

Spotify®

INTRODUCTION  
AUX  
STATISTIQUES

# RAPPORT D'ETUDE STATISTIQUE SUR SPOTIFY

MATHILDE GARDE, SARAH FARGES, VICTOR GELINEAU, PAUL ESCALIER

# SOMMAIRE

- 1.Introduction.....1
- 2.Description des jeux de données.....2
- 3.Application des tests de cohérence.....3
- 4.Problématique et Hypothèses.....4
- 5.Etude statistique.....6
- 6.Conclusion.....15
- 7.Bibliographie.....16

# INTRODUCTION

Spotify est une plateforme de streaming de musique en ligne qui permet aux utilisateurs d'accéder à des millions de chansons, de podcasts et d'autres contenus audio.

Elle a été créée en Suède en 2006 et est devenue très vite l'une des plus grandes plateformes de streaming musical au monde, avec plus de 365 millions d'utilisateurs actifs mensuels dans plus de 170 pays.

L'application dispose d'une version gratuite avec des publicités, ainsi que d'une version payante sans publicités et avec des fonctionnalités supplémentaires telles que la possibilité de télécharger de la musique pour une écoute hors ligne, une qualité audio supérieure et la possibilité de passer des chansons de façon illimitée.

La plateforme utilise également des algorithmes de recommandation avancés pour suggérer de la musique et des podcasts à ses utilisateurs en fonction de leurs préférences d'écoute.

En outre, elle offre des fonctionnalités sociales qui permettent aux utilisateurs de suivre les playlists et les artistes préférés de leurs amis, partager de la musique sur les réseaux sociaux et de découvrir de nouveaux artistes et genres musicaux en explorant les playlists et podcasts populaires.



Spotify rassemble des chansons pour créer des playlists suivies par des millions de personnes. Une grande majorité de personnes écoutent les musiques en vogue, notamment grâce aux playlists qui donnent les tops hits actuels. Mais alors, comment ces playlists sont constituées? Qu'est-ce qu'un hit et qu'est-ce qui fait que certaines chansons soient des hits et d'autres non? Ayant nous-mêmes Deezer ou Spotify sur nos appareils, nous étions curieux de savoir si ce que nous trouvions de commun parmi les hits du moment pouvaient être démontré scientifiquement. Grâce à R studio et à nos connaissances de cours, nous avons été à même d'appliquer différentes méthodes et d'apporter des preuves solides et concrètes à nos propos.

**365M**  
d'utilisateurs



**dans + de 170**  
pays



**des millions**  
de chansons,  
podcasts...



# DESCRIPTION DES JEUX DE DONNEES



Nos données proviennent de Kaggle, une plateforme internet qui accueille la plus grande communauté de Data Science au monde, avec plus de 536 000 membres actifs dans 194 pays.

## Jeu de données 1:

Dans notre premier jeu de données, nous retrouvons un total de 113 999 titres de musiques et 20 colonnes, chacune associée à une caractéristique. On considère donc un échantillon de 113 999 musiques (effectif).

Nous avons ainsi les informations suivantes pour chaque titre:

- l'**identifiant** Spotify du titre
- l'**artiste**
- le **nom de l'album** dans lequel la chanson apparaît
- le **titre** de la chanson
- un **indice de popularité**: entre 0 et 100, 100 étant le plus populaire. Cette popularité est calculée par un algorithme et basée en grande partie, sur le nombre total d'écoutes de la chanson et sur le caractère récent de ces écoutes. En règle générale, les chansons qui sont beaucoup jouées actuellement auront une popularité plus élevée que les chansons qui ont été beaucoup jouées dans le passé. Les titres en double (par exemple, le même titre d'un single et d'un album) sont évalués indépendamment. La popularité de l'artiste et de l'album est dérivée mathématiquement de la popularité de la piste.
- la **durée** (en millisecondes)
- l'**explicite**: les paroles ou les contenus d'une chanson ou d'un clip vidéo contiennent un ou plusieurs critères qui pourraient être considérés comme offensants ou inappropriés pour les enfants.
- **degré dansant**
- un **indice d'énergie**: l'énergie est une mesure comprise entre 0,0 et 1,0 et représente une mesure perceptive de l'intensité et de l'activité.
- la **clé**: la tonalité de la piste. Les nombres entiers correspondent aux hauteurs en utilisant la notation standard de la classe de hauteur. Par exemple, 0 = C, 1 = C # / D b , 2 = D, et ainsi de suite. Si aucune clé n'a été détectée, la valeur est -1.
- la **puissance sonore**: qualité du son
- le **mode**: majeur ou mineur
- un **degré de paroles**: détecte la présence de mots prononcés dans la chanson
- un **degré acoustique**: Une mesure de confiance de 0,0 à 1,0 pour savoir si la piste est acoustique.
- un **taux d'instrumentalité**: prédit si la piste ne contient pas de voix
- la **vivacité**: détecte la présence d'un public dans l'enregistrement
- **valence**: Une mesure de 0,0 à 1,0 décrivant la positivité musicale transmise par une piste.
- un **tempo**
- une **signature rythmique**
- un **genre musicale**



## Jeu de données 2:

Dans notre second jeu de données complémentaire, nous avons un total de 2 000 hits et 18 colonnes. Ce qui change notamment par rapport au jeu de données 1 est la date de sortie des chansons qui est une nouvelle variable.

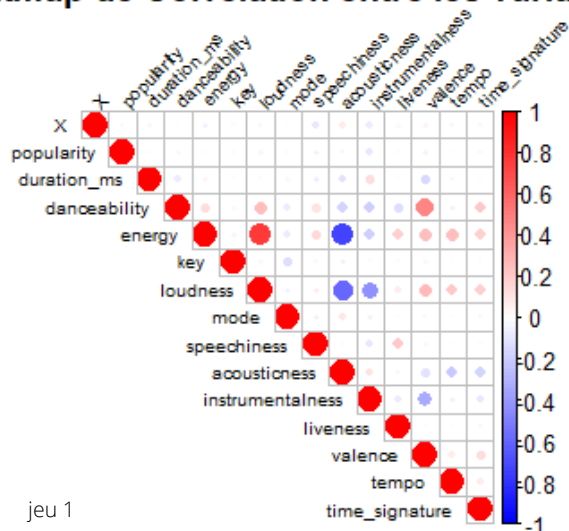
Voici un descriptif des intitulés des colonnes:

- l'**artiste**
- le **titre** de la chanson
- la **durée** en millisecondes
- l'**explicite**
- l'**année** de sortie
- la **popularité** (nous pouvons retrouver des indices de 0 même si toutes les chansons sont des hits puisque certaines ne sont plus écoutées aujourd'hui)
- **degré dansant**
- l'**énergie**
- l'**intensité sonore**
- le **mode**
- le **degré de paroles**
- le **degré acoustique**
- le **degré d'instrumentalité**
- la **vivacité**
- la **positivité**
- le **tempo**
- le **genre**

# APPLICATION DES TESTS DE COHERENCE

## Jeu 1:

### Heatmap de Corrélation entre les variables



jeu 1

A l'aide de R studio, nous avons pu représenter le graphique de corrélation entre les différentes variables de notre jeu de donnée numéro 1. Cela nous permet d'établir et de visualiser les liens ainsi que leur degré d'intensité entre les variables.

Par exemple, on peut voir que la corrélation entre l'énergie et l'intensité sonore s'approche de 1, ce qui indique que ces deux paramètres sont fortement liés. De la même manière, on voit que le degré dansant est relié à la valence, qui caractérise la positivité d'une chanson, avec un coefficient de 0,6. A contrario, les coefficients s'approchant de -1 indiquent que lorsqu'une variable est importante, l'autre variable est petite. L'intensité sonore et l'énergie partagent des coefficients très bas avec l'acoustique. Ceci est logique puisque plus la présence acoustique est importante, plus la chanson est "calme", donc manque d'énergie. Enfin, les valeurs qui s'approchent de zéro indiquent que les paramètres n'ont pas de lien entre eux.

