R_Project

By: Rasha Abu Rkab

ID: 220210069

Submitted to Dr: Samah Hammouda

Read Data

Code :

```
data = read.csv("C:\\Users\\pc\\Downloads\\dataset_car_seats.csv")
data = data.frame(data)
head(data,10)
```

```
   X Sales CompPrice Income Advertising Population Price ShelveLoc Age
## 1  0  9.50       138     73          11        276   120       Bad  42
## 2  1 11.22       111     48          16        260    83      Good  65
## 3  2 10.06       113     35          10        269    80    Medium  59
## 4  3  7.40       117    100           4        466    97    Medium  55
## 5  4  4.15       141     64           3        340   128       Bad  38
## 6  5 10.81       124    113          13        501    72       Bad  78
## 7  6  6.63       115    105           0         45   108    Medium  71
## 8  7 11.85       136     81          15        425   120      Good  67
## 9  8  6.54       132    110           0        108   124    Medium  76
## 10 9  4.69       132    113           0        131   124    Medium  76
```

Data Info

1) shape of data

Code :

```
dim(data) : 400  12
```

2) Columns name

Code : `str(data)`

```
'data.frame':    400 obs. of  12 variables:
$ X           : int  0 1 2 3 4 5 6 7 8 9 ...
$ Sales       : num  9.5 11.22 10.06 7.4 4.15 ...
$ CompPrice   : int  138 111 113 117 141 124 115 136 132 132 ...
$ Income      : int  73 48 35 100 64 113 105 81 110 113 ...
```

```
$ Advertising: int  11 16 10 4 3 13 0 15 0 0 ...
$ Population : int  276 260 269 466 340 501 45 425 108 131 ...
$ Price      : int  120 83 80 97 128 72 108 120 124 124 ...
$ ShelveLoc  : chr  "Bad" "Good" "Medium" "Medium" ...
$ Age        : int  42 65 59 55 38 78 71 67 76 76 ...
$ Education  : int  17 10 12 14 13 16 15 10 10 17 ...
$ Urban      : chr  "Yes" "Yes" "Yes" "Yes" ...
$ US         : chr  "Yes" "Yes" "Yes" "Yes" ...
```

3) Summary Of Data

```
Code : summary(data)
```

```
##        X              Sales          CompPrice       Income
##  Min.   :  0.00   Min.   : 0.000   Min.   : 77   Min.   : 21.00
##  1st Qu.: 99.75   1st Qu.: 5.390   1st Qu.:115   1st Qu.: 42.75
##  Median :199.50   Median : 7.490   Median :125   Median : 69.00
##  Mean   :199.50   Mean   : 7.496   Mean   :125   Mean   : 68.66
##  3rd Qu.:299.25   3rd Qu.: 9.320   3rd Qu.:135   3rd Qu.: 91.00
##  Max.   :399.00   Max.   :16.270   Max.   :175   Max.   :120.00

##   Advertising       Population        Price        ShelveLoc
##  Min.   : 0.000   Min.   : 10.0   Min.   : 24.0   Length:400
##  1st Qu.: 0.000   1st Qu.:139.0   1st Qu.:100.0   Class :character
##  Median : 5.000   Median :272.0   Median :117.0   Mode  :character
##  Mean   : 6.635   Mean   :264.8   Mean   :115.8
##  3rd Qu.:12.000   3rd Qu.:398.5   3rd Qu.:131.0
##  Max.   :29.000   Max.   :509.0   Max.   :191.0

##       Age          Education       Urban               US
##  Min.   :25.00   Min.   :10.0   Length:400         Length:400
##  1st Qu.:39.75   1st Qu.:12.0   Class :character   Class :character
##  Median :54.50   Median :14.0   Mode  :character   Mode  :character
##  Mean   :53.32   Mean   :13.9
##  3rd Qu.:66.00   3rd Qu.:16.0
##  Max.   :80.00   Max.   :18.0
```

Create a sub-data

Code :

```
subdata= data[,c(2,3,4,9,10,11,12)]
head(subdata,10)
```

```
##    Sales CompPrice Income Age Education Urban  US
## 1   9.50       138     73  42        17   Yes Yes
## 2  11.22       111     48  65        10   Yes Yes
```

```
## 3  10.06       113     35  59         12   Yes Yes
## 4   7.40       117    100  55         14   Yes Yes
## 5   4.15       141     64  38         13   Yes  No
## 6  10.81       124    113  78         16    No Yes
## 7   6.63       115    105  71         15   Yes  No
## 8  11.85       136     81  67         10   Yes Yes
## 9   6.54       132    110  76         10    No  No
## 10  4.69       132    113  76         17    No Yes
```

