Exercise 3

Part A

The equation is set up as the following:

$$\hat{y} = 50 + GPA * \beta_1 + IQ * \beta_2 + Gender * \beta_3 + (GPA * IQ) * \beta_4 + (GPA * Gender) * \beta_5$$

By putting in the beta values, we get:

$$\hat{y} = 50 + GPA * 20 + IQ * 0.07 + Gender * 35 + (GPA * IQ) * 0.01 + (GPA * Gender) * -10$$

From the updated $\hat{y}$, we know that i and ii are wrong because depending on the value of the GPA, females could make more.

Part B

```
In [1]: 50+20*4+110*0.07+1*35+110*4*0.01+4*1*-10
```

```
Out[1]: 137.1
```

Part C

This isn't true, LASSO regression incorporates variable selection by adding a coefficient of zero for predictors to are not statistically significiant. The p-value needs to be computed for each of the predictors first.

Exercise 4

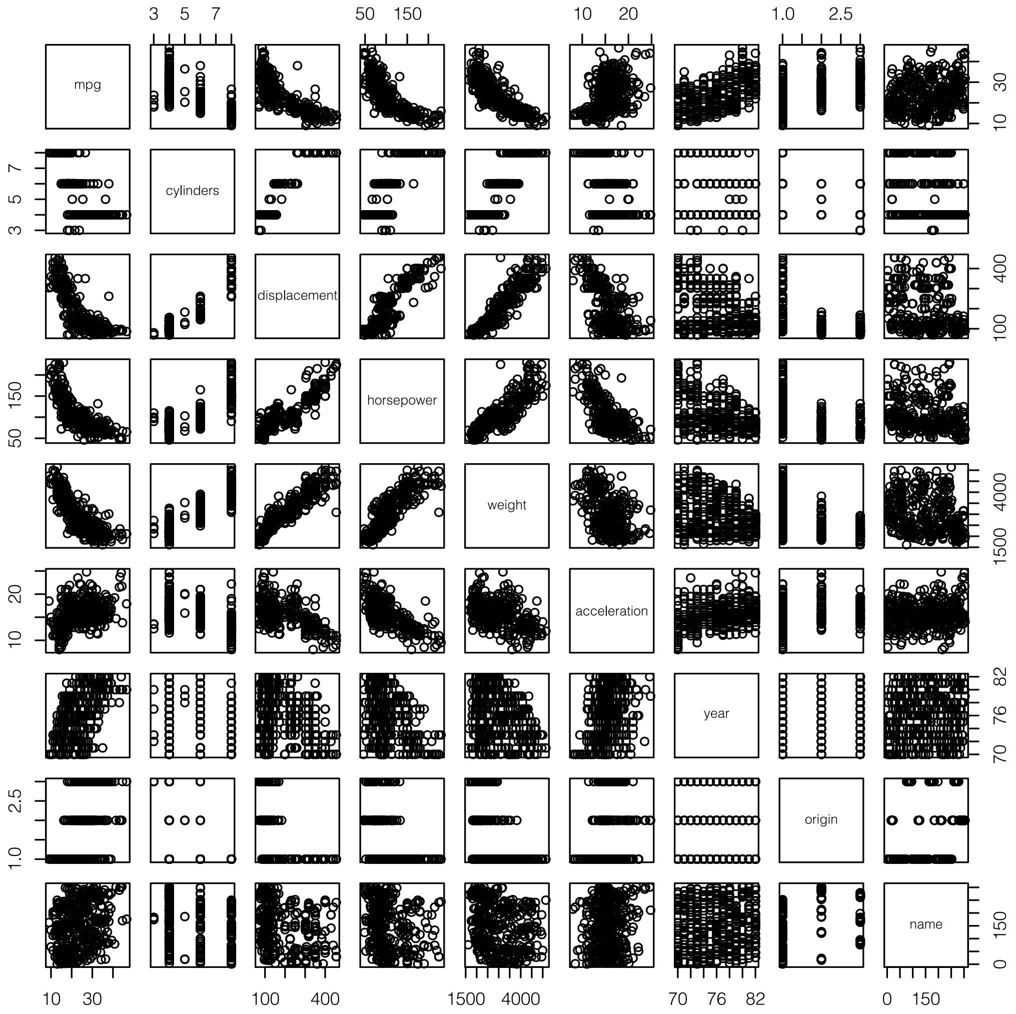
Exercise 9

```
In [2]: library(ISLR)
        data(Auto)
```

