



RAPPORT PROJET

Information Visualisation

Karim **CHARLEUX**
Damien **STENGEL**
Wassim **BARATLI**
Elias **MERDACI**
Romain **PELLEGRINI**

12 Novembre 2024 - Université Côte d'Azur - Polytech Nice

Sommaire

I. Introduction.....	3
1. Description des utilisateurs.....	3
2. Liste des tâches visuelles supportées par les utilisateurs et les objectifs de visualisation.....	3
3. Objectifs de visualisation.....	4
4. Workflows.....	4
II. Les différentes visualisations.....	5
1. Graphe de réseau.....	5
1.1 Aperçu.....	5
1.2 Objectif du graphe / Technique de visu.....	5
1.3 Données utilisées.....	6
1.4 Traitement des données.....	6
1.5 Données sorties et filtres.....	6
2. Tree.....	7
2.1 Aperçu.....	7
2.2 Objectif du graphe / Technique de visualisation.....	7
2.3 Données utilisées.....	7
2.4 Traitement des données.....	8
2.5 Données sorties et filtres.....	8
3. Graphique Stacked Area Chart.....	10
3.1 Aperçu.....	10
3.2 Objectif du graphe.....	11
3.3 Données utilisées.....	11
3.4 Traitement des données (traitement.py).....	11
3.5 Fonctionnalités et interaction.....	11
3.6 Données sorties et filtres.....	12
4. Carte Choroplèthe.....	12
4.1 Aperçu.....	12
4.2 Objectif du graphe / Technique de visualisation.....	12
4.3 Données utilisées.....	13
4.4 Traitement des données.....	13
4.5 Données sorties et filtres.....	13
5. Box diagram.....	14
5.1 Aperçu.....	14
5.2 Objectif du graphe.....	14
5.3 Données utilisées.....	15
5.4 Traitement des données.....	15
5.5 Données sorties et filtres.....	15
III. Conclusion.....	16

I. Introduction

1. Description des utilisateurs

Les utilisateurs de cette visualisation sont des **professionnels de la musique**, tels que les producteurs et les chercheurs en musicologie, ainsi que des **amateurs de musique** qui s'intéressent aux dynamiques entre genres musicaux. Ces utilisateurs cherchent à explorer des **tendances musicales**, à comprendre les **relations entre genres**, et à observer des **facteurs influents** comme la localisation géographique et les périodes historiques d'activité.

2. Liste des tâches visuelles supportées par les utilisateurs et les objectifs de visualisation

Les utilisateurs peuvent accomplir les tâches suivantes :

- **Identifier** : Reconnaître les genres musicaux dominants dans le réseau des genres.
- **Visualiser** : Voir l'arbre hiérarchique des genres.
- **Comparer** : Analyser les différences entre genres en fonction de la longévité et de la distribution géographique.
- **Associer** : Établir des liens entre genres musicaux et d'autres facteurs comme l'année ou le pays d'origine.
- **Catégoriser** : Organiser les artistes selon leur genre.
- **Corréler** : Étudier les relations entre genres et la durée moyenne des carrières.
- **Observer les tendances** : Analyser l'évolution des genres au fil du temps

3. Objectifs de visualisation

Les objectifs de cette visualisation sont :

- Offrir une vue d'ensemble des genres musicaux et de leurs connexions.
- Représenter l'évolution temporelle de certains genres musicaux.
- Montrer la répartition géographique des genres musicaux.
- Visualiser la relation entre genres et longévité des carrières artistiques.
- Visualiser les genres parents et sous genre d'un genre donné

4. Workflows

Notre site propose 2 scénarios :

Le premier scénario se focalise sur les genres. En cliquant sur nœud, cela va donner accès au Tree avec le genre sélectionné et ses sous-genre. Nous allons pouvoir aller encore plus dans le détail en allant sur la Choropleth Map qui montre la distribution géographique des genres. Il est possible de retourner au graphique précédent tout en gardant le genre sélectionné.

Le second scénario va quant à lui rajouter la date dans les paramètres qui sont passés entre les visualisations. Cette fois-ci, en cliquant sur un nœud, on va pouvoir se rendre au stack area graph. Cette visualisation va nous montrer l'évolution des genres selon la date sélectionnée et les artistes et albums populaires. Il est possible ensuite de se spécialiser pour aller observer la durée des carrières des différents artistes selon les paramètres définis au préalable. Il est possible de retourner au graphique précédent tout en gardant le genre et la date sélectionnée.