The Columns Name

Code : `str(subdata)`

```
'data.frame':    400 obs. of  7 variables:
$ Sales    : num  9.5 11.22 10.06 7.4 4.15 ...
$ CompPrice: int  138 111 113 117 141 124 115 136 132 132 ...
$ Income   : int  73 48 35 100 64 113 105 81 110 113 ...
$ Age      : int  42 65 59 55 38 78 71 67 76 76 ...
$ Education: int  17 10 12 14 13 16 15 10 10 17 ...
$ Urban    : chr  "Yes" "Yes" "Yes" "Yes" ...
$ US       : chr  "Yes" "Yes" "Yes" "Yes" ...
```

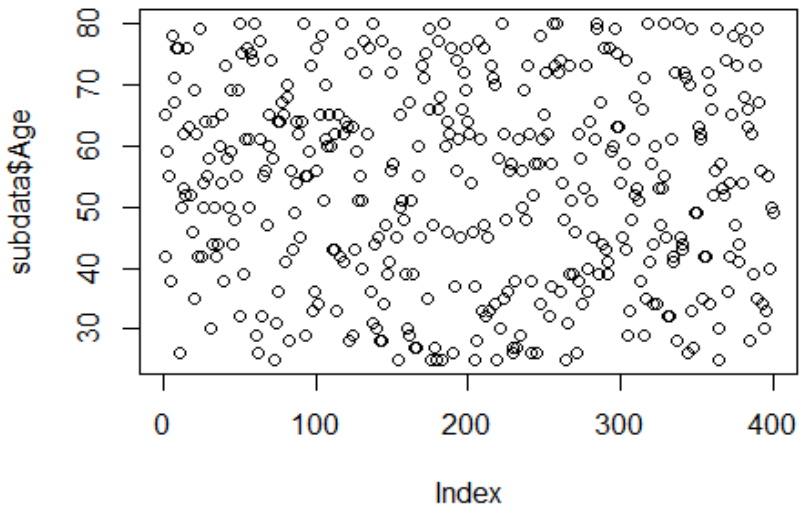
Convert string Data To Numerical

Code :

```
subdata$Urban = as.factor(subdata$Urban)
subdata$US = as.factor(subdata$US)
```

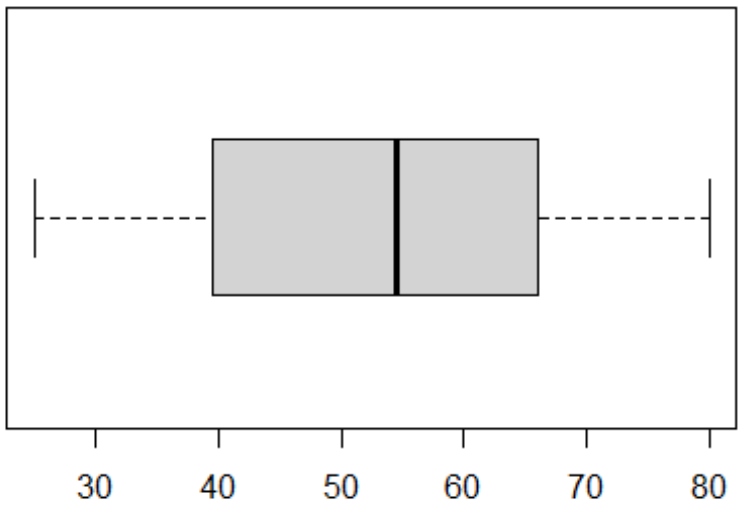
Plot Sub-Data

Code : `plot(subdata$Age)`



Box Plot Of Age

```
Code :  boxplot(subdata$Age,horizontal = TRUE)
```



calculate mean and variance and stander deviation for Age

```
Code:

mean(subdata$Age): 53.3225

var(subdata$Age): 262.4496

sd(subdata$Age): 16.2003

sum(subdata$Age): 21329

sum((subdata$Age - mean(subdata$Age))^2): 104717.4
```

Save Result in A Table

```
Code :

restable = data.frame(matrix(NA,ncol=2,nrow = 5))
restable[1,1] = 'Ages'
restable[2,1] = 'Mean'
restable[3,1] = 'Median'
restable[4,1] = 'SD'
restable[5,1] = 'Min,Max'
```

```
restable[2,2] = round(mean(subdata$Age),2)
restable[3,2] = median(subdata$Age)
restable[4,2] = round(sd(subdata$Age),2)
restable[5,2] = paste(min(subdata$Age),max(subdata$Age),sep = ',')

print(restable)
```

Ages	<NA>
Mean	53.32
Median	**54.5**
SD	**16.2**
Min,Max	**25,80**

Calculate The Range

Code :

```
RR= max(subdata$Age) - min(subdata$Age)
RR: 55
```

Calculate Z_Score

Print The First 10 Result

Code :

```
x= subdata$Age
xbar = mean(x)
sdd = sd(x)
z_score = (x - xbar)/sdd
head(z_score,10)
```

```
-0.6989069  0.7208201  0.3504565  0.1035475 -0.9458160  1.5232746
1.0911837  0.8442747  1.3998200  1.3998200
```

