

Lab3 Crime Statistics

Kenneth Chen, Peter Trenkwalder, Danielle Salah

7/15/2018

```
crime = read.csv("crime_v2.csv")
```

Introduction

We received a crime dataset on North Carolina and would like to explore crime statistics. We like to investigate crime statistics at hands to develop several viable approaches in order to reduce crime. The dataset has **97** observations and **25** variables. Our first approach is to investigate each of the variables and how they relate to the occurrence of crimes.

Exploratory Data Analysis

We listed all variables and their descriptions here.

variable	label
1 county	county identifier
2 year	1987
3 crmrte	crimes committed per person
4 prbarr	'probability' of arrest
5 prbconv	'probability' of conviction
6 prbpris	'probability' of prison sentence
7 avgsen	avg. sentence, days
8 polpc	police per capita
9 density	people per sq. mile
10 taxpc	tax revenue per capita
11 west	=1 if in western N.C.
12 central	=1 if in central N.C.
13 urban	=1 if in SMSA
14 pctmin80	perc. minority, 1980
15 wcon	weekly wage, construction
16 wtuc	weekly wage, trns, util, commun
17 wtrd	weekly wage, whlelse, retail trade
18 wfir	weekly wage, fin, ins, real est
19 wser	weekly wage, service industry
20 wmfgr	weekly wage, manufacturing
21 wfed	weekly wage, fed employees
22 wsta	weekly wage, state employees
23 wloc	weekly wage, local gov emps
24 mix	offense mix: face-to-face/other
25 pctymle	percent young male

Out of 25 variables, we set our dependent variable to be **crime rates, crmrte** because we believe this reflects the frequency of crimes in North Carolina.

Questions

1. Are crime rates affected by probability of arrest? or can we predict the crime rate by the probability of arrest in a given location?

Answer

2. Are crime rates dependent on the population of young males?

$$\widehat{\text{crime rate}} = \beta_0 + \beta_1 \cdot \text{prbarr} + \beta_2 \cdot \text{pctym1}$$

3. Omitted Variable Bias (OVB)

We are concerned that the variable we are currently interested, probability of arrest and population of young males, have other variables that are highly correlated to each other such as the location in North Carolina, and if the wages across the country has a causal effect on crimes based on monetary incentive. If that's the case, we will be missing a key variable in our analysis.

Preliminary investigation of each variable correlation

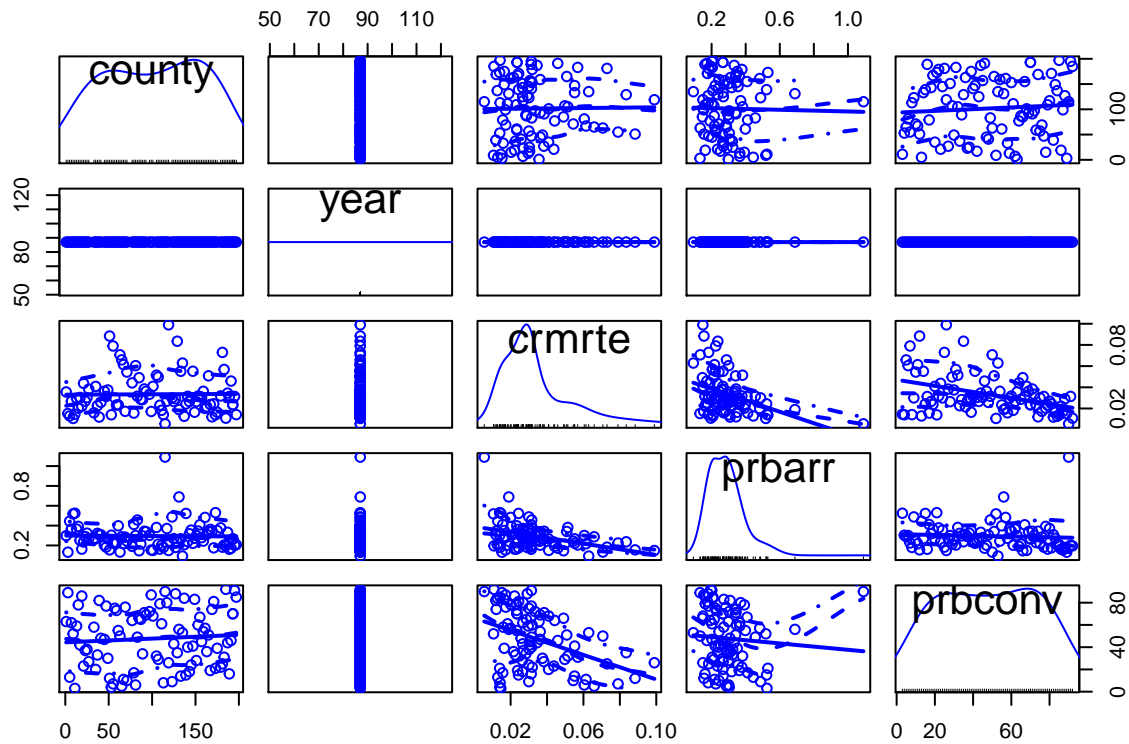
In order to retrieve all observations equally across all 25 variables, we first filtered the data with complete data on all 25 variables. Subsequently we checked each of the variable relationship by scatterplot matrix 5 at a time. In the last plot, we focused on

