Hotel Room Pricing In The Indian Market

NAME: James Rohan Gangavarapu

EMAIL: james.reckon@gmail.com

COLLEGE / COMPANY: Georgia State Univiersity

> ############## 1 Reading CSV file into R #################

> setwd("C:/Users/Nikki/Desktop/James/Data Science Internship/Assignments/Projects/CapStonProj\_Hotels")

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

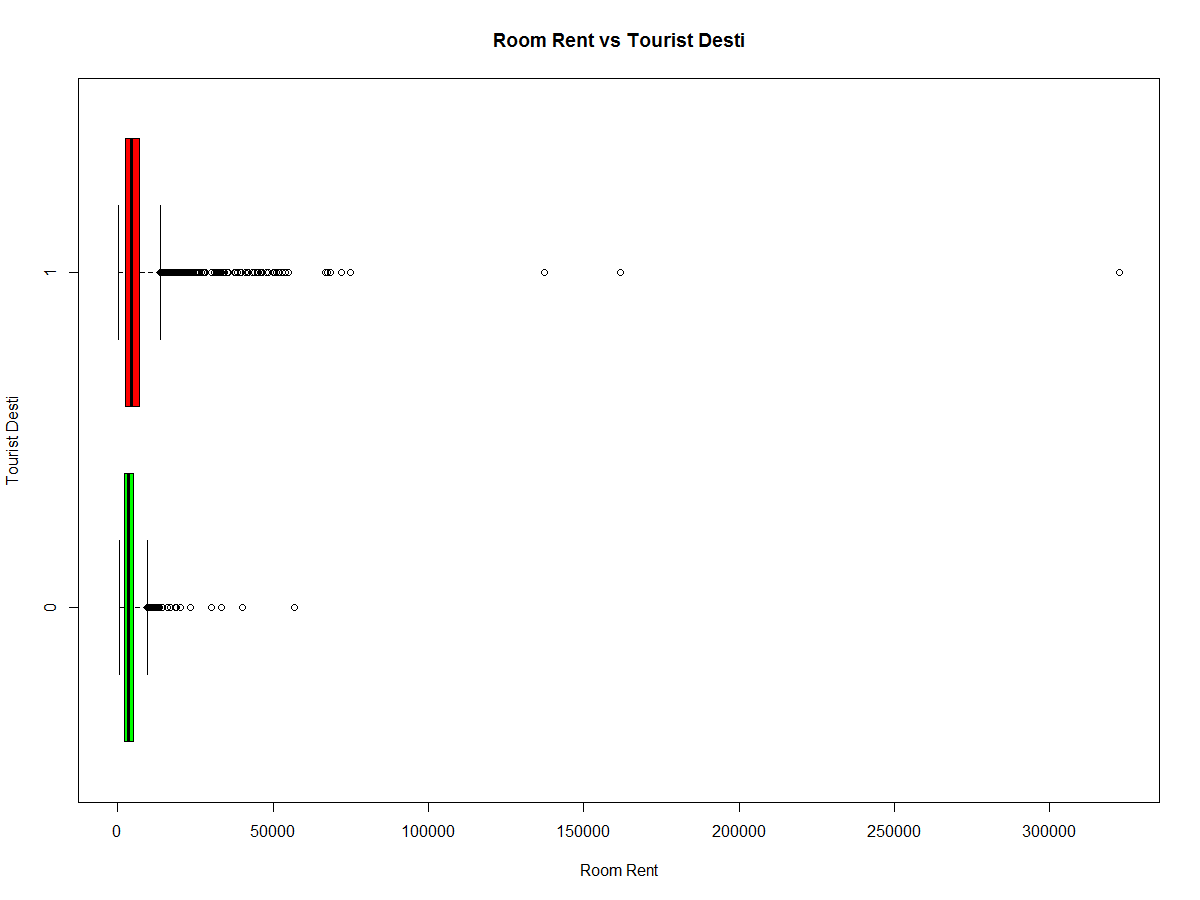
> attach(hotels.df)

Summary Stats

|  |
| --- |
| > ############ 2 Summary Stats ############  > summary(hotels.df)  CityName Population CityRank IsMetroCity IsTouristDestination IsWeekend IsNewYearEve  Delhi :2048 Min. : 8096 Min. : 0.00 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000  Jaipur : 768 1st Qu.: 744983 1st Qu.: 2.00 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000  Mumbai : 712 Median : 3046163 Median : 9.00 Median :0.0000 Median :1.0000 Median :1.0000 Median :0.0000  Bangalore: 656 Mean : 4416837 Mean :14.83 Mean :0.2842 Mean :0.6972 Mean :0.6228 Mean :0.1244  Goa : 624 3rd Qu.: 8443675 3rd Qu.:24.00 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000  Kochi : 608 Max. :12442373 Max. :44.00 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000  (Other) :7816  Date HotelName RoomRent StarRating Airport  Dec 21 2016:1611 Vivanta by Taj : 32 Min. : 299 Min. :0.000 Min. : 0.20  Dec 24 2016:1611 Goldfinch Hotel : 24 1st Qu.: 2436 1st Qu.:3.000 1st Qu.: 8.40  Dec 25 2016:1611 OYO Rooms : 24 Median : 4000 Median :3.000 Median : 15.00  Dec 28 2016:1611 The Gordon House Hotel: 24 Mean : 5474 Mean :3.459 Mean : 21.16  Dec 31 2016:1611 Apnayt Villa : 16 3rd Qu.: 6299 3rd Qu.:4.000 3rd Qu.: 24.00  Dec 18 2016:1608 Bentleys Hotel Colaba : 16 Max. :322500 Max. :5.000 Max. :124.00  (Other) :3569 (Other) :13096  HotelAddress HotelPincode HotelDescription  The Mall, Shimla : 32 Min. : 100025 3 : 120  #2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16 1st Qu.: 221001 Abc : 112  121, City Terrace, Walchand Hirachand Marg, Mumbai, Maharashtra : 16 Median : 395003 3-star hotel: 104  14-4507/9, Balmatta Road, Near Jyothi Circle, Hampankatta : 16 Mean : 397430 3.5 : 88  144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16 3rd Qu.: 570001 4 : 72  17, Oliver Road, Colaba, Mumbai, Maharashtra : 16 Max. :7000157 (Other) :12728  (Other) :13120 NA's : 8  FreeWifi FreeBreakfast HotelCapacity HasSwimmingPool  Min. :0.0000 Min. :0.0000 Min. : 0.00 Min. :0.0000  1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.: 16.00 1st Qu.:0.0000  Median :1.0000 Median :1.0000 Median : 34.00 Median :0.0000  Mean :0.9259 Mean :0.6491 Mean : 62.51 Mean :0.3558  3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.: 75.00 3rd Qu.:1.0000  Max. :1.0000 Max. :1.0000 Max. :600.00 Max. :1.0000    > library(psych)  > describe(hotels.df)  vars n mean sd median trimmed mad min max range skew kurtosis  CityName\* 1 13232 18.07 11.72 16 17.29 11.86 1.0 42 41.0 0.48 -0.88  Population 2 13232 4416836.87 4258386.00 3046163 4040816.22 3846498.95 8096.0 12442373 12434277.0 0.68 -1.08  CityRank 3 13232 14.83 13.51 9 13.30 11.86 0.0 44 44.0 0.69 -0.76  IsMetroCity 4 13232 0.28 0.45 0 0.23 0.00 0.0 1 1.0 0.96 -1.08  IsTouristDestination 5 13232 0.70 0.46 1 0.75 0.00 0.0 1 1.0 -0.86 -1.26  IsWeekend 6 13232 0.62 0.48 1 0.65 0.00 0.0 1 1.0 -0.51 -1.74  IsNewYearEve 7 13232 0.12 0.33 0 0.03 0.00 0.0 1 1.0 2.28 3.18  Date\* 8 13232 14.30 2.69 14 14.39 2.97 1.0 20 19.0 -0.77 1.92  HotelName\* 9 13232 841.19 488.16 827 841.18 641.97 1.0 1670 1669.0 0.01 -1.25  RoomRent 10 13232 5473.99 7333.12 4000 4383.33 2653.85 299.0 322500 322201.0 16.75 582.06  StarRating 11 13232 3.46 0.76 3 3.40 0.74 0.0 5 5.0 0.48 0.25  Airport 12 13232 21.16 22.76 15 16.39 11.12 0.2 124 123.8 2.73 7.89  HotelAddress\* 13 13232 1202.53 582.17 1261 1233.25 668.65 1.0 2108 2107.0 -0.37 -0.88  HotelPincode 14 13232 397430.26 259837.50 395003 388540.47 257975.37 100025.0 7000157 6900132.0 9.99 249.76  HotelDescription\* 15 13224 581.34 363.26 567 575.37 472.95 1.0 1226 1225.0 0.11 -1.25  FreeWifi 16 13232 0.93 0.26 1 1.00 0.00 0.0 1 1.0 -3.25 8.57  FreeBreakfast 17 13232 0.65 0.48 1 0.69 0.00 0.0 1 1.0 -0.62 -1.61  HotelCapacity 18 13232 62.51 76.66 34 46.03 28.17 0.0 600 600.0 2.95 11.39  HasSwimmingPool 19 13232 0.36 0.48 0 0.32 0.00 0.0 1 1.0 0.60 -1.64  se  CityName\* 0.10  Population 37019.65  CityRank 0.12  IsMetroCity 0.00  IsTouristDestination 0.00  IsWeekend 0.00  IsNewYearEve 0.00  Date\* 0.02  HotelName\* 4.24  RoomRent 63.75  StarRating 0.01  Airport 0.20  HotelAddress\* 5.06  HotelPincode 2258.86  HotelDescription\* 3.16  FreeWifi 0.00  FreeBreakfast 0.00  HotelCapacity 0.67  HasSwimmingPool 0.00 |
|  |
| |  | | --- | | > | |

