R Notebook

1. Download a dataset from the web. You may use any source, but specify the source in your code. Also ensure that the data has a mix of quantitative and qualitative (categorical) variables.
2. Import the dataset into R

```
library(readr)
insurance <- read_csv("C:/Trading detail/STUDY/01_MSBA/02 MSBA ML/Dataset/Machine-Learning-with-R-datasets-master/insurance.csv")
```

```
## Parsed with column specification:
## cols(
##   age = col_double(),
##   sex = col_character(),
##   bmi = col_double(),
##   children = col_double(),
##   smoker = col_character(),
##   region = col_character(),
##   charges = col_double()
## )
```

```
View(insurance)
```

3. Print out descriptive statistics for a selection of quantitative and categorical variables.

```
summary(insurance)
```

```
##       age            sex                 bmi           children
##  Min.   :18.00   Length:1338        Min.   :15.96   Min.   :0.000
##  1st Qu.:27.00   Class :character   1st Qu.:26.30   1st Qu.:0.000
##  Median :39.00   Mode  :character   Median :30.40   Median :1.000
##  Mean   :39.21                      Mean   :30.66   Mean   :1.095
##  3rd Qu.:51.00                      3rd Qu.:34.69   3rd Qu.:2.000
##  Max.   :64.00                      Max.   :53.13   Max.   :5.000
##     smoker             region             charges
##  Length:1338        Length:1338        Min.   : 1122
##  Class :character   Class :character   1st Qu.: 4740
##  Mode  :character   Mode  :character   Median : 9382
##                                        Mean   :13270
##                                        3rd Qu.:16640
##                                        Max.   :63770
```

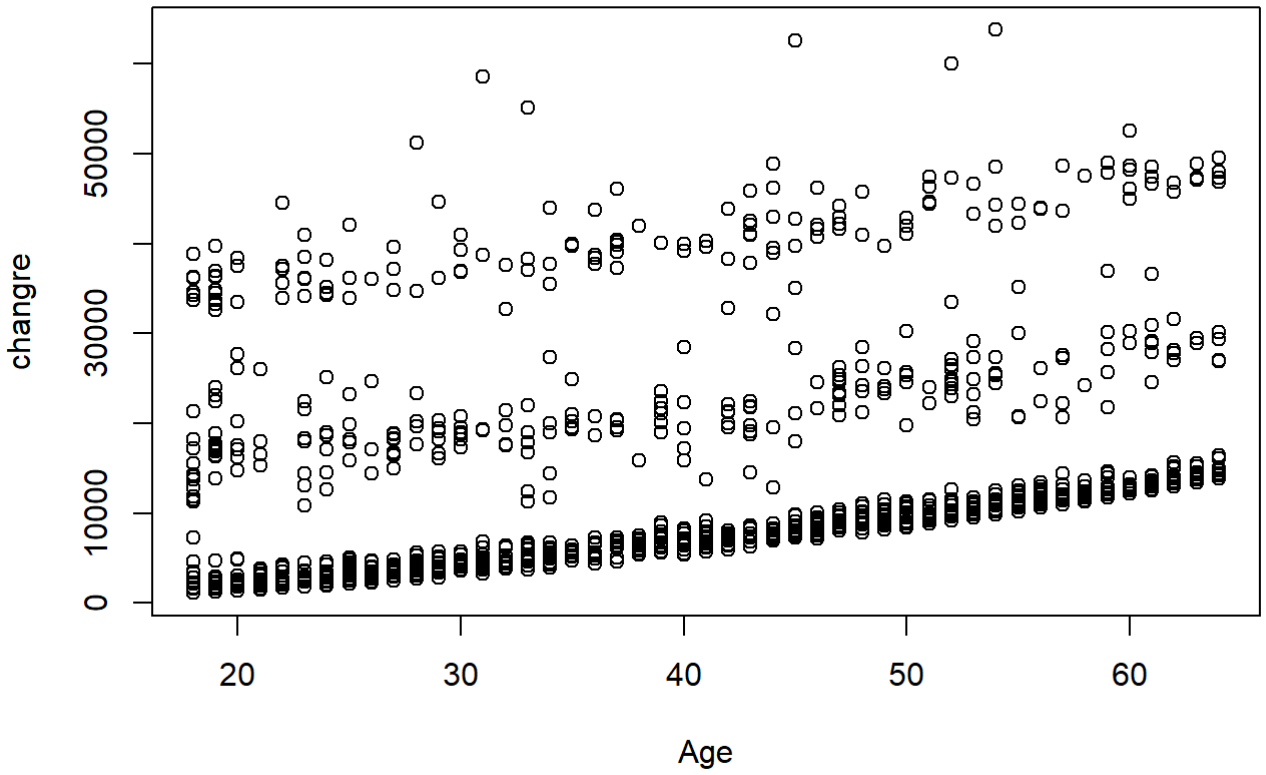
4. Transform at least one variable. It doesn't matter what the transformation is.

```
loginsurance = log(insurance$age)
summary(loginsurance)
```

