BSF Group 2 Team 10
Members: Paras 1801035, Karthigeyan 1801093, Shanmugapriya 1801115, Niranjan 1801122

Assignment 1

Contents

Question 1

R Command:

```
> quest1 <- data.frame (price=housing$price, lotsize=housing$lotsize,
bedrooms=housing$bedrooms)
```

R output:

```
> quest1
```

	price	lotsize	bedrooms
1	42000	5850	3
2	38500	4000	2
3	49500	3060	3
4	60500	6650	3
5	61000	6360	2
6	66000	4160	3
7	66000	3880	3
8	69000	4160	3
9	83800	4800	3
10	88500	5500	3
11	90000	7200	3
12	30500	3000	2
13	27000	1700	3
14	36000	2880	3
15	37000	3600	2
16	37900	3185	2
17	40500	3300	3
18	40750	5200	4
19	45000	3450	1
20	45000	3986	2

```
[ reached getOption("max.print") -- omitted 526 rows ]
```

Question 2

R Command:

```
> price_4beds <- subset(housing, bedrooms==4)

> mean(price_4beds$price)
```

R output:

[1] 81853.68

Question 3

R Command:

```
> price_per_lotsize <- housing$price/housing$lotsize

> summary(price_per_lotsize)
```

R output:

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.863 10.648 13.613 14.194 16.917 37.714

Question 4

R Command:

```
> stories.summary <- aggregate( housing$price, by=list(housing$stories), FUN = function(x)
c(count=length(x), avg=mean(x),median=median(x),max=max(x)))

> stories.summary
```

R output:

Group.1	x.count	x.avg	x.median	x.max
1	1	227.00	59580.84	55000.00	155000.00
2	2	238.00	68058.19	61550.00	175000.00
3	3	40.00	81240.00	83450.00	190000.00
4	4	41.00	102977.85	105000.00	175000.00

Question 5

R Command:

```
> lot_bn_4k_6k<- subset(housing, lotsize >= 4000 & lotsize <= 6000)

> length(lot_bn_4k_6k$lotsize)
```

R output:

[1] 193

Question 6

R Command:

```
> highest_10_lots <- housing[rev(order(housing$lotsize)),]

> highest_10_lots [1:10,2]
```

R output:

 # A tibble: 10 x 1

	lotsize
	<dbl>
1	16200
2	15600
3	13200
4	13200
5	12944
6	12900
7	12090
8	11460
9	11440
10	11410

Question 7

R Command:

```
> coeff_variation <- aggregate(housing$price, by= list(housing$airco), FUN = function(x)
c(cov=sd(x)/mean(x)))

> coeff_variation
```

R output:

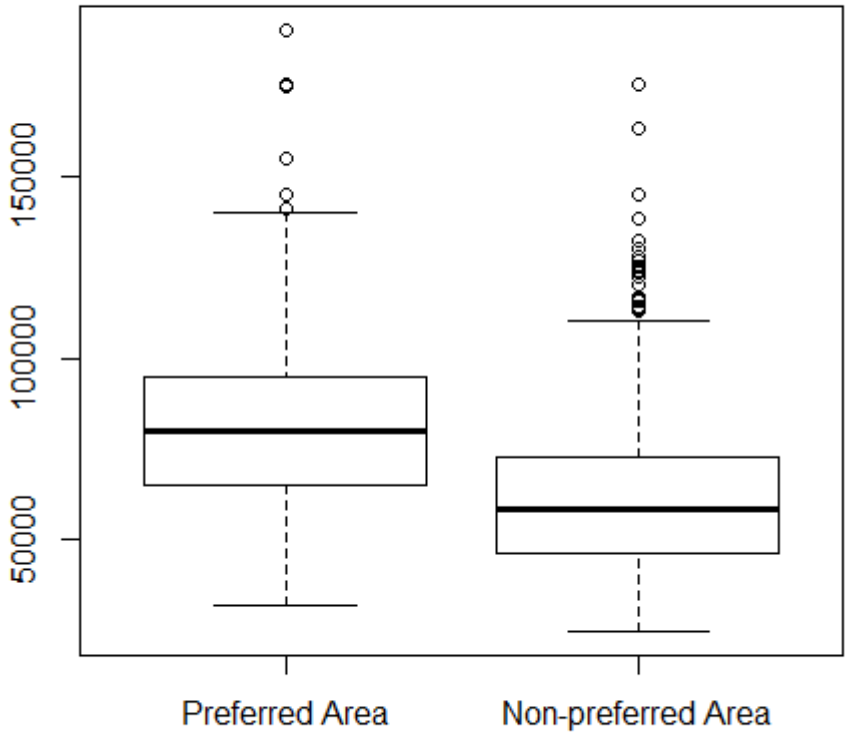
	Group.1	x
1	no	0.3563295
2	yes	0.3314304

