

# Music Atlas (@music\_atlas)

Debjyoti Paul, Nishant Agarwal, Shweta Singhal,

December 3, 2016

## 1 Basic Info

### Team Members:

Name	UID	Email
Debjyoti Paul	u0992708	deb@cs.utah.edu
Nishant Agarwal	u1010232	nishant@cs.utah.edu
Shweta Singhal	u1011542	shweta@cs.utah.edu

### 1.1 Code Repository:

<https://github.com/MusicAtlas/musicatlas>

### 1.2 Website:

<https://musicatlas.github.io/musicatlas>

## 2 Overview and Motivation

History shows music has always been a source of inspiration for the creative minds by the creative minds. Last couple of decades with the advent of information technology and social networking music has reached from one corner of the world to other. Exchange of such aesthetic medium has produced phenomenal creation. Recently some websites show music related information, they are related to latest hit songs lacking any form of exploration feature. Our aim for this project is to build visualization for exploration and analysis of information related to music. Music enthusiasts want to explore songs of diverse taste. Our visualization will enable the end users (music enthusiast) to explore and learn facts about music of different geographical areas, cultures and genres. This will make an ultimate search experience for a music enthusiast. Currently this kind of information can only be accessed after an extensive google search. For example it is elusive to find any information about an artist or a genre based on locations. We would like to present MusicAtlas (@music\_atlas) as a one-stop exploration and analysis tool for Music lovers.

The aim of this project is to provide user interface to educate, explore and analyze information related to music. The user should have the flexibility to explore the music, artist or release data and see what the trend is with respect to location. We also want to show artist history, popularity etc. to see how the trend is changing by changing the parameter, the same could be observed for tracks or company. The musicbrainz dataset also has time dimension attached to various entities, we want to leverage that to see the timeline trend for given entity (like company, artist, track etc).

The design was bit inspired from the assignments in class as well as **SiggraphPubVis** from last year's hall of fame. We strive to provide something as simple and useful to the users.

### 3 Questions

Our project aims to answer the following questions:

- Which country has maximum track released history till date.
- When the country was more active in the music releases(albums) filtered according to the year and track length.
- Tracks released by which artist in which release with other information like when(which year), length of the track and language in which the track was recorded.
- User should be able to search the artist, track and album in various popular sites like youtube, wiki and last.fm respectively (which has the massive database ).
- User should be able to see the number of tracks released by an artist.
- Genre which are most popular in a country also filtered to see for a particular year range.

### 4 Data

We have taken the music data from MusicBrainz( <http://musicbrainz.org/>) open source database to answer interesting questions about the music world. We have collected the data from the bitbucket instruction page by following the instructions provided by the MusicBrainz for setting up the Postgres database of the entire dataset.

**Database:** [https://musicbrainz.org/doc/MusicBrainz\\_Database](https://musicbrainz.org/doc/MusicBrainz_Database)

**Server Repository:** <https://bitbucket.org/lalinsky/mbslave>

Some of the dimensions from MusicBrainz database we are using are Artists, Company/Label, Tracks, Release/Album, Area etc. This database is able to reveal interesting patterns and trends in the past that can help user to know better or play around with music records.

**Source:**

- index.html
- main.js
- choroplethmap.js
- yearchart.js
- tracklength.js
- wordcloud.js
- tablechart.js
- scaleslider.js

Below is a list of libraries and files that we downloaded from the internet to beautify and implement some functionality.

**Libraries:**

- d3\_v3.js
- d3\_v3.layout.cloud.js
- jquery-3.1.1.min.js
- modernizr-2.6.2.min.js
- pdfobject.min.js
- colorbrewer.js
- d3.v4.min.js
- topojson.js

Flow of data that we used from the entire database in our project.



## 5 Exploratory Data Analysis

We expect our raw data to require a substantial cleanup process. So far we have created the whole database of the MusicBrainz, totalling compressed 9.1GB of data ( after expending more of around 50GB). The dataset which we have chosen is huge and in the form of relational tables which are linked with each other using keys. Also, the tables have various fields and meta-data information about the entities which we might not be using for our visualization. For extracting the data from relational tables, we have created the Python data server (using Flask) and wrote required APIs to get the relevant data in the JSON format. The data server APIs are written in Python using libraries like pickle, ujson etc.

Finding the relation between the tables and the column structures was the most difficult part of this process. As this dataset is from 18th century to present, we are unaware about the authenticity of the data, so we also did some YouTube and Google verification of the data. The schema ERD(Entity relationship diagram) defined on the MusicBrainz website gave us initial pump to represent everything in timeline manner as proposed in the project proposal, but after analysis we figured out that most of the "Year" data is missing with respect to the main entities like Artist, Label, Track etc.

Our data acquisition and processing pipeline is as follows:

- We started by going through the instruction page step by step for downloading the data, creating the schema and importing the data into the schema.
- We wrote various joined complex relational data queries to get the required data in single query.
- To make the queries faster, we have created temporary tables in the existing database with customized indexing and aggregate tables and groups for faster results. (eg. the joined query consists of 5-6 tables with each having around 1 million record, therefore the queries were slow at first).
- Next, we created the python WSGI server and wrote APIs in Flask in JSON format .
- In the APIs we have to filter out the data which had an empty year column and had to randomize the track length if the length is not available.
- As our whole data is linked to github and it supports SSL connection, so we had to update our server too to support SSL. Now all of our APIs are secured.

At the moment, we have finished extracting the essential information( artists, track, release, language relation) from our database, and produced working JSON APIs that have been used in the prototype. Below is an excerpt from artist, track, release relation JSON file:

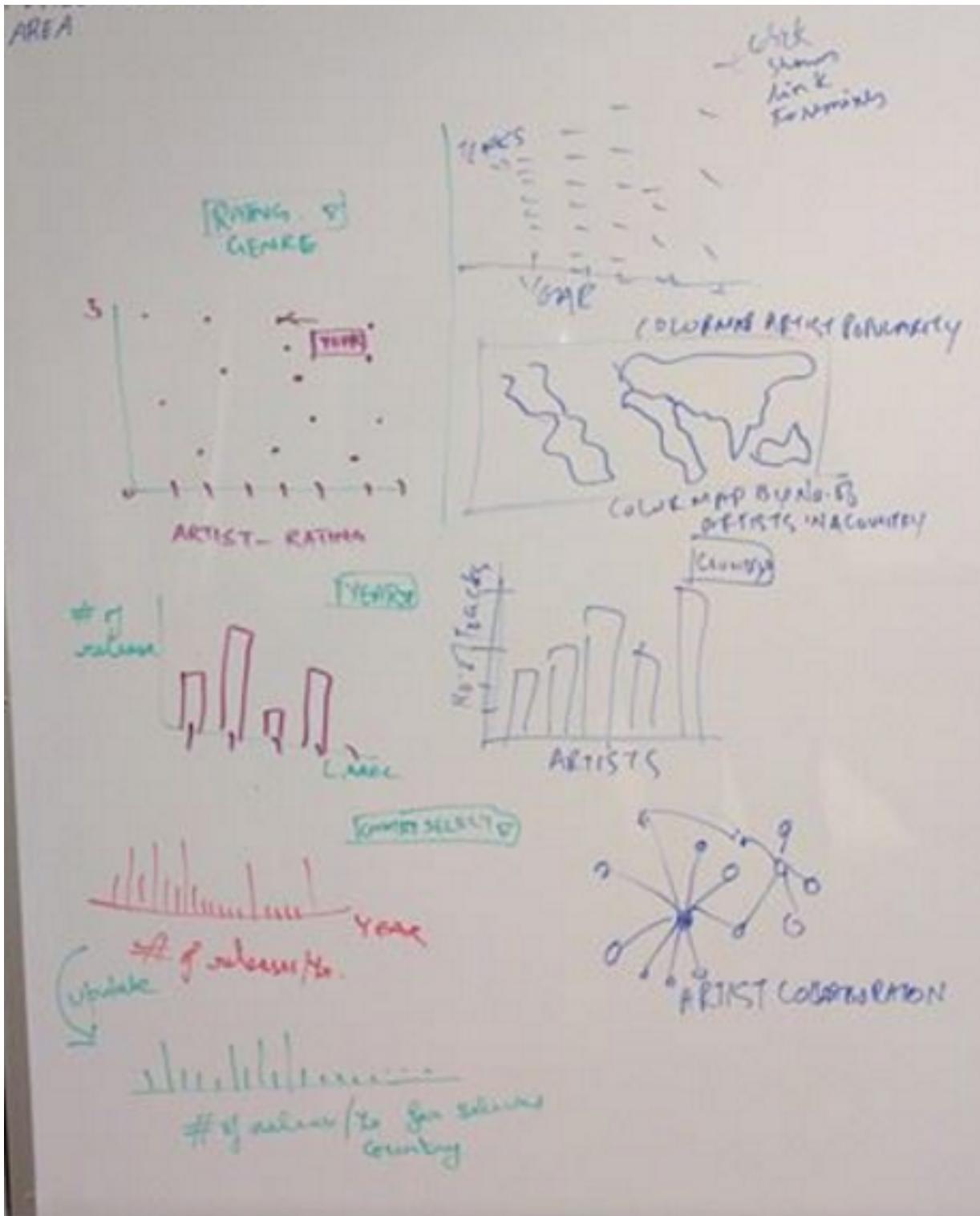
```
{"release_id":446162,"language":"English","gender":null,"country_id":"81","artist_name":"Crazy Frog","track_id":7189994,"country":"Germany","length":2.85,"track_name":"Axel F","year":2005,"artist_id":248914,"release_name":"Jamba Hits Volume 4"}
```

Apart from world map our design has changed almost completely as mentioned in the process book after the Peer feedback. Initially we were interested in doing the analysis as mentioned below:

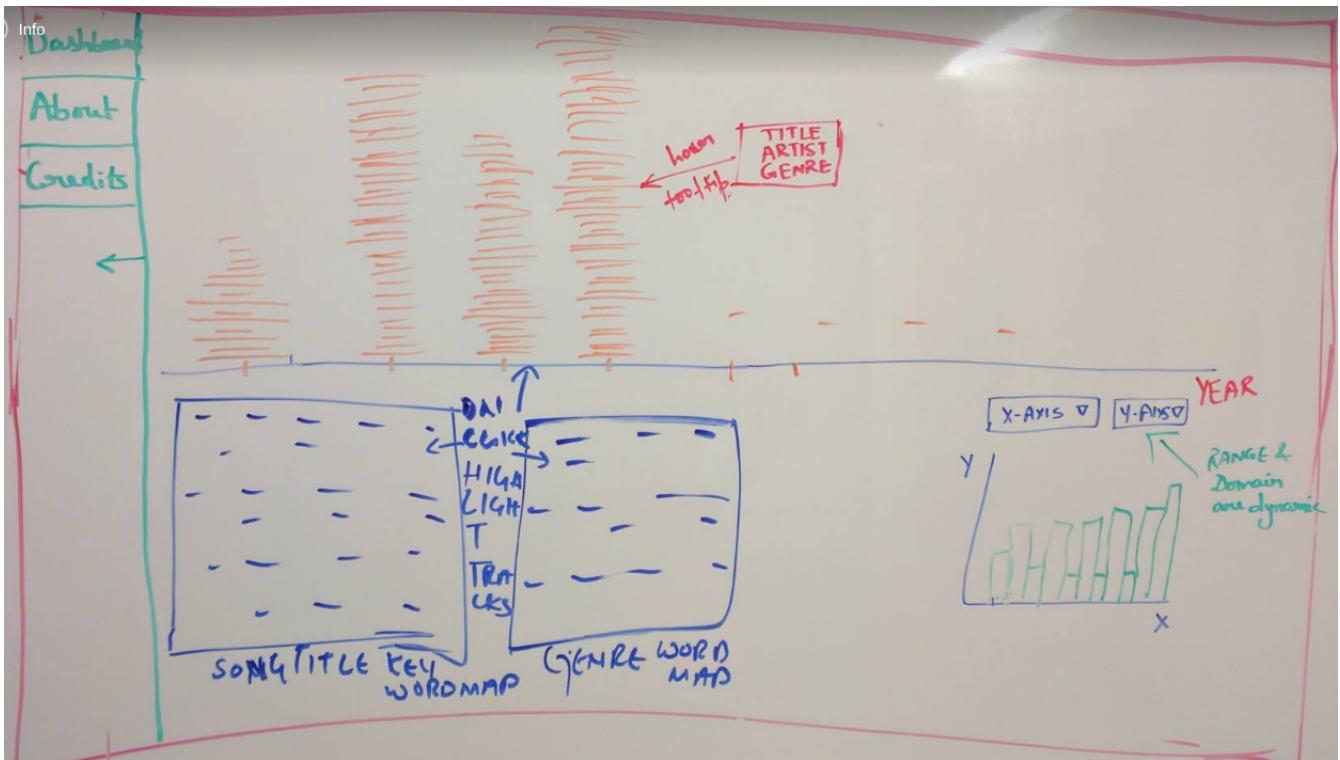
- A map which will be colored according to the popularity of the artist in a country
- A bar-chart representing the number of tracks released by an artist till date.
- A bar-chart showing the number of albums released by a label.
- An scatter plot which will show the artist rating with respect to the popularity and the genre, with artists on the horizontal axis and popularity with respect to year in the vertical axis represented as a dot.

