“How weather conditions influence what we listen”

Information Visualization — Final Report

MEIC – A · Instituto Superior Técnico · December 2018

Group 17

Hélio Domingos	Miguel Regouga	João Pina
83473	83530	85080

ABSTRACT

Everybody listens to music, and various factors may influence the music we choose to listen. Weather conditions could be one of them. In order to find out, we decided to create a visualization that shows if weather conditions can influence what we listen. This report expounds in detail the decisions, procedures and techniques used to create such visualization, as well as the steps taken in order to make it understandable, intuitive and concise.

Author Keywords

Information; visualization; infovis; music; songs; Spotify; streaming; streams; charts; songs; countries; weather; conditions; rain; sunny; fog; thunder; hail; tornado; d3; JavaScript; HTML; CSS

INTRODUCTION

Music is a fundamental part of everyone’s lives. As Shakespeare once told, “When words fail, music speaks.” Everybody listens to music and now, more than ever, people are listening to more and more music, due to the rise of streaming services.

It wasn’t always like this. In the early days – that is, in 1877 – phonographic cylinders were the earliest commercial medium for recording and reproducing sound. In 1895, came the Vinyl record; in 1963, the cassette; in 1982, the compact disc (CD). Then, the digital revolution came, and in 1993 the .mp3 format was created; since then, we’ve had the iPod and its iTunes Store, and then came the streaming services. With this evolution in music formats, listening to music became much easier and cheaper, at a point where you can enjoy your favorite songs without paying anything, while still doing it in a legal way.

However, people don’t listen to the same music all the time. There are various factors that can influence what people choose to listen. For example, if a group of friends is attending a party, they might listen to more “happy” songs, with higher tempo and energy. But if you just came home after work and you’re tired, you might want to listen to more “slow” songs, with smooth guitar riffs and easygoing voice.

One of those factors could be the weather conditions outside. It is scientifically proven that your mood can change based on the weather outside. If your temper changes, it could mean that what you want to listen may also change. In sunnier and hotter days, people may want to listen to happy songs that remind them of summer, but if it’s raining, they might want to stick with Prince’s “Purple Rain”.

Based on our absolute love of music, and in the lack of information about this subject online, we decided to explore this concept in a profounder way, in order to get a glimpse if our theory can be proven right. We want to know, for instance, if people hear more music if it was a sunny day or if it rained, how the top charts can change based on that, the favorite artists for when it’s a foggy day, and other thought-provoking questions.

Spotify is easily the most popular and robust streaming service right now, with 180 million users worldwide. It is currently available in 53 countries as of 2018. That being said, we can get the most accurate description of what people are listening by getting a list of the most listened tracks of each day. As both this and weather conditions of each country are public data available to everyone, we were able to conjugate both sides of information, in order to visualize possible listening patterns.

Tasks to be supported

As stated in the early stages of the development of the visualization (Checkpoint I of the Visualization Information course), we defined four different tasks that our visualization should allow to perform:

Discover — find music to listen based on weather conditions;

Identify — given a weather condition, identify which artists are most likely to be streamed and vice-versa;

Compare — select multiple countries to see how each country’s weather conditions influence what people stream;

Analyze — study how music habits change with the weather.

Example questions

Given such tasks, these questions can be answered while exploring the final visualization:

- On a sunny day, which song is the most listened worldwide?
- If it’s raining, what artists do people listen the most in Ecuador?

- In Portugal, do people tend to listen to more music when it's raining or when it's sunny?
- What artists do people listen when it's a foggy day in Spain?
- How many times was a song on the top 10 streamed, given it was a rainy day in Argentina?
- What were the top 10 streamed artists in Christmas day in the United States?

RELATED WORK

While searching for similar projects online, we didn't find anything alike what we'd want to visualize, and that aroused even more our desire of doing something about it.

However, we could indeed find some information that prove us that weather conditions change people's mood (http://bit.do/eDKsb). In fact, there is a study conducted by Spotify and Accuweather that shows how users' listening habits respond to the weather conditions, based on different audio attributes, like the energy or danceability of a song (http://bit.do/eDKsk).

Our main inspiration to our visualization was, however, our own experience — being music listeners as we are, we were able to find out that our music habits change based on the weather outside. For instance, we may even create playlists to listen to when it's rainy or to when it's sunny.

