Hotel Room Pricing in Indian Cities

Aman Gupta
February 23, 2018

Project Title - Hotel Room Pricing in Indian Cities

Name - Aman Gupta

Email - amangupta151299@gmail.com

College - IIT Kanpur

Introduction

In this project we will analyse the pricing strategy of hotels in the Indian hotel industry and effect of some factors on room rent. The objective of this project is to identify the factors that matter the most. For example, internal factors like free breakfast, free wi-fi and presence of a swimming pool may be responsible for increase in room rent but external factors like Star Rating of hotel and whether the city is metro city or not, will be more significant in determining the room rent.

Overview

The specific objective of this study was to investigate the pricing strategy employed by hotels located in different cities of India. This study analyzed effect of different factors on room rent. By regression analysis, it is discovered that factors like StarRating, IsMetroCity, IsTouristDestination and IsNewYearEve are quite effective for room rent in hotels while factors like distance from airport do not have considerable effect on room rent.

Data

For this study, we collected data from hotel website(http://www.hotel.in/). This dataset includes details of different hotels in 42 different cities in India. Data contains room rent of each hotel on different occasions, e.g., on a weekend, on a new year eve and on a normal working day. Also factors like distance from airport, star rating, hotel capacity, free wi-fi, free breakfast and swimming pool are clear in each row of data. Room rent also depends on city rank, whether the city is metro city/populated or not, whether city is a tourist destination or not. So all these details are available from dataset which makes our analysis more accurate and real.

Read the data

```
hotel <- read.csv(paste("Cities42.csv", sep=""))
View(hotel)

dim(hotel)

## [1] 13232   19
```

Attach the dataframe

```
attach(hotel)
```

Summarize the data

```
summary(hotel)

##      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
##

library(psych)
describe(hotel)

vars   n   mean   sd median   trimmed
CityName*   1 13232   18.07   11.72   16   17.29
Population   2 13232 4416836.87 4258386.00 3046163
4040816.22
CityRank   3 13232   14.83   13.51   9   13.30
IsMetroCity   4 13232   0.28   0.45   0   0.23
IsTouristDestination   5 13232   0.70   0.46   1   0.75
IsWeekend   6 13232   0.62   0.48   1   0.65
IsNewYearEve   7 13232   0.12   0.33   0   0.03
Date*   8 13232   14.26   2.82   14   14.39
HotelName*   9 13232   841.19   488.16   827   841.18
RoomRent   10 13232   5473.99   7333.12   4000   4383.33
StarRating   11 13232   3.46   0.76   3   3.40
Airport   12 13232   21.16   22.76   15   16.39
HotelAddress*   13 13232   1202.69   582.00   1261   1233.41
HotelPincode   14 13232 397430.26 259837.50 395003
388540.47
HotelDescription*   15 13224   582.40   362.10   569   576.47
FreeWifi   16 13232   0.93   0.26   1   1.00
FreeBreakfast   17 13232   0.65   0.48   1   0.69
HotelCapacity   18 13232   62.51   76.66   34   46.03
HasSwimmingPool   19 13232   0.36   0.48   0   0.32

```
##                        mad       min      max      range skew
## CityName*            11.86      1.0       42       41.0 0.48
## Population      3846498.95   8096.0 12442373 12434277.0 0.68
## CityRank            11.86      0.0       44       44.0 0.69
## IsMetroCity          0.00      0.0        1        1.0 0.96
## IsTouristDestination 0.00      0.0        1        1.0 -0.86
## IsWeekend            0.00      0.0        1        1.0 -0.51
## IsNewYearEve         0.00      0.0        1        1.0 2.28
## Date*                2.97      1.0       20       19.0 -1.05
## HotelName*         641.97      1.0     1670     1669.0 0.01
## RoomRent          2653.85    299.0   322500   322201.0 16.75
## StarRating           0.74      0.0        5        5.0 0.48
## Airport             11.12      0.2      124      123.8 2.73
## HotelAddress*      668.65      1.0     2108     2107.0 -0.37
## HotelPincode    257975.37 100025.0  7000157  6900132.0 9.99
## HotelDescription*  471.47      1.0     1226     1225.0 0.11
## FreeWifi             0.00      0.0        1        1.0 -3.25
## FreeBreakfast        0.00      0.0        1        1.0 -0.62
## HotelCapacity       28.17      0.0      600      600.0 2.95
## HasSwimmingPool      0.00      0.0        1        1.0 0.60
##                    kurtosis       se
## CityName*             -0.88     0.10
## Population            -1.08 37019.65
## CityRank             -0.76     0.12
## IsMetroCity          -1.08     0.00
## IsTouristDestination -1.26     0.00
## IsWeekend            -1.74     0.00
## IsNewYearEve          3.18     0.00
## Date*                 2.93     0.02
## HotelName*           -1.25     4.24
## RoomRent            582.06    63.75
## StarRating            0.25     0.01
## Airport               7.89     0.20
## HotelAddress*        -0.88     5.06
## HotelPincode        249.76  2258.86
## HotelDescription*    -1.25     3.15
## FreeWifi              8.57     0.00
## FreeBreakfast        -1.61     0.00
## HotelCapacity        11.39     0.67
## HasSwimmingPool      -1.64     0.00
```

Data Types

str(hotel)

```
## 'data.frame':    13232 obs. of  19 variables:
##  $ CityName        : Factor w/ 42 levels "Agra","Ahmedabad",..: 26 26 26
26 26 26 26 26 26 26 ...
##  $ Population       : int  12442373 12442373 12442373 12442373
12442373 12442373 12442373 12442373 12442373 12442373 ...
##  $ CityRank         : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ IsMetroCity       : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ IsTouristDestination: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ IsWeekend         : int  1 0 1 1 0 1 0 1 1 0 ...
##  $ IsNewYearEve      : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Date              : Factor w/ 20 levels "18-Dec-16","21-Dec-16",..: 11 12
13 14 15 16 17 18 11 12 ...
##  $ HotelName         : Factor w/ 1670 levels "14 Square Amanora",..: 1635
1635 1635 1635 1635 1635 1635 1635 1409 1409 ...
##  $ RoomRent          : int  12375 10250 9900 10350 12000 11475 11220
9225 6800 9350 ...
##  $ StarRating        : num  5 5 5 5 5 5 5 5 4 4 ...
##  $ Airport           : num  21 21 21 21 21 21 21 21 20 20 ...
##  $ HotelAddress      : Factor w/ 2108 levels " H.P. High Court Mall Road,
Shimla",..: 925 928 930 933 935 937 940 941 695 744 ...
##  $ HotelPincode      : int  400005 400006 400007 400008 400009
400010 400011 400012 400039 400040 ...
##  $ HotelDescription   : Factor w/ 1226 levels "#NAME?","10 star hotel
near Queensroad, Amritsar",..: 1030 1030 1030 1030 1030 1030 1030 1030
1006 1006 ...
##  $ FreeWifi          : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ FreeBreakfast     : int  0 0 0 0 0 0 0 0 1 1 ...
##  $ HotelCapacity     : int  287 287 287 287 287 287 287 287 28 28 ...
##  $ HasSwimmingPool   : int  1 1 1 1 1 1 1 1 0 0 ...
```

Variable Analysis

One way contingency tables
```
mytable <- with(hotel,table(CityName))
mytable

## CityName
##        Agra      Ahmedabad        Amritsar       Bangalore
##         432            424             136             656
##    Bhubaneswar      Chandigarh         Chennai      Darjeeling
##         120            336             416             136
##       Delhi        Gangtok           Goa        Guwahati
##        2048            128             624              48
##      Haridwar       Hyderabad          Indore          Jaipur
##          48            536             160             768
##      Jaisalmer        Jodhpur          Kanpur           Kochi
##         264            224              16             608
##       Kolkata        Lucknow         Madurai          Manali
##         512            128             112             288
##      Mangalore         Mumbai          Munnar          Mysore
##         104            712             328             160
##      Nainital          Ooty       Panchkula          Pune
##         144            136              64             600
##        Puri         Rajkot       Rishikesh          Shimla
##          56            128              88             280
##       Srinagar         Surat Thiruvanthipuram        Thrissur
```

```
##            40           80          392            32
##      Udaipur       Varanasi
##         456           264
```

mytable1 <- with(hotel,table(FreeWifi))
mytable1

```
## FreeWifi
##    0    1
##  981 12251
```

mytable2 <- with(hotel,table(FreeBreakfast))
mytable2

```
## FreeBreakfast
##    0    1
## 4643 8589
```

mytable3 <-with(hotel,table(HasSwimmingPool))
mytable3

```
## HasSwimmingPool
##    0    1
## 8524 4708
```

mytable4 <- with(hotel, table(StarRating))
mytable4

```
## 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
##   16    8  440  632 5953    8   16    8 1752    8   24   16   32 2463   24
##  4.3  4.4  4.5  4.7  4.8    5
##   16    8  376    8   16 1408
```

mytable5 <- with(hotel, table(IsMetroCity))
mytable5