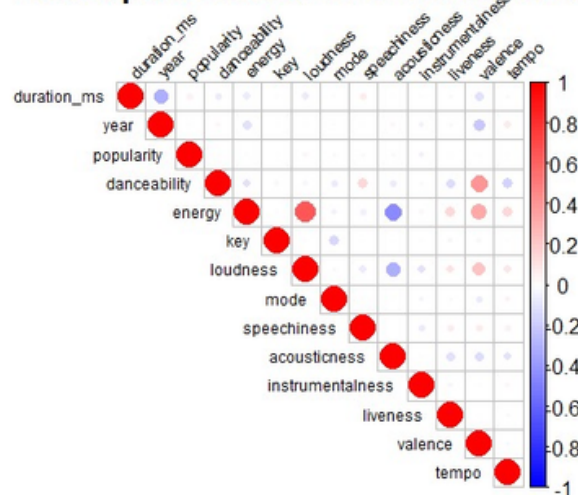
L'utilisation de la colonne "artist\_name" (nom de l'artiste) n'est pas pertinente ici car la répartition n'est pas du tout homogène et le poids de chaque artiste est disproportionné dans le jeu de données. C'est pourquoi dans la suite de l'étude, nous ne prendrons pas en compte cette donnée qui pourrait fausser nos résultats.

## Jeu 2:

Nous avons également appliqué des tests de corrélation pour notre deuxième jeu de données. Nous observons des valeurs similaires de corrélation entre les variables.

De plus, dans le jeu 2, seules 42 chansons sont des titres datant de 1998, 1999 ou 2020. L'objectif de l'ensemble de données étant d'analyser les données de la période 2000-2019, on peut donc se permettre de négliger les titres relatifs à ces trois années et les supprimer de notre base pour les études à suivre.

### Heatmap de Corrélation entre les variables



Appliquer des tests de corrélation nous permet donc dès le départ de nous orienter sur le choix des caractéristiques d'étude, les liens à mettre en avant, et les hypothèses que nous allons pouvoir faire par rapport à notre problématique. C'est un outil très utile pour répertorier les données exploitables et celles qui ne le sont pas.



# Problématique et hypothèses



## Problématique

On parle de plusieurs milliards nombre d'écoutes pour les chansons les plus populaires sur Spotify. Chaque chanson du top 10 a été écoutée entre 2,4 et 3,5 milliards de fois. De tels chiffres nous amènent donc à nous demander comment un tel phénomène est possible.

Par ailleurs, on constate très rapidement que toutes les chansons sont en anglais et que les années de publication sont toutes comprises entre 2016 et 2021. Il y aurait-il un lien quelconque avec la popularité de ces titres? Si oui, comment peut-on déterminer si une chanson est populaire?

Le but de nos recherches est de répondre à la question suivante: **Les chansons à succès partagent-elles des caractéristiques similaires et pouvons-nous utiliser ces similarités pour prédire quelles chansons deviendront des hits ?**

Pour ce faire, nous allons d'abord définir le terme "populaire" et "hits", puis nous évoquerons les hypothèses que nous avons faites en amont de notre étude. Dans un second temps, nous aborderons notre étude et montrerons si l'exploitation des données permet de corroborer nos hypothèses.

Enfin, nous finirons par résumer ce qui a pu être observé et présenter nos éventuelles difficultés.

## Qu'est-ce qu'un hit? :

Nous essayons ici d'estimer si une chanson sera un "hit" ou non et les tendances musicales communes à ces hits. Pour cela, nous prenons le parti de considérer que si une chanson obtient un score de popularité supérieur ou égal à 70/100, alors c'est un hit.

Dans notre jeu de données 1, nous répertorions donc 5 472 hits et dans le jeu de données 2 on en compte 727.

## Hypothèses:

Au préalable de notre étude, nous nous sommes mis d'accord sur des hypothèses concernant les tendances communes aux hits musicaux et les facteurs qui font qu'une chanson est populaire. Ces hypothèses se basent surtout sur le bon sens, la logique et nos connaissances sur la musique. En d'autre terme, on se base sur notre environnement, les modes observées autour de nous, notre vécu personnel...

Ceci nous a mené à la liste non exhaustive suivante:

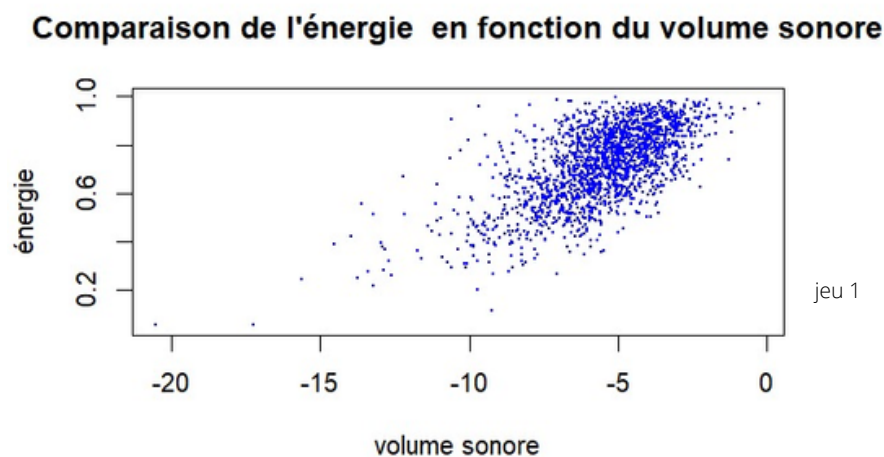
- ce qui fonctionne le plus depuis un certain temps ce sont les musiques avec beaucoup de rythme et une intensité sonore forte
- parmi les genres musicaux les plus écoutés, c'est celui des chansons pop qui est prédominant
- le tempo des musiques est plus rapide qu'avant
- de manière générale, les personnes ont tendance à écouter des musiques en tonalité majeur et plus joyeuses
- on retrouve moins d'acoustique et plus de sons artificiels (comme l'auto tune)
- les paroles comportent plus de contenu explicite
- les chansons qui fonctionnent le mieux sont celles sur lesquelles on peut danser facilement
- il n'y a pas forcément besoin de beaucoup de paroles pour qu'un titre fonctionne bien

## Pourquoi deux jeux de données ? :

Assez rapidement après avoir effectué le test de corrélation sur le jeu 1, nous avons constaté que l'indice de popularité ne présentait pas de fortes corrélations avec les autres caractéristiques. Or, nous pensions pouvoir nous baser sur cet indice pour établir des liens tangibles. Nous nous sommes donc demandés comment contourner/résoudre ce problème et ce qui faisait que les corrélations avec cet indice étaient si faibles. Une des pistes a été d'étudier les tendances et caractéristiques communes des hits mais aussi les autres variables qui présentaient des liens plus importants. L'autre piste a été de s'aider d'un autre jeu de données. En effet, le jeu de données 1 présente une lacune assez importante puisque l'année de sortie des chansons répertoriées n'est pas indiquée. Il était donc impossible de faire intervenir une temporalité dans notre étude. C'est pourquoi, nous avons trouvé un autre jeu de données, cette fois-ci disposant de la date de sortie des chansons afin de pouvoir prendre en compte exclusivement les hits, la popularité et les évolutions des autres caractéristiques au cours du temps. L'objectif étant de faire ressortir des tendances au fur et à mesure des années.

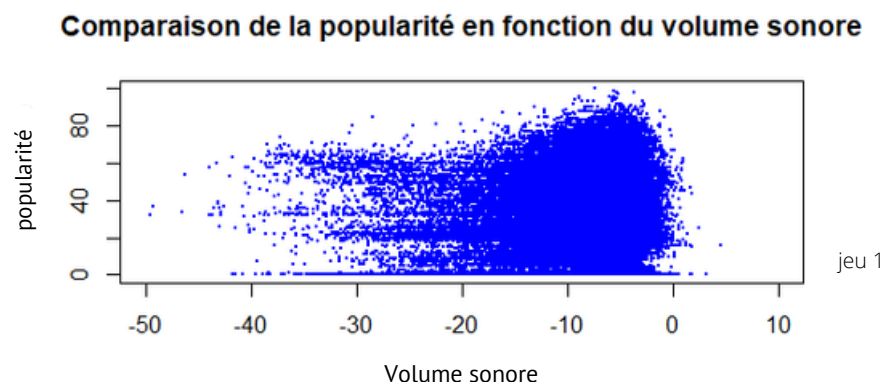
## Hypothèse n°1: Les musiques qui plaisent le plus ont tendance à être plus dansantes, énergiques, rythmées et de forte intensité sonore:

Nous sommes enclins à penser que les musiques avec beaucoup d'énergie ont plus tendance à être populaires. Or, d'après notre tableau de corrélation des variables, l'énergie est directement liée avec le niveau sonore d'une chanson. Nous avons pu tracer le graphique suivant:



Le volume sonore varie dans la grande majorité des chansons entre -10 et 2,5. Plus il s'approche de 0, plus la majorité des points possède une énergie proche de 1. On en conclut donc que plus une chanson possède un son élevé, plus elle transmet un niveau d'énergie élevé.