1. crime rates,
2. probability of arrest,
3. police per capita,
4. weekly wages, federal employee,
5. weekly wages, state,
6. percent of young males

```
# Since we're going to look at the all the variables across,  
# I'm dropping all the empty roles here. So 97 observations  
# end up at 91 observations.  
crime_fullrow = crime[complete.cases(crime), ]
```

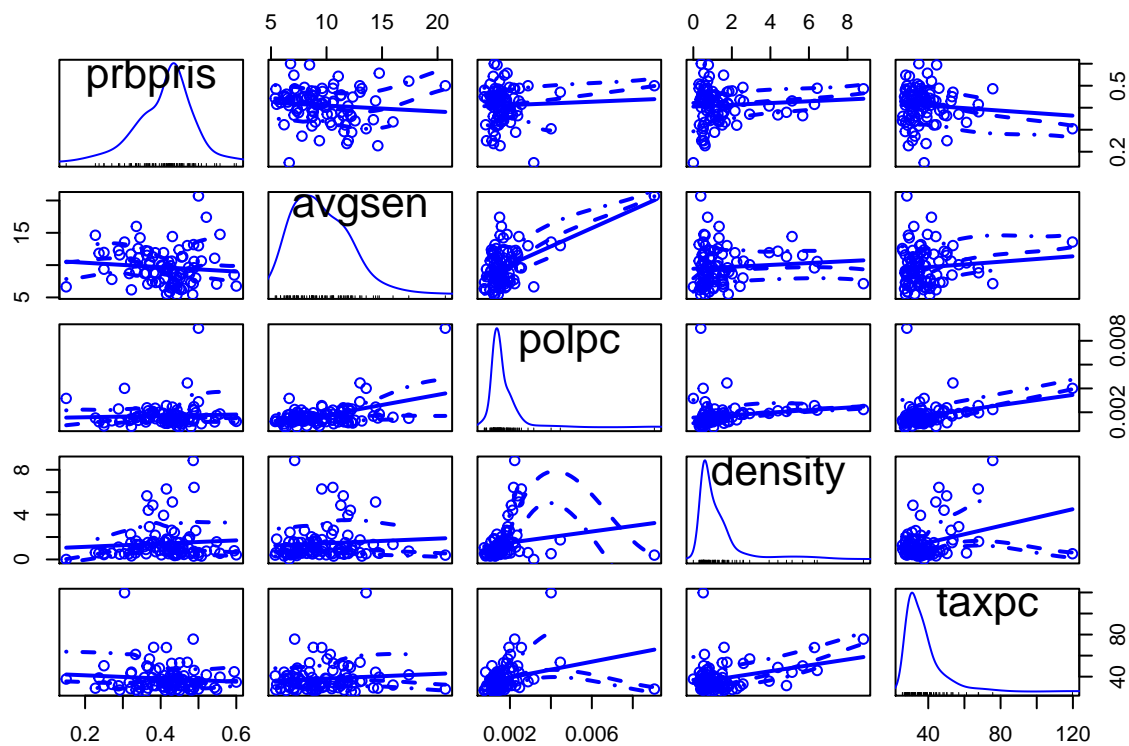
Checking the first 5 variables

```
scatterplotMatrix(crime_fullrow[1:5])  
  
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread  
  
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread  
  
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread  
  
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread
```



