Lab 5 – MATH 240 – Computational Statistics

Jake Schneider
Colgate University
Department of Mathematics
jdschneider@colgate.edu

2/25/2025

Abstract

In this lab, we aim to find out if Manchester Orchestra, The Front Bottoms, or All Get Out contributed the most to a song that they all collaborated on called "Allentown". In attempt to answer this question we downloaded all releases before "Allentown" totaling to 180 tracks that we analyze various metrics on. Using these metrics we showcased multiple box plots that indicate to us that Manchester Orchestra had the biggest impact on this collaboration.

Keywords: Objects; Loops; Libraries; Data Frames

1 Introduction

In this lab we start attempting to analyze the contributing influences on the song "Allentown" by The Front Bottoms, All Get Out and Manchester Orchestra. The 180 songs that we analyzed consisted of all releases before "Allentown" except for joint albums, live albums, and single releases contained in a full album or an Extended Play. We used Essentia (Bogdanov et al., 2013), Essentia Models (Alonso-Jiménez et al., 2020) and The Lingusitic Inquiry and Word Count/LIWC (Boyd et al., 2022) to aid our analysis of each song. Essentia , an open source musical analysis tool, allowed us do a spectrogram analysis. Essentia Models was used to collect information about the sound of these songs in human terms (happy, sad, angry, etc.). These different tools allow us to analyze a broader array of the impact that each band might have had.

After analyzing all 180 songs and "Allentown" we combined all of the data into one data frame aiming to identify 12 metrics that would help show which band could have had the biggest influence. With these features we created a five number summary which we would use to compare to "Allentown".

2 Methods

The data that we collected on our 180 songs contained both numerical and categorical variables. The Essentia that we extracted from these songs was contained in `.JSON` files, while the other two were in `CSV` files. In order to extract the Essentia (Bogdanov et al., 2013) data from the `.JSON` files we utilized the stringr package (Wickham, 2023) and jsonlite package (Ooms, 2014). Using the `fromJSON()` function gave us a large list. Using an empty vector and a `for` loop allows us to create a data frame with the `.JSON` file as the row and extract the following variables: overall loudness, spectral energy, dissonance, pitch salience, tempo in beats per minute (bpm), beat loudness, danceability, and tuning frequency. After loading both `EssentiaModelOutput.csv` and `LIWCOutput.csv` we now have three data frames of information to work with.

2.1 Cleaning and Merging `CSV` File Data

To tidy up our data in the `EssentiaModelOutput.csv` we took the averages of the following columns: valence, arousal, happy mood, party mood, relaxed mood, sad mood, acoustic, electric, and instrumental. Keeping these averages we removed every other column except for the artist, album and track. The `LIWCOutput.csv` had information regarding thoughts, feelings and personality traits based on lyrics and we left that `CSV` file as is.

Using the `merge()` function we were able to merge our original data frame and the two `CSV` files that we downloaded. To ensure that we did not duplicate any rows during this process we set the artist, album and track columns equal to each others which made sure that when the data frames merged we were still left with 181 rows. With all of the information combined we create two new `CSV` files, one with only the song "Allentown" and one with all the other songs except "Allentown".

2.2 Creating Summary Stats

Using the new `CSV` file without "Allentown" we created a five number summary for each band to determine whether "Allentown" is out of range, unusual or within the range for each band. It is important to note that before creating our five number summary we proceed to keep only the numerical columns in our data frame so that we could compute our five number summary. With all the numerical variables summarized we used the `mutate()` function to create two new columns indicating whether our features value for "Allentown" was out of range, unusual or within range. We define out of range as being less and the minimum or greater than the maximum, unusual as being between the minimum and the 25 percentile or between the maximum and the 75 percentile and within range if it is within the IQR.

