

# Brazilian/British Music Variations during the week

Arthur Costa

July 4, 2018

## Abstract

This paper is intended as a project for the discipline of Data Science in the UFPE. There are no scientific purposes on this article, however there will be used some methods to confirm the observations, and hopefully it will generate some fun trivia.

## 1 Obtaining Data

To obtain data about music trends in these countries, with a more in depth data than the most basic ones, I had to choose a platform that would allow me to get this data somehow. The one I chose was LastFM, to crawl users, and get the musical history from each one. To get more data about each music I chose to query Spotify API as they had more data on tracks.

### 1.1 Finding users on LastFM

To start, the initial thing was getting a few users as seed, and then for each user query its friends. There were 3 queues for users, 1 for brazilians, 1 for Uk users, and 1 for other users. On each round, one user was taken from one of the queues(brazilian one first, if it was empty or had 10 more brazilians processed then would choose Uk one, and would only get from the 3rd one if there was no one from br or uk), would see 100 of his friends, and query his nationalities, and put on each queue according to each nationality(if it wasn't processed before).

### 1.2 Getting musical history from each user on LastFM

When each user is taken from the queue, if he is from Brazil or UK the crawler will query LastFM for it's last 1000 musics. Then this musics are queried on Spotify to get their features.

### 1.3 Getting music features info about the musics from Spotify

Each music was searched by the name on Spotify, that gives a little of noise on musics with same name, but not enough to show into calculations. The Features extracted were 9: acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence.

### 1.4 Aggregating musical features by time in the week and country

As it was known the country of each user, it was needed to put all data from each country summed up aggregated by hour of week.

### 1.5 Statistics about the data crawled

200 hundred users were processed, being roughly 90 from each country. 76061 Songs from brazilian users, and 66295 from UK users. Over 150k requests were made during this data gathering.

## 2 Data Observations

### 2.1 Overall comparisons

Getting a comparison per category, on which tops which to then get some conclusions about the feature averages over each country.

Feature	Brazil	United Kingdom
Acousticness	0.3070522150555476	0.314761650420092
Danceability	0.5181635910650663	0.5409404253714458
Energy	0.6183142958835671	0.6010829950991778
Instrumentalness	0.1632221673711889	0.20346305205068257
Liveness	0.20991806655184653	0.1985800980466098
Loudness	-9.042842889259939	-9.250683792141187
Speechiness	0.07570553108689079	0.08827292254317823
Tempo	120.14557384204781	118.99128139377027
Valence	0.4547096533045844	0.4479696824798251

### 2.1.1 Observations on the averages

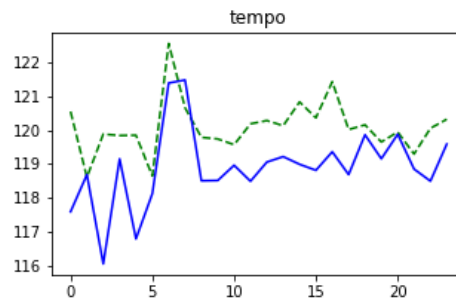
Brazil and United Kingdom are very close on almost all averages however a few things were strange, and here are they:

1. The instrumentalness was close to 25% bigger on the UK, what might imply that brazilian musics are more about singing, or since the measure of instrumentalness is a probability the only thing that can be that both countries are not very into instrumental music.
2. Loudness is 0.20 less in the UK than in BR, and remember that loudness is based on logarithm so UK music is 20% calmer than the musics that brazilians listen to.
3. Danceability is bigger in the UK by 5%, what is against my initial conceptions before this study.

## 2.2 Exploring hours of the day

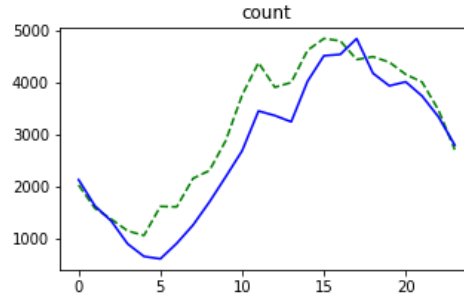
Here there will be graphs showing the data, and will use the dashed green lines to represent brazil, and the continuous blue line for the UK.

