

# EPPS6323 Knowledge Mining

## Assignment 2

1. Lab04 in R
2. Review Chapters 3 and 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. 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"))
```

4. What problems do you encounter when dealing with the dataset?
5. How to deal with missing values?
6. Explore the relationship between Tondu and other variables including female, DPP, age, income, edu, Taiwanese and Econ\_worse. What methods would you use?
7. How about the votetsai variable (vote for DPP candidate Tsai Ing-wen)?