

# Lab 3 – MATH 240 – Computational Statistics

Pierce Leclerc  
Colgate University  
Department of Mathematics  
pleclerc@colgate.edu

## Abstract

Musical data from The Front Bottoms, Manchester Orchestra, and All Get Out was collected, cleaned, and merged with lyrical data. Data from the collaborative track “Allentown” was isolated to determine which artist was most closely aligned with its production and sound. Box plots of musical and lyrical parameters suggest that “Allentown” aligns closest with Manchester Orchestra, with statistical analysis suggesting a similar conclusion.

**Keywords:** Music Analysis, Batch Files, Data Compilation, Data Cleaning

## 1 Introduction

In 2018, The Front Bottoms, Manchester Orchestra, and All Get Out collaborated on a song entitled “Allentown”. Our goal in this project is to determine which band contributed more to the song by analyzing the music from both bands and comparing data from the new song to their respective collections. In Lab 2, we set up the code for processing each individual track, building a batch file containing the specific lines of code for each song that we can run to collect data using executable files of Essentia (Bogdanov and Wack, 2013). In Lab 3, we compiled data from Essentia and combined the musical data with lyrical data into one data frame. Parameters were plotted using box plots to compare data between each artist and the track “Allentown”. In Lab 5, a function was created to determine whether a given numerical descriptor for “Allentown” was in range, outlying, or out of range of the descriptors from each individual artist’s songs.

## 2 Methods

In Lab 2, we began by utilizing tools from the `stringr` package for R to collect the names of tracks, albums, and artists (Wickham, 2023). We combined them to create unique names for each track’s output `.json` file. We then utilized the `writeLines()` function to write a line of code specific to each track into a batch file. Lastly, we took the sample song Au Revoir (Adios) by The Front Bottoms and analyzed specific characteristics of the song with the `jsonlite` package for R (Ooms, 2014).

### 2.1 Compiling and Cleaning Essentia Data

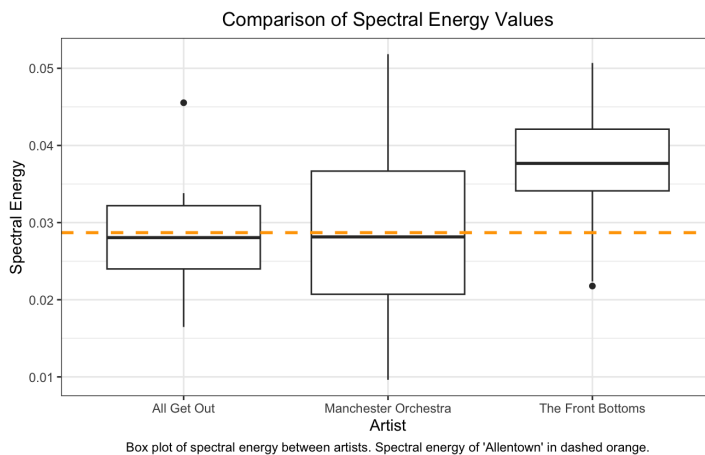
In Lab 3, we took each `.json` file and extracted musical data from each track. From the Essentia models, we

