

SpendWise: Visual Analytics for Spend Insight and Action



Monash University

By:

Chunlei Chang

Supervisor:

Tim Dwyer

This thesis is presented in partial fulfilment of the requirements for the Degree of
Master of Information of Technology (Honours)
At Monash University, Australia

June, 2014

DECLARATION

I declare that this thesis contains no material that has been accepted for the award of any degree or diploma in any university, and that to the best of my knowledge, the thesis contains no material previously published or written by any other person except where due reference is made in the text.

Chunlei Chang

16 June 2014

Caulfield School of Information Technology

Monash University

Caulfield East, Vic. 3145

Australia

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude first and foremost to my supervisor Dr. Tim Dwyer for his constant guidance, innovative ideas, encouragement, proofreading of many thesis drafts, and helping me throughout this research.

I thank Adam Hart from ANZ who have supported our research project.

I acknowledge my fellow honours students for making the honours year indefinitely enjoyable.

I am very grateful to Ella Dobbyn, Rosemary Moore, and Neythra Weerakkody for commenting and proofreading this work.

Finally, I dedicate this thesis to my parents and my fiancée Rui Wang in appreciation of their love and support.

ABSTRACT

This thesis explores techniques for creating a collaborative information visualisation system for tabular spend data (actual spend transactions of fiscal year 2013 from ANZ). Many real-world analysis tasks can be performed with the combined efforts of a group of data analysts. Research has been done on information visualisation for visual analytic tasks in a collaborative environment. Several novel interactive visualisations are proposed and developed to help analysts gain better insight through easy operations. The applicability of network analysis techniques and tools has been tested on corporate spend data involving geographical, and temporal properties, within heterogeneous data types. Another goal of the design is to support both individual as well as group collaborative analysis activities. SpendWise, a tabletop visual analytics tool, has been developed to support individual and group visual analytics. Users are able to work together in a co-location environment (e.g. Tabletop) or work separately on their own devices and update their findings to a shared workspace. These concepts are explored in the development of a prototype and heuristic evaluation by ANZ domain experts.

TABLE OF CONTENTS

Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
List of Figures	vii
List of Tables	ix
1 Introduction	1
1.1 Overview	1
1.2 Research Background	1
1.3 Research Objectives.....	2
1.3.1 Significance and innovation	2
1.4 Research Methodologies	4
1.4.1 Stage 1: Prototype development	4
1.4.2 Stage 2: Iterative development on increasing feedback from domain experts	5
1.4.3 Stage 3: Qualitative evaluation.....	5
2 Literature Review	7
2.1 Introduction	7
2.2 Review of Collaborative Interaction	7
2.2.1 Tabletop visualisation	8
2.2.2 Collaboration tradeoffs.....	9
2.3 Review of Information Visualisation.....	11
2.3.1 History of information visualisation.....	11
2.3.2 Network visualisation.....	13
2.3.3 Interactive visualisation	16
2.3.4 Scalability	19
2.3.5 Current visualisation on the Web	20
2.3.6 Other relevant topics	21

2.4	User-Centred Design Methodology	24
3	System Design	26
3.1	Establish Requirements From Domain Experts	26
3.2	System Architecture.....	29
3.3	Data Exploration	30
3.4	Data Pre-processing	30
3.4.1	Cleaning the raw data	30
3.4.2	Loading data.....	31
3.5	Data Service	32
3.6	Data Presentation	33
3.6.1	Views and design choices.....	34
3.6.2	Operations.....	47
3.7	Collaborative Scenarios.....	48
3.7.1	Co-location collaboration.....	48
3.7.2	Collaboration from multi-devices	49
4	Heuristic Evaluation	50
4.1	Introduction	50
4.2	Design of Evaluation	50
4.3	Method	51
4.4	Applying The Evaluation	51
4.5	Evaluators	52
4.6	Tasks And Scenarios.....	52
4.6.1	Task set 1 on collaborative panel.....	53
4.6.2	Task set 2 on CFO View	53
4.6.3	Task set 3 on collaborative panel.....	53
4.6.4	Task set 4 on private/public workspace.....	54
4.7	Use-Case	54
4.7.1	The usage flow	56

4.8	Analysis of Results.....	57
4.8.1	Visualisation improvements.....	57
4.8.2	Functionality improvements.....	58
4.8.3	Finding on bugs.....	58
5	Contribution.....	60
6	Future work.....	62
7	Conclusion.....	64
8	References.....	65
9	Appendix.....	69
9.1	Glossary.....	69
9.2	Transcripts Of Heuristic Evaluation.....	70

LIST OF FIGURES

Figure 1-1 (a) Prototype Architecture (left) (b) Prototype development process (right).....	5
Figure 1-2 SpendWise mock-up design: Multiple user interactive console.....	6
Figure 2-1 Two visualisations of the same undirected graph containing 50 vertices and 400 edges. The node-link diagram (A) is computed using the 'neato' program and the matrix representation (B) is computed using our VisAdj program (Ghoniem et al., 2005).....	14
Figure 2-2: Orion Visualisations. From left-to-right: (a) User interface (b) Node-link diagram. (c) Matrix view of the connection. (Heer & Perer, 2011)	16
Figure 2-3: Dynamic queries in FilmFinder (Ahlberg & Shneiderman, 1994)	17
Figure 2-4. Pliability envisioned in Sens-A-Patch (Löwgren, 2007).....	17
Figure 2-5: VisGets (Dork et al., 2008)	18
Figure 2-6: Node-and-link diagram versus PivotGraph (Wattenberg, 2006).....	19
Figure 2-7: Social network visualisations. (a) A node link representation of a very simple 8 node network (b) An adjacency matrix representation of the same network, notes that the number of grey cells is equal to the number of links in (a). (van Ham et al., 2009).....	20
Figure 2-8: The Tableau interface	21
Figure 2-9: Phase of the iterative design process (Van Velsen et al., 2008).....	24
Figure 3-1: Prototype architecture	29
Figure 3-2 Prototype data class diagram	31
Figure 3-3: Real-time communication framework in SpendWise with SignalR	33
Figure 3-4: Passing message structure by the SignalR.....	33
Figure 3-5: Early visualisations: organisational force-directed graph (upper left), simplified force-directed graph (upper right), ANZ hierarchy (bottom left), Vendor spending history (bottom right) .	35
Figure 3-6: Spend tree map - Data display based on ANZ business units and its total spends	36
Figure 3-7: Spend tree map - Data exploration by clicking the tree map.....	37
Figure 3-8: Left - Data service - Spend hierarchy web service Right: Response format	38
Figure 3-9: Spend Heat map – geographic spend data visualisation	39
Figure 3-10: Data service: Geographic spend web service	40
Figure 3-11: Data Service: Geographic data format	40
Figure 3-12: Vendor and spend network visualisation - the relationships between business unit, related vendors, and spend categories.....	41
Figure 3-13: Hovering (upper left); selecting a vendor (upper right); selecting a spend category (bottom left), finding related vendors for multiple spend categories (bottom right).....	42

Figure 3-14: Time Range Slider enables user to filter data in a specific time range.....	42
Figure 3-15: Data service: (left) Vendor and spend category (right) response format	43
Figure 3-16 Line chart: Detailed vendor spend history.....	43
Figure 3-17: Date service - Response format of spend history of the selected vendor	44
Figure 3-18: Overall interactive panel.....	45
Figure 3-19: CFO Spend report panel	46
Figure 3-20: CFO Spend report panel: CSV data format	47
Figure 3-21: Drag and Drop to make a new query.....	48
Figure 3-22: Collaborative interaction on a tabletop (multiple users and multi-touch experience)....	48
Figure 3-23: Collaboration from multiple devices (private work space)	49
Figure 4-1: Initial interface of SpendWise evaluation.....	54
Figure 4-2: Usage flow: Regular user work flow	56
Figure 4-3: Usage flow: The most frequent work flow	56
Figure 6-1: Graph Attribute Space Explorer.....	62

LIST OF TABLES

Table 3-1 Operation supported in SpendWise.....	26
Table 3-2 Scenario 1: Interactive scenario.....	28
Table 3-3 Scenario 2: CFO summary	28
Table 3-4 Scenario 3: Collaborative interaction.....	28
Table 3-5 Scenario 4: Collaboration from different devices.	29
Table 3-6 JavaScript libraries used in SpendWise.....	34
Table 3-7 Operation supported in SpendWise.....	47
Table 4-1 Observations of heuristic evaluation	59
Table 9-1 Terminology	69
Table 9-2 Transcripts of heuristic evaluation.....	84

1 INTRODUCTION

1.1 OVERVIEW

ANZ generates and tracks data about where the organisation spends money each minute, hour and day. Analysis of this data can help ANZ save millions of dollars by making smarter decisions on spend and cost. Raw data stored on computer discs is of no value, rather it is the information that can be extracted from the data that is valuable. Extracting information from this data often requires an analyst with database or data processing knowledge to query and analyse in an often tedious and time-consuming manner. To provide better tools for these professionals, ANZ is looking at major investments in information management capabilities, as well as Business Intelligence (BI) and visual analytics. ANZ seeks more advanced analysis and visualisation technology to provide rich user experience and gain better spend insight. Thus, ANZ is engaged with Monash University to explore and verify certain concepts.

In this project, a novel visual presentation method for exploring ANZ spend analytics is developed. Another innovative aspect of the project is to develop tools allowing multiple analysts to explore a dataset together in a shared workspace on a large display and across different devices. Analysts at large financial institutions routinely use data visualisations like time series charts and organisation hierarchy charts to analyse their data. However, these visuals are usually an output of analysis introduced in spreadsheets or database reporting tools. In this project, we make steps toward developing an analysis tool that is completely visual and supports a high degree of interactivity. Furthermore, sophisticated visualisations have been made that are not typically used by financial analysts, such as network visualisations to display connections between data entities (i.e. business unit, spend category, and vendors).

1.2 RESEARCH BACKGROUND

The challenge of working with massive datasets is how to present results of data exploration and analysis in a way that is not overwhelming. Although tables often describe multivariate data without explicit network semantics, it may be advantageous to explore the data modelled as a graph or network for analysis (Mazza, 2009) of spend “hubs”, “clusters” and other connectivity.

In recent years network visualisation techniques have been developed to meet this challenge. Across these techniques, many of the same basic operations are routinely performed. One such technique is aggregated over the vertices and edges to compute summary statistics or computing shortest

paths between vertices. The proposed network discerns important relationships through advanced analysis and data-visualisation. Network visualisation is an effective means to understand the patterns of interaction between entities (Hu, Snitkin, & DeLisi, 2008). It is particularly relevant when you are trying to find relationships among thousands of employees, vendors, and locations. For example a vendor involved in many transactions show up as a highly connected “hub”. Similarly, groups of highly connected entities appear as clusters in the visualisation.

Visual Analytics give analysts the ability to directly explore data in a way that could be awkward with other tools. Sometimes, by giving the data a visual mapping that is readily parsed you can identify important patterns that may not have been suspected otherwise. If sophisticated analyses can be performed quickly and easily, and results visually presented in ways that showcase these patterns, then more can be learned from existing data sources. Furthermore this ease of querying and exploration allows people with little or no database knowledge, across all levels to make faster, and more effective decisions. People who really need to understand data are not likely to be database experts or particularly savvy users. Decisions are often made with data presented by analysts to communicate effectively with other analysts and business executives. Therefore, a collaborative interface is strongly needed for these non-experts.

In ANZ’s scenario, multiple analysts often sit together in a meeting room to explore and analyse the dataset and communicate on the result. The ideal visualisation tool needs to support multiple user workflows in a collaborative manner (so called collaborative interaction) in the same location (so called co-located), specifically, on a tabletop (a table size multi-touch screen). Therefore, much work was reviewed on the design considerations and key concepts in the area of collaborative interaction in section 2.2.

1.3 RESEARCH OBJECTIVES

1.3.1 Significance and innovation

The objective of this research project is to investigate the applicability of network visualisation to the analysis of corporate spend data involving geographical, and temporal properties, within heterogeneous data types. To better understand ANZ’s needs, we interviewed analysts and had them discuss typical tasks and queries that they would need to perform using this dataset. A detailed list of such tasks and scenarios are given in section 4.6. To achieve this goal, we develop SpendWise, a prototype of a user interface suitable for ANZ spending data analysis. SpendWise supports general querying and business analytics for spend data.

Another contribution is to test the applicability of the network analysis techniques and tools such as centrality metrics (Chen, 2004). The techniques are useful in helping to identify important structural aspects of the network, which can then be highlighted in the visualisation work. Our investigation presents a view of the connection between business units, vendors, and spend categories. Exploration from one node to another node to find out the relationship between those entities is made easier by this work. For example, by dragging from one organisational spend view and dropping it to another network view, the prototype would quickly switch to another view which displays the related vendors of this selected business unit and the related spend categories. By clicking on a node (representing a vendor company) in the network view, the break down spend history of the selected vendor will be updated in the line chart. As a result, analysts can gain insight and new understanding of the spend data by quickly traversing from different nodes. Using the predefined view of fragmentation of spend data, non-expert users are able to quickly learn how to operate the application.

In addition, the prototype aims at offering views of a number of predefined analyses which analysts typically want to understand. Such as:

- Organisational hierarchical spend analysis - a view based on a hierarchical tree map to display spend figure and percentage based on their organisational structure. Analysts can quickly check the spend total of each business unit at different levels of company organisational hierarchy (see Figure 3-6).
- Geographic spend heat map analysis - a geographic view that displays the spend data of each area on the map to help analysts understand the spend data geographically (see Figure 3 8).
- Business units, vendors, and spend categories analysis - a network view of the spend schema of the business unit and its related vendors. This also carries the information of the total spend amount and transaction details (see Figure 3-12). It helps analysts to gain insight into the relationships between three entities: business units, vendors, and spend categories.
- Temporal analysis – temporal filtration is important to analysts who are willing to narrow down the analysis to a specific period. We used a slider gadget that enables users to filter data to a specific time period (see Figure 3-14, Figure 3-6).
- Detailed spend history analysis - a view that displays the detailed spend history in the previous fiscal year (see Figure 3 15). This assists analysts to check the detailed transactions of a vendor company on each day.

- Corporate level spend overview analysis – ANZ hope to have a summary view to identify patterns and trends of the spend data. We implemented a high level summary view of overall global spend (see Figure 3-19).

Another contribution is for easy deployment and portability of the tools and to explore some new techniques in collaborative interactions. SpendWise is developed in HTML5 and JavaScript standard with a widely used open-source web visualisation toolkit (see section 3.6.1).

1.4 RESEARCH METHODOLOGIES

We followed an agile user-centred design approach, paying attention to the needs and requirements of target users during the design of the visualisation and interaction model. The prototype development was iterative and incremental in the sense that feedback was collected from interviews and monthly sync up meetings with ANZ to improve SpendWise.

There were three stages involved in the SpendWise development:

1.4.1 Stage 1: Prototype development

In Stage 1, a prototype was developed to support rapid and flexible iterative agile development. There were three steps in developing this prototype. First, data pre-processing techniques were applied to the spend data to make it suitable for being accessed from the ASP.NET. Second, we devised data services to handle the requests from the front end. The final step involved displaying the data in an appropriate visualisation format and providing interaction supports for multiple users. Each of the following three sections discusses one of the three steps mentioned above in more detail.

1.4.1.1 *Data pre-processing*

The ANZ spend dataset is less than 100MB with around one million spend transactions and no more than 10 tables. For this data scale, it can be easily loaded into memory as binary data instead of installing a SQL server at the backend. A data service layer is required to handle the user queries, extract information, and return required data formats. Design consideration of data services layer are examined in next section.

1.4.1.2 *Data service*

In our project, web services provide a cross-platform solution for exchanging data between systems. Although web services are normally used to allow different operating systems, object models and programming language to send and receive data, they can also be used to dynamically inject data into an Asynchronous JavaScript and XML (AJAX) page or send data from a page to a back end

system. This meets the communication requirement of our project. A data service was built to handle requests from front end, and return the response in JSON/XML/CSV format. As shown in Figure 1-1 (a), we planned to employ WCF Services or ASP.NET services which can be exposed to AJAX clients by adding the appropriate endpoints either through configuration or by using a service host factory customised to generate a service host that configures the AJAX endpoint automatically.

1.4.1.3 Data Presentation & Interaction

With the data service layer ready in the back end, a front end application with JavaScript and D3JS library to present data from data services are built. Furthermore, techniques were investigated that enable multiple views for multiple users to work together with one large screen. Shown in the mock-up picture in Figure 1-1 (b), we also combined the spend data with geographic information to generate a spend heat map which makes it easier to understand. This enables a playful and dynamic experience for data exploration. The proposed visualisation allow users to traverse the nodes by touch or click: to browse the node information through the manipulation of the node-link diagram. Operations includes link traversal, filtering, aggregation, etc.

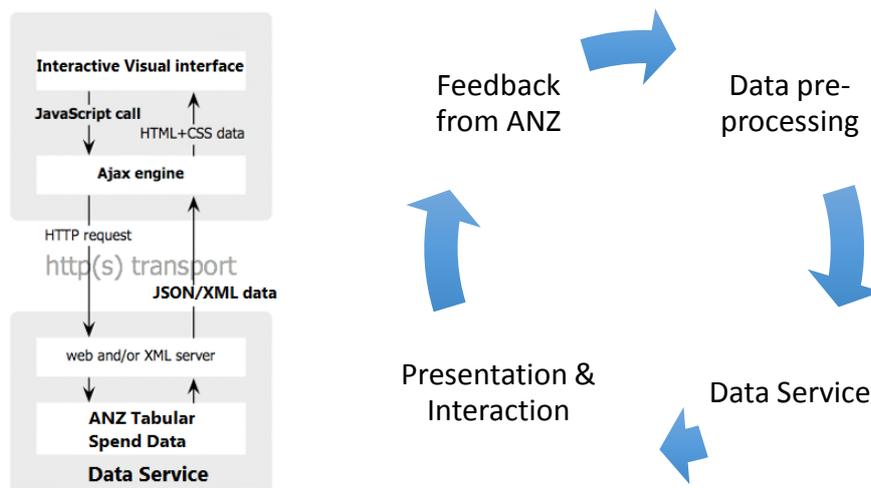


Figure 1-1 (a) Prototype Architecture (left) (b) Prototype development process (right)

1.4.2 Stage 2: Iterative development on increasing feedback from domain experts

In this stage, we had frequent communication and evaluation with the ANZ team to collect their feedback on a monthly basis. By collecting feedback, we incrementally improved our prototype as stated in section 1.4.1.

1.4.3 Stage 3: Qualitative evaluation.

We conducted a heuristic evaluation (Nielsen & Molich, 1990) on the user interface of the prototype at the final phase (see section 4).

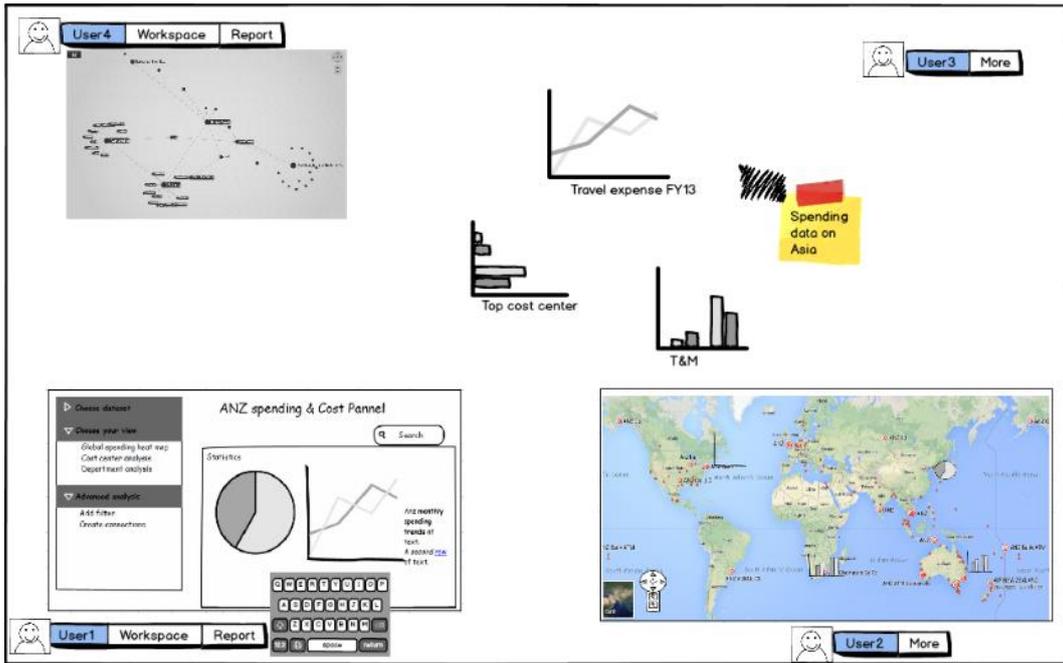


Figure 1-2 SpendWise mock-up design: Multiple user interactive console

For collaborative interactions, the initial idea is to utilise the new characteristics of HTML5 and SVG to create a collaborative platform where multiple users (up to 4 due to the size of tabletop) have their own consoles and are able to operate independently and share their results and graphs through simple drag and drop. In the HTML5 standard, canvas and SVG can be rotated through any angle. Thus, users can have their own views and consoles as shown in mock-up pictures in Figure 1-2. Every user is able to create their own analysis and share their results with others by drag-and-drop and enlarging/zoom operations.

