R Notebook

Code

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Hide

```
library(openxlsx)
xlsxFile <- system.file("extdata","data_table_B15.xlsx", package = "openxlsx")
df <- read.xlsx(xlsxFile = "data_table_B15.xlsx", sheet = 1, skipEmptyRows = FALSE)
df
```

	City <chr>	MORT <dbl>	PRECIP <dbl>	EDUC <dbl>	NONWHITE <dbl>	NOX <dbl>	SO2 <dbl>
1	San Jose, CA	790.73	13	12.2	3.0	32	3
2	Wichita, KS	823.76	28	12.1	7.5	2	1
3	San Diego, CA	839.71	10	12.1	5.9	66	20
4	Lancaster, PA	844.05	43	9.5	2.9	7	32
5	Minneapolis, MN	857.62	25	12.1	3.0	11	26
6	Dallas, TX	860.10	35	11.8	14.8	1	1
7	Miami, FL	861.44	60	11.5	11.5	1	1
8	Los Angeles, CA	861.83	11	12.1	7.8	319	130
9	Grand Rapids, MI	871.34	31	10.9	5.1	3	10
10	Denver, CO	871.77	15	12.2	4.7	8	28

1-10 of 60 rows Previous **1** 2 3 4 5 6 Next

Hide

```
ml <- lm(MORT~PRECIP+EDUC+NONWHITE+NOX+SO2, data=df)
summary(ml)
```

```
Call:
lm(formula = MORT ~ PRECIP + EDUC + NONWHITE + NOX + SO2, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-91.38 -18.97  -3.56  16.00  91.83

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 995.63646   91.64099  10.865 3.35e-15 ***
PRECIP        1.40734    0.68914   2.042 0.046032 *
EDUC        -14.80139    7.02747  -2.106 0.039849 *
NONWHITE      3.19909    0.62231   5.141 3.89e-06 ***
NOX          -0.10797    0.13502  -0.800 0.427426
SO2           0.35518    0.09096   3.905 0.000264 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.09 on 54 degrees of freedom
Multiple R-squared:  0.6746,    Adjusted R-squared:  0.6444
F-statistic: 22.39 on 5 and 54 DF,  p-value: 4.407e-12
```

Hide

```
#anova(ml)
```

Hide

```
AIC(ml)
```

```
[1] 611.5551
```

Hide

```
BIC(ml)
```

```
[1] 626.2155
```

Hide

```
confint(ml,level=0.95)
```

Hide

```
#install.packages("olsrr")
library(olsrr)
K=ols_step_all_possible(ml)
K[order(K$rsquare),]
```

	Index	N	Predictors	R-Square	Adj. R-Square	Mallow's Cp
	<int>	<int>	<chr>	<dbl>	<dbl>	<dbl>
4	5	1	NOX	0.005731729	-0.01141083	108.980121
5	4	1	SO2	0.181435802	0.16732263	79.825334
15	15	2	NOX SO2	0.257386457	0.23132984	69.222752
1	3	1	PRECIP	0.259582453	0.24681663	66.858367
2	2	1	EDUC	0.261104467	0.24836489	66.605818
11	14	2	EDUC NOX	0.262948911	0.23708747	68.299767
8	13	2	PRECIP NOX	0.298159792	0.27353382	62.457183
6	12	2	PRECIP EDUC	0.349356056	0.32652644	53.962126
12	11	2	EDUC SO2	0.360314920	0.33786983	52.143709
24	25	3	EDUC NOX SO2	0.376902004	0.34352175	51.391394

