Final Project

Hunter Garrison, Kevin Smith, Reese Madsen, Giulio Martini

2024-04-26

Section 1: Introduction

In this project, we are working with a dataset comprising of the observed prices of houses and corresponding attributes of each house. This dataset is from the database website Kaggle and is comprised of 545 observations of houses in India with 13 variables recorded for each specific house. Below are the attributes of each house recorded in the dataset:
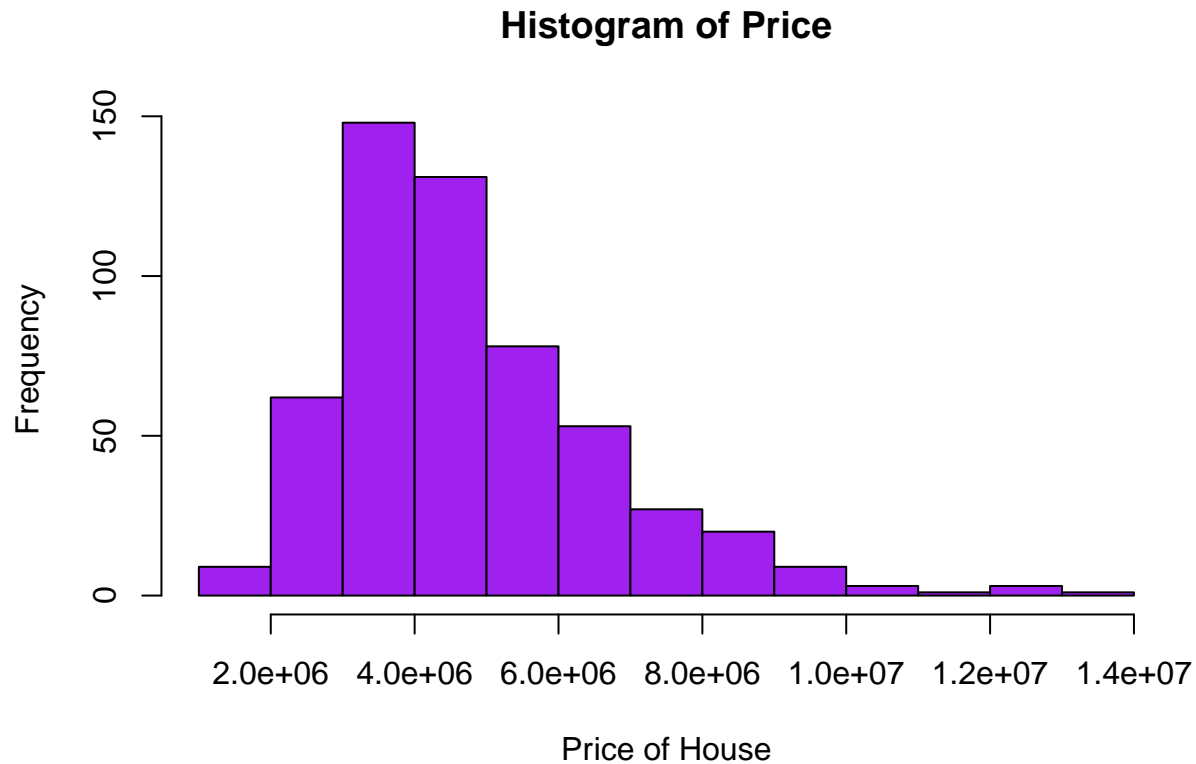
Attribute	Description	Data Type
price	Price of the houses in rupees. (this is our response variable)	Integer
area	Area of a house in square feet	Integer
bedrooms	Number of bedrooms in house	Integer
bathrooms	Number of bathrooms in house	Integer
stories	Number of stories in house	Integer
mainroad	Whether or not house is connected to main road	Boolean
guestroom	Whether or not house has a guest room	Boolean
basement	Whether or not house has a basement	Boolean
hotwaterheating	Whether or not house has a hot-water heater	Boolean
airconditioning	Whether or not house has air conditioning	Boolean
parking	Number of parking spots at house	Integer
prefarea	Whether or not the house is in a preferred area	Boolean
furnishingstatus	Furnishing status of the house (furnished, semi-furnished, unfurnished)	String

The majority of our predictors in this dataset are of type boolean, with the remainders being integers and one categorical predictor comprising of strings. Below is the first 5 lines of our dataset:

```
##       price area bedrooms bathrooms stories mainroad guestroom basement
## 1 13300000 7420        4         2       3      yes        no       no
## 2 12250000 8960        4         4       4      yes        no       no
## 3 12250000 9960        3         2       2      yes        no      yes
## 4 12215000 7500        4         2       2      yes        no      yes
## 5 11410000 7420        4         1       2      yes       yes      yes
##   hotwaterheating airconditioning parking prefarea furnishingstatus
## 1              no             yes       2      yes        furnished
## 2              no             yes       3       no        furnished
## 3              no              no       2      yes   semi-furnished
## 4              no             yes       3      yes        furnished
## 5              no             yes       2       no        furnished
```
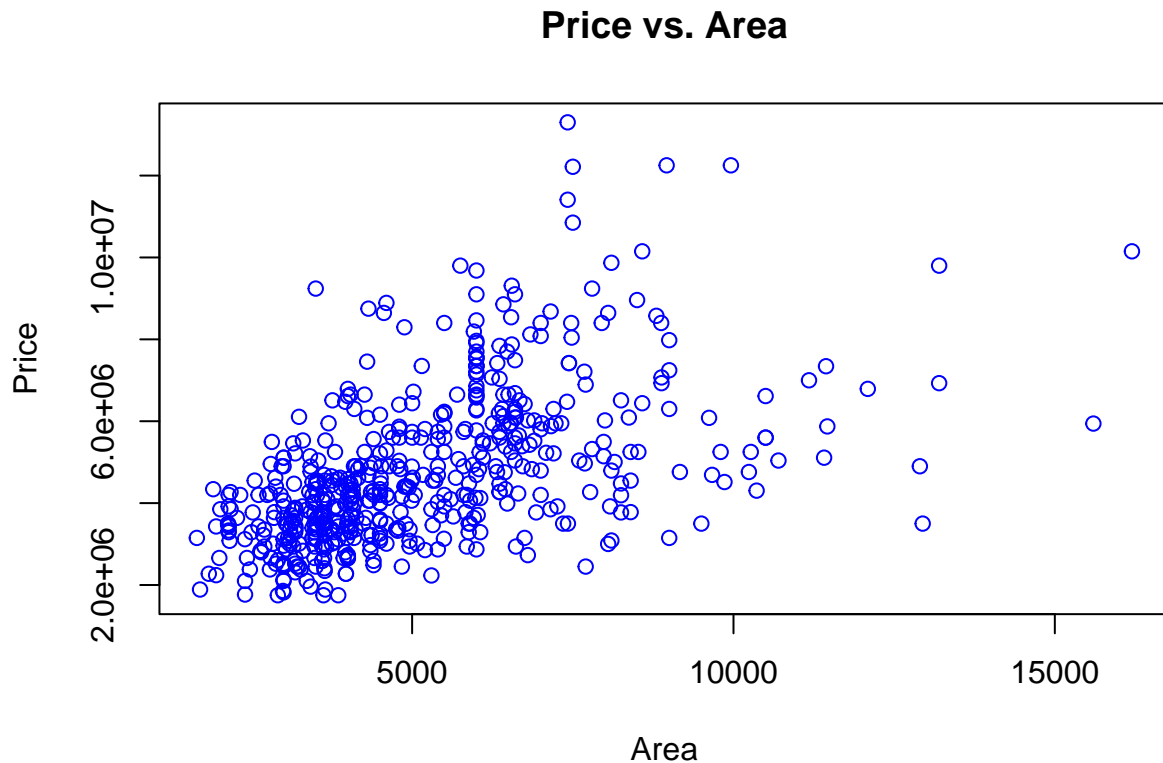
We are attempting to create a linear regression model that best explains the variance in the price of the houses using the predictors in our dataset. In this report, we will explore our data by finding collinearity, outliers, influential points, and by checking error assumption violations. We will also discover which variables are the most important by using model selection, and test OLS as well as GLS and WLS regressions. After doing this, we will combine all the knowledge we have gained about our data set in order to find the model in which we believe best explains the variance in the price of the houses.

