Music Genres And Associated Bands

Bernardo Faria
MEIC
Lisbon, Portugal
bernardo.faria@tecnico.ulisboa.pt

Samuel Vicente
MEIC
Lisbon, Portugal
samuel.vicente@tecnico.ulisboa.pt

Sofia Martins
MEIC
Lisbon, Portugal
sofia.carvalho.martins@tecnico.ulisoa.pt

ABSTRACT
Music is a passion for us. We are interested in discovering new genres, how they evolved through time, which artists were the most famous of a genre or location, where are these artists from, what others are connected to them. All these questions motivate us to work on this domain, presenting several interactive idioms that unlock the answers to these questions. To achieve this, we extracted information from the most diverse sources that are widely used nowadays to listen to music: from Spotify to AllMusic. The combination of all this data, together with an intensive work of lamination to have all the information needed and changes to accommodate the desired idioms, all this work resulted in the interactions that our project presents. With this, we are able to answer the questions above (and others), in an interactive, illustrative and accurate way.

Author Keywords
Music; genre; d3; visualization; idioms; artists.

INTRODUCTION
Music is a passion for us, the authors of this paper and is also a world of choices: there are more than a thousand music genres that have been created until today. Nowadays, genres are becoming more blended, and even Spotify has labeled over 3,000 genres to their available content. There is a lot to choose from. Each one of us has an interest in different musical genres. This difference, together with our passion for this art, helped us choose the domain of musical genres and to present popular artists within each one.

Music is not just listening to it: it is far beyond. It is also knowing the background, what is behind it. It is relevant for us to learn when a certain genre was more popular. It may be more recent, but it did not cause a major impact; it may be older and have recently return to the top of the lists, a resurgence. It is very interesting to watch the evolution of this art: this is our proposition.

Recently, it is increasingly common to use platforms to listen to music: these concentrate almost every single proof of each genre. We can say they have everybody's cup of tea. Therefore, we found interesting using these platforms' data to obtain information about the most recent decades. With our project, we intend to have an overview of the evolution of the biggest music genres, their origin, and their popularity; which are the most famous artists; what is the geographic distribution of artists compared to a genre: is it a genre popular in a certain country, but the artists who elevated them to the tops are from the other side of the globe? This also takes us into a journey around the world, to discover these crossings in this art history.

Our First Questions
Finally, we wrote some initial questions, which we, as a group, were interested in answering. Based on these, we started developing our work: how has a genre popularity evolved over the decades? What are the most popular artists of this same genre? And where did they come from? What more artists, whether on the top or not, are related to these? How has a particular genre evolved in a particular country? And did an artist always kept the same genre, or did it try new paths, risking different or even opposite genres? There were our initial questions that motivated us to carry out this project

RELATED WORK
Music is a common art in the daily basis of most people, thus is an art well studied and investigated, for business purposes or simply to admire how music works. For our study, we searched how visualizations of these studies are being created and what could we improve.

This art could turn into a general study of timelines or in a deeper way, how different characteristics like volume, mood, melody, instruments, tempo etc. that can be extracted from the audio file change depending its genres, existing multiple visualizations about it.

Due to technology, there are various authors that generate a visual studies of specific sound properties that enable users to interact and visualize the music in a way that is more understandable, meaningful, and entertaining. One of the interactive interfaces which is Visualizing Genres [1] is presented for how visualizing genre data could help us understand its music landscape with publicly available data. The interface consulted Spotify's API to get their assigned genres, which is one of the objectives for our visualization, even lacking interactivity, since the only interaction available is selecting a square, which represents a genre.

In order to design and develop music visualization it is possible to have real-time animation where animation is generated based on the music data when it is played, like the Windows Media Player (1998) and preprocessed animation that is developed by the author according to specific ideas about the purpose of the animation. One related work that

