Visualising the Private Property Market in Singapore

Kong Song Wei, Peh Jing Yuan, Tan Zhi Chong

Abstract— Real estate investment has always been a relatively safe choice of investment for people seeking to gain decent returns. Many people dream of retiring early and earning a good passive rental income through their invested property. They tend to have the misconception that real estate investment is an easy way to get rich but in fact, real estate investment is more complicated than most other investment vehicles. This has led to the development of various analytical tools globally to help analysts and investors gain better insights of the property market. All existing map visualisations of the property market in Singapore use a choropleth map for analysis. However, the use of choropleths has certain limitations in property market analysis. As such, we intend to fill this gap by developing a more effective map visualisation for the non-landed private property market through the use of an interactive hexagon binning map. Through our visualisation, we hope to provide analysts and investors with valuable insights of the private property market which cannot be discovered using existing map visualisation tools in the market.

Key Words—Real Estate, Private Property, Investment, Map Visualisation, Hexagon Binning Map, Geographic Information System (GIS)

1 Introduction

Over the past 50 years, real estate investment has become a very popular investment vehicle. This is attributed to the fact that real estate are tangible assets that typically appreciates in value over time. For example, based on the Private Residential Properties Price Index in Singapore, the prices of private residential properties had rose by 44.1 index points from 2006 to 2016. [1] Thus, we often hear of people making huge profits from real estate investment through the media.

As a result, many people have been hopping on the bandwagon to invest in real estate properties, thinking that it is an easy way to get rich. However, real estate investment is actually much more complexed than other investment vehicles such as stocks and bonds. It involves careful consideration of geographical factors, finance factors and legal factors.

As a result, various information systems have been developed to help analysts and investors in identifying potential investment opportunities.

2 Motivation

In Singapore, various property websites have developed visualisation tools that allows users to discover trends and patterns in the property market, which then enables them to identify good investment opportunities.

One of the key factor in influencing the price of a housing unit is the location. Buyers are willing to pay a premium for a housing unit that has greater accessibility to the city, amenities, convenience, prestige. [2] Due to the importance of geographical factors in the housing market, a map visualisation is needed for analysts and investors to identify patterns and correlations related to location.

The map visualisation that many property websites have used is a choropleth map.

A choropleth is a map used by cartographers to show the quantitative variation of a phenomenon among enumeration units such as countries and states. [3] The distribution is shown using graded shading to indicate the density per unit area of a phenomenon – the darker the shading is, the higher the density per unit area.

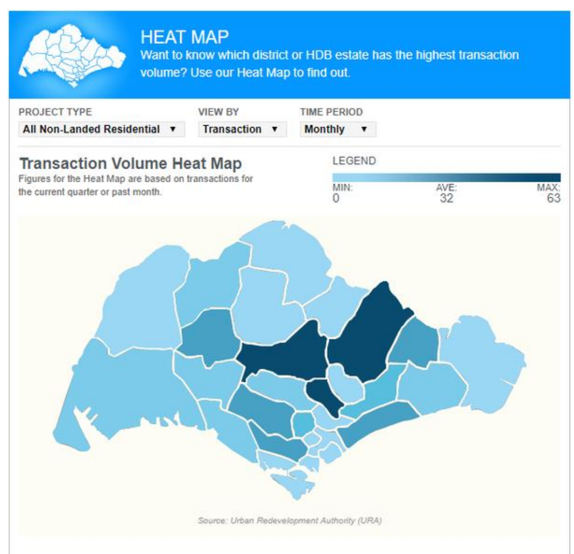

Figure 1

Figure 1 shows an example of a choropleth map from a Singapore property website. In the figure, the transaction volume in the past month has been aggregated to the Singapore planning region level. Each segment illustrated in the map shows a planning region; the darker the segment, the higher the number of sales transaction in the previous month. This visualisation offers a clear spatial overview of the number of housing transactions in the past month. However, it offers an

- *Kong Song Wei is an undergraduate student at the School of Information Systems, Singapore Management University, E-Mail: swkong.2014@sis.smu.edu.sg*
- *Peh Jing Yuan is an undergraduate student at the School of Information Systems, Singapore Management University, E-Mail: jypeh.2015@sis.smu.edu.sg*
- *Tan Zhi Chong is an undergraduate student at the School of Economics, Singapore Management University, E-Mail: zctan.2015@economics.smu.edu.sg*

overly generalised distribution which may mask finer details [4].

