

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 data can be easily accessed and stored.

2.1 Organization

In order to do organize the data, this lab organizes the data into one file where songs are organized and labeled by artist, album, and track.

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 with the Essentia output, organizing everything by artist, album, and track name. This structured dataset makes future analysis more efficient and accessible. The result of this merge of data creates a very large database of information.

4 Data Analysis - Coding Challenge

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, such as valence, for all songs in the file. Through these box plots, the median level of valence is visible, and easily compared to the *Allentown* level of Valence.

5 Table of results

```
library(xtable)
comparison.output.table <- xtable(comparison.output.df, label="comparison.tab")

## Error: object 'comparison.output.df' not found

print(comparison.output.table,
      table.placement= "H", include.rownames=TRUE, size="small")

## Error: object 'comparison.output.table' not found
```