1 Introduction

In 2018, two of my favorite bands – The Front Bottoms and Manchester Orchestra – released a song they collaborated on called "Allentown." In a statement to Noisey (Ross, 2018) – the music arm of Vice – Andy Hull of Manchester Orchestra recalled that the creation of this track started when Nate Hussey of All Get Out sent him the first four lines of the track. Andy Hull worked out the melody and music and shared it with Brian Sella of The Front Bottoms, who then helped develop the chorus.

This brings us to an interesting question: which band contributed most to the song?

To attempt to answer this question, we purchased all releases before "Allentown" except joint albums, live albums, and single releases contained in a full album or an Extended Play (EP), a release that is more than a single release but shorter than a full album. We have 42 tracks by All Get Out, 77 by Manchester Orchestra, and 61 by The Front Bottoms. This totals 180 tracks; there are 181, including "Allentown."

We aimed to use Essentia (Bogdanov et al., 2013) – an open-source program for music analysis, description, and synthesis – to create data about what each band's tracks "sound like." We also used Essentia models (Alonso-Jiménez et al., 2020) to collect additional "sounds like" features (e.g., moods, contexts, instrumentation, etc.).

We also include an additional 122 features created by the Linguistic Inquiry and Word Count (LIWC) engine version 1.9.0 (Boyd et al., 2022). The LIWC is a text analysis tool that uses psychometric dictionaries to provide features that describe the lyrical content with features that describe the linguistics (e.g., word count, and grammar usage) and psychological processes (e.g., power, certitude, tone, and emotion).

In the GitHub folder, I have provided an extended dataset containing 67 features from Essentia's music extractor, 14 features from Essentia models, 118 features from LIWC, and two additional variables from the bing sentiment lexicon (Hu and Liu, 2004). In the coding task below, you will explore these data to find the most discriminating features between the three bands. This lab will enable you to practice (1) using `tidyverse` syntax, (2) summarizing data both numerically and visually, and (3) using those summaries to answer the question at hand.

2 Lab Coding Task: Summarize the Data

Step 1: Write a function that uses `tidyverse` to determine whether Allentown is out of range, unusual, or within the range for each band. Make sure to get your approach working for `overall_loudness` before you attempt to complete the work for the entire dataset. You can create such a function by completing the following.

1. Use `group_by()` to group the data by `artist`.
2. Use `summarize()` to compute the following statistics, where x is the feature of interest:

$$\begin{aligned}
\texttt{min} &= x_{(1)} = \text{the minimum} \\
\texttt{LF} &= Q_1 - 1.5\text{IQR} = \text{the lower fence} \\
\texttt{UF} &= Q_3 - 1.5\text{IQR} = \text{the upper fence} \\
\texttt{maz} &= x_{(n)} = \text{the maximum}
\end{aligned}$$

   Note that the lower and upper fence are boundaries that statisticians often use to determine outlying observations. These quantites are often attributed to John Tukey, who introduced the boxplot (Tukey, 1977).

3. Use `mutate()` to create two new columns:
   - `out.of.range` to be `TRUE` when the feature value for Allentown is less than the `minimum` or more than the `maximum` and `FALSE` otherwise
   - `unusual` to be `TRUE` when the feature value for Allentown is less than the lower fence (`LF`) or more than the upper fence (`UF`) and `FALSE` otherwise
4. Use `mutate` to create a new column called `description` that is `"Out of Range"` when `out.of.range` is `TRUE`, `"Outlying"` when `unusual` is `TRUE`, and `"Within Range"` otherwise.

Note: When you write the function, you'll need to take something like `feature` as an argument. This will replace `overall_loudness` to make the function generalize to any numeric feature we would like to assess. While we know we want `feature` to be a character object containing the column name we would like to assess, `R` and `tidyverse` will not. To specify this, use the `get()` function. For example, if `feature` is a character object containing a column name, you can reference that column using `get(feature)`.
Step 2: Apply this function to the data to determine where Allentown differs from the music each band makes. Note that some of the columns in the extended dataset are not numeric, so it would not make sense to do this for *every* column (e.g., `artist`, `album`, `track`, `chords_scale`, `chords_key`, `key`, and `mode`). Independently investigate the categorical variables for any relevant information.

Step 3: Create a table in LaTeX that summarizes select features for making a decision about which band has the largest imprint on the collaborative track "Allentown." Remember that the `xtable` library can be rather helpful for completing this task (Dahl et al., 2019).

Step 4: Create a graph or series of graphs that summarize the selected features in Step 2 for making a decision about which band has the largest imprint on the collaborative track "Allentown." Ensure to neatly combine any plots using `facet_wrap()` or the `patchwork` library (Pedersen, 2024).

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. IEEE.

Bogdanov, D., Wack, N., Gómez Gutiérrez, E., Gulati, S., Boyer, H., Mayor, O., Roma Trepat, G., Salamon, J., Zapata González, J. R., Serra, X., et al. (2013). Essentia: An audio analysis library for music information retrieval. In *Britto A, Gouyon F, Dixon S, editors. 14th Conference of the International Society for Music Information Retrieval (ISMIR); 2013 Nov 4-8; Curitiba, Brazil.[place unknown]: ISMIR; 2013. p. 493-8.* International Society for Music Information Retrieval (ISMIR).

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.

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.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Pedersen, T. L. (2024). *patchwork: The Composer of Plots*. R package version 1.2.0.

Ross, A. R. (2018). Manchester Orchestra and The Front Bottoms are finally together on "Allentown". https://www.vice.com/en/article/manchester-orchestra-and-the-front-bottoms-are-finally-together-on-allentown/.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.