THE DATA

Original data

In order to achieve our goals, finding the data is a key part of the process of the design of our visualization. Fortunately, it was fairly easy to find it.

Our hunt for the data could be divided in two parts: the music part and the weather slice.

For the music slice, and as said before, Spotify makes public the information about the top charts of each day. For the sake of simplicity, we gathered the information for the year of 2017 only. Our music dataset comprised of the top 200 songs streamed each day on Spotify, for each of the 53 countries it is available in.

The weather part was tougher as we had to address a key issue to our visualization — although Spotify makes public the chart information for each country, it does not reveal the top streamed songs of each city. Although the weather conditions won't vary that much in a small country (for example, in Singapore), in a big country like the United States the weather conditions can be totally different — in New York it might be snowing, while in Los Angeles it might be a sunny and hot day.

Due to the lack of data, we decided to work with the weather conditions of the most populous city of each country, which represents the greatest number of listeners possible.

Therefore, we went ahead to the U.S. government's own climate database and we obtained 53 datasets, one for the most populous city for each of the 53 countries Spotify is available in. Each dataset contains the weather data for each day of 2017.

Data cleaning and processing

The first step to the processing of our datasets was to remove all the attributes we didn't need from each dataset. For instance, in the weather datasets, we removed the columns that didn't provide any valuable information to our visualization. Some columns were also renamed in order to make easier their use on the code and to keep everything clean and organized.

We added some columns as well to make the combination of both the weather and music dataset easier. For instance, we added one column to each of the 53 weather datasets, corresponding to its country.

After the pre-processing for each weather dataset, we combined all the 53 files into 1 single weather.csv file, containing all the weather data in just 1 file. Then, it was time to combine both the weather dataset and the music dataset into 1 combined.

This led up to a few problems. Firstly, and most importantly, we had a huge amount of data (more than 4 million rows, to be precise); although in first hand it might seem appealing, it can cause some trouble, especially when searching through the whole dataset. To solve this, we considered only the top 10 songs for each country.

On the other hand, we found out that Spotify only ranks tracks on its daily song ranking if it has at least 1,000 streams in that given day. Hence, in smaller countries where Spotify might not be broadly used, like Luxemburg, where it's hard for songs to have that number of streams, there might be some unavailable data. We also found out that Spotify's API was down during 3 days in 2017, resulting in the same consequence.

All of the data processing was accomplished using Pentaho.

Final data

We ended up with a single .csv file, containing all the data we'd want to visualize. Although, in later stages of the development of the prototype, we found out that the dataset could be too big, affecting the performance (that is, how quickly the prototype would search on the dataset). That being said, we had to divide the "main" dataset by each country, in order to access the data of the countries the users select to visualize.

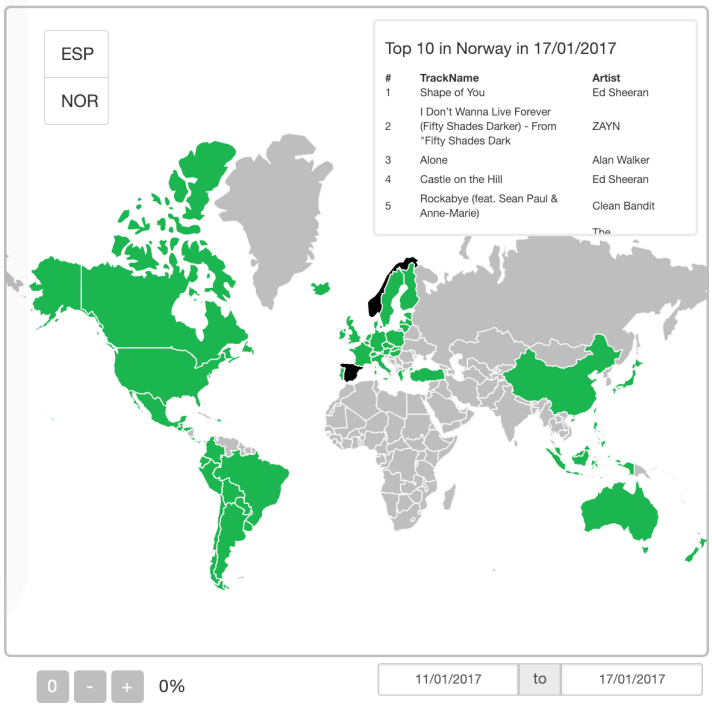


Figure 1 – General aspect of the visualization developed

Although it improved the performance, we weren't still totally satisfied of how quickly the prototype reacted. To solve these problems, we derived each dataset based on the idiom it was going to be used, in order to dynamically access the data, and therefore making the visualization fluid enough. We had to do the extra effort to convert from .csv to .json, in the case of the sunburst idiom, which will be explained later in the report.

We ended up with 212 .csv files (53 countries multiplied by 4 idioms) and 53 .json files.

VISUALIZATION

Overall description

The visualization we conceived is made up of five different idioms, allowing the user to explore the data related to how the weather conditions influence the music people stream on Spotify.

Choropleth map

The screen is divided in half. The main idiom, which takes up the left half of the screen, is the choropleth map, where the user can explore the different countries where Spotify is available in.

When clicking in a country, it allows to see the top 10 songs of a given day (or days) on that country. It is also added to the other visualization idioms on the right, which are all updated as the user adds or changes the selected countries or dates.

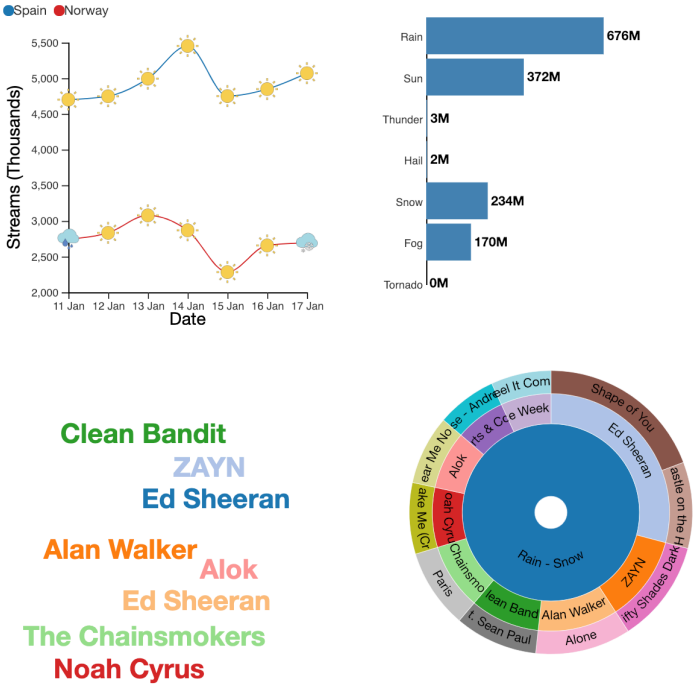


Figure 2 – Choropleth map

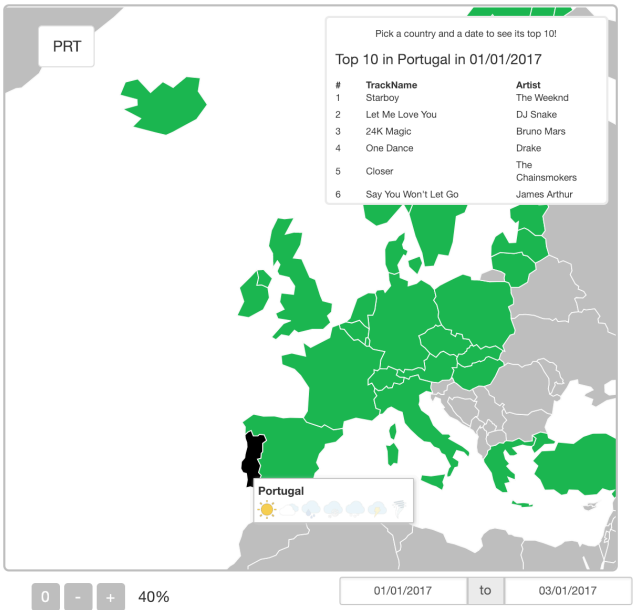


Figure 3 – Top 10 chart in the selected country

By hovering on a country, it allows to see the weather conditions of the selected day (or days) on the selection bar on the bottom-right corner of the map, much like a weather forecast map as shown in newscasts programs.