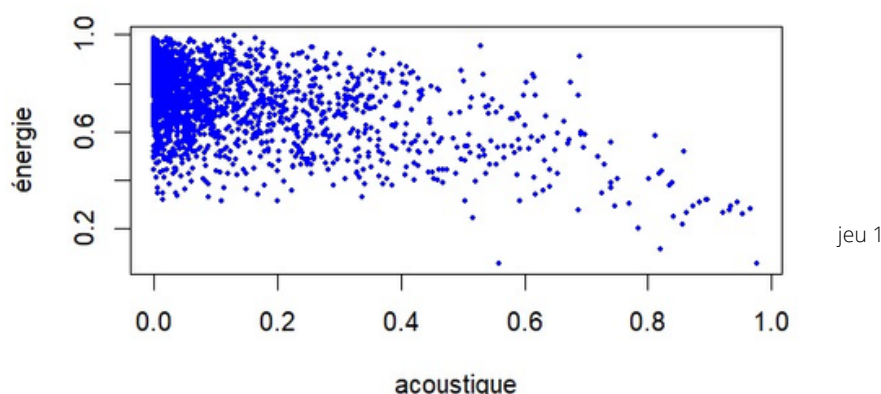
Regardons maintenant si les chansons au volume sonore important ont tendance à être populaires:



Nous pouvons observer que les points qui atteignent une popularité d'au moins 80 ont un volume sonore entre -10 et 0. On peut donc affirmer que les hits ont un volume sonore élevé, et par conséquent une énergie importante. Cette énergie apportée par le volume sonore est particulièrement appréciée car elle permet de relâcher le stress et de transmettre plus d'émotions car on entend plus de détails.

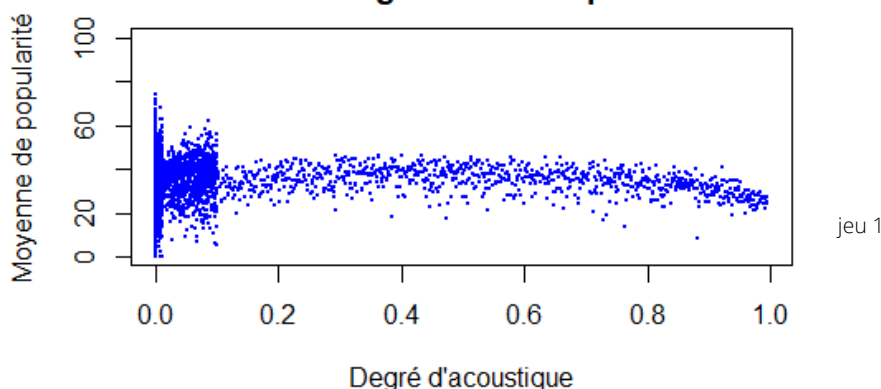
Avec le tableau de corrélation, nous avons observé que l'énergie était aussi liée à l'acoustique :

### Comparaison de la énergie en fonction de l'acoustique



Comme le volume sonore, l'énergie a tendance à être importante lorsque l'indice acoustique n'est pas élevé. Cela est logique puisque les chansons qui comportent des instruments acoustiques sont assez calmes et ne transmettent pas beaucoup d'énergie. Cependant, il faut démontrer que la grande majorité des hits sont des chansons dynamiques avec de l'énergie et non des ballades acoustiques qui ont gagné en popularité ces dernières années pour leur côté intimiste.

### Moyenne de popularité en fonction du degré d'acoustique

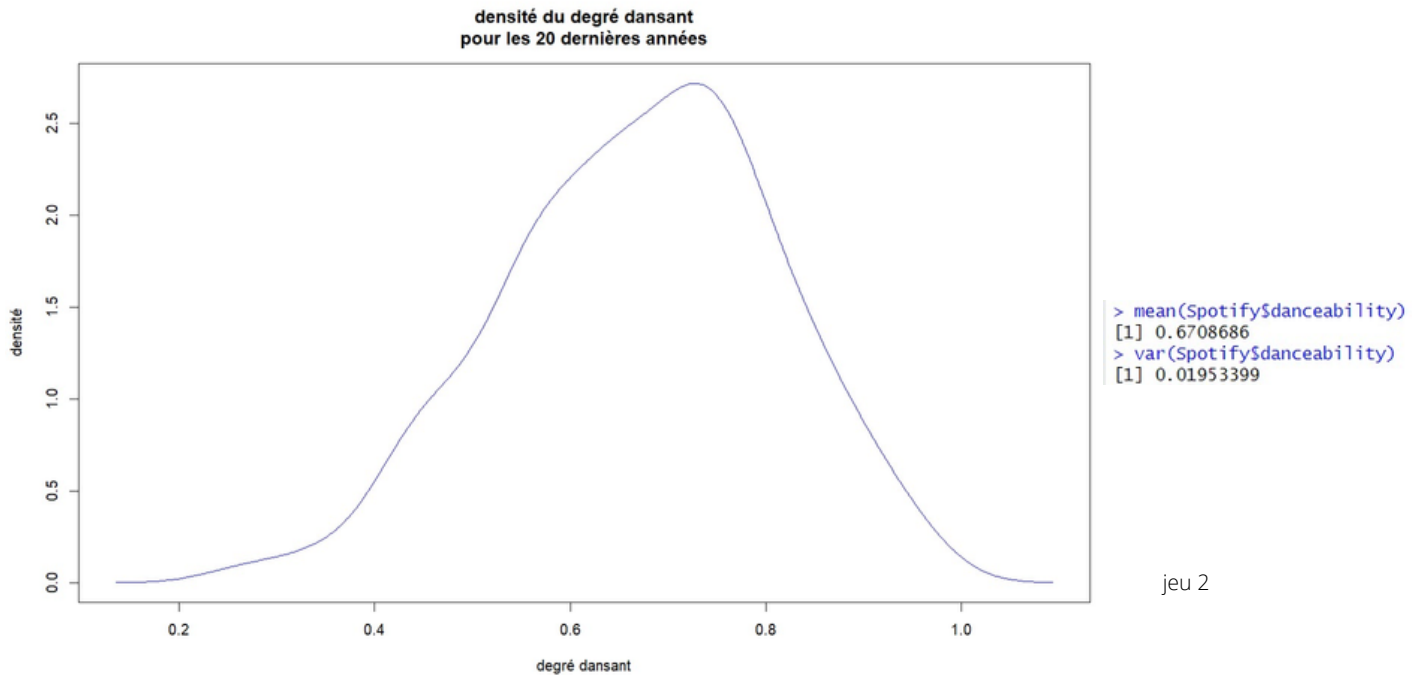


Le graphique nous permet d'observer un regroupement des chansons vers un degré acoustique faible entre 0 et 0,1. On constate d'ailleurs que moins il y a d'acoustique, plus la moyenne de popularité est élevée. Ainsi, on voit que certaines chansons ayant un degré 0 en acoustique ont une moyenne de popularité qui peut monter jusqu'à 80.

Ceci nous permet donc de valider notre hypothèse selon laquelle les musiques les plus populaires sont celles avec le plus d'énergie et le moins d'acoustique. La tendance devient donc d'écouter des titres plus rythmés qu'avant et moins calmes.

En utilisant le deuxième jeu de données avec uniquement les hits ces vingt dernières années, avec le graphique ci-dessous, nous observons que la moyenne de l'indice dansant dans les chansons populaires est d'environ 0,7. Ainsi, la majorité des chansons sont dansantes et donc positives (car le degré dansant et la valence ont un lien de corrélation élevé, avec la valence qui décrit si une chanson émane du positif).





On remarque ici que la courbe s'apparente à celle d'une loi normale, on essaie donc de vérifier notre hypothèse sur la répartition du degré dansant, et il se trouve que si l'on retire les chansons avec une popularité inférieure à 78, le test de Shapiro-Wilk montre ces résultats:

#### Shapiro-wilk normality test

data: Spotify\$danceability  
W = 0.99046, p-value = 0.124

#### Application du cours: Calcul de l'intervalle de confiance à 95 de la moyenne du degré dansant

Selon le test de Shapiro-Wilk, la répartition du degré dansant suit une loi normale, en effet la p-value est supérieure à 0,05 ce qui est satisfaisant pour ce test. De plus, l'effectif étant très grand, on peut effectuer une approximation pour une loi normale. On dispose de la moyenne et de la variance, donc de l'écart type et on peut appliquer un test d'intervalle de confiance à 95% du degré dansant moyen. L'effectif est de 236 titres puisqu'on ne considère ici que les titres avec un indice de popularité  $\geq 78$ .

