# Analysis of Airbnb Prices in Amsterdam - R Appendix

Eric Cheng, Nina Sun, Jeanny Zhang

## R Code Appendix

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggResidpanel)
library(broom)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(knitr)
```

```
# import dataset
airbnb <- read.csv("amsterdam_weekends.csv")
airbnb_mutate <- airbnb %>% mutate(case = row_number())

# EDA
summary(airbnb)
```

```
##        X          realSum        room_type         room_shared
##  Min.   :  0   Min.   : 165.9   Length:977         Length:977
```

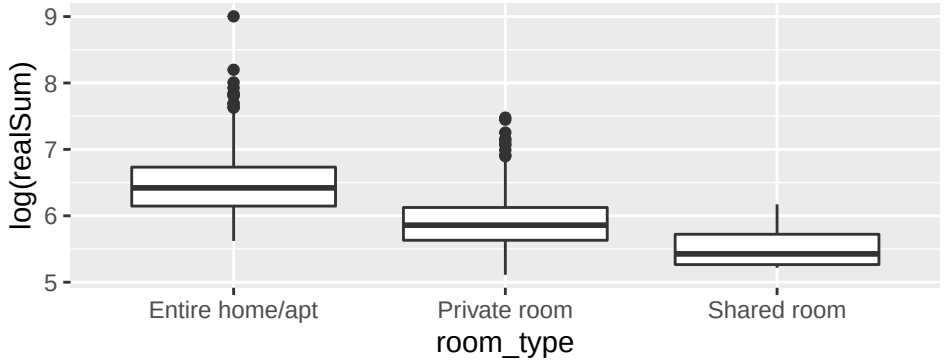
```
##  1st Qu.:244   1st Qu.: 362.8   Class :character   Class :character
##  Median :488   Median : 491.6   Mode  :character   Mode  :character
##  Mean   :488   Mean   : 604.8
##  3rd Qu.:732   3rd Qu.: 716.6
##  Max.   :976   Max.   :8130.7
##  room_private       person_capacity host_is_superhost      multi
##  Length:977         Min.   :2.00    Length:977         Min.   :0.0000
##  Class :character   1st Qu.:2.00    Class :character   1st Qu.:0.0000
##  Mode  :character   Median :2.00    Mode  :character   Median :0.0000
##                     Mean   :2.77                       Mean   :0.2549
##                     3rd Qu.:4.00                       3rd Qu.:1.0000
##                     Max.   :6.00                       Max.   :1.0000
##       biz          cleanliness_rating guest_satisfaction_overall
##  Min.   :0.00000   Min.   : 2.000     Min.   : 20.00
##  1st Qu.:0.00000   1st Qu.: 9.000     1st Qu.: 93.00
##  Median :0.00000   Median :10.000     Median : 96.00
##  Mean   :0.09417   Mean   : 9.471     Mean   : 94.69
##  3rd Qu.:0.00000   3rd Qu.:10.000     3rd Qu.: 99.00
##  Max.   :1.00000   Max.   :10.000     Max.   :100.00
##     bedrooms          dist            metro_dist        attr_index
##  Min.   :0.000   Min.   : 0.01504   Min.   :0.03652   Min.   :  40.89
##  1st Qu.:1.000   1st Qu.: 1.41004   1st Qu.:0.46714   1st Qu.: 132.72
##  Median :1.000   Median : 2.31455   Median :0.87617   Median : 208.90
##  Mean   :1.303   Mean   : 2.80634   Mean   :1.08929   Mean   : 266.87
##  3rd Qu.:2.000   3rd Qu.: 3.61017   3rd Qu.:1.50086   3rd Qu.: 366.69
##  Max.   :5.000   Max.   :11.19593   Max.   :4.41190   Max.   :1888.74
##  attr_index_norm     rest_index      rest_index_norm        lng
##  Min.   :  2.165   Min.   :  50.82   Min.   :  4.431   Min.   :4.776
##  1st Qu.:  7.027   1st Qu.: 170.68   1st Qu.: 14.879   1st Qu.:4.869
##  Median : 11.061   Median : 266.96   Median : 23.272   Median :4.888
##  Mean   : 14.129   Mean   : 329.13   Mean   : 28.692   Mean   :4.890
##  3rd Qu.: 19.414   3rd Qu.: 438.11   3rd Qu.: 38.193   3rd Qu.:4.906
##  Max.   :100.000   Max.   :1147.10   Max.   :100.000   Max.   :5.011
##       lat
##  Min.   :52.29
##  1st Qu.:52.35
##  Median :52.37
##  Mean   :52.36
##  3rd Qu.:52.38
##  Max.   :52.42
```

```
# fit a rich model
airbnb_lm <- lm(realSum ~ room_type+room_shared+room_private+person_capacity+host_is_superhost+multi+bi
summary(airbnb_lm)
```

```
##
## Call:
## lm(formula = realSum ~ room_type + room_shared + room_private +
##     person_capacity + host_is_superhost + multi + biz + cleanliness_rating +
##     guest_satisfaction_overall + bedrooms + dist + metro_dist +
##     attr_index + attr_index_norm + rest_index + rest_index_norm +
##     lng + lat, data = airbnb)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -745.6 -138.3  -35.5   76.2 6772.0
##
## Coefficients: (4 not defined because of singularities)
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -6.109e+04  4.857e+04  -1.258   0.2088
## room_typePrivate room       -1.626e+02  2.561e+01  -6.349 3.35e-10 ***
## room_typeShared room        -3.139e+02  1.707e+02  -1.839   0.0663 .
## room_sharedTrue                     NA         NA      NA       NA
## room_privateTrue                    NA         NA      NA       NA
## person_capacity              1.275e+02  1.449e+01   8.799  < 2e-16 ***
## host_is_superhostTrue       -9.423e-01  2.559e+01  -0.037   0.9706
## multi                        2.155e+01  2.653e+01   0.812   0.4167
## biz                         -3.615e+01  3.920e+01  -0.922   0.3567
## cleanliness_rating           1.263e+01  1.780e+01   0.710   0.4780
## guest_satisfaction_overall   1.942e+00  2.259e+00   0.860   0.3903
## bedrooms                     1.589e+02  2.013e+01   7.893 8.03e-15 ***
## dist                        -1.654e+01  1.133e+01  -1.459   0.1448
## metro_dist                  -8.253e+00  1.797e+01  -0.459   0.6461
## attr_index                   1.806e-01  1.072e-01   1.684   0.0925 .
## attr_index_norm                     NA         NA      NA       NA
## rest_index                   2.279e-01  1.111e-01   2.052   0.0404 *
## rest_index_norm                     NA         NA      NA       NA
## lng                          4.348e+02  2.928e+02   1.485   0.1379
## lat                          1.121e+03  9.230e+02   1.214   0.2249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 335.2 on 961 degrees of freedom
## Multiple R-squared:  0.438,  Adjusted R-squared:  0.4292
## F-statistic: 49.93 on 15 and 961 DF,  p-value: < 2.2e-16
```

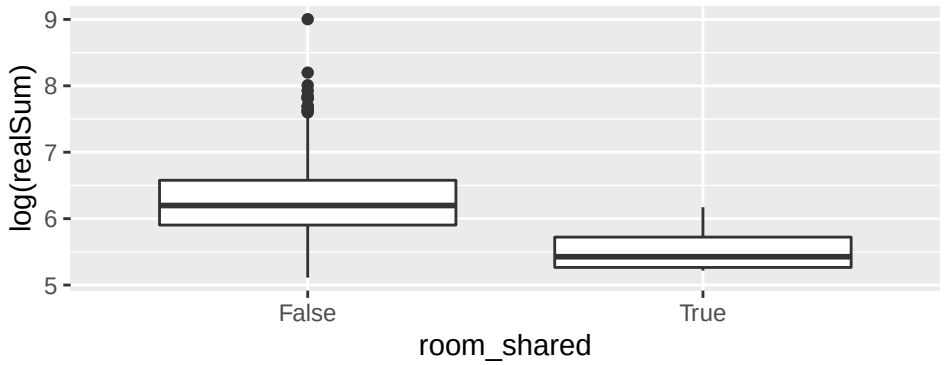
```
# check model transformation
# room_type vs. realSum
ggplot(airbnb, aes(x=room_type, y=log(realSum))) + geom_boxplot()
```




```
table(airbnb$room_type)
```

```
##
## Entire home/apt    Private room     Shared room
##             588             385               4
```

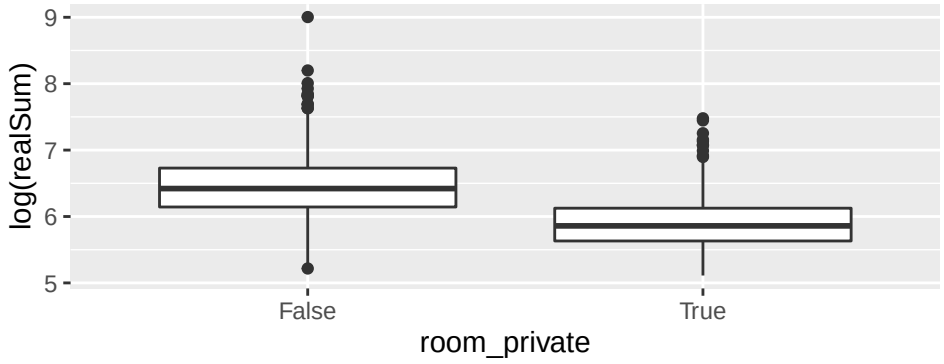
```
# room_shared vs. realSum
ggplot(airbnb, aes(x=room_shared, y=log(realSum))) + geom_boxplot()
```




```
table(airbnb$room_shared)
```

```
##
## False  True
##   973     4
```

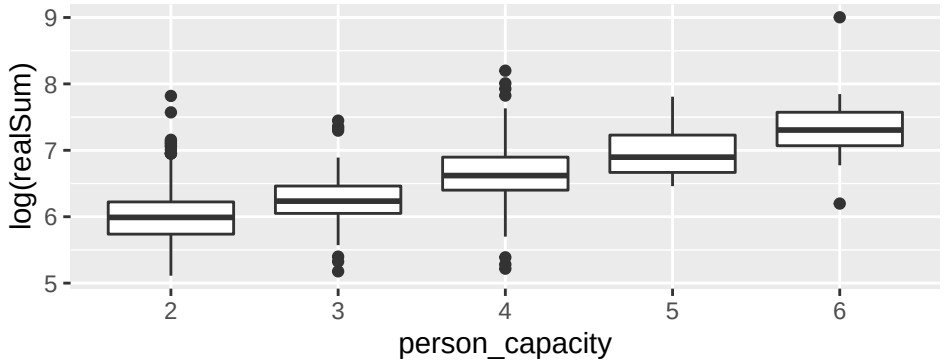
```
# room_private vs. realSum
ggplot(airbnb, aes(x=room_private, y=log(realSum))) + geom_boxplot()
```




```
table(airbnb$room_private)
```

```
##
## False  True
##   592   385
```

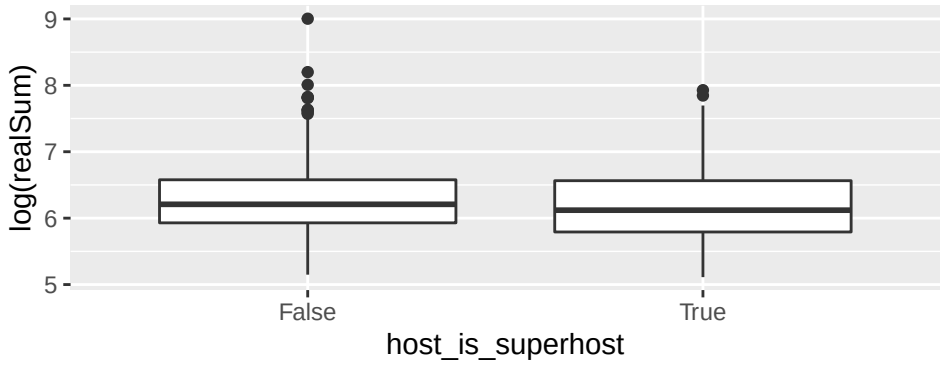
```
# person_capacity vs. realSum
airbnb$person_capacity <- as.factor(airbnb$person_capacity)
ggplot(airbnb, aes(x=person_capacity, y=log(realSum))) + geom_boxplot()
```




```
table(airbnb$person_capacity)
```

```
##
##   2   3   4   5   6
## 591  60 301  10  15
```

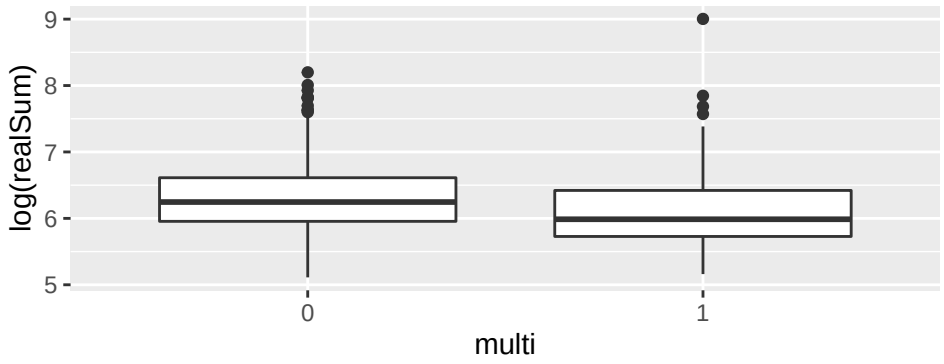
```
# host_is_superhost vs. realSum
ggplot(airbnb, aes(x=host_is_superhost, y=log(realSum))) + geom_boxplot()
```



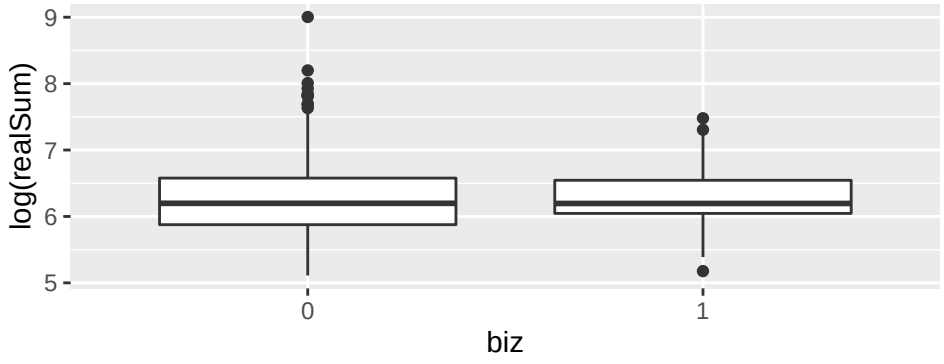

