Recent Sold House Data in San Jose, Scraping from Realtor.com

Objective

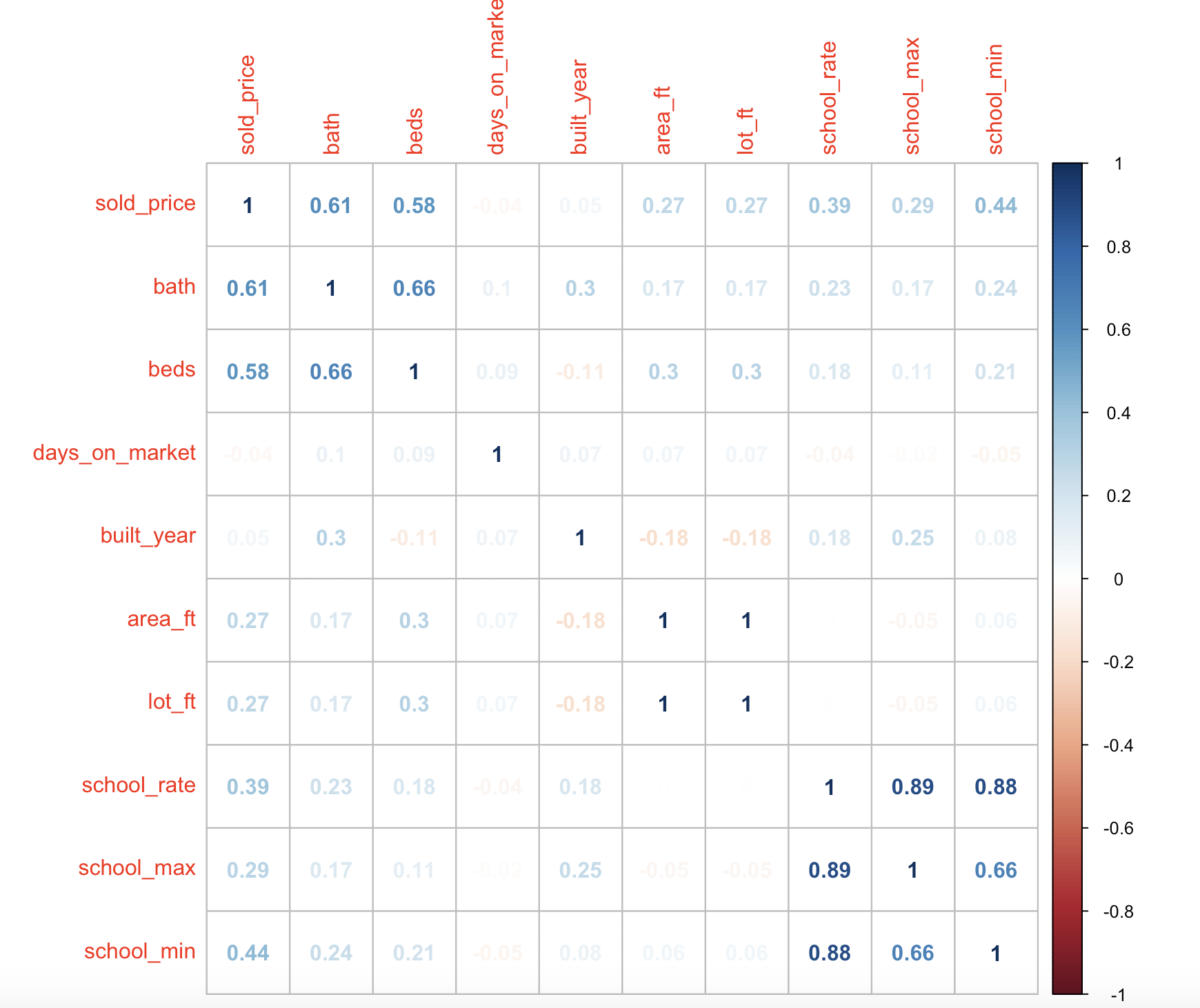
House marketing is going crazy these years, especial in the Bay area. It costs at least 1M to buy a house in the bay area. As a new buyer, it’s important to know which house is worth buying. San Jose is a typical city in the bar area. it can represent more cities in the bay area. The objective of this project is knowing how the house marketing in the San Jose perform. And provide suggestions to the new buyer depends on their requirements.

Method

On the Realtor.com, it shows most recent sold house in the San Jose. It also shows variety data for the house, like the sold price, house size, house type and so on. With web scraping, we are able to get those data by go through each house’s link and transfer into a large database. Then more analysis can be done with those databases.

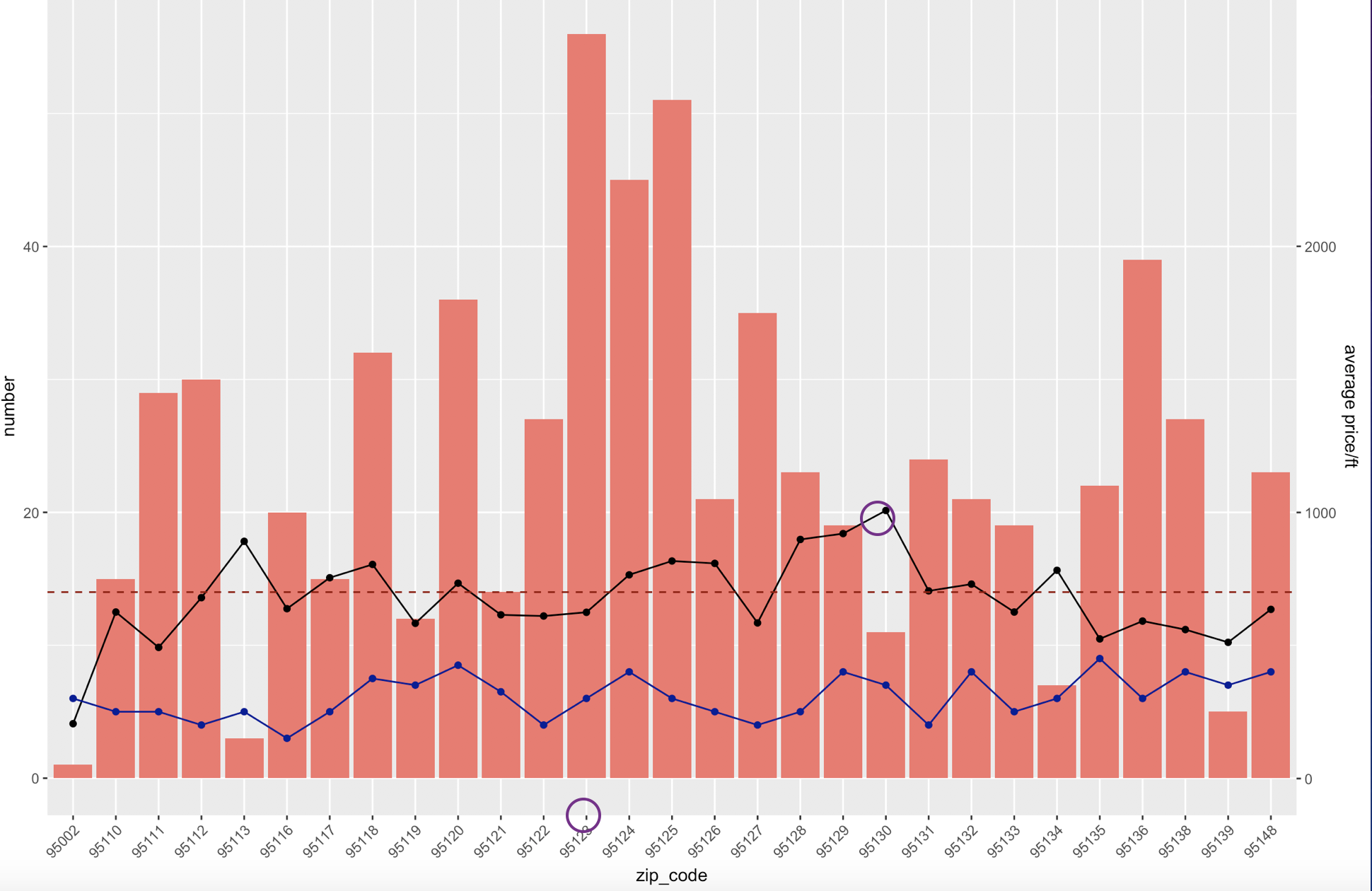
Analysis

First question I want to answer is, which factors affect the price of house. A correlation plot shows how the factors related to the sold price.



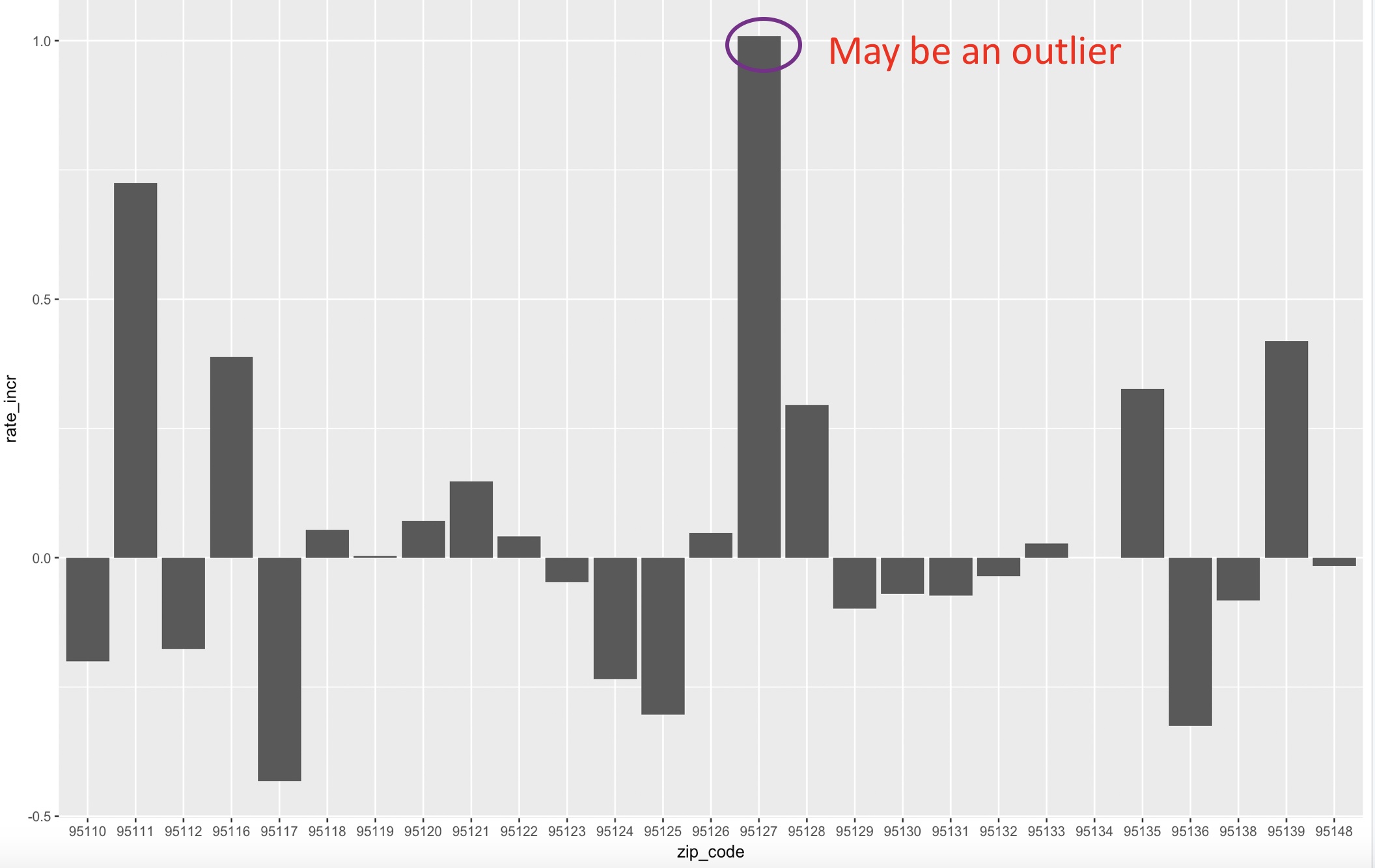
From the correlation plot, we can clearly see that the house structure (bath & bath) and school rate drive high the house price most, with correlation rate around 0.5. What’s more, we can also find that built year is negative correlate to the beds, area ft and lot ft. Which means old house has more area and better structure. So an old house with good school rate with lead to a high price, better structure house.

After knowing which factors affect the house price most, I also want to know how sales different in different area. This can be analysis by showing how the average price/ft and house sales number across among the zip \_code.



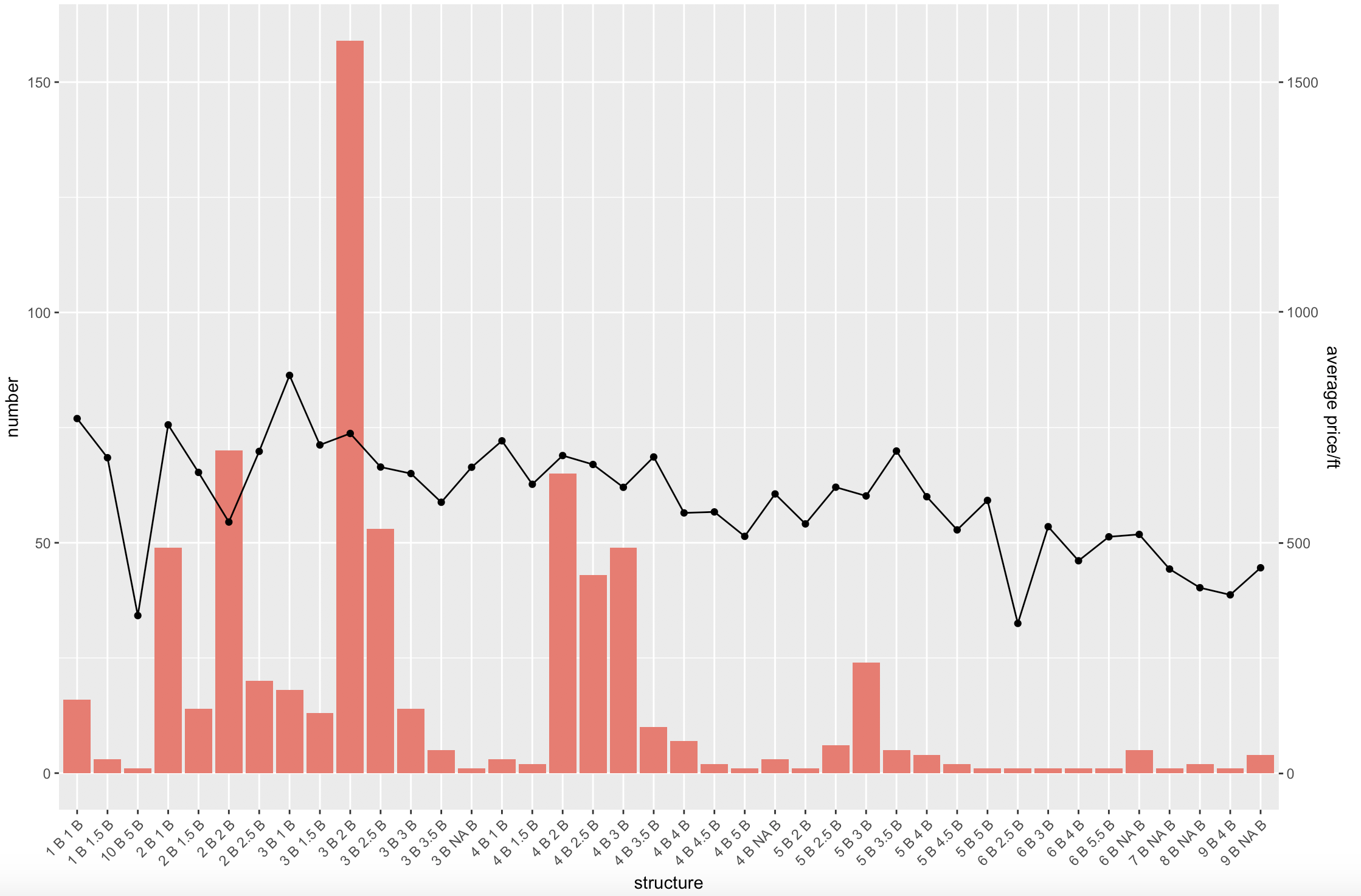
The bar chart above shows how many house sold in different zip code. The black line is the average price/ft and blue line is the average school rate. The highest average price/ft is in area 95130, which most schools have a rate of 7 or 8. And the black line and blue line have almost same trend. It again prove that school rate affect price. Meanwhile, the 95123 has highest sold house number. It has an average price/ft and average school rate. The average price/ft is around $700/ft.

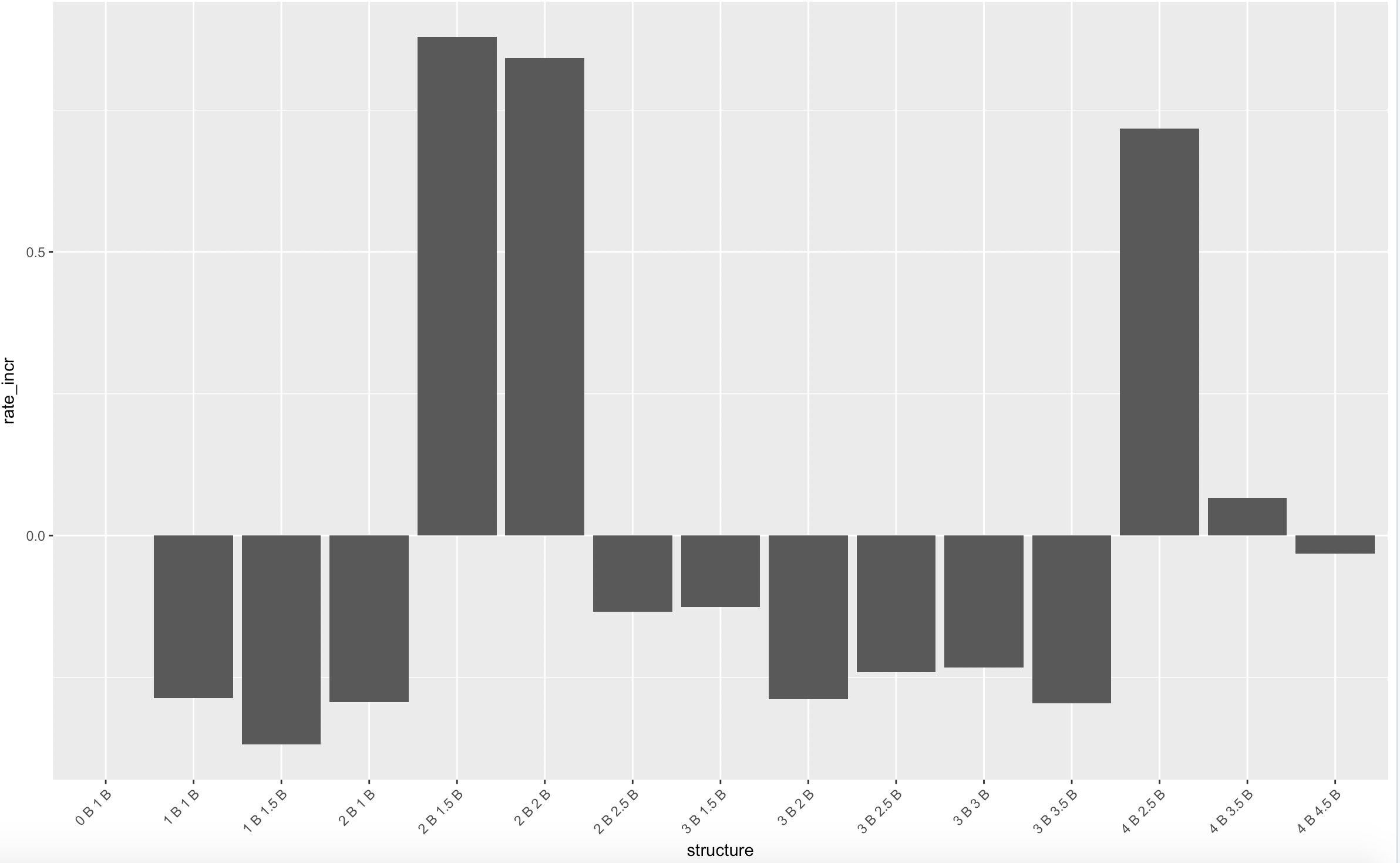
I believe buyers are more concern with how the rate increase in different area. Below is a plot to show how the single family house rate increase in different zip code. There is a point shows an increase rate for 100%, which might be an outlier because it’s too large. Besides that, 95111 has highest increase rate. An area with high rate increase will be a good investment.



We can do the same type analysis with the house structure, below are diagram shows how the average price/ft and sold house number different in different house structure. 3B2B is the most popular house type, and more beds will lead to lower price/ft.

Different then the price/ft, 2 bedrooms house increase most, second is the 4 bedrooms, while other structure house are all has negative increase rate.





Suggestion

To summarize all the analysis below, we can conclude that school rate and house structure affect the house price most, and old houses may be a good choice. For investment purpose, we can also conclude that 95111 and 2 bedrooms house are elements that has highest increase rate. There are more analysis can be done on those data and more data should be collected for more accuracy analysis.