Meher Shrishti Nigam
20BRS1193

EDA LAB – 6
10 / 2 / 23

 # Meher Shrishti Nigam

 # 20BRS1193

 # EDA Lab 6

```
options(prompt="MEHERSHRISHTI>", continue =" ")
```

 # options(prompt=">", continue =" ")

 # EDA-LAB-EXPERIMENT-6 (Date-10/2/2023)

```
library(ISLR)
```

 # Q1

```
data <- Weekly
```

```
data
```

```
MEHERSHRISHTI># Q1
MEHERSHRISHTI>data <- Weekly
MEHERSHRISHTI>data
    Year   Lag1   Lag2   Lag3   Lag4   Lag5    Volume  To
day Direction
1   1990  0.816  1.572 -3.936 -0.229 -3.484 0.1549760 -0.
270      Down
2   1990 -0.270  0.816  1.572 -3.936 -0.229 0.1485740 -2.
576      Down
3   1990 -2.576 -0.270  0.816  1.572 -3.936 0.1598375  3.
514        Up
4   1990  3.514 -2.576 -0.270  0.816  1.572 0.1616300  0.
712        Up
```

(a)

Correlation Matrix

```
df <- subset(data, select = -c(Direction) )
```

```
library(corrplot)
```

```
corr<-cor(df)
```

```
corrplot(corr, method="circle")
```

