# Lab3 YZ Draft

*Yulia Zamriy*

*March 29, 2018*

```
#setwd("/home/yulia/Documents/MIDS/W203/Lab_3/")
crime <- read.csv("crime_v2.csv", stringsAsFactors = FALSE)
crime <- na.omit(crime)
#str(crime)
```

```
crime$prbconv <- as.numeric(crime$prbconv)
crime$county <- NULL
crime$year <- NULL
```

```
crime_summary <- data.frame(t(mapply(summary, crime)))
#str(crime_summary)
crime_summary <- crime_summary[,c("Min.","Mean","Max.")]
crime_summary$Min. <- round(crime_summary$Min.,5)
crime_summary$Mean <- round(crime_summary$Mean,4)
crime_summary$Max. <- round(crime_summary$Max.,4)
```

```
crime_summary <- data.frame(t(mapply(summary, crime)))
#str(crime_summary)
crime_summary <- crime_summary[,c("Min.","Mean","Max.")]
crime_summary$Min. <- round(crime_summary$Min.,5)
crime_summary$Mean <- round(crime_summary$Mean,4)
crime_summary$Max. <- round(crime_summary$Max.,4)
```

```
kable(crime_summary, booktabs = TRUE) %>%
  kable_styling(font_size = 7)
```

| | Min. | Mean | Max. |
|---|---|---|---|
| crmrte | 0.00553 | 0.0334 | 0.0990 |
| prbarr | 0.09277 | 0.2949 | 1.0909 |
| prbconv | 0.06838 | 0.5513 | 2.1212 |
| prbpris | 0.15000 | 0.4108 | 0.6000 |
| avgsen | 5.38000 | 9.6468 | 20.7000 |
| polpc | 0.00075 | 0.0017 | 0.0091 |
| density | 0.00002 | 1.4288 | 8.8277 |
| taxpc | 25.69287 | 38.0551 | 119.7615 |
| west | 0.00000 | 0.2527 | 1.0000 |
| central | 0.00000 | 0.3736 | 1.0000 |
| urban | 0.00000 | 0.0879 | 1.0000 |
| pctmin80 | 1.28365 | 25.4955 | 64.3482 |
| wcon | 193.64316 | 285.3585 | 436.7666 |
| wtuc | 187.61726 | 411.6680 | 613.2261 |
| wtrd | 154.20900 | 211.5529 | 354.6761 |
| wfir | 170.94017 | 322.0982 | 509.4655 |
| wser | 133.04306 | 275.5642 | 2177.0681 |
| wmfg | 157.41000 | 335.5887 | 646.8500 |
| wfed | 326.10001 | 442.9007 | 597.9500 |
| wsta | 258.32999 | 357.5220 | 499.5900 |
| wloc | 239.17000 | 312.6808 | 388.0900 |
| mix | 0.01961 | 0.1288 | 0.4651 |
| pctymle | 0.06216 | 0.0840 | 0.2487 |

```
crime$region <- ifelse(crime$west == 1, "west",
                       ifelse(crime$central == 1, "central", "other"))
aggregate(crmrte ~ region, data = crime, mean)
```

```
##    region     crmrte
## 1 central 0.03699627
## 2   other 0.03739491
## 3    west 0.02216183
```

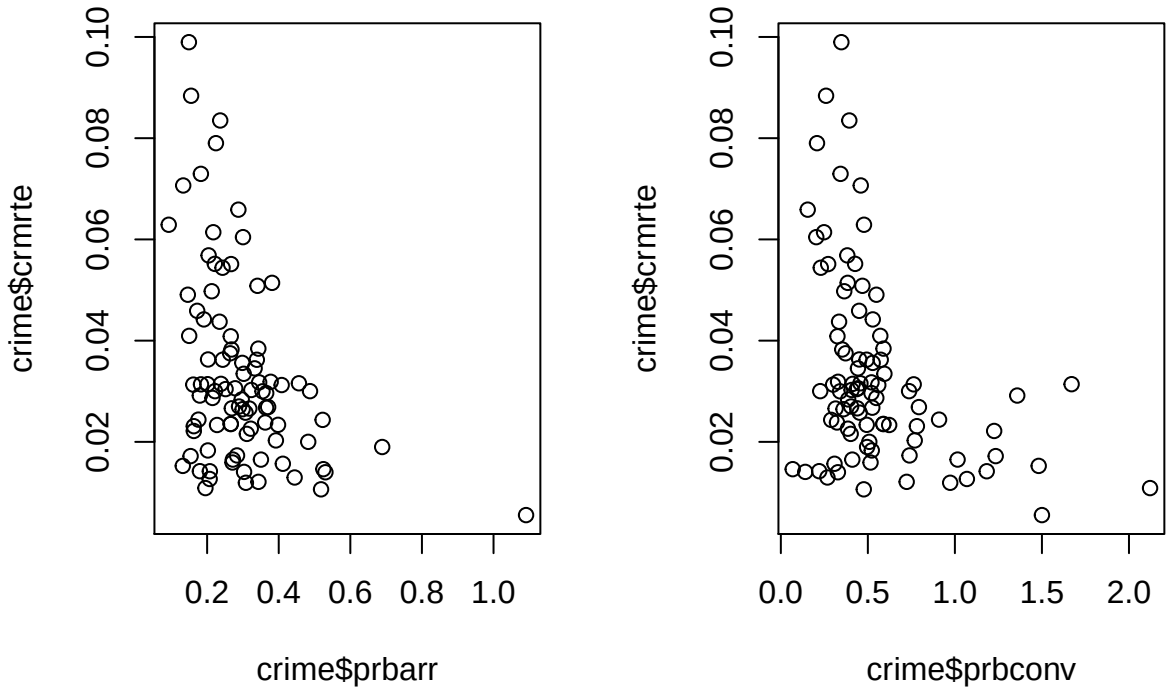
```
nrow(crime[crime$prbarr >= 1,])
```

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

```
nrow(crime[crime$prbconv >= 1,])
```

```
## [1] 10
```

```
par(mfrow=c(1,2))
plot(crime$prbarr, crime$crmrte)
plot(crime$prbconv, crime$crmrte)
```




```
crime$prbarr_imp <- ifelse(crime$prbarr > 1, mean(crime$prbarr), crime$prbarr)
summary(crime$prbarr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09277 0.20568 0.27095 0.29492 0.34438 1.09091
```

```
summary(crime$prbarr_imp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09277 0.20568 0.27095 0.28617 0.34323 0.68902
```

