airbnb in New York City

Carole Mattmann und Jonas Zuercher

13 Februar 2020

Packages used:

```
library(dplyr)
library(tidyverse)
library(geosphere)
library(ggplot2)
```

Summary

We are exploring a dataset of airbnb listings in New York City in 2019.

Analyses on the prices of the listings were run and models were created to predict prices of the listings.The best model to calculate the price that was found includes 5 variables:

- room type (factor with 3 levels, entire appartment being the highest and shared room the lowest)
- distance to timessquare (negative effect)
- availability (positive effect)
- neighbourhood group (factor with 5 levels, Manhattan being the highest and Bronx the lowest)
- minimum nights (negative effect)

The airbnb dataset was merged with a dataset concerning incidents (e.g. crimes) in the concerning neigbourhoods.

The airbnb dataset is vizualized on a map in the last chapter.

Data import and cleaning

airbnb dataset

The dataset was downloaded from: https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data

Import

```
AB_NYC <- read.csv("../01_data/AB_NYC_2019.csv", header=TRUE)
```

Overview of dataset

```
str(AB_NYC,width=80,strict.width="cut")
```

```
## 'data.frame':    48895 obs. of  16 variables:
##  $ id                            : int  2539 2595 3647 3831 5022 5099 5121 517..
##  $ name                          : Factor w/ 47906 levels "","'Fan'tastic",..:..
##  $ host_id                       : int  2787 2845 4632 4869 7192 7322 7356 896..
##  $ host_name                     : Factor w/ 11453 levels "","'Cil","-TheQuee"..
##  $ neighbourhood_group           : Factor w/ 5 levels "Bronx","Brooklyn",..: 2..
##  $ neighbourhood                 : Factor w/ 221 levels "Allerton","Arden Hei"..
##  $ latitude                      : num  40.6 40.8 40.8 40.7 40.8 ...
##  $ longitude                     : num  -74 -74 -73.9 -74 -73.9 ...
```

```
##  $ room_type                     : Factor w/ 3 levels "Entire home/apt",..: 2 ..
##  $ price                         : int  149 225 150 89 80 200 60 79 79 150 ...
##  $ minimum_nights                : int  1 1 3 1 10 3 45 2 2 1 ...
##  $ number_of_reviews             : int  9 45 0 270 9 74 49 430 118 160 ...
##  $ last_review                   : Factor w/ 1765 levels "","2011-03-28",..: 1..
##  $ reviews_per_month             : num  0.21 0.38 NA 4.64 0.1 0.59 0.4 3.47 0...
##  $ calculated_host_listings_count: int  6 2 1 1 1 1 1 1 1 4 ...
##  $ availability_365              : int  365 355 365 194 0 129 0 220 0 188 ...
```

Following changes are made to the dataset:

remove price 0

remove all listings with price 0

```
AB_NYC <-AB_NYC[AB_NYC$price > 0,]
```

add log price

add logarithmic price for analysis purposes

```
AB_NYC <- cbind(AB_NYC,price_log = log(AB_NYC$price))
```

remove inactive listings

remove inactive listings and make new dataset

```
AB_NYC_available <- AB_NYC %>%
  filter(availability_365 > 0)
```

add distance to Times Square to model

We want to make a statement about how central the place is. Therefore the distance to Times Square is caculated using the latitude and longitude of the listings. The package "geosphere" is used.

Times Square, Manhattan, NY, USA, Latitude and longitude coordinates are: 40.758896, -73.98513

```
coord <- cbind(AB_NYC_available$longitude,AB_NYC_available$latitude)
dist.timessquare <- distGeo(p1=coord, p2=c(-73.985130, 40.758896))
AB_NYC_available <- cbind(AB_NYC_available,dist.timessquare)
```

Prepare dataset for merging with the second dataset

```
# write neighbourhood group entries in lower case
AB_NYC_available$neighbourhood_group<-tolower(AB_NYC_available$neighbourhood_group)

#remove spaces from neighbourhood groups
AB_NYC_available$neighbourhood_group <-gsub(" ","", AB_NYC_available$neighbourhood_group)

# neighbourhood group as factor
AB_NYC_available$neighbourhood_group<-factor(AB_NYC_available$neighbourhood_group)
```

incidents dataset

The dataset was downloaded from: https://data.cityofnewyork.us/City-Government/Agency-Performance-Mapping-Indicators-gsj6-6rwm

```r
Ind_NYC<- read.csv("../01_data/Indicators_NYC.csv")
head(Ind_NYC)
```

```
##   Agency    Geographic.Unit Geographic.Identifier
## 1    DCA Community District       Staten Island 3
## 2    DCA Community District       Staten Island 2
## 3    DCA Community District       Staten Island 1
## 4    DCA Community District             Queens 14
## 5    DCA Community District             Queens 13
## 6    DCA Community District             Queens 12
##                      Indicator FY2011 FY2012 FY2013 FY2014 FY2015 FY2016
## 1 Resolved Consumer Complaints     44     40     53     38     38     33
## 2 Resolved Consumer Complaints     46     57     56     43     29     63
## 3 Resolved Consumer Complaints     75     56     29     61     42     65
## 4 Resolved Consumer Complaints     17     25      9      8      8     11
## 5 Resolved Consumer Complaints     64     36     22     41     44     61
## 6 Resolved Consumer Complaints    125    144    113    113    112    122
##   FY2017 FY2018 FY2019
## 1     22     29     14
## 2     23     25     26
## 3     46     28     34
## 4     14     23     25
## 5     36     45     40
## 6     94     59     66
```

