

ST362 Final project - Crime Data

Diya Bajaj, Joel Huebner

2025-07-28

DATA

```
R <- c(79.1,163.5,57.8,196.9,123.4,68.2,96.3,155.5,85.6,70.5,
      167.4,84.9,51.1,66.4,79.8,94.6,53.9,92.9,75.0,122.5,
      74.2,43.9,121.6,96.8,52.3,199.3,34.2,121.6,104.3,69.6,
      37.3,75.4,107.2,92.3,65.3,127.2,83.1,56.6,82.6,115.1,
      88.0,54.2,82.3,103.0,45.5,50.8,84.9)
Age <- c(151,143,142,136,141,121,127,131,157,140,
        124,134,128,135,152,142,143,135,130,125,
        126,157,132,131,130,131,135,152,119,166,
        140,125,147,126,123,150,177,133,149,145,
        148,141,162,136,139,126,130)
S <- c(1,0,1,0,0,0,1,1,1,0,
      0,0,0,0,1,1,0,1,0,0,
      0,1,0,0,0,0,0,0,0,1,
      0,0,1,0,0,0,1,0,1,1,
      0,0,1,0,1,0,0)
Ed <- c(91,113,89,121,121,110,111,109,90,118,
      105,108,113,117,87,88,110,104,116,108,
      108,89,96,116,116,121,109,112,107,89,
      93,109,104,118,102,100,NA,104,88,104,
      122,109,99,121,88,104,121)
Ex0 <- c(58,103,45,149,109,118,82,115,65,71,
      121,75,67,62,NA,81,66,123,128,113,
      74,47,87,78,63,160,69,82,166,58,
      55,90,63,97,97,109,58,51,61,82,
      72,56,75,95,46,106,90)
Ex1 <- c(56,95,44,141,101,115,79,109,62,68,
      116,71,60,61,53,77,63,115,128,105,
      67,44,83,73,57,143,71,76,157,54,
      54,81,64,97,87,98,56,47,54,74,
      66,54,70,96,41,97,91)
LF <- c(510,583,533,577,591,547,519,542,553,632,
      580,595,624,595,530,497,537,537,536,567,
      602,512,564,574,641,631,540,571,521,521,
      535,586,560,542,526,531,638,599,515,560,
      601,523,522,574,480,599,623)
M <- c(950,1012,969,994,985,964,982,969,955,1029,
      966,972,972,986,986,956,977,978,934,985,
      NA,962,953,1038,984,1071,965,1018,938,973,
      1045,964,972,990,948,964,974,1024,953,981,
      998,968,996,1012,968,989,1049)
```

```

N <- c(33,13,18,157,18,25,4,50,39,7,
      101,47,28,22,30,33,10,31,51,78,
      34,22,43,7,14,3,6,10,168,46,
      6,97,23,18,113,9,24,7,36,96,
      9,4,40,29,19,40,3)
NW <- c(301,102,219,80,30,44,139,179,286,15,
      106,59,10,46,72,321,6,170,24,94,
      12,423,92,36,26,77,4,79,89,254,
      20,82,95,21,76,24,349,40,165,126,
      19,2,208,36,49,24,22)
U1 <- c(108,96,94,102,91,84,97,79,81,100,
      77,83,77,77,92,116,114,89,78,130,
      102,97,83,142,70,102,80,103,92,NA,
      135,105,76,102,124,87,76,99,86,88,
      84,107,73,111,135,78,113)
U2 <- c(41,36,33,39,20,29,38,35,28,24,
      35,31,25,27,43,47,35,34,34,58,
      33,34,32,42,21,41,22,28,36,26,
      40,43,24,35,50,38,28,27,35,31,
      20,37,27,37,53,25,40)
W <- c(394,557,318,673,578,689,620,472,421,526,
      657,580,507,529,405,427,487,631,627,626,
      557,288,513,540,486,674,564,537,637,396,
      453,617,462,589,572,559,382,425,395,488,
      590,489,496,622,457,593,588)
X <- c(261,194,250,167,174,126,168,206,239,174,
      170,172,206,190,264,247,166,165,135,166,
      195,276,227,176,196,152,139,215,154,237,
      200,163,233,NA,158,153,254,225,251,228,
      144,170,224,162,249,171,160)

```

```

#s indicator variable = 1 if southern state, 0 if not
crime = data.frame(R, Age, S, Ed, Ex0, Ex1, LF, M, N, NW, U1, U2, W, X)
head(crime)

```

```

##      R Age S  Ed Ex0 Ex1  LF    M    N  NW  U1 U2   W   X
## 1  79.1 151 1   91  58  56 510  950  33 301 108 41 394 261
## 2 163.5 143 0  113 103  95 583 1012  13 102  96 36 557 194
## 3  57.8 142 1   89  45  44 533  969  18 219  94 33 318 250
## 4 196.9 136 0  121 149 141 577  994 157  80 102 39 673 167
## 5 123.4 141 0  121 109 101 591  985  18  30  91 20 578 174
## 6  68.2 121 0  110 118 115 547  964  25  44  84 29 689 126

```

```

#note models will remove rows of data that include NA so we can clean data to get the data that will be
crime2 = na.omit(crime)
head(crime2)

```

```

##      R Age S  Ed Ex0 Ex1  LF    M    N  NW  U1 U2   W   X
## 1  79.1 151 1   91  58  56 510  950  33 301 108 41 394 261
## 2 163.5 143 0  113 103  95 583 1012  13 102  96 36 557 194
## 3  57.8 142 1   89  45  44 533  969  18 219  94 33 318 250
## 4 196.9 136 0  121 149 141 577  994 157  80 102 39 673 167

```

```
## 5 123.4 141 0 121 109 101 591 985 18 30 91 20 578 174
## 6 68.2 121 0 110 118 115 547 964 25 44 84 29 689 126
```

#i think we might need to use crime2 for part a to be representative of models in b that will exclude r

a.) Exploratory data analysis (e.g., summary statistics, graphical displays of the data such as boxplots, and preliminary conclusions. No models are assumed in this section).

```
summary(crime2)
```

```
##           R           Age           S           Ed
## Min.      : 34.20   Min.      :119.0   Min.      :0.0000   Min.      : 88.0
## 1st Qu.: 59.67   1st Qu.:130.0   1st Qu.:0.0000   1st Qu.:100.5
## Median : 84.90   Median :135.5   Median :0.0000   Median :108.5
## Mean     : 91.78   Mean     :137.3   Mean     :0.3095   Mean     :106.6
## 3rd Qu.:113.12   3rd Qu.:143.0   3rd Qu.:1.0000   3rd Qu.:115.2
## Max.     :199.30   Max.     :162.0   Max.     :1.0000   Max.     :122.0
##           Ex0          Ex1          LF          M
## Min.      : 45.00   Min.      : 41.00   Min.      :480.0   Min.      : 934.0
## 1st Qu.: 63.50   1st Qu.: 61.25   1st Qu.:531.5   1st Qu.: 964.0
## Median : 81.50   Median : 75.00   Median :560.0   Median : 974.5
## Mean     : 86.93   Mean     : 82.00   Mean     :560.5   Mean     : 983.2
## 3rd Qu.:108.25   3rd Qu.: 97.75   3rd Qu.:589.8   3rd Qu.: 995.5
## Max.     :166.00   Max.     :157.00   Max.     :641.0   Max.     :1071.0
##           N           NW           U1           U2
## Min.      : 3.00   Min.      : 2.00   Min.      : 70.00   Min.      :20.00
## 1st Qu.: 9.25   1st Qu.: 24.50   1st Qu.: 81.50   1st Qu.:27.25
## Median : 24.00   Median : 76.50   Median : 93.00   Median :34.50
## Mean     : 37.36   Mean     : 96.31   Mean     : 96.26   Mean     :34.10
## 3rd Qu.: 42.25   3rd Qu.:121.00   3rd Qu.:106.50   3rd Qu.:38.75
## Max.     :168.00   Max.     :423.00   Max.     :142.00   Max.     :58.00
##           W           X
## Min.      :288.0   Min.      :126.0
## 1st Qu.:475.5   1st Qu.:163.5
## Median :538.5   Median :174.0
## Mean     :532.5   Mean     :190.5
## 3rd Qu.:611.0   3rd Qu.:224.8
## Max.     :689.0   Max.     :276.0
```

```
sapply(crime2, sd)
```

```
##           R           Age           S           Ed           Ex0           Ex1           LF
## 40.6847243 10.4421419 0.4679011 10.5810296 30.4297615 28.5247005 39.6891595
##           M           N           NW           U1           U2           W           X
## 31.1178215 40.1234810 96.9529652 18.4113691 8.6920472 94.9110593 38.7539066
```

```
if(!require("corrgram")) {install.packages("corrgram", dependencies = TRUE)}
```

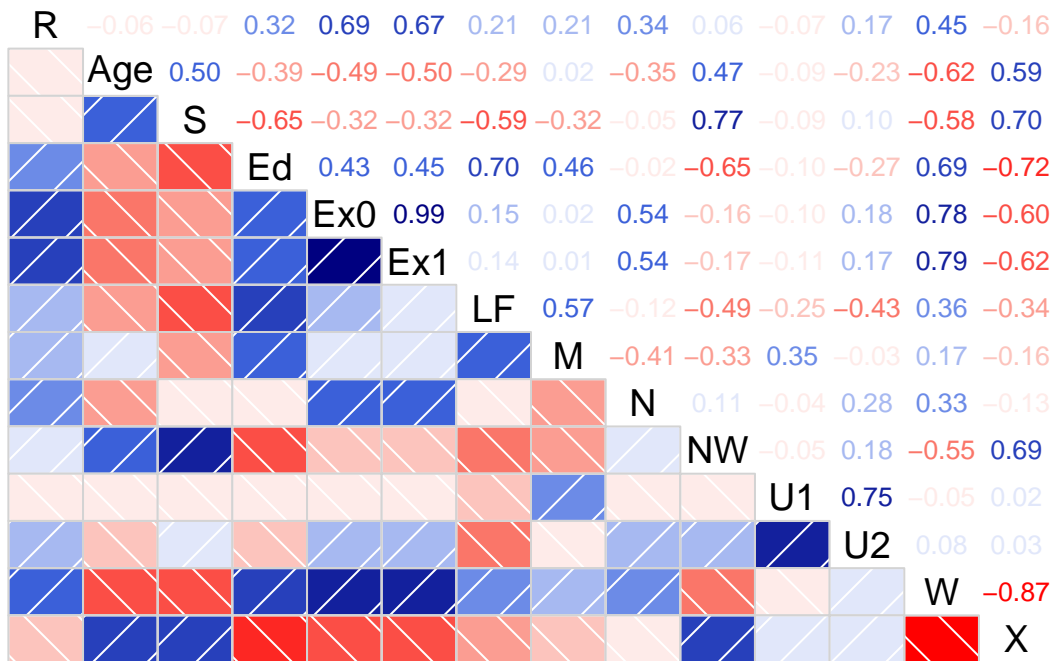
```
## Loading required package: corrgram
```

```
## Warning: package 'corrgram' was built under R version 4.5.1
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
```

[illegible]

[illegible]



a

##		R	Age	S	Ed	Ex0	Ex1
##	R	1.00000000	-0.06385252	-0.07019063	0.32329860	0.68928252	0.668464257
##	Age	-0.06385252	1.00000000	0.49563186	-0.39091483	-0.49264875	-0.502693296
##	S	-0.07019063	0.49563186	1.00000000	-0.64747633	-0.32045788	-0.323455258
##	Ed	0.32329860	-0.39091483	-0.64747633	1.00000000	0.43282086	0.448093527
##	Ex0	0.68928252	-0.49264875	-0.32045788	0.43282086	1.00000000	0.993734036
##	Ex1	0.66846426	-0.50269330	-0.32345526	0.44809353	0.99373404	1.000000000
##	LF	0.20705534	-0.29152317	-0.58723686	0.69798059	0.15264744	0.137772867
##	M	0.21121021	0.01526966	-0.31959407	0.46296602	0.02165309	0.009534882
##	N	0.34127174	-0.35436566	-0.05410083	-0.01910621	0.54442047	0.535792054
##	NW	0.05956120	0.47094928	0.76829134	-0.64603429	-0.16258262	-0.169233485
##	U1	-0.07289827	-0.08793529	-0.09174550	-0.09731568	-0.10466577	-0.114386133
##	U2	0.17420312	-0.23140870	0.10052251	-0.27269674	0.18260950	0.166347617
##	W	0.44673029	-0.61824271	-0.57843178	0.68823067	0.77793369	0.786065251
##	X	-0.16028052	0.59009977	0.69969529	-0.72343269	-0.60017237	-0.618469311
##		LF	M	N	NW	U1	U2
##	R	0.2070553	0.211210205	0.34127174	0.05956120	-0.07289827	0.17420312
##	Age	-0.2915232	0.015269658	-0.35436566	0.47094928	-0.08793529	-0.23140870
##	S	-0.5872369	-0.319594067	-0.05410083	0.76829134	-0.09174550	0.10052251
##	Ed	0.6979806	0.462966021	-0.01910621	-0.64603429	-0.09731568	-0.27269674
##	Ex0	0.1526474	0.021653093	0.54442047	-0.16258262	-0.10466577	0.18260950
##	Ex1	0.1377729	0.009534882	0.53579205	-0.16923348	-0.11438613	0.16634762
##	LF	1.0000000	0.565086214	-0.12012675	-0.48842912	-0.24579517	-0.43291231
##	M	0.5650862	1.000000000	-0.41148463	-0.33211203	0.34656091	-0.03443387

```
## N    -0.1201267 -0.411484630  1.00000000  0.10696571 -0.03757046  0.28474593
## NW   -0.4884291 -0.332112027  0.10696571  1.00000000 -0.05142211  0.17986948
## U1   -0.2457952  0.346560909 -0.03757046 -0.05142211  1.00000000  0.75410791
## U2   -0.4329123 -0.034433867  0.28474593  0.17986948  0.75410791  1.00000000
## W     0.3563029  0.171175479  0.33141816 -0.55176568 -0.04905061  0.07625084
## X    -0.3391087 -0.162806574 -0.13203942  0.68518986  0.02051815  0.02605951
##      W      X
## R     0.44673029 -0.16028052
## Age  -0.61824271  0.59009977
## S    -0.57843178  0.69969529
## Ed    0.68823067 -0.72343269
## Ex0   0.77793369 -0.60017237
## Ex1   0.78606525 -0.61846931
## LF    0.35630295 -0.33910873
## M     0.17117548 -0.16280657
## N     0.33141816 -0.13203942
## NW   -0.55176568  0.68518986
## U1   -0.04905061  0.02051815
## U2    0.07625084  0.02605951
## W     1.00000000 -0.86974138
## X    -0.86974138  1.00000000
```

```
#see Ex0 and Ex1 have strongest correlation with R and Ex0 and Ex1 have very strong correlation 0.99 wi
#correlations in order
cor = sort(a, decreasing = TRUE)
cor
```

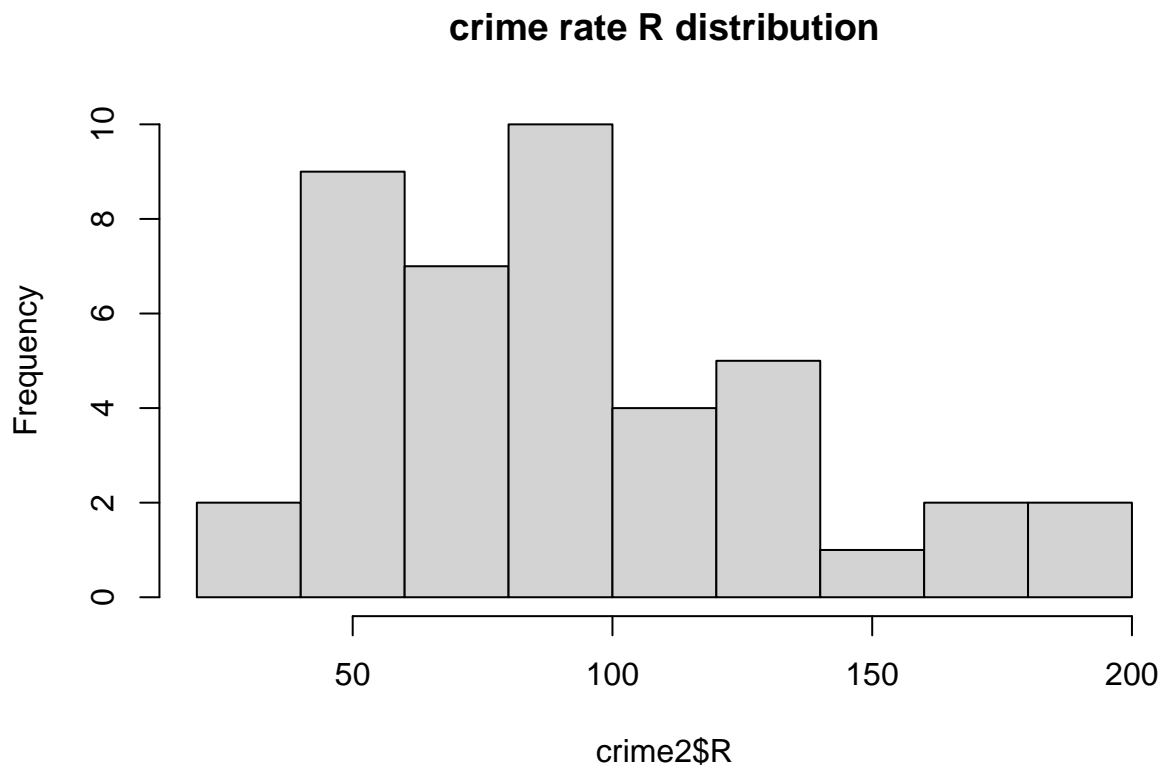
```
## [1] 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
## [6] 1.000000000 1.000000000 1.000000000 1.000000000 1.000000000
## [11] 1.000000000 1.000000000 1.000000000 1.000000000 0.993734036
## [16] 0.993734036 0.786065251 0.786065251 0.777933692 0.777933692
## [21] 0.768291340 0.768291340 0.754107907 0.754107907 0.699695291
## [26] 0.699695291 0.697980593 0.697980593 0.689282524 0.689282524
## [31] 0.688230666 0.688230666 0.685189855 0.685189855 0.668464257
## [36] 0.668464257 0.590099765 0.590099765 0.565086214 0.565086214
## [41] 0.544420469 0.544420469 0.535792054 0.535792054 0.495631860
## [46] 0.495631860 0.470949277 0.470949277 0.462966021 0.462966021
## [51] 0.448093527 0.448093527 0.446730295 0.446730295 0.432820860
## [56] 0.432820860 0.356302949 0.356302949 0.346560909 0.346560909
## [61] 0.341271739 0.341271739 0.331418161 0.331418161 0.323298601
## [66] 0.323298601 0.284745928 0.284745928 0.211210205 0.211210205
## [71] 0.207055338 0.207055338 0.182609503 0.182609503 0.179869475
## [76] 0.179869475 0.174203116 0.174203116 0.171175479 0.171175479
## [81] 0.166347617 0.166347617 0.152647438 0.152647438 0.137772867
## [86] 0.137772867 0.106965706 0.106965706 0.100522513 0.100522513
## [91] 0.076250839 0.076250839 0.059561200 0.059561200 0.026059506
## [96] 0.026059506 0.021653093 0.021653093 0.020518146 0.020518146
## [101] 0.015269658 0.015269658 0.009534882 0.009534882 -0.019106210
## [106] -0.019106210 -0.034433867 -0.034433867 -0.037570455 -0.037570455
## [111] -0.049050611 -0.049050611 -0.051422110 -0.051422110 -0.054100834
## [116] -0.054100834 -0.063852524 -0.063852524 -0.070190626 -0.070190626
## [121] -0.072898268 -0.072898268 -0.087935294 -0.087935294 -0.091745504
## [126] -0.091745504 -0.097315678 -0.097315678 -0.104665773 -0.104665773
## [131] -0.114386133 -0.114386133 -0.120126747 -0.120126747 -0.132039418
```

```
## [136] -0.132039418 -0.160280516 -0.160280516 -0.162582617 -0.162582617
## [141] -0.162806574 -0.162806574 -0.169233485 -0.169233485 -0.231408699
## [146] -0.231408699 -0.245795167 -0.245795167 -0.272696736 -0.272696736
## [151] -0.291523171 -0.291523171 -0.319594067 -0.319594067 -0.320457881
## [156] -0.320457881 -0.323455258 -0.323455258 -0.332112027 -0.332112027
## [161] -0.339108734 -0.339108734 -0.354365664 -0.354365664 -0.390914833
## [166] -0.390914833 -0.411484630 -0.411484630 -0.432912310 -0.432912310
## [171] -0.488429120 -0.488429120 -0.492648749 -0.492648749 -0.502693296
## [176] -0.502693296 -0.551765675 -0.551765675 -0.578431778 -0.578431778
## [181] -0.587236864 -0.587236864 -0.600172370 -0.600172370 -0.618242710
## [186] -0.618242710 -0.618469311 -0.618469311 -0.646034293 -0.646034293
## [191] -0.647476333 -0.647476333 -0.723432689 -0.723432689 -0.869741376
## [196] -0.869741376
```

```
upper_cor = cor[-(1:14)]
upper_cor = upper_cor[seq(1, length(crime$R), 2)]
upper_cor
```

```
## [1] 0.9937340 0.7860653 0.7779337 0.7682913 0.7541079 0.6996953 0.6979806
## [8] 0.6892825 0.6882307 0.6851899 0.6684643 0.5900998 0.5650862 0.5444205
## [15] 0.5357921 0.4956319 0.4709493 0.4629660 0.4480935 0.4467303 0.4328209
## [22] 0.3563029 0.3465609 0.3412717
```

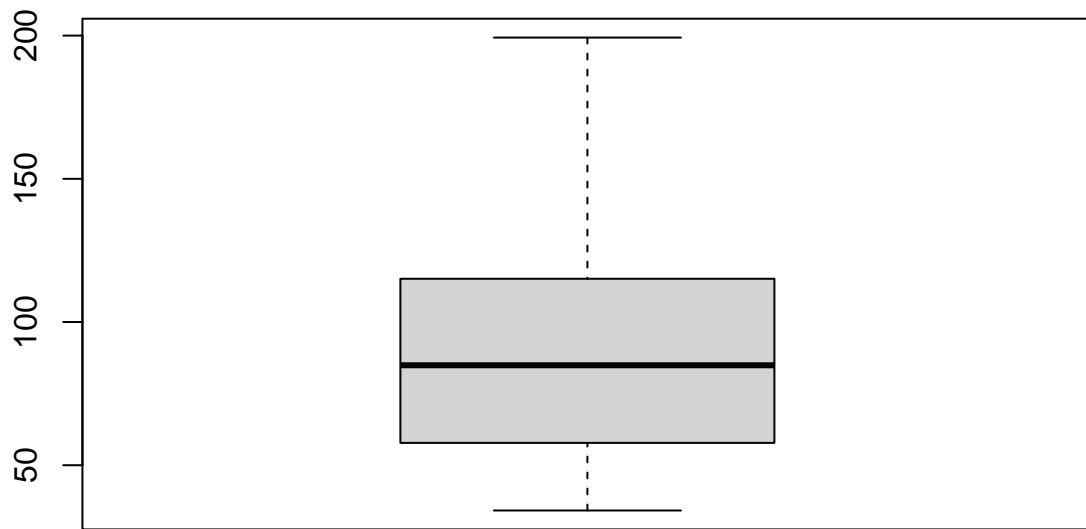
```
#distrubution of crime rate
hist(crime2$R, main = "crime rate R distribution")
```



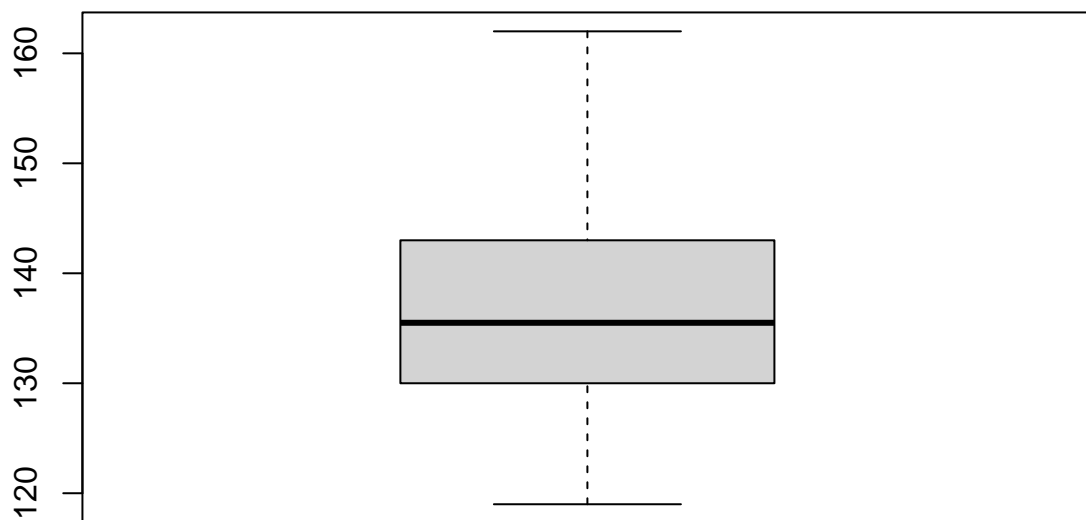

```

for (var in names(crime2)){
  Outlier = boxplot(crime2[[var]])$out
  if (length(Outlier) == 0){
    cat("No potential outliers in", var, "\n")
  } else {
    cat("potential outliers in", var, crime2[[var]][which(crime2[[var]] %in% Outlier)], "\n")
  }
}

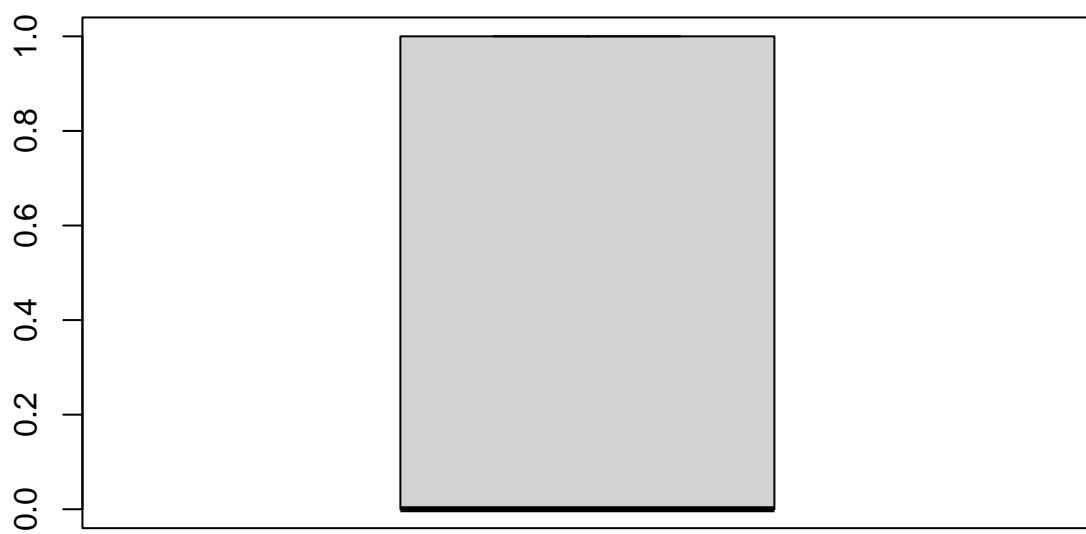
```