```
## IsMetroCity
##    0    1
## 9472 3760
```

mytable6 <- with(hotel, table(IsTouristDestination))
mytable6

```
## IsTouristDestination
##    0    1
## 4007 9225
```

Two way contingency tables
mytable <- xtabs(~ StarRating + FreeBreakfast, data=hotel)
mytable

```
##          FreeBreakfast
## StarRating    0    1
##      0    16    0
##      1     0    8
##      2   216  224
##      2.5 296  336
##      3  1789 4164
##      3.2   0    8
##      3.3   8    8
##      3.4   0    8
##      3.5 661 1091
##      3.6   8    0
##      3.7   0   24
##      3.8   8    8
##      3.9  16   16
##      4   783 1680
##      4.1   0   24
##      4.3  16    0
##      4.4   0    8
##      4.5 224  152
##      4.7   8    0
##      4.8   0   16
##      5   594  814
```

mytable1<- xtabs(~ StarRating + FreeWifi, data=hotel)
mytable1

```
##          FreeWifi
## StarRating    0    1
##      0     0   16
##      1     0    8
##      2    80  360
##      2.5 104  528
##      3   336 5617
##      3.2   0    8
##      3.3   0   16
##      3.4   0    8
##      3.5  96 1656
##      3.6   0    8
##      3.7   0   24
##      3.8   0   16
##      3.9   0   32
##      4   231 2232
##      4.1   0   24
##      4.3   0   16
##      4.4   0    8
##      4.5  24  352
##      4.7   0    8
##      4.8   0   16
##      5   110 1298
```

```
mytable3 <- xtabs(~ IsMetroCity+IsTouristDestination, data=hotel)
mytable3

##            IsTouristDestination
## IsMetroCity    0    1
##          0 3352 6120
##          1  655 3105

library(gmodels)
CrossTable(hotel$FreeWifi,hotel$FreeBreakfast)

##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  13232
##
##
##              | hotel$FreeBreakfast
## hotel$FreeWifi |        0 |        1 | Row Total |
## ---------------|-----------|-----------|-----------|
##             0 |      606 |      375 |      981 |
##               |  199.074 |  107.614 |          |
##               |    0.618 |    0.382 |    0.074 |
##               |    0.131 |    0.044 |          |
##               |    0.046 |    0.028 |          |
## ---------------|-----------|-----------|-----------|
##             1 |     4037 |     8214 |    12251 |
##               |   15.941 |    8.617 |          |
##               |    0.330 |    0.670 |    0.926 |
##               |    0.869 |    0.956 |          |
##               |    0.305 |    0.621 |          |
## ---------------|-----------|-----------|-----------|
##   Column Total |     4643 |     8589 |    13232 |
##               |    0.351 |    0.649 |          |
## ---------------|-----------|-----------|-----------|
##
##

CrossTable(hotel$FreeWifi,hotel$HasSwimmingPool)

##
##
```

```
##    Cell Contents
## |-----------------------|
## |                     N |
## | Chi-square contribution |
## |           N / Row Total |
## |             N / Col Total |
## |           N / Table Total |
## |-----------------------|
##
##
## Total Observations in Table:  13232
##
##
##               | hotel$HasSwimmingPool
## hotel$FreeWifi |        0 |        1 | Row Total |
## ---------------|----------|----------|-----------|
##            0 |      592 |      389 |      981 |
##              |    2.526 |    4.574 |          |
##              |    0.603 |    0.397 |    0.074 |
##              |    0.069 |    0.083 |          |
##              |    0.045 |    0.029 |          |
## ---------------|----------|----------|-----------|
##            1 |     7932 |     4319 |    12251 |
##              |    0.202 |    0.366 |          |
##              |    0.647 |    0.353 |    0.926 |
##              |    0.931 |    0.917 |          |
##              |    0.599 |    0.326 |          |
## ---------------|----------|----------|-----------|
##   Column Total |     8524 |     4708 |    13232 |
##              |    0.644 |    0.356 |          |
## ---------------|----------|----------|-----------|
##
##
```

Visualization

```r
boxplot(hotel$StarRating,
    xlab="Star Rating of the hotel",
    main="Box plot of Star Rating of hotel",
    col = "yellow",
    horizontal = TRUE)
```

Box plot of Star Rating of hotel

Star Rating of the hotel
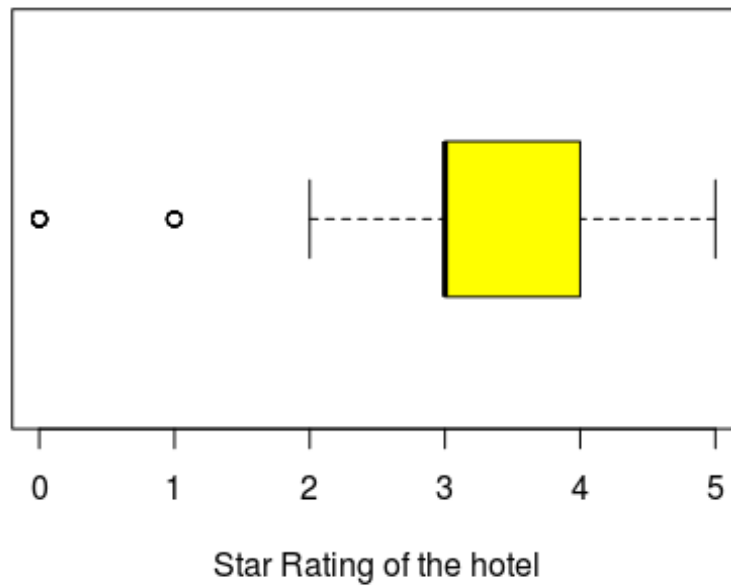
```
boxplot(hotel$Airport,
     xlab="Distance between Hotel and closest major Airport(in km)",
     main="Box plot of Airport Distance of hotel",
     col = "yellow",
     horizontal = TRUE)
```

Box plot of Airport Distance of hotel
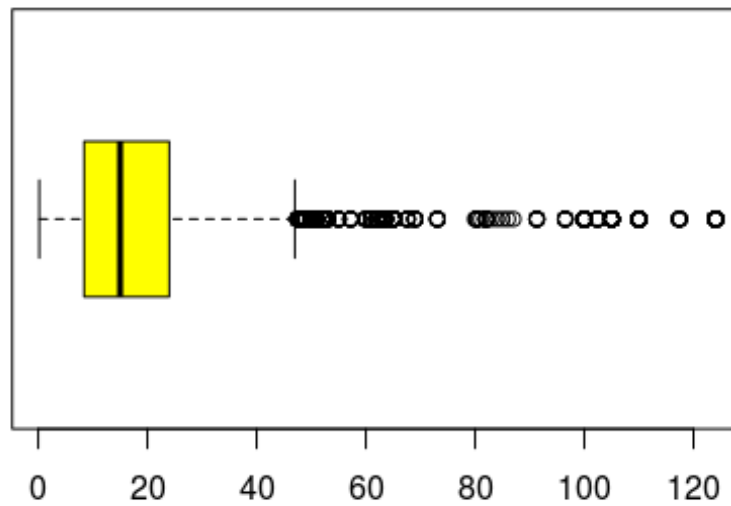
Distance between Hotel and closest major Airport(in km)

