

# Lab 05 – MATH 240 – Computational Statistics

Danny Molyneux  
Colgate University  
Mathematics  
dmolyneux@colgate.edu

Feb 10 2024

## Abstract

In this lab we are tasked with using tidyverse (Wickham et al., 2019) to gather stats for each musical feature for all three artists (The Front Bottoms, Manchester Orchestra, and All Get Out). We also extract these feature values for their combined song, Allentown. Then, use this information to answer the overarching question: Which band contributed most to the song Allentown?

**Keywords:** tidyverse; summarizing; analyzing; plotting

## 1 Introduction

The goal here is to address the question “Which band contributed most to the song Allentown?”. We have over 200 different features, for 181 different songs. The idea is to extract information on each band by looking at their batch of songs, and compare that information to the features of the song Allentown. Hopefully, we will see significant differences, and that one band has clearly more similar traits to that of Allentown. I will be sure to make various plots/graphs to get various insights and eventually come to a conclusion.

## 2 Methods

Last lab, I was working with a music directory that contains two authors, each having multiple albums, containing multiple tracks. Knowing how to create that batch file was really helpful for this lab, due to the fact that I had 181 .JSON files to process and extract data from.

The new methods that I utilized for this lab were mainly methods that modify data frames. Deleting/adding columns, changing column names, adding columns together, and finally, merging data frames. In order to clean and modify all of the data, I had to install the stringr (Wickham, 2023) package for R. And in order to process .JSON files using fromJSON(), I installed the jsonlite (Ooms, 2014) package for R.

Overall, I ended up working with three data frames that each had 181 rows, and a combined 146 columns (140 after merging due to duplications).

To wrap up this lab, we finally start to work towards answering the overarching question: “Which band contributed most to the song?” I do this by creating plots that give potential insights to the answers of this question.

## 3 Results

Although I didn’t really get any concrete results to the overarching question, now I know how to extract all of these various attributes of a .JSON file, as well as doing it for 180 .JSON files. This data could indicate how much someone contributes to a song, which we will dive deeper into next lab. For example, the length of time each band plays, or how loud they are in relation to each other. Now that we have a lot of data from each band individually, in addition to their song together, we will hopefully be able to get a lot of good insights. The plots I made at the end didn’t give me much information, but it did indicate that in regards to happiness, sadness, overall loudness, and beats loudness, Allentown was more similar to the average Manchester Orchestra song, than the average song by The Front Bottoms.

## 4 Discussion

### Bibliography:

## 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.
- 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