Getting Your Data Into *R*

from Doing LVC with *R*[*]

Matt Hunt Gardner

2022-09-27

Table of contents

The best way to organize your tokens is in a spreadsheet in Microsoft *Excel*. There are a lot of useful tools in Microsoft *Excel* for automatically coding and re-coding your data. Some of those tools overlap with what I present here for *R*. If you want to explore *Excel*'s functionality, I recommend this website as a springboard: https://support.office.com/en-us/article/Excel-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb. There are other programs for spreadsheet management: *Numbers* in OSX, or Google's *Sheets*, for example. They generally function similarly to *Excel*. *R* does not easily read *Excel*'s default file type (**.xls** or **.xlsx**, though it can be done[1]) therefore, you must save your token file as a tab-delimited-text file (**.txt**) or a comma-separated-values (**.csv**) file.

> ⚠ Warning
>
> I recommend NOT saving your token files as a **.csv file**. Often your token files will include a column of cells containing the broader context the token was extracted from. For example, the sentence in which it appears in a transcript. This broader context usually includes commas. If this is the case, when you save your file as a **.csv file**, it will appear as if there are column breaks in the middle of your broader context because commas are used as the column delimiter.

There are four or five ways to read data into *R*. Which method you choose is really up to you, but because I'm advocating the use of *R* script files and maximum replicability, I suggest using the following function at the top of your script file:

```
td <- read.delim("Data/deletiondata.txt")
```

The function creates an *R* "object" called `td` and then uses the assignment operator `<-` to specify what that object is. In this case `td` is the contents of the tab-delimited text file called `deletiondata.txt` that is located in a folder called `Data`, in the same directory as my script file. If your data is saved somewhere else on your computer, replace `'Data/deletiondata.txt"` with the full path of your data file. The `deletiondata.txt` is a tab-delimited text file and is the data file I will use to teach you about *R*. You can download this same file here[2]. Wherever you save this file, write that file path in quotation marks inside the `read.delim()` function. On a PC this file path will likely begin `"C:/..."`. You can actually just read the file directly into *R* from the web link using the following function:

[*] https://lingmethodshub.github.io/content/R/lvc_r/

[1] https://cran.r-project.org/web/packages/xlsx/xlsx.pdf

[2] https://www.dropbox.com/s/pi8xz1kuo6cz60l/deletiondata.txt?dl=1

```
td <-read.delim("https://www.dropbox.com/s/pi8xz1kuo6cz60l/deletiondata.txt?dl=1")
```

This data file contains tokens of word-final (t, d) and was created for a project studying different pronunciations of word-final (t, d), including deletion. The data comes from a corpus collected for Gardner (2010; 2013; 2017) among English speakers on Cape Breton Island, Nova Scotia, Canada[3].

If you prefer to save your data files as comma-separated-values files (even though you shouldn't, see above), you can read them into *R* using the function `read.csv2()`. If you find it tricky to figure out the file path of your data file you can instead write `file.choose()` (OS X) or `choose.files()` (PC) instead of the file path inside the `read.delim()`/`read.csv2()` function, with no quotation marks. This will create a pop-up window where you can browse through your files and select one. While this *seems* easier, it is not worth it. By not explicitly writing out the file path you introduce a non-replicable element in your script file because there is no record of what you browse to in the actual script file. This means that if you return to your project a year later, or someone else is looking over your code, it might not be clear which data file is supposed to be used. If you choose to use an *R* script file you can actually just drag and drop your data file (or any file) into the script window itself and the full file path will be automatically inserted.

You can also copy a file's filepath to the clipboard on a Mac by pressing **Control** while clicking on the file, then pressing **Option** and selecting **Copy " [your file] " as Pathname**. On a PC you can do the same thing by right-clicking on a file, or (if using Windows 10) using the **Copy path** button on the **Home** tab ribbon in *Windows File Explorer*.

For more information about reading files into *R*, go here[4]

References

Gardner, Matt Hunt. 2010. *Oat and a Boat: Identity creation through regional speech features in Industrial Cape Breton*. St. John's, NL, Canada: Memorial University dissertation.

Gardner, Matt Hunt. 2013. *Variable word-final (t,d) in Cape Breton English*. Toronto: University of Toronto dissertation.

Gardner, Matt Hunt. 2017. *Grammatical variation and change in Industrial Cape Breton*. University of Toronto dissertation.

[3] https://en.wikipedia.org/wiki/Cape_Breton_Island

[4] https://stat.ethz.ch/R-manual/R-devel/library/utils/html/read.table.html