Following changes have been made to the dataset:

```r
#Filter Data from 2019
Ind_NYC_2019<-data.frame("neighbourhood_group2"= Ind_NYC$Geographic.Identifier, "Indicator"=Ind_NYC$Ind
head(Ind_NYC_2019)
```

```
##   neighbourhood_group2                    Indicator Incidents
## 1      Staten Island 3 Resolved Consumer Complaints        14
## 2      Staten Island 2 Resolved Consumer Complaints        26
## 3      Staten Island 1 Resolved Consumer Complaints        34
## 4            Queens 14 Resolved Consumer Complaints        25
## 5            Queens 13 Resolved Consumer Complaints        40
## 6            Queens 12 Resolved Consumer Complaints        66
```

```r
Ind_NYC_2019_cleaned<-Ind_NYC_2019

#remove numbers
Ind_NYC_2019_cleaned$neighbourhood_group <-gsub("[0-9]","", Ind_NYC_2019_cleaned$neighbourhood_group2 )
#remove empty spaces
Ind_NYC_2019_cleaned$neighbourhood_group <-gsub(" ","", Ind_NYC_2019_cleaned$neighbourhood_group )
#lowercases
Ind_NYC_2019_cleaned$neighbourhood_group<-tolower(Ind_NYC_2019_cleaned$neighbourhood_group)
#factor
Ind_NYC_2019_cleaned$neighbourhood_group<-factor(Ind_NYC_2019_cleaned$neighbourhood_group)

#overview
head(Ind_NYC_2019_cleaned$Incidents)
```

```
## [1] 14 26 34 25 40 66
```

```r
head(Ind_NYC_2019_cleaned$neighbourhood_group)
```

```
## [1] statenisland statenisland statenisland queens       queens
## [6] queens
## Levels:  bronx brooklyn manhattan queens statenisland
```

```r
summary(Ind_NYC_2019_cleaned)
```

```
##  neighbourhood_group2
##          : 177
##  Bronx 1 :  35
##  Bronx 10:  35
##  Bronx 11:  35
##  Bronx 2 :  35
##  Bronx 3 :  35
##  (Other) :3307
##                                                             Indicator
##                                                                  : 177
##  Average Response Time to crimes in progress - Critical (minutes):  77
##  Burglary                                                        :  77
##  Crime related to domestic violence - Felonious assault          :  77
##  Crime related to domestic violence - Murder                     :  77
##  Crime related to domestic violence - Rape                       :  77
##  (Other)                                                         :3097
##    Incidents          neighbourhood_group
##  Min.   :     0.0               :1633
##  1st Qu.:    12.6   bronx       : 424
##  Median :    85.6   brooklyn    : 616
##  Mean   :  2319.2   manhattan   : 400
##  3rd Qu.:   322.8   queens      : 480
##  Max.   :424490.0   statenisland: 106
##  NA's   :1181
```

```r
summary(Ind_NYC_2019_cleaned$Indicator)
```

```
##
##                                                                                177
##                                                             Air complaints received
##                                                                                 59
##                                                        Asbestos complaints received
##                                                                                 59
##                                                           Average Daily Attendance
##                                                                                 32
##                                                 Average expenditure per student ($)
##                                                                                 32
##                              Average Response Time to crimes in progress - Critical (minutes)
##                                                                                 77
##          Average response time to life-threatening medical emergencies by ambulance units
##                                                                                  5
##              Average response time to life-threatening medical emergencies by fire units
##                                                                                  5
##                                                 Average response time to structural fires
##                                                                                  5
##                                                                            Burglary
##                                                                                 77
```

```
##              Children in the public schools who have completed required immunizations (%)
##                                                                                       32
##        Citywide acceptability rating for the cleanliness of small parks and playgrounds (%)
##                                                                                       59
## Citywide acceptability rating for the overall condition of small parks and playgrounds (%)
##                                                                                       59
##                                                                  Civilian fire fatalities
##                                                                                       59
##                                        Crime related to domestic violence - Felonious assault
##                                                                                       77
##                                                   Crime related to domestic violence - Murder
##                                                                                       77
##                                                     Crime related to domestic violence - Rape
##                                                                                       77
##                                  Curbside and containerized mixed paper recycled tons per day
##                                                                                       59
##                                           Curbside and Containerized Recycled Tons Per Day
##                                                                                       59
##                                         Curbside and Containerized Recycling Diversion Rate
##                                                                                       59
##                                                 Deaths from unintentional drug overdose (CY)
##                                                                                       59
##                                                        Domestic Violence Related Radio Runs
##                                                                                       77
##                                                                         Felonious assault
##                                                                                       77
##                                                                             Forcible rape
##                                                                                       77
##                                                                             Grand larceny
##                                                                                       77
##                                                                        Grand larceny auto
##                                                                                       77
##                                                        Hate Crime Related Felonious Assault
##                                                                                       77
##                                                                   Hate Crime Related Murder
##                                                                                       77
##                                                                        Hate Crimes (total)
##                                                                                       77
##                                                                    Intentionally set fires
##                                                                                       59
##                                                                        Major felony crime
##                                                                                       77
##                                                        Medical Emergencies (fire unit only)
##                                                                                       59
##                                                         Murder and non-negligent manslaughter
##                                                                                       77
##                          New Cases Requiring Environmental Intervention For Lead Poisoning
##                                                                                       59
##                                                                  Noise complaints received
##                                                                                       59
##                                                                        Nonstructural Fires
##                                                                                       59
##                                           Number of Priority A (emergency) complaints received
##                                                                                       59
```

