R CODE OUTPUT

1) Reading Data in R

```
setwd("D:/Master/data analytics")

cities.df <- read.csv(paste("Cities42.csv"))

library("utils", lib.loc="C:/Program Files/R/R-3.3.3/library")

View(cities.df)
```

2) Summary of data

a) View(summary(cities.df))

Var1	Var2	Freq
1	X	Min. : 1
2	X	1st Qu.: 3309
3	X	Median : 6616
4	X	Mean : 6616
5	X	3rd Qu.: 9924
6	X	Max. :13232
7	X	*NA*
8	CityName	Delhi :2048
9	CityName	Jaipur : 768

Var1	Var2	Freq
10	CityName	Mumbai : 712
11	CityName	Bangalore: 656
12	CityName	Goa : 624
13	CityName	Kochi : 608
14	CityName	(Other) :7816
15	Population	Min. : 8096
16	Population	1st Qu.: 744983
17	Population	Median : 3046163
18	Population	Mean : 4416837
19	Population	3rd Qu.: 8443675
20	Population	Max. :12442373
21	Population	*NA*
22	CityRank	Min. : 0.00
23	CityRank	1st Qu.: 2.00
24	CityRank	Median : 9.00

Var1	Var2	Freq
25	CityRank	Mean :14.83
26	CityRank	3rd Qu.:24.00
27	CityRank	Max. :44.00
28	CityRank	*NA*
29	IsMetroCity	Min. :0.0000
30	IsMetroCity	1st Qu.:0.0000
31	IsMetroCity	Median :0.0000
32	IsMetroCity	Mean :0.2842
33	IsMetroCity	3rd Qu.:1.0000
34	IsMetroCity	Max. :1.0000
35	IsMetroCity	*NA*
36	IsTouristDestination	Min. :0.0000
37	IsTouristDestination	1st Qu.:0.0000
38	IsTouristDestination	Median :1.0000

Var1	Var2	Freq
39	IsTouristDestination	Mean :0.6972
40	IsTouristDestination	3rd Qu.:1.0000
41	IsTouristDestination	Max. :1.0000
42	IsTouristDestination	*NA*
43	IsWeekend	Min. :0.0000
44	IsWeekend	1st Qu.:0.0000
45	IsWeekend	Median :1.0000
46	IsWeekend	Mean :0.6228
47	IsWeekend	3rd Qu.:1.0000
48	IsWeekend	Max. :1.0000
49	IsWeekend	*NA*
50	IsNewYearEve	Min. :0.0000
51	IsNewYearEve	1st Qu.:0.0000

Var1	Var2	Freq
52	IsNewYearEve	Median :0.0000
53	IsNewYearEve	Mean :0.1244
54	IsNewYearEve	3rd Qu.:0.0000
55	IsNewYearEve	Max. :1.0000
56	IsNewYearEve	*NA*
57	Date	Dec 21 2016:1611
58	Date	Dec 24 2016:1611
59	Date	Dec 25 2016:1611
60	Date	Dec 28 2016:1611
61	Date	Dec 31 2016:1611
62	Date	Dec 18 2016:1608
63	Date	(Other) :3569
64	HotelName	Vivanta by Taj : 32
65	HotelName	Goldfinch Hotel : 24
66	HotelName	OYO Rooms : 24

Var1	Var2	Freq
67	HotelName	The Gordon House Hotel: 24
68	HotelName	Apnayt Villa : 16
69	HotelName	Bentleys Hotel Colaba : 16
70	HotelName	(Other) :13096
71	RoomRent	Min. : 299
72	RoomRent	1st Qu.: 2436
73	RoomRent	Median : 4000
74	RoomRent	Mean : 5474
75	RoomRent	3rd Qu.: 6299
76	RoomRent	Max. :322500
77	RoomRent	*NA*
78	StarRating	Min. :0.000
79	StarRating	1st Qu.:3.000
80	StarRating	Median :3.000
81	StarRating	Mean :3.459

Var1	Var2	Freq
82	StarRating	3rd Qu.:4.000
83	StarRating	Max. :5.000
84	StarRating	*NA*
85	Airport	Min. : 0.20
86	Airport	1st Qu.: 8.40
87	Airport	Median : 15.00
88	Airport	Mean : 21.16
89	Airport	3rd Qu.: 24.00
90	Airport	Max. :124.00
91	Airport	*NA*
92	HotelAddress	The Mall, Shimla : 32
93	HotelAddress	#2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16
94	HotelAddress	121, City Terrace, Walchand Hirachand Marg, Mumbai, Maharashtra : 16
95	HotelAddress	14-4507/9, Balmatta Road, Near Jyothi Circle, Hampankatta : 16

Var1	Var2	Freq
96	HotelAddress	144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16
97	HotelAddress	17, Oliver Road, Colaba, Mumbai, Maharashtra : 16
98	HotelAddress	(Other) :13120
99	HotelPincode	Min. : 100025
100	HotelPincode	1st Qu.: 221001
101	HotelPincode	Median : 395003
102	HotelPincode	Mean : 397430
103	HotelPincode	3rd Qu.: 570001
104	HotelPincode	Max. :7000157
105	HotelPincode	*NA*
106	HotelDescription	3 : 120
10	HotelDescrip	Abc : 112

