Coal

October 9, 2016

Grab US Coal Production data which is measured in thousand short tons.

```
Coal <- read.csv("https://www.eia.gov/totalenergy/data/browser/csv.cfm?tbl=T06.01")

summary(Coal)
```

```
##       MSN           YYYYMM                  Value       Column_Order
##  CLEXPUS: 591   Min.   :194913   Not Available: 244   Min.   :1.00
##  CLIMPUS: 591   1st Qu.:198207   816.667      :   6   1st Qu.:2.75
##  CLLUPUS: 591   Median :199312   2            :   3   Median :4.50
##  CLNIPUS: 591   Mean   :199301   3            :   3   Mean   :4.50
##  CLPRPUS: 591   3rd Qu.:200504   -4657        :   2   3rd Qu.:6.25
##  CLSCPUS: 591   Max.   :201608   114          :   2   Max.   :8.00
##  (Other):1182                    (Other)      :4468
##                           Description                     Unit
##  Coal Consumption               : 591   Thousand Short Tons:4728
##  Coal Exports                   : 591
##  Coal Imports                   : 591
##  Coal Losses and Unaccounted for: 591
##  Coal Net Imports               : 591
##  Coal Production                : 591
##  (Other)                        :1182
```

Load the dplyr package.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Select only the Coal Production Figures and save it as CoalProduction. There is a cheat sheet for dplyr built into R. Look under the help menu.

```
CoalProduction <- Coal %>% filter(MSN == "CLPRPUS")
```

Grab all of the annual observations with month equal to 13, remove unnecessary columns and rename the remaining columns as RawYear and ProductionKShortTon.

```r
library(stringr)

CoalProduction <- CoalProduction %>% filter(str_sub(as.character(YYYYMM),5 ) == "13")

CoalProduction <- CoalProduction %>% select(YYYYMM, Value)

names(CoalProduction) <- c("RawYear", "ProductionKShortTon")

summary(CoalProduction)
```

```
##     RawYear        ProductionKShortTon
##  Min.   :194913   1000048.758: 1
##  1st Qu.:196563   1016458.418: 1
##  Median :198213   1029075.527: 1
##  Mean   :198213   1032973.77 : 1
##  3rd Qu.:199863   1033504.288: 1
##  Max.   :201513   1063855.51 : 1
##                   (Other)    :61
```

The ProductionKShortTon variable shows a count rather than a numerical summary which is considered a factor rather than number by R. Convert the factor, which is an integer, to the real value as a character and then convert that to numeric.

```r
CoalProduction$ProductionKShortTon <- as.numeric(as.character(CoalProduction$ProductionKShortTon))

#$ = access column
#  <-    assignment operator
#as.numeric (as.character(CoalProduction$ProductionKShortTon)) =  turn into number if possible
```

Create a column for the year and make it a numeric value.

```r
CoalProduction <- CoalProduction  %>%mutate(Year = as.numeric(str_sub(as.character(RawYear),0,4 )))


summary(CoalProduction)
```

```
##     RawYear       ProductionKShortTon      Year
##  Min.   :194913   Min.   : 420423     Min.   :1949
##  1st Qu.:196563   1st Qu.: 558547     1st Qu.:1966
##  Median :198213   Median : 829700     Median :1982
##  Mean   :198213   Mean   : 796953     Mean   :1982
##  3rd Qu.:199863   3rd Qu.:1033239     3rd Qu.:1998
##  Max.   :201513   Max.   :1171809     Max.   :2015
```

Install the Quandl library and grab US Price data for coal from 1949-2005.

```r
library(Quandl)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```r
Prices <- Quandl("EPI/152")

summary(Prices)
```

```
##       Year            Price (U.S. Dollars)
##  Min.   :1949-01-01   Min.   :16.78
##  1st Qu.:1963-01-01   1st Qu.:20.19
##  Median :1977-01-01   Median :25.02
##  Mean   :1976-12-31   Mean   :27.85
##  3rd Qu.:1991-01-01   3rd Qu.:31.52
##  Max.   :2005-01-01   Max.   :50.92
```

Convert the information in the Year column of the Prices data frame to a numeric value and simplify the names of the columns.

```r
Prices$Year <- as.numeric(str_sub(Prices$Year,0,4))

#showing how to merge two dataframes . takes the column prices to the forth digit and taking the years

names(Prices) <- c("Year", "PriceShortTon")
```

Merge the data frames Prices and CoalProduction into a new data frame called CoalMarket by the colums named Year.