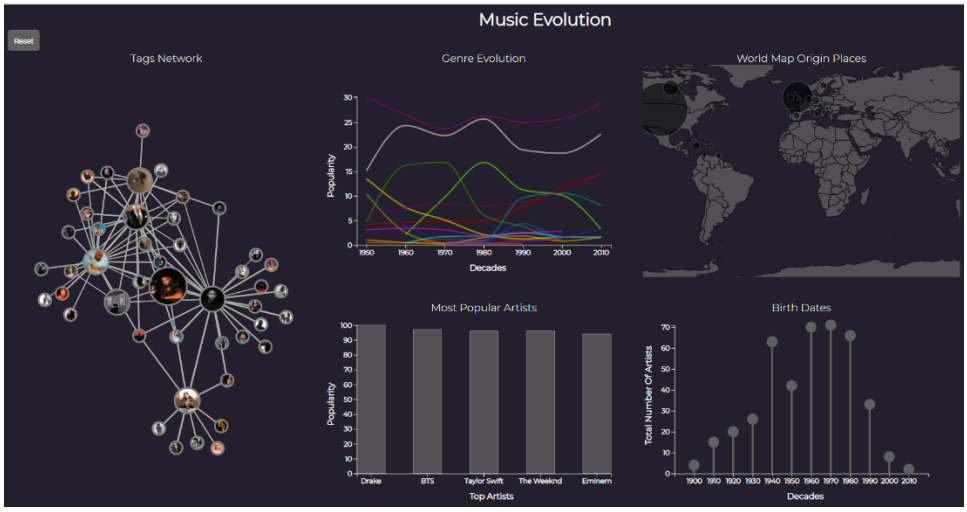

Figure 1. Layout

helped our visualization is Musicmap [2] that attempts to provide a genealogy of popular music genres, including their relations and history, making easy to understand despite its complexity. The authors choose a specific number of genres, determining forms of hierarchy and analogy and ordering everything in a logical way.

MusicPopcorn [3] displays in a network idiom how genres and sub genres are related with inserting the notion of genre hierarchy. This introduction of a network to understand the relations with genres gave us the idea to display how the genres and artists can also relate.

These visualizations had the genre in common but the artists and how they are connected even though the genre it is not the same are not displayed, we found that this type of relations were not being used as a form of data visualization. Therefore, we decided to use as an idea these sources and join them in a visualization that could also answer us to the initial questions.

THE DATA

Since we did not find any complete dataset on the internet that we though was useful for our use case we had to build our own; this turned out to be a whole project in and out of itself. For this we made use of several websites that we scraped using Python and Beautiful Soup 4 [4] and an API to gather the data.

We began building our dataset based of one we found regarding the top 100 artists from each decade [5]. From this dataset we only took the artists and decades fields. Since this data alone was not enough, we had to complement it with other fields.

We chose our source websites/services because they are the most used platforms for music data or consumption.

To this dataset we also added information that we found online [6], for completeness sake.

We scraped MusicBrainz [7] to gather information about:

The artists:

- Origin;
- Creation/birth date;
- Tags;
- Genres;

The albums:

- Album name;
- Release date;
- Tags;
- Genres;

We scraped AllMusic [8] to gather information about:

The artists:

- Origin;
- Styles (related to genre);
- Genres;

The albums:

- Album name;
- Release date;
- Themes; (what themes one associates with the album);
- Moods; (what moods one associates with the album);
- Styles; (related to genre);
- Genres;

We scraped LastFm [9] to gather information about:

The artists:

- Origin;
- Creation/birth date;
- Tags;
- Genres;
- Listener Count;
- Scrobble Count; (count of how many plays and artist has)

- Name;
- Image;

We used Spotify's API [10] to gather information about:

The artists:

- Popularity;
- Genres;

The albums:

- Album name;
- Release date;
- Popularity;

In the end we ended up discarding all information regarding the albums as it was too difficult to process, multiple names for the same album across the different sources were the issue, also the information about the album's genres was too difficult to process and our method of processing would generalize so much that it would become irrelevant, so all the questions that used it were also discarded.

In the end we only ended up using information about the artist origin, genre/styles, tags, creation date, popularity, image and name.

The following tables illustrate our primary datasets, we derived some more from these only to not compute them at runtime in D3.

Fields	Type
artist	nominal
decade	continuous

Table 1. Decades dataset

Fields	Type
artist	nominal
popularitSpotify	ratio
displayName	nominal
imageURL	nominal
Tag	nominal
Genre	nominal
country	nominal
creationDate	continuous

Table 2. Artist dataset

Since our project had to run offline, we also downloaded every imageURL from the Artist dataset and manipulated it using ImageMagick [11] to resize and cut into a square to use on the network idiom.

The derived datasets were:

- A json for the network that contained information (aside from the mandatory id, source and targets):
    - nodes: displayName and picture (name of the downloaded file for that artist);
    - links: weight (count of common tags) and tags;

- A CSV for the line chart that contained. for each decade. the popularity of all the genres. This popularity was derived from the percentage of artists on the decades dataset that contained that genre for each decade.

