Homework #5
Data Science I

For this assignment we will use 'listings.csv' and 'listings_extended.csv' to investigate data for AirBnB hosts in New York city. The data was accessed from http://insideairbnb.com/get-the-data.html on July 11, 2019. Please answer the questions below using the dataset. For all plots, make sure your plot has (1) labeled x and y axis, (2) a title, and (3) a legend when appropriate.

1. Use describe() to create a table that describes the price of AirBnBs in the 5 boroughs (the 5 neighbourhood groups). What borough has the highest average price? The lowest average price? Do you see any potential outliers in the dataset?
2. Make an ordered barplot of the median price in each borough.
3. Make an ordered bar plot of the median prices for all of the neighborhoods in the borough Manhattan. Which neighborhood has the highest median price and what is this price?
4. Make a barplot of the median prices in each bourough stratified by room type.
5. Make a plot of the location (latitude and longitude) of all of the AirBnB locations in New York city. Color the points in the plot by the bourough that the AirBnB belongs to. Make sure you have a legend identifying the colors of the bouroughs.
6. Make a table of the percentage of AirBnBs that are of each room type by Bourough. Plot this table as a stacked barplot
7. Count the number of listings for each individual 'host_id'. Show that the answer you got is the same as the column 'calculated_host_listings_count' in the original data.
8. Explore the listings for your favorite New York neighborhood (I picked the neighborhood that I live in!). Make a creative visualization that includes that latitude, longitude, and price for the AirBnBs in that neighborhood. Filter out any outliers. Make histograms of the prices of AirBnBs for the neighborhood that are stratified by the room type.
9. Read in 'listings_extended.csv'. Merge the dataset with 'listings.csv'. Now we have additional information about the listings, including the number of bedrooms that each listing has. Use the 'bedrooms' variable to make a plot of the number of bedrooms versus the median price stratified by borough. Filter out listings with more than 6 bedrooms for your plot (i.e. only go up to 6 bedrooms).
10. Interesting! The median price for zero bedroom listings appears to be higher than that of one bedroom listings! Why is this? Explore the data and come up with a reason for this trend.
11. Do ratings appear to impact the price of AirBnBs? Investigate if there is a relationship between the price of an AirBnB and the ratings that the AirBnB received. Use the variable 'review_scores_rating' for the rating of the AirBnB. Make at least 2 plots and 2 tables to explain your results.