Lab4_RMD_Layer2

Kim Vignola

8/5/2017

Add libraries

```
library(car)
library(corrplot)
```

new variables based on proportions of wages. Not sure that we need these but including in case you would like to review in greater detail.

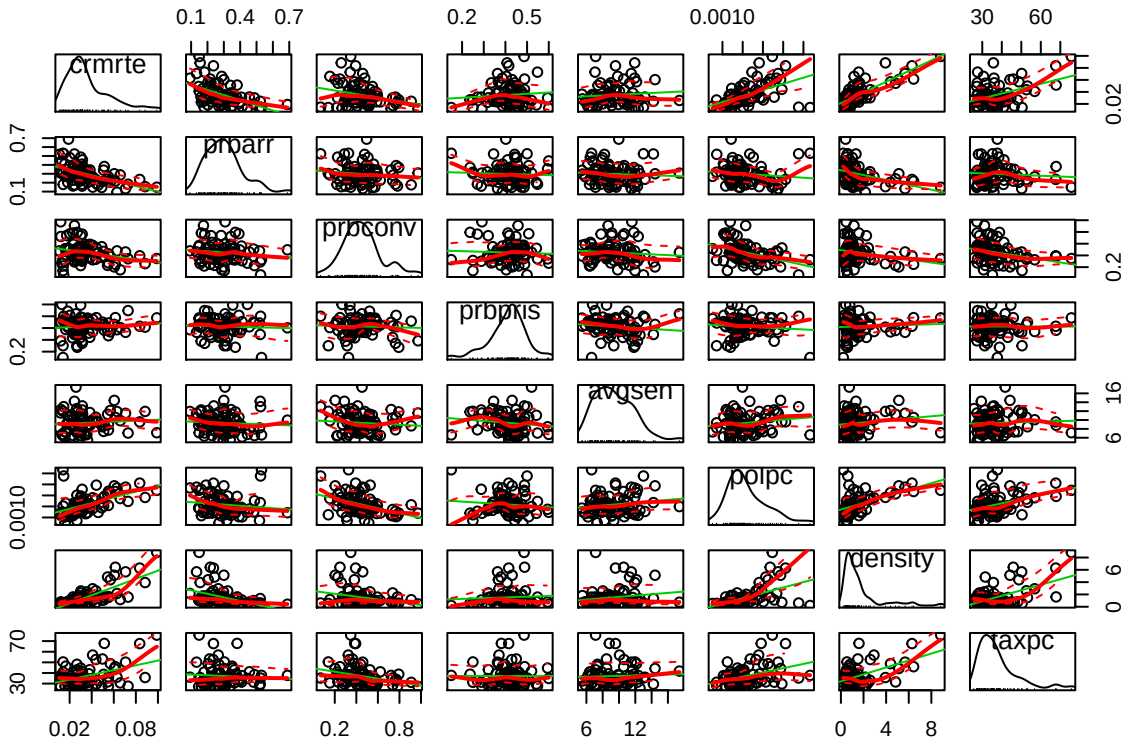
```
crime$tot_wages <- (crime$wcon + crime$wtuc + crime$wtrd + crime$wfir + crime$wser + crime$wmfg + crime$
crime$gov_wages <- (crime$wfed + crime$wsta + crime$wloc)
crime$bus_wages <- (crime$wcon + crime$wtuc + crime$wtrd + crime$wfir + crime$wser + crime$wmfg)
crime$low_wages <- as.numeric(crime$tot_wages < mean(crime$tot_wages))
crime$high_wages <- as.numeric(crime$tot_wages > mean(crime$tot_wages))
crime$wcon_p <- crime$wcon/crime$tot_wages
crime$wtuc_p <- crime$wtuc/crime$tot_wages
crime$wtrd_p <- crime$wtrd/crime$tot_wages
crime$wfir_p <- crime$wfir/crime$tot_wages
crime$wser_p <- crime$wser/crime$tot_wages
crime$wmfg_p <- crime$wmfg/crime$tot_wages
crime$wfed_p <- crime$wfed/crime$tot_wages
crime$wsta_p <- crime$wsta/crime$tot_wages
crime$wloc_p <- crime$wloc/crime$tot_wages
crime$bus_wage_p <- crime$wcon_p + crime$wtuc_p + crime$wtrd_p + crime$wfir_p + crime$wser_p + crime$wm
crime$govt_wage_p <- crime$wfed_p + crime$wfed_p + crime$wfed_p
crime$sta_loc_p <- crime$wsta_p + crime$wloc_p
```

