Question 01

(a) Upload the Advertising dataset and explore it.

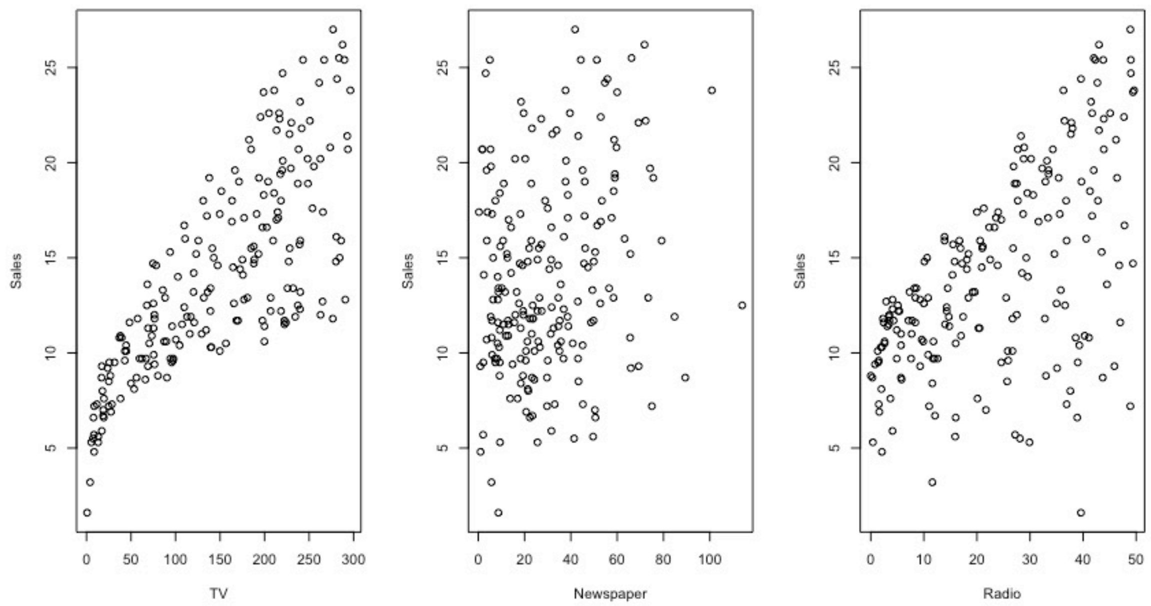
```
advertising <-read.csv('Advertising.csv')
 attach(advertising)
 View(advertising)
names(advertising)
dim(advertising)
```

(b) Construct scatter plots to visualize the relationship between following variables:

• Sales and TV • Sales and Radio • Sales and Newspaper

```
par(mfrow=c(1,3))
plot(Sales~TV,ylab='Sales',xlab='TV')
plot(Sales~Newspaper,ylab='Sales',xlab='Newspaper')
plot(Sales~Radio,ylab='Sales',xlab='Radio')
```

Output:



(c) Find the Correlation Coefficient to measure the strength of the linear relationship of Sales and TV

cor(Sales,TV,method=c("pearson","kendall","spearman"))

<u>Output:</u>

```
[1] 0.7822244
```

(d) Find the least square estimates of the linear model of Sales in terms of TV and give the resulting model

model =(lm(Sales~TV))
summary(model)