1-10 of 31 rows Previous **1** 2 3 4 Next

Hide

```
K2=ols_step_best_subset(ml)
K2
```

```
          Best Subsets Regression
-----------------------------------------
Model Index    Predictors
-----------------------------------------
     1         NONWHITE
     2         EDUC NONWHITE
     3         PRECIP NONWHITE SO2
     4         PRECIP EDUC NONWHITE SO2
     5         PRECIP EDUC NONWHITE NOX SO2
-----------------------------------------

                                                        Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
                      Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC        SBC           M
SEP           FPE         HSP       APC
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
  1        0.4193      0.4092      0.3765    40.3644    638.3043    466.2374    644.5873    1371
44.8298    2361.8927    40.1000    0.6208
  2        0.5668      0.5516      0.5122    17.8851    622.7202    451.4009    631.0975    1041
32.9791    1821.7307    30.9818    0.4788
  3        0.6406      0.6213      0.5808     7.6416    613.5172    443.2926    623.9889     879
67.8598    1562.8627    26.6397    0.4108
  4        0.6707      0.6468       0.598     4.6394    610.2614    440.9506    622.8274     820
82.3794    1480.5959    25.3093    0.3892
  5        0.6746      0.6444      0.5933     6.0000    611.5551    442.5911    626.2155     826
52.4128    1513.2955    25.9570    0.3978
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria
```

Hide

```
#plot(K)
```

Hide

```
K2
```

```
          Best Subsets Regression
-----------------------------------------
Model Index    Predictors
-----------------------------------------
     1         NONWHITE
     2         EDUC NONWHITE
     3         PRECIP NONWHITE SO2
     4         PRECIP EDUC NONWHITE SO2
     5         PRECIP EDUC NONWHITE NOX SO2
-----------------------------------------

                                                        Subsets Regression Summary
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC        SBC           M
SEP           FPE         HSP       APC
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
  1        0.4193      0.4092      0.3765    40.3644    638.3043    466.2374    644.5873    1371
44.8298    2361.8927    40.1000    0.6208
  2        0.5668      0.5516      0.5122    17.8851    622.7202    451.4009    631.0975    1041
32.9791    1821.7307    30.9818    0.4788
  3        0.6406      0.6213      0.5808     7.6416    613.5172    443.2926    623.9889     879
67.8598    1562.8627    26.6397    0.4108
  4        0.6707      0.6468       0.598     4.6394    610.2614    440.9506    622.8274     820
82.3794    1480.5959    25.3093    0.3892
  5        0.6746      0.6444      0.5933     6.0000    611.5551    442.5911    626.2155     826
52.4128    1513.2955    25.9570    0.3978
-------------------------------------------------------------------------------------------------------------
-----------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria
```

