Lab3_YZ_EDA

Yulia Zamriy

March 18, 2018

```r
#install.packages("kableExtra")
#install.packages("viridisLite")
#install.packages("viridis")
#install.packages("Hmisc")
library(knitr)
library(kableExtra)
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```
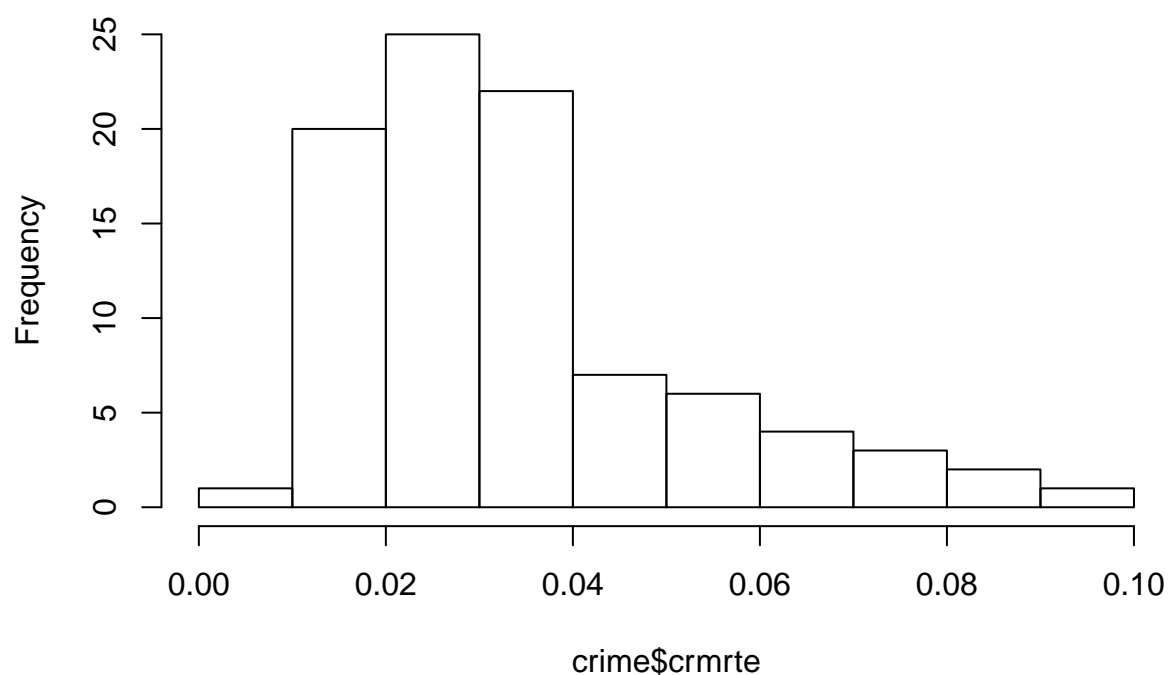
```r
library(reshape2)
library(ggplot2)
```

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

```r
summary(crime$crmrte)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.005533 0.020927 0.029986 0.033400 0.039642 0.098966
```

```r
hist(crime$crmrte)
```

Histogram of crime$crmrte


```
crime$prbconv <- as.numeric(crime$prbconv)
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$prbconv)
```

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

```
summary(crime$prbpris)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.1500  0.3648  0.4234  0.4108  0.4568  0.6000
```

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

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

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

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

```
crime$exclude <- 0
crime[crime$prbarr > 1,]$exclude <- 1
crime[crime$prbconv > 1,]$exclude <- 1
table(crime$exclude)
```

```
## 
##  0  1
## 81 10
```

```r
summary(crime$avgsen)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.380   7.340   9.100   9.647  11.420  20.700
```

```r
summary(crime$polpc)
```

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

```r
summary(crime$density)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00002 0.54741 0.96226 1.42884 1.56824 8.82765
```

```r
summary(crime$taxpc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.69   30.66   34.87   38.06   40.95  119.76
```

```r
mean(crime$west)
```

```
## [1] 0.2527473
```

```r
mean(crime$central)
```

```
## [1] 0.3736264
```

```r
mean(crime$urban)
```

```
## [1] 0.08791209
```

```r
summary(crime$pctmin80)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.284   9.845  24.312  25.495  38.142  64.348
```

```r
summary(crime$wcon)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   193.6   250.8   281.4   285.4   314.8   436.8
```

```r
summary(crime$wtuc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   187.6   374.6   406.5   411.7   443.4   613.2
```

```r
summary(crime$wtrd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   154.2   190.9   203.0   211.6   225.1   354.7
```

```r
summary(crime$wfir)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   170.9   286.5   317.3   322.1   345.4   509.5
```

```r
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$wmfg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   157.4   288.9   320.2   335.6   359.6   646.9
summary(crime$wfed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   326.1   400.2   449.8   442.9   478.0   598.0
summary(crime$wsta)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   258.3   329.3   357.7   357.5   382.6   499.6
summary(crime$wloc)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   239.2   297.3   308.1   312.7   329.2   388.1
summary(crime$mix)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01961 0.08073 0.10186 0.12884 0.15175 0.46512
summary(crime$pctymle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06216 0.07443 0.07771 0.08396 0.08350 0.24871
crime[crime$wser > 2000,]$exclude <- 1
crime_sub <- subset(crime, exclude == 0)
crime_sub$exclude <- NULL
```

```r
# Prepare a .RData for easier sharing and usage.
ind_variables <- c(
  'prbarr', 'prbconv', 'prbpris', 'avgsen',
  'polpc', 'density', 'taxpc', 'west', 'central', 'urban', 'pctmin80', 'wcon',
  'wtuc', 'wtrd', 'wfir', 'wser', 'wmfg', 'wfed', 'wsta', 'wloc', 'mix',
  'pctymle'
)
var_labels <- c(
  '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("Negative" , "Negative", "Negative", "Negative",
            "Negative", "Positive", "Negative",
            "Unclear", "Unclear", "Unclear", "Unclear",
            "Negative","Negative","Negative",
```

```
              "Negative", "Negative", "Negative", "Negative",
              "Negative", "Negative", "Unclear","Positive")
control <- c("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_sub)[3,-c(1,2,3)],2)
```

```
## Warning in cor(crime_sub): the standard deviation is zero
```

```
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?
prbarr	probability of arrest	Negative	-0.51	Yes
prbconv	probability of conviction	Negative	-0.31	Yes
prbpris	probability of prison sentence	Negative	0.05	Yes
avgsen	avg. sentence, days	Negative	0.14	Yes
polpc	police per capita	Negative	0.56	Yes
density	people per sq. mile	Positive	0.72	No
taxpc	tax revenue per capita	Negative	0.48	Yes
west	=1 if in western N.C.	Unclear	-0.34	No
central	=1 if in central N.C.	Unclear	0.16	No
urban	=1 if in SMSA	Unclear	0.62	No
pctmin80	perc. minority, 1980	Unclear	0.18	No
wcon	weekly wage, construction	Negative	0.36	Yes
wtuc	wkly wge, trns, util, commun	Negative	0.26	Yes
wtrd	wkly wge, whlesle, retail trade	Negative	0.39	Yes
wfir	wkly wge, fin, ins, real est	Negative	0.34	Yes
wser	wkly wge, service industry	Negative	0.33	Yes
wmfg	wkly wge, manufacturing	Negative	0.39	Yes
wfed	wkly wge, fed employees	Negative	0.47	Yes
wsta	wkly wge, state employees	Negative	0.19	Yes
wloc	wkly wge, local gov emps	Negative	0.42	Yes
mix	offense mix: face-to-face/other	Unclear	-0.23	No
pctymle	percent young male	Positive	0.29	No

```r
crime_sub$othergeo <- ifelse(crime_sub$west == 0 & crime_sub$central == 0, 1, 0)
mean(crime_sub$west)
```

```
## [1] 0.2345679
```

```r
mean(crime_sub$central)
```

```
## [1] 0.3950617
```

```r
mean(crime_sub$othergeo)
```

```
## [1] 0.382716
```

```r
crime_cor <- cor(crime_sub)[3,-c(1,2,3)]
```

```
## Warning in cor(crime_sub): the standard deviation is zero
```

```r
crime_cor <- crime_cor[order(crime_cor)]
```
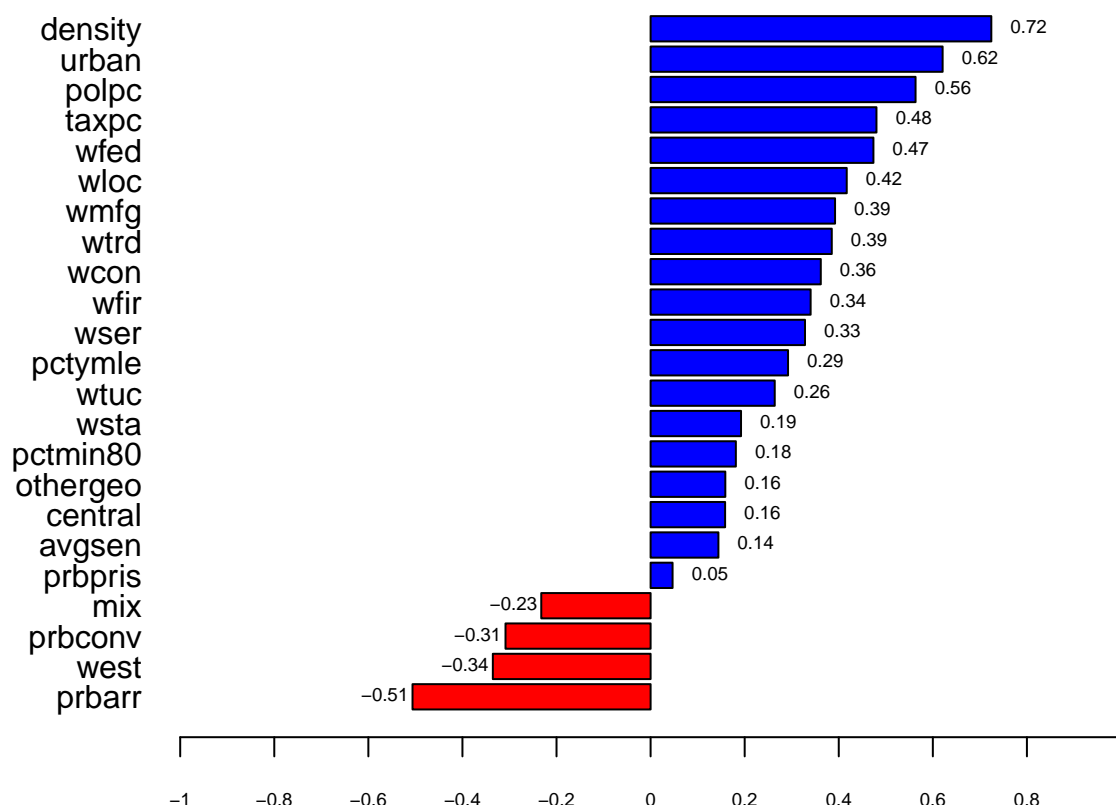
```r
crime_cor_lab <- ifelse(crime_cor < 0, crime_cor-0.15, crime_cor)

par(mar = c(2,8,1,0))
b <- barplot(crime_cor,
        col = ifelse(crime_cor < 0, "red","blue"),
        horiz = TRUE,
        las = 1,
        xaxt = "n",
        xlim = c(-1,1),
        main = "Correlation of Crime Rate with Other Variables")
text(x = crime_cor_lab,
     y = b,
     label = round(crime_cor,2),
     pos = 4,
     cex = 0.6)
axis(1,
     at = seq(-1,1, by = 0.2),
     labels = seq(-1,1, by = 0.2),
     cex.axis = 0.6)
```

