House Price Analysis

Darshan Nathani

2023-08-04

First we load the data from CSV file

```
getwd()
```

```
## [1] "C:/Users/Asus/Documents"
```

```
setwd('D:/Codes/R codes/Data/house')
data <- read.csv("D:/Codes/R codes/Data/house/kc_house_data.csv")
```

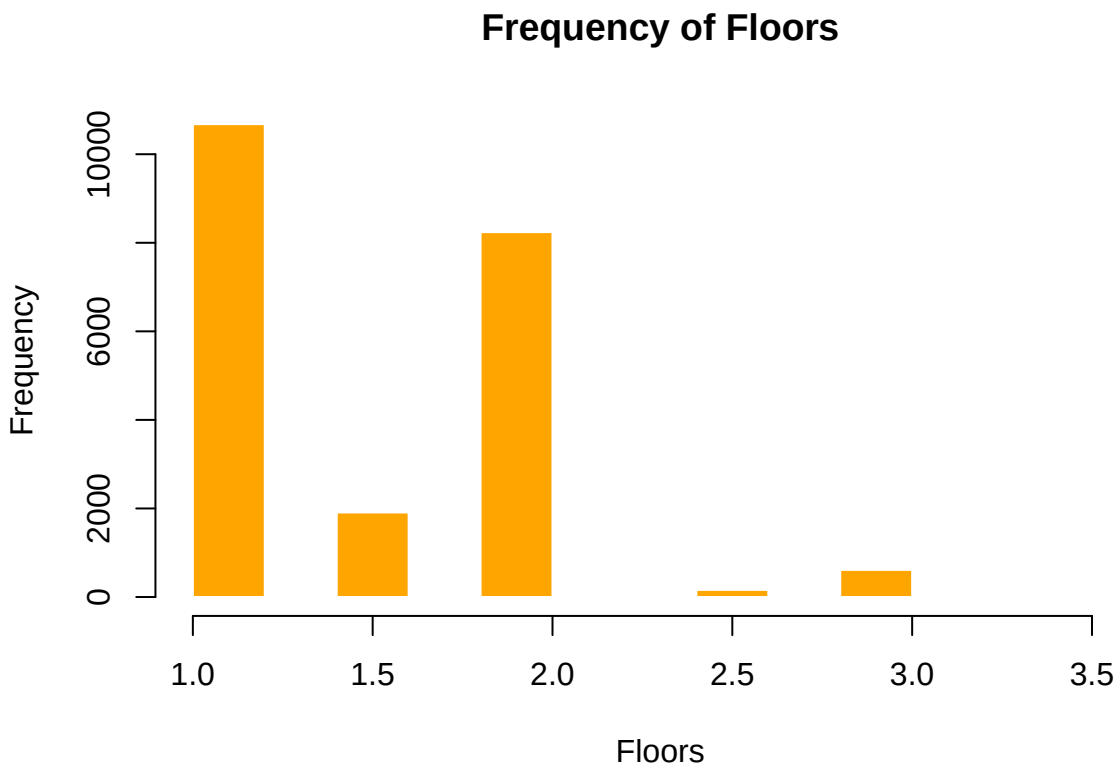
First let us look the houses that are built after 2000

```
new <- subset(data,data$yr_built>2000)
write.csv(new,"new.csv")
print(head(new))
```

```
##            id            date   price bedrooms bathrooms sqft_living sqft_lot
## 6  7237550310 20140512T000000 1225000        4      4.50        5420    10193
## 10 3793500160 20150312T000000  323000        3      2.50        1890     6560
## 30 1873100390 20150302T000000  719000        4      2.50        2570     7173
## 31 8562750320 20141110T000000  580500        3      2.50        2320     3980
## 32 2426039314 20141201T000000  280000        2      1.50        1190     1265
## 43 7203220400 20140707T000000  861990        5      2.75        3595     5639
##    floors waterfront view condition grade sqft_above sqft_basement yr_built
## 6       1          0    0         3    11       3890          1530     2001
## 10      2          0    0         3     7       1890             0     2003
## 30      2          0    0         3     8       2570             0     2005
## 31      2          0    0         3     8       2320             0     2003
## 32      3          0    0         3     7       1190             0     2005
## 43      2          0    0         3     9       3595             0     2014
##    yr_renovated zipcode     lat     long sqft_living15 sqft_lot15
## 6             0   98053 47.6561 -122.005          4760     101930
## 10            0   98038 47.3684 -122.031          2390       7570
## 30            0   98052 47.7073 -122.110          2630       6026
## 31            0   98027 47.5391 -122.070          2580       3980
## 32            0   98133 47.7274 -122.357          1390       1756
## 43            0   98053 47.6848 -122.016          3625       5639
```

The frequency of floors is

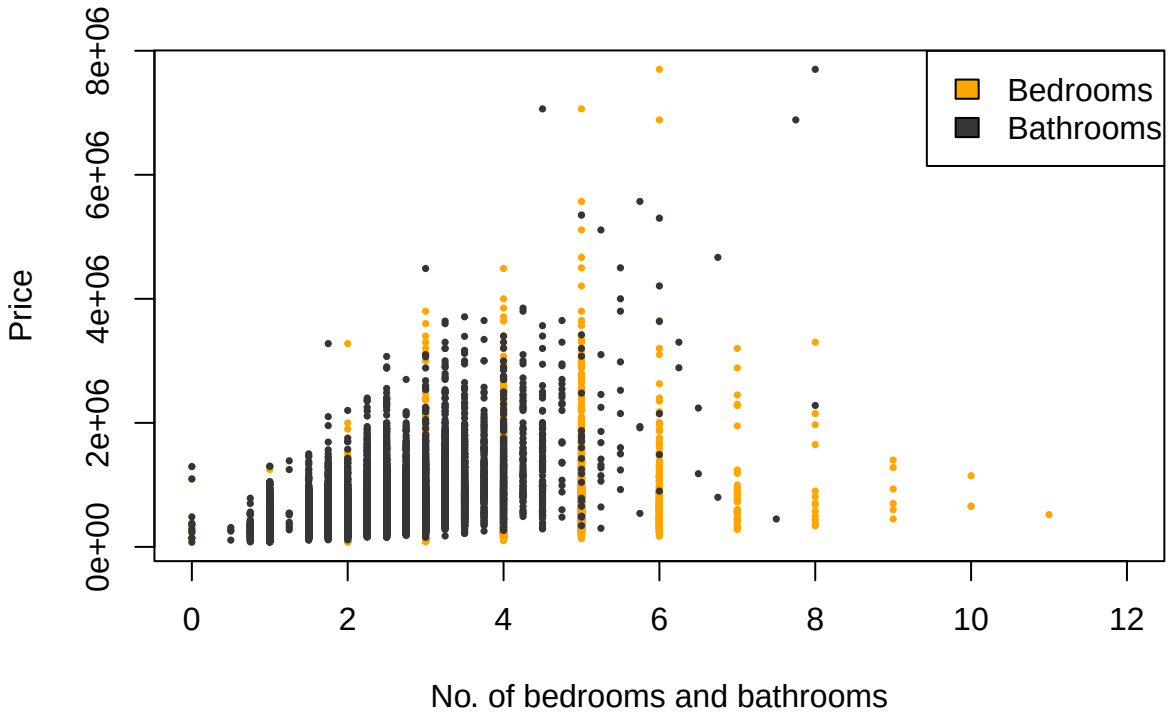
```r
hist(data$floors,xlab = "Floors",col ="orange",border = "white",main="Frequency of Floors")
```




```r
plot(data$sqft_living,data$price,xlab = "Living space (sqft.)", ylab = "Price",col="orange", main="Pric
lines(lowess(data$sqft_living, data$price), col = "black")
```

As we can see here as square feet of living space increases, the price of the following property also increases
Now Lets look what affects the price more