For the network json we had to compromise and only save the top 5 links of each artist (most tags in common), if we had not done this our network would become a ball of air with all the links it would have. The first json we generated had around 170k links, with this compromise we turned it into around 2k links.

For processing all these data we used Python Pandas [12] and R [13]. Since we had multiple sources, we could correlate it for a more correct final dataset. For the missing data we just manually entered it, after searching Wikipedia [14], as it was a small amount of missing data.

To clean and merge things we just normalized everything we could and if things matched great; else, we would manually check and select the correct one.

One of the biggest problems we faced ourselves with was how to deal with the genre information. There are probably thousands of different genres and our merged dataset really showed that, and since we wanted to visualize the evolution of these we had to compromise and filter only the main genres from our dataset.

To do this we first though of manually researching each artist and cross the information we had to the one on Wikipedia, that turned out to be too much of a hassle so we had to find a way to automate this. Our method wasn't perfect but we had to compromise, after all, genre analysis wasn't the topic of this project and that alone could probably be a theme for a master thesis.

The method we came up with and used to reduce the genre space was to substitute the more specific genres into the parent genres, to do this we used one list we found on Wikipedia [15] and mapped all the subgenres on our dataset to the main genres on that page. This reduced our space to only 17 genres.

VISUALIZATION

Overall Description

The solution achieved is a data visualization that is divided into 5 sections (see Figure 1). Each section displays different type of information and it is possible to interact with it and between sections. There is also a button that resets to the initial state of the visualization.

At first, the visualization will display all genres on the Genre Evolution, one line per genre, as well for all the artists locations in the World Map and the Birth Dates. The Most Popular Artists initiate with the most popular artists. The Tags Network is initialized with the main node as the most popular artist.

Data can be filtered by interacting with each section. Besides selecting there is also the possibility to use hovering to read what is selected.

Next, we will present how each idiom works and how to interact with it.

Genre Evolution

The line chart (see Figure 2) displays the genres and how they evolve throughout the decades, using the popularity attribute. This idiom starts by representing all genres, one per each line, throughout the decades (from 1950 to 2010). Hovering on a line, a small box appears indicating which genre is being hovered. Also, while hovering, this same line turns off its opacity (brighter), while the other lines turn on their opacity (less bright), making the visualization clearer about which mark is being selected. Clicking on a line, a genre is selected, and three interactions occur on three idioms: the bar chart now shows up to the top 5 artists of the selected genre, the lollipop chart represents the total number of artists of the selected genre that were born in each decade, and the network selects the most popular artist of the genre, it places it on the central node, and linked to it are related artists (in terms of tags).

If the user selects a location on the World Map, the line chart displays the genres for that location. If an artist is selected in our visualization (bar chart or network), the line chart displays the line referring to the artist' genres.

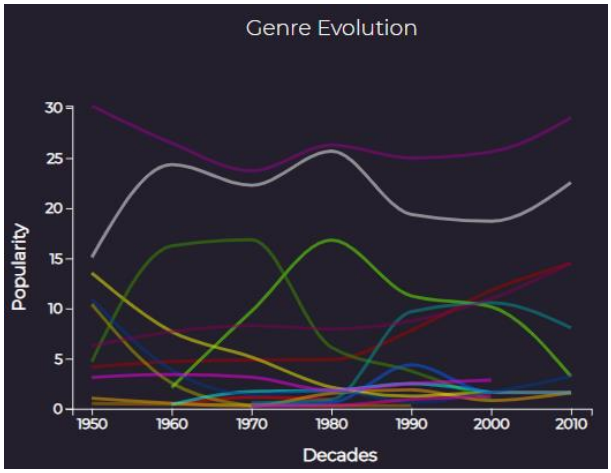

Figure 1. Line chart.

World Map Origin Places

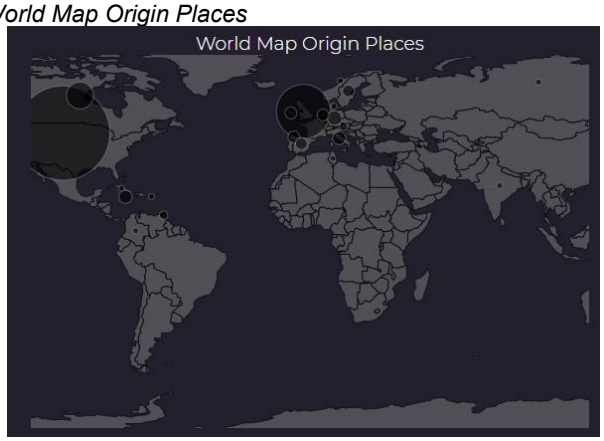

Figure 2. Bubble map.

