

VISUALIZATION OF AIRBNB LISTINGS

Basic Information

Andrew Liu (awl323)

Anna Skarpalezou (as11304)

Brandon Eap (be678)

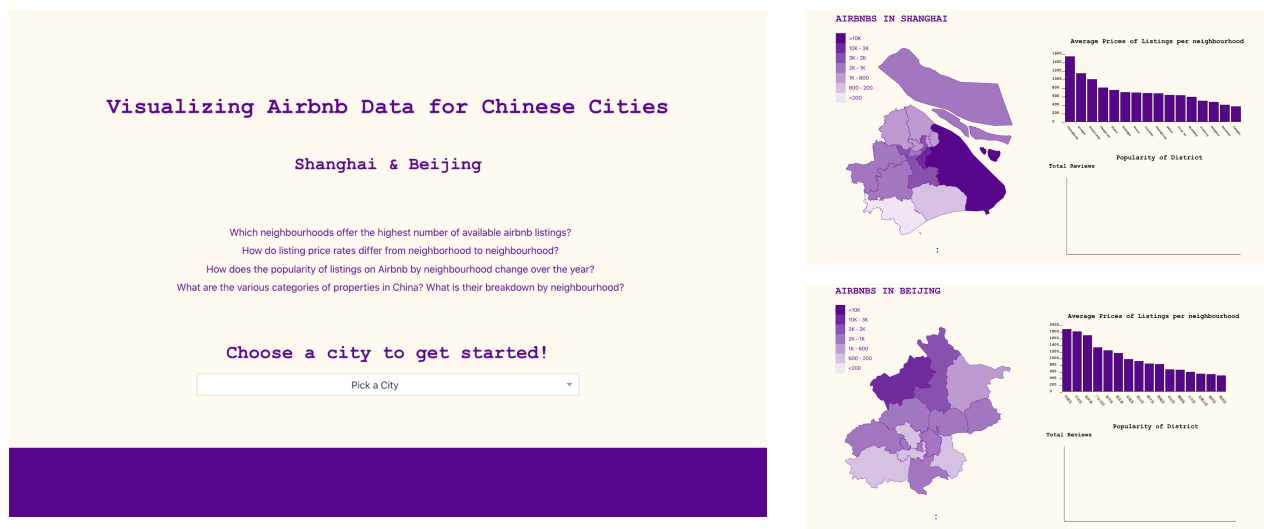
[Github Repository](#)

Overview

Airbnb is a digital platform through which users can organize and deliver hospitality and tourism experiences. This hospitality service can be accessed via their website or smartphone applications. The aim of this visualisation project was to analyze and estimate the most common rental neighbourhoods in different cities in China, as well as how much it would on average cost for someone considering a trip there. We are also exploring the popularity of each neighbourhood throughout the year, as well as offering some aggregate statistics by neighbourhood.

The visualizations we created are intended to assist users in exploring common neighbourhoods based on their preferences in Shanghai or Beijing. Our data and visualizations are mainly aimed at travelers looking for accommodations based on cost or popularity levels at a certain point in the year. People looking for a cheap place to stay for a single night or a place to host a small gathering/event are among our secondary community.

The cover page and visualization without selection look as follows:



Description of the data

The data can be accessed through the Airbnb Open Data Sources page. We utilized 3 types of files for each of the cities, making for a total of 6 files used:

- Listings.csv contains basic information about the listings including:
 - id (categorical , primary key of the table) , name (categorical), host_id (categorical) , host_name (categorical) , neighbourhood_group (categorical), neighbourhood (categorical) , latitude (quantitative), longitude (quantitative) , room_type (categorical), price (quantitative) , minimum_nights (categorical) ,number_of_reviews (quantitative), reviews_per_month (quantitative) ,calculated_host_listings_count (quantitative) ,availability_365 (quantitative).

- Neighborhoods.geojson is a GeoJSON file of neighborhoods of the city, a map of Shanghai separated into districts in both English and Chinese.
- Reviews.csv contains the date when each listing was reviewed and the listing_id

We utilised the *price* and *neighbourhood* information, grouping the listings together by neighbourhood to visualize the average price distribution across on a bar chart. To find the popularity of a district, we merged reviews.csv and listings.csv to get all Airbnb listings with their reviews and review dates and grouped the districts. The result was a file that conveys the popularity of each district for every month. Since we wanted to get aggregate statistics by neighborhood to show on our tooltip, we decided to do some data preprocessing and munging in python. We mainly used the library pandas, using the function `group_by` to get by neighbourhood info. We then used the `.mean()`, `.max()`, `.min()` and `.count()` functions to get the average, maximum and minimum price of listings per neighbourhood, the number of each type of listings (Entire home/apartments, Private rooms, Shared rooms) per neighbourhood and the average number of reviews per neighbourhood per month. The relevant code can be found on `data_munging.ipynb`.

Goals and Tasks

The main goal for our visualizations is to visualize the number of Airbnb listings in their respective neighbourhoods of cities. By using three different views to visualize the data, we hope to aid users in answering the following questions:

- Which neighbourhoods offer the highest number of available airbnb listings?
- How do listing price rates differ from neighborhood to neighbourhood?
- How does the popularity of neighbourhoods for Airbnbs change in a year?
- What are the various categories of properties in China? What is their breakdown by neighbourhood?