Hide

```
##forwards

K3=ols_step_forward_p(ml,pent=0.05,progress=TRUE,details=TRUE)
```

```
Forward Selection Method
--------------------------

Candidate Terms:

1. PRECIP
2. EDUC
3. NONWHITE
4. NOX
5. SO2

We are selecting variables based on p value...


Forward Selection: Step 1

+ NONWHITE

                         Model Summary
-----------------------------------------------------------------
R                       0.647       RMSE                 47.809
R-Squared               0.419       Coef. Var             5.084
Adj. R-Squared          0.409       MSE                2285.703
Pred R-Squared          0.376       MAE                  37.873
-----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
---------------------------------------------------------------------------
                 Sum of
                Squares        DF    Mean Square      F         Sig.
---------------------------------------------------------------------------
Regression     95704.645         1      95704.645    41.871    0.0000
Residual      132570.751        58       2285.703
Total         228275.396        59
---------------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
-----------------------------------------------------------------------------------------------
(Intercept)    886.722        10.334                85.803    0.000    866.035    907.408
   NONWHITE      4.525         0.699        0.647     6.471    0.000      3.125      5.925
-----------------------------------------------------------------------------------------------



Forward Selection: Step 2

+ EDUC
```

```
                        Model Summary
----------------------------------------------------------------
R                       0.753       RMSE                41.653
R-Squared               0.567       Coef. Var            4.430
Adj. R-Squared          0.552       MSE               1734.982
Pred R-Squared          0.512       MAE                 31.977
----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
--------------------------------------------------------------------------
                 Sum of
                Squares        DF    Mean Square      F         Sig.
--------------------------------------------------------------------------
Regression    129381.442         2      64690.721    37.286    0.0000
Residual       98893.954        57       1734.982
Total         228275.396        59
--------------------------------------------------------------------------

                                    Parameter Estimates
-----------------------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
-----------------------------------------------------------------------------------------------------
(Intercept)    1210.681        74.081                16.343    0.000    1062.337    1359.025
   NONWHITE       3.951         0.623        0.565     6.342    0.000       2.703       5.199
       EDUC     -28.902         6.560       -0.393    -4.406    0.000     -42.039     -15.766
-----------------------------------------------------------------------------------------------------



Forward Selection: Step 3

+ SO2

                        Model Summary
----------------------------------------------------------------
R                       0.794       RMSE                38.847
R-Squared               0.630       Coef. Var            4.131
Adj. R-Squared          0.610       MSE               1509.061
Pred R-Squared          0.572       MAE                 28.772
----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
--------------------------------------------------------------------------
                 Sum of
                Squares        DF    Mean Square      F         Sig.
```

```
-----------------------------------------------------------------------
Regression    143767.961         3      47922.654    31.757    0.0000
Residual       84507.435        56       1509.061
Total         228275.396        59
-----------------------------------------------------------------------

                                Parameter Estimates
-----------------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
-----------------------------------------------------------------------------------------------
(Intercept)    1155.479        71.365                16.191    0.000    1012.518    1298.440
   NONWHITE       3.736         0.585        0.535     6.385    0.000       2.564       4.908
       EDUC     -24.890         6.255       -0.338    -3.979    0.000     -37.420     -12.361
        SO2       0.255         0.083        0.260     3.088    0.003       0.090       0.421
-----------------------------------------------------------------------------------------------



Forward Selection: Step 4

+ PRECIP

                        Model Summary
-----------------------------------------------------------------
R                       0.819       RMSE                 36.969
R-Squared               0.671       Coef. Var             3.931
Adj. R-Squared          0.647       MSE                1366.704
Pred R-Squared          0.598       MAE                  26.016
-----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
-----------------------------------------------------------------------
                 Sum of
                Squares        DF    Mean Square      F         Sig.
-----------------------------------------------------------------------
Regression    153106.682         4      38276.670    28.007    0.0000
Residual       75168.714        55       1366.704
Total         228275.396        59
-----------------------------------------------------------------------

                                Parameter Estimates
-----------------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
-----------------------------------------------------------------------------------------------
(Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
   NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
       EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
        SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
     PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
```

```
-------------------------------------------------------------------------------------------



No more variables to be added.

Variables Entered:

+ NONWHITE
+ EDUC
+ SO2
+ PRECIP


Final Model Output
------------------

                         Model Summary
-----------------------------------------------------------------
R                       0.819       RMSE                 36.969
R-Squared               0.671       Coef. Var             3.931
Adj. R-Squared          0.647       MSE                1366.704
Pred R-Squared          0.598       MAE                  26.016
-----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
--------------------------------------------------------------------------
                 Sum of
                Squares        DF    Mean Square      F         Sig.
--------------------------------------------------------------------------
Regression    153106.682         4      38276.670    28.007    0.0000
Residual       75168.714        55       1366.704
Total         228275.396        59
--------------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------------
      model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
-----------------------------------------------------------------------------------------------
(Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
   NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
       EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
        SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
     PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
-----------------------------------------------------------------------------------------------
```

Hide

```
K3
```

```
                         Selection Summary
--------------------------------------------------------------------------
        Variable                 Adj.
Step    Entered     R-Square    R-Square     C(p)        AIC        RMSE
--------------------------------------------------------------------------
   1    NONWHITE      0.4193      0.4092    40.3644    638.3043    47.8090
   2    EDUC          0.5668      0.5516    17.8851    622.7202    41.6531
   3    SO2           0.6298      0.6100     9.4276    615.2876    38.8466
   4    PRECIP        0.6707      0.6468     4.6394    610.2614    36.9690
--------------------------------------------------------------------------
```

Hide

```
##backwards


K5=ols_step_backward_p(ml,prem=0.05)   ### BASED ON P-VALUES ###
K5
```

```
                         Elimination Summary
-------------------------------------------------------------------------
        Variable                 Adj.
Step    Removed     R-Square    R-Square     C(p)       AIC        RMSE
-------------------------------------------------------------------------
   1    NOX           0.6707      0.6468    4.6394    610.2614    36.9690
-------------------------------------------------------------------------
```

