Analysis of the Jorge Ben Jor data from the spotiyr packages

Ricardo Mühlstedt

2022-07-23

Abstract

Jorge Ben Jor is one of the greatest artists from Brazil, widely known for the MPB, Bossa nova, Samba-rock e Sambalanço genres.
I decided to realize an analysis of his tracks and albums on Spotify, utilizing the spotifyr package, to get an insight into his musical stats and share my findings.

Introduction

The spotifyr package lets me load all data from the Spotify API.

There‘s a lot of data here, but we can filter it out, only selecting the most interesting, like the cheeriest songs, measured by the valence value, the most danceable music, or even the liveliest, measured by the musical energy.

track_name	valence	album_name
Velhos, Flores, Criancinhas E Cachorros	0.984	Solta O Pavão
Gertrudes Bon Hausen	0.981	Homo Sapiens
Menina Mulher Da Pele Preta	0.979	Favourites: From Samba Esquema Novo 1963 To Africa Brasil 1976
Eu Vou Torcer	0.977	A Tabua De Esmeralda
Menina Mulher Da Pele Preta	0.976	A Tabua De Esmeralda

track_name	danceability
Emo	0.850
Gabriel, Rafael, Miguel	0.846
Gostosa	0.839
Alô Alô, Como Vai	0.839
Dorothy	0.837

track_name	energy
Bicho Do Mato - Live From Brazil/1964/Edit	0.984
Homem do espaço - Ao vivo	0.983
Homem do espaço - Ao vivo	0.983
Cavaleiro Do Cavalo Imaculado	0.975

track_name	energy
Bebete Vãobora / Criola / Cadê Tereza	0.973

Having seen all this data sparks a lot of questions about it. Which are the most common keys?

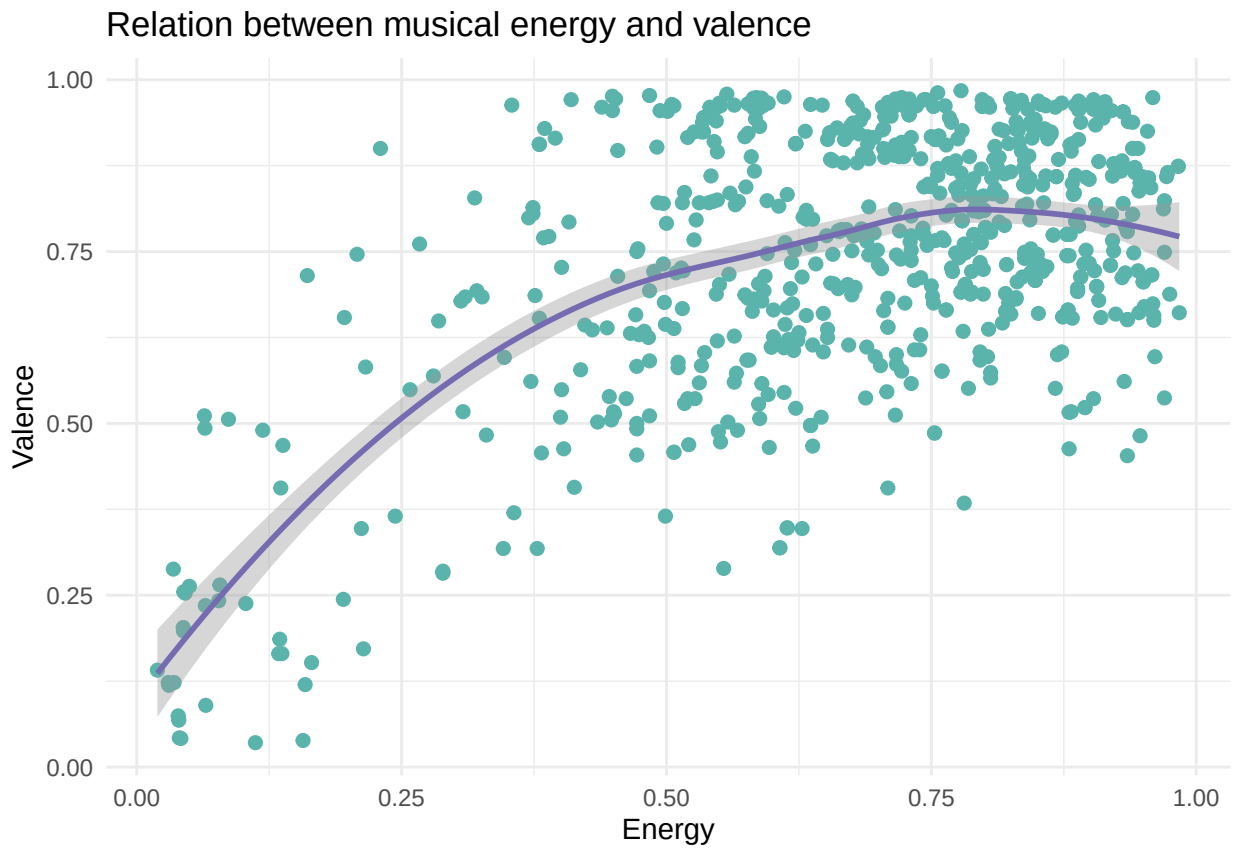
- Does the danceability vary with the change in valence?
- What is the relation between acousticnesss and valence?
- Is there a specific tempo that is more used?

