Code:

```
All<-read.csv(paste("All.csv",sep=""))
library(psych)
View(describe(All))
```

Output:

	vars	n	mean	sd	median	trimmed	mad
CityName*	1	13232	1.806651e+01	1.171740e+01	16	1.729152e+01	
Population	2	13232	4.416837e+06	4.258386e+06	3046163	4.040816e+06	38464
CityRank	3	13232	1.483374e+01	1.351246e+01	9	1.330266e+01	
IsMetroCity	4	13232	2.841596e-01	4.510303e-01	0	2.302097e-01	
IsTouristDestination	5	13232	6.971735e-01	4.594982e-01	1	7.464576e-01	
IsWeekend	6	13232	6.228083e-01	4.847018e-01	1	6.535046e-01	
IsNewYearEve	7	13232	1.243954e-01	3.300446e-01	0	3.051200e-02	
Date*	8	13232	1.429807e+01	2.693757e+00	14	1.438891e+01	
HotelName*	9	13232	8.418374e+02	4.881378e+02	834	8.420505e+02	6
RoomRent	10	13232	5.473992e+03	7.333117e+03	4000	4.383334e+03	26
StarRating	11	13232	3.458933e+00	7.562325e-01	3	3.401152e+00	
Airport	12	13232	2.115874e+01	2.275991e+01	15	1.639458e+01	

```
#Find top 3 independent variables for Agra
agra <- All[All$CityName=="Agra",]
```

Code:

```
#Hypothesis 1: The roomrent is independent of the weekend
mytable <- xtabs(~IsWeekend+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 91.377, df = 145, p-value = 0.9998
```

Code:

```
#Hypothesis 2: The roomrent is independent of new years eve
mytable <- xtabs(~IsNewYearEve+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 128.88, df = 145, p-value = 0.8276
```

Code:

```
#Hypothesis 3: The roomrent is independent of the star rating
mytable <- xtabs(~StarRating+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 2589.4, df = 1015, p-value < 2.2e-16
```

Code:

```
#Hypothesis 4: The roomrent is independent of the distance from the
 airport
mytable <- xtabs(~Airport+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 10881, df = 4205, p-value < 2.2e-16
```

Code:

```
#Hypothesis 5: The roomrent is independent of the service free wifi
mytable <- xtabs(~FreeWifi+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 299.95, df = 145, p-value = 7.42e-13
```

Code:

```
#Hypothesis 6: The roomrent is independent of the service free
 breakfast
mytable <- xtabs(~FreeBreakfast+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 365.4, df = 145, p-value < 2.2e-16
```

Code:

```
#Hypothesis 7: The roomrent is independent of the hotel capacity
mytable <- xtabs(~HotelCapacity+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 12673, df = 5075, p-value < 2.2e-16
```

Code:

```
#Hypothesis 8: The roomrent is independent of the swimming pool
 facility
mytable <- xtabs(~HasSwimmingPool+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 408.71, df = 145, p-value < 2.2e-16
```

Code:

```
#Hypothesis 9: The roomrent is independent of the date
All$Year <- substr(c(All$Date),nchar(c(All$Date))-
 2,nchar(c(All$Date)))
mylist <- list()
for(date in All$Date)
{
  mylist[length(mylist)+1] <- substr(date,nchar(date)-1,nchar(date))
}

All$Year <- as.numeric(mylist)
mylist <- NULL

mytable <- xtabs(~Year+RoomRent,data=agra)
addmargins(mytable)
chisq.test(mytable)
```

Output:

```
> chisq.test(mytable)

        Pearson's Chi-squared test

data:  mytable
X-squared = 408.71, df = 145, p-value < 2.2e-16
```

Code:

```
#From above results, the top 3 independent factors are IsWeekend,
 IsNewYear and Year
#Visualize these 3 with respect to RoomRent
table(All$IsWeekend)
table(All$IsNewYearEve)
table(All$Year
```

