Bruce Campbell ST-617 HW 1

Tue Jun 28 19:55:56 2016

Chapter 2

Problem 10

a)

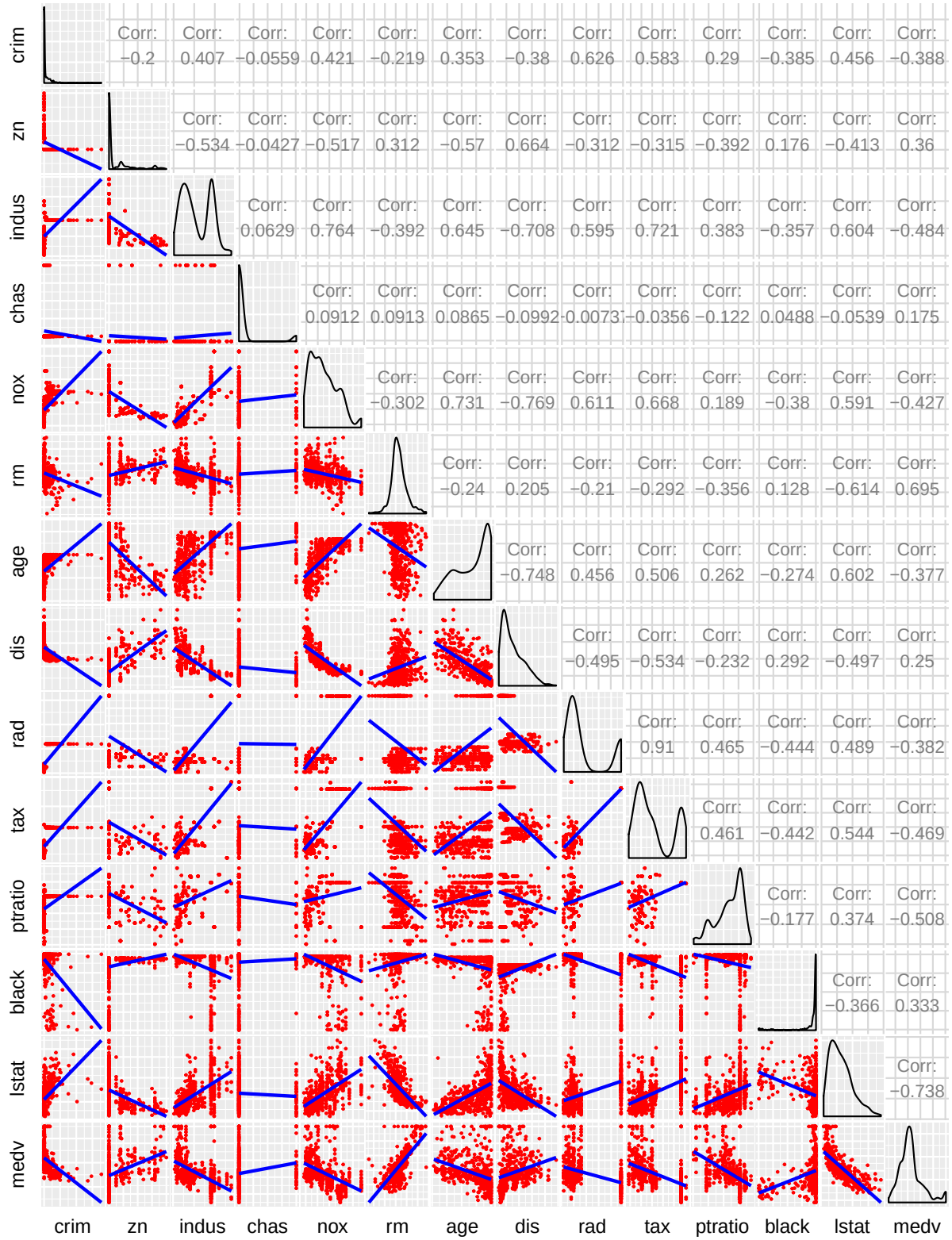
There are 506 observations in the Boston data set which contain a variety of variables porported to affect housing values in the suburbs of Boston.

This data frame contains the following 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 mean of 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.
- black : 1000(Bk - 0.63)ˆ2 where Bk is the proportion of blacks by town.
- lstat : lower status of the population (percent).
- medv : median value of owner-occupied homes in $1000s

b)

```
library(MASS)
# library(ggplot2) require(GGally)
library(pander)
DF <- Boston
```



We note * several bimodal variables : indus,rad,tax * some linear relationships : (dis,age) (rm,mdev) * a few mildly non linear relationships (nox,age) (nox,dis)

c)

chas has a higher crime rate at a value of 0

certain values of nox (nox >.57) are associated with a higher crime rate

older towns have a higher crime rate

Additionally there is a spike in crime at a indus value of 19

d)

The crime rate variable has a long tail.

There are 11 obersevations with a value higher than 25.

The tax rate appears bimodal with 137 towns having a rate above 650 and the rest below 450

There are some towns with a low putpil teacher ratio. We note that the tax rate for these towns does not fall in the upper bracket.

We use the summary function to look at the ranges of the variables

```
S <- summary(DF)
pander(S)
```

Table 1: Table continues below

crim	zn	indus	chas	nox
Min. : 0.00632	Min. : 0.00	Min. : 0.46	Min. :0.00000	Min. :0.3850
1st Qu.: 0.08204	1st Qu.: 0.00	1st Qu.: 5.19	1st Qu.:0.00000	1st Qu.:0.4490
Median : 0.25651	Median : 0.00	Median : 9.69	Median :0.00000	Median :0.5380
Mean : 3.61352	Mean : 11.36	Mean :11.14	Mean :0.06917	Mean :0.5547
3rd Qu.: 3.67708	3rd Qu.: 12.50	3rd Qu.:18.10	3rd Qu.:0.00000	3rd Qu.:0.6240
Max. :88.97620	Max. :100.00	Max. :27.74	Max. :1.00000	Max. :0.8710

Table 2: Table continues below

rm	age	dis	rad	tax
Min. :3.561	Min. : 2.90	Min. : 1.130	Min. : 1.000	Min. :187.0
1st Qu.:5.886	1st Qu.: 45.02	1st Qu.: 2.100	1st Qu.: 4.000	1st Qu.:279.0
Median :6.208	Median : 77.50	Median : 3.207	Median : 5.000	Median :330.0
Mean :6.285	Mean : 68.57	Mean : 3.795	Mean : 9.549	Mean :408.2
3rd Qu.:6.623	3rd Qu.: 94.08	3rd Qu.: 5.188	3rd Qu.:24.000	3rd Qu.:666.0
Max. :8.780	Max. :100.00	Max. :12.127	Max. :24.000	Max. :711.0

ptratio	black	lstat	medv
Min. :12.60	Min. : 0.32	Min. : 1.73	Min. : 5.00
1st Qu.:17.40	1st Qu.:375.38	1st Qu.: 6.95	1st Qu.:17.02
Median :19.05	Median :391.44	Median :11.36	Median :21.20
Mean :18.46	Mean :356.67	Mean :12.65	Mean :22.53
3rd Qu.:20.20	3rd Qu.:396.23	3rd Qu.:16.95	3rd Qu.:25.00

ptratio	black	lstat	medv
Max. :22.00	Max. :396.90	Max. :37.97	Max. :50.00

e)

```
countBoundingCharles <- nrow(DF[DF$chas == 1, ])
```

There are 35 towns identified as bordering the Charles.

f)

From the summary table above we see that the median pupil to teacher ration is 19.05.

g)

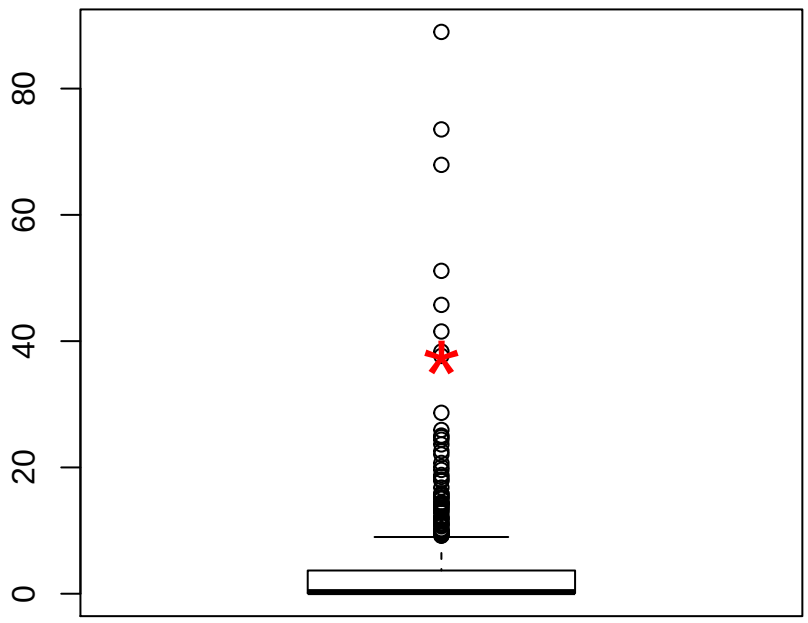
```
lowMed <- DF[which.min(DF$medv), ]
pander(lowMed)
```