I will address all these questions and show some phases of my exploratory data analysis and my thought process throughout the entirety of the exploration.

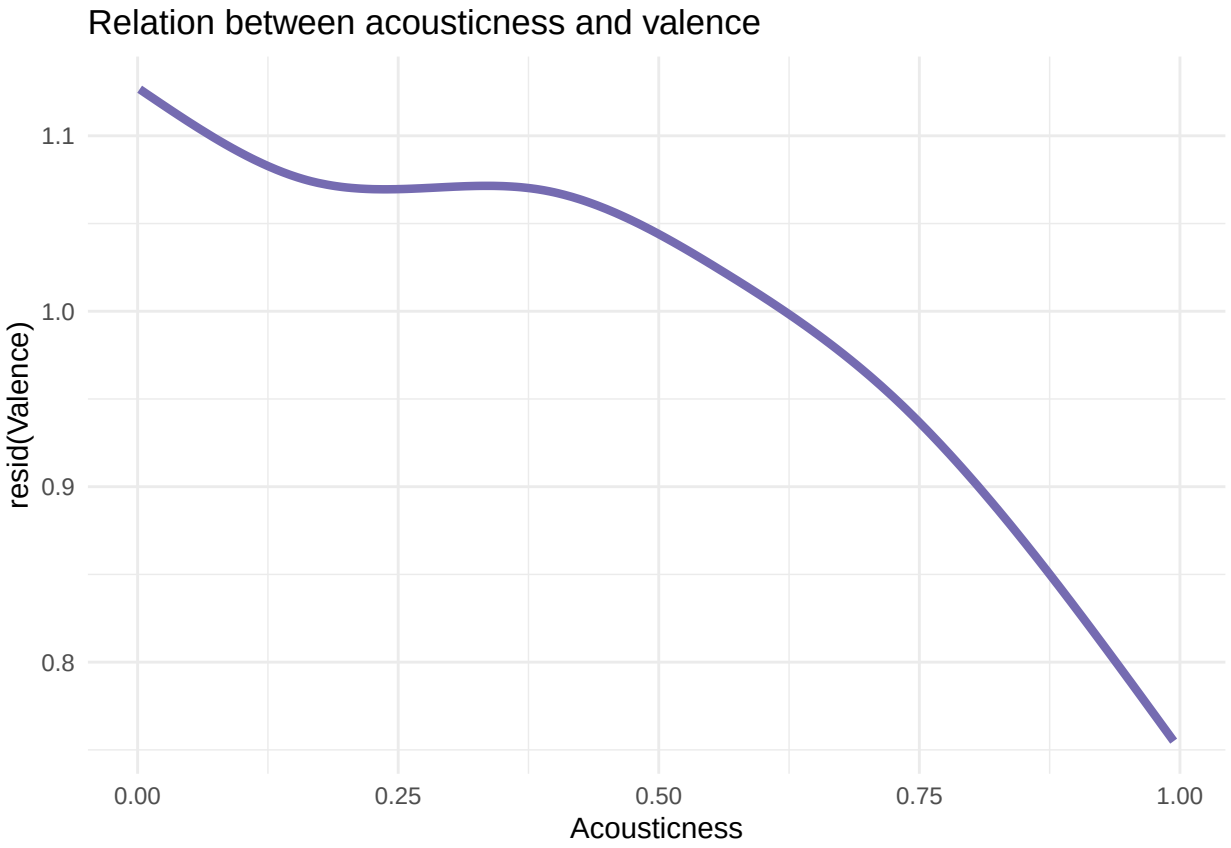
Results

When exploring, I found an interesting pattern, a strong correlation of variables. The musical energy seems to influence the valence value as the energy levels rise, and so does the valence. Strong covariation impact other relations, so how the variables behave, I must remove strong correlation to precisely how the data varies.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

