project prototype update

Data cleaning was done in python file. Firstly, all NaNs were dropped. Secondly, set alpha to 0.05 and dropped all outliers. Finally, generated samples of random data from a 50% distribution, with the base of 1. After the data cleaning was done, there were 17357 columns of data left with the maximum price about 300 dollars.

When customers are deciding which Airbnb to choose, two important factors are often considered: location and price. Thus, the first view is a geometric map showing Airbnb locations. Each neighborhood is bounded with white line and would show name of neighborhood when being clicked on.

To further help users to know what Airbnb properties, represented as circles, to select on the map, a color scale based on price was performed. Since most of the prices were under 150 dollars, for properties within that price domain, we chose colors more contrasting. We categorized data into 5 groups altogether. For Airbnb with prices smaller than 50, the circles were colored blue; for prices; for Airbnb with prices between 50 and 100, they were colored green; for Airbnb with prices between 100 and 150, they were colored yellow; for Airbnb with prices between 150 and 200, they were colored orange; for the remaining Airbnb with prices larger than 200, they were colored red. The categories above would be illustrated as a small double-ended color map with price numbers shown on it.

The second view would be a view showing more detailed statistic based on chosen data. The interaction between two views is done by brushing. One bug about brushing, which is to be fixed later, is the map would not been reset after one brushing is done. After certain group of data is selected through brushing, we plan to use a parallel coordinates or scatterplot matrices to show the correlation between different attributes of selected data, including price, neighborhood, availability, room types and minimum nights.