Musical Borders, analysis of how music can contain and reflect aspects of societies from countries grouped by geographical location

Applied Data Science Capstone - IBM Data Science

Ion Garaicoechea

Zürich, 8th of August 2020

Content Index

1.- Introduction

Motivation. Nowadays, on this world so easily connected all around its globe, it is amazing how cultures can differ geographically by countries. This differences can be captured in different aspects of the society. It is not necessary even to leave the continent or go to the other corner of the world. The countries have their own manners and customs.

This also applies to the music. The way the population of a country listens to it and sets their musical affinities can reflect part of the country's lifestyle and way of thinking. Music styles, language of the lyrics, rhythm, key played,... and many more aspects that can show how some groups of population can express they selves through the music.

This study focuses in European countries, grouped and compared by its geographical position. The aim of it, is to verify and analyze this cultural differences stored in the most listened songs of each country. Clarify if it is true that the northern countries have a "saddest", "colder" or "depressant" impact on this aspect as the stereotypes estipulate. And if, for example, the Mediterranean countries have a "happier", "carefree" or "warmer" way of selecting the music they listen to.

Project Plan. Once the motivation of the project has been defined, this are the following tasks that have been identified to achieve a good argument and reasons to hold the statement above.

A trusted and accessible data source has to be selected. This, has to provide data which accomplish some requirements. It has to be splitted by countries and provide different features to work with and to evaluate different aspects of the country's lifestyle and culture.

After manipulating and filtering the raw data, once the desired data has been achieved a comparison between countries will be carried out, and conclusions will be deduct based in the performed analysis.

2.- Methods

Data Retrieval. The project focuses in statistics and features of different countries most listened songs ('*$country$* Top 50' playlists songs). For this purpose, the streaming music player *Spotify* has been chosen, as it is one of the most used software to reproduce music all around the world. This carries a very accurate tracking of the users and provides a good accessibility to developers. When accessing their data repository it can be obtained easily enough information to solve and expose arguments to clarify the motivation of this project.

In the very first step, two different repositories of data are used:

- *WDI:* List of the European countries: obtained from the WDI package by calling a list 2018 and filtering it for 'region == Europe & Central Asia' and selecting the 'Country' column.
- *Spotify API:* As we are working with Spotify, the data retrieval of the playlists will be done with the very useful 'SpotifyR' package.

Retrieval of 'Top 50' playlists for each country: Firstly, it has been taken from the user 'spotify' as many 'Top' playlists as possible, but as not all countries have one, other countries playlists have been obtained from the user 'top 50 playlists'. The retrieval process for both is practically the same:

- Get a list with the users playlists (using '*get_user_playlists()*').
- Filter this list for the desired countries or by continent, so we have the 'Top' playlists of interest.
- Get the tracks of every playlist contained in the filtered list (using 'get_*playlist_tracks()*').
- Get the features SpotifyR provides for each song of each playlist in the filtered list (using '*get_track_audio_features()*'). This provides for each song a score on danceability, valence, energy, speechiness, acousticness, instrumentalness, liveness and tempo.

Data Processing. The data obtained through this steps is huge if it is not filtered. Taking the scope of the project in mind, some features as danceability, valence and energy will be selected to make an analysis and comparison between countries.

After merging, filtering and grouping the data in the different obtained tables for each of the source-users, this is the head of the resulting data frame, showing the mean of every mentioned feature '*spotifyr*' provides for the songs contained in the 'Top' playlist of each country:

	playlist_name	valence_mean	danceability_mean	energy_mean	speechiness_mean	acousticness_mean	instrumentalness_mean	liveness_mean	tempo_mean
1	Austria Top 50	0.5930340	0.7374000	0.6780200	0.19758200	0.2515760	0.0048844896	0.1304580	120.4512
2	Belgium Top 50	0.4297740	0.6908800	0.6169600	0.12149000	0.3074070	0.0047432046	0.1762300	120.8098
3	Denmark Top 50	0.5056000	0.7314000	0.5946600	0.11975000	0.2781680	0.0049168778	0.1506580	119.8478
4	Estonia Top 50	0.4636440	0.7030200	0.5683200	0.14913200	0.2974872	0.0421011954	0.1612020	125.7036
5	Finland Top 50	0.4647800	0.6838600	0.6506400	0.11261400	0.2251935	0.0161018330	0.1605700	122.2950
6	France Top 50	0.4246140	0.7222600	0.6383800	0.12701200	0.3565454	0.0002292104	0.1418940	124.8059
7	Germany Top 50	0.5968740	0.7396600	0.6794600	0.19139600	0.2431640	0.0023228816	0.1316480	120.9258
8	Iceland Top 50	0.4838780	0.7151200	0.5096380	0.13452400	0.3324541	0.0187852538	0.1760940	117.9019
9	Ireland Top 50	0.4666460	0.6967600	0.5653200	0.13090400	0.3243662	0.0286476826	0.1575000	119.1785
10	Italy Top 50	0.4692000	0.6857000	0.6655200	0.11568200	0.2256980	0.0048457308	0.1619000	116.2678

