# Opioid death rate prediction of California

*Jiayi Liu*

*May 15, 2018*

## Install packages

```
library(MASS)
library(car)
```

```
## Loading required package: carData
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
library(carData)
library(corrplot)
```

```
## corrplot 0.84 loaded
```

## Load data and summary data

```
opioid = read.table("LRdata.csv", sep=',', header = TRUE)
str(opioid)
```

```
## 'data.frame':    58 obs. of  7 variables:
##  $ County        : Factor w/ 58 levels "Alameda","Alpine",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Death.Rate    : num  2.63 0 15.61 7.07 9.97 ...
##  $ Unemployment  : num  4.3 6.5 6 6.6 5.7 15.6 4.5 7.5 5.1 9.5 ...
##  $ Poverty       : num  10.7 18.6 11.3 19.5 13.1 11.1 8.7 23.7 8.7 25.5 ...
##  $ Median.Income : int  79831 62375 57032 44366 53502 54946 82881 42363 72586 45963 ...
##  $ Home.Price    : int  593500 329500 265900 228500 259000 205000 472900 183600 379200 204900 ...
##  $ no.high.school: num  13.6 9.6 11.9 12.3 7.2 30.6 11.2 20.9 6.8 26.9 ...
```

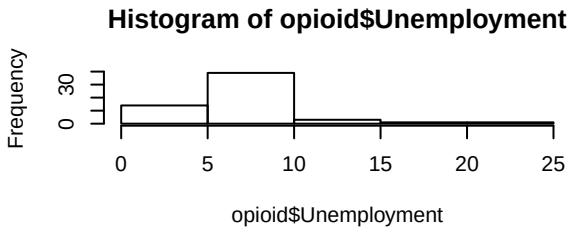
```
summary(opioid)
```

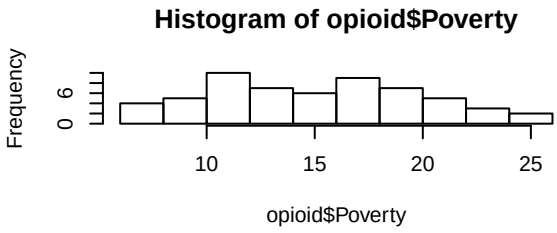
```
##        County     Death.Rate      Unemployment       Poverty
##  Alameda  : 1   Min.   : 0.000   Min.   : 3.000   Min.   : 6.60
##  Alpine   : 1   1st Qu.: 3.045   1st Qu.: 5.100   1st Qu.:11.10
##  Amador   : 1   Median : 5.740   Median : 6.350   Median :14.55
##  Butte    : 1   Mean   : 6.608   Mean   : 6.929   Mean   :15.08
##  Calaveras: 1   3rd Qu.: 8.238   3rd Qu.: 8.025   3rd Qu.:18.75
##  Colusa   : 1   Max.   :22.910   Max.   :23.600   Max.   :25.50
##  (Other)  :52
##  Median.Income      Home.Price     no.high.school
##  Min.   : 35270   Min.   :144600   Min.   : 5.70
##  1st Qu.: 44693   1st Qu.:204825   1st Qu.:11.45
```

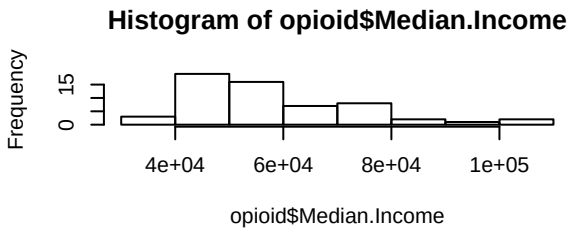
```
##  Median : 54996   Median :270700   Median :14.95
##  Mean   : 57999   Mean   :335724   Mean   :17.18
##  3rd Qu.: 66415   3rd Qu.:419650   3rd Qu.:21.77
##  Max.   :101173   Max.   :859400   Max.   :35.50
##
```

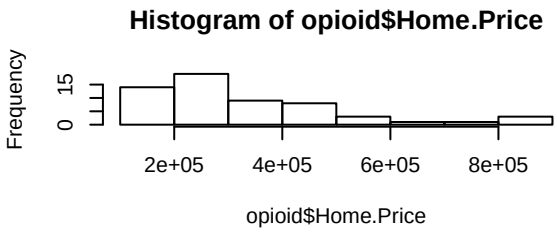
## Data visualization

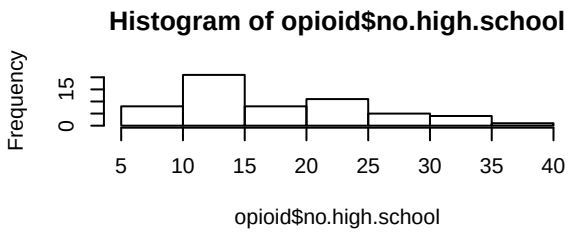
```
# histogram of factors

par(mfrow=c(3,2))
hist(opioid$Unemployment)
hist(opioid$Poverty)
hist(opioid$Median.Income)
hist(opioid$Home.Price)
hist(opioid$no.high.school)
```



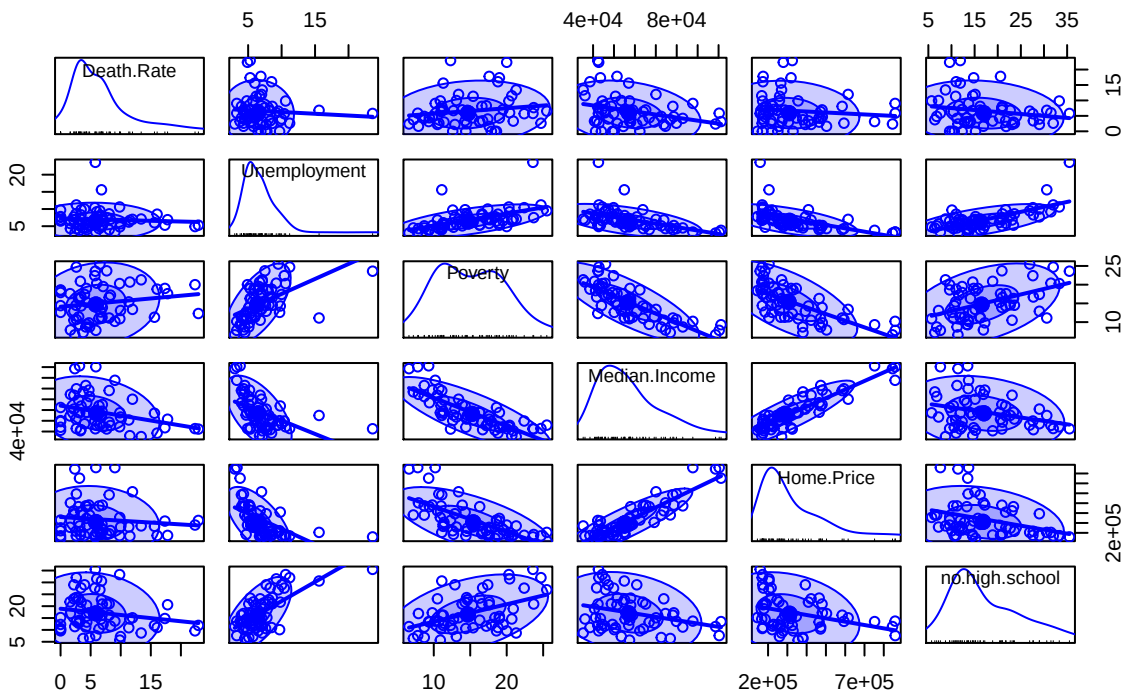



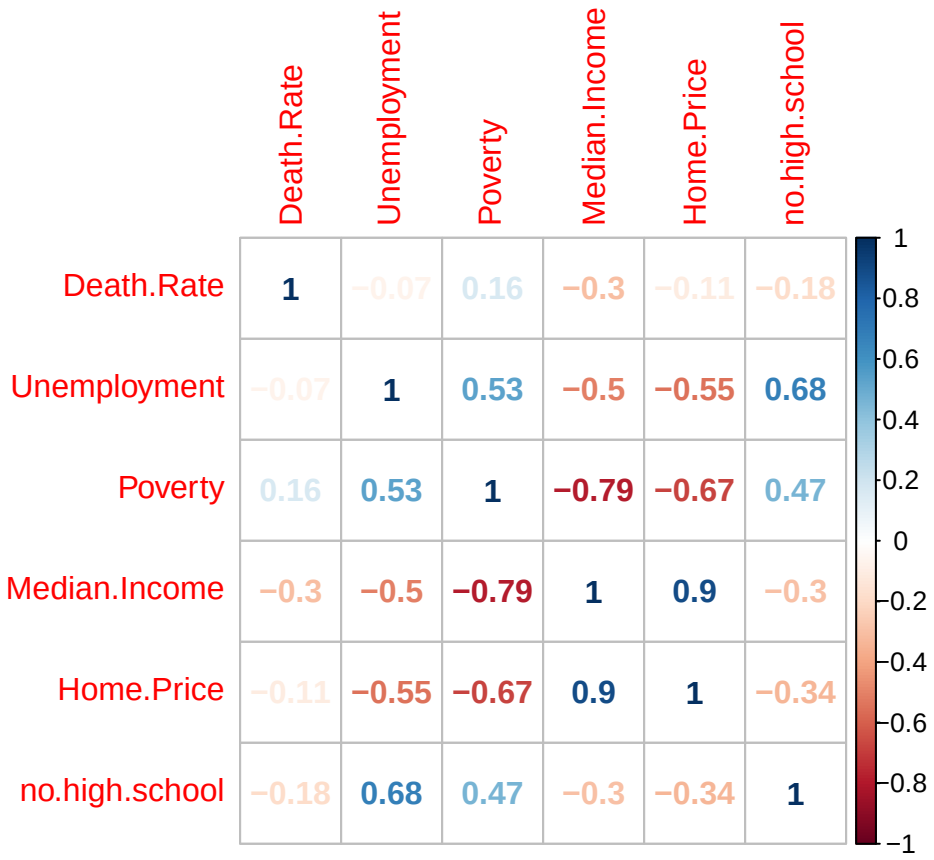

```
#look at correlations, scatterplots
scatterplotMatrix(~Death.Rate+Unemployment+Poverty+Median.Income+Home.Price+no.high.school,data=opioid,
```

```
## Warning in applyDefaults(ellipse, defaults = list(levels = c(0.5, 0.95), :
## unnamed ellipse arguments, will be ignored
```

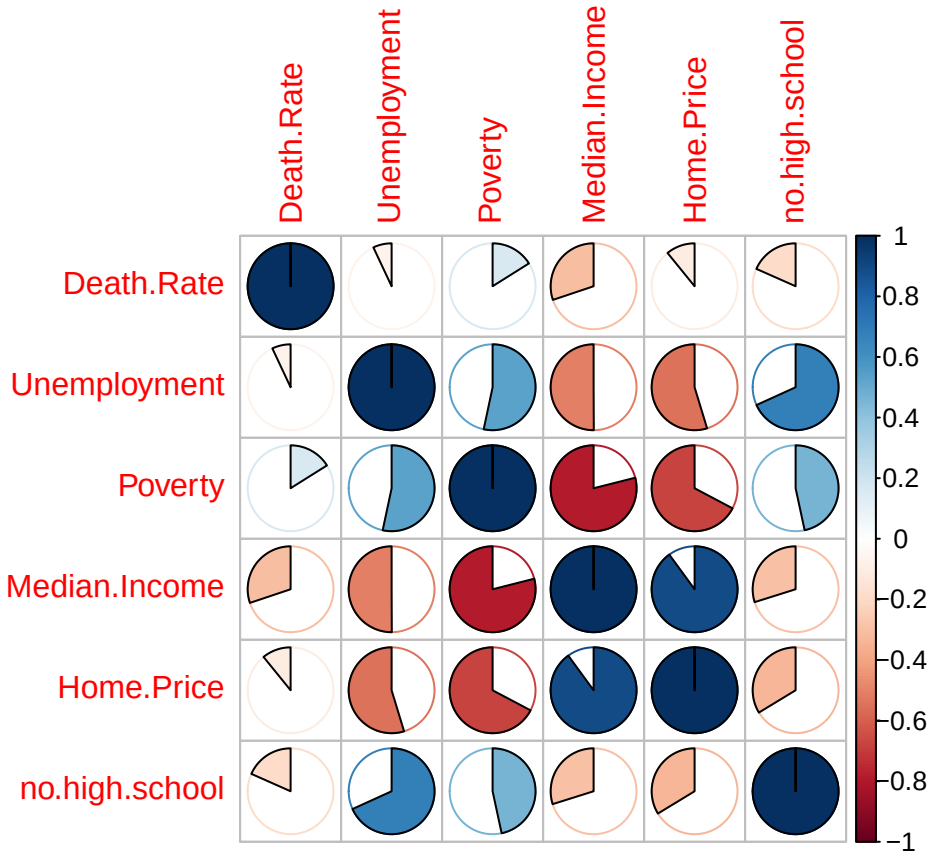



```
# cor(opioid[,2:7])
# corr.test(opioid[,2:7], y = NULL, use ="pairwise",method="pearson",adjust="holm",alpha=.05)
```