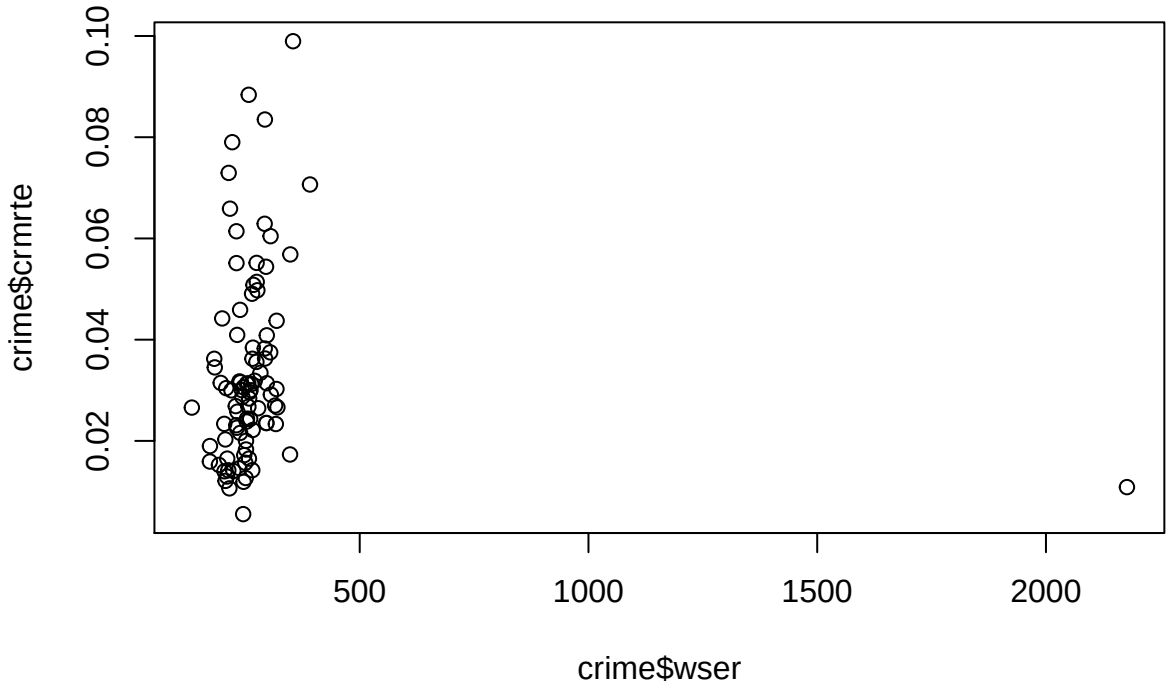
```
crime$prbconv_imp <- ifelse(crime$prbconv > 1, 1, crime$prbconv)
summary(crime$prbconv)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.55128 0.58886 2.12121
```

```
summary(crime$prbconv_imp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.50888 0.58886 1.00000
```

```
plot(crime$wser, crime$crmrte)
```




```
crime$wser_imp <- ifelse(crime$wser > 2000, mean(crime[crime$wser < 2000,]$wser), crime$wser)
summary(crime$wser)
```

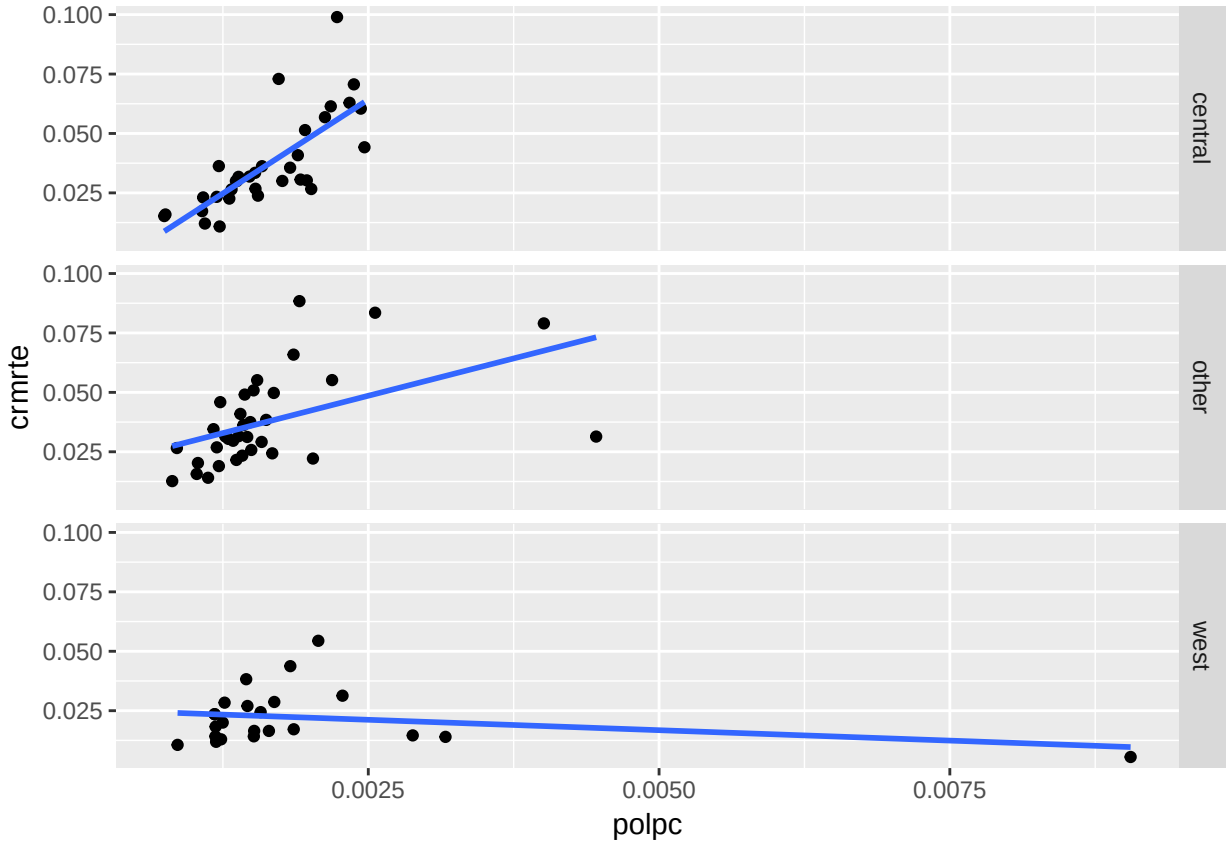
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   133.0   229.7   253.2   275.6   280.5  2177.1
```

