Untitled

Introduction A real estate agent wants to develop a model to predict the sale price of a house using data collected from 521 houses. He assumes that the following variables which may useful in predicting the sale price of a given house. 1. SalePrice - sale price of house
2. SqFeet - square footage of house
3. Beds - number of bed rooms
4. Baths - number of bath rooms
5. Air - whether the house has air conditioning or not.
6. Garage - number of garages
7. Pool - whether the house has pool or not.
8. Year - year of constriction
9. Quality - quality of the house (1 = Excellent, 2 = Good, 3 = Average)

```
RealEstate <- read.csv("C:/Users/acer/Downloads/RealEstate.csv")
```

```
str(RealEstate)
```

```
## 'data.frame':    521 obs. of  10 variables:
##  $ X        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ SalePrice: num  360 340 250 206 276 ...
##  $ SqFeet   : num  3.03 2.06 1.78 1.64 2.2 ...
##  $ Beds     : int  4 4 4 4 4 4 3 2 3 3 ...
##  $ Baths    : int  4 2 3 2 3 3 2 1 2 3 ...
##  $ Air      : int  1 1 1 1 1 1 1 1 1 0 ...
##  $ Garage   : int  2 2 2 2 2 5 2 1 2 1 ...
##  $ Pool     : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Year     : int  1972 1976 1980 1963 1968 1972 1972 1955 1975 1918 ...
##  $ Quality  : int  2 2 2 2 2 2 2 2 3 3 ...
```

```
head(RealEstate, n=7)
```

```
##   X SalePrice SqFeet Beds Baths Air Garage Pool Year Quality
## 1 1     360.0  3.032    4     4   1      2    0 1972       2
## 2 2     340.0  2.058    4     2   1      2    0 1976       2
## 3 3     250.0  1.780    4     3   1      2    0 1980       2
## 