

R as a tool for digital survey data

FOAR705 Proof of Concept presentation

Kathryn Phillips, 43686826

Macquarie University

May 27, 2017

Table of Contents

- 1 Introduction
- 2 Scoping
 - Selection
 - Elaboration
- 3 Implementation
 - Creating a Tool Chain
 - Elaborating on my Chain
- 4 Results
 - Pros
 - Cons
- 5 Conclusion

Introduction

My proof of concept for FOAR705 centred around constructing a tool chain using the data analysis software "R". As my MRes research topic is centred around examining the influence of Korean popular music (K-Pop) within the Japanese music market, I plan to collect large quantities of data on listening tastes and the recognition/interest in K-Pop from Japanese residents: especially within the Tokyo Region.

Manual collation of such data, or even the use of Excel spreadsheets could prove difficult and time-consuming. Therefore, I decided that attempting to create a simple tool chain, using R to collect and extrapolate CSV-formatted data sets and subsets, before producing easy-to-interpret visual data (in the form of graphs) would greatly reduce the amount of time spent on analysing the surveys in my actual research time frame.

Initially, I had several areas where using a digital tool could improve the flow and time-consumption of my work.

- Compiling sources for literature research
- Analysis of survey data
- Transcription and Translation of Interviews
- Writing and formatting my final Thesis

Out of these four areas, I considered survey data analysis to be the most time-consuming, so I decided to work on saving time here through implementing a digital tool.

In order to choose the correct tool to use, I had to refine my input and output sources. These went through many iterations, but eventually ended up like this:

- Input: Raw data collected from an online survey site that is capable of producing both qualitative and quantitative data and presenting it in a CSV format.
- Output: A dataset that can receive queries for specific subsets, answer these queries and plot the results in a graphical format.

Due to my choice of input and output, I was able to pick R as my intended tool-of-choice.

Implementation

Creating a Tool Chain

To begin with, I needed to prove that R could take a basic .csv input and create my desired output. To do this, I began by creating my own small table of data, before loading it into the R environment and constructing a plot.

This tool chain was a success, so I was then able to continue onto using more complex data, in a similar style to what I intended on using in my final thesis.

Implementation

Elaborating on my Chain

Now I knew that my tool chain worked with simple data, I decided to create a more complicated chain. This time, I would create a survey; mimicking question formats potentially suitable for my thesis.

For this, I created a simple survey via Google Forms. The number of questions were still kept to a minimum, and I asked my classmates and co-convenors to fill out this questionnaire. I then intended to construct normalised data grids within R, save them each as new, separate CSVs, before taking each and constructing graphs from them.

However, I have not yet completed this task. I intend to focus on this task in the next few weeks (until the date of the final formal presentation).

Results

Pros:

Introduction

Scoping

Selection
Elaboration

Implementation

Creating a Tool
Chain
Elaborating on
my Chain

Results

Pros
Cons

Conclusion

- General growth of my technical skillset
- Success in using R to create simple tool chains
- Increased knowledge of technical tools available for implementation in a range of tasks

Results

Cons:

- R support online is often outdated for the current version (code changes)
- Time is still currently an issue in terms of utilising this tool - however, I would say that if R is implemented properly, there will likely be fewer problems concerning data breaking under manipulation.

Conclusion

So far, my implementation of R as a tool for data analysis; taking simple raw csv data from online survey sites and producing clear graphical data for word processing and typesetting software has been successful. It still remains to be seen if I can effectively manipulate more complicated data. However, I am confident that with this technical skillset I have learnt in the process of undertaking my Proof of Concept, I will be more prepared to solve potential problems with my data, and will have saved time that will be valuable elsewhere during my MRes course.