Task: Data Visualization Spotify dataset tidytuesday – Zsolt Berecz

THE BRIEF EXPLANATION ABOUT THE DATASET 3

THE DATA DESCRIPTION OF THE DATASET: 3

WHAT IS THE MAIN GOAL WITH THE DATASET: 6

TOP 10 RAP SONGS IN THE STARTS OF THE 2020 AND THEIR PROPERTIES... 6

THE DISTRIBUTION OF ATTRIBUTES IN RAP MUSIC... 7

MEAN POPULARITY OF THE RAP AND OTHER GENRES IN 2020 JANUARY SORTED BY YEAR. 9

AVERAGE MUSIC DURATION THROUGHOUT THE YEARS IN RAP MUSIC 11

The brief explanation about the dataset

I love listening to music, I do it at work, at the gym or when I go running and I think a lot of people also really love to do it, that’s why streaming services nowadays like Spotify or Apple Music got popular. I chose the tidytuesday 2020.01.21 the Spotify Songs dataset. The data has a lot of interesting properties such as the songs characteristic or the subgenres. Which made it fun for me to visualize some fascinating facts.

The dataset available here: https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-01-21/readme.md

The data description of the dataset:

The dataset contains ~5000 or more songs from 6 big genre categories such as rock, pop, R&B, rap, Latin and EDM. These genres have subcategories as well. One of the most important things is the dataset has been saved in 2020 January, so all the latest songs are from that time. There are 23 properties or columns of the data, 12 are from the audio features for each track for example liveness, speechiness and danceability.

It seems to me that Spotify uses these features for finding out what are your preferences or what type of music is fitting for you. As your Discover Weekly or the Radio feature, Blend or just the Made for You section.

The other columns of the data are about the track itself, the name of the track, the id, artist, album name, and release date. One of my favorite data from that is the track popularity. That’s a good measurement of what music is viral or underground and not that famous.

There’s also a section about the playlist name, the genre of the music and the subgenre of the playlist. Here is a table that describes the data type, name and a little description of what the data wants to tell us.

