Data Visualization - Assignment 3 - Exploration

Alireza Mostafizi

May 27, 2018

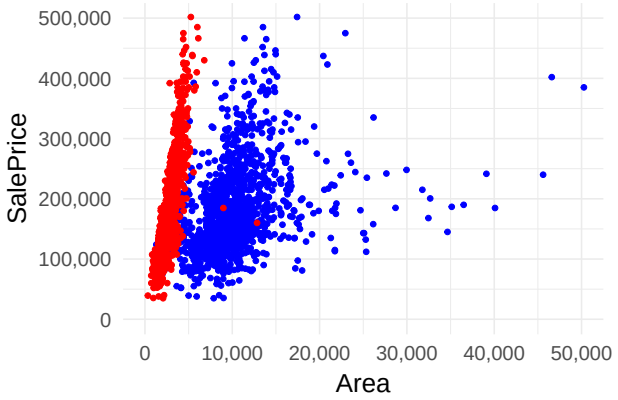
1 Data Description

I used the *House Price* dataset from *Kaggle*, known as *Ames Housing Dataset* compiled by *Dean De Cock*[1]. The data contains 79 features of 1460 houses sold in Ames, Iowa, as well as their prices. The data and its descriptions can be found here[1]. The features cover almost every aspect of a house. However, for the purpose of this assignment, I selected 20 of them that I was suspecting they show interesting findings. The features I chose mainly include the size of the house, number of bedrooms and bathrooms, type and style of the house, the overall quality, built, sold, and remodeled years, and specific conditions of a house.

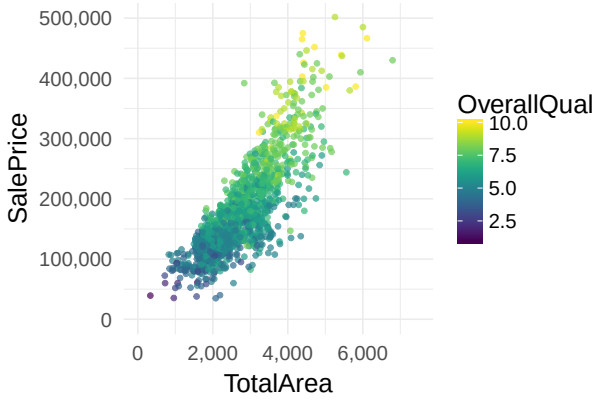
2 Findings

House Size and Price

I investigated the relationship between the price and the *Lot Area* as well as the size of the entire living area which is the summation of the size of all stories, garage, and basement, named *Total Area*. In addition, I concluded that there is a correlation between the size of the house and the overall quality marked for that house.



(a) Price ($) vs. Lot area in blue and Total Area in red (ft^2)



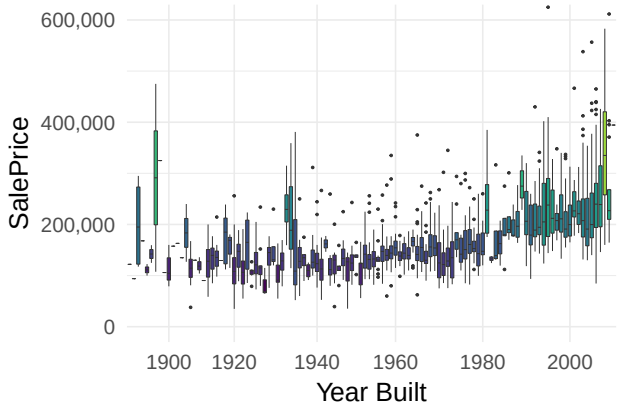
(b) Price ($) vs. Total Area and Overall Quality

Figure 1(a) shows the positive correlation between price and both total and lot area. However, it can be stated that the relationship with the total area of the house is much stronger and more linear. Figure 1(b) emphasizes on this linear trend. It also shows the correlation between overall quality of the house and both size and price. In other words, a bigger house generally is marked with higher quality and it's more expensive.

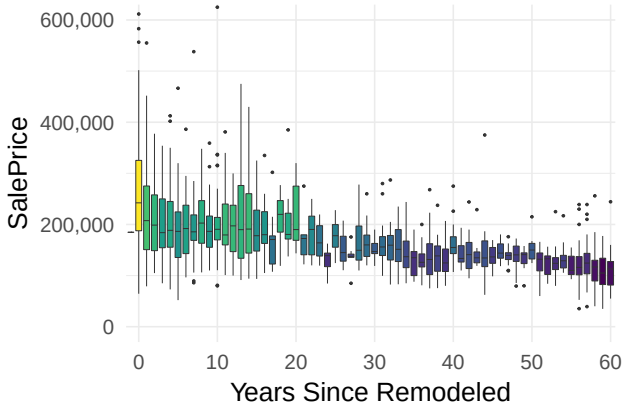
Years Built/Remodeled and Price

I suspected that there might be a relationship between the price and the year a house was built or remodeled. Figure 1(c) shows that the more recent the house has been built, the more expensive it has been sold. We can also see a jump in the prices of the houses built before and after $\tilde{1}990$. Houses built after 1990 has been sold for higher prices. Please note that the color is correlated with the average price of the house, and thus, the lighter the color, the more expensive the average price is. Another trend with the year is seen with the number of years since the house was remodeled. Figure 1(d) shows that the more recent the house was remodeled, the more expensive and with higher variability it was sold.

[1] https://www.kaggle.com/c/house-prices-advanced-regression-techniques



(c) Price ($) vs. Built Year



(d) Price ($) vs. Years Since Remodeled

Different Conditions and Price

It's expected that positive features of a house, for instance proximity to a park or a public facility positively impact the price of a house. On the other hand, if a house is adjacent to railroad, it makes sense that the house is cheaper than normal due to noise. Figure 1 shows different positive and negative characteristics of a house. As mentioned for instance, closeness to railroad or a feeder street is generally negative (compared to Normal condition). On the other hand, proximity to a positive off-site feature or an arterial street positively affects the price. However, there are a couple exceptions. For example, being within 200' of North South Railroad is found to be positive. This could be due to transportation characteristics of Ames, Iowa. Residing next to north-south railroad may increase the accessibility and mobility, and therefore has a positive impact on the price. Also, this needs further investigation to see if there are enough samples in each category to justify if the conclusions are reasonable and generalizable.

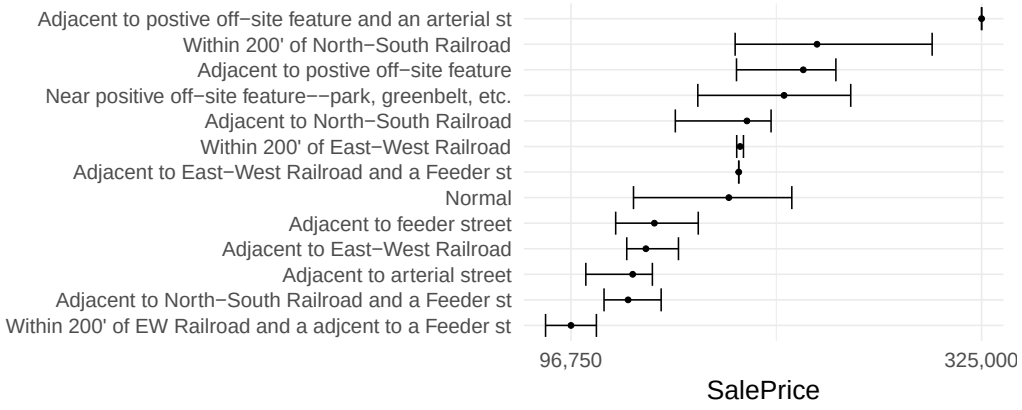


Figure 1: Average Price and 25- and 75-percentile values associated with each condition

Another important feature that changes the price of a house is the air conditioning system. Figure 2 shows the price distribution of the houses with or without central AC system. As expected, Central AC positively impact the price of a house. It can be seen that the average price for the houses with cnetral AC is slighty below 200k, whereas the same average for houses without it is closer to 100k.

References

[1] Dean De Cock. Ames, iowa: Alternative to the boston housing data as an end of semester regression project. *Journal of Statistics Education*, 19(3), 2011.