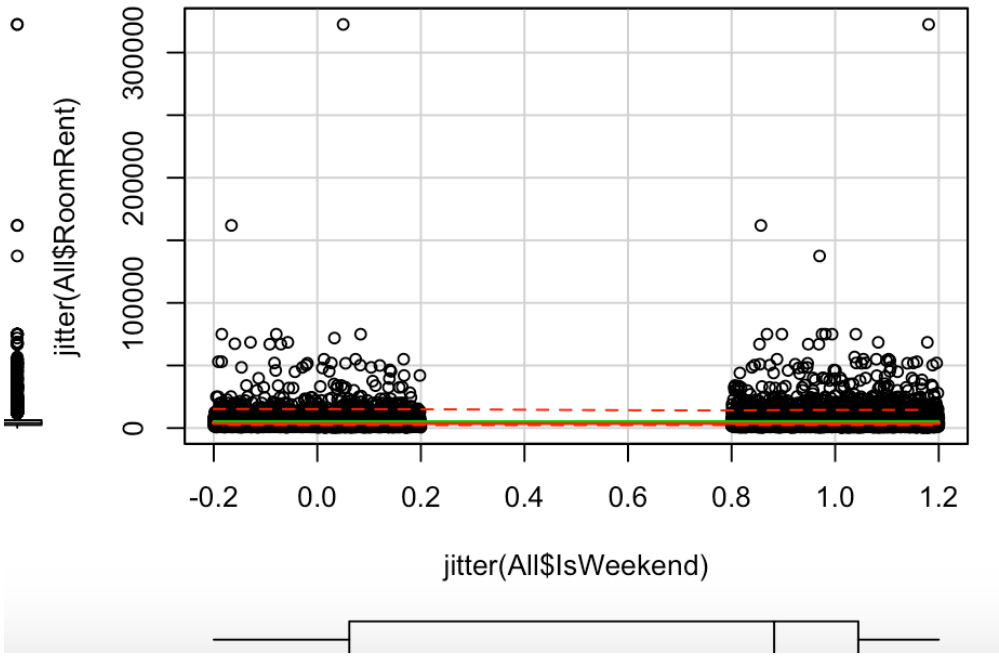
Output:

```
> table(All$IsWeekend)

   0    1
4991 8241
> table(All$IsNewYearEve)

    0     1
11586  1646
> table(All$Year)

  16   17
9989 3243
```

Code:

```
library(car)
scatterplot(jitter(All$IsWeekend), jitter(All$RoomRent))
```

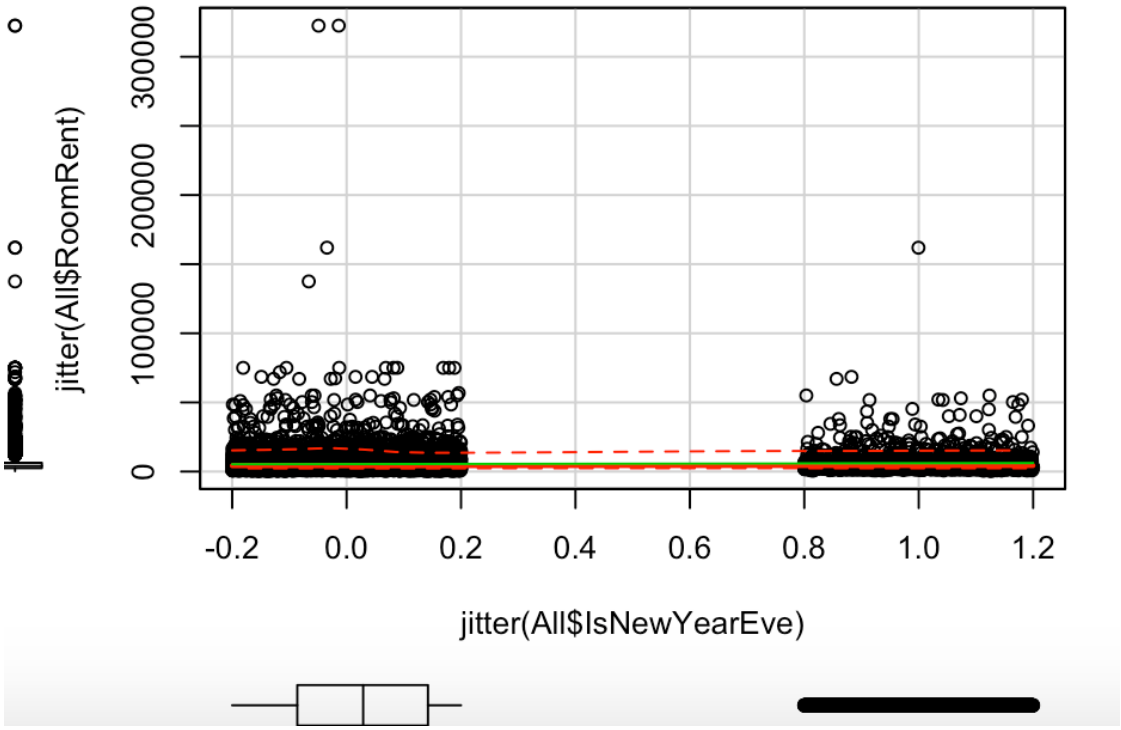
Output:



Code:

```
scatterplot(jitter(All$IsNewYearEve), jitter(All$RoomRent))
```

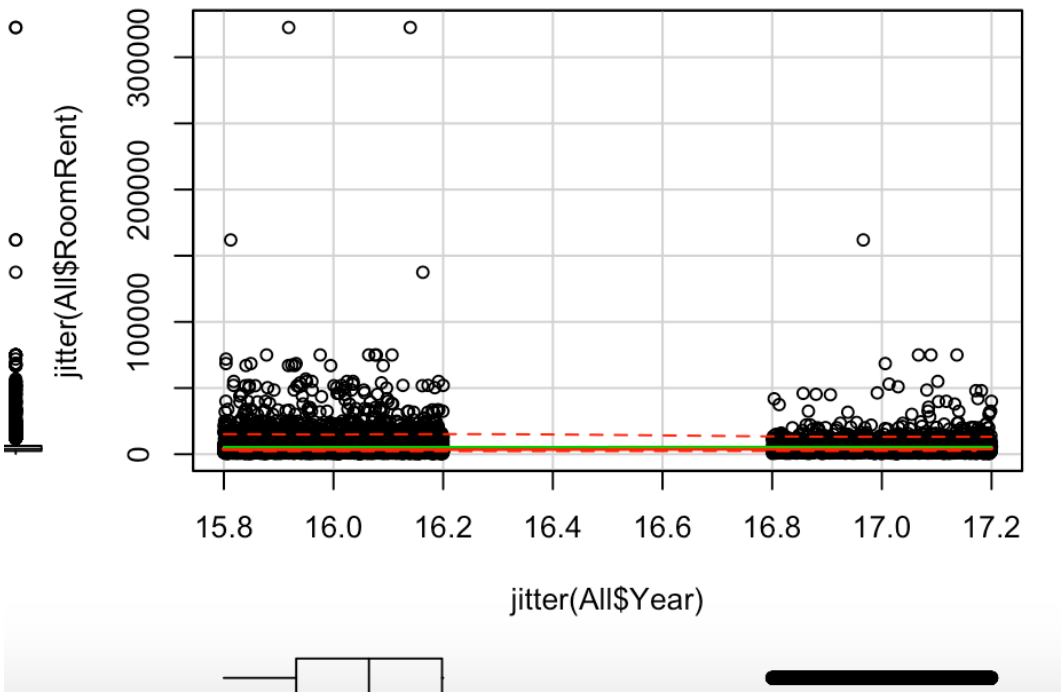
Output:



Code:

```
scatterplot(jitter(All$Year), jitter(All$RoomRent))
```

Output:



Code:

```
#Create a dataframe with only the 4 required variables
extra <- data.frame(All)
extra <- extra[,c(6,7,10,20)]
#extra <- extra[,c(1,2,4,3)]
View(extra)
```

Output:

	IsWeekend	IsNewYearEve	RoomRent	Year
1	1	0	2425	16
2	0	0	2425	16
3	1	0	2425	16
4	1	0	2425	16
5	0	0	2425	16
6	1	1	2425	16
7	0	0	2852	17
8	1	0	2852	17
9	1	0	3599	16
10	0	0	3329	16
11	1	0	6299	16
12	1	0	5399	16

Code:

```
#Corrgram
library(corrgram)
corrgram(extra,upper.panel=panel.cor)
```

Output:

Code:

```
#Compute Variance and Covariance matrices
var(extra)
cov(extra)
```

Output:

```
> var(extra)
               IsWeekend IsNewYearEve     RoomRent        Year
IsWeekend     0.23493587   0.04677330 1.627952e+01 -0.03444694
IsNewYearEve  0.04677330   0.10892942 9.315860e+01 -0.03049008
RoomRent     16.27951503  93.15859986 5.377460e+07 11.56990926
Year         -0.03444694  -0.03049008 1.156991e+01  0.18503369
> cov(extra)
               IsWeekend IsNewYearEve     RoomRent        Year
IsWeekend     0.23493587   0.04677330 1.627952e+01 -0.03444694
IsNewYearEve  0.04677330   0.10892942 9.315860e+01 -0.03049008
RoomRent     16.27951503  93.15859986 5.377460e+07 11.56990926
Year         -0.03444694  -0.03049008 1.156991e+01  0.18503369
```

