Data-Visualization Report

CSC_51052_EP

matteodns.github.io/MusicViz-Project

DENIS Matteo

Contents

1 Introduction

Music is a universal language that transcends cultural boundaries, weaving its way into the fabric of human society throughout history. The evolution of music genres has been shaped by a complex interplay of cultural, social, and technological factors, resulting in a diverse array of styles and characteristics that define different periods and regions. Understanding these genres and their unique traits provides valuable insights into the ever-changing landscape of music and its impact on audiences worldwide.

This report explores the characteristics and global trends of music genres using two comprehensive datasets. The first dataset includes detailed information about 15,000 songs spanning from 1950 to the present day, covering a wide variety of genres. Each entry in this dataset contains attributes such as tempo, energy, danceability, and instrumentalness, offering a rich perspective on the distinguishing features of musical styles. The second dataset complements this by providing geographical data on the most popular music genres in different countries, highlighting regional preferences and cultural influences.

The objective of this project is to examine how music genres differ in their attributes and how these differences manifest across regions and time periods. By analyzing the datasets, we aim to uncover patterns and relationships that reveal the unique identities of various genres. For instance, how do genres like jazz, rock, and electronic music compare in terms of energy and tempo? How have these characteristics evolved over decades, and what does this evolution tell us about societal changes? Furthermore, the geographical analysis allows us to investigate the global reach of certain genres and their local adaptations, providing a broader context for understanding the cultural impact of music.

2 Dataset

Two datasets were used for this project, both found on Kaggle, and created by independent users, transforming Spotify and other music platforms' data.

2.1 Dataset n°1: 15,000 classic hits

You can find this dataset on Kaggle: `https://www.kaggle.com/datasets/thebumpkin/10400-classic-hits-10-genres-1923-to-2023`.

For these 15,000 songs, covering a range from the 1950's to today, the dataset covers a wide range of features, including formal information about the song, but also more "fancy" characteristics. Here is a bunch of what can be found:

- Title
- Artist
- Genre
- Danceability
- Duration
- etc...

The dataset divides songs into 19 different genres, but the page only allows 17 to be visualized. Here's why:

- we filter out the genre "World", because it is too vague, and insignificant.
- in the dataset, a genre Today exists, and we can't filter it out as we did with "World", because it takes a really considerable part of the 21st century music. Basically, it is a mixed genre between *Pop*, *Rap* and *EDM*. We also observe that no song after 2000 have the label *Pop* which is quite strange. We could try to create a Machine-Learning classifier to help classify these songs into a specific genre, but as it is not the topic of this class, we decided to change all songs with genre "Today" into *Pop*.

2.2 Dataset n°2: Music genres by country

You can find this dataset on Kaggle: `https://www.kaggle.com/datasets/marshall13302/favorite-music-genres-by-country`.

For each country (72 are available in the dataset, other are displayed in `lightgrey` on the map), the dataset shows the number of songs for each genre that were present in Spotify's 2022 Top 50 (of the country). For example, Australia had 31 songs of the genre "HipHop/Rap/R&B" in its 2022's most listened songs.

It is important to note that the genres are not the same: in this dataset, the songs are only splitted in 6 genres:

- HipHop/Rap/R&B
- EDM (Electro Dance Music)
- Pop
- Rock/Metal
- Latin/Reggaeton (however, no music of this genre was present in the first dataset)

To have compatibility between the two datasets, we had to create a mapping between the genres. Keep in mind while exploring our page that multiple genres may have the same map, because they are mapped to the same genre on the 2nd dataset.

2.3 Combination and preprocessing of the two datasets

Finally, as we didn't use these datasets simultaneously, we did not need to "merge" them. Although, a small preprocessing was necessary for the second dataset: we created a .csv file for each genre, associating to each country its share in its 2022 Top 50. We did this short preprocessing task in Python for simplicity.

3 Structure

The page is longitudinal and must be explored by scrolling from top to bottom. It is divided in one home and 3 visualization sections:

1. Timeline
2. Characteristics
3. World Map

The sections each shows a visualization, according to the music genre selected in the header, in the right-hand top corner of the page. Every time a genre is chosen, these 3 visualizations get updated and the user is brought back at the home page. The user can explore these sections by scrolling, clicking on the arrows or clicking on the header shortcuts.

3.1 Timeline

This section shows the number of song of the selected genre, according to its release year. This can be of great value when showing the "peak time" of a genre.

Here is an example with *Disco*, and its peak in the 80s:

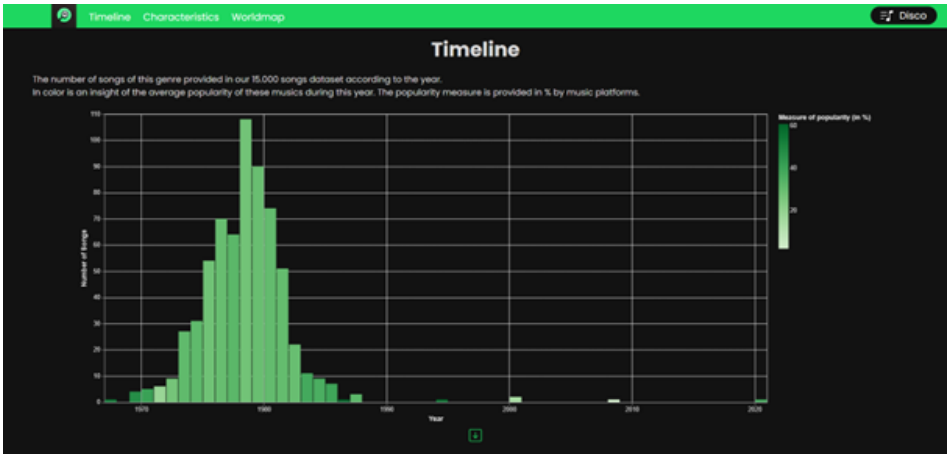


Figure 1: Timeline visualization of *Disco*

As a color encoder, we also have a "Popularity" measure. This is one of the features given by Spotify, and gives a score as a percentage. For this visualization, we compute the average popularity of every songs given the genre and the year. In other terms, there can be years when the genre is not really popular, so the bar would be low, but 2 or 3 songs made it to the great public, so overall popularity would be really high.

3.2 Characteristics

This sections shows multiple features about a genre. Each of these characteristics are again, computed as the average value for a given year (see Part 4.4 for more details). We decided to show:

- Valence
- Energy
- Danceability
- Liveness
- Acousticness
- Instrumentalness
- Speechiness

Here is an example with *Country* music, not really speeechy, but very danceable and acoustic:

Figure 2: Characteristics of *Country*

You will find more details about the visibility of this chart, and the difficulties that we encountered about it in Part 4.4.

3.3 World Map

This section provides a world map, showing where different genres are listened to. In particular the map is based on year 2022 only, with charts provided by Spotify, which covers 72 countries. The exact measure is the proportion of this genre's songs in its Top 50.

Here is an example with *Pop*, globally present, with more than 30% of the Top 50 in almost every northern countries:

Figure 3: Map for *Pop*

As we said in section 2.2, the genre provided here might not be the same than in the previous visualizations.

4 Technical overview

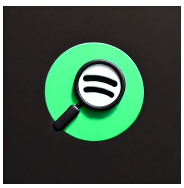
4.1 Implementation tools

The page was created with **D3.js**, mainly for the mapm and **Vega-Lite**, for the charts. We also implemented a strong CSS, for our page to be aesthetically convincing, which was quite a challenge to learn from scratch. The JavaScript and CSS are obviously linked to an HTML page, but we also used a bit of Python for the quick preprocessing of Dataset n°2.

4.2 General design choices

As you may have noticed, we are using the same colors as Spotify, and a similar logo, to emphasize the fact that we are effectively displaying data about music. Most users will immediately interpret these colors as those of Spotify, and thus, to music.