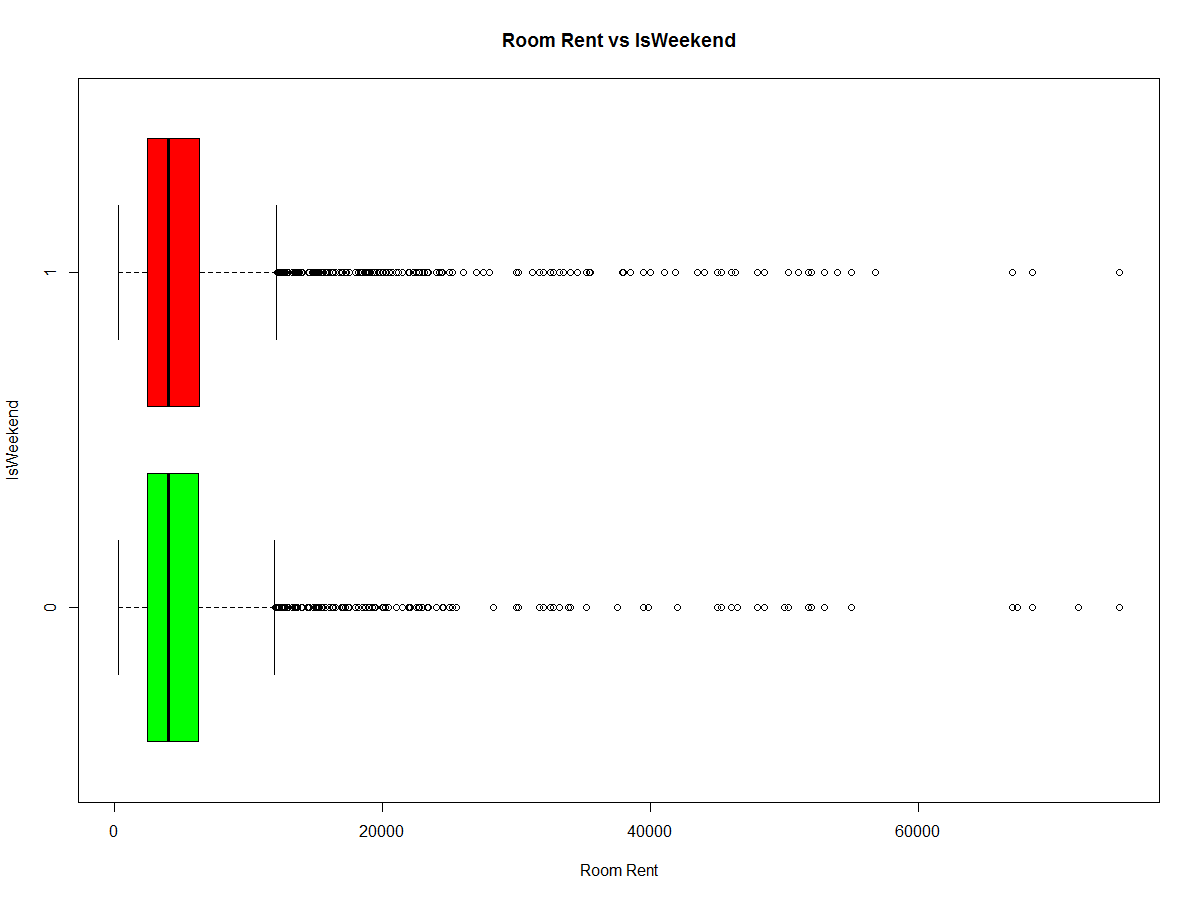
####### 3 4 5 6 7 ##########

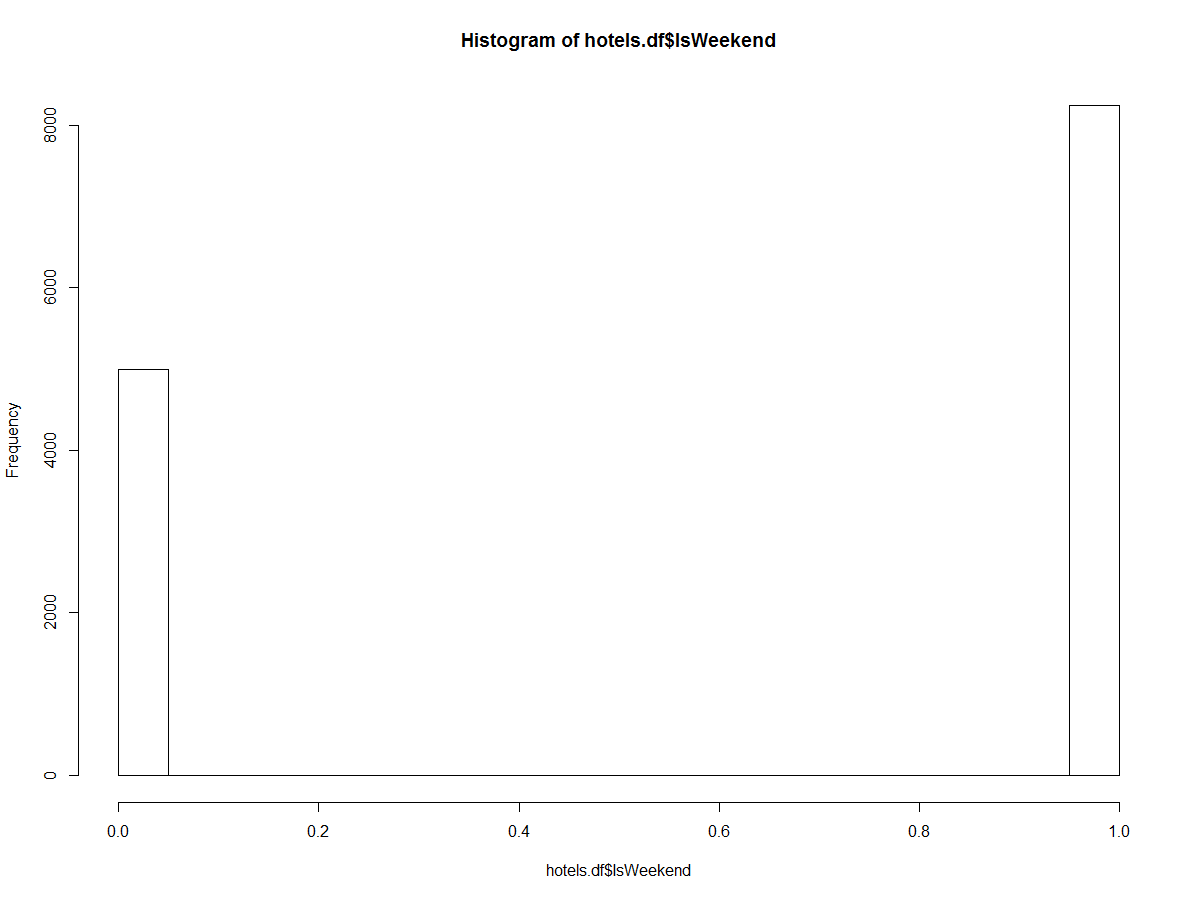
|  |
| --- |
| > boxplot(hotels.df$RoomRent~hotels.df$IsTouristDestination,horizontal = TRUE,main = "Room Rent vs Tourist Desti",  + xlab = "Room Rent"  + ,col = (c("green","red")), ylab = "Tourist Desti") |
|  |
| |  | | --- | | > | |

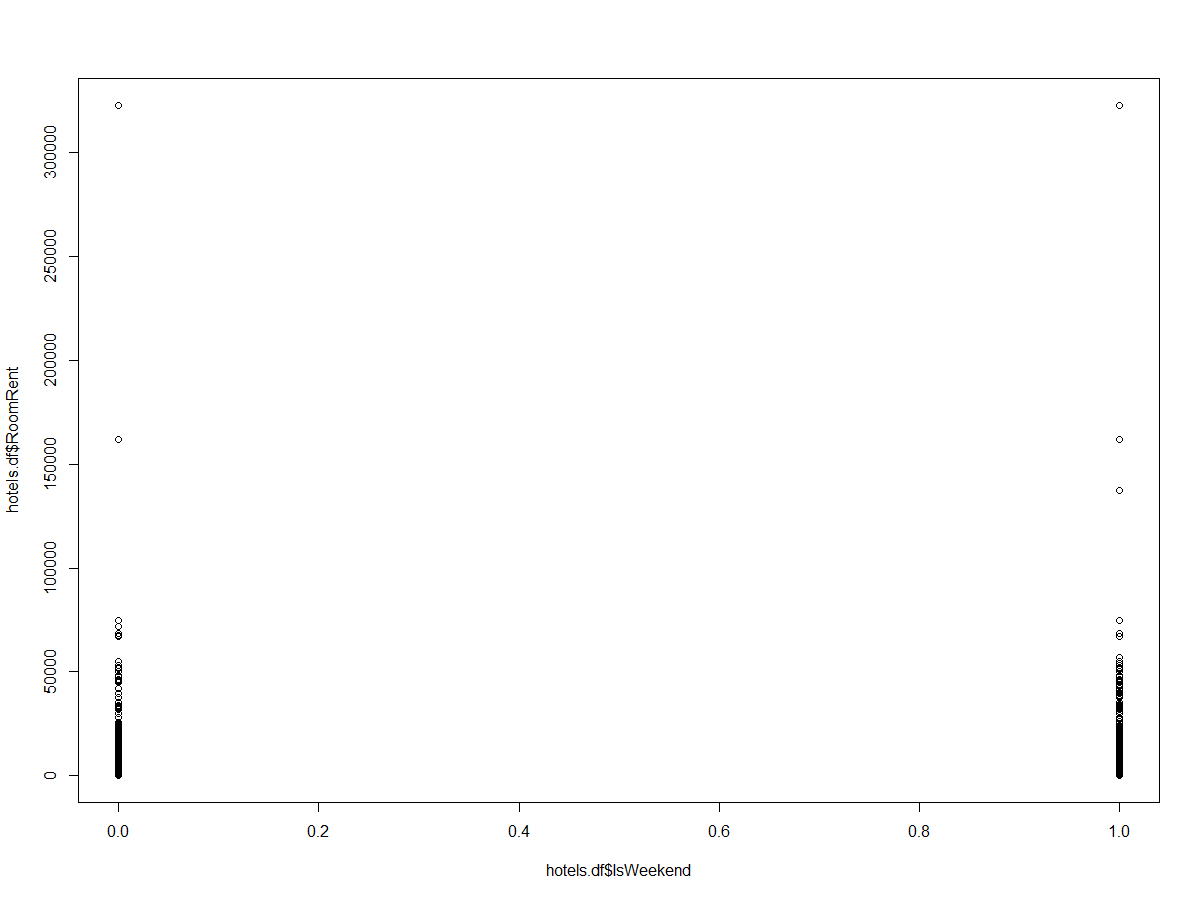


|  |
| --- |
| > hotels\_RoomRentLessThan.df <- hotels.df[ which(hotels.df$RoomRent <= 50000\*2) , ]  > View(hotels\_RoomRentLessThan.df) |
|  |
| |  | | --- | | > | |

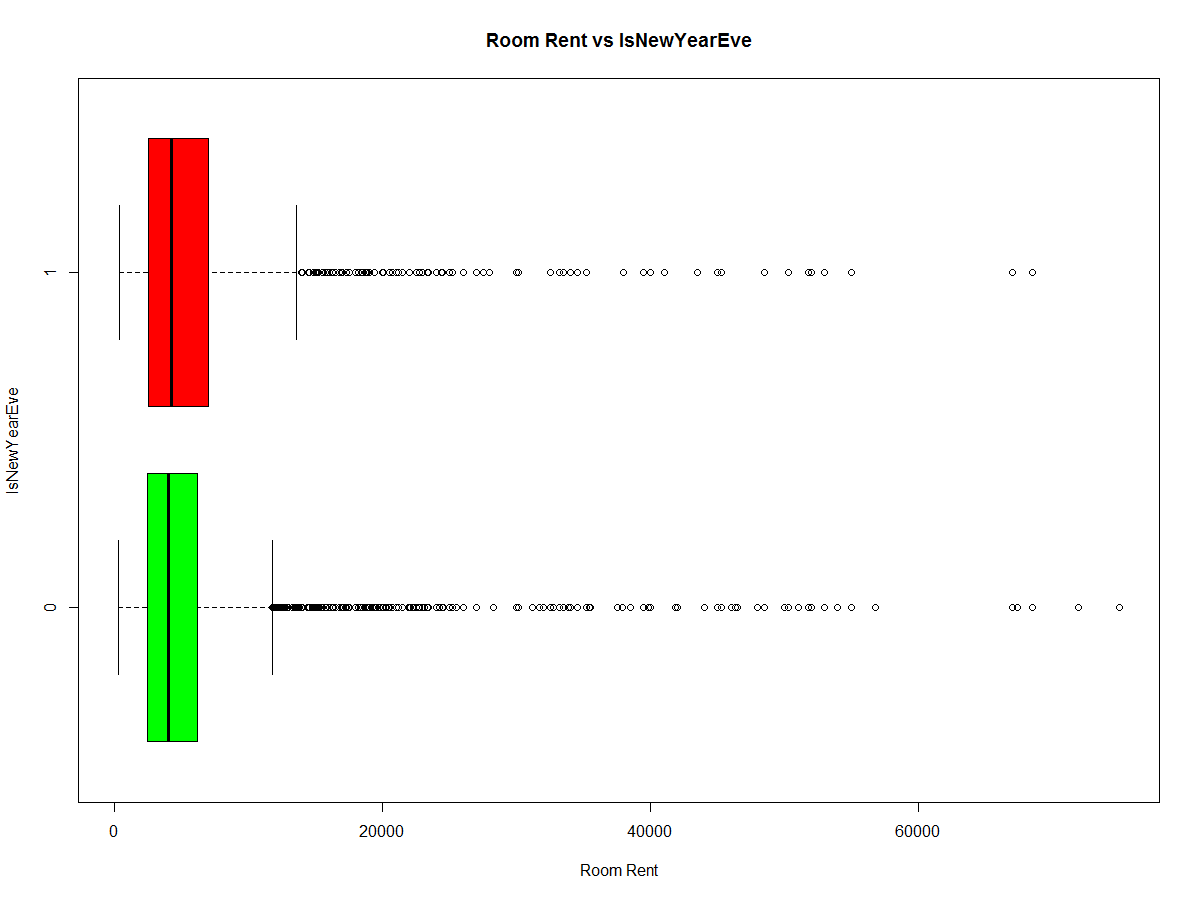
|  |
| --- |
| > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$IsWeekend,horizontal = TRUE,  + main = "Room Rent vs IsWeekend",xlab = "Room Rent",col = (c("green","red")), ylab = "IsWeekend")  > hist(hotels.df$IsWeekend)  > plot(hotels.df$IsWeekend,hotels.df$RoomRent) |
|  |
| |  | | --- | | > | |

