

Description de projet - Groupe 3

Information Visualisation
2024-2025

Avec :

Salma TALIB

Krysto DAGUES DE LA HELLERIE

Clément CHATELAIN

Lorenzo FROMENT

Marco LACLAVERE



Voici le document présentant notre projet en visualisation d'informations. Vous trouverez ci-dessous cinq propositions de visualisation d'information utilisant le dataset Wasabi. Chacune de ces propositions ont été faite par un membre du groupe différent. On y retrouve le type de visualisation, les utilisateurs ciblés, les Tâches/objectifs de visualisation qu'elles veulent atteindre, les attributs du dataset qui seront nécessaires et quels traitements de données vont devoir être effectués.

- **Visualisation 1** : Style de musique des artistes en fonction de leur lieu de naissance

- **membre du groupe** : Lorenzo FROMENT
- **utilisateurs** : Les utilisateurs ciblés sont les amateurs de musique et les chercheurs souhaitant explorer la distribution géographique des genres musicaux.
- **Tâches visuelles/ objectifs de visualisation** : visualiser les genres musicaux en fonction du pays de naissance des artistes à travers une carte interactive. L'interface permet aux utilisateurs de zoomer sur les pays, de sélectionner des genres spécifiques et de voir des représentations visuelles détaillées des genres par région.

1. **Exploration géographique** : Utilisation de la carte interactive pour visualiser la répartition des genres par pays.
2. **Filtrage par genres** : Sélection des genres à afficher pour une exploration ciblée.
3. **Affichage détaillé par pays** : Visualisation des répartitions de genres via des diagrammes circulaires superposés sur la carte, permettant aux utilisateurs de cliquer et d'obtenir des détails supplémentaires.

- **Attribut brut nécessaires** : `location.country`, `dbp_genre`
- **Traitement des données** :
Fichier "genreTraitement.js", fonction "groupGenresByCategory" => Il y a eu un premier gros traitement qui a consisté à récupérer tous les genres musicaux et à les classer dans des groupes plus généraux car ils étaient extrêmement nombreux et cela n'aurait donc pas été adéquat pour des pie chart (illisible).
Pour faire ce regroupement j'ai utilisé cette source: [List of music genres and styles - Wikipedia](#)

Fichier "gatherData.js", fonction "extractEssentialInfo" => Ensuite un second traitement a visé à récupérer les pays de naissances des artistes et créer un json regroupant cette information ainsi que le genre musical.

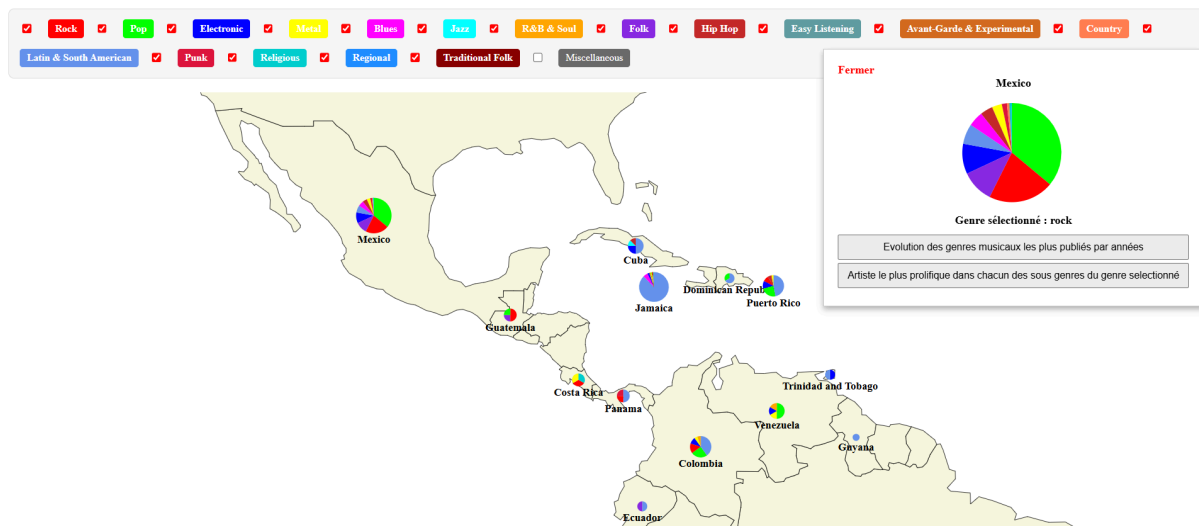
Fichier "countGenrePerCountry.js", fonction "aggregateGenresByCountry" =>

