Boston

October 23, 2019

[3]:
```
import numpy as np
```

[ ]:
```
import pandas as pd
```

[5]:
```
import matplotlib.pyplot as plt
```

[6]:
```
import matplotlib.mlab as mlab
```

[7]:
```
import seaborn as sns
```

[8]:
```
h_data=pd.read_csv('Boston.csv')
```

[13]:
```
h_data.head()
```

[13]:

	Unnamed: 0	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
0	1	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	2	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	3	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	4	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
4	5	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2

[17]:
```
h_data.shape
```

[17]:
```
(506, 15)
```

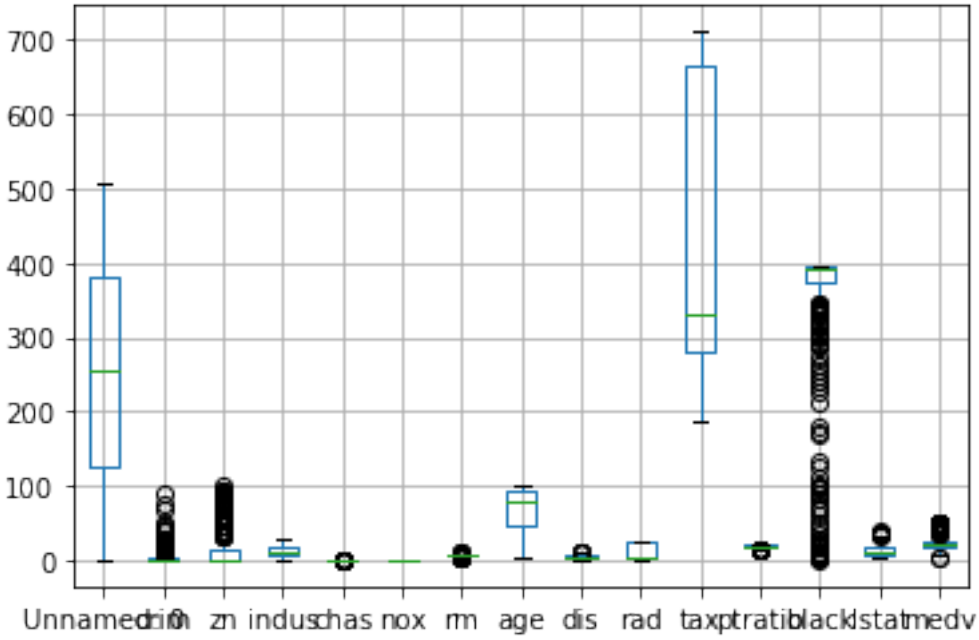
[18]:
```
h_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 506 entries, 0 to 505
Data columns (total 15 columns):
Unnamed: 0    506 non-null int64
```

```
crim          506 non-null float64
zn            506 non-null float64
indus         506 non-null float64
chas          506 non-null int64
nox           506 non-null float64
rm            506 non-null float64
age           506 non-null float64
dis           506 non-null float64
rad           506 non-null int64
tax           506 non-null int64
ptratio       506 non-null float64
black         506 non-null float64
lstat         506 non-null float64
medv          506 non-null float64
dtypes: float64(11), int64(4)
memory usage: 59.4 KB
```

[19]:
```
#figure 1
```

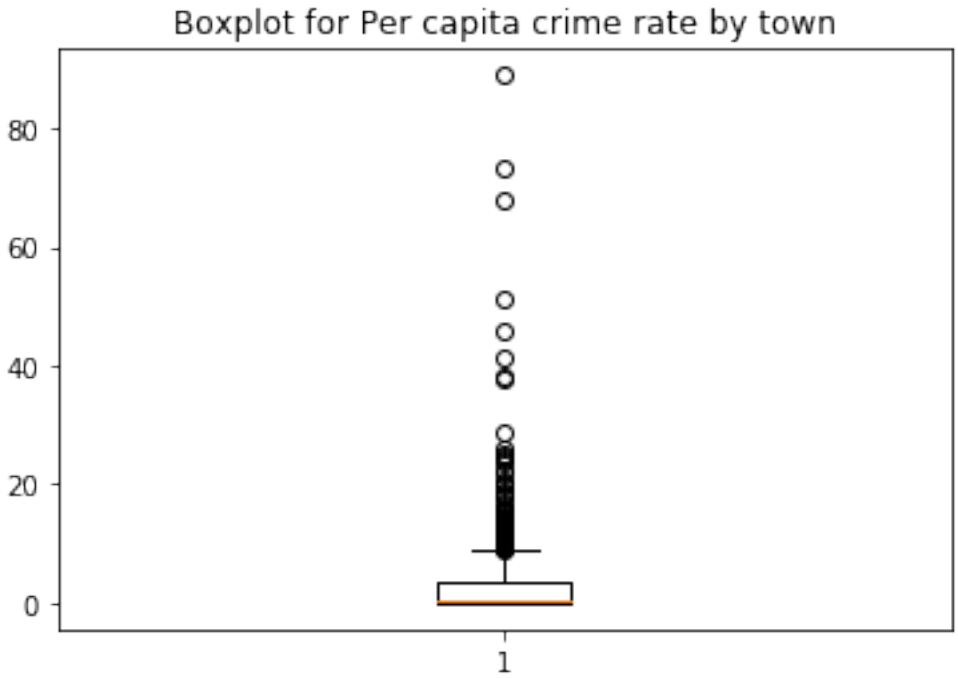
[21]:
```
h_data.boxplot()
plt.show()
```



[22]:
```
#figure 2
```

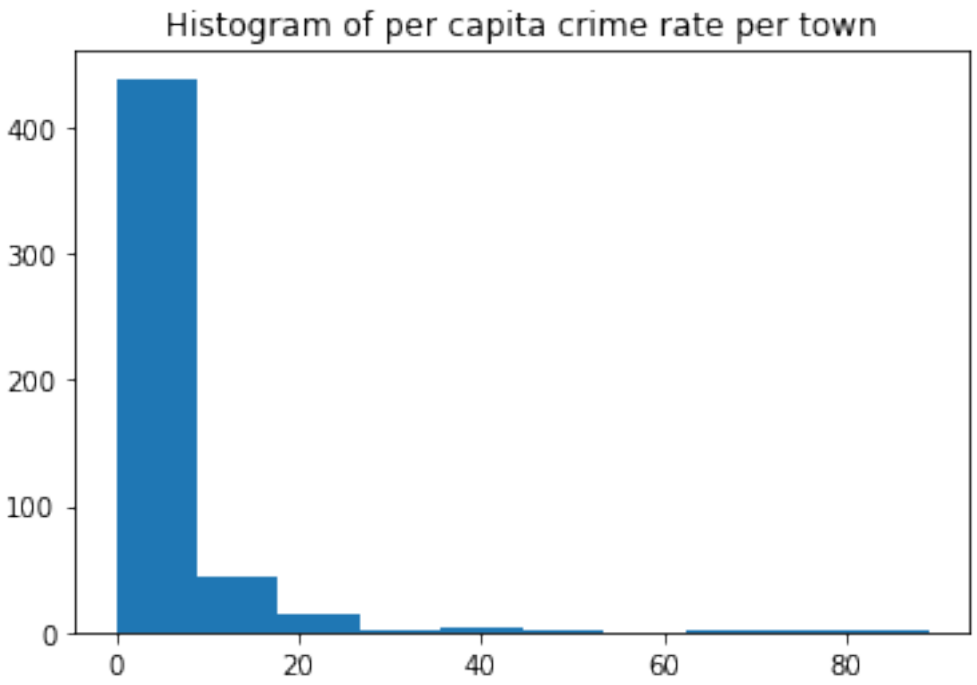
[23]:
```
plt.boxplot(h_data.crim)
plt.title('Boxplot for Per capita crime rate by town')
```

```
plt.show()
```

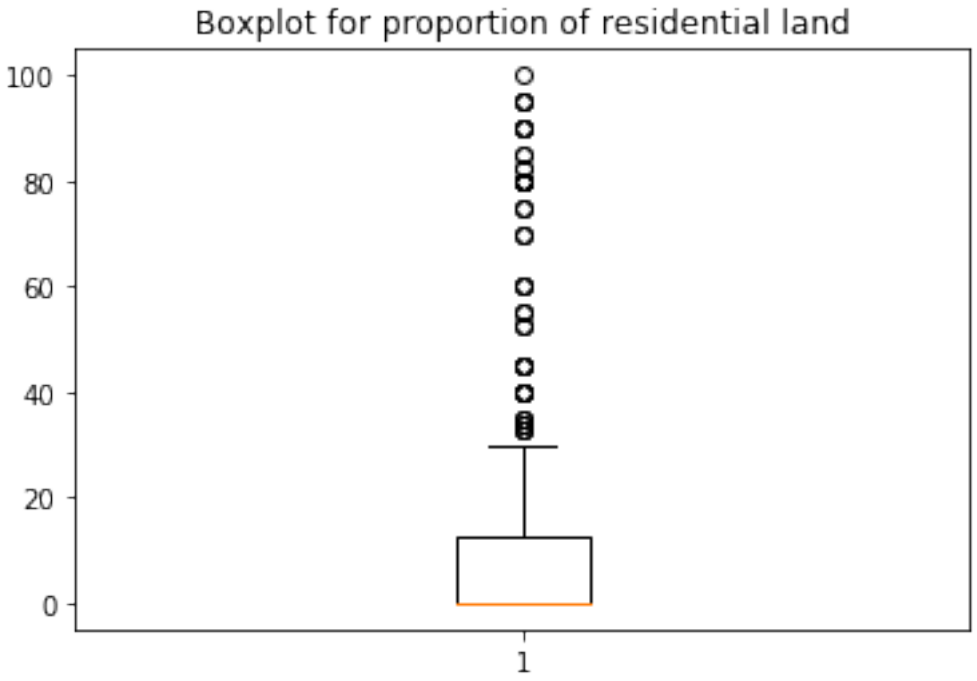


[24]:
```
#Figure 3
#Histogram
```

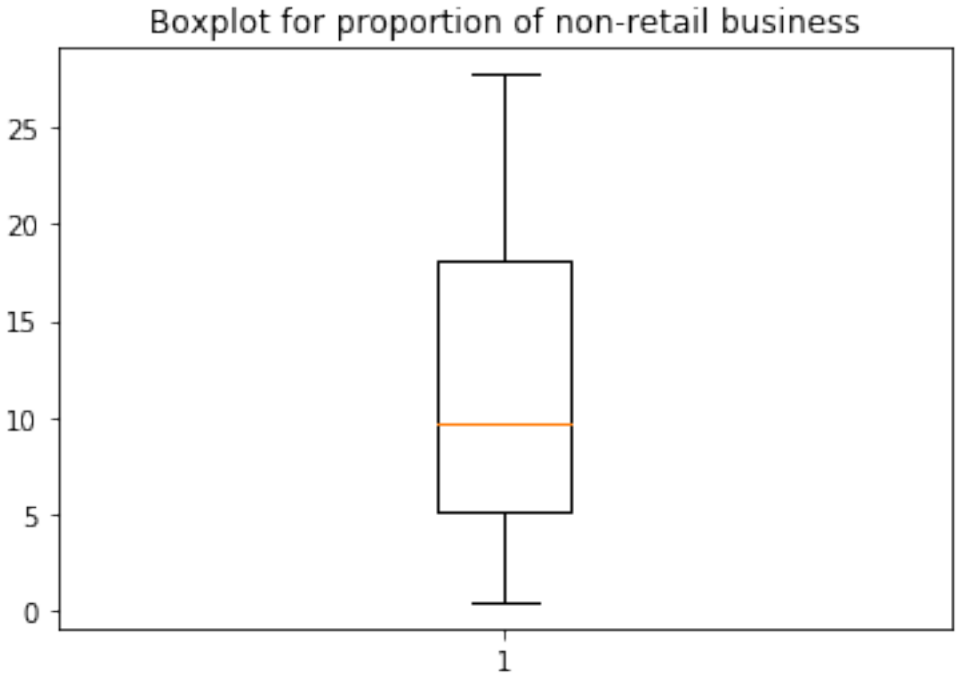
[33]:
```
plt.hist(h_data.crim, bins=10)
plt.title('Histogram of per capita crime rate per town')
plt.show()
```



[34]:
```
plt.boxplot(h_data.zn)
plt.title('Boxplot for proportion of residential land')
plt.show()
```



[35]:
```
plt.boxplot(h_data.indus)
plt.title('Boxplot for proportion of non-retail business')
plt.show()
```



[36]:
```
h_data['indus'].mean()
```

[36]: 11.13677865612648

[37]:
```
h_data['rm'].mean()
```

[37]: 6.284634387351779

[39]:
```
h_data['medv'].mean()
```

[39]: 22.532806324110677

[40]:
```
h_data['tax'].mean()
```

[40]: 408.2371541501976

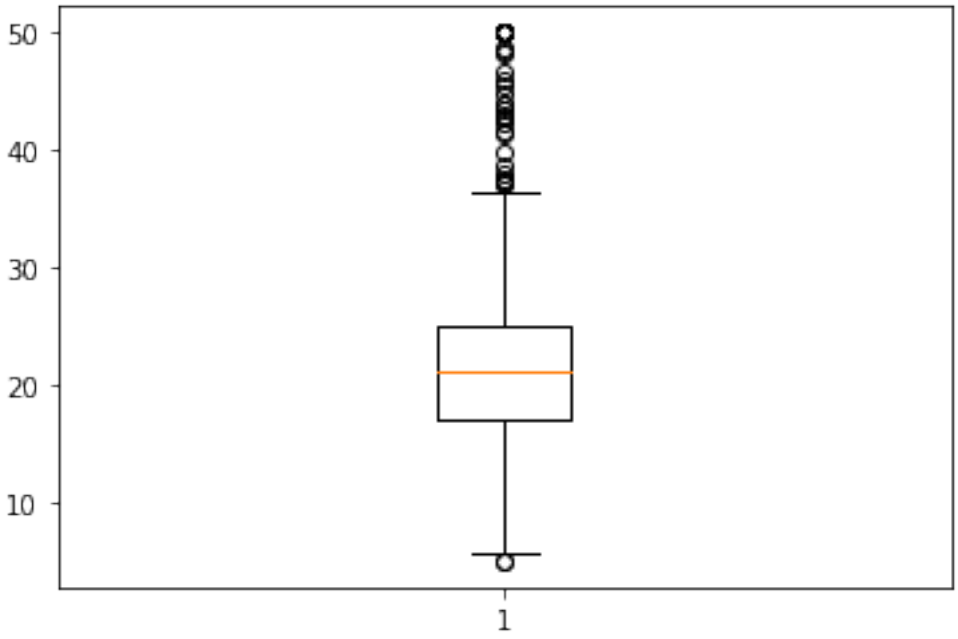
[41]:
```
h_data.describe()
```

[41]:
```
       Unnamed: 0        crim          zn       indus        chas         nox  \
count  506.000000  506.000000  506.000000  506.000000  506.000000  506.000000
mean   253.500000    3.613524   11.363636   11.136779    0.069170    0.554695
std    146.213884    8.601545   23.322453    6.860353    0.253994    0.115878
min      1.000000    0.006320    0.000000    0.460000    0.000000    0.385000
25%    127.250000    0.082045    0.000000    5.190000    0.000000    0.449000
50%    253.500000    0.256510    0.000000    9.690000    0.000000    0.538000
75%    379.750000    3.677082   12.500000   18.100000    0.000000    0.624000
max    506.000000   88.976200  100.000000   27.740000    1.000000    0.871000

              rm         age         dis         rad         tax     ptratio  \
count  506.000000  506.000000  506.000000  506.000000  506.000000  506.000000
mean     6.284634   68.574901    3.795043    9.549407  408.237154   18.455534
std      0.702617   28.148861    2.105710    8.707259  168.537116    2.164946
min      3.561000    2.900000    1.129600    1.000000  187.000000   12.600000
25%      5.885500   45.025000    2.100175    4.000000  279.000000   17.400000
50%      6.208500   77.500000    3.207450    5.000000  330.000000   19.050000
75%      6.623500   94.075000    5.188425   24.000000  666.000000   20.200000
max      8.780000  100.000000   12.126500   24.000000  711.000000   22.000000

            black       lstat        medv
count  506.000000  506.000000  506.000000
mean   356.674032   12.653063   22.532806
std     91.294864    7.141062    9.197104
min      0.320000    1.730000    5.000000
25%    375.377500    6.950000   17.025000
50%    391.440000   11.360000   21.200000
75%    396.225000   16.955000   25.000000
max    396.900000   37.970000   50.000000
```

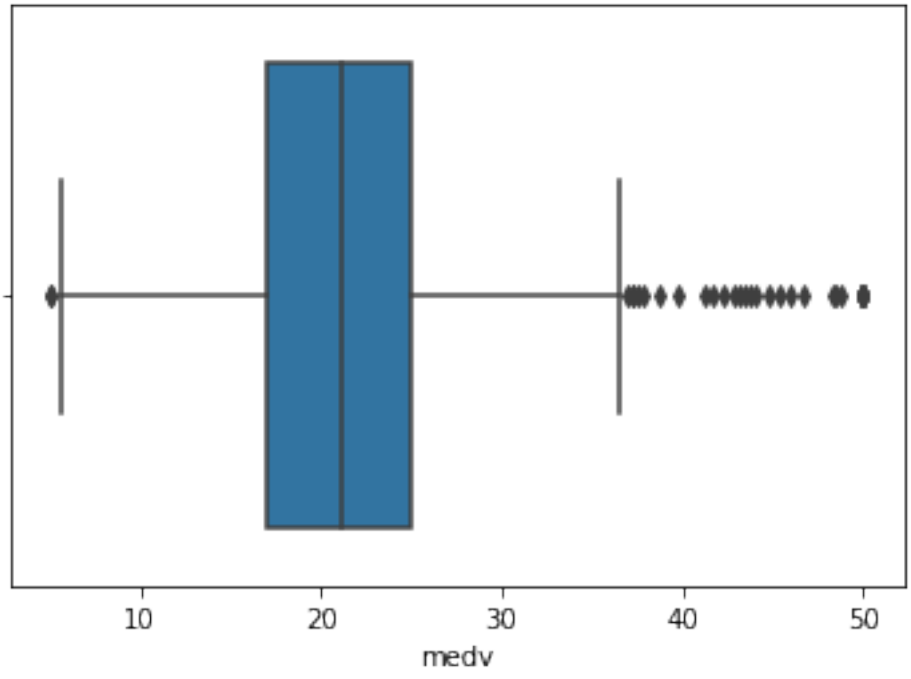
[42]:
```
h_data['medv'].describe()
```

[42]:
```
count    506.000000
mean      22.532806
std        9.197104
min        5.000000
25%       17.025000
50%       21.200000
75%       25.000000
max       50.000000
Name: medv, dtype: float64
```

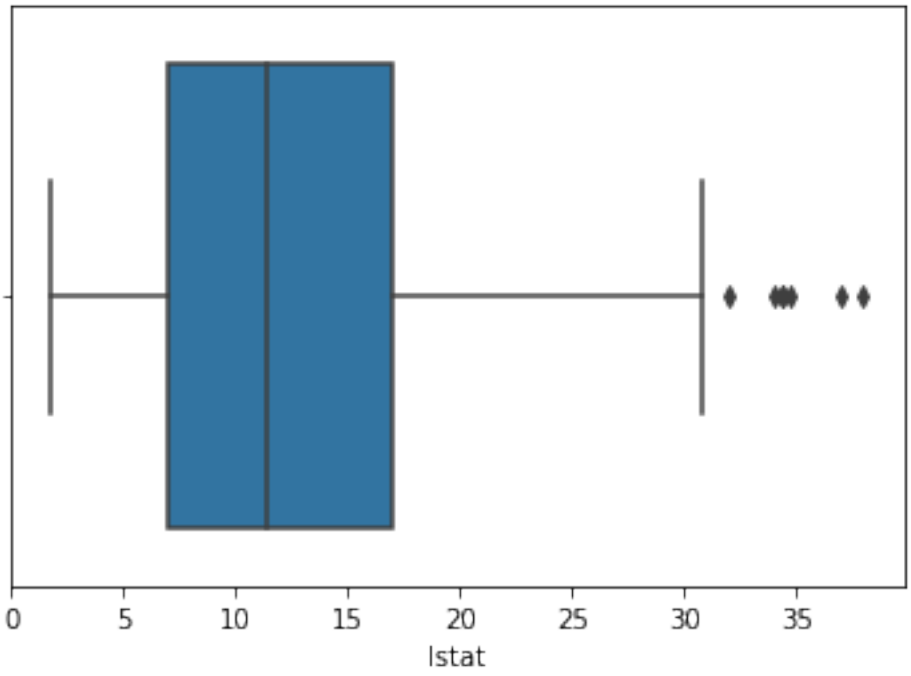
[43]:
```
plt.boxplot(h_data.medv)
plt.show()
```



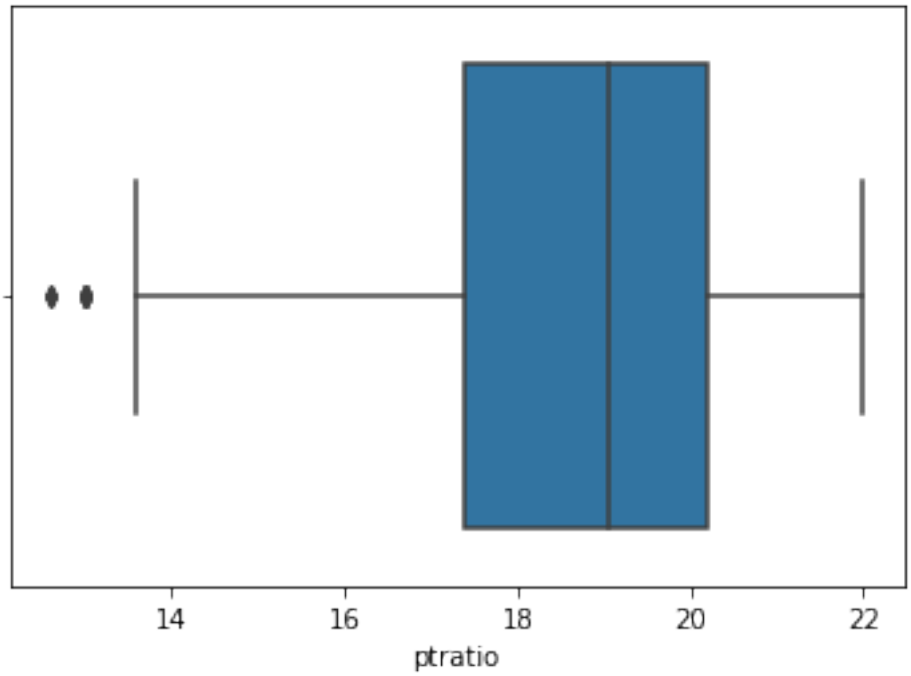
[45]:
```
sns.boxplot(h_data.medv)
plt.show()
```



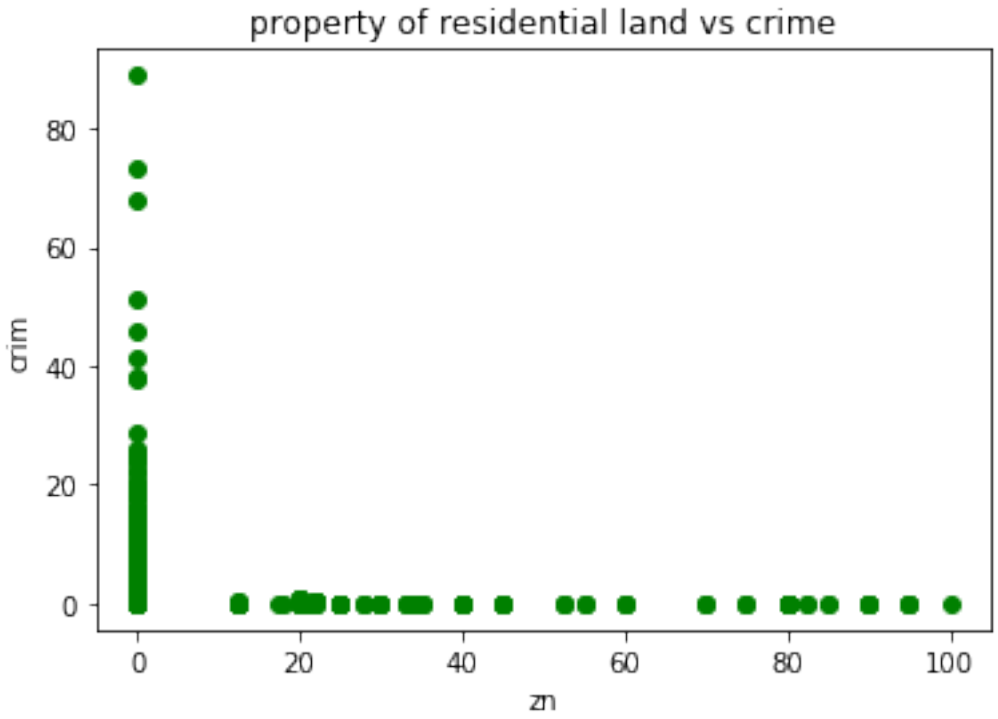
[47]:
```
sns.boxplot(h_data.lstat)
plt.show()
```



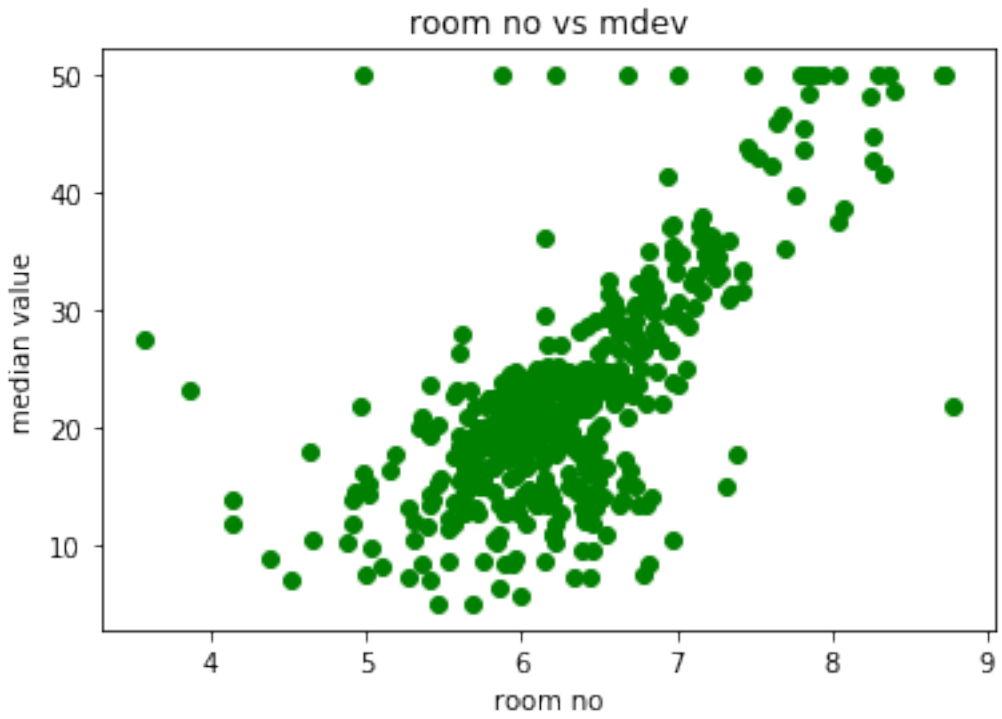
[49]:
```
sns.boxplot(h_data.ptratio)
plt.show()
```



[50]:
```
plt.scatter(h_data['zn'],h_data['crim'], color = 'g')
plt.xlabel('zn')
plt.ylabel('crim')
plt.title('proportion of residential land vs crime')
plt.show()
```

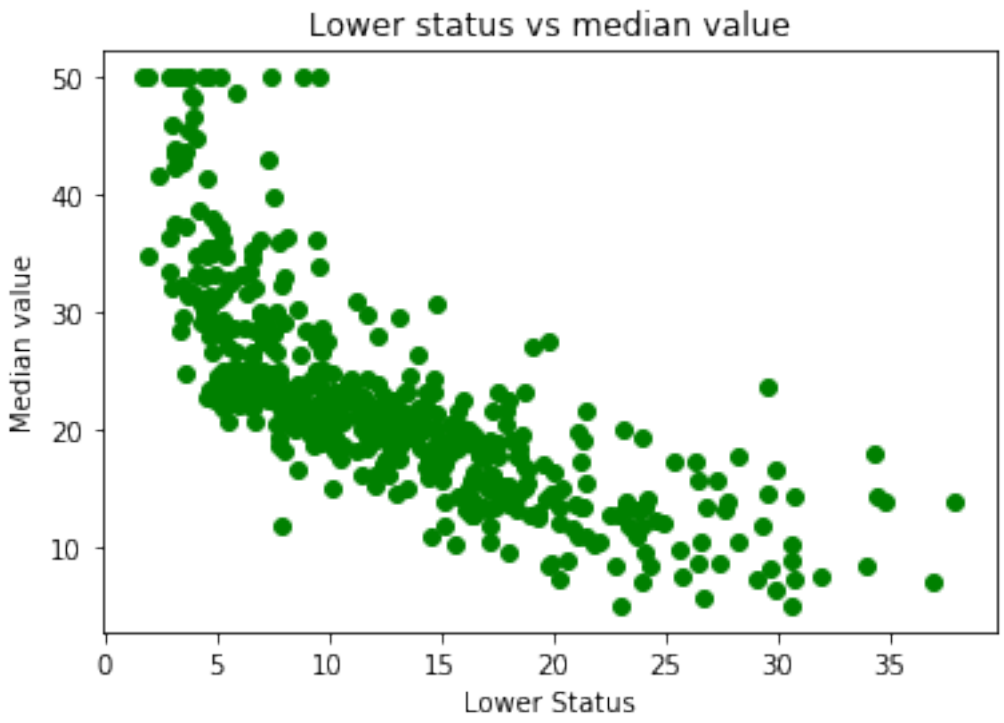


[11]:
```
#Scatter Diagram
plt.scatter(h_data['rm'],h_data['medv'], color = 'g')
plt.xlabel('room no')
plt.ylabel('median value')
plt.title('room no vs mdev')
plt.show()
```



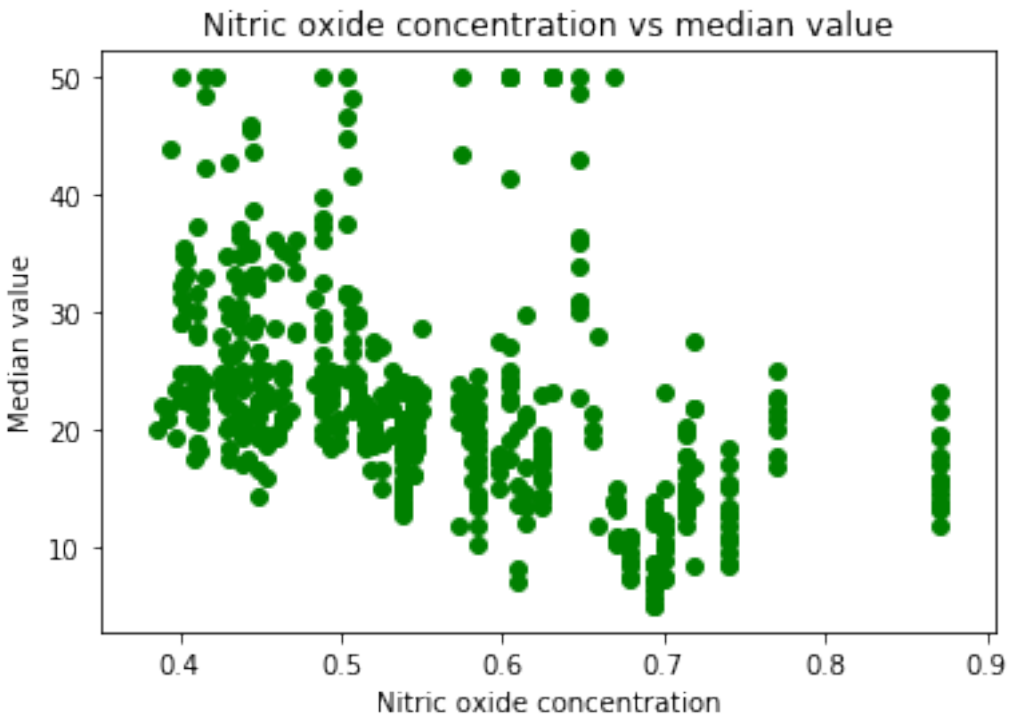
Strong positive coorelation, as the number of rooms increase/decrease, the housing price increases/decreases

[12]:
```
#Scatter Diagram
plt.scatter(h_data['lstat'],h_data['medv'], color = 'g')
plt.xlabel('Lower Status')
plt.ylabel('Median value')
plt.title('Lower status vs median value')
plt.show()
```



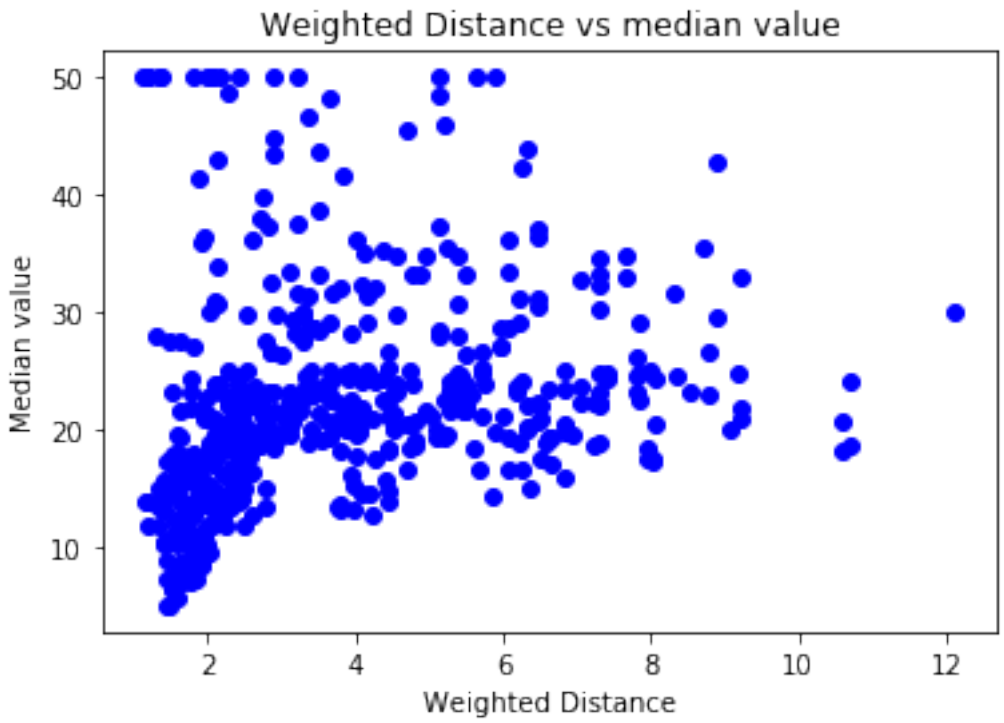
Strong negative coorelation, the more/less the population consists of lower status individuals,housing price decreases/increases.

[16]:
```
#Scatter Diagram
plt.scatter(h_data['nox'],h_data['medv'], color = 'g')
plt.xlabel('Nitric oxide concentration')
plt.ylabel('Median value')
plt.title('Nitric oxide concentration vs median value')
plt.show()
```



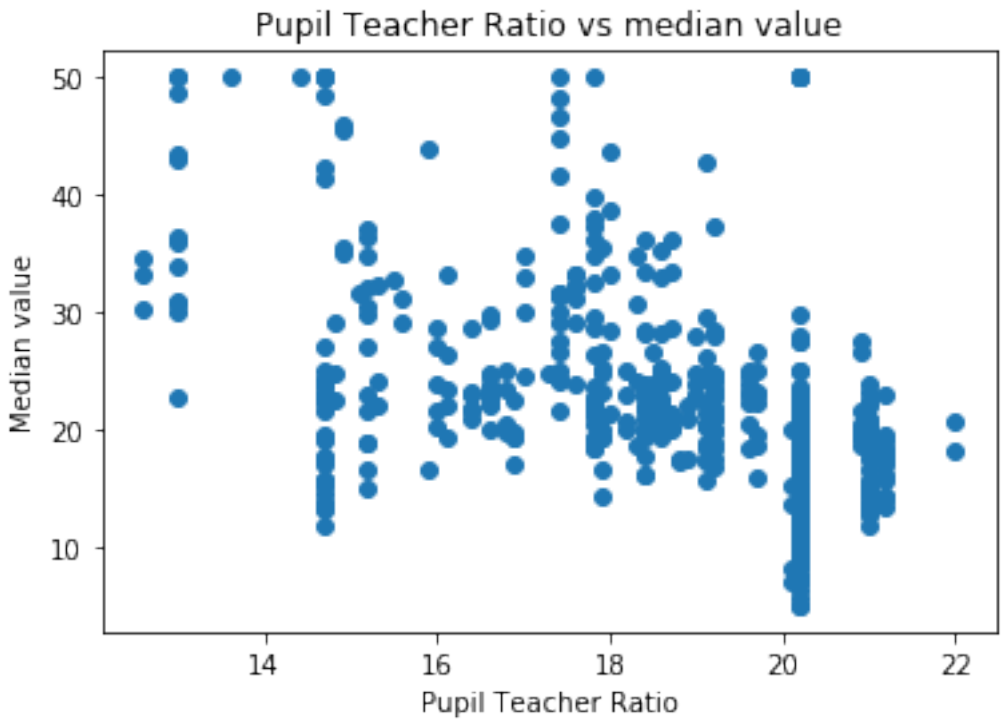
Strong negative coorelation, the more/less concentrated NOX is in the air, the lower/higher the price of housing

[18]:
```
#Scatter Diagram
plt.scatter(h_data['dis'],h_data['medv'], color = 'b')
plt.xlabel('Weighted Distance')
plt.ylabel('Median value')
plt.title('Weighted Distance vs median value')
plt.show()
```



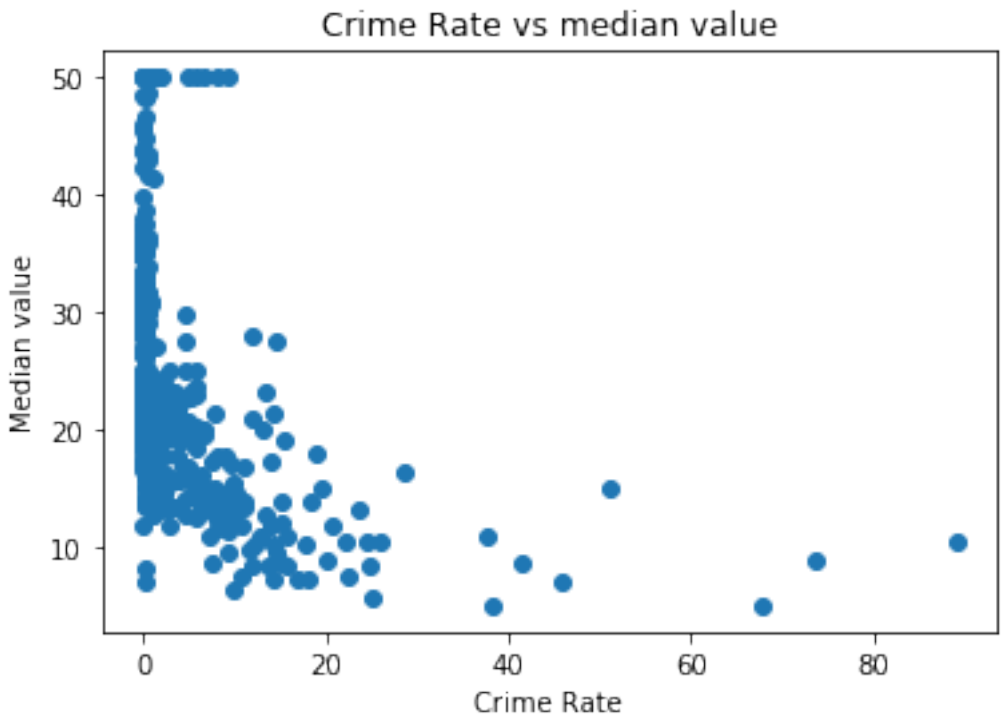
Strong positive coorelation, the closer/further the town is from employment centers, the higher/lower the housing price

[21]:
```
#Scatter Diagram
plt.scatter(h_data['ptratio'],h_data['medv'])
plt.xlabel('Pupil Teacher Ratio')
plt.ylabel('Median value')
plt.title('Pupil Teacher Ratio vs median value')
plt.show()
```



As the number of students increases for every teacher, the value of housing decreases

[22]:
```
#Scatter Diagram
plt.scatter(h_data['crim'],h_data['medv'])
plt.xlabel('Crime Rate')
plt.ylabel('Median value')
plt.title('Crime Rate vs median value')
plt.show()
```



As them crime rate decreases/increases, the housing price increases/decreases

[1]:
```
#Summary Measures
```

[9]:
```
h_data.describe()
```

[9]:

	Unnamed: 0	crim	zn	indus	chas	nox	\
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	
mean	253.500000	3.613524	11.363636	11.136779	0.069170	0.554695	
std	146.213884	8.601545	23.322453	6.860353	0.253994	0.115878	
min	1.000000	0.006320	0.000000	0.460000	0.000000	0.385000	
25%	127.250000	0.082045	0.000000	5.190000	0.000000	0.449000	
50%	253.500000	0.256510	0.000000	9.690000	0.000000	0.538000	
75%	379.750000	3.677082	12.500000	18.100000	0.000000	0.624000	
max	506.000000	88.976200	100.000000	27.740000	1.000000	0.871000	

	rm	age	dis	rad	tax	ptratio	\
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	
mean	6.284634	68.574901	3.795043	9.549407	408.237154	18.455534	
std	0.702617	28.148861	2.105710	8.707259	168.537116	2.164946	
min	3.561000	2.900000	1.129600	1.000000	187.000000	12.600000	
25%	5.885500	45.025000	2.100175	4.000000	279.000000	17.400000	
50%	6.208500	77.500000	3.207450	5.000000	330.000000	19.050000	
75%	6.623500	94.075000	5.188425	24.000000	666.000000	20.200000	

```
max      8.780000  100.000000   12.126500   24.000000  711.000000   22.000000

            black       lstat        medv
count  506.000000  506.000000  506.000000
mean   356.674032   12.653063   22.532806
std     91.294864    7.141062    9.197104
min      0.320000    1.730000    5.000000
25%    375.377500    6.950000   17.025000
50%    391.440000   11.360000   21.200000
75%    396.225000   16.955000   25.000000
max    396.900000   37.970000   50.000000
```

[10]:
```
h_data.mode()
```

[10]:
```
     Unnamed: 0      crim   zn  indus  chas    nox     rm    age     dis  \
0             1   0.01501  0.0   18.1   0.0  0.538  5.713  100.0  3.4952
1             2  14.33370  NaN    NaN   NaN    NaN  6.127    NaN     NaN
2             3       NaN  NaN    NaN   NaN    NaN  6.167    NaN     NaN
3             4       NaN  NaN    NaN   NaN    NaN  6.229    NaN     NaN
4             5       NaN  NaN    NaN   NaN    NaN  6.405    NaN     NaN
..          …       …  …    …   …    …    …    …     …
501         502       NaN  NaN    NaN   NaN    NaN    NaN    NaN     NaN
502         503       NaN  NaN    NaN   NaN    NaN    NaN    NaN     NaN
503         504       NaN  NaN    NaN   NaN    NaN    NaN    NaN     NaN
504         505       NaN  NaN    NaN   NaN    NaN    NaN    NaN     NaN
505         506       NaN  NaN    NaN   NaN    NaN    NaN    NaN     NaN

      rad    tax  ptratio  black  lstat  medv
0    24.0  666.0     20.2  396.9   6.36  50.0
1     NaN    NaN      NaN    NaN   7.79   NaN
2     NaN    NaN      NaN    NaN   8.05   NaN
3     NaN    NaN      NaN    NaN  14.10   NaN
4     NaN    NaN      NaN    NaN  18.13   NaN
..    …    …       …    …    …   …
501   NaN    NaN      NaN    NaN    NaN   NaN
502   NaN    NaN      NaN    NaN    NaN   NaN
503   NaN    NaN      NaN    NaN    NaN   NaN
504   NaN    NaN      NaN    NaN    NaN   NaN
505   NaN    NaN      NaN    NaN    NaN   NaN

[506 rows x 15 columns]
```

[11]:
```
h_data.median()
```

[11]:
```
Unnamed: 0    253.50000
crim            0.25651
zn              0.00000
```

```
indus           9.69000
chas            0.00000
nox             0.53800
rm              6.20850
age            77.50000
dis             3.20745
rad             5.00000
tax           330.00000
ptratio        19.05000
black         391.44000
lstat          11.36000
medv           21.20000
dtype: float64
```

[12]:
```
h_data.corr()
```

[12]:
```
            Unnamed: 0      crim        zn     indus      chas       nox  \
Unnamed: 0    1.000000  0.407407 -0.103393  0.399439 -0.003759  0.398736
crim          0.407407  1.000000 -0.200469  0.406583 -0.055892  0.420972
zn           -0.103393 -0.200469  1.000000 -0.533828 -0.042697 -0.516604
indus         0.399439  0.406583 -0.533828  1.000000  0.062938  0.763651
chas         -0.003759 -0.055892 -0.042697  0.062938  1.000000  0.091203
nox           0.398736  0.420972 -0.516604  0.763651  0.091203  1.000000
rm           -0.079971 -0.219247  0.311991 -0.391676  0.091251 -0.302188
age           0.203784  0.352734 -0.569537  0.644779  0.086518  0.731470
dis          -0.302211 -0.379670  0.664408 -0.708027 -0.099176 -0.769230
rad           0.686002  0.625505 -0.311948  0.595129 -0.007368  0.611441
tax           0.666626  0.582764 -0.314563  0.720760 -0.035587  0.668023
ptratio       0.291074  0.289946 -0.391679  0.383248 -0.121515  0.188933
black        -0.295041 -0.385064  0.175520 -0.356977  0.048788 -0.380051
lstat         0.258465  0.455621 -0.412995  0.603800 -0.053929  0.590879
medv         -0.226604 -0.388305  0.360445 -0.483725  0.175260 -0.427321

                   rm       age       dis       rad       tax   ptratio  \
Unnamed: 0 -0.079971  0.203784 -0.302211  0.686002  0.666626  0.291074
crim       -0.219247  0.352734 -0.379670  0.625505  0.582764  0.289946
zn          0.311991 -0.569537  0.664408 -0.311948 -0.314563 -0.391679
indus      -0.391676  0.644779 -0.708027  0.595129  0.720760  0.383248
chas        0.091251  0.086518 -0.099176 -0.007368 -0.035587 -0.121515
nox        -0.302188  0.731470 -0.769230  0.611441  0.668023  0.188933
rm          1.000000 -0.240265  0.205246 -0.209847 -0.292048 -0.355501
age        -0.240265  1.000000 -0.747881  0.456022  0.506456  0.261515
dis         0.205246 -0.747881  1.000000 -0.494588 -0.534432 -0.232471
rad        -0.209847  0.456022 -0.494588  1.000000  0.910228  0.464741
tax        -0.292048  0.506456 -0.534432  0.910228  1.000000  0.460853
ptratio    -0.355501  0.261515 -0.232471  0.464741  0.460853  1.000000
black       0.128069 -0.273534  0.291512 -0.444413 -0.441808 -0.177383
```

```
lstat      -0.613808  0.602339 -0.496996  0.488676  0.543993  0.374044
medv        0.695360 -0.376955  0.249929 -0.381626 -0.468536 -0.507787

                black     lstat      medv
Unnamed: 0 -0.295041  0.258465 -0.226604
crim       -0.385064  0.455621 -0.388305
zn          0.175520 -0.412995  0.360445
indus      -0.356977  0.603800 -0.483725
chas        0.048788 -0.053929  0.175260
nox        -0.380051  0.590879 -0.427321
rm          0.128069 -0.613808  0.695360
age        -0.273534  0.602339 -0.376955
dis         0.291512 -0.496996  0.249929
rad        -0.444413  0.488676 -0.381626
tax        -0.441808  0.543993 -0.468536
ptratio    -0.177383  0.374044 -0.507787
black       1.000000 -0.366087  0.333461
lstat      -0.366087  1.000000 -0.737663
medv        0.333461 -0.737663  1.000000
```

[13]:
```
h_data.cov()
```

[13]:
```
              Unnamed: 0        crim           zn       indus      chas  \
Unnamed: 0  21378.500000  512.381872  -352.578218  400.668663 -0.139604
crim          512.381872   73.986578   -40.215956   23.992339 -0.122109
zn           -352.578218  -40.215956   543.936814  -85.412648 -0.252925
indus         400.668663   23.992339   -85.412648   47.064442  0.109669
chas           -0.139604   -0.122109    -0.252925    0.109669  0.064513
nox             6.755757    0.419594    -1.396148    0.607074  0.002684
rm             -8.215627   -1.325038     5.112513   -1.887957  0.016285
age           838.722871   85.405322  -373.901548  124.513903  0.618571
dis           -93.045936   -6.876722    32.629304  -10.228097 -0.053043
rad           873.364356   46.847761   -63.348695   35.549971 -0.016296
tax         16427.306931  844.821538 -1236.453735  833.360290 -1.523367
ptratio        92.138119    5.399331   -19.776571    5.692104 -0.066819
black       -3938.380535 -302.381816   373.721402 -223.579756  1.131325
lstat         269.868842   27.986168   -68.783037   29.580270 -0.097816
medv         -304.723960  -30.718508    77.315176  -30.520823  0.409409

                   nox         rm          age         dis          rad  \
Unnamed: 0    6.755757  -8.215627   838.722871  -93.045936   873.364356
crim          0.419594  -1.325038    85.405322   -6.876722    46.847761
zn           -1.396148   5.112513  -373.901548   32.629304   -63.348695
indus         0.607074  -1.887957   124.513903  -10.228097    35.549971
chas          0.002684   0.016285     0.618571   -0.053043    -0.016296
nox           0.013428  -0.024603     2.385927   -0.187696     0.616929
rm           -0.024603   0.493671    -4.751929    0.303663    -1.283815
```

```
age          2.385927  -4.751929   792.358399  -44.329379   111.770846
dis         -0.187696   0.303663   -44.329379    4.434015    -9.068252
rad          0.616929  -1.283815   111.770846   -9.068252    75.816366
tax         13.046286 -34.583448  2402.690122 -189.664592  1335.756577
ptratio      0.047397  -0.540763    15.936921   -1.059775     8.760716
black       -4.020570   8.215006  -702.940328   56.040356  -353.276219
lstat        0.488946  -3.079741   121.077725   -7.473329    30.385442
medv        -0.455412   4.493446   -97.589017    4.840229   -30.561228
```

```
                    tax     ptratio        black       lstat        medv
Unnamed: 0  16427.306931   92.138119 -3938.380535  269.868842 -304.723960
crim          844.821538    5.399331  -302.381816   27.986168  -30.718508
zn          -1236.453735  -19.776571   373.721402  -68.783037   77.315176
indus         833.360290    5.692104  -223.579756   29.580270  -30.520823
chas           -1.523367   -0.066819     1.131325   -0.097816    0.409409
nox            13.046286    0.047397    -4.020570    0.488946   -0.455412
rm            -34.583448   -0.540763     8.215006   -3.079741    4.493446
age          2402.690122   15.936921  -702.940328  121.077725  -97.589017
dis          -189.664592   -1.059775    56.040356   -7.473329    4.840229
rad          1335.756577    8.760716  -353.276219   30.385442  -30.561228
tax         28404.759488  168.153141 -6797.911215  654.714520 -726.255716
ptratio       168.153141    4.686989   -35.059527    5.782729  -10.110657
black       -6797.911215  -35.059527  8334.752263 -238.667516  279.989834
lstat         654.714520    5.782729  -238.667516   50.994760  -48.447538
medv         -726.255716  -10.110657   279.989834  -48.447538   84.586724
```

[ ]: