

# Assignment 3 Description: Interactive Visualization Software

Yichao Wang, Leon Long, Haoran Pu

## Domain

As Seattle becomes one of the top fast-growing cities in the United States, we are interested in looking into the Greater Seattle's housing landscape and how various kinds of factors play different roles. For this assignment, we chose the housing sales data in as our domain and found a great dataset on Kaggle: [House Sales in King County between May 2014 and May 2015](https://www.kaggle.com/datasets/kingcounty/house-sales).

Since there are various attributes for a house including its location, size, price, number of bedrooms, bathrooms, floors, the year in which it was built, and etc, we thought that it would be very interesting and helpful to build a one-stop explorer that visualizes the distribution and relationships among different factors. It can be a meaningful tool for both house-buyers to explore assorted options and housing industry professionals and researchers to gain insights into underlying patterns and trends

## Techniques and Description of Storyboard



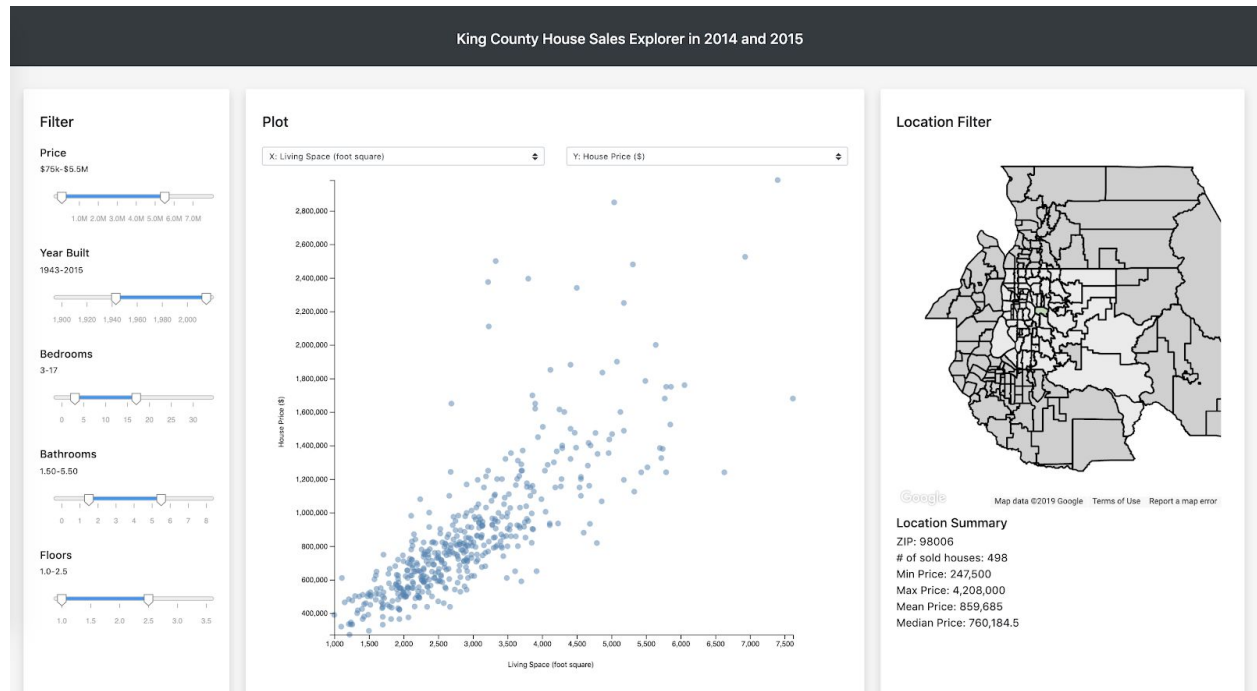
Considering that geolocation data is included in the dataset, we intuitively figured that map can be a great form of visual encoding that shows the location of the houses. Moreover, since house sales data is very location-sensitive, we thought that by utilizing the **direct manipulation** interaction technique, the users can directly click on specific region on the map and the data will be filtered to represent that region exclusively.

Besides, after inspecting the dataset, we found that some of the variables have a wide range. For example, the number of bedrooms ranges from 1 to 33 and the price from 75K to 7.7M, which means that the visualization potentially could look very different with different variable selections. Thus, we decided to provide a set of **filters (dynamic queries)** for the users to obtain the results based on their interests and needs.

With filters and map combined, we can generate the plot of either one or two variable of the user's choice. For one variable, we will plot a bar chart of the distribution of the variable. For two variables, we will generate a scatter plot showing the relationship between the two. The users can interact with it by selecting different ones. We also added **point selection tooltip** technique to display more information when user hovers over the point and bar on the visualization plot. In addition, we originally added a color or size encoding option but later removed it since it might cause confusion to non-technical users.

The filters, map, and visualization plot together employ the **brushing and linking** interaction technique, which is a great fit for our domain considering the number of variables presented in the dataset, the wide range of certain variables that could affect the result, and the users' need for the freedom to input choices of their interests and explore the trends dynamically.

## Final Interactive Visualization Application



Our final interactive visualization application is pretty similar to our initial storyboard in general. The left panel contains all the filters that the users can interact with, the right panel is a map with sub-regions separated by zip code. Hovering over a specific region, the user can see a summary of all houses in that region. Clicking on it will only include the data points in that area. The middle panel is the visualization plot: two dropdown menus to select the variables and a dynamically scaled bar/scatter plot based on the selected variables.

## Changes Between the Storyboard and the Final Implementation

When it comes to the difference between the initial storyboard and final application, we made a few adjustments as we work through the implementation. First, for the filters, as we inspected the attributes, we realized that dropdown menus might limit the users' ability to look at a particular range. Since many of the attributes are continuous, making the filters as double headed range sliders makes more sense as it gives the user not only more freedom to define custom ranges but also the ability to brush through and see how the visualization dynamically changes. On top of that, we removed the color and size encoding feature to eliminate confusions and distractions to users, since the current design is already very adequate and informative.

## Commentary on Development

Leon created the dynamically scaled filter sliders in the left panel and wrote the descriptions of the storyboard and final interactive visualization application. (Leon took about 6 hours)

Yichao created the location filter and stylized the final product. The most time consuming step is to load geojson file onto the map and link the house sale data with geojson data. (took about 8 hours).

Haoran created the data visualization (histogram and scatter plot), the choices of X-Y attributes, and the select technique for the histogram (display detail information when hover on the visual encoding). In this implementation, two parts are the most time-consuming: the first one is to alter scales of axis and visualizations based on the user's choice of x-y attributes, and the other one is to integrate the filter techniques from map/slider to actual data. (took about 10 hours)

Final product: <https://yichaow.github.io/info474a3/>

Source code: <https://github.com/YichaoW/info474a3>