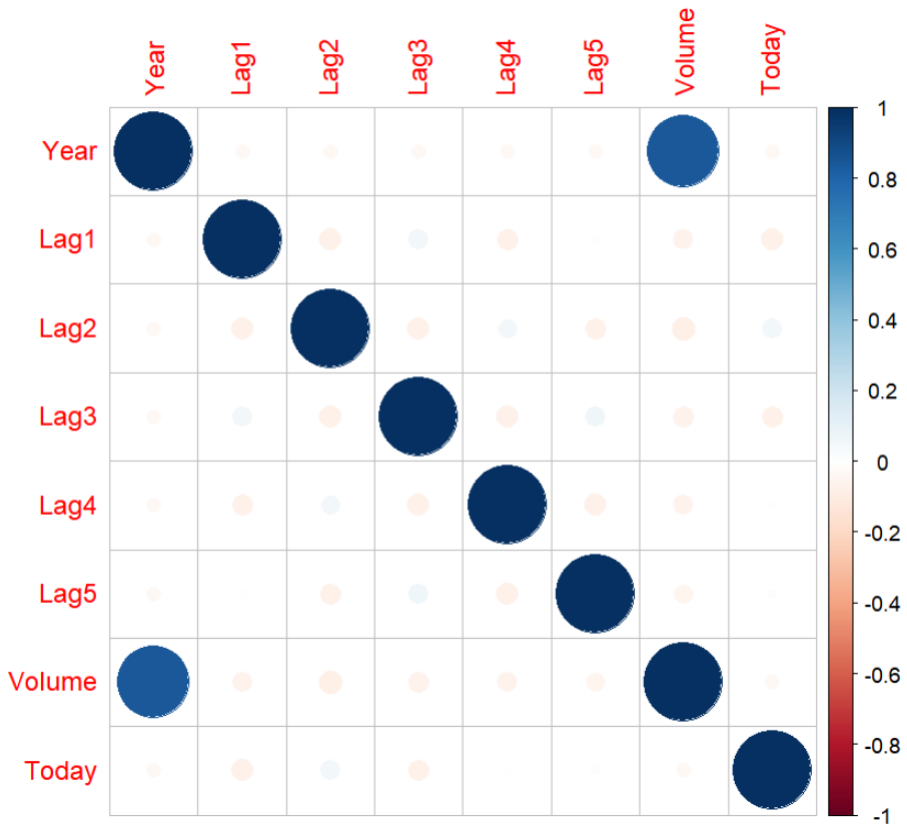
```
corr
```

Output:

```
MEHERSHRISHTI># Correlation Matrix
MEHERSHRISHTI>df <- subset(data, select = -c(Direction) )
MEHERSHRISHTI>library(corrplot)

MEHERSHRISHTI>corr<-cor(df)
MEHERSHRISHTI>corrplot(corr, method="circle")
```

```
MEHERSHRISHTI>corr
              Year          Lag1        Lag2
Year    1.00000000 -0.032289274 -0.03339001
Lag1   -0.03228927  1.000000000 -0.07485305
Lag2   -0.03339001 -0.074853051  1.00000000
Lag3   -0.03000649  0.058635682 -0.07572091
Lag4   -0.03112792 -0.071273876  0.05838153
Lag5   -0.03051910 -0.008183096 -0.07249948
Volume  0.84194162 -0.064951313 -0.08551314
Today  -0.03245989 -0.075031842  0.05916672
              Lag3          Lag4         Lag5
Year   -0.03000649 -0.031127923 -0.030519101
Lag1    0.05863568 -0.071273876 -0.008183096
Lag2   -0.07572091  0.058381535 -0.072499482
Lag3    1.00000000 -0.075395865  0.060657175
Lag4   -0.07539587  1.000000000 -0.075675027
Lag5    0.06065717 -0.075675027  1.000000000
Volume -0.06928771 -0.061074617 -0.058517414
Today  -0.07124364 -0.007825873  0.011012698
             Volume        Today
Year    0.84194162 -0.032459894
Lag1   -0.06495131 -0.075031842
Lag2   -0.08551314  0.059166717
Lag3   -0.06928771 -0.071243639
Lag4   -0.06107462 -0.007825873
Lag5   -0.05851741  0.011012698
Volume  1.00000000 -0.033077783
Today  -0.03307778  1.000000000
MEHERSHRISHTI>
```



Pairplot

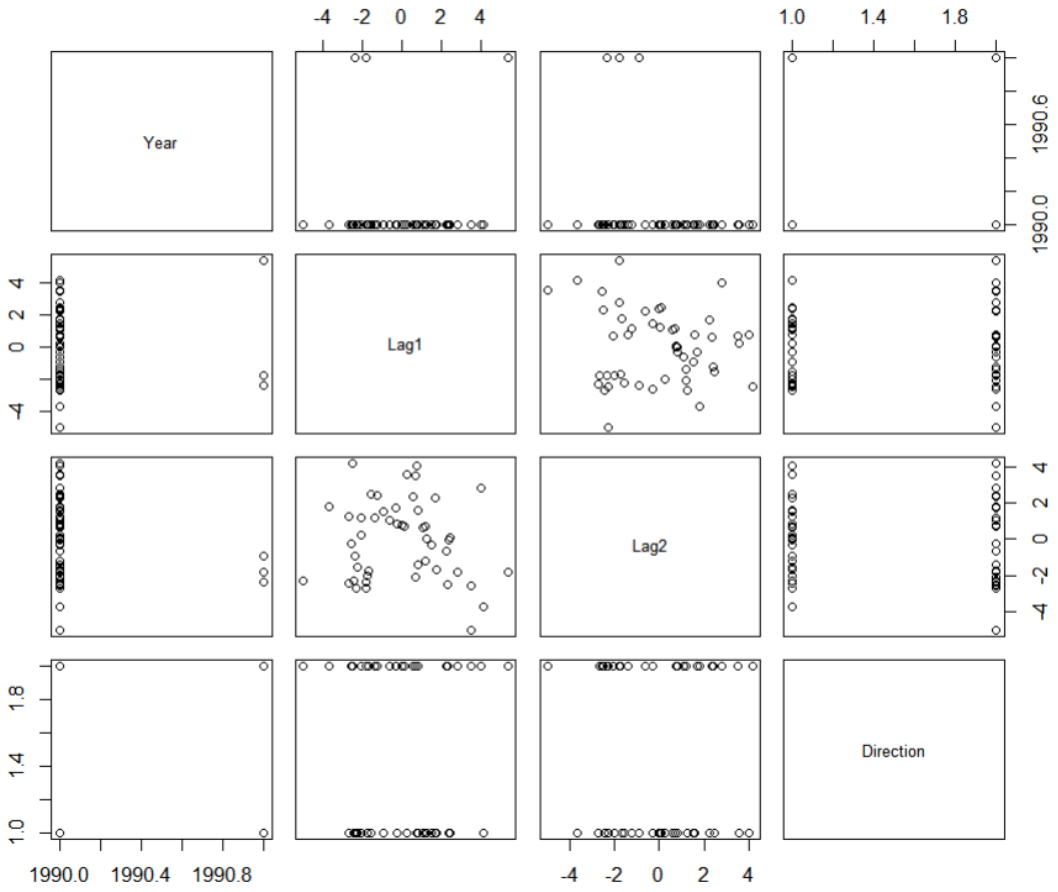
```
df2 <- subset(data, select = c(Year, Lag1, Lag2, Direction))

df2_head <- head(df2, 50)

pairs(df2_head)
```

Output:

```
MEHERSHRISHTI># Pairplot
MEHERSHRISHTI>df2 <- subset(data, select = c(Year, Lag1, Lag2, Directio
n))
MEHERSHRISHTI>df2_head <- head(df2, 50)
MEHERSHRISHTI>pairs(df2_head)
```



(b)

```
logistic_model <- glm(Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 + Volume, data = data, family="binomial")
logistic_model
```

Output:

```
MEHERSHRISHTI>## (b)
MEHERSHRISHTI>logistic_model <- glm(Direction ~ Lag1 + Lag2 + Lag3 +
 Lag4 + Lag5 + Volume,
                    data = data, family="binomial")
MEHERSHRISHTI>logistic_model

Call:  glm(formula = Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 +
    Volume, family = "binomial", data = data)

Coefficients:
(Intercept)         Lag1         Lag2         Lag3
    0.26686     -0.04127      0.05844     -0.01606
       Lag4         Lag5       Volume
   -0.02779     -0.01447     -0.02274

Degrees of Freedom: 1088 Total (i.e. Null);  1082 Residual
Null Deviance:      1496
Residual Deviance: 1486         AIC: 1500
MEHERSHRISHTI>
```

(c)

```
prediction <- predict(logistic_model,
                      data, type = "response")
cf <- rep("Down", length(prediction))
cf[prediction>0.5] = "Up" # Up
cf[prediction<=0.5] = "Down" # Down
table(cf, data$Direction)
```

Output:

```
MEHERSHRISHTI>prediction <- predict(logistic_model,
                           data, type = "response")
MEHERSHRISHTI>cf <- rep("Down", length(prediction))
MEHERSHRISHTI>cf[prediction>0.5] = "Up" # Up
MEHERSHRISHTI>cf[prediction<=0.5] = "Down" # Down
MEHERSHRISHTI>table(cf, data$Direction)

cf      Down  Up
  Down    54  48
  Up     430 557
MEHERSHRISHTI>
```

(d)

```
logistic_model <- glm(Direction ~ Lag2,
                      data = data, family="binomial")
logistic_model


prediction <- predict(logistic_model,
                      data, type = "response")
cf <- rep("Down", length(prediction))
cf[prediction>0.5] = "Up" # Up
cf[prediction<=0.5] = "Down" # Down
table(cf, data$Direction)
```

Output:

```
MEHERSHRISHTI>## (d)
MEHERSHRISHTI>logistic_model <- glm(Direction ~ Lag2,
                        data = data, family="binomial")
MEHERSHRISHTI>logistic_model

Call:  glm(formula = Direction ~ Lag2, family = "binomial", data = d
ata)

Coefficients:
(Intercept)         Lag2
    0.21473      0.06279

Degrees of Freedom: 1088 Total (i.e. Null);  1087 Residual
Null Deviance:      1496
Residual Deviance: 1490         AIC: 1494
MEHERSHRISHTI>prediction <- predict(logistic_model,
                        data, type = "response")
MEHERSHRISHTI>cf <- rep("Down", length(prediction))
MEHERSHRISHTI>cf[prediction>0.5] = "Up" # Up
MEHERSHRISHTI>cf[prediction<=0.5] = "Down" # Down
MEHERSHRISHTI>table(cf, data$Direction)

cf      Down  Up
  Down    33  26
  Up     451 579
MEHERSHRISHTI>
```

 # Q2

(a)

attach(Auto)

summary(Auto)

mpg01 <- rep(0, length(mpg))

mpg01[mpg > median(mpg)] <- 1

Auto = data.frame(Auto, mpg01)

Auto

Output:

```
MEHERSHRISHTI>mpg01 <- rep(0, length(mpg))
MEHERSHRISHTI>mpg01[mpg > median(mpg)] <- 1
MEHERSHRISHTI>Auto = data.frame(Auto, mpg01)
MEHERSHRISHTI>Auto
    mpg cylinders displacement horsepower weight acceleration year
1    18         8        307.0        130   3504         12.0   70
2    15         8        350.0        165   3693         11.5   70
3    18         8        318.0        150   3436         11.0   70
4    16         8        304.0        150   3433         12.0   70
5    17         8        302.0        140   3449         10.5   70
```

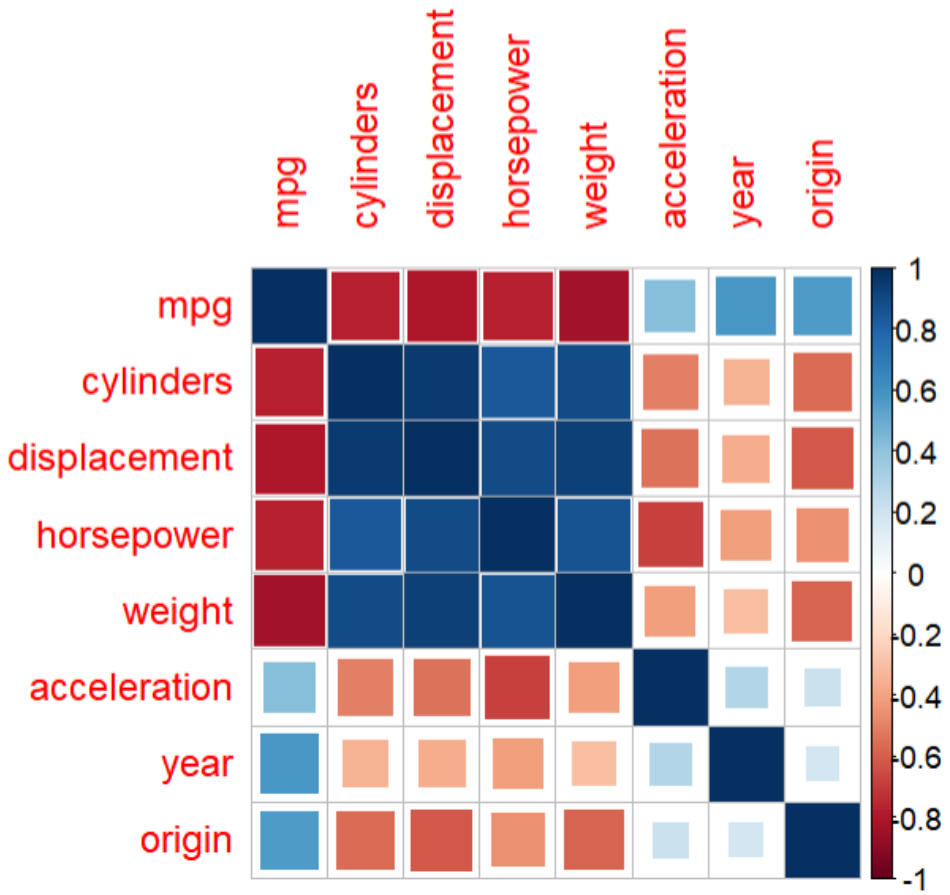
	origin	name	mpg01
1	1	chevrolet chevelle malibu	0
2	1	buick skylark 320	0
3	1	plymouth satellite	0
4	1	amc rebel sst	0
5	1	ford torino	0

(b)

```
corrplot(cor(Auto[,-9]), method="square")
```

Interpretation: The variables that appear to correlate strongly with mpg01 are Cylinders,Displacement, and Weight; these variables appear to correlate negatively with this variable. Also Horsepower and Origin appear to correlate moderately with mpg01.

Output:



(c)

```
train <- (year %% 2 == 0)

train.auto <- Auto[train,]

test.auto <- Auto[-train,]

auto.fit<-glm(mpg01~displacement+horsepower+weight+year+cylinders+origin, data=train.auto,family=binomial)

auto.fit

auto.probs = predict(auto.fit, test.auto, type = "response")

auto.pred = rep(0, length(auto.probs))
```

auto.pred[auto.probs > 0.5] = 1

table(auto.pred, test.auto$mpg01)

Output:

```
MEHERSHRISHTI>## (c)
MEHERSHRISHTI>train <- (year %% 2 == 0)
MEHERSHRISHTI>train.auto <- Auto[train,]
MEHERSHRISHTI>test.auto <- Auto[-train,]
MEHERSHRISHTI>auto.fit<-glm(mpg01~displacement+horsepower+weight+year+cylinders+origin, data=train.auto,family=binomial)
MEHERSHRISHTI>auto.fit

Call:  glm(formula = mpg01 ~ displacement + horsepower + weight + year +
    cylinders + origin, family = binomial, data = train.auto)

Coefficients:
 (Intercept)  displacement    horsepower        weight          year     cylinders        origin
  -15.314371     -0.010665     -0.057653     -0.005228      0.489888     -0.207389      0.471544

Degrees of Freedom: 209 Total (i.e. Null);  203 Residual
Null Deviance:      289.6
Residual Deviance: 57.66        AIC: 71.66
MEHERSHRISHTI>auto.probs = predict(auto.fit, test.auto, type = "response")
MEHERSHRISHTI>auto.pred = rep(0, length(auto.probs))
MEHERSHRISHTI>auto.pred[auto.probs > 0.5] = 1
MEHERSHRISHTI>table(auto.pred, test.auto$mpg01)

auto.pred   0   1
        0 174  12
        1  21 184
MEHERSHRISHTI>
```