liner.R

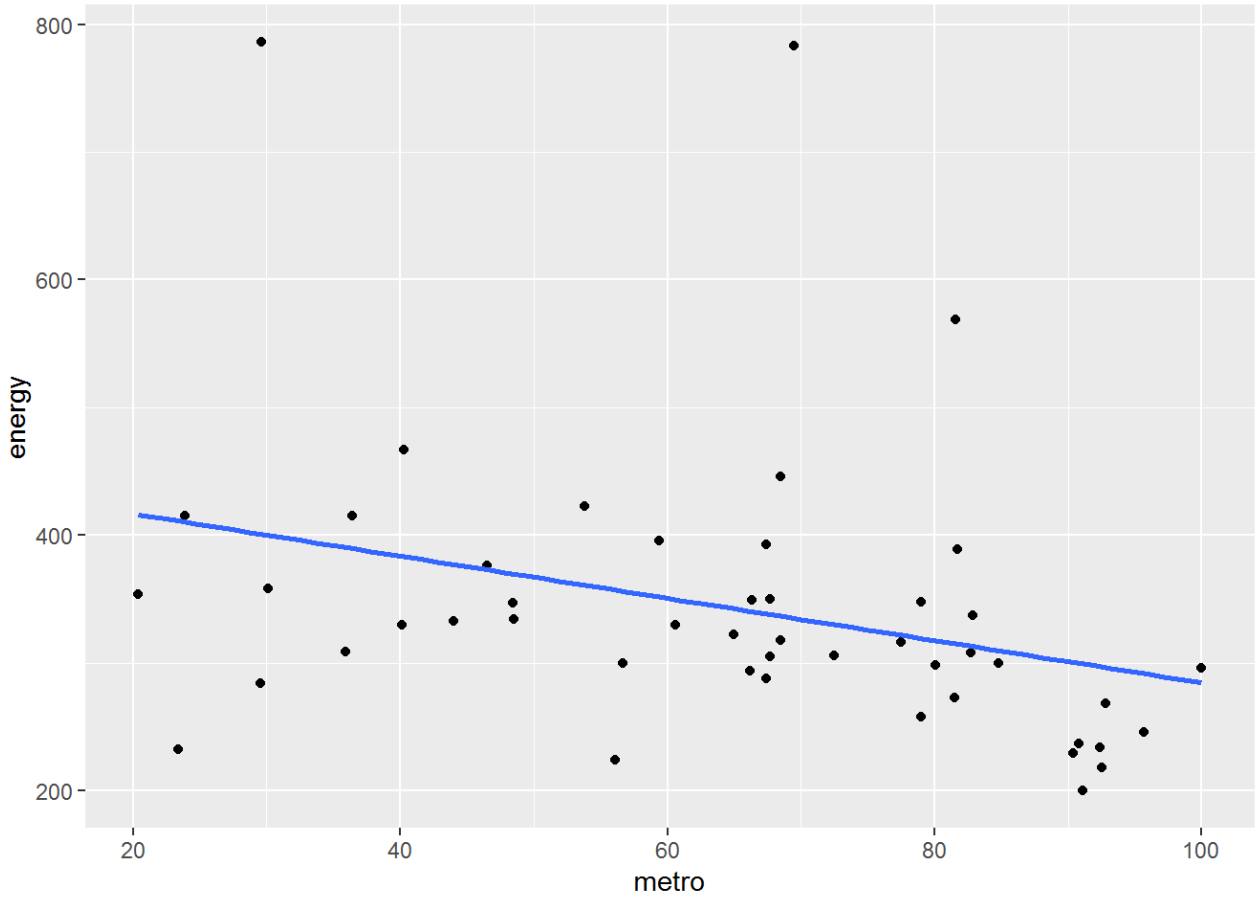
Pranshu

Wed Jun 22 21:10:19 2016

```
Exam <- readRDS("~/Research/TestRegression/Exam.rds")
View(Exam)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
states <- readRDS("~/Research/TestRegression/states.rds")
View(states)
states2<-na.omit(states)
View(states2)
model1<-lm(energy~metro,data = states2)
ggplot(states2,aes(x=metro,y=energy))+geom_point()+stat_smooth(method = "lm",se
=FALSE)
```



```
summary(model1)
```

```
##
## Call:
## lm(formula = energy ~ metro, data = states2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -179.17  -54.21  -21.64   15.07  448.02
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 449.8382    50.4472   8.917 1.37e-11 ***
## metro        -1.6526     0.7428  -2.225    0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 112.3 on 46 degrees of freedom
## Multiple R-squared:  0.09714,    Adjusted R-squared:  0.07751
## F-statistic: 4.949 on 1 and 46 DF,  p-value: 0.03105
```

