**# Project Title:  Room rent analysis for hotels**

**# NAME: Harieharaan Duraikannu**

**# EMAIL: harieharaan619@gmail.com**

**# COLLEGE / COMPANY: SSN college of Engineering**

**#Reading the dataset and creating a data frame**

cities42.df<-read.csv(paste("Cities42.csv",sep = ""))

**#Viewing the data**

View(cities42.df)

**#summarize the data**

summary(cities42.df)

CityName Population CityRank IsMetroCity IsTouristDestination

Delhi :2048 Min. : 8096 Min. : 0.00 Min. :0.0000 Min. :0.0000

Jaipur : 768 1st Qu.: 744983 1st Qu.: 2.00 1st Qu.:0.0000 1st Qu.:0.0000

Mumbai : 712 Median : 3046163 Median : 9.00 Median :0.0000 Median :1.0000

Bangalore: 656 Mean : 4416837 Mean :14.83 Mean :0.2842 Mean :0.6972

Goa : 624 3rd Qu.: 8443675 3rd Qu.:24.00 3rd Qu.:1.0000 3rd Qu.:1.0000

Kochi : 608 Max. :12442373 Max. :44.00 Max. :1.0000 Max. :1.0000

(Other) :7816

IsWeekend IsNewYearEve Date HotelName RoomRent

Min. :0.0000 Min. :0.0000 Dec 21 2016:1611 Vivanta by Taj : 32 Min. : 299

1st Qu.:0.0000 1st Qu.:0.0000 Dec 24 2016:1611 Goldfinch Hotel : 24 1st Qu.: 2436

Median :1.0000 Median :0.0000 Dec 25 2016:1611 OYO Rooms : 24 Median : 4000

Mean :0.6228 Mean :0.1244 Dec 28 2016:1611 The Gordon House Hotel: 24 Mean : 5474

3rd Qu.:1.0000 3rd Qu.:0.0000 Dec 31 2016:1611 Apnayt Villa : 16 3rd Qu.: 6299

Max. :1.0000 Max. :1.0000 Dec 18 2016:1608 Bentleys Hotel Colaba : 16 Max. :322500

(Other) :3569 (Other) :13096

StarRating Airport

Min. :0.000 Min. : 0.20

1st Qu.:3.000 1st Qu.: 8.40

Median :3.000 Median : 15.00

Mean :3.459 Mean : 21.16

3rd Qu.:4.000 3rd Qu.: 24.00

Max. :5.000 Max. :124.00

HotelAddress HotelPincode

The Mall, Shimla : 32 Min. : 100025

#2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16 1st Qu.: 221001

121, City Terrace, Walchand Hirachand Marg, Mumbai, Maharashtra : 16 Median : 395003

14-4507/9, Balmatta Road, Near Jyothi Circle, Hampankatta : 16 Mean : 397430

144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16 3rd Qu.: 570001

17, Oliver Road, Colaba, Mumbai, Maharashtra : 16 Max. :7000157

(Other) :13120

HotelDescription FreeWifi FreeBreakfast HotelCapacity HasSwimmingPool

3 : 120 Min. :0.0000 Min. :0.0000 Min. : 0.00 Min. :0.0000

Abc : 112 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.: 16.00 1st Qu.:0.0000

3-star hotel: 104 Median :1.0000 Median :1.0000 Median : 34.00 Median :0.0000

3.5 : 88 Mean :0.9259 Mean :0.6491 Mean : 62.51 Mean :0.3558

4 : 72 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.: 75.00 3rd Qu.:1.0000

(Other) :12728 Max. :1.0000 Max. :1.0000 Max. :600.00 Max. :1.0000

NA's : 8

**#Removing the repeated dates**

cities42.df$Date<-gsub("18-Dec-16", "Dec 18 2016", cities42.df$Date)

cities42.df$Date<-gsub("21-Dec-16", "Dec 21 2016", cities42.df$Date)

cities42.df$Date<-gsub("24-Dec-16", "Dec 24 2016", cities42.df$Date)

cities42.df$Date<-gsub("25-Dec-16", "Dec 25 2016", cities42.df$Date)

cities42.df$Date<-gsub("28-Dec-16", "Dec 28 2016", cities42.df$Date)

cities42.df$Date<-gsub("31-Dec-16", "Dec 31 2016", cities42.df$Date)

cities42.df$Date<-gsub("4-Jan-17", "Jan 04 2017", cities42.df$Date)

cities42.df$Date<-gsub("4-Jan-16", "Jan 04 2017", cities42.df$Date)

cities42.df$Date<-gsub("8-Jan-16", "Jan 08 2017", cities42.df$Date)

cities42.df$Date<-gsub("8-Jan-17", "Jan 08 2017", cities42.df$Date)

cities42.df$Date<-gsub("Jan 4 2017", "Jan 04 2017", cities42.df$Date)