Table 1. Sample of resultant data frame with all 'Top' playlists features mean

For the analysis, this data set will be splitted into two different ones by geographical location. (Remark: At a first moment the split was done in three groups (differentiating Central European countries as well, as seen below) but it was found that Southern and Central countries had similar records on the features so finally for an easier comparison it was decided to mix the Southern and Central in the same group:

- <u>Northern European countries (10):</u> Denmark, Estonia, Finland, Germany, Iceland, Latvia, Lithuania, Norway, Poland and Sweden.
- <u>Central European countries (13):</u> Austria, Belgium, Ireland, Luxembourg, Netherlands, Switzerland, United Kingdom, Czech Republic, Hungary, Liechtenstein, Luxembourg, Slovakia and Ukraine.
- <u>Southern European countries (12):</u> France, Italy, Portugal, Spain, Andorra, Bulgaria, Cyprus, Greece, Malta, Monaco, Romania and Turkey.

On this geographical split, each data frame is not grouped by playlist, so information of each song is contained and will be used in the following task for the analysis.

For a more detailed view of data processing, see '*main.R*' code in the Appendix.

3.- Results

Analysis. Two approaches have been done. First, there has been added a tuple with the overall mean for each feature to each of the grouped countries data frames (*'sc_countries'* and *'northern_countries',* see in Appendix 6.1.4).

Second, taking the information of every song retrieved for each group of countries in the data frames, '*stacked percent barplots*' have been computed for a better visual comparison of the data, which show the distribution of feature values for each song. The selected features for this purpose have been the *valence*, *danceability*, *energy* and *tempo* features (where all are ranked from 0-1 except the *tempo*). These have been chosen considering that are which provide more representative information about each group of countries preferences.
(Remarks: *SC* refers to southern and central European countries. This results were obtained with the songs in playlist the 16th of April of 2019).

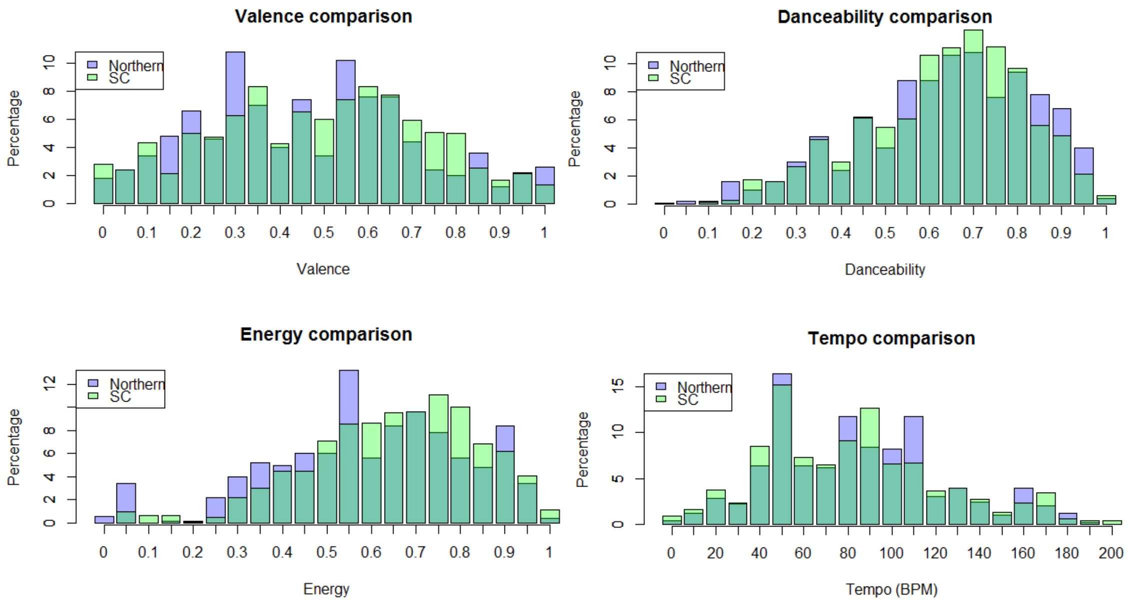

Figure 1. Stacked percent barplots of songs features for data visualisation

4.- Conclusion

First approach. It has been observed in Northern countries that in features as *valence* and *energy*, only 2 out of 10 countries are above the European overall mean value for each feature respectively. On the other side, in the southern countries it is the other way, where only 2 are below the European mean in terms of *danceability* (see Appendix 1.4).
Second approach. There is not a clear pattern that shows the prevalence of the central-southern countries as most *energetic* or *happiest*. Although, it can be seen in Figure 1 for the plots corresponding the *valence* and *energy*, how there is a slight tendency of south-central songs of highest values of this features. Also the SC songs, seem to be more danceable in average as the northern countries have a more splitted values along the feature.

