

# P8130 Final Project

## Abstract

## Introduction (brief context and background of the problem)

## Methods (data description and statistical methods)

## Results

## Conclusions/Discussion

```
library(tidyverse)
library(ggplot2)
library(GGally)
library(PerformanceAnalytics)
library(performance)
library(MASS)
```

## Read in dataset

```
cdi = read_csv("./cdi.csv") %>%
  janitor::clean_names()

## no missing value
cdi %>%
  dplyr::select(everything()) %>%
  summarise_all(funs(sum(is.na(.)))) %>%
  knitr::kable()
```

id	cty	state	area	pop	pop18	pop65	docs	beds	crimes	hsgrad	bagrad	poverty	unempp	pcincom	totalinc	region
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

## Data cleaning

```
# some normalization for better comparison
cdi =
  cdi %>%
  mutate(crm_1000 = crimes/pop*1000, # as indicated by the project prompt
         docs_1000 = docs/pop*1000, # every 1000 people how many doctors
         beds_1000 = beds/pop*1000, # ratio of beds per doctor
         pop_density = pop/area, # how many people per square miles
         region = factor(region)) %>%
  dplyr::select(-id, -crimes, -area, -docs, -beds, -totalinc, -pop)
```

## Data Exploration

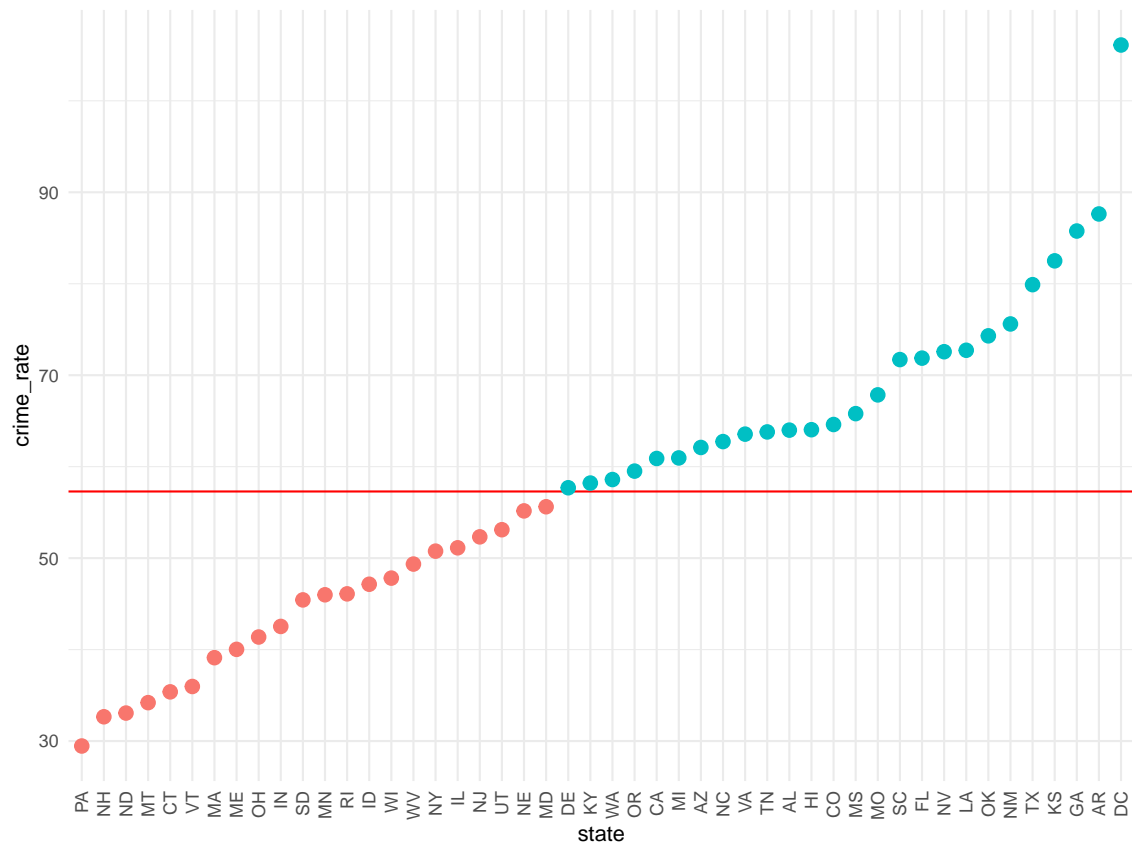
```
## summary statistics, tentative, NOT FINAL
```

```
sum_cdi =  
  cdi %>%  
  dplyr::select(-c(cty, state))  
summary(sum_cdi)
```

```
##      pop18      pop65      hsgrad      bagrad  
## Min.   :16.40  Min.   : 3.000  Min.   :46.60  Min.   : 8.10  
## 1st Qu.:26.20  1st Qu.: 9.875  1st Qu.:73.88  1st Qu.:15.28  
## Median :28.10  Median :11.750  Median :77.70  Median :19.70  
## Mean   :28.57  Mean   :12.170  Mean   :77.56  Mean   :21.08  
## 3rd Qu.:30.02  3rd Qu.:13.625  3rd Qu.:82.40  3rd Qu.:25.32  
## Max.   :49.70  Max.   :33.800  Max.   :92.90  Max.   :52.30  
##      poverty      unemp      pcincome      region      crm_1000  
## Min.   : 1.400  Min.   : 2.200  Min.   : 8899  1:103  Min.   : 4.601  
## 1st Qu.: 5.300  1st Qu.: 5.100  1st Qu.:16118  2:108  1st Qu.: 38.102  
## Median : 7.900  Median : 6.200  Median :17759  3:152  Median : 52.429  
## Mean   : 8.721  Mean   : 6.597  Mean   :18561  4: 77  Mean   : 57.286  
## 3rd Qu.:10.900  3rd Qu.: 7.500  3rd Qu.:20270      3rd Qu.: 72.597  
## Max.   :36.300  Max.   :21.300  Max.   :37541      Max.   :295.987  
##      docs_1000      beds_1000      pop_density  
## Min.   : 0.3559  Min.   : 0.1649  Min.   : 13.26  
## 1st Qu.: 1.2127  1st Qu.: 2.1972  1st Qu.: 192.34  
## Median : 1.7509  Median : 3.3287  Median : 335.91  
## Mean   : 2.1230  Mean   : 3.6493  Mean   : 888.44  
## 3rd Qu.: 2.4915  3rd Qu.: 4.5649  3rd Qu.: 756.55  
## Max.   :17.0377  Max.   :19.6982  Max.   :32403.72
```

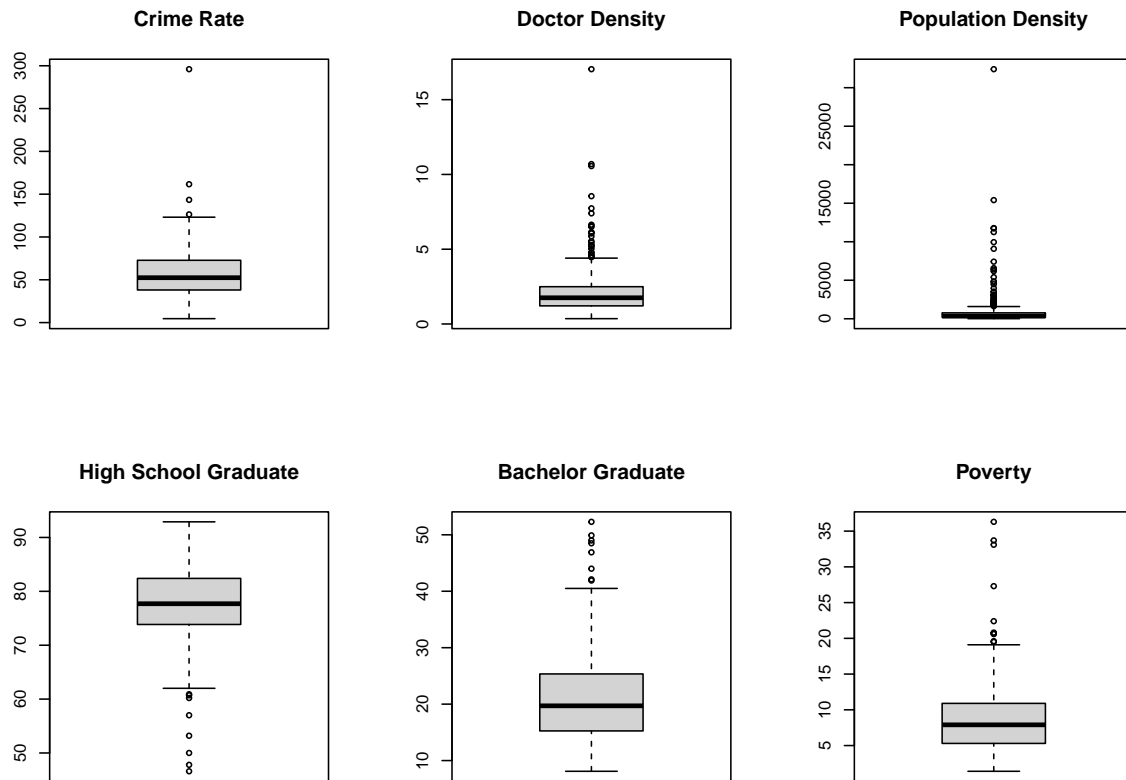
```
mean_crm = mean(sum_cdi$crm_1000)  
cdi_state = cdi %>%  
  group_by(state) %>%  
  summarize(crime_rate = mean(crm_1000)) %>%  
  mutate(low_high = ifelse(crime_rate>mean_crm, TRUE,FALSE))
```

```
cdi_state %>%  
  mutate(state = fct_reorder(state, crime_rate)) %>%  
  ggplot(aes(x = state, y = crime_rate))+  
  geom_hline(yintercept = mean_crm, color = "red")+  
  geom_point(aes(color = low_high),size = 3)+  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust= 1),  
        legend.position = "none")
```

