

# Multivariate Linear Regression Midterm Project

Diana Rueda

5/3/2021

```
#Loading and showing the head of the dataset containing dummy variables
library(readr)
df_dummies <- read_csv("airbnb_dummies.csv")

## Warning: Missing column names filled in: 'X1' [1]

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   price = col_double(),
##   minimum_nights = col_double(),
##   number_of_reviews = col_double(),
##   availability_365 = col_double(),
##   neighbourhood_group_Bronx = col_double(),
##   neighbourhood_group_Brooklyn = col_double(),
##   neighbourhood_group_Manhattan = col_double(),
##   neighbourhood_group_Queens = col_double(),
##   `neighbourhood_group_Staten Island` = col_double(),
##   `room_type_Entire home/apt` = col_double(),
##   `room_type_Private room` = col_double(),
##   `room_type_Shared room` = col_double()
## )
df_dummies <- df_dummies[, !(names(df_dummies) %in% c("X1"))]
head(df_dummies)

## # A tibble: 6 x 12
##   price minimum_nights number_of_reviews availability_365 neighbourhood_group_B~
##   <dbl>       <dbl>           <dbl>          <dbl>           <dbl>
## 1    225         1              45            355             0
## 2     89         1              270            194             0
## 3    200         3              74            129             0
## 4     79         2              430            220             0
## 5    150         1              160            188             0
## 6    135         5              53              6             0
## # ... with 7 more variables: neighbourhood_group_Brooklyn <dbl>,
## #   neighbourhood_group_Manhattan <dbl>, neighbourhood_group_Queens <dbl>,
## #   neighbourhood_group_Staten Island <dbl>, room_type_Entire home/apt <dbl>,
## #   room_type_Private room <dbl>, room_type_Shared room <dbl>
```

Building model with all variables to observe behavior

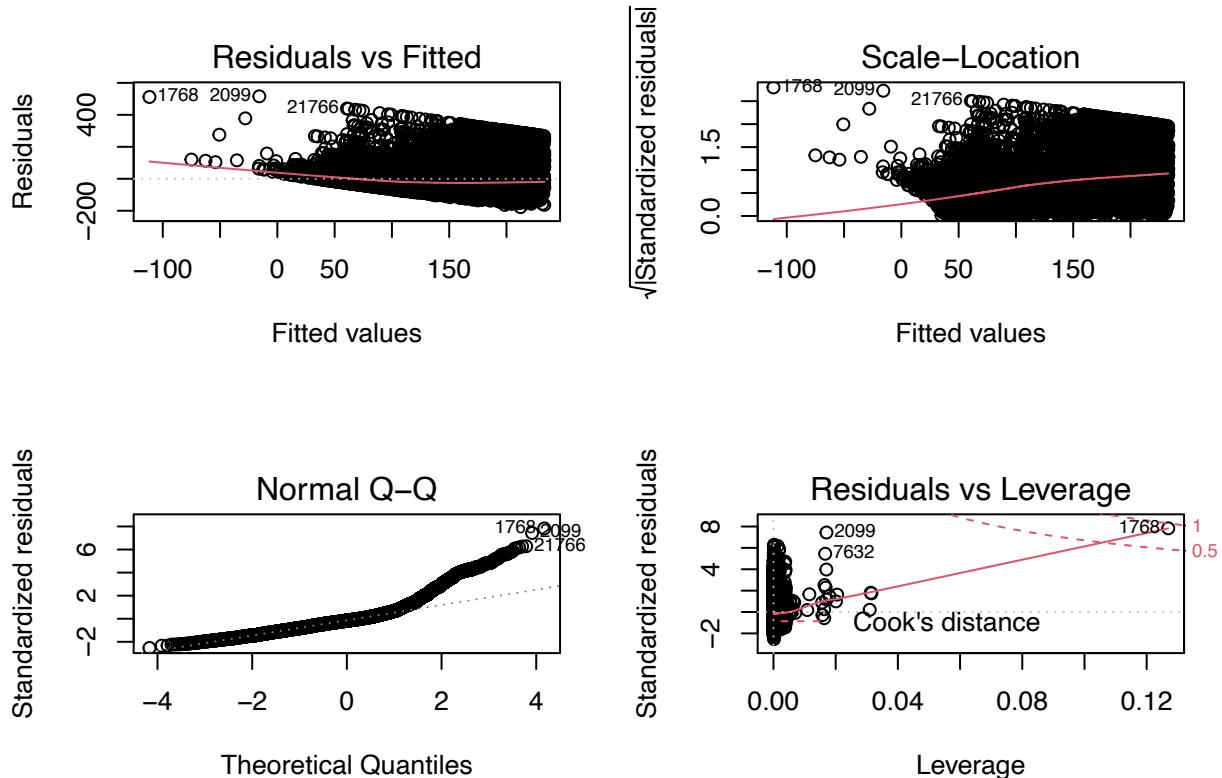
```
model <- lm(price ~ ., data = df_dummies)
summary(model)
```

```

## 
## Call:
## lm(formula = price ~ ., data = df_dummies)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -177.31  -40.54  -11.55   21.92  515.74 
## 
## Coefficients: (2 not defined because of singularities)
##                                     Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 14.941265  5.150107  2.901  0.00372 **  
## minimum_nights            -0.344781  0.025271 -13.643 < 2e-16 *** 
## number_of_reviews          -0.145603  0.007751 -18.786 < 2e-16 *** 
## availability_365           0.082668  0.003132  26.395 < 2e-16 *** 
## neighbourhood_group_Bronx    6.405901  4.861772   1.318  0.18764  
## neighbourhood_group_Brooklyn 32.370807  4.231385   7.650 2.06e-14 *** 
## neighbourhood_group_Manhattan 74.708492  4.228157  17.669 < 2e-16 *** 
## neighbourhood_group_Queens    18.247041  4.321685   4.222 2.43e-05 *** 
## `neighbourhood_group_Staten Island` NA         NA         NA         NA      
## `room_type_Entire home/apt`    113.576626 2.979913  38.114 < 2e-16 *** 
## `room_type_Private room`     20.458656  2.988997   6.845 7.80e-12 *** 
## `room_type_Shared room`      NA         NA         NA         NA      
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 69.94 on 32799 degrees of freedom
## Multiple R-squared:  0.3954, Adjusted R-squared:  0.3952 
## F-statistic:  2383 on 9 and 32799 DF,  p-value: < 2.2e-16 

# diagnostic plots
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(model)

```



Note: the dummy variables won't require all columns to be in the model since the last one will be a linear combination of the rest. If we include 2 out of 3 dummies for example in room\_type that will have carry the complete information from this variable. The correlation within dummy variables is what causes the NA values in the model above.

The following library performs diagnostics for different models in linear regression. It adds one variable at a time to evaluate which model has better performance. We input a model with all the possible variables and it gives us a final model.

```
# Stepwise Regression
library(MASS)
fit <- lm(price ~ ., data = df_dummies)
step <- stepAIC(fit, direction="both")

## Start:  AIC=278730.4
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##   neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##   `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##   `room_type_Private room` + `room_type_Shared room`
##
## Step:  AIC=278730.4
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##   neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##   `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##   `room_type_Private room`
```

```

## Step: AIC=278730.4
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##   neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##   `room_type_Entire home/apt` + `room_type_Private room`
##
##                                     Df Sum of Sq      RSS      AIC
## - neighbourhood_group_Bronx     1    8492 160442920 278730
## <none>                           160434428 278730
## - neighbourhood_group_Queens   1    87200 160521628 278746
## - `room_type_Private room`    1    229161 160663589 278775
## - neighbourhood_group_Brooklyn 1    286272 160720700 278787
## - minimum_nights                1    910476 161344904 278914
## - neighbourhood_group_Manhattan 1    1527125 161961553 279039
## - number_of_reviews              1    1726234 162160662 279080
## - availability_365              1    3407937 163842365 279418
## - `room_type_Entire home/apt`   1    7105714 167540142 280150
##
## Step: AIC=278730.2
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan +
##   neighbourhood_group_Queens + `room_type_Entire home/apt` +
##   `room_type_Private room`
##
##                                     Df Sum of Sq      RSS      AIC
## <none>                           160442920 278730
## + `neighbourhood_group_Staten Island` 1    8492 160434428 278730
## + neighbourhood_group_Bronx          1    8492 160434428 278730
## - neighbourhood_group_Queens        1    155770 160598690 278760
## - `room_type_Private room`         1    228427 160671347 278775
## - neighbourhood_group_Brooklyn      1    753310 161196230 278882
## - minimum_nights                   1    910673 161353593 278914
## - number_of_reviews                 1    1728107 162171027 279080
## - availability_365                 1    3403693 163846613 279417
## - neighbourhood_group_Manhattan     1    4844264 165287184 279704
## - `room_type_Entire home/apt`      1    7100140 167543060 280149
step$anova # display results