An overly generalised distribution without any drill down capabilities makes it impossible to identify the projects that are driving the sales, which in turn makes it impossible to identify potential investment opportunities. Analysts and investors require a higher level of detail for key insights. Therefore, this type of visualisation might be unsuitable for analysts and investors.

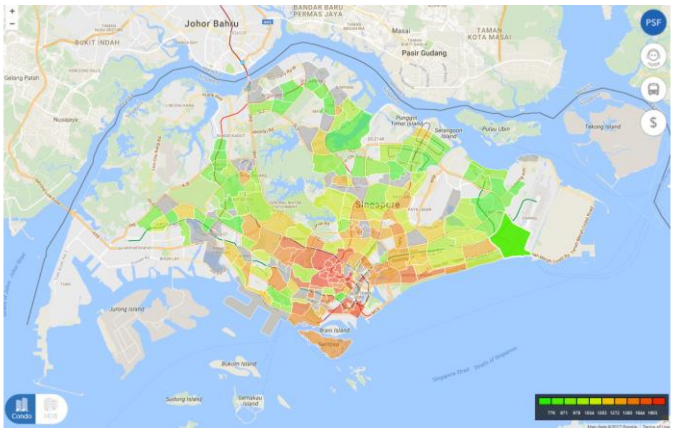

Figure 2

A possible solution is to increase the level of details by segmenting the areas into a finer level. In the case of Singapore, we could possible further segment the planning region into subzones, which has been done by a property website, as shown in Figure 2. An example of a subzone is Anchorvale from the Seng Kang planning region. This would certainly provide greater insights as compared to the previous visualisation. However, subzones in Singapore have a huge variation in size. It might be easier to pinpoint the projects that are driving the sales in small subzones as there are fewer projects in small subzones. However, in large subzones, it is impossible to identify the key drivers of prices or sales as there are more projects there. Hence, the use of a choropleth map with subzones still does not provide sufficient level of details for analysis.

In addition, choropleths can only display a single dimension of the data points within an enumeration unit, which is based on the shading / saturation of the color. If there are more than one dimension of interest within each enumeration unit, multiple choropleths have to be used to identify patterns and correlation between each enumeration unit, hence making it harder to identify correlations. For example, in the property market, analysts and investors are interested in the geographical patterns of the transaction volume as well as the price per square foot (psf). In order to discover patterns between transaction volume and psf, users will have to switch between two choropleths. Visually speaking, our eyes might fail to detect certain patterns because both visualisations are located a distance away from each other. As such, the use of choropleth maps in this situation makes it harder to detect patterns in the property market.

To summarise, the use of choropleth maps has certain limitations in visualising past transaction data in the property market because of the incapability to offer finer details needed for further analysis as well as the inability to show more than one dimension of an enumeration unit.

3 Objective

Our research and development efforts were motivated by a lack of suitable map visualisation in the Singapore housing market as discussed in the previous section. Through our project, we aim to provide analysts, investors and real estate agents with an analytical tool that allows users to discover the finer geographical patterns and time series trend of Singapore's non-landed private property market. More specifically, we aim to provide users with the ability to:

1. Identify projects that are driving the sales in a region
2. Explore time series price and transaction trend of the market to identify the market outlook
3. Explore pricing patterns across different region and different projects
4. Explore possible reasons for the trend in price and sales volume based on location

4 Visualisation Approach

To overcome the limitations of choropleth maps in visualising property transaction data, we have decided to adopt the use of a hexagon binning map.

Hexagon binning is a form of bivariate histogram used to visualise the structure of a large dataset. In a hexagon binning map, data points that are close to each other are grouped together into a hexagon shape bin. Hexagon shape bins are used because hexagon is the polygon with the maximum number of sides for a regular tessellation of a 2D plane. The two other shapes that could form regular tessellations are triangles and squares. [5]

However, a hexagon shape is used because it possesses a few properties that makes it a more effective binning polygon than squares and triangles. Firstly, the distance from the edge to the centre of a triangle and a square has a higher variation than the distance from the edge to the centre of a hexagon. Hence, points inside a hexagon are closer to the centre as compared to triangles and squares, which translates into a more efficient data aggregation around the centre. [6] Furthermore, hexagons are visually less biased for illustrating densities than other shapes. For example, our eyes tend to be drawn to the horizontal and vertical lines of a square. Hexagon, on the other hand, tend to break up the line between the polygons. This comparison can be viewed in Figure 3