Correlation of Crime Rate with Other Variables


```r
cor_mat <- round(cor(crime_sub[-c(1:3)]),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
```

```r
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()
```
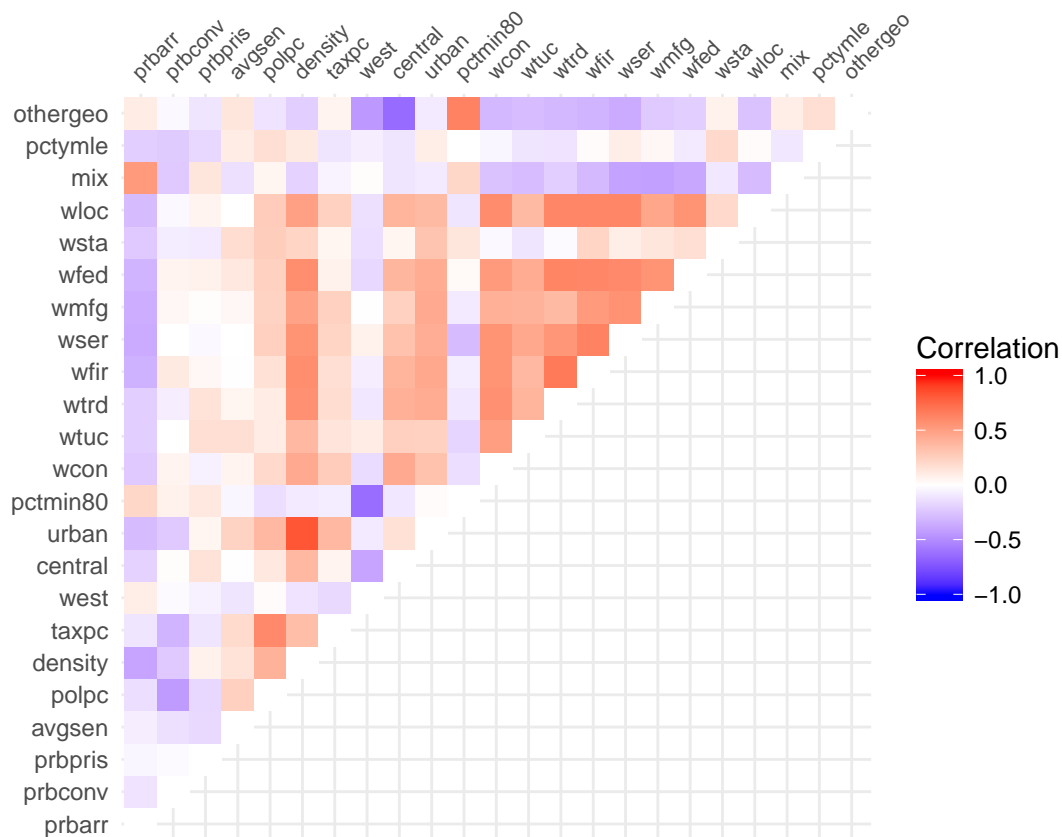
```r
crime_sub$region <- ifelse(crime_sub$west == 1, "west",
                           ifelse(crime_sub$central == 1, "central", "other"))
```

```r
crmrte_reg <- aggregate(crmrte ~ region, data = crime_sub, mean)
density_reg <- aggregate(density ~ region, data = crime_sub, mean)
urban_reg <- aggregate(urban ~ region, data = crime_sub, mean)
polpc_reg <- aggregate(polpc ~ region, data = crime_sub, mean)
cbind(crmrte_reg, density_reg[2], polpc_reg[2])
```

```
##    region     crmrte   density        polpc
## 1 central 0.03854043 2.133793 0.001679178
## 2   other 0.03914362 1.097235 0.001531251
## 3    west 0.02401637 1.158740 0.001626667
```