```

```

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##   neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##   `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##   `room_type_Private room` + `room_type_Shared room`
##
## Final Model:
## price ~ minimum_nights + number_of_reviews + availability_365 +
##   neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan +
##   neighbourhood_group_Queens + `room_type_Entire home/apt` +
##   `room_type_Private room`
##
```

```

##                                     Step Df Deviance Resid. Df Resid. Dev
## 1                                         32799 160434428
## 2 - `room_type_Shared room`  0     0.000    32799 160434428
## 3 - `neighbourhood_group_Staten Island` 0     0.000    32799 160434428
## 4 - neighbourhood_group_Bronx   1 8491.967    32800 160442920
##          AIC
## 1 278730.4
## 2 278730.4
## 3 278730.4
## 4 278730.1

```

To compare the three models in the process above I built the three model and placed them side to side in anova tables. This is not necessary at all but it helped me understand the difference of one to the other. Note that the number shown in Sum of Sq is the difference between the models and we prefer a lower RSS.

```

# compare models
fit2 <- lm(price ~ minimum_nights + number_of_reviews + availability_365 +
neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan +
neighbourhood_group_Queens + `room_type_Entire home/apt` +
`room_type_Private room`, data = df_dummies)

```

I used bootstrap to perform 10 fold crossvalidation for R-squared.

```

# Assessing R2 shrinkage using 10-Fold Cross-Validation

fit <- lm(price ~ minimum_nights + number_of_reviews + availability_365 +
neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan +
neighbourhood_group_Queens + `room_type_Entire home/apt` +
`room_type_Private room`, data = df_dummies)

library(bootstrap)
# define functions
theta.fit <- function(x,y){lsfit(x,y)}
theta.predict <- function(fit,x){cbind(1,x)%*%fit$coef}

# matrix of predictors
X <- as.matrix(df_dummies[c("availability_365", "number_of_reviews", "minimum_nights", "neighbourhood_group_Brooklyn", "neighbourhood_group_Manhattan", "neighbourhood_group_Queens", "room_type_Entire home/apt", "room_type_Private room")])
y <- as.matrix(df_dummies[c("price")])

results <- crossval(X,y,theta.fit,theta.predict,ngroup=10)
cor(y, fit$fitted.values)**2 # raw R2

##          [,1]
## price 0.3953225
cor(y, results$cv.fit)**2 # cross-validated R2

##          [,1]
## price 0.39463

```

This is an interesing plot because we get to see which variables have relatively more importance across the board.

```

# Calculate Relative Importance for Each Predictor
library(relaimpo)

```

```

## Loading required package: boot
## Loading required package: survey
## Loading required package: grid
## Loading required package: Matrix
## Loading required package: survival
##
## Attaching package: 'survival'

## The following object is masked from 'package:boot':
##
##     aml

##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart

## Loading required package: mitools

## This is the global version of package relaimpo.

## If you are a non-US user, a version with the interesting additional metric pmvd is available
## from Ulrike Groempings web site at prof.beuth-hochschule.de/groemping.

calc.relimp(fit,type=c("lmg","last","first","pratt"),
            rela=TRUE)

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arithmetic
##   Use c() or as.vector() instead.

## Response variable: price
## Total response variance: 8087.55
## Analysis based on 32809 observations
##
## 8 Regressors:
## minimum_nights number_of_reviews availability_365 neighbourhood_group_Brooklyn neighbourhood_group_Manhattan
## Proportion of variance explained by model: 39.53%
## Metrics are normalized to sum to 100% (rela=TRUE).
##
## Relative importance metrics:
##
##          lmg      last     first     pratt
## minimum_nights  0.003814839 0.047618406 0.001554394 -0.005295351
## number_of_reviews 0.013451611 0.090361427 0.007062759  0.015535625
## availability_365  0.024286800 0.177976598 0.009834979  0.025974495
## neighbourhood_group_Brooklyn 0.045954412 0.039390044 0.041922152 -0.068926266
## neighbourhood_group_Manhattan 0.141141989 0.253303036 0.137839740  0.319722489
## neighbourhood_group_Queens    0.037057384 0.008145094 0.037121800 -0.021514939
## room_type_Entire home/apt    0.398892016 0.371261123 0.397844638  0.887447053
## room_type_Private room      0.335400948 0.011944273 0.366819539 -0.152943106
##
## Average coefficients for different model sizes:
##

```

```

##                                     1X          2Xs          3Xs
## minimum_nights                  0.19939979   0.07161355 -0.03583623
## number_of_reviews                -0.13046973  -0.12406253 -0.12104865
## availability_365                 0.06161936   0.06233508  0.06408123
## neighbourhood_group_Brooklyn    -33.10574584 -25.34762768 -17.91966246
## neighbourhood_group_Manhattan    59.42663866   57.25072475  55.97605486
## neighbourhood_group_Queens      -45.90632120 -41.10657299 -35.01886783
## room_type_Entire home/apt       100.25679744  99.93231871 100.49758649
## room_type_Private room         -96.50639005 -80.01382824 -63.27503155
##                                     4Xs          5Xs          6Xs
## minimum_nights                  -0.12492756  -0.19829950 -0.25853458
## number_of_reviews                -0.12117760  -0.12399748 -0.12912105
## availability_365                 0.06659512   0.06974736  0.07349463
## neighbourhood_group_Brooklyn    -10.45673156 -2.58185886  6.10336970
## neighbourhood_group_Manhattan    55.82517622  56.98654786  59.62926051
## neighbourhood_group_Queens      -27.92774924 -19.87494711 -10.64576398
## room_type_Entire home/apt       101.91317312 104.08818484 106.88288172
## room_type_Private room         -46.40397877 -29.50506565 -12.67430554
##                                     7Xs          8Xs
## minimum_nights                  -0.30739816  -0.34481781
## number_of_reviews                -0.13638762  -0.14567849
## availability_365                 0.07782111   0.08260801
## neighbourhood_group_Brooklyn    16.02554204  27.62968273
## neighbourhood_group_Manhattan    63.91176000  69.96989809
## neighbourhood_group_Queens      0.23843753  13.50563219
## room_type_Entire home/apt       110.10882217 113.52044997
## room_type_Private room         3.99858170  20.42513103

# Bootstrap Measures of Relative Importance (1000 samples)
boot <- boot.relimp(fit, b = 1000, type = c("lmg",
  "last", "first", "pratt"), rank = TRUE,
  diff = TRUE, rela = TRUE)

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.
```

```

##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

booteval.relimp(boot) # print result

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arith
##   Use c() or as.vector() instead.

## Response variable: price
## Total response variance: 8087.55
## Analysis based on 32809 observations
##
## 8 Regressors:
## minimum_nights number_of_reviews availability_365 neighbourhood_group_Brooklyn neighbourhood_group_Manhattan
## Proportion of variance explained by model: 39.53%
## Metrics are normalized to sum to 100% (rela=TRUE).
##
## Relative importance metrics:
##
##          lmg      last     first      pratt
## minimum_nights  0.003814839  0.047618406  0.001554394 -0.005295351
## number_of_reviews 0.013451611  0.090361427  0.007062759  0.015535625
## availability_365  0.024286800  0.177976598  0.009834979  0.025974495
## neighbourhood_group_Brooklyn  0.045954412  0.039390044  0.041922152 -0.068926266
## neighbourhood_group_Manhattan  0.141141989  0.253303036  0.137839740  0.319722489
## neighbourhood_group_Queens    0.037057384  0.008145094  0.037121800 -0.021514939
## room_type_Entire home/apt    0.398892016  0.371261123  0.397844638  0.887447053
## room_type_Private room     0.335400948  0.011944273  0.366819539 -0.152943106
##
## Average coefficients for different model sizes:
##
##          1X       2Xs       3Xs
## minimum_nights  0.19939979  0.07161355 -0.03583623
## number_of_reviews -0.13046973 -0.12406253 -0.12104865
## availability_365  0.06161936  0.06233508  0.06408123
## neighbourhood_group_Brooklyn -33.10574584 -25.34762768 -17.91966246
## neighbourhood_group_Manhattan  59.42663866  57.25072475  55.97605486
## neighbourhood_group_Queens    -45.90632120 -41.10657299 -35.01886783
## room_type_Entire home/apt    100.25679744  99.93231871 100.49758649
## room_type_Private room      -96.50639005 -80.01382824 -63.27503155
##          4Xs       5Xs       6Xs