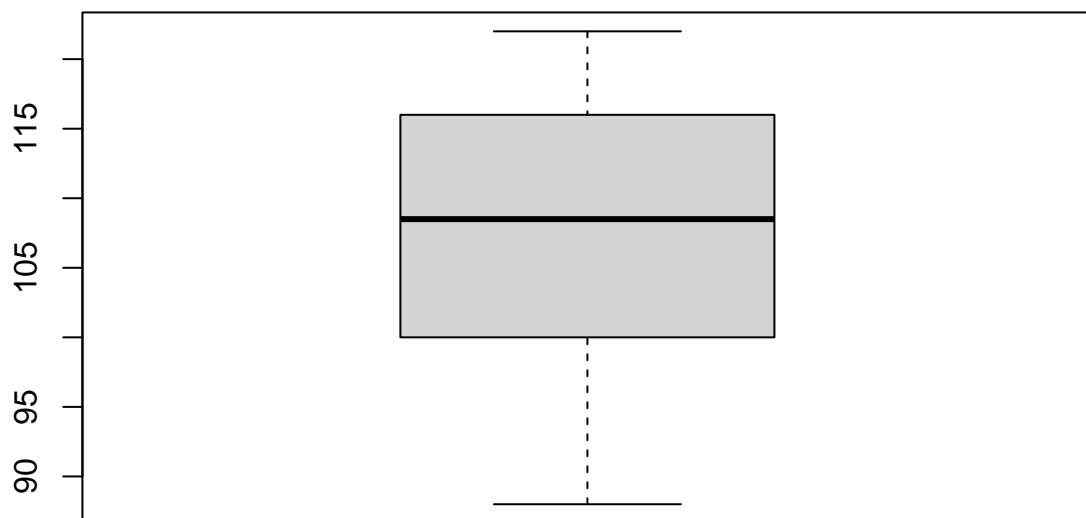
No potential outliers in R



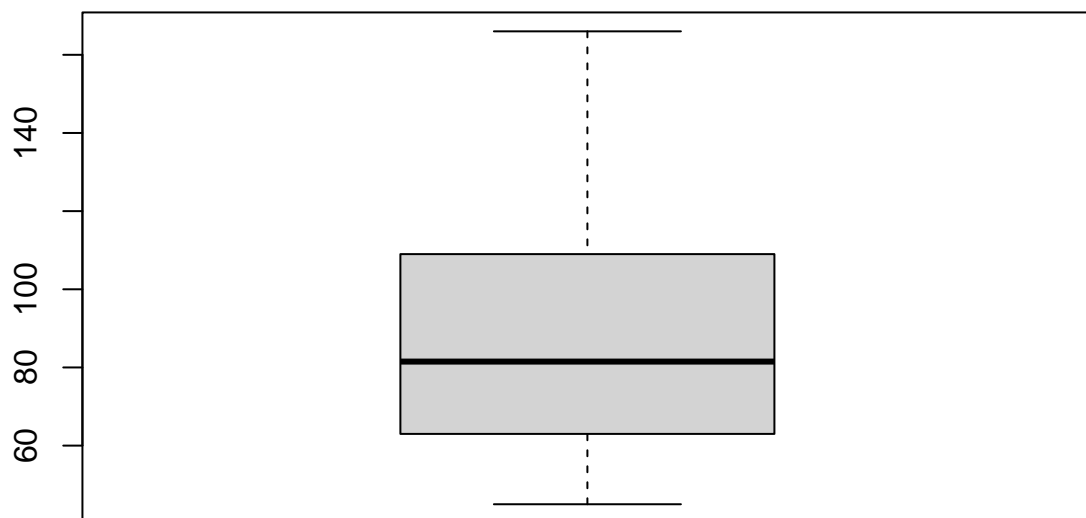
No potential outliers in Age



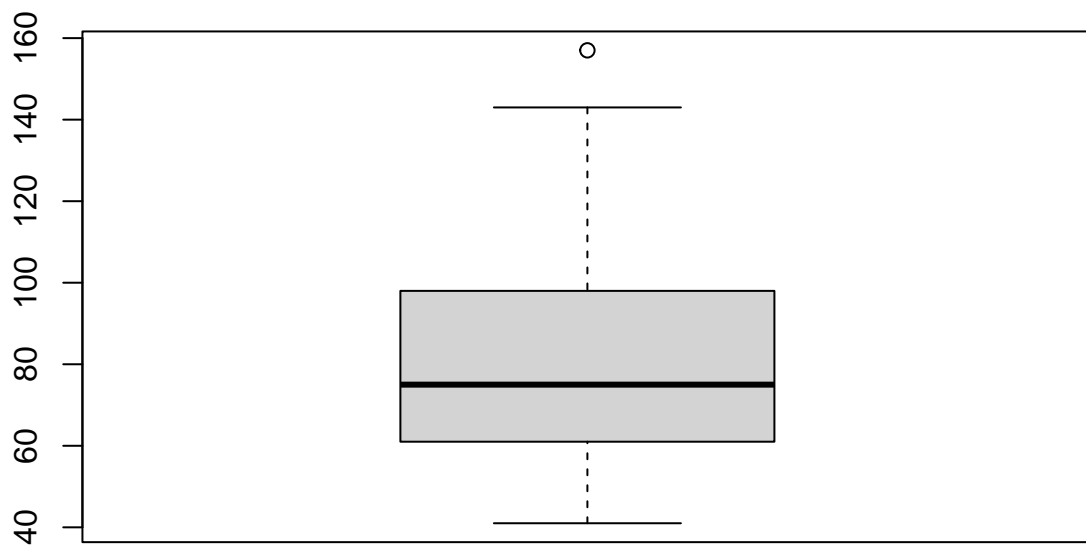
No potentail outliers in S



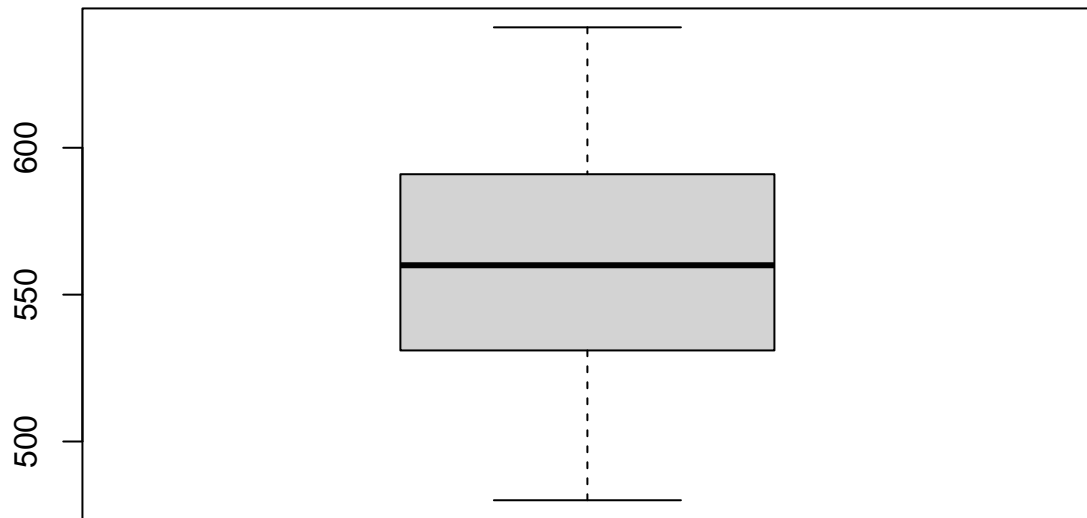
No potential outliers in Ed



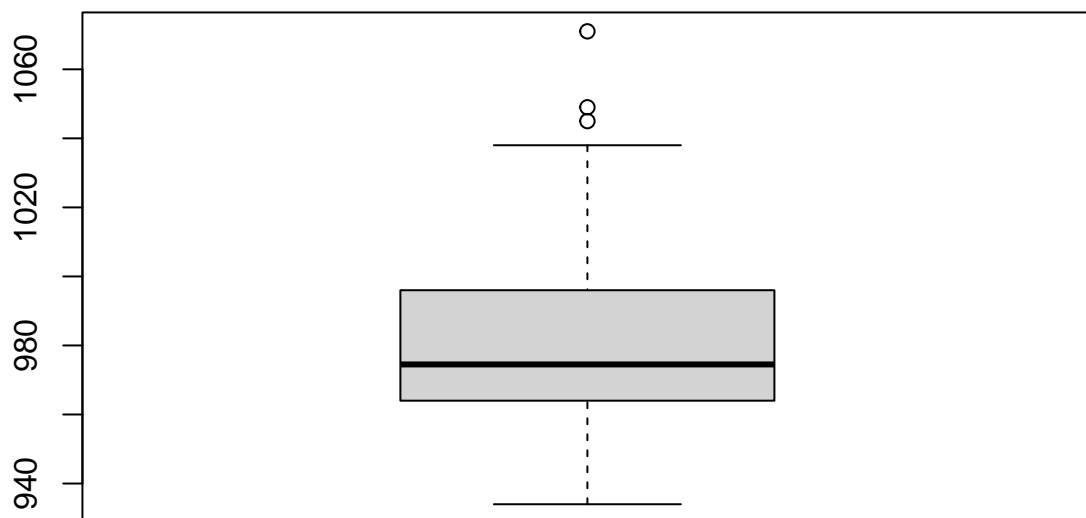
No potentail outliers in Ex0



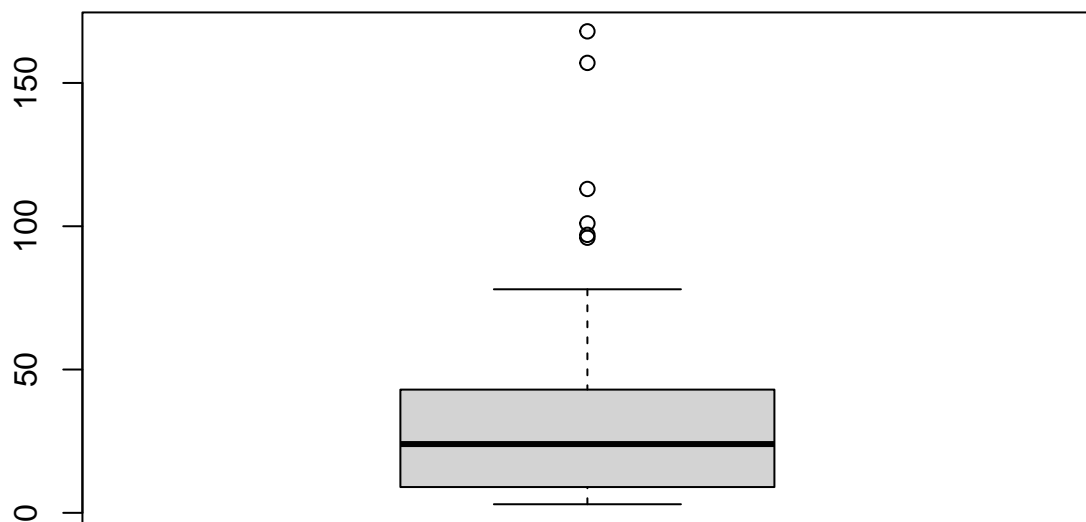
potential outliers in Ex1 157



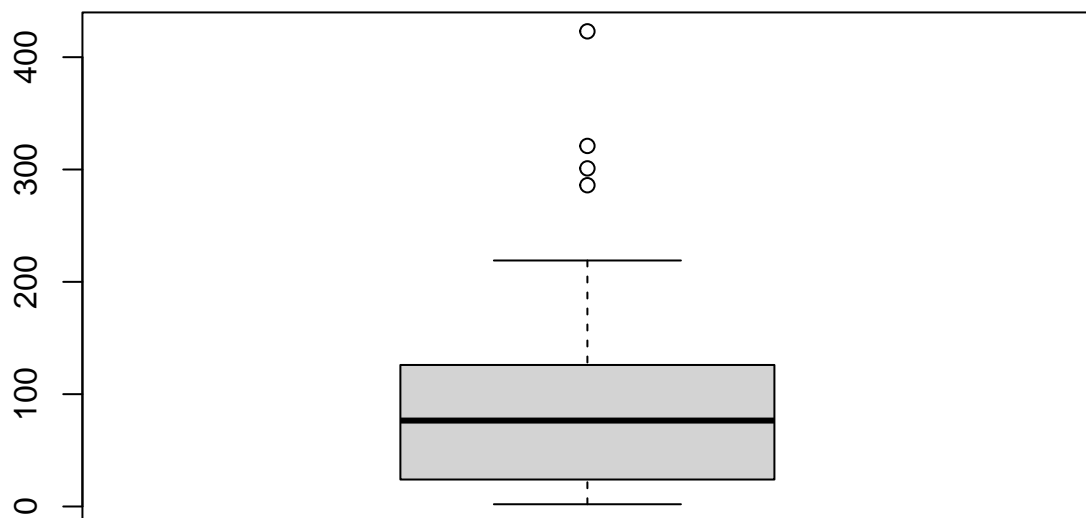
No potential outliers in LF



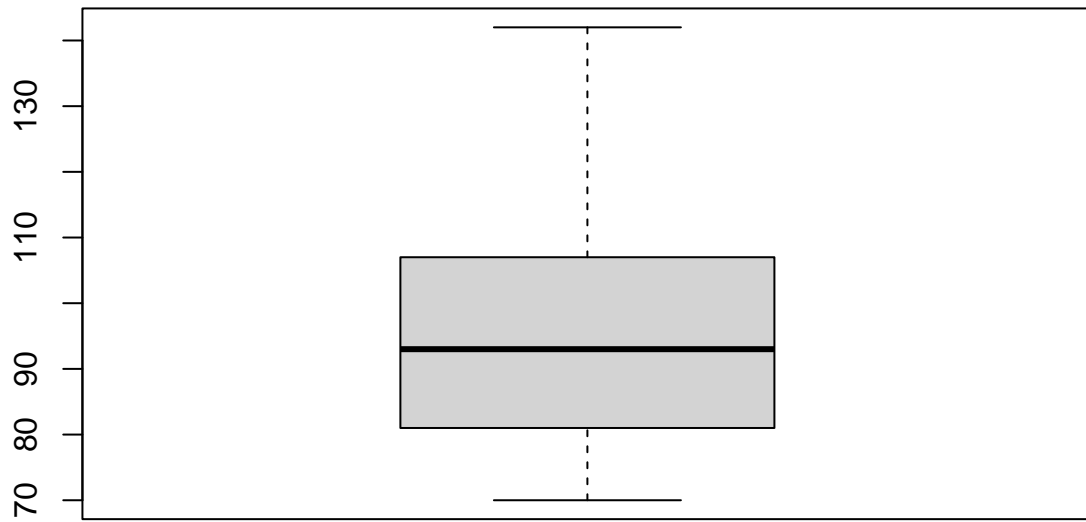
potential outliers in M 1071 1045 1049



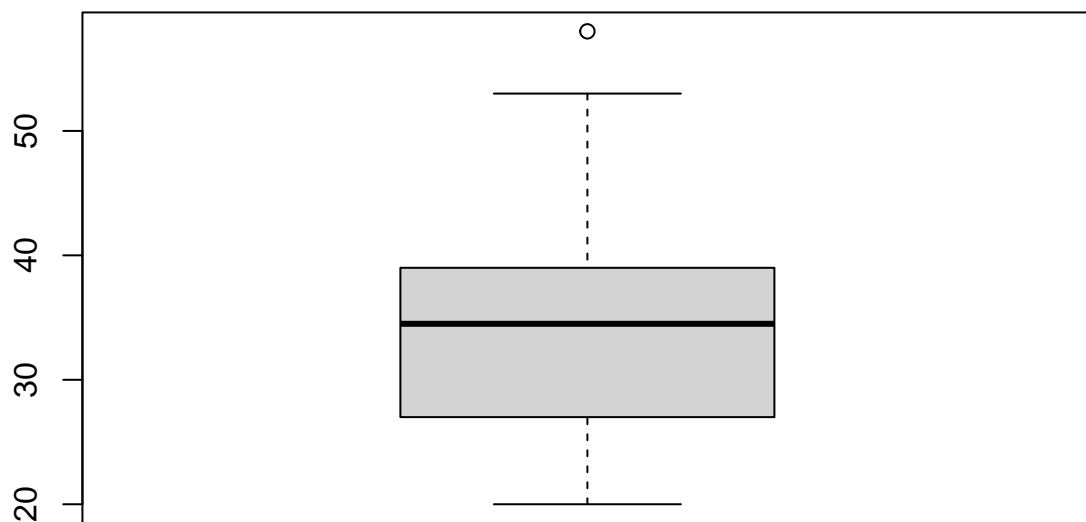
potential outliers in N 157 101 168 97 113 96



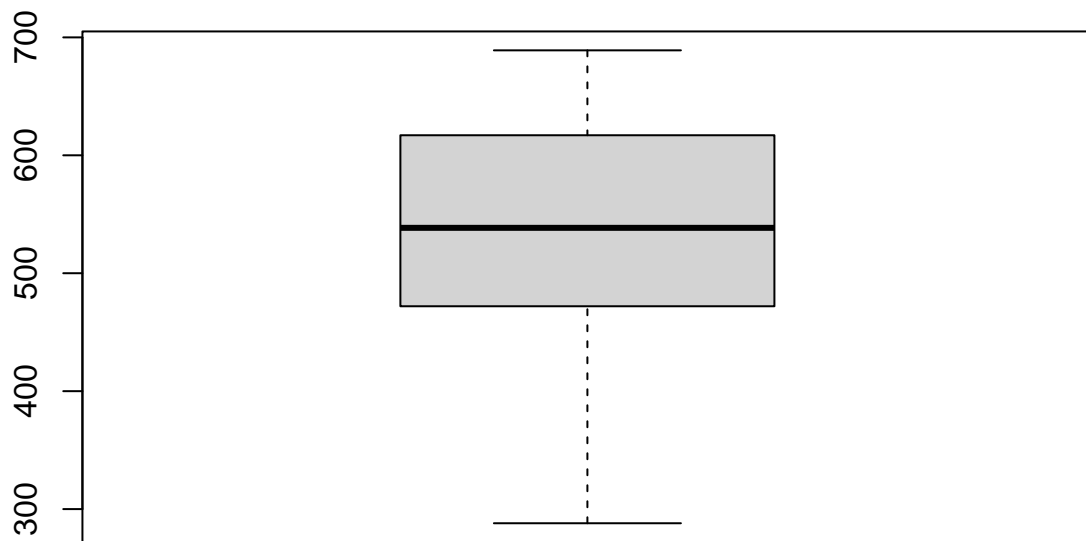
potential outliers in NW 301 286 321 423



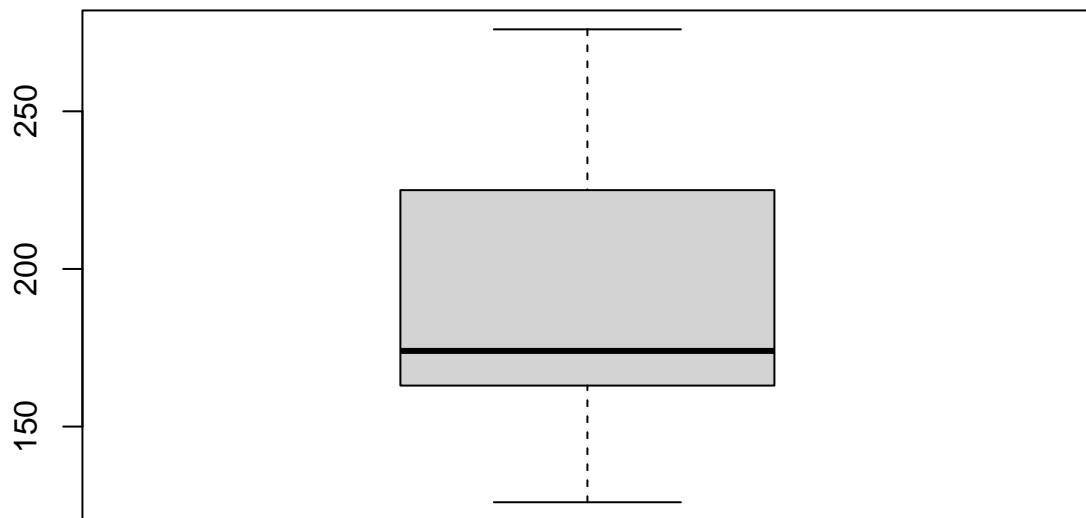
No potentail outliers in U1



potential outliers in U2 58

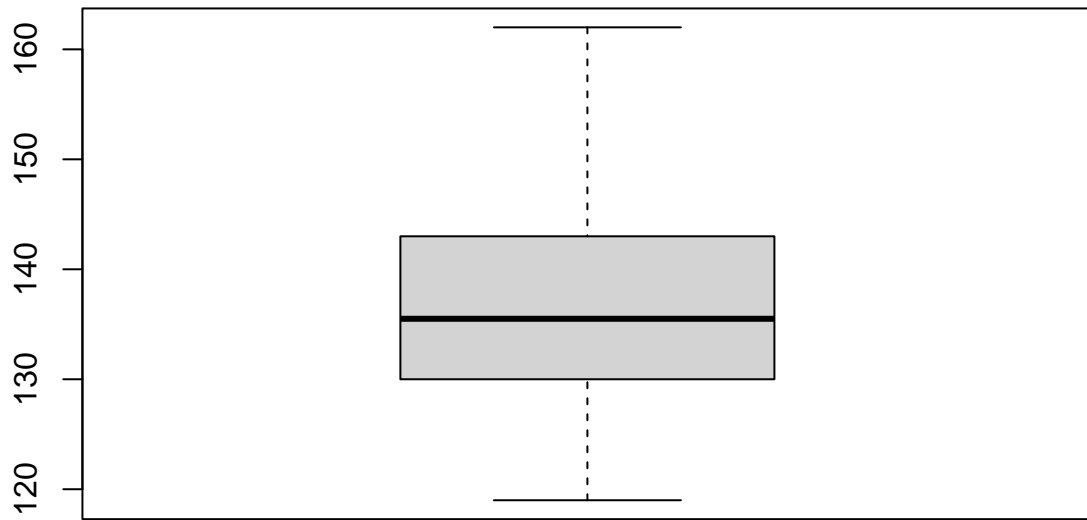


No potential outliers in W

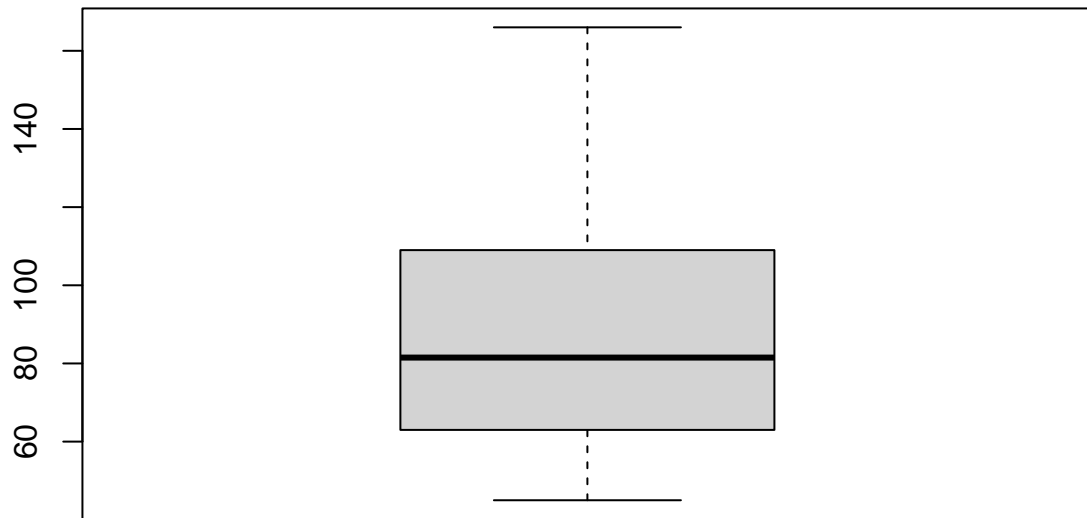


```
## No potential outliers in X
```

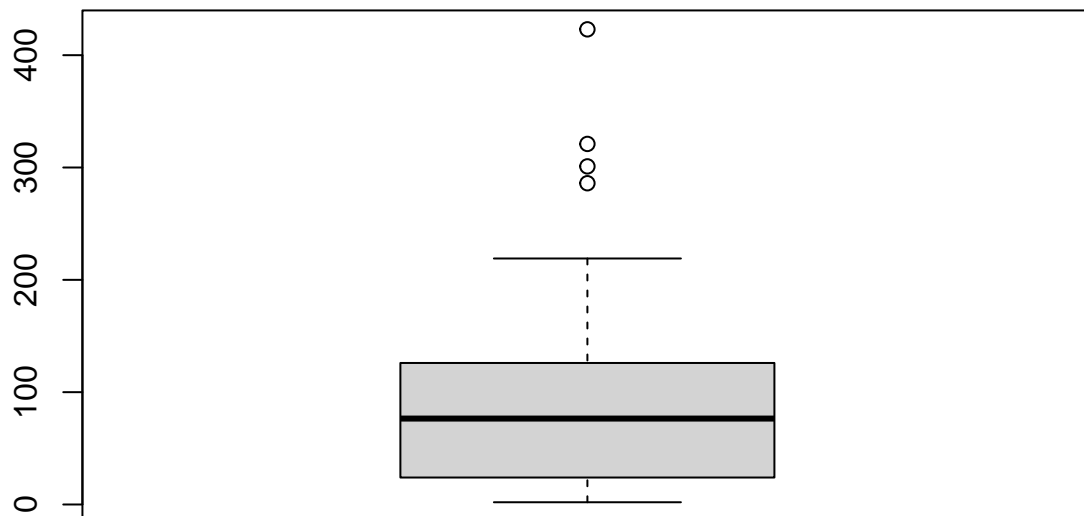
```
boxplot(crime2$Age)
```



```
boxplot(crime2$Ex0)
```



```
boxplot(crime2$NW)
```

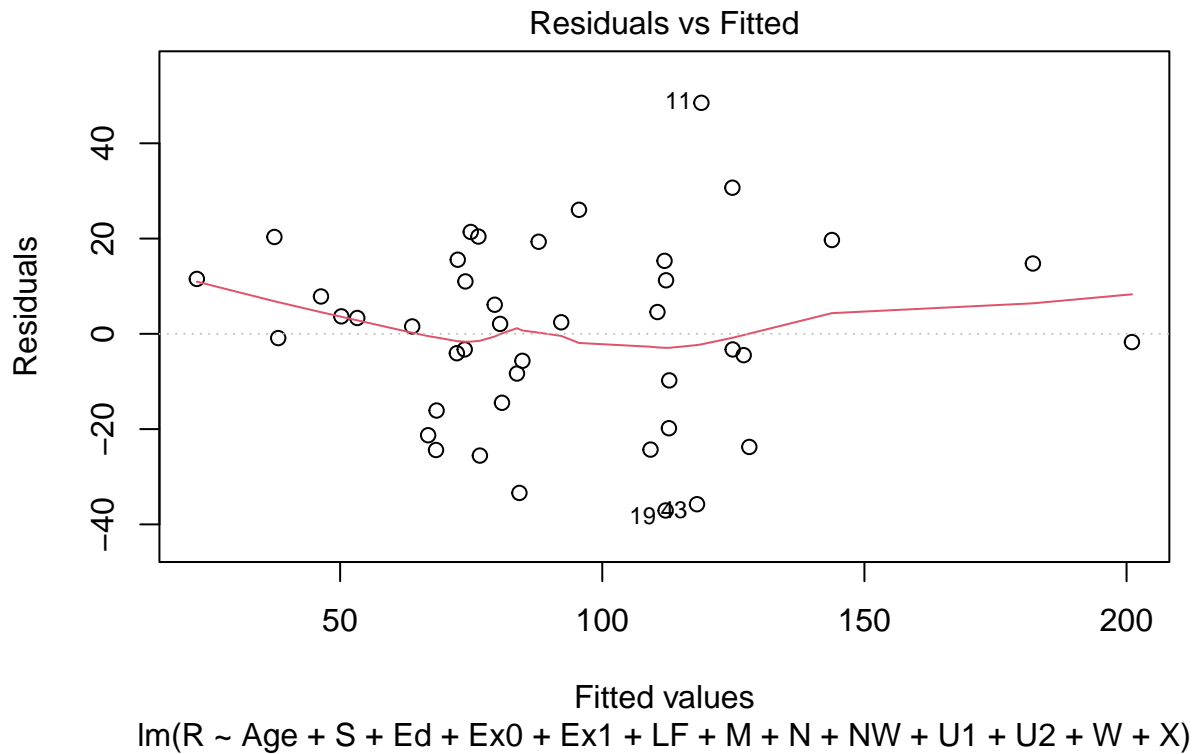
b.) Confirmatory data analysis (e.g., model selection, model diagnostics, justifications of models, and interpretations of results).