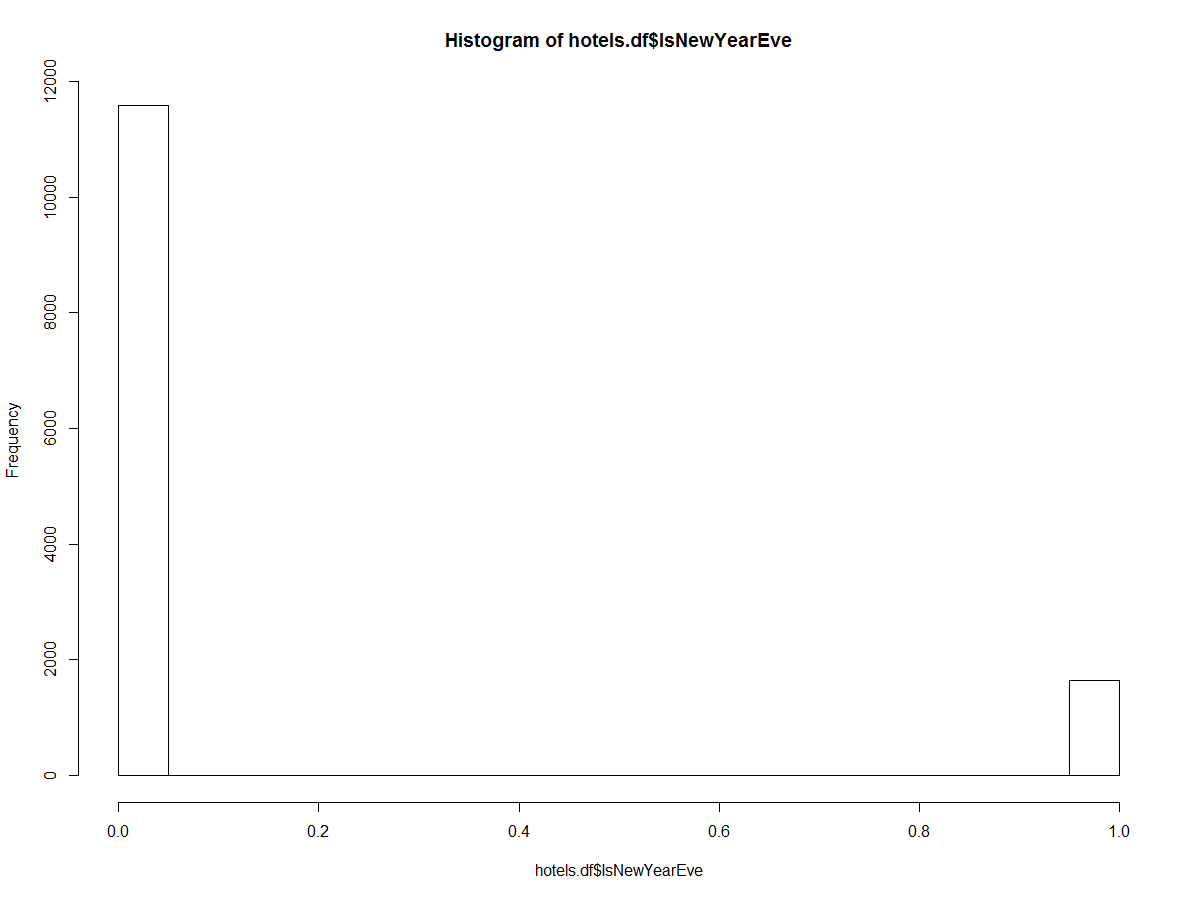


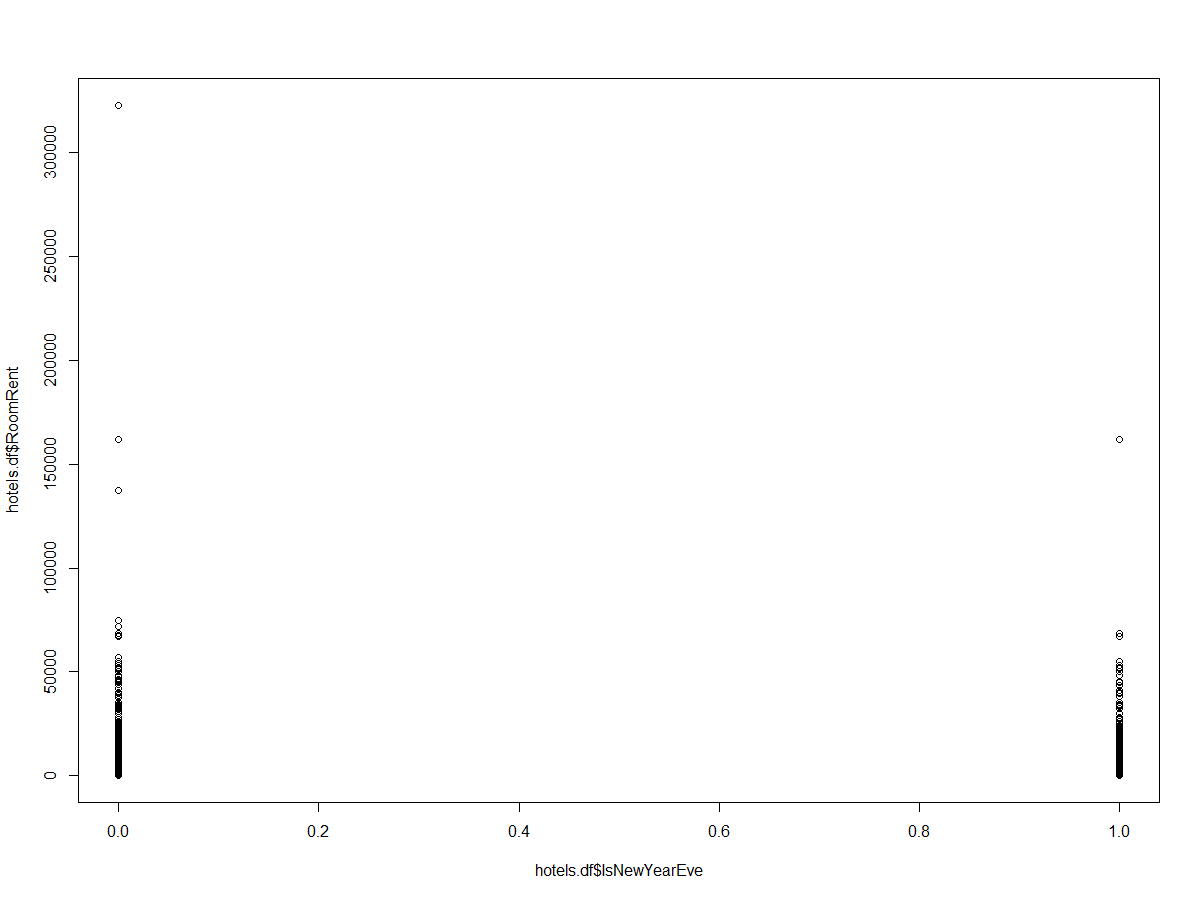




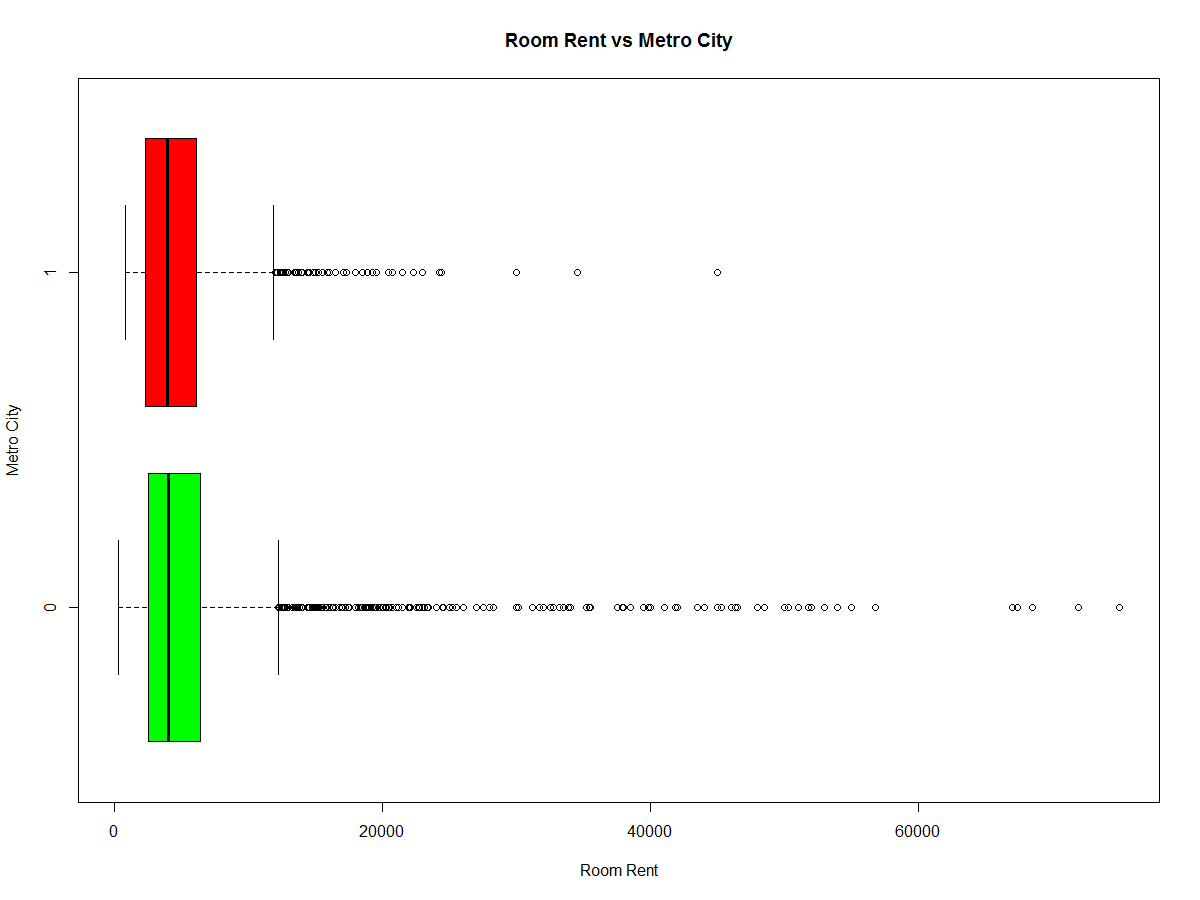
|  |
| --- |
| > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$IsNewYearEve,horizontal = TRUE,  + main = "Room Rent vs IsNewYearEve",xlab = "Room Rent",col = (c("green","red")), ylab = "IsNewYearEve")  > hist(hotels.df$IsNewYearEve)  > plot(hotels.df$IsNewYearEve,hotels.df$RoomRent) |
|  |
| |  | | --- | | > | |

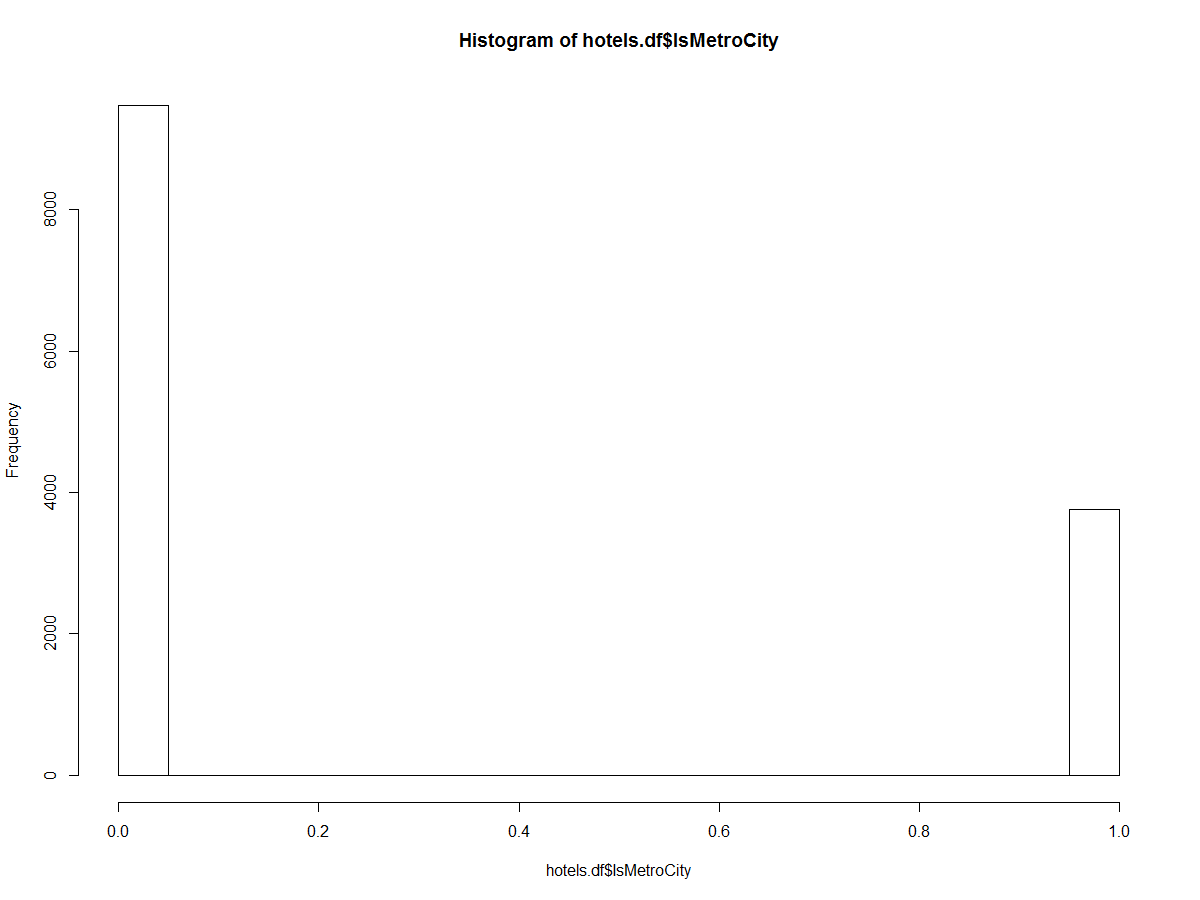


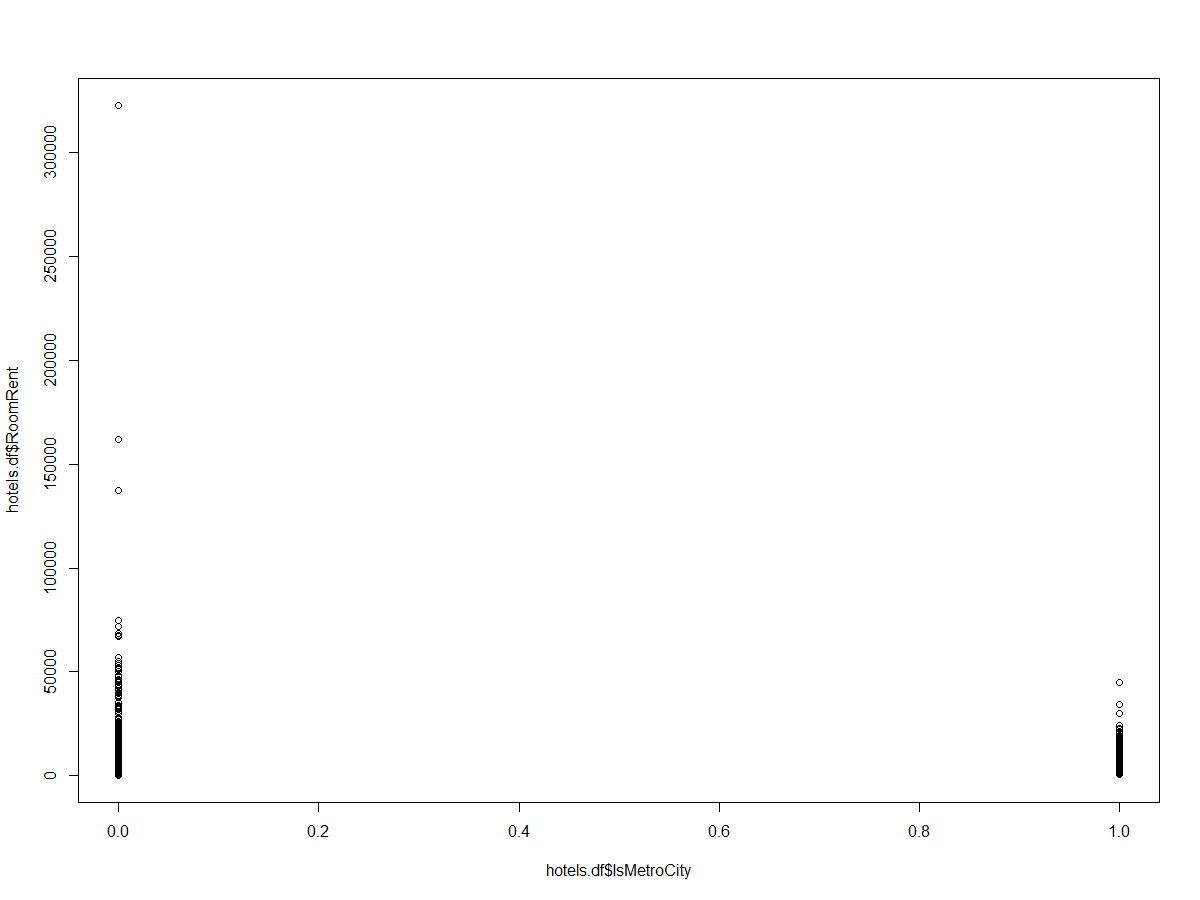




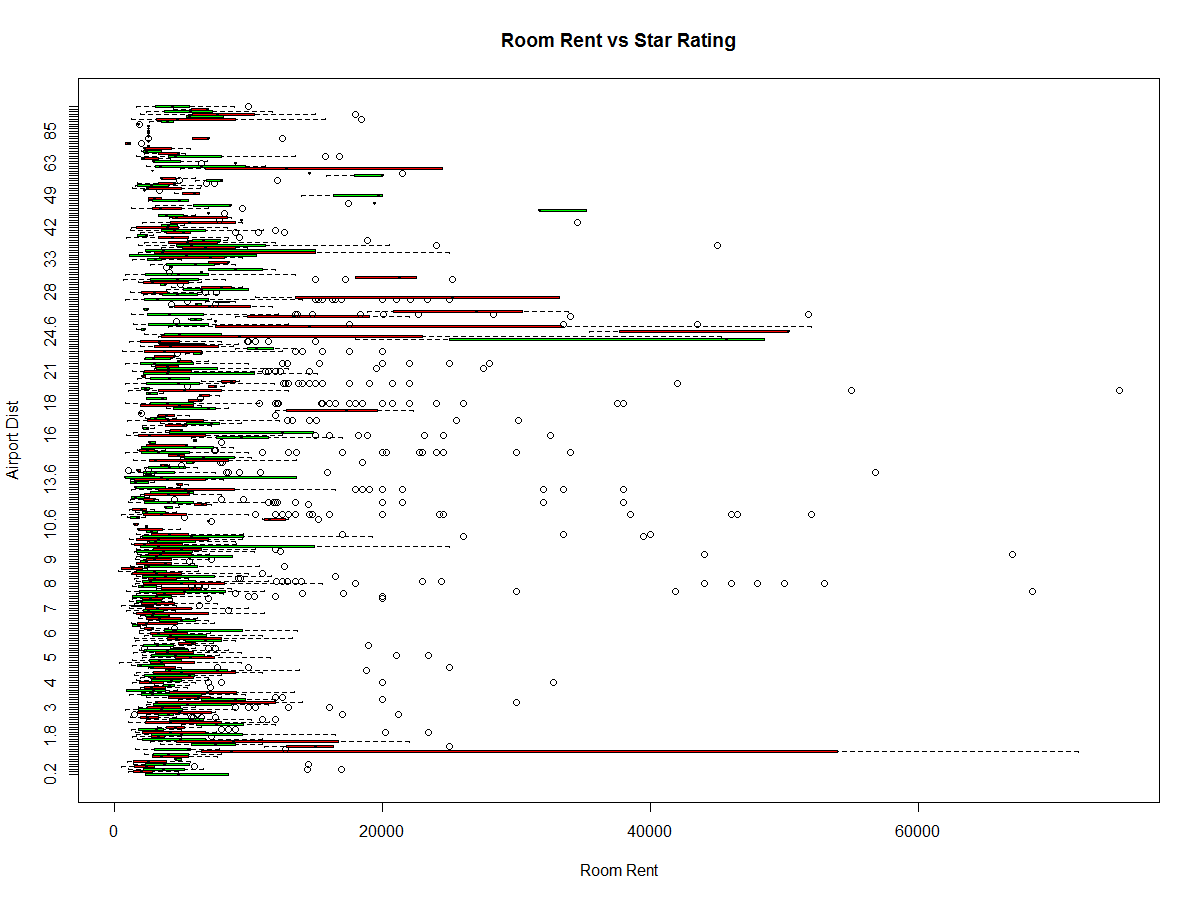
|  |
| --- |
| > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$IsMetroCity,horizontal = TRUE,  + main = "Room Rent vs Metro City",xlab = "Room Rent",col = (c("green","red")), ylab = "Metro City")  > hist(hotels.df$IsMetroCity)  > plot(hotels.df$IsMetroCity,hotels.df$RoomRent) |
|  |
| |  | | --- | |  | |

