MEMORANDUM

To:	Carole Voulgaris, Course Instructor, SES 5215
From:	Chunfeng Yang, Student, SES 5215
Date:	November 1, 2022
Subject:	Assignment 2, Describing and visualizing data

The purpose of this memo is to present and describe the variables in the dataset I am using the answer the question:

What is the effect of the distance to the nearest gallery on its listing price for each night, after accounting for the effects of the distance to the nearest subway station and room type in Manhattan, New York?

Over this semester, I will be addressing this question using the dataset found on a mission-driven project “Inside Airbnb” that provides data and advocacy about Airbnb's impact on residential communities, which I accessed through their website (http://insideairbnb.com/explore/) as well as public amenities information on NYC OpenData (https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49). My dataset includes 21,598 Airbnb listings in Manhattan, New York. The variables in my data set are:

(1) Airbnb Listing Price: The expense of one night stay in an Airbnb listing.
(2) Vicinity of the Gallery: The distance from one Airbnb listing to its nearest gallery.
(3) Vicinity of the Subway Station: The distance from one Airbnb listing to its nearest subway station.
(4) Room Type: The types of listing an Airbnb host can offer guests.

Table 1 presents basic descriptive statistics for each continuous variable in the dataset.

Table 1: Descriptive Statistics For Continuous Variables

	Airbnb Listing Price (USD$)	Vicinity of The Gallery (Meter)	Vicinity of The Subway Station (Meter)
Full range	10-2000	2-3658	1-1658
Interquartile range	95-220	130-395	233-512
Standard deviation	164	290	228
Mean	186	312	398
Median	150	223	360

Listing price in the sample have a minimum of $10, and a maximum of $2,000. Half of the Airbnb listings in the sample have a listing price between $95 and $220, representing an interquartile range of $125, which is less than the standard deviation of $164. The median

value of $150 is less than the average value of $186, which suggests some left skew in the distribution, as illustrated in Figure 1.