2 LITERATURE REVIEW

2.1 INTRODUCTION

One of the main challenges in our project was that the dataset constantly grew in size and complexity. Visualisation techniques has the potential to offer different ways of exploring and analysing the data. Past research in both collaborative interaction and information visualisation have been suggested as important areas in tackling these information complexity challenges (Thomas & Cook, 2005).

This section examines visualisation techniques, and some related research that produces interactive visualisation for tabular data, ranging from popular graphic user interface tools to some low-level languages such as D3.js. Furthermore, collaborative interaction was examined in terms of tabletop visualisation and collaboration trade-off between offering a single or multiple independent instances of data views (see section 2.2). A brief review of information visualisation are conducted, including network visualisation, review of database exploration frameworks, examination of scalable vector graphics, and case studies of popular web visualisation libraries in section 2.3. In addition, user-centred design was specified while rapid iterative testing and evaluation (RITE) methods are discussed and filtered in order to conduct our research in section 2.4. The gaps in knowledge that motivate our research are mentioned in every section.

2.2 REVIEW OF COLLABORATIVE INTERACTION

Collaborative information visualisation is a relatively new and still under explored research area. Activities as part of analytic reasoning tasks are often performed in a collaborative scenario. Collaboration can be co-located (in the same location) or in a distributed shared workspace. Furthermore, group activities can be categorised into “independent” where tasks can be done by individual and “shared” where tasks need to be done in parallel by group members. Thanks to multi-touch screen and tabletop displays, multiple users are able to have face to face collaboration, discussion, interpretation, and analysis around information displays. Large displays expand the possibilities of desktop-based data analysis environments as they allow multiple users to work co-located with sufficient room for individual work and group collaboration. Use of large multi-touch screen allows team members to manipulate both shared and individual instances of data representations concurrently. Realising this opportunity requires comprehensive considerations of the design of information visualisation workspaces, representations, and interaction techniques.

Visualisation tools designed for single user have been studied with co-located collaborative tasks (Scott, Grant, & Mandryk, 2003). Their work offered guidelines and general considerations on the effective co-located collaboration around a tabletop display. It also proposed a set of technologies and interactive scenarios which are beneficial to our research project.

2.2.1 Tabletop visualisation

In our research, a large size multi-touch tabletop is used as the collaborative platform where multiple analysts are able to conduct their work both individually and collaboratively. Generally, tasks and group work can be described as *“tightly coupled”* and *“loosely coupled”*. If a participant cannot do much work before having to interact, then it is tightly coupled; conversely, if a participant can work independently, the work is loosely coupled. Our research aims to offer the flexibility of a loosely coupled, co-located tabletop collaborative experience.

“Early work in collaborative interaction has indicated that many group activities, such as brainstorming or planning, involve phases of mixed-focus collaboration in which group members transition from loosely coupled parallel work, to closely coupled group work” (Gutwin & Greenberg, 1998). It is also known that these transitions require the coordination of group members’ activities. Tabletops have been used for these activities since the early 1990s. The Responsive Workbench (Wesche et al., 1997) was one of the first visualisation systems designed for co-located collaboration around a large horizontal surface. It is a virtual reality environment that requires shuttered glass to see 3D scene and a glove with Polhemus sensor (a motion tracking sensor determining the moves and gestures). These two works present general concepts of tabletop collaboration from two decades ago.

In a collaborative environment, group activities often happen at the same time. Mixed-focus collaboration has recently been applied to synchronous co-located collaboration with information visualisation over shared displays (Isenberg, Tang, & Carpendale, 2008; Tang, Tory, Po, Neumann, & Carpendale, 2006). In the context of collaborative tasks on a shared visualisation, two observational studies were presented (Tang et al., 2006) that examine three viewing techniques: lenses, filters, and ShadowBoxes to understand different types of group cohesion. Specifically, lenses show information in spatially localised areas, filters show information globally, and shadowboxes allow spatially localised areas to be displaced. The result indicates that individuals frequently engage and disengage with group activities through *“several distinct, and recognizable states”*. These states and its consequences are described for tabletop interface design.

The activities that groups perform together include the acquisition, analysis, or interpretation of information; sharing and presentation of analysis result; and decision making (Chuah & Roth, 2003;

Heer & Agrawala, 2008). General design considerations for collaborative analysis in a distributed setting was well presented by (Heer & Agrawala, 2008). These design considerations include:

- Asynchronous collaboration in visual analysis environment – this enables analysts to work together efficiently without interference in collocated collaboration (e.g., a shared work space). Greater scalability for group-oriented analysis is offered compared to synchronous distance work (e.g., real-time networked displays).
- Highlighting issues of work parallelisation, communication, and social organisation - the proposed approach encourages individual assessment and can reduce groupthink bias. In a cooperative scenario, it may involve tightly coupled collaboration, requiring awareness and communication among participants. This work benefits our collaborative visualisations that collaborators can immediately benefit from the analysis results of others.

The above mentioned work motivated us in the design of collaboration scenarios of our prototype.

2.2.2 Collaboration tradeoffs

It was noted that a tradeoff is necessary between offering a single or multiple independent instances of data views. With a single shared representation, productivity of the individual working independently may be compromised. Yet using separate copied views may initially prevent development of group collaborative dynamics. For example, an analyst may be frequently distracted by glimpsing others' results. Consequently, several collaborative information visualisation and analysis systems have been developed to deal with the problem of trading off individual and group work with data visualisation and have offered some solutions for coordinating both types of work.

A distributed collaborative visual analytics framework is offered that represents private viewpoints as distinct from shared viewpoints (Brennan et al., 2006). In this framework (Brennan et al., 2006), it uses a knowledge representation scheme based on annotated logic to generate visualisation. This enables not only tracking and fusing different viewpoints, but also unpacking them. Individual views can be explicitly shared with others and merged to assist group analysis.

In another distributed analysis system (Keel, 2006), a system of computational agents were employed to support the exchange of information among team members to converge their individual contributions. Once an individual has uncovered potential relationships in his workspace, computational agents will identify it, and then relay it to the larger group of collaborators.

In our research project, we offer explicit logical and graphical support for sharing information and translating among different views. However Keel's approach is quite different as individual viewpoints are emphasised and spontaneous interactions are not easily possible. For example, users

are not able to freely and quickly transit to other views of the data, but need to wait for the computational agent to identify it.

We would like to support coordination of views and interactions so that multiple analysts can follow a real-time insight, glance at another's view, and transition quickly between different views of data. In order to achieve our goal, we found two previous systems that were developed in co-located data analysis (Isenberg & Carpendale, 2007; Isenberg & Fisher, 2009). In the tree comparison system of Isenberg & Carpendale, group members can create multiple view instances and interact with these views as separate entities on a shared interactive tabletop display. They present a new system that facilitates hierarchical data comparison tasks for the co-located collaborative visualisation. However, in their proposed system, the coordination was limited to a tree comparison option, and the group members are not able to coordinate data annotations or other data modifications.

In Isenberg's later research, Cambiera (Isenberg & Fisher, 2009) was presented as a tabletop visual analytics tool that supports collaborative information foraging activities. The word "foraging" means it allows user to do casual, ad-hoc, and unrestrained searching to explore and gain new findings. Animap (Gobel, 2013) has further explored this topic on how to design the visualisation to support serendipitous discovery. Collaborative brushing and linking was defined as an awareness mechanism which enables analysts to focus on their own hypotheses in a collaborative session while still remaining aware of the group's activities. Analysts were provided with coordinated visualisation of their own search result through text document collections. Through collaborative brushing and linking, individuals linked their search results and views and search overlaps were explicitly visualised.

We extended the collaborative "brushing and linking" (Buja, McDonald, Michalak, & Stuetzle, 1991) into our prototype and test its applicability to the spend data. "Brushing" is the synchronisation of the different data views to reflect changes made in another data view. "Linking" enables a variable to be highlighted in all views when a particular variable is selected. Brushing and linking are implemented in SpendWise CFO summary view (see section 3.6.1.6).

Lark (Tobiasz, Isenberg, & Carpendale, 2009) presents a different approach to the coordination of activities through explicit coordination points on data, representation, presentation, and view levels that employ a representation of the information visualisation pipeline into the shared workspace. Working loosely coupled on visual analytic tasks using tools such as collaborative brushing and linking enables users to share common ground and awareness as they work.

Recent research guided us to design a collaborative environment (in a shared workspace) that supports a set of scenarios, such as collaborative brushing and linking, coordination of views and interactions. Additionally, we also made an attempt to support the collaboration across multiple devices (reviewed in section 2.3.6.4).

2.3 REVIEW OF INFORMATION VISUALISATION

New visualisation techniques assist analysts to gain a better sights from data by offering novel ways of exploring and analysing the data. Insights and understandings on spend data can help enterprises save millions of dollars by making smarter decisions on spend and cost. In order to gain better insight of the spend data, it is critical to visualise each instance of data and their relationships. Network visualisation is the key part of this research. Two common types of representation are node-link diagrams and adjacency matrix views. In node-link diagrams, the data organised in a network structure can be naturally represented by graphs, where boxes (called nodes) represent the instances and lines (called links, edges) represents the relationships between the instances (Ware, 2000). We review the development of the information visualisation in section 2.3.1. We discuss issues to do with creating node-link diagram in section 2.3.2, and also an alternative visualisation approach for drawing graphs using adjacency matrix. Section 2.3.3 introduces the state of the art in interactive visualisation; particularly looking at how complex queries can be formulated that simultaneously involve temporal, spatial, and topical data filters. In section 2.3.4, we also learn past work on tackling the scalability issue of visualisation with a large dataset. Related web visualisation projects that inspired our research are examined in section 2.3.5. Finally, the latest supporting technologies and frameworks that provide support to our research are introduced in section 2.3.6.

The following section examines the related works of information visualisation.

2.3.1 History of information visualisation

Information visualisation conveys abstract information in an intuitive way. Effective visual presentation allows users to see, explore, and understand large amounts of information at once. Similarly, when it comes to visualisation, effective visualisation maximise users' understanding through clear, unambiguous diagrams. *The Visual Display of Quantitative Information* (Tufte, 1985) discusses the importance of good design in the visual representation of data. Tufte's thesis postulates that visualisation is highly complex because it requires the designer to have cross-domain knowledge of mathematical, artistic, and psychological factors when constructing a graph. His book provides a thorough review of the history of trends in the field of data visualisation from the 19th century to the present to explore best practice for constructing visuals. He presents case studies

about how to incorporate a variety of factors into creating visualisations that balance data validity with clear presentation and aesthetics. Tufte also postulates the concept of data-ink minimisation that avoid superfluous decoration of graphs by removing the distracting excess. Tufte's conclusion regarding balancing is helpful to our project because some advanced visualisation methods (mentioned later in this section) make extensive use of computer graphics technology.

Seeking a better design of visualisation, many data exploration frameworks have been developed in past decades in this field. Projects such as VQE (Derthick, Kolojejchick, & Roth, 1997), Visage (Roth et al., 1996), Devise (Livny et al., 1997), Tioga (Stonebraker, Chen, Nathan, Paxson, & Wu, 1993), offer visualisation environments that directly support interactive database exploration by generating visual queries. DEVise (Livny et al., 1997) has been developed to allow users to easily develop, browse, and share visual presentations of large tabular datasets from multiple sources. In addition, it enables users to share some visual representations of the data, and dynamically explore it individually or concurrently, and seamlessly query and combine data from a variety of local and remote sources. Visualisations and queries can be directly generated by manipulating and interacting with the visualisation system interface.

These systems have flexible mechanisms for visualising query results to graphs. Furthermore all of the systems support mapping database records to grid properties of the marks in the graphs. However, none of aforementioned systems support table-based organisations of their visualisations. To solve the problem, Table Lens (Rao & Card, 1994), and later Polaris (Stolte, Tang, & Hanrahan, 2002) have been designed to support table-based visual explanation tabular data. Polaris is an interface for exploring large multi-dimensional databases that extends the well-known Pivot Table interface to directly generate a rich, expressive set of graphical displays. It includes an interface for constructing visual specifications of table-based graphic displays and has the ability to generate a set of relational queries from the visual specifications. However, none of aforementioned systems pays special attention to the potential of imposing user-defined relationships between attribute values in the form of networks.

Similarly, Jigsaw (Stasko, Görg, & Liu, 2008) represents documents and their entities visually to help analysts examine them more efficiently and develop theories about potential actions more quickly. It provides multiple coordinated views of document entities with a special emphasis on visually illustrating connections between entities across the different documents. However, since the connection model is centred on documents, allowing flexible exploration of entity relationship without having to explicitly click nodes and edges, the connection between table columns are less intuitive.

There are many forms of tabular data. Often tabular data can be interpreted as either multivariate data or attribute relationship graphs. Ploceus (Liu, Navathe, & Stasko, 2013) offers a general approach for performing multidimensional and multilevel network-based visual analysis on multivariate tabular data. It employs an underlying relational algebraic framework to support flexible construction and transformation of networks through a direct manipulation interface. It also integrates dynamic network manipulation with visual exploration through immediate feedback mechanisms. It also provides a graphical user interface for iterative network manipulation and visualisation. Ploceus' interface enables the rapid manipulation of large graphs, including the specification of complex link relationships, using simple drag-and-drop operations with the desired node types (Liu et al., 2013). Relational tables offer a flexible model for representing data, while network extraction involves creating linking queries that regularly include one or more join operations from those tables. Ploceus also offer a Network schema view which is a sandbox-like environment where user can construct and manipulate networks at a conceptual level. The network view offers the corresponding network visualisation and updates when it is being modified. This inspire our project to create a network schema for spend entities, such as vendors, business units, and locations.

Ploceus and its toolkits and libraries provide many valuable resources for our project. However, Ploceus was written entirely in Java on the NetBeans Rich Client platform which does not offer powerful graphical manipulation and collaborative interaction. To overcome this, our project employed HTML5 and JavaScript which provide a client and server framework that offers an engaging interactive user experience. Furthermore, network schema was applied in our prototype which enable users to move from one view to another by filtering and aggregating data.

2.3.2 Network visualisation

In recent years, network visualisation software is becoming an increasingly popular method. Network visualisation offers an intuitive way to visualise connectivity in data. In our project, we applied network visualisation to visualise spend data in order to provide a direct view of the relationships between those entities, such as business units, and vendors. To understand how such designs are implemented, the field of network visualisation are discussed in details.

In the social network analysis, two representations are commonly used: node-link diagrams (typically using force-directed placement) and adjacency matrix views (Ghoniem, Fekete, & Castagliola, 2005). As displayed in Figure 2-1, in the node-link diagram, the node represents instances of data where the link represents the relationship between the instances. In node-link diagrams, the data organised in a network structure can be naturally represented by graphs. In a node-link graph, nodes or vertices

represent instances of data whereas connections or edges represent relationships between the instances (Ware, 2000).

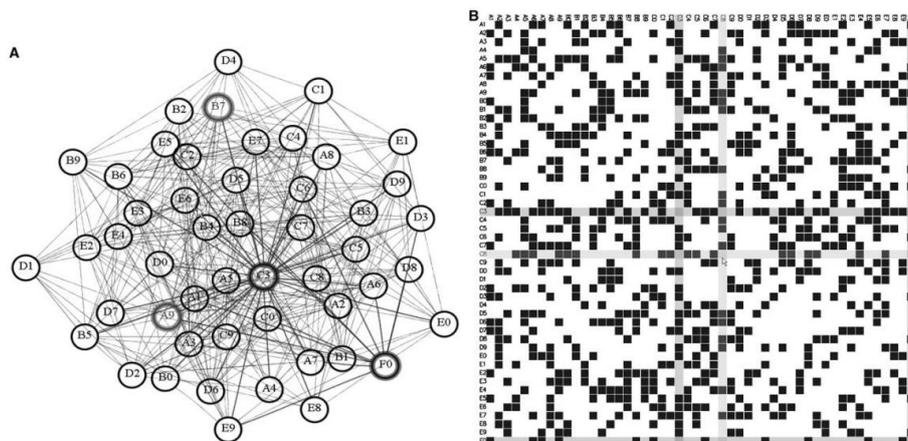


Figure 2-1 Two visualisations of the same undirected graph containing 50 vertices and 400 edges. The node-link diagram (A) is computed using the 'neato' program and the matrix representation (B) is computed using our VisAdj program (Ghoniem et al., 2005).

There is much research on network visualisation and dedicated conferences devoted to the topic (Shneiderman & Aris, 2006). Many visualisation techniques for network have been devised in recent years (Herman, Melançon, & Marshall, 2000). In a matrix-based visualisation, each cell shows the difference between two corresponding series: the row and the column. It is more obvious to see the relationship of adjacency node, however, it is difficult to see the paths or relationships from between nodes.

The literature on the network layout has been dominated by force directed strategies to produce the aesthetic spreading of the nodes and links. The force directed strategies are a class of techniques for drawing graphs in an aesthetically pleasing way, and tend to produce crossing-free layouts. In the network visualisation community, a variety of layout other techniques complying with the aesthetic rules has been devised. The layout techniques include minimizing the number of edge crossings, minimizing the ratio between the longest edge and the shortest edge, and revealing symmetries (Tollis, Eades, Di Battista, & Tollis, 1998).

Furthermore, a circular layout (Breitkreutz, Stark, & Tyers, 2003; Huffaker, Nemeth, & Claffy, 1999) can be adopted for nodes that produces an elegant presentation with crisscrossing lines through the centre of the circle. Multiple concentric circles are sometimes used. Its further variation, radial or egocentric layout, places an individual at the centre of a social network with closeness along radial lines to other nodes indicating strength of relationship.

However, Node-link diagrams has its own problems such as node occlusion, edge crossings, and edges tunnelling under nodes due to nodes and links having fixed places on the screen. On the other hand, spatial characteristics may become harder to perceive: it may be difficult to find nodes on a path and to identify clusters (Shneiderman & Aris, 2006). To avoid such problems, a different strategy has been proposed in (Ghoniem, Fekete, & Castagliola, 2004) to use matrix-based representations instead of node-link diagrams. Enhancement of network exploration with lists of nodes and links can clarify many tasks, especially when labels and attributes add helpful information (Lee, Czerwinski, Robertson, & Bederson, 2005). A hybrid of the two approaches has also been proposed by (Henry, 2007; Henry, Fekete, & McGuffin, 2007).

All of above aforementioned approaches organise the elements based on the linkage structure of the graph. Another approach is to plot the network data according to the attributes of the nodes (Lee et al., 2005; Shneiderman & Aris, 2006), with a scatter plot. Network links can be drawn between nodes. This approach is frequently used for assessing potential correlations between node attributes and network structure.

In related research, PaperLens (Lee et al., 2005) employs multiple coordinated views of network attributes to explore and analyse the dataset of conference paper publication. The NetLens system (Kang, Plaisant, Lee, & Bederson, 2007) generalises this approach to support networks that fit a “Content-actor” data model which allows users to pose a series of elementary queries and iteratively refine visual overviews and sorted lists. I.e., Bipartite networks such as publications and authors. This enables the support of complex queries that are traditionally hard to specify.

In Contrast to PaperLens and NetLens system, Orion (Heer & Perer, 2011) supports an arbitrary number of linking relationships both within and between data tables while facilitating integration with relational databases. Node types may be implicit within the attributes of a table; they provide methods to collect these values to generate a new table. Networks can be extracted from the foreign key relations among the tables. As display in Figure 2-2, analysts can import the data from multiple source, specify network model and generate graphic through drag-and-drop.

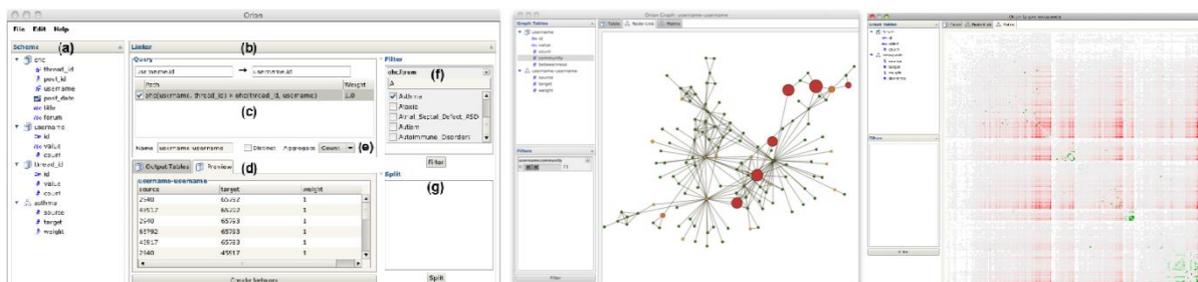


Figure 2-2: Orion Visualisations. From left-to-right: (a) User interface (b) Node-link diagram. (c) Matrix view of the connection. (Heer & Perer, 2011)

Orion's model and the aforementioned workflow language inspire our project in many aspects. Some notions such as the network visualisation and visual explorations using tabular, and node-link views were applied in our project. Moreover, we explored and evaluated the counterpart visualisation techniques in the prototype development.