The quantity of US coal production is reported in thousand short ton whereas the prices are reported in US dollar (base year 2000) per short ton.

```r
summary(CoalProduction)
```

```
##     RawYear       ProductionKShortTon      Year
##  Min.   :194913   Min.   : 420423     Min.   :1949
##  1st Qu.:196563   1st Qu.: 558547     1st Qu.:1966
##  Median :198213   Median : 829700     Median :1982
##  Mean   :198213   Mean   : 796953     Mean   :1982
##  3rd Qu.:199863   3rd Qu.:1033239     3rd Qu.:1998
##  Max.   :201513   Max.   :1171809     Max.   :2015
```

```r
summary(Prices)
```

```
##       Year      PriceShortTon  
##  Min.   :1949   Min.   :16.78  
##  1st Qu.:1963   1st Qu.:20.19  
##  Median :1977   Median :25.02  
##  Mean   :1977   Mean   :27.85  
##  3rd Qu.:1991   3rd Qu.:31.52  
##  Max.   :2005   Max.   :50.92
```

```r
CoalMarket <- inner_join(Prices, CoalProduction, by ="Year")

summary(CoalMarket)
```

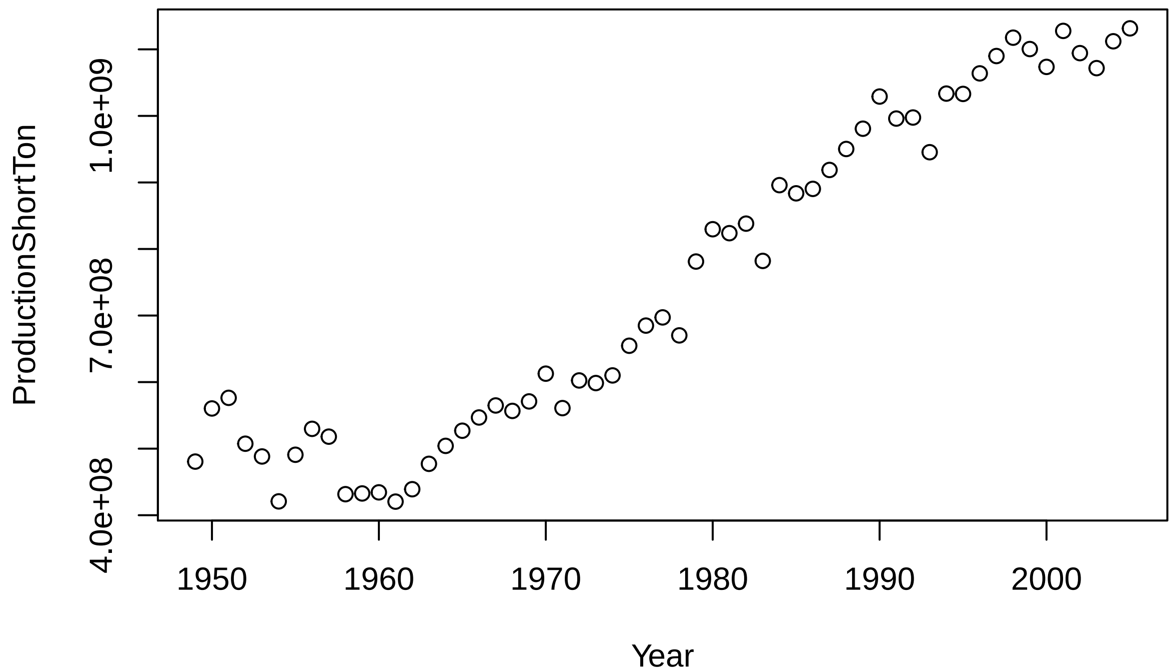
```
##       Year      PriceShortTon      RawYear       ProductionKShortTon
##  Min.   :1949   Min.   :16.78   Min.   :194913   Min.   : 420423    
##  1st Qu.:1963   1st Qu.:20.19   1st Qu.:196313   1st Qu.: 529774    
##  Median :1977   Median :25.02   Median :197713   Median : 684913    
##  Mean   :1977   Mean   :27.85   Mean   :197713   Mean   : 750201    
##  3rd Qu.:1991   3rd Qu.:31.52   3rd Qu.:199113   3rd Qu.: 995984    
##  Max.   :2005   Max.   :50.92   Max.   :200513   Max.   :1131498
```

Create new column called ProductionShortTon.

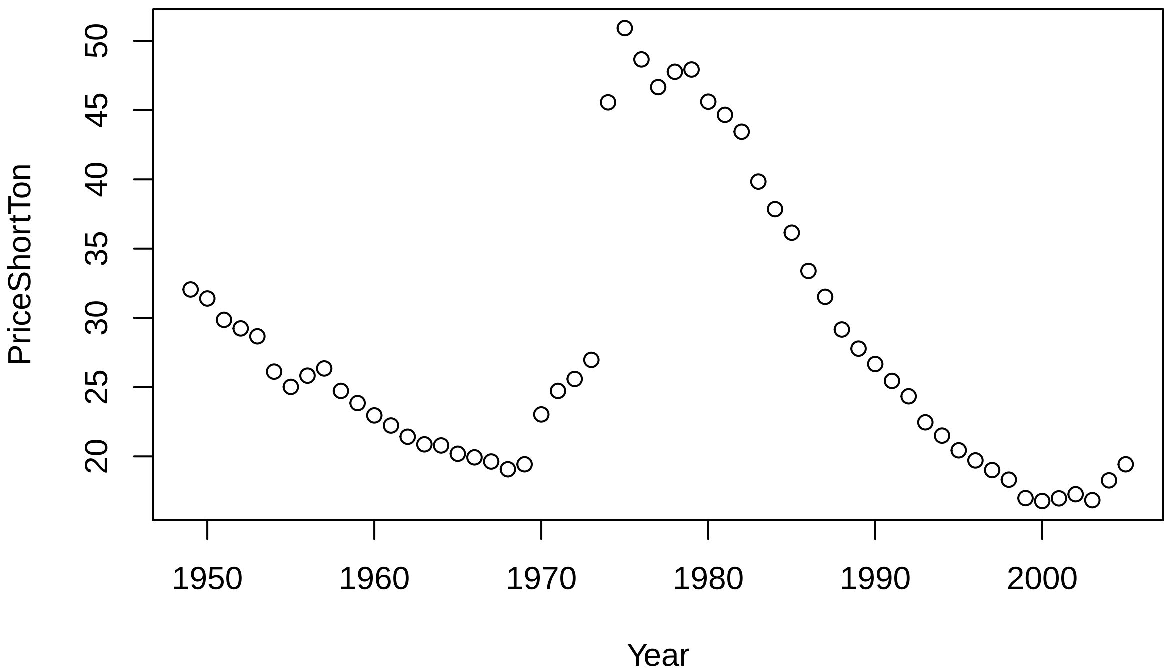
Convert the coal production quantity, which is reported in thousand short tons, into short tons by multiplying it with 1000 and assigning it to the new column ProductionShortTon.