Pour terminer, ces données ont été agrégées par pays dans le fichier countGenrePerCountry.js, où le nombre d'occurrences de chaque genre est comptabilisé pour chaque pays.

Tout ce traitement est effectué en une seule fois lors du lancement du fichier dataCleaning.js.

- **Technique de visualisation** : La technique utilisée est une carte interactive SVG avec des diagrammes circulaires (camemberts) générés en D3.js, montrant la répartition des genres dans chaque pays. Les couleurs spécifiques sont attribuées par genre, et les utilisateurs peuvent cliquer sur chaque diagramme pour obtenir des détails supplémentaires sur les genres dans le pays sélectionné.

Exemple :



- **Cartographie des variables visuelles** : Les variables traitées (country, genres) sont mappées sur une carte et des diagrammes circulaires en D3.js pour montrer la distribution par pays, ce qui offre aux utilisateurs une exploration intuitive de la répartition géographique des genres musicaux.

- **Visualisation 2** : Visualisation de **l'artiste le plus prolifique** dans chaque catégorie musicale (genre).

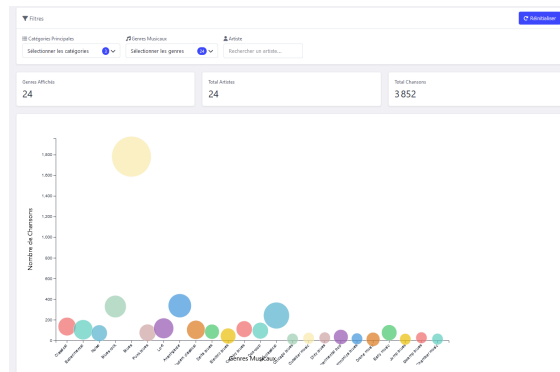
- **membre du groupe** : Salma Talib
- **utilisateurs** :
 - Amateurs de musique
 - Analystes musicaux
 - Plateformes de streaming
 - Chercheurs en musique
- **Tâches visuelles/ objectifs de visualisation** : Pouvoir voir quel est l'artiste qui a le plus produit de musique ou d'albums pour chaque genre musical.
- **Attribut brut nécessaires** : `id_artist` et `artist_name`, genre de la table **Albums**, `id_album` ou `albumTitle` de la table **Songs**.
- **Traitement des données** : Dans le script `generateGenreStats.js`, pour chaque musique, on regarde à quel album il appartient. Sur cet album, on regarde son genre musical et l'artiste qui l'a créé, avec un counter on compte le nombre de chansons et d'albums par artiste, de cette manière on peut obtenir le nombre de musiques créées pour chaque artiste par genre musical. A partir de cela on crée un nouvel `artiste_genre_stats.json` qui contient notre donnée traitée.
On a ajouté catégorie qui contient différents genre de wasabi afin de pouvoir classer les 399 genres en catégorie pour une bonne visualisation.
- **Technique de visualisation** : Bubble chart
Le script `bubbleChart.js` prend en charge la visualisation en utilisant la bibliothèque `D3.js`, en créant un graphique à bulles où chaque une représente un artiste, avec des dimensions proportionnelles au nombre d'album.
L'affichage des données peut être ajusté à l'aide de filtres, notamment :

- **Catégorie** (première couche)
- **Genre musical**, qui représente un sous-genre de la catégorie
- **Nom d'artiste**, permettant une sélection plus spécifique