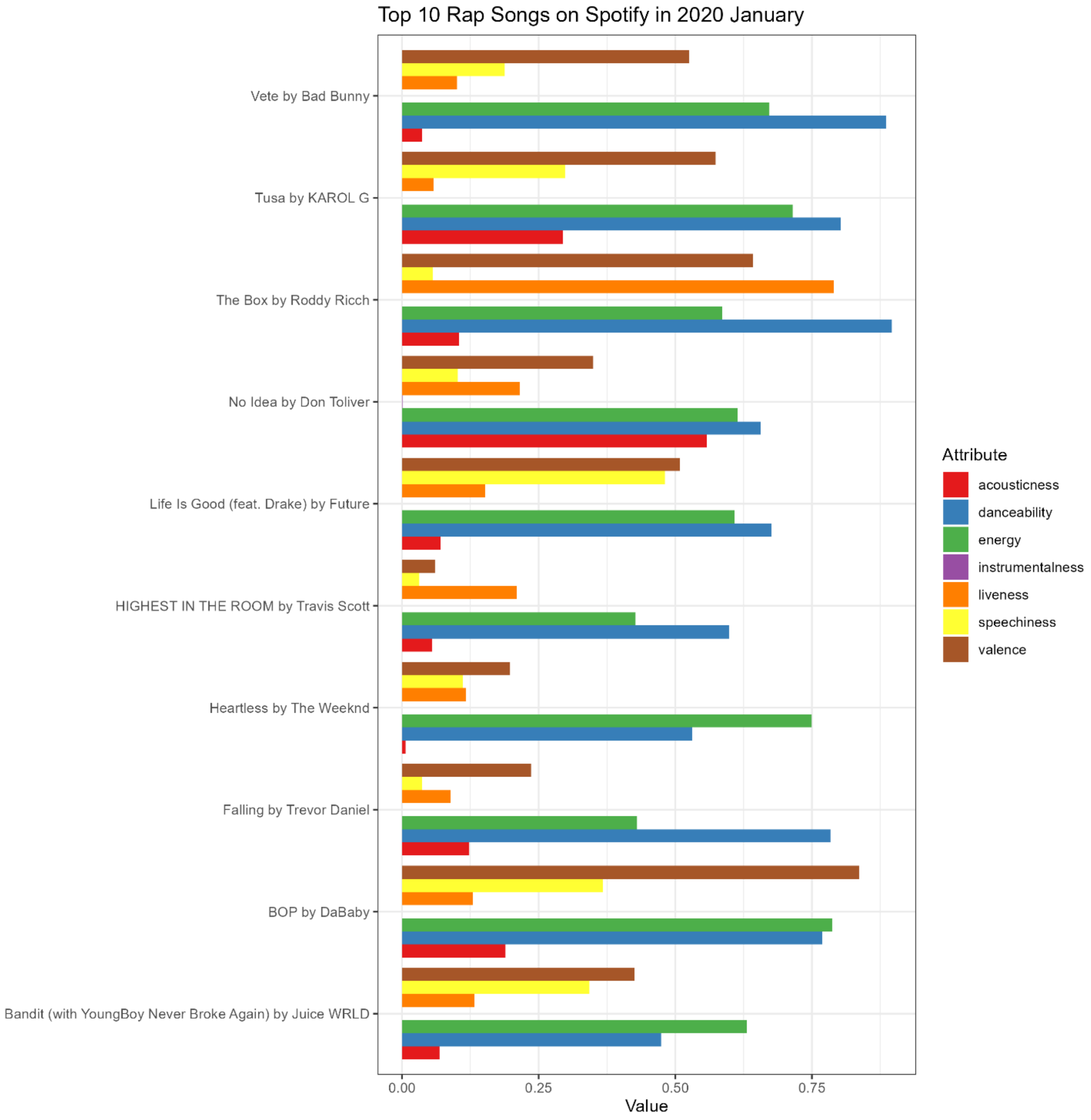
Variable name	Type of the data	Description
track_id	character	Unique ID of the song
track_name	character	Name of the song
track_artist	character	The artist who made the song
track_popularity	double	0-100 describing the popularity of the song
track_album_id	character	Album unique ID
track_album_name	character	The album name
track_album_release_date	character	Date when the album was released
playlist_name	character	Name of the playlist
playlist_id	character	Playlist ID
playlist_genre	character	Playlist genre
playlist_subgenre	character	Playlist subgenre
danceability	double	Danceability describes how suitable a track is for dancing scale is from 0.0 – 1.0
energy	double	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.

key	double	The estimated overall key of the track.
loudness	double	The overall loudness of a track in decibels (dB). Typical range is between -60db to 0db.
mode	double	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
speechiness	double	Speechiness detects the presence of spoken words in a track.
acousticness	double	Measure from 0.0 to 1.0 of whether the track is acoustic.
instrumentalness	double	Predicts whether a track contains no vocals.
liveness	double	Detects the presence of an audience in the recording.
valence	double	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.
tempo	double	The overall estimated tempo of a track in beats per minute (BPM).
duration_ms	double	Duration of song in milliseconds

What is the main goal with the dataset:

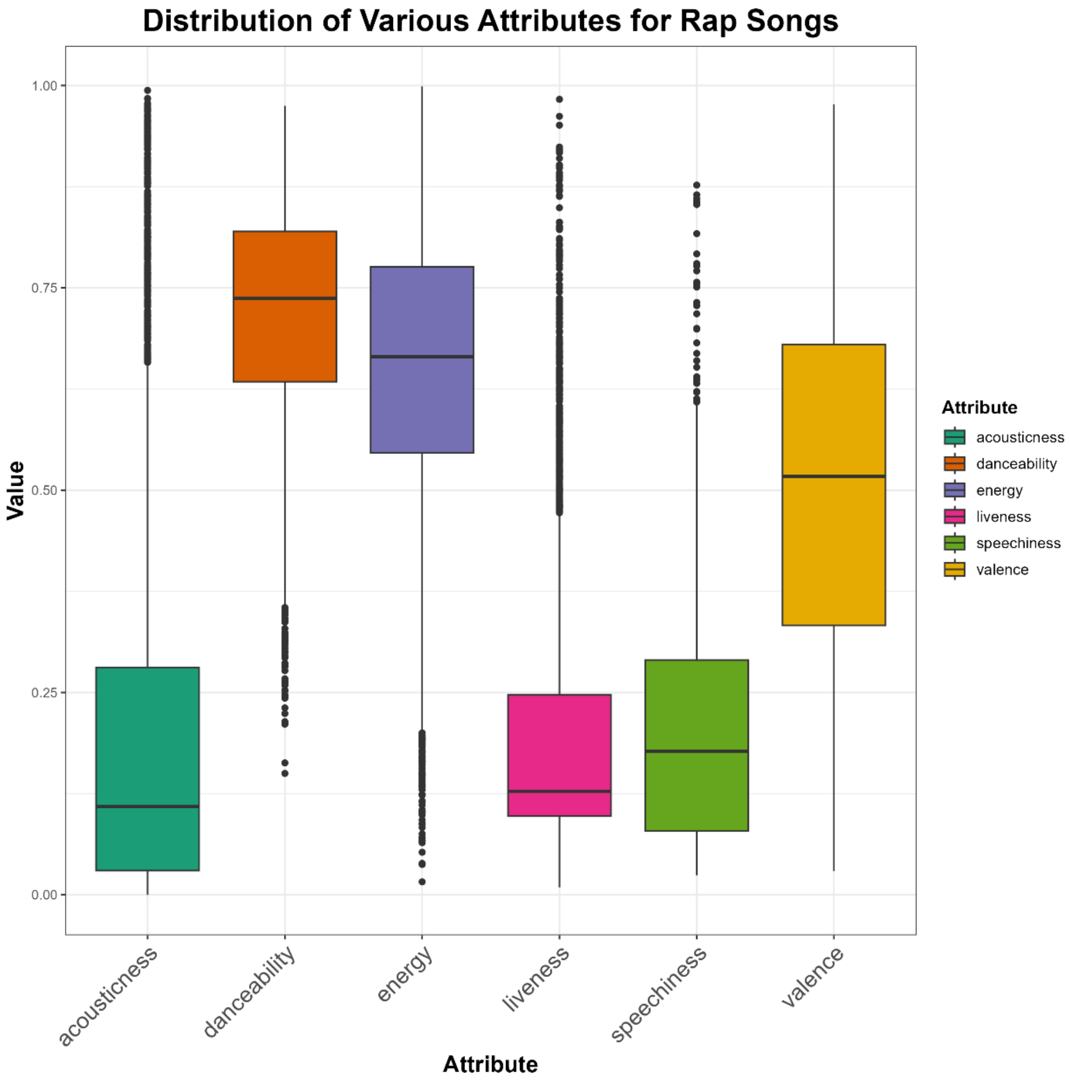
I love rap music, and I was interested in what plus information I can get from this dataset about mainly that genre. In this section I check first the top 10 rap songs and their properties, after that the whole rap genre properties and what we can tell about it. I want to show a very interesting fact that the years passing, and the duration of rap music is shrinking, almost a minute after 30 years. At the end I wanted to check the popularity of rap music throughout the years, because nowadays it's one of the biggest music industries in the world.