Voir annexe (page 16).

connexions grâce à la bibliothèque D3.js. Les utilisateurs peuvent ainsi naviguer dans les relations inter-genres, comprendre la proximité culturelle ou stylistique, et analyser l'intensité des connexions entre différents genres.

1.3 Données utilisées

Attributs utilisés:

- artist-without-members.json : id_artist_deezer/name
- song.json : id_artist_deezer/album_genre/publicationDate

1.4 Traitement des données

Le traitement des données inclut plusieurs étapes clés pour structurer les données en vue de leur utilisation dans la visualisation de réseau :

- **Nettoyage des genres** : Les noms de genres sont normalisés, avec conversion en minuscules et suppression de caractères spéciaux.
- **Calcul des connexions entre genres** : Une matrice de similarité est construite en identifiant les artistes ayant des productions dans plusieurs genres. Chaque lien est pondéré en fonction de la fréquence de production multi-genres.
- **Filtrage temporel** : Les genres peuvent être filtrés par période et par localisation pour pouvoir le filtre par date sur la visualisation.

1.5 Données sorties et filtres

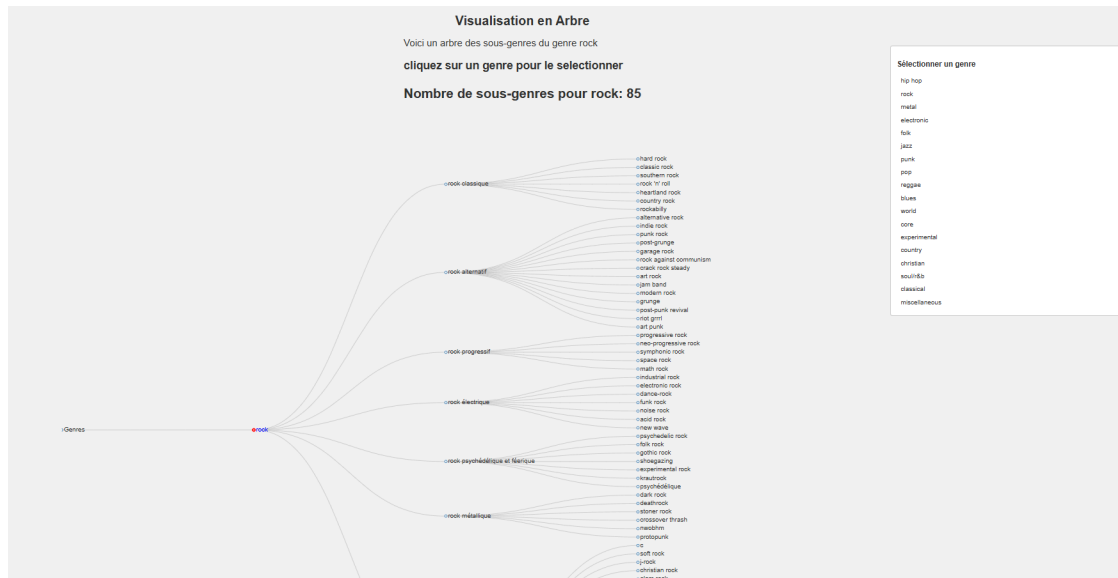
Les données finales utilisées pour le graphe sont structurées en objets JSON, avec :

- **Nœuds** : Chaque nœud représente un genre, dont la taille est proportionnelle au nombre de chansons associées.
- **Liens** : Les liens entre nœuds symbolisent les connexions entre genres, où l'épaisseur indique l'intensité de la relation, basée sur le nombre d'artistes communs.
- **Filtres et paramètres** : L'interface permet aux utilisateurs de définir des paramètres pour ajuster l'affichage, incluant :
 - **Filtre de genre central** pour se concentrer sur un genre particulier.
 - **Profondeur des connexions** pour afficher des relations plus ou moins éloignées.
 - **Nombre minimum de connexions** pour filtrer les liens selon leur intensité.
 - **Période d'activité** définie par des filtres de début et de fin d'année.

2. Tree

Responsable : Elias MERDACI

2.1 Aperçu



2.2 Objectif du graphe / Technique de visualisation

L'objectif de ce graphe est de structurer et visualiser l'organisation hiérarchique des genres et sous-genres musicaux. Le but est de représenter la relation entre chaque genre et sous-genre sous la forme d'un Tree, permettant de naviguer facilement à travers les différents niveaux de classification.

