Melbourne House Price Distribution

FIT5147 - Visualisation Project Report
Student ID: 30118891
Student Name: Kun Liang
Lab number: 14
Tutor Name: Mohammad Haqqani

Table of Contents

1. Introduction

As the economy develops and the number of immigrants increases year by year, the demand for real estate in Melbourne has inevitably increased a lot in the past decade. With the improvement of housing quality and the continuous expansion of housing resources in the real estate industry, both house sellers and buyers are facing fierce competition. In today's big data-oriented society, understanding the market insights through some professional analytical methods can effectively save their time and solve their problems.

The report will introduce how to use data visualization to effectively let the target group know the following conclusions from the analysis in a short time:

- How locations affect the sales and prices
- The popular suburbs and house type in each suburb
- The hot-selling season for the real-estate industry in Melbourne

2. Designing process

Basically, the issues to be discussed in the project can be grouped into three factors: location, time and type of house. Therefore, the entire design process is roughly divided into 3 stages.

2.1 Pre-design work

For the initial stage of design, brainstorming is the best method to choose the chart types, and then filter and combine. Firstly, all of the candidate graphs are listed out and categorised based on their function in visualisation. The process after categorisation is filtering. In this stage, charts got filtered as sort of them might not work well in the later work. For example, to discuss the location factors, map is the first choice but choropleth map which achieves the functions of heat map and geographic map undoubtfully be the most powerful and straightforward method. Moreover, the word cloud is kept but the bubble chart is removed because the former has a better-emphasized outline. Since different chart combinations will show different degrees of ability to transfer information, appropriate chart fusion or combination can make the result display interface more concise and clean. Hence, the outline of initial design is the idea of using choropleth map to illustrate the location factor and using line chart to answer the question about the hot-selling period.

2.2 Initial design

In the first version of the design, in order to allow customers to fully understand the trend of house prices and areas with high prices or high transaction volume, the display page will provide customers with a choice of house types to let them understand the sales and average price of different house types in various regions. However, since there are over 50

suburbs in Melbourne, then the multi-line chart would be very messy for drawing such number of lines in one chart which the highlight function cannot fix it very well.

So, in the second version of design, the line chart is simplified by setting the drop-down list to let users choose the suburb they prefer to look for. Also, the radar charts are included as they can significantly shows the users about the distribution of house structure, such as in Abbotsford, the houses with three rooms, 1 bathroom and 1 garage have the highest sales.

Although the settings of the pull-down menu and the combination of charts mentioned above can reflect how location factors and house type factors affect the house price, the effect of time is very implicit. In order not to make the interface too crowded and chaotic due to the use of too many charts, paginating the entire display interface not only makes it simple and clear, but also makes the data visualization more organized (Hurter C.,2016). Independently present to customers the impact performance of each major impression factor obtained through the previous EDA analysis.

2.3 Final design

Compared to d3, which can display more gorgeous effects, shiny not only has a similar visual interaction effect, but also has a better paging function. After clarifying the visualization system, the next issue to consider is the optimization of the chart.

In the final version of the design, only choropleth map is shown in location factor tab as the combination of it and buttons with filtering and switching functions is sufficient for showing users the geographical information about Melbourne's housing prices by showing the distribution of sales and housing prices on the map.

In the tab of time factor, the line graph was changed from the initial multi-line simultaneous display to a separate display after segmented coloring, so that it can better display the respective house price trends each year. In addition, a word cloud has been added to this panel because it can directly pass information to customers about the 3 months with the highest sales.

The last page is about house type information, so three radar charts and a histogram are set up on this page. The histogram also shows the price distribution of each type of house in the selected suburb, which allows users to make their own comparisons easier and more intuitive.

Please see Appendix A for detailed design board architecture of each sheet.

2.4 Implementation

As mentioned above, the visualization interface is implemented by Shiny in R. Drawing tools includes plotly, leaflet and ggplot2. In addition to graphic display, the analysis will be very understandable if user interaction technique is applied (Goyal N. et al., 2011). Except for the word cloud, all charts have added their own customer interaction functions, including specific information highlighting display, personalized check and applied filter, etc.

According to Sahay A. (2017), the function of each software or package used in the project are explained below:

- plotly: the chart construction platform for easier graphical interaction is compatible with the output of ggplot2.
- Leaflet: with a variety of map types to choose, better data interaction (zoom function)

3. User guide

3.1 Map tab

By opening the display interface, the user will first see the tab of location factor. On this page, a choropleth map with the switch function is displayed directly to the user. Normally, the questions that users will want to know the first is like which area has the highest or low house prices, and where is the popular area for buying a house (Kauko, T.,2004). As shown in Figure 1, the map can show customers the distribution of house prices and sales in Melbourne through mode switching, and the values can be intuitively interpreted to customers and answering these two questions in shades of one light color to another dark one. As the default setting, the first map shown to the user is about the sales distribution of the houses in Melbourne.

