

Rapport d'analyse des classements musicaux sur Spotify entre 2019 et 2020

Problématique: Les confinements liés au covid-19 influencent-ils le choix des musiques de français?

ZHANG Mingjie

2022-06-07

Table des matières

Introduction	2
présentation générale	2
présentation des données	2
Analyse des données	3
Pertinence de l'analyse	3
Les axes de l'analyse	5
Analyse des graphes	5
Conclusion	8
Synthèse	8
Bilan personnel	8

Introduction

présentation générale

Le covid-19 a grandement impacté la vie des gens, notamment lorsque les confinements ont été imposés, il est donc intéressant d'étudier leur comportement, et la musique étant très omniprésente dans leur quotidien, ce qui fait d'elle un bon sujet d'étude à mon avis.

Cette analyse de données a donc pour but d'étudier s'il existe, un lien entre les musiques qu'écoutent les français sur Spotify et les confinements liés au covid-19. Pour cela, on analysera les classements musicaux quotidiens francophones des années 2019 et 2020. On comparera les données de 2019 en tant que témoin avec celles de 2020.

présentation des données

Les fichiers téléchargés depuis le site "<https://charts.spotify.com/>" ne contiennent que les informations concernant le rang, le nombre de streaming, l'uri, le nom de la musique ainsi que le nom de l'artiste.

On s'aide donc de la librairie "spotifyr" pour obtenir les informations sur les musiques tels que:

-energy: énergie(l'intensité et l'activité, entre 0 et 1, une musique type "métal" a une valeur proche de 1)

-loudness: intensité sonore

-speechiness: élocution (la présence de paroles, entre 0 et 1)

-acoustic: acoustique (La présence d'instrument non électrifiés, entre 0 et 1)

-valence: valence (la positivité de la musique, entre 0 et 1)

-liveness: vivacité (la présence d'une audience dans l'audio, une valeur au dessus de 0.8 indique que la musique est probablement un enregistrement d'un concert)

-tempo: tempo

-instrumentalness: instrumentalité (la présence de voix, entre 0 et 1, plus la valeur est haute, moins la présence de voix est importante)

-danceability: incitation à danser (si la musique est faite pour danser, entre 0 et 1)

Analyse des données

Pertinence de l'analyse

Le but étant de savoir si les confinements ont eu un impact sur les musiques qu'écourent les français. C'est pourquoi, on doit d'abord comparer le tableau de cohérence des différentes variables décrites au dessus de 2019 avec celui de 2020, afin de savoir s'il est intéressant de comparer les données de 2019 avec celles de 2020. En effet, si les styles de musiques écoutées diffèrent d'une année à une autre, il serait impertinent de procéder au reste de l'analyse.

Un tableau de corrélation indique comment 2 variables sont corrélées, liées, une corrélation positive signifie que lorsque la valeur de l'une des 2 variables est grande relativement, la valeur de l'autre est grande aussi relativement. Et au contraire, une corrélation négative signifie que quand la valeur de l'une des 2 variables est grande relativement, l'autre est petite relativement. Ces rapports sont plus importants à mesure que la valeur absolue de corrélation s'approche de 1. Et une corrélation nulle signifie qu'il n'y a aucun lien entre les 2 variables. Par exemple, un retour à la mode du tecktonik augmenterait la corrélation entre le tempo et l'énergie, ou encore le tempo et l'intensité sonore.

```
##          danceability energy loudness speechiness acoustictness
## danceability          1.0    0.1    0.1          0.1          -0.2
## energy                0.1    1.0    0.7         -0.1          -0.3
## loudness              0.1    0.7    1.0         -0.2          -0.4
## speechiness           0.1   -0.1   -0.2          1.0           0.1
## acoustictness        -0.2   -0.3   -0.4          0.1           1.0
## instrumentalness      0.0   -0.1   -0.1          0.0           0.0
##          instrumentalness
## danceability          0.0
## energy               -0.1
## loudness             -0.1
## speechiness           0.0
## acoustictness         0.0
## instrumentalness      1.0
```

