

# A3 Report

Link to visualization: <https://info4310-a3.onrender.com/>

## Data Description

We chose the Zillow dataset of houses for sale or rent in Pittsburgh. The complete dataset contains more than 30 fields, of which we chose to focus on the neighborhood, property type, and sale amount as the variables of interest. Additionally, we included street address, number of bedrooms, and year built when interactively displaying information to the user. For the map, a GeoJSON file containing Pittsburgh neighborhoods was used, which was taken from <https://github.com/blackmad/neighborhoods>.

Some processing was done with the given dataset:

1. We used a python script to calculate the average prices for each neighborhood.
2. Neighborhood name discrepancies between the Zillow data and the GeoJSON file were corrected, e.g., mislabeled data points were relabeled and one house that was not actually located in the city was removed.

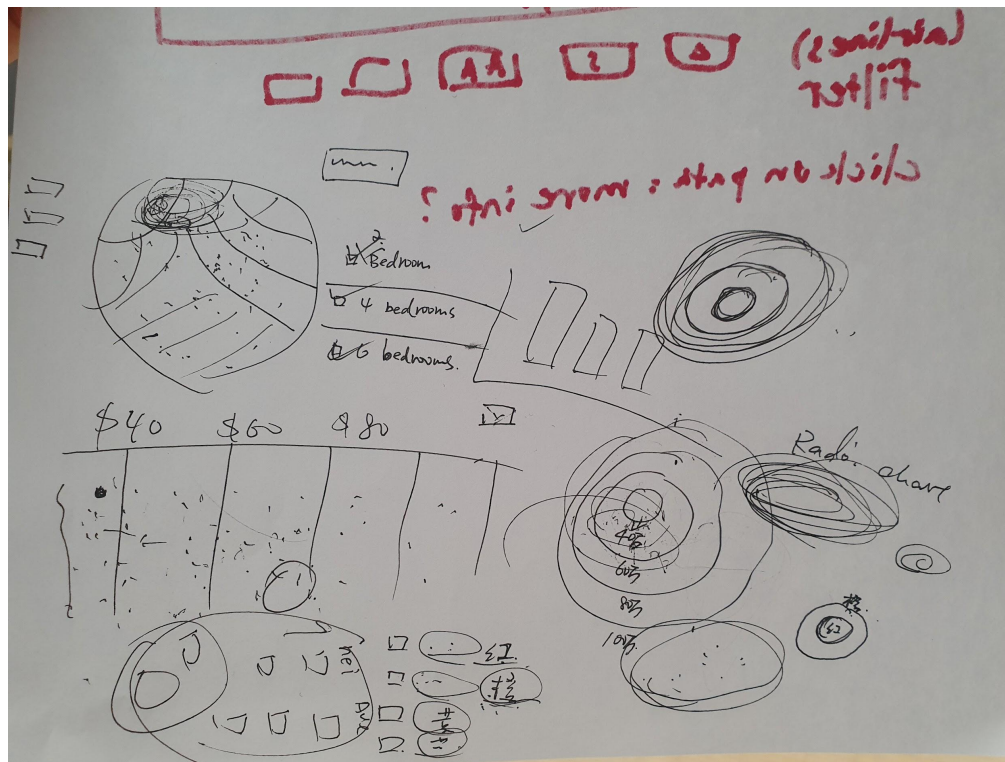
User needs we identified include:

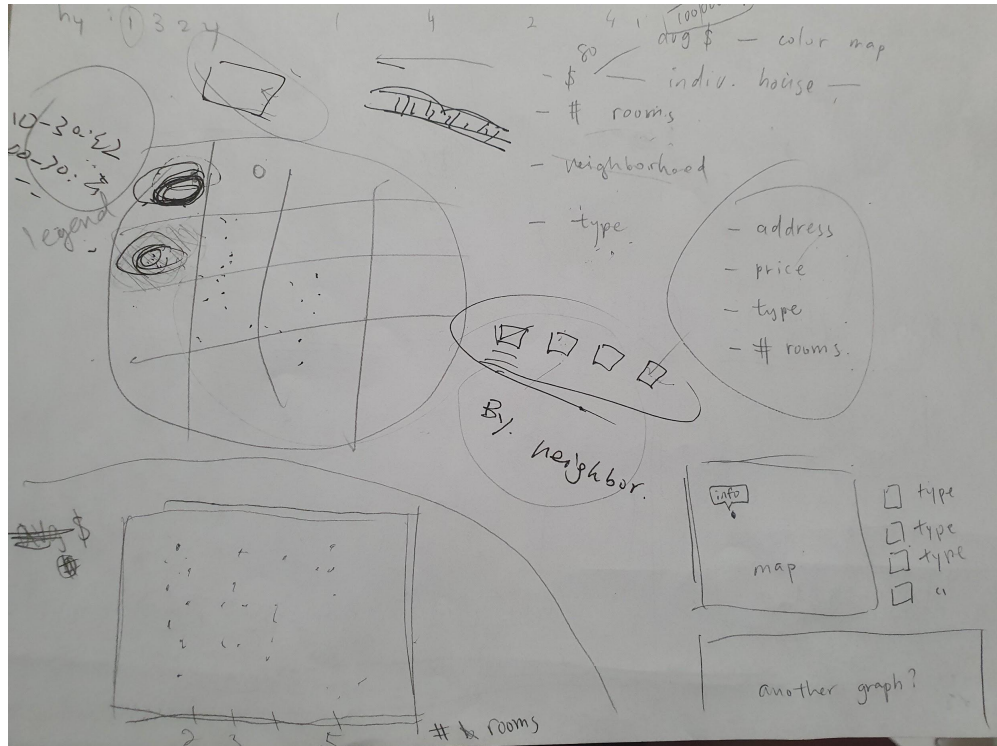
- An intuitive overview of all available properties.
- The ability to see further details on demand.
- Filtering by certain criteria to narrow down their choices.

## Storyboard

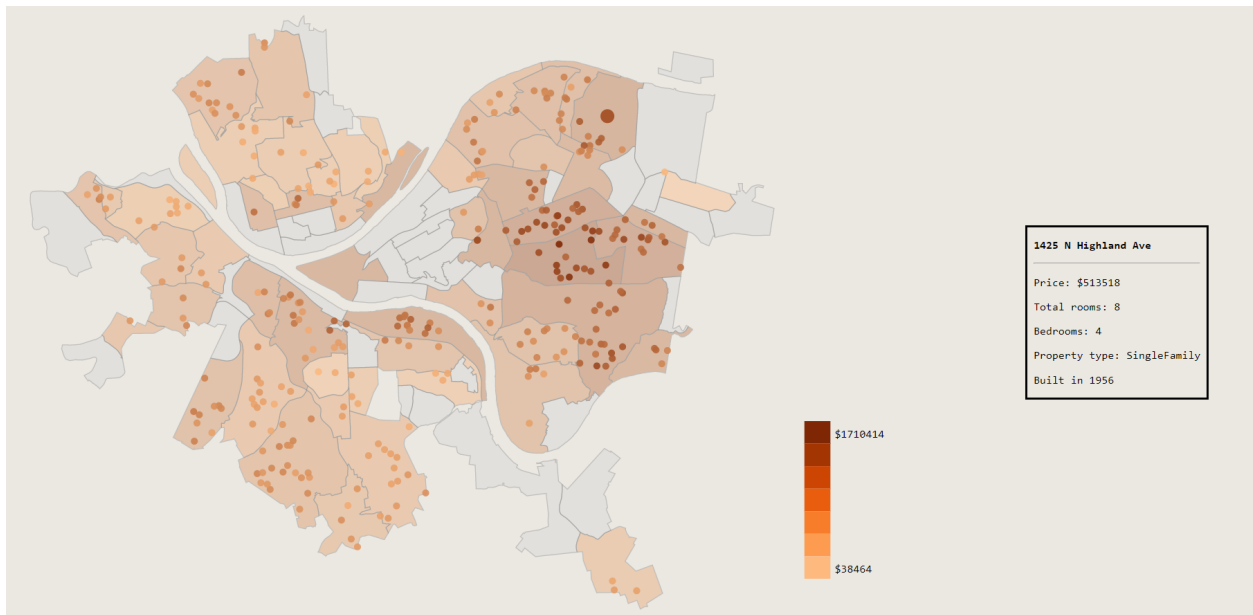
1. User opens the interactive map of Pittsburgh house prices
2. User sees a circle on the map representing a house with a certain price
3. User hovers over the circle with their mouse
4. The circle expands and a text box appears on the right side of the screen
5. Text box displays the price, number of rooms, address, property type, and year built of the house
6. The user can choose a set of checkboxes representing different property types,
7. The user selects one or more checkboxes, indicating which property types they want to filter by.
8. The web page displays interactive graphs representing the selected property types. **You can use Checkboxes filter functions on both scatterplots and maps.**
9. User reads the information and decides if they are interested in the house or not

Justification: This interaction is designed to provide users with basic information about a house when they hover over a circle on the map. This information is presented in a concise and visually appealing way, which makes it easy for users to quickly assess whether they are interested in the house or not. The option to click on the text box for more information also provides users with a way to learn more about a house without leaving the map.





## Visualization



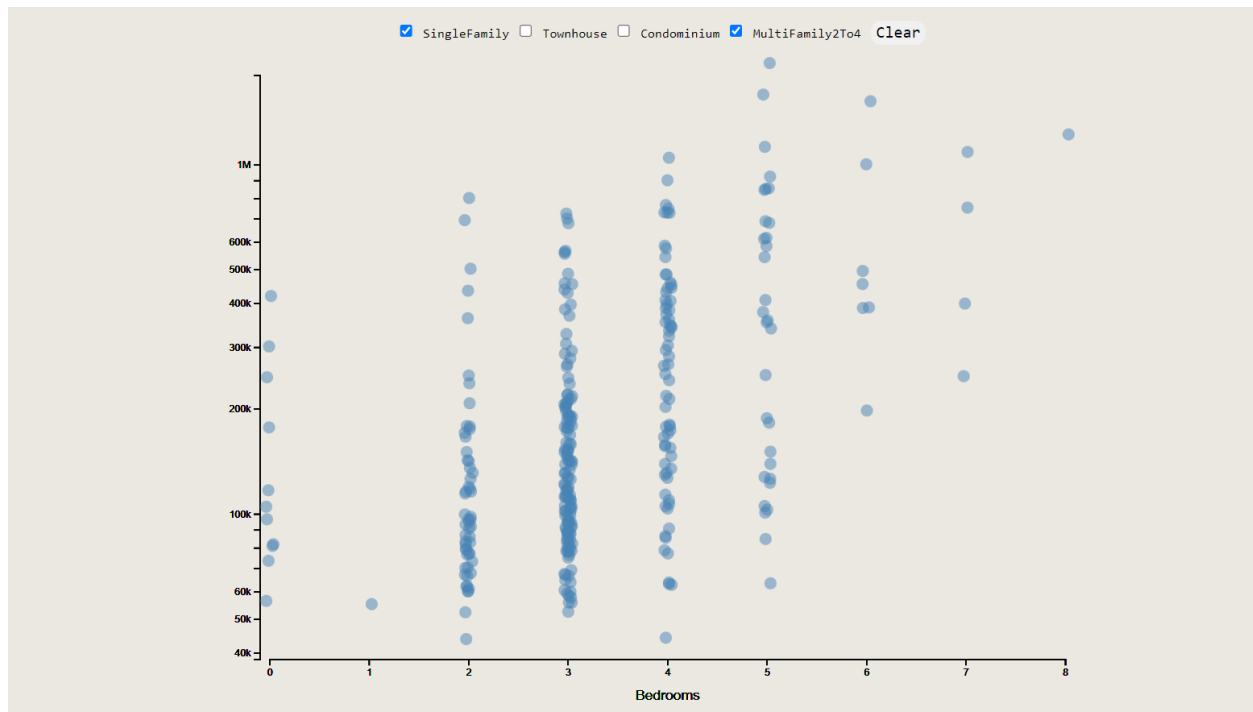
Our first interactive visualization application is a map of Pittsburgh house prices. The map displays circles on the locations of different properties in the city, with the color of each circle representing the price of the property. When a user hovers their mouse over a circle, it will be

enlarged, and related information about the property will be displayed on the right-hand side of the screen.

The information displayed will include the price of the property, the number of rooms, the address, the property type, and the year the property was built.

Using the checkboxes, users can filter the data and display graphs for specific property types, which are displayed and interactive both as maps and scatterplots.

Users can interact with these two graphs to learn more about the property and make informed decisions about their housing needs in Pittsburgh.



This interactive visualization is a scatterplot with checkboxes located above the graph. The checkboxes allow users to select different property types, such as single family houses, townhouses and condos. This functionality is shared with the map visualization above.

The x-axis of the scatterplot displays the number of bedrooms, while the y-axis shows the corresponding house price. By clicking on different checkboxes, users can filter the data and display a graph specific to the selected property type.

For example, if a user clicks on the "Single Family" checkbox, the scatterplot will only show data related to the single family option. If they then click on the "condos" checkbox, the scatterplot will update to show data related to condos.

Through this interactive visualization, users can easily compare house prices across different property types in a specific area, which will assist them in making informed decisions concerning the purchase of a residence and the number of bedrooms they require.

## Considerations

1. High density of information: in both of our graphs, there are dense clusters of data points. To avoid overwhelming users, we implemented checkbox options to create filtering functionality, which allows us to only present information that is relevant to users.
2. Map coloring: we colored both the individual dots (each representing a property) and the neighborhoods according to the sale amount. There are some disadvantages to this, such as not enough contrast for the lighter-colored circles. At the same time, it's clear at a glance which neighborhoods have no properties available and which have the highest prices. It also reflects the principle of redundant information being useful to a certain extent, especially as the neighborhood colorings don't change with the property type filter, providing a sort of overview even when the user "zooms in".

## Development Process

The visualizations and interactions remained relatively consistent between the storyboard and the final implementation. As part of the development process, some design choices were made, such as deciding which variables and filters to select and how to weigh them. Since a house's property type is an important factor affecting its price, we decided to use it as a filter. This also serves the purpose of avoiding too much information being presented all at once.

One thing that changed during implementation was the field that the circles' colors represented. Initially, we planned to color the data points by property type; however, we ultimately colored them by price because it is a more intuitive representation that addresses one of the biggest concerns people have when looking to buy a home. Additionally, the property type filters can be used to differentiate between types, so we were able to devote the colors for another field of interest.

## Work Breakdown

- Yvette: brainstorming & initial setup (2 hrs), data cleaning (~1 hr), map & interactivity (5 hrs), report (~2 hrs)
- Jiaxin: brainstorming & initial setup (2 hrs), data cleaning (~1hr), design choice & coding scatterplot chart (5 hrs), report (~2 hrs)