4 4     205.5  1.638    4     2   1      2    0 1963       2
## 5 5     275.5  2.196    4     3   1      2    0 1968       2
## 6 6     248.0  1.966    4     3   1      5    1 1972       2
## 7 7     229.9  2.216    3     2   1      2    0 1972       2
```

```
summary(RealEstate)
```

```
##        X           SalePrice        SqFeet           Beds           Baths
##  Min.   :  1   Min.   : 84.0   Min.   :0.980   Min.   :1.000   Min.   :1.000
##  1st Qu.:131   1st Qu.:180.0   1st Qu.:1.701   1st Qu.:3.000   1st Qu.:2.000
##  Median :261   Median :229.9   Median :2.061   Median :3.000   Median :3.000
##  Mean   :261   Mean   :277.4   Mean   :2.261   Mean   :3.478   Mean   :2.647
##  3rd Qu.:391   3rd Qu.:335.0   3rd Qu.:2.638   3rd Qu.:4.000   3rd Qu.:3.000
##  Max.   :521   Max.   :920.0   Max.   :5.032   Max.   :7.000   Max.   :7.000
##       Air           Garage           Pool            Year
```

```
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :1885
##  1st Qu.:1.0000   1st Qu.:2.000   1st Qu.:0.0000   1st Qu.:1956
##  Median :1.0000   Median :2.000   Median :0.0000   Median :1966
##  Mean   :0.8311   Mean   :2.098   Mean   :0.0691   Mean   :1967
##  3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:0.0000   3rd Qu.:1981
##  Max.   :1.0000   Max.   :7.000   Max.   :1.0000   Max.   :1998
##     Quality
##  Min.   :1.000
##  1st Qu.:2.000
##  Median :2.000
##  Mean   :2.186
##  3rd Qu.:3.000
##  Max.   :3.000
```

```r
boxplot(RealEstate$SalePrice, horizontal = TRUE)
```
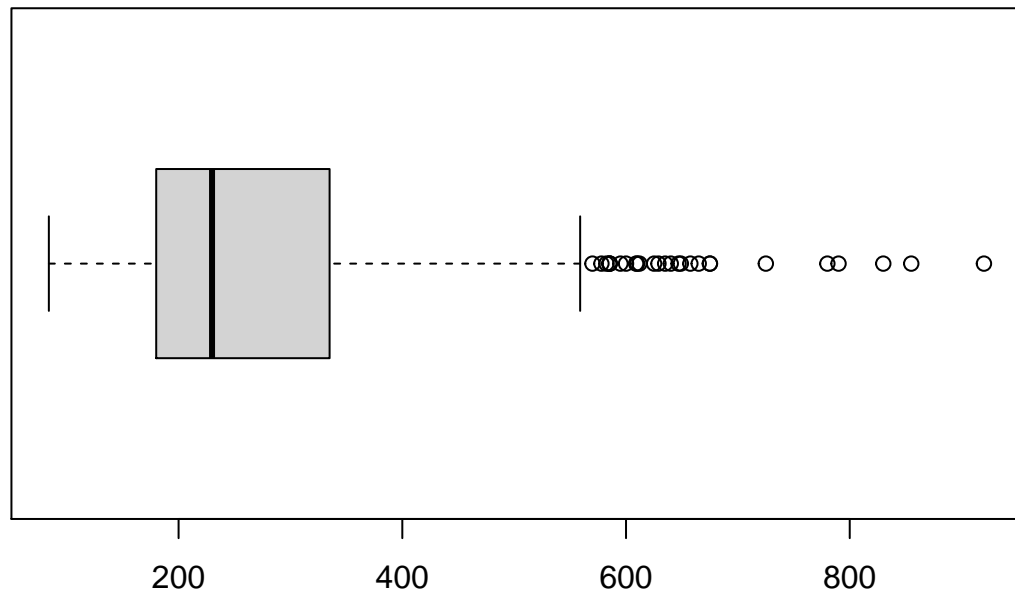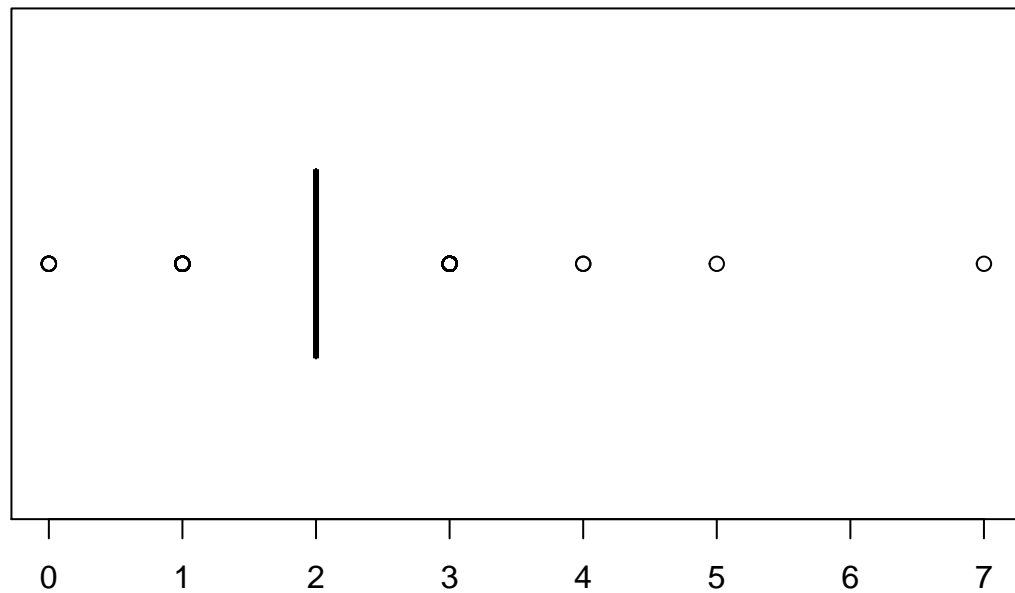