Table 4: Table continues below

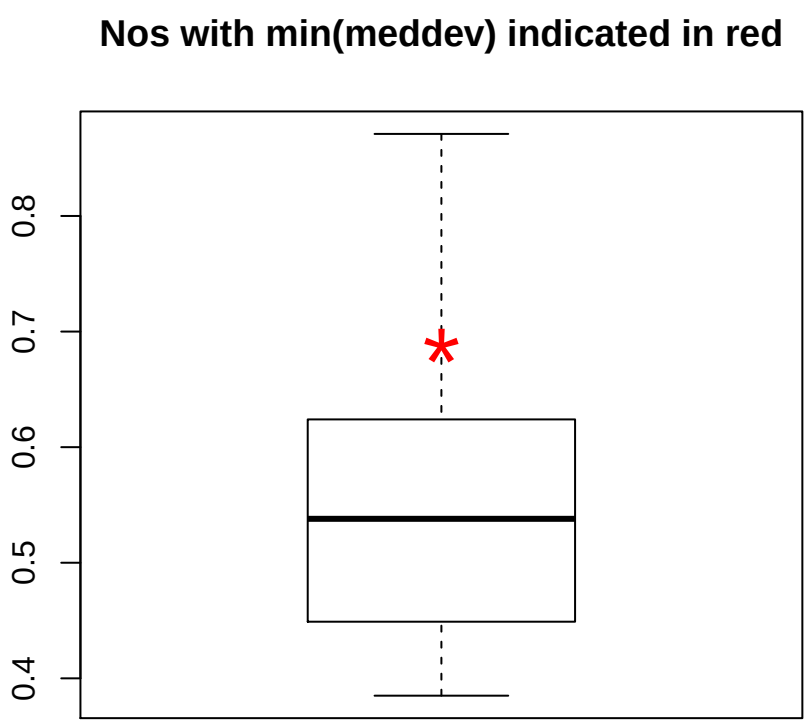
	crim	zn	indus	chas	nox	rm	age	dis	rad	tax
399	38.35	0	18.1	0	0.693	5.453	100	1.49	24	666

	ptratio	black	lstat	medv
399	20.2	396.9	30.59	5

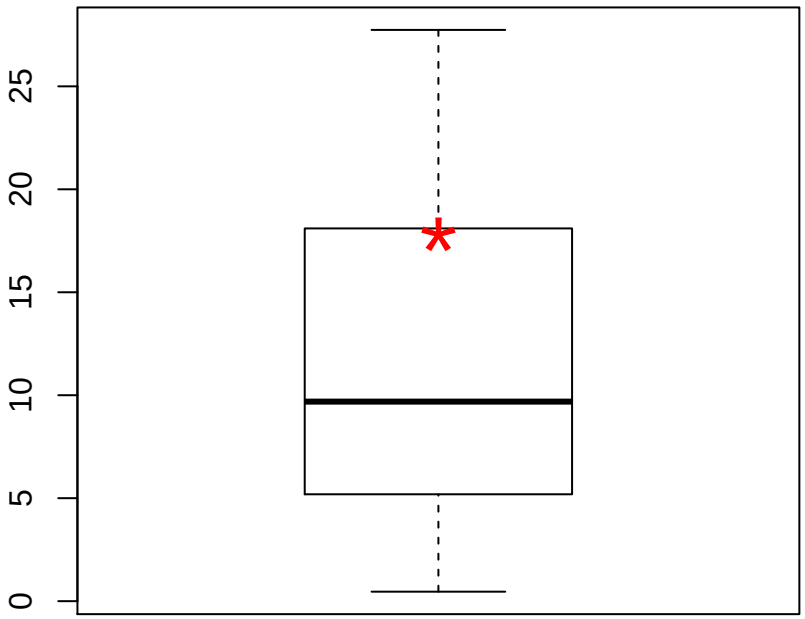
The town with the lowest median value of owner occupied houses has a high crime rate as indicated by the box plot below.

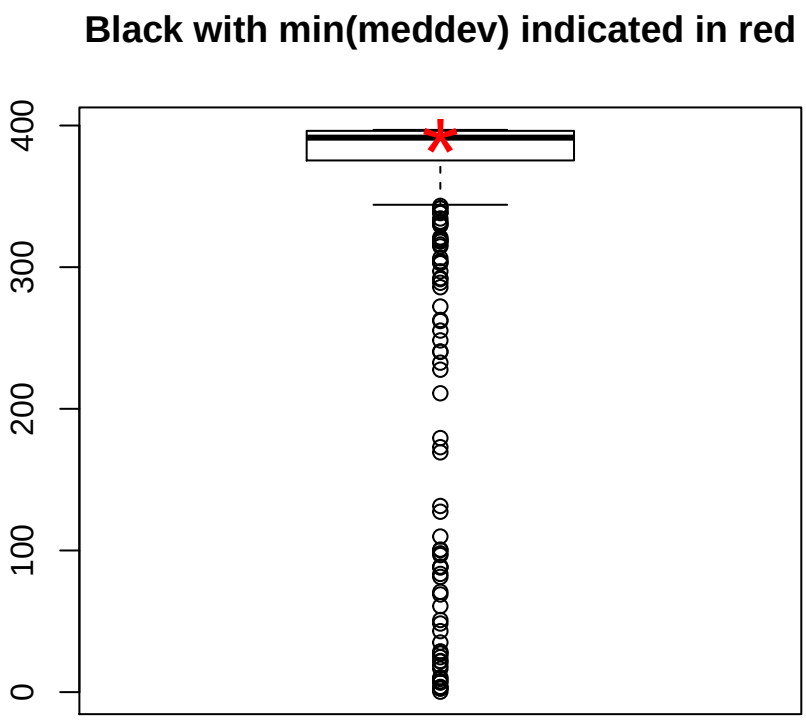


Nox is elevated for this town



Indus is elevated for this town


Indus with min(meddev) indicated in red
0
5
10
15
20
25



Black is elevated for this town

h)

```
sevenRooms <- DF[DF$rm > 7, ]

eightRooms <- DF[DF$rm > 8, ]
```

There are 64 dwellings with more than seven rooms, and there are 13 with more than eight rooms.

```
pander(eightRooms)
```

Table 6: Table continues below

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax
98	0.1208	0	2.89	0	0.445	8.069	76	3.495	2	276
164	1.519	0	19.58	1	0.605	8.375	93.9	2.162	5	403
205	0.02009	95	2.68	0	0.4161	8.034	31.9	5.118	4	224
225	0.3153	0	6.2	0	0.504	8.266	78.3	2.894	8	307
226	0.5269	0	6.2	0	0.504	8.725	83	2.894	8	307
227	0.3821	0	6.2	0	0.504	8.04	86.5	3.216	8	307
233	0.5753	0	6.2	0	0.507	8.337	73.3	3.838	8	307

	crim	zn	indus	chas	nox	rm	age	dis	rad	tax
234	0.3315	0	6.2	0	0.507	8.247	70.4	3.652	8	307
254	0.3689	22	5.86	0	0.431	8.259	8.4	8.907	7	330
258	0.6115	20	3.97	0	0.647	8.704	86.9	1.801	5	264
263	0.5201	20	3.97	0	0.647	8.398	91.5	2.288	5	264
268	0.5783	20	3.97	0	0.575	8.297	67	2.422	5	264
365	3.474	0	18.1	1	0.718	8.78	82.9	1.905	24	666

	ptratio	black	lstat	medv
98	18	396.9	4.21	38.7
164	14.7	388.4	3.32	50
205	14.7	390.6	2.88	50
225	17.4	385.1	4.14	44.8
226	17.4	382	4.63	50
227	17.4	387.4	3.13	37.6
233	17.4	385.9	2.47	41.7
234	17.4	378.9	3.95	48.3
254	19.1	396.9	3.54	42.8
258	13	389.7	5.12	50
263	13	386.9	5.91	48.8
268	13	384.5	7.44	50
365	20.2	354.6	5.29	21.9