The bubble map (see Figure 3) represents the origin places where an artist was born or created. In the initial state, displays a bubble on top of the countries of all the artists of our dataset. The size of the bubble indicates if that country has more or less artists: the bigger the bubble, the more artists that country has. Hovering on a bubble, a small box appears indicating the name of the country and the total number of artists of that country. Also, while hovering a bubble, this same bubble turns off its opacity (brighter), making the visualization clearer about which mark is being selected. Clicking on a bubble, the country the bubble represents is selected, and four interactions occur on four idioms: the bar chart now shows up to the top 5 artists born on the selected country, the lollipop chart represents the total number of artists of the selected country that were born in each decade, the line chart represents all the genres that are linked to this specific country, and finally the network selects the most popular artist of that country, it places it on the central node, and linked to it are related artists (in terms of tags).

When a genre is selected in our visualization (line chart), the bubble map displays the locations of the artists of that genre. If an artist is selected, the bubble that represents that artist' country will be highlighted

Most Popular Artists

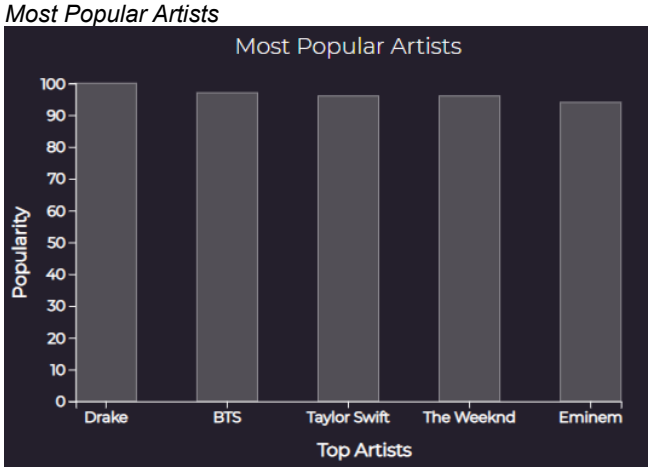

Figure 3. Bar chart.

The bar chart (see Figure 4) represents up to the 5 most popular artists of a specific music genre or a specific location, or an artist itself when selected on the network. The popularity is measured in a scale from 0 to 100 on y-axis. Hovering a bar, that bar gets highlighted and a small box appears showing the exact popularity of the artist that is being hovered. By clicking a bar, that artist is selected for the entire visualization, interacting with the four idioms: the line chart shows the genres associated with the artist selected; the lollipop chart highlights the lollipop of the decade of birth/origin of the artist; the bubble map highlights the bubble of the country where the artist was born/created; and the network places the artist on the central node, and link to it are related artists (in terms of tags).

When a location is selected, the bar chart will show the most popular artists of that location, but if a genre is selected, it will display only artists for that genre, the most popular ones. When an artist is selected, the bar gets highlighted.

Birth Dates

The lollipop chart (see Figure 5) represents the total number of artists born in each decade, from 1900 to 2010. In the initial state, it represents all the artists birthdates per decade. There are two different hover events: when no artist is selected, the hover shows a small box containing the total number of artists of a specific genre (this happens when clicking on the line chart), of a specific country (this happens when clicking on the bubble map), or the total number of all the artists (this happens in the initial state or when the "reset" button is pressed); when an artist is selected (this happens when clicking on the bar chart or the network), the lollipop that has that artist birthdate gets highlighted, and the hover for that lollipop shows the name of the artist and its exact birth year.