Later we proposed some other visualizations as our second and third draft for analyzing the same data, but after the Peer Feedback, we figured out that a user is not interested in looking at the number of tracks released by an artist or number of artists associated with an artist, they are more interested in knowing which country has most active music industry or what are the different languages were used in those songs or able to know that artist of their favorite track and last but not the least be able to listen or search the track. Therefore, we discarded our proposal statistics design and re-designed the whole visualization from scratch by going through the database once again and verifying the data according to the current need.

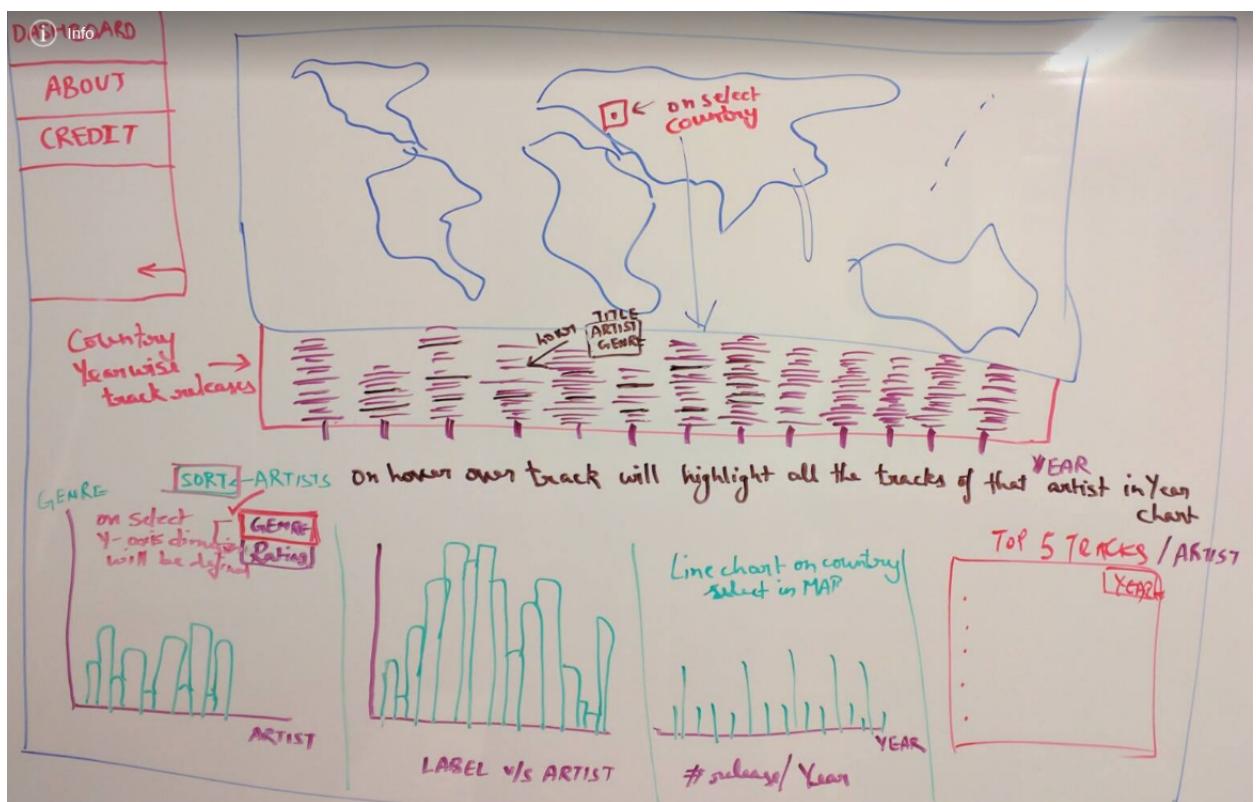
## 6 Design Evolution



This chart is the initial layout which shows multiple SVG's fulfilling different purposes. First, the Artist rating is shown in scatter-plot which on hover display the year in which that rating is marked. Second, On selecting the year from the drop-down, the bar chart will display the number of releases by the label companies in that year, by default one year details will be displayed. Third, On country select from the dropdown on the top-right side of that SVG, the number of releases done by any label aggregate is displayed with respect to year. By default the aggregate releases are displayed per year in line chart. Fourth, Track year scatterplot will be displayed in the top right side, on click, will show the link to the remix(but was complicated to get). Fifth, colored map with number of arists in country in world map and colorscale is proportional to the number of artists. Sixth, number of tracks released by artist in bar chart and at last the artists force map in node chart.



This design is more refined as compared to first design in managing the charts properly by providing the comparison axis selection in the right side in drop-down to render the below bar chart whose range and domain are dynamically loaded according to selection criteria in drop-down. Left-panel will have a Dashboard, About and Credits list which is collapsible. The main chart in this visualization is the year with track chart, where on hover on any track the tool-tip will be displayed with Full track title, artist and genre information and the track will be highlighted. Below, the song-title keyword map and Genre word map which will interact with the year-track chart mentioned above.

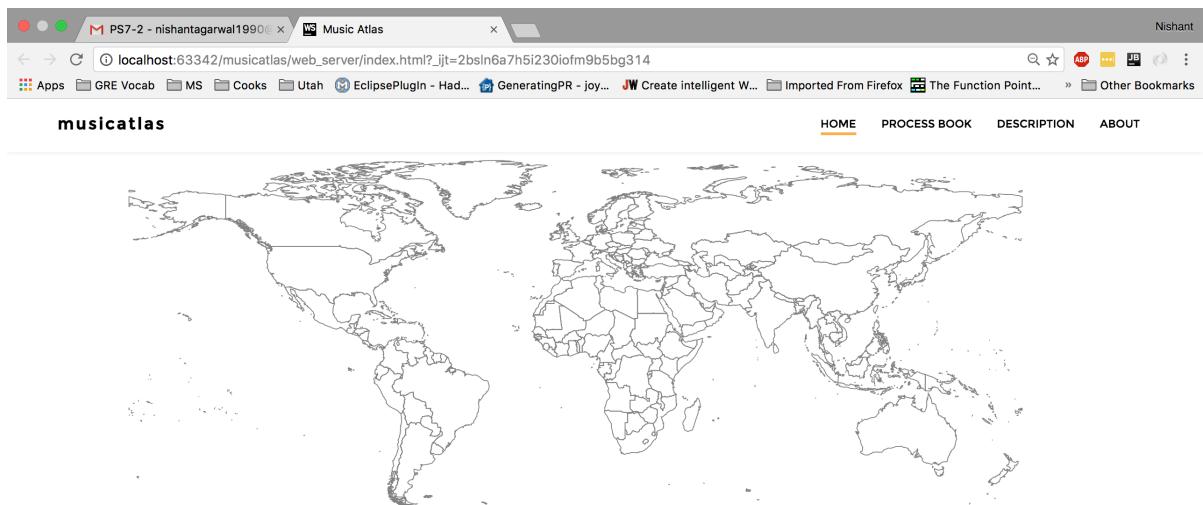


This design uses a map where countries are colored differently based on the number of tracks released. On selecting a country we display the tracks based on year below it. On clicking a track all tracks by that particular artist will get highlighted. Along with it we would show bar chart comparing artist based on genre and ratings, number of artists per label, number of releases per year for the selected country and the top 5 artists or tracks of that selected country. This is all aimed at providing the user with a burst of information which helps him/her decide on the kind of music they would possibly want to hear. We could also show a top 5 based on genre or other things. The ideas are quite infant at present.



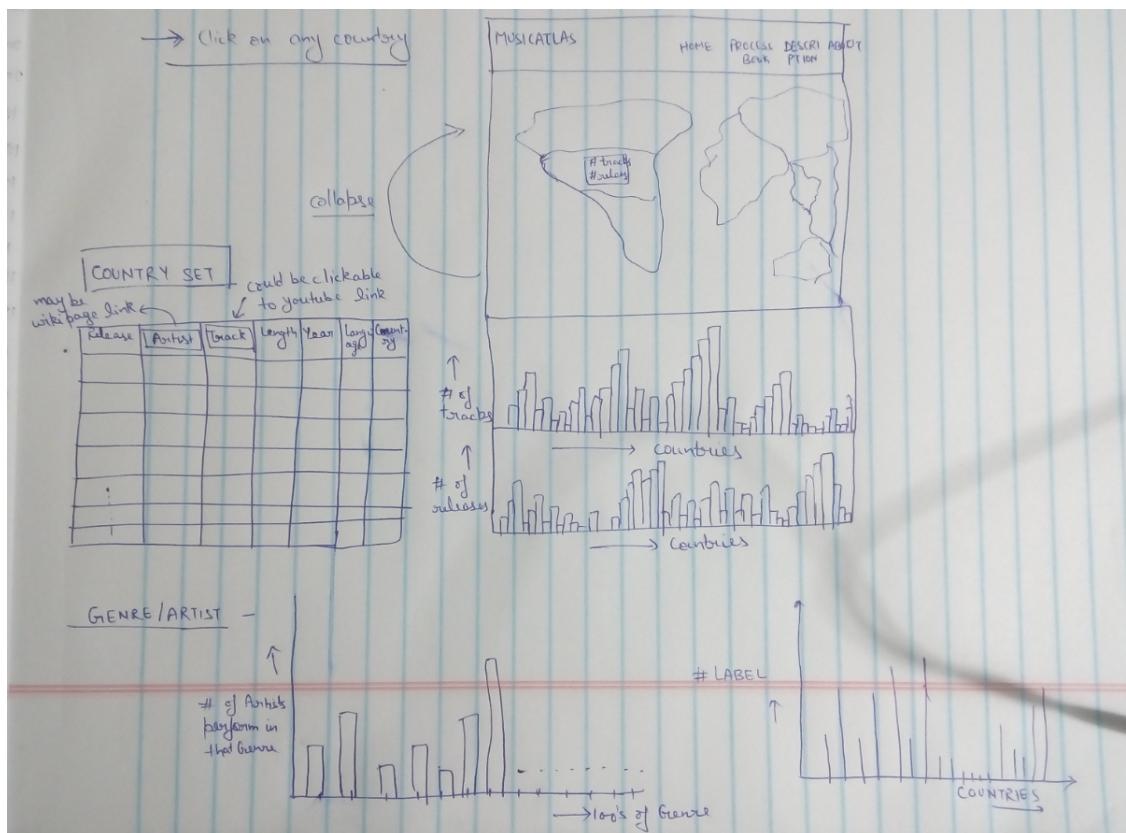
This design incorporates the best of the previous design to the best of our understanding. We would use a choropleth map displaying based on number of tracks per country. The bars below it is the representation of the number of tracks per country. On clicking the bar we would display the data in the charts below. Our display would show labels vs artists. A multi-dimension or choice bar chart for artist vs genre/popularity. We are yet to decide based on data if it would be possible to show a timeline for an artist. This data provides a good representation of the information and would amuse a music enthusiast. They would be able to see most sought after labels by artists, the genres most popular by artists etc.

## Initial WebPage



Expected design proposed in the project milestone :

This design is updated after the Peer review. The team which reviewed our project was not familiar with the dataset we have. So, they were kind of interested in listening to old songs or able to see sort and search. In this design, we took the feedback and updated our design by adding the table with youtube and wiki searchable tracks and artists, along with statistics mentioned in the earlier design.



## Final Design after Peer review and Mentor Feedback:

After the mentor feedback and letting her know about the peer review we have received, we all decided to remove the statistics and involve more interaction in the visualization. For new and innovative look, we have decided to involve interaction of the table with the word cloud. We have made all the columns in the table sortable. Also, we make our tracks, artists and albums searchable through external links. This design is more involved and inter-related as all the visualizations are inter linked and interact with each other.



## 7 Implementation

Our project includes multiple views like header, map view, year view, track length slider, table view and wordle view.

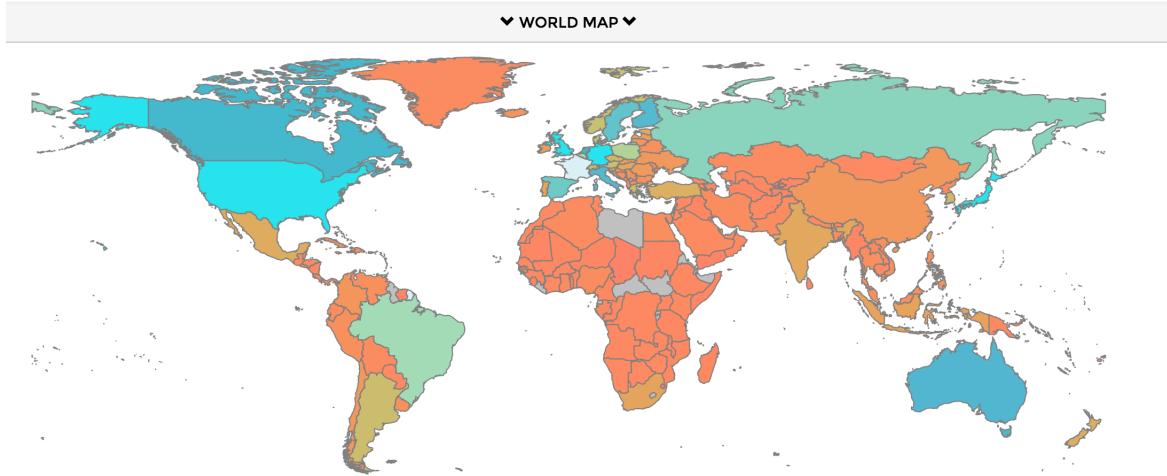
### 7.1 Header View

We decided to make it simple and space efficient header and cool as per our application. This design is mentioned below :

