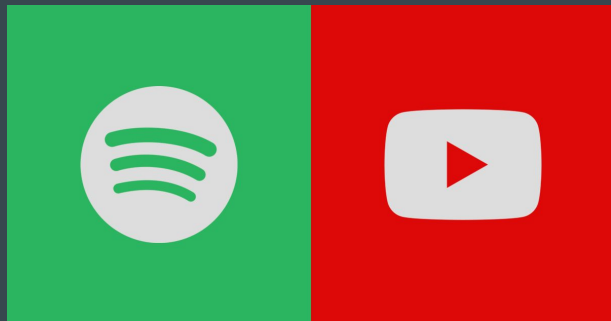


Analyse Exploratoire d'un jeu de données sur Spotify et Youtube

Équipe Zelda
IF36



Présentation du jeu de données

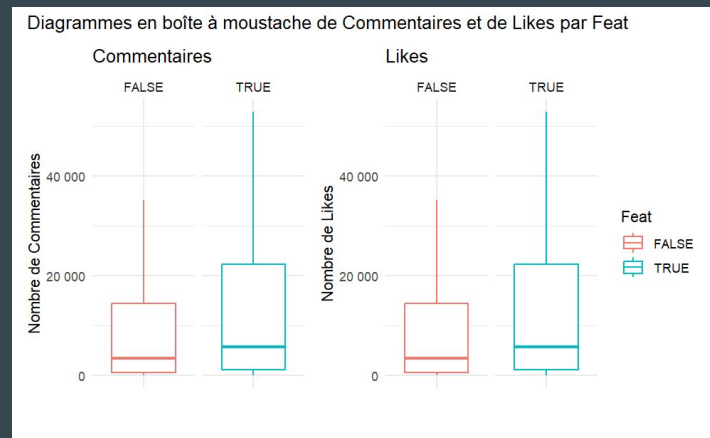
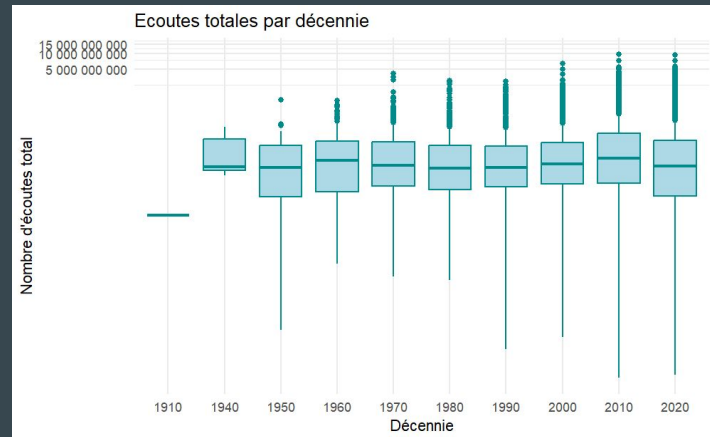
Source :

