

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math
```

```
In [2]: df = pd.read_csv('Boston.csv')
```

```
In [3]: df.rename(columns={'Unnamed: 0': 'i'}, inplace=True)
df.head()
# 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 - nitric 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
```

Out[3]:

	i	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	2
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	2
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	3
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	3
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	3

Problem 1

```
In [4]: df.shape
```

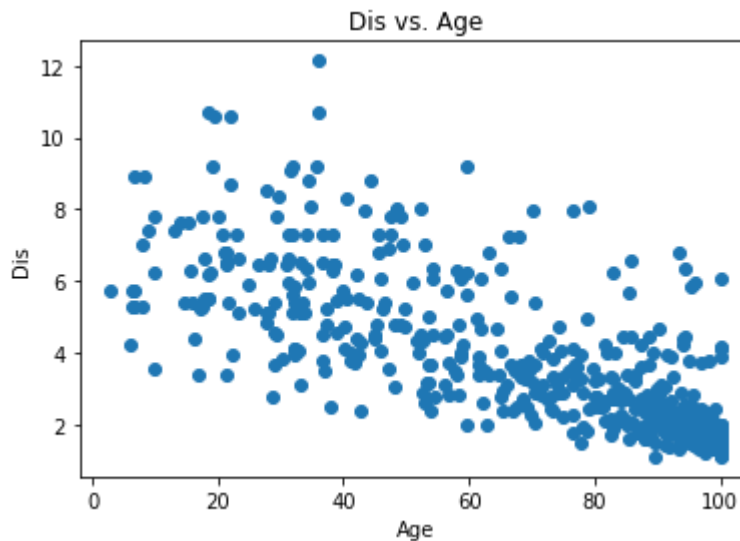
Out[4]: (506, 15)

Problem 2

```
In [5]: plt.scatter(df['age'], df['dis'])

plt.title("Dis vs. Age")
plt.xlabel("Age")
plt.ylabel("Dis")
```

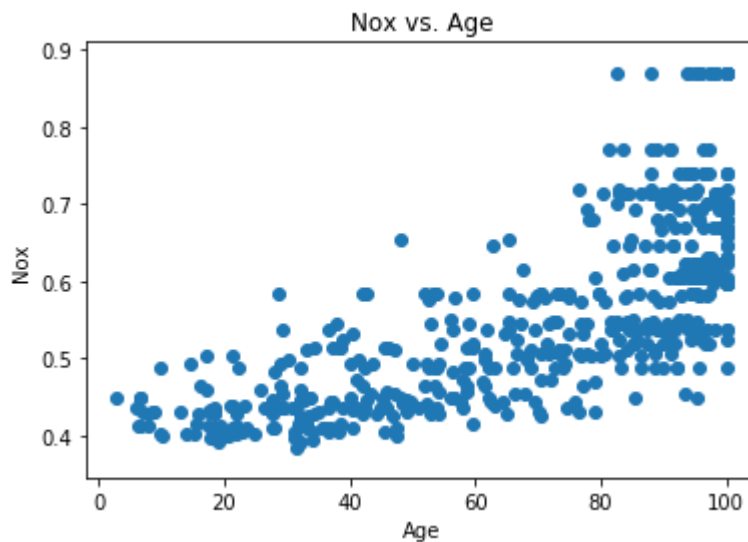
Out[5]: Text(0,0.5, 'Dis')



```
In [6]: plt.scatter(df['age'], df['nox'])