```

```

## minimum_nights           -0.12492756 -0.19829950 -0.25853458
## number_of_reviews        -0.12117760 -0.12399748 -0.12912105
## availability_365          0.06659512  0.06974736  0.07349463
## neighbourhood_group_Brooklyn -10.45673156 -2.58185886  6.10336970
## neighbourhood_group_Manhattan 55.82517622 56.98654786 59.62926051
## neighbourhood_group_Queens   -27.92774924 -19.87494711 -10.64576398
## room_type_Entire home/apt    101.91317312 104.08818484 106.88288172
## room_type_Private room      -46.40397877 -29.50506565 -12.67430554
##                                     7Xs          8Xs
## minimum_nights            -0.30739816 -0.34481781
## number_of_reviews         -0.13638762 -0.14567849
## availability_365          0.07782111  0.08260801
## neighbourhood_group_Brooklyn 16.02554204 27.62968273
## neighbourhood_group_Manhattan 63.91176000 69.96989809
## neighbourhood_group_Queens    0.23843753 13.50563219
## room_type_Entire home/apt    110.10882217 113.52044997
## room_type_Private room       3.99858170 20.42513103
##
##
## Confidence interval information ( 1000 bootstrap replicates, bty= perc ):
## Relative Contributions with confidence intervals:
##
##                                     Lower   Upper
##                                     percentage 0.95    0.95    0.95
## ## minimum_nights.lmg          0.0038  -----H  0.0023  0.0065
## ## number_of_reviews.lmg        0.0135  -----G_  0.0110  0.0161
## ## availability_365.lmg        0.0243  -----F_  0.0202  0.0294
## ## neighbourhood_group_Brooklyn.lmg 0.0460  ___D___  0.0428  0.0495
## ## neighbourhood_group_Manhattan.lmg 0.1411  __C____  0.1340  0.1488
## ## neighbourhood_group_Queens.lmg   0.0371  ___E___  0.0339  0.0403
## ## room_type_Entire home/apt.lmg  0.3989  A_____  0.3903  0.4063
## ## room_type_Private room.lmg    0.3354  _B_____  0.3286  0.3417
##
## minimum_nights.last          0.0476  ___EF__  0.0227  0.0788
## number_of_reviews.last        0.0904  __D____  0.0757  0.1045
## availability_365.last        0.1780  __C____  0.1553  0.2051
## neighbourhood_group_Brooklyn.last 0.0394  ___EF__  0.0308  0.0484
## neighbourhood_group_Manhattan.last 0.2533  _B_____  0.2289  0.2812
## neighbourhood_group_Queens.last  0.0081  _____GH  0.0043  0.0129
## room_type_Entire home/apt.last 0.3713  A_____  0.3272  0.4067
## room_type_Private room.last   0.0119  _____GH  0.0061  0.0182
##
## minimum_nights.first          0.0016  -----H  0.0006  0.0031
## number_of_reviews.first        0.0071  -----FG_  0.0054  0.0088
## availability_365.first        0.0098  -----FG_  0.0076  0.0127
## neighbourhood_group_Brooklyn.first 0.0419  __DE__  0.0374  0.0462
## neighbourhood_group_Manhattan.first 0.1378  __C____  0.1312  0.1452
## neighbourhood_group_Queens.first  0.0371  __DE__  0.0335  0.0406
## room_type_Entire home/apt.first 0.3978  A_____  0.3913  0.4038
## room_type_Private room.first   0.3668  _B_____  0.3610  0.3720
##
## minimum_nights.pratt         -0.0053  ___E__  -0.0067 -0.0038
## number_of_reviews.pratt        0.0155  __D____  0.0127  0.0184
## availability_365.pratt        0.0260  __C____  0.0215  0.0316

```

```

## neighbourhood_group_Brooklyn.pratt -0.0689      _____G_ -0.0776 -0.0600
## neighbourhood_group_Manhattan.pratt 0.3197      _B_____ 0.2998  0.3412
## neighbourhood_group_Queens.pratt    -0.0215      _____F__ -0.0273 -0.0157
## room_type_Entire home/apt.pratt   0.8874      A_______ 0.8364  0.9268
## room_type_Private room.pratt     -0.1529      _____H -0.1912 -0.1075
##
## Letters indicate the ranks covered by bootstrap CIs.
## (Rank bootstrap confidence intervals always obtained by percentile method)
## CAUTION: Bootstrap confidence intervals can be somewhat liberal.
##
##
## Differences between Relative Contributions:
##                                     Lower   Upper
##                                     difference
##                                     -0.0096
##                                     -0.0205
##                                     -0.0421
##                                     -0.1373
##                                     -0.0332
##                                     -0.3951
##                                     -0.3316
##                                     -0.0108
##                                     -0.0325
##                                     -0.1277
##                                     -0.0236
##                                     -0.3854
##                                     -0.3219
##                                     -0.0217
##                                     -0.1169
##                                     -0.0128
##                                     -0.3746
##                                     -0.3111
##                                     -0.0952
##                                     0.0089
##                                     -0.3529
##                                     -0.2894
##                                     0.1041
##                                     -0.2578
##                                     -0.1943
##                                     -0.3618
##                                     -0.2983
##                                     0.0635
##
##                                     -0.0427
##                                     -0.1304
##                                     0.0082
##                                     -0.2057
##                                     0.0395
##                                     -0.3236
##                                     0.0357
##                                     -0.0876
##                                     0.0510
##                                     -0.1629

```

```

## number_of_reviews-neighbourhood_group_Queens.last           0.0822
## number_of_reviews-room_type_Entire home/apt.last          -0.2809
## number_of_reviews-room_type_Private room.last            0.0784
## availability_365-neighbourhood_group_Brooklyn.last       0.1386
## availability_365-neighbourhood_group_Manhattan.last      -0.0753
## availability_365-neighbourhood_group_Queens.last          0.1698
## availability_365-room_type_Entire home/apt.last         -0.1933
## availability_365-room_type_Private room.last            0.1660
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.last -0.2139
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.last  0.0312
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.last -0.3319
## neighbourhood_group_Brooklyn-room_type_Private room.last    0.0274
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.last 0.2452
## neighbourhood_group_Manhattan-room_type_Entire home/apt.last -0.1180
## neighbourhood_group_Manhattan-room_type_Private room.last    0.2414
## neighbourhood_group_Queens-room_type_Entire home/apt.last   -0.3631
## neighbourhood_group_Queens-room_type_Private room.last       -0.0038
## room_type_Entire home/apt-room_type_Private room.last      0.3593
##
## minimum_nights-number_of_reviews.first                   -0.0055
## minimum_nights-availability_365.first                 -0.0083
## minimum_nights-neighbourhood_group_Brooklyn.first     -0.0404
## minimum_nights-neighbourhood_group_Manhattan.first    -0.1363
## minimum_nights-neighbourhood_group_Queens.first       -0.0356
## minimum_nights-room_type_Entire home/apt.first       -0.3963
## minimum_nights-room_type_Private room.first          -0.3653
## number_of_reviews-availability_365.first             -0.0028
## number_of_reviews-neighbourhood_group_Brooklyn.first  -0.0349
## number_of_reviews-neighbourhood_group_Manhattan.first -0.1308
## number_of_reviews-neighbourhood_group_Queens.first    -0.0301
## number_of_reviews-room_type_Entire home/apt.first    -0.3908
## number_of_reviews-room_type_Private room.first        -0.3598
## availability_365-neighbourhood_group_Brooklyn.first  -0.0321
## availability_365-neighbourhood_group_Manhattan.first -0.1280
## availability_365-neighbourhood_group_Queens.first    -0.0273
## availability_365-room_type_Entire home/apt.first     -0.3880
## availability_365-room_type_Private room.first        -0.3570
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.first -0.0959
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.first  0.0048
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.first -0.3559
## neighbourhood_group_Brooklyn-room_type_Private room.first   -0.3249
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.first 0.1007
## neighbourhood_group_Manhattan-room_type_Entire home/apt.first -0.2600
## neighbourhood_group_Manhattan-room_type_Private room.first   -0.2290
## neighbourhood_group_Queens-room_type_Entire home/apt.first  -0.3607
## neighbourhood_group_Queens-room_type_Private room.first     -0.3297
## room_type_Entire home/apt-room_type_Private room.first    0.0310
##
## minimum_nights-number_of_reviews.pratt              -0.0208
## minimum_nights-availability_365.pratt            -0.0313
## minimum_nights-neighbourhood_group_Brooklyn.pratt 0.0636
## minimum_nights-neighbourhood_group_Manhattan.pratt -0.3250
## minimum_nights-neighbourhood_group_Queens.pratt    0.0162
## minimum_nights-room_type_Entire home/apt.pratt    -0.8927