```
table(airbnb$host_is_superhost)
```

```
##
## False  True
##   709   268
```

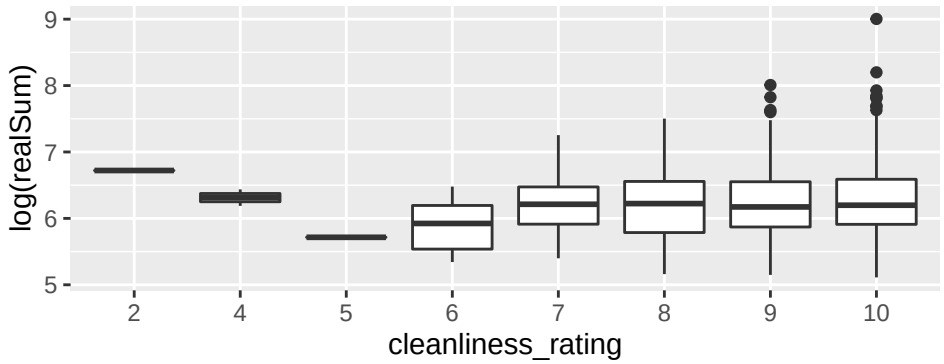
```
# multi vs. realSum
airbnb$multi <- as.factor(airbnb$multi)
ggplot(airbnb, aes(x=multi, y=log(realSum))) + geom_boxplot()
```




```
# biz vs. realSum
airbnb$biz <- as.factor(airbnb$biz)
ggplot(airbnb, aes(x=biz, y=log(realSum))) + geom_boxplot()
```



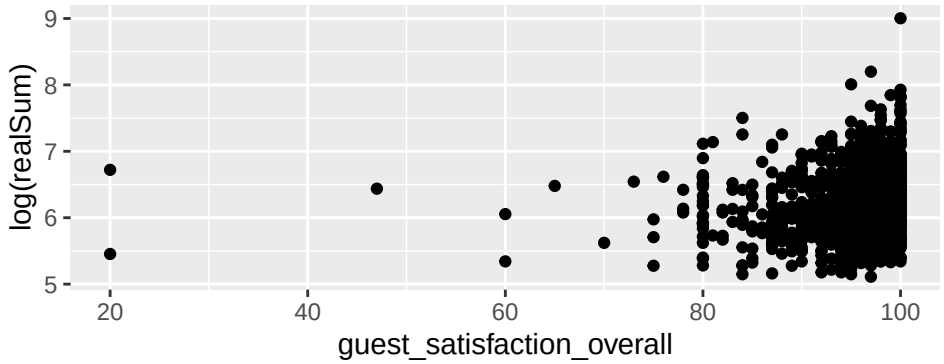
```
# cleanliness_rating vs. realSum
airbnb$cleanliness_rating <- as.factor(airbnb$cleanliness_rating)
ggplot(airbnb, aes(x=cleanliness_rating, y=log(realSum))) + geom_boxplot()
```



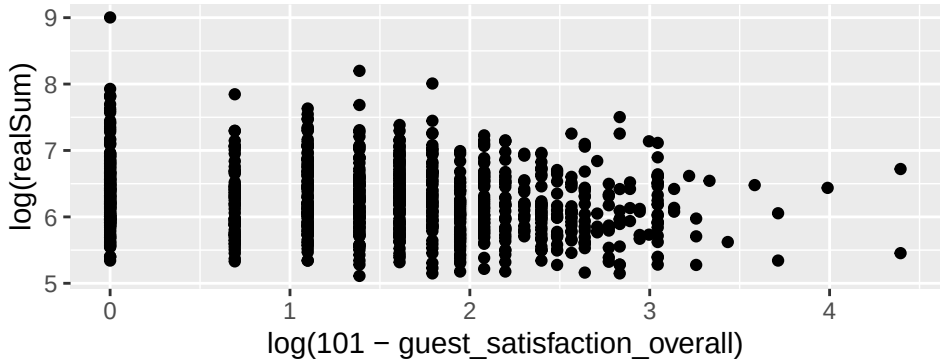

```
table(airbnb$cleanliness_rating)
```

```
##
##   2   4   5   6   7   8   9  10
##   1   2   1   7   8  81 278 599
```

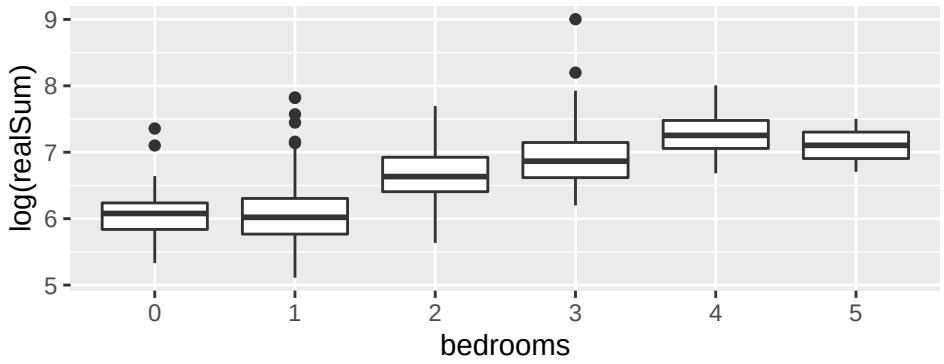
```
# guest_satisfaction_overall vs. realSum
ggplot(airbnb, aes(x=guest_satisfaction_overall, y=log(realSum))) + geom_point()
```




```
ggplot(airbnb, aes(x=log(101-guest_satisfaction_overall), y=log(realSum))) + geom_point()
```



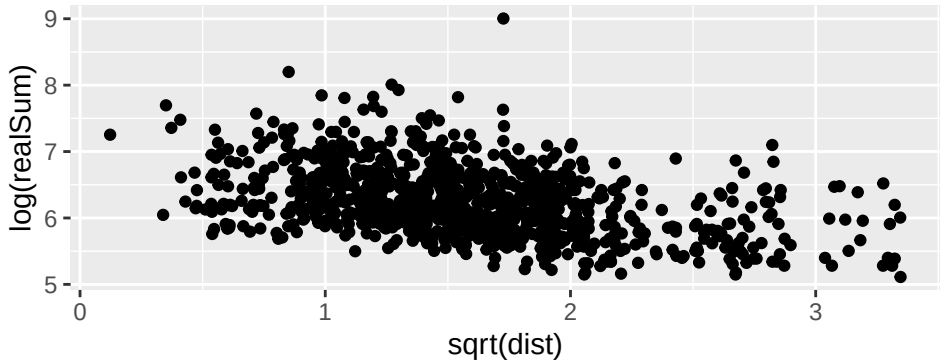

```
# bedrooms vs. realSum
airbnb$bedrooms <- as.factor(airbnb$bedrooms)
ggplot(airbnb, aes(x=bedrooms, y=log(realSum))) + geom_boxplot()
```



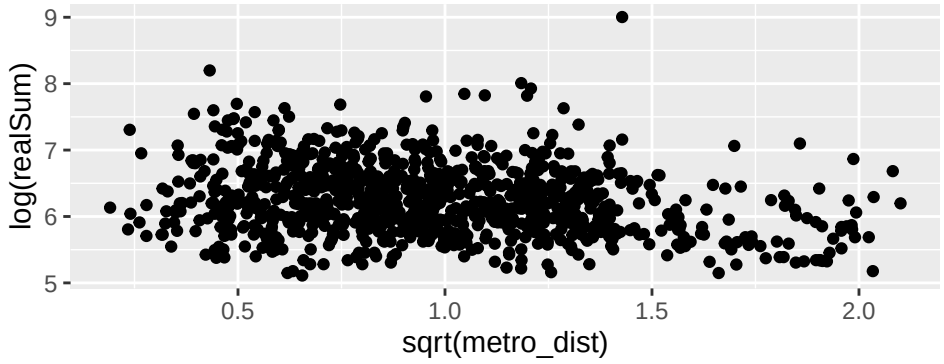

```
table(airbnb$bedrooms)
```

```
##
##   0   1   2   3   4   5
##  63 636 210  57   9   2
```

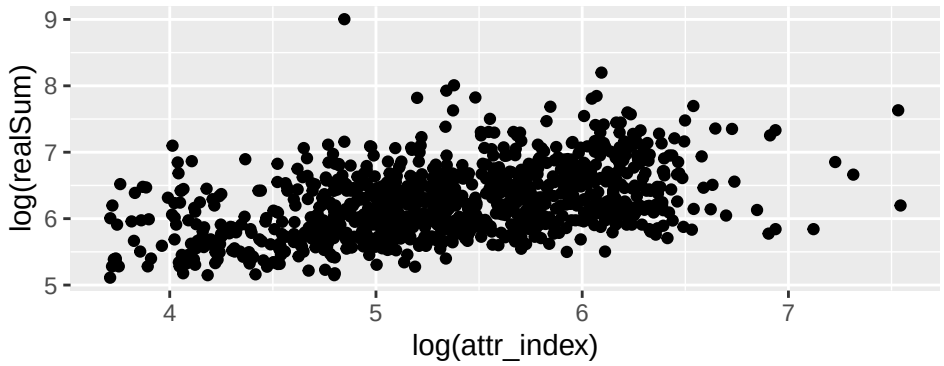
```
# dist vs. realSum
ggplot(airbnb, aes(x=sqrt(dist), y=log(realSum))) + geom_point()
```



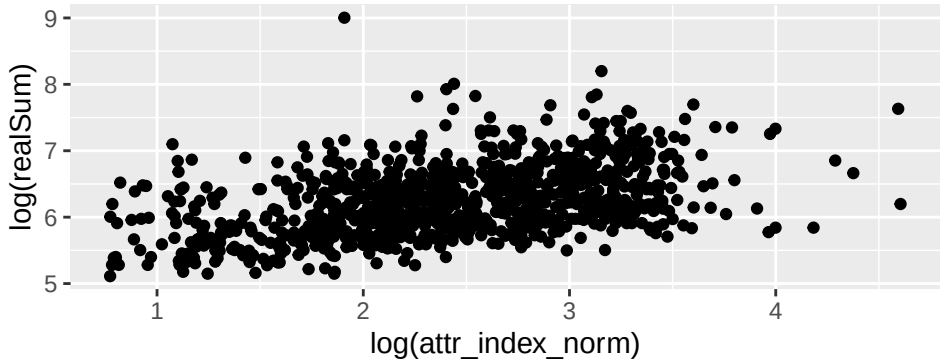

```
# metro_dist vs. realSum
ggplot(airbnb, aes(x=sqrt(metro_dist), y=log(realSum))) + geom_point()
```



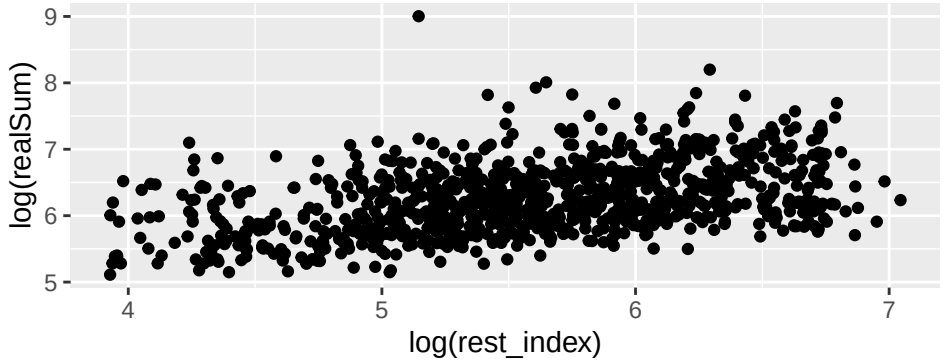

```
# attr_index vs. realSum
ggplot(airbnb, aes(x=log(attr_index), y=log(realSum))) + geom_point()
```



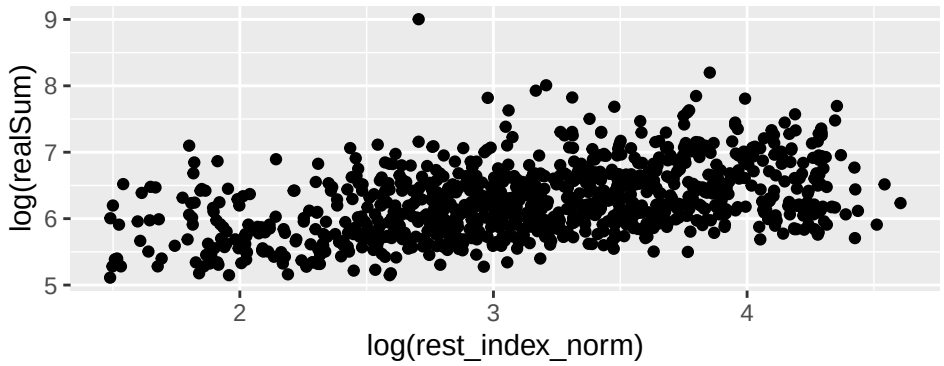

