# What Affects the Sale Price of Housing in Ames?

*Ruoqi Huang, Xuehan Liu, Yiran Xu*

## Introduction

Housing, as a necessity of living, has been a very important thing to people. In this project, we are interested in finding what have been affecting the sales price of housing in Ames in recent years. In general, we will look at the relation between sale prices and three aspects of houses: construction, structure, and quality. For each aspect, we will examine two variables. For construction of housing, we will study the type of exterior wall material and type of dwelling (house); for structure of housing, we will look at the above ground living area (in square feet), and the number of bedrooms above ground; for quality of the housing, we will look at the type of heating, and zoning classifications. In addition, we also investigate the interactions between variables. For interaction terms, we look at how the presence of the central air conditioning affect the relation between sale price and other variables.Intuitively, air conditioning should be one of the most important aspects to people to consider when they want to buy a house. We believe all these variables contribute to the sale price of housing, and our project is to check whether all these seven variables indeed affect the sale price remarkably.

## Summarizing and cleaning data

The data used for analysis in our project is found from Stanford University (URL:http://web.stanford.edu/class/stats191/data/). This data set contains 2000 observations and 80 variables that were directly related to property sales. We would only focus on seven variables and they are Bldg.Type (type of dwelling), Exterior.1st (exterior covering on house), MS.Zoning (general zoning classification of the sale), Heating (heating type), Central.Air (central air conditioning), Gr.Liv.Area (above ground living area (in square feet)), and Bedroom (the number of Bedrooms above grade).

We initially examined each data variable, converted them to the appropriate data type, and recoded NA variables to read as NA rather than as a number. Variables coded as factor variables are Bldg.Type, Exterior.1st, MS.Zoning, Heating, Bedrooms, and Central.Air. Variables coded as numeric variables are Gr.Liv.Area.

We also added a new variable called GroupPrice. This variable is created by regrouping the sales price according to its quantile. We grouped from the minimum 12790$ to the second quantile 129,00$ as “lowest”, from the second quantile 129,000$ to median 160,000$ as “Second lowest”, from the median 160,000$ to the third quantile 213,500$ as “second highest”, from the third quantile 213,500$ to maximum 755,000$ as “highest”. The reason we create this variable is to investigate the relationship between variables and sale price conveniently since we have a lot of categorical variables.

In addition, grouped some levels of variables together where there are only few cases in those levels. First, for variable Bedroom (number of bedrooms), we grouped houses that has 8 bedrooms together with houses that has 6 bedrooms as “Above 6”, because there is only one house with 8 bedrooms. Similarly, for variables Heating (Heating Type), there are only one data point in category of “floor” and “Wall”, so we combine them with “Grav” to form the new level of “other furnace”; we also combine “GasW” and “OthW” as “steam”, we leave “GasA” as “air furnace”.

## Analysis

**Question1: How does the Construction of a house affect its sale price?**

**1A. In ames, how does the type of exterior wall material affect the sale price?** There are 14 different kinds of materials that are used for the exterior walls. Our hypothesis is that the more expensive the material itself is, the higher the sale price of the house would be. Therefore, we think the houses that use Vinyl Siding(VinylSd), which we think is the most expensive material, as the exterior wall materials would have the highest price. Cinder block(CBlock) may have the lowest sale price.