```
##                                    Number of Priority B (nonemergency) complaints received
##                                                                                          59
##                                                         Persons receiving Cash Assistance
##                                                                                          59
##                                                           Persons receiving SNAP benefits
##                                                                                          59
##                                                          Private transfer station permits
##                                                                                          59
##                                                         Public Health Insurance enrollees
##                                                                                          59
##                                                             Recycling tons per truckshift
##                                                                                          59
##                                                Refuse Collected for Disposal (tons per day)
##                                                                                          59
##                                                                Refuse tons per truckshift
##                                                                                          59
##                                                              Resolved Consumer Complaints
##                                                                                          59
##                                                   Restaurants scoring an â\200\230Aâ\200\231 grade (
##                                                                                          59
##                                                                                     Robbery
##                                                                                          77
##                                     School Buildings in Good or Fair to Good Condition (%)
##                                                                                          32
##                                                       Sidewalks rated acceptably clean (%)
##                                                                                          59
##                                                                Sidewalks rated filthy (%)
##                                                                                          59
##                                      Streets maintained with a pavement rating of Good (%)
##                                                                                          59
##                                                         Streets rated acceptably clean (%)
##                                                                                          59
##                                                                  Streets rated filthy (%)
##                                                                                          59
##                                                                          Structural Fires
##                                                                                          59
##         Students in grades 3 to 8 meeting or exceeding standards - English Language Arts (%)
##                                                                                          32
##                             Students in grades 3 to 8 meeting or exceeding standards - Math (%)
##                                                                                          32
##                       Students in schools that exceed capacity (%)   - Elementary/middle schools
##                                                                                          32
##                                                               Tons of refuse collected (000)
##                                                                                          59
##                                                                Total housing starts (units)
##                                                                                          59
##                                                                Total Segment 1-8 Incidents
##                                                                                           5
##                                                                           Water main breaks
##                                                                                          59
```

```r
levels(Ind_NYC_2019_cleaned$neighbourhood_group)
```

```
## [1] ""              "bronx"         "brooklyn"      "manhattan"
## [5] "queens"        "statenisland"
```

```r
# sum of incidents per neighbourhood group and indicator
Summary_Ind_NYC_2019<-Ind_NYC_2019_cleaned %>%
  group_by(neighbourhood_group=Ind_NYC_2019_cleaned$neighbourhood_group,Indicator) %>%
  summarise(Observations=sum(Incidents,na.rm = TRUE))
summary(Summary_Ind_NYC_2019)
```

```
##    neighbourhood_group
##              :24
##  bronx       :38
##  brooklyn    :38
##  manhattan   :37
##  queens      :38
##  statenisland:38
##
##                                                                                  Indicator
##  Air complaints received                                                              :  5
##  Asbestos complaints received                                                         :  5
##  Average response time to life-threatening medical emergencies by ambulance units    :  5
##  Average response time to life-threatening medical emergencies by fire units         :  5
##  Average response time to structural fires                                            :  5
##  Citywide acceptability rating for the cleanliness of small parks and playgrounds (%):  5
##  (Other)                                                                              :183
##   Observations
##  Min.   :     0
##  1st Qu.:     6
##  Median :   273
##  Mean   : 26981
##  3rd Qu.:  2914
##  Max.   :556596
##
```

```r
# remove entries without neighbourhood group
Summary_Ind_NYC_2019<-filter(Summary_Ind_NYC_2019,neighbourhood_group != "")
summary(Summary_Ind_NYC_2019)
```

```
##    neighbourhood_group
##              : 0
##  bronx       :38
##  brooklyn    :38
##  manhattan   :37
##  queens      :38
##  statenisland:38
##
##                                                                                  Indicator
##  Air complaints received                                                              :  5
##  Asbestos complaints received                                                         :  5
##  Average response time to life-threatening medical emergencies by ambulance units    :  5
##  Average response time to life-threatening medical emergencies by fire units         :  5
##  Average response time to structural fires                                            :  5
##  Citywide acceptability rating for the cleanliness of small parks and playgrounds (%):  5
##  (Other)                                                                              :159
##   Observations
##  Min.   :     0.0
##  1st Qu.:     7.2
##  Median :   273.0
```

```
##   Mean   : 29370.4
##   3rd Qu.:   2617.5
##   Max.   :556596.0
##
```

```
head(Summary_Ind_NYC_2019)
```

```
## # A tibble: 6 x 3
## # Groups:   neighbourhood_group [1]
##   neighbourhood_gro~ Indicator                                Observations
##   <fct>              <fct>                                           <dbl>
## 1 bronx              Air complaints received                        536
## 2 bronx              Asbestos complaints received                   212
## 3 bronx              Average response time to life-threatenin~        7.44
## 4 bronx              Average response time to life-threatenin~        5.13
## 5 bronx              Average response time to structural fires        4.36
## 6 bronx              Citywide acceptability rating for the cl~     1137.
```

```
# nested indicators
NYC_nest<-Summary_Ind_NYC_2019 %>%
  nest(Indicator=c(Indicator, Observations))
head(NYC_nest)
```

```
## # A tibble: 5 x 2
##   neighbourhood_group data
##   <fct>               <list>
## 1 bronx               <tibble [38 x 2]>
## 2 brooklyn            <tibble [38 x 2]>
## 3 manhattan           <tibble [37 x 2]>
## 4 queens              <tibble [38 x 2]>
## 5 statenisland        <tibble [38 x 2]>
```