```
# attr_index_norm vs. realSum
ggplot(airbnb, aes(x=log(attr_index_norm), y=log(realSum))) + geom_point()
```



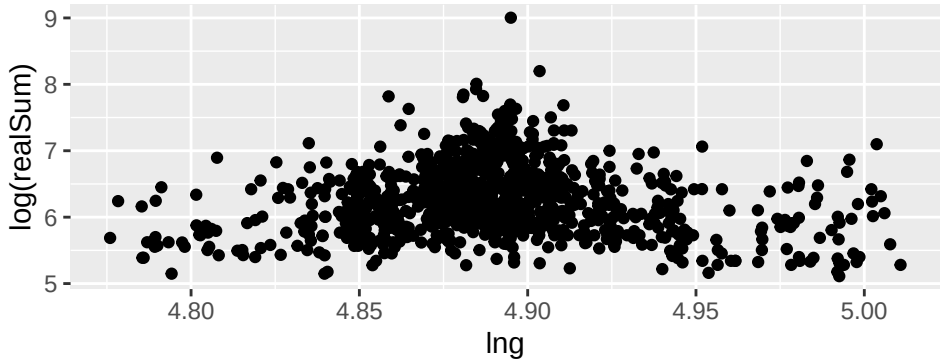

```
# rest_index vs. realSum
ggplot(airbnb, aes(x=log(rest_index), y=log(realSum))) + geom_point()
```



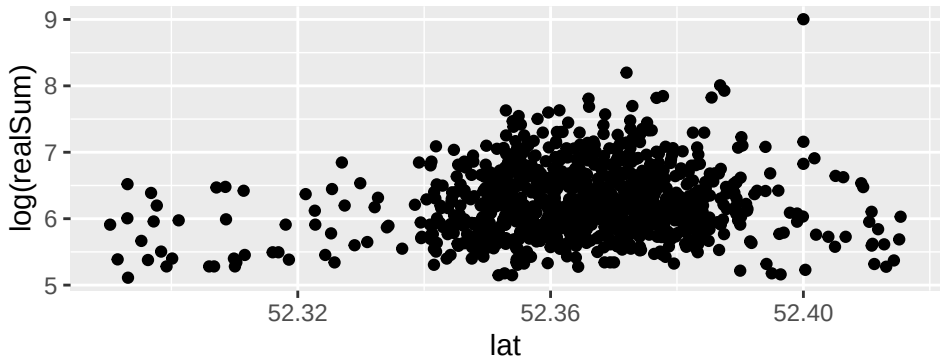

```
# rest_index_norm vs. realSum
ggplot(airbnb, aes(x=log(rest_index_norm), y=log(realSum))) + geom_point()
```



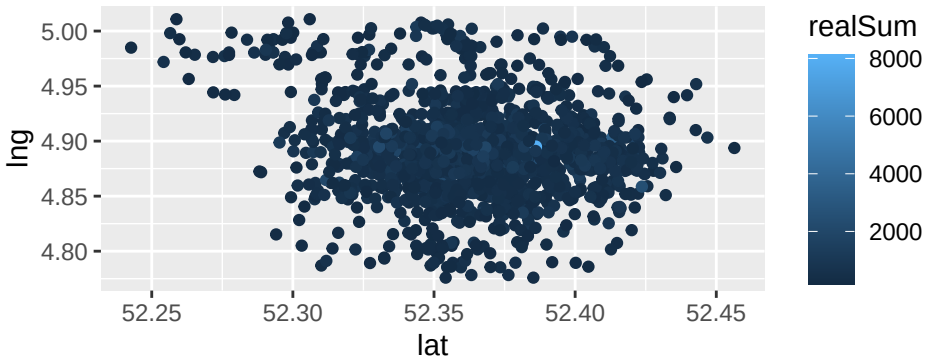

```
# lng vs. realSum
ggplot(airbnb, aes(x=lng, y=log(realSum))) + geom_point()
```




```
# lat vs. realSum
ggplot(airbnb, aes(x=lat, y=log(realSum))) + geom_point()
```




```
# lng vs. lat
ggplot(airbnb, aes(x=lat, y=lng, color = realSum)) + geom_point() + geom_jitter(width = 0.05)
```




```
# transformed model
airbnb_lm_transformed <- lm(log(realSum) ~ room_type+room_shared+room_private+person_capacity+host_is_su
summary(airbnb_lm_transformed)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + room_shared + room_private +
##     person_capacity + host_is_superhost + multi + biz + cleanliness_rating +
##     log(101 - guest_satisfaction_overall) + bedrooms + sqrt(dist) +
##     sqrt(metro_dist) + log(attr_index) + log(attr_index_norm) +
##     log(rest_index) + log(rest_index_norm), data = airbnb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96796 -0.18472 -0.02441  0.14165  1.72381
##
## Coefficients: (4 not defined because of singularities)
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          4.8137741  0.4835551   9.955  < 2e-16
## room_typePrivate room               -0.3250904  0.0222121 -14.636  < 2e-16
## room_typeShared room                -0.6211706  0.1483561  -4.187 3.09e-05
## room_sharedTrue                             NA         NA      NA       NA
## room_privateTrue                            NA         NA      NA       NA
## person_capacity3                     0.1933472  0.0411843   4.695 3.06e-06
## person_capacity4                     0.3266046  0.0311079  10.499  < 2e-16
## person_capacity5                     0.5413083  0.0972736   5.565 3.41e-08
## person_capacity6                     0.7711199  0.0852378   9.047  < 2e-16
## host_is_superhostTrue               -0.0006424  0.0222932  -0.029 0.977016
## multi1                              -0.0170297  0.0230957  -0.737 0.461088
## biz1                                 0.0061857  0.0340945   0.181 0.856071
## cleanliness_rating4                  0.2543658  0.3582216   0.710 0.477829
## cleanliness_rating5                 -0.4621271  0.4160099  -1.111 0.266911
## cleanliness_rating6                 -0.4156111  0.3128272  -1.329 0.184311
## cleanliness_rating7                 -0.3738294  0.3100169  -1.206 0.228181
## cleanliness_rating8                 -0.3189049  0.2954928  -1.079 0.280759
## cleanliness_rating9                 -0.3080339  0.2944443  -1.046 0.295756
## cleanliness_rating10                -0.2933455  0.2956053  -0.992 0.321277
## log(101 - guest_satisfaction_overall) -0.0394297  0.0122789  -3.211 0.001366
## bedrooms1                            0.0454986  0.0393075   1.158 0.247358
## bedrooms2                            0.2215028  0.0493335   4.490 8.00e-06
## bedrooms3                            0.4509334  0.0623491   7.232 9.76e-13
## bedrooms4                            0.7568799  0.1079190   7.013 4.41e-12
```

```
## bedrooms5                                0.7232943  0.2102449   3.440 0.000606
## sqrt(dist)                               0.0152180  0.0570692   0.267 0.789789
## sqrt(metro_dist)                         0.0059329  0.0297082   0.200 0.841752
## log(attr_index)                          0.0913190  0.0656785   1.390 0.164735
## log(attr_index_norm)                            NA         NA      NA       NA
## log(rest_index)                          0.2092932  0.0780708   2.681 0.007472
## log(rest_index_norm)                            NA         NA      NA       NA
##
## (Intercept)                             ***
## room_typePrivate room                   ***
## room_typeShared room                    ***
## room_sharedTrue
## room_privateTrue
## person_capacity3                        ***
## person_capacity4                        ***
## person_capacity5                        ***
## person_capacity6                        ***
## host_is_superhostTrue
## multi1
## biz1
## cleanliness_rating4
## cleanliness_rating5
## cleanliness_rating6
## cleanliness_rating7
## cleanliness_rating8
## cleanliness_rating9
## cleanliness_rating10
## log(101 - guest_satisfaction_overall) **
## bedrooms1
## bedrooms2                               ***
## bedrooms3                               ***
## bedrooms4                               ***
## bedrooms5                               ***
## sqrt(dist)
## sqrt(metro_dist)
## log(attr_index)
## log(attr_index_norm)
## log(rest_index)                         **
## log(rest_index_norm)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.289 on 950 degrees of freedom
## Multiple R-squared:  0.6997, Adjusted R-squared:  0.6915
## F-statistic: 85.14 on 26 and 950 DF,  p-value: < 2.2e-16
```

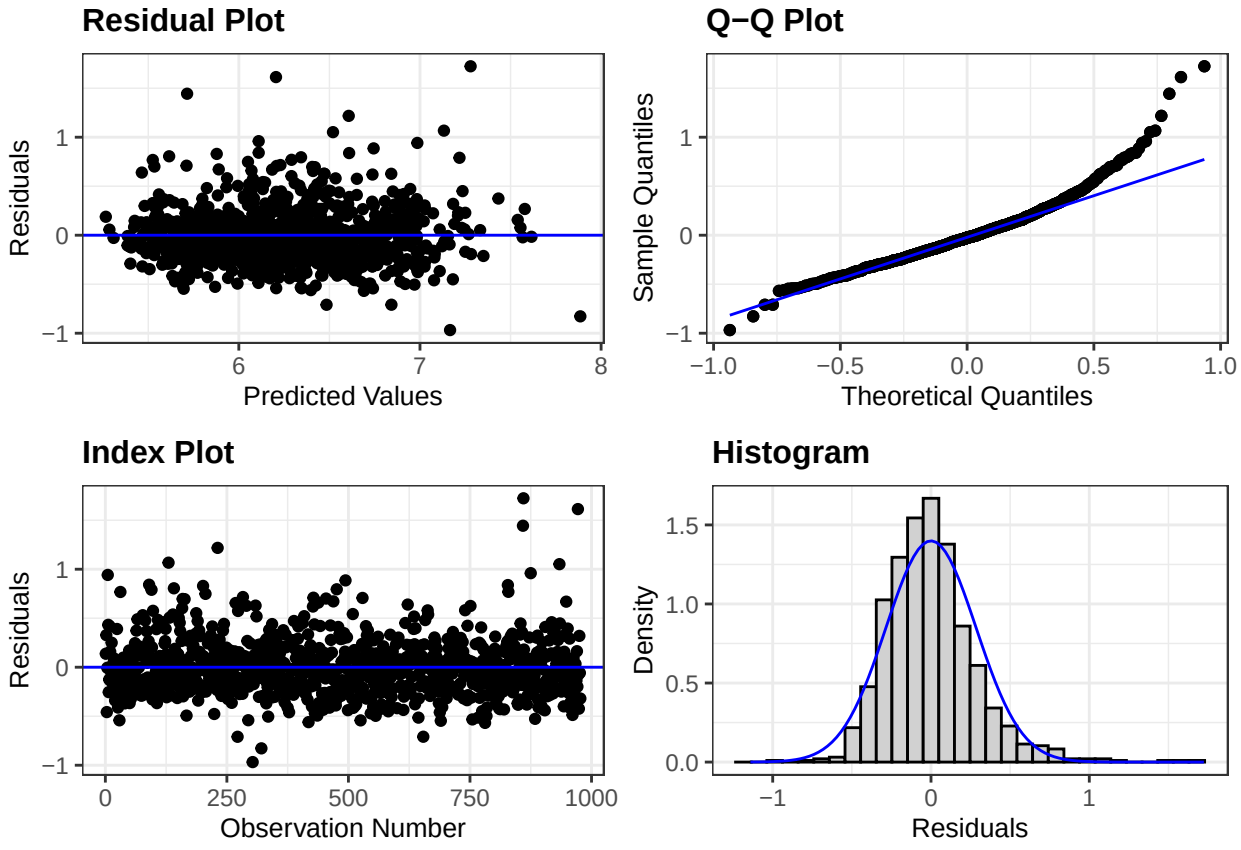
```
airbnb_lm_transformed <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+biz+cleanli
summary(airbnb_lm_transformed)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(dist) + sqrt(metro_dist) + log(attr_index) +
##     log(rest_index), data = airbnb)
```

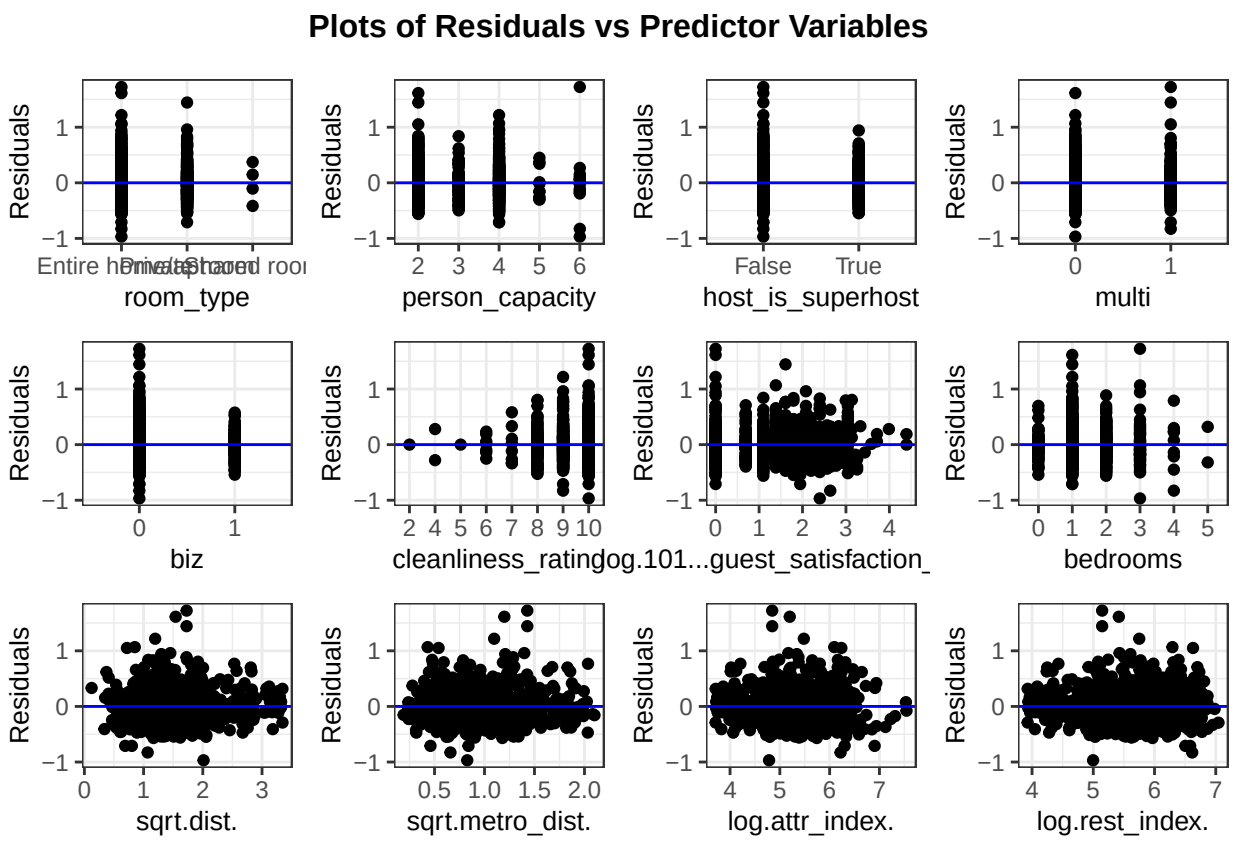
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96796 -0.18472 -0.02441  0.14165  1.72381
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         4.8137741  0.4835551   9.955  < 2e-16
## room_typePrivate room              -0.3250904  0.0222121 -14.636  < 2e-16
## room_typeShared room               -0.6211706  0.1483561  -4.187 3.09e-05
## person_capacity3                    0.1933472  0.0411843   4.695 3.06e-06
## person_capacity4                    0.3266046  0.0311079  10.499  < 2e-16
## person_capacity5                    0.5413083  0.0972736   5.565 3.41e-08
## person_capacity6                    0.7711199  0.0852378   9.047  < 2e-16
## host_is_superhostTrue              -0.0006424  0.0222932  -0.029 0.977016
## multi1                             -0.0170297  0.0230957  -0.737 0.461088
## biz1                                0.0061857  0.0340945   0.181 0.856071
## cleanliness_rating4                 0.2543658  0.3582216   0.710 0.477829
## cleanliness_rating5                -0.4621271  0.4160099  -1.111 0.266911
## cleanliness_rating6                -0.4156111  0.3128272  -1.329 0.184311
## cleanliness_rating7                -0.3738294  0.3100169  -1.206 0.228181
## cleanliness_rating8                -0.3189049  0.2954928  -1.079 0.280759
## cleanliness_rating9                -0.3080339  0.2944443  -1.046 0.295756
## cleanliness_rating10               -0.2933455  0.2956053  -0.992 0.321277
## log(101 - guest_satisfaction_overall) -0.0394297  0.0122789  -3.211 0.001366
## bedrooms1                           0.0454986  0.0393075   1.158 0.247358
## bedrooms2                           0.2215028  0.0493335   4.490 8.00e-06
## bedrooms3                           0.4509334  0.0623491   7.232 9.76e-13
## bedrooms4                           0.7568799  0.1079190   7.013 4.41e-12
## bedrooms5                           0.7232943  0.2102449   3.440 0.000606
## sqrt(dist)                          0.0152180  0.0570692   0.267 0.789789
## sqrt(metro_dist)                    0.0059329  0.0297082   0.200 0.841752
## log(attr_index)                     0.0913190  0.0656785   1.390 0.164735
## log(rest_index)                     0.2092932  0.0780708   2.681 0.007472
##
## (Intercept)                        ***
## room_typePrivate room              ***
## room_typeShared room               ***
## person_capacity3                   ***
## person_capacity4                   ***
## person_capacity5                   ***
## person_capacity6                   ***
## host_is_superhostTrue
## multi1
## biz1
## cleanliness_rating4
## cleanliness_rating5
## cleanliness_rating6
## cleanliness_rating7
## cleanliness_rating8
## cleanliness_rating9
## cleanliness_rating10
## log(101 - guest_satisfaction_overall) **
## bedrooms1
```

```
## bedrooms2                          ***
## bedrooms3                          ***
## bedrooms4                          ***
## bedrooms5                          ***
## sqrt(dist)
## sqrt(metro_dist)
## log(attr_index)
## log(rest_index)                    **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.289 on 950 degrees of freedom
## Multiple R-squared:  0.6997, Adjusted R-squared:  0.6915
## F-statistic: 85.14 on 26 and 950 DF,  p-value: < 2.2e-16
```

