

# Lab 02 – MATH 240 – Computational Statistics

Jack Schaeffer  
Professor Cipolli  
MATH 240  
jschaeffer@colgate.edu

## Abstract

This lab focused on using R code to procure data about various songs for analysis. Additionally, the lab focused on skills producing plots and tables to assist with interpretation.

**Keywords:** Lists; Objects in R, Coding Structures, Graphs, Tables

## 1 Introduction

This lab is focused on the song Allentown, a song released by The Front Bottoms and Manchester Orchestra (and also with additional writing credit to All Get Out). The main question of the lab is determining which artist had the greatest influence on Allentown. To do so, several data processing tools were used. These include Essentia for song analysis and a language analysis tool called LIWC, which utilized JSONLite to manipulate the data in R (Bogdanov et al., 2013) (Ooms, 2014).

This lab stems from determining what qualities or data about music makes up the total sound, specifically posed by the question of what artist had the greatest influence on the song Allentown. To do so, Essentia can be used to process the data from the songs and determine the characteristics of each individual song (Bogdanov et al., 2013). Our goal so far through lab has been to prepare executable lines of code that will produce expansive information on each song. Next, this information can be cleaned to prepare for isolating Allentown and comparing it to other songs to determine what artist had the most influence.

## 2 Methods

The brunt of the work done to prepare each song into an executable line of code was handled using various functions in the `stringr` package to manipulate character names of the file for our desired result (Wickham, 2023). To make sure this was done accurately for every track within a set of given music, two separate for loops were required. `stringr` was used to determine what folders in a given directory correspond to music albums, and within each song of the album various usage of `stringr` functions allowed names to be combined for our desired result. Once this manipulation was complete, each iteration ended with the executable line being appended onto a vector that contained each completed line of code.

## 2.1 Week 2

Our goals for week 2 consisted of an introduction in acquiring specific pieces from a list in R. A JSON file was loaded into R using `jsonlite`, and various objects were created to correspond to the average loudness, key, mean of spectral energy, danceability, bpm, key scale, and length (Ooms, 2014). Using a dataframe, this information can all be compiled together for each song in an easy to read format. From here, additional data was collected for each song. Essentia models were sifted to collect additional information on the differing mood values of each song. Finally, additional data was added to the dataframe that included analysis on the lyrics of each song. Once every song properly contained the information, we had a singular dataframe with every song containing the desired characteristics of the music. Finally, separate `.csv` files were made. In one file we included only the information for Allentown, and in the other file we included every other song.

## 3 Results

The results of this lab bring us very close to our desired goal of analyzing Allentown to determine what artist had the greater influence on the song. In our initial data compilation during week one, the acquired command lines made it easier to collect information on each song. Now, we have all of the information that is required for each song that will allow insight into Allentown specifically. Now that the information has been successfully pulled together into the dataframe, there is data on the technical construction of the music, the atmosphere or mood each song exhibits, and both technical and practical information considering the lyrical construction. Looking forward for completing the future tasks on this lab, the data will be analyzed in order to come to conclusions on Allentown's composition and other insight that can be gained from the data we collected this week.

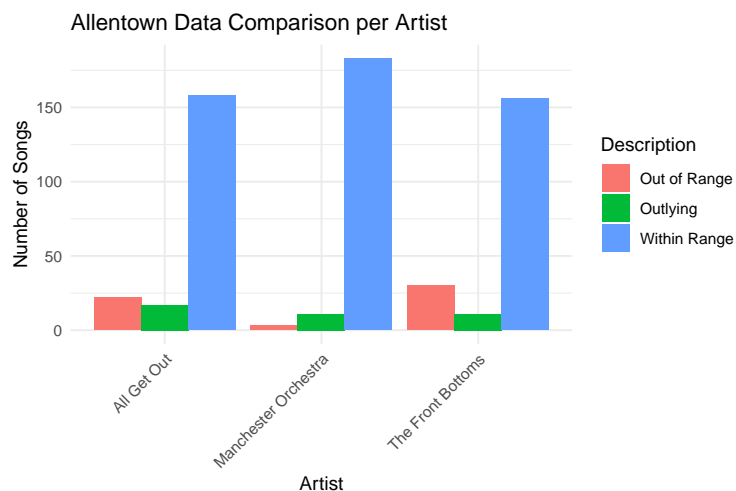


Figure 1: A comparison of Allentown’s data to each artist

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

**Bibliography:** Note that when you add citations to your bib.bib file *and* you cite them in your document, the bibliography section will automatically populate here.

## References

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., and Serra, X. (2013). *Essentia: An open-source library for sound and music analysis*. pages 855–858.

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.