

Lab 2 – MATH 240 – Computational Statistics

Henry Sun
Colgate University
Department of Mathematics
hlsun@colgate.edu

Abstract

This document provides a basic template for the 2-page labs we will complete each week. Here, briefly summarize what you did and why it might be helpful. Provide all the top-line conclusions, but avoid providing *all* the details. Results should be limited to “we show X, Y, and Z.”

Keywords: Installing, loading, and learning to use libraries; working with characters objects; coding `for()` loops; accessing elements of vectors and lists.

1 Introduction

In the first week of this lab, we built a batch file (`batfile.txt`) for data processing, and we processed a sample `.json` file.

We used the `stringr` Wickham (2023) and `jsonlite` Ooms (2014) packages to accomplish these tasks.

1.1 Intro Subsection

You might need/want to discuss the topics in subsections. Or, you may have multiple questions.

2 Methods

In the first week of this lab, we completed two main tasks: building a batch file (`batfile.txt`) for data processing, and we processed a sample `.json` file. We used the `stringr` Wickham (2023) and `jsonlite` Ooms (2014) files for each task, respectively.

2.1 Building a Batch File for Data Processing

For our first task, we first began by downloading the directory of songs onto the computer and downloading the `stringr` package. By using the functions in the `stringr` package, we were able to identify which directories were albums by using the `str_count()` function. For each album, we identified all the songs by using `str_count()` to count the occurrences of `.wav` in each file. By using a `for` loop through each album, we were able to create a vector called `code.to.process`, with each song being stored in the format as a `.json` and `.wav` file. We used `paste()` and other `stringr` functions to extract `artist`, `album`, and `track`, from each directory or file. Finally, we used `writelnLines()` to turn our vector of songs into a batch file.

2.2 Processing JSON Output

Our second task was to process JSON Output from a sample song, Au Revoir (Adios) on the Talon Of The Hawk album by The Front Bottoms. We first begin by installing the `jsonlite` package, and repeating a similar process in our first task, by extracting the `artist`, `album`, and `track` from our sample file. We used `fromJSON()` function from the `jsonlite` to load the file in as a list, where we could then extract the average loudness, beats per minute (bpm), musical key, song length, among other stylistic features.

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 phrase 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?

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.

5 Appendix

If you have anything extra, you can add it here in the appendix. This can include images or tables that don't work well in the two-page setup, code snippets you might want to share, etc.