Clean data (preliminary)

```
crime2 <- subset(crime, crime$prbarr < 1 & crime$prbconv <1 & crime$taxpc <100 & crime$pctymle < 20)
```

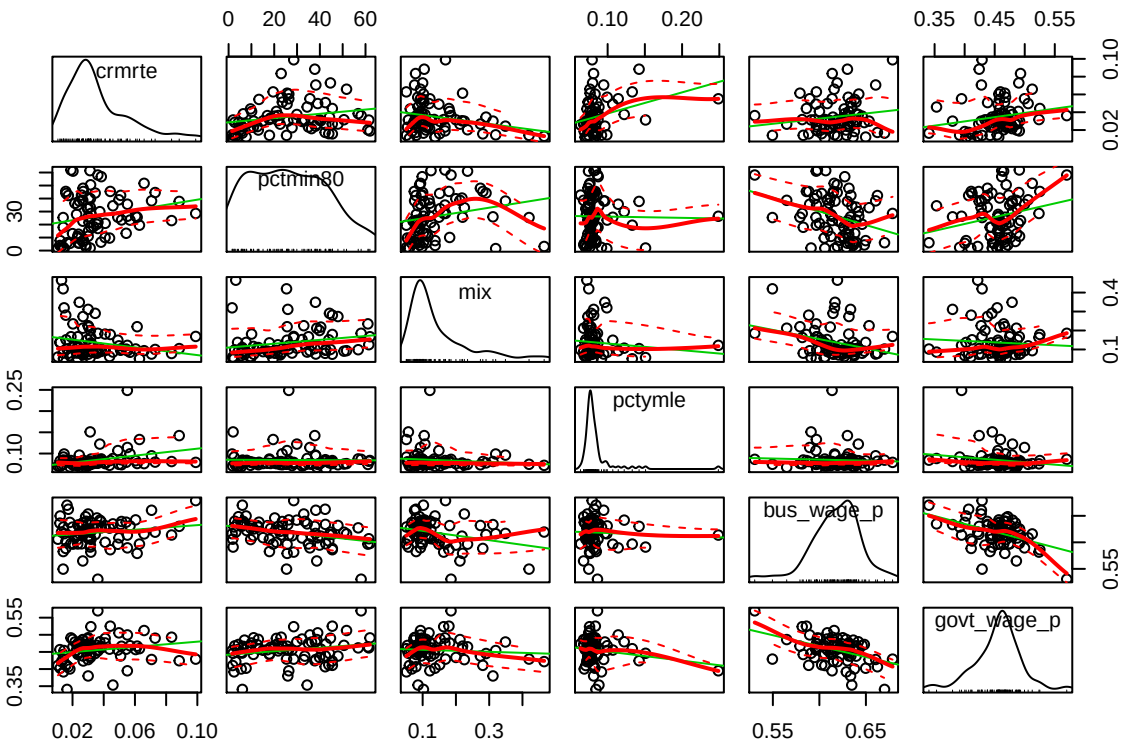
Scatterplot matrix #1 using cleaned (preliminary) dataset. (Had to do this in chunks due to the number of variables). The first one shows strong correlations between crime rate adn police per capita, density and taxpc. Density appears to have a parabolic relationship between taxes and police per capita. Of the prison metrics, prbarr appears to have the biggest impact on declining crime rates.

```
scatterplotMatrix(crime2[ , (names(crime2) %in% c("crmrte", "density", "prbarr", "prbconv", "prbpris",
```