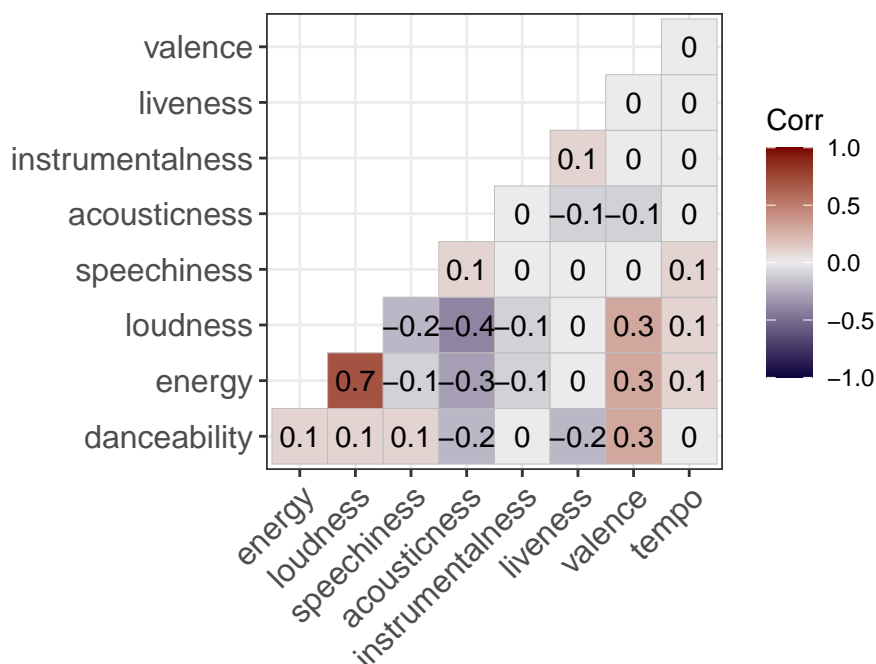


Tableau de corrélation de 2019

##	danceability	energy	loudness	speechiness	acousticness
## danceability	1.0	0.1	0.1	0.1	-0.2
## energy	0.1	1.0	0.7	0.0	-0.4
## loudness	0.1	0.7	1.0	-0.1	-0.3
## speechiness	0.1	0.0	-0.1	1.0	0.0
## acousticness	-0.2	-0.4	-0.3	0.0	1.0
## instrumentalness	-0.1	0.0	-0.1	0.0	0.0
##	instrumentalness				
## danceability	-0.1				
## energy	0.0				
## loudness	-0.1				
## speechiness	0.0				
## acousticness	0.0				
## instrumentalness	1.0				

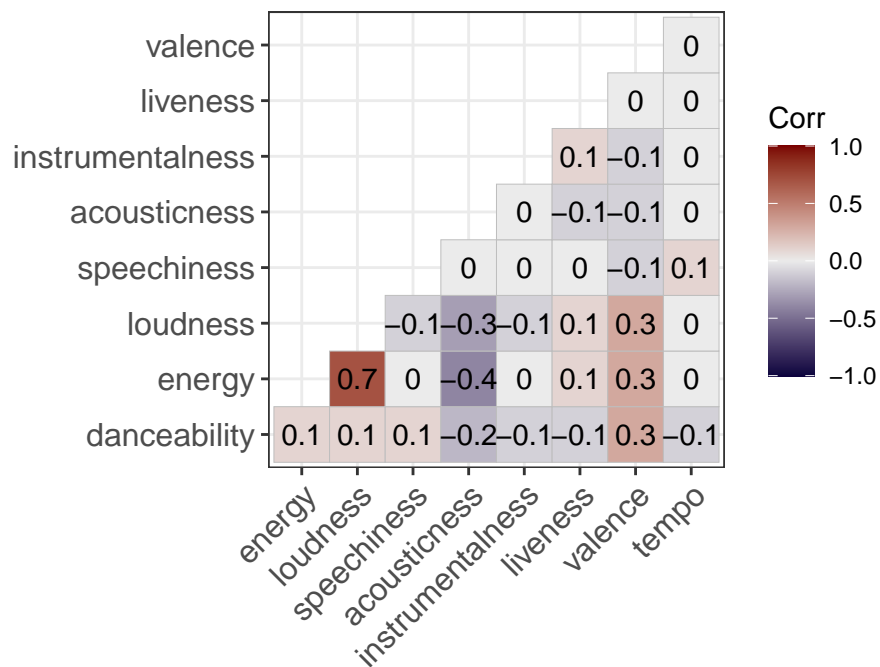


Tableau de corrélation de 2020

En comparant les 2 tableaux, on remarque dans la globalité la valeur de corrélation pour des mêmes variables est assez similaires entre 2019 et 2020, à ± 0.1 près, ce qui est assez négligeable pour l'échelle considérée, cela représente une différence de valeur de $\pm 5\%$.

On observe aussi des valeurs de corrélations non négligeables, comme par exemple entre énergie et l'intensité sonore où on a une valeur de 0.7, une valeur qui est normale, car plus une musique est énergétique plus elle a la tendance à avoir une intensité sonore élevée. De même, pour la valeur de corrélation entre l'intensité sonore et l'acoustique, une musique comportant des instruments non numériques a une intensité sonore plus élevée que des instruments numériques.

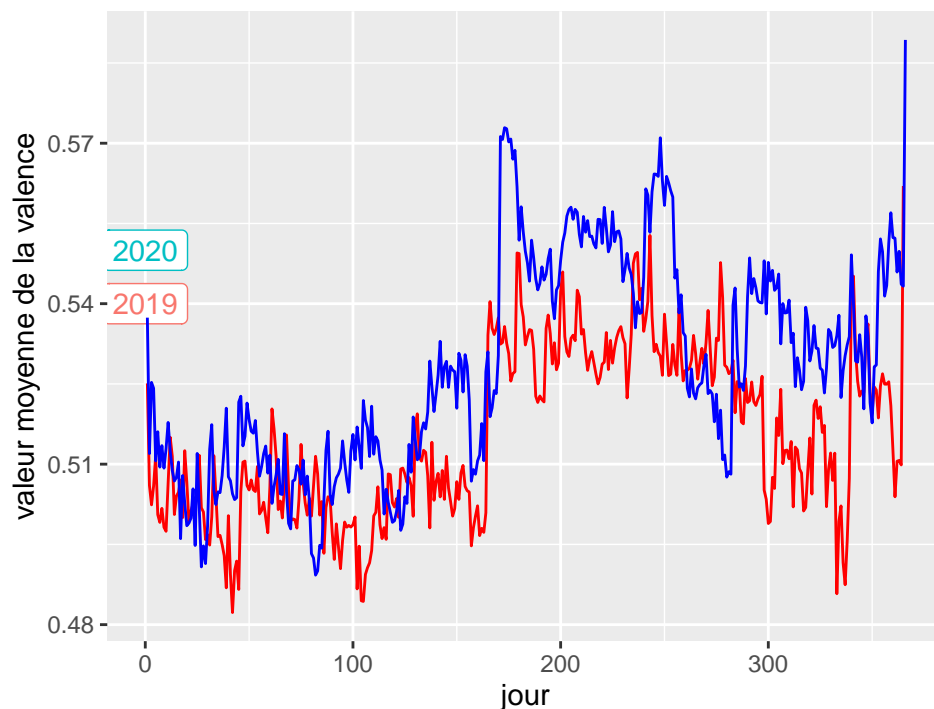
Les axes de l'analyse

On choisit d'étudier les graphes en fonction de 2 paramètres:

- La pertinence de la variable dans ce contexte.
- Sa valeur de cohérence dans le tableau avec les autres variables.

En effet, la vivacité d'une musique paraît peu pertinente à étudier, le fait qu'une musique soit ou non un enregistrement d'un concert n'est pas intéressant dans ce cadre, il serait aussi inutile d'étudier 2 graphes sur 2 variables dont la valeur absolue de cohérence est supérieure à 0.2, les allures des graphes sont assez similaires dans ces cas là. C'est pourquoi l'analyse se fera sur les graphes portant sur la valence, le tempo, instrumentalité, l'acoustique et l'élocution.

