

# BOSTON AIRBNB ANALYSIS

*EDA & Benford Analysis*

# Why I Chose Boston Airbnb Data ?



I have been using Airbnb for over three years and it has become a popular way of travelling. I have witness Airbnb develop from an unknown website to the most popular travelling website during the past several years. Many people choose Airbnb instead of hotels not only for its lower price and convenient location, but also for its humanness – travelers are able to make connections with people from all around the world. What's more, travelers are provided with more unique options compare to hotels - houses, condos, apartments, castles, houseboats, tree houses, barns, mansions, even caves!

Therefore, these unique properties of Airbnb inspired me to explore more about it. For example, what the factors may have an impact on the ratings, or, what is the relationship between the occupancy rate and the neighborhood of an Airbnb apartment, etc. What's more, I can also test if the reviews and the price follow the Benford Law.

# What is the goal of this project ?

1. Discover the factors (neighborhoods, room type etc.) that may have impact on ratings and price

2. Test if price and ratings follow Benford Law



## Future Implications

For host :  
To have a better understanding of customers' preference and taste.

For customers :  
To understand what ratings and price represents and help to choose a better place

# Description of Datasets

Dataset Source from Airbnb dataset website (<http://tomslee.net/airbnb-data-collection-get-the-data>). The data used in this project is a single survey for Boston with 3,864 observations with 14 variables (from 11/21/2016 dataset).

# Methodology for the study

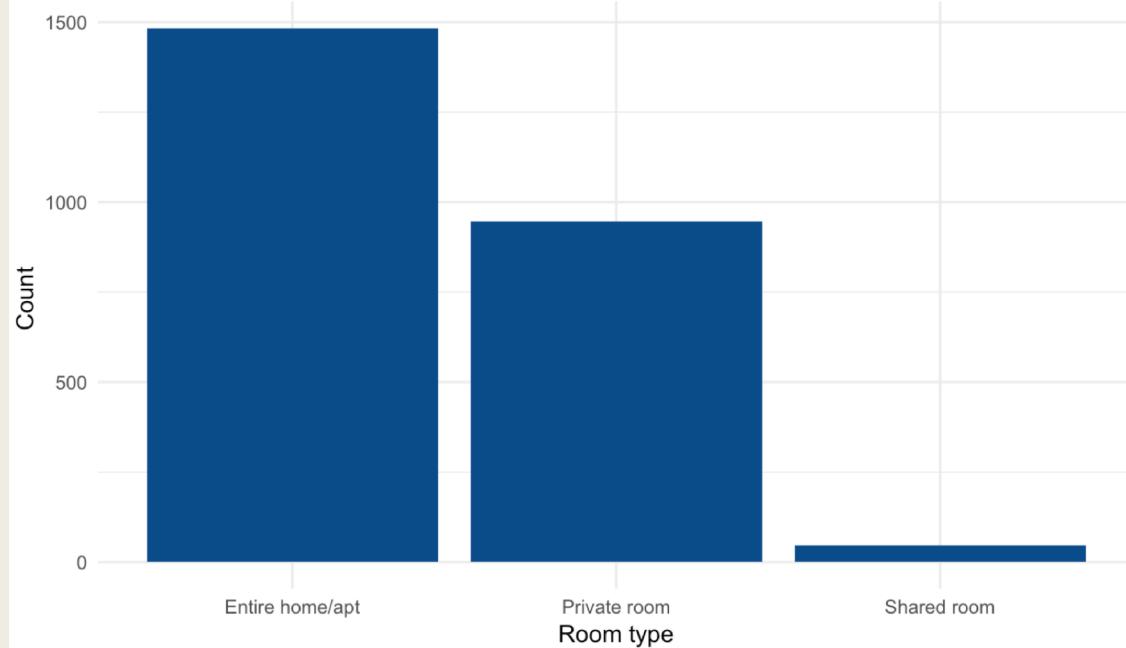
I will apply EDA, Benford Analysis to study the distribution of ratings within neighborhood and cross neighborhood, and explore other factors that may have significant impact on the ratings.

After fitting the model, I will assess the model fit by conducting residual and deviance analysis, check outliers, and measure parameter significance.

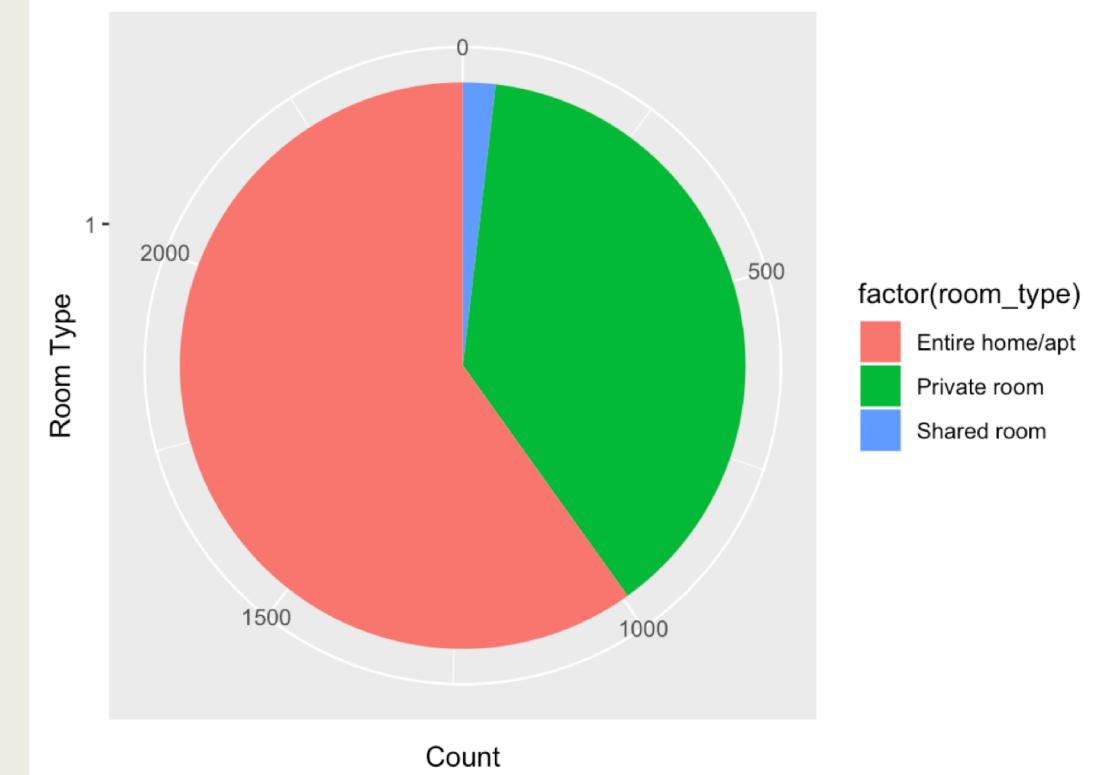
Names	Description
room_id:	A unique number identifying an Airbnb listing.
host_id	A unique number identifying an Airbnb host.
room_type	One of “Entire home/apt”, “Private room”, or “Shared room”
borough	A sub-region of the city or search area for which the survey is carried out. For some cities such as Boston, there is no borough information.
reviews	The number of reviews that a listing has received. The number of reviews can be used to estimate the number of visits.
overall_satisfaction	The average rating that the owner of the property has received.
accommodates	The number of guests a listing can accommodate.
bedrooms	The number of bedrooms a listing offers.
price	The price (in \$US) for a night stay. In early surveys, there may be some values that were recorded by month.
minstay	The minimum stay for a visit, as posted by the host.
Latitude and longitude	The latitude and longitude of the listing as posted on Airbnb web.

# EDA

Overview of Room Type

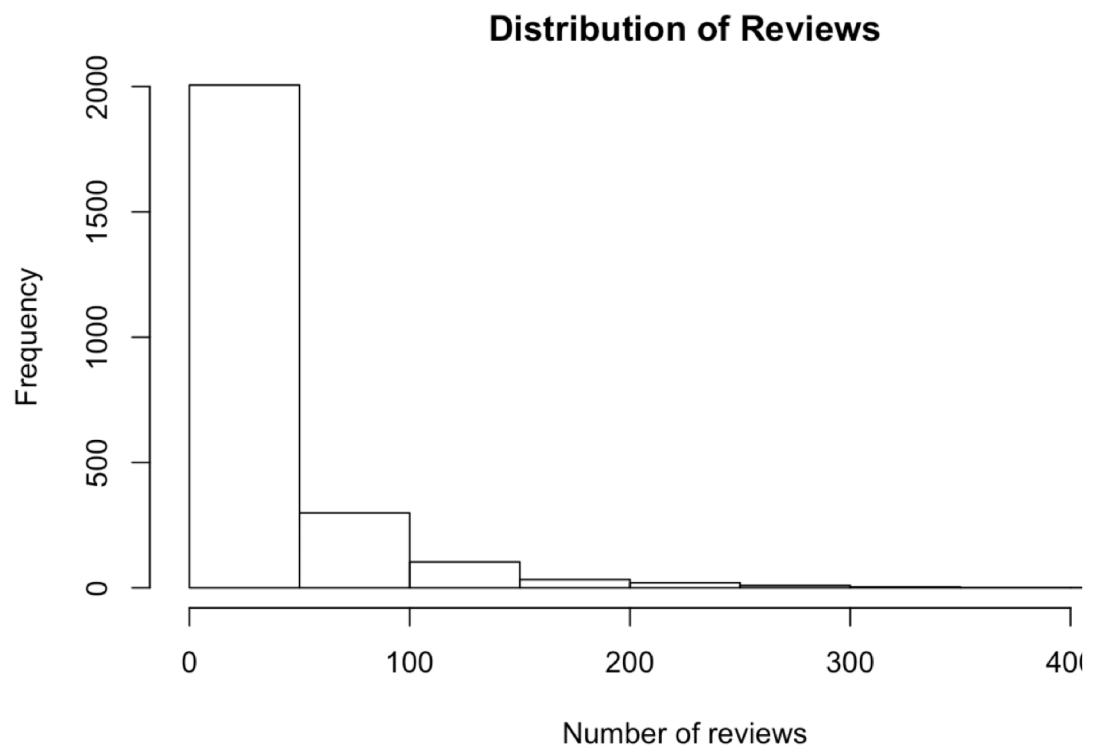


Room Type Distribution in Boston

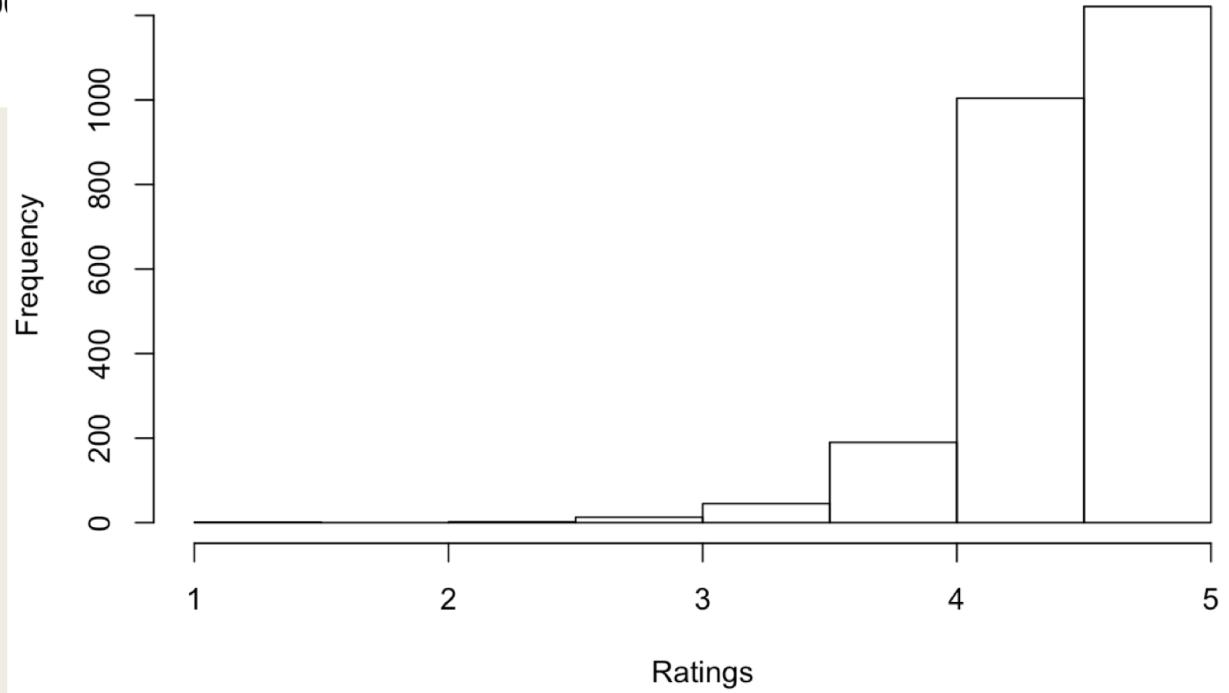


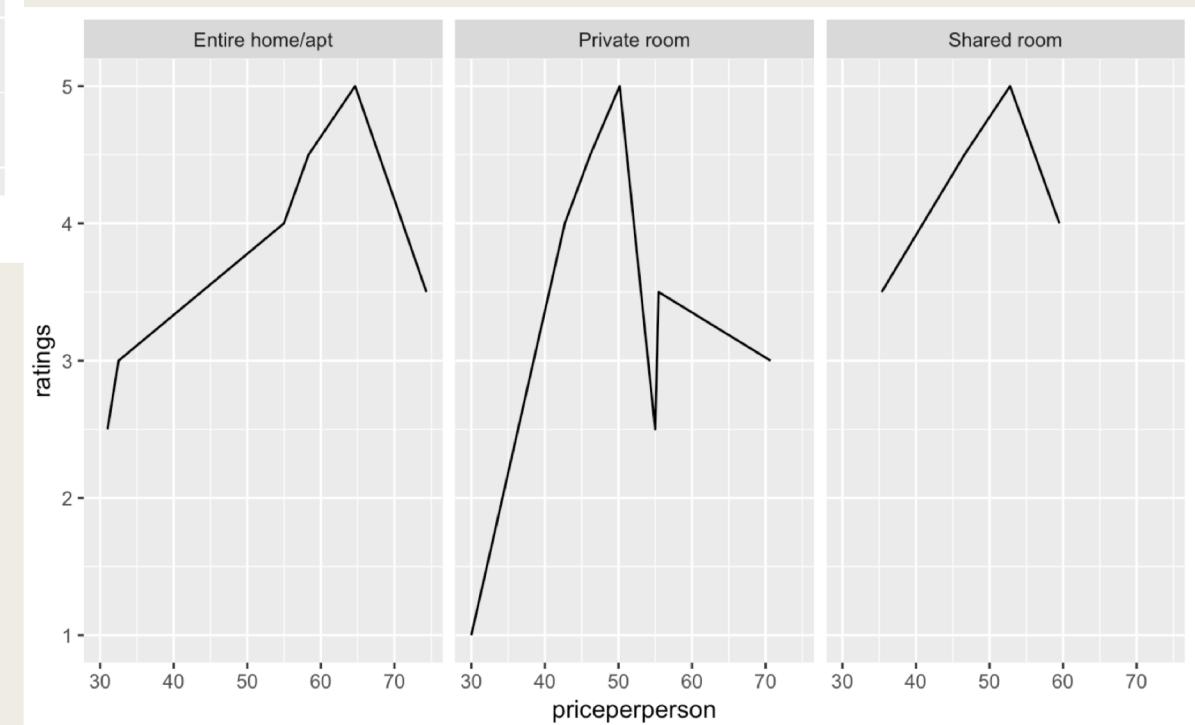
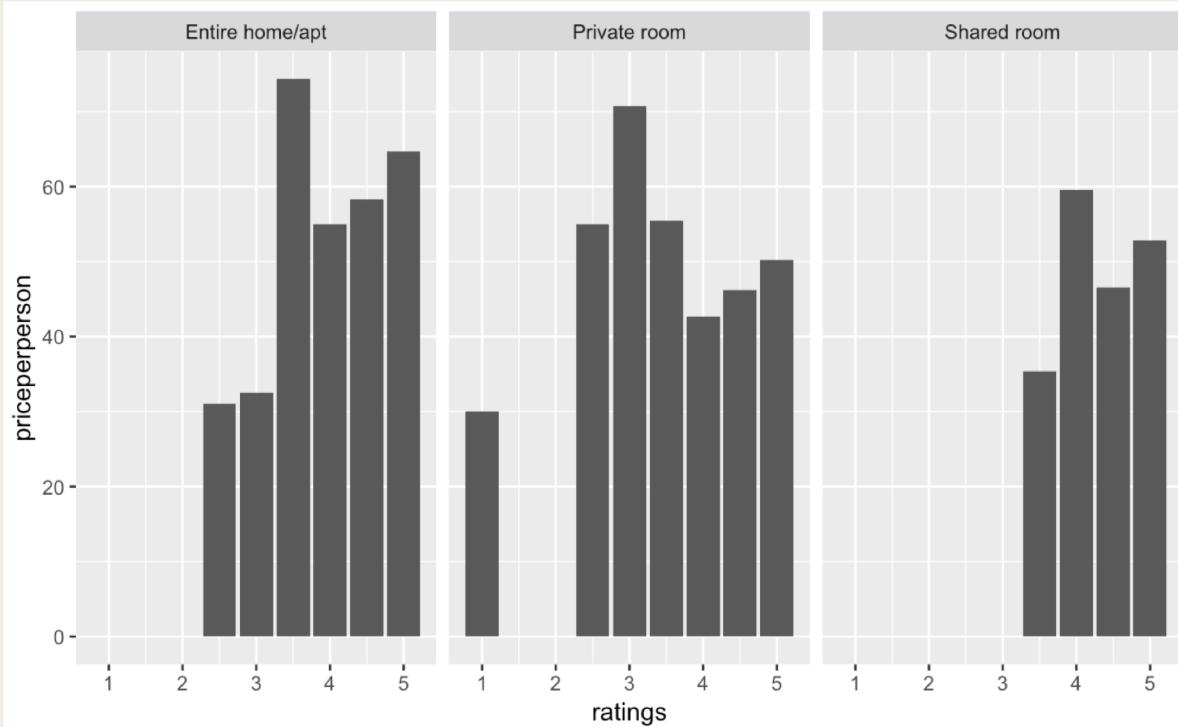
*Room Type Overview*

