Khochma חָכְמָה

NoCode ReCode

Last Updated 2025-08-17

Summary

This paper describes a meta-data approach to recoding/upcoding and translating data in R data frames. It generates an Excel workbook, which is edited and then used to generate the recode information in a later step. The paper assumes a working knowledge of R syntax.

Background

Some common reasons to change data are typos or variations (e.g., “auto” and “automobiles”), combining categories (e.g., treading “Apartments-Condo”, “Condo”, “Standard Condo”), etc. I developed this process to handle another reason: dealing with foreign characters in English-centric editors. It can be difficult to type Unicode characters with punctuation and quotations in an editor and have it work correctly.

The NoCode ReCode functions demonstrated below solved that issue.

Design Concept

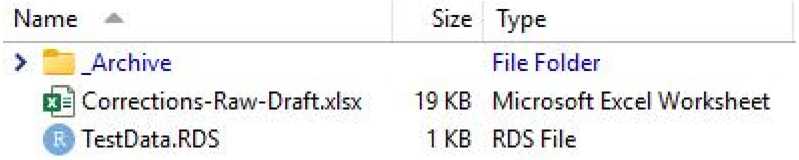
The process is generic in that it intentionally doesn’t need to know anything about the data, only that you want to generate lists of the frequencies and values in one or more fields. A content expert can then review the information to help with corrections, and the resulting information is then used by R to apply the changes.

Detail

Example code (Generating the metadata)	Explanation

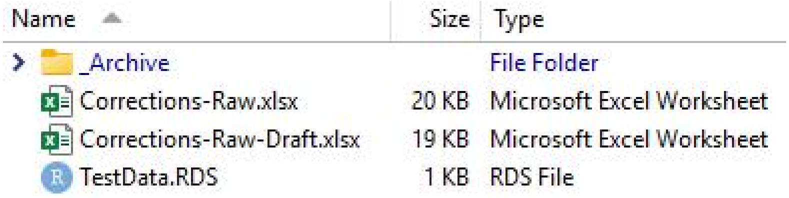
`# list of fields to explore, possibly translate`
`variableList <- colnames(TestData)`
`# Specify the workbook location and name`
`workbook <- paste0(workingDir,`
`                   "Corrections-Raw-Draft.xlsx")`
`# Generate the workbook`
`fnCreateMetaList(TestData,`
`                 workbook,`
`                 variableList)`
	The “variableList” is a vector of the fields in a data frame (“TestData”) that you want to examine. You can do all, but if you have a dataset with millions of unique IDs in a field, then it may exceed your spreadsheet’s maximum. In those cases, it is better to get a sense of what fields are worth exploring and explicitly identify them.
	“workbook “specifies the location and name of a workbook. You can name it however you like, but we recommend you put “-Draft” at the end to distinguish the generated version from the edited version.
	The “fnCreateMetaList()” function is then called.

The folder shows the generated Excel workbook and the saved input dataset.



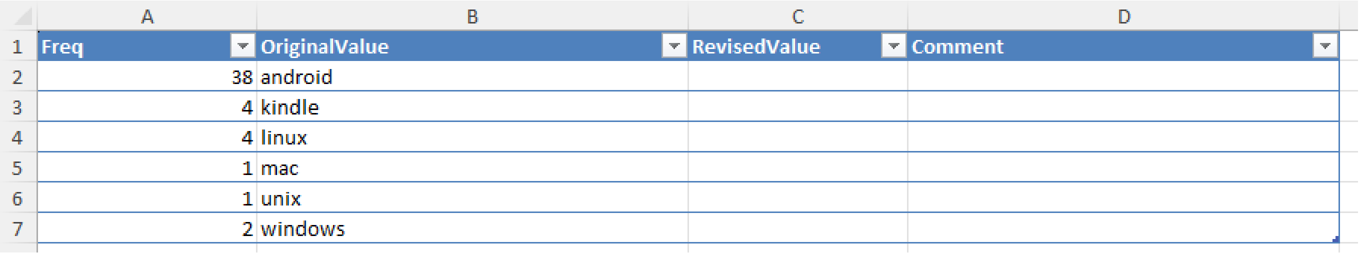
Name	Size	Type
_Archive		File Folder
Corrections-Raw-Draft.xlsx	19 KB	Microsoft Excel Worksheet
TestData.RDS	1 KB	RDS File

We copy and rename the workbook. The edits we want will be put in the renamed workbook. Renaming ensures it won't get overwritten if the code is rerun.



