Lab 5 – MATH 240 – Computational Statistics

Camilo
Colgate University
Mathematics Department
cgranadacossio@colgate.edu

2/25/25

Abstract

This lab applied statistical analysis techniques using `tidyverse` (Wickham et al., 2019) in `R` to determine which band has the largest influence on the collaborative song *Allentown* by the All Get Out, The Front Bottoms, and Manchester Orchestra. Using datasets sets from `Essentia` (Bogdanov et al., 2013) and `LIWC` (Pennebaker et al., 2015), I extract musical features to classify the song's similarity to each band. The analysis utilizes **outlier detection, summary statistics, and data visualization techniques** to provide insights

Keywords: Statistical analysis; Data visualization; Outlier detection; `tidyverse`; Musical feature

1 Introduction

In 2018, *Allentown*, a collaborative song by The Front Bottoms, Manchester Orchestra, and All Get Out, was released. The goal of this lab is to determine which band contributed more significantly to the song's composition using statistical techniques. The lab focuses on extracting numerical features from `Essentia`'s dataset, summarizing and comparing these features across the three bands, and using `ggplot2` (Wickham, 2016) for visualization. By implementing boxplots, scatter plots, and summary statistics, I identified patterns that highlight *Allentown's* alignment with each band's musical style.

2 Methods

The dataset includes key features such as `overall_loudness, tempo, danceability, and emotion`. The function `out()` was developed to compute summary statistics and classifying outliers. This was done using `group_by()` and `summarize()`. I computed minimum and maximum values per artist and determined the lower and upper fences $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$ to detect outliers. I applied `mutate()` to indicate whether *Allentown* was **out of range**, an **outlier**, or **within range**. A filtered Data Frame stored the statistical results for all the numerical features.

To begin the analysis of the data I used the `xtable` (Dahl et al., 2019) package to generate a summary table of numerical features. This table provides a detailed comparison of *Allentown's* values against the lower and upper fences showing whether the song falls within range, is an outlier, or is out of range for each feature across the three bands.

Visualization of the results were created through box plots comparing *Allentown's* feature values to each band's distribution and scatter plots highlighting *Allentown's* placement relative to the bands. The `facet_wrap()` function was used to arrange the plots together.

3 Results

The jittered scatter plots provide a visual representation of how *Allentown's* musical features compare to those of the three bands. For `danceability`, *Allentown's* aligns most closely with All Get Out, as indicated by its proximity to the band's typical range. The `emotion` feature, however, places *Allentown* closer to The Front Bottoms. The `overall_loudness` metric is positioned near Manchester Orchestra, but remains within the general range of all three bands. Lastly, *Allentown's* tempo is most similar to The Front Bottoms, thought it is also relatively close to Manchester Orchestra.

The box plots further clarify the extent to which *Allentown's* features deviate from each band's typical range. `Danceability` and `emotion` fall outside the interquartile range but remain within the overall spread, suggesting that *Allentown* exhibits moderate stylistic differences from each band. The `overall_loudness` and `tempo` features, on the other hand, are situated near the lower quartile across all three bands, reinforcing the idea that *Allentown* blends characteristics from multiple sources rather than aligning strongly with one particular group.

The final bar plot categorizes *Allentown's* features based on their classification as out of range, outlying, or within range. Most features fall within the expected distribution for all three bands, indicating a high level of stylistic integration. However, a small proportion of features are classified as outliers across all bands, suggesting that *Allentown* introduces unique elements that differentiate it from the contributing artists' historical compositions.

4 Discussion

Jittered Scatter Plot of Allentown's Features Across Artists

My analysis suggests that *Allentown* integrates elements from all three bands. The `tempo` and `emotion` metrics are closest to The Front Bottoms and Manchester Orchestra, while `danceability` aligns more with All Get Out. The box plots, scatter plots, and summary table reinforce that *Allentown* is a blend of multiple influences rather than strictly fitting within one band's historical musical style.

Boxplot Comparison of Allentown's Features

The bar plot results suggest that *Allentown's* features are least out of range when compared to Manchester Orchestra's historical compositions. This implies that while elements of all three bands are present, *Allentown's* overall structure and musical characteristics most closely align with Manchester Orchestra. The Front Bottoms also exhibit strong influence, particularly on `tempo` and `emotion`, while All Get Out's contribution is reflected in `danceability`. However, since Manchester Orchestra has the fewest features classified as "Out of Range ", it is the best overall match for *Allentown*.

Thus, based on the statistical alignment of *Allentown's* primary musical attributes, Manchester Orchestra appears to have contributed the most to the song's composition, though the song retains influences from all three bands.
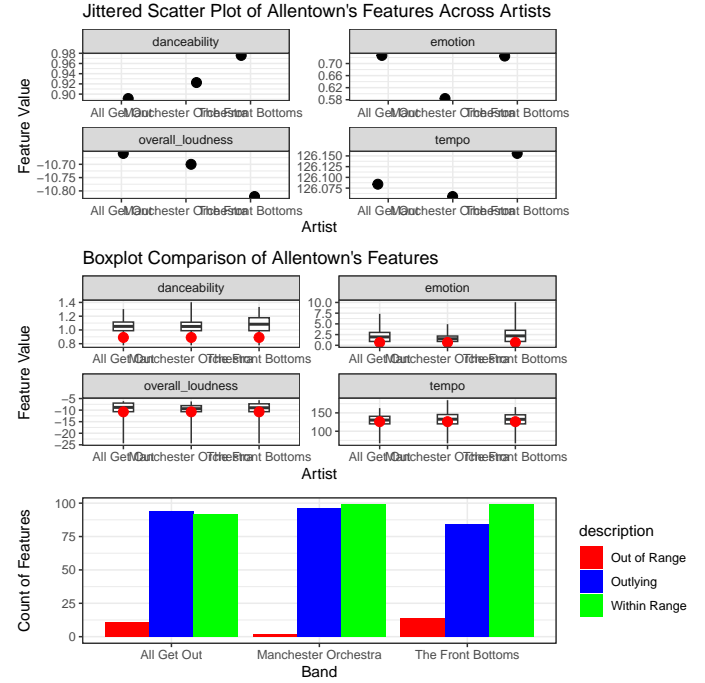
References

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Boyer, H., Mayor, O., et al. (2013). Essentia: an audio analysis library for music information retrieval. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, pages 493–498.

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.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., and Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Pennebaker Conglomerates, Austin, TX.

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

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.