Summarizing, it has been obtained some results that beyond some limitations of the analysis, there are some aspects stored in the music each country listens to, that reflects and represents different kind of societies with the stereotypes that were stated in the motivation of the project.

Limitations. This study has an important limitation, which is the scale of its application. The information stored in the 50 most listened songs for each country is not the best representation of a country or society, but it helps to make a picture and obtain some previous results of what could be a bigger scale study, showing the difference of features values along European countries. Also if some statistical concepts would have been applied to the data, there could have been obtained some more precise evidences of this research.

5.- References

- SpotifyR package - R documentation, "*https://www.rdocumentation.org/packages/spotifyr/versions/1.0.0* " , (last access on 16/04/2019)
- *Tracks Features - Spotify API Documentation*, "*https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/* ", (last access on 16/04/2019)
- *Barplot - R graph gallery*, "*https://www.r-graph-gallery.com/barplot/* " , (last access on 16/04/2019)

6.- Appendix

6.1.-Data Frames

6.1.1.- Europe

Data frame containing all the 'Top' playlists taking into account in the study. See for each of the playlists the mean of each feature for all the songs contained in the playlist. Also see 'EUROPEAN MEAN' in playlist_name where the mean of all is shown for each feature.

	playlist_name	valence_mean	danceability_mean	energy_mean	speechiness_mean	acousticness_mean	instrumentalness_mean	liveness_mean	tempo_mean
1	Austria Top 50	0.5890740	0.7416800	0.6732200	0.19834200	0.2497360	0.004884623	0.1441780	119.4081
2	Belgium Top 50	0.4301140	0.6925400	0.6090200	0.12561800	0.2926850	0.007162856	0.1742080	123.6217
3	Denmark Top 50	0.5106800	0.7338600	0.6018600	0.12031000	0.2685480	0.005328878	0.1547380	119.1591
4	Estonia Top 50	0.4546040	0.7063800	0.5711200	0.14949600	0.2877987	0.050721013	0.1564300	126.6132
5	Finland Top 50	0.4647800	0.6838600	0.6506400	0.11261400	0.2251935	0.016101833	0.1605700	122.2950
6	France Top 50	0.4337340	0.7280600	0.6332200	0.13271800	0.3516454	0.002067010	0.1400940	123.9268
7	Germany Top 50	0.5782740	0.7447200	0.6736200	0.18763400	0.2326100	0.002322924	0.1323240	119.7449
8	Iceland Top 50	0.4792380	0.7194400	0.5037980	0.13613000	0.3269141	0.018785254	0.1742320	117.4223
9	Ireland Top 50	0.4632060	0.6983600	0.5659000	0.13075000	0.3195982	0.028649571	0.1537460	119.3029
10	Italy Top 50	0.4719540	0.6840800	0.6656200	0.11766400	0.2367560	0.004859225	0.1641540	118.6625
11	Latvia Top 50	0.4587900	0.7170400	0.5768200	0.16654600	0.3159340	0.019325678	0.1652860	123.9115
12	Lithuania Top 50	0.4526260	0.7014400	0.5828800	0.16909000	0.2780140	0.031265568	0.1862360	119.6304
13	Luxembourg Top 50	0.4564540	0.7397000	0.5993800	0.12654800	0.3432880	0.016693927	0.1434240	118.2094
14	Netherlands Top 50	0.5101200	0.7349600	0.6322400	0.13061400	0.2317958	0.005345458	0.1860660	119.6665
15	Norway Top 50	0.4192400	0.6778200	0.6071800	0.10077400	0.2544120	0.005658749	0.1752160	118.3367
16	Poland Top 50	0.4779180	0.7125000	0.6287200	0.17125600	0.2873434	0.010459647	0.1873480	125.1258
17	Portugal Top 50	0.5028600	0.7565200	0.6101800	0.14505800	0.3079160	0.011011761	0.1428400	116.9428
18	Spain Top 50	0.6377400	0.7465600	0.6702400	0.12980400	0.3116560	0.002273517	0.1272260	121.2425
19	Sweden Top 50	0.4650000	0.6858600	0.5983660	0.12298800	0.2925388	0.005136101	0.1599460	116.4810
20	Switzerland Top 50	0.5564140	0.7302400	0.6594200	0.16666400	0.2696620	0.004803388	0.1429020	116.4931
21	Turkey Top 50	0.4738200	0.7079000	0.6316800	0.14287800	0.2820260	0.061031655	0.1591720	127.1889
22	United Kingdom Top 50	0.5334600	0.7350800	0.6027800	0.16163800	0.2912192	0.007740541	0.1394960	116.1689
23	ANDORRA Top 50	0.5114750	0.6969062	0.6327187	0.10427969	0.2921193	0.034052356	0.1668422	119.8814
24	BULGARIA Top 50	0.4546795	0.6809554	0.6442143	0.10318125	0.2403630	0.012999799	0.1687286	120.7933
25	CYPRUS Top 50	0.5033215	0.6870380	0.6739494	0.07984684	0.2025971	0.038481441	0.1802266	121.3387
26	CZECH REPUBLIC Top 50	0.4760065	0.6722742	0.6205323	0.09011452	0.2702170	0.028211068	0.1714210	119.1906
27	GREECE Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
28	HUNGARY Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
29	LIECHENSTEIN Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
30	LUXEMBOURG Top 50	0.5027377	0.6936230	0.6594918	0.08062459	0.2301851	0.034555800	0.1574557	118.9280
31	MALTA Top 50	0.5702200	0.6875600	0.7897000	0.08974800	0.1018364	0.014827487	0.2123120	120.5726
32	MONACO Top 50	0.3942960	0.6906000	0.7162200	0.14500200	0.1187618	0.084445617	0.1906680	128.0375
33	ROMANIA Top 50	0.5230400	0.6833000	0.6573400	0.09063200	0.2693286	0.002646221	0.1783440	117.4616
34	SLOVAKIA Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
35	UKRAINE Top 50	0.5392900	0.7071600	0.6826400	0.12005200	0.2522878	0.022736335	0.1621460	119.8450
36	EUROPEAN MEAN	0.4904068	0.7060348	0.6300203	0.12559311	0.2678918	0.019984653	0.1634311	120.4117