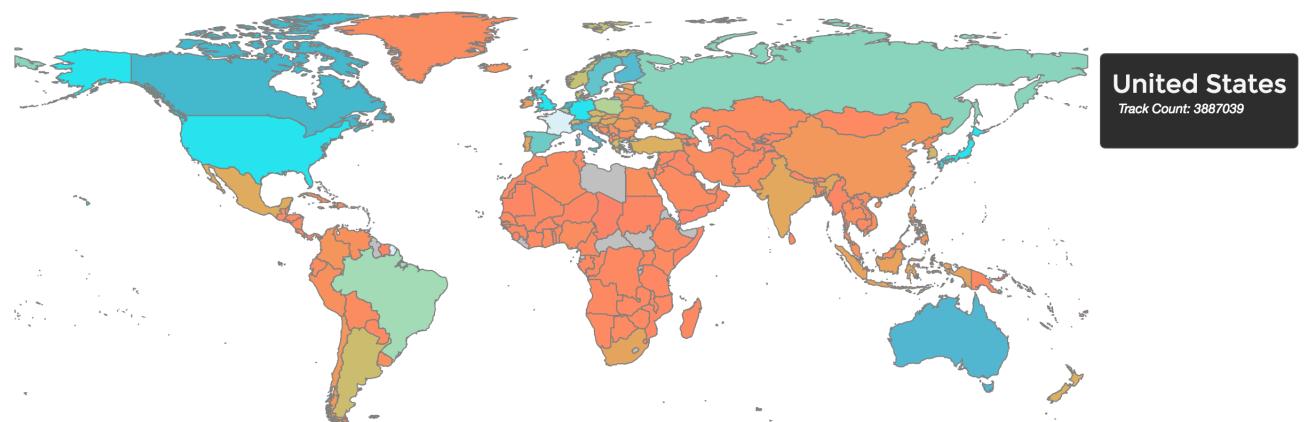


### 7.2 Map View

In this view, we have shown the world map with each country details. Each country is coloured according to the number of tracks released in that country till 2016. The track count not uniformly distributed among the countries, therefore, we have created multi range hues according to the number of track distribution and later created color scale including all of the above hues.



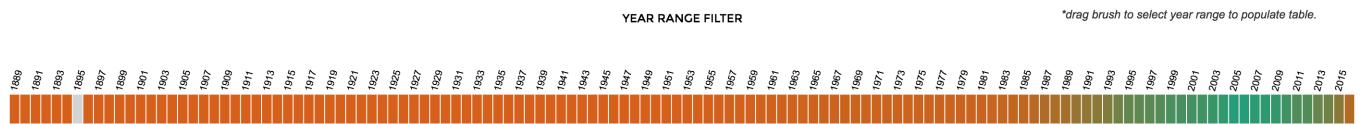
Later, we have added the tool-tip on the map on hover over any country with information like country name and track count released till date.



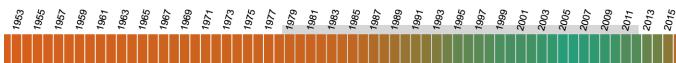
\*click on the country to display dashboard containing details of music.

### 7.3 Year View

In this view, after selection of country from the world map, the year stacked bar chart is created dynamically with respect to country from minimum year (when the first track release detail is available ) to maximum year ( until the track release detail is available). For coloring the bars which represent each year are based upon track count from the “colorbrewer.Dark2”, if year is available else its gray.

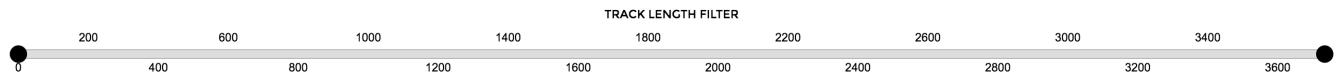


Later to make it more interactive, we have provided the brush feature over the year with year value mentioned over each bar.



## 7.4 Track Length Slider

In this view, after selection of country, for all the tracks released in that country, the slider is created with range from minimum to maximum track length. Also, it has length label with helps in selection for slider to drag and interact.



In the present study, the didactic material included a classification of the main features of the different types of



## 7.5 Table View

This view contains the whole information about the track, artist, release and other in the most easily readable format. The columns of the tables and their order are chosen very carefully, so that it will be easily understandable. Keeping in mind that the length of the tracks and artists are variable so the column will increase with the text, therefore all the integers like year and length are shifted on the right side. Language is pushed to the end of the column because there is usually less variation in a single country.

ARTIST-TRACK-RECORD TABLE

Track↑↓	Artist↑↓	Album↑↓	Length↑↓	Year↑↓	Language↑↓
Wish I Could Fly	Youtube Roxette	Wiki Fusion Syndicate	last.fm 4.75	2012	English
Queen of Rain	Youtube Roxette	Wiki Fusion Syndicate	last.fm 4.97	2012	English
Forever (radio edit)	Youtube Here [in]	Wiki Forever	last.fm 2.67	2009	English
Forever (a cappella)	Youtube Here [in]	Wiki Forever	last.fm 2.15	2009	English
雲に映る歌	Youtube 種ともご	Wiki The Dust That Made the Fire That Made the Light	last.fm 5.64	2011	English
ゼイタクにKiss+Tears	Youtube 種ともご	Wiki The Dust That Made the Fire That Made the Light	last.fm 3.03	2011	English
St. John's Day Songs: No. 3. Can't Be Caught With a Gun	Youtube Veijo Tormis	Wiki Califuneral	last.fm 1.2	1991	English
St. John's Day Songs: No. 4. Why St. John Is Awaited	Youtube Veijo Tormis	Wiki Califuneral	last.fm 1.68	1991	English
Three Songs From the Epic: No. 3. The Wave Rolls	Youtube Veijo Tormis	Wiki Califuneral	last.fm 5.32	1991	English
Autumn Landscapes: No. 2. Clouds Are Racing	Youtube Veijo Tormis	Wiki Califuneral	last.fm 1.6	1991	English
Autumn Landscapes: No. 5. Wind Over the Barrens	Youtube Veijo Tormis	Wiki Califuneral	last.fm 1.15	1991	English
The Singer's Closing Words	Youtube Veijo Tormis	Wiki Califuneral	last.fm 4.4	1991	English
Nostalgi	Youtube Miss B	Wiki Pushing It Hard	last.fm 8.67	2004	English
No Words (Promo Only clean edit)	Youtube Charlie Wilson	Wiki Promo Only: Urban Radio, May 2006	last.fm 4.18	2006	English
Believe (Xenomania mix)	Youtube Cher	Wiki Promo Only: Rhythm Radio, December 1998	last.fm 4.33	1998	English
Never Bought It	Youtube Dinosaur Jr.	Wiki Hand It Over	last.fm 3.71	1997	English
Sure Not Over You	Youtube Dinosaur Jr.	Wiki Hand It Over	last.fm 4.16	1997	English
Partida	Youtube Cesária Évora	Wiki Cabo Verde	last.fm 6.28	1997	[Multiple languages]
Bô é de meu cretcheu	Youtube Cesária Évora	Wiki Cabo Verde	last.fm 3.68	1997	[Multiple languages]
Mãe velha	Youtube Cesária Évora	Wiki Cabo Verde	last.fm 4.77	1997	[Multiple languages]

◀ PREVIOUS 2 NEXT ▶

After the successful rendering of the table view, we have made it more interesting by proving sorting feature on all of the columns.