Var1	Var2	Freq
7	tion	
108	HotelDescription	3-star hotel: 104
109	HotelDescription	3.5 : 88
110	HotelDescription	4 : 72
111	HotelDescription	(Other) :12728
112	HotelDescription	NA's : 8
113	FreeWifi	Min. :0.0000
114	FreeWifi	1st Qu.:1.0000
115	FreeWifi	Median :1.0000
116	FreeWifi	Mean :0.9259
117	FreeWifi	3rd Qu.:1.0000

Var1	Var2	Freq
118	FreeWifi	Max. :1.0000
119	FreeWifi	*NA*
120	FreeBreakfast	Min. :0.0000
121	FreeBreakfast	1st Qu.:0.0000
122	FreeBreakfast	Median :1.0000
123	FreeBreakfast	Mean :0.6491
124	FreeBreakfast	3rd Qu.:1.0000
125	FreeBreakfast	Max. :1.0000
126	FreeBreakfast	*NA*
127	HotelCapacity	Min. : 0.00
128	HotelCapacity	1st Qu.: 16.00

Var1	Var2	Freq
129	HotelCapacity	Median : 34.00
130	HotelCapacity	Mean : 62.51
131	HotelCapacity	3rd Qu.: 75.00
132	HotelCapacity	Max. :600.00
133	HotelCapacity	*NA*
134	HasSwimmingPool	Min. :0.0000
135	HasSwimmingPool	1st Qu.:0.0000
136	HasSwimmingPool	Median :0.0000
137	HasSwimmingPool	Mean :0.3558
138	HasSwimmingPool	3rd Qu.:1.0000
139	HasSwimmingPool	Max. :1.0000

Var1	Var2	Freq
140	HasSwimmingPool	*NA*

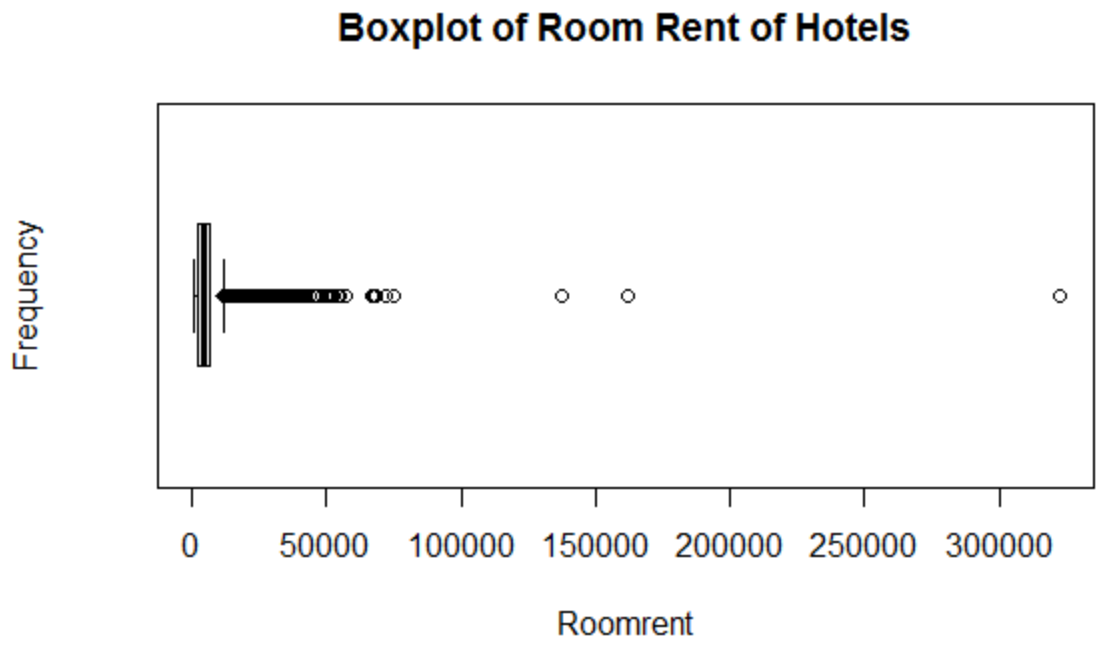
b) View(describe(cities.df))