```
summary(crime$wser_imp)
```

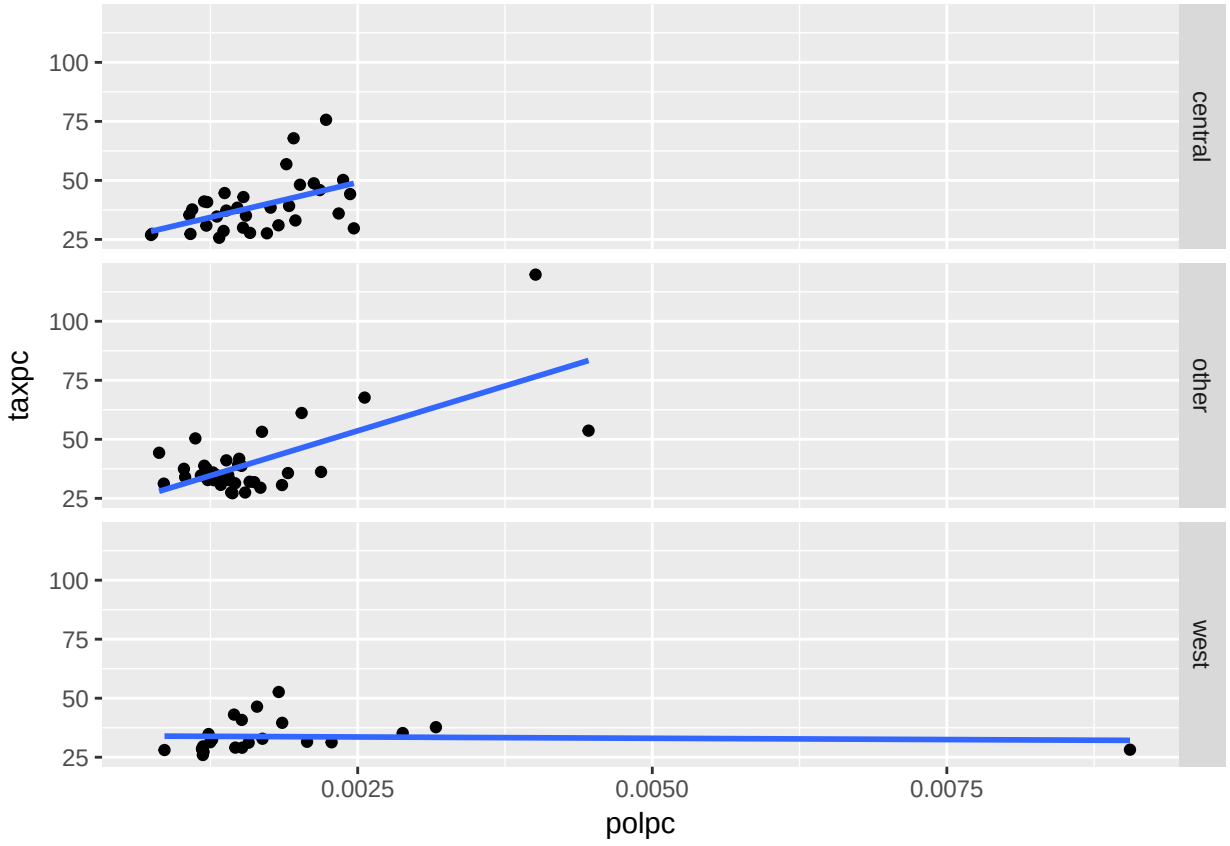
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   133.0   229.7   253.2   254.4   277.2   391.3
```

```
ggplot(crime, aes(polpc, crmrte)) +
  geom_point() +
  facet_grid(region~.) +
```

```
  geom_smooth(method = "lm", se = FALSE)
```




```
ggplot(crime, aes(polpc, taxpc)) +
  geom_point() +
  facet_grid(region~.) +
  geom_smooth(method = "lm", se = FALSE)
```




```
crime$polpc_imp <-
  ifelse(crime$polpc == max(crime$polpc), mean(crime[crime$west == 1,]$polpc), crime$polpc)
summary(crime$polpc)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0007459 0.0012308 0.0014853 0.0017022 0.0018768 0.0090543
```

```
summary(crime$polpc_imp)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0007459 0.0012308 0.0014853 0.0016239 0.0018768 0.0044592
```

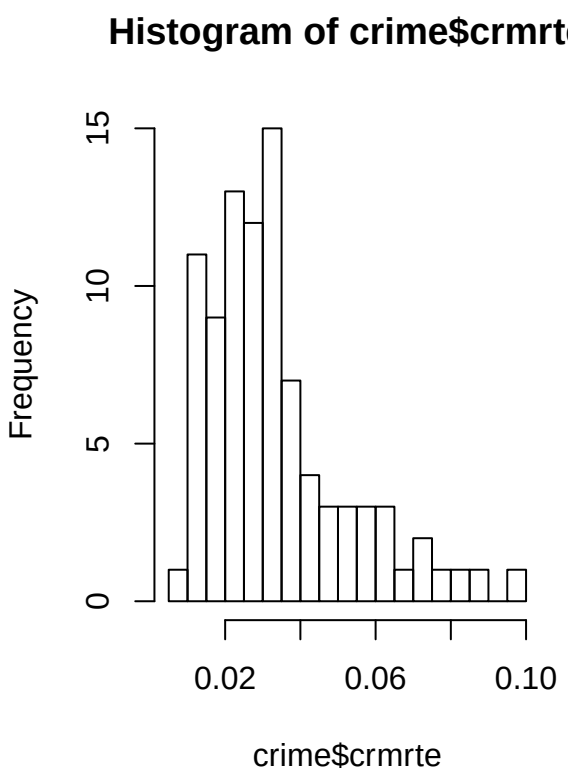
```
# Prepare a .RData for easier sharing and usage.
ind_variables <- c( 'crmrte',
  'prbarr_imp', 'prbconv_imp', 'prbpris', 'avgsen',
  'polpc_imp', 'density', 'taxpc', 'west', 'central', 'urban', 'pctmin80', 'wcon',
  'wtuc', 'wtrd', 'wfir', 'wser_imp', 'wmfg', 'wfed', 'wsta', 'wloc', 'mix',
  'pctymle'
)
var_labels <- c('crimes committed per person',
  'probability of arrest', 'probability of conviction',
  'probability of prison sentence', 'avg. sentence, days',
  'police per capita', 'people per sq. mile', 'tax revenue per capita',
  '=1 if in western N.C.', '=1 if in central N.C.', '=1 if in SMSA',
  'perc. minority, 1980', 'weekly wage, construction',
  'wkly wge, trns, util, commun', 'wkly wge, whlesle, retail trade',
  'wkly wge, fin, ins, real est', 'wkly wge, service industry',
  'wkly wge, manufacturing', 'wkly wge, fed employees',
```

