E1

Felix Shier

17/01/2022

```
tinytex::install_tinytex()
```

Data Exploration and Hypothesis Test
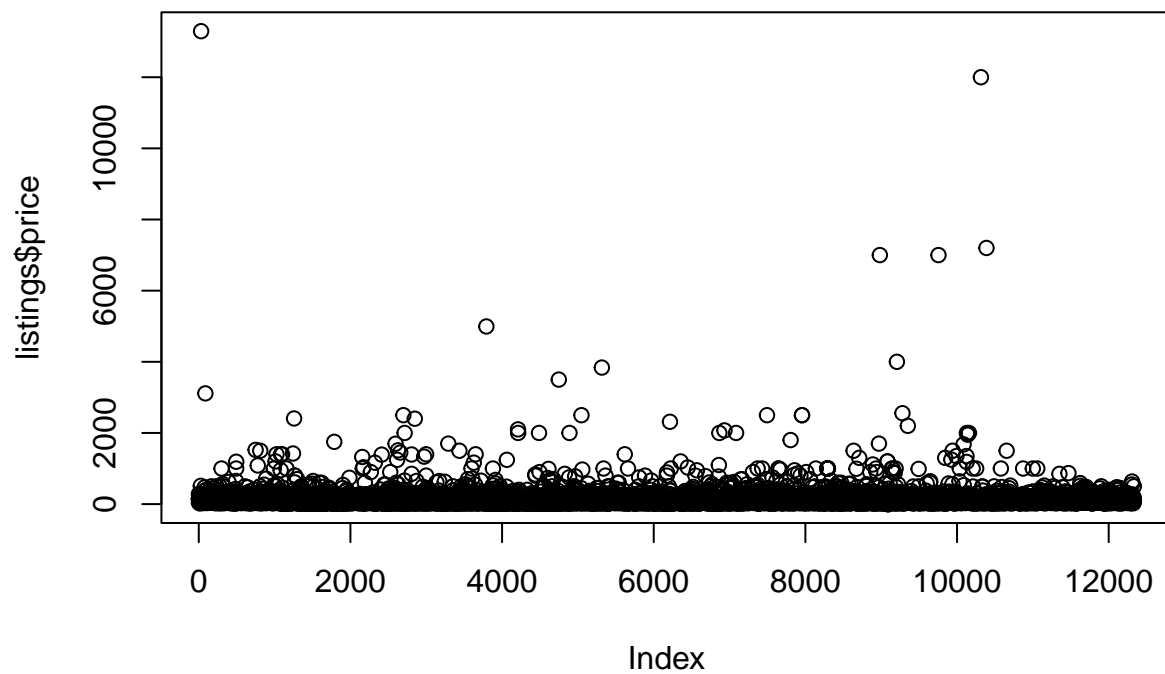
```
# import data
listings <- read.csv("listings.csv")
```

```
# show head
head(listings)
```

```
##      id                                          name host_id host_name
## 1 29059               Lovely studio Quartier Latin  125031  Maryline
## 2 29061      Maison historique - Quartier Latin  125031  Maryline
## 3 31847                       Downtown/Airport room     2319      Gail
## 4 34715             Chez Patrac ! MÃ©tro Beaubien  149769     Alain
## 5 36301         Romantic & peaceful Plateau  loft  381468    Sylvie
## 6 38118 Beautiful room  with a balcony in front of a parc  163569        M.
##   neighbourhood_group             neighbourhood latitude longitude
## 1                  NA                 Ville-Marie 45.51939 -73.56482
## 2                  NA                 Ville-Marie 45.51929 -73.56493
## 3                  NA               Le Sud-Ouest 45.48467 -73.57995
## 4                  NA Rosemont-La Petite-Patrie 45.53818 -73.60360
## 5                  NA    Le Plateau-Mont-Royal 45.53026 -73.58413
## 6                  NA                 Ville-Marie 45.52699 -73.55840
##          room_type price minimum_nights number_of_reviews last_review
## 1 Entire home/apt   152              2               333  2020-03-15
## 2 Entire home/apt   