Lab XX – MATH 240 – Computational Statistics

First Author
Affiliation
Department
email@domain

Abstract

This lab was split into two parts, one part taught me how to retrieve and process data by building a Batch file and the other part was about processing `JSON` files and working with their data.

Keywords: loops; conditionals; indexing, data

1 Introduction

This lab was the first of a series of labs that aim to answer an important question. This question is, who of the three bands, The Front Bottoms, Manchester Orchestra, and All Get Out, contributed the most to a famous song called "Allen Town" in which they all collaborated on. The focus of this lab was to get comfortable processing and evaluating data.

2 Methods

The lab began by requiring us to build a Batch file for data processing. I was given fake music data to work with, but I needed functions that would allow me to manipulate the titles of the files and directories in a way that allowed me to better organize the data for evaluation later on. Specifically, I needed an R package that would allow me to manipulate strings (directory names, file names, etc.) easily. I used the `stringr` package to gain access to the functions `str_sub()`, `str_count()` and `str_split()` which allowed me to work with the directory and file names to create the Batch file.

Later on, I had to work with a `JSON` file, which required me to download a package, `jsonlite`, that would allow me to read the data in the `JSON` file. This package gave me access to the `fromJSON()` function which I used to load the `JSON` file into my `R` script and then read the data within the file.

2.1 Methods Subsection

In collecting the data from the fake music file, I had to utilize for loops and conditional statements to specify what files I wanted to extract and what directories I wanted to look through.

3 Results

For the first part of the lab, I utilized the `stringr` functions, for loops, and conditional statements to isolate the band name, album name, and track name from the files in the fake music folder. Each file in the fake music folder consisted of sub-folders named after fake artists, and the WAV files within these sub-folders were named using fake album and track names. I had to extract the individual names of the band, album, and track and paste them into a vector according to this template: `streaming_extractor_music.exe / track_file / output_filename_with_the_combined_artist_album_track_names`. Each iteration of the for loop would repeat this process of pasting this template into a vector called code.to.process and then I finished by using the `writeLines()` function to write all the vector values into a batch file called batfile.txt.

The second part of the lab was split into two tasks. First, I had to do what I had previously done and split the `JSON` file name into components consisting of the artist, album, and track names using a for loop and a conditional statement. Second, I needed to utilize `fromJSON` to read through the contents of the `JSON` file and pull out specific metrics. These metrics included the `avergae_loundness`, `mean of spectral_energy`, `danceability`,`bpm`, `key_key`, `key_scale`, and `length`.

4 Discussion

Bibliography: Note that when you add citations to your bib.bib file *and* you cite them in your document, the bibliography section will automatically populate here.

5 Appendix