# 615 Final Project

Ningze Zu

12/7/2018

Abstract

In this project, I will use benford's law to analysis the distribution of Airbnb listings' price in Seattle and perform an exploratory data analysis figure out other features that are relevant to the rental price of Airbnb.

Price Distribution & Benford's Law

Figure 1 below shows the price distribution of Airbnb listings in Seattle. The plot displays a normal distribution but a little right skewness. It shows a high concentration of listing prices between 0 and 200 USD.

Figure 1. Airbnb Price Distribution in Seattle ($0 ~$1000)
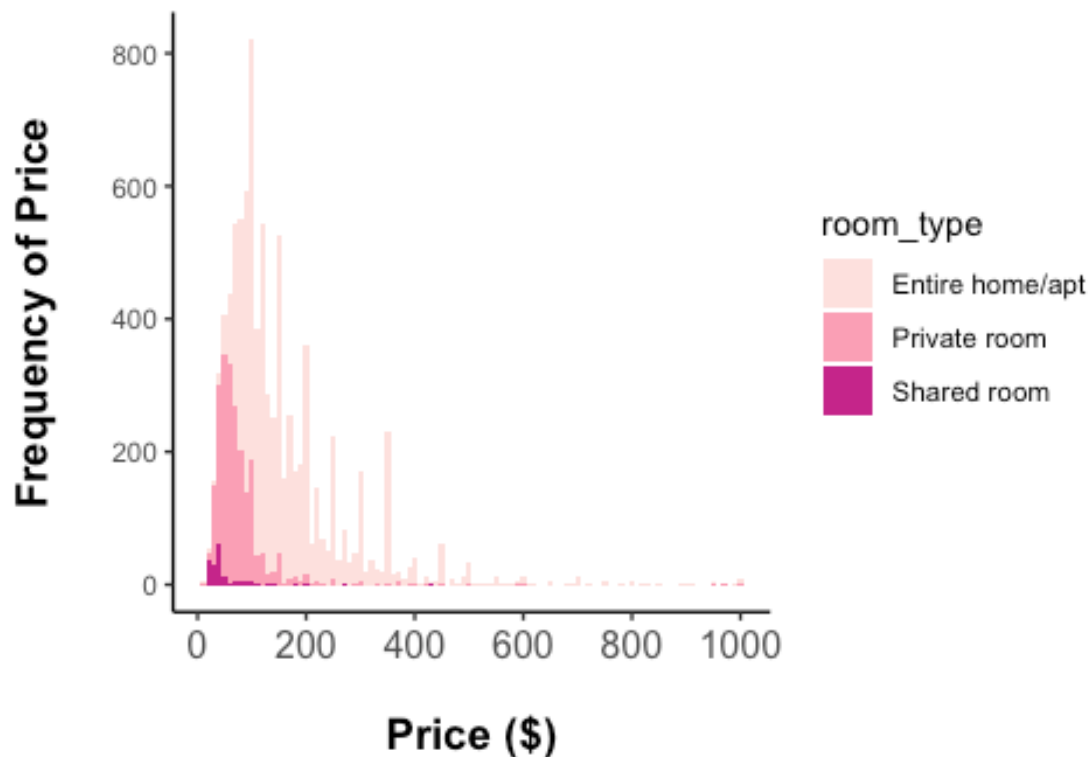
To validate the data against Benford's law, I using the first two digits to analysis 'benford'. The main results and plots of the analysis are shown below:

In the plots, the original data is in blue, and the expected Benford's distribution is in red. We can see from the first plot that the Airbnb rental price in Seattle do have a tendency to follow the Benford's Law. However, there are also some clear discrepancies.