- Kaggle
- API Spotify

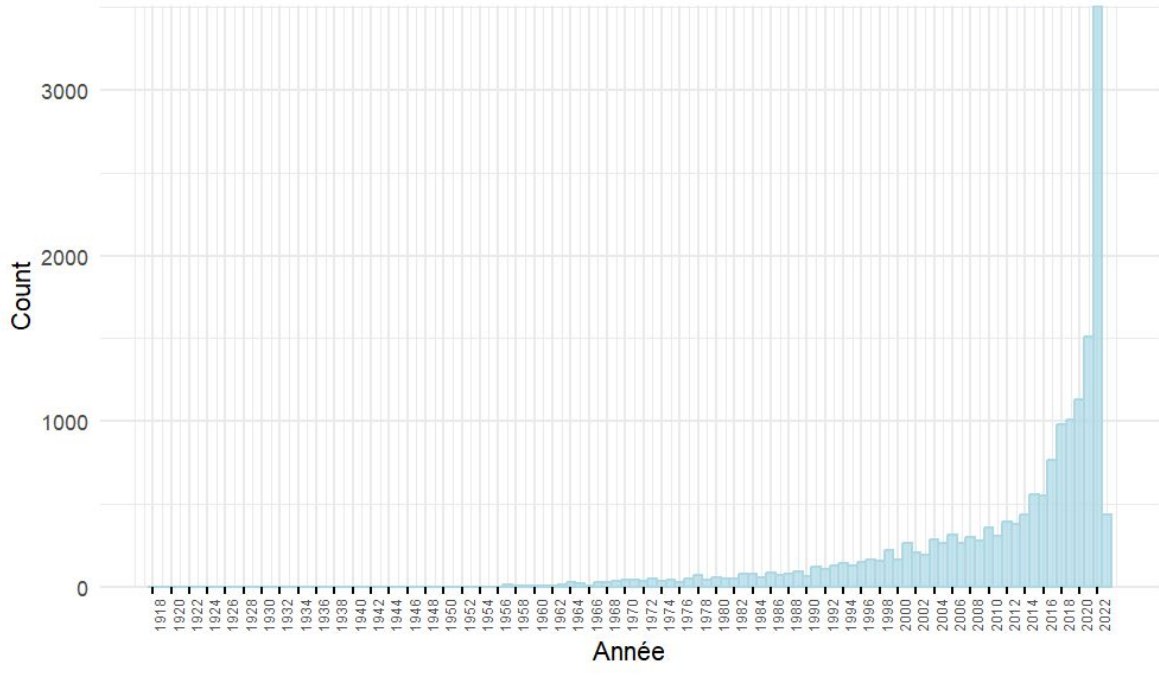
Données sur le titre	Données Spotify	Données Youtube
Artiste	URI : identifiant unique Spotify	Chaîne Youtube
Nom du titre	URL Spotify	Nombre de Vues
Nom de l'album	Nombre de Streams	Nombre de Likes
Type d'album (single, album,...)		Nombre de Commentaires
Date de publication		Description de la vidéo
+ Attributs divers (Energy, Key, ...)		Vidéo Officielle ou non

Pourquoi ce jeu de données ?