Plot ProductionShortTon to Year, PriceShortTon to Year, and Prices to ProductionShortTon to observe trends or patterns to the data.

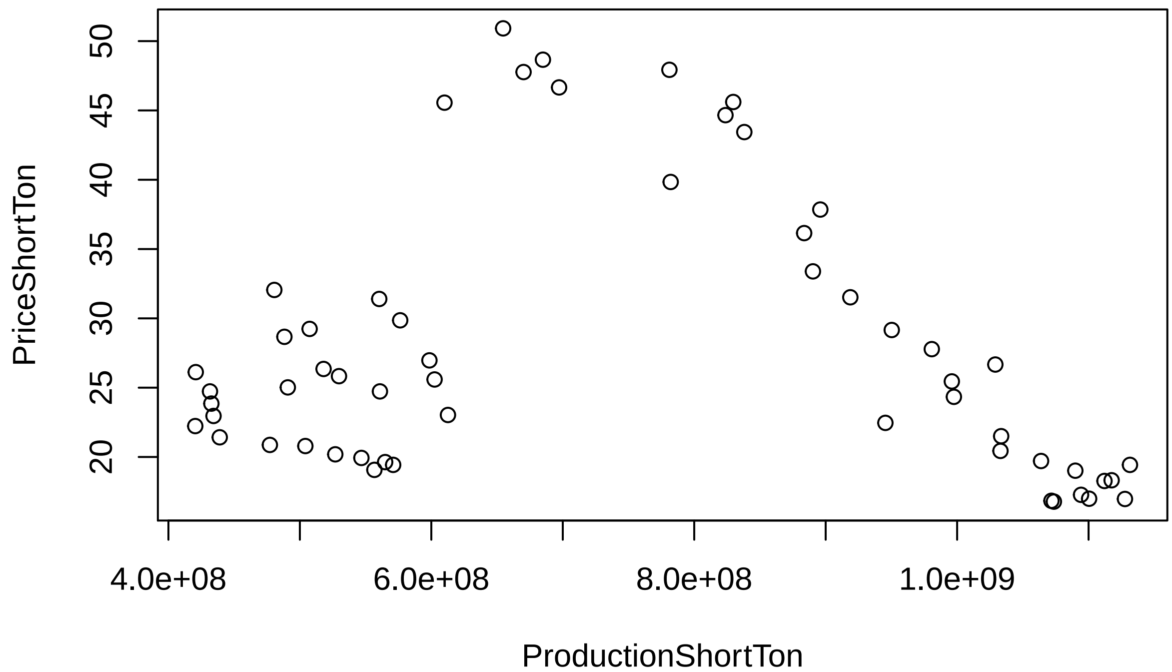
```r
CoalMarket <- CoalMarket  %>%mutate(ProductionShortTon = ProductionKShortTon * 1000)

plot(ProductionShortTon~Year, data = CoalMarket)
```




```
plot(PriceShortTon~Year, data = CoalMarket)
```




```
plot(PriceShortTon~ProductionShortTon, data = CoalMarket)
```



There is a sudden spike in coal prices in 1974.

Split data to before and after 1973 using a dummy variable called After73

```
CoalMarket$After73 <- FALSE

CoalMarket$After73[CoalMarket$Year > 1973] <- TRUE
```

Based on the scatter plot of Price vs Quantity, we notice that there is a downward sloping trend (for years after 1973). Thus, in this model, we assume that we have a demand curve. We can estimate the inverse demand by considering quantity (ProductionShortTon) as a function of Price (PriceShortTon) and they are linearly related.