Merge datasets

```
#Join both datasets
NYC<-left_join(AB_NYC_available,NYC_nest, by="neighbourhood_group")
```

```
## Warning: Column `neighbourhood_group` joining factors with different
## levels, coercing to character vector
```

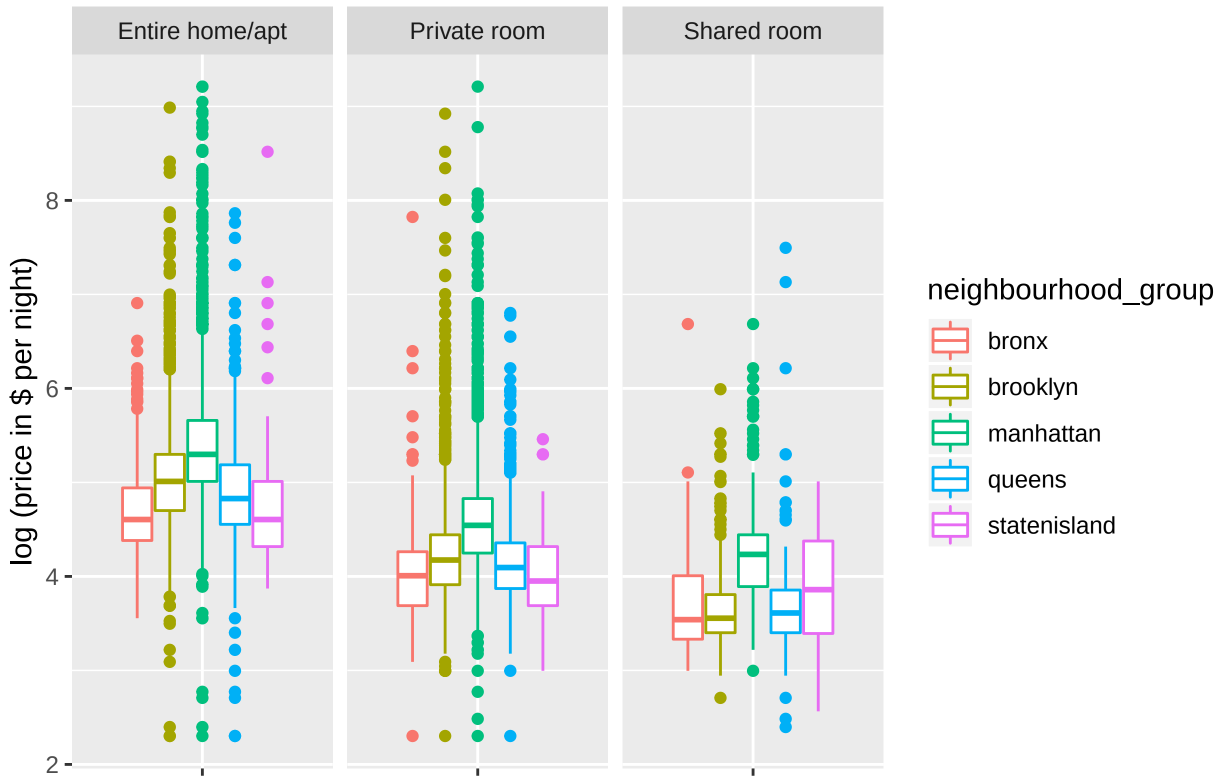
```
# neighbourhood group as factor
NYC$neighbourhood_group<-factor(NYC$neighbourhood_group)
```

Data visualisation

Distribution of prices by room types and neighbourhood

```
ggplot(data = AB_NYC_available,
       mapping = aes(y = price_log,
                     x = "",
                     group = neighbourhood_group,
                     colour = neighbourhood_group)) +
  geom_boxplot() +
  facet_wrap(. ~ room_type)+
```

```
  xlab("")+
  ylab("log (price in $ per night)")
```

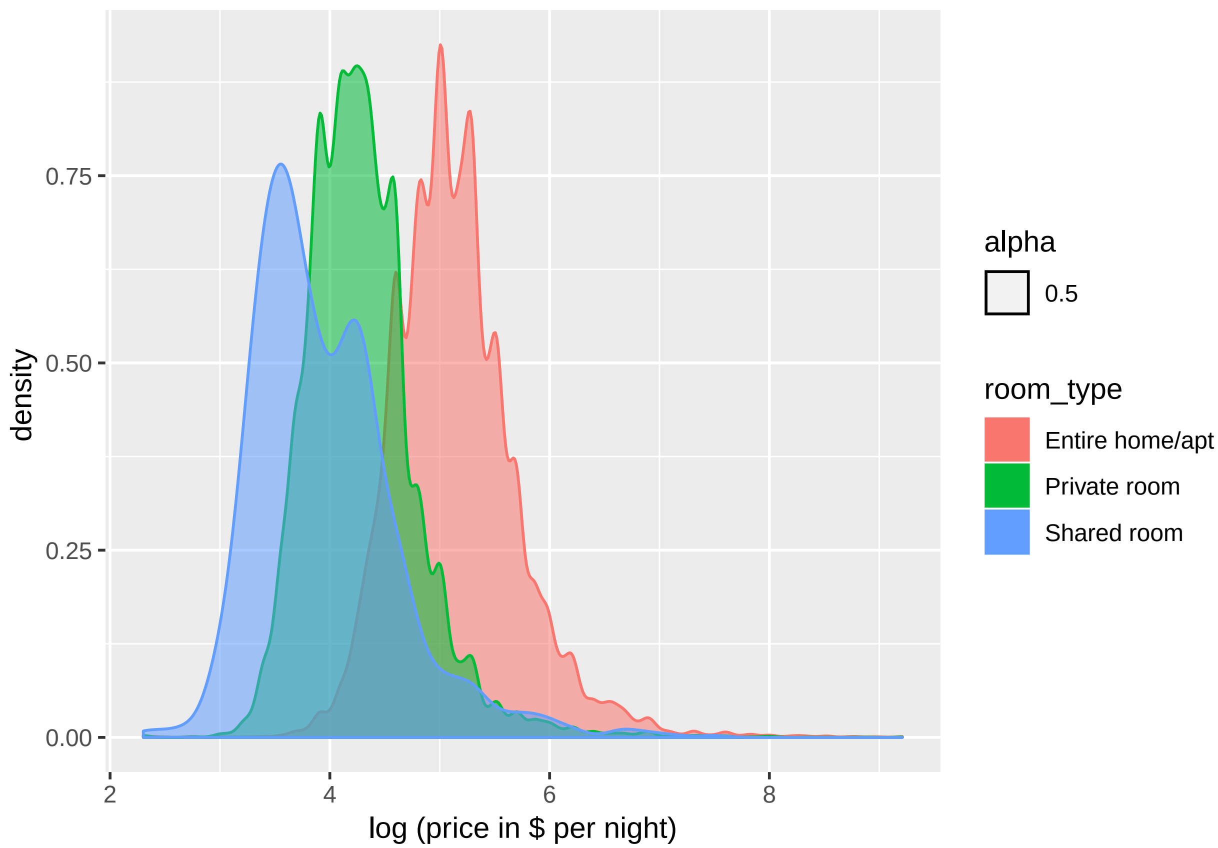


Prices of the room type “entire home/apt” have the highest median, followed by “private room”" and lastly “shared room”“, which is not surprising. 25. and 75. quantile for”entire home/apt" and “private room” are similarly distributed, for shared room there is no clear pattern.

For all room types, median prices in neighbourhood “Manhattan” are the highest. For for “entire home/apt” and “private room” the second highest mediam prices are in Manhattan.