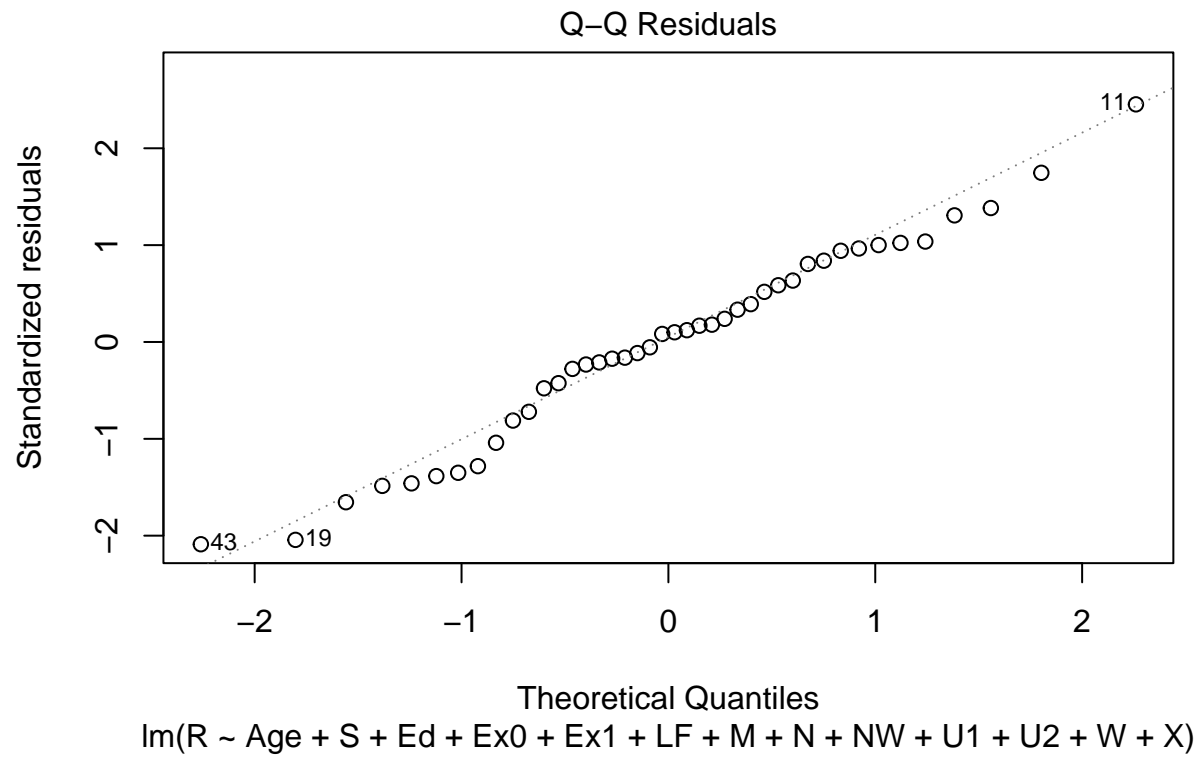
```
#general linear model
model = lm(R~Age+S+Ed+Ex0+Ex1+LF+M+N+NW+U1+U2+W+X, crime)
summary(model)
```

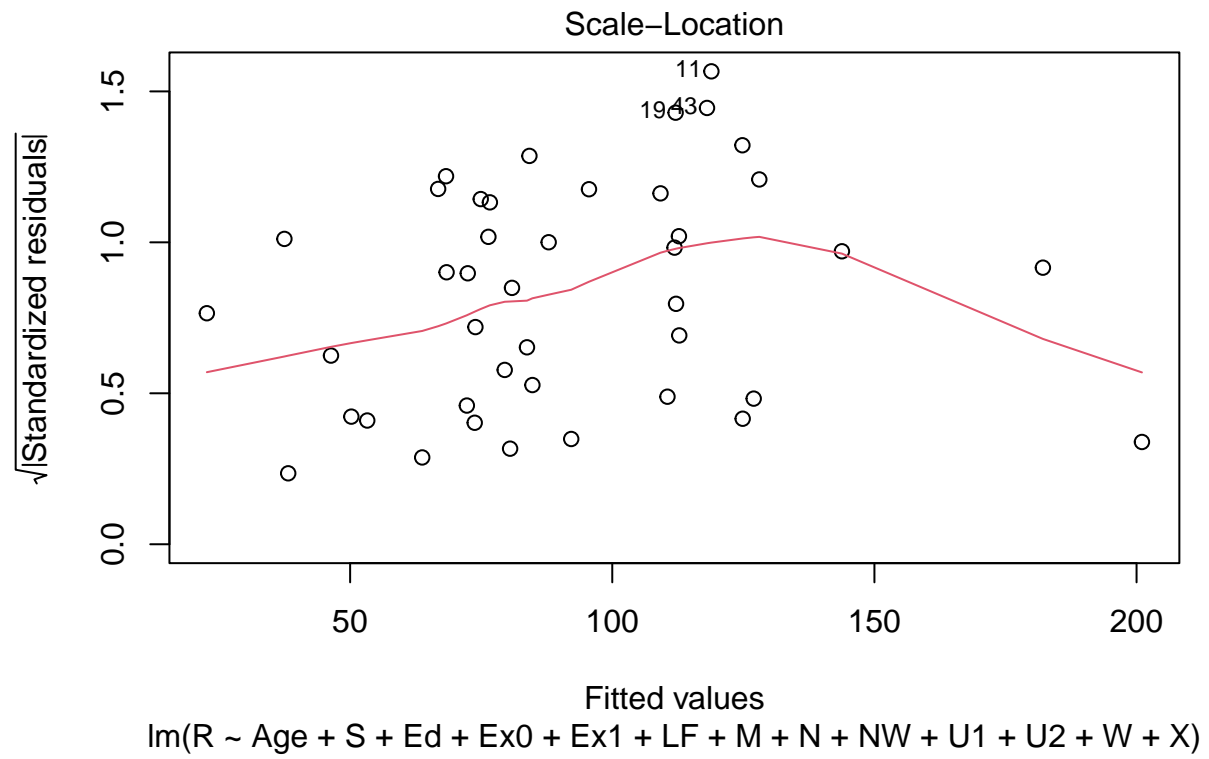
```
##
## Call:
## lm(formula = R ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW +
##      U1 + U2 + W + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.084 -13.299   1.818  13.953  48.498
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -753.94415   173.40191  -4.348  0.000164 ***
## Age           1.38771     0.58150   2.386  0.024011 *
## S            -3.77882    16.78753  -0.225  0.823539
## Ed             1.46481     0.80926   1.810  0.081030 .
## Ex0            1.13689     1.22329   0.929  0.360640
## Ex1           -0.12237     1.35167  -0.091  0.928510
## LF              0.06729     0.25164   0.267  0.791102
## M              0.15977     0.25792   0.619  0.540609
```

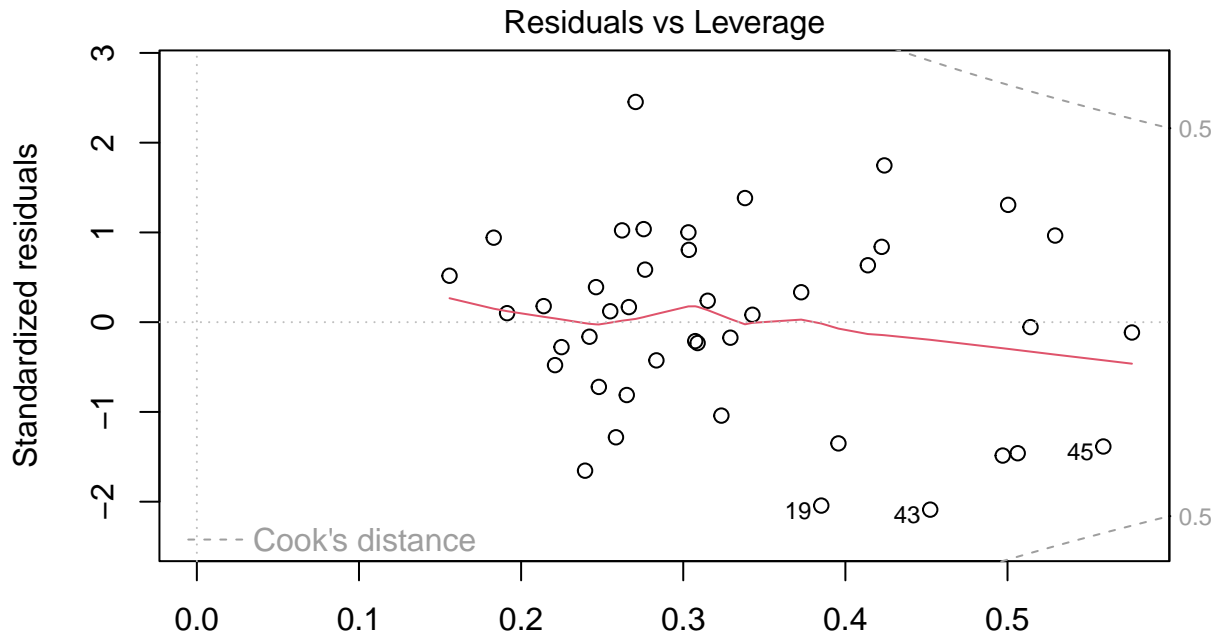
```
## N          -0.02862    0.14343   -0.200  0.843264
## NW         -0.02844    0.07612   -0.374  0.711450
## U1         -0.70749    0.52643   -1.344  0.189758
## U2          2.23189    0.98459    2.267  0.031319 *
## W           0.12587    0.11359    1.108  0.277217
## X           0.76347    0.27209    2.806  0.009025 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.14 on 28 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.779, Adjusted R-squared:  0.6764
## F-statistic: 7.592 on 13 and 28 DF, p-value: 4.026e-06
```

```
#x=>Age=>U2 most significant, Ed sig at 0.10 level
#R^2 of 77.9% with adjusted R^2 of 67.64%
plot(model)
```









Leverage
lm(R ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW + U1 + U2 + W + X)

*#does not seem to be outliers based on cooks distance
#residuals vs fitted seems to make double bow pattern indicating non linear pattern
#qqplot values follow line in middle but fall off in the tails
#scale location seems to make nonlinear pattern in the line of fit*

```
#model with significant factors based on pvalues
model2 <- lm(R ~ Age + U2 + X + Ed, data = crime)
summary(model2)
```

```
##
## Call:
## lm(formula = R ~ Age + U2 + X + Ed, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -47.073 -27.646  -4.101  21.694  90.767
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -347.8475   159.7200  -2.178  0.03537 *
## Age             0.6883     0.6529   1.054  0.29807
## U2              1.7471     0.7224   2.419  0.02022 *
## X               0.1769     0.2356   0.751  0.45717
## Ed              2.3611     0.8172   2.889  0.00621 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.3 on 40 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.2333, Adjusted R-squared:  0.1567
## F-statistic: 3.043 on 4 and 40 DF,  p-value: 0.0279
```

```
#worse r^2 and adjr^2
```

```
# Stepwise regression (both directions)
```

```
#backward
```

```
full_model <- lm(R ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW + U1 + U2 + W + X, data = crime)
step_model <- step(full_model, direction = "both")
```

```
## Start: AIC=274.88
```

```
## R ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW + U1 + U2 + W +
## X
```

```
##
##      Df Sum of Sq  RSS   AIC
## - Ex1  1      4.4 15003 272.89
## - N    1     21.3 15020 272.94
## - S    1     27.1 15026 272.95
## - LF   1     38.3 15037 272.99
## - NW   1     74.8 15074 273.09
## - M    1    205.6 15204 273.45
## - Ex0  1    462.7 15462 274.15
## - W    1    657.8 15657 274.68
## <none>          14999 274.88
## - U1    1    967.5 15966 275.50
## - Ed    1   1755.1 16754 277.53
## - U2    1   2752.5 17752 279.95
## - Age   1   3050.7 18050 280.65
## - X     1   4217.5 19216 283.29
##
```

```
## Step: AIC=272.89
```

```
## R ~ Age + S + Ed + Ex0 + LF + M + N + NW + U1 + U2 + W + X
##
```

```
##      Df Sum of Sq  RSS   AIC
## - N    1     21.6 15025 270.95
## - S    1     24.5 15028 270.96
## - LF   1     59.3 15063 271.06
## - NW   1     88.5 15092 271.14
## - M    1    201.2 15204 271.45
## - W    1    653.8 15657 272.68
## <none>          15003 272.89
## - U1    1    964.6 15968 273.51
## + Ex1   1      4.4 14999 274.88
## - Ed    1   1892.2 16896 275.88
## - U2    1   2827.2 17831 278.14
## - Age   1   3419.3 18423 279.51
## - X     1   4216.2 19220 281.29
## - Ex0   1   8044.7 23048 288.92
##
```

```

## Step: AIC=270.95
## R ~ Age + S + Ed + Ex0 + LF + M + NW + U1 + U2 + W + X
##
##      Df Sum of Sq  RSS    AIC
## - S      1      22.3 15047 269.01
## - LF      1      49.0 15074 269.09
## - NW      1      89.1 15114 269.20
## - M      1     359.2 15384 269.94
## - W      1     632.2 15657 270.68
## <none>                15025 270.95
## - U1      1    1080.6 16106 271.87
## + N      1       21.6 15003 272.89
## + Ex1     1        4.6 15020 272.94
## - Ed      1    1901.4 16926 273.96
## - U2      1    2851.5 17876 276.25
## - Age     1    3449.0 18474 277.63
## - X       1    4430.4 19455 279.81
## - Ex0     1    9427.3 24452 289.41
##
## Step: AIC=269.01
## R ~ Age + Ed + Ex0 + LF + M + NW + U1 + U2 + W + X
##
##      Df Sum of Sq  RSS    AIC
## - LF      1     136.8 15184 267.39
## - NW      1     170.8 15218 267.49
## - M      1     341.1 15388 267.95
## - W      1     610.6 15658 268.68
## <none>                15047 269.01
## - U1      1    1136.0 16183 270.07
## + S      1       22.3 15025 270.95
## + N      1       19.3 15028 270.96
## + Ex1     1        2.0 15045 271.01
## - Ed      1    1898.8 16946 272.00
## - U2      1    2839.1 17886 274.27
## - Age     1    3596.6 18644 276.01
## - X       1    4912.5 19960 278.88
## - Ex0     1   10406.1 25453 289.09
##
## Step: AIC=267.39
## R ~ Age + Ed + Ex0 + M + NW + U1 + U2 + W + X
##
##      Df Sum of Sq  RSS    AIC
## - NW      1     199.8 15384 265.94
## - W      1     622.3 15806 267.08
## <none>                15184 267.39
## - M      1     942.7 16127 267.92
## + LF      1     136.8 15047 269.01
## + S      1     110.1 15074 269.09
## + Ex1     1      35.3 15149 269.30
## + N      1        2.1 15182 269.39
## - U1      1    1604.4 16788 269.61
## - U2      1    2803.5 17988 272.51
## - Ed      1    3076.7 18261 273.14
## - Age     1    3730.2 18914 274.62

```

```
## - X      1      5696.8 20881 278.77
## - Ex0    1     10271.2 25455 287.09
##
## Step: AIC=265.94
## R ~ Age + Ed + Ex0 + M + U1 + U2 + W + X
##
##      Df Sum of Sq  RSS    AIC
## - W      1      749.8 16134 265.94
## <none>                15384 265.94
## - M      1     1050.2 16434 266.72
## + S      1      242.2 15142 267.28
## + NW     1      199.8 15184 267.39
## + LF     1      165.8 15218 267.49
## + Ex1    1       76.1 15308 267.73
## - U1     1     1464.7 16849 267.76
## + N      1        0.8 15383 267.94
## - U2     1     2610.2 17994 270.52
## - Ed     1     3284.9 18669 272.07
## - Age    1     3554.4 18938 272.67
## - X      1     5623.6 21007 277.03
## - Ex0    1     10726.5 26110 286.16
##
## Step: AIC=265.94
## R ~ Age + Ed + Ex0 + M + U1 + U2 + X
##
##      Df Sum of Sq  RSS    AIC
## <none>                16134 265.94
## + W      1      749.8 15384 265.94
## + NW     1      327.3 15806 267.08
## - M      1     1345.0 17479 267.31
## + S      1      206.4 15927 267.40
## + LF     1      189.8 15944 267.44
## + Ex1    1       78.0 16056 267.74
## + N      1       16.6 16117 267.90
## - U1     1     1972.7 18106 268.79
## - Age    1     3146.2 19280 271.42
## - U2     1     3400.3 19534 271.97
## - Ed     1     4161.6 20295 273.58
## - X      1     5559.9 21693 276.38
## - Ex0    1     18654.3 34788 296.21
```

```
summary(step_model)
```

```
##
## Call:
## lm(formula = R ~ Age + Ed + Ex0 + M + U1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.252 -10.029   0.014  15.006  54.445
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -713.4473   139.5666  -5.112 1.23e-05 ***
```

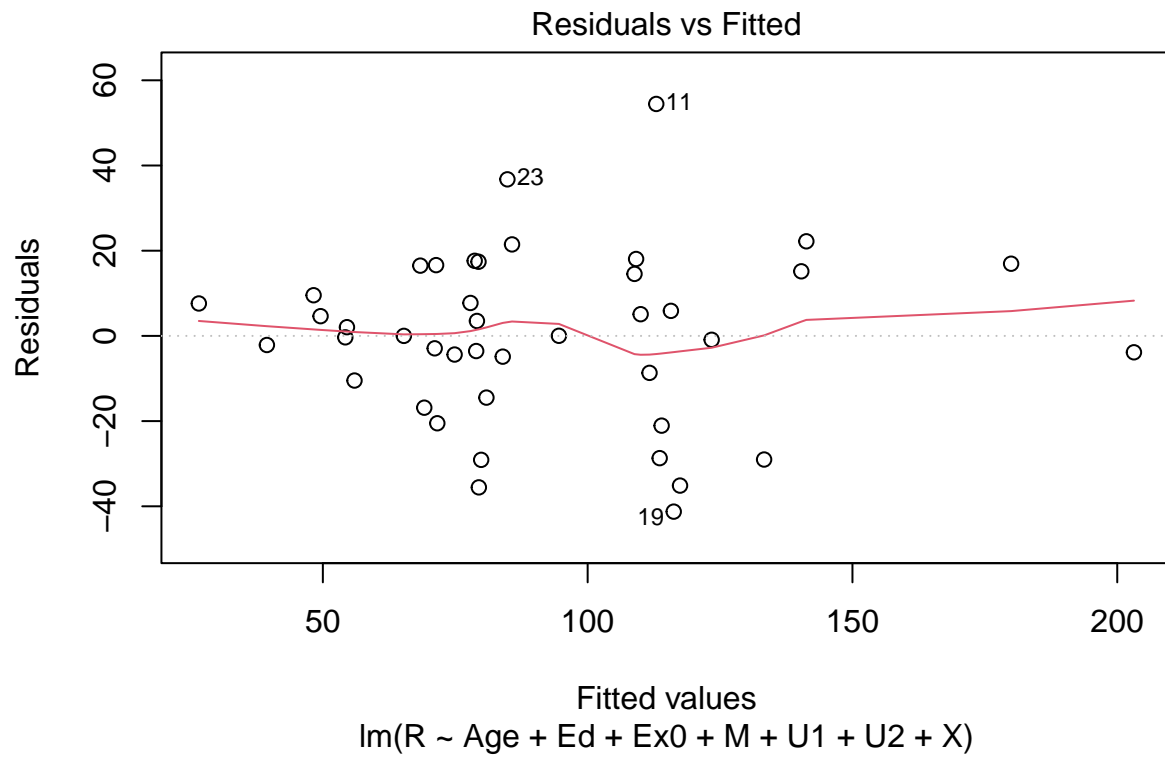


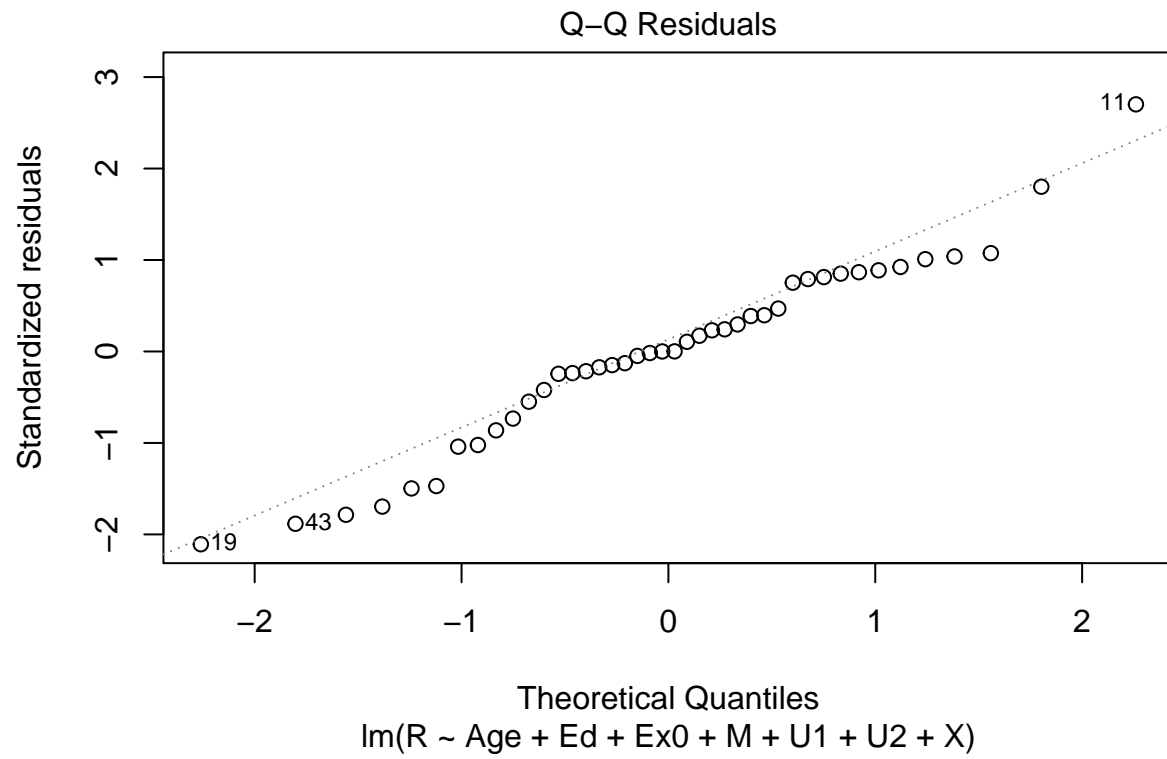
```
## Age          1.1238      0.4364    2.575  0.01455 *
## Ed           1.7967      0.6067    2.961  0.00555 **
## Ex0          1.0798      0.1722    6.270 3.85e-07 ***
## M            0.2662      0.1581    1.684  0.10142
## U1          -0.8218      0.4030   -2.039  0.04929 *
## U2           2.3159      0.8651    2.677  0.01135 *
## X            0.5461      0.1595    3.423  0.00163 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.78 on 34 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7133
## F-statistic: 15.57 on 7 and 34 DF,  p-value: 5.996e-09
```

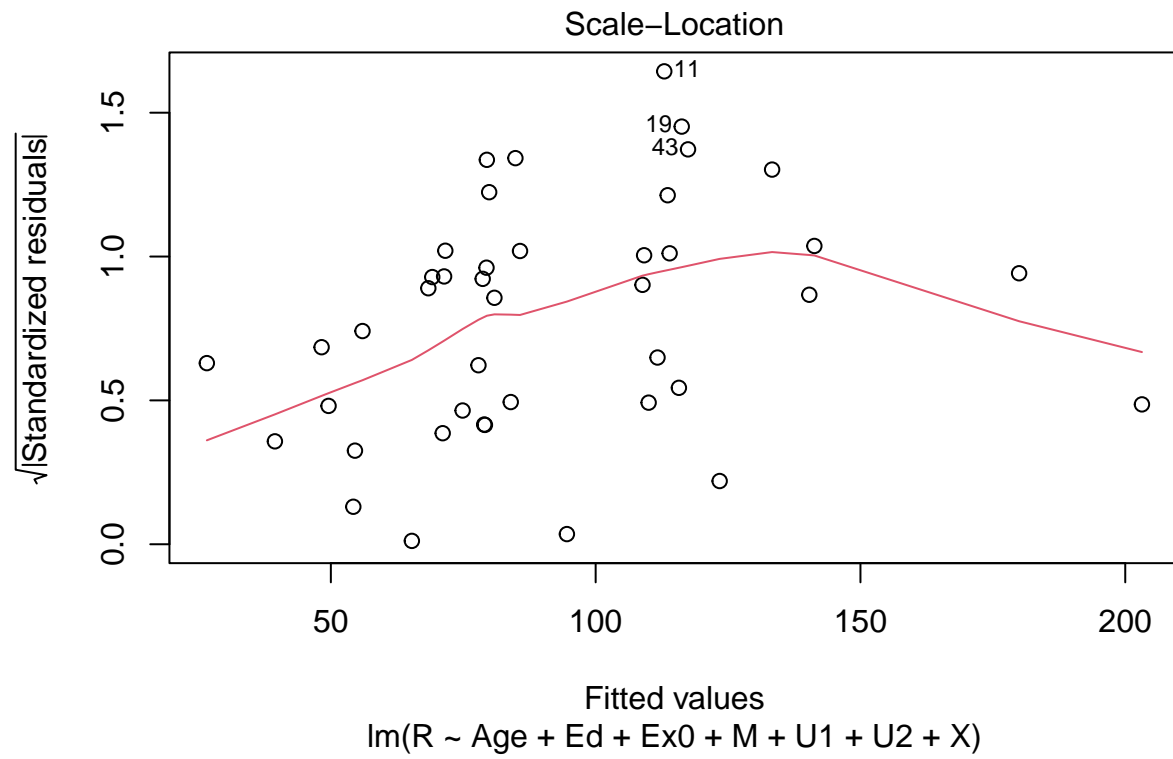
```
#from backward elimination best model uses age,ed,exo,m,u1,u2,x with r^2adj 0.7133 which is better than
model3 = lm(R~Age + Ed + Ex0 + M + U1 + U2 + X, data = crime)
summary(model3)
```