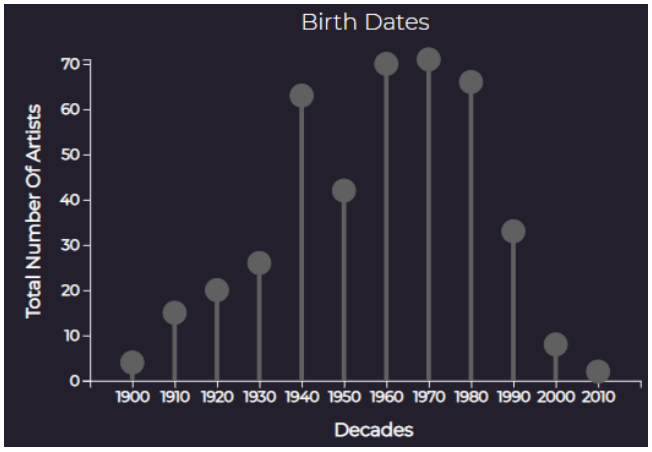

Figure 4. Lollipop chart.

Tags Network

Our last idiom is the network (see Figure 6). In the initial state, its central node has the most popular artist overall. The links represent the tags that connect the artists, and these are represented on each node, with a picture of the artist. There are three node sizes: the bigger node is the most popular artist selected; the medium nodes are the artists that are directly connected to the central artist; the small nodes represent artists directly connected with the medium nodes. So, in the end, we have a 2 depth-level network. This idiom has two hover events: when hovering a node, a small box appears indicating the name of the artist that is being hovered; when hovering a link, it displays the tags that formed that same link. Clicking on a node, four interactions occur in the other four idioms: the line chart displays the genres associated with the selected artists; the bubble map highlights the bubble of the country where the artist was formed; the bar chart displays only the bar of that same artists, showing its overall popularity; and finally, the lollipop chart highlights the lollipop that represents the decade which has the year that artist was born/created.

If a location or a genre is selected, it will show the most popular artist of that location or genre. If artist is selected, it will display the artist itself.

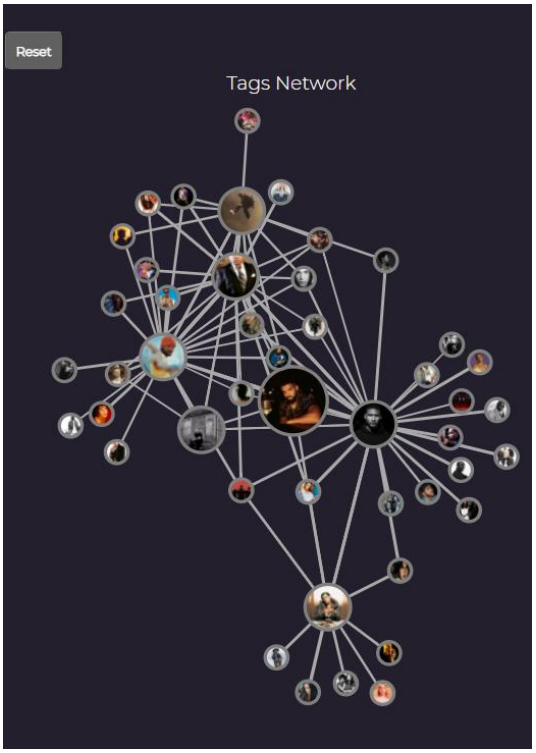

Figure 5. Network.

Rationale

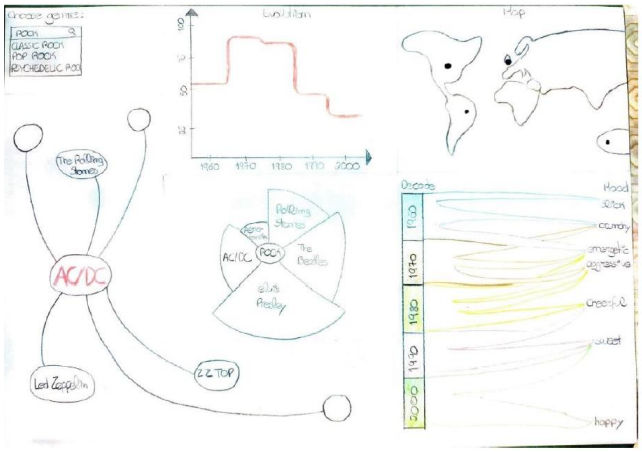

Figure 7. First prototype.

Initially, without thinking in the development process, we start searching for different methods to display data and also answering our initial questions.

From the solution to the first prototype (see Figure 7), only pen and paper, we change some charts and how the charts itself would work. This happened due to development issues and also lacking interaction between the different idioms.