Distribution of prices

```
ggplot(data = AB_NYC,
       mapping = aes(x = price_log,
                     group = room_type,
                     colour = room_type,
                     fill = room_type,
                     alpha = 0.5)) +
  geom_density() +
  xlab("log (price in $ per night)")+
  ylab("density ")
```

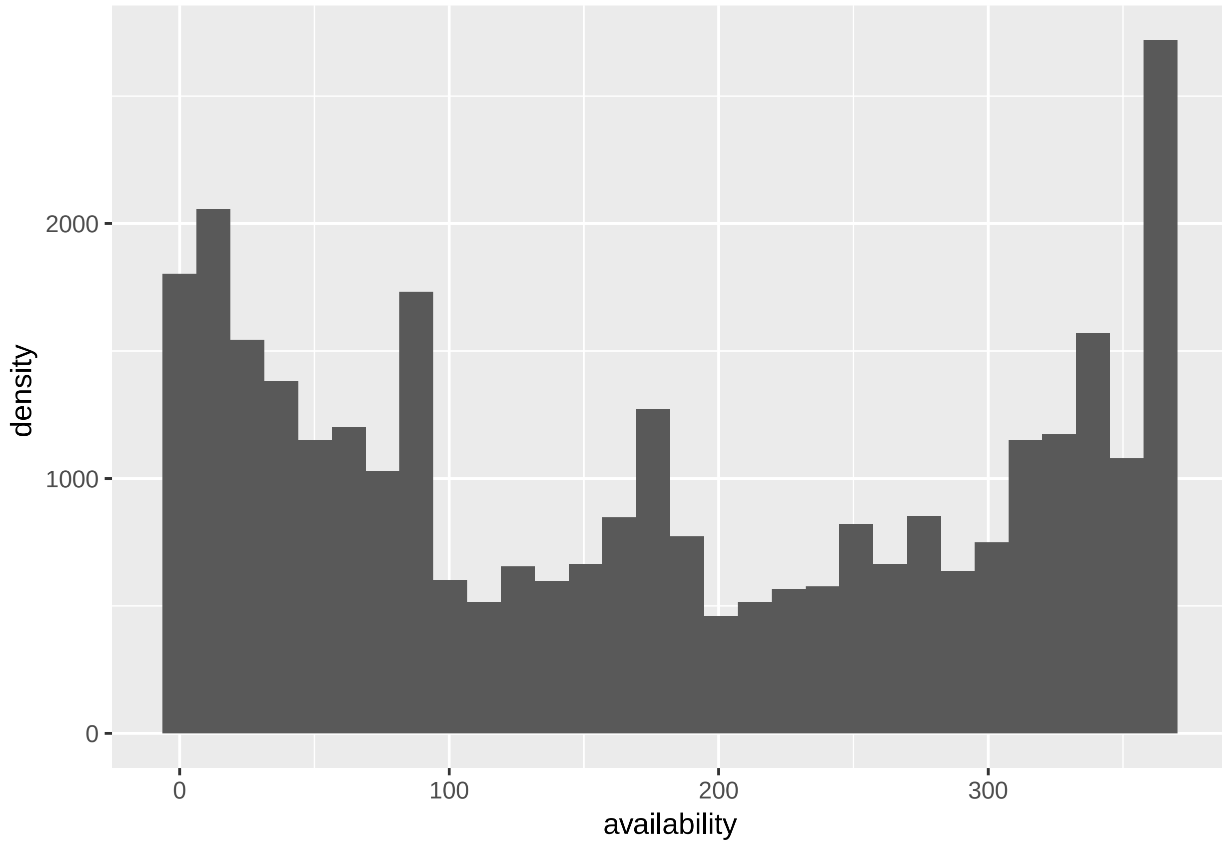


Prices for all room types are skewed to the right. Even with logarithmic display of prices, this is still clearly the case.

Availability of appartments

```
ggplot(data = AB_NYC_available,
       mapping = aes(x = availability_365)) +
  geom_histogram() +
  xlab("availability")+
  ylab("density ")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



There are a lot of listings with very low or very high (almost year round) availablity. Listings with no available days in 2019 were removed from the dataset. This distribution was not taken into account when looking at the prices.

Possible models to calculate the price of an airbnb

Simple linear models

Impact of several variables on the price are analysed. The highest R2 is reached with "room type".

```
##simple linear models

# price ~ neighbourhood group

lm.hood <- lm (data=AB_NYC_available, price_log~neighbourhood_group)
summary(lm.hood)
```

```
##
## Call:
## lm(formula = price_log ~ neighbourhood_group, data = AB_NYC_available)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7663 -0.4698 -0.0473  0.3886  4.3652
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                       4.25517    0.02199 193.516  < 2e-16 ***
## neighbourhood_groupbrooklyn      0.36688    0.02279  16.096  < 2e-16 ***
## neighbourhood_groupmanhattan     0.81367    0.02272  35.818  < 2e-16 ***
## neighbourhood_groupqueens        0.12670    0.02421   5.233 1.68e-07 ***
## neighbourhood_groupstatenisland  0.10551    0.04263   2.475   0.0133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6644 on 31349 degrees of freedom
## Multiple R-squared:  0.1491, Adjusted R-squared:  0.1489
## F-statistic:  1373 on 4 and 31349 DF,  p-value: < 2.2e-16
```