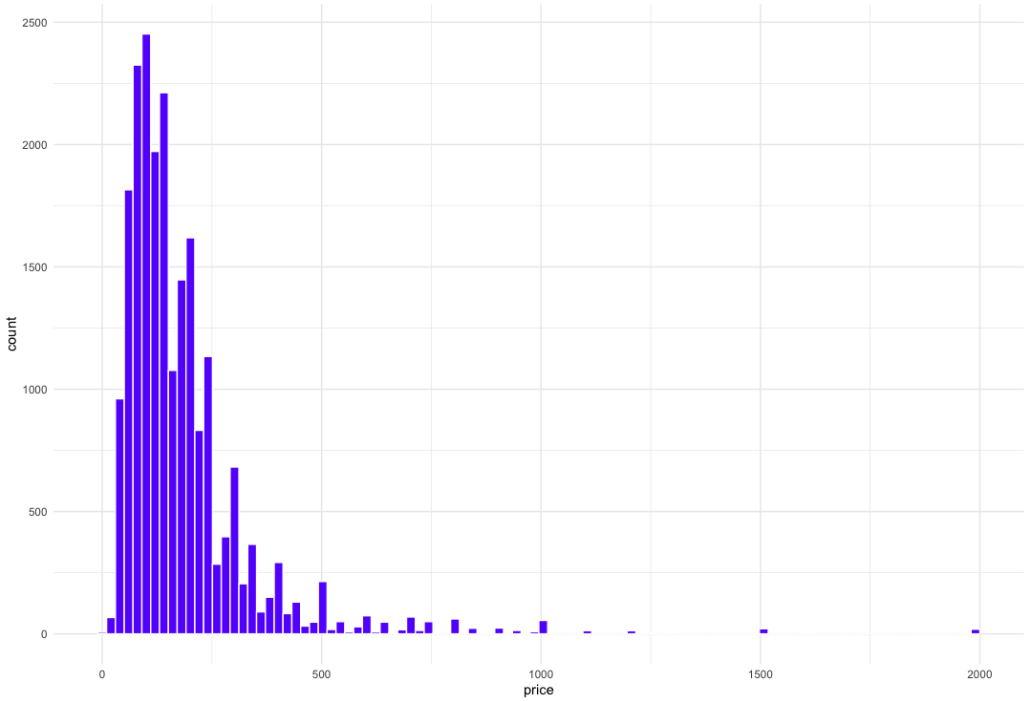
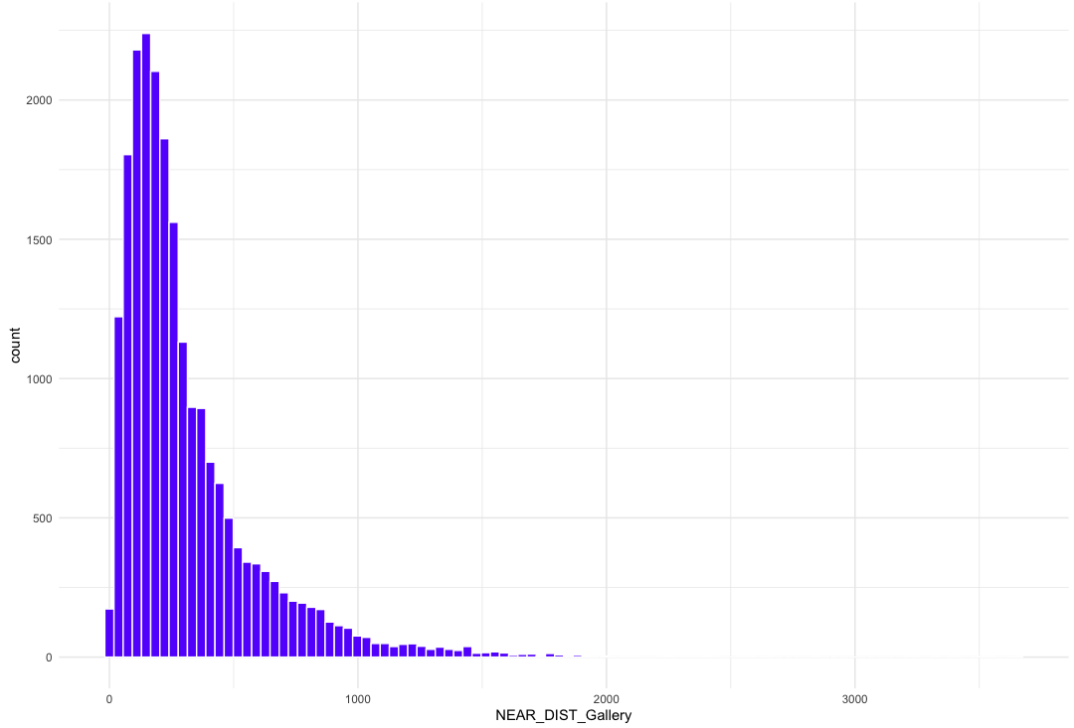


Figure 1: Distribution of Airbnb Listing Price

The distance from an Airbnb listing to its nearest gallery in the sample has a minimum of 2 meters and a maximum of 3,658 meters. Half of the Airbnb listings to their nearest galleries in the sample have distances between 233 and 512 meters, representing an interquartile range of 279 meters, larger than the standard deviation of 228 meters. The median value of 223 meters is less than the average value of $312, which suggests some left skew in the distribution, as illustrated in Figure 2.



Log transformation（try）

Figure 2: Distribution of Vicinity of The Gallery

The distance from an Airbnb listing to its nearest subway station in the sample has a minimum of 1 meter and a maximum of 1,658 meters. Half of the Airbnb listings to their nearest subway stations in the sample have distances between 130 and 395 meters, representing an interquartile range of 265 meters, slightly less than the standard deviation of 295 meters. The median value of 360 meters is less than the average value of $398, which suggests some left skew in the distribution, as illustrated in Figure 3.

