

# 260 project

Yulin He

10/28/2021

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

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

library(ggmap)

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
## Please cite ggmap if you use it! See citation("ggmap") for details.
df<-read.csv("listings.csv")
```

## Get Boston map by ggmap package

```
mapbos <- get_map(c(left = -71.17179, bottom = 42.23594, right =-71.00010, top = 42.38998),maptype = "t
```

## Boston airbnb location distribution

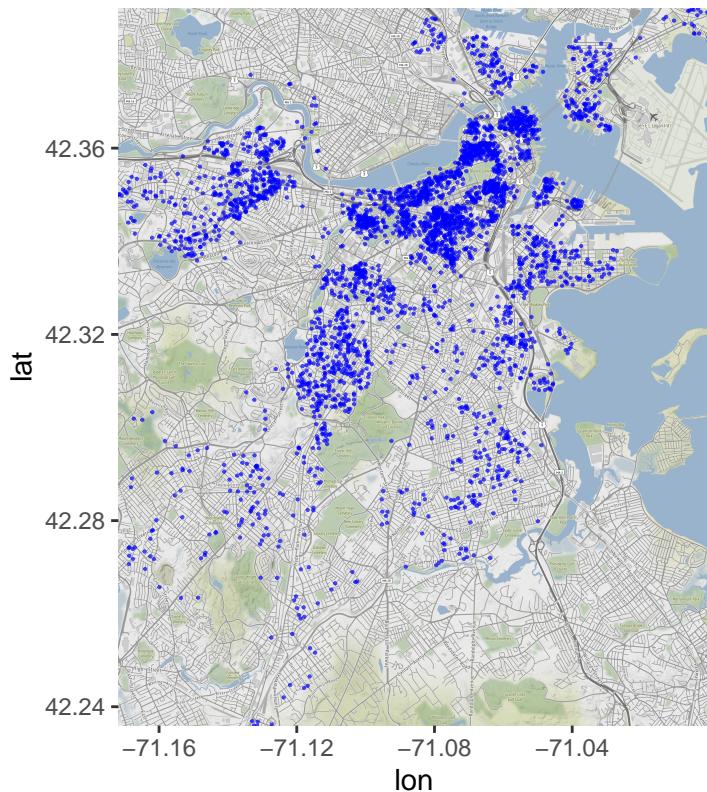
By longitude and latitude number in our dataset, we draw points in boston map. From the map, we can see that airbnb are mainly located in Boston Downtown, Backbay, Fenway, South End, Jamaica Plain, Seaport, East Boston, Charlestown, and Allston. And then we want to explore the price distribution. We hope to get a sense of in which area, the price is high, and in which area, price is low.

```
bos<-ggmap(mapbos) +
  geom_point(data = df,
             aes(x = longitude, y = latitude, size = I(0.5)),
             color = "blue", alpha = 0.7, size=0.08)

bos <- bos + labs(title = "Boston Airbnb Location Distribution")
bos

## Warning: Removed 2 rows containing missing values (geom_point).
```

## Boston Airbnb Location Distribution



## Price level heatmap

```
df$price = as.numeric(gsub("\\$", "", df$price))

## Warning: NAs introduced by coercion
summary(df$price)

##      Min.   1st Qu.   Median   Mean   3rd Qu.   Max.   NA's
##      10.0    85.0   150.0   169.1   220.0   999.0    12

