A Visual Analytics Framework for the Housing Estate Data

Mingzhao Li, Zhifeng Bao, Shi Yan*
RMIT University, Australia

Timos Sellis†
Swinburne University of Technology, Australia

ABSTRACT

The housing estate data contains multidimensional attributes which are highly geographically related. In this poster, after collecting data from different channels and integrating a comprehensive housing estate dataset, we propose a visual analytic framework to assist users in sense-making the housing estate data. Specifically, we claim two contributions in visualization research. First, we propose a method to efficiently visualize the geographically related multidimensional data by seamlessly linking parallel coordinates with a geo-coded scatter plot and a coloured boolean table. Secondly, we propose a visual analytics design with multiple coordinated views to help users find preferred properties and further understand the housing estate market. Our design can also be used to solve other geographically related multidimensional visualization problems.

Index Terms: visual analytics, multidimensional visualization, geographic, composite visualization design, housing estate

1 INTRODUCTION

People spend a lot of time and effort when seeking an appropriate property to live or rent, since there are many factors to consider, such as the price, number of bedrooms, location and the transportation nearby, public schools issues, etc. Today, many online platforms, such as Real Estate Australia[1], have been providing the access to the housing estate properties. After users' searching or filtering among numerous properties available in the housing market, the result is then displayed in form of list or map. However, the visual design of those platforms is poor to support presentation and comparison of properties in multiple aspects; and a visualized tool is needed for users to filter and compare the properties.

We consider it as a geographically related multidimensional visualization problem [4] based on (i) the considerable amount of property attributes and numerous available properties, and (ii) that the regional, transportational and educational profiles of the housing estate data are all highly geographically related (details presented at Section 2). The design challenge for visualizing this kind of data is to provide space-efficient presentations that neither clutter visually nor confuse cognitively [1, 3].

It has attracted a lot of effort from researchers to address related problems. One of the most general solutions is to use maps to depict the geographical variation in data [5]. However, this is only effective when where there are few attributes: since position- and size-based visual variables are already used by maps, retinal variables to depict other attributes are limited [6]. Another popular solution is to use multiple coordinated views. Turkay et al. [6] proposed attribute signatures to dynamically generate summarises in another view after selecting elements on a map view. Guo et al. [2] designed TripVista with multiple views to analyse the traffic at a road intersection. However, it is not always easy for users to link the elements in different views, especially when the mapping of position in other views is quite different with that in map-based views. In this poster, we propose a composite visualization design [3] to visualize both multidimensional and geographical attributes in a view, by integrating parallel coordinates with a geo-coded scatter plot and a coloured boolean table in a seamless way.

Based on the above observation, after integrating all forms of housing estate data from different channels, we propose a visual analytics framework to visualize the housing estate data (Figure 1). We claim two contributions in visualization research.

1. we design a composite multidimensional visualization method that seamlessly links parallel coordinates with a geo-coded scatter plot and a coloured boolean table, which is able to efficiently visualize different kinds of multiple attributes in a view.
2. we propose a visual analytics design with multiple coordinates views and user-friendly interactions to visualize the housing estate data that assists users in finding preferred properties and further understanding the housing estate market.

2 DATASET DESCRIPTION

We have crawled the housing estate related data from different channels, and the dataset generated for use contains 52, 918 properties in Melbourne metropolitan area that were sold in latest three years (2013-2015). It includes the following four profiles.

Basic Profile: it contains the basic information of the properties, including 16 attributes crawled directly from Real Estate Australia[1] and another 34 features extracted from the text description which are about whether a property has a specific facility, such as air conditioning, gas heating, intercom, etc.

Regional Profile: it includes the information of a SA1[2] (Statistical Area Level 1, the smallest unit for the processing and release of Australian census data) mapped to each property and the census statistics based on each SA1.

Transportational Profile: it includes the walking time from each property to its nearest train station and the travel time between each pair of train stations. We are then able to compute efficiently about how long it takes from each property to a specific place (such as a user's working place) at runtime.

Educational Profile: it contains an exact public secondary school associated to each property and ranking of all the schools[3].

More detailed dataset description is available online [4].

3 VISUALIZATION DESIGN

Based on different characteristics, we first divide all attributes associated with each property into the following categories:

- **Geographical Attributes**, i.e., the geo-information;
- **Numeric Attributes**, such as price, number of bedrooms, etc.;
- **Categorical Attributes**, such as the property type;
- **Boolean Attributes**, i.e., the 34 features generated from the text description;
- **Other Types of Attributes**, such as text and images.

We then identify the following Design Maxims (DMs) to accomplish the visualization tasks that aims to help users find appropriate properties and further understand the housing estate market.