```r
boxplot(hotel$HotelCapacity,
    xlab="Hotel Capacity",
    main="Box plot of Hotel Capacity",
    col = "yellow",
    horizontal = TRUE)
```

Box plot of Hotel Capacity

Hotel Capacity
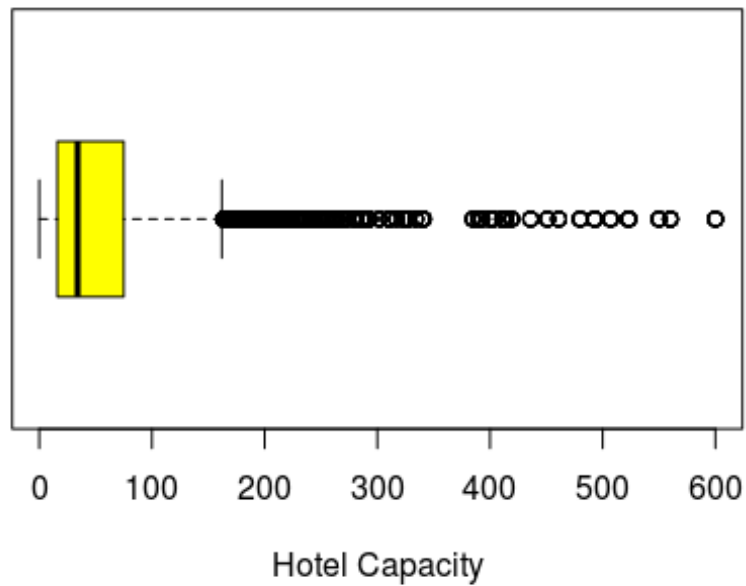
```
boxplot(hotel$StarRating ~ hotel$FreeBreakfast, horizontal=TRUE,
        ylab="breakfast avalability", xlab="Star ratings", las=1,
        main="Analysis of star rating and breakfast avalability",
        col=c("red","green")
        )
```

Analysis of star rating and breakfast avalability
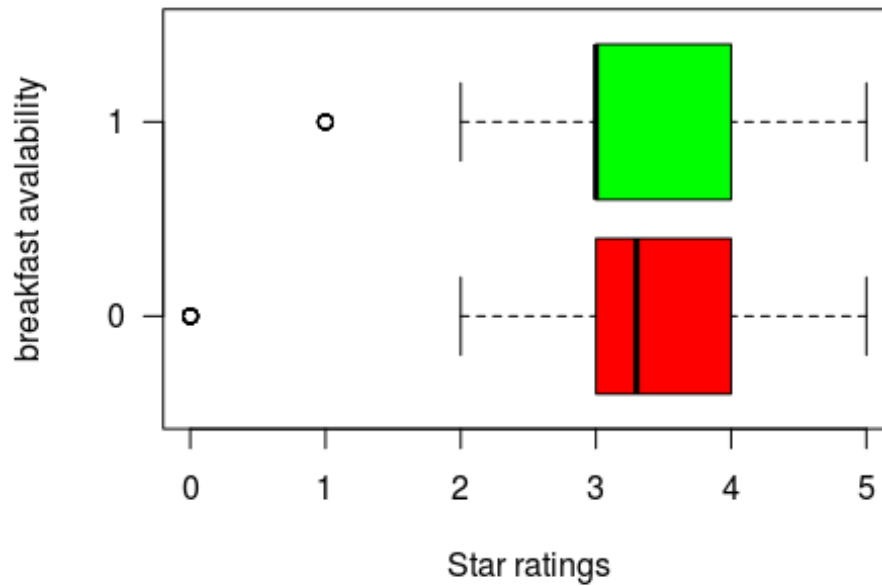

```
boxplot(hotel$RoomRent ~ hotel$IsMetroCity, horizontal=TRUE,
      ylab="City(metro=1,other=0)", xlab="Room rent", las=1,
      main="Analysis of type of city and room rent of hotels",
      col=c("red","blue"),
      ylim=c(0,30000)
      )
```

Analysis of type of city and room rent of hotels
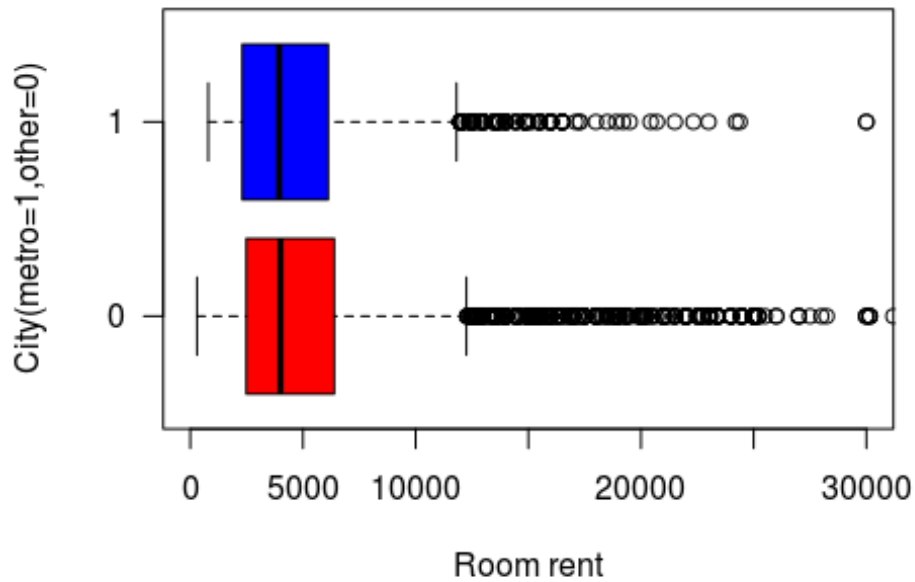

```
boxplot(hotel$RoomRent ~ hotel$IsNewYearEve, horizontal=TRUE,
    ylab="IsNewYearEve", xlab="Room rent", las=1,
    main="Analysis of room rent of hotels with New Year",
    ylim = c(0,30000),
    col = c("yellow","pink")
    )
```

Analysis of room rent of hotels with New Year
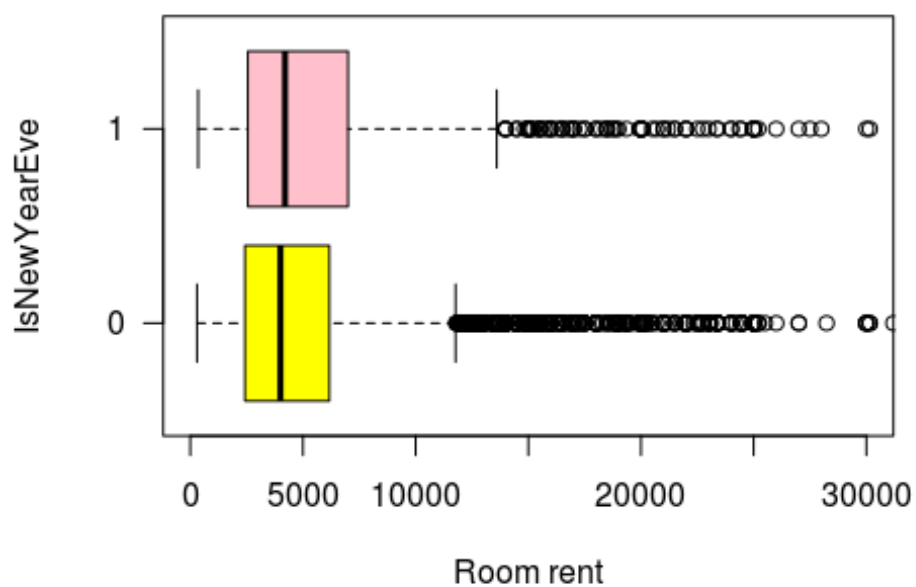