Checking the next 5 variables

```
scatterplotMatrix(crime_fullrow[6:10])
```



Checking the next 5 variables

```
scatterplotMatrix(crime_fullrow[11:15])
```

```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

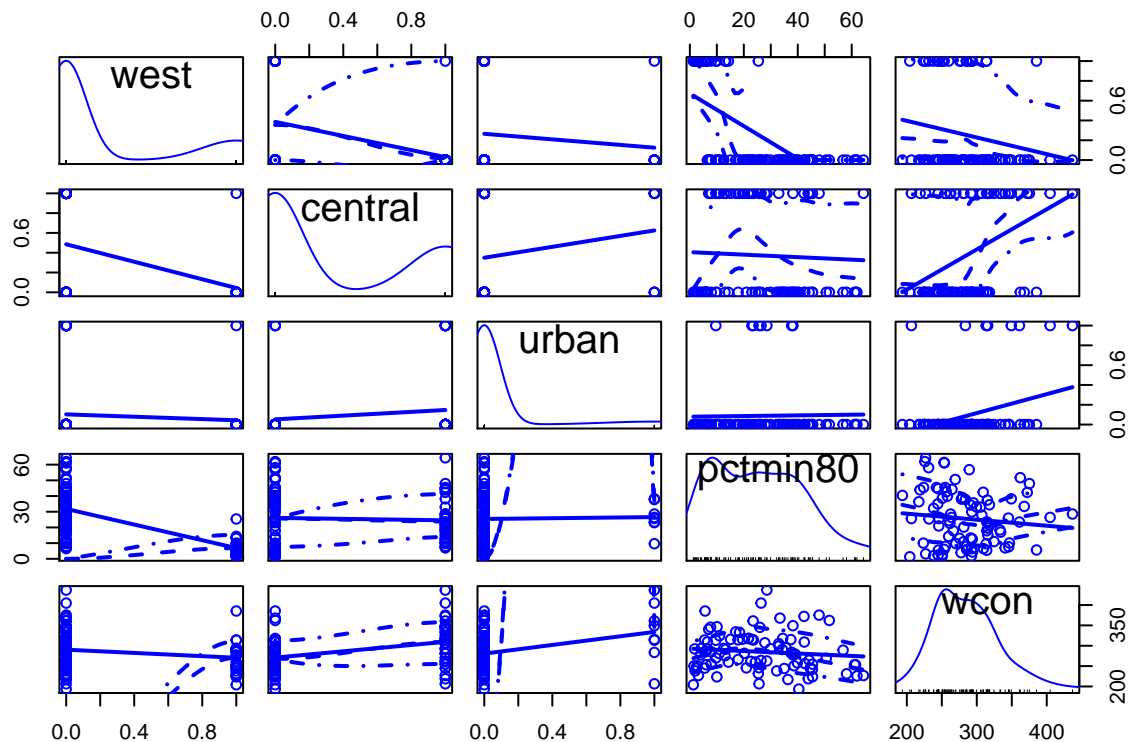
```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

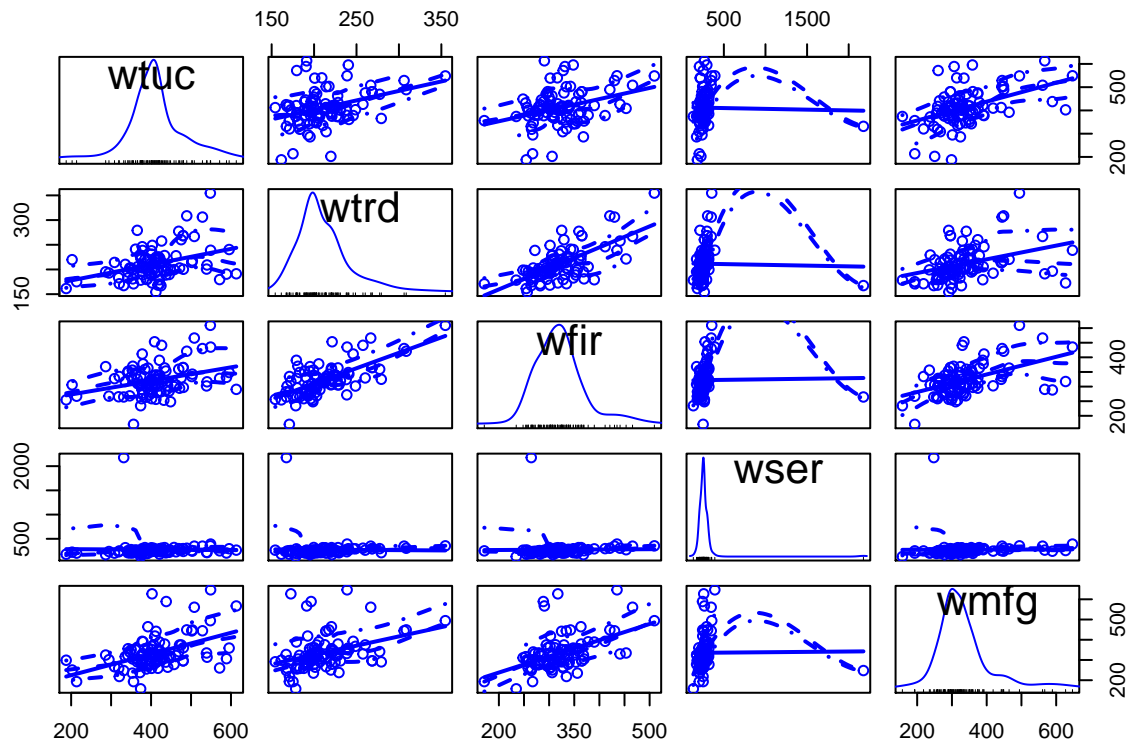
```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```

```
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit smooth
```



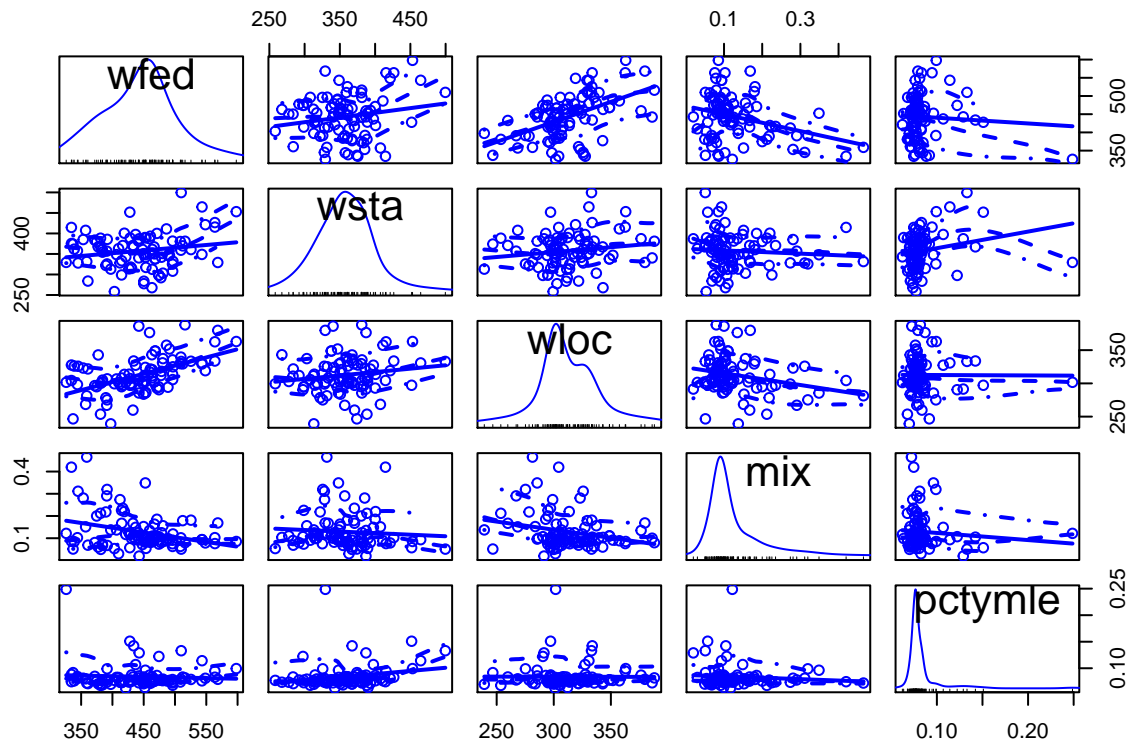
Checking the next 5 variables

```
scatterplotMatrix(crime_fullrow[16:20])
```



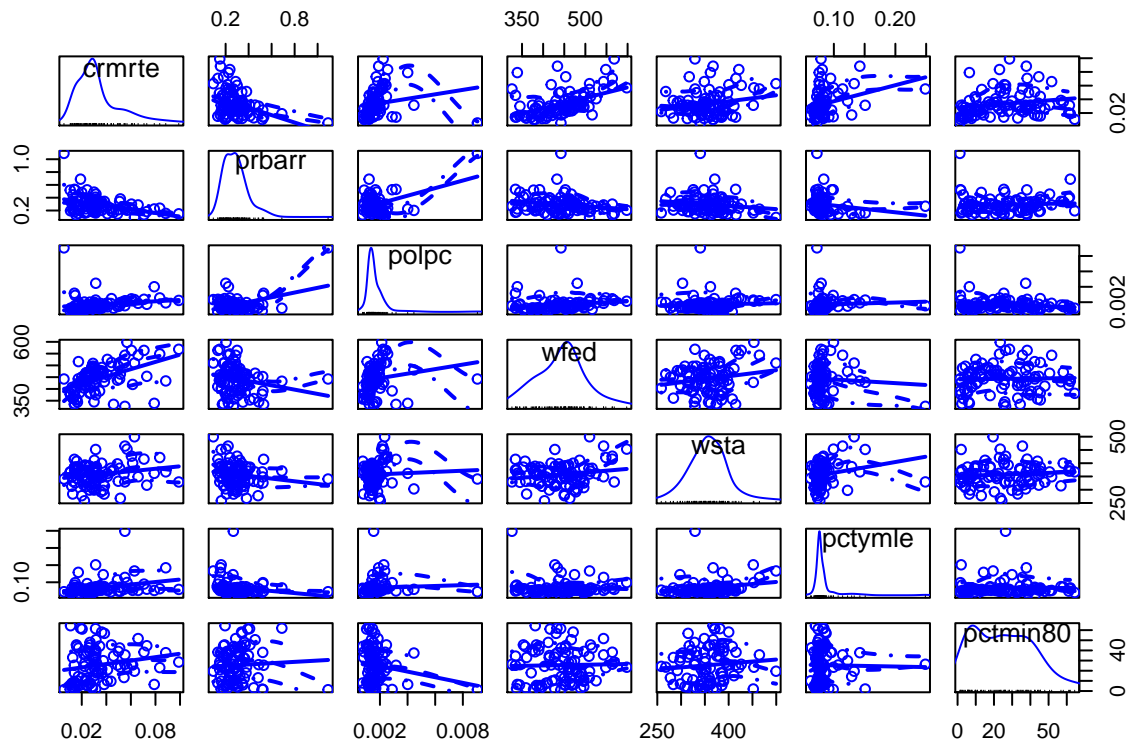
Checking the last 5 variables

```
scatterplotMatrix(crime_fullrow[21:25])
```



Looking at the crime rate, probability arrest and a few other variables or (Key Variables)

```
scatterplotMatrix(~crmte + prbarr + polpc + wfed + wsta + pctymle +
  pctmin80, data = crime_fullrow)
```



To use the stargazer later on `#stargazer {r, results='asis'} library(stargazer) stargazer(model1, model2, type = "latex", report = "vc", # Don't report errors, since we haven't covered them title = "Linear Models Predicting College GPA", keep.stat = c("rsq", "n"), omit.table.layout = "n") # Omit more output related to errors`