this is the boxplot for sale price of house

```r
boxplot(RealEstate$Beds, horizontal = TRUE)
```
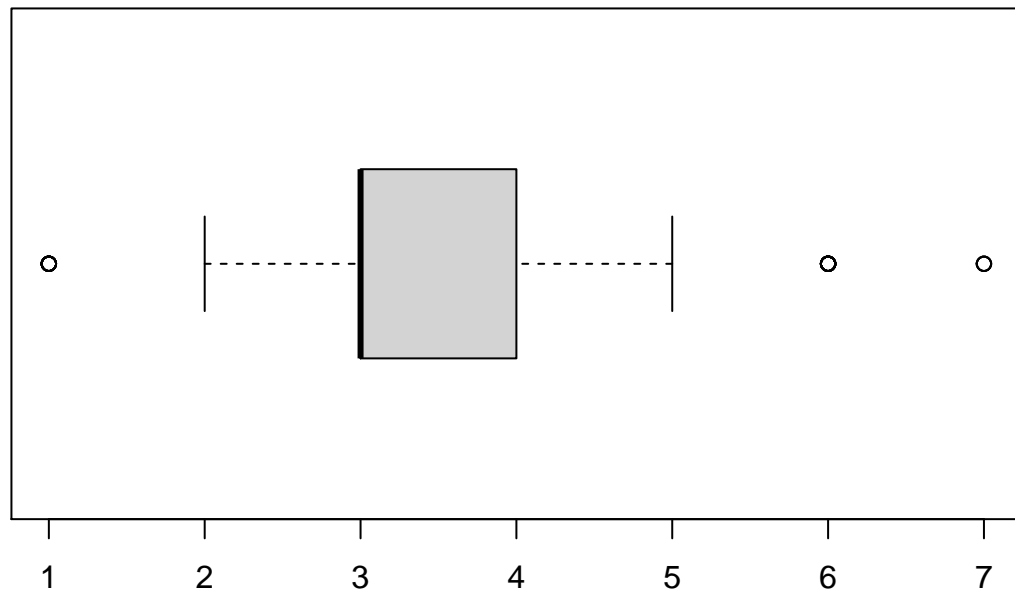
this the boxplot for number of bed rooms.Q1 is 2,Q3 is 5.median beds are 3.minimum beds value is 1 and maximum bed value is 7.
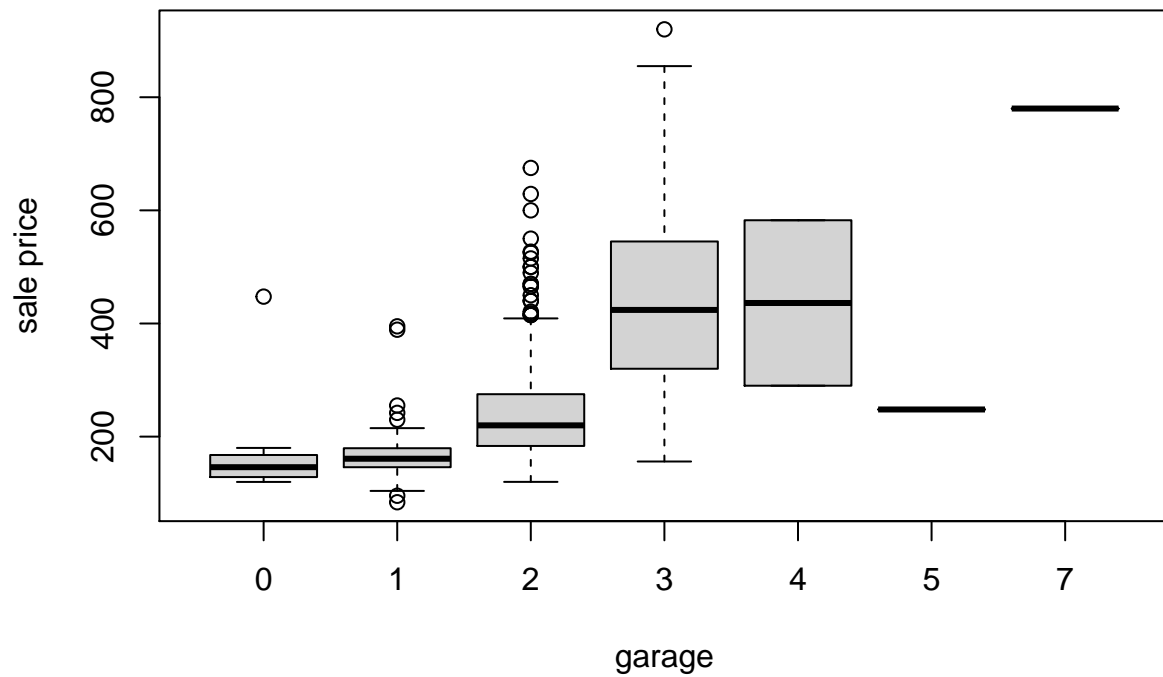
```r
boxplot(RealEstate$Garage, horizontal = TRUE)
```

