

EPPS6323 Knowledge Mining

Assignment 2 – Brush up R and Quarto

1. Lab01, Lab02 in R (Download the .qmd files in Teams)
 - a. Render and upload your files to GitHub website
 - b. Be sure to upload associated folders (e.g. Lab01.html, Lab01_files, Lab02.html, Lab02_files)
2. Review Chapters 3 to 7 in [R4DS](#), Section 3.5 in [Data programming](#). Run an exploratory data analysis with R using the TEDS2016 dataset (sample codes as follows):

```
library(haven)
TEDS_2016 <-
read_stata("https://github.com/datageneration/home/blob/master/DataProgramming
/data/TEDS_2016.dta?raw=true")
```

3. What problems do you encounter when working with the dataset?
4. How to deal with missing values?
5. (Next step) Explore the relationship between **Tondu** and other variables including **female**, **DPP**, **age**, **income**, **edu**, **Taiwanese** and **Econ_worse**. What methods would you use?
6. (Next step) How about the **votetsai** variable (vote for DPP candidate Tsai Ing-wen)?
7. (Next step) Generate frequency table and barchart of the **Tondu** variable. Assign labels to the variable using the following:

```
TEDS_2016$Tondu<-as.numeric(TEDS_2016$Tondu,labels=c("Unification now",
"Status quo, unif. in future", "Status quo, decide later", "Status quo
forever", "Status quo, indep. in future", "Independence now", "No response"))
```