```

## minimum_nights-room_type_Private room.pratt	0.1476
## number_of_reviews-availability_365.pratt	-0.0104
## number_of_reviews-neighbourhood_group_Brooklyn.pratt	0.0845
## number_of_reviews-neighbourhood_group_Manhattan.pratt	-0.3042
## number_of_reviews-neighbourhood_group_Queens.pratt	0.0371
## number_of_reviews-room_type_Entire home/apt.pratt	-0.8719
## number_of_reviews-room_type_Private room.pratt	0.1685
## availability_365-neighbourhood_group_Brooklyn.pratt	0.0949
## availability_365-neighbourhood_group_Manhattan.pratt	-0.2937
## availability_365-neighbourhood_group_Queens.pratt	0.0475
## availability_365-room_type_Entire home/apt.pratt	-0.8615
## availability_365-room_type_Private room.pratt	0.1789
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.pratt	-0.3886
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.pratt	-0.0474
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.pratt	-0.9564
## neighbourhood_group_Brooklyn-room_type_Private room.pratt	0.0840
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.pratt	0.3412
## neighbourhood_group_Manhattan-room_type_Entire home/apt.pratt	-0.5677
## neighbourhood_group_Manhattan-room_type_Private room.pratt	0.4727
## neighbourhood_group_Queens-room_type_Entire home/apt.pratt	-0.9090
## neighbourhood_group_Queens-room_type_Private room.pratt	0.1314
## room_type_Entire home/apt-room_type_Private room.pratt	1.0404
##	0.95 0.95
## minimum_nights-number_of_reviews.lmg	* -0.0124
## minimum_nights-availability_365.lmg	* -0.0254
## minimum_nights-neighbourhood_group_Brooklyn.lmg	* -0.0460
## minimum_nights-neighbourhood_group_Manhattan.lmg	* -0.1452
## minimum_nights-neighbourhood_group_Queens.lmg	* -0.0368
## minimum_nights-room_type_Entire home/apt.lmg	* -0.4029
## minimum_nights-room_type_Private room.lmg	* -0.3380
## number_of_reviews-availability_365.lmg	* -0.0165
## number_of_reviews-neighbourhood_group_Brooklyn.lmg	* -0.0368
## number_of_reviews-neighbourhood_group_Manhattan.lmg	* -0.1361
## number_of_reviews-neighbourhood_group_Queens.lmg	* -0.0276
## number_of_reviews-room_type_Entire home/apt.lmg	* -0.3936
## number_of_reviews-room_type_Private room.lmg	* -0.3290
## availability_365-neighbourhood_group_Brooklyn.lmg	* -0.0274
## availability_365-neighbourhood_group_Manhattan.lmg	* -0.1266
## availability_365-neighbourhood_group_Queens.lmg	* -0.0180
## availability_365-room_type_Entire home/apt.lmg	* -0.3839
## availability_365-room_type_Private room.lmg	* -0.3190
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.lmg	* -0.1005
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.lmg	* 0.0042
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.lmg	* -0.3627
## neighbourhood_group_Brooklyn-room_type_Private room.lmg	* -0.2978
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.lmg	* 0.0972
## neighbourhood_group_Manhattan-room_type_Entire home/apt.lmg	* -0.2712
## neighbourhood_group_Manhattan-room_type_Private room.lmg	* -0.2071
## neighbourhood_group_Queens-room_type_Entire home/apt.lmg	* -0.3708
## neighbourhood_group_Queens-room_type_Private room.lmg	* -0.3066
## room_type_Entire home/apt-room_type_Private room.lmg	* 0.0560
##	
## minimum_nights-number_of_reviews.last	* -0.0716
## minimum_nights-availability_365.last	* -0.1668

```

## minimum_nights-neighbourhood_group_Brooklyn.last           -0.0201
## minimum_nights-neighbourhood_group_Manhattan.last        * -0.2487
## minimum_nights-neighbourhood_group_Queens.last          *  0.0128
## minimum_nights-room_type_Entire home/apt.last            * -0.3784
## minimum_nights-room_type_Private room.last                *  0.0090
## number_of_reviews-availability_365.last                  * -0.1155
## number_of_reviews-neighbourhood_group_Brooklyn.last      *  0.0326
## number_of_reviews-neighbourhood_group_Manhattan.last     * -0.1953
## number_of_reviews-neighbourhood_group_Queens.last         *  0.0666
## number_of_reviews-room_type_Entire home/apt.last         * -0.3233
## number_of_reviews-room_type_Private room.last             *  0.0620
## availability_365-neighbourhood_group_Brooklyn.last       *  0.1097
## availability_365-neighbourhood_group_Manhattan.last      * -0.1172
## availability_365-neighbourhood_group_Queens.last          *  0.1447
## availability_365-room_type_Entire home/apt.last          * -0.2420
## availability_365-room_type_Private room.last              *  0.1409
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.last * -0.2359
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.last *  0.0254
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.last * -0.3724
## neighbourhood_group_Brooklyn-room_type_Private room.last    *  0.0165
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.last *  0.2224
## neighbourhood_group_Manhattan-room_type_Entire home/apt.last * -0.1695
## neighbourhood_group_Manhattan-room_type_Private room.last   *  0.2135
## neighbourhood_group_Queens-room_type_Entire home/apt.last    * -0.4002
## neighbourhood_group_Queens-room_type_Private room.last       -0.0116
## room_type_Entire home/apt-room_type_Private room.last      *  0.3188
##
## minimum_nights-number_of_reviews.first                     * -0.0075
## minimum_nights-availability_365.first                   * -0.0113
## minimum_nights-neighbourhood_group_Brooklyn.first       * -0.0447
## minimum_nights-neighbourhood_group_Manhattan.first      * -0.1438
## minimum_nights-neighbourhood_group_Queens.first         * -0.0393
## minimum_nights-room_type_Entire home/apt.first          * -0.4022
## minimum_nights-room_type_Private room.first              * -0.3710
## number_of_reviews-availability_365.first                 -0.0063
## number_of_reviews-neighbourhood_group_Brooklyn.first     * -0.0394
## number_of_reviews-neighbourhood_group_Manhattan.first    * -0.1382
## number_of_reviews-neighbourhood_group_Queens.first       * -0.0340
## number_of_reviews-room_type_Entire home/apt.first        * -0.3970
## number_of_reviews-room_type_Private room.first            * -0.3656
## availability_365-neighbourhood_group_Brooklyn.first      * -0.0373
## availability_365-neighbourhood_group_Manhattan.first     * -0.1361
## availability_365-neighbourhood_group_Queens.first        * -0.0316
## availability_365-room_type_Entire home/apt.first         * -0.3943
## availability_365-room_type_Private room.first            * -0.3632
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.first * -0.1007
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.first -0.0017
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.first * -0.3654
## neighbourhood_group_Brooklyn-room_type_Private room.first   * -0.3334
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.first *  0.0937
## neighbourhood_group_Manhattan-room_type_Entire home/apt.first * -0.2719
## neighbourhood_group_Manhattan-room_type_Private room.first  * -0.2408
## neighbourhood_group_Queens-room_type_Entire home/apt.first   * -0.3684
## neighbourhood_group_Queens-room_type_Private room.first     * -0.3370