Name	Size	Type
_Archive		File Folder
Corrections-Raw.xlsx	20 KB	Microsoft Excel Worksheet
Corrections-Raw-Draft.xlsx	19 KB	Microsoft Excel Worksheet
TestData.RDS	1 KB	RDS File

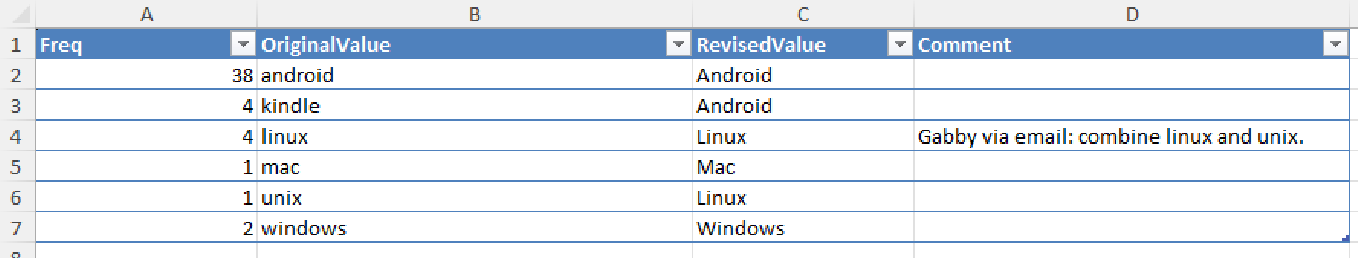
The pre-edited sheet for the "platform" field appears below.



	A	B	C	D
1	Freq	OriginalValue	RevisedValue	Comment
2	38	android		
3	4	kindle		
4	4	linux		
5	1	mac		
6	1	unix		
7	2	windows		

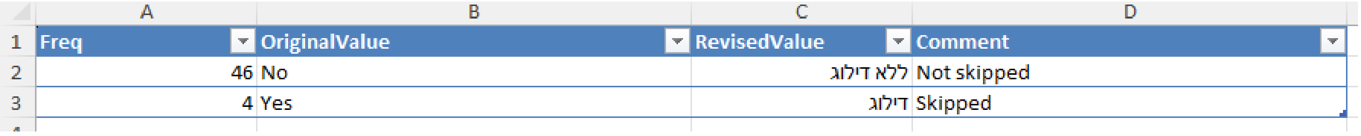
The "Freq" contains the number of occurrences in the data, "OriginalValue"column has the actual values and "RevisedValue" will contain the values we want use instead. "Comment" is an optional field where you can put notes or additional information (such as why the choice was made).

When finished, the sheet now has the recodes in the "RevisedValue" column.



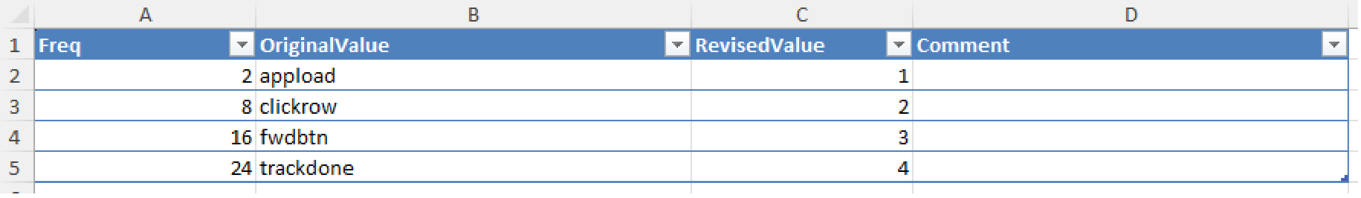
	A	B	C	D
1	Freq	OriginalValue	RevisedValue	Comment
2	38	android	Android	
3	4	kindle	Android	
4	4	linux	Linux	Gabby via email: combine linux and unix.
5	1	mac	Mac	
6	1	unix	Linux	
7	2	windows	Windows	

In the field "Skipped" we see an example of changing the Yes/No to the foreign language equivalent of Skipped/Not skipped. Not only does this show the connection between the original and the revised, but it avoids the difficulty of typing foreign characters in an English-centric environment.