```
predictmodel=predict(model1,data=states2)
predictmodel
```

```
##        1        3        4        5        6        7        8       10
## 338.4532 319.2831 383.5691 291.6848 315.1516 297.1383 340.2711 299.7825
##       11       13       14       15       16       17       18       19
## 342.4195 416.1252 313.1685 336.6354 377.1240 360.9285 372.9925 334.9828
##       20       21       22       23       24       25       26       27
## 390.5100 296.4773 300.4435 317.4653 337.9574 400.0951 340.4363 410.3412
##       28       29       30       31       32       33       34       35
## 369.6873 312.8380 357.1276 284.5786 369.8525 299.2867 356.1360 383.2386
##       36       37       38       39       40       41       42       43
## 319.2831 351.6740 336.6354 309.6980 296.9731 349.6909 401.0866 337.9574
##       44       45       46       47       48       49       50       51
## 314.9864 321.7620 411.1675 330.0250 314.8211 389.6837 338.4532 400.9214
```

```
SSE=sum((states2$energy-predictmodel)^2)
SST=sum((states2$energy-mean(states2$energy))^2)
1-SSE/SST
```

```
## [1] 0.09714146
```

```
model2<-lm(energy~metro+density,data=states2)
summary(model2)
```

```
##
## Call:
## lm(formula = energy ~ metro + density, data = states2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -160.56  -50.86  -25.89   27.96  435.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 420.44462   54.05903   7.778 7.28e-10 ***
## metro        -0.88062    0.91532  -0.962    0.341
## density      -0.11958    0.08453  -1.415    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 111.1 on 45 degrees of freedom
## Multiple R-squared:  0.1356, Adjusted R-squared:  0.09717
## F-statistic: 3.529 on 2 and 45 DF,  p-value: 0.03769
```

```
model3<-lm(energy~metro+density+waste, data=states2)
summary(model3)
```

```
##
## Call:
## lm(formula = energy ~ metro + density + waste, data = states2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -166.74  -50.53  -21.33   33.92  419.76
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 473.24711   72.77260   6.503 6.14e-08 ***
## metro        -0.42262    1.00701  -0.420    0.677
## density      -0.12273    0.08442  -1.454    0.153
## waste       -82.82686   76.59647  -1.081    0.285
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 110.9 on 44 degrees of freedom
## Multiple R-squared:  0.158,  Adjusted R-squared:  0.1006
## F-statistic: 2.751 on 3 and 44 DF,  p-value: 0.05385
```

```
region<-factor(states2$region)
model4<-lm(energy~region,data=states2)
summary(model4)
```

```
##
## Call:
## lm(formula = energy ~ region, data = states2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -143.13  -50.13  -23.62   17.36  418.82
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     367.18      33.25  11.044 2.87e-14 ***
## regionN. East  -118.07      49.56  -2.382   0.0216 *
## regionSouth      12.94      43.19   0.300   0.7658
## regionMidwest   -23.18      46.03  -0.504   0.6170
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 110.3 on 44 degrees of freedom
## Multiple R-squared:  0.1677, Adjusted R-squared:  0.111
## F-statistic: 2.956 on 3 and 44 DF,  p-value: 0.04268
```

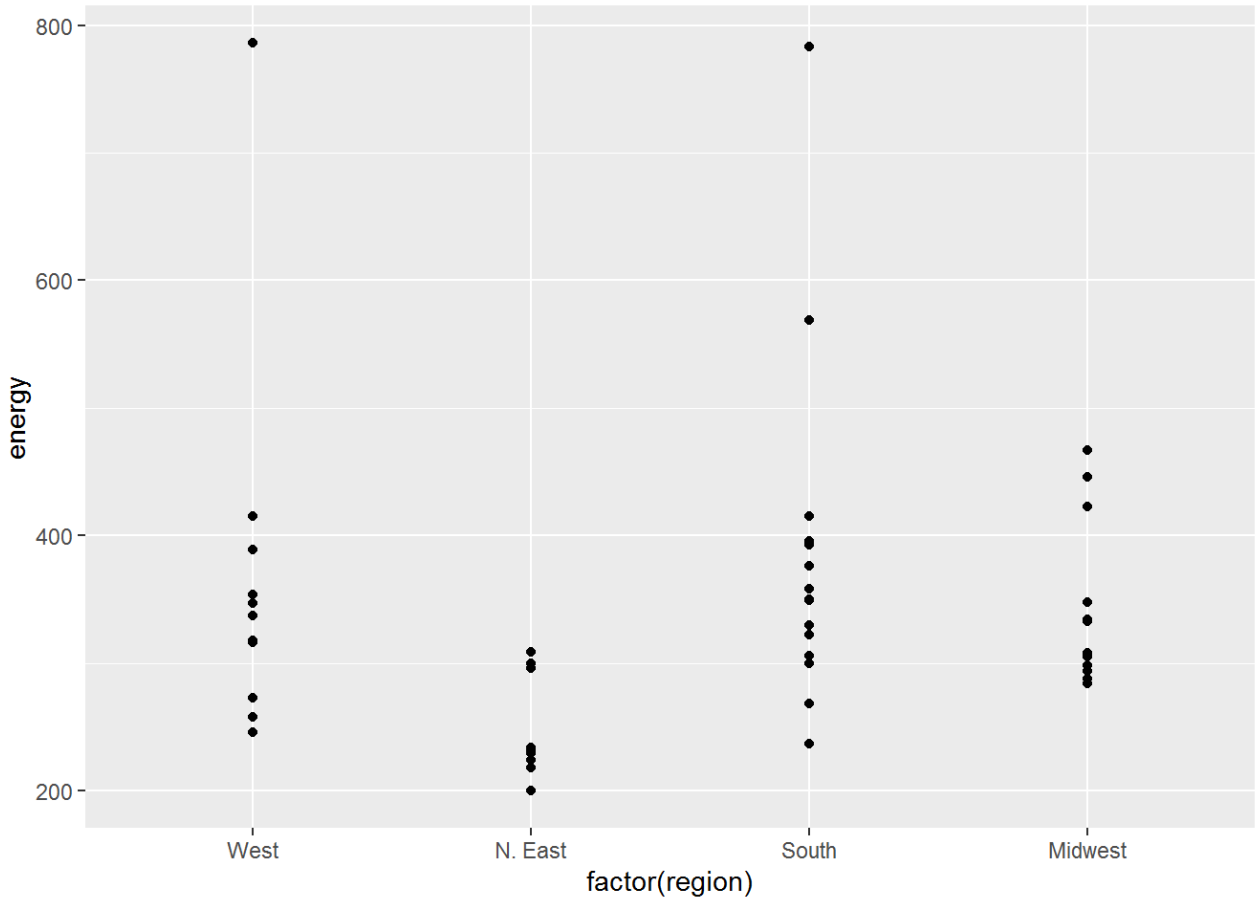
```
anova(model4)
```

```
## Analysis of Variance Table
##
## Response: energy
##           Df Sum Sq Mean Sq F value  Pr(>F)
## region     3 107818   35939  2.9555 0.04268 *
## Residuals 44 535042   12160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov1<-aov(energy~region,data=states2)
summary(aov1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## region       3 107818   35939   2.956 0.0427 *
## Residuals   44 535042   12160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(states2,aes(x=factor(region),y=energy))+geom_point()
```



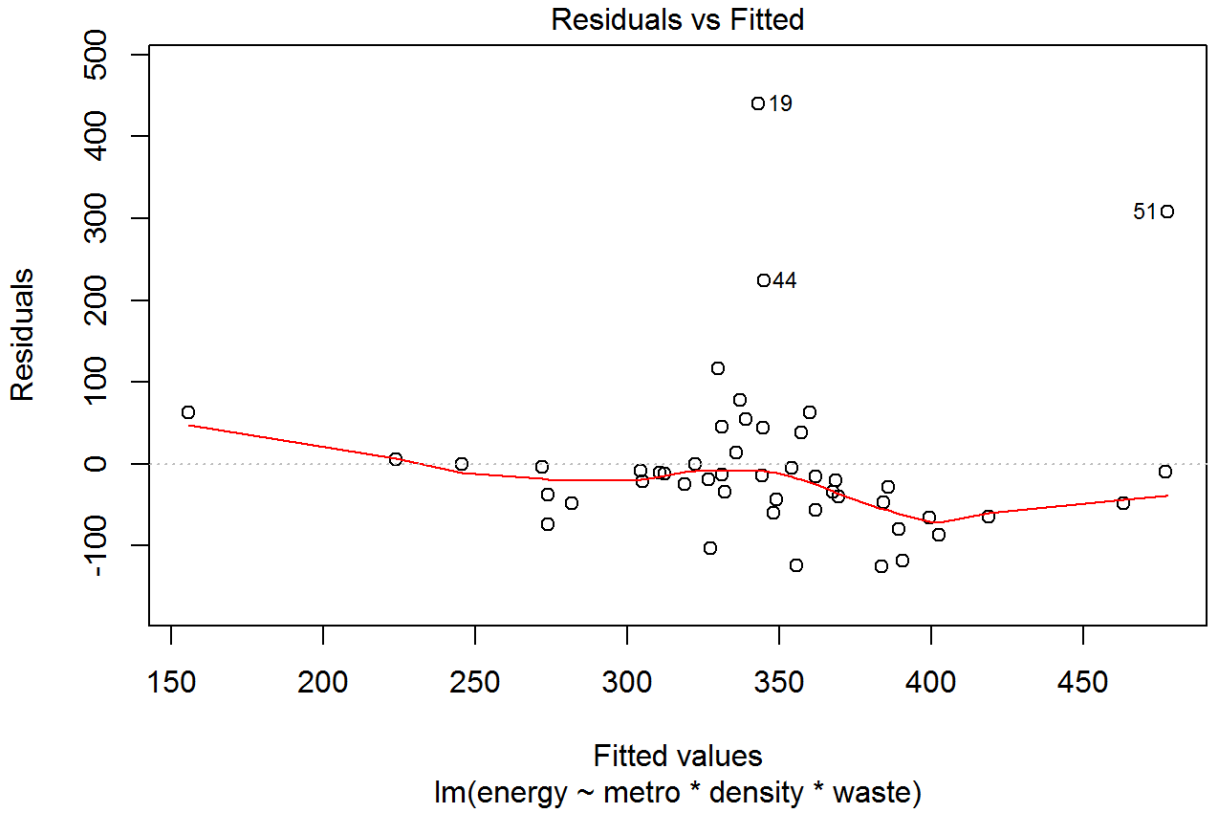

```