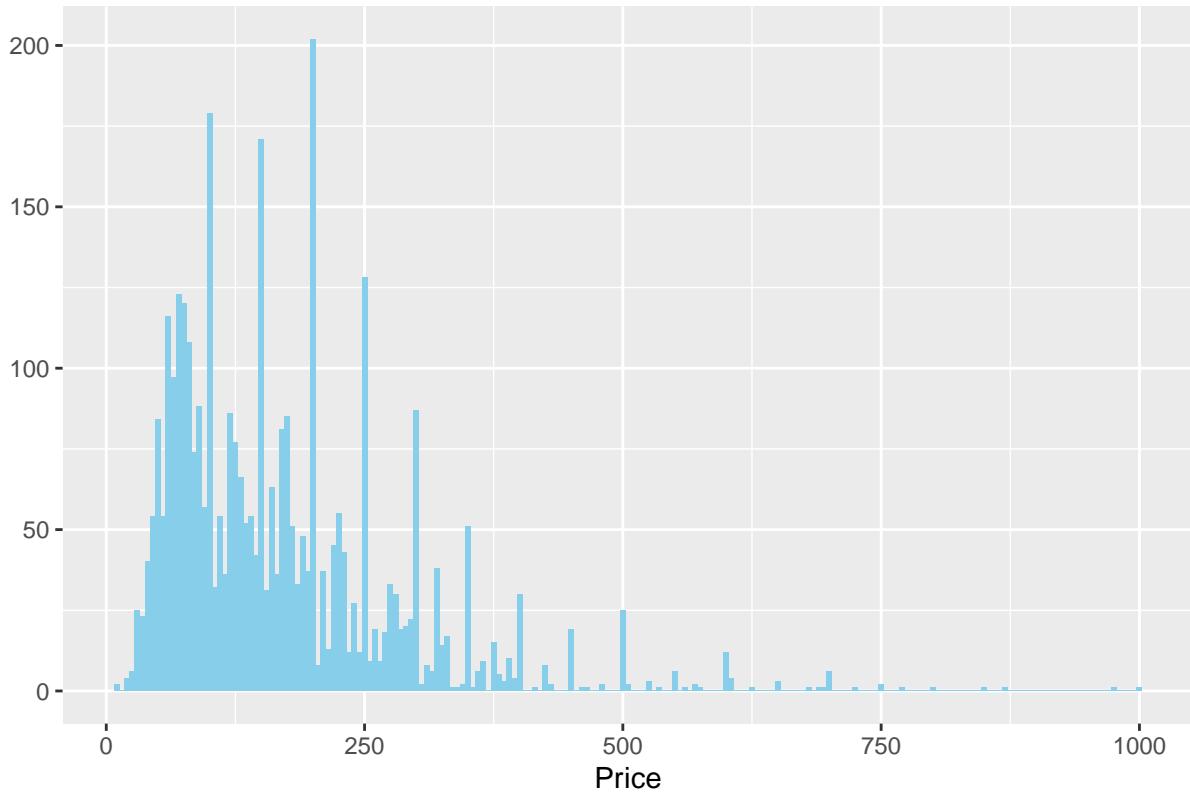
dfnew<- df %>%
  mutate(pricelevel = case_when(price <=85 ~ 1,
                                price>85 & price<=150 ~ 2,
                                price>150 & price<=220 ~ 3,
                                price>220 ~4))

dfnew$pricelevel<-as.factor(dfnew$pricelevel)

qplot(dfnew$price,
      geom="histogram",
      binwidth=5,
      main="Histogram of Price",
      xlab="Price",
      fill=I("skyblue"))

## Warning: Removed 12 rows containing non-finite values (stat_bin).
```

## Histogram of Price



By statistical analysis, we get the scale that the first quatile is 85 dollars, the median is 150 dollars, and the third quatile is 220 dollars. The min is 10 dollars and the max is 999 dollars. There are 12 NA values that we just ignore since the number is small compare to the sample size. We divided all the airbnb into 4 group 10 - 85, 86 - 150, 151 - 220, and 221 - 999, and draw a heatmap. The darker color indicates the higher price. From the heatmap below, we can see that high price airbnb mainly located in Fenway, Backbay, Downtown, and South End, some are located in Seaport and Charlestown. Airbnbs in Jamaica Plain, East Boston, and Allston are cheaper.

```
bosprice<-ggmap(mapbos) +
  geom_point(data = dfnew,
             aes(x = longitude, y = latitude, size = I(0.5), color=pricelevel),
             alpha = 0.7, size=0.08)+ 
  scale_color_manual(values=c("yellow","orange","red","firebrick4"))

bosprice

## Warning: Removed 14 rows containing missing values (geom_point).
```