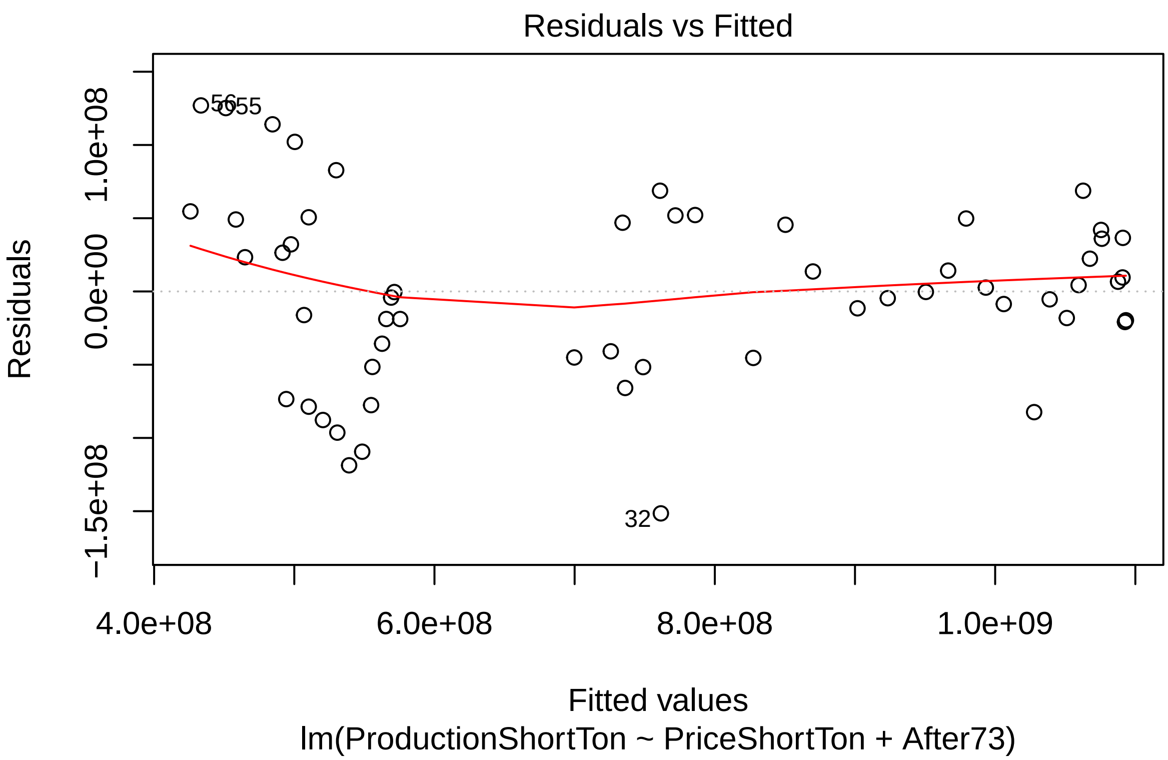
Run a linear regression on data for after 1973.

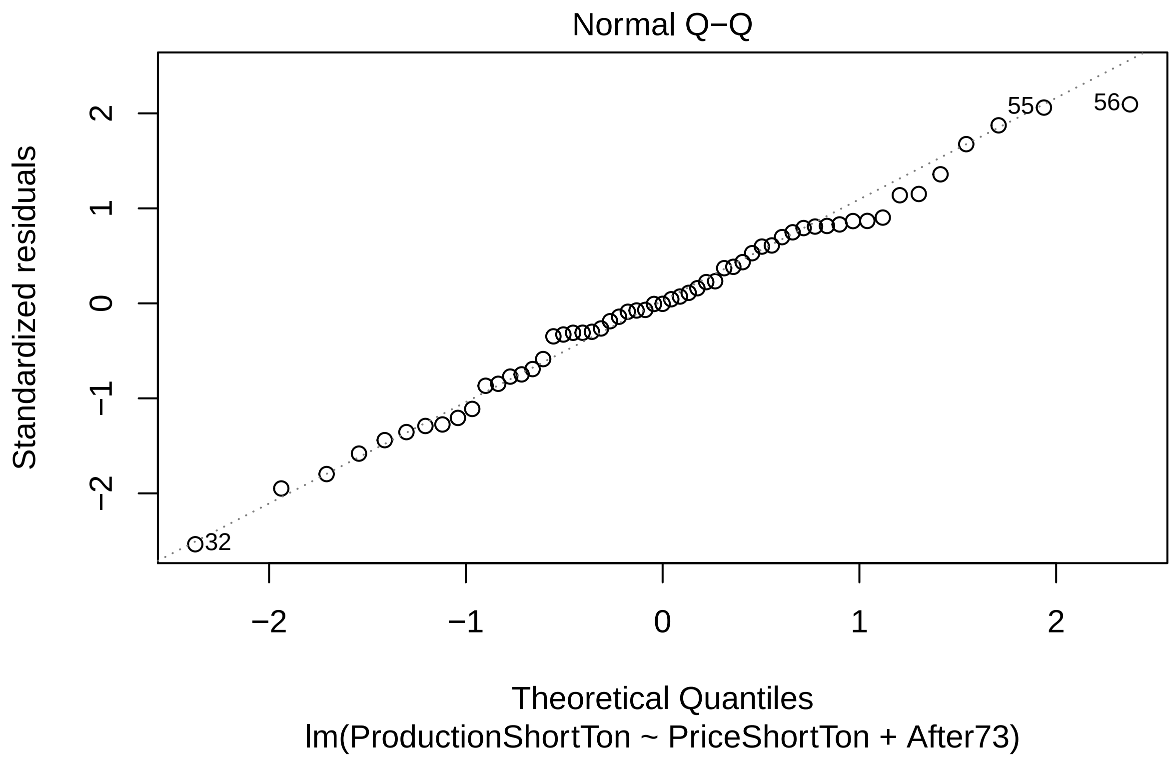
```
Regression <- lm(ProductionShortTon ~ PriceShortTon + After73, data = CoalMarket)

summary(Regression)
```