this is a boxplot for garage.median is 2.

Side by side boxplots can be used to make comparison between groups.

```r
boxplot(RealEstate$SalePrice~RealEstate$Garage,xlab = "garage" , ylab = "sale price")
```

this is the side boxplots are used to comparison between garage and sales price.when there is 7 garage has maximum median sales price.maximum midean is increase when number of garage are increasing.but when number of garages are 5 ,their median is low.highest sale price is occur for 3 garages.

```r
hist(RealEstate$Beds, main = "number of beds in house",
xlab = "number of beds")
```

number of beds in house
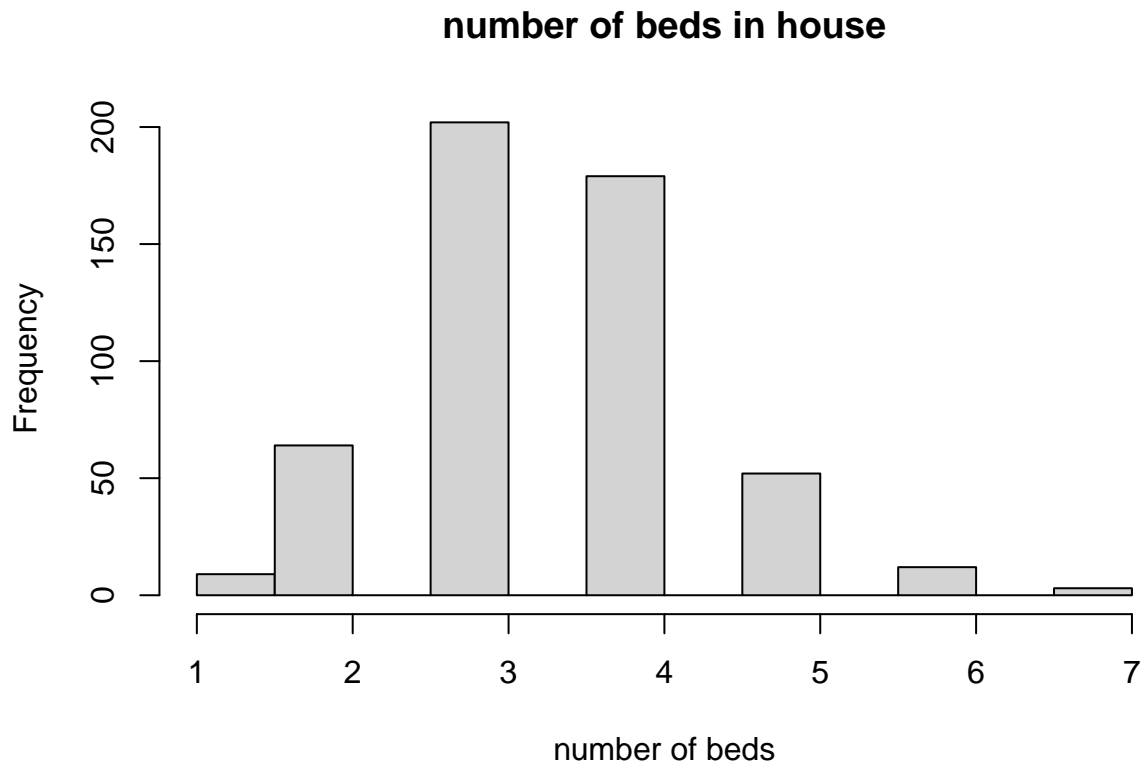
many houses have 3 beds.when number of beds are mimimum or maximum,minimum number of houses used it.in this histrogram we can conclude many of houses are do not use more beds.

```
hist(RealEstate$Air, main = "using air conditioner for  house",
xlab = "number of air conditioners")
```