Hide

```
# stepwise using p value
K7=ols_step_both_p(ml, details = TRUE, pent=0.05, prem=0.1)
K7
```

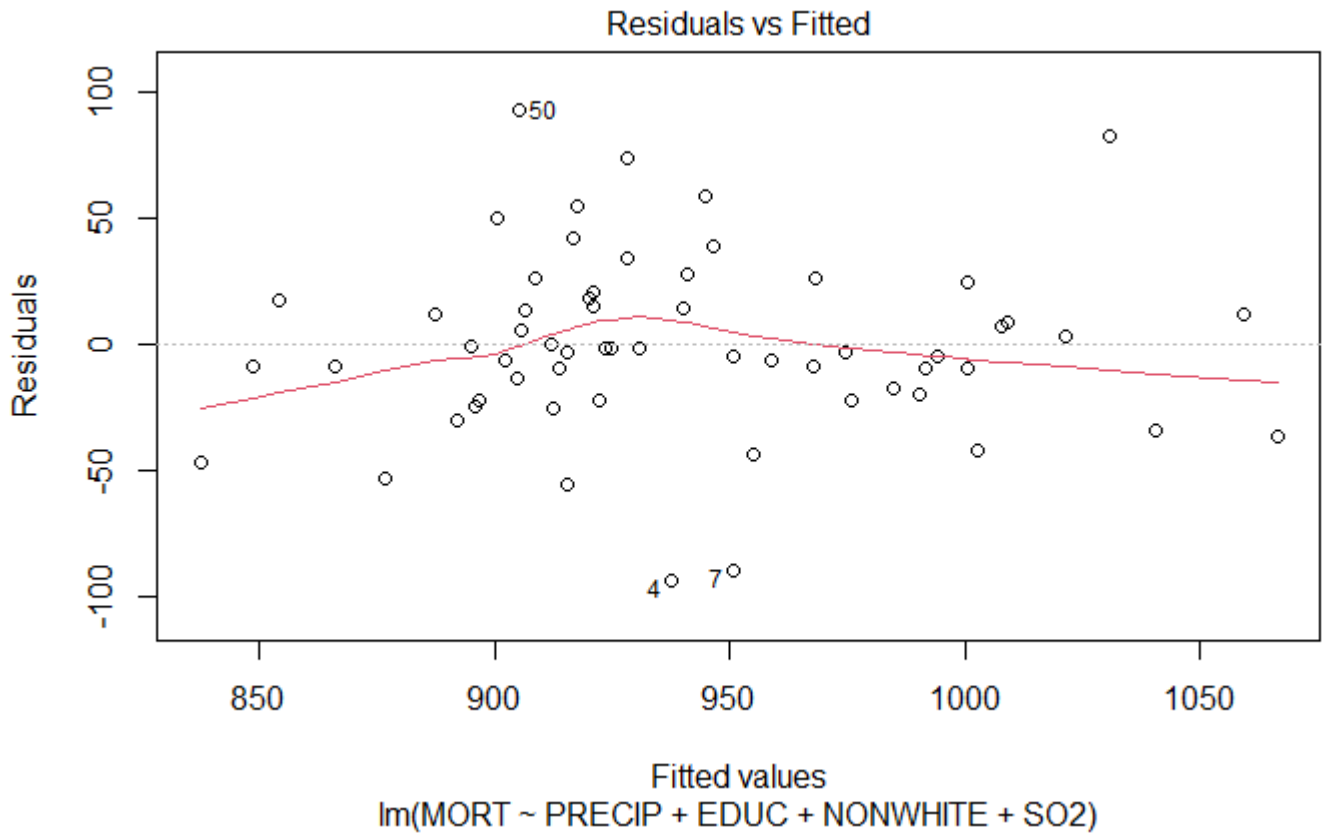
