Homework_2

Ritwik Bhriguvanshi

2/4/2020

```
knitr::opts_chunk$set(echo = TRUE)
data <- readRDS("br.rds")
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

#Question 1a Does price per square footage (price/sqft) vary by number of bedrooms (bedrooms) of a house? Report an appropriate scatterplot, and fit a simple linear regression model to answer this question. Interpret the output, specifically the coefficient for the effect of bedrooms on price/sqft. Also, report any concern you may have about the model output.

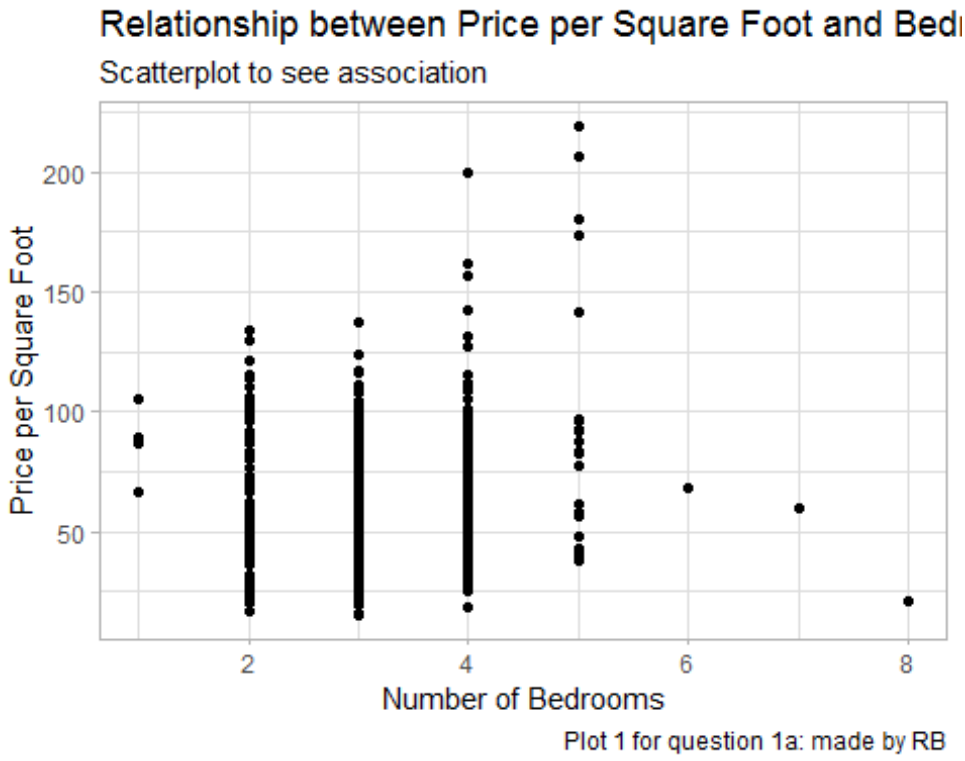
```
PriceBySqft <- data$price/data$sqft  #Define a new variable for PriceBySqft
data$PricePerSqft <- PriceBySqft    #New column for Price per Square feet in the dataframe
dim(data)       # The dimensions will change to 12 columns in the dataframe now
```

```
## [1] 1080   12
```

```
head(data)      # Printing the top 5 rows of the data
```

```
##    price sqft bedrooms baths age occupancy pool style fireplace waterfront
## 1  66500  741        1     1  18         1    1     1         1          0
## 2  66000  741        1     1  18         2    1     1         0          0
## 3  68500  790        1     1  18         1    0     1         1          0
## 4 102000 2783        2     2  18         1    0     1         1          0
## 5  54000 1165        2     1  35         2    0     1         0          0
## 6 143000 2331        2     2  25         1    0     1         1          0
##   dom PricePerSqft
## 1   6     89.74359
## 2  23     89.06883
## 3   8     86.70886
```