```r
ggplot(crime_sub, aes(density, crmrte)) +
  geom_point() +
  facet_grid(region~.) +
  geom_smooth(method = "lm", se = FALSE)
```
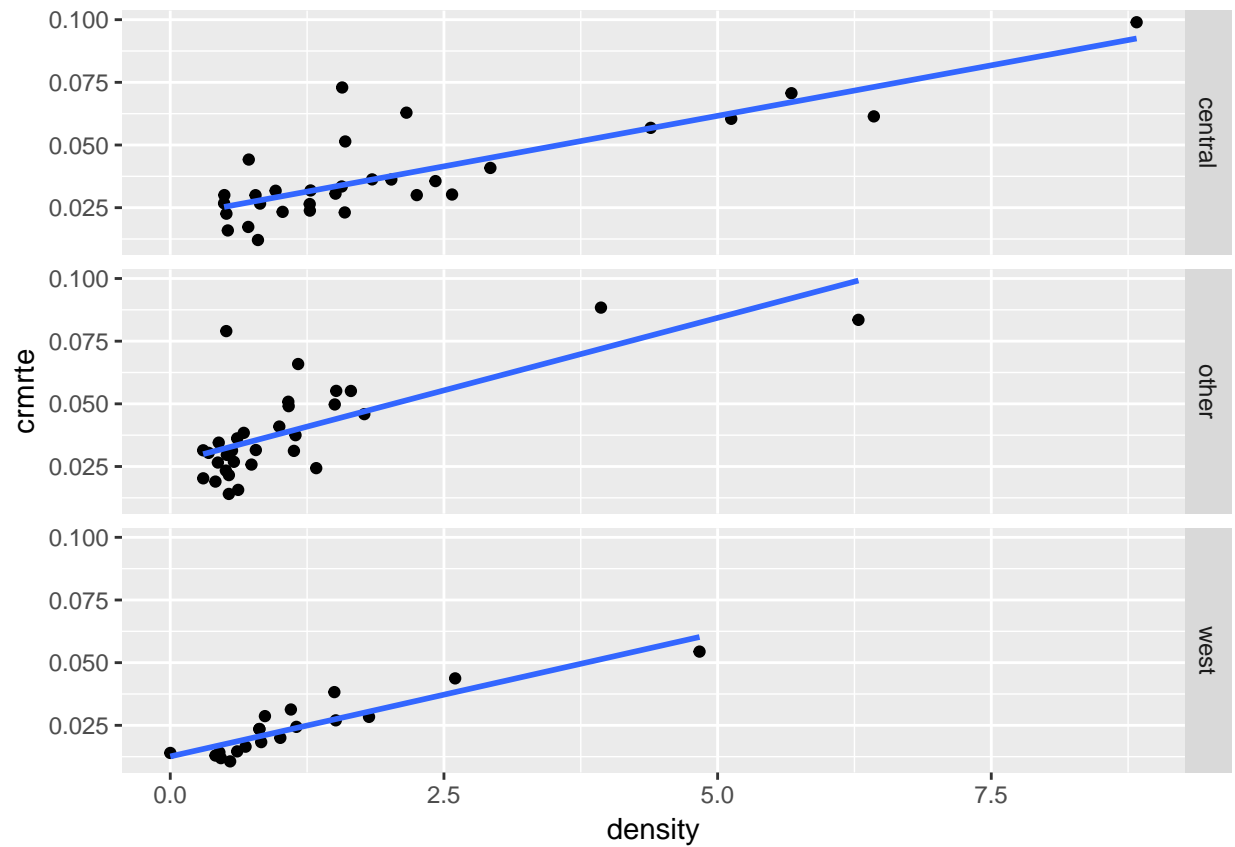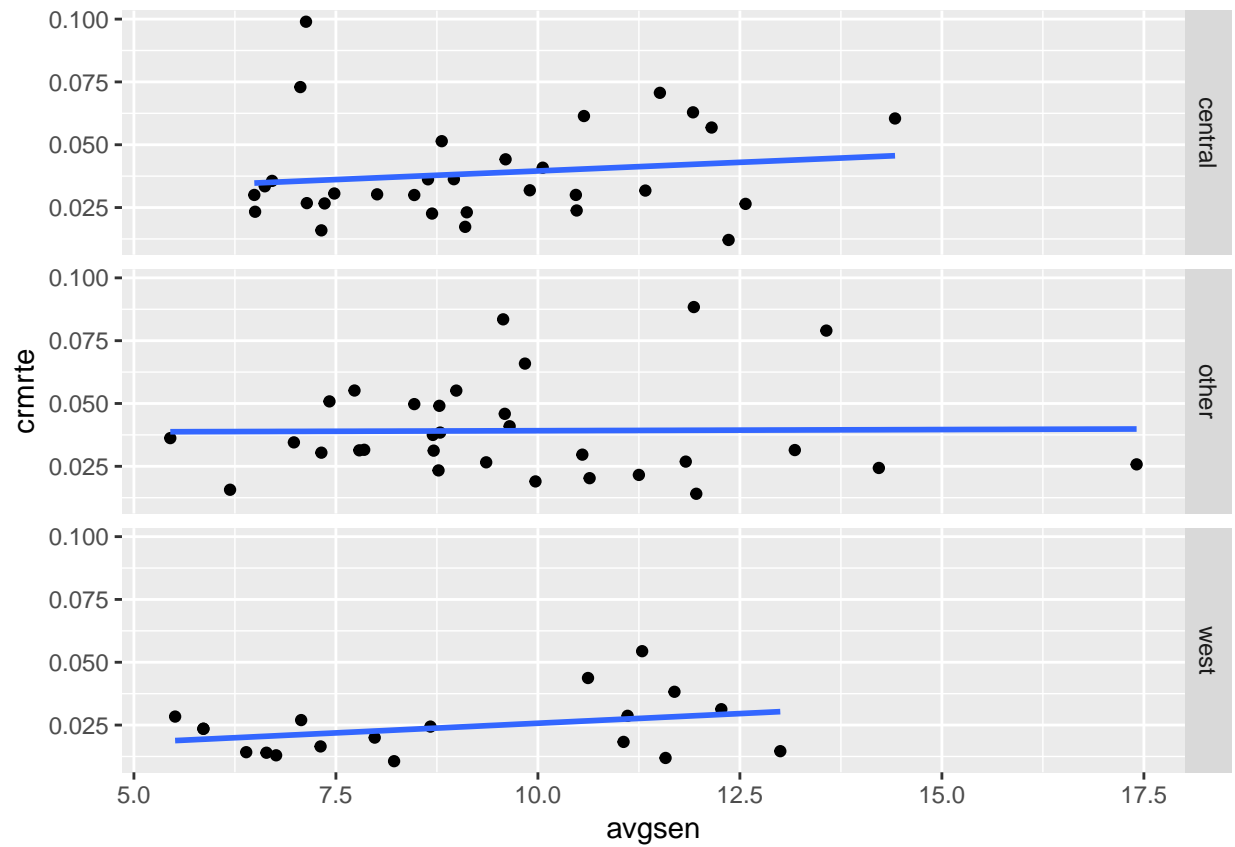
```
ggplot(crime_sub, aes(polpc, crmrte)) +
  geom_point() +
  facet_grid(region~.) +
  geom_smooth(method = "lm", se = FALSE)
```
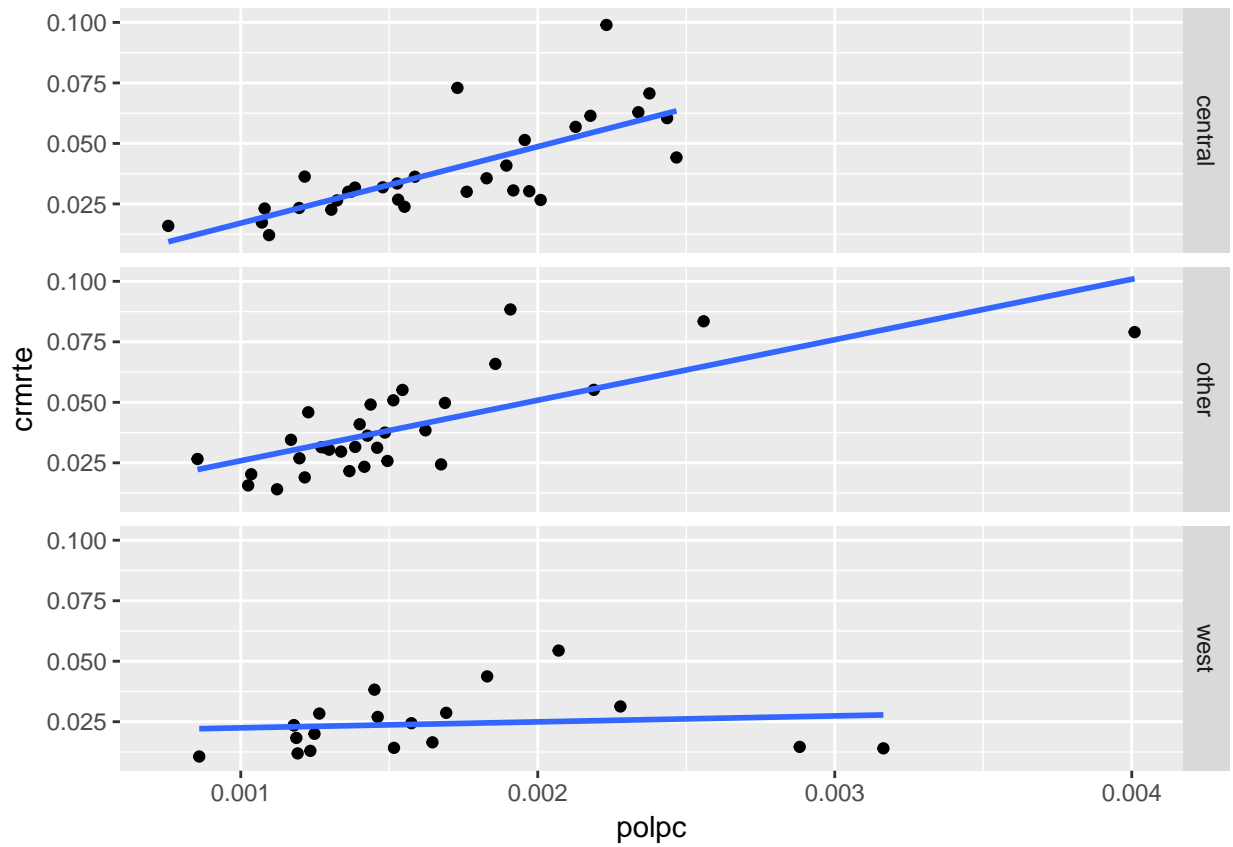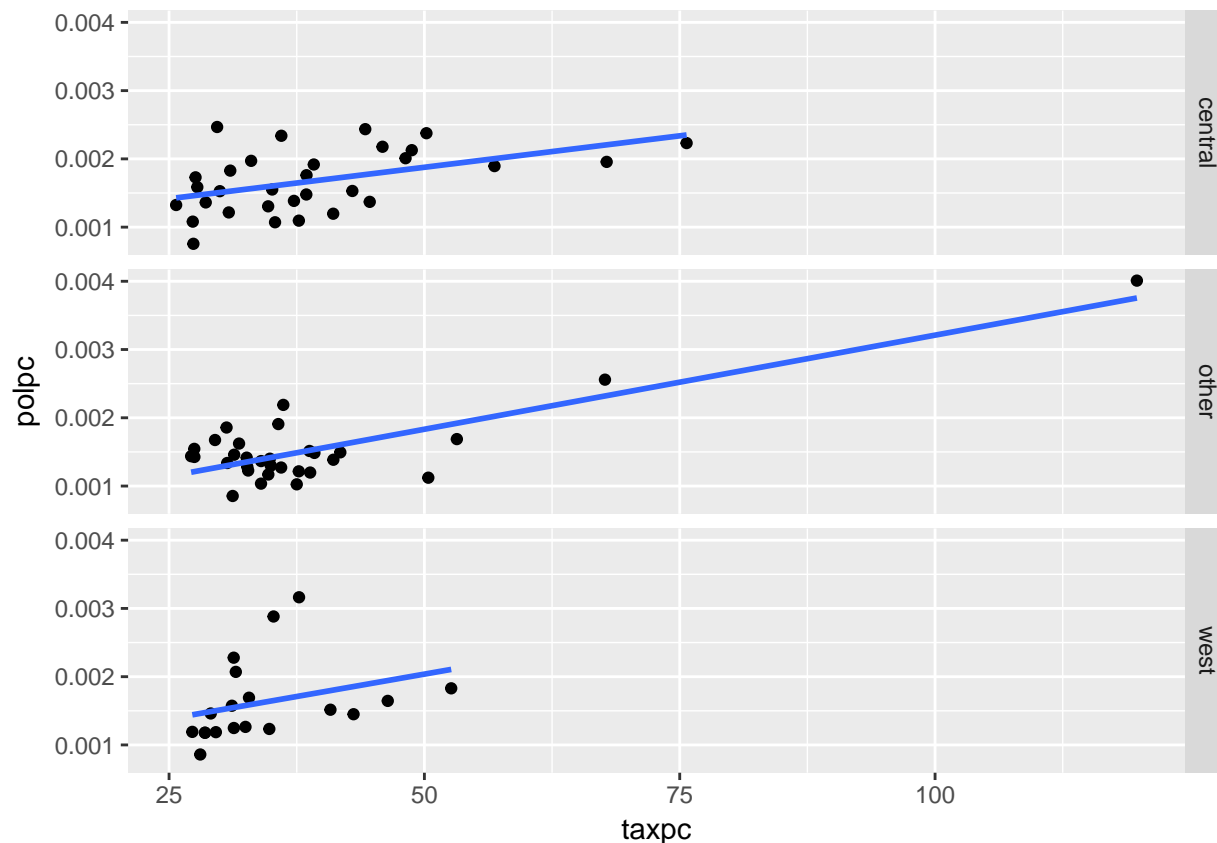
```
ggplot(crime_sub, aes(taxpc, crmrte)) +
  geom_point() +
  facet_grid(region~.) +
  geom_smooth(method = "lm", se = FALSE)
```
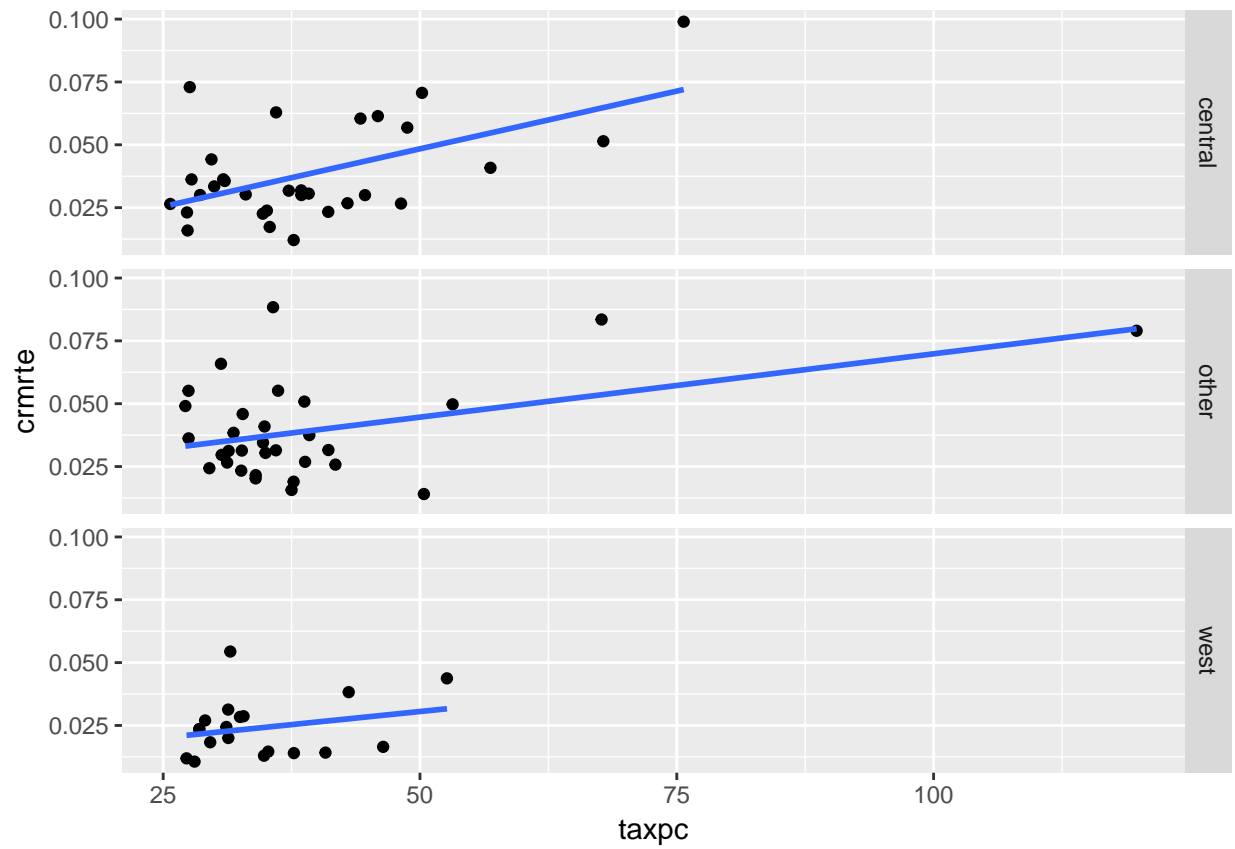
```
ggplot(crime_sub, aes(avgsen, crmrte)) +
  geom_point() +
  facet_grid(region~.) +
  geom_smooth(method = "lm", se = FALSE)
```

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

```
showScatterPlotMatrices <- function(dependent, independents, data) {
  make.formula <- function(independent) {
    f <- paste(c(dependent, independent), collapse = ' ~ ')
    return(as.formula(f))
  }
  formulas <- mapply(make.formula, independents)
  plot.relation <- function(f) {
    model <- lm(f, data = data)
    r2 <- round(summary(model)$r.squared,2)
    description <-
      as.character(desc$Explanation[as.character(desc$`Explanatory Variables`) == as.character(f[3])])
    test <- paste(deparse(substitute(data)),names(model$coefficients[2]),sep="$")
    xval <-
      min(eval(parse(text = paste(deparse(substitute(data)),names(model$coefficients[2]),sep="$"))))
    par(mar = c(4,4,4,4))
    plot(f,
         data = data,
         main = paste(c('Effect of', description, 'on crime rate')))
    abline(model, col = "darkblue")
    text(x = xval*1.1, y = 0.09, bquote(R^2 == .(r2)), col = "darkblue")
  }
  mapply(plot.relation, formulas)
}
wage_vars <- colnames(crime_sub)[grep("w", colnames(crime_sub))][-1]
plots <- showScatterPlotMatrices('crmrte', wage_vars, crime_sub)
```
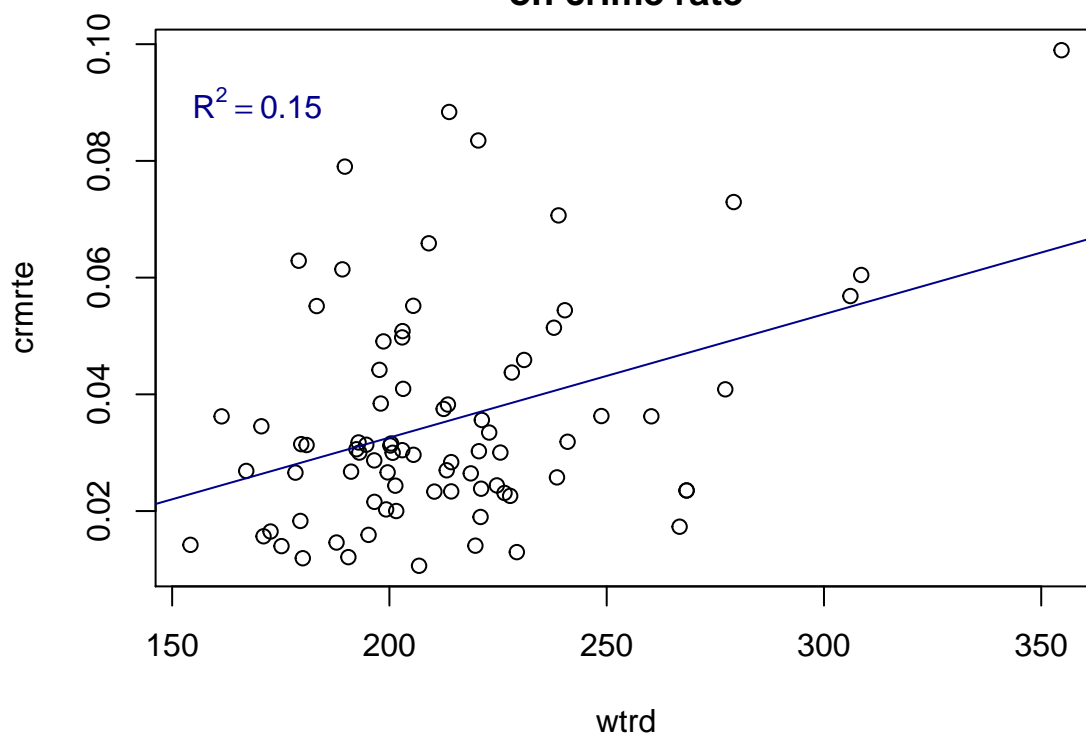
**Effect of
weekly wage, construction
on crime rate**

