

# SCC0252 - Visualização Computacional (2022)

## Introdução

A base de dados escolhida contém informações sobre diversas músicas, sendo os campos (traduzidos): gênero musical, nome do artista, nome da música, id da música, popularidade, acústico, dançabilidade, duração em ms, instrumentalidade, nota (nota musical principal, o sistema utilizado é bem diferente do nosso, sendo as notas sendo classificadas em letras e o tipo com alguma nomeação, por exemplo, # significa sharp, um tipo específico de nota), vivacidade, volume, modo (variável categórica, Minor e Major), fala, tempo (no sentido musical, a pronúncia é um pouco diferente), assinatura de tempo e valência. As traduções podem ter ficado um pouco quebradas, por isso recomendo ler um pouco sobre no primeiro link.[\[1\]](#)

O objetivo deste relatório é mostrar as diversas características a serem descobertas por meio de visualização.

## Visualizações

### 0 - Tratamento dos dados

Existia uma duplicação de gêneros musicais valores na base, foi algo simples de tratar, simplesmente substituindo todas as colunas com o gênero "Children's Music" por "Children's Music"

### 1 - Popularidade e gênero

Foi plotada a relação popularidade com o gênero, a partir desse gráfico é possível entender as seguintes informações:

- Pop é o mais popular.
- A Capella é o menos popular.

### 2 - Emoção e gênero

Foram pegas as principais variáveis envolvendo emoção que são contínuas, sendo elas : acousticness, danceability, instrumentalness, liveness, loudness, speechiness, valence.

Porém, resolvi remover instrumentalness e speechiness por entender que algumas categorias seriam muito favorecidas ou prejudicadas, como Opera e Classical.

A primeira análise é bem simples, eu simplesmente printei as médias globais para ter uma ideia de quem está acima ou abaixo da média nas próximas tabelas.

Depois fiz uma análise usando uma série de plots-violino para entender principalmente a distribuição de um valor dentro de um gênero.

## 2.1 - Análise Violino

### Acousticness

- Opera possui um acústico muito alto, e diferente dos outros não varia tanto. A maior parte dos valores de Opera ,Classical, Capella e Movie estão no topo de acústico. A maior parte dos valores de Electronic e Ska estão na parte mais baixa de acústico. Não dá para analisar tanto a máximo e mínimo pois diversos gêneros possuem valores tanto no máximo quanto no mínimo.

### Danceability

- Opera e comedy possuem as menores variabilidades. Opera, Classical, Soundtrack possuem pouca dançabilidade. Movie possui a mais média dançabilidade. Hip-Hop, Rap, Reggae, Reggaeton possuem as maiores dançabilidades. Impressionantemente Dance não tem uma dançabilidade média superior aos citados como maiores.

### Loudness

- A maior parte dos gêneros possui a box do violino entre -10 e -15. World, Classical, Soundtrack, Opera e Blues possuem variabilidade altas, e tirando World suas boxes também estão abaixo das outras.

### Liveness

- Todos possuem valores médios de liveness baixo, sendo apenas a variabilidade e a distribuição geral diferente, o único valor diferente disso é Capella, com uma variabilidade muito baixa.

### Valence

- Capella possui claramente o máximo e mínimo nessa categoria. Opera, Classical, Soundtrack e World possuem novamente os valores mais diferentes.

## 2.2 - Análise Spider

Para fazer um contraponto com o violino eu tentei fazer um outro plot utilizando um spider-chart, minha ideia é mostrar sobreposto os 5 menos populares gêneros para apresentar quais são suas similaridades.

Infelizmente não consegui fazer uma plotagem legal, tentei usando for e outras ideias, mas no final não ficou bom realmente, a única informação útil foi de que todos possuem um loudness alto. Tentei montar de outras formas, porém no final vi que a única maneira viável parece ser copiando e colando o código diversas vezes, algo imprático. Foi necessário minimizar o espaço dos dados senão basicamente todos os gráficos ficariam iguais.[\[2\]](#)

A única característica comum a todos é a loudness alta.

## 3 - Reanálise

Podemos perceber a partir das duas análises que alguns dos valores citados como os mais diferentes nas análises de valores também são diferentes da norma em popularidade,

por exemplo, Capella, Movie e Opera. Outros valores fora da norma possuem popularidade abaixo da média, como : Soundtrack e World.

Dançabilidade alta não necessariamente significa popularidade alta, sendo o caso do Reggae e Reggaeton, mas não é determinante de popularidade, com outros valores com alta dançabilidade sendo populares.

#### 4 - Análise de artistas

Sabendo já algumas informações sobre gênero é interessante entender a relação disso com os artistas. Foram feitos dois plots, um mostrando os 20 mais populares artistas nos dados, e outro mostrando os 20 mais populares, porém para cada um dos estilos musicais tocados. Ambas as visualizações usam o lollipop-plot, foi escolhida essa técnica pois ficou mais visível do que os gráficos de barras. A primeira visualização serve como base para a segunda.

A primeira nos mostra que os 5 mais populares artistas são respectivamente : Drake, Chris Brown, Nobuo Uematsu, Future e Hams Zimmer. Agora com um agrupamento por gênero toda distribuição dos 20 mudaram, os novos 5 primeiros são: Hans Zimmer (Soundtrack), Mozart (Classical), Nobuo Uematsu (Anime), Verdi (Opera), Bach (Classical).

Olhando mais a fundo esses dados percebemos o seguinte :

- Alguns artistas muito populares na primeira não se apresentam na segunda, como Chris Brown, isso pode ser um erro do tratamento de dados, porém é mais provável ser por conta dos gêneros e quantidades de músicas. Chris Brown deve ter mais de um gênero principal, e somando todos os gêneros ele se torna popular, contudo, o segundo plot foca em especialistas e não em generalistas.
- Embora Drake tenha caído diversas posições no segundo plot, ele apresenta dados interessantes, estando com 3 gêneros diferentes na classificação, sendo eles : Hip-Hop, Rap e Pop.
- A primeira visualização nos mostra a popularidade dos gêneros, porém, alguns dos gêneros “não-populares” estão presentes em diversos artistas da lista de 20 maiores, isso significa que a popularidade do artista infere a do gênero. Outra observação possível é de que esses gêneros não possuem tantos artistas em comparação com os outros.

#### 5 - Análise popularidade do gênero

Para entender a popularidade de um gênero é necessário observar a distribuição dele dentro dos dados, para fazer isso utilizei um histograma, este contando as repetições do campo gênero no banco de dados.

A partir da visualização foi possível analisar o seguinte:

- A Capella possui o menor número de músicas na base, com apenas 119, enquanto a maior parte dos gêneros passa dos 8000. Faz sentido nesse caso ele ser o menos popular.
- Children's Music possui o maior número de músicas com mais de 14 mil, porém não é a mais popular, isso indica que embora muitas músicas infantis sejam produzidas nem todas possuem tanta relevância quanto em outros estilos.

- Tirando essas duas exceções não existe anormalidade quanto à distribuição de gêneros.

## Referências

[1]<https://towardsdatascience.com/is-my-spotify-music-boring-an-analysis-involving-music-data-and-machine-learning-47550ae931de>

[2]<https://towardsdatascience.com/how-to-create-a-radar-chart-in-python-36b9ebaa7a64>