Part A

In [3]:

```
pairs(Auto)
```



Part B

In [4]:

```
cor(Auto[,-ncol(Auto)])
```

Out[4]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
mpg	1.0000000	-0.7776175	-0.8051269	-0.7784268	-0.8322442	0.4233285	0.5805410	0.5652088
cylinders	-0.7776175	1.0000000	0.9508233	0.8429834	0.8975273	-0.5046834	-0.3456474	-0.5689316
displacement	-0.8051269	0.9508233	1.0000000	0.8972570	0.9329944	-0.5438005	-0.3698552	-0.6145351
horsepower	-0.7784268	0.8429834	0.8972570	1.0000000	0.8645377	-0.6891955	-0.4163615	-0.4551715
weight	-0.8322442	0.8975273	0.9329944	0.8645377	1.0000000	-0.4168392	-0.3091199	-0.5850054
acceleration	0.4233285	-0.5046834	-0.5438005	-0.6891955	-0.4168392	1.0000000	0.2903161	0.2127458
year	0.5805410	-0.3456474	-0.3698552	-0.4163615	-0.3091199	0.2903161	1.0000000	0.1815277
origin	0.5652088	-0.5689316	-0.6145351	-0.4551715	-0.5850054	0.2127458	0.1815277	1.0000000

Part C

```
In [5]: auto.lm = lm(mpg~.,data=Auto[,-ncol(Auto)])
        summary(auto.lm)
```

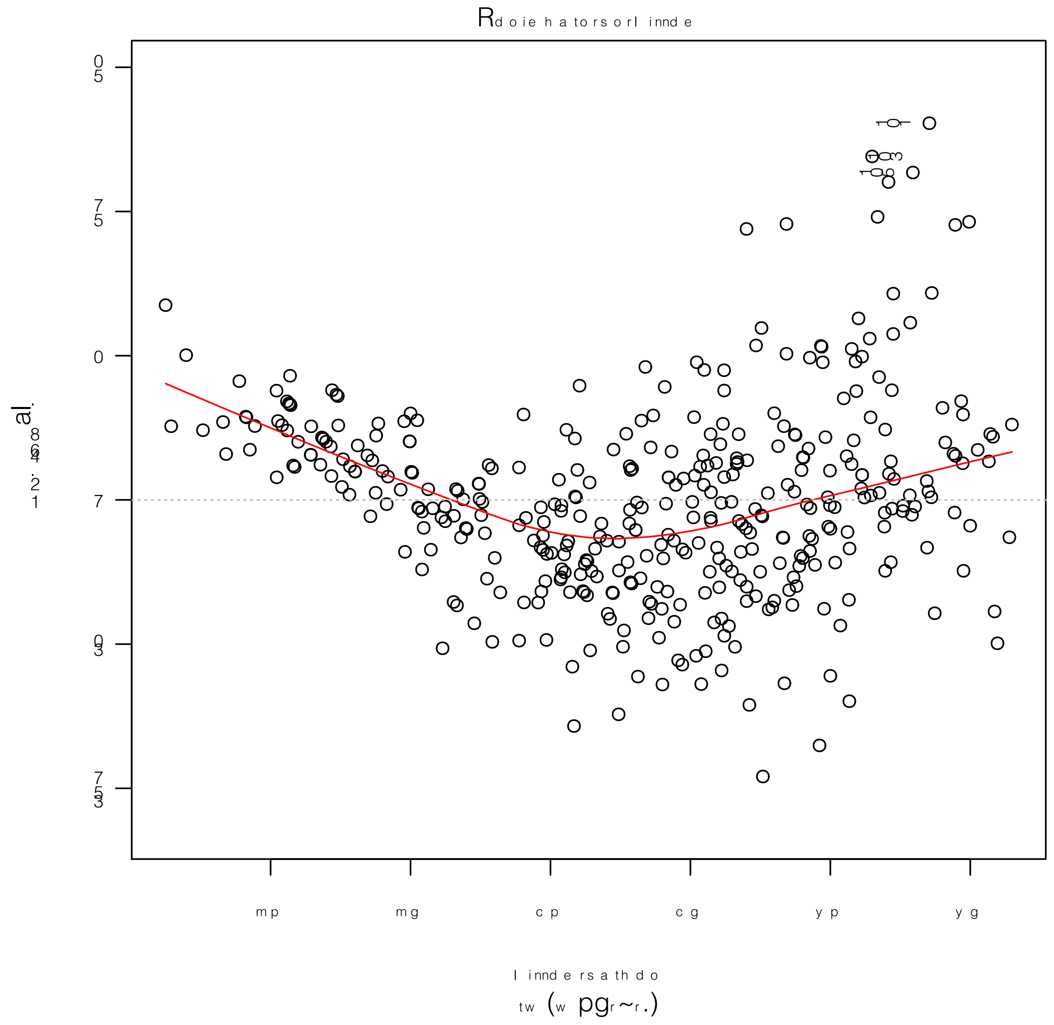
```
Out[5]: Call:
        lm(formula = mpg ~ ., data = Auto[, -ncol(Auto)])

        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
```