$R^2 = 0.13$
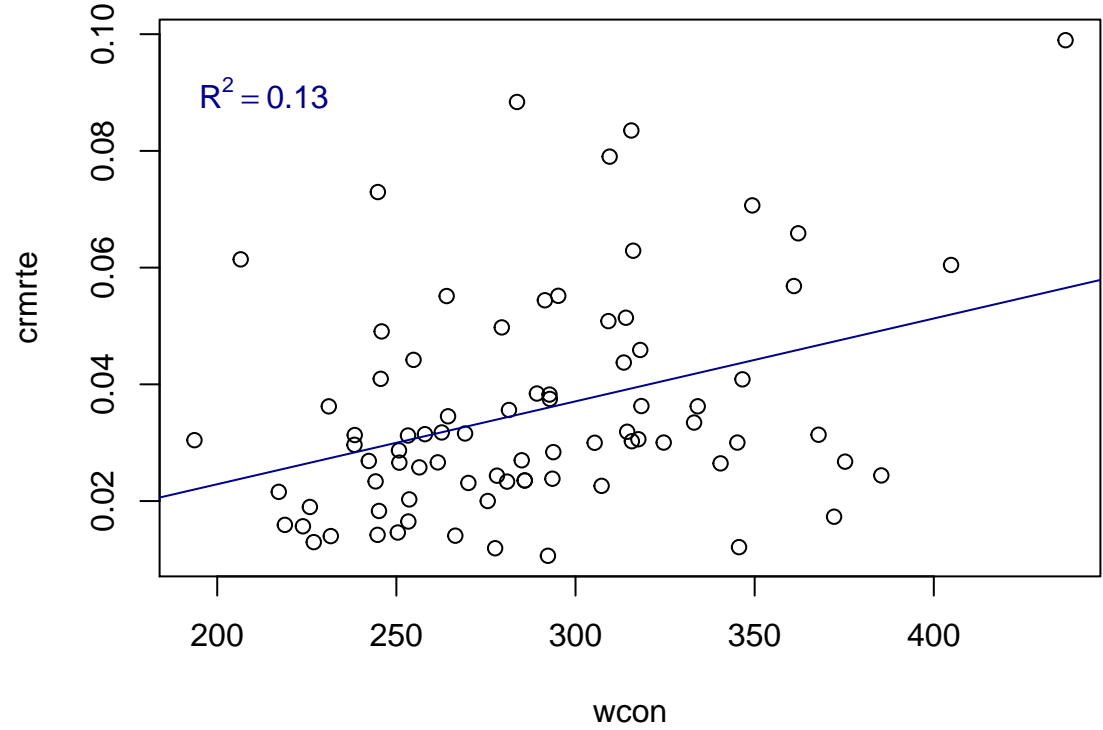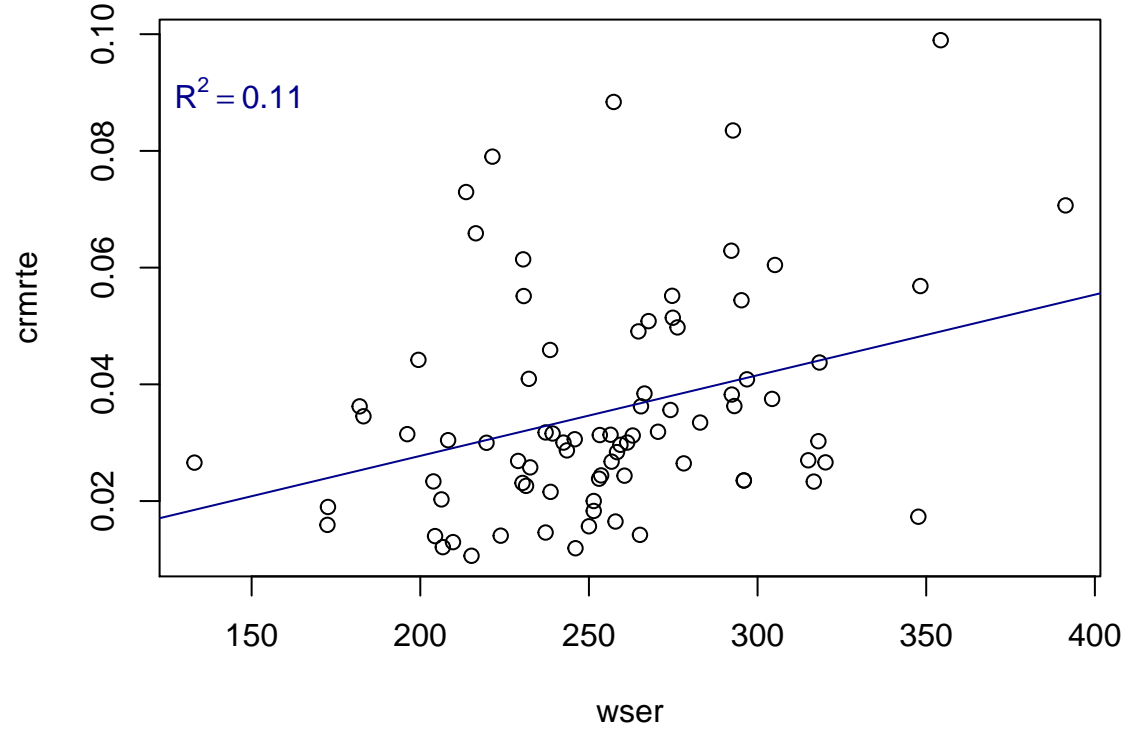
**Effect of
wkly wge, trns, util, commun
on crime rate**