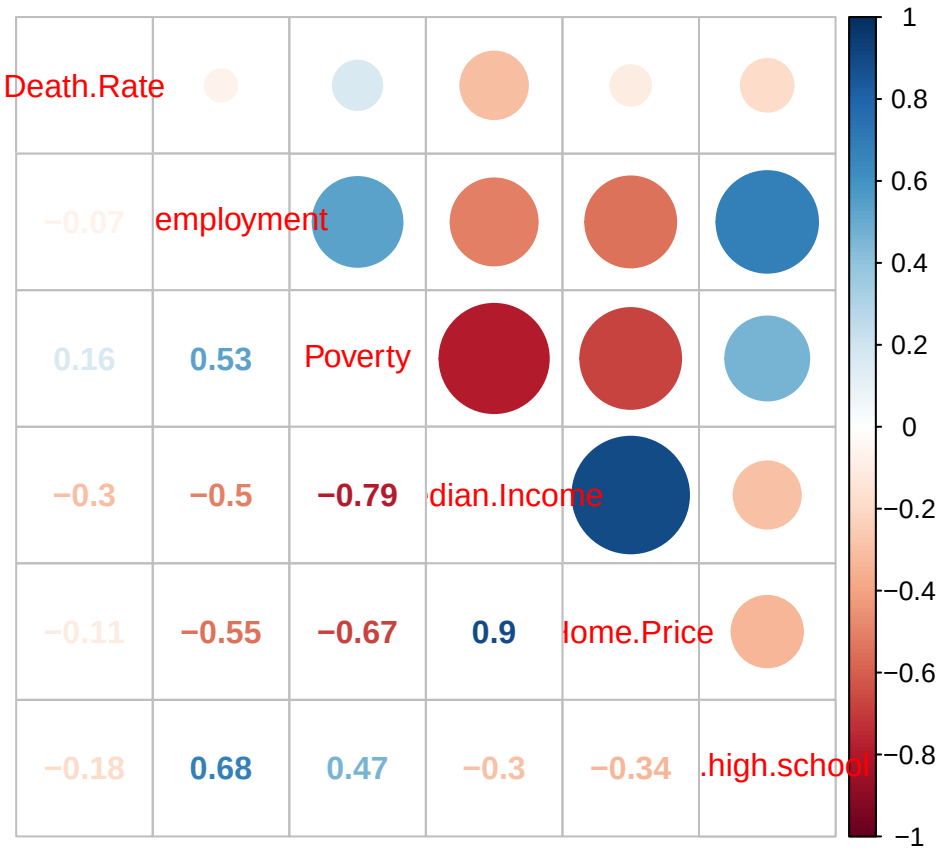
```
##more on correlation plots with p-values

M <- cor(opioid[,2:7])
corrplot(M,method="number")
```




```
corrplot(M,method="pie")
```




```
corrplot.mixed(M)
```




```
# matrix of the p-value of the correlation
p.mat <- cor.mtest(opioid[,2:7], conf.level=0.95)
corrplot(M, p.mat = p.mat$p, sig.level=0.05, insig="p-value")
```


## Build models

```
# use AIC to select features and find out that model with Median.Income, Home.Price and no.high.school
```

```
null=lm(Death.Rate~1, data=opioid)
summary(null)
```

```
##
## Call:
## lm(formula = Death.Rate ~ 1, data = opioid)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6079 -3.5629 -0.8679  1.6296 16.3021
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.6079     0.6939   9.523 2.19e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.285 on 57 degrees of freedom
```

```
full = lm(Death.Rate~.-County, data=opioid)
summary(full)
```

```
##
## Call:
## lm(formula = Death.Rate ~ . - County, data = opioid)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2550 -2.5748  0.3439  1.9636 13.1923
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.716e+01  7.169e+00   3.788 0.000396 ***
## Unemployment   -5.544e-02  2.956e-01  -0.188 0.851960
## Poverty        -1.786e-01  2.328e-01  -0.767 0.446392
## Median.Income  -3.991e-04  1.103e-04  -3.617 0.000673 ***
## Home.Price      2.326e-05  8.236e-06   2.823 0.006718 **
## no.high.school -1.240e-01  1.168e-01  -1.061 0.293534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.66 on 52 degrees of freedom
## Multiple R-squared:  0.2905, Adjusted R-squared:  0.2223
## F-statistic: 4.258 on 5 and 52 DF,  p-value: 0.00255
```

```
step(null, scope=list(lower=null, upper=full),direction="forward")
```

```
## Start:  AIC=194.11
## Death.Rate ~ 1
##
##                  Df Sum of Sq    RSS    AIC
## + Median.Income   1   144.685 1447.2 190.58
## <none>                        1591.8 194.11
## + no.high.school  1    53.668 1538.2 194.12
## + Poverty         1    41.147 1550.7 194.59
## + Home.Price      1    18.783 1573.0 195.42
## + Unemployment    1     7.767 1584.1 195.82
##
## Step:  AIC=190.58
## Death.Rate ~ Median.Income
##
##                  Df Sum of Sq    RSS    AIC
## + Home.Price      1   220.495 1226.7 182.99
## + no.high.school  1   130.833 1316.3 187.09
## + Unemployment    1   103.573 1343.6 188.27
## <none>                        1447.2 190.58
## + Poverty         1    25.222 1421.9 191.56
##
## Step:  AIC=182.99
## Death.Rate ~ Median.Income + Home.Price
##
##                  Df Sum of Sq    RSS    AIC
## + no.high.school  1    83.270 1143.4 180.92
## + Poverty         1    51.497 1175.2 182.51
## + Unemployment    1    43.582 1183.1 182.90
## <none>                        1226.7 182.99
##
```

```
## Step:  AIC=180.92
## Death.Rate ~ Median.Income + Home.Price + no.high.school
##
##                Df Sum of Sq    RSS    AIC
## <none>                      1143.4 180.92
## + Poverty       1   13.1797 1130.2 182.24
## + Unemployment  1    1.1571 1142.2 182.86
```

```
##
## Call:
## lm(formula = Death.Rate ~ Median.Income + Home.Price + no.high.school,
##     data = opioid)
##
## Coefficients:
##    (Intercept)   Median.Income      Home.Price  no.high.school
##      2.216e+01      -3.466e-04       2.215e-05      -1.683e-01
```

```
# build the model with Median.Income, Home.Price and no.high.school and find out the no.high.school is
rline0<-lm(formula=opioid$Death.Rate ~ opioid$Median.Income+opioid$Home.Price+opioid$no.high.school)
summary(rline0)
```

```
##
## Call:
## lm(formula = opioid$Death.Rate ~ opioid$Median.Income + opioid$Home.Price +
##     opioid$no.high.school)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9083 -2.6929  0.2688  2.0475 13.0616
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.216e+01  3.492e+00   6.347 4.71e-08 ***
## opioid$Median.Income   -3.466e-04  8.616e-05  -4.022  0.00018 ***
## opioid$Home.Price       2.215e-05  7.752e-06   2.858  0.00605 **
## opioid$no.high.school  -1.683e-01  8.487e-02  -1.983  0.05245 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.602 on 54 degrees of freedom
## Multiple R-squared:  0.2817, Adjusted R-squared:  0.2418
## F-statistic:  7.06 on 3 and 54 DF,  p-value: 0.0004341
```

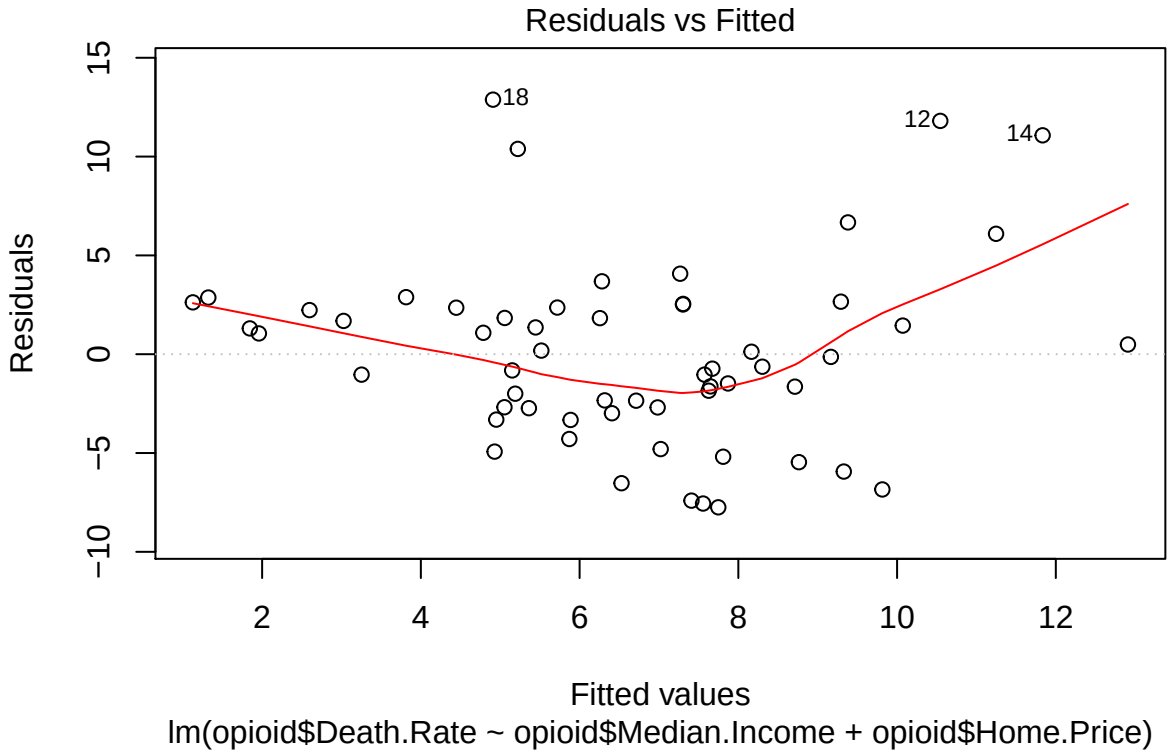
```
# build model with Median.Income and Home.Price only and find the features are still significant and th

rline2<-lm(formula = opioid$Death.Rate ~ opioid$Median.Income+opioid$Home.Price)
summary(rline2)
```

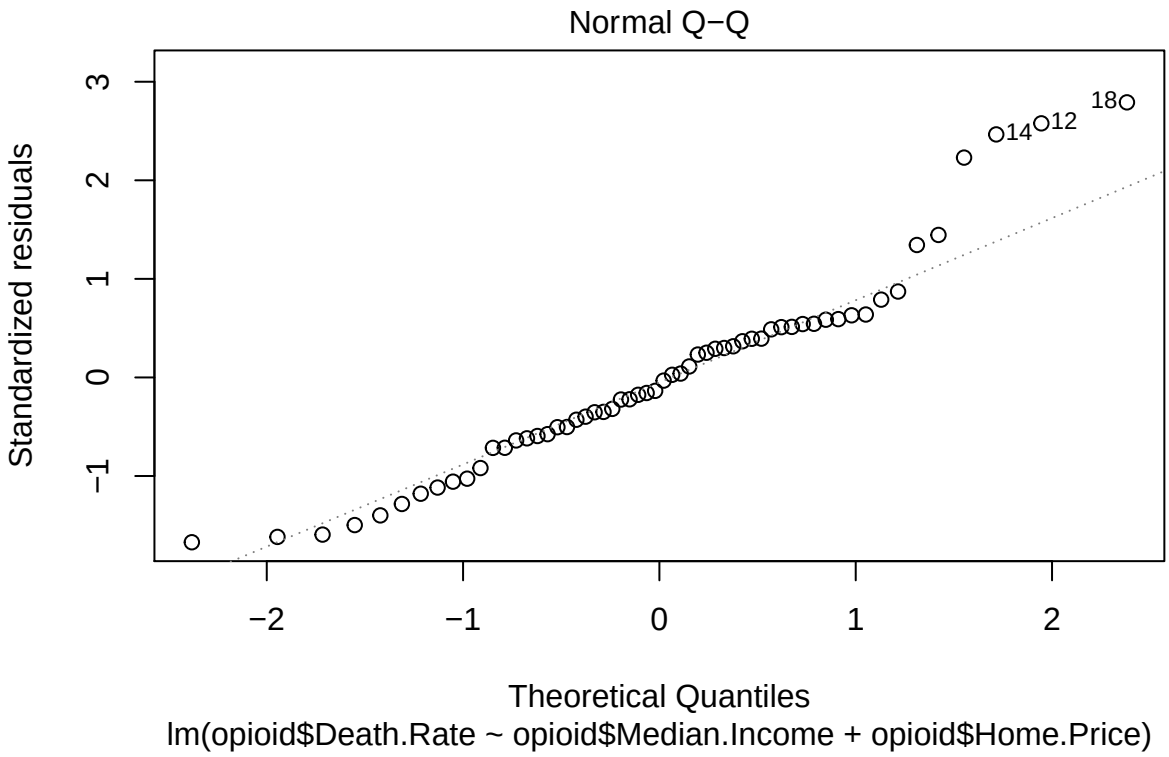
```
##
## Call:
## lm(formula = opioid$Death.Rate ~ opioid$Median.Income + opioid$Home.Price)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.747 -2.722 -0.384  2.359 12.881
##
```

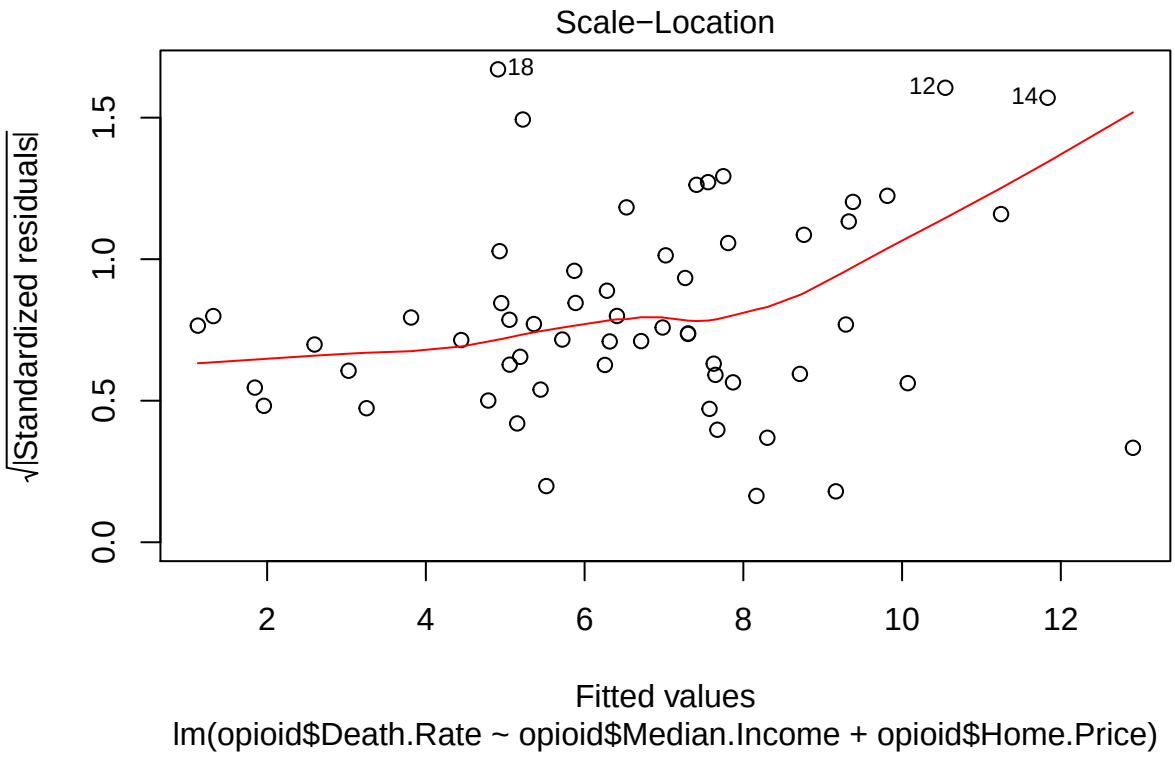
```
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.853e+01  3.052e+00   6.072 1.23e-07 ***
## opioid$Median.Income -3.485e-04  8.842e-05  -3.941 0.000231 ***
## opioid$Home.Price     2.468e-05  7.848e-06   3.144 0.002684 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.723 on 55 degrees of freedom
## Multiple R-squared:  0.2294, Adjusted R-squared:  0.2014
## F-statistic: 8.187 on 2 and 55 DF,  p-value: 0.0007721
```