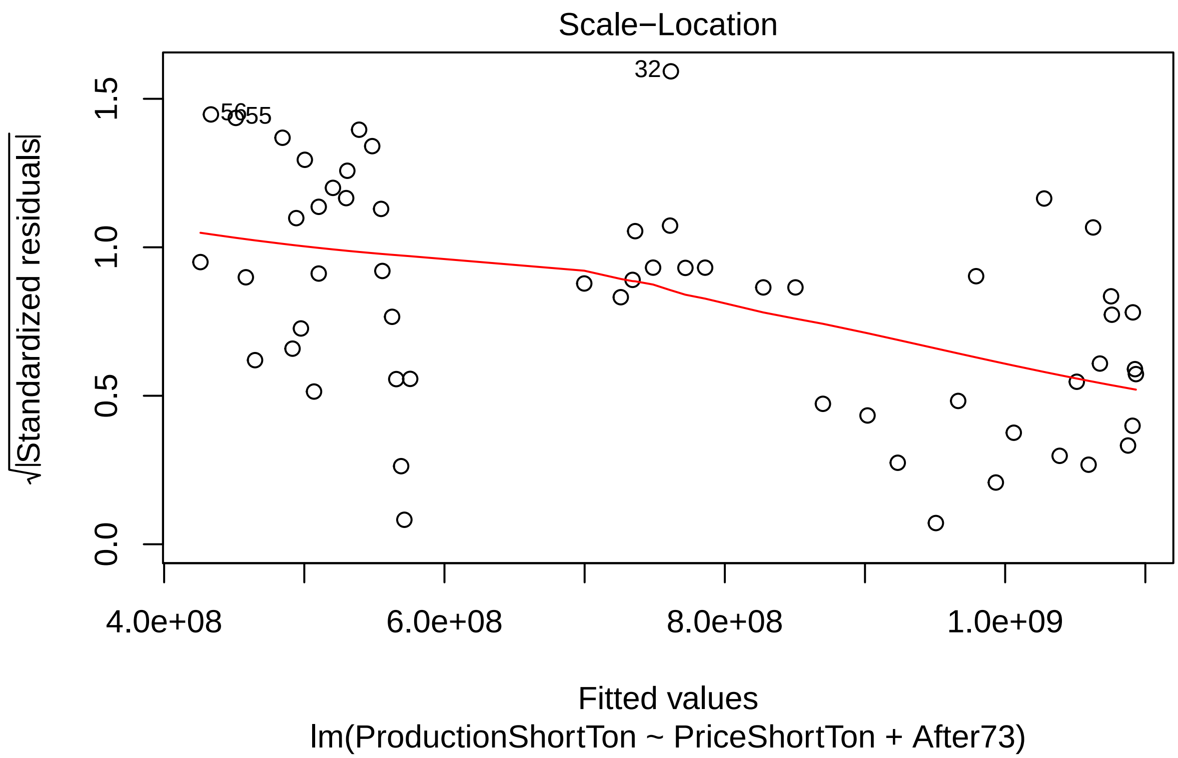
```
##
## Call:
## lm(formula = ProductionShortTon ~ PriceShortTon + After73, data = CoalMarket)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -151492580  -40865316    -311953   45523155  126989611
```