```
hist(StarRating , col = "green")
```

Histogram of StarRating
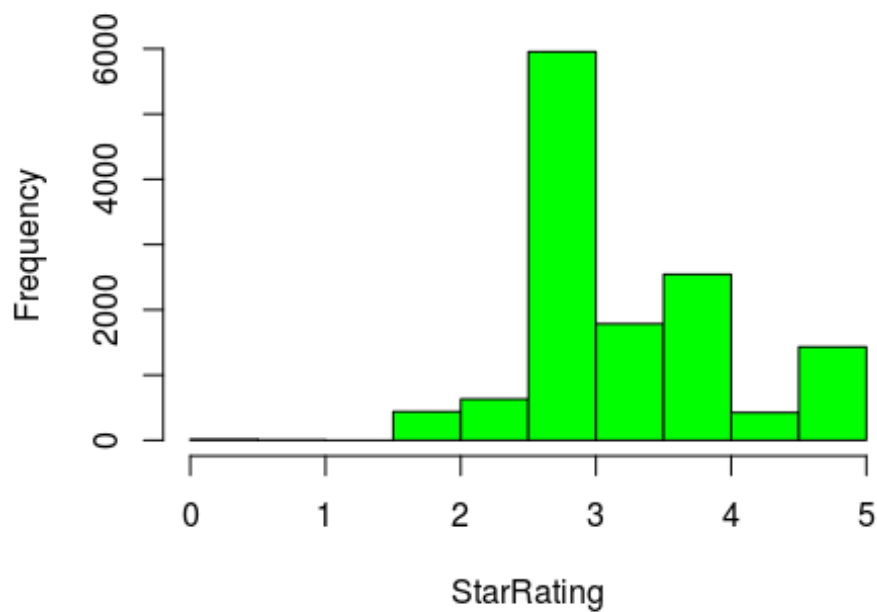
```r
hist(Airport , col = "magenta" , xlab = "Distance between Hotel and closest
major Airport in Km" , main = " Airport to Hotel")
```

Airport to Hotel

Distance between Hotel and closest major Airport in Km
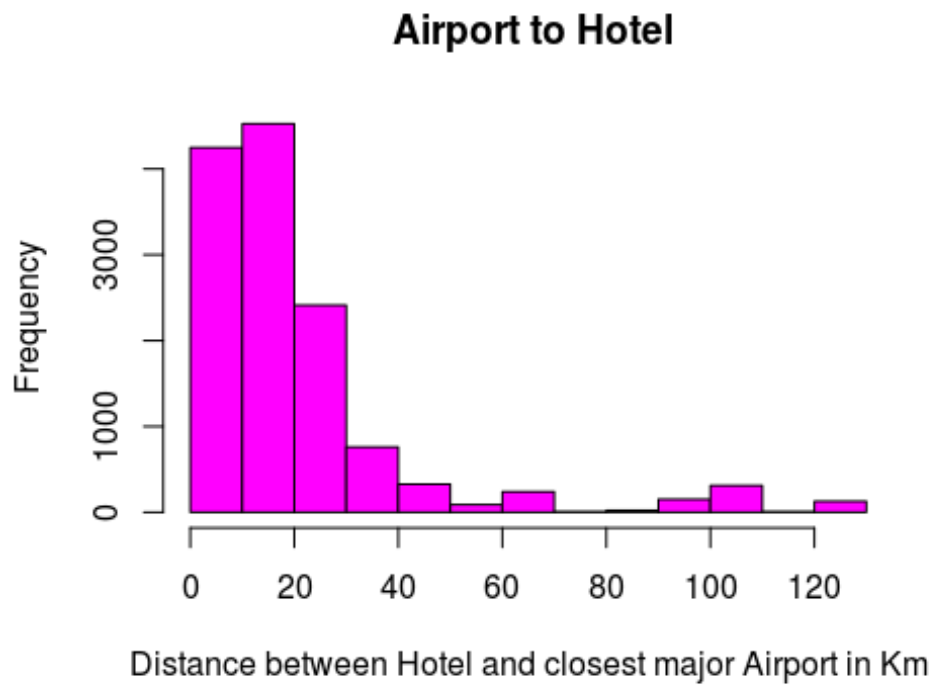
```r
hist(HotelCapacity , col = "blue")
```

Histogram of HotelCapacity
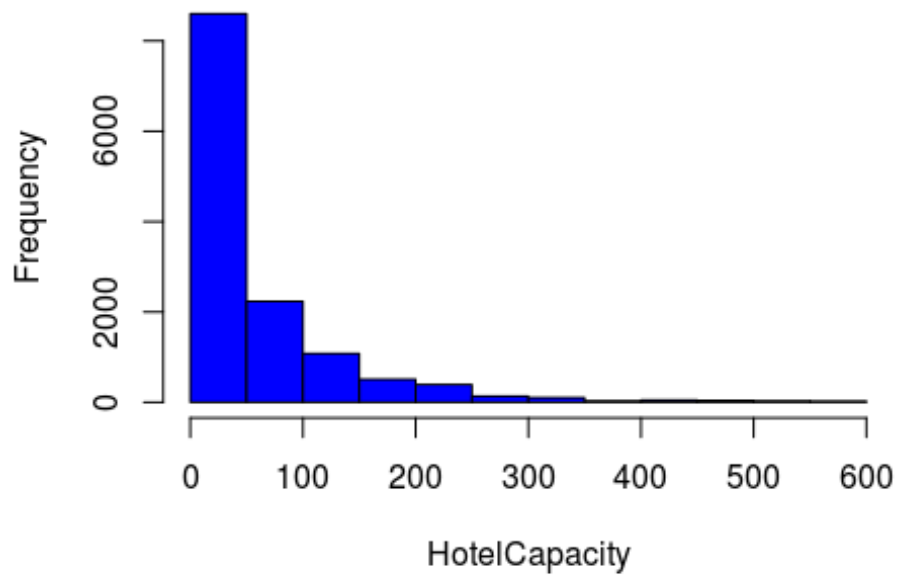

```
hist(hotel$RoomRent,
    main="Analysis of room rents of hotels",
    xlab="Rents of room",ylab="Relative frequency",
    xlim = c(0,30000),breaks=30,
    col="green")
```

Analysis of room rents of hotels
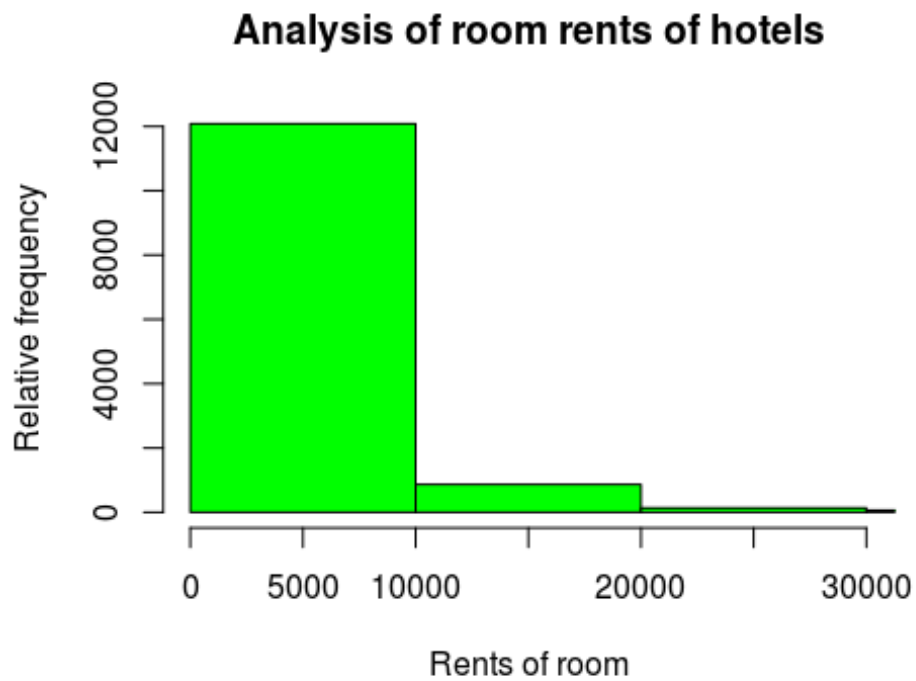

```
hotel$FreeWifi=factor(hotel$FreeWifi, levels=c(0,1), labels=c("No","Yes"))
plot(hotel$FreeWifi,col="yellow",main="Has Wifi?")
```

Has Wifi?
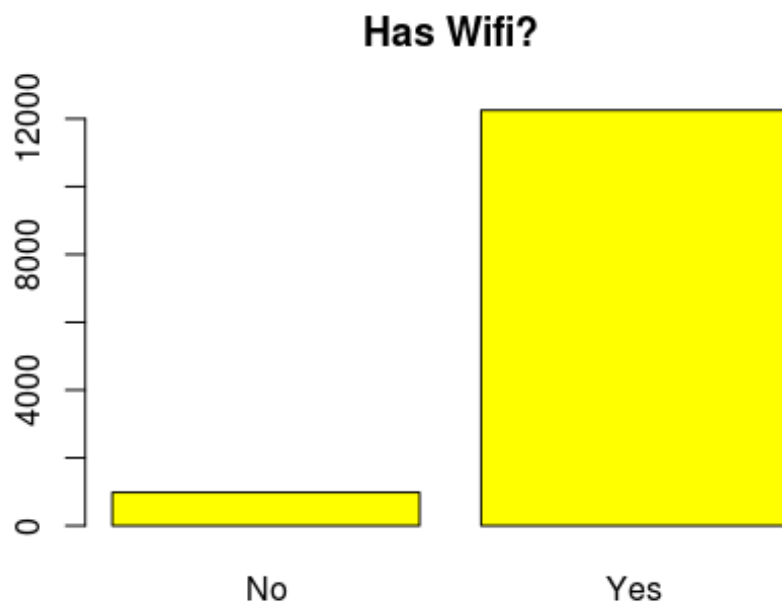
```
hotel$FreeBreakfast=factor(hotel$FreeBreakfast, levels=c(0,1),
labels=c("No","Yes"))
plot(hotel$FreeBreakfast,col="blue",main="Has Free Breakfast?")
```

Has Free Breakfast?
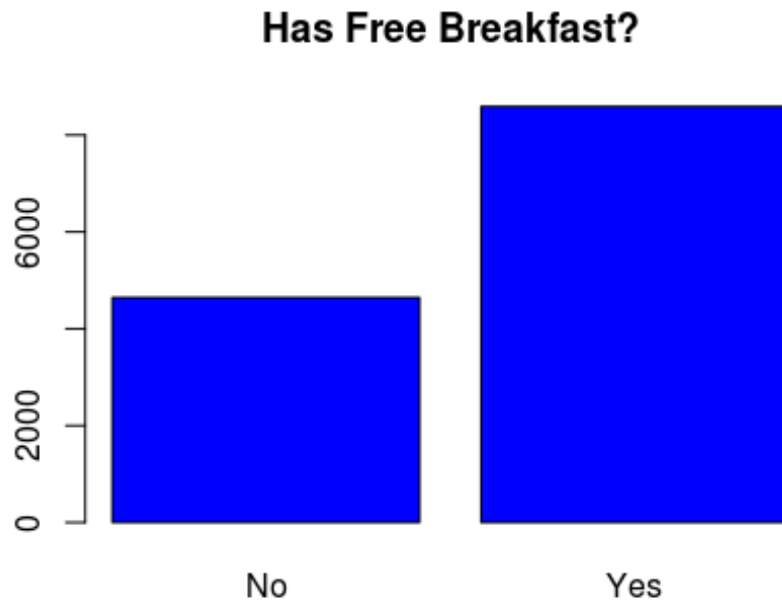