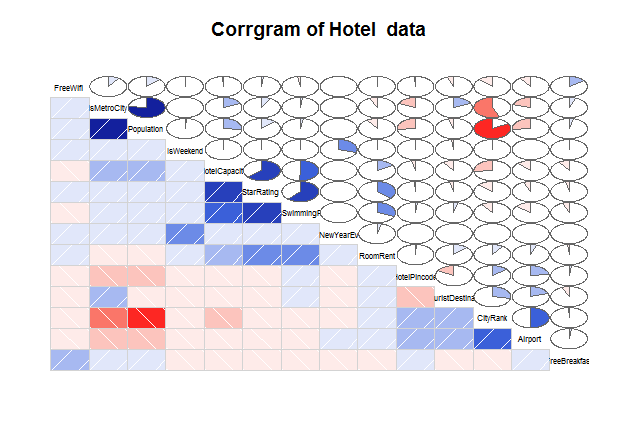
cities42.df$Date<-gsub("Jan 8 2017", "Jan 08 2017", cities42.df$Date)

**#dates to factors for labelling**

cities42.df$Date<-factor(cities42.df$Date)

is.factor(cities42.df$Date)

**#Taking Y = RoomRent, identifying the most relevent predictor variables by correlation corrgram**

**#Corrgram**

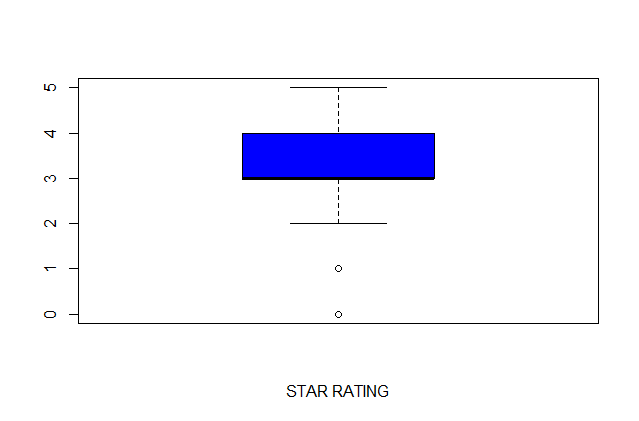
**##through corrgram HasSwimming, StarRating, HotelCapital are very well correlated to RoomRent**

**##so we can take them as predicto**rs

**#boxplot for predictors**

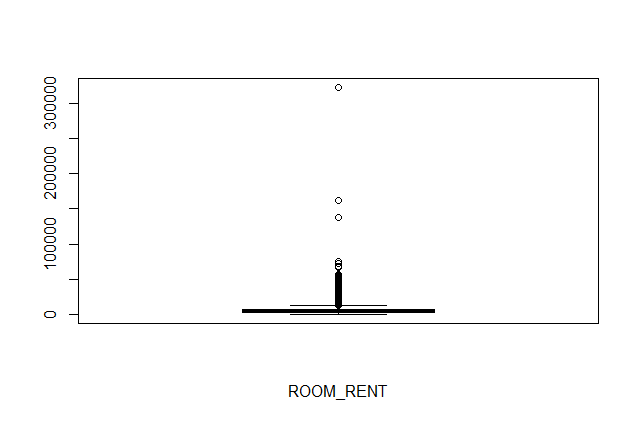
**#boxplot for STAR\_RATING**

boxplot(cities42.df$StarRating,xlab="STAR RATING",col="blue")



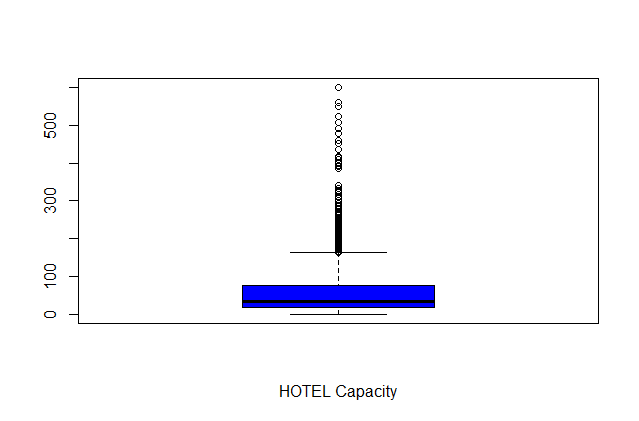
**#boxplot for ROOM\_RENT**

boxplot(cities42.df$RoomRent,xlab="ROOM\_RENT",col="blue")