Hide

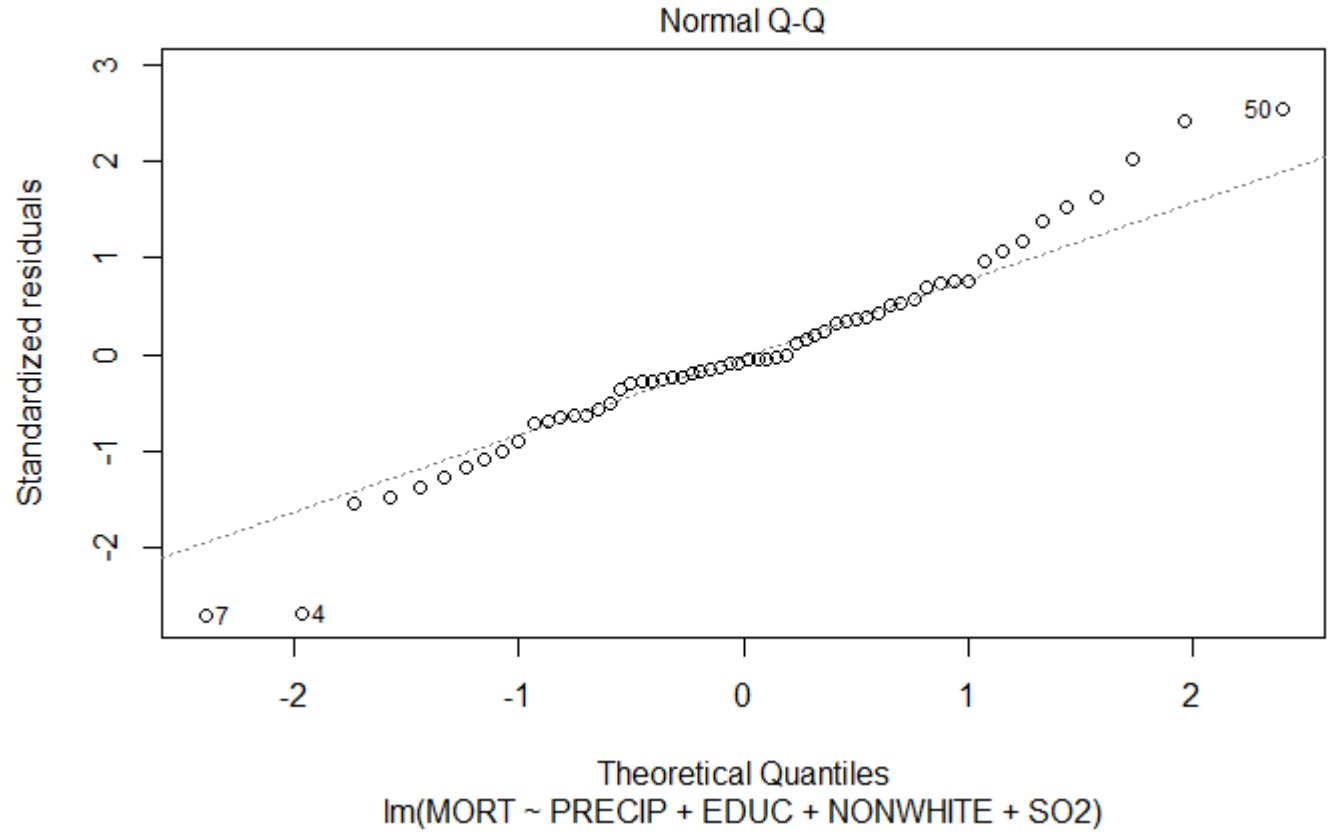
```
ml <- lm(MORT~PRECIP+EDUC+NONWHITE+SO2, data=df)
summary(ml)
```

