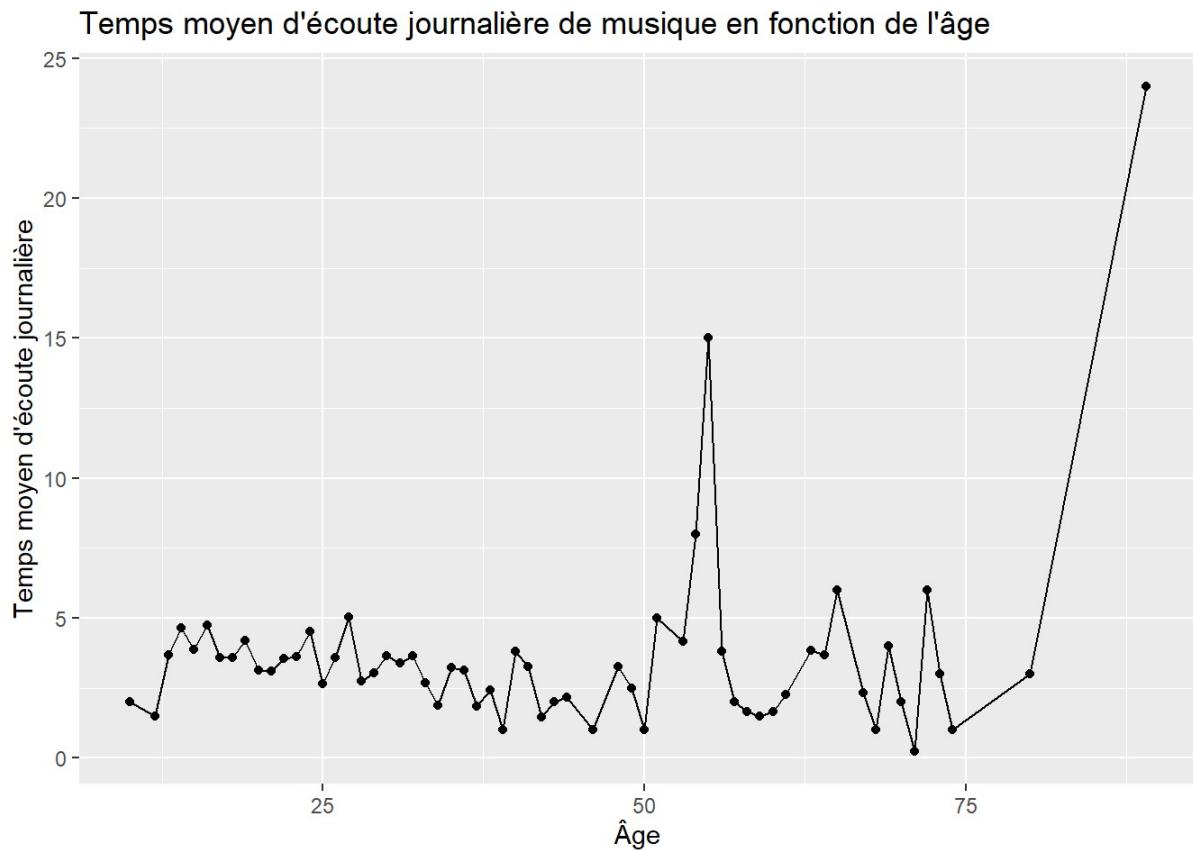
Conclusion

Cette analyse suggère que le style de musique favori a une influence sur les habitudes d'écoute d'autres genres musicaux. Cependant, il est important de noter que ces résultats sont basés sur des données auto-déclarées, et les conclusions doivent être interprétées avec prudence. D'autres facteurs tels que l'âge, le contexte culturel et les expériences individuelles peuvent également jouer un rôle dans les préférences musicales et les habitudes d'écoute.

7) Est-ce que l'âge est corrélé avec le temps d'écoute quotidien de musique ?

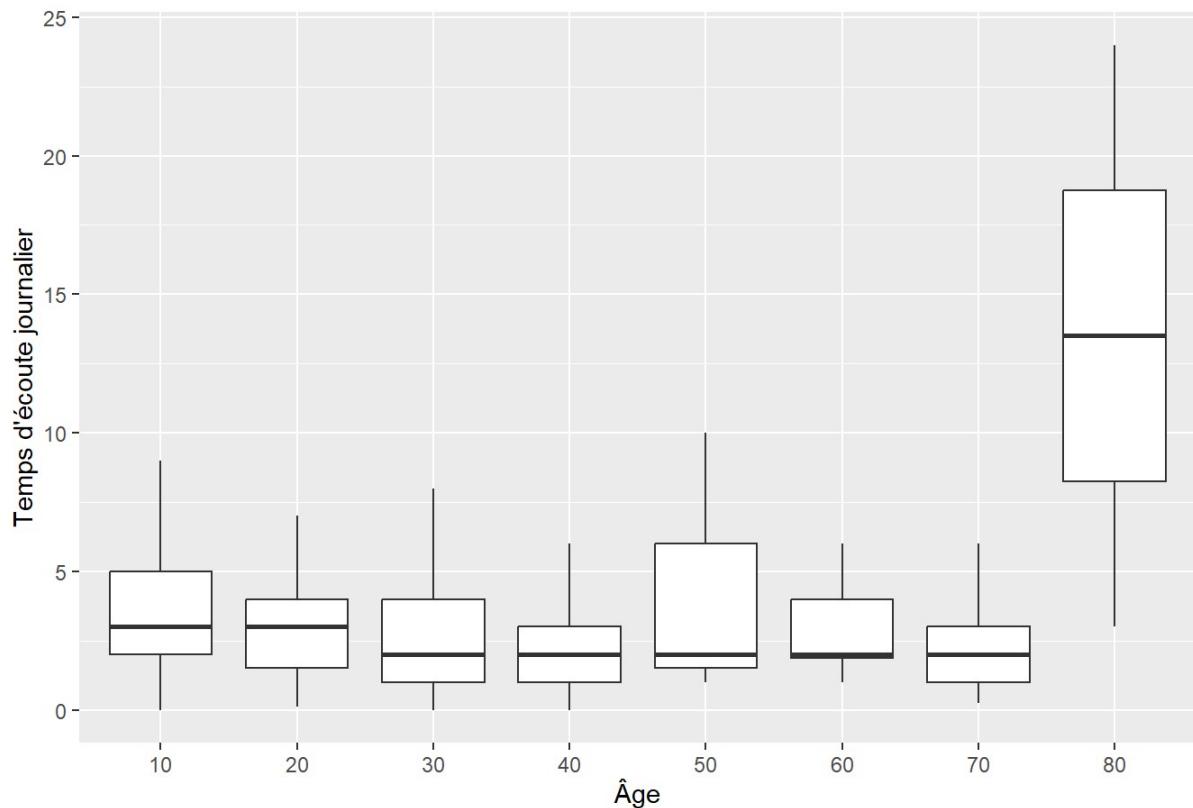
Analyse

Nous pouvons ensuite nous demander si l'âge est corrélé avec le temps d'écoute quotidien. On peut très bien imaginer que oui : il arrive souvent qu'on voit des personnes jeunes avec des écouteurs ou des casques dans les transports en commun, mais on imagine moins des personnes agées écouter de la musique à tout instant.