We found the Spotify graphic chart on: `https://developer.spotify.com/documentation/design`, and made the logo and the icon with ChatGPT (Dall-E).



4.3 The page as a whole

We implemented transition buttons to move smoothly from section to section. Fonts and colors are unified, and you can move from section to section either by scrolling,

clicking on the arrows or using the header menu. The user only chooses the genre once, by selecting it in the drop-down menu on the right-hand side. Once the user chooses a genre, every visualization are updated simultaneously and the users is brought back at the top of the page.

4.4 Difficulties for characteristics chart

For better visibility of characteristics, the user can hover over a curve to highlight it. Here is the highlighted "Energy" curve for *Blues*:

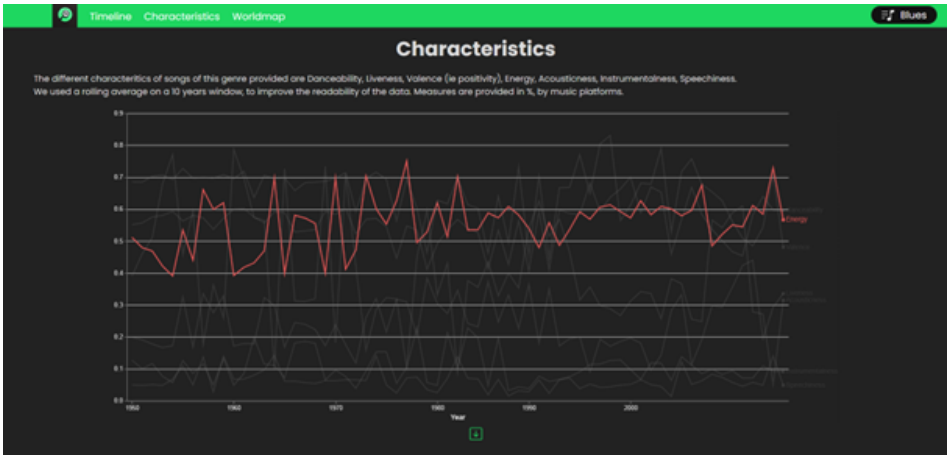


Figure 4: Characteristics of *Blues*, with "Energy" highlighted

This is a first tool to improve visibility on this chart. We also observed that the displayed data was very noisy. This is due to the poor number of data for certain years, which translate in very different data one year to the next. We then decided to artificially reduce this noise, by applying a rolling-average on a 5-years window. This helped reduce the noise, but we didn't want to enlarge the window, as it would mean losing data, and sight of short-term tendencies.

With hindsight, we now believe that a histogram-like chart could have been more pertinent, as the temporal dimension is not so important in this visualization.

4.5 Difficulties for map

Other than the time it took to actually display a proper map with **d3-geo**, and the preprocessing task that we had to do, an additionnal difficulty was the mismatch of genres in the two diffrnt datasets. For this purpose, we had to implement a mapping to link each genre of dataset n°1 to its corresponding genre in dataset n°2, going from 17 genres to 6. As a result, we ultimately loose information on this visualization. For example, many genres are mapped as "Other", even though they are very different (*Gospel* and *Ska* for example). To reduce this confusion and to be fully transparent with the user, the corresponding genre is displayed in the title above the map.

Here is the example, for *Ska* pointing to "Other":

Figure 5: Map for *Ska*, effectively "Other"

5 Conclusion

In conclusion, this project allowed us to learn many techniques in data-visualization, but also in HTML/JavaScript/CSS page creation, which was very interesting. We would have liked to go further in the visualization, with more detailed ones, or with a research bar. Being able to navigate more precisely through genres, but also albums or artists would have been a great step.

This project demonstrates the power of data visualization in uncovering meaningful insights about music genres and their evolution over time and across regions. By analyzing detailed song characteristics and geographical listening trends, we revealed the unique traits that define each genre and highlighted their cultural significance. The integration of visual tools, such as timelines, feature-specific charts, and world maps, provided an interactive and intuitive platform to explore these patterns.