### 2.2.1 Tempo over the day



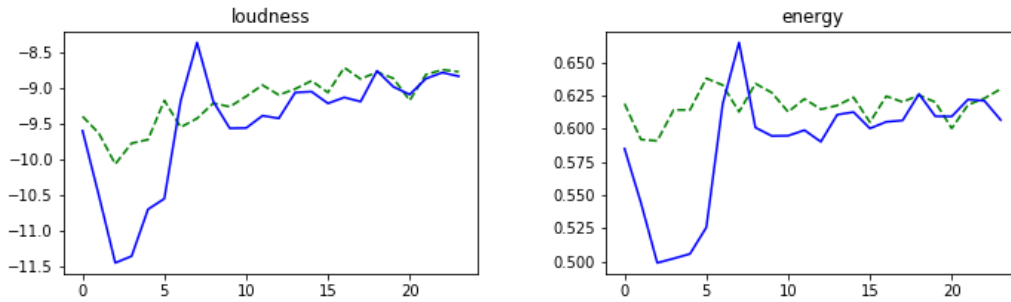
Brazil songs stay 2 beats per minute above the UK almost all the day, there is a strange spike in the morning where might be caused by people trying to start the day more agitated.

### 2.2.2 Count over the day



Although both countries seem similar on the count along the day, the fact that we can see the number of musics increasing by 5x from 5 AM to 5 PM, still is impressive. Caused by the sleep time, no doubt, however knowing the hours music is listened to is useful on ads targeting.

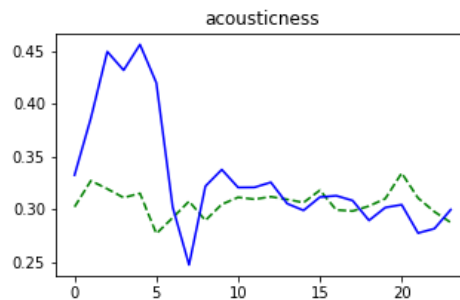
### 2.2.3 Loudness and Energy over the day



There is a clear correlation (plotting actual correlation got me 0.923330) between uk loudness and energy, what might be as one of the parameters to calculate energy might be the loudness.

There is a drop on the energy and loudness between 2 AM and 5 AM in the UK, what might suggest that people from the UK prefer calmer songs during the night period.

### 2.2.4 Acousticness over the day

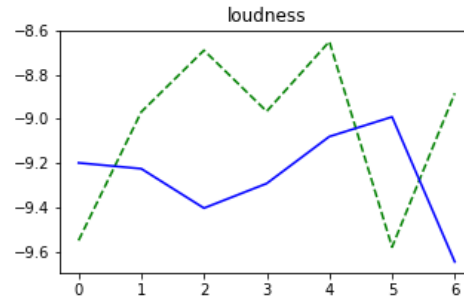


There is a rise on the acousticness on the same time the loudness/energy goes down. So led to the thought that the correlation is inverted, and what was found by calculating correlation confirmed the idea, there was a -0.966364 energy-acousticness correlation. That might be caused as acoustic songs tend to be way calmer.

## 2.3 Exploring days of the week

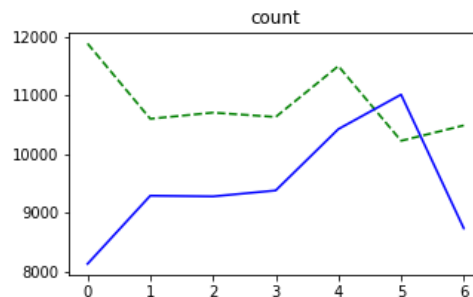
The week graphs were made in a way that Monday is the day 0, and 6 is Sunday.

### 2.3.1 Loudness over days



While music in the UK gets louder closer to the weekends, while brazilians get calmer on the saturdays.

### 2.3.2 Count over days



Finally the count during the week, while brazil goes down along the week, and UK goes up as weeks passes by, while there is a clear pattern on the UK one. UK pattern is aligned with the expectations that people listen more closer to weekends, where they will go out. However there is no clear pattern on the brazil count.

## 3 Problems

Here I will point the mistakes done in this article, as again it was not meant to be used into other scientific purposes, and is actually meant to learn on data analysis, I decided to point out mistakes done during this study.

### 3.1 Gathering data

#### 3.1.1 Handling errors

The APIs were failing on some of the calls, and the error handling was not perfect so there were some losses on a few of the users.

#### 3.1.2 Granularity

I assumed I didn't need a bigger granularity than the hour so I didn't get more granular than that on collecting data, that should be done on the processing, not on getting data, as I would want use more granular data later.

### **3.1.3 Raw Data**

I didn't save the raw data, and if I had saved the whole raw data I would not have the problem 3.1.2, and would be able to test things using p tests.

### **3.1.4 Taking care of timezones**

When I collected the data it was not considering timezones, so I had to take that into account later.

## **3.2 Analyzing Data**

### **3.2.1 Colors only graph(fixed later)**

During the creation I chose color based graphs, so I realized the mistake that it would be awful for color blind people, and fixed it.

### **3.2.2 Count Normalization**

The raw count wasn't the best option to visualize frequencies as there was 10k more from brazil. I should have normalized it.

### **3.2.3 Treating data in wrong formats**

The fact that instrumentality was a probability was not well taken into account during some parts, as it didn't make much sense to use an average for it.