

401 HW9

Sylvia

11/16/2019

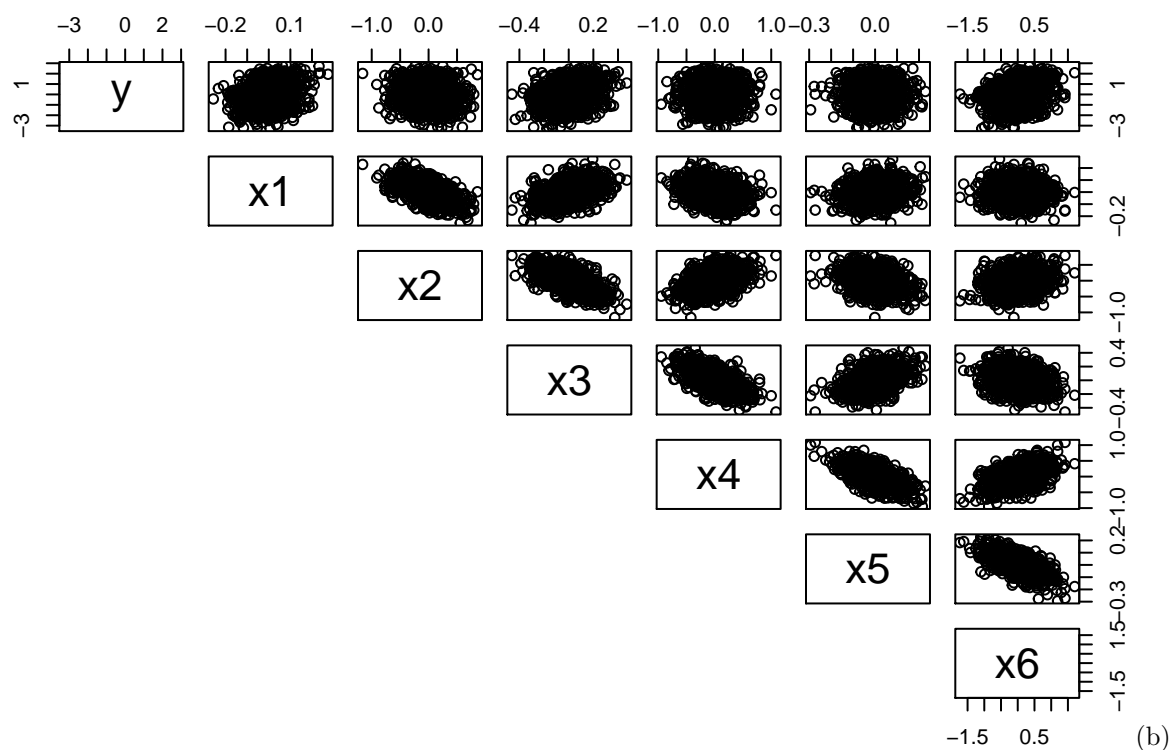
Problem #1

(a)

```
Rpdata <- read.csv("Rpdata.csv", header = TRUE)
head(Rpdata)
```

```
##      X          y          x1          x2          x3          x4          x5          x6
## 1 1  0.31680 -0.034611 -0.0635220  0.0437570 -0.030642  0.1303200 -0.56870
## 2 2 -0.58091 -0.028196  0.2368700 -0.1884600  0.190510 -0.0818980  0.29123
## 3 3 -0.38789  0.059727  0.0884460 -0.0946590  0.022736  0.0073608  0.40923
## 4 4 -0.31245  0.052236  0.0032441 -0.1222200  0.127780  0.0471430 -0.64441
## 5 5  0.96025 -0.057986  0.0659260 -0.0070171  0.107700  0.1265700 -0.71757
## 6 6  1.18400  0.189340 -0.4765500  0.1547300 -0.321130  0.0726280 -0.13887
```

```
pairs(~y+x1+x2+x3+x4+x5+x6, data = Rpdata, lower.panel = NULL)
```



```
model.full <- lm(y~x1+x2+x3+x4+x5+x6, data = Rpdata)
summary(model.full)
```