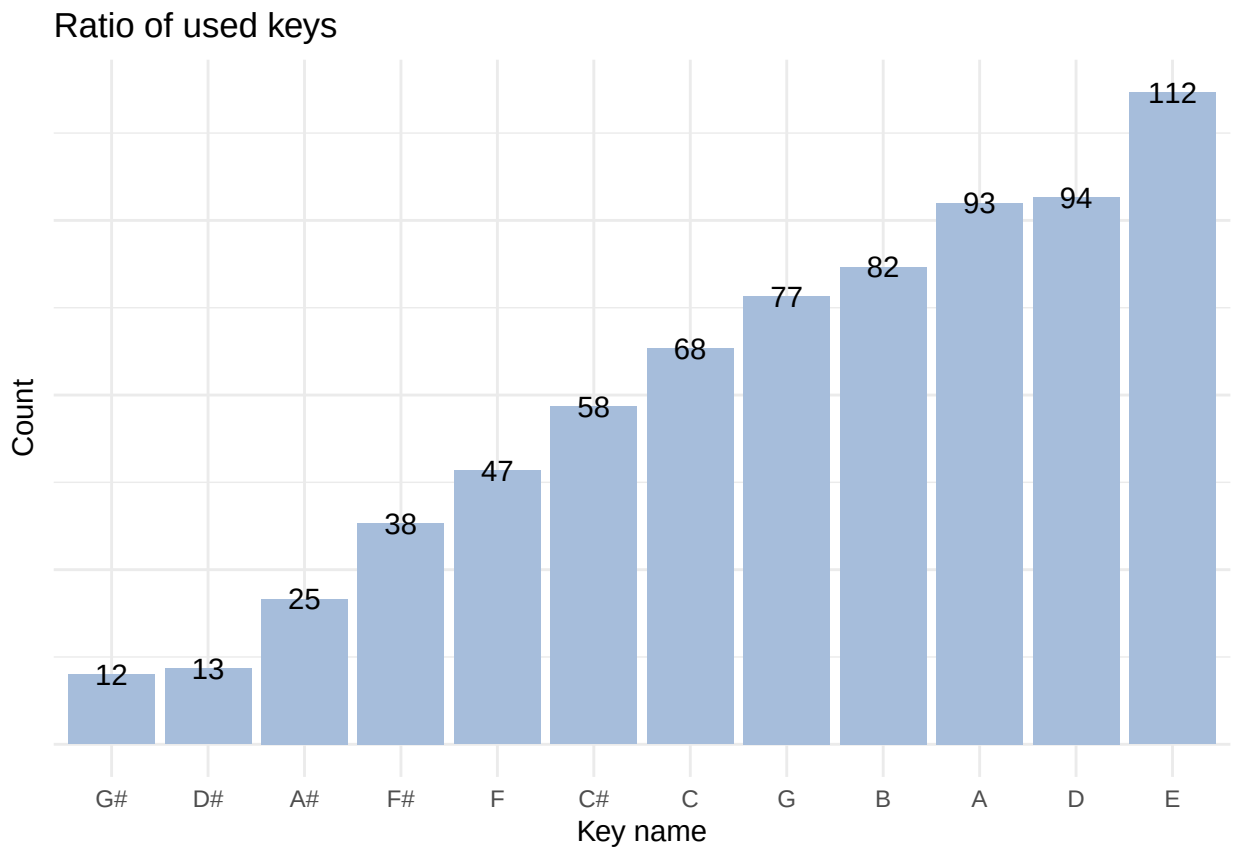


As you can see, the pattern is easily recognized. To remove the pattern, I inserted a model that generalizes the variation and then used the residuals as my new dataset. The residuals are all the data that don't fit into the model and are not displayed, utilizing the residuals as my new data set, essentially making the data unbiased. With the relation removed, defining a model showing the pattern between acousticness and valence becomes possible.