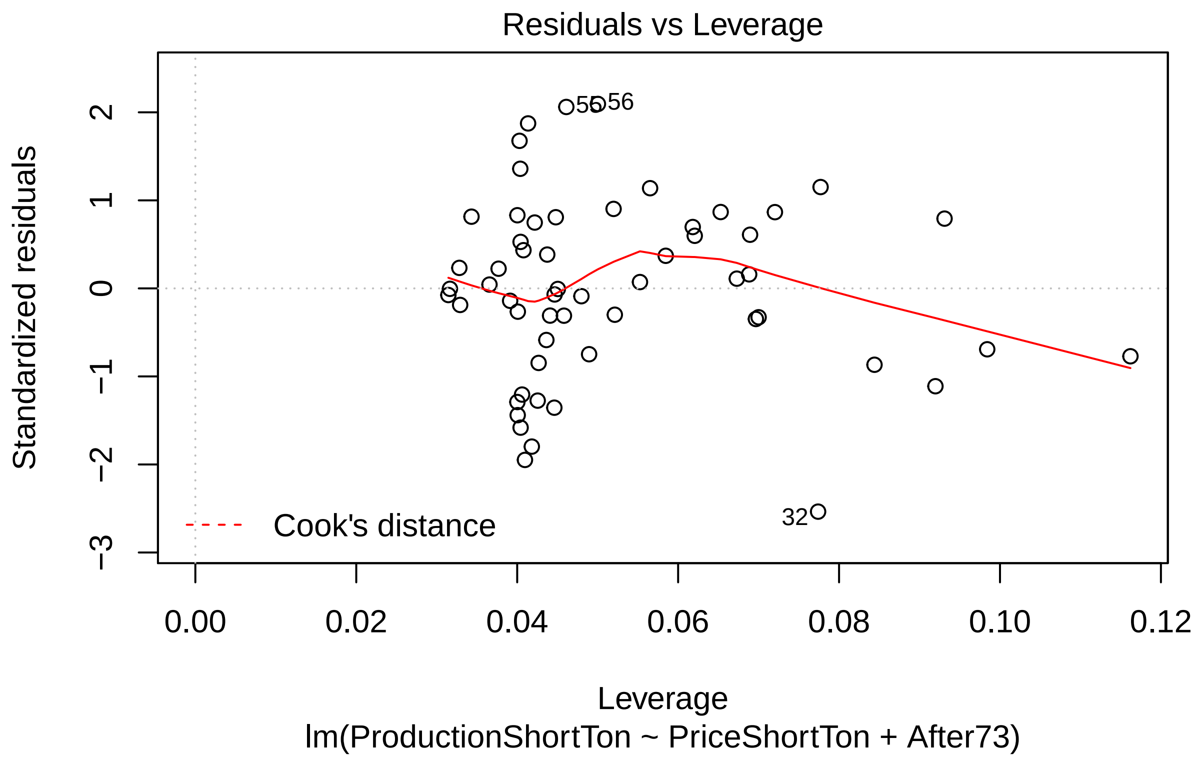
```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   795382282   25017637   31.79   <2e-16 ***
## PriceShortTon -11528150     889650  -12.96   <2e-16 ***
## After73TRUE   491355788   17478227   28.11   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62200000 on 54 degrees of freedom
## Multiple R-squared:  0.9375, Adjusted R-squared:  0.9352
## F-statistic: 404.8 on 2 and 54 DF,  p-value: < 2.2e-16
```

```r
plot(Regression)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(ProductionShortTon ~ PriceShortTon + After73)
−2
−1
0
1
2
32
55
56


Scale−Location
32
56
55
√|Standardized residuals|
0.0
0.5
1.0
1.5
4.0e+08
6.0e+08
8.0e+08
1.0e+09
Fitted values
lm(ProductionShortTon ~ PriceShortTon + After73)



The R squared value indicates how well the data fits the model. The closer it is to 1, the better the data fits.

Find elasticity through regression by using the log scale. The coefficients inform us of the elasticity of Coal demand.

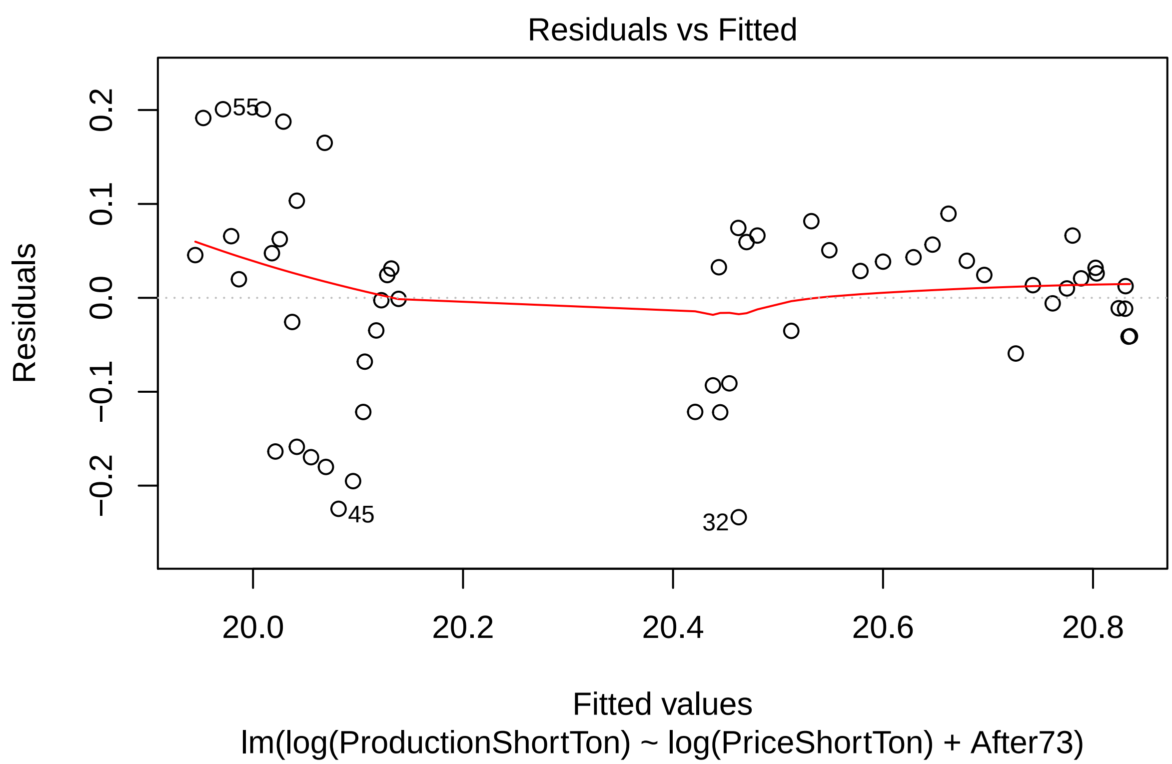
For each dollar increase in coal price, the demand for coal increases by the coefficient (in short tons).

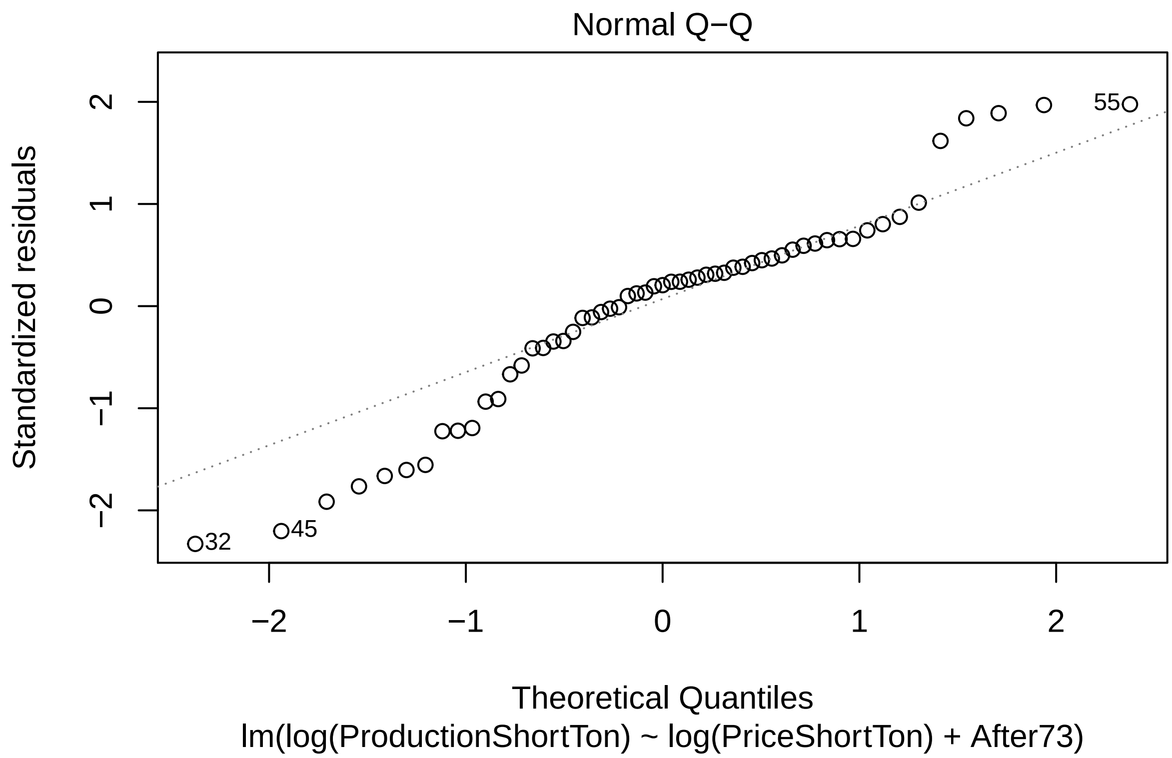
```
ElastForm <- lm(log(ProductionShortTon) ~ log(PriceShortTon) + After73, data=CoalMarket)

