

Crime Reduction thoughts

Kiersten Henderson

8/5/2017

Big Picture Thoughts

My understanding is that we are trying to inform a political candidate for local government on how to reduce crime locally.

We don't know what "kind" of county the candidate is wondering about. Hopefully, the candidate is running in NC, otherwise the dataset is probably not appropriate to apply to other counties in other states.

There are actually 100 counties in NC (all were created before 1911). So this dataset does not include data for all NC counties and is hopefully random.

I am wondering - do we want to understand overall crime rates for NC or do we want to understand crime rate in the specific kind of locality that the politician is from?

-We may want to give recommendations for the cases where the candidate is running in an urban county versus a rural one.

I guess that indicator variables would help take care of this - but you would know a priori what kind of county the candidate is from... so would you really use indicator variables or would you just subset the dataset and use counties that are similar in terms of the indicators?

We will probably discuss this!

Regardless of the answer to the above, it will be good to know which county characteristics actually affect crime rate.

That's what I did what is below. The analysis led me to see how being rural versus urban affects the variables that seem correlated with crime. And the punchline is that being urban seems to affect most of the variables positively correlated with crime rate. In retrospect, this could be due to density (not due to urban, necessarily).

```
data = read.csv("crime.csv")
#View(data)
```

Only 8 of the 90 counties in NC that are included in our dataset are urban.

```
sum(data$urban == 1)
```

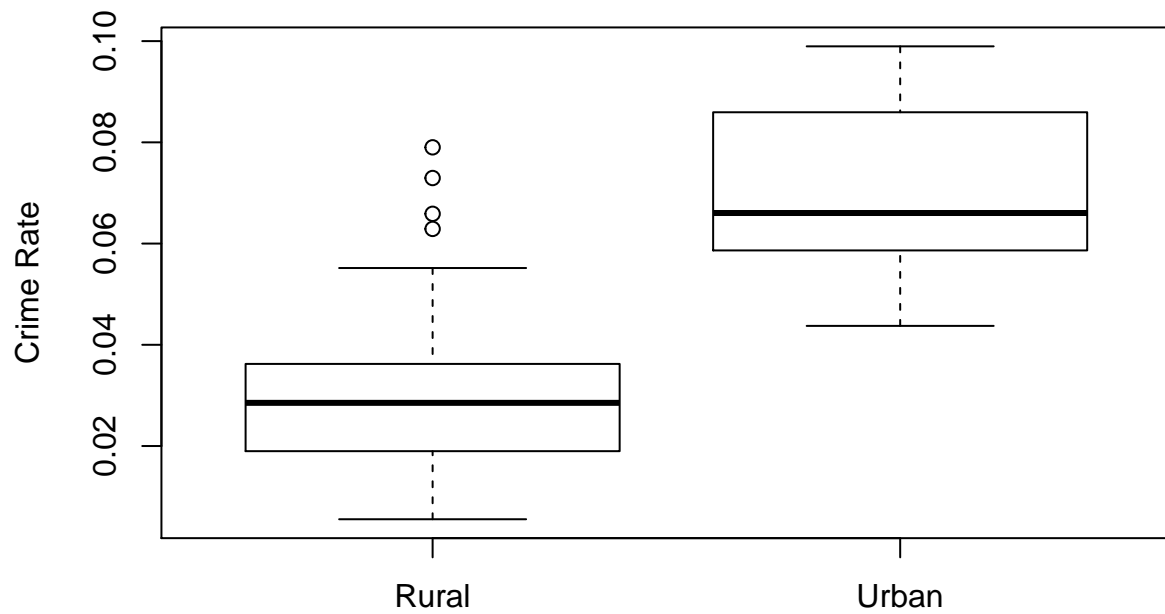
```
## [1] 8
```

Crime Rate in Different kinds of Counties

Crime rate appears about 3 times higher in urban counties than rural ones. However, note the urban sample is only 8 observations of 90.

```
boxplot(data$crmrte~data$urban, ylab = "Crime Rate", main = "Crime Rate in Rural versus Urban NC Counties")
```

Crime Rate in Rural versus Urban NC Counties



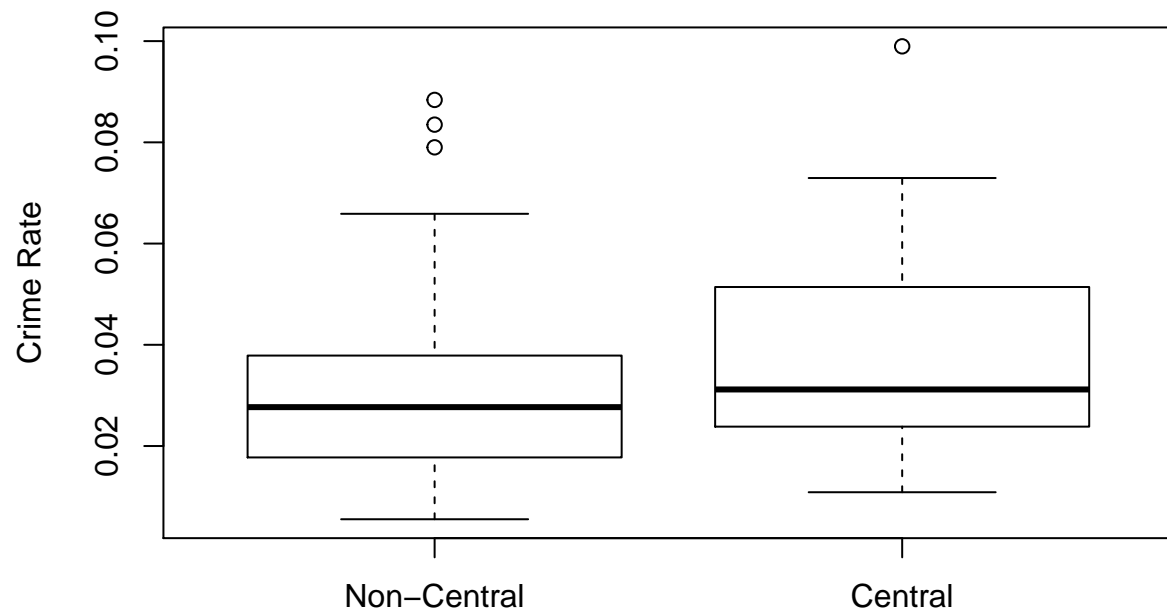
```
sum(data$central == 1)
```

```
## [1] 34
```

One third of counties are “central”. Crime doesn’t seem different between central and non-central counties.

```
boxplot(data$crmrte~data$central, ylab = "Crime Rate", main = "Crime Rate in Central vs. Non-Central NC
```

