An interactive information visualisation of house price with regional information with user cantered design methodology

Submitted September 2017, in partial fulfilment of
the conditions of the award of the degree **MSc Human Computer Interaction** .

Zheran Xu
4296875
School of Computer Science
University of Nottingham

I hereby declare that this dissertation is all my own work, except as indicated in the text:

Signature ______________________

Date _____/_____/_____

I hereby declare that I have all necessary rights and consents to publicly distribute this dissertation via the University of Nottingham's e-dissertation archive.*

Public access to this dissertation is restricted until: Date ___/___/___**

Contents

Abstract

This dissertation mainly designed and implemented an interactive information visualization system that combines regional information with house price information, with a user-centred development methodology. The user test is adopted to evaluate the system with different visual encodings. The prototype provides a potential visualisation method for the existing price information search platform especially for map search.

Acknowledgement

I would like to express my gratitude to all those who have helped me during the period of writing my dissertation.

My deepest gratitude goes first and foremost for my supervisor, Ke Zhou, for his guidance and help. He is the lecturer of my Information Visualisation Module, where I build a solid theoretical foundation. During this time, I have a strong interest in this aspect. In the process of completing the dissertation, he provided me with the theme for me to choose, provide me with advice on the design plan, help me sort out the research methods, modify my test plan and. Without his help, the dissertation is impossible for me.

Secondly, I would like to express my heartfelt gratitude to all the Participants, especially my classmates. They supported my research and gave me design direction, provided advice during the process, and participated in the user research. Without your support, my so-called User cantered would become empty talk.

Thirdly, I would like to thank my dear parents and family for their financial and psychological support over the years as my strong backing. I am very grateful for what they have done for me.

1. INTRODUCTION

With the development and popularization of Internet technology, more and more users choose to find the house information online. Various platforms have emerged in around the world. With the support of big data and multiple information source, users are provided with a powerful function of searching for ideal house. They choose make right decisions on the platform according to their own needs and the available info [1]. Driven by the advancement of web technology and the growth of user demand, higher demand has been placed on the development of the housing price information platform. One is the ability to process large-scale data, including the accuracy and speed of retrieval. The other is the need to present users with the information they need in an easier way to understand. On the other hand, the decision-making factors that affect users are not only the price information, but also the information about the area where the house is located. This information is relatively scarce on existing platforms. With the increasing complexity of the information structure, therefore, presenting a large amount of information in an understandable and effect method has become a new demand.

Based on the text encoding, information visualisation is consider as a good method to present information. Its high performance allows users to quickly perceive, and its aesthetics could also improve the user experience. However, there is no clear path that how to correctly and efficiently handle multiple types of information such as house prices in a visual manner.

In this project, the aim is to design and develop an information visualisation system, presenting house price information and regional information to real users at the same time and researching the visualisation method, and finally evaluating the system. As an outcome, the system can be potentially used as part of the existing online platform, and participate in the user's actual use case.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Information visualisation

Information Visualisation is a technology that uses computer-based representations to express abstract information in visual form and to reveal the relationships and features between abstract information. Visualisation technology originated from the visualisation of scientific computing that emerged in the 1980s [2]. The term information visualisation first appeared in the 1989 article "Cognitive Coprocessor for Interactive User Interface" by Robertson [3]. Information visualisation is the application of visualisation technology in the field of non-spatial data. It is the process of transforming data information into visual form. It can enhance the data rendering effect, allowing users to observe and browse data in an intuitive and interactive way, thus discovering characteristics, relationships and patterns in the hidden data. The graphical form of information visualisation first appeared in the 18th century. Historical and political scientist Playfair and mathematician Lamber created visualisation charts for the first time. They thought that turning complex data into charts could help people understand the data. In the 19th century, French scientists Minard and Marey first used non-pure manual methods to make charts [4]. In the 20th century, advances in computer technology expanded the capabilities of data processing and could provide a variety of interactive methods, making it easier for users to observe the data they are interested. Now Visual applications are also more extensive. The main areas involve data mining visualisation and networking, data visualisation, social visualisation, traffic visualisation, text visualisation, biomedical visualisation, and so on.

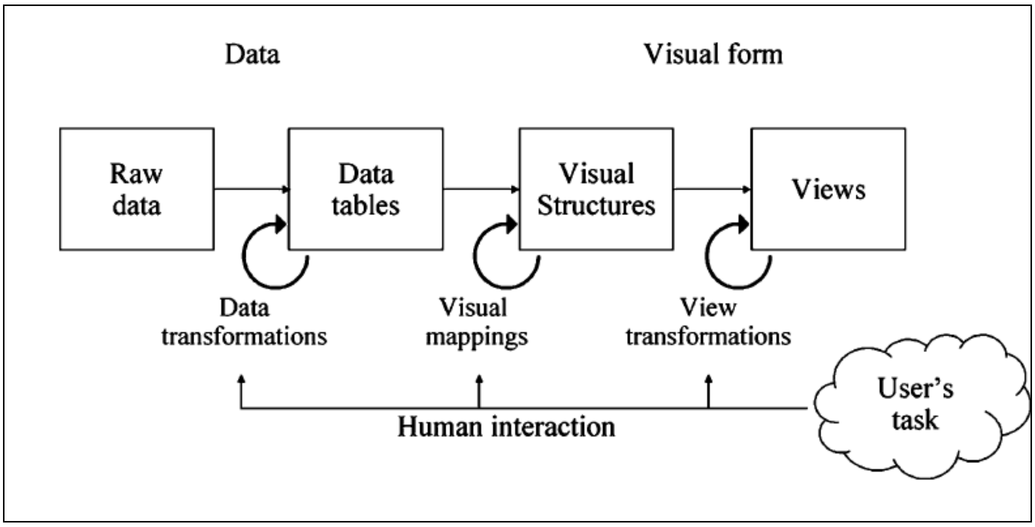


Figure 2.1 Information Visualisation Reference Model (from Card et al., 1999).

The Card visualisation model [5] can divide the process of information visualisation into the following stages: data pre-processing, drawing, display and interaction. The data pre-processing processes the collected information to make it easy to be understood and easy to be input into the display visualisation module. The pre-processing content includes data format, data standardization, data transformation technology, data compression and decompression. Some data also need to be processed such as outlier detection, clustering, and dimensionality reduction. The function of drawing is to complete the conversion of data to geometric images. A complete graphical description needs to be integrated with various visual rendering techniques based on user needs. The function of display and interactive display is to output the image data generated by the drawing module according to the requirements specified by the user. In addition to completing the image information output function, the user's feedback information needs to be transmitted to the software layer to realize human-computer interaction. For the main tasks of visualisation, namely overview, zoom, filter, details-on-demand, relate, etc., interactive technology mainly includes dynamic filtering, global + detailed, translation +Scale, focus + context and deformation, multi-view association coordination, and etc. [5,6]

According to the classification of Schneiderman, information visualisation data is divided into the following categories, one-dimensional data, two-dimensional data, three-dimensional data, multi-dimensional data, temporal data, hierarchical data, and network data [8]. Multidimensional data visualisation methods mainly include geometry-based methods, icon methods, and animation methods. The most classic of the geometry-based visualisation methods is the parallel coordinate method.

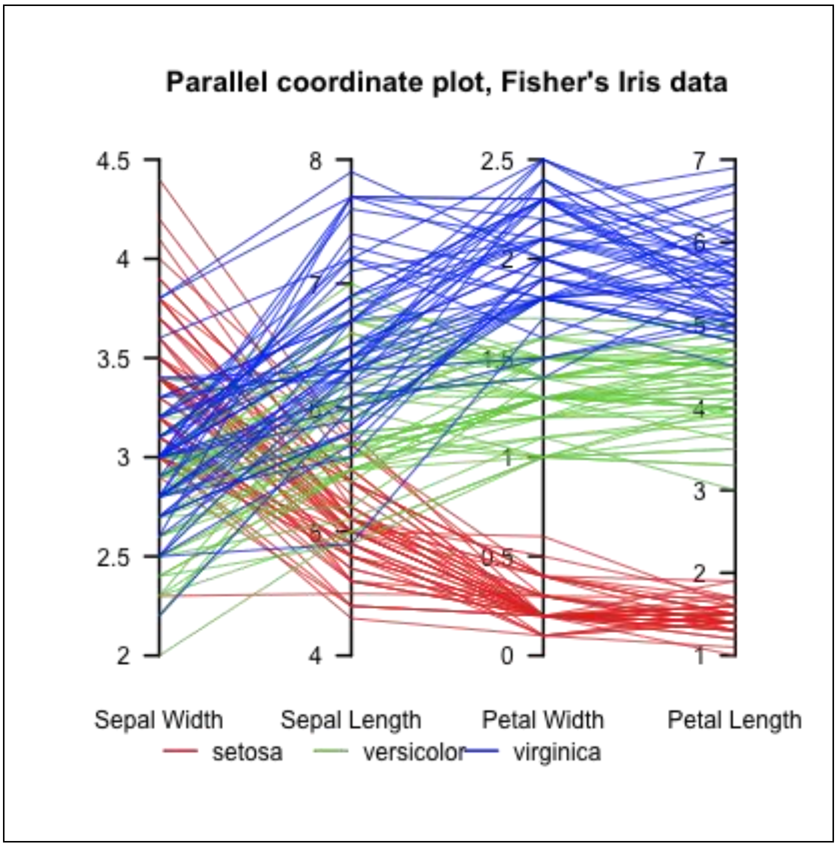


Figure 2.2 Parallel coordinates.

Parallel coordinates [8] use parallel vertical axes to represent dimensions, by plotting the values of the multidimensional data on the axis and connecting the data points on the axes with polylines to display the multidimensional data. The parallel coordinate system method can display multidimensional data in a simple and fast manner and develop many improved techniques. But when the size of the data set becomes very large, dense polylines can cause "visual clutter", which includes dimensional rearrangement, interaction methods, clustering, filtering, animation, and so on. Other geometric-based methods include the Radviz method [9] using a circular coordinate system to present visualisation results.

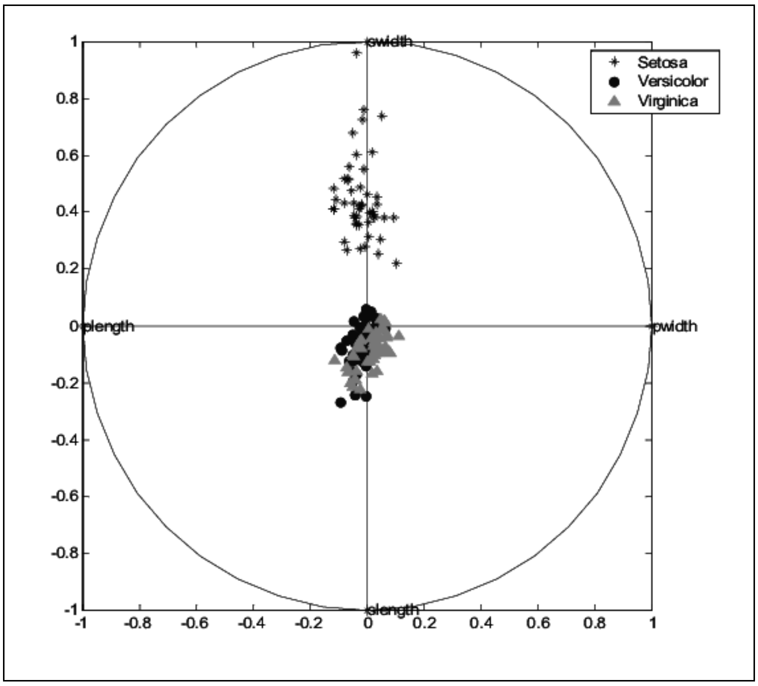


Figure 2.3 Parallel coordinates.

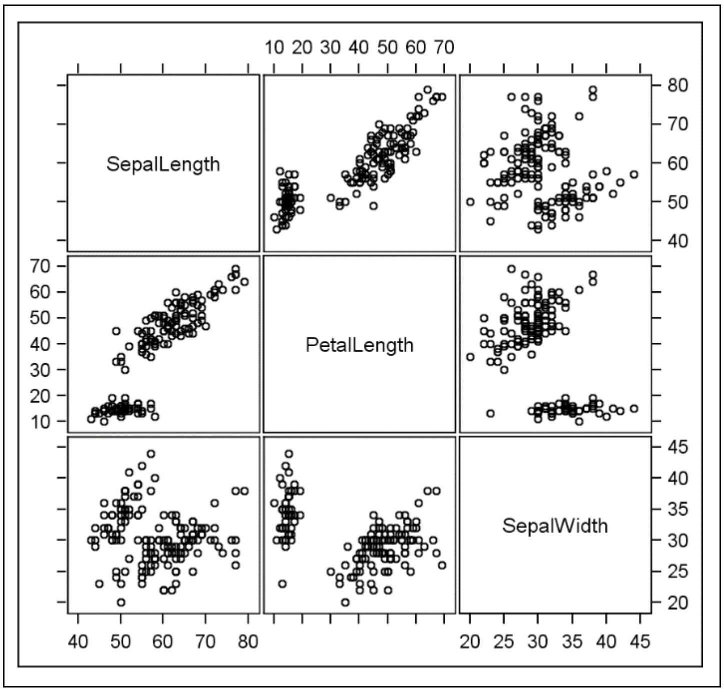


Figure 2.4 Scatter plot matrix.

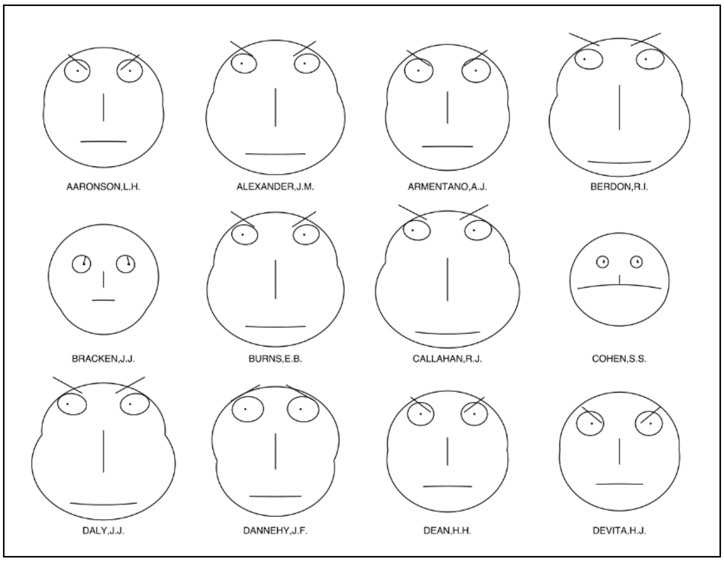


Figure 2.5 Chernoff face.

The scatter plot matrix [10] plots the dimensions of the multidimensional data in pairs to form a series of regular arrangements. The icon-based visualisation method scribbles data uses geometric features such as size, length, shape, colour, etc., with representative features, including star mapping and Chernoff face methods [11]. Animation methods can be used to improve interactivity and comprehension in visualisation. The disadvantages include distracting attention, causing user misunderstandings, and generating chart garbage. [12]

Time series data refers to a data set with time attributes. The visualisation methods for time series data are as follows: line graph, stack graph, animation, horizon graph [13], timeline.

Hierarchical data has a hierarchical or hierarchical relationship. The visualisation method of hierarchical data mainly includes two methods: node link graph and tree graph. The treemap is a series of nested rings and blocks to display hierarchical data. In order to display more node content, some interactive methods based on focus + context technology were developed, including fisheye technology, geometric deformation, semantic scaling, node clustering technology away from focus method and etc. [14].Network data has a mesh structure. The automatic layout algorithm is the core of network data visualisation. At present, there are three main types: one is force-directed layout; the other is hierarchical layout; the third is grid layout. . When there are many connections of data nodes, edge crossing is easy to occur, resulting in visual confusion. The edge bundle technology that solves the edge-crossing phenomenon can be divided into the following categories: force-oriented bundle edge technology, hierarchical bundle edge technique, geometry-based edge clustering technique, multi-layer agglomerative bundle edge technique, and grid-based Method et al. [15].

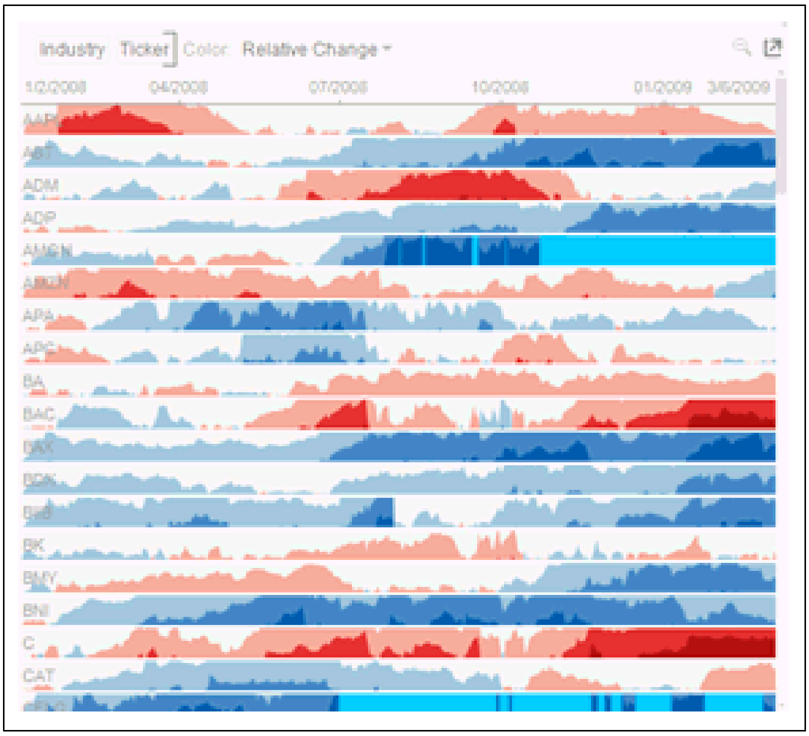


Figure 2.6 Horizon gragh.

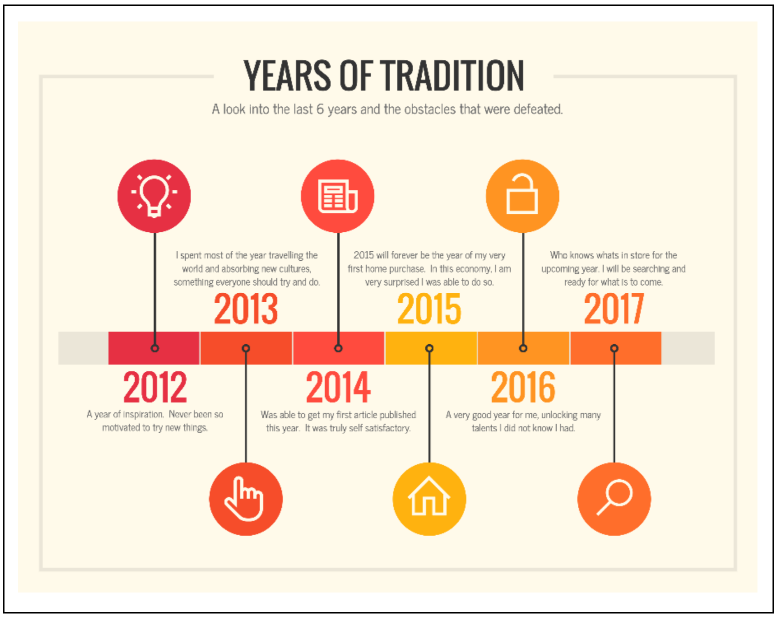


Figure 2.7 Time line

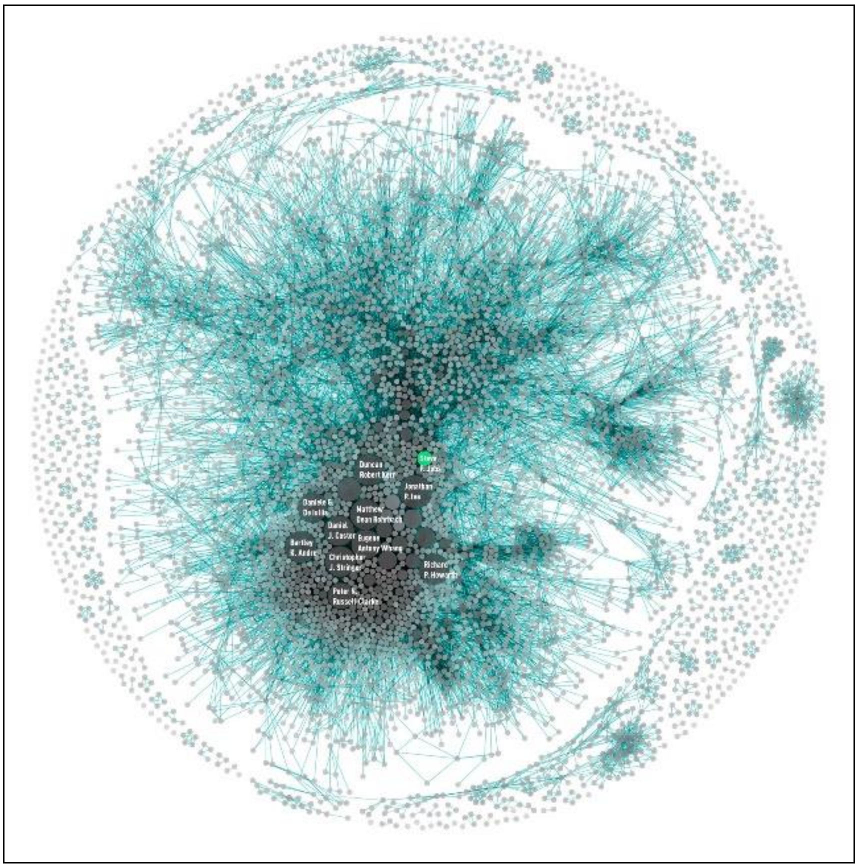


Figure 2.8 Node link graph.

.

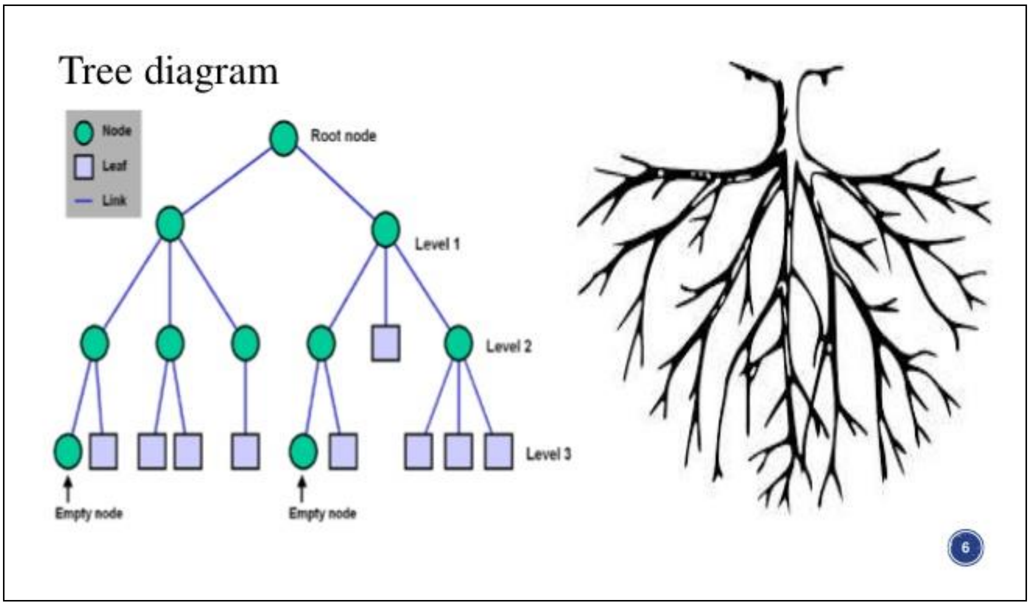


Figure 2.8 Tree graph.

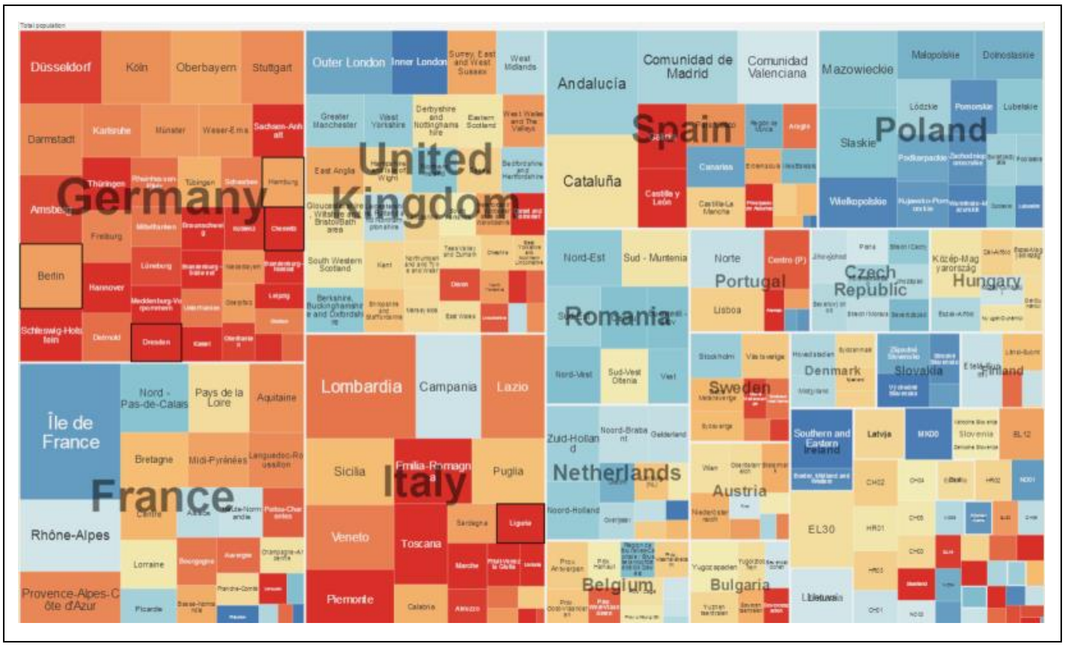


Figure 2.9 Tree map.

2.2 Geospatial visualisation

In information visualisation of geospatial data, points, lines, surface, and stereoscope are classic visual elements, which are widely used in the process of visual analysis of geospatial data.

Spatial geographical position is the basic attributes, which refer to entity' specific location. Points or dots are usually designed to these attribute, and sometimes by position distribution.

Jerry et al. [16] designed the Georgia initiative for community housing data visual analysis system, which uses dots to describe the geospatial location of the community housing, colours and locations on map represent the information such as the community housing condition. As shown in image, users can intuitively analyse and discover information such as community location and house condition.

Marking dots on the map can effectively display geospatial location information, but as the amount of data increases, the dots are prone to mutual coverage and other problems, which hinders the effective perception of data. Walker et al. [17] proposed a human terrain for data exploration. (The US military refers to the cultural intelligence that serves tactical operations as human terrain). The visual analysis method uses scatter plots, density maps, and dot plots to map the values of different attributes of the human terrain to maps, by analysing these points. The distribution of attribute values and their relationships to reveal the source and trend of intelligence to help military experts make appropriate strategic decisions.

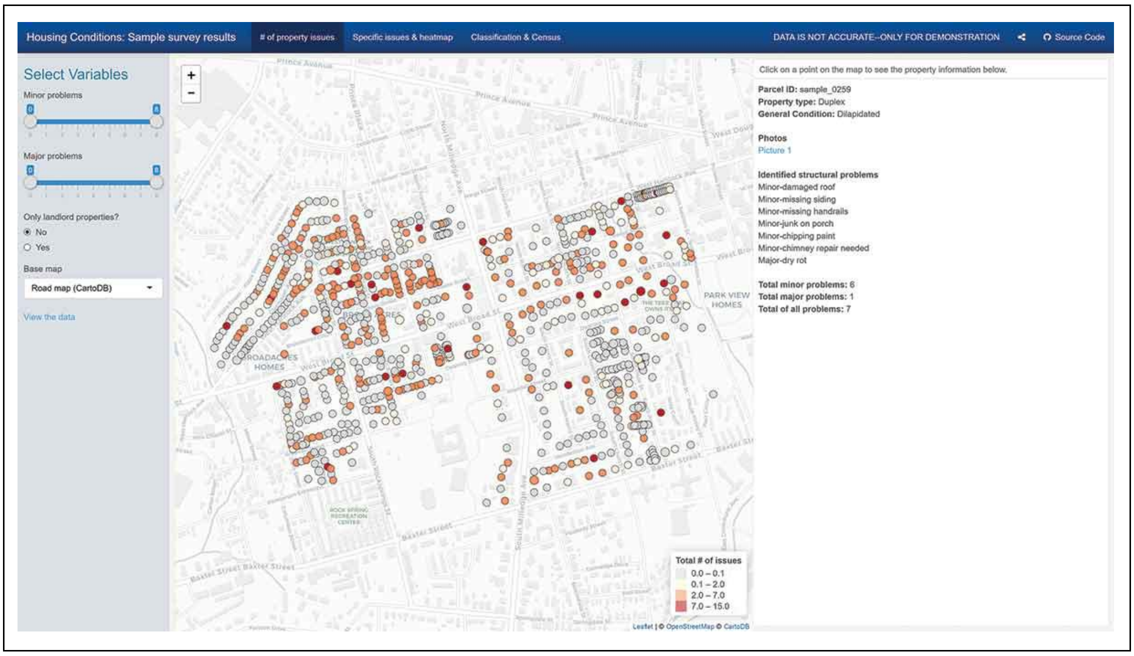


Figure 2.10 Shiny application for viewing collected data on housing conditions.

Marking dots on the map can effectively display geospatial location information, but as the amount of data increases, the dots are prone to mutual coverage and other problems, which hinders the effective perception of data. Walker et al. [17] proposed a human terrain for data exploration. (The US military refers to the cultural intelligence that serves tactical operations as human terrain). The visual analysis method uses scatter plots, density maps, and dot plots to map the values of different attributes of the human terrain to maps, by analysing these points. The distribution of attribute values and their relationships to reveal the source and trend of intelligence to help military experts make appropriate strategic decisions.

Figure 2.11 human terrain visual analysis method.

Line is a visualisation method that connects points. It usually represents the relationship between two or more visual elements. Many valuable information in the visualisation method can be described by means of line visualisation, for example, path, flow, trend, etc. Rae[18] For large-scale census datasets, use the line segments with arrows to indicate the flow direction and migration distance of the population, so as to visually analyse the geospatial characteristics of population migration. As the amount of data increases, the visual distribution of the lines is complicated and will interfere with people. Eye visual perception hinders the identification of geospatial features. Therefore, a large amount of research work optimizes line visualisation. for example, line clustering and line bundling.

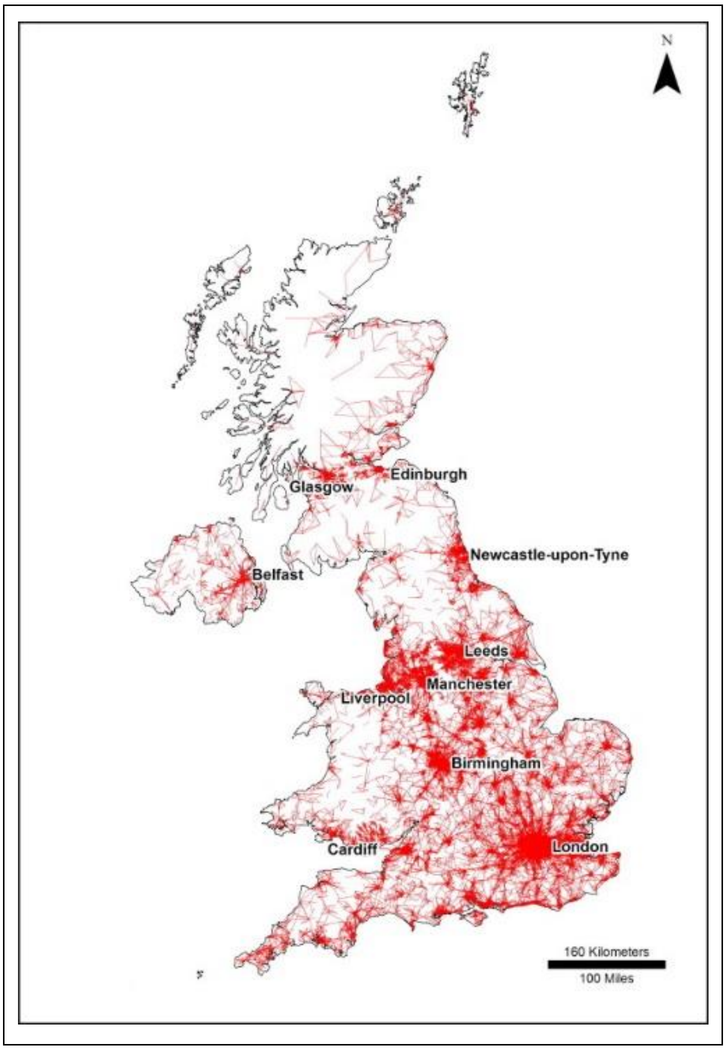


Figure 2.12 Flows of 50 km or less and 12 or more person.

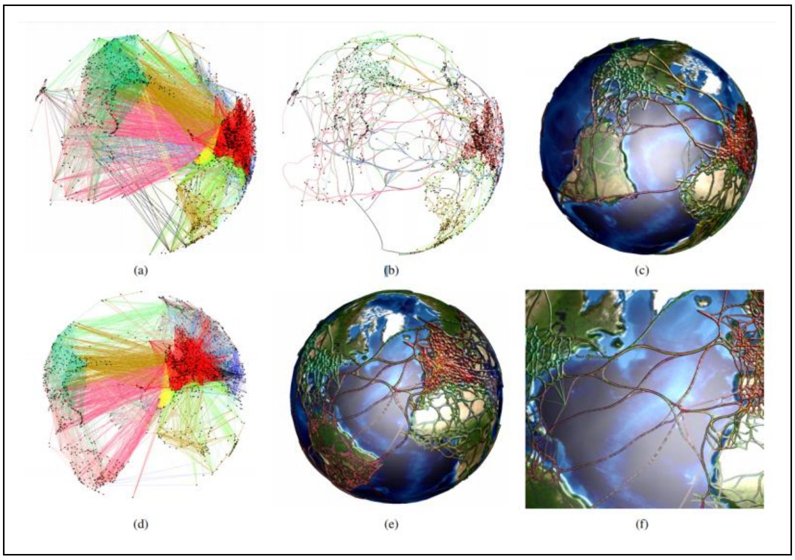


Figure 2.13 2000 international air interconnections network, containing 1524 airports and 16397 flights, embedded on the globe.

Andrienko et al. [19] designed a progressive clustering analysis method for large-scale ship transportation trajectory. Based on the similarity of ship destinations, the original trajectory was clustered and displayed using the OPTICS algorithm.

Lambert et al. [20] carried out 3D line bundling of the initial tens of thousands of aircraft routes for the global air traffic network and further optimized the display on the earth background to help users understand the visual complexity of the network and reduce the occlusion caused by line confusion. Line bundling technology can effectively avoid the problem of line visual layout confusion. Although it can reflect the overall trend of data to a certain extent, it has with certain limitations in describing the relationship between data objects in detail.

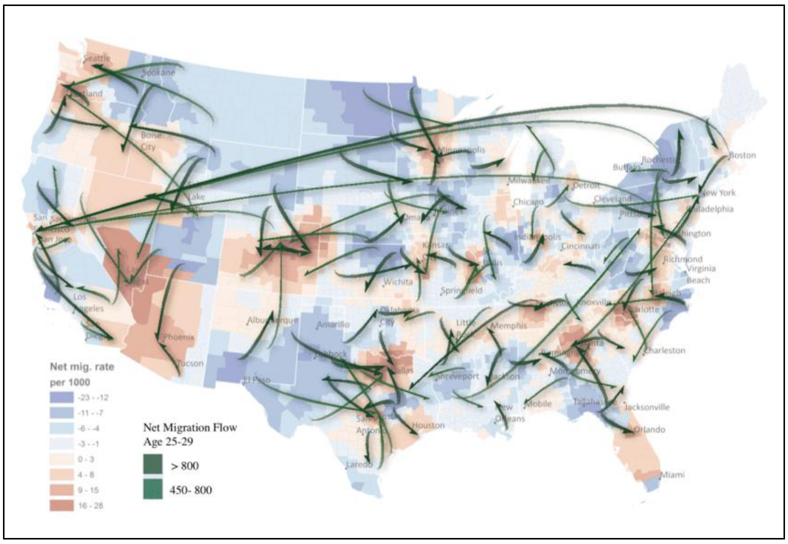
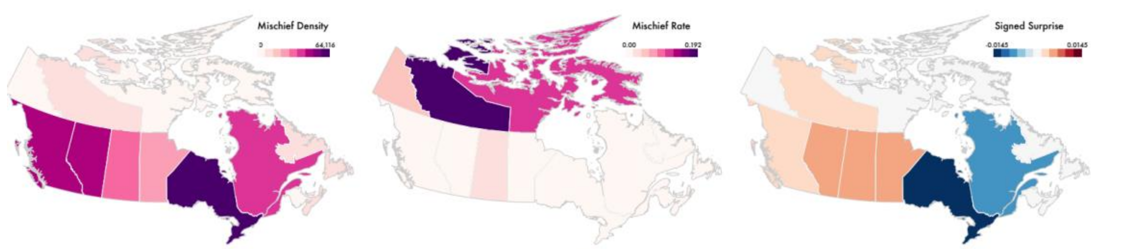


Figure 2.14 Smoothed net migration flows.

Guo et al. [21] proposed a smoothing and mapping algorithm for large-scale geographic flow data. The flow density estimation method and the flow graph generalization method were used to smooth the magnitude, and a representative algorithm was used to select representative flows and map them. Compared with the general wire binding technology, the algorithm can extract and separate important flow directions and connections well, avoiding the problems that the connections belonging to different flow directions are bundled together when reducing visual confusion.

A region is a geographical range of geospatial data with adjacent spatial locations or similar attributes, such as administrative regions of various countries, provinces, cities, etc. in the map, or attribute regions based on human social behaviour, etc. visible in specific geospatial data. In the analysis and research, the definition of the region can be roughly divided into fixed partition regions, attribute cluster regions and interactive designated regions.

Correll et al. [22] used the Bayesian Surprise method to design a Surprise map, combined the expected model with the observed data, and performed a Bayesian update step to re-estimate the traditional thematic maps that may be misleading due to underlying ratio deviations or sample anomalies. Its rationality is to increase the importance of accidents and reduce the weight of expected events. On the map of fixed geographical regions, the visual comparison results of crime event density and per capita crime rate distribution in Canadian provinces are displayed. Attribute clustering area refers to clustering geospatial locations with the same or similar attributes to obtain regional features with specific analytical significance.



(a) The **Event Density** of "mischief" in Canada. (b) The per-capita **Event Rate** of mischief. (c) The **Surprise Map** of mischief.

Figure 2.15 visual comparison results of crime event density and per capita crime rate distribution in Canadian provinces.

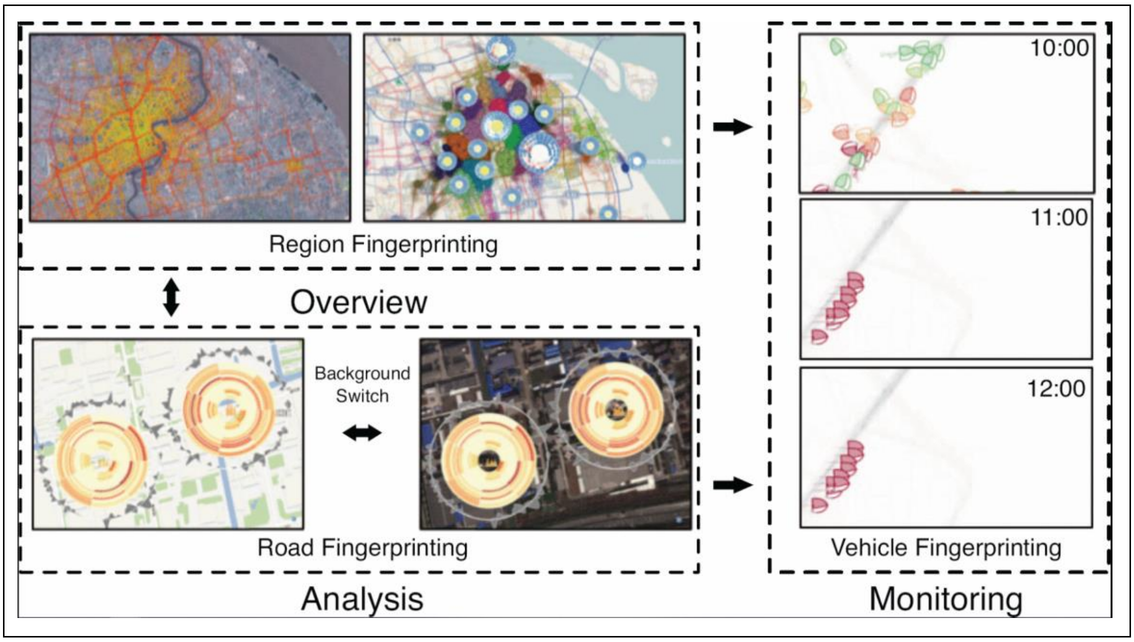


Figure 2.16 Interface of visual analysis system T-Watcher.

Pu et al. [23] proposed a visual analysis system T-Watcher, which divides the map into grid points according to GPS data, and clusters them to form a regional view. The colour brightness indicates the traffic flow, and the taxis can get in and out of the dense area. Effectively present the location distribution of hotspots in the city, and then help the transportation department to monitor and analyse the complex traffic conditions of the big cities. The interactive designated areas refer to the user's selection of visual objects in a specific area, using filtering, brushing and other interactive technologies.

Ferreira et al. [24] developed an analysis system TaxiVis that quickly queries and displays taxi origin-destination (OD) data. Users can quickly define visual query models through interactive interfaces and introduce adaptive LOD methods to reduce the amount of data is used to reduce the overlapping coverage of taxi OD points. At the same time, the thermal map is used to show the macroscopic distribution of taxi OD points, and grid maps are constructed in different ways to help users understand the OD distribution of taxis in various regions of the city.

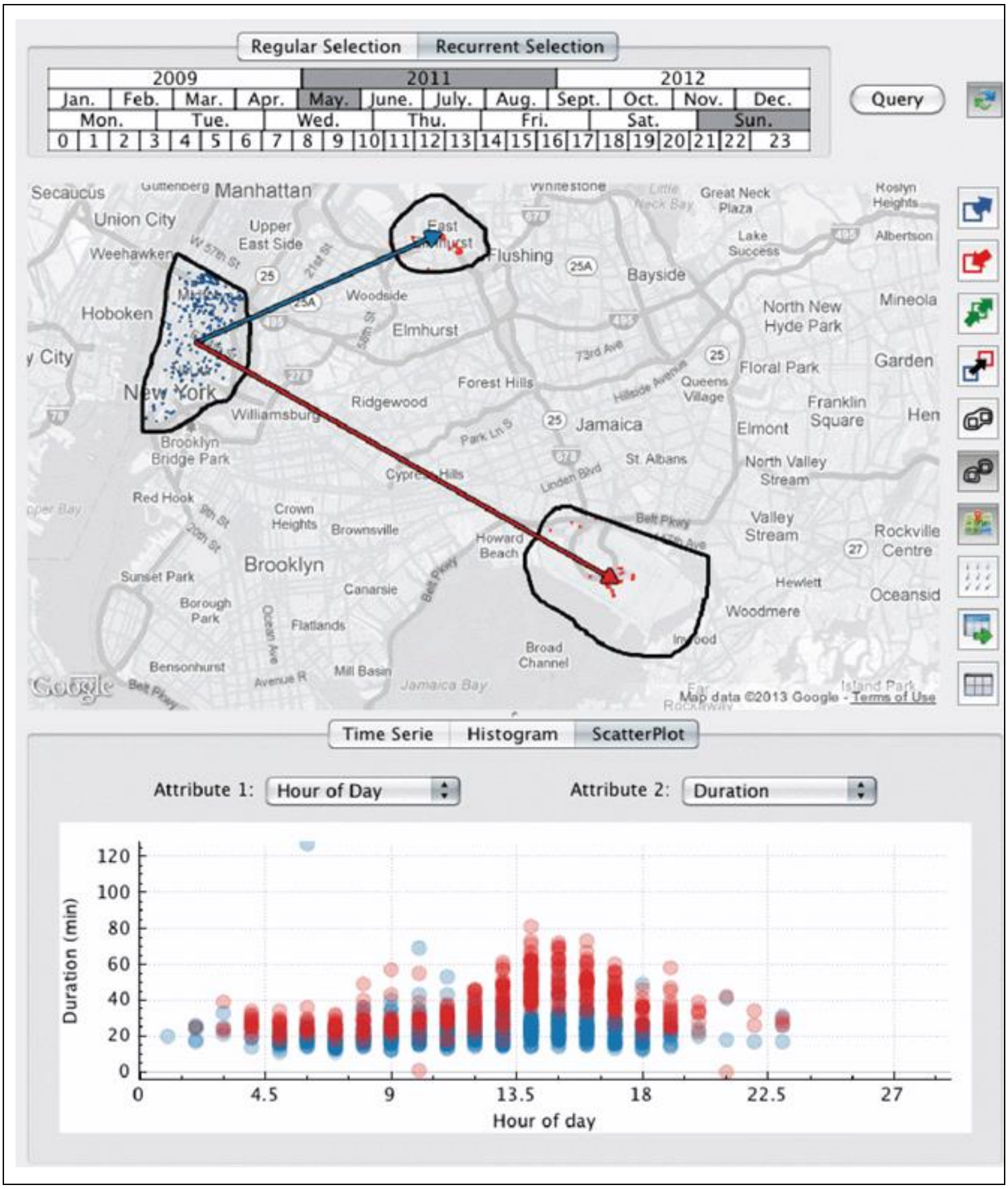


Figure 2.17 TaxiVis interface with taxi origin-destination (OD) data.

Spatial stereoscopic element visualisation refers to drawing entities or superimposing attributes on a spatial dimension based on a 2D planar map.

Reddy et al. [25] proposed a three-dimensional terrain visualisation method, which uses virtual reality modelling language (VRML) to approximate the realistic terrain of a specific area through three-dimensional modelling.

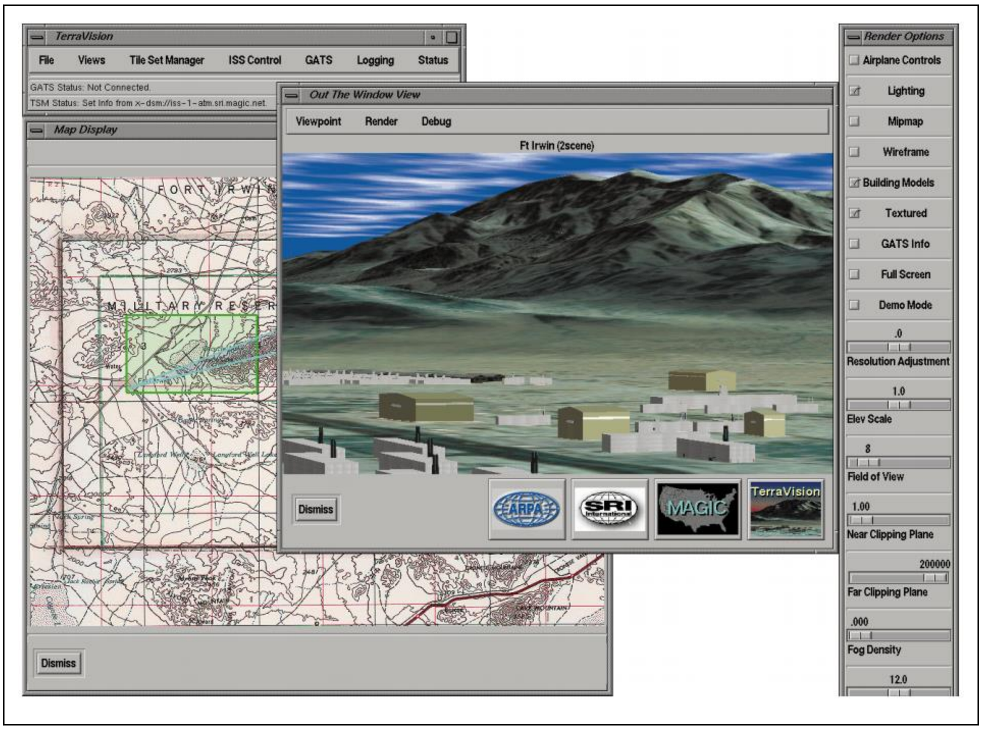


Figure 2.18 Three-dimensional terrain visualisation system interface.

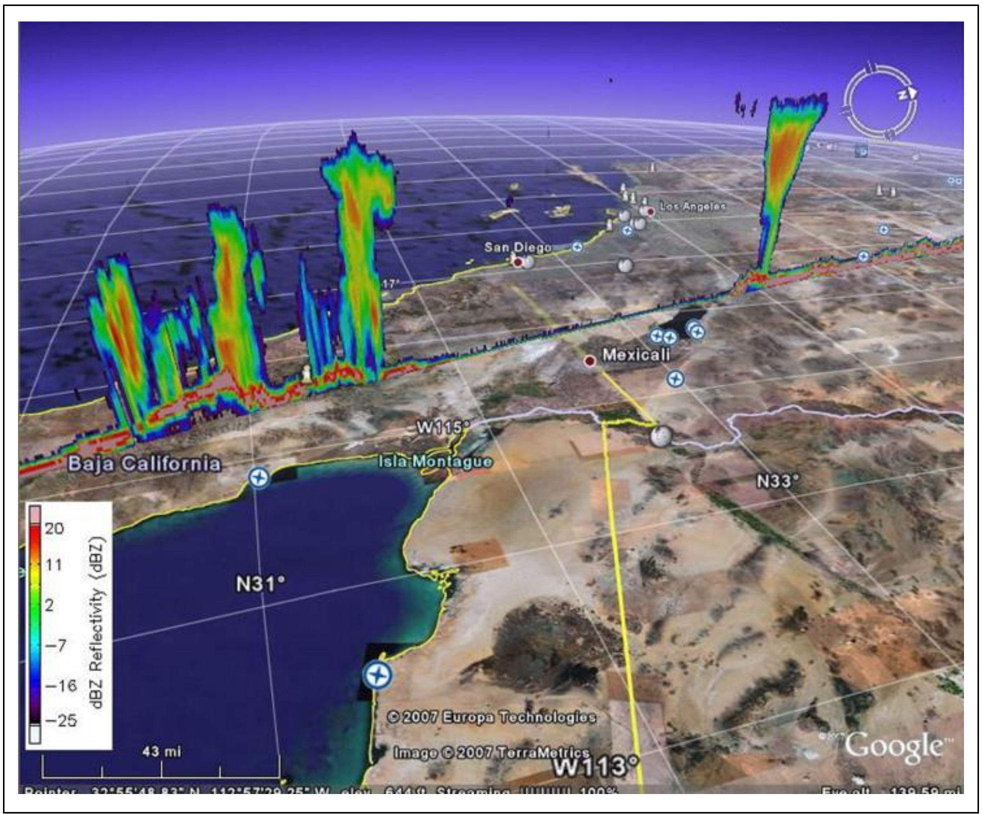


Figure 2.19 a meteorological research visualisation platform based on Google Earth.

Chen et al. [26] designed a meteorological research visualisation platform based on Google Earth. Based on the visualisation of geospatial view, the atmospheric satellite data was drawn from the vertical direction of the space. The NASA Goddard interactive online visual analysis infrastructure system was used to show the cloud reflectivity. The change of vertical profile information.

2.3 Geospatial cognition

Geospatial cognition refers to how human beings gradually understand geospatial space in daily life, conduct geographic analysis and decision-making, including a series of psychological processes of perception, coding, storing, memorizing and decoding of geographic information. [27] The research content of geospatial cognition includes the location of geography in geospatial (where) and the nature of the geography itself (what). [28] Geospatial cognition as an interdisciplinary subject between cognitive science and geoscience, needs to do specialized research of cognitive science based on geospatial related issues. The complexity of the real world and the limitation of human cognition determine that the perception of the geographic world can only be simplified and selective. Gestalt psychology perception theory is the study of the general principles of perceptual organization. [29] Gestalt psychology is the theory that reveals the most influential theory about the processes of the relativity, integrity, constancy, and organization of cognition. It also summarizes the series of perceptual organization principles called organizational law.

The lens model [30] and the supply model [31] are the most influential general models in the current perception field. The opposition between the two theories is the contradiction of whether information processing exists in cognition, and the information processing view is the basis of the emergence and development of contemporary cognitive psychology. The foundation of the emergence and development of cognitive psychology. The lens model describes the individual as an active information processor, emphasizing the construction of perception in the interaction between the current sense and the past experience; the supply model emphasizes the role of the environment itself and the biological nature of the individual's perceptual response, ignoring the role of viewer's personal experience and knowledge.

The basic principle of processing geographic information is the separation principle of location information (Where) and the object information (What) are separately perceived, processed and memorized in the cognitive system [32, 33]. The location system processes spatial information to determine the position, size, and direction of the object in space, and to encode the object [38]. The coding mode is mainly in the form of propositions for abstract categorical qualitative coding [34].

More precise spatial knowledge encodes the specific coordinate space relationships between objects in the form of representations. The object system processes various information for spatial object recognition, including shape, colour, texture, and so on. Among them, the “trigger feature” is mainly encoded in the form of representation, and other non-shape features for object recognition, such as colour and texture, are mainly encoded in propositional form [34]. “Trigger feature” refers to the information that remains unchanged under different observational conditions and can be used for object shape recognition [35]. Further research has also shown that the two are not completely separate, but can be interconnected and interacted in some way [36].

Marr's sketch model [37] is the most influential theory of geographic perception processes and steps. The study begins with the sensory registration of the image memory, and ends with the scene being identified as a series of objects, concepts, and spaces. The process is divided into three stages: first, the stage forms the expression of the original sketch, the brightness of the light and its changes are deployed into a series of simplifications and miniatures of spots, lines, curves, etc. in all directions, forming a psychological expression of local geometric forms. The second stage forms the expression of a 2.5-dimensional sketch. The original sketch is processed through a number of psychological processes, such as stereoscopic images, shadows, textures, contours, etc., to form the visual surface width, direction, boundary, and discontinuity obtained from a certain observation point. The third stage forms the expression of three-dimensional sketches, and each object is extracted and identified from the environment. The first and the second stage coordinate systems are centered on the observer, while the third stage is centered on the object. Information includes the volume, the space occupied, and the shape of each part. On this basis, the object becomes a type by feature analysis or prototype matching. The instance of the concept is used to complete the perception process of the scene. The Marr model is a complete bottom-up information processing course. Although the original sketch stage of perception is completely bottom-up, but the latter two stages contain a top-down analysis process [38, 39]. In the top-down process, the abstract knowledge stored in the long-term memory can guide the recognition process of the object in the form of a hypothetical expectation or schema. The high-level perception process consists of interconnected bottom-up and top-down hypothesis testing processes [39]. The processes include receiving information, forming and testing hypotheses, and then receiving information, testing hypotheses again until verification.

The scale problem is one of the cognitive methods related to geographic information science [40]. The difference in perception patterns is the main basis for spatial scale division. Psychology divides space into figures, vista, environment and geographic [41] according to different scales of spatial perception. The image is smaller than the human body and can be fully perceived from one observation point; The human body can be perceived from an observation point by changing the angle of view (such as scanning); the environmental space cannot be fully perceived from an observation point, and needs to move through the space to obtain a series of "observation views" at multiple observation points to form a plurality of observation points, and then to splice the views to form a cognitive space in the form of a collage (Cognitive space to obtain a complete perception of space [42]; geospatial scale is much larger than environmental space. It cannot be symbolized by self-experience and direct perception. It is necessary to symbolize the reduction of space into a graphical space through symbolic representations such as maps and 3D models. Based on the disposition, mobility, and size, space can be divided into manipulable objects, non-manipulable object, environmental geographic panorama, and map space. Map space is a large-scale space, which reduce and simplify spatial information in symbolic form. It can be perceived without movement [43]. Partitioned space concept is of great significance to the visual design.

2.4 Visualisation evaluation

Nivala et al. [44] use the methods of expert evaluations and user tests, evaluate four types of web maps and found more than four hundred problems in user interfaces, map visualisation and map tools.

Morse et al. pointed out that there are few studies on information visualisation evaluation, and a small number of studies have not proposed direct and universal visualisation evaluation methods [45]. Meyer argue that there is no generally accepted standard for optimal data presentation, in part because of the lack of empirical evidence and the variety of visualisation methods [46]. Purchase believes that it is not advisable to evaluate based on aesthetic factors or efficiency, but should consider the ability of visualisation to improve user performance. Studies have shown that the relationship between aesthetic traits and performance of interfaces is complex [47]. Although aesthetic are often associated with perceived ease of use, perceived ease of use and actual availability may not be relevant. Freitas advocates the use of two criteria, expressiveness and effectiveness, to evaluate the effects of information visualisation. He proposes two criteria: cognitive complexity and spatial organization [48].

Dowell et al. consider these evaluations to be evaluations in specific areas, which are less popular [49]. A more direct and more general evaluation method is to apply cognitive psychology theory, use different visualisation methods to present the questions to the subjects, and compare the performance of different visualisation methods by evaluating the accuracy, confidence and time spent on the responses [50].

In general, advanced and there is no perfect evaluation method and system for information visualisation, but most evaluation methods are focus on the user test.

3. METHODOLOGY

3.1 Development process

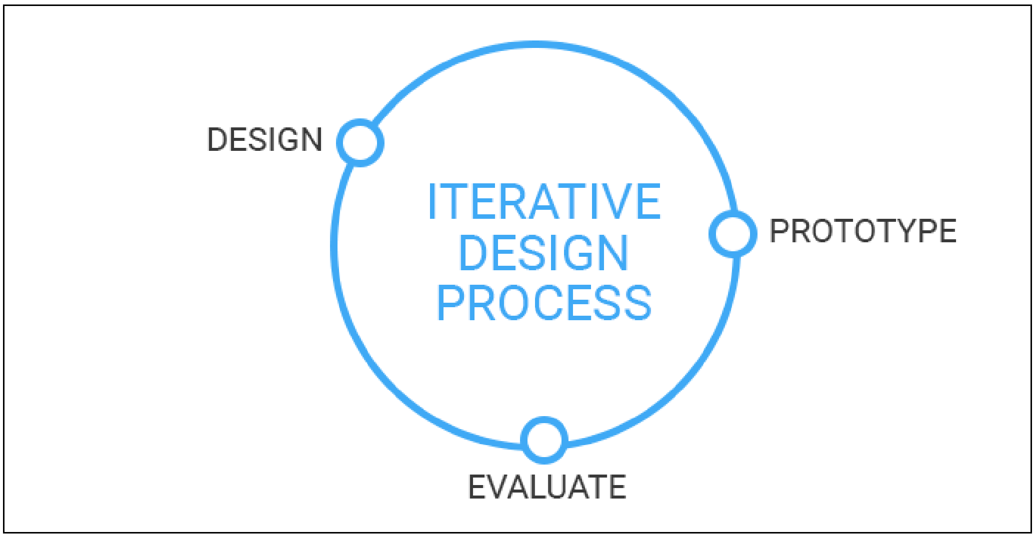


Figure 3.1 Iterative design methodology process.

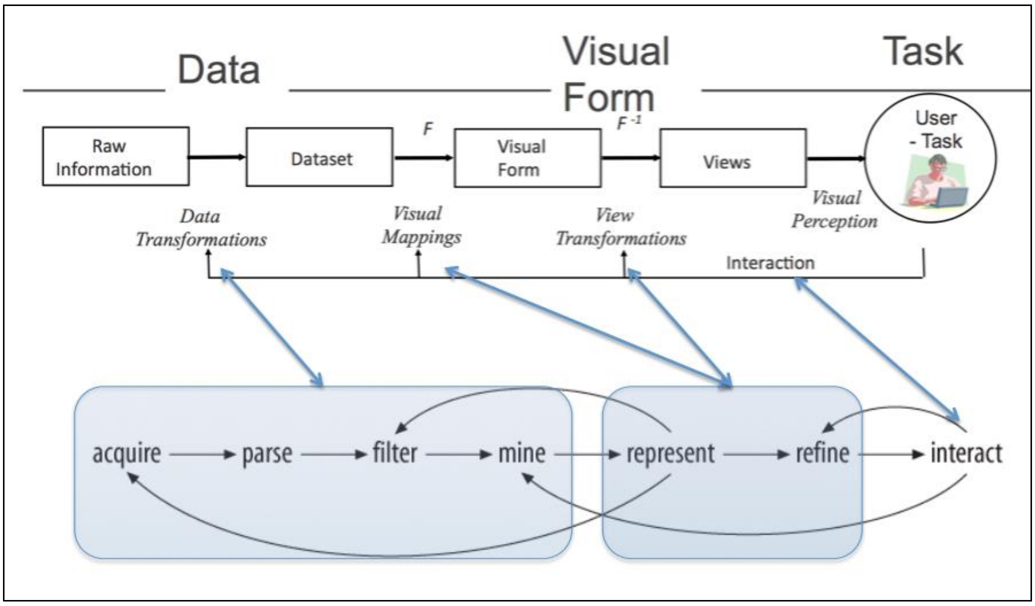


Figure 3.2 Seven states of visualisation.

In the project, the process follows the iterative design methodology [52], which based on cyclic design process, including user research, requirements analysis, design and implements, user testing, and continuously improving the prototype. The principle of the method is keeping user involved in the process of design by constantly evaluation. From the very beginning of the requirement to the prototype usability test, the real user would give advice and idea to the project design. Therefore, the prototypes mentioned in this paper refer to most typical one during the process, which means there are many versions with differences in details. From the development perspectives, the process is conducted in a more specific way, an overall has been given as below.

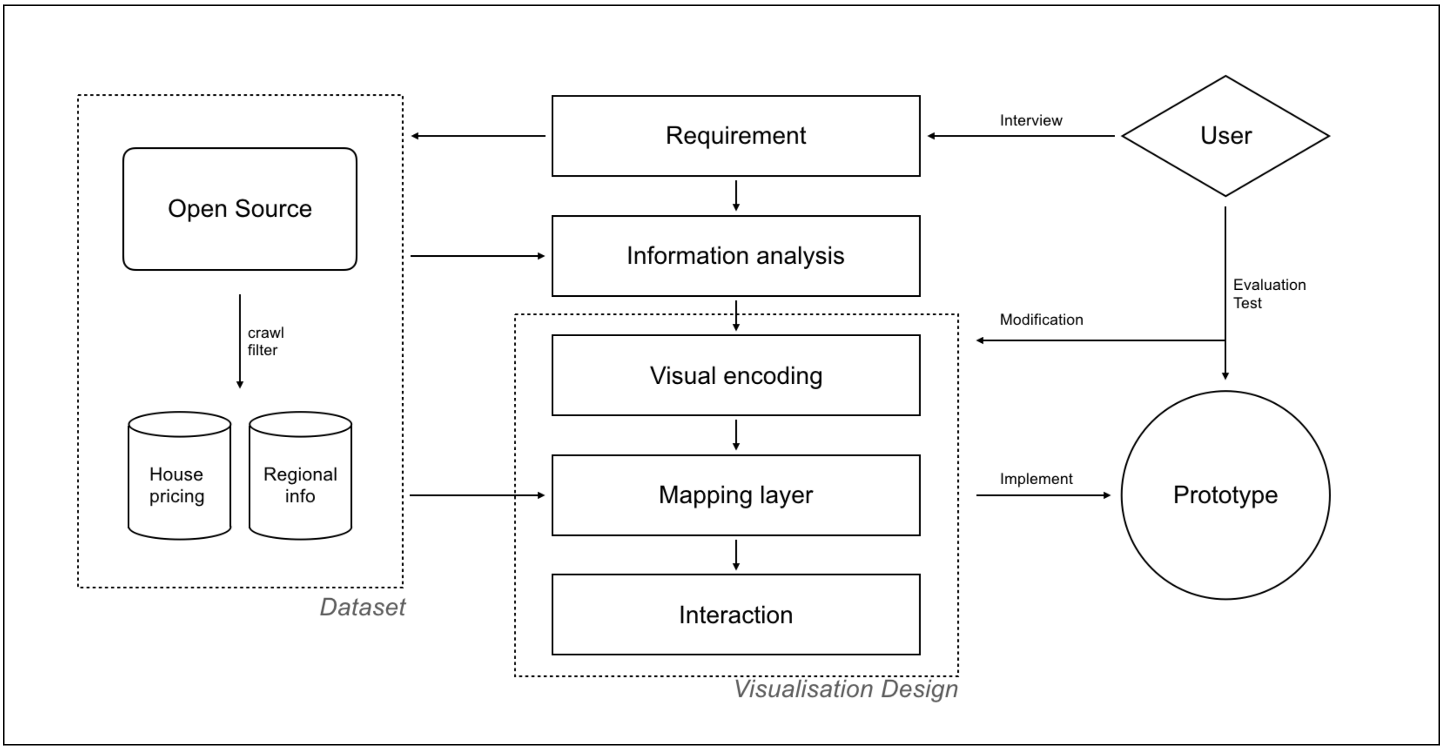


Figure 3.3 Development process.

The following content would also be illustrated in this structure.

There are two main prototype, Prototype I is a visualisation system of Nottingham renting house visualisation with regional info, while Prototype II is built upon the dataset of London selling house pricing with regional info.

3.2 Requirement

The requirement plans an important role in whole process, which guide the visualisation design and usability task setting. The process of development is built upon the information gather from the potential real users from online informal interview. By analysing the scripts of answers, which towards specific aspects, and user experience on existing house pricing platform, the requirement is generated.

3.2.1 Online informal interview

There six interviewees in total, with three females and three males in respectively, aged from 22 to 26. Four of them have experience in renting house while the remains are currently seeking for a house or accommodation. Currently, they live in China, UK and Australia. All interviewee are friends or classmates and be recruited personally. Interview is conducted online on an instant messaging tool (WeChat) in a semi-structured method. The question is raised during discussion, which are set as follow.

Q1. How do you rent a house?
Q2. What about rent online?
Q3. Could you try to rent house on (Rightmove, Lianjia, Realestate) based on your current need? Please tell me the steps.
Q4. How do you compare the pricing and locations between properties
Q5. How do you think about map searching? And please evaluation this method.
Q6. Do you want to know the regional information such as crime rate to make decisions?
Q7: What kind of regional information do you want to know?
Q8: Do you want to gain an insight of an area where you live or decide to buy a house?
Q9: What kind of information support your decision of buy a house?
Q10: How do you think of the interaction method during using the previous website?

Table 3.1 Interview questions.

Q1 and Q2 is designed to investigate interviewee's experiencing in renting house.

Q3 to Q5 is to find out workflow in the online platform of renting a house by completing a task (The platform selecting of depends on preference and current living city)

Q6 and Q7 is to investigate the potential need of the regional info. (An open question without clear definition of regional)

Q8 and Q9 find out the issue of a house.

Q10 is an overall evaluation of the website interaction.

The feedbacks are taken from the chat records and simplified as transcripts.

3.2.2 Requirements analysis

The purpose of the requirements analysis is to determine the work flow when user is looking for a house property and the demand for the regional information. By exploring these two aspects, the requirement of visual encoding and interaction would be correspondingly designed.

From the feedback of transcripts, for a given request, the six interviewee firstly set the main search criteria (usually price, the distance to target location,.etc.), the website would display the filtered properties by list or grid. After that, they compare properties with each other and click into the properties to find the detail information (position on map, image and decoration) and make decisions. They may go back to compare and repeat these actions.

Two interviewees have used the map search to find the properties. In this method, they finish searching and comparing properties within the map, which mapping target properties by positions. The other four interviewees are asked to experience map search function. After having a try, they think it has some practicability. As for the cognition performance, interviewee was asked to evaluate method in understanding information on the map search. They all thought that the geographical location is clearer, and the rest of info is the same as the grid or list, especially in interaction.

As all website platform show similarity in workflow. There takes Rightmove as an example.

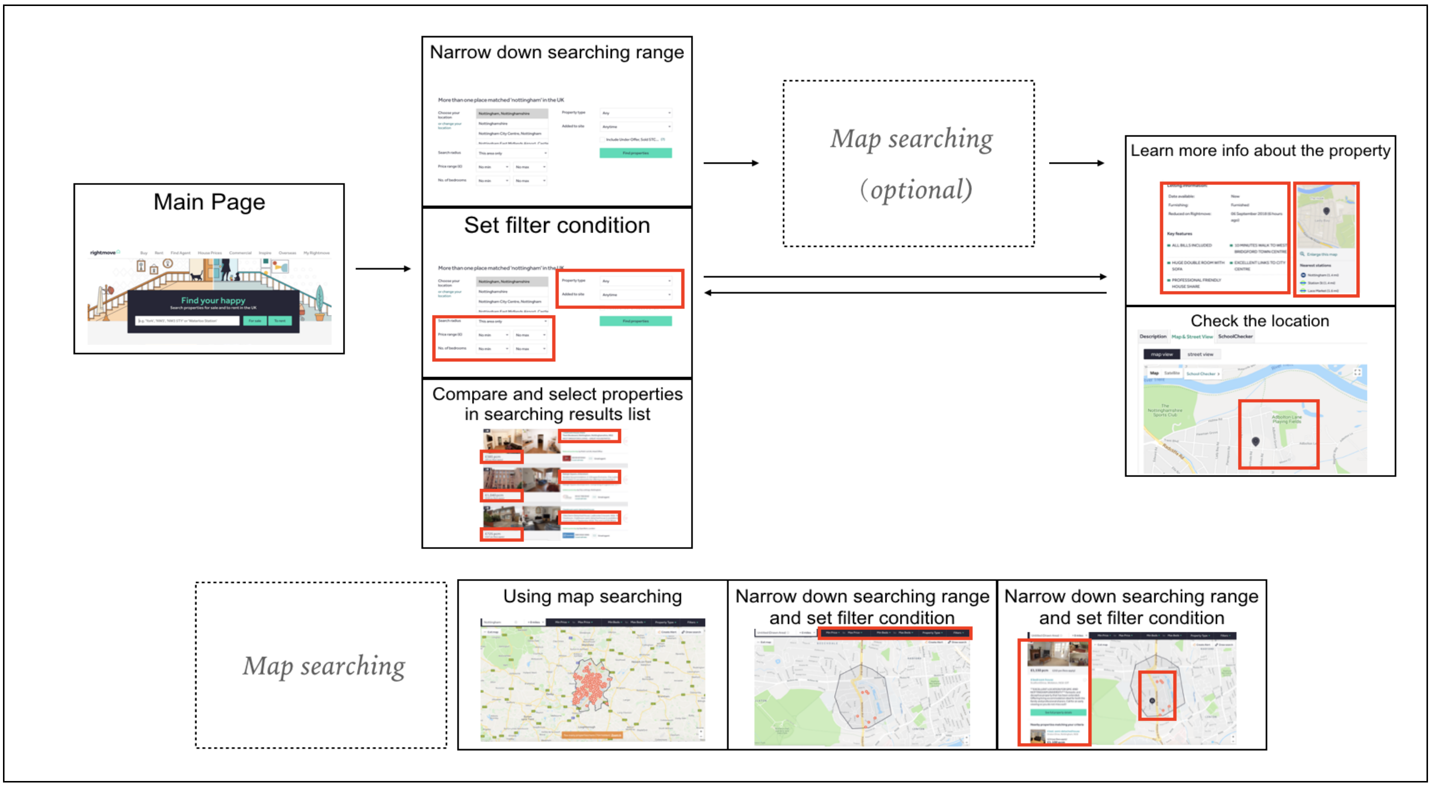


Figure 3.4 Workflow of searching properties.

As for regional information requirements, all female interviewees and one male interviewee indicated that they want to know the crime rate or in another word, regional security situation. When asking about other types of regional info, the answers shows individual preferences, for example, some want to learn more about environment while other care about transportation convenience. The remaining two feel that it doesn't matter.

In addition, in view of the topic of buying housing price and visualisation, although the six respondents neither plan to buy a house, nor able to buy a house at present, they show great interests, talking about the future with for buying a house. The major issue that is health discussed is way of making money. As for visualisation, they want to understand the of housing prices and regional distribution, while considering various factors such as location, school district, environment, and public health access. Two of interviewees evaluate the visualisation system on Lianjia.com. They believed that the visual encoding could clearly show house info, but the stack of the text hides each other, which made the reading less convenient, and the price was not intuitive to display.

Finally, when asking about the overall evaluation information visualisation several website platforms, most of them said that the system is now good enough, but it is more troublesome to find the property on the map, and the information about the regional information is relatively less. For the interaction design, they said that it is easy enough to use.

From the above analysis, the requirements generate as follow.

No.	Requirements	Source
1	Users can quickly locate the target location.	Interviewee 1, 2, 3, 4, 5, 6
2	Users can quickly find the house properties in the area.	Interviewee 2, 3, 5
3	Users can quickly understand and price information in the area.	Interviewee 1, 3, 4, 5, 6
4	Users can quickly compare price information in the area	Interviewee 4, 6
5	Users can quickly learn about regional information in the area, which include several aspects.	Interviewee 3, 4, 6
6	Users can clearly recognize the average price of the area.	Interviewee 1, 2, 3, 4, 5, 6
7	Users can filter information and properties as they wish.	Interviewee 4, 5, 6

Table 3.2 Requiments.

3.3 Design and implement

The requirement plans an important role in whole process, which guide the visualisation design and usability task setting. The process of development is built upon the information gather from the potential real users from online informal interview. By analysing the scripts of answers, which towards specific aspects, and user experience on existing house pricing platform, the requirement is generated.

3.3.1 Approach

Data collecting

a. Data source

There are two main types of data used in this project. One is a regional information called The English Indices of Deprivation 2015, which is divided into LSOA (A Lower Layer Super Output Area (LSOA) is a GEOGRAPHIC AREA). The indices describe the relative development level of the region, which is distinguished by the deprivation of several dimensions. Among them, the least deprived is better developed, and the indices score are higher. The other one is house price information, which includes location and price information for house properties in the area. The pricing refers to selling price and rent price and depends on the type of system.

Regional data acquisition is from

https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015'

Official Statistics - English indices of deprivation 2015. Statistics on relative deprivation in small areas in England.

Geographic data acquisition comes from LSOA

https://github.com/gausie/LSOA-2011-GeoJSON

contains all the data from UK LSOA (2011) in GeoJSON format. GeoJSON is an open standard format designed for representing simple geographical features, along with their non-spatial attributes. It is based on JSON.

The house price data is mainly from online house property portals, Rightmove and Zoopla. Nottingham's house pricing data comes from Rightmove. The rent search page is obtained by conditional filtering and paged crawling. London's house pricing data comes from Zoopla’s sale pricing, which is also obtained through page index.

As for data cleaning and integration, in the Deprivation table, the data of Nottingham and London areas are separately filtered, and column LSOA (2011) is used as a header to form a hash table. Query the hash table in the GeoJSON file and insert the corresponding information. This work is done by the python scripts. As for the housing pricing, the data also reshape the price by crawler.

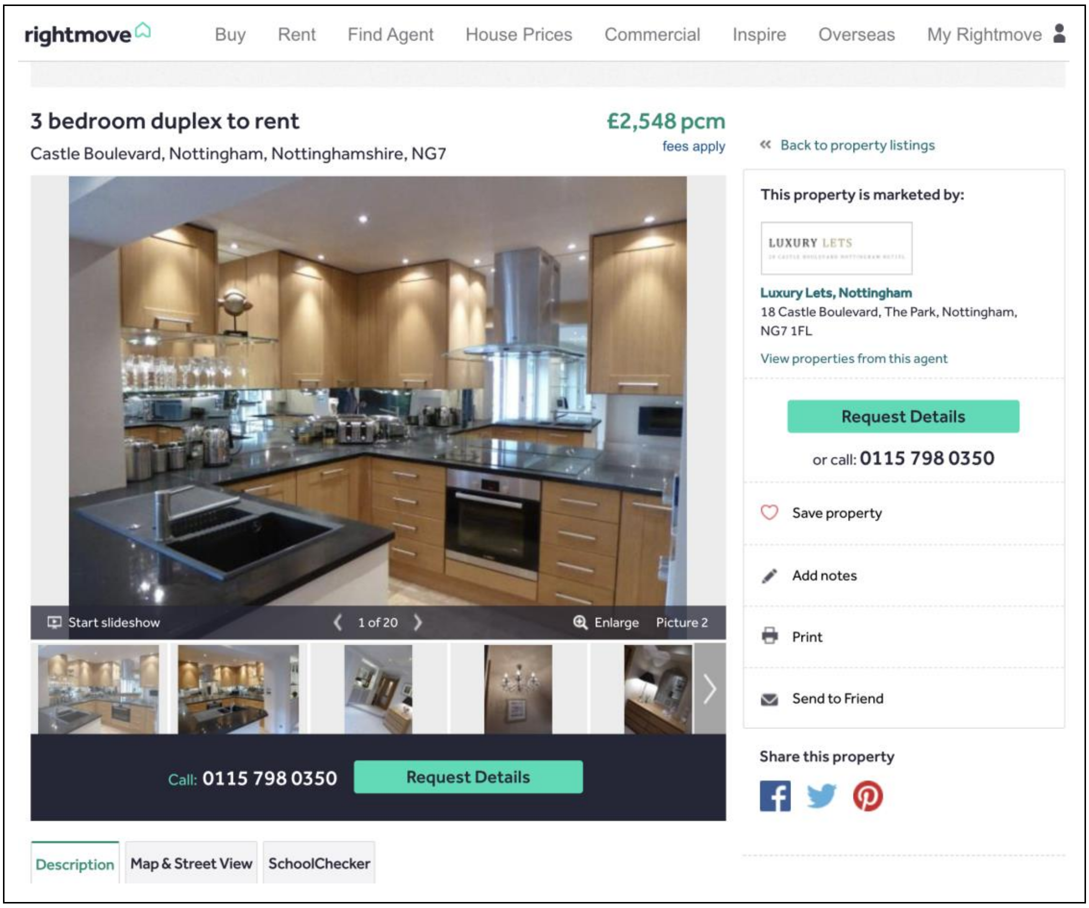


Figure 3.4 Online house website.

b. Data crawl

The basic principle of web crawler is to simulate the behaviour of human web browsing.

The first step is data crawling, which is mainly to obtain raw data through the Internet and and done by spiders. The web crawler downloads the HTML text of the webpage to the local as client, by sending an HTTP request to the web server. In theory, web pages that are accessible to humans can be obtained through crawlers. Subsequently, the content screening is performed. Since the HTML text content of the webpage is usually disordered, the data extraction of the webpage content is filtered according to the specific situation, and the useless information is removed to retain the effective content. This step also includes work such as deduplication and content analysis.

The crawler in the project are built on Scrary. The process are to iterate through the houses in the list one by one and collect their information, starting with the index on the home page. The crawler will then automatically go to the next page and repeat the former steps.

Scrapy is an open source web crawler framework written in python. It includes seven main components, Engine, Scheduler, Downloader, Spiders, Item Pipeline, Downloader middleware and Spider middleware. The engine is the core component of the entire framework, responsible for controlling the data transfer between the components, triggering according to the event, calling other components. The customisation of the function is carried out in other components besides the engine, and the development is mainly in the writing of spiders and the selector of Items.

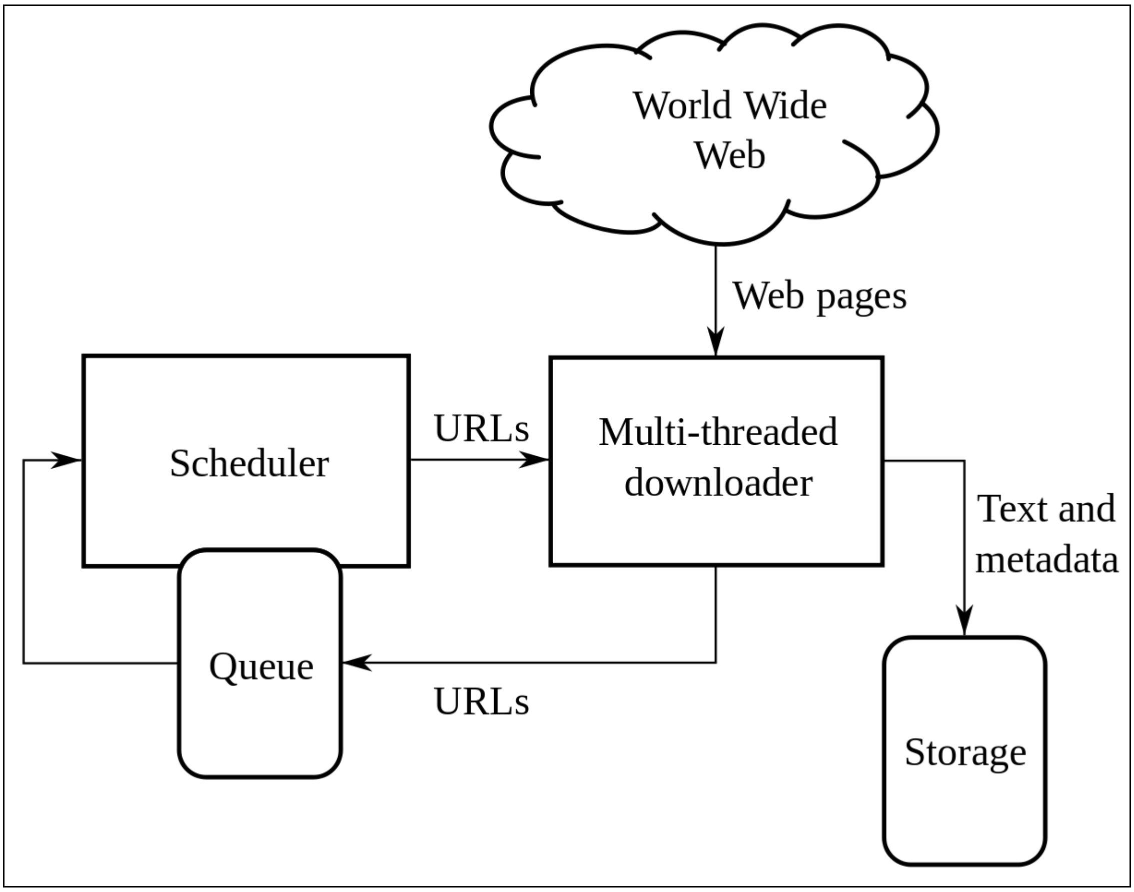


Figure 3.5 Web crawling principle.

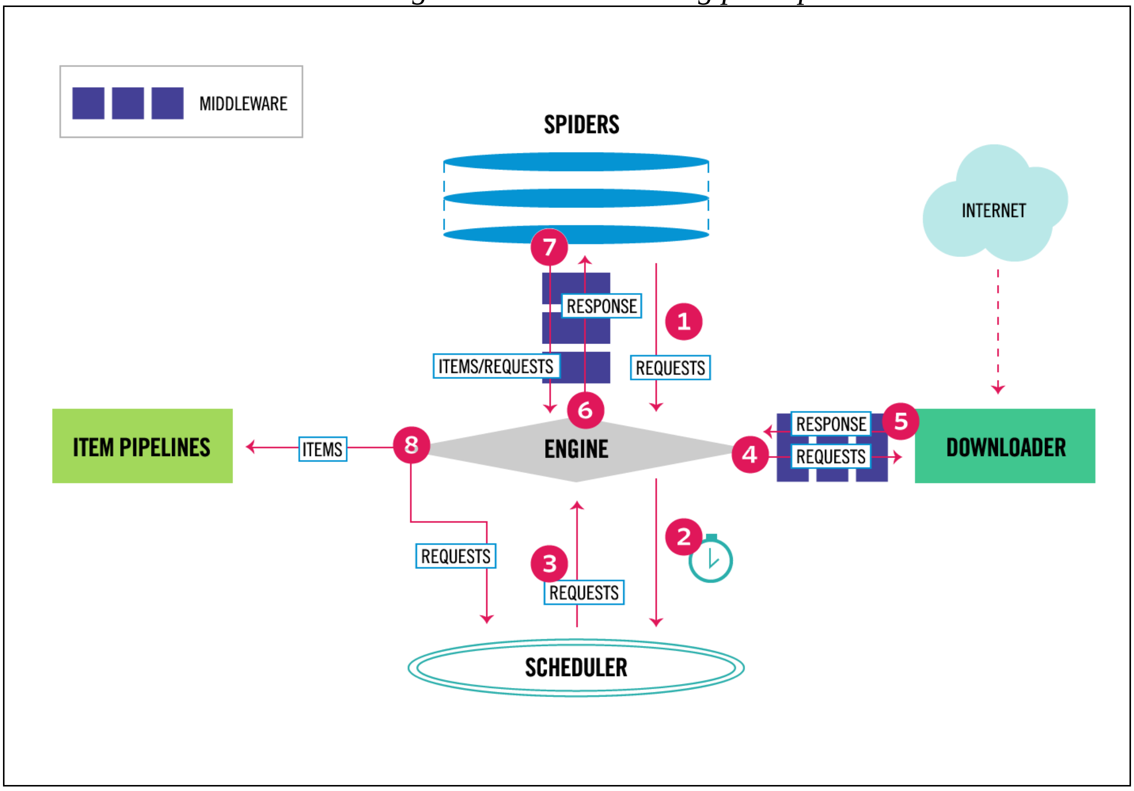


Figure 3.6 Scrapy structure.

System development

The system is built on Kepler.gl and launched on local server.

Kepler.gl is a data-agnostic, high-performance web-based application for visual exploration of large-scale geolocation data sets, which is a React component built on the Uber open source internal visualisation toolkit deck.gl that supports the CSV, JSON and GeoJSON formats.

Deck.gl is one of the main frameworks in the vis.gl framework suite， which is Uber Visualisation’s open-source frameworks. It is designed to make visualisation of large data sets simple. It enables users to quickly get impressive visual results with limited effort through composition of existing layers, while offering a complete architecture for packaging advanced WebGL based visualisations as reusable JavaScript layers.

In the web interface, users can browse through the data by manipulating different layers, performing filtering or aggregation. All layers are accelerated by the GPU due to support for the WebGL development mechanism deck.gl. Even for web applications, kepler.gl can successfully draw millions of data points, which means it can render thousands of itineraries and spatial aggregation, so it has high performance. The feature is that it can be used for visual exploration of large-scale geolocation data sets.

The project is developed with Node.js as a platform, which built on Chrome's JavaScript runtime for easily building fast and scalable network applications. JS packages are managed by npm.

3.3.2 Dataset

Crawler developer

a. Website Structure analysis

According to the needs of the project, the spider would main crawl the data of pricing and location, namely latitude and altitude. Given that there are no open datasets for both sites, it needs to develop the crawler script for data capture. First, use chrome DevTool to analyze the webpage. After parsing the webpage, use xpath and regular expression to extract the content, and complete the crawler function in the scrapy system framework.

Rightmove

On the search page, url

https://www.Rightmove.co.uk/property-to-rent/find.html?keywords=&sortType=6&viewType=LIST&channel=RENT&index=0&radius=0.0&locationIdentifier=USERDEFINEDAREA%5E%7B%22id%22%3A5002099%7D

It can be seen that the number corresponding to index is the page number. Thereby establishing an index table of the crawl page;

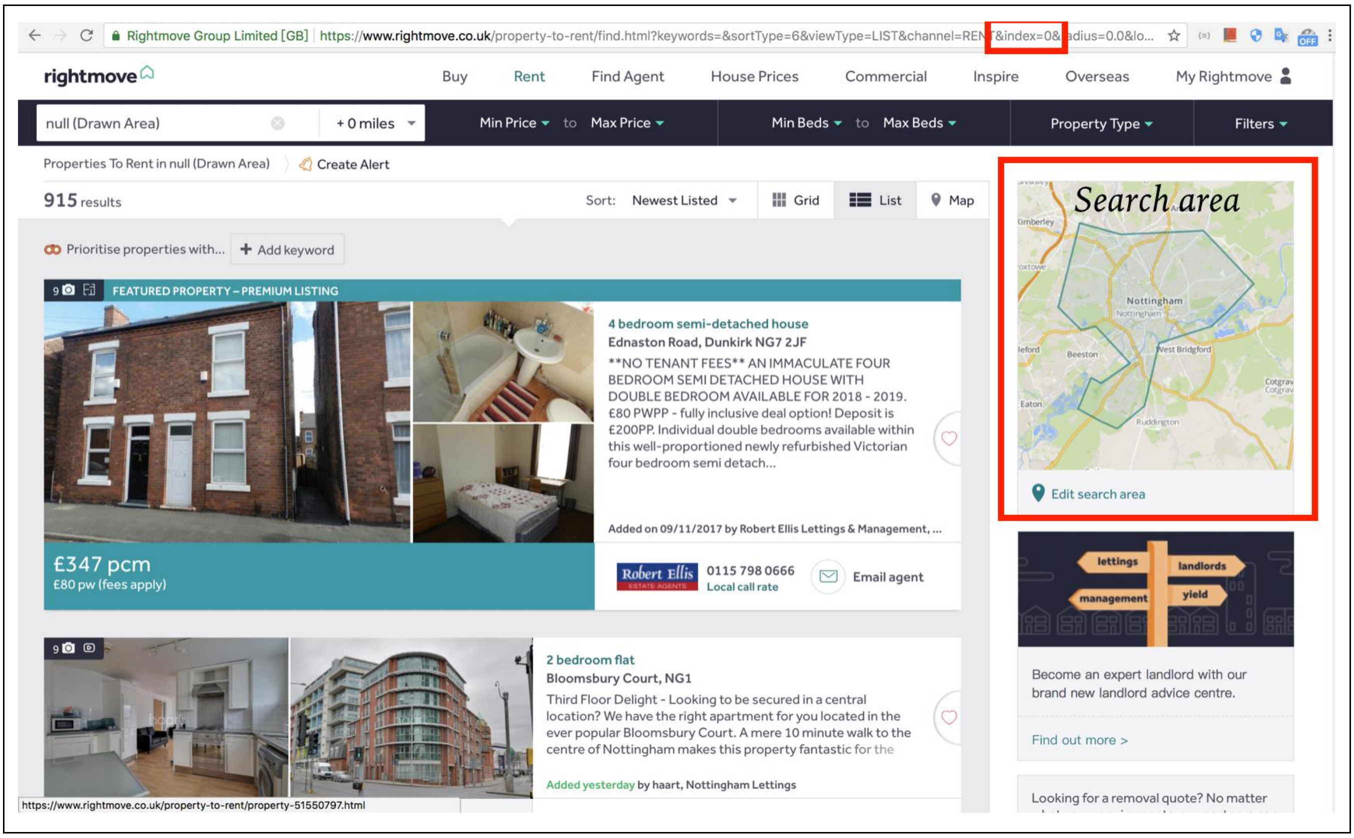


Figure 3.7 Rightmove website.

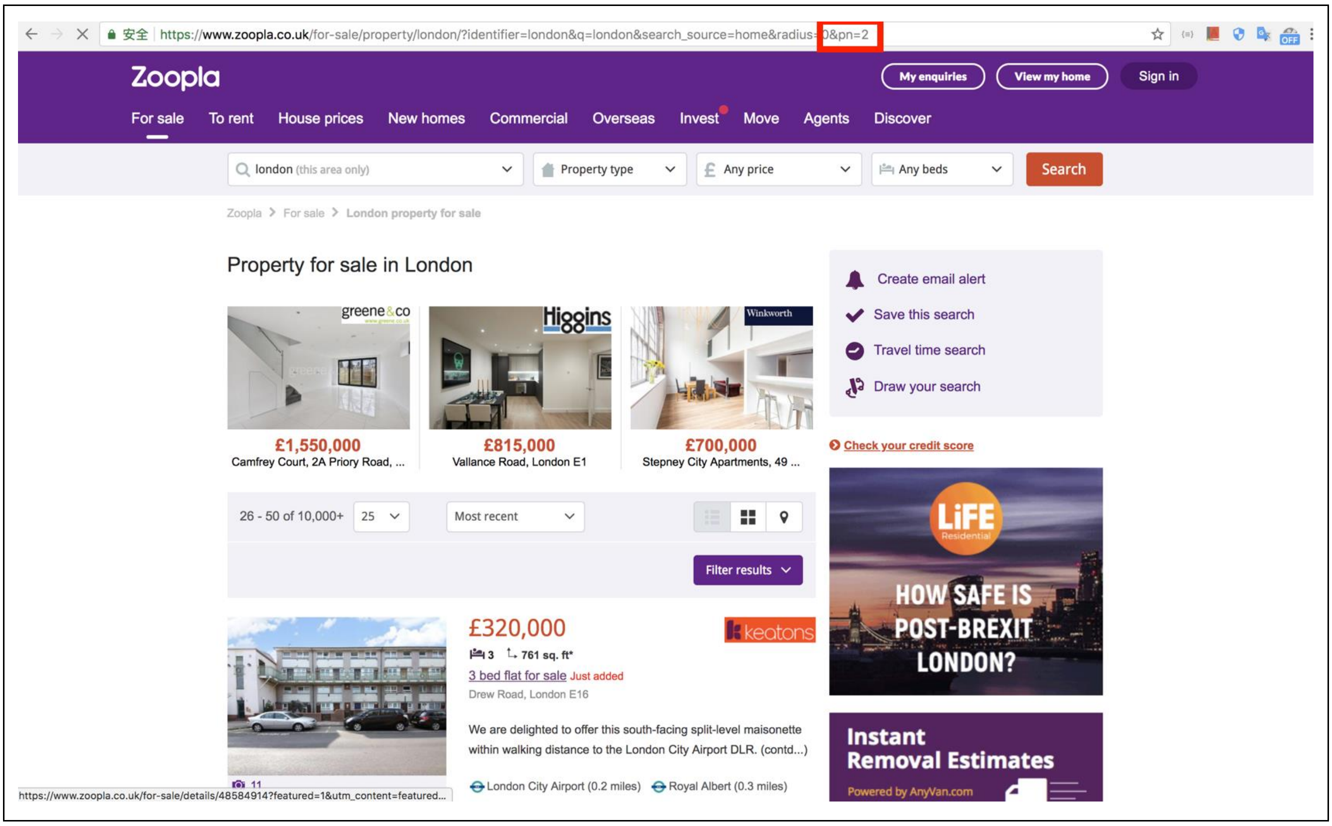


Figure 3.8 Zoopla website.

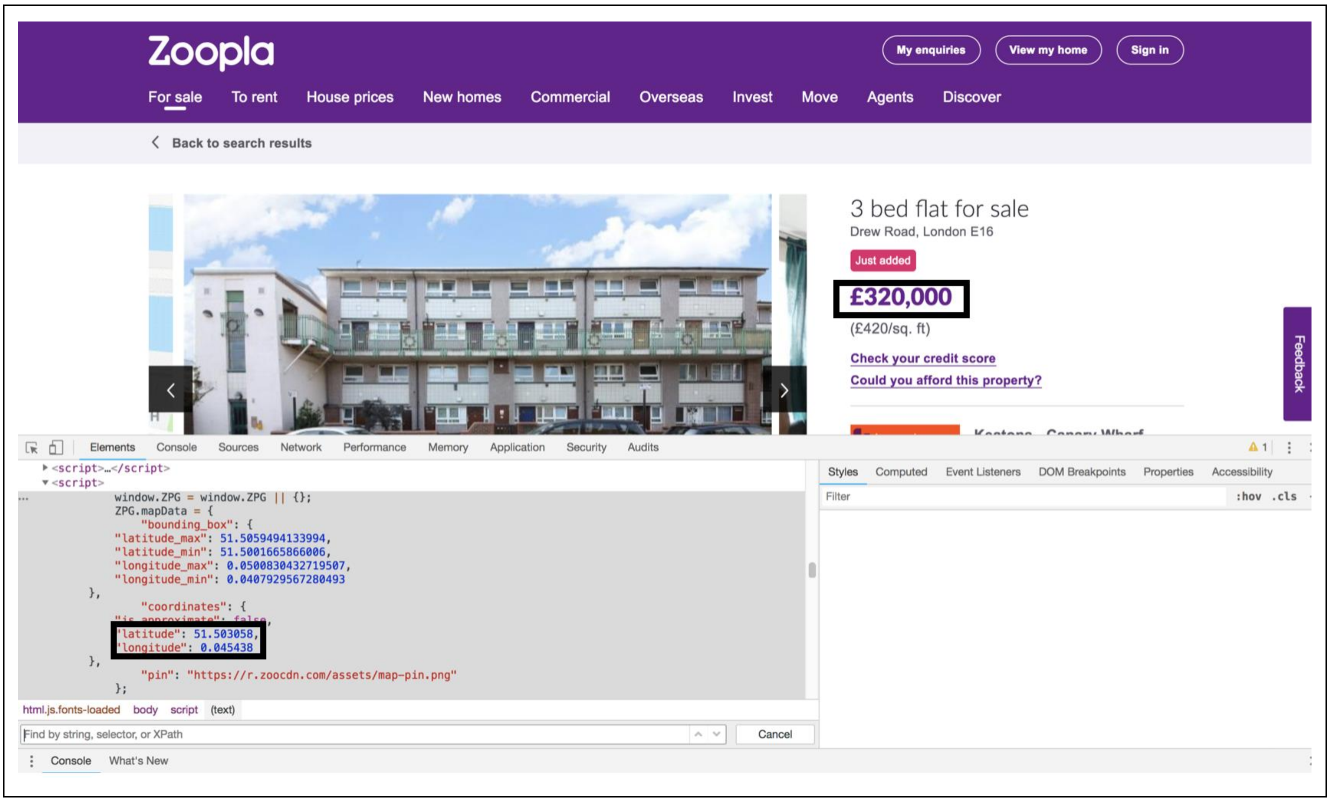


Figure 3.8 Zoopla website.

On the property page, url

(https://www.Rightmove.co.uk/property-to-rent/property-<propertyID>.html)

According to DevTool, data exists in two types on the page, one is the text of html elements, this kind of information can be directly obtained by xpath, and the other is loaded into the function of JavaScript by json object. This type uses xpath. A regular expression matching is required after extraction.

Similarly, Zoopla's web page structure is relatively simple.

b. Selector

XPath (XML Path Language) is a query language for selecting nodes from an XML document. In addition, XPath may be used to compute values (e.g., strings, numbers, or Boolean values) from the content of an XML document. XPath was defined by the World Wide Web Consortium (W3C). [53]

A regular expression, is, in theoretical computer science and formal language theory, a sequence of characters that define a search pattern. Usually this pattern is then used by string searching algorithms for "find" or "find and replace" operations on strings, or for input validation.

Since this project is based on pricing, there are not many requirements for the property's remaining variables (Number of bedrooms, Number of bathroom.etc.), and the variables have great similarities, so only the description is taken in the Nottingham system. Unique text variables. Only the pricing and location information is kept in the London system.

Variables	Xpath	Regular Expression
Rightmove(https://www.Rightmove.co.uk/property-to-rent/property-<propertyID>）		
Description	//title/text()	/
ID	//div[@class ="clearfix main"]/script[3]	propertyId":(\d+),"viewType
Price	//p [@class="property-header-price "]/strong/text()	(\d+)
Latitude	//div[@class ="clearfix main"]/script[3]	latitude":(\d.+),"longitude
Longitude	//div[@class ="clearfix main"]/script[3]	longitude":(-\d.+)\},"
Zoopla(https://www.Zoopla.co.uk/for-sale/details/<propertyID>)		
Price	//*[@id="main-content"]/div[2]/div[1]/div[5]/section[2]/div/span[2]/text()	/
Latitude	/html/body/script[4]/text()	"latitude": "(\d.+)"
Longitude	/html/body/script[4]/text()	"longitude": "(\d.+|-\d.+)

Table 3.4 Selector paths

Data Analysis

After collecting data, the analysis such as distribution and feature is the key for visualisation design. The data is analysed by python scripts which based on library pandas(pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools) and numpy(NumPy is the fundamental package for scientific computing). There are also some plots by matplotlib and run on Jupyter Notes.

```
# histogram
price1 = df1['price']
plt.hist(price1,bins=40)
plt.xlabel('price in Nottingham')
plt.ylabel('Count')
# plt.xticks([1000,2000,3000,4000,5000],['20%','40%','60%','80%','100%'])
plt.show()
```

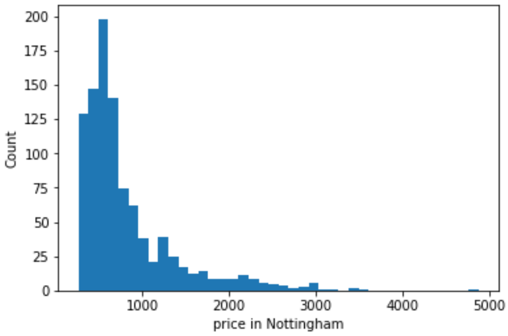


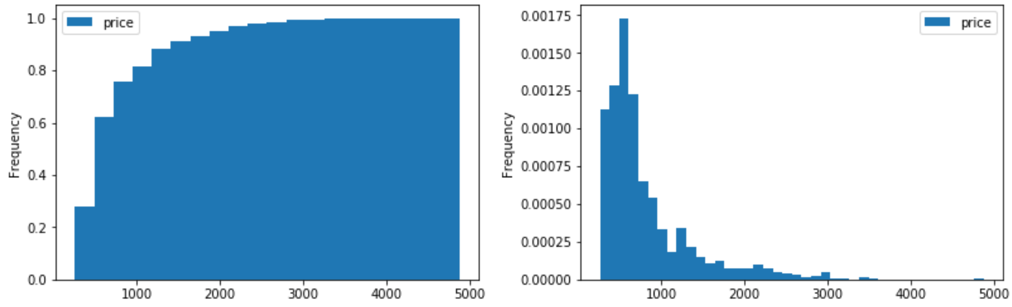
Figure 3.9 Plot scripts.

After filtering , the data set include 991 properties renting in Nottingham with 182 LSOAs and 8943 properties selling in London with 4833 LSOAs.

Variables	Type	Example
Nottingham – Prototype I		
Price	Interval	640 (£pm)
ID	Category	74754821
Latitude	Geospatial	52.94321
Longitude	Geospatial	-1.17868
Desription	Text	2 bedroom house to rent in Old Lenton, Nottingham, NG7 - P2503, NG7
Crime Score	Vertical	62
Education, Skills and Training Score	Vertical	60
Health Deprivation and Disability Score	Vertical	52
Income Score	Vertical	55
Living Environment Score	Vertical	44
LSOA Name	Text	Nottingham 018A
LSOA Coordinates	Geospatial	/
London - Prototype		
Price	Vertical	450000 (£)
Latitude	Geospatial	51.4658
Longitude	Geospatial	-0.08819
IMB Score	Vertical	18
LSOA Name	Text	Westminster 001A
LSOA Coordinates	Geospatial	/

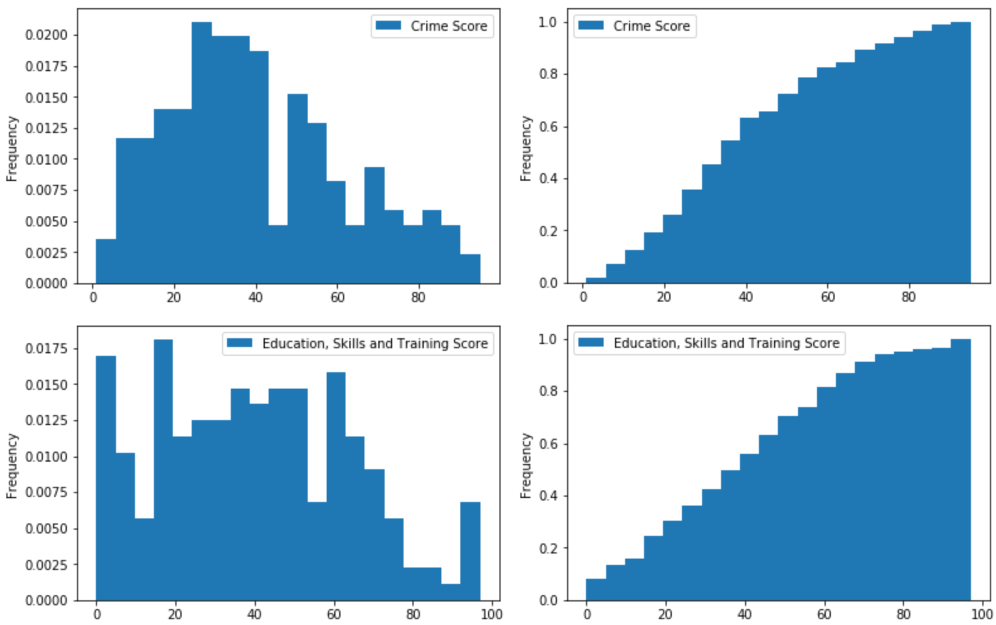
Table 3.4 Selector paths

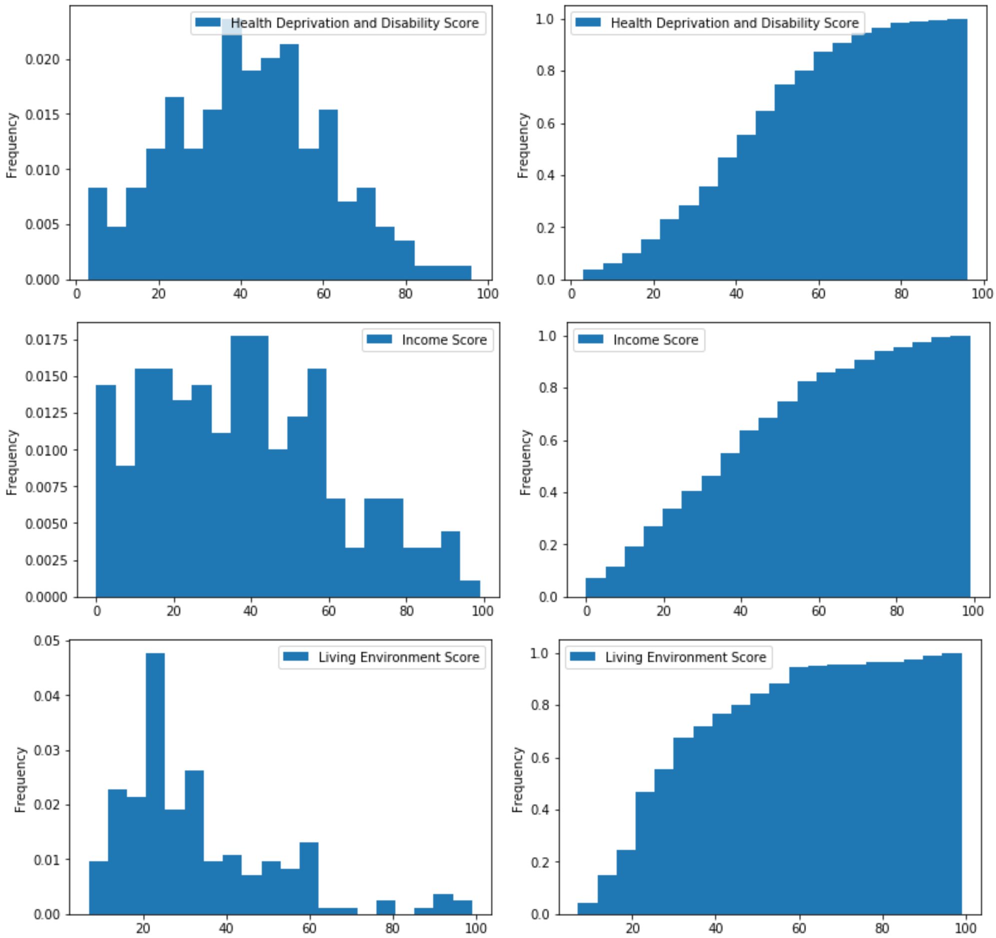
Before visualisation design, frequency distribution and cumulative distribution could be used to analyse the vertical data and find feature. For geospatial info, density and distribution analysis are required at mapping on coordinating location.



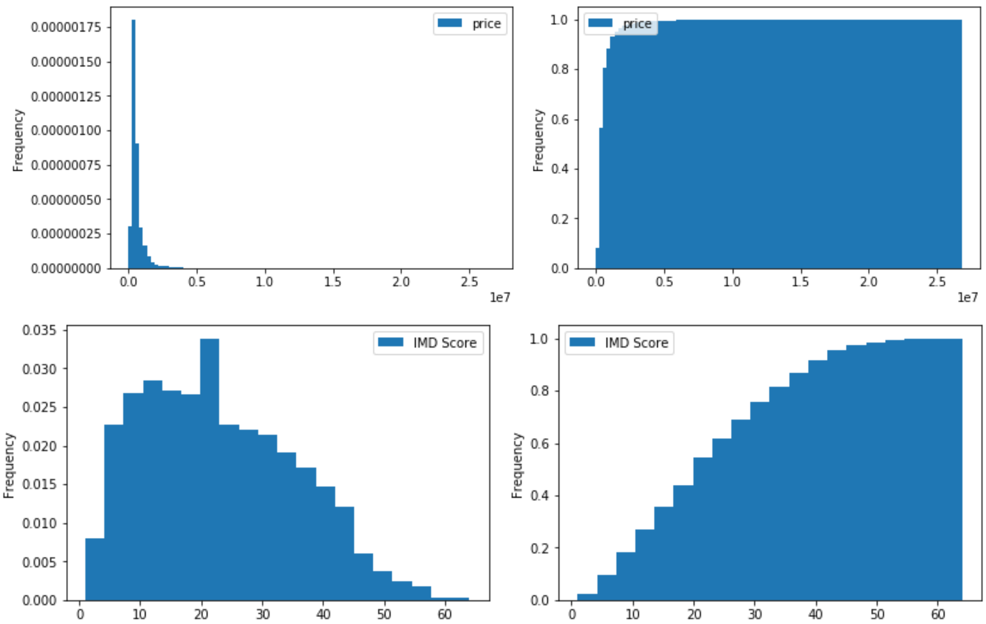
From the perspective of Nottingham rent pricing, frequency distribution indicates that the major of house properties are approximately nearby the 600£. The trend shows that the higher the price, the lower the frequency, with the peak price around 5000£. The distribution feature could also could be draw from the cumulative distribution that is extremely uneven, and more than 80% of the properties are in the interval of 1,200£.

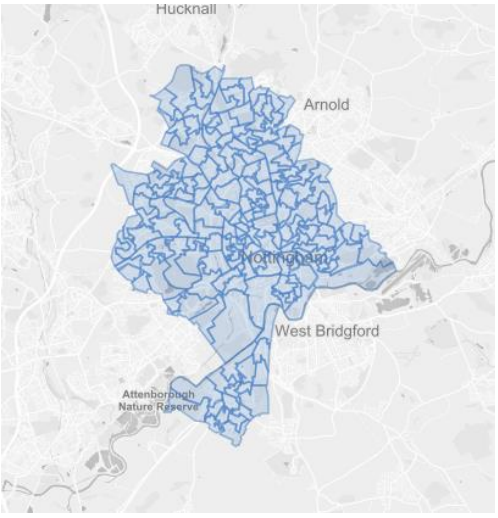
From the perspective of Nottingham regional info, the score distribution in five dimensions is quite from the pricing, which is uneven with value range from 0 to 100. However, the cumulative distribution goes up relatively steadily.

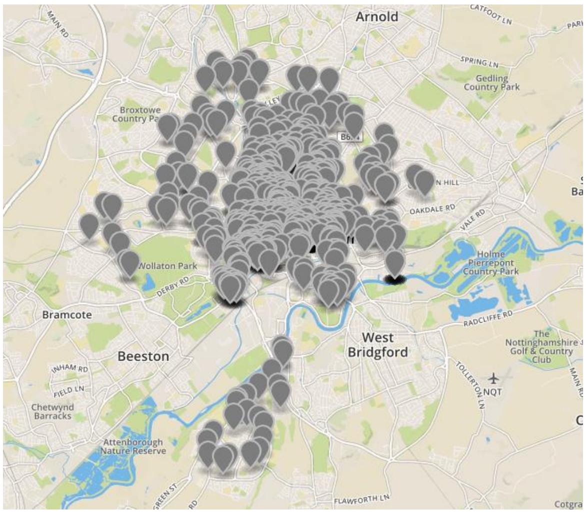




As for the info in London, it is similar to Nottingham. For the sale of information, the difference between the highest value and the lowest value is greater, and the price range is more widely distributed, with dramatically climbing cumulative distribution. The regional information is approximately the same.







3.3.3 Visualisation design

Visual encoding

There are several aspects need to be considered in to visualisation design on map. For example, the visual encoding should allow users to find their information better and faster on real scenario case and provide a to clear vision of information. When see at a glance the metrics of the data in the geographical area, users could understand the feature of the area as well. As for the info in London, it is similar to Nottingham. For the sale of information, the difference between the highest value and the lowest value is greater, and the price range is more widely distributed, with dramatically climbing cumulative distribution. The regional information is approximately the same.

According to the results of the data analysis, visualising the geospatial data take a large part in the project.

First consideration is that the visualisation of point data. Visualising a single data point is fundamental, with the geographic location of a property as the basis for the point, using colour and radius to represent the attributes. Based on a single data point, a map style of point density can be used to represent regional features. At the same time, since the dot density map is effective, it is usually used for volume or mode in other geographical phenomena. Displaying the condition and density of an area is simple but effective and can quickly bring visibility. The dot density map relies on the clustering method, so the appropriate values must be determined when segmenting the data. The Jenks optimization method [55] is often used to calculate and combine the best values to switch colours or proportions. In the dot density map, a region having more dots indicates a high density value, while a region having fewer dots indicates a lower concentration value. At the same time, the mean can be visualized based on the data of the points in the area. As for the map style of dot density, consider using hexagonal binning (hexbin), grid, heatmap, etc., which are not bound by the boundary. For example, hexbin can cluster results in a structured form. A scatter plot of several points, converted to a hexagonal fit. At the same time, the coverage of the area can be fitted according to the coverage of the shape.

Also for visualisation of regional data Choropleth method could be used based on statistics from pre-defined regions. Choropleth is a theme map style that can be mapped as a polygon for regions with well-defined boundaries. Depending on the intensity of the data displayed on the map, the corresponding appears on the shadow colour or pattern texture. In general, hierarchical charts can represent two types of data with the breadth of space, and the intensity of space.

Colour

In designing a Choropleth, hexbin, heat map, even the single dot, usually have two requirements. First, the darker colour values are higher, and secondly, the human eye can only easily distinguish the limited colours. There are several different types of colours that can be selected.

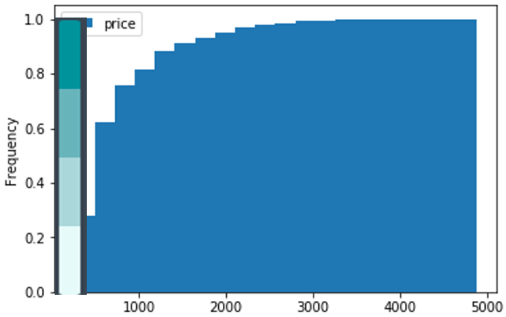
Single hue progression: The colour fades from the dark colour of the selected colour to the light or white of the same colour. The darkest colour represents the largest number in the dataset, and the lightest shade represents the smallest number. 

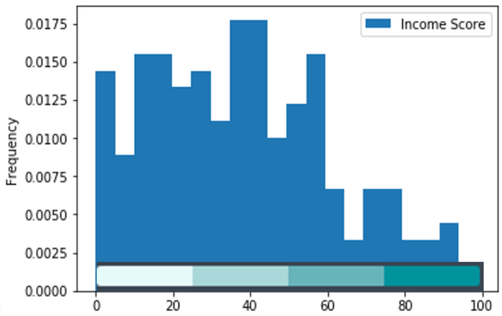
Bipolar progression: Two opposite tones are typically used to show a change from negative to centre to positive. These types of maps show the size of the values associated with each other. 

Partial spectral tonal gradient: used to blend and map two different sets of data. This technique combines the adjacent two opponent tones and shows the size of the mixed data category. Partial spectral tone gradient

The variables would be divide into groups by the steps of colour. In order to separate data with better distinguishing. The progression would cover separated groups evenly.

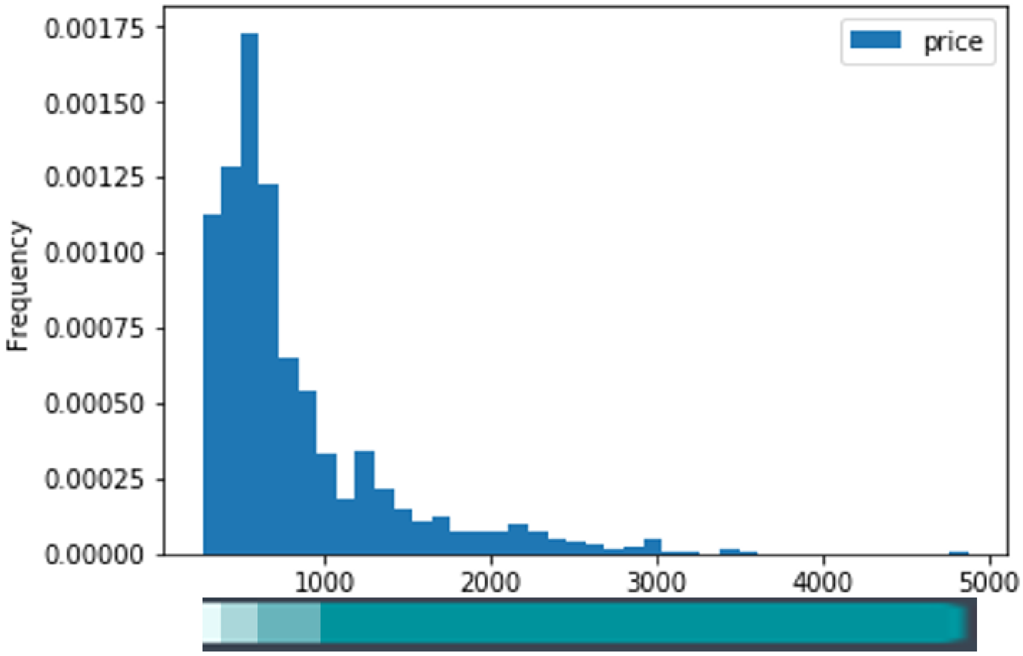
For the variables of pricing, the group would be divided in quantile (even in proportion). For the variables of regional scores, the group would be divided in quantize (even in score sequence).



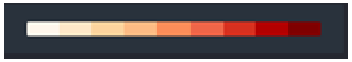
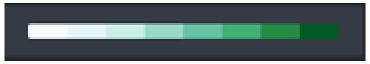
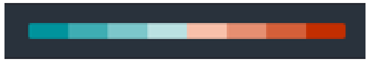
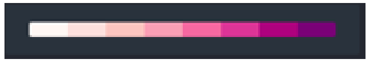


An advanced visual encoding

During the design, an early version prototype, which use only colour in 6-step-single hue progression, is evaluated by a potential user in a dirty way. The feedback shows that it need to increase the colour steps to encoding large properties. Another issue is that when viewing the highest group (in darkest colour), it is hard to distinguish the price as there are too many dots or hexbin in same hue. Therefore, a dual encoding, height, which is based on price is applied in further prototype.



Since height is a 3D variables, the interaction method would change from traditional 2 D view method. Under the 3d condition of the system, whether the user can adapt to the visualisation, whether it is better or faster to complete the cognitive search task is still unknown. So the project studied the usability performance under the 2d and 3d visual encoding. Please see the part of Evaluation.

Geospatial Visualisation					Layer		
Variables	Visual encoding			Comments	2D		3D
House Pricing							
Pricing	Single hue progression	9-steps-quantile		Coverage radius manually adjusted by show and hide button in various zooming scope	2D-hexbin	2D-cluster	3D-stack
		Stack - Height			/	/	3D-stack
		Cluster - Radius			Cluster	Cluster	/
Regional information							
Environment score	Single hue progression	8-steps-quantize		Variables show as one layer. Using hide and show button to select.	2D Polygons according to LSOA boundaries		
Crime score	Bipolar progression	8-steps-quantize					
Income score	Single hue progression	8-steps-quantize					
Education score	Single hue progression	8-steps-quantize					
Health score	Single hue progression	8-steps-quantize					

3.3.4 Interaction design

There are mainly three types of interactions, which is based on User Intent, applied in the project.

1. Selecting

Users could mark the interacting object as they wish, and see learn more info about it. It’s is usually done by mouse select and the attributes would show after interaction. In house property, user could check the pricing and the feature which is texted by description. In regional info, user could see the name of region and other detailed data.

2. Explore – Zooming and shifting

It enable users to examine a different aspect of data with other displayed method. Users are always interested in exploring on digital map, and find the objects as they with. The view range could be adjusted easily. Zooming and shifting, especially in geospatial, take an important role in interaction as it is the main method for searching and focusing. Adjusting the zoom level within a finite-size window directly determines the amount of information that can be visualized.

3. Reconfigure and encode

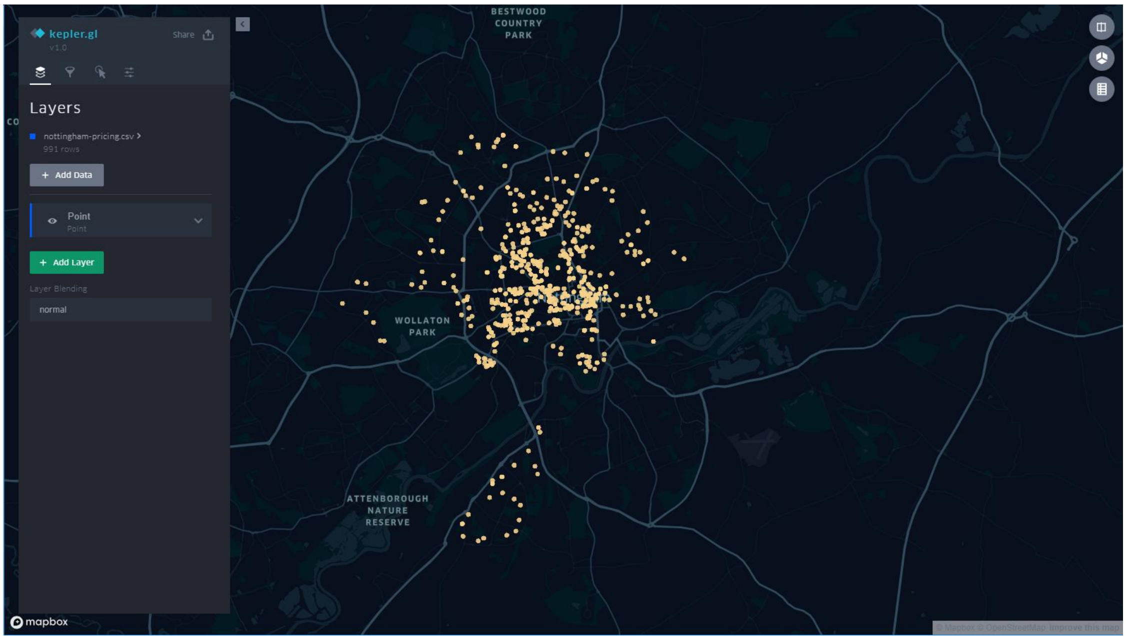
Reconfigure provide different perspective of same dataset according different task requirement by sorting and rearrange visual encoding method, which is determined by users. For example, the visual encoding and the size of point forming cluster could be applied in various manner as to reveal more potential information.

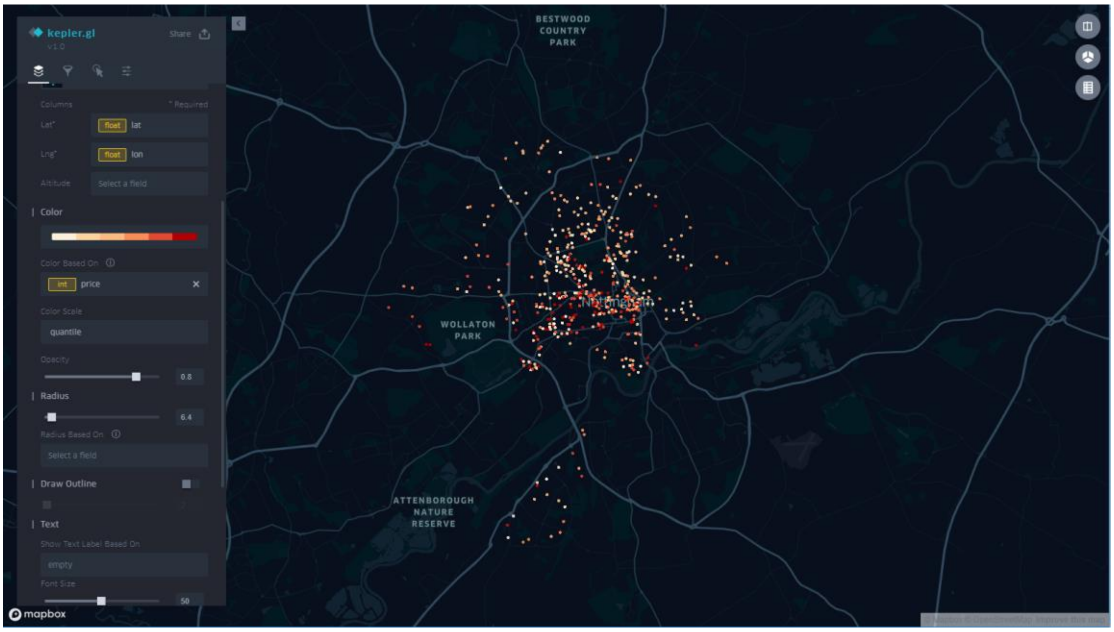
3.3.5 Implement

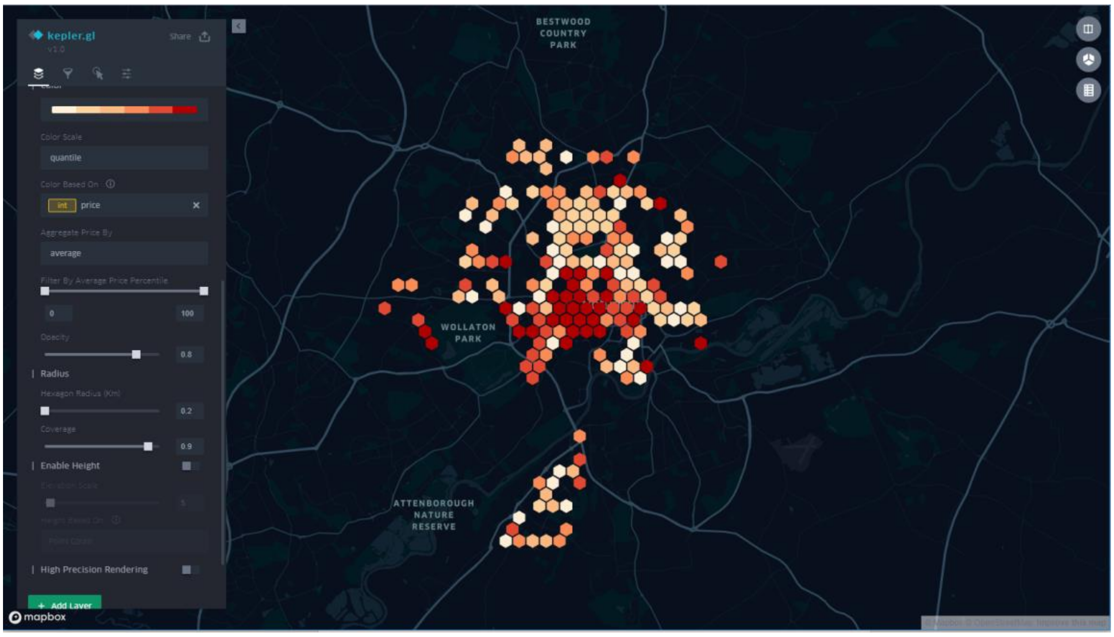
Regards of the above consideration, the Kepler.gl provide an efficient toolkit to implement. Take developing prototype I as example.

Layers

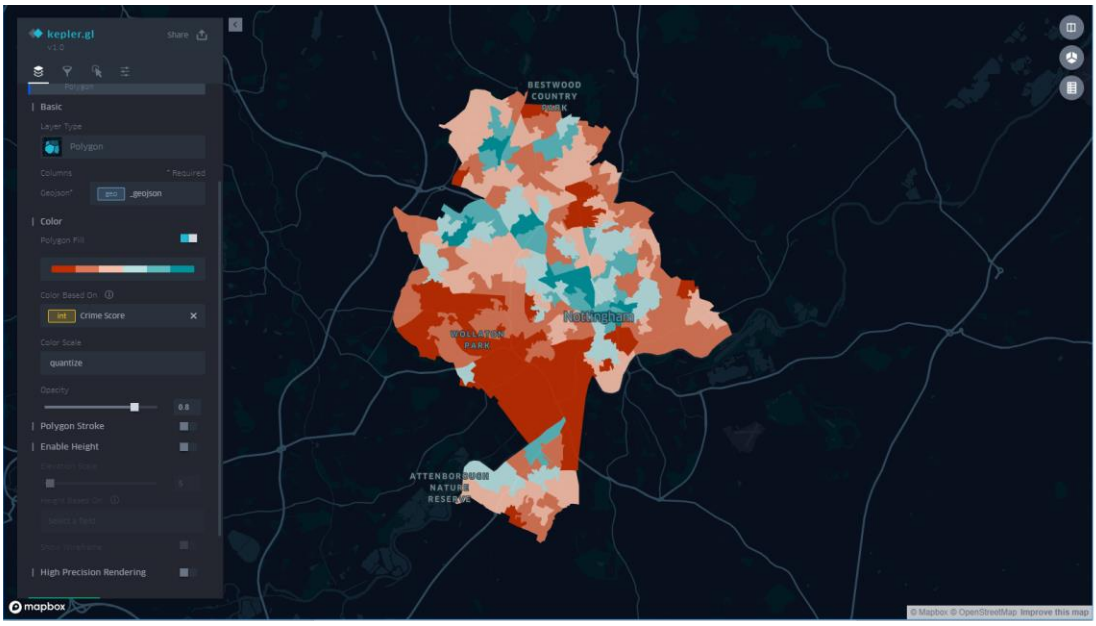
The visualisation is shown as layer on the original map. For each layer, they could be edited and interacted individually. To create layer, the first step is import data. It initially show as location or boundaries. Then adjust the configure according to visual design and dataset. After that, create next layer and repeat these process. The top and bottom of the layer need to be set to avoid occlusion.





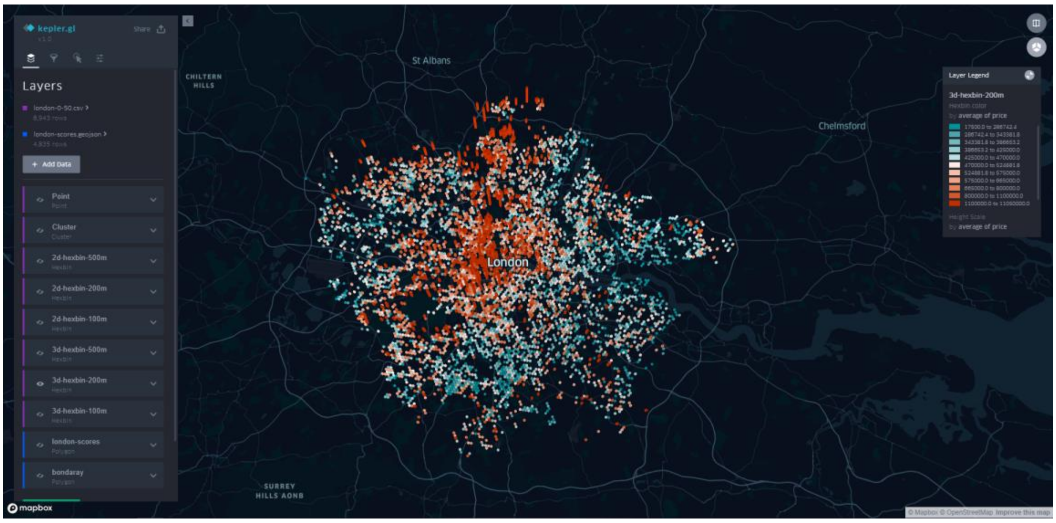

kepler.gl
Share
Color Scale
quantile
Color Based On
int price
Aggregate Price By
average
Opacity
Radius
Coverage
Enable Height
High Precision Rendering
+ Add Layer
BESTWOOD
COUNTRY
PARK
WOLLATON
PARK
ATTENBOROUGH
NATURE
RESERVE
mapbox


kepler.gl
Share
Layers
new dataset
+ Add Data
new layer
+ Add Layer
normal
BESTWOOD
COUNTRY
PARK
Nottingham
WOLLATON
PARK
ATTENBOROUGH
NATURE
RESERVE
mapbox
© Mapbox © OpenStreetMap Improve this map

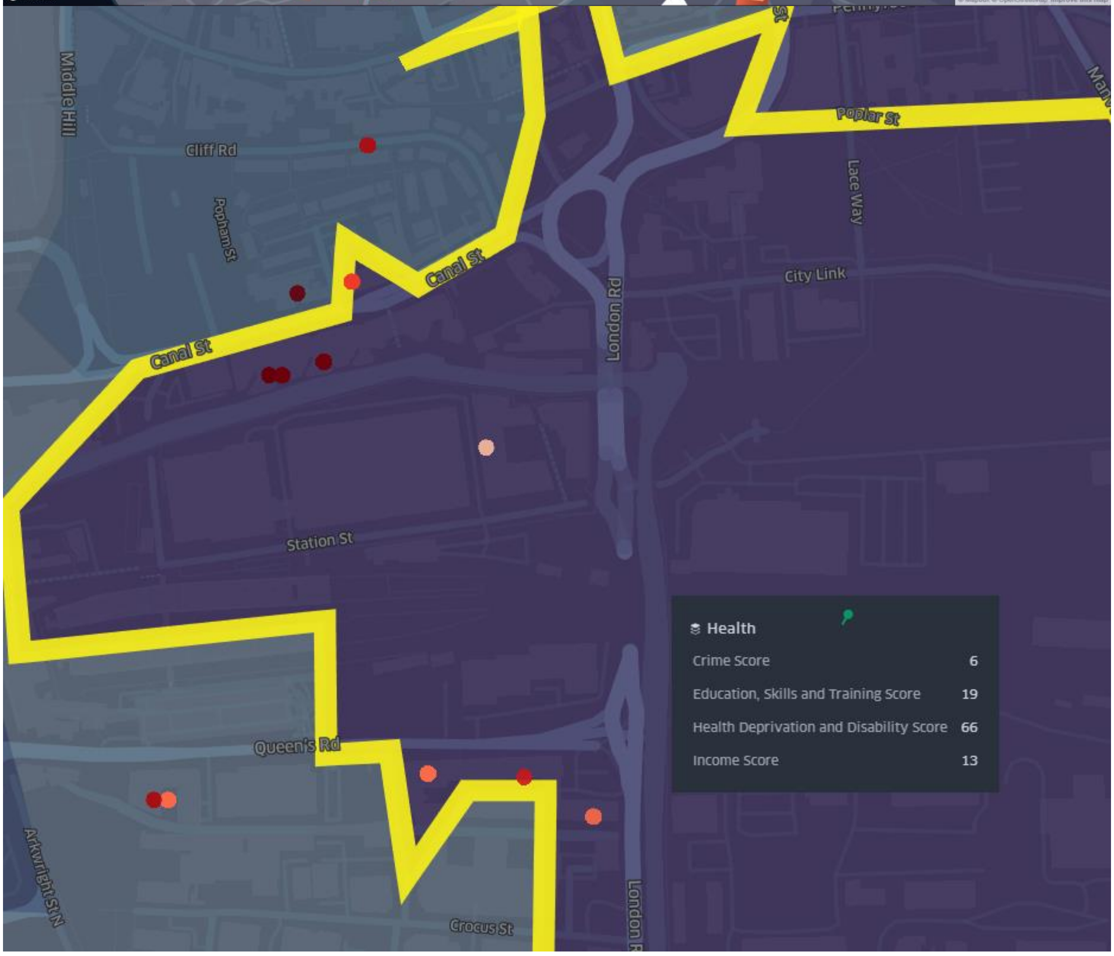


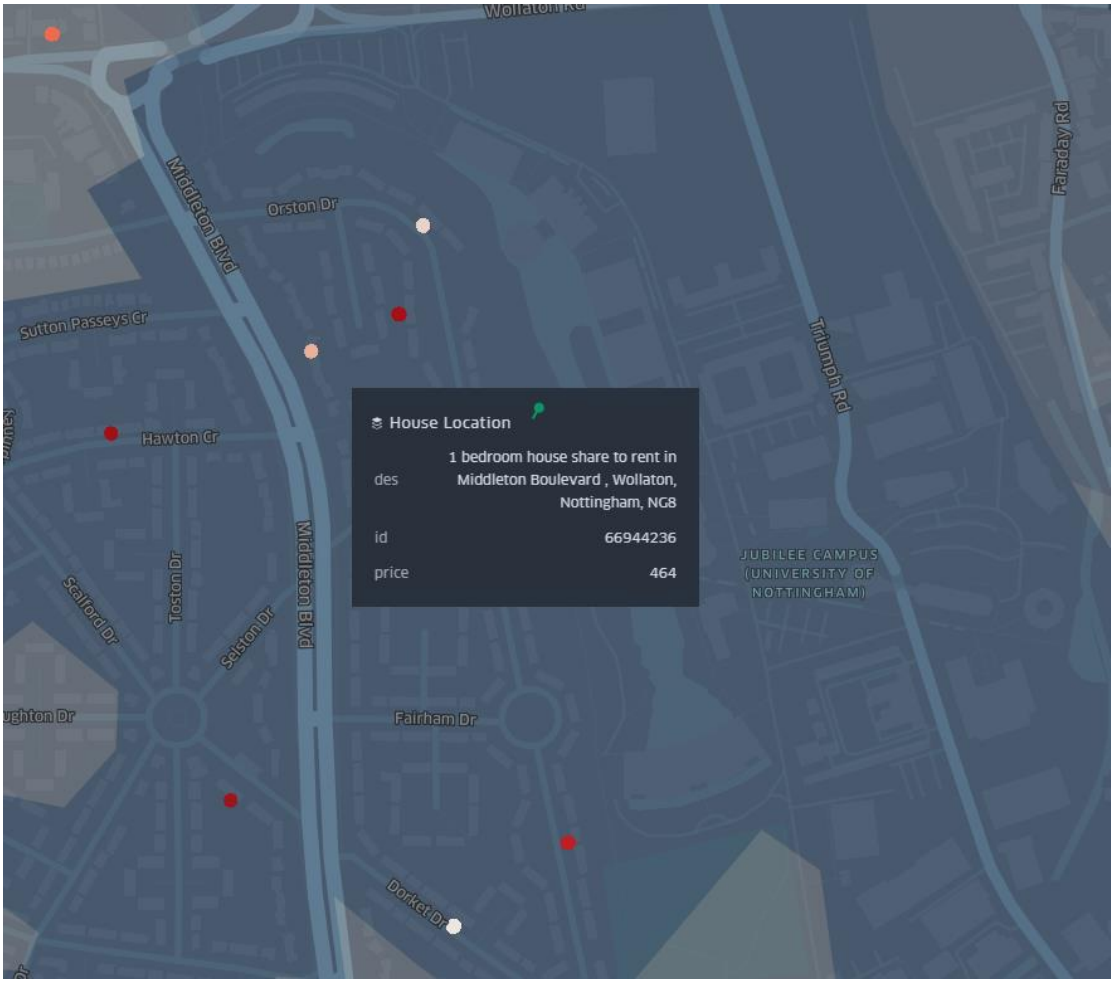
Interaction

Visual interactive functions are implemented by the functional components of Kepler. Although they have certain limitations, they generally meet the visual interaction requirements. In addition to the basic mouse scroll and other operations, the tooltip and the layer show and hide button are also provided in the framework. On the left side of the layout, as the layer area, the user can hide the layers according to the layer name and the corresponding button, and the main window in the middle of the layout. The mouse will go through the corresponding area and highlight the object. At the same time, the specific data information will be displayed in the tooltip, or user can click the fixed tooltip. There are three buttons on the right side of the layout, one is to show legend, which displays the distribution range, and the other two are switched between 3d view and between the dual map view mode.




Nottingham
Thorneywood
BROXTOWE COUNTRY PARK
PHOENIX PARK
mapbox


Middle Hill
Cliff Rd
Popham St
Canal St
Canal St
London Rd
Poplar St
Lace Way
City Link
Station St
Queen's Rd
Arkwright St N
Crocus St
London Rd
Health
Crime Score 6
Education, Skills and Training Score 19
Health Deprivation and Disability Score 66
Income Score 13



Interaction	Mouse Action
Zooming the map view	Scroll
Shift the map view	Left click on and drag
Rotate / Adjust the angle of view	Right click on and drag *3D map only
Show the corresponding layers	Click the ‘eye’ icon as able
Hide the corresponding layers	Click the ‘eye’ icon as disable
Check the object info	Mouse over the object
Mark the object	Mouse over the object and click on
Show 3D map	Enable the 3D map view on the button
Show legend	Enable the legend on the button

The specific interaction method

Local deployment

1.clone repo from

https://github.com/uber/kepler.gl/tree/master/examples/demo-app

2.set package.json & webpack.config.js to adjust local setting

3.install packages by npm command line

4. add mapbox access token to node env which is

<pk.eyJ1IjoieHpyYW5ubm4iLCJhIjoiY2pmaXZxbjIxMGEzaDJ3cThnaTc1a25tNCJ9.Ny6q1ACtjLDEANDbJEIIRw>

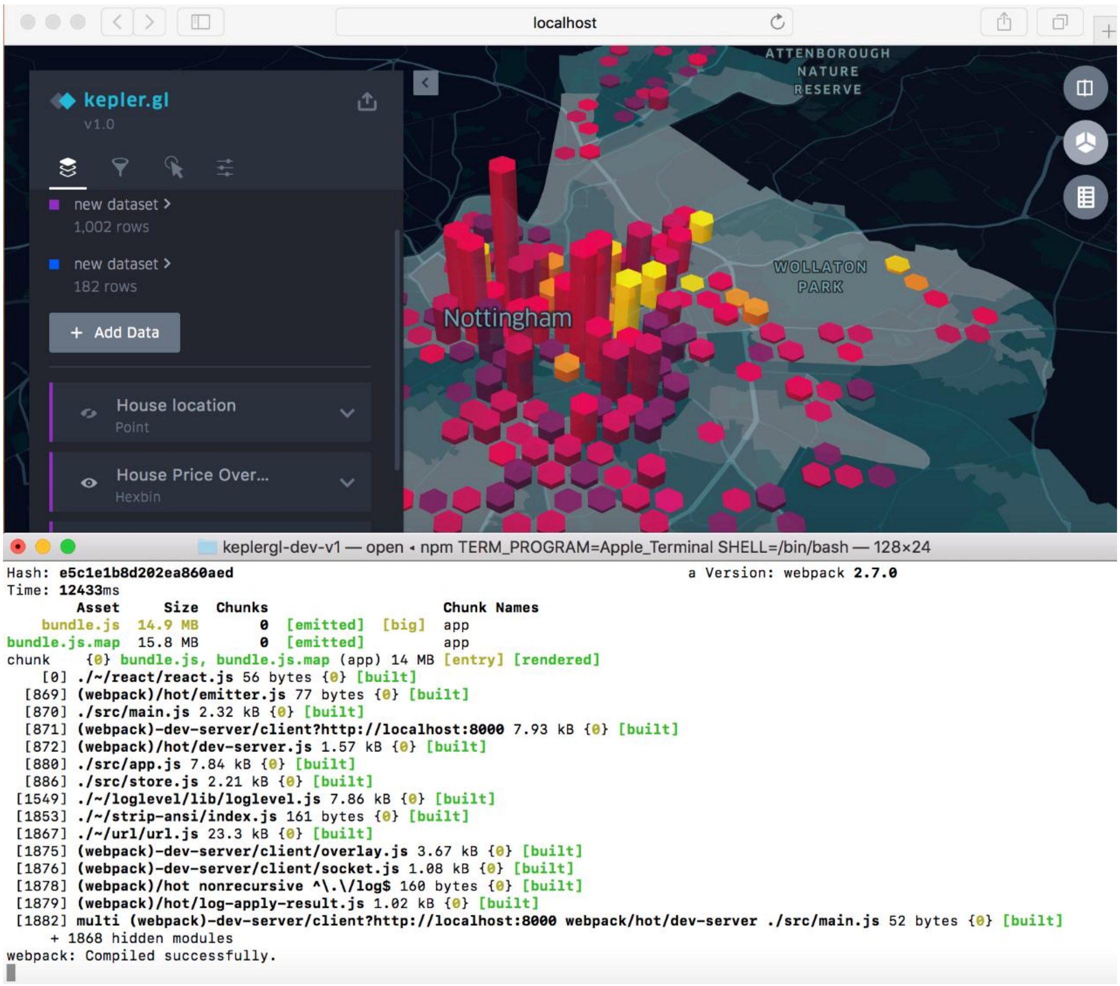
3.recode .src/app.js to import data file & config from local

4.using npm command line to compile

5.After success, the visualisation will run on local sever as a webstie

```
import React, {Component} from 'react';
import {connect} from 'react-redux';
import AutoSizer from 'react-virtualized/dist/commonjs/AutoSizer';
import KeplerGl from 'kepler.gl';

// Kepler.gl actions
import {addDataToMap} from 'kepler.gl/actions';
// Kepler.gl Data processing APIs
import Processors from 'kepler.gl/processors';

// Kepler.gl Schema APIs
import KeplerGlSchema from 'kepler.gl/schemas';

// Component and helpers
import Button from './button';
import downloadJsonFile from "./file-download";

// Sample data
import property from './data/house-property-csv';
import nottinghamConfig from './data/keplergl.json';
import lsoa from './data/lsoa-csv';

const MAPBOX_TOKEN = process.env.MapboxAccessToken; // eslint-disable-line

class App extends Component {

  componentDidMount() {
    // Use processCsvData helper to convert csv file into kepler.gl structure {fields, rows}
    const data_p = Processors.processCsvData(property);
    const data_g = Processors.processCsvData(lsoa);
    var data = data_p
    // Create dataset structure
    const dataset_p = {
      data,
      info: {
        // this is used to match the dataId defined in nyc-config.json. For more details see API documentation.
        // It is paramount that this id matches your configuration otherwise the configuration file will be ignored.
        id: '2z77qxzrv'
```

Import data file & config

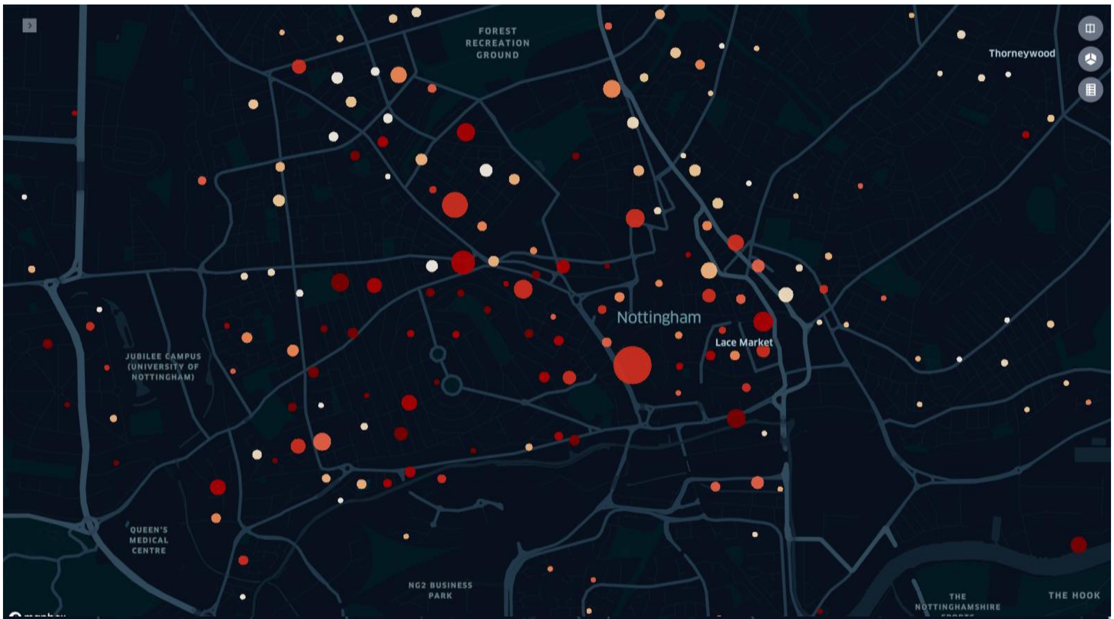


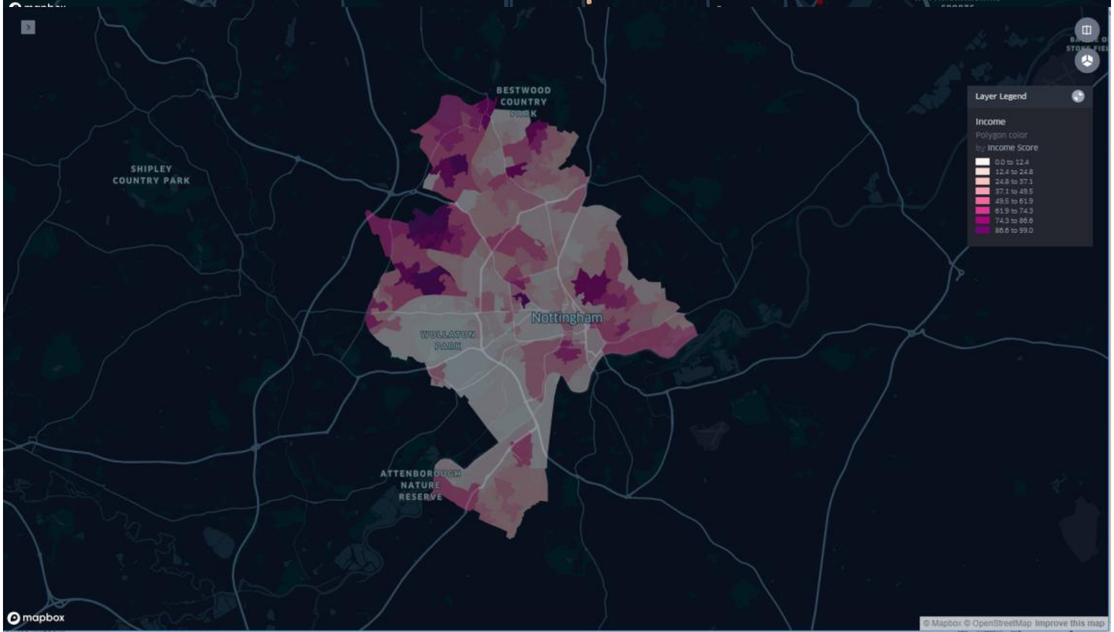
3.4 Prototyping

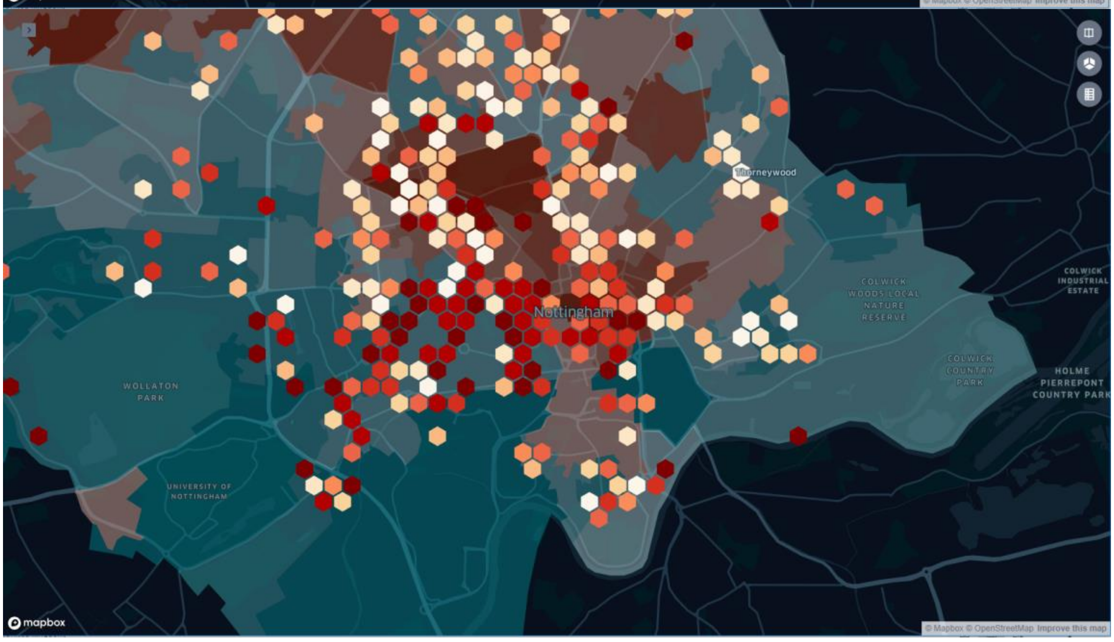
3.4.1 Prototype I – Nottingham

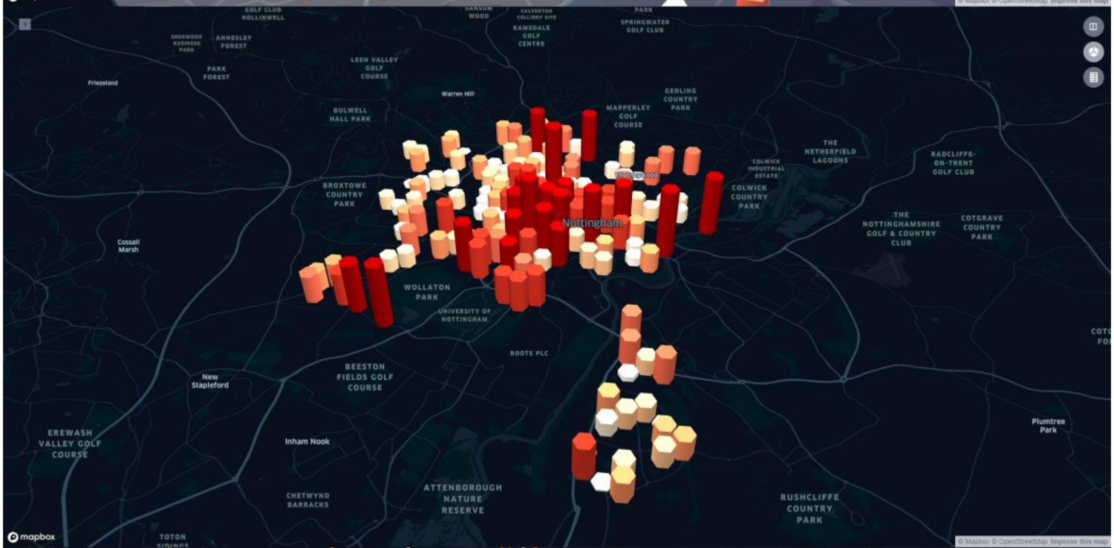
Layer introduction

Layer	Information	Description
House Location	Pricing	Single house property with location and pricing
House Cluster		Clusters of properties, Gather or split by zooming level
2d-hexbin-10m		A block cluster of house properties, with side length 10m hexagon coverage. Topological stitching
2d-hexbin-50m		A block cluster of house properties, with side length 50m hexagon coverage. Topological stitching
2d-hexbin-100m		A block cluster of house properties, with side length 100m hexagon coverage. Topological stitching
2d-hexbin-200m		A block cluster of house properties, with side length 200m hexagon coverage. Topological stitching
3d-hexbin-10m		A block cluster of house properties, with side length 10m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-50m		A block cluster of house properties, with side length 50m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-100m		A block cluster of house properties, with side length 100m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-200m		A block cluster of house properties, with side length 200m hexagon coverage. Topological stitching with 3D stacks.
Crime Score	Regional(LSOA)	Polygon with boundary of Nottingham LSOA, filed by encoding colour.
Income Score		Polygon with boundary of Nottingham LSOA, filed by encoding colour.
Education, Skills and Training Score		Polygon with boundary of Nottingham LSOA, filed by encoding colour.
Health Deprivation and Disability Score		Polygon with boundary of Nottingham LSOA, filed by encoding colour.
Living Environment Score		Polygon with boundary of Nottingham LSOA, filed by encoding colour.
Boundary		White mark line by the boundary of Nottingham LSOA.


FOREST RECREATION GROUND
Thorneywood
Nottingham
Lace Market
JUBILEE CAMPUS (UNIVERSITY OF NOTTINGHAM)
QUEEN'S MEDICAL CENTRE
NG2 BUSINESS PARK
THE NOTTINGHAMSHIRE
THE HOOK


BESTWOOD COUNTRY PARK
SHIPLEY COUNTRY PARK
Layer Legend
Income
Polygon color
by Income Score
0.0 to 12.4
12.4 to 24.8
24.8 to 37.1
37.1 to 49.5
49.5 to 61.9
61.9 to 74.3
74.3 to 86.6
86.6 to 99.0
Nottingham
WOLLATON PARK
ATTENBOROUGH NATURE RESERVE
mapbox
© Mapbox © OpenStreetMap Improve this map


Thorneywood
Nottingham
COLWICK WOODS LOCAL NATURE RESERVE
COLWICK INDUSTRIAL ESTATE
COLWICK COUNTRY PARK
HOLME PIERREPONT COUNTRY PARK
WOLLATON PARK
UNIVERSITY OF NOTTINGHAM
mapbox
© Mapbox © OpenStreetMap Improve this map


Nottingham
Hexbin-100m
total points
average of price
1666.142857142857
THE HOOK
Thorneywood
BROXTOWE COUNTRY PARK
PHOENIX BUSINESS PARK
PHOENIX PARK
mapbox

Nottingham
WOLLATON PARK
BEESTON FIELDS GOLF COURSE
ATTENBOROUGH NATURE RESERVE
RUSHCLIFFE COUNTRY PARK
EREWASH VALLEY GOLF COURSE
Inham Nook
New Stapleford
mapbox

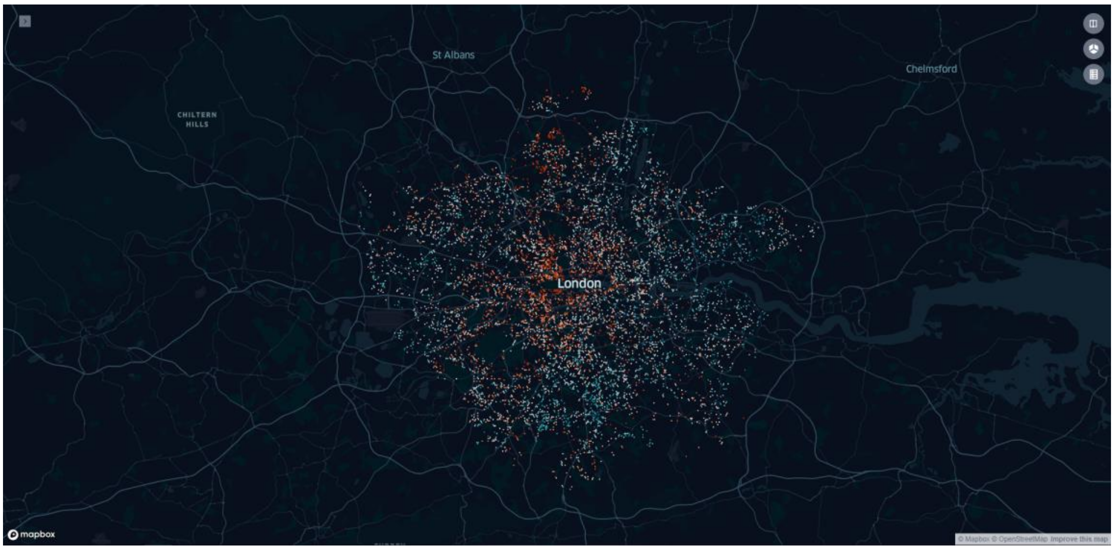
Lace Market
FOREST RECREATION GROUND
THE ARBORETUM
NOTTINGHAM TRENT UNIVERSITY
JUBILEE CAMPUS (UNIVERSITY OF NOTTINGHAM)
mapbox

3.4.2 Prototype II – London

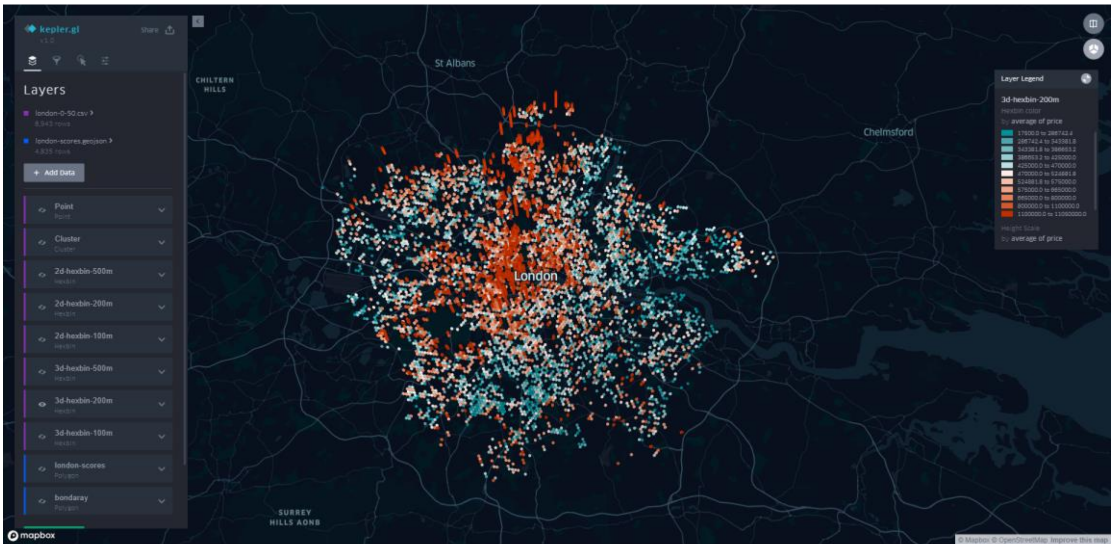
Prototype II is migrated by Prototype with a large data set of London. There is only one regional info variable, IMB score. And 11-step- Bipolar progression encode the pricing as greater interval.

Layer introduction

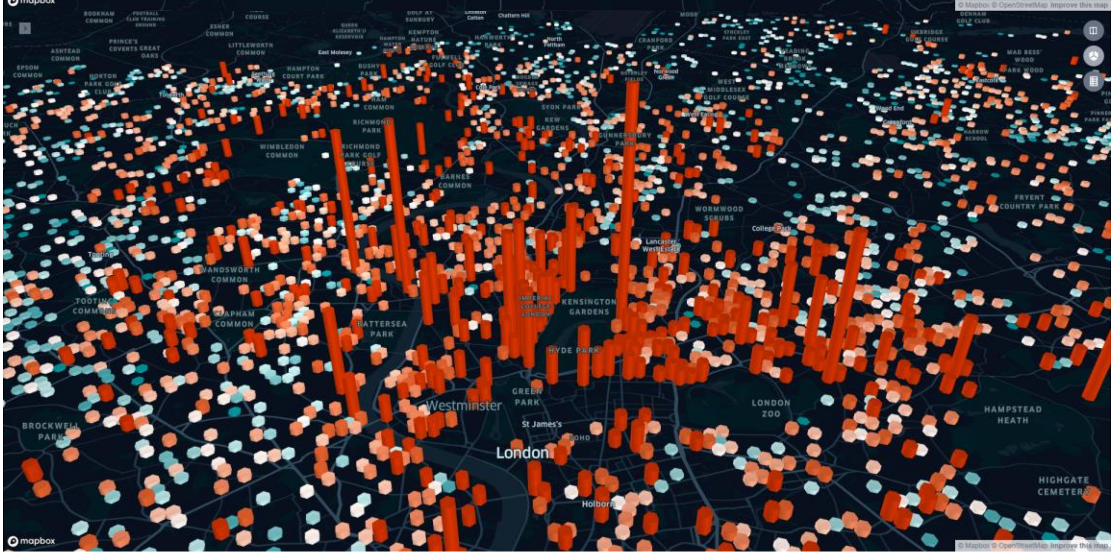
Layer	Information	Description
House Location	Pricing	Single house property with location and pricing
House Cluster		Clusters of properties, Gather or split by zooming level
2d-hexbin-500m		A block cluster of house properties, with side length 500m hexagon coverage. Topological stitching
2d-hexbin-200m		A block cluster of house properties, with side length 200m hexagon coverage. Topological stitching
2d-hexbin-100m		A block cluster of house properties, with side length 100m hexagon coverage. Topological stitching
2d-hexbin-50m		A block cluster of house properties, with side length 50m hexagon coverage. Topological stitching
3d-hexbin-500m		A block cluster of house properties, with side length 500m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-200m		A block cluster of house properties, with side length 200m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-100m		A block cluster of house properties, with side length 100m hexagon coverage. Topological stitching with 3D stacks.
3d-hexbin-50m		A block cluster of house properties, with side length 50m hexagon coverage. Topological stitching with 3D stacks.
London IMD Score	Regional(LSOA)	Polygon with boundary of London LSOA, filed by encoding colour.
Boundary		White mark line by the boundary of London LSOA.