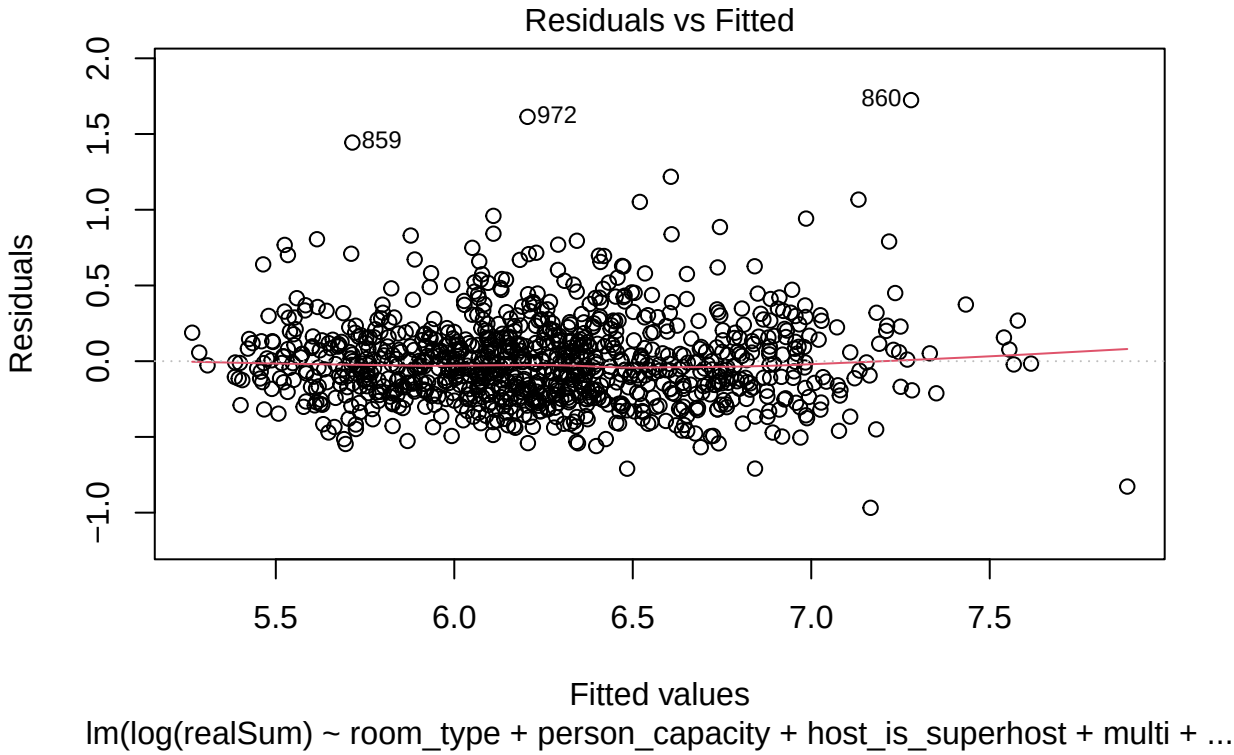
```
# check assumptions
resid_panel(airbnb_lm_transformed)
```




```
resid_xpanel(airbnb_lm_transformed)
```

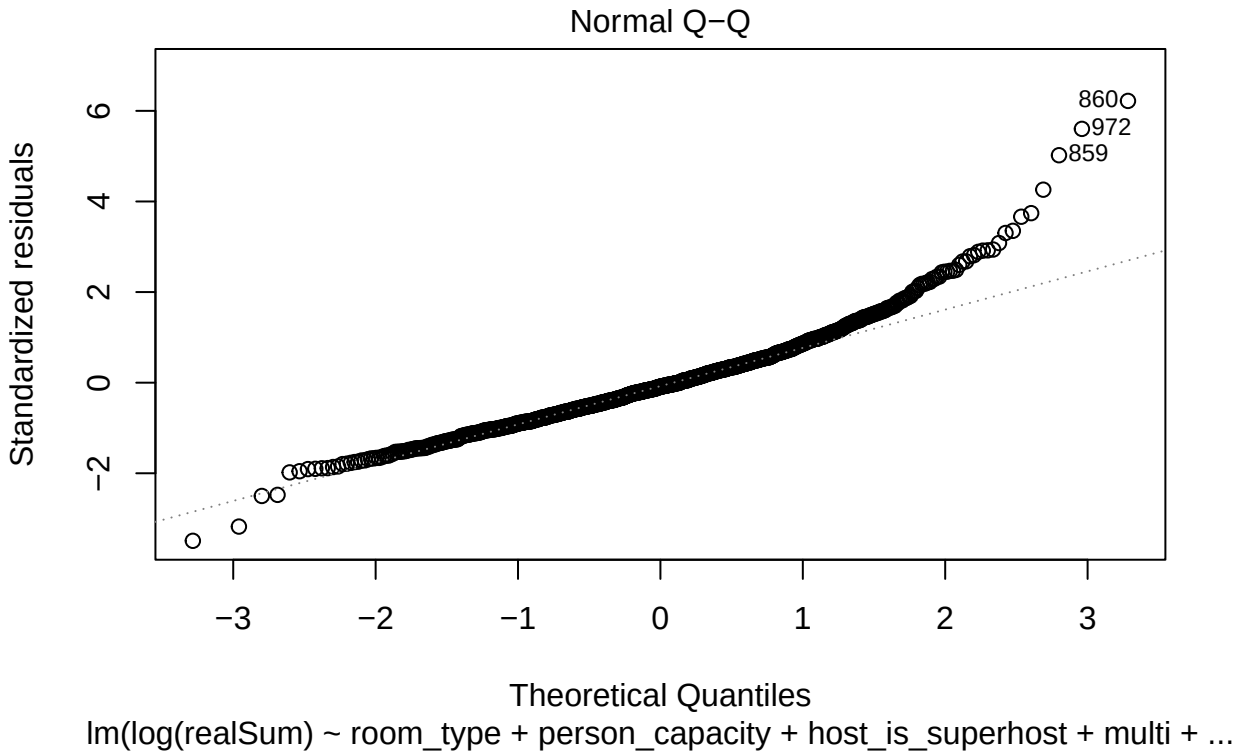



```
plot(airbnb_lm_transformed, which = 1)
```

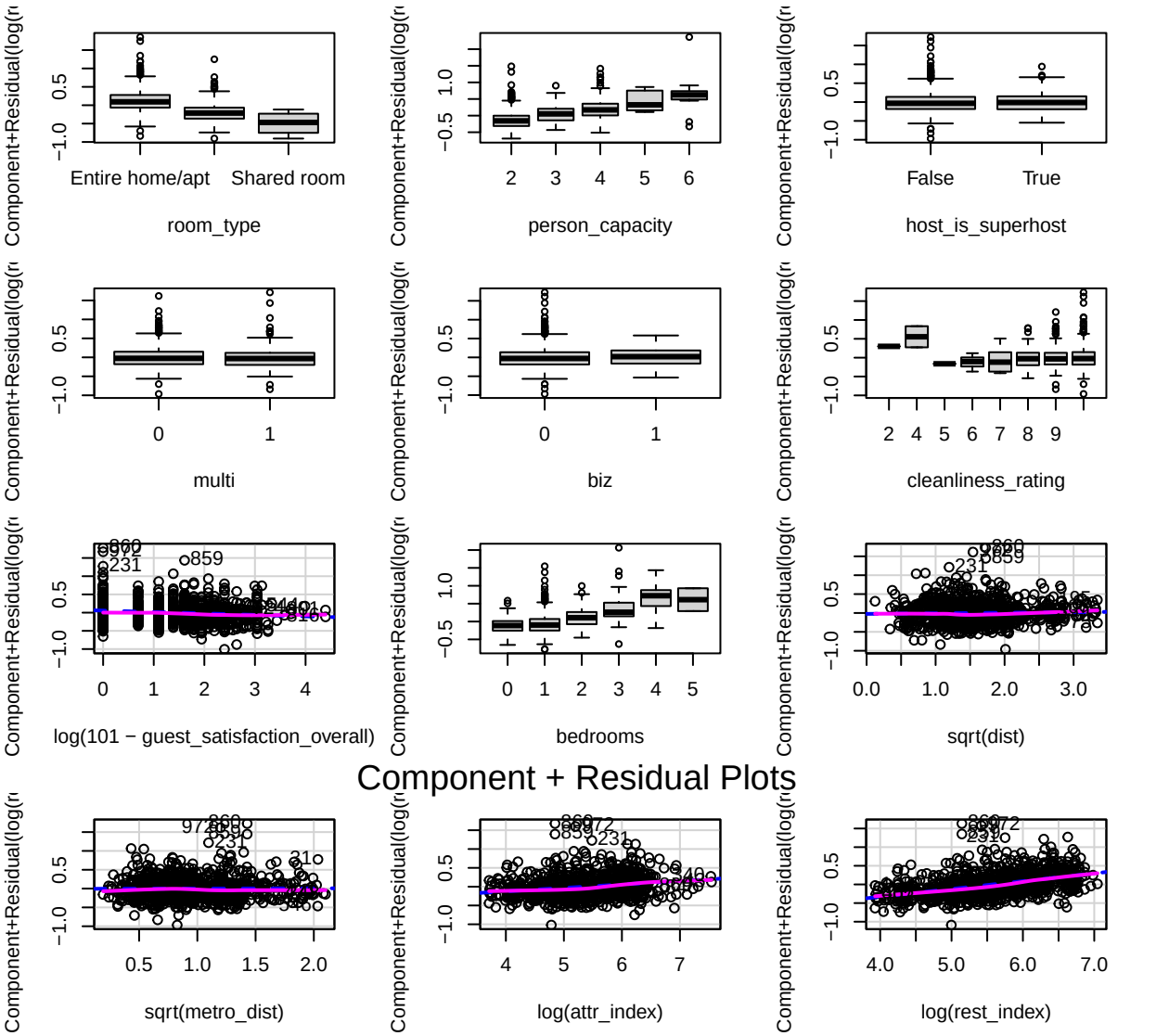


```r
plot(airbnb_lm_transformed, which = 2)
```

```
## Warning: not plotting observations with leverage one:
##   816
```