À chaque étape, les informations suivantes sont affichées :

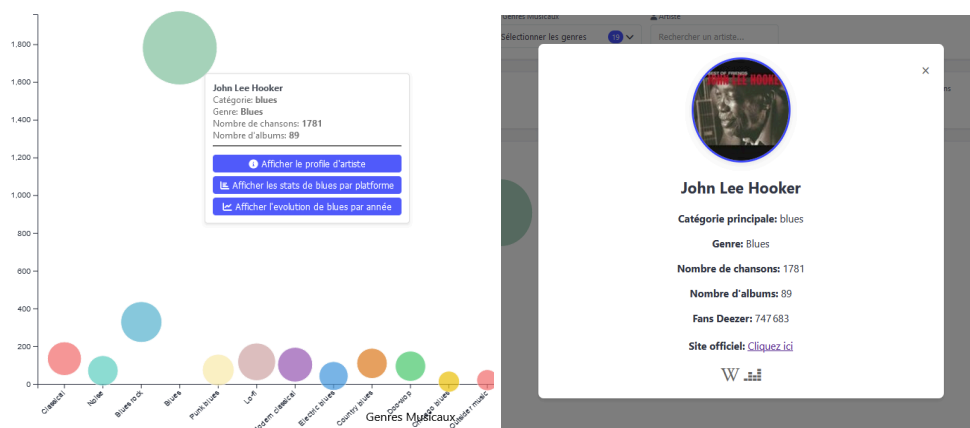
- Le **nombre total d'artistes** correspondant aux filtres appliqués
- Le **nombre de genres musicaux** affichés
- Le **nombre total de chansons**

Chaque fois qu'un filtre est modifié, la page est mise à jour automatiquement pour refléter les nouvelles données. De plus, un bouton de **réinitialisation** permet d'afficher toutes les catégories, genres et artistes disponibles en supprimant les filtres appliqués.



La redirection peut se faire sous trois formes principales :

- Affichage des statistiques du blues par plateforme (visu 3)
- Affichage de l'évolution du blues par année(visu4)
- Affichage du profil d'un artiste



- **Cartographie des variables visuelles** : En axe X : Représente les différents genres musicaux. En axe Y : Représente le nombre total d'albums ou de chansons créées par l'artiste. La taille des bulles sera proportionnelle au nombre d'albums produits par l'artiste dans chaque genre.

- **Visualisation 3** : Évolution des genres musicaux les plus publiés par année.
 - **Membre du groupe** : Clément Chatelain
 - **Utilisateurs** : Les analystes de l'industrie musicale, les maisons de disques, et les historiens de la musique qui souhaitent étudier l'évolution des tendances musicales au fil des années, pour comprendre quels genres ont dominé différentes périodes et quelles tendances émergent.
 - **Tâches visuelles/ objectifs de visualisation** : Visualiser quels genres musicaux ont été les plus populaires ou publiés chaque année pour repérer les tendances dominantes, les genres émergents, ou le déclin d'autres genres au fil du temps.
 - **Attributs bruts nécessaires** : `genre` , et `releaseDate` de la table `Album`
 - **Traitement des données** :

Étape 1 : Préparation des genres

1. **Point de départ** : fichier `album.json`.
2. **Script** : `albumGenresTraitement.js`
 - **Méthode utilisée** : `groupGenresByCategory` (exportée pour être exécutée dans `dataCleaning.js`).
 - **But** : Catégoriser et simplifier les genres pour correspondre aux nouvelles catégories (utilisation du même traitement que Lorenzo mais pour les albums cette fois-ci).
3. **Résultat** : fichier `album-with-new-genres.json`.

Étape 2 : Extraction des données essentielles

1. **Point de départ** : fichier `album-with-new-genres.json`.
2. **Script** : `gatherAlbumData.js`
 - **Méthode utilisée** : `extractEssentialInfo` (également exportée pour être exécutée dans `dataCleaning.js`).
 - **But** : Réduire les informations au minimum nécessaire (titre, genre, et date de parution).
3. **Résultat** : fichier `album-simplified.json`.