The print result of the analysis displays the main statistics of the data. We can see that the mean value is close to 0.5. It shows the five largest discrepancies and some results of statistical tests like Chi-squared test as well. The p-value is pretty small, but we cannot focus on p-value only.

```
bf <- benford(airbnb_se_m$price)
plot(bf)
```
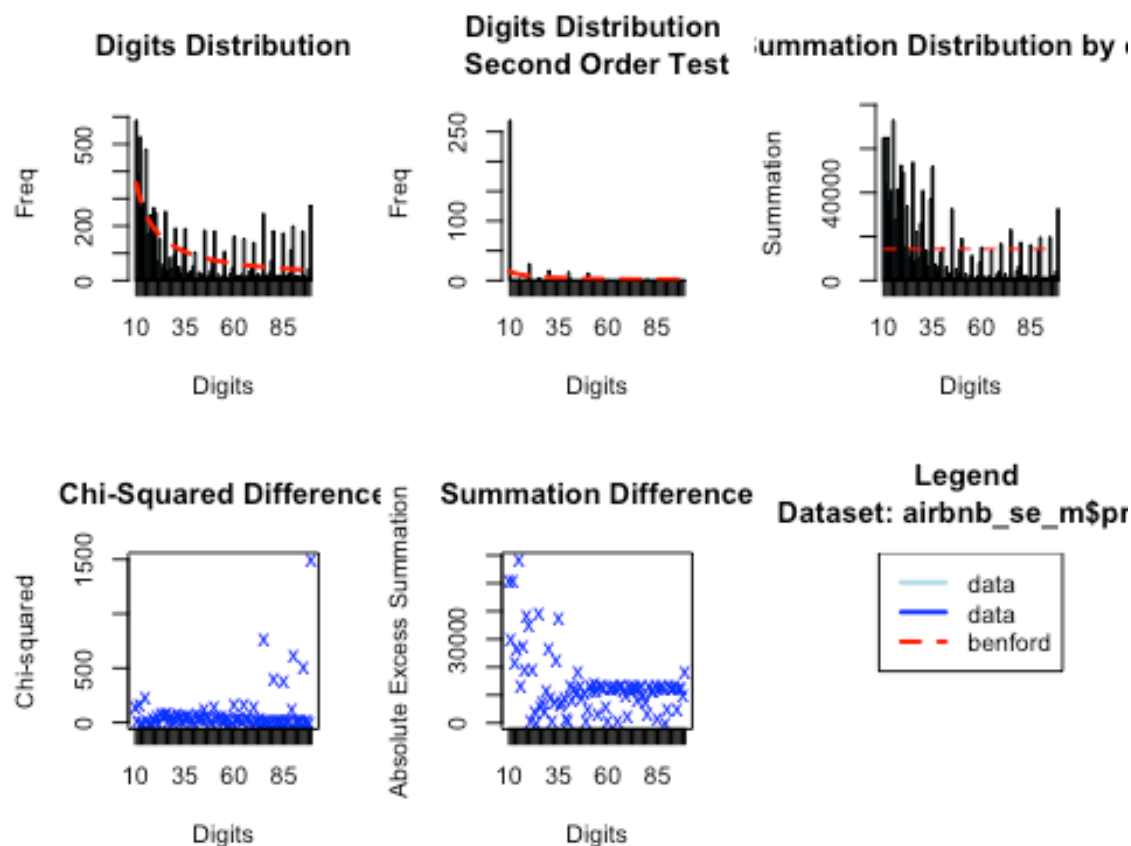


```
bf

##
## Benford object:
##
## Data: airbnb_se_m$price
## Number of observations used = 8610
## Number of obs. for second order = 364
## First digits analysed = 2
##
## Mantissa:
```

```
## 
##     Statistic Value
##          Mean  0.48
##           Var  0.11
##   Ex.Kurtosis -1.43
##      Skewness  0.15
## 
## 
## The 5 largest deviations:
## 
##    digits absolute.diff
## 1      15        239.67
## 2      99        237.42
## 3      10        230.61
## 4      12        227.70
## 5      75        195.47
## 
## Stats:
## 
##   Pearson's Chi-squared test
## 
## data:  airbnb_se_m$price
## X-squared = 7826.1, df = 89, p-value < 2.2e-16
## 
## 
##   Mantissa Arc Test
## 
## data:  airbnb_se_m$price
## L2 = 0.057452, df = 2, p-value < 2.2e-16
## 
## Mean Absolute Deviation: 0.007389969
## Distortion Factor: 2.066781
## 
## Remember: Real data will never conform perfectly to Benford's Law. You
## should not focus on p-values!
```

Next, I extract the observations with the largest discrpencies and plot the price
distribution.

```
suspects <- getSuspects(bf, airbnb_se_m)
head(suspects)

##      X      id                                     name host_id
## 1: 30  24212     ★PERFECT Getaway w/ Views★SALE 10/8, 11    74305
## 2: 52 103908               Green Lake Gem - Walk to Lake   541406
## 3: 61 132160         Green Lake Guest Suite, private apt   558743
## 4: 81 202174           Modern Townhome in Superb Location    58235
## 5: 83 202260 Cozy Seattle Apartment Near UW & Children's   990195
## 6: 84 208356          contemporary art loft downtown [7] 1025723
##    host_is_superhost                      street neighbourhood_cleansed
```

```
## 1:                      f Seattle, WA, United States                      Belltown
## 2:                      f Seattle, WA, United States                    Green Lake
## 3:                      t Seattle, WA, United States                    Green Lake
## 4:                      f Seattle, WA, United States International District
## 5:                      f Seattle, WA, United States                     View Ridge
## 6:                      t Seattle, WA, United States International District
##         city state zipcode country_code latitude longitude property_type
## 1: Seattle    WA      NA               US 47.61540 -122.3487   Condominium
## 2: Seattle    WA      NA               US 47.68427 -122.3266     Apartment
## 3: Seattle    WA      NA               US 47.68226 -122.3434   Guest suite
## 4: Seattle    WA      NA               US 47.59605 -122.3179     Townhouse
## 5: Seattle    WA      NA               US 47.68004 -122.2840     Apartment
## 6: Seattle    WA      NA               US 47.59908 -122.3256          Loft
##            room_type accommodates bathrooms bedrooms beds
## 1: Entire home/apt              6       1.0        2    3
## 2: Entire home/apt              4       1.0        1    1
## 3: Entire home/apt              6       1.0        2    3
## 4: Entire home/apt              4       2.0        2    2
## 5: Entire home/apt              2       1.0        1    1
## 6:    Private room              2       1.5        1    1
##
amenities
## 1:
{TV,"Cable TV",Internet,Wifi,"Air conditioning","Wheelchair
accessible",Pool,Kitchen,"Free parking on premises",Gym,Elevator,"Hot
tub",Heating,"Family/kid friendly",Washer,Dryer,"Smoke detector","Carbon
monoxide detector","Fire extinguisher",Essentials,Shampoo,"24-hour check-
in",Hangers,"Hair dryer",Iron,"Laptop friendly workspace","Hot water","Long
term stays allowed","Host greets you"}
## 2:
{TV,"Cable TV",Internet,Wifi,Kitchen,"Free street parking","Buzzer/wireless
intercom",Heating,"Family/kid friendly",Washer,Dryer,"Smoke detector","Carbon
monoxide detector","First aid kit","Safety card","Fire
extinguisher",Essentials,Shampoo,Hangers,"Hair dryer",Iron,"Laptop friendly
workspace","Private entrance","High chair",Crib,"Hot water",Other}
## 3: {TV,Internet,Wifi,"Air conditioning",Kitchen,"Free street
parking",Heating,Washer,Dryer,"Smoke detector","Carbon monoxide
detector","First aid kit","Fire extinguisher",Essentials,Shampoo,"24-hour
check-in",Hangers,"Hair dryer",Iron,"Laptop friendly workspace","translation
missing: en.hosting_amenity_49","Self check-in",Keypad,"Private
entrance","Room-darkening shades","Hot water","Bed linens","Extra pillows and
blankets",Microwave,"Coffee maker",Refrigerator,Dishwasher,"Dishes and
silverware","Cooking basics",Oven,Stove,"BBQ grill","Patio or
balcony","Luggage dropoff allowed","Wide hallway clearance","Wide
doorway","Well-lit path to entrance","Lake access","Handheld shower head"}
## 4:
{TV,"Cable TV",Internet,Wifi,Kitchen,"Free parking on premises","Smoking
allowed","Free street parking",Heating,Washer,Dryer,"Smoke detector","Carbon
monoxide detector","Safety card",Essentials,Shampoo,Hangers,"Hair
dryer",Iron,"Self check-in",Lockbox,"Private entrance",Bathtub,"Hot
```

water","Bed linens","Extra pillows and blankets","Coffee
maker",Refrigerator,Dishwasher,"Dishes and silverware","Cooking
basics",Oven,Stove,"BBQ grill","Patio or balcony","Garden or
backyard","Luggage dropoff allowed","Long term stays allowed","Step-free
access","Flat path to front door","Well-lit path to entrance","Handheld
shower head"}
5:
{Internet,Wifi,Kitchen,"Free parking on premises","Pets live on this
property",Dog(s),"Indoor fireplace",Heating,Washer,Dryer,"Smoke
detector","Carbon monoxide detector","First aid kit","Safety card","Fire
extinguisher",Essentials,Shampoo,"24-hour check-in",Hangers,"Hair
dryer",Iron,"Laptop friendly workspace","translation missing:
en.hosting_amenity_49","translation missing: en.hosting_amenity_50","Private
entrance"}
6:
{TV,"Cable TV",Internet,Wifi,"Air conditioning",Kitchen,"Paid parking off
premises",Heating,"Smoke detector","Carbon monoxide detector","First aid
kit","Safety card","Fire extinguisher",Essentials,Shampoo,"Lock on bedroom
door","24-hour check-in",Hangers,"Hair dryer",Iron,"Laptop friendly
workspace","translation missing: en.hosting_amenity_49","translation missing:
en.hosting_amenity_50","Self check-in",Lockbox,"Private living room","Hot
water","Bed linens","Extra pillows and blankets",Microwave,"Coffee
maker",Refrigerator,"Dishes and silverware","Luggage dropoff allowed"}
price security_deposit cleaning_fee minimum_nights maximum_nights
1: 157   500   130   1   365
2: 150   400   125   3   120
3: 150   400   NA   3   10
4: 159   200   140   2   730
5:   99   185   NA   30   200
6:   99   250   20   4   14
availability_365 number_of_reviews review_scores_rating
## 1:   134   76   94
2:   315   17   94
3:   164   237   97
4:   0   18   97
5:   0   143   98
6:   0   702   95
cancellation_policy   Area   neighbourhood
1: strict_14_with_grace_period Seattle   Downtown
2: strict_14_with_grace_period Seattle Other neighborhoods
3: strict_14_with_grace_period Seattle Other neighborhoods
4:   moderate Seattle   Downtown
5:   moderate Seattle Other neighborhoods
6: strict_14_with_grace_period Seattle   Downtown

Figure 2. Airbnb Price Distribution in Seattle ($0 ~$1000)
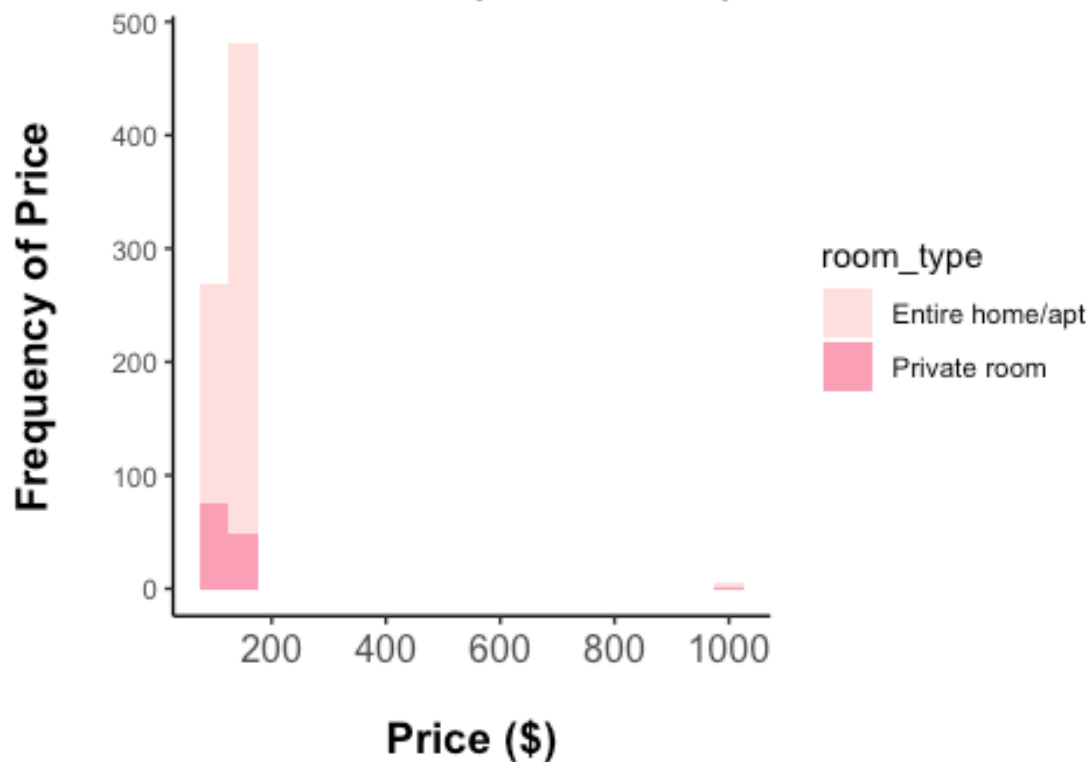
Other Features

There are many features that are relevant to the Airbnb rental price such as neighbourhood, number of bedrooms and number of accommodates.

Neighbourhood

Speaking of housing, most of the rental prices depend on the geographic feature. Houses in downtown are more expensive than those in suburb area. Further, the distance to shopping center or traffic station is also inneggligible if we try to list a fair rental price.

Leaflet mapping below shows the prices and names of different rental rooms in different neighbourhoods.

To give a better understanding of the price in different neighbourhoods, I plotted a heatmap (1) to show the average room prices in different neighbourhoods with different room types.

Next, from the heatmap(2), with the increase in the number of bedrooms, the average price of Airbnb listings in Seattle increase. Although it depends upon the neighbourhoods as well.
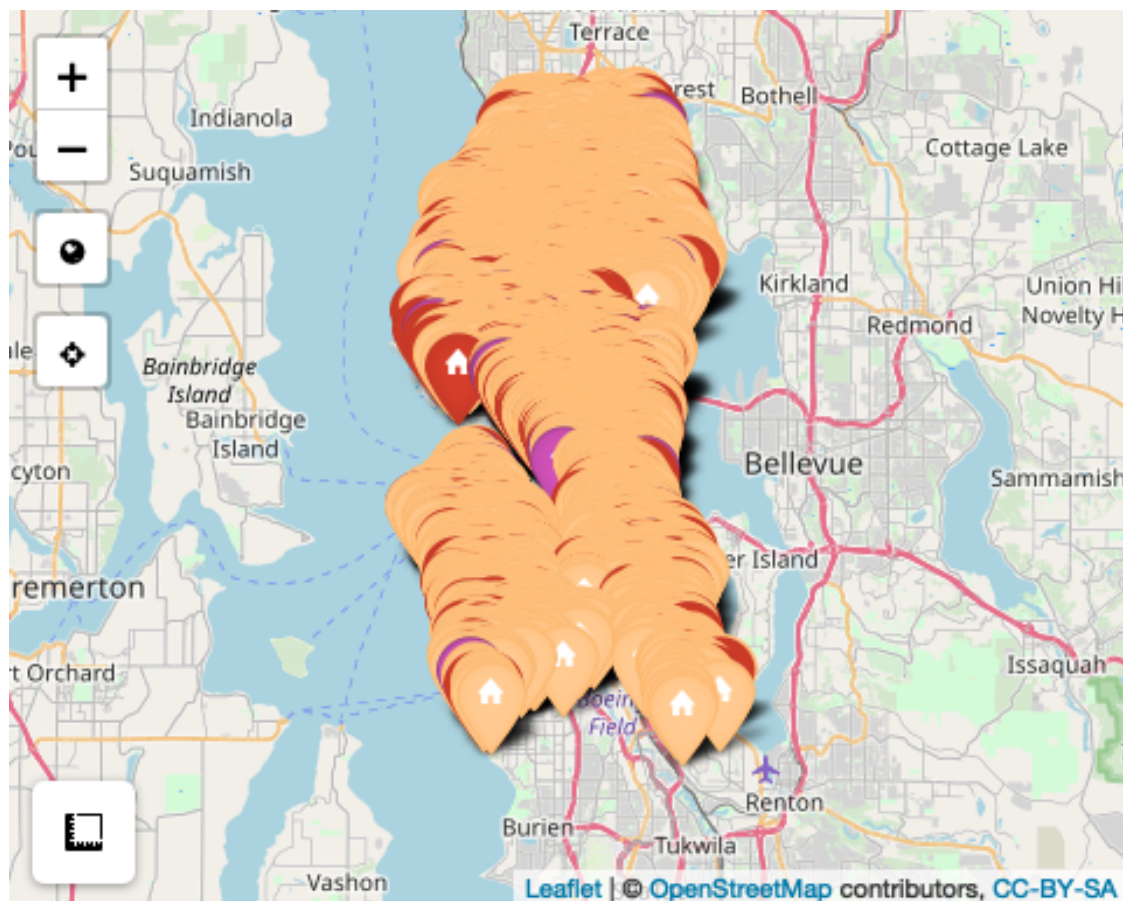
1. Neighbourhood ~ Room type

Neighbourhood	Entire home/apt	Private room	Shared room
West Seattle	169	78	
University District	135	82	43
Seward Park	166	76	
Rainier Valley	162	61	115
Queen Anne	200	91	73
Other neighborhoods	154	74	69
Northgate	133	72	
Magnolia	207	85	
Lake City	145	82	72
Interbay	129	45	26
Downtown	211	112	71
Delridge	126	68	55
Central Area	170	83	54
Cascade	187	108	
Capitol Hill	171	98	34
Beacon Hill	135	71	34
Ballard	165	71	71

Average Price

200
150
100
50

Room Type

2. Neighbourhood ~ Bedroom

Neighbourhood	0.0		2.5		5.0		7.5		10.0
West Seattle	94	97	163	257	300	432			
University District	95	82	161	191	332	450	899	29	600
Seward Park	114	81	140	233	286	340	440		
Rainier Valley	91	73	159	211	326	458	680		
Queen Anne	115	120	210	319	344	516	512	541	
Other neighborhoods	104	90	148	223	307	436	456	450	
Northgate	88	77	131	190	306				
Magnolia	103	99	161	310	382	899			
Lake City	69	86	168	251	276				
Interbay	76	103	160	140					
Downtown	149	187	282	431					
Delridge	77	70	117	198	195	190			
Central Area	100	96	162	250	347	544	725	650	
Cascade	158	167	214	295	424				
Capitol Hill	115	114	189	278	387	460	650	650	
Beacon Hill	139	80	126	179	217	310			
Ballard	98	94	175	249	328	365	891		