Question 8

R Command:

```
> boxplot(housing$price[housing$prefarea=='yes'],  housing$price[housing$prefarea=='no'],
names = c('Preferred Area', 'Non-preferred Area'))
```

R output:



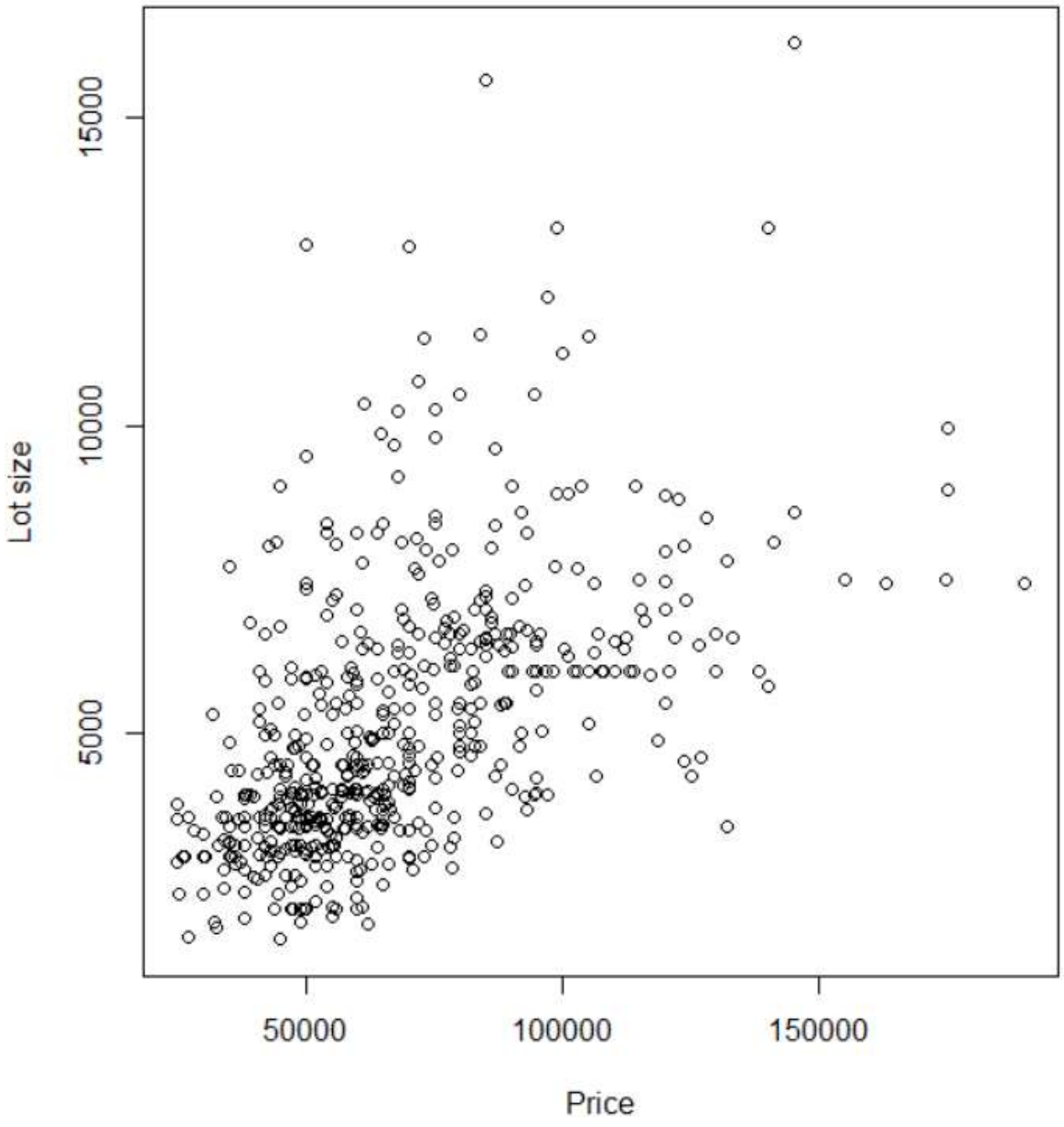
Question 9

R Command:

```
> plot(housing$price, housing$lotsize, xlab = "Price", ylab = "Lot size")

> cor(housing$price,housing$lotsize)
```

R output:

[1] 0.5357957

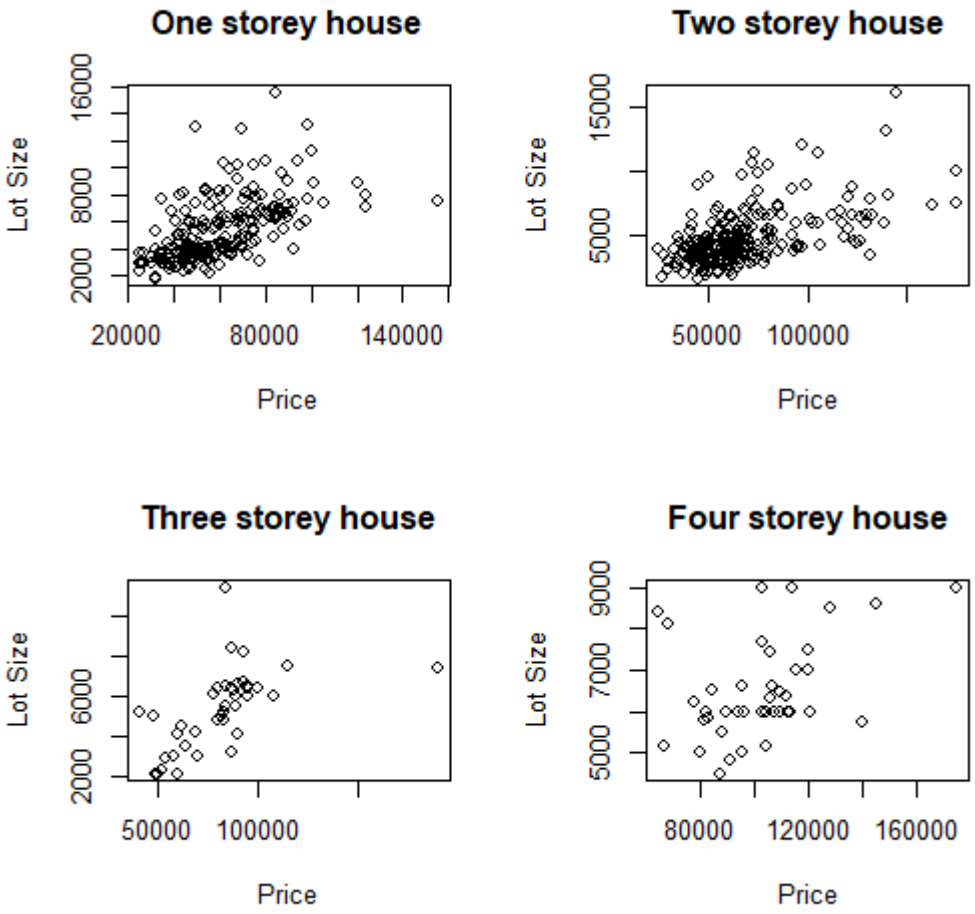


Question 10

R Command:

```
> par(mfrow=c(2,2))

> plot(housing$price[housing$stories==1],housing$lotsize[housing$stories==1], xlab = "Price", ylab = "Lot Size", main="One storey house")

> plot(housing$price[housing$stories==2],housing$lotsize[housing$stories==2], xlab = "Price", ylab = "Lot Size", main="Two storey house")

> plot(housing$price[housing$stories==3],housing$lotsize[housing$stories==3], xlab = "Price", ylab = "Lot Size", main="Three storey house")

> plot(housing$price[housing$stories==4],housing$lotsize[housing$stories==4], xlab = "Price", ylab = "Lot Size", main="Four storey house")
```

R output:



Subjective arguments:

```
> cor(housing$price[housing$stories==1],housing$lotsize[housing$stories==1
])
[1] 0.5824472
> cor(housing$price[housing$stories==2],housing$lotsize[housing$stories==2
])
[1] 0.5466179
> cor(housing$price[housing$stories==3],housing$lotsize[housing$stories==3
])
[1] 0.5717948
> cor(housing$price[housing$stories==4],housing$lotsize[housing$stories==4
])
[1] 0.3959304
```

We can see from above plots and the correlation factors that correlation between housing prices and lot size is better correlated for 1, 2 and 3 storied buildings compared 4 storied buildings. The concentration of houses is also less with 3 or 4 storied buildings.

More houses are concentrated before 80k price for 1 storied, 100k for 2 storied buildings. For 3 storied most houses are concentrated between 50k and 100k and for 4 storied it's between 80k and 120k.