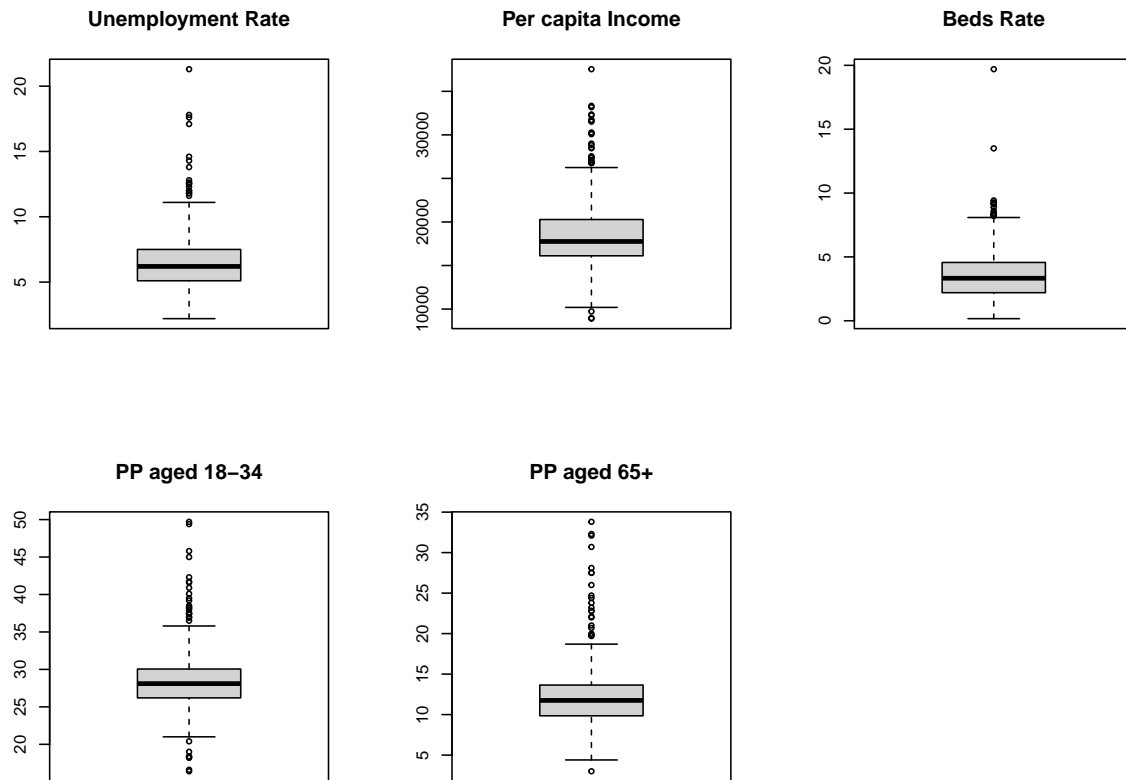


boxplot for each variable

```
par(mfrow=c(2,3))
boxplot(sum_cdi$crm_1000, main='Crime Rate')
boxplot(sum_cdi$docs_1000, main='Doctor Density')
boxplot(sum_cdi$pop_density, main='Population Density' )
boxplot(sum_cdi$hsgrad, main='High School Graduate')
boxplot(sum_cdi$bagrad, main='Bachelor Graduate')
boxplot(sum_cdi$poverty, main='Poverty')
```

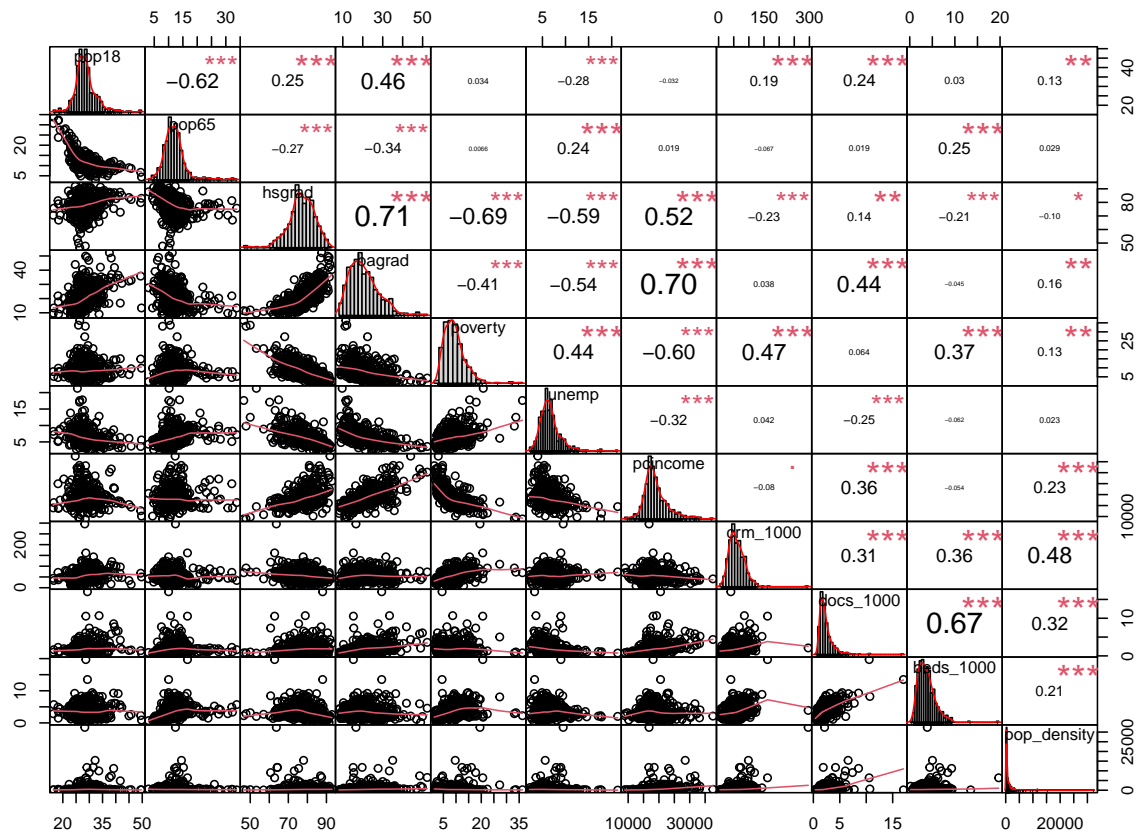


```
par(mfrow=c(2,3))
boxplot(sum_cdi$unemp, main='Unemployment Rate')
boxplot(sum_cdi$pcincome, main='Per capita Income')
boxplot(sum_cdi$beds_1000, main='Beds Rate')
boxplot(sum_cdi$pop18, main='PP aged 18-34')
boxplot(sum_cdi$pop65, main='PP aged 65+')
```



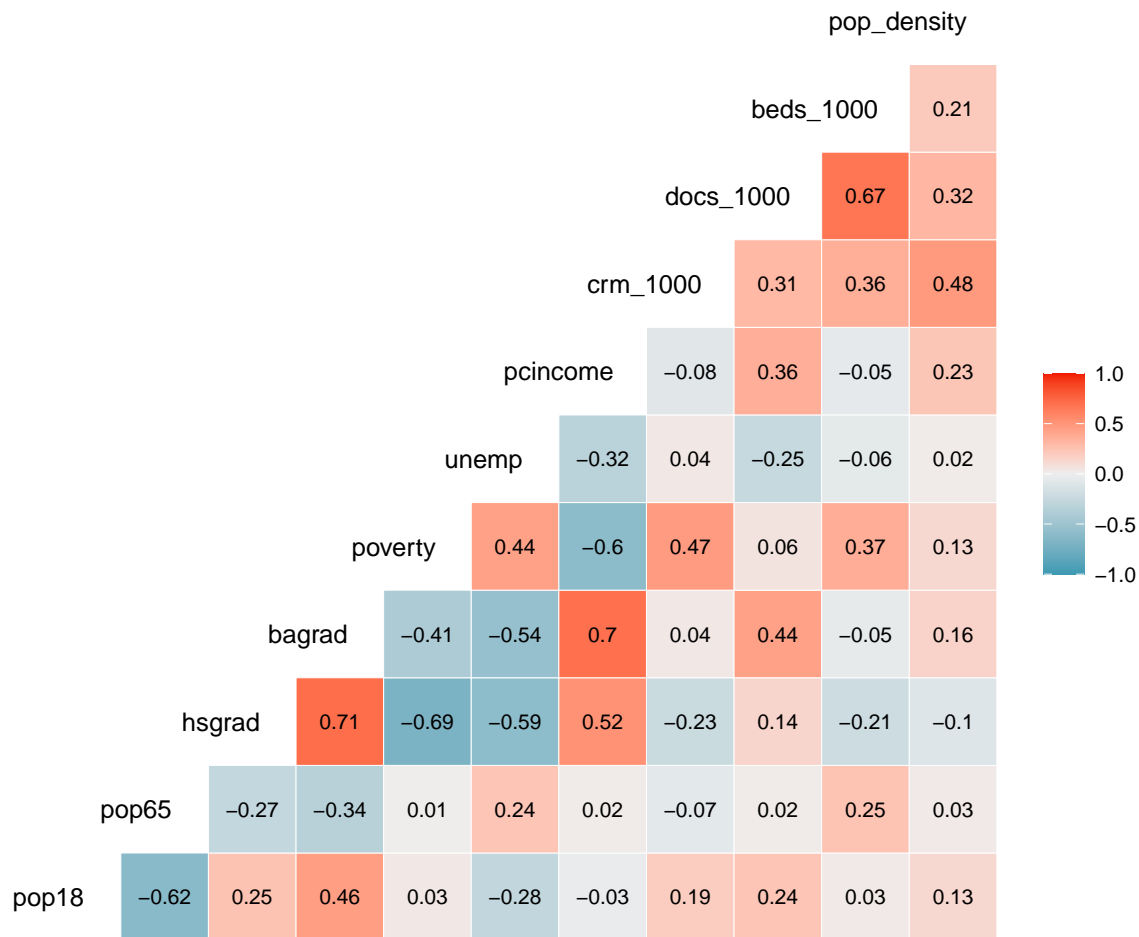
## Marginal Correlation and Correlation martix

```
corr_matrix =
  cdi %>%
  dplyr::select(-state, -region, -cty) %>%
  #sum_cdi %>%
  chart.Correlation(histogram = TRUE, method = "pearson")
```



Correlation Heatmap

```
cdi %>%
  dplyr::select(-state, -cty) %>%
  # sum_cdi %>%
  ggcorr(label=TRUE, hjust = 0.9, layout.exp = 2, label_size = 3, label_round = 2)
```



```
#corrplot(cor(cdi_1), type = "upper", diag = FALSE)
```

## Build Model

### Backward Elimination

```
mult_fit = lm(crm_1000 ~ ., data = sum_cdi)
summary(mult_fit)
```

```
##
## Call:
## lm(formula = crm_1000 ~ ., data = sum_cdi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.755 -11.637  -0.547  10.682  74.653
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.326e+01  2.772e+01  -2.643 0.008515 **
## pop18        8.338e-01  3.323e-01   2.509 0.012480 *
## pop65       -1.637e-01  3.092e-01  -0.529 0.596875
## hsgrad       5.573e-01  2.719e-01   2.049 0.041048 *
## bagrad      -5.565e-01  3.001e-01  -1.854 0.064402 .
```

```
## poverty      1.987e+00  3.895e-01  5.102 5.08e-07 ***
## unemp        5.221e-01  5.376e-01  0.971 0.332059
## pcincome     1.409e-03  4.673e-04  3.015 0.002722 **
## region2      9.607e+00  2.761e+00  3.479 0.000555 ***
## region3      2.795e+01  2.693e+00  10.379 < 2e-16 ***
## region4      2.299e+01  3.121e+00  7.365 9.24e-13 ***
## docs_1000    -5.658e-01  1.032e+00 -0.548 0.583889
## beds_1000    2.970e+00  8.023e-01  3.701 0.000243 ***
## pop_density  5.217e-03  4.502e-04  11.588 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.03 on 426 degrees of freedom
## Multiple R-squared:  0.5774, Adjusted R-squared:  0.5645
## F-statistic: 44.77 on 13 and 426 DF,  p-value: < 2.2e-16
multi_back = step(mult_fit, direction='backward')
```

```
## Start:  AIC=2559
## crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty + unemp +
##      pcincome + region + docs_1000 + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - pop65      1         91 138649 2557.3
## - docs_1000   1         98 138656 2557.3
## - unemp       1        307 138865 2558.0
## <none>                138558 2559.0
## - bagrad      1        1118 139677 2560.5
## - hsgrad      1        1366 139924 2561.3
## - pop18       1        2047 140606 2563.4
## - pcincome    1        2957 141515 2566.3
## - beds_1000   1        4456 143014 2570.9
## - poverty     1        8466 147024 2583.1
## - region      3       40239 178797 2665.2
## - pop_density 1       43675 182234 2677.6
##
## Step:  AIC=2557.29
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + unemp + pcincome +
##      region + docs_1000 + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - docs_1000   1         94 138744 2555.6
## - unemp       1        271 138920 2556.2
## <none>                138649 2557.3
## - bagrad      1        1119 139768 2558.8
## - hsgrad      1        1415 140065 2559.8
## - pcincome    1        3029 141679 2564.8
## - pop18       1        3331 141980 2565.7
## - beds_1000   1        4461 143110 2569.2
## - poverty     1        9179 147829 2583.5
## - region      3       40288 178937 2663.5
## - pop_density 1       43627 182277 2675.7
##
## Step:  AIC=2555.59
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + unemp + pcincome +
```



```
##      region + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - unemp      1      266 139010 2554.4
## <none>                138744 2555.6
## - bagrad      1     1367 140111 2557.9
## - hsgrad      1     1487 140230 2558.3
## - pcincome    1     2939 141683 2562.8
## - pop18       1     3247 141991 2563.8
## - beds_1000   1     7528 146271 2576.8
## - poverty     1     9264 148008 2582.0
## - region      3    40196 178940 2661.5
## - pop_density 1    43700 182444 2674.1
##
## Step:  AIC=2554.43
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
##      beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## <none>                139010 2554.4
## - hsgrad      1     1308 140317 2556.6
## - bagrad      1     1685 140695 2557.7
## - pop18       1     3230 142239 2562.5
## - pcincome    1     3349 142358 2562.9
## - beds_1000   1     7339 146349 2575.1
## - poverty     1    11263 150273 2586.7
## - region      3    41638 180647 2663.7
## - pop_density 1    43482 182491 2672.2
```

## Forward Selection

```
multi_forward = step(mult_fit, direction = 'forward')
```

```
## Start:  AIC=2559
## crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty + unemp +
##      pcincome + region + docs_1000 + beds_1000 + pop_density
```

## Both direction

```
multi_both = step(mult_fit, direction = "both")
```

```
## Start:  AIC=2559
## crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty + unemp +
##      pcincome + region + docs_1000 + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - pop65      1      91 138649 2557.3
## - docs_1000   1      98 138656 2557.3
## - unemp       1     307 138865 2558.0
## <none>                138558 2559.0
## - bagrad      1     1118 139677 2560.5
## - hsgrad      1     1366 139924 2561.3
## - pop18       1     2047 140606 2563.4
## - pcincome    1     2957 141515 2566.3
```

```

## - beds_1000      1      4456 143014 2570.9
## - poverty        1      8466 147024 2583.1
## - region         3     40239 178797 2665.2
## - pop_density    1     43675 182234 2677.6
##
## Step: AIC=2557.29
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + unemp + pcincome +
##      region + docs_1000 + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - docs_1000      1         94 138744 2555.6
## - unemp          1        271 138920 2556.2
## <none>                        138649 2557.3
## - bagrad         1        1119 139768 2558.8
## + pop65          1         91 138558 2559.0
## - hsgrad         1        1415 140065 2559.8
## - pcincome       1        3029 141679 2564.8
## - pop18          1        3331 141980 2565.7
## - beds_1000      1        4461 143110 2569.2
## - poverty        1        9179 147829 2583.5
## - region         3       40288 178937 2663.5
## - pop_density    1       43627 182277 2675.7
##
## Step: AIC=2555.59
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + unemp + pcincome +
##      region + beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## - unemp          1         266 139010 2554.4
## <none>                        138744 2555.6
## + docs_1000      1         94 138649 2557.3
## + pop65          1         88 138656 2557.3
## - bagrad         1        1367 140111 2557.9
## - hsgrad         1        1487 140230 2558.3
## - pcincome       1        2939 141683 2562.8
## - pop18          1        3247 141991 2563.8
## - beds_1000      1        7528 146271 2576.8
## - poverty        1        9264 148008 2582.0
## - region         3       40196 178940 2661.5
## - pop_density    1       43700 182444 2674.1
##
## Step: AIC=2554.43
## crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
##      beds_1000 + pop_density
##
##           Df Sum of Sq    RSS    AIC
## <none>                        139010 2554.4
## + unemp          1         266 138744 2555.6
## + docs_1000      1         89 138920 2556.2
## + pop65          1         53 138957 2556.3
## - hsgrad         1        1308 140317 2556.6
## - bagrad         1        1685 140695 2557.7
## - pop18          1        3230 142239 2562.5
## - pcincome       1        3349 142358 2562.9

```

```
## - beds_1000      1      7339 146349 2575.1
## - poverty        1      11263 150273 2586.7
## - region         3      41638 180647 2663.7
## - pop_density    1      43482 182491 2672.2
```

## Residuals vs. Fitted && QQ Plots

### Check Multicollinearity

```
check_collinearity(multi_forward)
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
##      Term  VIF Increased SE Tolerance
##      pop18 2.57      1.60      0.39
##      pop65 2.06      1.43      0.49
##      hsgrad 3.32      1.82      0.30
##      bagrad 3.70      1.92      0.27
##      poverty 2.52      1.59      0.40
##      unemp 1.89      1.38      0.53
##      pcincome 1.02      1.01      0.98
##      region 1.99      1.41      0.50
##      docs_1000 2.75      1.66      0.36
##      beds_1000 3.30      1.82      0.30
##      pop_density 1.01      1.00      0.99
```

```
check_collinearity(multi_back)
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
##      Term  VIF Increased SE Tolerance
##      pop18 1.90      1.38      0.53
##      hsgrad 3.17      1.78      0.32
##      bagrad 3.34      1.83      0.30
##      poverty 2.20      1.48      0.45
##      pcincome 1.03      1.01      0.97
##      region 1.62      1.27      0.62
##      beds_1000 1.32      1.15      0.76
##      pop_density 1.01      1.00      0.99
```

```
check_collinearity(multi_both)
```

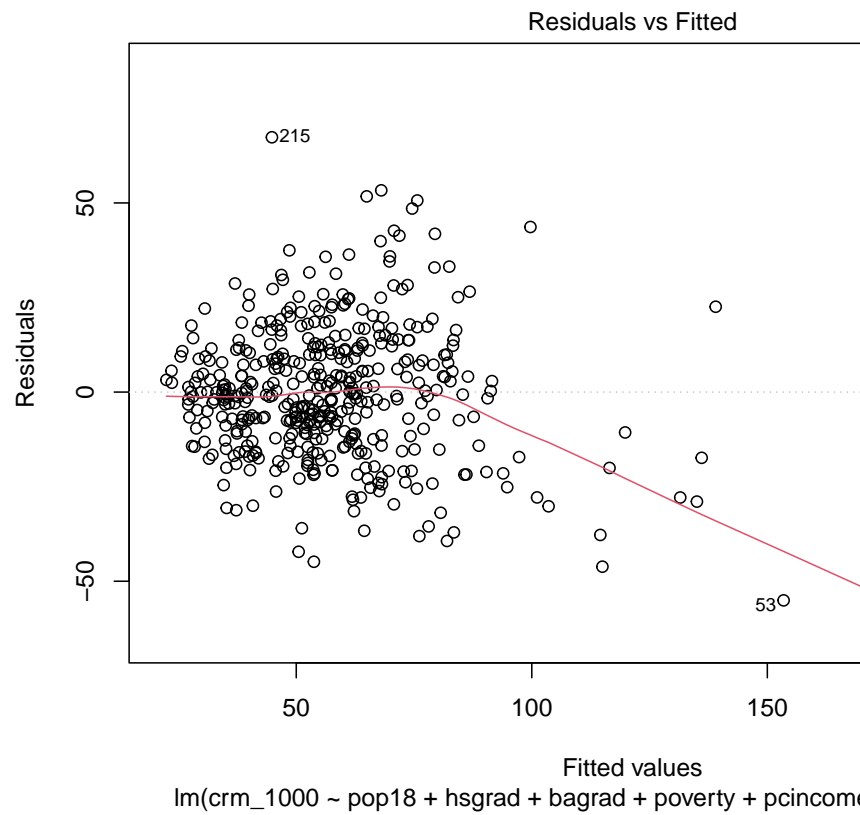
```
## # Check for Multicollinearity
##
## Low Correlation
##
##      Term  VIF Increased SE Tolerance
##      pop18 1.90      1.38      0.53
##      hsgrad 3.17      1.78      0.32
##      bagrad 3.34      1.83      0.30
##      poverty 2.20      1.48      0.45
##      pcincome 1.03      1.01      0.97
```

```
##      region 1.62      1.27      0.62
##    beds_1000 1.32      1.15      0.76
##   pop_density 1.01      1.00      0.99
```

## Model Diagnostics

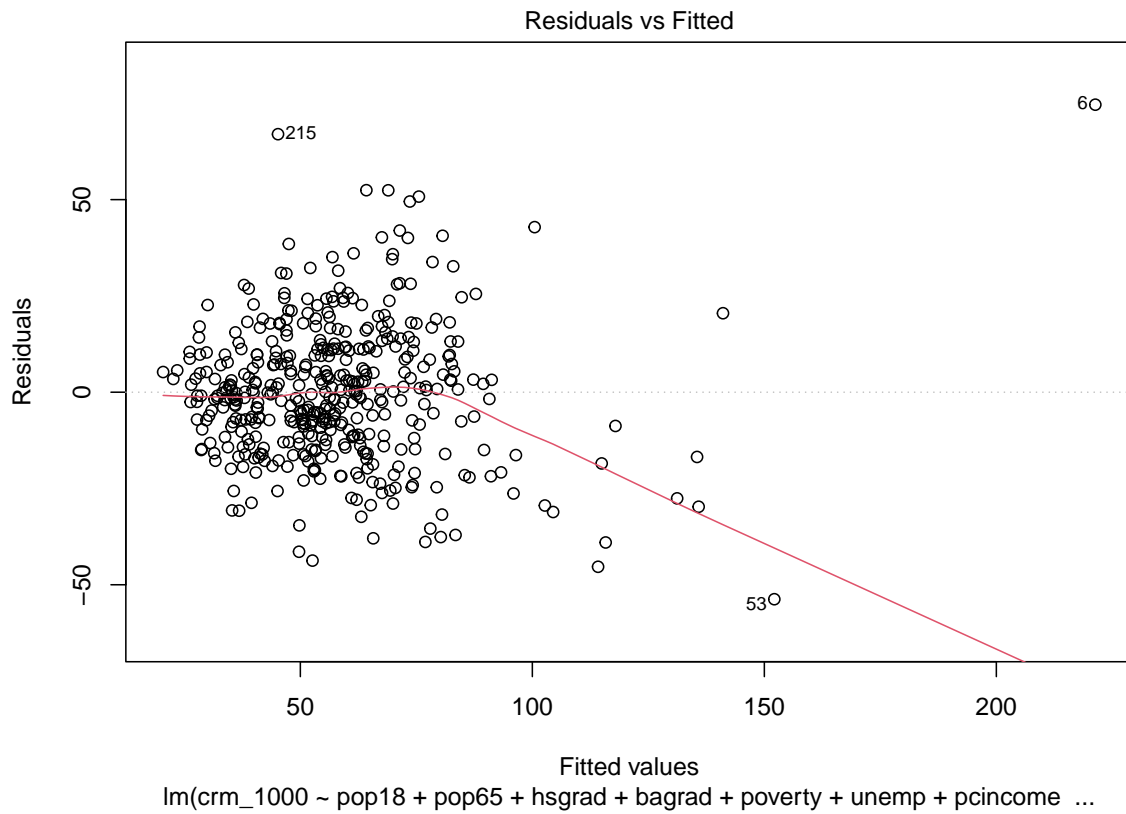
### Residuals VS Fitted/Predicted Values

```
plot(multi_back, which = 1)
```

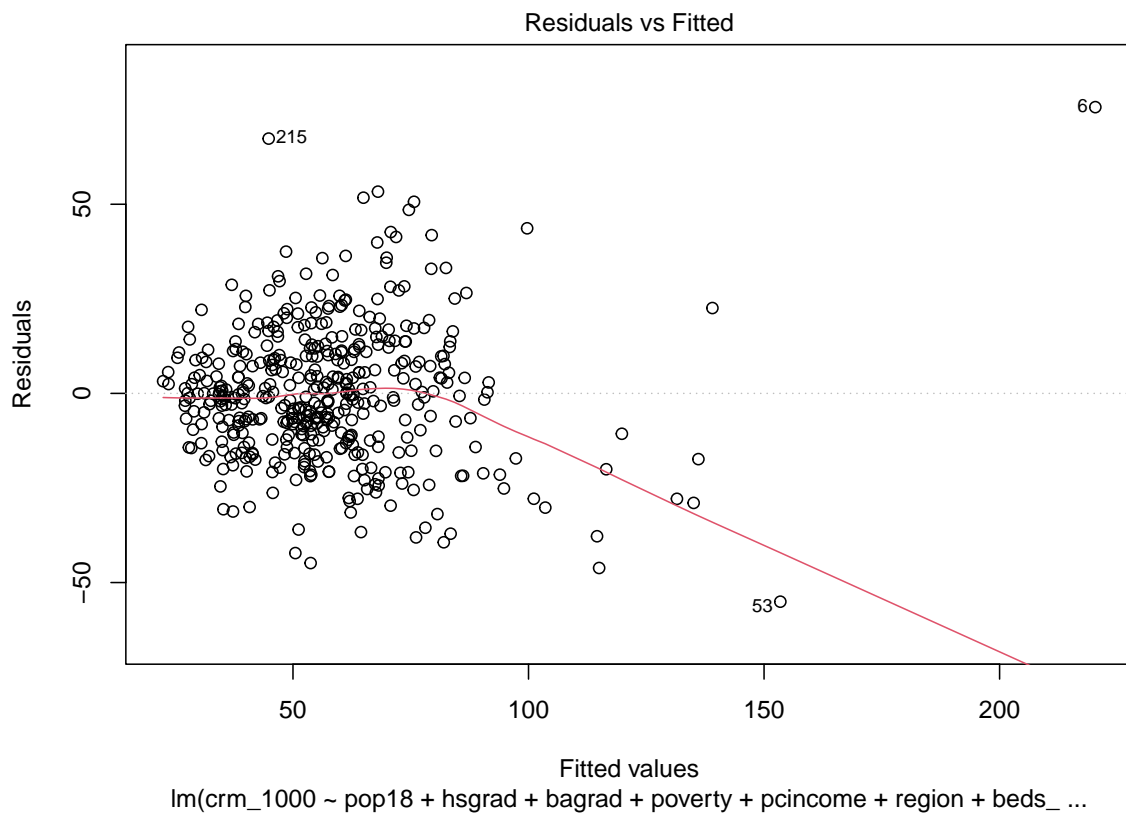


To detect unequal error varince and Outliers

```
plot(multi_forward, which = 1)
```



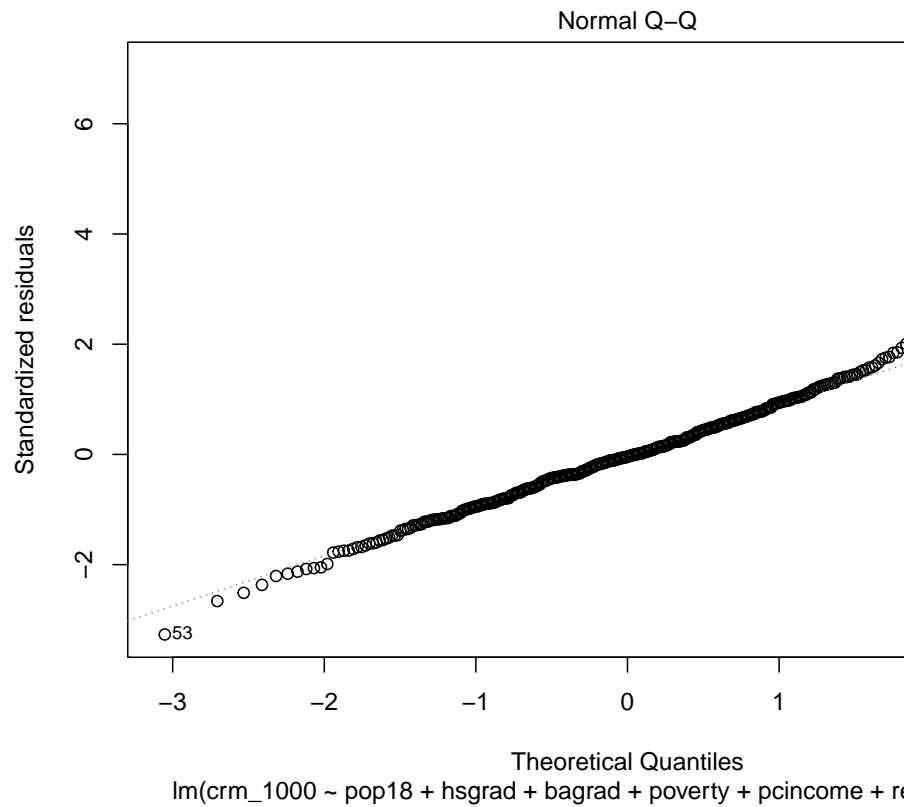
```
plot(multi_both, which = 1)
```



```
#library(olsrr)
#olsrr::ols_plot_resid_fit(multi_forward)
```

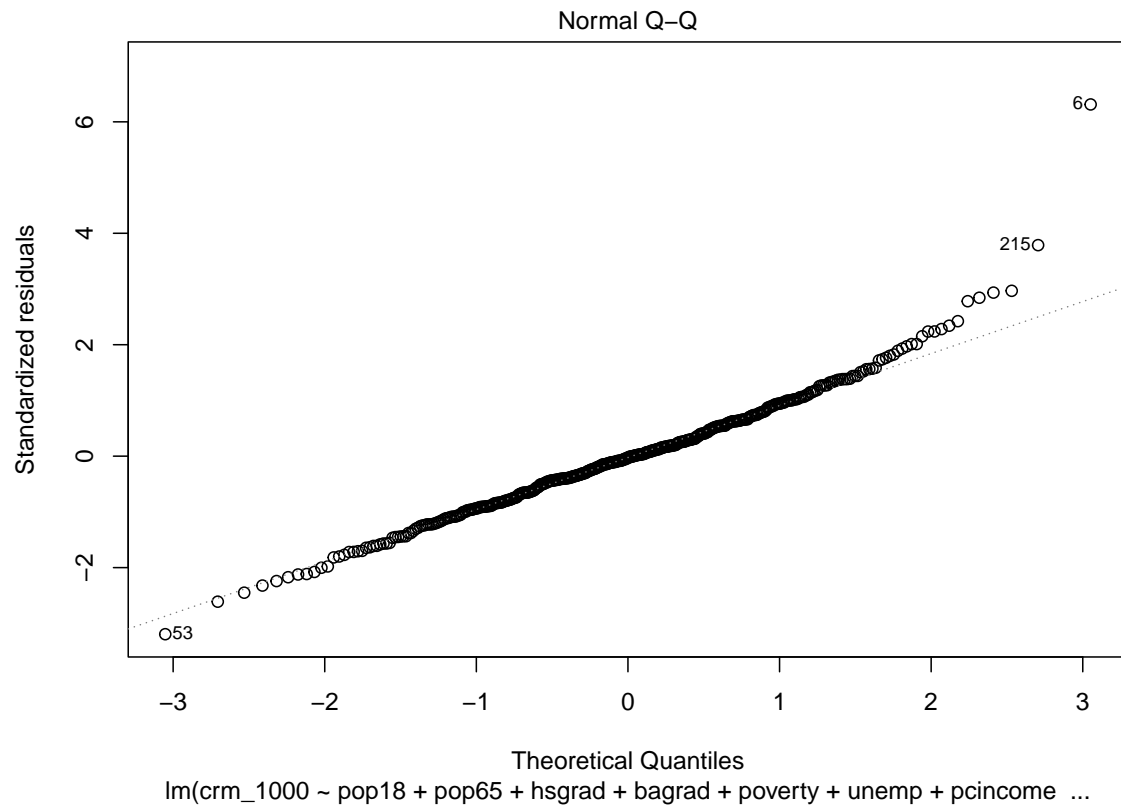
QQ plot

```
plot(multi_back, which = 2)
```

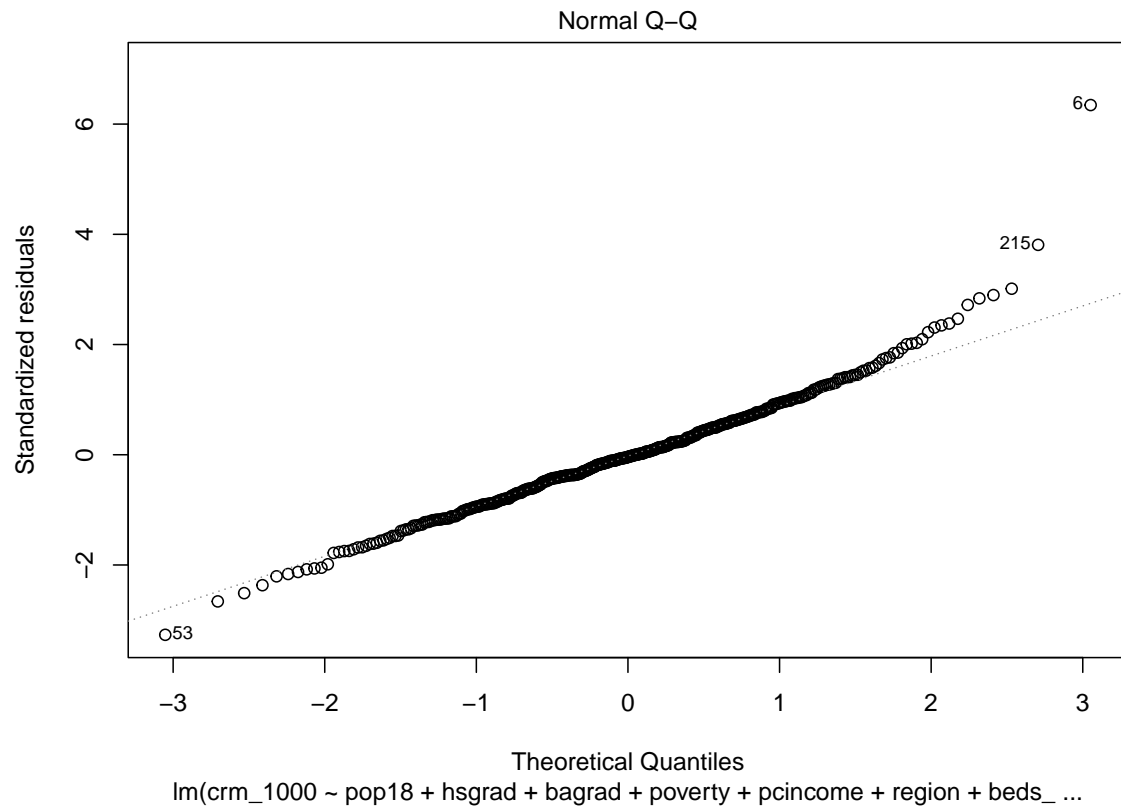


(non) normality of residuals and outliers

```
plot(multi_forward, which = 2)
```

