

EPPS6323 Knowledge Mining

Assignment 3: Exploratory Data Analysis

1. Lab03 in R
2. Review ISLR Chapters 3
3. Use the TEDS2016 dataset to run a multiple regression model. 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")
```

4. Write a function called regplot to plot a regression line

```
## Creating a function: regplot
## Combine the lm, plot and abline functions to create a regression fit plot
function
regplot=function(x,y){
  fit=lm(y~x)
  plot(x,y)
  abline(fit,col="red")
}
```

5. Run a regplot on the dependent variable using:
 - a. Age
 - b. Education
 - c. Income
6. What is the problem? Why? (hint: how many categories in the DV?)
7. What can be done to improve prediction of the dependent variable?