



**Serviço Público Federal
Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Faculdade de Estatística**

Projeto da Disciplina
Estatística Computacional (EST1008)
2021.3 - ERE - Local: PL-II (24/06/2021 - 07/10/2021)

Gabriel Menezes Gaia¹ - 201806840009

Prof. Dr. rer. nat. João Marcelo Brazão Protázio

Análise das tendências musicais no mundo entre 2010 e 2019 no Spotify

Resumo

Este trabalho tem como intuito apresentar como se comportaram as tendências sobre música no mundo entre 2010 e 2019. O estudo foi feito com músicas que figuraram entre as 10 mais tocadas durante cada ano do período citado no app de streaming de músicas mundialmente aclamado Spotify. A partir das variáveis presentes no trabalho - tais como gênero musical, artista, batidas por minuto etc. - o objetivo é mostrar quais foram as constantes durante o intervalo de tempo e, no fim, modelar uma música que seria a ideal para a década. Os resultados mostram que as músicas *pop* predominaram a década, com sons enérgicos e com um astral mais positivo, o que atinge mais pessoas por ser algo mais palatável ao ouvido e, sendo assim, gera melhores resultados financeiros.

Abstract

This paper intends to present how the music tendencies behaved between 2010 and 2019 around the world. The study has been made with the songs that appeared among the 10 most played on each year of the said period on the world famous music streaming app, Spotify. From the variables present on the work, such as music genre, artist, beats per minute, etc, the goal is to show what were the constants along the time interval and, in the end, model a song that would fit anywhere on the decade. The results show that pop songs predominated the decade, with energetic sounds and a more positive mood, which reaches more people because it is something more palatable to the ear and, therefore, generates more financial results.

¹ gaiagabriel22@gmail.com

1 Introdução

A música sempre é uma forma de arte que sempre foi – e deve continuar sendo – muito presente na humanidade. Com vários propósitos e formas, ela se faz fundamental tanto para descontração, como para transmitir mensagens e críticas, para disseminar sentimentos e contagiar o âmago de cada ouvinte com cada componente de sua estrutura, da letra à melodia.

Tendo isso em vista, foi completamente natural a mudança na forma como a música é acessada nos tempos atuais. Com a consolidação do mercado de smartphones e os avanços na área de telecomunicações (provendo melhores serviços de acesso à internet), o streaming de músicas passou a ser uma alternativa cada vez mais forte nesse âmbito. Assim como o streaming se tornou uma realidade no quesito audiovisual (filmes, séries, documentários), ele também tomou conta de podcasts, entrevistas e, claro, da música. Hoje, ao invés de ser necessário levar consigo um player individual com fitas e/ou CDs, basta pesquisar a música desejada (ou apenas selecionar o que está em alta ou o que é sugerido pela plataforma) e apertar o play para escutar qualquer produção musical de qualquer canto do planeta.

Sendo assim, vários apps se destacam no mercado, com o Spotify sendo o escolhido como fonte de dados para o presente estudo. Com mais de 350 milhões de usuários ativos no mundo (Exame, 2021), ele se torna um excelente parâmetro para o objetivo deste trabalho, que visa mostrar uma visão geral do comportamento musical nos últimos 10 anos, tendo como base várias informações sobre músicas que apareceram entre as mais tocadas no referido aplicativo em cada ano de 2010 a 2021. Com isso, é possível mostrar quais foram as tendências de gêneros e artistas, sejam elas positivas ou negativas, além de tentar modelar qual seria a música ideal que seria bem aceita por todos os anos, utilizando as variáveis como musicalidade e energia.

2 Materiais e Métodos

2.1 Materiais

Neste trabalho, foi escolhido o dataset *Top Spotify songs from 2010-2019 – BY YEAR* (Melhores músicas do Spotify 2010-2019 – Por ano) encontrado no website *Kaggle*, mundialmente famoso como um lar para trabalhos de ciência de dados e machine learning. Este dataset contém o nome de cada música que figurou entre as 10 mais tocadas em cada ano de 2010 e 2021 e um ID para cada uma, além de 13 variáveis referentes a cada música, como o ano de lançamento e informações mais técnicas como altura, duração, popularidade etc. As informações são retiradas do site *Organize Your Music*, que gera esses valores para cada música fazendo comparações com outras em seu banco de dados. No total, são 15 colunas, sendo 3 com valores de tipo *string* e as demais de tipo *int*.