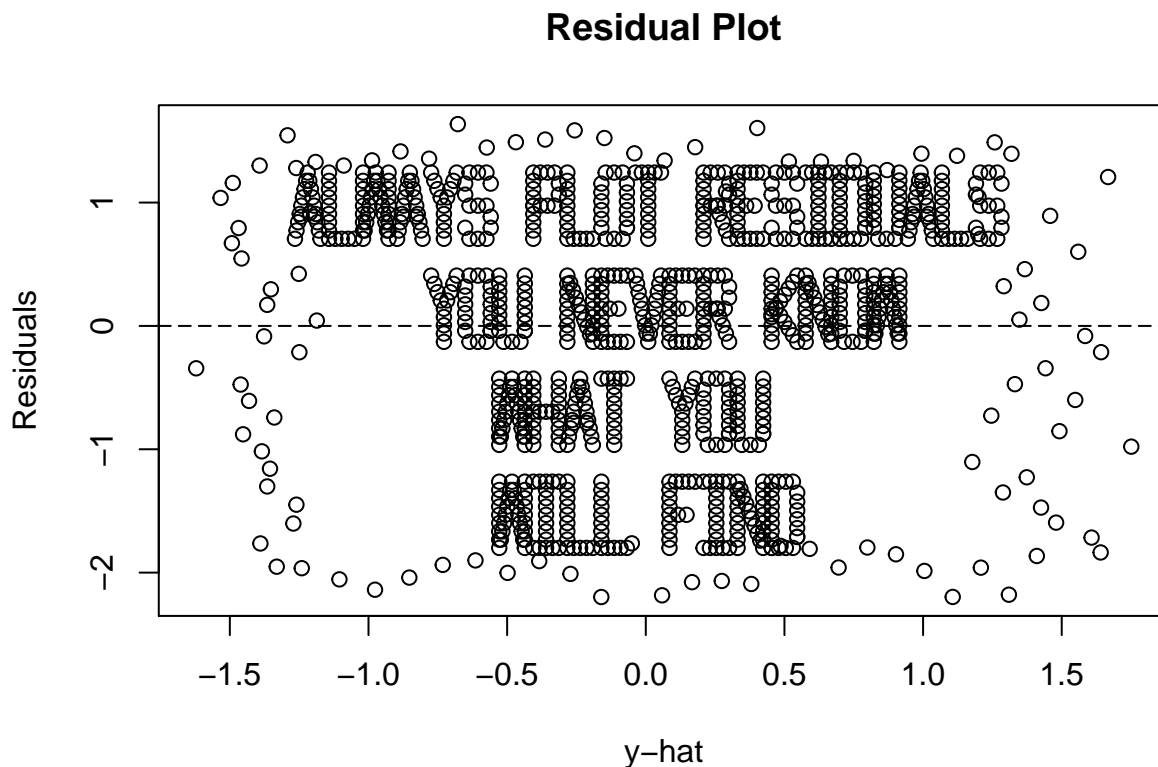
```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6, data = Rpdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -2.1977 -0.7631 0.1729 0.8851 1.6359
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02481    0.03188   0.778   0.437
## x1           4.14061    0.50954   8.126 1.32e-15 ***
## x2           1.01233    0.15522   6.522 1.11e-10 ***
## x3           3.99614    0.32663  12.234 < 2e-16 ***
## x4           0.96045    0.16657   5.766 1.09e-08 ***
## x5           3.75122    0.64726   5.796 9.17e-09 ***
## x6           0.95390    0.08561  11.142 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.003 on 983 degrees of freedom
## Multiple R-squared:  0.3112, Adjusted R-squared:  0.307
## F-statistic: 74.03 on 6 and 983 DF, p-value: < 2.2e-16
```

(c)

Residual vs Fitted Value

```
y.hat <- fitted(model.full)
ep.hat <- resid(model.full)
plot(y.hat, ep.hat, main = "Residual Plot", ylab = "Residuals", xlab = "y-hat")
abline(h=0, lty = 5)
```



Problem #3

(a)

```
Fuel2001 <- read.csv("Fuel2001.csv", header = TRUE)
```

```
Fuel2001$Dlic = (Fuel2001$Drivers/Fuel2001$Pop)*1000
```

```
Fuel2001$Fuel = (Fuel2001$FuelC/Fuel2001$Pop)*1000
```

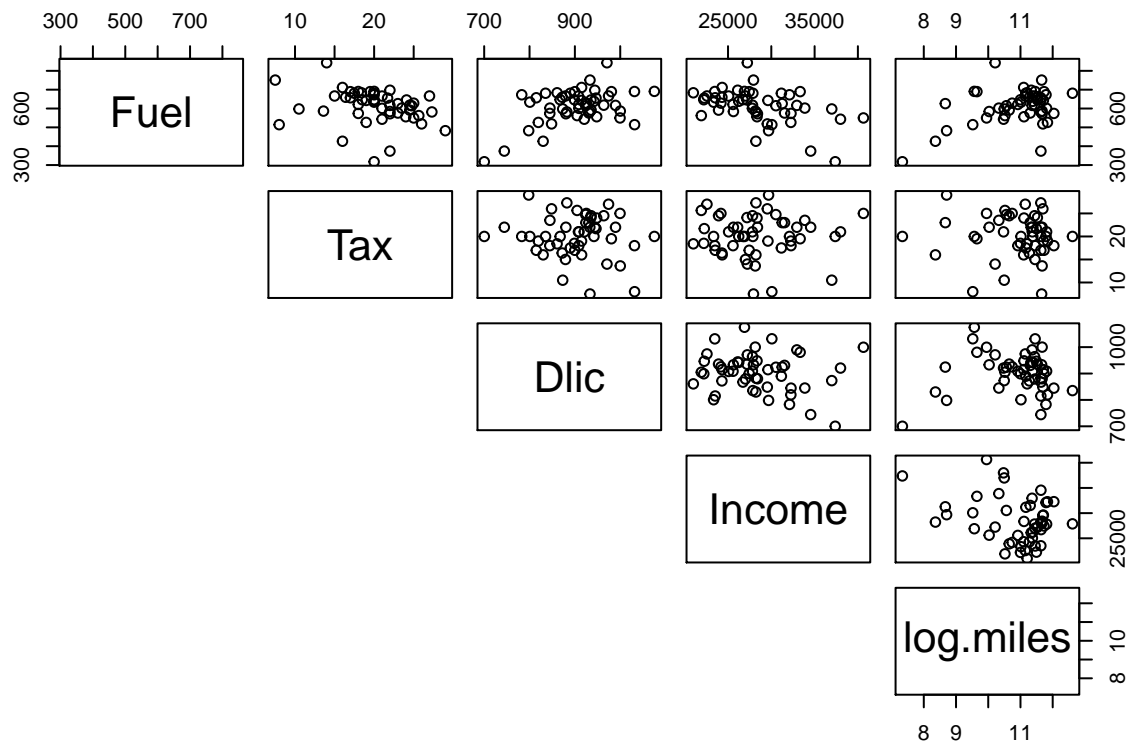
```
Fuel2001$log.miles = log(Fuel2001$Miles)
```

```
head(Fuel2001)
```

```
##   State Drivers   FuelC Income Miles      Pop Tax      Dlic      Fuel
## 1    AL 3559897 2382507 23471 94440 3451586 18.0 1031.3801 690.2644
## 2    AK  472211  235400 30064 13628  457728  8.0 1031.6411 514.2792
## 3    AZ 3550367 2428430 25578 55245 3907526 18.0  908.5972 621.4751
## 4    AR 1961883 1358174 22257 98132 2072622 21.7  946.5706 655.2927
## 5    CA 21623793 14691753 32275 168771 25599275 18.0  844.7033 573.9129
## 6    CO 3287922 2048664 32949 85854 3322455 22.0  989.6062 616.6115
##   log.miles
## 1 11.455720
## 2  9.519882
## 3 10.919533
## 4 11.494069
## 5 12.036298
## 6 11.360403
```

(b)

```
pairs(~Fuel+Tax+Dlic+Income+log.miles, data = Fuel2001, lower.panel = NULL)
```



(c)

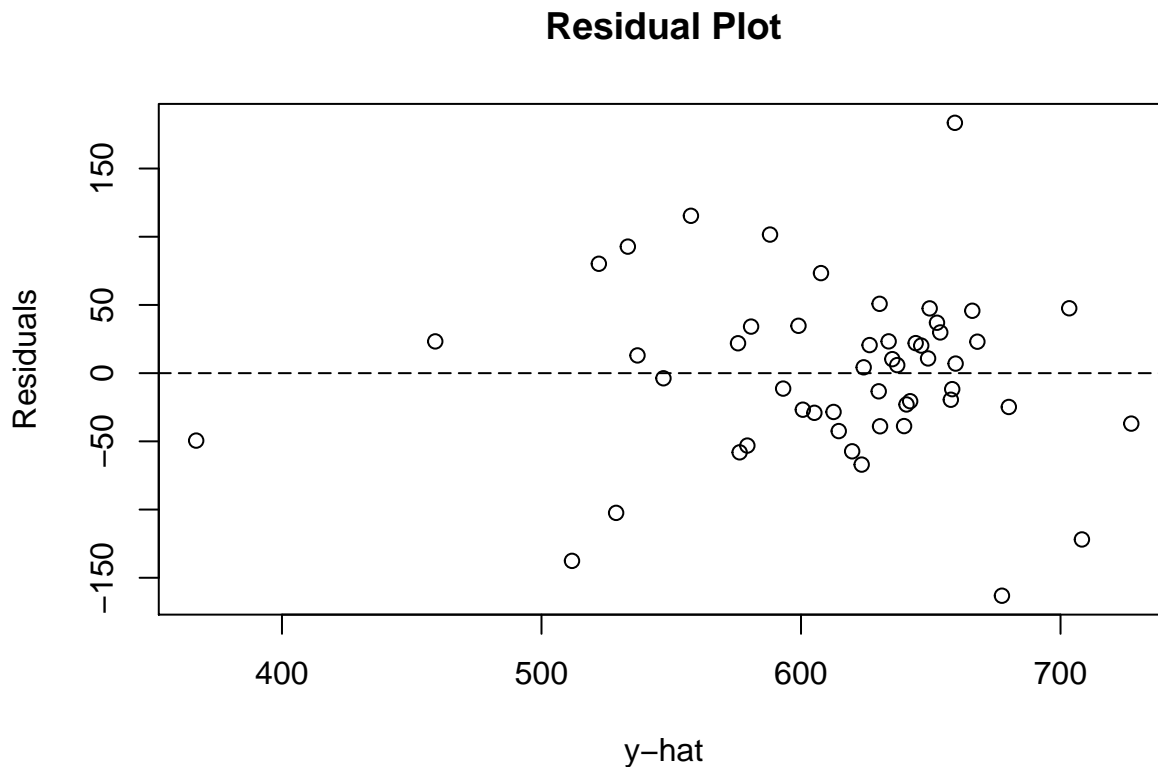
```
model <- lm(Fuel~Tax+Dlic+Income+log.miles, data = Fuel2001)
```

```
summary(model)
```

```
##
## Call:
## lm(formula = Fuel ~ Tax + Dlic + Income + log.miles, data = Fuel2001)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -163.145  -33.039   5.895   31.989  183.499
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 154.192845 194.906161  0.791 0.432938
## Tax         -4.227983   2.030121 -2.083 0.042873 *
## Dlic          0.471871   0.128513  3.672 0.000626 ***
## Income      -0.006135   0.002194 -2.797 0.007508 **
## log.miles    26.755176   9.337374  2.865 0.006259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64.89 on 46 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.4679
## F-statistic: 11.99 on 4 and 46 DF,  p-value: 9.331e-07
```

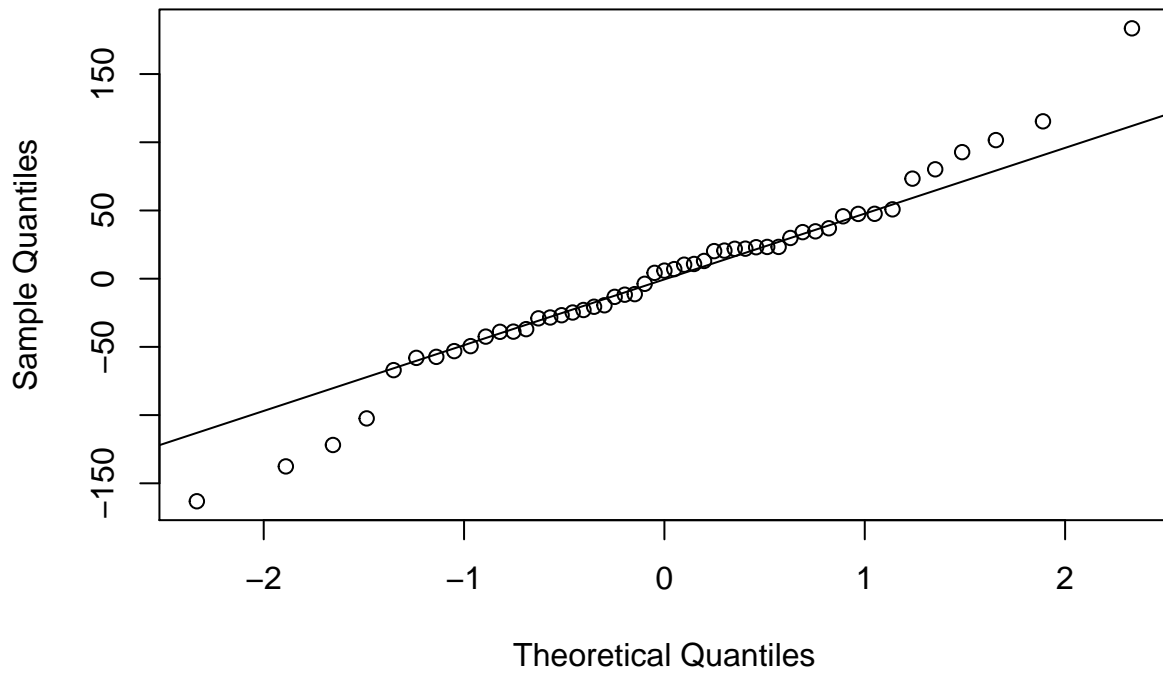
(d)

```
y.hat1 <- fitted(model)
ep.hat1 <- resid(model)
plot(y.hat1, ep.hat1, main = "Residual Plot", ylab = "Residuals", xlab = "y-hat")
abline(h=0, lty = 5)
```



```
qqnorm(ep.hat1)
qqline(ep.hat1)
```

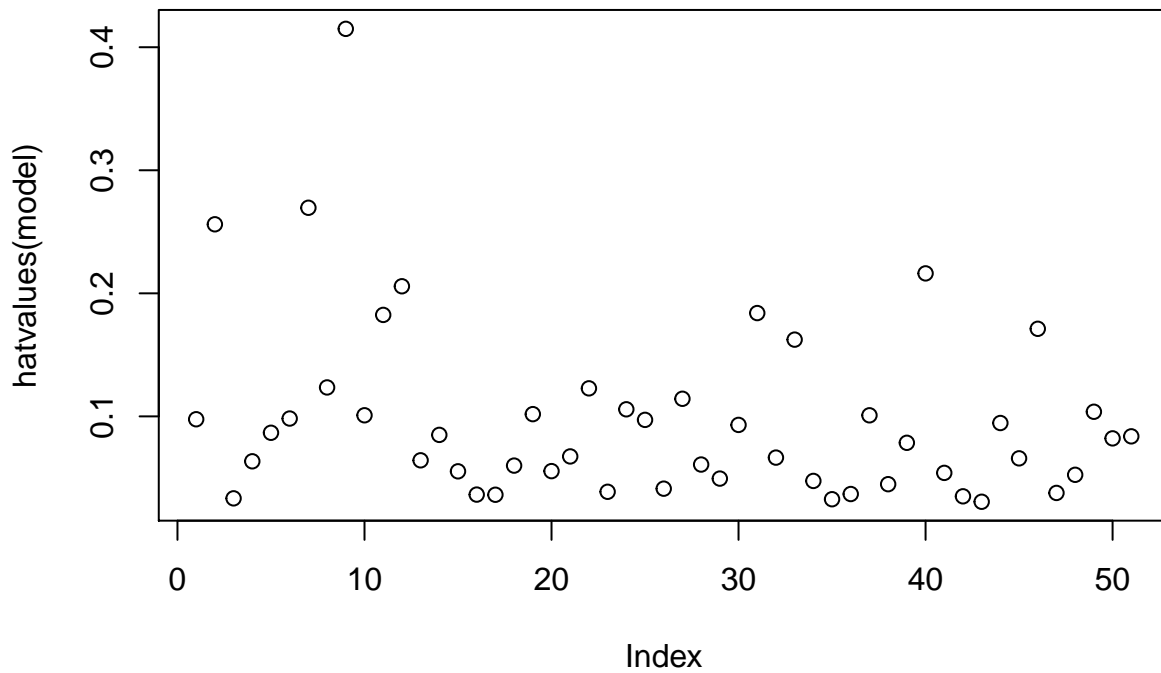
Normal Q-Q Plot



(e)

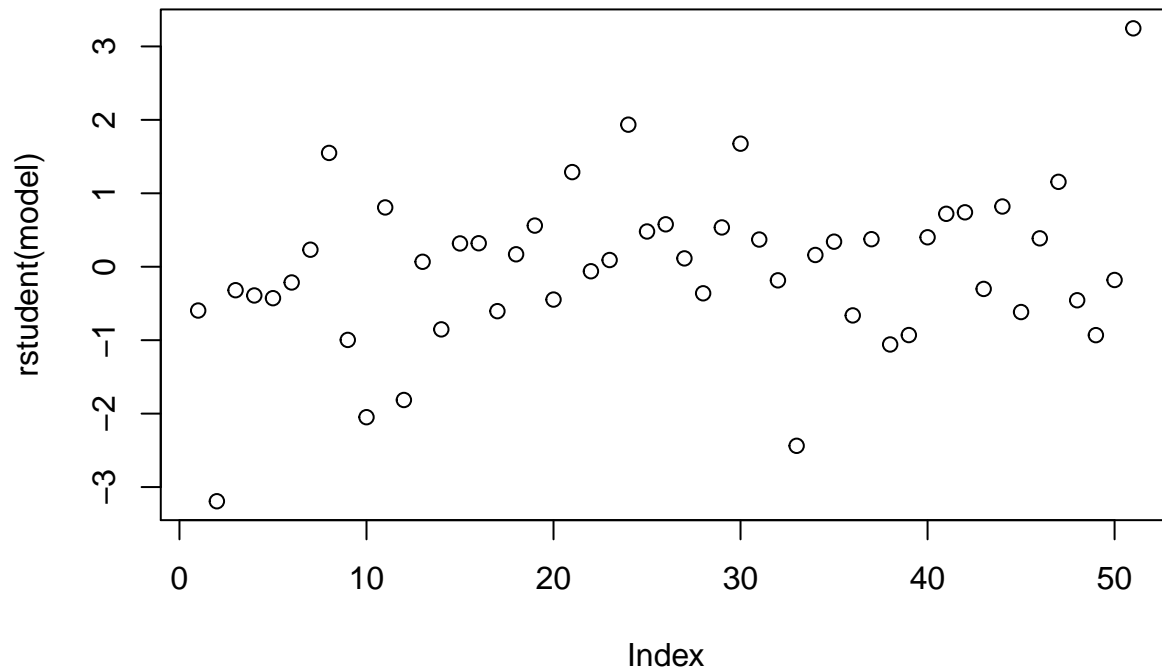
```
plot(hatvalues(model), main = "hii vs index")
```

hii vs index



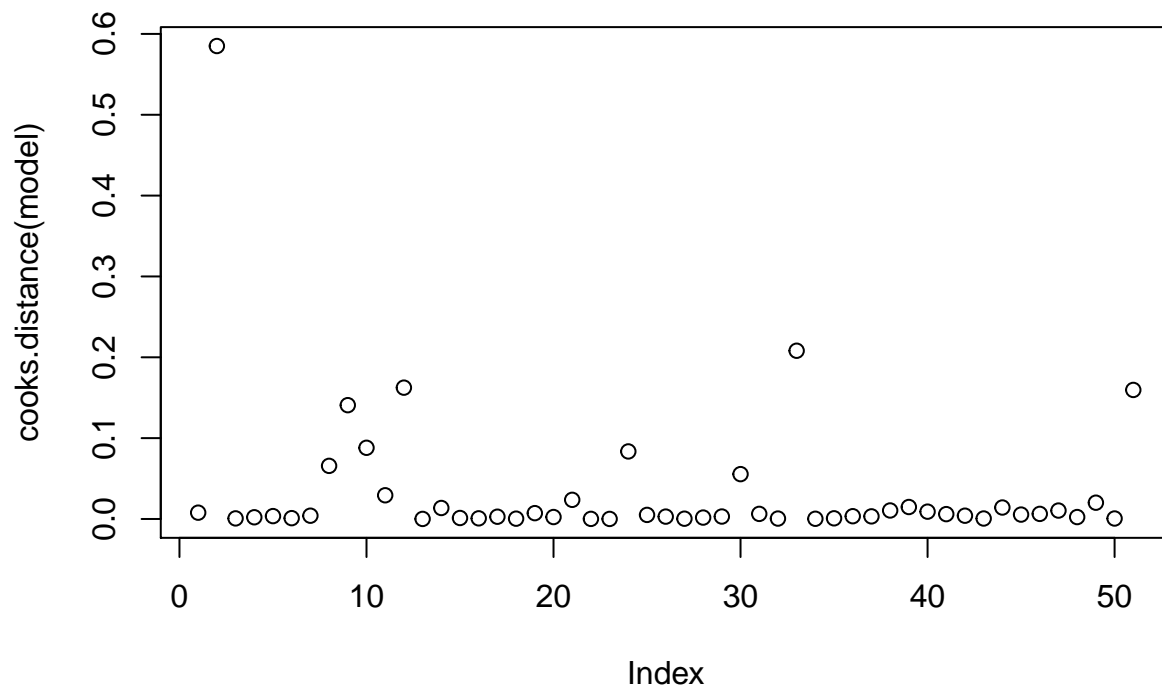
```
plot(rstudent(model), main = "student.res vs index")
```

student.res vs index



```
plot(cooks.distance(model), main = "Cook's D vs index")
```

Cook's D vs index



```
rownames(Fuel2001) <- Fuel2001$State  
case.analysis <- cbind(rownames(Fuel2001), rstudent(model), cooks.distance(model))  
case.analysis
```

| ## | [,1] | [,2] | [,3] |
|-------|------|-----------------------|------------------------|
| ## 1 | "AL" | "-0.596042485993333" | "0.00780045852114796" |
| ## 2 | "AK" | "-3.19302217075954" | "0.585026016676098" |
| ## 3 | "AZ" | "-0.31940485601251" | "0.00071888219116757" |
| ## 4 | "AR" | "-0.391014980940815" | "0.00211167256675987" |
| ## 5 | "CA" | "-0.428027576924223" | "0.00354062288733056" |
| ## 6 | "CO" | "-0.214050435234415" | "0.00101961390379869" |
| ## 7 | "CT" | "0.231972968182381" | "0.00405454506860829" |
| ## 8 | "DE" | "1.54976004212964" | "0.065686332159253" |
| ## 9 | "DC" | "-0.996210235832272" | "0.140779849591751" |
| ## 10 | "FL" | "-2.04875497248923" | "0.0880742704538602" |
| ## 11 | "GA" | "0.807404526670859" | "0.0293205837216657" |
| ## 12 | "HI" | "-1.81436530704376" | "0.162436671129185" |
| ## 13 | "ID" | "0.0674582668198293" | "6.39175324396592e-05" |
| ## 14 | "IL" | "-0.852733940352972" | "0.0135842718701548" |
| ## 15 | "IN" | "0.317342974946862" | "0.0012036419785702" |
| ## 16 | "IA" | "0.319848530901146" | "0.000789183507346487" |
| ## 17 | "KS" | "-0.605021515664667" | "0.0027979523792731" |
| ## 18 | "KY" | "0.169427110018665" | "0.000374380833063231" |
| ## 19 | "LA" | "0.56022238730954" | "0.00722369601977901" |
| ## 20 | "ME" | "-0.447127226675198" | "0.00239234460166189" |
| ## 21 | "MD" | "1.28876754814011" | "0.0236874819911553" |
| ## 22 | "MA" | "-0.0612614671462933" | "0.000107374510950894" |
| ## 23 | "MI" | "0.0916513038634789" | "6.91669753497943e-05" |
| ## 24 | "MN" | "1.93472315028076" | "0.0835371533032919" |
| ## 25 | "MS" | "0.480000815117261" | "0.00504363774668347" |
| ## 26 | "MO" | "0.57755618602351" | "0.0029129848585571" |
| ## 27 | "MT" | "0.113683514199963" | "0.00034088305217595" |
| ## 28 | "NE" | "-0.36157527412908" | "0.00172708590191034" |
| ## 29 | "NV" | "0.535282407417611" | "0.00302830451112673" |
| ## 30 | "NH" | "1.67591799114767" | "0.0554615146617763" |
| ## 31 | "NJ" | "0.370253662280363" | "0.00629865177076245" |
| ## 32 | "NM" | "-0.185137419124969" | "0.000498888304262042" |
| ## 33 | "NY" | "-2.43822459554556" | "0.208109881148508" |
| ## 34 | "NC" | "0.160286467119527" | "0.00026208932009073" |
| ## 35 | "ND" | "0.341352792925739" | "0.000802597628418467" |
| ## 36 | "OH" | "-0.66279376093001" | "0.00340761655822384" |
| ## 37 | "OK" | "0.374959052224893" | "0.00321459594729061" |
| ## 38 | "OR" | "-1.05843155955933" | "0.0105037108435948" |
| ## 39 | "PA" | "-0.930304912383491" | "0.0147964385959963" |
| ## 40 | "RI" | "0.401239260174248" | "0.00904679423466858" |
| ## 41 | "SC" | "0.720821071077901" | "0.00600427192004031" |
| ## 42 | "SD" | "0.741157565333648" | "0.00402975074116317" |
| ## 43 | "TN" | "-0.30215152297771" | "0.000588454131168029" |
| ## 44 | "TX" | "0.820131897051737" | "0.0141569366705321" |
| ## 45 | "UT" | "-0.616914030843366" | "0.00544283674563854" |
| ## 46 | "VT" | "0.386738119687449" | "0.00629424764959809" |
| ## 47 | "VA" | "1.15620603047519" | "0.010425578059541" |
| ## 48 | "WA" | "-0.456354424668765" | "0.00234934914970489" |
| ## 49 | "WV" | "-0.931356063976436" | "0.0201412157075794" |
| ## 50 | "WI" | "-0.180295702577323" | "0.000594275803718709" |
| ## 51 | "WY" | "3.24608994718709" | "0.159616897489066" |

```
F.50 <- qf(0.5, 5, 46)
which(case.analysis[,3] > F.50)
```

```
## 13 23
## 13 23
```

(f)

```
model.reduced <- lm(Fuel~Dlic+log.miles, data = Fuel2001)
anova(model.reduced)
```

```
## Analysis of Variance Table
##
## Response: Fuel
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Dlic        1  86854    86854  17.156 0.0001387 ***
## log.miles    1  65834    65834   13.004 0.0007382 ***
## Residuals  48 243006     5063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(model)
```

```
## Analysis of Variance Table
##
## Response: Fuel
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Tax         1  26635    26635   6.3254 0.0154602 *
## Dlic         1  79378    79378  18.8506 7.692e-05 ***
## Income       1  61408    61408  14.5833 0.0003997 ***
## log.miles    1  34573    34573   8.2104 0.0062592 **
## Residuals  46 193700     4211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.