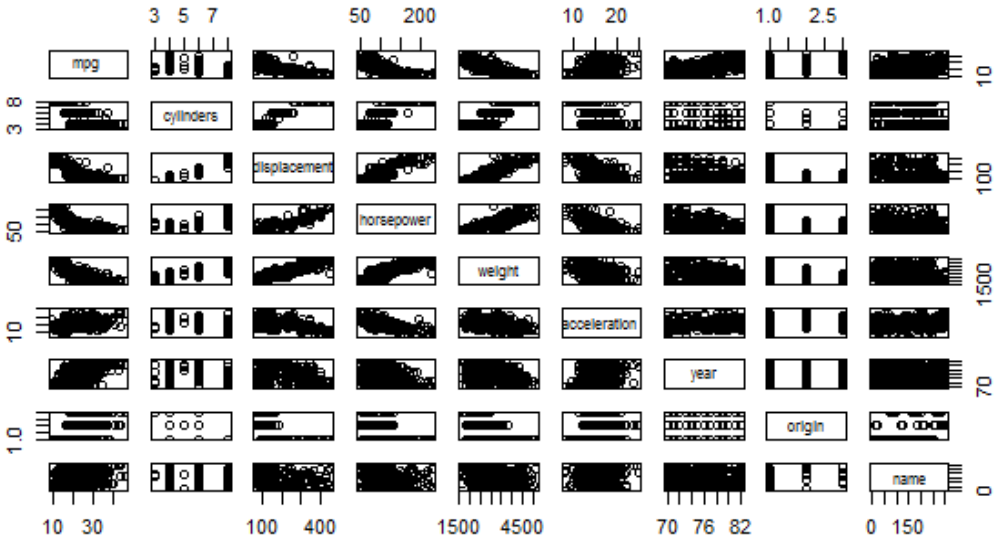
2. Problem 9, Chapter 3

(a) *R code:*

```
> library(ISLR)
> data(Auto)
> pairs(Auto)
```



(b) *R code:*

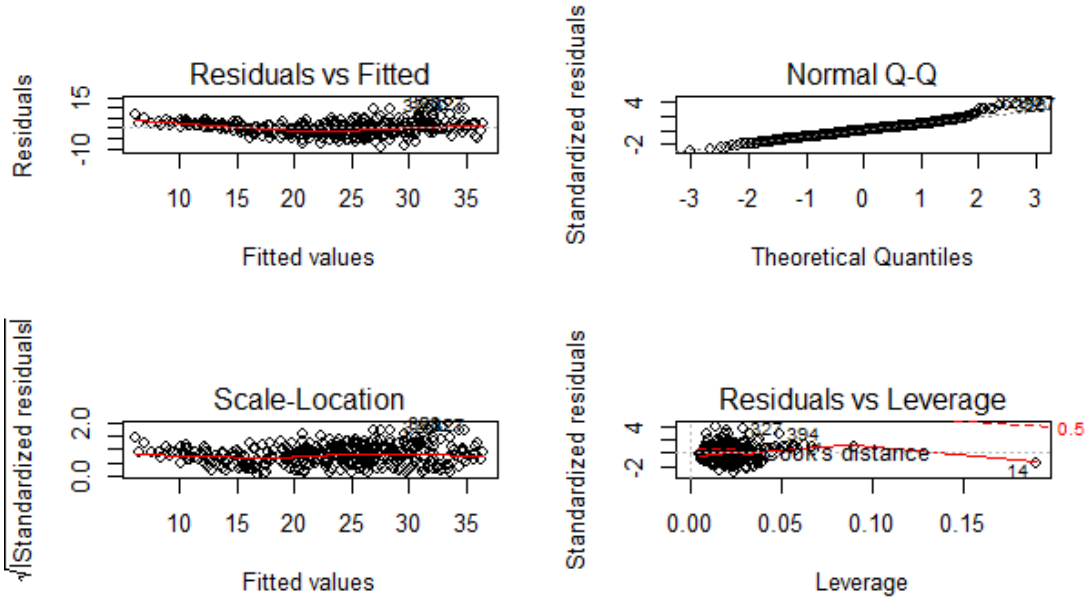
```
> names(Auto)
[1] "mpg"          "cylinders"    "displacement" "horsepower"   "weight"
[6] "acceleration" "year"         "origin"       "name"
> cor(Auto[1:8])
                     mpg  cylinders displacement horsepower     weight acceleration
mpg           1.0000000 -0.7776175    -0.8051269 -0.7784268 -0.8322442    0.4233285
cylinders    -0.7776175  1.0000000     0.9508233  0.8429834  0.8975273   -0.5046834
displacement -0.8051269  0.9508233     1.0000000  0.8972570  0.9329944   -0.5438005
horsepower   -0.7784268  0.8429834     0.8972570  1.0000000  0.8645377   -0.6891955
weight       -0.8322442  0.8975273     0.9329944  0.8645377  1.0000000   -0.4168392
acceleration  0.4233285 -0.5046834    -0.5438005 -0.6891955 -0.4168392    1.0000000
year          0.5805410 -0.3456474    -0.3698552 -0.4163615 -0.3091199    0.2903161
origin        0.5652088 -0.5689316    -0.6145351 -0.4551715 -0.5850054    0.2127458
                   year     origin
mpg           0.5805410  0.5652088
cylinders    -0.3456474 -0.5689316
displacement -0.3698552 -0.6145351
horsepower   -0.4163615 -0.4551715
weight       -0.3091199 -0.5850054
acceleration  0.2903161  0.2127458
year          1.0000000  0.1815277
origin        0.1815277  1.0000000
```

(c) *R code:*

```
> fit2 = lm(mpg ~ . - name, data = Auto)
> summary(fit2)
Call:
lm(formula = mpg ~ . - name, data = Auto)
Residuals:
     Min       1Q   Median       3Q      Max
-9.5903  -2.1565  -0.1169   1.8690  13.0604
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435   4.644294  -3.707  0.00024 ***
cylinders    -0.493376   0.323282  -1.526  0.12780
displacement  0.019896   0.007515   2.647  0.00844 **
horsepower   -0.016951   0.013787  -1.230  0.21963
weight       -0.006474   0.000652  -9.929  < 2e-16 ***
acceleration  0.080576   0.098845   0.815  0.41548
year          0.750773   0.050973  14.729  < 2e-16 ***
origin        1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,  Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

(d)*R code:*

```
> par(mfrow=c(2,2))
> plot(fit2)
```

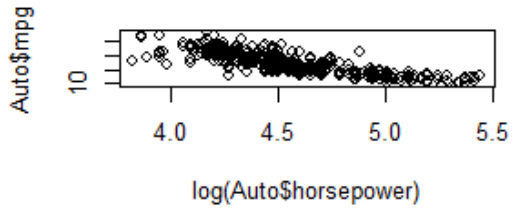


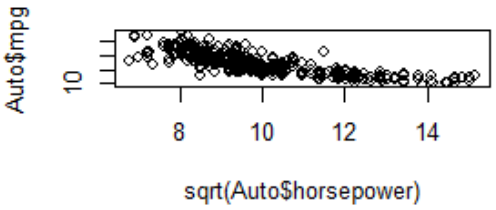
(e) *R code:*

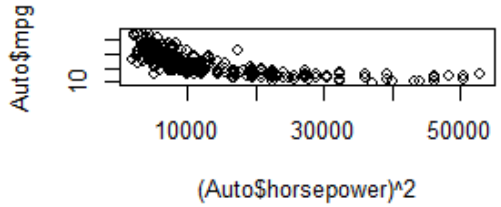
```
> fit3 <- lm(mpg ~ cylinders * displacement + displacement * weight, data = Auto[, 1:8])
> summary(fit3)
Call:
lm(formula = mpg ~ cylinders * displacement + displacement *
    weight, data = Auto[, 1:8])

Residuals:
     Min       1Q   Median       3Q      Max
-13.2934  -2.5184  -0.3476   1.8399  17.7723

Coefficients:
                              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                  5.262e+01   2.237e+00   23.519   < 2e-16 ***
cylinders                    7.606e-01   7.669e-01    0.992     0.322
displacement                -7.351e-02   1.669e-02   -4.403  1.38e-05 ***
weight                      -9.888e-03   1.329e-03   -7.438  6.69e-13 ***
cylinders:displacement      -2.986e-03   3.426e-03   -0.872     0.384
displacement:weight          2.128e-05   5.002e-06    4.254  2.64e-05 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 4.103 on 386 degrees of freedom
Multiple R-squared:  0.7272,  Adjusted R-squared:  0.7237
F-statistic: 205.8 on 5 and 386 DF,  p-value: < 2.2e-16
```

(f) *R code:*

```
> par(mfrow = c(2, 2))
> plot(log(Auto$horsepower), Auto$mpg)
> plot(sqrt(Auto$horsepower), Auto$mpg)
> plot((Auto$horsepower)^2, Auto$mpg)
```







4. Problem 8, Chapter 5

(a)*R code:*

```
> set.seed(1)
> y = rnorm(100)
> x = rnorm(100)
> y = x - 2 * x^2 + rnorm(100)
```

(b)plot(x,y)

(c)

```
> library(boot)
> set.seed(1)
> Data = data.frame(x, y)
```

i.

```
> glm.fit = glm(y ~ x)
> cv.glm(Data, glm.fit)$delta
[1] 5.890979 5.888812
```

ii.

```
> glm.fit = glm(y ~ poly(x, 2))
> cv.glm(Data, glm.fit)$delta
[1] 1.086596 1.086326
```

iii.

```
> glm.fit = glm(y ~ poly(x, 3))
> cv.glm(Data, glm.fit)$delta
[1] 1.102585 1.102227
```

iv.

```
> glm.fit = glm(y ~ poly(x, 4))
> cv.glm(Data, glm.fit)$delta
[1] 1.114772 1.114334
```

(d)

```
> set.seed(100)
> glm.fit = glm(y ~ x)
> cv.glm(Data, glm.fit)$delta
[1] 5.890979 5.888812
> glm.fit = glm(y ~ poly(x, 2))
> cv.glm(Data, glm.fit)$delta
[1] 1.086596 1.086326
> glm.fit = glm(y ~ poly(x, 3))
> cv.glm(Data, glm.fit)$delta
[1] 1.102585 1.102227
> glm.fit = glm(y ~ poly(x, 4))
> cv.glm(Data, glm.fit)$delta
[1] 1.114772 1.114334
```

(f)

```
> summary(glm.fit)

Call:
glm(formula = y ~ poly(x, 4))

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-2.8914  -0.5244   0.0749   0.5932   2.7796

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.8277     0.1041 -17.549   <2e-16 ***
poly(x, 4)1   2.3164     1.0415   2.224   0.0285 *
poly(x, 4)2 -21.0586     1.0415 -20.220   <2e-16 ***
poly(x, 4)3  -0.3048     1.0415  -0.293   0.7704
poly(x, 4)4  -0.4926     1.0415  -0.473   0.6373
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for gaussian family taken to be 1.084654)

    Null deviance: 552.21  on 99  degrees of freedom
Residual deviance: 103.04  on 95  degrees of freedom
AIC: 298.78

Number of Fisher Scoring iterations: 2
```