Cette hiérarchie débute avec le genre principal, qui se subdivisent en sous-genres, chacun pouvant contenir un ensemble de sous genres et ainsi de suite. Ce graphe facilitera l'exploration des genres musicaux en montrant comment les genres ont évolué dans une structure organisée, de sorte que l'utilisateur puisse identifier cette évolution en un coup d'œil.

2.3 Données utilisées

La donnée utilisée est l'attribut genre contenu dans les sons répertorié dans le [song.json](#). En plus de cet attribut on utilise aussi un sous-genre.json contenant une liste des genres et de leur sous-genre trouvée sur internet qu'on doit créer et que l'on va formater ensuite

2.4 Traitement des données

Après avoir récupéré tous les genres contenus dans les sons , je dois ensuite formater le tout afin de convertir les caractères spéciaux en leur version

alphabétique , les entités HTML en des caractères normés et les majuscules en minuscules afin d'avoir une liste de genre traités contenue dans genre.json

Par exemple "Country Pop" devient "country pop"

Ensuite cette liste de genre est comparée avec la liste de sous-genre créer préalablement avec des informations trouvées sur Internet afin de vérifier que tous les genres des sons sont contenues dans la liste des sous-genre et si non les ajoutés avec leur sous-genres et/ou genre parent, assurant ainsi la communication avec les autres visualisations car chaque genre qu'il pourrait m'envoyer est forcément dans le json des sous-genre après le traitement.

Voici un exemple de la structure du sous-genre.json :

```
{
  "genre": "jazz",
  "subgenres": [
    {
      "genre": "jazz traditionnel",
      "subgenres": [
        "ragtime",
        "swing",
        "new orleans rhythm and blues",
        "boogie woogie",
        "gypsy jazz"
      ]
    },
    {
      "genre": "jazz moderne",
      "subgenres": [
        "smooth jazz",
        "contemporary jazz",
        "soul jazz",
        "cool jazz",
        "hard bop",
        "bebop",
        "free jazz",
        "latin jazz",
        "jazz fusion",
        "vocal jazz",
        "nu jazz",
        "acid jazz"
      ]
    },
    {
      "genre": "jazz électrique et fusion",
      "subgenres": [
        "jazz-funk",
        "indietronica",
        "trip hop"
      ]
    }
  ]
}
```

