



Exploratory Data Analysis: Music

November 26, 2021

The Dataset

For this assignment we will be using the 'Spotify' dataset. Each row represents a song.

There are 16 columns. 13 of which are song attributes, one column for song name, one for artist, and a column called "target" which is the label for the song.

Here are the 13 track attributes: acousticness, danceability, durationms, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, timesignature, valence.

Information on what those traits mean can be found here: <https://developer.spotify.com/web-api/get-audio-features/>

To Receive full points for this assignment you must summarize your findings in a latex report, and send to Mr. Merrick **by February 16**.

1. What is the average danceability of songs from Spotify?
2. Is the average danceability of songs a good representation of a typical observation? Use the data to present an argument.
3. Visualize the average speechiness of a song using a histogram.
4. What proportion of songs are from the year 1970 or earlier?
5. Visualize the distribution of durations **in minutes** for songs.
6. Mr. Merrick claims that a typical song has a duration of about 3.9 minutes. What is concerning about Mr. Merrick's claim?
7. What is the average energy of all Justin Bieber songs?
8. Visualize the distribution for the speechiness of Eminem's songs.
9. Visualize the distribution for the danceability of all Justin Timberlake songs.