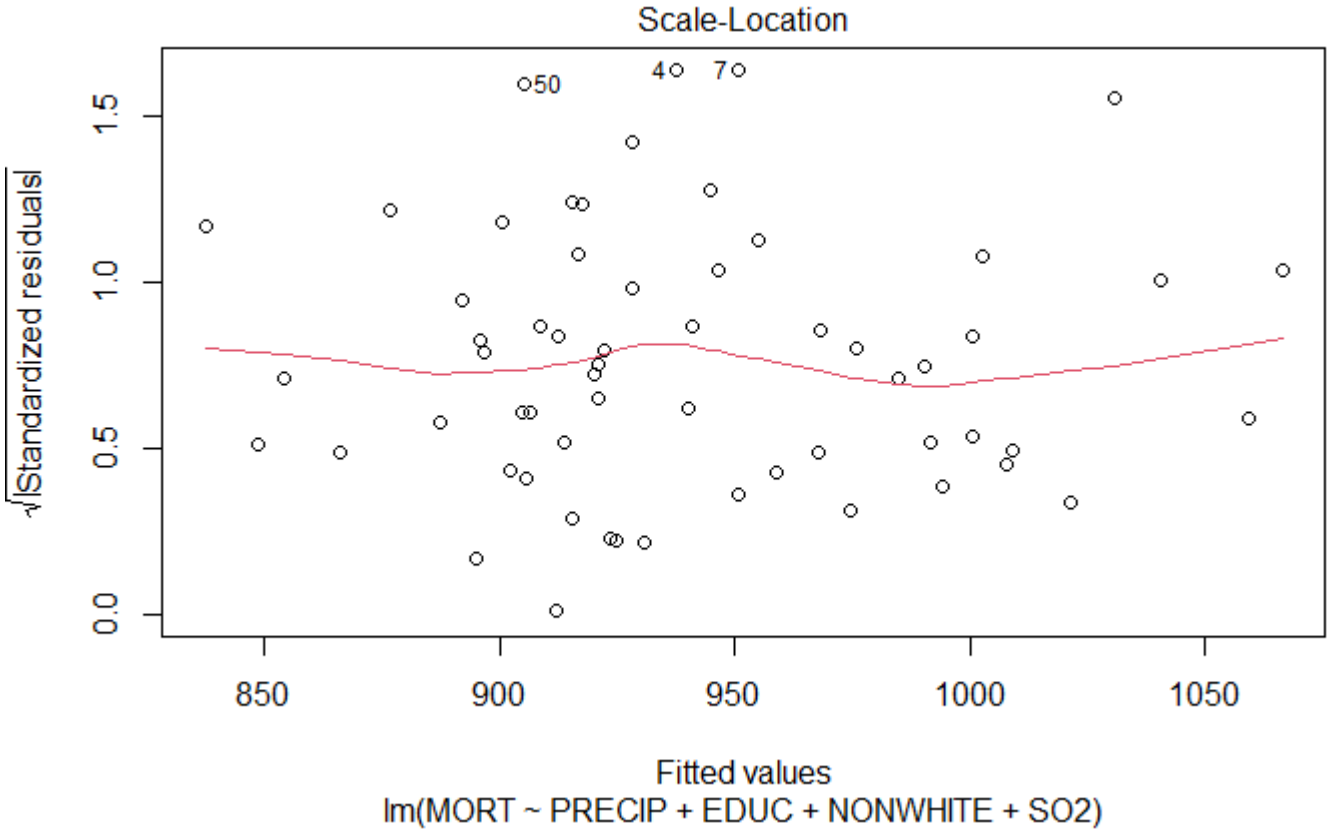
```
Call:
lm(formula = MORT ~ PRECIP + EDUC + NONWHITE + SO2, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-93.600 -20.499  -2.443  17.891  92.521

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 995.82238   91.33980  10.902 2.31e-15 ***
PRECIP        1.63505    0.62550   2.614 0.011522 *
EDUC        -15.56968    6.93862  -2.244 0.028883 *
NONWHITE      3.09979    0.60779   5.100 4.33e-06 ***
SO2           0.32634    0.08323   3.921 0.000247 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.97 on 55 degrees of freedom
Multiple R-squared:  0.6707,    Adjusted R-squared:  0.6468
F-statistic: 28.01 on 4 and 55 DF,  p-value: 1.052e-12
```