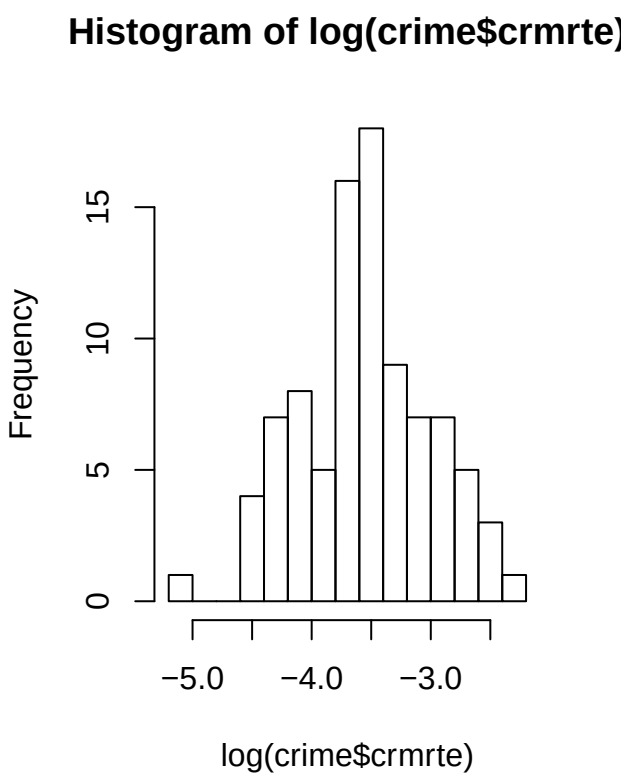
```
  'wkly wge, state employees', 'wkly wge, local gov emps',
  'offense mix: face-to-face/other', 'percent young male'
)
impact <- c("Dependent",
  "Negative" , "Negative", "Negative", "Negative",
             "Negative", "Positive", "Negative",
             "Unclear", "Unclear", "Unclear", "Unclear",
             "Negative","Negative","Negative",
             "Negative", "Negative", "Negative", "Negative",
             "Negative", "Negative", "Unclear","Positive")
control <- c("NA","Yes", "Yes", "Yes", "Yes",
              "Yes", "No", "Yes",
              "No", "No", "No","No",
              "Yes", "Yes", "Yes",
              "Yes", "Yes", "Yes", "Yes",
              "Yes", "Yes", "No", "No")
cor_w_crimerate <- round(cor(crime[,ind_variables])[1,],2)
desc <- data.frame(ind_variables, var_labels, impact, cor_w_crimerate, control,
                   row.names = NULL)
colnames(desc) <- c("Explanatory Variables",
                    "Explanation",
                    "Expected Impact on Crime Rate",
                    "Correlation w/ Crime Rate",
                    "Can Gov Impact This?")
```

```
kable(desc, booktabs = TRUE, align = c("llccc")) %>%
  kable_styling(latex_options = c("scale_down"),
                full_width = FALSE) %>%
  row_spec(0, bold = TRUE) %>%
  column_spec(1, width = "8em") %>%
  column_spec(3, width = "10em") %>%
  column_spec(4, width = "8em") %>%
  column_spec(5, width = "9em")
```

| Explanatory Variables | Explanation | Expected Impact on Crime Rate | Correlation w/ Crime Rate | Can Gov Impact This? |
|---|---|---|---|---|
| crmrte | crimes committed per person | Dependent | 1.00 | NA |
| prbarr_imp | probability of arrest | Negative | -0.38 | Yes |
| prbconv_imp | probability of conviction | Negative | -0.42 | Yes |
| prbpris | probability of prison sentence | Negative | 0.05 | Yes |
| avgsen | avg. sentence, days | Negative | 0.03 | Yes |
| polpc_imp | police per capita | Negative | 0.47 | Yes |
| density | people per sq. mile | Positive | 0.73 | No |
| taxpc | tax revenue per capita | Negative | 0.45 | Yes |
| west | =1 if in western N.C. | Unclear | -0.35 | No |
| central | =1 if in central N.C. | Unclear | 0.17 | No |
| urban | =1 if in SMSA | Unclear | 0.62 | No |
| pctmin80 | perc. minority, 1980 | Unclear | 0.19 | No |
| wcon | weekly wage, construction | Negative | 0.39 | Yes |
| wtuc | wkly wge, trns, util, commun | Negative | 0.23 | Yes |
| wtrd | wkly wge, whlesle, retail trade | Negative | 0.41 | Yes |
| wfir | wkly wge, fin, ins, real est | Negative | 0.33 | Yes |
| wser_imp | wkly wge, service industry | Negative | 0.34 | Yes |
| wmfg | wkly wge, manufacturing | Negative | 0.35 | Yes |
| wfed | wkly wge, fed employees | Negative | 0.49 | Yes |
| wsta | wkly wge, state employees | Negative | 0.20 | Yes |
| wloc | wkly wge, local gov emps | Negative | 0.35 | Yes |
| mix | offense mix: face-to-face/other | Unclear | -0.13 | No |
| pctymle | percent young male | Positive | 0.29 | No |

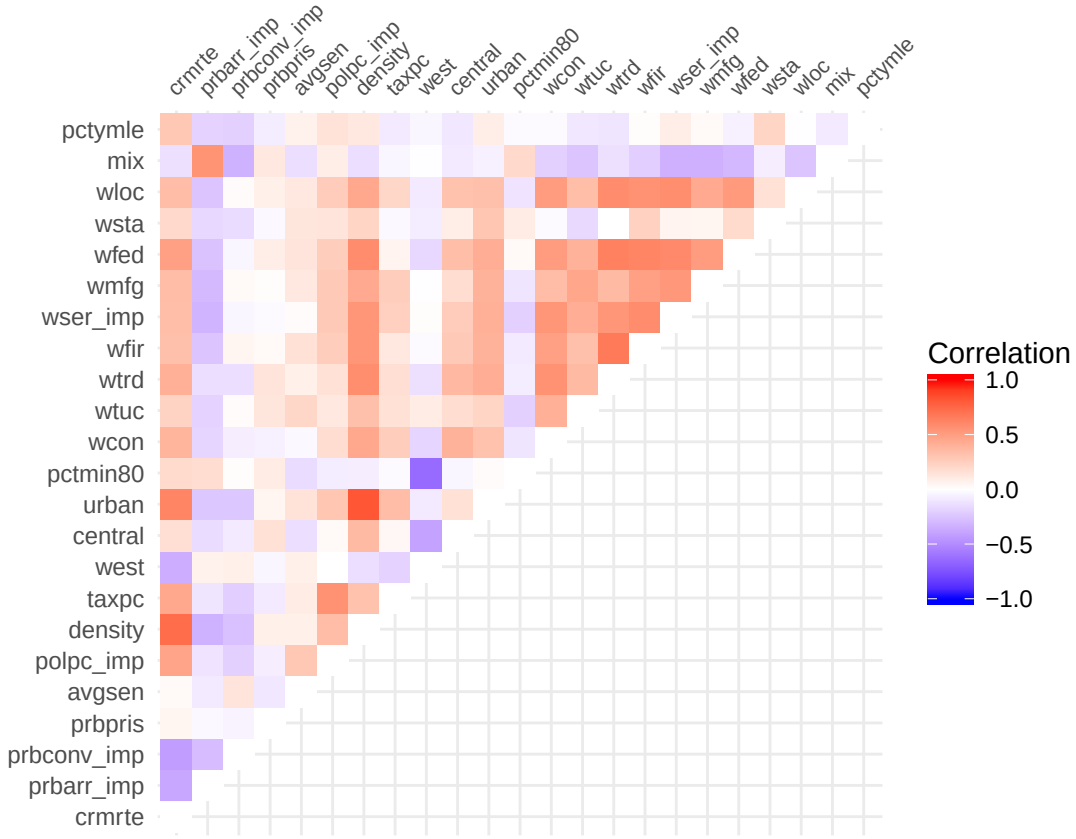
```
par(mfrow=c(1,2))
hist(crime$crmrte, breaks = 15)
hist(log(crime$crmrte), breaks = 15)
```




```
cor_mat <- round(cor(crime[,ind_variables]),2)
get_upper_tri <- function(cor_mat){
    cor_mat[lower.tri(cor_mat)]<- NA
    return(cor_mat)
}
cor_mat_upper <- get_upper_tri(cor_mat)
cor_mat_upper2 <- melt(cor_mat_upper, na.rm = TRUE)
cor_mat_upper2[cor_mat_upper2$value == 1,]$value <- 0
```

```
ggplot(data = cor_mat_upper2, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                       midpoint = 0, limit = c(-1,1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  scale_x_discrete(position = "top") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, size = 8, hjust = 0),
        axis.title.x=element_blank(),
        axis.title.y=element_blank()) +
  coord_fixed()
```




```
ind_vars_transf <- c("wcon", "wtuc", "wtrd", "wfir",
              "wser", "wser_imp", "wmfg", "wfed", "wsta", "wloc",
              "prbarr", "prbarr_imp", "prbconv", "prbconv_imp", "prbpris", "avgsen",
                "polpc_imp", "density", "taxpc", "pctmin80", "mix", "pctymle")
crime[paste(ind_vars_transf,"ln",sep = ".")] <- log(crime[ind_vars_transf])
crime[paste(ind_vars_transf,"sq",sep = ".")] <- (crime[ind_vars_transf])**2
```

```
ind_vars1 <- c("density", "west", "prbarr", "prbconv")
crmrte_formula1 <- as.formula(paste("log(crmrte) ~", paste(ind_vars1, collapse = "+"), sep = ""))
crmrte_lm1 <- lm(crmrte_formula1, data = crime)
coeftest(crmrte_lm1)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -2.989700   0.119319 -25.0564 < 2.2e-16 ***
## density      0.150219   0.023519   6.3870 8.317e-09 ***
## west        -0.356472   0.076327  -4.6703 1.100e-05 ***
## prbarr      -1.269137   0.255149  -4.9741 3.320e-06 ***
## prbconv     -0.552467   0.096285  -5.7378 1.404e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm1)$adj.r.squared
```

```
## [1] 0.6770812
```

```r
ind_vars2 <- c("density", "west", "prbarr", "prbconv", "polpc_imp.ln", "pctmin80",
               "west*polpc_imp.ln")
crmrte_formula2 <- as.formula(paste("log(crmrte) ~", paste(ind_vars2, collapse = "+"), sep = ""))
crmrte_lm2 <- lm(crmrte_formula2, data = crime)
coeftest(crmrte_lm2)
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         1.1292993  0.6820650  1.6557 0.1015577
## density             0.1026480  0.0208024  4.9344 4.083e-06 ***
## west               -3.6778777  1.2944306 -2.8413 0.0056513 **
## prbarr             -1.4036147  0.2075950 -6.7613 1.775e-09 ***
## prbconv            -0.5657364  0.0772900 -7.3197 1.446e-10 ***
## polpc_imp.ln        0.6549475  0.1024599  6.3922 9.059e-09 ***
## pctmin80            0.0077037  0.0020841  3.6964 0.0003911 ***
## west:polpc_imp.ln -0.5394537  0.1999601 -2.6978 0.0084527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(crmrte_lm2)$adj.r.squared
```

```
## [1] 0.7954901
```

```r
ind_vars_all <- c("prbarr_imp", "prbconv", "prbpris", "avgsen", "polpc_imp.ln", "density", "taxpc.ln",
               "west", "central", "urban", "pctmin80", "wcon.ln", "wtuc.ln", "wtrd.ln", "wfir.ln",
               "wser_imp.ln", "wmfg.ln", "wfed.ln", "wsta.ln", "wloc.ln", "mix", "pctymle")
crmrte_formula_all <- as.formula(paste("log(crmrte) ~", paste(ind_vars_all, collapse = "+"), sep = ""))
crmrte_lm_all <- lm(crmrte_formula_all, data = crime)
coeftest(crmrte_lm_all)
```

```
##
## t test of coefficients:
##
##                 Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  -1.7998743  3.2030152 -0.5619  0.576010
## prbarr_imp   -1.7045460  0.3234426 -5.2700 1.520e-06 ***
## prbconv      -0.6657768  0.0913576 -7.2876 4.293e-10 ***
## prbpris      -0.1352744  0.3381287 -0.4001  0.690361
## avgsen       -0.0257841  0.0109404 -2.3568  0.021322 *
## polpc_imp.ln  0.5566838  0.1186111  4.6934 1.346e-05 ***
## density       0.1091624  0.0373346  2.9239  0.004692 **
## taxpc.ln     -0.0462678  0.1411971 -0.3277  0.744159
## west         -0.2297256  0.1056246 -2.1749  0.033117 *
## central      -0.1739208  0.0725520 -2.3972  0.019275 *
## urban        -0.0943715  0.1737469 -0.5432  0.588799
## pctmin80      0.0077879  0.0025133  3.0987  0.002826 **
## wcon.ln       0.3429113  0.2146021  1.5979  0.114704
## wtuc.ln       0.1633127  0.1613146  1.0124  0.314942
## wtrd.ln       0.2496244  0.2862947  0.8719  0.386323
## wfir.ln      -0.1628161  0.2404463 -0.6771  0.500614
## wser_imp.ln  -0.4981123  0.2229067 -2.2346  0.028732 *
## wmfg.ln      -0.0390599  0.1502282 -0.2600  0.795646
## wfed.ln       0.6942787  0.3240088  2.1428  0.035713 *
```

```
## wsta.ln      -0.3323233  0.2598635 -1.2788  0.205301
## wloc.ln       0.0468408  0.3942515  0.1188  0.905777
## mix          -0.6559210  0.4466844 -1.4684  0.146601
## pctymle       2.0885832  1.3051536  1.6003  0.114177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm_all)$adj.r.squared
```

```
## [1] 0.8193607
```

```
AIC(crmrte_lm1)
```

```
## [1] 52.14778
```

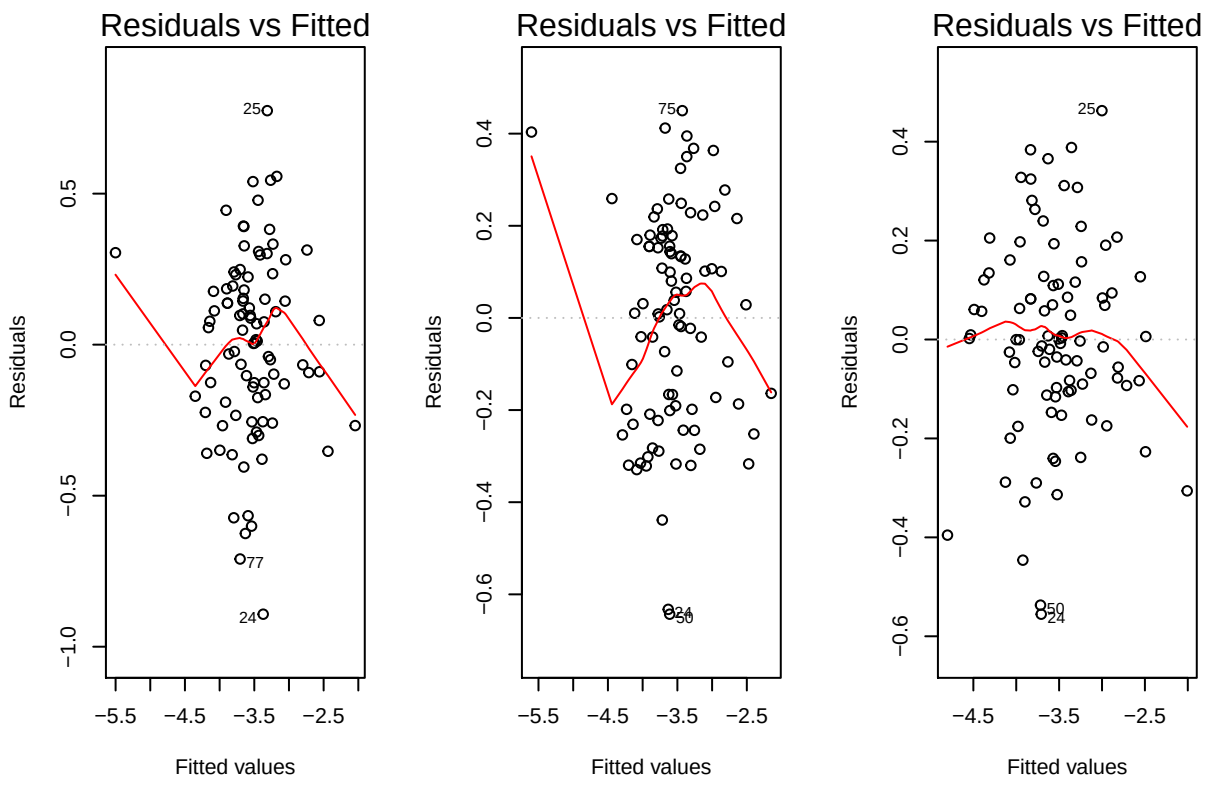
```
AIC(crmrte_lm2)
```

```
## [1] 13.34928
```

```
AIC(crmrte_lm_all)
```

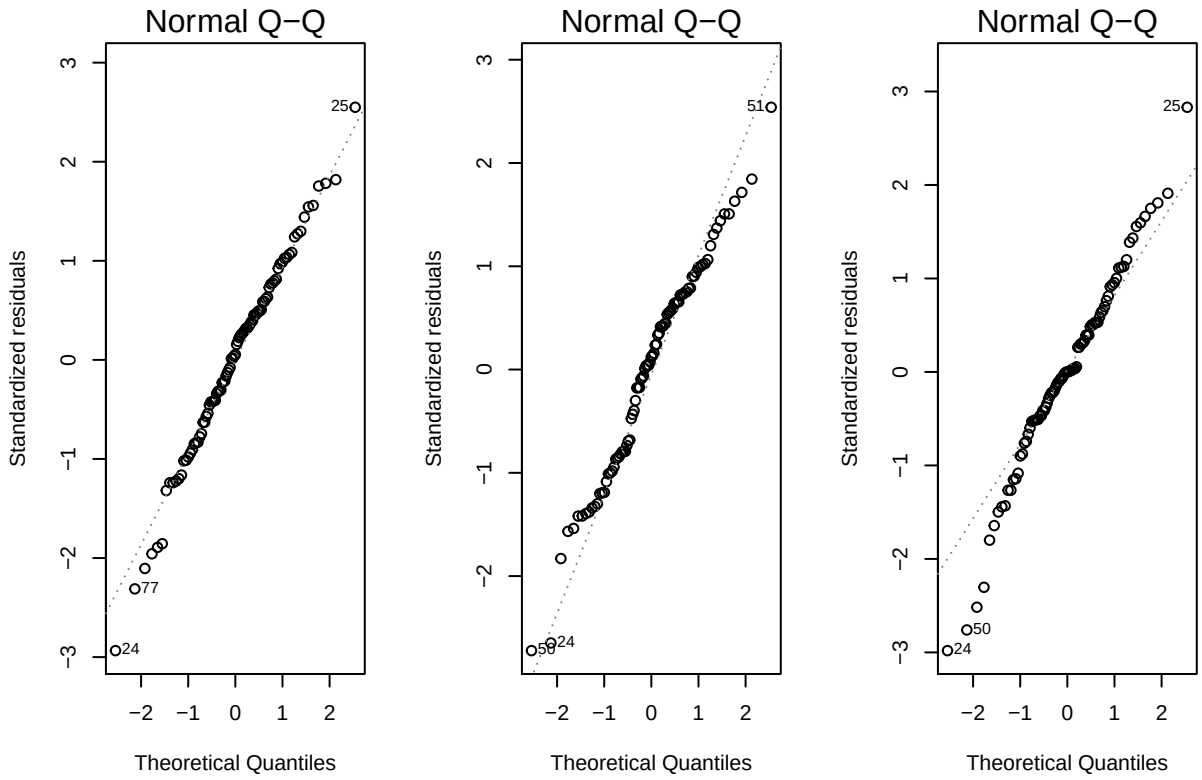
```
## [1] 13.91558
```

```
par(mfrow=c(1,3))
plot(crmrte_lm1, which = 1)
plot(crmrte_lm2, which = 1)
plot(crmrte_lm_all, which = 1)
```