<u>Output:</u>

```
Call:
lm(formula = Sales ~ TV)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.3860 -1.9545 -0.1913  2.0671  7.2124 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.032594   0.457843   15.36   <2e-16 ***
TV          0.047537   0.002691   17.67   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.259 on 198 degrees of freedom
Multiple R-squared:  0.6119,    Adjusted R-squared:  0.6099 
F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

(e) Assess the accuracy of the parameter estimates
The formula is Y = a +Bx, so we need to assess a and B for the parameter estimates and we get the result by sumarry (model) ,and the intercept and slope is lower than Standard error

Residual standard error: 3.259 on 198 degrees of freedom

Standard Error for Intercept: 0.457843 of 7.032594

Standard Error for Slope: 0.002691 of 0.047537

(f) Test the significance of the slope of the linear model

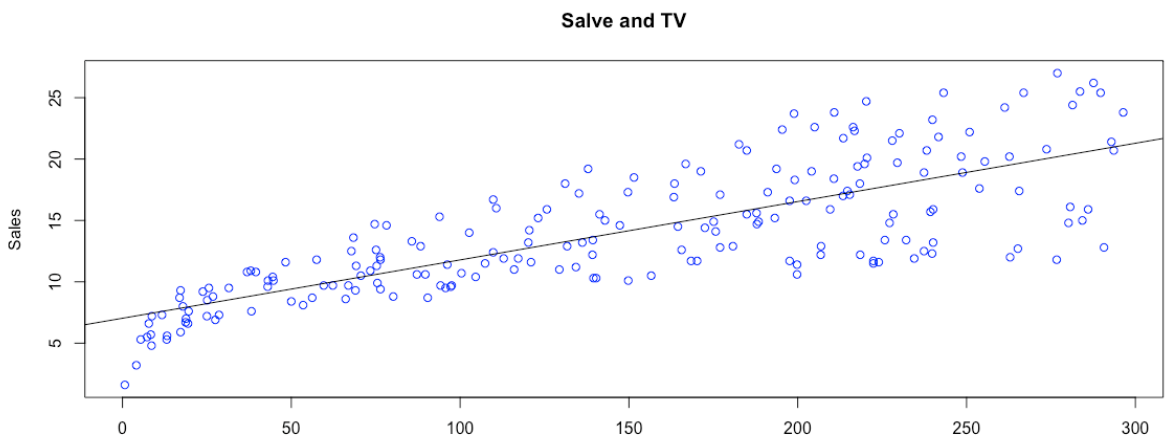
To test the significance of the slope of the linear model, and consider the hypotheses testing:

Assume the B1 = 0, the there is not linear relationship between X and Y , but the p-value = 2.2e-16 so < 0.05 so slope of linear model is significance and B1 != 0.

(g) Plot the straight line within the scatter plot and comment

```
plot(TV,Sales,col="blue",main="Salve and TV",ylab='Sales', xlab='TV')
abline(lm(Sales~TV))
```

<u>Output:</u>



(h) Assess the overall accuracy of the model
anova(model)

<u>Output:</u>

```
Analysis of Variance Table

Response: Sales
           Df Sum Sq Mean Sq F value    Pr(>F)
TV          1 3314.6  3314.6  312.14 < 2.2e-16 ***
Residuals 198 2102.5    10.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(i) Use the model to make predictions
result <-predict(model,list(TV=c(100,101,11)))
Output:

```
        1         2         3
11.786258 11.833794  7.555497
```

Question 02

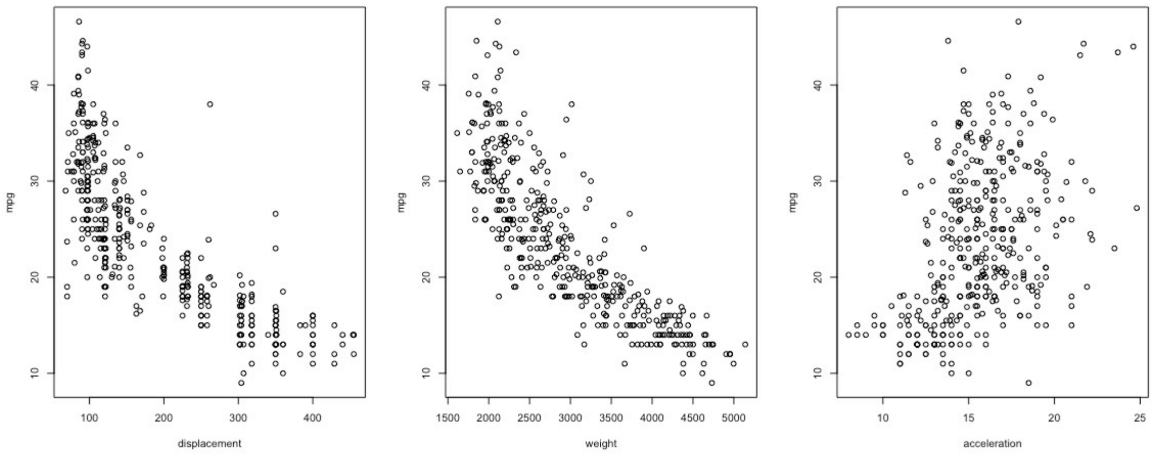
(a) Upload the Auto Dataset and explore it.

```
Auto <- read.csv('Auto.csv')
attach(Auto)
View(Auto)
```