```
hotel$HasSwimmingPool=factor(hotel$HasSwimmingPool, levels=c(0,1),
labels=c("No","Yes"))
plot(hotel$HasSwimmingPool,col="green",main="Swimming pool?")
```

Swimming pool?
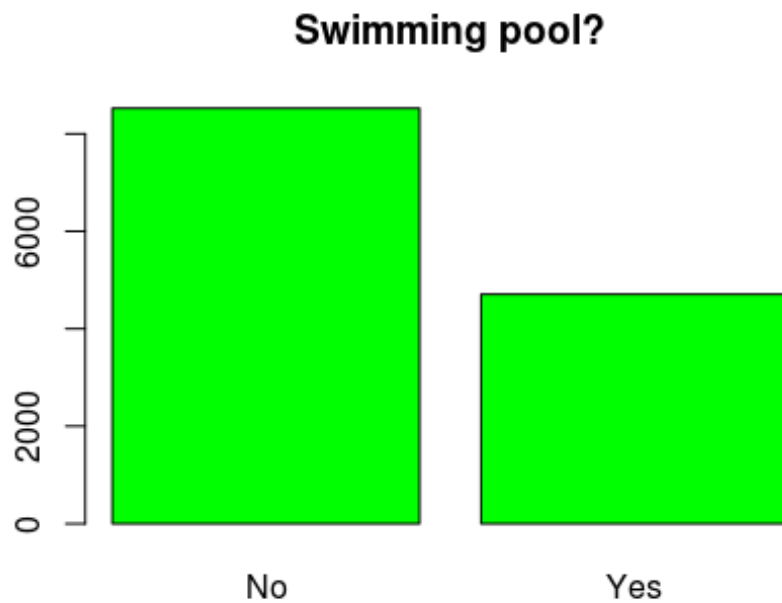
Scatterplots

```r
library(car)

##
## Attaching package: 'car'

## The following object is masked from 'package:psych':
##
##     logit

scatterplot(RoomRent~StarRating,    data=hotel,
        spread=FALSE, smoother.args=list(lty=2),
        main="Scatter plot of Star Rating vs Room rent",
        ylab="Room Rent",
        xlab="Star Rating")
```
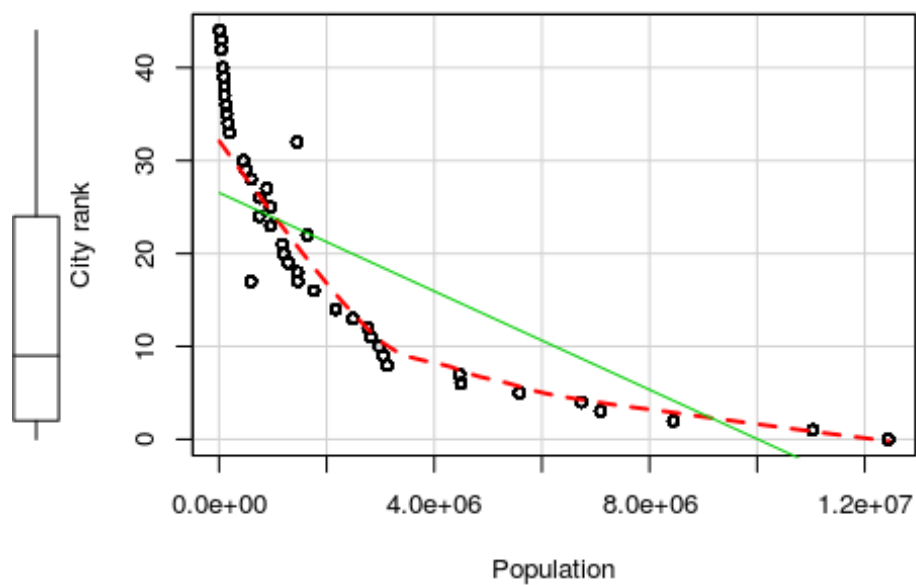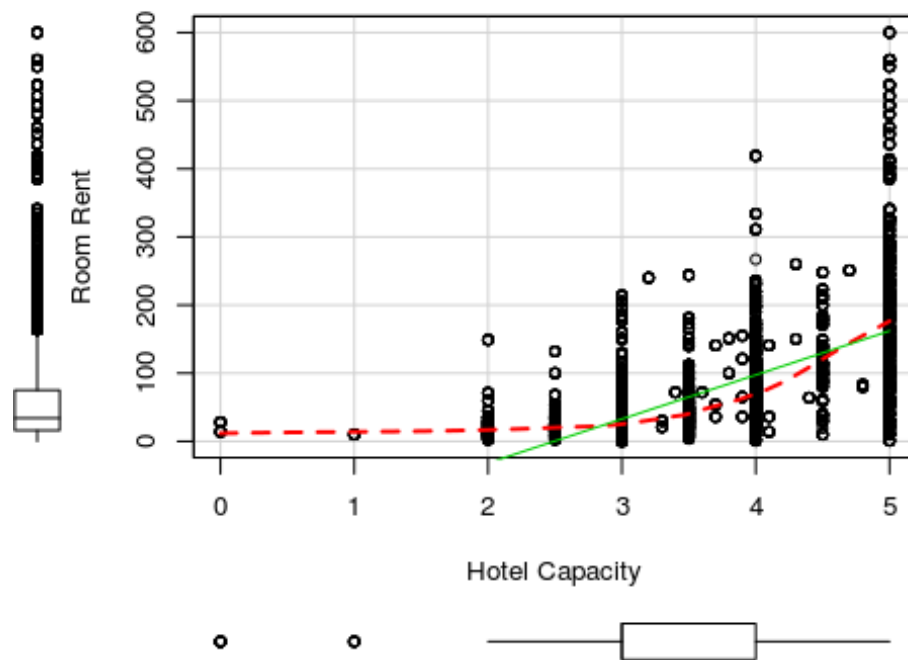
Scatter plot of Star Rating vs Room rent
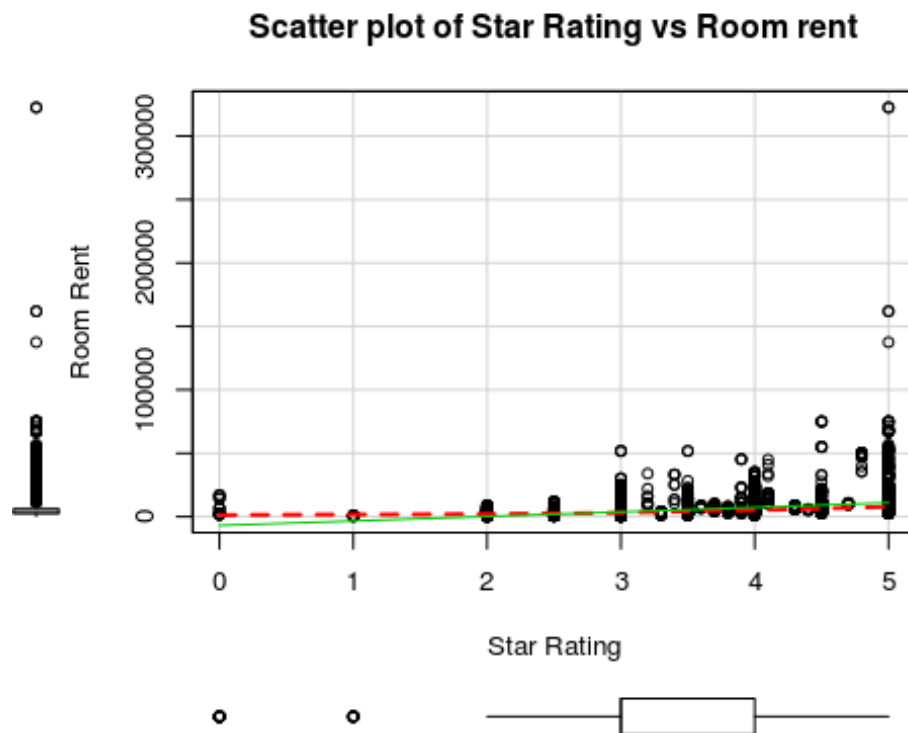

```
scatterplot(x = hotel$Population, y = hotel$CityRank,
        spread=FALSE, smoother.args=list(lty=2),
        main="Population Vs City Rank " ,
        xlab="Population", ylab="City rank")
```

Population Vs City Rank



```
scatterplot(HotelCapacity~StarRating,      data=hotel,
        spread=FALSE, smoother.args=list(lty=2),
        main="Scatter plot of Star Rating vs Hotel Capacity",
        ylab="Room Rent",
        xlab="Hotel Capacity")
```

Scatter plot of Star Rating vs Hotel Capacity