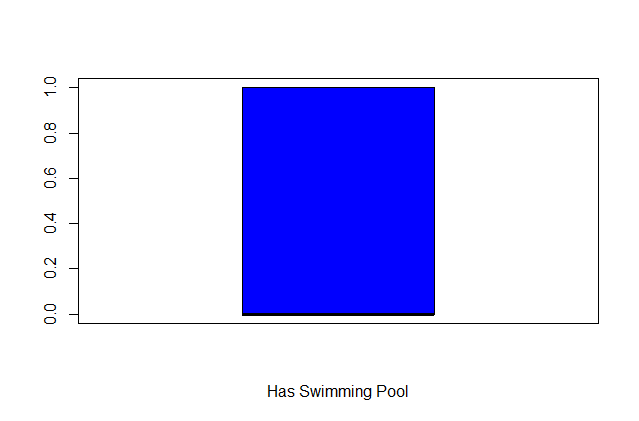
**#boxplot for HOTEL\_CAPACITY**

boxplot(cities42.df$HotelCapacity,xlab="HOTEL Capacity",col="blue")



**#boxplot for SWIMMING POOL**

boxplot(cities42.df$HasSwimmingPool,xlab="Has Swimming Pool",col="blue")



**##Visualizing data for Y as Room rent and X1,X2,X3 as HasSwimmingPool, StarRating and HotelCapacity respectively**

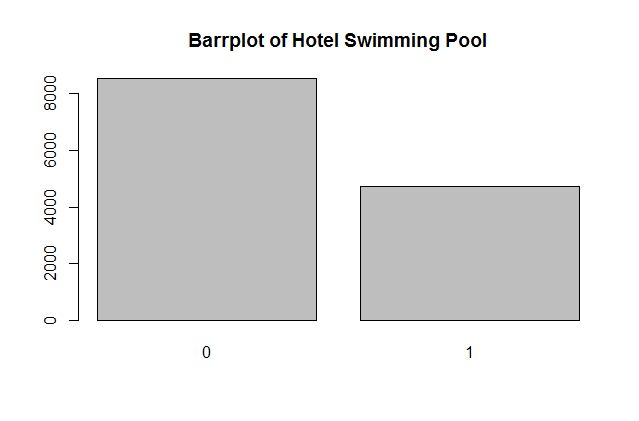
**#Table for HasSwimmingPool**

table(cities42.df$HasSwimmingPool)

0 1

8516 4708

barplot(table(cities42.df$HasSwimmingPool),main="Barrplot of Hotel Swimming Pool")



**#Table for StarRating**

table(cities42.df$StarRating)

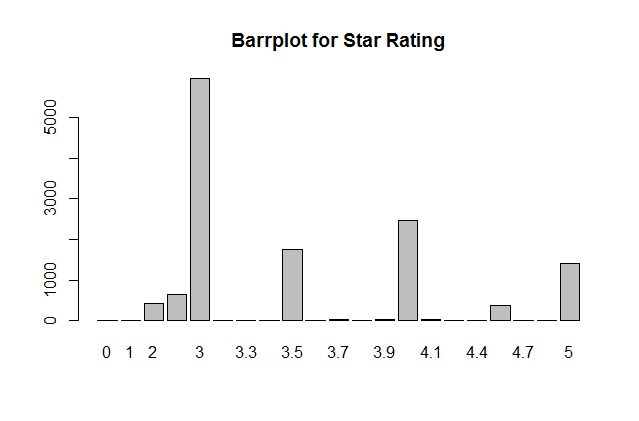
0 1 2 2.5 3 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1 4.3

