Lab 5 Report – Math240 – Compuational Statistics

Jeremy Artiga
Colgate University
No Department
`jartiga@colgate.edu`

Abstract

Music Analysis can be a daunting task which can take up hours, even days of your time! Thankfully, we have `R` and it's libraries to help us perform this music analysis. We will be analyzing "Allentown" and it's contributing bands to find out who contributed more to the song.

Keywords: This lab will enable us to practice (1) utilizing libraries; (2) working with numerical and character objects; (3) writing `for()` loops; (4) using `tidyverse` syntax; (5) summarize data numerically and visually; (6) using summaries to answer questions.

1 Introduction

The objective for this lab is to use `R` and it's libraries to figure out which band contributed most to the song "Allentown": Manchester Orchestra, All Get Out, or The Front Bottoms?

To accomplish this, we will:

1. Create batch files to process song data using Essentia.

2. Use `jsonlite` (Ooms, 2014) to extract song data from processed `.JSON` files.

3. Collect/Extract data from Essentia (Bogdanov et al., 2013) and LIWC (Pennebaker et al., 2022) models.

4. Combine our Data into a Master data frame for analysis.

5. Summarize/Analyze our data utilizing `tidyverse` (Wickham et al., 2019) and plots.

2 Methodology

2.1 Processing Song Data via Essentia Batch Files and Extracting Data Using `jsonlite`

By utilizing packages and base `R` coding structures such as `for()` loops, objects, and accessing elements within vectors/lists, we were able to create an Essentia command batch file from examination of `.WAV` files in the music directory. This not only enabled us to practice our abilities to utilize the said coding structures, but it also allowed us to build an `R` program which could analyze any sort of directory and pull the `.WAV` files for the sake of data processing into Essentia. We also figured out how to extract data from .JSON files using the `jsonlite` package to gather information on specific albums/songs, which will allow us to collect even more data which could be used in data analysis..

2.2 Scaling up Data Extraction and Data Summarization via Boxplots

The next step in accomplishing our objective is to upscale our data extraction process to collect data from all the `.json` files we now possess from the music processing we did earlier in Essenita. By utilizing the `jsonlite` package and `tidyverse`, we collected a variety of data points and built a data frame. We also furthered our data collection by extracting data from LIWC and Essenita models, which contain additional information on the songs we want to analyze. We constructed additional data frames using the same processes. By the end of this process, we were able to combine all of our data frames into a single "Master" data frame using additional `R` functions, this will be used for our data analysis.

As a start to our data summary/analysis, we plotted three possible song traits between artists that could give us insights on our initial question (i.e. objective); Dissonance and Emotion (One sound and lyrical trait).

2.3 Full Data Summary via `tidyverse` and More Plots

The last task we need to perform to accomplish our objective is to summarize our data and analyze it using a plot. To do so, we will determine how similar Allentown is to each artist (Manchester Orchestra, All Get Out, The Front Bottoms) by song traits. To do this, we will do a mathematical version of the box plot and determine whether Allentown's traits is within range(similar), outlying(marginal), and out of range(not similar) using features in `tidyverse`. By Repeating this process for each song trait for each artist, we will be able to use `tidyverse` to combine the analyzed traits into a data frame. All we need to do now is to count the number of within ranges, outlying, and out of ranges for each artists and put them into a summarized data frame, this will be the data used for our column plots and table which will give us our answer to the initial question.

3 Results

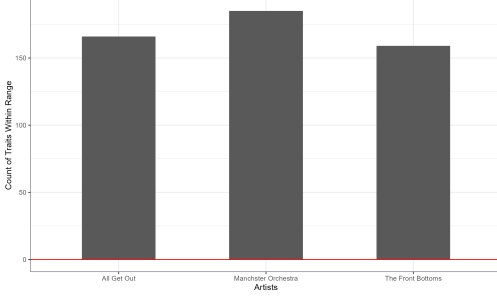| Artists	In Range	Outlying	Out of Range
All Get Out	166.00	17.00	14.00
Manchster Orchestra	185.00	11.00	1.00
The Front Bottoms	159.00	11.00	27.00