$R^2 = 0.07$

crmrte

wtuc
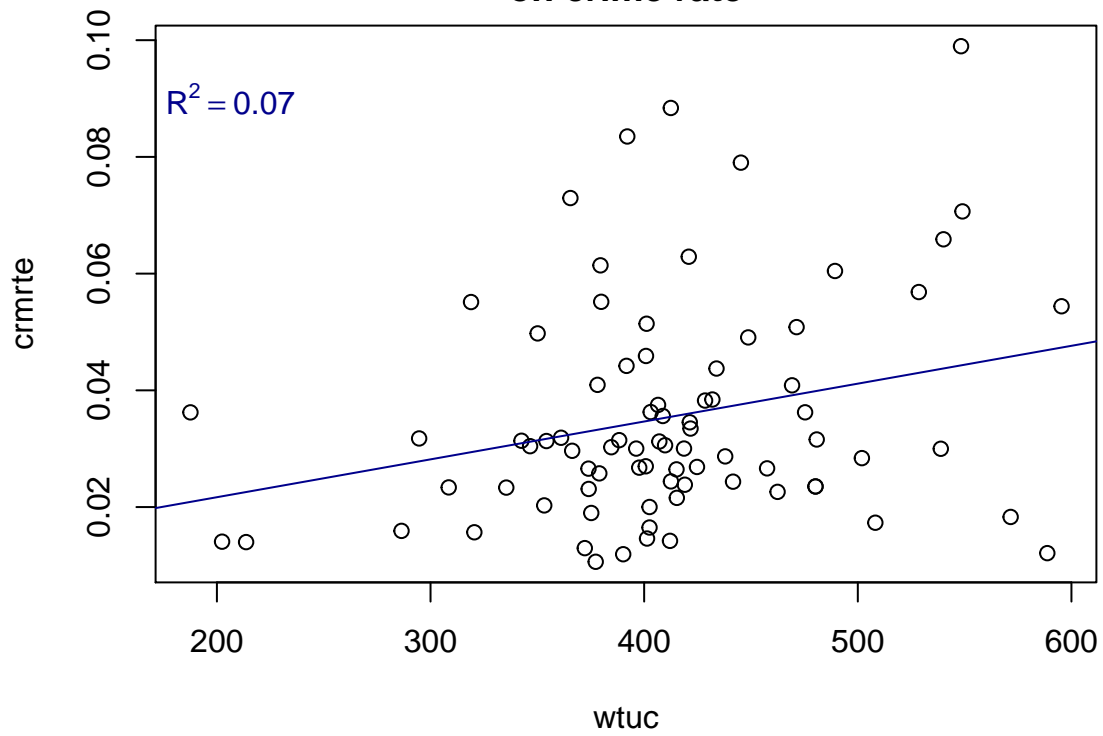
**Effect of
wkly wge, whlesle, retail trade
on crime rate**

