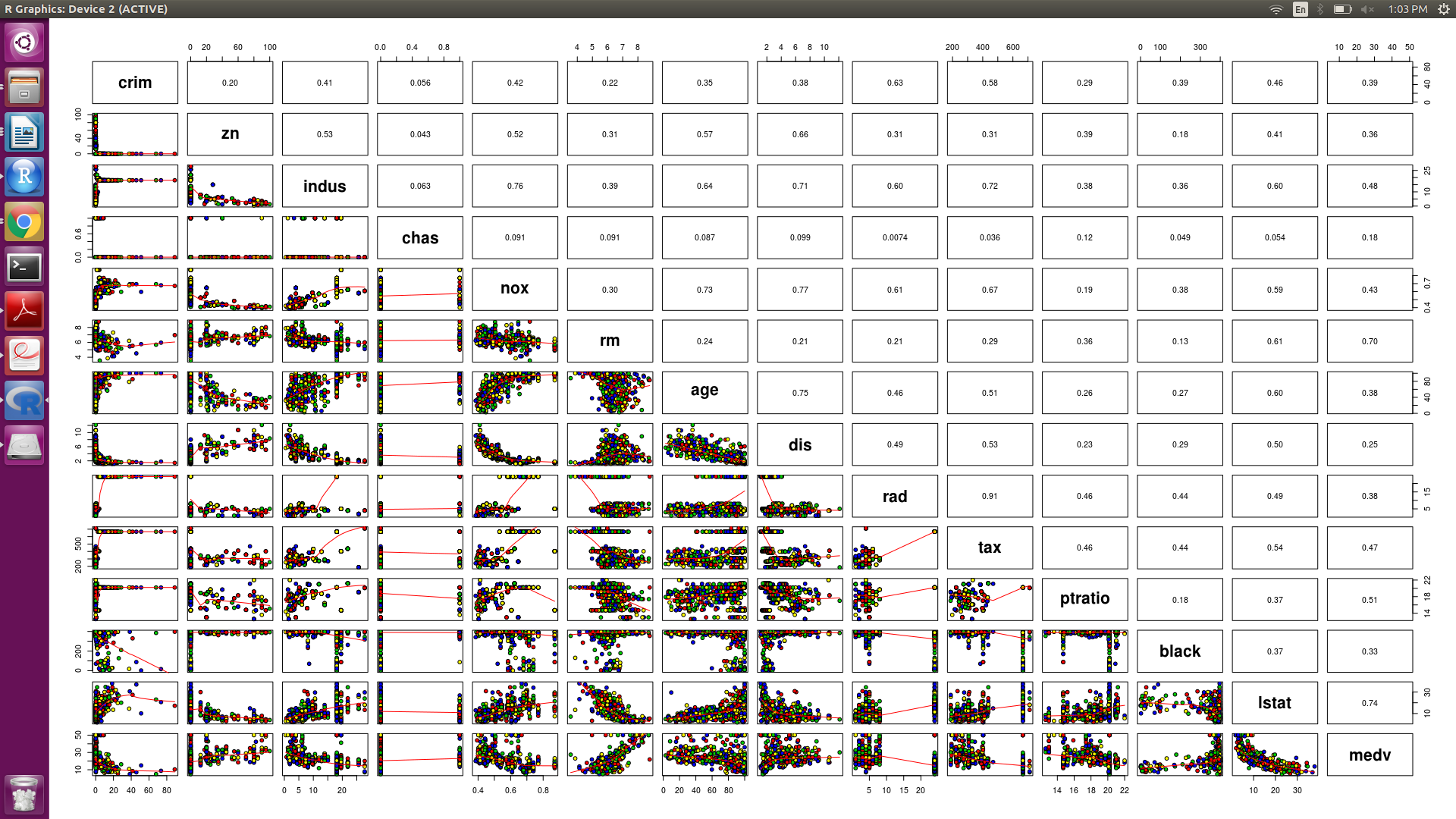
**4) Tasks:**

**a) Make Pairwise scatterplots of predictors:**



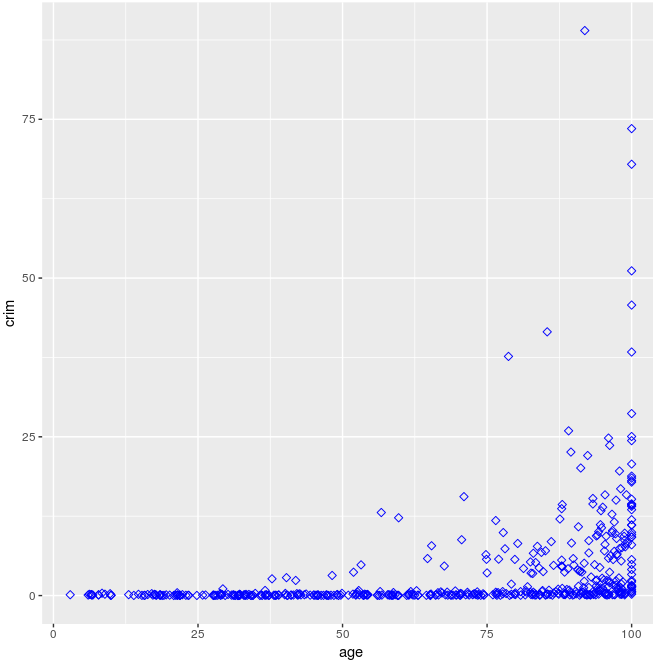
The above image contains relationship between all the predictors for Boston Dataset. The upper diagonal values are the Correlation among all the predictors. From these very descriptive graph we can give below analysis:

* ‘**rad**’ and ‘**tax**’ has the maximum correlation of 0.91 which implies they are strongly associated with each other with positive relationship.
* ‘**dis**’ and ‘**nox**’ has the correlation of 0.77 with -ve response so as we move away from the employment area the Nitric Oxide concentration in air starts reducing.
* ‘**Age**’ and ‘**Distance**’ has strong relationship we move away from the employment city areas the age starts decreasing.
* Per capita crime rate is fairly associated with Distance, Tax, Medv.
* Perc Lower status of popoulation i.e. lstat is inversely related to the distance of the employment areas.

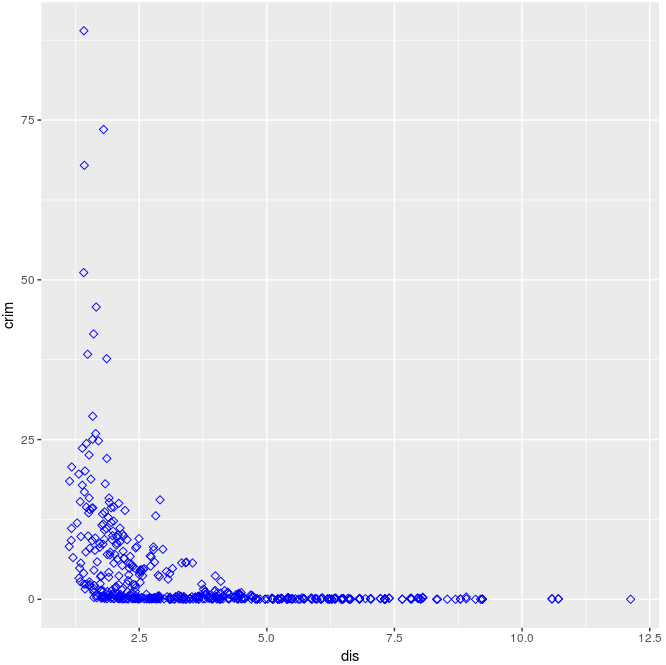
**b) Association with per capita crime rate:**

From the graph it is implied that the Crime rate correlates with rad, tax, age,distance and ptratio.

Crime vs Age: As the age of houses is increasing the crime rate is also increasing.



Crime vs Distance from employed cities:

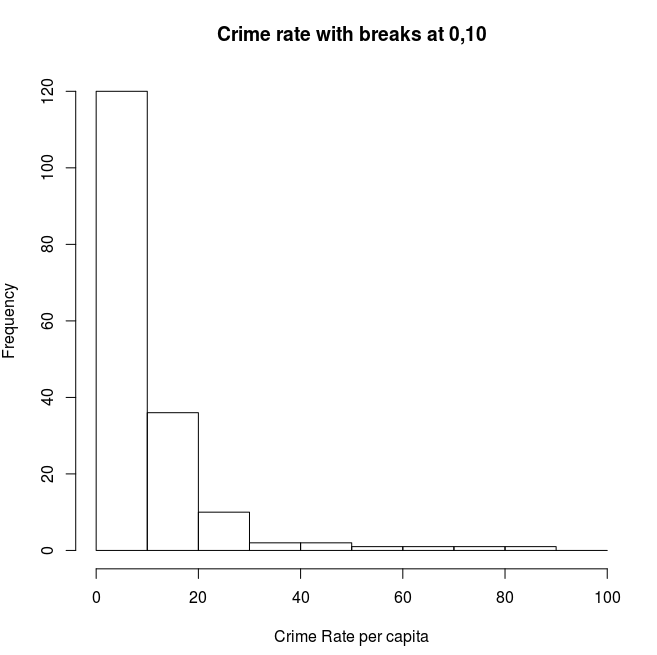


It can be said that the Crime rate is more for the localities near the employed cities, and decreases in the outer areas.

**c)** Here, we need to find which suburb has the high crime rates,Tax rates and Pupil-teacher ratios. From dataset we have suburbs as if Boston$chas=1 then near river else away from river.

**Crime rates:**

hist(Boston$crim[Boston$crim>1], breaks = 0 + (0:10)\*10, xlab ="Crime Rate", main = "Boston crime rate with breaks at 0,10")



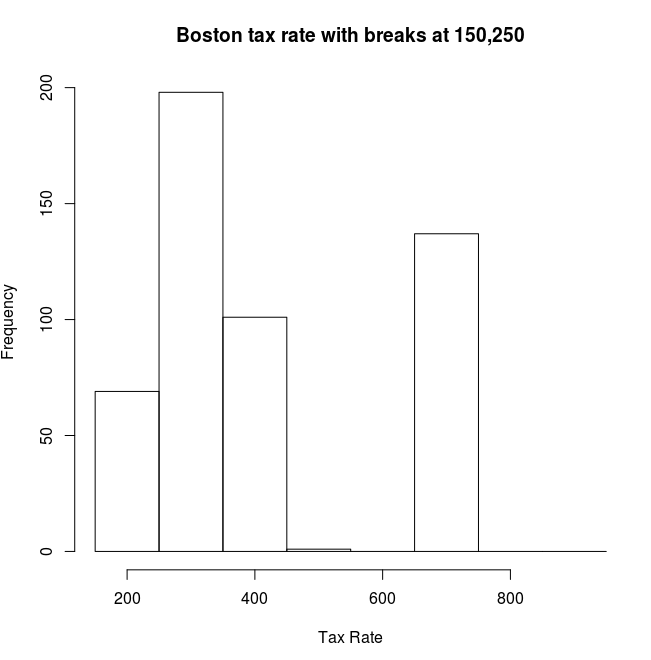
Keeping 20 as the mark point we are considering that the crime rates above 20 are high, so now we are finding suburbs for high crime rates.

length(Boston$chas[Boston$crim>20] & Boston$chas==1)  
[1] 0

length(Boston$chas[Boston$crim>20] & Boston$chas==0)  
[1] 18  
> min(Boston$crim[Boston$crim>1])  
[1] 1.00245  
> max(Boston$crim[Boston$crim>1])  
[1] 88.9762

The range is 1-88. And from the above results we can conclude that the high crime rates is absent for the suburbs near the river but 1 suburbs which are away from river are having high crime rates going from 20 till 90.

**Tax:**



Keeping 400 as the mark point we are considering that the crime rates above 400 are high, so now we are finding suburbs for high Tax rates.

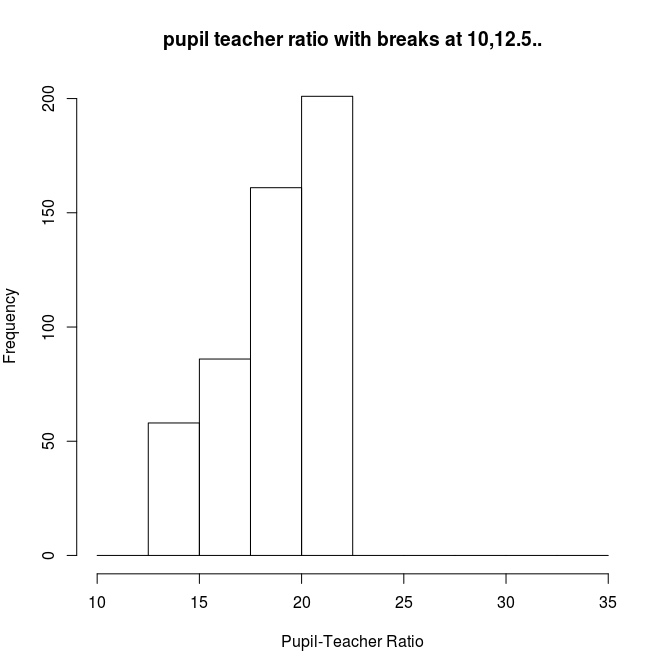
length(Boston$chas[Boston$tax>400] & Boston$chas==1)  
[1] 15

length(Boston$chas[Boston$tax>400] & Boston$chas==0)  
[1] 185  
> max(Boston$tax[Boston$tax>1])  
[1] 711  
> min(Boston$tax[Boston$tax>1])  
[1] 187

The range is 187-711. So, mostly all the suburbs have low tax rate i.e below 400. There are 185 suburbs which are away from river and has a very high tax rate i.e above 400 and reaching till 711. There are only 15 suburbs which are near river and having high tax rate.

**Pupil-teacher ratio:**

hist(Boston$ptratio[Boston$ptratio>1], breaks = 10 + (0:10)\*2.5, xlab ="Pupil-Teacher Ratio", main = "pupil teacher ratio with breaks at 10,12.5..")



> min(Boston$ptratio[Boston$ptratio>1])  
[1] 12.6  
> max(Boston$ptratio[Boston$ptratio>1])  
[1] 22

The range is 12.6 to 22 and from the graph we can see that beyond 27 there is no higher ptratio.

**d)**

nrow((data[data$rm > 7 & data$chas == 0,])) >> 56

nrow((data[data$rm > 7 & data$chas == 1,])) >> 8

nrow((data[data$rm > 8 & data$chas == 0,])) >> 11

nrow((data[data$rm > 8 & data$chas == 1,])) >> 2

So there are 56 rooms away rooms away from the river and having more than 7 rooms per dwelling.  
There are 8 rooms near the river and having more than 7 rooms per dwelling  
There are 11 rooms away from the river and having more than 8 rooms per dwelling.  
There are only 2 rooms near the river which are having more than 8 rooms per dwelling.