

Development of an Ontology-Based Visual Approach for Property Data Analytics

Xin Janet Ge*, Jinson Zhang

School of Built Environment, DAB, University of Technology Sydney, Sydney, Australia

*Corresponding author's email: XinJanet.Ge@uts.edu.au

Article history: Received: 29 August 2020 Received in revised form: 19 November 2020
Accepted: 27 December 2020 Published online: 23 June 2021

Abstract

Real estate is a complex market that consists of many layers of social, financial and economic data, including but not limited to price, rental, location, mortgage, demographic and housing supply data. The sheer number of real estate properties around the world means that property transactions produce an extraordinary amount of data that is increasing exponentially. Most of the data are presented through thousands of rows on a spreadsheet or described in long paragraphs that are difficult to understand. The emergent data visualisation techniques are intended to allow data to be processed and analytics to be displayed visually to enable an understanding of complex information and the identification of new patterns from the data. However, not all visualisation techniques can achieve such a thing. Most techniques are able to display only visual low-dimensional data. This paper introduces an ontology visualisation methodology to explore the ontologies of property data behaviour for multidimensional data. The visualisation combines real estate data statistical analysis with several high-dimensional data visualisation techniques, including parallel coordinates and stacked area charts. By using six residential suburbs in Sydney as a demonstration, we find that the developed data visualisation methodology can be applied effectively and efficiently to analyse complex real estate market behaviour patterns.

Keywords: Data visualisation, real estate property market, property data behaviour ontologies, high-dimensional data, ratio behaviour pattern

© 2021 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

The residential property market is a very complicated industry that is closely integrated with broader economic trends, as well as being affected by the demand for and supply of housing. The behaviours of households—including buyers, renters and sellers—are shaped by the complex, multidimensional and heterogeneous peculiarities and the ‘neighbourhood’ characteristics of property, including location, size and quality of construction, among other factors. Understanding how the demand for and supply of housing come together to shape household behaviour is important when predicting market changes and assisting the government to make informed decisions. However, the analysis and communication of property market performances are challenged by the extensive high-dimensional, multivariate characteristics of property data that are generated every day (Dawidowicz et al., 2014).

Data are information that can be used for making decisions only once they have been processed and analysed in some manner. However, most property data are presented through thousands of rows on a spreadsheet or described in long paragraphs. These types of unprocessed raw data show a collection of information without revealing the relationships between them or displaying data patterns; they are frequently difficult to understand, and new knowledge cannot be added by the users. The ontology data model, on the other hand, displays the data relationships and enables automated reasoning about data. The ontology data model can also be applied to a set of data to create a visual representation of the data, such as a knowledge graph, to enable the viewer to grasp the true meaning of the insights (Li et al., 2016). Data visualisation can be defined as the use of images or charts to present data. A good data visualisation can enable the user to process complex information, remove the noise from data and facilitate the identification of new patterns, as well as present analytics visually (Chen et al., 2007). The benefits of data visualisation include, but are not limited to, communicating trends, analysing and processing information efficiently, and showing multiple attributes in one graph when high-dimensional visualisation techniques are used. However, not all visualisation techniques can achieve such a thing. Most visualisation techniques only visualise low-dimensional data.

Data visualisation techniques have been widely used in many industries and by many organisations, such as NASA (National Aeronautics and Space Administration, 2018), who have combined 3D-visualisation techniques, animations and images to develop the NASA Visualization Explorer, which visualises our sun and the universe (NASA, 2018). Hans Rosling famously used five-dimensional data visualisation techniques in a TED Talk in 2007 to illustrate the relationship between a country's regional classification, life expectancy, income per person and total population over a 200-year time series (see Gapminder Foundation, 2009).

Property data visualisations are increasingly utilised in the real estate industry. For example, Sun et al. (2013) combined geography mapping and web techniques to analyse Hangzhou's real estate market data. Their approach combined geographical maps to illustrate

property location, clustered pixel bar charts to demonstrate property prices and stacked line area charts to indicate other data trends. Another application can be found in research from Li et al. (2018), who analysed real estate data in Australia and proposed a system to provide suburban property information to buyers. They used a variety of different visualisation tools for multidimensional data analytics, including geographic maps to illustrate the property's locations and prices, parallel coordinates to illustrate the relationship between property types and prices, line charts to demonstrate property price trends over time, word clouds to display the frequency of each property's characteristics and spider charts to measure distances between each property and the nearest school zone.

To date, property data visualisations have only focused on the property's location, price trends or the relationship between real estate demand and supply. To the best of our knowledge, no previous work has used property data visualisations to develop a systematic methodology to examine and establish property behavioural patterns. The study of property behavioural patterns is necessary to develop an understanding of changes in property market conditions due to changes in household consumption and rental behaviours, as well as housing supply behaviours affected by government policies, planning regulations, investor appetite and residential development appetite.

This paper focuses on developing an ontology-based visualisation methodology for property data analytics using visualisation. Using an effective data visualisation technique may provide a systematic way of presenting and analysing property data trends and patterns more efficiently. By identifying and creating a dataset consisting of multidimensional property data, we can categorise property data into a number of groups based on the characteristics of property markets. We can then create behavioural ratios such as property price and household financial status in order to understand property data patterns that reflect household behavioural changes in property markets. Developing this ontology-based visualisation methodology is the main contribution of this paper. The second contribution of this paper is to classify property data based on property data behaviour ontologies, which can help simplify the process of data analysis. Thirdly, the introduction of behavioural property ratios can help the analytical process. In addition, this paper also demonstrates that the data visualisation developed from the methodology can identify the main variable changes at any particular point of time and analyse behavioural patterns associated with these variable changes.

The next section reviews recent works and applications of data visualisation and analytics in the property industry. We then develop a methodology of property data visualisation, including: a) exploring property data as a matrix dataset, b) classifying data behaviour ontologies into six data dimensions, c) creating nine behaviour ratios that measure the behavioural relationship between the variables and d) visualising these behaviour ratios using parallel coordinates and stacked bar charts. Our empirical study will then apply the developed methodology to analyse the relationships of two sets of variables in a real-life example: dwelling prices and household financial attributes for six residential suburbs in Sydney. The demonstrated visualisation techniques can be applied to illustrate new and otherwise hidden behaviour patterns, and significantly improve the analytical method for property data visualisation.

■ 2.0 RELATED WORKS

Most contemporary research on property data visual analytics has focused on either geographic visualisation or time series visualisation. Geographic visualisation overlays property data onto a geographic map demonstrating the property's location and other data characteristics. Time series visualisation illustrates changes in a single property attribute over a period of time, and is often used for property price predictions and market forecasting.

2.1 Geographic Visualisation

The earliest academia on property data visualisation was from Williamson and Shneiderman (1992), who established a query system called HomeFinder which explored real estate data through scatter maps and provided a visualisation of both query formulation and the corresponding results. A few years later, Hong (1999) developed a visualisation system called ReV (Real estate Visualizer), which displays an overview of all properties in different-coloured points on a geographical map. Each coloured point contains major property attributes for the property, such as price, number of bedrooms etc., which provides an extensible means of collating and visualising high-dimensional data in property listings in an intuitive and navigable way. Takatsuka and Gahegan (2002) developed a visualisation tool called GeoVISTA Studio, which uses a Java-based visual development environment to generate geographic maps for property location, and then applies parallel coordinates to visualise data relationships between properties. Sun et al. (2013) used geographic maps to show property information in combination with other visualisation techniques such as treemaps and bar charts. Daana and Huang (2013) forecasted real estate property prices using geographic maps by combining sensitivity analysis with visual interaction to analyse the most influential predictors in the property information system and make predictions on the market prices of similar properties. Similarly, Li et al. (2016) developed a geographic mapping system called HouseSeeker, an interactive real estate visualisation system that assists users in understanding the real estate market by finding candidate properties that satisfy their requirements and visually comparing candidate properties. Cheung (2017) proposed a model to value property prices based on attribute differences with their neighbours and used geographic mapping at both macro and micro levels. Chang (2017) analysed housing and rental demand in Hong Kong's housing market through both geographic mapping and line charts. Adebara et al. (2020) used the mapping technique to show the spatial patterns between the rental value and cultural heritage sites in a study of Nigerian residential property values. However, these previous works either provide property information only for the purpose of searching for properties without processing the information (e.g. Li et al., 2016; Williamson & Shneiderman, 1992) or providing only low-dimensional visual charts (e.g. Chang, 2017; Cheung, 2017). None of the papers has used property data visualisations to develop a systematic methodology to examine and establish property behavioural patterns.

2.2 Time Series Visualisation

Time series data visualisation analyses a single dataset over a period in order to discover underlying patterns, identify property price trends and predict future developments. Specifically, time series property data visualisation displays periodic behaviour for better estimating future price trends, property demand and periodic characteristics. Time series visualisation is one of the most common methodologies for analysing property data, most commonly with time on the horizontal axis and price or rent etc. on the vertical axis. Miles (2016) analysed the Hong Kong housing market between April 1973 and January 2014 for twelve regional markets and used fifteen line charts to measure the annual returns for each submarket. Yang and Zhang (2016) analysed the US real estate cycle between 1990 and 2012, including the US subprime-related crisis, and used 19 line charts to present multiple property and finance datasets, such as interest rates, mortgage originations, mortgage security issuances, conventional loan issuances, consumer credit expansion, housing supply, house price indices, housing affordability indices and consumer sentiment indices. The above works focused on low-dimensional visualisation (line charts) that display limited data information and do not show the relationships between the listed variables.