```

## room_type_Entire home/apt-room_type_Private room.first	*	0.0269
##		
## minimum_nights-number_of_reviews.pratt	*	-0.0243
## minimum_nights-availability_365.pratt	*	-0.0369
## minimum_nights-neighbourhood_group_Brooklyn.pratt	*	0.0547
## minimum_nights-neighbourhood_group_Manhattan.pratt	*	-0.3462
## minimum_nights-neighbourhood_group_Queens.pratt	*	0.0105
## minimum_nights-room_type_Entire home/apt.pratt	*	-0.9318
## minimum_nights-room_type_Private room.pratt	*	0.1020
## number_of_reviews-availability_365.pratt	*	-0.0169
## number_of_reviews-neighbourhood_group_Brooklyn.pratt	*	0.0757
## number_of_reviews-neighbourhood_group_Manhattan.pratt	*	-0.3260
## number_of_reviews-neighbourhood_group_Queens.pratt	*	0.0300
## number_of_reviews-room_type_Entire home/apt.pratt	*	-0.9113
## number_of_reviews-room_type_Private room.pratt	*	0.1232
## availability_365-neighbourhood_group_Brooklyn.pratt	*	0.0852
## availability_365-neighbourhood_group_Manhattan.pratt	*	-0.3152
## availability_365-neighbourhood_group_Queens.pratt	*	0.0399
## availability_365-room_type_Entire home/apt.pratt	*	-0.9009
## availability_365-room_type_Private room.pratt	*	0.1334
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.pratt	*	-0.4172
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.pratt	*	-0.0525
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.pratt	*	-0.9968
## neighbourhood_group_Brooklyn-room_type_Private room.pratt	*	0.0370
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.pratt	*	0.3159
## neighbourhood_group_Manhattan-room_type_Entire home/apt.pratt	*	-0.6147
## neighbourhood_group_Manhattan-room_type_Private room.pratt	*	0.4250
## neighbourhood_group_Queens-room_type_Entire home/apt.pratt	*	-0.9484
## neighbourhood_group_Queens-room_type_Private room.pratt	*	0.0865
## room_type_Entire home/apt-room_type_Private room.pratt	*	0.9474
##		0.95
## minimum_nights-number_of_reviews.lmg		-0.0062
## minimum_nights-availability_365.lmg		-0.0157
## minimum_nights-neighbourhood_group_Brooklyn.lmg		-0.0377
## minimum_nights-neighbourhood_group_Manhattan.lmg		-0.1297
## minimum_nights-neighbourhood_group_Queens.lmg		-0.0294
## minimum_nights-room_type_Entire home/apt.lmg		-0.3850
## minimum_nights-room_type_Private room.lmg		-0.3241
## number_of_reviews-availability_365.lmg		-0.0061
## number_of_reviews-neighbourhood_group_Brooklyn.lmg		-0.0283
## number_of_reviews-neighbourhood_group_Manhattan.lmg		-0.1201
## number_of_reviews-neighbourhood_group_Queens.lmg		-0.0197
## number_of_reviews-room_type_Entire home/apt.lmg		-0.3758
## number_of_reviews-room_type_Private room.lmg		-0.3144
## availability_365-neighbourhood_group_Brooklyn.lmg		-0.0159
## availability_365-neighbourhood_group_Manhattan.lmg		-0.1080
## availability_365-neighbourhood_group_Queens.lmg		-0.0068
## availability_365-room_type_Entire home/apt.lmg		-0.3632
## availability_365-room_type_Private room.lmg		-0.3017
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.lmg		-0.0900
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.lmg		0.0133
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.lmg		-0.3418
## neighbourhood_group_Brooklyn-room_type_Private room.lmg		-0.2804
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.lmg		0.1114

```

## neighbourhood_group_Manhattan-room_type_Entire home/apt.lmg      -0.2419
## neighbourhood_group_Manhattan-room_type_Private room.lmg        -0.1813
## neighbourhood_group_Queens-room_type_Entire home/apt.lmg        -0.3517
## neighbourhood_group_Queens-room_type_Private room.lmg        -0.2893
## room_type_Entire home/apt-room_type_Private room.lmg         0.0703
##
## minimum_nights-number_of_reviews.last                          -0.0106
## minimum_nights-availability_365.last                         -0.0936
## minimum_nights-neighbourhood_group_Brooklyn.last            0.0415
## minimum_nights-neighbourhood_group_Manhattan.last          -0.1605
## minimum_nights-neighbourhood_group_Queens.last             0.0718
## minimum_nights-room_type_Entire home/apt.last              -0.2598
## minimum_nights-room_type_Private room.last                  0.0690
## number_of_reviews-availability_365.last                      -0.0625
## number_of_reviews-neighbourhood_group_Brooklyn.last          0.0693
## number_of_reviews-neighbourhood_group_Manhattan.last         -0.1321
## number_of_reviews-neighbourhood_group_Queens.last           0.0978
## number_of_reviews-room_type_Entire home/apt.last            -0.2308
## number_of_reviews-room_type_Private room.last                0.0951
## availability_365-neighbourhood_group_Brooklyn.last          0.1687
## availability_365-neighbourhood_group_Manhattan.last         -0.0350
## availability_365-neighbourhood_group_Queens.last            0.1983
## availability_365-room_type_Entire home/apt.last             -0.1315
## availability_365-room_type_Private room.last                 0.1964
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.last -0.1942
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.last   0.0373
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.last    -0.2845
## neighbourhood_group_Brooklyn-room_type_Private room.last       0.0397
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.last 0.2713
## neighbourhood_group_Manhattan-room_type_Entire home/apt.last   -0.0622
## neighbourhood_group_Manhattan-room_type_Private room.last      0.2727
## neighbourhood_group_Queens-room_type_Entire home/apt.last     -0.3180
## neighbourhood_group_Queens-room_type_Private room.last         0.0038
## room_type_Entire home/apt-room_type_Private room.last         0.3920
##
## minimum_nights-number_of_reviews.first                        -0.0034
## minimum_nights-availability_365.first                       -0.0056
## minimum_nights-neighbourhood_group_Brooklyn.first          -0.0359
## minimum_nights-neighbourhood_group_Manhattan.first         -0.1296
## minimum_nights-neighbourhood_group_Queens.first            -0.0315
## minimum_nights-room_type_Entire home/apt.first             -0.3892
## minimum_nights-room_type_Private room.first                 -0.3587
## number_of_reviews-availability_365.first                     0.0002
## number_of_reviews-neighbourhood_group_Brooklyn.first        -0.0303
## number_of_reviews-neighbourhood_group_Manhattan.first       -0.1242
## number_of_reviews-neighbourhood_group_Queens.first          -0.0261
## number_of_reviews-room_type_Entire home/apt.first           -0.3838
## number_of_reviews-room_type_Private room.first              -0.3536
## availability_365-neighbourhood_group_Brooklyn.first        -0.0272
## availability_365-neighbourhood_group_Manhattan.first        -0.1209
## availability_365-neighbourhood_group_Queens.first           -0.0226
## availability_365-room_type_Entire home/apt.first            -0.3805
## availability_365-room_type_Private room.first                -0.3503
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.first -0.0913

```

```

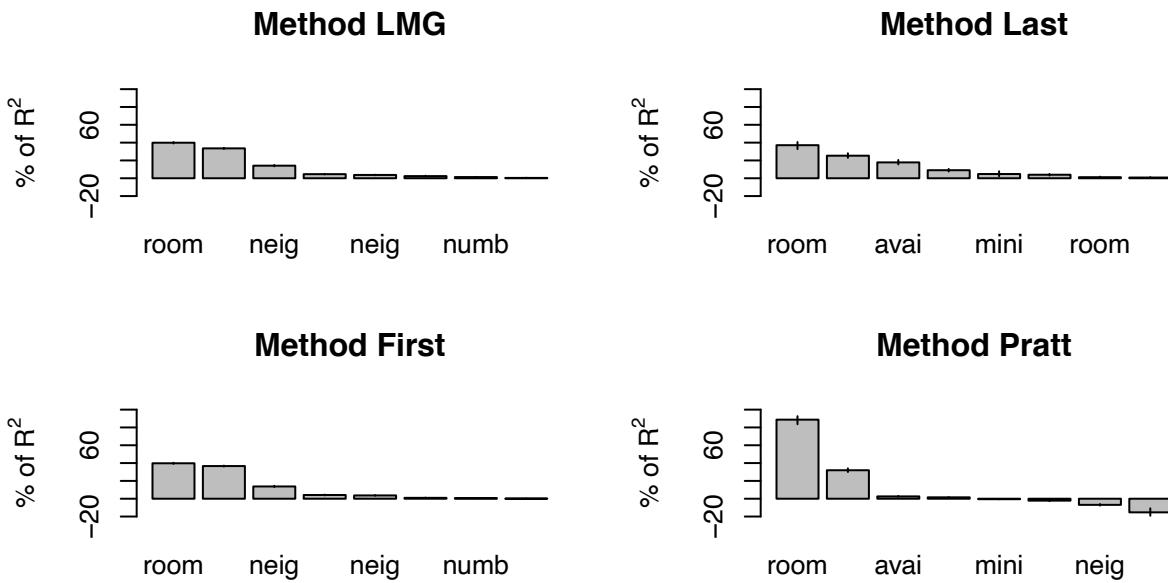
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.first      0.0111
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.first        -0.3456
## neighbourhood_group_Brooklyn-room_type_Private room.first           -0.3156
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.first     0.1080
## neighbourhood_group_Manhattan-room_type_Entire home/apt.first       -0.2464
## neighbourhood_group_Manhattan-room_type_Private room.first          -0.2160
## neighbourhood_group_Queens-room_type_Entire home/apt.first          -0.3526
## neighbourhood_group_Queens-room_type_Private room.first             -0.3220
## room_type_Entire home/apt-room_type_Private room.first              0.0346
##
## minimum_nights-number_of_reviews.pratt                            -0.0175
## minimum_nights-availability_365.pratt                           -0.0262
## minimum_nights-neighbourhood_group_Brooklyn.pratt                0.0725
## minimum_nights-neighbourhood_group_Manhattan.pratt               -0.3047
## minimum_nights-neighbourhood_group_Queens.pratt                  0.0221
## minimum_nights-room_type_Entire home/apt.pratt                   -0.8423
## minimum_nights-room_type_Private room.pratt                      0.1859
## number_of_reviews-availability_365.pratt                         -0.0049
## number_of_reviews-neighbourhood_group_Brooklyn.pratt              0.0934
## number_of_reviews-neighbourhood_group_Manhattan.pratt             -0.2844
## number_of_reviews-neighbourhood_group_Queens.pratt                0.0436
## number_of_reviews-room_type_Entire home/apt.pratt                 -0.8203
## number_of_reviews-room_type_Private room.pratt                    0.2063
## availability_365-neighbourhood_group_Brooklyn.pratt              0.1052
## availability_365-neighbourhood_group_Manhattan.pratt             -0.2733
## availability_365-neighbourhood_group_Queens.pratt                 0.0551
## availability_365-room_type_Entire home/apt.pratt                  -0.8096
## availability_365-room_type_Private room.pratt                     0.2175
## neighbourhood_group_Brooklyn-neighbourhood_group_Manhattan.pratt -0.3597
## neighbourhood_group_Brooklyn-neighbourhood_group_Queens.pratt      -0.0431
## neighbourhood_group_Brooklyn-room_type_Entire home/apt.pratt       -0.9072
## neighbourhood_group_Brooklyn-room_type_Private room.pratt          0.1234
## neighbourhood_group_Manhattan-neighbourhood_group_Queens.pratt    0.3666
## neighbourhood_group_Manhattan-room_type_Entire home/apt.pratt      -0.5087
## neighbourhood_group_Manhattan-room_type_Private room.pratt          0.5144
## neighbourhood_group_Queens-room_type_Entire home/apt.pratt          -0.8572
## neighbourhood_group_Queens-room_type_Private room.pratt             0.1712
## room_type_Entire home/apt-room_type_Private room.pratt            1.1174
##
## * indicates that CI for difference does not include 0.
## CAUTION: Bootstrap confidence intervals can be somewhat liberal.

plot(booteval.relimp(boot,sort=TRUE)) # plot result

## Warning in rev(variances[[p]]) - variances[[p + 1]]: Recycling array of length 1 in vector-array arithmetic
##   Use c() or as.vector() instead.