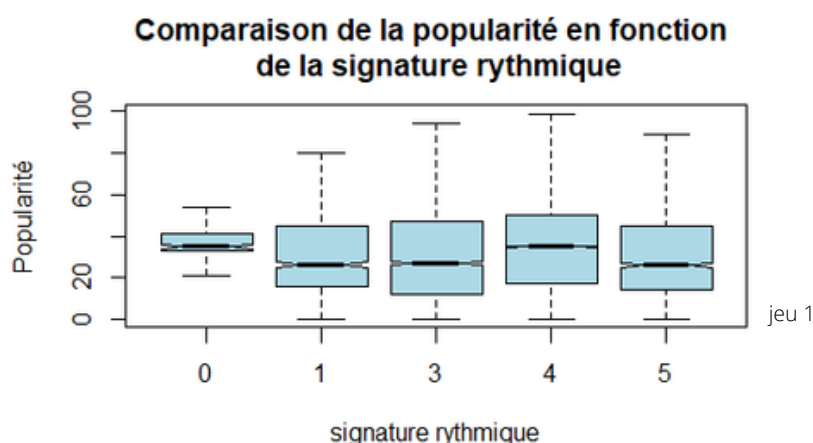
Application de la loi normale à la moyenne de la densité du degré dansant, calcul intervalle de confiance à 95% :

$$IC_{95\%} = \left[ \mu - x \times \frac{\sigma}{\sqrt{n}}, \mu + x \times \frac{\sigma}{\sqrt{n}} \right]$$

$$IC_{95\%} = \left[ 0,67 - 1,96 \times \frac{0,14}{\sqrt{236}}, 0,67 + 1,96 \times \frac{0,14}{\sqrt{236}} \right]$$

$$IC_{95\%} = [0,6521 ; 0,6878]$$

On voit bien que la moyenne est comprise dans cet intervalle de confiance et que le degré dansant de n'importe quel titre populaire aura 95% de chance d'être compris dans l'intervalle de [0,6598 ; 0,6801]. En terme d'interprétation des résultats, cela nous permet de voir que le degré dansant des chansons les plus populaires est assez élevé. En d'autres termes, les chansons qui fonctionnent le mieux ont en commun le fait d'avoir un son entraînant permettant de danser dessus aisément.

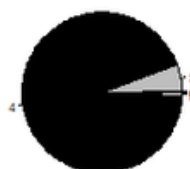


Avant d'étudier ce diagramme, il est important de préciser que la signature rythmique d'un morceau donne une indication sur la mesure et la valeur du temps. Plus la signature rythmique est élevée, plus le tempo (pulsation) sera rapide.

Ce diagramme de boîtes à moustaches est très intéressant pour l'étude du rythme des chansons. Si on l'observe bien, on remarque que la médiane des chansons d'intensité rythmique 4 est plus élevée que la médiane des autres intensités. Autrement dit, les chansons d'intensité 4 ont une popularité plus importante que celles d'autres signatures rythmiques puisque 50% des chansons se trouvent au-dessus et 50% au-dessous des médianes sur de tels diagrammes. En outre, l'étude des quartiles nous permet de voir que les valeurs de popularité des chansons de forte signature rythmique sont beaucoup plus étalées et ont des amplitudes beaucoup plus grandes. Ainsi, pour les chansons de signature rythmique 3 ou 5, l'indice de popularité des chansons monte jusqu'à 90 et 85 et pour les chansons de signature rythmique 4, cela peut même aller jusqu'à 98. D'un autre côté, si on se focalise sur les chansons de signature rythmique 0, on remarque que les valeurs sont beaucoup plus regroupées et l'indice de popularité varie seulement entre 20 et 60.

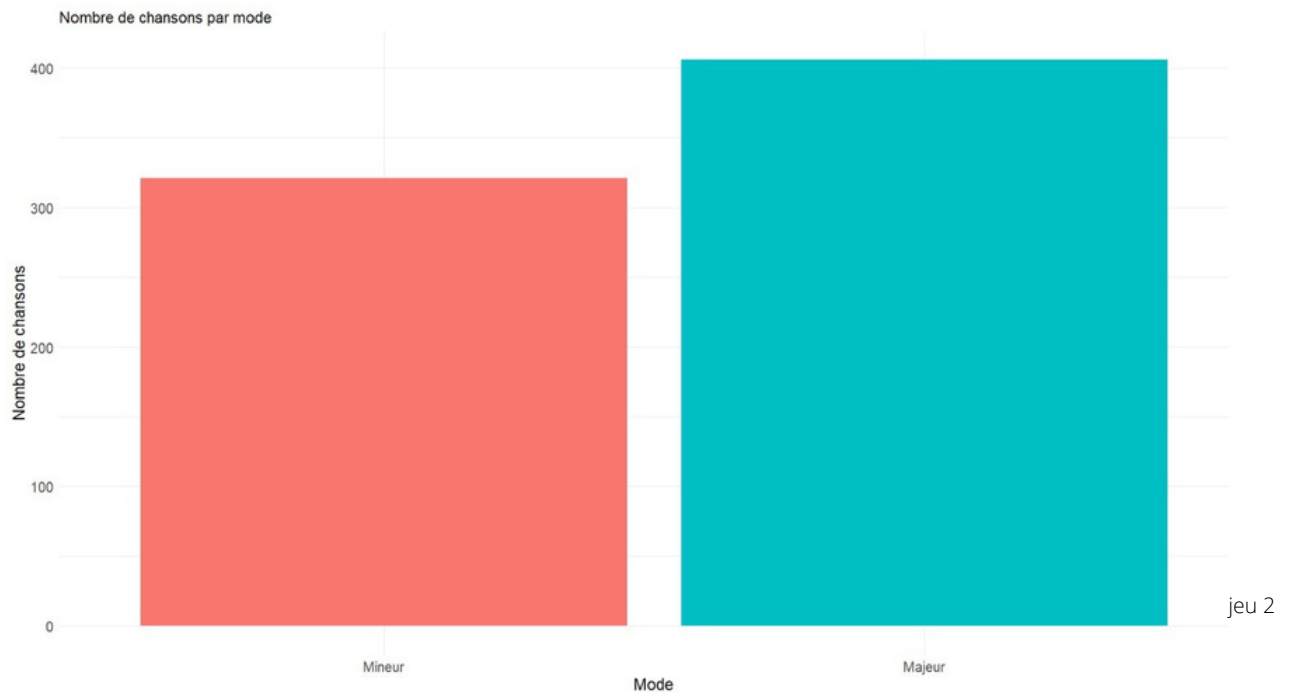
Finalement, cela nous permet de voir que les musiques les plus populaires ont en commun le fait d'être très rythmées, ce qui rejoint l'indice d'énergie et joue sur le caractère dansant et entraînant qu'une musique peut avoir.

### Signatures rythmiques dans les musique à forte popularité ( $\geq 70$ )

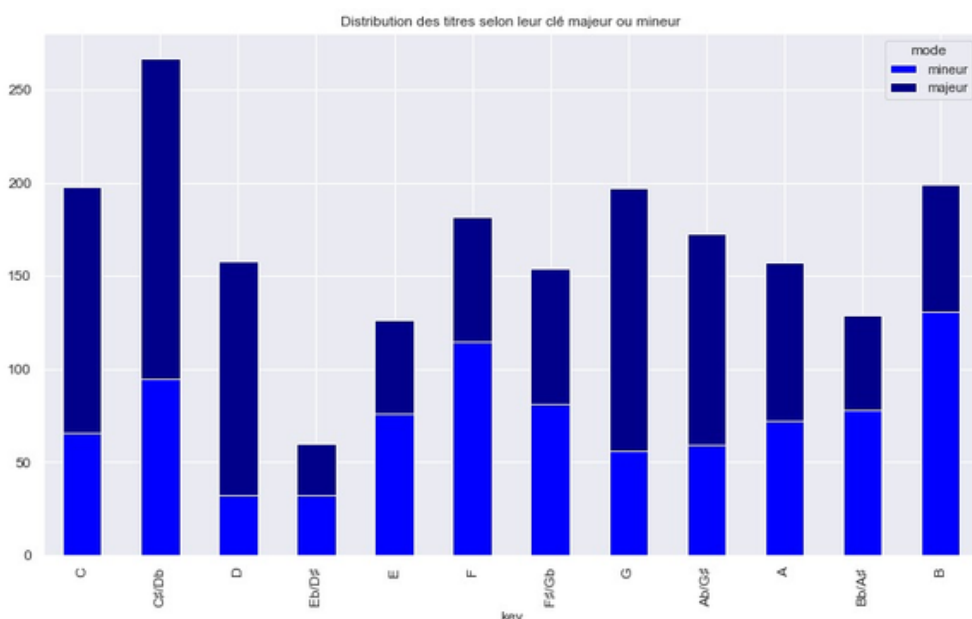


Pour préciser nos propos, nous pouvons nous pencher sur les chansons avec un indice de popularité supérieur ou égal à 70. Il devient alors très clair que ces dernières disposent presque toutes d'une signature rythmique de 4, soit une signature très élevée. Donc, les musiques qui font partie des tendances ont en commun le fait d'être rapides, rythmées et cadencées.

