House price in Melbourne analysis

Yafeng Zhou

12/4/2019

Introduction

This report is an analysis of the overall situation of house prices in Melbounrne. In this analysis, i will analyze the effects of diffrent variables on housing prrices, such as bedrooms number, cap praking spaces for each property. In order to get the factors that influence the housing price, this article will focus on the correlation between diffrent variables.

The dataset this article use is from the kaggle. The dataset introduces a variety of important information about Melbourne's real estate market up to June 2018.

Some important varibles in the dataset:

-Suburb: Suburb

-Address: Address

-Rooms: Number of rooms

-Bedroom2 : Scraped # of Bedrooms

-Bathroom: Number of Bathrooms

-Car: Number of carspots

-Landsize: Land Size in Metres

-BuildingArea: Building Size in Metres

-YearBuilt: Year the house was built

-Lattitude: Self explanitory

-Longtitude: Self explanitory

Firstly, we read the csv file that we need.

The package we mainly use in this data analysis

-dplyr

-tidyverse

-ggplot2

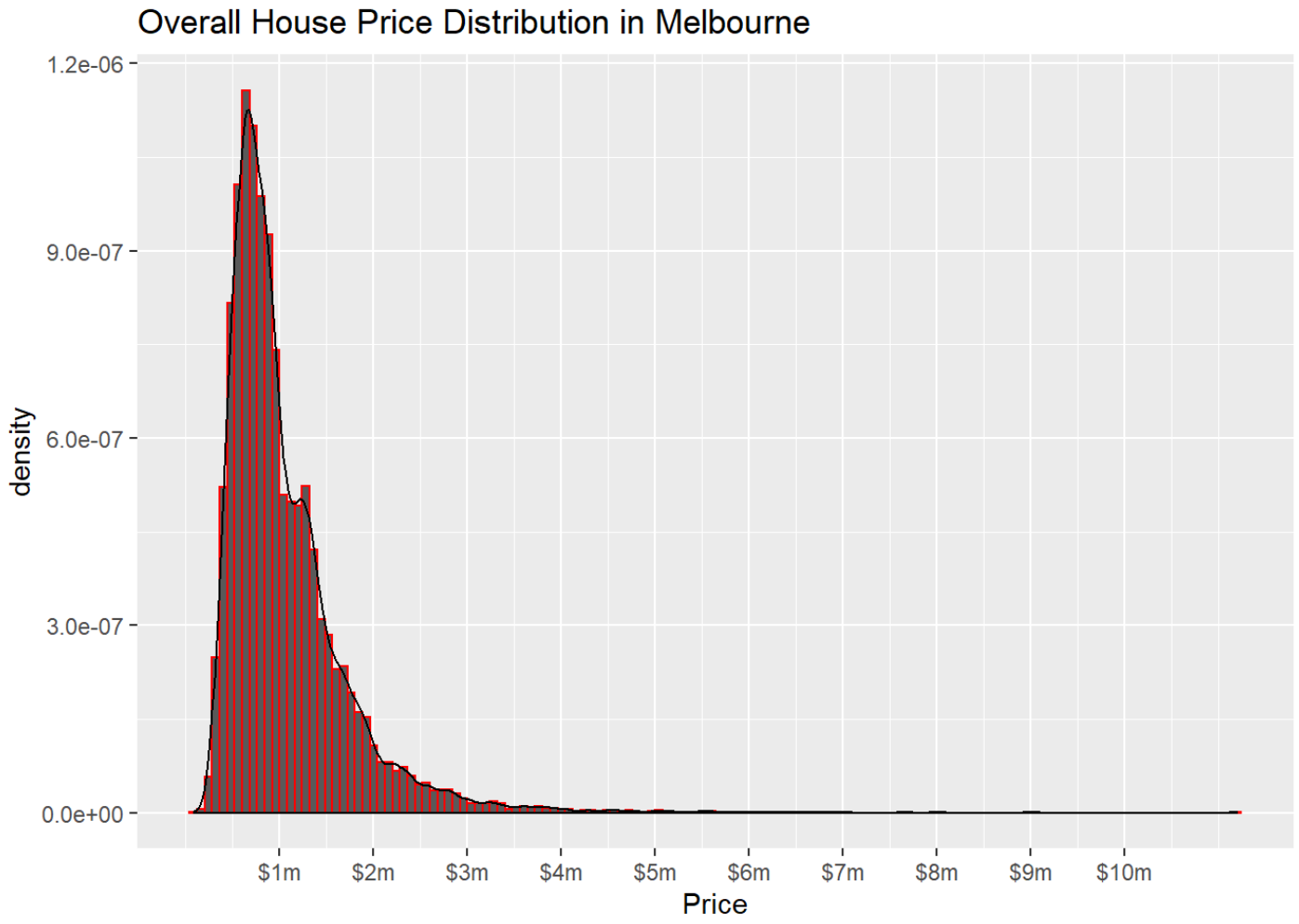
-corrplot

-gridExtra

---

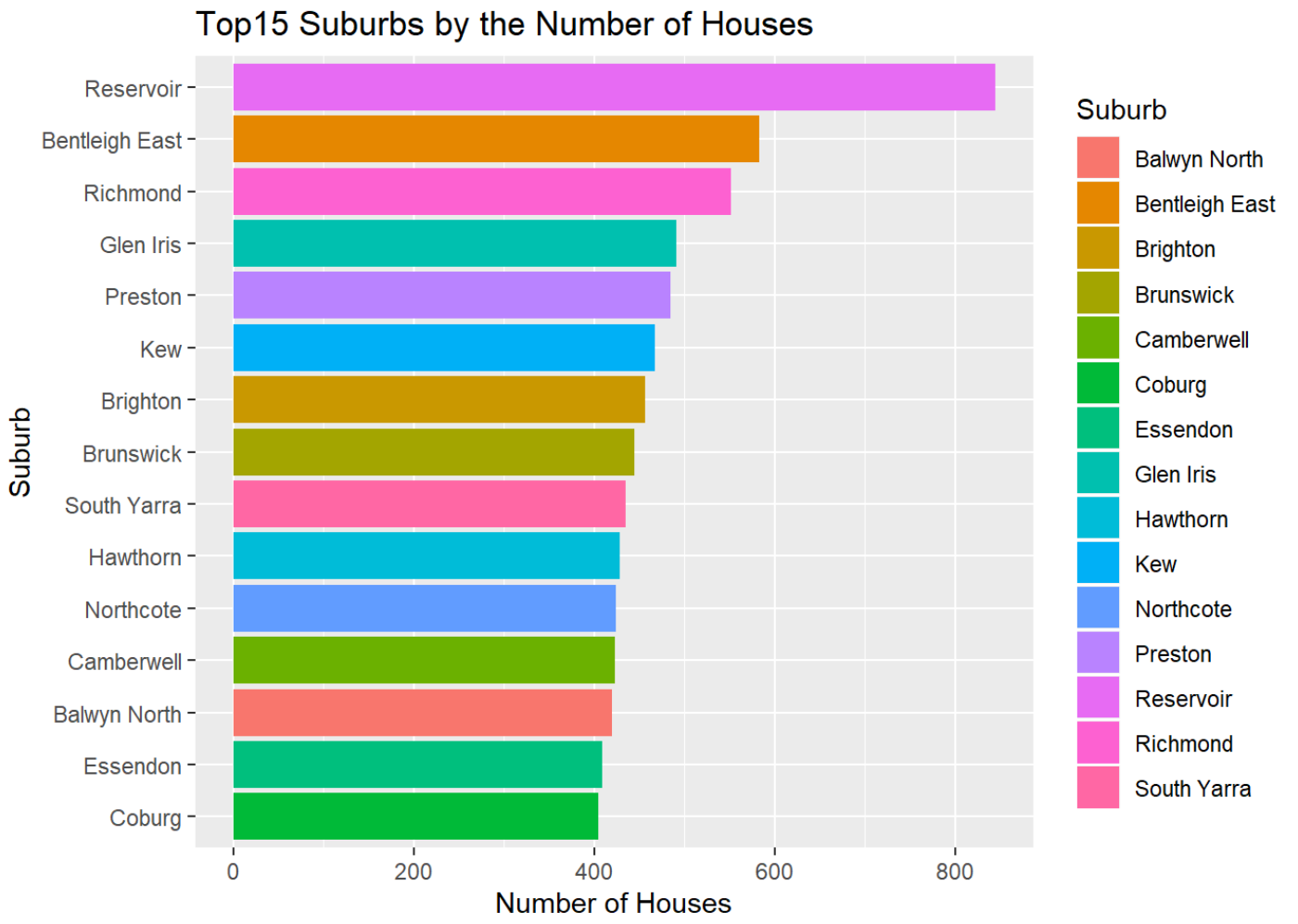
gg-freqpoly

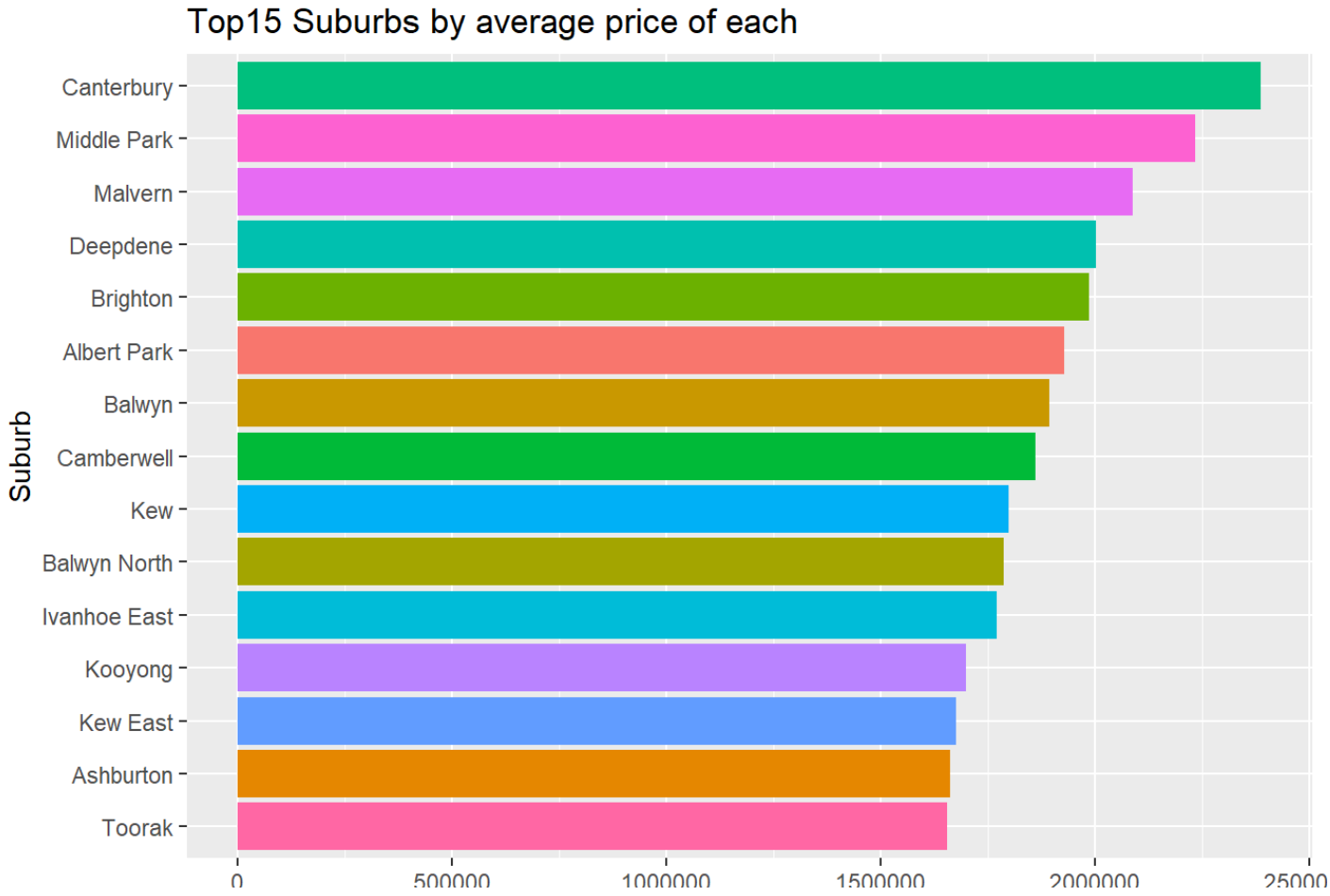
Let we overview the house price distribution in Melbourne.



From the graph, we could see that more than half houses in Melbourne is less than $1millon. But the distribution of house prices in Melbourne is very wide , from $0.25m to more than $15m.

Let we see the Top 15 areas which have the most number of houses. Then, we will compare it with the Top 15 areas which the price are the most expensive.

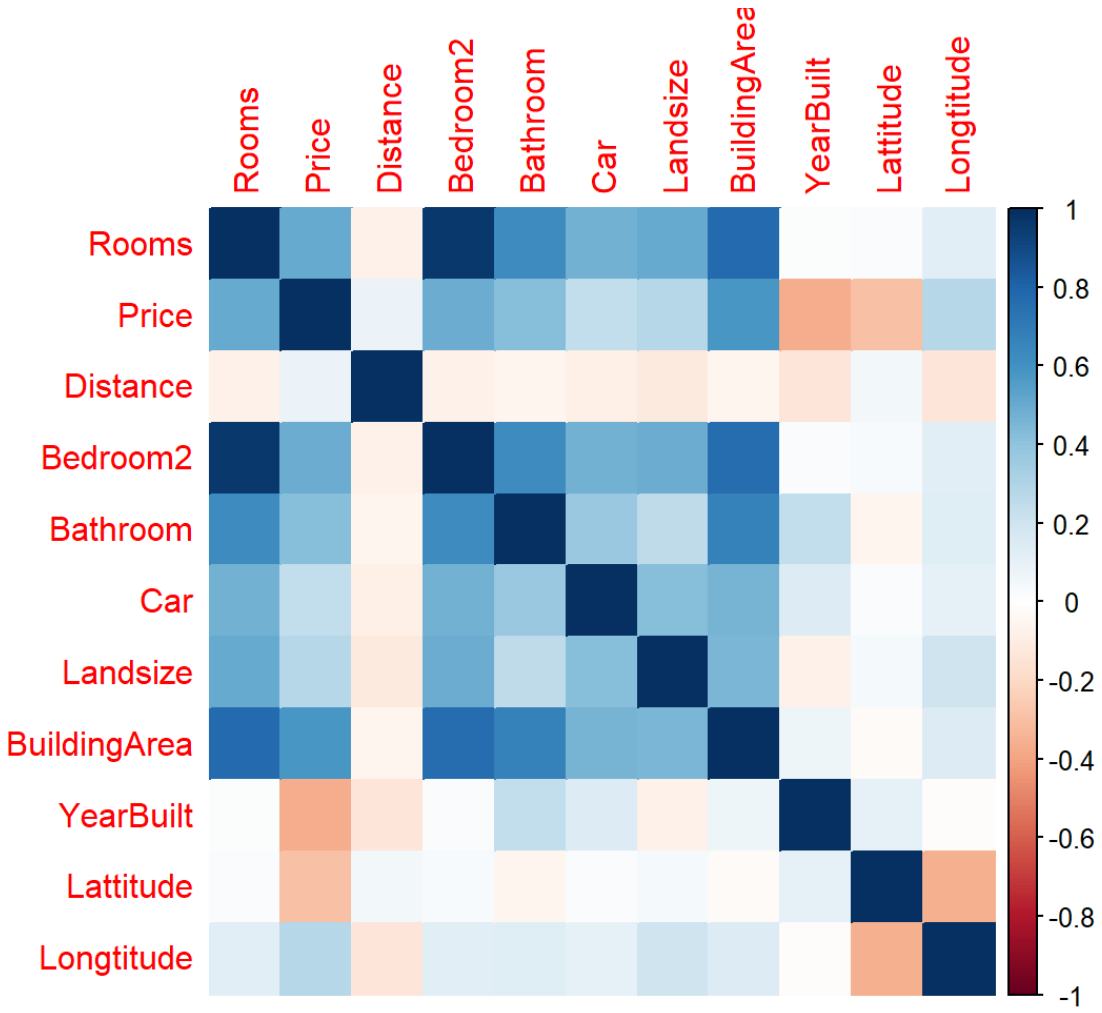

Top15 Suburbs by the Number of Houses
Reservoir
Bentleigh East
Richmond
Glen Iris
Preston
Kew
Brighton
Brunswick
South Yarra
Hawthorn
Northcote
Camberwell
Balwyn North
Essendon
Coburg
Suburb
0
200
400
600
800
Number of Houses
Suburb
Balwyn North
Bentleigh East
Brighton
Brunswick
Camberwell
Coburg
Essendon
Glen Iris
Hawthorn
Kew
Northcote
Preston
Reservoir
Richmond
South Yarra


Top15 Suburbs by average price of each
Canterbury
Middle Park
Malvern
Deepdene
Brighton
Albert Park
Balwyn
Camberwell
Kew
Balwyn North
Ivanhoe East
Kooyong
Kew East
Ashburton
Toorak
Suburb
0
500000
1000000
1500000
2000000
25000

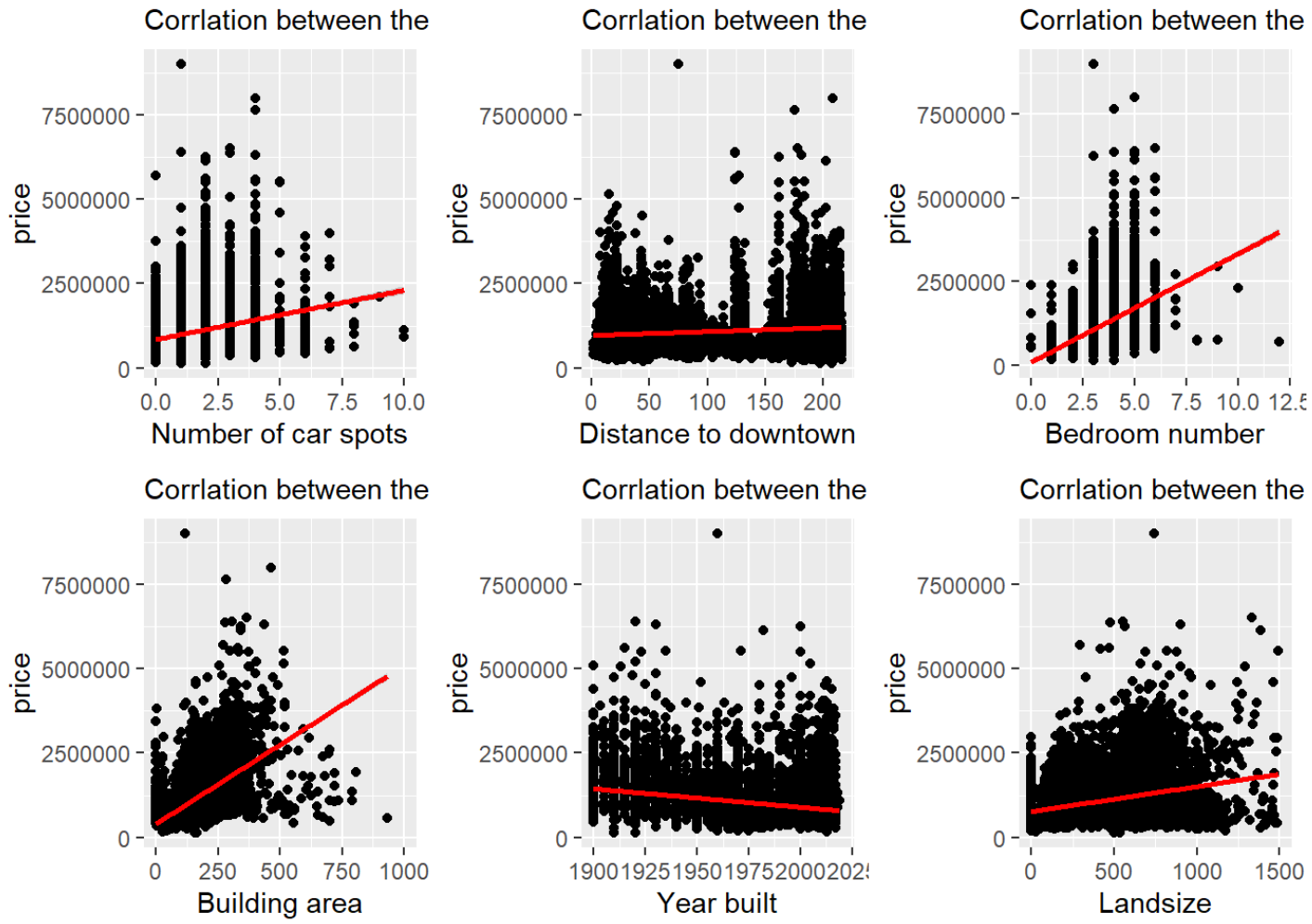
Average Price of House

From the graph, we clearly see that the area “Reservoir” has the most number of houses, following by th eare “Bentleigh East”. The least number of houses area is the “Coburg” among the Suburbs above. The graph “Top 15 Suburbs by average price” follow. From decreasing order, the Centerbury has the largest mean value, which is more than $2.25m. The Suburb with the lowest mean price among the top 15 is “Toorak”, attaching the mean more than $0.5m.

After knowing the overall distribution and ranking of top 15 in Melbourne, let we start to analysis the correlation in different factors that affect the housing prices in Merbourne.



From the correlation matrix above, we would clearly see that only the factor, “YearBuilt” and “Lattitude” have a negative relationship between house price. Because Melbourne is located at the southern hemishere, so the houses with the large lattitude may sell at a lower price. Therefore, we may defer that the southern part of Melbourne might be relatively lower. It is worth noting that floor area has the most positive correlation with the house prices, followed by the factors “Number of rooms” and “Number of bedrooms”. It is also could be seen that the distance to downtown is also has a slighly correlation with house price. For this part, i consider that the Number of car spots of the houses might be a major factor that effect house price.



From the six correlation plot that generated above, we could get more detailed information that different factors between the different factors with the house price. The individual graph of each factors are clearer reponse the correlation than the correlation matrix. We could clear see that there are some very obvious positive relations among housing prices and Building are a and also the landsize. It makes sense because in the case of approxiamtely same unit price, larger building area means more expensive.And with the larger land size the building area may be bigger. Although we could see some outlier points which are far fromthe cluster of data.

We could also observe from the graph:

-1.The bedroom numbers have a positive relationship with the house price.

-2.Number of carspots gets a slight positive correlation with the house price.

-3.The distance to downtown does not have a clear relationship with house price.

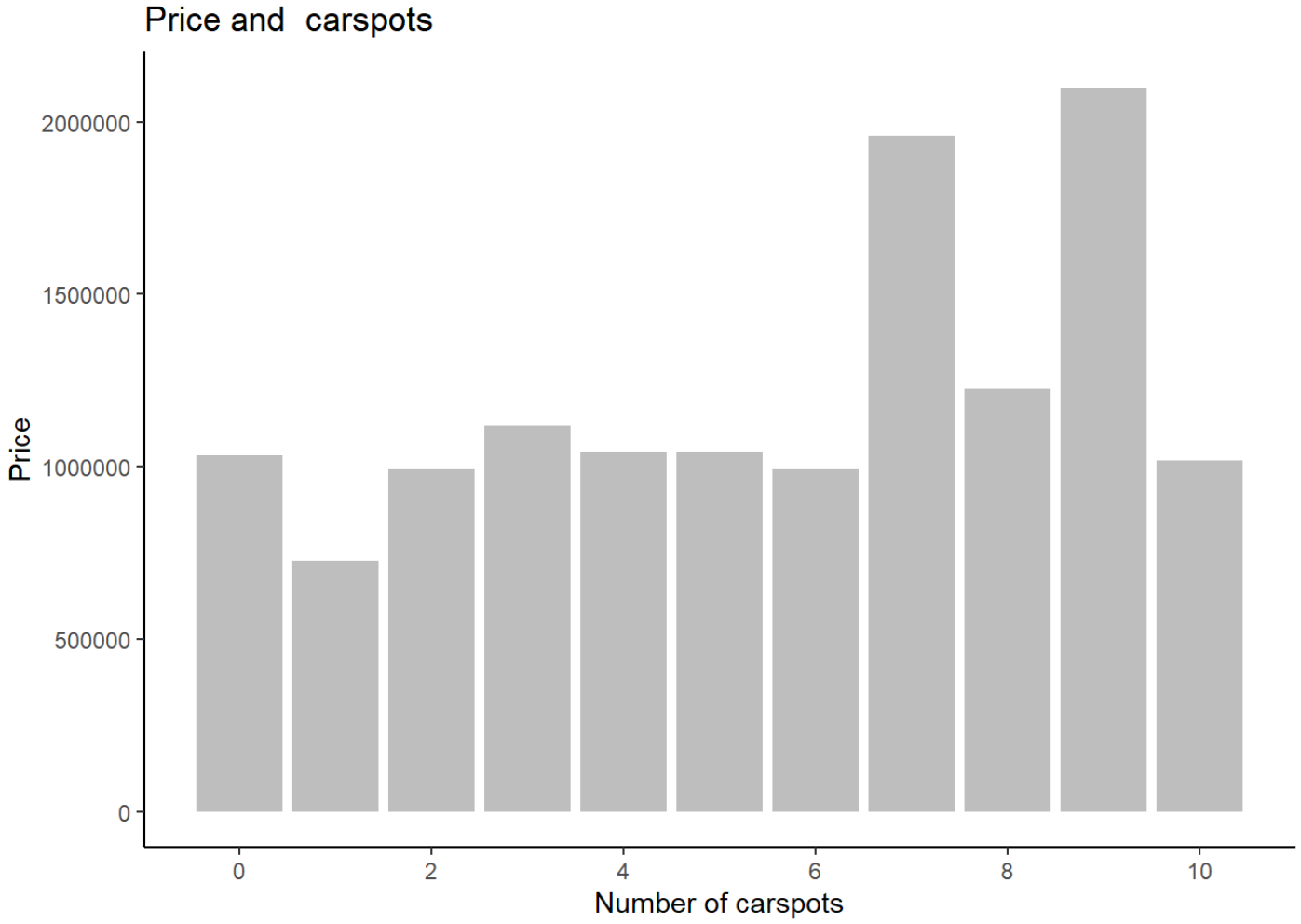
Home variables (carspots, rooms, bedrooms) with price analysis.

After the correlation analysis above, we need to get a more specific relationship between different of house's situations.

- rooms

- bedrooms
- Car spot

How does the number of car spots influence the price of house?

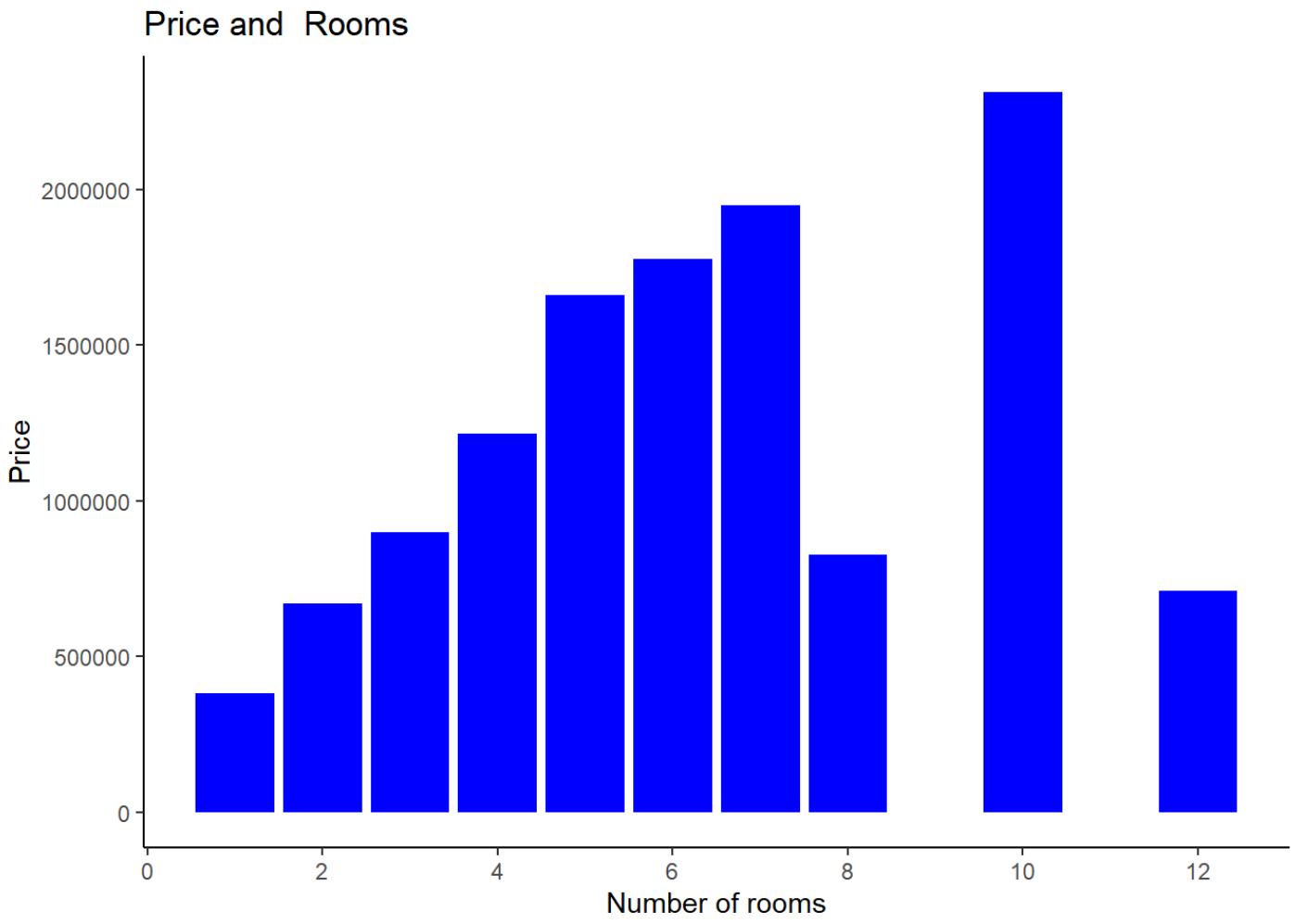


By the bar graph above we could conclude that

-1.The houses with 7,9 carspots get the highest average price.

-2.The remaining number of garages does not have much effect on housing price.

Does number of rooms affect house price?



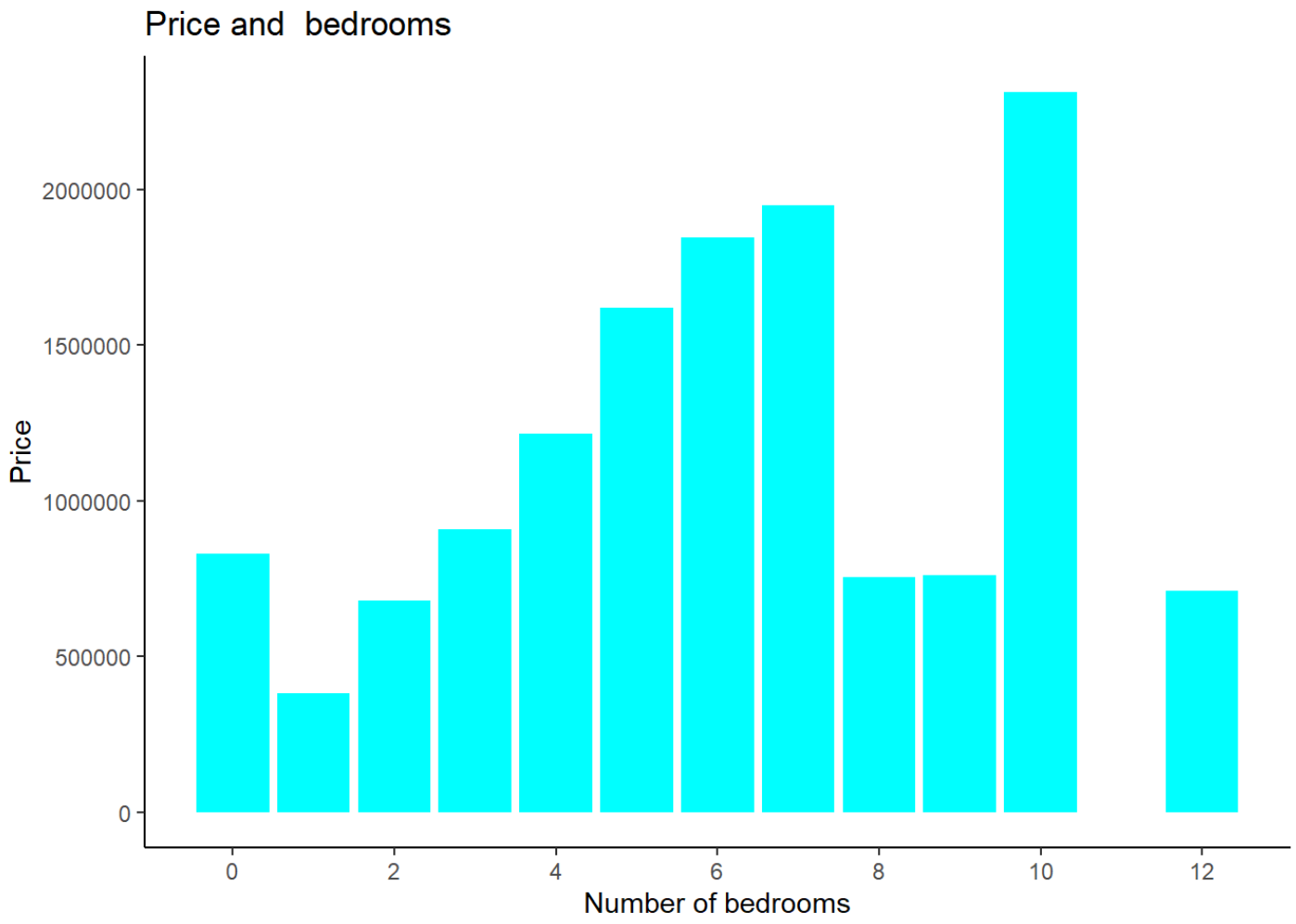
From the bargraph above, we could conclude that:

-1.In houses with 1-7 roomes, prices tend to increase with the number of rooms.

-2.The average price of a house with 7-10 rooms fell at a significant interval.

-3.The number of rooms in houses do affect house price.

What about number of bedrooms?



According to the bar graph above we could figure out that:

-1.It has a slightly decilining between the house with 1 bedrooms and house with 2 bedrooms.

-2.The overrall trend of the house price are increasing with the number of bed rooms

-3.The number of bedrooms are positive influencing the house price.

Correlation between Rooms and Bedroom Correlation between Rooms and car spots

```
## The correlation between the rooms and bedrooms:  0.9644817
##  The correlation between the rooms and the car parking spots: 0.4012512
```

From the correlation between the number of rooms and the number of bedrooms is nearly 1, we may defer that more rooms in a house leads to more bedrooms in the house, which leads to a higher price for the house itself.

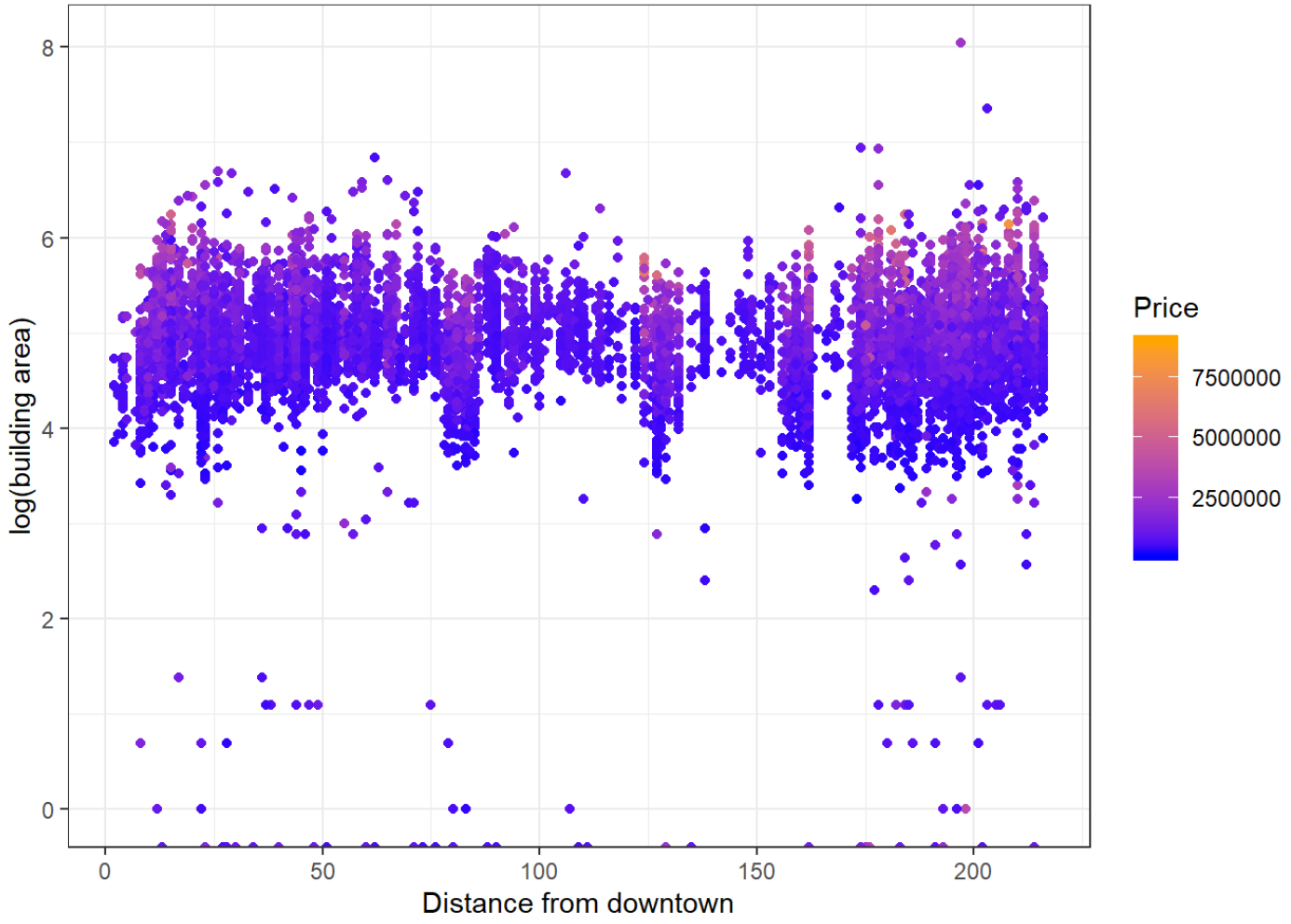
On the other hand, the relationship between the number of parking spaces and the number of rooms is not as obvious as the former. This may be why parking spaces do not significantly affect housing prices.

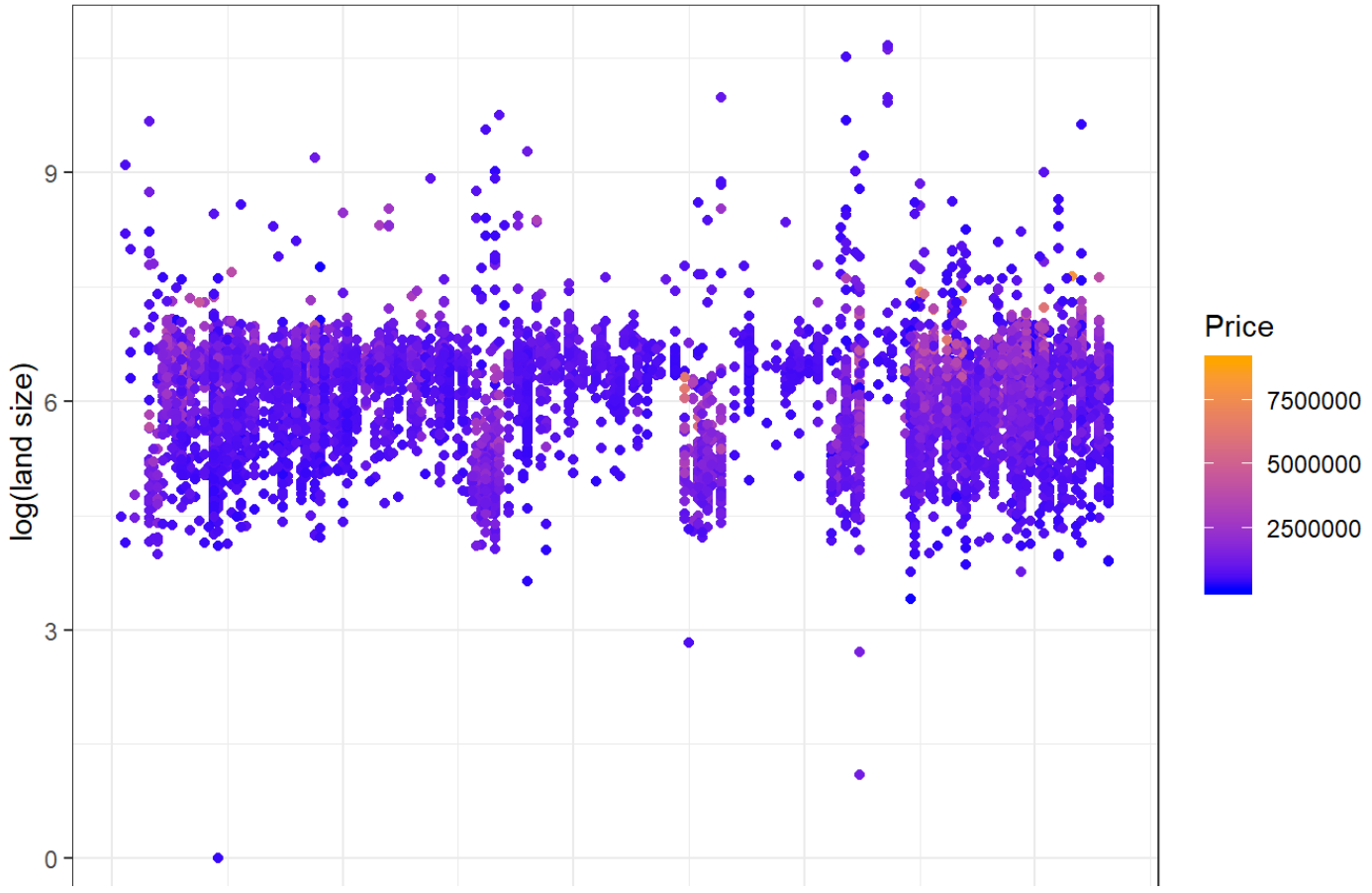
Is location really not the determinat of housing prices?

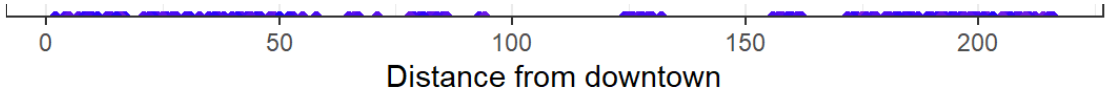
From the analysis above, we can not get a clearly conclusion between the distance to downtown and house price

We may assume that the houses' landsize and building area may be affected by their distance to downtown.

To show more the relationship clearly, we need to take the log value of the 2 variables







The graph show that:

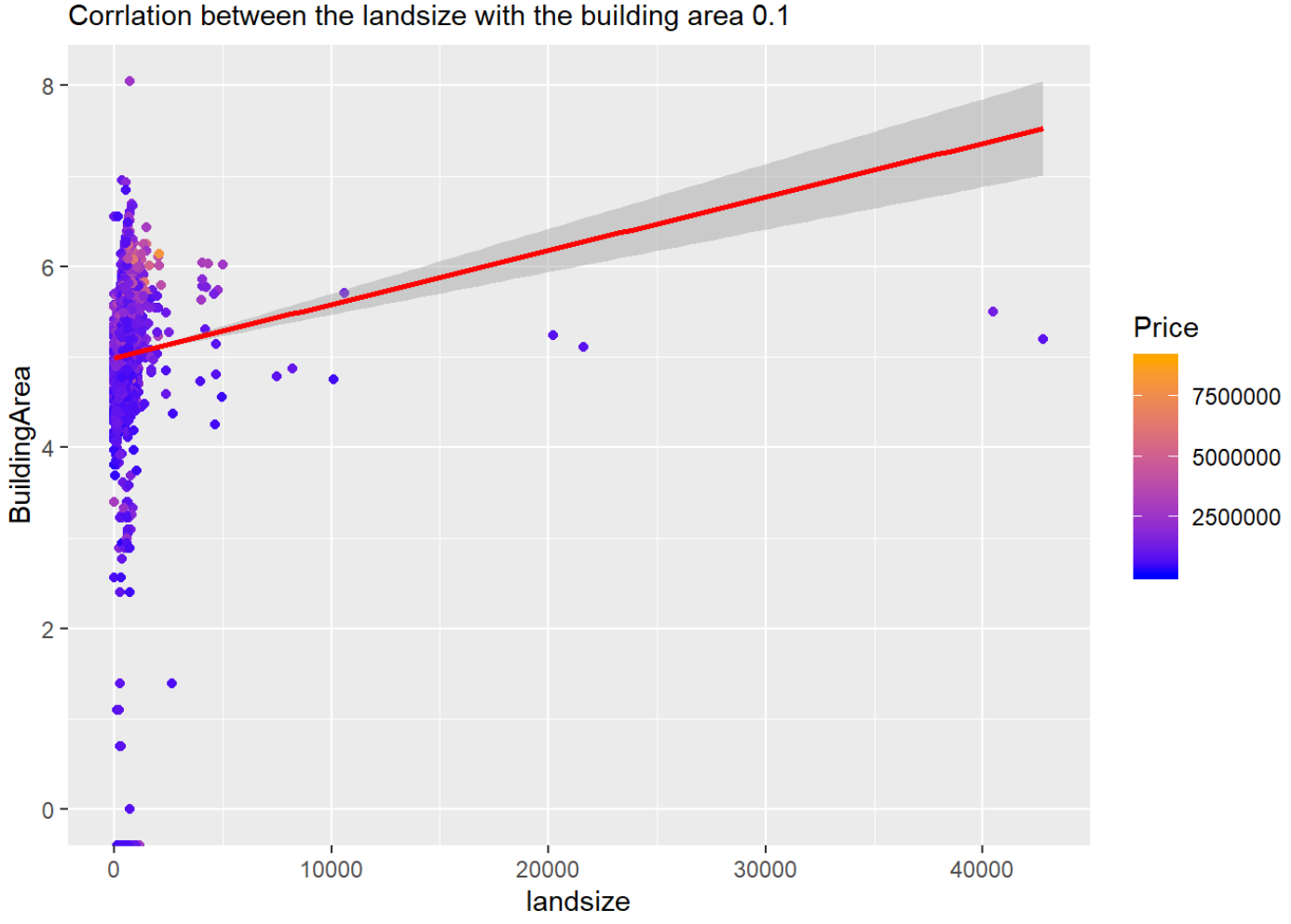
-1.In general more floor space means higher prices. However the larger of the landsize does not mean the higher price of the house

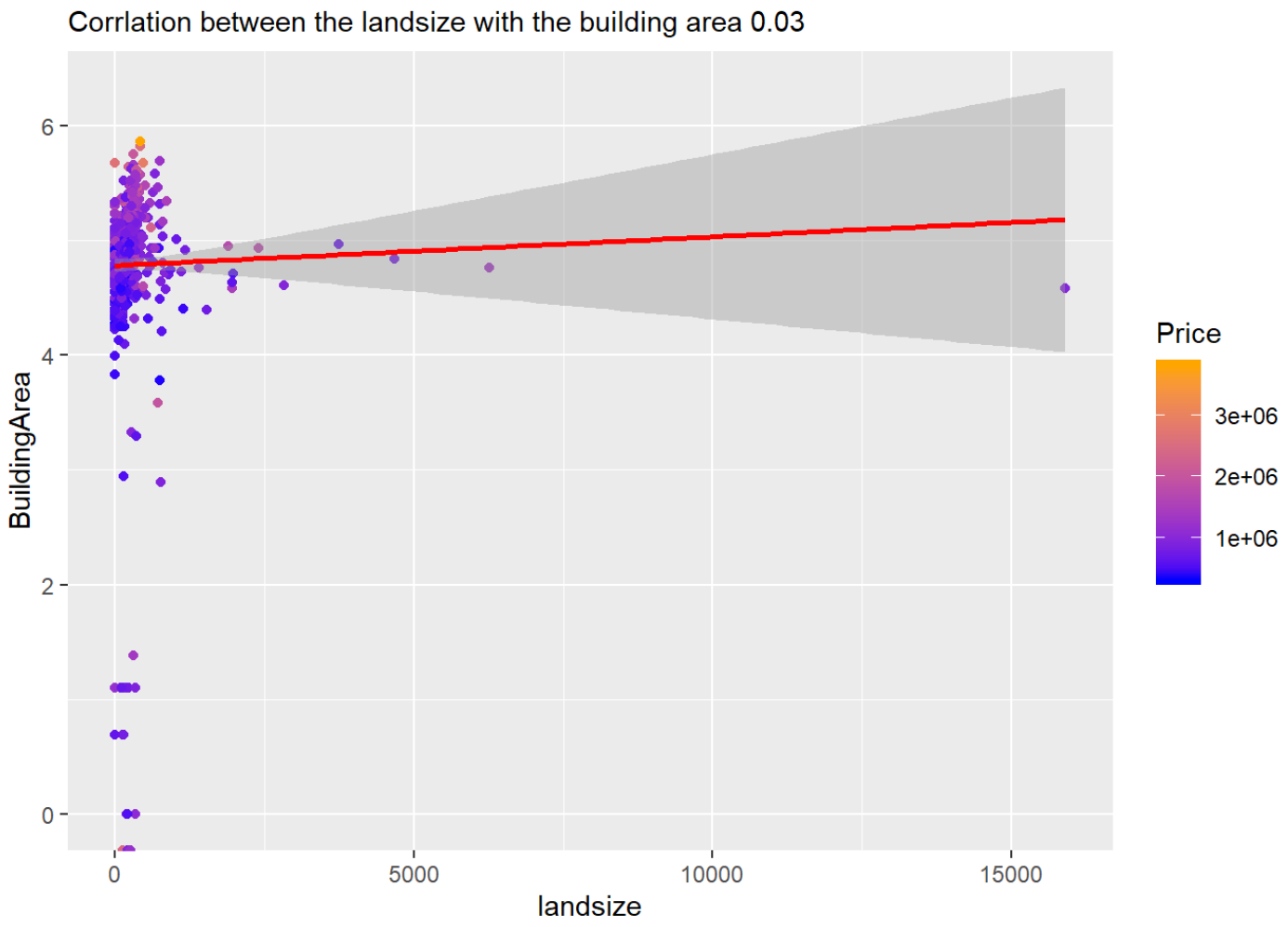
-2.Diffrent in the distance of houses from downtown have similar price distribution.

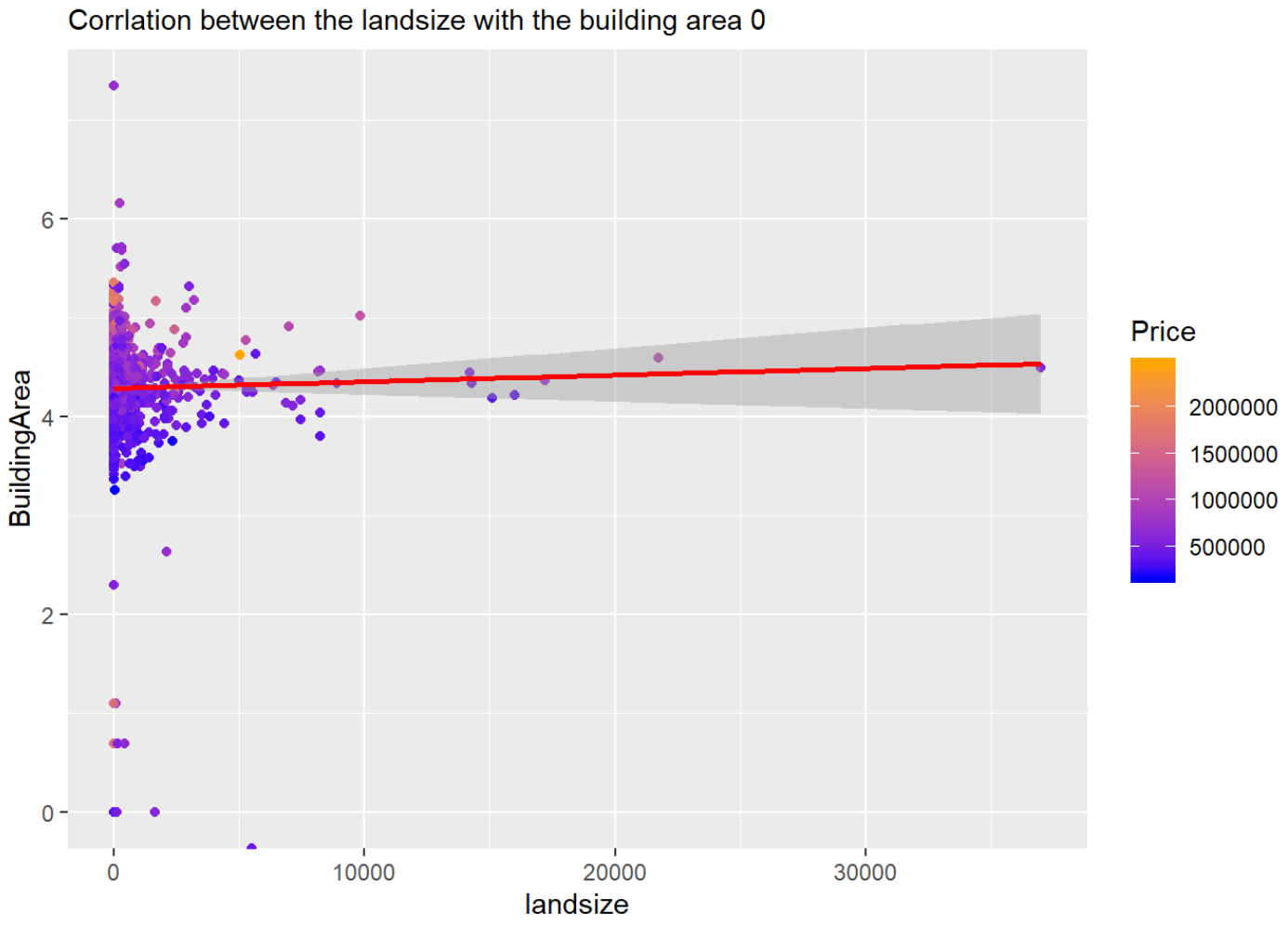
-3.It is worth noting that houses of the same building area and landsize seem to cost more than those closer to the city center, especially for the house more than 150km to downtown.

Besides the building area, we notice that the other 2 important factors are the “Landsize” and building types.

To determine how landsize a affect the building area in different building type, we will plot the correlation between different building types and show the gradient of the price under this situation.




Corrlation between the landsize with the building area 0.03
BuildingArea
landsize
Price
3e+06
2e+06
1e+06
0
2
4
6
0
5000
10000
15000



From 3 facet graph above, we could conclude that:

1.In the type of “house”, we could once again proce that more floor space leads to higher house prices.

2.The correlation between the landsize and building area seems like not clearly.

3.The “unit” type seems have only small variance in the building area, whatever the landsize. The possible reason is that the structure of the apartment is fixed and the difference in the units area is not large.

Conclusion

In the Melbourne housing market:

1. Building area is the main factor affecting building prices, especially in “house” type.
2. More rooms that a building has, the more expensive it will be.
3. The number of parking spaces hardly affects the price of a building.
4. Shorter distance to downtown does not mean higher house price, some regions which are far away from dowmtown also with reletive high price.

What need to be improved in this analysis?

Not get a clear relation between the distance to downtown and the housing prices.

Since the distance to the city center is not the only factor that affects housing prices, we need more infomation, such as whether the building in a convenient area. Is there a good school nearby? These are important factors in determine housing prices.

After all above information being collected, a more reliable conclusion how the location affects house price, can be drawn.