**Number of bedroom**

Average Price

- 750
- 500
- 250

Number of Bedrooms

Number of bedrooms will always be a major factor to the rental price of Airbnb. I made a barplot to show the average prices over different number of bedroom. In this case, we can see from Figure 2 that there is a certain relationship between the average price of Airbnb listings and number of bedrooms.
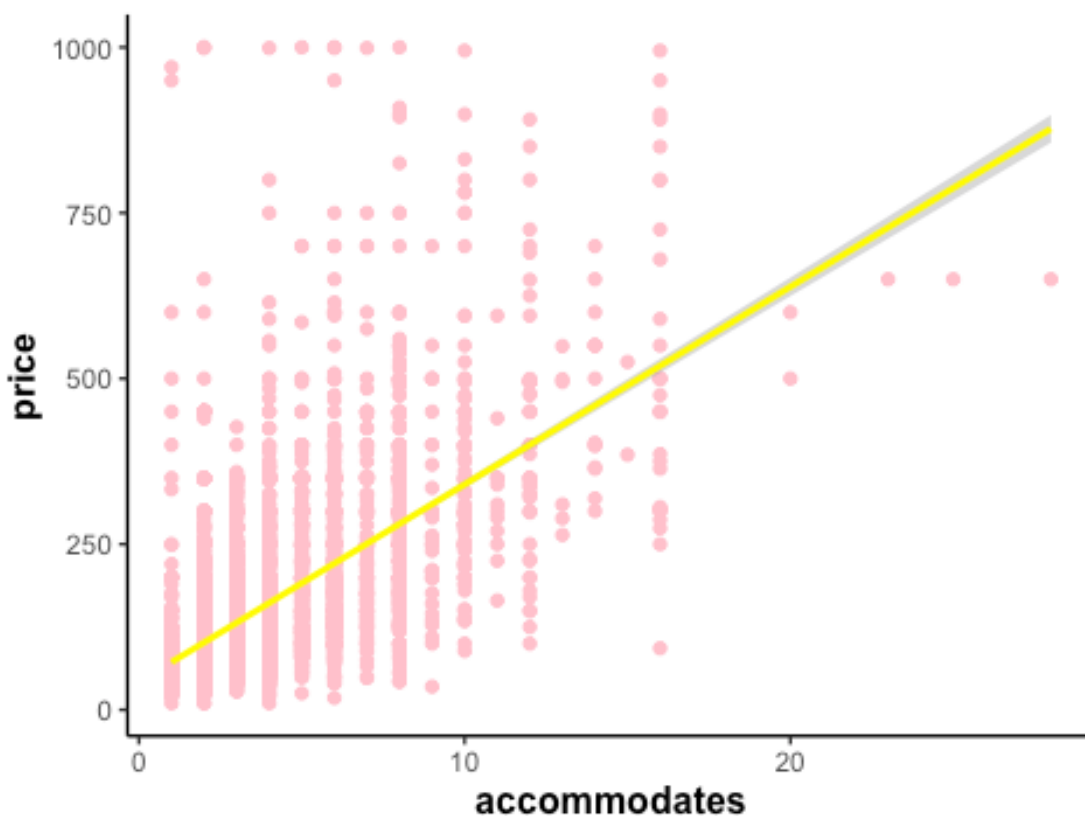
Figure 2. Price over different number of bedrooms
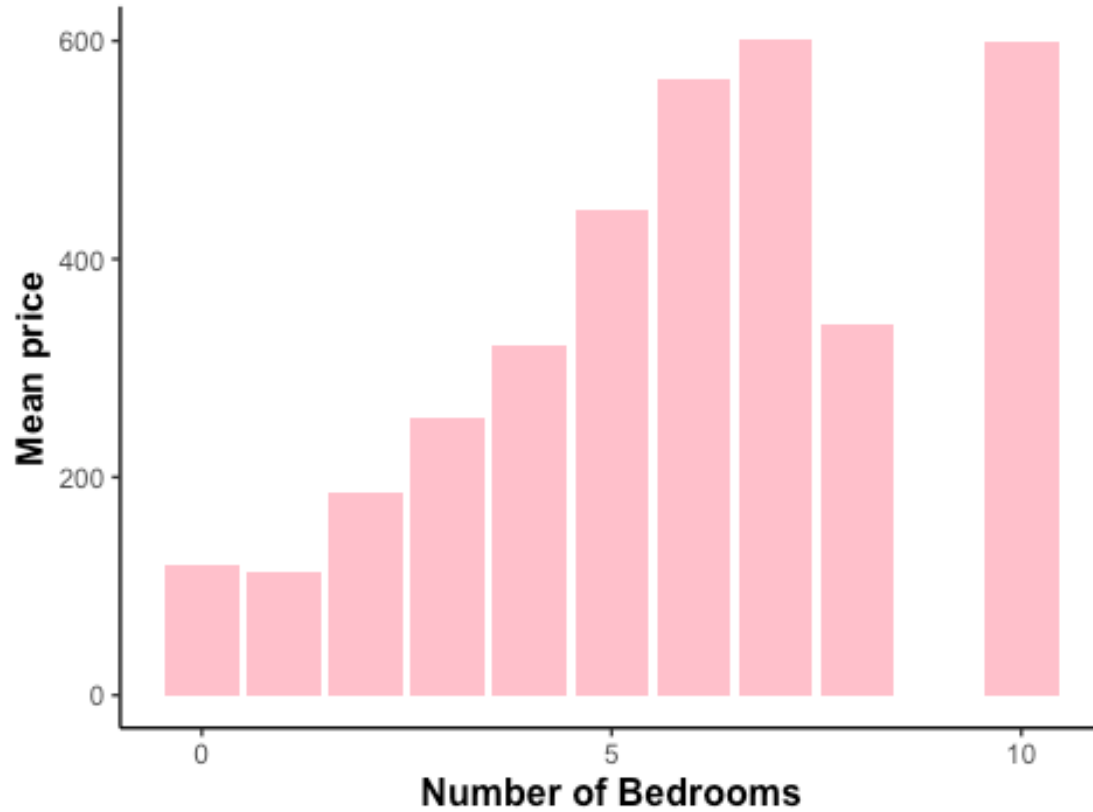
Number of Accommodats