## Hypothèse n°2: Nous sommes portés à écouter des musiques joyeuses le plus souvent de tonalité majeure:



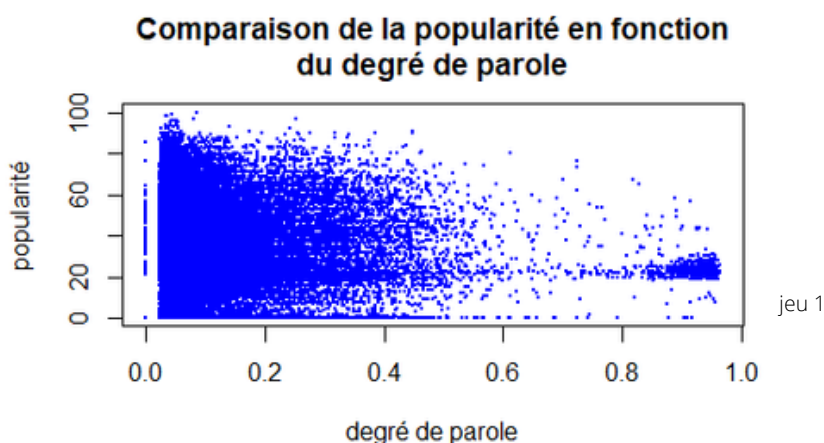
Ici, nous voyons que le nombre de chansons à indice de popularité important étant de tonalité majeure est plus important que celui de tonalité mineure. Ceci s'explique notamment par le fait que les personnes écoutent de la musique pour se faire plaisir et se détendre. Souvent, on aura tendance à écouter des titres plus joyeux que tristes.



Le graphique ci-contre nous permet quant à lui d'étudier les tendances d'apparition de certaines clés. Ainsi, on distingue une clé qui ressort particulièrement du lot, celle de C # / D ♭ pour la majorité des top hits entre la période 2000-2019. De la même manière que précédemment, il s'agit d'une autre représentation qui montre bien encore une fois que la plupart des hits sont à l'échelle majeure.

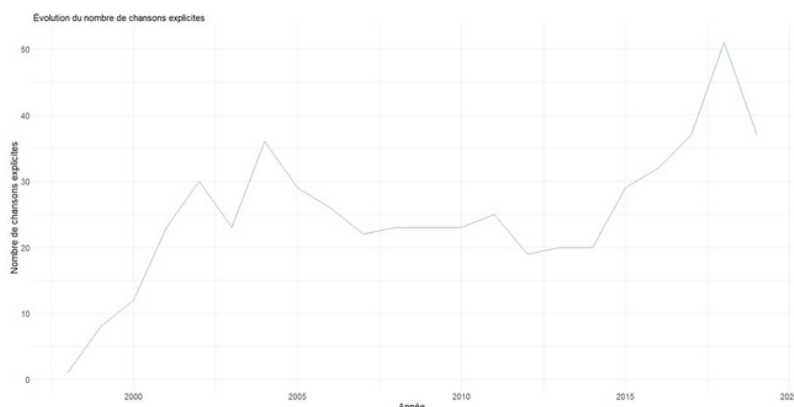
Ceci s'explique par le fait que l'écoute de musique génère des hormones telles que la dopamine (liée au bonheur et plaisir/récompense) et est intrinsèquement liée à notre activité cérébrale et physiologique. La musique est un vrai médicament! Elle peut aider à réduire l'anxiété et la douleur, baisser la tension artérielle, diminuer certains effets secondaires et même aider à la réparation cérébrale chez des grands prématurés.

Hypothèse n°3: Il n'y a pas forcément besoin de beaucoup de paroles pour qu'un titre fonctionne bien et généralement ces dernières sont peu explicites:



Le niveau d'élocution des titres phares n'est pas nécessairement conséquent. Autrement dit, il n'y a pas besoin de beaucoup de paroles pour qu'une chanson plaise et fonctionne bien. On remarque bien d'ailleurs que très peu de chansons ont un degré de parole élevé supérieur à 0,4. Il y a une petite concentration entre 0,8 et 1 mais avec une popularité relativement faible de 20 alors que les titres avec une popularité plus grande ont un degré de parole compris entre 0 et 0,4. Ceci peut en partie s'expliquer par le fait que de nos jours, beaucoup de musiques sont de type électro ou dance house et ne sont constituées que de beats et sans ou avec très peu de paroles. Un autre facteur est celui de l'attention portée aux paroles. Dans les années 40,50,60,70, les chansons (notamment françaises) étaient constituées de paroles plus élaborées. Ces paroles avaient autant d'importance et de profondeur que la musique en elle-même. Au fur et à mesure des décennies, écrire bien est une valeur qui s'est perdue. Aujourd'hui, on trouve peu de chansons avec des paroles recherchées.

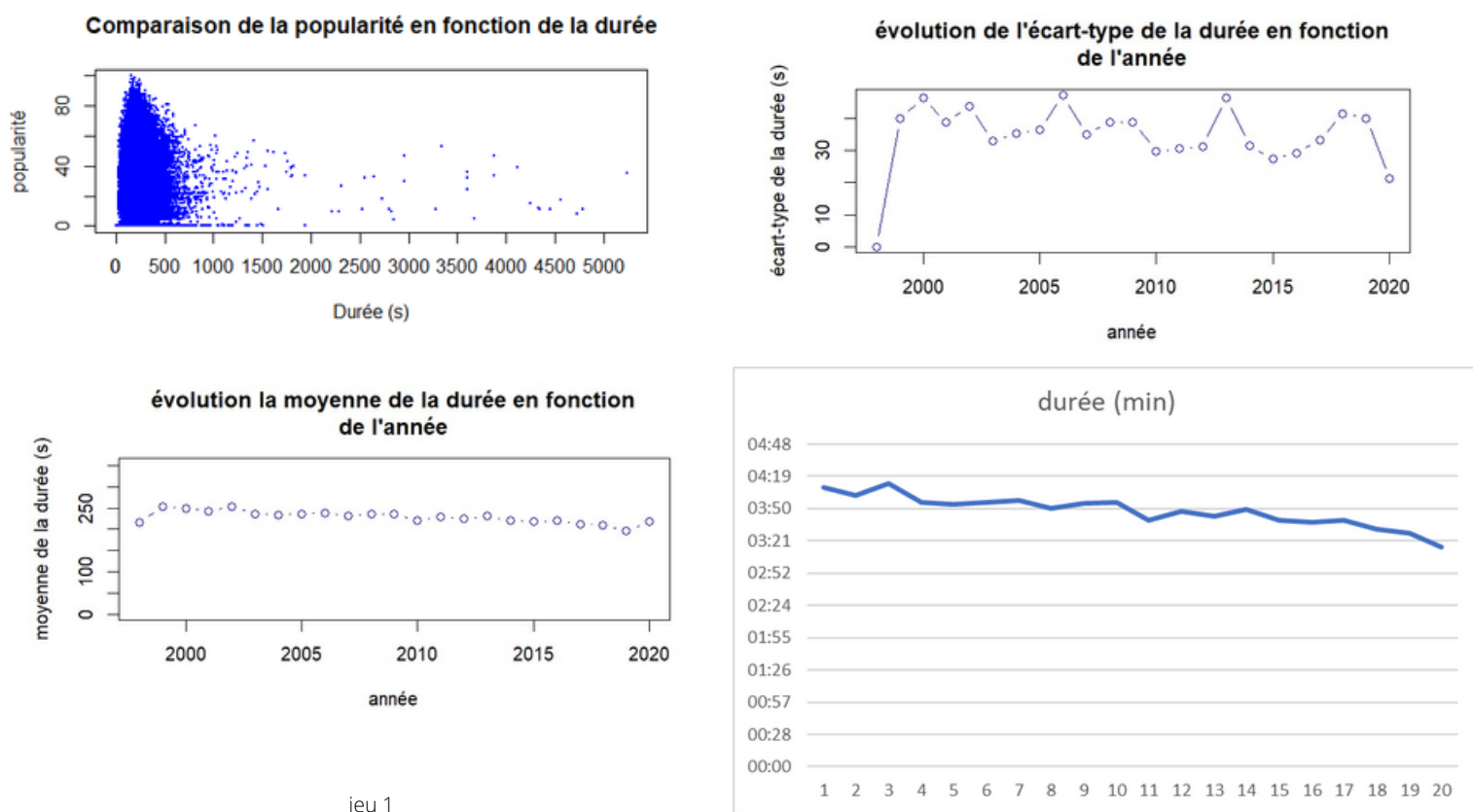
Depuis plusieurs années, nous avons pu remarquer que de plus en plus de chansons comportaient des contenus explicites. Nous avons donc voulu représenter le nombre de chansons explicites selon les années et on voit clairement que ce nombre a considérablement augmenté. Nous n'avons pas étudié le graphe pour l'année 2020 car le jeu de données 2 ne contenait que trois chansons pour cette année: les données n'auraient pas été représentatives de la réalité. Cependant, nous remarquons que ce nombre de chansons explicites reste minoritaire en prenant en compte l'ensemble des chansons populaires. Avec le diagramme camembert, on voit que plus de 3/4 des chansons à forte popularité comportent des paroles non explicites. Cela suit une certaine logique car on peut partir du principe qu'elles toucheront un public plus large si ces dernières sont entendables pour une population plus jeune. Pour obtenir ce graphique, nous avons supprimé toutes les chansons qui avaient un indice de popularité inférieur à 70 et nous concentrer uniquement sur les hits.



### Paroles explicites dans les musique à forte popularité ( $\geq 70$ )



## Hypothèse n°4: Une grande partie des musiques ont la même durée:



jeu 1

Ce graphique permet de constater qu'au moins 90% des chansons se concentrent autour d'une même durée.

Au-delà de la question de la popularité, ceci semble donc toucher presque l'intégralité des titres. Ainsi, une caractéristique évidente de la popularité d'une chanson et valeur sûre va être sa durée qui doit ne pas dépasser 500 secondes, soit 8 minutes, même si la popularité est bien plus importante (presque 100) pour la moitié de ce temps-là: soit 3min50-4 min.

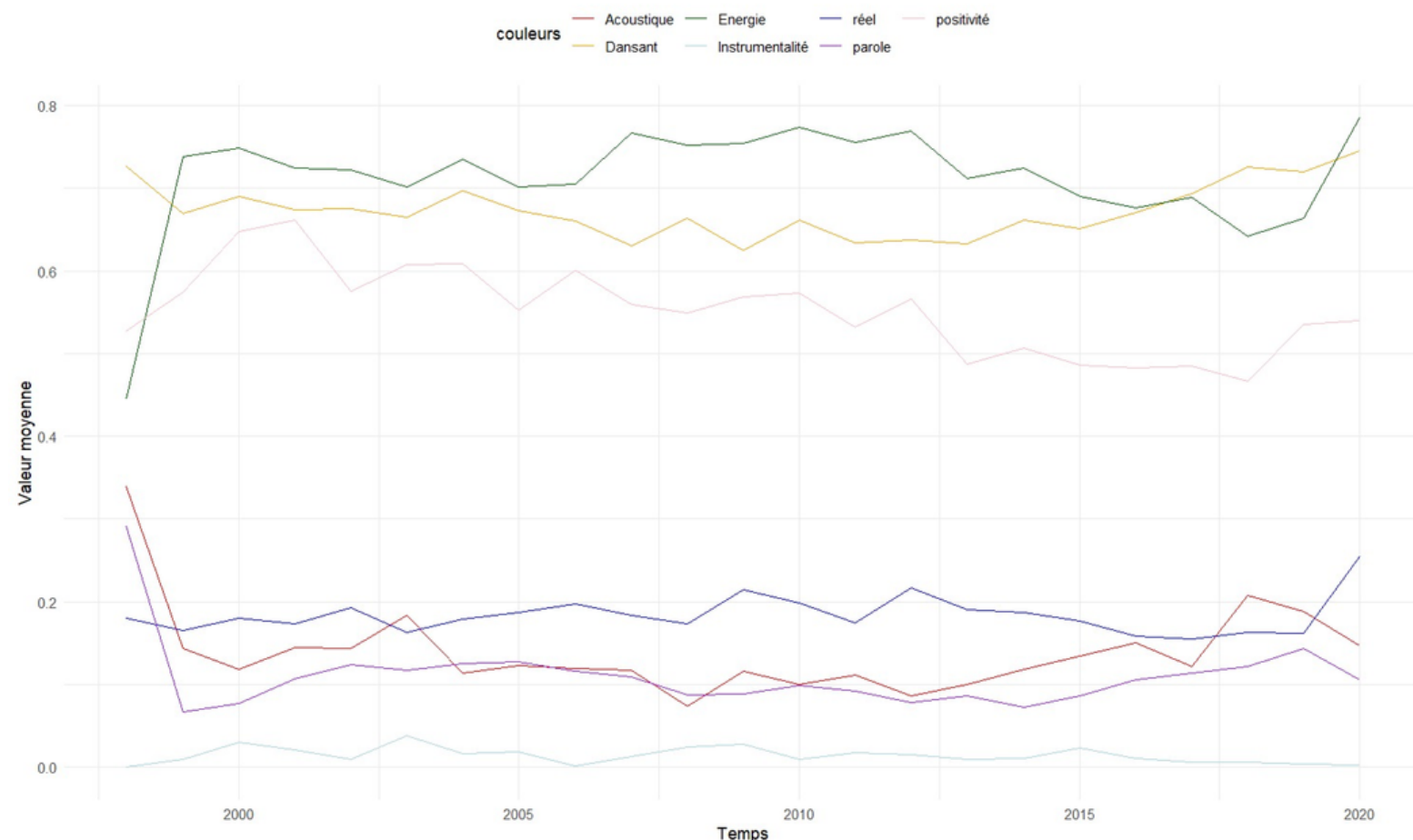
Le second graphique met en parallèle avec le troisième permet de montrer que bien qu'il y ait des variations de la durée en fonction de l'année, la durée moyenne des titres restent globalement linéaire. On peut donc penser que même si la durée a tendance à diminuer légèrement, toutes les musiques varient en même temps et les artistes s'adaptent à la tendance, ce qui explique les différences de variation faibles.

Enfin, le quatrième graphique permet de voir qu'il y a une légère diminution de la durée (cette fois-ci on ne considère pas la moyenne mais la durée dans son ensemble) des titres au cours des années. La différence visuelle par rapport au graphique 3 s'explique par le changement d'échelle (temps en minutes et non plus en secondes sur l'axe des abscisses). De plus, on s'intéresse à une durée moyenne sur le graphique 3 alors que le graphique 4 on prend en compte la durée de chaque titre. Nous pouvons en conclure que la durée des titres est passée de plus de 4min en 2001 à 3min16 en 2020. Les chansons se raccourcissent donc au cours du temps.



## Hypothèse n°5: Au fur et à mesure des années, les tendances changent et les musiques tendent à s'uniformiser :

Caractéristiques musicales en fonction du temps jeu 2

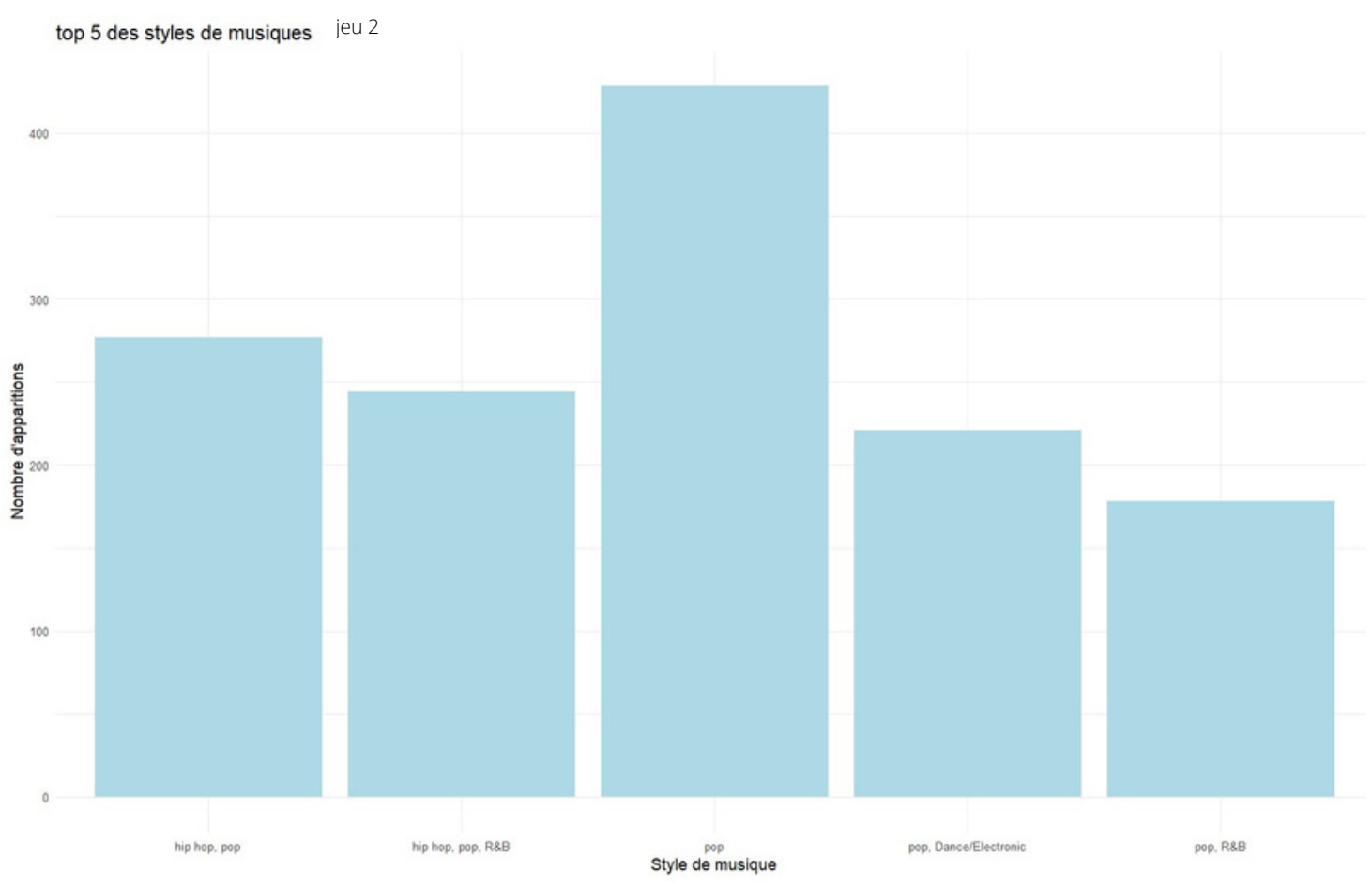
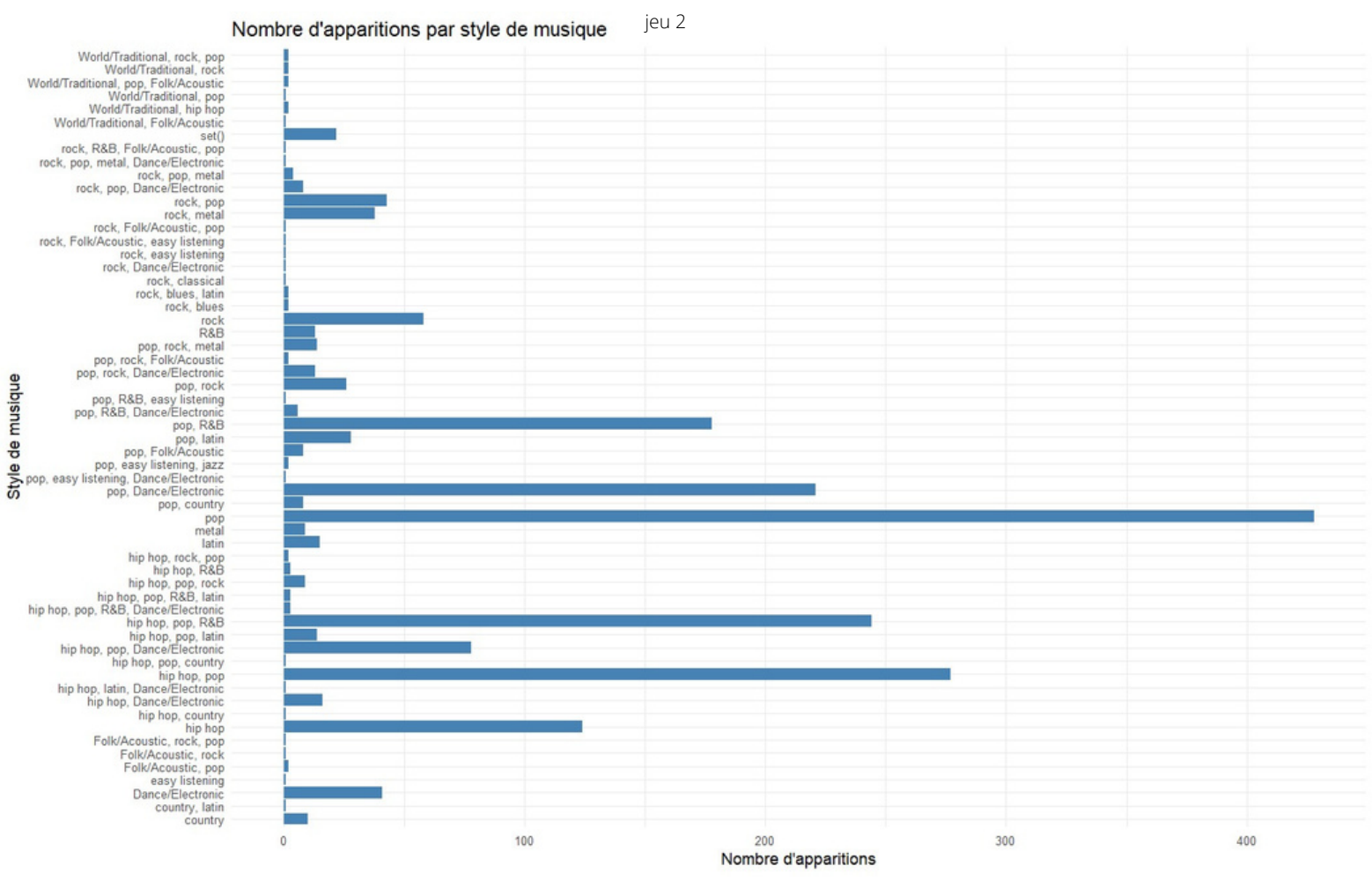


Nous ne disposons pas de données datant d'avant 1998 mais il serait intéressant de noter l'évolution des caractéristiques des chansons sur une échelle temporelle plus grande. Toutefois, nous pouvons noter qu'aux abords des années 2000, beaucoup de caractéristiques ont changé. Le niveau d'énergie a brusquement augmenté tandis que les paroles et le degré acoustique ont baissé. On peut penser alors à l'apparition des styles techno, électro...

