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).In the rest of this report, I will explain how these tools produced helpful data and how this data was analyzed to obtain a proposed answer to what artist most influenced Allentown.

2 Methods

A majority of the work that went into the lab was focused on acquiring data for future analysis of each artist and the song Allentown. The package `stringr` was very helpful for taking our song files and prepping them for data extraction (Wickham, 2023). The data was then collected through JSON files loaded by `jsonlite` as well as Essentia (Ooms, 2014) (Bogdanov et al., 2013). The last set of information included lyrical analysis through LIWC. Once the data was collected, information on Allentown was compared to each song written by the three specified artists to determine what parameters were out of the standard range of each artist. `xtable` and `ggplot` were used to form the table and plot featured below for analysis. These were the main resources used to make interpretations on what artist had the greatest influence on Allentown (Dahl et al., 2019) (Wickham, 2016).

3 Results

The initial result from our lab was an extensive list of information that would be extremely difficult to analyze as it consisted of over 100 different qualities and their values for each song. Using a summary of information for each artist, the code produced results for Allentown and whether it remained within the bounds of each artist for various features. The data was counted and produced the following table detailing the number of times Allentown exhibited qualities that were outside the bounds of 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

Table 1: Count of Allentown in Comparison to Each Artist

The results provide helpful information to compare Allentown to the other artists. Unlike the more complicated information, this table provides a far more understandable group of information that shows that Allentown tends to align more closely to certain artists. An even better way to compare the values of each artist can be done with the following plot:

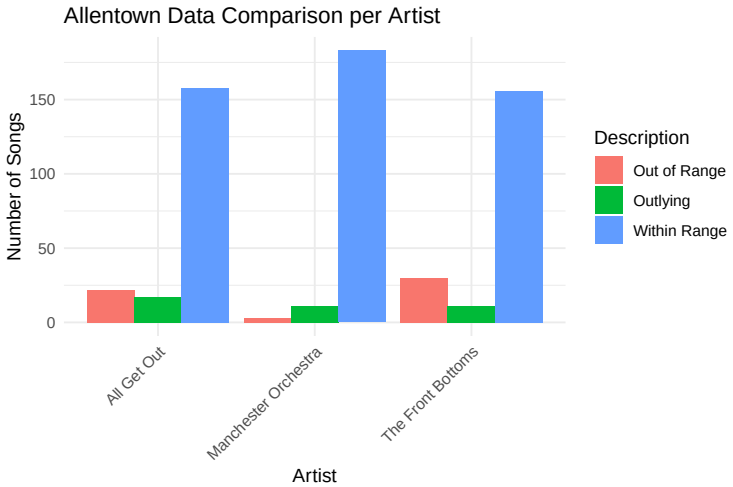

Figure 1: A comparison of Allentown's data to each artist

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.

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.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.