ACTL4305 Assignment 1 - z5194905

Contents

Task 1: The iterative cycle of EDA

Exploratory data analysis (EDA) is a procedure to systematically produce descriptive summaries of a dataset by investigating any data problems and understand the relationships between its predictors. EDA includes importing and cleaning data as well as preparing for future modelling or predictions. It provides a summary of the data through descriptive statistics and visualisations and should reveal data flaws such as missing values.

The EDA procedure is a 3 step process.

1. **Develop a question**

- Questions are at the heart of data analytics and they direct data exploration. Some examples of good exploratory questions are the following-
  - What does my data look like?
  - Are some of my variables correlated?

2. **Find answers**

- Import data
  - To begin exploring, the data is needed in a workable environment, such as R or python. Merging of required datasets should also be done at this stage.

- Check data quality
  - check data quality by summarising each predictor for quick inspection and cleanse any missing values by either imputting or removing them. The following common data problems should also be controlled -
    - Duplicated data
    - Unlikely extreme values
    - Inconsistently formatted entries; such as dates
- Manipulate data
  - Data transformation is important in preparing it for modelling and visualisation. This can include creating new variables from the current data, such as interaction terms, or changing variables types, for example from numeric to a factor. If there are illegal variables in your dataset, such as an individuals Tax File Number or similar private information, these should be removed. Numerical transformations of variables such as logarithmic transformations should be taken when appropriate.
- Visualise data
  - Probably the most important step, visualised data can be easily interpreted and can summarise quickly. It can reveal any further explorations needed before modelling, any missing values or any large outliers that may need cleaning. This can include-
    - plotting correlations between variables in a correlation matrix
    - probability density functions of variables
    - histograms of factor variables
    - scatterplots to check for linearity
    - Box and whisker plots

3. **Refine question and form a new question**

- Using the insights from above, the original question should now either be answered, need further investigation or found to be inappropriate. This question should be refined accordingly and a new one will be made. Repeat steps 1 - 3 until all relevant questions are answered.

Task 2: The EDA objectives in this report

There are three main objectives in this report-

1. Observe data
2. Assess missing values
3. Prepare for modelling.

These objectives will be explored through a number of questions in the iterative cycle of EDA.

- **Observe data**
  - Observing the data is a simple objective, but important in noticing any prominent data errors before continuing forward. This step can highlight any inherent data problems that need to be addressed.
  - A good observation of the data includes *understanding the format* of the data. This is critical in analysing the data itself and can reduce interpretability and errors at future stages.
  - Any *abnormal values* should be found prior to any indepth exploration. Significant outliers or inconsistent entries can create a number of issues including duplicated data or large unexpected deviances.
- **Assess missing values**
  - Missing values cause a range of issues. Explicitly missing values can cause a range of computational errors in R functions. They can also introduce bias which can lead to invalid conclusions. However, missing values can also be informative when they can predict a change in price. Therefore, they need to be validly assessed.
  - Missing values need to be either imputed or removed before analysis as they can create modelling errors. They can either be removed, losing information in the process, or imputed. The treatment of missing values can depend on the proportion of missing values. If 50% of a dataset is missing, it would not be appropriate to remove these values.
- **Prepare for modelling**
  - *Multicollinearity* leads to unstable estimates and higher variance. Therefore removing any highly correlated terms is necessary for effective modelling.
  - *Non-linear data* will not model well under linear regression. Therefore, some transformations of the data may be needed to appropriately visualise predictors and understand any underlying relationships.
  - *Interaction terms* can be informative in many modelling techniques. Since they can vastly improve some model's predictive capabilities, they should be considered. Interaction terms can also increase the interpretability of a model.

Task 3: EDA

Objective 1: Observe Data

Question 1: What does my data look like?

The first step was to load the data into Rstudio through the read.csv function. The packages "tidyverse", "ggpubr", "kableExtra" and "corrplot" were also loaded and will be used for the remainder of this report for calculations and visualisations. A quick inspection of the data shows that both the first and last entries have no missing values and that there are 13 columns in the data.

Table 1: First 6 rows

Price	PostedBy	UnCon	RERA	NoR	ToP	Size	RtM	RoN	Longitude	Latitude	NoS	NoP
418	Dealer	0	0	3	BHK	1668.6627	1	1	30.75232	76.77282	0	3
132	Owner	0	0	1	BHK	587.9733	1	1	21.68810	72.27630	1	0
1500	Dealer	0	1	2	BHK	1211.8274	1	1	19.18490	72.83470	4	2
330	Dealer	0	0	3	BHK	1100.0000	1	1	26.86260	75.76330	3	2
200	Owner	0	0	3	BHK	800.0000	1	1	28.66667	77.41667	1	2
450	Dealer	1	1	1	BHK	700.0622	0	1	21.56128	74.21238	1	0

Table 2: Last 6 rows

	Price	PostedBy	UnCon	RERA	NoR	ToP	Size	RtM	RoN	Longitude	Latitude	NoS	NoP
19904	180	Owner	0	0	1	BHK	609.9627	1	1	16.05654	73.46875	1	0
19905	579	Dealer	0	1	3	BHK	1400.2418	1	1	26.90093	75.77593	3	2
19906	1000	Dealer	0	0	2	BHK	1508.7508	1	1	15.43852	74.99037	1	2
19907	185	Owner	0	0	2	BHK	777.3109	1	1	15.96393	73.99890	1	2
19908	416	Dealer	1	0	1	BHK	657.0842	0	1	18.51042	73.78194	0	1
19909	350	Owner	0	0	2	BHK	916.2304	1	1	18.74478	73.82365	1	2

Duplicate values

There were only 397 duplicated rows in the data. These were not removed from the dataset since it was assumed that there could be apartment blocks which may be seperate listings but are otherwise identical.

Answer: The data has 13 columns and 19909 rows. It is relatively clean, with no significant data duplicates and understandable formatting.

Question 2: Are there any abnormal values?

All columns were summarised using R's "summary" function and plotted to show general relationships and potential outliers (results in appendix).

Answer: These summaries showed the following abnormal entries -

- Values of 3 in column UnCon
- There are both "BHK" and "bHK" entries in the ToP column. These were assumed to both be BHK.
- There are extreme values of 999999 for both NoS and NoP, these were assumed to be errors
- There are explicit missing values across a number of variables

Objective 1: Conclusion

The data is relatively clean. There are a few missing values and a number of entries entered incorrectly, however, these values are an overwhelming minority. In the second objective these inconsistent entries will need to be altered.

Objective 2: Assess missing values

Question 3: How should the missing values be treated?

The errors found in the observing the data will be altered before any assessment of missing data to get a full depiction of the missing entries. These were the alterations-

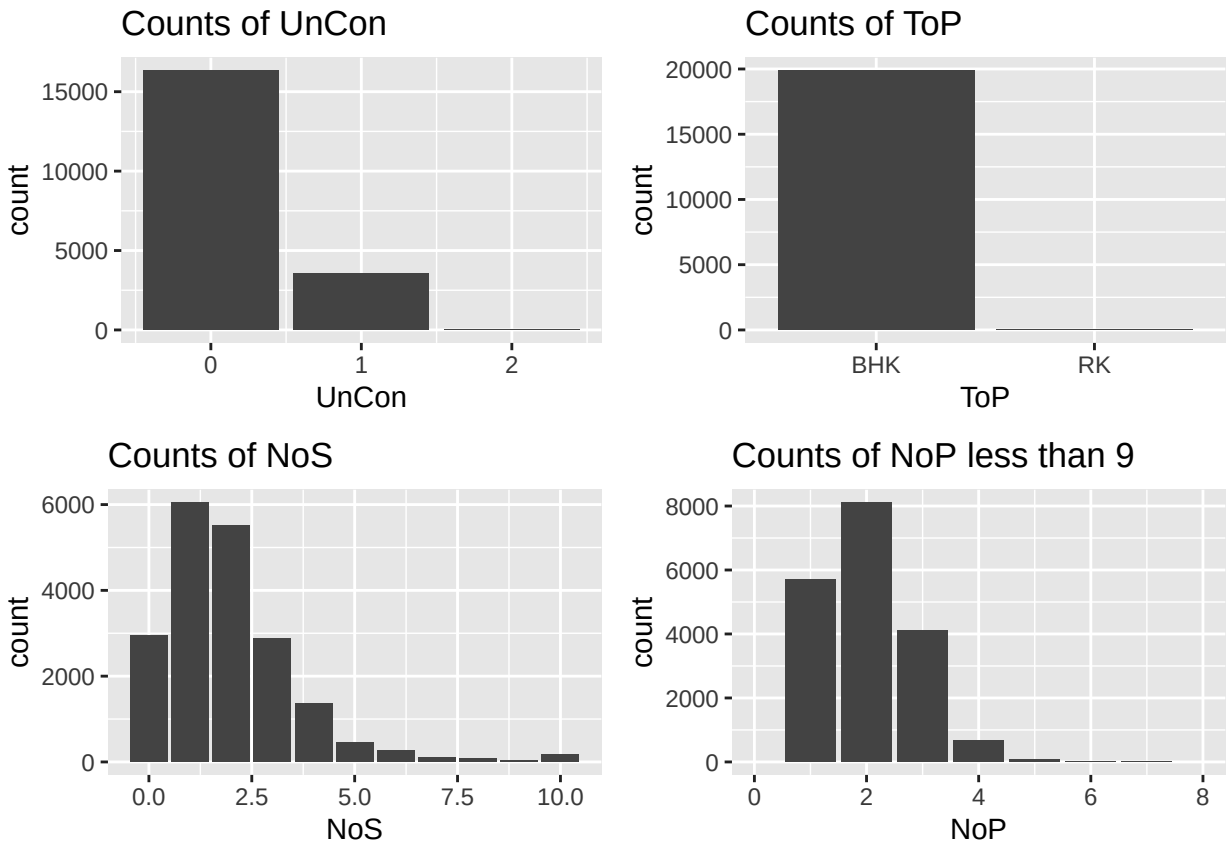
- All UnCon values of either 3 or NA were changed to 2 to represent missing entries
- "bHK" entries were corrected to "BHK"
- NoS and NoP values greater than 1000 were changed to NA, since it was assumed that this is an unreasonable amount of nearby parks and supermarkets

Now that all data corrections have been made, the sample below shows some rows with at least one missing value. There does not seem to be any relationships between missing data.

Table 3: First 13 rows of missing data

	Price	PostedBy	UnCon	RERA	NoR	ToP	Size	RtM	RoN	Longitude	Latitude	NoS	NoP
11	NA	Owner	0	0	2	BHK	919.1176	1	1	12.77803	77.63219	1	2
178	260	Dealer	0	0	1	BHK	NA	1	1	18.96611	73.14828	0	0
456	420	Dealer	2	NA	2	BHK	1065.1788	0	1	12.84580	77.67270	0	2
457	5400	Dealer	2	NA	4	BHK	5655.0424	1	1	18.55119	73.94994	5	3
458	5200	Dealer	2	NA	5	BHK	5267.4230	1	1	22.54111	88.33778	7	5
773	310	Owner	0	0	1	NA	674.9401	1	1	12.82257	77.66735	0	1
1444	NA	Owner	1	1	3	BHK	960.7894	0	1	26.83235	75.84175	0	3
1778	630	Dealer	0	0	2	BHK	NA	1	1	18.84129	73.14879	1	1
4355	850	Owner	0	0	2	BHK	2000.0000	1	1	18.52361	NA	NA	1
4918	3400	NA	0	0	3	BHK	3433.9966	1	1	28.40314	77.07005	5	2
5997	545	Dealer	0	1	2	BHK	1095.0372	NA	1	30.68452	76.69173	1	1
7564	NA	Dealer	0	0	3	BHK	1484.2418	1	1	19.10383	72.83736	6	2
7888	1100	Dealer	0	1	3	BHK	NA	1	1	28.60233	77.40199	2	2

The plots below show the resulting cleaned data and how the data is distributed across a number of factors. *Note: there are only 14 NoP values greater than 8.*



Answer: There appears to be no underlying trend between missing values. They are not clustered at any particular location of the dataset and do not appear to be informative. Since there seems to be no trend amongst missing values and there are only 27 rows missing (0.14% of all data), all missing data was removed.

Objective 2: Conclusion

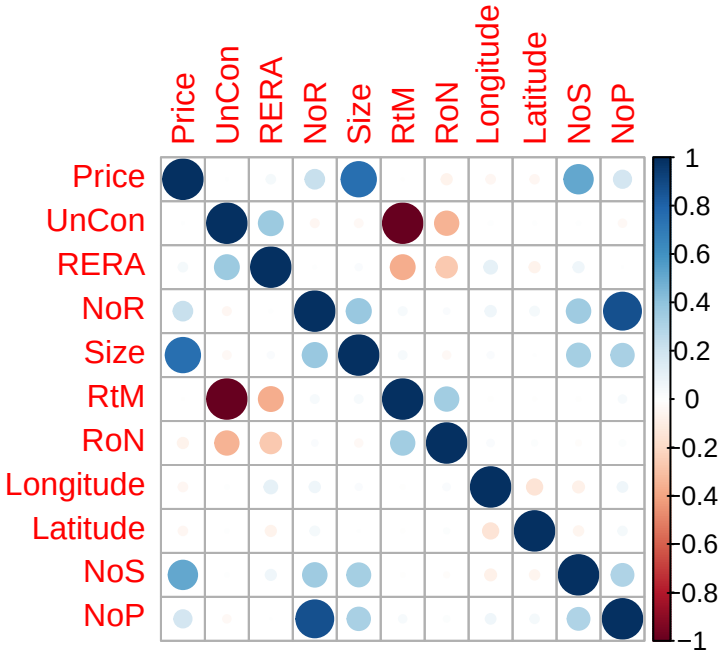
There were some missing entries in the dataset, however, since they were few, all were removed from the dataset. All variables now have consistent data entries and have no missing values.

Objective 3: Prepare for modelling

Question 4: Are there any collinear terms?

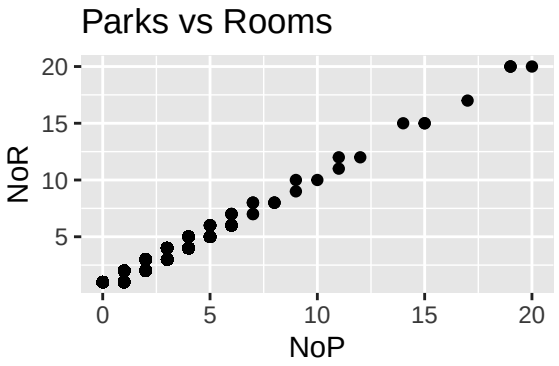
To understand how terms are correlated with each other, a correlation plot was used. From this figure, two pairs of variables stand out-

1. UnCon and RtM with a correlation of -1
2. NoP and NoR with a correlation of 0.87



The relationship between UnCon and RTM is expected, as buildings under construction should not be moved into. Filtering the data reveals that there are no buildings which are both under construction and ready to move, the only discrepency between the two is that UnCon has more missing values. Therefore it is removed as it is the least informative.

The second pair of highly correlated variables is unexpected, the number of nearby parks is closely related to the number of rooms. This relationship is shown below.



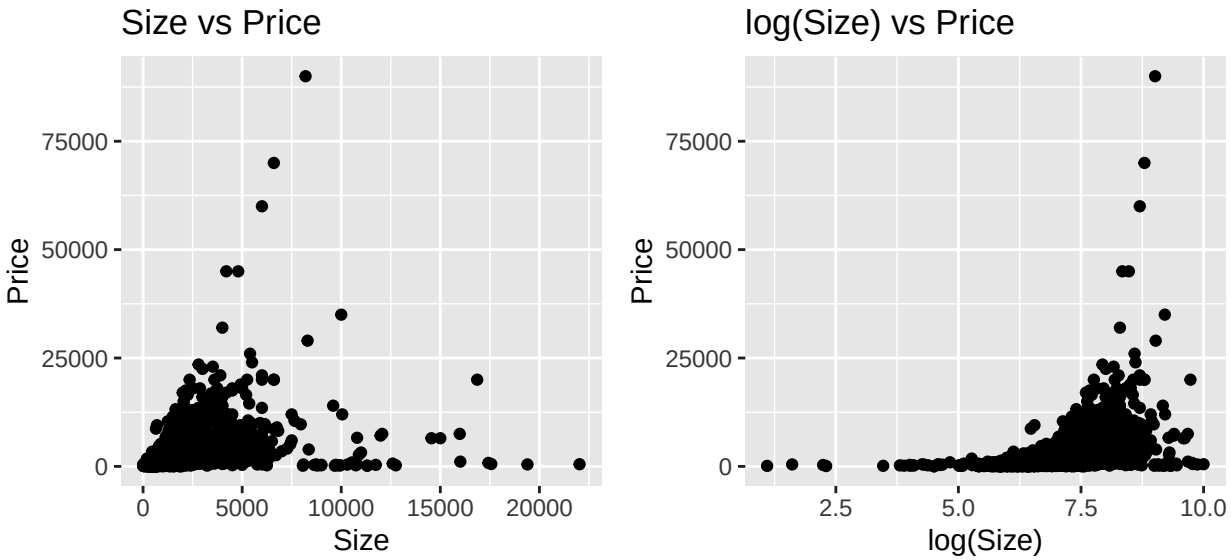
Since the number of rooms is a better indicator of price with a correlation of 0.22 compared to 0.19 the number of parks was removed.

Answer: There were some collinear terms. Houses that were under construction were never ready for moving into and houses with more rooms usually had more surrounding parks. To avoid collinearity problems the UnCon and NoP columns were dropped.

Question 5: Is my data linear?

Price will be used as the response variable in future modelling, therefore all continuous variables will be plotted against it to check for linearity. Aside from price, there are only 3 other continuous variables, size, longitude and latitude. As expected, the logarithmic transformations of longitude and latitude were not helpful but can be found in the appendix.

Since size has a long tail, its range was restriced to 30000. Both size and its log transformation seem to be relatively poor linear fits, however there may be some non-linear trend in log(size).



Answer: There seems to be no appropriate transformation of house size that can improve upon the current predictor. There are no transformations required in the dataset.

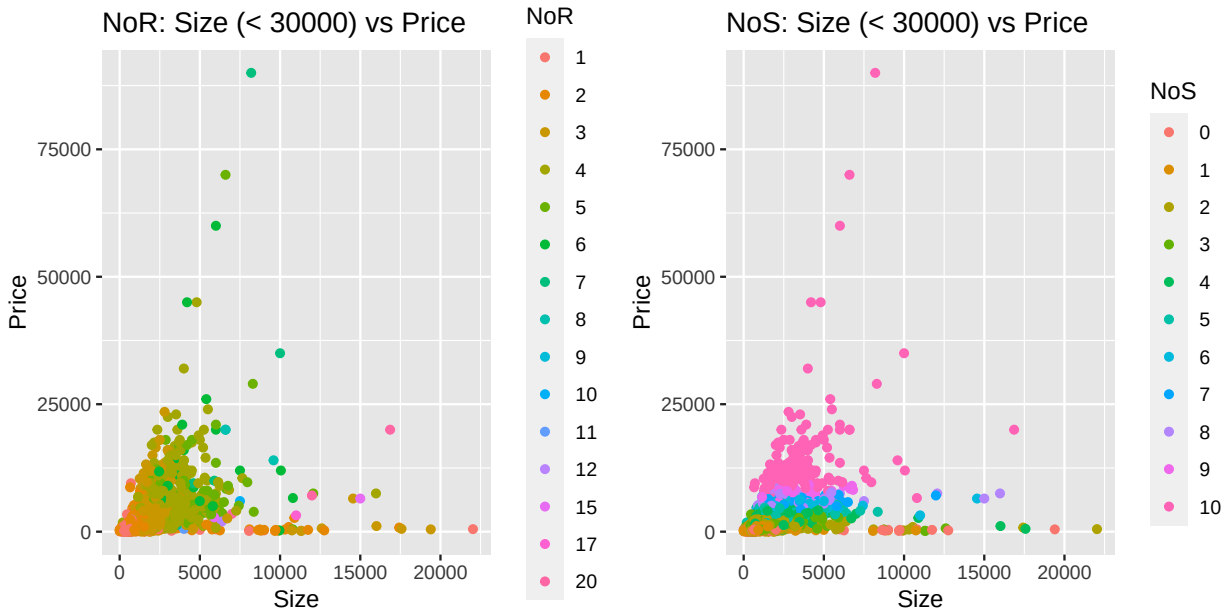
Question 6: Are there any interaction terms?

When present, interaction terms can be a powerful tool in modelling and are easily interpretable. By plotting variables against one another, it seems that there is no obvious interaction between any of the predictors. Since there is no interaction of note, all results of this can be found in the appendix.

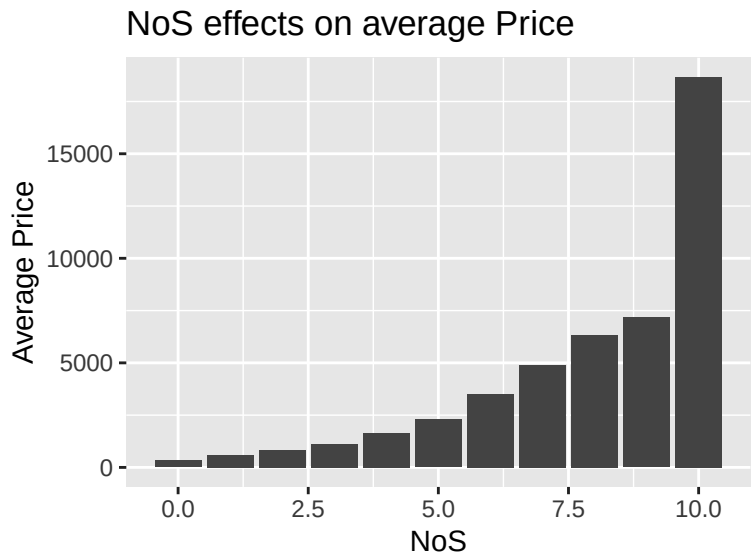
Answer: There seems to be no obvious interaction effects between terms, however in future modelling, such as tree-based methods, a relationship may become clear. Therefore, no interaction terms will be introduced in the EDA.

Question 7: How do prices differ between house characteristics?

The change in price relative to other house characteristics can be seen by plotting the two most informative continuous variables against each other - size vs price. This produced some very interesting results (see appendix for all figures). Here we will discuss the effects of 2 factors- Number of Supermarkets (NoP) and Number of Rooms (NoR).



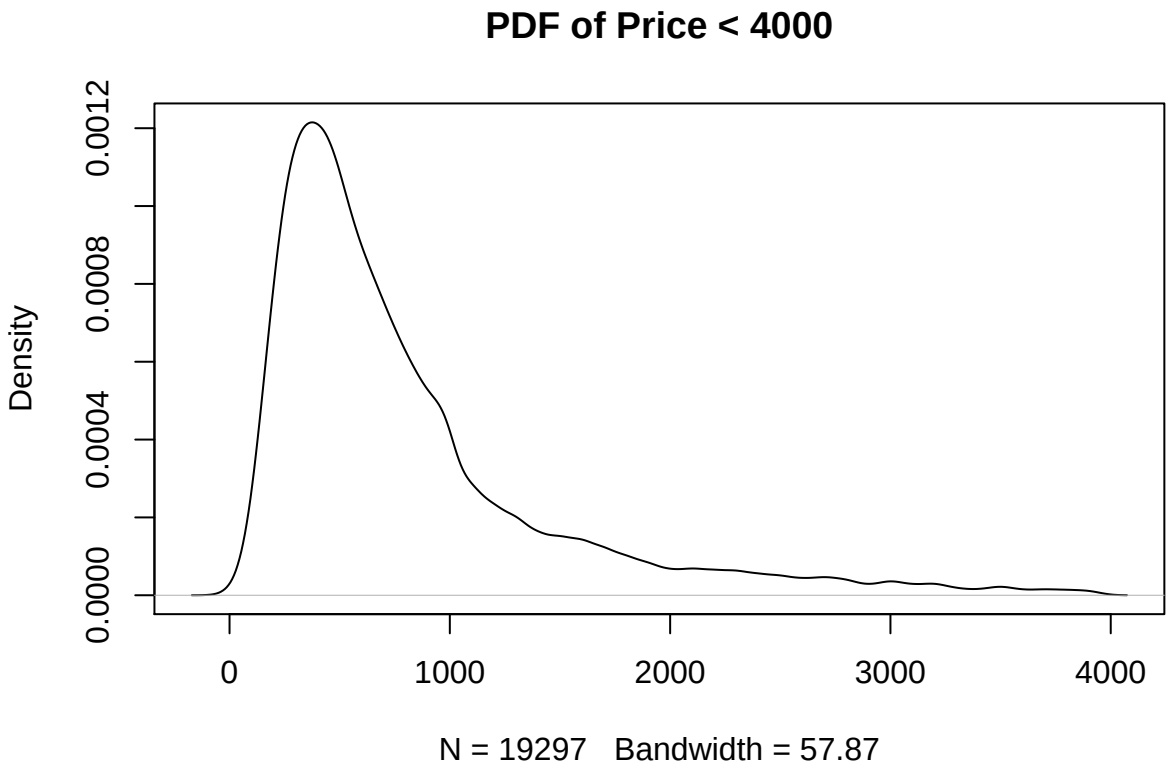
As expected, the number of rooms and the size of the house was somewhat correlated as seen from above. However, the close relationship between the number of nearby supermarkets and price is unexpected. This is again demonstrated by the graph below.



Answer: There seems to be a strong relationship between the number of nearby supermarkets and the house price. This should be kept in consideration for future modelling.

Question 8: How is price distributed?

For constructing Genereralisted linear models and to get an overview of the spread of the prices, its helpful to understand how the price is distributed.



Answer: Price is heavily positively skewed with a mean of 1069.51 and a median of 610. It also has a very large tail with 18 houses costing more than 4000% of the average home.

Objective 3: Conclusion

The data is now prepared for modelling. Important factors to consider when beginning the modelling stage is that price is highly positively skewed and that the number of nearby supermarkets affects housing prices.

Task 4: Conclusion of EDA

Recommeded Models

For further analysis, a number of models should be used to fit the pricing data including; linear regression, a generalised linear model (GLM), regression trees and a neural network. Since one of the main aims of this report is to distinguish the features that affect buyers behaviour, the interpretability of a model is prioritised as well as its predictive power.

Linear Regression: Linear regression is effective in that it is simple, easily interpretable and can be a powerful predictor. Since the number of predictors is small, best subsets selection may be able to be used to identify the best regression model. Rectangularization methods such as lasso, ridge and elastic net should also be considered for feature selection and to choose the optimal bias-variance trade-off through an appropriate lambda choice.

GLM: Since price has such a large tail, the gamma distribution would be the most appropriate response distribution. The log link is also recommended for easily interpreting the coefficients. Unlike linear regression, GLM's can capture non-linear trends and so may provide some insights into the pricing behaviour of Cydney that linear regression could not.

Regression Trees: Trees can be a helpful visualisation of how terms are interacting with one another and can also have strong predictive capabilities. Therefore this report recommends using tree based methods such as Random forest, bagging and boosting to model price.

Neural Network: Although neural networks are usually a much less interpretive model than others, it can be a powerful predictive tool. Since the aim of this study is to both predict prices and understand the underlying factors of the Cydney housing market, a neural network model would not be sufficient to interpret buyer behaviour. But, it should still be used to compare its predictive power relative to other models.

Conclusion:

This EDA has imported and cleaned the data, removed UnCon and number of parks as collinear terms, and shown how price changes with some predictors.

The data was first imported into RStudio. After a quick inspection it was clear that the data had some problems including missing values and inconsistent entries. After replacing mislabled values and removing some extreme outliers where the number of nearby supermarkets or parks exceeded 1000, 0.14% of the data was missing. This was assumed to be negligible and so all missing values were removed.

After creating a correlation matrix, it was found that buildings under construction were never ready to move in, and so the variable UnCon was removed from the dataset. Similarly, the number of parks was removed as it was highly correlated with the number of rooms. After checking that all continuous data was linear, it was also found that there were not any immediately obvious interaction terms and so none were introduced into the dataset. The price distribution also had long tails and was positively skewed. Lastly, it was found that the houses with more supermarkets less than 3km away were on average much more expensive.

Appendix

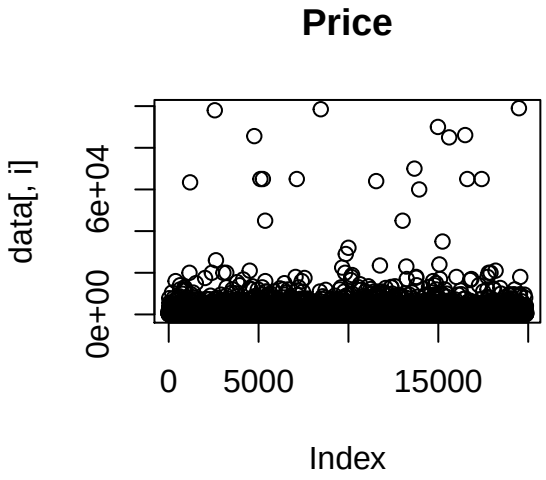
Assumptions

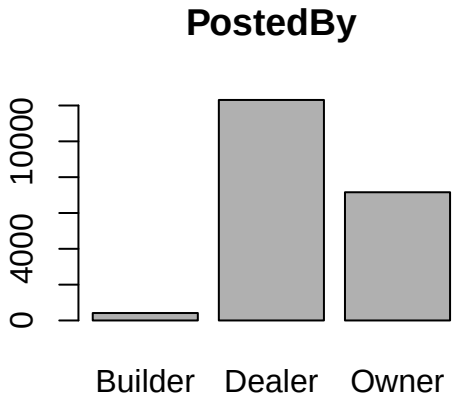
- The number of Parks in a 5km radius of a home is not greater than 1000.
- The number of supermarkets in a 3km radius of a home is not greater than 1000.
- BHK and bHK values were equivalent.
- All pricing values are legitimate and large values are indicative of the Cydney housing market.
- Longitude and Latitude were not used in this report for any area locating.
- Different units in apartment complexes fall under different listings and so can account for any duplicates in the data

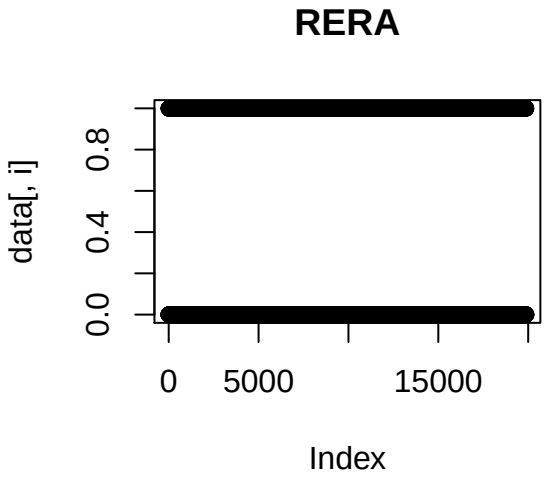
Summary of columns

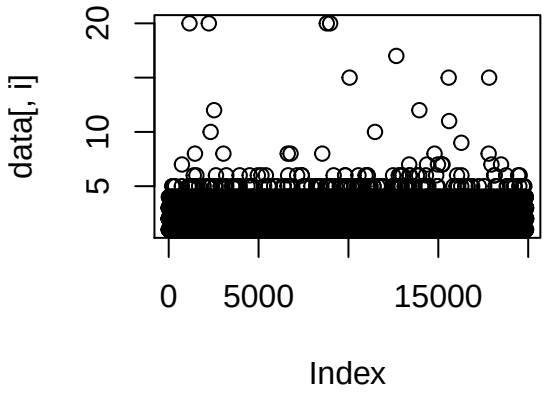
```
## [1] "Price"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     2.5   380.0   610.0  1069.5  1000.0 99000.0
## [1] "PostedBy"
## Builder  Dealer   Owner
##     415   12310    7157
## [1] "RERA"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.3146  1.0000  1.0000
## [1] "NoR"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.392   3.000  20.000
## [1] "ToP"
##   BHK    RK
## 19869    13
## [1] "Size"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       3     900    1170    1351    1545   49690
## [1] "RtM"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  1.0000  1.0000  0.8211  1.0000  1.0000
## [1] "RoN"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  1.0000  1.0000  0.9332  1.0000  1.0000
## [1] "Longitude"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -37.71   18.45   20.91   21.29   26.90   59.91
## [1] "Latitude"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -121.76   73.80   77.32   76.90   77.79  152.96
## [1] "NoS"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   2.000   1.939   3.000  10.000
```

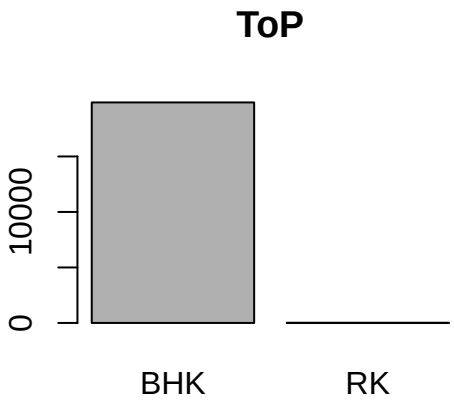
Outlier plots

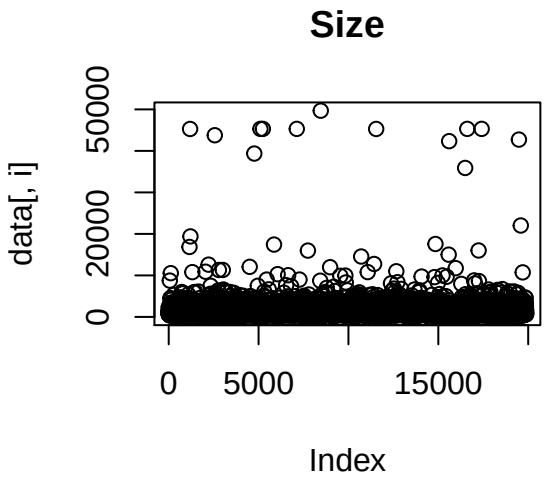


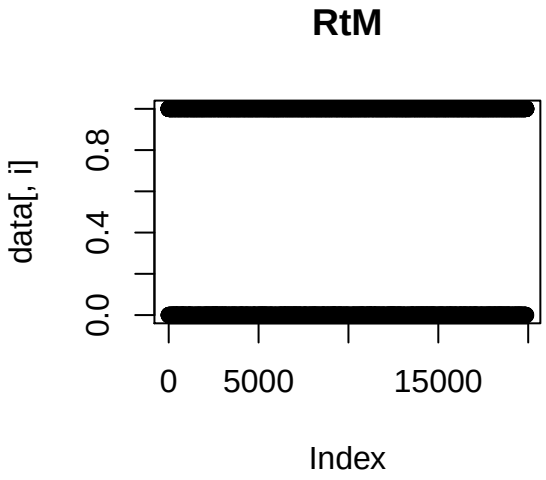


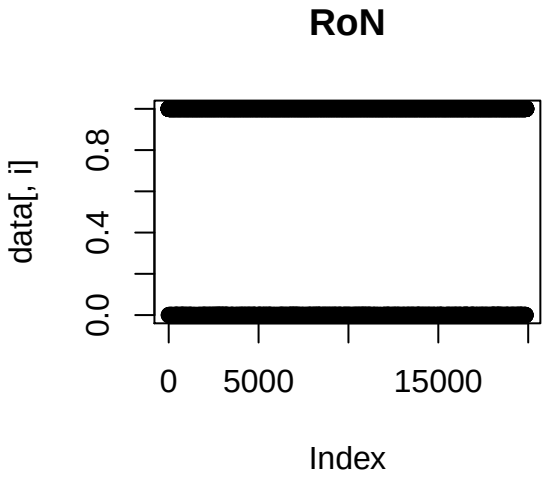


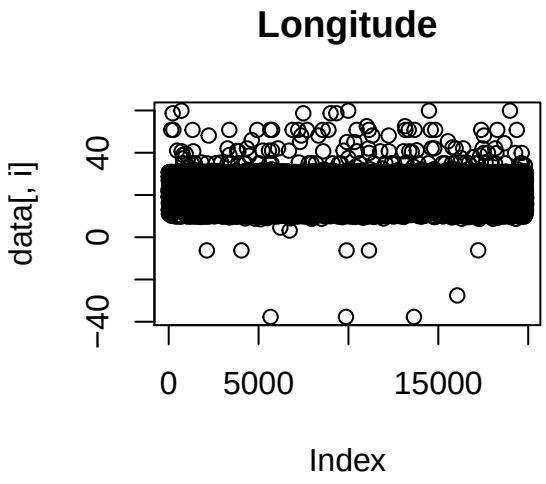


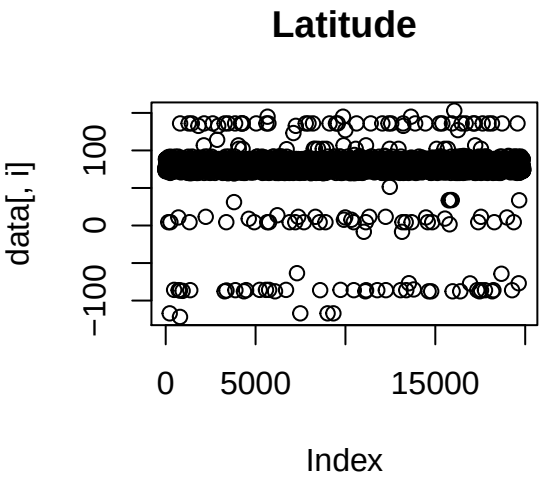

ToP
0
10000
BHK
RK

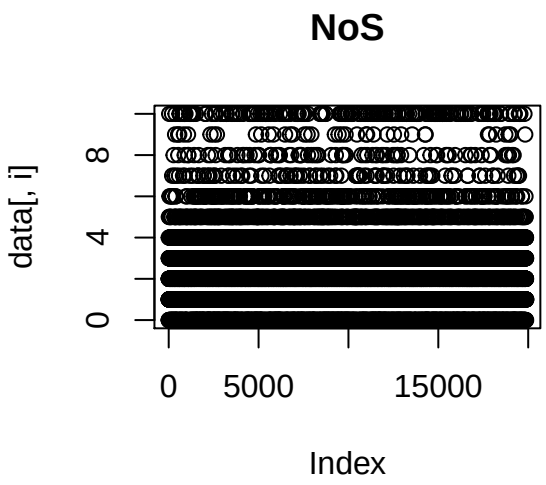

Size
data[, i]
0
20000
50000
0
5000
15000
Index


RtM
data[, i]
0.0
0.4
0.8
0
5000
15000
Index

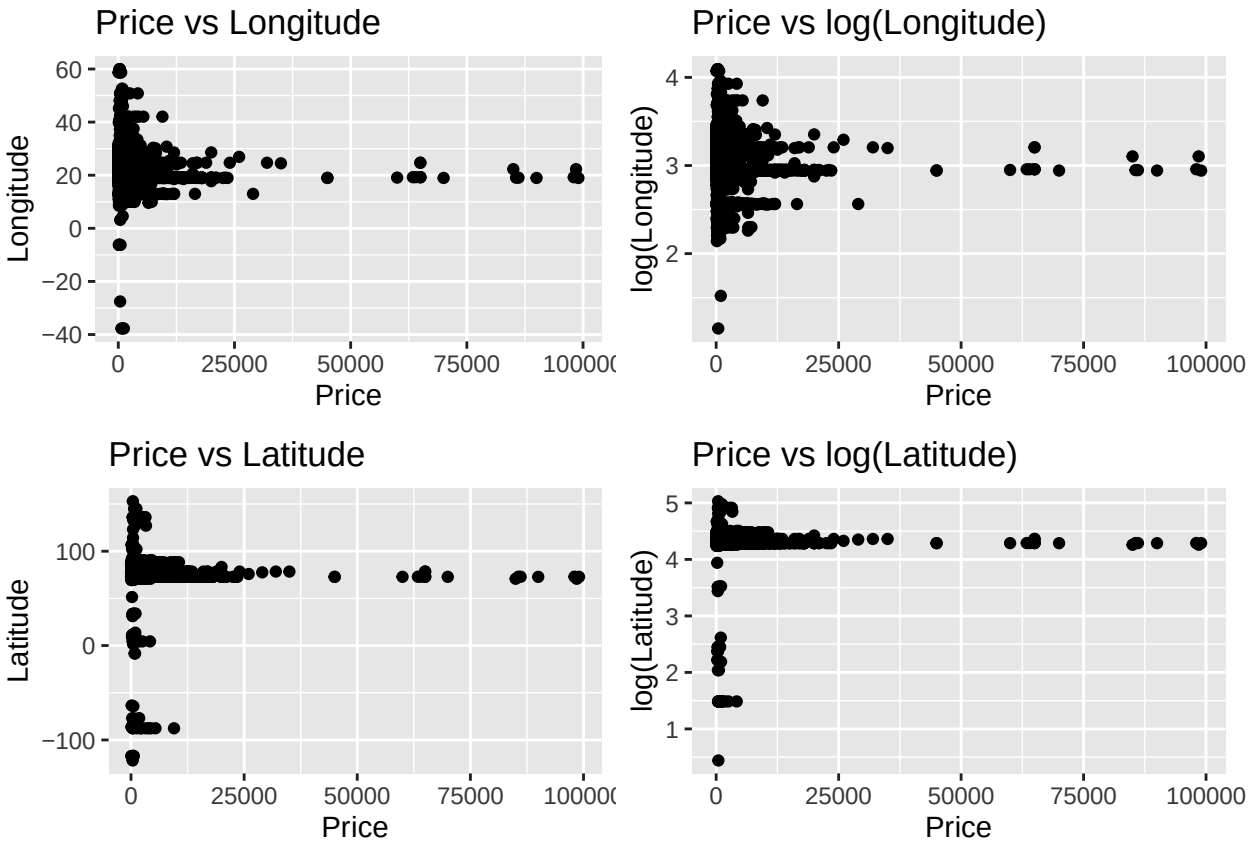

RoN
data[, i]
0.0
0.4
0.8
0
5000
15000
Index


Longitude
data[, i]
−40
0
40
0
5000
15000
Index

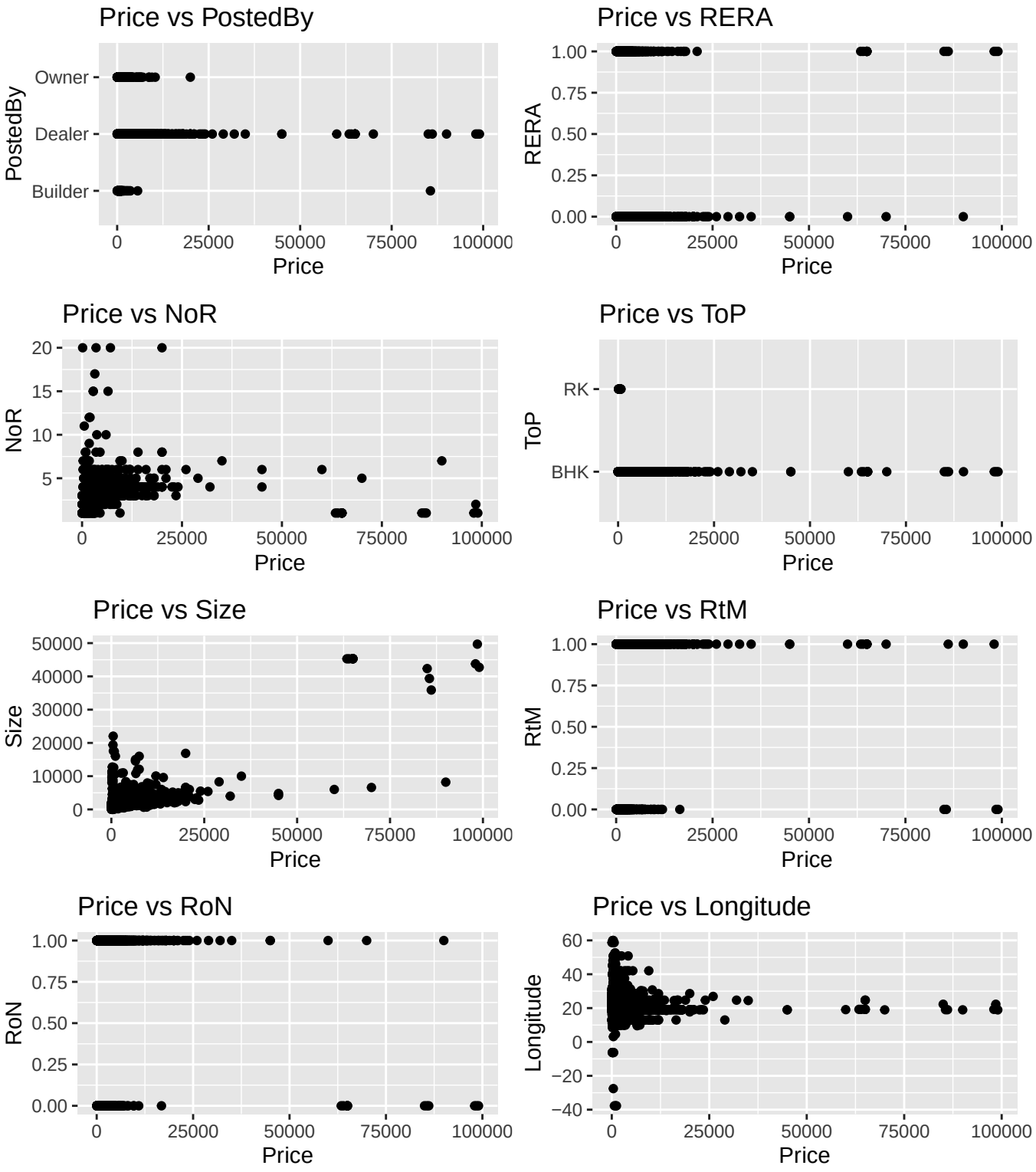

Latitude
data[, i]
−100
0
100
0
5000
15000
Index

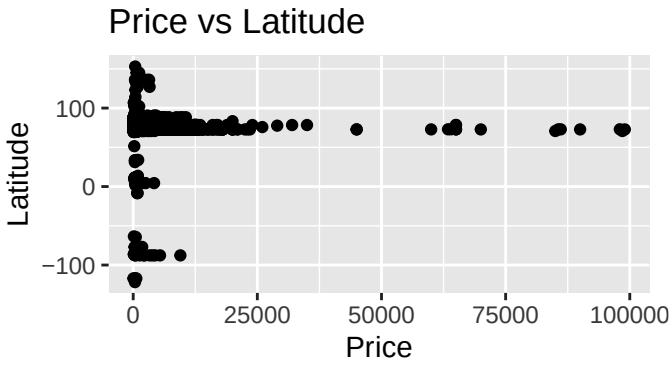

NoS
data[, i]
0
4
8
0
5000
15000
Index

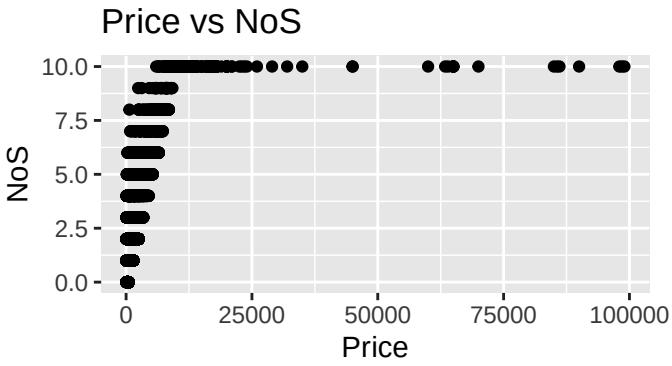
Longitude and Latitude vs Price

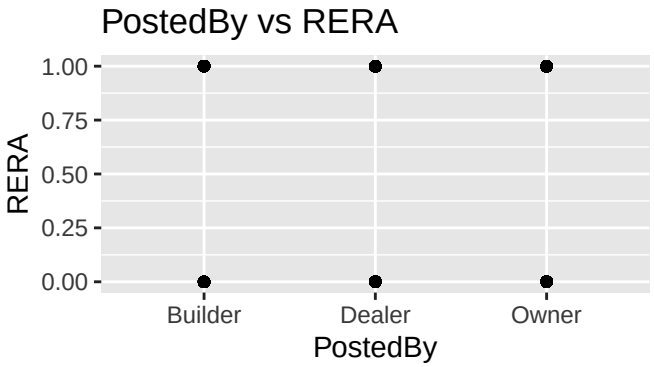


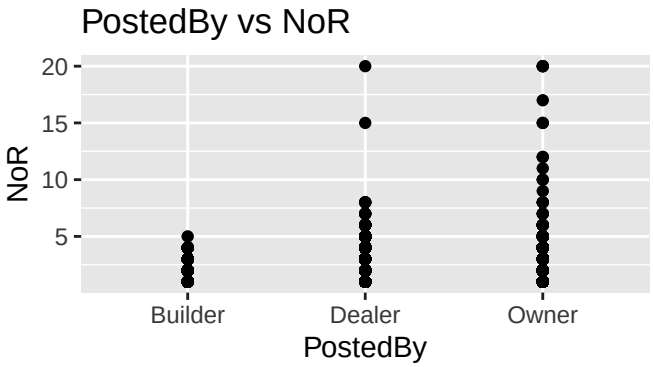
Interaction plots between variables

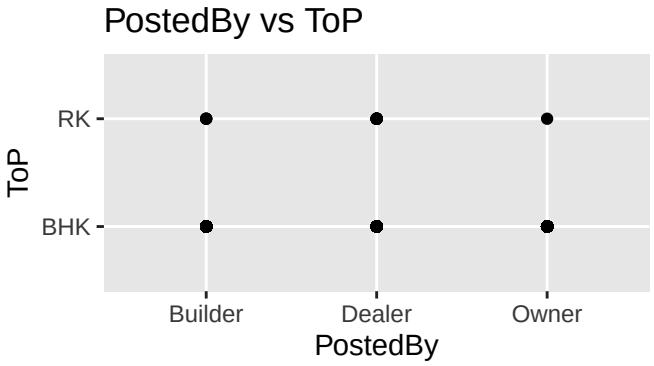


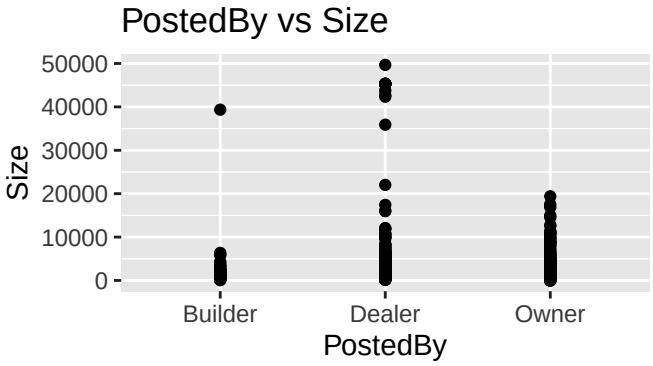

Price vs Latitude
Latitude
100
0
−100
0
25000
50000
75000
100000
Price

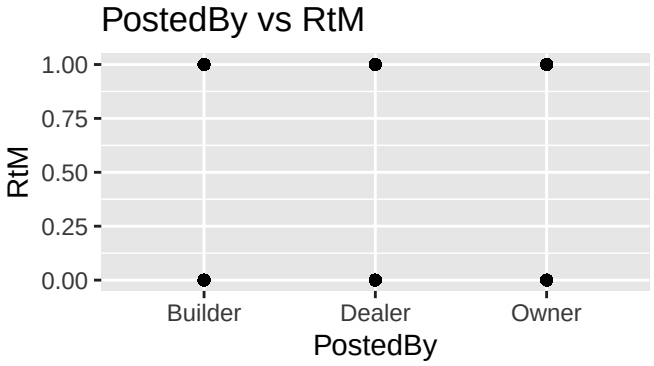

Price vs NoS
NoS
10.0
7.5
5.0
2.5
0.0
0
25000
50000
75000
100000
Price

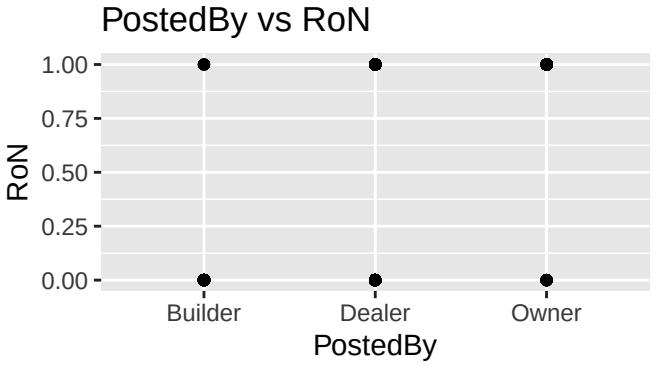

PostedBy vs RERA
RERA
1.00
0.75
0.50
0.25
0.00
Builder
Dealer
Owner
PostedBy

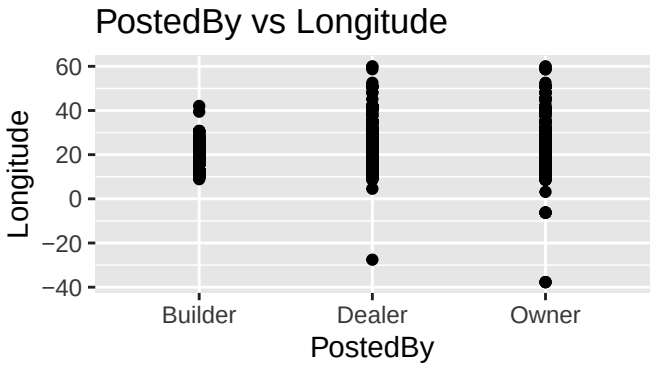

PostedBy vs NoR
NoR
20
15
10
5
Builder
Dealer
Owner
PostedBy

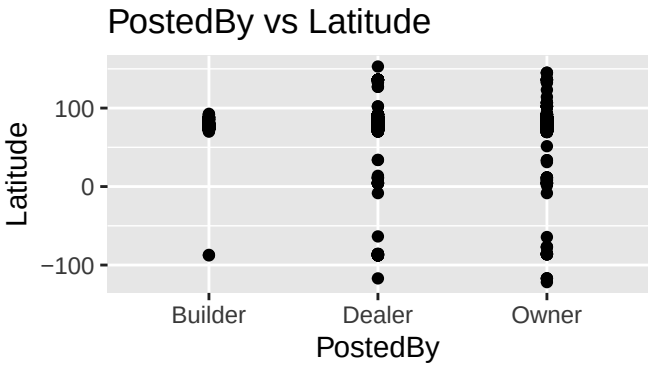

PostedBy vs ToP
ToP
RK
BHK
Builder
Dealer
Owner
PostedBy

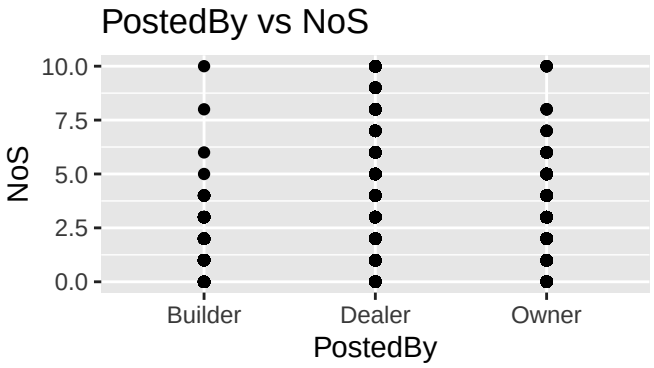

PostedBy vs Size
Size
50000
40000
30000
20000
10000
0
Builder
Dealer
Owner
PostedBy

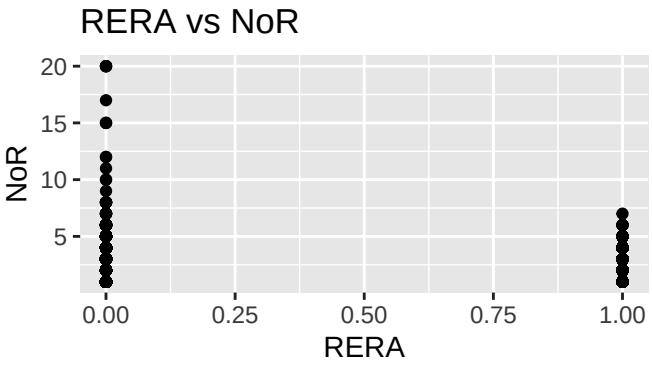

PostedBy vs RtM
RtM
1.00
0.75
0.50
0.25
0.00
Builder
Dealer
Owner
PostedBy

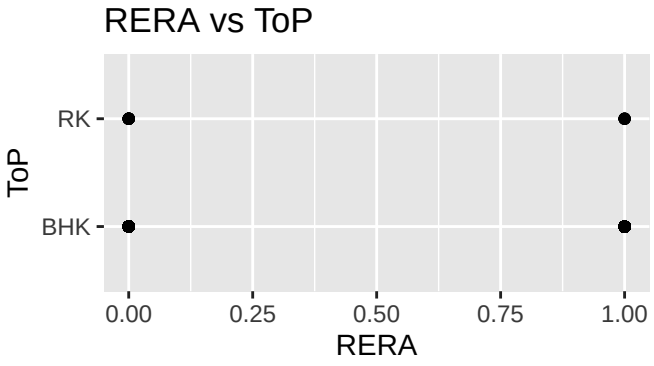

PostedBy vs RoN
RoN
1.00
0.75
0.50
0.25
0.00
Builder
Dealer
Owner
PostedBy

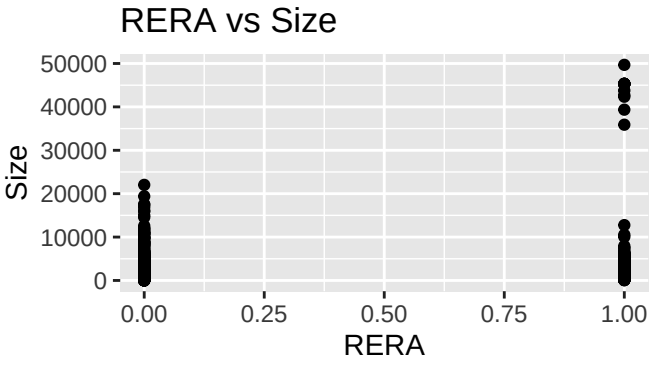

PostedBy vs Longitude
Longitude
60
40
20
0
−20
−40
Builder
Dealer
Owner
PostedBy

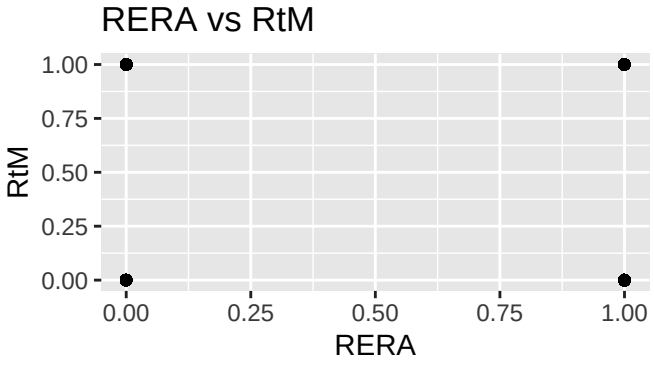

PostedBy vs Latitude
Latitude
100
0
−100
Builder
Dealer
Owner
PostedBy

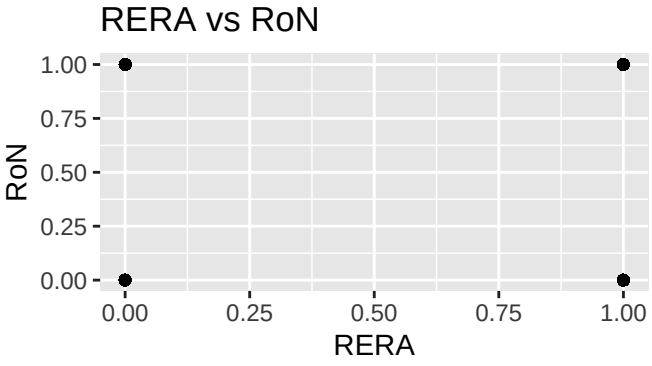

PostedBy vs NoS
NoS
10.0
7.5
5.0
2.5
0.0
Builder
Dealer
Owner
PostedBy

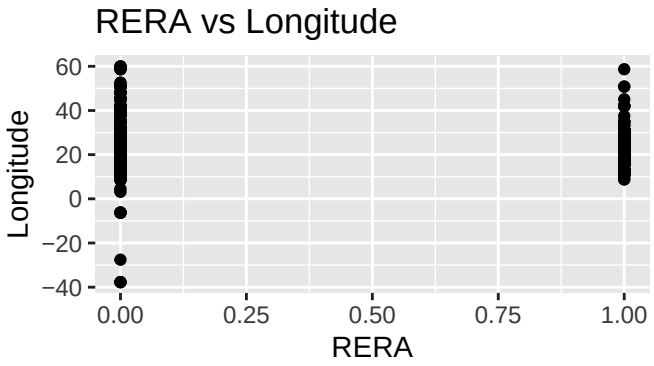

RERA vs NoR
NoR
20
15
10
5
0.00
0.25
0.50
0.75
1.00
RERA

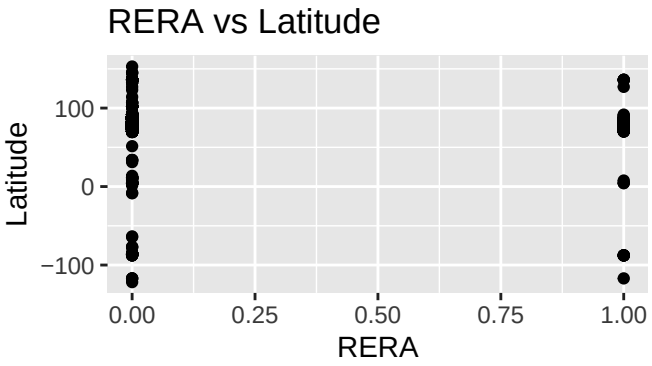

RERA vs ToP
ToP
RK
BHK
0.00
0.25
0.50
0.75
1.00
RERA

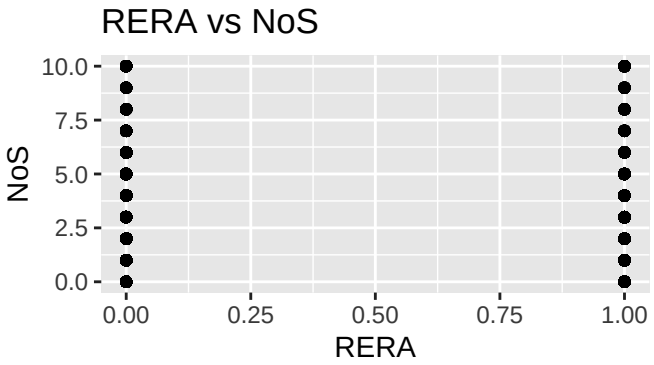

RERA vs Size
Size
50000
40000
30000
20000
10000
0
0.00
0.25
0.50
0.75
1.00
RERA

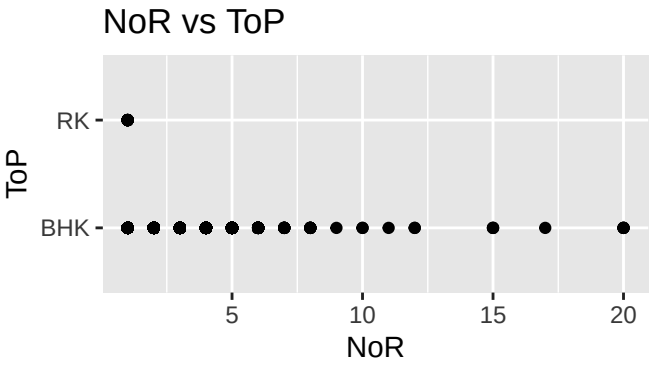

RERA vs RtM
RtM
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
RERA

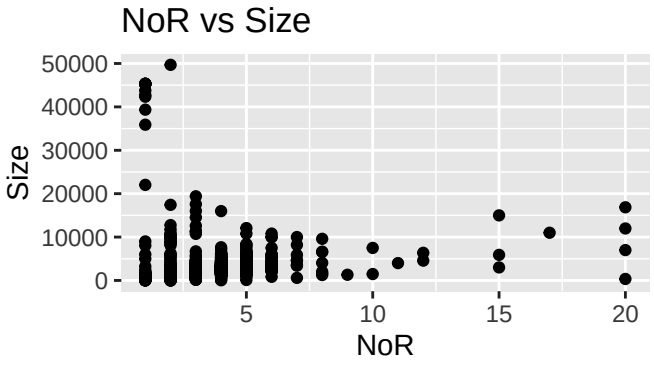

RERA vs RoN
RoN
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
RERA

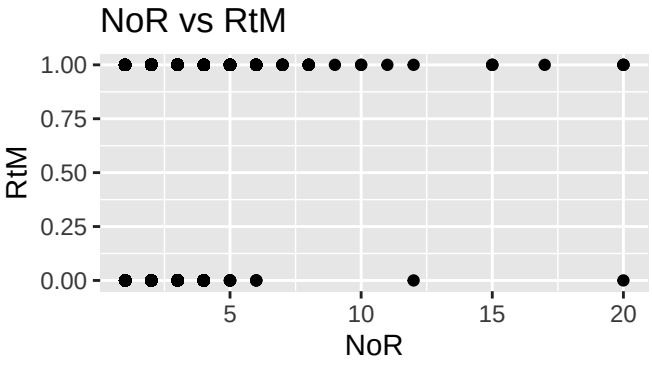

RERA vs Longitude
Longitude
RERA
60
40
20
0
−20
−40
0.00
0.25
0.50
0.75
1.00

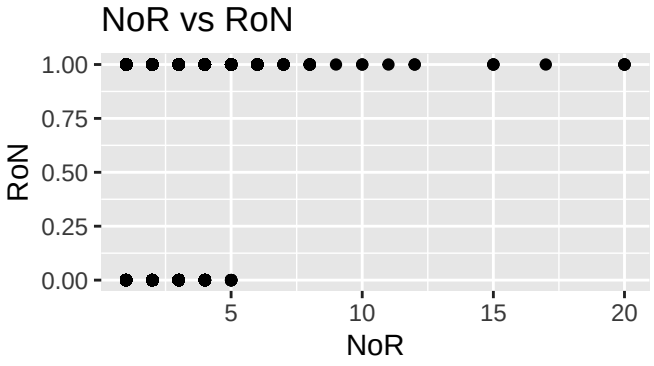

RERA vs Latitude
Latitude
RERA
100
0
−100
0.00
0.25
0.50
0.75
1.00

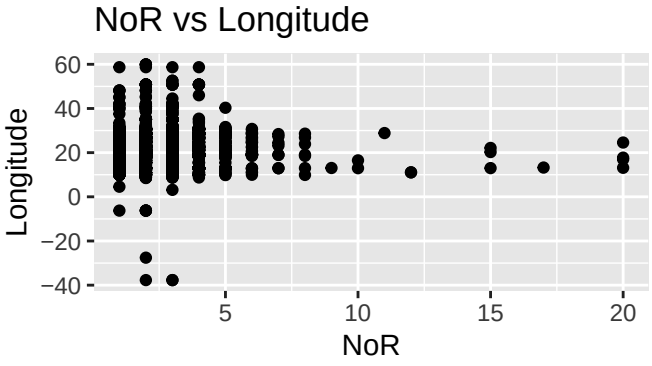

RERA vs NoS
NoS
RERA
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00

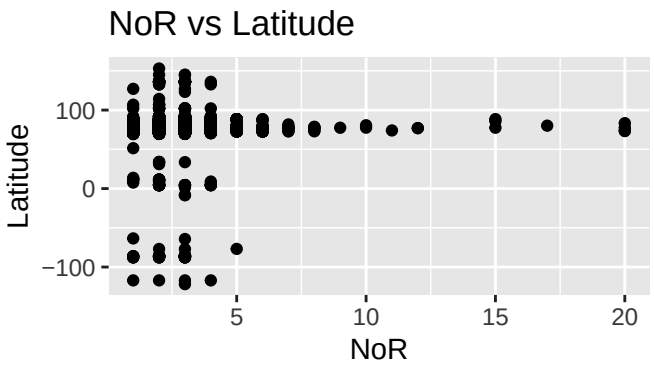

NoR vs ToP
ToP
NoR
RK
BHK
5
10
15
20

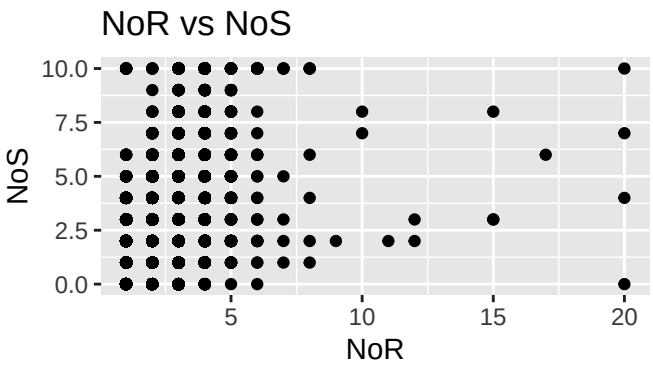

NoR vs Size
Size
NoR
50000
40000
30000
20000
10000
0
5
10
15
20

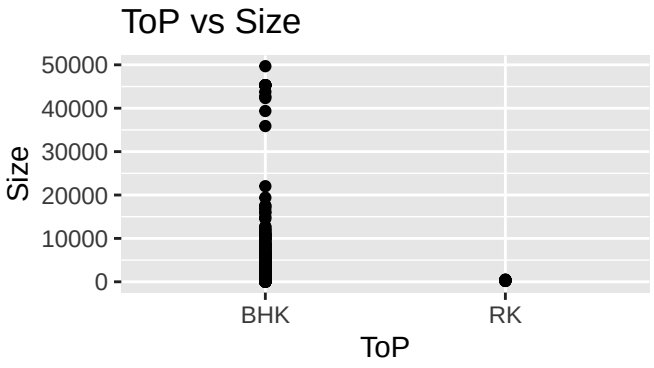

NoR vs RtM
RtM
NoR
1.00
0.75
0.50
0.25
0.00
5
10
15
20

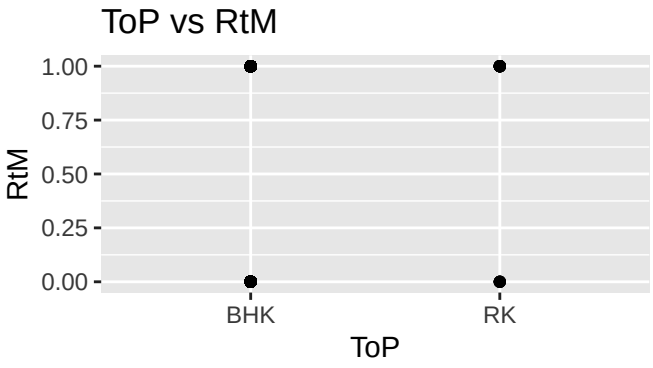

NoR vs RoN
RoN
NoR
1.00
0.75
0.50
0.25
0.00
5
10
15
20

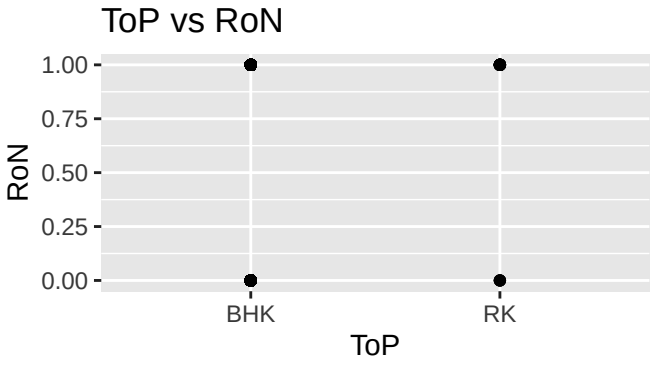

NoR vs Longitude
Longitude
NoR
60
40
20
0
−20
−40
5
10
15
20

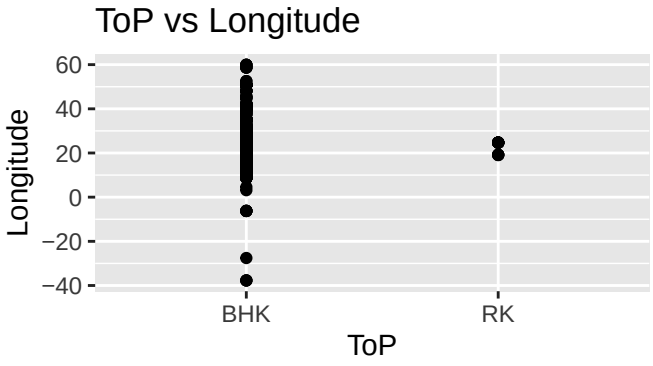

NoR vs Latitude
Latitude
100
0
−100
5
10
15
20
NoR

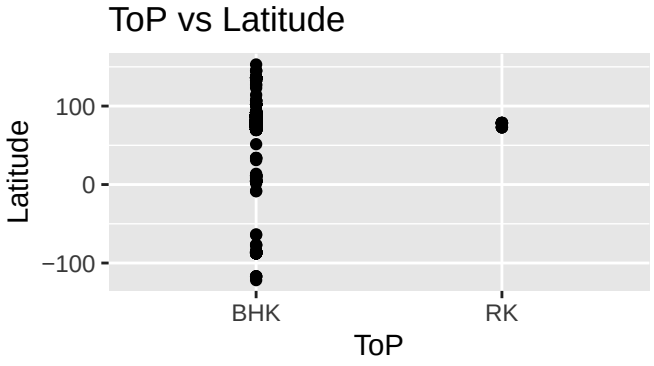

NoR vs NoS
NoS
10.0
7.5
5.0
2.5
0.0
5
10
15
20
NoR

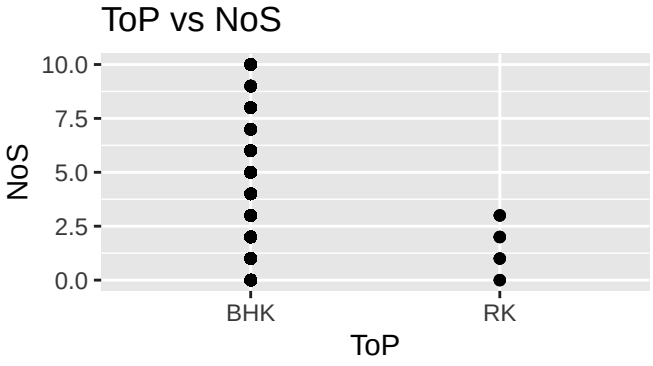

ToP vs Size
Size
50000
40000
30000
20000
10000
0
BHK
RK
ToP

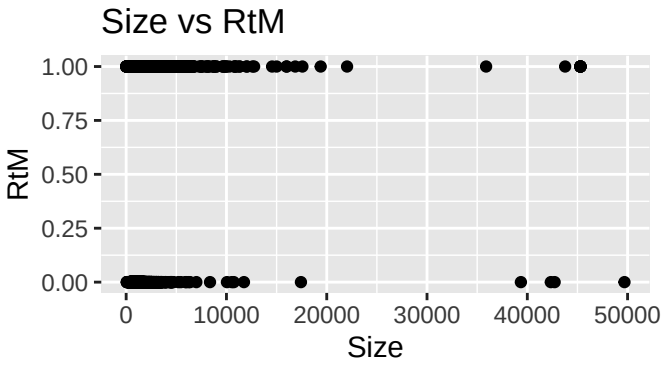

ToP vs RtM
RtM
1.00
0.75
0.50
0.25
0.00
BHK
RK
ToP

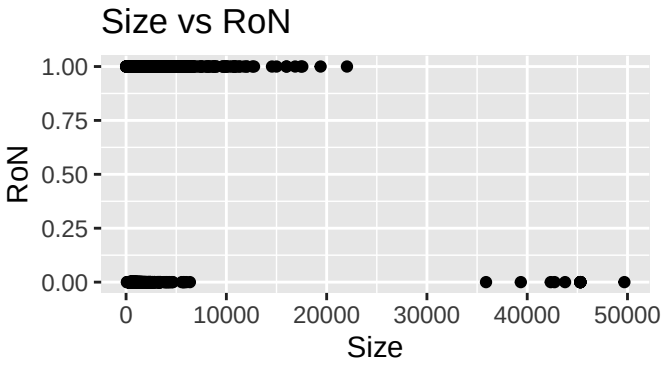

ToP vs RoN
RoN
1.00
0.75
0.50
0.25
0.00
BHK
RK
ToP

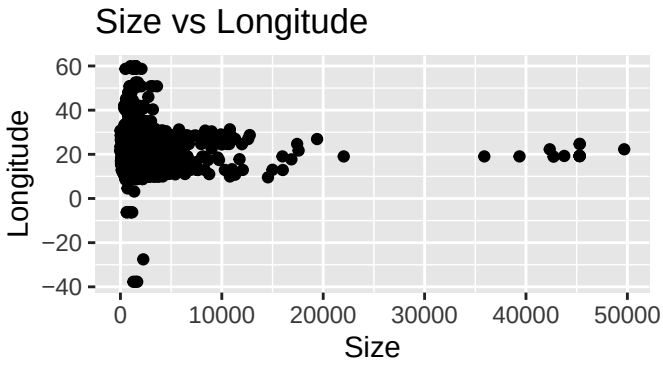

ToP vs Longitude
Longitude
60
40
20
0
−20
−40
BHK
RK
ToP

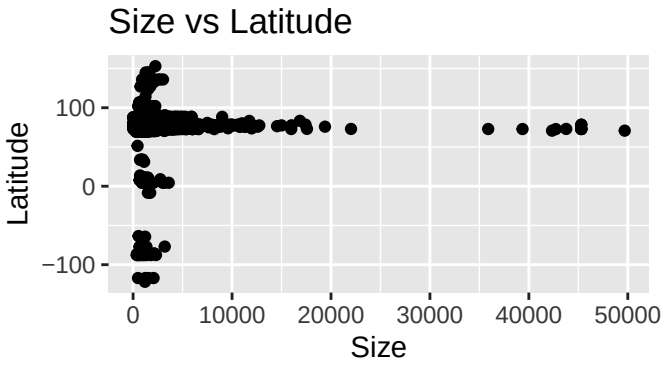

ToP vs Latitude
Latitude
100
0
−100
BHK
RK
ToP

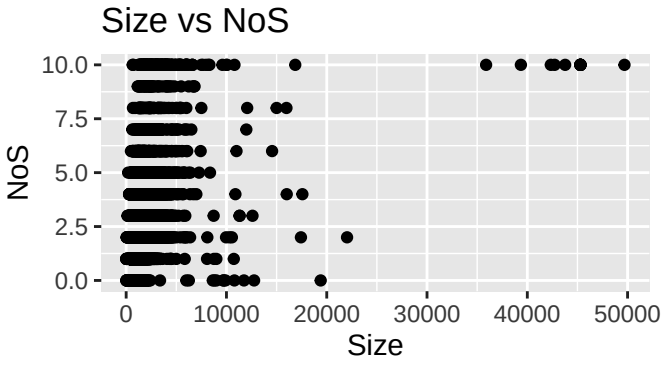

ToP vs NoS
NoS
10.0
7.5
5.0
2.5
0.0
BHK
RK
ToP

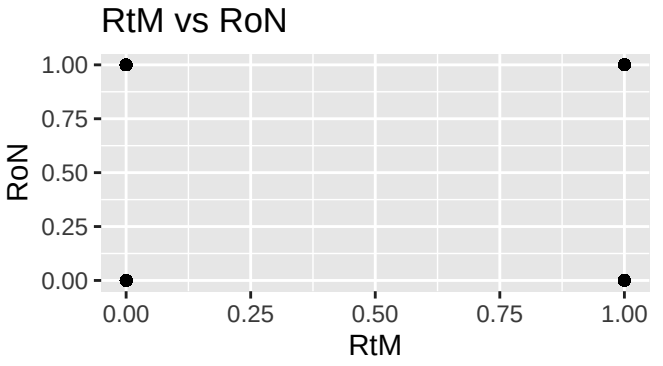

Size vs RtM
RtM
Size
1.00
0.75
0.50
0.25
0.00
0
10000
20000
30000
40000
50000

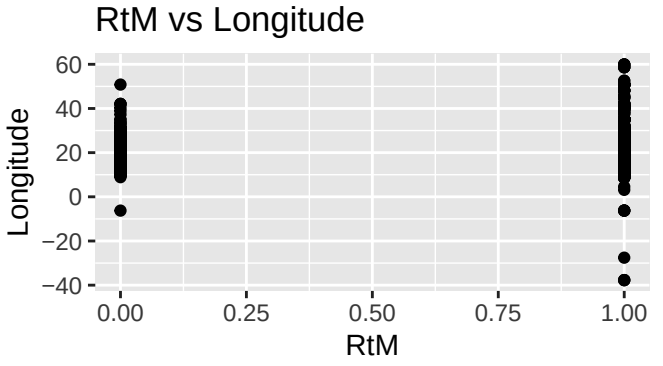

Size vs RoN
RoN
Size
1.00
0.75
0.50
0.25
0.00
0
10000
20000
30000
40000
50000

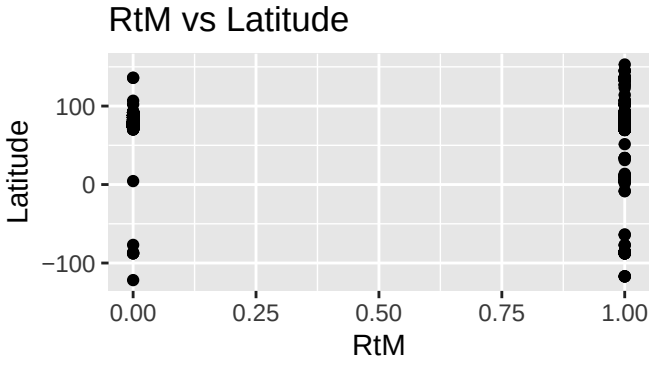

Size vs Longitude
Longitude
Size
60
40
20
0
−20
−40
0
10000
20000
30000
40000
50000

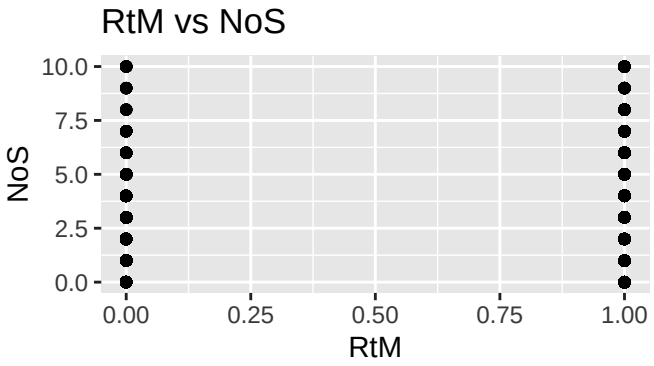

Size vs Latitude
Latitude
Size
100
0
−100
0
10000
20000
30000
40000
50000

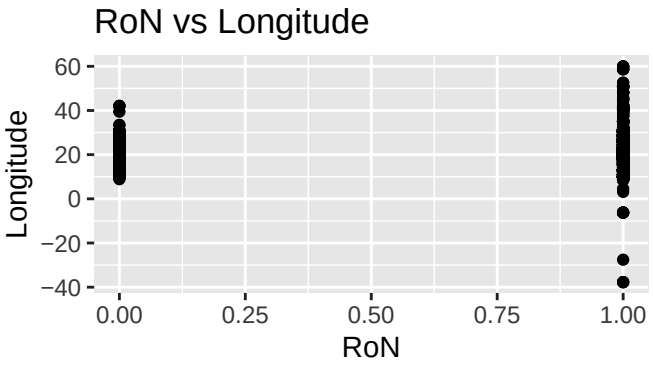

Size vs NoS
NoS
Size
10.0
7.5
5.0
2.5
0.0
0
10000
20000
30000
40000
50000

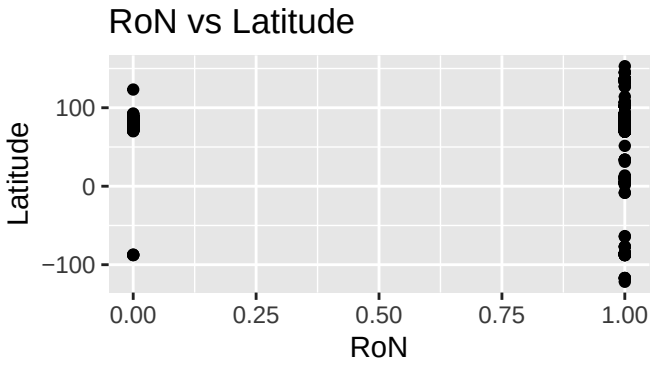

RtM vs RoN
RoN
RtM
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00

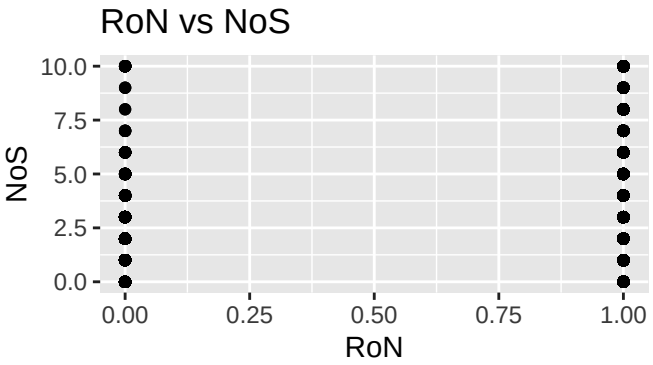

RtM vs Longitude
Longitude
RtM
60
40
20
0
−20
−40
0.00
0.25
0.50
0.75
1.00

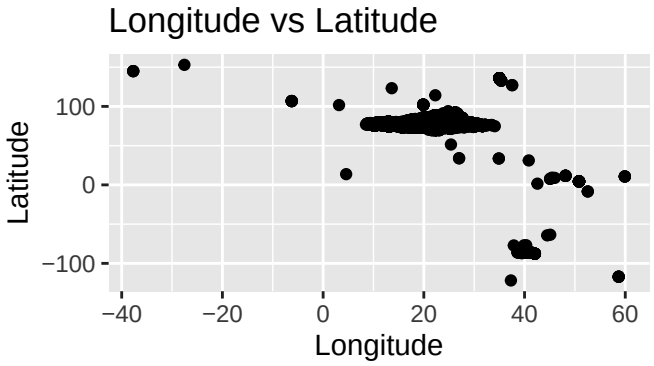

RtM vs Latitude
Latitude
RtM
100
0
−100
0.00
0.25
0.50
0.75
1.00

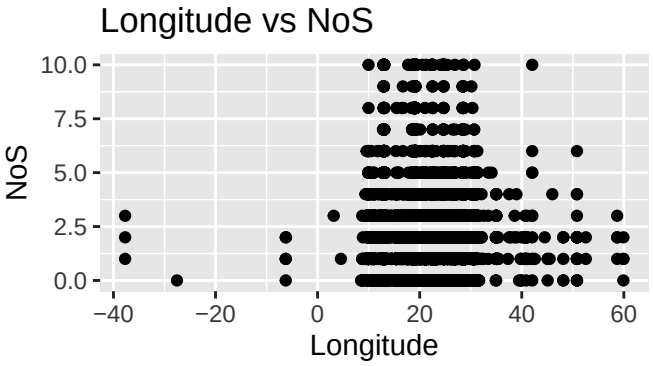

RtM vs NoS
NoS
RtM
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00

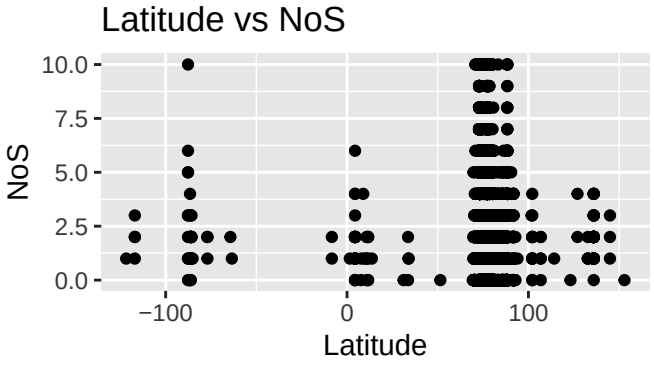

RoN vs Longitude
Longitude
RoN
60
40
20
0
−20
−40
0.00
0.25
0.50
0.75
1.00


RoN vs Latitude
Latitude
RoN
100
0
−100
0.00
0.25
0.50
0.75
1.00

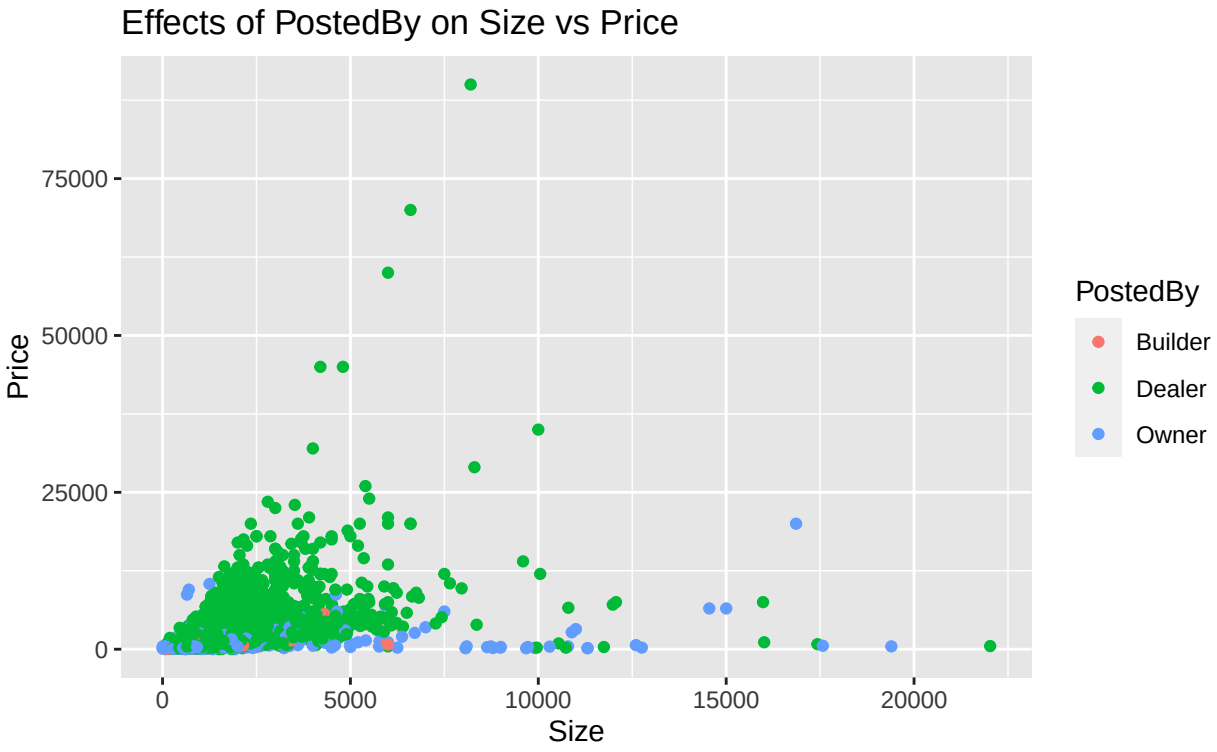

RoN vs NoS
NoS
RoN
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00

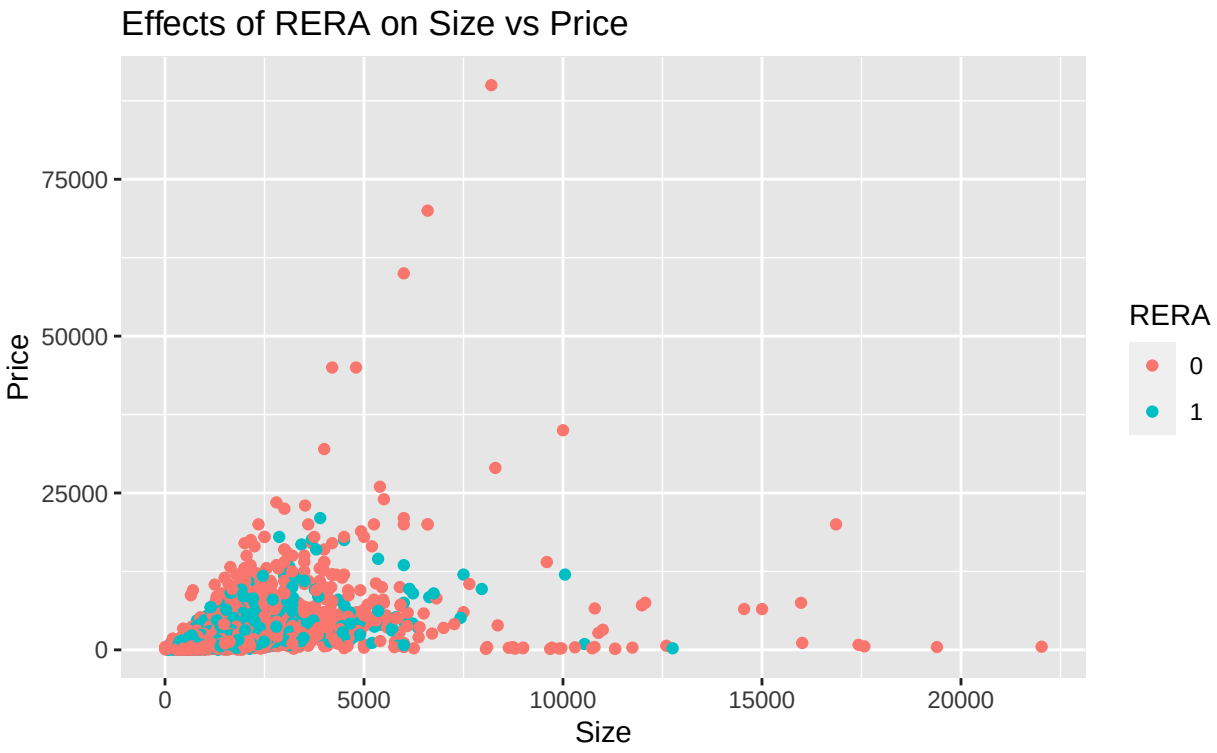

Longitude vs Latitude
Latitude
Longitude
100
0
−100
−40
−20
0
20
40
60

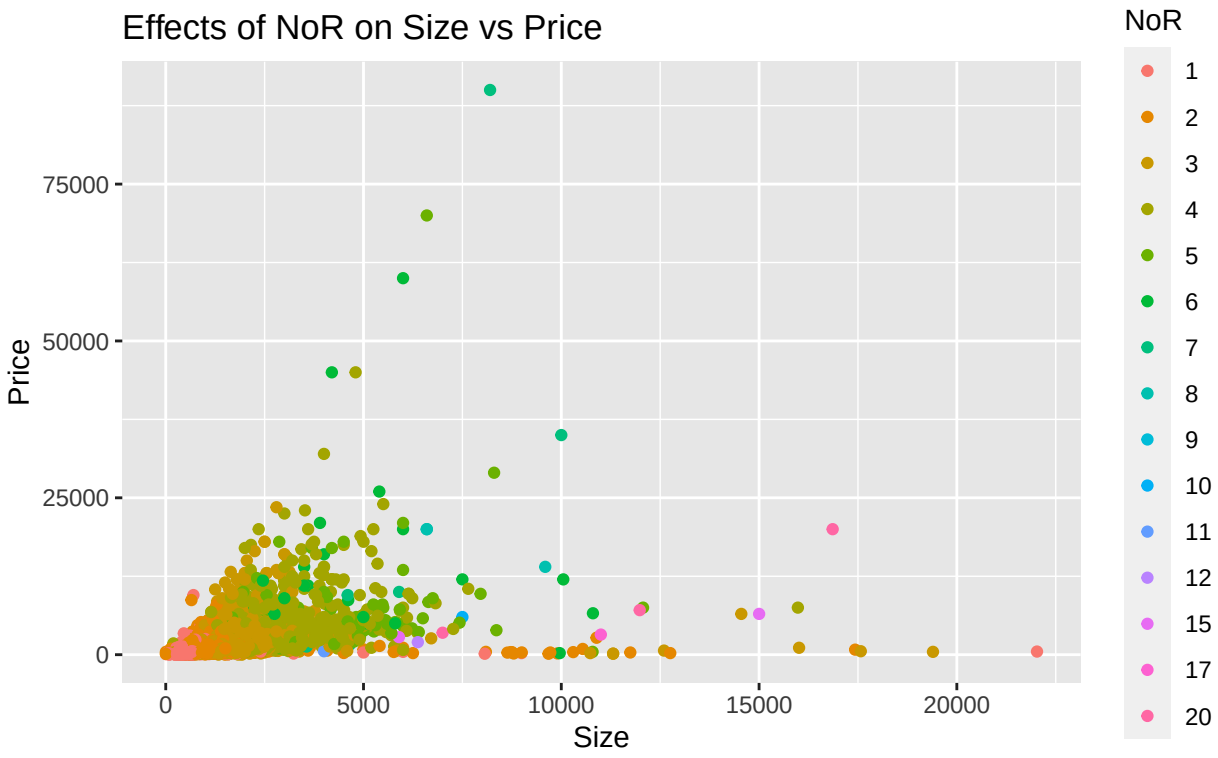

Longitude vs NoS
NoS
Longitude
10.0
7.5
5.0
2.5
0.0
−40
−20
0
20
40
60

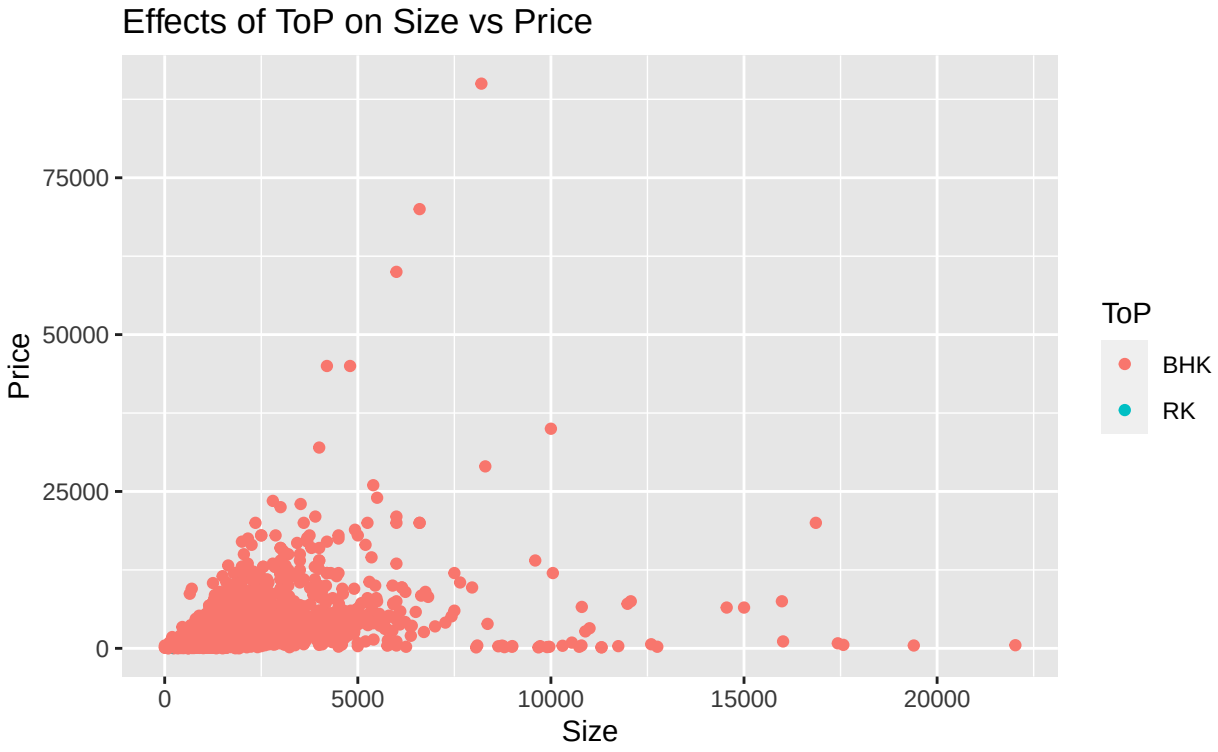

Latitude vs NoS
NoS
Latitude
10.0
7.5
5.0
2.5
0.0
−100
0
100

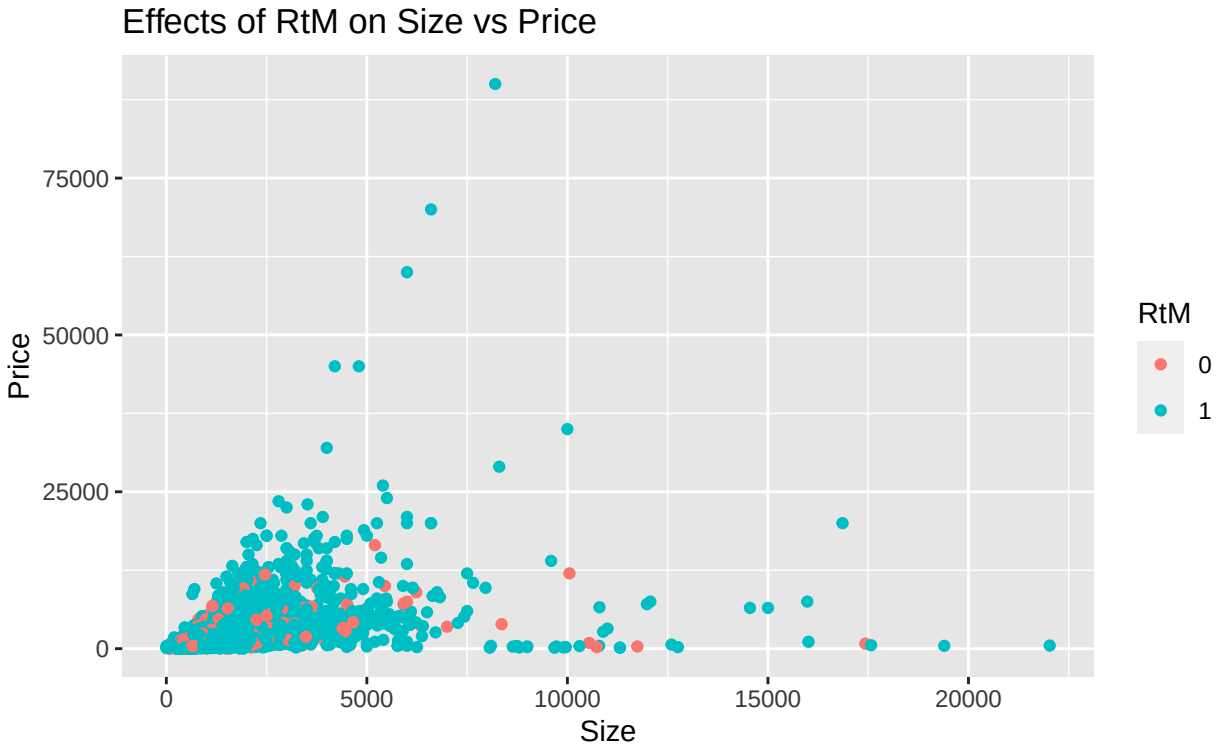
Effects of house characteristics on size and price

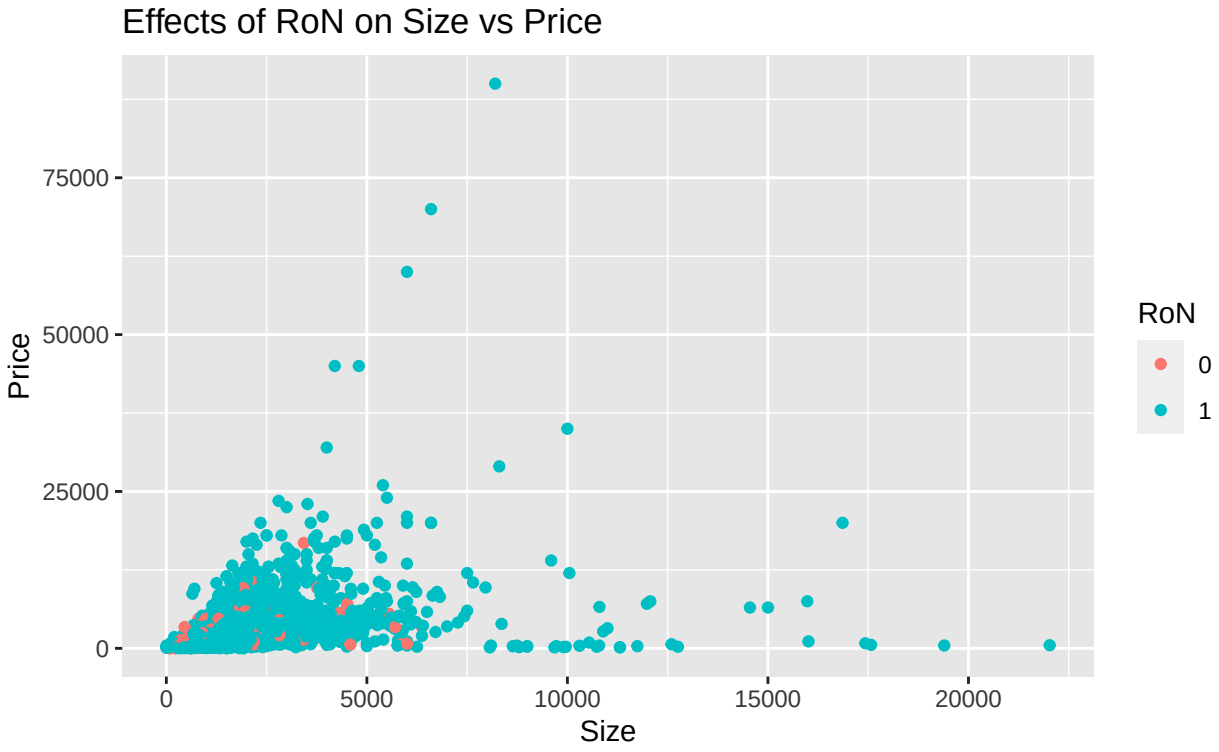


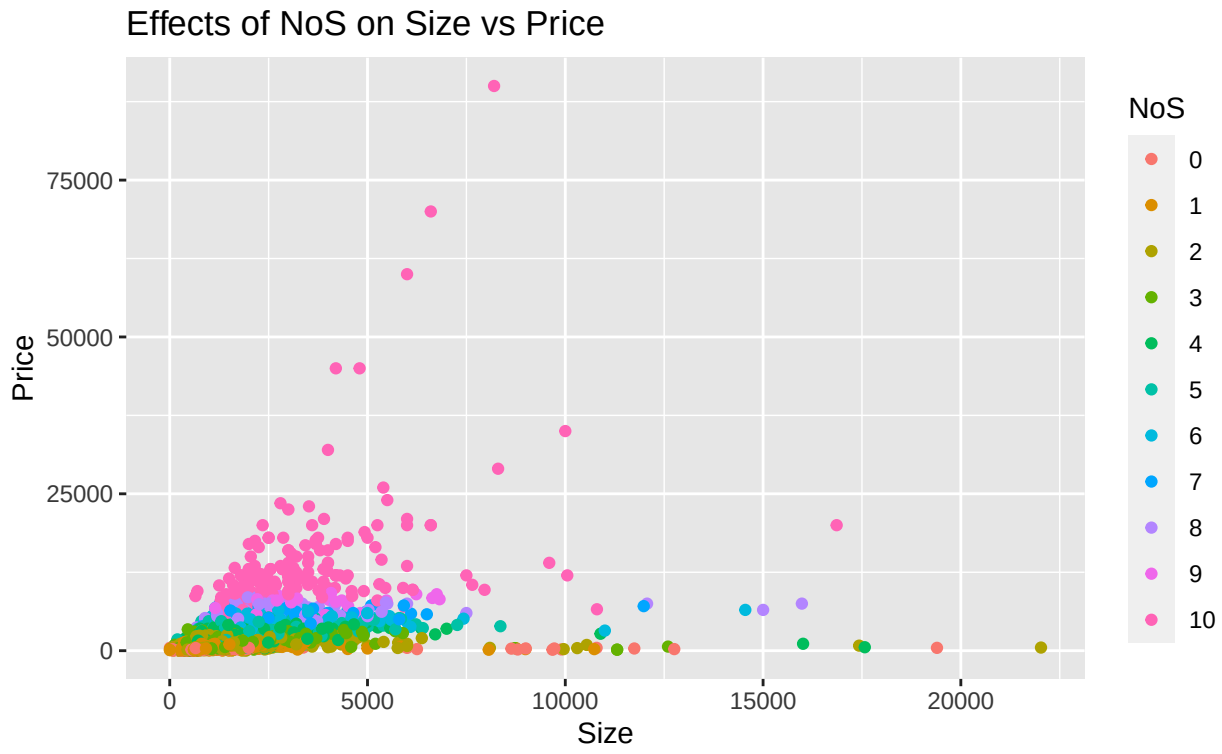



Effects of NoR on Size vs Price
Price
75000
50000
25000
0
0
5000
10000
15000
20000
Size
NoR
1
2
3
4
5
6
7
8
9
10
11
12
15
17
20


Effects of ToP on Size vs Price
Price
75000
50000
25000
0
0
5000
10000
15000
20000
Size
ToP
BHK
RK


Effects of RtM on Size vs Price
Price
75000
50000
25000
0
0
5000
10000
15000
20000
Size
RtM
0
1


Effects of RoN on Size vs Price
Price
75000
50000
25000
0
0
5000
10000
15000
20000
Size
RoN
0
1


Effects of NoS on Size vs Price
Price
75000
50000
25000
0
0
5000
10000
15000
20000
Size
NoS
0
1
2
3
4
5
6
7
8
9
10