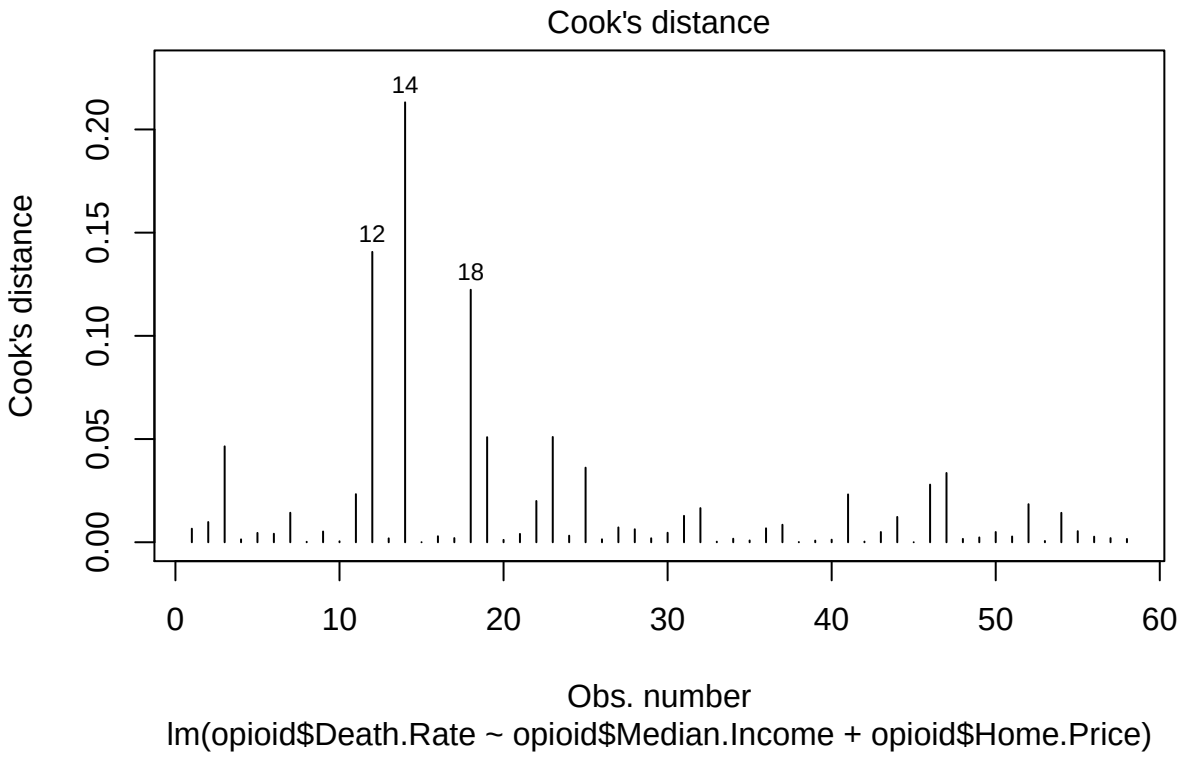
### Check assumptions

```
#diagnostics on residuals

plot(rline2,which=1:4)
```

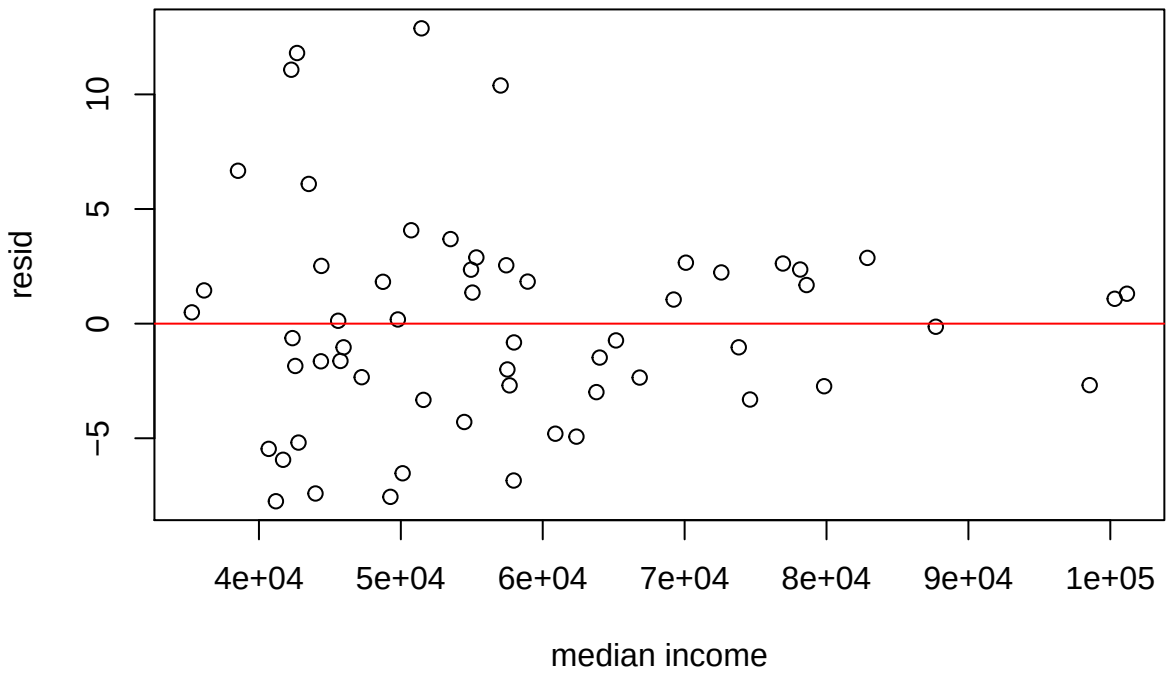



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(opioid$Death.Rate ~ opioid$Median.Income + opioid$Home.Price)
18
14
12

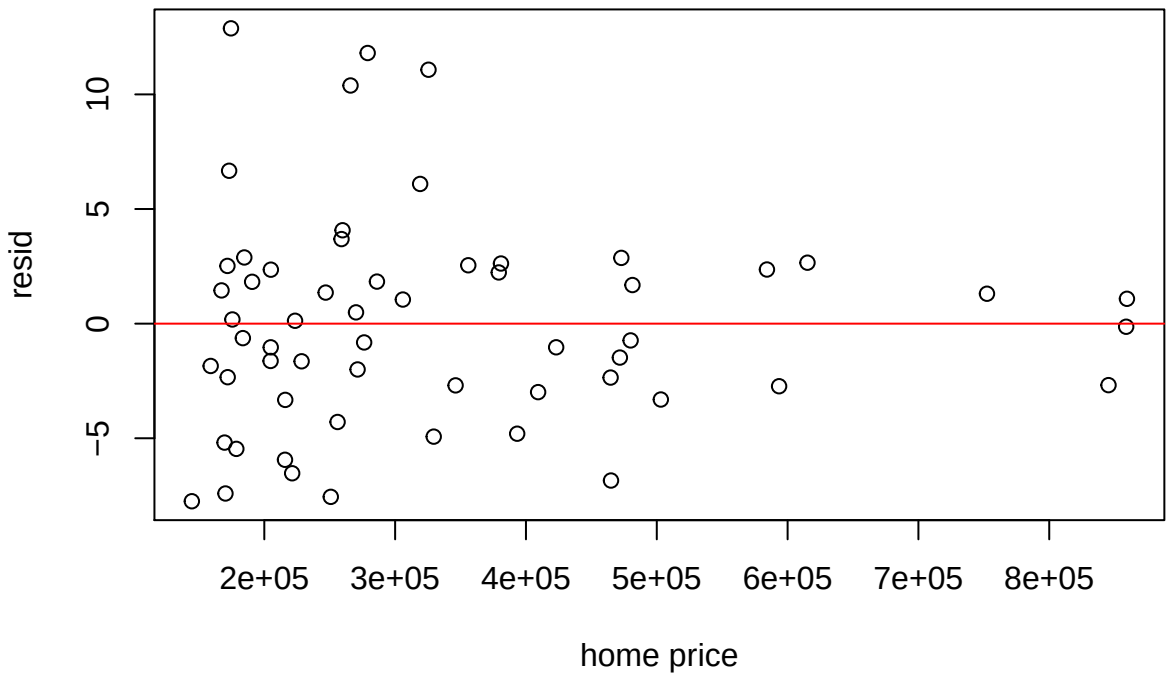

Scale−Location
18
12
14
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
4
6
8
10
12
Fitted values
lm(opioid$Death.Rate ~ opioid$Median.Income + opioid$Home.Price)




```
e=residuals(rline2)

plot(opioid$Median.Income, e, xlab = "median income", ylab =
              "resid")
abline(0,0, col="red")
```




```
plot(opioid$Home.Price, e, xlab = "home price", ylab =
              "resid")
abline(0,0, col="red")
```




```
# normality and constant variance models
shapiro.test(rline2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rline2$residuals
## W = 0.94465, p-value = 0.01041
```

```
ncvTest(rline2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 5.51864, Df = 1, p = 0.018815
```

```
# normality condition passed but the constant variance test failed and there are some outliers
```

## Transformation

```
# Attempts at transformations to get constance of variance box cox, Have to artificially inflate rate=0
opioid_boxcox<-opioid
opioid_boxcox[,2][opioid_boxcox[,2] == 0] <- 0.0001
opioid_boxcox
```

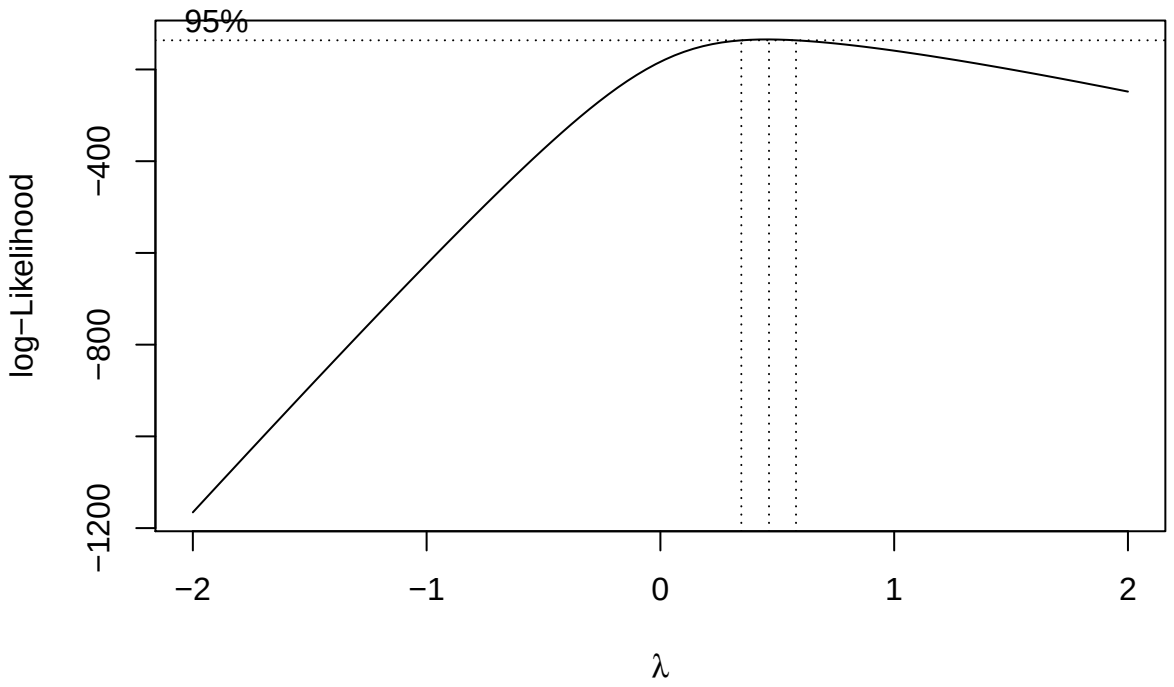
```
##             County Death.Rate Unemployment Poverty Median.Income
## 1          Alameda     2.6300          4.3    10.7         79831
## 2           Alpine     0.0001          6.5    18.6         62375
```

```
## 3           Amador    15.6100          6.0    11.3         57032
## 4            Butte     7.0700          6.6    19.5         44366
## 5        Calaveras     9.9700          5.7    13.1         53502
## 6           Colusa     6.8000         15.6    11.1         54946
## 7     Contra Costa     4.1900          4.5     8.7         82881
## 8        Del Norte     7.6700          7.5    23.7         42363
## 9        El Dorado     4.8300          5.1     8.7         72586
## 10          Fresno     6.5400          9.5    25.5         45963
## 11           Glenn     3.3900          8.3    16.7         41699
## 12        Humboldt    22.3500          4.9    20.0         42685
## 13        Imperial     5.7800         23.6    23.6         42560
## 14            Inyo    22.9100          5.3    12.3         42278
## 15            Kern     5.7000         10.4    22.4         49788
## 16           Kings     3.9800         10.0    17.4         47241
## 17            Lake    11.5200          6.7    20.7         36132
## 18          Lassen    17.7900          6.9    17.6         51457
## 19     Los Angeles     2.9700          5.3    16.3         57952
## 20          Madera     6.0200          9.2    20.4         45742
## 21           Marin     5.8700          3.3     7.8        100310
## 22        Mariposa     0.0001          6.9    17.9         49265
## 23       Mendocino    17.3400          5.3    19.0         43510
## 24          Merced     9.8200         10.6    20.3         44397
## 25           Modoc     0.0001          7.8    18.4         41194
## 26            Mono     6.8900          5.4    11.3         58937
## 27        Monterey     2.2200          7.7    12.9         60889
## 28            Napa     1.6400          4.3     7.9         74609
## 29          Nevada     9.8500          4.8    10.9         57429
## 30          Orange     8.0800          4.0    11.1         78145
## 31          Placer     3.7500          4.5     7.2         76926
## 32          Plumas     0.0001          9.7    12.6         50125
## 33       Riverside     4.3300          6.1    15.3         57972
## 34      Sacramento     3.1900          5.4    16.3         57509
## 35      San Benito     2.2200          6.8    10.5         73814
## 36  San Bernardino     1.5800          5.8    17.6         54469
## 37       San Diego     6.7000          4.7    12.7         55322
## 38   San Francisco     9.0300          3.3    10.2         87701
## 39     San Joaquin     6.8000          8.1    14.6         55045
## 40 San Luis Obispo     6.3900          4.3    11.0         64014
## 41       San Mateo     2.3700          3.0     6.6         98546
## 42   Santa Barbara     6.9400          5.1    13.9         65161
## 43     Santa Clara     3.1500          3.8     9.3        101173
## 44      Santa Cruz    11.9500          7.0    13.4         70088
## 45          Shasta     8.2900          7.0    17.2         45582
## 46          Sierra     0.0001          7.6    14.4         43984
## 47        Siskiyou    16.0500          8.5    18.8         38524
## 48          Solano     3.0100          5.5    11.4         69227
## 49          Sonoma     4.3600          4.0     9.4         66833
## 50      Stanislaus     2.5600          8.6    14.5         51591
## 51          Sutter     3.4200          9.8    14.3         63783
## 52          Tehama     3.3000          7.1    20.9         40687
## 53         Trinity    13.4000          7.1    21.3         35270
## 54          Tulare     2.6200         11.2    24.7         42789
## 55        Tuolumne    11.3400          6.2    15.4         50731
## 56         Ventura     4.7100          5.2     9.8         78593
```

```
## 57            Yolo     4.2900          5.9    19.0         57663
## 58            Yuba     8.0800          8.6    16.6         48739
##    Home.Price no.high.school
## 1      593500           13.6
## 2      329500            9.6
## 3      265900           11.9
## 4      228500           12.3
## 5      259000            7.2
## 6      205000           30.6
## 7      472900           11.2
## 8      183600           20.9
## 9      379200            6.8
## 10     204900           26.9
## 11     215800           27.7
## 12     279000            9.6
## 13     159000           35.5
## 14     325500           11.9
## 15     175600           27.6
## 16     172000           29.0
## 17     167300           14.6
## 18     174500           20.6
## 19     465000           23.4
## 20     204800           31.5
## 21     859400            7.6
## 22     250800           12.3
## 23     319110           14.8
## 24     171800           33.3
## 25     144600           15.2
## 26     286100           16.2
## 27     393300           29.0
## 28     503100           16.9
## 29     355900            5.7
## 30     584200           16.2
## 31     380900            6.4
## 32     221300           10.1
## 33     276300           20.4
## 34     271300           14.1
## 35     423100           23.1
## 36     256000           21.8
## 37     184700           14.5
## 38     858800           13.7
## 39     246900           22.7
## 40     471800           10.4
## 41     845300           11.4
## 42     480000           20.9
## 43     752400           13.5
## 44     615200           15.1
## 45     223500           11.6
## 46     170300           10.6
## 47     173100           11.0
## 48     305900           12.8
## 49     464700           13.3
## 50     216000           23.6
## 51     409300           21.7
```

```
## 52     178600           18.9
## 53     270100            8.7
## 54     169600           32.0
## 55     259800           10.9
## 56     481400           17.2
## 57     346200           15.7
## 58     190700           21.0
```

```
rline_boxcox<-lm(formula = opioid_boxcox$Death.Rate~opioid_boxcox$Median.Income+opioid_boxcox$Home.Price)
boxcox(rline_boxcox)
```