We can see thatnumber of accommodates have a certain effect on the rental price of Airbnb listings.

Figure 3. Price over Number of Accommodates

Room Type

In Airbnb listings, there are three types of room: Entire home/apt, Shared room and Private room. Room type is also a major factor of the price. Figure 4 shows the proportion of each of these three room type and we can see that there are nearly 74% of rooms are 'Entire Room/Apt'. Further, Figure 5 below shows that 'Entire room/apt' have a higher average price than 'Shared room' and 'Private room'.

Figure 4 Room Type Proportion

Legend:
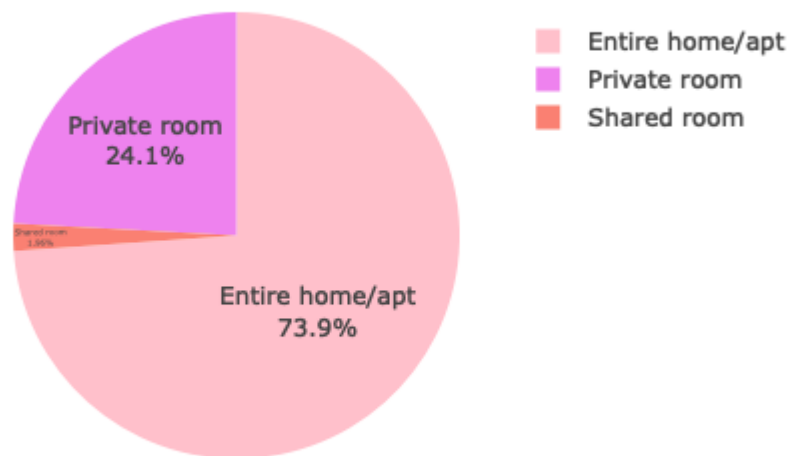- Entire home/apt
- Private room
- Shared room

Private room
24.1%

Shared room
1.96%

Entire home/apt
73.9%

Figure 5 Price ~ Room type

price($)

Different Room Type

room_type
- Entire home/apt
- Private room
- Shared room