Le résultat est assez étonnant : l'écoute quotidienne de musique a tendance à baisser un peu avec l'âge, mais ce n'est pas aussi marqué que ce que nous aurions pu penser. Il y a évidemment des valeurs aberrantes sur ce graphique, notamment la personne de 89 ans qui écoute de la musique 24/24h. On peut aussi regarder faire des box plots, pour pouvoir mieux comprendre la répartition. Pour éviter de faire chaque box plot à partir de seulement quelques données, on va regrouper les personnes par tranche de 10.

Temps d'écoute journalier de musique en fonction de l'âge des répondants



Avec ce graphique, on comprend mieux la répartition des temps d'écoute. On peut voir que jusqu'à 50 ans, le temps d'écoute a tendance à baisser, puis il ré-augmente légèrement pendant la cinquantaine avant de lentement redescendre.

Conclusion

Finalement, la corrélation est moins forte que nous le pensions. On a cependant vu qu'il y a une petite tendance à la baisse. Il y a ensuite un nouveau pic d'écoutes aux alentours de 50 ans.

8) Existe-t-il une corrélation entre le niveau d'un trouble mental, l'effet de la musique auto-perçu, et le temps d'écoute ?

Analyse

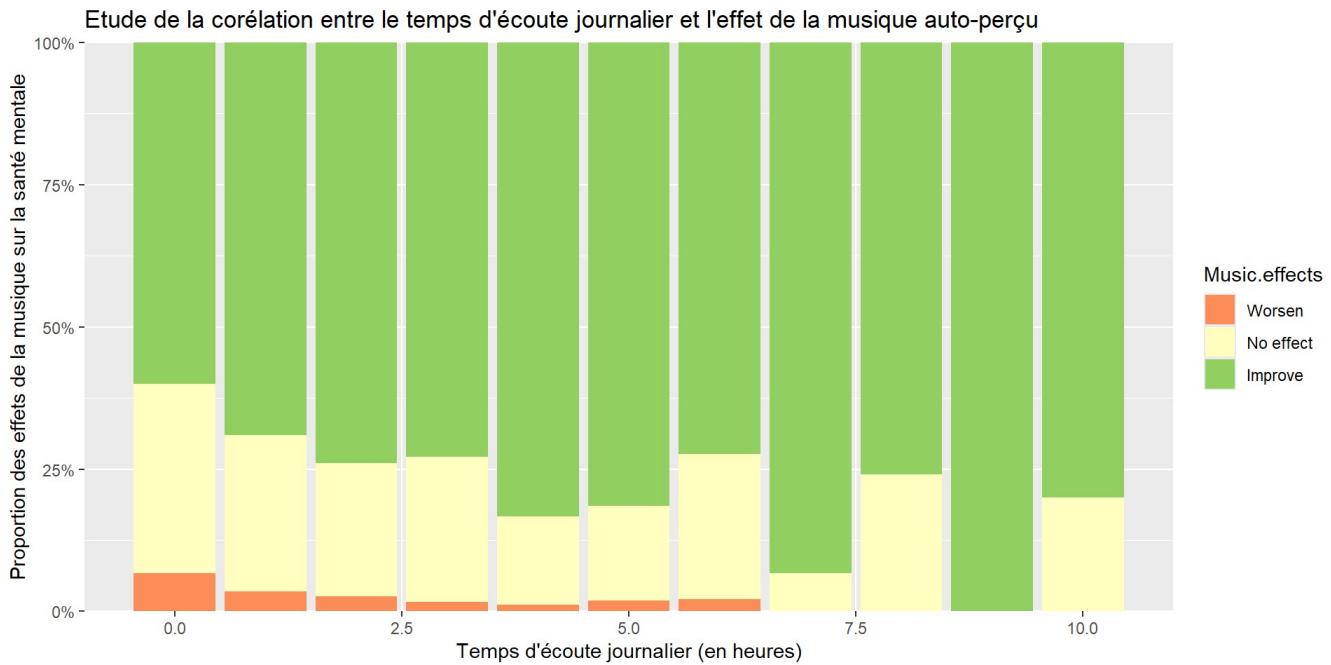
Posons-nous maintenant la question suivante : Existe-t-il une corrélation entre le niveau d'un trouble mental, l'effet de la musique auto-perçu, et le temps d'écoute ? Personnellement, je ne pense pas que le trouble puisse être lié à l'impact de la musique ou au temps d'écoute, alors que je pense le temps d'écoute corrélé avec l'impact. Pour vérifier cela, nous allons d'abords regarder par trouble la proportion de personnes qui considèrent l'impact positif, négatif ou neutre par niveau d'intensité de chaque trouble mental.

Etude de la corrélation entre le niveau d'un trouble mental et l'effet de la musique auto-perçu



La première chose que l'on remarque est que pour la plupart des sondés, la musique a un impact positif. Cet impact semble indépendant du niveau d'insomnie et d'OCD, mais semble cependant légèrement corrélé négativement au niveau d'anxiété (même si l'on pourrait émettre une réserve sur le fait qu'au contraire le nombre d'impact négatif plus neutre lui diminue avec l'intensité) et davantage au niveau de dépression, avec 13% de personnes avec une dépression de niveau 10 considérant que la musique a un impact négatif. Cependant, il ne faut pas oublier que l'effet de la musique reste un terme assez vague, cela ne signifie pas par exemple qu'écouter de la musique aide à contrer l'insomnie. Et peut-être que ce que l'on observe est lié au temps d'écoute.

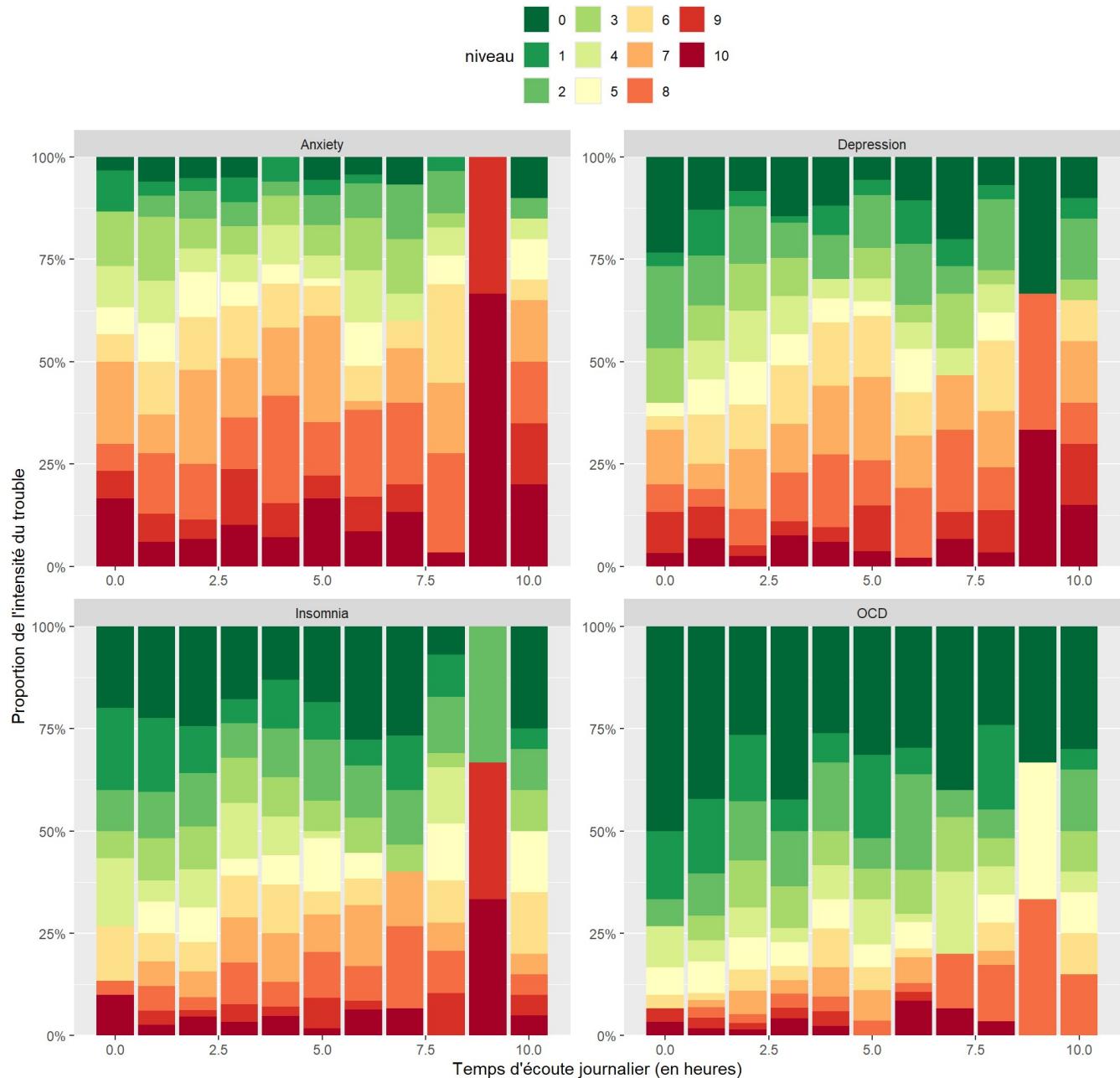
Ainsi, il semblerait intéressant d'incorporer dans cette analyse le temps d'écoute. Pour ce faire, nous allons regarder la corrélation de cette dernière avec l'intensité de chaque trouble ainsi qu'avec l'effet de la musique auto-perçu.



On remarque une légère tendance montrant une diminution de l'impact négatif et une augmentation de l'impact positif avec l'augmentation du temps d'écoute, ce qui semblerait montrer qu'au global, plus on écoute de la musique, mieux l'on se porte. Mais comme cela ne montre qu'une corrélation, ce pourrait très bien être l'inverse : les personnes considérant que la musique a un impact négatif pourraient avoir tendance à en écouter moins justement, ce qui serait plus logique d'après moi. À noter que l'on a retiré du graphique les personnes avec un temps d'écoute supérieur à l'IQR, car il y avait quelques valeurs aberrantes comme 24h d'écoute par jour.

Maintenant, vérifions si par exemple pour la dépression, les personnes présentant une haute intensité montre un temps d'écoute plus faible. Si oui, alors c'est uniquement temps d'écoute qui définirait l'impact de la musique. Si non (c'est-à-dire si on observe une corrélation positive), alors la dépression aurait bien un impact sur l'effet de la musique, plus fort que celui du temps d'écoute, car on observerait deux tendances qui sont opposées. Pour les autres maladie (à part peut-être l'anxiété), nous ne devrions pas observer de tendance.

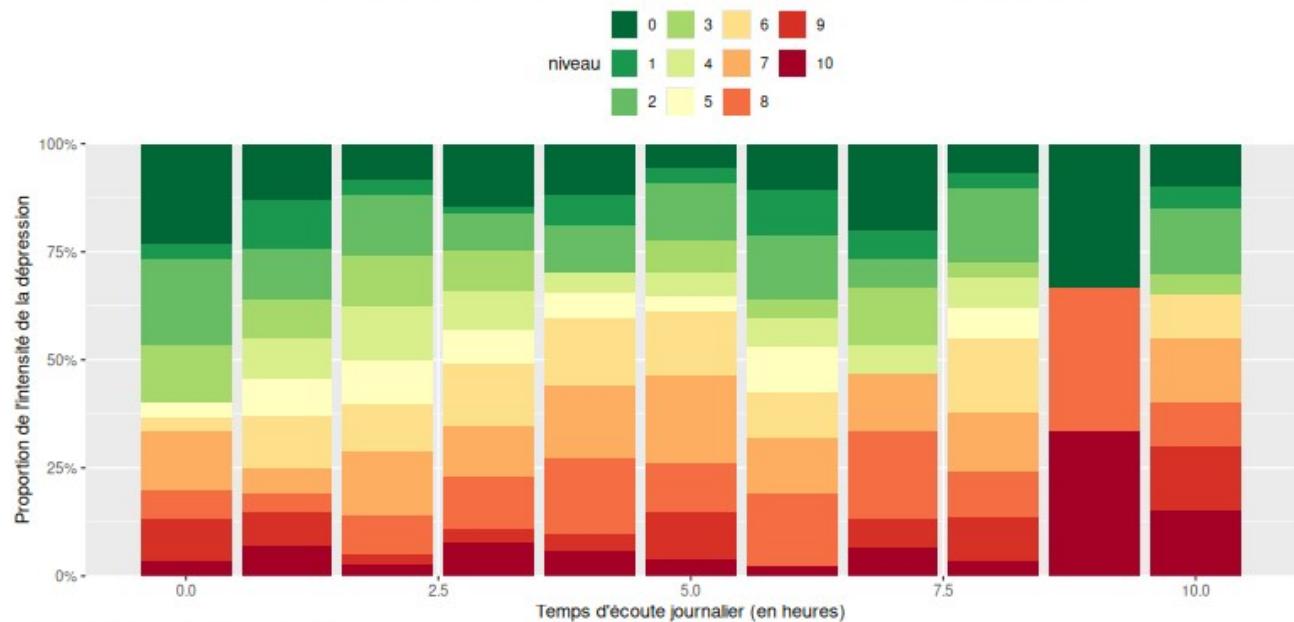
Etude de la corrélation entre le niveau d'un trouble mental et le temps d'écoute



On peut ici aussi créer un graphe interactif pour se focaliser sur les informations qui nous intéressent. On pourrait par exemple ne pas vouloir se limiter à 10h d'écoute au maximum, ou vouloir moins de couleurs différentes en arrondissant plus les valeurs des troubles.

Relation entre le temps d'écoute quotidienne de musique et certains problèmes médicaux

Etude de la corrélation entre le niveau de dépression et le temps d'écoute



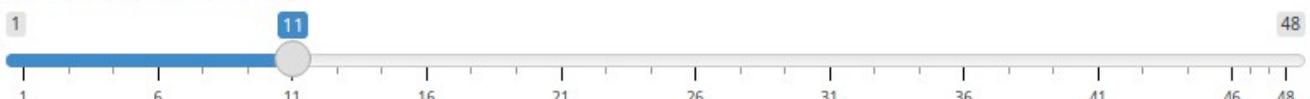
Sélectionner un problème médical

Dépression

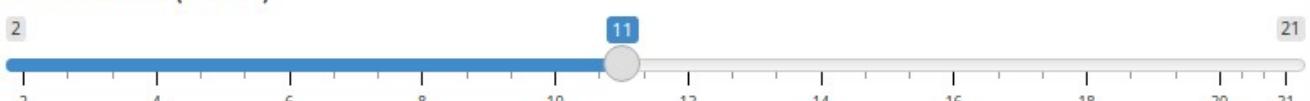
Plage du temps d'écoute



Nombre de bins (horizontal)



Nombre de bins (vertical)



Conclusion

On remarque déjà que la colonne de 9h ne possède que 3 personnes, nous n'en tiendrons donc pas compte. Ensuite, comme attendu, l'anxiété, l'insomnie et l'OCD n'ont pas l'air de présenter de tendance particulière. Pour ce qui est de la dépression, c'est assez complexe de conclure avec notre échantillon. Il semblerait que la corrélation soit positive, ce qui serait donc opposée au graphique précédent, mettant en évidence une légère corrélation négatif de la dépression avec l'effet de la musique, qui serait donc plus fort que celui du temps d'écoute. Mais s'il n'y a pas de tendance, alors on aurait une incohérence. Ainsi, au moins l'une des deux corrélations trouvées précédemment seraient fausses. Or, celle entre le temps d'écoute et l'effet de la musique semble assez solide, contrairement à celle entre la dépression et l'effet de la musique. On en conclurait donc que seule la première est possible.

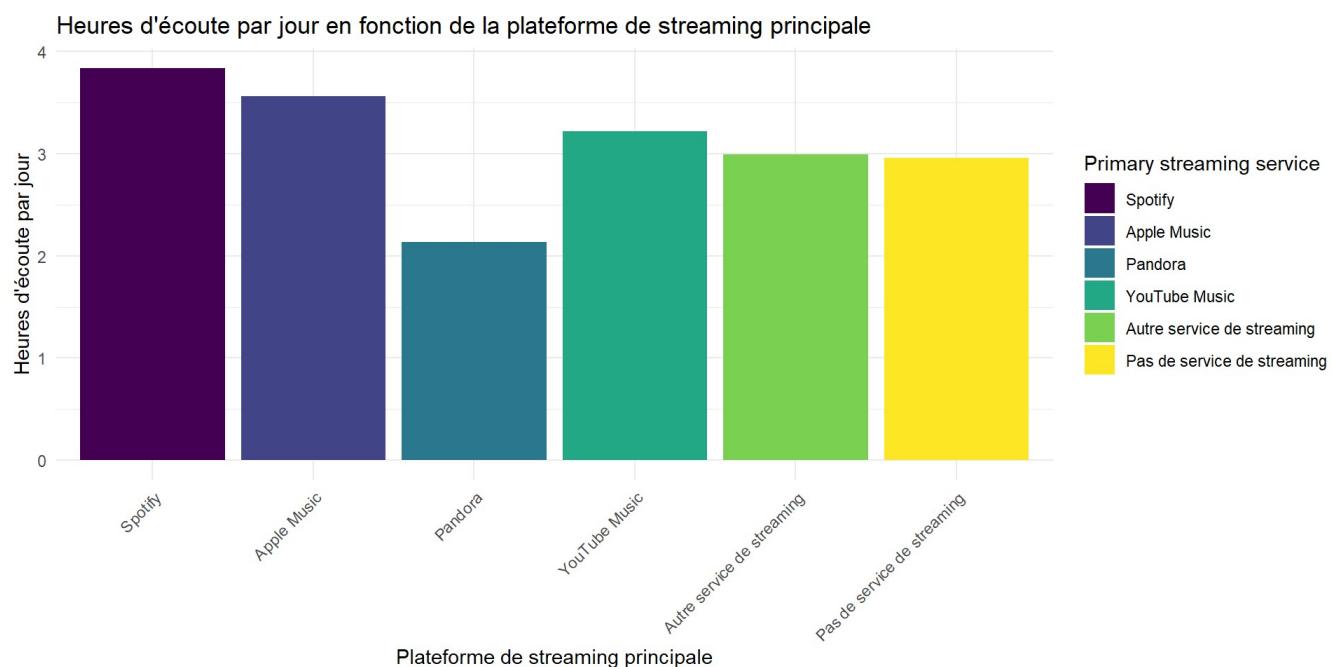
9) Y a-t-il des différences notables dans les scores de troubles

mentaux ou les heures d'écoute par jour en fonction de la plateforme de streaming principale utilisée ?