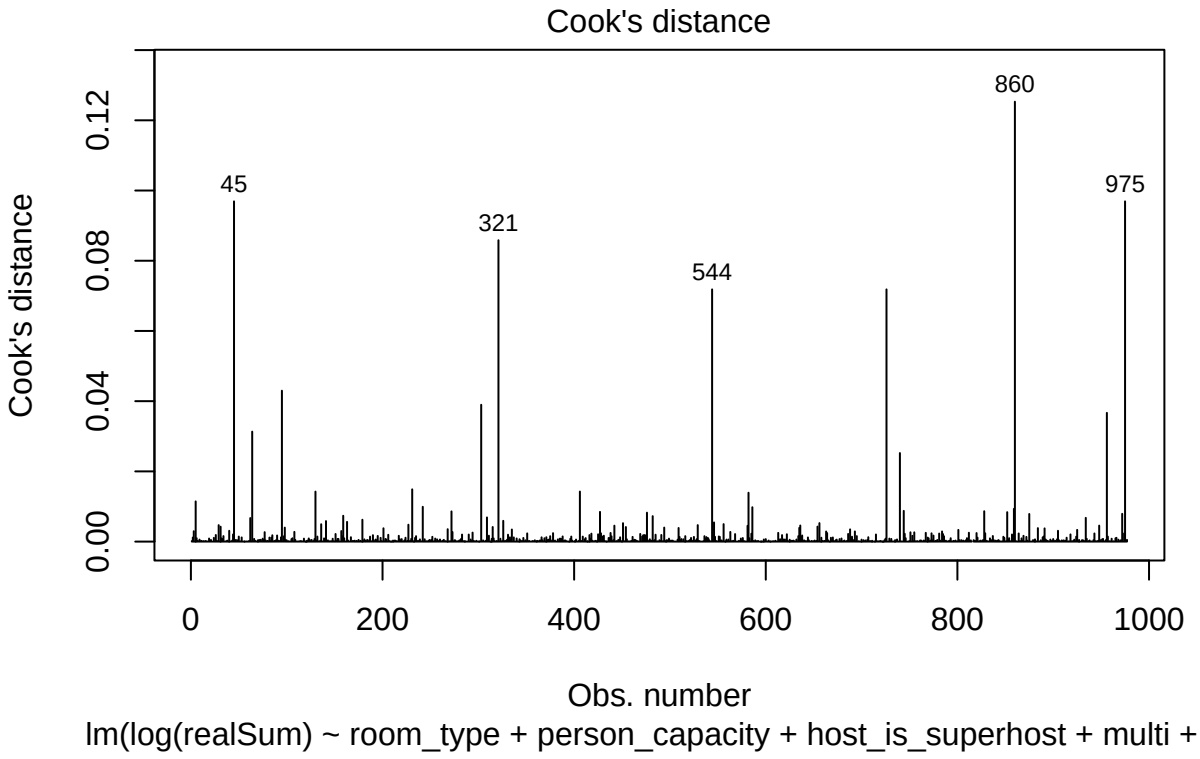



```r
# partial residuals
crp(airbnb_lm_transformed, id = list(n = 4))
```




Component + Residual Plots

```
# check outliers
plot(airbnb_lm_transformed, which = 4, id.n = 5)
```