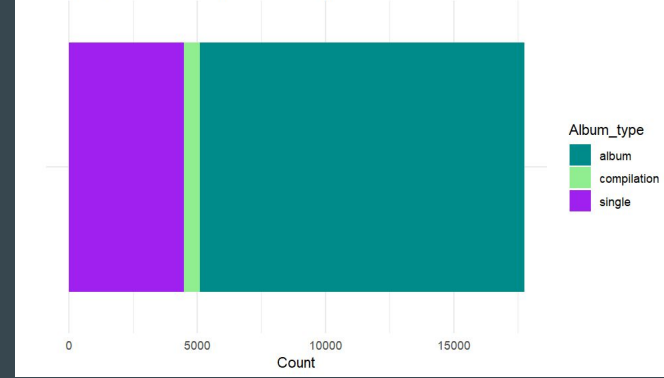
- Diversité et richesse des données
 - challenge : représenter autant de données de manière concise et exploitable (boxplot ou autres visualisations statistiques)
- Intérêt personnel
 - "Music is the universal language of mankind."
- H. Wadsworth



Distribution des titres du jeu de données en fonction de l'année de publication



Graphique à barres empilées des Types d'album



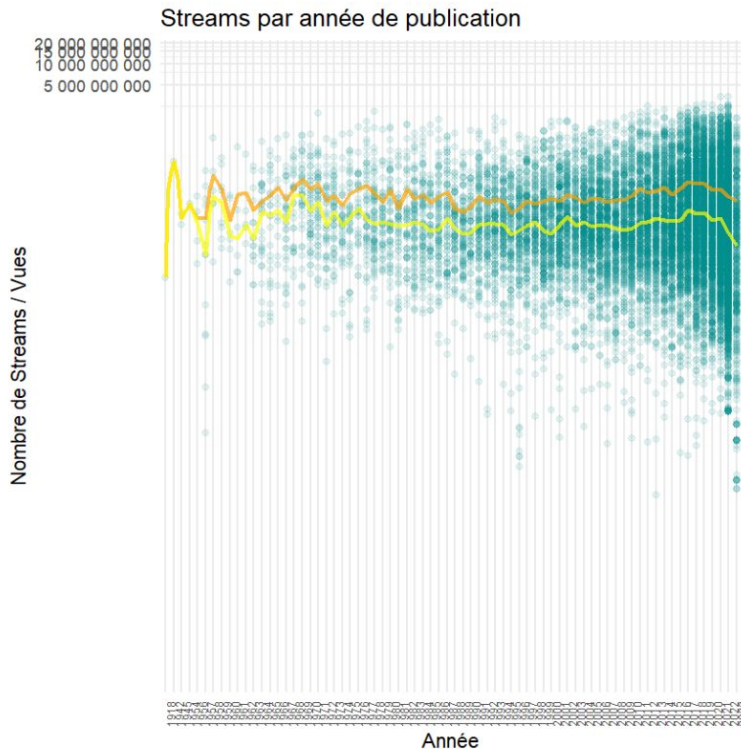
Echantillon :

- Concentré sur 2022
- Concentré sur les albums

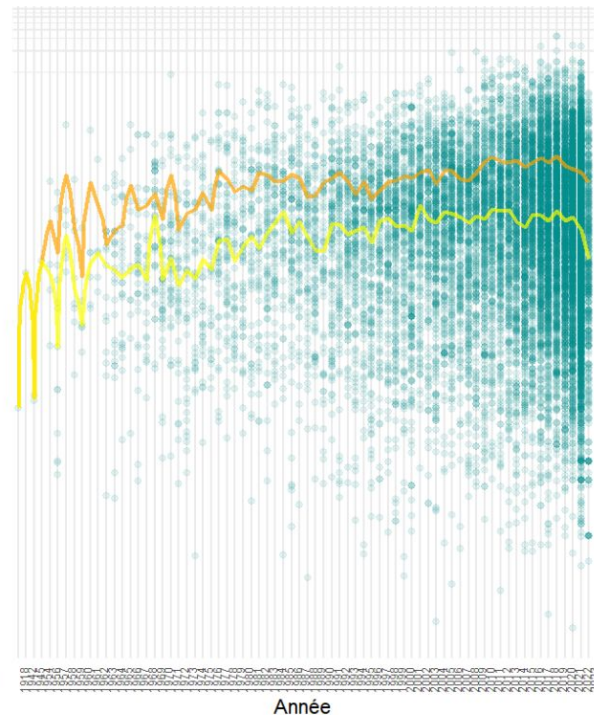
Choix d'échantillonnage du dataset

Ecoutes en fonction de la date de publication

Graphiques en échelle logarithmique

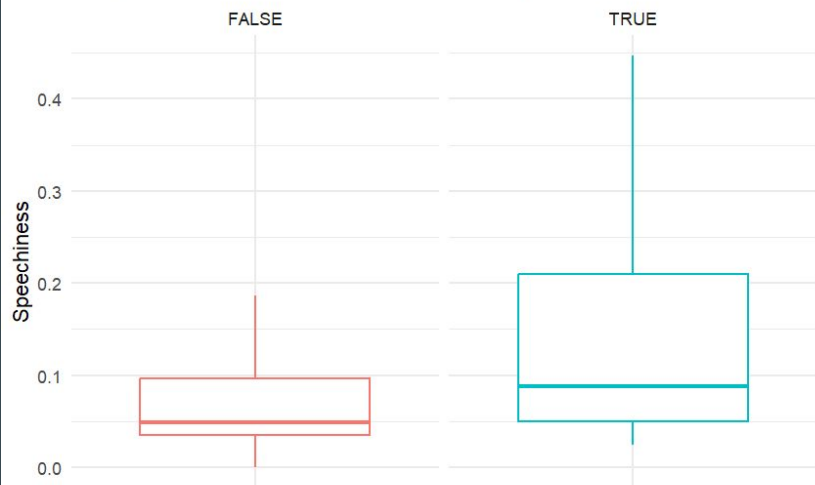


Vues par année de publication

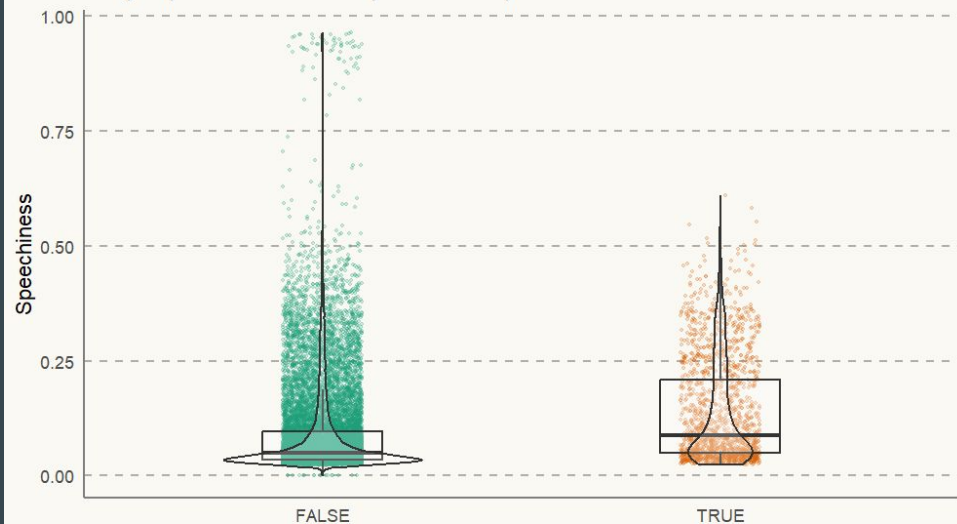


Mesures — Médiane — Moyenne

Diagrammes en boîte à moustache de Speechiness en fonction de Feat



Graphique en Violon de Speechiness par Feat

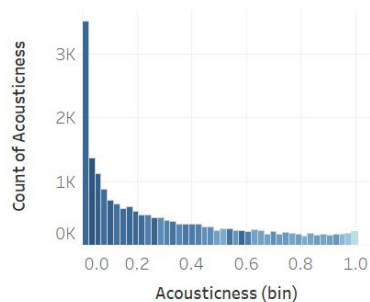


Les titres avec des collaborations (feat) sont-ils plus 'parlés' ?

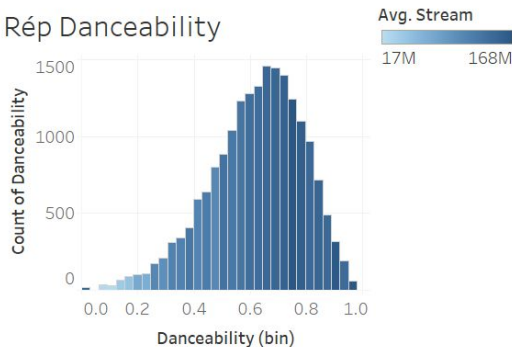


- Facilité de composition (Dashboard)
- Certains aspect du 'data cleaning' plus faciles
- Tester une visualisation avant de la faire avec Rstudio
- Exploration 'à l'aveugle' plus rapide

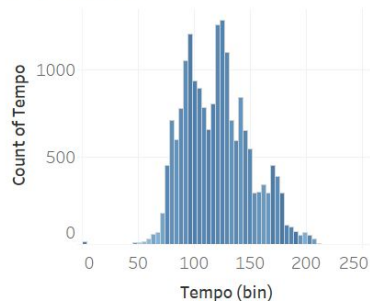
Rép Acousticness



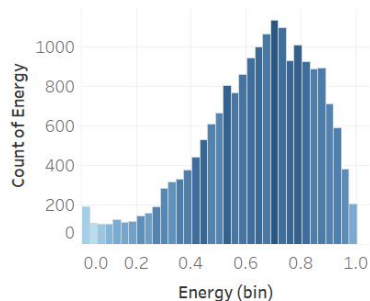
Rép Danceability



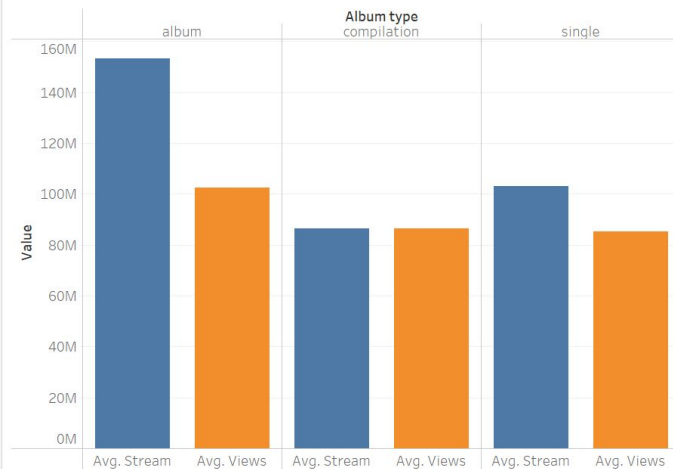
Rép Tempo



Rép Energy

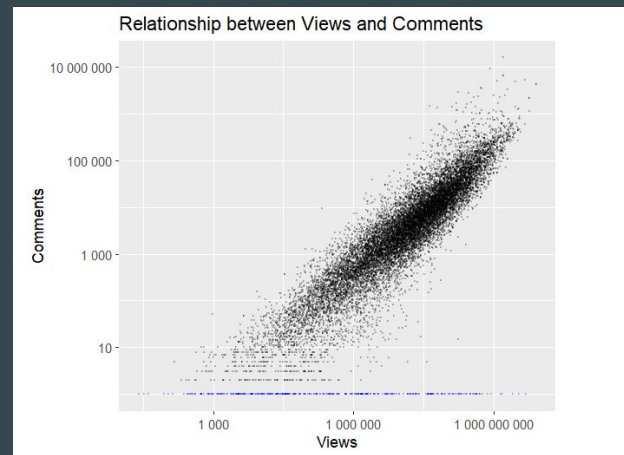
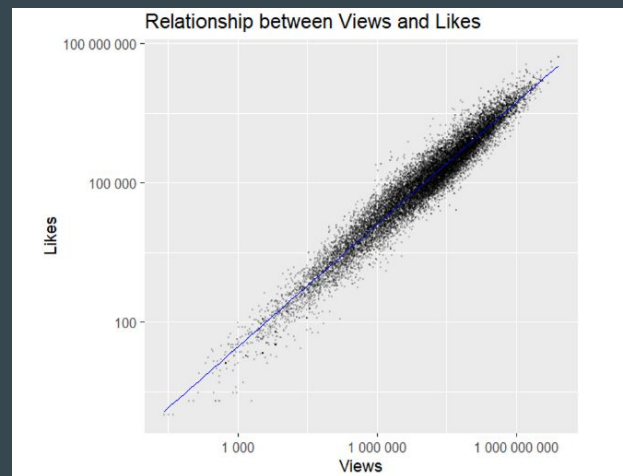
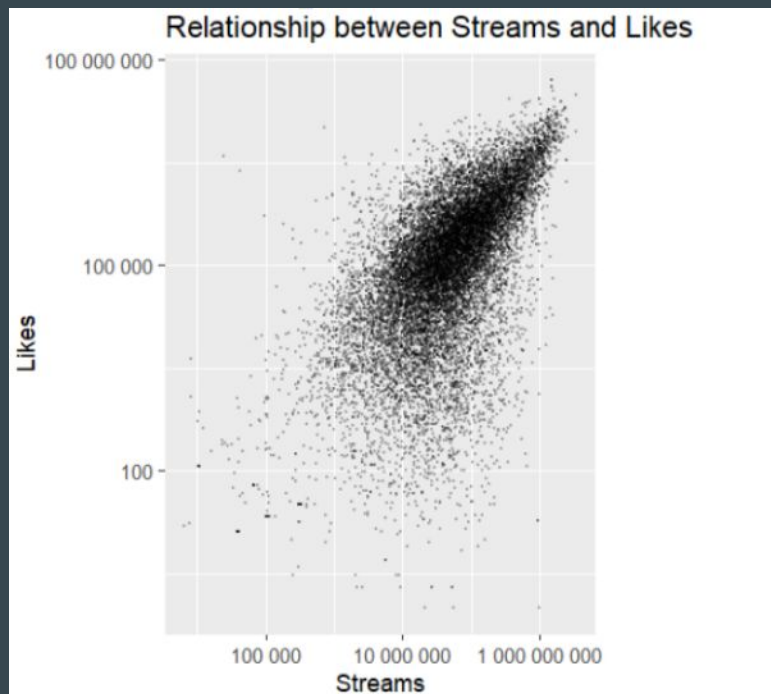


Est-ce que les singles sont plus écoutés/apprécies des auditeurs que les albums ?



Measure Names

Avg. Stream
Avg. Views

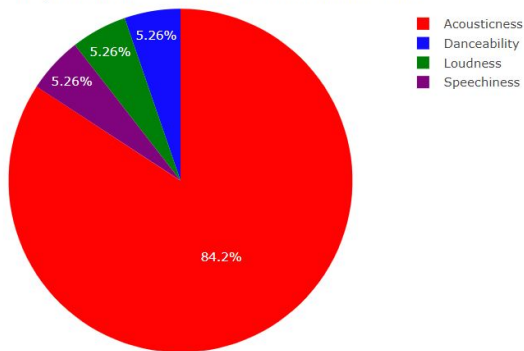


La relation entre le trafic, les vues, les likes et les commentaires

Les attributs des chansons populaires

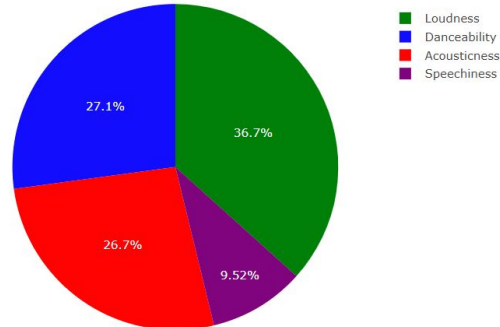
LASD

Répartition des attributs de chanson en 1962



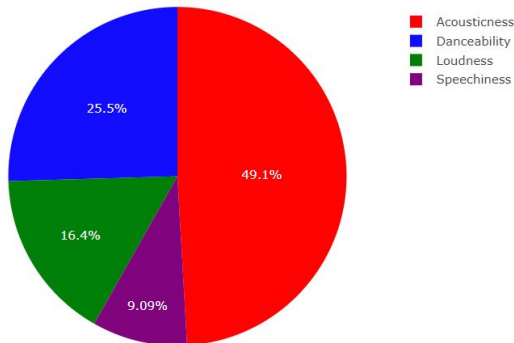
LASD

Répartition des attributs de chanson en 2000



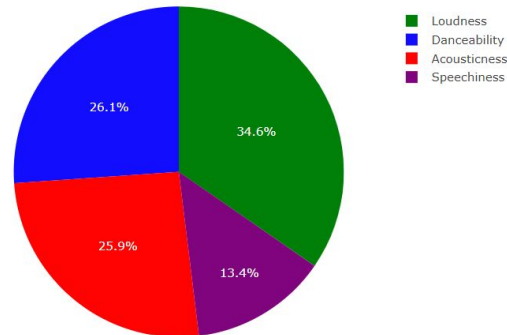
LASD

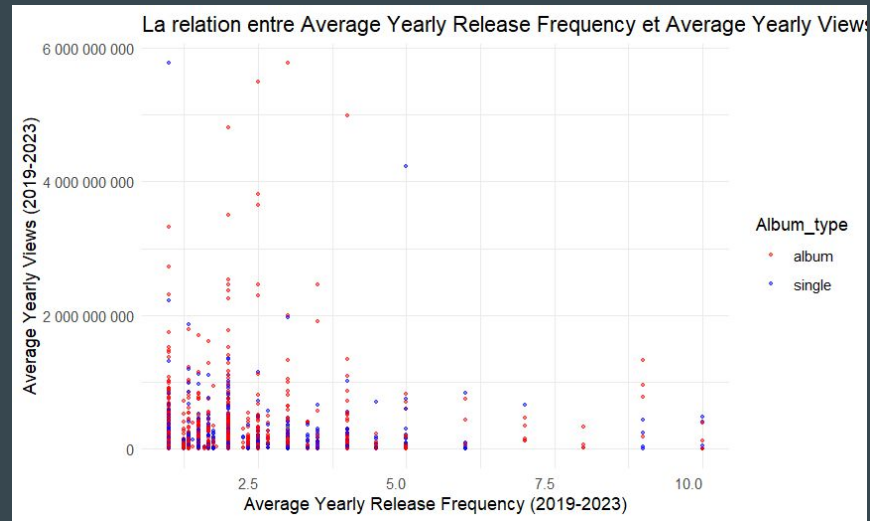
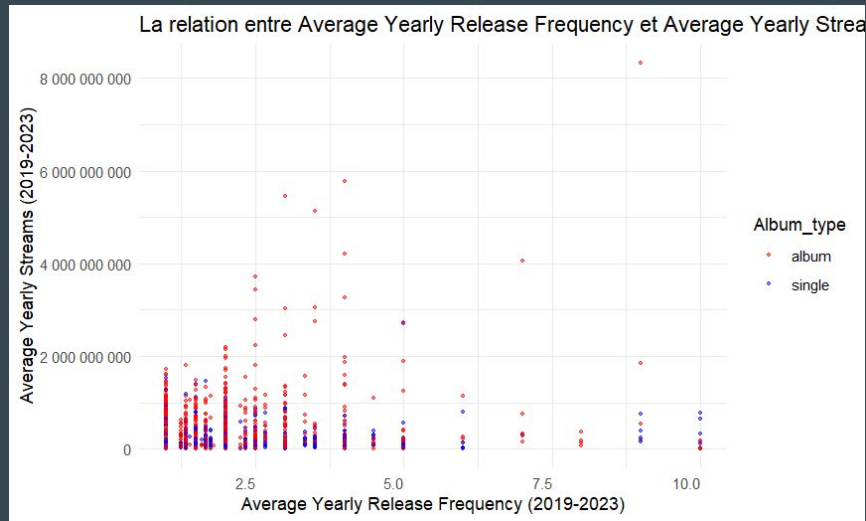
Répartition des attributs de chanson en 1975



LASD

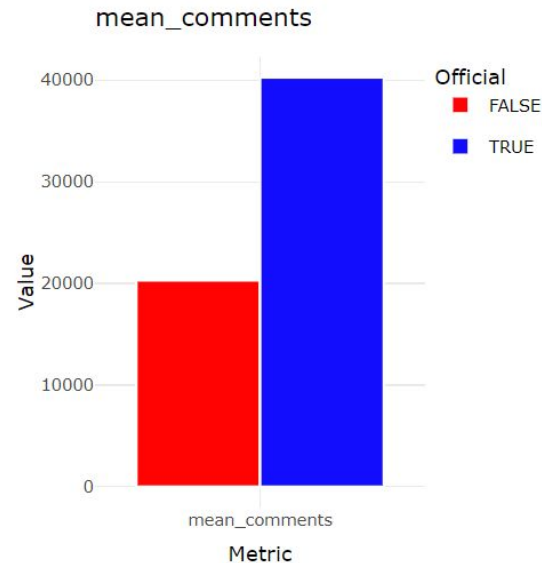
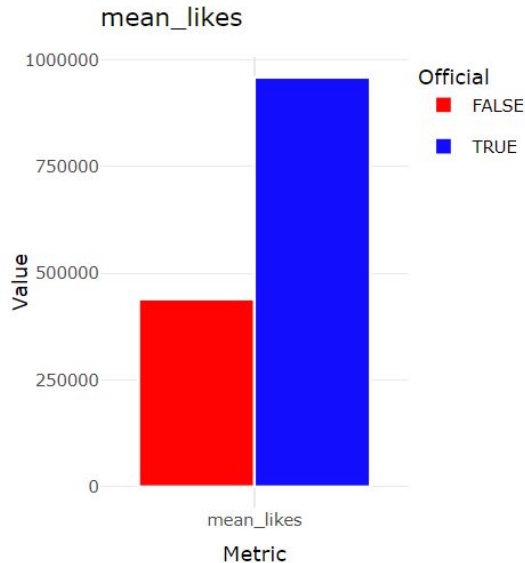
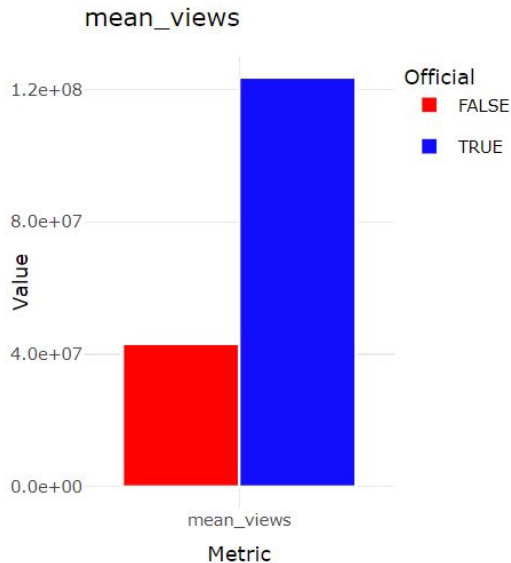
Répartition des attributs de chanson en 2012



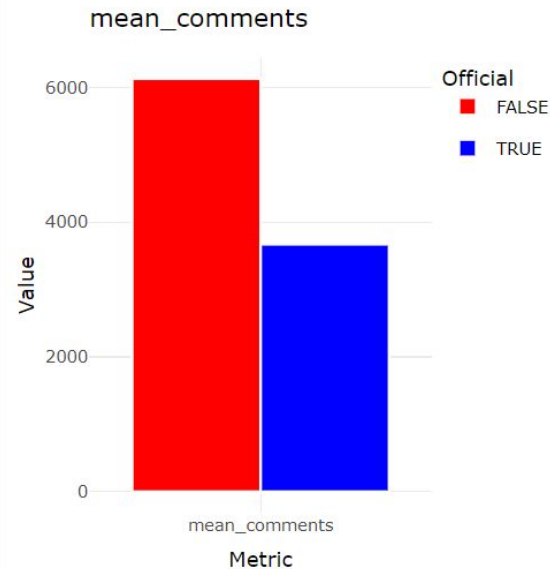
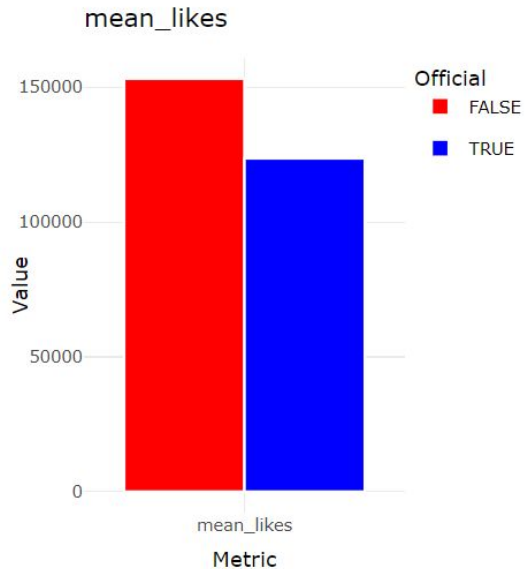
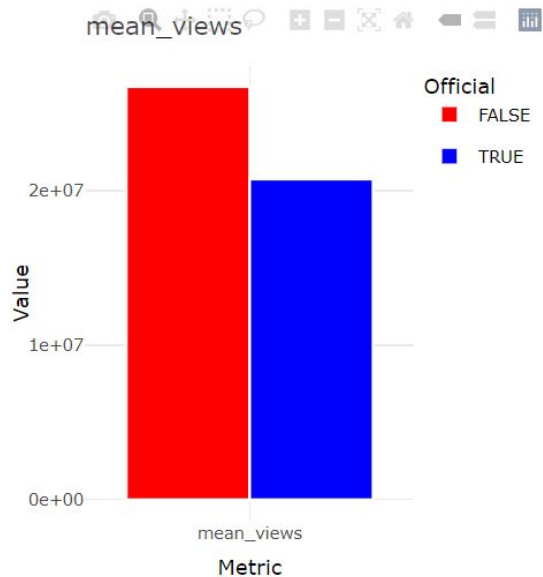


Existe-t-il une relation entre le volume annuel moyen des œuvres publiées et leur diffusion et visualisation ? La forme des œuvres les affecte-t-elle ?

officiel ou non officiel ? qui a plus de succès ?



Parfois ça peut changer...



Étude concentrée sur les attributs

"Danceability", "Energy", "Key", "Loudness", "Speechiness", "Acousticness", "Instrumentalness", "Liveness", "Valence", "Tempo"

