



Process Book Vizir

Trends in Music History

Alessandro Tempia Calvino
Alexandre Hutter
Florian Singer

4 June 2021

Contents

1	Introduction	2
2	Data	2
2.1	Data Description	2
2.2	Audio Features	2
3	Visualizations	3
3.1	Initial plan	3
3.2	The Most Popular Tracks per Year	4
3.3	Evolution of genres and music production over the years	4
3.4	Audio features of genres	5
4	Website	6
5	Peer Assessment	6
6	Conclusion	7

1 Introduction

Data visualization aims to graphically visualize data so that main features and trends can be easily extracted and examined. It is an effective way of communicating a message appealingly and interactively. Data visualization crosses the classic frontiers of computer science by meeting **art**.

In this book, we will go through our ideas and implementation steps emphasizing the choices we made during the creative process. Our work analyzes different trends in music history in a wide time range, from 1923 to 2021, by using the available Spotify data. In our visualization, we aim to show: 1) the trends of music production showing the most popular tracks per year and investigating their audio features and genres; 2) the evolution of genres in history and their production; 3) Audio features of genres and the genre evolution through time (e.g. popular music varies consistently through the years). The website is structured on a single page containing three interactive visualizations that all compare the evolution of a particular aspect over the years.

2 Data

In our initial brainstorm, we decided that music is a right way to go for a visualization. We started by finding some datasets that allowed us to visualize interesting trends and the taste of people. On Kaggle, we found a large-scale and up-to-date dataset extracted from the Spotify API which contains enough data to cover the last century.

2.1 Data Description

The Spotify dataset contains two dataset. One contains artists, the other song tracks. The artists dataset provides details on 1.1 million artists with their names, genres, popularity and number of followers. The tracks dataset on the other hand contains information about 600 thousands songs, such as their name, release date, the artist(s) who created the song, the duration and many other audio features described below.

2.2 Audio Features

The Spotify API gives the possibility to extract several audio features that are used in our representation to find similarities and trends. The features that were present in the dataset and that we used are:

- **Danceability**: describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
- **Energy**: represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.

- **Valence:** describes the musical positiveness conveyed by a track. Tracks with high valence sound more positive, while tracks with low valence sound more negative.
- **Acousticness:** measures if a track is acoustic. A track with a low acousticness has few electronic instruments (e.g. electric guitars, synthesizers).
- **Instrumentalness:** measures the probability that a track does not contain vocals.
- **Liveness:** detects the presence of an audience in the recording, live performance.

3 Visualizations

Our idea was to focus on finding interesting trends to share with the audience so that people can learn more about music and discover nice songs in history. In this section we start describing our initial plan and then we present the final result.

3.1 Initial plan

The datasets contain a large set of tracks and artists together with audio features and genres. From this starting point, after a first brainstorming session, we collected a series of ideas related to music trends and audio features. In Figure 1, we show the first sketches and basic implementations. Our main ideas were:

- Show the best tracks for every year, in Figure 1a. This is an excellent way of discovering new songs and relive memories. This is implemented in a chart that shows the popularity through time. Circles elements show with tooltips the name of the song and the artist.
- Show the evolution of genres per decade, in Figure 1b. Which were the most popular genres in the '70s? How popular were they? In which years R&B became significantly popular? We planned to show this with a grouped and stacked barplot.
- Show the features of genres, in Figure 1c. We wanted to compare different genres through the years by their audio features such as danceability and energy, and see how music has evolved through these features.

From the original plan, we wanted to extend it with more interactivity and music. We chose to implement also the following ideas in the first visualization in Figure 1a:

- Connect to the Spotify API to reproduce the best tracks
- Compare features of the best tracks
- Add a genre bar to visualize at which genre the best tracks belong to

3.2 The Most Popular Tracks per Year

This visualization shows the most popular tracks that are considered a symbol of their publishing year according to Spotify. We show the tracks from 1923 to 2021 based on the current popularity. The visualization allows us to interactively visualize the popular trend, the genres, and the feature trends. All the elements are interactive and show the information using tooltips, colors, and charts. We integrated a Spotify playlist that allows playing all the songs included in the visualization. The visualization is composed of the following elements:

- **Interactive chart:** the chart shows and compares the popularity and features of the most popular tracks per year. It is possible to switch from a plot to another using the selection element above the chart. The chart contains interactive circles that show the track name and artist in a tooltip. It is possible to zoom in and pan using the brushing element. Some interesting trends are visible such as an increase in energy average, a decrease in acoustiness through time, and a pretty constant alternation of happy and sad tracks (valence feature).
- **Genres bar:** the genre bar shows using colors the genres of the best tracks. In particular, it is an interesting measure to have an idea of the most popular and trending genre through the years and its evolution. From the visualization we can distinguish genre periods in history, in order: blues, jazz, adult standards¹, rock, hip hop, dance-pop, and latino.
- **Spotify playlist:** the Spotify playlist contains all the best tracks that are included in the visualization so that the audience can listen to the music or discover new famous songs while enjoying the website.

