EPPS6323 Knowledge Mining

Assignment 4

1. Lab06 in R
2. Review ISLR Chapters 4
3. Use the TEDS2016 dataset to run a logit (logistic regression) model using female as sole predictor. Access the data set using the following codes:

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

   Hint: glm.vt=glm(votetsai~female, data=TEDS_2016,family=binomial)

   Are female voters more likely to vote for President Tsai? Why or why not?

4. Add party ID variables (KMT, DPP) and other demographic variables (age, edu, income) to improve the model.

   What do you find? Which group of variables work better in explaining/predicting votetsai?

5. Try adding the following variables:

   Independence – Supporting Taiwan's Independence (vs. Unification with China)
   Econ_worse – Evaluations of economy (Negative)
   Govt_dont_care – Political Efficacy (Government does not care about people)
   Minnan_father – Descendent of local Taiwanese
   Mainland_father – Descendent of mainland China (migrated from mainland circa or after 1949)
   Taiwanese – Self-identified Taiwanese

6. Run the model in Stata (available in EPPS labs)

   Hint: Use this command:

```
use
"https://github.com/datageneration/home/blob/master/DataProgramming/data/TEDS_20
16.dta?raw=true"
logit votetsai Independence Econ_worse Govt_dont_care Minnan_father
Mainland_father Taiwanese KMT DPP age edu female
```

7. Compare the results from R and Stata
8. Use the mrobust command in Stata to determine what the most important variable is. (Warning: do mrobust progressively since big models take some time to run)