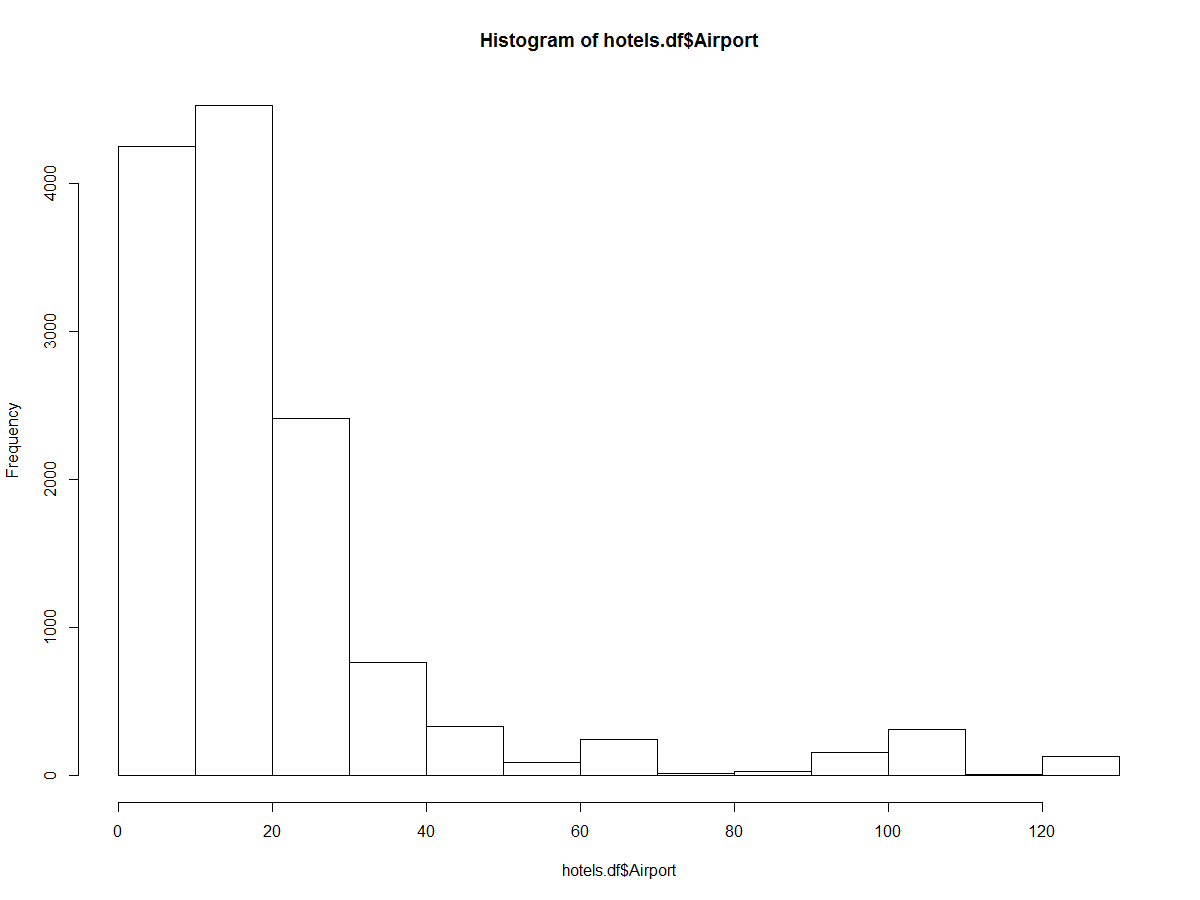


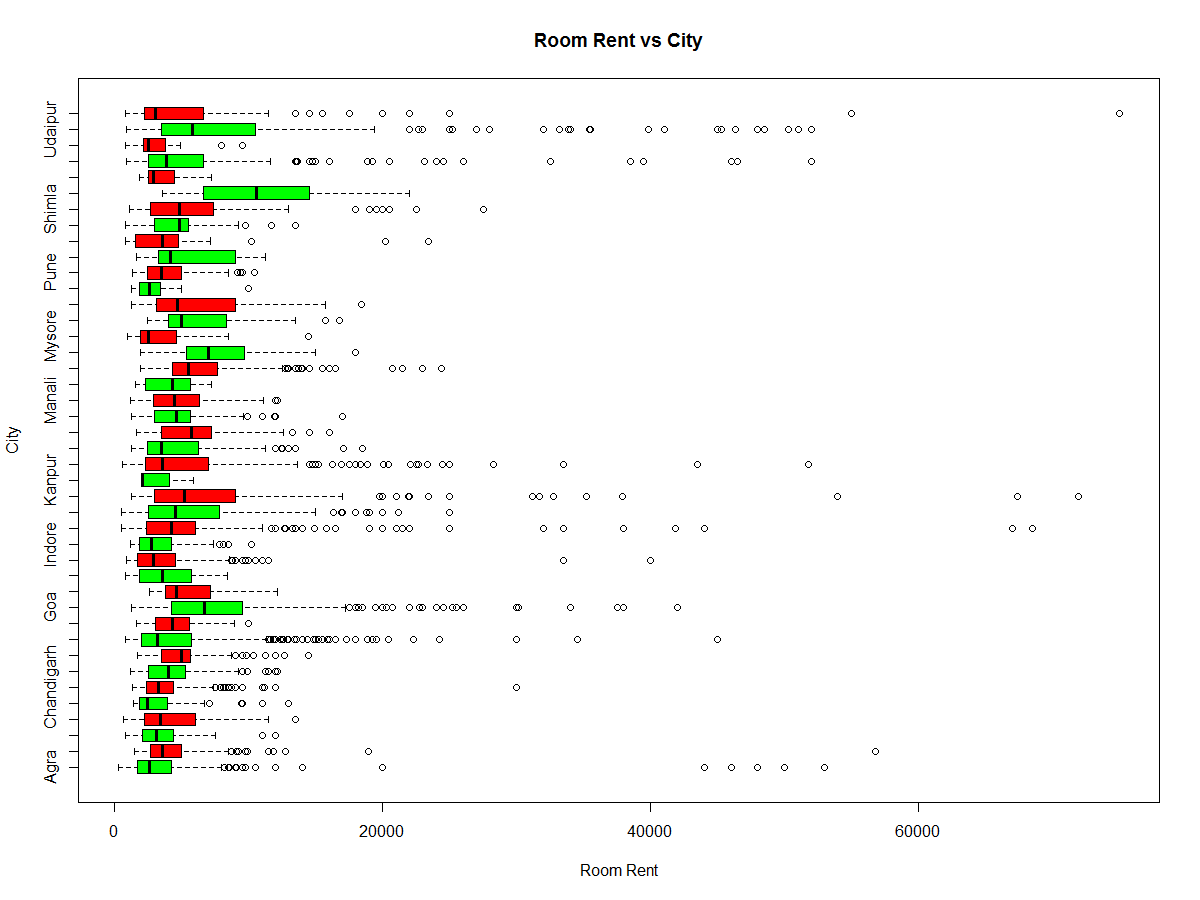


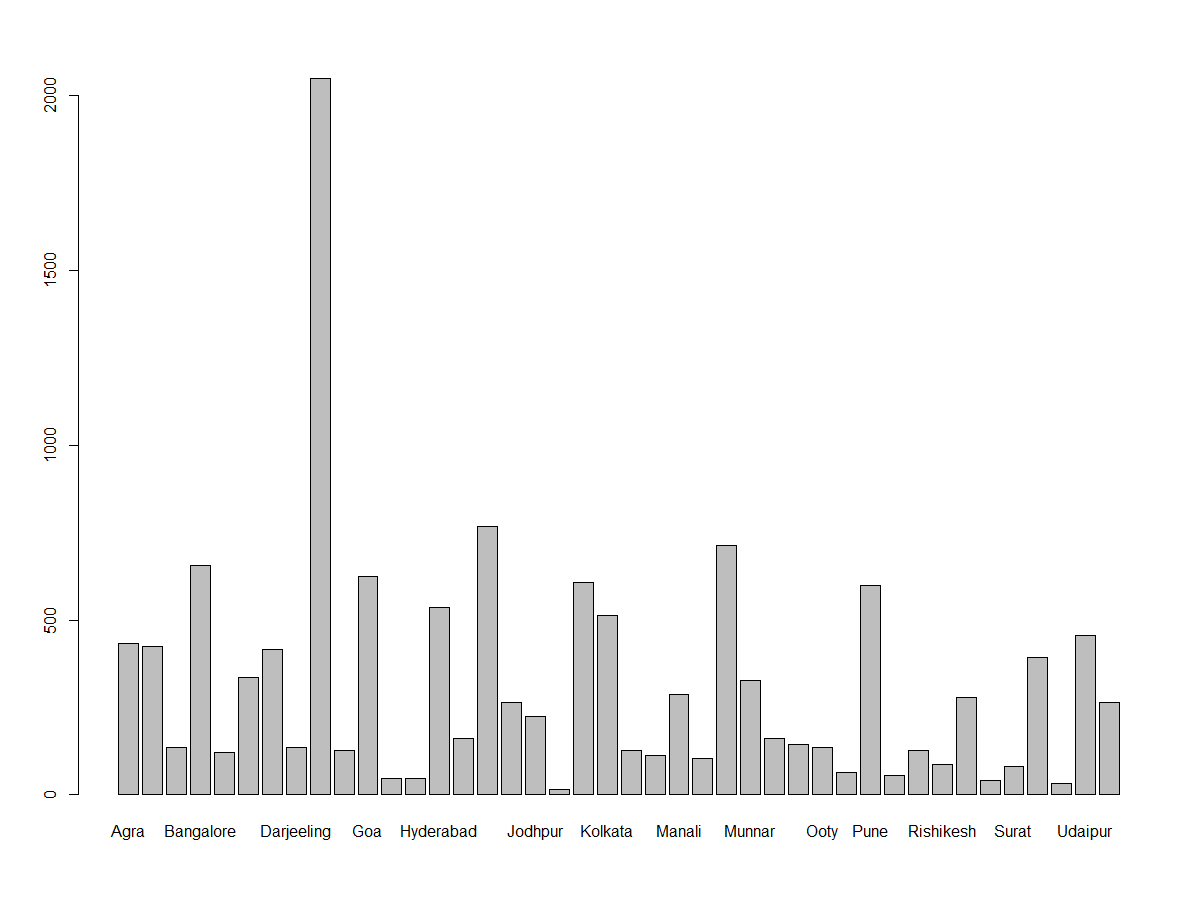


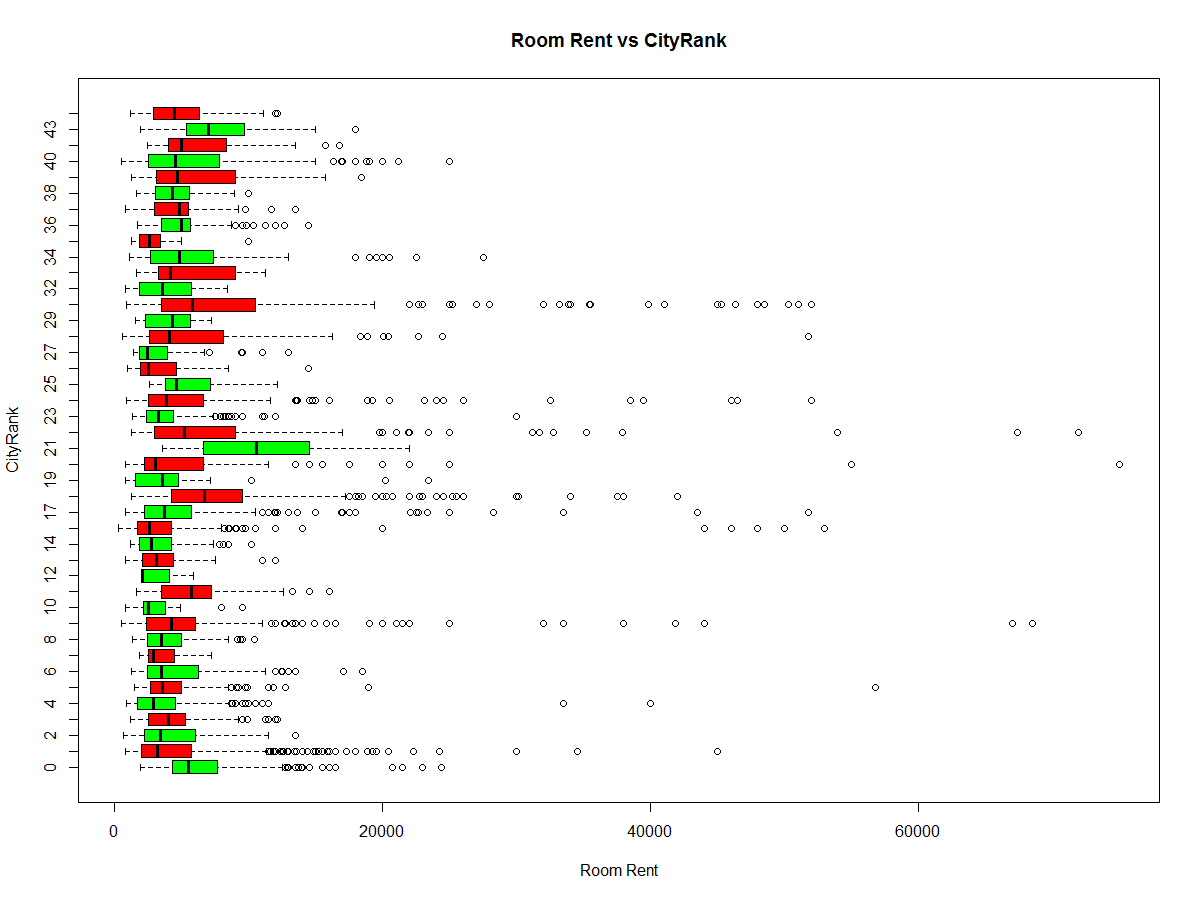
|  |
| --- |
| > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$Airport,horizontal = TRUE,  + main = "Room Rent vs Star Rating",xlab = "Room Rent",col = (c("green","red")), ylab = "Airport Dist")  > hist(hotels.df$Airport)  >  >  > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$CityName,horizontal = TRUE,  + main = "Room Rent vs City",xlab = "Room Rent",col = (c("green","red")), ylab = "City")  > plot(hotels.df$CityName)  >  > boxplot(hotels\_RoomRentLessThan.df$RoomRent~hotels\_RoomRentLessThan.df$CityRank,horizontal = TRUE,  + main = "Room Rent vs CityRank",xlab = "Room Rent",col = (c("green","red")), ylab = "CityRank")  > hist(hotels.df$CityRank) |
|  |
| |  | | --- | |  | |