2.3 Plotting Data

After creating the summary stats for each band we then filtered each feature to see whether it was unusual or out of range. The reason for this is because if all three bands are within range on a feature then that is not indicative of much and on the same token if all three bands are out of range then that doesn't tell us much either. However, the reason for filtering this way allowed us to see which band might be in range while the other two bands are out of range for certain features indicating that they might have had more of an influence on that feature. Once filtering this way we searched for features that appeared twice which told us that the third feature had to be in range. This left us with 42 features to choose from. We selected 12 features to create box plots for where we also ploted the value for "Allentown".
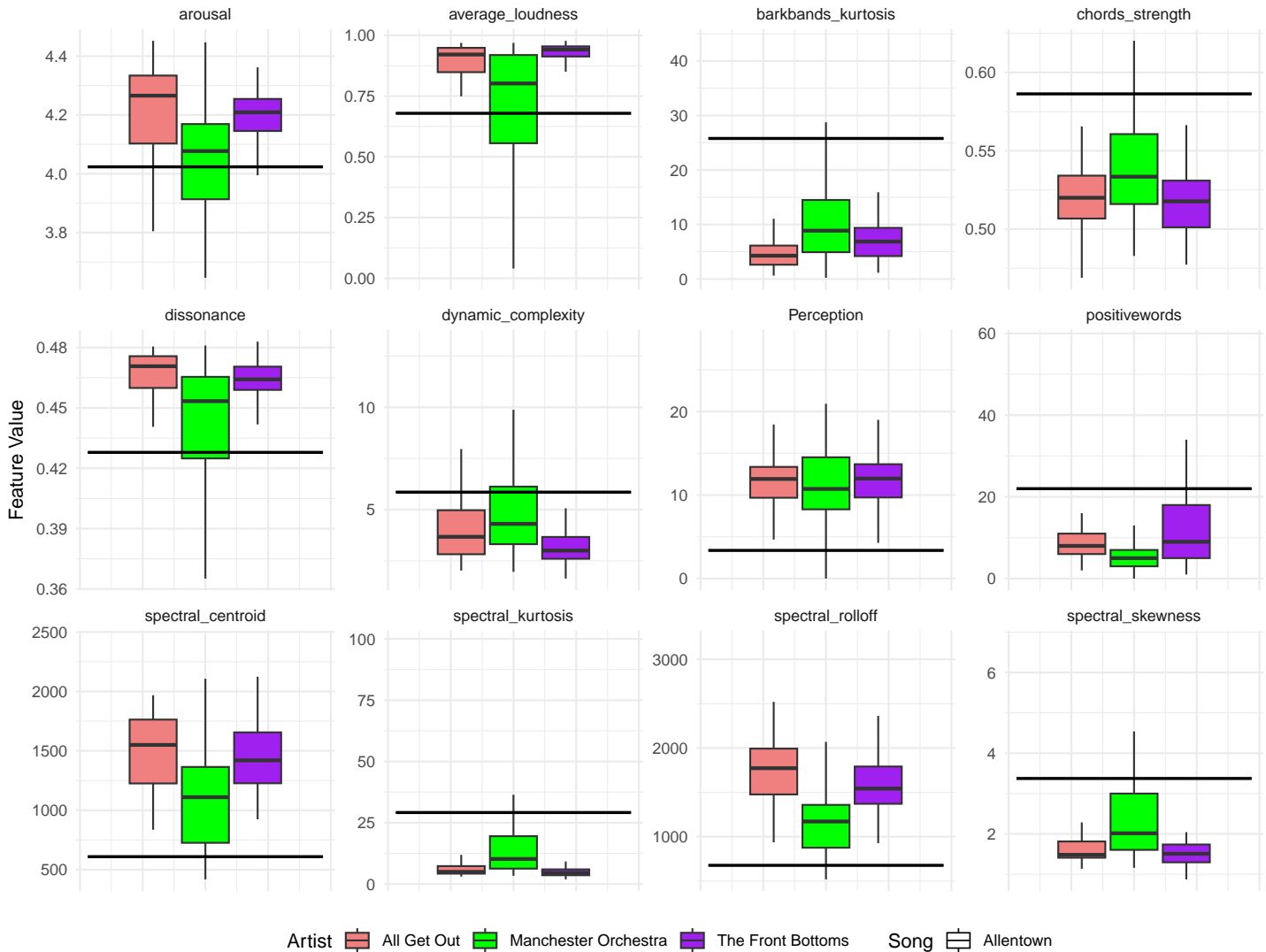
3 Results

To answer our question of which band had the biggest impact on the song "Allentown" have analyzed

4 Discussion

References

Alonso-Jiménez, P., Bogdanov, D., Pons, J., and Serra, X. (2020). Tensorflow audio models in essentia. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 266–270.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., Serra, X., et al. (2013). Essentia: An audio analysis library for music information retrieval. In *ISMIR*, volume 13, pages 493–498.

Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10:1–47.