contrasts(states2$region)
```

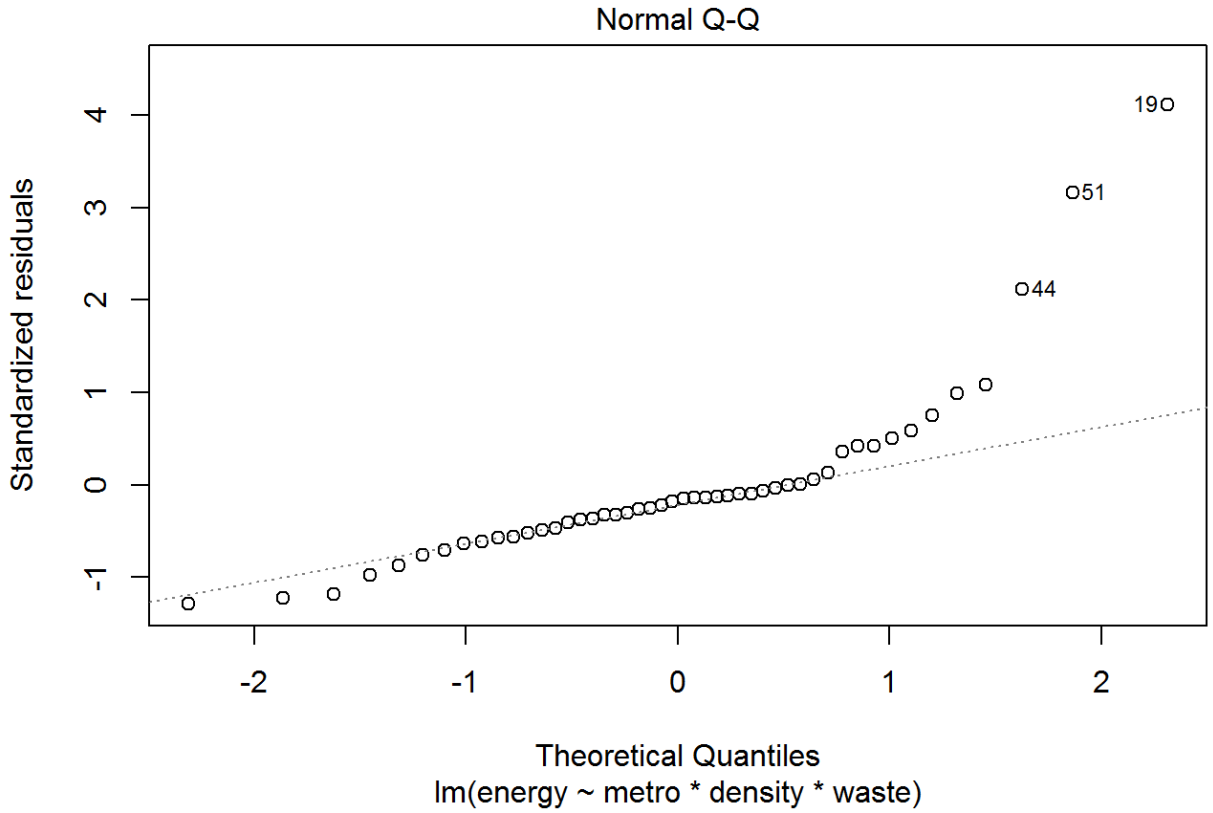
```
##         N. East South Midwest
## West          0     0       0
## N. East       1     0       0
## South         0     1       0
## Midwest       0     0       1
```

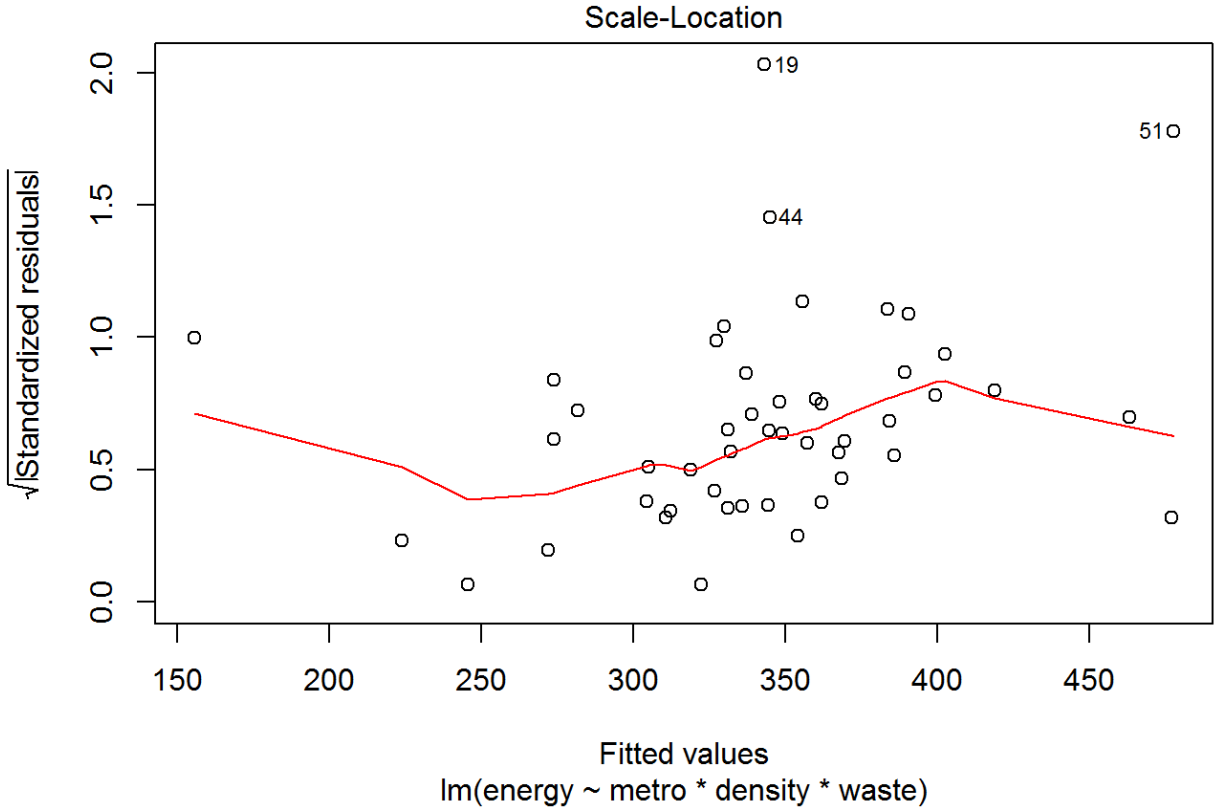
```
model5<-lm(energy~metro*density*waste, data=states2)
summary(model5)
```

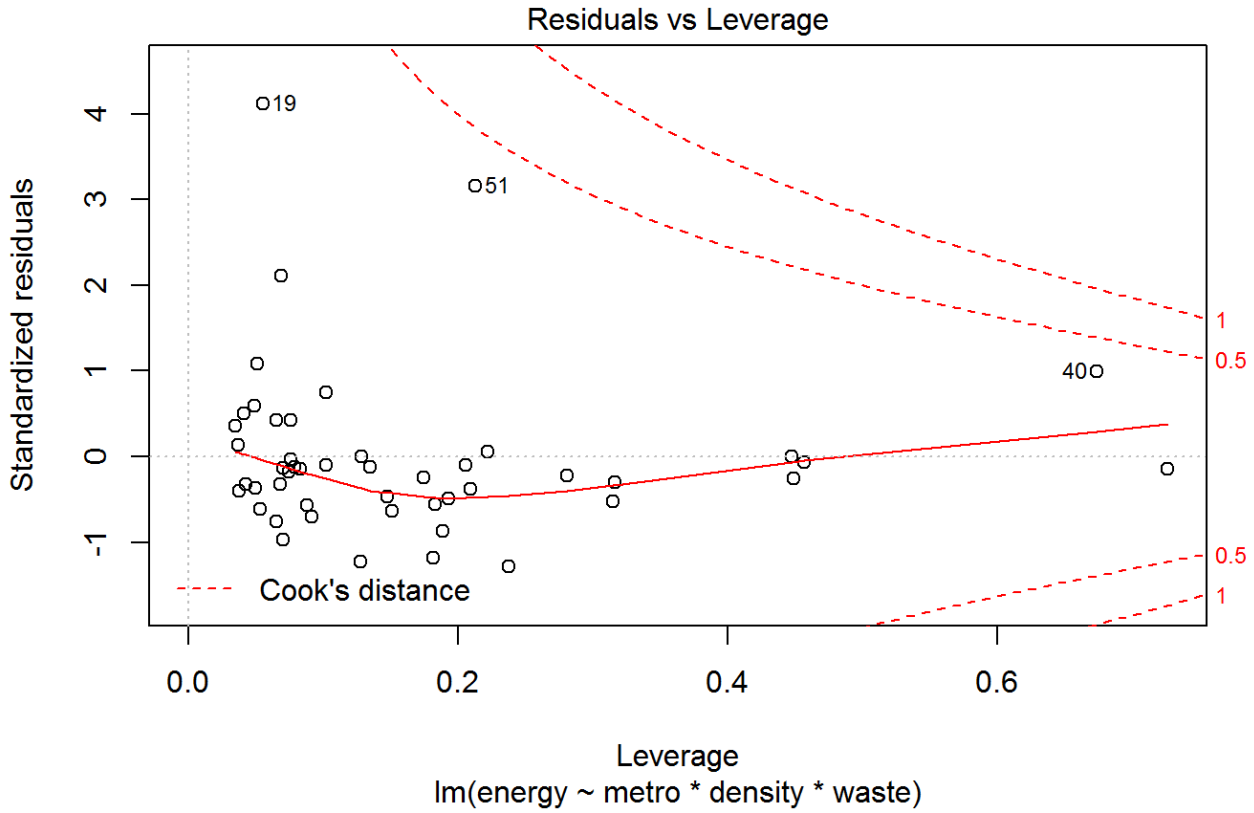
```
##
## Call:
## lm(formula = energy ~ metro * density * waste, data = states2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -125.65  -47.72  -16.79    7.37  439.80
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          928.69939  298.85885   3.107  0.00347 **
## metro                 -5.07130    4.10121  -1.237  0.22347
## density               -9.21044    4.30087  -2.142  0.03838 *
## waste               -566.14583  342.79437  -1.652  0.10645
## metro:density          0.09989    0.04615   2.165  0.03643 *
## metro:waste            5.07814    4.53405   1.120  0.26939
## density:waste          8.75201    4.23590   2.066  0.04533 *
## metro:density:waste   -0.09665    0.04592  -2.105  0.04165 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 109.9 on 40 degrees of freedom
## Multiple R-squared:  0.2491, Adjusted R-squared:  0.1177
## F-statistic: 1.895 on 7 and 40 DF,  p-value: 0.09592
```

```
plot(model5)
```


Residuals vs Fitted
Residuals
Fitted values
lm(energy ~ metro * density * waste)
19
44
51
-100
0
100
200
300
400
500
150
200
250
300
350
400
450


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(energy ~ metro * density * waste)
19
51
44
-1
0
1
2
3
4
-2
-1
0
1
2


Scale-Location
19
51
44
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
150
200
250
300
350
400
450
Fitted values
lm(energy ~ metro * density * waste)


Residuals vs Leverage
Standardized residuals
Leverage
lm(energy ~ metro * density * waste)
Cook's distance
19
51
40
1
0.5
0.5
1
-1
0
1
2
3
4
0.0
0.2
0.4
0.6