2.3 Visual Analytics and Comparison of Property Attributes

Property data visual analysis often combines multiple attributes of property prices or market analysis, such as demand prediction or affordability analysis, for the purpose of comparing related attributes and extracting new trends. These comparisons often utilise both geographic information and time series data in creating a multi-dimensional dataset. Chang and Chen (2018) studied Taiwan's house demand based on housing affordability, and analysed various metrics such as housing price–income ratio, owner–occupier ratio and housing vacancy ratio using line charts and bar charts. Mottaleb et al. (2016) compared household income and house prices in Bangladesh between 1961 and 2011 using time series line charts. Bourassa and Haurin (2017) used several line charts to show US house prices and various forms of purchaser costs, including mortgage interest rates, property tax rates, property insurance, transaction costs, and depreciation and maintenance. Deng et al. (2018) developed global public real estate price indices and returns for seven countries (US, HK, Japan, Australia, Singapore, UK and France), creating eighteen line charts for comparison analytics between different countries. Van Order (2018) studied the US mortgage securitisation and structure through line charts and area charts to compare mortgage originations, loans delinquents and loans issuances that may have caused the 2007 Global Financial Crisis.

However, the visual analytics outlined above only focus on low-dimensional data, comparing a limited number of attributes such as property location, property price, household income, household income–price ratio or mortgage interest rate. Currently, there has been no academic work studying high-dimensional property data and behaviour patterns, such as the behaviour pattern between the property dwelling prices and household finance status across both different time series and different geographic locations. This is the key gap that we seek to address in this paper.

3.0 DEVELOPING AN ONTOLOGY-BASED VISUAL METHODOLOGY ANALYSING PROPERTY DATA

Property data frequently contains multi-dimensional data extracted from different data sources and presented in different data formats with different data measurements. This property data may include a multitude of attributes, such as property type, property location, property ownership status, dwelling status, price trend, household finance information, regional and national population, and so on. These different categories of data contain many hidden patterns and causal relationships. Developing a systematic visual methodology enhances the user's ability to discover data behaviours and hidden patterns. Four steps are included in developing the methodology (Figure 1), namely: identifying property datasets, classifying datasets based on behaviour ontologies, establishing behavioural property ratios for analysis and visualising the data behaviours.

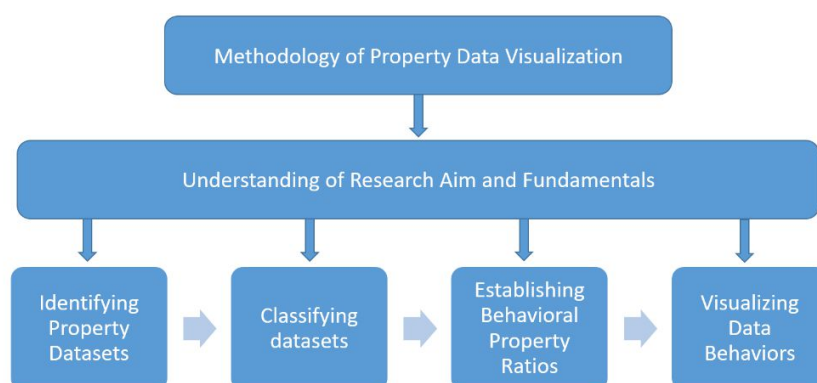


Figure 1 Steps of property data visualisation

3.1 Property Dataset

Property data may be either qualitative data (property type, dwelling status, property location etc.) or quantitative data (property price, household financial information etc.) that can be presented in different formats. This makes property data difficult to analyse and visualise using typical statistics methods. For example, the Australian Bureau of Statistics (ABS) Census website profiles the Sydney suburb of Bondi Beach as follows:

Bondi Beach in 2016: Population 11656. Families 2468. Total private dwellings 6498. Median weekly household income \$2253. Median monthly mortgage repayment \$2708. Median weekly rent \$630. Birth in Australia 45.3%, England 8.7%, New Zealand 3.4%, South Africa 2.3%, USA 2.0%, Brazil 2.0%. Separate house 4.5%. Semi-detached 11.2%. Unit 83.9%. Fully owned 16.5%, owned with mortgage 17.7%, rented 61.7%. ... etc.

Within this short profile, there are many different types of data presented as whole numbers, prices and percentages. However, the suburb profile on its own is unable to show the relationship between these datasets. For instance, the relationships between household income and mortgage repayments, population and the type of dwellings, property price and mortgage loans, and the correlations are not presented and cannot be analysed.

To address this problem, we create a dataset D that contains the entire multi-dimensional property data, with n time incidents with m dimensions. This dataset can be denoted as:

$$D = \begin{pmatrix} d_{1,1} & d_{1,2} & d_{1,3} & \cdots & d_{1,j} & \cdots & d_{1,m} \\ d_{2,1} & d_{2,2} & d_{2,3} & \cdots & d_{2,j} & \cdots & d_{2,m} \\ d_{3,1} & d_{3,2} & d_{3,3} & \cdots & d_{3,j} & \cdots & d_{3,m} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ d_{i,1} & d_{i,2} & d_{i,3} & \cdots & d_{i,j} & \cdots & d_{i,m} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ d_{n,1} & d_{n,2} & d_{n,3} & \cdots & d_{n,j} & \cdots & d_{n,m} \end{pmatrix} \quad (1)$$

The first property incident data dI contains m -dimension attributes, meaning that the data node can then be illustrated as: $d_{1m} = \{d_{11}, d_{12}, d_{13}, d_{1j}, \dots, d_{1m}\}$. Within this data node, j indicates the j th dimension and attribute d_{1j} illustrates the 1st data incident in the j th dimension. Therefore, $j=1,2,3, \dots, m$ indicates the number of dimensions and $i=1,2,3, \dots, n$ indicates the number of time incidents. For example, if $n=1, 2, 3, \dots, 20$ represents a 20-year time serial from 1996 to 2016, and $m=1, 2, 3, \dots, 10$ represents 10 data dimensions, such as property location, property type, property price, suburb population, household income, mortgage payment, ownership status, mortgage status, rental status and distance to school, then the total number of attributes in the dataset will be $n \times m = 20 \times 10 = 200$. For complex datasets, the number of attributes can reach hundreds, even thousands of attributes, with thousands of data points for each data node representing thousands of properties.

3.2 Property Data Behaviour Ontologies

In the Oxford Dictionary, the word ‘behaviour’ is defined as “the way in which one acts or conducts oneself, especially towards others”. Human behaviour generally consists of actors, operations or actions, interactions between the actors, and the property or object acted upon. In a real estate market context, property market behaviour can therefore be defined as a set of characteristics of the property and operations acted on the property, such as the property location, dwelling status, demand and supply status, property price, buyer and seller’s personal status, buyer and seller’s finance status etc. The breadth of potential property behaviours means that the different types and formats of these attributes need to be categorised.

To efficiently visualise the property dataset, we classify the behavioural property data into 5W+1H categories (When, Where, What, Why, Who, How) based on the property characteristics. Each behavioural property data contains those 5W+1H dimensions, which stand for: ‘When’ the property data occurred, ‘Where’ the property is located, ‘What’ the property’s characteristics are, ‘Why’ pay (ask) the particular price, ‘Who’ buys (sells) the property and ‘How’ to pay for the property. These 5W+1H dimensions are illustrated by using six subsets showing in Table 1:

Table 1 5W+1H Six subsets in ontology and attributes

Subset	Data series	Ontologies	Property attributes
T	$\{t_1, t_2, t_3, \dots\}$	When	<i>Time series</i>
P	$\{p_1, p_2, p_3, \dots\}$	Where	<i>Location attributes</i>
X	$\{x_1, x_2, x_3, \dots\}$	What	<i>Dwelling attributes</i>
Q	$\{q_1, q_2, q_3, \dots\}$	Why	<i>Price attributes</i>
Z	$\{z_1, z_2, z_3, \dots\}$	Who	<i>Ownership attributes</i>
Y	$\{y_1, y_2, y_3, \dots\}$	How	<i>Financial attributes</i>

Each data node represents one of the 5W+1H dimensions, which collects all attributes belonging to that category. For example, $X=\{x_1, x_2, x_3, \dots\}$ represents the property’s physical characteristics, such as $x_1=house, x_2=two-storey house, x_3=4x bedroom, x_4=2x bathroom, x_5=1x carpark, \dots$ etc. $Y = \{y_1, y_2, y_3, \dots\}$ represents how the property was paid for, such as $y_1=fully paid purchaser, y_2=60 per cent mortgage loan, y_3=monthly mortgage payments, y_4=30-year loan, y_5=5 per cent interest rate, \dots$ etc. $Z = \{z_1, z_2, z_3, \dots\}$ represents who buys (sells) the property, such as $z_1=fully owned, z_2=owned with mortgage, z_3=owner-occupier, z_4=family with children, z_5=dual income household, \dots$ etc. $Q = \{q_1, q_2, q_3, \dots\}$ represents why pay (ask) the particular price, such as $q_1=auction process, q_2=house auction price, q_3=auction reserve price, q_4=number of bids made, q_5=neighbourhood median house price, \dots$ etc. Figure 2 shows an example of a behaviour ontology diagram.

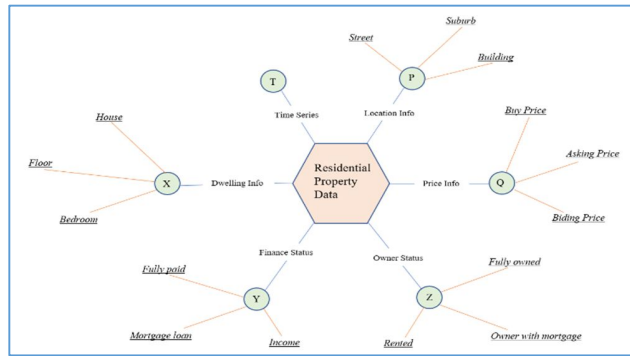


Figure 2 An example of residential property data in an ontology diagram

Therefore, the dataset D can be denoted using the 5W+1H dimension dataset as follows:

When, Where, What, Why, Who, How

$$D = \begin{pmatrix} d_{1T} & d_{1P} & d_{1X} & d_{1Q} & d_{1Z} & d_{1Y} \\ d_{2T} & d_{2P} & d_{2X} & d_{2Q} & d_{2Z} & d_{2Y} \\ d_{3T} & d_{3P} & d_{3X} & d_{3Q} & d_{3Z} & d_{3Y} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{iT} & d_{iP} & d_{iX} & d_{iQ} & d_{iZ} & d_{iY} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{nT} & d_{nP} & d_{nX} & d_{nQ} & d_{nZ} & d_{nY} \end{pmatrix} \tag{2}$$

The 5W+1H dimensions have significantly simplified the process of property dataset classification for all forms of property data. Most property data visualisations focus mainly on location dimension P and price dimension Q without considering attributes $d_{(x, y, z)}$ as they are constricted by the limited capabilities of low-dimension charts such as bar charts and line charts. The 5W+1H coordinates enable us not only to focus closely on relevant factors such as the property price and location, but also combine together behavioural attributes such as household finance situation, ownership status and dwelling information to maximise the effectiveness of the property data analysis and visualisation process.