```
## 4  50     36.65110
## 5 190     46.35193
## 6  86     61.34706

lm1 <- lm(PricePerSqft ~ bedrooms, data = data); lm1  #Linear model for
PricePerSqft by bedrooms

##
## Call:
## lm(formula = PricePerSqft ~ bedrooms, data = data)
##
## Coefficients:
## (Intercept)     bedrooms
##    65.44764     -0.05086

summary(lm1)

##
## Call:
## lm(formula = PricePerSqft ~ bedrooms, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.736 -12.516  -2.772  10.281 153.659
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 65.44764    3.02015  21.670   <2e-16 ***
## bedrooms    -0.05086    0.92706  -0.055    0.956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.61 on 1078 degrees of freedom
## Multiple R-squared:  2.792e-06,  Adjusted R-squared:  -0.0009248
## F-statistic: 0.00301 on 1 and 1078 DF,  p-value: 0.9563

mygraph1 <- ggplot(data, aes(x = bedrooms, y = PricePerSqft)) +
        geom_point()

    mygraph1 + theme_light() +
    labs (x = "Number of Bedrooms", y = "Price per Square Foot",
    title = "Relationship between Price per Square Foot and Bedrooms",
    subtitle = "Scatterplot to see association",
    caption = "Plot 1 for question 1a: made by RB")
```



Response: The data contains bedrooms for ranging from 1 bedrooms to 8 bedrooms. For 2, 3, and 4 bedroom apartments, the price per square feet is within the same range and does vary greatly. However, there are a few outliers present for apartments with 4 and 5 bedrooms. For apartments with 1, 6, 7, and 8 bedrooms, we do not have enough data points to make an inference or conclusion about the price per square feet.

The model produces a R-sqaure of 2.7 x 10^-6 which is negligible. This illustrates that the variable 'bedrooms' is not able explain the variability for Price per square feet in the model.

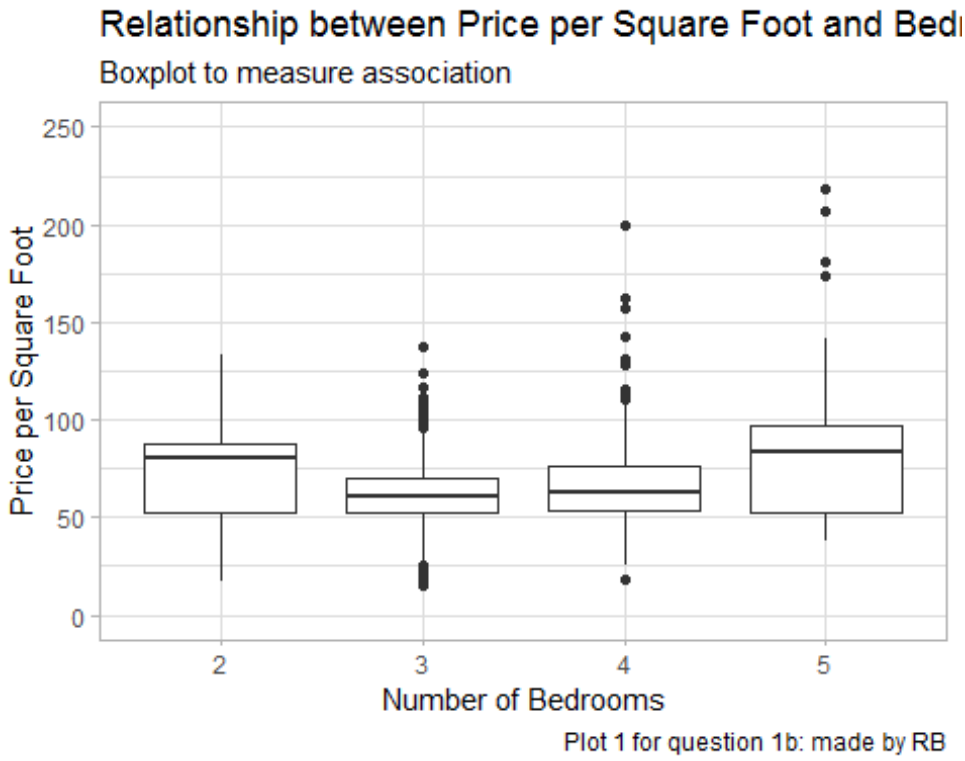
The coefficient of -0.05 indicates that for every increase in bedrooms in an apartment will decrease the price per square foot by $0.05.

The model is not an accurate indication for price per square feet because four bedroom types(1,6,7,8) does not contain enough data points to make inference. Additionally, we would be interested in finding the coefficients for each bedroom separately helping to ascertain the changes in price per square foot meaningfully rather than treating all different types of bedrooms as a single variable. Since, bedrooms are stored as an integer type, the linear model will generate only a single coefficient.

#Question 1b Repeat (a), but unlike in (a), treat the number of bedrooms as a categorical variable (factor). Report an appropriate plot to display this information. Also, fit a simple linear regression model and use the output to describe the association between the two variables. Describe your conserns, if any. Note: You will notice that the number of houses with 1, 6, 7, or 8 bedrooms are very few. You may have to drop these, specially the ones with only one house (6,7,8) since the model would not have enough degrees for reliable estimation and inference.

