MATH 4322 Lab 2

Ayub Farah

Task 1

```r
library(ISLR2)
head(Boston)
```

```
     crim zn indus chas   nox    rm  age    dis rad tax ptratio lstat medv
1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3  4.98 24.0
2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8  9.14 21.6
3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8  4.03 34.7
4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7  2.94 33.4
5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7  5.33 36.2
6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7  5.21 28.7
```

Question 1

For the 6th suburb of Boston what is the median house value and the average number of rooms per dwelling?
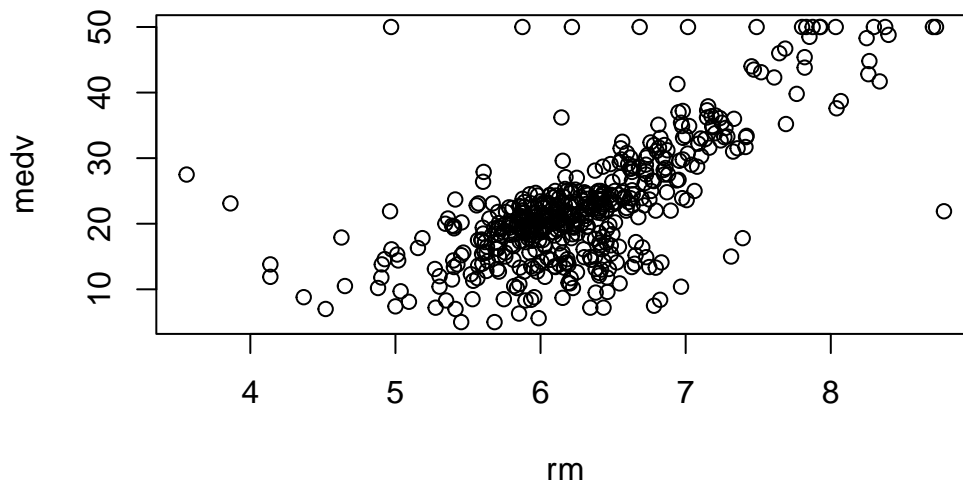
```r
Boston[6, "rm"]
```

```
[1] 6.43
```

```r
Boston[6, "medv"]
```

```
[1] 28.7
```

Task 2

1

```
plot(Boston$rm,Boston$medv,xlab = "rm",ylab = "medv")
```


Question 2

According to the plot what is the relationship between median value of homes and average number of rooms per dwelling?

The relationship is somewhat strong, somewhat linear.

Task 3

Question 3

Which observation has the largest average number of rooms per dwelling? What is the largest average number of rooms per dwelling?

```
max_rm <- which.max(Boston$rm)
max_rm
```

```
[1] 365
```

```
Boston[max_rm, "rm"]
```

```
[1] 8.78
```

Question 4

Which observation has the smallest average number of rooms per dwelling?
What is the smallest average number of rooms per dwelling?

```
min_rm <- which.min(Boston$rm)
min_rm
```

```
[1] 366
```

```
Boston[min_rm, "rm"]
```

```
[1] 3.561
```

Task 4

```
lm.fit <- lm(medv ~ rm, data = Boston)
summary(lm.fit)
```

```
Call:
lm(formula = medv ~ rm, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-23.346  -2.547   0.090   2.986  39.433

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -34.671      2.650  -13.08   <2e-16 ***
rm             9.102      0.419   21.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.616 on 504 degrees of freedom
Multiple R-squared:  0.4835,    Adjusted R-squared:  0.4825
F-statistic: 471.8 on 1 and 504 DF,  p-value: < 2.2e-16
```

Question 5

Give the linear model equation.

$$\hat{y} = 9.102x - 34.671$$

where $\hat{y}$ is predicted `medv` and x is `rm`.

Question 6

What is the percent of variation of `medv`that can be explained by this model?

$$48.35\%$$

Question 7

Is rm a good predictor for `medv`? Justify your answer.

The R^2 for this model is fairly low, since it explains less than half the variation in observed `medv` .

Task 5

Question 8

What is the 95% confidence interval for the slope β_1 of this model?

```
confint.lm(lm.fit, level=0.9)
```

```
                  5 %        95 %
(Intercept) -39.037185 -30.304057
rm            8.411602   9.792616
```
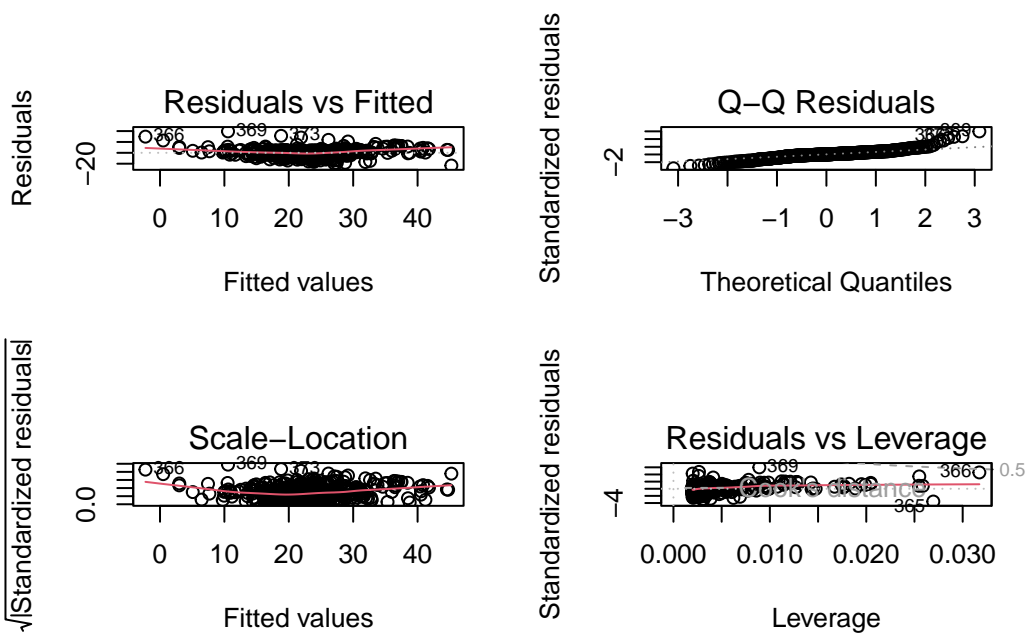
Task 6

Question 9

What is the predicted median value of homes where the average number of
rooms per dwelling is 5?

```
(9.102*5) - 34.671
```

```
[1] 10.839
```

Task 7

```
par(mfrow = c(2,2))
plot(lm.fit)
```


Question 10

Do there appear to be extreme values?

Yes.

Question 11

Which row has the largest leverage?

```
max_leverage <- which.max(hatvalues(lm.fit))
Boston[max_leverage, ]
```

```
       crim zn indus chas   nox    rm  age    dis rad tax ptratio lstat medv
366 4.55587  0  18.1    0 0.718 3.561 87.9 1.6132  24 666    20.2  7.12 27.5
```

Question 12

How many average number of rooms per dwelling and what is the median
value of the homes in this suburb?

```
Boston[max_leverage, "rm"]
```

```
[1] 3.561
```

```
Boston[max_leverage, "medv"]
```

```
[1] 27.5
```