```
scatterplot(Airport~StarRating,data=hotel,
        spread=FALSE, smoother.args=list(lty=2),
        main="Star Rating vs Distance from Airport",
        ylab="Airport Distance",
        xlab="Star Rating")
```

Star Rating vs Distance from Airport
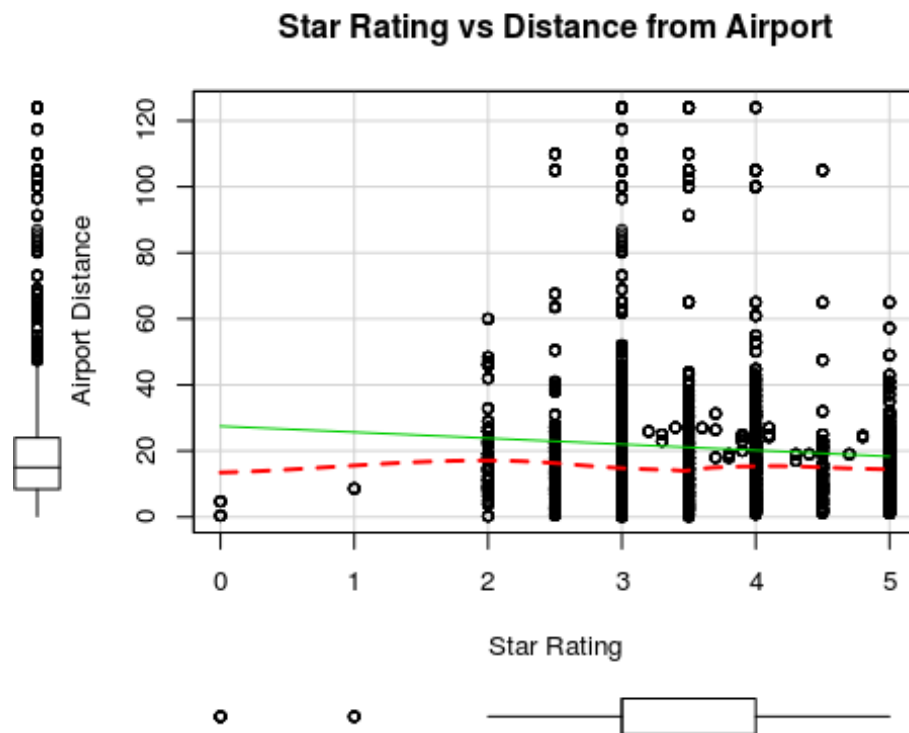

```
plot(y=hotel$RoomRent, x=hotel$Airport,
    col="green",
    ylim=c(0, 100000), xlim=c(0, 150),
    main="Relationship Btw Room Rent and Airport Distance",
    ylab="Hotel Rent", xlab="Airport Distance")
```
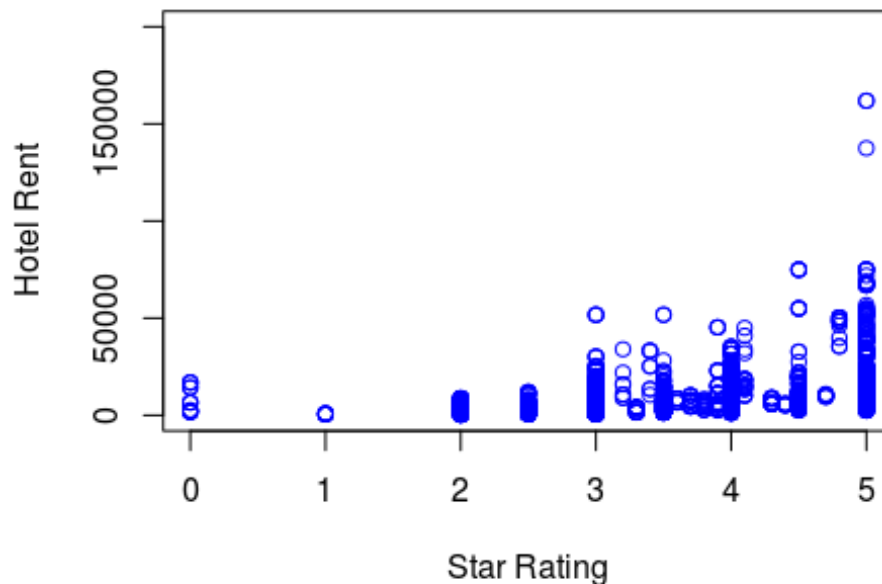
Relationship Btw Room Rent and Airport Distance
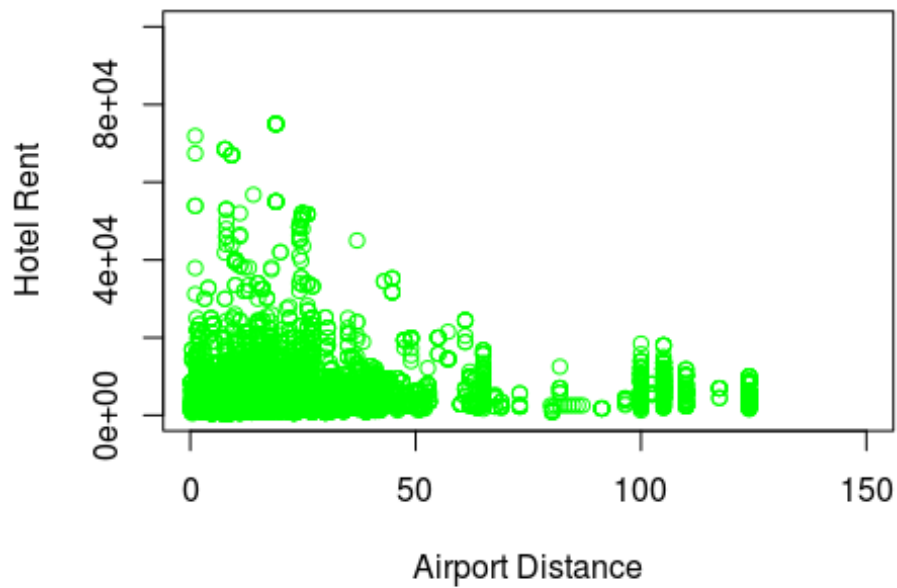

```
plot(y=hotel$RoomRent, x=hotel$StarRating,
    col="blue",
    ylim=c(0, 200000), xlim=c(0, 5),
    main="Relationship Btw Room Rent and Star Rating of Hotel",
    ylab="Hotel Rent", xlab="Star Rating")
```

Relationship Btw Room Rent and Star Rating of Ho

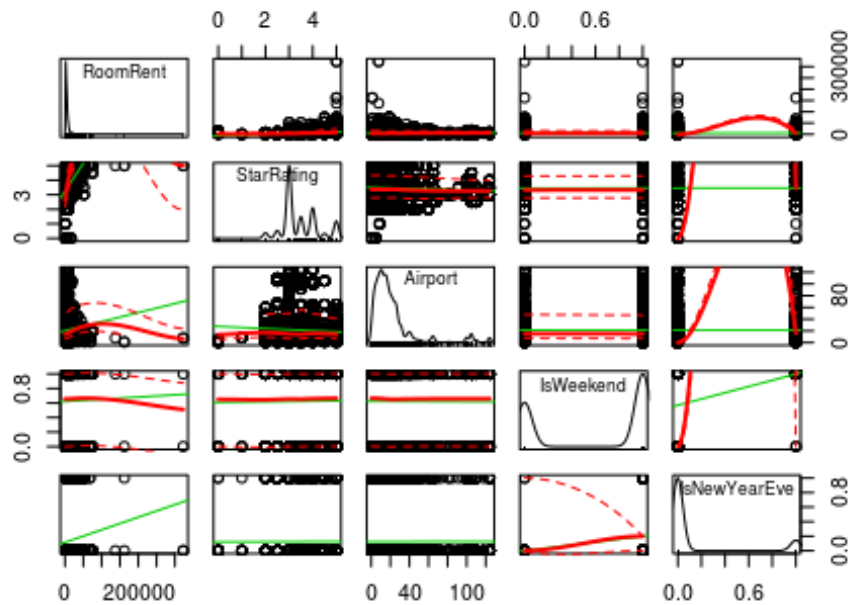## Scatterplot Matrix

Effect of external factors on room rent