The following are the tasks that each of these visualizations sought to tackle:

- Choropleth Map: visualize the number and distribution of airbnb listings in the neighbourhoods of relevant cities (Beijing and Shanghai)
- Bar Chart: provide an alternative visualization for the number of listings filtered by neighbourhood, however making it easier to easily compare values. Differentiated by height channel rather than color saturation channel
- Time Series: visualize the popularity of the selected neighbourhood through the course of a year.

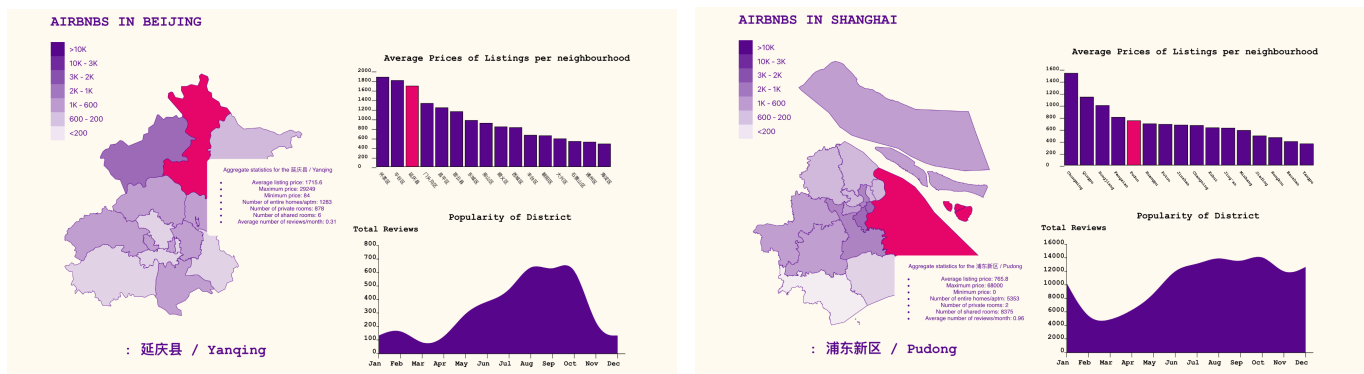
Visualization

Our app is split into two main components, the cover page and the visualizations of the data. The cover page introduces the main objectives we hope to convey through our visualizations. Initially the cover page pops up, to navigate to the visualizations there is a dropdown UI widget that allows the user to choose among the cities for which visualisations are available, filtering the dataset used. For each city there are 3 main views, a choropleth map, a barchart and a time series.

The map's interaction is simple and intuitive, using the channel of saturation to differentiate the sections of airbnb listings. Users can hover over each of the respective city's neighbourhoods to see a toolbar with the name of the neighbourhood as well as a tooltip with aggregate statistics for that neighbourhood, which follows the cursor. The selected neighbourhood is distinguished by its changed color.

Interactions with the bar graph allow the user to select a neighbourhood, which changes color upon selection. Both the choropleth map and bar graph views are coordinated with linked highlighting bidirectionally, with a hover over a neighbourhood in the barchart triggering the same neighbourhood on the map and vice versa. The time series is filtered based on the selected district and displays the data related to the district. A tooltip also appears, which does not follow the cursor as not to cover the information dense bar chart.

Lastly, an area map appears upon the selection of a neighbourhood either through the map or the barchart. This is unidirectionally linked to the other views. The three views with selection for each city are as follows:



Reflection

A major difference from our initial plan was the inclusion of a second city in our visualisation. We decided to use an interaction that dynamically responds to the users to enrich our visualization and show different aspects of the dataset adaptively. One downside to our dropdown layout is with respect to comparatively looking at the two cities, as it may impose cognitive load when viewers are required to remember previous views. However, we believe that the second city makes for an overall more informative visualization experience.

We were initially planning on showcasing each individual listing on the choropleth map with a red dot, making for a dot density map. After reviewing the datasets however, we realized there were hundreds of thousands of listings in each neighborhood and such a representation would clutter the map and reduce the visualization's effectiveness. We decided to instead use the channel of saturation as it works well for perceptual ordering. We also wanted to represent ranges of values and not discrete categories so we believe the higher number of bins used does not impede the effectiveness of the visualization. The Choropleth map is a straightforward and easy way to understand visualisation for spatial data, however it takes up position channels for geographic information only, leaving the size and color channels to convey the rest of the information. Even though color and area interact so that human perceptual consistency is not perfect, due to the size of our visualizations we believe it is still effective.

We decided to use a sorted by price barchart, to allow for a data driven ordering that makes dataset trends more obvious and allows for easier lookup by price for the user. The line marks are aligned within a common frame so that the highest-accuracy aligned position channel is used. In retrospect, we could have used stacked bar charts to show the price variation by neighbourhood type. Given that the types are only three, such a representation would not make for a too cluttered view and would allow part-to-whole comparisons.

We used a measurement temporal dataset for the visualization of popularity. The dataset was sequentially structured, but through our preprocessing we grouped the data by month turning it to a cyclical structure for our visualization. Initially, we were planning on using a multiple line chart to encode more of the statistics information included in the tooltip, however the majority of information did not show interesting variations throughout the year. Therefore, we used an area chart to show popularity only, as it makes it easy to perceive the shape of the curve. The interaction between the 3 views on each city, imposes a lower cognitive load but requires more display area.

NOTE: the source code can be found on our github repo hyperlinked on the cover page of this paper or at <https://github.com/AnnaSkarpalezou/InfoViz-Final-Project>