



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/>

1. What is the average danceability of a song? Interpret the average in context.

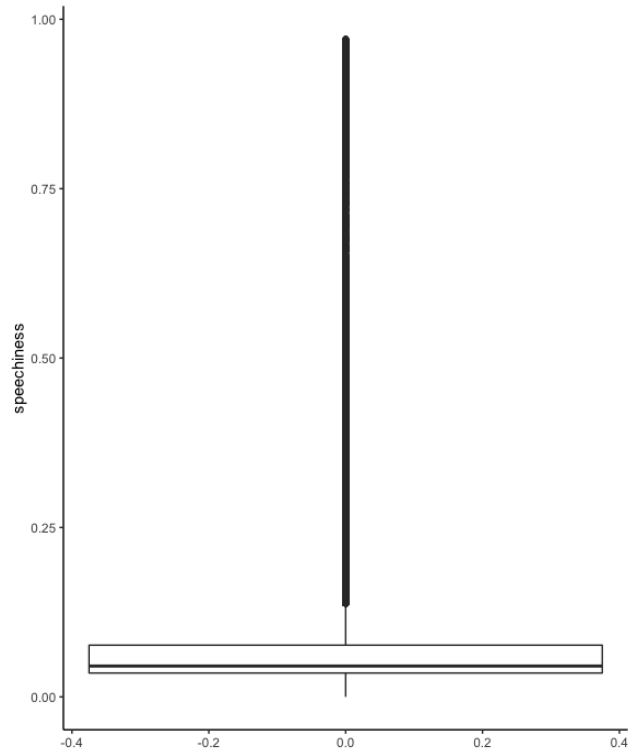
Solution: 0.5367576

2. What is the variance for the variable 'tempo'. Interpret the variance in context.

Solution: 915.3153

3. Visualize the distribution of music speechiness using a boxplot.

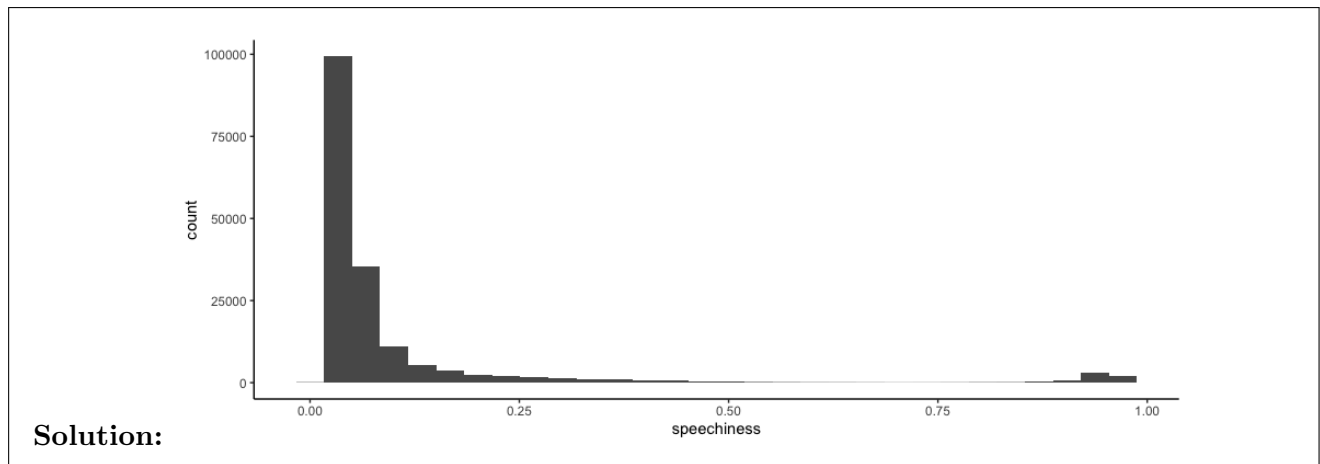
Solution:



4. Does the variable Speechiness have any outliers?

Solution: YES, outliers are shown as dots above the boxplot. There are many outliers on the top end, suggesting this is a skew right distribution.

5. Visualize the distribution of song speechiness using a histogram.



6. Describe the distribution of of music speechiness using the descriptors below:

Shape:

Outliers/Unusual Points:

Center:

Spread:

7. For the remaining class time, create a visualization for 3 more variables within the Spotify dataset and describe **using the same framework**.