# *Frequency Distributions*

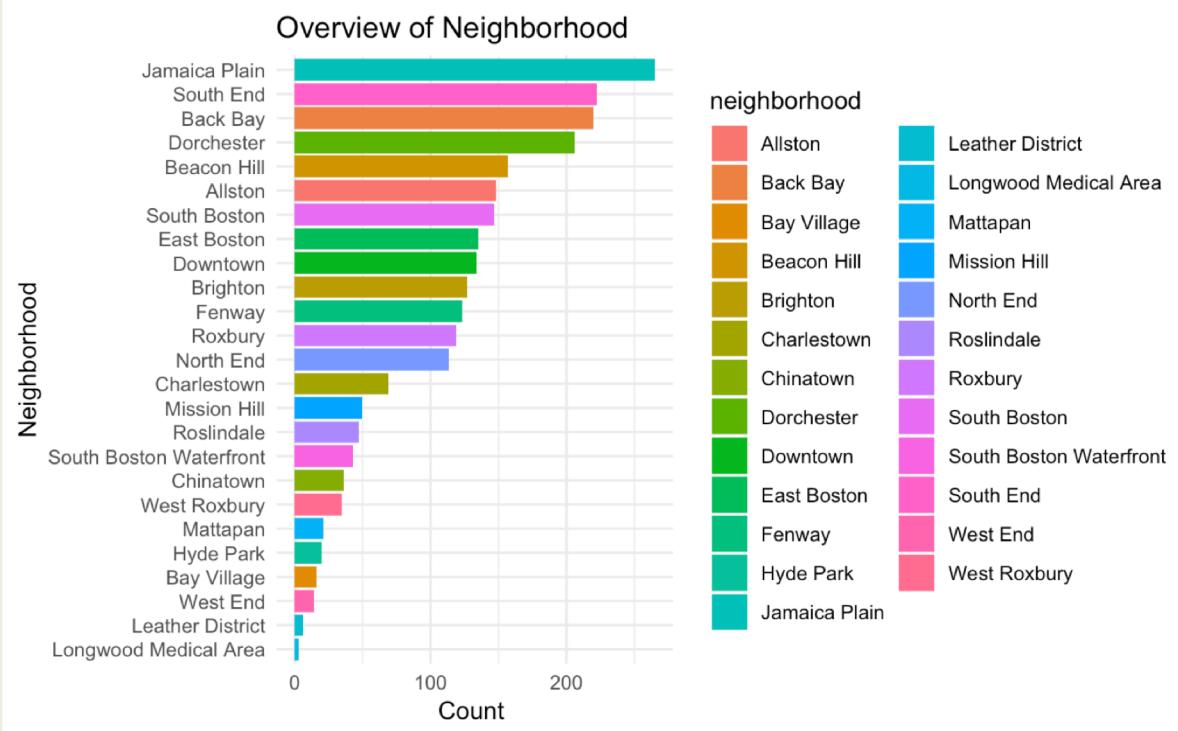


### Distribution of Ratings

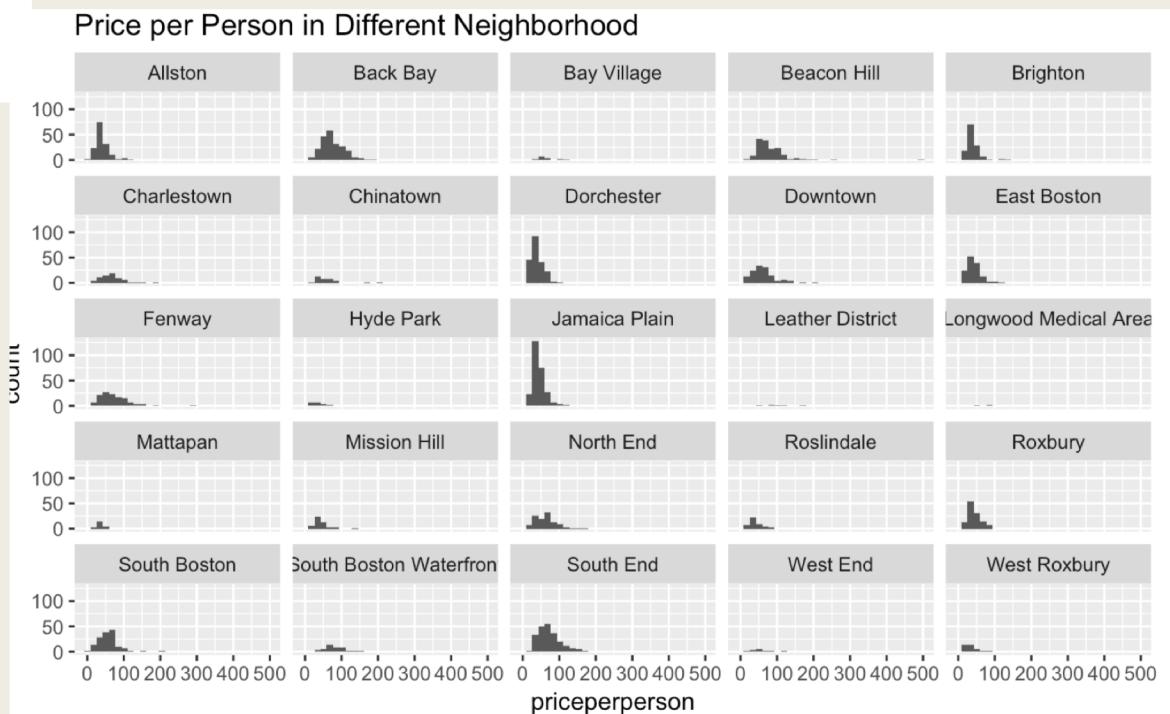


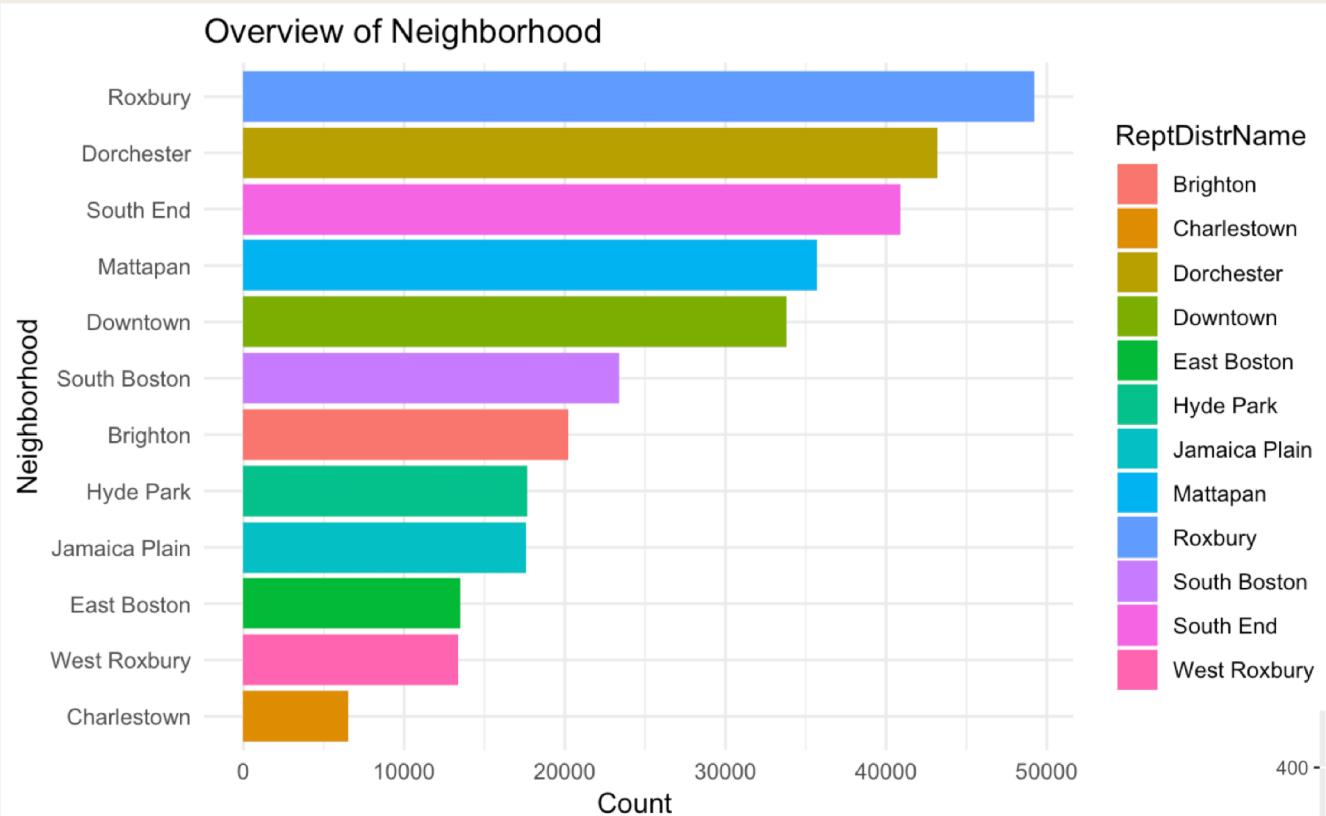


# <= Properties Overview in Boston