Crime Rate in Central vs. Non-Central NC Counties



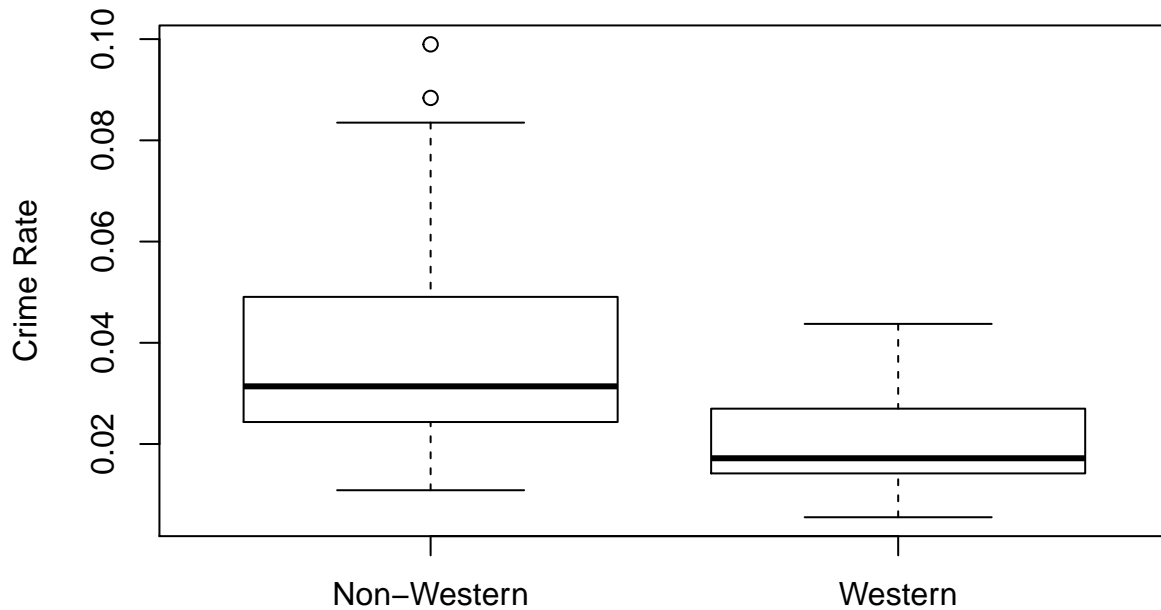
There are 21 western counties. Western counties may have a slightly lower crime rate than non-western counties. Are non-western countries more urban? I didn't check but can.

```
sum(data$west == 1)
```

```
## [1] 21
```

```
boxplot(data$crmrte~data$west, ylab = "Crime Rate", main = "Crime Rate in Western vs. Non-Western NC Counties")
```

Crime Rate in Western vs. Non-Western NC Counties



Basically, if the candidate is from a rural area in NC, all's well with the sample. If the candidate wants info on determinants of crime in an urban county of NC - the sample or urban counties probably isn't big enough. It would have made more sense to start with a random sample of counties in the US with similar population densities.

But if use "urban" as an indicator variable, then we will probably have interaction terms using it.

??Question - do we need to find statistically significant difference in crime rate between urban and non-urban to use it as an indicator variable?

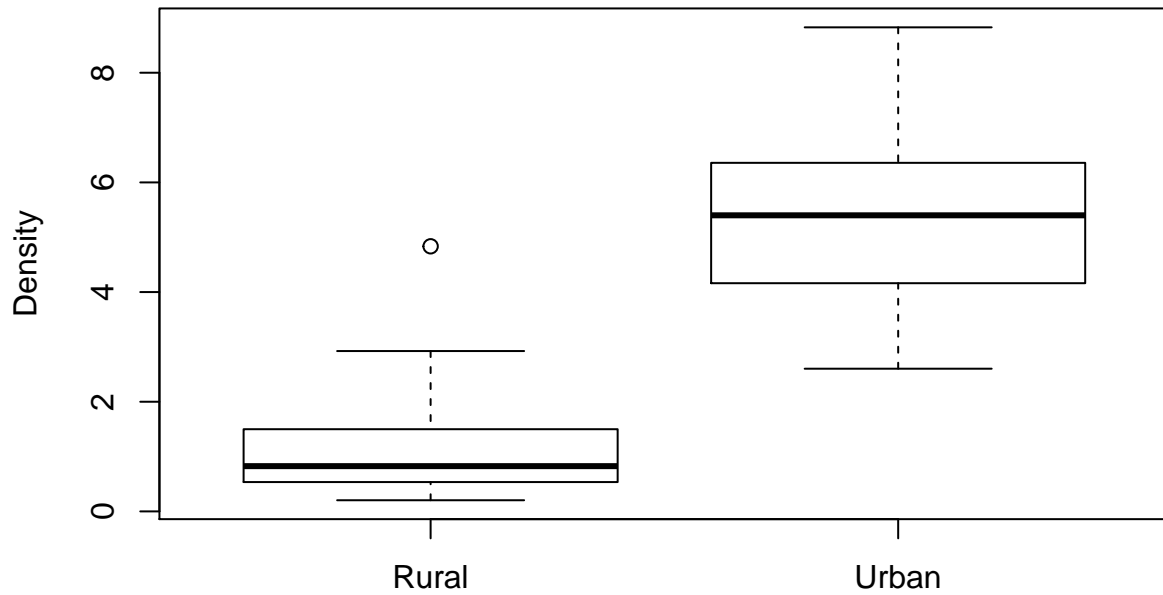
We'll have to do bivariate analyses with urban and other variables to see which ones urban influences. -wages are probably influenced (they'll be higher in urban environments because of cost-of-living adjustment).

Ok, i'll look at those that came out as correlated in Aaron's analysis.

Density makes perfect sense, more dense living in urban environments.

```
boxplot(data$density~data$urban, ylab = "Density", names = c("Rural", "Urban"), main = "Density in Urban")
```

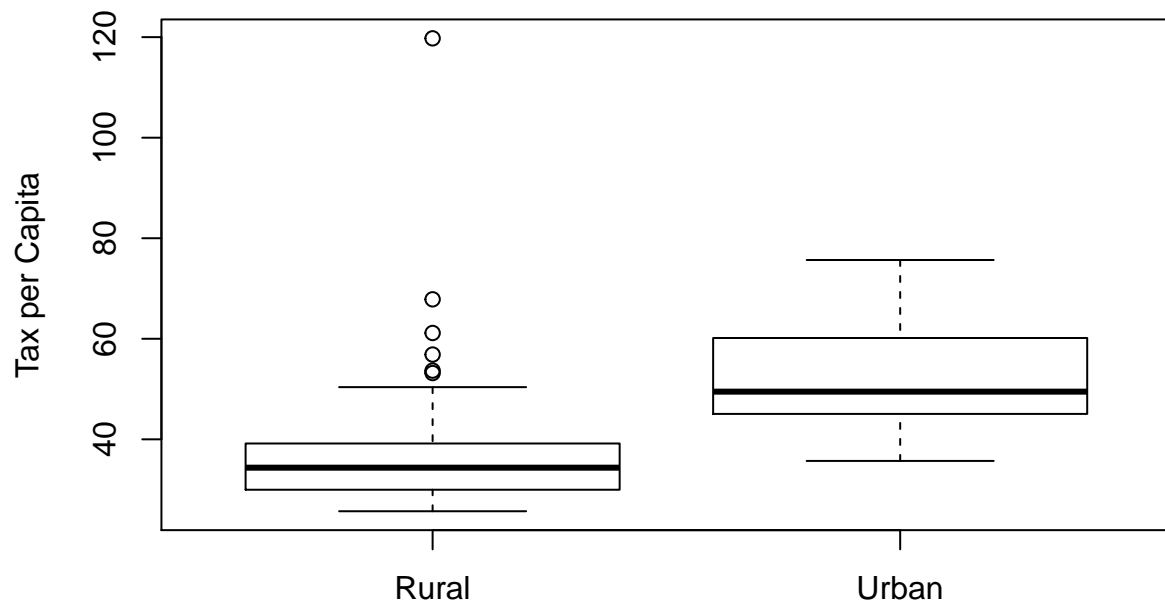
Density in Urban vs. Rural Counties



Tax per capita being higher in urban areas makes sense if wages are higher in urban areas (cost-of-living adjustment), or it is possible that percent taxation is also higher in urban areas.