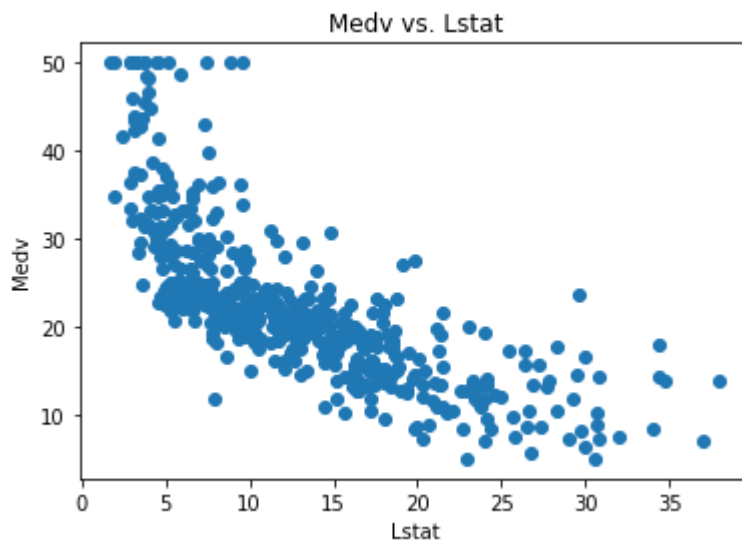
plt.title("Nox vs. Age")
plt.xlabel("Age")
plt.ylabel("Nox")
```

Out[6]: Text(0,0.5, 'Nox')



```
In [7]: plt.scatter(df['lstat'], df['medv'])  
  
plt.title("Medv vs. Lstat")  
plt.xlabel("Lstat")  
plt.ylabel("Medv")
```

```
Out[7]: Text(0,0.5,'Medv')
```



Problem 3

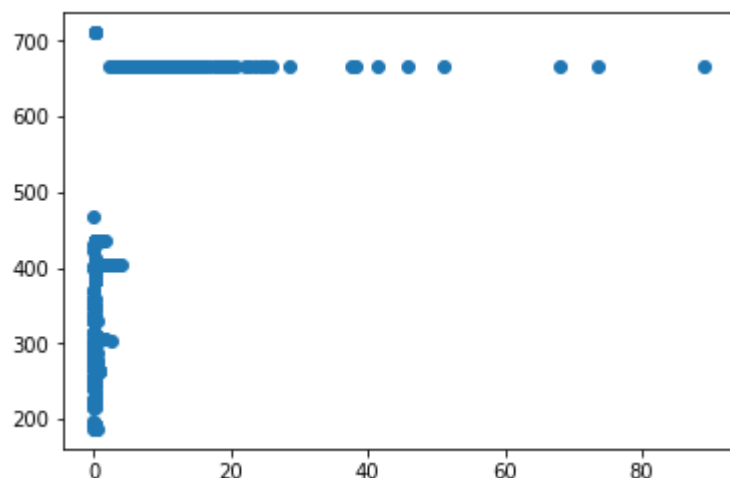
In [8]: `df.corr()`

Out[8]:

	i	crim	zn	indus	chas	nox	rm	age	
i	1.000000	0.407407	-0.103393	0.399439	-0.003759	0.398736	-0.079971	0.203784	-0
crim	0.407407	1.000000	-0.200469	0.406583	-0.055892	0.420972	-0.219247	0.352734	-0
zn	-0.103393	-0.200469	1.000000	-0.533828	-0.042697	-0.516604	0.311991	-0.569537	0
indus	0.399439	0.406583	-0.533828	1.000000	0.062938	0.763651	-0.391676	0.644779	-0
chas	-0.003759	-0.055892	-0.042697	0.062938	1.000000	0.091203	0.091251	0.086518	-0
nox	0.398736	0.420972	-0.516604	0.763651	0.091203	1.000000	-0.302188	0.731470	-0
rm	-0.079971	-0.219247	0.311991	-0.391676	0.091251	-0.302188	1.000000	-0.240265	0
age	0.203784	0.352734	-0.569537	0.644779	0.086518	0.731470	-0.240265	1.000000	-0
dis	-0.302211	-0.379670	0.664408	-0.708027	-0.099176	-0.769230	0.205246	-0.747881	1
rad	0.686002	0.625505	-0.311948	0.595129	-0.007368	0.611441	-0.209847	0.456022	-0
tax	0.666626	0.582764	-0.314563	0.720760	-0.035587	0.668023	-0.292048	0.506456	-0
ptratio	0.291074	0.289946	-0.391679	0.383248	-0.121515	0.188933	-0.355501	0.261515	-0
black	-0.295041	-0.385064	0.175520	-0.356977	0.048788	-0.380051	0.128069	-0.273534	0
lstat	0.258465	0.455621	-0.412995	0.603800	-0.053929	0.590879	-0.613808	0.602339	-0
medv	-0.226604	-0.388305	0.360445	-0.483725	0.175260	-0.427321	0.695360	-0.376955	0

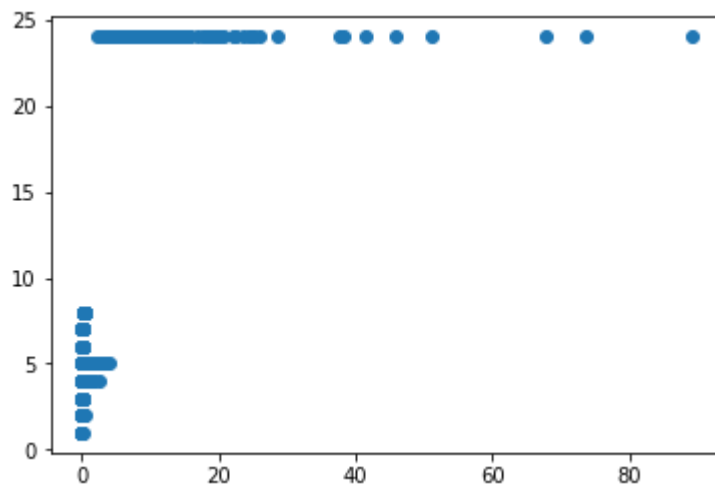
In [9]: `plt.scatter(df['crim'], df['tax'])`

Out[9]: `<matplotlib.collections.PathCollection at 0x7fac972d9128>`



```
In [10]: plt.scatter(df['crim'], df['rad'])
```

```
Out[10]: <matplotlib.collections.PathCollection at 0x7fac972a3f98>
```



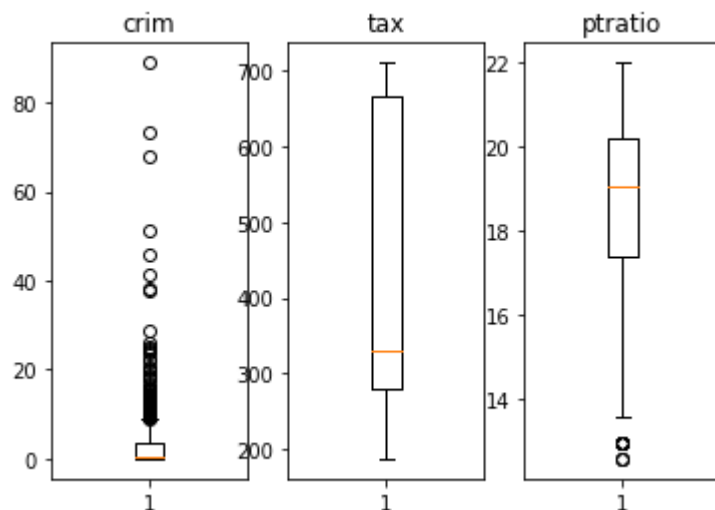
Problem 4

```
In [11]: # Problem D
fig, (ax1, ax2, ax3) = plt.subplots(1, 3)
ax1.boxplot(df['crim'])
ax1.set_title('crim')

ax2.boxplot(df['tax'])
ax2.set_title('tax')

ax3.boxplot(df['ptratio'])
ax3.set_title('ptratio')
```

```
Out[11]: Text(0.5,1,'ptratio')
```



```
In [12]: df.max() - df.min()
```

```
Out[12]: i          505.00000  
        crim       88.96988  
        zn        100.00000  
        indus     27.28000  
        chas       1.00000  
        nox       0.48600  
        rm        5.21900  
        age       97.10000  
        dis       10.99690  
        rad       23.00000  
        tax      524.00000  
        ptratio    9.40000  
        black     396.58000  
        lstat     36.24000  
        medv      45.00000  
        dtype: float64
```

Problem 5

```
In [13]: df.sum()
```

```
Out[13]: i          128271.00000  
        crim       1828.44292  
        zn        5750.00000  
        indus     5635.21000  
        chas       35.00000  
        nox       280.67570  
        rm        3180.02500  
        age       34698.90000  
        dis       1920.29160  
        rad       4832.00000  
        tax      206568.00000  
        ptratio    9338.50000  
        black     180477.06000  
        lstat     6402.45000  
        medv     11401.60000  
        dtype: float64
```

Problem 6

```
In [14]: df.median()
```

```
Out[14]: i          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
```

Problem 7

```
In [15]: df['medv'].min()
```

```
Out[15]: 5.0
```

```
In [16]: df['medv'].idxmin()
```

```
Out[16]: 398
```

```
In [17]: df.iloc[398]
```

```
Out[17]: i          399.0000
        crim        38.3518
        zn          0.0000
        indus       18.1000
        chas        0.0000
        nox         0.6930
        rm          5.4530
        age        100.0000
        dis         1.4896
        rad         24.0000
        tax        666.0000
        ptratio     20.2000
        black      396.9000
        lstat       30.5900
        medv         5.0000
        Name: 398, dtype: float64
```

```
In [18]: df.iloc[398] - (df.max()-df.min())
```

```
Out[18]: i          -106.00000  
        crim       -50.61808  
        zn        -100.00000  
        indus      -9.18000  
        chas       -1.00000  
        nox        0.20700  
        rm         0.23400  
        age        2.90000  
        dis       -9.50730  
        rad        1.00000  
        tax       142.00000  
        ptratio    10.80000  
        black      0.32000  
        lstat      -5.65000  
        medv      -40.00000  
        dtype: float64
```

Problem 8


```
In [19]: df.loc[df['rm'] > 7]
```

Out[19]:

	i	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black
2	3	0.02729	0.0	7.07	0	0.4690	7.185	61.1	4.9671	2	242	17.8	392.83
4	5	0.06905	0.0	2.18	0	0.4580	7.147	54.2	6.0622	3	222	18.7	396.90
40	41	0.03359	75.0	2.95	0	0.4280	7.024	15.8	5.4011	3	252	18.3	395.62
55	56	0.01311	90.0	1.22	0	0.4030	7.249	21.9	8.6966	5	226	17.9	395.93
64	65	0.01951	17.5	1.38	0	0.4161	7.104	59.5	9.2229	3	216	18.6	393.24
88	89	0.05660	0.0	3.41	0	0.4890	7.007	86.3	3.4217	2	270	17.8	396.90
89	90	0.05302	0.0	3.41	0	0.4890	7.079	63.1	3.4145	2	270	17.8	396.06
97	98	0.12083	0.0	2.89	0	0.4450	8.069	76.0	3.4952	2	276	18.0	396.90
98	99	0.08187	0.0	2.89	0	0.4450	7.820	36.9	3.4952	2	276	18.0	393.53
99	100	0.06860	0.0	2.89	0	0.4450	7.416	62.5	3.4952	2	276	18.0	396.90
161	162	1.46336	0.0	19.58	0	0.6050	7.489	90.8	1.9709	5	403	14.7	374.43
162	163	1.83377	0.0	19.58	1	0.6050	7.802	98.2	2.0407	5	403	14.7	389.61
163	164	1.51902	0.0	19.58	1	0.6050	8.375	93.9	2.1620	5	403	14.7	388.45
166	167	2.01019	0.0	19.58	0	0.6050	7.929	96.2	2.0459	5	403	14.7	369.30
180	181	0.06588	0.0	2.46	0	0.4880	7.765	83.3	2.7410	3	193	17.8	395.56
182	183	0.09103	0.0	2.46	0	0.4880	7.155	92.2	2.7006	3	193	17.8	394.12
186	187	0.05602	0.0	2.46	0	0.4880	7.831	53.6	3.1992	3	193	17.8	392.63
189	190	0.08370	45.0	3.44	0	0.4370	7.185	38.9	4.5667	5	398	15.2	396.90
192	193	0.08664	45.0	3.44	0	0.4370	7.178	26.3	6.4798	5	398	15.2	390.49
195	196	0.01381	80.0	0.46	0	0.4220	7.875	32.0	5.6484	4	255	14.4	394.23
196	197	0.04011	80.0	1.52	0	0.4040	7.287	34.1	7.3090	2	329	12.6	396.90
197	198	0.04666	80.0	1.52	0	0.4040	7.107	36.6	7.3090	2	329	12.6	354.31
198	199	0.03768	80.0	1.52	0	0.4040	7.274	38.3	7.3090	2	329	12.6	392.20
200	201	0.01778	95.0	1.47	0	0.4030	7.135	13.9	7.6534	3	402	17.0	384.30
202	203	0.02177	82.5	2.03	0	0.4150	7.610	15.7	6.2700	2	348	14.7	395.38
203	204	0.03510	95.0	2.68	0	0.4161	7.853	33.2	5.1180	4	224	14.7	392.78
204	205	0.02009	95.0	2.68	0	0.4161	8.034	31.9	5.1180	4	224	14.7	390.55
224	225	0.31533	0.0	6.20	0	0.5040	8.266	78.3	2.8944	8	307	17.4	385.05
225	226	0.52693	0.0	6.20	0	0.5040	8.725	83.0	2.8944	8	307	17.4	382.00
226	227	0.38214	0.0	6.20	0	0.5040	8.040	86.5	3.2157	8	307	17.4	387.38
...
233	234	0.33147	0.0	6.20	0	0.5070	8.247	70.4	3.6519	8	307	17.4	378.95
237	238	0.51183	0.0	6.20	0	0.5070	7.358	71.6	4.1480	8	307	17.4	390.07
253	254	0.36894	22.0	5.86	0	0.4310	8.259	8.4	8.9067	7	330	19.1	396.90
256	257	0.01538	90.0	3.75	0	0.3940	7.454	34.2	6.3361	3	244	15.9	386.34

	i	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black
257	258	0.61154	20.0	3.97	0	0.6470	8.704	86.9	1.8010	5	264	13.0	389.70
258	259	0.66351	20.0	3.97	0	0.6470	7.333	100.0	1.8946	5	264	13.0	383.29
260	261	0.54011	20.0	3.97	0	0.6470	7.203	81.8	2.1121	5	264	13.0	392.80
261	262	0.53412	20.0	3.97	0	0.6470	7.520	89.4	2.1398	5	264	13.0	388.37
262	263	0.52014	20.0	3.97	0	0.6470	8.398	91.5	2.2885	5	264	13.0	386.86
263	264	0.82526	20.0	3.97	0	0.6470	7.327	94.5	2.0788	5	264	13.0	393.42
264	265	0.55007	20.0	3.97	0	0.6470	7.206	91.6	1.9301	5	264	13.0	387.89
266	267	0.78570	20.0	3.97	0	0.6470	7.014	84.6	2.1329	5	264	13.0	384.07
267	268	0.57834	20.0	3.97	0	0.5750	8.297	67.0	2.4216	5	264	13.0	384.54
268	269	0.54050	20.0	3.97	0	0.5750	7.470	52.6	2.8720	5	264	13.0	390.30
273	274	0.22188	20.0	6.96	1	0.4640	7.691	51.8	4.3665	3	223	18.6	390.77
276	277	0.10469	40.0	6.41	1	0.4470	7.267	49.0	4.7872	4	254	17.6	389.25
280	281	0.03578	20.0	3.33	0	0.4429	7.820	64.5	4.6947	5	216	14.9	387.31
282	283	0.06129	20.0	3.33	1	0.4429	7.645	49.7	5.2119	5	216	14.9	377.07
283	284	0.01501	90.0	1.21	1	0.4010	7.923	24.8	5.8850	1	198	13.6	395.52
284	285	0.00906	90.0	2.97	0	0.4000	7.088	20.8	7.3073	1	285	15.3	394.72
291	292	0.07886	80.0	4.95	0	0.4110	7.148	27.7	5.1167	4	245	19.2	396.90
299	300	0.05561	70.0	2.24	0	0.4000	7.041	10.0	7.8278	5	358	14.8	371.58
304	305	0.05515	33.0	2.18	0	0.4720	7.236	41.1	4.0220	7	222	18.4	393.68
306	307	0.07503	33.0	2.18	0	0.4720	7.420	71.9	3.0992	7	222	18.4	396.90
341	342	0.01301	35.0	1.52	0	0.4420	7.241	49.3	7.0379	1	284	15.5	394.74
364	365	3.47428	0.0	18.10	1	0.7180	8.780	82.9	1.9047	24	666	20.2	354.55
370	371	6.53876	0.0	18.10	1	0.6310	7.016	97.5	1.2024	24	666	20.2	392.05
375	376	19.60910	0.0	18.10	0	0.6710	7.313	97.9	1.3163	24	666	20.2	396.90
453	454	8.24809	0.0	18.10	0	0.7130	7.393	99.3	2.4527	24	666	20.2	375.87
482	483	5.73116	0.0	18.10	0	0.5320	7.061	77.0	3.4106	24	666	20.2	395.28

64 rows × 15 columns



In [20]: `df.loc[df['rm'] > 8]`

Out[20]:

	i	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstr
97	98	0.12083	0.0	2.89	0	0.4450	8.069	76.0	3.4952	2	276	18.0	396.90	4.2
163	164	1.51902	0.0	19.58	1	0.6050	8.375	93.9	2.1620	5	403	14.7	388.45	3.3
204	205	0.02009	95.0	2.68	0	0.4161	8.034	31.9	5.1180	4	224	14.7	390.55	2.8
224	225	0.31533	0.0	6.20	0	0.5040	8.266	78.3	2.8944	8	307	17.4	385.05	4.1
225	226	0.52693	0.0	6.20	0	0.5040	8.725	83.0	2.8944	8	307	17.4	382.00	4.6
226	227	0.38214	0.0	6.20	0	0.5040	8.040	86.5	3.2157	8	307	17.4	387.38	3.1
232	233	0.57529	0.0	6.20	0	0.5070	8.337	73.3	3.8384	8	307	17.4	385.91	2.4
233	234	0.33147	0.0	6.20	0	0.5070	8.247	70.4	3.6519	8	307	17.4	378.95	3.9
253	254	0.36894	22.0	5.86	0	0.4310	8.259	8.4	8.9067	7	330	19.1	396.90	3.5
257	258	0.61154	20.0	3.97	0	0.6470	8.704	86.9	1.8010	5	264	13.0	389.70	5.1
262	263	0.52014	20.0	3.97	0	0.6470	8.398	91.5	2.2885	5	264	13.0	386.86	5.9
267	268	0.57834	20.0	3.97	0	0.5750	8.297	67.0	2.4216	5	264	13.0	384.54	7.4
364	365	3.47428	0.0	18.10	1	0.7180	8.780	82.9	1.9047	24	666	20.2	354.55	5.2



In []:

In []:

In []: