Lab 03 – MATH 240 – Computational Statistics

Anya Suko
asuko@colgate.edu

13 Feb. 2024

Abstract

This lab seeks to identify which of the three collaborating artists—The Front Bottoms, All Get Out, or Manchester Orchestra—had the greatest influence on the song *Allentown*.

1 Introduction

By analyzing the musical, rhythmic, lyrical, and other characteristics of each artist's discography and comparing them to the corresponding traits of *Allentown*, we can determine which artist contributed the most to the song.

2 Data Structure

In this lab, the first step is creating a place where all of the Essentia output data can be easily accessed and stored, as well as the output data from the lyrics analysis.

2.1 Organization

Songs are labeled by artist, album, and track name.

2.2 Essentia

Processing the various traits of each artist's discography, as well as those of *Allentown*, is a crucial step in determining which artist contributed the most to the song. These traits—including loudness, dissonance, beats per minute, and more—are generated by Essentia, an open-source program for music analysis, description, and synthesis. Essentia outputs a file containing various trait measurements for each inputted song. Our task is to clean and prepare this data, making it suitable for comparing the characteristic levels of each artist's music to those found in *Allentown*.

2.3 Lyrics

Another important factor that helps indicate which artist contributed the most to *Allentown* is understanding the lyrics.

3 Merging Data

In this lab, we combine lyric analysis data from each artist's discography (from before the release of Allentown) with the Essentia output, organizing everything by artist, album, and track name. This structured dataset makes analysis more efficient and accessible. The result of this merge of data creates a very large database of information.

4 Data Analysis

Analyzing this large database of information is not a simple task, and there are a multitude of ways to analyze the information at hand. One way to analyze the data is to generate a box-plot for each artist, for any given trait.

In order to determine which traits to analyze, a variable called *describe* was created to represent whether the specific song's trait levels were within the range of Allentown, out of the range of *Allentown*, or if they were completely outlying. I chose to analyze the outlying traits because there were far more traits that were within range that out of range, and it would be much more straight forward to find the artist that remained in range, while the others did not.

The traits that I used were overall and average loudness, spectral entropy, dissonance, authentic, and conj. Every one of these traits was out of the range of *Allentown* for at least one artist. I created a box plot for all six traits, and added a red line representing *Allentown*. If the red line went through the interquartile range section on the boxplot, it demonstrated that the artist's level of the trait was similar to that of *Allentown*.

I determined that the artist with the least "out of range" features was Manchester Orchestra, and so it is likely that they contributed most in the development of *Allentown*.

5 Table of Results

feature	The Front Bottoms	All Get Out	Manchester O
overall_loudness	Outlying	Within Range	Within Range
spectral_entropy	Out of Range	Outlying	Within Range
dissonance	Out of Range	Outlying	Within Range
average_loudness	Outlying	Outlying	Within Range
Authentic	Within Range	Outlying	Within Range

6 Graphs

Figure 1: WHAT DO I PUT HERE TO GET THE PLOT FROM CODE FILE????

7 Bibliography

I am trying to cite xtable and essentia, and it is not working. xtable: Duncan Murdoch. *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4, 2021. Available at: https://cran.r-project.org/web/packages/xtable/index.html. Essentia: Cyril Debruyne and David J. Hockman. *Essentia: An open-source library for audio and music analysis, description, and processing*. Accessed May 14, 2022. Available at:https://essentia.upf.edu/.