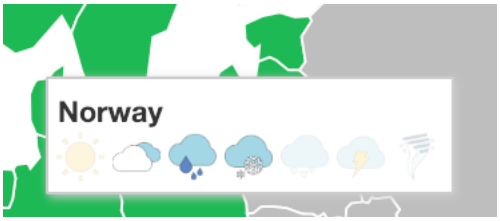


Figure 4 - Tooltip

The user can also zoom in on the map by clicking on the correspondent buttons on the bottom-left corner of the map.

Line chart

On the right half of the visualization there are 4 distinct idioms. The line chart, located on the top-left, compares the total number of streams of the selected countries and dates with the reported weather conditions on those dates and countries selected. This allows to see how the weather influences the amount of music people listen.

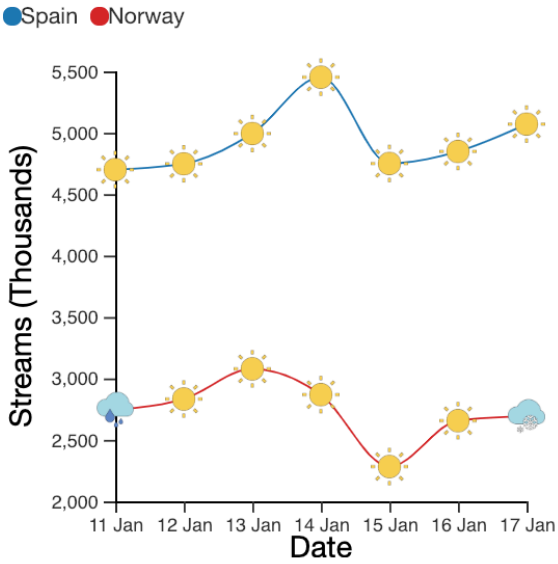
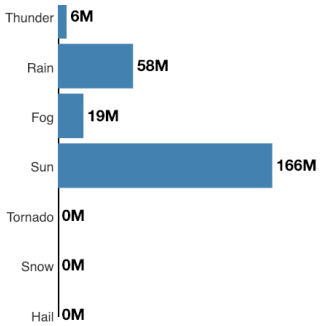


Figure 5 – Line chart with 3 days and 3 countries selected

Bar chart

Figure 6 – Bar chart with Portugal selected



The bar chart, located on the top-right, compares directly each weather condition with the number of streams (in millions) on 2017, on the selected countries. For instance, if the user selects Portugal, the bar chart shows the number of streams in 2017 on days where the day was sunny, rainy, foggy and other weather factors. If a determined weather factor has zero streams, then it means that in that year it wasn't registered any occurrence of that weather condition.

Word cloud

The word cloud, located on the bottom-left, shows the most listened artists on the countries and dates selected by the user. It's a simple manner to see how the music tastes evolves with the weather conditions.

Figure 7 – Word cloud

Sunburst

Finally, the sunburst, located on the bottom-right, allows the analyzation of the most listened songs by weather condition, in the countries selected. If the user wants to analyze in detail, the user simply needs to click on an attribute (or "slice" of the sunburst") to increase its detail and show more information.

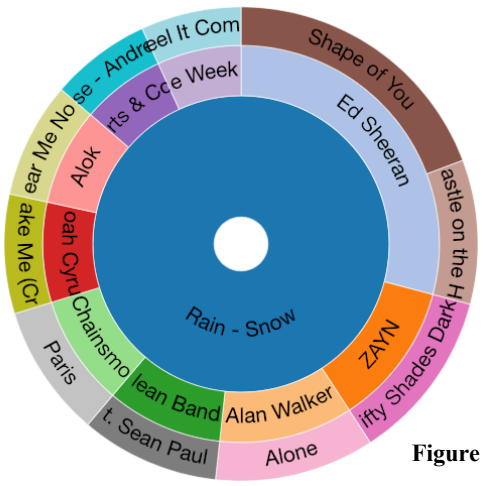


Figure 8 - Sunburst

By hovering on each slice, the user can also get the number of streams of that artist or song, and the corresponding percentage of the day's total streams. The sunburst is ordered clockwise by number of streams and allows to see how the weather conditions of a given country (or countries) influence its song ranking.

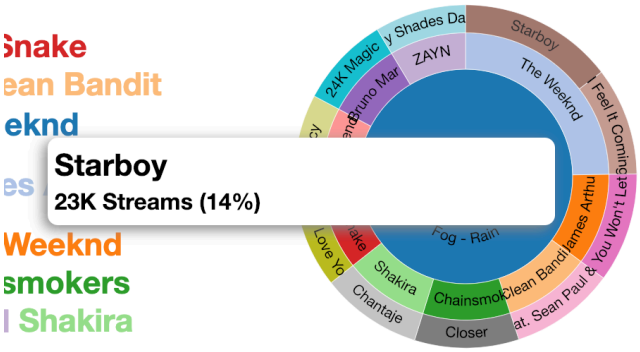


Figure 9 – Sunburst tooltip

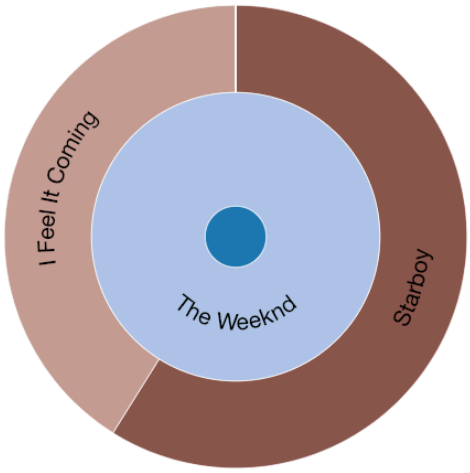


Figure 10 – Detail in sunburst

Rationale

In the early stages of the development, we started by drawing how the visualization should look like, as represented in the figure below.

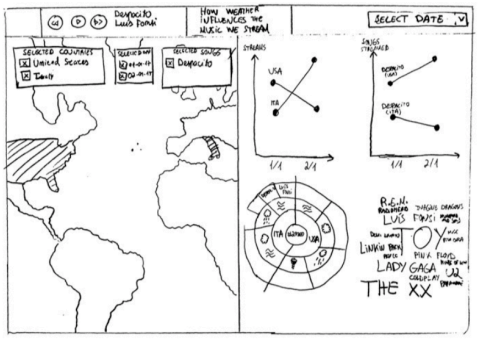


Figure 11 – Initial drawing

Initially, we considered embedding a Spotify player within the visualization. Although, during the development of the project, this became tough to implement, and since it didn't bring any kind of information to our visualization, we decided not to implement it.

Also, in the first drawings, we thought in adding two line charts. After some deliberation, we though that this could be repetitive, and the difference between these two idioms wouldn't be clear enough.

That being said, we decided to implement just one line chart, in addition to a bar chart. The bar chart shows useful information in a more concise and simple way, while the line chart allows the user to easily compare information that otherwise could become difficult to interpret.

We have also considered choosing a simpler list of countries instead of the choropleth map. Although this could be a simple approach, we chose the choropleth map because we wanted to implement an idiom that would resemble the meteorology section of a newscast, giving our visualization a more realistic and familiar interface to the user. We consider it to be more graphically appealing as well.

Finally, we chose to implement the sunburst due to its ease of use to show big chunks of data in an intuitive and fun way.

The visual encodings as stated in Checkpoint III are the following:

- **Choropleth map:** The countries Spotify is available in are given by green on the map and the countries selected by a user are given by black on the map;
- **Line chart:** the number of streams is given by the height in relation to the left axis; the dates selected are given by the bottom axis; the number of streams of a country on a day is given by a weather icon; The difference between streams is given by each line. Value attribute with aligned vertical position. Separate key attribute with horizontal position. Line connecting the attribute values.
- **Bar chart:** the weather conditions are given by the left axis and the number of streams of a given weather condition on a country (or countries) in 2017 is given by the bar with the width in relation to the bottom axis;
- **Sunburst:** The weather conditions and its combinations are given on the first ring; the artist information is given on the second ring with a different colour tone; the song information is given on the third ring with a different colour tone and the given data of each weather condition, artist or song is given by one different colour tone. The arc length is proportional to attribute value. Slices color and label provide additional attribute and key

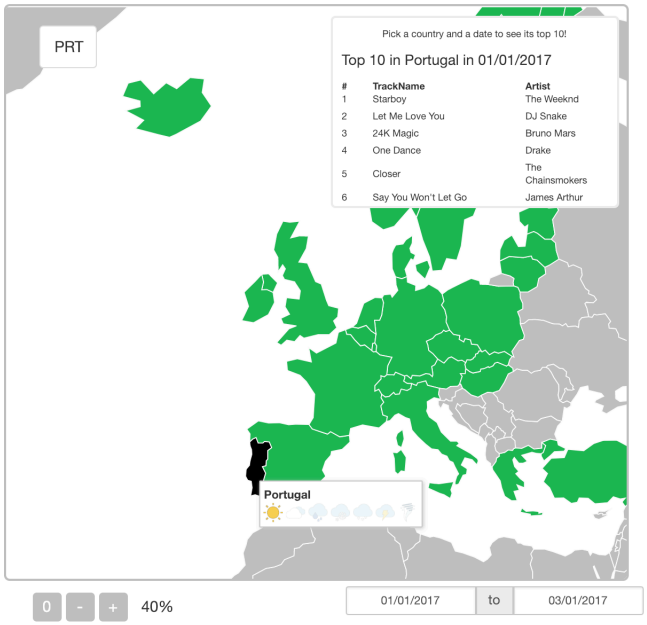
- **Word cloud:** artists with a similar number of streams are distinguished by the colour range.

Demonstrate the potential

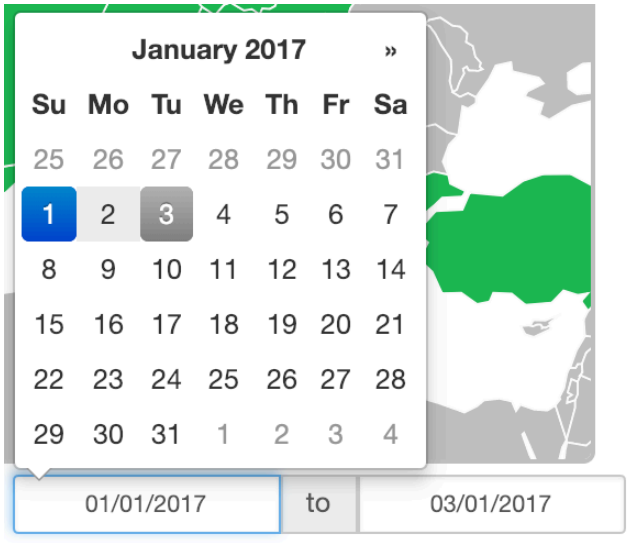
In order to better understand how our visualization works as a whole, we provide answers to two of our questions.

In Portugal, do people tend to listen to more music when it's raining or when it's sunny?

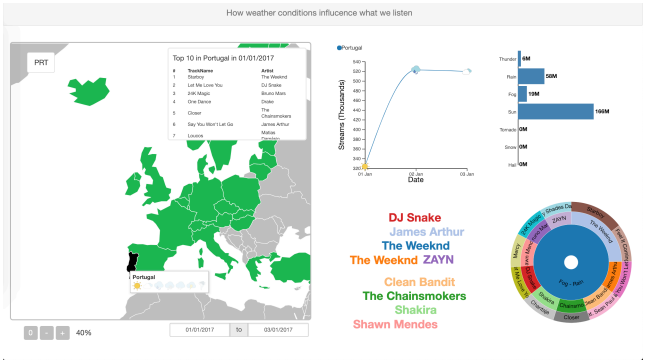
Firstly, the user selects the country (Portugal) on the map:



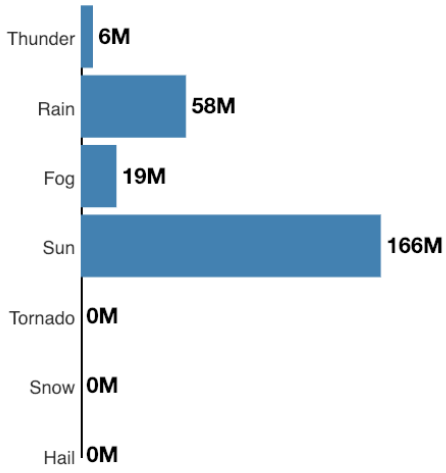
Then, the user selects the dates he wishes to visualize, on the bottom of the map:



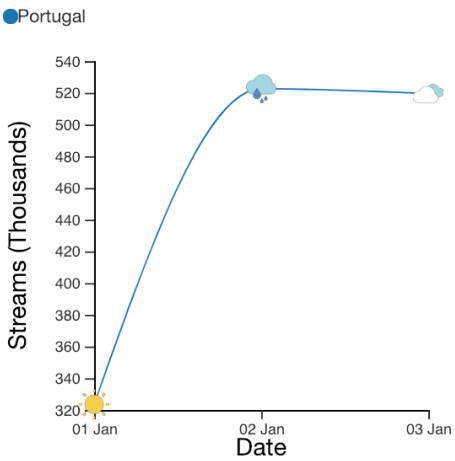
The user can then see all the other idioms on the right half of the screen being updated.



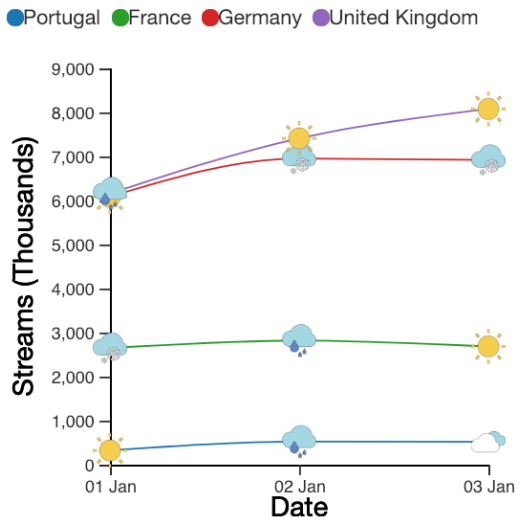
The user can observe the bar chart to directly answer this question, as the bar chart shows the number of streams of the days when it rained, or it was sunny:



It is possible to see more information, based on the question. Looking at the line chart, we can observe users didn't follow the general trend in 2017 — in January 2nd, where it rained, people listened to more music than in January 1st, where it was a sunny day.

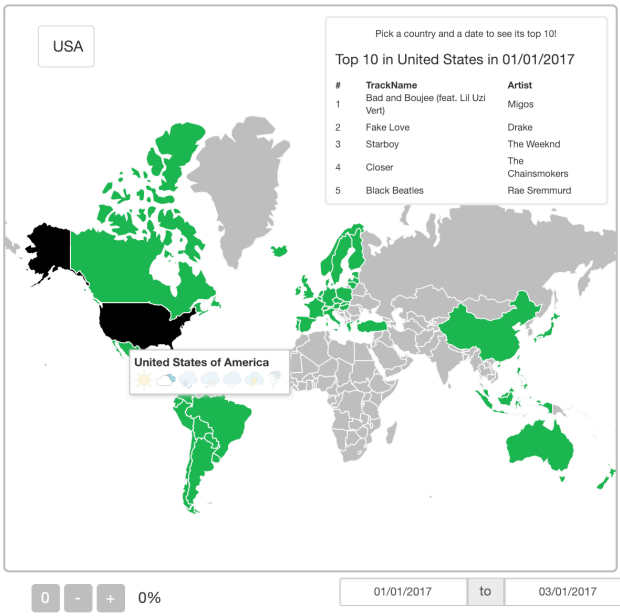


The user can even select more countries to compare, being the line chart updated as more countries are selected:



What were the top 10 streamed artists in Christmas day in the United States?

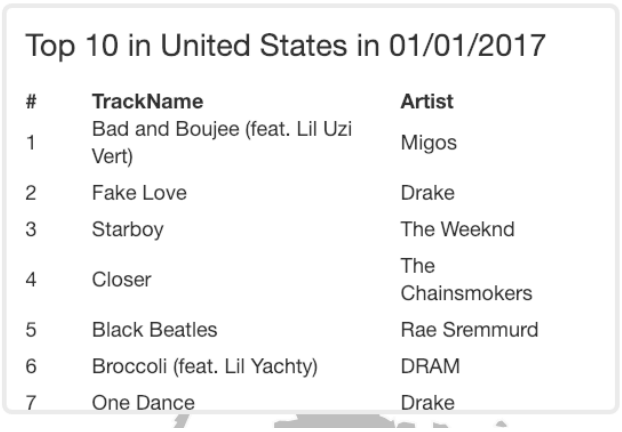
Firstly, the user selects the country (United Sates) on the map. The user can also see the weather conditions on that day by hovering on the country.