Figure 2. 'Europe' data frame with all features mean for each country.

6.1.2.- northern_countries

Same as in 'Europe' data frame but only taking the Northern European countries. 'European Mean' is maintained with the same values for comparison aspects.

	playlist_name	valence_mean	danceability_mean	energy_mean	speechiness_mean	acousticness_mean	instrumentalness_mean	liveness_mean	tempo_mean
1	Denmark Top 50	0.5106800	0.7338600	0.6018600	0.1203100	0.2685480	0.005328878	0.1547380	119.1591
2	Estonia Top 50	0.4546040	0.7063800	0.5711200	0.1494960	0.2877987	0.050721013	0.1564300	126.6132
3	Finland Top 50	0.4647800	0.6838600	0.6506400	0.1126140	0.2251935	0.016101833	0.1605700	122.2950
4	Germany Top 50	0.5782740	0.7447200	0.6736200	0.1876340	0.2326100	0.002322924	0.1323240	119.7449
5	Iceland Top 50	0.4792380	0.7194400	0.5037980	0.1361300	0.3269141	0.018785254	0.1742320	117.4223
6	Latvia Top 50	0.4587900	0.7170400	0.5768200	0.1665460	0.3159340	0.019325678	0.1652860	123.9115
7	Lithuania Top 50	0.4526260	0.7014400	0.5828800	0.1690900	0.2780140	0.031265568	0.1862360	119.6304
8	Norway Top 50	0.4192400	0.6778200	0.6071800	0.1007740	0.2544120	0.005658749	0.1752160	118.3367
9	Poland Top 50	0.4779180	0.7125000	0.6287200	0.1712560	0.2873434	0.010459647	0.1873480	125.1258
10	Sweden Top 50	0.4650000	0.6858600	0.5983660	0.1229880	0.2925388	0.005136101	0.1599460	116.4810
11	EUROPEAN MEAN	0.4904068	0.7060348	0.6300203	0.1255931	0.2678918	0.019984653	0.1634311	120.4117

Figure 3. 'Northern_countries' data frame with all features mean for each country.