Calculate Percentiles

Create A Function and Print Data As A Table

Code :

```
mystat = function(x,nom){
  resTable = data.frame(matrix(NA,ncol=2,nrow = 5))
  resTable[1,1] = nom
  resTable[2,1] = 'Mean ± SD'
  resTable[3,1] = 'MD(Q1-Q3)'
  resTable[4,1] = 'Min,Max'

  resTable[2,2]= paste(mean(x),'±',sd(x),sep = '')
```

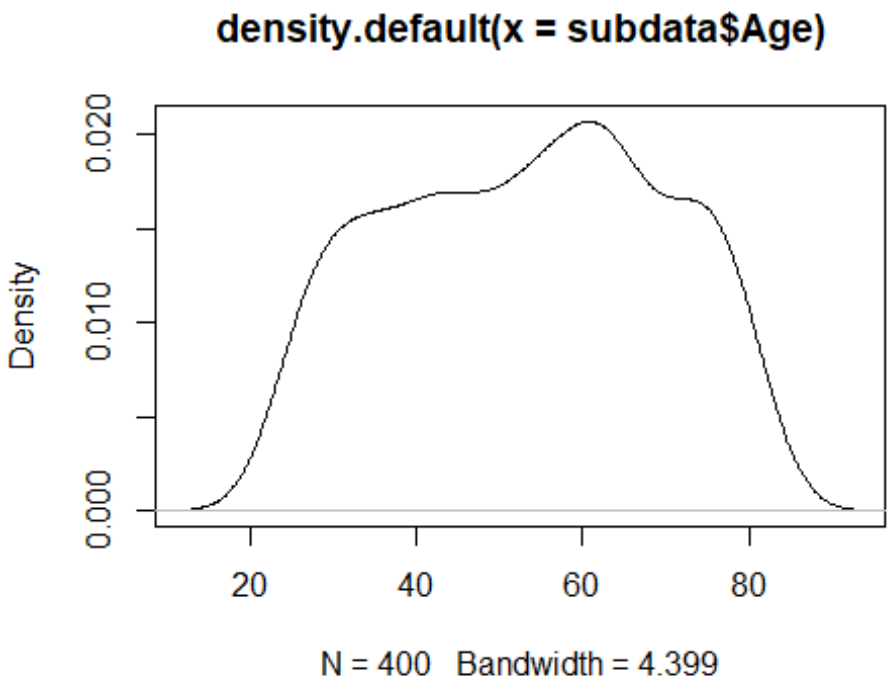
```
  resTable[3,2]= paste(median(x),'(',quantile(x,0.25),'-',quantile(x,0.75),')
',sep = '')
  resTable[4,2]= paste(min(x),max(x),sep = ',')

  return(resTable)
   }

 mystat(x,'Age')
```

Age	<NA>
Mean ± SD	53.3225±16.2002968427892
MD(Q1-Q3)	54.5(39.75-66)
Min\Max	25\80

skewness

Code :

```
library(moments)
skewness(subdata$Age): -0.076892

plot(density(subdata$Age))
```



T-Test

Code :

```
library(BSDA)
```

```
t.test(subdata$Age):
```

```
data:  subdata$Age
t = 65.829, df = 399, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
51.73007 54.91493
sample estimates:
mean of x: 53.3225
```

Code :

```
t.test(subdata$Age,subdata$Sales):
```

```
data:  subdata$Age and subdata$Sales
t = 55.734, df = 423.23, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0 95
```

```
percent confidence interval: 44.21001 47.44234
sample estimates:
mean of x mean of y: 53.322500  7.496325
```

Code :

```
wilcox.test(subdata$Age , subdata$Sales):
```

```
data:  subdata$Age and subdata$Sales
W = 160000, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Calculate The Correlation And Regression

Find The Correlation For All Numerical Columns And Company Price

Code :

```
ResTable = data.frame(matrix(NA,ncol=2,nrow = 5))
ResTable[1,1]= 'Columns'
ResTable[2,1] = "Age"
ResTable[3,1] = 'Sales'
ResTable[4,1] = 'Income'
ResTable[5,1] = 'Education'

ResTable[1,2]= 'Resule'
ResTable[2,2]= round(cor(subdata$CompPrice, subdata$Age, method = "pearson",
use = "complete.obs"),2)
```

```
ResTable[3,2]= round(cor(subdata$CompPrice, subdata$Sales, method = "pearson"
, use = "complete.obs"),2)
ResTable[4,2]= round(cor(subdata$CompPrice, subdata$Income, method = "pearson
", use = "complete.obs"),2)
ResTable[5,2]= round(cor(subdata$CompPrice, subdata$Education, method = "pear
son", use = "complete.obs"),2)