Then, the user selects the dates he wishes to visualize, on the bottom of the map (25th of December).

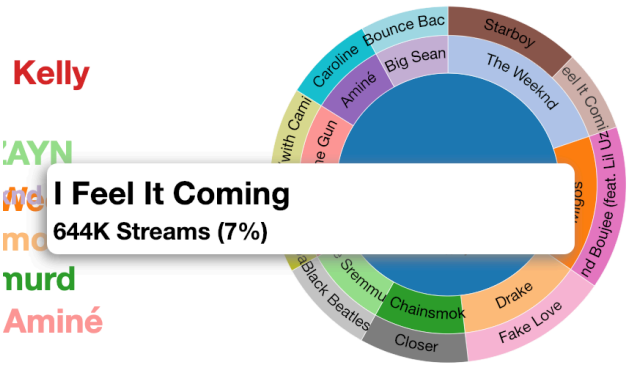
The user can then see all the other idioms on the right half of the screen being updated.

The user can scroll on the box on the top-right corner of the map to see the top 10 streamed songs on that day.



The user can also see the most streamed artists by looking at the word cloud:

Finally, the user can also see exactly how many streams a given song or artist has, by hovering and clicking on the sunburst:



IMPLEMENTATION DETAILS

The implementation of the visualization was done through the use of HTML5 technology (HTML, JavaScript and CSS). We also used the Data Driven Documents JavaScript library (D3.js), as well as a couple other libraries derived from it, such as the DataMaps library (used to develop the map) and the D3-Cloud library (used to develop the word cloud idiom).

The use of global variables allowed us to develop the interaction between all the idioms. Whenever these global variables are changed, so are the idioms, in order to accommodate the information to visualize. The selection of both the country on the map and the dates on the date picker lead to all the other idioms changing, as they are all sensitive to these selections.

We also included the use of hover and animations throughout the visualization, in order to give it a more appealing design and interface, and to make something the users aren't frightened to play around with.

As with all coding projects, we faced some challenges when developing our visualization. For instance, we had some trouble reading the datasets dynamically, as explained earlier in the report. Some of the idioms that we wanted to implement also required a specific version of the D3 library, and we had to adapt them to run on the most recent version. For instance, we had to rewrite the code that was used on the DataMaps library mentioned earlier, as it only worked on v3 of the D3 library.

CONCLUSION

Information visualization is undoubtedly an interesting topic to investigate. This project allowed us not only to learn how to code a lush visualization that shows information in a concise and educational way, but also all the steps before the implementation itself that enrich the visualization.

Although not everything planned was implemented due to some challenges we faced during the development, we consider this project a success overall, as we improved our knowledge on this matter.

With more time and resources, we would like to add more functionality to our idioms, as well as more animations and, if possible, to show more information.

We'd also like to implement more features, such as the possibility to select a song and start playing it on Spotify, or even embed a player on the website itself. Something great to implement would also be live updates of the visualization, that is, show in real-time the weather conditions of a given country and the music people are listening and relate both.

The interface itself also could be improved. For instance, we would like to apply a responsive web design that would allow the visualization to be seen in different window and screen sizes.

REFERENCES

1. https://www.webmd.com/balance/features/can-rainy-days-really-get-you-down#1
2. https://insights.spotify.com/us/2017/02/07/spotify-accuweather-music-and-weather/
3. https://www.kaggle.com/edumucelli/spotifys-worldwide-daily-song-ranking
4. https://www.kaggle.com/edumucelli/spotifys-worldwide-daily-song-ranking
5. https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&resolution=40
6. http://datamaps.github.io/
7. https://github.com/jasondavies/d3-cloud
8. https://medium.freecodecamp.org/learn-to-create-a-line-chart-using-d3-js-4f43f1ee716b
9. https://medium.freecodecamp.org/how-to-create-your-first-bar-chart-with-d3-js-a0e8ea2df386
10. https://github.com/d3/d3/wiki/Gallery