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, three bands, All Get Out, The Front Bottoms and Manchester Orchestra, collaborated on the song *Allentown*. The goal is the determine which band contributed more significantly to the song's composition using statistical techniques.

This lab focuses on:

1. Extracting numerical features from Essentia's dataset.
2. Summarizing and comparing these features across the two bands.
3. Using `ggplot2` (Wickham, 2016) for visualization

By implementing box plots, scatter plots, and summary statistics I identified patterns that highlight *Allentown's* similarity with each band's musical style.

2 Methods

2.1 Data Processing

- The dataseet includes features such as **overall_loudness, tempo, danceability, and spectral_energy**
- Using `group_by()` and `summarize()`, I computed:
  - **Minimum and maximum values** per artist.
  - **Lower and upper fences** $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$ to detect outliers.
- I applied `mutate()` to detect whether *Allentown* was **out of range** or **an outlier** for each feature.

2.2 Statistical Computation

- The function `out()` was created to compute **summary statistics** and **outliers** for each feature.
- A filtered Dataframe stored the results for all the features.

2.3 Visualization

- Box plots: Compare *Allentown's* feature statistical values to each bands distribution.
- Scatter plots: Highlight *Allentown's* placement relative to the bands

3 Results

3.1 Key Findings

- Overall_loudness, tempo, and energy, align more closely with Manchester Orchestra's data

The Data Frame I created compiled 181 songs with extracted musical and sentiment features. The boxplot indicated differences in loudness across bands. The danceability vs BPM scatter plot showed variations in BPM between artists. These insights begin to provide a foundation for determining which band's style aligns most with "Allentown".

4 Discussion

In the future, plots with more variables could be made to better visualize the correlations between artists differing audio features.

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.

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.

5 Appendix