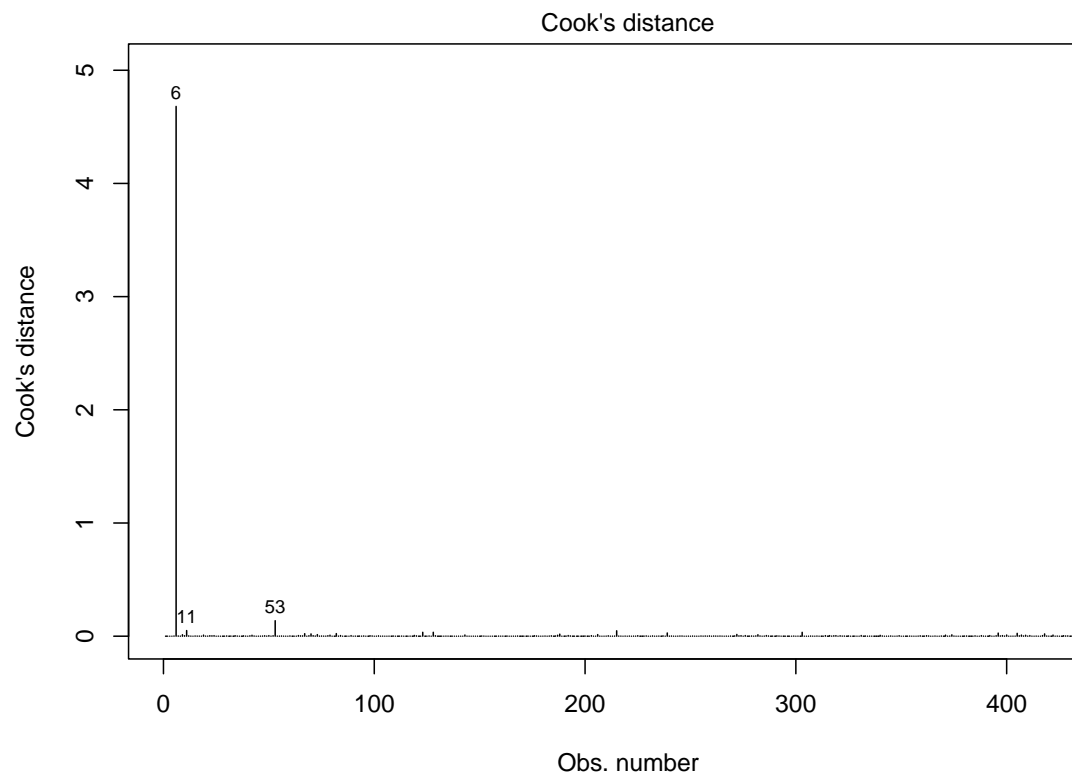


```
plot(multi_both, which = 2)
```



residuals vs leverage plot

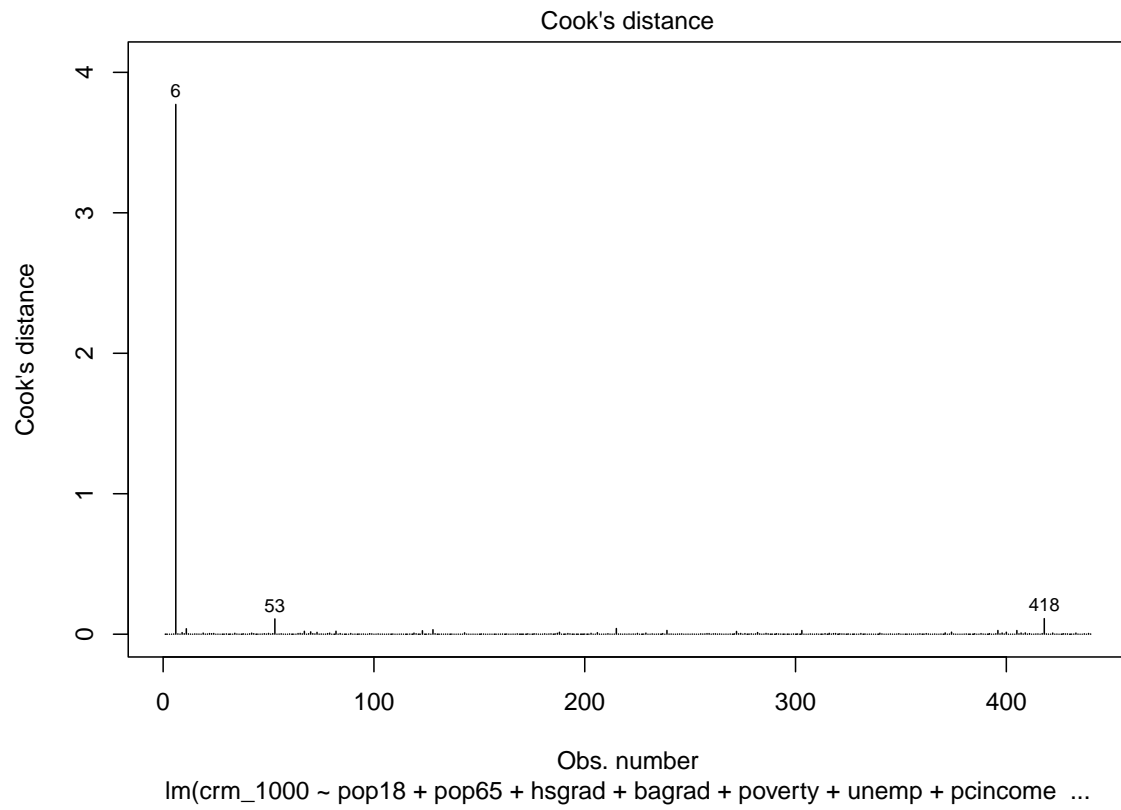
```
plot(multi_back, which = 4)
```



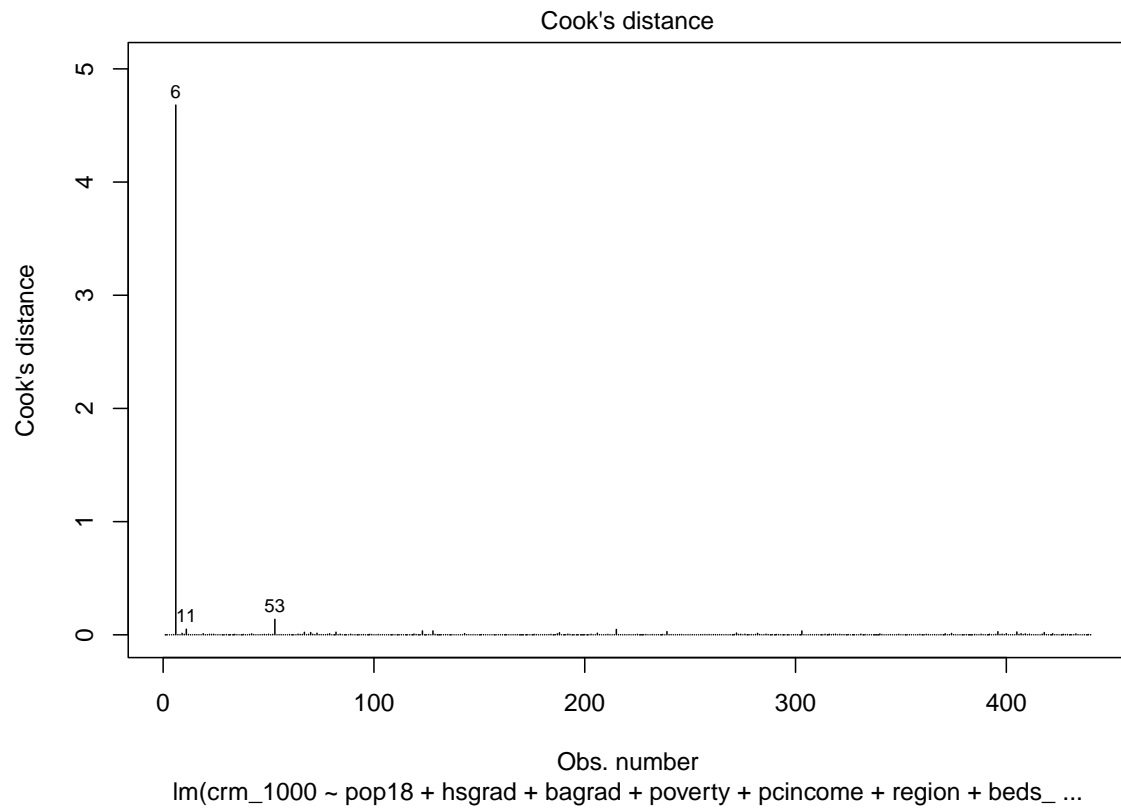
To identify influential cases

```
plot(multi_forward, which = 4)
```



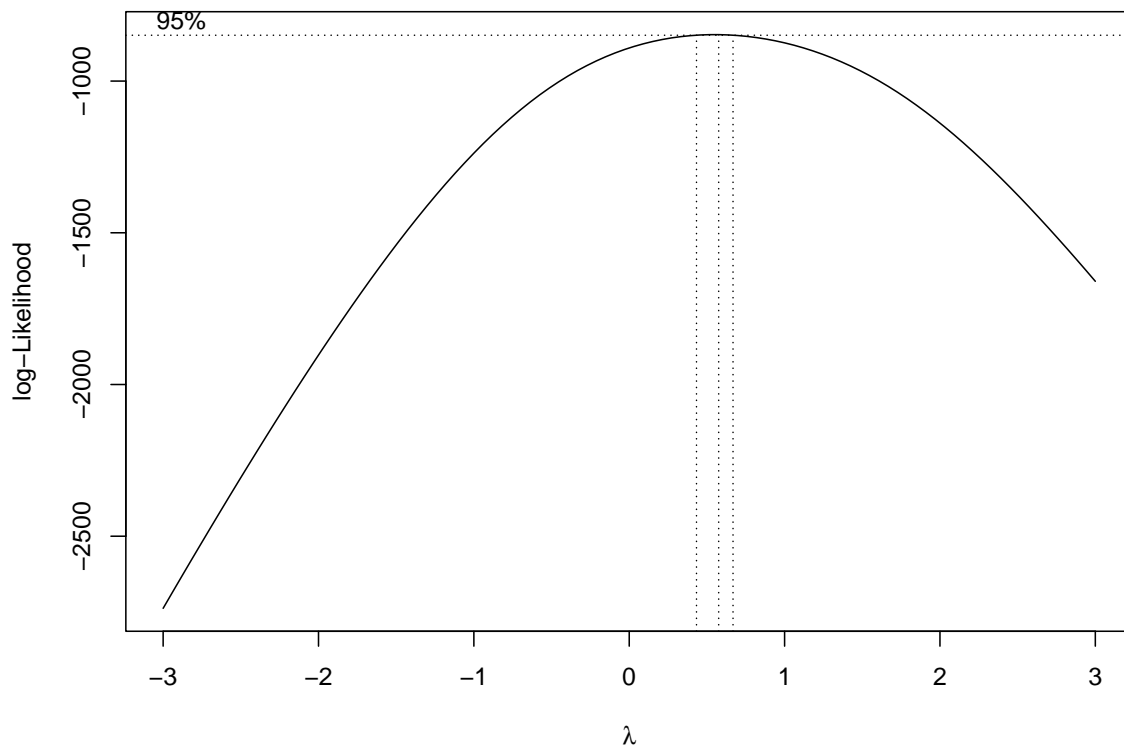


```
plot(multi_both, which = 4)
```

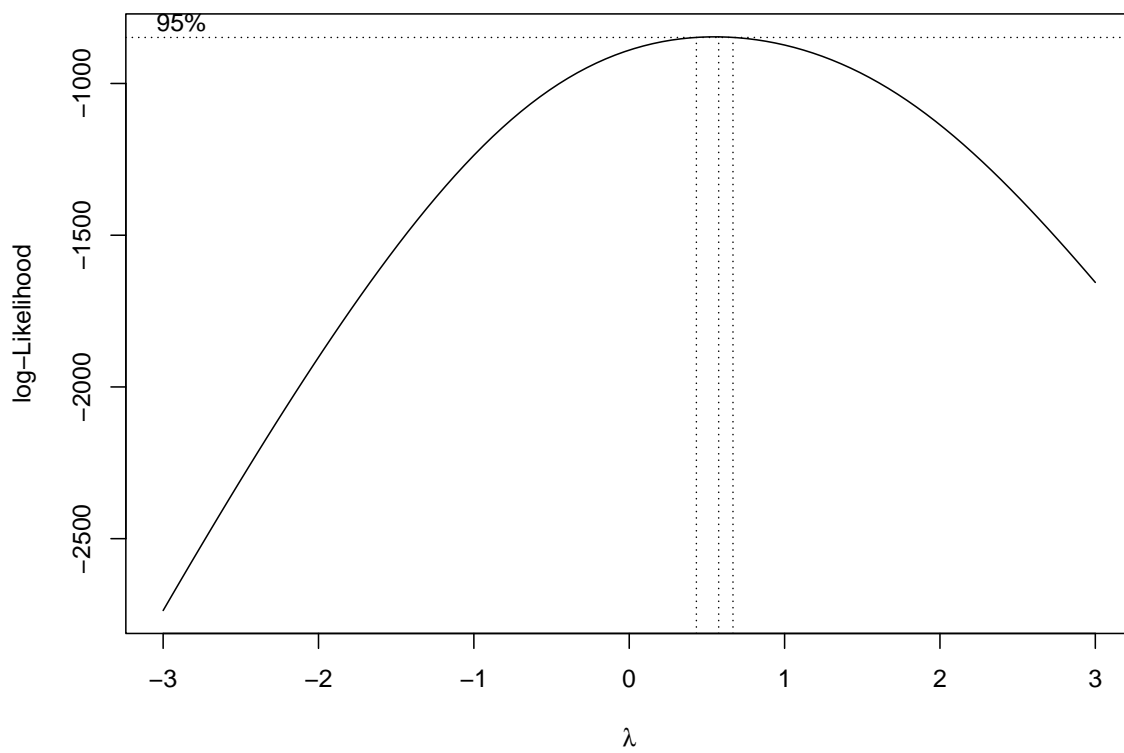


## Box-Cox Transformation

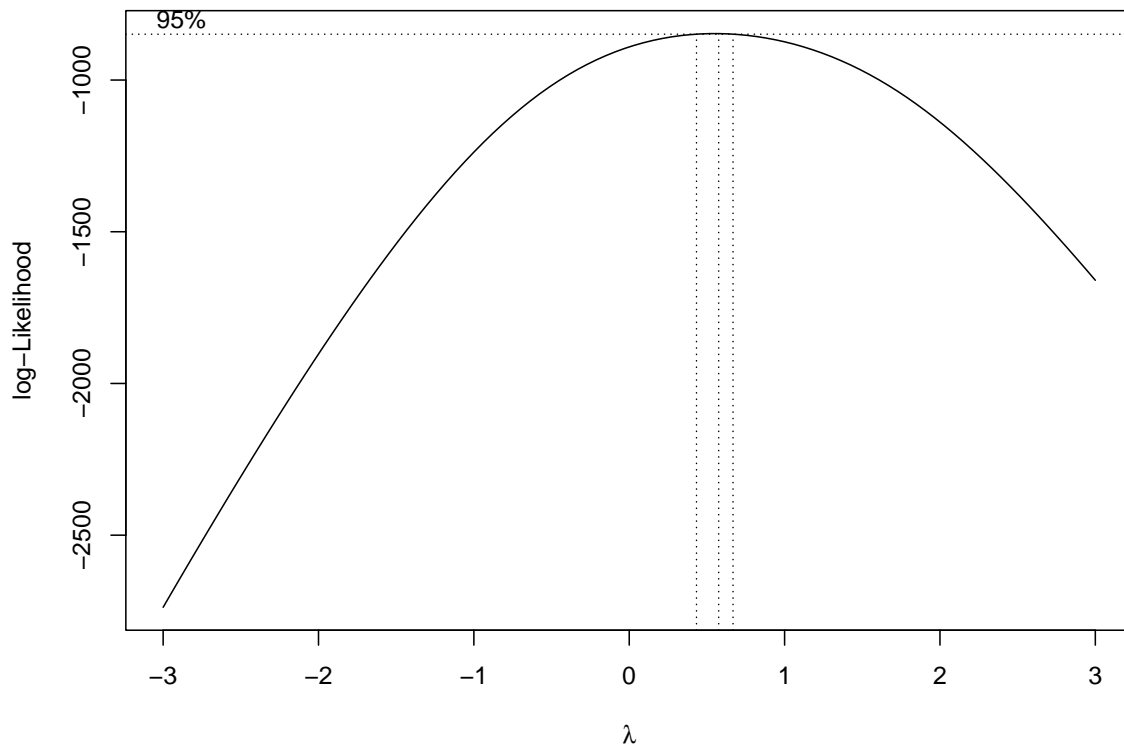
```
boxcox(multi_back, lambda = seq(-3, 3, by = 0.25))
```



```
boxcox(multi_forward, lambda = seq(-3, 3, by = 0.25))
```



```
boxcox(multi_both, lambda = seq(-3, 3, by = 0.25))
```