Étape 3 : Comptage des albums par genre et par année

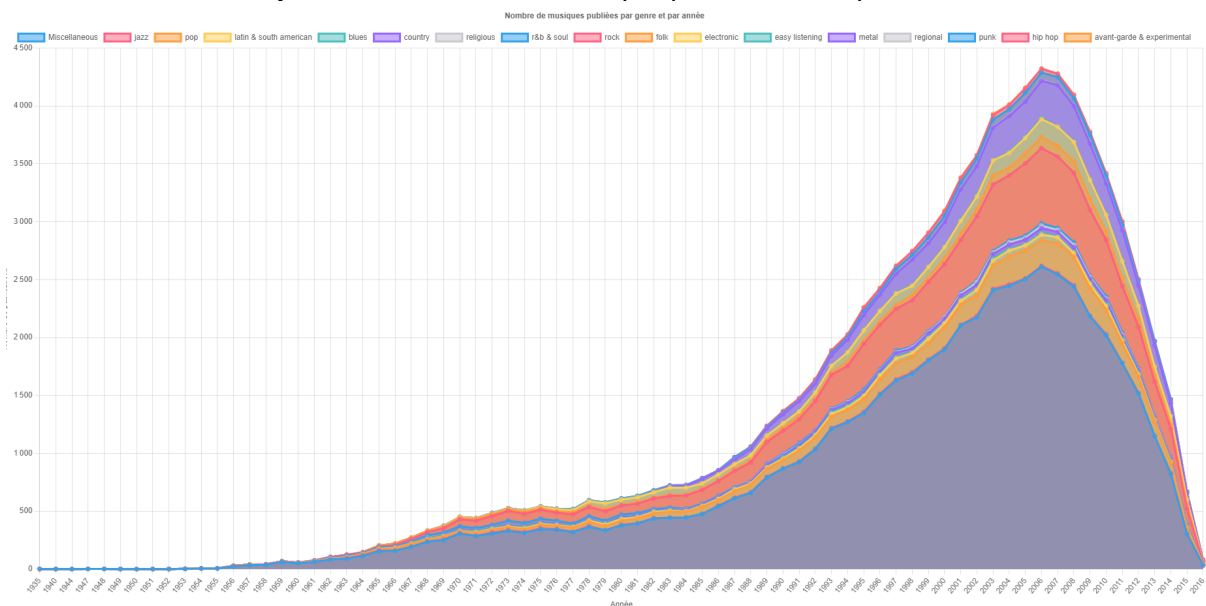
1. **Point de départ** : fichier `album-simplified.json`.
2. **Script** : `script.js`
 - Charger le fichier JSON.
 - Extraire l'année de chaque album et le genre associé.
 - Compter le nombre d'albums pour chaque combinaison d'année et de genre, en excluant les années invalides ("`0000`" ou données non numériques).

- Sauvegarder les résultats sous forme de comptages annuels par genre dans un nouveau fichier.
3. **Résultat** : fichier `year-genre-counts.json` contenant les comptages par année et par genre

Récapitulatif

- **Scripts principaux :**
 1. `dataCleaning.js` (pour exécuter les méthodes `groupGenresByCategory` et `extractEssentialInfo`).
 2. `script.js` (pour le traitement final des années et genres).
- **Fichiers générés à chaque étape :**
 1. `album-with-new-genres.json`
 2. `album-simplified.json`
 3. `year-genre-counts.json`