Analyse des graphes

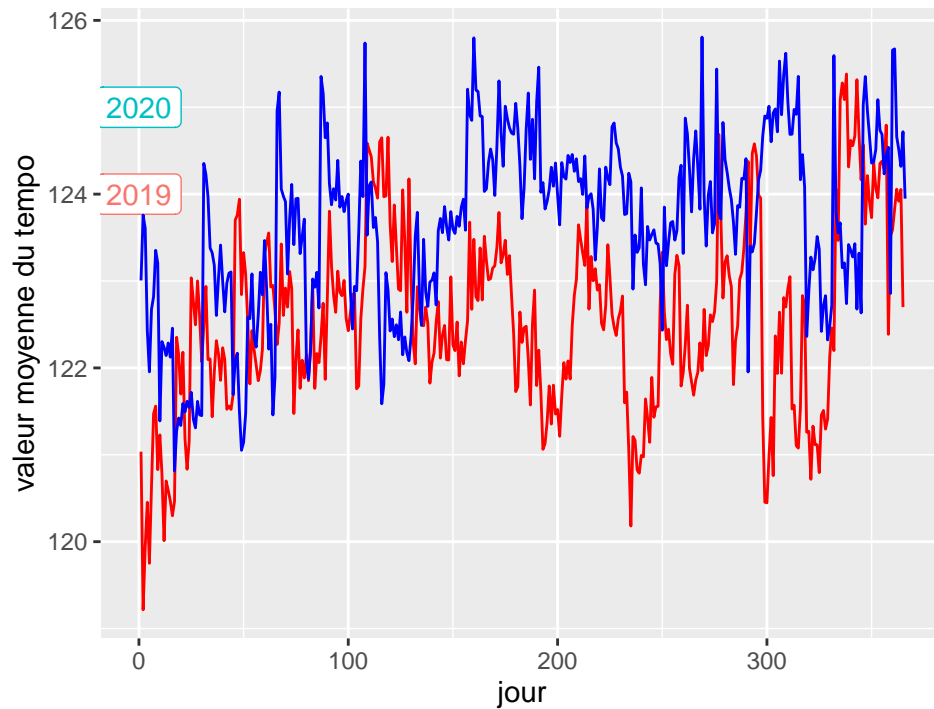


Graphe montrant l'évolution de la valeur moyenne de la valence en fonction du jour des années 2019 et 2020

On constate que dans la globalité, la courbe de 2020 est au dessus de celle de 2019, et qu'il y a une montée brusque des valeurs vers le 160ème jour pour les 2 courbes. La montée brusque peut s'expliquer par le fait que vers le 160ème jour, c'est à dire vers début mai, on a généralement une amélioration de la météo, avec un climat plus chaud. Donc cela justifie que les gens ont plus tendance à écouter des musiques avec une valence plus grande, donc plus positive, plus joyeuse, mais ce saut est encore plus important pour la courbe de 2020, cela pourrait s'expliquer par la joie que procure le fait que le premier confinement s'est fini le 11 mai 2019.

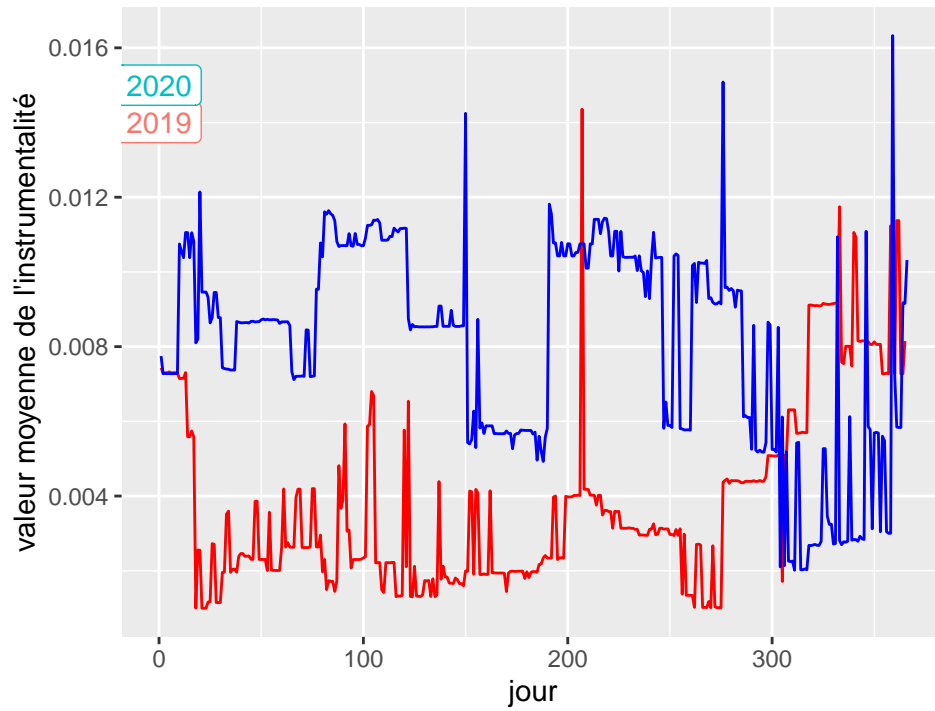
De plus, on observe une certaine périodicité avant le saut des valeurs, en effet, pour la courbe de 2019, la valeur moyenne de la valence semble varier périodiquement avec une période de 60 jours dont la valeur maximale ne dépasse pas 0.53, et ça semble être le cas aussi pour la courbe de 2020 à une exception près, la période qui commence vers le 120ème jour a un maximal plus grand que les autres périodes, et à cette date,