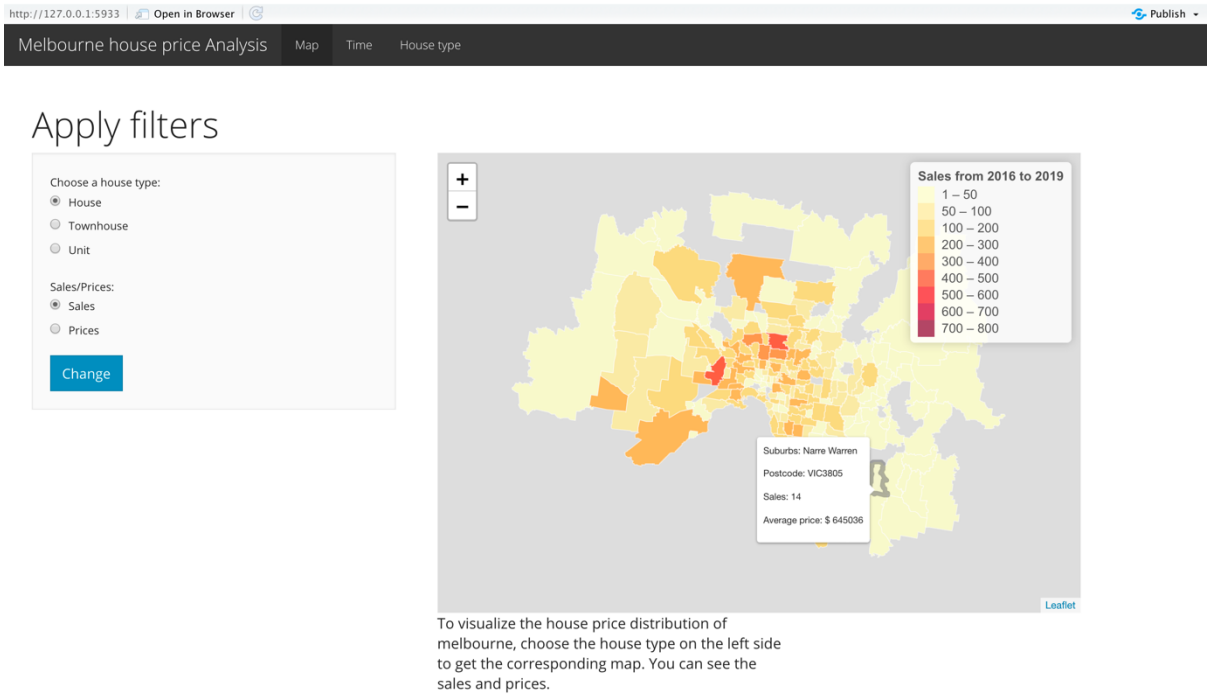


Figure 1 Choropleth map interaction

The user can learn the sales or price distribution of different house types by selecting the button correspond to their preferred type on the left side first, then clicking the change

button right down below to submit the change. Also, by putting the mouse on different modules in the map, the user can directly get the corresponding detailed information: the postcode, suburbs, the average house price and sales from 2016 to 2019 of this area.

3.2 Time tab

To jump to the other page, as Figure 2 shown, the user can click the tab on the upper bar anytime.

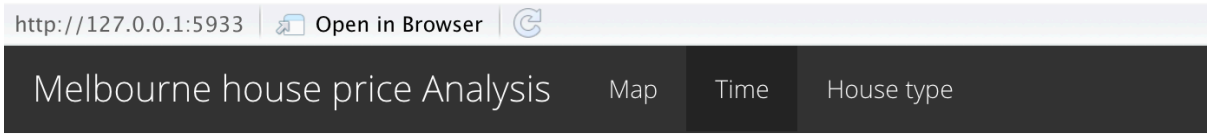


Figure 2 Switching pages

The user will come to the time panel if they click on the 'Time' tab of the upper bar. In Figure 3, the outline of this tab, the 3 months that with the highest sales in the Northern Metropolitan of Melbourne is significantly emphasized with the darker color in the word cloud. On its right side is the line chart segmented into different colors by years that automatically shows the user about the overall trend of house price throughout the period from the first quarter of 2016 to the first quarter of 2019 (latest data).