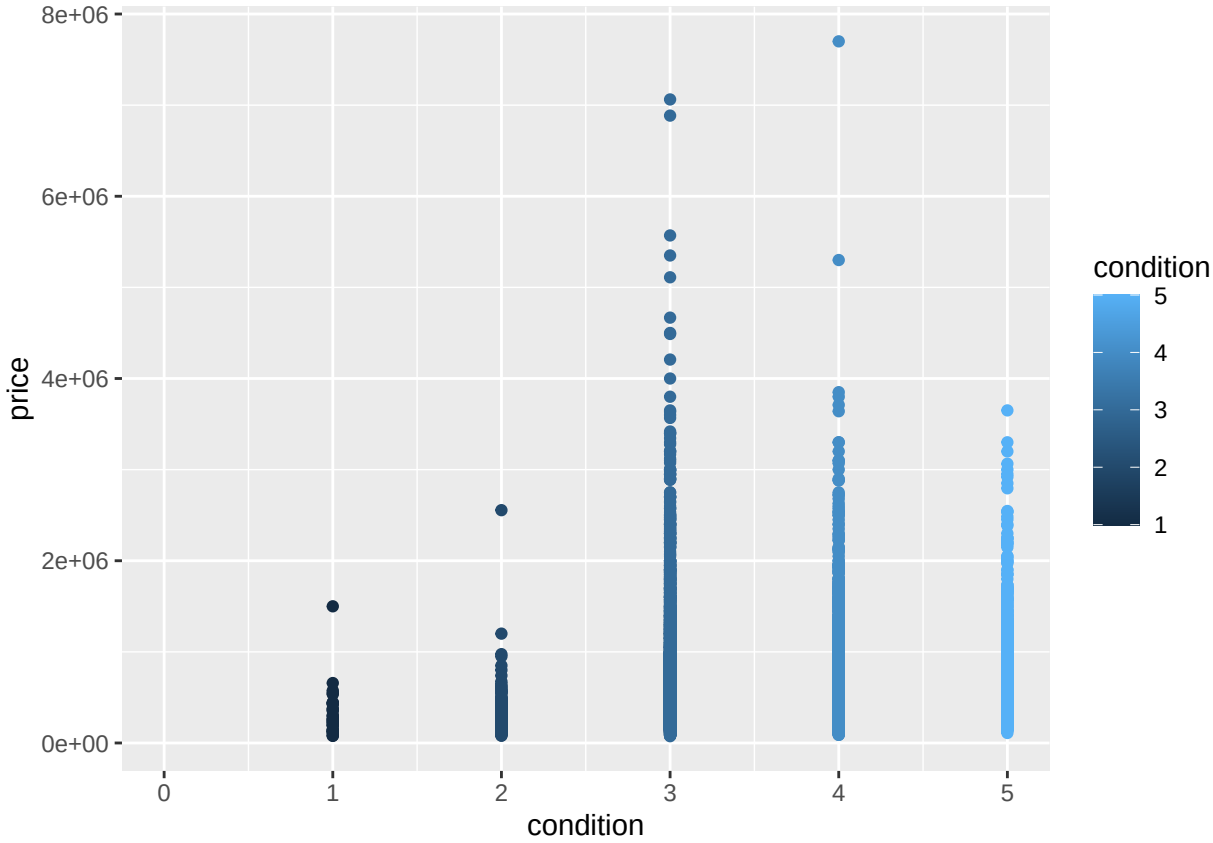
```
plot(data$bedrooms, data$price, col = "orange", pch = 16,cex=0.5, main="",xlim = c(0,12),xlab="No. of b
legend(9.432, 8e+06, legend=c("Bedrooms", "Bathrooms"),
       fill = c("orange","#353436"))
points(data$bathrooms, data$price, col = "#353436",cex=0.5,pch=16)
```



Here we can see that more number of bedrooms do not influence the price of house, whereas price of house is carried with number of bathrooms.

