房价数据分析

Ryan

August 9, 2015

# 房价数据分析

## 1 数据

本案例数据来源于[Zillow.com](http://zillow.com)，数据收集于2015年8月8日，包含了30个房屋的位置、价格、面积、卧室数量和洗手间数量的信息。

前5个房价信息如下所示:

data <- read.csv("HousePriceData.csv", stringsAsFactors = F)  
head(data)

## Address Price Sq.Feet Beds Baths  
## 1 61 Hiawatha Ave. 409000 1293 3 1  
## 2 37 Wilmot Rd. 579000 1872 3 2  
## 3 25 Marlborough 489900 2040 3 1  
## 4 18 Wildwood Ln 499000 1763 4 2  
## 5 78 Lake St. 399500 1600 4 2  
## 6 47 Pine Vale Rd. 457000 1582 2 2

## 2 变量小结

### 2.1 房屋价格

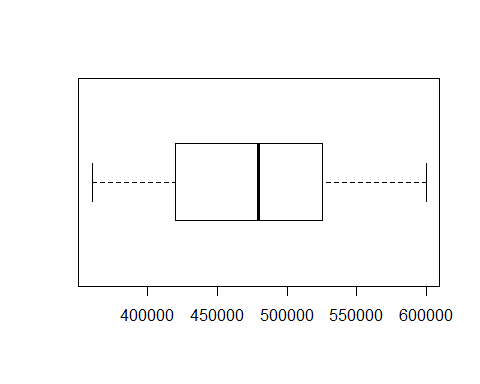
房屋价格分布

summary(data$Price)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 359900 419900 479400 477600 521200 599900

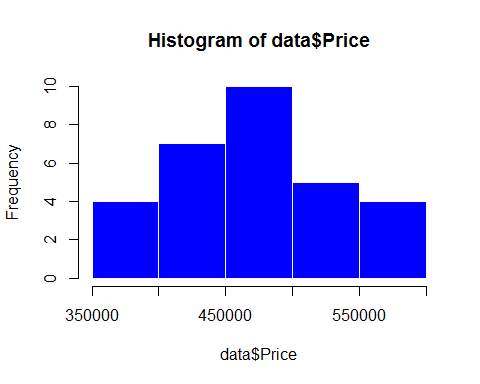
房屋价格分布的盒子图

boxplot(data$Price, horizontal = T)



房屋价格分布的直方图

hist(data$Price, border = 0, col = "blue")



## 2.2 房屋面积分布

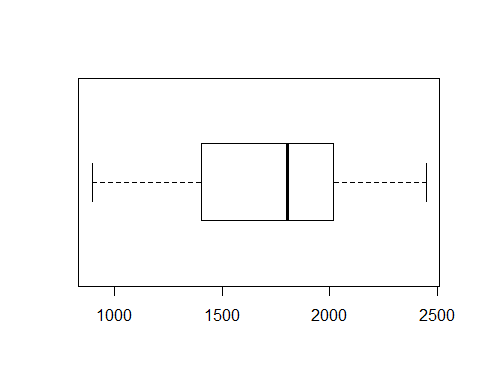
房屋面积分布小结

summary(data$Sq.Feet)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 896 1448 1803 1753 2015 2448

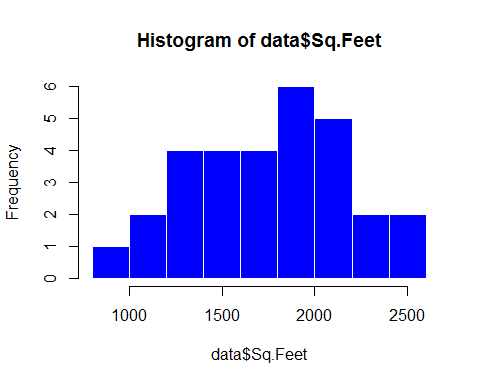
房屋面积分布的盒子图

boxplot(data$Sq.Feet, horizontal = T)



房屋面积分布的直方图

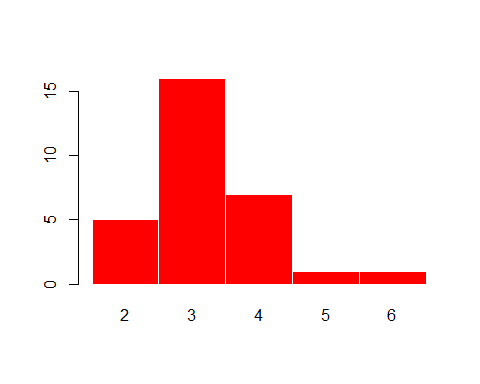
hist(data$Sq.Feet, border = 0, col = "blue")