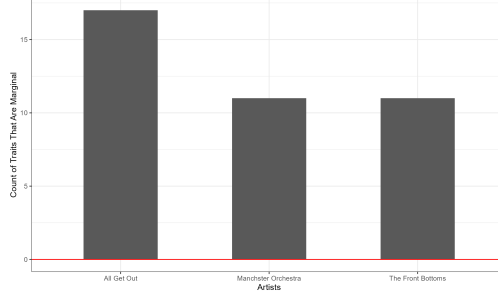Figure 1: Distribution by Similar Song Traits

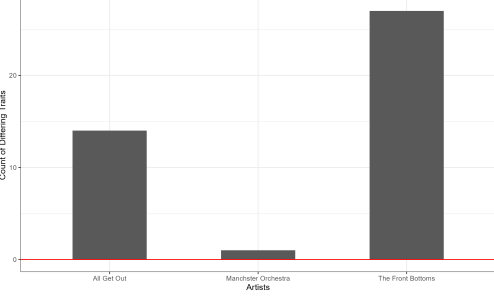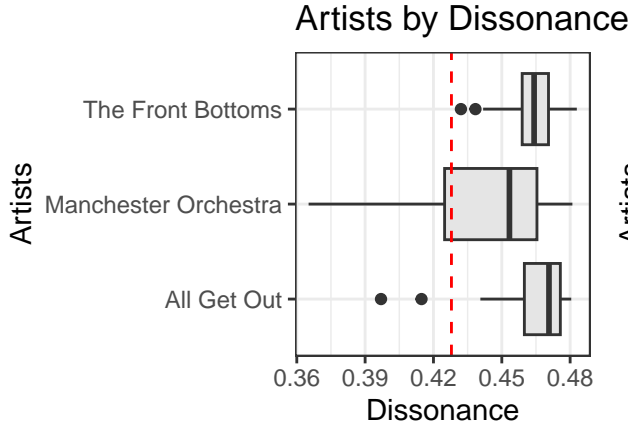Figure 2: Distribution by Marginal Song Traits

Figure 3: Distribution by Different Song Traits

We can now answer the question at hand for this lab, who contributed more to the song "Allentown", Manchester Orchestra; The Front Bottoms; or All Get Out? With the data collection/extraction, data cleaning, and data summarizing/analysis we did earlier, we were able to derive a couple of figures which should allow us to answer the question. The results of our data analysis are shown above. The table we made using xtables (Dahl et al., 2019) figure above illustrates the in range, outlying, and out of range song traits for each artist in a numerical form, we got this from our summarized data frame. The next three column plots are the graphical representation of our table, which help us compare the magnitude of contribution each band had on "Allentown" and allows us to compare artists easier. The last two plots are the box plots from the selected features we chose earlier in our first data extraction. This will serve as additional supporting evidence for our results.

4 Discussion

With the creation of our plots and tables, we can now accurately rank the artists in terms of contribution. Based on the table and the three corresponding column plots, it seems that Manchester Orchestra has the greater contribution on the song "Allentown"! Manchester Orchestra has the most in range song traits while numbering less in outlying and out of range song traits in comparison to the other artists. All Get Out is the second biggest contributor since it has less in range song traits against Manchester Orchestra but more against The Front Bottoms. This would put The Front Bottoms as the smallest contributor to "Allentown". These results are also reinforced by the select features we chose for box plot analysis. The red line on the box plots mark the song trait value for "Allentown". As seen on the plots, "Allentown" falls much more in range to the Manchester Orchestra box plots than the other two artists. Thus, the biggest contributor on "Allentown" is Manchester Orchestra!

References

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. *Proceedings of the 14th Conference of the International Society for Music Information Retrieval (ISMIR)*, 13:493–498.

Dahl, D. B., Graves, S., and Maechler, M. (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]*.

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2022). *LIWC-22: Linguistic Inquiry and Word Count*. Pennebaker Conglomerates, Austin, TX. https://www.liwc.app.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.