The layout changed, starting by the dropdown menu that would filter the genres, and then all idioms would change. This button turned out to be a gateway for our interactions not working, in other words, obsolete with our current development. A reset button replaced it.

For the line chart, we were only showing a line, for one genre, which would work, but would not give us an overview of how the genres evolved though time.

Our initial intentions for the map were kept, as for the network. An addition to the dataset was made, with the images for the artists, to replace in the nodes of the network, instead showing only the name.

The circular bar chart turned into a normal bar chart. The information would not be well displayed, also the design of the circular bar was planned to be a sound column. The lollipop chart was not intended initially, and the Sankey chart was discarded.

A timeline chart was in our initial plans, but since the line chart will display all genres at the same time, this turned out to be a duplicate chart to see the same data.

Demonstrate the Potential

Let us answer some of the initial questions: How as a genre popularity evolved over the decades? What are the most popular artists of this same genre? These two questions are answered on Figure 8. We selected the genre Pop (the hover on the line indicates the genre we selected), and we can see its popularity, it decreased in the beginning of the 1950, but in 2010 it emerged on the tops. The most popular Pop artists are displayed on the bar chart. Drake has a popularity of 100, being the most famous Pop artist.

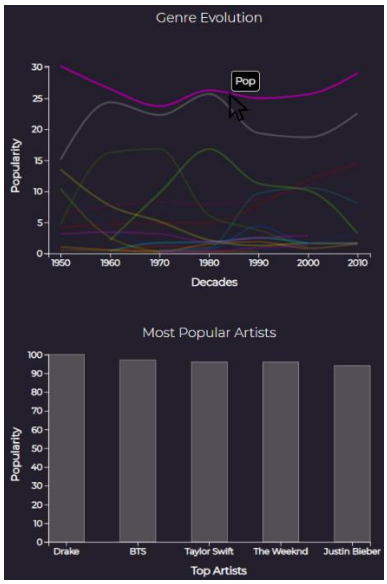

Figure 8. First example.

In other question, we want to know where an artist came from. Taking the example above, from the artists on the bar chart, we want to know where Taylor Swift was born. We click on that artist, and the bubble map highlights the bubble of the country where she was born – United States of America (Figure 9).

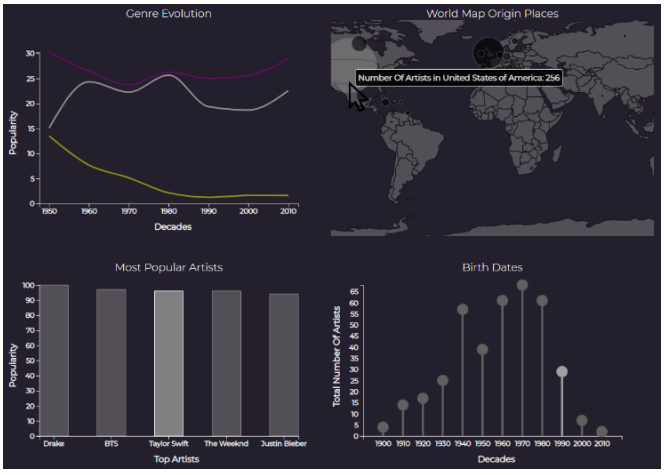

Figure 9. Second example.

Since we clicked on an artist, another question can be answered: if an artist always kept the same music genre: from the line chart, we see that Taylor Swift has three different genres associated, which indicates she did not keep her music in only one genre. Note that the bar chart highlights the artist selected. From this figure, we can also answer some other

questions not presented above. The hover on the bubble map also indicates the total number of artists born on the United States: 256. Also, since we clicked on an artist, the lollipop highlights the lollipop that represents the decade when Taylor Swift was born.

Figure 10 answers the question: what more artists, whether on top or not, are related to others? Since we clicked on Taylor Swift (the bar highlighted indicates this), the network on the left placed her at the center node. All the other nodes linked to it are artist related to her in music tags. Hovering on the nodes shows the name of the artist, and hovering on the links shows the tags that form that link.