on est à vers 3 semaines après le début du confinement, l'augmentation de la valeur moyenne de la valence peut s'expliquer par la joie de rester chez soi et de ne pas travailler, et/ou l'écoute des musiques positives comble l'ennui.



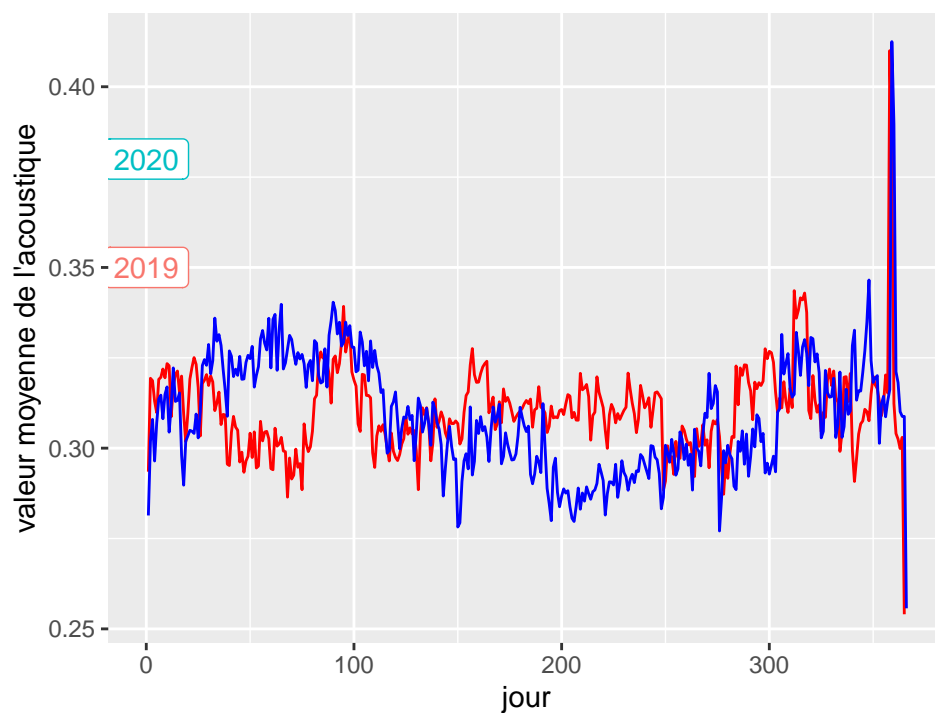
Graphique montrant l'évolution de la valeur moyenne du tempo en fonction du jour des années 2019 et 2020

On observe que jusqu'au 120ème jour, les valeurs des courbes de 2019 et 2020 restent assez proches avec une différence inférieure à 2bpm, et à partir de cette date, les valeurs de la courbe de 2020 sont supérieures à celles de 2019 jusqu'au 250ème jour.