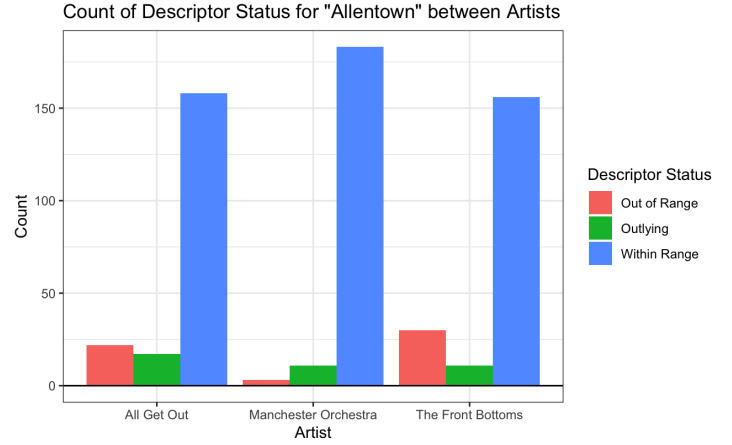
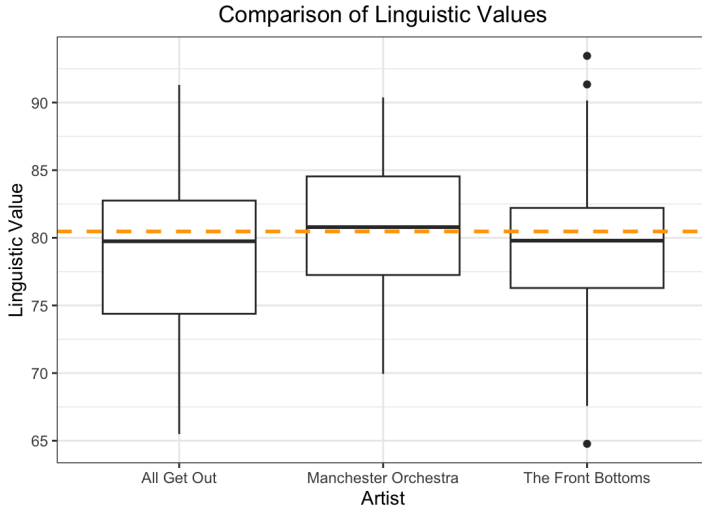
then extracted musical characteristics of each song, including overall loudness, spectral energy, dissonance, pitch salience, tempo (BPM), beat loudness, danceability, and tuning frequency. We compiled the data from all the songs into a data frame, and included additional data from the Essentia models. `valence`, `arousal`, and `timbre` were included, alongside moods such as `aggressive`, `happy`, `party`, `relaxed`, and `sad`. Sound characteristics like `acoustic`, `electric`, and `instrumental` were accounted for as well. Lyrical data from the text analysis tool LIWC was subsequently analyzed and merged into the existing data frame.

## 3 Results

Once the data frame was assembled, two `.csv` files were created to separate “Allentown” from the rest of the songs. Sound and lyrical value of the songs from each artist were grouped in box plots and compared to the spectral energy and linguistic value from “Allentown”.

### 3.1 Box Plots: Spectral Energy, Linguistic Value





In Lab 5, a table was created to store the quantities of each descriptor status by artist. It is clear numerically that Manchester Orchestra has the lowest number of descriptors that are out of range from “Allentown” and the largest number of descriptors in range.

	Artist	Out of Range	Outlying	Within Range
1	All Get Out	22	17	158
2	Manchester Orchestra	3	11	183
3	The Front Bottoms	30	11	156

Table 1: Count of descriptor status for ”Allentown” between artists.

A bar graph was also created to visualize the data from the above table.

## 4 Discussion

From the spectral energy box plot, it is evident that “Allentown” is much more closely aligned to the medians of All Get Out and Manchester Orchestra, while distant from The Front Bottoms. Lyrically, the box plot of linguistic value indicates relation to all three artists, although most closely aligned to the median of Manchester Orchestra once again.

In evaluation of both musical and lyrical characteristics, it appears that “Allentown” is more closely related to Manchester Orchestra than The Front Bottoms. By grouping the descriptors by artist, and comparing their summary values to the descriptor values of “Allentown,” it is evident that Manchester Orchestra has the closest alignment overall, as only 3 of the 197 descriptors are out of range.

## References

- Bogdanov, D. and Wack, N., e. a. (2013). *Essentia: An audio analysis library for music information retrieval*. *International Society for Music Information Retrieval (ISMIR)*, pages 493–8.
- Ooms, J. (2014). *The jsonlite package: A practical and consistent mapping between json data and r objects*. arXiv:1403.2805 [stat.CO].
- Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.