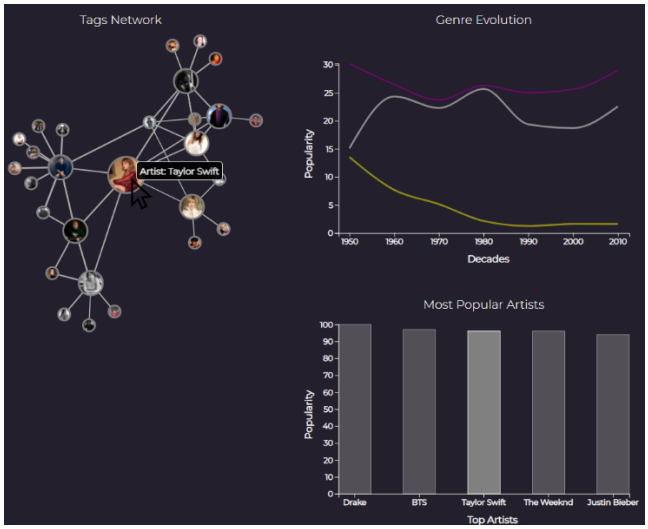

Figure 10. Third example.

IMPLEMENTATION DETAILS

We had a few challenges with the datasets. An example: to implement the network, we needed a specific format of dataset we did not have, consisting of an array of links and nodes. To have this format, we needed to format and work with the artist dataset, mostly with the tags attribute. We found all the combinations of artists that shared an equal tag, and that forms a link. We did this to all the artists we had, which gave us an enormous number of links. To reduce this, and also to not overload the network, we cut the links to 5 links per artist.

For the line chart, we needed to calculate every popularity of every genre of every decade. To to this on a javascript function, it would consume a lot, so we organized a new dataset that had only three attributes: decade, genre, and popularity. This way, it was much easier to work with the data and to implement the idiom.

And for the bubble map, we needed a new dataset that could implement the map, with each countrys' coordinates and names. This dataset was obtained from the World Atlas TopoJSON github repository [16]. We used the countries-110m.json, on a 1:110m small scale to adapt to our visualization.

As mentioned earlier, at first, we had a dropdown box in which we could choose the genre we wanted to visualize. But since this approach was not helpful during our implementation, we discarded it and decided to present all the genres on the initial state of the line chart, each genre per line and a different colour. We decided not to have a legend of all the genres but having this information on a hover event, because we have a lot of genres, which would make the legend too big to appear on screen, and some colours could be misunderstood. So, the hover even facilitates this problem, because when one line is being hovered, all the other lines are turned down on the colour and the one selected gets highlighted, and the hover box with the genre name is much more appealing.

CONCLUSION AND FUTURE WORK

Implementing this visualization was facilitated by the planning that we had, from the data scrapping to thinking what the interactions between the charts would be, and every detail was helpful to achieve our working solution.

Unfortunately, not all the first questions were answered (but new ones appeared, like the number of artists in a specific country and the number of artists per decade), but if we had to do it all over again, we would gather more data about the music industry and how it affects the tops, and more precise questions with the data we would have.

The skills that we gathered were mainly with the implementation and using D3 as a library, this took some time to understand and implement.

More time and more skills would combine in a visualization that would still display all the genres and artists, but with an enrich design, a design that we had in mind when we first thought of our subject for the visualization. The dataset would also enrich to use years, instead of decades, we would be able to deal with scalability, a problem with having so many years of this art.

ACKNOWLEDGMENTS

Sample text: We thank all the volunteers, and all

REFERENCES

1. Plantinga, Bo. Visualizing Genres. Jan, 2020. Retrieved from https://tabsoft.co/2SCcgd0
2. Crauwels, Kwinten. Musicmap. Oct, 2020. Retrieved from https://musicmap.info/
3. Lamere, Paul. MusicPopcorn Sep, 2013. Retrieved from https://musicmachinery.com/2013/09/22/5025/
4. Beautiful Soup 4 https://www.crummy.com/software/BeautifulSoup/bs4/doc/
5. Original Dataset https://www.crummy.com/software/BeautifulSoup/bs4/doc/

6. Complementary data https://www.crummy.com/software/BeautifulSoup/bs4/doc/
7. MusicBrainz https://musicbrainz.org/
8. AllMusic https://www.allmusic.com
9. LastFm https://www.last.fm/
10. Spotify's API https://developer.spotify.com/
11. Image Magick https://imagemagick.org/index.php
12. Python Pandas https://pandas.pydata.org/
13. R https://www.r-project.org/
14. Wikipedia https://www.wikipedia.org/
15. List of popular music genres https://en.wikipedia.org/wiki/List_of_popular_music_genres
16. World Atlas TopoJSON. Retrieved from https://github.com/topojson/world-atlas