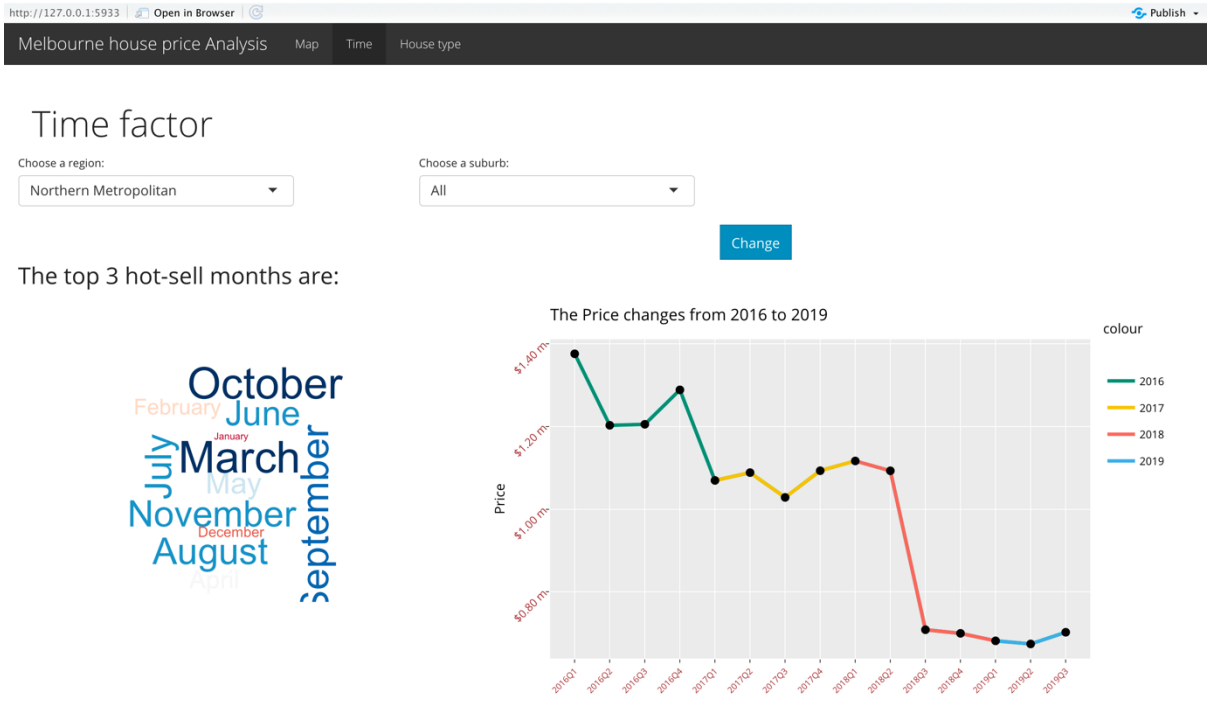


Figure 3 Outline of time factor interface

Likewise, the both of these two charts can be changed region by region (the state council in 2013) and suburb by suburb respectively by choosing the prefer one in the drop-down list and pressing the 'Change' button. Besides, the user can get the details including the suburb name, the average price in each quarter of each year in the period by hovering over on the points in line chart.

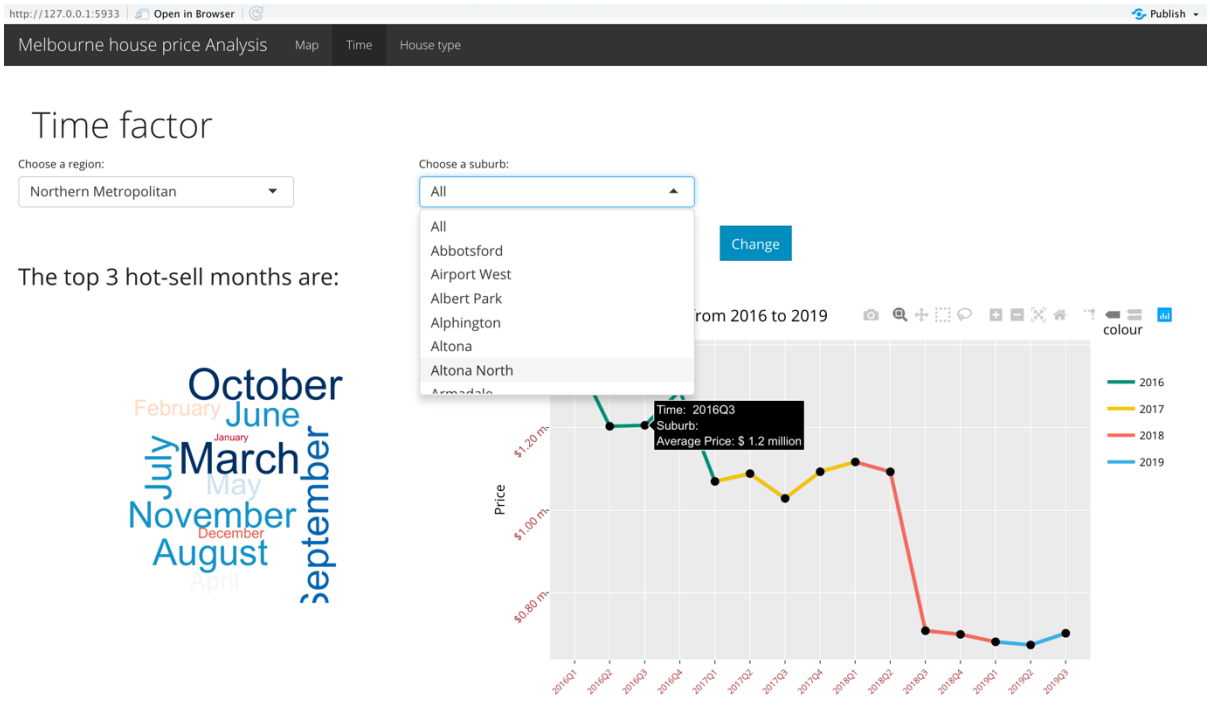


Figure 4 The interaction in time tab

3.3 House type tab

The third page shows the user about the house structure and the corresponding price distribution in each suburb. There are 3 radar charts and each is respect to the number of rooms, bathrooms and garages distribution, and when putting the mouse upon the categorical points, the message box will pop up and provide two parameters to the user, r and theta, where r stands for the number of house that equipped the corresponding rooms/bathroom/garage and theta is the categorical number of these factors. By checking these three radar charts, the user can have a better comprehensive understanding about the popular house structure in the user-selected suburb.