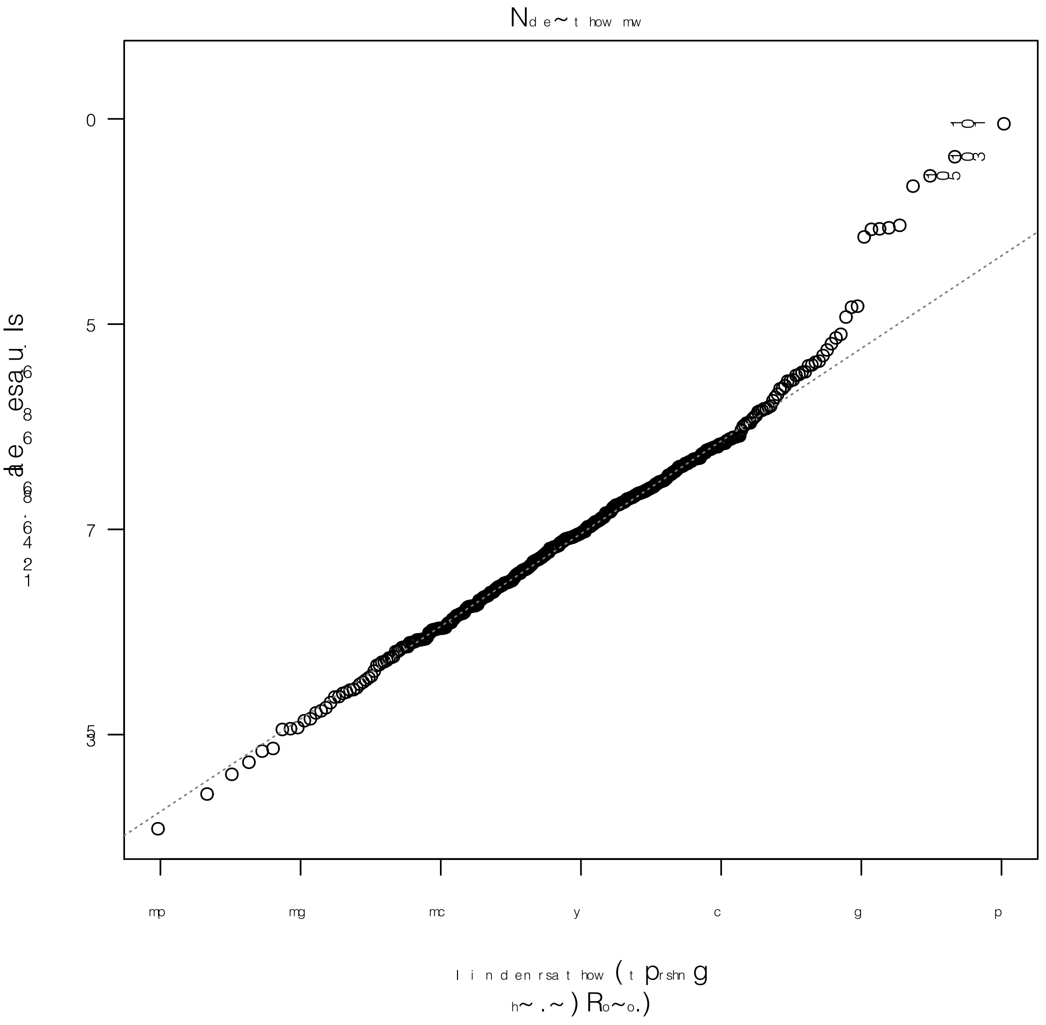
i:

Part D

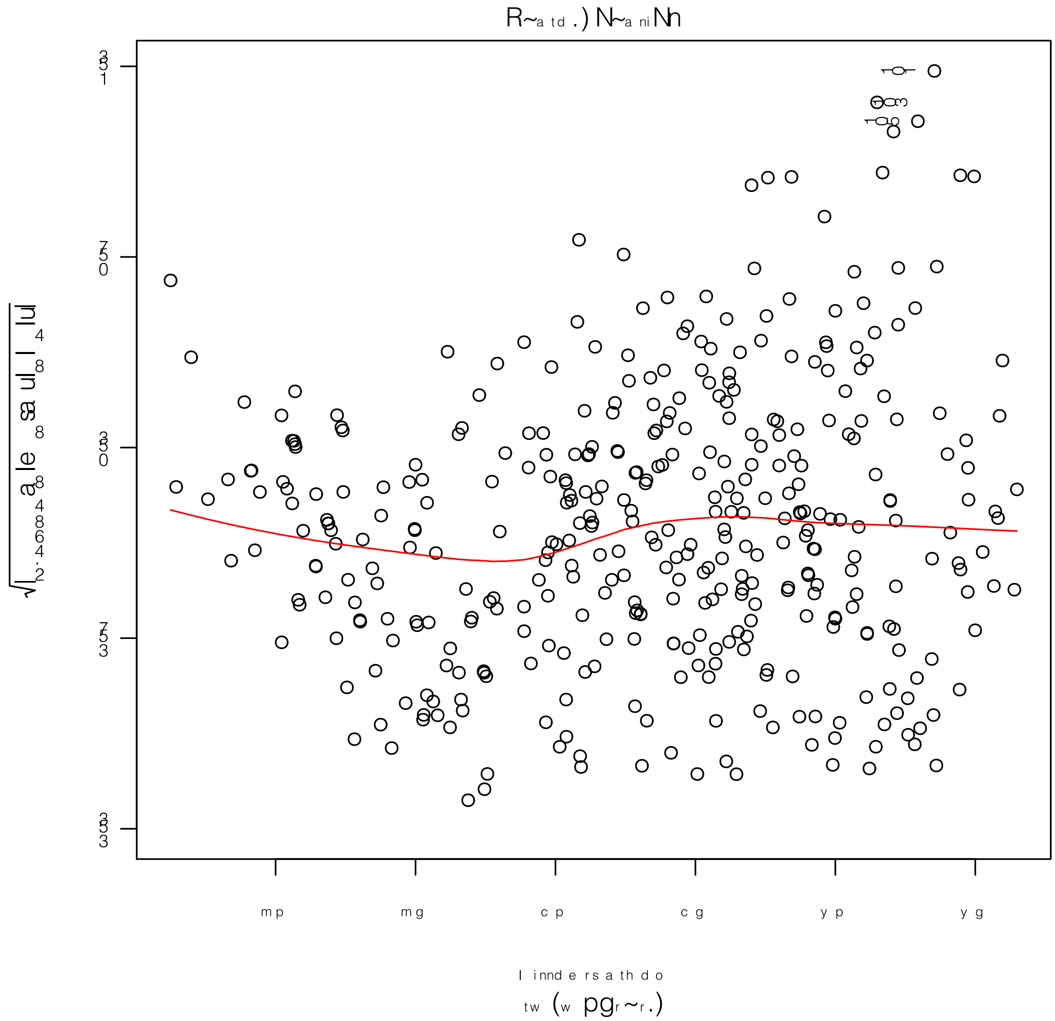
In [6]:

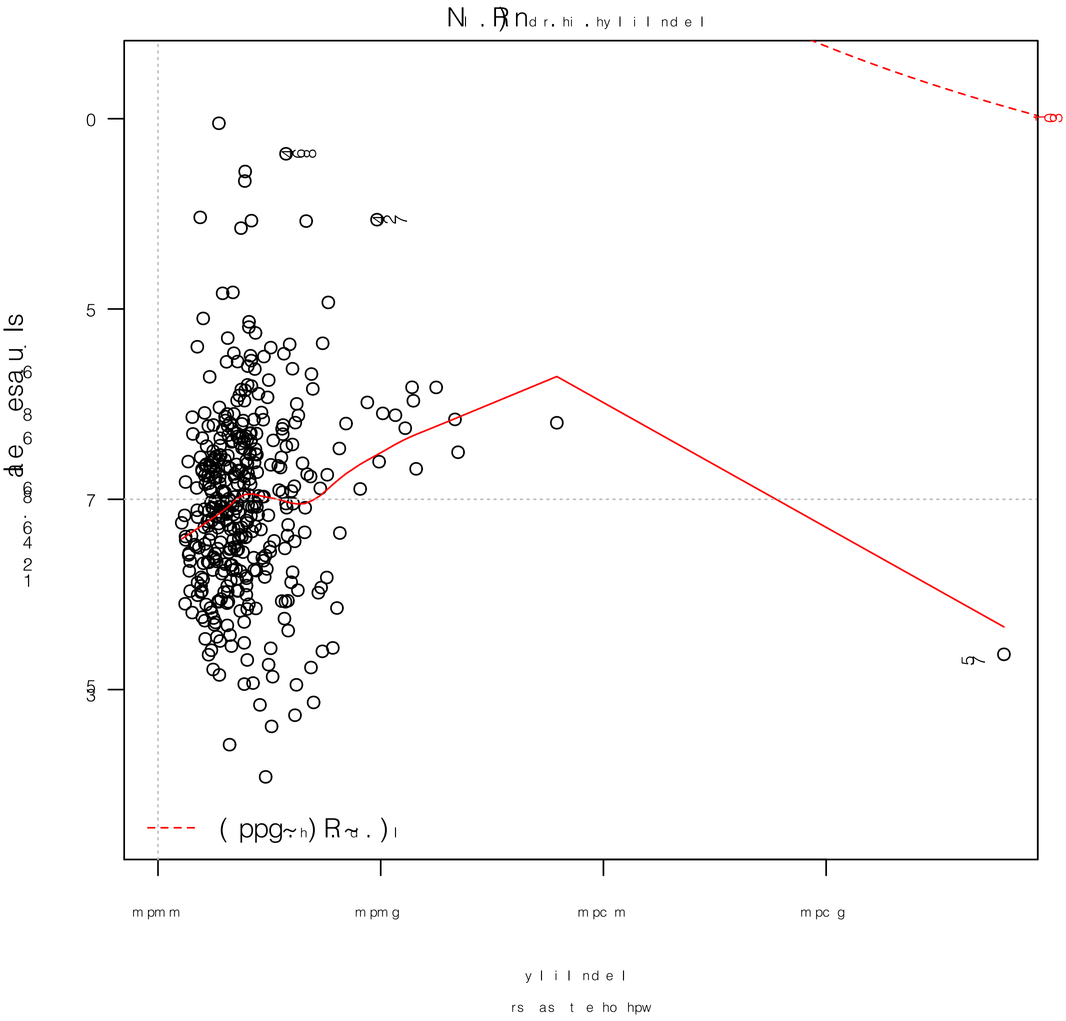
```
plot(auto.lm)
```


Rdoie hatorsorl innde
0
5
7
5
0
7
0
3
7
5
3
al.
8
6
4
2
1
1o
1o3
1o5
mp
mg
cp
cg
yp
yg
l innd e rs a th d o
tw (w pgr~r.)