Analyse

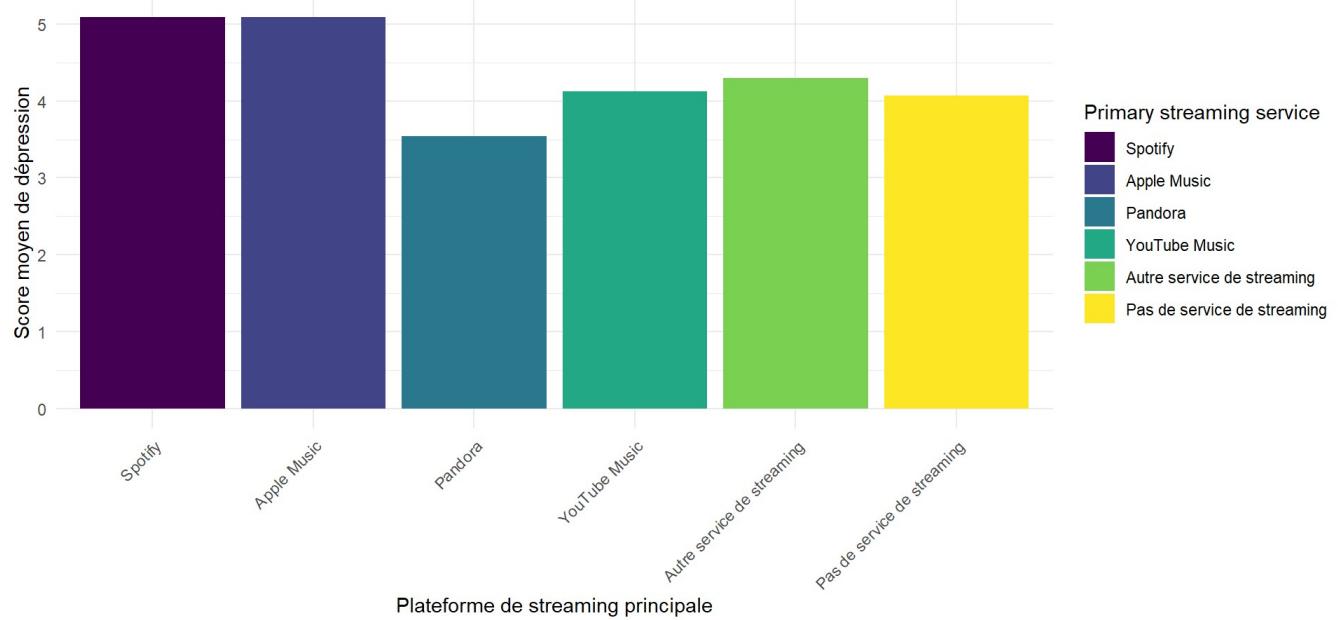
On peut maintenant se demander si les scores de troubles mentaux ou les heures d'écoute par jour varient en fonction de la plateforme de streaming principale utilisée. Pour répondre à cette question, nous allons d'abord examiner les différences dans les heures d'écoute par jour en fonction de la plateforme de streaming principale, puis nous examinerons les différences dans les scores de troubles mentaux en fonction de la plateforme de streaming principale.

On peut émettre l'hypothèse que les personnes n'utilisant pas de service de streaming écoutent moins de musique que les autres, car elles n'ont pas accès à une bibliothèque musicale aussi étendue. On pourrait également penser que la plateforme d'écoute n'a pas d'impact sur les scores de troubles mentaux car il n'y a pas de raison que l'utilisation d'une plateforme de streaming en particulier ait un impact sur la santé mentale.

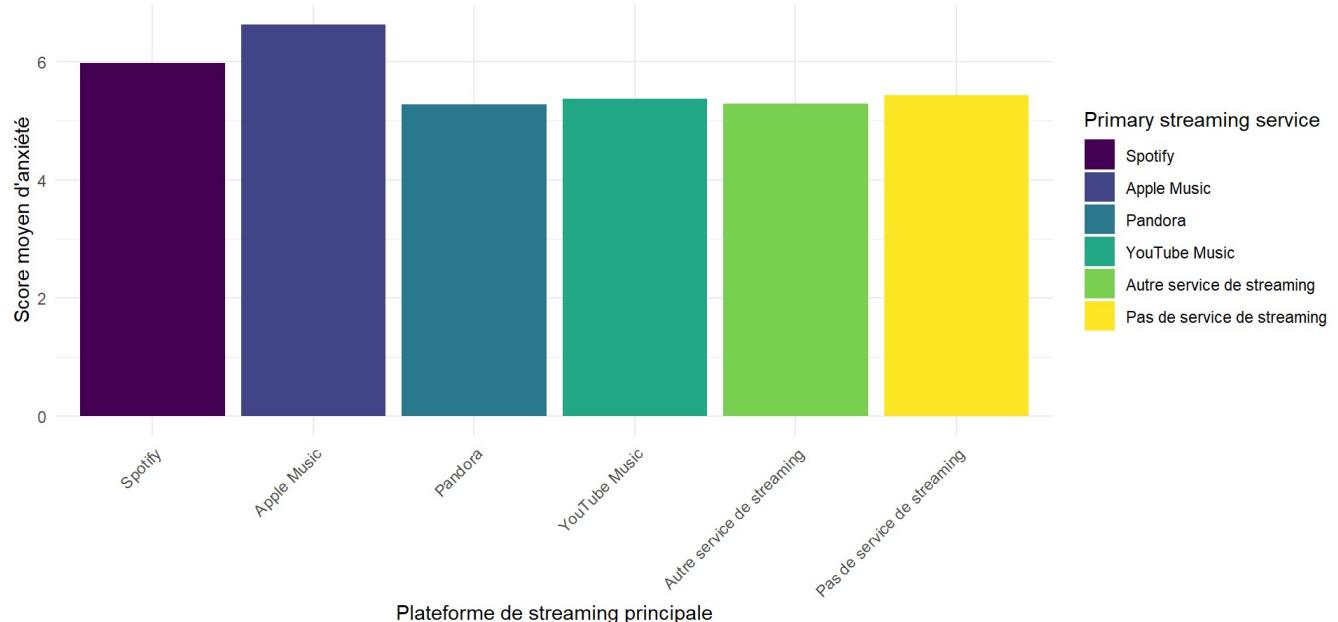


On peut voir que les personnes utilisant Spotify semblent écouter plus de musique par jour que les personnes utilisant les autres plateformes de streaming. Les personnes utilisant Pandora semblent écouter moins de musique par jour que les autres ce qui pourrait s'expliquer par le faible nombre de répondant. Les personnes utilisant Apple Music, YouTube Music et d'autres plateformes de streaming semblent avoir des habitudes d'écoute similaires. Enfin contrairement à ce que l'on aurait pu penser, les personnes n'utilisant pas de service de streaming semblent écouter autant de musique que les autres.

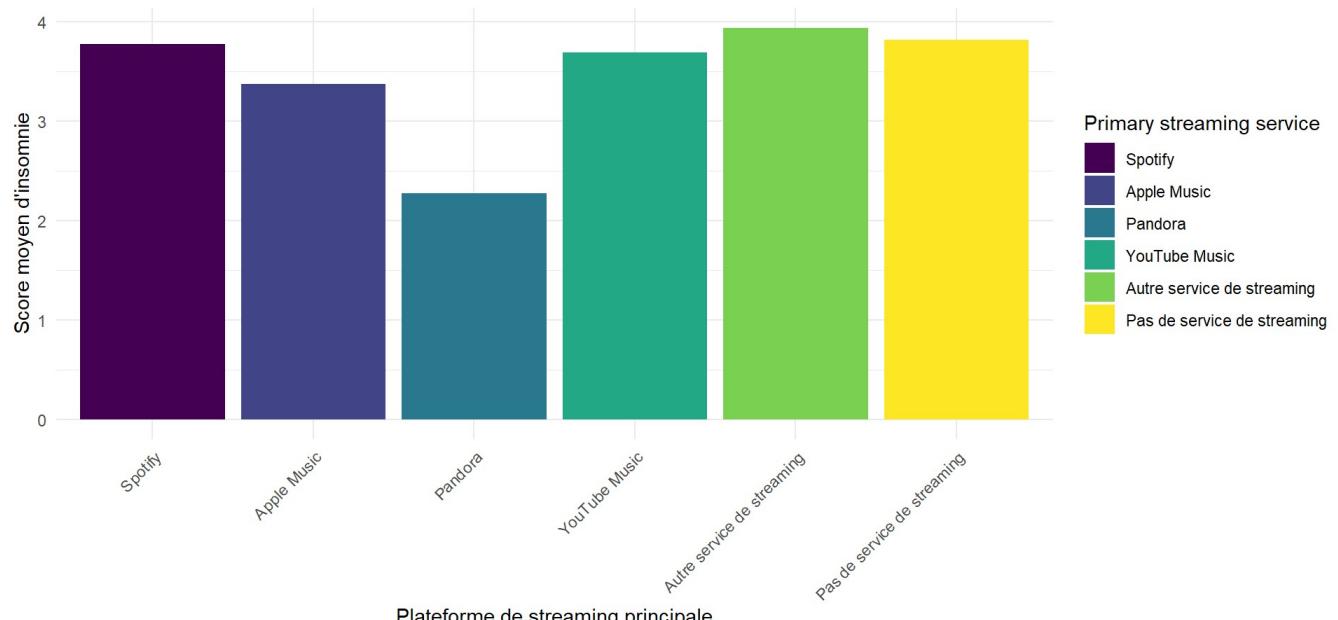
Score moyen de dépression en fonction de la plateforme de streaming principale



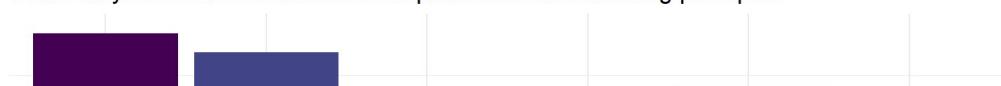
Score moyen d'anxiété en fonction de la plateforme de streaming principale



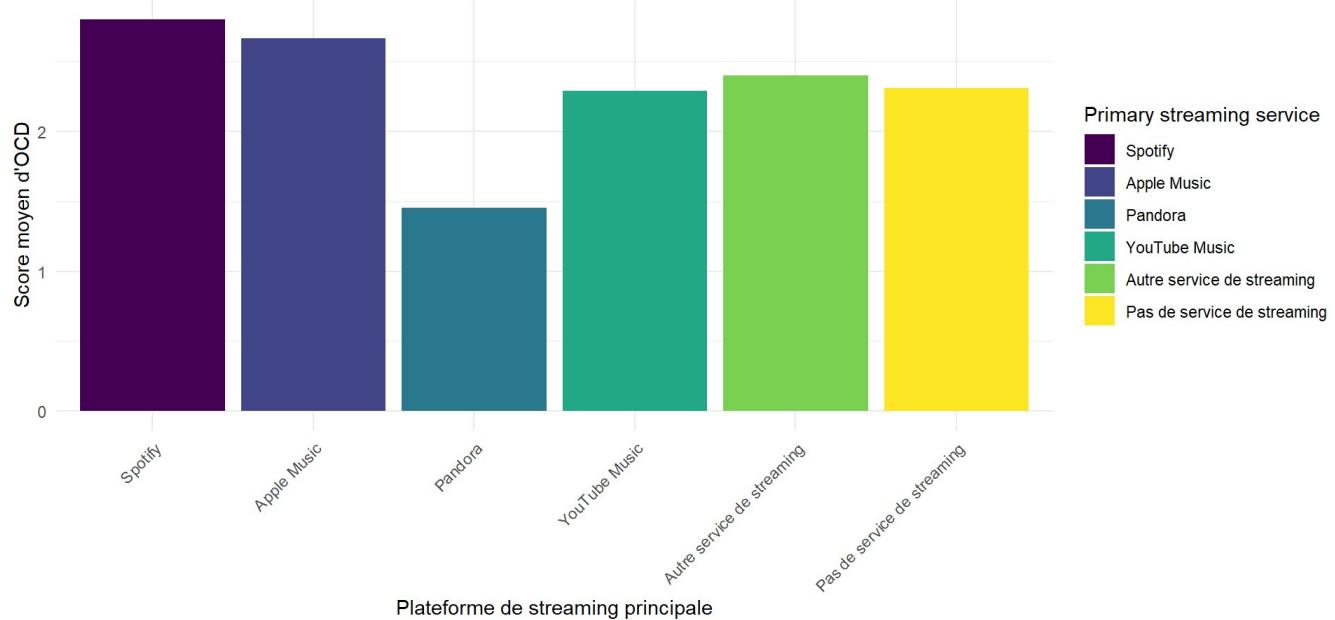
Score moyen d'insomnie en fonction de la plateforme de streaming principale



Score moyen d'OCD en fonction de la plateforme de streaming principale



Score moyen d'OCD en fonction de la plateforme de streaming principale



Si l'on regarde maintenant les scores de troubles mentaux en fonction de la plateforme de streaming principale, on peut voir que les scores de dépression, d'anxiété, d'insomnie et d'OCD sont relativement similaires pour les différentes plateformes de streaming. Cela suggère qu'il n'y a pas de différences notables dans les scores de troubles mentaux en fonction de la plateforme de streaming utilisée.

On peut toutefois noter que les scores d'insomnie et d'OCD semblent plus faibles pour les personnes utilisant Pandora ce qui peut s'expliquer par le faible nombre de personnes utilisant cette plateforme de streaming dans notre échantillon.

Conclusion

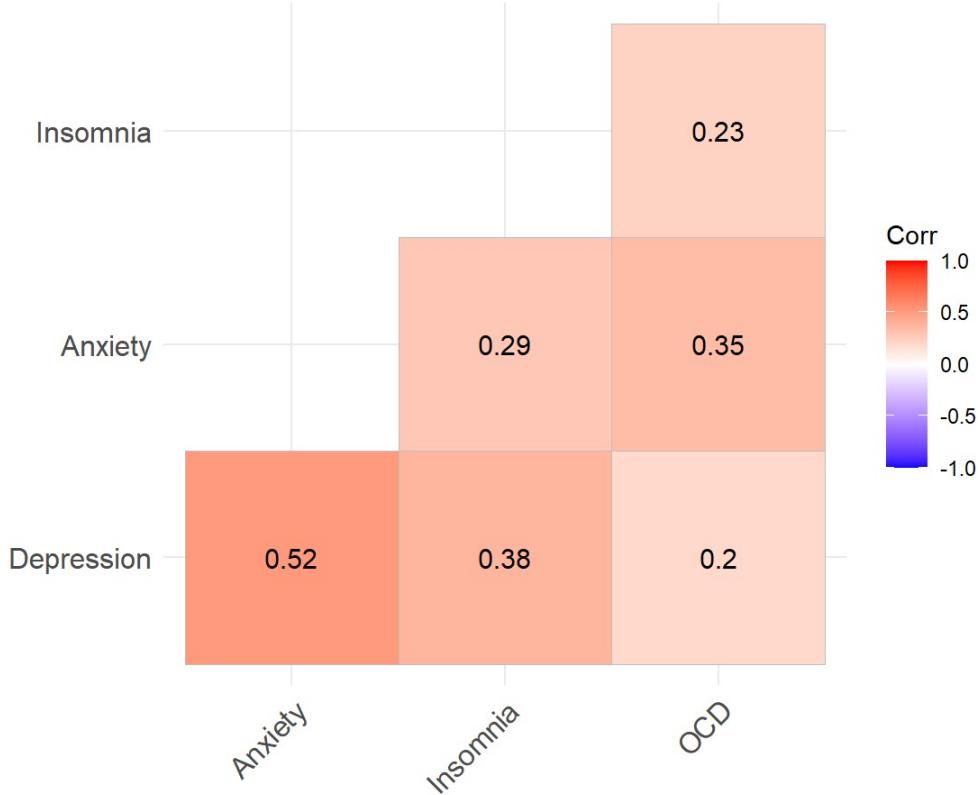
En conclusion, il semble que les personnes utilisant Spotify écoutent plus de musique par jour que les personnes utilisant les autres plateformes de streaming. Les scores de troubles mentaux ne semblent pas varier en fonction de la plateforme de streaming principale utilisée.

10) Comment les scores d'anxiété, de dépression, d'insomnie et d'OCD sont-ils corrélés entre eux ?

Analyse

Puisque les troubles mentaux sont souvent liés les uns aux autres, il est intéressant de voir comment les scores d'anxiété, de dépression, d'insomnie et d'OCD sont corrélés entre eux. Pour cela, nous allons calculer la corrélation entre ces scores et les visualiser à l'aide d'un graphique.

Corrélation entre les scores de troubles mentaux



En regardant le graphique ci-dessus, on peut remarquer qu'il n'y a pas de corrélation négative entre les scores de troubles mentaux, ce qui est cohérent car les troubles mentaux sont souvent liés les uns aux autres.

On peut également voir que les scores de dépression, d'anxiété, d'insomnie et d'OCD sont pas forcément corrélés entre eux. En effet, les scores de dépression et d'anxiété sont plutôt corrélés positivement, ce qui est cohérent car ces deux troubles mentaux sont souvent liés. On remarque que les autres scores sont moins corrélés entre eux. On peut tout de même noter une corrélation positive entre les scores de dépression et d'insomnie et une autre entre les scores d'anxiété et d'OCD.

Conclusion

Pour conclure, il semble que la dépression et l'anxiété ont la corrélation la plus élevée parmi les troubles mentaux étudiés ce qui indique selon nos résultats qu'une personne souffrant de dépression aurait plus de chances de souffrir d'anxiété et vice versa. Cependant, il est important de noter que la corrélation n'implique pas de causalité et que d'autres facteurs peuvent influencer les scores de troubles mentaux. Il serait donc intéressant d'intéresser à des études plus approfondies pour mieux comprendre les liens entre ces troubles mentaux.

11) Les différentes tranches d'âge écoutent-elles plus ou moins de musique en langue étrangère ?