286              3                82  2021-09-06
## 3     Private room    40              2               157  2019-10-02
## 4 Entire home/apt   140              4                 4  2011-09-12
## 5 Entire home/apt   130             31                42  2021-08-31
## 6     Private room    50             10                14  2019-09-15
##   reviews_per_month calculated_host_listings_count availability_365
## 1              2.82                              2              323
## 2              0.71                              2              249
## 3              1.37                              3              255
## 4              0.03                              1              214
## 5              0.71                              7              118
## 6              0.14                              3              342
```

```
##   number_of_reviews_ltm license
## 1                     4
## 2                    14
## 3                     4
## 4                     0 261 026
## 5                     3
## 6                     0
```
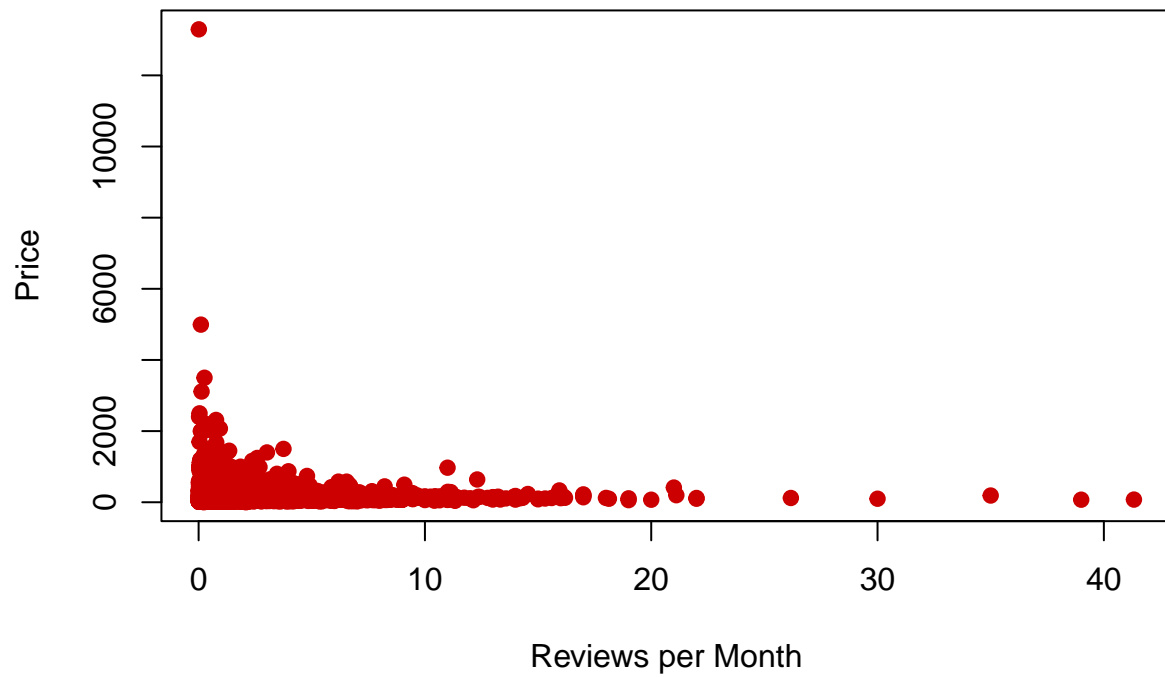
```
# plot prices
plot(listings$price)
```
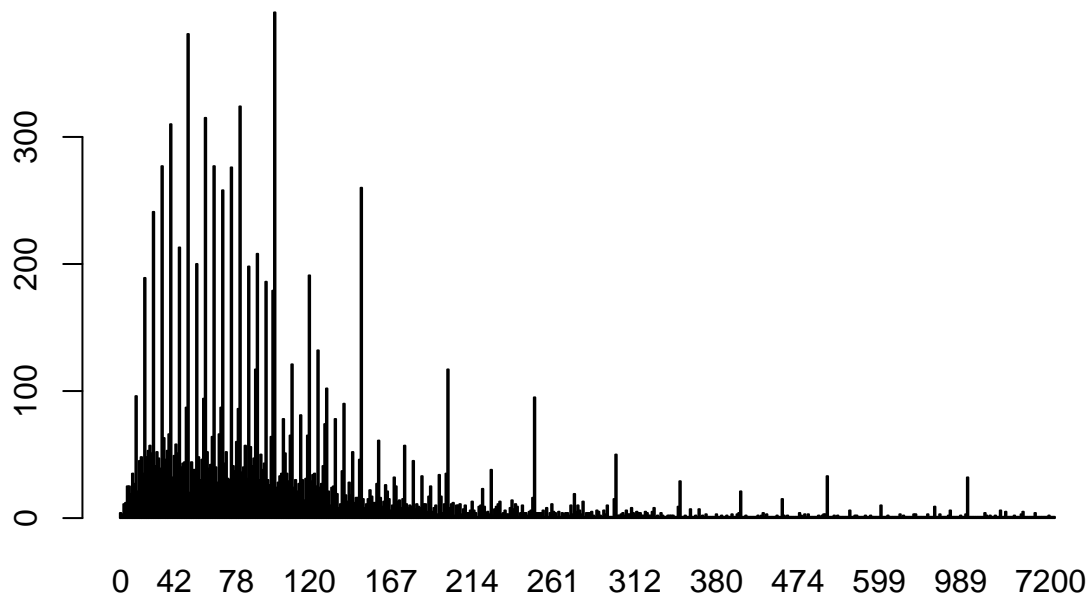


```
# plot price vs. reviews per month
plot(listings$reviews_per_month, listings$price,
     col = "#cc0000",
     pch = 19,
     main = "Listing Price vs. Reviews per Month",
     xlab = "Reviews per Month",
     ylab = "Price")
```