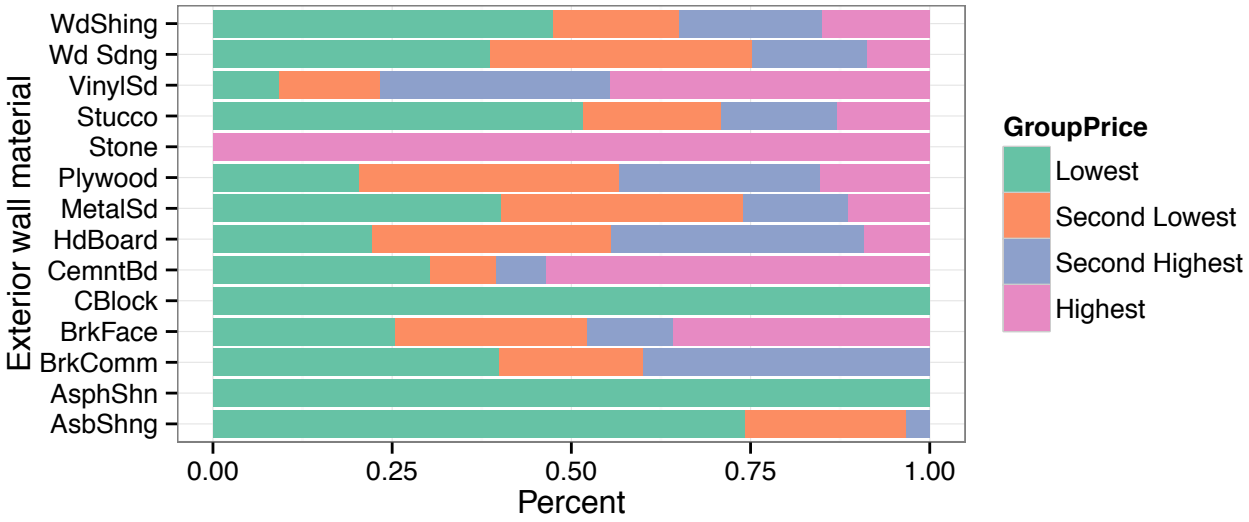


**Figure 1. Mosaic plot of Sale Price vs. Exterior Wall Material Type to see the relationship between exterial wall material and sale price**

From Figure1 we can tell that Asphalt Shingles(AsphShn) and Cinder Block(CBlock) have the lowest sale price because they only have one green color. It supports our hypothesis. The material stone is all red which is the highest price, which means the houses whose exterior walls material is stone have the highest sale price. However,after deeper research, we found there is only one data point for stone which is the reason why Stone seems to have the highest sale price. Since there are not enough data for us to analyze stone,it makes the material Cement Board(CemntBd) has the highest sale price. Therefore, due to the big variation of the Sale Price,exterior wall material is an important factor for sale price. Houses with different types of exterior wall material have different sale prices.

**1B. In ames, how does the type of dwelling affect the sale price?** There are five levels of dwelling. Our hypothesis is that Two-family Conversion(2fmCon) may have the lowest sale price.We also guess that Single-family Detached(1Fam) has the highest sale price.

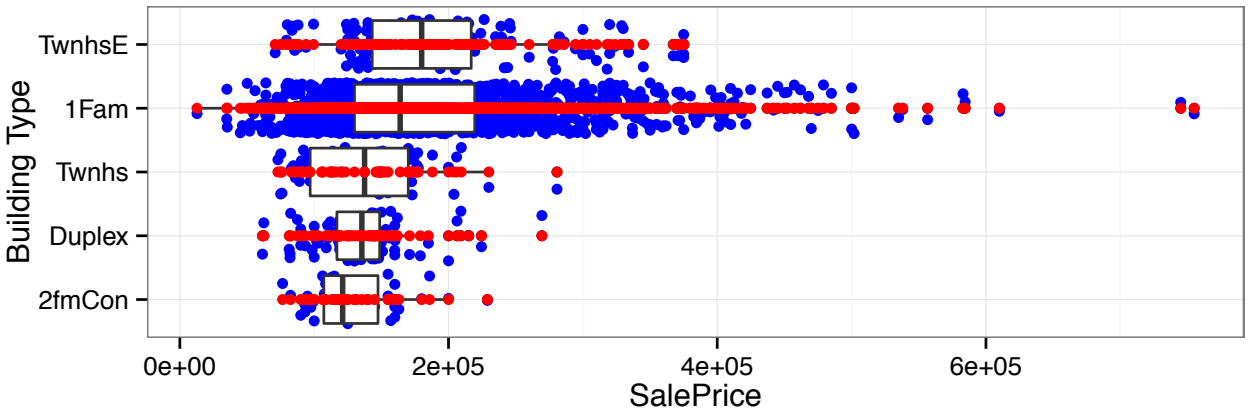


**Figure 2. Boxplot of Sale Price vs. Types of Dwelling to see the relationship between type of dwelling and sale price**

From the box plot we can tell that Townhouse End Unit(TwnhsE) has the highest mean sale price, but the second highest Single-family Detached(1Fam) has the largest maximum and more outliers on the upper bound.The cheapest type of dwelling is Two-family Conversion(2fmCon) which fits our hypothesis. Since different types of dwelling have various prices, we can say that dwelling is a factor that affects sale price.

**Question2: How does the structure of a house affect its sale price?**

**2A.In Ames, how does above ground living area affect the sale price?** With more living area above ground, owner of the house would have more space to do whatever they want. Therefore, we hypothesize that the larger above ground living area is, the higher the sale price would be.

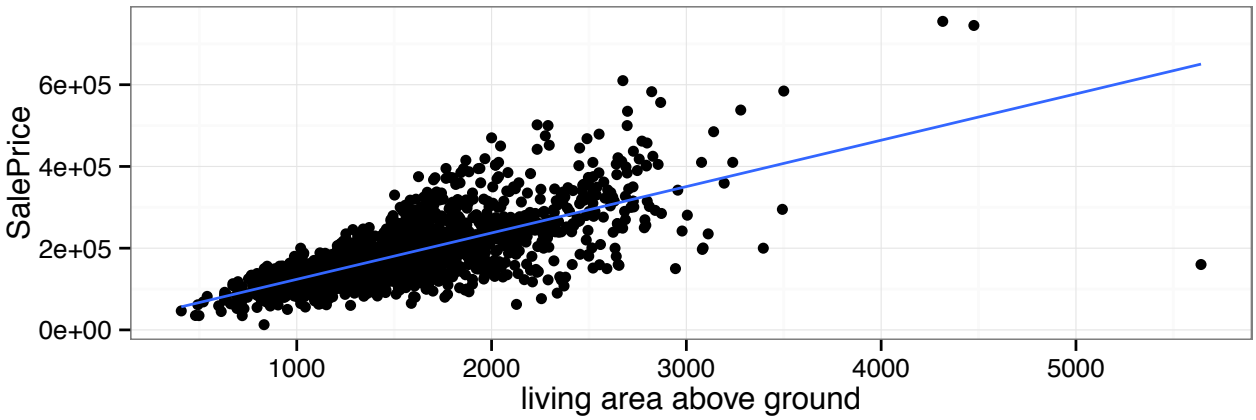


**Figure 3. Scatterplot of Sale Price vs. living area above ground to see the relationship between living area and sale price**

Figure 3 is the scatter plot of above ground living area and sale price with a least squared linear regression line. From the graph, we can see that sale price and area above ground living area are positively related, except an outlier in living area above 5000 square feet. The correlation between them is around 0.735 excluding the outlier. R-square is about 0.52. Considering there are so many other factors to influence sale price, the R square of 0.52 is reasonable. Look more closely, the residual seems to get larger as living area increases. It might due to that when people feel they have enough living area, living area does not contribute a lot to the sale price anymore.Other characteristic of house become important in determining the sale price.

**2B.In Ames, how does the number of bedrooms above ground affect the sale price?** Intuitively, more bedrooms above ground suggest more rooms and space in a house. Therefore, we hypothesize that the greater the number of bedrooms are, the higher the sale price would be.

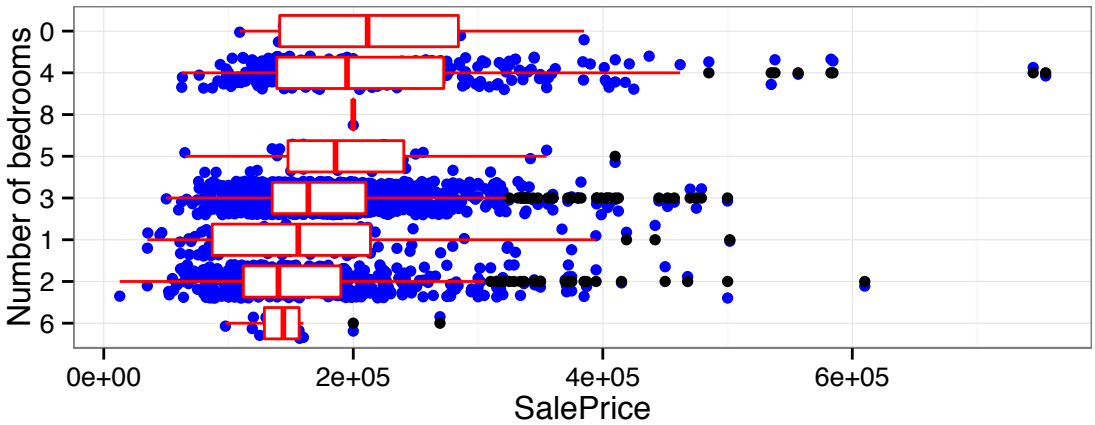

**Figure 4. Boxplot of Sale Price vs. Number of Bedrooms to see the relationship between number of bedrooms and sale price**

Figure 4 shows the box plot of sale price and the number of houses reordered by mean value of sale price. From the graph, we can see that on average, houses with six or more bedrooms tend to be sold at lowest price.Average sale price of houses with tow bedrooms are the second lowest. Average price of houses with one bedroom are the third lowest. Then, houses with three or five bedrooms have higher average sale price. Average price for houses with four bedrooms are the second highest. To our surprise, mean sale price for houses with no bedroom are the highest. This is not consistent with our hypothesis. Furthermore, if we only look at houses with two or three or four bedrooms, which are the types of houses that are sold most. Although four-bedrooms houses do have highest mean,maximum price and two-bedrooms houses do have lowest minimum and mean price,we can see that the prices in each type are widely distributed.The range of price in each bedroom level are very close,excluding outliers. Therefore, number of bedrooms are not a significant factor affecting sale price according to our analysis.

**Question3: How does the quality of a house affect its sale price?**

**3A. In Ames, how does sale price varies among different zonings?** In our data set, there are seven different zoning classifications in Ames. They are agriculture, commercial, floating village residential, industrial, residential high density, residential low density, and residential medium density. Intuitively, most people prefer to stay in the residential areas, so houses there should have a relatively high price; on the other hand, in the industrial zone, they tend to have no residential housing built, because it is near factory, and may get pollution. Houses there should be the least popular.

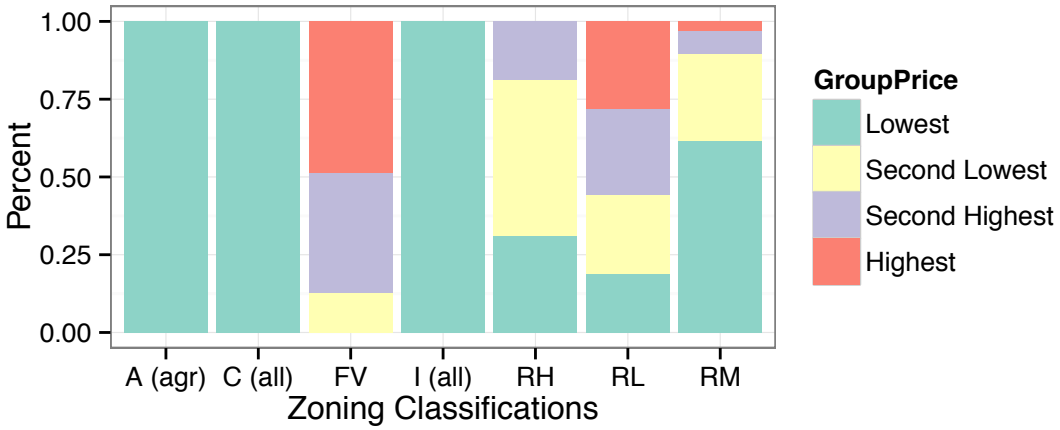


**Figure 5. Mosaic plot of Sale Price vs. Zoning classification to see the relationship between zoning and sale price**

From figure 5,we observe that for agriculture, commercial, and industrial zones, houses are only classified in the lowest sale price group. This result is not surprise because nobody want to buy a housing these types of areas, so the sale price of houses should be very low. For floating village residential zone, the biggest fraction is the highest sale price; second highest sale price has the smaller proportion; and the second lowest sale price group has even smaller fraction, and there is no houses in lowest sale price group. For residential high density zone, no houses are recorded in the highest sale price group. The proportion of houses in the second lowest sale price group is the largest, and the second highest sale price group is the smallest. The proportion of second lowest sale price group is in the middle. For residential low density zone, the biggest proportion is the lowest sale price group, and the smallest proportion is the highest sale price group. In conclusion, we can detect that the distribution of sale price is very different for each level of zoning. So the variable zoning classifications (MS.Zoning) affects the sale price in Ames a lot. However we should also notice that there only

a few data points for zoning of industrial and agriculture, so this limitation may make our analysis a bit imprecise.

**3B. How does heating type affect the sale price of housing in Ames?** In general, most common heating type is the use air as a medium. So the houses with "air furnace" may have a relatively high price. Heating type of "other furnace" may be the least popular, so houses with such heating type should have relatively low sale price.

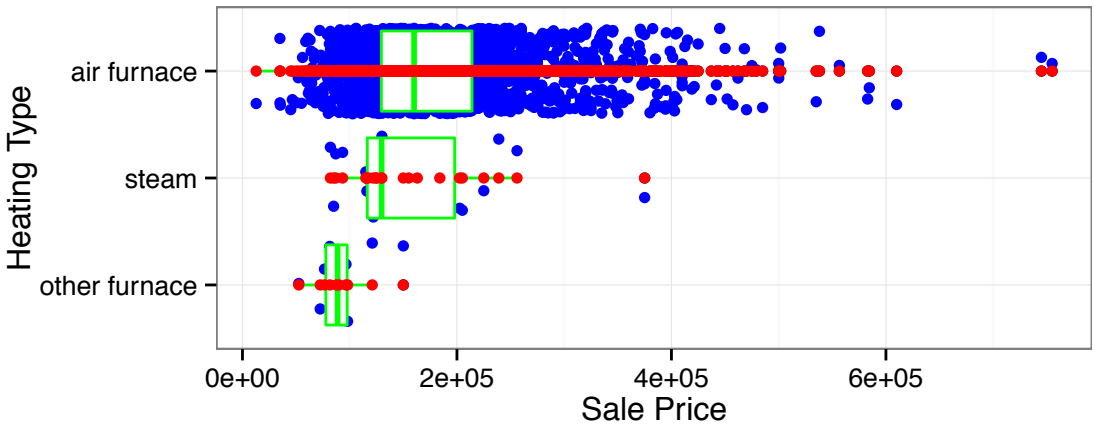


**Figure 6. Boxplot of Sale Price vs. Heating Type to see the relationship between heating type and sale price**

Based on figure 6 we observe that the house with heating type of "air furnace" has the highest mean sale price; the house with heating type of "steam" has the second highest mean sale price; the house with heating type of "other furnace" obtains the Lowest mean sale price. In addition, we can see that the sale price changes a lot as the type of heating changes. So the variable Heating Type (Heating) affects the sale price significantly. We should also notice that for heating type "other furnace", there are still only a few data points after regrouping. This limitation of data may occur, as we mentioned before, because that "other furnace" such as floor furnace or wall furnace was not indeed popular at that time.

**4.Interaction**

**4.0. Is having air conditioning really important to the sale price of housing?** As we mentioned in the introduction, having air conditioning is regarded as an important factor to the sale price of housing. We first ascertain the relation between the sale price and the presence of central air conditioning.

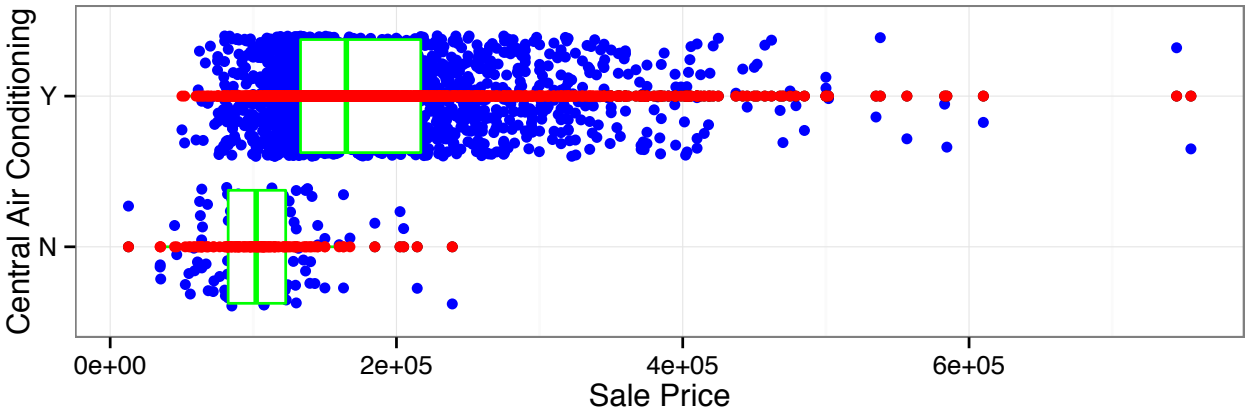

**Figure 7. Boxplot of Sale Price vs. the Presence of Air Conditioning to see the relationship between presences of air conditinging and sale price**

Figure 7 gives us a clear picture of the distribution of sale price of housing with or without central air conditioning. It is obvious that the median of sale price for houses with central air conditioning is much higher than that for houses without central air conditioning, the same happens to maximum. Also notice that the number of houses sold with central air conditioning is much larger than the number of houses sold without central air conditioning. In conclusion, air conditioning affects the sale price of housing remarkably.

**4A.How does Air Conditioning change the relation between Sale Price and the Exterior Wall Material Type?** Because exterior wall might have different insulation ability, we hypothesis that presence of air conditioning might change the relationship between sale price and exterior wall material type.

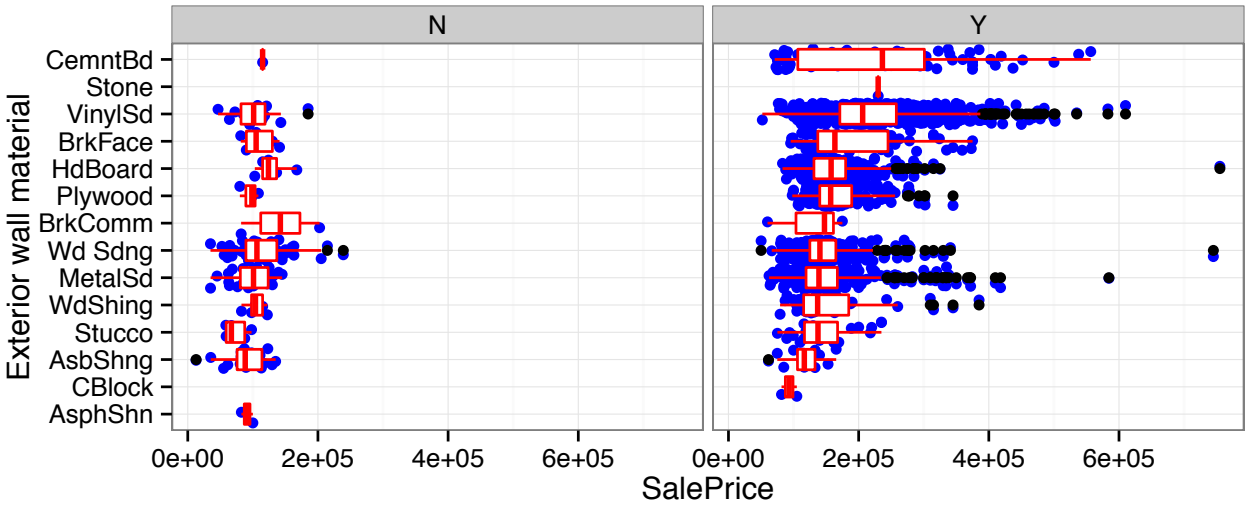


**Figure 8. Boxplot of Sale Price vs. Exterior Wall Material Facet by Air Conditioning to see whether the relationship between sale price and exterior wall material changes as the presence of air conditioning**

We can see from the Figure 8 that the distribution of Sale Price with Central air conditioning is about the same as before. For houses without air conditioning, the relationship has changed. For instance, bricks become the highet median, maximum and mean price. We figured that the reason might be due to different thermal mass properties. Bricks have exceptional thermal mass properties to store heat and release slowly. Therefore, as we have hypothesized, relationship between sale price and exterior wall material indeed changes as the presence of air conditiong.

**4B. Dose having air conditioning influence the relation between number of bedrooms and sale price?** Because air conditiong and number of bedrooom seem to be two totally differnet variables, we hypothesis presence of air conditiong might not change the relationship between number of bedrooms and sale price.

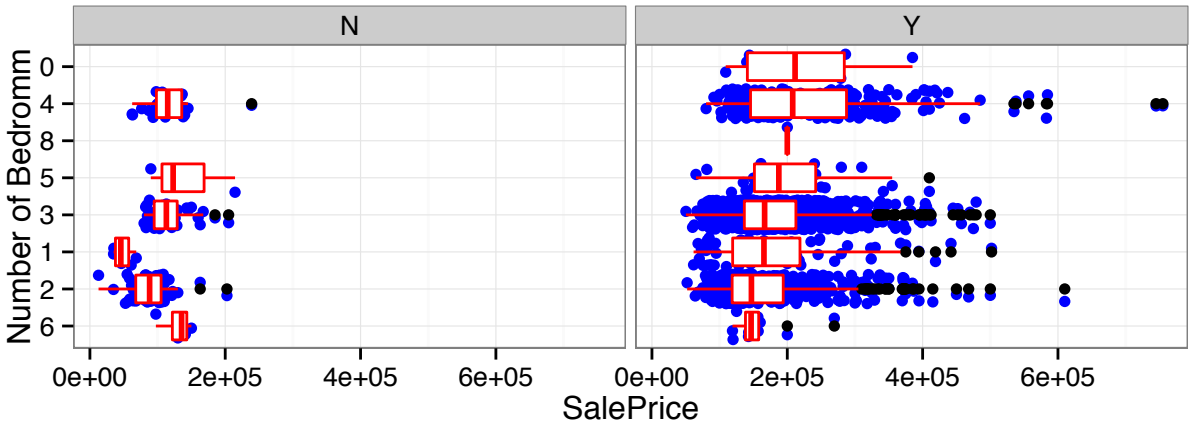


**Figure 9. Boxplot of Sale Price vs. Number of bedrooms facet by Air Conditioning to see whether the relationship between sale price and number of bedrooms changes as the presence of air conditioning**

The left panel of figure 9 indicates the relationship between number of bedrooms and Sale Price for houses without air conditioner. We can see that houses with one bedroom has the lowest mean price, minimum price as well as maximum prices. Houses with two bedroom have second lowest mean price.Houses with three bedrooms have higher mean price than houses with two bedrooms. Therefore, the left faset panel suggest that for houses without air conditioner, the more bedrooms it has, the higher sale price would be. The right panel of figure 8 indicates the relationship between number of bedrooms and Sale Price for houses with air conditioner.The graph is very similar to what we analyzed before for general case. Therefore, unlike our hypothesis, the presence of air conditioning change the relationship between number of bedrooms and sale price.

**4C.How having air conditionning affects the relation between sale price of the housing and heating type?** Air conditionging and heating can play similar role in winter. Therefore, we hypothesis the relationship between sale price and heating type will change as the presence of air conditioning.

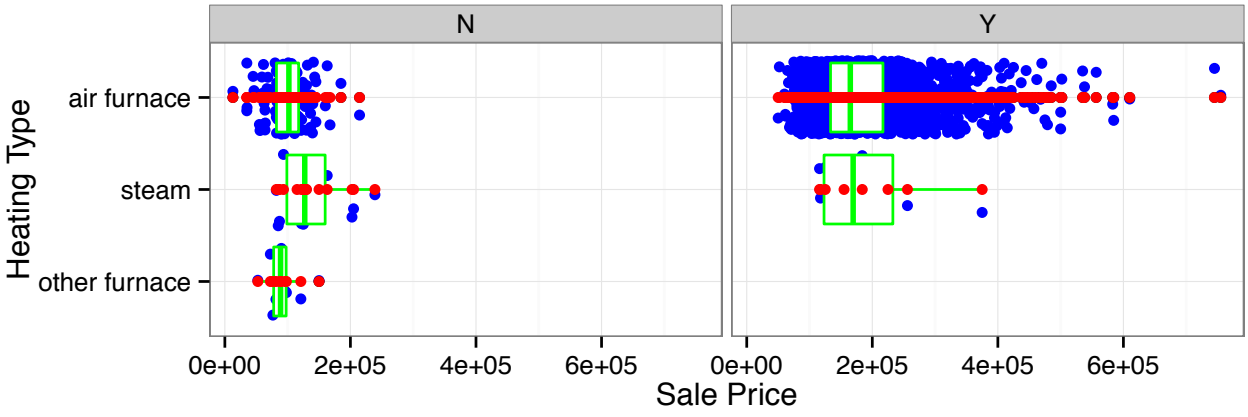


**Figure 10. Boxplot of Sale Price vs. Heating type facet by Air Conditioning to see whether the relationship between sale price and heating type changes as the presence of air conditioning**

Based on the figure 10, we notice that given the housing does not have central air conditioning, the median sale price of houses with "steam" as heating type is higher than the median sale price of houses with "air furnace". However, this is not quite the same for houses with central air conditioning. Given houses with

central air conditioning, the median sale price for houses using air furnace is almost the same as the median sale price for houses using other type furnace. This happens may because air conditioning can do both heating and cooling for the house, so the individual heating machine is not that important. Overall, having air conditionning does affect the relation between sale price of the housing and heating type. This is consistens with our hypothesis.

## Summary of Results/Conclusions:

For construction of housing, type of exterior wall material is an important factor for the sale price. Fancier and more expensive material makes a house have higher price. Type of dwelling has large impact on sale price too. Townhouse End Unit and Single-family Detached are more expensive. For interaction, having Air Conditioning changes the relation between Sale Price and the Exterior Wall Material Type except due to different thermal mass property. For the two variables we investigate in structure, we found that they contribute to sale price to various degrees. For living area above ground, the more living area a house has, the higher the sale price tend to be. For number of bedrooms, we found that there is no obvious relationship between number of bedrooms and sale price in general. However, when we facet the relationship by air conditioning, we found that for houses without air conditioning, the more bedrooms it has, the higher the price would be. For houses with air conditioning, there is no significant relationship between number of bedrooms and sale prices. For quality of the house, first, we found that the sale price varies a lot in different zones in Ames. In addition, houses with different heating types also have pretty different sale prices. Overall, both two variables zoning classifications and heating type affect the sale price in a remarkable way. Last, for interaction, in general, having air contioning or dose not affect the relation between sale price of housing and housing's heating type.