To achieve this visualization, we used two datasets, one containing tracks and features, the other containing artists. Genres are attributes of artists, and thus we merged the tables to have the list of genres accessible for each track. Genres needed considerable elaboration and filtering to remove niche genres and converge to a correct genre assignment. Out of all the tracks, we selected the most popular ones for each year. Using these tracks, we created a playlist using the Spotify API.

3.3 Evolution of genres and music production over the years

For this second visualization, we decided to focus on the evolution of music production per genre over the years. At first, we wanted to depict the evolution of the popularity of the genres, as seen on the second plot on Figure 1, but it was not really relevant. This is due to the fact that the popularity score of the

¹Since pop music stands for popular, the pop genre should include all the popular music in history, which differ considerably in style. To differentiate between contemporary pop and older pop, adult standards have been introduced to refer to older pop.

artists/tracks is the value as of today and not at the time when the songs came out. Therefore, we found that it was more meaningful to represent the evolution of the genres in terms of the mean number of tracks that came out during each decade. This allows us to also analyze the actual evolution in music production.

The visualization contains the following parts :

- **The chart** : This is the first and main part of the visualization. It uses stacked bar charts to display the mean number of tracks. On the x-axis, we have the decades and the number of tracks on the y-axis. Each colored stacked bar represent the mean number of tracks for the corresponding genre during the corresponding decade.
- **The buttons** : The second part of the visualization is the block of buttons. They allow the user to interact with the visualization and add/remove genres from it and dynamically adjust the scales.

Using the buttons, it is possible to keep only one genre and analyze its evolution. Taking a look at Folk for example, we see that it became more popular in the 70's up until the 90's as more artists produced songs of this genre. However, we see a rapid decline in the early 2000's which indicates that fewer people listen to this kind of music today than before.

On the other hand, it is also easy to compare different genres. Selecting Rock and Pop for example, we see that both genre has grown a lot in production quantity over the years and never really dipped that much compared to other genres - except for Rock in the early 2000's. At the same time, we can see that Rock seem to be a genre which is far more productive than Pop. Indeed, a lot of Rock tracks come out each year. However, this should be analyzed carefully as many rock songs could have been created by very small independent groups and most Pop songs today are probably produced by big and very popular artists.

Finally, please note that the genre of the tracks was extracted from the genre of the corresponding artists, given by spotify. Therefore, there may be some anomalies in the data for tracks that were released before that an artist got tagged as a given genre. For example, a old song could be tagged as being of the R&B just because its artist created lots of R&B songs later.

3.4 Audio features of genres

The goal of this visualization is to study more in depth the audio features for all the tracks present in the dataset.

The first step was to process the dataset, by grouping for every year all tracks associated to the desired genres, and computing the mean value of the audio features. There is also an aggregate of all the tracks irrespective of the genre. This was done in a Jupyter notebook, and the result was stored in a JSON file.

Then, the visualization was made in three parts :

- **The chart** : A simple line chart with smooth curves, presenting the value of the audio feature (from 0 to 100) throughout the past century. It is possible to zoom on a specific period using the brush context below.
- **The genres** : The user can click on a set of buttons to activate / deactivate the corresponding genre. Each is represented as a new colored line on the chart, with the same color as the button. The name of the genre also appears as a tooltip when passing the mouse over a line.
- **The features** : A dropdown list allows the user to select the feature to display on the chart, with a smooth transition.

Interesting observations can be made from this chart, and the buttons let the user focus on specific genres without cluttering the visualization.

Note that the set of genres had to be carefully selected in order to have both well-known and disjoint genres. A challenge that we faced when processing the data was that the genres are linked to the artists, not the tracks themselves. This means that the genres assigned to the songs may not always be representative, for instance if an artist makes a song different from what they usually make. This also means that an artist linked with two similar genres (e.g. rap and hip-hop) will have all of their songs associated to these two genres, and it will create a strong overlap for the set of songs contained in rap and hip-hop.