Listing Price vs. Reviews per Month


```r
# barplot prices
prices <- table(listings$price)
barplot(prices)
```

```
summary(listings)
```

```
##       id              name             host_id            host_name
## Min.   :   29059   Length:12333      Min.   :     2319   Length:12333
## 1st Qu.:19020060   Class :character   1st Qu.: 19751308   Class :character
## Median :30414140   Mode  :character   Median : 72761895   Mode  :character
## Mean   :30260472                      Mean   :118227401
## 3rd Qu.:43294350                      3rd Qu.:195416989
## Max.   :52773199                      Max.   :426878210
##
## neighbourhood_group neighbourhood        latitude        longitude
## Mode:logical        Length:12333      Min.   :45.40   Min.   :-73.96
## NA's:12333          Class :character   1st Qu.:45.50   1st Qu.:-73.60
##                     Mode  :character   Median :45.52   Median :-73.58
##                                        Mean   :45.51   Mean   :-73.59
##                                        3rd Qu.:45.53   3rd Qu.:-73.57
##                                        Max.   :45.70   Max.   :-73.48
##
##   room_type            price         minimum_nights    number_of_reviews
## Length:12333       Min.   :    0.0   Min.   :    1.00   Min.   :  0.00
## Class :character   1st Qu.:   52.0   1st Qu.:    1.00   1st Qu.:  1.00
## Mode  :character   Median :   85.0   Median :    2.00   Median :  5.00
##                    Mean   :  125.6   Mean   :   15.81   Mean   : 21.54
##                    3rd Qu.:  130.0   3rd Qu.:   15.00   3rd Qu.: 20.00
##                    Max.   :13294.0   Max.   :11684.00   Max.   :627.00
```

4

```
## 
##  last_review       reviews_per_month calculated_host_listings_count
##  Length:12333      Min.   : 0.010    Min.    :  1.00
##  Class :character  1st Qu.: 0.120    1st Qu.:  1.00
##  Mode  :character  Median : 0.460    Median :  1.00
##                    Mean   : 1.244    Mean    :  8.73
##                    3rd Qu.: 1.550    3rd Qu.:  5.00
##                    Max.   :41.330    Max.    :140.00
##                    NA's   :2655
##  availability_365 number_of_reviews_ltm   license
##  Min.   :  0.0    Min.   :  0.000        Length:12333
##  1st Qu.:  0.0    1st Qu.:  0.000        Class :character
##  Median : 48.0    Median :  0.000        Mode  :character
##  Mean   :112.1    Mean   :  3.985
##  3rd Qu.:220.0    3rd Qu.:  3.000
##  Max.   :365.0    Max.   :125.000
## 
```

```r
# price histogram
hist(listings$price,
     col = "skyblue3",
     breaks = c(0,50,100,150,200,250, 300, 350, 400, 450, 500, 13294.0))
```

Histogram of listings$price