Vars	n	mean	sd	median	trimmed	mad	min	max	Range	skew	kurtosi
X	1	13232	6.616500e+03	3.819894e+03	6616.5	6.616500e+03	4904.4408	1.0	13232	13231.0	0.0000000
CityName*	2	13232	1.806651e+01	1.171740e+01	16.0	1.729152e+01	11.8608	1.0	42	41.0	0.4758382
Population	3	13232	4.416837e+06	4.258386e+06	3046163.0	4.040816e+06	3846498.9528	8096.0	12442373	12434277.0	0.6780918
CityRank	4	13232	1.483374e+01	1.351246e+01	9.0	1.330266e+01	11.8608	0.0	44	44.0	0.6859426
IsMetroCity	5	13232	2.841596e-01	4.510303e-01	0.0	2.302097e-01	0.0000	0.0	1	1.0	0.9570270
IsTouristDestination	6	13232	6.971735e-01	4.594982e-01	1.0	7.464576e-01	0.0000	0.0	1	1.0	-0.8581475
IsWeekend	7	13232	6.228083e-01	4.847018e-01	1.0	6.535046e-01	0.0000	0.0	1	1.0	-0.5066993
IsNewYearEve	8	13232	1.243954e-01	3.300446e-01	0.0	3.051200e-02	0.0000	0.0	1	1.0	2.2759120
Date*	9	13232	1.425816e+01	2.817908e+00	14.0	1.438891e+01	2.9652	1.0	20	19.0	-1.0489570
HotelName*	10	13232	8.411877e+02	4.881570e+02	827.0	8.411780e+02	641.9658	1.0	1670	1669.0	0.0097381
RoomRent	11	13232	5.473992e+03	7.333117e+03	4000.0	4.383334e+03	2653.8540	299.0	322500	322201.0	16.7546833
StarRating	12	13232	3.458933e+00	7.562325e-01	3.0	3.401152e+00	0.7413	0.0	5	5.0	0.4835933
Airport	13	13232	2.115874e+01	2.275991e+01	15.0	1.639458e+01	11.1195	0.2	124	123.8	2.7326605
HotelAddress*	14	13232	1.202531e+03	5.821660e+02	1261.0	1.233254e+03	668.6526	1.0	2108	2107.0	-0.3721332

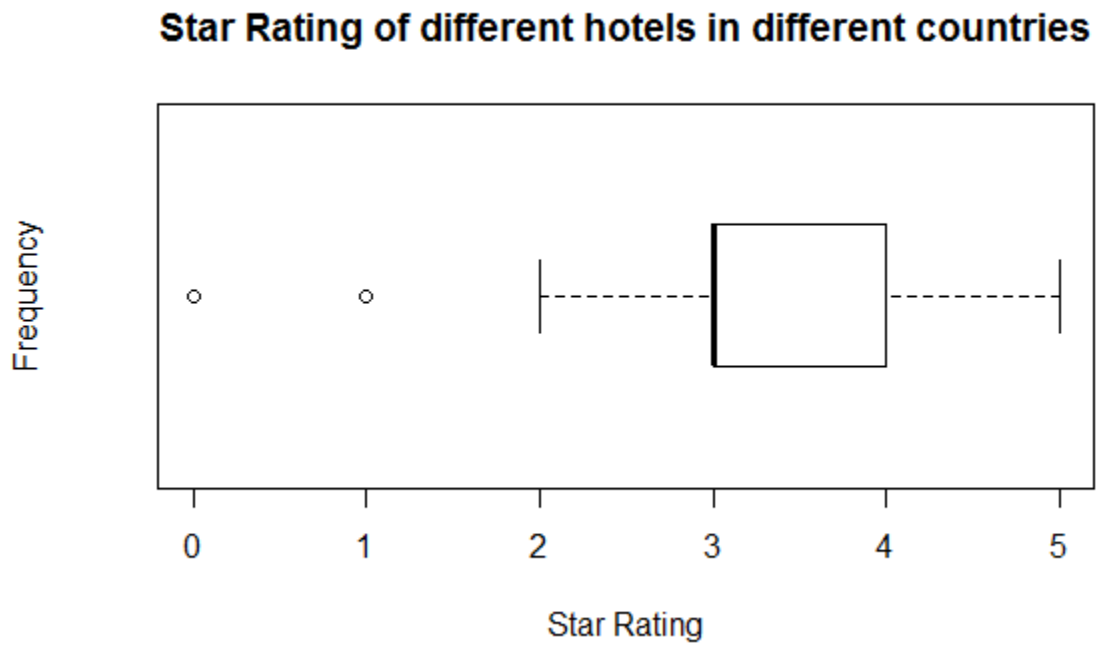
Vars	n	mean	sd	median	trimmed	mad	min	max	Range	skew	kurtosi
HotelPincode	15	13232	3.974303e+05	2.598375e+05	395003.0	3.885405e+05	257975.3652	100025.0	7000157	6900132.0	9.9889631
HotelDescription*	16	13224	5.813384e+02	3.632607e+02	567.0	5.753715e+02	472.9494	1.0	1226	1225.0	0.1116900
FreeWifi	17	13232	9.258615e-01	2.620060e-01	1.0	1.000000e+00	0.0000	0.0	1	1.0	-3.2505317
FreeBreakfast	18	13232	6.491082e-01	4.772672e-01	1.0	6.863782e-01	0.0000	0.0	1	1.0	-0.6247945
HotelCapacity	19	13232	6.251164e+01	7.666334e+01	34.0	4.603382e+01	28.1694	0.0	600	600.0	2.9518438
HasSwimmingPool	20	13232	3.558041e-01	4.787743e-01	0.0	3.197619e-01	0.0000	0.0	1	1.0	0.6023087

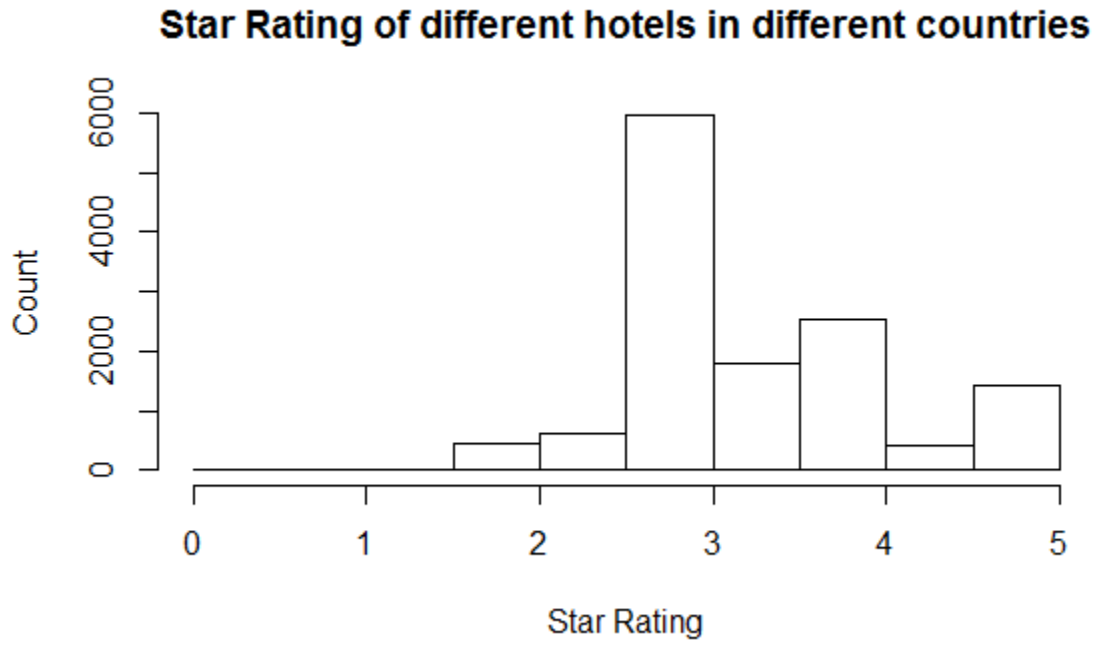
6) Boxplots and Histograms

a) Room Rent

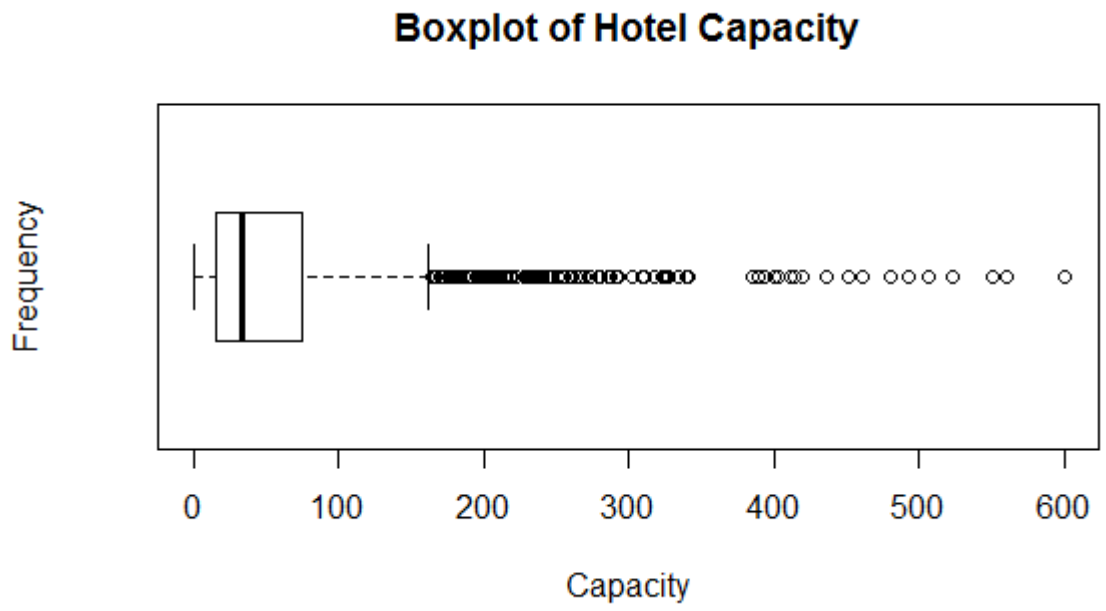


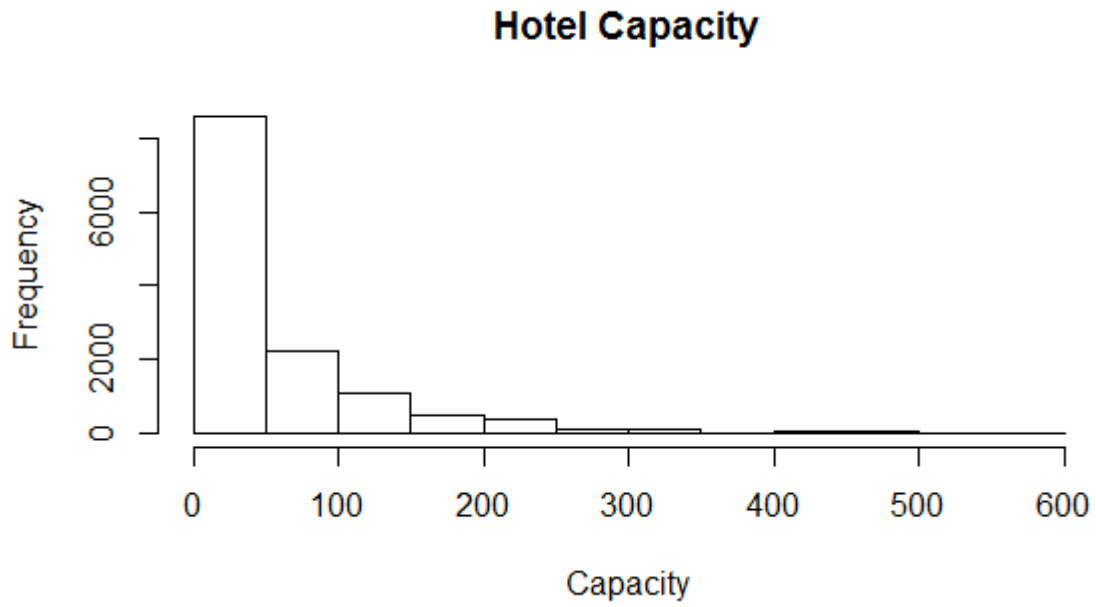
b) Star Rating