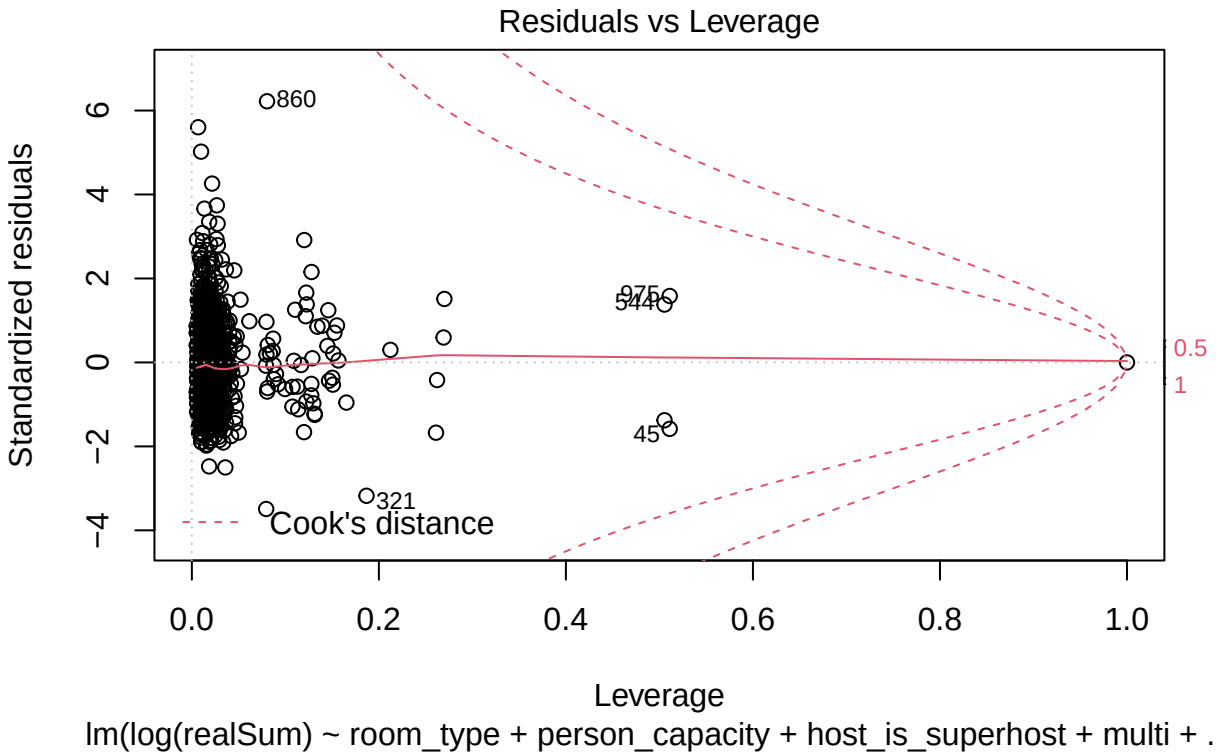



```
plot(airbnb_lm_transformed, which = 5, id.n = 5)
```

```
## Warning: not plotting observations with leverage one:
##   816
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
# refit model without potential influential cases
airbnb_lm_transformed_no_inf <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+biz+
# bedrooms5 is not significant anymore, only has 2 samples
summary(airbnb_lm_transformed_no_inf)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(dist) + sqrt(metro_dist) + log(attr_index) +
##     log(rest_index), data = airbnb, subset = -c(975))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96477 -0.18614 -0.02309  0.14060  1.72310
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.825770   0.483231   9.986  < 2e-16 ***
## room_typePrivate room           -0.323184   0.022227 -14.540  < 2e-16 ***
## room_typeShared room            -0.617570   0.148256  -4.166 3.39e-05 ***
## person_capacity3                 0.193258   0.041152   4.696 3.04e-06 ***
## person_capacity4                 0.326863   0.031084  10.516  < 2e-16 ***
## person_capacity5                 0.541535   0.097197   5.572 3.29e-08 ***
## person_capacity6                 0.770055   0.085173   9.041  < 2e-16 ***
## host_is_superhostTrue           -0.001519   0.022282  -0.068 0.945661
## multi1                          -0.016696   0.023078  -0.723 0.469575
## biz1                             0.010443   0.034173   0.306 0.759988
## cleanliness_rating4              0.249953   0.357949   0.698 0.485166
## cleanliness_rating5             -0.465473   0.415685  -1.120 0.263094
```

```
## cleanliness_rating6                   -0.418250   0.312584  -1.338 0.181203
## cleanliness_rating7                   -0.375817   0.309774  -1.213 0.225356
## cleanliness_rating8                   -0.326964   0.295302  -1.107 0.268480
## cleanliness_rating9                   -0.312138   0.294222  -1.061 0.289009
## cleanliness_rating10                  -0.297743   0.295384  -1.008 0.313718
## log(101 - guest_satisfaction_overall) -0.040749   0.012297  -3.314 0.000956 ***
## bedrooms1                              0.045765   0.039277   1.165 0.244235
## bedrooms2                              0.222333   0.049297   4.510 7.29e-06 ***
## bedrooms3                              0.451472   0.062301   7.247 8.84e-13 ***
## bedrooms4                              0.757066   0.107834   7.021 4.20e-12 ***
## bedrooms5                              0.395990   0.294705   1.344 0.179372
## sqrt(dist)                             0.013939   0.057030   0.244 0.806961
## sqrt(metro_dist)                       0.006623   0.029688   0.223 0.823514
## log(attr_index)                        0.094123   0.065650   1.434 0.151985
## log(rest_index)                        0.205588   0.078044   2.634 0.008569 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2888 on 949 degrees of freedom
## Multiple R-squared:  0.6987, Adjusted R-squared:  0.6905
## F-statistic: 84.65 on 26 and 949 DF,  p-value: < 2.2e-16
```

```
airbnb_lm_transformed_no_inf2 <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+bi
# bedrooms5 is gone
summary(airbnb_lm_transformed_no_inf2)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(dist) + sqrt(metro_dist) + log(attr_index) +
##     log(rest_index), data = airbnb, subset = -c(975, 45))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96477 -0.18625 -0.02339  0.14064  1.72310
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           4.825770   0.483231   9.986  < 2e-16 ***
## room_typePrivate room                -0.323184   0.022227 -14.540  < 2e-16 ***
## room_typeShared room                 -0.617570   0.148256  -4.166 3.39e-05 ***
## person_capacity3                      0.193258   0.041152   4.696 3.04e-06 ***
## person_capacity4                      0.326863   0.031084  10.516  < 2e-16 ***
## person_capacity5                      0.541535   0.097197   5.572 3.29e-08 ***
## person_capacity6                      0.770055   0.085173   9.041  < 2e-16 ***
## host_is_superhostTrue                -0.001519   0.022282  -0.068 0.945661
## multi1                               -0.016696   0.023078  -0.723 0.469575
## biz1                                  0.010443   0.034173   0.306 0.759988
## cleanliness_rating4                   0.249953   0.357949   0.698 0.485166
## cleanliness_rating5                  -0.465473   0.415685  -1.120 0.263094
## cleanliness_rating6                  -0.418250   0.312584  -1.338 0.181203
## cleanliness_rating7                  -0.375817   0.309774  -1.213 0.225356
## cleanliness_rating8                  -0.326964   0.295302  -1.107 0.268480
## cleanliness_rating9                  -0.312138   0.294222  -1.061 0.289009
```

```
## cleanliness_rating10                  -0.297743   0.295384  -1.008 0.313718
## log(101 - guest_satisfaction_overall) -0.040749   0.012297  -3.314 0.000956 ***
## bedrooms1                              0.045765   0.039277   1.165 0.244235
## bedrooms2                              0.222333   0.049297   4.510 7.29e-06 ***
## bedrooms3                              0.451472   0.062301   7.247 8.84e-13 ***
## bedrooms4                              0.757066   0.107834   7.021 4.20e-12 ***
## sqrt(dist)                             0.013939   0.057030   0.244 0.806961
## sqrt(metro_dist)                       0.006623   0.029688   0.223 0.823514
## log(attr_index)                        0.094123   0.065650   1.434 0.151985
## log(rest_index)                        0.205588   0.078044   2.634 0.008569 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2888 on 949 degrees of freedom
## Multiple R-squared:  0.6985, Adjusted R-squared:  0.6905
## F-statistic: 87.94 on 25 and 949 DF,  p-value: < 2.2e-16
```

```
airbnb_lm_transformed_no_inf3 <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+bi
# bedrooms5 is gone, no significant change
summary(airbnb_lm_transformed_no_inf3)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(dist) + sqrt(metro_dist) + log(attr_index) +
##     log(rest_index), data = airbnb, subset = -c(975, 45, 544,
##     726))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96162 -0.18657 -0.02314  0.14045  1.72403
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           4.801261   0.483313   9.934  < 2e-16 ***
## room_typePrivate room                -0.323445   0.022217 -14.558  < 2e-16 ***
## room_typeShared room                 -0.617442   0.148183  -4.167 3.37e-05 ***
## person_capacity3                      0.193065   0.041132   4.694 3.08e-06 ***
## person_capacity4                      0.326619   0.031069  10.513  < 2e-16 ***
## person_capacity5                      0.540814   0.097150   5.567 3.38e-08 ***
## person_capacity6                      0.768837   0.085135   9.031  < 2e-16 ***
## host_is_superhostTrue                -0.001383   0.022272  -0.062 0.950500
## multi1                               -0.016080   0.023071  -0.697 0.485986
## biz1                                  0.014562   0.034284   0.425 0.671114
## cleanliness_rating5                  -0.464509   0.415480  -1.118 0.263849
## cleanliness_rating6                  -0.419596   0.312431  -1.343 0.179591
## cleanliness_rating7                  -0.376327   0.309621  -1.215 0.224499
## cleanliness_rating8                  -0.328626   0.295159  -1.113 0.265825
## cleanliness_rating9                  -0.314079   0.294080  -1.068 0.285790
## cleanliness_rating10                 -0.300169   0.295243  -1.017 0.309563
## log(101 - guest_satisfaction_overall) -0.041387   0.012300  -3.365 0.000797 ***
## bedrooms1                             0.046476   0.039261   1.184 0.236801
## bedrooms2                             0.223182   0.049277   4.529 6.68e-06 ***
## bedrooms3                             0.452142   0.062272   7.261 8.01e-13 ***
```

```
## bedrooms4                                0.757246   0.107780   7.026 4.06e-12 ***
## sqrt(dist)                               0.017512   0.057059   0.307 0.758986
## sqrt(metro_dist)                         0.007753   0.029684   0.261 0.794006
## log(attr_index)                          0.095472   0.065625   1.455 0.146050
## log(rest_index)                          0.207846   0.078022   2.664 0.007854 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2886 on 948 degrees of freedom
## Multiple R-squared:  0.6991, Adjusted R-squared:  0.6914
## F-statistic: 91.75 on 24 and 948 DF,  p-value: < 2.2e-16
```

```
# check collinearity
vif(airbnb_lm_transformed_no_inf2)
```

```
##                                      GVIF Df GVIF^(1/(2*Df))
## room_type                        1.428492  2        1.093250
## person_capacity                  2.825552  4        1.138644
## host_is_superhost                1.157108  1        1.075690
## multi                            1.184242  1        1.088229
## biz                              1.155490  1        1.074937
## cleanliness_rating               1.789273  7        1.042433
## log(101 - guest_satisfaction_overall)  1.576988  1        1.255782
## bedrooms                         2.965420  4        1.145541
## sqrt(dist)                      12.362496  1        3.516034
## sqrt(metro_dist)                 1.449851  1        1.204098
## log(attr_index)                 23.828914  1        4.881487
## log(rest_index)                 32.714445  1        5.719654
```

```
# transformed model - removing dist
airbnb_lm_transformed_1 <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+biz+clean
summary(airbnb_lm_transformed_1)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index) + log(rest_index),
##     data = airbnb, subset = -c(975, 45))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96597 -0.18627 -0.02322  0.13955  1.72091
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       4.913499   0.323386  15.194  < 2e-16 ***
## room_typePrivate room            -0.322826   0.022168 -14.563  < 2e-16 ***
## room_typeShared room             -0.615948   0.148034  -4.161 3.46e-05 ***
## person_capacity3                  0.193467   0.041122   4.705 2.92e-06 ***
## person_capacity4                  0.326752   0.031065  10.518  < 2e-16 ***
## person_capacity5                  0.540862   0.097109   5.570 3.32e-08 ***
## person_capacity6                  0.769682   0.085117   9.043  < 2e-16 ***
## host_is_superhostTrue            -0.001447   0.022269  -0.065 0.948219
## multi1                           -0.016646   0.023066  -0.722 0.470669
## biz1                              0.009801   0.034056   0.288 0.773556
```

```
## cleanliness_rating4                     0.251165   0.357737   0.702 0.482792
## cleanliness_rating5                    -0.467376   0.415407  -1.125 0.260829
## cleanliness_rating6                    -0.416314   0.312329  -1.333 0.182872
## cleanliness_rating7                    -0.375736   0.309620  -1.214 0.225226
## cleanliness_rating8                    -0.326106   0.295135  -1.105 0.269467
## cleanliness_rating9                    -0.310796   0.294025  -1.057 0.290763
## cleanliness_rating10                   -0.296547   0.295197  -1.005 0.315358
## log(101 - guest_satisfaction_overall) -0.040662   0.012286  -3.310 0.000969 ***
## bedrooms1                               0.045354   0.039221   1.156 0.247828
## bedrooms2                               0.222247   0.049272   4.511 7.27e-06 ***
## bedrooms3                               0.451765   0.062258   7.256 8.26e-13 ***
## bedrooms4                               0.757603   0.107758   7.031 3.93e-12 ***
## sqrt(metro_dist)                        0.004537   0.028421   0.160 0.873199
## log(attr_index)                         0.092578   0.065313   1.417 0.156678
## log(rest_index)                         0.195448   0.066070   2.958 0.003171 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2886 on 950 degrees of freedom
## Multiple R-squared:  0.6985, Adjusted R-squared:  0.6908
## F-statistic: 91.69 on 24 and 950 DF,  p-value: < 2.2e-16
```

```
# anova test
anova(airbnb_lm_transformed_1, airbnb_lm_transformed_no_inf2)
```

```
## Analysis of Variance Table
##
## Model 1: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index) + log(rest_index)
## Model 2: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(dist) + sqrt(metro_dist) + log(attr_index) +
##     log(rest_index)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    950 79.139
## 2    949 79.134  1 0.0049815 0.0597  0.807
```

```
# collinearity between attr_index and rest_index is high after removing sqrt(dist)
vif(airbnb_lm_transformed_1)
```

```
##                                        GVIF Df GVIF^(1/(2*Df))
## room_type                          1.420219  2        1.091664
## person_capacity                    2.820576  4        1.138393
## host_is_superhost                  1.156903  1        1.075594
## multi                              1.184149  1        1.088186
## biz                                1.148677  1        1.071764
## cleanliness_rating                 1.775364  7        1.041853
## log(101 - guest_satisfaction_overall)  1.575671  1        1.255257
## bedrooms                           2.948508  4        1.144723
## sqrt(metro_dist)                   1.330032  1        1.153270
## log(attr_index)                   23.607847  1        4.858791
## log(rest_index)                   23.469291  1        4.844511
```

```
# transformed model - removing attr_index
airbnb_lm_transformed_2 <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+biz+clean
```

```
summary(airbnb_lm_transformed_2)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(rest_index), data = airbnb,
##     subset = -c(975, 45))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96708 -0.18494 -0.02415  0.14086  1.71727
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          4.917225   0.323547  15.198  < 2e-16 ***
## room_typePrivate room               -0.322895   0.022180 -14.558  < 2e-16 ***
## room_typeShared room                -0.611118   0.148073  -4.127 4.00e-05 ***
## person_capacity3                     0.193625   0.041144   4.706 2.90e-06 ***
## person_capacity4                     0.329172   0.031034  10.607  < 2e-16 ***
## person_capacity5                     0.542202   0.097156   5.581 3.12e-08 ***
## person_capacity6                     0.771936   0.085147   9.066  < 2e-16 ***
## host_is_superhostTrue               -0.000113   0.022261  -0.005  0.99595
## multi1                              -0.019445   0.022994  -0.846  0.39795
## biz1                                 0.008167   0.034054   0.240  0.81051
## cleanliness_rating4                  0.240282   0.357844   0.671  0.50208
## cleanliness_rating5                 -0.481796   0.415502  -1.160  0.24652
## cleanliness_rating6                 -0.427650   0.312392  -1.369  0.17134
## cleanliness_rating7                 -0.387241   0.309678  -1.250  0.21144
## cleanliness_rating8                 -0.338553   0.295161  -1.147  0.25167
## cleanliness_rating9                 -0.321782   0.294079  -1.094  0.27414
## cleanliness_rating10                -0.309194   0.295219  -1.047  0.29521
## log(101 - guest_satisfaction_overall) -0.040611   0.012293  -3.304  0.00099 ***
## bedrooms1                            0.045930   0.039240   1.170  0.24210
## bedrooms2                            0.221624   0.049296   4.496 7.79e-06 ***
## bedrooms3                            0.450972   0.062289   7.240 9.25e-13 ***
## bedrooms4                            0.770713   0.107417   7.175 1.45e-12 ***
## sqrt(metro_dist)                     0.001025   0.028327   0.036  0.97116
## log(rest_index)                      0.286335   0.015941  17.962  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2888 on 951 degrees of freedom
## Multiple R-squared:  0.6978, Adjusted R-squared:  0.6905
## F-statistic: 95.49 on 23 and 951 DF,  p-value: < 2.2e-16
```

```
# anova test
anova(airbnb_lm_transformed_1, airbnb_lm_transformed_2)
```

```
## Analysis of Variance Table
##
## Model 1: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index) + log(rest_index)
```

```
## Model 2: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(rest_index)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    950 79.139
## 2    951 79.307 -1  -0.16737 2.0092 0.1567
```

```
# transformed model - removing rest_index
airbnb_lm_transformed_3 <- lm(log(realSum) ~ room_type+person_capacity+host_is_superhost+multi+biz+clean
summary(airbnb_lm_transformed_3)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index), data = airbnb,
##     subset = -c(975, 45))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96985 -0.18250 -0.02421  0.14363  1.72845
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          4.991055   0.323632  15.422  < 2e-16 ***
## room_typePrivate room               -0.324203   0.022253 -14.569  < 2e-16 ***
## room_typeShared room                -0.633116   0.148522  -4.263 2.22e-05 ***
## person_capacity3                     0.191190   0.041282   4.631 4.14e-06 ***
## person_capacity4                     0.322806   0.031163  10.359  < 2e-16 ***
## person_capacity5                     0.543614   0.097500   5.576 3.21e-08 ***
## person_capacity6                     0.765685   0.085452   8.960  < 2e-16 ***
## host_is_superhostTrue               -0.004363   0.022338  -0.195 0.845195
## multi1                              -0.010307   0.023060  -0.447 0.654986
## biz1                                 0.015493   0.034140   0.454 0.650062
## cleanliness_rating4                  0.260581   0.359178   0.725 0.468329
## cleanliness_rating5                 -0.459788   0.417088  -1.102 0.270577
## cleanliness_rating6                 -0.404090   0.313571  -1.289 0.197826
## cleanliness_rating7                 -0.361713   0.310843  -1.164 0.244856
## cleanliness_rating8                 -0.310411   0.296288  -1.048 0.295058
## cleanliness_rating9                 -0.297257   0.295185  -1.007 0.314182
## cleanliness_rating10                -0.279799   0.296343  -0.944 0.345322
## log(101 - guest_satisfaction_overall) -0.040923   0.012336  -3.317 0.000943 ***
## bedrooms1                            0.041021   0.039353   1.042 0.297507
## bedrooms2                            0.219498   0.049463   4.438 1.02e-05 ***
## bedrooms3                            0.447713   0.062496   7.164 1.57e-12 ***
## bedrooms4                            0.730524   0.107805   6.776 2.16e-11 ***
## sqrt(metro_dist)                     0.002186   0.028525   0.077 0.938935
## log(attr_index)                      0.280083   0.015814  17.711  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2898 on 951 degrees of freedom
## Multiple R-squared:  0.6957, Adjusted R-squared:  0.6883
## F-statistic: 94.53 on 23 and 951 DF,  p-value: < 2.2e-16
```

```r
# anova test
anova(airbnb_lm_transformed_1, airbnb_lm_transformed_3)
```

```
## Analysis of Variance Table
##
## Model 1: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index) + log(rest_index)
## Model 2: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(attr_index)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    950 79.139
## 2    951 79.868 -1  -0.72899 8.7509 0.003171 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# no significant collinearity
vif(airbnb_lm_transformed_2)
```

```
##                                       GVIF Df GVIF^(1/(2*Df))
## room_type                         1.419437  2        1.091513
## person_capacity                   2.811435  4        1.137931
## host_is_superhost                 1.154838  1        1.074634
## multi                             1.175472  1        1.084192
## biz                               1.147361  1        1.071149
## cleanliness_rating                1.769499  7        1.041606
## log(101 - guest_satisfaction_overall) 1.575657  1        1.255252
## bedrooms                          2.919040  4        1.143286
## sqrt(metro_dist)                  1.319920  1        1.148878
## log(rest_index)                   1.364777  1        1.168236
```

```r
# anova test
airbnb_lm_transformed_reduced <- lm(log(realSum) ~ room_type+person_capacity+log(101-guest_satisfaction_
summary(airbnb_lm_transformed_reduced)
```