**Effect of
wkly wge, fin, ins, real est
on crime rate**

$R^2 = 0.12$

crmrte
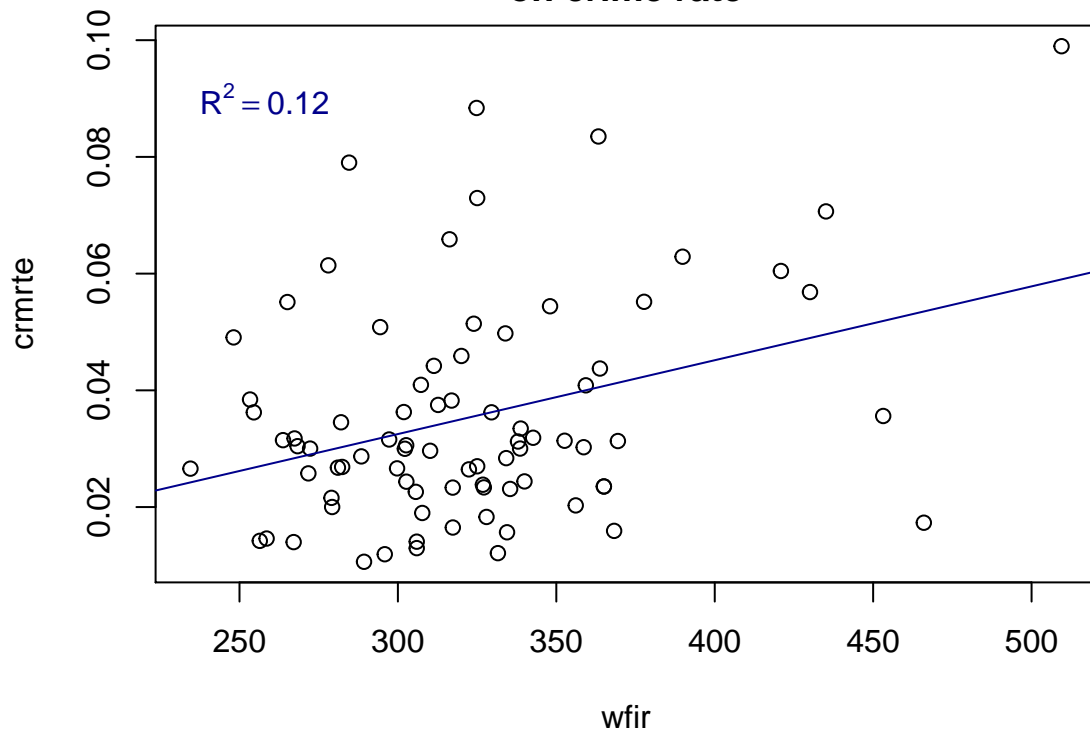
wfir

**Effect of
wkly wge, service industry
on crime rate**

$R^2 = 0.11$

crmrte

wser

**Effect of
wkly wge, manufacturing
on crime rate**

$R^2 = 0.15$

crmrte
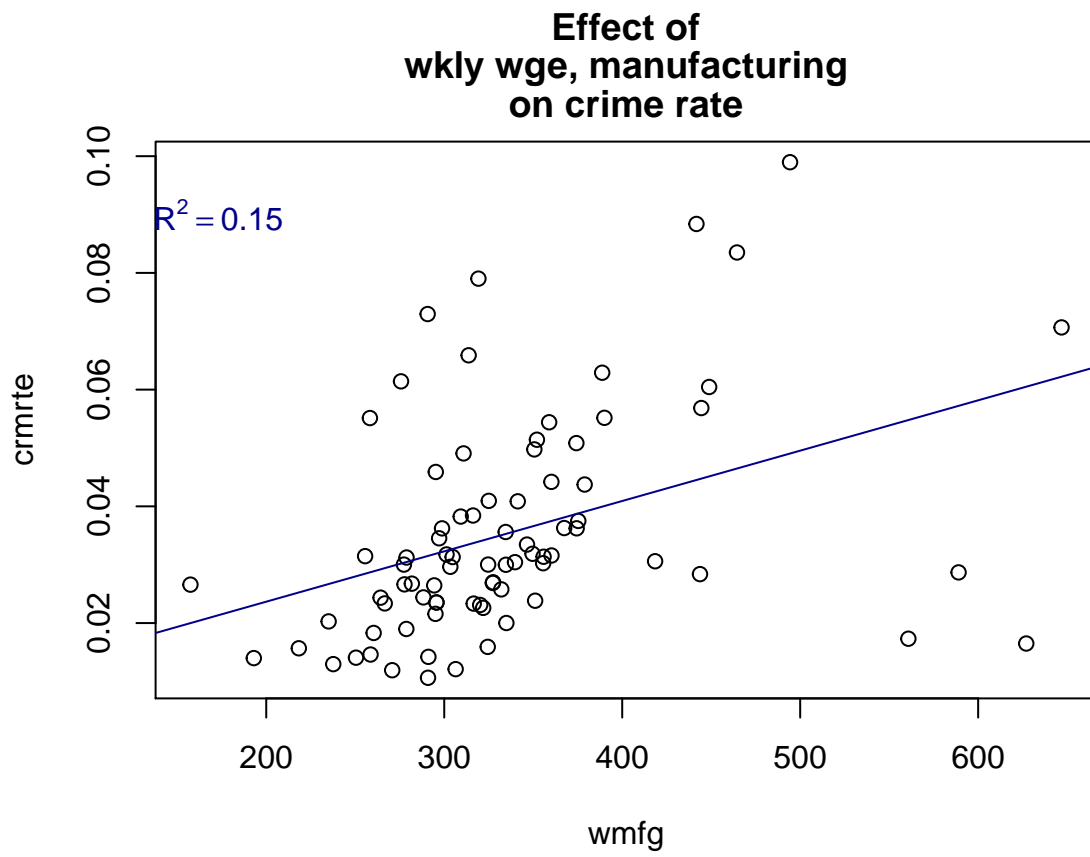
wmfg

Effect of
wkly wge, fed employees
on crime rate

$R^2 = 0.22$

crmrte
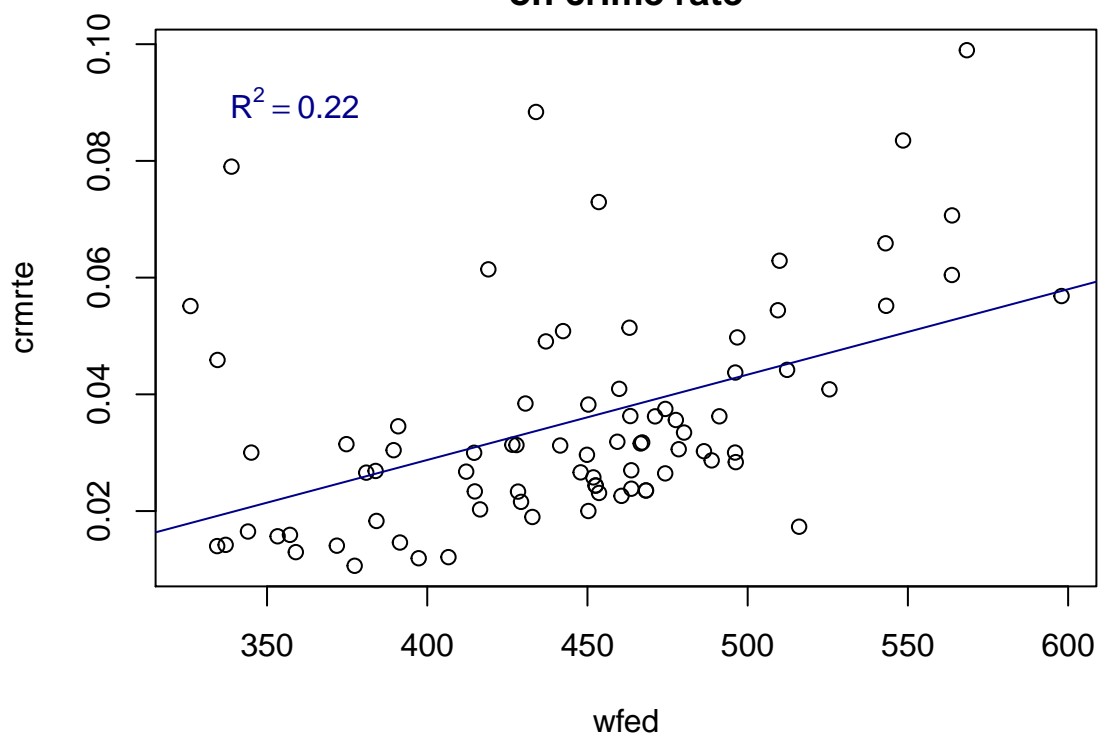
wfed

**Effect of
wkly wge, state employees
on crime rate**

$R^2 = 0.04$

crmrte

wsta

Effect of
wkly wge, local gov emps
## on crime rate

$R^2 = 0.17$
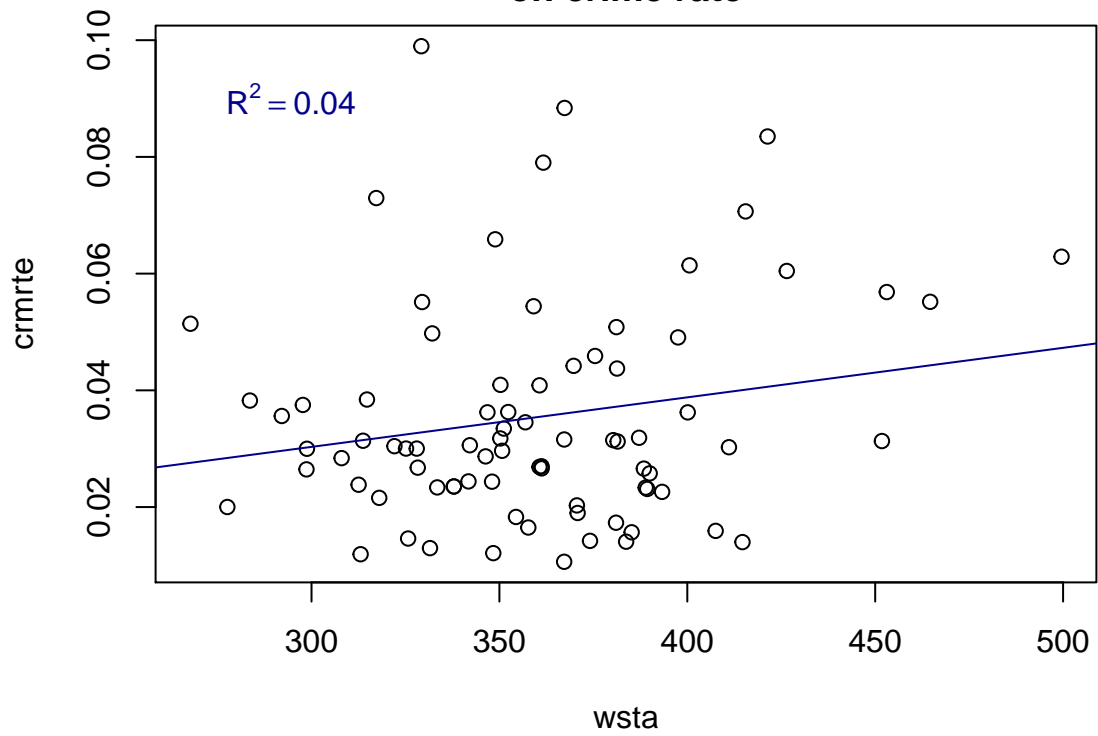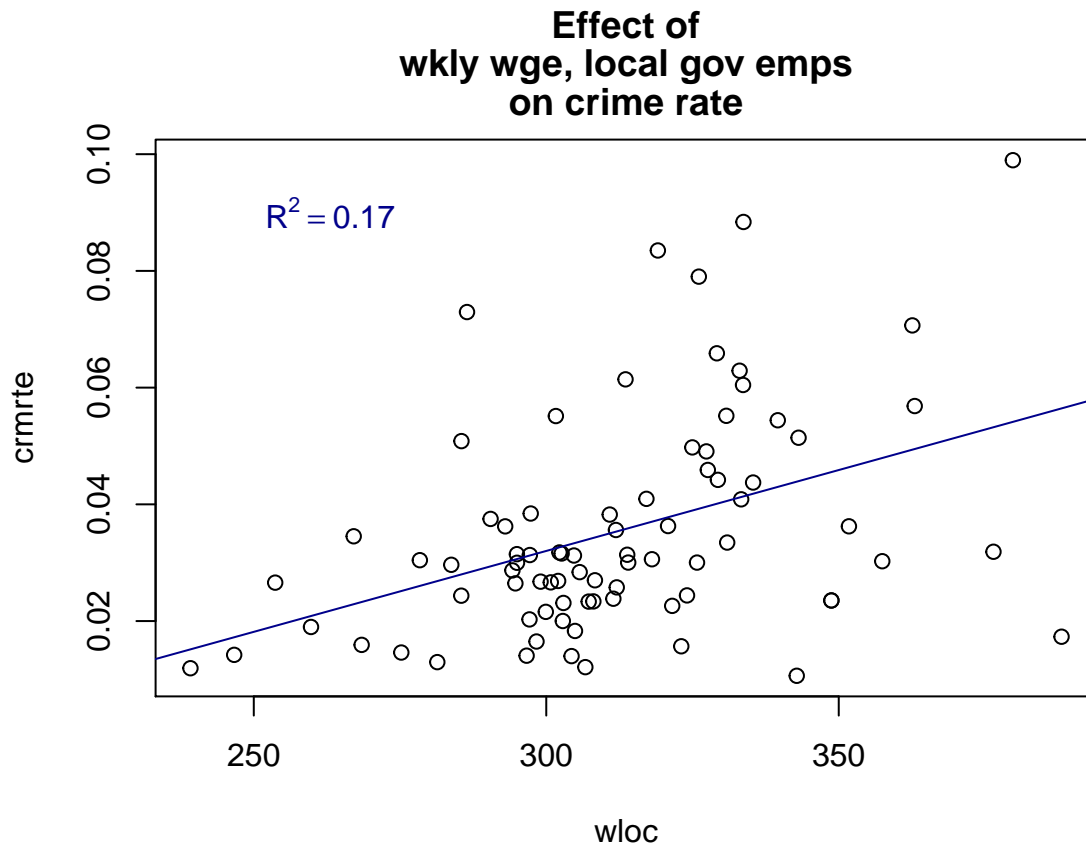
crmrte