crmrte
prbarr
prbconv
prbpris
avgsen
polpc
density
taxpc

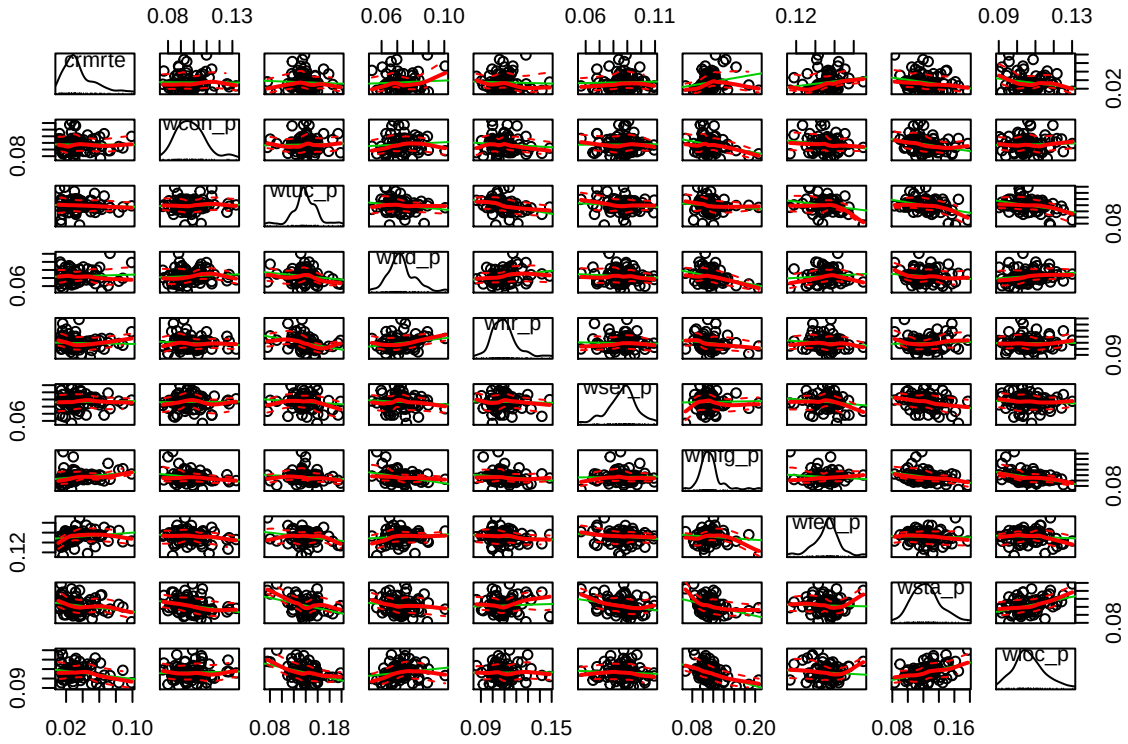
Scatterplot #2: As Kristen noted, percent male has a strong correlation with crime rate. There is some relation (pattern looks concave) between percent minority and crime mix. Counties with higher proportions of business wages are inversely correlated with those with government wages. Wages overall are linked with higher crime.

```
scatterplotMatrix(crime2[ , (names(crime2) %in% c("crmrte", "pctmin80", "mix", "pctymle", "bus_wage_p",
```