ARTIST-TRACK-RECORD TABLE

Track↑↓	Artist↑↓	Album↑↓	Length↑↓	Year↑↓	Language↑↓
The Marching Blue Ants	Youtube (Young) Pioneers	Wiki ABCs of Punk	last.fm 1.27	1997	English
Short Back & Sides	Youtube 2 Cold	Wiki Turntables on the Hudson, Volume 3	last.fm 5.08	2001	English
Inside Knowledge	Youtube 25 ta Life	Wiki Psycho Civilized	last.fm 3.43	1995	English
Hush	Youtube Acappella	Wiki Live From Paris	last.fm 5.18	2002	English
I Feel Good	Youtube Acappella	Wiki Live From Paris	last.fm 2.91	2002	English
Amazing Grace	Youtube Acappella	Wiki Live From Paris	last.fm 7.68	2002	English
Shut De Do	Youtube Acappella	Wiki Live From Paris	last.fm 2.96	2002	English
Victory in Jesus	Youtube Acappella	Wiki Live From Paris	last.fm 5.39	2002	English
Love is the Power	Youtube Acappella	Wiki Live From Paris	last.fm 2.51	2002	English
Roll Jordan Roll	Youtube Acappella	Wiki Live From Paris	last.fm 5.2	2002	English
Peace Be Still	Youtube Acappella	Wiki Live From Paris	last.fm 4.22	2002	English
Jesus Knows	Youtube Acappella	Wiki Live From Paris	last.fm 3.29	2002	English
Instrupella	Youtube Acappella	Wiki Live From Paris	last.fm 4.21	2002	English
Sweet Fellowship	Youtube Acappella	Wiki Live From Paris	last.fm 3.91	2002	English
This Little Light of Mine	Youtube Acappella	Wiki Live From Paris	last.fm 3.77	2002	English
Everybody Said	Youtube Acappella	Wiki Live From Paris	last.fm 3.11	2002	English
Holy City	Youtube Acappella	Wiki Live From Paris	last.fm 2.75	2002	English
Benign (live)	Youtube Action Patrol	Wiki ABCs of Punk	last.fm 1.54	1997	English
Riding the Waves (Mark Sinclair mix)	Youtube Afro Celt Sound System	Wiki Earth Dance 2000: The Global Dance Party for Peace	last.fm 5.33	2000	English
Aint No Pleasing You	Youtube Alex Campbell	Wiki Life Is Ugly So Why Not Kill Yourself	last.fm 4.27	2001	English

After getting the peer feedback in the peer review session, we decided to provide a searchable link for track, artist and album in the most popular websites like Youtube, Wiki and Last.fm respectively. These sites are chosen because they had vast database of information from all over the world. Below is the searchable youtube link view inside the track cell:

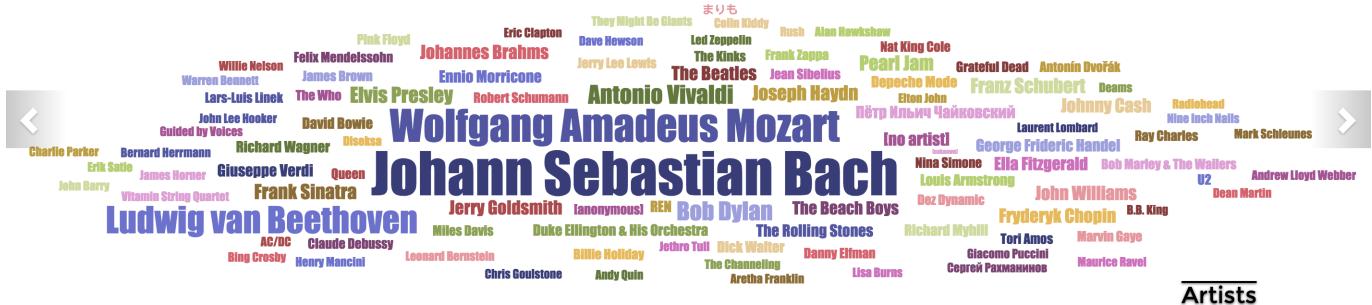
St. John's Day Songs: No. 4. Why St. John Is Awaited	Youtube	Veijo Tormis	Wiki	Califuneral	last.fm	1.68	1991	English
Three Songs From the Epic: No. 3. The Wave Rolls	Youtube	Veijo Tormis	Wiki	Califuneral	last.fm	5.32	1991	English
Autumn Landscapes: No. 2. Clouds Are Racing	Youtube	Veijo Tormis	Wiki	Califuneral	last.fm	1.6	1991	English

## 7.6 Wordle View

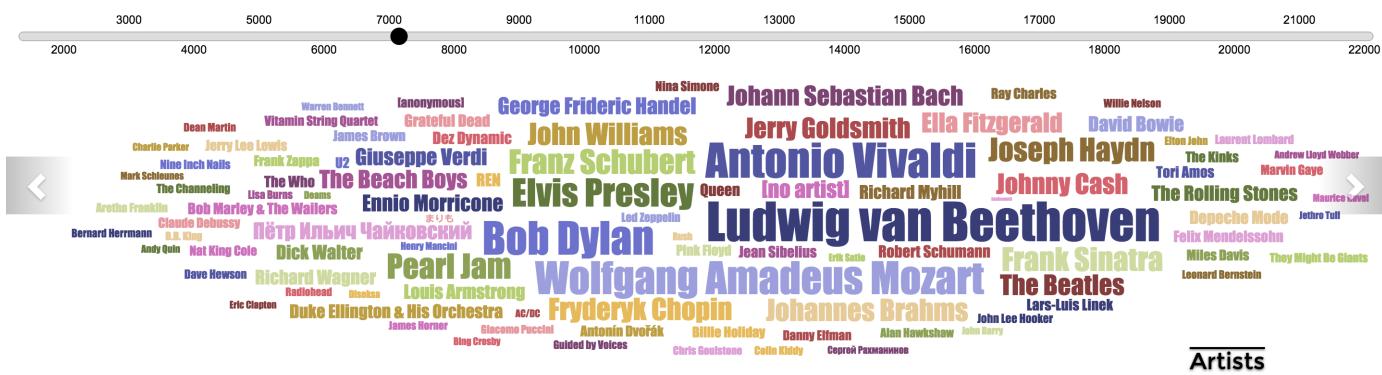
### 7.6.1 Artist Wordle

This view is most interesting for us, as this was new and has various interesting functionalities but also equally difficult to make it dynamic.