Top 10 Rap songs in the starts of the 2020 and their properties

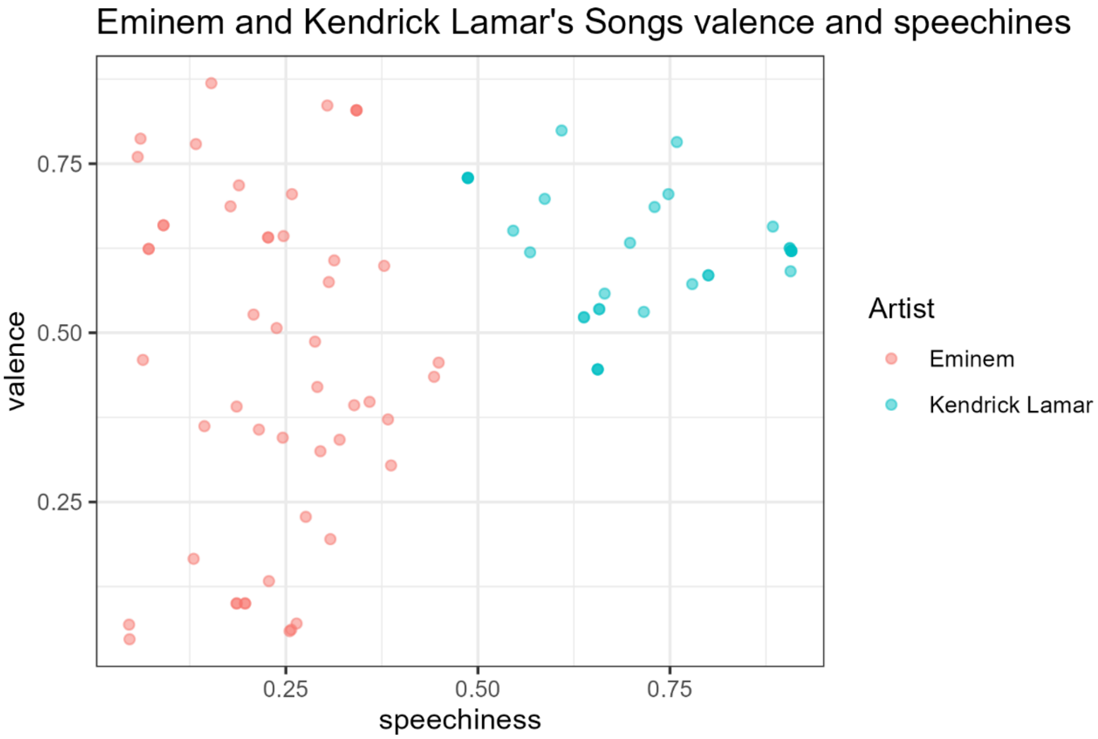


The first plot I did was the top 10 latest rap songs in 2020 January and their attributes and what we can see my goal is to find the most influential attributes in top rap songs. The most important feature that every song has is danceability and energy, I think for a lot of people that is the selling point of a song. The instrumentalness is a non-existent value for rap music that is viral, because it is always some kind of speech or talk is involved, however there are rap songs that are more instrumental. The positiveness of rap music varies, in this dataset almost half of them are positive and half of them are more melancholic. One thing catched my eye, that is the rap songs more tend to be not just about delivering a message, but more tend to use music elements to be catchier. My main problems were the ggplot and the song labels were too big also I wanted to make more distinct colors and to make it not overwhelming. Because my first tries were horrendous. The solution was using horizontal bar plot with RColorBrewer, to make better colors.

The distribution of attributes in rap music



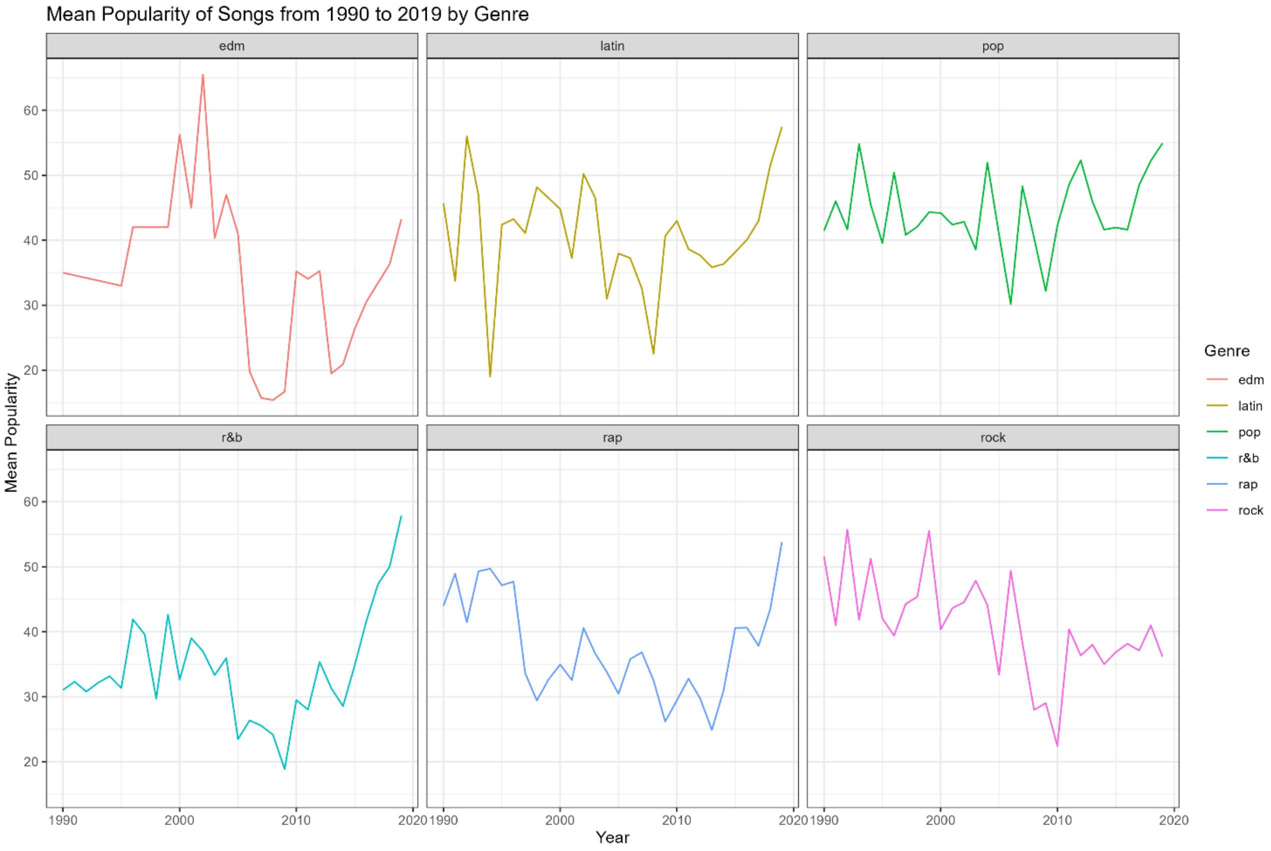
I’ve used a box plot to show from this dataset how much different the values are from the latest top 10 music and in general rap music. We can see from all the rap songs tend to be danceable. Because the median is 0.7 and rap music tends to be catchy as we seen from the top 10 songs also. Energy is also a big slice in this genre. I assumed that in valence we can see that it is balanced that the music is positive or more melancholic and sadder. These are just the combined values, keep in mind that every artist has their own style. For example, two of my favorite artist Kendrick Lamar and Eminem have much different styles.



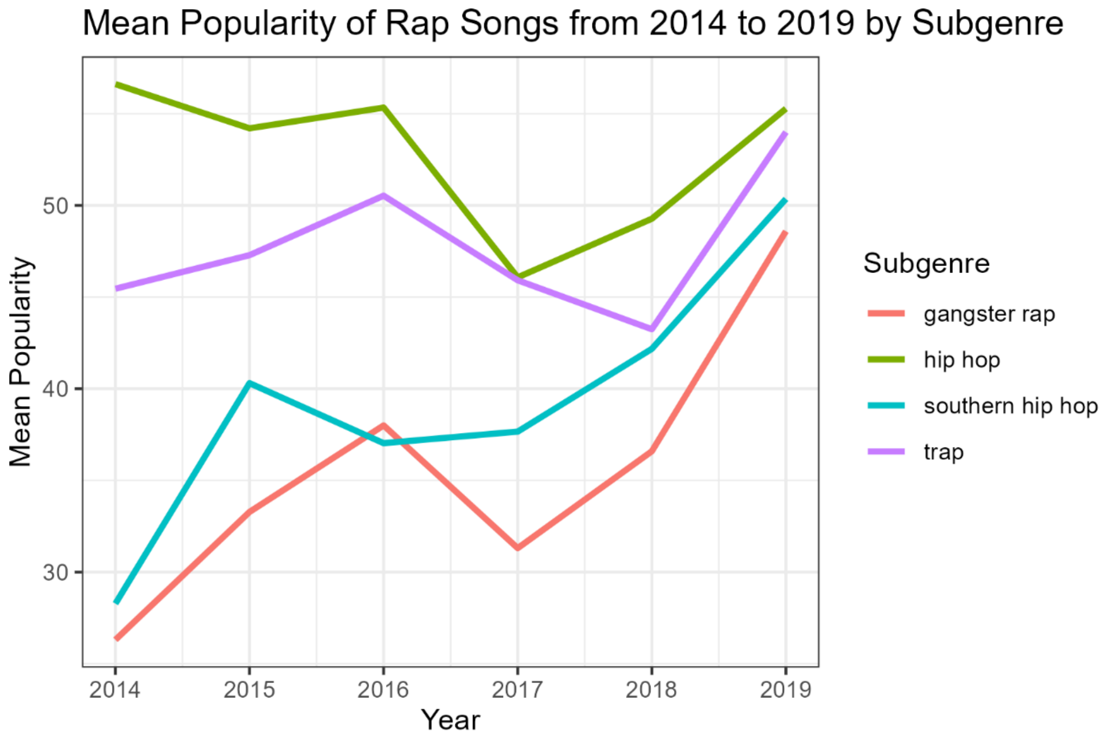
Eminem is more focused on the valence as we can see his songs are volatile, more positive or more depressed, but it uses a lot of music element from the speechiness, Kendrick Lamar more focused on the speechiness and more positive and consistent with his songs.

But in general, the acoustics and liveness and speechiness is not a big factor in rap music, that means it is easy to identify rap songs, because it’s not varying a lot. My problems were first to find out the good values because at first, for example loudness had strange values for me. For me it helped a lot to use dplyr package, also readr is helpful to read the csv data.

Mean popularity of the rap and other genres in 2020 January sorted by year.



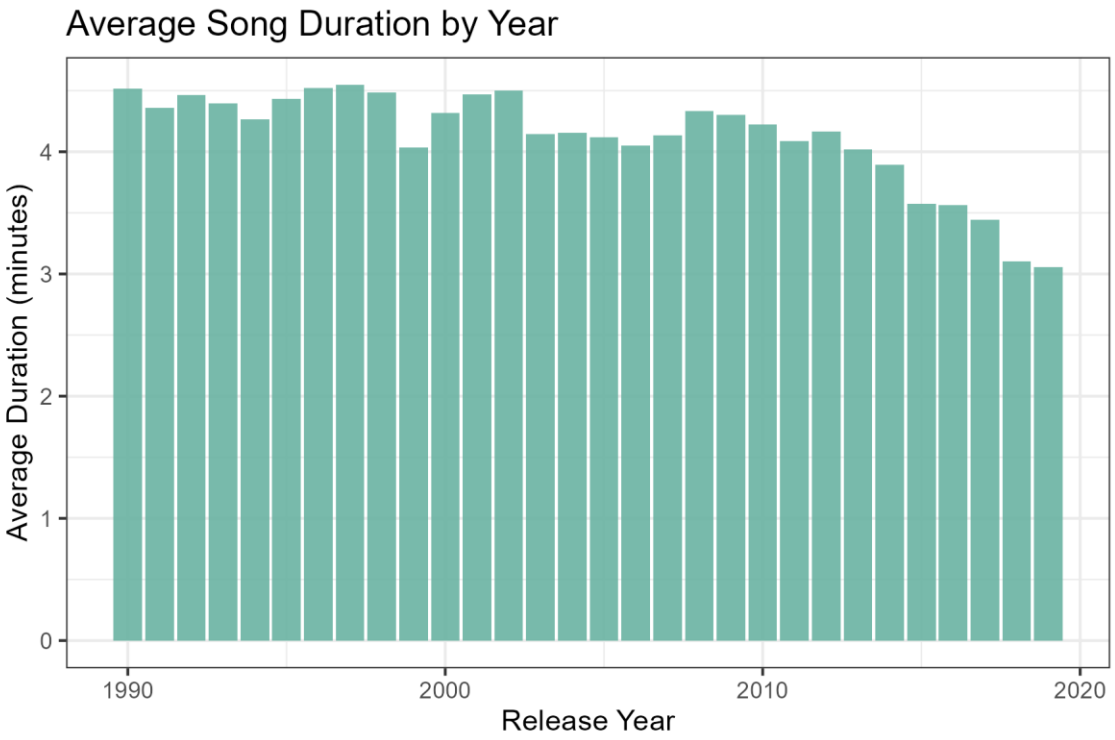
I was curious how popular other genres are by years, and I was surprised that rock music is not as popular as it used to be. The data shows also that the latest music is obviously more popular than the older ones. From the rap music perspective, I saw that a lot of people loved the 90’s rap music or as I can say the golden era of the genre. The same can apply to EDM music, the 2000’s were the time when EDM got popular and had well known music. R&B music got popular after 2010’s before that the genre just not that known, but some tracks are great and well known in the end of the 90’s and 00’s. Popular music or Pop for me not a big surprise, if a star had a great album and tracks it got really known. The next year probably the not known artist made tracks or just stayed with the older tracks, that’s why the big spikes in the line plot. Latin music mainly known in Spain and Latin America, but because a lot of people listen to it, it is mainly constant and has big rating, but the latest songs are even more famous and popularity rating almost hit 55 and ~60.



2014 was a good year for hip hop and rap music genres in general. But mainly hip hop and trap music was known and played everywhere, gangster rap and southern hip hop subgenres were considered underground. Throughout the years these genres got more popular and listened to and, we can see a tendency for a lot of people to get to know rap music and the ratings got higher. There can be also a bias that the latest songs in 2020 January are more played so for Spotify the popular rating is much higher. I just wanted to show that from year to year the ratings got better and better.

Hip hop is the most known style and of course from the previous visualizations I can say because it has the energy and danceability and performances and parties more willing to play these songs. Furthermore, that can apply to trap music. My main problem was with the genres first that trap music values only existed from just 2013 and not before, also I was afraid my data could be biased because of course these data from just the top playlists and representing what people listened in 2020. For packages it helped me a lot after some days I found out tidyverse is a whole package in r for data visualization and the most useful packages in this project. Moreover, when I compared all the genres, I thought at first, it’s also biased because one of my favorite albums is from 2012 for rap. But I think I live in a bubble, and I tried to not believe that.

Average music duration throughout the years in rap music



The last plot I made is also interesting for me. From this dataset I've used all the rap songs and their playlist release date and their duration_ms and I found out throughout the years the songs duration got much shorter, from 1990 to 2019 almost a half a minute or a minute. I think nowadays the music must be catchy or from delivering a long message to the listeners went to just purely have a good time to listen music or try to shorten the time to express their feelings.

Other than that, I think nowadays people have much shorter attention span, and the music industry try to follow it also that trend. That's why the rap music mean duration went to 3 minutes.
In this plot my message was to find out that the music industry changed the 4 minutes long rap tracks to almost 3 minutes. At first my problem was to make out the histogram by year, I had to change the dates and refactor it to years. At first, I wanted to show how much tracks have 3 minutes long or more longer tracks, even there's an 8-minute-long song. But I think this deliver more message.