```

## Relative importances for price with 95% bootstrap confidence intervals



$R^2 = 39.53\%$ , metrics are normalized to sum 100%.

Below are the individual detailed anova tables of all three model, but note that we have chosen the model: `lm(formula = price ~ minimum_nights + number_of_reviews + availability_365 + neighbourhood_group_Bronx + neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan + neighbourhood_group_Queens + neighbourhood_group_Staten Island + room_type_Entire home/apt + room_type_Private room + room_type_Shared room, data = df_dummies)`

for our analysis. For the purpose of identification it goes by “fit” and “fit2”

```
summary(fit2)
```

```
##
## Call:
## lm(formula = price ~ minimum_nights + number_of_reviews + availability_365 +
##     neighbourhood_group_Brooklyn + neighbourhood_group_Manhattan +
##     neighbourhood_group_Queens + `room_type_Entire home/apt` +
##     `room_type_Private room`, data = df_dummies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -177.29   -40.54  -11.57   21.90  515.73
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.736276  3.644207  5.416 6.14e-08 ***
## minimum_nights -0.344818  0.025272 -13.645 < 2e-16 ***
## number_of_reviews -0.145678  0.007751 -18.796 < 2e-16 ***
## availability_365    0.082608  0.003132  26.379 < 2e-16 ***
```

```

## neighbourhood_group_Brooklyn 27.629683 2.226447 12.410 < 2e-16 ***
## neighbourhood_group_Manhattan 69.969898 2.223416 31.470 < 2e-16 ***
## neighbourhood_group_Queens 13.505632 2.393297 5.643 1.68e-08 ***
## `room_type_Entire home/apt` 113.520450 2.979641 38.099 < 2e-16 ***
## `room_type_Private room` 20.425131 2.988922 6.834 8.42e-12 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 69.94 on 32800 degrees of freedom
## Multiple R-squared: 0.3953, Adjusted R-squared: 0.3952
## F-statistic: 2680 on 8 and 32800 DF, p-value: < 2.2e-16

```

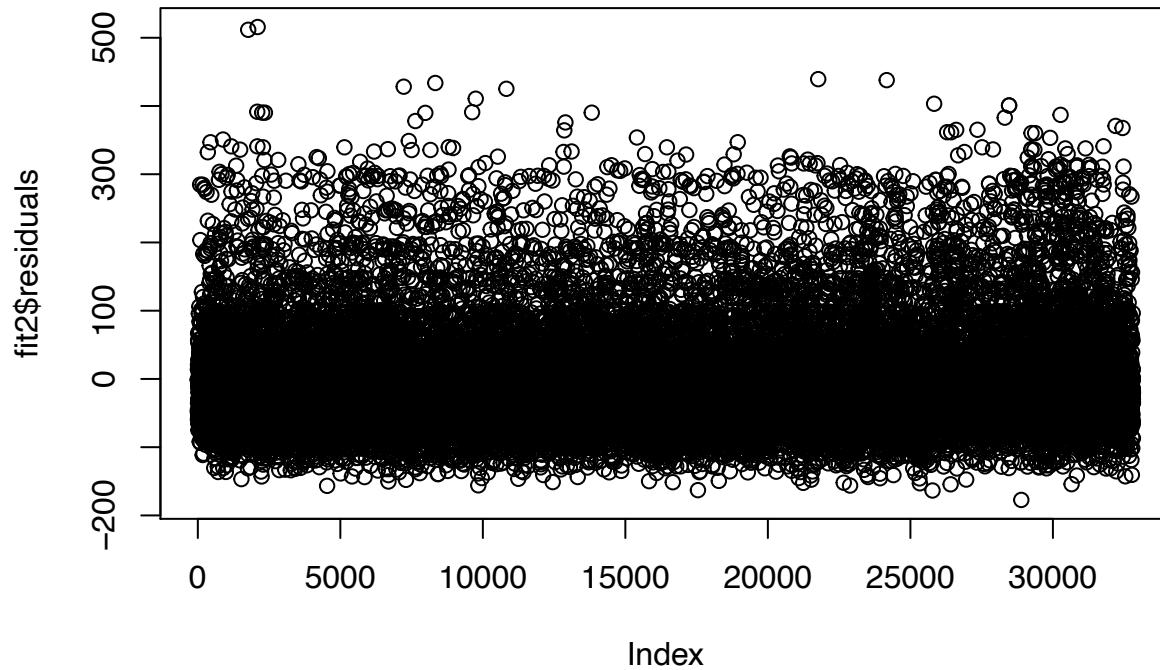
The following are the confidence intervals for the estimates of our regression model. Note that none of them contain zero which is good for us. We have chosen good model.

```
confint(fit2, level=0.95)
```

	2.5 %	97.5 %
## (Intercept)	12.5934982	26.87905306
## minimum_nights	-0.3943510	-0.29528466
## number_of_reviews	-0.1608699	-0.13048709
## availability_365	0.0764699	0.08874612
## neighbourhood_group_Brooklyn	23.2657664	31.99359908
## neighbourhood_group_Manhattan	65.6119225	74.32787370
## neighbourhood_group_Queens	8.8146841	18.19658026
## `room_type_Entire home/apt`	107.6802446	119.36065533
## `room_type_Private room`	14.5667348	26.28352723

The following two plots don't look great which is not encouraging but they are not the worst?

```
plot(fit2$residuals)
```