This is achieved through using high-dimensional visual charts such as parallel coordinates, which are able to illustrate the multidimensional relationship between the 5W+1H axes and uncover visual patterns that would otherwise be hidden within complex property data (Zhang & Huang, 2016). In order to best use parallel coordinates to measure 5W+1H behaviour patterns, we will create various behaviour ratios to analyse the relationship between any two behavioural dimensions in the 5W+1H behaviour axes.

3.3 Analysis of Behaviour Ratio Sets between Dimensions

Analysing multidimensional data can be achieved by creating behaviour ratios. As an example, we will focus on the relationship between the Q dimension, which represents the property’s price characteristics, and the Y dimension, which represents the household’s finance characteristics. Our analysis will be framed within the constraints provided by the T dimension, which is a timestamp using time serials, and the P dimension, which represents the property’s location characteristics. To compare the relationship between a property’s price attributes $Q = \{q_1, q_2, q_3, \dots, q_i, \dots, q_n\}$ and the household’s financial attributes dimension $Y = \{y_1, y_2, y_3, \dots, y_j, \dots, y_m\}$, we will create a new ratio set that measures the relationship between the price attribute (Q) and the financial attributes (Y) within the specified time and location restrictions $d_{(t, p)}$. In other words, the ratio R aims to illustrate how the property’s price attributes are influenced by the owner’s financial attributes within the specified time and location. The new ratio set can be denoted as R :

$$R_{\{Q \text{ vs } Y, d(t, p)\}} = \begin{matrix} & \mathbf{q(1),} & \mathbf{q(2),} & \mathbf{q(3),} & \dots & \mathbf{q(i),} & \dots & \mathbf{q(n)} \\ \left. \begin{matrix} r_{(q_1, y_1)} & r_{(q_2, y_1)} & r_{(q_3, y_1)} & \vdots & r_{(q_i, y_1)} & \vdots & r_{(q_n, y_1)} \\ r_{(q_1, y_2)} & r_{(q_2, y_2)} & r_{(q_3, y_2)} & \vdots & r_{(q_i, y_2)} & \vdots & r_{(q_n, y_2)} \\ r_{(q_1, y_3)} & r_{(q_2, y_3)} & r_{(q_3, y_3)} & \vdots & r_{(q_i, y_3)} & \vdots & r_{(q_n, y_3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{(q_1, y_j)} & r_{(q_2, y_j)} & r_{(q_3, y_j)} & \vdots & r_{(q_i, y_j)} & \vdots & r_{(q_n, y_j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{(q_1, y_m)} & r_{(q_2, y_m)} & r_{(q_3, y_m)} & \vdots & r_{(q_i, y_m)} & \vdots & r_{(q_n, y_m)} \end{matrix} \right\} & \begin{matrix} \mathbf{Y(1)} \\ \mathbf{Y(2)} \\ \mathbf{Y(3)} \\ \vdots \\ \mathbf{Y(j)} \\ \vdots \\ \mathbf{Y(m)} \end{matrix} \end{matrix} \quad (3)$$

where

$$r_{(q(i), y(j))} = \frac{d\{q(i), d(t, p)\}}{d\{y(j), d(t, p)\}} \quad (4)$$

The ratio set $R_{\{Q \text{ vs } Y, d(t, p)\}}$ represents the ratio of Q dimension, $Q = \{q_1, q_2, q_3, \dots, q_n\}$, against Y dimension, $Y = \{y_1, y_2, y_3, \dots, y_m\}$, at condition $d(t, p)$ where $T = \{t_1, t_2, t_3, \dots\}$ and $P = \{p_1, p_2, p_3, \dots\}$. Because $i=1, 2, 3, \dots, n$ indicates the elements of the Q attribute, and $j=1, 2, 3, \dots, m$, which indicates the elements of the Y attribute, then the ratio set $R_{\{Q \text{ vs } Y, d(t, p)\}}$ contains $n \times m$ attributes, and $R_{\{Q \text{ vs } Y, d(t, p)\}} = R \{r(1), r(2), r(3), \dots, r(n \times m)\}$. This ratio set contains $n \times m$ combinations that uncover the behavioural patterns between the property’s price attributes and the household’s financial attributes through property data visual analytics.

Similarly, the ratio of Q dimension to Y dimension can reflect housing cost behaviours (e.g. price to household income) which assesses how easily a typical household can purchase a typical dwelling. Fox and Finlay (2012) advised that household net worth is closely linked to dwelling prices, whereby the higher the ratio, the higher the cost of housing.

This paper demonstrates a methodology for multidimensional data visualisation. Thus, we adopt the behaviour patterns of price and household finance as an empirical study to test the use of different visualisation techniques. There is a large variety of other behaviour patterns that we have excluded from this paper, such as the behavioural relationships between the Z axis and X axis as a visualisation of the relationship between the property’s ownership attributes and the property physical attributes.

4.0 EMPIRICAL STUDY

As indicated, this empirical study applies the developed visualisation methodology to analyse behaviour patterns between house prices and household finance. The sample dataset (Figure 3) was collected by Ge (2018) from various editions of the Australian Census, and contains data for six suburbs of Sydney—Epping, Leichhardt, Manly, Parramatta, Rhodes and Westmead—between 2001 and 2016. The dataset has eight attributes: suburb location, property prices, ethnic and population demographics, household finance statuses, physical attributes, suburban unemployment rate, ownership statuses and year of census data. It is difficult to identify patterns and trends by looking through the rows and columns on a spreadsheet. By using the above visualisation methodology, these eight categories can be classified into the 5W+1H, dimensions, which contain 32 elements with 768 data attributes (refer to Figure 3).

[D0]-Location	Parramatta	Parramatta	Parramatta	Parramatta	Westmead	Westmead	Westmead	Westmead	Epping	Epping	Epping	Epping	Rhodes	Rhodes	Rhodes	Rhodes	Manly	Manly	Manly	Manly	Leichhardt	Leichhardt	Leichhardt	Leichhardt
[D31]-Time	Y2001	Y2006	Y2011	Y2016	Y2001	Y2006	Y2011	Y2016	Y2001	Y2006	Y2011	Y2016	Y2001	Y2006	Y2011	Y2016	Y2001	Y2006	Y2011	Y2016	Y2001	Y2006	Y2011	Y2016
[D1]-TransactionHouseNO	59	38	62	43	25	40	37	48	147	129	149	211	6	5	11	27	42	37	43	86	98	127	53	251
[D2]-MedianHousePrice	\$370	\$460	\$542.50	\$1,012	\$383	\$499	\$615	\$1,176	\$520	\$705	\$935	\$1,651	\$408	\$600	\$1,030	\$2,532	\$914.40	\$1,300	\$1,885	\$2,582	\$451.50	\$627	\$820	\$1,378
[D3]-TransactionUnitNO	393	217	770	677	85	126	319	220	105	80	220	724	3	84	604	281	199	188	363	344	37	83	81	130
[D4]-MedianUnitPrice(000)	\$230.00	\$495.00	\$425.00	\$670.00	\$255.00	\$302.50	\$379.00	\$551.00	\$329.00	\$416.50	\$580.50	\$790.00	\$504.90	\$600.00	\$868.00	\$465.00	\$642.50	\$695.00	\$706.00	\$399.00	\$408.80	\$694.50	\$620.00	\$620.00
[D5]-AverageHouseholdSize	2.4	2.4	2.4	2.6	2.4	2.5	2.6	2.8	2.7	2.8	2.8	2.9	2.7	2.4	2.3	2.4	2	2.1	2.1	2.2	2.2	2.2	2.3	2.5
[D6]-MedianHouseholdWeeklyIncome	\$900	\$1,022	\$1,314	\$1,739	\$900	\$1,064	\$1,475	\$1,866	\$1,100	\$1,432	\$1,683	\$1,973	\$1,100	\$1,565	\$1,617	\$1,693	\$1,100	\$1,591	\$2,084	\$2,449	\$1,100	\$1,516	\$1,924	\$2,256
[D7]-MedianMortgageWeeklyPayment	\$253	\$362	\$449	\$433	\$253	\$379	\$460	\$462	\$345	\$449	\$526	\$510	\$345	\$548	\$598	\$508	\$391	\$529	\$690	\$665	\$345	\$506	\$648	\$692
[D8]-MedianWeeklyRent	\$225	\$230	\$350	\$430	\$225	\$230	\$360	\$430	\$550	\$300	\$420	\$530	\$275	\$375	\$560	\$620	\$325	\$380	\$530	\$650	\$325	\$340	\$460	\$560
[D9]-Population	18292	18447	19745	25798	10210	9486	14171	16309	18347	18969	20227	23688	743	1671	5679	11906	14922	13949	15072	15866	12608	12249	13520	14625
[D10]-BirthplaceInAustralia	6631	5991	5427	6263	4542	3364	4271	1068	10870	10174	10020	9778	500	778	1349	2235	7800	7207	7137	8213	7955	7803	8791	9395
[D11]-BirthplaceInOverseas	9157	10633	12529	17422	4512	5368	9030	10858	6567	7840	9595	13227	181	646	3711	9004	4497	4806	5818	6288	3744	3430	4002	4136
[D12]-BirthplaceUnknown	2504	1823	1789	2113	1156	754	870	3124	910	955	612	718	62	247	619	670	2625	1936	2117	1492	869	1016	1727	1135
[D13]-ParentsBirth-India	1376	2776	4260	7885	772	1745	4670	6751	564	909	1173	1822	5	29	296	675	74	85	91	90	118	103	173	235
[D14]-ParentsBirth-PRChina	2799	3526	4187	4766	870	992	1374	1384	3124	4327	5812	8834	46	205	2010	5848	237	228	256	301	331	348	489	634
[D15]-ParentsBirth-UK	2507	2141	2108	2242	1566	1145	1382	1193	5445	4305	4262	3835	213	290	633	933	5458	5049	5932	6463	3852	3487	4265	4817
[D16]-ParentsBirth-NewZealand	120	106	74	81	55	53	29	38	140	152	111	102	3	18	22	19	215	237	255	245	144	174	199	218
[D17]-ParentsBirth-Italy	279	253	224	246	190	152	163	152	551	445	505	476	50	71	123	169	316	393	451	663	1409	1374	1445	1508
[D18]-AustralianCitizenNo	11716	11111	10546	12026	7088	6190	8589	8943	15140	15237	16179	16815	620	1130	2514	4606	9708	8963	10055	10792	10467	10028	11275	11945
[D19]-MedianAge	32	30	30	31	33	33	31	33	36	37	38	36	39	32	28	29	35	35	35	36	34	35	36	36
[D20]-FullyOwned	1,378	976	870	958	698	511	606	545	3,328	2,724	2,819	2646	115	115	287	484	1,875	1,562	1,550	1,734	1,597	1,191	1,243	1271
[D21]-OwnedWithMortgage	930	1,444	1,685	1,556	372	688	1,138	1,107	1,235	1,826	2,084	2,345	56	192	627	1,302	740	978	1,297	1,247	1,308	1,657	1,907	1855
[D22]-Rented	4,018	4,361	4,389	6,101	1,978	2,007	3,018	3,385	1,784	2,044	2,639	69	270	1,210	2,810	3,020	2,797	3,285	3,323	2,191	2,046	2,165	2,165	2,165
[D23]-SeparateHouse-persons	4,148	3,491	3,023	2,855	2,699	2,722	2,958	2,837	12,999	13,092	13,414	14,105	504	475	538	516	2,224	2,586	2,777	2,800	6,217	5,176	5,712	5,126
[D24]-SemiDetached-persons	1,699	2,038	1,232	1,646	347	363	1,597	1,697	1,247	1,363	1,807	2,766	159	198	253	208	1,601	1,441	1,774	1,924	3,780	4,177	4,337	5,703
[D25]-FlatUnitApartment-persons	11,153	11,213	12,916	18,749	4,299	5,329	8,279	10,381	3,733	3,668	4,415	6,401	13	723	4,150	10,829	8,874	7,292	8,513	9,786	1,988	1,750	2,237	2,160
[D26]-SeparateHouse-dwellings	1444	1196	993	912	977	938	966	891	4300	4236	4447	4362	168	198	176	793	940	985	739	2526	2109	2224	1904	1904
[D27]-SemiDetached-dwellings	711	759	461	596	151	149	614	596	500	522	661	935	78	91	103	83	695	581	704	723	1732	1830	1880	2231
[D28]-FlatUnitApartment-dwellings	5109	5008	5645	7278	2155	2251	3319	3713	1868	1711	1980	2527	6	330	1838	4456	5010	4004	4597	5011	1178	1056	1293	1175
[D29]-Other-Dwellings	739	110	602	104	410	0	324	49	450	52	429	505	11	0	323	0	1249	52	1250	24	459	40	487	46
[D30]-Unemployment-%	9.5	8.1	8.2	9.6	8.8	8.3	7.5	9.3	4.6	4.3	6.1	7.4	4.9	6.3	7.8	9.4	4.6	3.3	3.7	3.9	4.4	3.2	4	4

Figure 3 Property sample dataset combined 6x suburbs between 2001 and 2016
(Source: ABS—Census data)

Epping is a suburb located 25 kilometres north-west of the Sydney Central Business District (CBD). The population in Epping increased from 18,347 in 2001 to 23,688 in 2016, while the median house price more than tripled from \$520,000 in 2001 to \$1,651,000 in 2016. The number of separate house dwellings remained at around 4330, while the number of semi-detached houses increased from 500 in 2001 to 935 in 2016, and the number of unit apartment dwellings increased from 1868 in 2001 to 2527 in 2016. The median household weekly income increased from \$1100 in 2001 to \$1973 in 2016, while the median mortgage weekly payment increased from \$350 in 2001 to \$510 in 2016. However, the median rental weekly payment remained stable at around \$540.

The suburb of Leichhardt, located five kilometres west of the Sydney CBD, experienced moderate population growth from 12,608 in 2001 to 14,625 in 2016, but the median house price more than tripled from \$451,500 in 2001 to \$1,378,000 in 2016. The number of separate house dwellings dropped from 2536 in 2001 to 1904 in 2016 as more houses were converted to semi-detached houses, which increased from 1830 in 2001 to 2231 in 2016. The number of unit apartment dwellings remained at the same level—around 1175. Leichhardt has the largest Italian community, with 1508 households in 2016. The median household weekly income increased from \$1100 in 2001 to \$2256 in 2016, while the median mortgage weekly payment increased from \$345 in 2001 to \$692 in 2016. The median rental weekly payment also increased from \$325 in 2001 to \$560 in 2016.

Manly, a beautiful beach suburb located 18 kilometres north of the Sydney CBD, experienced only slight population growth from 14,922 in 2001 to 15,866 in 2016. However, the median house price increased from \$914,400 in 2001 to \$2,582,000 in 2016. The number of separate homes, semi-detached houses and apartments remained stable—around 860, 700 and 5011 respectively. Manly has the largest community of a UK background, with more than 40 per cent of the population having parents of UK origin. Manly is a wealthy suburb, and the median household weekly income increased from \$1100 in 2001 to \$2449 in 2016. The median mortgage weekly payment increased from \$391 in 2001 to \$665 in 2016, offset by the median rental weekly payment also increasing from \$325 in 2001 to \$650 in 2016.

The suburb of Rhodes, located 16 kilometres west of the Sydney CBD, sits on a peninsula on the southern bank of the Parramatta River. The built environment of Rhodes has changed significantly in recent years, with an extraordinary number of new apartment complexes having been built in recent years. The population in Rhodes increased more than 15 times, from 743 in 2001 to 11,906 people in 2016, while the median house price increased from \$408,000 in 2001 to \$2,532,000 in 2016. Although the number of separate and semi-detached houses remained at around 176 and 83 respectively, the number of unit apartment dwellings dramatically increased, from six in 2001 to 4456 in 2016, an increase of more than 4450 new units. Because more young professionals are choosing to live in apartments, the median age dropped from 39 years in 2001 to 29 years in 2016, while the median household weekly income increased from \$1100 in 2001 to \$1693 in 2016. The median mortgage weekly payment increased from \$345 in 2001 to \$508 in 2016, and the median rental weekly payment increased from \$275 in 2001 to \$620 in 2016.

The suburb of Parramatta, located 23 kilometres west of the Sydney CBD, is Sydney's second-largest business centre after the Sydney CBD and a major government, business and commercial centre. The population increased from 18,292 in 2001 to 25,798 in 2016. As a result, the median house price increased from \$370,000 in 2001 to \$1,012,000 in 2016. The decline in the number of separate houses from 1444 in 2001 to 993 in 2016 and the fall in the number of semi-detached houses from 711 in 2001 to 596 in 2016 was offset by a significant increase in the number of high-density unit apartment buildings, which increased from 5109 in 2001 to 7278 in 2016. Parramatta had a very high unemployment rate of 9.6 per cent in 2016, although the household weekly income increased from \$900 in 2001 to \$1,739 in 2016. Median mortgage weekly payments increased from \$253 in 2001 to \$433 in 2016, although this was matched by an increase in median rental weekly payments from \$225 in 2001 to \$430 in 2016.

The suburb of Westmead, located 26 kilometres west of the Sydney CBD, contains the largest government hospital in Australia, which serves around 1.85 million people each year (Wikipedia 2019). The population increased moderately from 10,210 in 2001 to 16,309 in 2016. However, the median house price increased dramatically from \$383,000 in 2001 to \$1,176,000 in 2016. The number of separate house dwellings dropped from 977 in 2001 to 891 in 2016, while the number of semi-detached houses increased from 151 in 2001 to 596 in 2016, and the number of unit apartments increased from 2155 in 2001 to 3713 in 2016. The median household weekly income doubled from \$900 in 2001 to \$1866 in 2016. Similarly, the median mortgage weekly payment increased from \$253 in 2001 to \$462 in 2016, and the median rental weekly payment increased from \$225 in 2001 to \$430 in 2016.

Although descriptive, the above suburb profiles are not particularly informative as they contain a large series of information with no relationships or analysis presented. In addition, the dataset contains six different forms of data, including dollar, number and percentage formats, and even within the same format (e.g. dollar). The dataset also contains major differences in scale (e.g. house prices in \$ millions while weekly rent in \$ 000s). Therefore, these data attributes are difficult to analyse without the rescaling and classification of the data. As pointed out previously, data visualisation is an instrument for reasoning about quantitative information, and it allows us to analyse data behaviours by understanding data patterns, trends, correlations and causal analysis. The visual technique can address the drawbacks of the tabular and descriptive form data presented above and catch data behaviour that could not be detected by the traditional text-based data. The next section discusses the data visualisation using parallel coordinates.

4.1 Data Visualisation without Behaviour Classification

Unlike other visualisation charts that undertake a point-to-point mapping of only two-dimensional data, or standard three-dimensional coordinate systems, the use of parallel coordinates is a visualisation technique which can visualise multi-dimensional geometry data (McDonnell & Mueller, 2008) so that there is no loss of information. This implies that parallel coordinates can visualise many attributes in a two-dimensional chart. The parallel coordinates can also clearly show data relationships and patterns (Tory et al., 2005). In addition, parallel coordinates could be used for data statistics through order, scaling and rotation of the axes (Heinrich & Weiskopf, 2009). If the attributes of the sample dataset (Figure 3) do not apply to the developed visualisation methodology and are not grouped into similar behaviours, then the sample dataset contains 32 attributes, requiring 32 axes (refer to Figure 4).

The very left axis in Figure 4 indicates suburb data denoted (D0) (in this case, six suburbs). The four different colours represent each timestamp for 2001 (pink), 2006 (green), 2011 (orange) and 2016 (blue) located at the right axis, denoted (D31). Each axis records the original data values of 32 attributes. For example, on the population axis (D9), Parramatta in 2016 shows the highest value of 25,798, and Rhodes in 2001 depicts the lowest value of 743. By observing the colours, we can identify that many of the attributes follow the colour pattern from pink, green and orange to blue, which represents data from 2001 to 2016, indicating an increasing trend. Due to many lines and colours crossing, meaning that information overlaps itself, 24 points can be found on each of the axes in this case. It is self-evident that this makes it difficult to identify trends and compare housing market performance between suburbs.

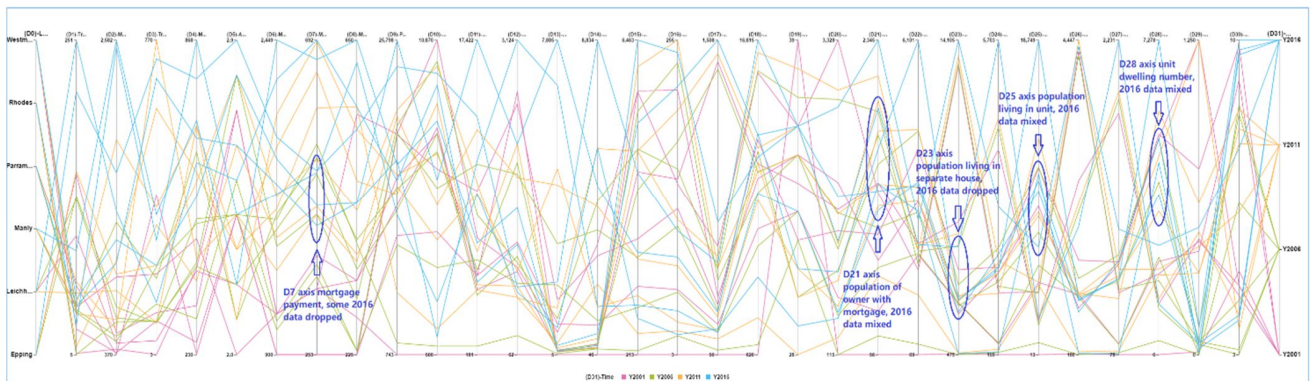


Figure 4 Parallel coordinate for sample dataset without data behaviour classification

This issue may be overcome by highlighting one suburb (refer to Figure 5), using Leichhardt as an example to study the visual results. In this way, only four points are identified on each of the axes. Among the 32 attributes, most of the points follow the pattern of pink, green, orange and blue from the bottom to the top, indicating increasing value over the periods. For example, Leichhardt had the highest population of people living in a semi-detached dwelling (D24) and had the highest supply of semi-detached dwellings (D27) in 2016. However, the median price of units (D4) dropped in value in 2011. The population of renters (D22) and the number of units (D28) remained the same over the last 15 years.

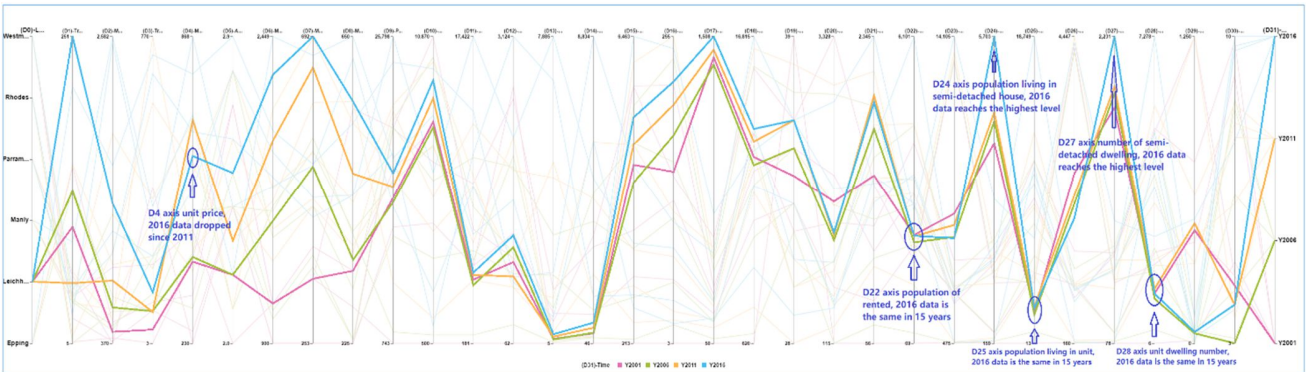


Figure 5 Parallel coordinate with Leichhardt highlighted

As demonstrated, parallel coordinates are able to include all attributes in one graph to show the performances of each attribute for each suburb across multiple years. The visual graph may improve the drawback from tablet or descriptive presentation by showing the characters and trends of each attribute. However, the relationship between the attributes or behavioural patterns has not been shown. In the next section, we adopt the relationship between property price status and household financial status as an example, demonstrating how the developed visualisation methodology is used to explore the relationships and patterns between those two property characteristic behaviours.

4.2 Behaviour Pattern between Property Price Status and Household Finance Status

In this part, we apply the visualisation methodology developed in Section 3 to demonstrate how visual analytics may be used to analyse behavioural patterns. We use the same dataset as shown in Figure 6 and categorised according to the 5W+1H model shown in Formula (2) to analyse the behavioural pattern between property price status and household financial status.

Next, we create the ratio of the behavioural pattern based on Formulas (3) and (4). For this demonstration, the behavioural pattern contains the dimensions T, P and behavioural ratio $R_{(Q \text{ vs } Y)}$. The Q dimension contains 2x elements (house price, unit price) while the Y dimension also contains 3x elements (median weekly income, median weekly mortgage payment, median weekly rent payment). As a result, six ratios (2x3) have been created to measure the behavioural patterns between different types of price status and financial status for condition $d_{(t,p)}$ where $T = \{t_1=2001, t_2=2006, t_3=2011 \text{ and } t_4=2016\}$ and $P = \{p_1=Epping, p_2=Leichhardt, p_3=Manly, p_4=Parramatta, p_5=Rhodes \text{ and } p_6=Westmead\}$. The set of behavioural ratios is shown in Table 2.

Table 2 Ratio set measuring the behaviour patterns for price status vs household finance status

$R_{(k)} = \frac{d\{q(t), d(t,p)\}}{d\{y(j), d(t,p)\}}$	$q_{(1)} = \text{House-Price}$	$q_{(2)} = \text{Unit-Price}$
$y_{(1)} = \text{Weekly-Income}$	$r_{(1)} = \frac{d\{q(1), d(t,p)\}}{d\{y(1), d(t,p)\}}$	$r_{(4)} = \frac{d\{q(2), d(t,p)\}}{d\{y(1), d(t,p)\}}$
$y_{(2)} = \text{Mortgage-Payment}$	$r_{(2)} = \frac{d\{q(1), d(t,p)\}}{d\{y(2), d(t,p)\}}$	$r_{(5)} = \frac{d\{q(2), d(t,p)\}}{d\{y(2), d(t,p)\}}$
$y_{(3)} = \text{Rent-Payment}$	$r_{(3)} = \frac{d\{q(1), d(t,p)\}}{d\{y(3), d(t,p)\}}$	$r_{(6)} = \frac{d\{q(2), d(t,p)\}}{d\{y(3), d(t,p)\}}$

4.2.1 The Ratios between House Price and Household Finance Status

In Table 2, ratios one, two and three measure the suburb’s behavioural patterns between the house price and the suburb’s median household weekly income, median weekly mortgage payment and median weekly rent payment at T time period. These ratios can be denoted as:

$$r(1, t, p) = \frac{\text{House Price } (\$)}{\text{Household Weekly Income } (\$)}$$

$$r(2, t, p) = \frac{\text{House Price } (\$)}{\text{Mortgage Weekly Payment } (\$)} \quad (5)$$

$$r(3, t, p) = \frac{\text{House Price } (\$)}{\text{Rent Weekly Payment } (\$)}$$

The value of these three ratios measures the impact that a dollar change in the suburb's median weekly income, median weekly mortgage payment or median weekly rent payment has on the house price in a suburb (p) at a particular time period (t). These ratios have the advantage of presenting the impact of changes in household financial status on the house price as a numerical, rescaled ratio that is easier to compare between suburbs and time series.

4.2.2 The Ratios between Unit Price and Household Finance Status

Similarly, in Table 2, ratios four, five and six measure the behavioural patterns between the unit price in a suburb and the suburb's median household weekly income, median weekly mortgage payment and median weekly rent payment. These ratios can be denoted as:

$$r(4, t, p) = \frac{\text{Unit Price } (\$)}{\text{Household Weekly Income } (\$)}$$

$$r(5, t, p) = \frac{\text{Unit Price } (\$)}{\text{Mortgage Weekly Payment } (\$)} \quad (6)$$

$$r(6, t, p) = \frac{\text{Unit Price } (\$)}{\text{Rent Weekly Payment } (\$)}$$

The value of these three ratios measures the impact that a dollar change in the suburb's median weekly income, median weekly mortgage payment or median weekly rent payment has on the unit price in a suburb (p) at a particular time (t). These ratios have the advantage of presenting the impact of changes in household financial status on the unit price as a numerical, rescaled ratio that is easier to compare between suburbs and time series. There are many behaviour patterns hidden inside this ratio set. For example, the higher the price to household income ratio, the higher the price or the lower the household income, the less affordable the housing. The higher price to household income ratio also indicates that the increase in price is faster than the increase in household income. We will visualise these behaviour patterns in the following sections.