Neighborhoods

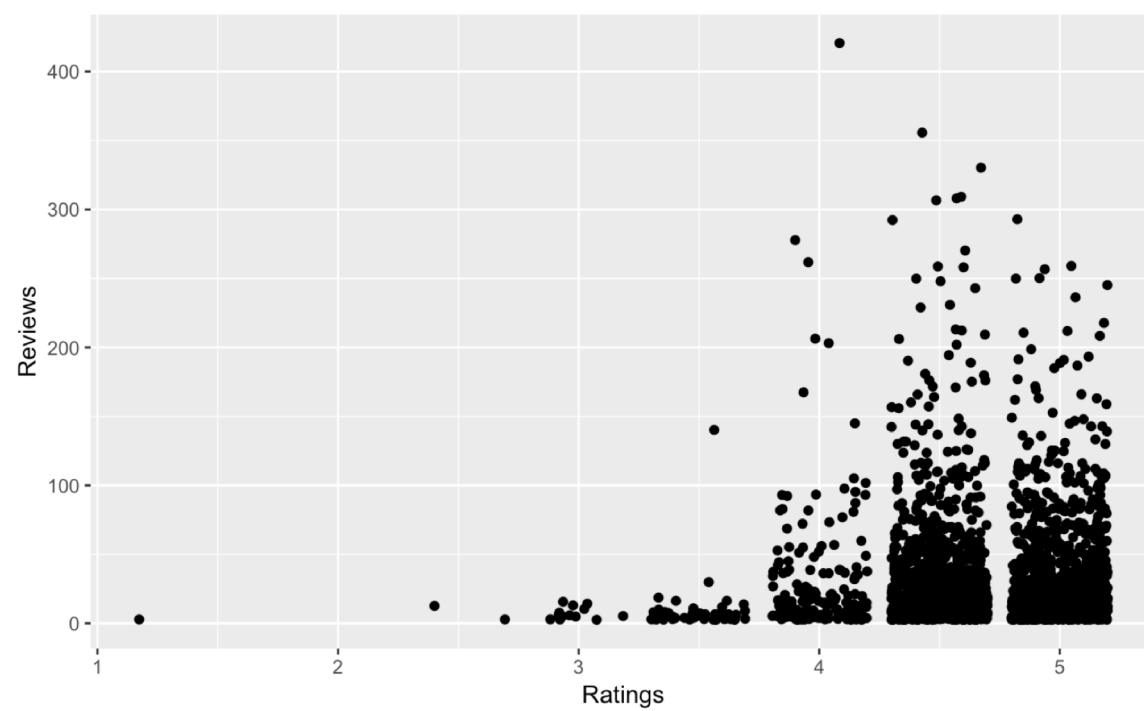


*Facet by Price per Person  
in Boston Neighborhoods =>*

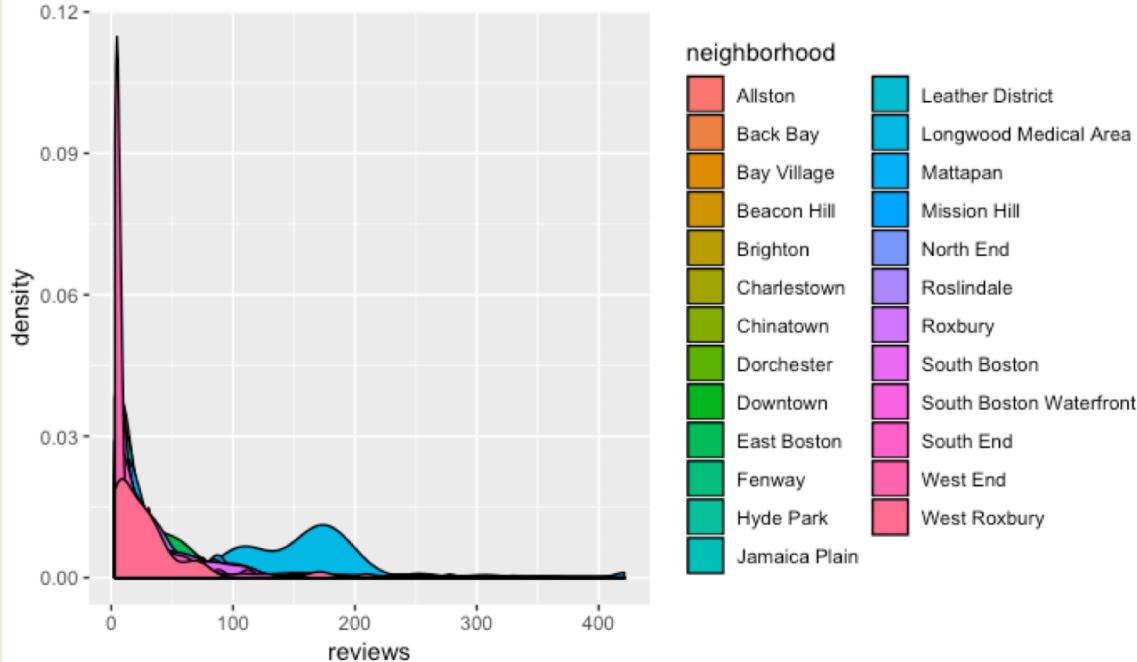




*<= Reported Crime Incidents  
in Boston Neighborhoods*