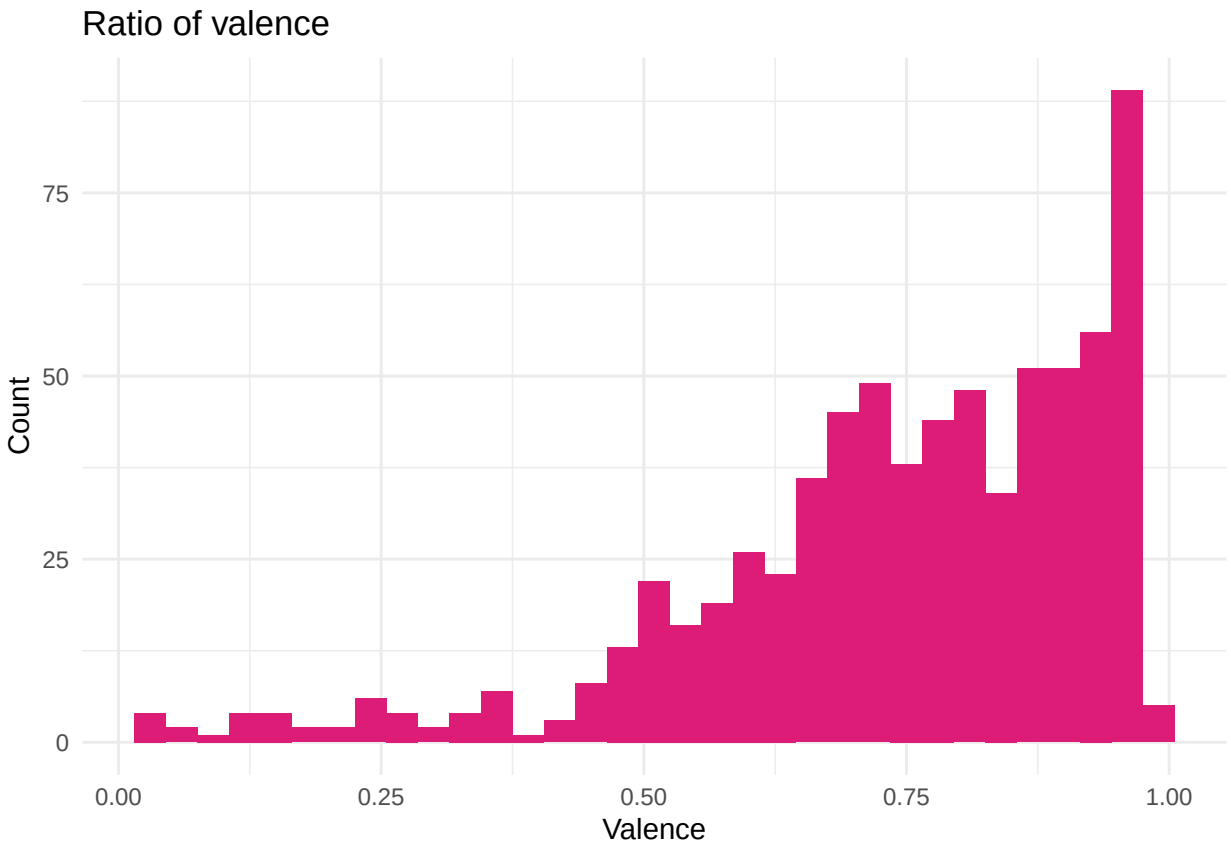
The reason the pattern exists may be because the artist uses acousticness to represent sadness, sorrow, or any negative emotion, as most of the work by Ben Jor is gleeful and bright with a decent level of acousticness. There is no way of determining it by analyzing only the work of Jorge Ben Jor. The conclusion would be biased and not scientific.