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

Feature Distributions Across Artists

Artist: All Get Out, Manchester Orchestra, The Front Bottoms   Song: Allentown

	artist	min	LF	UF	max	out.of.range	unusual	description	feature
1	All Get Out	1.14	0.81	2.42	4.12	FALSE	TRUE	Outlying	spectral_skewness
2	Manchester Orchestra	1.16	-0.49	5.09	6.75	FALSE	FALSE	Within Range	spectral_skewness
3	The Front Bottoms	0.87	0.63	2.40	2.04	TRUE	TRUE	Out Of Range	spectral_skewness
4	All Get Out	935.91	701.91	2767.30	2520.04	TRUE	TRUE	Out Of Range	spectral_rolloff
5	Manchester Orchestra	518.87	151.27	2083.17	2566.67	FALSE	FALSE	Within Range	spectral_rolloff
6	The Front Bottoms	927.04	740.58	2421.46	3190.29	TRUE	TRUE	Out Of Range	spectral_rolloff
7	All Get Out	3.03	-0.36	11.90	40.10	FALSE	TRUE	Outlying	spectral_kurtosis
8	Manchester Orchestra	3.36	-13.58	39.36	98.58	FALSE	FALSE	Within Range	spectral_kurtosis
9	The Front Bottoms	1.89	0.00	9.55	10.47	TRUE	TRUE	Out Of Range	spectral_kurtosis
10	All Get Out	836.21	419.54	2569.37	1967.62	TRUE	FALSE	Out Of Range	spectral_centroid
11	Manchester Orchestra	418.64	-231.76	2322.30	2106.83	FALSE	FALSE	Within Range	spectral_centroid
12	The Front Bottoms	924.40	584.67	2298.03	2412.48	TRUE	FALSE	Out Of Range	spectral_centroid
13	All Get Out	2.02	-0.40	8.19	7.96	FALSE	FALSE	Within Range	dynamic_complexity
14	Manchester Orchestra	1.96	-0.90	10.34	13.21	FALSE	FALSE	Within Range	dynamic_complexity
15	The Front Bottoms	1.63	1.00	5.26	6.50	FALSE	TRUE	Outlying	dynamic_complexity
16	All Get Out	0.40	0.44	0.50	0.48	FALSE	TRUE	Outlying	dissonance
17	Manchester Orchestra	0.37	0.36	0.53	0.48	FALSE	FALSE	Within Range	dissonance
18	The Front Bottoms	0.43	0.44	0.49	0.48	TRUE	TRUE	Out Of Range	dissonance
19	All Get Out	0.62	-2.65	11.38	17.68	TRUE	TRUE	Out Of Range	barkbands_kurtosis
20	Manchester Orchestra	0.22	-9.46	28.87	43.71	FALSE	FALSE	Within Range	barkbands_kurtosis
21	The Front Bottoms	1.14	-3.58	17.15	23.90	TRUE	TRUE	Out Of Range	barkbands_kurtosis
22	All Get Out	0.16	0.70	1.10	0.97	FALSE	TRUE	Outlying	average_loudness
23	Manchester Orchestra	0.00	0.01	1.46	0.97	FALSE	FALSE	Within Range	average_loudness
24	The Front Bottoms	0.55	0.85	1.02	0.98	FALSE	TRUE	Outlying	average_loudness
25	All Get Out	0.47	0.47	0.58	0.59	FALSE	TRUE	Outlying	chords_strength
26	Manchester Orchestra	0.48	0.45	0.63	0.62	FALSE	FALSE	Within Range	chords_strength
27	The Front Bottoms	0.48	0.46	0.58	0.57	TRUE	TRUE	Out Of Range	chords_strength
28	All Get Out	3.80	3.76	4.68	4.45	FALSE	FALSE	Within Range	arousal
29	Manchester Orchestra	3.65	3.53	4.55	4.45	FALSE	FALSE	Within Range	arousal
30	The Front Bottoms	3.88	3.98	4.42	4.44	FALSE	FALSE	Within Range	arousal
31	All Get Out	4.67	4.14	18.91	20.89	TRUE	TRUE	Out Of Range	Perception
32	Manchester Orchestra	0.00	-1.06	23.87	28.37	FALSE	FALSE	Within Range	Perception
33	The Front Bottoms	4.27	3.74	19.66	22.56	TRUE	TRUE	Out Of Range	Perception
34	All Get Out	2.00	-1.50	18.50	58.00	FALSE	TRUE	Outlying	positivewords
35	Manchester Orchestra	0.00	-3.00	13.00	27.00	FALSE	TRUE	Outlying	positivewords
36	The Front Bottoms	1.00	-14.50	37.50	34.00	FALSE	FALSE	Within Range	positivewords

Table 1: Five Number Summary of Selected Features