```r
data1 <- data
data1$bedrooms <- as.factor(data1$bedrooms)
newdata <- data1[which(data1$bedrooms %in% c(2,3,4,5)),] #Subsetting for
selected bedrooms

PriceBySqft <- newdata$price / newdata$sqft
newdata <- cbind(newdata, PriceBySqft)      #Creating a new column in the
dataframe
lm2 <- lm(PriceBySqft ~ bedrooms, data = newdata);lm2   #Linear Model for
PriceBySqft by Bedroom

##
## Call:
## lm(formula = PriceBySqft ~ bedrooms, data = newdata)
##
## Coefficients:
## (Intercept)    bedrooms3    bedrooms4    bedrooms5
##      73.689      -11.752       -7.404       19.373

summary(lm2)

##
## Call:
## lm(formula = PriceBySqft ~ bedrooms, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.550 -11.382  -1.586   9.770 133.158
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   73.689      1.793  41.105  < 2e-16 ***
## bedrooms3    -11.752      1.975  -5.951 3.60e-09 ***
## bedrooms4     -7.404      2.181  -3.394 0.000713 ***
## bedrooms5     19.373      4.699   4.123 4.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.83 on 1068 degrees of freedom
## Multiple R-squared:  0.06958,    Adjusted R-squared:  0.06697
## F-statistic: 26.62 on 3 and 1068 DF,  p-value: < 2.2e-16

mygraph2 <- ggplot(newdata, aes(x = bedrooms, y = PriceBySqft, group =
bedrooms)) +
  geom_boxplot() + scale_y_continuous(limit = c(0,250)) +
  theme_light()

mygraph2 +
    labs (x = "Number of Bedrooms", y = "Price per Square Foot",
    title = "Relationship between Price per Square Foot and Bedrooms",
```