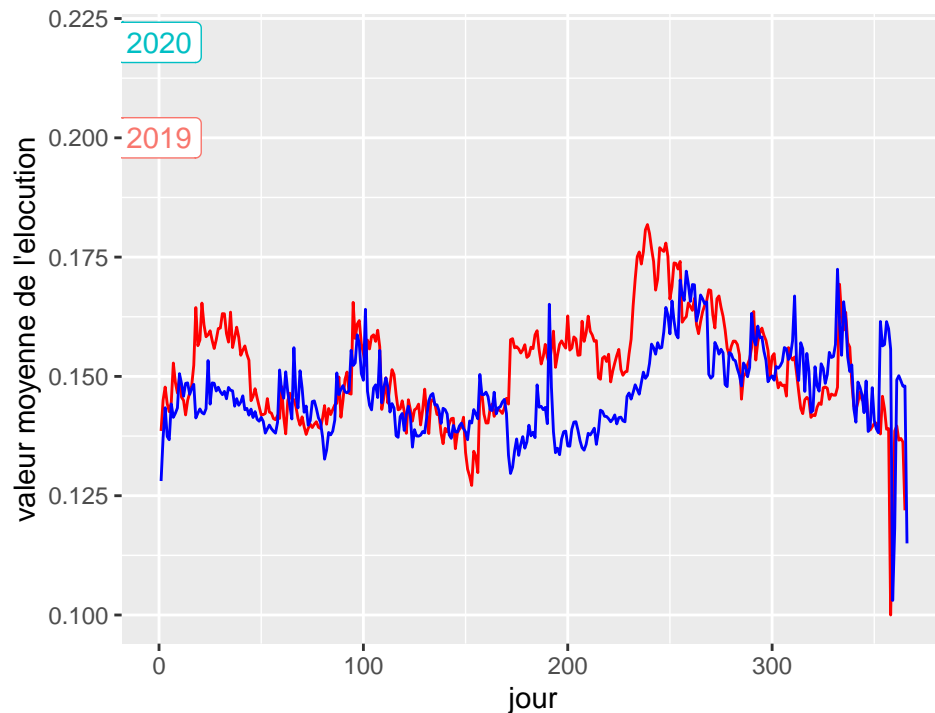
Graphique montrant l'évolution de la valeur moyenne de l'instrumentalité en fonction du jour des années 2019 et 2020

On constate que la courbe de 2020 est généralement au dessus de celle de 2019, et ce même avant le début des confinements, par conséquent, cela ne nous permet pas d'établir de lien entre le confinement et l'instrumentalité des musiques écoutées.



Graphe montrant l'évolution de la valeur moyenne de l'acoustique en fonction du jour des années 2019 et 2020

On observe que jusqu'au 150ème jour, les valeurs de la courbe de 2020 sont généralement supérieures à celles de la courbe de 2019, et par la suite on assiste à une inversion de cette tendance jusqu'au 250ème jour, où les 2 courbes se rejoignent. Cette analyse n'est pas concluante.



Graphe montrant l'évolution de la valeur moyenne de l'élocution en fonction du jour des années 2019 et 2020

On remarque que l'évolution des courbes ne semblent pas avoir de lien logique avec les dates spécifique des confinements. Par conséquent, on ne peut rien en déduire.

Conclusion

Synthèse

Sur les 5 graphes analysés, seuls 2 graphes nous permettent d'établir un potentiel lien avec les dates de confinements, celui portant sur la valence et celui sur le tempo, et la valence ayant un coefficient de corrélation non négligeable de 0.3 avec l'intensité sonore, l'énergie et l'incitation à danser, on peut alors conclure qu'effectivement les confinements liés au covid-19 ont affecté les choix musicaux en 2020, les musiques ayant des valeurs hautes pour les variables ci-dessus se voient attribuées une position au classement top200 de France.

Bilan personnel

Dans la globalité, ce travail m'a été très instructif, le fait de découvrir un nouveau langage de programmation, ou encore le fait que les confinements ont affecté les musiques qu'écoutent les gens. Je trouve que cette

analyse n'est pas complète mais assez précise, en effet les données analysées sont des données des classements quotidiens de 2019 et 2020, soit 731 fichiers avec 200 musiques chacun, mais le mieux aurait été de pouvoir analyser les données de plus d'année, comme celles de 2016,2017 et 2021.