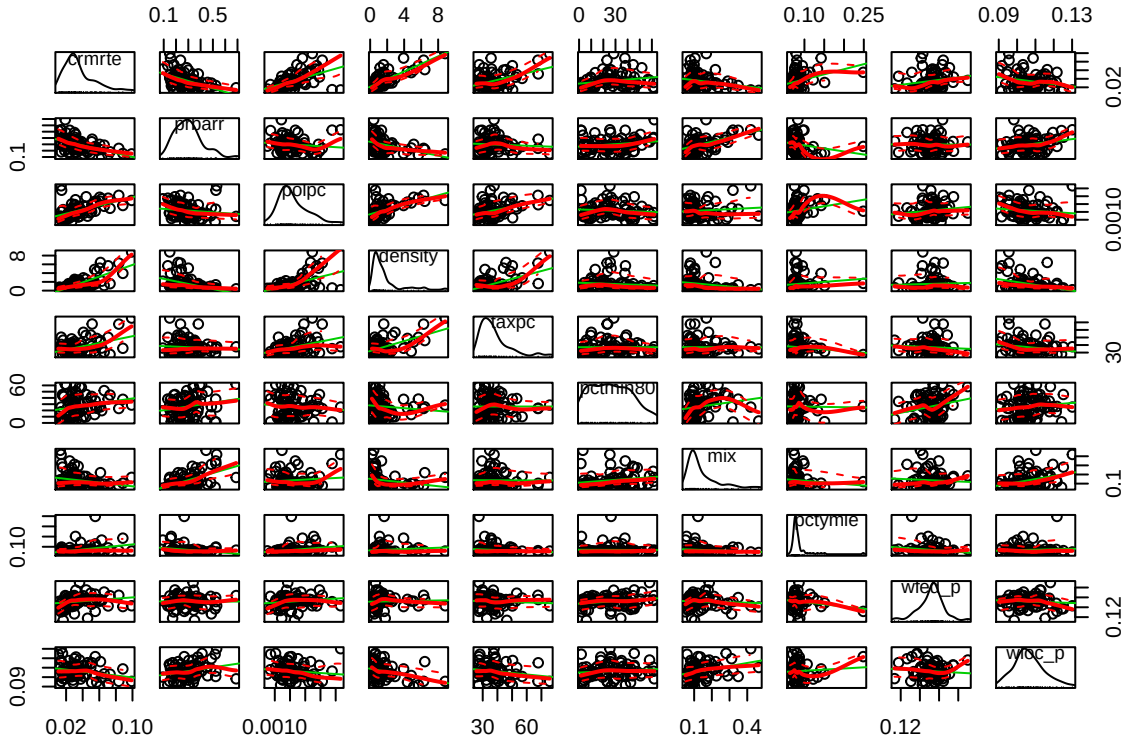
Scatterplot 3: Specific industries. Retail, manufacturing and federal wages are positively correlated with crime rates while there is a descending relationship between state and local government wages and crime. (These are based on proportions of wages in each county.)

```
scatterplotMatrix(crime2[ , (names(crime2) %in% c("crmrte", "wcon_p", "wtuc_p", "wtrd_p", "wfir_p", "ws
```



Pulling out key metrics we can also see that percentage male has a relationship with policepc (positive) and probability of arrest (negative).

```
scatterplotMatrix(crime2[ , (names(crime2) %in% c("crmrte", "density", "prbarr", "polpc","taxpc", "pcty
```