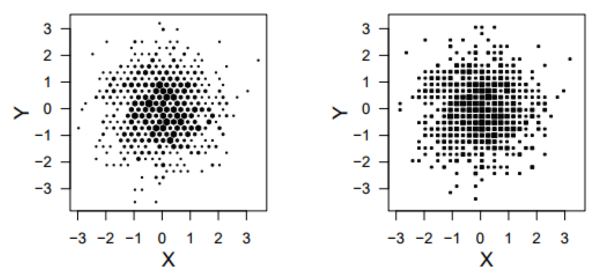

Figure 3

Conventional hexagon binning only visualises a single property of the data points which is done through the shading/ color of the hexagon – the darker the shading/ color, the higher the count of the data points. However, we can also add in another property of the data points through the size of the hexagons. For instance, the size of the hexagon could indicate the standard deviation – where the higher the standard deviation, the bigger the size. [7]

In the context of the property market, we could visualise both psf and transaction volume in a hexagon bin where psf can be indicated by the shading while transaction volume can be indicated by the size of the hexagon. Hence, this allows us to overcome the choropleth's inability to visualise more than one dimension.

A slightly creative approach is required to offer finer level of details of the property market. To allow users to discover the finer details, we will require the use of another hexagon binning map which shows a more detailed view. In total, we will have two hexagon binning maps – one for the Overview Map and one for the Detailed Map.

The Overview Map shows the big picture of the property market across the entire area of Singapore. The entire Singapore will be grouped into different hexagons, forming a honeycomb. Data points that are within the same hexagons are aggregated together and a hexagon will be reflected on the hexagon bin. The size and the color of the hexagon depicted will reflect both the past transaction volume and the average psf of the aggregated units within the hexagon bin.

On the other hand, the Detailed Map shows the hexagon binng map pattern of the selected hexagon from the Overview Map. Similarly, the area within the selected hexagon will be grouped into hexagons forming a honeycomb and a hexagon will be reflected on the hexagon bin to reflect two dimensions from the aggregated data points. Users can further zoom-in on the Detailed Map to view hexagons that are aggregated on a finer level of detail. The Detailed Map acts as a drill down feature to allow users to identify the drivers of price and volume in each hexagon. This feature can only be implemented in segments with the same area, such as a hexagon bin. This feature will fail to work in choropleths as there is a variation in the area high enumeration units, making it hard to detect patterns on enumeration units with higher areas. Therefore, with the use of another hexagon binning map, users will now be able to drill down to explore and discover the key drivers of price and volume in an area.

As such, hexagon binning map overcomes the limitations of choropleths, making it a more effective map visualisation for the property market.

As map visualisations lack the ability to show a time series trend, we have decided to integrate additional charts with our hexagon binning maps to allow users to discover time series trends of the property market. Thus, time series line chart and time-series bar chart will be used to show the trend over time. The line chart will be used to show the average psf over time while the bar charts will be used to show the trend of the housing transaction volume over time. These charts will be integrated with the hexagon binning map to provide a drill down feature to discover the time series trend of the private property market.

5 DATA CLEANING

The first phase for our research is data collection. We will require a dataset that contains all the past transaction data of the housing market in Singapore. The Urban Redevelopment Authority (URA) of Singapore has a Real Estate Information System (REALIS) which contains a huge database that provides comprehensive datasets for this analysis. [8]

In our analysis, we will be focusing on transactions of apartments, condominiums and executive condominiums from 2012 to 2017. The rationale for focusing on transactions between 2012 to 2017 is that the property prices in the past 5 years will be sufficient and more relevant in helping analysts to detect the current trend.

As REALIS only allows user to download a few thousand rows of data each time, we had to download multiple datasets as Comma-Separated Values File (CSV). Thus, it took us some time to download the entire private property data from the past 5 years. After downloading all the necessary rows of data, we wrote a Python script to combine these CSV into one CSV as it would be inefficient to manually combine our data. We have decided to compile our final data in a CSV format as it is easier to process hence reducing the workload on the application. This is because D3.js library uses an in-built method to automatically delimit the dataset and store it into a list of rows. As a result, we can simply access every row of delimited data by calling a loop. On the other hand, alternative file types require more processing to make it suitable for use, which means it will take a longer time for the application to load.

