

Homework 3

Group 1

Contents

1	Introduction	2
2	Statement of the Problem	2
3	Data Exploration	2

Prepared for:

Dr. Nathan Bastian

City University of New York, School of Professional Studies - Data 621

Prepared by:

Group 1

Senthil Dhanapal

Yadu Chittampalli

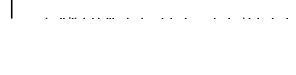
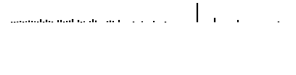
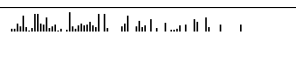
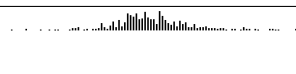
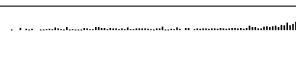
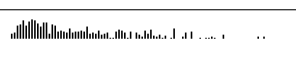
Christophe Hunt

1 Introduction

2 Statement of the Problem

3 Data Exploration

Table 1 : Descriptive Statistics
13 Variables 466 Observations

zn													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	26	0.61	12	0	0	0	0	16	45	80		
lowest : 0 12 18 18 20, highest: 82 85 90 95 100													
indus													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	73	0.98	11	2	3	5	10	18	20	21		
lowest : 0.5 0.7 1.2 1.2 1.2, highest: 18.1 19.6 21.9 25.6 27.7													
chas													
n	missing	unique	Info	Sum	Mean								
466	0	2	0.2	33	0.07								
nox													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	79	1	0.6	0.4	0.4	0.4	0.5	0.6	0.7	0.8		
lowest : 0.4 0.4 0.4 0.4 0.4, highest: 0.7 0.7 0.7 0.8 0.9													
rm													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	419	1	6	5	6	6	6	7	7	8		
lowest : 4 4 4 5 5, highest: 8 8 9 9 9													
age													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	333	1	68	18	26	44	77	94	99	100		
lowest : 3 6 6 6 7, highest: 99 99 99 99 100													
dis													
n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95		
466	0	380	1	4	1	2	2	3	5	7	8		
lowest : 1 1 1 1 1, highest: 9 9 11 11 12													
rad													
n	missing	unique	Info	Mean									
466	0	9	0.96	10									
Frequency 1 2 3 4 5 6 7 8 24													
17 20 36 103 109 25 15 20 121													
% 4 4 8 22 23 5 3 4 26													

tax

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
466	0	63	0.98	410	222	233	281	334	666	666	666

lowest : 187 188 193 198 216, highest: 432 437 469 666 711

ptratio

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
466	0	46	0.98	18	15	15	17	19	20	21	21

lowest : 13 13 14 14 15, highest: 21 21 21 21 22

black

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
466	0	331	0.99	357	88	295	376	391	396	397	397

lowest : 0.3 2.5 2.6 3.5 3.6
highest: 396.3 396.3 396.3 396.4 396.9

lstat

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
466	0	424	1	13	4	5	7	11	17	23	27

lowest : 2 2 2 2 3, highest: 34 34 35 37 38

medv

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
466	0	218	1	23	10	13	17	21	25	35	43

lowest : 5 6 6 7 7, highest: 46 47 48 49 50

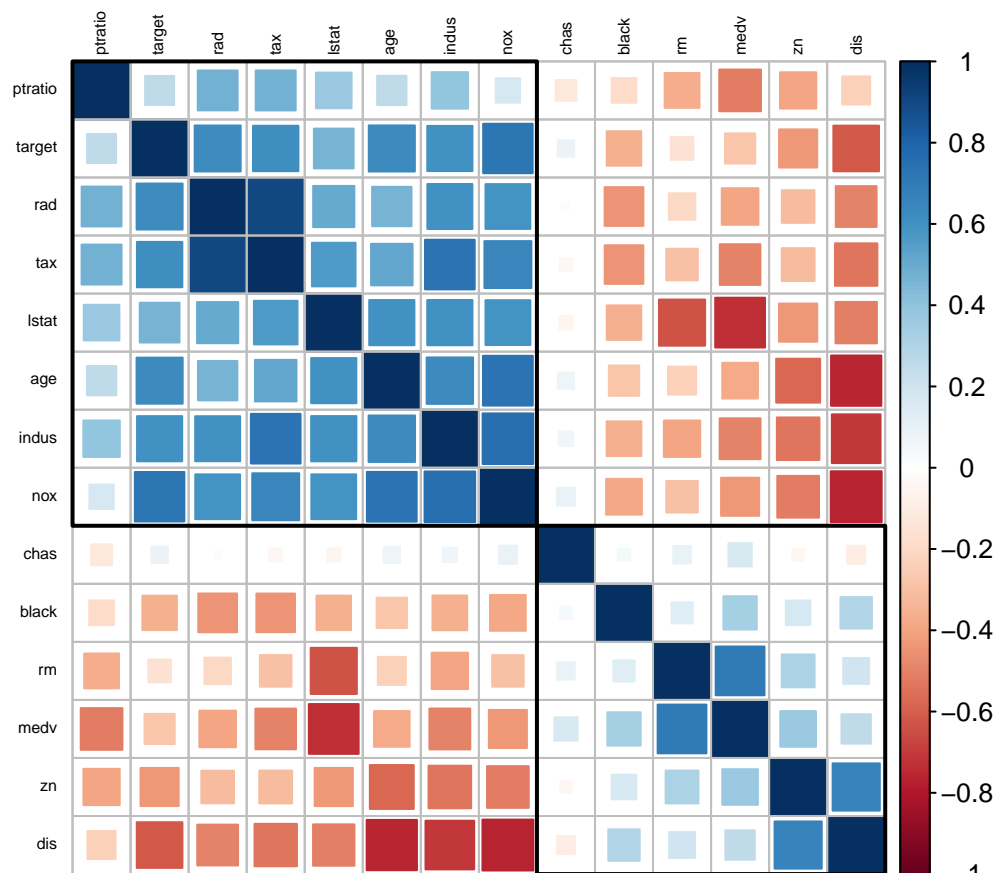


Figure 1: Correlation Plot of Training Data Set