c) Hotel Capacity



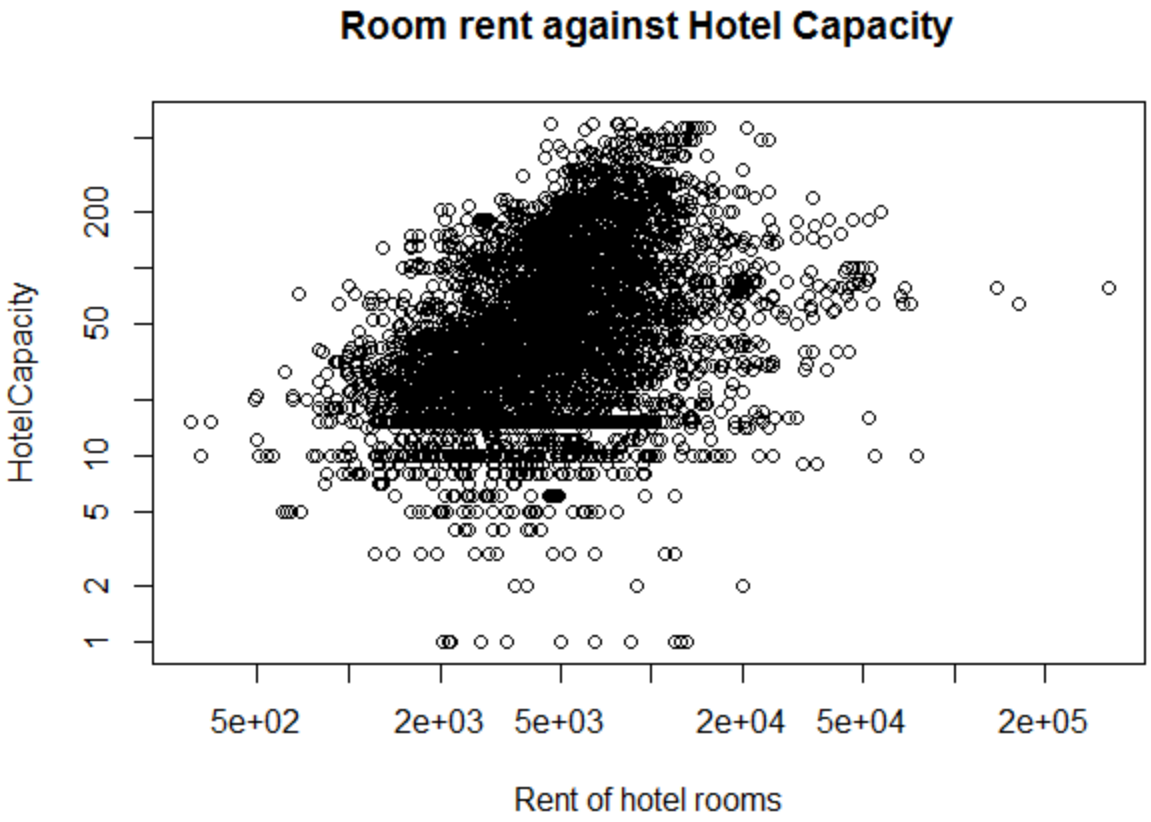


d) Table of Has a Swimming Pool

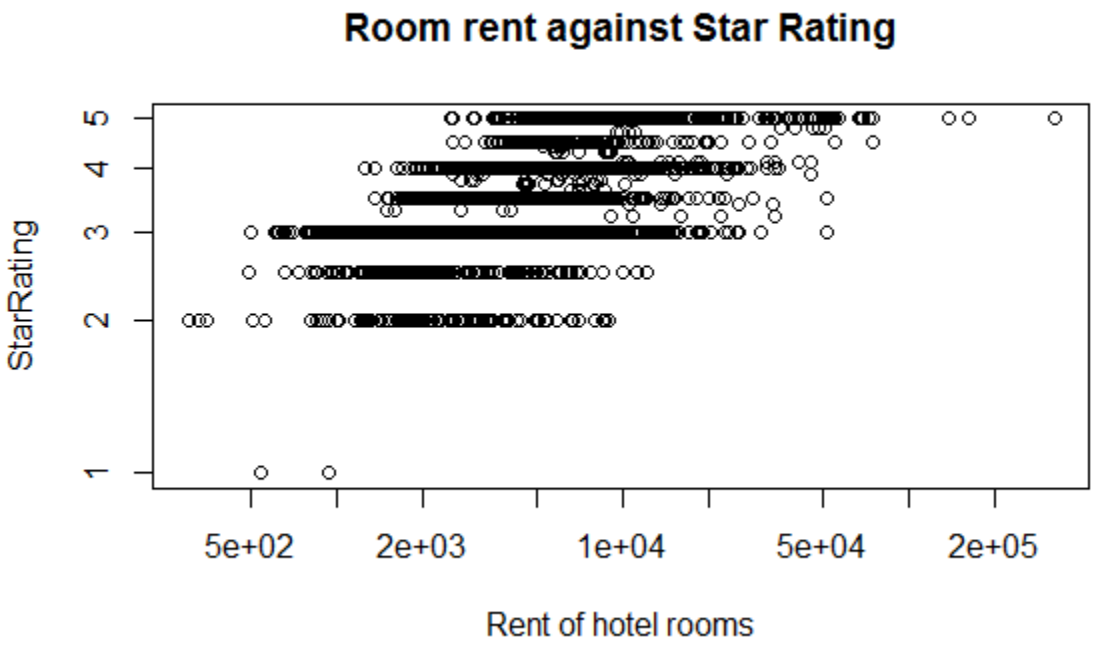
```
HasSwimmingPool
   0     1
8524  4708
```