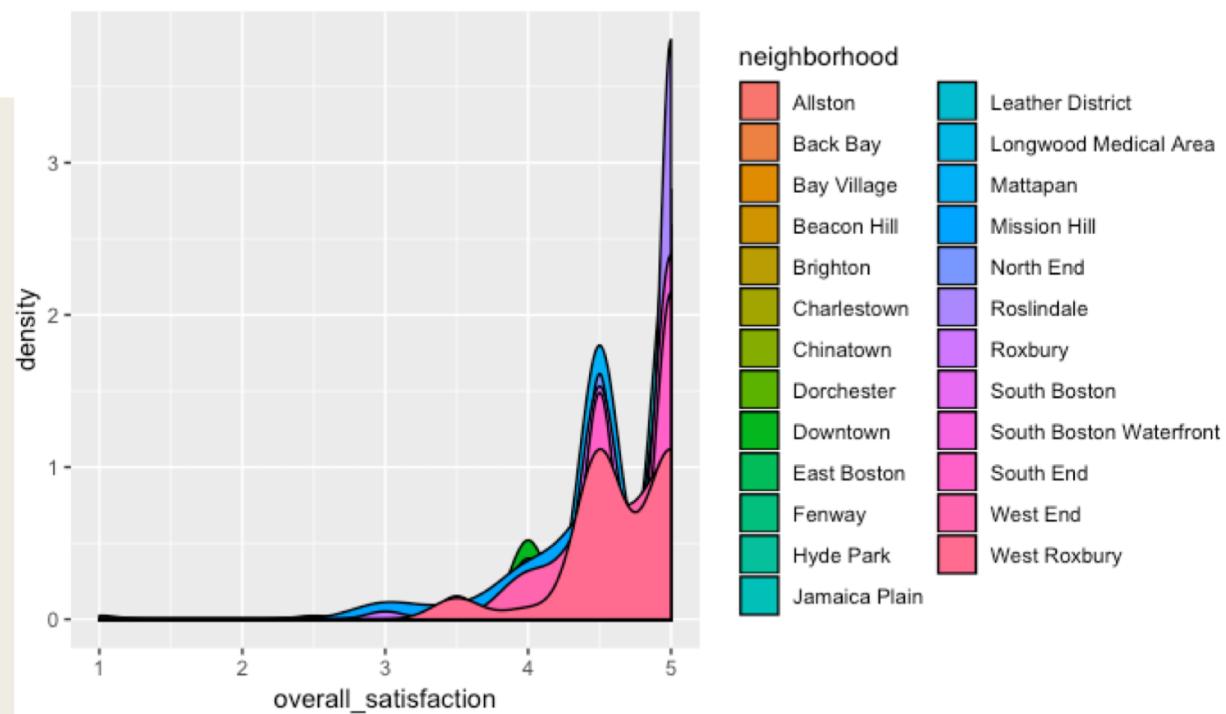


### Reviews Density Plot by Neighborhoods

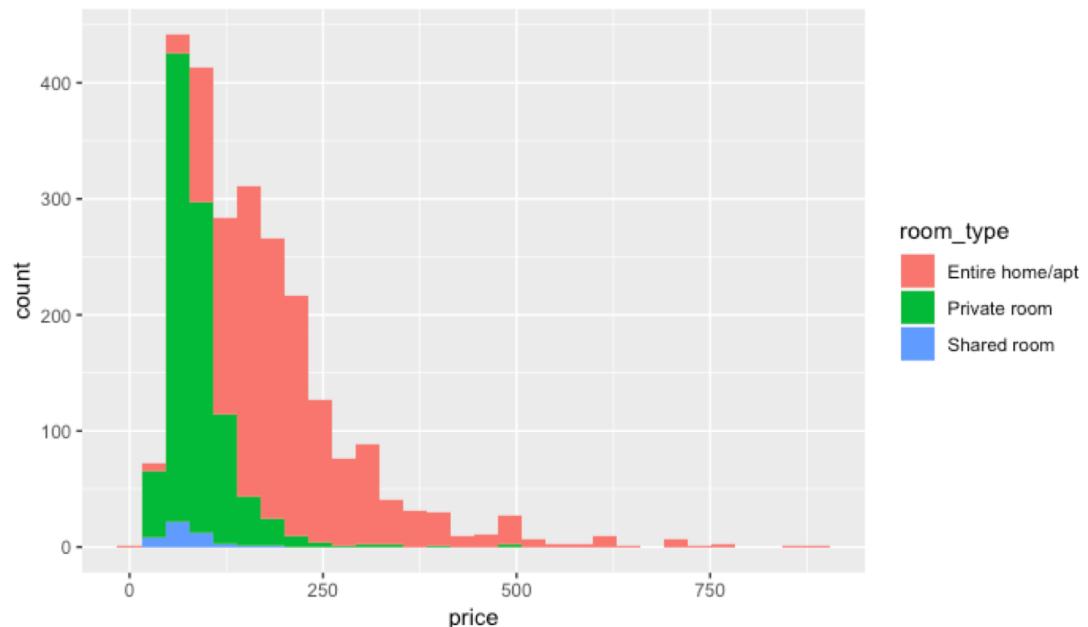


# Density Plot

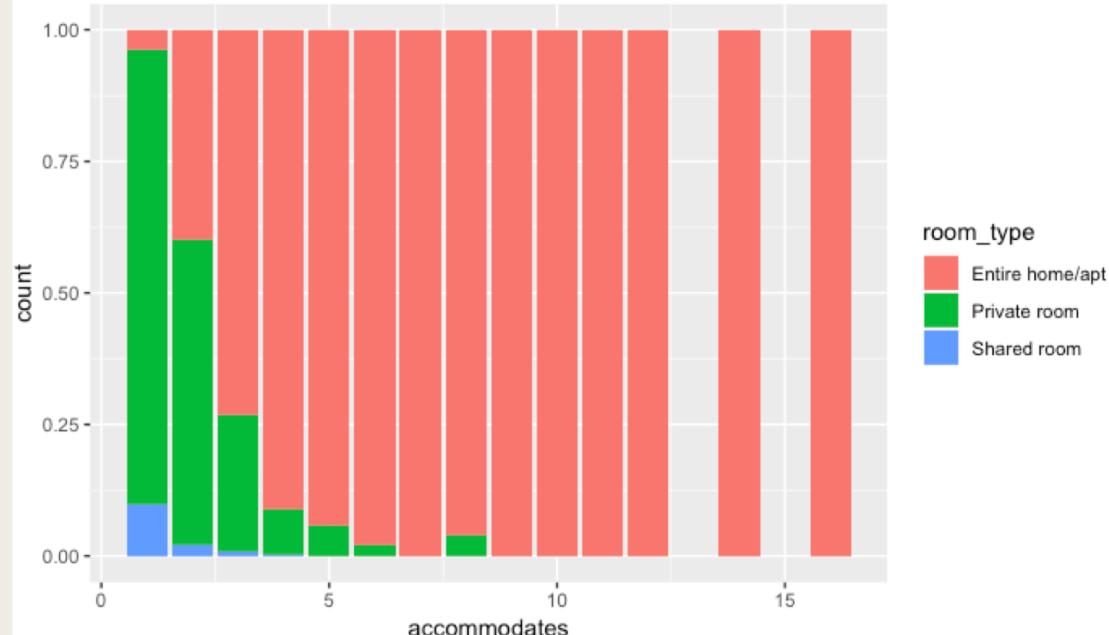
### Ratings Density Plot by Neighborhoods