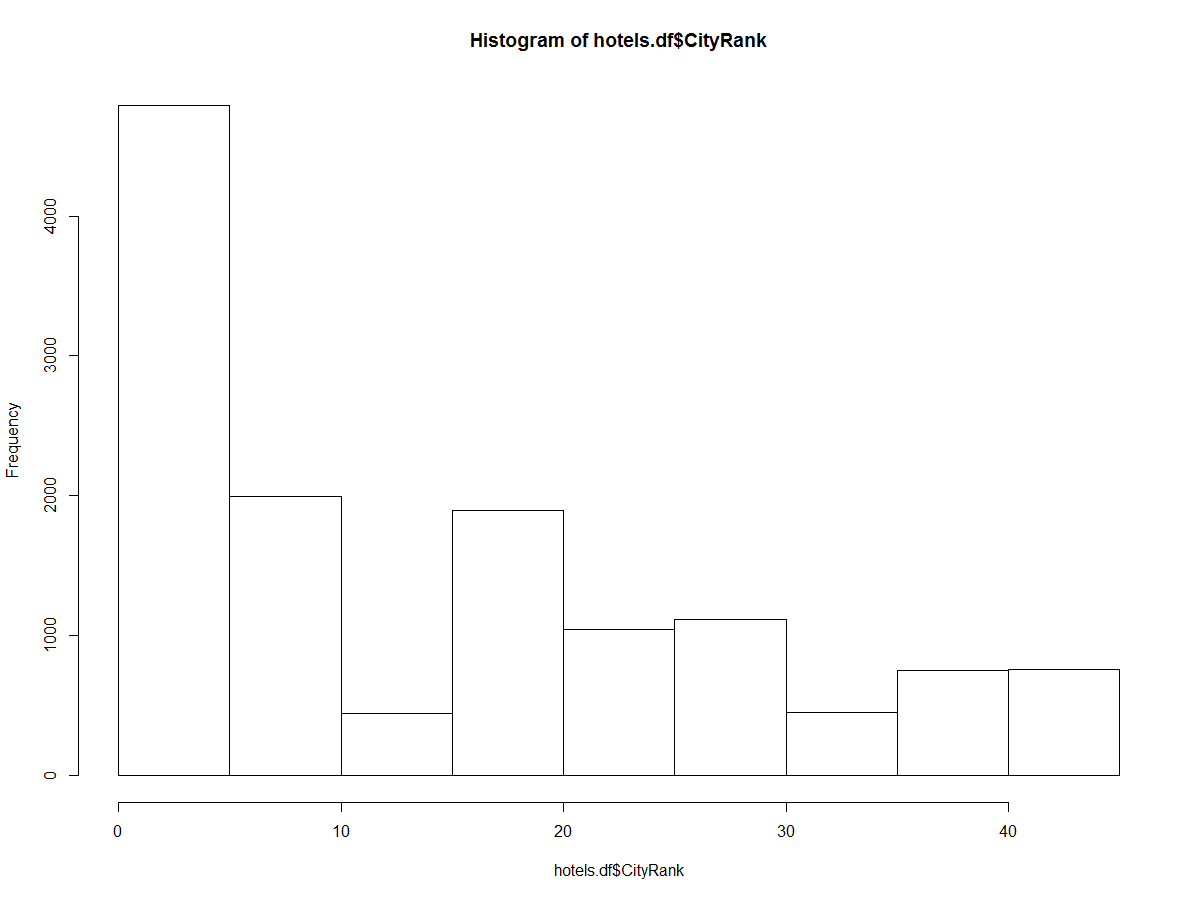




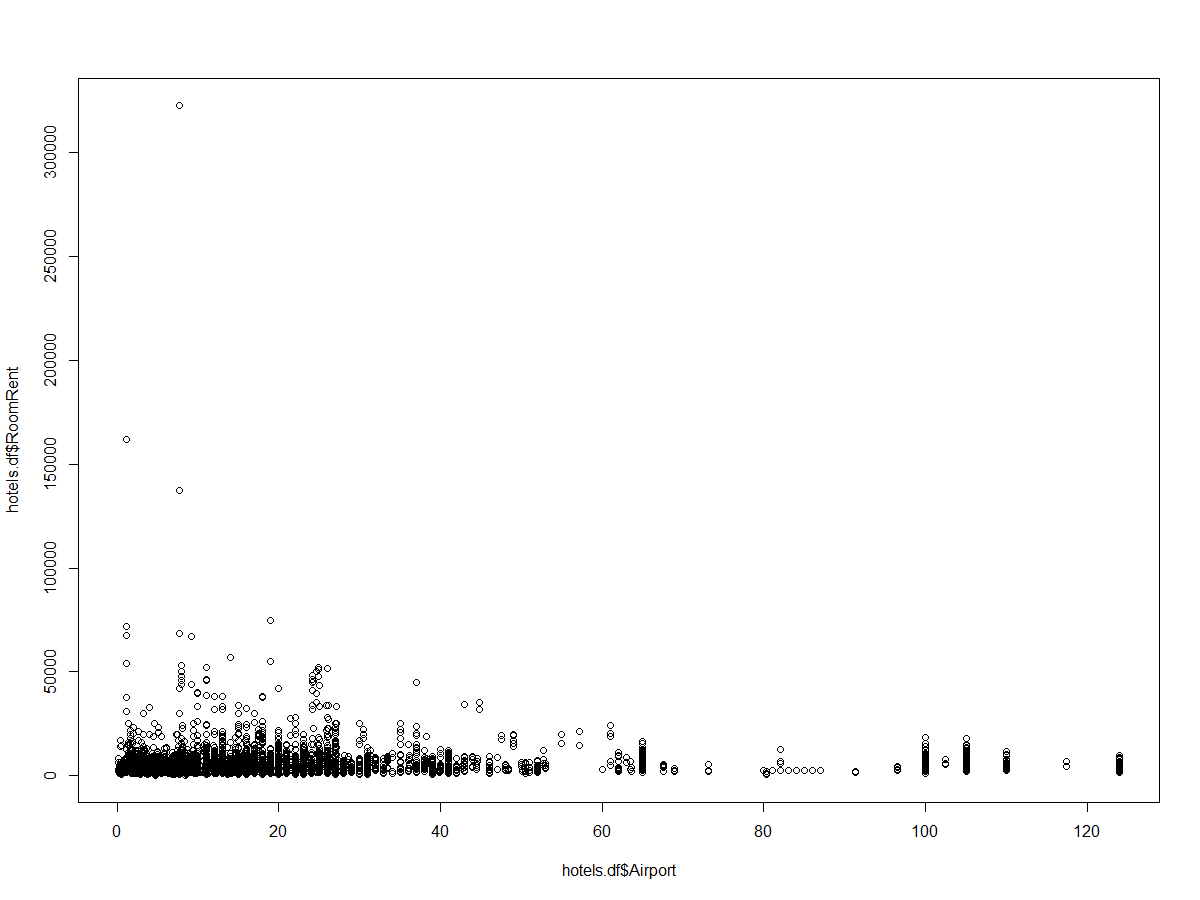


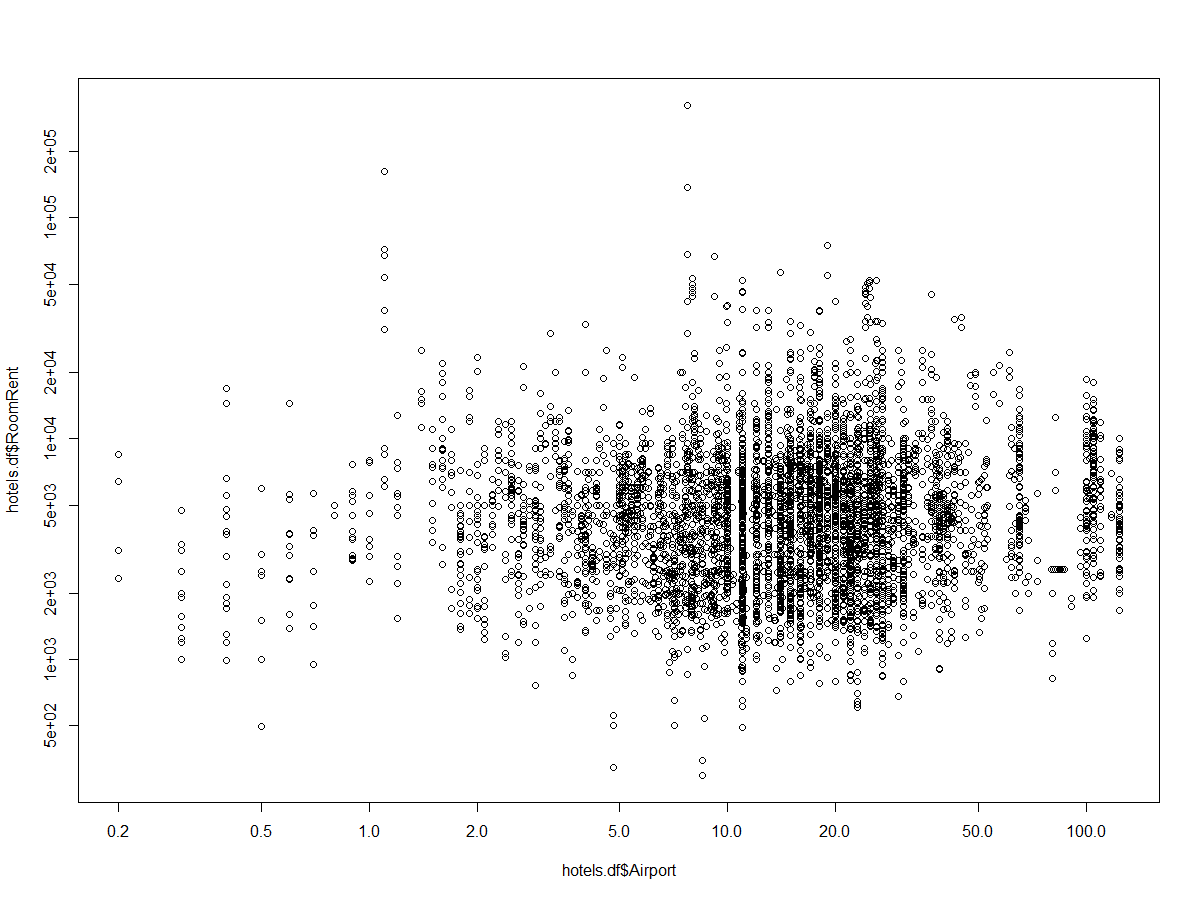


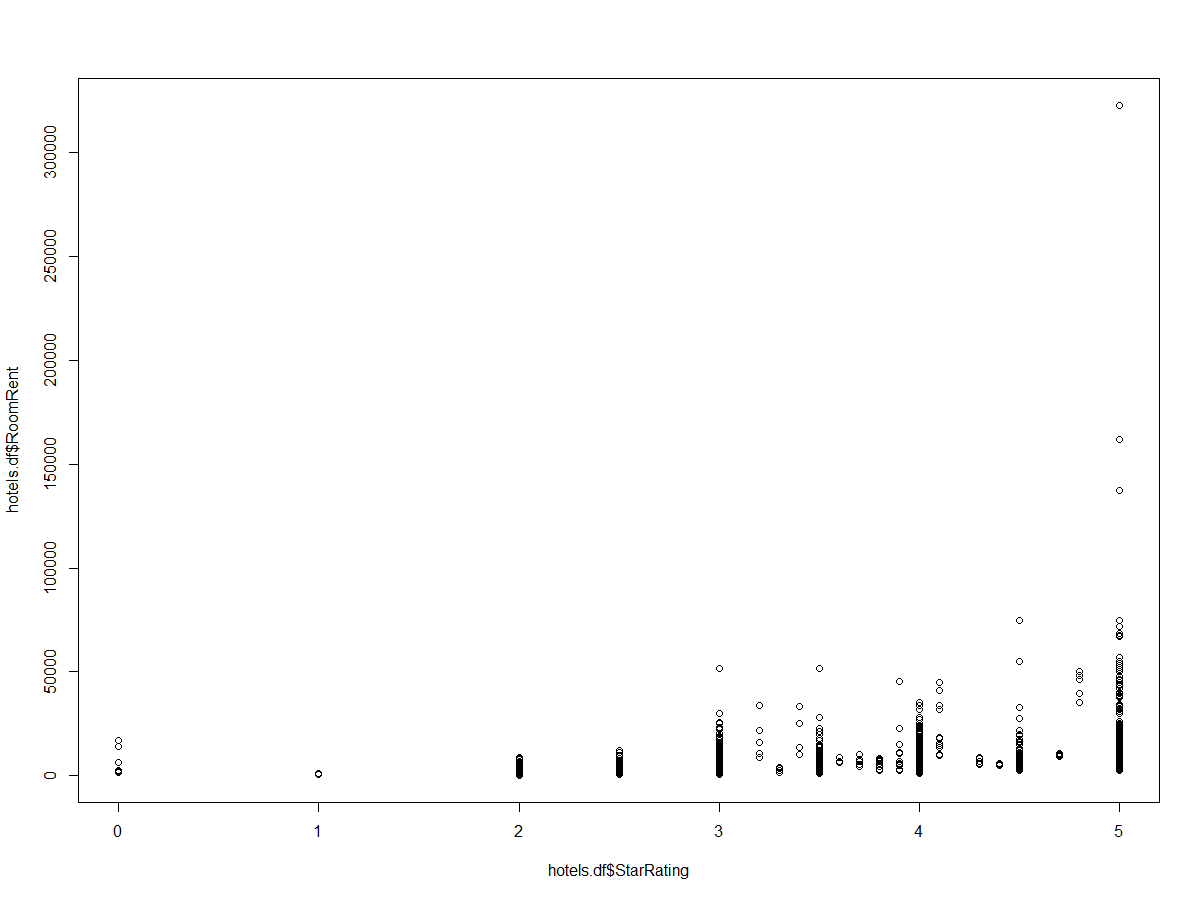


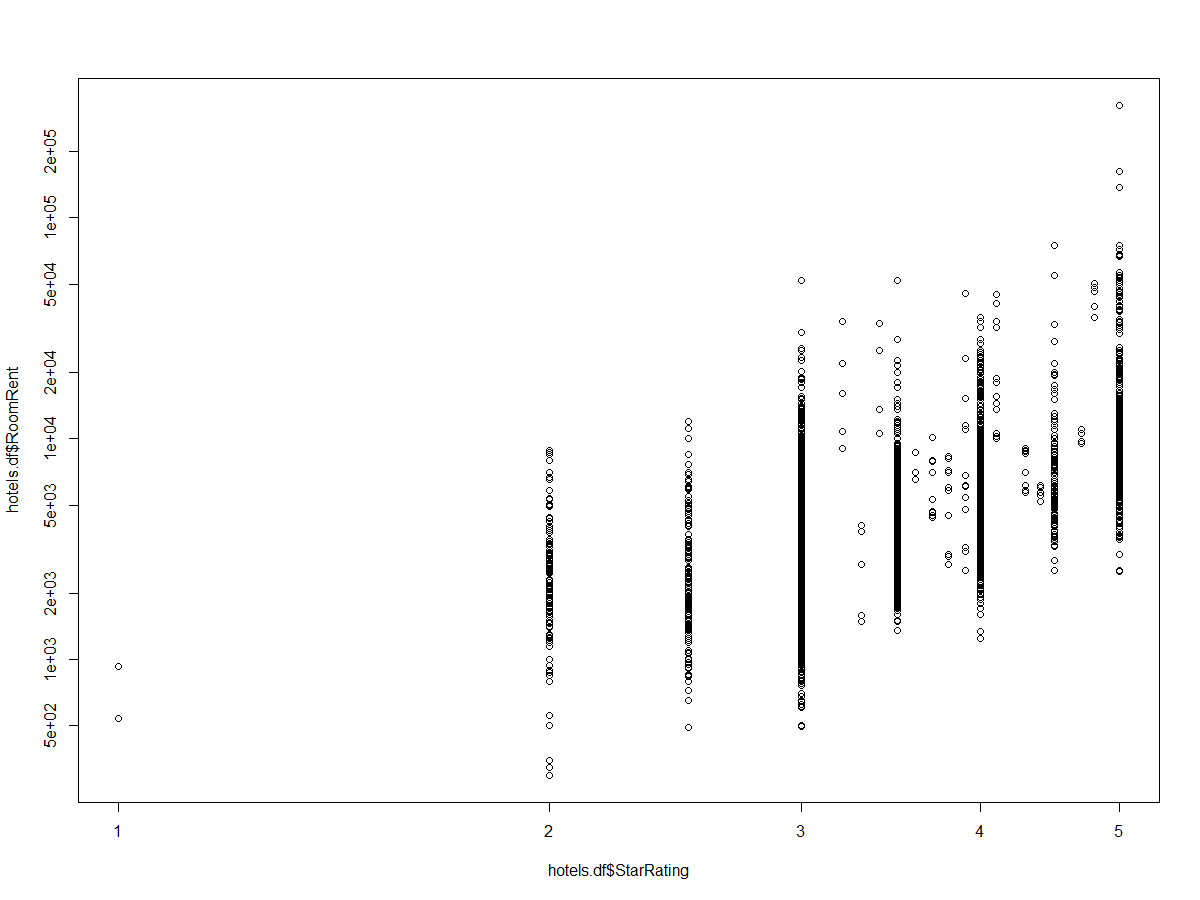


|  |
| --- |
| > plot(hotels.df$Airport,hotels.df$RoomRent)  > plot(hotels.df$Airport,hotels.df$RoomRent,log = 'xy')  >  >  > plot(hotels.df$StarRating,hotels.df$RoomRent)  > plot(hotels.df$StarRating,hotels.df$RoomRent,log = 'xy')  Warning message:  In xy.coords(x, y, xlabel, ylabel, log) :  16 x values <= 0 omitted from logarithmic plot |
|  |
| |  | | --- | | > | |









|  |
| --- |
| > IsWeekend.df <- aggregate(RoomRent ~ IsWeekend, data=hotels.df, mean)  > IsWeekend.df  IsWeekend RoomRent  1 0 5430.835  2 1 5500.129  > plot(IsWeekend.df)  > IsNewYearEve.df <- aggregate(RoomRent ~ IsNewYearEve, data=hotels.df, mean)  > IsNewYearEve.df  IsNewYearEve RoomRent  1 0 5367.606  2 1 6222.826  > IsMetroCity.df <-aggregate(RoomRent ~ IsMetroCity, data=hotels.df, mean)  > IsMetroCity.df  IsMetroCity RoomRent  1 0 5782.794  2 1 4696.073  > IsTouristDestination.df <- aggregate(RoomRent ~ IsTouristDestination, data=hotels.df, mean)  > IsTouristDestination.df  IsTouristDestination RoomRent  1 0 4111.003  2 1 6066.024  > FreeWifi.df <- aggregate(RoomRent ~ FreeWifi, data=hotels.df, mean)  > FreeWifi.df  FreeWifi RoomRent  1 0 5380.004  2 1 5481.518  > HasSwimmingPool.df <-aggregate(RoomRent ~ HasSwimmingPool, data=hotels.df, mean)  > HasSwimmingPool.df  HasSwimmingPool RoomRent  1 0 3775.566  2 1 8549.052  > StarRating.df <- aggregate(RoomRent ~ hotels.df$StarRating, data=hotels.df, mean)  > StarRating.df  hotels.df$StarRating RoomRent  1 0.0 7237.125  2 1.0 686.625  3 2.0 2783.166  4 2.5 2520.816  5 3.0 3694.811  6 3.2 15937.500  7 3.3 2841.062  8 3.4 23437.500  9 3.5 4843.346  10 3.6 7769.500  11 3.7 6701.958  12 3.8 5400.062  13 3.9 13062.750  14 4.0 6393.105  15 4.1 19075.000  16 4.3 7423.125  17 4.4 5563.500  18 4.5 8699.920  