- **Technique de visualisation** : Graphique en aire empilée

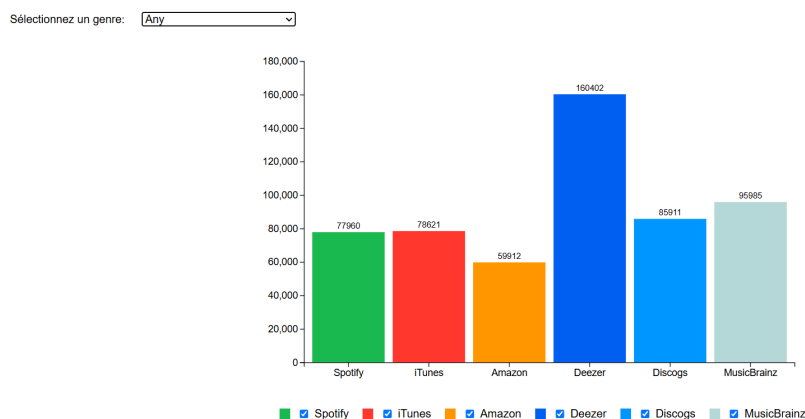


- **Cartographie des variables visuelles :**
 - **En axe X** : les années (en fonction de `releaseDate`).
 - **En axe Y** : le nombre d' albums publiés.
 - Chaque genre est représenté par une aire montrant l'évolution des genres dominants au fil des ans

- **Visualisation 4** : plateformes proposant le plus d'albums à l'achat ou à l'écoute
 - **membre du groupe** : Krysto Dagues de La Hellerie
 - **utilisateurs** : Des chercheurs/analystes en marketing qui souhaitent comprendre la stratégie de distribution des albums.
 - **Tâches visuelles/ objectifs de visualisation** : But : Avoir une répartition des albums disponibles sur chaque plateformes et pouvoir comparer les différentes plateformes en fonction du nombre d'albums qui y sont disponibles.
On peut trier les albums par genre musical, et filtrer les plateformes qu'on veut afficher ou non.
 - **Attribut brut nécessaires** : `urlITunes urlDeezer urlSpotify urlAmazon urlDiscogs urlMusicBrainz` de la table **Albums**
 - **Traitement des données** :
Scripts de traitement : `Clement_Vlsu/gatherAlbumData.js` (modifié) et `./Krysto_Visu/getPlatformStats.js`

On modifie le script de génération des données simplifiées pour les albums (fonction `extractEssentialInfo` dans `gatherAlbumData.js`, qui va permettre de générer `album-simplified.json`) de Clément pour y ajouter les plateformes, et on y ajoute un attribut `True` ou `False` pour chaque plateforme en fonction de si l'attribut est présent dans les données de base ou non. A partir de là, on génère des données (`getPlatformStats.js`) sur le nombre d'albums pour chaque genre, en prenant soin de mettre un genre "any" englobant tous les genres. Les statistiques sur les albums par plateforme sont ensuite stockées dans `stats-platform.json`.

- **Technique de visualisation** : Graphique à barres



- **Cartographie des variables visuelles** : En axe X : le nom de chaque plateforme. En axe Y : le nombre d'albums disponible sur la plateforme. Il y a donc une barre pour chaque plateforme dont la hauteur varie en fonction du nombre d'albums disponibles.

- **Visualisation 5:** Analyse du nombre d'artistes que possèdent les maisons de disques par genres musicaux.

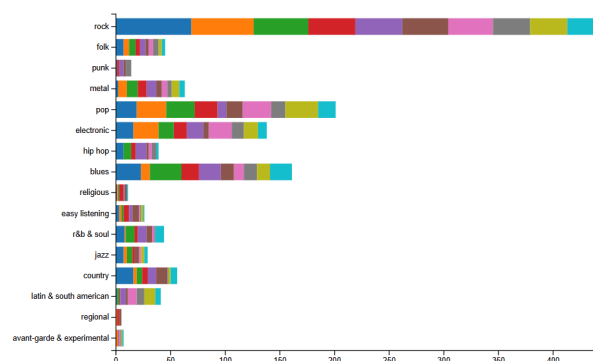
- **membre du groupe:** Marco Laclavere
- **utilisateurs:** les artistes cherchent une maison de disques ou voulant voir quels sont les genres de musiques les plus produits, les maisons de disques voulant comparer leurs nombres de publications avec d'autres maisons, des analystes ou des chercheurs afin d'en ressortir des données à utiliser dans leurs travaux.
- **Tâches visuelles/ objectifs de visualisation:** But : Visualiser la répartition des artistes pour chaque maison de disque par genre musical. Identifier les genres de musique les plus publiés au sein d'une maison de disque ou parmi toutes les maisons de disques.