```
# price ~ room type

lm.type <- lm (data=AB_NYC_available, price_log~room_type)
summary(lm.type)
```

```
##
## Call:
## lm(formula = price_log ~ room_type, data = AB_NYC_available)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8872 -0.3695 -0.0658  0.2816  4.8867
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.189793   0.004377 1185.79   <2e-16 ***
## room_typePrivate room -0.866270   0.006468 -133.92   <2e-16 ***
## room_typeShared room  -1.280409   0.019660  -65.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5627 on 31351 degrees of freedom
## Multiple R-squared:  0.3895, Adjusted R-squared:  0.3895
## F-statistic: 1e+04 on 2 and 31351 DF,  p-value: < 2.2e-16
```

```
# price ~ dist.timessquare

lm.dist <- lm (data=AB_NYC_available, price_log~dist.timessquare)
summary(lm.dist)
```

```
##
## Call:
## lm(formula = price_log ~ dist.timessquare, data = AB_NYC_available)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8052 -0.4752 -0.0408  0.3890  4.4443
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.211e+00  6.900e-03  755.27   <2e-16 ***
## dist.timessquare -5.913e-05  7.753e-07  -76.26   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.6615 on 31352 degrees of freedom
## Multiple R-squared:  0.1565, Adjusted R-squared:  0.1565
## F-statistic:  5816 on 1 and 31352 DF,  p-value: < 2.2e-16
```

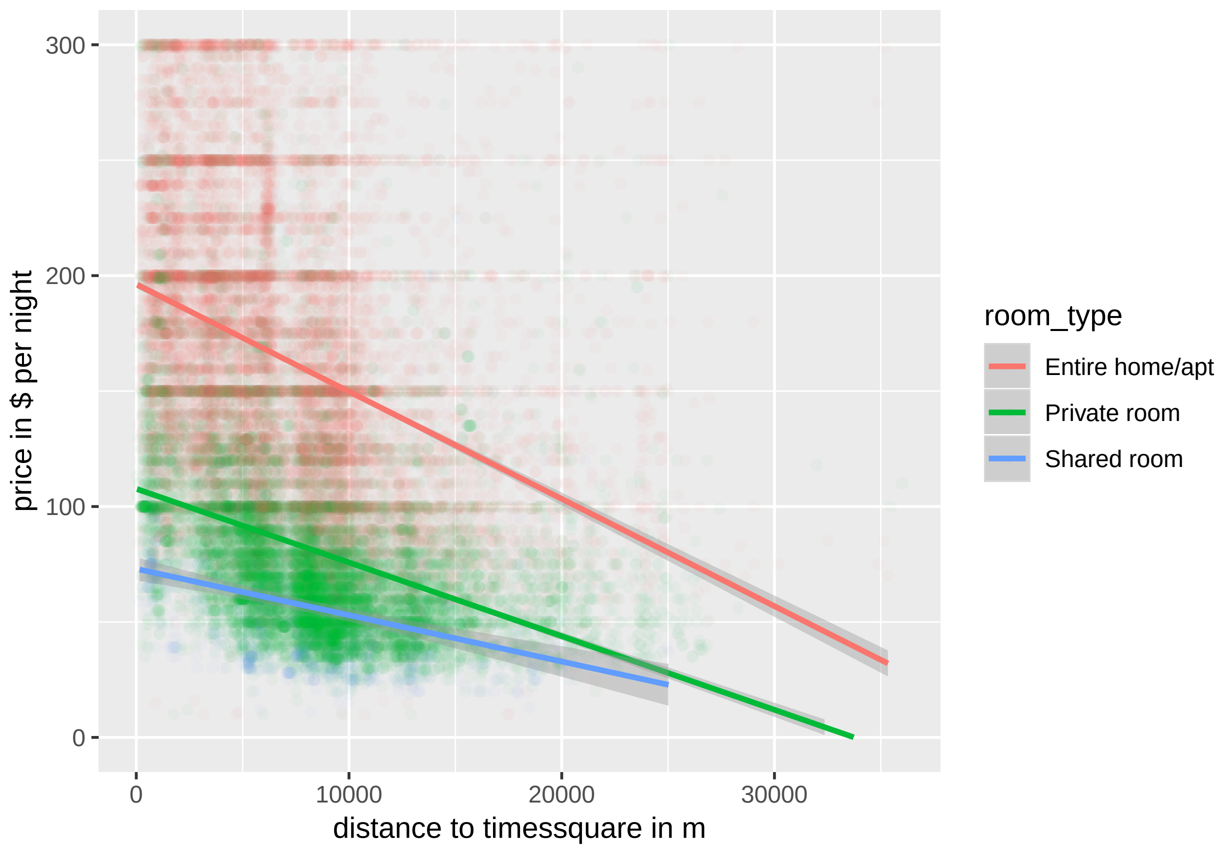
Multiple linear model

A linear model using "distance to timessquare" and "room type" (as factor) with interactions to calculate the price is applied.