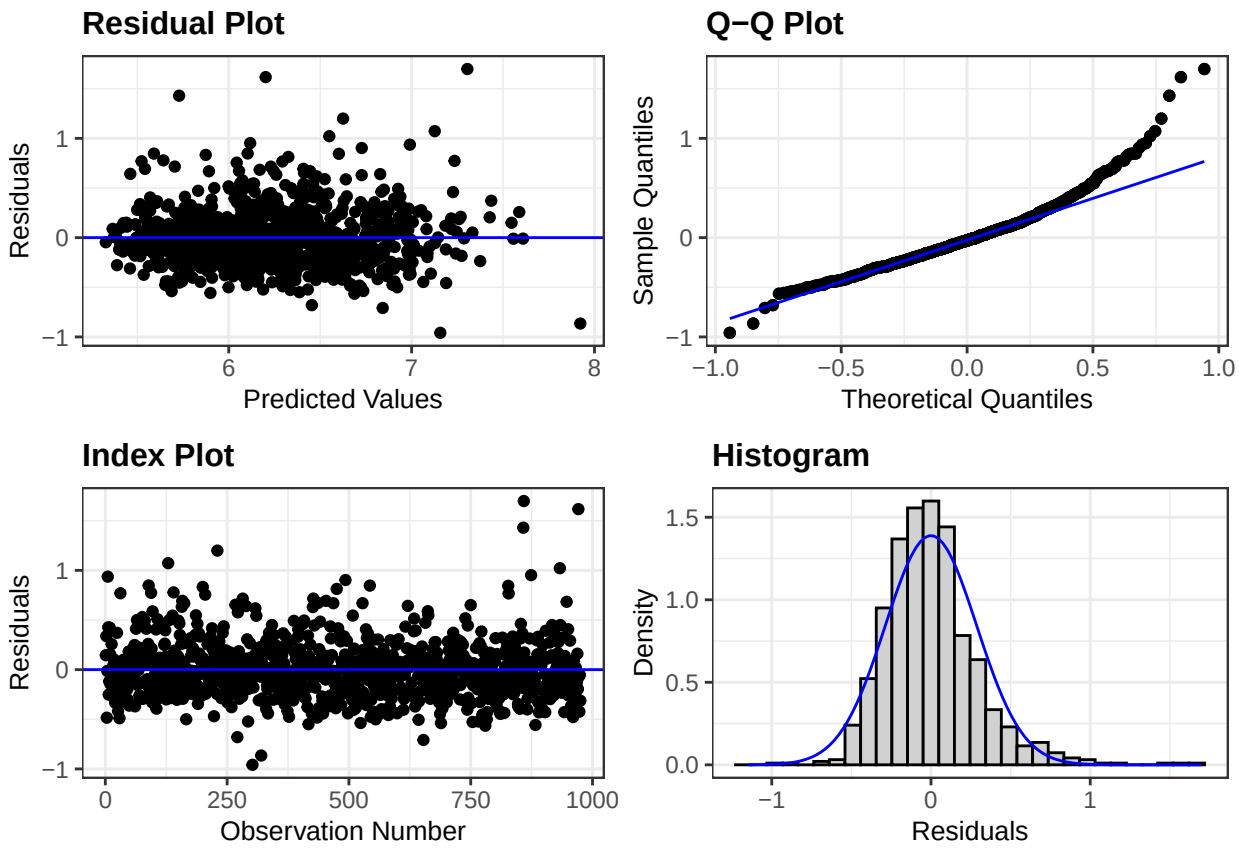
```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + log(101 -
##     guest_satisfaction_overall) + bedrooms + log(rest_index),
##     data = airbnb, subset = -c(975, 45))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95869 -0.18670 -0.02772  0.13846  1.69852
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         4.59366    0.09305  49.366  < 2e-16 ***
## room_typePrivate room              -0.32412    0.02076 -15.610  < 2e-16 ***
## room_typeShared room               -0.62969    0.14651  -4.298 1.90e-05 ***
## person_capacity3                    0.19913    0.04074   4.888 1.19e-06 ***
## person_capacity4                    0.32786    0.03061  10.709  < 2e-16 ***
## person_capacity5                    0.53585    0.09648   5.554 3.61e-08 ***
## person_capacity6                    0.77503    0.08428   9.196  < 2e-16 ***
```

```
## log(101 - guest_satisfaction_overall) -0.04402    0.01008  -4.366 1.40e-05 ***
## bedrooms1                              0.05141    0.03868   1.329    0.184
## bedrooms2                              0.22553    0.04832   4.667 3.49e-06 ***
## bedrooms3                              0.45785    0.06163   7.429 2.41e-13 ***
## bedrooms4                              0.76923    0.10681   7.202 1.20e-12 ***
## log(rest_index)                        0.28734    0.01397  20.563  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.289 on 962 degrees of freedom
## Multiple R-squared:  0.6939, Adjusted R-squared:   0.69
## F-statistic: 181.7 on 12 and 962 DF,  p-value: < 2.2e-16
```

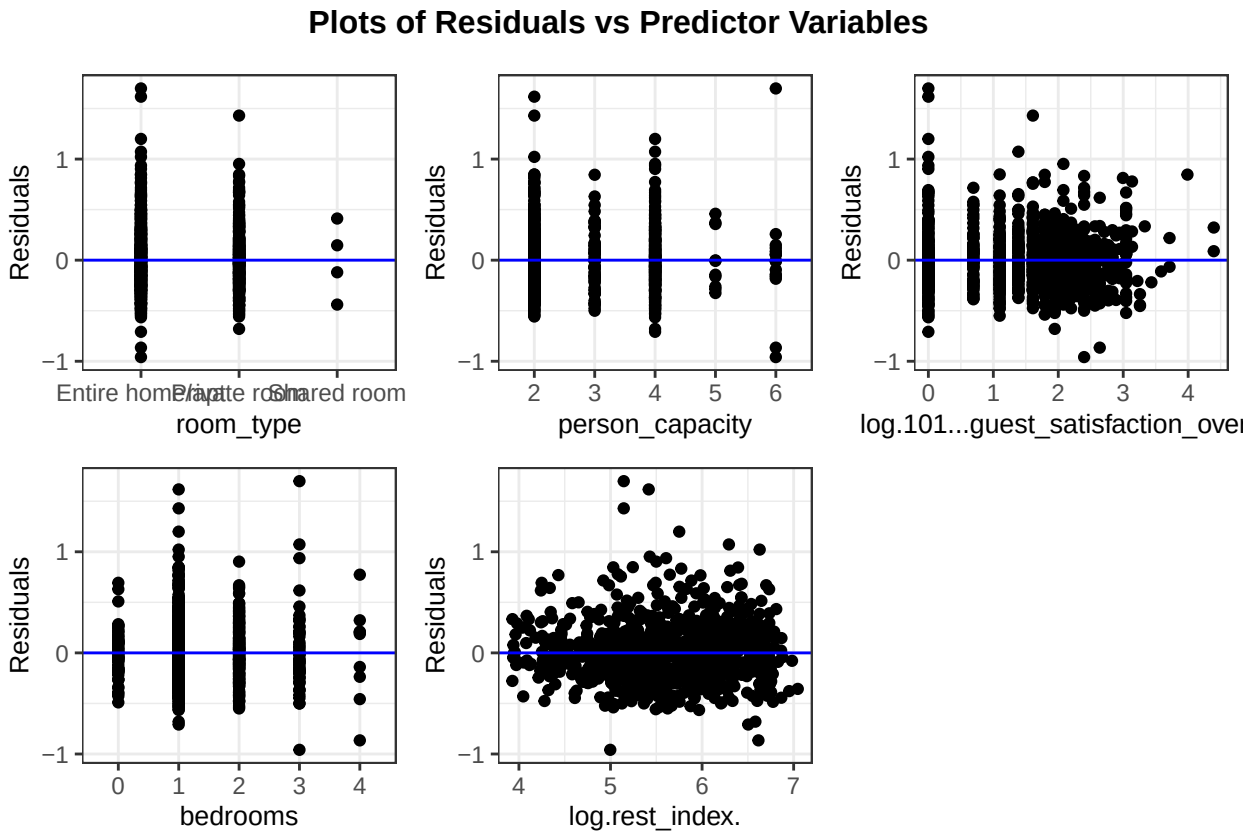
```
anova(airbnb_lm_transformed_reduced, airbnb_lm_transformed_2)
```

```
## Analysis of Variance Table
##
## Model 1: log(realSum) ~ room_type + person_capacity + log(101 - guest_satisfaction_overall) +
##     bedrooms + log(rest_index)
## Model 2: log(realSum) ~ room_type + person_capacity + host_is_superhost +
##     multi + biz + cleanliness_rating + log(101 - guest_satisfaction_overall) +
##     bedrooms + sqrt(metro_dist) + log(rest_index)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    962 80.346
## 2    951 79.307 11    1.0392 1.1329 0.3315
```

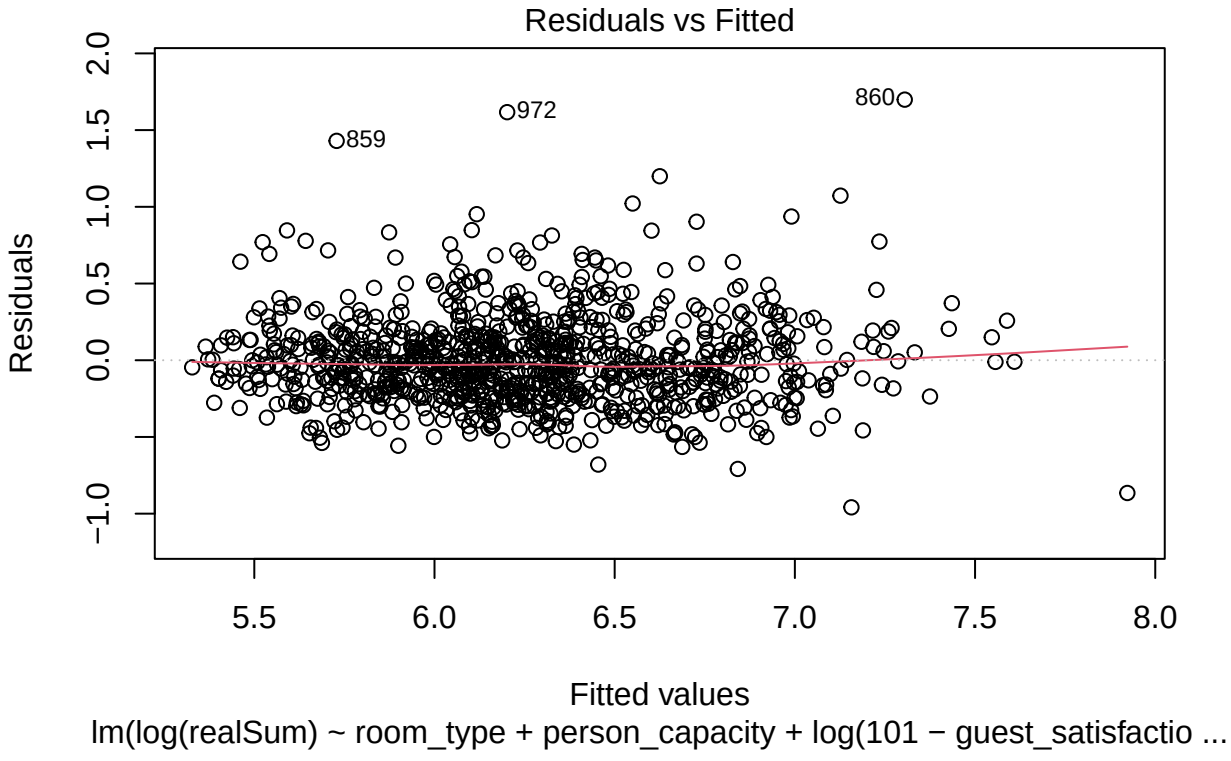
```
# check assumptions
resid_panel(airbnb_lm_transformed_reduced)
```



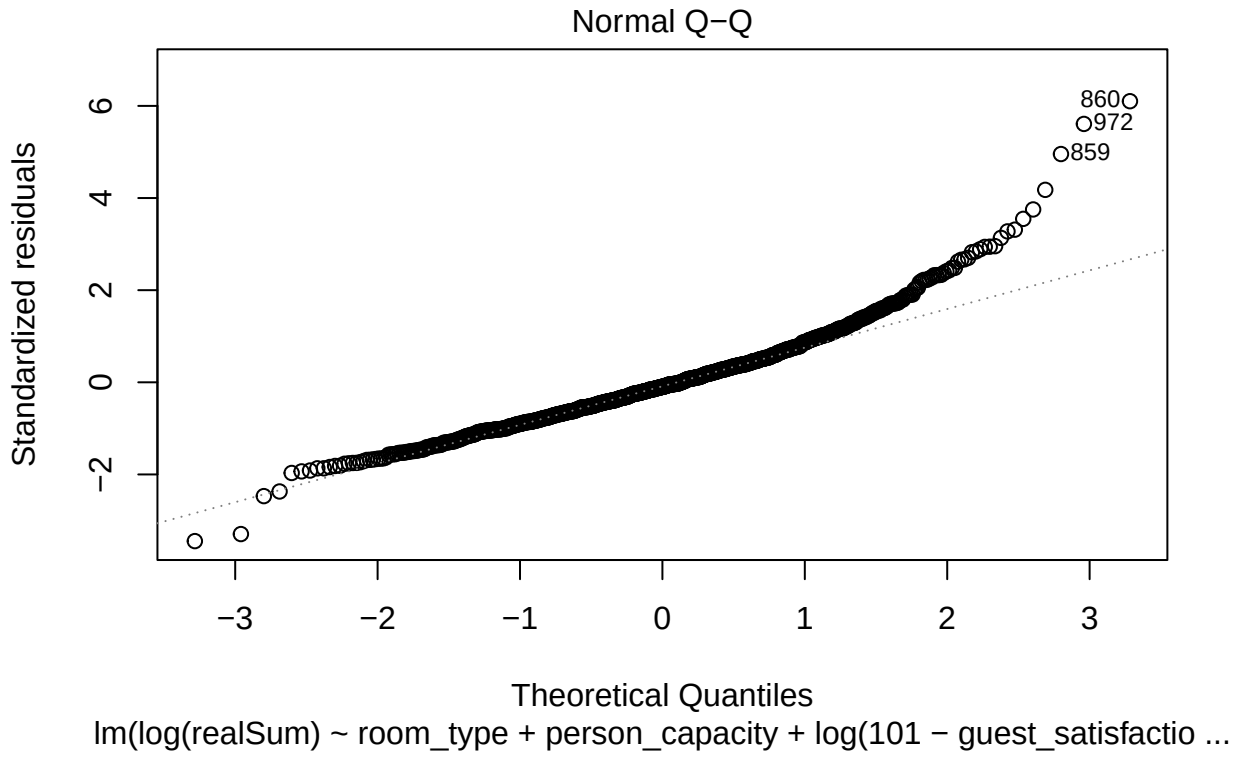
```
resid_xpanel(airbnb_lm_transformed_reduced)
```



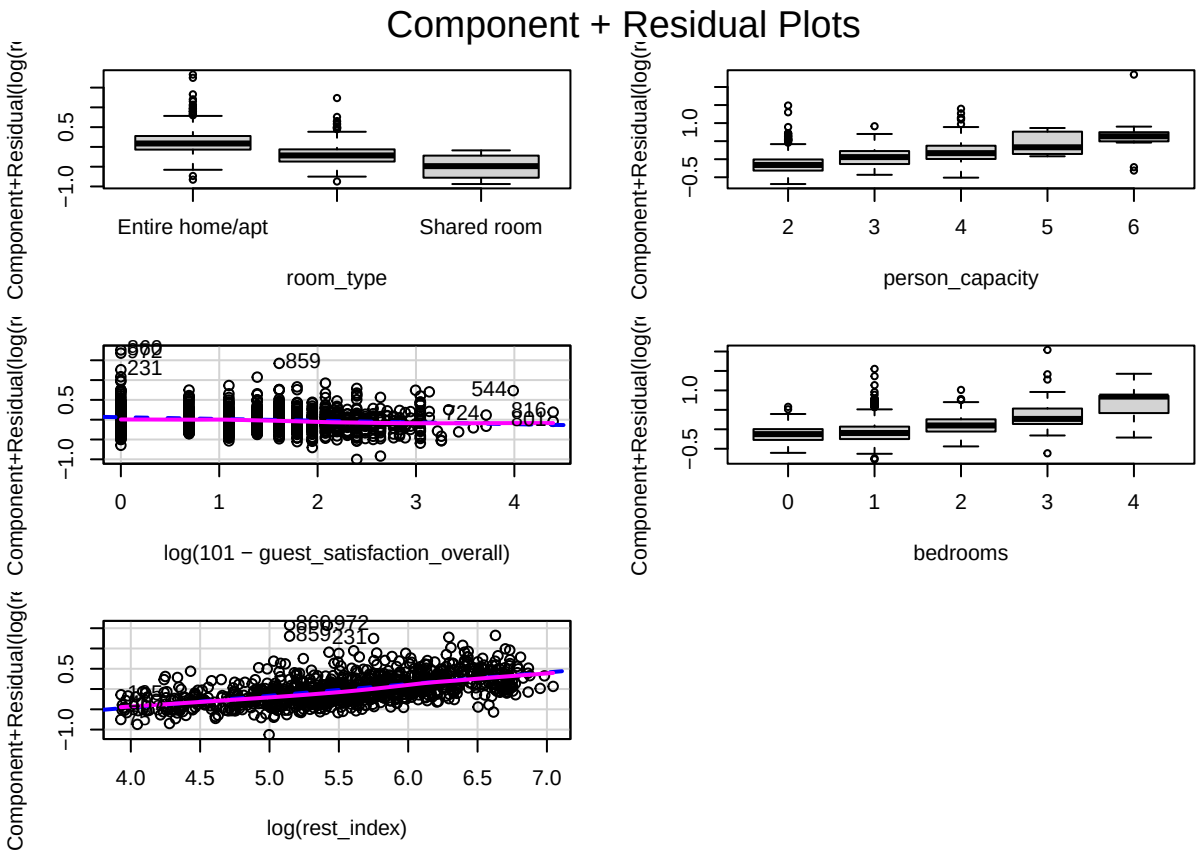

```
plot(airbnb_lm_transformed_reduced, which = 1)
```



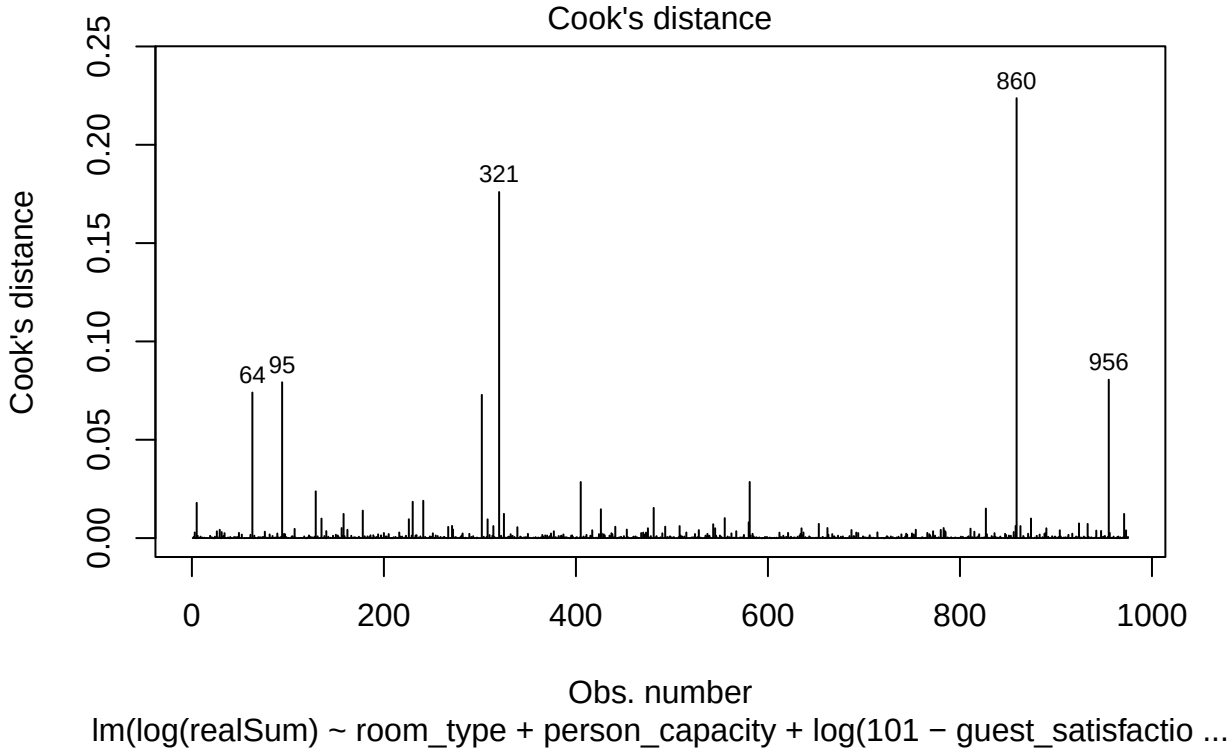
```
plot(airbnb_lm_transformed_reduced, which = 2)
```



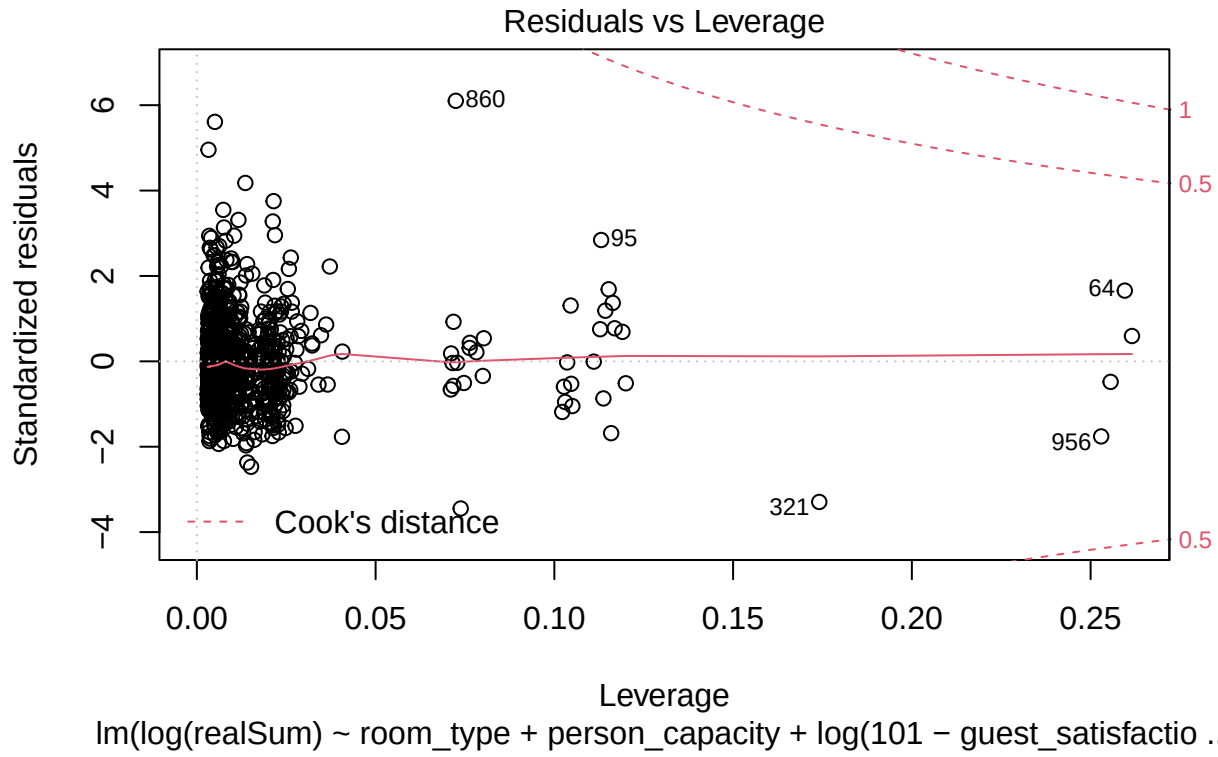

```
# partial residuals
crp(airbnb_lm_transformed_reduced, id = list(n = 4))
```




```
# check outliers
plot(airbnb_lm_transformed_reduced, which = 4, id.n = 5)
```



```r
plot(airbnb_lm_transformed_reduced, which = 5, id.n = 5)
```




```r
# refit model without potential influential cases
airbnb_lm_transformed_reduced_no_inf <- lm(log(realSum) ~ room_type+person_capacity+log(101-guest_satisf
# no change in significance
summary(airbnb_lm_transformed_reduced_no_inf)
```

```
##
## Call:
## lm(formula = log(realSum) ~ room_type + person_capacity + log(101 -
##     guest_satisfaction_overall) + bedrooms + log(rest_index),
##     data = airbnb, subset = -c(975, 45, 321, 956, 860, 95, 64))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88852 -0.18304 -0.02872  0.13789  1.62212
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          4.576224   0.090793  50.403  < 2e-16 ***
## room_typePrivate room               -0.324917   0.020222 -16.068  < 2e-16 ***
## room_typeShared room                -0.617941   0.201574  -3.066  0.00223 **
## person_capacity3                     0.201259   0.039726   5.066 4.87e-07 ***
## person_capacity4                     0.331446   0.029881  11.092  < 2e-16 ***
## person_capacity5                     0.542840   0.093902   5.781 1.00e-08 ***
## person_capacity6                     0.728302   0.086770   8.393  < 2e-16 ***
## log(101 - guest_satisfaction_overall) -0.040699   0.009832  -4.139 3.79e-05 ***
## bedrooms1                            0.053011   0.037632   1.409  0.15926
## bedrooms2                            0.225328   0.047058   4.788 1.95e-06 ***
## bedrooms3                            0.434056   0.060092   7.223 1.04e-12 ***
```

```
## bedrooms4                                0.781077   0.115223   6.779 2.12e-11 ***
## log(rest_index)                          0.289307   0.013647  21.199  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2812 on 957 degrees of freedom
## Multiple R-squared:  0.6974, Adjusted R-squared:  0.6936
## F-statistic: 183.8 on 12 and 957 DF,  p-value: < 2.2e-16
```

```
# check collinearity
vif(airbnb_lm_transformed_reduced)
```

```
##                                       GVIF Df GVIF^(1/(2*Df))
## room_type                         1.220335  2        1.051041
## person_capacity                   2.584994  4        1.126049
## log(101 - guest_satisfaction_overall) 1.058207  1        1.028692
## bedrooms                          2.722158  4        1.133350
## log(rest_index)                   1.047144  1        1.023300
```

```
#  interpreation stats
(exp(-0.324) - 1)*100
```

```
## [1] -27.67498
```

```
(exp(-0.630) - 1)*100
```

```
## [1] -46.74082
```

```
(exp(0.199) - 1)*100
```

```
## [1] 22.0182
```

```
(exp(0.328) - 1)*100
```

```
## [1] 38.8189
```

```
(exp(0.536) - 1)*100
```

```
## [1] 70.91565
```

```
(exp(0.775) - 1)*100
```

```
## [1] 117.0592
```

```
(2^-0.044 - 1)*100
```

```
## [1] -3.003809
```

```
(exp(0.226) - 1)*100
```

```
## [1] 25.35757
```

```
(exp(0.458) - 1)*100
```

```
## [1] 58.0909
```

```
(exp(0.769) - 1)*100
```

```
## [1] 115.7608
```

```
(2^0.287 - 1)*100
```

```
## [1] 22.01005
```

```r
# confidence interval
(exp(-0.365) - 1)*100
```

```
## [1] -30.58033
```

```r
(exp(-0.283) - 1)*100
```

```
## [1] -24.64802
```

```r
(exp(-0.917) - 1)*100
```

```
## [1] -60.02836
```

```r
(exp(-0.342) - 1)*100
```

```
## [1] -28.96518
```

```r
(exp(0.119) - 1)*100
```

```
## [1] 12.63699
```

```r
(exp(0.279) - 1)*100
```

```
## [1] 32.18073
```

```r
(exp(0.268) - 1)*100
```

```
## [1] 30.73471
```

```r
(exp(0.388) - 1)*100
```

```
## [1] 47.40298
```

```r
(exp(0.347) - 1)*100
```

```
## [1] 41.48167
```

```r
(exp(0.725) - 1)*100
```

```
## [1] 106.4731
```

```r
(exp(0.610) - 1)*100
```

```
## [1] 84.04314
```

```r
(exp(0.940) - 1)*100
```

```
## [1] 155.9981
```

```r
(2^-0.064 - 1)*100
```

```
## [1] -4.339184
```

```r
(2^-0.024 - 1)*100
```

```
## [1] -1.649793
```

```r
(exp(0.131) - 1)*100
```

```
## [1] 13.99678
```

```r
(exp(0.320) - 1)*100
```

```
## [1] 37.71278
```

```r
(exp(0.337) - 1)*100
```

```
## [1] 40.07391
```

```r
(exp(0.579) - 1)*100
```

```
## [1] 78.42533
```

```r
(exp(0.560) - 1)*100
```

```
## [1] 75.06725
```

```r
(exp(0.979) - 1)*100
```

```
## [1] 166.1793
```

```r
(2^0.260 - 1)*100
```

```
## [1] 19.74787
```

```r
(2^0.315 - 1)*100
```

```
## [1] 24.40117
```