6.1.3.- sc_countries

Same as in 'Europe' data frame but only taking the Central and Southern European countries. 'European Mean' is maintained with the same values for comparison aspects.

	playlist_name	valence_mean	danceability_mean	energy_mean	speechiness_mean	acousticness_mean	instrumentalness_mean	liveness_mean	tempo_mean
1	ANDORRA Top 50	0.5114750	0.6969062	0.6327187	0.10427969	0.2921193	0.034052356	0.1668422	119.8814
2	Austria Top 50	0.5890740	0.7416800	0.6732200	0.19834200	0.2497360	0.004884623	0.1441780	119.4081
3	Belgium Top 50	0.4301140	0.6925400	0.6090200	0.12561800	0.2926850	0.007162856	0.1742080	123.6217
4	BULGARIA Top 50	0.4546795	0.6809554	0.6442143	0.10318125	0.2403630	0.012999799	0.1687286	120.7933
5	CYPRUS Top 50	0.5033215	0.6870380	0.6739494	0.07984684	0.2025971	0.038481441	0.1802266	121.3387
6	CZECH REPUBLIC Top 50	0.4760065	0.6722742	0.6205323	0.09011452	0.2702170	0.028211068	0.1714210	119.1906
7	France Top 50	0.4337340	0.7280600	0.6332200	0.13271800	0.3516454	0.002067010	0.1400940	123.9268
8	GREECE Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
9	HUNGARY Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
10	Ireland Top 50	0.4632060	0.6983600	0.5659000	0.13075000	0.3195982	0.028649571	0.1537460	119.3029
11	Italy Top 50	0.4719540	0.6840800	0.6656200	0.11766400	0.2367560	0.004859225	0.1641540	118.6625
12	LIECHENSTEIN Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
13	Luxembourg Top 50	0.4564540	0.7397000	0.5993800	0.12654800	0.3432880	0.016693927	0.1434240	118.2094
14	LUXEMBOURG Top 50	0.5027377	0.6936230	0.6594918	0.08062459	0.2301851	0.034555800	0.1574557	118.9280
15	MALTA Top 50	0.5702200	0.6875600	0.7897000	0.08974800	0.1018364	0.014827487	0.2123120	120.5726
16	MONACO Top 50	0.3942960	0.6906000	0.7162200	0.14500200	0.1187618	0.084445617	0.1906680	128.0375
17	Netherlands Top 50	0.5101200	0.7349600	0.6322400	0.13061400	0.2317958	0.005345458	0.1860660	119.6665
18	Portugal Top 50	0.5028600	0.7565200	0.6101800	0.14505800	0.3079160	0.011011761	0.1428400	116.9428
19	ROMANIA Top 50	0.5230400	0.6833000	0.6573400	0.09063200	0.2693286	0.002646221	0.1783440	117.4616
20	SLOVAKIA Top 50	0.4672680	0.6833000	0.6065000	0.08678600	0.2853067	0.026219391	0.1655280	119.7021
21	Spain Top 50	0.6377400	0.7465600	0.6702400	0.12980400	0.3116560	0.002273517	0.1272260	121.2425
22	Switzerland Top 50	0.5564140	0.7302400	0.6594200	0.16666400	0.2696620	0.004803388	0.1429020	116.4931
23	Turkey Top 50	0.4738200	0.7079000	0.6316800	0.14287800	0.2820260	0.061031655	0.1591720	127.1889
24	UKRAINE Top 50	0.5392900	0.7071600	0.6826400	0.12005200	0.2522878	0.022736335	0.1621460	119.8450
25	United Kingdom Top 50	0.5334600	0.7350800	0.6027800	0.16163800	0.2912192	0.007740541	0.1394960	116.1689
26	EUROPEAN MEAN	0.4904068	0.7060348	0.6300203	0.12559311	0.2678918	0.019984653	0.1634311	120.4117

Figure 4. 'sc_countries' data frame with all features mean for each country.

6.1.4.- First Approach

6.1.4.1.- Northern countries

	playlist_name	valence_mean
4	Germany Top 50	0.5782740
1	Denmark Top 50	0.5106800
11	EUROPEAN MEAN	0.4904068
5	Iceland Top 50	0.4792380
9	Poland Top 50	0.4779180
10	Sweden Top 50	0.4650000
3	Finland Top 50	0.4647800
6	Latvia Top 50	0.4587900
2	Estonia Top 50	0.4546040
7	Lithuania Top 50	0.4526260
8	Norway Top 50	0.4192400

Figure 5. Energy mean comparison

	playlist_name	energy_mean
4	Germany Top 50	0.6736200
3	Finland Top 50	0.6506400
11	EUROPEAN MEAN	0.6300203
9	Poland Top 50	0.6287200
8	Norway Top 50	0.6071800
1	Denmark Top 50	0.6018600
10	Sweden Top 50	0.5983660
7	Lithuania Top 50	0.5828800
6	Latvia Top 50	0.5768200
2	Estonia Top 50	0.5711200
5	Iceland Top 50	0.5037980