Descrição do dataset

ATRIBUTO	COLUNA	TIPO	DESCRIÇÃO
ID	#	int	Identificação da música no dataset
TÍTULO	title	string	Nome da música
ARTISTA	artist	string	Artista ou banda
GÊNERO PRINCIPAL	top genre	string	Gênero principal da música
ANO	year	int	Ano de lançamento
BPM (BATIDAS POR MINUTO)	bpm	int	O tempo da música
ENERGIA	nrgy	int	Quanto maior, mais energia tem a música
FACILIDADE PARA DANÇA	dnce	int	Quanto maior, mais fácil é para dançar a música
ALTURA	dB	int	Quanto maior, mais alta é a música
É AO VIVO	live	int	Quanto maior, mais chance de ser uma gravação ao vivo
POSITIVIDADE	val	int	Quanto maior, mais alegre é a música
COMPRIMENTO	dur	int	Quanto maior, mais longa é a música
ACÚSTICA	acous	int	Quanto maior, mais acústica a música é
FALA	spch	int	Quanto maior, mais palavras são ditas na música

POPULARIDADE	pop	int	Quanto maior, mais popular é a música
--------------	-----	-----	---------------------------------------

2.2 Métodos

O conjunto de dados escolhido já provém os dados preparados para uso, portanto não foi necessário nenhum tipo de manipulação ou pré-processamento de qualquer tipo, já que os dados já estavam bem colocados e com classes adequadas. Dessa forma, o projeto foca na visualização e análise dos dados - com gráficos de barra, por exemplo, utilizando referências como WICKHAM & GROLEMUND, 2019 - buscando inferir e verificar as tendências musicais no período. Logo, o trabalho não tem a intenção de entrar no mérito musical, para entender os porquês de um elemento ter mais sucesso ou não.

3 Resultados e Discussão

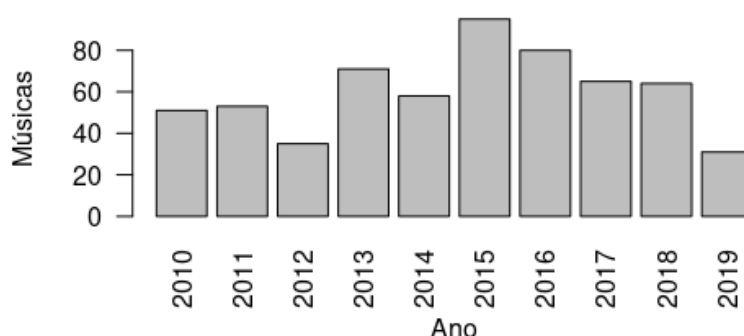
3.1 Estatística Descritiva

Como exemplo da descrição das colunas, as colunas *dançabilidade* e *fala* foram selecionadas para a apresentação. Na Tabela 1, são mostrados os números relativos às colunas.

Coluna	Média	Desvio Padrão	Mínimo	Máximo	1 Quartil	Mediana	3 Quartil
Dançabilidade	64.38	13.38	0	97	57	66	73
Fala	8.35	7.48	0	48	4	5	9

Na tabela, é possível notar, por exemplo, que a coluna *fala* possui um desvio padrão muito alto, porém seu terceiro quartil é 9 e sua média é 8.35, ou seja, a maioria dos valores estão abaixo de 9 mas o desvio padrão indica muita variação - infere-se, então, que existem outliers influenciando nos cálculos. Já na coluna *dançabilidade*, nota-se que a distribuição é mais homogênea, com a média ficando bem próxima da mediana.

Outra descrição importante, é mostrar a quantidade de músicas em cada ano, o que é feito na imagem 1.



Na imagem, é possível ver a distribuição das 603 músicas do dataset por ano. Pode-se perceber tanto a homogeneidade de alguns anos como 2010 e 2011, assim como a discrepância de outros, como 2015 - quando várias músicas entraram entre as 10 mais tocadas do mundo em algum momento - e 2019, quando ocorreu o contrário.

Por fim, é válido verificar como é feita a caracterização de cada música no dataset. Assim, na tabela 2 é vista a primeira linha do dataset - a música *Hey, Soul Sister* - e suas informações, modelo que é replicado para as outras entradas do conjunto.

title	artist	top.genre	year	bpm	nrgy	dnce
-------	--------	-----------	------	-----	------	------

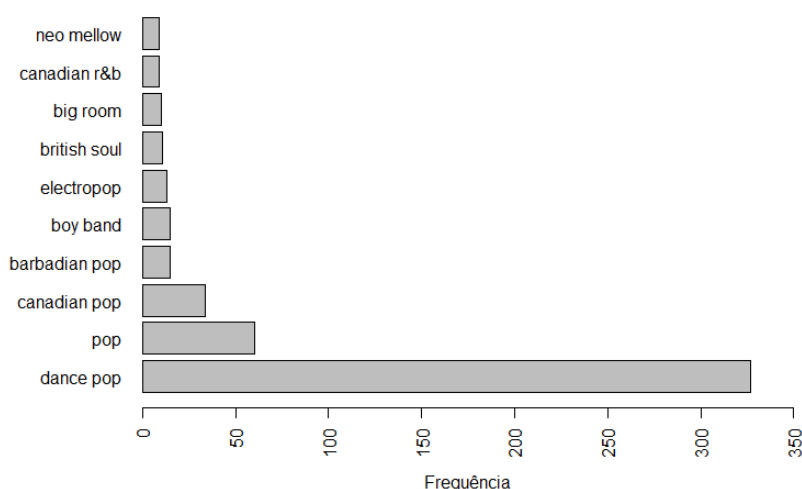
Hey, Soul Sister	Train	neo mellow	2010	97	89	67
------------------	-------	------------	------	----	----	----

dB	live	val	dur	acous	spch	pop
-4	8	80	217	19	4	83

3.2 Resultados

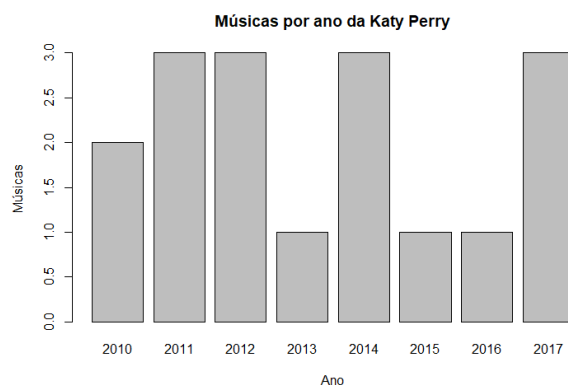
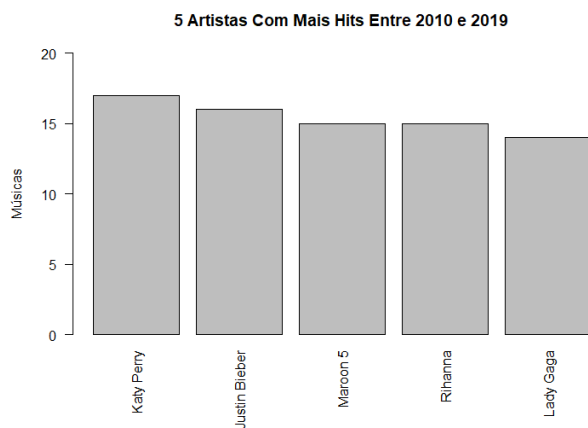
Inicialmente, é possível ver uma tendência no que mais faz sucesso no mundo em relação a música: o *pop*. O *pop* - como o próprio nome diz - vem da palavra popular, o que é bem aplicado para esse tipo de música, que faz sucesso com alguma facilidade por ser feito para isso, músicas para agradarem a todos. Isso pode ser visualizado na imagem 2.

10 Gêneros Mais Tocados Entre 2010 e 2019

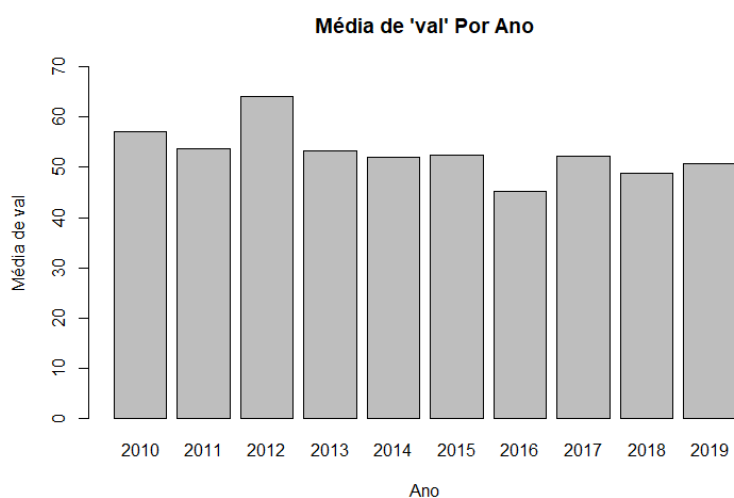


Na imagem, é possível constatar que as músicas *pop* têm dominado as músicas na década, somando mais 70% de todas as entradas do conjunto em todas as suas subcategorias. Com muita vantagem, foi o gênero que mais se destacou no período mostrado. Em contrapartida, alguns gêneros não se saíram tão bem, chegando a ter no máximo 2 hits no período, como *electro house* (ritmo de sucesso nos anos 80 e 90) com apenas uma música.

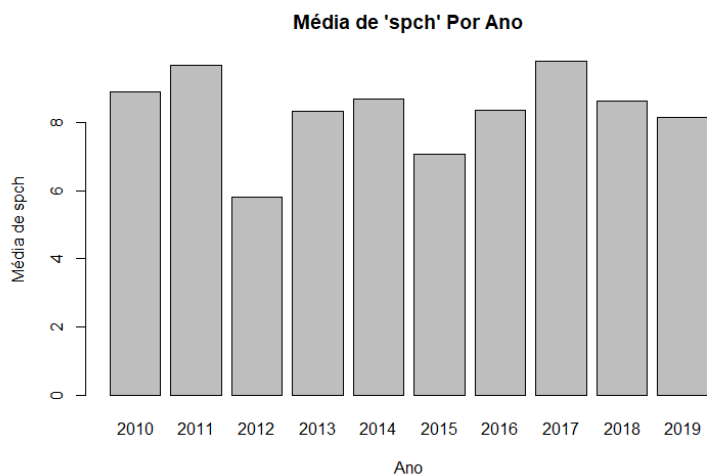
Em sequência, é possível ver na imagem 3 os 5 artistas com mais hits durante a década e, a partir disso, pode-se reafirmar que o gênero *pop* foi o que mais fez sucesso, já que os 5 artistas com mais hits cantam músicas nesse estilo. Na imagem 4, verifica-se a distribuição das músicas da cantora Katy Perry, a qual foi a que mais teve hits durante a década. Percebe-se uma uniformidade muito grande, com pelo menos uma música por ano entre 2010 e 2017, chegando até três em quatro dos anos mostrados. Assim, tem-se um vislumbre do porquê a música *pop* é tão bem sucedida, com uma regularidade considerável durante vários anos em sequência.



Além disso tudo, ainda temos as variáveis numéricas referentes às características mais técnicas das músicas. Como são muitas, a análise ficará restrita a apenas duas para elucidação, a positividade (*val*) e a fala (*spch*), mostradas nas imagens 5 e 6.



Primeiramente, pode-se ver que músicas mais positivas, com um bom clima, têm sido uma tendência na última década, com pouquíssimas variações durante o período. Já em relação à *spch*, há mais variações como visto na imagem a seguir.



Com base no gráfico, é possível atestar as variações na média, o que indica a instabilidade dos valores de *schp*. Dessa forma, pode-se inferir que, no geral, rap e ritmos similares entraram

poucas vezes para as mais tocadas, visto que músicas que têm mais palavras ditas são os *outliers* e não o padrão das músicas mais tocadas na década.

Por fim, com esses dados é possível modelar uma música ideal para a década, utilizando os valores do conjunto de dados para fazer uma média simples, que resultou na seguinte saída:

bpm	nrgy	dnce	dB	live
118.5	70.5	64.38	-5.6	17.7

val	dur	acous	spch	pop
52.2	224.7	14.33	8.3	66.52

Com esses números, pode-se dizer que a música ideal que poderia entrar entre as mais tocadas de cada ano do período escolhido é uma música enérgica e rápida, durando por volta de 3 minutos - devido a valores altos de *nrgy*, *bpm* e *dur* - e com preferência por sons mais elétricos e gravados em estúdio - já que os valores de *acous* e *live* foram os menores. Além disso, a música não deve tender para o rap, já que o valor de *spch* foi mediano, e deve tender mais pro *pop*, devido ao valor um pouco mais alto de *pop*.

4 Conclusão

Ante o exposto, conclui-se que as tendências musicais na década de 2010 foram mais voltadas para as músicas *pop*, com boa energia e som contagiante, sem exagero de palavras nem duração. Isso é justificável, entre outros motivos, pela demanda de rádios e produtoras de fazerem músicas mais abrangentes para seu público, pois quanto mais pessoas foram atingidas por essas produções, mais retorno financeiro é gerado.

Além disso, é preciso destacar a qualidade dos artistas desse gênero que se destacaram com várias músicas entre as mais tocadas, principalmente levando em conta que não basta a música ser divertida se ela não for bem performada em todos os sentidos. A tendência deve ser a manutenção desse cenário, haja vista que por mais que por 10 anos o *pop* tenha dominado, ele ainda fornece muitos resultados bons.

5 Anexo I – Repositório no GitHub contendo o código em R desenvolvido e o Dataset

- https://github.com/gabrielgaia/projeto_spotify

6 Referências Utilizadas

- R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- LAVADO, Thiago. Spotify continua a crescer e aumenta número de assinantes em 21%. **Exame**, 2021. Disponível em: <https://exame.com/tecnologia/spotify-continua-a-crescer-e-aumenta-numero-de-assinantes-em-21/>. Acesso em: 18 de ago. de 2021.
- Top 10 Spotify songs from 2010-2019 – BY YEAR. **Kaggle**, 2019. Disponível em: <https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year>. Acesso em: 10 de ago. de 2021.
- Organize Your Music. **Organize Your Music**, 2021. Disponível em: <http://organizeyourmusic.playlistmachinery.com>. Acesso em: 10 de ago. de 2021.
- WICKHAM, Hadley; GROLEMUND, Garrett. **R para data science: Importe, arrume, transforme, visualize e modele dados**. 1ª edição. Alta Books, 2019.