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

**# NAME: Harish Ravi**

**# EMAIL: hharish001@gmail.com**

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

#TASK 1 and 2

> summary(cities42.df)

CityName Population CityRank IsMetroCity

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

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

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

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

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

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

(Other) :7816

IsTouristDestination IsWeekend IsNewYearEve Date

Min. :0.0000 Min. :0.0000 Min. :0.0000 Dec 21 2016:1611

1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 Dec 24 2016:1611

Median :1.0000 Median :1.0000 Median :0.0000 Dec 25 2016:1611

Mean :0.6972 Mean :0.6228 Mean :0.1244 Dec 28 2016:1611

3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 Dec 31 2016:1611

Max. :1.0000 Max. :1.0000 Max. :1.0000 Dec 18 2016:1608

(Other) :3569

HotelName RoomRent StarRating

Vivanta by Taj : 32 Min. : 299 Min. :0.000

Goldfinch Hotel : 24 1st Qu.: 2436 1st Qu.:3.000

OYO Rooms : 24 Median : 4000 Median :3.000

The Gordon House Hotel: 24 Mean : 5474 Mean :3.459

Apnayt Villa : 16 3rd Qu.: 6299 3rd Qu.:4.000

Bentleys Hotel Colaba : 16 Max. :322500 Max. :5.000

(Other) :13096

Airport

Min. : 0.20

1st Qu.: 8.40

Median : 15.00

Mean : 21.16

3rd Qu.: 24.00

Max. :124.00

HotelAddress

The Mall, Shimla : 32

#2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16

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

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

144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16

17, Oliver Road, Colaba, Mumbai, Maharashtra : 16

(Other) :13120

HotelPincode HotelDescription FreeWifi FreeBreakfast

Min. : 100025 3 : 120 Min. :0.0000 Min. :0.0000

1st Qu.: 221001 Abc : 112 1st Qu.:1.0000 1st Qu.:0.0000

Median : 395003 3-star hotel: 104 Median :1.0000 Median :1.0000

Mean : 397430 3.5 : 88 Mean :0.9259 Mean :0.6491

3rd Qu.: 570001 4 : 72 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :7000157 (Other) :12728 Max. :1.0000 Max. :1.0000

NA's : 8

HotelCapacity HasSwimmingPool

Min. : 0.00 Min. :0.0000

1st Qu.: 16.00 1st Qu.:0.0000

Median : 34.00 Median :0.0000

Mean : 62.51 Mean :0.3558

3rd Qu.: 75.00 3rd Qu.:1.0000

Max. :600.00 Max. :1.0000

**Task 3:y(room rent) =F(star rating , capacity , swimming pool , date , hotel name , city)**

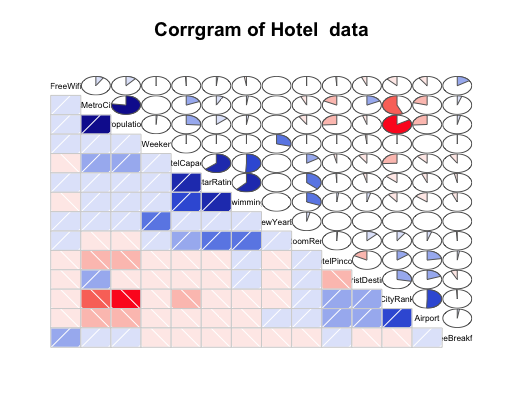
**Task 4: Dependent variable = roomrent;**

**#TASK 5**

# Three most important predictors

##from the corrgram it is understood that HasSwimming, StarRating, HotelCapital are very well correlated to RoomRent

##therefore we take it as predictors



**# Visualize Y, x1, x2, x3 individually. Ignore other variables for now.**

> table(cities42.df$HasSwimmingPool)

0 1

8524 4708

> table(cities42.df$StarRating)