## Checking Outliers and Influential Points

```
sum_cdi_wo = sum_cdi[-c(6,53,418),]
# for backward
without_back = lm(crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
  beds_1000 + pop_density, data = sum_cdi_wo)
with_back = lm(crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
  beds_1000 + pop_density, data = sum_cdi)
summary(with_back); summary(without_back)
```

```
##
## Call:
## lm(formula = crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome +
##   region + beds_1000 + pop_density, data = sum_cdi)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-55.058	-11.358	-0.699	10.514	75.665

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
## (Intercept)	-7.103e+01	2.473e+01	-2.872	0.004276	**
## pop18	9.017e-01	2.856e-01	3.157	0.001707	**
## hsgrad	5.330e-01	2.653e-01	2.009	0.045162	*
## bagrad	-6.489e-01	2.845e-01	-2.280	0.023071	*
## poverty	2.140e+00	3.630e-01	5.896	7.57e-09	***
## pcincome	1.460e-03	4.542e-04	3.215	0.001404	**
## region2	9.615e+00	2.680e+00	3.588	0.000372	***

```

## region3      2.728e+01  2.549e+00  10.702 < 2e-16 ***
## region4      2.264e+01  3.055e+00   7.412 6.70e-13 ***
## beds_1000    2.412e+00  5.069e-01   4.759 2.66e-06 ***
## pop_density  5.154e-03  4.449e-04  11.584 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18 on 429 degrees of freedom
## Multiple R-squared:  0.576, Adjusted R-squared:  0.5661
## F-statistic: 58.28 on 10 and 429 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome +
##     region + beds_1000 + pop_density, data = sum_cdi_wo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -46.86 -11.25  -0.90   10.69   61.61
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.882e+01  2.354e+01  -3.349 0.000884 ***
## pop18        1.223e+00  2.766e-01   4.422 1.25e-05 ***
## hsgrad       3.494e-01  2.532e-01   1.380 0.168373
## bagrad      -6.850e-01  2.708e-01  -2.530 0.011780 *
## poverty      2.070e+00  3.472e-01   5.961 5.25e-09 ***
## pcincome     2.125e-03  4.471e-04   4.753 2.75e-06 ***
## region2      1.072e+01  2.551e+00   4.202 3.23e-05 ***
## region3      2.819e+01  2.425e+00  11.624 < 2e-16 ***
## region4      2.506e+01  2.936e+00   8.533 2.52e-16 ***
## beds_1000    3.268e+00  5.119e-01   6.385 4.49e-10 ***
## pop_density  2.348e-03  6.946e-04   3.380 0.000791 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.09 on 426 degrees of freedom
## Multiple R-squared:  0.5372, Adjusted R-squared:  0.5264
## F-statistic: 49.45 on 10 and 426 DF,  p-value: < 2.2e-16

# for forward
without_for = lm(crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty + unemp +
  pcincome + region + docs_1000 + beds_1000 + pop_density, data = sum_cdi_wo)
with_for = lm(crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty + unemp +
  pcincome + region + docs_1000 + beds_1000 + pop_density, data = sum_cdi)
summary(with_for); summary(without_for)

##
## Call:
## lm(formula = crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty +
##     unemp + pcincome + region + docs_1000 + beds_1000 + pop_density,
##     data = sum_cdi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max

```

```

## -53.755 -11.637 -0.547 10.682 74.653
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.326e+01 2.772e+01 -2.643 0.008515 **
## pop18        8.338e-01 3.323e-01  2.509 0.012480 *
## pop65       -1.637e-01 3.092e-01 -0.529 0.596875
## hsgrad       5.573e-01 2.719e-01  2.049 0.041048 *
## bagrad      -5.565e-01 3.001e-01 -1.854 0.064402 .
## poverty     1.987e+00 3.895e-01  5.102 5.08e-07 ***
## unemp       5.221e-01 5.376e-01  0.971 0.332059
## pcincome    1.409e-03 4.673e-04  3.015 0.002722 **
## region2     9.607e+00 2.761e+00  3.479 0.000555 ***
## region3     2.795e+01 2.693e+00 10.379 < 2e-16 ***
## region4     2.299e+01 3.121e+00  7.365 9.24e-13 ***
## docs_1000   -5.658e-01 1.032e+00 -0.548 0.583889
## beds_1000   2.970e+00 8.023e-01  3.701 0.000243 ***
## pop_density 5.217e-03 4.502e-04 11.588 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.03 on 426 degrees of freedom
## Multiple R-squared:  0.5774, Adjusted R-squared:  0.5645
## F-statistic: 44.77 on 13 and 426 DF, p-value: < 2.2e-16
##
## Call:
## lm(formula = crm_1000 ~ pop18 + pop65 + hsgrad + bagrad + poverty +
##      unemp + pcincome + region + docs_1000 + beds_1000 + pop_density,
##      data = sum_cdi_wo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.885 -11.702  -0.672  10.146  63.426
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.042e+01 2.646e+01 -3.039 0.002520 **
## pop18        1.118e+00 3.241e-01  3.449 0.000618 ***
## pop65       -9.255e-02 2.965e-01 -0.312 0.755054
## hsgrad       4.336e-01 2.582e-01  1.679 0.093857 .
## bagrad      -7.410e-01 2.856e-01 -2.594 0.009814 **
## poverty     1.886e+00 3.733e-01  5.052 6.51e-07 ***
## unemp       6.442e-01 5.105e-01  1.262 0.207671
## pcincome    1.887e-03 4.621e-04  4.083 5.31e-05 ***
## region2     1.137e+01 2.626e+00  4.330 1.86e-05 ***
## region3     2.925e+01 2.555e+00 11.451 < 2e-16 ***
## region4     2.458e+01 2.984e+00  8.237 2.23e-15 ***
## docs_1000   1.998e+00 1.195e+00  1.672 0.095323 .
## beds_1000   2.642e+00 7.741e-01  3.413 0.000705 ***
## pop_density 2.116e-03 7.104e-04  2.978 0.003066 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```
## Residual standard error: 17.06 on 423 degrees of freedom
## Multiple R-squared:  0.5418, Adjusted R-squared:  0.5278
## F-statistic: 38.48 on 13 and 423 DF,  p-value: < 2.2e-16

# for both
without_both = lm(crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
  beds_1000 + pop_density, data = sum_cdi_wo)
with_both = lm(crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome + region +
  beds_1000 + pop_density, data = sum_cdi)
summary(with_both); summary(without_both)

##
## Call:
## lm(formula = crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome +
##     region + beds_1000 + pop_density, data = sum_cdi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -55.058 -11.358  -0.699   10.514   75.665
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.103e+01  2.473e+01  -2.872 0.004276 **
## pop18        9.017e-01  2.856e-01   3.157 0.001707 **
## hsgrad       5.330e-01  2.653e-01   2.009 0.045162 *
## bagrad      -6.489e-01  2.845e-01  -2.280 0.023071 *
## poverty     2.140e+00  3.630e-01   5.896 7.57e-09 ***
## pcincome     1.460e-03  4.542e-04   3.215 0.001404 **
## region2      9.615e+00  2.680e+00   3.588 0.000372 ***
## region3      2.728e+01  2.549e+00  10.702 < 2e-16 ***
## region4      2.264e+01  3.055e+00   7.412 6.70e-13 ***
## beds_1000    2.412e+00  5.069e-01   4.759 2.66e-06 ***
## pop_density  5.154e-03  4.449e-04  11.584 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18 on 429 degrees of freedom
## Multiple R-squared:  0.576, Adjusted R-squared:  0.5661
## F-statistic: 58.28 on 10 and 429 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = crm_1000 ~ pop18 + hsgrad + bagrad + poverty + pcincome +
##     region + beds_1000 + pop_density, data = sum_cdi_wo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -46.86 -11.25  -0.90   10.69   61.61
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.882e+01  2.354e+01  -3.349 0.000884 ***
## pop18        1.223e+00  2.766e-01   4.422 1.25e-05 ***
## hsgrad       3.494e-01  2.532e-01   1.380 0.168373
## bagrad      -6.850e-01  2.708e-01  -2.530 0.011780 *
```

```

## poverty      2.070e+00  3.472e-01   5.961 5.25e-09 ***
## pcincome     2.125e-03  4.471e-04   4.753 2.75e-06 ***
## region2      1.072e+01  2.551e+00   4.202 3.23e-05 ***
## region3      2.819e+01  2.425e+00  11.624 < 2e-16 ***
## region4      2.506e+01  2.936e+00   8.533 2.52e-16 ***
## beds_1000    3.268e+00  5.119e-01   6.385 4.49e-10 ***
## pop_density  2.348e-03  6.946e-04   3.380 0.000791 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.09 on 426 degrees of freedom
## Multiple R-squared:  0.5372, Adjusted R-squared:  0.5264
## F-statistic: 49.45 on 10 and 426 DF,  p-value: < 2.2e-16

```