## 2.3 卧室数量

卧室数量的分布情况

barplot(table(as.factor(data$Beds)), space = 0,border = 0, col = "red")



## 2.4 洗手间数量

洗手间数量的分布情况

table(as.factor(data$Baths))

##   
## 1 1.5 2 2.5 3   
## 7 7 10 4 2

## 2.5 均值汇总

apply(data[,2:5], MARGIN = 2, mean)

## Price Sq.Feet Beds Baths   
## 4.775963e+05 1.753433e+03 3.233333e+00 1.783333e+00

## 2.6 方差、标准差、变异系数汇总

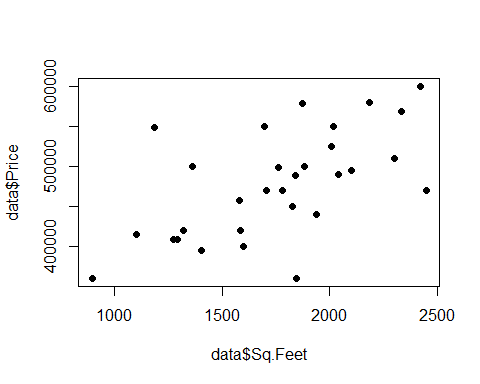
dt <- rbind(apply(data[,2:5], MARGIN = 2, var),  
 apply(data[2:5], MARGIN = 2, sd),  
 apply(data[2:5], MARGIN = 2,function(data){sd(data)/mean(data)}))  
row.names(dt) <- c("var","sd","cv")  
dt

## Price Sq.Feet Beds Baths  
## var 4.425952e+09 1.601946e+05 0.8057471 0.3566092  
## sd 6.652783e+04 4.002432e+02 0.8976342 0.5971676  
## cv 1.392972e-01 2.282626e-01 0.2776188 0.3348604

## 3 变量间关系

房价和面积的关系

plot(data$Sq.Feet, data$Price, pch = 19)



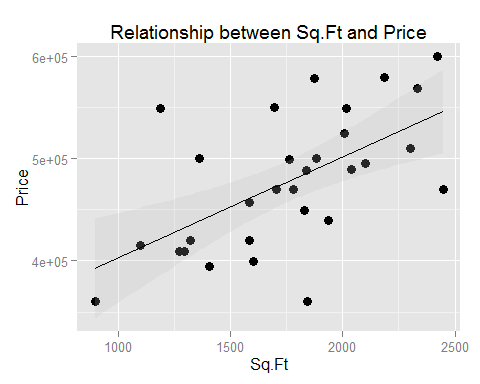
房价和面积的相关系数

cor(data$Sq.Feet, data$Price)

## [1] 0.5965192

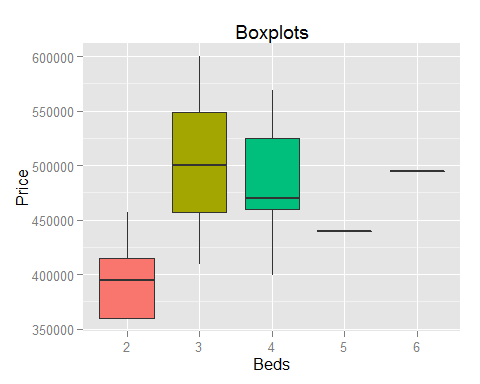
## 4 更好看点的图

ggplot(data = data, aes(x = Sq.Feet, y = Price))+  
 geom\_point(size = 3)+  
 geom\_smooth(method ="lm",alpha = 0.2, colour = 'black', fill = 'grey')+  
 labs(title = "Relationship between Sq.Ft and Price",  
 x = "Sq.Ft",  
 y = "Price")



data$Beds <- as.factor(data$Beds)  
data$Baths <- as.factor(data$Baths)

ggplot(data = data, aes(x = Beds, y = Price), fill = Beds)+  
 geom\_boxplot(aes(fill=Beds))+  
 labs(title="Boxplots",xlab="Beds")+  
 theme(legend.position="none")



ggplot(data = data, aes(x = Baths, y = Price), fill = Baths)+  
 geom\_boxplot(aes(fill=Baths))+  
 labs(title="Boxplots",xlab="Baths")+  
 theme(legend.position="none")