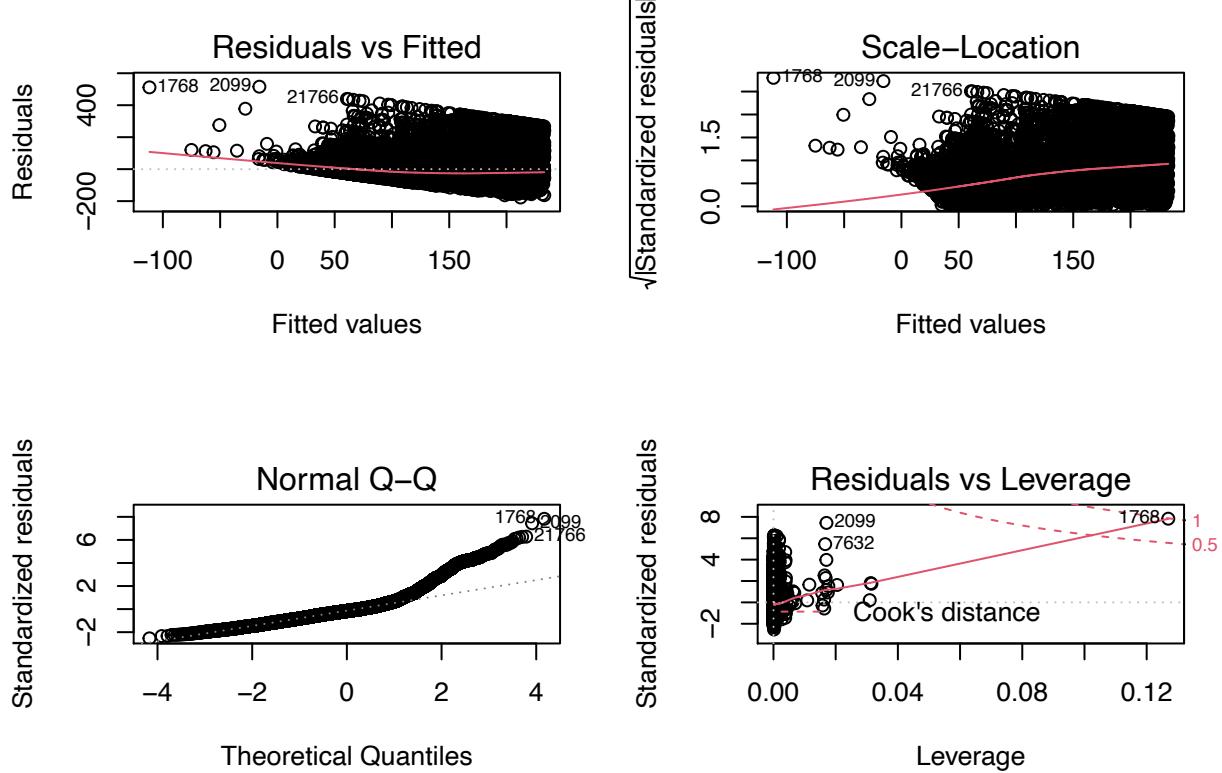


```

# diagnostic plots
layout(matrix(c(1,2,3,4), 2, 2)) # optional 4 graphs/page

```

```
plot(fit2)
```



In the following section of my code I perform hypothesis testing for the model parameters through multivariable statistical analysis.

Parameter estimation

```
# least square estimates
beta_hat <- solve(t(Z)%*%Z)%*%t(Z)%*%Y
beta_hat

##                                     price
## minimum_nights           -0.34481781
## number_of_reviews         -0.14567849
## availability_365          0.08260801
## neighbourhood_group_Brooklyn 27.62968273
## neighbourhood_group_Manhattan 69.96989809
## neighbourhood_group_Queens   13.50563219
## room_type_Entire home/apt 113.52044997
## room_type_Private room     20.42513103

# R^2 statistic
R_square <- 1 - sum((Y - Z%*%beta_hat)^2)/sum((Y-mean(Y))^2)
R_square

## [1] 0.3953225

# sigma_hat_square
sigma_hat_square <- sum((Y - Z%*%beta_hat)^2)/(n-r-1)
```

```

sigma_hat_square

## [1] 4891.552

# estimated covariance of hat{beta}
#not to use but I printed to confirm
cov_est = sigma_hat_square * solve(t(Z) %*% Z)
cov_est

##                                     minimum_nights number_of_reviews
## 13.2802423999    7.894186e-04     -6.274183e-04
## minimum_nights      0.0007894186    6.386513e-04     2.679952e-05
## number_of_reviews   -0.0006274183    2.679952e-05     6.007130e-05
## availability_365    -0.0017868664   -1.294307e-05    -4.610716e-06
## neighbourhood_group_Brooklyn -4.5731372442  -1.522766e-03    -2.903183e-04
## neighbourhood_group_Manhattan -4.6659203735  -2.744881e-03    -4.362309e-05
## neighbourhood_group_Queens   -4.5710168036  -5.224631e-04    -1.943930e-04
## room_type_Entire home/apt   -8.5450229216  -2.815935e-03    -5.991664e-04
## room_type_Private room     -8.6118183673  -1.464337e-03    -6.010012e-04
##                                     availability_365 neighbourhood_group_Brooklyn
##                                         -1.786866e-03          -4.5731372442
## minimum_nights                  -1.294307e-05          -0.0015227656
## number_of_reviews                -4.610716e-06          -0.0002903183
## availability_365                 9.807102e-06          0.0006220928
## neighbourhood_group_Brooklyn     6.220928e-04          4.9570648324
## neighbourhood_group_Manhattan    5.401792e-04          4.5875634763
## neighbourhood_group_Queens       2.666893e-04          4.5619467394
## room_type_Entire home/apt       2.367404e-04          -0.0856271194
## room_type_Private room          2.492751e-04          -0.0568367915
##                                     neighbourhood_group_Manhattan
##                                         -4.665920e+00
## minimum_nights                  -2.744881e-03
## number_of_reviews                -4.362309e-05
## availability_365                 5.401792e-04
## neighbourhood_group_Brooklyn     4.587563e+00
## neighbourhood_group_Manhattan    4.943578e+00
## neighbourhood_group_Queens       4.556092e+00
## room_type_Entire home/apt       -1.810648e-02
## room_type_Private room          8.210834e-02
##                                     neighbourhood_group_Queens
##                                         -4.5710168036
## minimum_nights                  -0.0005224631
## number_of_reviews                -0.0001943930
## availability_365                 0.0002666893
## neighbourhood_group_Brooklyn     4.5619467394
## neighbourhood_group_Manhattan    4.5560918280
## neighbourhood_group_Queens       5.7278682009
## room_type_Entire home/apt       0.0016215945
## room_type_Private room          -0.0276402666
##                                     room_type_Entire home/apt room_type_Private room
##                                         -8.5450229216     -8.6118183673
## minimum_nights                  -0.0028159347     -0.0014643371
## number_of_reviews                -0.0005991664     -0.0006010012
## availability_365                 0.0002367404     0.0002492751

```

```

## neighbourhood_group_Brooklyn           -0.0856271194
## neighbourhood_group_Manhattan        -0.0181064837
## neighbourhood_group_Queens          0.0016215945
## room_type_Entire home/apt          8.8782629315
## room_type_Private room            8.5918173409
Omega <- solve(t(Z) %*% Z)

```

## Hypothesis testing

We want to reject the hypothesis  $H_0: \beta_j = 0$ , for this we need a t-statistic ( $t_{\text{stat}}$ ) larger than a critical value ( $cval_t$ ). Unfortunately we fail to reject the hypothesis for the first two betas which is not enough evidence to eliminate them but it does not support the integrity of our model positively.

```

# t-test for single coefficient
# H_0: beta_j = 0, H_a: beta_j != 0

j <- 1
t_stat <- (beta_hat[j+1] - 0) / sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])

alpha <- 0.05
cval_t <- qt(1-alpha/2, n-r-1)
t_stat

## [1] -13.64451
cval_t

## [1] 1.960036

```

From t-test for  $\hat{\beta}_1$  we obtained a t-statistic of -13.6445052 and a 1.9600363

Therefore we can not reject the null hypothesis that the estimate  $\hat{\beta}_1$  is zero at level alpha = 0.05 .

```

# t-test for single coefficient
# H_0: beta_j = 0, H_a: beta_j != 0

j <- 2
t_stat <- (beta_hat[j+1] - 0) / sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])

alpha <- 0.05
cval_t <- qt(1-alpha/2, n-r-1)
t_stat

## [1] -18.79585
cval_t

## [1] 1.960036

```

From t-test for  $\hat{\beta}_2$  we obtained a t-statistic of -18.7958472 and a 1.9600363

Therefore we can not reject the null hypothesis that the estimate  $\hat{\beta}_2$  is zero at level alpha = 0.05 .

```

# t-test for single coefficient
# H_0: beta_j = 0, H_a: beta_j != 0

j <- 3
t_stat <- (beta_hat[j+1] - 0) / sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])

alpha <- 0.05

```

```
cval_t <- qt(1-alpha/2, n-r-1)
t_stat
```

```
## [1] 26.3786
```

```
cval_t
```

```
## [1] 1.960036
```

From t-test for  $\hat{\beta}_3$  we obtained a t-statistic of 26.3786045 and a 1.9600363

Therefore we can reject the null hypothesis that the estimate  $\hat{\beta}_3$  is zero at level alpha = 0.05 .

```
# t-test for single coefficient
```

```
# H_0: beta_j = 0, H_a: beta_j != 0
```

```
j <- 4
```

```
t_stat <- (beta_hat[j+1] - 0)/sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])
```

```
alpha <- 0.05
```

```
cval_t <- qt(1-alpha/2, n-r-1)
```

```
t_stat
```

```
## [1] 12.40977
```

```
cval_t
```

```
## [1] 1.960036
```

From t-test for  $\hat{\beta}_4$  we obtained a t-statistic of 12.4097662 and a 1.9600363