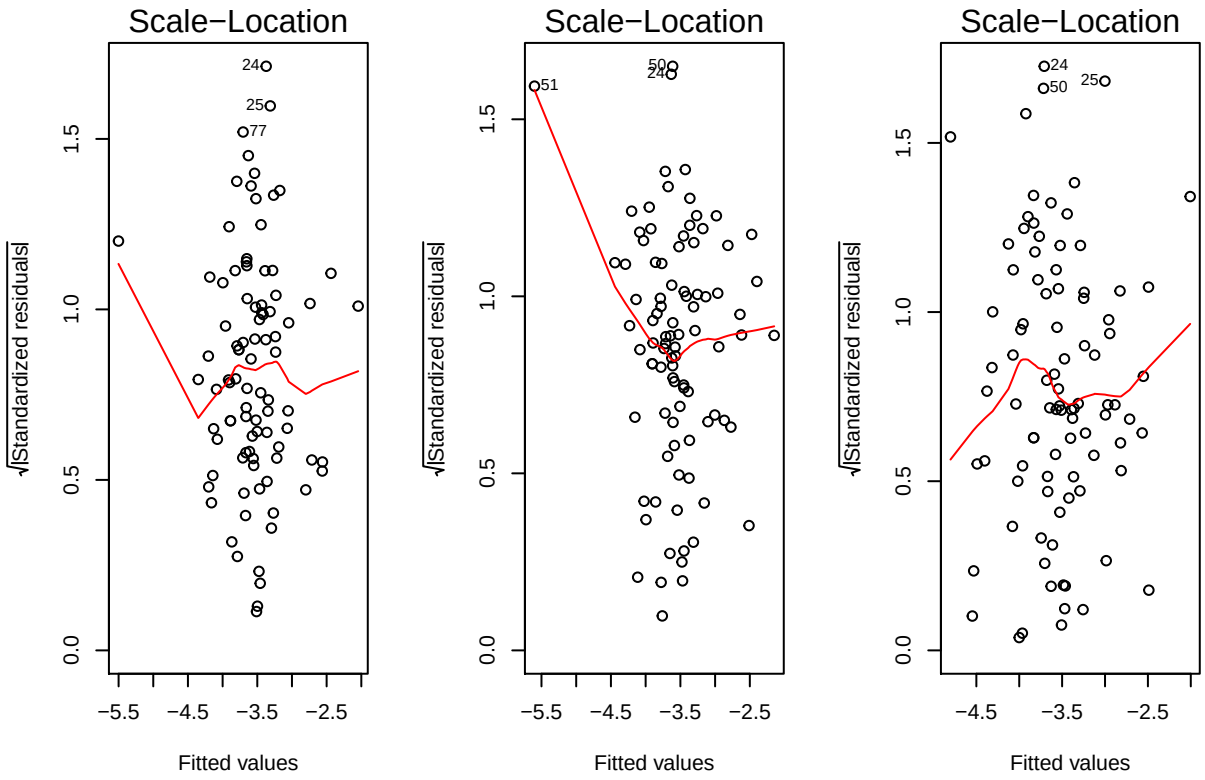



```
par(mfrow=c(1,3))
plot(crmrte_lm1, which = 2)
```

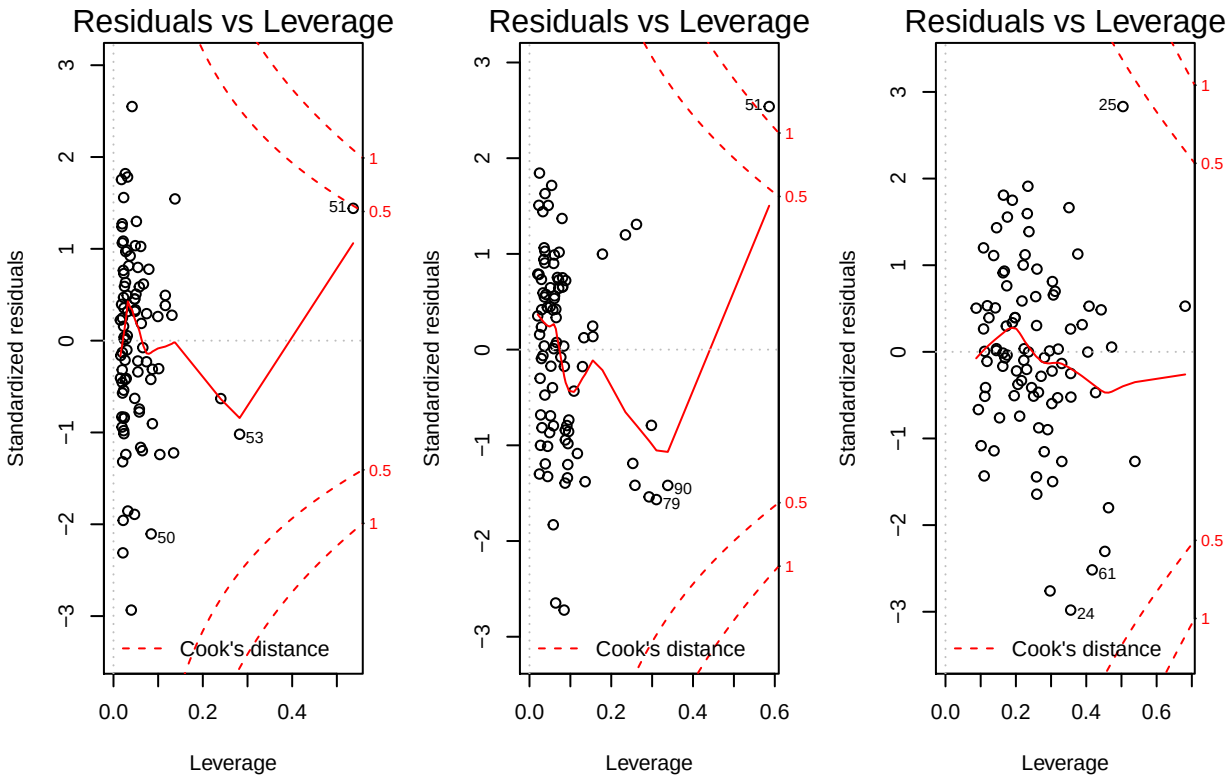
```
plot(crmrte_lm2, which = 2)
plot(crmrte_lm_all, which = 2)
```