using air conditioner for house
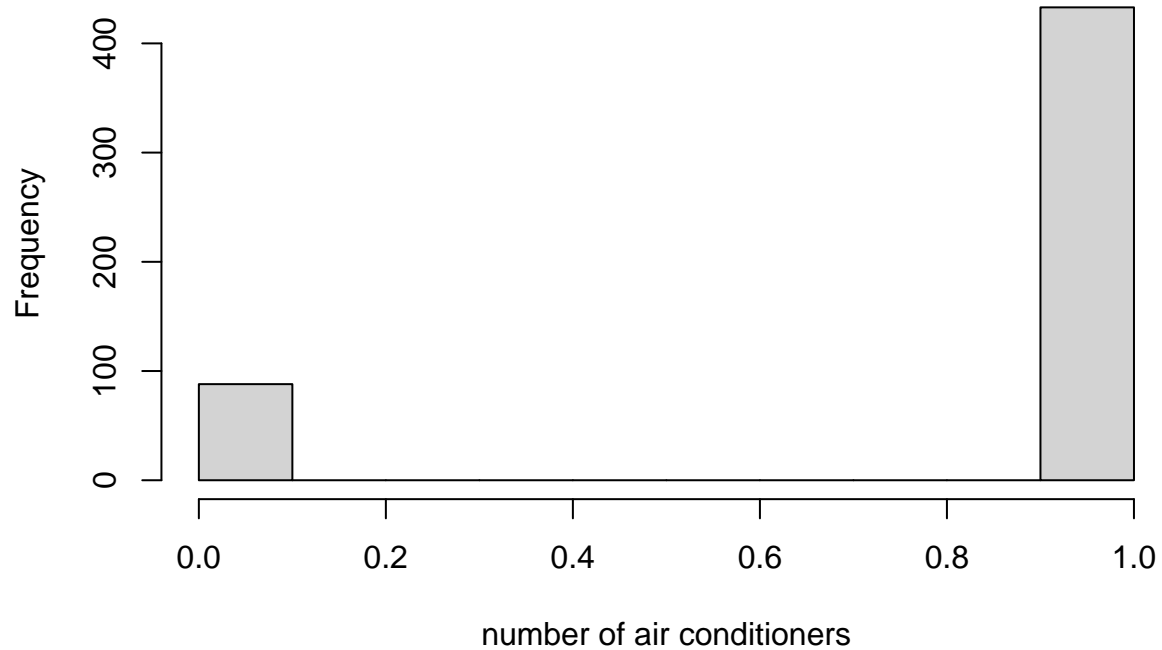
there are two options here.that is using air conditioner or not. 400 of houses are use air conditioner.100 of houses are not used air conditioners.

```r
plot(RealEstate$SalePrice,RealEstate$SqFeet,
xlab = "sqfeet of a house", ylab = "sale price of a house",main = "relation ship between sqfeet and sal
```