summary(ElastForm)
```

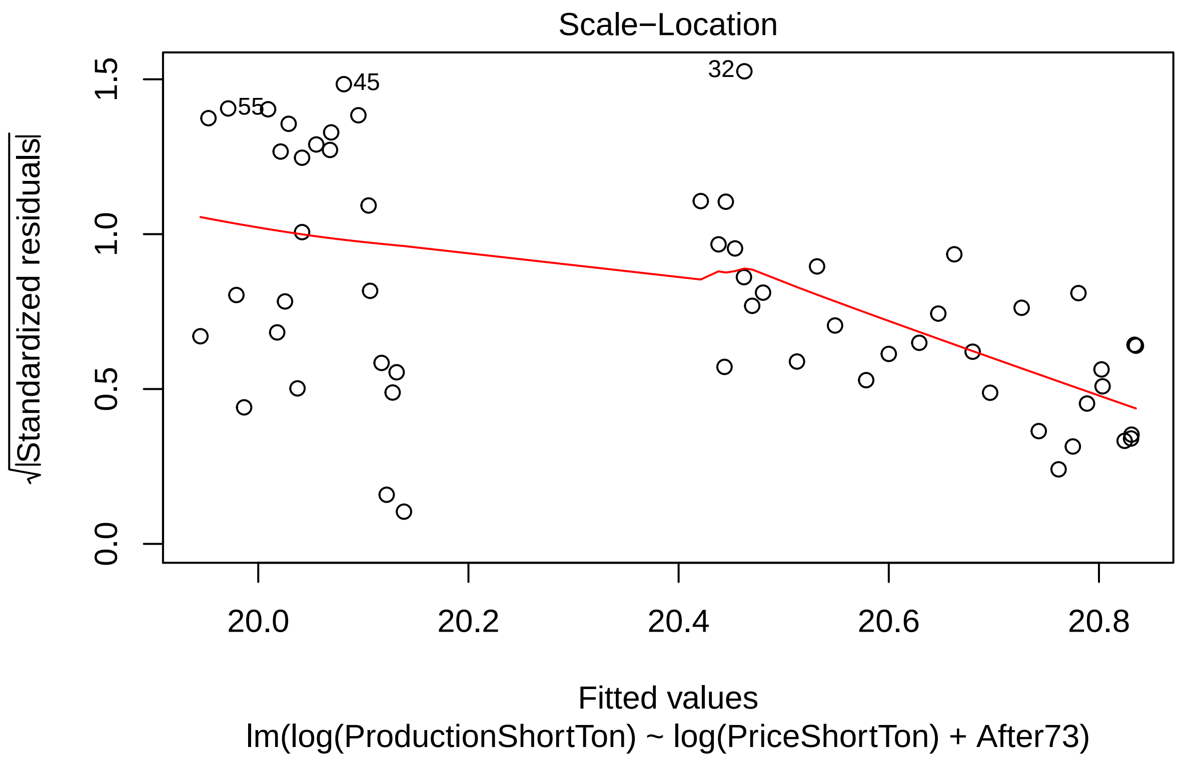
```
##
## Call:
## lm(formula = log(ProductionShortTon) ~ log(PriceShortTon) + After73,
##     data = CoalMarket)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23357 -0.04130  0.02074  0.05668  0.20083
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        21.23843    0.14211 149.452  < 2e-16 ***
## log(PriceShortTon) -0.37305    0.04417  -8.446 1.89e-11 ***
## After73TRUE         0.64882    0.02868  22.626  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