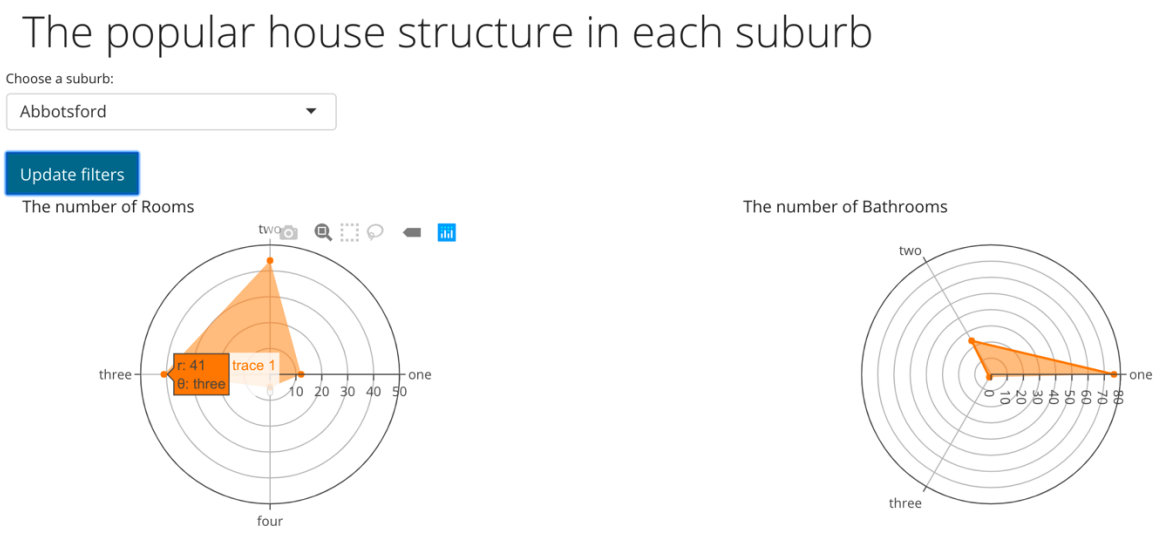


Figure 5 The radar charts illustrate house structure

Besides the radar charts, the house price distribution of each type of house is introduced to the user by the overlaid histogram in the lower right corner of the page. Like the previous two panels, the default histogram is drawn by the overall data of Melbourne surrounding area. As shown in Figure 6, in this graph, the price distribution of each house type is distinguished and overlapped by different colors, and the average price of the corresponding color is displayed as a dotted line. In other words, the user can learn the distribution of all type of houses are right-skewed and approximately concentrated in the range that between 1.3 to 1.7 million dollars, and in Melbourne, most type of selling properties are houses and unit is relatively rare. From this, for users with low economic strength, compared with the other two types, the number of units is large and the price is low, which is a good choice.

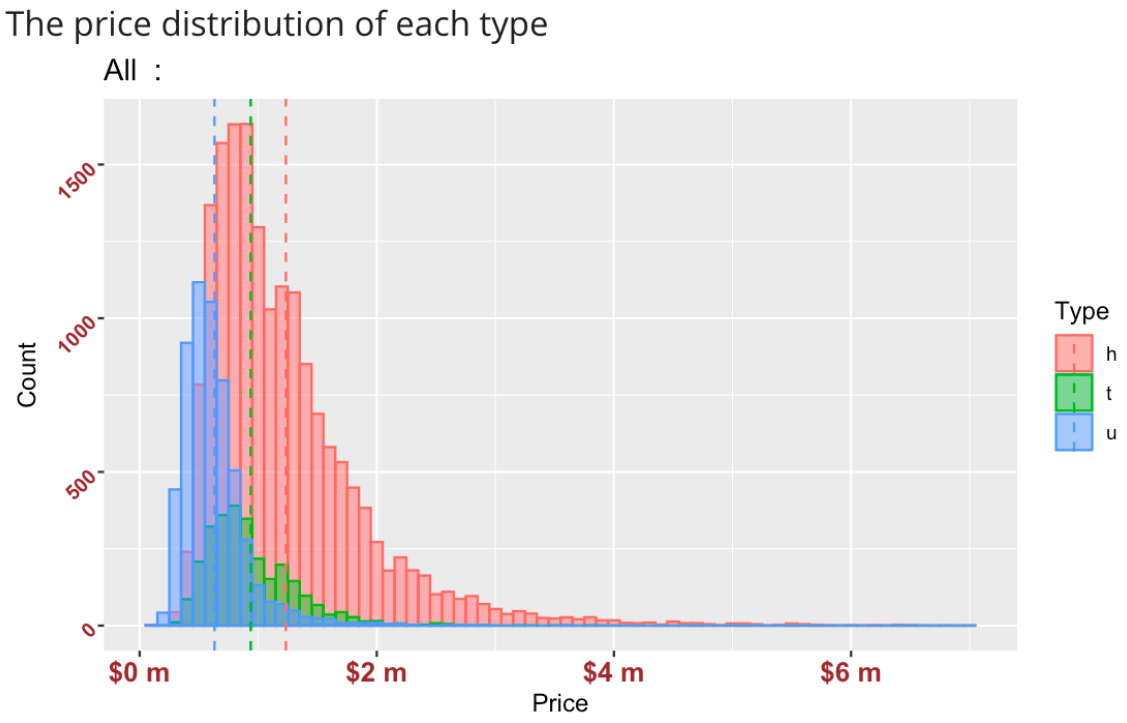


Figure 6 The price distribution of each type of house

4. Conclusion

From this project, not only the skill of using R-shiny to achieve the design of the user interface for visualisation got improved, but also have a better understanding about how to make a plan to complete the visualisation task and the purpose of the visualisation. For example, how to develop the initial designs by using five design sheets at the beginning of the project, how to improve the interpretation level of the graph by combining other kinds of charts, categorizing the chart via implementing some user interaction functions for better graphic explanation.