On peut filtrer par plateforme et choisir si on veut cacher ou non le genre "miscellaneous" et afficher ou non seulement les 10 premières maison de disque.

Nous pouvons aussi filtrer le genre de musique "miscellaneous" et le recordlabel "Other". Car dans les deux cas ce sont des classes très majoritaires par rapport aux autres données et nous souhaitons à la fois avoir des données exploitables mais aussi permettre à l'utilisateur de voir l'ensemble de données dans son entièreté.

- **Attributs brut nécessaires:** recordlabel et genre de la table **Artiste**.
- **Traitement des données:** On reprend le json sur les artistes généré pour la visualisation 1 (gatherData.js, fonction extractEssentialInfo) et on y ajoute l'attribut recordlabel contenant la/les maison(s) de disques avec lequel l'artiste a travaillé (si aucune valeur n'est présente on met cette valeur à "Other"). De plus nous rajoutons aussi les plateformes sur lesquelles on peut retrouver l'artiste (True ou False pour chaque plateforme). De cette manière on est capable de compter le nombre d'artistes pour chaque maison de disques, par genre (attribut qui était déjà présent), et de les trier par plateforme.
- **Technique de visualisation:** Stacked bar graph

Sélectionnez une plateforme : ☒ Cacher le genre "Miscellaneous" ☒ Afficher seulement les 10 premiers recordLabels



- **Cartographie des variables visuelles:** Cette technique de visualisation permet à la fois de voir pour chaque genres musicaux quels sont les maisons de disques les plus populaires ainsi que les genres musicaux les plus populaires
 - axe X: Les genres musicaux.
 - axe Y : le nombre d'artistes pour chaque maison de disques.

LIAISON DES VISUALISATIONS

Visu 1 => lié avec Visu 2. Quand on a l'évolution des styles musicaux on peut cliquer dessus pour voir les artistes les plus prolifiques dans ce style musical.

Visu 1 => lié avec Visu 3. On peut à partir d'un genre sur un diagramme passer sur l'évolution du genre par année.

Visu 2 => lié avec Visu 3. Si on clique sur un artiste avec un genre musical (par exemple "rock"), on peut choisir de passer sur l'évolution de genre par année

Visu 2 => lié avec Visu 4. Comme pour le lien précédent, on peut choisir de passer sur la répartition du nombre d'albums par plateforme pour le genre musical voulu.

Visu 3 => lié avec Visu 4. En cliquant sur un des genres, on peut obtenir la répartition des albums du genre musical sélectionné pour chaque plateforme.

Visu 4 => lié avec Visu 5. En cliquant sur une des barres liée à une plateforme, on obtient les barres empilées représentant les artistes par maison de disques pour chaque genre sur cette plateforme.

Répartition du travail

NOM	Travail effectué
FROMENT Lorenzo	Visualisation 1 (carte avec diagrammes circulaires), lien entre la visualisation 1 et 2, lien entre la visualisation 1 et 3, établissement du mapping des genres, application du mapping des genres pour les artistes
TALIB Salma	Visualisation 2 (Bubble Chart), lien entre la visualisation 2 et 3, lien entre la visualisation 2 et 4, traitement statistiques sur les artistes par genre
CHATELAIN Clément	Visualisation 3 (Graphe en aire empilées), traitement sur les albums, (retrait des doublons et application du mapping des genres)
DAGUES DE LA HELLERIE Krysto	Visualisation 4 (Graphe à Barres), traitement sur les plateformes pour les albums pour obtenir des statistiques, lien entre la visualisation 3 et 4, lien entre la visualisation 4 et 5.
LACLAVERE Marco	Visualisation 5 (Graphe à barres empilées), traitement sur les plateformes et les maison de disques pour les artistes