19 4.7 10125.000  20 4.8 46752.812  21 5.0 12398.221  > #AirpotDist.df <- aggregate(RoomRent ~ hotels.df$Airport, data=hotels.df, mean)  > #AirpotDist.df  > CityName.df <- aggregate(RoomRent, list(CityName),data= hotels.df,mean)  > CityName.df  Group.1 x  1 Agra 4124.287  2 Ahmedabad 4175.045  3 Amritsar 3444.029  4 Bangalore 4112.803  5 Bhubaneswar 3587.442  6 Chandigarh 4030.940  7 Chennai 4323.647  8 Darjeeling 5458.088  9 Delhi 4318.606  10 Gangtok 4629.648  11 Goa 8170.801  12 Guwahati 5325.812  13 Haridwar 3919.938  14 Hyderabad 3852.175  15 Indore 3414.594  16 Jaipur 7292.022  17 Jaisalmer 5986.072  18 Jodhpur 10661.371  19 Kanpur 3008.562  20 Kochi 6039.609  21 Kolkata 4528.986  22 Lucknow 5879.070  23 Madurai 4768.223  24 Manali 4858.285  25 Mangalore 4110.337  26 Mumbai 6343.730  27 Munnar 7543.500  28 Mysore 3320.869  29 Nainital 6409.833  30 Ooty 6144.257  31 Panchkula 2813.500  32 Pune 3897.652  33 Puri 5708.429  34 Rajkot 4107.078  35 Rishikesh 4943.670  36 Shimla 5780.604  37 Srinagar 10572.025  38 Surat 3660.850  39 Thiruvanthipuram 6726.796  40 Thrissur 3387.844  41 Udaipur 10145.252  42 Varanasi 8675.042  > CityRank.df <- aggregate(RoomRent, list(CityRank),data= hotels.df,mean)  > CityRank.df  Group.1 x  1 0 6343.730  2 1 4318.606  3 2 4112.803  4 3 4323.647  5 4 3852.175  6 5 4175.045  7 6 4528.986  8 7 3660.850  9 8 3897.652  10 9 7292.022  11 10 3387.844  12 11 5879.070  13 12 3008.562  14 13 3444.029  15 14 3414.594  16 16 4124.287  17 17 5342.444  18 18 8170.801  19 19 4107.078  20 20 8675.042  21 21 10572.025  22 22 10661.371  23 23 4030.940  24 24 6726.796  25 25 5325.812  26 26 3320.869  27 27 3587.442  28 28 6664.368  29 29 4110.337  30 30 10145.252  31 32 3919.938  32 33 5708.429  33 34 5780.604  34 35 2813.500  35 36 5458.088  36 37 4943.670  37 38 4629.648  38 39 6144.257  39 40 5986.072  40 42 6409.833  41 43 7543.500  42 44 4858.285 |
|  |
| |  | | --- | | > | |

> library(car)

Attaching package: ‘car’