```
scatterplotMatrix(formula = ~ RoomRent + StarRating +Airport+ IsWeekend
+ IsNewYearEve ,
          data = hotel ,
          main="Scatter plot matrix of external factors")
```

Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
spread = spread, : could not fit smooth

Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
spread = spread, : could not fit smooth

Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
spread = spread, : could not fit smooth

Scatter plot matrix of external factors
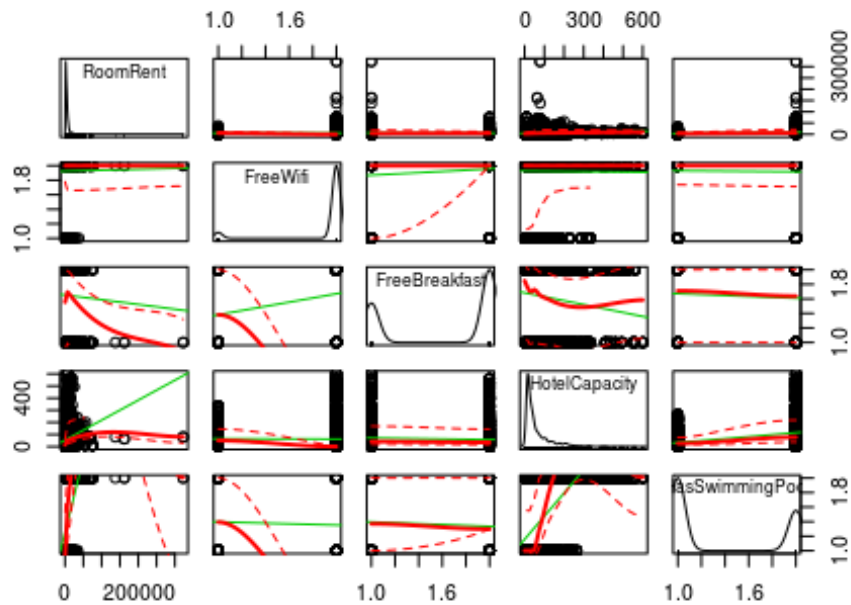

Effect of internal factors on room rent

```
scatterplotMatrix(formula = ~ RoomRent + FreeWifi + FreeBreakfast +
HotelCapacity + HasSwimmingPool,
          data = hotel ,
          main="Scatter plot matrix of internal factors")
```

```
## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
## spread = spread, : could not fit positive part of the spread
```
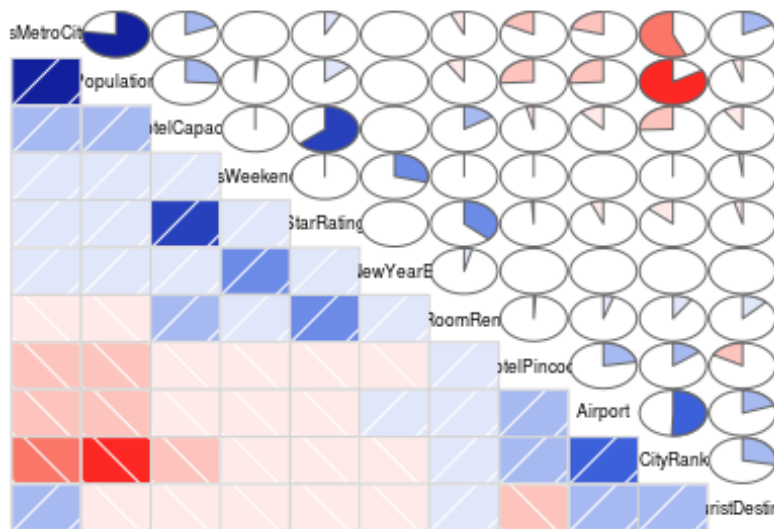
Scatter plot matrix of internal factors

Corrgram

```
library(corrgram)
corrgram(hotel, order=TRUE, lower.panel=panel.shade,
upper.panel=panel.pie, text.panel=panel.txt,
main="Hotel Pricing Corrgram")
```

Hotel Pricing Corrgram

Variance- Covariance Matrix

VarianceCovariancematrix <- var(hotel[,c(4:7,10:12,16:19)])

Warning in var(hotel[, c(4:7, 10:12, 16:19)]): NAs introduced by coercion

round(VarianceCovariancematrix, 2)

```
##                    IsMetroCity IsTouristDestination IsWeekend
## IsMetroCity              0.20                 0.04      0.00
## IsTouristDestination     0.04                 0.21      0.00
## IsWeekend                0.00                 0.00      0.23
## IsNewYearEve             0.00                 0.00      0.05
## RoomRent              -221.07               412.78     16.28
## StarRating               0.03                -0.01      0.00
## Airport                 -2.13                 2.03     -0.03
## FreeWifi                   NA                   NA        NA
## FreeBreakfast              NA                   NA        NA
## HotelCapacity            6.47                -3.32      0.23
## HasSwimmingPool            NA                   NA        NA
##                    IsNewYearEve   RoomRent StarRating Airport FreeWifi
## IsMetroCity                0.00    -221.07       0.03   -2.13       NA
## IsTouristDestination       0.00     412.78      -0.01    2.03       NA
## IsWeekend                  0.05      16.28       0.00   -0.03       NA
## IsNewYearEve               0.11      93.16       0.00    0.00       NA
## RoomRent                  93.16 53774601.81    2048.38 8287.18      NA
## StarRating                 0.00    2048.38       0.57   -1.05       NA
```

```
## Airport              0.00    8287.18    -1.05  518.01      NA
## FreeWifi               NA       NA       NA     NA      NA
## FreeBreakfast          NA       NA       NA     NA      NA
## HotelCapacity         0.03   88753.41    36.96 -205.32     NA
## HasSwimmingPool         NA       NA       NA     NA      NA
##               FreeBreakfast HotelCapacity HasSwimmingPool
## IsMetroCity             NA      6.47          NA
## IsTouristDestination    NA     -3.32          NA
## IsWeekend               NA      0.23          NA
## IsNewYearEve            NA      0.03          NA
## RoomRent                NA   88753.41         NA
## StarRating              NA     36.96          NA
## Airport                 NA   -205.32          NA
## FreeWifi                NA       NA           NA
## FreeBreakfast           NA       NA           NA
## HotelCapacity           NA    5877.27         NA
## HasSwimmingPool         NA       NA           NA
```

Correlation Matrix

```
x<- hotel[,c("Population",
"RoomRent","StarRating","Airport","HotelCapacity")]
y<- hotel[,c("Population",
"RoomRent","StarRating","Airport","HotelCapacity")]
correlationmatrix <- cor(x,y,method="pearson")
round(correlationmatrix,digits = 2)

##             Population RoomRent StarRating Airport HotelCapacity
## Population       1.00    -0.09      0.13   -0.26       0.26
## RoomRent        -0.09     1.00      0.37    0.05       0.16
## StarRating       0.13     0.37      1.00   -0.06       0.64
## Airport         -0.26     0.05     -0.06    1.00      -0.12
## HotelCapacity    0.26     0.16      0.64   -0.12       1.00

#Correlation matrix of Room Rent with IsWeekend and IsnewyearEve
round(cor(hotel$RoomRent, hotel[,c("IsWeekend","IsNewYearEve")]),4)

##     IsWeekend IsNewYearEve
## [1,]   0.0046      0.0385

#correlation Matrix of RoomRent with CityRank, IsMetroCity,
IsTouristDestination
round(cor(hotel$RoomRent,
hotel[,c("CityRank","IsMetroCity","IsTouristDestination")]),4)

##     CityRank IsMetroCity IsTouristDestination
## [1,]   0.094     -0.0668           0.1225
```

Pearson's correlation Test

```
cor.test(hotel$RoomRent , hotel$StarRating)
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  hotel$RoomRent and hotel$StarRating
## t = 45.719, df = 13230, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3545660 0.3839956
## sample estimates:
##       cor
## 0.3693734

cor.test(hotel$RoomRent , hotel$Airport)

## 
##  Pearson's product-moment correlation
## 
## data:  hotel$RoomRent and hotel$Airport
## t = 5.7183, df = 13230, p-value = 1.099e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03264192 0.06663581
## sample estimates:
##       cor
## 0.04965324

cor.test(hotel$RoomRent , hotel$IsTouristDestination)

## 
##  Pearson's product-moment correlation
## 
## data:  hotel$RoomRent and hotel$IsTouristDestination
## t = 14.197, df = 13230, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1056846 0.1392512
## sample estimates:
##       cor
## 0.122503

cor.test(hotel$RoomRent , hotel$HotelCapacity)

## 
##  Pearson's product-moment correlation
## 
## data:  hotel$RoomRent and hotel$HotelCapacity
## t = 18.389, df = 13230, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1412142 0.1744430
## sample estimates:
```

```
##      cor
## 0.1578733
```

T-tests

RoomRent and IsMetroCity

H0: There is no significant difference between the Room Rent of Hotels in non-metro cities and metro cities. H1: Hotels in non-metro cities are more expensive than that in metro cities.