Figure 6. Valence mean comparison

6.1.4.2.- Southern countries

	playlist_name	energy_mean
7	MALTA Top 50	0.7897000
8	MONACO Top 50	0.7162200
3	CYPRUS Top 50	0.6739494
11	Spain Top 50	0.6702400
6	Italy Top 50	0.6656200
10	ROMANIA Top 50	0.6573400
2	BULGARIA Top 50	0.6442143
4	France Top 50	0.6332200
1	ANDORRA Top 50	0.6327187
12	Turkey Top 50	0.6316800
13	EUROPEAN MEAN	0.6300203
9	Portugal Top 50	0.6101800
5	GREECE Top 50	0.6065000

Figure 7. Energy mean comparison in 'southern_countries' data frame

6.2.-R Code

6.2.1.- 'main.R'

6.2.1.1.- #Import packages and access spotify API

```
options(stringsAsFactors = FALSE)
library(magrittr)
library(dplyr)
library(spotifyr)
library(tidyverse)
library(knitr)
library(WDI)
library(RJSONIO)
library(syuzhet)
library(data.table)
library(rAmCharts)

Sys.setenv(SPOTIFY_CLIENT_ID = "XXXXXXXXXXXXXXXXXXXXX")
Sys.setenv(SPOTIFY_CLIENT_SECRET = "XXXXXXXXXXXXXXXXXXXXXXXXX")

access_token <- get_spotify_access_token()

european_countries <- WDI(start = 2018, end = 2018, extra = TRUE) %>%
  subset(region == "Europe & Central Asia") %>%
    select(country)
```

6.2.1.2.- #Source 1 for playlist data retrieval

```
spotify_lists <- get_user_playlists('spotify')
spotify_top_lists <- subset(spotify_lists, grepl("(^| )top 50", playlist_name, ignore.case = TRUE))

char_country <- as.character(c(european_countries$country))
char_playlist <- as.character(c(spotify_top_lists$playlist_name))

european_top_lists <- c()

for (country in char_country)
{
  for (playlist_name in char_playlist)
  {
    if ((paste(country, "Top 50", sep=" ") %in% playlist_name) == TRUE)
    {european_top_lists <- rbind(european_top_lists, c(playlist_name))

    }
  }
}
colnames(european_top_lists)<- 'playlist_name'
european_top_lists<-merge(european_top_lists, spotify_top_lists)

europe50_tracks <- get_playlist_tracks(european_top_lists)
europe50_features <- get_track_audio_features(europe50_tracks)

europe50 <- merge(europe50_tracks, europe50_features ) %>%
  select(playlist_name,track_name, artist_name, danceability, energy, speechiness, acousticness, instrumentalness,
liveness, tempo, valence)

europe50_resume <- europe50 %>%
  group_by(playlist_name) %>%
  summarise(valence_mean = mean(valence),
        danceability_mean = mean(danceability),
        energy_mean = mean(energy),
```

```
        speechiness_mean = mean (speechiness),
        acousticness_mean = mean (acousticness),
        instrumentalness_mean = mean (instrumentalness),
        liveness_mean = mean (liveness),
        tempo_mean = mean(tempo)
  )
```

6.2.1.3.- #Source 2 for playlist data retrieval

```
country_toplists <- get_user_playlists('dgeizkipsrsk26i9krksl8wav')

char_country_2 <- as.character(c('UKRAINE', 'ROMANIA', 'MONACO', 'MALTA', 'LUXEMBOURG', 'LIECHENSTEIN', 'CZECH REPUBLIC', 'CYPRUS','ANDORRA','BULGARIA', 'GREECE', 'HUNGARY', 'SLOVAKIA'))

char_playlist_2 <- as.character(c(country_toplists$playlist_name))
european_top_lists_2 <- c()

for (country in char_country_2)
{playlist_count <- 0
 for (playlist_name in char_playlist_2)
 {playlist_count <-playlist_count + 1
  if (grepl(paste("TOP 50", country, sep=" "), playlist_name) == TRUE)
  {european_top_lists_2 <- rbind(european_top_lists_2, c(paste(country, "Top 50", sep=" "),playlist_name, country_toplists$playlist_uri[playlist_count], country_toplists$playlist_tracks_url[playlist_count], country_toplists$playlist_num_tracks[playlist_count], country_toplists$playlist_img[playlist_count]))

  }
 }
}
colnames(european_top_lists_2)<- c('playlist_name2','playlist_name', 'playlist_uri', 'playlist_tracks_url', 'playlist_num_tracks', 'playlist_img')

europe50_tracks_2 <- get_playlist_tracks(country_toplists)
europe50_tracks_2 <- merge(european_top_lists_2, europe50_tracks_2)
europe50_features_2 <- get_track_audio_features(europe50_tracks_2)

europe50_2 <- merge(europe50_tracks_2, europe50_features_2 ) %>%
 select(playlist_name2,track_name, artist_name,valence, danceability, energy, speechiness, acousticness, instrumentalness, liveness, tempo)
colnames(europe50_2)<-(c('playlist_name','track_name', 'artist_name', 'valence', 'danceability', 'energy', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'tempo'))

europe50_resume_2<-europe50_2 %>%
 group_by(playlist_name) %>%
 summarise(valence_mean = mean(valence),
        danceability_mean = mean(danceability),
        energy_mean = mean(energy),
        speechiness_mean = mean (speechiness),
        acousticness_mean = mean (acousticness),
        instrumentalness_mean = mean (instrumentalness),
        liveness_mean = mean (liveness),
        tempo_mean = mean(tempo)
 )
colnames(europe50_resume_2)<- c('playlist_name', 'valence_mean', 'danceability_mean', 'energy_mean', 'speechiness_mean', 'acousticness_mean', 'instrumentalness_mean', 'liveness_mean', 'tempo_mean')

#BIND BOTH SOURCES LISTS
europe <- rbind(europe50_resume, europe50_resume_2)
```

6.2.1.4.- #Create row with mean values of europe for each feature

```
valence_mean <- mean(europe [["valence_mean"]])
danceability_mean <- mean(europe [["danceability_mean"]])
energy_mean <- mean(europe [["energy_mean"]])
```

```
speechiness_mean <- mean(europe [["speechiness_mean"]])
acousticness_mean <- mean(europe [["acousticness_mean"]])
instrumentalness_mean <- mean(europe [["instrumentalness_mean"]])
liveness_mean <- mean(europe [["liveness_mean"]])
tempo_mean <- mean(europe [["tempo_mean"]])

playlist_name<-'EUROPEAN MEAN'
means_europe<-data.frame(playlist_name, valence_mean, danceability_mean, energy_mean, speechiness_mean, acousticness_mean, instrumentalness_mean, liveness_mean, tempo_mean)
europe <- rbind(europe, means_europe)
```

6.2.1.5.- #Geographical split of countries into groups

```
all_songs <- rbind(europe50, europe50_2)

#NORTHERN COUNTRIES
north_countries <- data.frame(rbind('Denmark Top 50', 'Estonia Top 50', 'Finland Top 50', 'Germany Top 50', 'Iceland Top 50', 'Latvia Top 50', 'Lithuania Top 50', 'Norway Top 50', 'Poland Top 50','Sweden Top 50'))
colnames(north_countries)<- 'playlist_name'
north_countries <- merge(north_countries,europe)
north_countries <- rbind(north_countries, means_europe)
north_songs <- merge(north_countries, all_songs) %>%
 select(playlist_name,track_name, artist_name, valence, danceability, energy, speechiness, acousticness, instrumentalness, liveness, tempo)

#CENTRAL COUNTRIES
central_countries <- data.frame(rbind('Austria Top 50','Belgium Top 50','Ireland Top 50', 'Luxembourg Top 50', 'Netherlands Top 50', 'Switzerland Top 50', 'United Kingdom Top 50', 'CZECH REPUBLIC Top 50', 'HUNGARY Top 50', 'LIECHENSTEIN Top 50', 'LUXEMBOURG Top 50', 'SLOVAKIA Top 50', 'UKRAINE Top 50'))
colnames(central_countries)<- 'playlist_name'
central_countries <- merge(central_countries,europe)
central_countries <- rbind(central_countries, means_europe)
central_songs <- merge(central_countries, all_songs) %>%
 select(playlist_name,track_name, artist_name, valence, danceability, energy, speechiness, acousticness, instrumentalness, liveness, tempo)

#SOUTHERN COUNTRIES
south_countries <- data.frame(rbind('France Top 50','Italy Top 50', 'Portugal Top 50','Spain Top 50', 'Turkey Top 50', 'ANDORRA Top 50', 'BULGARIA Top 50', 'CYPRUS Top 50', 'GREECE Top 50', 'MALTA Top 50', 'MONACO Top 50', 'ROMANIA Top 50'))
colnames(south_countries)<- 'playlist_name'
south_countries <- merge(south_countries,europe)
south_countries <- rbind(south_countries, means_europe)
south_songs <- merge(south_countries, all_songs) %>%
 select(playlist_name,track_name, artist_name, valence, danceability, energy, speechiness, acousticness, instrumentalness, liveness, tempo)

#SOUTHERN-CENTRAL COUNTRIES
sc_countries <- data.frame(rbind('France Top 50','Italy Top 50', 'Portugal Top 50','Spain Top 50', 'Turkey Top 50', 'ANDORRA Top 50', 'BULGARIA Top 50', 'CYPRUS Top 50', 'GREECE Top 50', 'MALTA Top 50', 'MONACO Top 50', 'ROMANIA Top 50', 'Austria Top 50','Belgium Top 50','Ireland Top 50', 'Luxembourg Top 50', 'Netherlands Top 50', 'Switzerland Top 50', 'United Kingdom Top 50', 'CZECH REPUBLIC Top 50', 'HUNGARY Top 50', 'LIECHENSTEIN Top 50', 'LUXEMBOURG Top 50', 'SLOVAKIA Top 50', 'UKRAINE Top 50'))
colnames(sc_countries)<- 'playlist_name'
sc_countries <- merge(sc_countries,europe)
sc_countries <- rbind(sc_countries, means_europe)
sc_songs <- merge(sc_countries, all_songs) %>%
 select(playlist_name,track_name, artist_name, valence, danceability, energy, speechiness, acousticness, instrumentalness, liveness, tempo)
```