One limitation of the REALIS dataset is that it lacks the coordinates required to plot the data on a map visualisation. Thus, we will require the use of a geocoder to obtain the longitude and latitude of each transaction unit. Using Javascript, we wrote a geocoding script that calls OneMap's API to geocode over 5000 unique postal codes in our dataset to obtain the longitude and latitude of each transaction. Due to the large number of postal codes required, we had to write a script to obtain the longitude and latitude instead of manually deriving it from Google Map or OneMap.

6 USER INTERFACE DESIGN

The next phase of our research is the implementation of our application. Before developing our application, we first researched on various charting libraries and designs to find out the optimal way to create our visualisations.

Eventually, we decided to use the D3.js library [9] due to its flexibility in customizing charts, the large pool of resources available and its ease of loading CSV files.

Firstly, to implement our hexagon binning maps, we used Hexbin.js, a D3.js plugin to create the hexagon grids. Next, we used Leaflet.js [10] to integrate the Singapore Map with the hexagon bins created in the previous step. Lastly, we used DC.js [11] library's MapReduce functionality to create charts that could be easily filtered based on the selected parameters.

The following description shows the design considerations in this visualization tool.

A) Filtering Elements

1) Property and Sales Type Filter

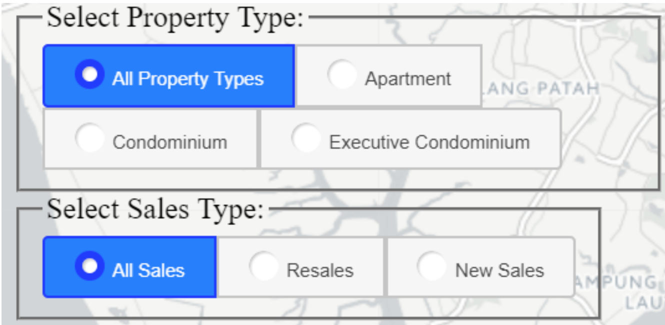

Figure 4

Filters with radio buttons are created to provide users with the functionality to filter the dataset according to the selected property and/or sales type. By default, all property and sales type are selected as seen in Figure 4. If analysts are interested in discovering the map visualisation of another Property Type or Sales Type, they can simply select their preferred parameters by clicking on the radio buttons.

2) Date Range Filter

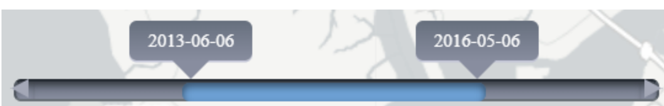

Figure 5

The functionality of the date slicer is to filter the dataset according to the selected date range. The Date Slider can be seen in Figure 5. By default, it is ranged from the beginning of 2012 to 2017. Users can adjust the slider to change to their preferred time frame for analysis.

3) Filter by Hexagon

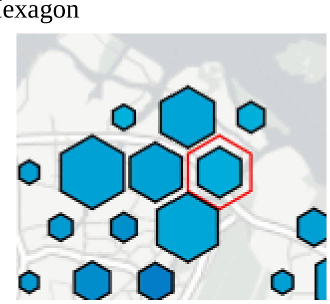

Figure 6

Users can select the hexagons from the Detailed Map to filter the dataset according to the selected bin. By default, the line charts and bar charts will show the price trend and transaction volume across Singapore. When analysts select a hexagon bin, the line and bar chart will only reflect transactions that has occurred within the area of the hexagon bin, which is indicated by a red outline as shown in Figure 6. This allows analysts to drill down a specific area to analyse the price and volume instead of a generic planning area.

B) Charts and View

1) Overview of visualization

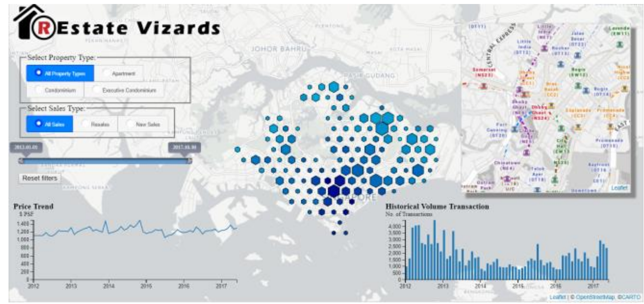

Figure 7: Overview of Visualisation

Our visualisation tool is designed to fit into a web browser without the need to scroll, thus enhancing readability and convenience.

