

# Week 4 Programming Assignment

Instructor: Chien-Hao Fu

3/14/2019

**Due: 3/20/2019**

## Data

- 縣市重要統計指標 <http://statdb.dgbas.gov.tw/pxweb/Dialog/statfile9.asp>

## Prerequisite

1. Go to the course website and download the following file:  
`counties.xlsx`
2. Install the packages: `readxl` and `tidyverse`

## Programming Assignment

### Instruction

Create an R script and follow the instructions to draw a simple plot

1. Load the package `readxl` and `tidyverse`
2. Import the file that you have just downloaded using the `read_excel` function and save the imported data into a `data.frame` object
  - a. When you load the file, remember to specify the missing value indicator as “-”
  - b. You might need to specify the working directory by using the function: `setwd()`

3. Examine the data and answer the following questions:
  - a. Q1: What was the share of elderly people (variable: **elder\_share**) living in New Taipei City in 1998?
  - b. Q2: What is the data type for variable **Year**?
4. Process the **Year** variable
  - a. Re-code the variable **Year** as a numeric variable using the **as.numeric()** function
  - b. Rename the variable from **Year** to **year**
5. Process the **Area** variable
  - a. Rename the variable from **Area** to **area**
  - b. Re-code the variable **area** as a factor variable (using the function **factor()** to recode the variable)
  - c. Q3: Answer the question: how many levels does the new **area** variable has? [hint: use the function **levels()**]
6. Create a new data frame. In this new data frame, **exclude** observations that belong to **Total** and **Taiwan** [hint: you can use the function **filter()** ]
7. Create another data frame based on Step 6. In this new data frame, keep only the observations from year 2015
8. Use the function **plot()** to draw a scatter plot for your data frame created in Step 7. In this plot, use the variable **self\_funding** (縣市自籌財源比率) on the x-axis, and use **elder\_share** (65 歲以上老年人口比率) on the y-axis
9. Q4: Is there a relationship between the two variables? Can you spot and identify the outlier(s)?
10. Submit your answers to Q1-4, the plot (from step 8), and your source code using the file: “W4 Student Submission.docx”.

Your plot should look similar to the following graph:

