UC DAVIS
GRADUATE SCHOOL
OF MANAGEMENT

Zillow Housing Information Data Visualization

BAX 431 FQ 2018

Data Visualization Final Assignment

Team Member: Xinyan (Sidney) Lyu

Zhenyu Fan

Jiahui (Hazel) Liang

Executive Summary

A home is often the largest and most expensive purchase a person makes in his or her lifetime. Ensuring homeowners have a trusted way to monitor this asset is incredibly important. *[1]* Our visualizations are created for average-income married-couple families to choose a suitable house in Orange County, CA.

In this report, we use dataset from Kaggle competition: Zillow Prize. It includes full information for all real estate listings in three Los Angeles-area counties. We only focus on data in Orange county(Regionid 1286) in 2017. Before visualization design, we perform data cleaning process to make our data presentable. We use Kepler as our visualization tool since we have a larger geographic dataset. This report also includes our imagination user case and our explanation for our visualization design.

Data Cleaning

From the original dataset, we choose parcelid, latitude, longitude, bathroomcnt, bedroomcn, yearbuilt, taxvaluedollarcnt, calculatedfinishedsquarefeet, garagecarcnt, fireplacecnt, hashottuborspa, and regionidzip as our future filters for housing choice. The variable names from the original dataset are ambiguous for our future visualization design. Therefore, we transform them into ID, Latitude, Longitude, Bathroom, Bedroom, Yearbuilt, Price, Squarefeet, GarageCar, Fireplace, Spa, and RegionidZip. The meaning of Bathroom and Bedroom stands for the number of rooms in house. GarageCar, Fireplace and Spa means whether the house has those facilities. For the null values in Price (house price) and Squarefeet (house area), we deleted respective rows because they indicate the lack of date and will contribute less to our future visualization. However, for the null values in GarageCar, Fireplace and Spa, we fill null values with zero, which means that the house doesn't have garage, fireplace or spa. What's more, we add one column called 'Price/Sqft' to indicate the house price of every square feet.

Filters and Metrics

To build an useful tool, we consider the purchasing power and the family structure of married-couple families. A house should first satisfy the requirement of accomodation. So the numbers of bedrooms and bathrooms stand out as key metrics, which may weigh more importance than total square feet . Moreover, price always accounts for most important part in purchasing decision, as well as price per square feet.

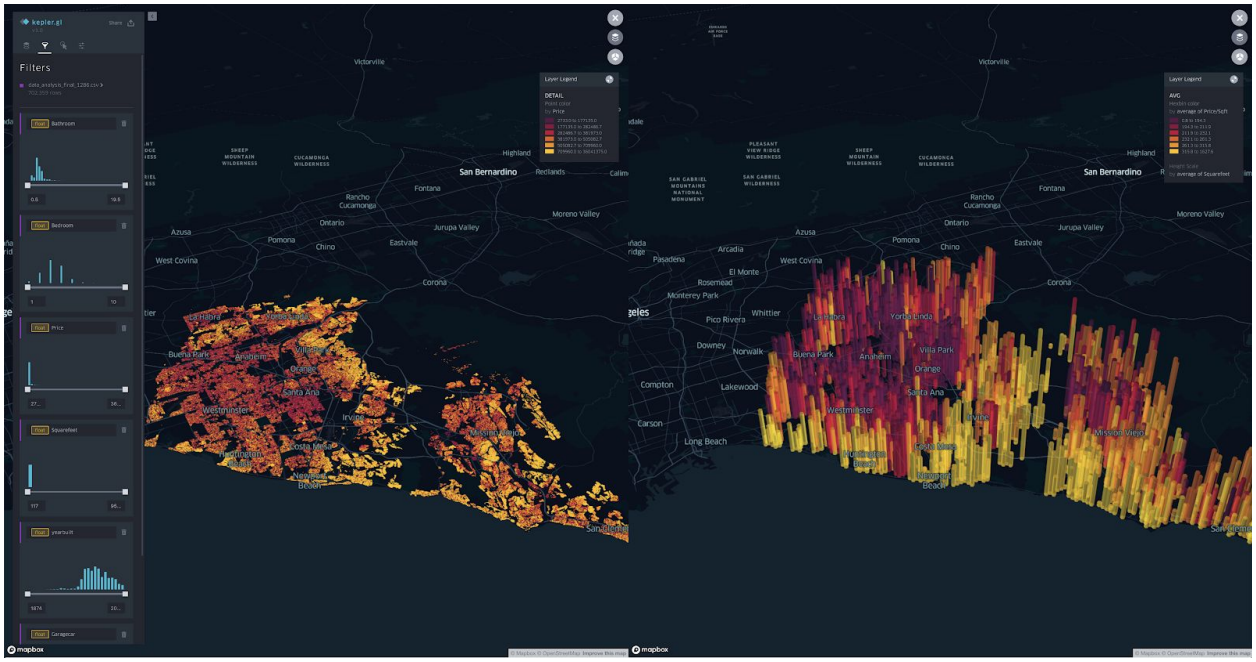
When it comes to house overall condition and facility, "yearbuilt" could represents the overall maintenance and house condition. Apart from bedrooms and bathrooms, garage is another concern that most families have since Americans own an average of 2.28 vehicles per household nationwide. Especially in California, cars are the most convenient transportation. Therefore, better parking capacity definitely gains competence for a house. As for fireplace, considering that Orange County located at West Coast and the weather is moderate, it may be a minor significant feature.