Apart from node-link diagram and matrix representation, displaying the nodes on a map (Becker, Eick, & Wilks, 1995) to generate familiar and comprehensible layout is another common design strategy. Becker's research visualised the network links geographically; this matrix arrangement gives equal emphasis to all network links. (Silva & Catarci, 2000) present a survey on visual techniques for interactive exploration of time-oriented information. In a later research, (Viégas & Donath, 2004) incorporate temporal data into visualisation. Their visualisation depicts the temporal rhythms of interactions in social networks.

In our project, combining the geographic and temporal data into visualisation goes beyond the traditional graph paradigm and offer better insight for spend data.

2.3.3 Interactive visualisation

Interactive visualisation generated dynamic queries at the time when users are interacting with the interface. It allows user to have in-depth and real-time exploration of the data, and make new queries on the fly. There is a wealth of literature on navigation and exploration of information spaces. Dynamic visualisation can help users interact with visualised networks and have a direct way to explore and understand datasets. User study of the original dynamic query interface (Ahlberg & Shneiderman, 1994) supports the claim that 'tight coupling' facilitates the exploratory behaviour and take it as a principle for visual information seeking. For example, in Figure 2-3, with (alpha) sliders and genre buttons.

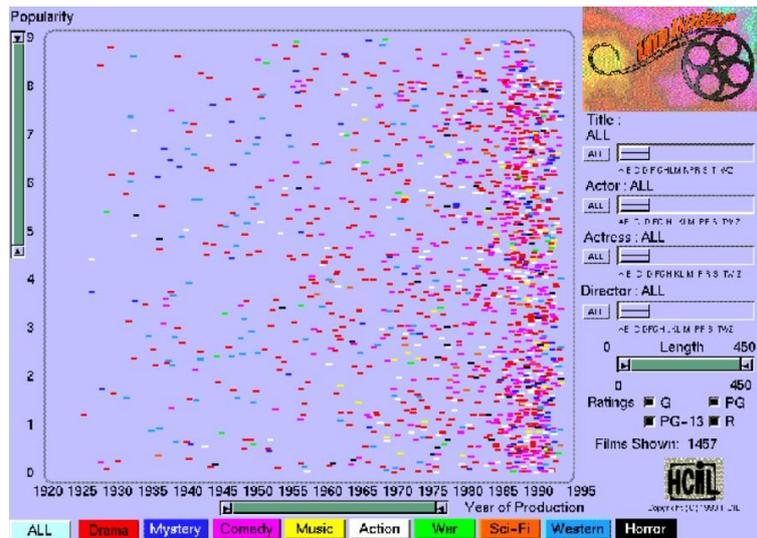


Figure 2-3: Dynamic queries in FilmFinder (Ahlberg & Shneiderman, 1994)

Advancing interaction design, Lowgren (Löwgren, 2007) introduced the term ‘pliability’ through Sens-A-Patch. This is highly related to tight coupling (displayed in Figure 2-4). He suggested that compare with static information visualisation, such function can be seen as a different interaction design genre. According to him, since *‘tight connection between action and outcome, the pseudo-tactile sense of manipulating the interface and shaping the information, the sense of being drawn into the material under exploration - all of this points to a rather highly involved and immediate experience at the focus of attention’* (Löwgren, 2007), pliable interaction engages the user, immerses them in the tasks.

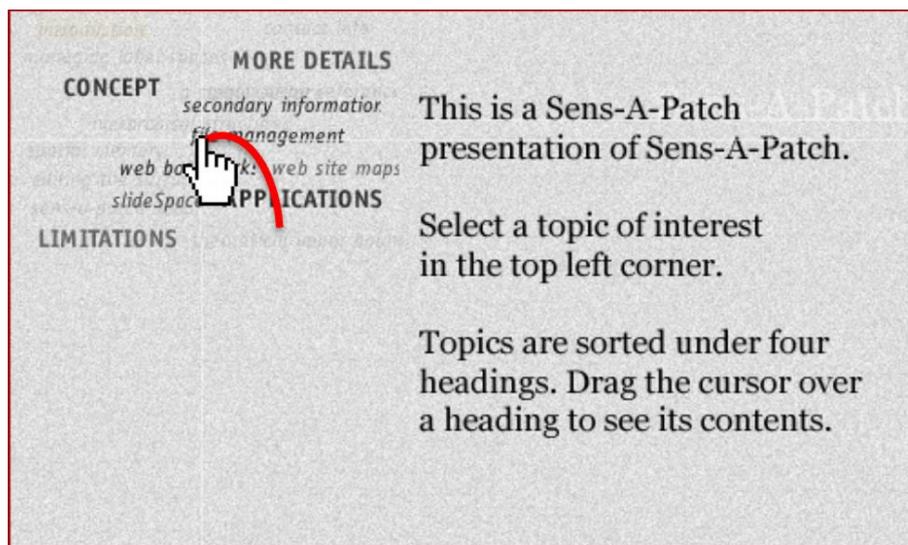


Figure 2-4. Pliability envisioned in Sens-A-Patch (Löwgren, 2007)

Referring to pliability and tight coupling, Spence (Spence & Press, 2000) comments in further research that ‘real-life problems invoke dynamic exploration thanks to lack of knowledge and

therefore formulating a problem is as important as solving it.’ According to him, visualisations that represent dynamic queries in a clear and relevant way enables user to see possible formulations and solutions of the problems on the fly.

In common web search interfaces, it can be fairly complex to formulate queries that simultaneously involve temporal, spatial, and topical data filters. Dork (Dork, Carpendale, Collins, & Williamson, 2008) also mentioned dynamic queries as an integral part of interactive visualisation, and argued that this genre improves the exploration of data with a swift and playful approach. Their prototype, VisGets, elaborates on coordinated visualisation for web-based information, explorations and discoveries (Dork et al., 2008). The prototype is web-based and offers multiple visualisations (Thudt, Hinrichs, & Carpendale, 2012) that facilitate temporal, spatial and topical data filters. The purpose of this prototype was to enable ‘information seekers orient themselves within online information spaces and to incrementally build complex filtering queries’ (Dork et al., 2008). As a result, the focus is more on interacting (exploratory search and filtering) than (explicit) serendipitous discoveries. As seen in Figure 2-5, their main contribution to interactive visualisation and interaction design is ‘weighted brushing’ to represent varying degrees of relatedness between items.

These works offered a variety of interactive techniques and design considerations to our research. In SpendWise, we designed and implemented a set of interactive components that allows users to formulate queries that involves temporal, spatial and topical data filters (see section 3.6.1.2 and section 3.6.1.6).



Figure 2-5: VisGets (Dork et al., 2008)

2.3.4 Scalability

ANZ's spend data is a large dataset with approximately 1,000,000 records. If we try to put it into one network graph, it would generate one million nodes which make nonsense to users. Problems occur when dealing with large graphs that thousands of nodes (too many edges crossing makes the graph unreadable). For example, our first attempt on network visualisation was failed (see Figure 3-5). Visual analysis tools need to handle the massive scale of network data. It is a challenge to visualise massive spend data, discover patterns, and enhance the bank's decision making process. The scalability problem is a protracted challenge for information visualisation.

Some common strategies, including filtering and aggregation have been proposed. Filtering involves using queries to reduce a large set of data to a smaller set, while aggregating is about gathering small sets of data together. Fisheye (Furnas, 1986) postulates an early concept for the display of large structures. The basic strategy uses a Degree-of-interest function which assigns to each point in the structure. It assigns a number to define how interested the user is in seeing that point in the current task. (Heer & Perer, 2011) introduced a degree-of-interest function that reduces a graph to a small connected subset of nodes based on an input set of foci (e.g., search result). PivotGraph (Wattenberg, 2006) aggregate networks by displaying of edges according to node attributes; for example, an analyst can collapse a social network of corporate employees to show the summed connection strengths between workers' geographic locations. As Figure 2-6 displayed, the PivotGraph diagram based on a roll-up view where node size corresponds to the number of nodes, and edge thicknesses represent the number of edges being aggregated. While the node-and-link diagram makes the topology clear, the PivotGraph diagram makes it immediately obvious that there are connections between genders/division pairs, with the exception of men and women in division2.

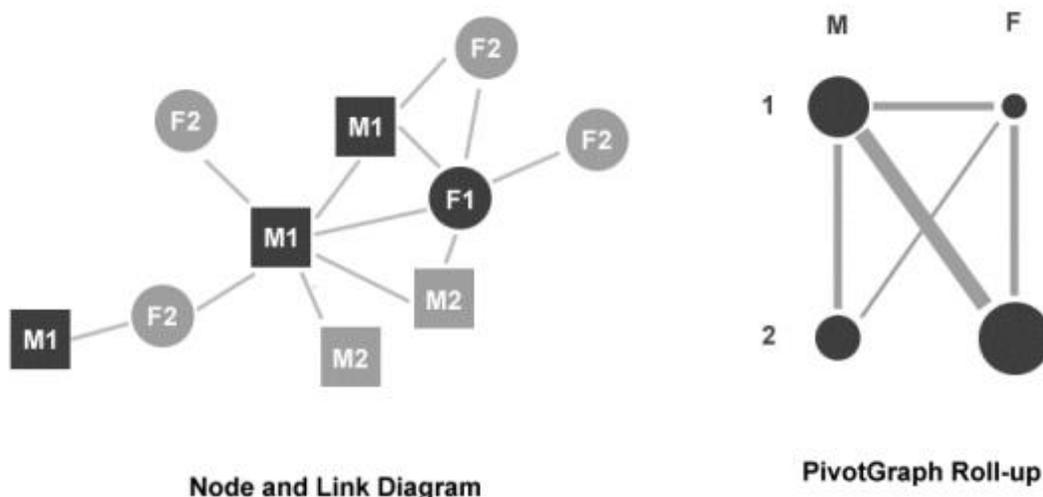


Figure 2-6: Node-and-link diagram versus PivotGraph (Wattenberg, 2006)

As displayed in Figure 2-7, Honeycomb (van Ham, Schulz, & Dimicco, 2009) groups the nodes in the network into clusters and renders the aggregated network instead. It is a matrix based tool for exploring large social networks using a similar approach to PivotGraph. Honeycomb also has the ability to perform temporal analysis. Temporal views provide more insight into how a network changes over time and allows us to create better network models and predictions. In our research project, we also incorporated temporal displays of spend data to gain insight into how the relationships between business units and vendors change over time.

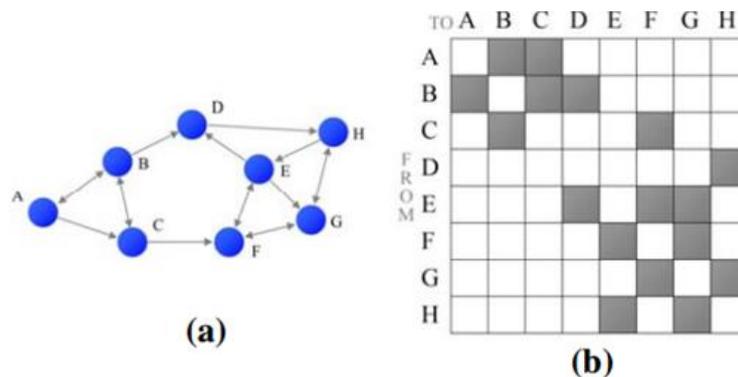


Figure 2-7: Social network visualisations. (a) A node link representation of a very simple 8 node network (b) An adjacency matrix representation of the same network, notes that the number of grey cells is equal to the number of links in (a). (van Ham et al., 2009)

ManyNet (Freire, Plaisant, Shneiderman, & Golbeck, 2010) offers a new approach that enables analysts to work on multiple networks simultaneously. It offers a comparison among multiple networks using a tabular view of summary graph statistics. Network can be presented as a row in the tabular visualisation with attributes such as edge density, number of vertices etc. It can then be inspected, sorted, and filtered based on these attributes. However, users can still view traditional node-link diagrams if they want.

2.3.5 Current visualisation on the Web

In our research, we choose web-based framework to develop our prototype. Since the advantages of publicity and accessibility, many visualisation tool has been developed on the web. Web visualisation is relatively accessible to a casual user familiar with a point and click interface.

Developed by IBM, Many Eyes (Eyes & Hoag, 2009) is a visualisation graphic user interface tool that is entirely online. It requires no download or install beyond Java and Flash. The dataset uploaded and visualisation created is automatically published publicly online which makes it easy to share and

communicate. However, it means it cannot be made private and there are limitations on accepted data format. The additional effort is often required to convert the data into the right format.

Tableau Public, based on Polaris, is an interface for the exploration of multidimensional databases that extends the Pivot Table interface to generate a rich, expressive set of graphical displays (Stolte, Tang, & Hanrahan, 2008). It was later commercialised as Tableau. It requires a client application to be downloaded and installed. Rather than accepting specific data formats, Tableau accepts an entire database and allows the user to explore the variables in the data via a variety of potential plots. As displayed in Figure 2-8, the dashboard allows us to place multiple graphics on the same page. In the right panel, Tableau Public suggests some appropriate graph types for the dataset. This work influence our project in terms of the layout style and selection of visualisations.

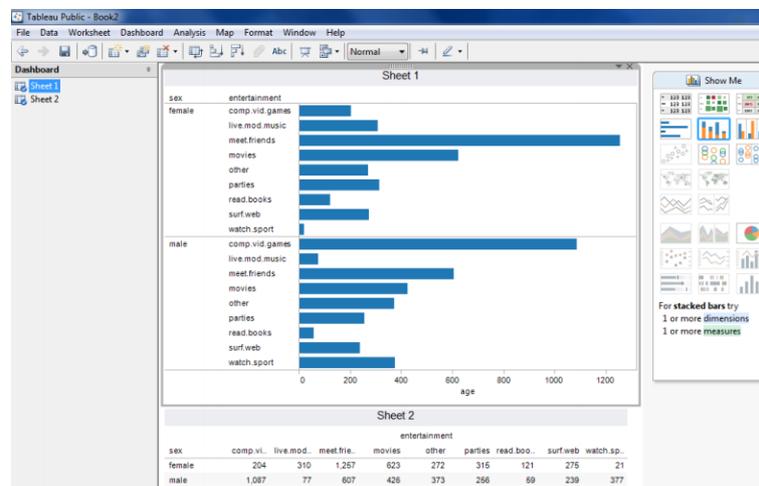


Figure 2-8: The Tableau interface

2.3.6 Other relevant topics

Our prototype implementation sits on top of novel technologies, such as HTML5, JavaScript, D3, and Hammer libraries. We employed the D3 library to present the query results to the scalable vector graphics, and evaluate the how effective D3 can be for in spend data analysis.

In this section, we review related technologies and software libraries that are beneficial to our prototype. Software libraries such as hammer.js and D3.js have simplified the development of interactive visualisations. Those libraries offer an extra layer of abstraction over the direct manipulation of graphical elements, allowing the developer to focus on the higher level application design.

2.3.6.1 HTML5 canvas element

The HTML5 Canvas Element can be understood to be a drawing surface for Raster graphics, where the drawing can be specified by scripts. Objects cannot be separately identified or modified once

drawn; dynamic graphics can be achieved by redrawing the Canvas from scratch. This also means that object-based interactivity is no easier than any other interactivity. The interactivity is accomplished via coordinate matching - the coordinate of the mouse is checked to detect if it 'hit' any 'object'. This hit detection is trivial for rectangular objects, but becomes more difficult for increasingly complex shapes. Another disadvantage is the raster property of the image makes it scale poorly. This can be overcome by forcing a redraw of the image.

2.3.6.2 Scalable Vector Graphics (SVG)

The Scalable Vector Graphics (Ferraiolo, Jun, & Jackson, 2000) is an image that is specified via a XML file with an extension of SVG. This standard can be considered to be a 'low-level graphics language' that can be used to draw anything from a rectangle to professionally typeset text. In addition, the SVG image can be scaled without pixelation. Thus, object-based interactivity such as clicking an object is very easy. However, any interactivity that encompasses multiple objects may still require coordinate matching to work. In our research, D3.js is the tools that use SVG.

2.3.6.3 jQuery

jQuery is a lightweight JavaScript library that offering rich features. It makes operations like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers (Chaffer, 2013). It offers a pluggable framework for controls and other libraries from its online repository (Wang, 2006).

jQuery libraries played an important role in SpendWise because it made the manipulation on DOM elements easy.

2.3.6.4 SignalR

SignalR is a framework that facilitates the creation of real-time applications, such as online collaboration tools, live information services, whose development has traditionally been quite complex (Ingebrigtsen, 2013). In order to support collaboration from multiple devices, SignalR was implemented in SpendWise to provide a real-time communication framework (see section 3.5 and Figure 3-3).

2.3.6.5 Data Driven Documents (D3.js)

Data-Driven Documents (Bostock, 2011), is a JavaScript library that uses digital data to drive the creation and control of dynamic and interactive graphical forms that run in web browsers. It allows the direct inspection and manipulation of the document object model (DOM) (Bostock, Ogievetsky, & Heer, 2011). D3 enables programmers to interact directly with SVG. Programmer can bind an external dataset to DOM elements and apply properties to them that allow the visualisation to be

interactive. This ability for the DOM element to be interactive in real-time largely reduces a developer's work as they can avoid enumerating all use cases when creating a visualisation. The D3 library improves browser-based visualisation in terms of accessibility, efficiency, and expressiveness. D3 is written in JavaScript and utilised by writing client-side JavaScript, which is the dominant scripting language on the internet and is relatively easy to learn (Goodman, Morrison, & Eich, 2007). Furthermore, D3 claims to have abundant documentation. The efficiency advantage is that learning and debugging in D3 is quick and easy. Thanks to direct interaction with DOM, the expressiveness of D3 is essentially unlimited. The developer *"has full access to the native representation"* (Bostock et al., 2011). D3 avoids all the unnecessary restrictive layers of abstraction and simplifies the tedious and error-prone nature of working directly with SVG elements, which helps to keep the runtime efficient.

D3.js was used to support all visualisations in SpendWise (see section 3.6).

2.3.6.6 NVD3

NVD3 (<http://nvd3.org/>) aims to build re-usable charts and chart components for d3.js. It is a small collection of components, with the goal of keeping these components highly customisable, staying away from the standard cookie cutter solutions.

NVD3 was implemented to visualise the spend history of a vendor company into a line chart (see section 3.6.1.4).

2.3.6.7 Hammer.js

Hammer is a small JavaScript Library that triggers multi-touch gesture events on a web page. It is simple to use with a jQuery-like API. You don't need to add a new keyword, and the instance methods are chainable. Events can be added or removed with the on and off methods, just like in jQuery. Event delegation is also possible when you use the jQuery plugin.

The following gestures are available in Hammer.js:

- Hold
- Tap
- Double tap
- Drag, drag start, drag end, drag up, drag down, draglift, drag right
- Swipe, swipe up, swipe down, swipe left, swipe right
- Transform, transform start, transformed
- Rotate
- Pinch, pinch in, pinch out

- Touch (gesture detection starts)
- Release (gesture detection ends)

Hammer.js offers an excellent application programming interface for multi-touch gestures in web development. However, to our current knowledge, it has not been developed for use in collaborative settings. SpendWise employed hammer.js to support multi-touch gestures such as drag and drop, pinch, zoom, rotate etc.

In summary, SVG via D3, with Hammer.js to supply sophisticated gestures, was selected for developing SpendWise (see section 3.7.1).

2.4 USER-CENTRED DESIGN METHODOLOGY

Due to the tight schedule and highly user-centric prototype development, we follow user-centred design (UCD) methods to develop our prototype. User-centred evaluation (UCE) serves three goals: detecting problems, supporting decisions, and verifying the quality of a product (Van Velsen, Van Der Geest, Klaassen, & Steehouder, 2008). These functions make UCE a valuable tool for developers of software systems to justify their efforts, improve upon a system or help developers to decide which version of a system to release. As a result, this may lead to higher adoption due to a more easy to use and pleasant user experience. Figure 2-9 links the goals of UCE to the phases of the iterative design process.

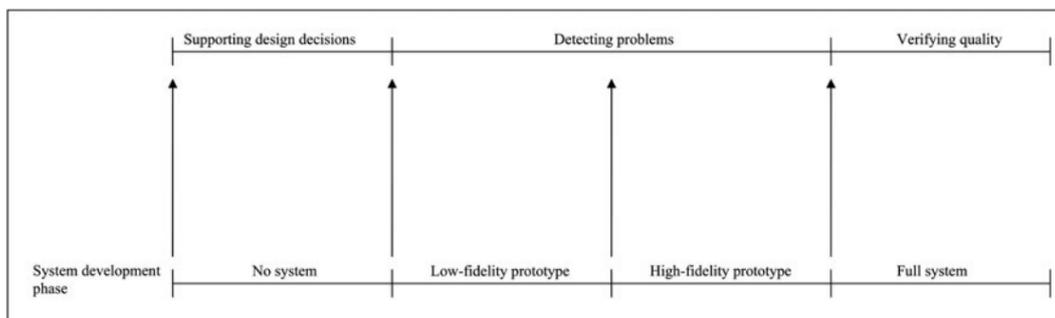


Figure 2-9: Phase of the iterative design process (Van Velsen et al., 2008)

Usability, perceived usefulness, and appropriateness of adaptation are the three most commonly assessed variables (Van Velsen et al., 2008). Questionnaires, interviews, data log analysis, focus groups and group discussions, think-aloud protocols, and expert reviews are frequently used methods in the evaluations. Currently, there is no consensus as to which evaluation methods are best suited for assessing the perceived quality of unique output of personalised systems. Questionnaires can be useful for measuring appreciation of personalisation, user satisfaction, and general opinions. The interview is a feasible method for assessing system adoption (van Velsen, van

der Geest, & Klaassen, 2007). It may also be used to access the usability issues that are typical for personalised system since it allows detailed feedback. Interviews may assist us in identifying problems, perceived causes and solutions.

Two common user-centred design methods are comparisons and laboratory or real-life observations. Comparisons try to identify the difference between a personalised version and one without personalisation. In the laboratory and real-life observation, a researcher watches a participant working with a personalised system, noting interesting events, or recording the whole session. Using a laboratory allows the researcher have better control of the environment. Outside influences can be excluded bringing focus to the variables they wants to assess. However, the downside of using a laboratory is loss of real-life interaction.

In order to evaluate a system, subjects need to interact with a prototype of the system. Otherwise, it is difficult for subjects to image how the system will work and how it will relate to their everyday activities (Weibelzahl, Jedlitschka, & Ayari, 2006). Types of prototype include working prototype, computer simulation, paper prototype, mock-up simulation, Wizard-of-Oz prototype, demonstration of the system. In our project, using the paper mock-up prototype at early phase of the system development to present idea and create a working prototype is cheaper than to create a full system, and it could help to verify the quality of the research result. Creating a prototype is a wise use of money and effort (Field, Hartel, & Mooij, 2001). However, the use of prototypes doesn't yield the same results as one would obtain using the full system (Field et al., 2001). An excessively simplified version may therefore not to be the right method for evaluation.

Rapid Iterative Testing and Evaluation (RITE) (Medlock, Wixon, Terrano, Romero, & Fulton, 2002), is an iterative usability method. This method finds a comparatively a high frequency of problems found and make fixes, and then empirically verifies the efficacy of the fixes. This method is highly effective in terms of finding and fixing problems. The tester and team must define a target population for testing, schedule participants to come into the lab, decide on what and how to collect and measure of users' behaviours. Once the data have been collected, the team would decide if they would make any changes to the prototype prior to the next participant.

In the next section, we explain how we follow RITE to iteratively develop SpendWise based on the feedback from domain experts.

3 SYSTEM DESIGN

SpendWise was developed by following RITE method, which requires frequent interviews and evaluations with the target users. User requirements are established, as well as personas, user scenarios, and design considerations are defined in section 3.1. Section 3.2 explains the system architecture. Section 3.3 and 3.4 explore the data types and load data into memory. Section 3.5 describes how SpendWise extract useful information from spend data. Section 3.6 demonstrates all the visualisations in SpendWise and the design choices we made throughout the research. In section 3.7, approaches for collaborative scenarios were implemented in such ways that support group collaboration on a tabletop or from multiple devices.

3.1 ESTABLISH REQUIREMENTS FROM DOMAIN EXPERTS

Feedback from domain experts was collected at all stages of the prototype development. Earlier design ideas and prototypes were evaluated in an interactive process. The results were used for improvements. Throughout the development phase, five meetings were held to gain feedback from three ANZ contacts (Adam Hart, Catherine Thompson, and Melissa O’Neill). At each meeting, several new visualisations were proposed to ANZ and the domain experts offered their feedback on the improved prototype. By collecting feedback, we started to learn more about the spend data and user requirements and were able to incrementally improve our prototype.

Since the prototype was developed gradually, early visualisations were static or implemented on the static JSON data. After several meetings, we had a greater understanding of the target users and what visualisations would be meaningful for them. Three personas were defined including data analyst, domain expert, and non-expert user.

Persona	Description
David (Data Analyst)	A person has a strong technology background or database knowledge.
Adam (Domain Expert)	A person has an authority on the spend data analysis.
Patrick (Non-expert User)	A person without specialised knowledge in a data analysis.

Table 3-1 Operation supported in SpendWise

According to the personas in Table 3-1, several user scenarios and scenario components were designed for component design.

Scenario 1: Interactive scenario

	<p>Scenario</p> <p>David needs to understand the spend data of a specific business unit: Technology. He opens the interactive panel of SpendWise and browses the organisational tree map to find “Technology” at the second level. He starts dragging the “Technology” icon from the tree map into different visualisations to see how the information is translated by the different parameters in these visualisations. He wants to understand which vendor companies are associated with this business unit and where this business unit spends its money. He then drops the “technology” icon onto the network view. The network view updates itself automatically with a list of nodes which represent vendors and spend categories. He selects a node by clicking on it and a filtered view is generated to display only this node and its related nodes (see Figure 3-12). Each node is labelled and looks like a bubble. The size of the node reflects the spend figure. He then drags the "Technology" icon from the organisational tree map and drop it to another view which display geographical information. This map returns a layer of translucent bubbles to present the location of related vendors and the total spend of the area. He then places a time constraint to update all visualisations. When he clicks a specific vendor node in the network view, a line chart view shows the spend history of the selected vendor.</p>
<p>Scenario Components</p>	<p>Organisational tree Map</p> <p>A view of spend summary by the level of ANZ organisation hierarchy. Users can quickly check the spend total of each business unit at different levels (see Figure 3-6)</p>
	<p>Network view</p> <p>A network view of the spend schema of the business unit and its related vendors. This also carries the information of the total spend amount and transaction details (see Figure 3-12).</p>
	<p>Geographic spend heat map</p> <p>A geographic view that displays the spend data of each area on the map (see Figure 3-9).</p>
	<p>Spend Detail line chart</p> <p>A line chart that displays the detailed spend history in the previous fiscal year (see Figure 3-16).</p>
	<p>Time range selector</p> <p>A slider gadget that enables users to filter all data into a specific time period (see Figure 3-14, Figure 3-6).</p>
	<p>SVG</p> <p>A communication feature enables users to pass over the visualisation</p>

	communications	parameters and make queries by drag and drop (see section 3.6.2).
--	----------------	---

Table 3-2 Scenario 1: Interactive scenario

Scenario 2: CFO Summary scenario		
Scenario		Patrick needs to understand the spend data of the whole ANZ group. He opens the summary panel of SpendWise. He clicks the PLAY button to watch a global map change along with each month. He then wants to understand spend trend in the area of Sydney. He searches the city name in the input box. The map only displays the spend data of Sydney. He then moves the cursor to the Melbourne on the map, and the map only displays the data of the Melbourne area.
Scenario Components	Play Button	An animation to display the spend data in a temporal format.
	Geographic spend heat map	A geographic view that displays the spend data of each area on the map (see Figure 3-19).
	Spend trend line chart	A line chart that displays the spend trend in the previous fiscal year (see Figure 3-19).
	Search	An input box where user can conduct searches (see Figure 3-19).

Table 3-3 Scenario 2: CFO summary

Scenario 3: Collaborative interaction		
Scenario		(All following operations were designed to support touch gestures) Patrick, Adam, and David need to work together to explore the spend data. They lock themselves into a meeting room and use a tabletop. Patrick focuses on the network view while Adam is checking the organisational tree map. Then Adam rotates the tree map and shows his findings to Patrick. Patrick then drags the business unit that Adam was working on from the tree map and drops it onto the network view. The network view updates quickly.
Scenario Components	Touch	Touch enables analysts use touch gestures to control the SpendWise.
	Tap	Tapping explores data by changing the level of the graph e.g. (Links together related nodes and filters out unrelated nodes in network visualisation)
	Zoom	Enlarging/Zooming an SVG graph
	Pinch	Using pinch to rotate the SVG
	Drag/Drop	Dragging from source SVG, and drop it on target SVG to make queries

Table 3-4 Scenario 3: Collaborative interaction

Scenario 4: Collaboration from different devices.

Scenario		In the meeting room, Adam uses his own laptop (a private workspace) to analyse the data, whereas Patrick and David work together on a tabletop (a shared workspace). Adam finds some results and he clicks send to the public. Patrick and David immediately see the update on the public work space.
Scenario Components	Private Workspace	A private device with a modern browser (normally a laptop, iPad or personal computer)
	Public workspace	A shared workspace allows 2-4 users to work together (normally a large tabletop display)
	Real-time Communication framework	A framework enables real-time communication between multiple machines.

Table 3-5 Scenario 4: Collaboration from different devices.

The following sections explain how we design the architecture to meet these system requirements.

3.2 SYSTEM ARCHITECTURE

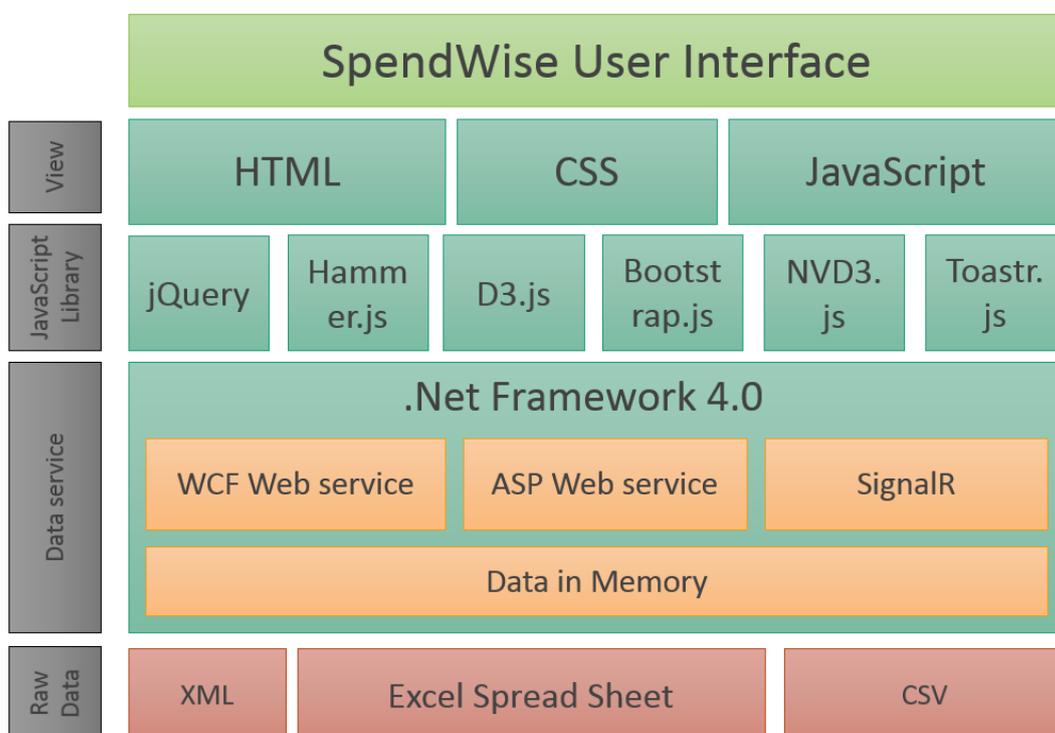


Figure 3-1: Prototype architecture

Two main software platform options are available to build a prototype for spend data visualisation: on a local computer or web application. In SpendWise, we chose to employ a web application for the main advantages listed as follows. First, web applications are easy to maintain and can be used on any system that has a web browser. Second, web applications are well-suited for collaborative scenarios, whereas local hosts do not meet the requirements of multi-device collaboration. Third, the visualisation techniques (see section 2.3.6) we intended to explore were JavaScript-based libraries.

Figure 3-1 shows that the system architecture is composed of four layers: raw data, data service, JavaScript libraries, and view. The detailed functions of each layer are introduced in the following sections.

3.3 DATA EXPLORATION

The dataset was in the format of an excel spreadsheet. We retrieved location information data in the semi-structured format (e.g. XML files) and Comma-separated values (CSV).

The dataset contains all spend transactions from ANZ in the fiscal year 2013. The data shown in the visualisation has been improved by filtering dirty data. For example, transactions without a specific vendor ID have been removed to clear the dirty data.

3.4 DATA PRE-PROCESSING

In the system initialisation phase, SpendWise loads the data from different file formats such as CSV, XML, and Excel spreadsheet. To display spend data on the map, geographical data including latitude and longitude need to be collected from internet.

Since no geographical data (latitude and longitude value) were provided from ANZ raw data, we need to develop a method to retrieve the geographical data. We managed to download the public domain geographical dataset from Natural Earth Data website (<http://www.naturalearthdata.com/>). Then the geographical data were loaded into the memory and mapped by the postal code of vendor address.

3.4.1 Cleaning the raw data

Not every row of data was able to be processed as tabular data. Before loading the raw data into memory, the data needed to be filtered to make sure it did not contain dirty data, such as excel formula. Furthermore, some data fields were processed in different ways (i.e., string may be trimmed or converted to DateTime format).

3.4.2 Loading data

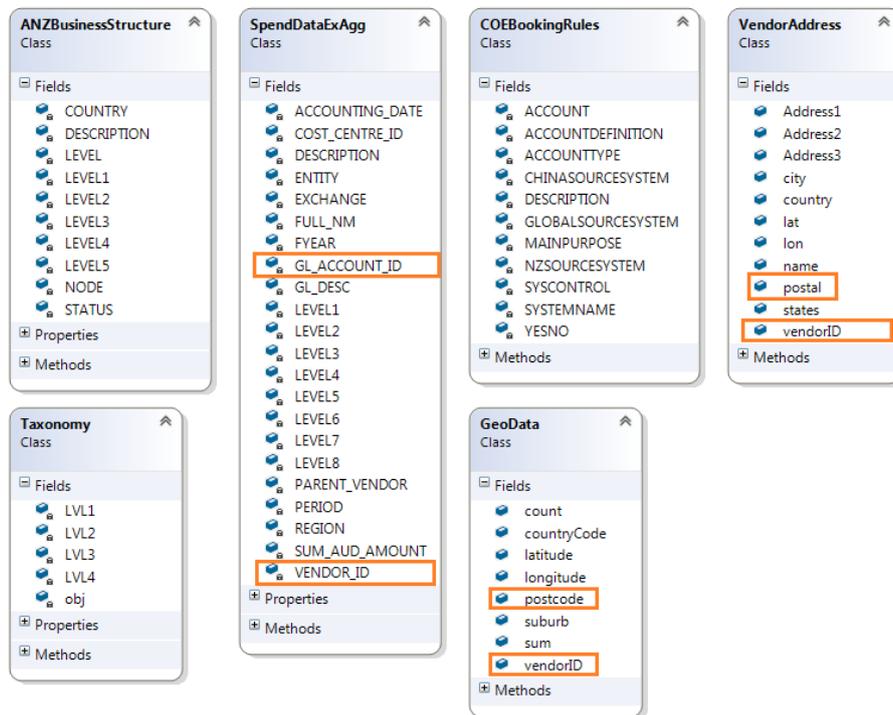


Figure 3-2 Prototype data class diagram

The raw dataset is about 1 million transactions which was about 96 MB size in Excel spreadsheet. Thus, it can be loaded into the memory rather than installing a database. The data was loaded in the form of a List or Dictionary (see Figure 3-2).

In Figure 3-2, we created 6 classes to source data from 3 different data files and to map all the tables. Each row in the raw data matches an instance of a relative class. A supportive .NET library, Excel Data Reader (Ian1971, 2013), has been used to load Excel files into the memory.

ANZBusinessStructure is a list of hierarchical structures with unique IDs of each business unit. The visualisation of this hierarchy table is discussed in section 3.6.1.1.

Taxonomy is a list of all the four levels of organisational taxonomy. The visualisation of this hierarchy table is shown in section 3.6.1.1

SpendDataExAgg stores about 1 million organisational spend transactions from the fiscal year 2013. This is the main data table of the dataset. Almost every visualisation needs to extract information from this table and mapping related information to other tables. Visualisations of this table can be found in section 3.6.1.5.

GeoData is a list of the geographic data we have retrieved from other data sources. GeoData was used in the geographic heat map (see section 3.4).

The VendorAddress class keeps all the vendor information in a separate file. The vendor address table has been used for the geographic heat map and CFO summary view (see section 3.6.1.2 and 3.6.1.6).

The COEBookingRules class stores the account definition and description of business units. This table was not visualised due to insufficient data fields.

The highlighted data fields in Figure 3-2 are unique IDs that represent a row of another table for referencing purpose when extracting data from multiple tables.

In section 3.5, we explain how SpendWise extracts information and the relationships between given tables in each visualisation design.

3.5 DATA SERVICE

After reviewing several technologies, ASP.NET web services (Bai, 2010) were chosen to write the data services layer. However, it proved to be insufficient because it wraps all the data to SOAP format, but some JavaScript libraries require JavaScript Object Notation (JSON) format and CSV. Therefore, Windows Communication Foundation (WCF) (Mackey, 2010) was added into the framework to handle the requests from the user interface, perform the relative analysing tasks, and return the results. The data formats requested by the visualisation are normally CSV and JSON. JSON allows for the serialisation of a data structure to be transmitted between applications. JSON data can easily be parsed into a JavaScript object which can then be projected to the visualisation. Comma-Separated Values (CSV) is another format which is often used in the D3 visualisations. While JSON works for a variety of data types, CSV has the highest efficiency to transmit tabular data. Additionally, CSV format enables data to be easily downloadable for other software packages. Thus, power users can download data and use it in other tools such as Excel, web applications, or MATALAB.

Group collaboration from multiple devices requires real-time synchronisation between the devices. Web Sockets and long polling are the popular methods of communicating and displaying real-time data. Microsoft released SignalR (Aguilar, 2014) to its web development technology stack to facilitate the creation of real-time communications (see section 2.3.6.4). SignalR was applied to the data service layer to offer real-time bi-directional communication between server and client (see Figure 3-3). In addition, it provides both web sockets and hub functionality which allows asynchronous multicast communication via HTTP protocol. Basically, numerous clients can connect to a SignalR Hub. When one message has been sent to the hub from a device, SignalR broadcasts it to all clients that are connected to the hub.

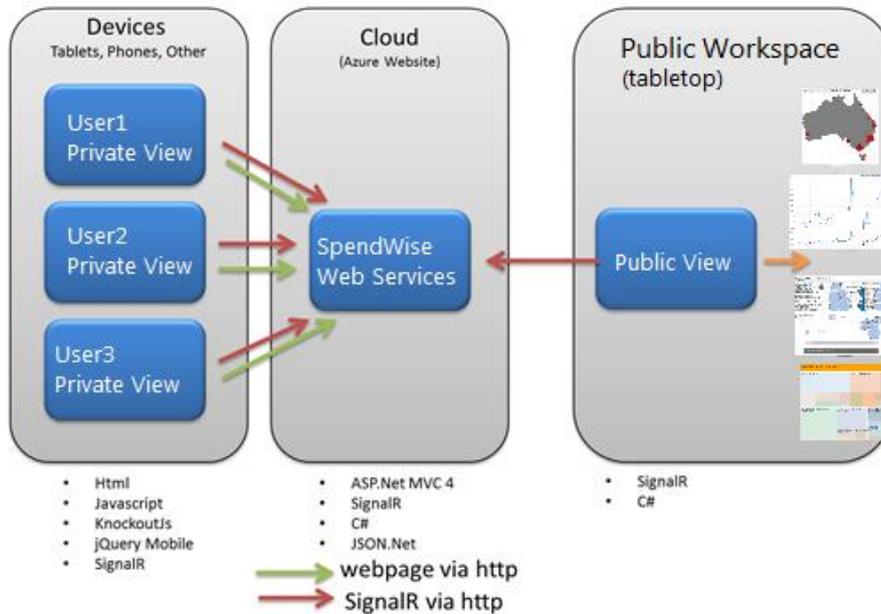


Figure 3-3: Real-time communication framework in SpendWise with SignalR

SpendWise contains a SignalR JavaScript Client to allow a client to establish a Web Sockets connection to the SignalR hub. Once connected to the hub, they receive broadcasted messages from all connected clients. Figure 3-4 shows the message sent from an individual device. It contains the user id and the parameters of the visualisation such as height, weight, query string, and data service.

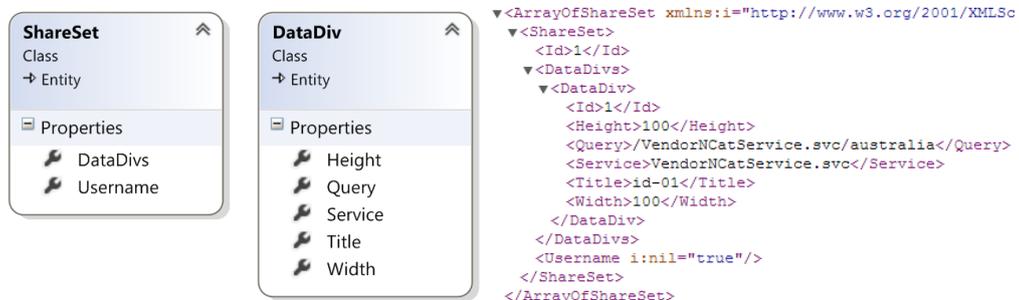


Figure 3-4: Passing message structure by the SignalR

3.6 DATA PRESENTATION

As mentioned in section 2.2.5, we decided to develop a prototype that sits on top of HTML5, CSS3, and JavaScript. Thus, visualisations in SpendWise are JavaScript based web components written using the powerful D3 visualisation library (see section 2.2.5).

The JavaScript libraries used in SpendWise are listed in Table 3-6.

Library	Purpose
jQuery	DOM element manipulation
D3	Brings data to life using HTML, SVG and CSS
Hammer	Offers Multi-Touch support
Bootstrap	Offers design templates for typography, forms, buttons, charts, navigation and other interface components.
NVD3	Reusable charts for d3.js
Toastr	JavaScript library for Gnome/Growl type non-blocking notifications

Table 3-6 JavaScript libraries used in SpendWise

3.6.1 Views and design choices

In the early phase of development, we made attempts on a variety of visualisations (see Figure 3-5). In the first four months, we took the improved prototypes to the monthly meetings with ANZ and asked domain experts to evaluate several new visualisations. Figure 3-5 (upper left) shows the network visualisation of ANZ's business structure in which each node represents a business unit and each linkage represents their superior or inferior level. Apparently, domain experts got scared of the overcrowded network views. Consequently they suggested to simplify it. Figure 3-5 (upper right) shows a simplified version of the business structure network. As a result, early network view failed to show the relationship between data entities. Accordingly, we started to seek better presentation for the network data.

Domain experts often analyse spend data by business unit. There were 8 levels of organisational hierarchy in the SpendDataExAgg table (see Figure 3-2). Inspired by the feedback, we developed a tree view which enables users to explore their organisational structure by clicking on the nodes of the tree view. A rectangle tree map was implemented to show the spend details of selected business unit. This design change offered better user experience compared to the overcrowded network visualisation.

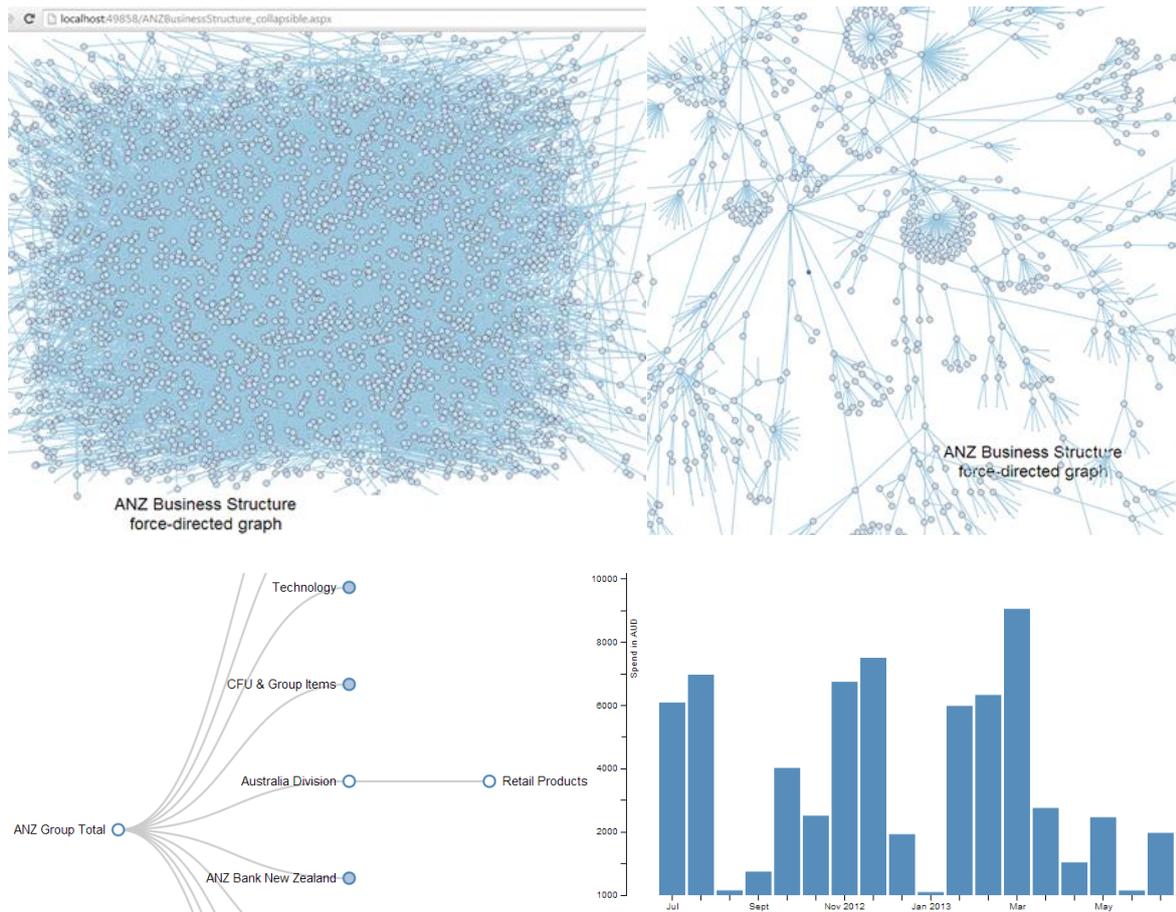


Figure 3-5: Early visualisations: organisational force-directed graph (upper left), simplified force-directed graph (upper right), ANZ hierarchy (bottom left), Vendor spending history (bottom right)

On the basis of the proposed scenarios and attempts on previous visualisations, the final version of SpendWise was designed with the following user interfaces.

3.6.1.1 Spend tree map

Domain experts are really keen to analyse data from an organisational hierarchy perspective. They often compare the total spend of one business unit with others in the same level of hierarchy. Therefore, we made an attempt of a tree map view as displayed in the Figure 3-5. We found the design of Figure 3-6 could have better ways to explore and present the spend data.

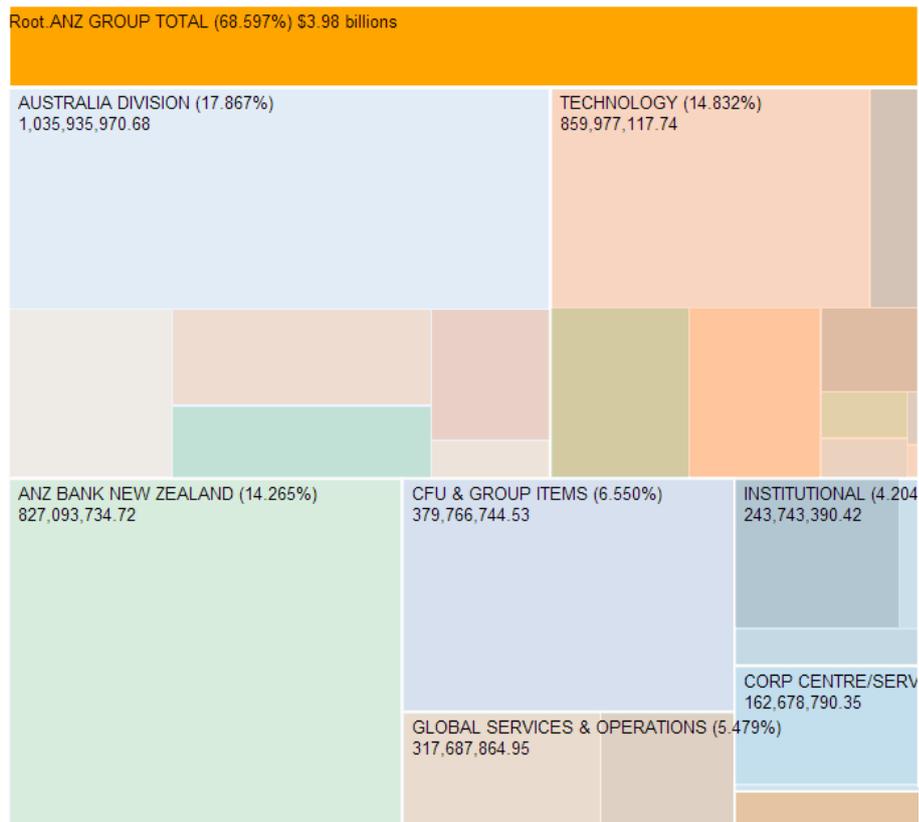


Figure 3-6: Spend tree map - Data display based on ANZ business units and its total spends

Treemapping (Shneiderman, 1998) is widely used for displaying corporate hierarchical data with its spend figures by using nested rectangles. The spend tree map displayed in Figure 3-6 allows hierarchical representation of the overall organisational spend data. Data for the organisation as a whole remains on view in the top bar, whilst displaying data for successive subordinate business units as a proportion of the whole. Figure 3-6 shows the spend tree map expanded through two levels allowing concurrent visualisation of two further levels of subordinate business units. Users can change the level of detail shown on the fly as they pursue queries. The size of the business unit rectangles proportionally represent spend figures.

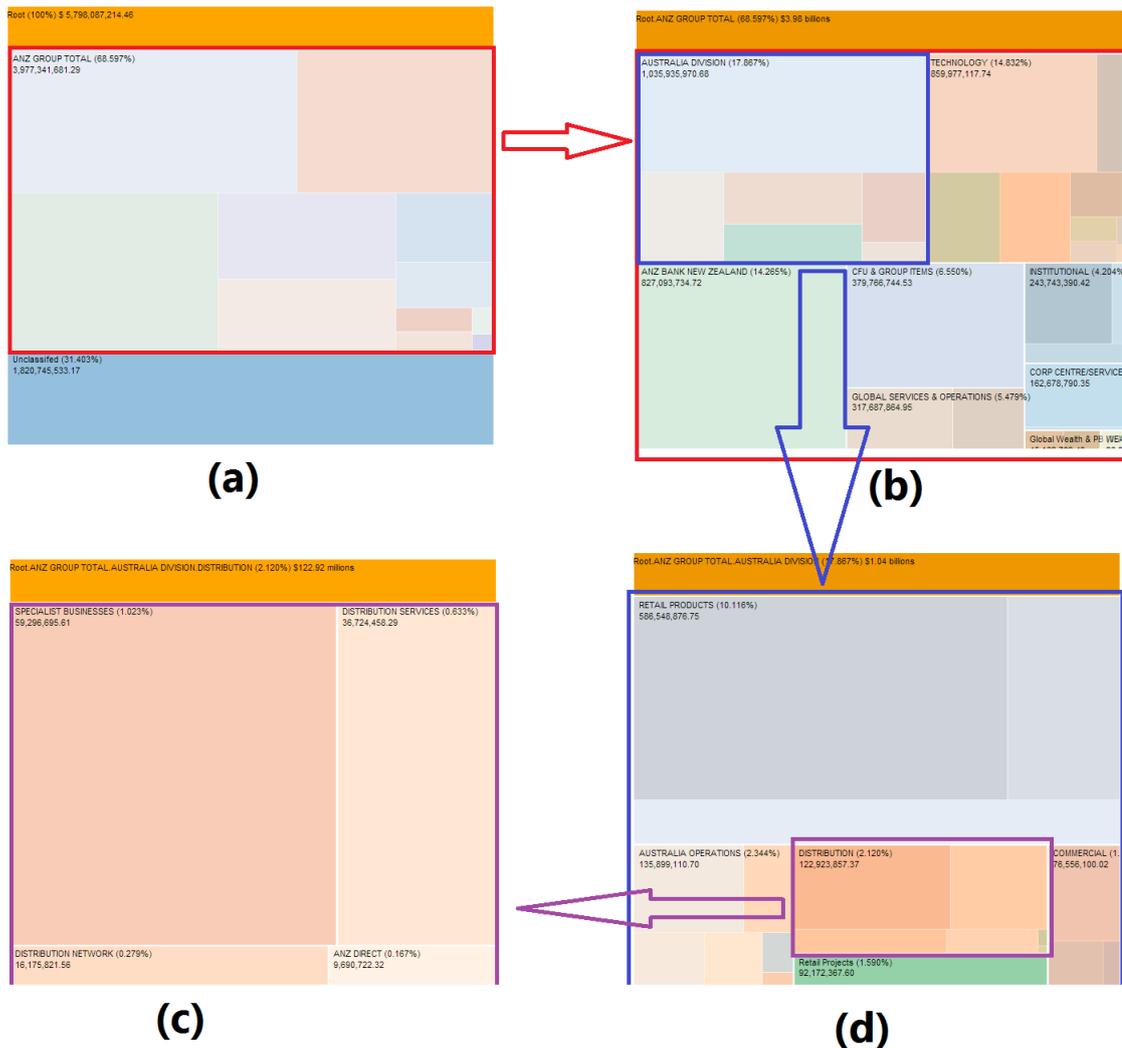


Figure 3-7: Spend tree map - Data exploration by clicking the tree map

Figure 3-7 demonstrates the process of exploring data in the spend tree map of the organisational hierarchy. By clicking a business unit in Figure 3-7 (a) , it drills down and updates the business unit to display the next level of business units as displayed in Figure 3-7 (b). Users can drill down according to the organisational hierarchy and go back to the upper level by clicking the top rectangle. The level of organisational hierarchy is separated by the breadcrumbs.

SpendHierarchyWebService (see Figure 3-7) was created to extract hierarchical information from raw data and create a tree structure of the hierarchy. In the initialisation phase, it goes through the spend data and adds tree nodes and its linkages into the Dictionary (C# data structure). Then it calculates the percentage of total spend and generates a spend summary for each node. When a user clicks on a rectangle on the tree map view in Figure 3-7 (a), JavaScript sends an asynchronous request to the web service with parameters including the name of the business unit and its level in the organisation. The web service generates the tree node of the selected business unit and returns

its subordinate nodes in JSON format. Finally AJAX update the tree map once it has received the response.

The layout of tree map has been proposed by Bostock et al., (2011). Several attributes were populated on each node including: parents, children, value, depth, etc. as shown in Figure 3-8 (left). It is convenient to load the node hierarchy using d3.json, and represent the input hierarchy as a nested JSON object. Figure 3-8 (right) shows a fragment of JSON format of the HTML response. The value was used to set the rectangle of each node proportionally to the spend figures. The percentage was calculated to reflect the percentage of the selected business unit spend in terms of the whole company.



Figure 3-8: Left - Data service - Spend hierarchy web service Right: Response format

3.6.1.2 Geographic heat map

Geographic visualisation facilitate user's understanding on the combined view of spend data and geospatial data. It is more direct to understand. Spatial data serves as the common variable to link various datasets together. In the early stage of feedback collection, Analysts showed a strong interest to understand the data displayed in the geographic view. To understand the spend in a geographic context (see section 1.3), we selected D3 map (Bostock et al., 2011) to visualise the spend data. Figure 3-9 displays spend data in an Australian map. The map has a customised pop-up to display town name, postcode and total area spend when hovering the mouse over the map. The radius of bubbles reflects the spend amount of the area. The map can scale to any size since it is SVG based. In addition, it has live updating of bubbles when users make new queries. The map can be updated by operations such as search, drag and drop.

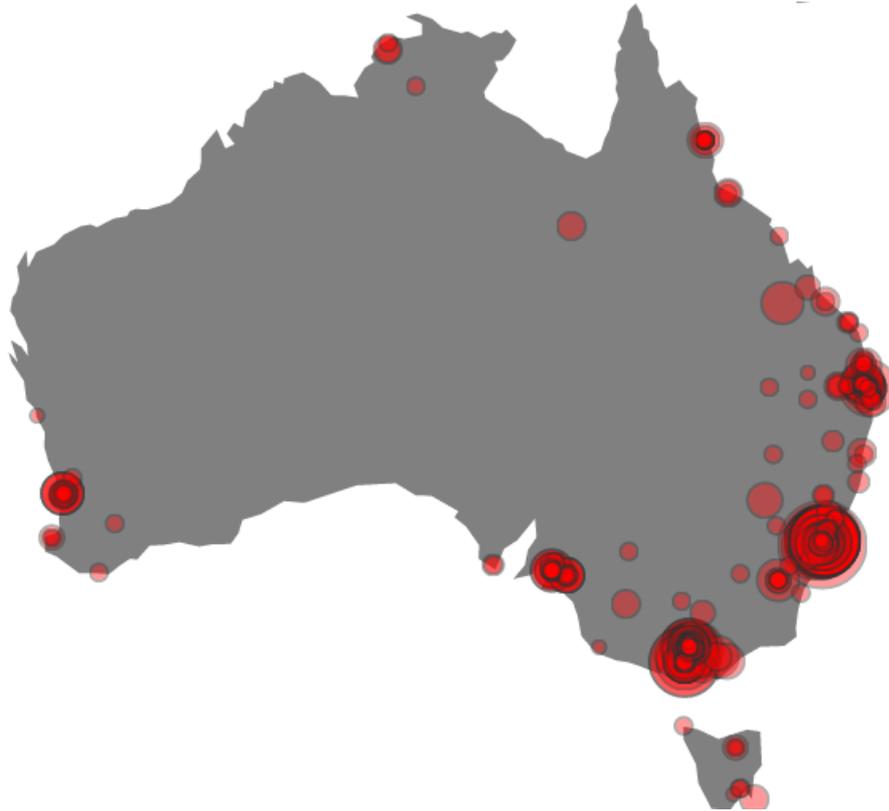


Figure 3-9: Spend Heat map – geographic spend data visualisation

Once users make a query, AJAX sends an HTTP request with several parameters (i.e., the name of business unit, start date, and end date). By default, the time range was set as start and end date of the fiscal year. The data service extracts the vendors of the selected business unit, calculates its total spend, and then maps the list of vendor addresses with GeoData. It then returns a list of results with the geographic data, postcode, town name, and the total spend figures. The invocation process was asynchronous.

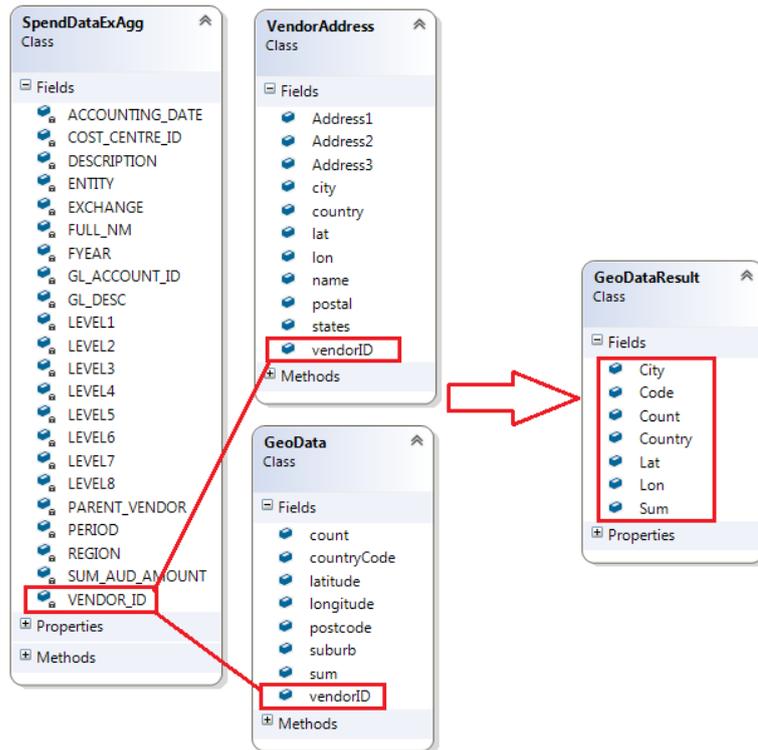


Figure 3-10: Data service: Geographic spend web service

ASP.NET web service returns a CSV format file (see Figure 3-11) to the user interface. Then D3 loads the data and updates the projection of each node on the geographic map. Circles can be sized by the amount of spend. However, if the range of sizes become too big, the chart would be cluttered and unreadable. To tackle the cluttering issue, spend figures were scaled from 1px to 30px. An update function has also been developed to clear the current data projection and load new data on the interface.

code	city	country	lat	lon	sum	reserve
AU	NORTH SY	2059	-33.8028	151.1049	996541.68	ReservedStr
AU	ST LEONAI	2065	-33.8282	151.1966	981227.33	ReservedStr
AU	SYDNEY	2001	-33.7949	151.2681	24025860.91	ReservedStr
AU	SOUTH ME	3205	-37.9329	145.0337	882915	ReservedStr
AU	BROOKVA	2100	-33.7677	151.2661	5200313.4	ReservedStr
AU	NORTH SY	2060	-33.84	151.1958	468384.2	ReservedStr
AU	MACQUAF	2113	-33.7968	151.1244	486595.42	ReservedStr
AU	LANE COV	2066	-33.8256	151.1599	845163.85	ReservedStr

Figure 3-11: Data Service: Geographic data format

3.6.1.3 Vendor and spend network

Vendor and Spend category

click a node to focus, drag and drop to make new query.
drop here!

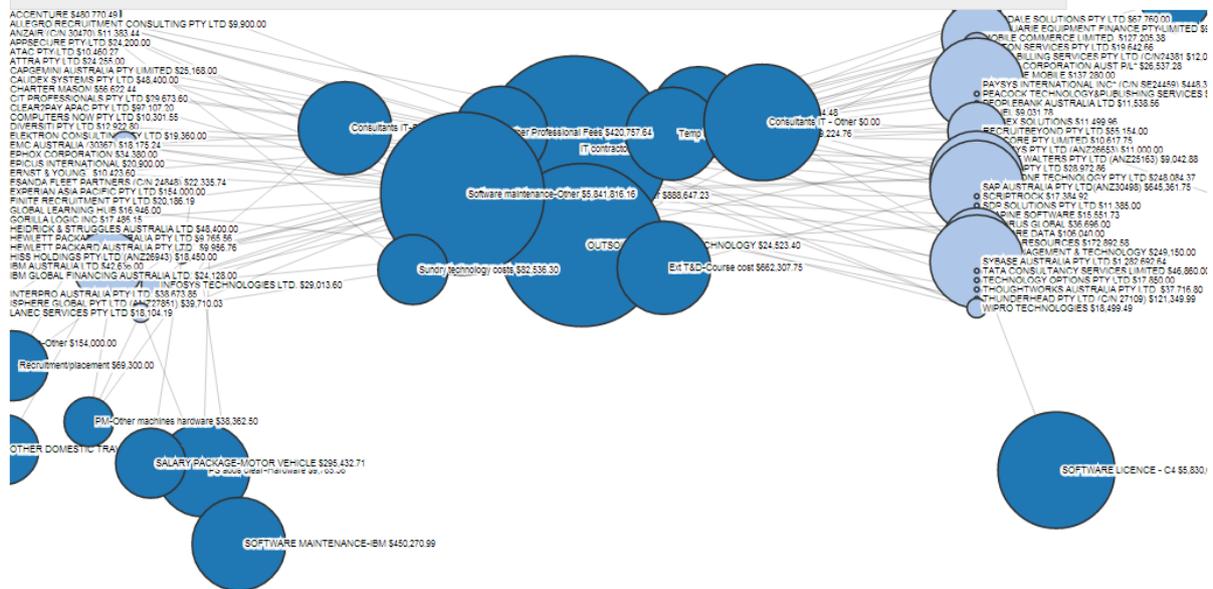


Figure 3-12: Vendor and spend network visualisation - the relationships between business unit, related vendors, and spend categories.

Network visualisation offers an intuitive way to visualise connectivity in the spend data. It can offer a direct view of the relationships between entities such as business units, and vendors. Domain experts are required to identify spend patterns from the relationships between business units, vendors, and spend categories. A vendor and spend network can be easily understood and explored in a network format, using vendors as nodes, relationships as edges and additional information (e.g., spend figures, vendor name, spend category name, etc.) as properties. Figure 3-12 was inspired by (Bostok) which is a force directed node to visualise the spend categories and related vendors of the selected business unit. The diagram was used to browse the spend categories by connecting vendors, rather than hierarchically. Users can make real time queries for a business unit and update the networks.

When hovering on a particular node, related spend categories and vendors are highlighted in red as displayed in Figure 3-13 (upper left). Users can focus on a specific node by a single click which filters out unrelated nodes as displayed in Figure 3-13 (upper right and bottom left). More relationships can be added with a single click on the others nodes as shown in Figure 3-13 (bottom right). This enables users to gain insight on the spend categories and related vendors. In addition, users can update line chart (see Figure 3-16) by double clicking on any vendor node.

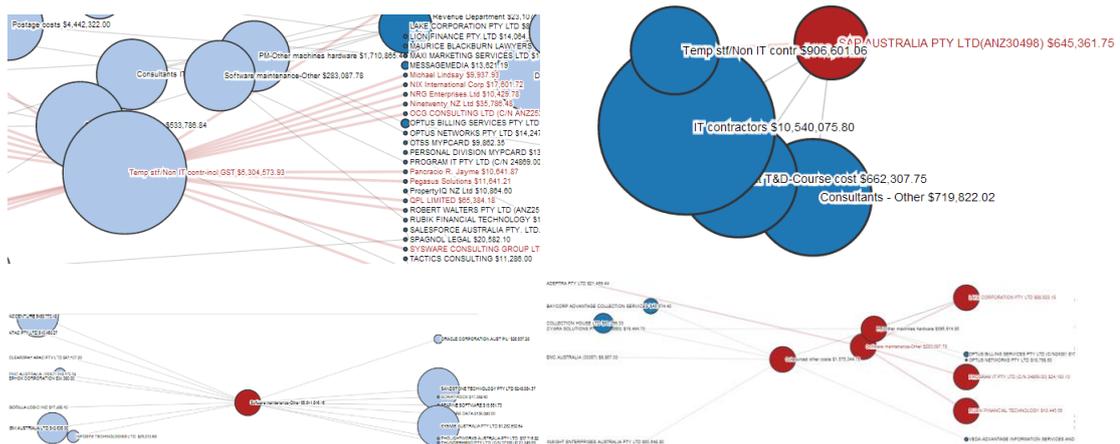


Figure 3-13: Hovering (upper left); selecting a vendor (upper right); selecting a spend category (bottom left), finding related vendors for multiple spend categories (bottom right).

To filter data by temporal properties, a time range selector was added to the interface. According to Kapler & Wright (2005), time slider is the most intuitive way for users to specify a time range for a dataset. In SpendWise, jQRangeSlider.js (Derrick Snider, 2013) has been implemented to constrain data into a timeframe (see Figure 3-14). Users can select a time frame to update the network visualisation and geographic heat map. With a touch-enabled device, users can select a range by touching and swiping on the screen. They can drag and drop to move fixed nodes which represent vendors and group them in new ways.



Figure 3-14: Time Range Slider enables user to filter data in a specific time range.

Queries can be made by clicking or drag and drop. Each query includes the name of the selected business unit, the start date and end date. The time range is set using the time range slider (see Figure 3-15). Then AJAX sends an HTTP request to VendorNCatService (written in WCF). It filters the transactions and lists the related vendors of the selected business unit. In the meantime, it retrieves the category of the transaction, and lists this spend information under IT contr each node (vndor). Figure 3-15 (left) shows the response format of each query. The JSON format response data is an array of spend categories which contains an list of vendors. The network view makes connections between topics by finding vendors that are common across each spend category array. The key is a unique value that is used for matching spend categories. The size of the node reflects the spend figure. It is worth taking into account that scale function has been used to map from an input domain to an output range. This avoids the circles from overlapping. However, the scale function needs to be polished as overlapping is still occurring.

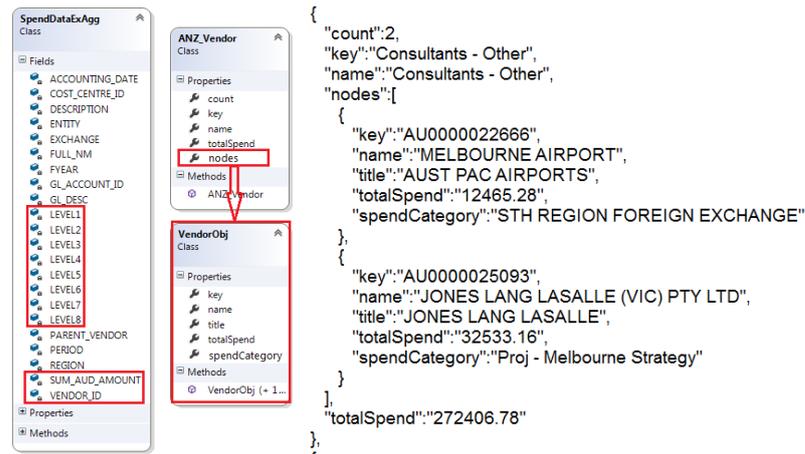


Figure 3-15: Data service: (left) Vendor and spend category (right) response format

3.6.1.4 Detailed vendor spend history

In the spend data analysis, it is critical to narrow down the analysis into specific details. SpendWise enables users to break down analytical data into specific transactional detail. A line chart was best used to track the spend data of a specific vendor over time.

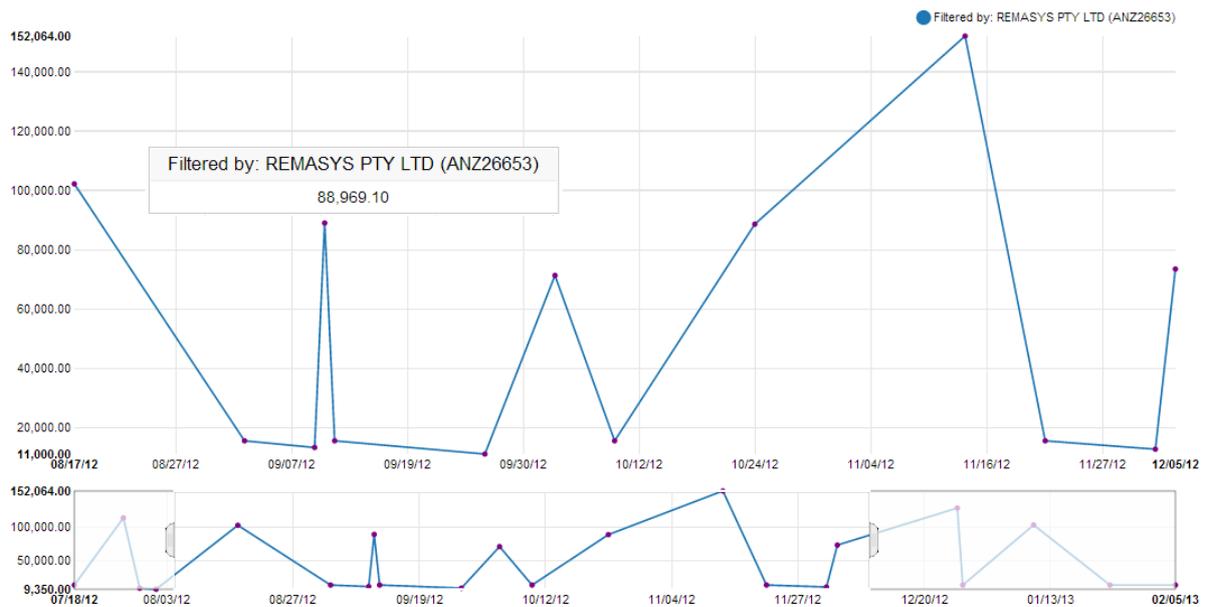


Figure 3-16 Line chart: Detailed vendor spend history

Figure 3-10 shows a line chart displaying the spend history of a selected vendor in the fiscal year 2013. It was done with the library of NVD3 (Partners). It has the ability to zoom/focus on a selection of data. More information is available on mouse over. The underneath chart is a brush component to implement focus and context zooming. Click and drag operations are supported in the small chart below to pan and zoom.

By double-clicking a specific vendor company node in the network view, an HTTP request with vendor ID is sent to the data service. The service filters the transaction list by the vendor ID and calculates the total spend. It then returns a list of dates and spend figures (see Figure 3-17). The date is converted into a UTC date format by JavaScript.

```
{
  "spendSum":102806.22,
  "date":"\VDate(1368972000000+1000)\V"
},
{
  "spendSum":16144.92,
  "date":"\VDate(1369836000000+1000)\V"
},
{
  "spendSum":18810,
  "date":"\VDate(1365080400000+1100)\V"
},
{
  "spendSum":57860,
  "date":"\VDate(1367848800000+1000)\V"
},
{
  "spendSum":37537.5,
  "date":"\VDate(1368453600000+1000)\V"
},
{
  "spendSum":33825.09,
  "date":"\VDate(1368712800000+1000)\V"
},
}
```

Figure 3-17: Date service - Response format of spend history of the selected vendor

3.6.1.5 Collaborative panel

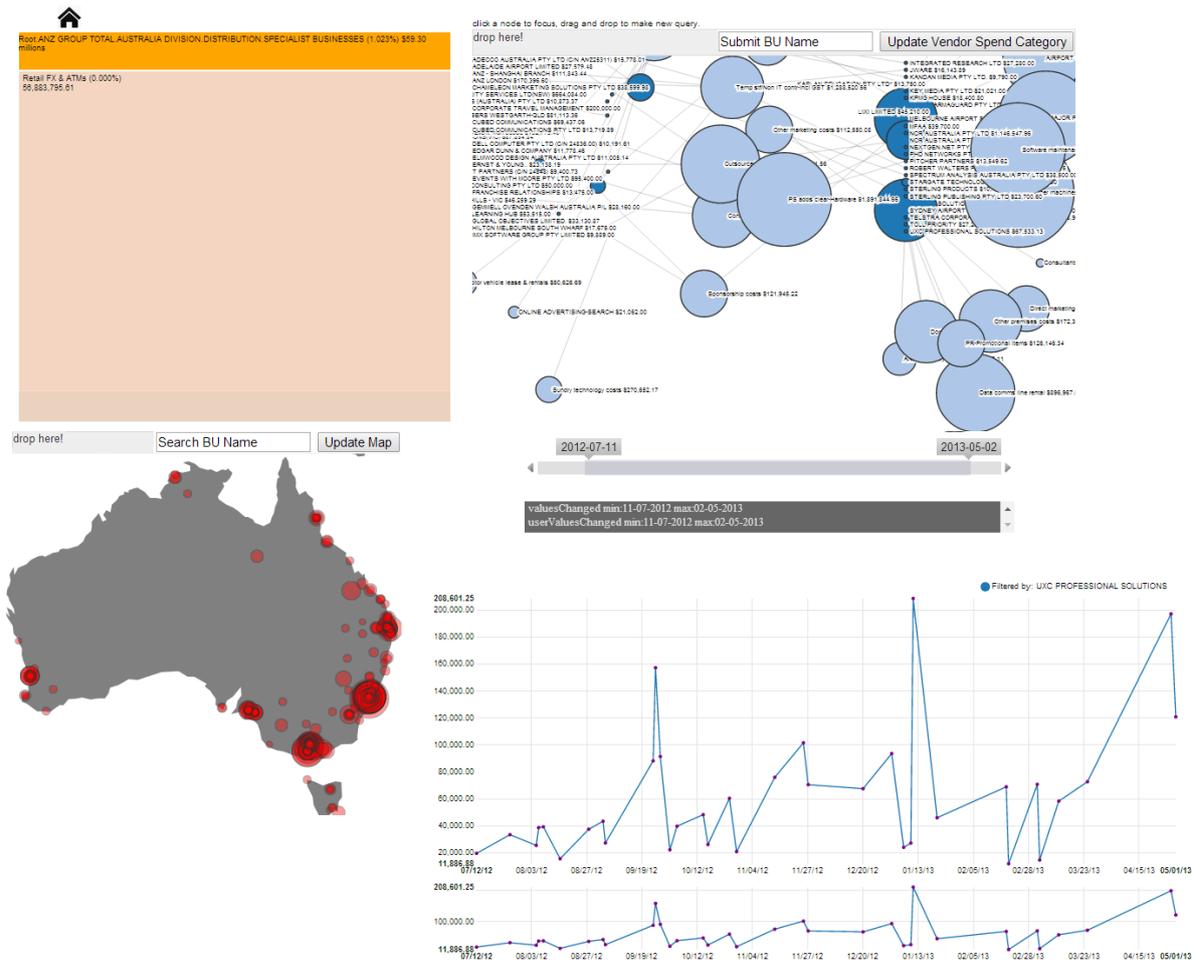


Figure 3-18: Overall interactive panel

Our research was experimented on a 46 inch display (3M™ Multi-Touch Display C4667PW). The size of the display was well suited for two to four users. To support two to four users to work together on the tabletop, collaborative interactions need to be considered in the interface design. In the collaborative view (see Figure 3-11), each of the data visualisations can communicate with each other. Section 3.7.1 describes how users can make queries from one SVG to another by drag and drop.

3.6.1.6 ANZ CFO spend report panel

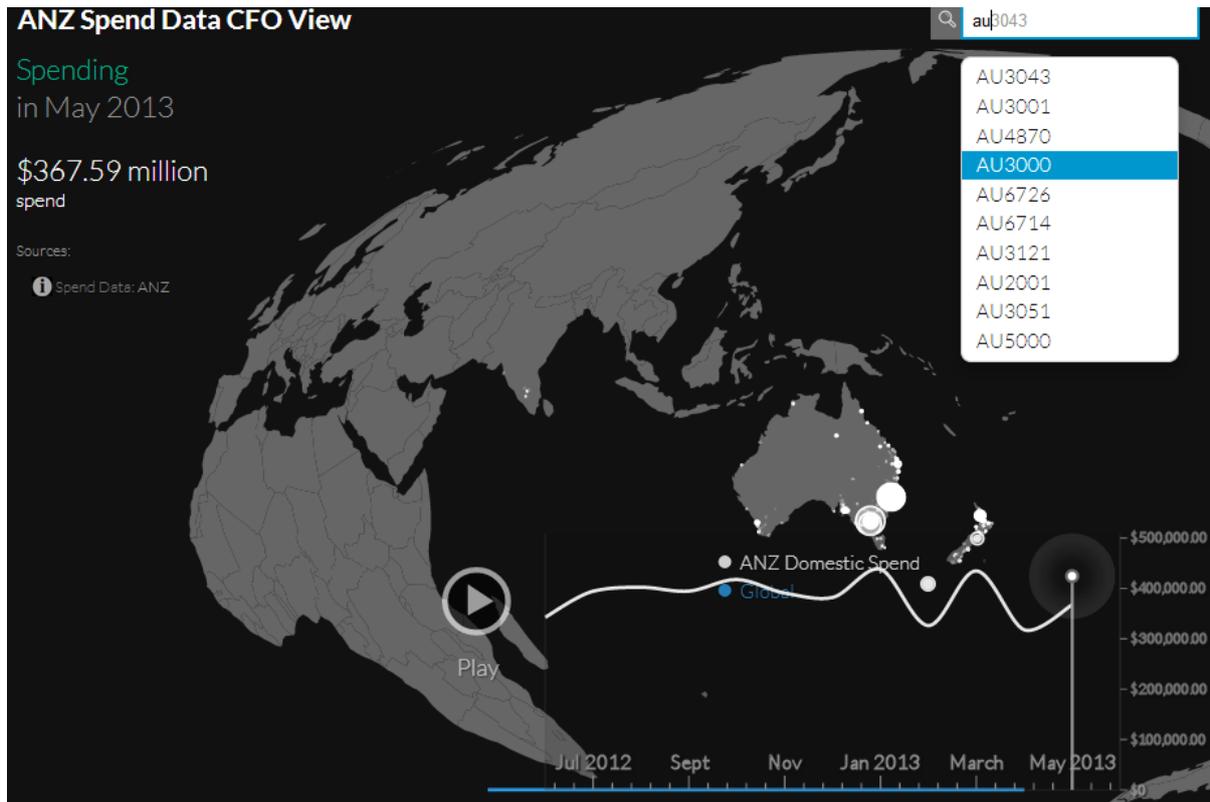


Figure 3-19: CFO Spend report panel

Ever since Hans Rosling demonstrated his motion chart (Rosling, 2009) to visualise the wealth and health data of nations, there has been an affinity for proportional bubbles on an x-y axis. However, bubble charts can be helpful if combined with geographic data and temporal animation display. Inspired by Sander, Abel, Bauer, & Schmidt, (2014) and <http://www.visualizing.org/full-screen/54850>, 2013, we also developed CFO spend report panel (in Figure 3-19) which is a high level summary view of global spend data. The user clicks Play to see changes over the temporal flow of cities on the geographic heat map. Users can also view the spend data of a specific city by entering the country code and post code of the city. The radius of the circle reflects its total spend figure in the area. In this visualisation, users can learn about the spend pattern from the spend data with a geographical and temporal view. The process of extracting data is similar to the heat map in section 3.6.1.2 and Figure 3-10. The data for CFO visualisation is stored as JSON format (see Figure 3-20).

2012.07	2012.08	2012.09	2012.1	2012.11	2012.12	2013.01	2013.02	2013.03	2013.04	2013.05	2013.06	lat	lon	name	postcode
504531	539487.5	524921.6	508753.3	575501.7	448305.5	484860.3	517783.2	475428.4	434636.7	519335.2	511390.1	-37.6866	144.8827	TULLAMAI	AU3043
14867233	16414269	18281226	15043103	18406626	18132420	28681850	26706277	16707173	17099455	15033964	16866300	-38.365	144.7659	MELBOUR	AU3001
196952.8	249573.7	327386	312066.7	262648.6	261613.4	242518.3	236743.7	196436.6	187776.4	272985.8	253380.9	-16.9043	145.7245	WHITFIELD	AU4870
25516762	25029107	28726656	27960523	24699300	29897855	18276191	26876974	24905526	23637054	22347136	28609814	-37.8146	144.9703	MELBOUR	AU3000
9113.5	10323.5	2948	28358	24002	13409	15664	20790	16104	13212	16346	14960	-17.9502	122.1964	CABLE BE/	AU6726
46092	51967	47038	45425	39583	49483	37987.5	46257.5	35553	36499.5	38638	40408.5	-20.7574	116.8495	STOVE HIL	AU6714
1430118	779576.9	1131344	1260886	2885241	397910.7	1337290	2525817	528661.8	1175554	494238.1	666062.7	-37.815	144.9914	RICHMON	AU3121
29753153	25259190	24142669	33581440	31306481	31574920	26798652	54999031	21864206	32782159	31531094	21020635	-33.7949	151.2681	SYDNEY	AU2001
21690.6	199926	146358.9	164554.1	148239.2	123453.8	115263.4	142751.3	23345.45	189204.3	184428.9	129785.3	-37.906	145.0563	NORTH MI	AU3051
273624.7	586457.6	459845.1	2477445	1171597	605657.1	341638.3	416974.8	1712144	494766	411504.5	1158997	-34.9333	138.5935	STURT STR	AU5000
1620815	590387.2	1038274	1514214	1137975	1214801	856995.6	1271961	1242367	1621672	1326309	1387095	-27.631	153.31	WATERFR	AU4001
1151141	1535430	2036018	3491016	4286207	1959087	1387223	1782447	1147828	2949802	1466641	1735044	-37.8373	144.9763	MELBOUR	AU3004
33574.79	239007.4	324067	136753.3	177400	835508.8	119533.2	261875.9	314915.6	104576.4	152008.2	638810.7	-37.8009	144.9792	FITZROY	AU3065

Figure 3-20: CFO Spend report panel: CSV data format

3.6.2 Operations

In SpendWise, the visualisations are alive to the underlying data. In Table 3-7, listed operations are supported to help users explore data in a touch-enabled collaborative environment. Each visualisation has been embedded in a DIV tag with the touch event listener. Hammer.js, a JavaScript library for multi-touch gestures, has been implemented in our prototype to support tap, double tap, swipe, drag, pinch, and rotate gestures. Each gesture triggers related events as explained in the table below.

Operation	Action
Hover	Hovering displays more information/focuses on particular nodes e.g. (Highlights related nodes and links them together in network visualisation)
Click/Tap	Clicking explores data by changing the level of the graph e.g. (Links together related nodes and filters out unrelated nodes in network visualisation)
Double Click/Tap	Double clicking /tapping makes queries and updates other visualisation
Drag/Drop	Dragging from source SVG, and drop it on target SVG to make queries
Zoom	Enlarging/Zooming an SVG graph
Pinch	Using pinch to rotate the SVG

Table 3-7 Operation supported in SpendWise

The related work in section 2.3.3 inspired the communication feature in SpendWise which enables the SVGs to talk to each other on the same panel. The user can make queries by simply dragging from one SVG, and dropping it on another. Specifically, on a dragging-start event, the parameters are stored in a global variable, and contents are displayed until the dropping event (see Figure 3-21). The user drags a parameter from a source view and drops it on other views. On dropping event, update functions are invoked to form a new query and update the target view.

Hammer.js supports desktop browsers. It also works on Android, BlackBerry, iOS and Windows touch devices.



Figure 3-21: Drag and Drop to make a new query

3.7 COLLABORATIVE SCENARIOS

3.7.1 Co-location collaboration

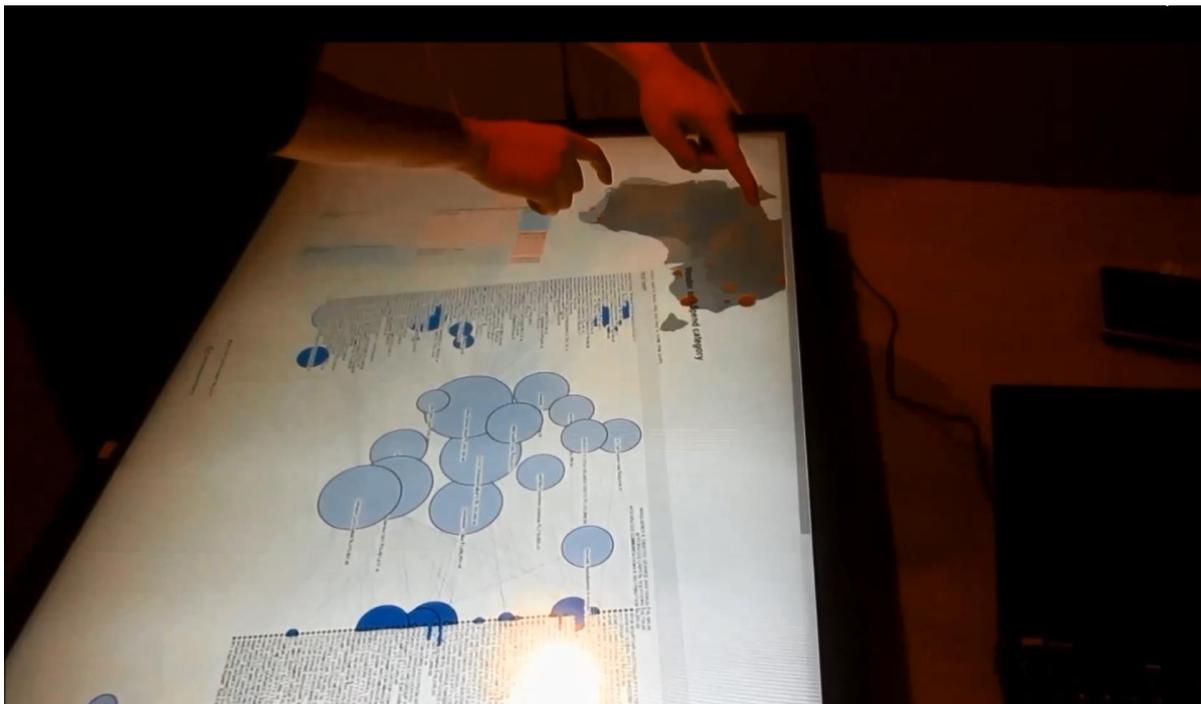


Figure 3-22: Collaborative interaction on a tabletop (multiple users and multi-touch experience)

As shown in Figure 3-22, the collaborative view is displayed on a large-size touch-enabled display. It enables two to four users to work together on the same view and having discussions on the result (see Table 3-4). Each SVG is contained by an html <div> tag, and data interactions are supported through the gestures made directly on the data visualisation. The <div> frame provides resizing,

rotating and translation operations. Multi-touch gestures are supported throughout the entire frame. Through interactions with the SVGs, data interactions are scoped and linked.

3.7.2 Collaboration from multi-devices

In another group collaboration scenario (see Table 3-5), a group of data analysts lock themselves in a meeting room. They often work separately on their laptops, periodically posting their results on the tabletop. Thus, we implemented a public/shared workspace feature to enable the collaboration of different people working on different devices. By using Restful Interfaces to build a collaborative interface, users can load their private workspace on any type of device with a modern browser. As Figure 3-23 shows, the user can send their results to the public workspace by selecting a view and sending it to the server. The server broadcasts a message and updates the public view at the same time. A description of the real-time communication feature can be found in section 3.5, Figure 3-3, and Figure 3-4.

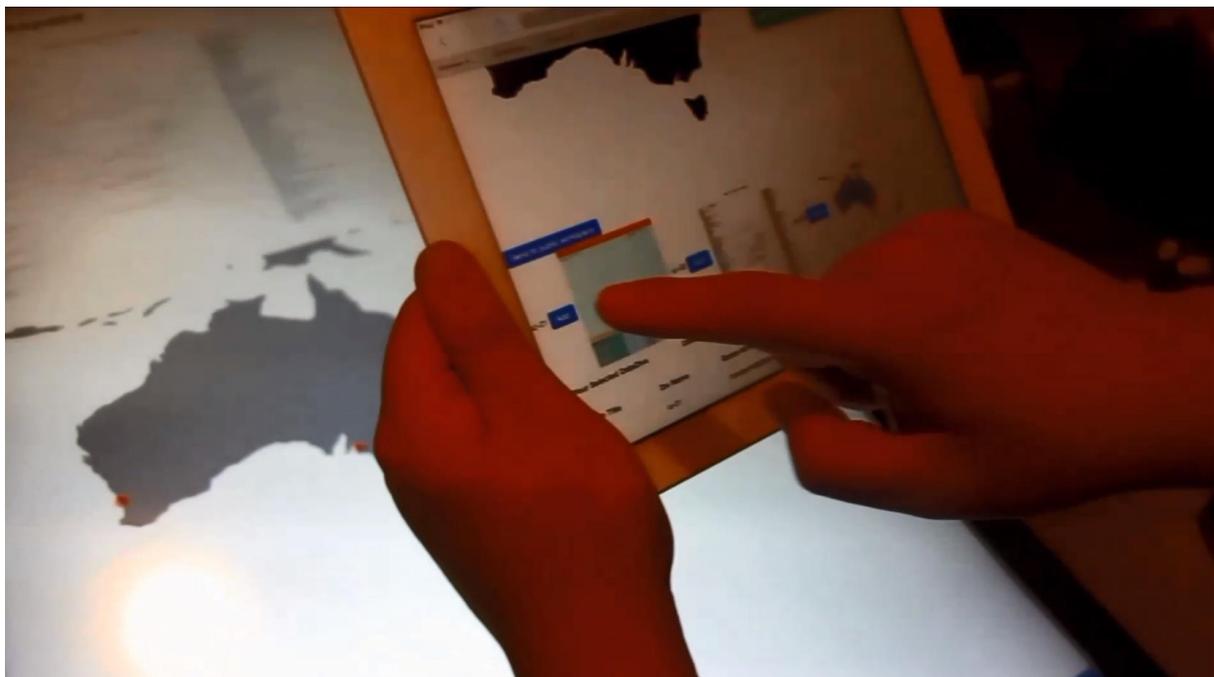


Figure 3-23: Collaboration from multiple devices (private work space)

Section 4 discusses the final phase of the system development in which a heuristic evaluation was conducted.

4 HEURISTIC EVALUATION

4.1 INTRODUCTION

The prototype developed for ANZ spend data was evaluated in interviews with domain experts. The interview design was based around heuristic evaluation methodology. It is widely used in prototype development as conducting such evaluations does not require predefined measures of performance or a rigid protocol. It was chosen in our research because it is a cost-effective, quick, intuitive, and the method is simple to administer. In a survey of user-centred design amongst usability practitioners, it was rated one of the most popular methods (Vredenburg, Mao, Smith, & Carey, 2002).

The purpose of heuristic evaluation in our research is to find usability problems in the prototype and gain feedback from domain data analysts. Evaluation has been a useful development tool at all stages of the project. Evaluators gave feedback on early design ideas and prototypes, at this formative stage results were used for improvements in the interactive process and to guide the design. Later evaluations of more developed prototypes were used as appraisals.

Qualitative evaluation was the most practical approach because the prototype is designed for analysing data from a specific domain, and because of the nature of the collaborative environment. It was useful for gathering overall impressions and feedback on the prototypes and collaborative design. This evaluation also provided feedback, enabling improvements to be made to the system's usability. Transcriptions of typical interviews are given in Appendix 9.2. The remaining part of this section describes this evaluation method further, and details how the evaluation was conducted in our research.

4.2 DESIGN OF EVALUATION

To evaluate our interface and compare it to traditional table-based interfaces, we conducted an initial heuristic evaluation based on inspection techniques developed by Nielsen (1994).

We prepared some tasks that were designed to determine whether the prototype met our objectives such as helping users understand the related operations in the context of data exploration. Additionally, the heuristics offer a basis for discussion with evaluators. However, our aim was to explore qualitative issues such as strengths, weaknesses, and usefulness of each visualisation and ideas for the future development. The evaluation results determine the issues that should be taken into account in future research.

Heuristic evaluation is informal and subjective. A group of domain experts from ANZ enterprise Architect team were invited to join the evaluation. The evaluation was undertaken in several rounds; e.g. in the first round, we played the introduction video for the evaluators and the second round, we invited them to explore the interface. Additional rounds were conducted in which each visualisation was inspected and judged using heuristics. The evaluators identify general usability problems that users can be expected to encounter by themselves. Sometimes the evaluators also provide suggestions to overcome the problems and improve overall usability. The results of the evaluations were recorded.

4.3 METHOD

Inspired by Tory & Moller (2005), we created a set of task lists (as stated in a section 4.7). After using the prototype to solve a series of tasks, the evaluators offered feedback through discussion. We designed a task list to guide the evaluators as they went through the necessary tasks to assess their experience with the visualisation prototype. A questionnaire was also used as an interview guide when discussing the outcome of the evaluation with each evaluator.

We conducted five discussion meetings with the ANZ team prior to the final evaluation. At the first meeting, a set of sketches was offered on paper (with its functionality listed on the side) depicting the general concepts of the prototype. In subsequent meetings, the prototype was iteratively developed to a fully functioning system.

During the evaluation, it is important that the evaluation manager remains neutral. When interacting with the evaluators, the communication was conducted in an unbiased manner. This motivated the evaluators to give honest responses. In addition, we assured all evaluators that they should not feel any pressure to respond positively because the evaluation manager is the developer of the system. On the contrary, they were encouraged to point out problems and give constructive feedback since obtaining such feedback was the purpose of the evaluation and the feedback would eventually improve the prototype.

4.4 APPLYING THE EVALUATION

To evaluate and improve the implementation, SpendWise was set up on a tabletop for three domain experts, consisting of both professional data analysts and database expert from the ANZ Enterprise Architect. These domain experts were interviewed and were encouraged to explore and analyse the data using SpendWise. Their feedback was gathered in the process. The example task list given below is typical of the data explorations that were performed by the evaluators.

During the interview, the evaluators spoke about the result, and their conversations were recorded. This reduced the load for the evaluator as he/she is doing the inspections and the evaluation manager (the author) can review the findings at a later date. The evaluation was arranged in the following interview process:

- At the beginning of each evaluation interview, an introduction video and a short demonstration are given to the evaluators to obtain a general feel of SpendWise.
- Conversations are recorded for later analysis and archiving purposes.
- The evaluation manager guides evaluators through different user scenarios
- Each evaluator is asked to complete a set of pre-defined tasks.
- Evaluators are asked to do the inspection individually to assure unbiased results.
- At the end of the interview, all evaluators form a discussion group to work together in a collaborative environment. The results are compiled into one report.
- Therefore, duplicated problems can be removed and a count given for the number of occasions that the particular feedback was given.

It is important to note that traditionally evaluators are not allowed to discuss with each other until they are finished with the evaluation in order not to influence each other. However, collaboration can be useful if experts in usability and experts in the domain can work together. In this case, such groups may discover more problems than they would normally have had if they were working individually. The collaborative approach can also overcome the risk of evaluators getting bored and feeling unmotivated, which may be one cause for false alarms and missed problems (Helen Petrie, 2010). This approach generates less duplicate feedback and evaluators can have a more straightforward way to come to an agreement on how to classify their results. This again saves time in aggregating the final list of problems.

4.5 EVALUATORS

Three data analysts and users were invited to join the interview. They were the experts in the actual data domain visualised by the system and are real users in this specific context. One of them was an expert both in usability and in the data domain.

4.6 TASKS AND SCENARIOS

Traditionally in a heuristic evaluation, the evaluators decide themselves how to proceed during the inspection. They are only required to inspect it freely and are not supposed to use the system to perform real tasks. However, another approach has been widely used which provides evaluators

with typical user scenarios and tasks to direct the inspection. It is aimed to collect improved accuracy of feedback on the overall usability of the visualisations. Furthermore, the evaluators are domain experts without much of an IT background. Therefore, it is necessary to assist them to use it, such as listing the steps a real user would do when performing typical tasks, and archiving the certain tasks which SpendWise has been designed for.

Datasets and the prototype were set up ahead of the evaluation. The evaluators were first introduced to the interfaces of the prototype and the concept of information visualisation. Then the evaluators were asked to perform two sample tasks:

- 1) Explore the dataset
- 2) Analyse a particular set of spend data for a vendor or business unit.

During the evaluation, the evaluation manager used contextual inquiry techniques to encourage participants to discuss problems, and share their ideas and opinions.

Figure 4-1 is the start point of the task. The evaluators were asked to finish the tasks below.

4.6.1 Task set 1 on collaborative panel

- Understanding when and where the money has been spent and in what category.
- E.g. “ANZ Group Total” - “Australia Operations” - “Collection”
- Make a query on the map to understand the geographical spread of spend data.
- Check what are the most significant categories of total spend (in terms of \$ value). Also find out how much money was spent on “postage costs”. Then to figure out what is the amount of money spent on the Australia Post and other vendors.
- Filter the data from 20-10-2012 to 20-03-2013 and check the updated visualisation again.
- Check the detail spend on Australia Post VIC using the line chart.

4.6.2 Task set 2 on CFO View

- Play overall spend map
- Check spend by postcode au3000
- Check spend by postcode nz1140
- Move cursor over the city and select city to understand the spends of the city.

4.6.3 Task set 3 on collaborative panel

- Use the touch screen to conduct queries in task 1.

4.6.4 Task set 4 on private/public workspace

- Send a task from a private workspace (e.g. iPad/another computer) to public work space.

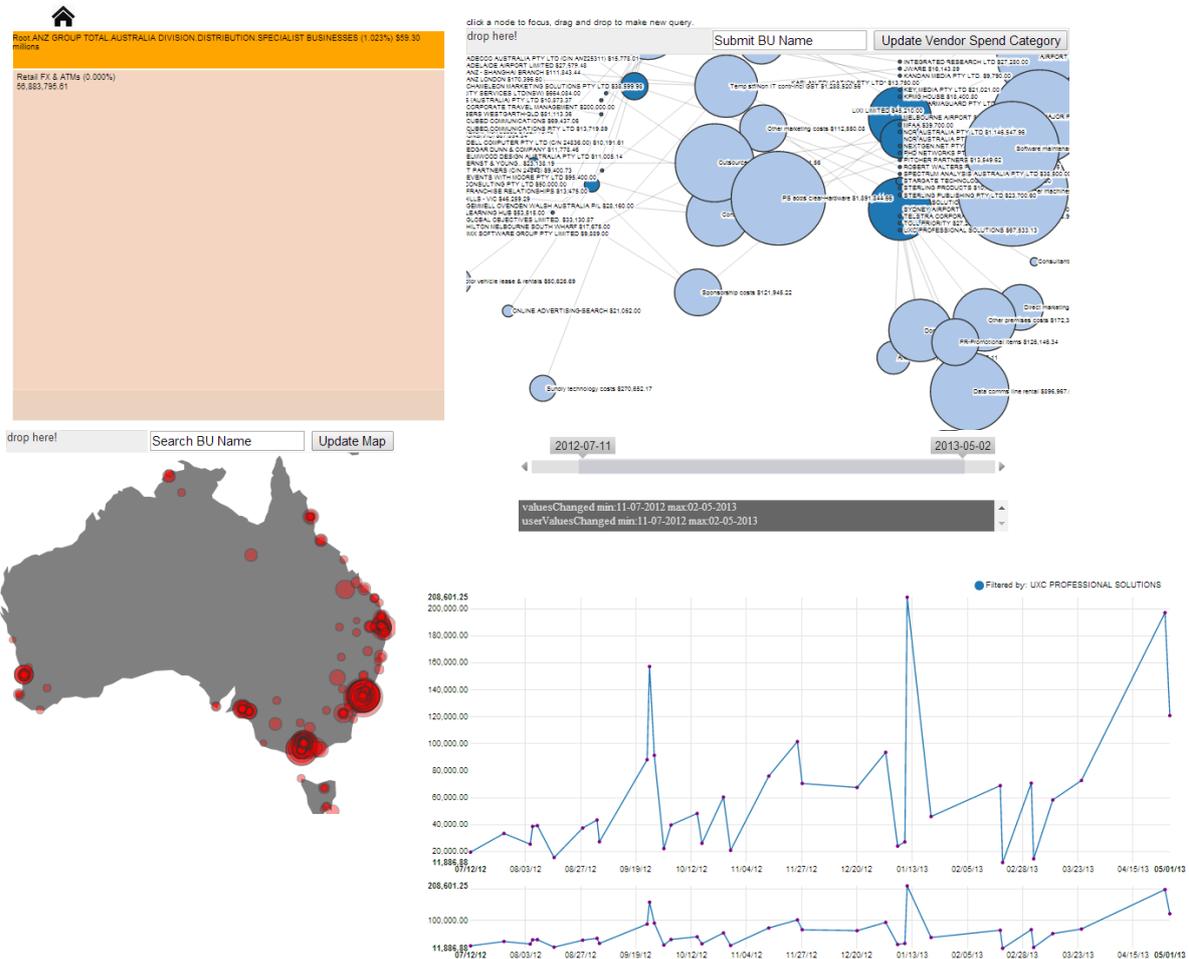


Figure 4-1: Initial interface of SpendWise evaluation

4.7 USE-CASE

Figure 4-1 is a screenshot of the initial user interface after loading the default view. Twelve months of accumulated data from June 2012 to May 2013 is shown. This is the starting point of use-case, as shown to the evaluators. By default the system loads the data with the first level of the organisational hierarchy. Then the user is able to familiarise themselves with those pre-defined views. Some additional features are shown in Figure 4-1 that were not described previously. These include an input box used to identify business units for map updates, and an input box for updating the network visualisation by business unit. Another feature is the custom zoom operation: using multi-touch functions to zoom, rotate or drag and drop each view. Evaluators identified several patterns immediately based on the visualisation as shown in bold text below.

Some initial observations noted by the evaluators include:

- Hierarchical spend overview (tree map) - an intuitive view to understand the level of the organisational hierarchy and its spend figure. Users can quickly check the spend total of each business unit at different levels by clicking the different areas of the view (see Figure 3-7). The evaluator gave feedback about adding revenue data as a comparison view to help put the figures in context. Closer views on the proportion of total spend enables them to quickly compare some specific business units and start a discussion. **In the demonstration, a spend pattern was immediately identified and confirmed a flaw in a credit card payroll system which leads to about 30% of transactions becoming untraceable (unclassified data).**
- Geographical spend heat map view - a geographic view that displays the spend data of each area on the map to provide a picture of where the money has been spent (see Figure 3-17).
- Network visualisation on a selected business unit - a network view of the spend schema of the business unit and its related vendors. It also carries information regarding total spend amount and spend details (see Figure 3-12). The network visualisation was the most popular feature among the evaluators. **All evaluators immediately searched their own business units and reviewed their spend categories and their related vendors. They further explored several of their counterpart business units and began discussion on the usefulness of network visualisation to help them in understanding the spend data. The attention of the evaluators was drawn to the relationship between the categories and the spend amount of a particular vendor in a given time period.** Then they explored the time range slider gadget (see Figure 3-14) to help them narrow down the analysis into a specific timeframe. After trying to compare the expenditure of several business units in different timeframes, they started mapping the data with the transactions/events from that period. Next, the line chart (see Figure 3-16) was noted. They double clicked a node in the network view to update the line chart. Then they checked the detailed spend history of a selected vendor in the line chart. They also clicked and dragged the small chart to enable a more detailed view using zoom and pan. In another scenario, **two evaluators specifically went back to the network visualisation to check the “unclassified” data and reviewed each of the related vendor companies. They were trying to interpret the visualisation with their knowledge of past events.** They also explained why there were peaks in March after figures dropped in December and January (people on holidays, and people came back and spend went up). These variations are due to common vacation periods. Many discussions and interpretations were made based on their analysis of the network view.
- CFO spend summary view - evaluators checked the high level summary view of overall global spend (see Figure 3-19) and searched the postcode of Sydney and Melbourne to see the

animation report. In this summary view, **they immediately spotted a few spend patterns. For instance, in December the spend goes down because many employees are on holiday, whereas they noticed a sudden spend rise at the end of the fiscal year.**

- Collaborative scenarios (co-location) - the group evaluation involved four people (including the evaluator manager), the evaluators were excited about the multi-touch function and had many discussions on how collaborations would work using such functions within the ANZ protocol (see section 4.8).
- Collaborative scenarios (multi-device) - the evaluators were asked to send analysis results (views) from an iPad (private workspace) to the tabletop (public workspace). More description can be found in section 4.8.

4.7.1 The usage flow

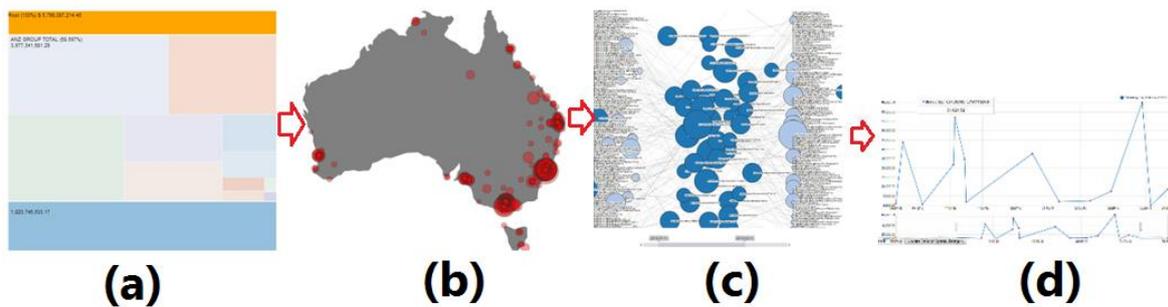


Figure 4-2: Usage flow: Regular user work flow

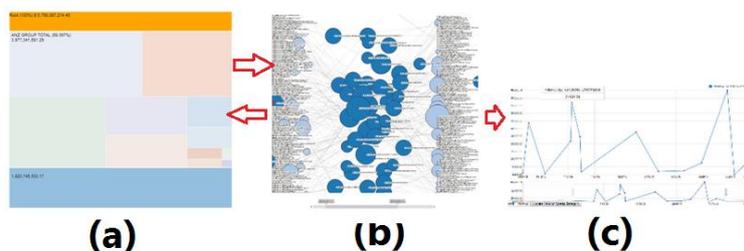


Figure 4-3: Usage flow: The most frequent work flow

All evaluators were asked to go through the regular work flow as shown in Figure 4-1. They checked the spend tree map in Figure 4-1 (a) and the geographic heat map in Figure 4-1 (b) for a while, then their attention quickly moved to the hierarchical tree map and network visualisation. Based on our observations on the frequency of their usage, they often check the hierarchical tree map in Figure 4-3 (a) and then update the network view Figure 4-3 (b) to identify patterns and relationships between vendors and spend categories. They spent a large portion of time going back and forth making comparisons between the different views and began to discuss the findings. They sometimes

narrowed down the time range and checked the detailed spend (see Figure 4-3 (c)) of a vendor company to verify their findings.

4.8 ANALYSIS OF RESULTS

By finishing the above tasks, evaluators began to learn about the data, and the potential of these visualisation tools for analysis. Detailed transcripts are available in Appendix 9.2. Overall, we received very positive feedback and the ANZ evaluators seemed enthusiastic about the potential applications of this work. In this section we mainly analyse what we can improve in future work. After analysing the audio record and transcripts using a coding method (Law & Hvannberg, 2004), some feedback noted by the evaluators include:

4.8.1 Visualisation improvements

Putting the data in context and offering comparison analysis (mentioned on five different occasions) - users want to see spend data and revenue data which assists analysis in the same view by offering a comparison. Having large spend numbers is not the whole story, you need to see the revenue and be able to put these numbers into context.

Improving the mapping experience by adding zoom capability (mentioned on three different occasions) - The circles on the map are difficult to see because so many circles overlap each other. A suggestion was made to use Google map API to allow zoom and provide bigger views behind the circles to improve the overall layout.

Enhancing the colour scheme (mentioned on two different occasions) - adjusting the colours of the tree map and adding a strong border were suggested as ways to better distinguish view level. One user also suggested adding a sidebar showing the current position in the hierarchy.

Improving the network layout (mentioned on three different occasions) - one evaluator said the network is too busy. Another commented he did not find the network view to be very intuitive, because at first glance, he thought the left nodes have some relationship with the right nodes with the linkages representing the relationship.

Minimizing the time of animation (mentioned on one occasion) - one evaluator mentioned the animation transition is too long, suggesting it last no more than 3 seconds. (The current setting is 5 seconds)

Changing or enlarging the icon for breadcrumbs (mentioned on one occasion) - Current breadcrumbs are shown, however, they are too small to pick up at a glance.

4.8.2 Functionality improvements

Annotation (mentioned on seven different occasions) - evaluators noted that a feature will be required to allow users to add comments on the data. In bank work flows, documentation of analysis and decision making is essential. The software must incorporate the capacity to annotate decisions as they are made so they can be justified as they pass up the management chain.

Preserving a particular view (mentioned on six different occasions) - All the evaluators strongly suggested adding the capacity for users to store states of analysis, allowing them to return to that view.

Pattern identification (mentioned on three different occasions) - evaluators suggested extending the system to help them to identify spend patterns based on the datasets over time. (e.g., in other years is this comparable?)

Sharing the views (mentioned on two different occasions) - evaluators hope to share the result of analysis through email

Exporting data into a spreadsheet (mentioned on two different occasions) - Accountants prefer looking at numeric results in the familiar format of spreadsheets, they still want to export the results into Excel/CSV files.

Advanced search (mentioned on one occasion) - One evaluator suggested having a search for different types, i.e. the capacity to select categories for a search and have that view returned.

4.8.3 Finding on bugs

The inconsistency of the query history (mentioned on one occasion) in the geographic heat map view the history of queries is displayed, but in the spend category network view it is not.

Display the percentage error of the spend hierarchy tree view (mentioned on one occasion) - it is supposed to be larger than 0.000% when the view goes deeper than 6th level.

Categories	Observations	No. of mentions
Visualisation	Putting the data in context and offering comparison analysis	5
Improvements	Improving the mapping experience by adding zooming	3
	Enhancing the colour scheme	2
	Improving the network layout	3
	Minimizing the time of the animation	1
	Changing or enlarging the icon for bread crumbs	1
Functionality Improvements	Annotation	7
	Preserving a particular view	6
	Pattern identification	3
	Sharing the views	2
	Exporting data into spreadsheet	2
	Advanced search	1
Finding on bugs	The inconsistency of the query history	1
	The percentage error of the spend hierarchy tree view	1

Table 4-1 Observations of heuristic evaluation

The evaluators paid much attention on analysis tool itself and did not give much comments on the collaborative design. More evaluation can be done on collaborative features in future work.

Six meetings were held during the project development. Suggestions from the initial five meetings have been incorporated into the design of SpendWise, with some large scale ideas and feedback from the sixth meeting noted for future investigation and implementation. Suggestions to be implemented in the future include putting the data in context and offering comparison analysis, improving the mapping experience, enabling data archiving by adding annotations and preserving a particular view.

5 CONTRIBUTION

The contribution of our research project is as follows:

- We have proposed and developed a prototype implementation of an interface that sits on top of the HTML5, CSS3, JavaScript, WCF, D3 Technologies, and SignalR which offers a rich experience of collaborative interaction and group collaboration activities (see section 3.7). Easy deployment and portability of the tools were tested and some new techniques in collaborative interactions have been explored. The proposed framework offers a novel interface based on the novel representation for querying and visualising data in property graphs (see section 3.6).
- We have investigated the applicability of network visualisation to the analysis of corporate spend data involving geographical, and temporal properties, within heterogeneous data types (see section 4.6). An extensible prototype, SpendWise, tailored for ANZ spend data analysis has been developed. The prototype supports general querying and business analytics for spend data.
- We have tested the applicability of the network analysis techniques and tools such as centrality metrics (Chen, 2004). The techniques are useful in helping to identify important structural aspects of the network, which can then be highlighted in the visualisation work (see Figure 3-13). The network visualisation presents a view of the connection between business units, vendors, and places (see Figure 3-12). Exploration from one node to another node to find out the relationship between those entities is made easier by this work. For example, by clicking a node on the network visualisation the relationships between a spend category and its vendors is displayed. The network visualisation enables users to quickly grasp the overall spend paradigm of a business unit and understand where the money has been spent and which vendors take the money (see Figure 3-18). Furthermore, SpendWise is able to quickly narrow down to a line chart visualisation which displays the breakdown of expenses of a specific vendor (see Figure 3-16). By quickly traversing from different nodes, analysts can gain insight and new understanding of the spend data.
- We have developed several visualisations of a number of predefined analyses which analysts typically want to understand. These include:
 - ANZ hierarchy spend overview (tree map) - a view of spend summary by the level of ANZ organisation hierarchy. Users can quickly check the spend total of each business unit at different levels (see Figure 3-6).

- Geographic spend heat map view - a geographic view that displays the spend data of each area on the map (see Figure 3-9).
- Network visualisation for a selected business unit - a network view of the spend schema of the business unit and its related vendors. This also carries information of the total spend amount and transaction details (see Figure 3-12).
- A time range slider gadget - a slider gadget that enables users to filter the data into a specific time period (see Figure 3-14, Figure 3-6).
- Vendor spend detail visualisation - a line chart that displays the detailed spend history in the previous fiscal year (see Figure 3-16).
- A CFO spend summary view - a high level summary view of overall global spend (see Figure 3-19).

By using the predefined view of fragmentation of spend data, non-expert users are able to quickly learn how to operate the application.

- We have defined and developed the collaborative interaction experience to support team collaboration scenarios
 - Co-location collaboration on a tabletop device - SpendWise supports two to four users working together on a large-size multi-touch screen. Users can conduct analysis by touch gestures such as drag and drop, pinch, zoom and so on (see section 3.7.1).
 - Collaboration from multi-devices - SpendWise supports users working from private workspaces and posting their results on the public workspace such as a tabletop (see section 3.7.2).
- We have also extended the widely used open-source web visualisation toolkit (<http://d3js.org>) with tables and network layout tools required to support the above interface (see section 3.6).

In summary, the primary contribution of our research is that we have proposed and developed SpendWise, a visual analytic framework, to support network visualisation and collaborative interaction.

We outline our future research directions in the next section.

6 FUTURE WORK

As stated in section 4.7 and 4.8, this research opens up several avenues for future work:

- Better geographic API can be used to provide a clearer overlay of useful data and support zooming (see section 4.8.1).
- Improvement can be made on the algorithm and design of the network visualisation. For example, our current network view becomes overcrowded when it displays more than 120 nodes (see section 4.8.1). More work can be done on rearranging the network, filter, and in exploring the connected data. (E.g. filter and sort functionality can be considered to help the user focus on useful information. Searches are often needed when displaying large datasets.)
- Archiving the results, preserving results into links and exporting the analysed data offers possibilities for collaborative analysis where multiple users analyse the same dataset. For example, a documentation system can be added allowing users to collaboratively annotate the visualisation (see section 4.8.2).
- Graph Attribute Space Explorer is an ongoing visualisation project, which has been inspired by our research. It extends the network visualisation and enables users to explore the data relationships between multiple entities such as business unit, geographic data, and total spend amount. Figure 6-1 is a screenshot of this project.

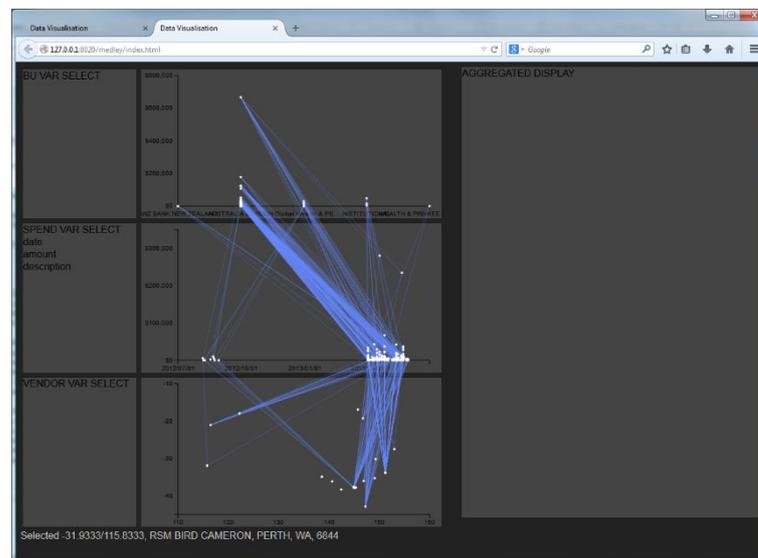


Figure 6-1: Graph Attribute Space Explorer

- Collaborative design can be made more sophisticated, according to evaluation feedback, e.g. public/private workspace is currently only implemented as a limited proof of concept (see section 4.8.2).
- Since the prototype system proposed should be applicable to any forms of tabular spend/revenue dataset, it is hoped that SpendWise can be tested in other application domains. The SpendWise prototype should be applicable to any application domain involving spend or revenue relationships. The current prototype has a hardcoded data structure specific to the ANZ Spend data format, however, future work can solve rapid construction and manipulation visualisation from any form of tabular data.
- In the feedback from ANZ, they also suggested a number of predefined analyses which analysts typically want to understand. However, due to source data been insufficient, we did not implement such analyses in our research. Nevertheless, it can be useful in future work. These include:
 - Supplier fragmentation analysis - a view based on the data of respective suppliers' percentage of the separate parts of a project.
 - Contract leakage (variance with agreed terms) - a view of the uncertain and unpredictable costs in the contract.
 - Savings goals by spend type - a summary view of the actual savings compared with the saving goals under each spend type.
 - Cross-supplier price comparison (same spend different supplier) - a view of the prices offered by different suppliers.
 - Deviation to industry commodity indices - a view of the prices from supplier versus industry commodity indices.
 - Spend pattern deviation from industry ratio classification - a view of current spend pattern VS industry ratio classification.

7 CONCLUSION

In this thesis, we have focused on a system design and prototype for performing visual analytics on a large spend dataset, and we have argued the new collaborative design for multi-user scenario offers potential for a new collaborative experience received enthusiastically by our industry partners. The contribution of this work includes the following:

- A conceptual framework offering easy operations for constructing and transforming visualisations of multivariate tabular data.
- The design and implementation of a prototype based on the framework, which integrates data manipulation with the visual exploration process.
- An architecture supporting collaborative visual analytics by multiple analysts across multiple devices.
- A user interface supporting collaborative multi-touch operations for the above framework on tabletop and personal tablet devices.
- A heuristic evaluation of SpendWise that may have implications for future work on visual analytics.

The prototype of the conceptual framework, SpendWise, is demonstrated. By providing a novel and collaborative interface, we demonstrate how ANZ data can be quickly visualised and analysed in a collaborative environment. The main visualisation panel offers novel and interactive visualisations of the entire ANZ spend dataset which allows users to explore their spend data quickly using simple operations like drag and drop. The collaborative design and public/private workspace enables users to share the individual analysis results with each other. Particularly, it allows an analyst to send live results from his/her own device to a shared tabletop collaborative environment.

A broad overview and definition of these scenarios is provided. A concrete prototype system has been developed and evaluated with domain data analysts. An evaluation report describes some enhancements made to this system based on their feedback.

In summary, SpendWise is a system for collaborative visual analytics on a multi-touch tabletop display. Collaborative panels allow users to work individually and collaboratively on visual analytic tasks. A demonstration video was made (Chang, 2014) to offer a brief introduction to SpendWise.

Our industry partner, ANZ, have been very positive about the potential for the concepts and prototype that we have demonstrated to them. We hope to continue this collaboration.

8 REFERENCES

- Aguilar, J. M. (2014). SignalR Programming in Microsoft ASP .NET.
- Ahlberg, C., & Shneiderman, B. (1994). *Visual information seeking: tight coupling of dynamic query filters with starfield displays*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Bai, Y. (2010). ASP. NET Web Services. *Practical Database Programming with Visual C#. NET*, 721-893.
- Becker, R. A., Eick, S. G., & Wilks, A. R. (1995). Visualizing network data. *Visualization and Computer Graphics, IEEE Transactions on*, 1(1), 16-28.
- Bostock, M. (2011). d3.js Data-Driven Documents.
- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12), 2301-2309.
- Bostok, M. Site Concept Browser
- Breitkreutz, B.-J., Stark, C., & Tyers, M. (2003). Osprey: a network visualization system. *Genome Biol*, 4(3), R22.
- Brennan, S. E., Mueller, K., Zelinsky, G., Ramakrishnan, I., Warren, D. S., & Kaufman, A. (2006). *Toward a multi-analyst, collaborative framework for visual analytics*. Paper presented at the Visual Analytics Science And Technology, 2006 IEEE Symposium On.
- Buja, A., McDonald, J. A., Michalak, J., & Stuetzle, W. (1991). *Interactive data visualization using focusing and linking*. Paper presented at the Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on.
- Chaffer, J. (2013). *Learning jQuery*: Packt Publishing Ltd.
- Chang, C. (2014) Demonstration video for SpendWise (Attached to minor thesis submission).
- Chen, C. (2004). *Information visualization: Beyond the horizon*: Springer.
- Chuah, M. C., & Roth, S. F. (2003). *Visualizing common ground*. Paper presented at the Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on.
- Derrick Snider, P. P., Steve Swedler, Scott Lewis (2013).
- Derthick, M., Kolojejchick, J., & Roth, S. F. (1997). *An Interactive Visualization Environment for Data Exploration*. Paper presented at the KDD.
- Dork, M., Carpendale, S., Collins, C., & Williamson, C. (2008). Visgets: Coordinated visualizations for web-based information exploration and discovery. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6), 1205-1212.
- Eyes, I. M., & Hoag, T. (2009). Many Eyes. *Many Eyes*.
- Ferraiolo, J., Jun, F., & Jackson, D. (2000). *Scalable vector graphics (SVG) 1.0 specification*: iuniverse.
- Field, A., Hartel, P., & Mooij, W. (2001). *Personal DJ, an architecture for personalised content delivery*. Paper presented at the Proceedings of the 10th international conference on World Wide Web.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3), 75-174.
- Freire, M., Plaisant, C., Shneiderman, B., & Golbeck, J. (2010). *ManyNets: an interface for multiple network analysis and visualization*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Furnas, G. W. (1986). *Generalized fisheye views* (Vol. 17): ACM.
- Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2004). *A comparison of the readability of graphs using node-link and matrix-based representations*. Paper presented at the Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on.
- Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2005). On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Information Visualization*, 4(2), 114-135.

- Gobel, B. (2013). AniMap: An Interactive Visualization Supporting Serendipitous Discovery of Information about Anime.
- