

EAS506 - Statistical Data Mining I

Homework 1 – Question 3

Paul M Girdler

09/14/18

Content

1	Introduction.....	3
2	Method.....	3
3	Results	4
3.1	Pairwise Scatterplots of Predictors.....	4
3.2	Corrplot	7
3.3	Histograms	8
4	Discussion	8
4.1	Part A.....	8
4.2	Part B.....	8
4.3	Part C.....	9
4.3.1	Crime Rate.....	9
4.3.2	Property Tax Rate	9
4.3.3	Pupil-teacher Ratio	9
4.4	Part D.....	9

1 Introduction

This dataset contains information collected by the U.S Census Service concerning housing in the area of Boston Mass. It was obtained from the **StatLib** archive and has been used extensively throughout the literature to benchmark algorithms. The dataset is small in size with only 506 cases.

The following report summarises my results and answers to Homework Question 3.

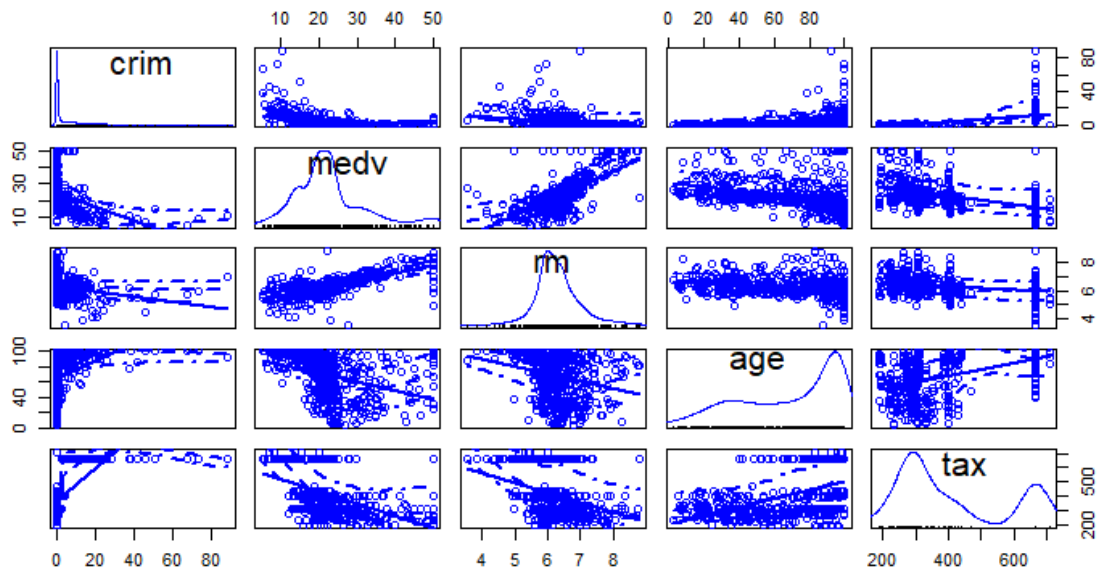
2 Method

Heavily annotated code has been provided to provide the method. Observations are included in the code under each graph as a companion.

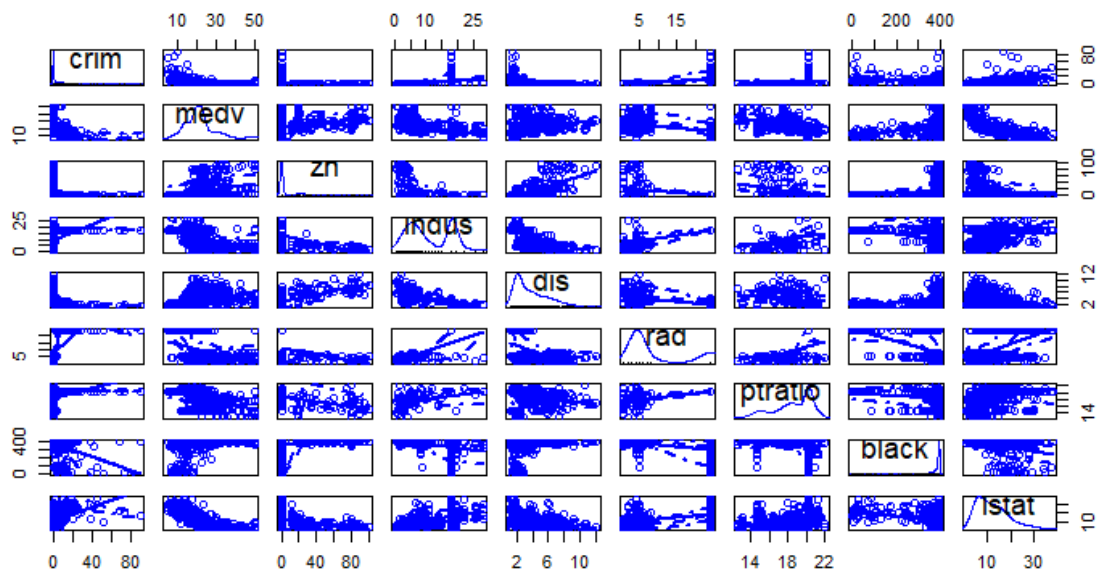
3 Results

3.1 Pairwise Scatterplots of Predictors

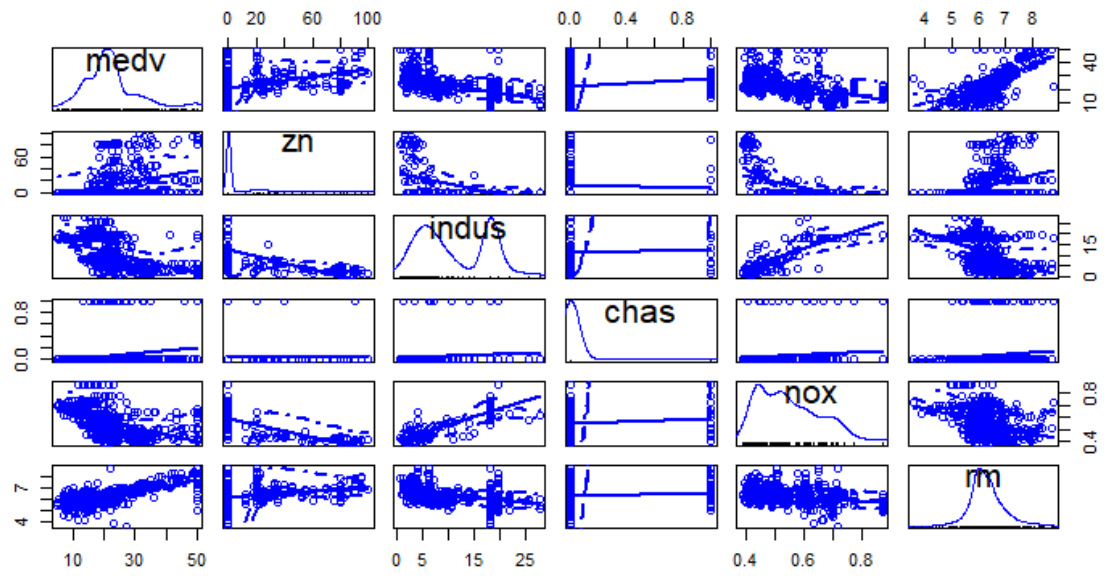
Enhanced Scatterplot Matrix: Comp Feature



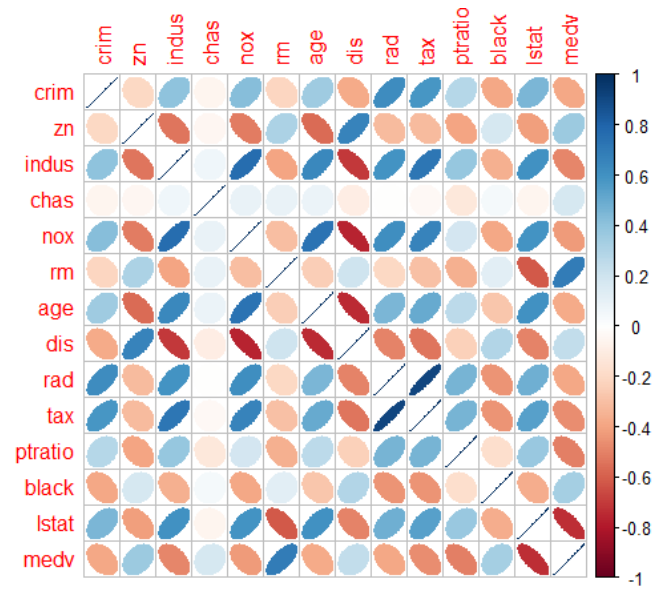
Enhanced Scatterplot Matrix: Neighbourhood Feature



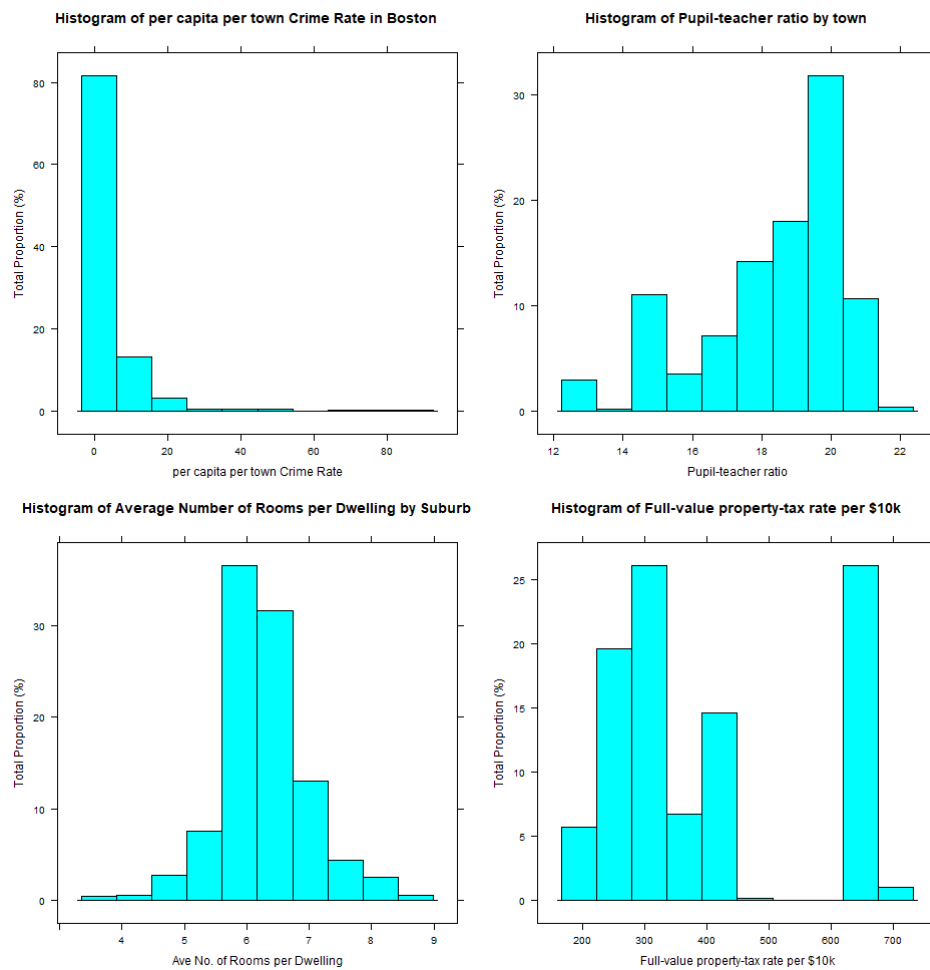
Enhanced Scatterplot Matrix



3.2 Corrplot



3.3 Histograms



4 Discussion

4.1 Part A

Make pairwise scatterplots of the predictors, and describe your findings.

4.2 Part B

Are any of the predictors associated with per capita crime rate?

- RAD (index of accessibility to radial highways), TAX - full-value property-tax rate per \$10,000 shows a significant positive correlation.
- Black (the proportion of blacks by town), DIS (weighted distances to five Boston employment centres), and MEDV (Median value of owner-occupied homes in \$1000's) shows a significant negative correlation.

4.3 Part C

Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.

4.3.1 Crime Rate

Yes. Top 20 crime suburbs have a per capita Crime Rate of between 18 % and 89%.

4.3.2 Property Tax Rate

Yes. The histogram shows a bi or tri modal distribution of full-value property-tax rate in the suburbs of Boston.

The most heavily taxed 20% pay \$666 and up to \$711 on every \$10k in property value.

4.3.3 Pupil-teacher Ratio

Yes. The histogram shows skewed distribution. The top 20 suburbs with worst pupil-teacher ratio, have a ratio around 21-22.

4.4 Part D

In this data set, how many of the suburbs average more than seven rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.

64 suburbs in Boston average more than 7.

Only 13, however, average more than 13.

These dwellings appear to be close to Boston's employment centres, and are in suburbs with high median value of owner-occupied homes.