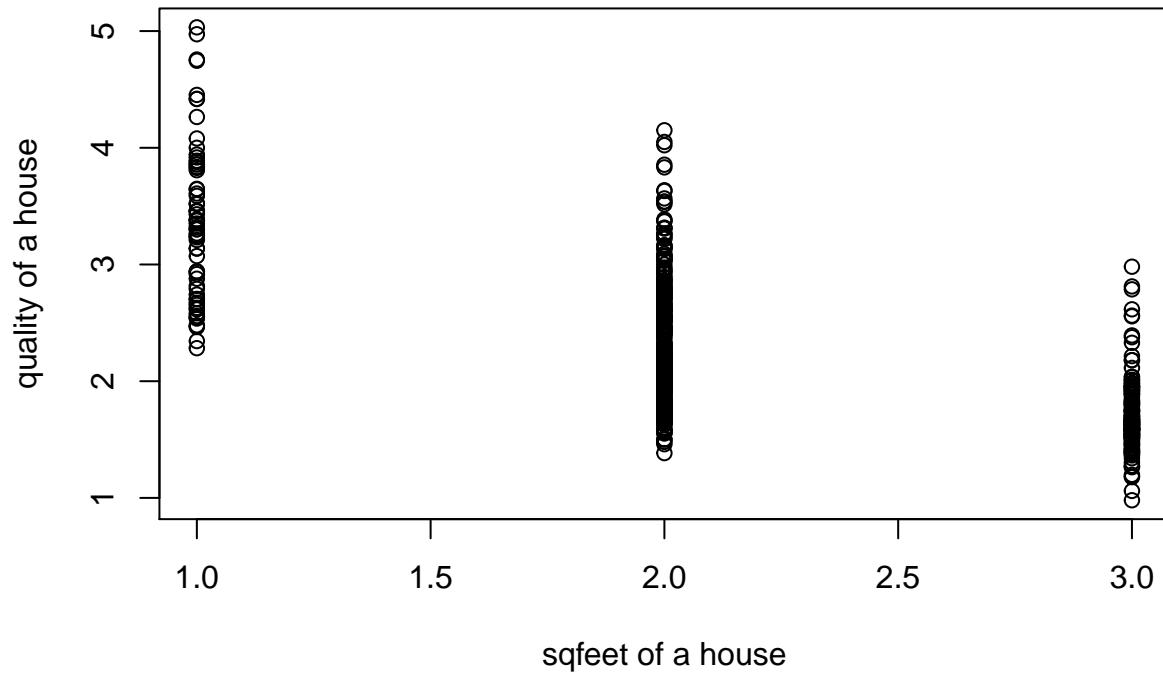# relation ship between sqfeet and sale price of a house
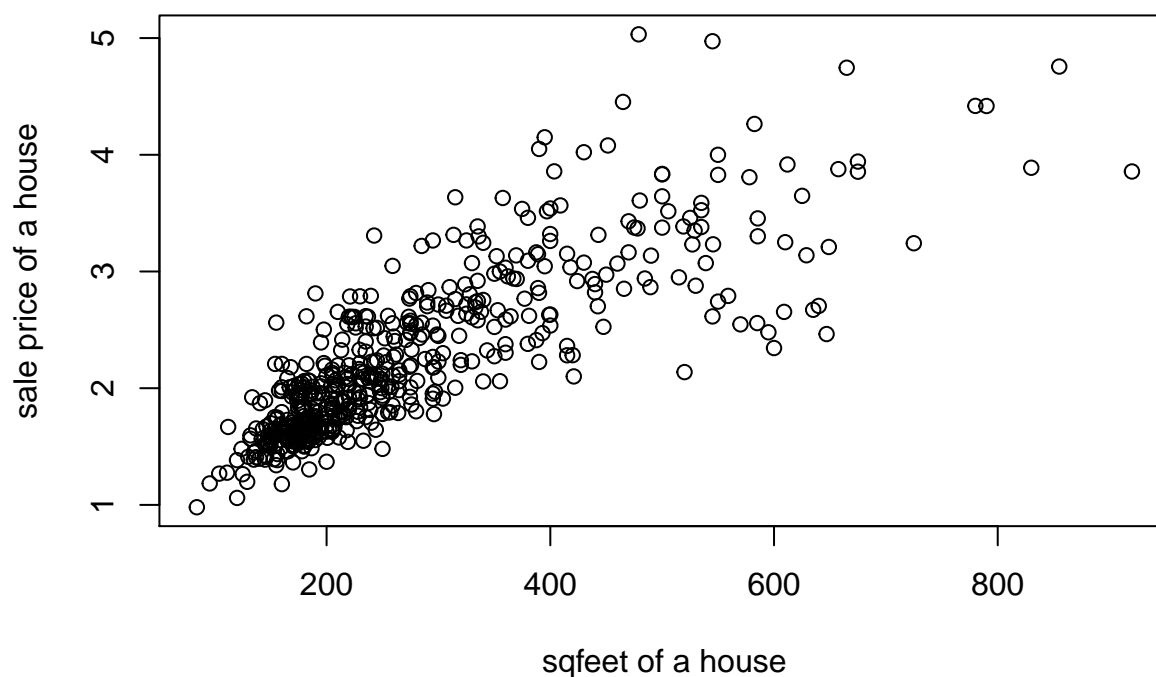
this is a posotive correlation between between sqfeet of a house and sale price of house.in here both variables are increasing togrther.

```
plot(RealEstate$Quality,RealEstate$SqFeet,
xlab = "sqfeet of a house", ylab = "quality of a house",main = "relation ship between sqfeet and quality
```
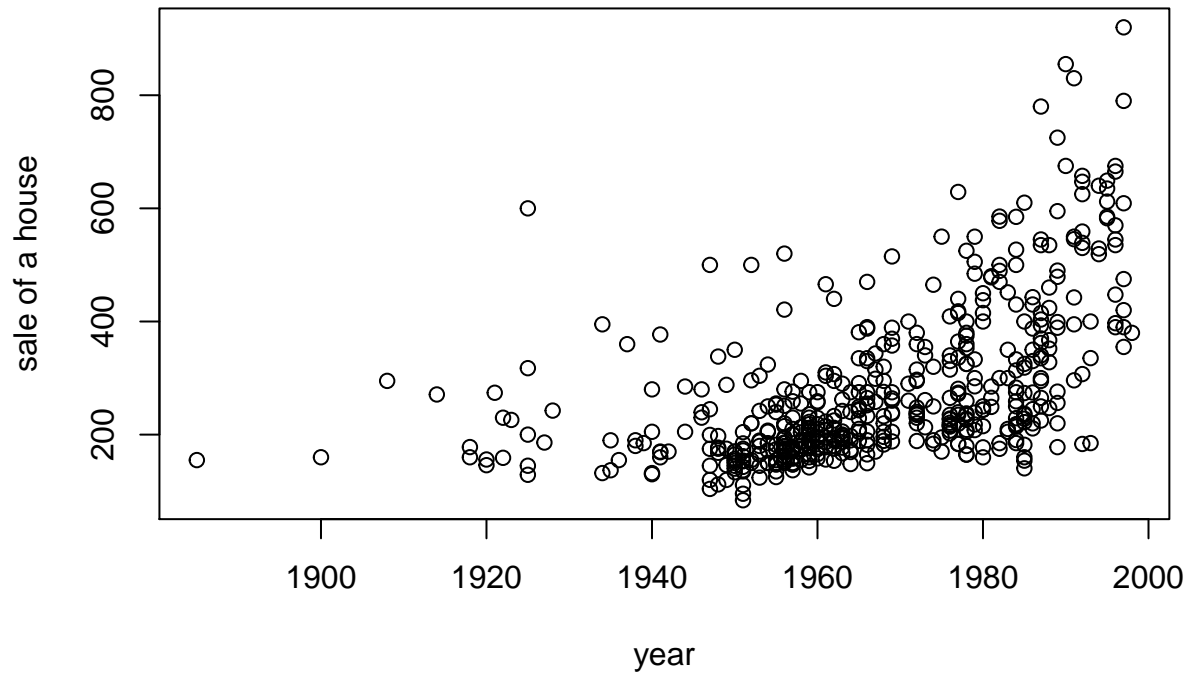
relation ship between sqfeet and quality of a house

sqfeet of a house

there is no relation positive or negetive relation between sqfeet and quality

```
plot(RealEstate$Year,RealEstate$SalePrice,
xlab = "year", ylab = "sale of a house",main = "relation ship between year and sale price of a house")
```

relation ship between year and sale price of a house

when time is modern sale price of house increasing.

DISCUSSION

In this survay , we observed data from 522 houses.sample size is large.in here we looked at how many houses were sold between 1900 and 2000.in this survey we learned that selling price of a house will increase acording to the number of beds,baths,air conditioner garage and pool in the house .by the time ,the selling price of a house , number of beds ,pools,garages and baths are increased.there is a positive correlation between between sqfeet of a house and sale price of house.over time, the use of air conditioner in homes has increased.quality of the house is increased with the time.