6.2.1.6.- #Plots

```
#VALENCE (C-S)
```

```
h1 <- central_songs$valence
h2 <- south_songs$valence

xLimits <- range(c(h1,h2))
breakPoints <- seq(xLimits[1], xLimits[2], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
mp<-barplot(hist2Percentage, col = "#00ff0050", add = T, main = "Valence comparison: Central-Southern Europe",
xlab = "Valence", ylab = "Percentage")
legend("topleft", legend = c("Central", "Southern"), fill = c("#0000ff50", "#00ff0050"))
axis(1,at=mp,labels=seq(0,1,0.05))


#VALENCE(N-C-S)
h1 <- north_songs$valence
h2 <- central_songs$valence
h3 <- south_songs$valence
xLimits <- range(c(h1,h2, h3))
breakPoints <- seq(xLimits[1], xLimits[2],xLimits[3], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

hist3 <- hist(h3, breaks = breakPoints, plot = F)
hist3Percentage = hist3$counts/sum(hist3$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
barplot(hist2Percentage, col = "#ff000050", add=T)
mp<-barplot(hist3Percentage, col = "#00ff0050", add = T, main = "Valence comparison: N-C-S Europe", xlab =
"Valence", ylab = "Percentage")
legend("topleft", legend = c("Northern", "Central", "Southern"), fill = c("#0000ff50", "#ff000050", "#00ff0050"))
axis(1,at=mp,labels=seq(0,1,0.05))

#VALENCE (N-CS)
h1 <- north_songs$valence
h2 <- sc_songs$valence

xLimits <- range(c(h1,h2))
breakPoints <- seq(xLimits[1], xLimits[2], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
mp<-barplot(hist2Percentage, col = "#00ff0050", add = T, main = "Valence comparison", xlab = "Valence", ylab =
"Percentage")
legend("topleft", legend = c("Northern", "SC"), fill = c("#0000ff50", "#00ff0050"))
axis(1,at=mp,labels=seq(0,1,0.05))


#DANCEABILITY (N-CS)
```

```
h1 <- north_songs$danceability
h2 <- sc_songs$danceability

xLimits <- range(c(h1,h2))
breakPoints <- seq(xLimits[1], xLimits[2], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
mp<-barplot(hist2Percentage, col = "#00ff0050", add = T, main = "Danceability comparison", xlab = "Danceability",
ylab = "Percentage")
legend("topleft", legend = c("Northern", "SC"), fill = c("#0000ff50", "#00ff0050"))
axis(1,at=mp,labels=seq(0,1,0.05))


#ENERGY (N-CS)
h1 <- north_songs$energy
h2<-sc_songs$energy

xLimits <- range(c(h1,h2))
breakPoints <- seq(xLimits[1], xLimits[2], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
mp<-barplot(hist2Percentage, col = "#00ff0050", add= T,  main = "Energy comparison", xlab = "Energy", ylab =
"Percentage")
legend("topleft", legend = c("Northern", "SC"), fill = c("#0000ff50", "#00ff0050"))
axis(1,at=mp,labels=seq(0,1,0.05))

#TEMPO (N-CS)
h1 <- north_songs$tempo
h2<-sc_songs$tempo

xLimits <- range(c(h1,h2))
breakPoints <- seq(xLimits[1], xLimits[2], length.out = 22)

hist1 <- hist(h1, breaks = breakPoints, plot = F)
hist1Percentage = hist1$counts/sum(hist1$counts)*100

hist2 <- hist(h2, breaks = breakPoints, plot = F)
hist2Percentage = hist2$counts/sum(hist2$counts)*100

barplot(hist1Percentage, col = "#0000ff50")
mp<-barplot(hist2Percentage, col = "#00ff0050", add= T,  main = "Tempo comparison", xlab = "Tempo (BPM)", ylab
= "Percentage")
legend("topleft", legend = c("Northern", "SC"), fill = c("#0000ff50", "#00ff0050"))
axis(1,at=mp,labels=seq(0,200,10))
```