Exploratory Analysis: Boston Housing

Cool Beans Programming

2023-04-23

Load the data

```
library(ISLR2)
data(Boston)
```

Explore the data

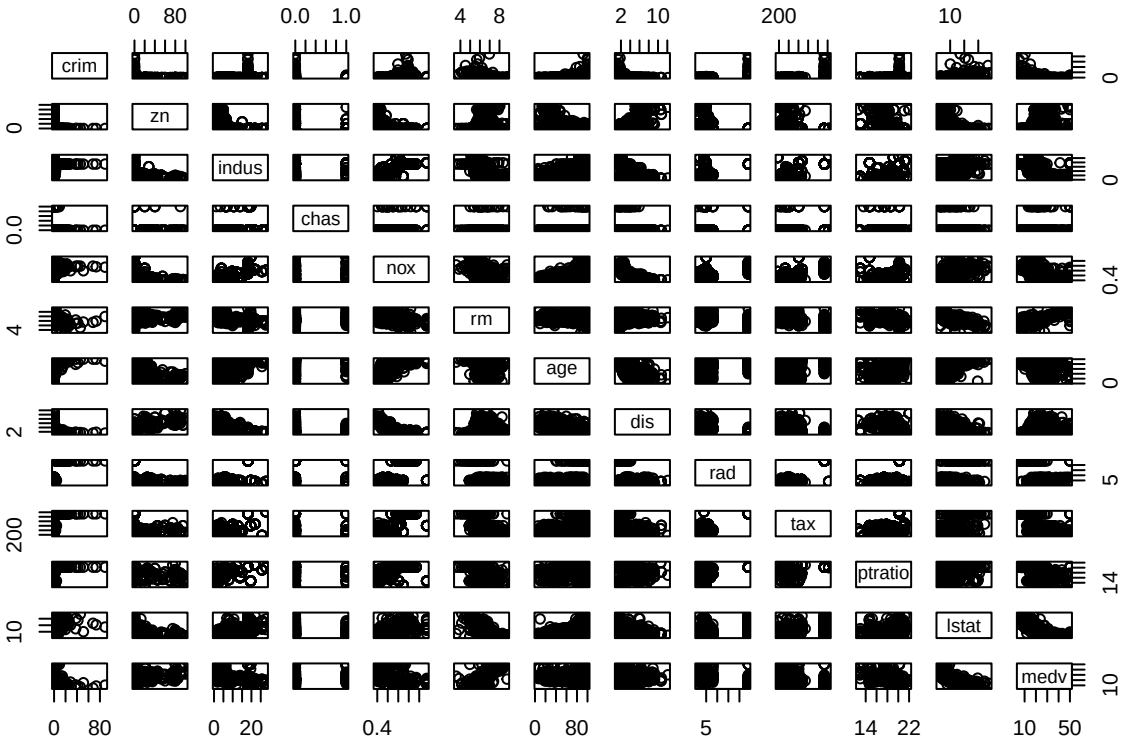
```
dim(Boston)
```

```
## [1] 506  13
```

There are 13 variables and 506 observations in the Boston data set. Each row represents a different neighborhood in Boston and the columns represent the predictor variables and relate to crime rate, average number of rooms, and pupil-teacher ratio by town among others.

Relationship to per capita crime rate

```
pairs(Boston)
```



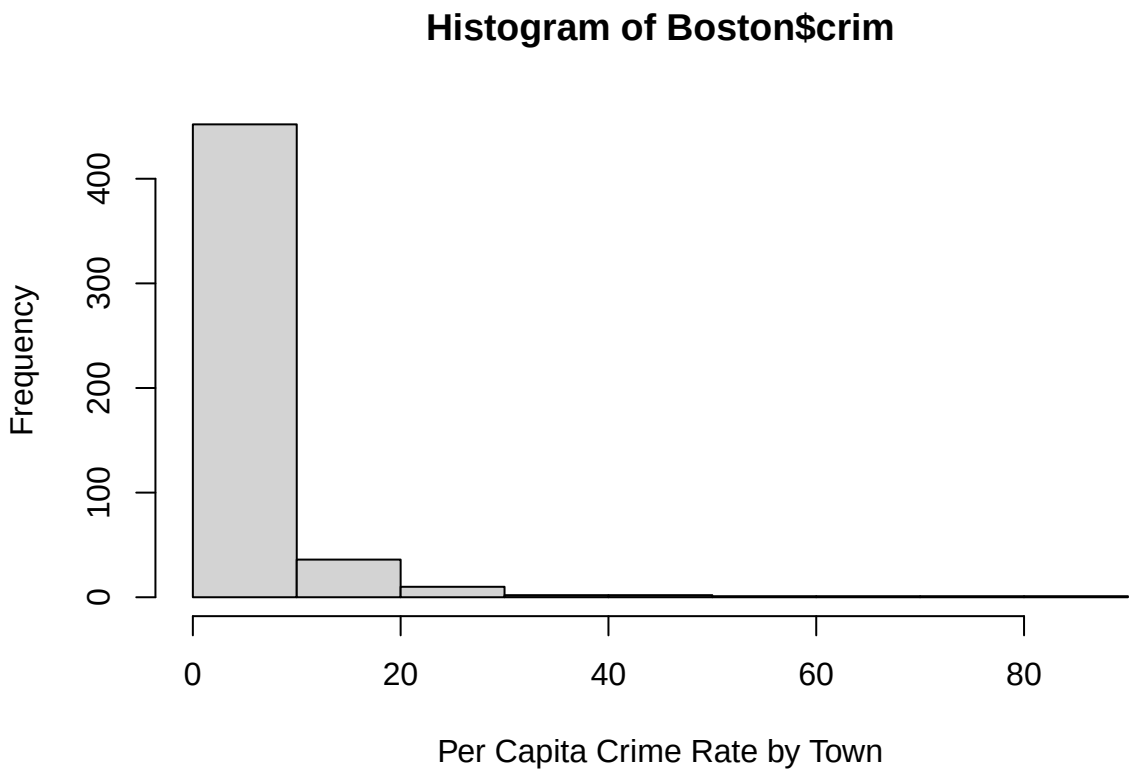
Some variables like rm (average number of rooms per dwelling) and lstat (lower status of the pop.) appear to have a linear relationship with the median value of owner-occupied homes (in $1000). Overall it is difficult to tell from the scatter plot matrix.

```
cor(Boston)[,1]
```

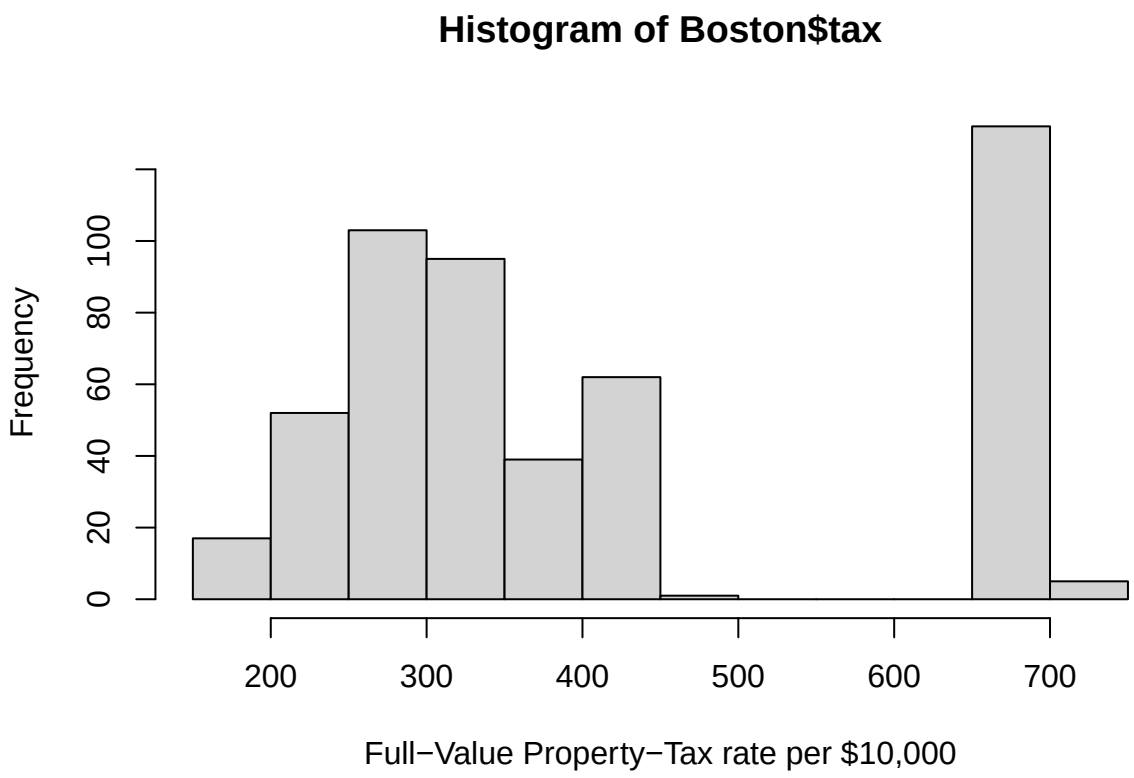
```
##       crim         zn      indus       chas        nox         rm
##  1.0000000 -0.2004692  0.4065834 -0.0558916  0.4209717 -0.2192467
##        age        dis        rad        tax    ptratio      lstat
##  0.3527342 -0.3796701  0.6255051  0.5827643  0.2899456  0.4556215
##       medv
## -0.3883046
```

Rad and Tax have a strong and positive correlation to per capita crime rate while Indus, Nox, Lstat, and Age have a moderate and positive correlation with Crim. Dis and Medv have a moderate and negative correlation with Crim.

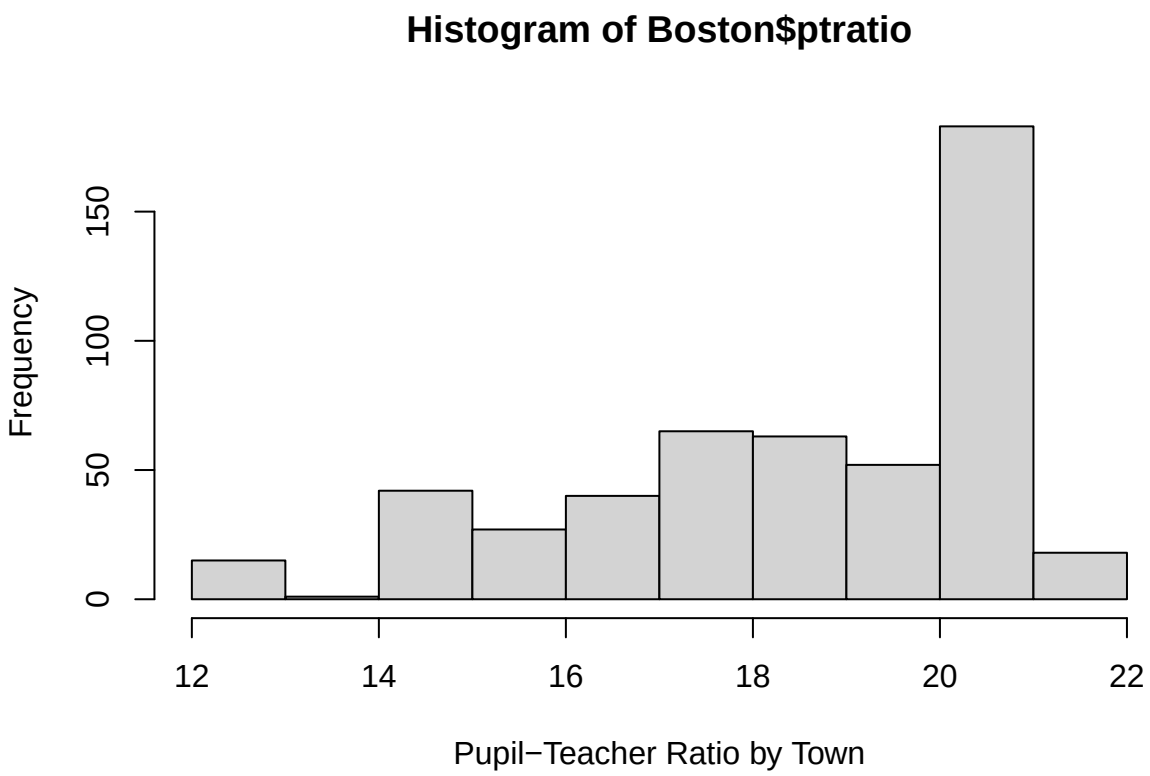
```
hist(Boston$crim, xlab="Per Capita Crime Rate by Town")
```




```
hist(Boston$tax, xlab="Full-Value Property-Tax rate per $10,000")
```




```
hist(Boston$ptratio, xlab="Pupil-Teacher Ratio by Town")
```



The census tracts do not appear to have particularly high crime rates. The data is slightly left skewed from the pupil-teacher ratio by town.

```
sum(Boston$chas == 1)
```

```
## [1] 35
```

There are 35 census tracts bound the Charles river.

```
median(Boston$ptratio)
```

```
## [1] 19.05
```

The median pupil-teacher ratio by town is 19.05.

```
t(subset(Boston, Boston$medv==min(Boston$medv)))
```

```
##             399      406
## crim     38.3518  67.9208
## zn        0.0000   0.0000
## indus    18.1000  18.1000
## chas      0.0000   0.0000
## nox       0.6930   0.6930
## rm        5.4530   5.6830
```

```
## age     100.0000 100.0000
## dis       1.4896   1.4254
## rad      24.0000  24.0000
## tax     666.0000 666.0000
## ptratio  20.2000  20.2000
## lstat    30.5900  22.9800
## medv      5.0000   5.0000
```

Census tracts 399 and 406 had the lowest median value of owner occupied homes.

```
sum(Boston$rm > 8)
```

```
## [1] 13
```

Thirteen census tracts average more than 8 rooms per dwelling.