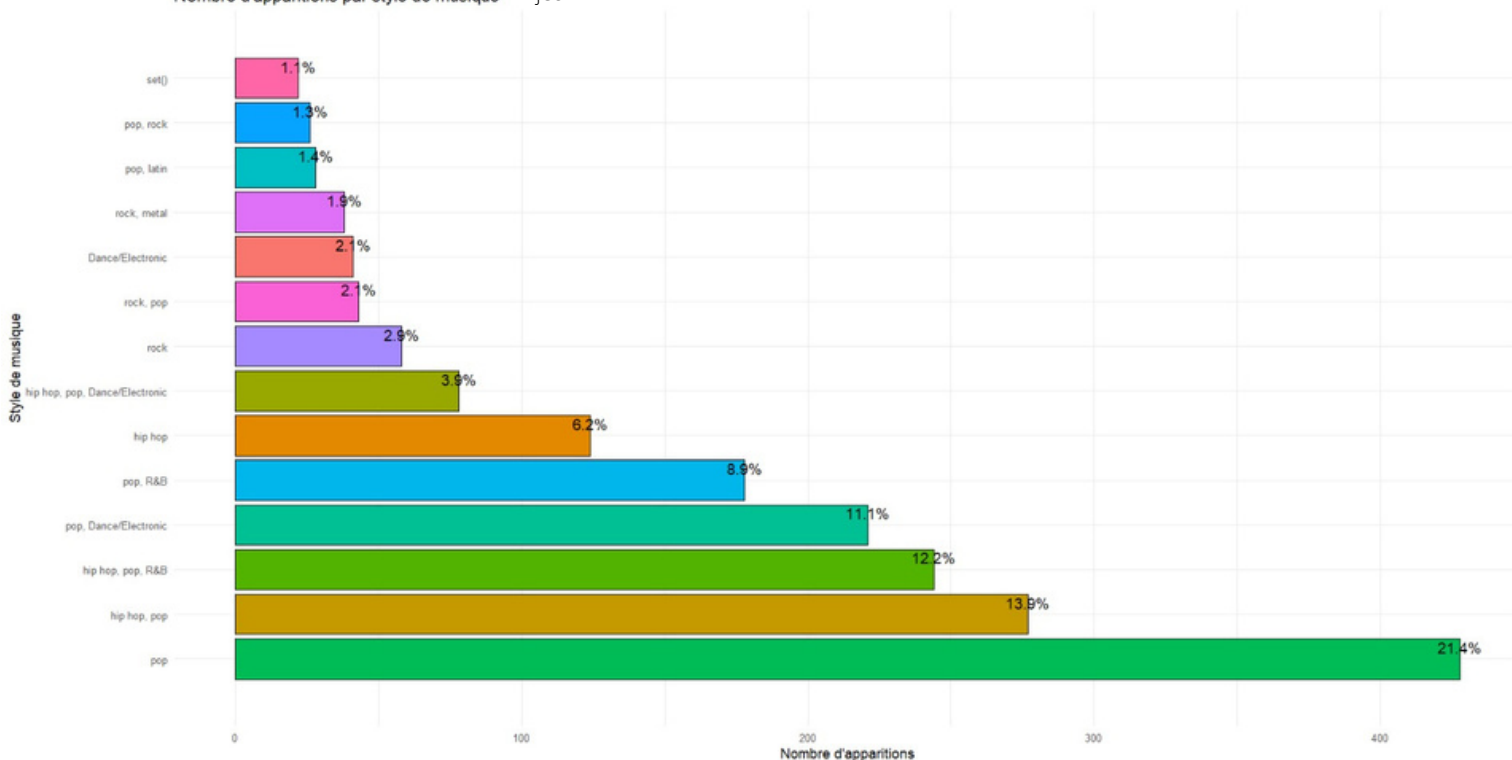
Si on étudie certaines caractéristiques une par une, voici ce qu'on peut en tirer:

- L'instrumentalité est restée globalement la même au cours du temps, bien que si on la regarde depuis 2015, elle est en légère diminution constante.
- À l'approche de l'année 2020, le taux de parole a diminué, particulièrement durant les années 2000, comme on a pu le faire remarquer plus haut.
- Depuis les années 2007, le degré dansant des musiques a été en augmentation malgré certaines variations. L'augmentation la plus marquée se fait d'ailleurs depuis 2015. On peut faire le parallèle avec la popularisation des musiques électro, des DJ, des festivals, des réseaux sociaux (TikTok avec le phénomène des danses sur certaines musiques)...
- L'énergie augmente beaucoup également depuis 2018 environ.

Globalement, on pourrait se figurer un intervalle fictif pour chaque caractéristique dans lequel toutes les valeurs rentreraient. Ce qui corrobore bien cette uniformisation que l'on avait supposée. Les artistes ont tendance à se répéter et à faire la même chose car c'est ce qui marche et plaît. Peu de musiques sortent du lot et se distinguent par leur originalité.



Nombre d'apparitions par style de musique jeu 2



Pour poursuivre sur l'étude des tendances, certains genres musicaux font très clairement leur effet et sont beaucoup plus largement représentés que d'autres. Ceci n'est pas nouveau mais c'est le cas de la pop qui apparaît dans le top 5 des styles musicaux les plus écoutés. Cela est dû notamment à sa polyvalence. En effet, la pop peut être mixée avec d'autres styles et donc s'adapte presque à tous les goûts. Il y a le hip hop, le hip hop pop et R&B, la pop et Dance Electronique et la pop et le R&B par exemple (d'après le graphique ci-dessus du top 5 des styles musicaux). Son nombre d'apparitions combinées total est de 1330 et la pop (seule) de 420 environ contre moins de 50 apparitions pour 51 autres styles tels que le métal ou la musique country.

La pop apparaît donc à elle seule huit fois de plus que 86% des autres styles ensemble. Le graphique 3 nous permet d'ailleurs de voir qu'elle représente à 21,4% des styles musicaux apparus et combinées à d'autres styles (ceux du top 5), ce taux monte même jusqu'à 73,7%.

On a ainsi une très forte prédominance de la pop et ce n'est d'ailleurs pas pour rien que l'on parle de "pop culture". Cela tend également à rejoindre notre hypothèse d'uniformisation des écoutes, styles et tendances. La majeure partie de la population écoute donc la même chose.



Le graphique ci-contre nous permet de faire un parallèle entre nos résultats et une étude qui a déjà été publiée par statista en 2018. On observe que nos résultats pour la pop sont cohérents puisque dans l'étude statista, la pop représente 64% des genres musicaux écoutés dans le monde. Nous trouvons de notre côté 73,7 % de pop (pouvant être combinée à d'autres styles). L'écart s'explique par rapport à la population sondée, la différence de temporalité et le nombre de pays représentés.



# Conclusion

En somme, il est impossible de pouvoir prédire avec exactitude si une chanson va devenir un hit ou non. Toutefois, notre étude nous a permis de faire ressortir plusieurs caractéristiques communes aux chansons d'indice de popularité élevé. Ceci peut permettre de s'orienter et d'avoir plus de chance qu'une chanson rencontre un certain succès. En regroupant ces caractéristiques, l'artiste s'assure avec une plus grande probabilité que sa chanson soit bien accueillie par le public.

Il faut retenir que les chansons les plus populaires sont des chansons rythmées, énergiques et dansantes. Leur rythme est assez rapide, elles ne présentent pas une grande variété instrumentale mais plutôt des sons artificiels. Il n'y a pas forcément beaucoup de paroles et leur intensité sonore est élevée. À ceci s'ajoute le fait qu'elles ont toutes à peu près la même durée oscillant autour des 3min15-3min50.

L'étude de nos données a permis de valider les hypothèses que nous avons faites, sauf celle sur le fait que nous pensions que les chansons populaires avaient plutôt des paroles explicites, ce qui a été contredit avec les résultats obtenus.

Un point marquant qu'il est important de retenir est l'omniprésence de la pop dans notre culture musicale depuis maintenant plusieurs années (début des années 2000).

Nous avons aussi observé une uniformisation des genres musicaux et des styles de musique.

De nos jours, on peut retrouver ce phénomène d'uniformisation dans plusieurs aspects de notre société, comme la manière de s'habiller ou de filmer. Nous avons de plus en plus tendance à nous conformer à ce qui nous entoure et à faire comme tout le monde. Cela est peut-être aussi lié à la poursuite du succès et donc à la reproduction de ce qui marche déjà.

Les difficultés que nous avons rencontrées:

- se rendre compte que notre jeu de données 1 n'était pas suffisant pour notre étude et qu'il fallait le compléter avec un autre jeu en raison de l'absence de la date de sortie des chansons
- savoir comment faire le lien entre les données dont on dispose dans l'Excel et ce qu'on veut prouver
- déterminer la meilleure manière de représenter les données (choix du type de graphique)

## Partie Problématique et Hypothèses:

- Tableau du TOP 10 des chansons les plus écoutées sur spotify: Wikipedia  
[https://en.wikipedia.org/wiki/List\\_of\\_most-streamed\\_songs\\_on\\_Spotify](https://en.wikipedia.org/wiki/List_of_most-streamed_songs_on_Spotify)

## Partie Jeux de données:

- Jeu de données n°1:  
<https://www.kaggle.com/datasets/tomigelo/spotify-audio-features>
- Jeu de données n°2:  
<https://www.kaggle.com/code/ashvanths/statistical-analysis-of-top-hits-in-spotify/input>

## Parallèles avec étude déjà menées:

- les genres musicaux les plus écoutés dans le monde: Statista  
<https://fr.statista.com/infographie/15840/genres-de-musique-preferes-dans-le-monde/>
- Prédire le succès d'une musique du Spotify | Jedha Bootcamp  
<https://www.jedha.co/blog/predire-le-succes-dune-chanson-sur-spotify>
- Les bienfaits de la musique: radio france  
<https://www.radiofrance.fr/franceculture/frisson-musical-c-est-votre-memoire-qui-fait-tout-le-boulot-5564168>
- Sciences et Avenir:  
[https://www.sciencesetavenir.fr/sante/les-bienfaits-averes-de-la-musique\\_143967](https://www.sciencesetavenir.fr/sante/les-bienfaits-averes-de-la-musique_143967)

