Lab 02 – MATH 240 – Computational Statistics

Danny Molyneux
Affiliation
Department
dmolyneux@colgate.edu

Abstract

In this lab we are tasked with building a batch file to improve efficiency when dealing with many files. In this case, these files are tracks. I use many different functions including paste(), str sub(), str split(), and many more in order to extract and display certain information for each track. The other task is to process the output of a .JSON file. This is helpful because it makes it possible to extract attributes of songs such as average loudness, danceability, etc.

Keywords: Accessing; directories; files; loops

1 Introduction

The goal here is to address the question "Which band contributed most to the song?". We don't go on to answer this question in this lab, but we make progress towards being able to answer that kind of question. This comes from a song in 2018 that The Front Bottoms and Manchester Orchestra collaborated on, called Allen Town.

Like I said, we don't answer the overarching question about who contributed more, but we do build a batch file that can process all of the tracks, as well as process the output of a JSON file.

2 Methods

Here I am working with a music directory that contains two authors, each having multiple albums, containing multiple tracks.

First I had to download the directory of the songs as well as install the stringr (Wickham, 2023) package for R. Then I did some practice with listing directories and using string count() to make a subset of album subdirectories. I used additional methods such as paste(), str.split(), and str.sub() to modify and extract from the various albums and track files. This is helpful when there is a specific desired output like "artist name - album name - trackname". Finally, I used the writeLines() method to create the batch file.

The other main task involved installing the jsonlite (Ooms, 2014) package for R, and using the similar fucntions in order to modify .JSON files. I used fromJSON() to process the output and eventually extract various attributes from the file such as loudness, bpm, and danceability.

3 Results

Tie together the Introduction – where you introduce the problem at hand – and the methods – what you propose to do to answer the question. Present your data, the results of your analyses, and how each reported aspect contributes to answering the question. This section should include table(s), statistic(s), and graphical displays. Make sure to put the results in a sensible order and that each result contributes a logical and developed solution. It should not just be a list. Avoid being repetitive.

3.1 Results Subsection

Subsections can be helpful for the Results section, too. This can be particularly helpful if you have different questions to answer.

4 Discussion

You should objectively evaluate the evidence you found in the data. Do not embellish or wish-terpet (my made-up phase for making an interpretation you, or the researcher, wants to be true without the data *actually* supporting it). Connect your findings to the existing information you provided in the Introduction.

Finally, provide some concluding remarks that tie together the entire paper. Think of the last part of the results as abstract-like. Tell the reader what they just consumed – what's the takeaway message?

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.