7) Scatterplots

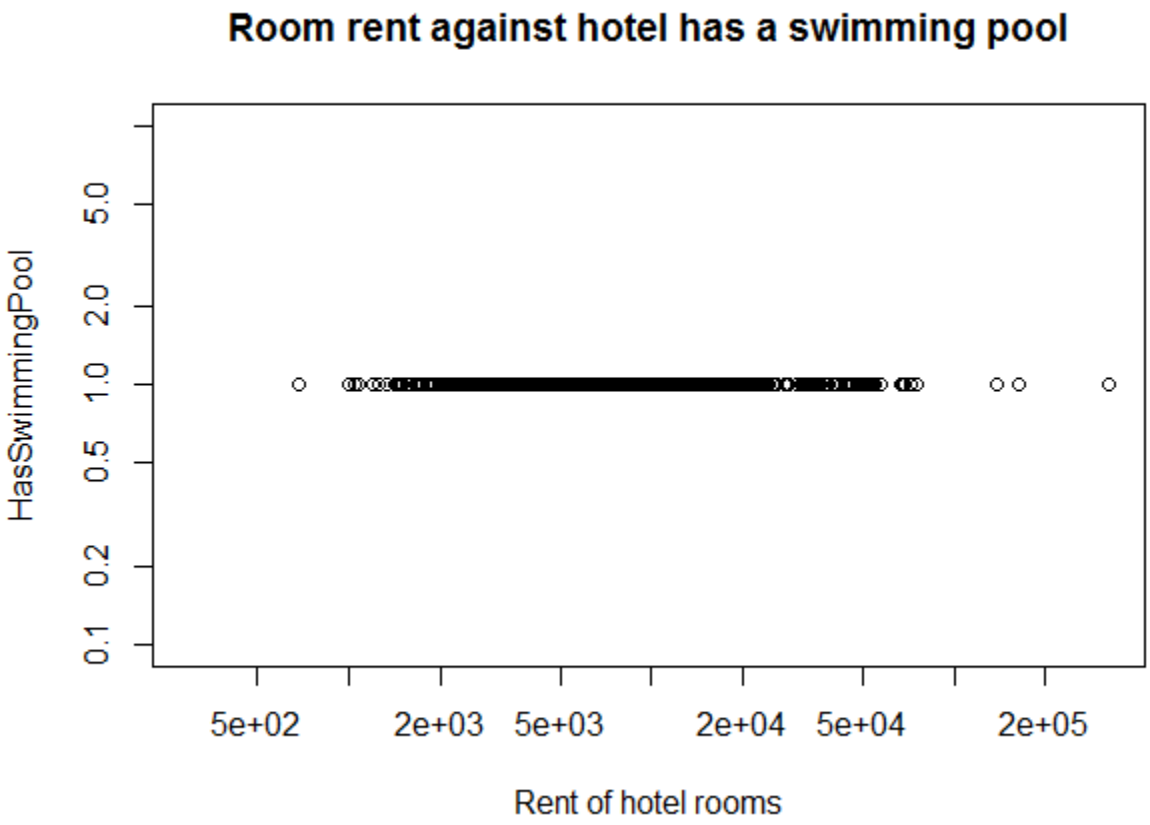
a) For Room Rent against Hotel Capacity



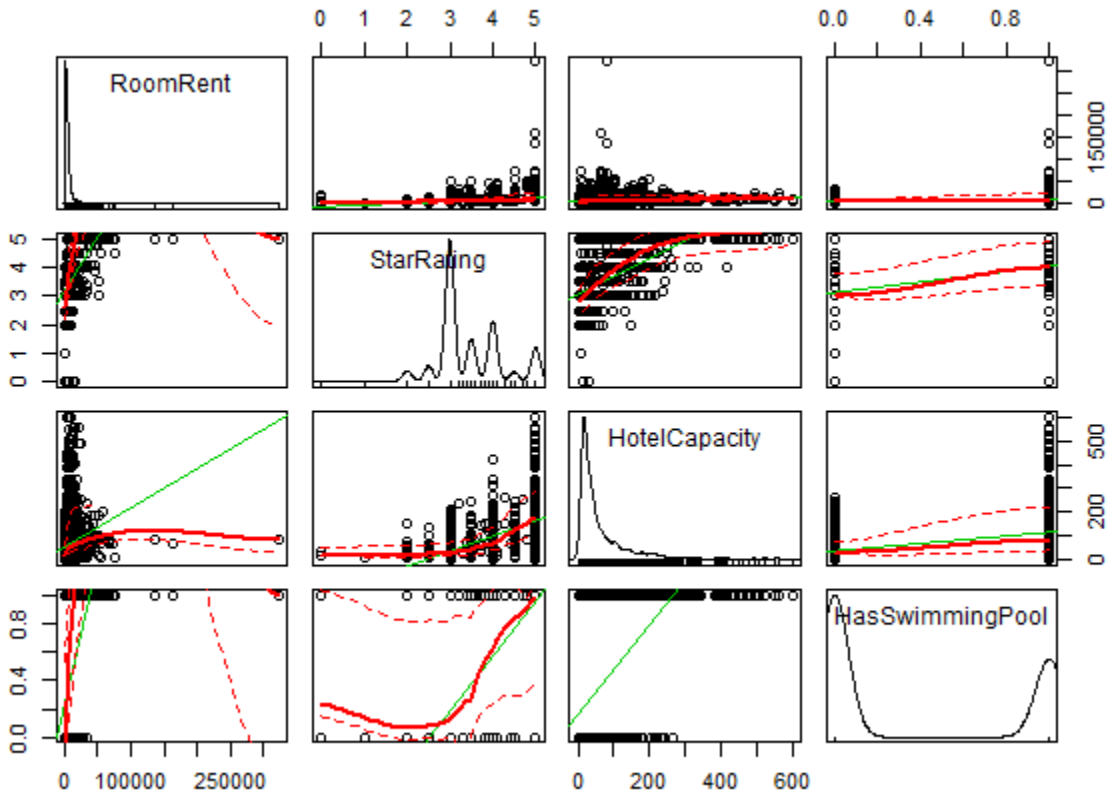