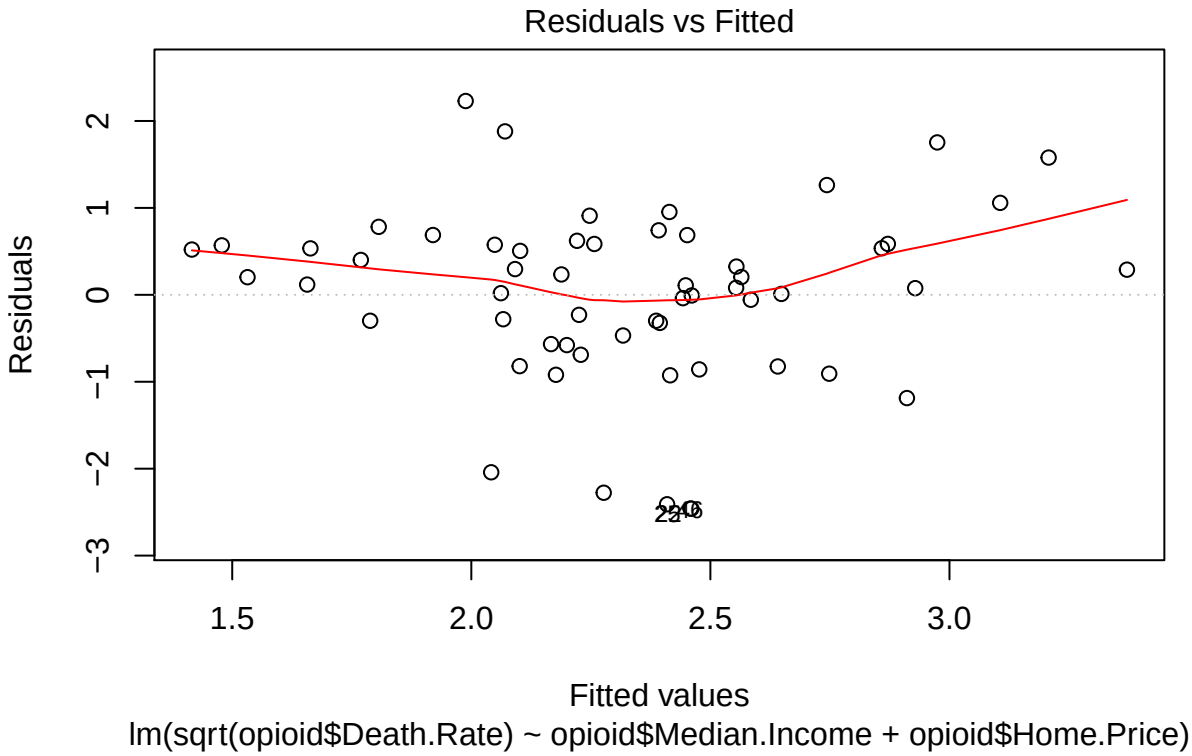

```
# the best suggested transfromation is about 0.5 for y.
```

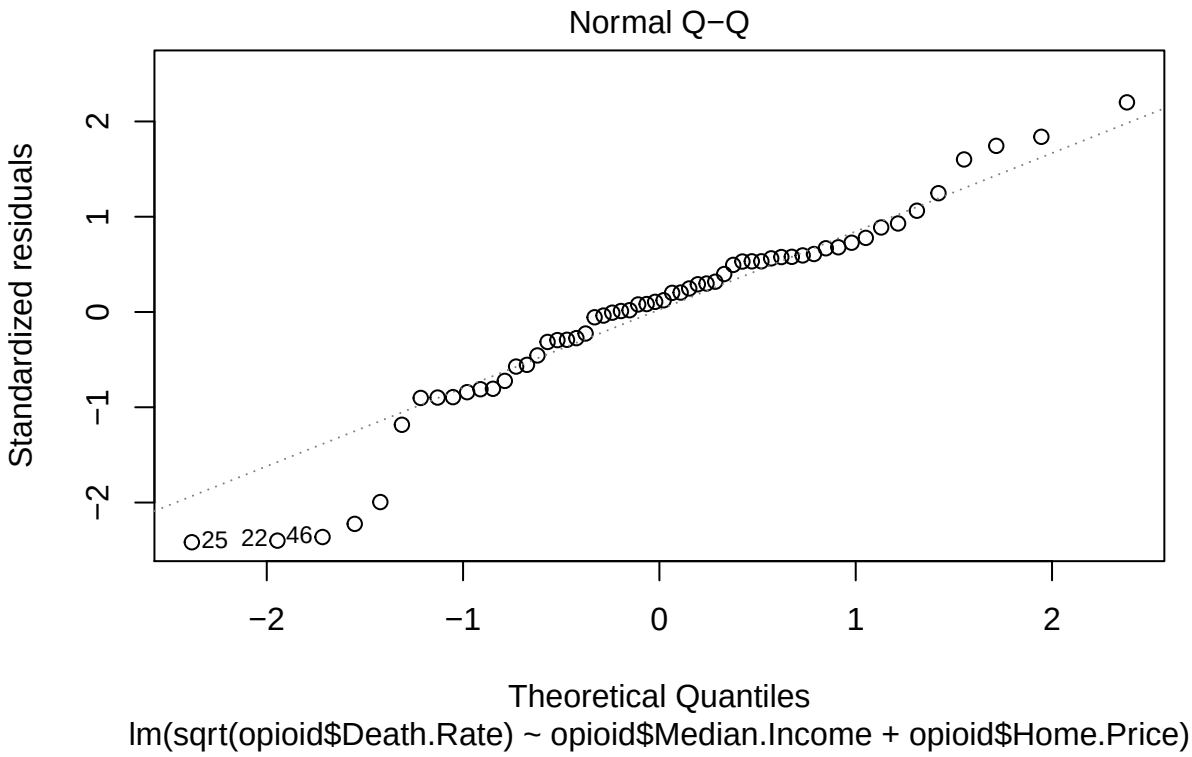
```
# build transform model and check assumptions

rline_transform<-lm(formula = sqrt(opioid$Death.Rate) ~ opioid$Median.Income+opioid$Home.Price)

summary(rline_transform)
```

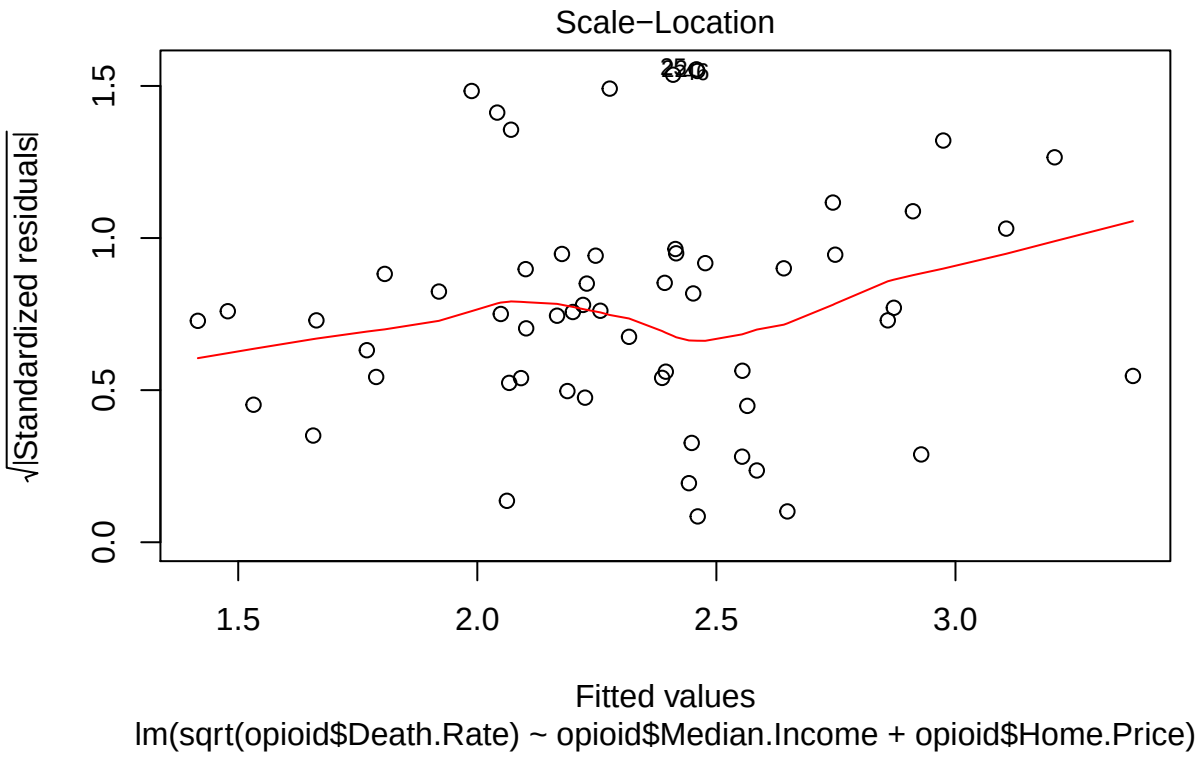
```
##
## Call:
## lm(formula = sqrt(opioid$Death.Rate) ~ opioid$Median.Income +
##     opioid$Home.Price)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4601 -0.5421  0.1135  0.5829  2.2296
##
```