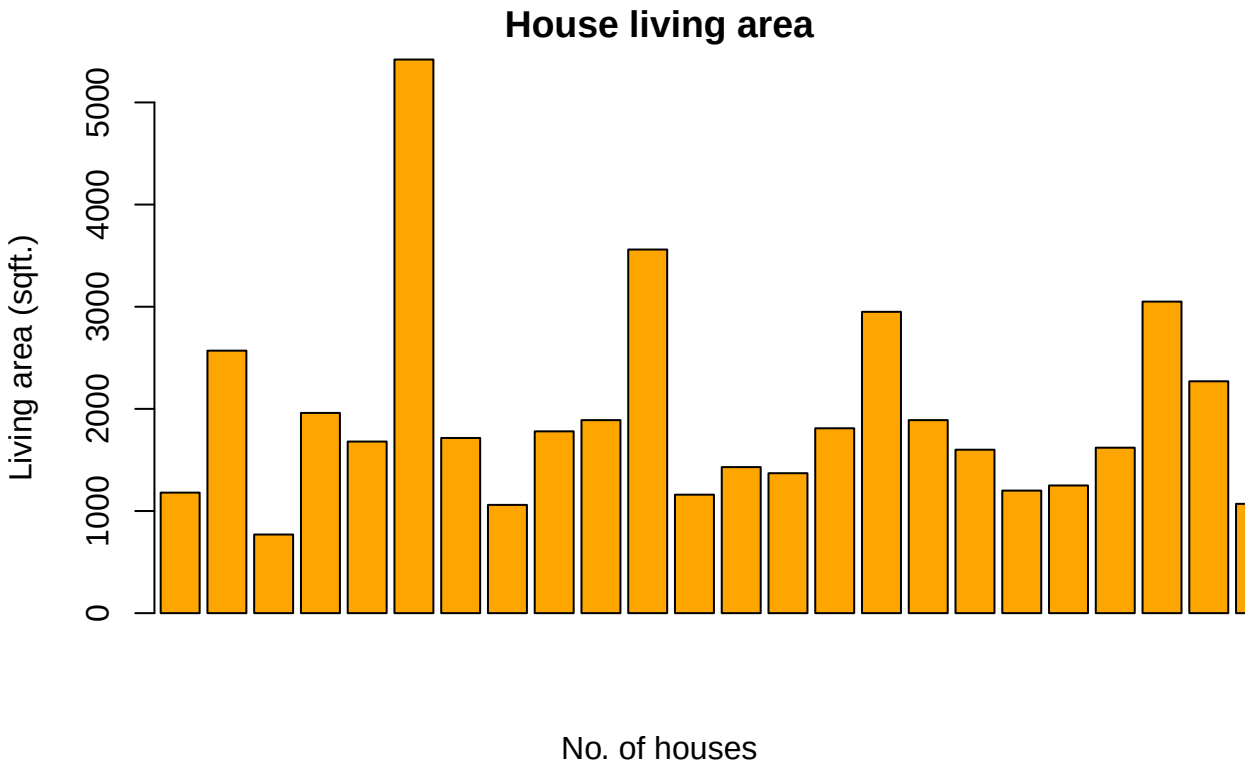
Thus, Number of Bathrooms in a house has more influence than Number of bedrooms.

```
library(ggplot2)
 ggplot(data = data, aes(x = condition, y = price, col=condition)) +
                        geom_point()+
                        coord_cartesian(xlim = c(0, 5))
```



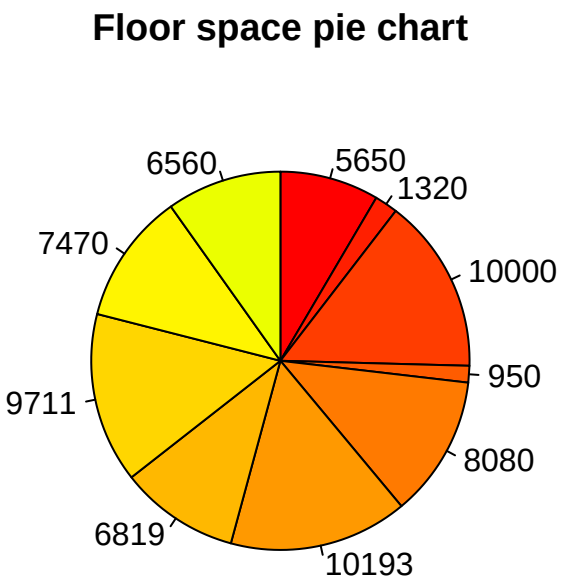
Here We observe that condition of house peaks price at condition degree 3 and then gradually slopes down. Houses are more concentrated at condition degree 1 and 2.

```
barplot(data$sqft_living,xlab="No. of houses",ylab="Living area (sqft.)",col="orange",
        main="House living area",xlim = c(1,25),ylim = c(0,5000))
```



Now we see pie chart of floor space

```
new_data <- head(data,n=10)
pie(new_data$sqft_lot,labels = new_data$sqft_lot, col = rainbow(50),main = "Floor space pie chart",cloc
```



Now lets see details of the house which is costliest and which is cheapest

```
max<-max(data$price)
max1<-subset(data,data$price==max)
print(max1)
```

```
##              id            date   price bedrooms bathrooms sqft_living sqft_lot
## 7253 6762700020 20141013T000000 7700000        6         8       12050    27600
##      floors waterfront view condition grade sqft_above sqft_basement yr_built
## 7253    2.5          0    3         4    13       8570          3480     1910
##      yr_renovated zipcode     lat     long sqft_living15 sqft_lot15
## 7253         1987   98102 47.6298 -122.323          3940       8800
```

The cheapest option from data is

```
min<-min(data$price)
min1<-subset(data,data$price==min)
print(min1)
```

```
##              id            date price bedrooms bathrooms sqft_living sqft_lot
## 1150 3421079032 20150217T000000 75000        1         0         670    43377
##      floors waterfront view condition grade sqft_above sqft_basement yr_built
## 1150      1          0    0         3     3        670             0     1966
##      yr_renovated zipcode     lat     long sqft_living15 sqft_lot15
## 1150            0   98022 47.2638 -121.906          1160      42882
```