wloc

```r
vars_wages <- c("wcon", "wtuc", "wtrd", "wfir",
                "wser", "wmfg", "wfed", "wsta", "wloc")
crime_sub[paste(vars_wages,"ln",sep = ".")] <- log(crime_sub[vars_wages])
crime_sub[paste(vars_wages,"sq",sep = ".")] <- (crime_sub[vars_wages])**2

crime_sub_w1 <- subset(crime_sub, select = c("crmrte", vars_wages))
crime_sub_w2 <- subset(crime_sub, select = c("crmrte", paste(vars_wages,"ln",sep = ".")))
crime_sub_w3 <- subset(crime_sub, select = c("crmrte", paste(vars_wages,"sq",sep = ".")))
wages_cor <- cbind(round(cor(crime_sub_w1)[1,-1],2),
                   round(cor(crime_sub_w2)[1,-1],2),
                   round(cor(crime_sub_w3)[1,-1],2))
colnames(wages_cor) <- c("As is", "Log", "Squared")
wages_cor <- melt(wages_cor)
colnames(wages_cor) <- c("Wages", "Transformation", "Correlation")

ggplot(data = wages_cor, aes(Wages, Correlation, fill = Transformation)) +
  geom_col(position = "dodge")
```
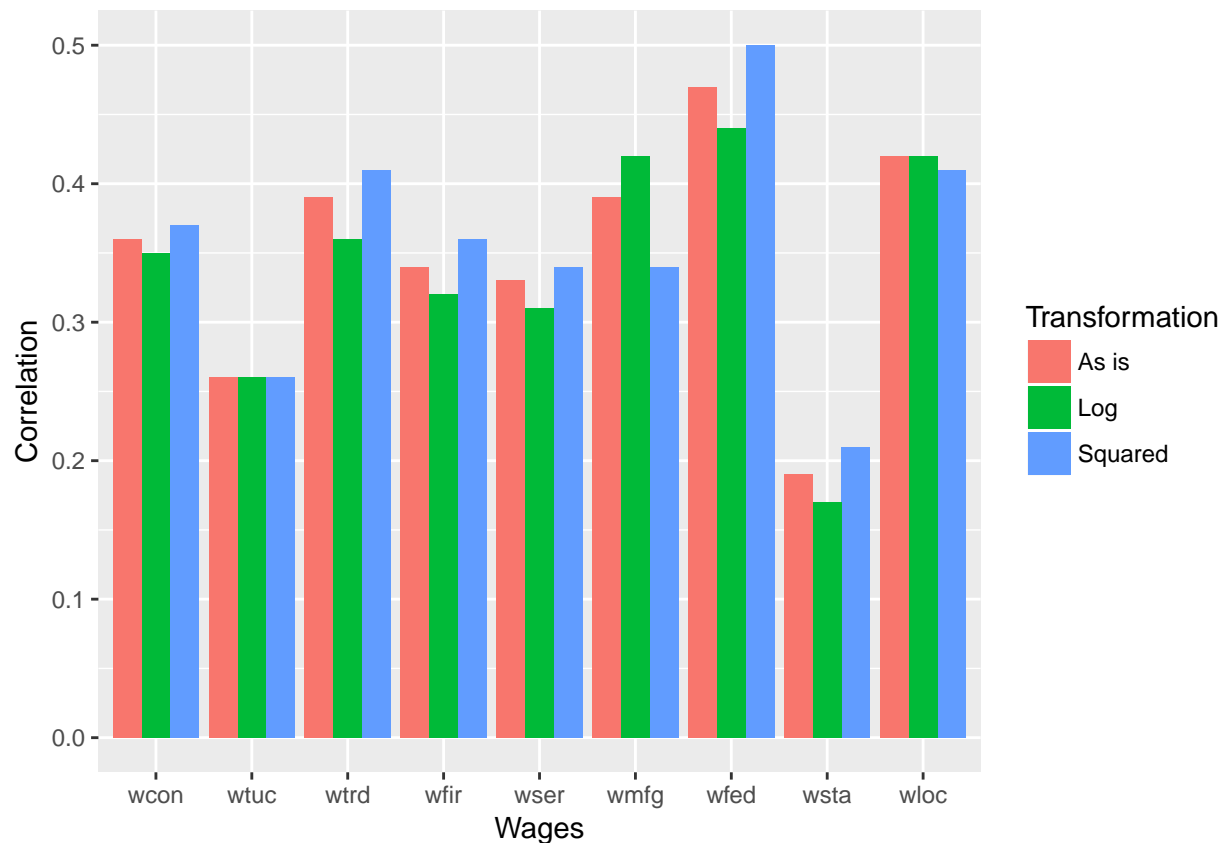
```
vars_other <- c("prbarr", "prbconv", "prbpris", "avgsen", "polpc", "density", "taxpc",
                "pctmin80", "mix", "pctymle")
crime_sub[paste(vars_other,"sq",sep = ".")] <- (crime_sub[vars_other])**2

crime_sub_o1 <- subset(crime_sub, select = c("crmrte", vars_other))
crime_sub_o2 <- subset(crime_sub, select = c("crmrte", paste(vars_other,"sq",sep = ".")))
others_cor <- cbind(round(cor(crime_sub_o1)[1,-1],2),
                    round(cor(crime_sub_o2)[1,-1],2))
colnames(others_cor) <- c("As is", "Squared")
others_cor <- melt(others_cor)
colnames(others_cor) <- c("OtherVars", "Transformation", "Correlation")

ggplot(data = others_cor, aes(OtherVars, Correlation, fill = Transformation)) +
  geom_col(position = "dodge")
```
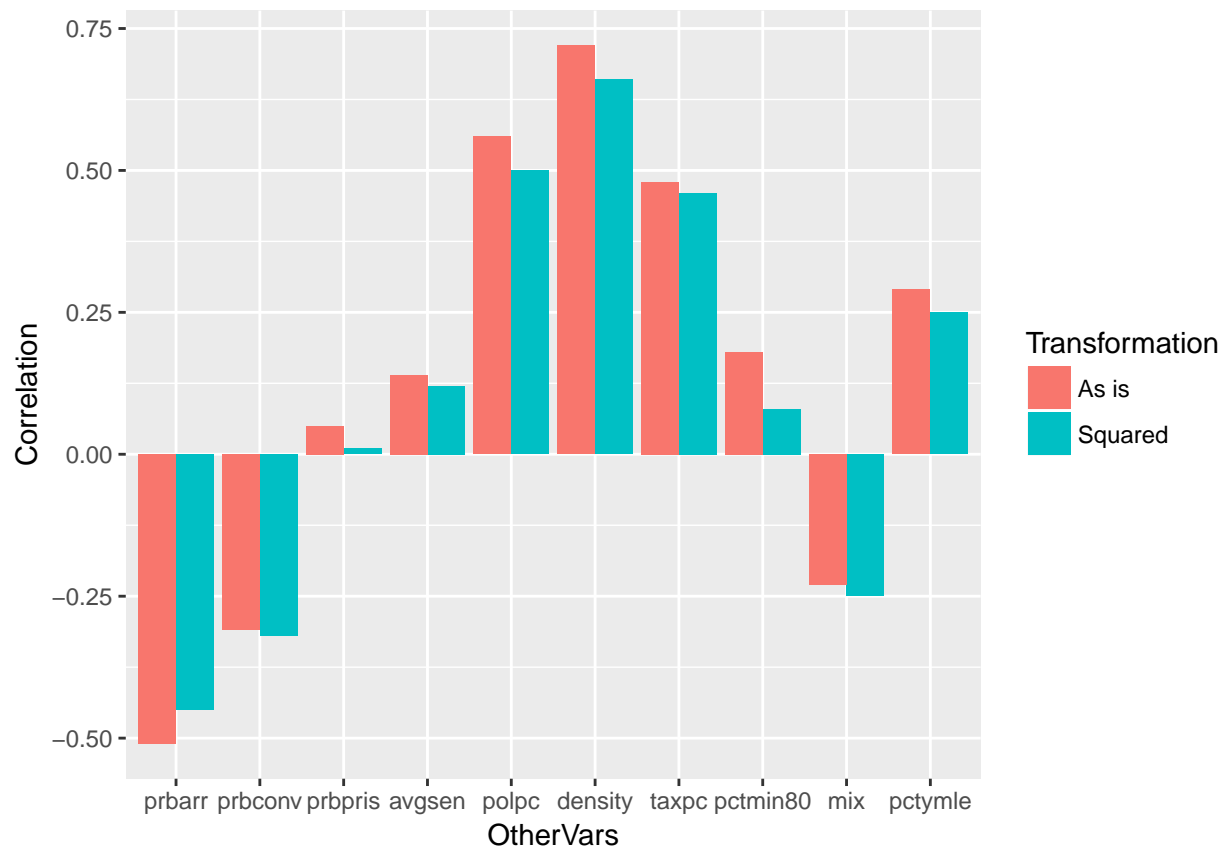
```
ind_vars1 <- c("density", "prbarr", "polpc", "taxpc","pctymle","pctmin80")
crmrte_formula1 <- as.formula(paste("crmrte ~", paste(ind_vars1, collapse = "+"), sep = ""))
crmrte_lm1 <- lm(crmrte_formula1, data = crime_sub)
summary(crmrte_lm1)
```

```
##
## Call:
## lm(formula = crmrte_formula1, data = crime_sub)
##
## Residuals:
##        Min        1Q     Median        3Q        Max
## -0.0150184 -0.0059256  0.0000288  0.0047033  0.0265733
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.070e-03  6.067e-03  -0.176 0.860483
## density      5.391e-03  6.998e-04   7.703 4.71e-11 ***
## prbarr      -5.382e-02  9.688e-03  -5.555 4.15e-07 ***
## polpc        8.408e+00  2.344e+00   3.587 0.000598 ***
## taxpc        2.764e-04  9.181e-05   3.011 0.003561 **
## pctymle      1.256e-01  4.194e-02   2.995 0.003733 **
## pctmin80     3.723e-04  5.717e-05   6.513 7.83e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008341 on 74 degrees of freedom
```

```
## Multiple R-squared:  0.8189, Adjusted R-squared:  0.8042
## F-statistic: 55.76 on 6 and 74 DF,  p-value: < 2.2e-16
```

```r
ind_vars_all <- c("prbarr", "prbconv", "prbpris", "avgsen", "polpc", "density", "taxpc",
                  "west", "central", "urban", "pctmin80", "wcon", "wtuc", "wtrd", "wfir",
                  "wser", "wmfg", "wfed", "wsta", "wloc", "mix", "pctymle")
                  # paste(vars_wages,"ln",sep = "."),
                  # paste(vars_wages,"sq",sep = "."),
                  # paste(vars_other,"sq",sep = "."))

crmrte_formula_all <- as.formula(paste("crmrte ~", paste(ind_vars_all, collapse = "+"), sep = ""))
# crmrte_lm0 <- lm(crmrte ~ 1,
#                  data = crime_sub)
crmrte_lm_all <- lm(crmrte_formula_all,
                    data = crime_sub)
crmrte_lm_step <- step(crmrte_lm1, scope=list(lower=crmrte_lm1, upper=crmrte_lm_all),
                       direction="both",
                       trace = FALSE)
summary(crmrte_lm_step)
```

```
##
## Call:
## lm(formula = crmrte ~ density + prbarr + polpc + taxpc + pctymle +
##     pctmin80 + wsta + prbconv + mix + wser + wfed + wloc + central +
##     wfir + avgsen + wcon + wtrd, data = crime_sub)
##
## Residuals:
##        Min        1Q     Median        3Q       Max
## -0.013045 -0.004003 -0.001198  0.003880  0.018825
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.446e-02  1.542e-02   0.938 0.352050
## density      5.555e-03  7.941e-04   6.995 2.04e-09 ***
## prbarr      -5.667e-02  9.502e-03  -5.964 1.22e-07 ***
## polpc        9.407e+00  2.352e+00   4.000 0.000169 ***
## taxpc        2.413e-04  8.949e-05   2.696 0.008993 **
## pctymle      1.456e-01  4.015e-02   3.625 0.000579 ***
## pctmin80     3.620e-04  5.495e-05   6.588 1.04e-08 ***
## wsta        -5.249e-05  2.193e-05  -2.393 0.019686 *
## prbconv     -9.038e-03  5.884e-03  -1.536 0.129549
## mix         -2.121e-02  1.304e-02  -1.626 0.108938
## wser        -8.551e-05  2.843e-05  -3.007 0.003783 **
## wfed         4.192e-05  2.316e-05   1.810 0.075116 .
## wloc         5.446e-05  4.266e-05   1.277 0.206446
## central     -4.111e-03  1.890e-03  -2.175 0.033420 *
## wfir        -5.645e-05  2.616e-05  -2.158 0.034750 *
## avgsen      -6.384e-04  3.618e-04  -1.765 0.082474 .
## wcon         3.500e-05  2.393e-05   1.462 0.148584
## wtrd         5.199e-05  3.935e-05   1.321 0.191185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006998 on 63 degrees of freedom
## Multiple R-squared:  0.8915, Adjusted R-squared:  0.8622
```

```
## F-statistic: 30.44 on 17 and 63 DF,  p-value: < 2.2e-16
```

```r
sel_vars_pvals <- summary(crmrte_lm_step)$coefficients[-1,4]
sel_vars <- names(sel_vars_pvals[sel_vars_pvals < 0.05])

crmrte_formula2 <- as.formula(paste("crmrte ~", paste(sel_vars, collapse = "+"), sep = ""))
crmrte_lm2 <- lm(crmrte_formula2, data = crime_sub)
summary(crmrte_lm2)
```

```
##
## Call:
## lm(formula = crmrte_formula2, data = crime_sub)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0140651 -0.0059552 -0.0004406  0.0039566  0.0221718
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.394e-02  1.216e-02    2.791  0.00677 **
## density      6.384e-03  8.042e-04    7.938 2.37e-11 ***
## prbarr      -6.180e-02  9.467e-03   -6.527 8.98e-09 ***
## polpc        9.988e+00  2.318e+00    4.310 5.24e-05 ***
## taxpc        2.368e-04  8.859e-05    2.673  0.00936 **
## pctymle      1.262e-01  4.083e-02    3.090  0.00287 **
## pctmin80     3.863e-04  5.740e-05    6.730 3.87e-09 ***
## wsta        -6.514e-05  2.323e-05   -2.804  0.00653 **
## wser        -2.670e-05  2.823e-05   -0.946  0.34744
## central     -2.095e-03  2.034e-03   -1.030  0.30665
## wfir        -1.402e-05  2.526e-05   -0.555  0.58064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007905 on 70 degrees of freedom
## Multiple R-squared:  0.8461, Adjusted R-squared:  0.8241
## F-statistic: 38.49 on 10 and 70 DF,  p-value: < 2.2e-16
```

```r
sel_vars_cut <- c("density",  "prbarr",   "polpc",    "taxpc",
            "pctymle", "pctmin80", "wsta")

crmrte_formula3 <- as.formula(paste("crmrte ~", paste(sel_vars_cut, collapse = "+"), sep = ""))
crmrte_lm3 <- lm(crmrte_formula3, data = crime_sub)
summary(crmrte_lm3)
```

```
##
## Call:
## lm(formula = crmrte_formula3, data = crime_sub)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.013305 -0.005632 -0.001064  0.004422  0.022028
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.105e-02  9.588e-03    2.195  0.03134 *
```

```
## density       5.530e-03  6.691e-04    8.265 4.47e-12 ***
## prbarr       -5.899e-02  9.410e-03   -6.269 2.29e-08 ***
## polpc         1.002e+01  2.304e+00    4.349 4.36e-05 ***
## taxpc         2.433e-04  8.830e-05    2.755  0.00740 **
## pctymle       1.360e-01  4.016e-02    3.387  0.00114 **
## pctmin80      4.066e-04  5.579e-05    7.288 3.03e-10 ***
## wsta         -6.645e-05  2.297e-05   -2.892  0.00503 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007954 on 73 degrees of freedom
## Multiple R-squared:  0.8375, Adjusted R-squared:  0.8219
## F-statistic: 53.75 on 7 and 73 DF,  p-value: < 2.2e-16
```

```r
sel_vars_inter <- c("density",  "prbarr",   "polpc",
               "pctymle", "pctmin80", "wsta", "west*polpc")

crmrte_formula3 <- as.formula(paste("crmrte ~", paste(sel_vars_inter, collapse = "+"), sep = ""))
crmrte_lm3 <- lm(crmrte_formula3, data = crime_sub)
summary(crmrte_lm3)
```

```
##
## Call:
## lm(formula = crmrte_formula3, data = crime_sub)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0160659 -0.0053533  0.0001909  0.0044421  0.0190092
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.518e-02  9.485e-03    1.601  0.11386
## density      5.225e-03  6.422e-04    8.137 8.52e-12 ***
## prbarr      -5.197e-02  9.221e-03   -5.636 3.18e-07 ***
## polpc        1.858e+01  2.145e+00    8.662 8.90e-13 ***
## pctymle      1.232e-01  3.695e-02    3.333  0.00136 **
## pctmin80     4.007e-04  7.082e-05    5.657 2.92e-07 ***
## wsta        -6.241e-05  2.173e-05   -2.872  0.00536 **
## west         2.314e-02  6.822e-03    3.392  0.00113 **
## polpc:west  -1.553e+01  3.760e+00   -4.131 9.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007522 on 72 degrees of freedom
## Multiple R-squared:  0.8567, Adjusted R-squared:  0.8407
## F-statistic: 53.79 on 8 and 72 DF,  p-value: < 2.2e-16
```