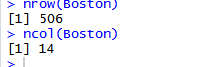
1. **Boston Housing Analysis**

I loaded the Boston dataset using MASS library.

The data describes attributes for towns in Boston.



There are data for 506 towns.

There are 14 columns:

CRIM - per capita crime rate by town

ZN - proportion of residential land zoned for lots over 25,000 sq.ft.

INDUS - proportion of non-retail business acres per town.

CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)

NOX - nitrogen oxides concentration (parts per 10 million)

RM - average number of rooms per dwelling

AGE - proportion of owner-occupied units built prior to 1940

DIS - weighted distances to five Boston employment centres

RAD - index of accessibility to radial highways

TAX - full-value property-tax rate per $10,000

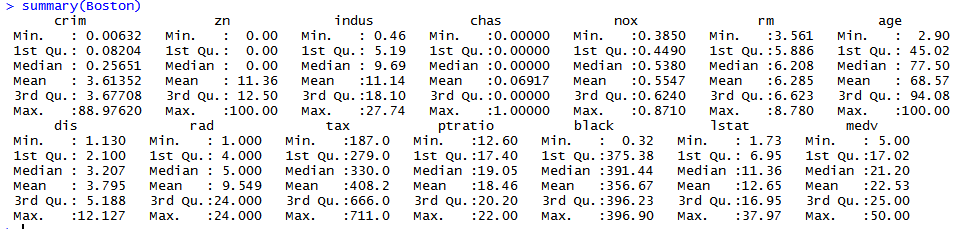
PTRATIO - pupil-teacher ratio by town

B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

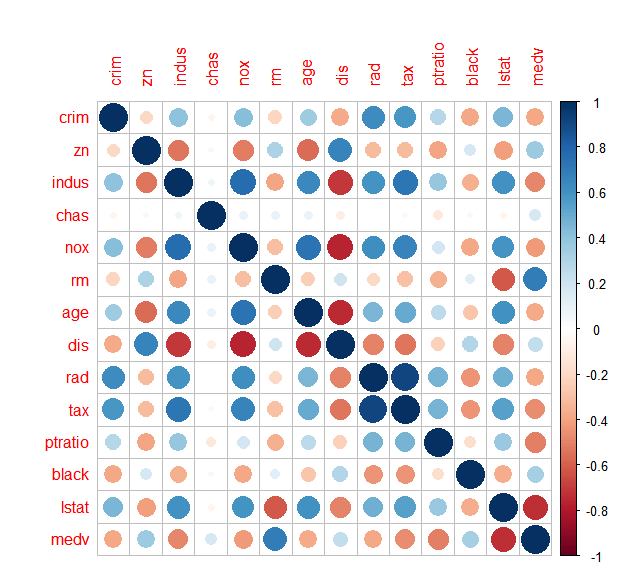
LSTAT - % lower status of the population

MEDV - Median value of owner-occupied homes in $1000's

Below is the summary statistics of all columns. We will be referring to this as we analyse further.



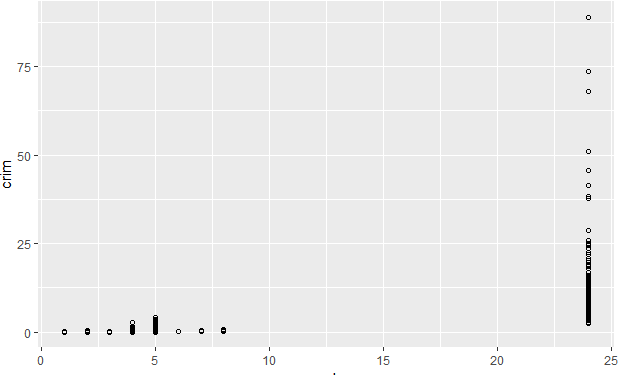
Another metric we’ll be using is the correlation matrix. I generated this using corrplot library:



A big blue circle indicates positive correlation and a big red circle indicates negative correlation. Smaller circles indicate not much correlation. These values however are not definite, and further analysis may be required to fully unearth the relations. However, they are very helpful in general.

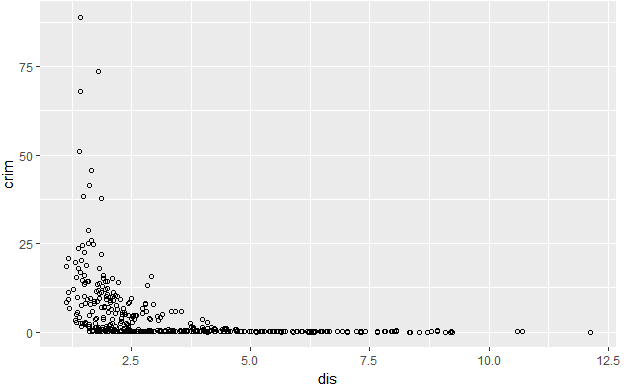
A few interesting scatter plots:

**Rad vs Crim**



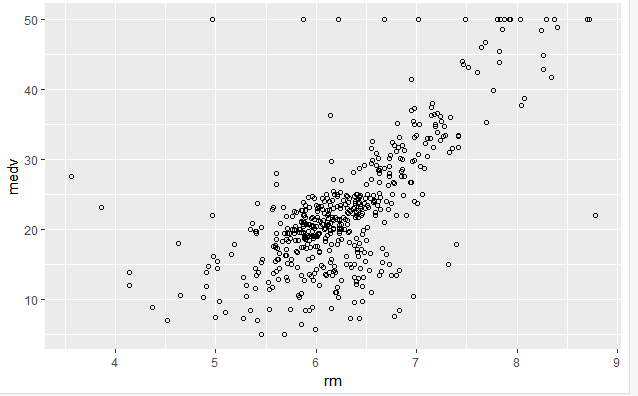
We see that as connectivity increases the crime rate has increased too. We see a crime rate of more than 5 only when RAD is 24. More access to Radial Highways mean an urbanized town. More urbanization leading to more crimes may be the case here.

**Dis vs crime**



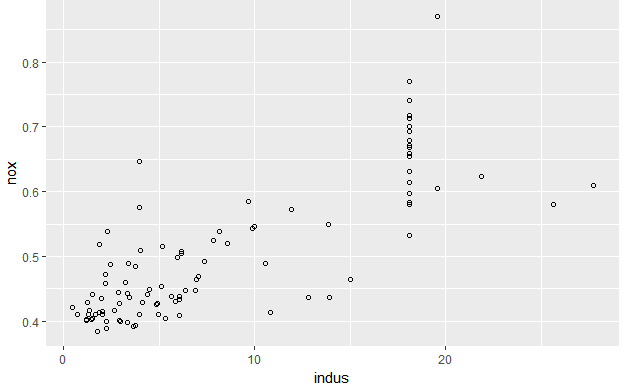
We see that in towns with high crime rate the distance to employment centers is low. Perhaps the introduction of more employment centers may be thought of as a solution to high crime rate.

**RM vs MEDV**

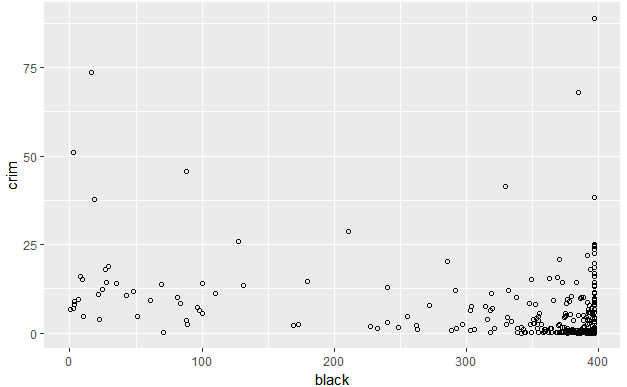


The plot of RM vs MEDV shows that as the average number of rooms in a dwelling increase the median value of houses also increase. This is a very straightforward conclusion. However, we can see that there is strong linear correlation beween the two.

**INDUS vs NOX**

The plot shows that as the proportion of industries or non-retial businesses increase the Nitrogen oxide concentration(a pollutant) increases.

**Black vs Crim**

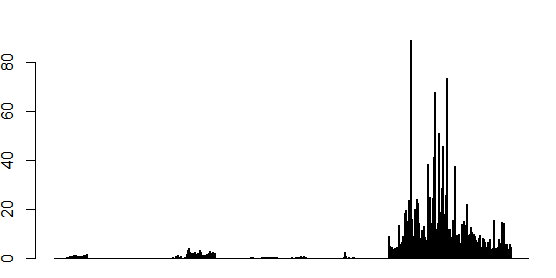


There is a stereotype that is maintained by some people that black communities are prone to high crime rate. The above graph debunks this claim. We can see that crime rate has no dependency on black population. We see the lower right of the graph thickly populated indicating that high black population with low crime rate is prevalent in Boston.

**Predictors associated with per capita crime rate.**

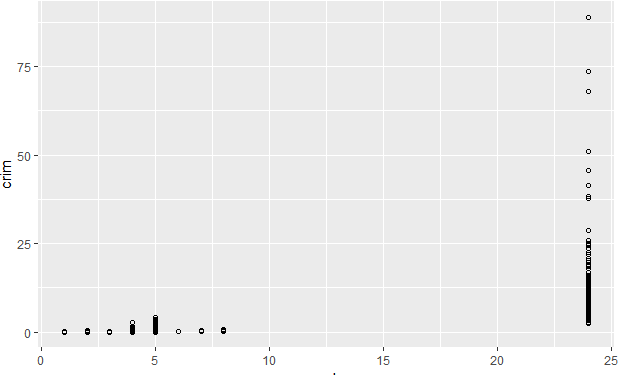
Given below is the bar plot of Crime Rate . On the X-axis are each town and on the Y-axis the crime rates.

From this we understand crime rates are low for 75% of the towns and it is exceptionally high for some. The order in which the data was collected also seem important because we see high crime rates toward the end.

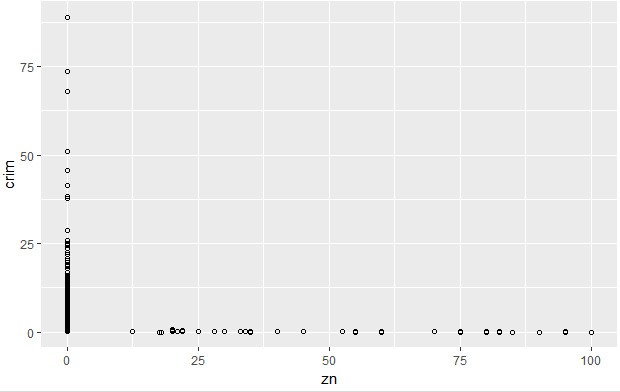


Now let’s see what can influence Crime rate.

RAD vs CRIM

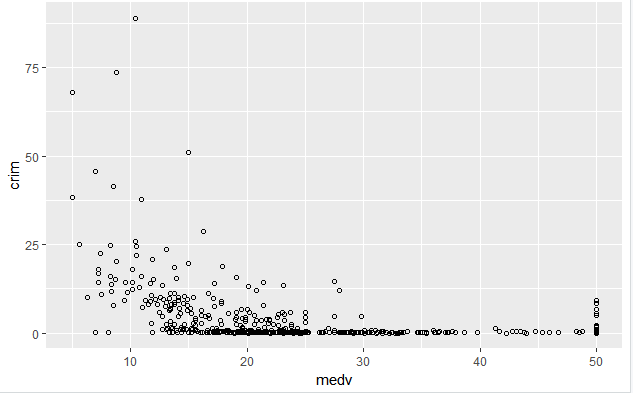


We can count access to radial highways, that is , higher urbanization, as a factor that increases crime rate.



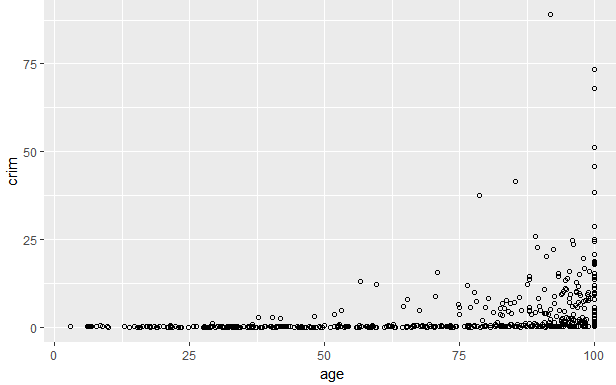
This plot shows that crime rate higher than normal happens only when residential zoning has no been done at all. Hence we can say this affects crime rate too. More than the number of zones we can see that whether zoning is done or not can be a good measure.

Another good predictor is median value of homes.

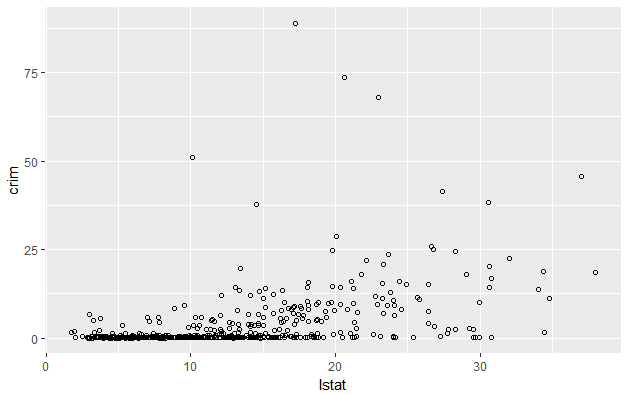


From above we see that a very high crime rate happens only in towns where **medv** is lower than 25 (25,000 $).

Below is a scatter plot of **Age vs Crim.**  We see that crime rate is low when the newr buildings (homes) are built in the town.



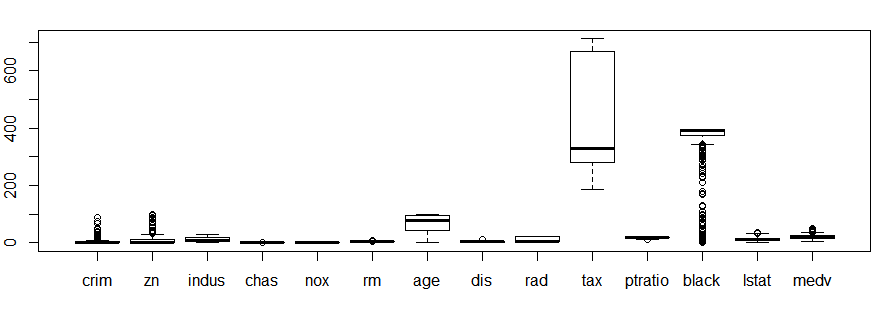
Finally, here is the plot for lstatvs crim. It shows that the percentage of people in lower status increases crime rate. We can interpret this as - more poverty gives rise to more crime.



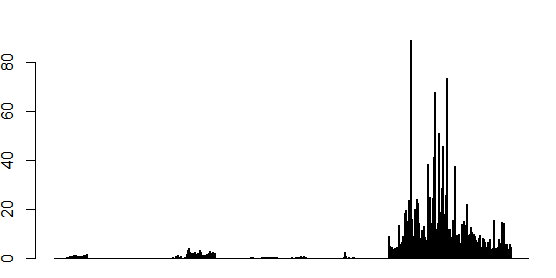
So to conclude, some of the factors that can help predict crime rate are - Access to highways,zoning, value of homes, whether new dwellings are built, number of people in lower status.

**Do any towns have higher crime rates? Tax Rates? PT ratios?**

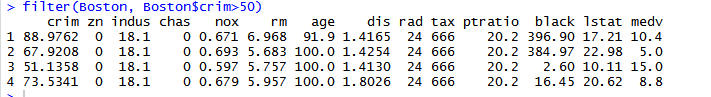
Here s the box plot for all columns:



We see that there many towns with unusual crime rates. With crime rate , it would seem that, a few towns have very high crime rate, while most towns enjoy a very low crime rate as is shown below.

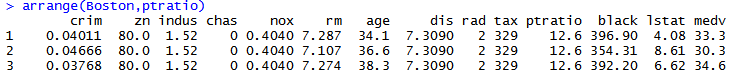


We filter out towns with crime rate greater than 50. We can observe that they have low median value for homes, good access to highways, no zoning, and close to 100 percent buildings built prior to 1940



All towns are within the interquartile when it comes to taxes.

Only three towns have an unusual PT ratio. It is lower than the interquartile range for these two. We sort the data by this column to see these rows.



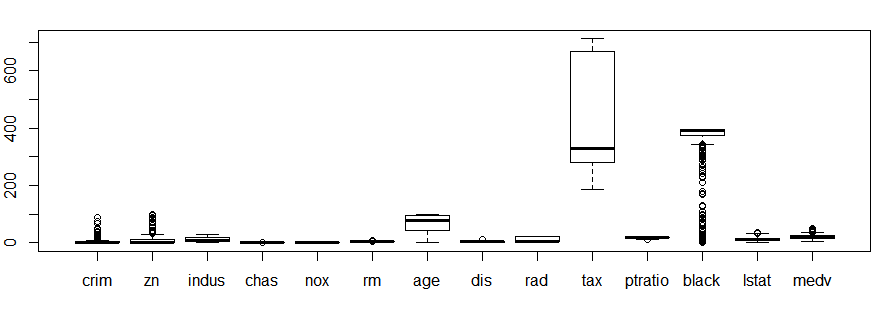
We can see that they are quite developed towns with low crime rate, high value for houses, and, a good proportion of new buildings.

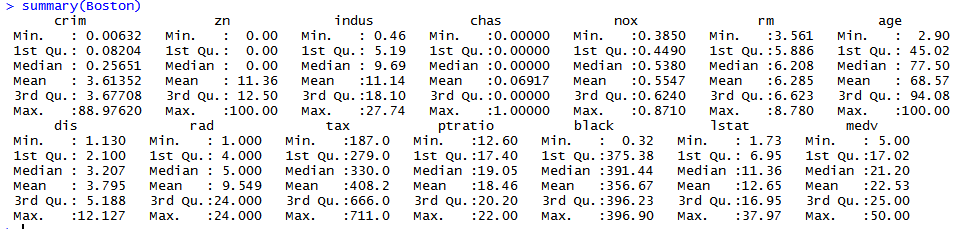
Coming to the range of other predictors, we see that :

1. Most towns have low crime rates and no zoning with a few exceptions.
2. Nitrogen Oxide concentration(betwen 0 and 1) is distributed among the towns.
3. The column **chas** indicates that only few towns bound the Charles river.
4. The column **INDUS** is well distributed between 0.46 and 27.74
5. The columns **RM, MEDV,LSTAT,DIS** has a few outliers.
6. The black population is very low for quite a number of towns.But for the most towns this number is near the median value.
7. **RAD** is an index and is not continuous. It ranges between 1 and 24. We can see from the plot that. We can find that there is a lot of variance in these values.



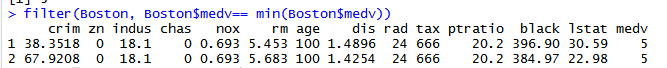
Box plot and Summary of the columns given below for reference:





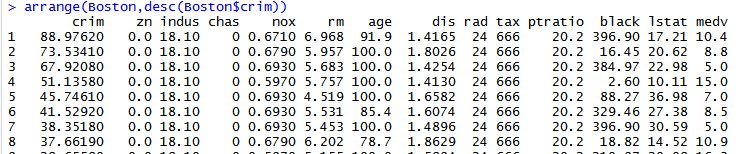
**Which Boston suburbs have lowest median value of owner occupied homes? Hows does the predictors of these suburbs compare to the general data.**

The records with the lowest median value of owner occupied homes are below:



Let’s analyze each predictor:

1. Wee see that the crime rate in these towns are very high.

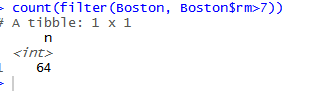


From the above data , ordered in descending order of crime rates , we see that the towns in question stand at 3rd and 7th in crime rate per capita.

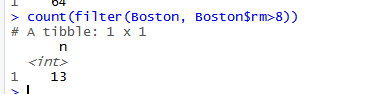
1. We see that no areas have been marked for residential zoning
2. There are many non-retail business per acre in this town. It is equal to the 3rd QR value(18.1)
3. Both towns are off the Charles river, and have the same high nitrogen oxide levels of 0.693. We can make the case that the two towns are neighbours.
4. The following columns - Black population proportion, RM(average number of rooms) column doesn’t deviate much from the normal
5. We see that all of the owner occupied homes were built before 1940s.
6. Employment centers are quite close by, compared to the other towns.
7. Access to radial highways is maximum among the dataset. Which indicates well connectivity.
8. We see that Tax Rate and PT ratio are also very high. Both are equal to third quartile value. These are bad indicators. Andd the fact that they are high indicates bad news for the towns.
9. The proportion of people with low status is high in these areas.

On the whole we can get the idea that the above towns experience high crime rate, is polluted, was settled in largely before 1940s, and with high tax rate and less number of teachers available,they are not a great option for newer settlers.

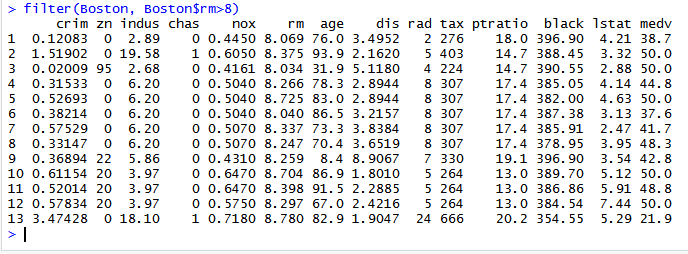
**How many suburbs average more than 7,8 rooms per dwelling?**



Number of suburbs with room average >7 is 64.



Number of suburbs with room average >8 is 13.



We can make two observations from the above data:

1. The median value of homes increase as number of rooms on average increase. The highest medv value of 50 also appears here.
2. The proportion of lower status people decrease when RM is 8. All the entries have a value less than medium. This indicates a general trend of prosperity.