ECO239 R- Homework 1
[Due: Nov.20 (Friday) 2015 at 16:30]
Start now and don't panic at midnight on 12th!!!

Task1: Downloading R, Loading packages
Task2: Importing data using R Console Window
    Task 2-1: Using housingdataSRT.csv or .xlsx data
    Task 2-2: Using your own data download from TUIK webpage
Task3: Describing Data [Numerically, Graphically] using R Console Window
    Task 3-1: Using housingdataSRT.csv or .xlsx data
    Task 3-2: Using your own data download from TUIK webpage
Task4: Use AER / ECDAT data packages

The objective of this homework is to be familiar with R program, using R Console and conduct the basic data analysis which is discussed during the class (Ch.1 and 2 contents).

Task 1: Downloading R, Loading packages

    a.   Download R program to your compute from http://www.r-project.org/. From the left-hand-side menu, click CLAN under "download, packages". Select one of the mirror sites, follow the instruction guidelines.

    b.   Once you download R program, open R. RGui window will open. On the top of the menu [File Edit View Misc Packages Windows Help], click Packages => Install packages => Select one location as CRAN mirror => Scroll down the list of packages and find " XLSX ". (This library is used for downloading data files with .xlsx extension.)

    c.   On the top of the RGui menu, click Packages => load packages => select XLSX.

        Note: If you have encountered an error message "JAVA_HOME cannot be determined from the Registry", you have to download Java program from the following site:
        http://www.java.com/en/download/manual.jsp
        Restart after installing, then re-open R and try to load xlsx library. It should work.

Task2: Importing data using R Console Window
Task 2-1: Work with housingdataSRT.csv and .xls

Download housingdataSRT.csv and/or .xlsx from course webpage. Import the data file as instructed in the in-class R demonstration conducted on Oct. 22, 2015 [R_Demo_2015.R], or refer to the code below.

 # Downloading data in .csv format

```
housingdata <-read.table("C:/yourfilelocation.csv", header=TRUE, sep=",")
ls(housingdata) # list the names of variables
```

 # Downloading data in .xlsx format
library(xlsx)
housingdata2 <-read.xlsx("C:/yourfile location.xlsx",1)

(The last entry "1" is referring to the first worksheet of the excel file. If you want to load the data on the second worksheet of the file, simply specify it as "2".)

You have to type a command directly to the window in order to import the data file. If you use .xlsx extension, make sure to download library named "XLSX" and use it as instructed in the demo.

Task 2-2: Work with your own data downloaded from [www.tuik.gov.tr]

In order to complete this task, you are required to find your own data set from [www.tuik.gov.tr]. Download and save the data set of your interest on your PC as .csv or .xlsx extentions. Import "[your data].csv" or "[your data].xlsx" by using R Console Window.

For Task 2 and Task 3, report both your codes and results.
Task3: Describing Data [Numerically, Graphically] using R Console Window
Task 3-1: Using housingdataSRT.csv or .xlsx data
Task 3-2: Using your own data set.

For both data set, complete the following numerical data analysis.
*You have to first define the variables to complete the following task.

example: #Define each variable. If the variables on the data file is organized as follows:
ID   COUNTY   DPRICE   LOTACR   BATHN   GRGSQF   AGE

COUNTY<-housingdata[,2]
DPRICE<-housingdata[,3]
LOTACR<-housingdata[,4]

a. Find the following summary statistics for three variables in your data set. Comment on the distributional shape (symmetric, right-/left-skewed, spread etc.).
- min
- median
- mean
- max
- variance
- standard deviation
- IQR

b. Select two pairs of variables which you consider to have high correlations. Then find correlation coefficient for each pair. (You need to calculate two correlation coefficient for each pair of variables.)
e.g. COR(AGE, PRICE), COR(LOTACR, GRGSQ). Comment on your findings.

c. Create Histogram for two numerical variables of your choice. Comment on the distributional shape.

d. Pick *two variables <u>which are highly correlated</u>* and create a scatter plot. Comment on your findings.

Task 4: Use AER / ECDAT data packages

Step1 : On GRui window, click R Console window. Tabs including "Packages" appears on RGui window. Click on Packages => Set Cran mirror (select one location), again click on "Packages" tab => Install packages => Select AER and ECDAT packages and wait till download finishes.
The contents of these data packages are found in AER.pdf and ECDAT.pdf on the course web.

Step2 : Select one data file you like. Go to R commander window. Data => Data in packages => Read dataset from an attached package.. => Double-click ECDAT or AER which contains the dataset of your choice => Select Data set. Confirm that the selected data set is now active (Data set: the name of the file on R commander)

Step3 : Repeat Task 2 and 3.
NOTES:
- Just staple your report. DO NOT put it into a plastic cover. I just don't want to do "taking it out from the cover, putting back into the cover" action 120 times.
- Attach your data for Task 2 and 4.
- Copy-Paste codes and outputs to Word for your report.
- Please try to make the size of graphs small (but visible) to save some paper.
- Printing with colors is NOT necessary.
- Come to my office if you need any assistance.
- **Late homework is NOT accepted. (Even if you are late just for 5 minutes, it won't be accepted.)**
- **You should work all by yourself. If you copy someone's work, even partially, you will get 0.** It's better to do whatever you can and submit your own work than copying somebody's perfect homework since for the latter case, you will get 0 for sure.