	A	B	C	D
1	Freq	OriginalValue	RevisedValue	Comment
2	46	No	ללא דילוג	Not skipped
3	4	Yes	דילוג	Skipped

In the field “reason_start” we show a recode of words with number values instead of text. ****IMPORTANT**** The process cannot change the underlying type of the field during this process because of how the changes are applied.   If you need to both recode AND change the field type, then after you have finished the recode process, you can use simple code to revise the field type: (e.g., dplyr) `mutate (var = as.numeric(var))`



	A	B	C	D
1	Freq	OriginalValue	RevisedValue	Comment
2	2	appload	1	
3	8	clickrow	2	
4	16	fwdbtn	3	
5	24	trackdone	4	

To apply the metadata, we run the following code:

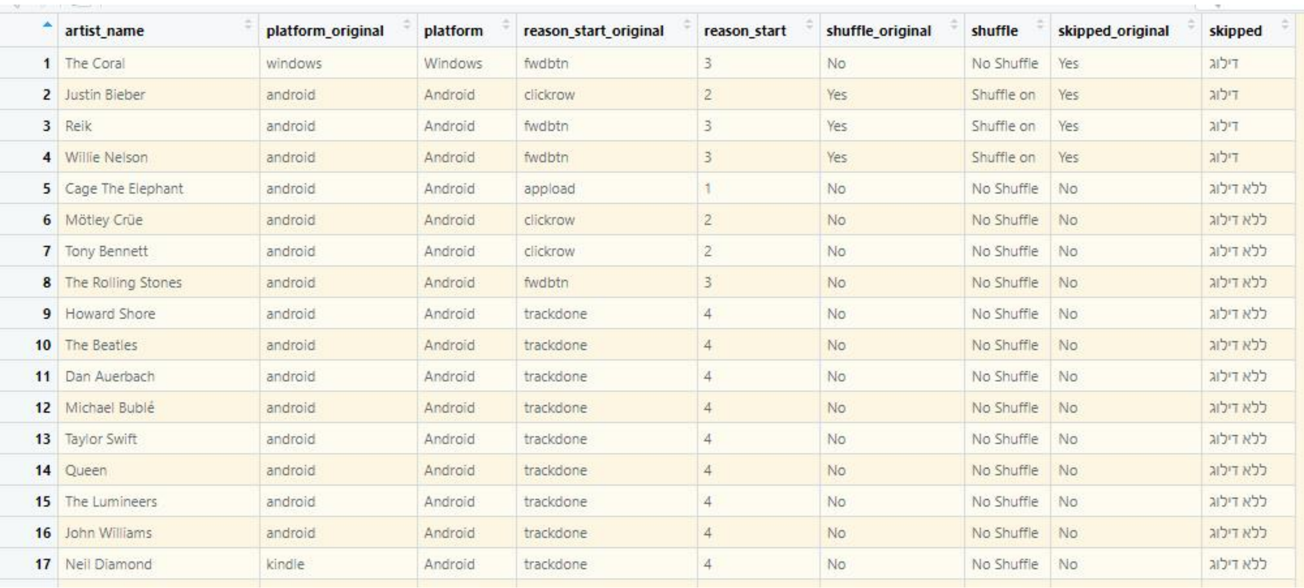
Example code (Applying the metadata)	Explanation

`# list of fields to correct after examination`
`variableList <- c("reason_start", "platform", "shuffle", "skipped")`
	The “variableList” contains just those fields to be recoded.

`# designate the workbook name and location`
`workbook <- paste0(workingDir, "Corrections-Raw.xlsx")`
	The workbook is the edited version, not the “-Draft-“

`# Apply the corrections`
`TestDataUpdated <- fnCorrectFields(TestData,`
`                                   workbook,`
`                                   variableList,`
`                                   "_original")`
	There is an optional parameter where we can specify what suffix to place on the original values. The default is “_old” but you can put anything you want.
	Later, you can want to remove all the original variables with this suffice, just use that text pattern.

The finished dataset:



	artist_name	platform_original	platform	reason_start_original	reason_start	shuffle_original	shuffle	skipped_original	skipped
1	The Coral	windows	Windows	fwdbtn	3	No	No Shuffle	Yes	דילוג
2	Justin Bieber	android	Android	clickrow	2	Yes	Shuffle on	Yes	דילוג
3	Reik	android	Android	fwdbtn	3	Yes	Shuffle on	Yes	דילוג
4	Willie Nelson	android	Android	fwdbtn	3	Yes	Shuffle on	Yes	דילוג
5	Cage The Elephant	android	Android	appload	1	No	No Shuffle	No	ללא דילוג
6	Mötley Crüe	android	Android	clickrow	2	No	No Shuffle	No	ללא דילוג
7	Tony Bennett	android	Android	clickrow	2	No	No Shuffle	No	ללא דילוג
8	The Rolling Stones	android	Android	fwdbtn	3	No	No Shuffle	No	ללא דילוג
9	Howard Shore	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
10	The Beatles	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
11	Dan Auerbach	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
12	Michael Bublé	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
13	Taylor Swift	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
14	Queen	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
15	The Lumineers	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
16	John Williams	android	Android	trackdone	4	No	No Shuffle	No	ללא דילוג
17	Neil Diamond	kindle	Android	trackdone	4	No	No Shuffle	No	ללא דילוג