4.3 Visual Ratios between Property Price Status and Household Finance Status

Based on Formulas (5) and (6), there are six different ratios that represent behaviour patterns between property price status and household finance status for T time series and different location P . The value of the ratio set is shown in Figure 6. The ratio set has combined six ratios, six suburbs and four time series. There are 144 ratios in this ratio set (refer to Figure 6) that contain many relationships and patterns to be uncovered. We have deployed parallel coordinates to demonstrate those behaviour ratio patterns, as shown in the next section.

Location	Time	r(1) HousePrice/Income	r(2) HousePrice/MortgagePayment	r(3) HousePrice/RentPayment	r(4) UnitPrice/Income	r(5) UnitPrice/MortgagePayment	r(6) UnitPrice/RentPayment
Parramatta	Y2001	7.91	28.12	31.62	4.91	17.48	19.66
Parramatta	Y2006	8.66	24.44	38.46	9.31	26.30	41.39
Parramatta	Y2011	7.94	23.24	29.81	6.22	18.20	23.35
Parramatta	Y2016	11.19	44.95	45.26	7.41	29.76	29.96
Westmead	Y2001	8.18	29.11	32.74	5.45	19.38	21.79
Westmead	Y2006	9.02	25.32	41.72	5.47	15.35	25.29
Westmead	Y2011	8.02	25.71	32.85	4.94	15.84	20.25
Westmead	Y2016	12.12	49.00	52.59	5.68	22.96	24.64
Epping	Y2001	9.09	28.99	18.18	5.75	18.34	11.50
Epping	Y2006	9.47	30.20	45.19	5.59	17.84	26.70
Epping	Y2011	10.68	34.18	42.81	6.63	21.22	26.58
Epping	Y2016	16.09	62.25	59.91	7.70	29.79	28.66
Rhodes	Y2001	7.13	22.74	28.53	7.09	22.60	28.36
Rhodes	Y2006	7.37	21.06	30.77	6.20	17.72	25.89
Rhodes	Y2011	12.25	33.12	35.37	7.14	19.30	20.60
Rhodes	Y2016	28.76	95.91	78.54	9.86	32.88	26.92
Manly	Y2001	15.99	44.97	54.11	8.13	22.87	27.51
Manly	Y2006	15.71	47.26	65.79	7.77	23.36	32.52
Manly	Y2011	17.39	52.54	68.40	6.41	19.37	25.22
Manly	Y2016	20.28	74.66	76.39	5.54	20.41	20.89
Leichhardt	Y2001	7.89	25.17	26.72	6.98	22.24	23.61
Leichhardt	Y2006	7.95	23.83	35.46	5.19	15.54	23.12
Leichhardt	Y2011	8.20	24.34	34.28	6.94	20.61	29.03
Leichhardt	Y2016	11.75	38.28	47.32	5.29	17.22	21.29

Figure 6 Behaviour ratios between property price status and household finance status

4.3.1 Visual Behaviour Patterns in Parallel Coordinate

Figure 7 is a graph of parallel coordinates we produced to analyse six derived ratios. The left-hand-side axis represents the location attributes, which contain six suburbs: Epping, Leichhardt, Manly, Parramatta, Rhodes and Westmead. The right-hand side axis represents the time series for 2001, 2006, 2011 and 2016, marked by colours. The first ratio is next to the location axis and the sixth ratio is next to the time axis. The other four axes represent the ratios between the second ratio and the fifth ratio.

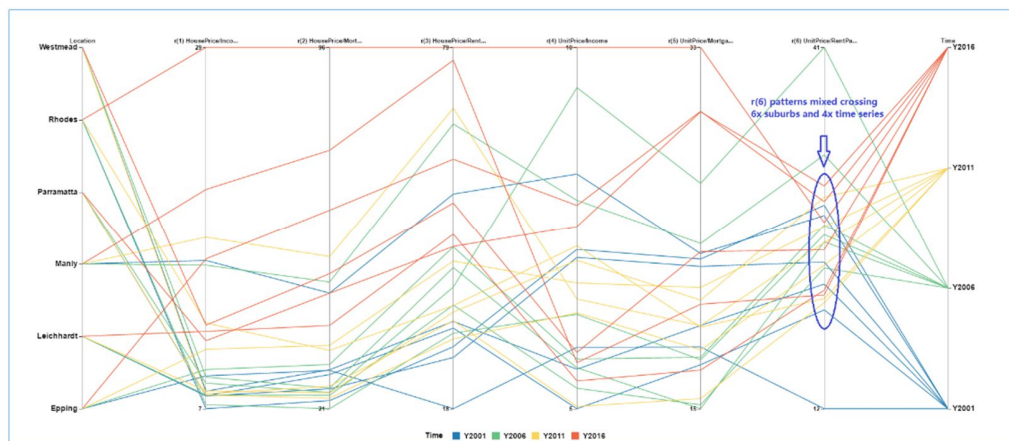


Figure 7 Parallel coordinate for the behaviour ratio pattern between property price status and household finance status between six suburbs and four time series

The first ratio $r(1, t, p)$ represents house price vs income. By looking at the pattern, we observe that the ratios appear to follow an increasing trend, particularly between 2011 and 2016, for all studied suburbs. The highest point is 28.76 for Rhodes in 2016, which is a 135 per cent increase from 2011 (value of 12.25) and a 300 per cent increase from 2001 (value of 7.13). Correspondingly, the population of migrants increased more than 140 per cent from 2011 to 2016, which may be the reason that house prices rose so dramatically. Additionally, these migrants tended to be younger, bringing the median age down from 39 in 2001 to 29 in 2016. The first ratio of Leichhardt also increased by 43 per cent, from 8.20 in 2011 to 11.75 in 2016. Given that the median households' income increased 4.7 per cent and 17.2 per cent from 2011 to 2016 for Rhodes and Leichhardt respectively, we may conclude that the increase in property prices was faster than the increase in household income for these two suburbs, meaning that these suburbs became less affordable in terms of owning property.

Empirical evidence (see Megbolugbe & Cho, 1993) has suggested a correlation between housing price and mortgage. The value of the second ratio $r(2, t, p)$ represents house price compared with mortgage payments, which is the amount of money used to pay the interest and principal of the house's loan. The higher the ratio, the less money used to pay interest and principal, the smaller the mortgage and the greater the homeowner's ability to pay off the loan. All studied suburbs exhibit an increasing trend from 2001 to 2016, with Rhodes the highest at 95.91 and Leichhardt the lowest at 38.28 in 2016. We observe that weekly mortgage payments in Rhodes increased from \$345 in 2001 to \$508 in 2016, while Leichhardt's increased even more from \$345 in 2001 to \$692 in 2016. Additionally, the dwelling price of Rhodes increased faster than Leichhardt, from \$408,000 and \$451,500 in 2001 respectively to \$2,532,000 and \$1,378,000 in 2016 respectively. Based on this analysis, we can deduce that households in Rhodes borrowed a smaller percentage of loan relative to the dwelling price than in Leichhardt.

The third ratio $r(3, t, p)$ represents house price to rent ratio, from which it may be deduced whether or not it is better to rent or own property in the suburb. Trulia, a US company, established thresholds for these house price/rent payment ratios, and suggested that for US households it is better to rent than own if the ratio is less than 15, and better to own than rent if the ratio is above 16. In Australia, the ratio was estimated at 109.2 in the fourth quarter of 2019 by Granwal (2020). Rhodes and Manly recorded the highest value of 78.45 and 76.39 respectively, with Parramatta and Leichhardt recording the lowest ratio of 45.26 and 47.32 in 2016. This means that relative to Rhodes and Manly, it was better to rent in Parramatta and Leichhardt than own, although in all suburbs it was better to own property than rent in the long term.

The fourth ratio $r(4, t, p)$ focuses on rental markets for units using the same ratio sequence as for houses discussed above. The unit price/household income ratios range from 5.29 in Leichhardt to 9.86 in Rhodes in 2016, although interestingly the ratios do not follow the same trend over time. The ratio Manly decreased from 8.13 in 2001 to 5.54 in 2016, implying that units in Manly were gradually more affordable to purchase. However, the ratios of Epping increased from 5.75 in 2001 to 7.70 in 2016, indicating that units were becoming more expensive relative to household income. The ratios of Rhodes decreased from 7.09 in 2001 to 6.20 in 2006, before increasing significantly to 9.86 in 2016, the highest amongst the studied suburbs. Parramatta, Westmead and Leichhardt displayed relatively small fluctuations that did not display a clear trend over time.

The fifth ratio $r(5, t, p)$ represents unit price to mortgage payment. The ratios reveal an increased trend from 2006 to 2016 for the suburbs of Westmead, Epping and Rhodes. For Parramatta and Manly, the ratios increased only from 2011 to 2016. However, in Leichhardt, the ratio decreased from 22.24 in 2001 to 15.54 in 2006, and increased again to 20.61 in 2011, before falling to the lowest value of 17.22 in 2016, implying that unit households in Leichhardt borrowed more to purchase units than other studied suburbs. The highest ratio is 32.88 in 2016 in Rhodes, suggesting households took out more equity to purchase units.

The sixth ratio $r(6, t, p)$ represents unit price to rent payments. The ratios were at their highest in 2006 and declined in 2016. The ratios of six suburbs in 2016 ranged from 29.96 in Parramatta to 21.29 in Leichhardt, implying that for all studied suburbs it was better to own units than to rent.

Aggregating these ratios can help us to understand the suburb profiles. For example, in the case of Leichhardt, it is worth noting that the unit price to rent ratio dropped 27 per cent from 29.03 in 2011 to 21.29 in 2016, which was the lowest among the studied suburbs, as indicated in Figure 8. To investigate the reasons, we found that the median unit price dropped by 11 per cent between 2011 and 2016 (refer to Figure 8), which may be due to decreased demand or increased supply. Further analysis shows that in 2001, there were 1988 people or around 904 households compared to 1178 units available. In 2016, however, the number of households reduced to 864 and the supply of units remained the same at 1175, suggesting that the excess of supply increased. Given that the weekly household income was \$2256 in Leichhardt, the highest among the studied suburbs in 2016, and that Leichhardt had the lowest ratio of unit price to household income ratio, units in Leichhardt may be considered affordable compared to other suburbs for the time period. However, mortgage payment of units was high at \$692 per week, indicating that households in Leichhardt used more debt to purchase units than in other suburbs.

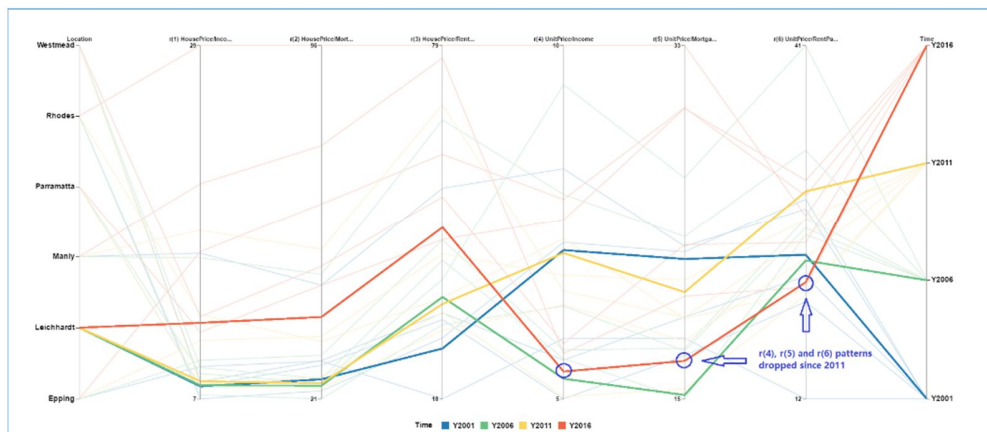


Figure 8 Parallel coordinates for the behaviour ratio pattern between property price status and household finance status for Leichhardt

By using the developed visualisation methodology, we have successfully visualised multidimensional property data using a parallel coordinate visual graph. Overall, parallel coordinates can illustrate the relationships and measurement of multidimensional property data that combines time series, location series and ratio series, which a normal graph cannot achieve. The successful presentation of multidimensional property data in a graph may be attributed mainly to the use of ratio analysis and an ontology-based classification of the dataset.

4.3.2 Visual Behaviour Patterns in Stacked Area Chart

Though a parallel coordinates visualisation has the advantage of enabling the user to visualise multi-dimensional property data with trends and relationships for different locations and time in one graph, it has a limitation in that it may be difficult to compare the attributes among the suburbs as many lines fall across each other. Thus, a stacked area visualisation chart is applied to display the multiple attributes on top of each other and present the evolution of the variables of several suburbs on the same graphic. According to Indratmo et al. (2018), the stacked area visualisation method could be used to compare and explore rankings of variables and identify the relative importance of each attribute to the totals.

When multiple attributes are included, the first attribute is plotted as a line with colour, followed by the second attribute, and so on. The stacked area charts shown in Figure 9 visualise the various behavioural patterns between property price status and household finance status for each of the six Sydney suburbs by colour. The area graph has four data points on the horizontal axis to show how these ratios change between 2001, 2006, 2011 and 2016. Importantly, the area graph stacks the six ratios on top of each other, such that the height of the graph shows the cumulative impact of household financial status on property price. The darkest colour, located on the top layer, represents the first ratio measuring the relationship between house price and the change in median household income. The lightest colour, located on the bottom layer, represents the sixth ratio measuring the relationship between unit price and the change in median rental payments.

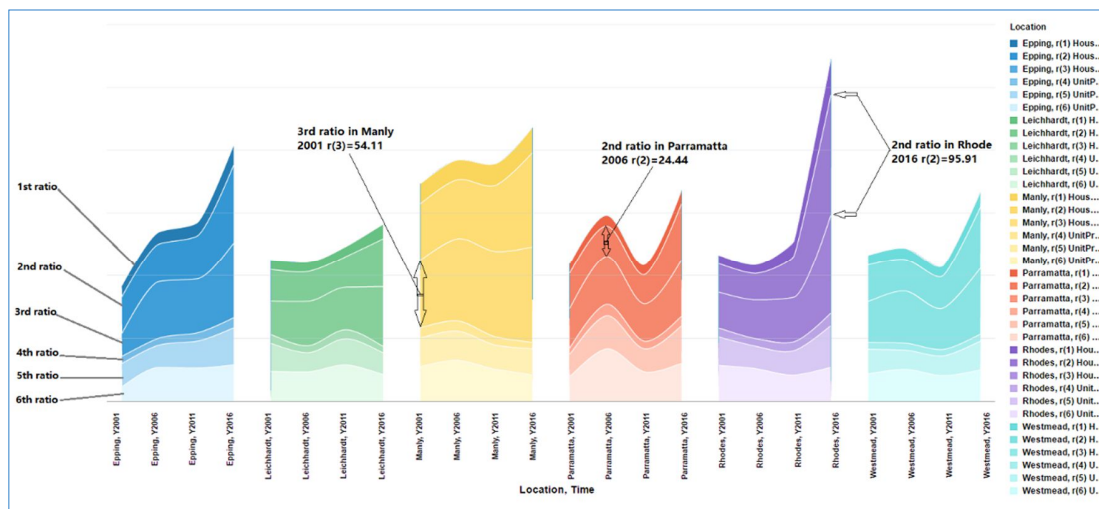


Figure 9 Stacked area chart for behaviour ratio pattern between property price status and household finance status between six suburbs and four time series

The height of each area represents the value of the corresponding ratio. For example, in the purple graph, which represents Rhodes, the second area from the top represents the second ratio, which is house price/mortgage payment $r(2, 2016, Rhodes) = 95.91$. We can see that this ratio increased dramatically from 2011. In the orange graph, which represents Parramatta, the second ratio $r(2, 2006, Parramatta) = 24.44$ indicates the ratio of Parramatta's house price/mortgage payment, which had similar values in both 2006 and 2011. In the yellow graph, which represents Manly, the third ratio $r(3, 2001, Manly) = 54.11$ indicates the ratio of house price to rent in 2001, which increased consistently over the next 15 years.

Epping (blue) had the smallest stacked value for 2001 compared to the other five suburbs, implying that household finance status had less effect on property prices. It shows also the lowest third ratio $r(3, 2001, Epping) = 18.18$ and the lowest sixth ratio $r(6, 2001, Epping) = 11.50$ across the six suburbs at any point in time. Furthermore, all six ratios consistently increased from 2001 to 2016. By 2016, the highest ratio area in Epping was the third ratio $r(3, T, Epping)$, which increased more than 149 per cent from 2006, indicating that the house price strongly increased. The lowest ratio area was the fourth ratio $r(4, T, Epping)$, which means Epping's unit price was more affordable than houses.

Leichhardt (green) had the smallest total stacked areas across the entire time series compared to the other five suburbs, which means that overall, Leichhardt's property prices were relatively stable during the studied period. In 2006, the fifth ratio $r(5, 2006, Leichhardt)$ decreased 30 per cent from 2001, which means that mortgage payments were higher for unit purchasers. In 2016, the fifth ratio $r(5, 2016, Leichhardt)$ dropped 17 per cent and the sixth ratio $r(6, 2016, Leichhardt)$ dropped 27 per cent, which indicates higher mortgage payments related to unit price and that owning units was better than renting. However, in the same period, the second ratio $r(2, 2016, Leichhardt)$ increased by 57 per cent and the third ratio $r(3, 2016, Leichhardt)$ increased by 38 per cent, implying houses were more expensive and renting was better than owning.

Manly (yellow) had the largest total stacked areas across the entire time series relative to the other five suburbs. The areas for the first ratio $r(1, T, Manly)$, second ratio $r(2, T, Manly)$ and third ratio $r(3, T, Manly)$ were the largest areas compared to the other five suburbs in the same period. That indicates that prices in Manly were high and houses were expensive to own. Between 2006 and 2016, the value of the sixth ratio dropped by 36 per cent, which suggests that owning a unit was better than renting for that year.

Parramatta (red) had different behaviour patterns compared to the other five suburbs. From 2001 to 2006, the stacked areas of the six ratios increased by more than 35 per cent. In the next five years from 2006 to 2011, the stacked areas dropped by 41 per cent, and then went up by 55 per cent. It had the lowest value of fourth ratio $r(4, 2001, Parramatta) = 4.91$ crossing all six suburbs in 2001, which indicates that unit prices were very affordable relative to household income in the suburb. In 2006, it had the highest value in the sixth ratio $r(6, 2006, Parramatta) = 41.39$, which means that Parramatta's unit prices increased strongly and they became more expensive to own.

Westmead (light blue) had similar patterns compared to Parramatta. The scale of dropping and rising ratios was much smaller. Between 2006 and 2011, Westmead's fifth ratio $r(5, 2006, Westmead) = 15.35$ and $r(5, 2011, Westmead) = 15.84$ remained at the lowest level compared to the other five suburbs. This means that mortgage payments were high relative to unit prices in Westmead during that period. From 2011 to 2016, Westmead's second ratio $r(2, 2016, Westmead) = 49.00$ increased by more than 90 per cent, and its third ratio $r(3, 2016, Westmead) = 52.59$ increased by more than 60 per cent. This led its sixth ratio's stacked areas to increase by 55 per cent. Houses in Westmead were therefore more expensive, and renting was more affordable than owning.

For Rhodes (purple), the highest stacked areas occurred for 2016, with the five highest ratios of all the suburbs, including its first ratio $r(1, 2016, Rhodes) = 28.76$, second ratio $r(2, 2016, Rhodes) = 95.91$, third ratio $r(3, 2016, Rhodes) = 78.54$, fourth ratio $r(4, 2016, Rhodes) = 9.86$ and fifth ratio $r(5, 2016, Rhodes) = 32.88$. This indicates that, in 2016, Rhodes' property prices increased strongly. In 2001, Rhodes had the lowest first ratio $r(1, 2001, Rhodes) = 7.13$, indicating that Rhodes's houses were very affordable. In 2006, Rhodes had the lowest second ratio $r(2, 2006, Rhodes) = 21.06$, which means that purchasers borrowed less relative to house prices in Rhodes.

Overall, a stacked area chart can display multiple dimensional property data, combining time series, location series and ratio series. In time series, an increased pattern could be found for most of the six ratios in Epping, implying that both house and unit prices increased consistently and became unaffordable over the years. Parramatta, on the other hand, showed up (2001 to 2006) and down (2006 to 2011),

then up (2011 to 2016) patterns for all ratios. The patterns suggest changes to household income, mortgage payments and rent payments, along with changes to house and unit prices.

In the location series, we found that the higher the stacks, the higher the ratios and the higher the prices, therefore, the lower the affordability. The highest total stacked ratios in Manly were caused by the highest stacked first ratio $r(1, T, Manly)$, second ratio $r(2, T, Manly)$ and third ratio $r(3, T, Manly)$. Rhodes had the second-highest total stacked ratios and Leichhardt had the lowest total stacked ratios, with the lowest second ratio $r(2, T, Leichhardt)$ and third ratio, $r(3, T, Leichhardt)$.

The demonstrations above have shown that, similar to parallel coordinates, the stacked area charts can also visualise multidimensional property behaviour patterns in a single graph. The main difference between the two graphs is that one uses lines and another uses stacked areas to demonstrate behavioural patterns. By using the parallel coordinates, the highest and lowest ratios could be found among the studied suburbs. When stacked area charts are used, the behaviour patterns of each suburb may be easily recognised.

5.0 CONCLUSION AND FUTURE WORKS

In this paper, we have developed a methodology to visualise and analyse property data behaviour and its patterns. The new methodology not only provides the ability to analyse property data and its behaviour ontologies, but also create new ratio sets that effectively analyse, process and interpret different series for property data visualisation and analysis. In contrast to Chang and Chen (2018), who displayed low-dimensional data, our new analytic model has significantly improved the visualisation for multiple high-dimensional property data by combining different data types and formats between property price and household financial attributes. The use of high-dimensional ratio sets can uncover behavioural patterns and relationships that normal low-dimensional analytical methods cannot.

We have applied both parallel coordinates to display overall visual patterns and stacked area charts to discover specific trends between three different series. Based on our developed visualisation methodology and the results of the analysis, we can conclude that when comparing many attributes and studying the patterns between them, multidimensional visual techniques such as parallel coordinates and stacked area charts can be used. The application of our developed methodology can reduce many individual attributes to six groups of datasets through data classification to make the visual representations more meaningful. Our visual approach could also be adopted by governments and property industry professions for training purposes. The visual representations could provide valid information to governments, enabling them to make informed policy decisions. In addition, the visual results are easily understood by laypersons, narrowing down the property market information asymmetries.

Nowadays, people's activities create quintillion bytes of data every day. Artificial intelligence (AI) machine learning techniques process data quickly and efficiently and become more widespread applications for learning and analysing the data. However, the results from the 'black box' approach are hard to be convinced as the data insights could not be explained. The developed ontology-based visual approach provides a scientific way for data analytics that study data and draw patterns. Integrating AI and data analytics using visualisation could not only improve data process efficiency, but also reveal deeper and better insights beyond what human analysts can do. Our future work will focus on causality analysis utilising our data analytics and visualisation tools on the prevalent issue of housing affordability and price forecasting to identify visual behavioural patterns and predict trends within the broader residential property market cycle. We also seek to use the visual dashboard technique to generate easy-to-understand property data visualisations.

References

- Adebara, O. B., Adebara, T. M., Badiora, A. I., & Ojo, D. B. (2020). Exploring spatial pattern of residential property value around cultural heritage sites in Ile-Ife, Nigeria. *International Journal of Real Estate Studies*, 14(1), 110-122.
- Bourassa, S. C., & Haurin, D. R. (2017). A dynamic housing affordability index. *International Real Estate Review*, 20(2), 251-286.
- Chang, C.-O., & Chen, S.-M. (2018). Dilemma of housing demand in Taiwan. *International Real Estate Review*, 21(3), 397-418.
- Chang, Z. (2017). Non-local students, housing demand and rental impact: Evidence from mainland students in Hong Kong. *International Real Estate Review*, 20(4), 525-548.
- Chen, C.-h., Härdle, W., & Unwin, A. (Eds.). (2007). *Handbook of data visualization*. Heidelberg: Springer.
- Cheung, S. K. C. (2017). A localized model for residential property valuation: Nearest neighbor with attribute differences. *International Real Estate Review*, 20(2), 221-250.
- Daana, M., & Huang, M. L. (2013, August). Visual sensitivity analysis in real estate prediction system. In E. Banissi & M. Sarfraz (Eds.), *CGiV 2013. Proceedings of the 10th International Conference Computer Graphics, Imaging and Visualization* (pp. 100-105). Los Alamitos, CA: IEEE.
- Dawidowicz, A., Radzewicz, A., & Renigier-Bilozor, M. (2014). Algorithm for purposes of determining real estate markets efficiency with help of land administration system. *Survey Review*, 46(336), 189-204.
- Deng, Y., Bao, H. X. H., & Gong, P. (2018). Increased tail dependence in global public real estate markets. *International Real Estate Review*, 21(2), 145-168.
- Fox, R., & Finlay, R. (2012). Dwelling prices and household income. RBA Bulletin, December Quarter, 13-22. Retrieved from <https://www.rba.gov.au/publications/bulletin/2012/dec/2.html>
- Gapminder Foundation. (2009, March 22). *200 years that changed the world (with Hans Rosling)* [Video]. Youtube. Retrieved on 3 April 2019 from <https://www.youtube.com/watch?v=BPt8EITQMlg&t=10s>.
- Ge, X. J. (2018). Effects of ethnic changes on house prices: Sydney cases. *International Journal of Housing Markets and Analysis*, 13(1), 96-119.
- Granwal, L. (2020). House price to rent ratio in Australia from 2nd quarter 2016 to 4th quarter 2019. Retrieved on 08 April 2020 from <https://www.statista.com/aboutus/our-research-commitment>
- Heinrich, J., & Weiskopf, D. (2009). Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1531-1538.
- Hong, T. (1999, July). Visualizing real estate property information on the web. In E. Banissi, F. Khosrowshahi, M. Sarfraz, E. Tatham & A. Ursyn (Eds.), *Proceedings of the IEEE International Conference on Information Visualization* (pp. 182-187). London: IEEE.
- Indratno, Howorko, L., Boedianto, J. M., & Daniel, B. (2018). The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. *Visual Informatics*, 2(3), 155-165.
- Li, M., Bao, Z., Sellis, T., & Yan, S. (2016). Visualization-aided exploration of the real estate data. In M. Cheema, W. Zhang & L. Chang (Ed.), *ADC 2016. Databases Theory and Applications [Lecture Notes in Computer Science]* (Vol. 9877, pp. 435-439). Cham: Springer.

- Li, M., Bao, Z., Sellis, T., Yan, S., & Zhang, R. (2018). HomeSeeker: A visual analytics system of real estate data. *Journal of Visual Languages & Computing*, 45, 1-16.
- McDonnell, K. T., & Mueller, K. (2008). Illustrative parallel coordinates. *Computer Graphics Forum*, 27(3), 1031-1038.
- Megbolugbe, I. F., & Cho, M. (1993). An empirical analysis of metropolitan housing and mortgage markets. *Journal of Housing Research*, 4(2), 191-224.
- Miles, W. (2016). Contagion, interdependence and diversification across regional UK housing markets. *International Real Estate Review*, 19(3), 327-351.
- Mottaleb, K. A., Sene, S. O., & Mishra, A. K. (2016). Impact of remittance income on house prices: Evidence from Bangladesh. *International Real Estate Review*, 19(1), 98-119.
- National Aeronautics and Space Administration (NASA). (2018). NASA visualization explorer. Retrieved on 6 June 2019 from <https://svs.gsfc.nasa.gov/nasaviz/index.html>.
- Sun, G.D., Liang, R.H., Wu, F.L., & Qu, H.M. (2013). A web-based visual analytics system for real estate data. *Science China Information Sciences*, 56(5), 1-13.
- Takatsuka, M., & Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, 28(10), 1131-1144.
- Tory, M., Potts, S., & Moller, T. (2005). A parallel coordinates style interface for exploratory volume visualization. *IEEE Transactions on Visualization and Computer Graphics*, 11(1), 71-80.
- Van Order, R. (2018). Mortgage securitization, structuring and moral hazard: Some evidence and some lessons from the Great Crash. *International Real Estate Review*, 21(4), 521-547.
- Wikipedia. (2019). Westmead hospital. Retrieved on 3 April 2019 from https://en.wikipedia.org/wiki/Westmead_Hospital
- Williamson, C., & Shneiderman, B. (1992, June). The dynamic HomeFinder: Evaluating dynamic queries in a real-estate information exploration system. In N. Belkin, P. Ingwersen & A. M. Pejtersen (Eds.), *SIGIR '92. Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 338-346). New York, NY: Association for Computing Machinery.
- Yang, T. T., & Zhang, J. Y. (2016). Contagious real estate cycle: Case of the US subprime related crisis. *International Real Estate Review*, 19(2), 171-196.
- Zhang, J., & Huang, M. L. (2016). Data behaviours model for Big Data visual analytics. *International Journal of Big Data Intelligence*, 3(1), 1-17.