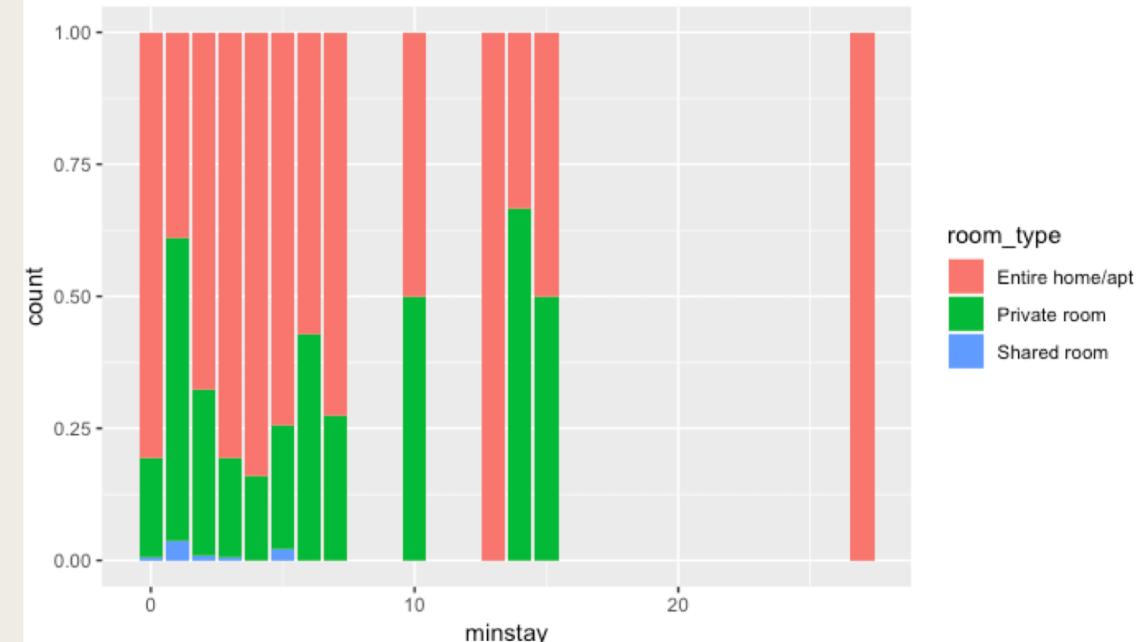
### Price and Room Type



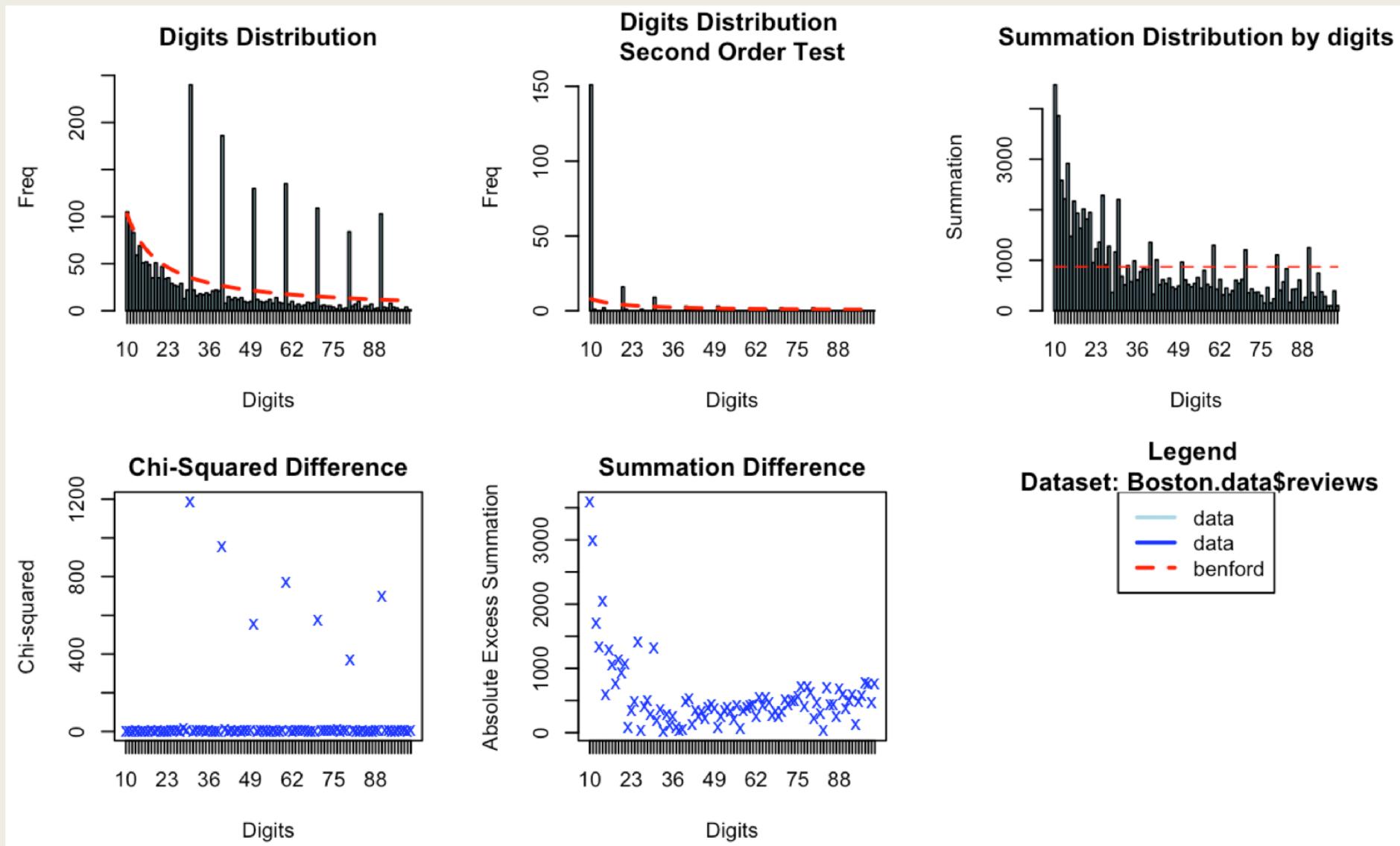
### Accommodates and Room Type