(b) Construct scatter plots to visualize the relationship between following variables:
mpg and displacement
mpg and weight
mpg and acceleration

```
par(mfrow=c(1,3))
plot(mpg~displacement,ylab='mpg',xlab='displacement')
plot(mpg~weight,ylab='mpg',xlab='weight')
plot(mpg~acceleration,ylab='mpg',xlab='acceleration')
```

Output:



(D) Find the correlation coefficient to measure the strength of the linear relationship of Sales and acceleration
cor(mpg,acceleration,method=c("pearson","kendall","spearman"))
Output:
[1] 0.4222974

(d)Find the least square estimates
model_auto=(lm(mpg~acceleration))
summary(model_auto)

Output:

```
Call:
lm(formula = mpg ~ acceleration)

Residuals:
    Min      1Q  Median      3Q     Max
-18.054  -5.646  -1.238   4.753  23.194

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.8218     2.0502   2.352   0.0192 *
acceleration   1.2018     0.1298   9.259   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.103 on 395 degrees of freedom
Multiple R-squared:  0.1783,    Adjusted R-squared:  0.1763
F-statistic: 85.73 on 1 and 395 DF,  p-value: < 2.2e-16
```

(e)Assess the accuracy of the parameter estimates
The formula is Y = a +Bx, so we need to assess a and B for the parameter estimates and we get the result by sumarry (model_auto) ,and the intercept and slope is lower than Standard error

Residual standard error: 7.103 on 395 degrees of freedom

Standard Error for Intercept: 2.0502 of 4.8218

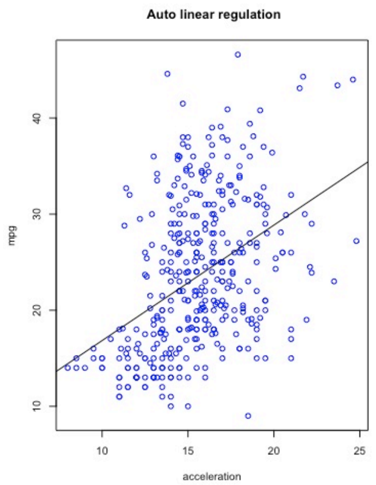
Standard Error for Slope: 0.1298 of 1.2018

(f)Test the significance of the slope of the linear model

To test the significance of the slope of the linear model, and consider the hypotheses testing:

Assume the B1 = 0, the there is not linear relationship between X and Y , but the p-value = 2.2e-16 so < 0.05 so slope of linear model is significance and B1 != 0.

(g)Plot the straight line within the scatter plot and comment
plot(acceleration,mpg,col="blue",main=" Auto linear regulation",ylab='mpg', xlab='acceleration')
abline(lm(mpg~acceleration))



(h)Assess the overall accuracy of the model
anova(model_auto)
<u>Output:</u>

```
Analysis of Variance Table

Response: mpg
              Df Sum Sq Mean Sq F value    Pr(>F)
acceleration   1   4325  4325.0  85.731 < 2.2e-16 ***
Residuals    395  19927    50.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(i) Use the model to make predictions

result <-predict(model_auto,data.frame(acceleration=100))

Output:

```
> result
       1
124.9972
```