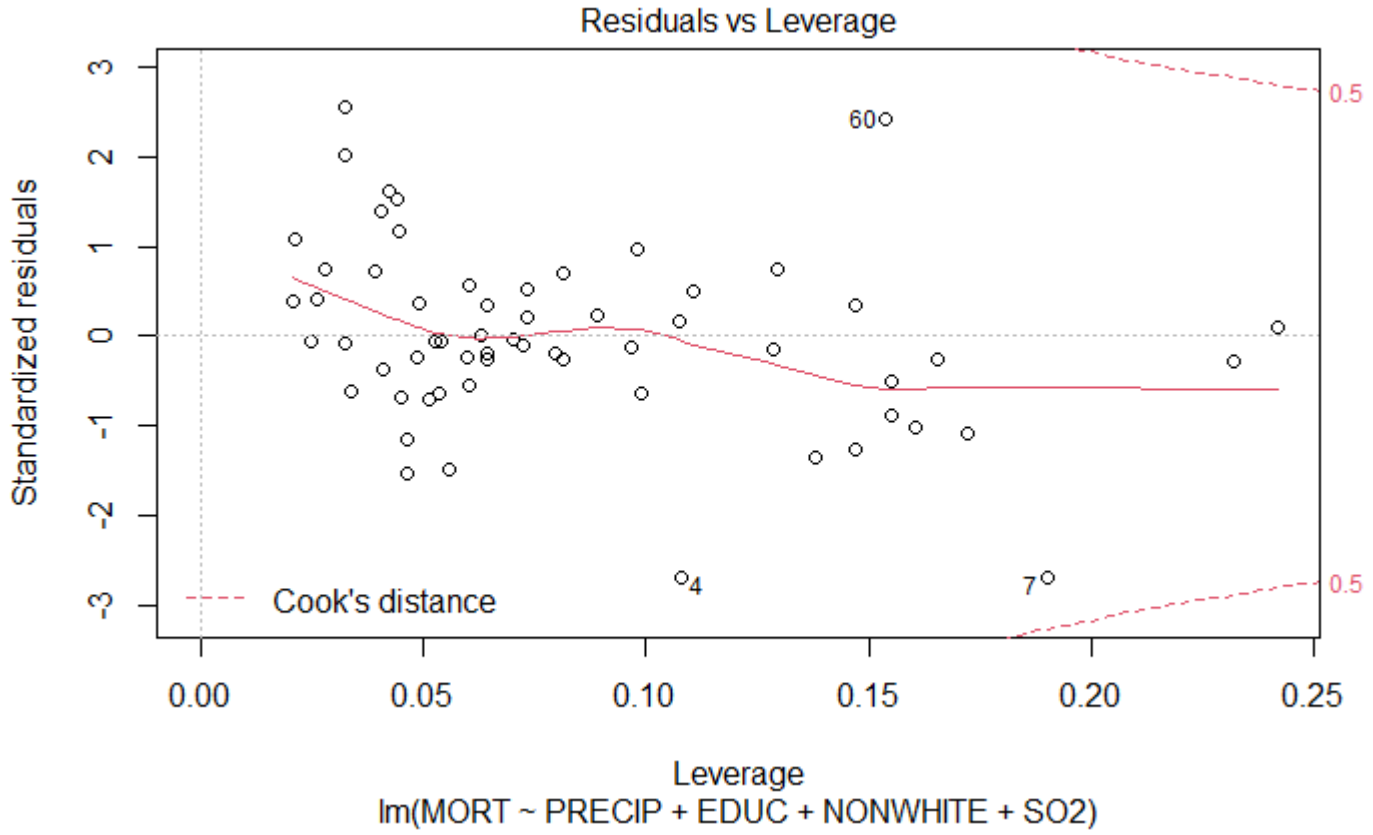
Hide

```
plot(ml)
```




Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
50
7
4
-2
-1
0
1
2
Theoretical Quantiles
lm(MORT ~ PRECIP + EDUC + NONWHITE + SO2)


Scale-Location
√|Standardized residuals|
1.5
1.0
0.5
0.0
50
4
7
850
900
950
1000
1050
Fitted values
lm(MORT ~ PRECIP + EDUC + NONWHITE + SO2)



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.