Therefore we can reject the null hypothesis that the estimate  $\hat{\beta}_4$  is zero at level alpha = 0.05 .

```
# t-test for single coefficient
```

```
# H_0: beta_j = 0, H_a: beta_j != 0
```

```
j <- 5
```

```
t_stat <- (beta_hat[j+1] - 0)/sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])
```

```
alpha <- 0.05
```

```
cval_t <- qt(1-alpha/2, n-r-1)
```

```
t_stat
```

```
## [1] 31.46955
```

```
cval_t
```

```
## [1] 1.960036
```

From t-test for  $\hat{\beta}_5$  we obtained a t-statistic of 31.4695522 and a 1.9600363

Therefore we can reject the null hypothesis that the estimate  $\hat{\beta}_5$  is zero at level alpha = 0.05 .

```
# t-test for single coefficient
```

```
# H_0: beta_j = 0, H_a: beta_j != 0
```

```
j <- 6
```

```
t_stat <- (beta_hat[j+1] - 0)/sqrt(sigma_hat_square * solve(t(Z) %*% Z)[j+1, j+1])
```

```
alpha <- 0.05
```

```
cval_t <- qt(1-alpha/2, n-r-1)
```

```
t_stat
```



```

cval_f <- qf(1-alpha, 2, n-r-1)
critical = cval_f * df_1 * sigma_hat_square
f_stat

##          price
## price 104893409
critical

## [1] 117241

```

with a level alpha = 0.05 we find that the f-statistic is  $1.0489341 \times 10^8$  while the critical value is  $1.1724096 \times 10^5$ . Therefore we can reject the null hypothesis that the estimates  $\hat{\beta}_1 = \hat{\beta}_2 = \dots = \hat{\beta}_8 = 0$  at level alpha = 0.05 .

```

#These are here again only for easy access to comparison
confint(fit2, level=0.95)

```

	2.5 %	97.5 %
## (Intercept)	12.5934982	26.87905306
## minimum_nights	-0.3943510	-0.29528466
## number_of_reviews	-0.1608699	-0.13048709
## availability_365	0.0764699	0.08874612
## neighbourhood_group_Brooklyn	23.2657664	31.99359908
## neighbourhood_group_Manhattan	65.6119225	74.32787370
## neighbourhood_group_Queens	8.8146841	18.19658026
## `room_type_Entire home/apt`	107.6802446	119.36065533
## `room_type_Private room`	14.5667348	26.28352723

---

```

df_bronx = df_dummies[df_dummies$neighbourhood_group_Bronx == 1,]
head(df_bronx)

```

```

## # A tibble: 6 x 12
##   price minimum_nights number_of_reviews availability_365 neighbourhood_group_B~
##   <dbl>        <dbl>            <dbl>           <dbl>           <dbl>
## 1     40            1             219            353            1
## 2     45            1             138            323            1
## 3     90            2              0             349            1
## 4     90           30              4             346            1
## 5     77            1             197            309            1
## 6     37            4             117            232            1
## # ... with 7 more variables: neighbourhood_group_Brooklyn <dbl>,
## #   neighbourhood_group_Manhattan <dbl>, neighbourhood_group_Queens <dbl>,
## #   neighbourhood_group_Staten Island <dbl>, room_type_Entire home/apt <dbl>,
## #   room_type_Private room <dbl>, room_type_Shared room <dbl>
fit_bronx <- lm(price ~ ., data = df_bronx)
summary(fit_bronx)

```

```

##
## Call:
## lm(formula = price ~ ., data = df_bronx)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -87.60 -26.79 -10.52  10.05 421.10 
## 
```

```

## Coefficients: (6 not defined because of singularities)
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 50.44787   11.71433   4.307 1.87e-05 ***
## minimum_nights              -0.26487    0.12539  -2.112   0.035 *
## number_of_reviews             -0.16684    0.04141  -4.029 6.15e-05 ***
## availability_365              0.06873    0.01479   4.646 3.96e-06 ***
## neighbourhood_group_Bronx          NA         NA       NA      NA
## neighbourhood_group_Brooklyn         NA         NA       NA      NA
## neighbourhood_group_Manhattan        NA         NA       NA      NA
## neighbourhood_group_Queens            NA         NA       NA      NA
## `neighbourhood_group_Staten Island` NA         NA       NA      NA
## `room_type_Entire home/apt`        67.70292   11.94263   5.669 2.01e-08 ***
## `room_type_Private room`           6.49534    11.81394   0.550   0.583
## `room_type_Shared room`            NA         NA       NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.67 on 790 degrees of freedom
## Multiple R-squared:  0.2637, Adjusted R-squared:  0.259
## F-statistic: 56.59 on 5 and 790 DF,  p-value: < 2.2e-16

# Stepwise Regression
library(MASS)
step <- stepAIC(fit_bronx, direction="both")

## Start:  AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##     neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##     neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##     `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##     `room_type_Private room` + `room_type_Shared room`
##
##
## Step:  AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##     neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##     neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##     `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##     `room_type_Private room`
##
##
## Step:  AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##     neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##     neighbourhood_group_Manhattan + `room_type_Entire home/apt` +
##     `room_type_Private room`
##
##
## Step:  AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##     neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##     neighbourhood_group_Manhattan + `room_type_Entire home/apt` +
##     `room_type_Private room`

```

```

## Step: AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##      `room_type_Entire home/apt` + `room_type_Private room`
##
##
## Step: AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      neighbourhood_group_Bronx + `room_type_Entire home/apt` +
##      `room_type_Private room`
##
##
## Step: AIC=6316.59
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      `room_type_Entire home/apt` + `room_type_Private room`
##
##                                     Df Sum of Sq    RSS     AIC
## - `room_type_Private room`     1      838 2192002 6314.9
## <none>                           2191164 6316.6
## - minimum_nights               1     12375 2203539 6319.1
## - number_of_reviews             1     45016 2236180 6330.8
## - availability_365              1     59870 2251034 6336.1
## - `room_type_Entire home/apt`   1     89138 2280302 6346.3
##
## Step: AIC=6314.9
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      `room_type_Entire home/apt`
##
##                                     Df Sum of Sq    RSS     AIC
## <none>                           2192002 6314.9
## + `room_type_Private room`      1      838 2191164 6316.6
## + `room_type_Shared room`       1      838 2191164 6316.6
## - minimum_nights                1     12173 2204176 6317.3
## - number_of_reviews              1     44263 2236266 6328.8
## - availability_365              1     60106 2252109 6334.4
## - `room_type_Entire home/apt`   1     718314 2910317 6538.5
step$anova # display results

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      neighbourhood_group_Bronx + neighbourhood_group_Brooklyn +
##      neighbourhood_group_Manhattan + neighbourhood_group_Queens +
##      `neighbourhood_group_Staten Island` + `room_type_Entire home/apt` +
##      `room_type_Private room` + `room_type_Shared room`
##
## Final Model:
## price ~ minimum_nights + number_of_reviews + availability_365 +
##      `room_type_Entire home/apt`
##
##
##                                     Step Df Deviance Resid. Df Resid. Dev

```

```

## 1
## 2      - `room_type_Shared room` 0 0.0000 790 2191164
## 3 - `neighbourhood_group_Staten Island` 0 0.0000 790 2191164
## 4      - neighbourhood_group_Queens 0 0.0000 790 2191164
## 5      - neighbourhood_group_Manhattan 0 0.0000 790 2191164
## 6      - neighbourhood_group_Brooklyn 0 0.0000 790 2191164
## 7      - neighbourhood_group_Bronx 0 0.0000 790 2191164
## 8      - `room_type_Private room` 1 838.4199 791 2192002
##          AIC
## 1 6316.594
## 2 6316.594
## 3 6316.594
## 4 6316.594
## 5 6316.594
## 6 6316.594
## 7 6316.594
## 8 6314.899

fit_bronx <- lm(price ~ minimum_nights + number_of_reviews + availability_365 +
  `room_type_Entire home/apt`, data = df_bronx)
summary(fit_bronx)

##
## Call:
## lm(formula = price ~ minimum_nights + number_of_reviews + availability_365 +
##     `room_type_Entire home/apt`, data = df_bronx)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -87.64 -26.80 -10.51  10.06 421.41 
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 56.56652   3.65557 15.474 < 2e-16 ***
## minimum_nights -0.26254   0.12527 -2.096  0.0364 *  
## number_of_reviews -0.16472   0.04121 -3.997 7.03e-05 ***
## availability_365   0.06886   0.01479  4.657 3.76e-06 ***
## `room_type_Entire home/apt` 61.48186   3.81876 16.100 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.64 on 791 degrees of freedom
## Multiple R-squared:  0.2634, Adjusted R-squared:  0.2597 
## F-statistic: 70.72 on 4 and 791 DF,  p-value: < 2.2e-16

```