2.5 Données sorties et filtres

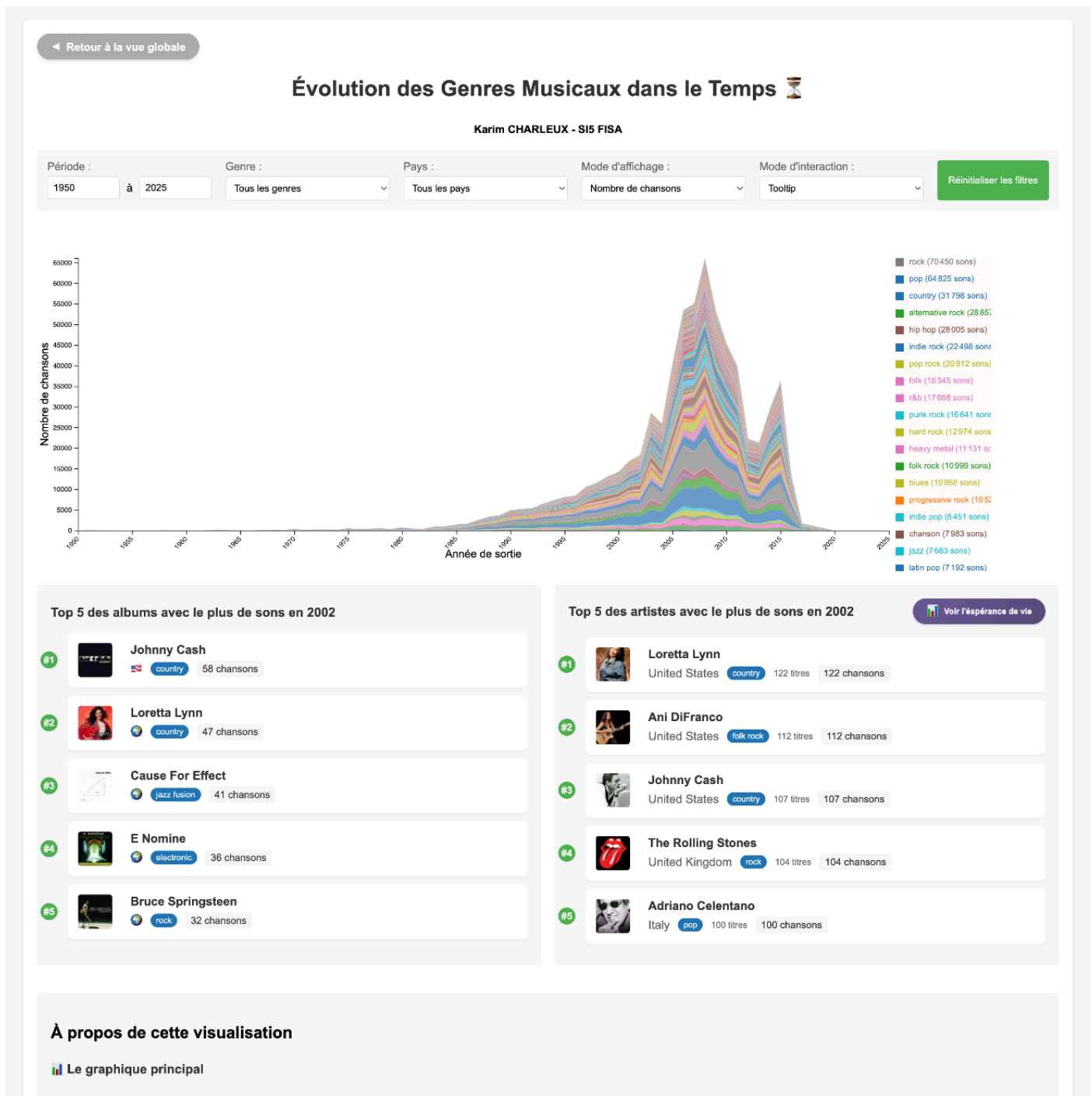
Les données finales utilisées pour l'arbre de genres musicaux sont structurées en objets JSON, avec :

- **Nœuds** : Chaque nœud représente un genre ou sous-genre musical
- **Liens** : Les liens entre les nœuds symbolisent les relations hiérarchiques entre les genres et sous-genres
- **Filtres et paramètres** :
 - L'utilisateur peut cliquer sur les nœuds de l'arbre pour naviguer dans l'arborescence des genres et choisir de se focaliser sur un nœud en filtrant ainsi les données à seulement les sous-genres et genre parents du genre sélectionné.

3. Graphique Stacked Area Chart

Responsable : Karim CHARLEUX

3.1 Aperçu



3.2 Objectif du graphe

Le Stacked Area Chart est une visualisation interactive qui permet de suivre l'évolution temporelle des genres musicaux entre 1950 et 2024. Cette visualisation offre une vue d'ensemble des tendances musicales au fil du temps, permettant aux utilisateurs d'observer la popularité relative des différents genres musicaux et leur évolution. En bas de la page on peut retrouver toutes les descriptions des visualisations présentes et les interactions possibles.

3.3 Données utilisées

album.json Détails des albums	artist-without-members.json Informations sur les artistes	song.json Informations sur les chansons
<ul style="list-style-type: none">• _id• id_artist• country• publicationDate• cover > small	<ul style="list-style-type: none">• _id• name• country• picture > small	<ul style="list-style-type: none">• _id• publicationDate• album_genre• rank (popularité)• id_album

3.4 Traitement des données (traitement.py)

1. **Nettoyage des données :**
 - a. Normalisation des noms de genres
 - b. Conversion des caractères spéciaux
 - c. Filtrage des données invalides ou incomplètes
2. **Agrégation des données :**
 - a. Regroupement par année et par genre
 - b. Calcul des métriques :
 - i. Nombre total de chansons
 - ii. Score de popularité moyen
 - iii. Statistiques par artiste et par album
3. **Structuration des données :**
 - a. Organisation hiérarchique année > genre > statistiques
 - b. Création des liens entre artistes, albums et genres

3.5 Fonctionnalités et interaction

Mode d'affichage : 2 choix soit par nombre de sons ou par popularité

Mode d'interaction :

- Mode Tooltip : affichage détaillé au survol
- Mode Zoom/Brush : sélection de périodes spécifiques

Top 5 dynamique : Top des albums et artistes en fonction du mode d'affichage

Légende interactive : Au hover d'un élément de la légende on voit l'area correspondante au genre et au clique on peut passer à la prochaine visualisation.

