Reading in Data

Dr Andrew J. Stewart

Fellow of the Software Sustainability Institute

E: drandrewjstewart@gmail.com
T: @ajstewart_lang
G: ajstewartlang

Reading in .csv Data Stored Locally

```
Andrew_project
|--Andrew_project.Rproj
|--raw_data
        |--data.csv
|--tidied_data
        |--tidy_data.csv
|--R_scripts
        |--data_tidy.R
        |--data_vis.R
        |--analysis.R
|--markdown_scripts
        |--analysis.Rmd
|--output_reports
        |--analysis.html
        |--analysis.pdf
```

Imagine your folder structure looks like what we have on the left. Inside your folder you have a number of sub-folders. As you can see, we have one folder entitled “raw_data”. In this folder we have one data file entitled “data.csv”. Our script is in the folder “R_scripts”.

To read in the “data.csv” data file in our data_tidy.R script and assign it a variable called my_raw_data we would write:

```
my_raw_data <- read_csv("raw_data/data.csv")
```

This mean go to the “raw_data” folder in the project folder (where the .Rproj file sits) and the file “data.csv”. The expression `"raw_data/data.csv"` is the file path.

Reading in .csv Data from a Webpage

Quite a lot of the data we'll be working with in this unit is hosted on GitHub. Reading in data from GitHub is easy (assuming you have a working internet connection).

Imagine we want to read in a Tidy Tuesday dataset such as ratings for the US version of "The Office". The data are stored in a .csv file here:

```
https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-17/office_ratings.csv
```

To read in the data to R and assign it to the variable "office_data", we'd simply type:

```
office_data <-
read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-17/office_ratings.csv")
```