2) Overview Map

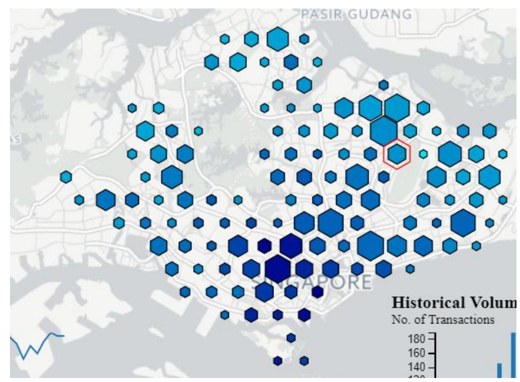

Figure 8: Overview Map

The Overview Map shows the past transaction data points grouped into hexagon bins. The size of the hexagon indicates the number of transactions in the selected period; the higher the number of transactions, the larger the hexagon. On the other hand, the color indicates the average psf over the selected period; the darker to color, the more expensive the psf.

3) Tooltips for Overview Map

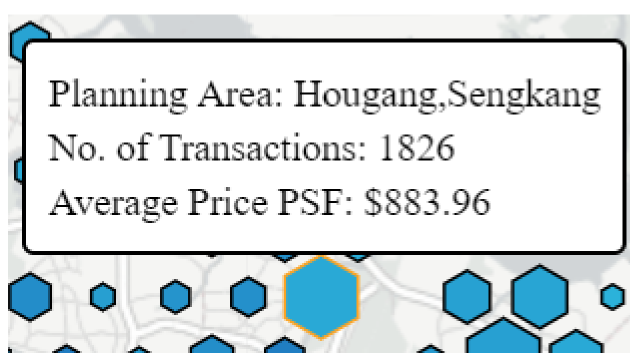

Figure 9: Tooltip for Overview Map

Tooltip is included for each hexagon bin on the Overview Map. When analysts mouse over a hexagon bin, they are able to see the planning area(s), number of transactions as well as the average psf of the data points in the hexagon bin area. Thus, it provides additional information not seen in the map for further analysis.

4) Detailed Map

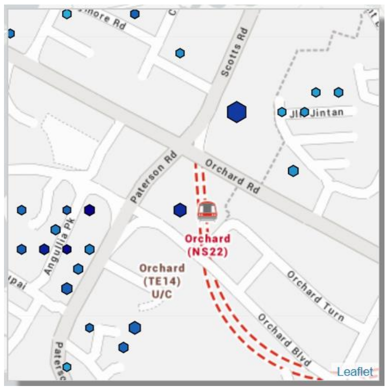

Figure 10: Detailed Map

Hexagon binning map showing the Detailed View is implemented on the top right corner of this visualization tool. When analysts click on a hexagon bin in the Overview Map, the Detailed Map will zoom into that selected bin area and show all the projects and transactions in that bin. In this Detailed Map, analysts can then explore the price and transaction patterns in that specific area. Users can also zoom-in further to explore hexagon bins that are aggregated at a finer level and also to explore the amenities in the region. This Detailed Map allows analysts to drill down from the Overview Map to explore what are the factors that may influence the price and volume of a region.

5) Tooltip for Detailed Map

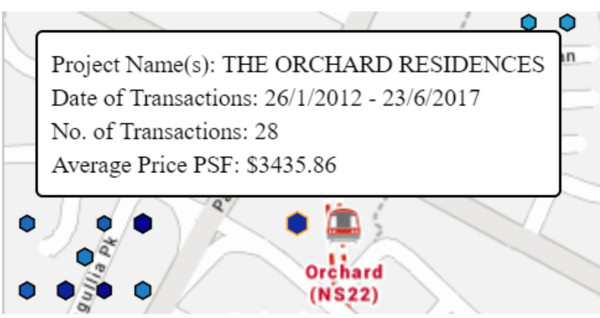

Figure 11: Tooltip for Detailed Map

Tooltip is included for each hexagon bin on the Detailed Map. When analysts mouse over a bin, they can see the project names, date of transactions, number of transactions as well as the average psf, thus allowing analysts to identify the details of the project.

6) Price Trend Chart

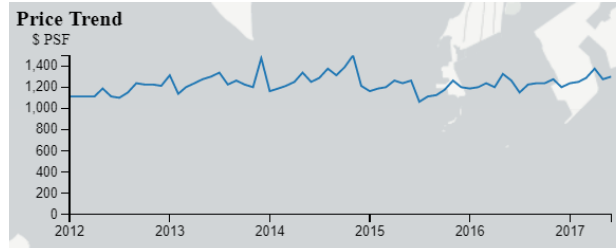

Figure 12: Price Trend Chart (line chart)

A time series trend of the average psf is shown using a line chart. If no hexagons are selected, it shows the psf trend over time for the entire property market. However, if hexagons are selected from the Overview Map or the Detailed Map, then the line chart will reflect the pricing trend of the selected hexagon

7) Historical Volume Chart

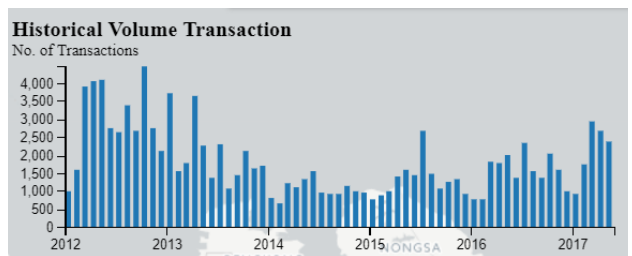

Figure 13: Historical Volume Chart

A time series trend of the volume transaction is shown using a bar chart. If no hexagons are selected, it shows the transaction trend over time for the entire property market. However, if hexagons are selected from the Overview Map or the Detailed Map, then the bars will reflect the transaction trend of the selected hexagon.

7 WEB APPLICATION ARCHITECTURE

This visualization tool is designed and developed using JavaScript programming language. This is because it has a high flexibility in customizing charts and integrating them. Our code uses hexbin.js to create hexagon binning grids and leaflet.js to add this grid layer on top of the Singapore map. The code then uses DC.js to create the line and bar chart. Finally, it uses Crossfilter.js to allow the charts and maps to interact between one another using map reduce method.

For the Singapore map in the visualization tool, the main code also calls API provided by OneMap to display the map. In addition, there is a separate geocoding that calls OneMap's API to geocode all the postal codes in our dataset.

8 CASE STUDY

To demonstrate how this application could be used by analysts, we will look at a case study.

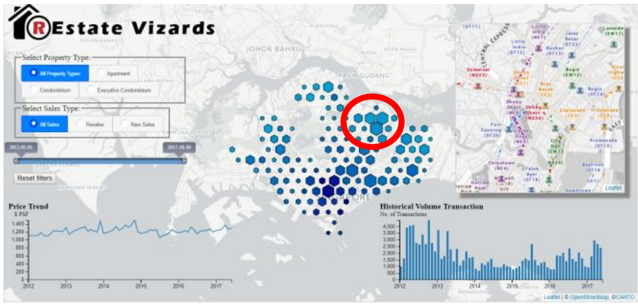

Figure 14

In the initial overview of the visualisation from Figure 14, we can immediately observe some patterns in the market. Firstly, over the past 5 years, there appears to be many transactions in the North-east region of Singapore, specifically in Hougang, Sengkang and Punggol region. These huge transactions could be driven by either newly built condominiums or by huge number of resale units. To identify what is driving the transaction in the region, we could make use of the 'Sales Type' filter.

After switching the sales type filter between New Sales and Resales, we can observe that these transactions are mainly driven by New Sales, which indicates that there has been many newly built condominium in the North-east Region.

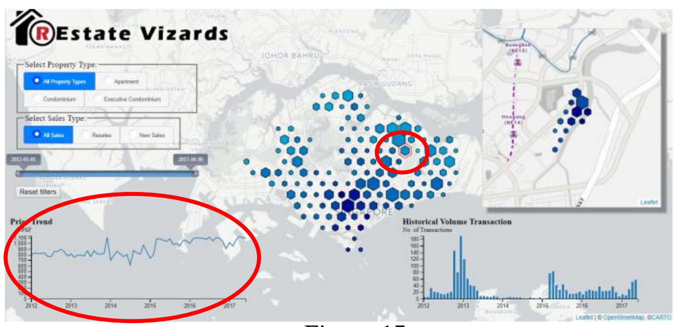

Figure 15

After discovering that there have been many newly built condominiums in the North-east region, the analyst might be interested to identify any interesting patterns in these regions. Upon selecting a hexagon, the selected hexagon will have a red outline to indicate that it has been selected as seen in Figure 15. The outline of the hexagon also shows user the boundary of the selected hexagon.

After selecting one of the hexagon in the North-east area, there seems to be a surge in psf in the region over the past 5 years as seen in Figure 15. After clicking on each hexagon in the Detailed Map to explore the patterns of each condominium, we can observe that the increase in psf in the region is due to the new sales of the more expensive newly built condominium, Kingsford Waterbay.

However, for most projects within this cluster, the price seems to be facing an upward trend. This might indicate that demand is increasing for the housing units in this vicinity, possibly due to future developments. As such, this provides a starting point for further analysis in discovering what is causing the increase in price. Analysts will have to supplement this visualisation with news reports of future development in this area to offer an explanation to the rising trend. If analysts were to use existing visualisations in the market, they would fail to discover interesting insights like this.

9 FURTHER WORK

This web application could be extended further to include the private property market rental data. Prices of private properties and rental of private properties go hand in hand as price reflects the cost of investment while rental of private properties reflect the returns of investment. Thus, analysts and investors will also be interested in discovering patterns of the rental market and how it correlates to the private property housing market. Therefore, a possible extension to this project is to include a toggle that allows analysts to switch between hexagon binning map of a housing market or hexagon binning map of the housing rental market.

10 CONCLUSION

Through this project, we hope to provide analysts, investors and real estate agents with a more effective visualization tool for the private property market as compared to the existing tools in the market. Our visualization tool enables analysts to explore and discover valuable insights in the non-landed private property market regarding price and transaction volume across time and locations. Through our visualization, we hope analysts will be able to detect useful patterns in the property market and thus guide them into doing further analysis. This application mainly serves as an exploratory platform and will therefore require qualitative research such as news reports of recent developments to understand the reason for certain patterns.

ACKNOWLEDGMENTS

The authors wish to thank Professor Kam Tin Seong and Instructor Prakash Chandra Sukhwal for their kind support and guidance throughout this project

REFERENCES

[1] Department Of Statistics Singapore. (n.d.). Retrieved November 20, 2017, from http://www.singstat.gov.sg/statistics/visualising-data/charts/private-residential-properties-price-index

[2] Is Location Really Everything When It Comes to Buying a House in Singapore? (2015, July 27). Retrieved November 20, 2017, from https://blog.moneysmart.sg/property/is-location-really-everything-when-it-comes-to-buying-a-house-in-singapore/

[3] Stewart, J., & Kennelly, P. (2010). Illuminated Choropleth Maps. *Annals of the Association of American Geographers, 100*(3), 513-534.

[4] Mayhew, S. (2009). Choropleth. *A Dictionary of Geography,* A Dictionary of Geography.

[5] Lewin-Koh, N (2011) Hexagon Binning: An Overview. Cran.rproject.org/web/packages/hexbin/).

[6] Graham. (n.d.). An Introduction to Hexagonal Geometry. Retrieved November 20, 2017, from https://hexnet.org/content/hexagonal-geometry

[7] Nelli, F. (n.d.). Hexagonal Binning - a new method of visualization for data analysis. Retrieved November 20, 2017, from http://www.meccanismocomplesso.org/en/hexagonal-binning/

[8] Real Estate Information System (REALIS). (n.d.). Retrieved November 21, 2017, from https://spring.ura.gov.sg/lad/ore/login/index.cfm

[9] Bostock, M. (n.d.). Data-Driven Documents. Retrieved November 20, 2017, from https://d3js.org/

[10] Agafonkin, V. (n.d.). An Open-source JavaScript Library for Interactive Maps. Retrieved November 20, 2017, from http://leafletjs.com/

[11] Nick, Q. (n.d.). Dimensional Charting Javascript Library. Retrieved November 20, 2017, from https://dc-js.github.io/dc.js/