```
    subtitle = "Boxplot to measure association",
    caption = "Plot 1 for question 1b: made by RB")
```



Response: After removing bedrooms 1,6,7,8 and developing a linear model, we find the intercepts and coefficients of bedrooms. The linear model uses '2 bedrooms' as a reference point. The p-values for the coefficients are below 5% which indicates that there is significant difference on Price per Square Foot for different bedroom sizes.

The intercept of 73.69 indicates that the average mean price per square foot for 2 bedroom apartment is $73.69. Similarly, for bedroom 3 the price per square feet will increase by approximately $ 61.94 (decrease by $ 11.75 relative to bedroom 2). The price per sqaure feet will increase by $66.29 (decrease by $7.4 relative to bedroom 2) for 4 bedroom house. The price per square feet will increase by $93.1 (increase by $19.37 relative to bedroom 2) for a 5 bedroom house.

However, the model explains only about 7% of the variation in the effect for Price per Square footage with respect to different bedroom sizes.

Referring to the boxplot, the median price per square foot for 2,3,4,and 5 bedroom units varies between $50 and $75 approximately. Bedrooms units comprising of 4 and 5 bedrooms have a few right skewed outliers depicting higher price per sqaure foot as compared to the two other bedroom units. Overall, the price per square feet does not vary greatly among the different bedroom units leaving a few outliers.

Question 1c

Here you want to understand the effect of square footage (sqft) and number of bedrooms (bedrooms)on price of a house (price). Fit a mulitple linear regression model of price on sqft and bedrooms to answer this question. Report and interpret the estimated coefficients corresponding to sqft and bedrooms, and indicate whether the coefficients are significant. Report any concerns you may have about the results you get from this model.

```
lm3 <- lm(price ~ sqft + bedrooms, data = data);lm3   # Linear Model

##
## Call:
## lm(formula = price ~ sqft + bedrooms, data = data)
##
## Coefficients:
## (Intercept)         sqft     bedrooms
##      -16842          103       -21379

summary(lm3)

##
## Call:
## lm(formula = price ~ sqft + bedrooms, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -384007  -32970   -1772   29299  916451
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16842.109  11337.256  -1.486    0.138
## sqft           103.048      3.277  31.450  < 2e-16 ***
## bedrooms    -21379.382   4655.504  -4.592  4.9e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79090 on 1077 degrees of freedom
## Multiple R-squared:  0.5867, Adjusted R-squared:  0.586
## F-statistic: 764.6 on 2 and 1077 DF,  p-value: < 2.2e-16
```

Response: 'Sqft' has a coefficient of 103.05 which indicates that Price of a house unit will increase by $103.05 with for each additional square foot, keeping 'bedrooms' as fixed. 'Bedrooms' has a coefficient of -21379.38 which demonstrates that price of a house will decrease by $21,379.38 for each additional bedroom, keeping 'sqft' as fixed.

The model interprets the coefficient of bedrooms as a whole rather than breaking it down into different number of bedrooms for a house.

Both these variables are significant since they vary greatly and have a p-value of less than 5%.

Question 1d

Repeat (c), except now you will treat number of bedrooms (bedrooms) as a categorical variable. Note: You will notice that the number of houses with 1, 6, 7, or 8 bedrooms are very few. You may have to drop these.

```
data1$bedrooms <- as.factor(data1$bedrooms)
newdata <- data1[which(data1$bedrooms %in% c(2,3,4,5)),] # Subsetting for
selected bedrooms
lm4 <- lm(price ~ sqft + bedrooms, data = newdata);lm4 # Linear Model

##
## Call:
## lm(formula = price ~ sqft + bedrooms, data = newdata)
##
## Coefficients:
## (Intercept)         sqft    bedrooms3    bedrooms4    bedrooms5
##   -33957.11        97.82    -42213.70    -57337.96     71863.70

summary(lm4)

##
## Call:
## lm(formula = price ~ sqft + bedrooms, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -504923  -30036    1700   23926  893812
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33957.105   7727.729  -4.394 1.22e-05 ***
## sqft            97.820      3.187  30.693  < 2e-16 ***
## bedrooms3   -42213.703   7611.796  -5.546 3.69e-08 ***
## bedrooms4   -57337.958   9788.703  -5.858 6.25e-09 ***
## bedrooms5    71863.697  20302.955   3.540 0.000418 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 75680 on 1067 degrees of freedom
## Multiple R-squared:  0.6214, Adjusted R-squared:   0.62
## F-statistic: 437.9 on 4 and 1067 DF,  p-value: < 2.2e-16
```

Response: The price of a house will decrease by $42,213 for a three bedroom unit keeping 'sqft' as fixed. The price of a house will decrease by $57,338 for a four bedroom unit keeping 'sqft' as fixed. The price of a house will increase by $71,864 for a five bedroom unit keeping 'sqft' as fixed.

All the p-values are significant since they are less than the 5% threshold.

The linear model performs a litle better compared to the model in question 1c as the R-square increases by 4% explaining a little more variation in the model.

Making the number of bedrooms as categorical, we can compare the mean difference for 3,4,5 bedrooms as compared to 2 bedrooms units keeping sqft as fixed.

#Question 1e

Repeat (d), except now you will include the interaction term in your model. Does effect of sqft on price differ by number of bedrooms? If so, use your model output to characterize effect of sqft on price for each bedrooms level that you include in the model. Describe the differences you see in detail.

```
lm5 <- lm(price ~ sqft + bedrooms + sqft:bedrooms , data = newdata);lm5

##
## Call:
## lm(formula = price ~ sqft + bedrooms + sqft:bedrooms, data = newdata)
##
## Coefficients:
##    (Intercept)            sqft       bedrooms3       bedrooms4
##       49428.68           33.91       -44529.82      -180190.18
##      bedrooms5  sqft:bedrooms3  sqft:bedrooms4  sqft:bedrooms5
##     -541294.20           25.35           76.61          175.38

summary(lm5)

##
## Call:
## lm(formula = price ~ sqft + bedrooms + sqft:bedrooms, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -843468  -24355   -1235   17854  832983
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      49428.68   21798.29   2.268   0.0236 *
## sqft                33.91      16.11   2.105   0.0355 *
## bedrooms3       -44529.82   23774.60  -1.873   0.0613 .
## bedrooms4      -180190.18   25971.14  -6.938 6.89e-12 ***
## bedrooms5      -541294.20   47054.85 -11.503  < 2e-16 ***
## sqft:bedrooms3      25.35      16.68   1.520   0.1288
## sqft:bedrooms4      76.61      16.68   4.592 4.92e-06 ***
## sqft:bedrooms5     175.38      18.10   9.688  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67290 on 1064 degrees of freedom
## Multiple R-squared:  0.7015, Adjusted R-squared:  0.6996
## F-statistic: 357.3 on 7 and 1064 DF,  p-value: < 2.2e-16
```

```r
# OR
lm6 <- lm(price ~ sqft + bedrooms + sqft*bedrooms, data = newdata);lm6

##
## Call:
## lm(formula = price ~ sqft + bedrooms + sqft * bedrooms, data = newdata)
##
## Coefficients:
##    (Intercept)            sqft       bedrooms3       bedrooms4
##       49428.68           33.91       -44529.82      -180190.18
##      bedrooms5  sqft:bedrooms3  sqft:bedrooms4  sqft:bedrooms5
##     -541294.20           25.35           76.61          175.38

summary(lm6)

##
## Call:
## lm(formula = price ~ sqft + bedrooms + sqft * bedrooms, data = newdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -843468  -24355   -1235   17854  832983
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      49428.68   21798.29   2.268   0.0236 *
## sqft                33.91      16.11   2.105   0.0355 *
## bedrooms3       -44529.82   23774.60  -1.873   0.0613 .
## bedrooms4      -180190.18   25971.14  -6.938 6.89e-12 ***
## bedrooms5      -541294.20   47054.85 -11.503  < 2e-16 ***
## sqft:bedrooms3      25.35      16.68   1.520   0.1288
## sqft:bedrooms4      76.61      16.68   4.592 4.92e-06 ***
## sqft:bedrooms5     175.38      18.10   9.688  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67290 on 1064 degrees of freedom
## Multiple R-squared:  0.7015, Adjusted R-squared:  0.6996
## F-statistic: 357.3 on 7 and 1064 DF,  p-value: < 2.2e-16
```

Response: The model is better explained as it can accurately predict the price for the house now:

Price for a 2 bedroom house: 49428.68 + 33.91 * sqft Interpretation: For a 2 bedroom unit, the price of a house will increase by $33.91 for each additional sqaure foot.

Price for a 3 bedroom house: 49428.68 + 33.91 * sqft + (1 x -44529.82) + 25.35 * sqft Interpretation: For a 3 bedroom unit, the price of a house will increase by $59.26 for each additional sqaure foot.

Price for a 4 bedroom house: 49428.68 + 33.91 * sqft + (1 x -180190) + 76.61 * sqft
Interpretation: For a 4 bedroom unit, the price of a house will increase by $110.52 for each additional sqaure foot.

Price for a 5 bedroom house: 49428.68 + 33.91 * sqft + (1 x -541294) + 175.38 * sqft
Interpretation: For a 5 bedroom unit, the price of a house will increase by $209.29 for each additional sqaure foot.

```
q2 <- read.csv("ca_hospitals_market_share_data_homework2.csv")
q2 <- q2[which(q2$Year %in% c(2008)),]
```

#Question 2a Develop a mulitple linear regression model to understand association of Market_Share (response) withthe other variables (predictors). Write a paragraph summarizing your findings regarding the association of the predictors with the response, and comment on the significance of each of the predictors in your model.

```
q2a <- q2
q2a <- subset(q2a, select = -c(Year, Hospital_ID))

lm7 <- lm(Market_Share~.,data = q2a);lm7

##
## Call:
## lm(formula = Market_Share ~ ., data = q2a)
##
## Coefficients:
##                 (Intercept)      Overall_hospital_rating
##                  -1.224e-01                    2.366e-01
## Level_of_Market_Competition                  Size_in_Beds
##                    2.397e-03                    1.374e-03
##              Mortality_rate            Readdmission_rate
##                  -5.771e-01                   -1.993e-01
## Average_Length_Of_Stay_days           Cost_Per_Discharge
##                  -1.803e-03                   -2.496e-07
##                    TeachYES
##                   6.500e-03

summary(lm7)

##
## Call:
## lm(formula = Market_Share ~ ., data = q2a)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44936 -0.05887 -0.00615  0.05503  0.34688
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -1.224e-01  6.140e-02  -1.993 0.047122 *
## Overall_hospital_rating      2.366e-01  6.359e-02   3.720 0.000235 ***
```

```
## Level_of_Market_Competition  2.397e-03  2.966e-02   0.081 0.935617
## Size_in_Beds                 1.374e-03  4.965e-05  27.671  < 2e-16 ***
## Mortality_rate              -5.771e-01  3.756e-01  -1.536 0.125449
## Readdmission_rate           -1.993e-01  9.979e-02  -1.997 0.046678 *
## Average_Length_Of_Stay_days -1.803e-03  5.252e-04  -3.434 0.000673 ***
## Cost_Per_Discharge          -2.496e-07  1.634e-07  -1.528 0.127539
## TeachYES                     6.500e-03  2.651e-02   0.245 0.806447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1006 on 322 degrees of freedom
## Multiple R-squared:  0.8172, Adjusted R-squared:  0.8126
## F-statistic: 179.9 on 8 and 322 DF,  p-value: < 2.2e-16
```

Response: Year and Hospital ID were excluded from the model while all the other variables were incorporated in the linear model. Half of the coefficients are positive while the other half are negative. The four significant coefficient with p-value less than 5 percent are Overall Hospital Rating, Size in Beds, Re-admission rate, and Average Length of days stayed. The two driving coefficients to predict market share are overall hospital rating and mortality rate. While Overall hospital rating has the highest coefficient value, mortality rate has the lowest coefficient value, depicting a change in the these two predictor variables would greatly affect the response variable (market share). The variables in the model explain 82% of the variation in market share. The average value for market share is 0.279. Hence, the RMSE of 0.1 is good for the values which are above the average market share and is bad for lower market share values.

#Question 2b Rebuild the model in (a) including the interaction term between Teach and Size_in_Beds. Is the effect of additional bed in a hospital on market share different for teaching non-teaching hospitals? If so, how? Also, investigate if Teach interacts with any of the other predictors in the model (with respect to its effect on market share.

```
lm8 <- lm(Market_Share ~ Overall_hospital_rating +
Level_of_Market_Competition +
          Size_in_Beds +  Mortality_rate + Readdmission_rate +
          Average_Length_Of_Stay_days + Cost_Per_Discharge + Teach +
          Teach*Size_in_Beds, data = q2a); lm8

##
## Call:
## lm(formula = Market_Share ~ Overall_hospital_rating +
Level_of_Market_Competition +
##     Size_in_Beds + Mortality_rate + Readdmission_rate +
Average_Length_Of_Stay_days +
##     Cost_Per_Discharge + Teach + Teach * Size_in_Beds, data = q2a)
##
## Coefficients:
##                  (Intercept)      Overall_hospital_rating
##                   -1.235e-01                    2.309e-01
## Level_of_Market_Competition                 Size_in_Beds
```

```
##                   5.899e-03                    1.403e-03
##              Mortality_rate            Readdmission_rate
##                  -5.636e-01                   -1.991e-01
## Average_Length_Of_Stay_days           Cost_Per_Discharge
##                  -1.854e-03                   -2.235e-07
##                    TeachYES        Size_in_Beds:TeachYES
##                   7.769e-02                   -1.700e-04
```

```
summary(lm8)     # Teach has an interaction with Size in Beds
```

```
##
## Call:
## lm(formula = Market_Share ~ Overall_hospital_rating +
Level_of_Market_Competition +
##     Size_in_Beds + Mortality_rate + Readdmission_rate +
Average_Length_Of_Stay_days +
##     Cost_Per_Discharge + Teach + Teach * Size_in_Beds, data = q2a)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45824 -0.05583 -0.00486  0.05730  0.34273
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -1.235e-01  6.135e-02  -2.013 0.044932 *
## Overall_hospital_rating      2.309e-01  6.368e-02   3.627 0.000334 ***
## Level_of_Market_Competition  5.899e-03  2.975e-02   0.198 0.842955
## Size_in_Beds                 1.403e-03  5.462e-05  25.690  < 2e-16 ***
## Mortality_rate              -5.636e-01  3.754e-01  -1.501 0.134238
## Readdmission_rate           -1.991e-01  9.969e-02  -1.997 0.046649 *
## Average_Length_Of_Stay_days -1.854e-03  5.261e-04  -3.523 0.000489 ***
## Cost_Per_Discharge          -2.235e-07  1.645e-07  -1.358 0.175329
## TeachYES                     7.769e-02  6.153e-02   1.263 0.207625
## Size_in_Beds:TeachYES       -1.700e-04  1.326e-04  -1.282 0.200837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1005 on 321 degrees of freedom
## Multiple R-squared:  0.8181, Adjusted R-squared:  0.813
## F-statistic: 160.4 on 9 and 321 DF,  p-value: < 2.2e-16
```

```
lm9 <- lm(Market_Share ~ .*Teach, data = q2a);lm9  # New Model
```

```
##
## Call:
## lm(formula = Market_Share ~ . * Teach, data = q2a)
##
## Coefficients:
##                           (Intercept)
##                            -1.445e-01
##              Overall_hospital_rating
```

```
##                              2.433e-01
##          Level_of_Market_Competition
##                              3.437e-03
##                          Size_in_Beds
##                              1.403e-03
##                         Mortality_rate
##                             -5.570e-01
##                       Readdmission_rate
##                             -1.525e-01
##          Average_Length_Of_Stay_days
##                             -1.958e-03
##                     Cost_Per_Discharge
##                             -1.568e-07
##                               TeachYES
##                              6.852e-01
##     Overall_hospital_rating:TeachYES
##                             -2.565e-01
## Level_of_Market_Competition:TeachYES
##                              4.649e-01
##                 Size_in_Beds:TeachYES
##                             -1.116e-04
##                Mortality_rate:TeachYES
##                             -8.871e+00
##            Readdmission_rate:TeachYES
##                              4.428e-01
## Average_Length_Of_Stay_days:TeachYES
##                             -5.102e-02
##          Cost_Per_Discharge:TeachYES
##                             -2.459e-06
```

```
summary(lm9)
```

```
##
## Call:
## lm(formula = Market_Share ~ . * Teach, data = q2a)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46413 -0.05229 -0.00524  0.05665  0.34357
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                         -1.445e-01  6.068e-02  -2.382
## Overall_hospital_rating              2.433e-01  6.276e-02   3.877
## Level_of_Market_Competition          3.437e-03  2.927e-02   0.117
## Size_in_Beds                         1.403e-03  5.363e-05  26.165
## Mortality_rate                      -5.570e-01  3.695e-01  -1.507
## Readdmission_rate                   -1.525e-01  9.934e-02  -1.535
## Average_Length_Of_Stay_days         -1.958e-03  5.185e-04  -3.776
## Cost_Per_Discharge                  -1.568e-07  1.634e-07  -0.959
```

```
## TeachYES                                   6.852e-01  7.304e-01   0.938
## Overall_hospital_rating:TeachYES          -2.565e-01  7.985e-01  -0.321
## Level_of_Market_Competition:TeachYES       4.649e-01  4.331e-01   1.073
## Size_in_Beds:TeachYES                     -1.116e-04  1.393e-04  -0.801
## Mortality_rate:TeachYES                   -8.871e+00  6.273e+00  -1.414
## Readdmission_rate:TeachYES                 4.428e-01  1.023e+00   0.433
## Average_Length_Of_Stay_days:TeachYES      -5.102e-02  1.932e-02  -2.641
## Cost_Per_Discharge:TeachYES               -2.459e-06  1.100e-06  -2.236
##                                          Pr(>|t|)
## (Intercept)                              0.017801 *
## Overall_hospital_rating                  0.000129 ***
## Level_of_Market_Competition              0.906591
## Size_in_Beds                              < 2e-16 ***
## Mortality_rate                           0.132686
## Readdmission_rate                        0.125701
## Average_Length_Of_Stay_days              0.000190 ***
## Cost_Per_Discharge                       0.338114
## TeachYES                                 0.348900
## Overall_hospital_rating:TeachYES         0.748282
## Level_of_Market_Competition:TeachYES     0.283902
## Size_in_Beds:TeachYES                    0.423623
## Mortality_rate:TeachYES                  0.158288
## Readdmission_rate:TeachYES               0.665503
## Average_Length_Of_Stay_days:TeachYES     0.008678 **
## Cost_Per_Discharge:TeachYES              0.026036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09855 on 315 degrees of freedom
## Multiple R-squared:  0.8282, Adjusted R-squared:  0.8201
## F-statistic: 101.3 on 15 and 315 DF,  p-value: < 2.2e-16
```

Response: The variables size in beds and Teach does not interact with each other as the p-value is less than five percent and it is not statistically significant. So, adding an interaction term between size in beds and Teach will not help us to better predict the response variable.

For the new model, we try to create an interaction term of Teach with all the variables. There can be numerous combinations formed by making an interaction between teach and other variables while building a linear model. However, the new model provides a significant interaction of Teach with Average Length of Stay in Days and Cost per Discharge. The interaction of Teach with these two variables would only be significant if we build a linear model incorporating an interaction of Teach with all the variables in the dataframe. The newer model explains the variation little better and has a slightly lower root mean squared error from the previous model.