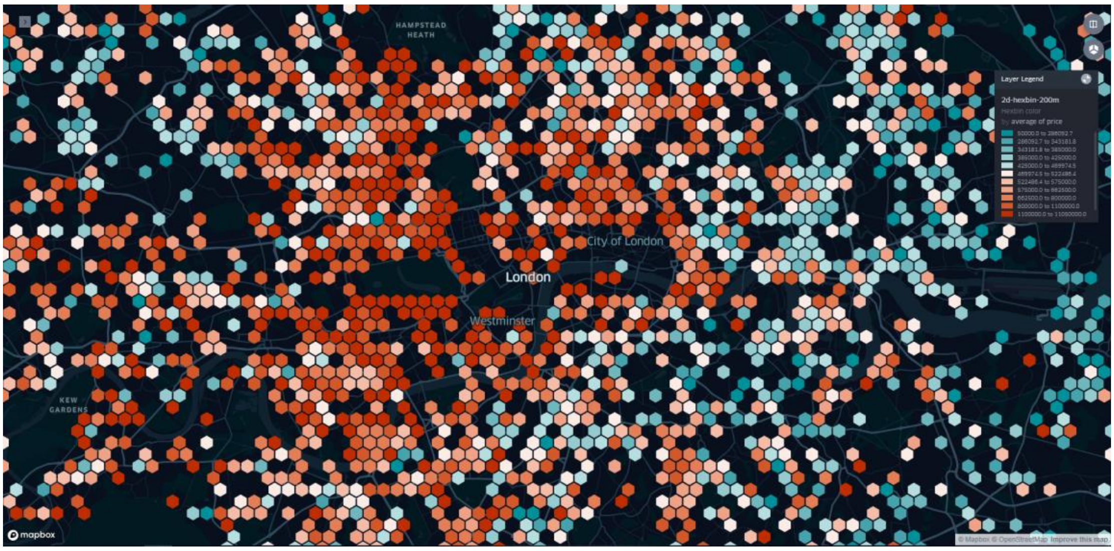
St Albans
Chelmsford
CHILTERN HILLS
London
mapbox



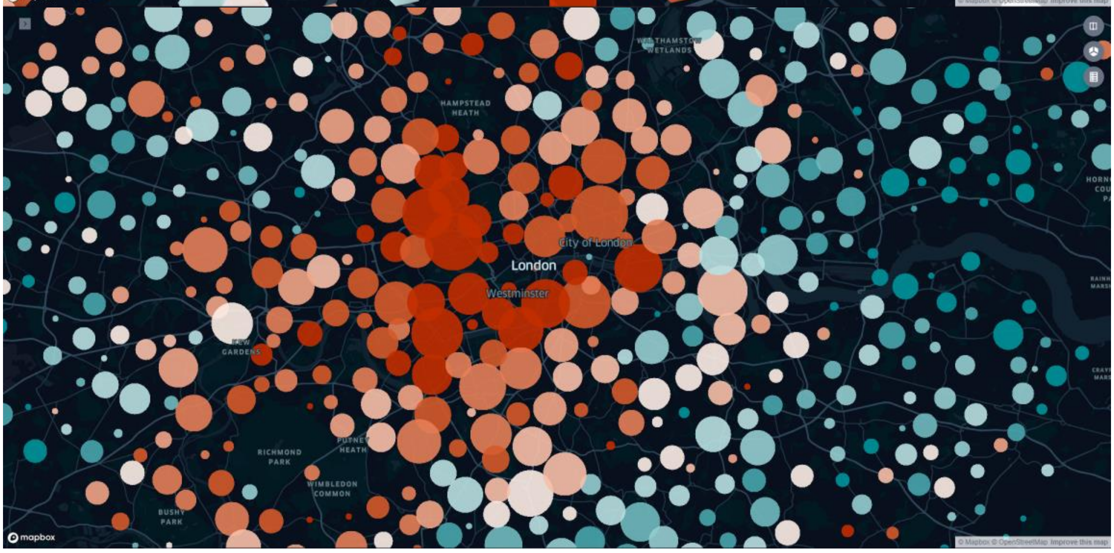
kepler.gl
Share
Layers
london-0-50.csv
london-scores.geojson
Add Data
Point
Cluster
2d-hexbin-500m
2d-hexbin-200m
2d-hexbin-100m
3d-hexbin-500m
3d-hexbin-200m
3d-hexbin-100m
london-scores
bondaray
St Albans
CHILTERN HILLS
Chelmsford
London
SURREY HILLS AONB
Layer Legend
3d-hexbin-200m
by average of price
mapbox



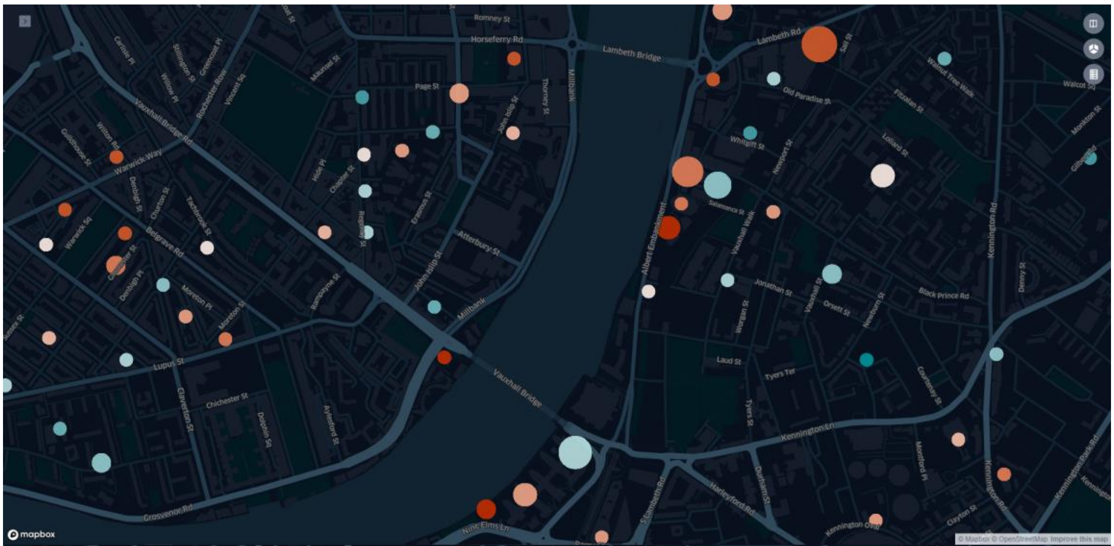
WIMBLEDON COMMON
RICHMOND PARK
WANDSWORTH COMMON
CLAPHAM COMMON
BATTERSEA PARK
KENSINGTON GARDENS
HYDE PARK
GREEN PARK
Westminster
St James's
London
Holborn
LONDON ZOO
HAMPSTEAD HEATH
HIGHGATE CEMETERY
BROCKWELL PARK
mapbox

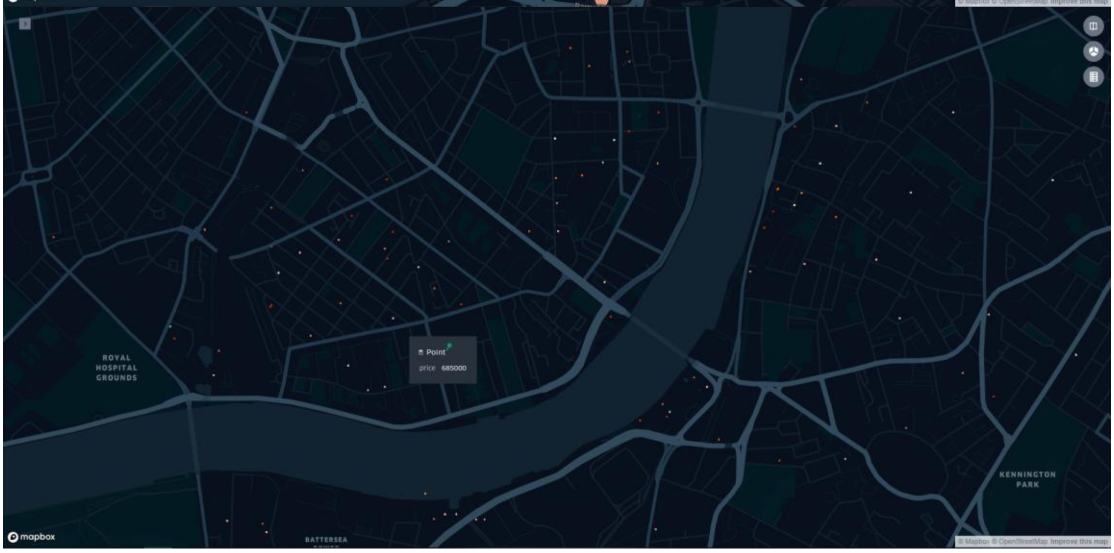

HAMPSTEAD HEATH
Layer Legend
2d-hexbin-200m
average of price
City of London
London
Westminster
KEW GARDENS
mapbox

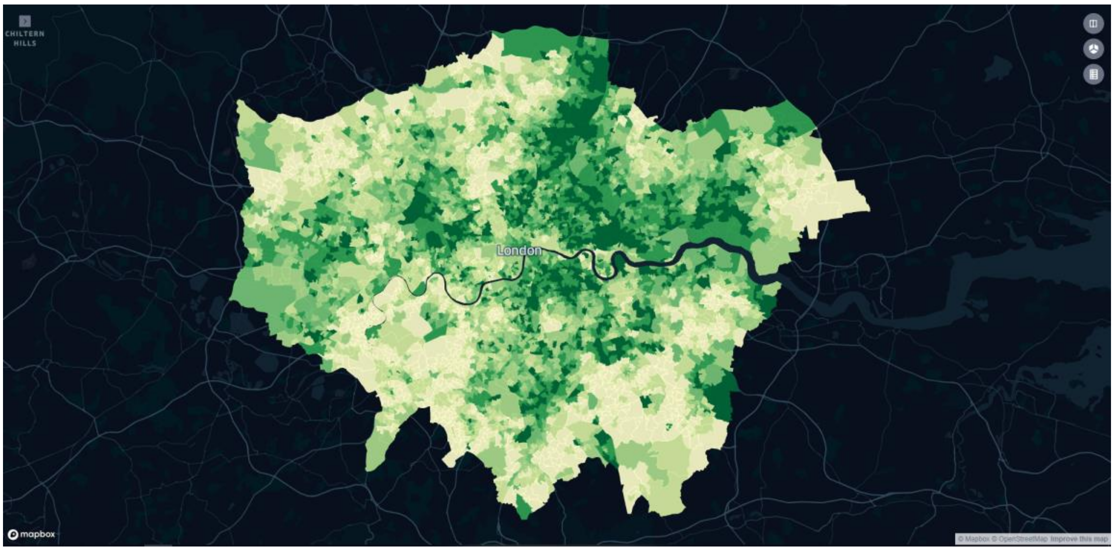

3d-Hexbin-100m
total points
1
average of price
5500000
average of price
5500000
Holborn
St Clement Danes
St James's
London
ST. JAMES'S PARK
City of London
Layer Legend
3d-hexbin-100m
average of price
mapbox


HAMPSTEAD HEATH
City of London
London
Westminster
RICHMOND PARK
PUTNEY HEATH
WIMBLEDON COMMON
BUSHY PARK
mapbox


KENSINGTON GARDENS
HYDE PARK
SOHO
Holborn
St Clement Danes
City of London
St James's
London
GREEN PARK
ST. JAMES'S PARK
GARDEN AT BUCKINGHAM PALACE
IMPERIAL COLLEGE LONDON
Westminster
BURGESS PARK
mapbox


Lambeth Bridge
Lambeth Rd
Vauxhall Bridge
Lupus St
Kennington Ln
mapbox


ROYAL HOSPITAL GROUNDS
Point
price 685000
KENNINGTON PARK
mapbox


London


FOREST
RECREATION
GROUND
THE
ARBORETUM
Health
Crime Score 33
Education, Skills and Training Score 1
Health Deprivation and Disability Score 6
Income Score 0
Alfreton Rd

4. EVALUATION

The aim of evaluation is to guide the design of the information visualisation system prototypes during the process. It is conducted by subjective assessment and controlled experiment from five aspects of usability. (Learnability, efficiency, memorability, error, satisfaction, Nielson 1993). Due to correlation between error and efficiency, they would be combined in this experiment. Subjective assessments are conducted by questionnaires. Controlled experiments focus on efficiency by task-oriented usability testing and compare the performance between two versions of visual encodings, 2D and 3D, which is based on Prototype I.

4.1 Methodology

The aim of evaluation is to guide the design of the information visualisation system prototypes during the process. It is conducted by subjective assessment and controlled experiment from five aspects of usability. (Learnability, efficiency, memorability, error, satisfaction, Nielson 1993). Due to correlation between error and efficiency, they would be combined in this experiment. Subjective assessments are conducted by questionnaires. Controlled experiments focus on efficiency by task-oriented usability testing and compare the performance between two versions of visual encodings, 2D and 3D, which is based on Prototype I.Figures and tables should to be integrated into the text and <u>not</u> put at the end of the paper. Boxes around the figures are optional. If there are many figures, drawings etc that need documentation, then these can go in the appendices.

Figures should be suitable for reproduction in black and white, ie. colour images may be used but be aware that they may be photocopied. You must ensure that colour shadings also make sense in black and white!

4.1.1 Ethic

All work in this study was approved and authorized by the University of Nottingham College Research Ethics Committee (REC).

4.1.2 Controlled experiments

There is a spilt version from prototype I, in which house pricing is in two types of visual encodings, 2d and 3d. The goal of experiments is to study whether is there is performance difference between them.

Hypothesis

H0: There is no difference in reported comfort between the 2d and 3d visual encoding versions.

H1: There is a difference in reported comfort between the 2d and 3d visual encoding versions.

Participants

Participants are students recruited from University of Nottingham, aged from 23 to 31, with experience in living in Nottingham Before the start of the study, the content of the experimental test and the purpose of the study were informed. After obtaining their consent, they clearly indicates that their participation was completely voluntary and signed the consent form. All participants in this study are Chinese.There are 10 participants in total. They are group into two groups and each group do test on only one version. (Within subjects)

IV: Visual encodings - 2 levels (2d and 3d)

DV: Task performance mainly based time cost

Experiment environment

The experiment location are the lab at computer science school. The system is deployed on the Windows10 with Google Chrome version 69.0.3.

Experimental process

1. Inform the goal and the background of study.

2. Ask participants to sign the consent form.

3. Introduce the system and explain the experiment to participants.

4. To avoid the potential bias caused by the familiarity, the participants would familiarize with the system at first, learn the basic operation of the interaction, and adapt to the geographical mapping at beginning.

5. Ask participants to complete test tasks on Prototype I.

6. Ask participants to experience Prototype II.

Task No.	Task Description	Source Requirement No.
A	Search regional information (A1) and house property info (A2) based on a given location. (Find location and Obtaining data)	1, 2, 3, 4, 5
B	Search house property info (B2) based on given condition (B1). (A task include map exploration)	1, 2, 3, 4, 5
C	Find the area with highest average pricing.(A broad view of perceiving)	4, 6

Task No.	Procedure	Goal
A1	1.Give participants a certain location (E.g. UoN Jubilee Campus) 2. Let participants find the location and select the LSOA on map. 3. Using tooltip to find the specific figure on tag.	The regional data
A2	1. Find and check the house properties inside the boundary of LSOA. 2. Find the certain range pricing properties with a given figure.	House properties
B1	Find the region which has highest score in aspect of crime.	The LSOA
B2	1. Find and comparing the house properties inside the boundary of LSOA. 2. Find the highest pricing properties and give the pricing figure.	House property with pricing
C	1. Compare and find the highest average pricing area which is a hexbin. (E.g. with coverage 0.1 km) 2. Using tooltip to find the average pricing figure on tag.	The average price

Measurement

Performance is measured from three dimensions as follow:

	Dimensions	Metrics	Description
1	Speed	Time cost	$t_{cost} = T_{stop} – T_{start}$ Count manually by timer.
2	Accuracy	Bool	If success, P = 1 If fail, P = 0 $R_{a} = P_{sum} / N_{Group}$ Count after finish task.
3	Help times	Number of times	$n_{helptimes}$ Count manually.

4.1.3 Questionnaire

Design

The questionnaire in this study (appendix). It is true that the participant completes the test for prototype I and has done it after experiencing Prototypes.

The questionnaire is divided into five parts. Most of the questions use the standardized scale of the seven-point Likert scale [55], which is designed to examine the reasons for user preferences and their preferences.

The first part focuses on the basic information of users and the experience of using similar systems.

The second part uses USE [56] subjective assessment of the learnability of the system usability and the efficiency of the interaction.

The third part uses QUIS (Questionnaire for User Interaction Satisfaction)

QUIS was created to assess user satisfaction with specific aspects of the human-machine interface. The scale includes several dimensions of software outlook, terminology and system information, learnability, and system performance. [57] This section of the question is designed to assess how easily the subjective evaluation of information visualisation of system experience.

The fourth part establishes a standardized scale that uses seven distributions to allow users to evaluate the visual preferences used in the system.

The fifth part uses CSAT (Customer Satisfaction Score) to evaluate user satisfaction with specific events/experiences, using a seven-point scale. [58] In order to assess the satisfaction of the overall system use experience.

Pilot

After the first draft of questionnaire was completed, two students were given pilots to conduct a preliminary assessment of the content and settings of the questions. They think that the description needs to be modified on section 3, and the question should be set based on more task cases. In addition, on section 4, there is a certain question in the choice of interaction mode. The comparison control of variables should be related to the actual application scenario to evaluate the interaction mode. Another problem is that the questioning perspective of the questionnaire should be based on the user’s prospect, especially for that conduction of questionnaire is after the test. The user's answer is based on the experience, and the question should help to recall the experience.

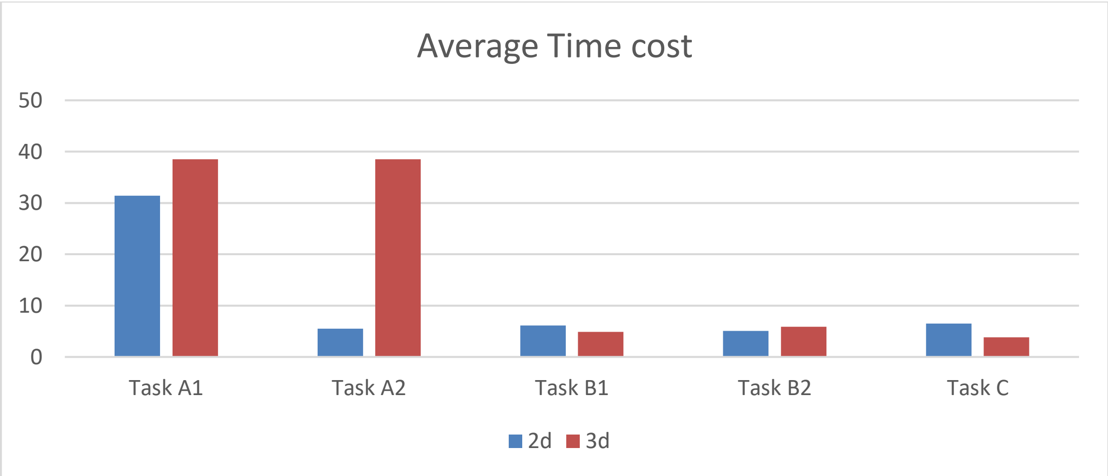
Administration

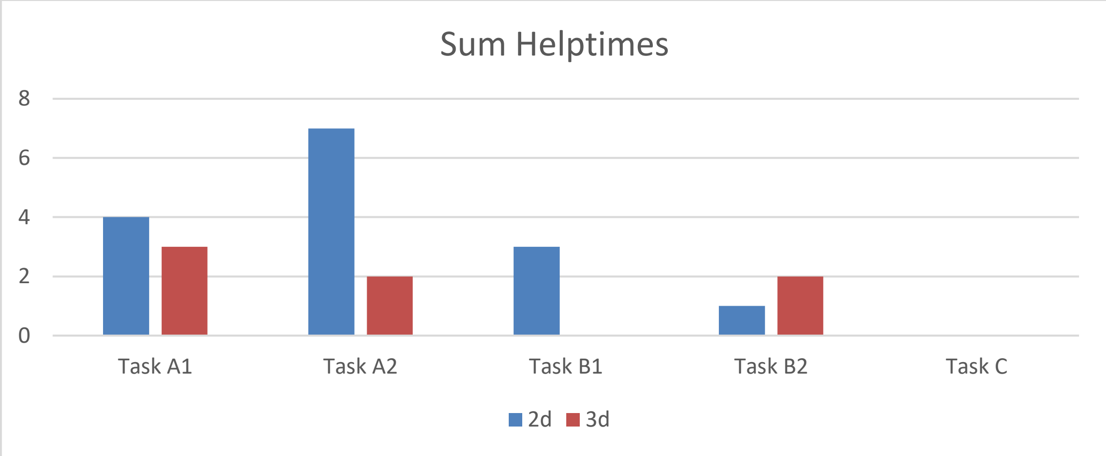
After The administration of questionnaire is deliver after the experimental task of Prototype I and using of Prototype II. There ten participants in total and the method of analysis is collect questionnaire data and transform into digital format, then import to SPSS.

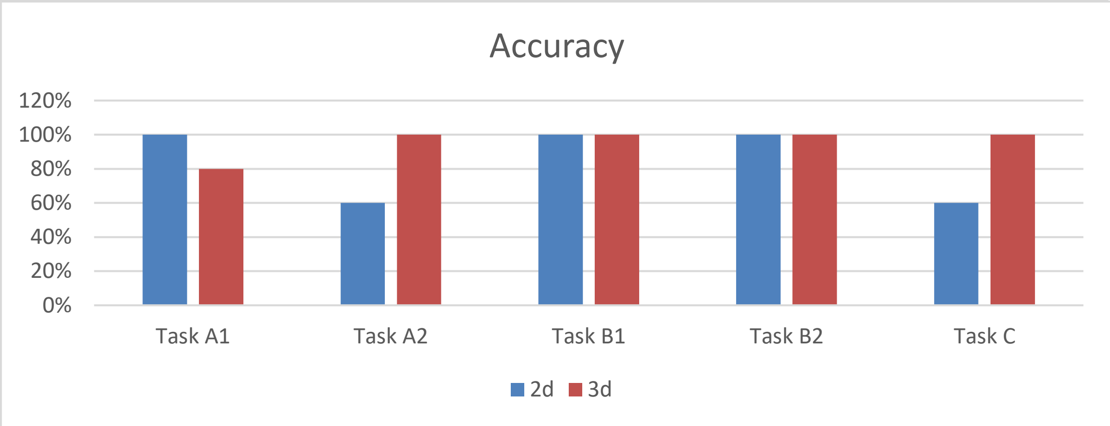
4.2 Result and analysis

3.1.1 Controlled experiments

Regard only 5 participants for each groups, Non-parametric method is choose. There are 2 level independent variables. Therefore, Mann Whitney U is used to analyse tcost (time cost) in SPSS. The statistics of tcost [Appendix] shows that these is no significant difference between the two groups (2d vs. 3d) in all tasks. As for other measurements, the results as follow.







Regarding the results of SPSS, there is no significance between groups, it is hard to tell which version is better. But most of the task can be complete expect five cases. One in A1 3d group, participant do not know the location. However, the other four wrongly perceiving the data.

As far as the experimental process is concerned, in the Task A, more participants have spent a long time familiar with the system. There are also two participants who are not very good at using maps. It takes a long time to find the corresponding location with helps. Before familiarizing tooltip, it took some time to check information and finish the task.

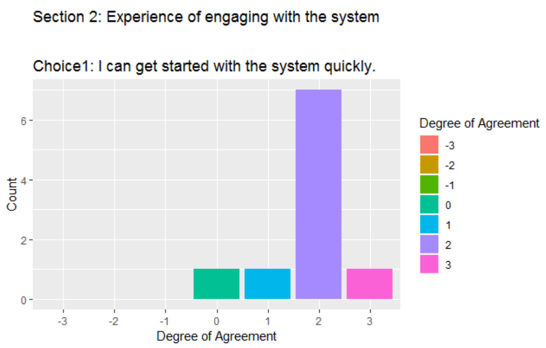
With the experience of Task A, the time cost in Task B is much less. However, web loading usually takes time during the process, which is more obvious in 3d version.

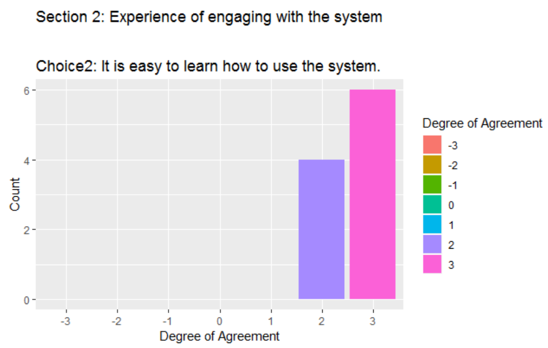
In addition, a large part of the participant is indicated that, in the task C (the area with the highest average price in many areas), the 3d version is easier to find the peak value, and after adding the visual encoding, compared to the design of the single colour, dual encoding could well distinguishes the data. However, there is also a participant, who takes a long time, for sake of viewing angle (minimum 45 degrees) cannot perceive the difference in height, and thus have troubling perceive the value. Another issue is that stacking with height in 3d, the front stacks would cover the back one, and the drag is not as smooth as 2d.

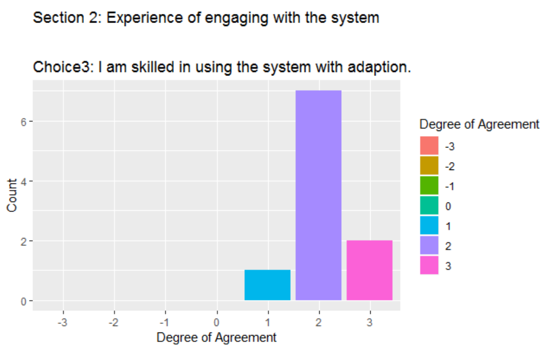
The main limitation of controlled experiment as follows. First, the number of participants is too small, thus with limited sample, which may cause large bias. Second, the timing method relies on manual timers and the boundaries between the start and end of the task are not clear enough, which will bring decrease the reliability of data. The third point is the instability of the system platform. Since the front-end platform rendering data occasionally gets stuck, such systematic deviations cannot be controlled when the sample size is small. Also, the influence of irrelevant variables, such as the cognitive ability of the subjects, the familiarity of use, etc., and these variables will greatly influence the results of the data in such size of participants.

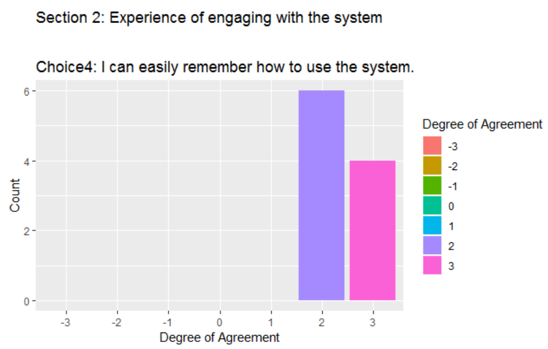
3.1.2 Questionnaire

The first question show that all the participants are students with experiencing of digital map. They age from 24 to 31, with 3 males and 7 females..



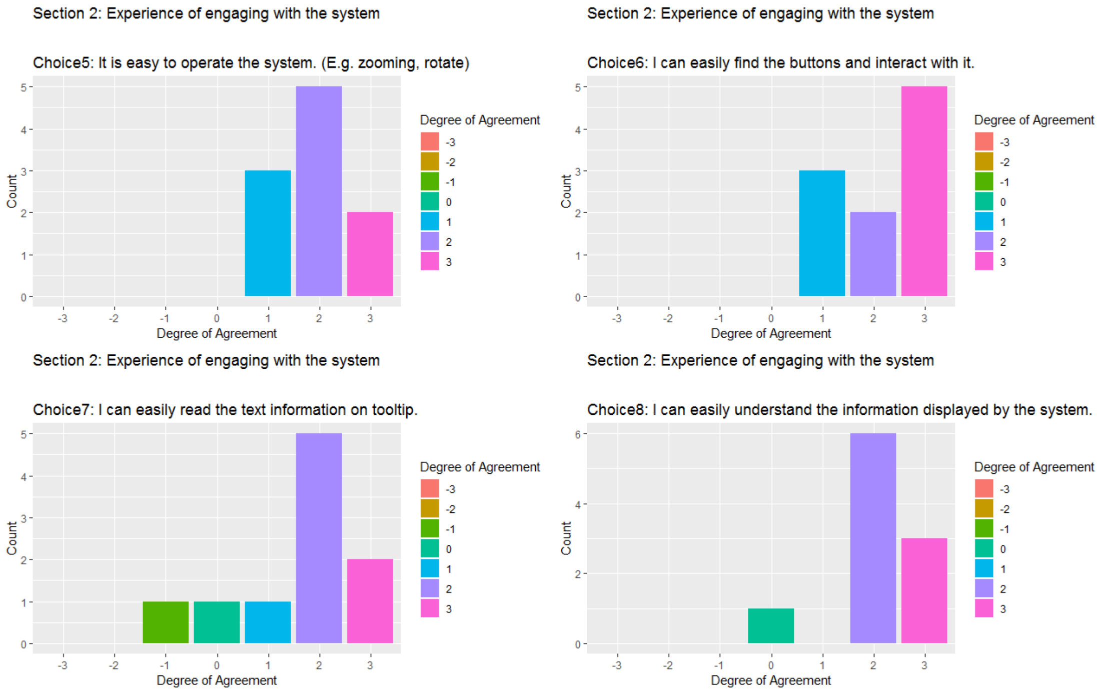




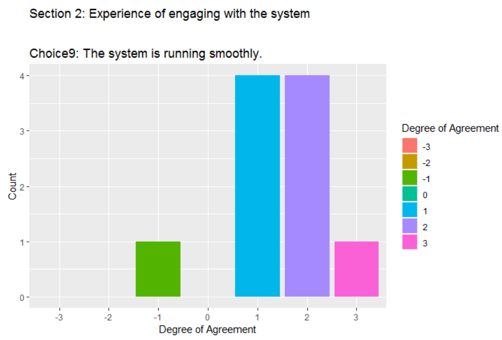


The first four question in section 2 is to evaluate the learnability and memorability. The result show that user could easily learn using the system with average score 2.6 and 2.1 in choice2&3 respectively. While the start is no as easily as learning, with 1.8 in choice.

The memorability is considerable good with average score 2.4 in choice4.

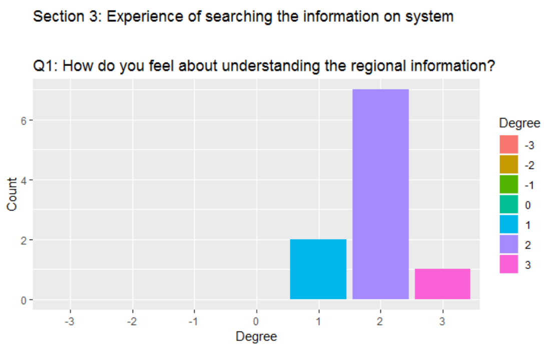


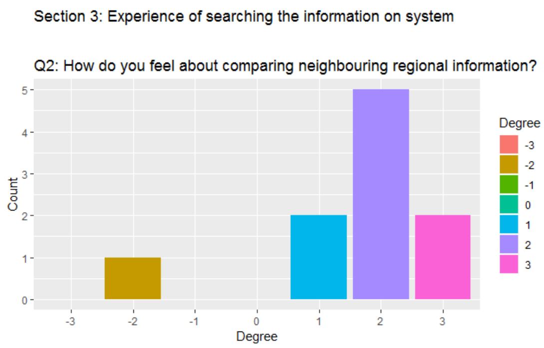
The next four question evaluation the aspects on interaction section. User generally feel that it is easy to opera the system (1.9 in choice5). However, the text info on tooltip is hard to read (1.6 in choice7). As for the button, user feels it better than tooltip (2.2 in c hoice6). The overall info displayed could be perceived by participants (2.1 in choice8)

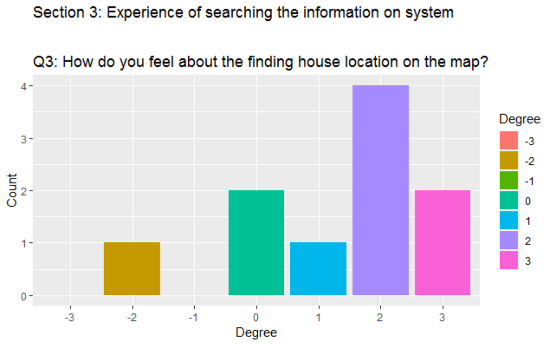


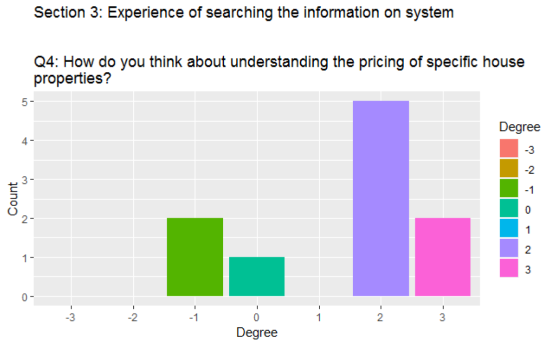
The next four question evaluation the aspects on interaction section. User generally feel that it is easy to opera the system (1.9 in choice5). However, the text info on tooltip is hard to read (1.6 in choice7). As for the button, user feels it better than tooltip (2.2 in c hoice6). The overall info displayed could be perceived

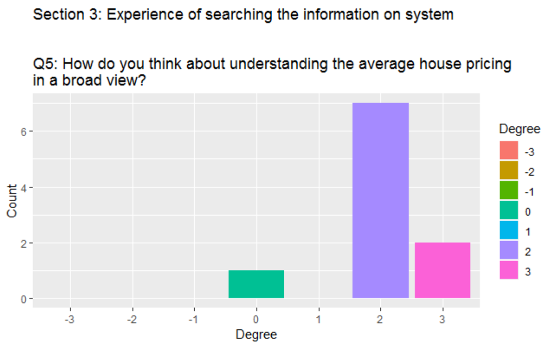
by participants (2.1 in choice8). The last question reflect that the system cannot run very smoothly with only 1.4 average score.

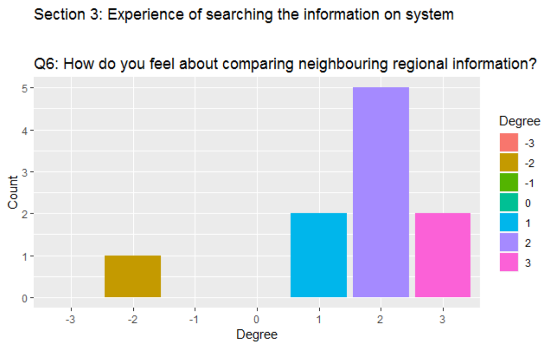




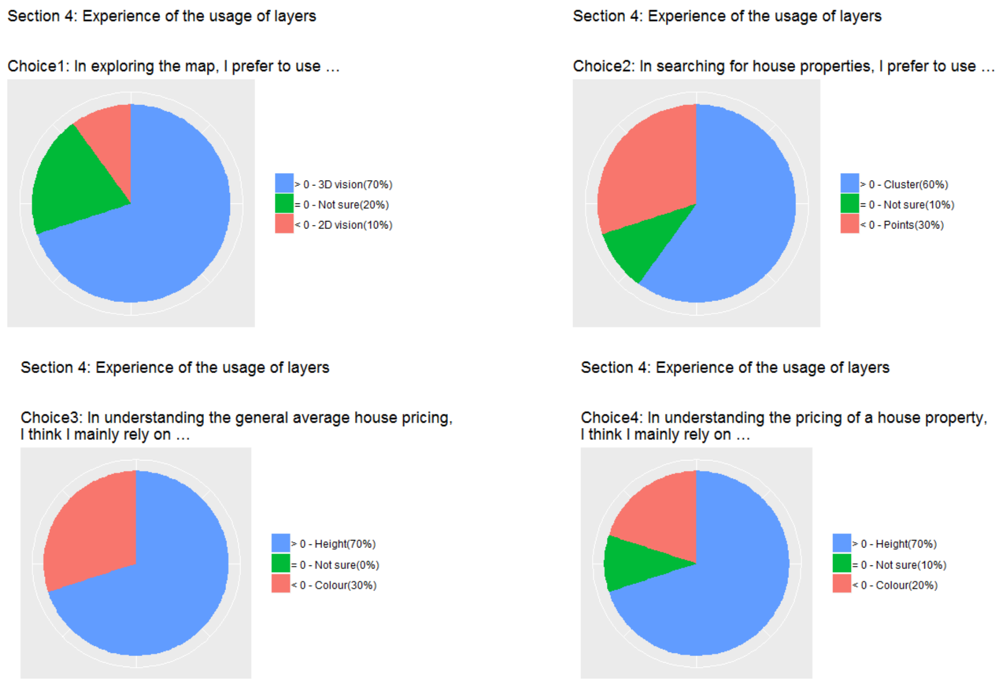




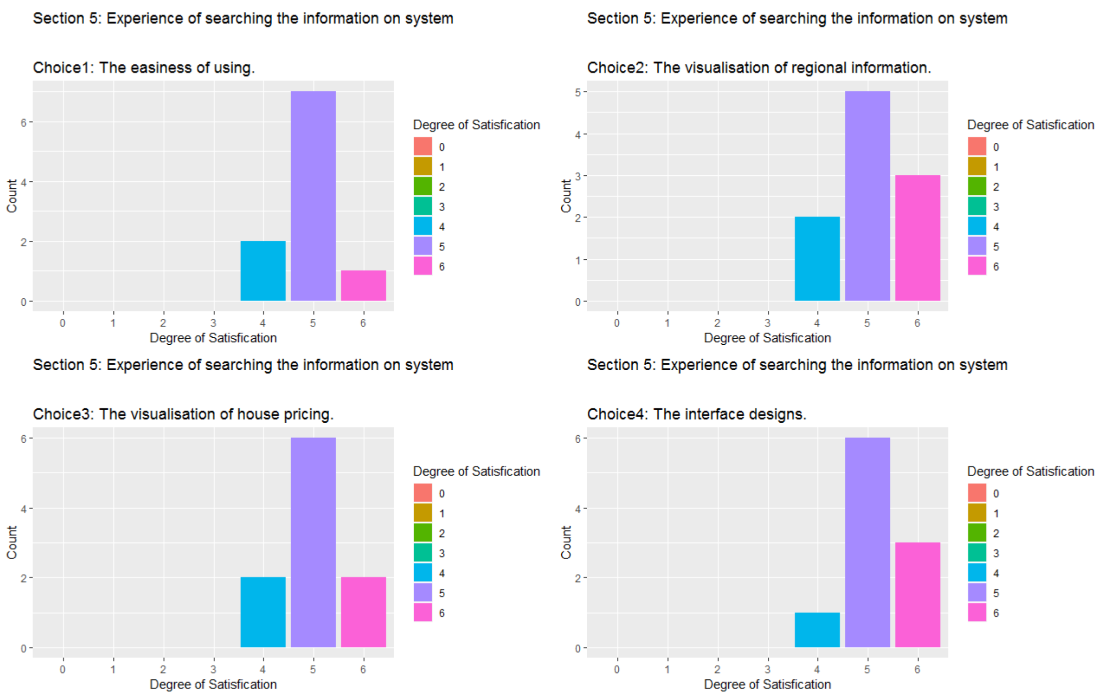


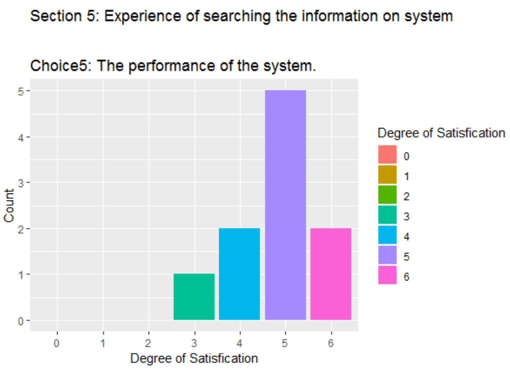


In this section, participants answer the question according to their experience in doing the task. The visualisation of regional info is considerably easily understood with 1.9 average score in Q1 and 2 average score in Q5. The search and finding properties is not that easy as only 1.3 in Q4, 1.4 in Q5. Although, some participant feel it is hard to compare the info scores, there are still 1.6 average score both in Q2 and Q6.



From the perspective of encoding preference, 70% prefer to explore map on 3D. Most of participants would like to view properties rather than single points. And they also think height as visual encoding is the major way for them to perceive and compare the house pricing.





This section reveal the satisfaction of the system. From the aspects of ease of use, regional info visualisation, house pricing visualisation, interface design, and the performance, the average score are quite high, with average score 4.9, 5.1, 5, 5.2, 4.8.

5. CONCLUSIONS AND DISCUSSION

All in all, this project uses the iterative design method to complete the seven stages of to achieve information visualisation. Based on the existing online house pricing platform, it is found that the function like mapping searching can better use the information visualisation method to display data. Regard on this point, the potential user's task flow is built and the requirements are established by informal interview. Then the web front-end framework and package are used for development, and the obtained regional info and house pricing are obtained as data source, and successfully implement an information visualisation platform. After the deployment is completed, user evaluation is performed. Using the controlled experiment method, the 2d and 3d visual encodings are quantitatively compared through the usability test of completing task. The results show no significance. The method of questionnaire is also applied to subjectively evaluate system. The analysis indicates that the system has value in several aspects, such as ease of use, aesthetics, and .etc.

The limitations of the project are also obvious, mainly focused on the following points. First of all, due to limited capabilities, it is not possible to develop a complete system from zero. The front-end framework used has great limits, which makes the interactive functions of visualisation imperfect, such as the interaction of buttons and the implementation of the filter function. Play an important role in search properties in real case. At the same time, the fluency of the web page is not so good. This can be improved, but it takes more time and effort. Due to the limitations of the framework, many functions, such as jumping to a specific information page, displaying a picture, etc. cannot be completed. Secondly, many users pointed out that although the task setting has certain rationality, it still has a very large distance from the actual use. In the specific information visualisation of the house, such as the number of rooms, the number of years of use, etc. is very insufficient. Moreover, the value displayed by the price is not reasonable enough. Some houses are superimposed on multiple rooms, while others are auction rooms that are not actual prices. But these factors have homogenous properties that can be solved by adding specific housing information. In addition, in the questionnaire survey and experiment process, the number of participants is small, the sample is small, and the deviation of the results may be large.

Finally, as for the extension of the prototype, most of ideas come in the process of doing user evaluation. The participants presented a lot of valuable reference. For example, interactive button like the panel, or tooltip, the text could uses the colour labeling consistent with the area information, which is easier for distinction. Using the icon to standard information on the house and design a more convenient and more natural layer with switching methods, etc. In theory, these can be achieved through technology.

Designing better visualize information is not only my dissertation topic, but also the direction that interaction designers, web developers, and user researchers work together.

6. REFERENCES

[1] Darley, William K., Charles Blankson, and Denise J. Luethge. "Toward an integrated framework for online consumer behavior and decision-making process: A review." Psychology & marketing 27, no. 2 (2010): 94-116.

[2] DEFANTI T A, BROWN M D, MCCORMICK B H.Visualisation:Expanding scientific and enginering research opportunities[J].Com- puter, 1989, 22(8):12-16.

[3] ROBERTSON G, CARD S K, MACKINLAY J D.The cognitive coprocesor architecture for interactive user interfaces[A].Procedings of the 2nd Annual ACM SIGGRAPH Symposium on User interface Software and Technology[C].New York:ACM, 1989.10-18.

[4] SINDIY O, LITOMISKY K, DAVIDOFF S, et al.Introduction to information visualisation (InfoVis)techniques for Model-Based sys- tems enginering[J].Procedia Computer Science, 2013(16):49-58.

[4] CARD S K, MACKINLAY J D, SHNEIDERMAN B.Readings in Information Visualisation:Using Vision to Think[M].San Francisco: Morgan Kaufmann, 1999.

[5] REN Lei, WANG Weixin, ZHOU Mingjun, et al.A model driven development method for interactive information visualisation[J].Jour- nal of Software, 2008, 19(8):1947-1967.

[6] LIU Dahai.The Research of Large-scale Data Visualisation[D].Tianjin:Tianjin University, 2009.

[7] Shneiderman, B., 1996, September. The eyes have it: A task by data type taxonomy for information visualizations. In Visual Languages, 1996. Proceedings., IEEE Symposium on (pp. 336-343). IEEE.

[8] INSELBERG A.The plane with paralel coordinates[J].The Visual Computer , 1985, 1(2):69-91.

[9] KANDOGAN E.Visualizing multi-dimensional clusters, trends, and outliers using star coordinates[A].Procedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C].New York:ACM, 2001.107-116.

[10] CRAIG P, KENNEDY J.Coordinated graph and scater-plot views for the visual exploration of microaray time-series data[A].Pro- cedings of the Ninth Annual IEEE Conference on Information Visualisation[C].Washington DC:IEEE Computer Society, 2003. 173-180.

[11] Chernoff, H., 1973. The use of faces to represent points in k-dimensional space graphically. Journal of the American statistical Association, 68(342), pp.361-368.

[12] TUFTE E.The Visual Display of Quantitative Information[M].Cheshire:Graphics Pres, 2001.

[13] Saito, T., Miyamura, H.N., Yamamoto, M., Saito, H., Hoshiya, Y. and Kaseda, T., 2005. Two-tone pseudo coloring: Compact visualization for one-dimensional data.

[14] CARD S K, SUH B, PENDLETON B A, et al.Time tre:Exploring time changing hierarchies[A].2006IEEE Symposium on Visual

Analytics Science and Technology[C].Washington DC:IEEE Computer Society, 2006.3-10.

[15] SELASSIE D, HELLER B, HEER J.Divided edge bundling for directional network data[J].IEEE Trans Visualisation and Graphics, 2011, 17(12):2354-2363.

[16] Shannon, J. and Walker, K., 2018. Opening GIScience: A process-based approach. International Journal of Geographical Information Science, pp.1-16.
[19] Walker R, Singsby A, Dykes J, et al. An extensible framework for provenance in human terrain visual analytics [J]. IEEE Transactions on Visualisation & Computer Graphics, 2013, 19(12): 2139-2148

[17] Walker, R., Slingsby, A., Dykes, J., Xu, K., Wood, J., Nguyen, P.H., Stephens, D., Wong, B.W. and Zheng, Y., 2013. An extensible framework for provenance in human terrain visual analytics. IEEE Transactions on Visualization and Computer Graphics, 19(12), pp.2139-2148.

[18] Rae A. From spatial interaction data to spatial interaction information? geovisualisation and spatial structures of migrationfrom the 2001 UK census[J]. Computers Environment & Urban Systems, 2009, 33(3): 161–178

[19] Andrienko N, Andrienko G. Visual analytics of movement: an overview of methods, tools and procedures[J]. Information Visualisation, 2013, 12(1): 3-24

[20] Lambert A, Bourqui R, Auber D. 3D edge bundling for geographical data visualisation[C] //Proceedings of the 14th International Conference Information Visualisation. Los Alamitos: IEEE Computer Society Press, 2010: 329-335

[21] Guo D S, Zhu X. Origin-destination flow data smoothing and mapping[J]. IEEE Transactions on Visualisation & Computer Graphics, 2014, 20(12): 2043-2052

[22] Correll M, Heer J. Surprise! Bayesian weighting for de-biasing thematic maps[J]. IEEE Transactions on Visualisation & Computer Graphics, 2017, 23(1): 651-660

[23] Pu J S, Liu S Y, Ding Y, et al. T-Watcher: a new visual analytic system for effective traffic surveillance[C] //Proceedings of the 14th IEEE International

Conference on Mobile Data Management. Los Alamitos: IEEE Computer Society Press, 2013: 127-136

[24] Ferreira N, Poco J, Vo H T, et al. Visual exploration of big spatio- temporal urban data: a study of New York city taxi trips[J]. IEEE Transactions on Visualisation & Computer Graphics, 2013, 19(12): 2149-2158

[25] Reddy M, Leclerc Y, Iverson L, et al. TerraVision II: visualizing massive terrain databases in VRML[J]. IEEE Computer Graphics & Applications, 1999, 19(2): 30-38 & Computer Graphics, 2010, 16(6): 1413-1420

[26] Chen A, Leptoukh G, Kempler S, et al. Visualisation of a-train vertical profiles using google earth[J]. Computers & Geosciences, 2009, 35(2): 419-427

[27] Lloyd, R., 1997. Spatial cognition: Geographic environments (Vol. 39). Springer Science & Business Media.

[28] Goodchild, M.F., 1992. Geographical information science. International journal of geographical information systems, 6(1), pp.31-45.

[29] Koffka, K., 1922. Perception: an introduction to the Gestalt-Theorie. Psychological Bulletin, 19(10), p.531.

[30] Brunswik, E., 1956. Perception and the representative design of psychological experiments. Univ of California Press.

[31] Gibson, J.J., 2014. The ecological approach to visual perception: classic edition. Psychology Press.

[32] Müller, H.J. and Rabbitt, P.M., 1989. Spatial cueing and the relation between the accuracy of "where" and "what" decisions in visual search. The Quarterly Journal of Experimental Psychology, 41(4), pp.747-773.

[33] Mennis, J.L., Peuquet, D.J. and Qian, L., 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. International Journal of Geographical Information Science, 14(6), pp.501-520.

[34] Kosslyn, S.M., Flynn, R.A., Amsterdam, J.B. and Wang, G., 1990. Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. Cognition, 34(3), pp.203-277.

[35] Lloyd, R. and Steinke, T., 1984. Recognition of disoriented maps: The cognitive process. The Cartographic Journal, 21(1), pp.55-59.

[36] Carlson-Radvansky, L.A., Covey, E.S. and Lattanzi, K.M., 1999. "What" effects on "where": Functional influences on spatial relations. Psychological Science, 10(6), pp.516-521.

[37] Marr, D., 1982. Vision: A computational investigation into the human representation and processing of visual information. MIT Press. Cambridge, Massachusetts.

[38] Gardner, H., 1987. The mind's new science: A history of the cognitive revolution. Basic books.

[39] Jackendoff, R., 1995. Languages of the mind: Essays on mental representation. mit Press.

[40] Mark, D.M., Freksa, C., Hirtle, S.C., Lloyd, R. and Tversky, B., 1999. Cognitive models of geographical space. International journal of geographical information science, 13(8), pp.747-774.

[41] Montello, D.R., 1993, September. Scale and multiple psychologies of space. In European conference on spatial information theory (pp. 312-321). Springer, Berlin, Heidelberg.

[42] Tversky, B., 1993, September. Cognitive maps, cognitive collages, and spatial mental models. In European conference on spatial information theory (pp. 14-24). Springer, Berlin, Heidelberg.

[43] Freundschuh, S.M. and Egenhofer, M.J., 1997. Human conceptions of spaces: implications for GIS. Transactions in GIS, 2(4), pp.361-375.

[44] Nivala, Annu-Maaria, Stephen Brewster, and Tiina L. Sarjakoski. "Usability evaluation of web mapping sites." The Cartographic Journal 45, no. 2 (2008): 129-138.

[45] Morse, E., Lewis, M. and Olsen, K.A., 2000. Evaluating visualisations: using a taxonomic guide. International Journal of Human-Computer Studies, 53(5), pp.637-662.

[46] Meyer, J., 2000. Performance with tables and graphs: Effects of training and a visual search model. Ergonomics, 43(11), pp.1840-1865.

[47] Ahn, J.W. and Brusilovsky, P., 2013. Adaptive visualisation for exploratory information retrieval. Information Processing & Management, 49(5), pp.1139-1164.

[48] Freitas, C.M., Luzzardi, P.R., Cava, R.A., Winckler, M., Pimenta, M.S. and Nedel, L.P., 2002, May. On evaluating information visualisation techniques. In Proceedings of the working conference on Advanced Visual Interfaces (pp. 373-374). ACM.

[49] Dowell, J. and Long, J., 1998. Target paper: conception of the cognitive engineering design problem. Ergonomics, 41(2), pp.126-139.

[50] Purchase, H.C., 1998, November. The effects of graph layout. In Computer Human Interaction Conference, 1998. Proceedings. 1998 Australasian (pp. 80-86). IEEE.

[51] Cleveland, W.S., 1993. Visualizing data (Vol. 2). Summit, NJ: Hobart Press.

[52] Nielsen, J., 1993. Iterative user-interface design. Computer, (11), pp.32-41.

[53] "XML and Semantic Web W3C Standards Timeline" (PDF). 2012-02-04.

[54] Brewer, C.A. and Pickle, L., 2002. Evaluation of methods for classifying epidemiological data on choropleth maps in series. Annals of the Association of American Geographers, 92(4), pp.662-681.

[55] Gliem, J.A. and Gliem, R.R., 2003. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

[56] Bangor, A., Kortum, PT and Miller, JT, 2008. An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), pp.574-594.

[57] Norman, K.L., Shneiderman, B., Harper, B. and Slaughter, L., 1998. Questionnaire for user interaction satisfaction. University of Maryland (Norman, 1989) Disponível em.

[58] Gliem, J.A. and Gliem, R.R., 2003. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Appendix 1:

Consent form
Information sheet
Interview transcripts
Questionnaire
Questionnaire response
User evaluation result
SPSS result