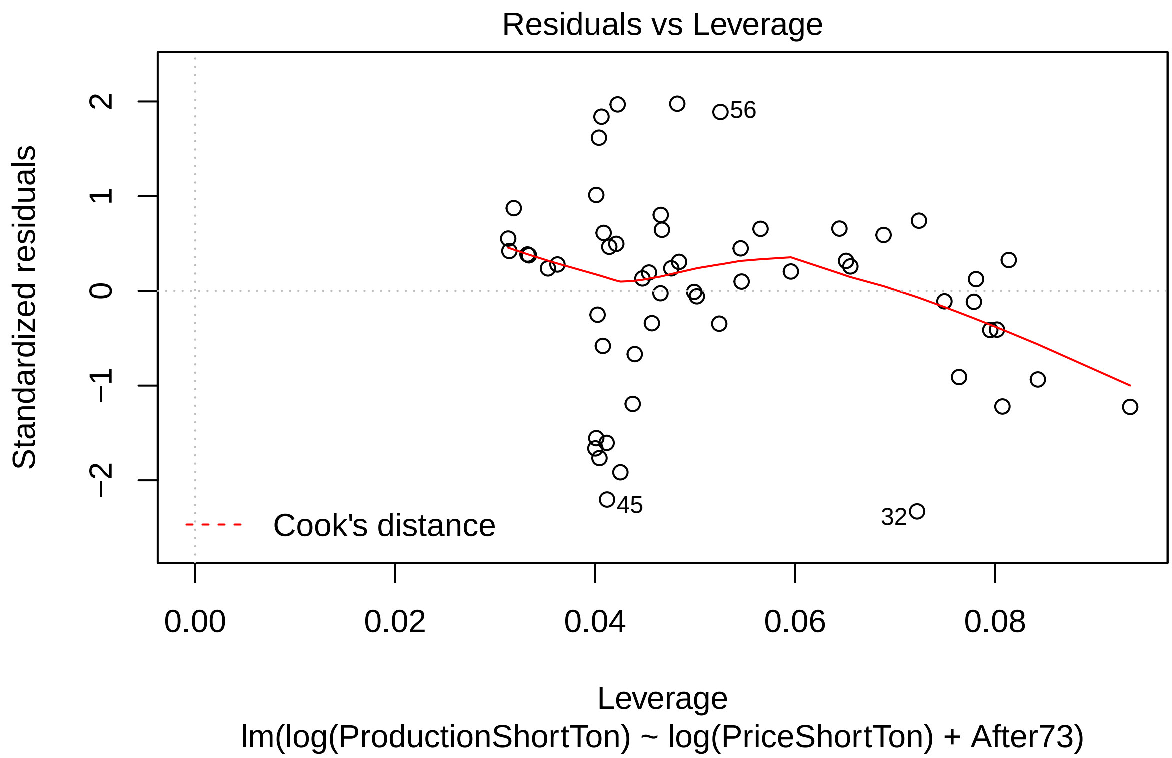
```
##
## Residual standard error: 0.1042 on 54 degrees of freedom
## Multiple R-squared:  0.9061, Adjusted R-squared:  0.9026
## F-statistic: 260.5 on 2 and 54 DF,  p-value: < 2.2e-16
```

```r
plot(ElastForm)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(ProductionShortTon) ~ log(PriceShortTon) + After73)
−2
−1
0
1
2
32
45
55


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20.0
20.2
20.4
20.6
20.8
55
45
32
Fitted values
lm(log(ProductionShortTon) ~ log(PriceShortTon) + After73)


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
56
45
32
Cook's distance
0.00
0.02
0.04
0.06
0.08
Leverage
lm(log(ProductionShortTon) ~ log(PriceShortTon) + After73)