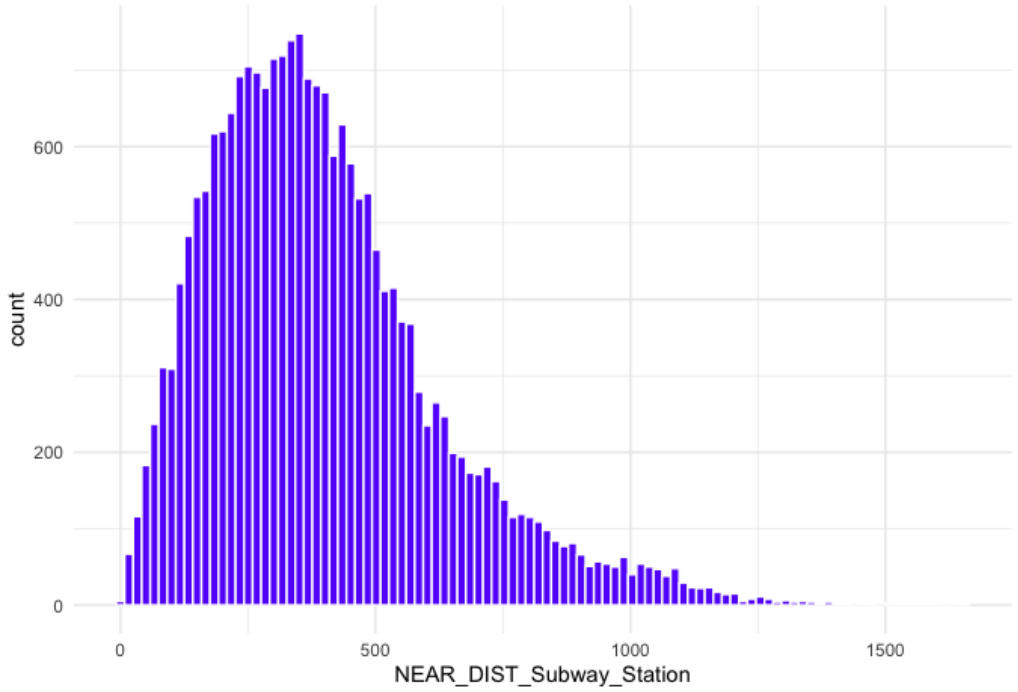


Figure 3: Distribution of Vicinity of The Subway Station

Figure 4 illustrates the sample proportions for different Airbnb listing room types. More than half of the listing sample (61 percent) is an entire home or apartment. The remaining listing sample is mostly private rooms (37 percent) only 2 percent of the listing sample is shared rooms.

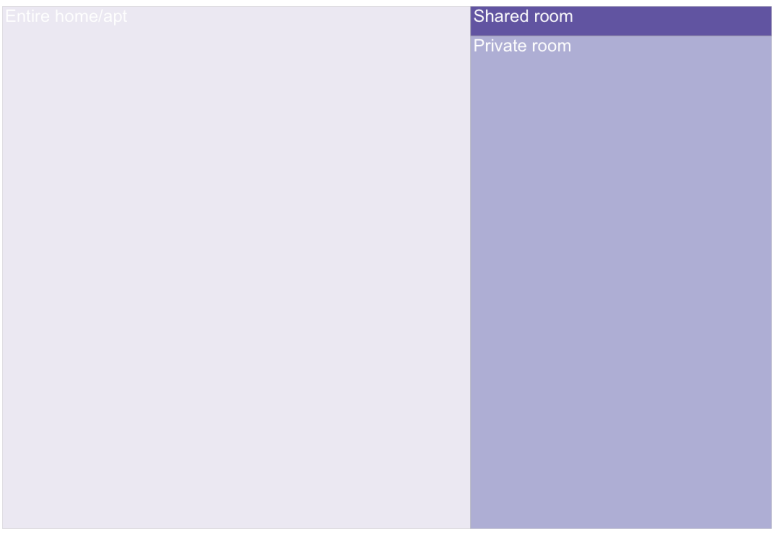


Figure 4: Proportions of Sample Listing By Room Type