Potential Models

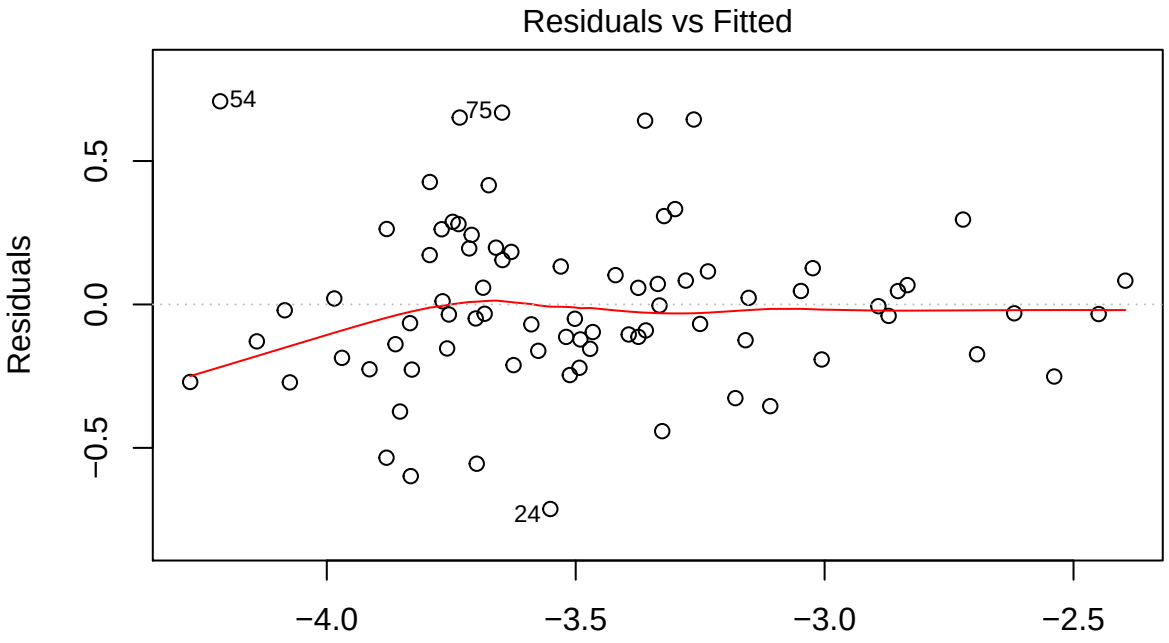
```
model2 <- lm(log(crmrte) ~ density + taxpc + density*taxpc + pctymle + wfed_p + prbarr, data = crime2)
```

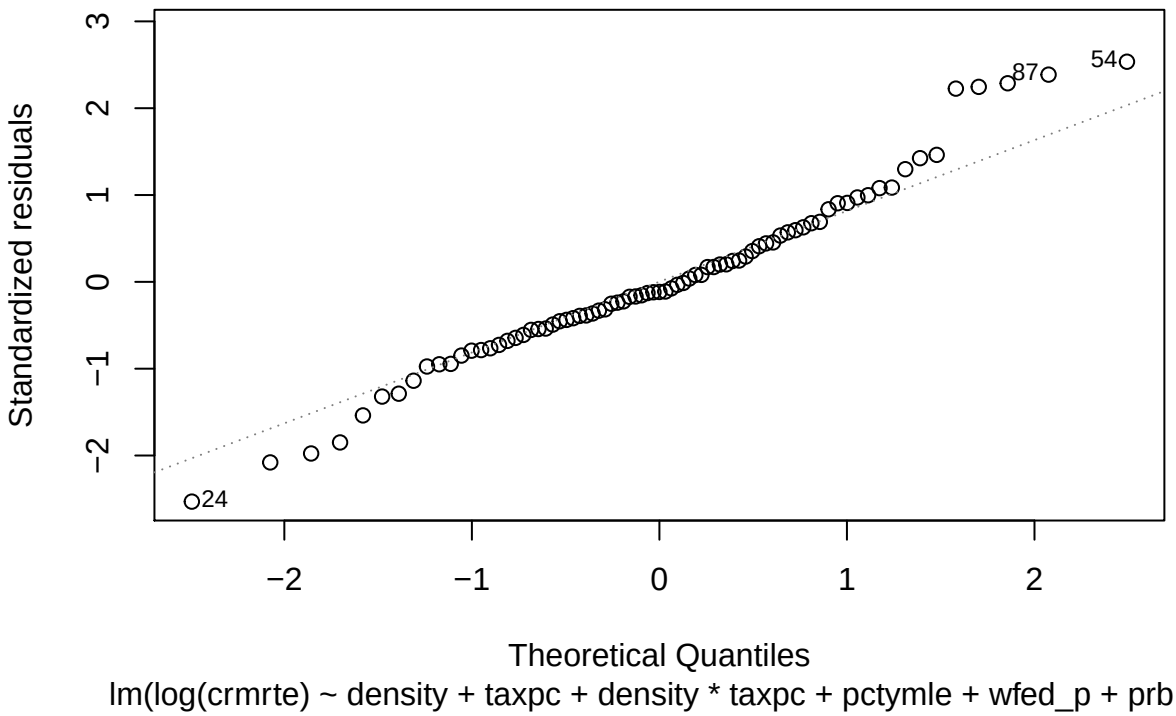
Model 2

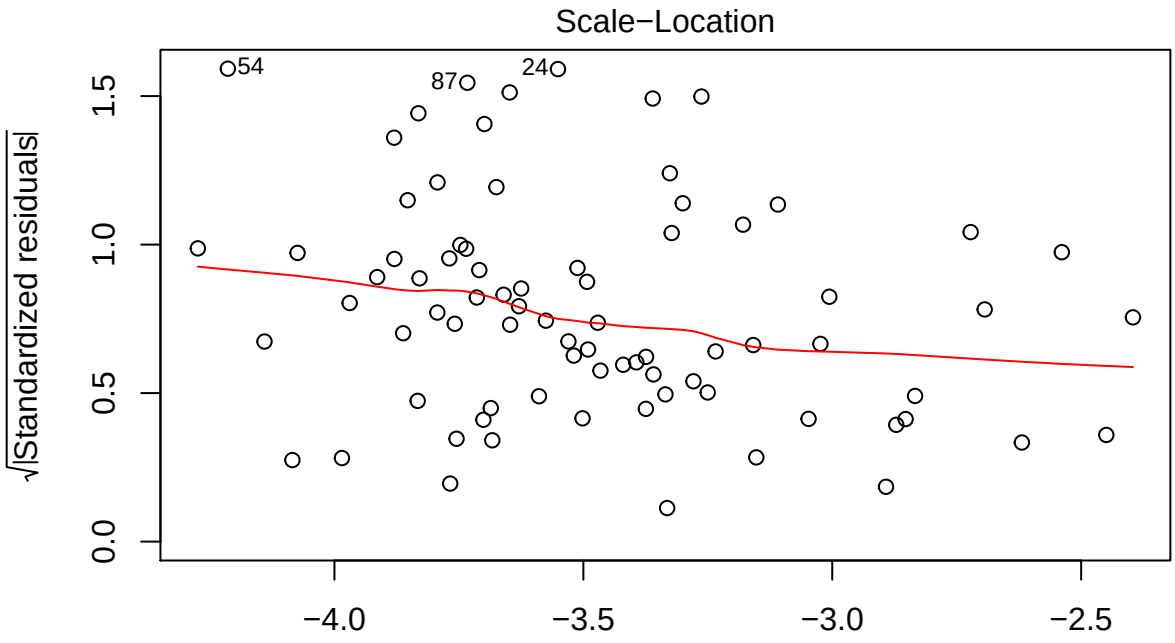
```
summary(model2)
```

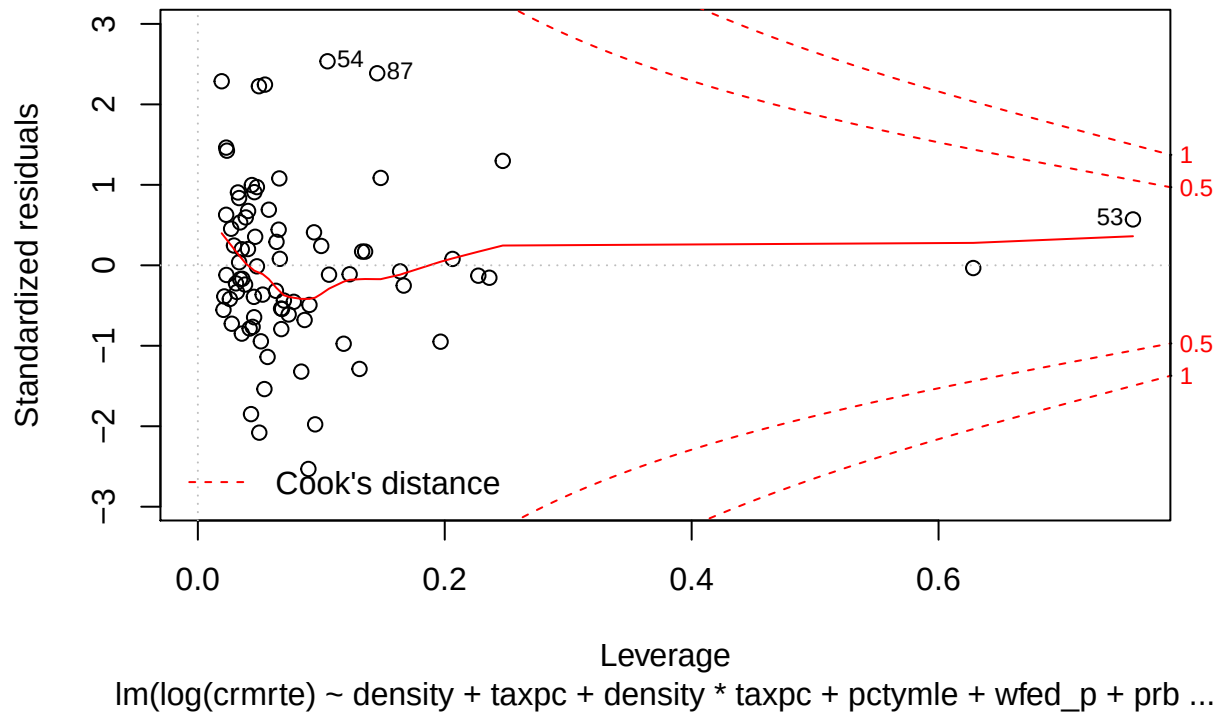
```
##
## Call:
## lm(formula = log(crmrte) ~ density + taxpc + density * taxpc +
##     pctymle + wfed_p + prbarr, data = crime2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.71306 -0.15824 -0.03247  0.14351  0.70822
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.237784   0.545399 -11.437  < 2e-16 ***
## density        0.236809   0.077755   3.046 0.003244 **
## taxpc          0.011606   0.005718   2.030 0.046076 *
## pctymle        5.250586   1.475482   3.559 0.000665 ***
## wfed_p        13.325171   2.812317   4.738 1.06e-05 ***
## prbarr        -1.266591   0.342499  -3.698 0.000421 ***
## density:taxpc -0.001885   0.001434  -1.314 0.193073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2953 on 72 degrees of freedom
## Multiple R-squared:  0.678,  Adjusted R-squared:  0.6512
## F-statistic: 25.27 on 6 and 72 DF,  p-value: 6.369e-16
```

```
plot(model2)
```


Residuals vs Fitted
Residuals
0.5
0.0
−0.5
54
75
24
−4.0
−3.5
−3.0
−2.5
Fitted values
lm(log(crmrte) ~ density + taxpc + density * taxpc + pctymle + wfed_p + prb ...


Normal Q−Q
Standardized residuals
3
2
1
0
−1
−2
87
54
24
−2
−1
0
1
2
Theoretical Quantiles
lm(log(crmrte) ~ density + taxpc + density * taxpc + pctymle + wfed_p + prb ...


Scale−Location
√|Standardized residuals|
54
87
24
1.5
1.0
0.5
0.0
−4.0
−3.5
−3.0
−2.5
Fitted values
lm(log(crmrte) ~ density + taxpc + density * taxpc + pctymle + wfed_p + prb ...


Residuals vs Leverage
Standardized residuals
54
87
53
3
2
1
0
−1
−2
−3
1
0.5
0.5
1
Cook's distance
0.0
0.2
0.4
0.6
Leverage
lm(log(crmrte) ~ density + taxpc + density * taxpc + pctymle + wfed_p + prb ...