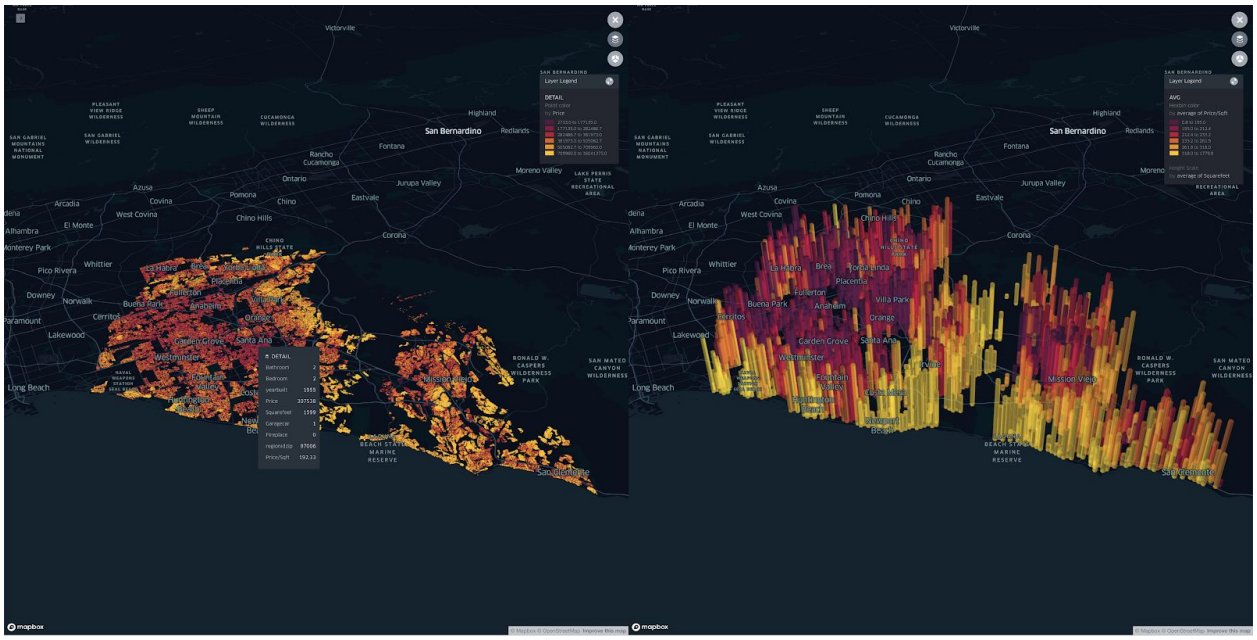
In summary, we choose "bedroom", "bathroom", "yearbuilt", "price" and "garagecar" as five main filters in our Kepler map, we also add other filters in case there are special customer needs. Other factors, such as, total square feet, price per square feet, zip code(which may related to security and school district) and number of fireplace are shown in the detail.

Visualization Design

Our visualization design is presented by two parts. In the right map, user can locate detailed information about specific houses. In the left map, user can compare the neighborhood information. We use color to represent house price, the darker the color, the higher the price.

In the right map, we choose to show the total price for the house, as well as the information about the house including bedroom number, bathroom number, total square foot, year built, garage situation, etc. So users can have access for information of the houses they are interested in. In the left map, we choose to show the average of price/sqft and the average of square foot in that neighborhood(0.5km). The height shows us the average square foot. To exclude the extreme value, and also to serve our purpose better, which is visualization for average-income married-couple families, we choose to show the data of 25 percentile to 75 percentile. After finishing layer building, we add various filters to create a interaction visualization.





User Case

For instance, a typical middle-class couple family, John and Amy; they have two children, and expect to buy a well functional house at least with 3 or 4 bedrooms and one garage.

According to US Census Bureau, 80% Married-couple families in Orange County's incomes range from $50,000- $200,000 or more, and median income is estimated as $107,728. Meanwhile. More than 85% of

owner-occupied housing values range from $300,000 to $1,000,000 or more, with median of $620,500.[2] So, we assume that they have a budget of lower than $650,000.

First, we filter the bedroom number from 3 to more, and set total price lower than $650,000. Consider there are two children in the family, we then filter out houses which have more than two bathroom and two garages for cars.

Then for the age of the house, considering their limited budget, they want to look for a house that is well-functional, which means they don't need to make any major maintenance to the house they buy. So they would want a house that's relatively new. We assume that they may want a house that's built after the year 2010.

Then, they will consider the neighborhood. Generally speaking, Orange County has safety environment and neighbourhood as long as you avoid living in so-called high-crime area: Santa Ana and Anaheim. Our map also shows the average house price of the neighbourhood which can help this couple to identify the better living area.

Reference:

[1] Zillow Prize: Zillow's Home Value Prediction : https://www.kaggle.com/c/zillow-prize-1
[2] US Census Bureau:
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF
[3] 2019 Best School Districts in Orange County
https://www.niche.com/k12/search/best-school-districts/c/orange-county-ca/
[4]Top 10 Safest Cities In Orange County by Santa Ana Bail Bonds
https://ocbailnow.com/top-10-safest-cities-in-orange-county/