Code:

```
#Hypothesis 1: Room Rents are equal on weekends and weekdays
t.test(All$RoomRent~All$IsWeekend)
```

Output:

```
> t.test(All$RoomRent~All$IsWeekend)

        Welch Two Sample t-test

data:  All$RoomRent by All$IsWeekend
t = -0.51853, df = 9999.4, p-value = 0.6041
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -331.2427  192.6559
sample estimates:
mean in group 0 mean in group 1
       5430.835        5500.129
```

Code:

```
#Hypothesis 2: Room Rents are same on new years and other days of
 the year
```

```
t.test(All$RoomRent~All$IsNewYearEve)
```

Output:

```
> t.test(All$RoomRent~All$IsNewYearEve)

        Welch Two Sample t-test

data:  All$RoomRent by All$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
```

Code:

```
#Hypothesis 3: Room Rents are same in 2016 and 2017
t.test(All$RoomRent~All$Year)
```

Output:

```
> t.test(All$RoomRent~All$Year)

        Welch Two Sample t-test

data:  All$RoomRent by All$Year
t = -0.48142, df = 7090.6, p-value = 0.6302
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -317.1399  192.0826
sample estimates:
mean in group 16 mean in group 17
        5458.667         5521.195
```

Code:

```
price_predict <-
 lm(RoomRent~Population+CityRank+IsMetroCity+IsTouristDestination+Is
 Weekend

 +IsNewYearEve+StarRating+Airport+FreeWifi+FreeBreakfast+HotelCapaci
 ty
                     +HasSwimmingPool+Year, data=All)

coefficients(price_predict)
```

Output:

```
> price_predict <- lm(RoomRent~Population+CityRank+IsMetroCity+IsTouristDestina
tion+IsWeekend
+
+IsNewYearEve+StarRating+Airport+FreeWifi+FreeBreakfast+HotelCapacity
+                     +HasSwimmingPool+Year, data=All)
>
> coefficients(price_predict)
         (Intercept)           Population             CityRank
       -1.171606e+04        -1.181016e-04         1.811204e+00
         IsMetroCity IsTouristDestination            IsWeekend
       -6.701936e+02         1.924041e+03        -7.192006e+01
        IsNewYearEve           StarRating              Airport
        9.278289e+02         3.592449e+03         9.539383e+00
            FreeWifi        FreeBreakfast        HotelCapacity
        5.501057e+02         1.690076e+02        -1.027884e+01
     HasSwimmingPool                 Year
        2.153465e+03         1.903758e+02
>
```

Code:

```
summary(price_predict)
```

Output:

```
> summary(price_predict)

Call:
lm(formula = RoomRent ~ Population + CityRank + IsMetroCity +
    IsTouristDestination + IsWeekend + IsNewYearEve + StarRating +
    Airport + FreeWifi + FreeBreakfast + HotelCapacity + HasSwimmingPool +
    Year, data = All)

Residuals:
   Min     1Q Median     3Q    Max
-11844  -2358   -682   1030 309734

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -1.172e+04  2.292e+03  -5.113 3.22e-07 ***
Population          -1.181e-04  3.592e-05  -3.288  0.00101 **
CityRank             1.811e+00  1.035e+01   0.175  0.86103
IsMetroCity         -6.702e+02  2.165e+02  -3.096  0.00197 **
IsTouristDestination  1.924e+03  1.481e+02  12.993  < 2e-16 ***
IsWeekend           -7.192e+01  1.246e+02  -0.577  0.56380
IsNewYearEve         9.278e+02  1.847e+02   5.023 5.15e-07 ***
StarRating           3.592e+03  1.108e+02  32.437  < 2e-16 ***
Airport              9.539e+00  3.171e+00   3.009  0.00263 **
FreeWifi             5.501e+02  2.242e+02   2.454  0.01416 *
FreeBreakfast        1.690e+02  1.233e+02   1.370  0.17063
HotelCapacity       -1.028e+01  1.033e+00  -9.949  < 2e-16 ***
HasSwimmingPool      2.153e+03  1.616e+02  13.329  < 2e-16 ***
Year                 1.904e+02  1.375e+02   1.385  0.16610
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6600 on 13218 degrees of freedom
Multiple R-squared:  0.1907,    Adjusted R-squared:  0.1899
F-statistic: 239.6 on 13 and 13218 DF,  p-value: < 2.2e-16
```