

Visualization of New York City Airbnb Listings

Chandan Nagaraj Naik, Tetsumichi Umada

{University of Colorado Boulder}, {University of Colorado Boulder}

ABSTRACT

Airbnb is one of the most popular websites to book a room to stay for a short time. The number of listings and users have been exponentially increasing in the past few years. Our project aims to build a website that visualizes the Airbnb listings in New York City so that the users can analyze the distribution of the prices and room types depending on the neighborhood area in New York. At the same time, we plot the listings on the map so that the analysis can see the geographical spread.

1 INTRODUCTION

Airbnb was started as Airbedandbreakfast.com on August 11, 2008. Since then Airbnb is growing so much so that in recent times it affects city's housing supply and affordability. Airbnb is popular, as of 2019 there are over 7 million [1] listings worldwide, over one hundred thousand cities with Airbnb listings and it is now one of the most popular options for travelers to stay. Those numbers are very interesting, and as part of this project, we are exploring and creating a visualization using all the airbnb listings in New York which happens to be one of the busiest cities in the world.

The users on the Airbnb websites can see the list of rooms that are currently available in the given area and days. However, the users would not be able to see the rooms that have been available in the past and their prices. For this project, we create a webpage that allows users to choose one of the five different neighborhoods in New York City area and see the popular room type, distributions of prices, and popularity based on the number of reviews the listing has. The goal of this project is to analyze the price and room types in New York City on the Airbnb

2 RELATED WORK

The typical visualization for geographical data would be a weather map, especially representing the snow or wind data. In particular, a governmental agency, National Oceanic and atmospheric administration, NOAA, provides weather and climate toolkit in which an analyst can check the archived data, including radar, satellite, and model data. The toolkits allow seeing the change of temperatures over time laying over the map with a heat map or change of the wind.[2]

With the geographical information, the scatter plot is one of the visualization methods to emphasize the spatial distributions of the two-dimension spaces. It is relatively simple and flexible to customize and map each attributes (Sarikaya and Gleicher)[3]. For example, the size of each marker can be used to represent the magnitude of the data point. By using different color coding, they can represent categorical data as well. As mentioned by Cui et al.,[4] the scatter plot cannot handle too much data. If there is too much data in a certain space, the plots can be overdrawn and they can be potentially plotted repeatedly. If several markers with

different colors or information are plotted, some of the visual information cannot be represented. To avoid overdraw, reducing the number of data points can be one of the options. It's possible to reduce the data by cleaning up the data points. Also, the data can be reduced by binding the data points. It means that if several data points are close to each other, it's possible to draw them with one marker.

Another method of visualizing data would be a bar graph and histogram. The bar graphs are often used to represent categorical data while the histogram is used for plotting a distribution of the continuous data. The bar graphs allow the analysis or readers to compare values between the categorical groups to see how different each other(Weissgerber et al.)[5]. On the other hand, a histogram can be used for continuous data to see the distribution of the data. Both the bar chart and histogram are able to see the distributions of the continuous data or differences of categorical data. That can be more informative than the summary statistics.

In addition to different types of visualization, the color is also a key point to create a visualization that can represent the data. In a visualization, colors help to clarify the data or enhance a data point. To keep the visualization organized, it needs to be grouped by categories and ordered by importance(Stone)[6]. Then, color needs to be mapped to the categorical items in a proportion to importance.

To build a visualization effectively, we use the Five Design-Sheet methodologies proposed by Roberts[7]. The five states involve 1) consider a task, 2) sketch small ideas, 3) plan three alternative designs, 4) consider a solution with client, and 5) implement a prototype. This methodology helps to develop a visualization effectively. At the same time, the info show-case at the end of the semester would be an opportunity to generalize the idea and have feedback from the users so that we can have another iteration to improve the visualization.

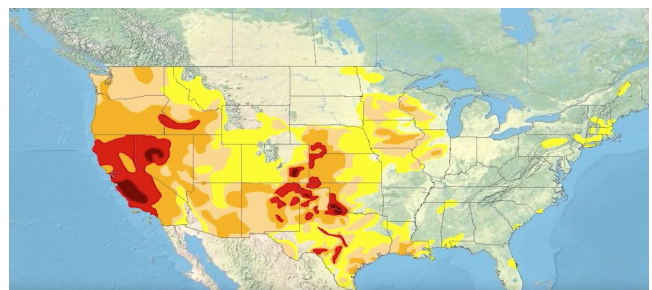


Figure 1: An example of a typical weather map visualization

3 DETAILED DESCRIPTION OF PROJECTS AND DESIGN CHOICES

For this project, we visualize the Airbnb listing in New York City area in 2019. The data set is available on Kaggle[8]. The data set contains the 49081 listings in five different neighborhood groups including Manhattan, Brooklyn, Queens, Bronx, and Staten Island. As a pre-processing, we remove the listings whose price is 0. We also removed the listings if the names of some listings did not contain alphabetical characters. Some listings did not contain their room types, so we also remove them. After cleaning up the data, there are 49056 listings in total.

To visualize the location of the listings, we decide to plot each data point using a WebGL framework called deck.gl [9] on the map of New York City built using Mapbox [10] [11]. Each mark on the map represents a room listed on the Airbnb website. The scatter plot on the map works a function of a heat map, meaning that if there are multiple listings are listed in the given latitude and longitude, then the color of the marker is close to red. If it is the only listing, then the color would be blue. This visualization allows the users to see how the Airbnb listings are geographically distributed in the New York City area and how dense they are.

On the left side panel, we plot the distribution of the prices with given max value. The user is able to choose the area of New York and specify the max value of the listings in the dataset and be able to see the distributions of the price from 0 to the given max value as a histogram. The x-axis represents a price and the y-axis denotes a count of the listings. The bar with the highest count is colored with slightly darker so that the users able to identify the most available listings in the given price range. The price selection is also correlated with the scatter plot on the map as well as the bar chart that represents the room types. For each room types that are available in the New York City area, we show the counts that are listed within the given price and the area. At the bottom of the page, we list the most popular rooms in the given neighborhood area and prices. The popularity is determined based on the number of reviews. The first one would be the one with the highest number of reviews.

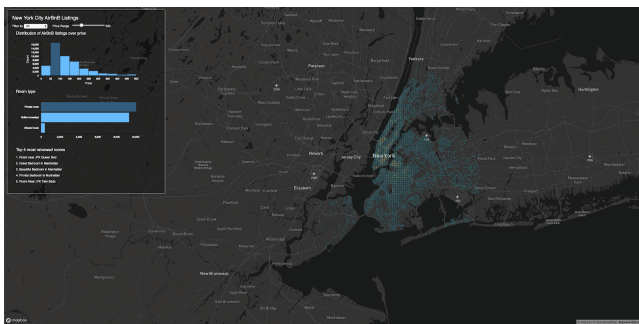


Figure 2: Overview of our interactive visualization

4 DISCUSSION

The visualization allows the users to see how the Airbnb listings are spread in the New York area. Many listings are available in New York City. Especially, in Manhattan, the area is mostly covered by the listings. If we take a look closely, the rooms located very close to each other in the area near Times Square and East Village. On the other hand, there are not so many rooms are listed in Staten Island and listings are spread out. If we also look

at the outside of the Manhattan area, the listing counts are less as the location becomes far from the city.

The room type in the New York Area, most of them are entire home or apartment. Comparing to the other areas in the United States, New York is a city place, and it's very likely for anyone to assume that people who try to choose Airbnb in New York would more likely choose a shared place over private space given how expensive New York is. However, we can see that New York has more private space listings compared to the shared space which makes us rethink.

The distribution of the prices across the New York looks very similar between the neighborhood areas. Most of the price is between 50 and 100. By looking at the histogram, it has a long tail for the higher prices. It indicates that some of the listings are very pricey. It would be interesting to see if we can take a look at each location to investigate how the room looks actually looks like.

5 CONCLUSION

With our map visualization, we tried to explain how a given data can be explored in our case it's Airbnb listing data for New York City. In our work, we've tried to explore different parameters and visualizations we can for our dataset.

Although we're showing the scatter plot and analysis for a given region, we could make the map interactive by showing more data on hovering which could be using. We can also have our visualization to show how the Airbnb listing changed over time. This helps us gain insight about how popular a given city is when it comes to travellers.

REFERENCES

- [1] Fast Facts <https://news.airbnb.com/fast-facts/>
- [2] Visualizing Weather and Climate Data <https://www.ncei.noaa.gov/news/visualizing-weather-climate-data>
- [3] Scatterplots: Tasks, Data, and Designs Alper Sarikaya, Student Member, IEEE and Michael Gleicher, Member, IEEE
- [4] Measuring Data Abstraction Quality in Multiresolution Visualizations. Qingguang Cui, Matthew Ward, Elke Rundensteiner, Jing Yang
- [5] Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. Tracey L. Weissgerber , Natasa M. Milic, Stacey J. Winham, Vesna D. Garovic
- [6] Color things, Maureen Stone
- [7] Jonathan C. Roberts, Chris Headleand, and Panagiotis D. Ritsos, "Sketching Designs Using the Five Design-Sheet Methodology," Visualization and Computer Graphics, IEEE Transactions on , vol.PP, no.99
- [8] New York City Airbnb Open Data <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>
- [9] Deck.Gl <https://deck.gl/#/>
- [10] Mapbox <https://docs.mapbox.com/mapbox-gl-js/example/setstyle/>
- [11] deck.gl and Mapbox GL JS: Better Together <https://medium.com/vis-gl/deckgl-and-mapbox-better-together-47b29d6d4fb1>