The difficulty of the project is mainly reflected in the following aspects:

1. Use the least amount of text and the most intuitive graphs to answer the three core questions of the customer

2. Different from general visualization, in order to better explain the problem, the chart has been optimized and structurally deformed. For example, dimensionality of the line chart is reduced by categorising the data into segments by years and use different colors to distinguish them. Besides, three original histograms corresponding to each room type are combined into three so that users can compare more conveniently.

3. For choropleth map, high workload of pre-processing including browsing and determine the most suitable geographical data, and the bins selection to illustrate the distribution by the shade of color.

5. Reference

Goyal, N., Rachapalli, V., Burns, H., & Lloyd, D. C. . (2011). Cervical spine imaging in trauma: Does the use of grid and filter combination improve visualisation of the cervicothoracic junction? Radiography, 17(1), 39–42. https://doi.org/10.1016/j.radi.2010.04.005

Hurter, C. (2016). Image-based visualization: interactive multidimensional data exploration.

Kauko, T. (2004). Towards Infusing Institutions and Agency into House Price Analysis. Urban Studies, 41(8), 1507–1519. https://doi.org/10.1080/0042098042000226975

Sahay, A., & Business Expert Press. (2017). Data visualization. Volume 1, Recent trends and applications using conventional and big data.

6. Datasets of the project

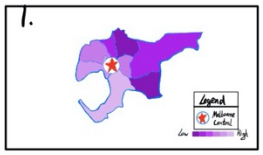
Melbourne housing clearance data from Jan 2016
https://www.kaggle.com/anthonypino/melbourne-housing-market?select=Melbourne_housing_FULL.csv

Victorian Property Sales Report - Median House by Suburb Quarterly
https://discover.data.vic.gov.au/dataset/victorian-property-sales-report-median-house-by-suburb

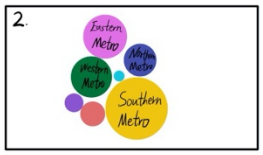
Postcode Boundaries (Polygon) - Vicmap Admin
https://discover.data.vic.gov.au/dataset/postcode-boundaries-polygon-vicmap-admin

7. Appendix A

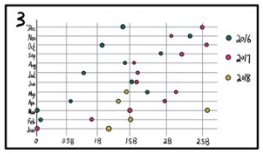
Sheet 1 Brainstorming Charts

1. 

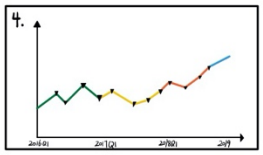
choropleth map

2. 

Bubble chart

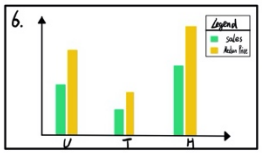
3. 

Scatter plot

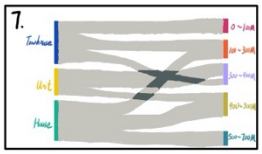
4. 
Multi-coloured line chart

5. 

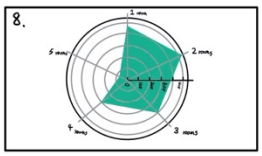
Donut chart

6. 

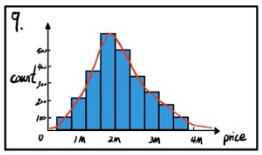
Combined bar chart

7. 

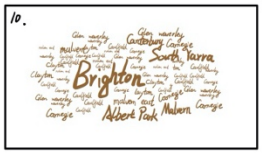
Sankey graph

8. 

Radar chart

9. 

Histogram chart

10. 

Word cloud

Data type:

- time series data (2016.01 - 2019.3)
- Locations (coordinates, zip code, suburb, region)
- numeric (sales, price, growth rate)

Categorize :

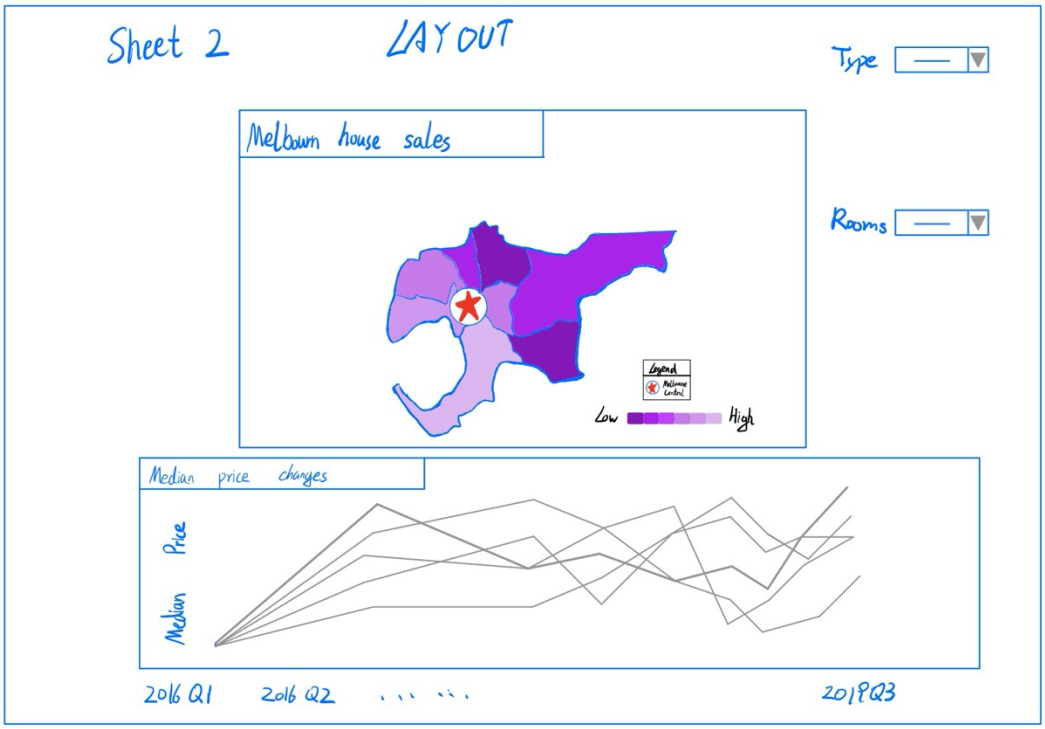
1. Dominant element selection:
   - bubble chart, Word cloud, donut chart
2. Showing trend:
   - Line chart
3. Distribution:
   - scatter chart, sankey graph, map, histogram, radar chart
4. Categorical comparison:
   - bar chart

Filter:

Bubble chart – word cloud can better emphasize the main components, the bubble chart would be hard to read when the data is similar

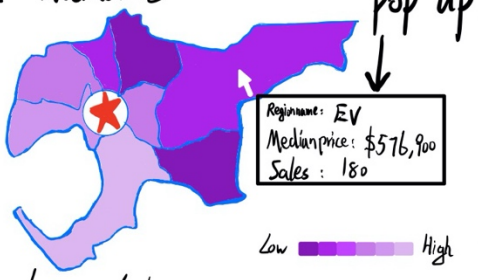
Scatter plot – compared with other distribution graph, this chart is difficult to getting the insights

Sankey graph – not suitable for the data type



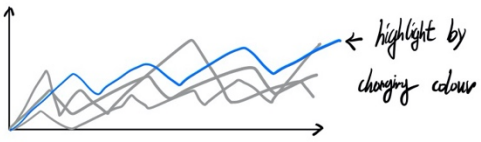
Operation:

1. mouse interaction



2. drop-down list

3. highlight after region selection

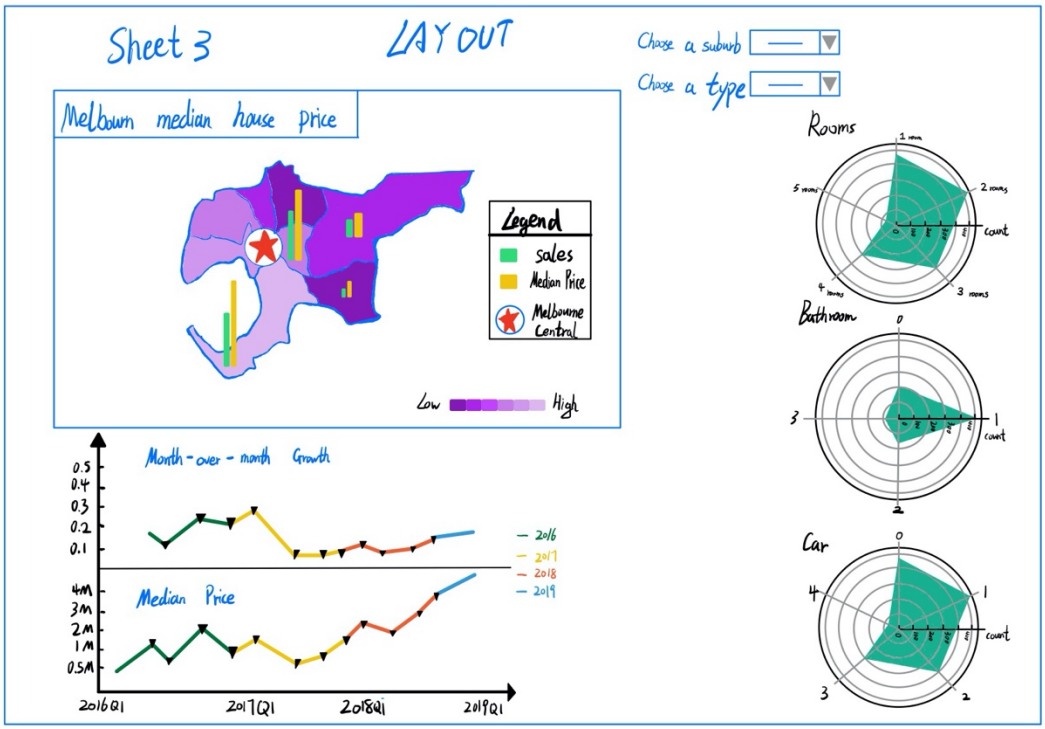


Advantage:

Combine heat map and choropleth map to directly show the sales distribution and allow user to get the price information by selecting their prefer house type

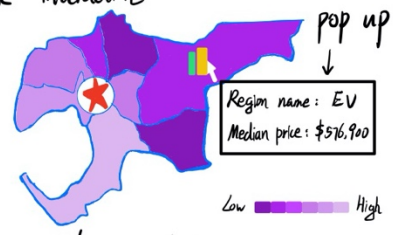
Disadvantage:

The multiple-line chart is a little bit messy

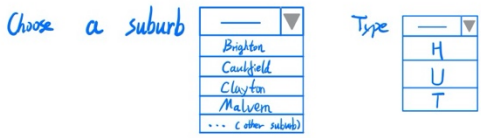


Operation:

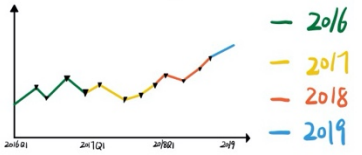
1. mouse interaction



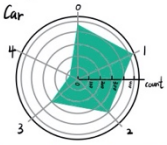
2. drop-down list



3. use color to categorize the trend in years



4. show the house type details



Shaw below message when moving the mouse to the category

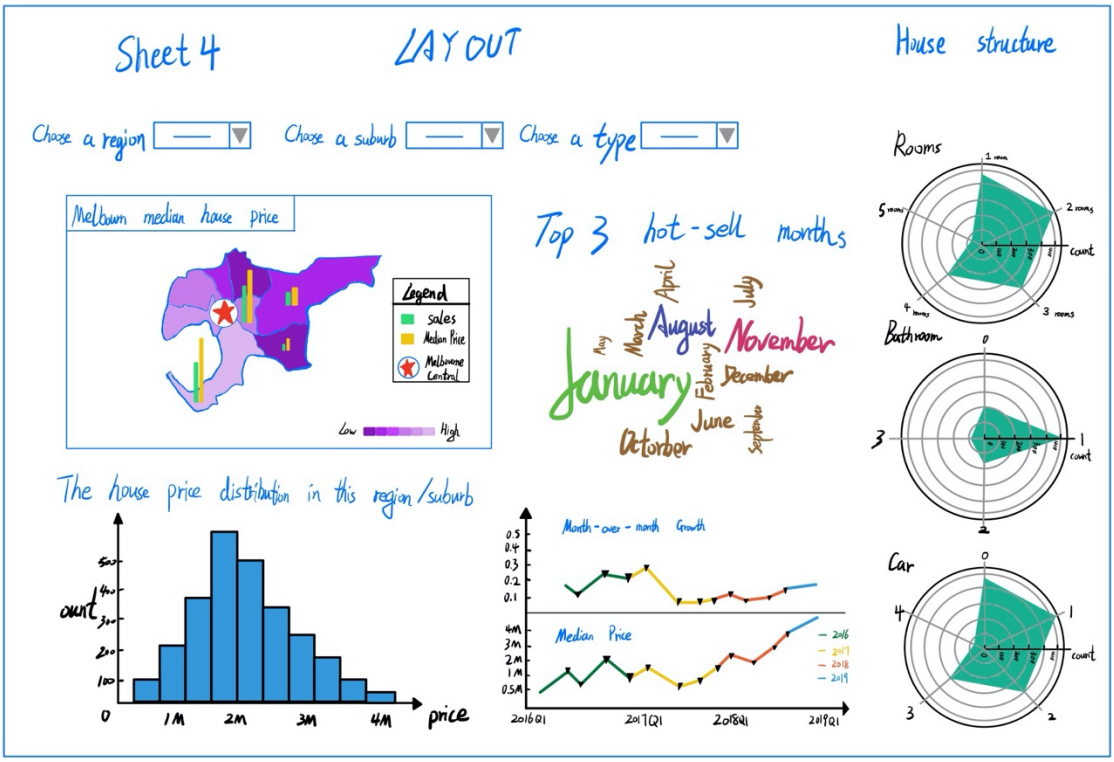
count: 349

Advantage:

- Emphasize the comparison of sales and median price between regions
- Use radar chart to illustrate the distribution of the house type and the house structure
- Show the statistic information (month-over-month growth)

Disadvantage:

The relationship between sales and time is insignificant by using line chart and bar chart



Operation:

1. mouse interaction

2. drop-down list

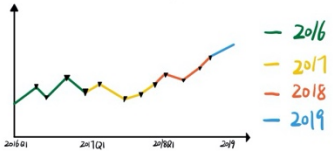
Choose a suburb

- Brighton
- Caulfield
- Clayton
- Malvern
- ... (other suburb)

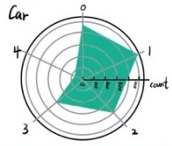
Type

- H
- U
- T

3. use color to categorize the trend in years



4. show the house type details



Show below message when moving the mouse to the category

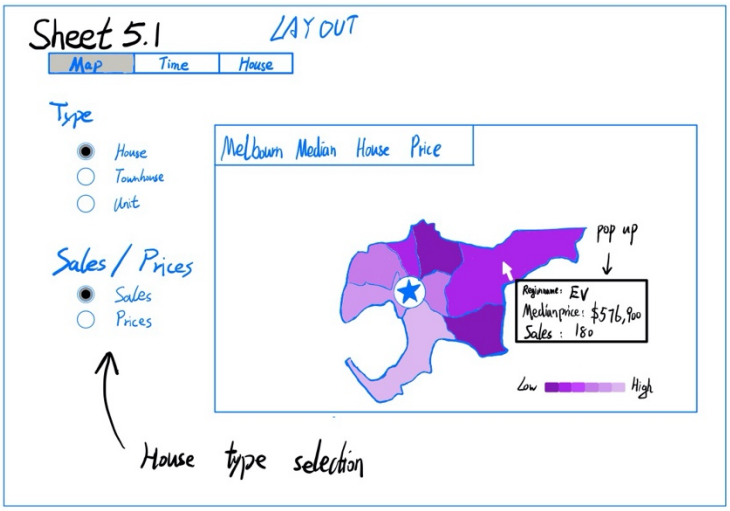
count: 349

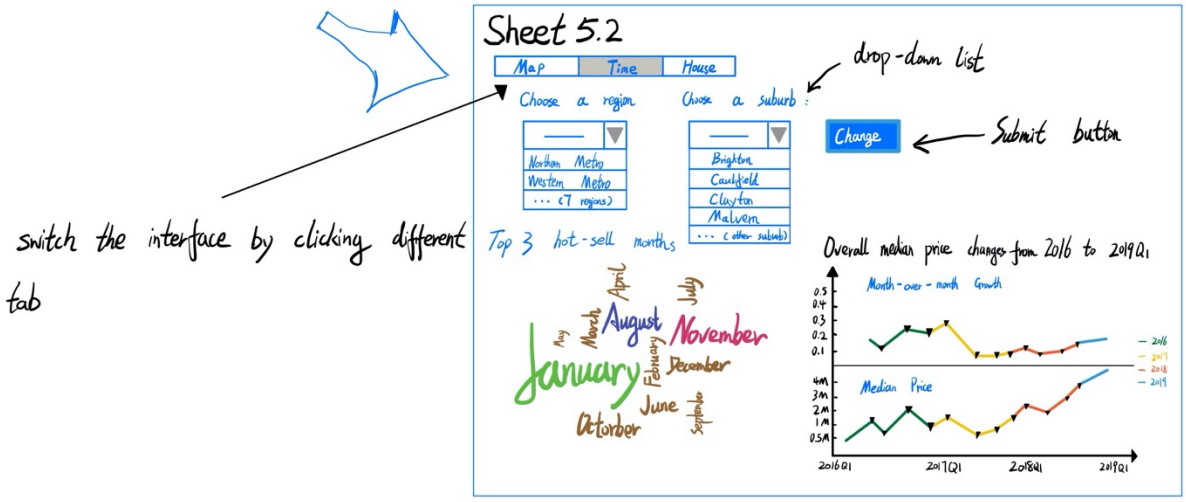
Advantage:

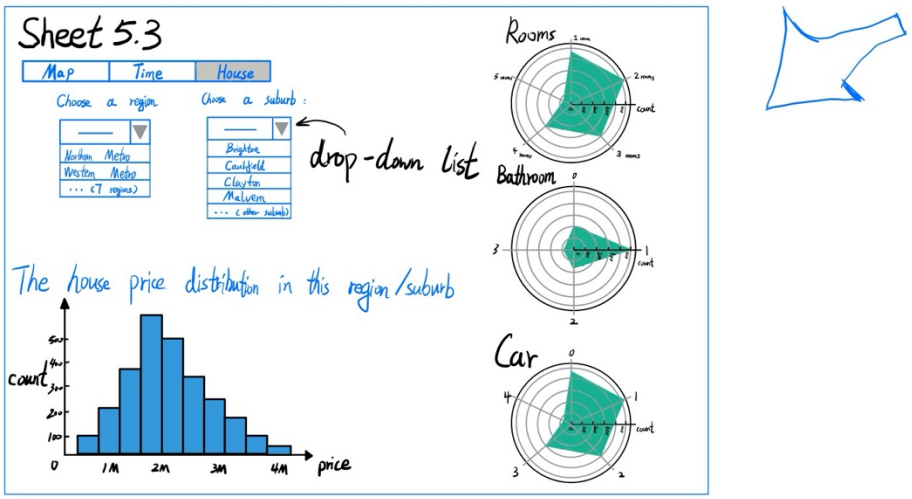
- Word cloud directly shows the dominant component
- Comprehensive
- Provides the relationship between house price and time, location, house type and structure

Disadvantage:

Too crowded and messy


Sheet 5.1
LAYOUT
Map
Time
House
Type
House
Townhouse
Unit
Sales / Prices
Sales
Prices
Melbourn Median House Price
pop up
Regionname: EV
Median price: $576,800
Sales : 180
Low
High
House type selection


switch the interface by clicking different tab
Sheet 5.2
Map
Time
House
drop-down list
Choose a region
Choose a suburb :
Change
Submit button
Northern Metro
Western Metro
... (7 regions)
Brighton
Caulfield
Clayton
Malvern
... (other suburb)
Top 3 hot-sell months
April
July
May
March
August
November
January
February
December
June
October
September
Overall median price changes from 2016 to 2019Q1
Month-over-month Growth
Median Price
2016Q1
2017Q1
2018Q1
2019Q1


Sheet 5.3
Map
Time
House
Choose a region
Choose a suburb :
drop-down list
Northern Metro
Western Metro
... (7 regions)
Brighton
Caulfield
Clayton
Malvern
... (other suburb)
The house price distribution in this region/suburb
count
price
0
1M
2M
3M
4M
Rooms
Bathroom
Car
count