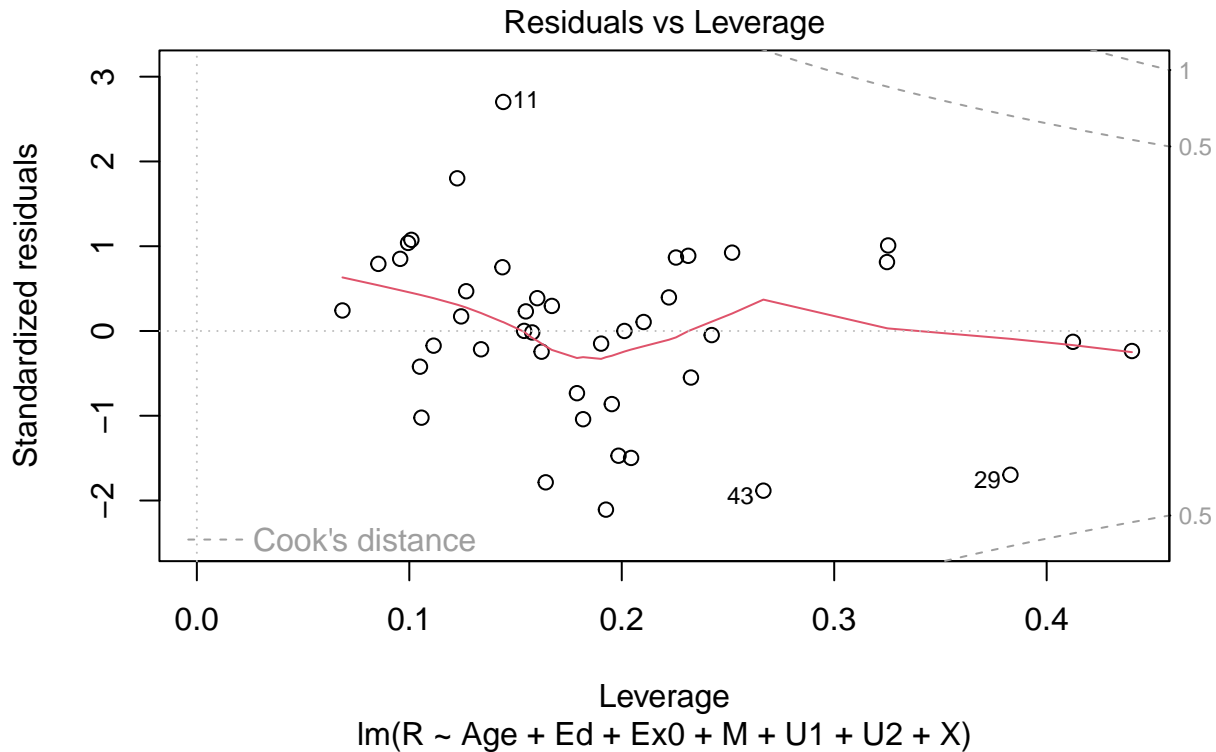
```
##
## Call:
## lm(formula = R ~ Age + Ed + Ex0 + M + U1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.252 -10.029   0.014  15.006  54.445
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -713.4473    139.5666  -5.112 1.23e-05 ***
## Age          1.1238      0.4364    2.575  0.01455 *
## Ed           1.7967      0.6067    2.961  0.00555 **
## Ex0          1.0798      0.1722    6.270 3.85e-07 ***
## M            0.2662      0.1581    1.684  0.10142
## U1          -0.8218      0.4030   -2.039  0.04929 *
## U2           2.3159      0.8651    2.677  0.01135 *
## X            0.5461      0.1595    3.423  0.00163 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.78 on 34 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7133
## F-statistic: 15.57 on 7 and 34 DF,  p-value: 5.996e-09
```

```
plot(model3)
```









```
#residuals vs fitted still looks double bow
#tail residuals still dont fit line
#non linear scale location pattern
#no cooks distance outliers
```

```
#forward
empty_model = lm(R~1, crime2)
step_for = step(empty_model, scope = formula(full_model), direction = "forward")
```

```
## Start: AIC=312.28
## R ~ 1
##
##      Df Sum of Sq  RSS   AIC
## + Ex0  1    32243 35622 287.21
## + Ex1  1    30325 37540 289.41
## + W    1    13544 54321 304.93
## + N    1     7904 59961 309.08
## + Ed   1     7093 60772 309.64
## <none>          67865 312.28
## + M    1     3027 64838 312.36
## + LF   1     2910 64956 312.44
## + U2   1     2059 65806 312.99
## + X    1     1743 66122 313.19
## + U1   1        361 67504 314.06
## + S    1        334 67531 314.07
## + Age  1         277 67588 314.11
```

```

## + NW      1      241 67624 314.13
##
## Step: AIC=287.21
## R ~ Ex0
##
##      Df Sum of Sq  RSS    AIC
## + Age   1    6812.7 28809 280.29
## + X      1    6811.6 28810 280.29
## + M      1    2615.9 33006 286.00
## + NW     1    2053.3 33568 286.71
## + S      1    1717.5 33904 287.13
## <none>                35622 287.21
## + Ex1    1    1478.8 34143 287.43
## + W      1    1376.4 34245 287.55
## + LF     1     720.6 34901 288.35
## + U2     1     164.0 35458 289.01
## + N      1     111.4 35510 289.08
## + Ed     1      52.0 35570 289.15
## + U1     1       0.0 35622 289.21
##
## Step: AIC=280.29
## R ~ Ex0 + Age
##
##      Df Sum of Sq  RSS    AIC
## + X      1    2761.73 26047 278.06
## + LF     1    2419.33 26390 278.61
## + M      1    2372.38 26437 278.68
## <none>                28809 280.29
## + Ex1    1     759.81 28049 281.17
## + U2     1     719.37 28090 281.23
## + Ed     1     707.56 28101 281.25
## + U1     1     176.37 28633 282.04
## + NW     1      75.60 28733 282.18
## + S      1      69.85 28739 282.19
## + W      1       3.25 28806 282.29
## + N      1       0.67 28808 282.29
##
## Step: AIC=278.06
## R ~ Ex0 + Age + X
##
##      Df Sum of Sq  RSS    AIC
## + Ed     1    5690.1 20357 269.71
## + LF     1    4014.1 22033 273.03
## + M      1    3823.4 22224 273.39
## + W      1    3388.3 22659 274.21
## + NW     1    1397.0 24650 277.75
## <none>                26047 278.06
## + S      1     841.5 25206 278.68
## + N      1     499.1 25548 279.25
## + Ex1    1     283.8 25763 279.60
## + U2     1     170.4 25877 279.78
## + U1     1     153.8 25894 279.81
##
## Step: AIC=269.71

```

```

## R ~ Ex0 + Age + X + Ed
##
##      Df Sum of Sq  RSS    AIC
## + U2   1   1997.42 18360 267.37
## + W    1   1681.17 18676 268.09
## <none>                20357 269.71
## + M    1    753.77 19603 270.12
## + U1   1    440.49 19917 270.79
## + Ex1  1    362.88 19994 270.95
## + LF   1    179.60 20178 271.34
## + NW   1    128.42 20229 271.44
## + N    1     60.65 20297 271.58
## + S    1     35.96 20321 271.63
##
## Step:  AIC=267.37
## R ~ Ex0 + Age + X + Ed + U2
##
##      Df Sum of Sq  RSS    AIC
## + W    1   1308.88 17051 266.26
## + LF   1   1043.54 17316 266.91
## + U1   1    881.10 17479 267.31
## <none>                18360 267.37
## + M    1    253.39 18106 268.79
## + NW   1    232.24 18128 268.84
## + Ex1  1    201.94 18158 268.91
## + N    1    102.05 18258 269.14
## + S    1     46.83 18313 269.26
##
## Step:  AIC=266.26
## R ~ Ex0 + Age + X + Ed + U2 + W
##
##      Df Sum of Sq  RSS    AIC
## + LF   1    817.54 16233 266.20
## <none>                17051 266.26
## + U1   1    616.90 16434 266.72
## + M    1    202.39 16849 267.76
## + N    1    196.19 16855 267.78
## + Ex1  1    178.12 16873 267.82
## + NW   1    114.73 16936 267.98
## + S    1     91.54 16959 268.04
##
## Step:  AIC=266.2
## R ~ Ex0 + Age + X + Ed + U2 + W + LF
##
##      Df Sum of Sq  RSS    AIC
## <none>                16233 266.20
## + U1   1    655.19 15578 266.47
## + N    1    115.30 16118 267.90
## + NW   1     47.98 16185 268.08
## + S    1     18.05 16215 268.15
## + Ex1  1      1.31 16232 268.20
## + M    1      0.04 16233 268.20

```

```
summary(step_for)
```

```
##
## Call:
## lm(formula = R ~ Ex0 + Age + X + Ed + U2 + W + LF, data = crime2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.326 -11.886  -2.043  11.935  51.397
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -693.9931   124.3150  -5.583 3.00e-06 ***
## Ex0           1.0808     0.1888   5.723 1.97e-06 ***
## Age           1.4520     0.4634   3.133 0.00355 **
## X              0.7334     0.2101   3.490 0.00136 **
## Ed            1.2114     0.6771   1.789 0.08251 .
## U2            1.0891     0.4941   2.204 0.03438 *
## W              0.1526     0.1013   1.506 0.14130
## LF            0.1878     0.1435   1.309 0.19946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.85 on 34 degrees of freedom
## Multiple R-squared:  0.7608, Adjusted R-squared:  0.7116
## F-statistic: 15.45 on 7 and 34 DF,  p-value: 6.628e-09
```

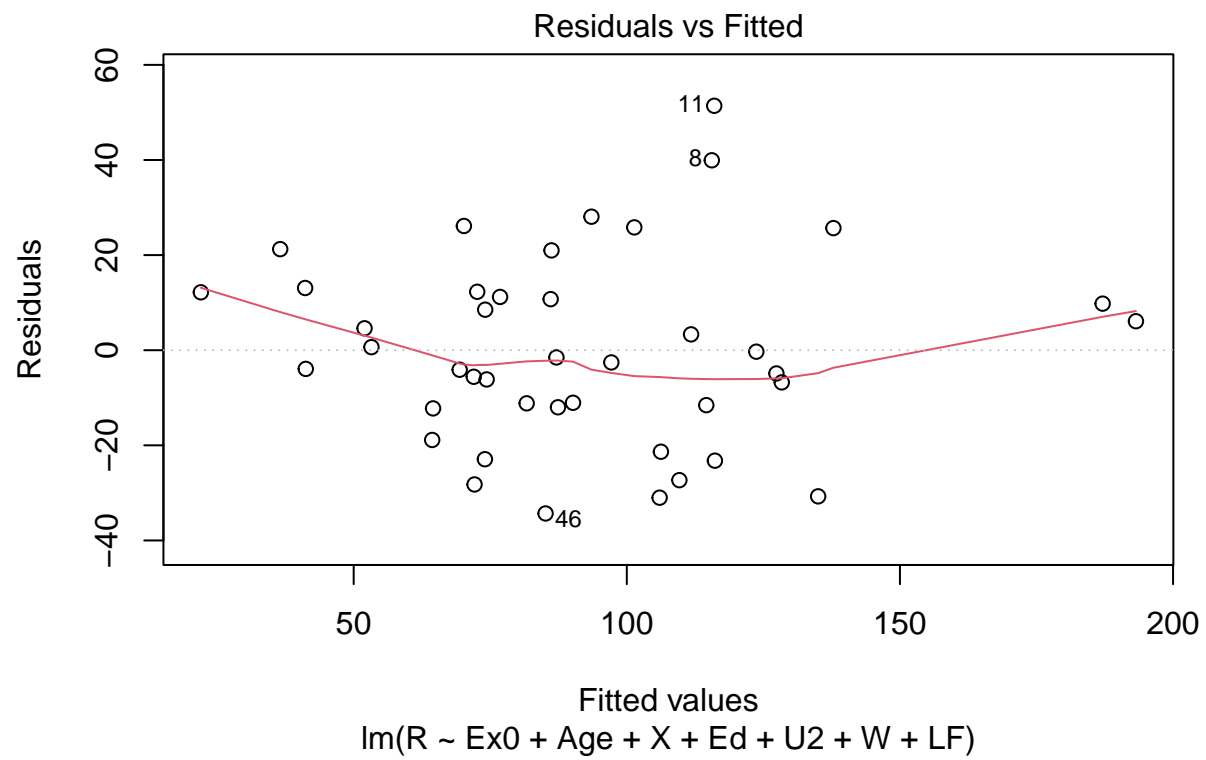
```
#from forward elimination best model uses Ex0, Age, X, Ed, U2, W, LF
model11 = lm(formula = R ~ Ex0 + Age + X + Ed + U2 + W + LF, data = crime2)
summary(model11)
```

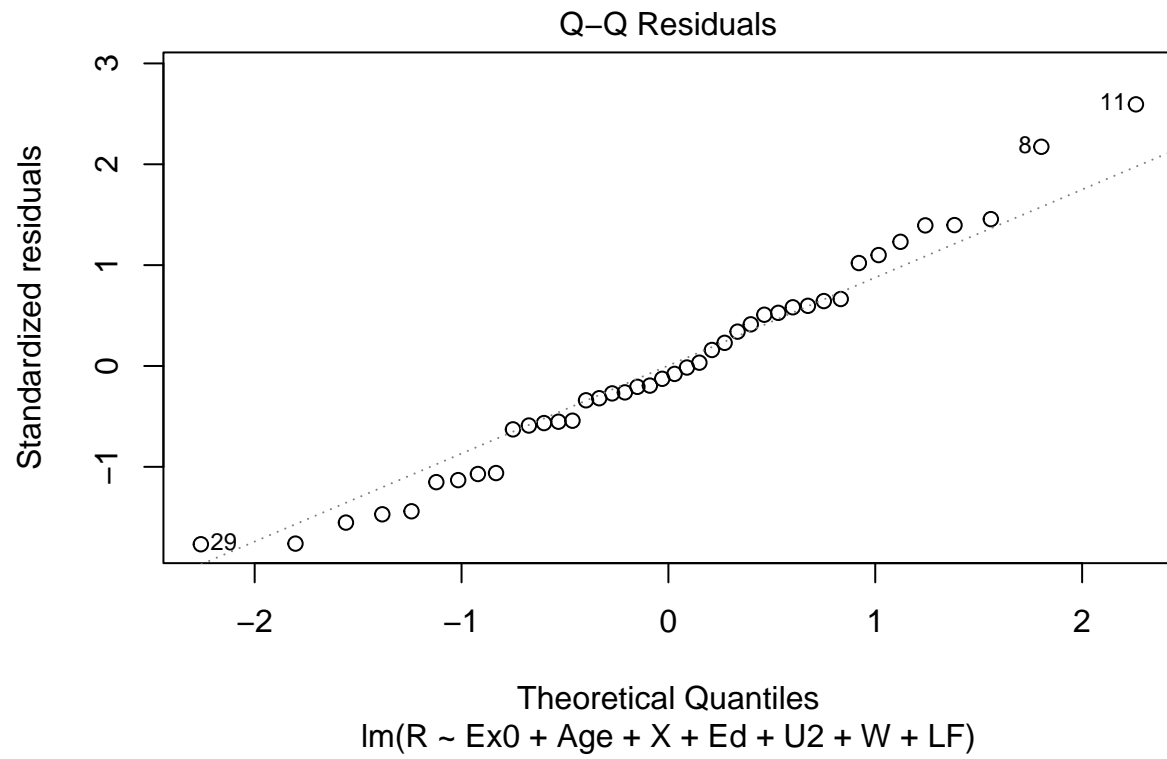
```
##
## Call:
## lm(formula = R ~ Ex0 + Age + X + Ed + U2 + W + LF, data = crime2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.326 -11.886  -2.043  11.935  51.397
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -693.9931   124.3150  -5.583 3.00e-06 ***
## Ex0           1.0808     0.1888   5.723 1.97e-06 ***
## Age           1.4520     0.4634   3.133 0.00355 **
## X              0.7334     0.2101   3.490 0.00136 **
## Ed            1.2114     0.6771   1.789 0.08251 .
## U2            1.0891     0.4941   2.204 0.03438 *
## W              0.1526     0.1013   1.506 0.14130
## LF            0.1878     0.1435   1.309 0.19946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

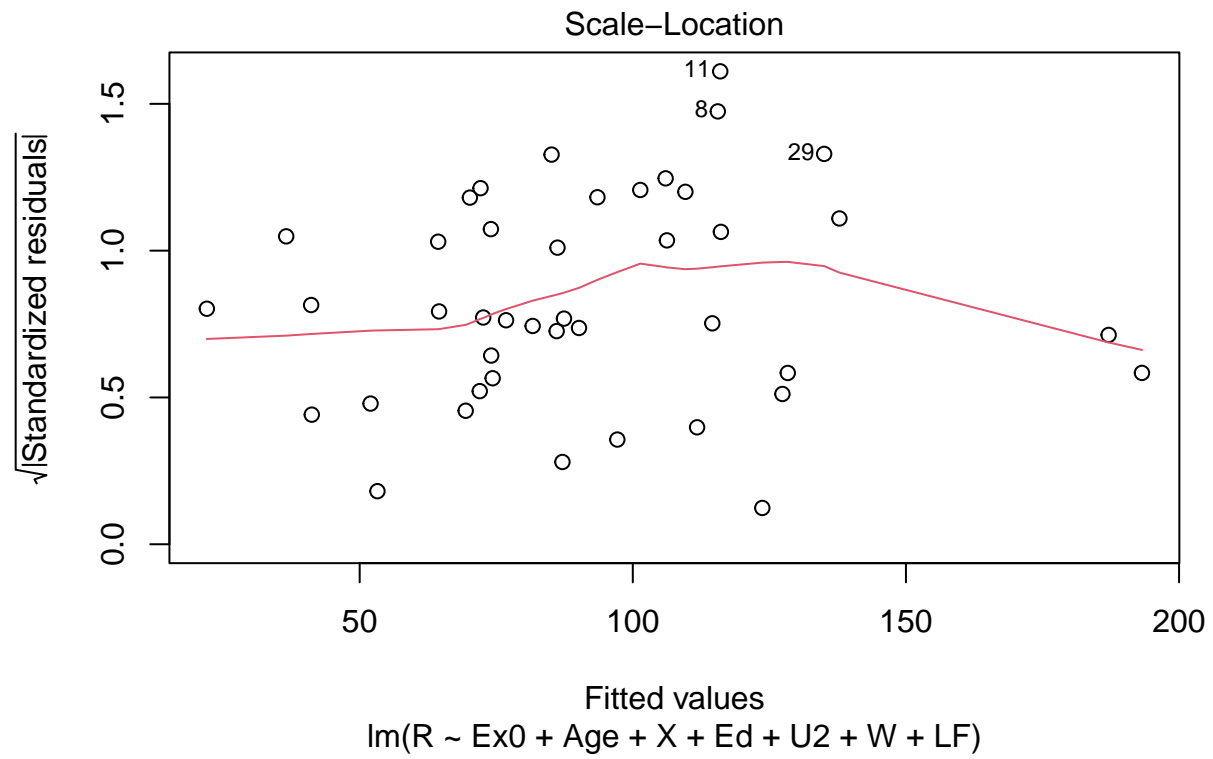


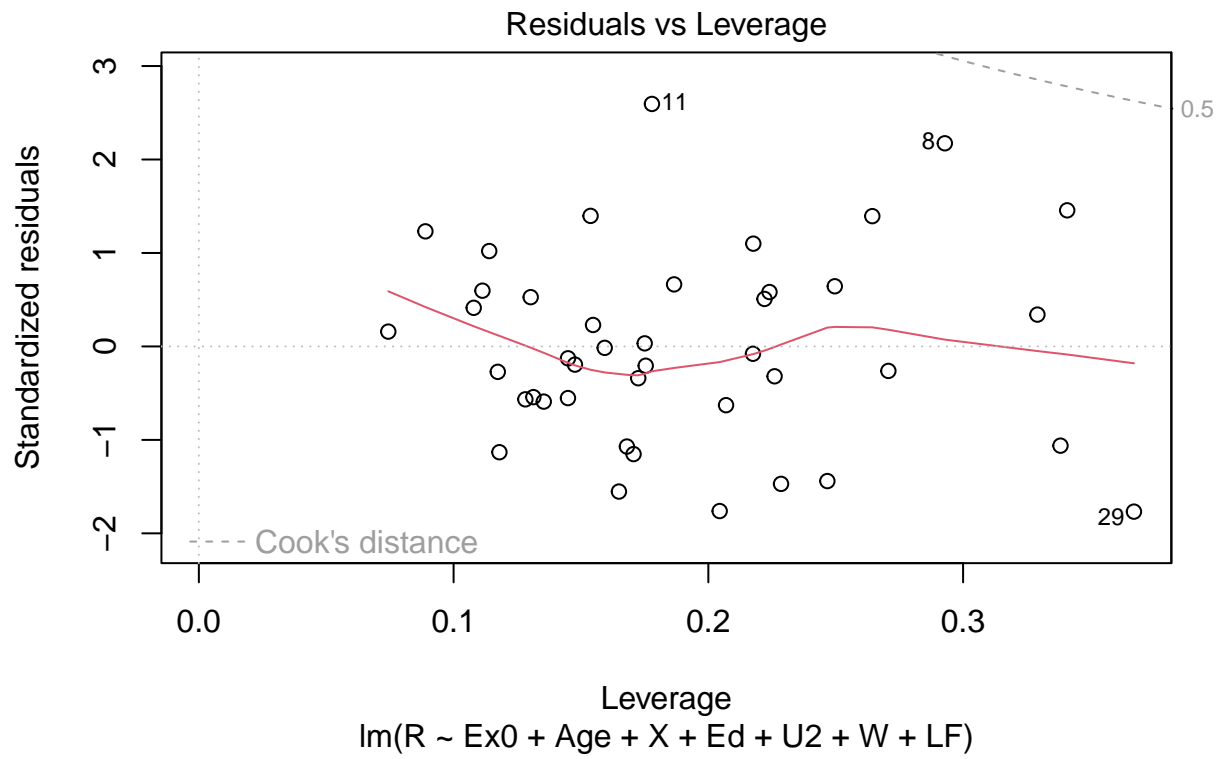
```
## Residual standard error: 21.85 on 34 degrees of freedom
## Multiple R-squared:  0.7608, Adjusted R-squared:  0.7116
## F-statistic: 15.45 on 7 and 34 DF,  p-value: 6.628e-09
```

```
plot(model11)
```

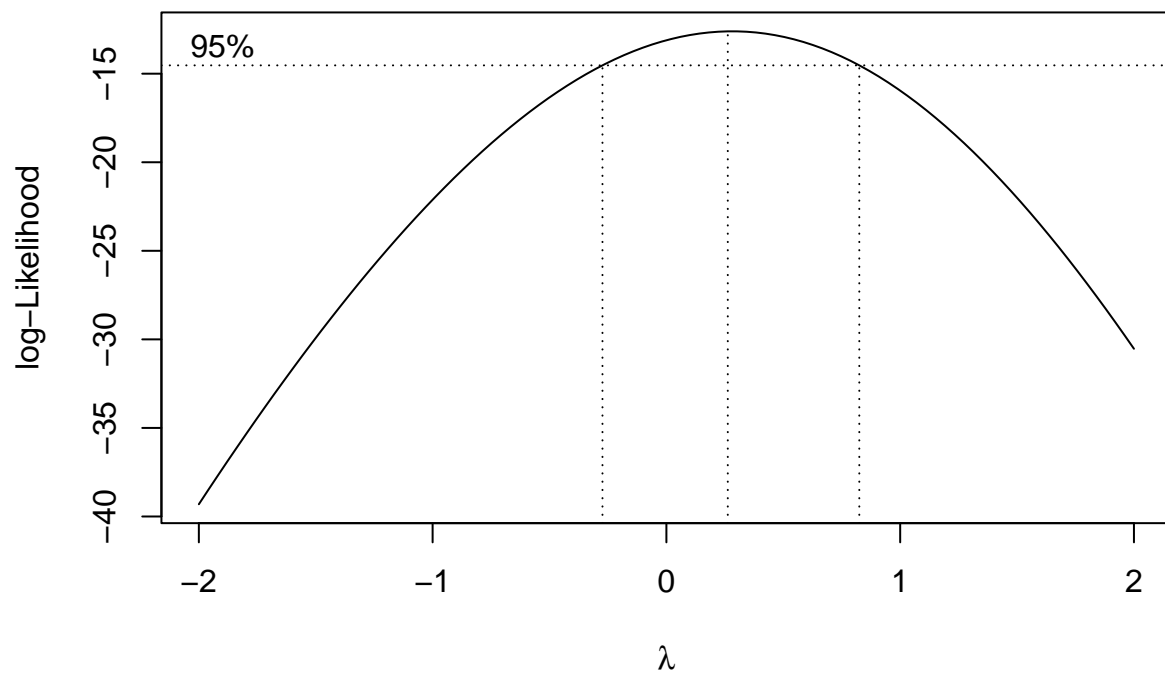








```
#since general models do not assume model assumptions, try box cox transformation
library(MASS)
bc = boxcox(model, lambda = seq(-2, 2, by = 0.1))
```



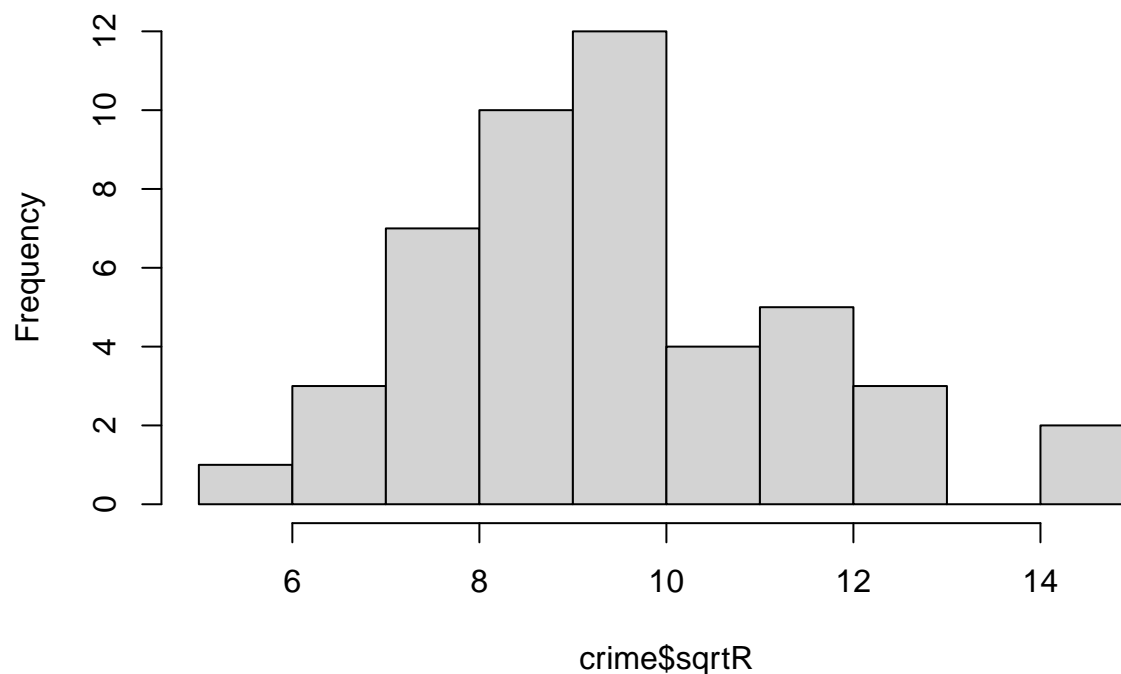
```
optimal_lambda = bc$x[which.max(bc$y)]
print(paste("optimal lambda:", optimal_lambda))
```

```
## [1] "optimal lambda: 0.262626262626263"
```

```
#since lambda = 0.26 closer to 0.5, try square root transformation
```

```
crime$sqrtR = sqrt(crime$R)
hist(crime$sqrtR)
```

Histogram of crime\$sqrtR

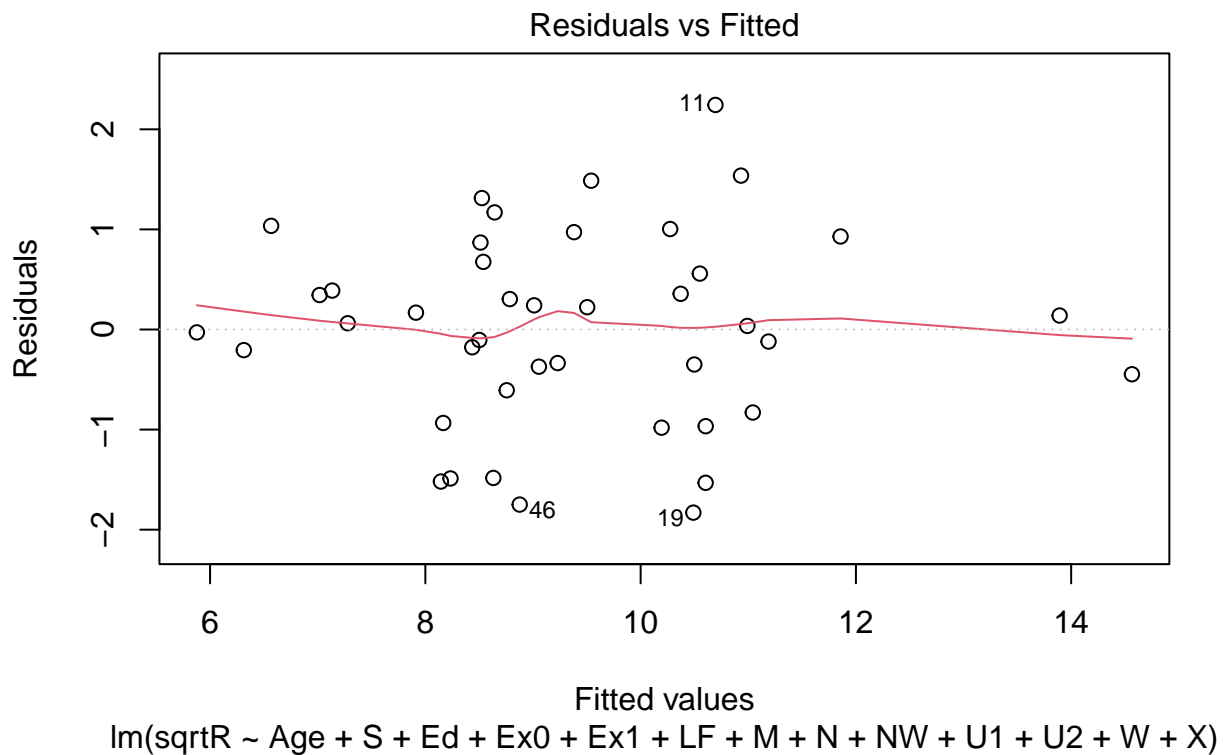


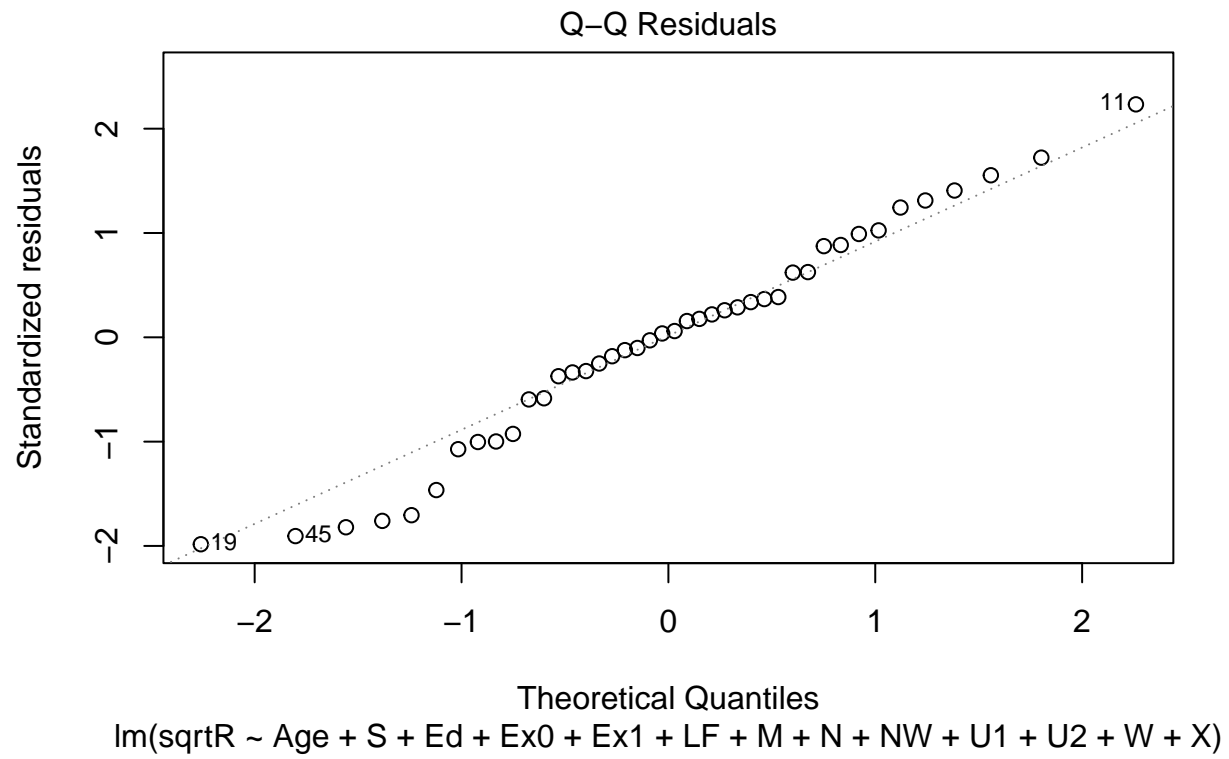
```
model4 = lm(sqrtR~Age+S+Ed+Ex0+Ex1+LF+M+N+NW+U1+U2+W+X, crime)
summary(model4)
```

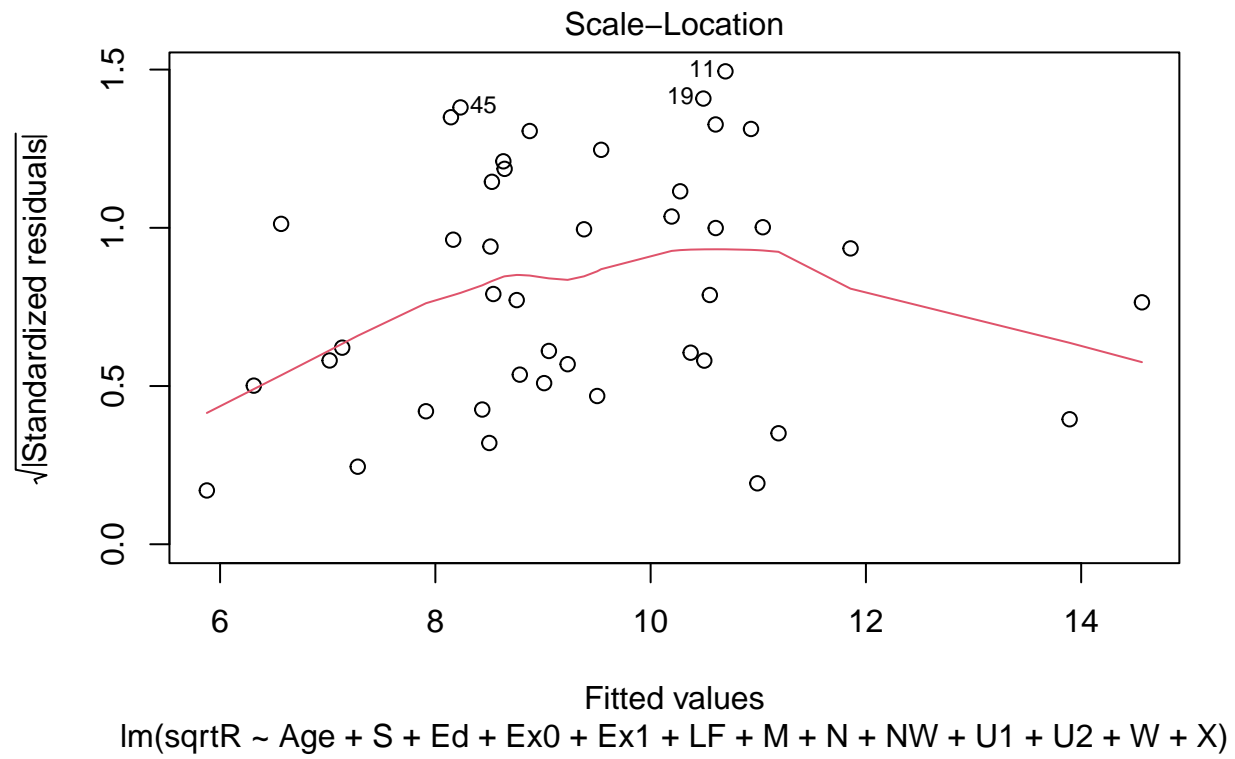
```
##
## Call:
## lm(formula = sqrtR ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N +
##     NW + U1 + U2 + W + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82957 -0.56724  0.04916  0.64648  2.24340
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -31.394105   8.812013  -3.563  0.00134 **
## Age          0.077095   0.029551   2.609  0.01441 *
## S            0.100683   0.853116   0.118  0.90690
## Ed           0.073750   0.041125   1.793  0.08373 .
## Ex0          0.046908   0.062166   0.755  0.45681
## Ex1          0.004438   0.068690   0.065  0.94894
## LF           0.007620   0.012788   0.596  0.55606
## M            0.001079   0.013107   0.082  0.93500
## N           -0.002597   0.007289  -0.356  0.72430
## NW          -0.001456   0.003868  -0.376  0.70954
## U1          -0.026632   0.026752  -0.996  0.32801
```

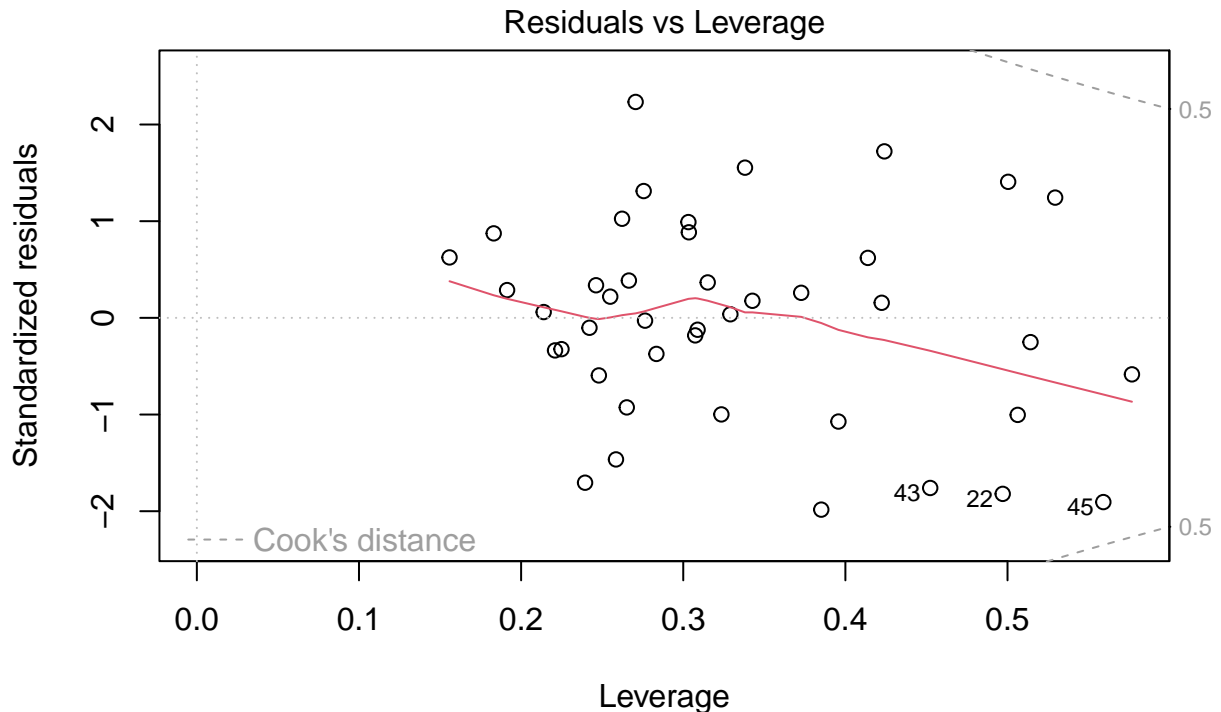
```
## U2          0.105913    0.050036    2.117  0.04330 *
## W           0.007992    0.005772    1.385  0.17712
## X           0.039075    0.013827    2.826  0.00860 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.176 on 28 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.7742, Adjusted R-squared:  0.6693
## F-statistic: 7.384 on 13 and 28 DF,  p-value: 5.277e-06
```

```
plot(model4)
```









$\text{lm}(\text{sqrtR} \sim \text{Age} + \text{S} + \text{Ed} + \text{Ex0} + \text{Ex1} + \text{LF} + \text{M} + \text{N} + \text{NW} + \text{U1} + \text{U2} + \text{W} + \text{X})$

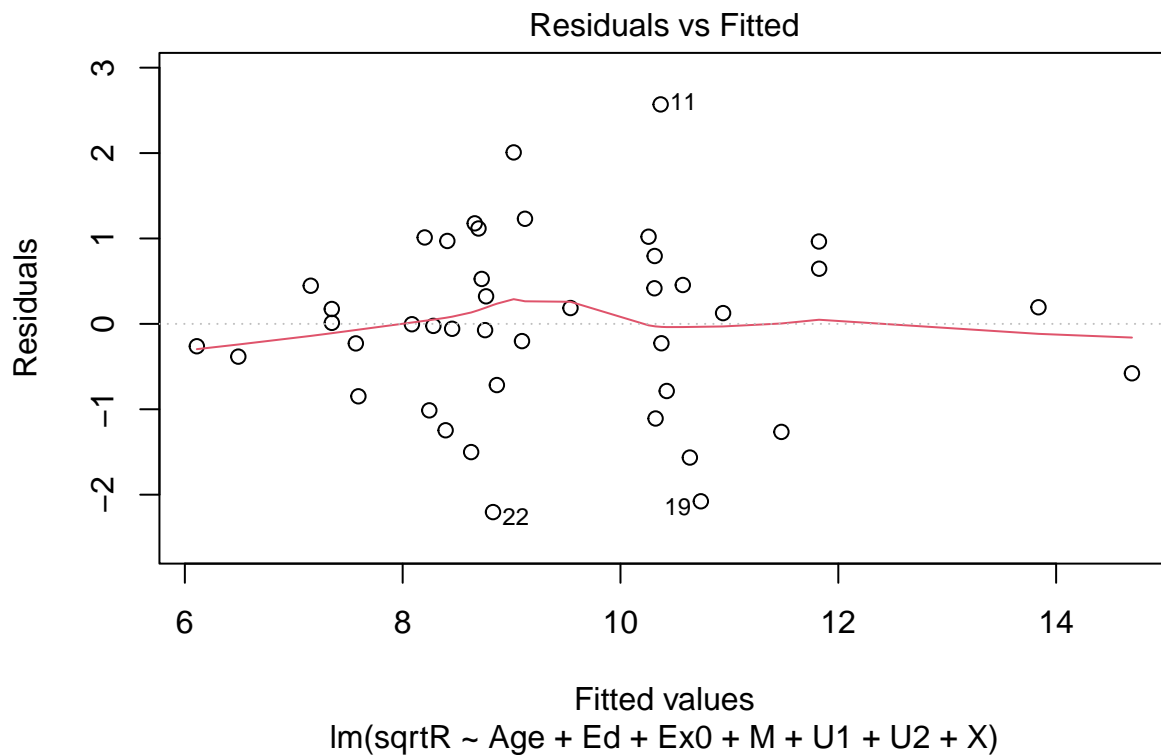
```
#adjusted R^2 = 66.93
#still double bow
#qq residuals fit better on upper tail
#non linear scale
#no residuals
```

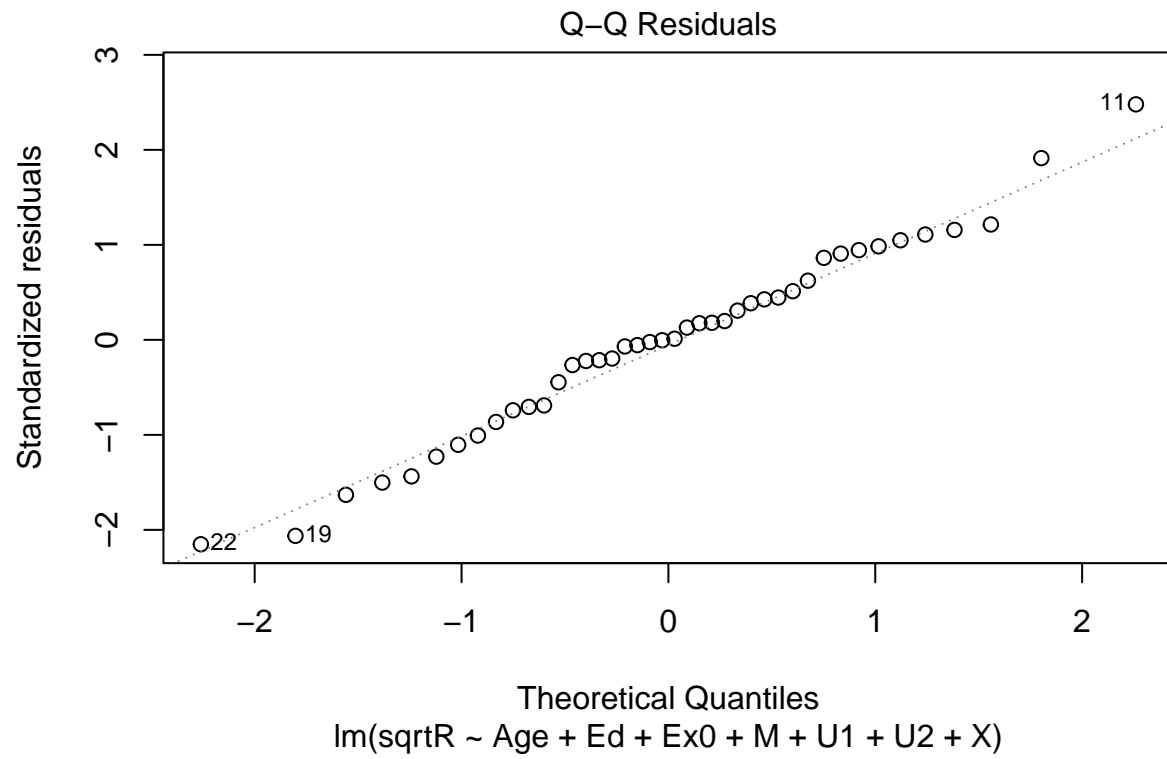
```
model5 = lm(sqrtR~Age + Ed + Ex0 + M + U1 + U2 + X, crime)
summary(model5)
```

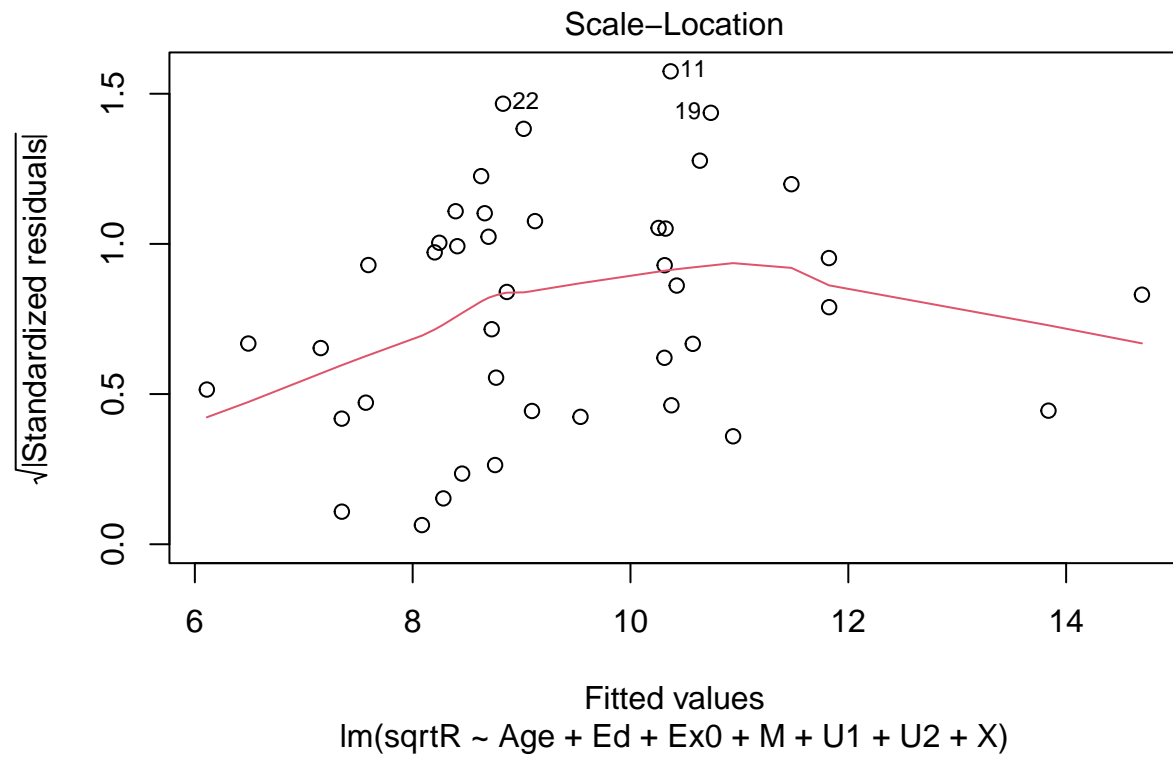
```
##
## Call:
## lm(formula = sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20388 -0.68198  0.00402  0.61598  2.56974
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29.271343   7.178983  -4.077 0.000259 ***
## Age          0.060497   0.022449   2.695 0.010865 *
## Ed           0.098533   0.031207   3.157 0.003329 **
## Ex0          0.053798   0.008859   6.073 6.93e-07 ***
## M            0.009824   0.008132   1.208 0.235378
## U1          -0.039929   0.020731  -1.926 0.062490 .
##
```

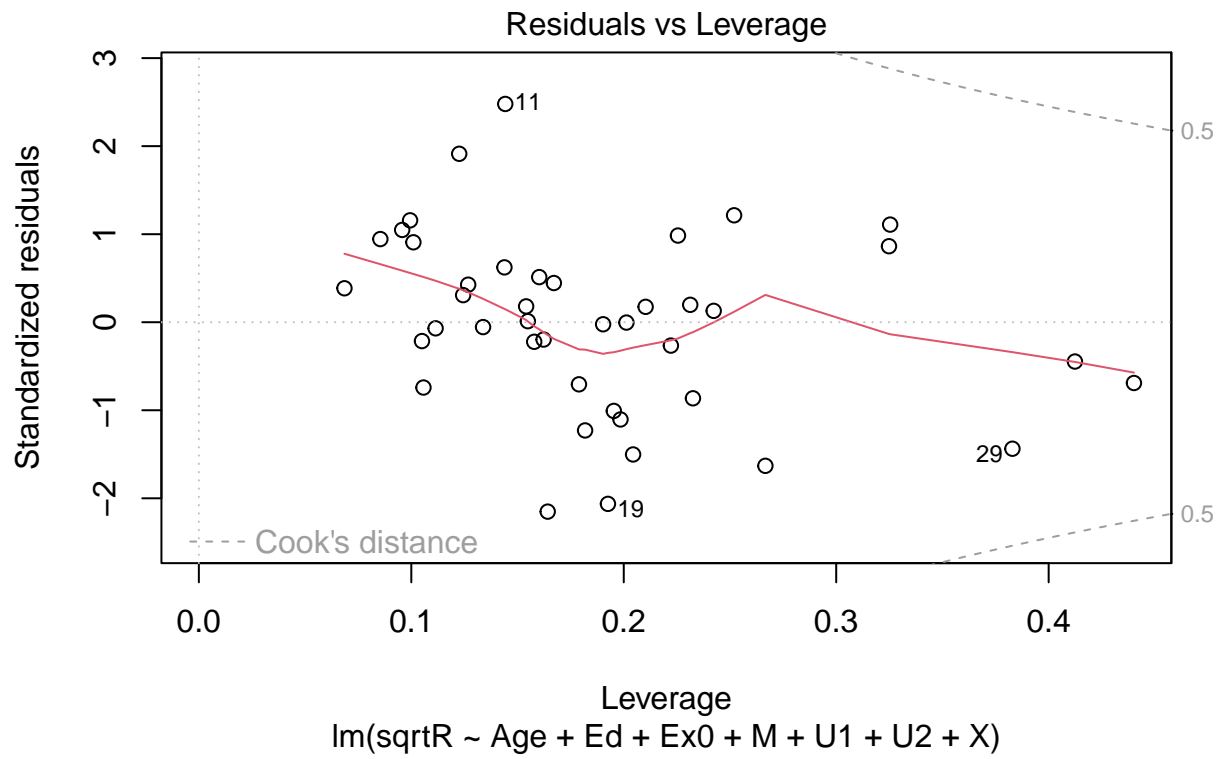
```
## U2          0.116941    0.044501    2.628 0.012802 *
## X           0.028085    0.008206    3.423 0.001633 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.12 on 34 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.7511, Adjusted R-squared:  0.6999
## F-statistic: 14.66 on 7 and 34 DF,  p-value: 1.262e-08

#adjusted R^2 = 69.99 better
plot(model5)
```





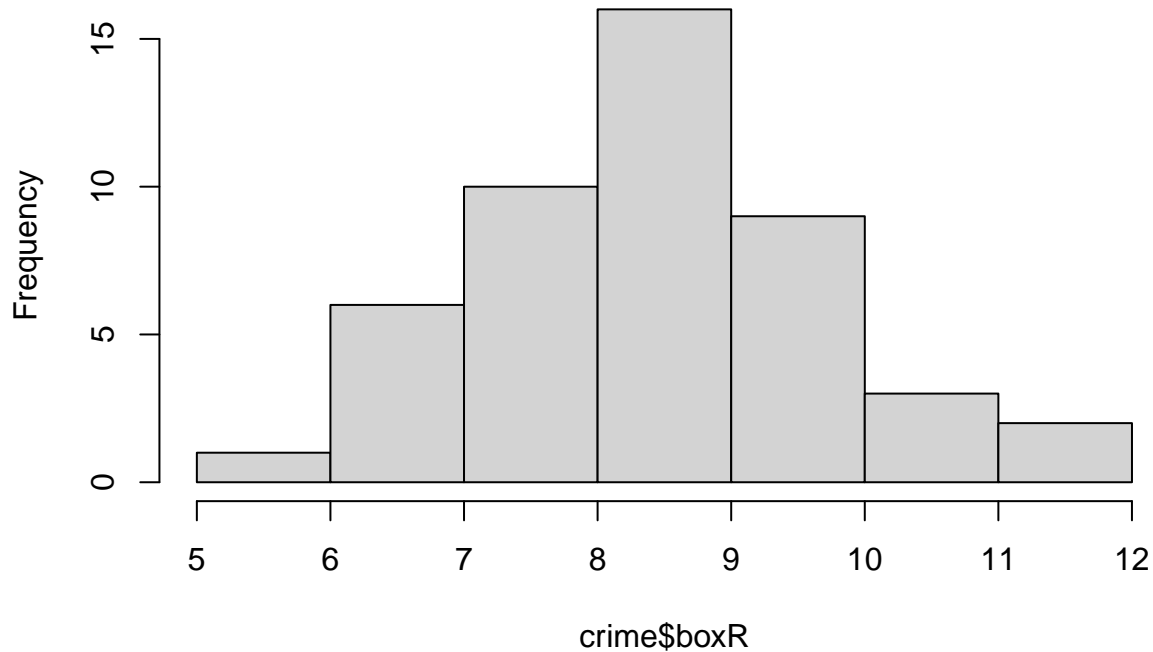




```
#double bow
#residuals fit line fairly well
#non linear scale
#no residuals

#boxcox trans
crime$boxR = (crime$R^optimal_lambda-1)/optimal_lambda
hist(crime$boxR)
```

Histogram of crime\$boxR

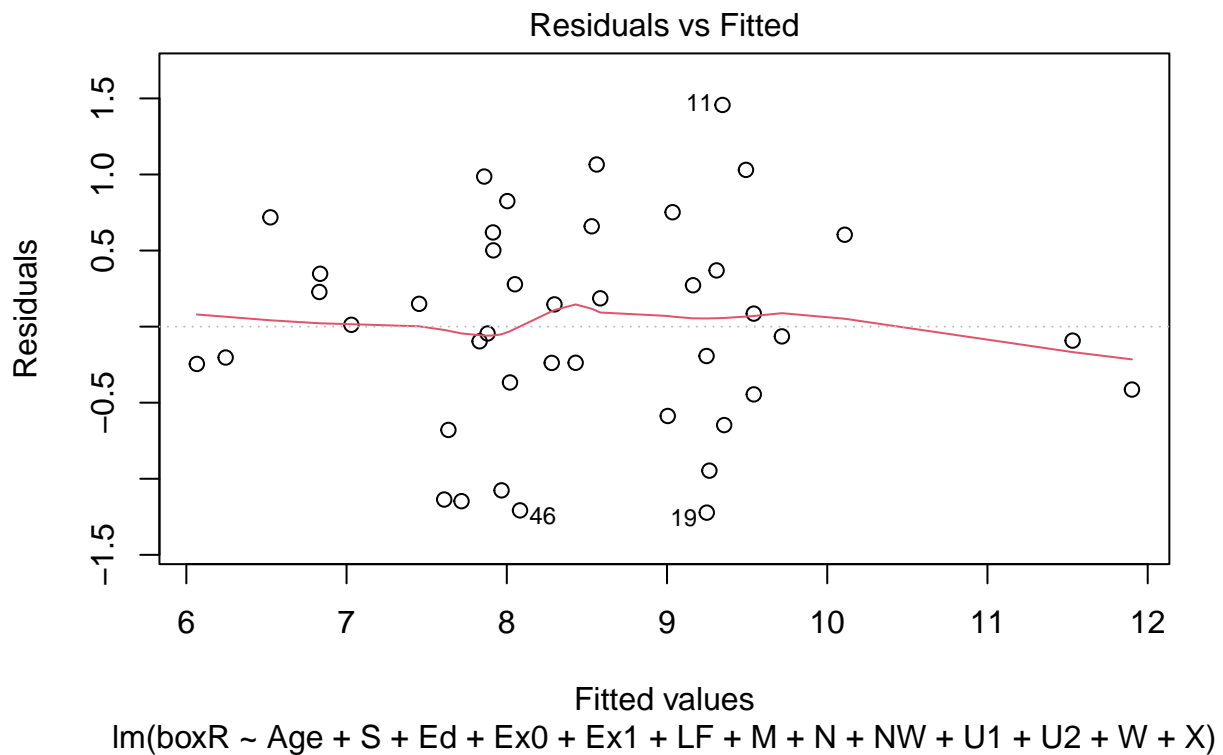


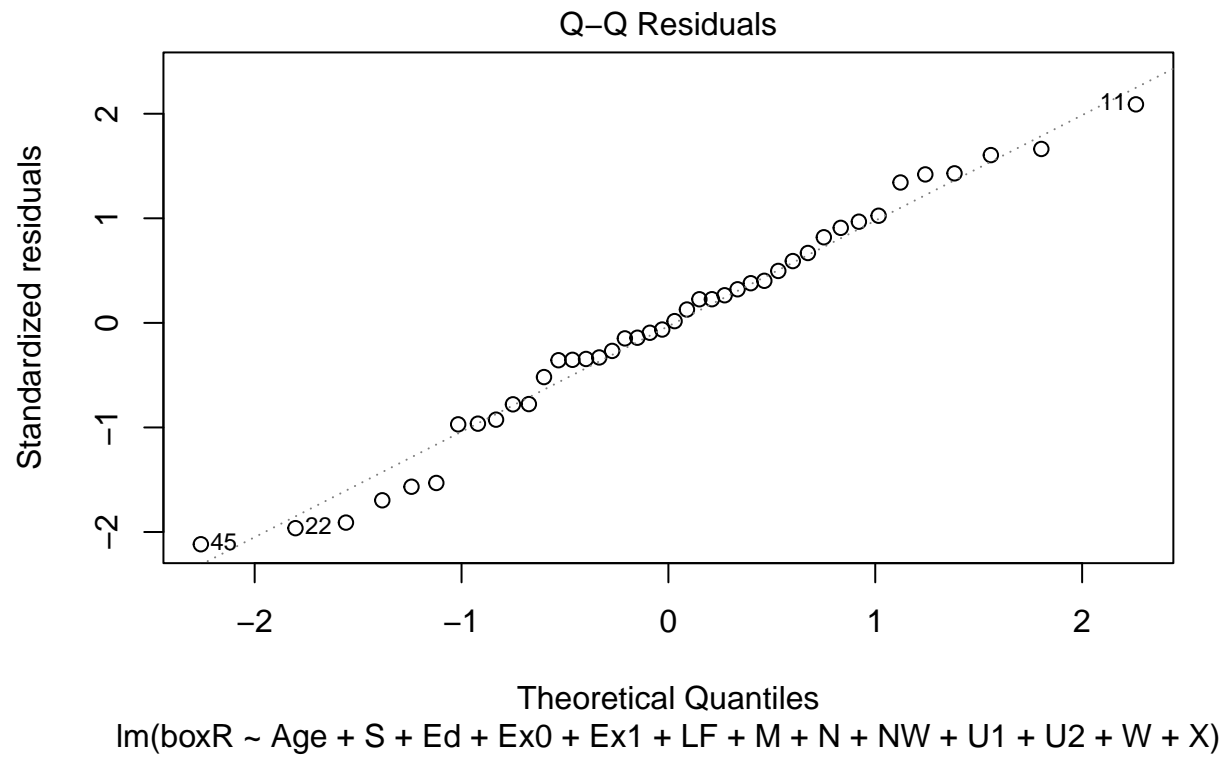
```
model6 = lm(boxR~Age+S+Ed+Ex0+Ex1+LF+M+N+NW+U1+U2+W+X, crime)
summary(model6)
```

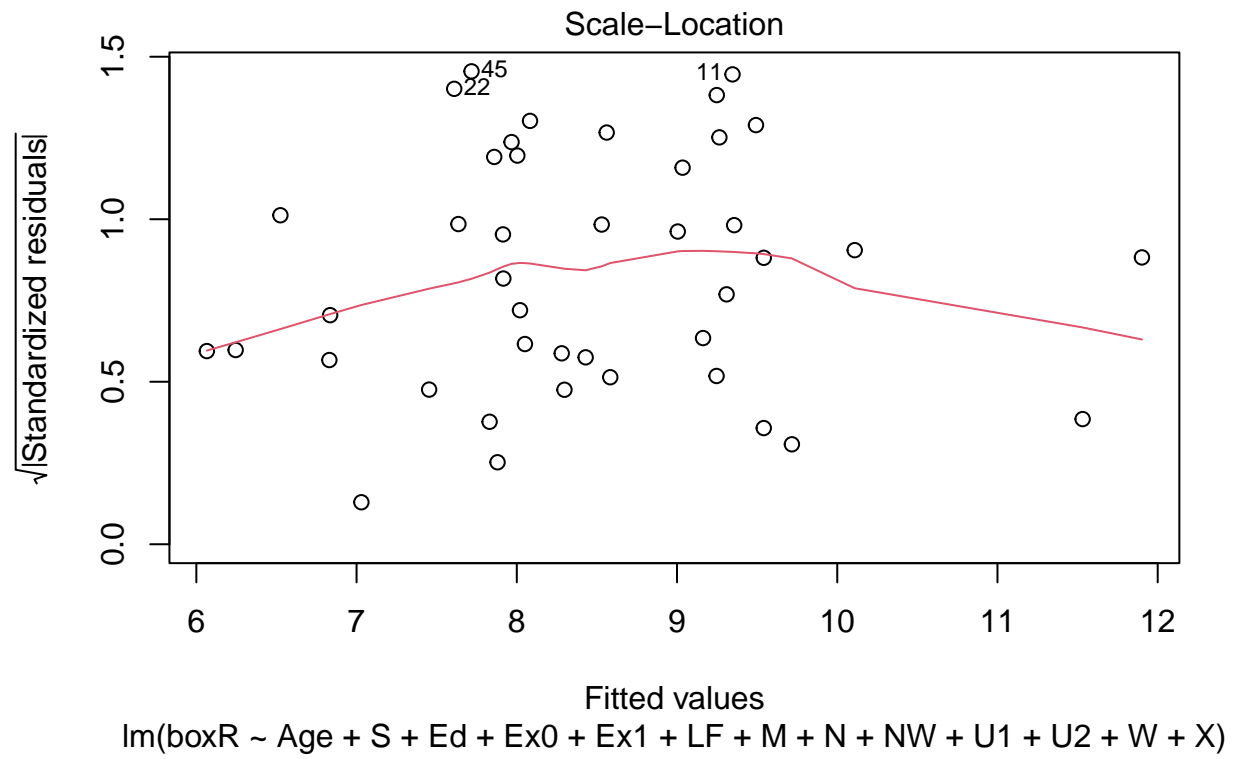
```
##
## Call:
## lm(formula = boxR ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW +
##      U1 + U2 + W + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22255 -0.40197 -0.01655  0.46825  1.45702
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.874e+01  6.116e+00  -3.065  0.00478 **
## Age          5.485e-02  2.051e-02   2.674  0.01236 *
## S            1.686e-01  5.921e-01   0.285  0.77788
## Ed           5.055e-02  2.854e-02   1.771  0.08745 .
## Ex0          2.900e-02  4.314e-02   0.672  0.50695
## Ex1          5.963e-03  4.767e-02   0.125  0.90135
## LF           6.677e-03  8.875e-03   0.752  0.45817
## M           -1.586e-03  9.097e-03  -0.174  0.86285
## N           -2.122e-03  5.059e-03  -0.419  0.67814
## NW          -9.791e-04  2.685e-03  -0.365  0.71806
## U1          -1.527e-02  1.857e-02  -0.823  0.41764
```

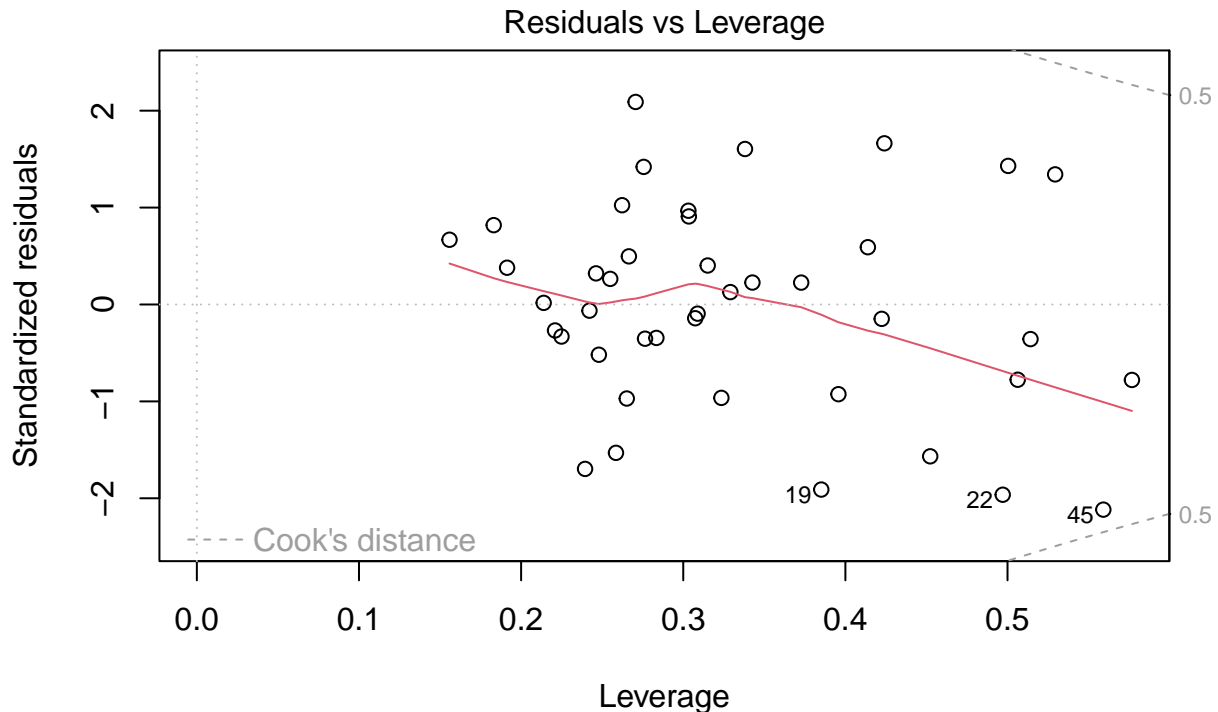
```
## U2          7.042e-02  3.473e-02  2.028  0.05217 .
## W           5.965e-03  4.006e-03  1.489  0.14766
## X           2.676e-02  9.597e-03  2.788  0.00941 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8163 on 28 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.7679, Adjusted R-squared:  0.6601
## F-statistic: 7.126 on 13 and 28 DF, p-value: 7.44e-06
```

```
plot(model6)
```









lm(boxR ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW + U1 + U2 + W + X)

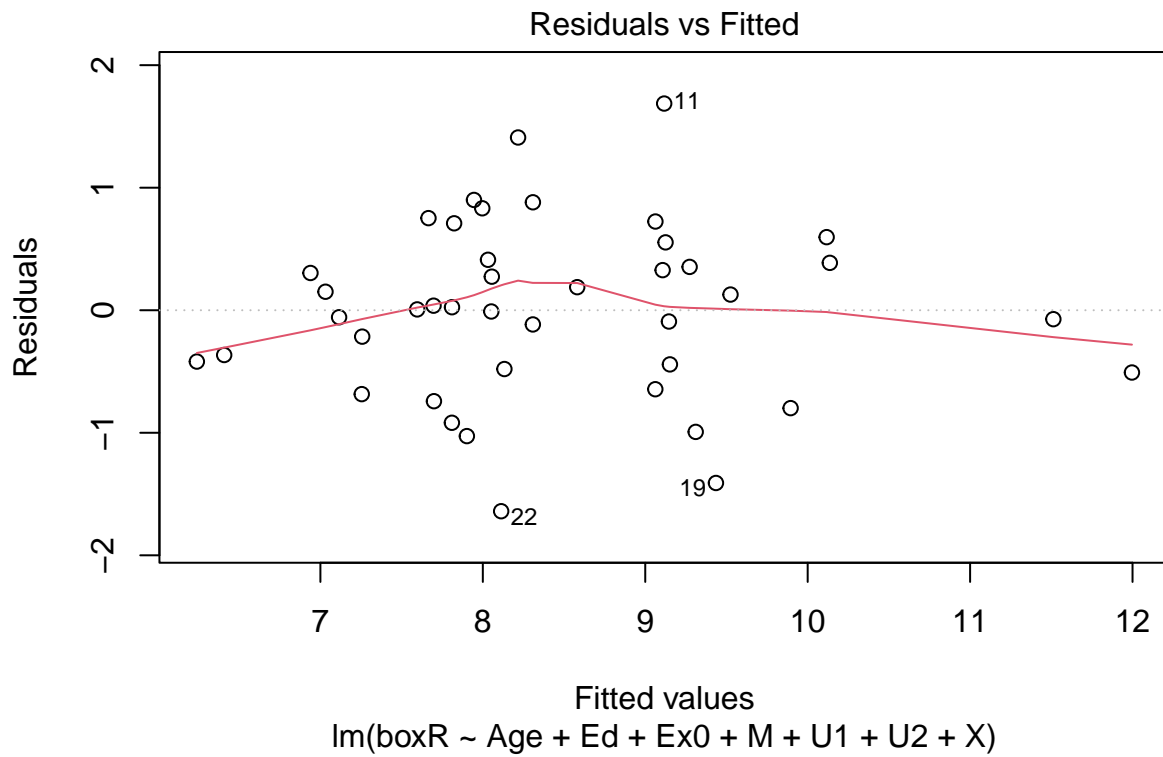
#R² adjusted 66.01%

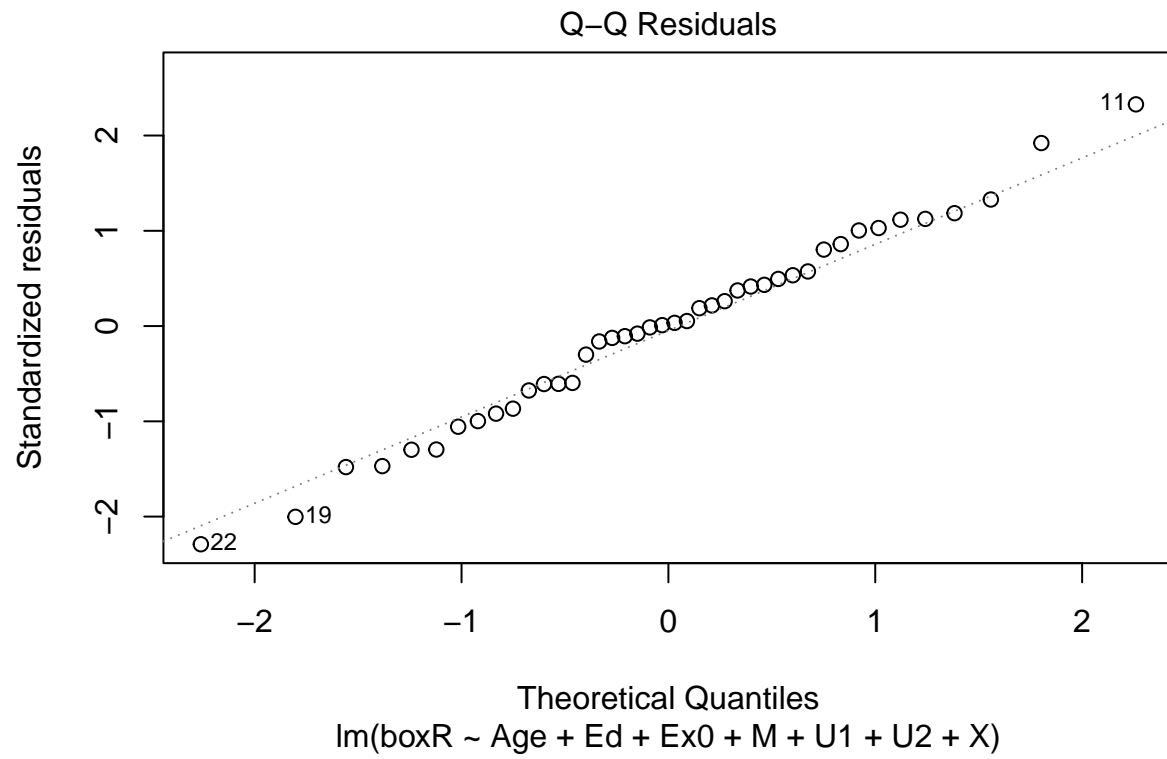
```
model7 = lm(boxR~Age + Ed + Ex0 + M + U1 + U2 + X, crime)
summary(model7)
```

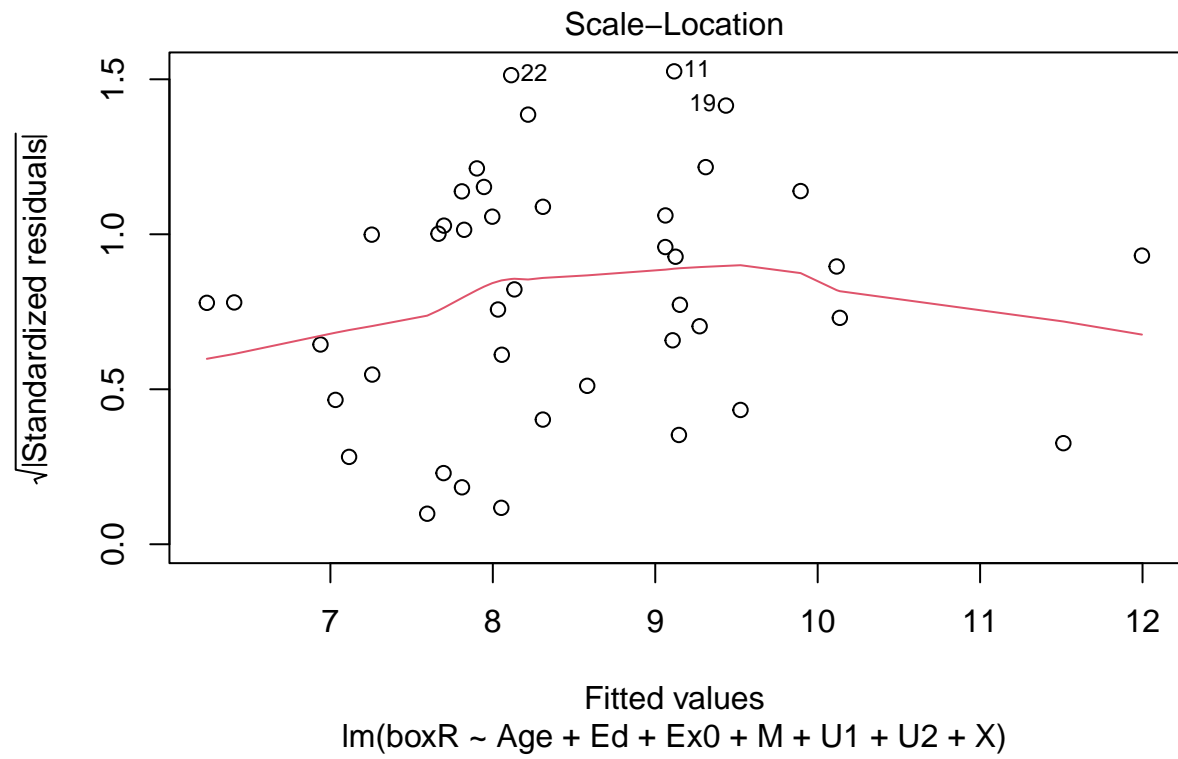
```
##
## Call:
## lm(formula = boxR ~ Age + Ed + Ex0 + M + U1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.64020 -0.47037  0.01569  0.40543  1.68671
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.289755   5.019263  -3.445  0.00154 **
## Age           0.042421   0.015696   2.703  0.01065 *
## Ed            0.070171   0.021818   3.216  0.00285 **
## Ex0           0.036416   0.006194   5.880 1.23e-06 ***
## M             0.005528   0.005686   0.972  0.33782
## U1           -0.026926   0.014494  -1.858  0.07189 .
## U2            0.080274   0.031114   2.580  0.01437 *
## X             0.019387   0.005737   3.379  0.00184 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

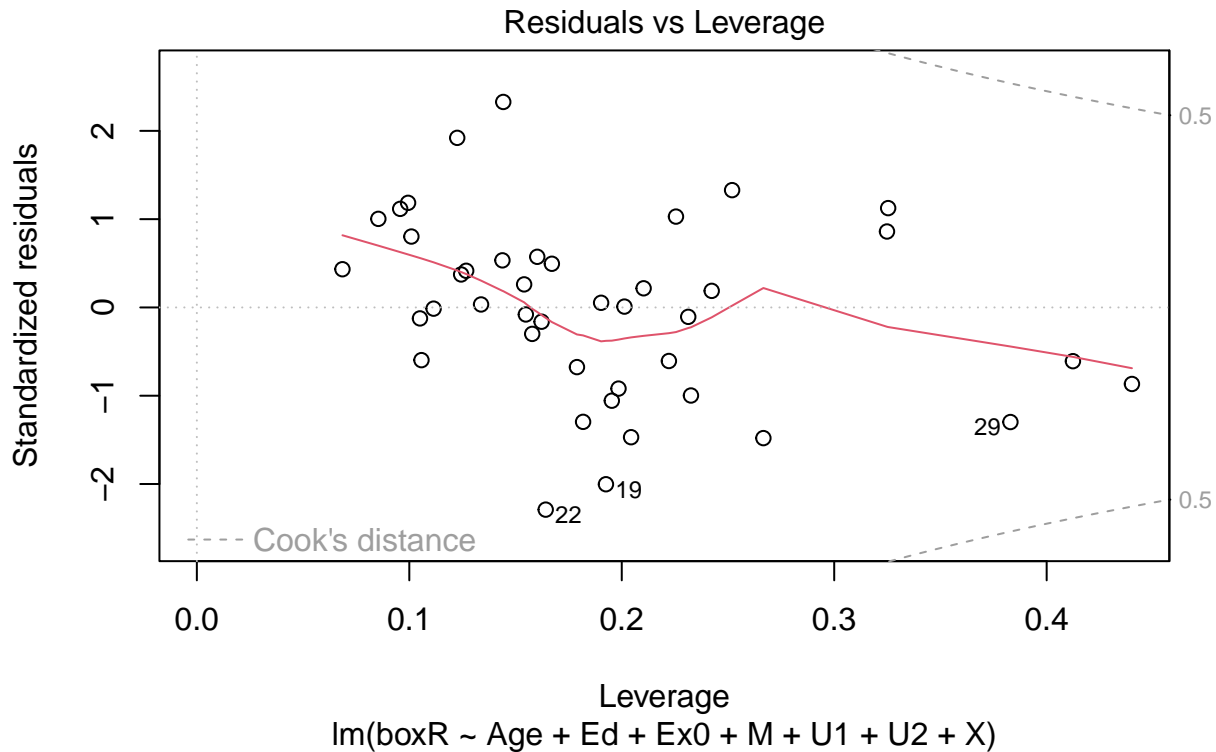
```
##
## Residual standard error: 0.7834 on 34 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared: 0.7404, Adjusted R-squared: 0.687
## F-statistic: 13.85 on 7 and 34 DF, p-value: 2.5e-08
```

```
plot(model7)
```









```
#R^2 adjusted 68.7%
```

```
#potential outliers
cooksD <- cooks.distance(model)
which(cooksD > 1)
```

```
## named integer(0)
```

```
if(!require("car")) {install.packages("car", dependencies = TRUE)}
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 4.5.1
```

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

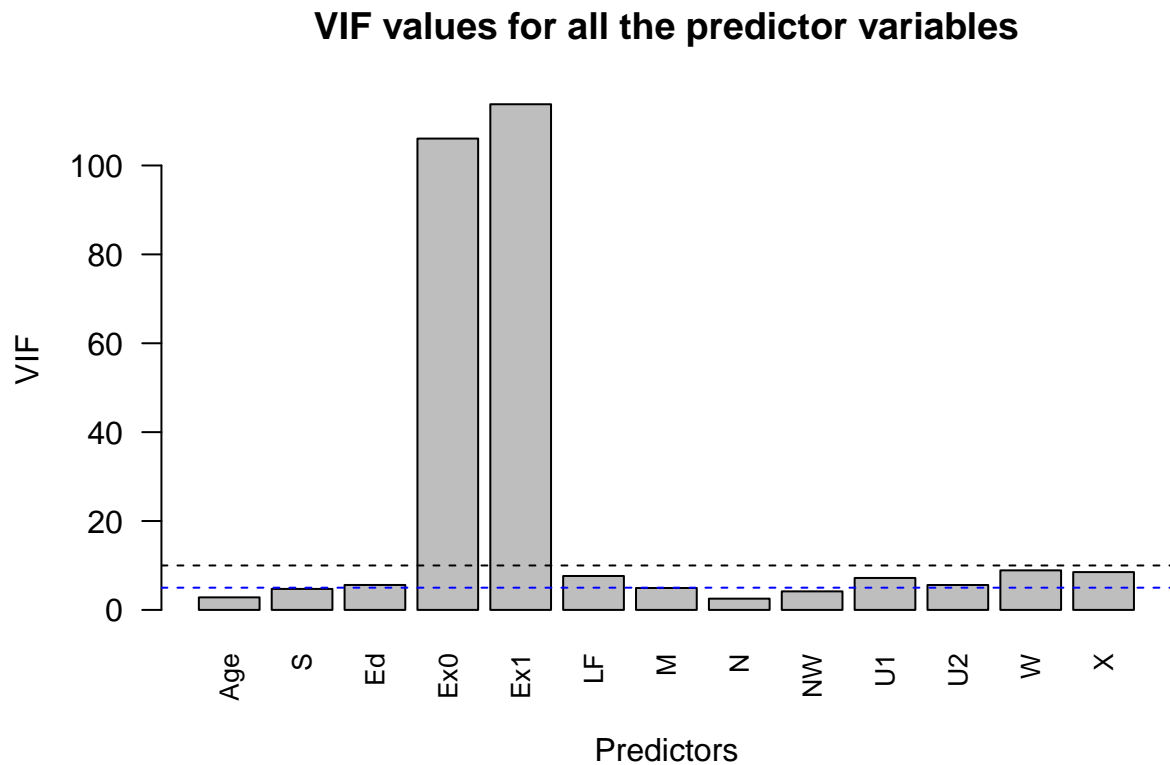
```
library(car)
print(vif(model))
```

```
##      Age      S      Ed      Ex0      Ex1      LF      M
## 2.822025 4.722406 5.611892 106.055900 113.779026 7.634529 4.930182
##      N      NW      U1      U2      W      X
## 2.534878 4.168548 7.190151 5.605829 8.895448 8.510396
```

```

vifval<- vif(model)
barplot(vifval, main = "VIF values for all the predictor variables",
        ylab = "VIF",
        xlab = "Predictors",
        col = "gray",
        las = 2,
        cex.names = 0.8)
abline(h = 10, col = "black", lty = 2)
abline(h = 5, col = "blue", lty = 2)

```



```

#multicollinearity between ex0 and Ex1 based on vif values
#note multicollinearity between Exo and ex1 and potentially other varaibles, havent learned how to deal

#try interactions to fix model not following assumptions, backwards selection using interaction terms
model_start <- lm(sqrtR ~ (Age + Ed + Ex0 + M + U1 + U2 + X)^2, data = crime)
model_step <- step(model_start, direction = "both")

```

```

## Start:  AIC=-2.47
## sqrtR ~ (Age + Ed + Ex0 + M + U1 + U2 + X)^2
##
##      Df Sum of Sq  RSS   AIC
## - M:U1    1    0.0128 9.9652 -4.4198
## - Ed:M    1    0.0185 9.9709 -4.3958
## - Age:U2   1    0.0421 9.9945 -4.2968

```



```

## - Age:Ex0 1 0.0984 10.0508 -4.0605
## - Age:M 1 0.3145 10.2669 -3.1673
## - Age:U1 1 0.3889 10.3413 -2.8639
## - U2:X 1 0.4339 10.3863 -2.6816
## <none> 9.9524 -2.4739
## - Ex0:U1 1 0.5943 10.5467 -2.0380
## - U1:X 1 0.8223 10.7747 -1.1397
## - M:X 1 0.8798 10.8322 -0.9160
## - Ex0:X 1 0.9057 10.8581 -0.8158
## - Ex0:M 1 1.0817 11.0341 -0.1406
## - M:U2 1 1.4317 11.3841 1.1711
## - Ex0:U2 1 1.7150 11.6674 2.2035
## - Ed:X 1 1.9743 11.9267 3.1266
## - Age:X 1 2.3456 12.2980 4.4141
## - U1:U2 1 3.0117 12.9641 6.6296
## - Ed:Ex0 1 3.2904 13.2428 7.5230
## - Ed:U2 1 4.0093 13.9617 9.7432
## - Ed:U1 1 4.0614 14.0138 9.8997
## - Age:Ed 1 5.3924 15.3448 13.7104
##
## Step: AIC=-4.42
## sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed + Age:Ex0 +
## Age:M + Age:U1 + Age:U2 + Age:X + Ed:Ex0 + Ed:M + Ed:U1 +
## Ed:U2 + Ed:X + Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X +
## U1:U2 + U1:X + U2:X
##
## Df Sum of Sq RSS AIC
## - Ed:M 1 0.0167 9.9819 -6.3495
## - Age:U2 1 0.0313 9.9965 -6.2882
## - Age:Ex0 1 0.0856 10.0509 -6.0604
## - Age:U1 1 0.3855 10.3508 -4.8256
## - Age:M 1 0.4351 10.4003 -4.6251
## - U2:X 1 0.4633 10.4285 -4.5114
## <none> 9.9652 -4.4198
## - U1:X 1 0.8625 10.8277 -2.9336
## - Ex0:X 1 0.9541 10.9193 -2.5796
## - M:X 1 0.9631 10.9283 -2.5452
## + M:U1 1 0.0128 9.9524 -2.4739
## - Ex0:U1 1 1.0335 10.9987 -2.2754
## - Ex0:M 1 1.5734 11.5386 -0.2628
## - Ed:X 1 2.0999 12.0651 1.6113
## - Age:X 1 2.8537 12.8189 4.1567
## - Ex0:U2 1 2.9899 12.9551 4.6004
## - U1:U2 1 3.0005 12.9657 4.6349
## - Ed:Ex0 1 3.5179 13.4832 6.2784
## - Ed:U2 1 5.1662 15.1314 11.1223
## - Ed:U1 1 5.1903 15.1555 11.1892
## - M:U2 1 5.3686 15.3338 11.6804
## - Age:Ed 1 5.4536 15.4188 11.9126
##
## Step: AIC=-6.35
## sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed + Age:Ex0 +
## Age:M + Age:U1 + Age:U2 + Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 +
## Ed:X + Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 +

```

```

##      U1:X + U2:X
##
##      Df Sum of Sq      RSS      AIC
## - Age:U2    1    0.0226 10.0046 -8.2544
## - Age:Ex0   1    0.1039 10.0859 -7.9144
## - Age:U1    1    0.3745 10.3565 -6.8025
## - U2:X      1    0.4478 10.4298 -6.5062
## <none>                9.9819 -6.3495
## - Age:M     1    0.4980 10.4799 -6.3048
## - U1:X      1    0.8978 10.8798 -4.7321
## - Ex0:X     1    0.9406 10.9225 -4.5674
## + Ed:M      1    0.0167  9.9652 -4.4198
## + M:U1      1    0.0110  9.9709 -4.3958
## - Ex0:U1    1    1.0335 11.0155 -4.2115
## - M:X       1    1.0720 11.0539 -4.0652
## - Ex0:M     1    1.9475 11.9295 -0.8636
## - Ed:X      1    2.2134 12.1953  0.0621
## - Age:X     1    2.8657 12.8477  2.2507
## - Ex0:U2    1    2.9745 12.9565  2.6048
## - U1:U2     1    3.0251 13.0071  2.7686
## - Ed:Ex0    1    3.6028 13.5848  4.5938
## - Ed:U2     1    5.4610 15.4429  9.9782
## - Age:Ed    1    5.4657 15.4477  9.9911
## - M:U2      1    5.5844 15.5664 10.3126
## - Ed:U1     1    6.7089 16.6908 13.2419
##
## Step:  AIC=-8.25
## sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed + Age:Ex0 +
##      Age:M + Age:U1 + Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 + Ed:X +
##      Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 + U1:X +
##      U2:X
##
##      Df Sum of Sq      RSS      AIC
## - Age:Ex0   1    0.2179 10.2224 -9.3496
## <none>                10.0046 -8.2544
## - Age:M     1    0.5554 10.5599 -7.9854
## - Age:U1    1    0.7365 10.7411 -7.2710
## - U2:X      1    0.7923 10.7969 -7.0534
## + Age:U2    1    0.0226  9.9819 -6.3495
## + Ed:M      1    0.0081  9.9965 -6.2882
## + M:U1      1    0.0022 10.0023 -6.2638
## - Ex0:U1    1    1.0354 11.0400 -6.1182
## - M:X       1    1.0709 11.0754 -5.9836
## - U1:X      1    1.2217 11.2262 -5.4156
## - Ex0:X     1    1.7370 11.7416 -3.5304
## - Ex0:M     1    1.9251 11.9296 -2.8631
## - Ed:X      1    2.6176 12.6222 -0.4931
## - U1:U2     1    3.1608 13.1654  1.2768
## - Ex0:U2    1    3.1648 13.1694  1.2894
## - Age:X     1    3.3397 13.3442  1.8435
## - Ed:Ex0    1    4.3341 14.3386  4.8621
## - M:U2      1    5.7542 15.7587  8.8284
## - Ed:U2     1    7.0749 17.0795 12.2088
## - Ed:U1     1    7.4486 17.4532 13.1178

```

```
## - Age:Ed    1      8.1848 18.1894 14.8531
##
## Step:  AIC=-9.35
## sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed + Age:M + Age:U1 +
##      Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 + Ed:X + Ex0:M + Ex0:U1 +
##      Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 + U1:X + U2:X
##
##           Df Sum of Sq    RSS    AIC
## <none>                10.222 -9.3496
## - Age:U1    1      0.6904 10.913 -8.6046
## + Age:Ex0   1      0.2179 10.005 -8.2544
## - Ex0:U1    1      0.8335 11.056 -8.0577
## - M:X       1      0.8556 11.078 -7.9738
## - Age:M     1      0.8630 11.085 -7.9456
## + Age:U2    1      0.1365 10.086 -7.9144
## + M:U1      1      0.0436 10.179 -7.5291
## + Ed:M      1      0.0151 10.207 -7.4117
## - U2:X      1      1.2141 11.437 -6.6361
## - U1:X      1      1.6904 11.913 -4.9224
## - Ex0:M     1      1.7317 11.954 -4.7770
## - Ed:X      1      2.4911 12.714 -2.1900
## - U1:U2     1      3.2476 13.470  0.2374
## - Ex0:U2    1      3.3742 13.597  0.6304
## - Ex0:X     1      3.4493 13.672  0.8618
## - Age:X     1      3.6550 13.877  1.4888
## - Ed:Ex0    1      4.7081 14.931  4.5610
## - M:U2      1      5.6150 15.837  7.0376
## - Ed:U2     1      7.8329 18.055 12.5423
## - Age:Ed    1      8.0498 18.272 13.0439
## - Ed:U1     1      8.0807 18.303 13.1148
```

```
summary(model_step)
```

```
##
## Call:
## lm(formula = sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed +
##      Age:M + Age:U1 + Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 + Ed:X +
##      Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 + U1:X +
##      U2:X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04259 -0.31191 -0.03354  0.28977  1.30621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.086e+02  1.319e+02   0.823 0.421750
## Age          -2.398e-01  8.869e-01  -0.270 0.790147
## Ed           -2.558e+00  6.293e-01  -4.065 0.000805 ***
## Ex0          -1.174e-01  1.991e-01  -0.590 0.563176
## M             2.045e-01  1.584e-01   1.290 0.214156
## U1           -7.984e-01  4.440e-01  -1.798 0.089940 .
## U2           -3.278e-01  9.347e-01  -0.351 0.730139
## X            -7.706e-02  3.869e-01  -0.199 0.844502
```

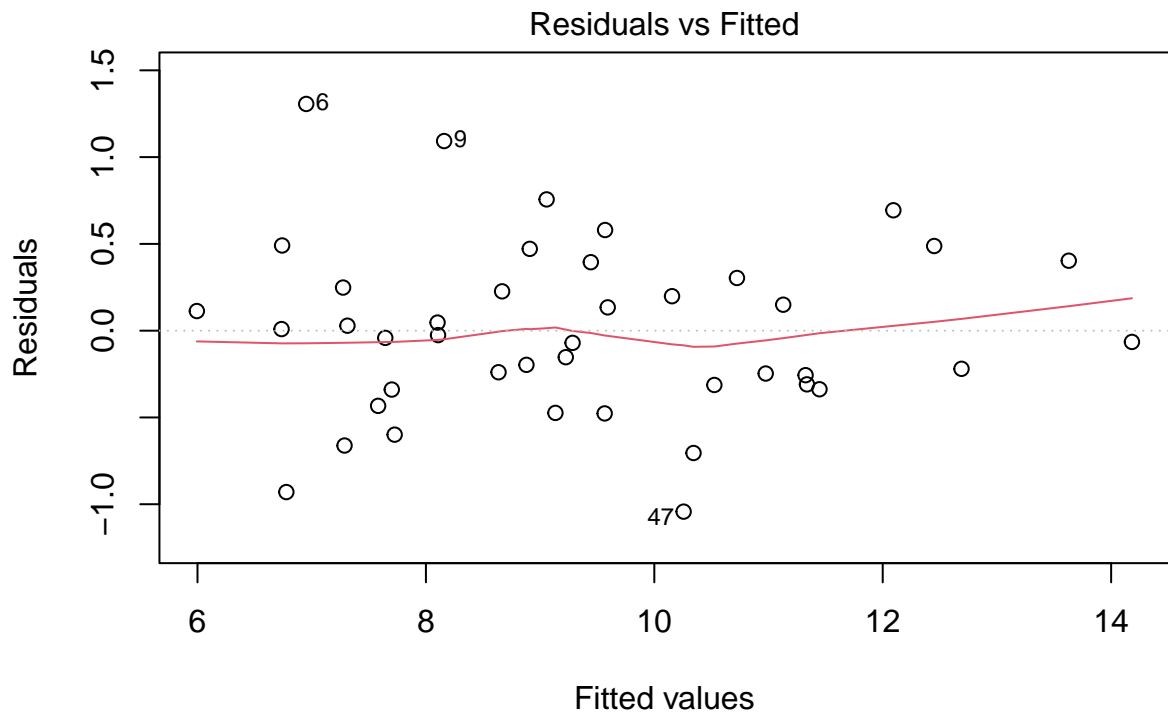
```
## Age:Ed      1.338e-02  3.657e-03   3.659 0.001944 **
## Age:M      -1.336e-03  1.115e-03  -1.198 0.247357
## Age:U1     -1.969e-03  1.838e-03  -1.072 0.298899
## Age:X       1.879e-03  7.623e-04   2.465 0.024628 *
## Ed:Ex0      3.667e-03  1.310e-03   2.798 0.012354 *
## Ed:U1       8.175e-03  2.230e-03   3.666 0.001915 **
## Ed:U2      -2.188e-02  6.062e-03  -3.609 0.002165 **
## Ed:X        2.270e-03  1.115e-03   2.035 0.057709 .
## Ex0:M      -4.303e-04  2.536e-04  -1.697 0.107926
## Ex0:U1     -9.032e-04  7.671e-04  -1.177 0.255289
## Ex0:U2      3.541e-03  1.495e-03   2.369 0.029953 *
## Ex0:X       7.608e-04  3.177e-04   2.395 0.028409 *
## M:U2        3.438e-03  1.125e-03   3.056 0.007150 **
## M:X        -5.377e-04  4.507e-04  -1.193 0.249323
## U1:U2      -3.435e-03  1.478e-03  -2.324 0.032779 *
## U1:X        1.627e-03  9.703e-04   1.677 0.111903
## U2:X       -2.479e-03  1.745e-03  -1.421 0.173423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7754 on 17 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.9404, Adjusted R-squared:  0.8563
## F-statistic: 11.18 on 24 and 17 DF, p-value: 2.419e-06
```

```
#best model found with interaction terms and improves R^2 adjusted
model9 = lm(sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed +
             Age:M + Age:U1 + Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 + Ed:X +
             Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 + U1:X +
             U2:X, data = crime)
summary(model9)
```

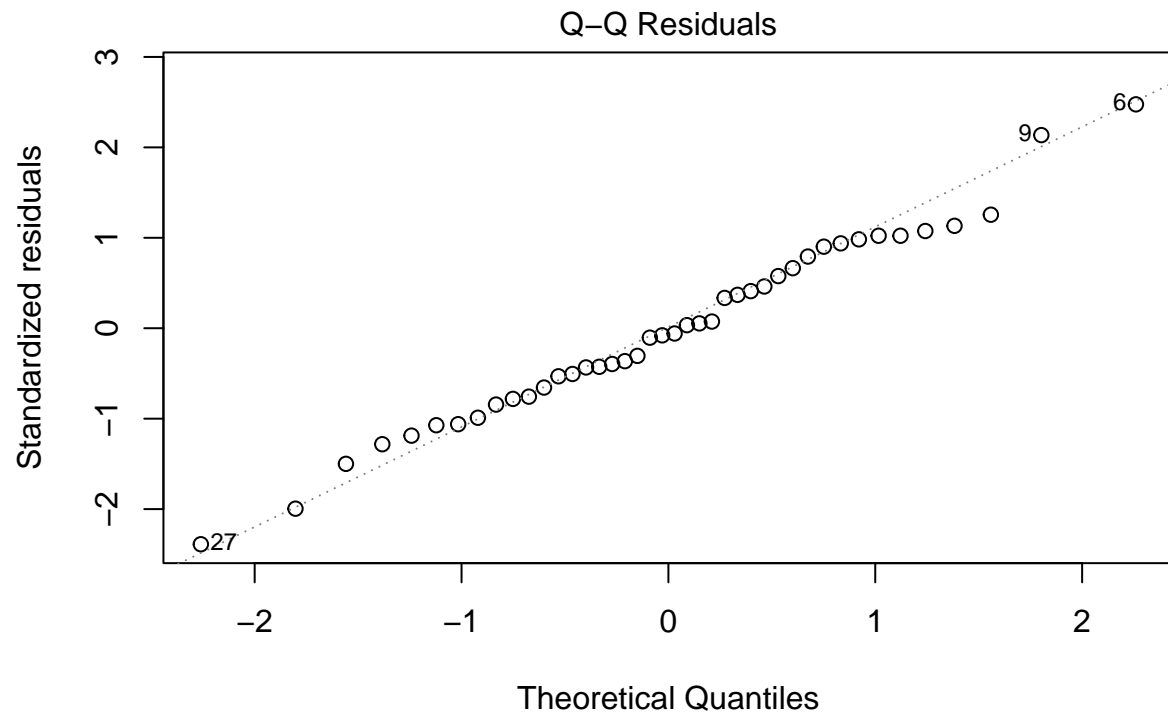
```
##
## Call:
## lm(formula = sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed +
##      Age:M + Age:U1 + Age:X + Ed:Ex0 + Ed:U1 + Ed:U2 + Ed:X +
##      Ex0:M + Ex0:U1 + Ex0:U2 + Ex0:X + M:U2 + M:X + U1:U2 + U1:X +
##      U2:X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04259 -0.31191 -0.03354  0.28977  1.30621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.086e+02  1.319e+02   0.823 0.421750
## Age         -2.398e-01  8.869e-01  -0.270 0.790147
## Ed          -2.558e+00  6.293e-01  -4.065 0.000805 ***
## Ex0         -1.174e-01  1.991e-01  -0.590 0.563176
## M            2.045e-01  1.584e-01   1.290 0.214156
## U1          -7.984e-01  4.440e-01  -1.798 0.089940 .
## U2          -3.278e-01  9.347e-01  -0.351 0.730139
## X           -7.706e-02  3.869e-01  -0.199 0.844502
## Age:Ed       1.338e-02  3.657e-03   3.659 0.001944 **
```

```
## Age:M      -1.336e-03  1.115e-03  -1.198  0.247357
## Age:U1     -1.969e-03  1.838e-03  -1.072  0.298899
## Age:X       1.879e-03  7.623e-04   2.465  0.024628 *
## Ed:Ex0      3.667e-03  1.310e-03   2.798  0.012354 *
## Ed:U1       8.175e-03  2.230e-03   3.666  0.001915 **
## Ed:U2      -2.188e-02  6.062e-03  -3.609  0.002165 **
## Ed:X        2.270e-03  1.115e-03   2.035  0.057709 .
## Ex0:M       -4.303e-04  2.536e-04  -1.697  0.107926
## Ex0:U1      -9.032e-04  7.671e-04  -1.177  0.255289
## Ex0:U2       3.541e-03  1.495e-03   2.369  0.029953 *
## Ex0:X        7.608e-04  3.177e-04   2.395  0.028409 *
## M:U2         3.438e-03  1.125e-03   3.056  0.007150 **
## M:X          -5.377e-04  4.507e-04  -1.193  0.249323
## U1:U2        -3.435e-03  1.478e-03  -2.324  0.032779 *
## U1:X         1.627e-03  9.703e-04   1.677  0.111903
## U2:X         -2.479e-03  1.745e-03  -1.421  0.173423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7754 on 17 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.9404, Adjusted R-squared:  0.8563
## F-statistic: 11.18 on 24 and 17 DF, p-value: 2.419e-06
```

```
#R2adjusted = 85.63%
plot(model9)
```

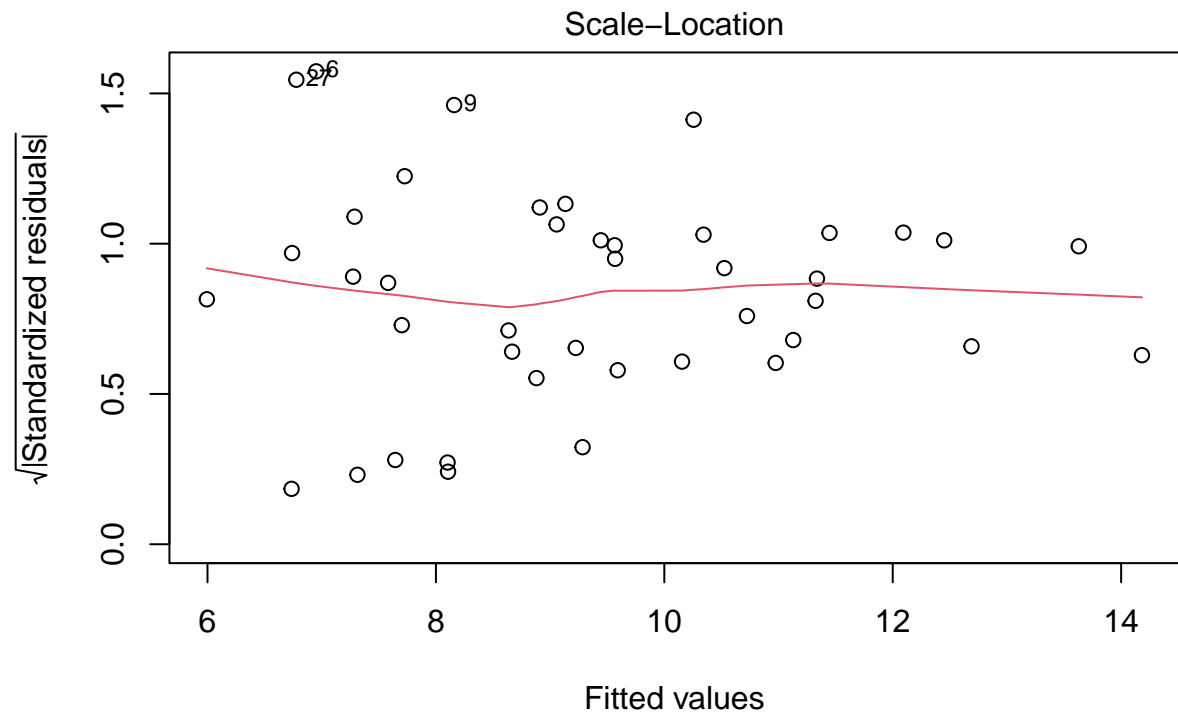


$\text{lm}(\text{sqrtR} \sim \text{Age} + \text{Ed} + \text{Ex0} + \text{M} + \text{U1} + \text{U2} + \text{X} + \text{Age:Ed} + \text{Age:M} + \text{Age:U1} + \text{Age} .$

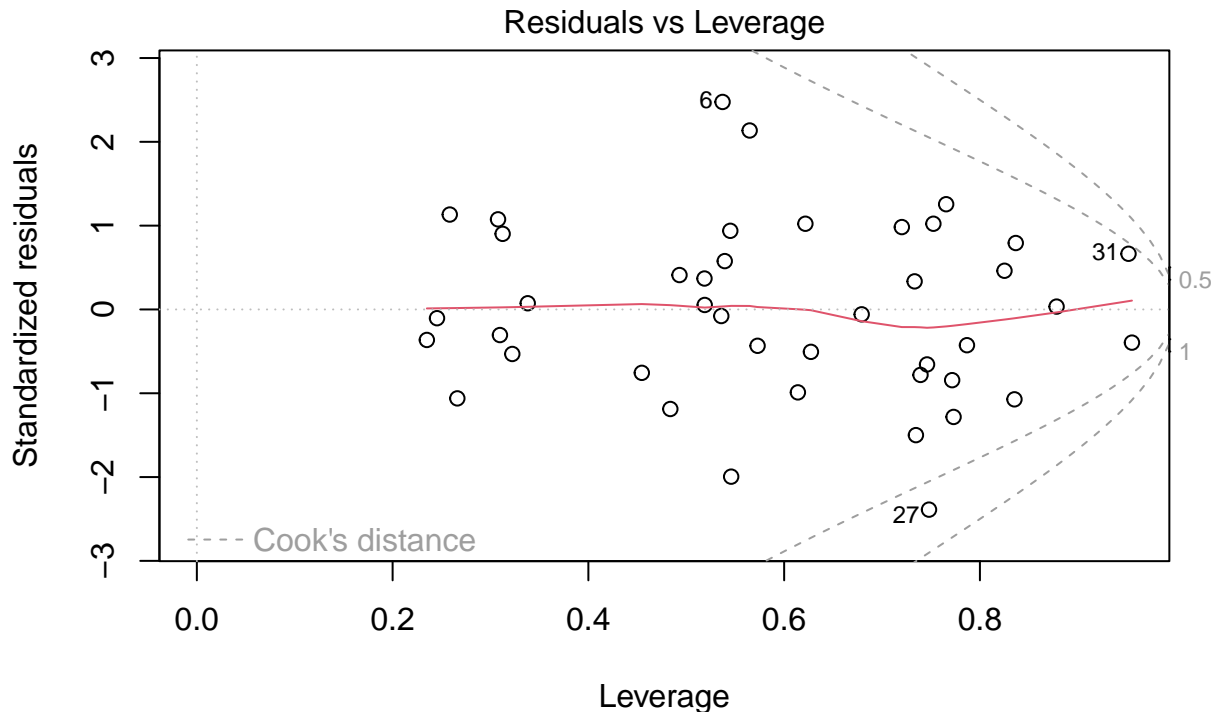


Theoretical Quantiles

$\text{lm}(\text{sqrtR} \sim \text{Age} + \text{Ed} + \text{Ex0} + \text{M} + \text{U1} + \text{U2} + \text{X} + \text{Age}:\text{Ed} + \text{Age}:\text{M} + \text{Age}:\text{U1} + \text{Age} .$



lm(sqrtR ~ Age + Ed + Ex0 + M + U1 + U2 + X + Age:Ed + Age:M + Age:U1 + Age .



```
#residuals appear scattered around 0
#qq residuals fit line better especiallu in bottom tail
#scale location appears more ranodm
# no outliers based on cooks, but improved model and better staisfys assumptions
```

```
#forward selection
crime2$sqrtR = sqrt(crime2$R)
model_start2 = lm(sqrtR ~ (Age+ Ed+Ex0+ LF+ U2+ W+ X)^2, crime2)
empty = lm(sqrtR~1, crime2)
step_for = step(empty, scope = formula(model_start2), direction = "forward")
```

```
## Start: AIC=61.1
## sqrtR ~ 1
##
##      Df Sum of Sq    RSS   AIC
## + Ex0  1    79.563  91.965 36.917
## + W    1    34.296 137.233 53.728
## + Ed   1    17.466 154.062 58.587
## <none>          171.528 61.097
## + LF   1     6.534 164.994 61.466
## + U2   1     4.976 166.553 61.861
## + X    1     4.041 167.487 62.096
## + Age  1     0.460 171.068 62.984
##
## Step: AIC=36.92
## sqrtR ~ Ex0
```



```

##
##      Df Sum of Sq    RSS    AIC
## + Age   1   18.2318 73.734 29.637
## + X     1   17.4701 74.495 30.069
## <none>                91.965 36.917
## + W     1    2.9694 88.996 37.539
## + LF    1    1.4612 90.504 38.245
## + U2    1    0.3746 91.591 38.746
## + Ed    1    0.1249 91.841 38.860
##
## Step:  AIC=29.64
## sqrtR ~ Ex0 + Age
##
##      Df Sum of Sq    RSS    AIC
## + Age:Ex0  1    9.3180 64.416 25.963
## + X       1    6.8618 66.872 27.535
## + LF      1    5.5632 68.171 28.342
## <none>                73.734 29.637
## + Ed      1    1.8379 71.896 30.577
## + U2      1    1.7858 71.948 30.607
## + W       1    0.0153 73.718 31.628
##
## Step:  AIC=25.96
## sqrtR ~ Ex0 + Age + Ex0:Age
##
##      Df Sum of Sq    RSS    AIC
## + X     1   14.3560 50.060 17.373
## <none>                64.416 25.963
## + U2    1    2.6200 61.796 26.219
## + LF    1    1.6371 62.779 26.882
## + W     1    1.6288 62.787 26.887
## + Ed    1    0.0000 64.416 27.963
##
## Step:  AIC=17.37
## sqrtR ~ Ex0 + Age + X + Ex0:Age
##
##      Df Sum of Sq    RSS    AIC
## + Ed    1    6.6976 43.362 13.341
## + Ex0:X  1    5.4530 44.607 14.529
## + W     1    4.5604 45.499 15.361
## + LF    1    3.1787 46.881 16.618
## <none>                50.060 17.373
## + Age:X  1    2.1347 47.925 17.543
## + U2    1    0.4459 49.614 18.997
##
## Step:  AIC=13.34
## sqrtR ~ Ex0 + Age + X + Ed + Ex0:Age
##
##      Df Sum of Sq    RSS    AIC
## + Ex0:X  1    4.0152 39.347 11.259
## + U2     1    3.4093 39.953 11.901
## + W      1    3.0002 40.362 12.329
## <none>                43.362 13.341
## + Age:Ed  1    1.6672 41.695 13.694

```

```

## + Age:X      1      1.4560 41.906 13.906
## + Ed:Ex0     1      0.4580 42.904 14.895
## + Ed:X       1      0.1166 43.245 15.227
## + LF         1      0.0009 43.361 15.340
##
## Step:  AIC=11.26
## sqrtR ~ Ex0 + Age + X + Ed + Ex0:Age + Ex0:X
##
##           Df Sum of Sq    RSS      AIC
## + U2       1      4.1691 35.178  8.5554
## + W        1      2.4154 36.932 10.5987
## <none>                39.347 11.2594
## + Ed:Ex0   1      0.8015 38.545 12.3950
## + Age:X    1      0.5776 38.769 12.6383
## + Ed:X     1      0.4897 38.857 12.7334
## + Age:Ed   1      0.3818 38.965 12.8499
## + LF       1      0.0031 39.344 13.2562
##
## Step:  AIC=8.56
## sqrtR ~ Ex0 + Age + X + Ed + U2 + Ex0:Age + Ex0:X
##
##           Df Sum of Sq    RSS      AIC
## + W        1      1.87041 33.307  8.2607
## <none>                35.178  8.5554
## + Age:U2   1      1.41270 33.765  8.8339
## + Ex0:U2   1      1.34663 33.831  8.9160
## + Ed:U2    1      1.29731 33.881  8.9772
## + LF       1      0.92422 34.254  9.4372
## + U2:X     1      0.81333 34.365  9.5729
## + Age:X    1      0.75591 34.422  9.6430
## + Age:Ed   1      0.68607 34.492  9.7282
## + Ed:Ex0   1      0.05888 35.119 10.4850
## + Ed:X     1      0.02635 35.152 10.5239
##
## Step:  AIC=8.26
## sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X
##
##           Df Sum of Sq    RSS      AIC
## + Ex0:U2   1      2.05974 31.248  7.5796
## + U2:W      1      1.71174 31.596  8.0448
## + Age:X     1      1.64692 31.661  8.1308
## + Ed:U2     1      1.58433 31.723  8.2138
## <none>                33.307  8.2607
## + Ex0:W    1      1.38903 31.918  8.4716
## + U2:X     1      1.20201 32.105  8.7169
## + W:X       1      1.01862 32.289  8.9562
## + Age:U2   1      0.89751 32.410  9.1134
## + LF       1      0.75686 32.551  9.2953
## + Ed:W     1      0.70053 32.607  9.3679
## + Age:Ed   1      0.47682 32.831  9.6551
## + Age:W    1      0.12916 33.178 10.0975
## + Ed:Ex0   1      0.11501 33.192 10.1154
## + Ed:X     1      0.03456 33.273 10.2171
##

```

```

## Step: AIC=7.58
## sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X + Ex0:U2
##
##      Df Sum of Sq  RSS    AIC
## + Age:X    1   1.86470 29.383 6.9954
## <none>                31.248 7.5796
## + Age:Ed    1   0.94853 30.299 8.2849
## + W:X       1   0.78515 30.463 8.5108
## + Ed:W      1   0.65255 30.595 8.6932
## + Ex0:W     1   0.56941 30.678 8.8072
## + Ed:U2     1   0.50520 30.742 8.8950
## + Ed:Ex0    1   0.28849 30.959 9.1900
## + Ed:X      1   0.22779 31.020 9.2723
## + U2:X      1   0.09972 31.148 9.4454
## + Age:U2    1   0.08277 31.165 9.4682
## + U2:W      1   0.06848 31.179 9.4875
## + LF       1   0.05713 31.191 9.5027
## + Age:W     1   0.03796 31.210 9.5285
##
## Step: AIC=7
## sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X + Ex0:U2 +
##      Age:X
##
##      Df Sum of Sq  RSS    AIC
## + Age:W     1   4.4407 24.942 2.1136
## + W:X       1   3.3330 26.050 3.9387
## + Ed:U2     1   1.6826 27.700 6.5186
## <none>                29.383 6.9954
## + Ed:W      1   1.1100 28.273 7.3780
## + U2:X      1   0.5365 28.846 8.2213
## + Ed:Ex0    1   0.3001 29.083 8.5642
## + Ed:X      1   0.2907 29.092 8.5777
## + Ex0:W     1   0.2626 29.120 8.6183
## + U2:W      1   0.2543 29.129 8.6303
## + Age:U2    1   0.1411 29.242 8.7931
## + Age:Ed    1   0.0169 29.366 8.9712
## + LF       1   0.0137 29.369 8.9757
##
## Step: AIC=2.11
## sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X + Ex0:U2 +
##      Age:X + Age:W
##
##      Df Sum of Sq  RSS    AIC
## <none>                24.942 2.1136
## + Ed:U2     1   1.01459 23.928 2.3695
## + Age:U2    1   0.81294 24.129 2.7219
## + W:X       1   0.57025 24.372 3.1422
## + Age:Ed    1   0.54535 24.397 3.1851
## + U2:X      1   0.48820 24.454 3.2834
## + Ed:Ex0    1   0.21149 24.731 3.7560
## + LF       1   0.14758 24.795 3.8644
## + U2:W      1   0.10998 24.832 3.9280
## + Ed:W      1   0.07505 24.867 3.9871
## + Ex0:W     1   0.07101 24.871 3.9939

```

```
## + Ed:X      1    0.00041 24.942 4.1129
```

```
summary(step_for)
```

```
##
## Call:
## lm(formula = sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age +
##      Ex0:X + Ex0:U2 + Age:X + Age:W, data = crime2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72634 -0.53748  0.07713  0.41759  1.52209
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.478e+02  5.387e+01  -2.744  0.01014 *
## Ex0          -1.594e-01  1.203e-01  -1.325  0.19512
## Age           1.010e+00  3.830e-01   2.638  0.01309 *
## X             3.679e-01  1.321e-01   2.786  0.00917 **
## Ed            6.252e-02  2.458e-02   2.544  0.01636 *
## U2           -4.849e-02  6.618e-02  -0.733  0.46941
## W            1.441e-01  5.835e-02   2.469  0.01946 *
## Ex0:Age       7.726e-04  8.904e-04   0.868  0.39244
## Ex0:X         4.371e-04  2.093e-04   2.088  0.04535 *
## Ex0:U2        1.006e-03  7.330e-04   1.372  0.18018
## Age:X        -2.567e-03  9.351e-04  -2.746  0.01011 *
## Age:W        -9.760e-04  4.223e-04  -2.311  0.02788 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9118 on 30 degrees of freedom
## Multiple R-squared:  0.8546, Adjusted R-squared:  0.8013
## F-statistic: 16.03 on 11 and 30 DF, p-value: 1.123e-09
```

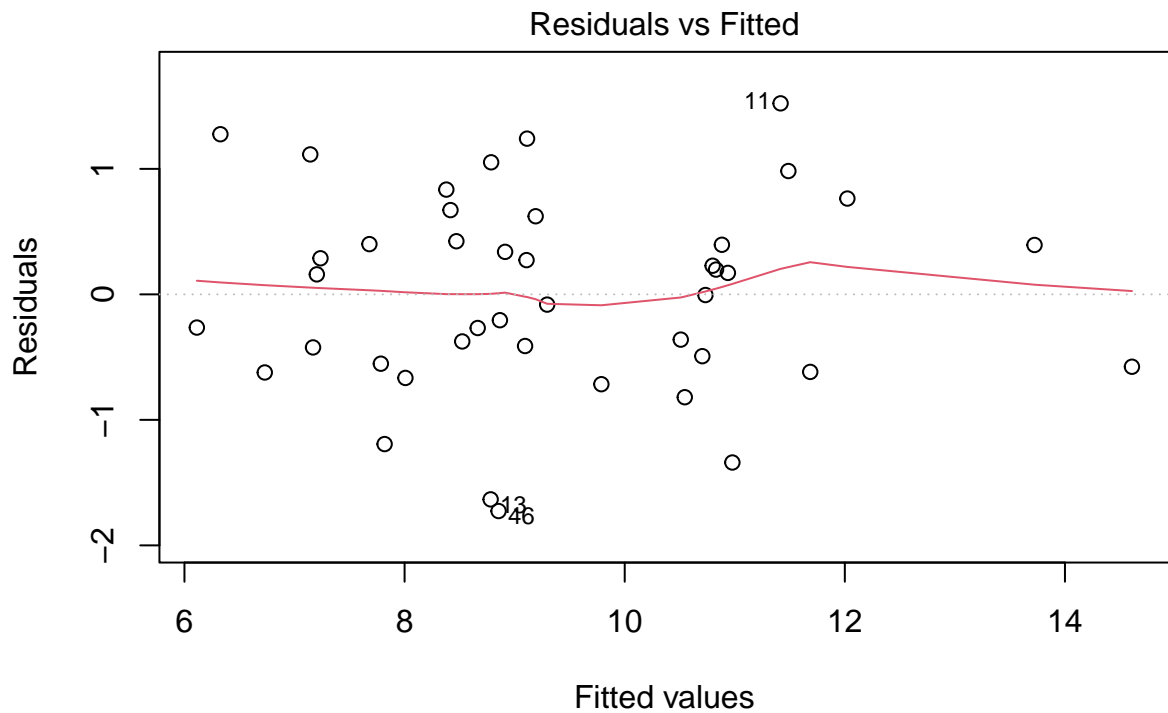
```
#Adjusted R-squared: 0.8013
```

```
model17 = lm(sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age +
              Ex0:X + Ex0:U2 + Age:X + Age:W, data = crime2)
summary(model17)
```

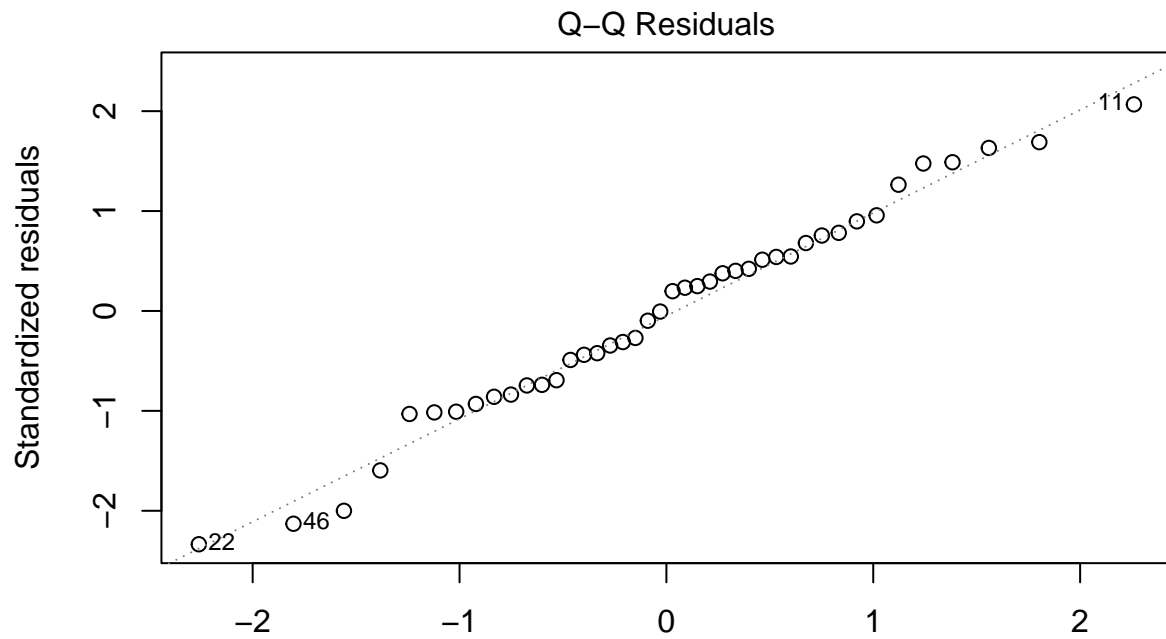
```
##
## Call:
## lm(formula = sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age +
##      Ex0:X + Ex0:U2 + Age:X + Age:W, data = crime2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72634 -0.53748  0.07713  0.41759  1.52209
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.478e+02  5.387e+01  -2.744  0.01014 *
## Ex0          -1.594e-01  1.203e-01  -1.325  0.19512
## Age           1.010e+00  3.830e-01   2.638  0.01309 *
```

```
## X          3.679e-01  1.321e-01  2.786  0.00917 **
## Ed         6.252e-02  2.458e-02  2.544  0.01636 *
## U2        -4.849e-02  6.618e-02 -0.733  0.46941
## W          1.441e-01  5.835e-02  2.469  0.01946 *
## Ex0:Age    7.726e-04  8.904e-04  0.868  0.39244
## Ex0:X      4.371e-04  2.093e-04  2.088  0.04535 *
## Ex0:U2     1.006e-03  7.330e-04  1.372  0.18018
## Age:X     -2.567e-03  9.351e-04 -2.746  0.01011 *
## Age:W     -9.760e-04  4.223e-04 -2.311  0.02788 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9118 on 30 degrees of freedom
## Multiple R-squared:  0.8546, Adjusted R-squared:  0.8013
## F-statistic: 16.03 on 11 and 30 DF,  p-value: 1.123e-09
```

```
plot(model17)
```

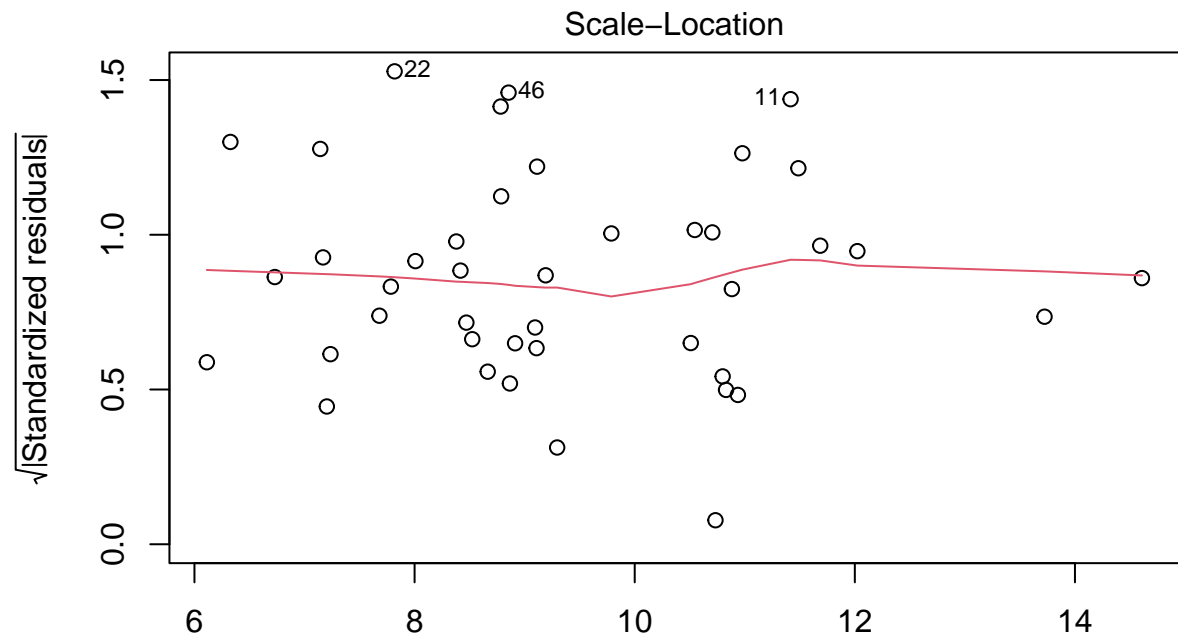


$\text{lm}(\text{sqrtR} \sim \text{Ex0} + \text{Age} + \text{X} + \text{Ed} + \text{U2} + \text{W} + \text{Ex0:Age} + \text{Ex0:X} + \text{Ex0:U2} + \text{Age:X} + \dots)$

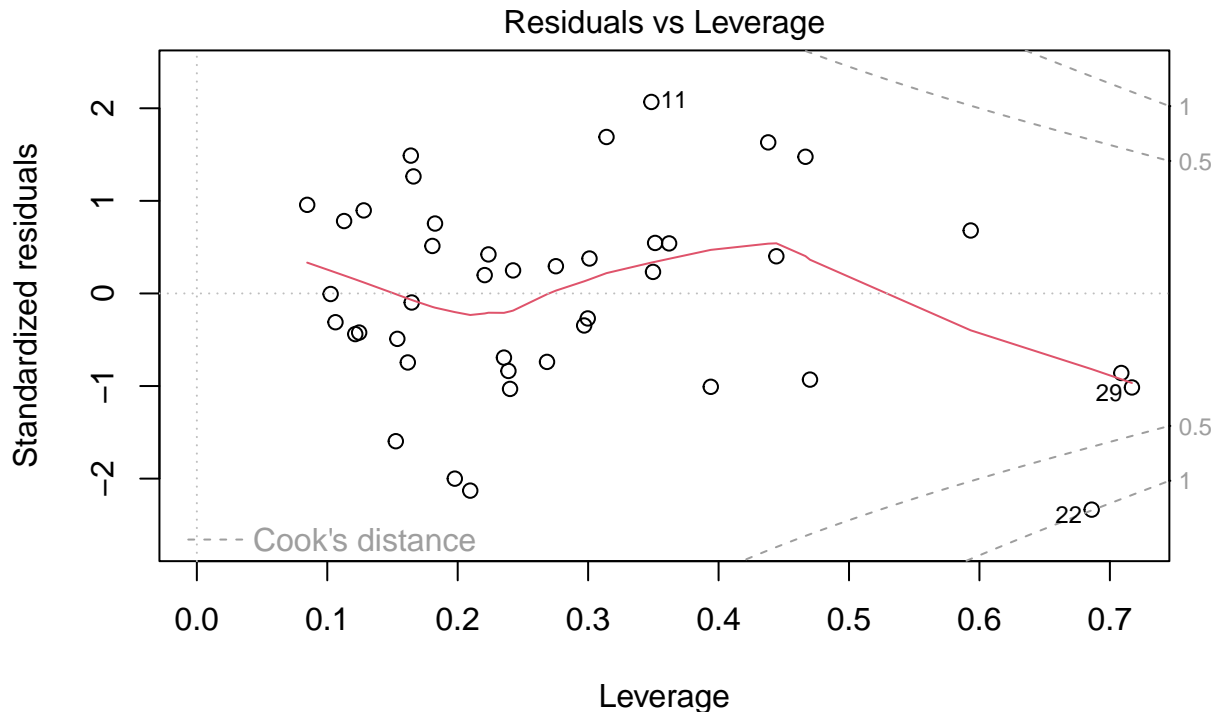


Theoretical Quantiles

lm(sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X + Ex0:U2 + Age:X + ..



Fitted values
 $\text{lm}(\text{sqrtR} \sim \text{Ex0} + \text{Age} + \text{X} + \text{Ed} + \text{U2} + \text{W} + \text{Ex0}:\text{Age} + \text{Ex0}:\text{X} + \text{Ex0}:\text{U2} + \text{Age}:\text{X} + \dots)$



lm(sqrtR ~ Ex0 + Age + X + Ed + U2 + W + Ex0:Age + Ex0:X + Ex0:U2 + Age:X + ..

```
#one way anova
```

```
aov1 <- aov(R ~ S, data = crime)
```

```
summary(aov1)
```

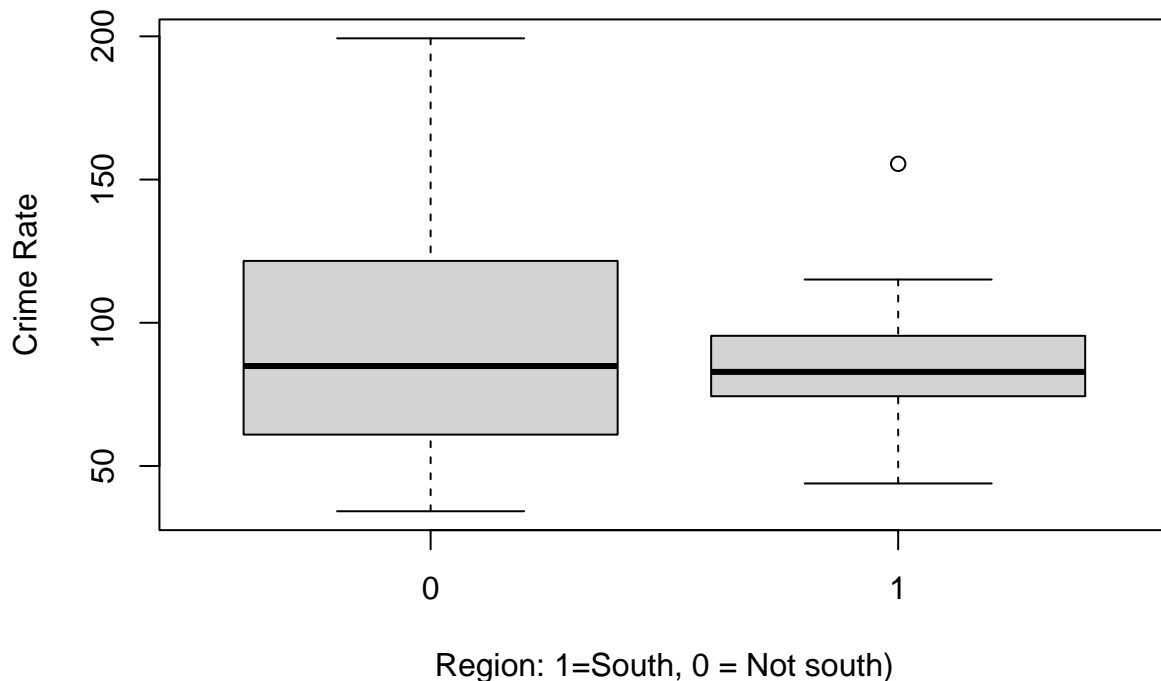
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## S           1    565    565.3    0.373  0.545
## Residuals   45  68244   1516.5
```

```
#H0: mean crime rate is the same for south and non south states, Ha: mean crime rate is not the same
```

```
#since p value =0.545 > 0.05 we can not reject H0 at 0.05 signifiance level
```

```
boxplot(R~S, crime, main = "Crime Rate by Region", xlab = "Region: 1=South, 0 = Not south)", ylab = "C
```


Crime Rate by Region



```
#transform rate into binary to use logistic regression
#create y binary variable "High rate" if R is > median
crime$High <- ifelse(crime$R > median(crime$R, na.rm = TRUE), 1, 0)
logit_model <- glm(High ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N + NW + U1 + U2 + W + X,
                   data = crime,
                   family = binomial(link = "logit"))
summary(logit_model)
```

```
##
## Call:
## glm(formula = High ~ Age + S + Ed + Ex0 + Ex1 + LF + M + N +
##      NW + U1 + U2 + W + X, family = binomial(link = "logit"),
##      data = crime)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.255e+02  7.203e+01  -1.742   0.0815 .
## Age          2.830e-01  1.760e-01   1.608   0.1078
## S            5.356e+00  4.015e+00   1.334   0.1822
## Ed           1.568e-01  1.692e-01   0.927   0.3541
## Ex0          3.476e-01  3.274e-01   1.062   0.2884
## Ex1         -1.614e-01  2.600e-01  -0.621   0.5347
## LF           1.082e-01  6.818e-02   1.588   0.1124
## M           -9.175e-02  7.924e-02  -1.158   0.2469
## N           -6.487e-02  4.839e-02  -1.340   0.1801
## NW          -8.466e-04  1.103e-02  -0.077   0.9388
```

```
## U1          1.606e-01  1.330e-01  1.207  0.2274
## U2          3.623e-02  1.671e-01  0.217  0.8283
## W           6.893e-02  4.982e-02  1.383  0.1665
## X           1.565e-01  1.163e-01  1.346  0.1783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 58.129  on 41  degrees of freedom
## Residual deviance: 26.173  on 28  degrees of freedom
## (5 observations deleted due to missingness)
## AIC: 54.173
##
## Number of Fisher Scoring iterations: 8
```

```
#null deviance 58.129 on 41df, residual deviance 26.173 on 28df
pvalue = 1-pchisq(logit_model$deviance, logit_model$df.residual)
pvalue
```

```
## [1] 0.5635497
```

```
#since p value = 0.56 > 0.05 we can say there is no significant evidence that the model does not fit we
exp(coef(logit_model))
```

```
## (Intercept)          Age          S          Ed          Ex0          Ex1
## 3.235077e-55 1.327141e+00 2.119640e+02 1.169744e+00 1.415626e+00 8.509430e-01
##          LF          M          N          NW          U1          U2
## 1.114304e+00 9.123373e-01 9.371917e-01 9.991537e-01 1.174161e+00 1.036898e+00
##          W          X
## 1.071361e+00 1.169435e+00
```

```
#interpretation of coefficients is for a unit increase in xi, there is a (1-Bi)*100% increase in R hold
```