We have started with designing a static wordle of the top 100 artists in that country ( top 100 artists are fetched according to the number of tracks they have released) and depending upon their position the size and color of the text is decided.



Later, we have decided to make it more interesting by adding a slider to chose from, the slider contains the range from minimum number of tracks released in that country by an artist to maximum number of tracks released. As the slider changes the wordle gets updated.



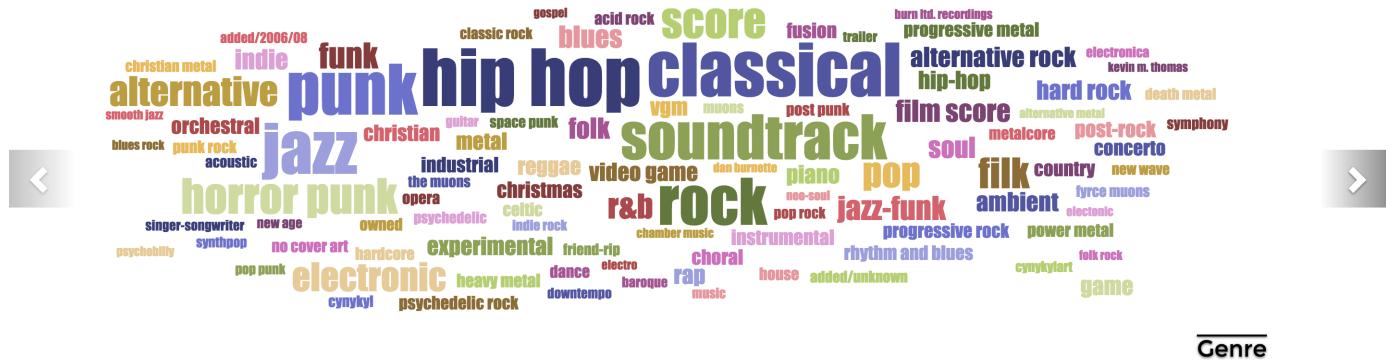
\*artist word cloud where size is scaled on track count and click updates table.

We have even provided a click functionality on artists appearing in the wordle to interact with the table view. The click allows for filtering of the table based on artist. This is possible with the other two filters working on top of it.

### 7.6.2 Genre Wordle

This view is the last minute evolution from the dataset to represent the popular genre in a country on the basis of albums released. Initially, we were unable to link the genre with the country as the details were not available to us directly. Then, after our visualization till the table view, we found out that we have country linked to releases and we have release labels with identifier as their release Ids.

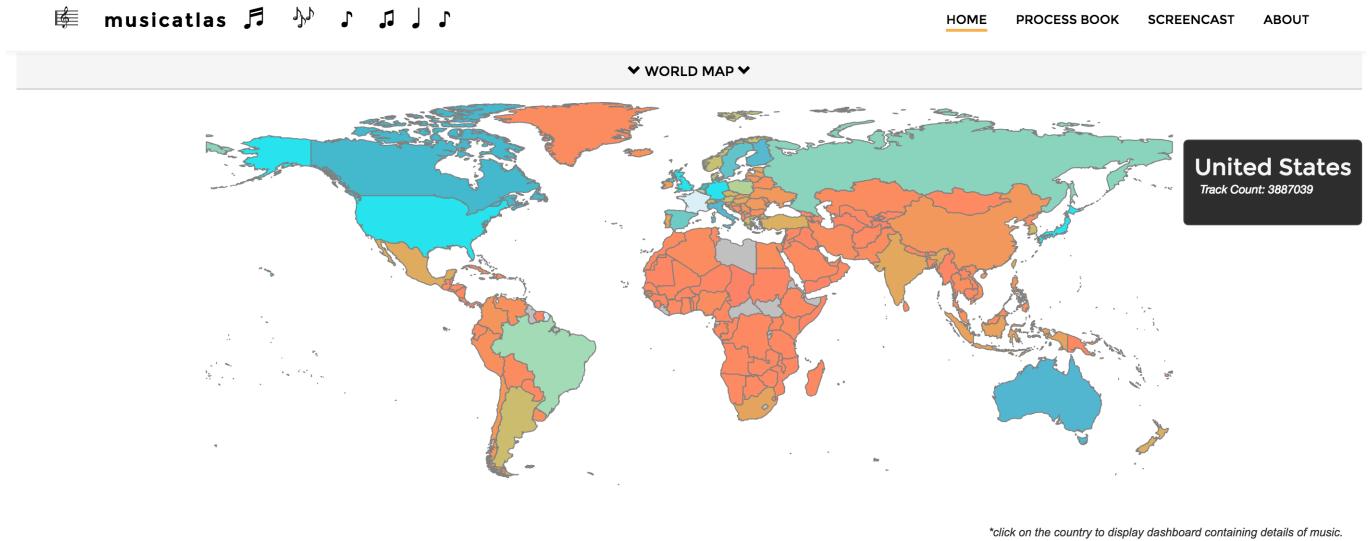
This wordle gets updated as the country changes. The size and the color of each genre is dynamically assigned on the basis of number of albums released in that category by that country till date.



## 8 Evaluation

In this section, we demonstrate a typical flow of our website, along with the details of the analysis and information extracted from the view.

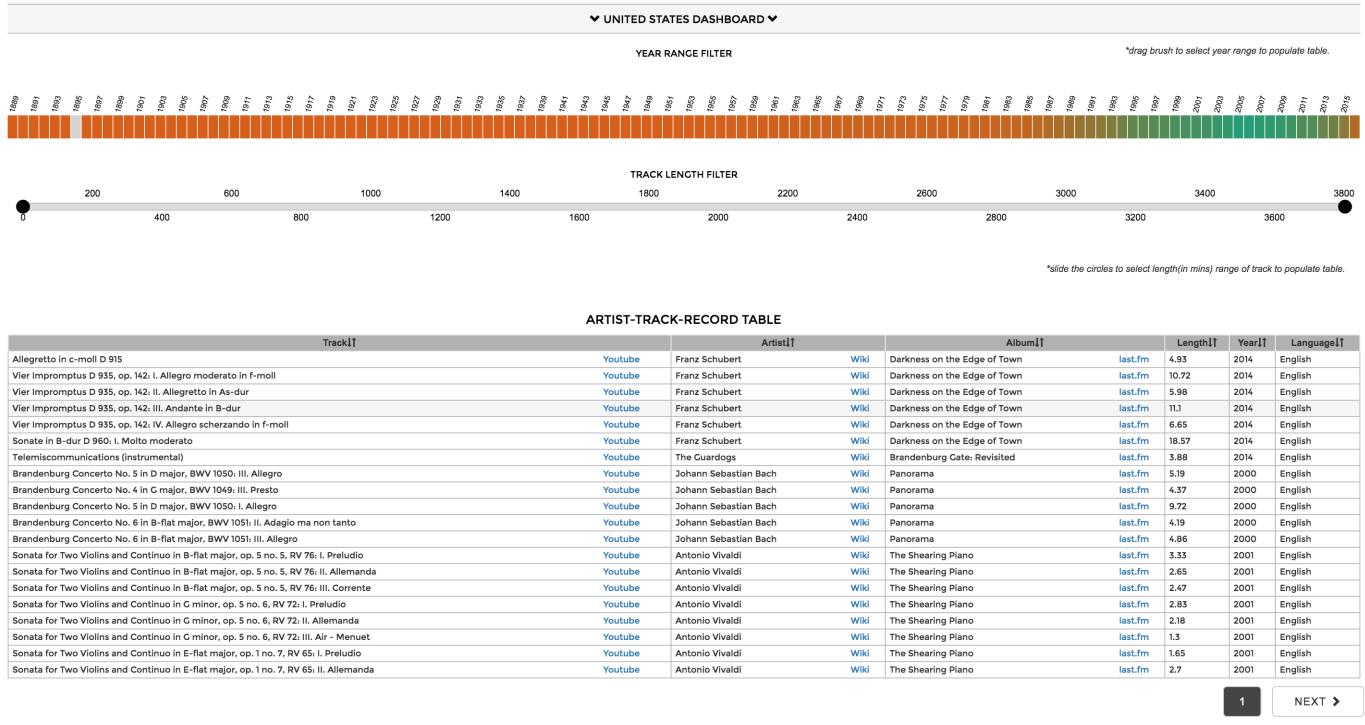
At first, the home page appears with the world map view, which has track count detail and country name on the tool-tip when mouse hovers. This view contain many information regarding the world music distribution according to the number of tracks produced. Coloring here is the most complicated part as the music industry is not uniform in the whole world. Therefore, as mentioned above various ranges are selected to color with hues and among all the range we have chosen different colors. The view can be seen below:



Second, after any country selection the dashboard display the name of the selected country and creates various visualizations which behave as dynamic filter over the tracks. Year stack bar, track length slider, table view represent the tracks released in the selected country.

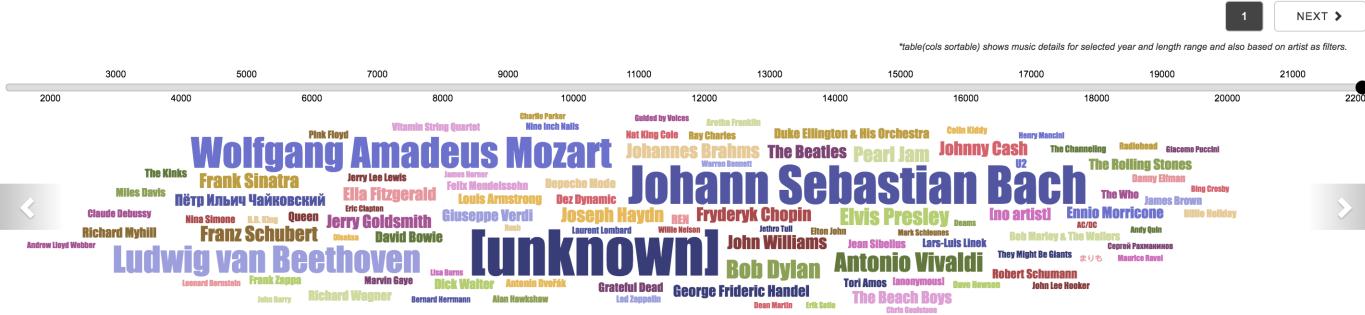
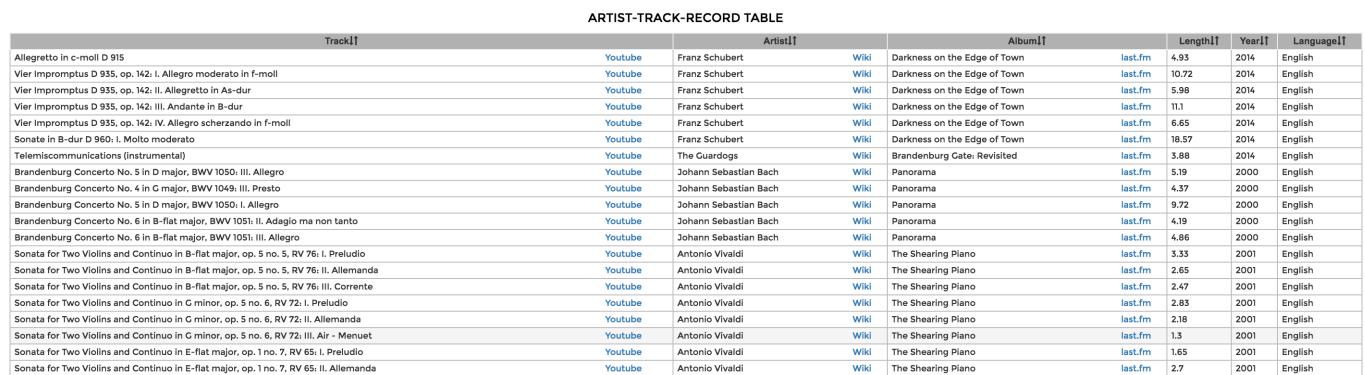
The initial table length is set as 20, to be easily visible on the screen. To see other records we have provided the pagination feature with previous, next button and a div contain the page number.

All the different visualizations are interactive with each other as filter over the table data and table has functionality of sorting over all the columns and external web links to search tracks, artists and albums. The view can be seen below:



Third, but not the last there is artist-track-release table interaction with the artist word cloud. Each text in the word cloud is the name of the top artist form the selected country with respect to tracks released. As the user click on any artist the table gets reloaded with the filtered artist.

This filter work with year and track length filter if the artists are clicked at last, but if the order of the filter changes then, the data is reloaded for all the artists. This is currently the limitation of the application because of the class object structure we have used to create the visualizations. Each object gets stuck in the deadlock.



The artist word cloud is associated in the carousel when click on the right arrow, the next word cloud of genre is displayed which is also designed dynamically according to the country selected. This word cloud is interesting as it gives directly the top genre preferred in a country.

## 9 Conclusion

We have shown that our tool provides several ways to navigate the space of music and artist that are related to a selected country. One can use either the filters or just browse through the data displayed on the table to know the interesting details about the track or country or artist and many more. On our way of exploring the data, we also learn more about the music popularity and taste of countries along with the artists origin and also find out same title song has released by many artists in almost the near by years. All of these can work together to help the user quickly get a high-level understanding of the music taste and can get more by going through the external links provided in the table view.

## 10 References

- <http://bl.ocks.org/mbostock/6452972>
- <http://techslides.com/d3-world-maps-tooltips-zooming-and-queue>
- Class homeworks and slides.