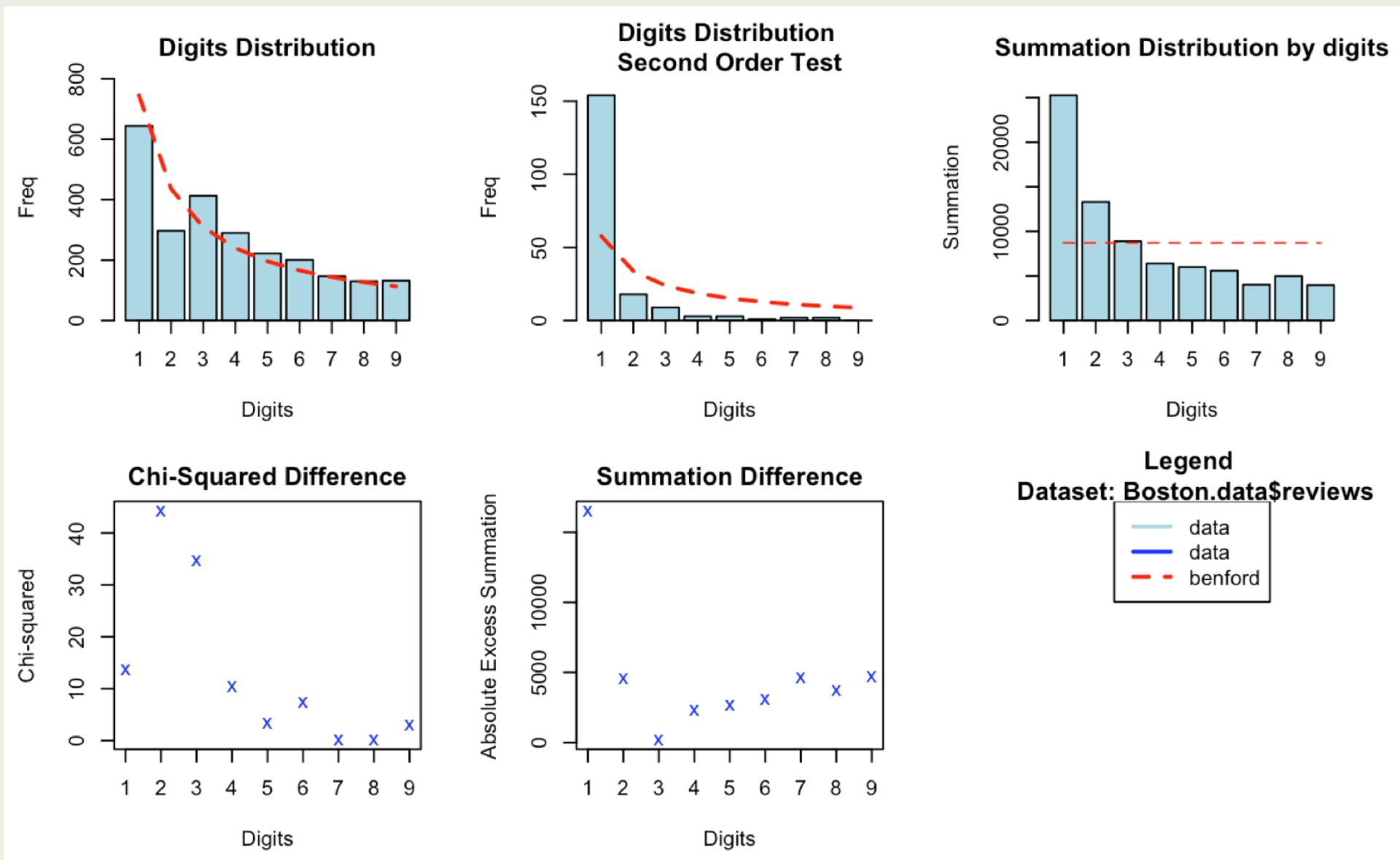
### Minstays and Room Type



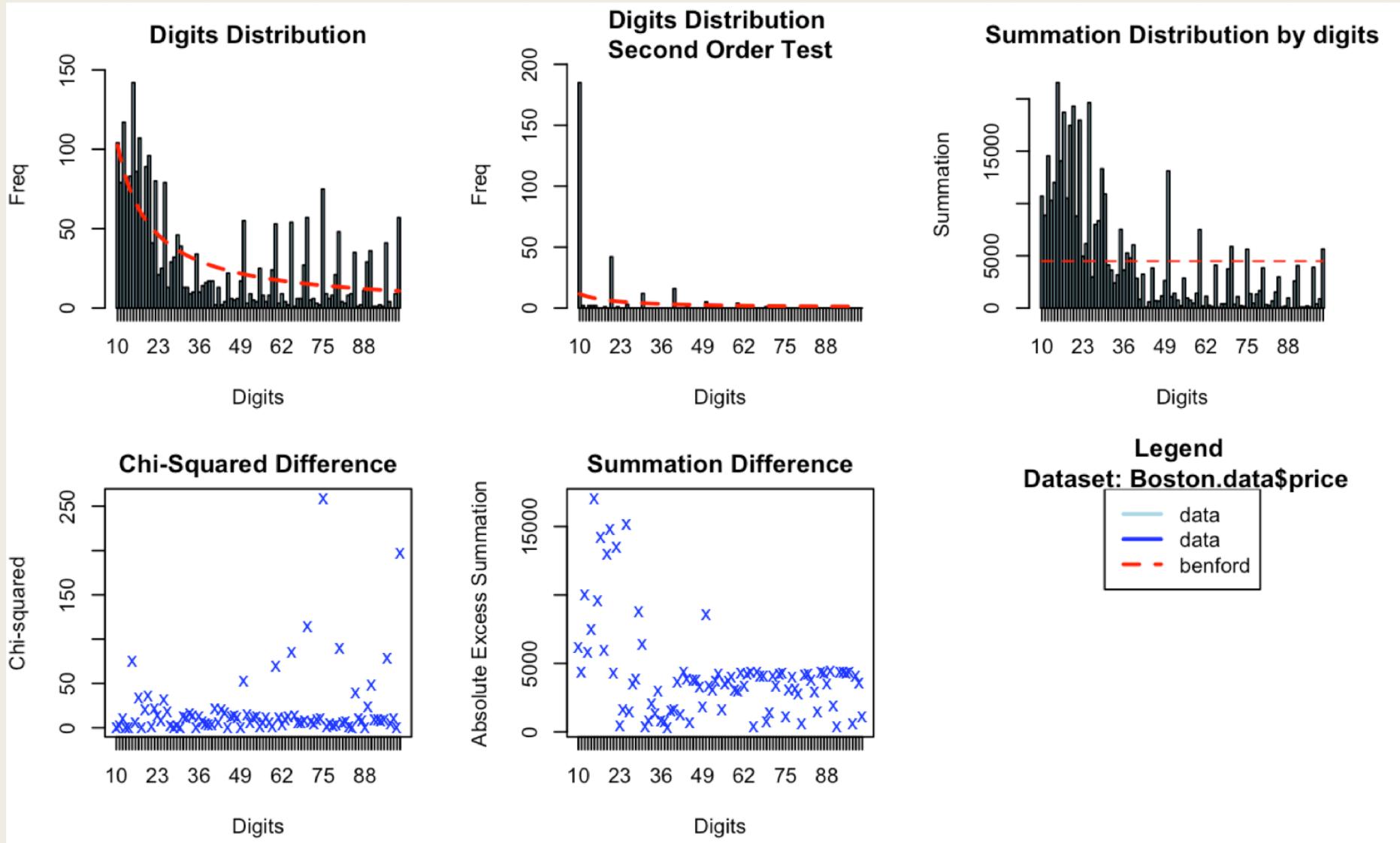
# Benford Analysis for Reviews (two digits)