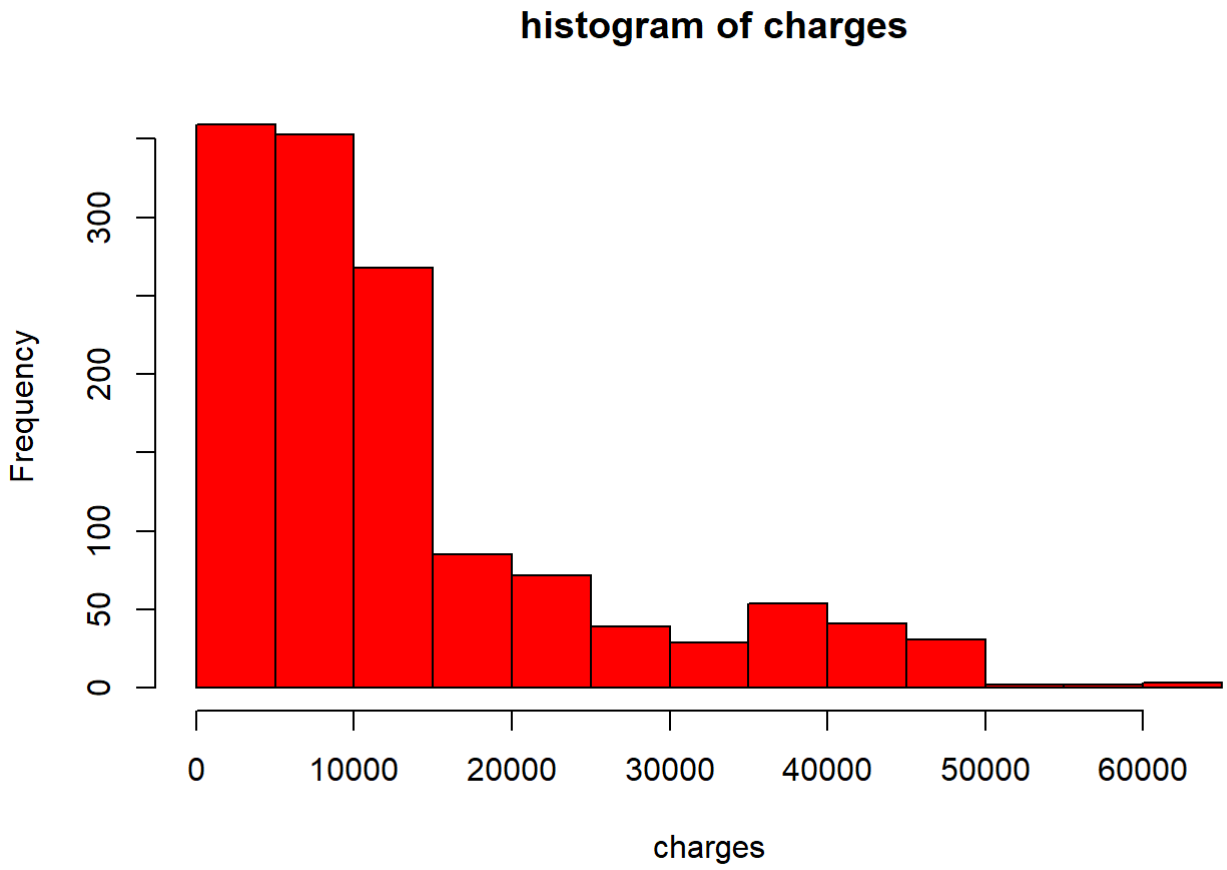
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.890   3.296   3.664   3.597   3.932   4.159
```

5.Plot at least one quantitative variable, and one scatterplot

```
library(plotrix)
plot ( insurance$age,insurance$charges,xlab ="Age",ylab="changre")
```




```
ht =hist(insurance$charges,main =" histogram of charges", xlab ="charges",col = 'red')
```


histogram of charges
Frequency
0
50
100
200
300
0
10000
20000
30000
40000
50000
60000
charges