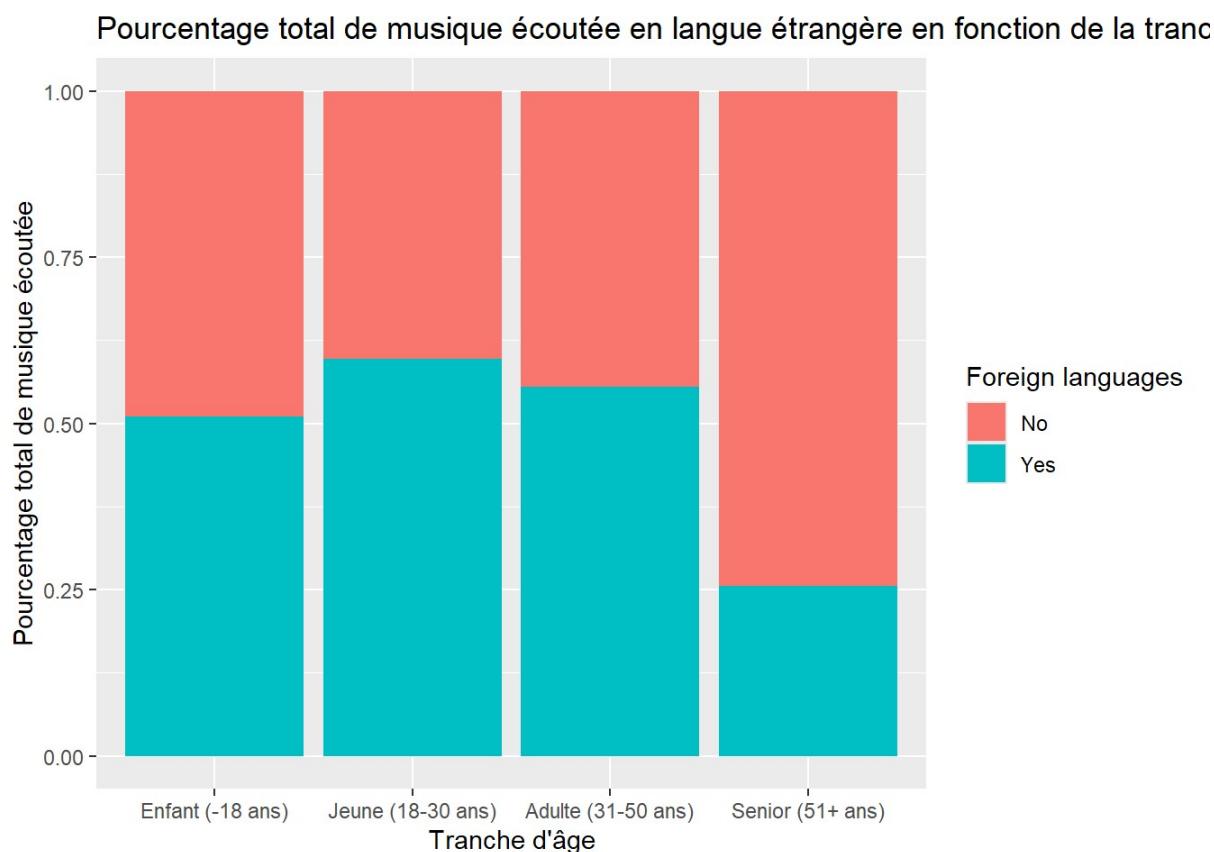
Analyse

Il est évident que la culture musicale varie d'une personne à une autre notamment en fonction de sa génération. Posons nous maintenant la question suivante : est-ce que l'âge influence le choix de la langue de la musique écoutée ?

Pour y répondre, nous allons, à cause de la faible quantité de données, regrouper les participants et reprendre les

tranches d'âge précédemment définies et les comparer à la langue de la musique écoutée.

On peut émettre l'hypothèse que les personnes plus jeunes écoutent plus de musique en langue étrangère que les personnes plus âgées. Pour vérifier cette hypothèse, nous allons comparer les heures d'écoute de musique en langue étrangère en fonction de la tranche d'âge.



On remarque sur le graphique ci-dessus que les jeunes entre 18 et 30 ans écoutent plus de musique en langue étrangère que les autres tranches d'âge. On observe également que les senior de plus de 51 ans écoutent moins de musique en langue étrangère que les autres tranches d'âge. Cependant, ce qui peut paraître surprenant, c'est que les enfants de moins de 18 ans écoutent moins de musique en langue étrangère que les adultes et les jeunes.

Conclusion

Pour conclure, on peut dire que les jeunes écoutent plus de musique en langue étrangère que les autres tranches d'âge. Cela peut être dû à l'ouverture d'esprit des jeunes et à leur curiosité pour d'autres cultures. Les seniors, eux, écoutent moins de musique en langue étrangère que les autres tranches d'âge, ce qui peut être dû à leur attachement à leur culture et à leur langue maternelle. Les enfants écoutent moins de musique en langue étrangère que les adultes et les jeunes, ce qui peut être dû à l'influence familiale et à la musique écoutée par leurs parents.

Cependant, il faudrait approfondir cette analyse en obtenant plus d'échantillons et en étudiant des tranches d'âge plus précises pour confirmer ces résultats.

Conclusion générale

Le jeu de données ne nous a finalement pas appris beaucoup de choses sur les corrélations entre le genre de musique

écouté et les différentes maladies ou problèmes mentaux que l'on a pu étudier. Il semble y en avoir assez peu. Sinon, le jeu de données nous a permis de confirmer certaines choses comme par exemple quelqu'un qui écoute du rap a aussi tendance à écouter du métal, et vice-versa, mais a aussi infirmé certaines choses que nous pensions, notamment que l'écoute de musique réduisait drastiquement avec l'âge.

Cependant, il serait nécessaire d'approfondir notre sujet pour réellement tirer des conclusions. D'une part, il faudrait essayer de récolter plusieurs milliers de réponses avec une meilleure représentation sociale (différentes cultures, âges, etc...). D'autre part, les réponses portant sur un sujet très sensible et étant auto-déclarées, il serait bien de réaliser une étude plus rigoureuse et scientifique avec des protocoles stricts permettant l'obtention de résultats plus fiables.

Concernant les problèmes qu'on a pu rencontrer, il y a tout d'abord les difficultés à concevoir la structure de données nécessaire au graphique que l'on souhaite réaliser. Il y a notamment toute la question de la corrélation entre le style favori d'une personne et l'intensité du trouble mental, qui, faute d'avoir pu trouver la bonne structure de données, s'est retrouvé à créer un graphique "manuellement" sur plus de 100 lignes. Également, nous avons eu beaucoup de difficultés pour gérer le grand nombre de styles de musique différents, qui rendait souvent l'interprétation des graphiques assez complexe à effectuer.

Pour terminer voici les retours de chaque membre du groupe :

Pierre-Emeric : Le projet m'a permis de bien comprendre le principe de R et également d'élargir mes capacités à choisir une visualisation adaptée à ce que l'on veut étudier. J'aurais cependant aimé avoir davantage de données, que ce soit de nouvelles colonnes, et surtout davantage de lignes.

Noé : Grâce à ce projet, j'ai pu comprendre comment construire des graphiques. Pour compléter ce projet, j'aurais bien aimé faire plus de graphiques sur Tableau ou sur Power BI car j'ai vraiment apprécié utiliser Tableau pour faire quelques graphes lors de la présentation orale.

Teddy : Le projet m'a permis de mieux comprendre le fonctionnement de R, et m'a permis de me rendre compte que faire une visualisation n'est pas toujours simple : manque de données, graphes illisibles, moins pertinents que prévu, ... J'aurais aimé aborder le même projet, mais en Python, car je me dirige dans l'intelligence artificielle, où ce langage est très utilisé.

Théo : Réaliser ce projet m'a permis de développer mes compétences en dataviz, et ainsi de réaliser des analyses relativement poussées sur un dataset. J'ai pu apprendre l'utilisation du langage R. Cependant, j'aurais aimé faire plus de Shiny car j'ai trouvé la techno très intéressante et le TD réalisé dessus m'a beaucoup intéressé.

Annexe

Répartition des tâches

Afin de réaliser notre projet, nous avions à l'origine comme idée que Pierre-Emeric s'intéresse aux liens entre les styles de musiques et les troubles mentaux, contrairement à Teddy qui lui devait regarder les liens entre les styles et les autres informations que nous avions. Noé quant à lui s'intéressait au temps d'écoute en fonction des activités de créations musicales. Théo finalement s'intéressait lui au jeu de données en lui-même. Nous nous sommes donc séparé le travail ainsi :

- Théo : Questions 1 et 6
- Teddy : Questions 3, 7 et Shiny Dashboard
- Pierre-Emeric : Questions 4, 5, 6 et 8
- Noé : Questions 2, 9, 10, 11 et Tableau