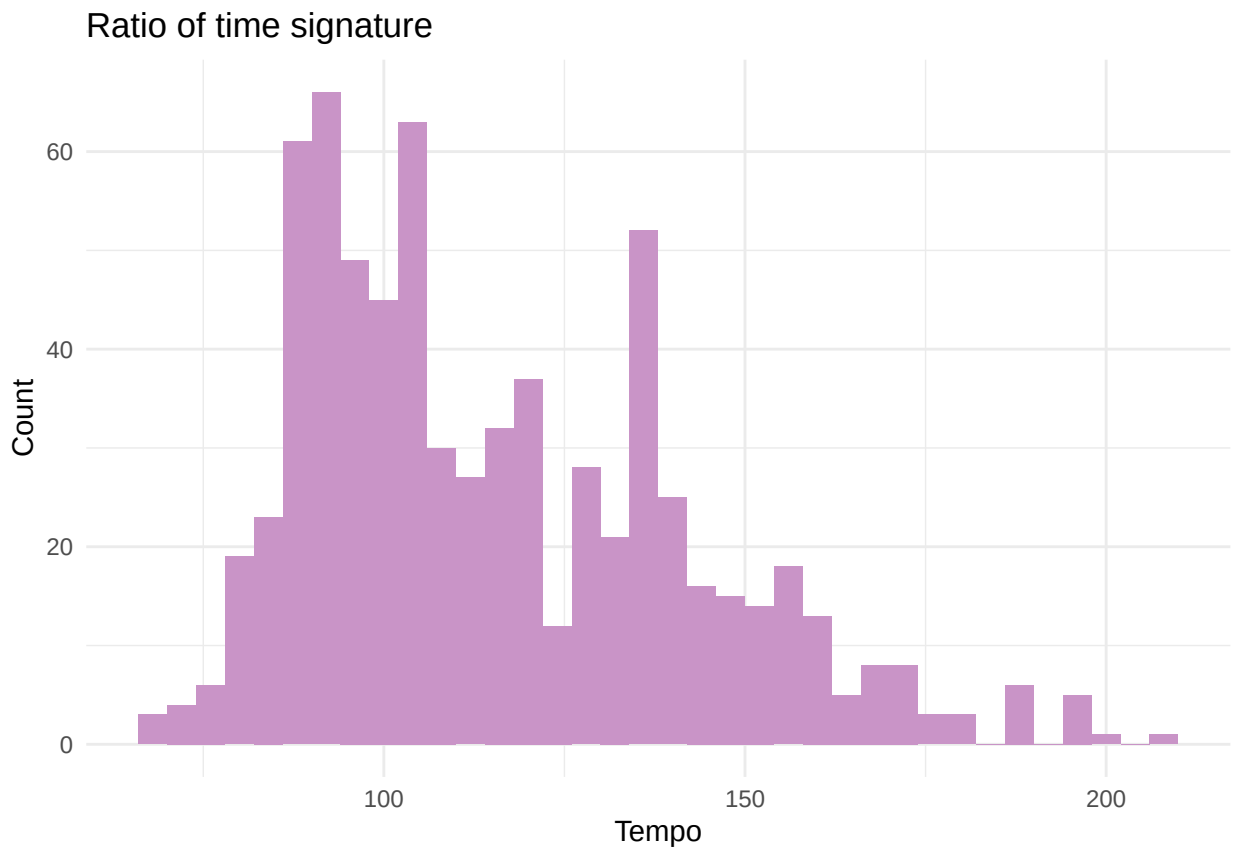
It is reasonable to find the most used keys, to demonstrate how he explores the musical theory in his pieces.



Ben Jor is a well roudned musician, managing to diverse a lot his use of key and instruments on his band and concerts, with a lot of experience and releases.

Diving into his style and finding the ratio of valence in his tracks is essential. Samba-rock, MPB, and Sambalanço genres are generally energetic, diverse, high tempo and happy. Those characteristics must show in his work.


Ratio of valence
Count
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Valence



All the characteristics are here, as we expected.