*e-mail: {mingzhao.li, zhifeng.bao}@rmit.edu.au, syanmust@gmail.com

†e-mail: tsellis@swin.edu.au

[1] http://www.realestate.com.au
[2] http://www.abs.gov.au/
[3] https://bettereducation.com.au/results
[4] www.dropbox.com/l/s/R3icuW827OEI8bSLySoyon

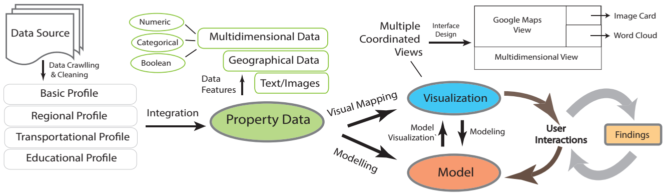
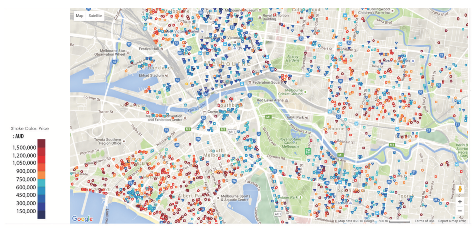


Figure 1: A framework for visualizing the housing estate data



(a)

(b)

Figure 2: A preliminary visualization result of the Google Maps view (a) and the multidimensional view (b)

- **DM1.** An overview of the housing estate data from different aspects.
- **DM2.** A user-friendly interaction design to allow users' filtering with the system responding timely.
- **DM3.** In-depth visualization for selected properties that helps users to compare the properties and understand the differences.

3.1 A Composite Multidimensional Visualization Design

We notice that all attributes except text and images can be converted into numeric attributes, which means that it is possible to visualize those attributes with axes-based visualization methods. Based on different characteristics of those attributes, we propose a composite visualization design using the strategy of overloaded views [3], to visualize the data by directly and sequentially linking a geo-coded scatter plot, Parallel Coordinates and a coloured boolean table in a seamless way (Figure 3).

Geo-coded Scatter Plot. We directly map each data item as a point in a 2D coordinate system based on its geographical attributes, i.e., the latitude and longitude.

Parallel Coordinates. Originally numeric and categorical attributes are displayed in the parallel coordinates, with categorical attributes mapped as segment in the axes instead of data points.

Coloured Boolean Table. It is a variation of Heatmap to visualize the boolean attributes. The attribute of a data item that has a value of true will be filled with the colour that is the same as that in parallel coordinates and the geo-code scatter plot.

We draw curved lines to directly link the geo-coded scatter plot with the leftest axis in parallel coordinates. With regard to connecting parallel coordinates with the boolean table, since boolean table expresses much less data items than parallel coordinates, we place the ranking of properties (based on user filtering) as the last axis in parallel coordinates, and then present the top several data items in the last axis and connect them directly with the boolean table.

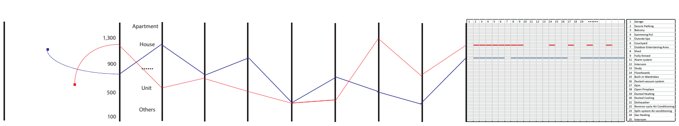


Figure 3: Illustration of the multidimensional view

3.2 Visual Analytics Design

Based on the data characteristics and our design maxims, we propose a visual analytics design with multiple coordinated views (Figure 1), including the Google Maps view, the multidimensional view (the design proposed in 3.1), the Word Cloud view and the Image Card view.

3.2.1 An overview of the housing estate data (DM1)

We give an overview of the housing estate data in different views. First, the Google Maps view provides users a vision of how the properties distribute in geographic and how the prices are different. We draw another layer on the top of Google Maps to present some key information of the properties (Figure 2(a)). Each property will be mapped as a regular polygon; and the shape and color of the polygon are determined by bedroom number and price of the property respectively (the determination factors are specified by users). Secondly, in the multidimensional view, users can get a general idea of what kind of attributes of properties we can provide and how the approximate distribution of each attribute is. Thirdly, the Word Cloud view and the Image Card view provides a rough description of what kinds of words and pictures to describe the housing estate data respectively.

3.2.2 Details via interactions (DM2)

We design several types of interactions in each view. As all the views are linked by default, any interactions in one view will affect the other views directly. For example, users can zoom in/out or select a suburb in the Google Maps view, those properties will be instantly updated in other views; users can also select an individual property and see how it is different with others in different dimensions in the multidimensional view. If the user is interested in the property, he/she could add this temporary selections into a **set**, which is defined as a collection of users' interests. The properties in a set can be individually visualized in depth (DM3). For another instance, users can also filter in parallel coordinates and see how the remainders distribute in the Google Maps view, and what the description of those properties is in the Word Cloud view. A preliminary visualization result is shown in Figure 2.

3.2.3 In-depth comparison of selected properties (DM3)

When there are only a few properties left after users' filtering and selections, or if users would like to visualize the properties saved in sets, we provide users an option to switch to the comparison mode. In this mode, the dimensions with larger variance will be highlighted, and we give more clear illustration of the comparison by changing the two ends of each coordinate axis to the maximum and minimum of each dimension among the remainders instead of the whole dataset.

References

[1] O. Dykes, C. Rooney, R. Beecham, J. Dykes, A. Slingsby, J. Wood, and W. Wong. Multi-Perspective Synopsis with Faceted Views of Varying Emphasis. *IEEE InfoVis Poster*, 2015.

[2] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *IEEE PacificVis*, 2011.

[3] W. Javed and N. Elmqvist. Exploring the design space of composite visualization. In *IEEE PacificVis*, 2012.

[4] M. Kreuseler. Visualization of geographically related multidimensional data in virtual 3D scenes. *Computers & Geosciences*, 26:101–108, 2000.

[5] A. D. Singleton and P. A. Longley. Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, 29(3):289–298, 2009.

[6] C. Turkay, A. Slingsby, H. Hauser, J. Wood, and J. Dykes. Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data. In *IEEE InfoVis*, 2014.