```
t.test(RoomRent~IsMetroCity,data=hotel)

##
##  Welch Two Sample t-test
##
## data:  RoomRent by IsMetroCity
## t = 10.721, df = 13224, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   888.0308 1285.4102
## sample estimates:
## mean in group 0 mean in group 1
##       5782.794       4696.073
```

Since p-value<0.05, we reject the null hypothesis,hence,the Room Rents of Hotels in non-metro cities is more than that of metro cities.

RoomRent and IsTouristDestination

H0: There is no significant difference between the Room Rent of Hotels in Tourist destinations and non tourist destinations. H1: The Room Rents of Hotels in Tourist destinations are greater than that in non tourist destinations

```
t.test(RoomRent~IsTouristDestination,data=hotel)

##
##  Welch Two Sample t-test
##
## data:  RoomRent by IsTouristDestination
## t = -19.449, df = 12888, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2152.059 -1757.983
## sample estimates:
## mean in group 0 mean in group 1
##       4111.003       6066.024
```

Since p-value<0.05, we reject the null hypothesis, hence, the Room Rents of Hotels in Tourist destinations are greater than that in non tourist destinations.

RoomRent and IsNewYearEve

H0: There is no significant difference between the Room Rent of Hotels on normal Eve and New Year's Eve. H1: The Room Rents of Hotels on normal Eve are cheaper than that on New Year's Eve

```
t.test(RoomRent~IsNewYearEve,data=hotel)

##
##  Welch Two Sample t-test
##
## data:  RoomRent by IsNewYearEve
## t = -4.1793, df = 2065, p-value = 3.046e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1256.5297  -453.9099
## sample estimates:
## mean in group 0 mean in group 1
##      5367.606      6222.826
```

Since p-value<0.05, we reject the null hypothesis, hence, the Room Rents of Hotels on normal Eve are cheaper than that on New Year's Eve.

RoomRent and FreeBreakfast

H0: There is no significant difference between the Room Rent of Hotels with free breakfast and hotels without free breakfast. H1: There is a significant difference between the Room Rent of Hotels with free breakfast and hotels without free breakfast

```
t.test(RoomRent~FreeBreakfast,data=hotel)

##
##  Welch Two Sample t-test
##
## data:  RoomRent by FreeBreakfast
## t = 0.98095, df = 6212.3, p-value = 0.3267
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -153.5017  460.9935
## sample estimates:
##  mean in group No mean in group Yes
##       5573.790       5420.044
```

Since p-value>0.05, we accept H0, hence, there is no significant difference between the Room Rent of Hotels with free breakfast and hotels without free breakfast.

RoomRent and FreeWifi

H0:-There is no significant difference between the Room Rent of Hotels providing free wifi and those which do not H1:-There is a significant difference between the Room Rent of Hotels providing free wifi and those which do not.

```
t.test(RoomRent~FreeWifi,data=hotel)

##
##  Welch Two Sample t-test
##
## data:  RoomRent by FreeWifi
## t = -0.76847, df = 1804.7, p-value = 0.4423
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -360.5977  157.5701
## sample estimates:
##  mean in group No mean in group Yes
##        5380.004          5481.518
```

Since p-vale>0.05, we accept H0, hence there is no significant difference between the Room Rent of Hotels providing free wifi and those which do not.

Regression Analysis

Consider the following Regression Model

$$RoomRent = \beta_0 + \beta_1 StarRating + \beta_2 Airport + \beta_3 FreeWifi + \beta_4 FreeBreakfast + \beta_5 HotelCa$$

```
fit <- lm(RoomRent ~ StarRating + Airport + FreeWifi + FreeBreakfast +
HotelCapacity + HasSwimmingPool + IsMetroCity + IsNewYearEve +
IsTouristDestination, data = hotel)
summary(fit)

##
## Call:
## lm(formula = RoomRent ~ StarRating + Airport + FreeWifi +
FreeBreakfast +
##     HotelCapacity + HasSwimmingPool + IsMetroCity + IsNewYearEve +
##     IsTouristDestination, data = hotel)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -11696  -2375   -701   1063 309539
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -8906.418    405.396 -21.970  < 2e-16 ***
## StarRating          3564.570    110.489 32.262  < 2e-16 ***
## Airport               11.265      2.710   4.157 3.24e-05 ***
## FreeWifiYes          485.597    224.134   2.167   0.0303 *
## FreeBreakfastYes     182.992    123.296   1.484   0.1378
## HotelCapacity        -10.990      1.026 -10.714  < 2e-16 ***
## HasSwimmingPoolYes  2227.069    159.327 13.978  < 2e-16 ***
## IsMetroCity        -1548.328    138.527 -11.177  < 2e-16 ***
## IsNewYearEve         844.123    174.046   4.850 1.25e-06 ***
## IsTouristDestination 2113.725    134.336 15.735  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6607 on 13222 degrees of freedom
## Multiple R-squared:  0.1887, Adjusted R-squared:  0.1881
## F-statistic: 341.7 on 9 and 13222 DF,  p-value: < 2.2e-16
```

beta coefficients
fit$coefficients

```
##         (Intercept)           StarRating             Airport
##        -8906.41842           3564.56963             11.26457
##         FreeWifiYes     FreeBreakfastYes        HotelCapacity
##           485.59699           182.99190            -10.98971
##   HasSwimmingPoolYes          IsMetroCity         IsNewYearEve
##          2227.06946          -1548.32763            844.12347
## IsTouristDestination
##          2113.72520
```

# confidence intervals
confint(fit)

```
##                          2.5 %       97.5 %
## (Intercept)         -9701.052106 -8111.784736
## StarRating           3347.994688  3781.144564
## Airport                 5.953022    16.576118
## FreeWifiYes            46.263019   924.930954
## FreeBreakfastYes      -58.686360   424.670157
## HotelCapacity         -13.000331    -8.979088
## HasSwimmingPoolYes   1914.764820  2539.374098
## IsMetroCity         -1819.860397 -1276.794869
## IsNewYearEve          502.967551  1185.279392
## IsTouristDestination 1850.407049  2377.043359
```

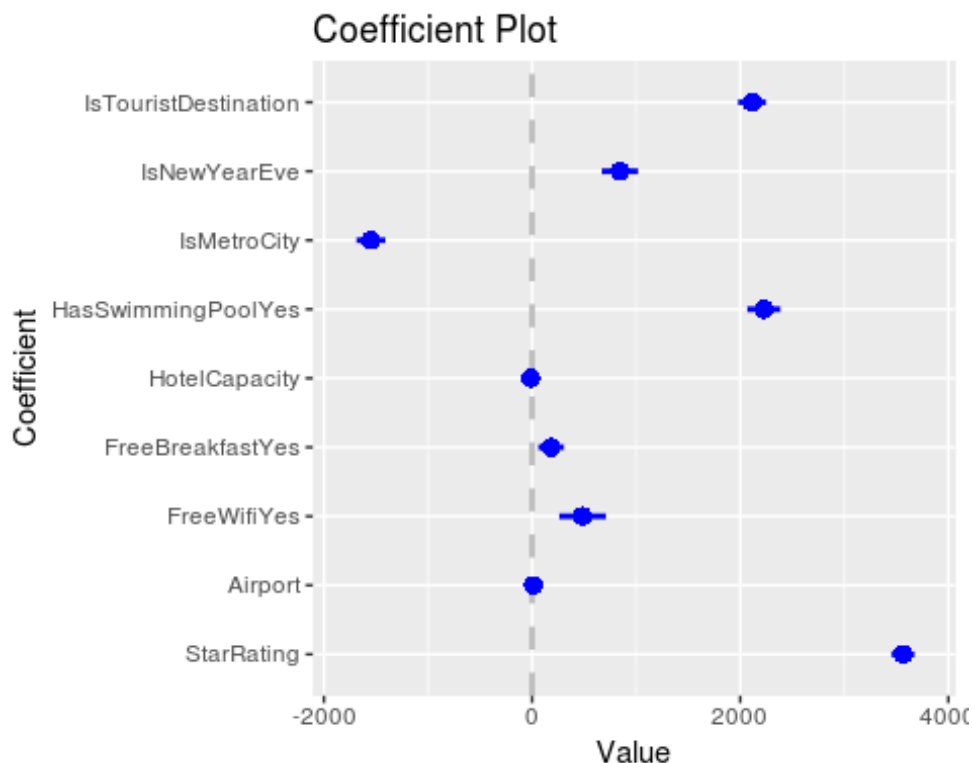## Visualizing the beta coefficients
library(coefplot)

```
## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

coefplot(fit, predictors=c("StarRating", "Airport", "FreeWifi", "FreeBreakfast",
"HotelCapacity", "HasSwimmingPool", "IsMetroCity", "IsNewYearEve",
"IsTouristDestination"))

Coefficient Plot

Conclusion

- Hence, the p-values and the coefficients suggest that the model is a good fit and the regression is good.
- Factors like StarRating, IsMetroCity, IsTouristDestination and IsNewYearEve affect RoomRent much much significantly.
- Factors like FreeWifi and FreeBreakfast have a very little effect on Roomrent as their confidence interval is closer to zero.
- Factors like Distance from Airport and Hotel Capacity are not statistically significant for RoomRent as their confidence interval includes zero.