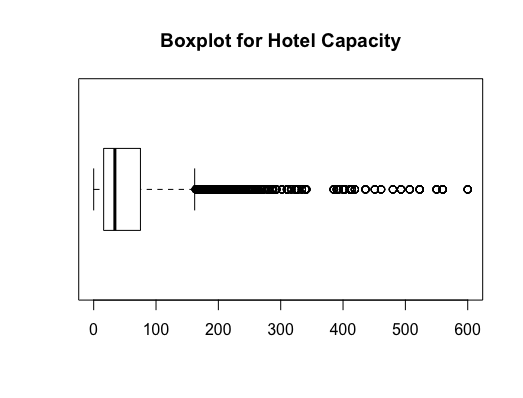
0 1 2 2.5 3 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1 4.3

16 8 440 632 5953 8 16 8 1752 8 24 16 32 2463 24 16

4.4 4.5 4.7 4.8 5

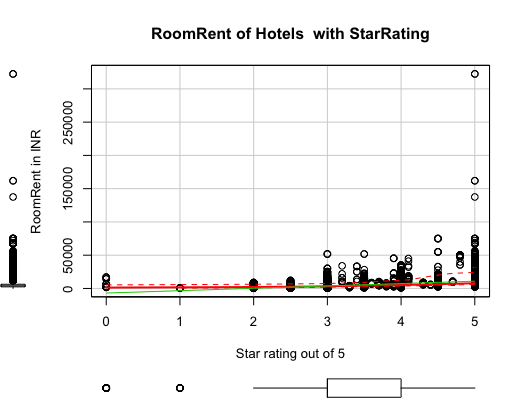
8 376 8 16 1408

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

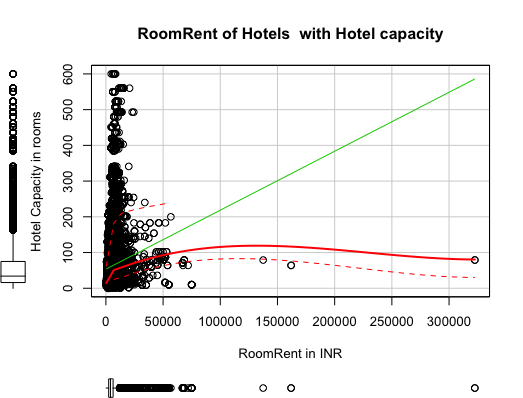


##Task 7: **Draw Scatter Plots to understand how are the variables correlated pair-wise**

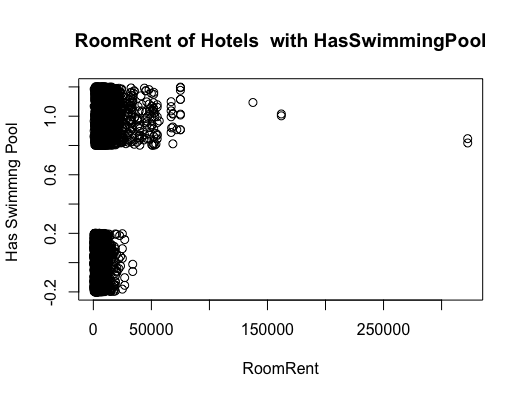
> scatterplot(cities42.df$StarRating,cities42.df$RoomRent,main="RoomRent of Hotels with StarRating",ylab = "RoomRent in INR", xlab="Star rating out of 5",cex=1.1



> 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)



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



library(lattice)

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



###8 **Draw a Corrgram of Y, x1, x2, x3  (Ignore other variables for now)**



## **9 .Create a Variance-Covariance Matrix for Y, x1, x2, x3**

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

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

> #Variance and covariance matrix

> cov(x,y)

[,1]

HasSwimmingPool 1094.202

StarRating 2048.375

HotelCapacity 88753.413

> var(x,y)

[,1]

HasSwimmingPool 1094.202

StarRating 2048.375

HotelCapacity 88753.413

> cor(x,y)

[,1]

HasSwimmingPool 0.3116577

StarRating 0.3693734

HotelCapacity 0.1578733