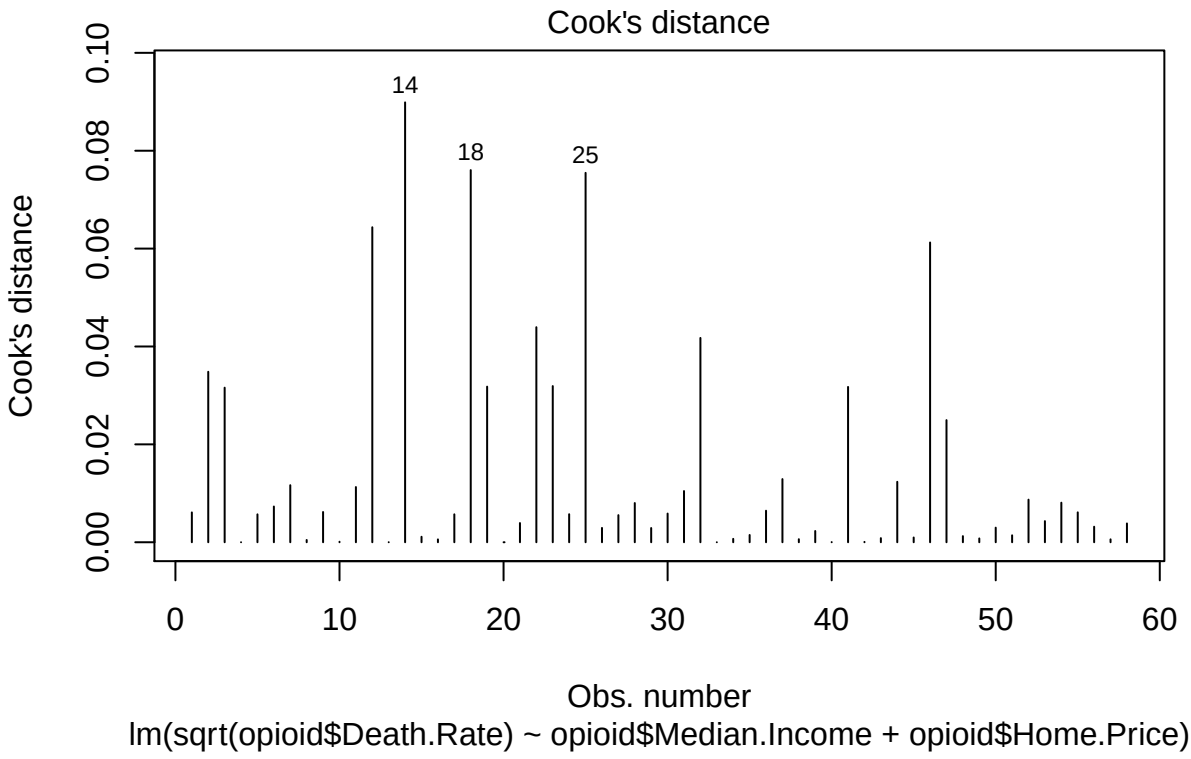
```
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.235e+00  6.701e-01   6.320 4.88e-08 ***
## opioid$Median.Income -5.890e-05  1.941e-05  -3.034  0.00368 **
## opioid$Home.Price     4.493e-06  1.723e-06   2.608  0.01171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.037 on 55 degrees of freedom
## Multiple R-squared:  0.1444, Adjusted R-squared:  0.1133
## F-statistic: 4.641 on 2 and 55 DF,  p-value: 0.01372
```

```
plot(rline_transform, which=1:4)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(sqrt(opioid$Death.Rate) ~ opioid$Median.Income + opioid$Home.Price)
25
22
46


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
Fitted values
lm(sqrt(opioid$Death.Rate) ~ opioid$Median.Income + opioid$Home.Price)



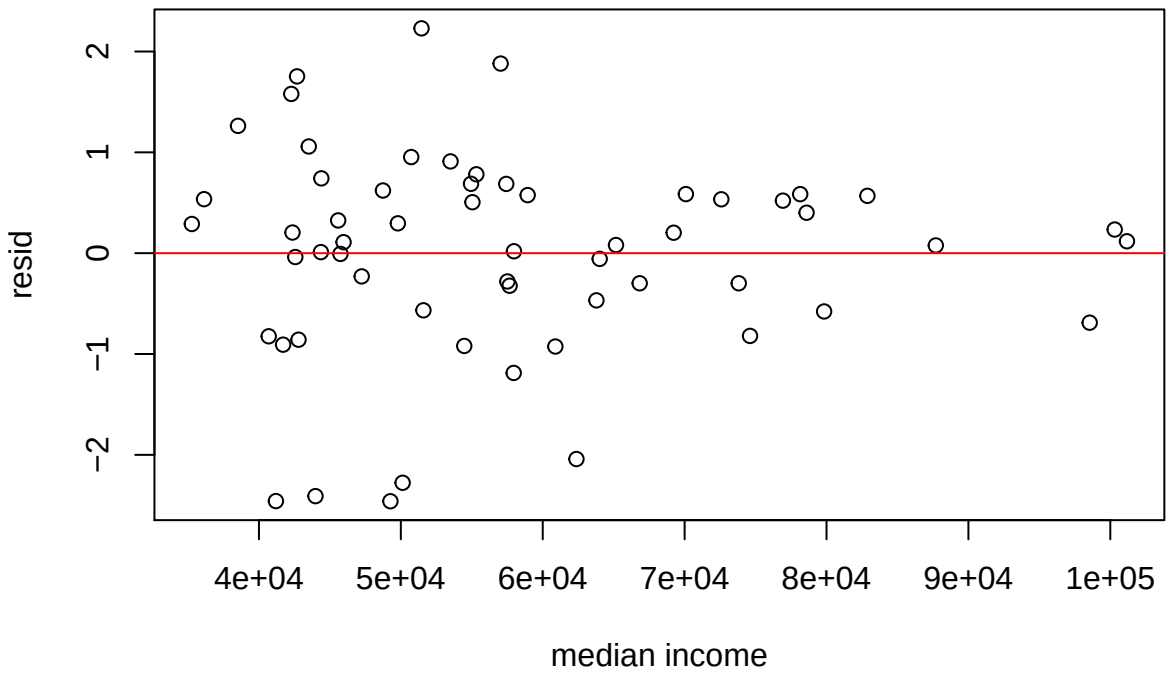

```
shapiro.test(rline_transform$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rline_transform$residuals
## W = 0.94761, p-value = 0.01415
```

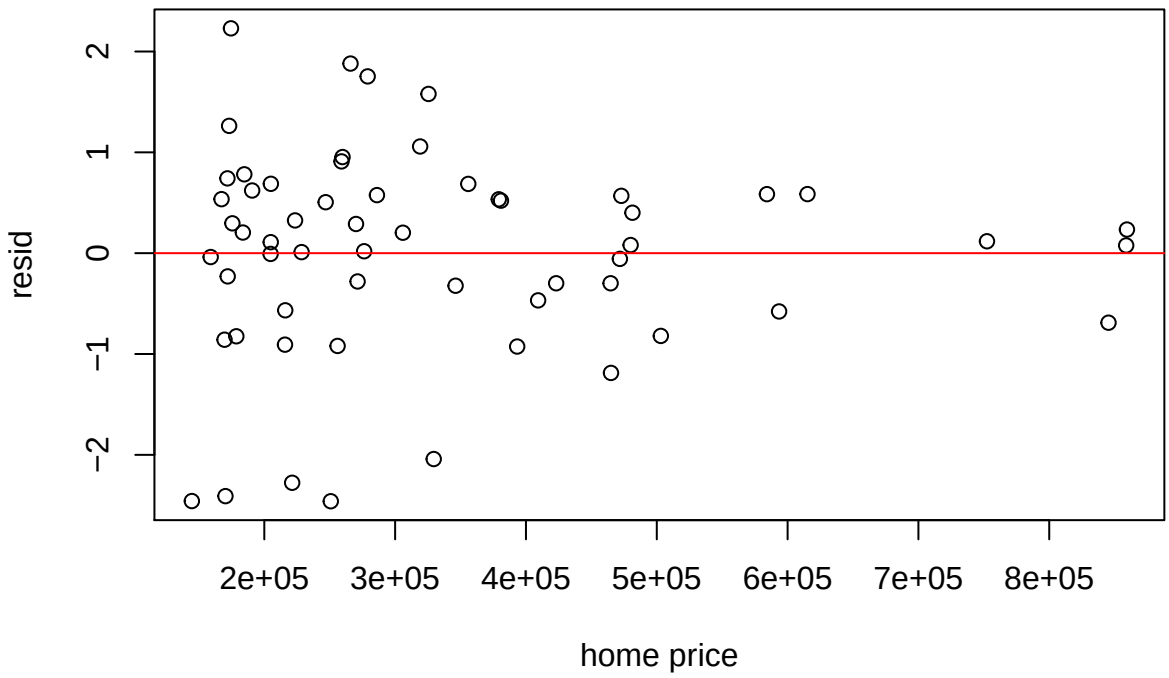
```
ncvTest(rline_transform)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.6981284, Df = 1, p = 0.40341
```

```
e2=residuals(rline_transform)
plot(opioid$Median.Income, e2, xlab = "median income", ylab =
             "resid")
abline(0,0, col="red")
```



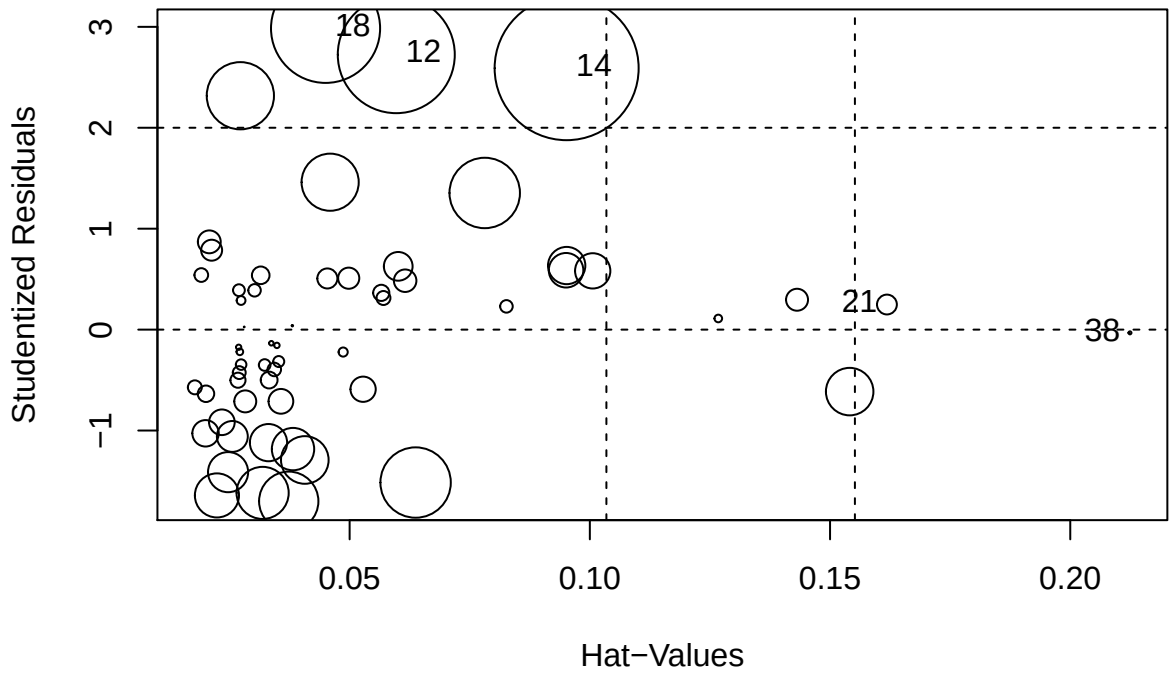

```
plot(opioid$Home.Price, e2, xlab = "home price", ylab =
               "resid")
abline(0,0, col="red")
```




```
# normality and constant variance both pass

# for Breusch-Pagan-Godfrey test, The null hypothesis for this test is that the error variances are all
```

### Check outliers

```
# for original model with median income and house price
influencePlot(rline2)
```

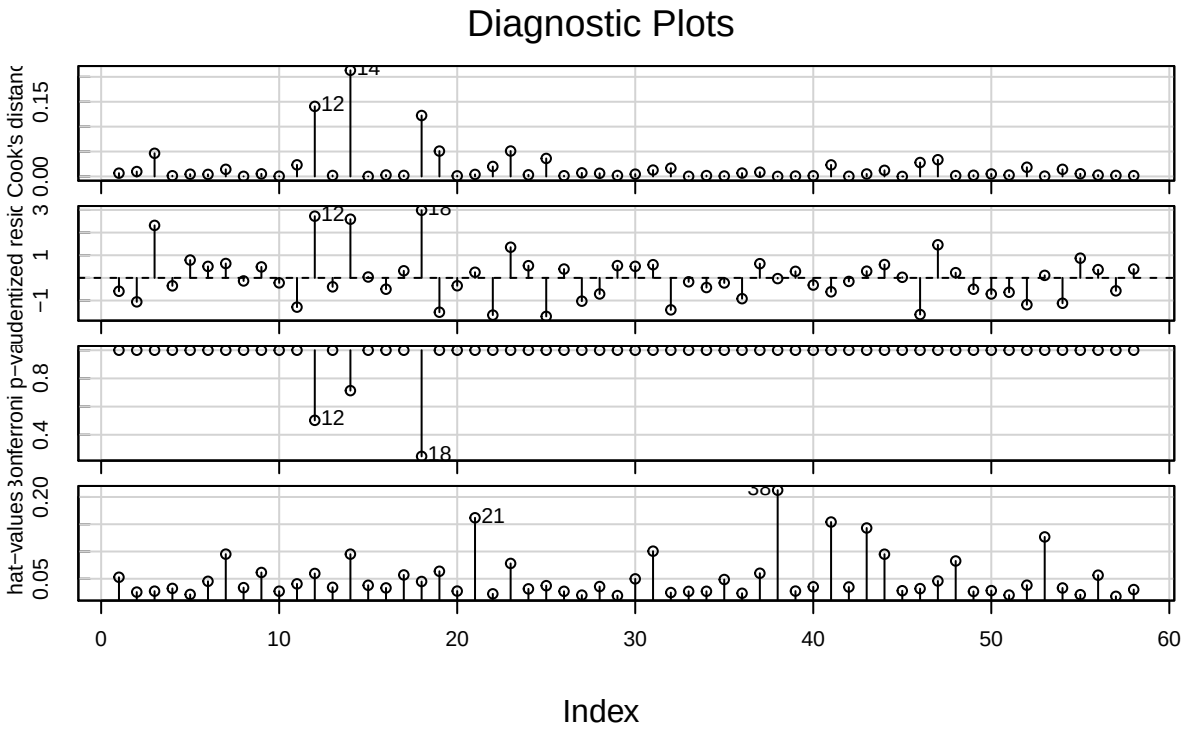



```
##        StudRes        Hat        CookD
## 12  2.72428416 0.05970936 1.406710e-01
## 14  2.59045900 0.09516861 2.131362e-01
## 18  2.98500168 0.04498952 1.223242e-01
## 21  0.24846961 0.16183919 4.042540e-03
## 38 -0.03203127 0.21242300 9.394989e-05
```

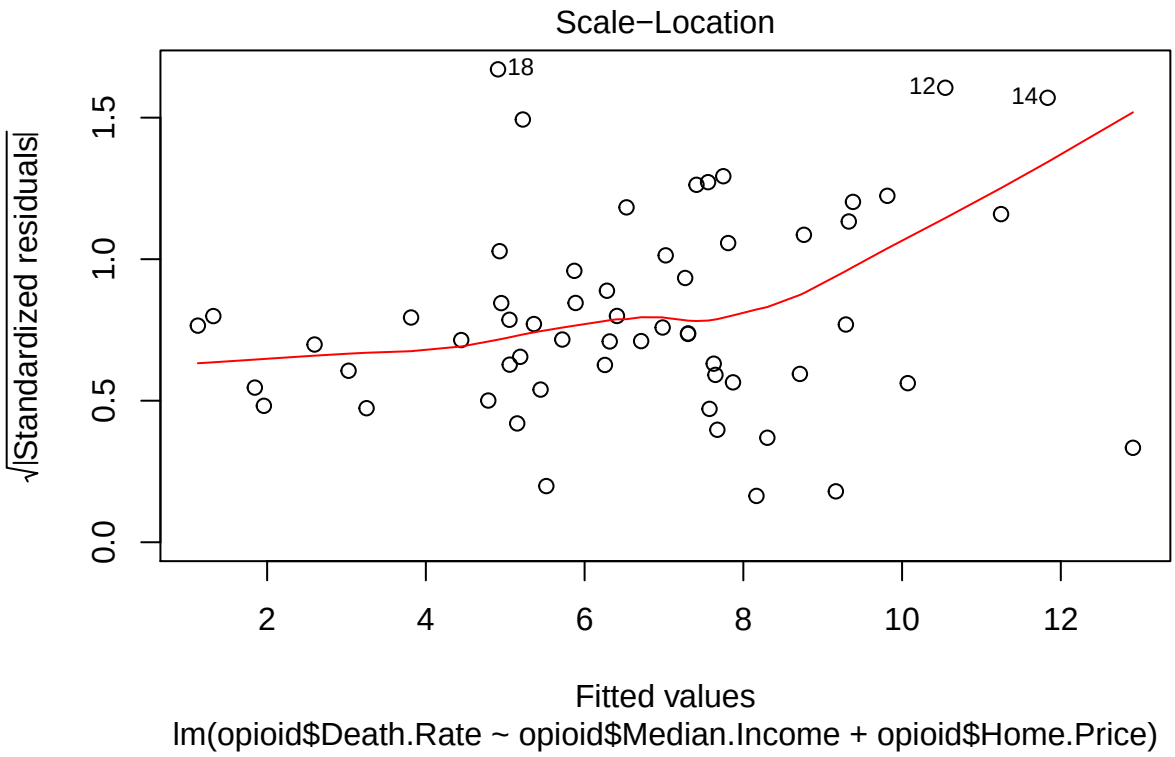
```
outlierTest(rline2)
```

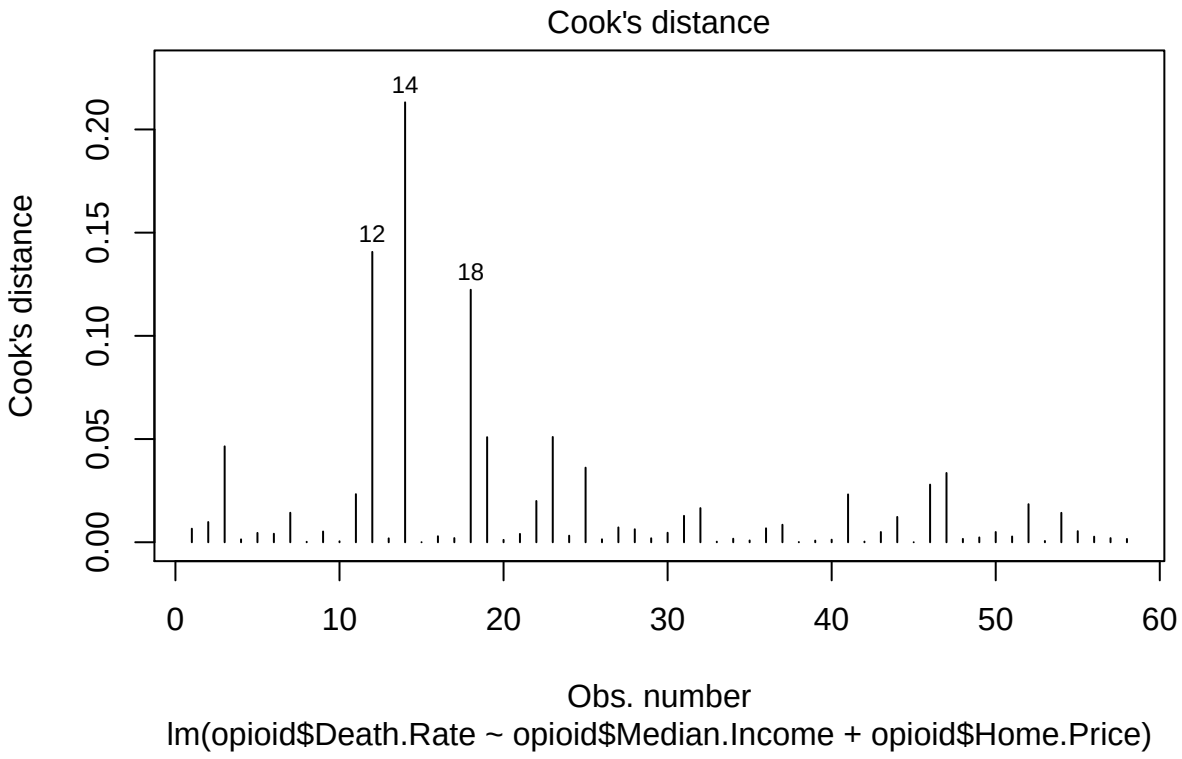
```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 18 2.985002          0.0042539      0.24673
```

```
influenceIndexPlot(rline2)
```

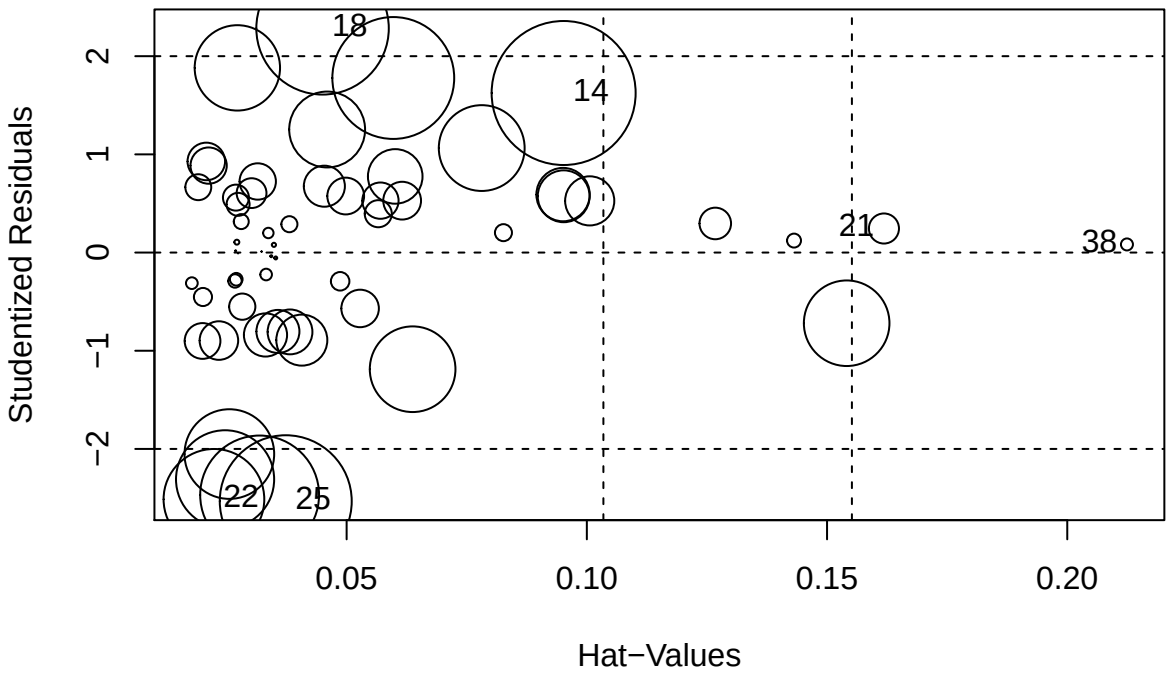



```
plot(rline2,which=3:4)
```


Scale−Location
18
12
14
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
4
6
8
10
12
Fitted values
lm(opioid$Death.Rate ~ opioid$Median.Income + opioid$Home.Price)




```
# for transform model with sqrt(y)

influencePlot(rline_transform)
```

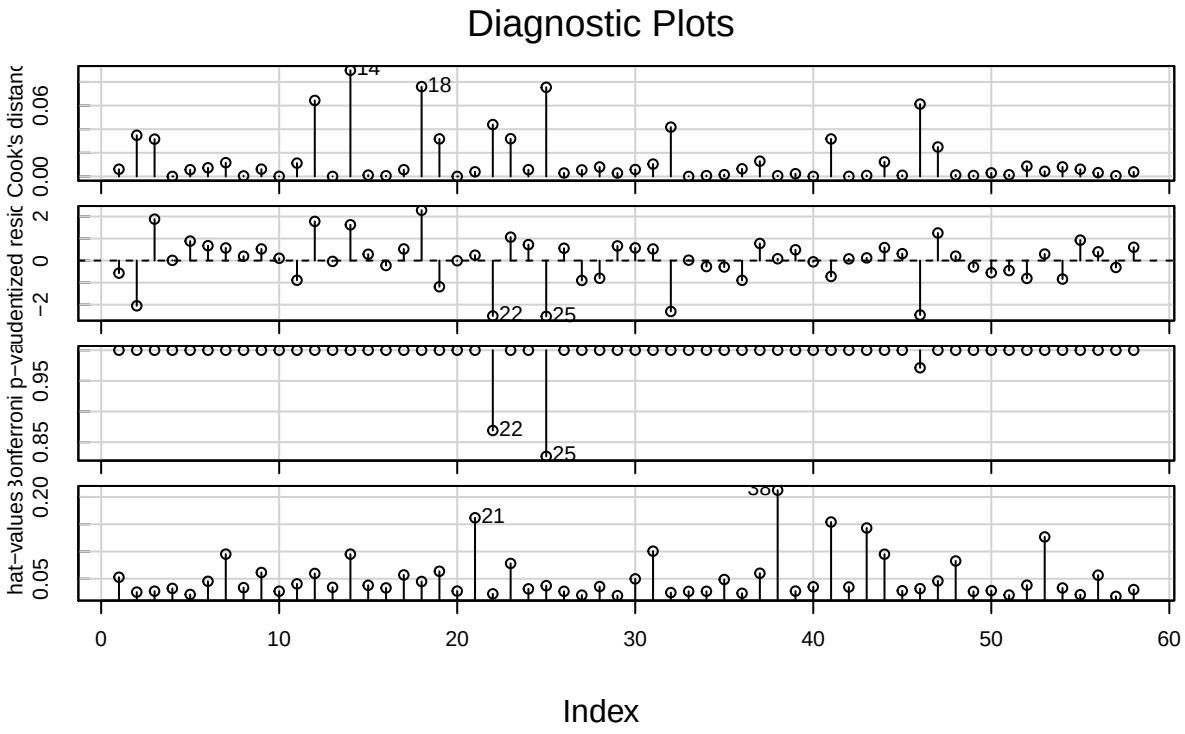



```
##        StudRes        Hat        CookD
## 14  1.62488000 0.09516861 0.0898845475
## 18  2.28330953 0.04498952 0.0760419844
## 21  0.24483921 0.16183919 0.0039254017
## 22 -2.51314678 0.02237902 0.0439456351
## 25 -2.53276902 0.03732672 0.0754795481
## 38  0.08255752 0.21242300 0.0006240432
```

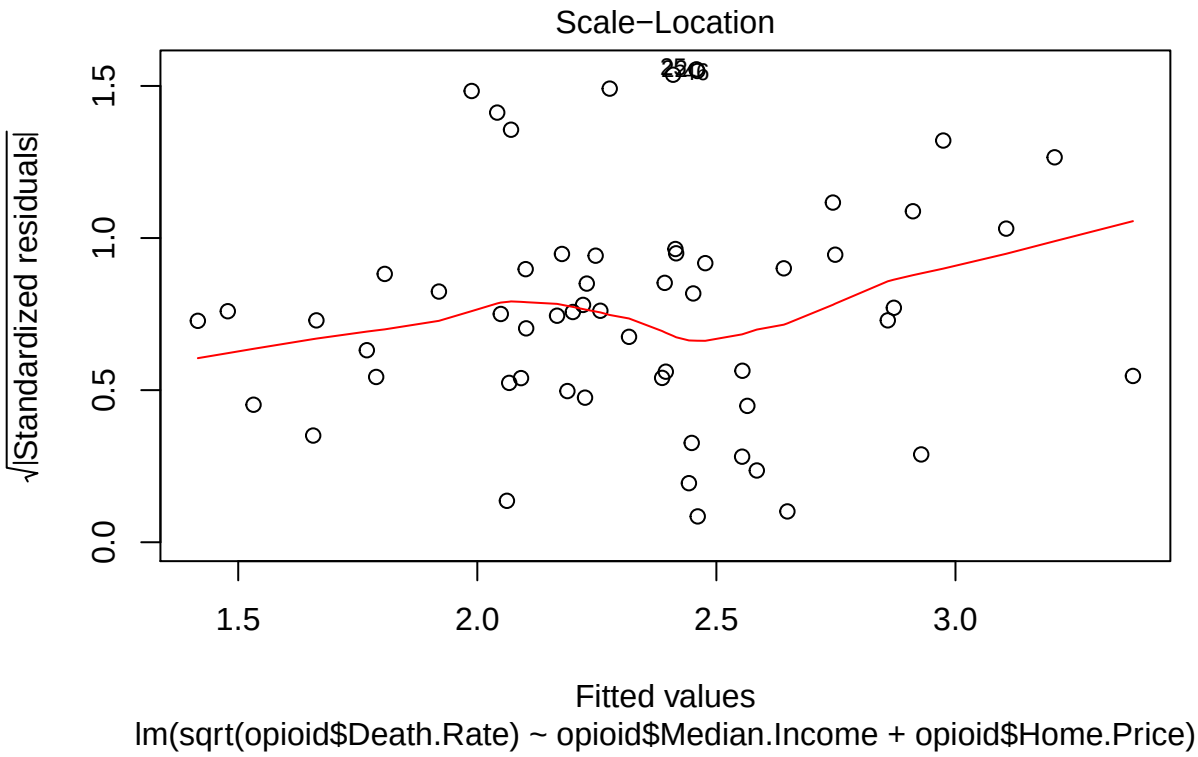
```
outlierTest(rline_transform)
```

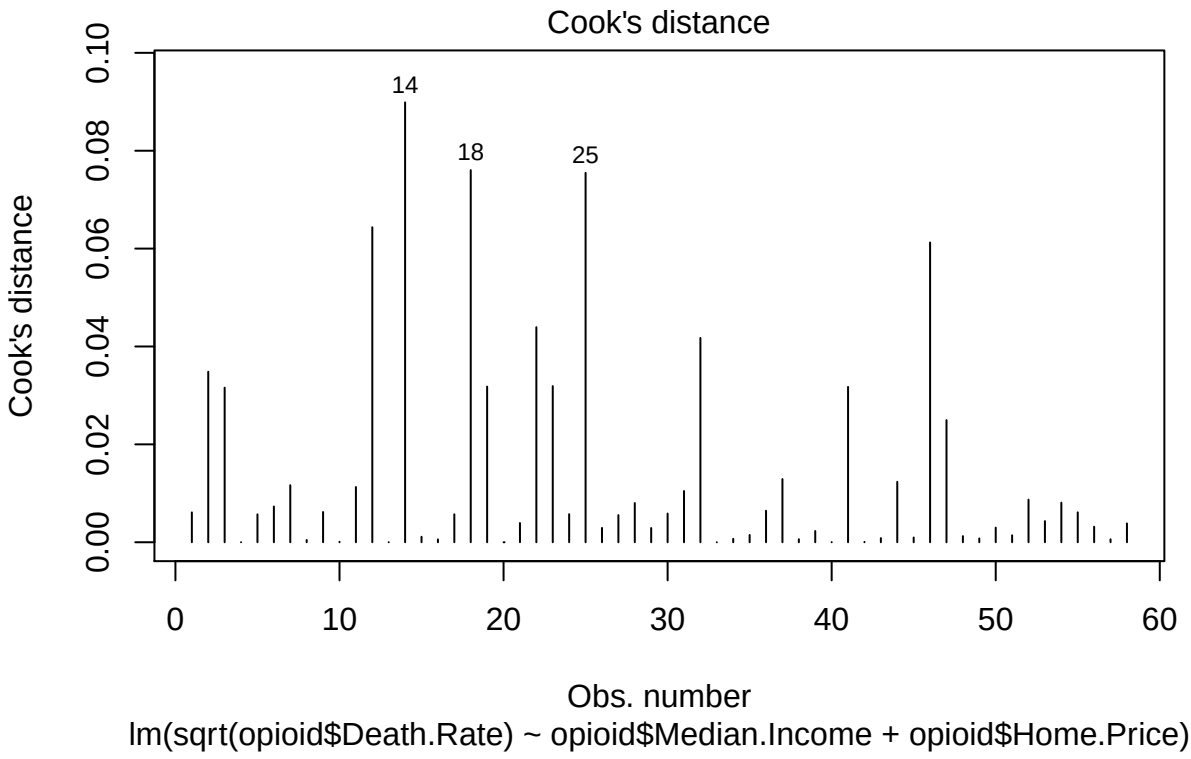
```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferroni p
## 25 -2.532769           0.014255       0.8268
```

```
influenceIndexPlot(rline_transform)
```



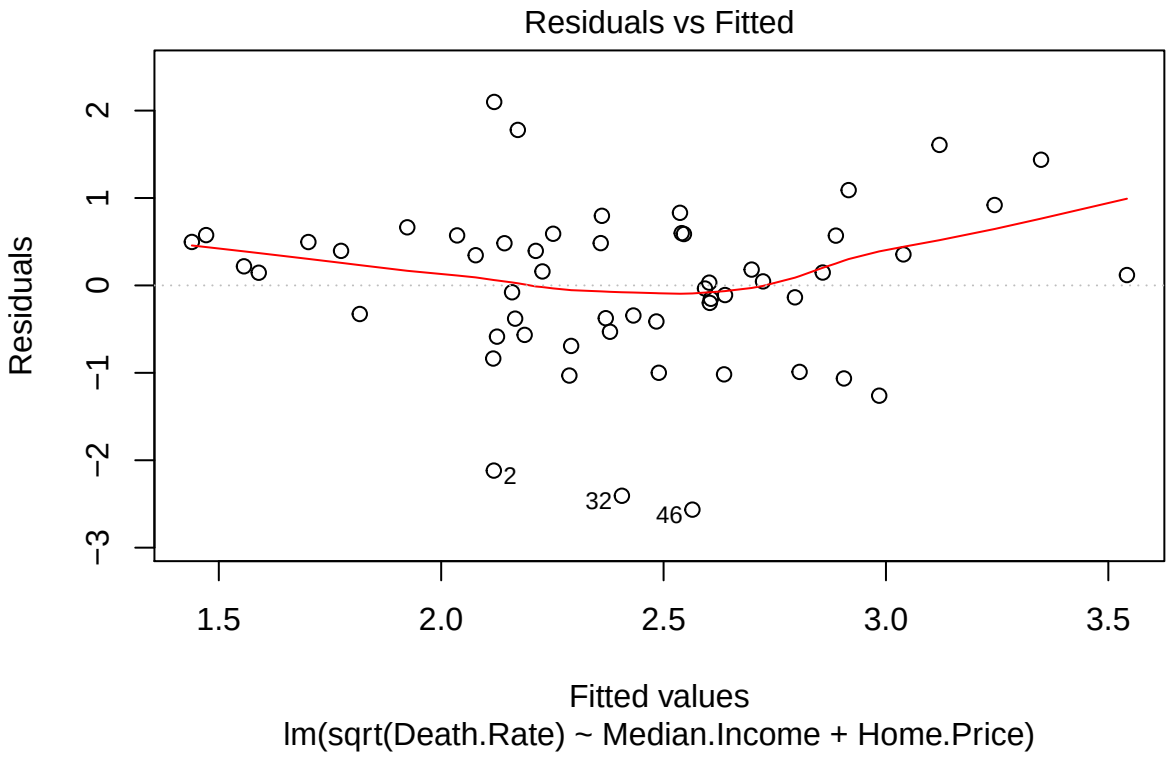

```
plot(rline_transform,which=3:4)
```

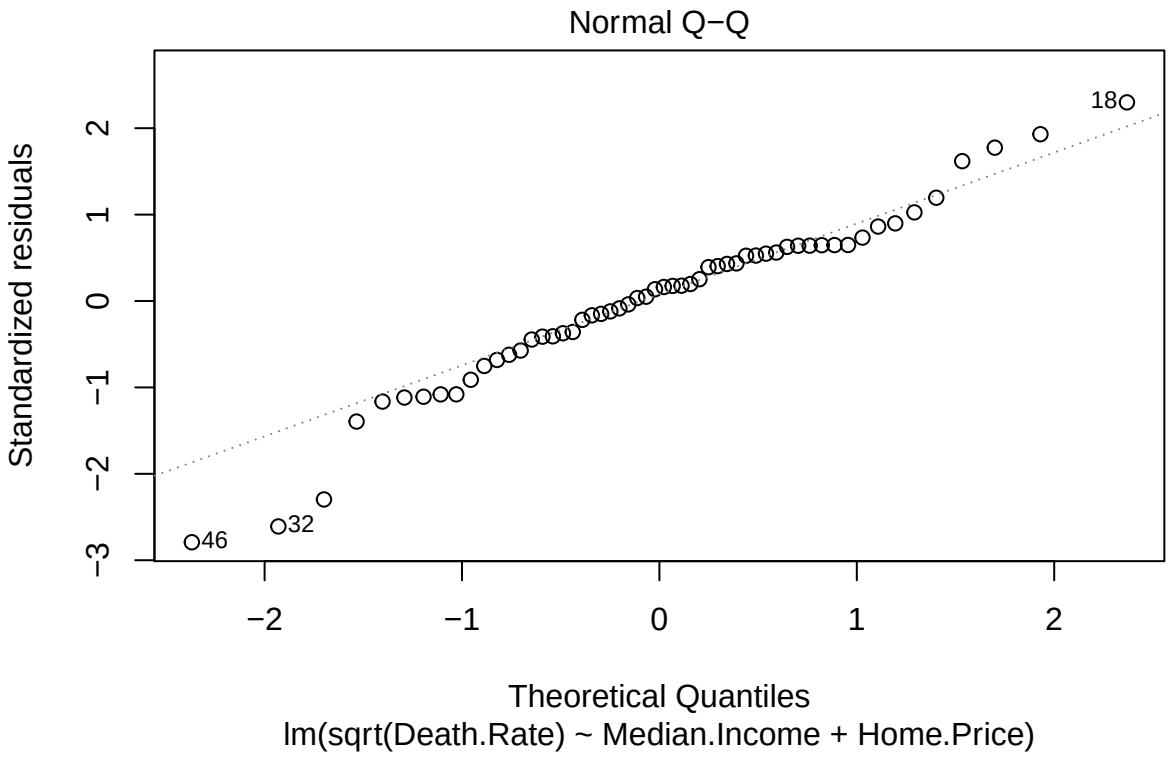

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
Fitted values
lm(sqrt(opioid$Death.Rate) ~ opioid$Median.Income + opioid$Home.Price)

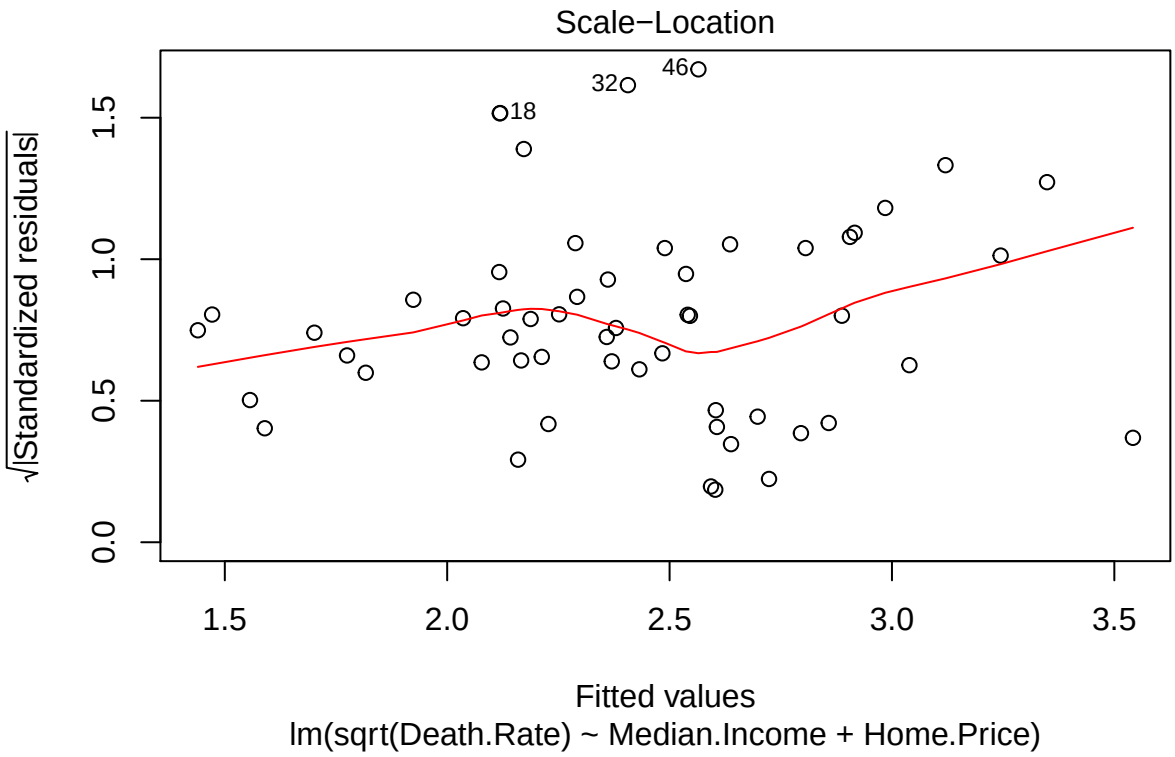


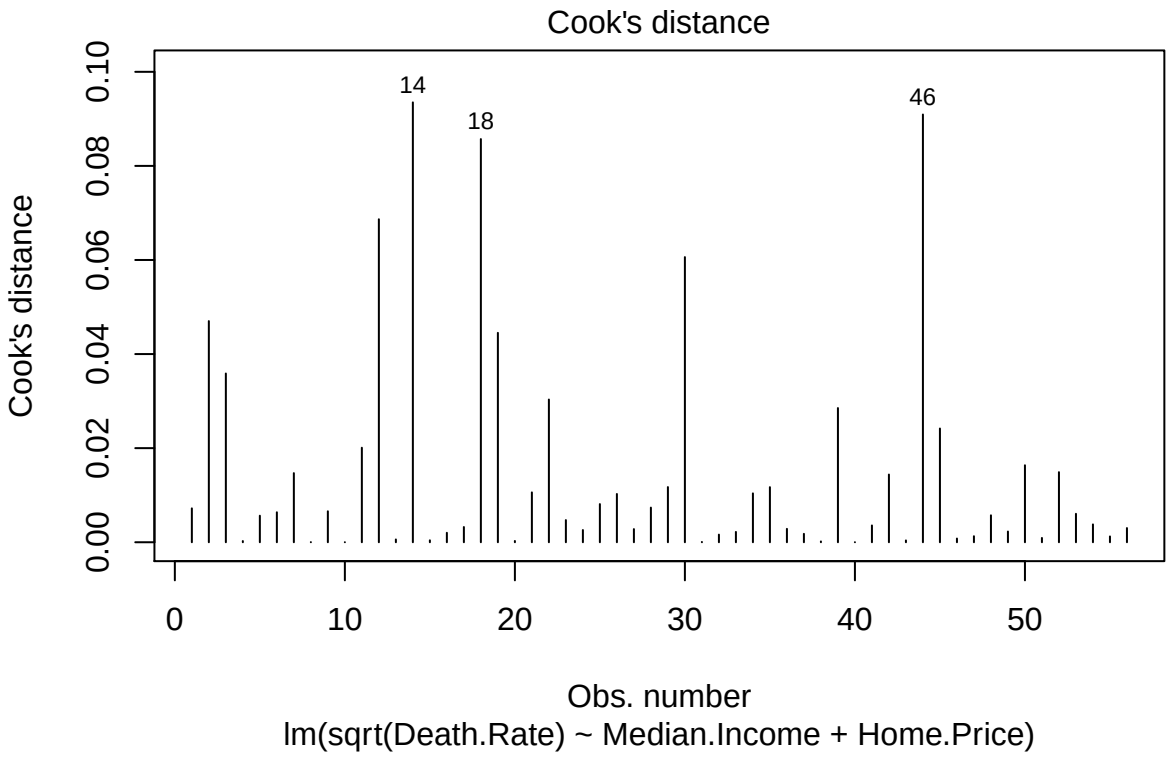

```
# 22 and 25 are significant outliers for transform model
```

```
# build transformation model without outliers and check assumptions


opioid_outliers<-opioid[-c(22,25),]

rline_outliers_transform<-lm(formula = sqrt(Death.Rate) ~ Median.Income+Home.Price, data=opioid_outliers

plot(rline_outliers_transform,which = 1:4)
```

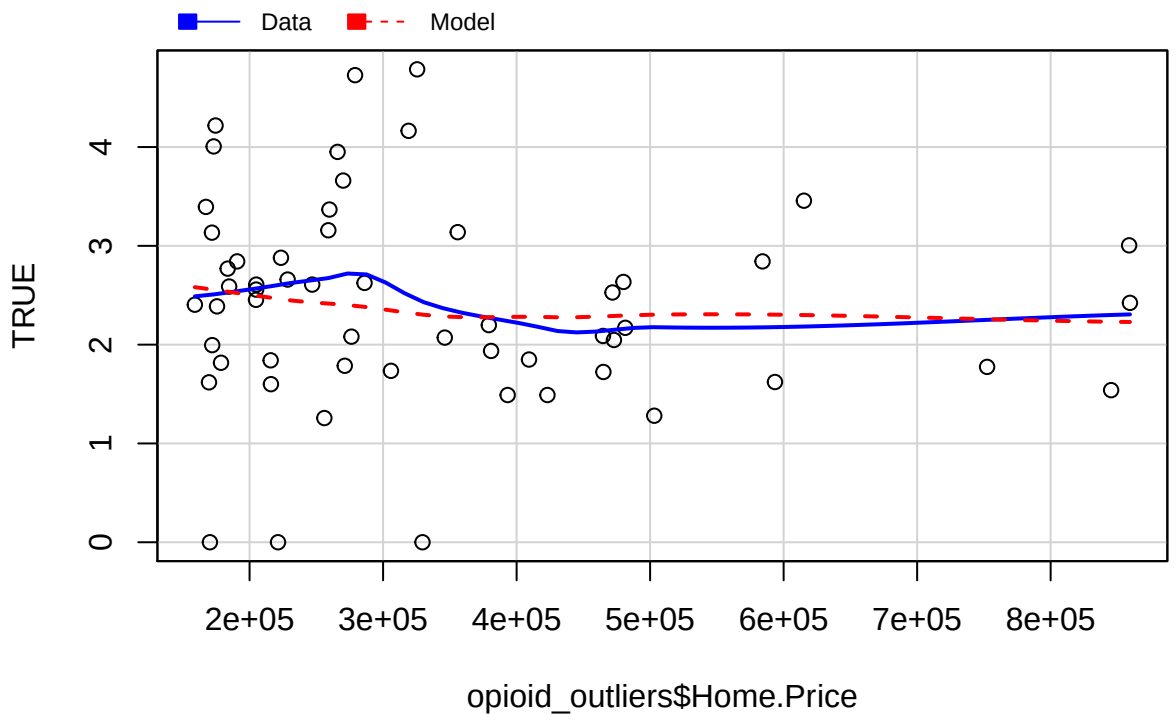

Residuals vs Fitted
Residuals
2
1
0
−1
−2
−3
1.5
2.0
2.5
3.0
3.5
2
32
46
Fitted values
lm(sqrt(Death.Rate) ~ Median.Income + Home.Price)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(sqrt(Death.Rate) ~ Median.Income + Home.Price)
18
32
46
−3
−2
−1
0
1
2

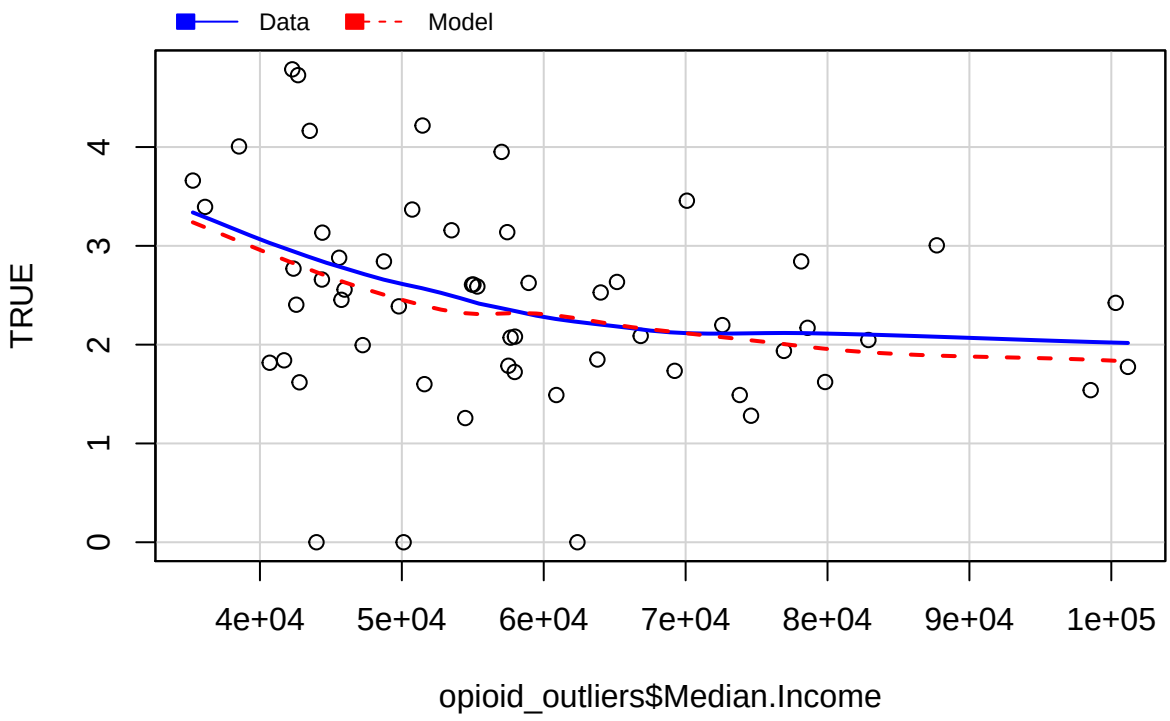

Scale−Location
46
32
18
√|Standardized residuals|
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
3.5
Fitted values
lm(sqrt(Death.Rate) ~ Median.Income + Home.Price)




```
mmp(rline_outliers_transform,opioid_outliers$Home.Price)
```




```
mmp(rline_outliers_transform,opioid_outliers$Median.Income)
```



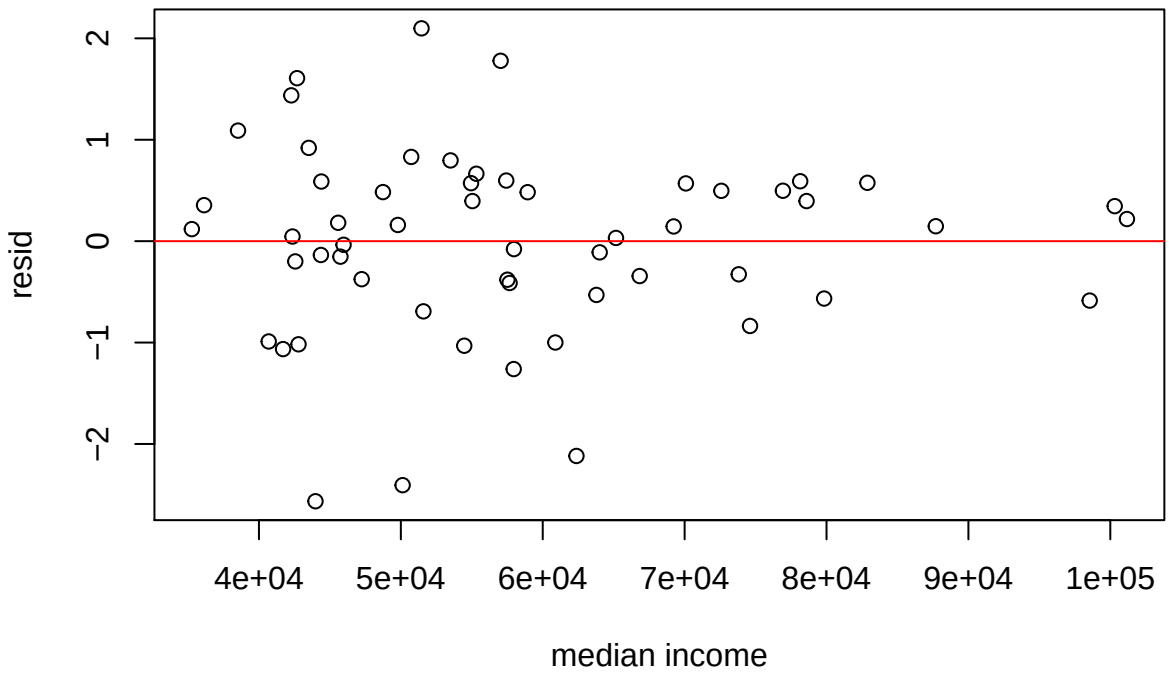

```
shapiro.test(rline_outliers_transform$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rline_outliers_transform$residuals
## W = 0.96192, p-value = 0.07443
```

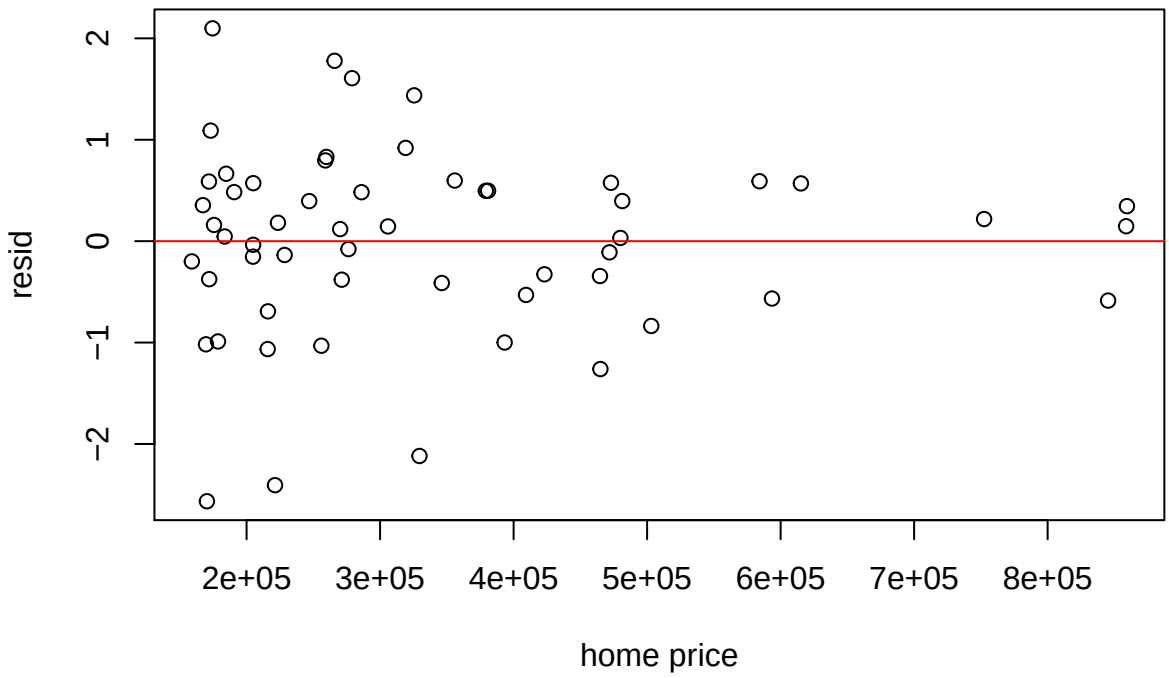
```
ncvTest(rline_outliers_transform)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.6681621, Df = 1, p = 0.41369
```

```
e3=residuals(rline_outliers_transform)
plot(opioid_outliers$Median.Income, e3, xlab = "median income", ylab =
             "resid")
abline(0,0, col="red")
```




```
plot(opioid_outliers$Home.Price, e3, xlab = "home price", ylab =
              "resid")
abline(0,0, col="red")
```




```
# we could see that both normality and costant variance conditions pass.
```

### Summary the final model

```
summary(rline_outliers_transform)
```

```
##
## Call:
## lm(formula = sqrt(Death.Rate) ~ Median.Income + Home.Price, data = opioid_outliers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5648 -0.4420  0.1319  0.5701  2.0986
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.552e+00  6.098e-01   7.465 8.06e-10 ***
## Median.Income -6.209e-05  1.752e-05  -3.544 0.000832 ***
## Home.Price     4.368e-06  1.554e-06   2.810 0.006916 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9346 on 53 degrees of freedom
## Multiple R-squared:  0.2001, Adjusted R-squared:  0.1699
## F-statistic: 6.629 on 2 and 53 DF,  p-value: 0.002694
```

```
# R^2 is 20% and p-value is 0.002694 which is small enough
```

```
# Design and implement linear regression modeling to perform statistical analysis for opioid epidemic a
# Collect and extract data from government websites, process raw data for regression, handle missing va
# Perform the statistical analyses including linear regression analysis, statistical modeling(GLM, Logi
```

### Future suggestions

```
# use VIF to check multicollinearity
# collect more data to emprove the R^2 such as use city replace county
# could use million unit to interprate the correlation since the correlation is very small
```

```
vif(rline_outliers_transform)
```

```
## Median.Income    Home.Price
##      5.122854      5.122854
```

```
# we see that the VIF is both smaller than 10, then the multicollinearity is not very serious.
```