b) For room rent against Star Rating



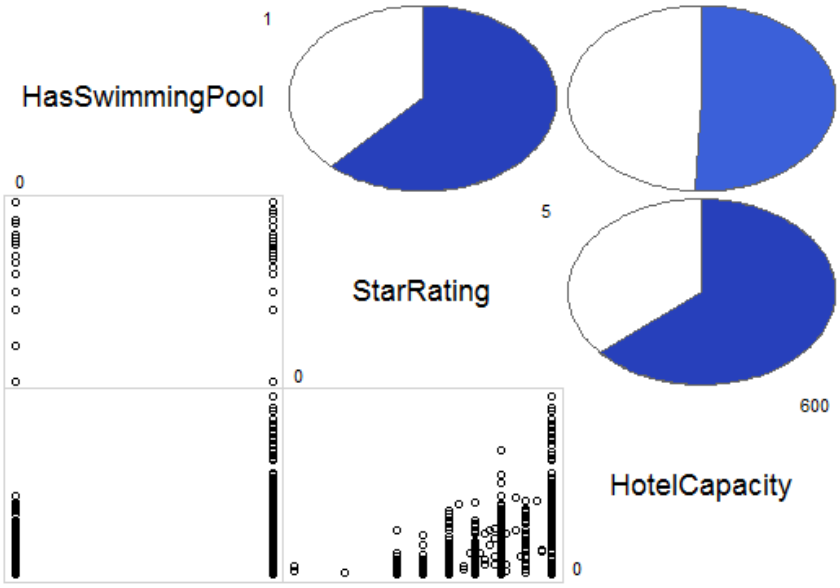
c) For room rating against Has a Swimming Pool