4 Website

The website is based on bootstrap for realizing the frontend. The visualization is all contained in one page. The website is available at <https://com-480-data-visualization.github.io/data-visualization-project-2021-vizir/>

5 Peer Assessment

Work of each team member:

- **Ideas, brainstorming**: Everybody
- **Writing and Process book**: Mainly written by Alessandro Tempia Calvino, but everybody took part in the review before the submissions.
- **Website**: Florian Singer : setup and design - Everybody : visualizations and tweaks
- **The most popular tracks per Year**: Alessandro Tempia Calvino
- **Evolutions of genres over the years**: Alexandre Hutter
- **Audio features of genres**: Florian Singer

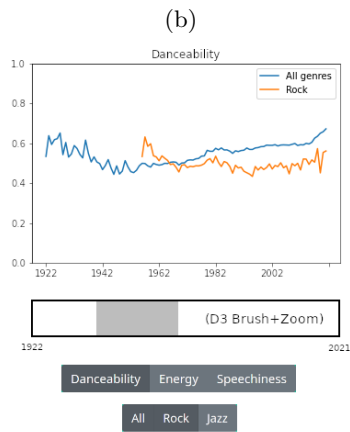
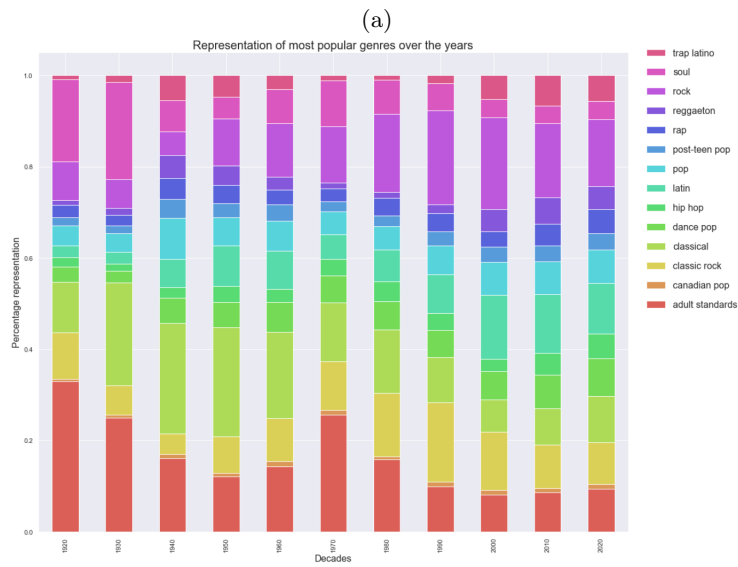
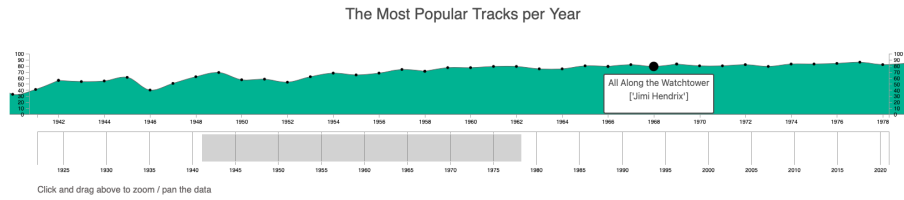
Roles in the project:

- **Alessandro Tempia Calvino:** writer of the milestones and process book. He realized the whole first visualization and the necessary data processing. He realized the screen cast.
- **Alexandre Hutter:** Came up with the dataset and first ideas of genres analysis, fully handled the second visualization along with the data processing required for it.
- **Florian Singer:** Responsible for setting up the skeleton of the website, and handling the third visualization and its corresponding data processing.

6 Conclusion

In this work, we presented our process that led to a data visualization website on music trends starting from brainstorming ideas. The final result can be found [here](#). We presented three interactive visualizations. The first one shows the tracks considered the best ones of their year based on the current popularity. It is possible to visualize the different audio features and genres. The audience can discover the tracks using the embedded playlist. In the second visualization, we showed the evolution of genres and music production over the years. Last, we compared the audio features of the most common genres through the years.

In this project, we hope you discovered famous songs that wrote the history of music or that you learned fascinating facts about audio features and genres.



(c)

Figure 1: Initial sketches and visualizations