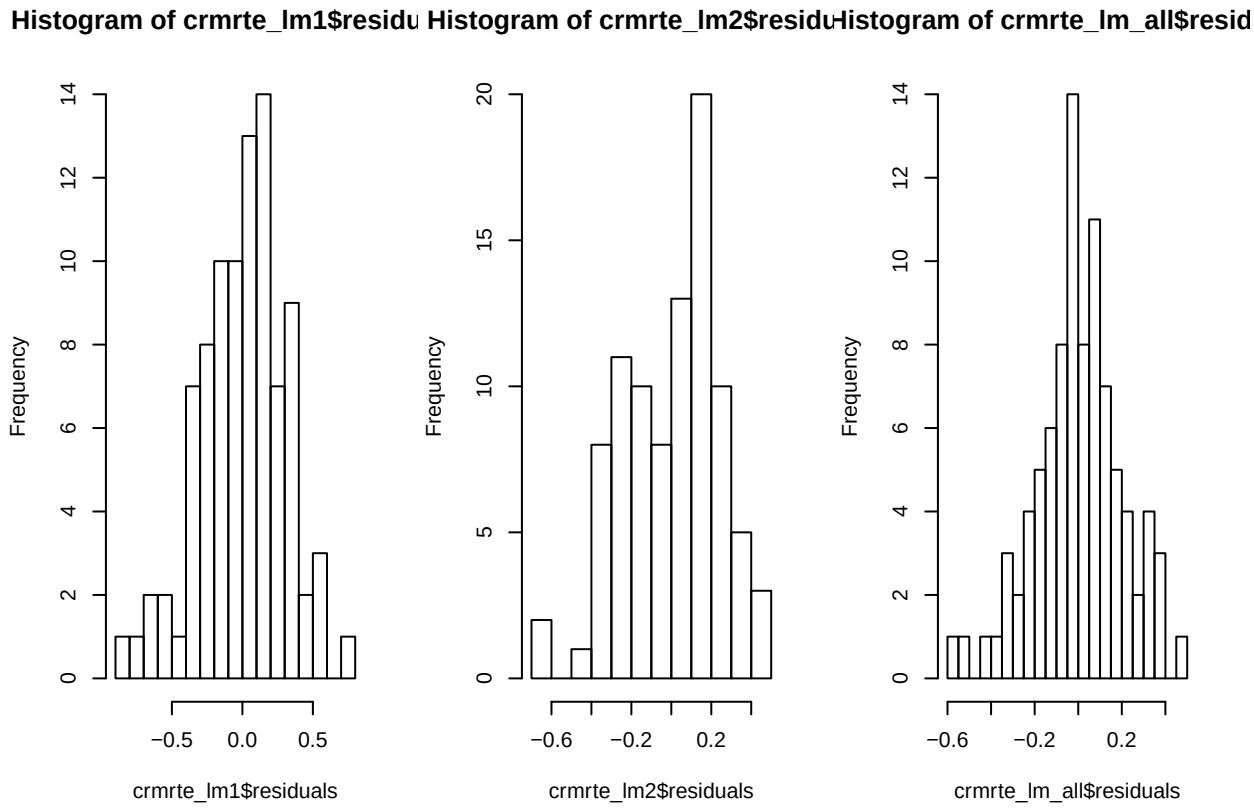

```
par(mfrow=c(1,3))
plot(crmrte_lm1, which = 3)
plot(crmrte_lm2, which = 3)
plot(crmrte_lm_all, which = 3)
```




```
par(mfrow=c(1,3))
plot(crmrte_lm1, which = 5)
plot(crmrte_lm2, which = 5)
plot(crmrte_lm_all, which = 5)
```




```
par(mfrow=c(1,3))
hist(crmrte_lm1$residuals, breaks = 15)
hist(crmrte_lm2$residuals, breaks = 15)
hist(crmrte_lm_all$residuals, breaks = 15)
```




```
se.crmrte_lm1 <- sqrt(diag(vcovHC(crmrte_lm1)))
se.crmrte_lm2 <- sqrt(diag(vcovHC(crmrte_lm2)))
se.crmrte_lm_all <- sqrt(diag(vcovHC(crmrte_lm_all)))
stargazer(crmrte_lm1, crmrte_lm2,crmrte_lm_all,
          type = "text", omit.stat = "f",
          se = list(se.crmrte_lm1, se.crmrte_lm2, se.crmrte_lm_all),
          star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## ==========================================================================
##                                    Dependent variable:
##                      ---------------------------------------------------
##                                          log(crmrte)
##                            (1)              (2)              (3)
## --------------------------------------------------------------------------
## prbarr_imp                                                   -1.705***
##                                                               (0.308)
##
## density                  0.150***         0.103***          0.109*
##                           (0.026)          (0.026)          (0.052)
##
## taxpc.ln                                                     -0.046
##                                                               (0.234)
##
## west                     -0.356***         -3.678*          -0.230
##                           (0.071)          (1.780)          (0.122)
```

```
##
## prbarr                -1.269**        -1.404*
##                        (0.394)         (0.559)
##
## central                                              -0.174*
##                                                      (0.078)
##
## urban                                                -0.094
##                                                      (0.220)
##
## prbconv               -0.552***       -0.566***      -0.666***
##                        (0.132)         (0.157)        (0.124)
##
## prbpris                                              -0.135
##                                                      (0.432)
##
## avgsen                                               -0.026
##                                                      (0.016)
##
## polpc_imp.ln                          0.655***       0.557**
##                                        (0.137)       (0.208)
##
## pctmin80                               0.008**       0.008**
##                                        (0.003)       (0.003)
##
## west:polpc_imp.ln                     -0.539*
##                                        (0.274)
##
## wcon.ln                                               0.343
##                                                      (0.243)
##
## wtuc.ln                                               0.163
##                                                      (0.285)
##
## wtrd.ln                                               0.250
##                                                      (0.306)
##
## wfir.ln                                              -0.163
##                                                      (0.339)
##
## wser_imp.ln                                          -0.498
##                                                      (0.307)
##
## wmfg.ln                                              -0.039
##                                                      (0.164)
##
## wfed.ln                                               0.694
##                                                      (0.433)
##
## wsta.ln                                              -0.332
##                                                      (0.326)
##
## wloc.ln                                               0.047
##                                                      (0.685)
```

```
##
## mix                                                    -0.656
##                                                        (0.559)
##
## pctymle                                                 2.089
##                                                        (1.307)
##
## Constant               -2.990***         1.129         -1.800
##                         (0.198)         (0.906)        (4.102)
##
## ------------------------------------------------------------------------
## Observations              91               91              91
## R2                       0.691            0.811           0.864
## Adjusted R2              0.677            0.795           0.819
## Residual Std. Error 0.310 (df = 86) 0.247 (df = 83) 0.232 (df = 68)
## ========================================================================
## Note:                                  *p<0.05; **p<0.01; ***p<0.001
```