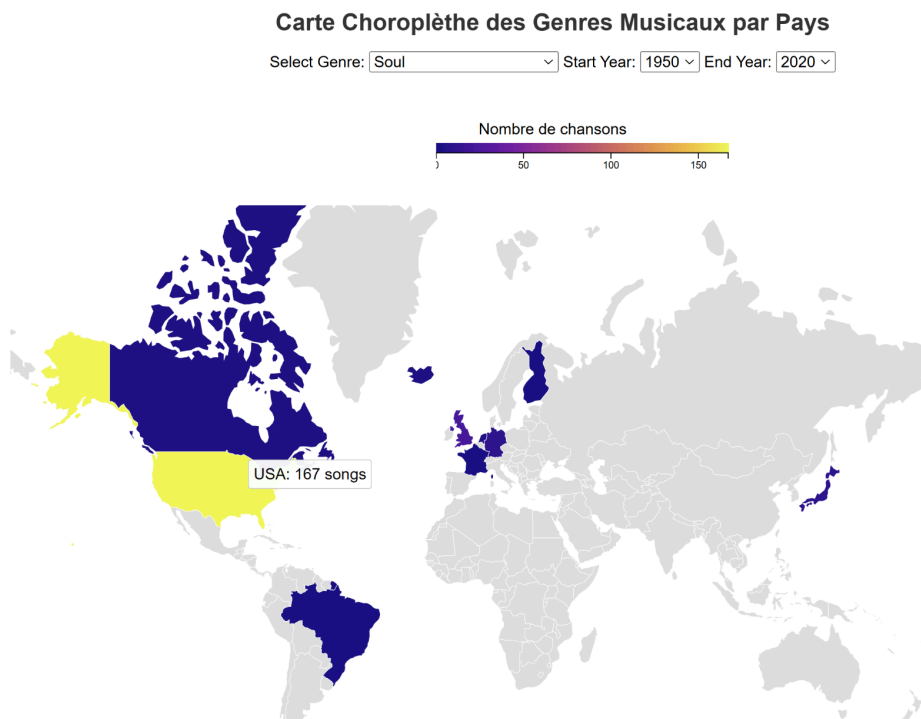
3.6 Données sorties et filtres

- Fichier de sortie : genre_evolution.json
- Axe X : Années.
- Axe Y : Fréquence des genres musicaux.
- Couleurs : Pour chaque genre.
- Filtres disponible : temporels (année début/fin), genre, pays

4. Carte Choroplèthe

4.1 Aperçu

Responsable : Wassim BARATLI



4.2 Objectif du graphe / Technique de visualisation

Objectif : Représenter visuellement la distribution géographique des genres musicaux dans différents pays.

Technique de visualisation : Carte choroplèthe interactive avec une échelle de couleurs pour illustrer soit le genre prédominant dans chaque pays, soit une mesure de diversité des genres. Des interactions supplémentaires permettent d'explorer les données de manière dynamique.

4.3 Données utilisées

Les données utilisées proviennent d'un fichier JSON contenant des informations sur les albums. Les attributs utilisés incluent :

- **Country** : Le pays d'origine de l'artiste ou du groupe.
- **Genre** : Le genre musical de chaque album.
- **Publication Date** : La date de sortie de chaque album.

4.4 Traitement des données

1. Importation des données

- **Entrée** : Fichiers JSON contenant les données des albums (fichier `albums.json`).
- **Traitement** : Lecture des fichiers JSON et extraction des attributs nécessaires (pays, genre, date de publication).
- **Sortie** : Données d'albums prêtes pour un traitement plus approfondi, incluant les informations sur le pays, le genre, et l'année de sortie.

2. Traitement des genres et des dates

- **Entrée** : Genre musical et date pour chaque album.
- **Traitement** : Nettoyage et standardisation des genres musicaux (suppression des espaces inutiles, conversion en minuscules). Extraction de l'année et normalisation des dates pour ne retenir que celles dans la plage d'intérêt (ex : 1950 à 2022).
- **Sortie** : Liste de genres nettoyée et standardisée avec l'année de publication.