16 8 432 632 5953 8 16 8 1752 8 24 16 32 2463 24 16

4.4 4.5 4.7 4.8 5

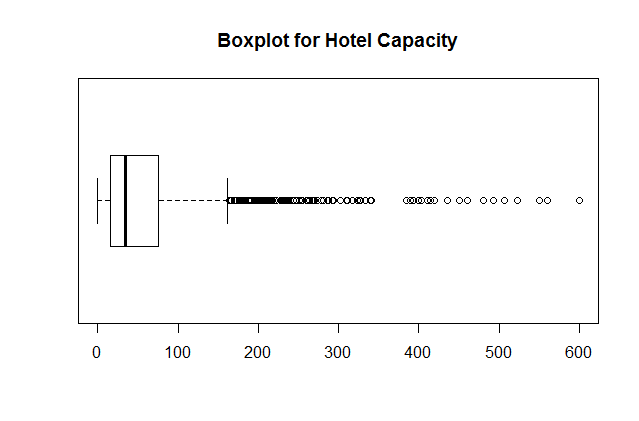
8 376 8 16 1408

barplot(table(cities42.df$StarRating),main = "Barrplot for Star Rating")



**#BoxPlot for HotelCapacity**

boxplot(cities42.df$HotelCapacity, main="Boxplot for Hotel Capacity",horizontal = TRUE)

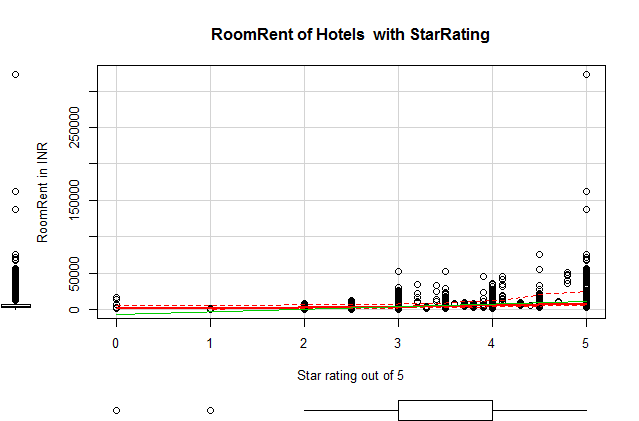


**#Scatterplot for predictor variable**

**#StarRating Vs RoomRent**

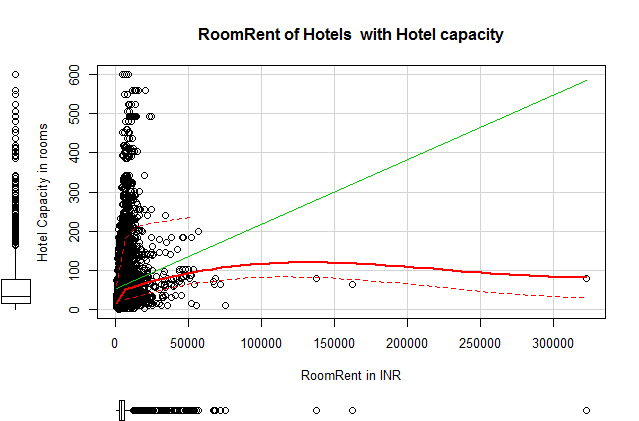
library(car)

scatterplot(cities42.df$StarRating,cities42.df$RoomRent,main="RoomRent of Hotels with StarRating",ylab = "RoomRent in INR", xlab="Star rating out of 5",cex=1.1)



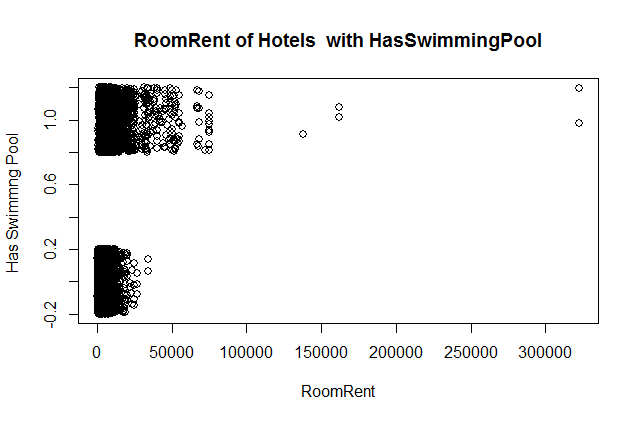
**#RoomRent Vs HotelCapacity**

scatterplot(cities42.df$RoomRent,cities42.df$HotelCapacity,main="RoomRent of Hotels with Hotel capacity",ylab = "Hotel Capacity in rooms", xlab="RoomRent in INR",cex=1.1)