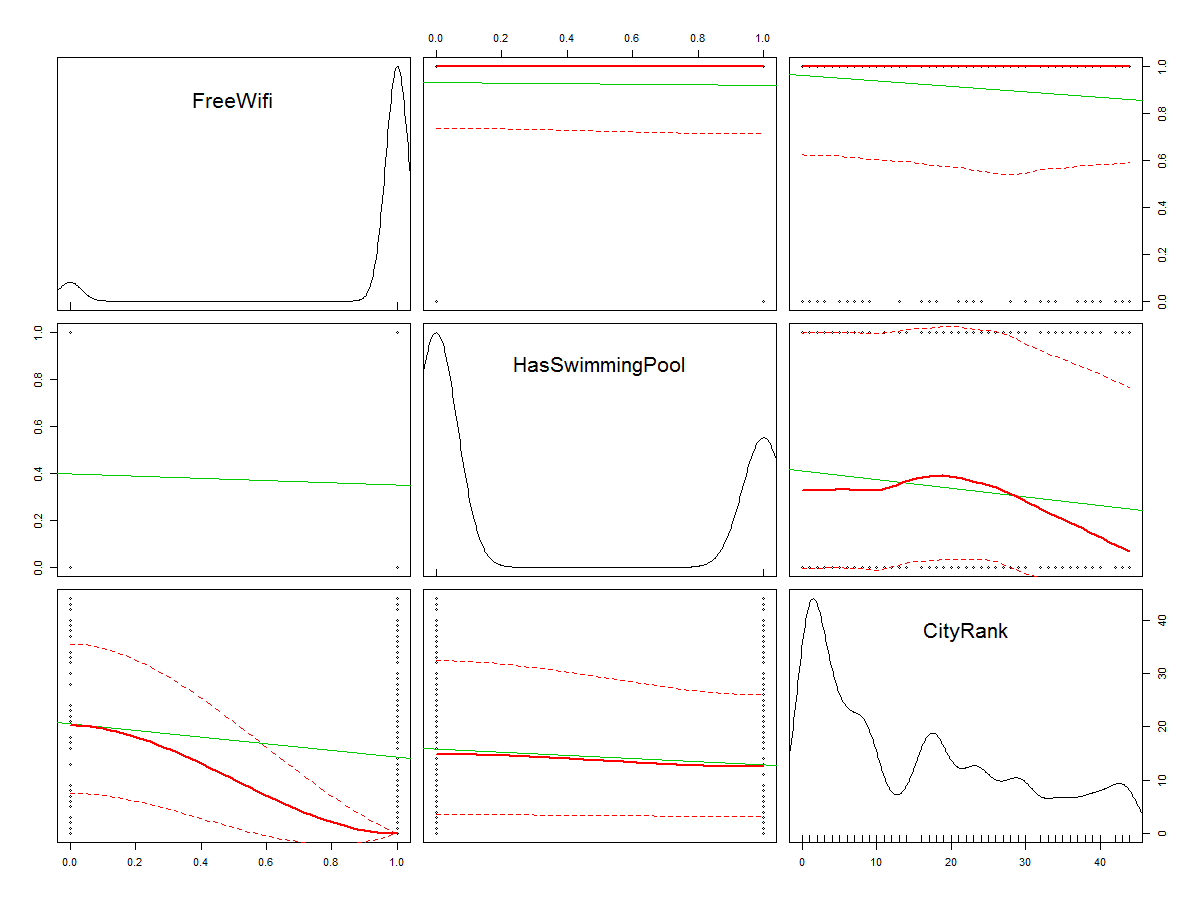
The following object is masked from ‘package:psych’:

logit

> scatterplotMatrix(formula = ~ FreeWifi+HasSwimmingPool+CityRank,

+ cex=0.6,data=hotels.df)

|  |
| --- |
| Warning messages:  1: In smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE, spread = spread, :  could not fit positive part of the spread  2: In smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE, spread = spread, :  could not fit smooth |
|  |
| |  | | --- | | > | |



> ############ 8 Corrgram ##################

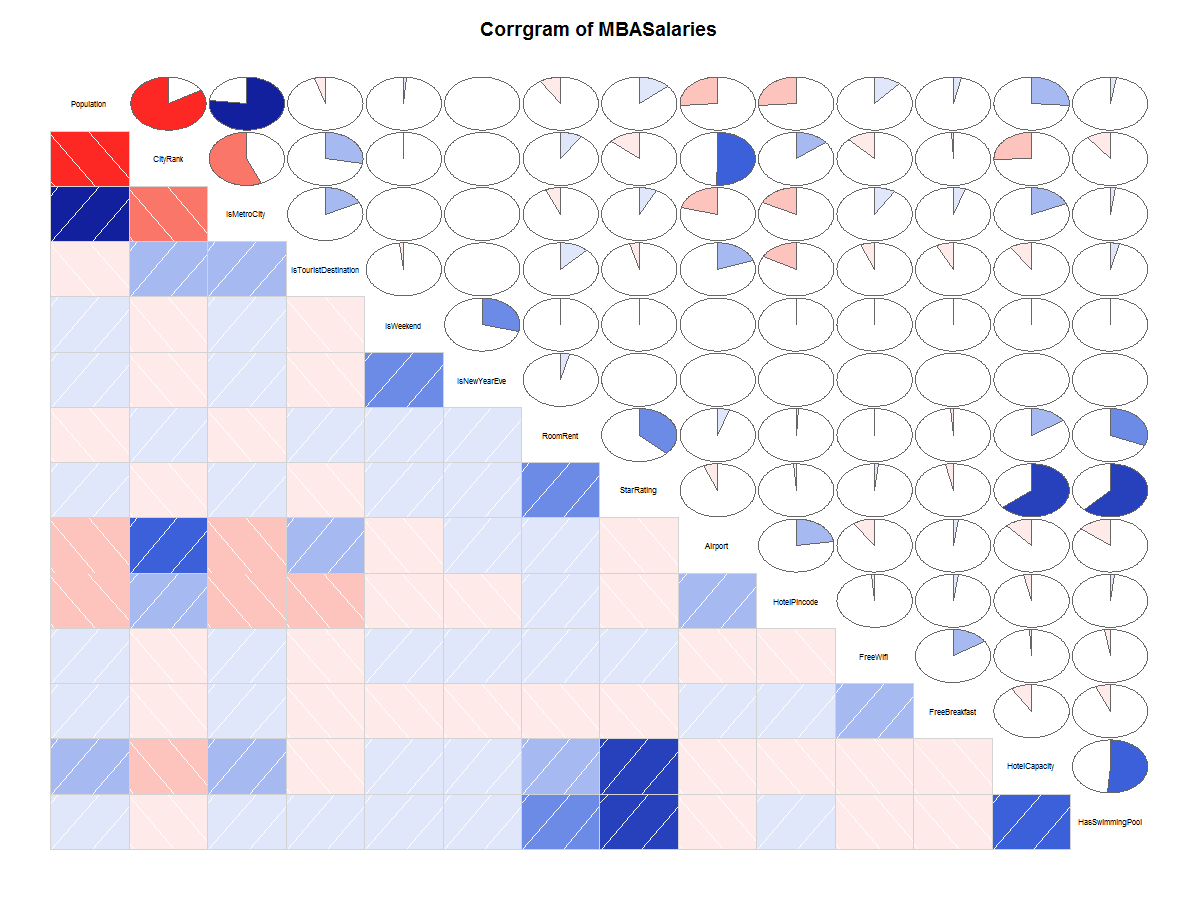
> library(corrgram) # install if needed

> corrgram(hotels.df, order=FALSE, lower.panel=panel.shade,

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

+ main="Corrgram of MBASalaries")

>



|  |
| --- |
| > ################ 9 Variance-Covariance Matrix ###################3  >  > cov(hotels.df)  Error: is.numeric(x) || is.logical(x) is not TRUE  > var(hotels.df)  CityName Population CityRank IsMetroCity IsTouristDestination IsWeekend IsNewYearEve Date  CityName NA NA NA NA NA NA NA NA  Population NA 1.813385e+13 -4.806539e+07 1.481264e+06 -9.431975e+04 2.392784e+04 1.030549e+03 NA  CityRank NA -4.806539e+07 1.825865e+02 -3.439713e+00 1.742936e+00 -4.752639e-02 -2.821413e-03 NA  IsMetroCity NA 1.481264e+06 -3.439713e+00 2.034283e-01 3.655261e-02 3.960871e-04 9.623437e-05 NA  IsTouristDestination NA -9.431975e+04 1.742936e+00 3.655261e-02 2.111386e-01 -4.338823e-03 -3.437089e-04 NA  IsWeekend NA 2.392784e+04 -4.752639e-02 3.960871e-04 -4.338823e-03 2.349359e-01 4.677330e-02 NA  IsNewYearEve NA 1.030549e+03 -2.821413e-03 9.623437e-05 -3.437089e-04 4.677330e-02 1.089294e-01 NA  Date NA NA NA NA NA NA NA NA  HotelName NA NA NA NA NA NA NA NA  RoomRent NA -2.770733e+09 9.312879e+03 -2.210697e+02 4.127803e+02 1.627952e+01 9.315860e+01 NA  StarRating NA 4.319641e+05 -1.362962e+00 2.646908e-02 -1.409235e-02 2.337998e-03 5.892573e-04 NA  Airport NA -2.517034e+07 1.555893e+02 -2.128620e+00 2.033292e+00 -3.005887e-02 3.454573e-03 NA  HotelAddress NA NA NA NA NA NA NA NA  HotelPincode NA -2.861988e+11 5.086601e+05 -2.058668e+04 -2.034654e+04 -8.116700e+02 -1.810728e+02 NA  HotelDescription NA NA NA NA NA NA NA NA  FreeWifi NA 1.260024e+05 -4.299074e-01 1.026080e-02 -7.412348e-03 3.760097e-04 2.410429e-06 NA  FreeBreakfast NA 7.403547e+04 -5.600196e-02 1.106138e-02 -1.572242e-02 -1.761081e-03 -4.105611e-04 NA  HotelCapacity NA 8.487461e+07 -2.653170e+02 6.471184e+00 -3.323850e+00 2.343426e-01 3.422592e-02 NA  HasSwimmingPool NA 5.353717e+04 -6.661803e-01 4.623728e-03 9.274210e-03 1.044390e-03 1.773436e-04 NA  HotelName RoomRent StarRating Airport HotelAddress HotelPincode HotelDescription  CityName NA NA NA NA NA NA NA  Population NA -2.770733e+09 4.319641e+05 -2.517034e+07 NA -2.861988e+11 NA  CityRank NA 9.312879e+03 -1.362962e+00 1.555893e+02 NA 5.086601e+05 NA  IsMetroCity NA -2.210697e+02 2.646908e-02 -2.128620e+00 NA -2.058668e+04 NA  IsTouristDestination NA 4.127803e+02 -1.409235e-02 2.033292e+00 NA -2.034654e+04 NA  IsWeekend NA 1.627952e+01 2.337998e-03 -3.005887e-02 NA -8.116700e+02 NA  IsNewYearEve NA 9.315860e+01 5.892573e-04 3.454573e-03 NA -1.810728e+02 NA  Date NA NA NA NA NA NA NA  HotelName NA NA NA NA NA NA NA  RoomRent NA 5.377460e+07 2.048375e+03 8.287179e+03 NA 1.764935e+07 NA  StarRating NA 2.048375e+03 5.718875e-01 -1.048528e+00 NA -1.890003e+03 NA  Airport NA 8.287179e+03 -1.048528e+00 5.180133e+02 NA 1.322589e+06 NA  HotelAddress NA NA NA NA NA NA NA  HotelPincode NA 1.764935e+07 -1.890003e+03 1.322589e+06 NA 6.751553e+10 NA  HotelDescription NA NA NA NA NA NA NA  FreeWifi NA 6.968634e+00 3.568375e-03 -5.636666e-01 NA -8.512421e+02 NA  FreeBreakfast NA -3.502085e+01 -1.187171e-02 2.637857e-01 NA 3.085468e+03 NA  HotelCapacity NA 8.875341e+04 3.695522e+01 -2.053202e+02 NA -6.989568e+05 NA  HasSwimmingPool NA 1.094202e+03 2.238337e-01 -1.543721e+00 NA 2.522999e+03 NA  FreeWifi FreeBreakfast HotelCapacity HasSwimmingPool  CityName NA NA NA NA  Population 1.260024e+05 7.403547e+04 8.487461e+07 5.353717e+04  CityRank -4.299074e-01 -5.600196e-02 -2.653170e+02 -6.661803e-01  IsMetroCity 1.026080e-02 1.106138e-02 6.471184e+00 4.623728e-03  IsTouristDestination -7.412348e-03 -1.572242e-02 -3.323850e+00 9.274210e-03  IsWeekend 3.760097e-04 -1.761081e-03 2.343426e-01 1.044390e-03  IsNewYearEve 2.410429e-06 -4.105611e-04 3.422592e-02 1.773436e-04  Date NA NA NA NA  HotelName NA NA NA NA  RoomRent 6.968634e+00 -3.502085e+01 8.875341e+04 1.094202e+03  StarRating 3.568375e-03 -1.187171e-02 3.695522e+01 2.238337e-01  Airport -5.636666e-01 2.637857e-01 -2.053202e+02 -1.543721e+00  HotelAddress NA NA NA NA  HotelPincode -8.512421e+02 3.085468e+03 -6.989568e+05 2.522999e+03  HotelDescription NA NA NA NA  FreeWifi 6.864713e-02 1.978499e-02 -1.748230e-01 -3.019890e-03  FreeBreakfast 1.978499e-02 2.277840e-01 -3.189287e+00 -1.405801e-02  HotelCapacity -1.748230e-01 -3.189287e+00 5.877268e+03 1.868424e+01  HasSwimmingPool -3.019890e-03 -1.405801e-02 1.868424e+01 2.292249e-01  Warning message:  In var(hotels.df) : NAs introduced by coercion |
|  |
| |  | | --- | | > | |