Linear Regression Eric Johnson

library(fpp) library(fpp2) library(corrplot)

fuel summary(fuel)

```
library(fpp)
```

```
## Loading required package: forecast

## Warning: package 'forecast' was built under R version 4.0.5

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Loading required package: fma

## Warning: package 'fma' was built under R version 4.0.5

## Loading required package: expsmooth

## Warning: package 'expsmooth' was built under R version 4.0.5

## Loading required package: lmtest

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.5

##
## Attaching package: 'zoo'

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

## Loading required package: tseries
```

```
library(fpp2)
```

```
## -- Attaching packages --------------------------------------------------- fpp2 2.4 --
```

```
## v ggplot2 3.3.2

##

##
## Attaching package: 'fpp2'

## The following objects are masked from 'package:fpp':
##
##     ausair, ausbeer, austa, austourists, debitcards, departures,
##     elecequip, euretail, guinearice, oil, sunspotarea, usmelec
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(data.table)
library(faraway)
```

```
##
## Attaching package: 'faraway'

## The following objects are masked from 'package:fma':
##
##     airpass, eggs, ozone, wheat
```

```
library(car)
```

```
## Loading required package: carData

## Registered S3 methods overwritten by 'car':
##   method                          from
##   influence.merMod                lme4
##   cooks.distance.influence.merMod lme4
##   dfbeta.influence.merMod         lme4
##   dfbetas.influence.merMod        lme4

##
## Attaching package: 'car'

## The following objects are masked from 'package:faraway':
##
##     logit, vif
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following objects are masked from 'package:fma':
##
##     cement, housing, petrol
```

Load data and summary of data:

- plan is to predict Carbon based on other variables:

```
fuel
```

```
##                                       Model Cylinders Litres Barrels City
## 20                           Chevrolet Aveo         4    1.6    12.2   25
## 21                         Chevrolet Aveo 5         4    1.6    12.2   25
## 27                         Chevrolet Cobalt         4    2.2    12.7   24
## 120                  Chevrolet Colorado 2WD         4    2.9    17.1   18
## 127                  Chevrolet Colorado 2WD         5    3.7    18.0   17
## 133 Chevrolet Colorado Cab Chassis inc 2WD         5    3.7    20.1   15
## 121         Chevrolet Colorado Crew Cab 2WD         4    2.9    17.1   18
## 128         Chevrolet Colorado Crew Cab 2WD         5    3.7    18.0   17
## 95                        Chevrolet HHR FWD         4    2.0    14.9   19
## 96                  Chevrolet HHR Panel FWD         4    2.0    14.9   19
## 43                         Chevrolet Malibu         4    2.4    13.2   22
## 48                         Chevrolet Malibu         4    2.4    13.7   22
## 15                  Chevrolet Malibu Hybrid         4    2.4    11.8   26
## 113                     Chrysler PT Cruiser         4    2.4    16.3   19
## 122                     Chrysler PT Cruiser         4    2.4    16.3   18
## 60                        Chrysler Sebring         4    2.4    14.3   21
## 74            Chrysler Sebring Convertible         4    2.4    14.9   20
## 61                           Dodge Avenger         4    2.4    14.3   21
## 40                           Dodge Caliber         4    2.0    14.3   23
## 71                           Dodge Caliber         4    2.4    14.9   21
## 110                      Dodge Journey 2WD         4    2.4    16.3   19
## 79                        Ford Escape FWD         4    2.5    14.9   20
## 4                  Ford Escape Hybrid FWD         4    2.5    10.7   34
## 28                         Ford Focus FWD         4    2.0    12.7   24
## 80                        Ford Fusion FWD         4    2.3    14.9   20
## 114                 Ford Ranger Pickup 2WD         4    2.3    16.3   19
## 123                        GMC Canyon 2WD         4    2.9    17.1   18
## 129                        GMC Canyon 2WD         5    3.7    18.0   17
## 134         GMC Canyon Cab Chassis Inc 2WD         5    3.7    20.1   15
## 124               GMC Canyon Crew Cab 2WD         4    2.9    17.1   18
## 130               GMC Canyon Crew Cab 2WD         5    3.7    18.0   17
## 62                           Honda Accord         4    2.4    14.3   21
## 63                     Honda Accord Coupe         4    2.4    14.3   21
## 19                            Honda Civic         4    1.8    11.8   25
## 2                      Honda Civic Hybrid         4    1.3     8.2   40
## 88                         Honda CR-V 2WD         4    2.4    14.9   20
## 91                      Honda Element 2WD         4    2.4    15.6   20
## 11                              Honda Fit         4    1.5    11.4   27
## 9                               Honda Fit         4    1.5    11.0   28
## 13                         Hyundai Accent         4    1.6    11.8   26
## 24                        Hyundai Elantra         4    2.0    12.2   25
## 39                Hyundai Elantra Touring         4    2.0    13.2   23
## 46                         Hyundai Sonata         4    2.4    13.7   22
## 92                     Hyundai Tucson 2WD         4    2.0    15.6   20
## 41                       Jeep Compass 2WD         4    2.0    14.3   23
## 72                       Jeep Compass 2WD         4    2.4    14.9   21
## 42                       Jeep Patriot 2WD         4    2.0    14.3   23
```

```
## 73                  Jeep Patriot 2WD          4    2.4    14.9    21
## 47                        Kia Optima          4    2.4    13.7    22
## 14                           Kia Rio          4    1.6    11.4    26
## 89                         Kia Rondo          4    2.4    15.6    20
## 30                       Kia Spectra          4    2.0    12.7    24
## 93                 Kia Sportage 2WD          4    2.0    15.6    20
## 49                           Mazda 3          4    2.0    13.7    22
## 54                           Mazda 3          4    2.3    14.3    22
## 68                           Mazda 5          4    2.3    14.9    21
## 64                           Mazda 6          4    2.5    14.3    21
## 115                 Mazda B2300 2WD          4    2.3    16.3    19
## 81                Mazda Tribute FWD          4    2.5    14.9    20
## 5          Mazda Tribute Hybrid 2WD          4    2.5    10.7    34
## 82              Mercury Mariner FWD          4    2.5    14.9    20
## 6        Mercury Mariner Hybrid FWD          4    2.5    10.7    34
## 83                    Mercury Milan          4    2.3    14.9    20
## 107              Mitsubishi Eclipse          4    2.4    15.6    19
## 108       Mitsubishi Eclipse Spyder          4    2.4    15.6    19
## 90                Mitsubishi Galant          4    2.4    14.9    20
## 55                Mitsubishi Lancer          4    2.0    14.3    22
## 69                Mitsubishi Lancer          4    2.4    14.9    21
## 70      Mitsubishi Lancer Sportback          4    2.4    14.9    21
## 94         Mitsubishi Outlander 2WD          4    2.4    15.6    20
## 36                    Nissan Altima          4    2.5    13.2    23
## 37              Nissan Altima Coupe          4    2.5    13.2    23
## 3              Nissan Altima Hybrid          4    2.5    10.1    35
## 131             Nissan Frontier 2WD          4    2.5    18.0    17
## 58                Nissan Rogue FWD           4    2.5    14.3    22
## 32                    Nissan Sentra          4    2.5    13.2    24
## 12                     Nissan Versa          4    1.8    11.8    27
## 31                     Nissan Versa          4    1.8    12.7    24
## 22                  Pontiac G3 Wave          4    1.6    12.2    25
## 23                Pontiac G3 Wave 5          4    1.6    12.2    25
## 29                       Pontiac G5          4    2.2    12.7    24
## 35                    Pontiac G5 GT          4    2.2    13.2    23
## 44                       Pontiac G6          4    2.4    13.2    22
## 50                       Pontiac G6          4    2.4    13.7    22
## 102                Pontiac Solstice          4    2.0    16.3    19
## 116                Pontiac Solstice          4    2.4    16.3    19
## 18                     Pontiac Vibe          4    1.8    12.2    26
## 65                     Pontiac Vibe          4    2.4    14.3    21
## 103            Saab 9-3 Convertible          4    2.0    15.6    19
## 97             Saab 9-3 Sport Sedan          4    2.0    15.6    19
## 104             Saab 9-3 SportCombi          4    2.0    15.6    19
## 125                  Saab 9-5 Sedan          4    2.3    17.1    17
## 126             Saab 9-5 SportCombi          4    2.3    17.1    17
## 33       Saturn Astra 2DR Hatchback          4    1.8    12.7    24
## 34       Saturn Astra 4DR Hatchback          4    1.8    12.7    24
## 45                      Saturn Aura          4    2.4    13.2    22
## 16               Saturn Aura Hybrid          4    2.4    11.8    26
## 105                      Saturn SKY          4    2.0    16.3    19
## 117                      Saturn SKY          4    2.4    16.3    19
## 109                  Saturn Vue FWD          4    2.4    15.6    19
## 25                Saturn Vue Hybrid          4    2.4    12.2    25
```

```
## 66                           Scion tC         4    2.4    14.3   21
## 56                           Scion xB         4    2.4    14.3   22
## 17                           Scion xD         4    1.8    12.2   26
## 132                Suzuki Equator 2WD         4    2.5    18.0   17
## 111               Suzuki Grand Vitara         4    2.4    16.3   19
## 51                         Suzuki SX4         4    2.0    13.7   22
## 38                   Suzuki SX4 Sedan         4    2.0    13.2   23
## 52                   Suzuki SX4 Sport         4    2.0    13.7   22
## 59                       Toyota Camry         4    2.4    13.7   21
## 7                 Toyota Camry Hybrid         4    2.4    10.1   33
## 10                     Toyota Corolla         4    1.8    11.4   27
## 53                     Toyota Corolla         4    2.4    13.7   22
## 26                      Toyota Matrix         4    1.8    12.2   25
## 67                      Toyota Matrix         4    2.4    14.3   21
## 1                        Toyota Prius         4    1.5     7.4   48
## 57                    Toyota RAV4 2WD         4    2.5    14.3   22
## 112                 Toyota Tacoma 2WD         4    2.7    16.3   19
## 8                        Toyota Yaris         4    1.5    11.0   29
## 75                  Volkswagen Jetta          5    2.5    14.3   20
## 76        Volkswagen Jetta SportWagen         5    2.5    14.3   20
## 77               Volkswagen New Beetle        5    2.5    14.9   20
## 84   Volkswagen New Beetle Convertible        5    2.5    14.9   20
## 78                  Volkswagen Rabbit         5    2.5    14.3   20
## 85                      Volvo C30 FWD         5    2.4    14.9   20
## 98                      Volvo C30 FWD         5    2.5    14.9   19
## 118             Volvo C70 Convertible         5    2.5    16.3   18
## 86                      Volvo S40 FWD         5    2.4    14.9   20
## 99                      Volvo S40 FWD         5    2.5    14.9   19
## 119                     Volvo S60 FWD         5    2.4    16.3   18
## 100                     Volvo S60 FWD         5    2.5    14.9   19
## 106                     Volvo S60 FWD         5    2.5    15.6   19
## 87                      Volvo V50 FWD         5    2.4    14.9   20
## 101                     Volvo V50 FWD         5    2.5    14.9   19
##     Highway Cost Carbon
## 20       34 1012    6.6
## 21       34 1012    6.6
## 27       33 1049    6.8
## 120      24 1418    9.2
## 127      23 1491    9.6
## 133      20 1667   10.8
## 121      24 1418    9.2
## 128      23 1491    9.6
## 95       29 1233    8.0
## 96       29 1233    8.0
## 43       33 1091    7.1
## 48       30 1134    7.3
## 15       34  978    6.3
## 113      24 1349    8.7
## 122      24 1349    8.7
## 60       30 1182    7.7
## 74       29 1233    8.0
## 61       30 1182    7.7
## 40       27 1182    7.7
## 71       25 1233    8.0
```

```
## 110       25 1349    8.7
## 79        28 1233    8.0
## 4         31  887    5.7
## 28        33 1049    6.8
## 80        28 1233    8.0
## 114       24 1349    8.7
## 123       24 1418    9.2
## 129       23 1491    9.6
## 134       20 1667   10.8
## 124       24 1418    9.2
## 130       23 1491    9.6
## 62        30 1182    7.7
## 63        30 1182    7.7
## 19        36  978    6.3
## 2         45  675    4.4
## 88        27 1233    8.0
## 91        25 1290    8.3
## 11        33  944    6.1
## 9         35  916    5.9
## 13        35  978    6.3
## 24        33 1012    6.6
## 39        30 1091    7.1
## 46        32 1134    7.3
## 92        25 1290    8.3
## 41        27 1182    7.7
## 72        25 1233    8.0
## 42        27 1182    7.7
## 73        25 1233    8.0
## 47        32 1134    7.3
## 14        35  944    6.1
## 89        27 1290    8.3
## 30        32 1049    6.8
## 93        25 1290    8.3
## 49        30 1134    7.3
## 54        28 1182    7.7
## 68        27 1233    8.0
## 64        30 1182    7.7
## 115       24 1349    8.7
## 81        28 1233    8.0
## 5         31  887    5.7
## 82        28 1233    8.0
## 6         31  887    5.7
## 83        28 1233    8.0
## 107       26 1290    8.3
## 108       26 1290    8.3
## 90        27 1233    8.0
## 55        28 1182    7.7
## 69        27 1233    8.0
## 70        27 1233    8.0
## 94        25 1290    8.3
## 36        31 1091    7.1
## 37        31 1091    7.1
## 3         33  833    5.4
## 131       22 1491    9.6
```

```
## 58        27 1182    7.7
## 32        30 1091    7.1
## 12        33  978    6.3
## 31        32 1049    6.8
## 22        34 1012    6.6
## 23        34 1012    6.6
## 29        33 1049    6.8
## 35        32 1091    7.1
## 44        33 1091    7.1
## 50        30 1134    7.3
## 102       27 1349    8.7
## 116       24 1349    8.7
## 18        31 1012    6.6
## 65        29 1182    7.7
## 103       27 1290    8.3
## 97        28 1290    8.3
## 104       27 1290    8.3
## 125       27 1418    9.2
## 126       27 1418    9.2
## 33        30 1049    6.8
## 34        30 1049    6.8
## 45        33 1091    7.1
## 16        34  978    6.3
## 105       27 1349    8.7
## 117       24 1349    8.7
## 109       26 1290    8.3
## 25        32 1012    6.6
## 66        29 1182    7.7
## 56        28 1182    7.7
## 17        32 1012    6.6
## 132       22 1491    9.6
## 111       25 1349    8.7
## 51        30 1134    7.3
## 38        31 1091    7.1
## 52        30 1134    7.3
## 59        31 1134    7.3
## 7         34  833    5.4
## 10        35  944    6.1
## 53        30 1134    7.3
## 26        31 1012    6.6
## 67        29 1182    7.7
## 1         45  615    4.0
## 57        28 1182    7.7
## 112       25 1349    8.7
## 8         35  916    5.9
## 75        29 1182    7.7
## 76        29 1182    7.7
## 77        29 1233    8.0
## 84        28 1233    8.0
## 78        29 1182    7.7
## 85        28 1233    8.0
## 98        28 1233    8.0
## 118       26 1349    8.7
## 86        28 1233    8.0
```

```
## 99       28 1233    8.0
## 119      26 1349    8.7
## 100      28 1233    8.0
## 106      27 1290    8.3
## 87       28 1233    8.0
## 101      28 1233    8.0
```

```
summary(fuel)
```

```
##                              Model      Cylinders        Litres    
##  Volvo S60 FWD                  :  3   Min.   :4.000   Min.   :1.30  
##  Chevrolet Colorado 2WD         :  2   1st Qu.:4.000   1st Qu.:2.00  
##  Chevrolet Colorado Crew Cab 2WD:  2   Median :4.000   Median :2.40  
##  Chevrolet Malibu               :  2   Mean   :4.157   Mean   :2.31  
##  Chrysler PT Cruiser            :  2   3rd Qu.:4.000   3rd Qu.:2.50  
##  Dodge Caliber                  :  2   Max.   :5.000   Max.   :3.70  
##  (Other)                        :121                                 
##     Barrels           City          Highway          Cost     
##  Min.   : 7.40   Min.   :15.00   Min.   :20.00   Min.   : 615  
##  1st Qu.:13.20   1st Qu.:19.00   1st Qu.:27.00   1st Qu.:1091  
##  Median :14.30   Median :21.00   Median :28.00   Median :1182  
##  Mean   :14.31   Mean   :21.97   Mean   :28.89   Mean   :1185  
##  3rd Qu.:15.60   3rd Qu.:23.75   3rd Qu.:31.00   3rd Qu.:1290  
##  Max.   :20.10   Max.   :48.00   Max.   :45.00   Max.   :1667  
##                                                                 
##      Carbon      
##  Min.   : 4.000  
##  1st Qu.: 7.100  
##  Median : 7.700  
##  Mean   : 7.671  
##  3rd Qu.: 8.300  
##  Max.   :10.800  
## 
```

Checking for correlation and visualizing data

```
library(GGally)
```

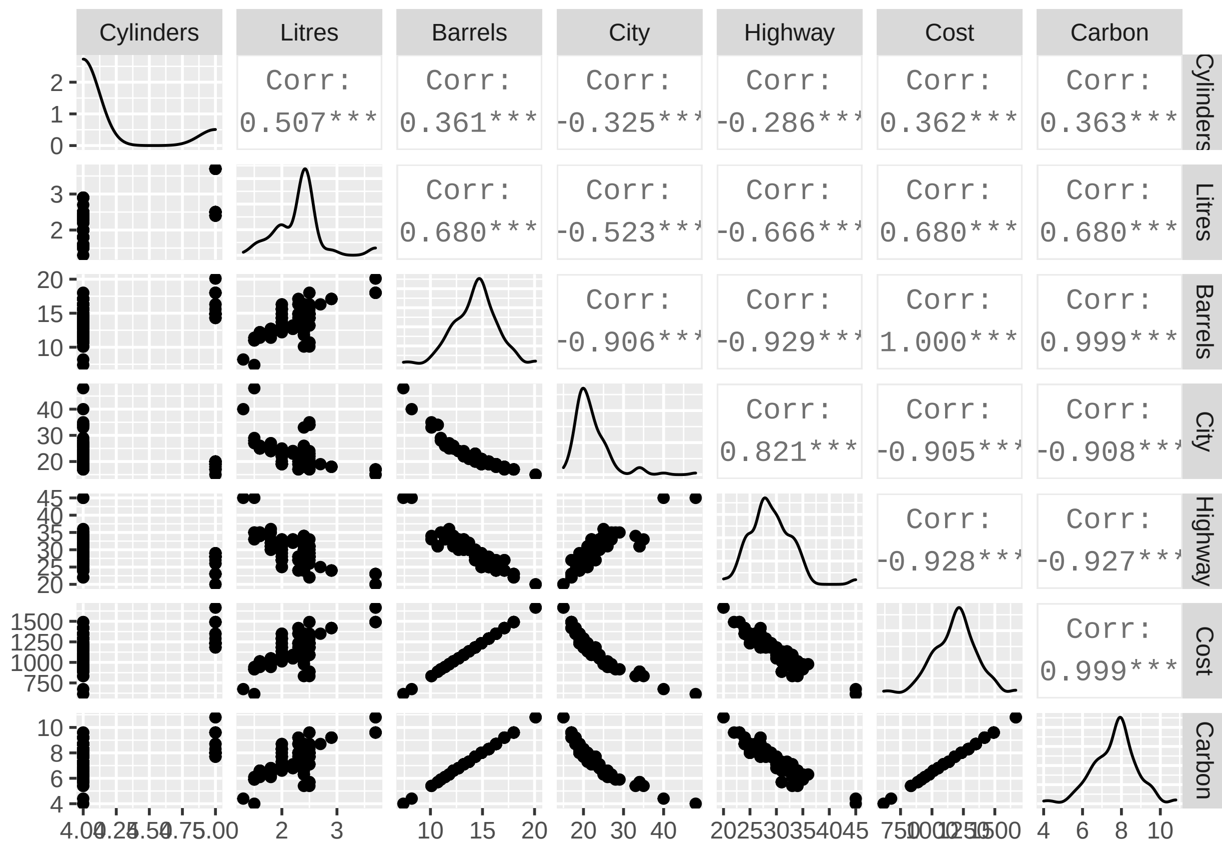
```
## Registered S3 method overwritten by 'GGally':
##   method from   
##   +.gg   ggplot2

## 
## Attaching package: 'GGally'

## The following object is masked from 'package:faraway':
## 
##     happy

## The following object is masked from 'package:fma':
## 
##     pigs
```

```
?ggpairs
```

```
## starting httpd help server ...
```

```
## done
```

```
ggpairs(fuel,c(2:8))
```




```
#lower = list(mapping = aes(color = Quarter), continuous = 'smooth', combo = 'facetdensity'))
```

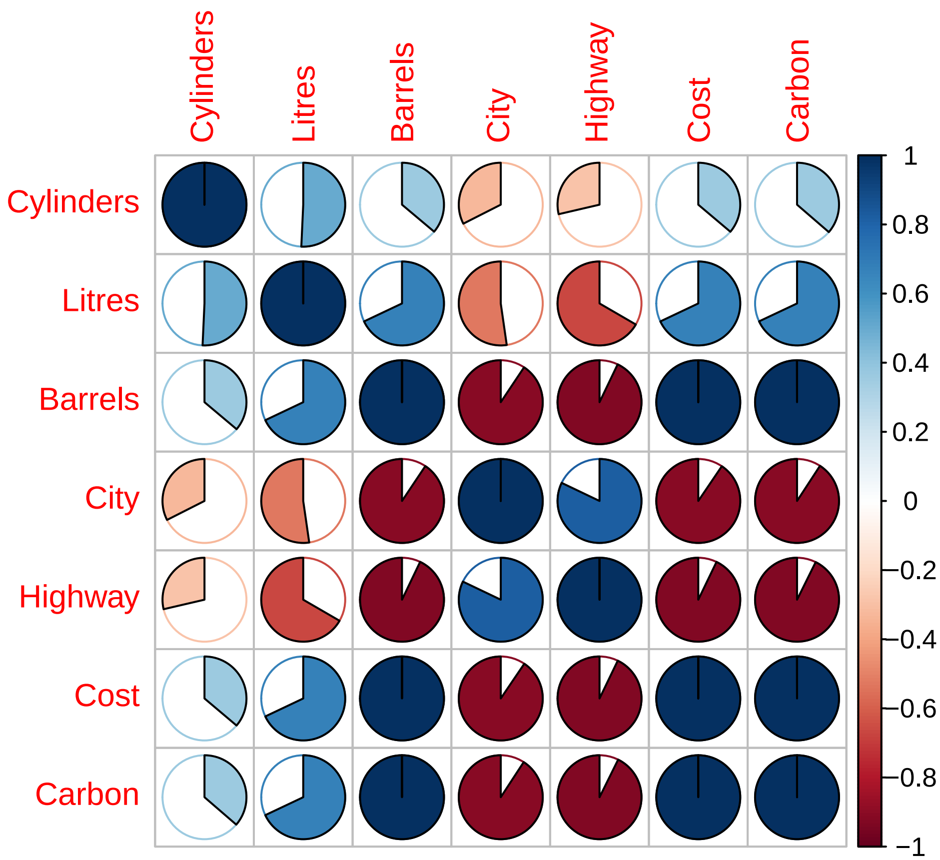
- above we see there are strong correlations with Carbon and Barrels, City, Highway, and Cost. We see weaker correlations between Carbon and Cylinders and Litres. There is likely colinearity with City and Highway when it comes to determining Carbon; probably only one of those variables is needed to predict Carbon. Further the use of the Cost or Barrels variable would be redundant since clearly as the mpg of the vehicle increases, the cost or barrels would decrease. We will use the City variable to predict Carbon on it's own.
- Further with the relationship of City and Carbon, we can see from the plot the relationship is a slightly curved line indicating an exponential decay. This is non-linear and the data needs transformation to be linear. We will use the natural log of both the independent variable (City) and the reponse variable (Carbon) for constructing our model.

```r
f1 <- fuel
fuelmat <- f1[,c(2:8)]

cor(fuelmat)
```

Correlation matrix and plot:

```
##             Cylinders     Litres    Barrels       City    Highway       Cost
## Cylinders  1.0000000  0.5072599  0.3612070 -0.3246731 -0.2863089  0.3615217
## Litres     0.5072599  1.0000000  0.6795817 -0.5228023 -0.6663893  0.6796274
## Barrels    0.3612070  0.6795817  1.0000000 -0.9062413 -0.9288892  0.9999567
## City      -0.3246731 -0.5228023 -0.9062413  1.0000000  0.8209571 -0.9052234
## Highway   -0.2863089 -0.6663893 -0.9288892  0.8209571  1.0000000 -0.9284566
## Cost       0.3615217  0.6796274  0.9999567 -0.9052234 -0.9284566  1.0000000
## Carbon     0.3633336  0.6798169  0.9994928 -0.9079411 -0.9267689  0.9994937
##               Carbon
## Cylinders  0.3633336
## Litres     0.6798169
## Barrels    0.9994928
## City      -0.9079411
## Highway   -0.9267689
## Cost       0.9994937
## Carbon     1.0000000
```

```r
fuelmat <- cor(fuelmat)
corrplot(fuelmat, method = 'pie')
```



- The above is a correlation matrix and plot to show the strength of relationships of each variable.

fitting a model to predict Carbon

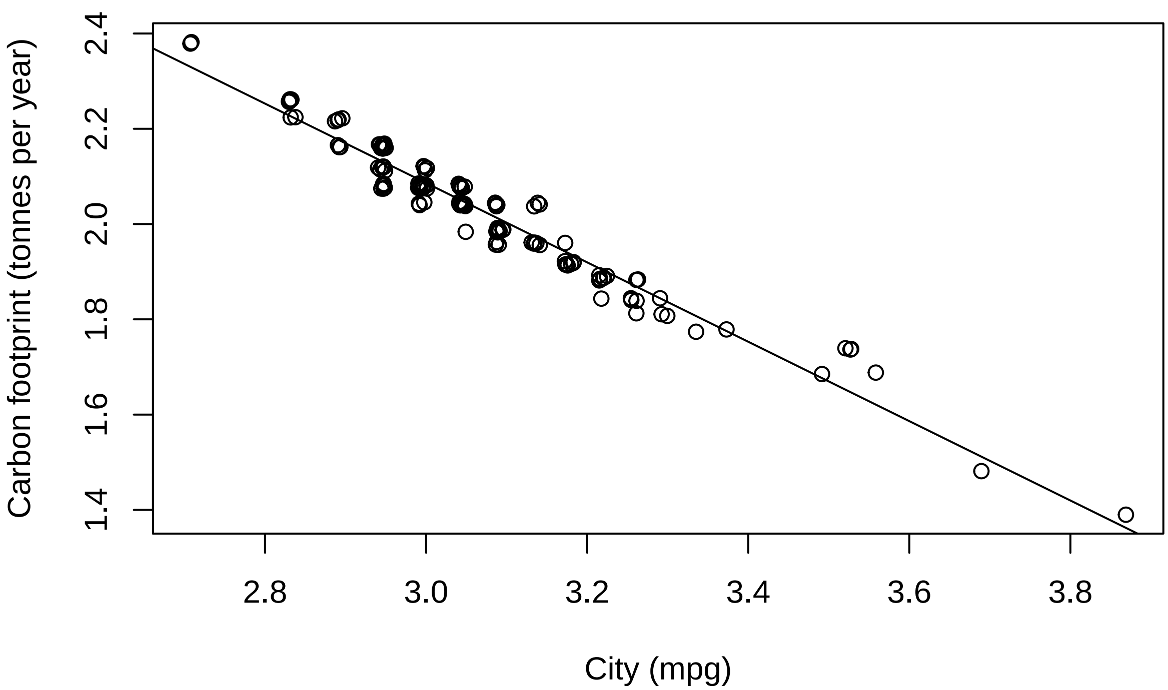
```
cfit <- lm(log(Carbon) ~ log(City), data = fuel)
summary(cfit)
```

```
##
## Call:
## lm(formula = log(Carbon) ~ log(City), data = fuel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06330 -0.02093 -0.01033  0.03070  0.09282
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.58581    0.05043   90.94   <2e-16 ***
## log(City)   -0.83320    0.01639  -50.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03408 on 132 degrees of freedom
```

```
## Multiple R-squared:  0.9514, Adjusted R-squared:  0.9511
## F-statistic:  2585 on 1 and 132 DF,  p-value: < 2.2e-16
```

- the above summary of the model shows a high F-statistic with a low p-value indicating the model can be used to predict Carbon.
- below is a plot of Carbon and City after transformation of the data to form a linear relationship along with a superimposed line generated from the fitted data of the model; we can see the model is overall a good fit:

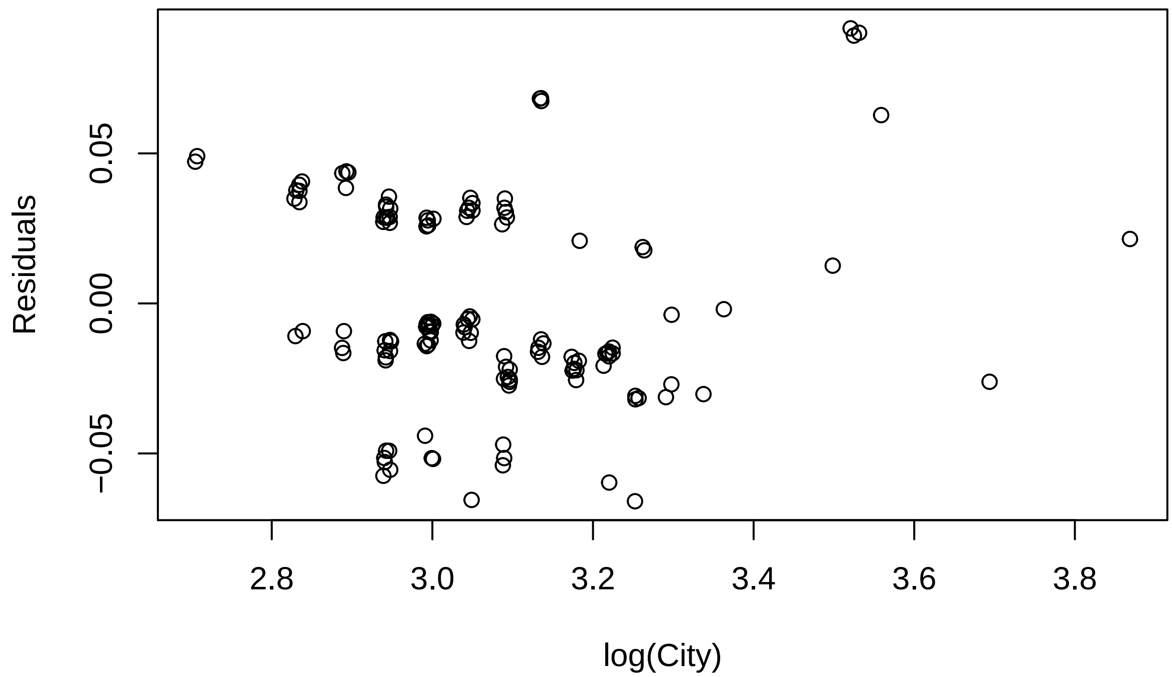
```
plot(jitter(log(Carbon)) ~ jitter(log(City)), xlab="City (mpg)",
ylab="Carbon footprint (tonnes per year)", data=fuel)
abline(cfit)
```



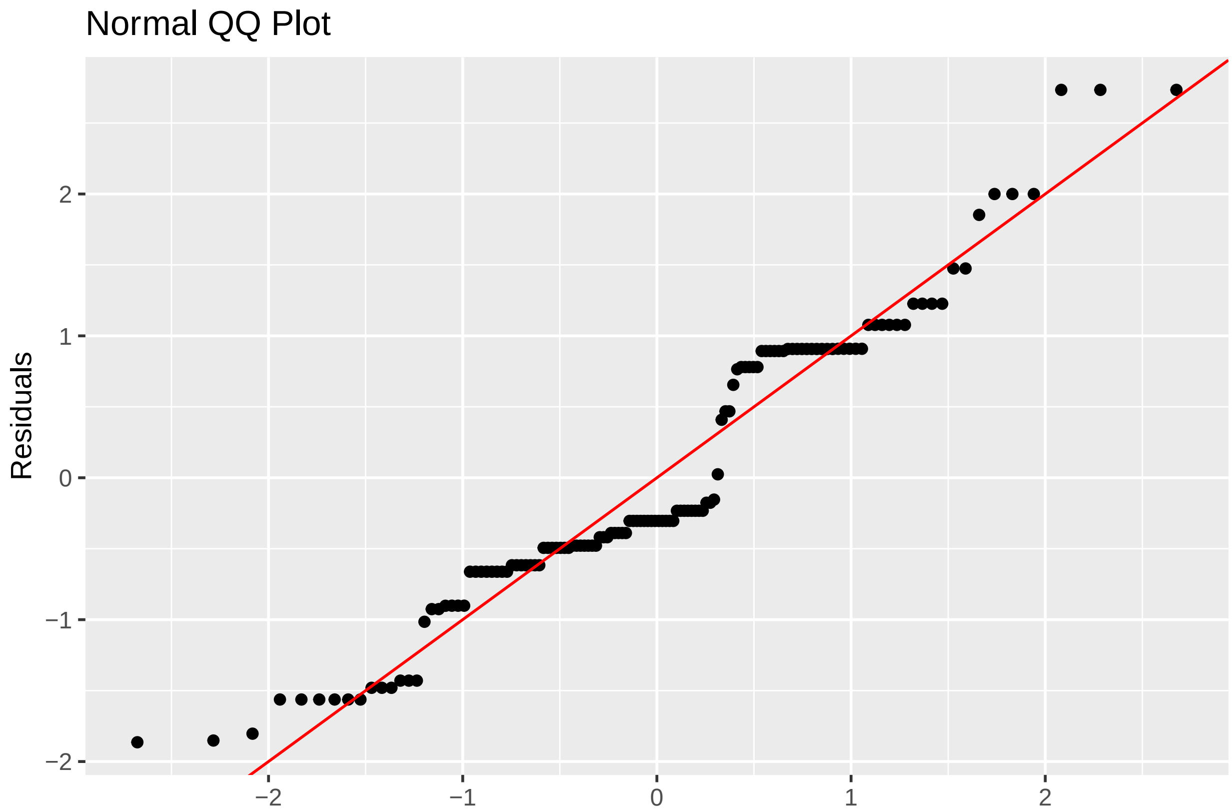
Let's check the residuals:

- the plots below show a fairly good fit of the model and there are no obvious patterns. The QQ-plot reveals a slightly longer right-tail so there is not a normal distribution.

```
res <- residuals(cfit)

plot(jitter(res, amount=.005) ~ jitter(log(City)),
 ylab="Residuals", xlab="log(City)", data=fuel)
```

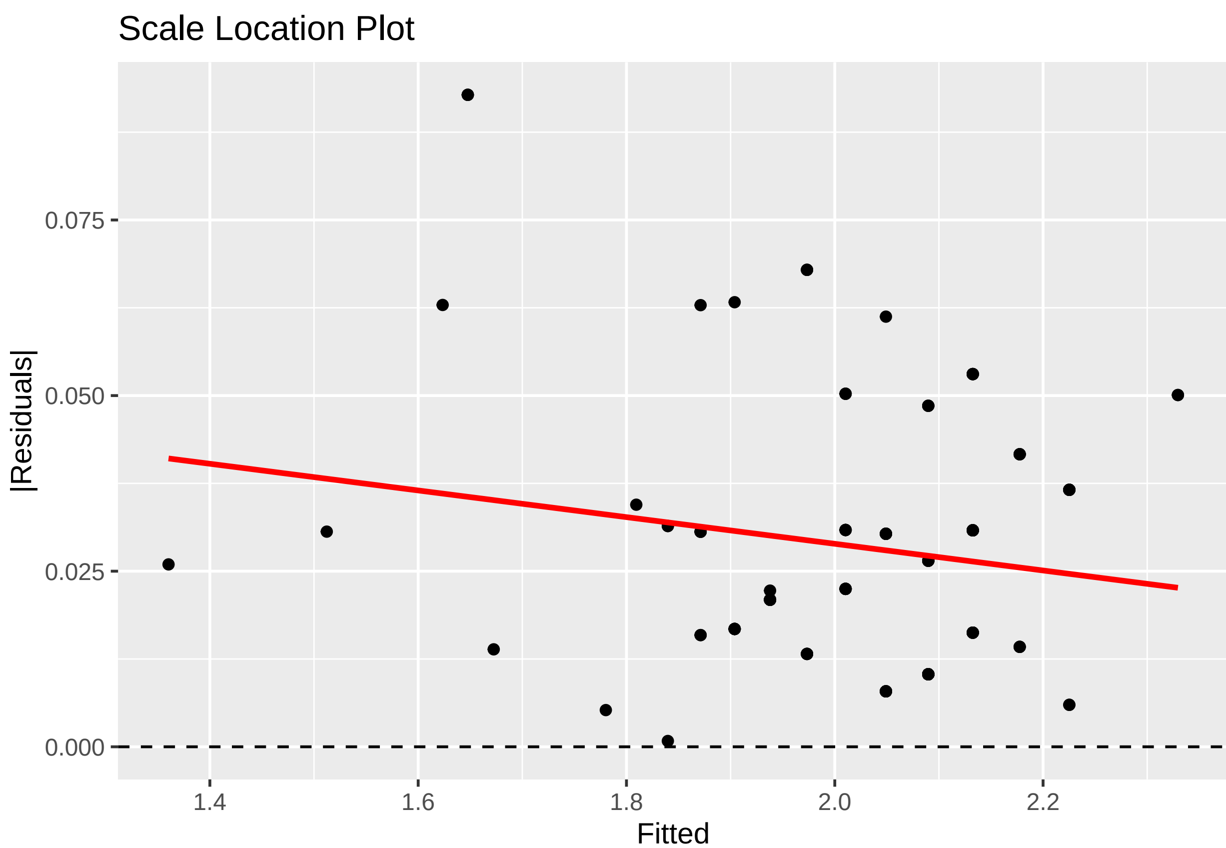



```
cfitmod <- fortify(cfit)
qplot(sample = scale(.resid), data = cfitmod) + geom_abline(intercept = 0, slope = 1, color = 'red') +
labs(title = 'Normal QQ Plot', y = 'Residuals' )
```




```
cfitmod <- fortify(cfit)
qplot(.fitted, abs(.resid), data = cfitmod) + geom_hline(yintercept = 0, linetype = 'dashed') +
labs(title = 'Scale Location Plot', x = 'Fitted', y = '|Residuals|' ) +
  geom_smooth(method = 'lm', color = 'red', se = F)
```

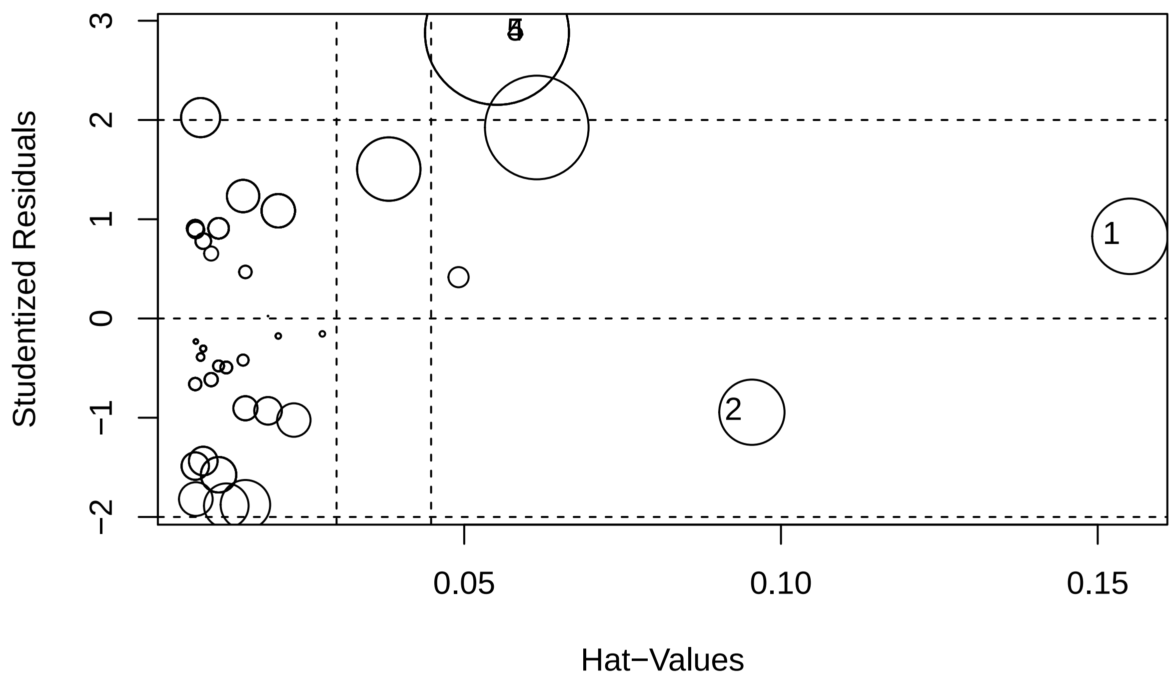
```
## 'geom_smooth()' using formula 'y ~ x'
```



Cook's distance:

- lets look if there are outliers that are influencing the model which we may want to remove:
- we can see what the highest Cook's distance values are with a maximum of .2292:

```
influencePlot(cfit)
```




```
##       StudRes        Hat      CookD
## 4  2.8779233 0.05517210 0.22917792
## 2 -0.9446127 0.09541329 0.04709672
## 5  2.8779233 0.05517210 0.22917792
## 1  0.8278611 0.15509206 0.06305240
```

```
cook <- cooks.distance(cfit)
summary(cook)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 5.850e-06 1.305e-03 2.708e-03 1.187e-02 8.926e-03 2.292e-01
```

- below we can combine the standard residual and cooks distance with the fuel data:

```
r <- stdres(cfit)
fuelcook <- cbind(fuel, cook, r)
```

- we can determine which rows in the data correspond to any cook's distance greater than 4/n: There are 10 rows that fit that criteria:

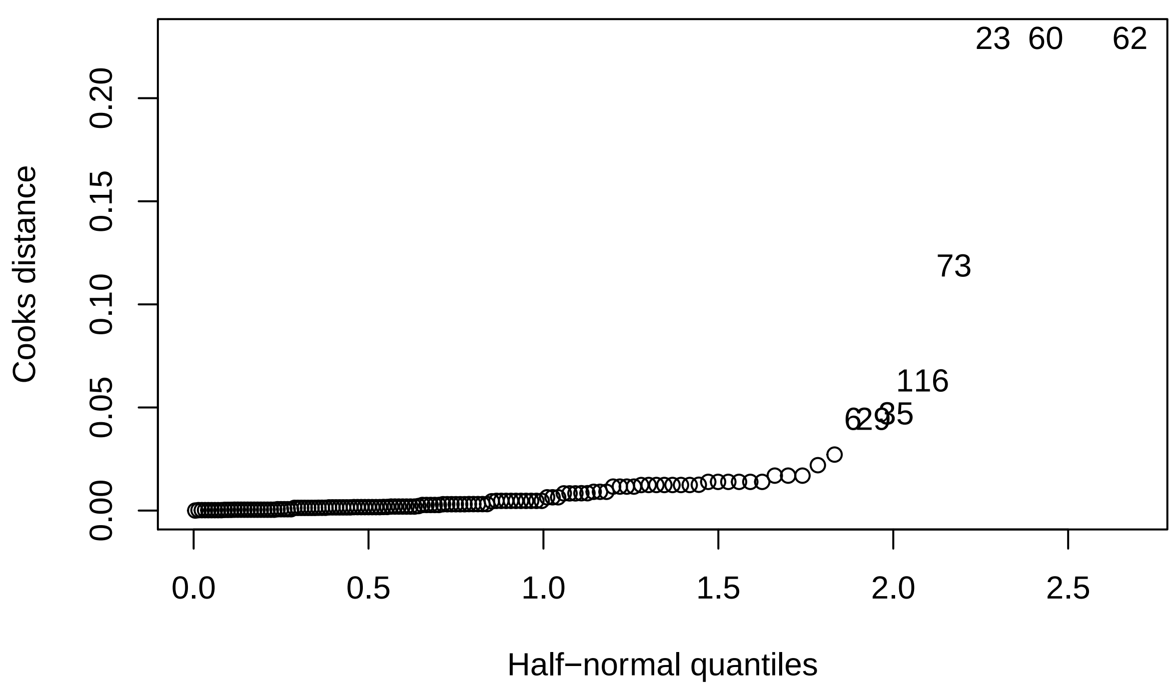
```
fuelcook[cook > 4/134,]
```

```
##                                       Model Cylinders Litres Barrels City
```

```
## 133 Chevrolet Colorado Cab Chassis inc 2WD         5    3.7    20.1    15
## 4                         Ford Escape Hybrid FWD         4    2.5    10.7    34
## 134           GMC Canyon Cab Chassis Inc 2WD         5    3.7    20.1    15
## 2                             Honda Civic Hybrid         4    1.3     8.2    40
## 5                       Mazda Tribute Hybrid 2WD         4    2.5    10.7    34
## 6                     Mercury Mariner Hybrid FWD         4    2.5    10.7    34
## 3                           Nissan Altima Hybrid         4    2.5    10.1    35
## 1                                   Toyota Prius         4    1.5     7.4    48
##     Highway Cost Carbon       cook          r
## 133      20 1667   10.8 0.04445043  1.4982420
## 4        31  887    5.7 0.22917792  2.8016766
## 134      20 1667   10.8 0.04445043  1.4982420
## 2        45  675    4.4 0.04709672 -0.9449983
## 5        31  887    5.7 0.22917792  2.8016766
## 6        31  887    5.7 0.22917792  2.8016766
## 3        33  833    5.4 0.11881856  1.9050509
## 1        45  615    4.0 0.06305240  0.8288496
```

- the plot below can also show which are the highest influencers on the model.

```
halfnorm(cook, 8, ylab = 'Cooks distance')
```



- further let's sort for Cook's distance from high to low:

```r
fuelcook <- cbind(fuel, cook, r)
sortedfuelcook <- fuelcook[order(-cook),]
```

- A look at the first 8 rows: These are the vehicles with Cook's distance > 4/n

```r
sortedfuelcook[1:8,c(1,9)]
```

```
##                                      Model       cook
## 4                 Ford Escape Hybrid FWD 0.22917792
## 5               Mazda Tribute Hybrid 2WD 0.22917792
## 6             Mercury Mariner Hybrid FWD 0.22917792
## 3                    Nissan Altima Hybrid 0.11881856
## 1                            Toyota Prius 0.06305240
## 2                      Honda Civic Hybrid 0.04709672
## 133 Chevrolet Colorado Cab Chassis inc 2WD 0.04445043
## 134         GMC Canyon Cab Chassis Inc 2WD 0.04445043
```

Fitting a model with outlier's (maximum Cook's distance) removed:

- below is a new model with the max cook's distance datapoints removed and a comparison of summary between the two models:

```r
cfit1 <- lm(log(Carbon) ~ log(City), data = fuel, subset = (cook < (4/134)))
summary(cfit)
```

```
##
## Call:
## lm(formula = log(Carbon) ~ log(City), data = fuel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06330 -0.02093 -0.01033  0.03070  0.09282
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.58581    0.05043   90.94   <2e-16 ***
## log(City)   -0.83320    0.01639  -50.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03408 on 132 degrees of freedom
## Multiple R-squared:  0.9514, Adjusted R-squared:  0.9511
## F-statistic:  2585 on 1 and 132 DF,  p-value: < 2.2e-16
```

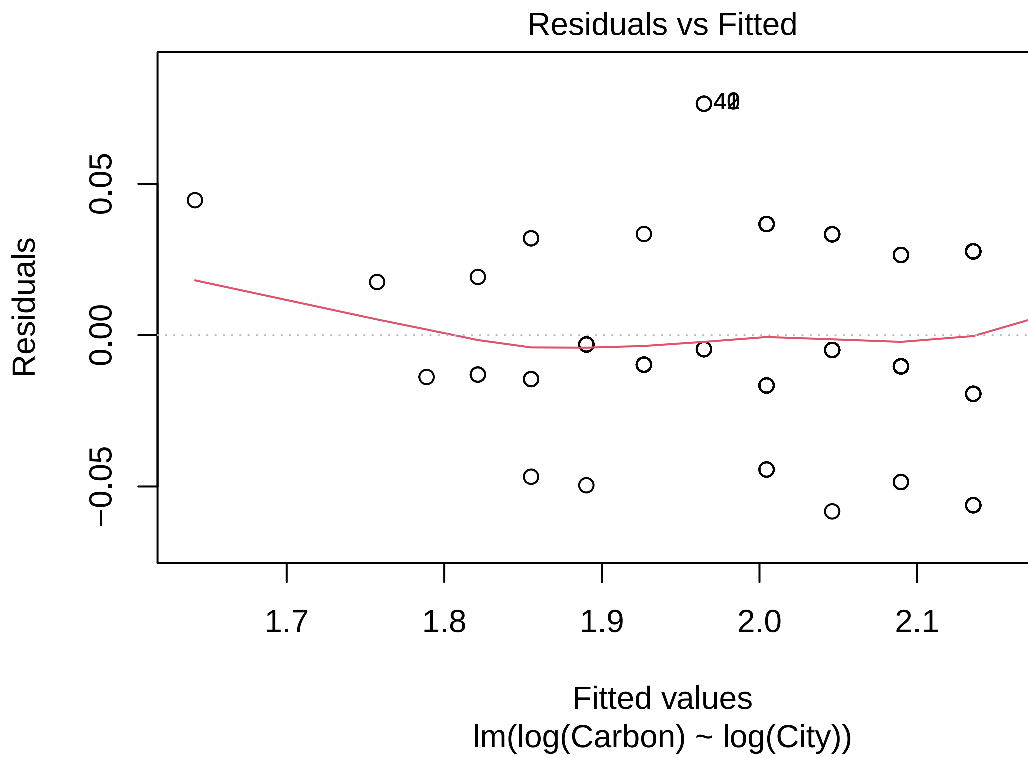
```r
summary(cfit1)
```

```
##
## Call:
## lm(formula = log(Carbon) ~ log(City), data = fuel, subset = (cook <
##     (4/134)))
```

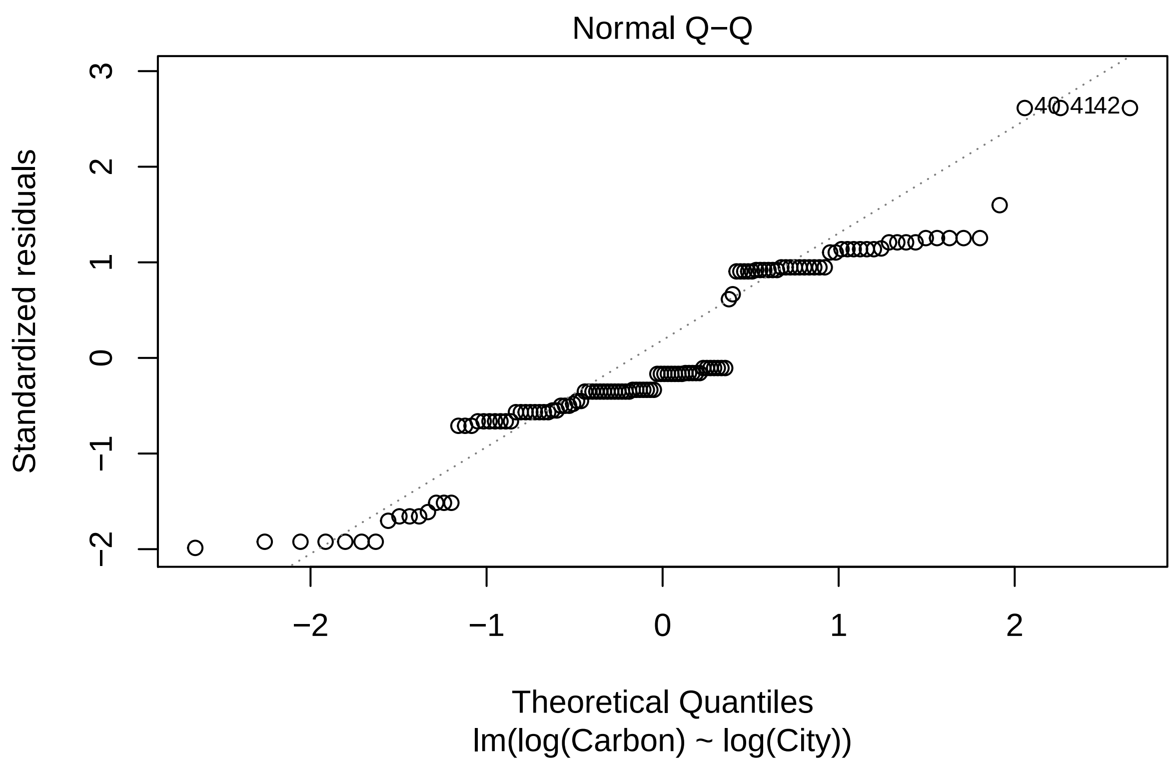
```
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.058230 -0.016617 -0.004884  0.027430  0.076491
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.76945    0.06147   77.59   <2e-16 ***
## log(City)   -0.89451    0.02012  -44.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02942 on 124 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.9405
## F-statistic:  1976 on 1 and 124 DF,  p-value: < 2.2e-16
```

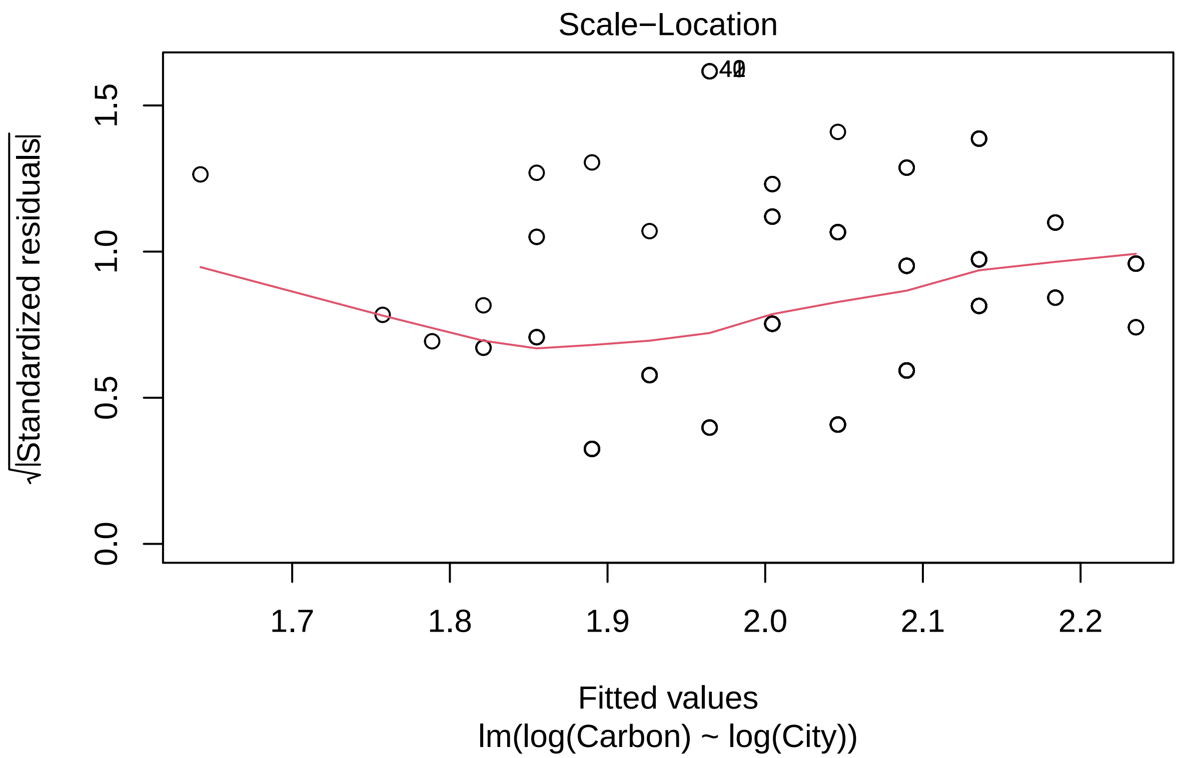
- we can see the second model has a high F-statistic with low p-value and the residual standard error is lower than the first model.

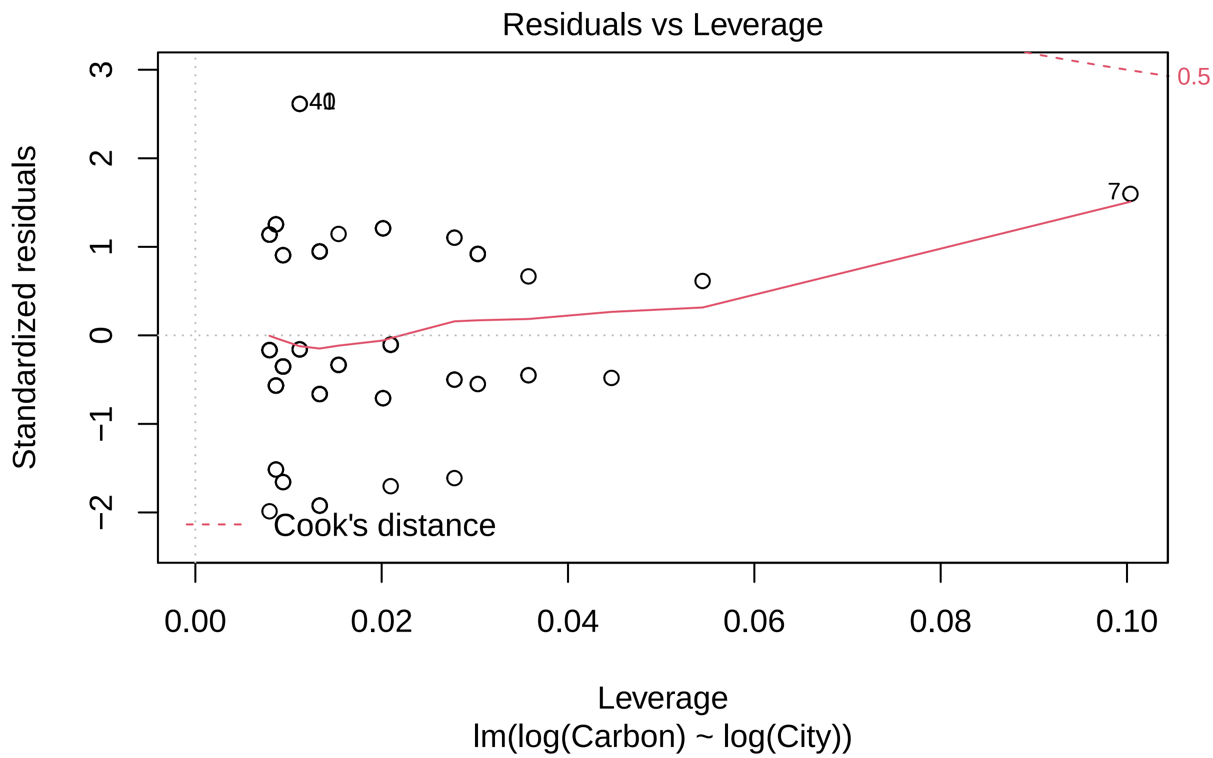
```
plot(cfit1)
```



Let's check the residuals now:




Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
41
40
1.7
1.8
1.9
2.0
2.1
2.2
Fitted values
lm(log(Carbon) ~ log(City))




```
Anova(cfit1, cfit)
```

Let's do an Anova test of the 2 models:

```
## Anova Table (Type II tests)
##
## Response: log(Carbon)
##             Sum Sq  Df F value    Pr(>F)
## log(City) 1.71072   1  1472.7 < 2.2e-16 ***
## Residuals 0.15334 132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- it shows the model with the outliers removed is a better fit with a high F-statistic and low p-value.