**#RoomRent Vs HasSwimmingPool**

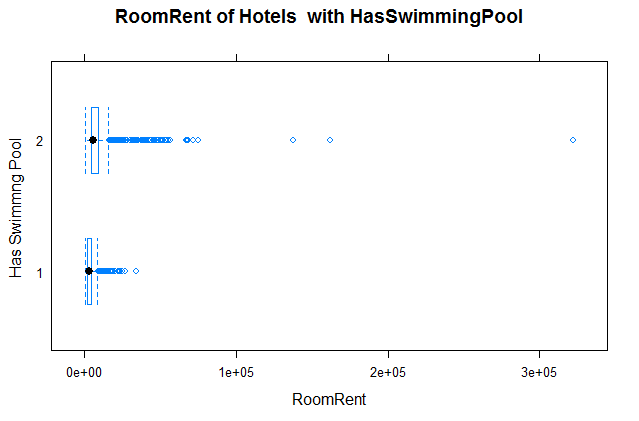
plot(jitter(cities42.df$RoomRent),jitter(cities42.df$HasSwimmingPool),main="RoomRent of Hotels with HasSwimmingPool",ylab = "Has Swimmng Pool ", xlab="RoomRent",cex=1.1)



**#RoomRent Vs HasSwimmingPool**

library(lattice)

bwplot(HasSwimmingPool~RoomRent, data = cities42.df,main="RoomRent of HotelsHasSwimmingPool",ylab = "Has Swimmng Pool ", xlab="RoomRent" )



\

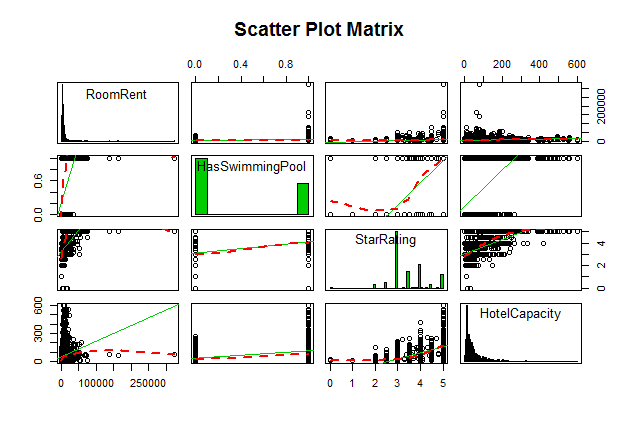
**#Scatterplot matrix**

scatterplotMatrix(

cities42.df[c("RoomRent","HasSwimmingPool","StarRating", "HotelCapacity")],

spread=FALSE, smoother.args=list(lty=2),

main="Scatter Plot Matrix", diagonal = "histogram")



**#Corrgram of Y, x1, x2, x3**

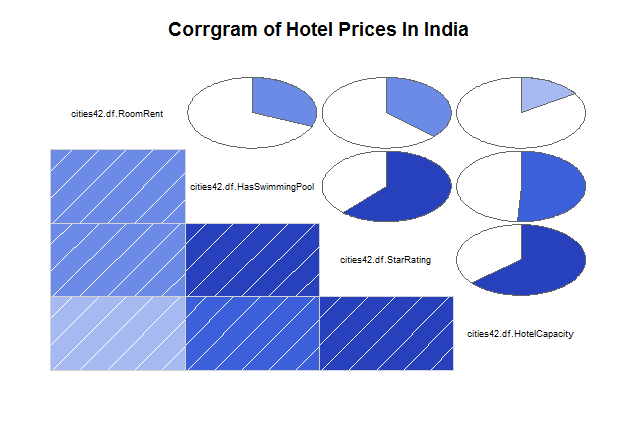
library(corrgram)

xyz<-data.frame(cities42.df$RoomRent, cities42.df$HasSwimmingPool, cities42.df$HotelCapacity, cities42.df$StarRating)

corrgram(xyz, order=TRUE, lower.panel=panel.shade,

upper.panel=panel.pie, text.panel=panel.txt,

main="Corrgram of Hotel Prices In India")



**#Variance-Covariance Matrix for Y, x1, x2, x3**

x<-cities42.df[,c("HasSwimmingPool","StarRating", "HotelCapacity")]

y<-cities42.df[,c("RoomRent")]

cov(x,y)

[,1]

HasSwimmingPool 1094.266

StarRating 2047.165

HotelCapacity 88745.795

var(x,y)

[,1]

HasSwimmingPool 1094.266

StarRating 2047.165

HotelCapacity 88745.795

cor(x,y)

[,1]

HasSwimmingPool 0.3115533

StarRating 0.3693641

HotelCapacity 0.1577820