Lab 3 – MATH 240 – Computational Statistics

Sankalp Ojha

Colgate University

Mathematics

`sojha@colgate.edu`

Abstract

Will fill out at the end of the lab once we are done with all the tasks.

1 Introduction

Through this lab, we will be exploring who contributed more to the song "Allentown". Was it The Front Bottoms or Manchester Orchestra? To answer the question at hand, we will analyze all the songs released, excluding all joint albums, live albums, and single releases contained in a full album or an Extended Play (EP), by both bands before Allentown. Through the analysis, we will aim to determine the style of each band which will allow us to compare Allentown's style with each band. The comparison will allow us to determine which band's style matches the most with Allentown, leading to the conclusion of which band contributed in a greater quantity.

2 Methods

2.1 Making A batfile.txt

To accomplish this lab, we will first need to create a `batfile.txt` file using the `R` package `stringr` (Wickham, 2023) which we can execute in the command line to produce a `.json` file. Using the `R` package `jsonlite` (Ooms, 2014) we will gather all relevant information needed, such as tempo in beats per minute, musical key or musical mode, to analyze the similarities between the each band's style and Allentown's style.

To create the `batfile.txt` file, I downloaded all the given songs which were assorted by artist with a sub folder for each album. Each album folder contains all the songs in that given album. We retrieved the artist name, album name, and each track's name from the file path to create a file name for each song's `.json` file. Each song's file path and `.json` file name were pasted into an execuatble `batfile.txt` file which will be used to create the necessary `.json` file for each song.

2.2 Gathering All Relevant Data

Once we had all 181 `.json` files for each song, it was time to sift through all the data and gather all relevant information needed for our analysis. We also were provided with Essentia models which came in the form of a `.csv` file. The grand outcome of this portion of the lab is to sift through the data and create a new data frame with only the relevant information we will need for the final portion of the lab which will be to analyze the data. Cleaning up the data first requires combining similar variables into one average variable. For example, for the party measurement, we took an average of the `eff party` and `nn party` varibles to get a mean average party variable. We used this combining process for all the measurements we wanted to analyze at the end. The final step of compling all the relevant informantion was creating a new data frame which only cotained the necesary information such as artist, album, track, and all the new varibles we created.

2.3 Merging The `.csv` Files

The next portion of the lab involved merging our new data frame which we made with all the necessary information and a `.csv` file which contains information about the lyrics and personality types of the songs. We used the `R` function `merge()` to merge the `streaming music extractor .csv`, the new `.csv` we made, and the languge `.csv` file. The new merged dataframe will allow us to extract various pieces of information which we will be able to use for making graphs and plots for analyizng which band contributed more to the making of Allentown.

3 Results

In the appendix there are two violin plots which show a Relaxed V. Artist and timbreBright V. Artist plot. Relaxed and timbreBright are both varibles which measure different aspects on the songs. The merged `.csv` file will is our gateway to creating plots for the analysis portion of the lab.

4 Discussion

As can be seen in the Relaxed V. Artist plot, Manchester Orchestra is more leftward, whereas The Front Bottoms is more rightward. As can be seen in the timbreBright V. Artist plot, Manchester Orchestra is more in the center, whereas The Front Bottoms is more leftward. We will next need to analyze where Allentown falls on these graphs to make any predictions about which band contributed more.

References

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.
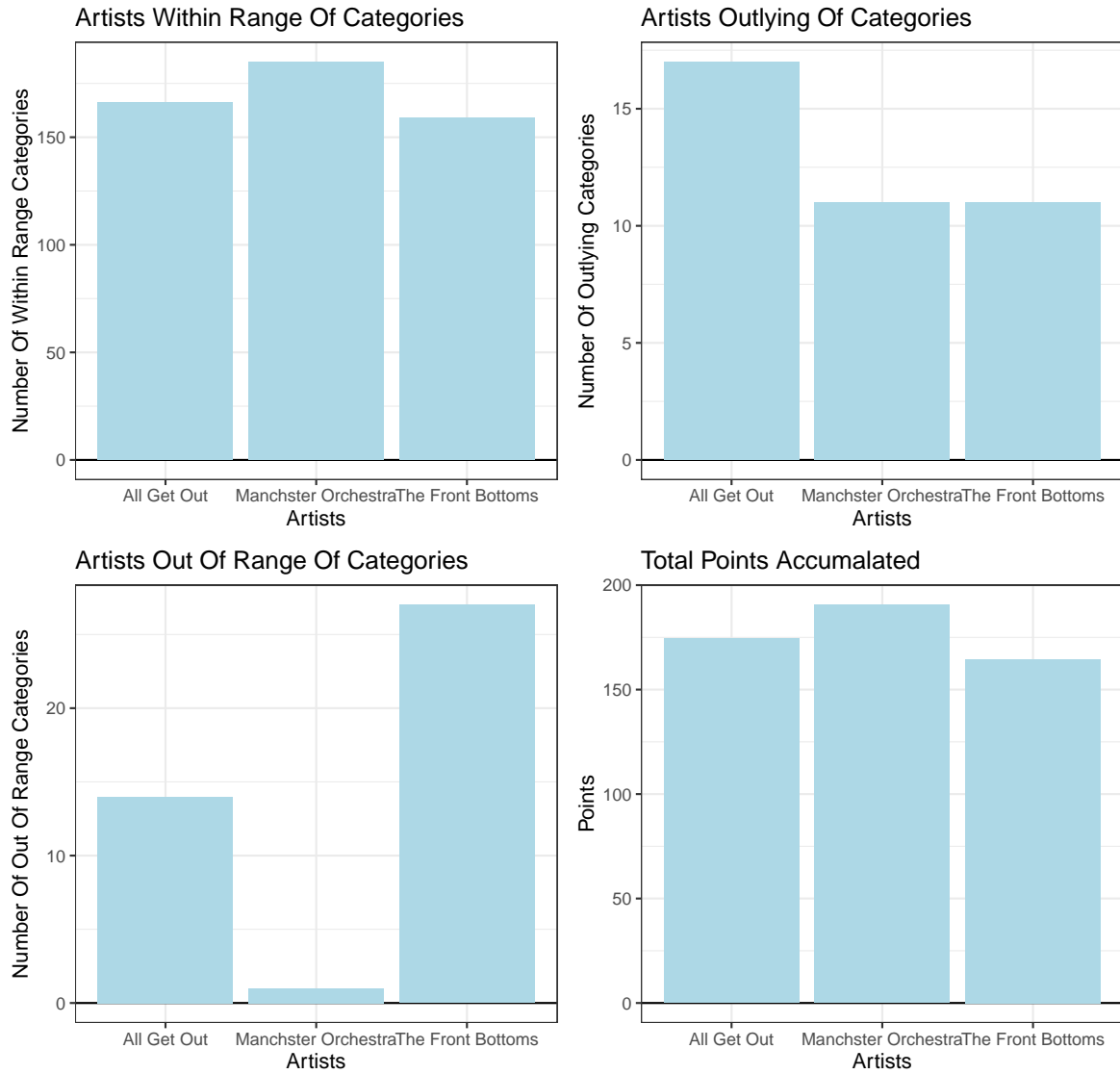
5 Appendix

Figure 1: Relaxed V. Artist

	artists	count.in.range	count.outlying	count.out.range	points
1	All Get Out	166.00	17.00	14.00	174.50
2	Manchster Orchestra	185.00	11.00	1.00	190.50
3	The Front Bottoms	159.00	11.00	27.00	164.50