3. Agrégation des données pour la visualisation

- **Entrée** : Données traitées incluant pays, genre, et année de publication.
- **Sortie** : Fichier JSON final contenant les données agrégées, structurées par `country > genre > year` avec les comptages des albums.

4.5 Données sorties et filtres

Régions : Chaque pays est représenté individuellement sur la carte.

Couleurs :

- **Genre dominant** : Les pays sont colorés selon le genre le plus représenté (ou dominant) parmi les artistes originaires de chaque pays.
- **Diversité des genres** : Les couleurs peuvent représenter une mesure de la diversité des genres par pays, illustrée par une variation de nuance ou d'intensité selon le nombre de genres présents.

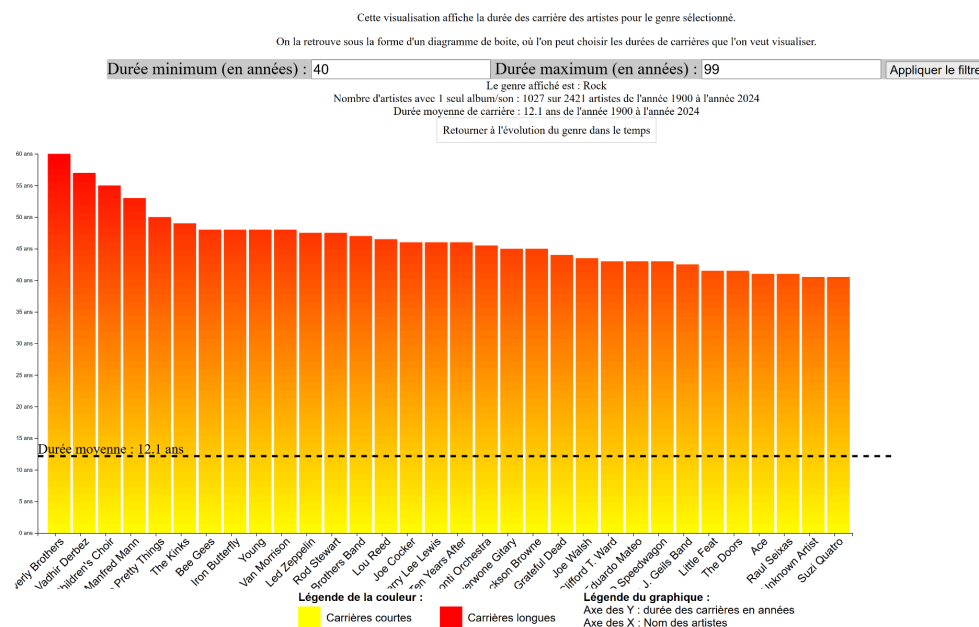
Filtres et Interactions :

- **Filtres** : Possibilité de filtrer par genre et par année pour explorer des périodes spécifiques ou se concentrer sur des genres musicaux particuliers.
- **Tooltips** : Affichage de détails supplémentaires lorsque l'utilisateur survole un pays, incluant le nombre d'albums par genre et par année.
- **Échelle de couleurs** : Permet de visualiser la répartition relative des genres musicaux entre les pays en fonction des données agrégées.

5. Box diagram

Responsable : Romain Pellegrini

5.1 Aperçu



5.2 Objectif du graphe

L'objectif de cette visualisation est de comparer la durée de carrière des artistes selon un genre sélectionné. Le graphique peut aussi se spécialiser selon un intervalle de date définis par les anciennes visualisations. On va pouvoir y retrouver un graphique avec le nom des artistes en abscisse et le nombre d'années de la carrière de l'artiste en ordonnée. La visualisation propose une ligne qui montre la durée moyenne de la carrière des artistes.

La couleur des boîtes va du jaune au rouge. Les carrières les plus longues vont tendre vers le rouge plus la carrière s'approche de la carrière la plus longue du graphique.

On va pouvoir retrouver le nombre d'artistes du genre qui n'ont proposé qu'un seul album ou un seul titre. Une moyenne de la durée des carrières est aussi affichée.

Pour un meilleur affichage, l'utilisateur peut choisir la durée des carrières qu'il souhaite affichée.

5.3 Données utilisées

Attributs utilisés:

- artist-without-members.json : id_artist_deezer/name
- song.json : id_artist_deezer/album_genre/publicationDate

5.4 Traitement des données

Le traitement des données inclut plusieurs étapes clés pour structurer les données en vue de leur utilisation dans la visualisation :

- **Traitement des dates** : cette étape se concentre sur les dates de publication présentes dans les informations des chansons. Le traitement inclut le nettoyage des dates incorrectes (par exemple, celles indiquées comme "0000-00-00").
- **Traitement des artistes** : À partir du fichier `song.json`, les noms des artistes sont récupérés et associés à leurs identifiants dans le fichier `artist-without-members.json`. Si un artiste ne possède pas d'identifiant, il est désigné comme "Unknown Artist".
- **Calcul de la durée de la carrière**, Une fois les listes de dates et d'artistes nettoyées, un fichier JSON est généré. Pour chaque artiste et genre, les champs `first_song` (date de la première chanson) et `last_song` (date de la dernière chanson) sont ajoutés. La durée de la carrière est calculée avec ces données.

5.5 Données sorties et filtres

- **Axe X** : Nom des artistes.
- **Axe Y** : Longévité des carrières.
- **Boîtes** : Distribution des durées de carrière selon les genres.
- **Texte** :
 - Nombre d'artistes n'ayant sorti qu'un album ou titre
 - Durée moyenne d'une carrière d'un artiste
- **Filtres et paramètres** : L'utilisateur peut ajuster :
 - **La durée de la carrière** : le minimum et maximum
 - **L'intervalle de date** : Si l'artiste a sorti un son dans cette intervalle de temps

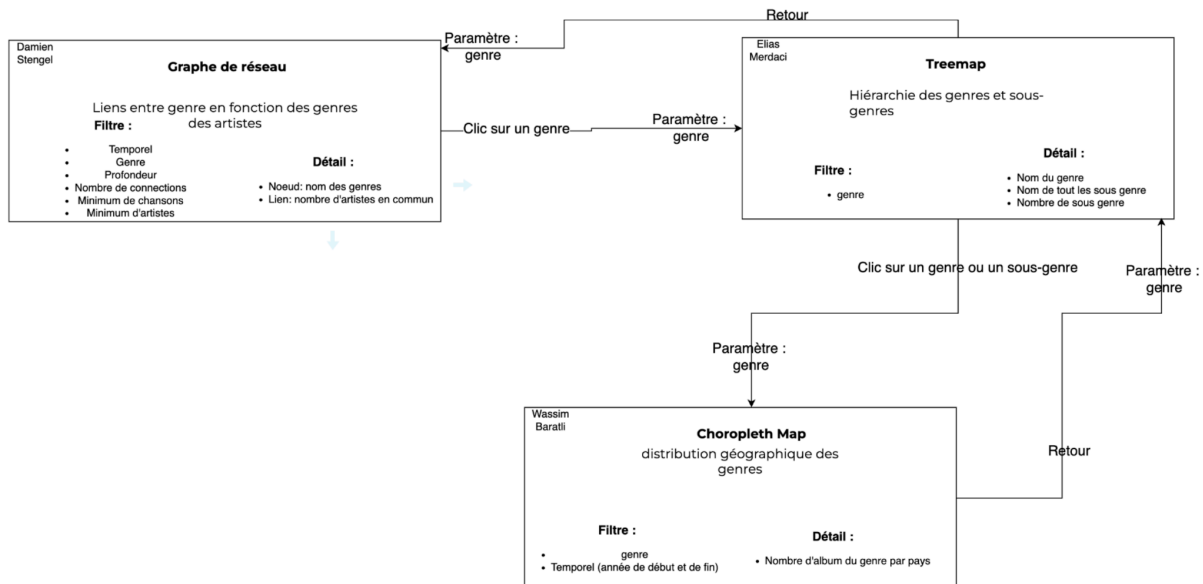
III. Conclusion

Pour conclure, ce projet de visualisation a permis de développer une série d'outils interactifs pour explorer les genres musicaux et leurs dynamiques. Chaque visualisation répond à un besoin spécifique : le graphe de réseau montre les connexions entre genres, l'arbre hiérarchique organise les genres et sous-genres, le graphique Stacked Area Chart suit les tendances dans le temps, la carte choroplèthe montre la répartition géographique, et le diagramme en boîte compare la longévité des carrières.

En combinant ces visualisations, les utilisateurs peuvent naviguer entre les différentes visualisations et ainsi analyser les tendances, identifier des relations entre genres et observer des influences culturelles par pays et par époque.

Annexes

Présentation du workflow des genres



Présentation du workflow des genres et dates