We will now perform some exploratory data analysis to show the distribution of our response variable as well as some key relationships in this dataset. Below is a histogram of our response variable, price, as well as a table showing some key statistical information about the response:

Histogram of Price


```
##       Min      Q1  Median    Mean      Q3      Max NA_Count
## 1 1750000 3430000 4340000 4766729 5740000 13300000        0
```

From our histogram and table, we see that our data is right skewed with a small amount of houses showing prices much farther than the mean. These houses may get removed as outliers later on in our report. Below, we also show a scatter plot of price vs. area:

Price vs. Area

We can see a visible linear correlation between the two variables, suggesting that price will be a large factor in the models we will test.
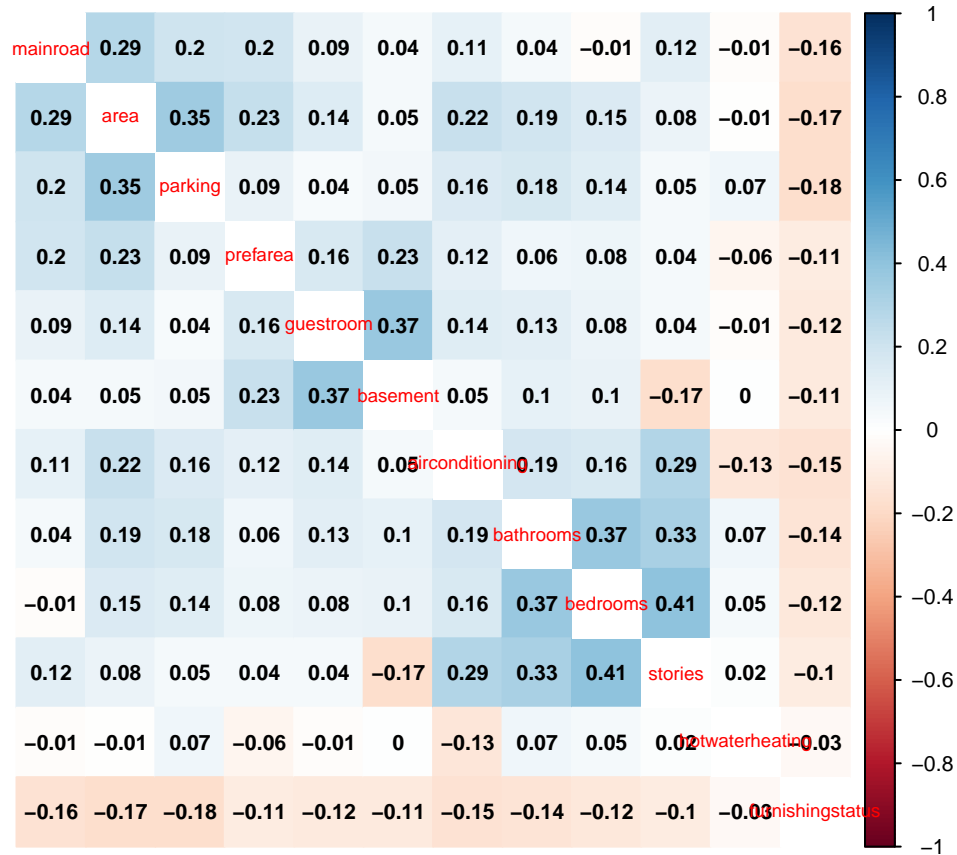
All in all, this dataset is very interesting to us because of how it can teach us the relationships between a house's price and attributes about the house and can help us better understand the housing market when we need to go out and purchase a house for ourselves.

Section 2: Regression Analysis

Collinearity

In our study of the dataset we must ask ourselves whether or not the data is plagued by the issue of collinearity. Do the predictors present a linear relationship, not just with the response, but within each other as well? Collinearity can hurt a model's performance; inflated R^2 scores and diminished p-values are but a few of the problems that arise alongside predictor collinearity... and so, the dataset must be tested for it.

Let us begin by creating a ***correlation heat-map*** for each predictor. This measures the correlation between predictors, and provides us with a visual method of estimating it. Each categorical column is *label encoded* - their class value is turned to a constant integer - and is added to the correlation matrix.

	mainroad	area	parking	prefarea	guestroom	basement	airconditioning	bathrooms	bedrooms	stories	hotwaterheating	furnishingstatus
mainroad	mainroad	0.29	0.2	0.2	0.09	0.04	0.11	0.04	−0.01	0.12	−0.01	−0.16
area	0.29	area	0.35	0.23	0.14	0.05	0.22	0.19	0.15	0.08	−0.01	−0.17
parking	0.2	0.35	parking	0.09	0.04	0.05	0.16	0.18	0.14	0.05	0.07	−0.18
prefarea	0.2	0.23	0.09	prefarea	0.16	0.23	0.12	0.06	0.08	0.04	−0.06	−0.11
guestroom	0.09	0.14	0.04	0.16	guestroom	0.37	0.14	0.13	0.08	0.04	−0.01	−0.12
basement	0.04	0.05	0.05	0.23	0.37	basement	0.05	0.1	0.1	−0.17	0	−0.11
airconditioning	0.11	0.22	0.16	0.12	0.14	0.05	airconditioning	0.19	0.16	0.29	−0.13	−0.15
bathrooms	0.04	0.19	0.18	0.06	0.13	0.1	0.19	bathrooms	0.37	0.33	0.07	−0.14
bedrooms	−0.01	0.15	0.14	0.08	0.08	0.1	0.16	0.37	bedrooms	0.41	0.05	−0.12
stories	0.12	0.08	0.05	0.04	0.04	−0.17	0.29	0.33	0.41	stories	0.02	−0.1
hotwaterheating	−0.01	−0.01	0.07	−0.06	−0.01	0	−0.13	0.07	0.05	0.02	hotwaterheating	−0.03
furnishingstatus	−0.16	−0.17	−0.18	−0.11	−0.12	−0.11	−0.15	−0.14	−0.12	−0.1	−0.03	furnishingstatus

As we can see by the correlation heat-map, correlations between each predictor are realtively small. The strongest correlation is between *stories* and *bedrooms*, with a value of 0.41, swiftly followed by *bedrooms* and *bathrooms* with a coefficient of 0.37, and by *parking* and *area*, with a value of 0.29. These coefficients are decently small, however we will proceed with a more thorough investigation to make sure collinearity is not one of this dataset's problems.

```
##      Eigenvalue Condition Index
## 1      11.4591          1.0000
## 2       0.5846          4.4275
## 3       0.2140          7.3182
## 4       0.1537          8.6333
## 5       0.1229          9.6550
## 6       0.1016         10.6223
## 7       0.0886         11.3721
## 8       0.0778         12.1358
## 9       0.0622         13.5738
## 10      0.0569         14.1939
## 11      0.0403         16.8673
## 12      0.0309         19.2715
## 13      0.0075         38.9969
```

The next step is viewing the dataset's condition numbers, based on a linear model's (that is using the dataset) values.

Eigenvalue	Condition Index
11.4591	1.0000
0.5846	4.4275
0.2140	7.3182
0.1537	8.6333
0.1229	9.6550
0.1016	10.6223
0.0886	11.3721
0.0778	12.1358
0.0622	13.5738
0.0569	14.1939
0.0403	16.8673
0.0309	19.2715
0.0075	38.9969

The condition number is 38.9969, and usually a value above 30 marks the dataset for collinearity. This value of condition number suggests that collinearity should be a problem in this dataset. To further confirm this hypothesis, we shall check each predictor's VIF value.

Variable	VIF
area	1.325208
bedrooms	1.367503
bathrooms	1.286559
stories	1.478029
mainroad	1.172661
guestroom	1.212687
basement	1.320749
hotwaterheating	1.039293
airconditioning	1.207262
parking	1.211959
prefarea	1.148598
furnishingstatus	1.095641

Only predictors with VIFs above 5 are considered problematic, however we can see that, in this dataset, the highest VIF value is 1.478029. We could stop, however we wish to make the investigation very thorough.

We can then check if the variables are orthogonal. An orthogonal variable is not plagued by multicollinearity, and has a R^2_k of less than 0.3. Below is a table of these values for each predictor

Variable	R^2_k
area	0.24540168
bedrooms	0.26874003
bathrooms	0.2273304
stories	0.32342322
mainroad	0.14723849
guestroom	0.17538517
basement	0.24285414
hotwaterheating	0.03780781
airconditioning	0.17167947
parking	0.17488936
prefarea	0.12937340

Variable	R_k^2
furnishingstatus	0.08729247

We can see that most predictors are orthogonal, with the exception of *stories* which has a value of 0.3234.

If, perchance, there was an issue of collinearity, we could check the individual condition numbers. They would have to be above 30. Each value above 30 would signify a singular instance of a problematic linear dependence between predictors. Luckily, there is none here, as you can see from the values below.

```
##    Eigenvalue Condition Index intercept  area bedrooms bathrooms stories
## 1      11.459           1.000     0.000 0.001    0.000     0.001   0.001
## 2       0.585           4.428     0.000 0.003    0.000     0.000   0.001
## 3       0.214           7.318     0.000 0.001    0.004     0.036   0.340
## 4       0.154           8.633     0.001 0.070    0.000     0.008   0.008
## 5       0.123           9.655     0.000 0.467    0.005     0.108   0.001
## 6       0.102          10.622     0.000 0.177    0.004     0.472   0.093
## 7       0.089          11.372     0.000 0.011    0.003     0.006   0.083
## 8       0.078          12.136     0.000 0.096    0.000     0.180   0.123
## 9       0.062          13.574     0.006 0.001    0.046     0.087   0.017
## 10      0.057          14.194     0.004 0.121    0.014     0.030   0.138
## 11      0.040          16.867     0.000 0.000    0.751     0.068   0.147
## 12      0.031          19.271     0.003 0.053    0.067     0.002   0.041
## 13      0.008          38.997     0.986 0.001    0.106     0.001   0.008
##    mainroad guestroom basement hotwaterheating airconditioning parking prefarea
## 1     0.000     0.001    0.001           0.000           0.001   0.002    0.001
## 2     0.000     0.001    0.001           0.000           0.000   0.775    0.001
## 3     0.001     0.012    0.058           0.001           0.013   0.005    0.014
## 4     0.000     0.030    0.046           0.007           0.011   0.093    0.032
## 5     0.004     0.033    0.128           0.001           0.018   0.052    0.029
## 6     0.001     0.020    0.007           0.003           0.156   0.040    0.058
## 7     0.002     0.029    0.000           0.003           0.475   0.005    0.419
## 8     0.002     0.271    0.001           0.009           0.233   0.000    0.265
## 9     0.027     0.363    0.047           0.157           0.007   0.014    0.097
## 10    0.028     0.184    0.586           0.108           0.005   0.000    0.021
## 11    0.091     0.040    0.107           0.011           0.002   0.000    0.012
## 12    0.519     0.001    0.010           0.404           0.020   0.001    0.041
## 13    0.325     0.015    0.009           0.296           0.060   0.014    0.010
##    furnishingstatus
## 1             0.001
## 2             0.007
## 3             0.034
## 4             0.423
## 5             0.014
## 6             0.004
## 7             0.013
## 8             0.006
## 9             0.205
## 10            0.105
## 11            0.000
## 12            0.003
## 13            0.184
```

With all the information provided, we then conclude, with confidence, that multicollinearity is not a problem of this dataset.

Variable Selection

We now move on to selecting variables to use in our final model. We will this using forward, backward, and step-wise selection using AIC, BIC, and RMSE LOOCV to determine which variables are the best for our model

```
n = nrow(data)
mod_all_preds = lm(price~., data=data)
mod_back_aic = step(mod_all_preds, direction='backward')
```

```
## Start:  AIC=15144.37
## price ~ area + bedrooms + bathrooms + stories + mainroad + guestroom +
##     basement + hotwaterheating + airconditioning + parking +
##     prefarea + furnishingstatus
##
##                    Df  Sum of Sq        RSS    AIC
## <none>                             6.0560e+14 15144
## - bedrooms          1  2.8512e+12 6.0845e+14 15145
## - guestroom         1  5.9376e+12 6.1153e+14 15148
## - mainroad          1  1.0006e+13 6.1560e+14 15151
## - basement          1  1.1494e+13 6.1709e+14 15153
## - furnishingstatus  2  1.6379e+13 6.2198e+14 15155
## - hotwaterheating   1  1.6760e+13 6.2236e+14 15157
## - parking           1  2.5568e+13 6.3116e+14 15165
## - prefarea          1  3.6178e+13 6.4178e+14 15174
## - stories           1  5.6299e+13 6.6190e+14 15191
## - airconditioning   1  7.2675e+13 6.7827e+14 15204
## - bathrooms         1  1.0413e+14 7.0973e+14 15229
## - area              1  1.1523e+14 7.2082e+14 15237
```

```
n = nrow(data)
mod_all_preds = lm(price~., data=data)
mod_back_bic = step(mod_all_preds, direction='backward', k=log(n))
```

```
## Start:  AIC=15204.58
## price ~ area + bedrooms + bathrooms + stories + mainroad + guestroom +
##     basement + hotwaterheating + airconditioning + parking +
##     prefarea + furnishingstatus
##
##                    Df  Sum of Sq        RSS    AIC
## - bedrooms          1  2.8512e+12 6.0845e+14 15201
## - guestroom         1  5.9376e+12 6.1153e+14 15204
## <none>                             6.0560e+14 15205
## - furnishingstatus  2  1.6379e+13 6.2198e+14 15206
## - mainroad          1  1.0006e+13 6.1560e+14 15207
## - basement          1  1.1494e+13 6.1709e+14 15208
## - hotwaterheating   1  1.6760e+13 6.2236e+14 15213
## - parking           1  2.5568e+13 6.3116e+14 15221
## - prefarea          1  3.6178e+13 6.4178e+14 15230
## - stories           1  5.6299e+13 6.6190e+14 15247
## - airconditioning   1  7.2675e+13 6.7827e+14 15260
## - bathrooms         1  1.0413e+14 7.0973e+14 15285
## - area              1  1.1523e+14 7.2082e+14 15293
```

```
## 
## Step:  AIC=15200.84
## price ~ area + bathrooms + stories + mainroad + guestroom + basement +
##     hotwaterheating + airconditioning + parking + prefarea +
##     furnishingstatus
## 
##                   Df  Sum of Sq        RSS    AIC
## - guestroom        1  5.7567e+12  6.1421e+14  15200
## <none>                            6.0845e+14  15201
## - mainroad         1  8.8739e+12  6.1732e+14  15202
## - furnishingstatus 2  1.7110e+13  6.2556e+14  15203
## - basement         1  1.3371e+13  6.2182e+14  15206
## - hotwaterheating  1  1.7015e+13  6.2546e+14  15210
## - parking          1  2.6907e+13  6.3536e+14  15218
## - prefarea         1  3.6622e+13  6.4507e+14  15226
## - airconditioning  1  7.2399e+13  6.8085e+14  15256
## - stories          1  7.4793e+13  6.8324e+14  15258
## - bathrooms        1  1.1780e+14  7.2625e+14  15291
## - area             1  1.1885e+14  7.2730e+14  15292
## 
## Step:  AIC=15199.67
## price ~ area + bathrooms + stories + mainroad + basement + hotwaterheating +
##     airconditioning + parking + prefarea + furnishingstatus
## 
##                   Df  Sum of Sq        RSS    AIC
## <none>                            6.1421e+14  15200
## - mainroad         1  9.3488e+12  6.2355e+14  15202
## - furnishingstatus 2  1.7781e+13  6.3199e+14  15203
## - hotwaterheating  1  1.7010e+13  6.3122e+14  15208
## - basement         1  2.2879e+13  6.3708e+14  15213
## - parking          1  2.5746e+13  6.3995e+14  15216
## - prefarea         1  3.7855e+13  6.5206e+14  15226
## - airconditioning  1  7.5704e+13  6.8991e+14  15257
## - stories          1  7.7295e+13  6.9150e+14  15258
## - bathrooms        1  1.2013e+14  7.3433e+14  15291
## - area             1  1.2414e+14  7.3834e+14  15294
```

```r
n = nrow(data)
mod_start = lm(price~1, data=data)
mod_forwd_aic = step(mod_start, scope=price~mainroad+area+parking+prefarea+guestroom+basement+aircondit
```

```
## Start:  AIC=15742.43
## price ~ 1
## 
##                   Df  Sum of Sq        RSS    AIC
## + area             1  5.4678e+14  1.3564e+15  15560
## + bathrooms        1  5.0978e+14  1.3934e+15  15574
## + airconditioning  1  3.9048e+14  1.5127e+15  15619
## + stories          1  3.3687e+14  1.5663e+15  15638
## + parking          1  2.8122e+14  1.6220e+15  15657
## + bedrooms         1  2.5563e+14  1.6476e+15  15666
## + prefarea         1  2.0698e+14  1.6962e+15  15682
## + furnishingstatus 2  1.7980e+14  1.7234e+15  15692
## + mainroad         1  1.6777e+14  1.7354e+15  15694
```

```
## + guestroom          1 1.2426e+14 1.7789e+15 15708
## + basement           1 6.6594e+13 1.8366e+15 15725
## + hotwaterheating    1 1.6487e+13 1.8867e+15 15740
## <none>                          1.9032e+15 15742
##
## Step:  AIC=15559.85
## price ~ area
##
##                      Df  Sum of Sq        RSS    AIC
## + bathrooms          1 3.3838e+14 1.0181e+15 15406
## + stories            1 2.7053e+14 1.0859e+15 15441
## + airconditioning    1 2.2303e+14 1.1334e+15 15464
## + bedrooms           1 1.5835e+14 1.1981e+15 15494
## + furnishingstatus   2 9.2458e+13 1.2640e+15 15525
## + prefarea           1 8.3772e+13 1.2727e+15 15527
## + parking            1 8.2837e+13 1.2736e+15 15528
## + guestroom          1 6.3125e+13 1.2933e+15 15536
## + basement           1 4.9839e+13 1.3066e+15 15541
## + mainroad           1 4.1907e+13 1.3145e+15 15545
## + hotwaterheating    1 1.8287e+13 1.3381e+15 15554
## <none>                          1.3564e+15 15560
##
## Step:  AIC=15405.46
## price ~ area + bathrooms
##
##                      Df  Sum of Sq        RSS    AIC
## + airconditioning    1 1.5149e+14 8.6656e+14 15320
## + stories            1 1.2531e+14 8.9275e+14 15336
## + prefarea           1 7.7575e+13 9.4048e+14 15364
## + furnishingstatus   2 5.6760e+13 9.6129e+14 15378
## + parking            1 4.8508e+13 9.6955e+14 15381
## + mainroad           1 4.5434e+13 9.7262e+14 15383
## + bedrooms           1 4.1866e+13 9.7619e+14 15385
## + guestroom          1 3.7173e+13 9.8088e+14 15387
## + basement           1 2.8511e+13 9.8954e+14 15392
## + hotwaterheating    1 8.9455e+12 1.0091e+15 15403
## <none>                          1.0181e+15 15406
##
## Step:  AIC=15319.65
## price ~ area + bathrooms + airconditioning
##
##                      Df  Sum of Sq        RSS    AIC
## + stories            1 7.0108e+13 7.9645e+14 15276
## + prefarea           1 6.4072e+13 8.0249e+14 15280
## + furnishingstatus   2 4.3440e+13 8.2312e+14 15296
## + mainroad           1 3.8079e+13 8.2848e+14 15297
## + parking            1 3.7158e+13 8.2940e+14 15298
## + bedrooms           1 2.9685e+13 8.3687e+14 15303
## + basement           1 2.5470e+13 8.4109e+14 15305
## + guestroom          1 2.4261e+13 8.4230e+14 15306
## + hotwaterheating    1 2.3146e+13 8.4341e+14 15307
## <none>                          8.6656e+14 15320
##
## Step:  AIC=15275.67
```

```
## price ~ area + bathrooms + airconditioning + stories
##
##                    Df  Sum of Sq         RSS    AIC
## + prefarea          1  6.3539e+13  7.3291e+14  15232
## + basement          1  5.1784e+13  7.4467e+14  15241
## + parking           1  4.1571e+13  7.5488e+14  15248
## + furnishingstatus  2  4.0027e+13  7.5642e+14  15252
## + mainroad          1  2.8378e+13  7.6807e+14  15258
## + guestroom         1  2.6429e+13  7.7002e+14  15259
## + hotwaterheating   1  2.0450e+13  7.7600e+14  15264
## + bedrooms          1  8.7141e+12  7.8774e+14  15272
## <none>                            7.9645e+14  15276
##
## Step:  AIC=15232.36
## price ~ area + bathrooms + airconditioning + stories + prefarea
##
##                    Df  Sum of Sq         RSS    AIC
## + parking           1  4.1282e+13  6.9163e+14  15203
## + furnishingstatus  2  3.5119e+13  6.9779e+14  15210
## + basement          1  3.0449e+13  7.0246e+14  15211
## + hotwaterheating   1  2.4405e+13  7.0851e+14  15216
## + mainroad          1  1.8094e+13  7.1482e+14  15221
## + guestroom         1  1.7369e+13  7.1554e+14  15221
## + bedrooms          1  7.0988e+12  7.2581e+14  15229
## <none>                            7.3291e+14  15232
##
## Step:  AIC=15202.76
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking
##
##                    Df  Sum of Sq         RSS    AIC
## + basement          1  2.9393e+13  6.6224e+14  15181
## + furnishingstatus  2  2.7313e+13  6.6432e+14  15185
## + hotwaterheating   1  1.9636e+13  6.7199e+14  15189
## + guestroom         1  1.9284e+13  6.7235e+14  15189
## + mainroad          1  1.2313e+13  6.7932e+14  15195
## + bedrooms          1  5.0708e+12  6.8656e+14  15201
## <none>                            6.9163e+14  15203
##
## Step:  AIC=15181.09
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement
##
##                    Df  Sum of Sq         RSS    AIC
## + furnishingstatus  2  2.1964e+13  6.4027e+14  15167
## + hotwaterheating   1  1.8635e+13  6.4360e+14  15168
## + mainroad          1  1.1363e+13  6.5087e+14  15174
## + guestroom         1  7.0630e+12  6.5517e+14  15177
## <none>                            6.6224e+14  15181
## + bedrooms          1  2.2753e+12  6.5996e+14  15181
##
## Step:  AIC=15166.71
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus
```

```
##
##                   Df   Sum of Sq        RSS    AIC
## + hotwaterheating  1  1.6719e+13  6.2355e+14  15154
## + mainroad         1  9.0579e+12  6.3122e+14  15161
## + guestroom        1  6.2197e+12  6.3405e+14  15163
## <none>                            6.4027e+14  15167
## + bedrooms         1  1.7516e+12  6.3852e+14  15167
##
## Step:  AIC=15154.29
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##       parking + basement + furnishingstatus + hotwaterheating
##
##                Df   Sum of Sq        RSS    AIC
## + mainroad     1  9.3488e+12  6.1421e+14  15148
## + guestroom    1  6.2316e+12  6.1732e+14  15151
## <none>                        6.2355e+14  15154
## + bedrooms     1  1.5499e+12  6.2200e+14  15155
##
## Step:  AIC=15148.06
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##       parking + basement + furnishingstatus + hotwaterheating +
##       mainroad
##
##                Df   Sum of Sq        RSS    AIC
## + guestroom    1  5.7567e+12  6.0845e+14  15145
## + bedrooms     1  2.6702e+12  6.1153e+14  15148
## <none>                        6.1421e+14  15148
##
## Step:  AIC=15144.93
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##       parking + basement + furnishingstatus + hotwaterheating +
##       mainroad + guestroom
##
##                Df   Sum of Sq        RSS    AIC
## + bedrooms     1  2.8512e+12  6.0560e+14  15144
## <none>                        6.0845e+14  15145
##
## Step:  AIC=15144.37
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##       parking + basement + furnishingstatus + hotwaterheating +
##       mainroad + guestroom + bedrooms
```

```r
n = nrow(data)
mod_start = lm(price~1, data=data)
mod_forwd_bic = step(mod_start, scope=price~mainroad+area+parking+prefarea+guestroom+basement+aircondit
```

```
## Start:  AIC=15746.74
## price ~ 1
##
##                     Df   Sum of Sq        RSS    AIC
## + area               1  5.4678e+14  1.3564e+15  15568
## + bathrooms          1  5.0978e+14  1.3934e+15  15583
## + airconditioning    1  3.9048e+14  1.5127e+15  15628
## + stories            1  3.3687e+14  1.5663e+15  15647
```

11

```
## + parking              1 2.8122e+14 1.6220e+15 15666
## + bedrooms             1 2.5563e+14 1.6476e+15 15674
## + prefarea             1 2.0698e+14 1.6962e+15 15690
## + mainroad             1 1.6777e+14 1.7354e+15 15703
## + furnishingstatus     2 1.7980e+14 1.7234e+15 15705
## + guestroom            1 1.2426e+14 1.7789e+15 15716
## + basement             1 6.6594e+13 1.8366e+15 15734
## <none>                            1.9032e+15 15747
## + hotwaterheating      1 1.6487e+13 1.8867e+15 15748
##
## Step:  AIC=15568.45
## price ~ area
##
##                       Df  Sum of Sq       RSS    AIC
## + bathrooms           1 3.3838e+14 1.0181e+15 15418
## + stories             1 2.7053e+14 1.0859e+15 15454
## + airconditioning     1 2.2303e+14 1.1334e+15 15477
## + bedrooms            1 1.5835e+14 1.1981e+15 15507
## + prefarea            1 8.3772e+13 1.2727e+15 15540
## + parking             1 8.2837e+13 1.2736e+15 15540
## + furnishingstatus    2 9.2458e+13 1.2640e+15 15543
## + guestroom           1 6.3125e+13 1.2933e+15 15549
## + basement            1 4.9839e+13 1.3066e+15 15554
## + mainroad            1 4.1907e+13 1.3145e+15 15558
## + hotwaterheating     1 1.8287e+13 1.3381e+15 15567
## <none>                           1.3564e+15 15568
##
## Step:  AIC=15418.36
## price ~ area + bathrooms
##
##                       Df  Sum of Sq       RSS    AIC
## + airconditioning     1 1.5149e+14 8.6656e+14 15337
## + stories             1 1.2531e+14 8.9275e+14 15353
## + prefarea            1 7.7575e+13 9.4048e+14 15382
## + parking             1 4.8508e+13 9.6955e+14 15398
## + furnishingstatus    2 5.6760e+13 9.6129e+14 15400
## + mainroad            1 4.5434e+13 9.7262e+14 15400
## + bedrooms            1 4.1866e+13 9.7619e+14 15402
## + guestroom           1 3.7173e+13 9.8088e+14 15404
## + basement            1 2.8511e+13 9.8954e+14 15409
## <none>                           1.0181e+15 15418
## + hotwaterheating     1 8.9455e+12 1.0091e+15 15420
##
## Step:  AIC=15336.85
## price ~ area + bathrooms + airconditioning
##
##                       Df  Sum of Sq       RSS    AIC
## + stories             1 7.0108e+13 7.9645e+14 15297
## + prefarea            1 6.4072e+13 8.0249e+14 15301
## + mainroad            1 3.8079e+13 8.2848e+14 15319
## + parking             1 3.7158e+13 8.2940e+14 15319
## + furnishingstatus    2 4.3440e+13 8.2312e+14 15321
## + bedrooms            1 2.9685e+13 8.3687e+14 15324
## + basement            1 2.5470e+13 8.4109e+14 15327
```

```
## + guestroom       1 2.4261e+13 8.4230e+14 15328
## + hotwaterheating 1 2.3146e+13 8.4341e+14 15328
## <none>                       8.6656e+14 15337
##
## Step:  AIC=15297.17
## price ~ area + bathrooms + airconditioning + stories
##
##                   Df  Sum of Sq        RSS    AIC
## + prefarea         1 6.3539e+13 7.3291e+14 15258
## + basement         1 5.1784e+13 7.4467e+14 15267
## + parking          1 4.1571e+13 7.5488e+14 15274
## + furnishingstatus 2 4.0027e+13 7.5642e+14 15282
## + mainroad         1 2.8378e+13 7.6807e+14 15284
## + guestroom        1 2.6429e+13 7.7002e+14 15285
## + hotwaterheating  1 2.0450e+13 7.7600e+14 15289
## <none>                        7.9645e+14 15297
## + bedrooms         1 8.7141e+12 7.8774e+14 15298
##
## Step:  AIC=15258.16
## price ~ area + bathrooms + airconditioning + stories + prefarea
##
##                   Df  Sum of Sq        RSS    AIC
## + parking          1 4.1282e+13 6.9163e+14 15233
## + basement         1 3.0449e+13 7.0246e+14 15241
## + furnishingstatus 2 3.5119e+13 6.9779e+14 15244
## + hotwaterheating  1 2.4405e+13 7.0851e+14 15246
## + mainroad         1 1.8094e+13 7.1482e+14 15251
## + guestroom        1 1.7369e+13 7.1554e+14 15251
## <none>                        7.3291e+14 15258
## + bedrooms         1 7.0988e+12 7.2581e+14 15259
##
## Step:  AIC=15232.87
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking
##
##                   Df  Sum of Sq        RSS    AIC
## + basement         1 2.9393e+13 6.6224e+14 15216
## + hotwaterheating  1 1.9636e+13 6.7199e+14 15224
## + furnishingstatus 2 2.7313e+13 6.6432e+14 15224
## + guestroom        1 1.9284e+13 6.7235e+14 15224
## + mainroad         1 1.2313e+13 6.7932e+14 15229
## <none>                        6.9163e+14 15233
## + bedrooms         1 5.0708e+12 6.8656e+14 15235
##
## Step:  AIC=15215.5
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement
##
##                   Df  Sum of Sq        RSS    AIC
## + hotwaterheating  1 1.8635e+13 6.4360e+14 15206
## + furnishingstatus 2 2.1964e+13 6.4027e+14 15210
## + mainroad         1 1.1363e+13 6.5087e+14 15212
## <none>                        6.6224e+14 15216
## + guestroom        1 7.0630e+12 6.5517e+14 15216
```

```
## + bedrooms          1 2.2753e+12 6.5996e+14 15220
##
## Step:  AIC=15206.24
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating
##
##                   Df  Sum of Sq        RSS    AIC
## + furnishingstatus  2 2.0048e+13 6.2355e+14 15202
## + mainroad          1 1.1616e+13 6.3199e+14 15203
## <none>                           6.4360e+14 15206
## + guestroom         1 7.0550e+12 6.3655e+14 15206
## + bedrooms          1 2.0006e+12 6.4160e+14 15211
##
## Step:  AIC=15201.6
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating + furnishingstatus
##
##              Df  Sum of Sq        RSS    AIC
## + mainroad    1 9.3488e+12 6.1421e+14 15200
## <none>                     6.2355e+14 15202
## + guestroom   1 6.2316e+12 6.1732e+14 15202
## + bedrooms    1 1.5499e+12 6.2200e+14 15206
##
## Step:  AIC=15199.67
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating + furnishingstatus +
##     mainroad
##
##              Df  Sum of Sq        RSS    AIC
## <none>                     6.1421e+14 15200
## + guestroom   1 5.7567e+12 6.0845e+14 15201
## + bedrooms    1 2.6702e+12 6.1153e+14 15204
```

```r
n = nrow(data)
mod_start = lm(price~1, data=data)
mod_step_aic = step(mod_start, scope=price~mainroad+area+parking+prefarea+guestroom+basement+airconditi
```

```
## Start:  AIC=15742.43
## price ~ 1
##
##                   Df  Sum of Sq        RSS    AIC
## + area             1 5.4678e+14 1.3564e+15 15560
## + bathrooms        1 5.0978e+14 1.3934e+15 15574
## + airconditioning  1 3.9048e+14 1.5127e+15 15619
## + stories          1 3.3687e+14 1.5663e+15 15638
## + parking          1 2.8122e+14 1.6220e+15 15657
## + bedrooms         1 2.5563e+14 1.6476e+15 15666
## + prefarea         1 2.0698e+14 1.6962e+15 15682
## + furnishingstatus 2 1.7980e+14 1.7234e+15 15692
## + mainroad         1 1.6777e+14 1.7354e+15 15694
## + guestroom        1 1.2426e+14 1.7789e+15 15708
## + basement         1 6.6594e+13 1.8366e+15 15725
## + hotwaterheating  1 1.6487e+13 1.8867e+15 15740
## <none>                          1.9032e+15 15742
```

14

```
##
## Step:  AIC=15559.85
## price ~ area
##
##                    Df  Sum of Sq        RSS    AIC
## + bathrooms         1 3.3838e+14 1.0181e+15 15406
## + stories           1 2.7053e+14 1.0859e+15 15441
## + airconditioning   1 2.2303e+14 1.1334e+15 15464
## + bedrooms          1 1.5835e+14 1.1981e+15 15494
## + furnishingstatus  2 9.2458e+13 1.2640e+15 15525
## + prefarea          1 8.3772e+13 1.2727e+15 15527
## + parking           1 8.2837e+13 1.2736e+15 15528
## + guestroom         1 6.3125e+13 1.2933e+15 15536
## + basement          1 4.9839e+13 1.3066e+15 15541
## + mainroad          1 4.1907e+13 1.3145e+15 15545
## + hotwaterheating   1 1.8287e+13 1.3381e+15 15554
## <none>                         1.3564e+15 15560
## - area              1 5.4678e+14 1.9032e+15 15742
##
## Step:  AIC=15405.46
## price ~ area + bathrooms
##
##                    Df  Sum of Sq        RSS    AIC
## + airconditioning   1 1.5149e+14 8.6656e+14 15320
## + stories           1 1.2531e+14 8.9275e+14 15336
## + prefarea          1 7.7575e+13 9.4048e+14 15364
## + furnishingstatus  2 5.6760e+13 9.6129e+14 15378
## + parking           1 4.8508e+13 9.6955e+14 15381
## + mainroad          1 4.5434e+13 9.7262e+14 15383
## + bedrooms          1 4.1866e+13 9.7619e+14 15385
## + guestroom         1 3.7173e+13 9.8088e+14 15387
## + basement          1 2.8511e+13 9.8954e+14 15392
## + hotwaterheating   1 8.9455e+12 1.0091e+15 15403
## <none>                         1.0181e+15 15406
## - bathrooms         1 3.3838e+14 1.3564e+15 15560
## - area              1 3.7537e+14 1.3934e+15 15574
##
## Step:  AIC=15319.65
## price ~ area + bathrooms + airconditioning
##
##                    Df  Sum of Sq        RSS    AIC
## + stories           1 7.0108e+13 7.9645e+14 15276
## + prefarea          1 6.4072e+13 8.0249e+14 15280
## + furnishingstatus  2 4.3440e+13 8.2312e+14 15296
## + mainroad          1 3.8079e+13 8.2848e+14 15297
## + parking           1 3.7158e+13 8.2940e+14 15298
## + bedrooms          1 2.9685e+13 8.3687e+14 15303
## + basement          1 2.5470e+13 8.4109e+14 15305
## + guestroom         1 2.4261e+13 8.4230e+14 15306
## + hotwaterheating   1 2.3146e+13 8.4341e+14 15307
## <none>                         8.6656e+14 15320
## - airconditioning   1 1.5149e+14 1.0181e+15 15406
## - bathrooms         1 2.6684e+14 1.1334e+15 15464
## - area              1 2.7663e+14 1.1432e+15 15469
```

```
##
## Step:  AIC=15275.67
## price ~ area + bathrooms + airconditioning + stories
##
##                    Df  Sum of Sq        RSS    AIC
## + prefarea          1  6.3539e+13 7.3291e+14 15232
## + basement          1  5.1784e+13 7.4467e+14 15241
## + parking           1  4.1571e+13 7.5488e+14 15248
## + furnishingstatus  2  4.0027e+13 7.5642e+14 15252
## + mainroad          1  2.8378e+13 7.6807e+14 15258
## + guestroom         1  2.6429e+13 7.7002e+14 15259
## + hotwaterheating   1  2.0450e+13 7.7600e+14 15264
## + bedrooms          1  8.7141e+12 7.8774e+14 15272
## <none>                            7.9645e+14 15276
## - stories           1  7.0108e+13 8.6656e+14 15320
## - airconditioning   1  9.6295e+13 8.9275e+14 15336
## - bathrooms         1  1.7457e+14 9.7102e+14 15382
## - area              1  2.8409e+14 1.0805e+15 15440
##
## Step:  AIC=15232.36
## price ~ area + bathrooms + airconditioning + stories + prefarea
##
##                    Df  Sum of Sq        RSS    AIC
## + parking           1  4.1282e+13 6.9163e+14 15203
## + furnishingstatus  2  3.5119e+13 6.9779e+14 15210
## + basement          1  3.0449e+13 7.0246e+14 15211
## + hotwaterheating   1  2.4405e+13 7.0851e+14 15216
## + mainroad          1  1.8094e+13 7.1482e+14 15221
## + guestroom         1  1.7369e+13 7.1554e+14 15221
## + bedrooms          1  7.0988e+12 7.2581e+14 15229
## <none>                            7.3291e+14 15232
## - prefarea          1  6.3539e+13 7.9645e+14 15276
## - stories           1  6.9575e+13 8.0249e+14 15280
## - airconditioning   1  8.6236e+13 8.1915e+14 15291
## - bathrooms         1  1.7307e+14 9.0599e+14 15346
## - area              1  2.1862e+14 9.5153e+14 15373
##
## Step:  AIC=15202.76
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking
##
##                    Df  Sum of Sq        RSS    AIC
## + basement          1  2.9393e+13 6.6224e+14 15181
## + furnishingstatus  2  2.7313e+13 6.6432e+14 15185
## + hotwaterheating   1  1.9636e+13 6.7199e+14 15189
## + guestroom         1  1.9284e+13 6.7235e+14 15189
## + mainroad          1  1.2313e+13 6.7932e+14 15195
## + bedrooms          1  5.0708e+12 6.8656e+14 15201
## <none>                            6.9163e+14 15203
## - parking           1  4.1282e+13 7.3291e+14 15232
## - prefarea          1  6.3249e+13 7.5488e+14 15248
## - stories           1  7.3958e+13 7.6559e+14 15256
## - airconditioning   1  7.6479e+13 7.6811e+14 15258
## - area              1  1.4678e+14 8.3841e+14 15306
```

```
## - bathrooms         1 1.5212e+14 8.4375e+14 15309
##
## Step:  AIC=15181.09
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement
##
##                       Df  Sum of Sq        RSS   AIC
## + furnishingstatus     2 2.1964e+13 6.4027e+14 15167
## + hotwaterheating      1 1.8635e+13 6.4360e+14 15168
## + mainroad             1 1.1363e+13 6.5087e+14 15174
## + guestroom            1 7.0630e+12 6.5517e+14 15177
## <none>                            6.6224e+14 15181
## + bedrooms             1 2.2753e+12 6.5996e+14 15181
## - basement             1 2.9393e+13 6.9163e+14 15203
## - parking              1 4.0225e+13 7.0246e+14 15211
## - prefarea             1 4.2263e+13 7.0450e+14 15213
## - airconditioning      1 6.9676e+13 7.3191e+14 15234
## - stories              1 9.3063e+13 7.5530e+14 15251
## - bathrooms            1 1.2845e+14 7.9069e+14 15276
## - area                 1 1.5129e+14 8.1353e+14 15291
##
## Step:  AIC=15166.71
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus
##
##                       Df  Sum of Sq        RSS   AIC
## + hotwaterheating      1 1.6719e+13 6.2355e+14 15154
## + mainroad             1 9.0579e+12 6.3122e+14 15161
## + guestroom            1 6.2197e+12 6.3405e+14 15163
## <none>                            6.4027e+14 15167
## + bedrooms             1 1.7516e+12 6.3852e+14 15167
## - furnishingstatus     2 2.1964e+13 6.6224e+14 15181
## - basement             1 2.4044e+13 6.6432e+14 15185
## - parking              1 3.3295e+13 6.7357e+14 15192
## - prefarea             1 4.0683e+13 6.8096e+14 15198
## - airconditioning      1 6.6456e+13 7.0673e+14 15218
## - stories              1 8.7396e+13 7.2767e+14 15234
## - bathrooms            1 1.2188e+14 7.6215e+14 15260
## - area                 1 1.4264e+14 7.8292e+14 15274
##
## Step:  AIC=15154.29
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus + hotwaterheating
##
##                       Df  Sum of Sq        RSS   AIC
## + mainroad             1 9.3488e+12 6.1421e+14 15148
## + guestroom            1 6.2316e+12 6.1732e+14 15151
## <none>                            6.2355e+14 15154
## + bedrooms             1 1.5499e+12 6.2200e+14 15155
## - hotwaterheating      1 1.6719e+13 6.4027e+14 15167
## - furnishingstatus     2 2.0048e+13 6.4360e+14 15168
## - basement             1 2.3403e+13 6.4696e+14 15172
## - parking              1 2.9641e+13 6.5320e+14 15178
## - prefarea             1 4.3507e+13 6.6706e+14 15189
```

```
## - airconditioning    1 7.5252e+13 6.9881e+14 15214
## - stories            1 8.4230e+13 7.0778e+14 15221
## - bathrooms          1 1.1587e+14 7.3943e+14 15245
## - area               1 1.4362e+14 7.6717e+14 15265
##
## Step:  AIC=15148.06
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus + hotwaterheating +
##     mainroad
##
##                      Df  Sum of Sq        RSS    AIC
## + guestroom          1  5.7567e+12 6.0845e+14 15145
## + bedrooms           1  2.6702e+12 6.1153e+14 15148
## <none>                              6.1421e+14 15148
## - mainroad           1  9.3488e+12 6.2355e+14 15154
## - furnishingstatus   2  1.7781e+13 6.3199e+14 15160
## - hotwaterheating    1  1.7010e+13 6.3122e+14 15161
## - basement           1  2.2879e+13 6.3708e+14 15166
## - parking            1  2.5746e+13 6.3995e+14 15168
## - prefarea           1  3.7855e+13 6.5206e+14 15179
## - airconditioning    1  7.5704e+13 6.8991e+14 15209
## - stories            1  7.7295e+13 6.9150e+14 15211
## - bathrooms          1  1.2013e+14 7.3433e+14 15243
## - area               1  1.2414e+14 7.3834e+14 15246
##
## Step:  AIC=15144.93
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus + hotwaterheating +
##     mainroad + guestroom
##
##                      Df  Sum of Sq        RSS    AIC
## + bedrooms           1  2.8512e+12 6.0560e+14 15144
## <none>                              6.0845e+14 15145
## - guestroom          1  5.7567e+12 6.1421e+14 15148
## - mainroad           1  8.8739e+12 6.1732e+14 15151
## - basement           1  1.3371e+13 6.2182e+14 15155
## - furnishingstatus   2  1.7110e+13 6.2556e+14 15156
## - hotwaterheating    1  1.7015e+13 6.2546e+14 15158
## - parking            1  2.6907e+13 6.3536e+14 15166
## - prefarea           1  3.6622e+13 6.4507e+14 15175
## - airconditioning    1  7.2399e+13 6.8085e+14 15204
## - stories            1  7.4793e+13 6.8324e+14 15206
## - bathrooms          1  1.1780e+14 7.2625e+14 15239
## - area               1  1.1885e+14 7.2730e+14 15240
##
## Step:  AIC=15144.37
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + furnishingstatus + hotwaterheating +
##     mainroad + guestroom + bedrooms
##
##                      Df  Sum of Sq        RSS    AIC
## <none>                              6.0560e+14 15144
## - bedrooms           1  2.8512e+12 6.0845e+14 15145
## - guestroom          1  5.9376e+12 6.1153e+14 15148
```

```
## - mainroad            1 1.0006e+13 6.1560e+14 15151
## - basement            1 1.1494e+13 6.1709e+14 15153
## - furnishingstatus    2 1.6379e+13 6.2198e+14 15155
## - hotwaterheating     1 1.6760e+13 6.2236e+14 15157
## - parking             1 2.5568e+13 6.3116e+14 15165
## - prefarea            1 3.6178e+13 6.4178e+14 15174
## - stories             1 5.6299e+13 6.6190e+14 15191
## - airconditioning     1 7.2675e+13 6.7827e+14 15204
## - bathrooms           1 1.0413e+14 7.0973e+14 15229
## - area                1 1.1523e+14 7.2082e+14 15237
```

```r
n = nrow(data)
mod_start = lm(price~1, data=data)
mod_step_bic = step(mod_start, scope=price~mainroad+area+parking+prefarea+guestroom+basement+airconditi
```

```
## Start:  AIC=15746.74
## price ~ 1
##
##                     Df  Sum of Sq        RSS   AIC
## + area               1 5.4678e+14 1.3564e+15 15568
## + bathrooms          1 5.0978e+14 1.3934e+15 15583
## + airconditioning    1 3.9048e+14 1.5127e+15 15628
## + stories            1 3.3687e+14 1.5663e+15 15647
## + parking            1 2.8122e+14 1.6220e+15 15666
## + bedrooms           1 2.5563e+14 1.6476e+15 15674
## + prefarea           1 2.0698e+14 1.6962e+15 15690
## + mainroad           1 1.6777e+14 1.7354e+15 15703
## + furnishingstatus   2 1.7980e+14 1.7234e+15 15705
## + guestroom          1 1.2426e+14 1.7789e+15 15716
## + basement           1 6.6594e+13 1.8366e+15 15734
## <none>                          1.9032e+15 15747
## + hotwaterheating    1 1.6487e+13 1.8867e+15 15748
##
## Step:  AIC=15568.45
## price ~ area
##
##                     Df  Sum of Sq        RSS   AIC
## + bathrooms          1 3.3838e+14 1.0181e+15 15418
## + stories            1 2.7053e+14 1.0859e+15 15454
## + airconditioning    1 2.2303e+14 1.1334e+15 15477
## + bedrooms           1 1.5835e+14 1.1981e+15 15507
## + prefarea           1 8.3772e+13 1.2727e+15 15540
## + parking            1 8.2837e+13 1.2736e+15 15540
## + furnishingstatus   2 9.2458e+13 1.2640e+15 15543
## + guestroom          1 6.3125e+13 1.2933e+15 15549
## + basement           1 4.9839e+13 1.3066e+15 15554
## + mainroad           1 4.1907e+13 1.3145e+15 15558
## + hotwaterheating    1 1.8287e+13 1.3381e+15 15567
## <none>                          1.3564e+15 15568
## - area               1 5.4678e+14 1.9032e+15 15747
##
## Step:  AIC=15418.36
## price ~ area + bathrooms
##
```

19

```
##                         Df  Sum of Sq        RSS    AIC
## + airconditioning   1 1.5149e+14 8.6656e+14 15337
## + stories           1 1.2531e+14 8.9275e+14 15353
## + prefarea          1 7.7575e+13 9.4048e+14 15382
## + parking           1 4.8508e+13 9.6955e+14 15398
## + furnishingstatus  2 5.6760e+13 9.6129e+14 15400
## + mainroad          1 4.5434e+13 9.7262e+14 15400
## + bedrooms          1 4.1866e+13 9.7619e+14 15402
## + guestroom         1 3.7173e+13 9.8088e+14 15404
## + basement          1 2.8511e+13 9.8954e+14 15409
## <none>                             1.0181e+15 15418
## + hotwaterheating   1 8.9455e+12 1.0091e+15 15420
## - bathrooms         1 3.3838e+14 1.3564e+15 15568
## - area              1 3.7537e+14 1.3934e+15 15583
##
## Step:  AIC=15336.85
## price ~ area + bathrooms + airconditioning
##
##                         Df  Sum of Sq        RSS    AIC
## + stories           1 7.0108e+13 7.9645e+14 15297
## + prefarea          1 6.4072e+13 8.0249e+14 15301
## + mainroad          1 3.8079e+13 8.2848e+14 15319
## + parking           1 3.7158e+13 8.2940e+14 15319
## + furnishingstatus  2 4.3440e+13 8.2312e+14 15321
## + bedrooms          1 2.9685e+13 8.3687e+14 15324
## + basement          1 2.5470e+13 8.4109e+14 15327
## + guestroom         1 2.4261e+13 8.4230e+14 15328
## + hotwaterheating   1 2.3146e+13 8.4341e+14 15328
## <none>                             8.6656e+14 15337
## - airconditioning   1 1.5149e+14 1.0181e+15 15418
## - bathrooms         1 2.6684e+14 1.1334e+15 15477
## - area              1 2.7663e+14 1.1432e+15 15482
##
## Step:  AIC=15297.17
## price ~ area + bathrooms + airconditioning + stories
##
##                         Df  Sum of Sq        RSS    AIC
## + prefarea          1 6.3539e+13 7.3291e+14 15258
## + basement          1 5.1784e+13 7.4467e+14 15267
## + parking           1 4.1571e+13 7.5488e+14 15274
## + furnishingstatus  2 4.0027e+13 7.5642e+14 15282
## + mainroad          1 2.8378e+13 7.6807e+14 15284
## + guestroom         1 2.6429e+13 7.7002e+14 15285
## + hotwaterheating   1 2.0450e+13 7.7600e+14 15289
## <none>                             7.9645e+14 15297
## + bedrooms          1 8.7141e+12 7.8774e+14 15298
## - stories           1 7.0108e+13 8.6656e+14 15337
## - airconditioning   1 9.6295e+13 8.9275e+14 15353
## - bathrooms         1 1.7457e+14 9.7102e+14 15399
## - area              1 2.8409e+14 1.0805e+15 15457
##
## Step:  AIC=15258.16
## price ~ area + bathrooms + airconditioning + stories + prefarea
##
```

```
##                       Df  Sum of Sq        RSS    AIC
## + parking            1  4.1282e+13  6.9163e+14  15233
## + basement           1  3.0449e+13  7.0246e+14  15241
## + furnishingstatus   2  3.5119e+13  6.9779e+14  15244
## + hotwaterheating    1  2.4405e+13  7.0851e+14  15246
## + mainroad           1  1.8094e+13  7.1482e+14  15251
## + guestroom          1  1.7369e+13  7.1554e+14  15251
## <none>                          7.3291e+14  15258
## + bedrooms           1  7.0988e+12  7.2581e+14  15259
## - prefarea           1  6.3539e+13  7.9645e+14  15297
## - stories            1  6.9575e+13  8.0249e+14  15301
## - airconditioning    1  8.6236e+13  8.1915e+14  15312
## - bathrooms          1  1.7307e+14  9.0599e+14  15367
## - area               1  2.1862e+14  9.5153e+14  15394
##
## Step:  AIC=15232.87
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking
##
##                       Df  Sum of Sq        RSS    AIC
## + basement           1  2.9393e+13  6.6224e+14  15216
## + hotwaterheating    1  1.9636e+13  6.7199e+14  15224
## + furnishingstatus   2  2.7313e+13  6.6432e+14  15224
## + guestroom          1  1.9284e+13  6.7235e+14  15224
## + mainroad           1  1.2313e+13  6.7932e+14  15229
## <none>                          6.9163e+14  15233
## + bedrooms           1  5.0708e+12  6.8656e+14  15235
## - parking            1  4.1282e+13  7.3291e+14  15258
## - prefarea           1  6.3249e+13  7.5488e+14  15274
## - stories            1  7.3958e+13  7.6559e+14  15282
## - airconditioning    1  7.6479e+13  7.6811e+14  15284
## - area               1  1.4678e+14  8.3841e+14  15332
## - bathrooms          1  1.5212e+14  8.4375e+14  15335
##
## Step:  AIC=15215.5
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement
##
##                       Df  Sum of Sq        RSS    AIC
## + hotwaterheating    1  1.8635e+13  6.4360e+14  15206
## + furnishingstatus   2  2.1964e+13  6.4027e+14  15210
## + mainroad           1  1.1363e+13  6.5087e+14  15212
## <none>                          6.6224e+14  15216
## + guestroom          1  7.0630e+12  6.5517e+14  15216
## + bedrooms           1  2.2753e+12  6.5996e+14  15220
## - basement           1  2.9393e+13  6.9163e+14  15233
## - parking            1  4.0225e+13  7.0246e+14  15241
## - prefarea           1  4.2263e+13  7.0450e+14  15243
## - airconditioning    1  6.9676e+13  7.3191e+14  15264
## - stories            1  9.3063e+13  7.5530e+14  15281
## - bathrooms          1  1.2845e+14  7.9069e+14  15306
## - area               1  1.5129e+14  8.1353e+14  15321
##
## Step:  AIC=15206.24
```

```
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating
##
##                    Df  Sum of Sq        RSS    AIC
## + furnishingstatus  2 2.0048e+13 6.2355e+14 15202
## + mainroad          1 1.1616e+13 6.3199e+14 15203
## <none>                          6.4360e+14 15206
## + guestroom         1 7.0550e+12 6.3655e+14 15206
## + bedrooms          1 2.0006e+12 6.4160e+14 15211
## - hotwaterheating   1 1.8635e+13 6.6224e+14 15216
## - basement          1 2.8392e+13 6.7199e+14 15224
## - parking           1 3.5651e+13 6.7925e+14 15229
## - prefarea          1 4.5300e+13 6.8890e+14 15237
## - airconditioning   1 7.9324e+13 7.2293e+14 15263
## - stories           1 8.9391e+13 7.3299e+14 15271
## - bathrooms         1 1.2166e+14 7.6526e+14 15294
## - area              1 1.5210e+14 7.9571e+14 15316
##
## Step:  AIC=15201.6
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating + furnishingstatus
##
##                    Df  Sum of Sq        RSS    AIC
## + mainroad          1 9.3488e+12 6.1421e+14 15200
## <none>                          6.2355e+14 15202
## + guestroom         1 6.2316e+12 6.1732e+14 15202
## - furnishingstatus  2 2.0048e+13 6.4360e+14 15206
## + bedrooms          1 1.5499e+12 6.2200e+14 15206
## - hotwaterheating   1 1.6719e+13 6.4027e+14 15210
## - basement          1 2.3403e+13 6.4696e+14 15215
## - parking           1 2.9641e+13 6.5320e+14 15221
## - prefarea          1 4.3507e+13 6.6706e+14 15232
## - airconditioning   1 7.5252e+13 6.9881e+14 15257
## - stories           1 8.4230e+13 7.0778e+14 15264
## - bathrooms         1 1.1587e+14 7.3943e+14 15288
## - area              1 1.4362e+14 7.6717e+14 15308
##
## Step:  AIC=15199.67
## price ~ area + bathrooms + airconditioning + stories + prefarea +
##     parking + basement + hotwaterheating + furnishingstatus +
##     mainroad
##
##                    Df  Sum of Sq        RSS    AIC
## <none>                          6.1421e+14 15200
## + guestroom         1 5.7567e+12 6.0845e+14 15201
## - mainroad          1 9.3488e+12 6.2355e+14 15202
## - furnishingstatus  2 1.7781e+13 6.3199e+14 15203
## + bedrooms          1 2.6702e+12 6.1153e+14 15204
## - hotwaterheating   1 1.7010e+13 6.3122e+14 15208
## - basement          1 2.2879e+13 6.3708e+14 15213
## - parking           1 2.5746e+13 6.3995e+14 15216
## - prefarea          1 3.7855e+13 6.5206e+14 15226
## - airconditioning   1 7.5704e+13 6.8991e+14 15257
## - stories           1 7.7295e+13 6.9150e+14 15258
```

```
## - bathrooms          1 1.2013e+14 7.3433e+14 15291
## - area               1 1.2414e+14 7.3834e+14 15294
```

```
calc_loocv_rmse = function(model) {
  sqrt(mean((resid(model) / (1 - hatvalues(model))) ^ 2))
}
calc_loocv_rmse(mod_back_aic)
```

```
## [1] 1087331
```

```
calc_loocv_rmse(mod_back_bic)
```

```
## [1] 1090310
```

```
calc_loocv_rmse(mod_forwd_aic)
```

```
## [1] 1087331
```

```
calc_loocv_rmse(mod_forwd_bic)
```

```
## [1] 1090310
```

```
calc_loocv_rmse(mod_step_aic)
```

```
## [1] 1087331
```

```
calc_loocv_rmse(mod_step_bic)
```

```
## [1] 1090310
```

```
library(leaps)
```

```
mod_exhaustive = summary(regsubsets(price ~ ., data = data, nvmax = 8))
```

```
best_r2_ind = which.max(mod_exhaustive$adjr2)
```

```
mod_exhaustive$which[best_r2_ind,]
```

```
##                    (Intercept)                        area
##                           TRUE                        TRUE
##                       bedrooms                   bathrooms
##                          FALSE                        TRUE
##                        stories                  mainroadyes
##                           TRUE                       FALSE
##                    guestroomyes                 basementyes
##                          FALSE                        TRUE
##             hotwaterheatingyes           airconditioningyes
##                          FALSE                        TRUE
##                        parking                  prefareayes
##                           TRUE                        TRUE
## furnishingstatussemi-furnished    furnishingstatusunfurnished
##                          FALSE                        TRUE
```

```
mod_exhaust_adjr2 <- lm(price~area+bathrooms+stories+basement+airconditioning+parking+prefarea+furnishin
summary(mod_exhaust_adjr2)
```

```
##
## Call:
## lm(formula = price ~ area + bathrooms + stories + basement +
##     airconditioning + parking + prefarea + furnishingstatus,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2832572  -684438   -51716   500702  5066685
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     510518.70  206990.03   2.466 0.013961 *
## area                               264.49      24.23  10.917  < 2e-16 ***
## bathrooms                      1035964.52  102657.48  10.091  < 2e-16 ***
## stories                         521603.49   61037.95   8.546  < 2e-16 ***
## basementyes                     470422.82  104952.53   4.482 9.04e-06 ***
## airconditioningyes              815568.08  109445.37   7.452 3.72e-13 ***
## parking                         312237.72   59197.50   5.275 1.94e-07 ***
## prefareayes                     683529.82  117234.62   5.830 9.57e-09 ***
## furnishingstatussemi-furnished  -58125.85  119191.87  -0.488 0.625986
## furnishingstatusunfurnished    -475893.22  128543.13  -3.702 0.000236 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1094000 on 535 degrees of freedom
## Multiple R-squared:  0.6636, Adjusted R-squared:  0.6579
## F-statistic: 117.3 on 9 and 535 DF,  p-value: < 2.2e-16
```

Model Diagnostics