ResTable
```

Columns	Result
Age	-0.1
Sales	0.06
Income	-0.08
Education	0.03

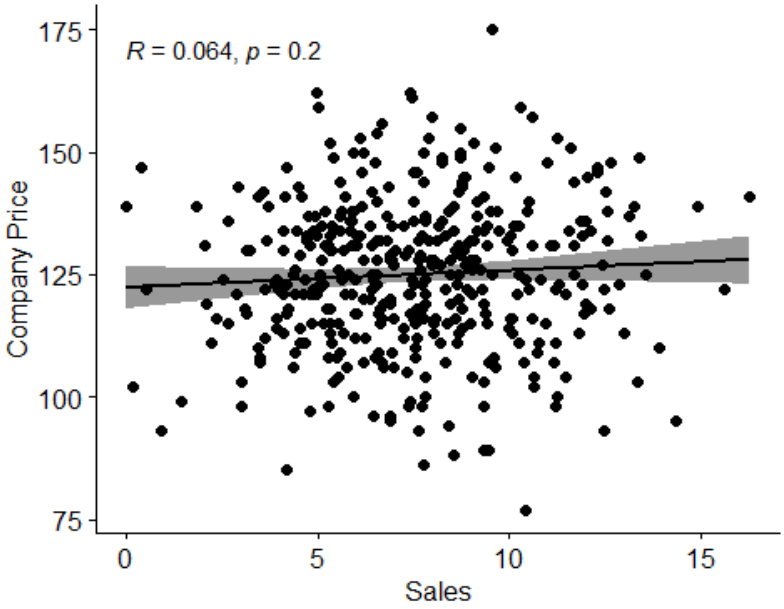
Correlation and Regression Analysis

Scatter plot with smooth fit curve With Positive Correlation

1) With Sales

Code :

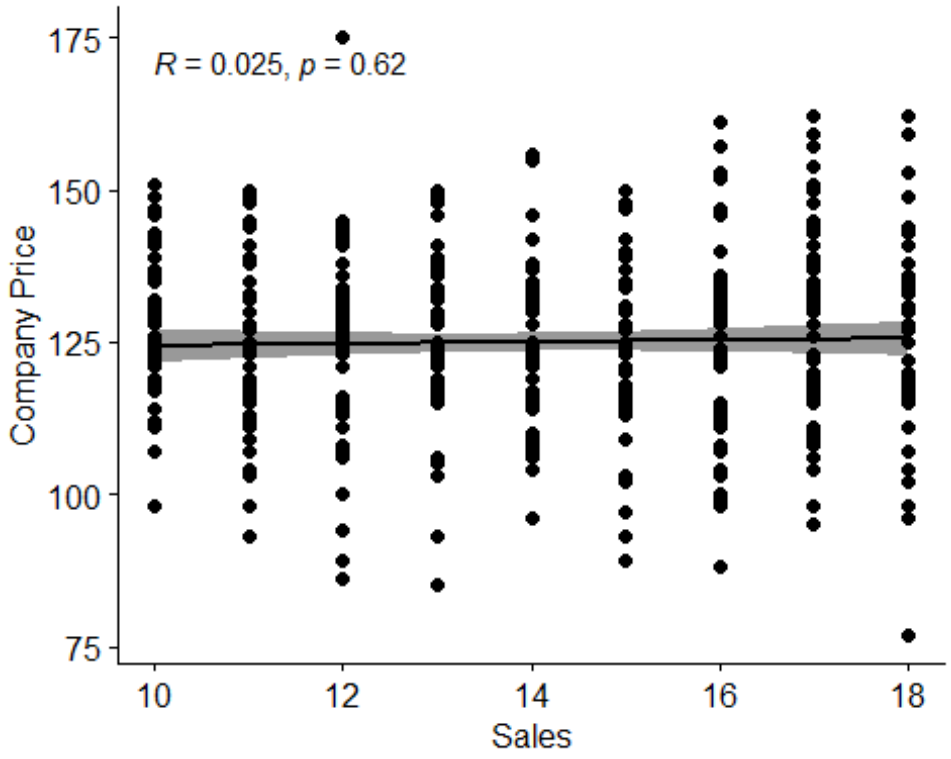
```
library("ggpubr")

ggscatter(subdata, x = "Sales", y = "CompPrice",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Sales", ylab = "Company Price")
```



2) With Education

Code :

```
library("ggpubr")
ggscatter(subdata, x = "Education", y = "CompPrice",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Sales", ylab = "Company Price")
```


R = 0.025, p = 0.62
Company Price
175
150
125
100
75
10
12
14
16
18
Sales