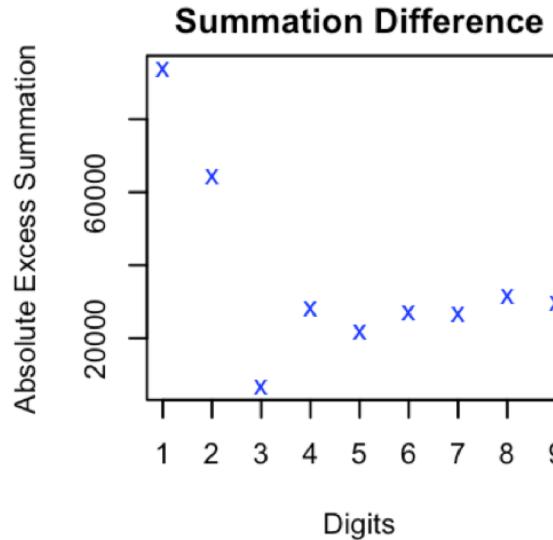
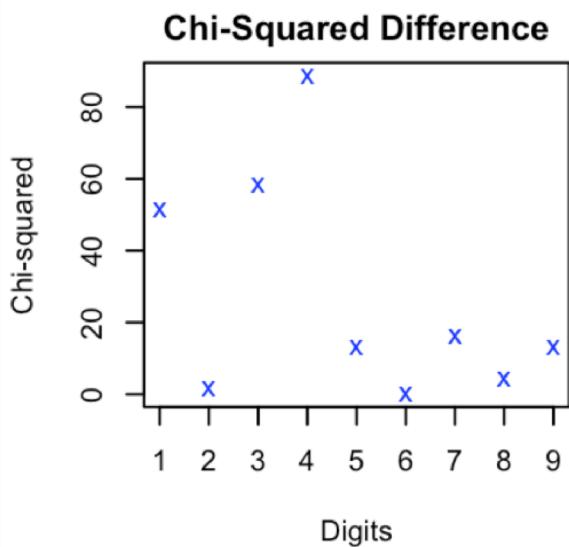
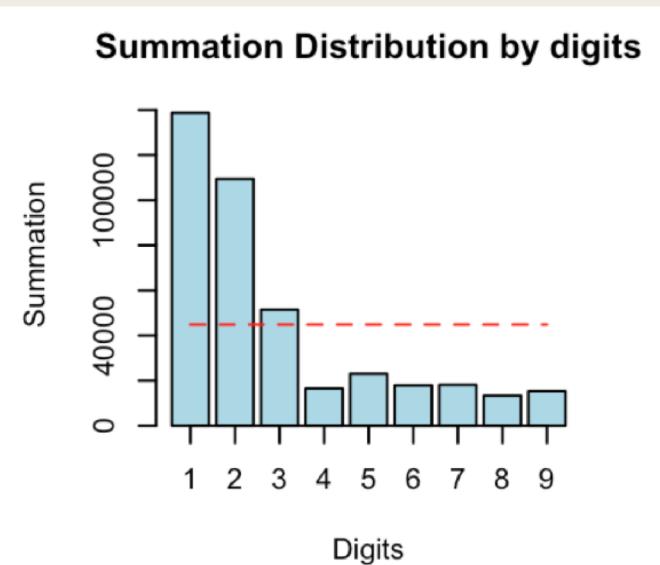
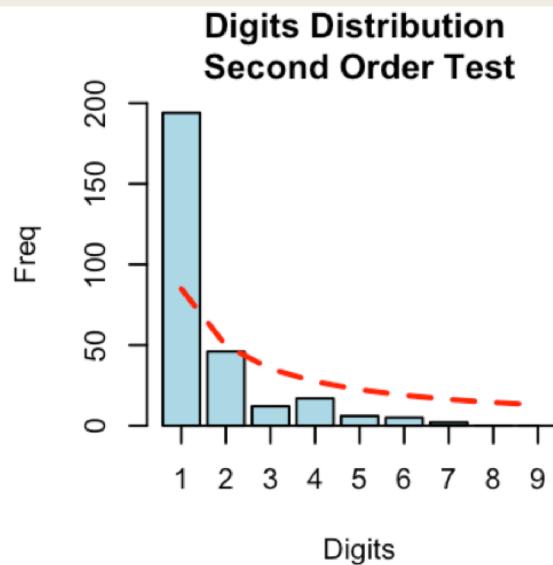
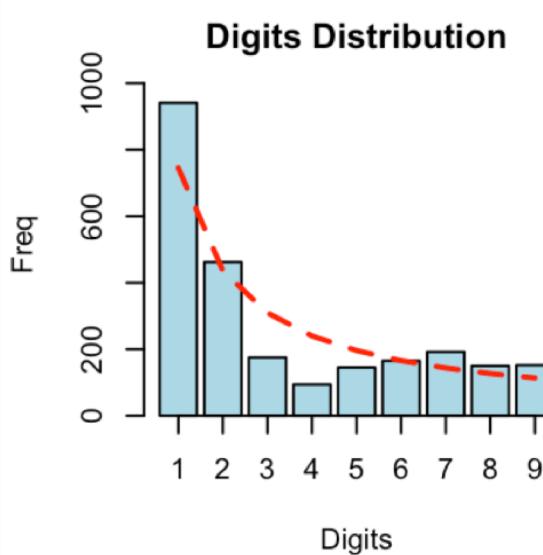
# Benford Analysis for Reviews (one digit)



# Benford Analysis for Price (two digits)



# Benford Analysis for Price (one digit)



**Legend**  
Dataset: Boston.data\$price

The legend box contains three entries:

- data (represented by a light blue bar)
- data (represented by a blue 'x')
- benford (represented by a red dashed line)

# Conclusion

In conclusion, we can say that the price and reviews in Boston Airbnb doesn't perfectly follow the Benford Law. However, it follows the trend in general. Accomodates and ratings have positive effect on the price, other variables have negative effect. Among all the variables, room type has the most significant effect on the price on Airbnb because it would cost a lot more when it comes to entire house/ apt than shared room or private room. Price, accomodates and neighborhood are the major factors that may influence the ratings on Airbnb. These all make sense since the neighborhood is related to crime rate in the area, properties in a nice place tend to have more high ratings. I think these are all useful points for Airbnb users or owners to know about.

# Shiny Dashboard Link

The Published Shiny Dashboard :

<https://shiyu95.shinyapps.io/benfordboston/>



Thanks !!