```r
#distance and room type on price (with interaction)
lm.dist.type.interact <- lm (data=AB_NYC_available, price_log~dist.timessquare*room_type)
summary(lm.dist.type.interact)
```

```
##
## Call:
## lm(formula = price_log ~ dist.timessquare * room_type, data = AB_NYC_available)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0389 -0.3348 -0.0639  0.2365  4.7750
##
## Coefficients:
##                                         Estimate Std. Error t value
## (Intercept)                            5.480e+00  6.991e-03 783.825
## dist.timessquare                      -4.357e-05  8.550e-07 -50.955
## room_typePrivate room                 -7.825e-01  1.148e-02 -68.163
## room_typeShared room                  -1.214e+00  3.403e-02 -35.682
## dist.timessquare:room_typePrivate room -7.460e-07  1.274e-06  -0.586
## dist.timessquare:room_typeShared room  -2.747e-07  3.568e-06  -0.077
##                                       Pr(>|t|)
## (Intercept)                             <2e-16 ***
## dist.timessquare                        <2e-16 ***
## room_typePrivate room                   <2e-16 ***
## room_typeShared room                    <2e-16 ***
## dist.timessquare:room_typePrivate room   0.558
## dist.timessquare:room_typeShared room    0.939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5229 on 31348 degrees of freedom
## Multiple R-squared:  0.4729, Adjusted R-squared:  0.4728
## F-statistic:  5625 on 5 and 31348 DF,  p-value: < 2.2e-16
```

```r
ggplot(data = AB_NYC_available,
       mapping = aes(y = price,
                     x = dist.timessquare,
                     colour = room_type,
                     group = room_type)) +
  geom_point(alpha = 0.03) +
  xlab("distance to timessquare in m")+
  ylab("price in $ per night")+
  ylim(0,300)+
  geom_smooth(method="lm")
```



There is a tendency for all room types that the price is lower if the place is further from Times Square. The interactions are not significant.

Multiple linear model by choosing smallest RSS

A multiple linear model is created. The best model to calculate the price we can find includes 5 variables:

- room type (factor with 3 levels, entire appartment being the highest and shared room the lowest)
- distance to timessquare (negative effect)
- availability (positive effect)
- neighbourhood group (factor with 5 levels, Manhattan being the highest and Bronx the lowest)
- minimum nights (negative effect)

```
#full model
lm.full <- lm (data=AB_NYC_available, price_log~room_type
               +neighbourhood_group
               +minimum_nights
               +number_of_reviews
               +calculated_host_listings_count
               +availability_365
               +dist.timessquare)

#empty model
lm.empty <- lm (data=AB_NYC_available, price_log~NULL)
add1(lm.empty,scope=lm.full)

#choose value with smallest RSS
```

```
lm.1 <- update(lm.empty,.~.+room_type)
add1(lm.1,scope=lm.full)
lm.2 <- update(lm.1,.~.+dist.timessquare)
add1(lm.2,scope=lm.full)
lm.3 <- update(lm.2,.~.+availability_365)
add1(lm.3,scope=lm.full)
lm.4 <- update(lm.3,.~.+neighbourhood_group)
add1(lm.4,scope=lm.full)
lm.5 <- update(lm.4,.~.+minimum_nights)
add1(lm.5,scope=lm.full)
```

```
summary(lm.5)
```

```
##
## Call:
## lm(formula = price_log ~ room_type + dist.timessquare + availability_365 +
##     neighbourhood_group + minimum_nights, data = AB_NYC_available)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0328 -0.3241 -0.0652  0.2332  4.8822
##
## Coefficients:
##                                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)                      5.082e+00  2.077e-02  244.710  < 2e-16
## room_typePrivate room           -7.823e-01  5.995e-03 -130.499  < 2e-16
## room_typeShared room            -1.235e+00  1.785e-02  -69.173  < 2e-16
## dist.timessquare                -3.073e-05  8.321e-07  -36.926  < 2e-16
## availability_365                 6.388e-04  2.311e-05   27.645  < 2e-16
## neighbourhood_groupbrooklyn      1.415e-01  1.779e-02    7.956 1.83e-15
## neighbourhood_groupmanhattan     3.210e-01  1.908e-02   16.821  < 2e-16
## neighbourhood_groupqueens        4.895e-02  1.863e-02    2.627  0.00861
## neighbourhood_groupstatenisland  1.754e-01  3.306e-02    5.304 1.14e-07
## minimum_nights                  -1.930e-03  1.226e-04  -15.741  < 2e-16
##
## (Intercept)                     ***
## room_typePrivate room           ***
## room_typeShared room            ***
## dist.timessquare                ***
## availability_365                ***
## neighbourhood_groupbrooklyn     ***
## neighbourhood_groupmanhattan    ***
## neighbourhood_groupqueens       **
## neighbourhood_groupstatenisland ***
## minimum_nights                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5089 on 31344 degrees of freedom
## Multiple R-squared:  0.5009, Adjusted R-squared:  0.5008
## F-statistic:  3496 on 9 and 31344 DF,  p-value: < 2.2e-16
```

Interactive map with the leaflet package

```
df_exp<-filter(NYC,price == max(price))
df_cheap<-filter(NYC,price == min(price))
```