

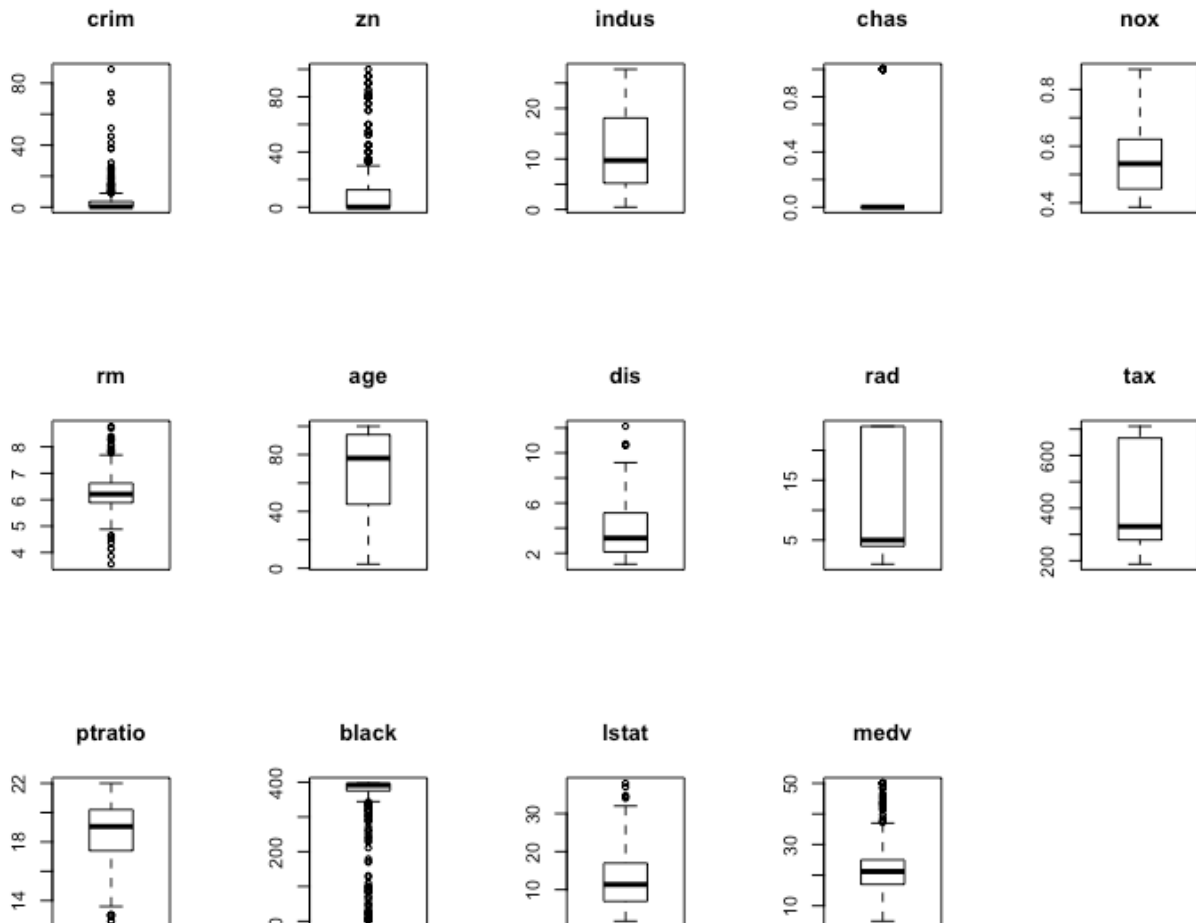
1.

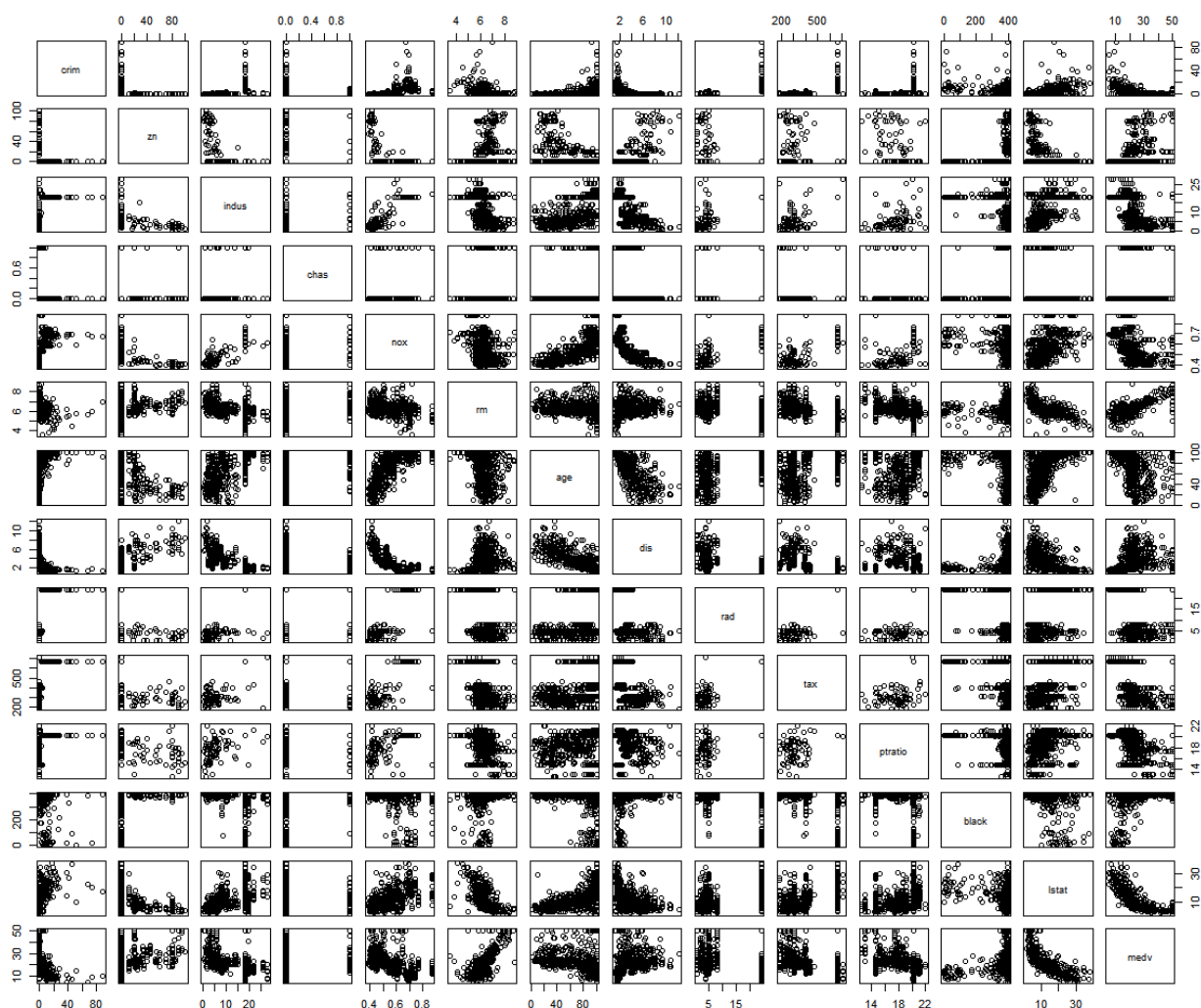
```
> summary(df)
```

crim	zn	indus	chas	nox	rm	age
Min. : 0.00632	Min. : 0.00	Min. : 0.46	Min. : 0.00000	Min. : 0.3850	Min. : 3.561	Min. : 2.90
1st Qu.: 0.08204	1st Qu.: 0.00	1st Qu.: 5.19	1st Qu.: 0.00000	1st Qu.: 0.4490	1st Qu.: 5.886	1st Qu.: 45.02
Median : 0.25651	Median : 0.00	Median : 9.69	Median : 0.00000	Median : 0.5380	Median : 6.208	Median : 77.50
Mean : 3.61352	Mean : 11.36	Mean : 11.14	Mean : 0.06917	Mean : 0.5547	Mean : 6.285	Mean : 68.57
3rd Qu.: 3.67708	3rd Qu.: 12.50	3rd Qu.: 18.10	3rd Qu.: 0.00000	3rd Qu.: 0.6240	3rd Qu.: 6.623	3rd Qu.: 94.08
Max. : 88.97620	Max. : 100.00	Max. : 27.74	Max. : 1.00000	Max. : 0.8710	Max. : 8.780	Max. : 100.00

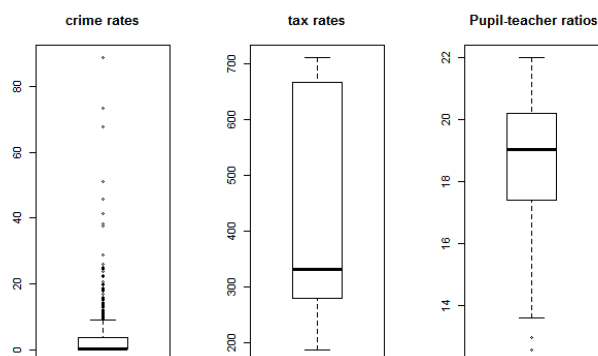
dis	rad	tax	ptratio	black	lstat	medv
Min. : 1.130	Min. : 1.000	Min. : 187.0	Min. : 12.60	Min. : 0.32	Min. : 1.73	Min. : 5.00
1st Qu.: 2.100	1st Qu.: 4.000	1st Qu.: 279.0	1st Qu.: 17.40	1st Qu.: 375.38	1st Qu.: 6.95	1st Qu.: 17.02
Median : 3.207	Median : 5.000	Median : 330.0	Median : 19.05	Median : 391.44	Median : 11.36	Median : 21.20
Mean : 3.795	Mean : 9.549	Mean : 408.2	Mean : 18.46	Mean : 356.67	Mean : 12.65	Mean : 22.53
3rd Qu.: 5.188	3rd Qu.: 24.000	3rd Qu.: 666.0	3rd Qu.: 20.20	3rd Qu.: 396.23	3rd Qu.: 16.95	3rd Qu.: 25.00
Max. : 12.127	Max. : 24.000	Max. : 711.0	Max. : 22.00	Max. : 396.90	Max. : 37.97	Max. : 50.00

2.





除了幾對變數有明顯的正相關、負相關之線性關係以外，age與lstat兩變數之分散圖可以發現，age越高lstat的值與範圍即越高，若要進行迴歸分析，則建議對lstat取log後，降低同age之lstat數值變異，避免R-square太小。另外以最右下角兩變數lstat與medv，其關係可能不僅是負相關，亦有可能有倒數關係或指數關係，可以另外以其他廣義線性迴歸分析該數據。



Crime rates 大多呈城市偏低，多在10以內，但是也不少城市有屬於離群值，遠高於10，最多有超過60的城鎮。稅率的部分約介於200至700，且無離群值。師生比的部分大多介於14至22之間，有兩個城鎮屬於偏低之離群值。

3. 所有連續的變數在常態檢定下皆為不顯著，故不服從常態分佈。

```
> conti.variable.p
      p-value of the Lillie test
crim      1.655823e-165
zn         1.214282e-261
indus      1.562590e-67
nox         1.006836e-14
rm          1.252791e-08
age         4.928761e-30
dis         1.043784e-23
tax          8.538468e-61
ptratio     1.097664e-43
black       1.020562e-173
lstat       1.639952e-10
medv        1.671245e-30
```

4. medv與rm有接近高度正相關，與lstat有高度負相關。

```
> corr.medv
      correlation coefficient
crim      -0.3883046
zn         0.3604453
indus      -0.4837252
nox         -0.4273208
rm          0.6953599
age         -0.3769546
dis         0.2499287
tax         -0.4685359
ptratio     -0.5077867
black       0.3334608
lstat       -0.7376627
```

5. 所有連續變數皆有顯著的影響，故拒絕係數為0之虛無假設。其餘變數見R code。

```
> cont.p
      p.value of the continuous predictor
crim      1.173987e-19
zn         5.713584e-17
indus      4.900260e-31
nox         7.065042e-24
rm          2.487229e-74
age         1.569982e-18
dis         1.206612e-08
tax         5.637734e-29
ptratio     1.609509e-34
black       1.318113e-14
lstat       5.081103e-88
```