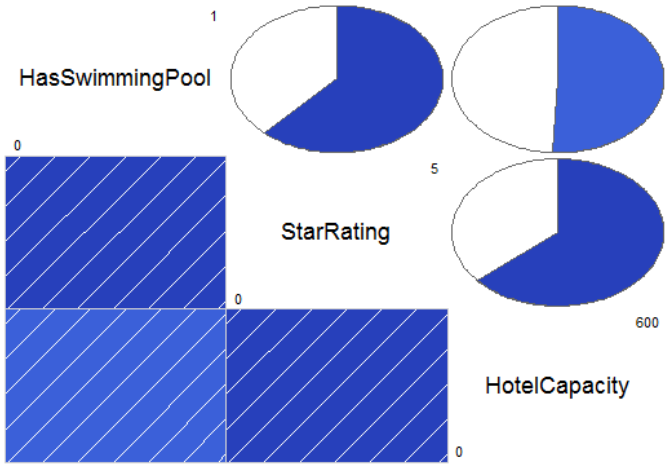


d) Scatterplot matrix of room rent against


RoomRent
StarRating
HotelCapacity
HasSwimmingPool
0 1 2 3 4 5
0.0 0.4 0.8
0 150000
0 100000 250000
0 200 400 600
0 200 500

8) Corrgram





10) Covariance – Correlation Matrix

a) Covariance Matrix

	RoomRent	StarRating	HotelCapacity	HasSwimmingPool
RoomRent	53774601.806	2048.3754792	88753.41284	1094.2016799
StarRating	2048.375	0.5718875	36.95522	0.2238337
HotelCapacity	88753.413	36.9552206	5877.26810	18.6842420

```
HasSwimmingPool     1094.202    0.2238337      18.68424       0.2292249
```

b) Correlation Matrix

```
                RoomRent   StarRating    HotelCapacity   HasSwimmingPool
RoomRent        1.0000000  0.3693734     0.1578733       0.3116577
StarRating      0.3693734  1.0000000     0.6374303       0.6182147
HotelCapacity   0.1578733  0.6374303     1.0000000       0.5090458
HasSwimmingPool 0.3116577  0.6182147     0.5090458       1.0000000
```

11) T – Tests

a) Room rent against Star Rating

```
        Welch Two Sample t-test

data:  hotel.df$RoomRent and hotel.df$StarRating
t = 85.813, df = 13231, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5345.575 5595.491
sample estimates:
  mean of x   mean of y
5473.991838    3.458933
```

b) Room rent against Hotel Capacity

```
        Welch Two Sample t-test

data:  hotel.df$RoomRent and hotel.df$HotelCapacity
t = 84.882, df = 13234, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5286.515 5536.445
sample estimates:
 mean of x  mean of y
5473.99184   62.51164
```

c) Room rent against HasSwimmingPool

```
        Welch Two Sample t-test

data:  hotel.df$RoomRent and hotel.df$HasSwimmingPool
t = 85.862, df = 13231, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5348.678 5598.594
sample estimates:
   mean of x    mean of y
5473.9918380    0.3558041
```

12) Regression Models

a) Room Rent against Star Rating

```
Residuals:
   Min     1Q Median     3Q    Max
 -8494  -2312  -1130    811 311506

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6915.15     277.38  -24.93   <2e-16 ***
StarRating   3581.78      78.34   45.72   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6815 on 13230 degrees of freedom
Multiple R-squared:  0.1364,   Adjusted R-squared:  0.1364
F-statistic:  2090 on 1 and 13230 DF,  p-value: < 2.2e-16
```

b) Room rent against Hotel Capacity

```
Residuals:
   Min     1Q Median     3Q    Max
 -8916  -2682  -1559    522 316777

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   4529.9952    81.2288   55.77   <2e-16 ***
HotelCapacity   15.1011     0.8212   18.39   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7241 on 13230 degrees of freedom
Multiple R-squared:  0.02492,  Adjusted R-squared:  0.02485
F-statistic: 338.2 on 1 and 13230 DF,  p-value: < 2.2e-16
```

c) Room rent against HasSwimmimgPool

```
Residuals:
   Min     1Q Median     3Q    Max
 -7870  -2250  -1144    724 313951

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      3775.57      75.47   50.02   <2e-16 ***
HasSwimmingPool  4773.49     126.53   37.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6968 on 13230 degrees of freedom
Multiple R-squared:  0.09713,  Adjusted R-squared:  0.09706
F-statistic:  1423 on 1 and 13230 DF,  p-value: < 2.2e-16
```

d) Room Rent against all variables

```
Residuals:
   Min     1Q Median     3Q    Max
-10804  -2295   -946   1002 310110

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -6896.154    340.549  -20.25   <2e-16 ***
StarRating       3597.322    111.670   32.21   <2e-16 ***
HotelCapacity     -15.558      1.006  -15.47   <2e-16 ***
HasSwimmingPool  2528.885    157.894   16.02   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6710 on 13228 degrees of freedom
Multiple R-squared:  0.1628,   Adjusted R-squared:  0.1626
F-statistic: 857.5 on 3 and 13228 DF,  p-value: < 2.2e-16
```