Nd e~ t how mw
0
5
7
5
3
ale esa u. ls
6
8
6
6
8
6
4
2
1
1
3
5
mp
mg
mc
y
c
g
p
I i n d en rsa t how (t prshn g
h~ .~) Ro~o.)


R~atd.) N~aniNh
3
5
1
7
5
0
3
5
0
7
5
3
3
3
√|₀: a le ₈ ₈ sa ul ₈ l ₄ lul 4 8 6 4 2
mp
mg
cp
cg
yp
yg
linndersathdo
tw (w pgr~r.)



Part E

```
In [ ]:
```

Part F

Exercise 12

Part A

Part B

```
In [9]: X = rnorm(100)
        Y = rpois(n = 100, lambda = 2)
        train = data.frame(X,Y)
```

```
In [12]: summary(lm(Y~X, data = train))
```

```
Out[12]: Call:
         lm(formula = Y ~ X, data = train)

         Residuals:
             Min      1Q  Median      3Q     Max 
         -2.1529 -1.0049 -0.0166  0.9011  5.2616 

         Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
         (Intercept)   1.9968     0.1398  14.288   <2e-16 ***
         X            -0.1312     0.1289  -1.018    0.311    
         ---
         Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

         Residual standard error: 1.396 on 98 degrees of freedom
         Multiple R-squared:  0.01046,	Adjusted R-squared:  0.00036 
         F-statistic: 1.036 on 1 and 98 DF,  p-value: 0.3113
```

```
In [13]: summary(lm(X~Y, data = train))
```

```
Out[13]: Call:
         lm(formula = X ~ Y, data = train)

         Residuals:
             Min      1Q  Median      3Q     Max 
         -2.9381 -0.7046 -0.0222  0.7164  2.7078 

         Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
         (Intercept)  0.21037    0.19005   1.107    0.271
         Y           -0.07969    0.07831  -1.018    0.311

         Residual standard error: 1.088 on 98 degrees of freedom
         Multiple R-squared:  0.01046,	Adjusted R-squared:  0.00036 
         F-statistic: 1.036 on 1 and 98 DF,  p-value: 0.3113
```

Part C

```
In [14]: X = rnorm(100)
         Y = rnorm(100)
         train = data.frame(X,Y)
```

```
In [15]: summary(lm(Y~X, data = train))
```

```
Out[15]: Call:
         lm(formula = Y ~ X, data = train)

         Residuals:
              Min       1Q   Median       3Q      Max 
         -1.85932 -0.52532  0.00114  0.53170  2.91866 

         Coefficients:
                     Estimate Std. Error t value Pr(>|t|)  
         (Intercept) -0.17468    0.09356  -1.867   0.0649 .
         X           -0.12833    0.09151  -1.402   0.1639  
         ---
         Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

         Residual standard error: 0.934 on 98 degrees of freedom
         Multiple R-squared:  0.01968,	Adjusted R-squared:  0.009672 
         F-statistic: 1.967 on 1 and 98 DF,  p-value: 0.1639
```

```
In [16]: summary(lm(X~Y, data = train))
```

```
Out[16]: Call:
         lm(formula = X ~ Y, data = train)

         Residuals:
             Min      1Q  Median      3Q     Max 
         -3.1919 -0.6560  0.0268  0.7437  2.0877 

         Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
         (Intercept) -0.08637    0.10370  -0.833    0.407
         Y           -0.15331    0.10932  -1.402    0.164

         Residual standard error: 1.021 on 98 degrees of freedom
         Multiple R-squared:  0.01968,   Adjusted R-squared:  0.009672 
         F-statistic: 1.967 on 1 and 98 DF,  p-value: 0.1639
```