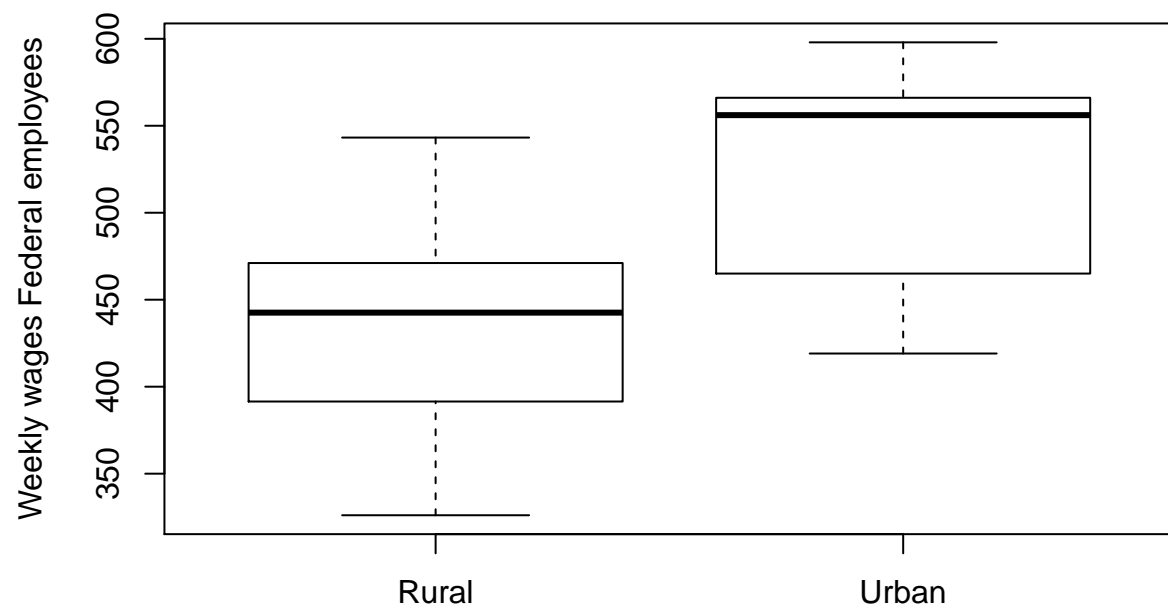
```
boxplot(data$taxpc~data$urban, ylab = "Tax per Capita", names = c("Rural", "Urban"), main="Tax per Capita by Urban vs Rural")
```

Tax per Capita in Urban vs. Rural Counties

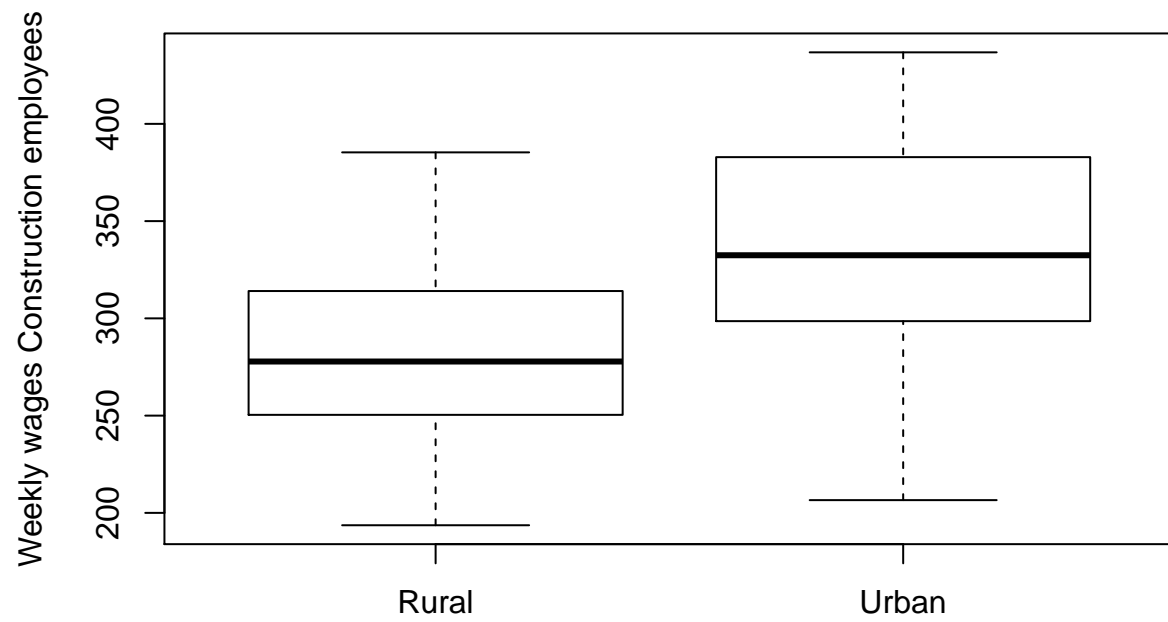


All forms of wage are below..sorry i didn't beautify the charts, but you should be able to glean the trends. In general, wages across categories are higher in urban areas. A notable exception is service employees - and there is no correlation between wage of service employees and crimerate.

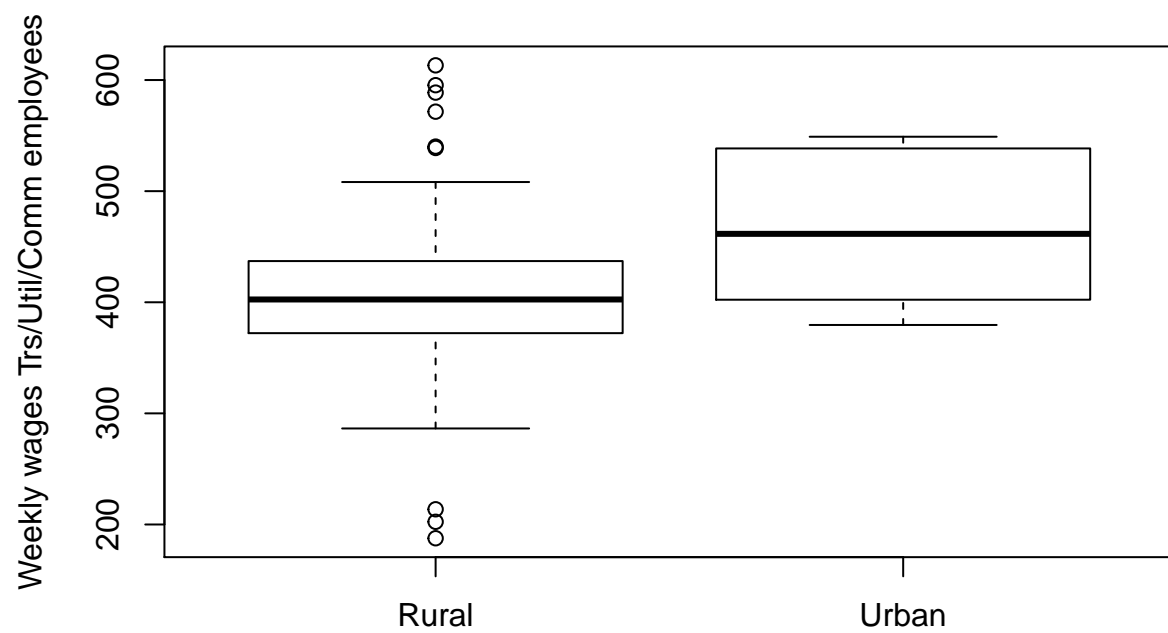
```
boxplot(data$wfed~data$urban, ylab = "Weekly wages Federal employees", names = c("Rural", "Urban"))
```



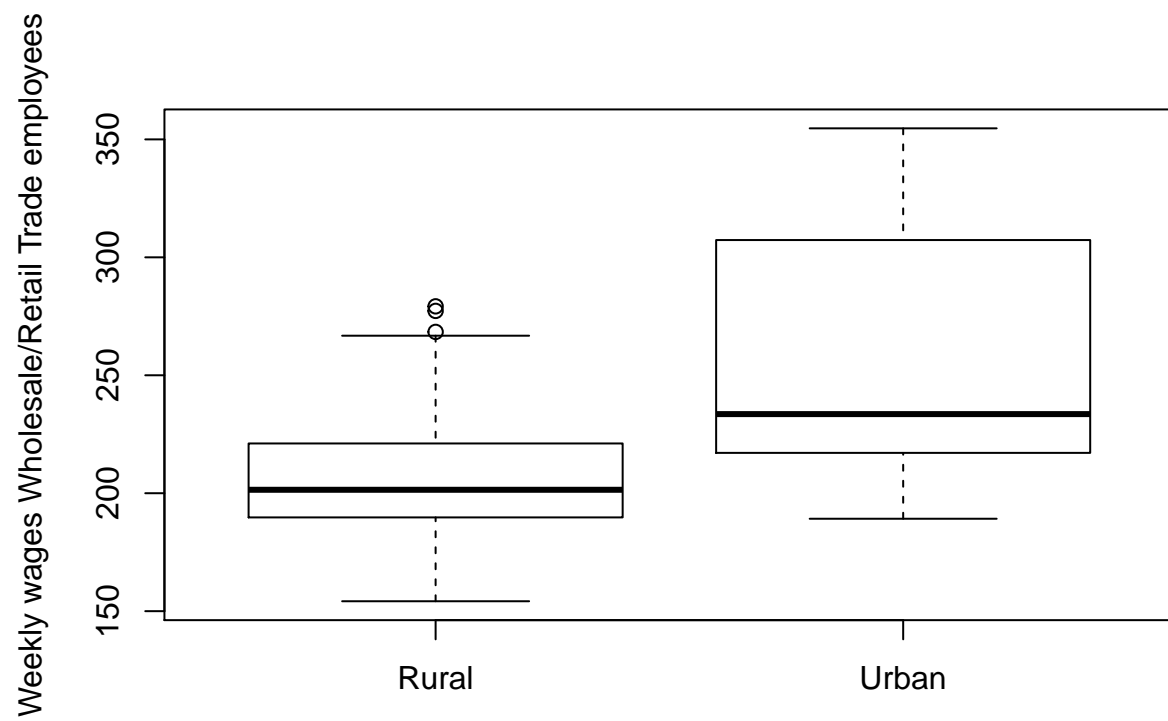
```
boxplot(data$wcon~data$urban, ylab = "Weekly wages Construction employees", names = c("Rural", "Urban"))
```



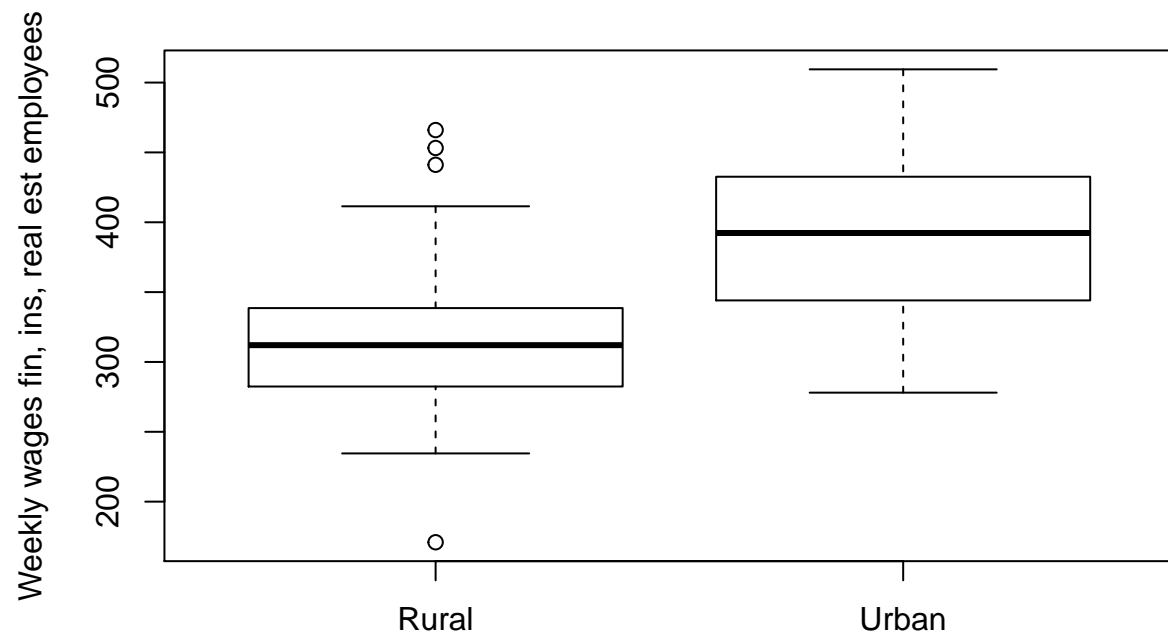
```
boxplot(data$wtuc~data$urban, ylab = "Weekly wages Trs/Util/Comm employees", names = c("Rural", "Urban"))
```

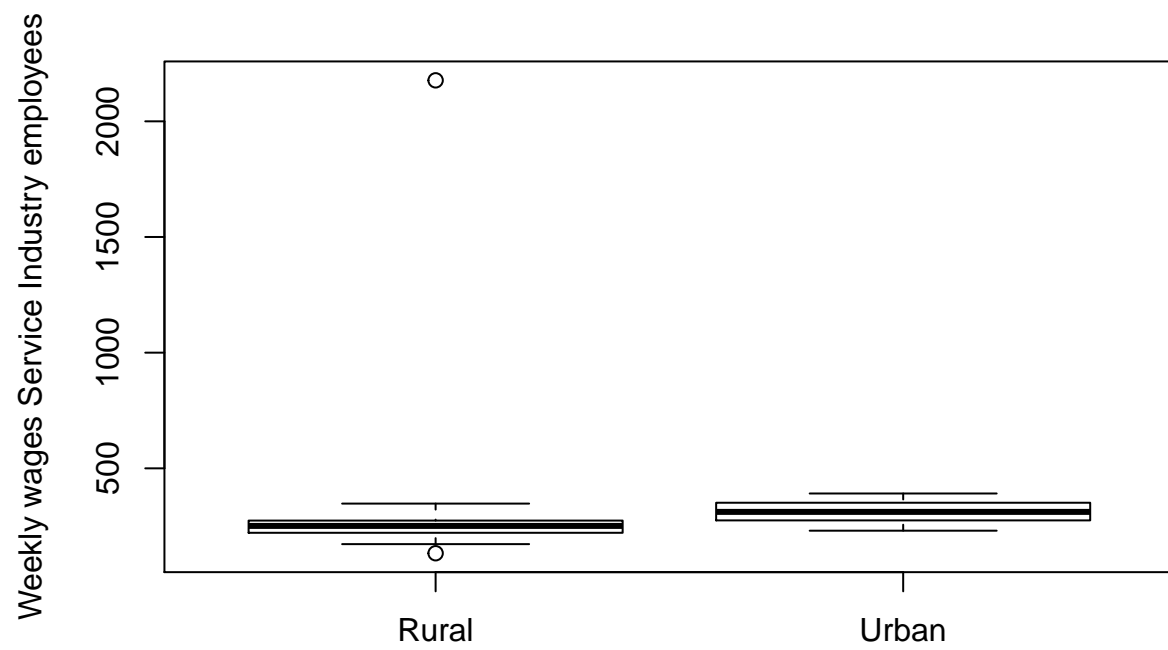
```
boxplot(data$wtrd~data$urban, ylab = "Weekly wages Wholesale/Retail Trade employees", names = c("Rural"
```



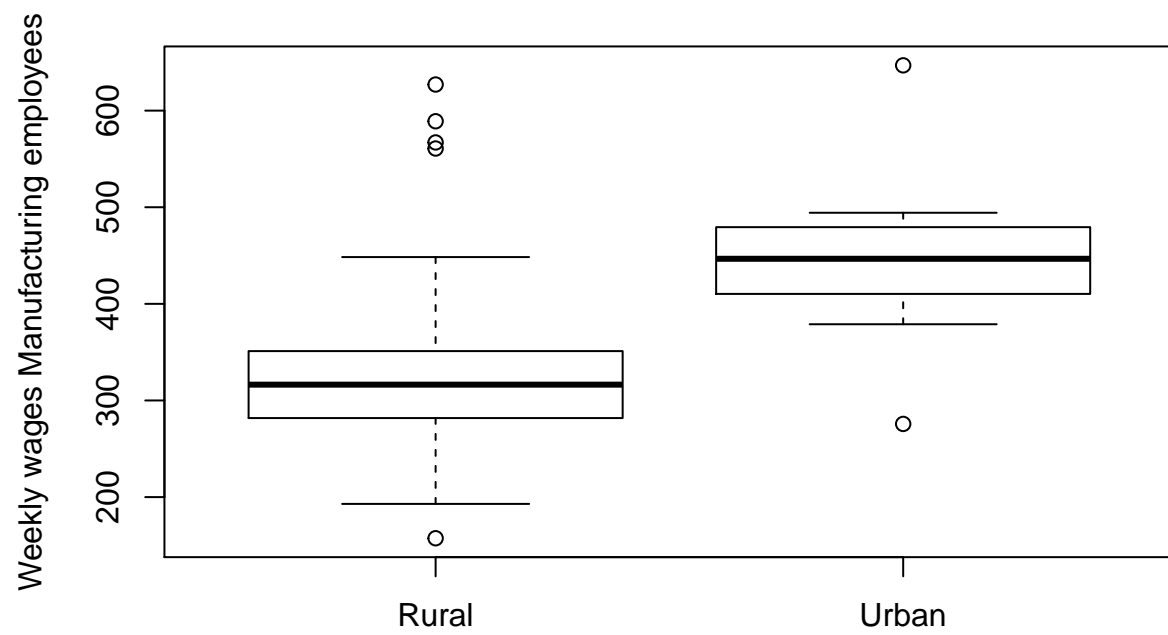
```
boxplot(data$wfirm~data$urban, ylab = "Weekly wages fin, ins, real est employees", names = c("Rural", "U
```



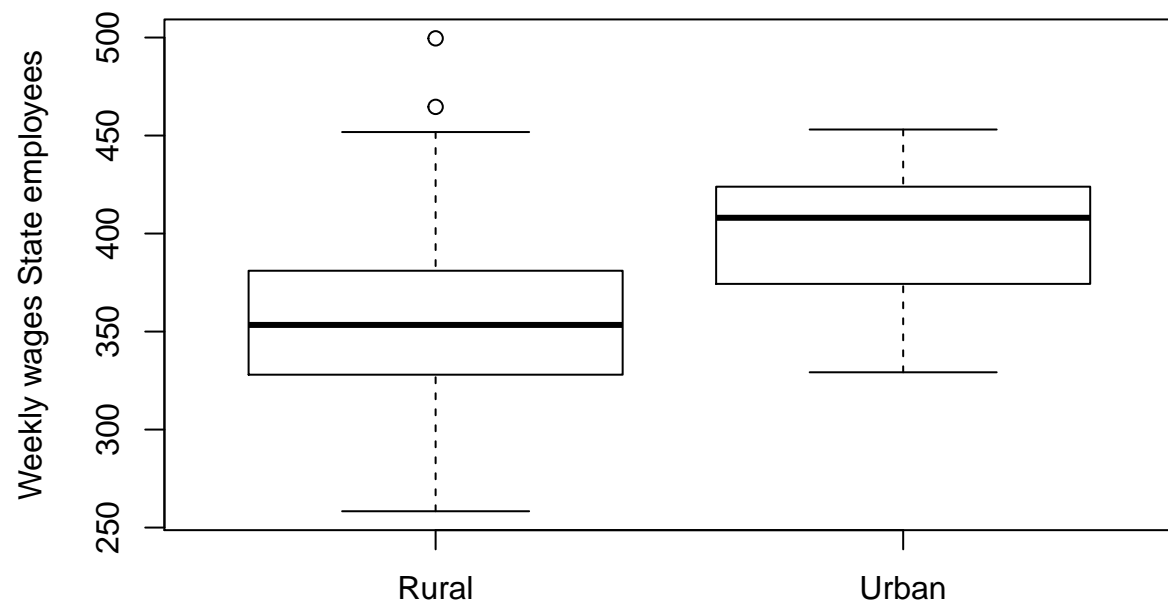
```
boxplot(data$wser~data$urban, ylab = "Weekly wages Service Industry employees", names = c("Rural", "Urban"))
```



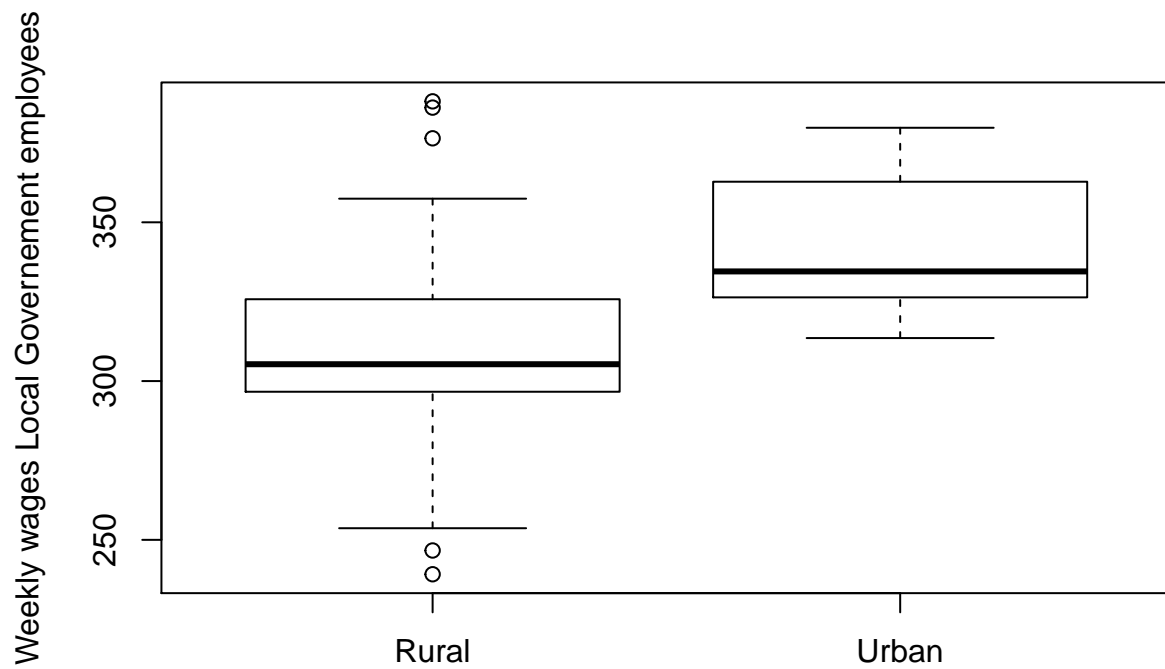
```
boxplot(data$wmfg~data$urban, ylab = "Weekly wages Manufacturing employees", names = c("Rural", "Urban"))
```



```
boxplot(data$wsta~data$urban, ylab = "Weekly wages State employees", names = c("Rural", "Urban"))
```



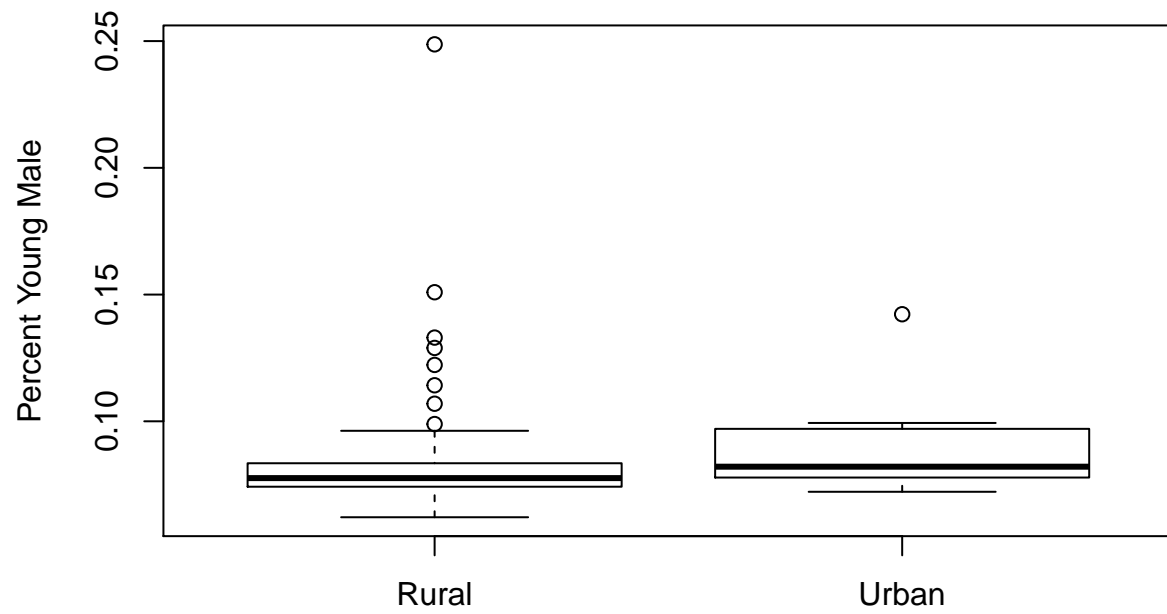
```
boxplot(data$wloc~data$urban, ylab = "Weekly wages Local Government employees", names = c("Rural", "Urban"))
```



I cannot see why percent young male would have anything to do with urban versus rural. But i'll check it and all other variables too. Nothing's too different based on urban versus rural. So the correlation of these variables with crimerate seems unrelated to anything that has to do with rural versus urban.

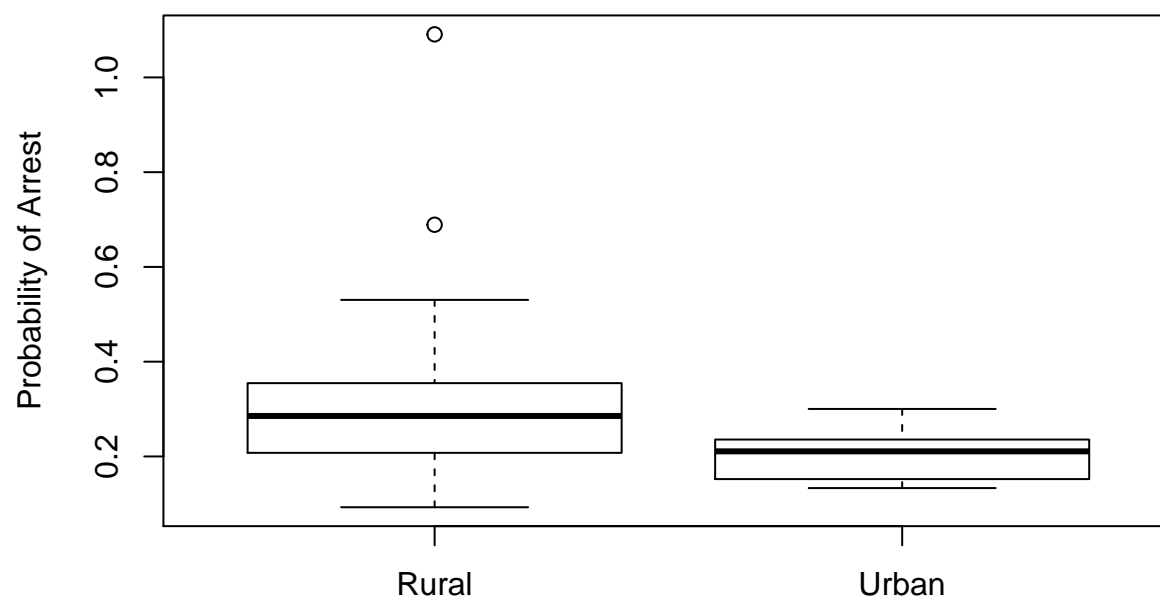
```
boxplot(data$pctymle~data$urban, ylab = "Percent Young Male", names = c("Rural", "Urban"), main = "Perce
```

Percent young male in Urban versus Rural Counties



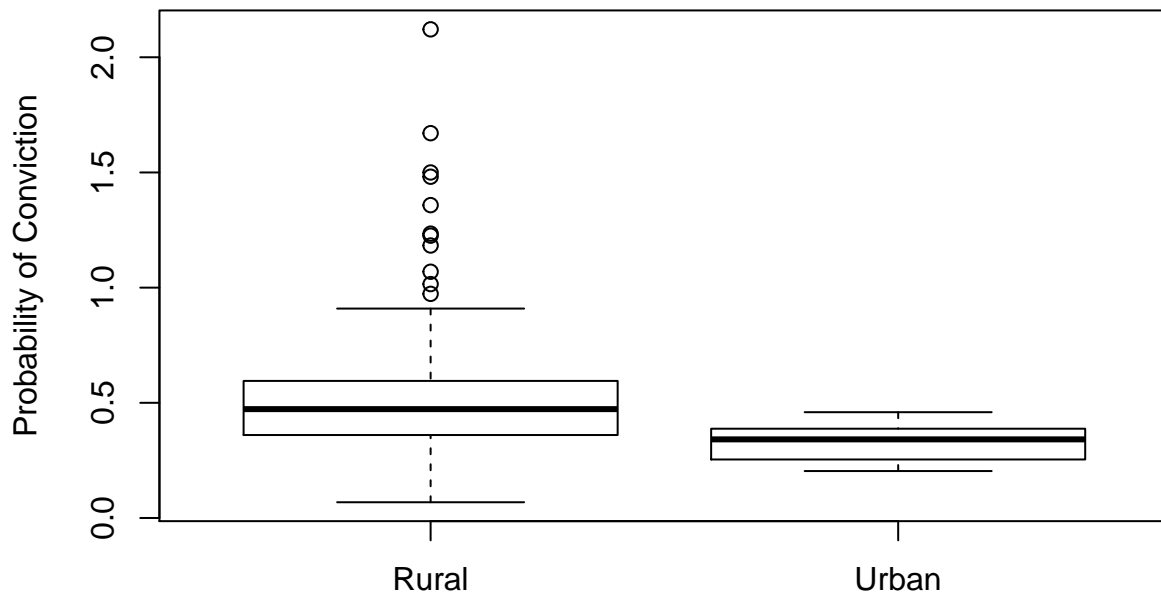
```
boxplot(data$prbarr~data$urban, ylab = "Probability of Arrest", names = c("Rural", "Urban"), main = "Pr
```


Probability of arrest in Urban vs. Rural Counties



```
boxplot(data$prbcon~data$urban, ylab = "Probability of Conviction", names = c("Rural", "Urban"), main =
```

Probability of Conviction in Urban vs. Rural Counties



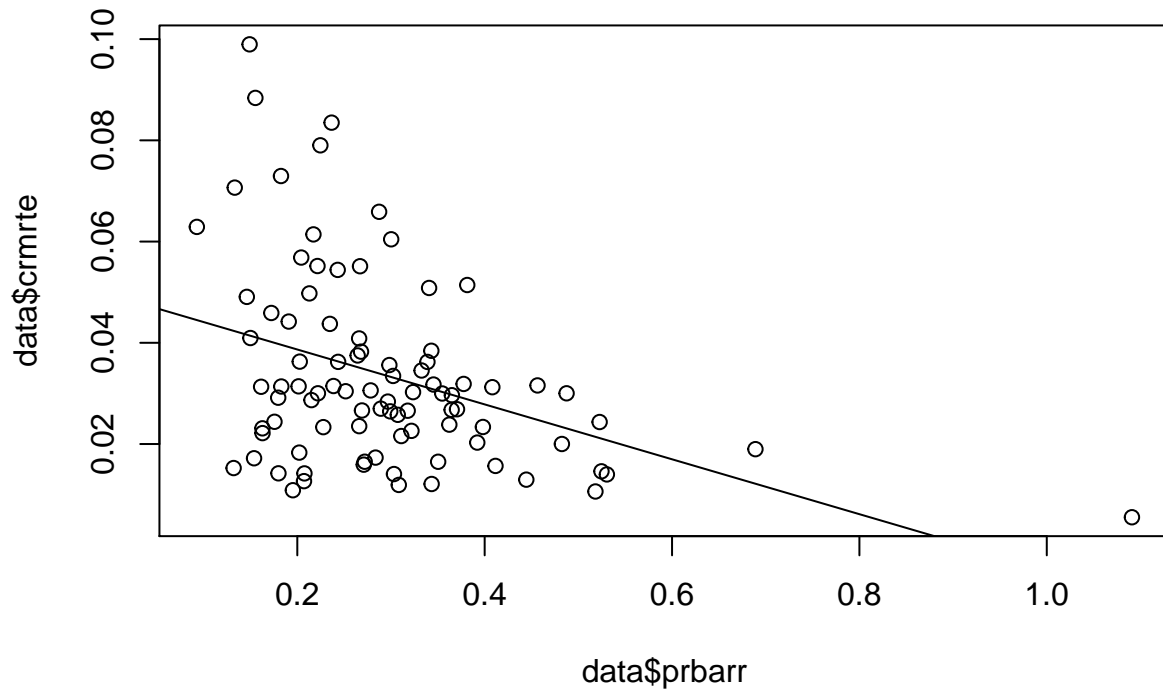
In general, urban areas have a higher crime rate. And it seems like the trappings of urbanity (higher salaries, higher density, and higher tax per capita) then correlate with crime rate. Perhaps it is the density part of urban areas that is causative for higher crime, but that's not something a politician can alter.

I also just wanted to look more closely at the bivariate relationships between crime rate and those variables that correlate with crime rate...just to see if the relationships look linear. I cannot really tell. We can discuss!

Below (last graph), it seems like there is a weak correlation between police per capita and crime -it shows up in your corr plots. But it is the opposite of what we expect, so we should probably address it. And maybe we put it in our model actually?

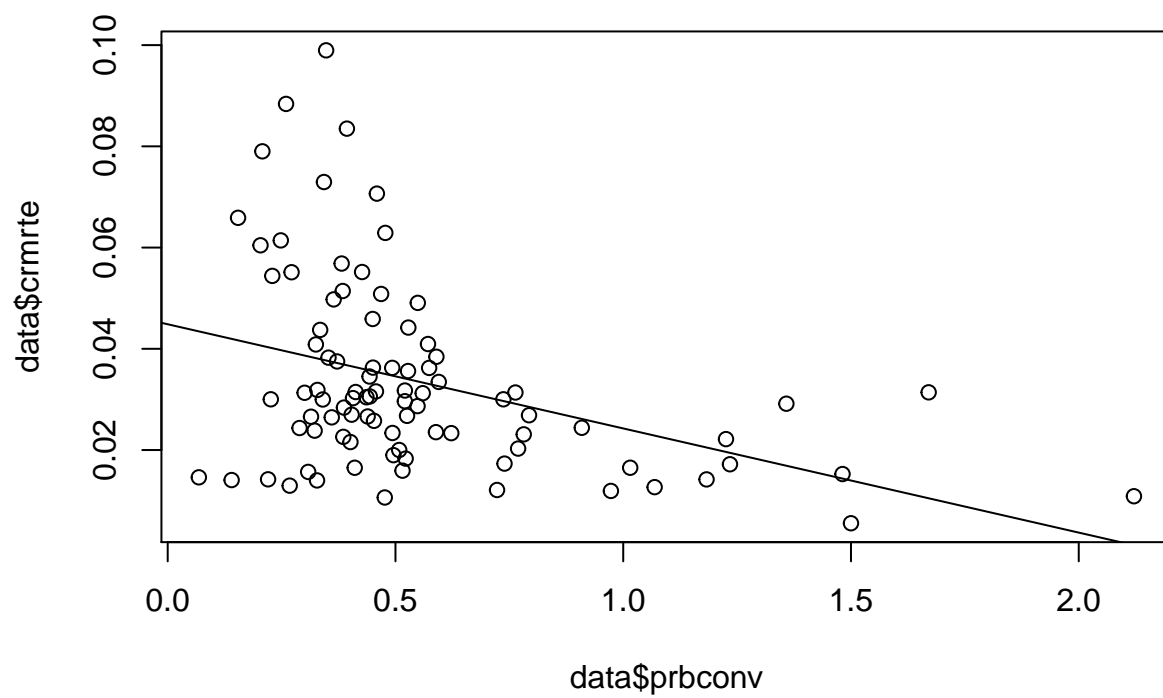
```
plot(data$prbarr,data$crmrte, main="Crime rate v. Prob. Arrest")
abline(lm(data$crmrte ~ data$prbarr))
```

Crime rate v. Prob. Arrest



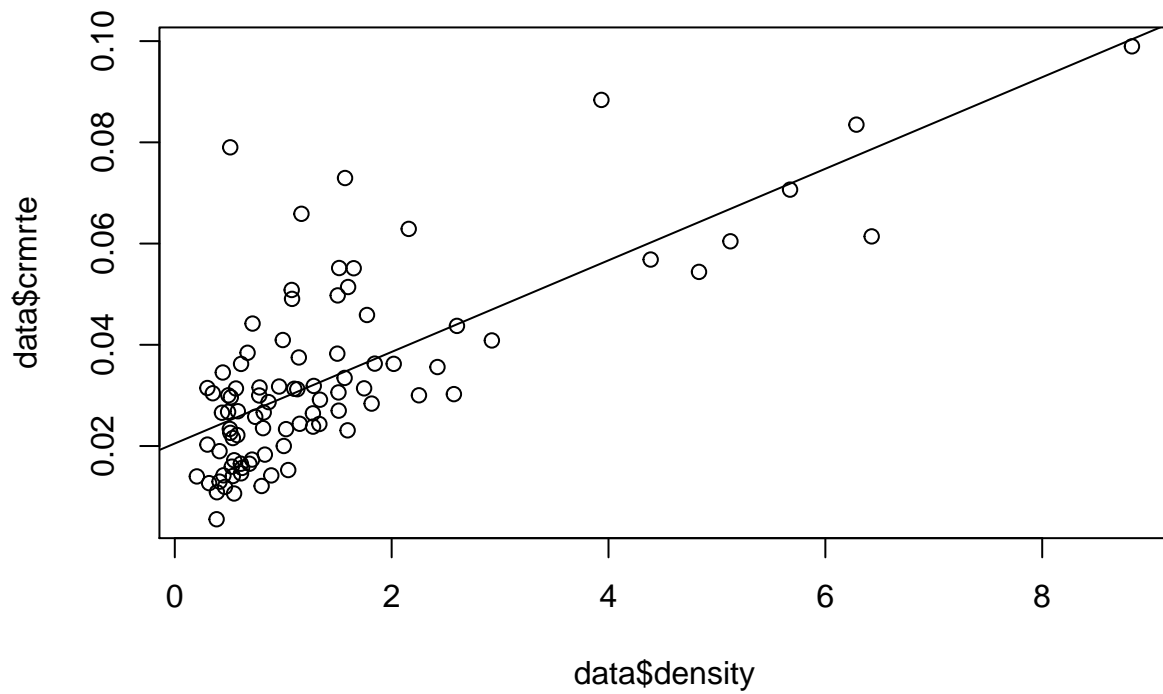
```
plot(data$prbconv,data$crmrte, main="Crime rate v.Prob. Conviction")
abline(lm(data$crmrte ~ data$prbconv))
```

Crime rate v.Prob. Conviction



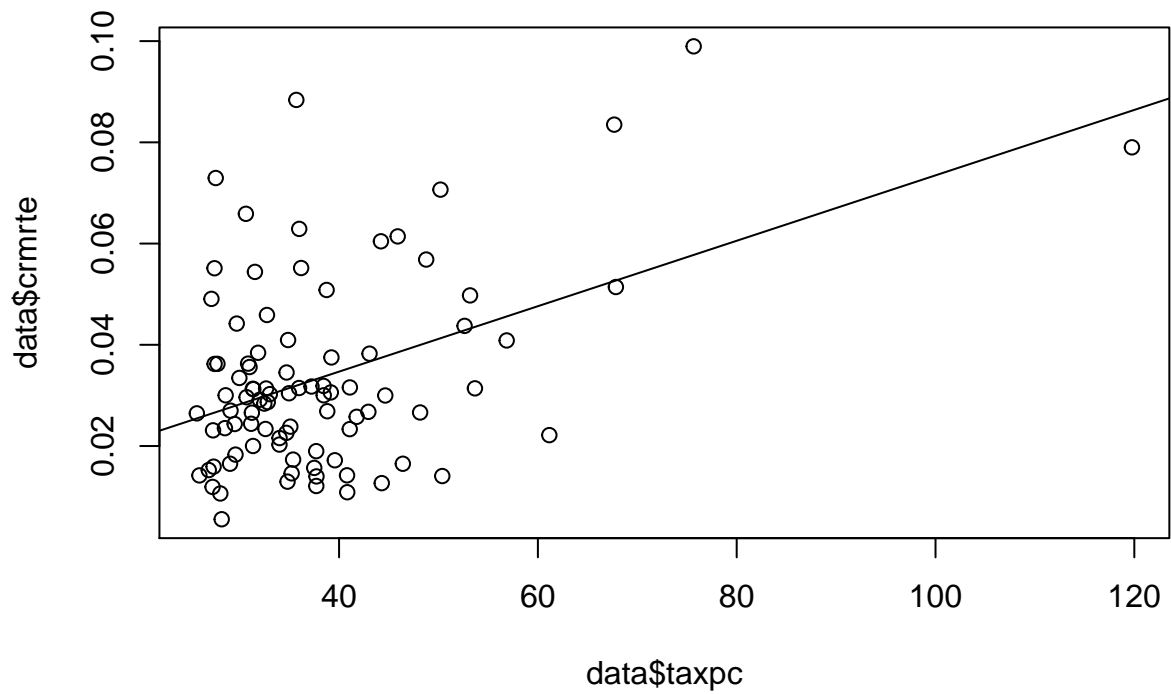
```
plot(data$density,data$scrmrte, main="Crime rate v. Density")
abline(lm(data$scrmrte ~ data$density))
```

Crime rate v. Density



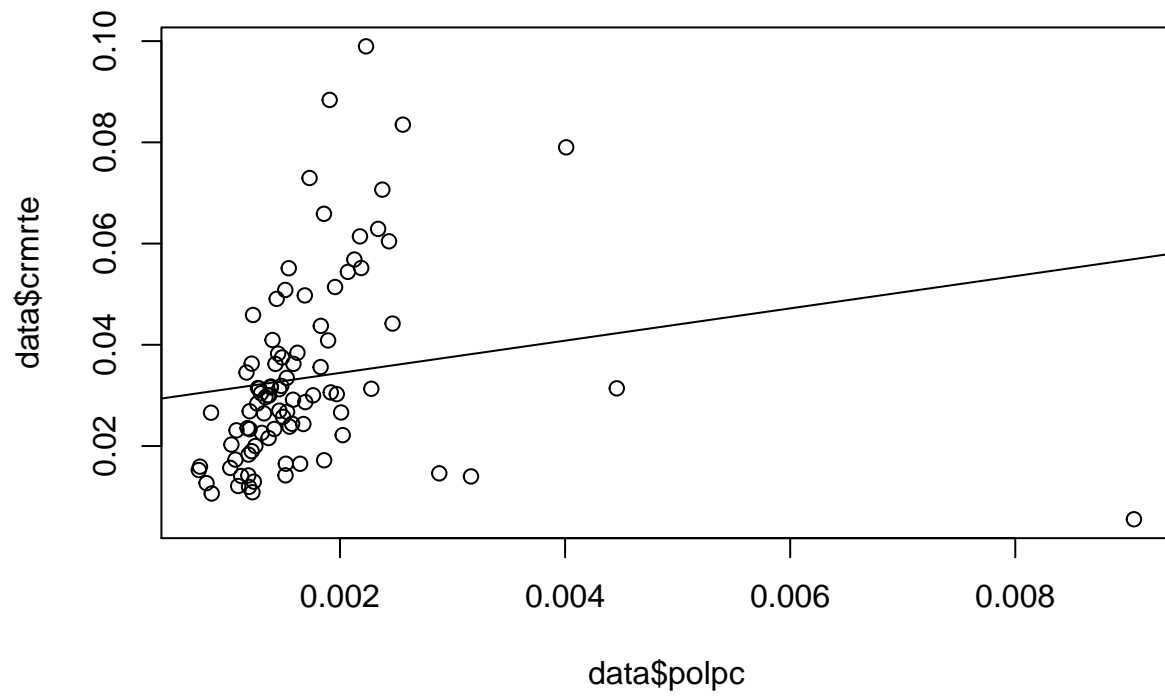
```
plot(data$taxpc, data$scrmrte, main="Crime rate v. tax per capita")  
abline(lm(data$scrmrte ~ data$taxpc))
```

Crime rate v. tax per capita



```
plot(data$polpc,data$scmrte, main="Crime rate v. Police per capita")  
abline(lm(data$scmrte ~ data$polpc))
```

Crime rate v. Police per capita



```
cor(data$crmrte,data$polpc )
```

```
## [1] 0.1672816
```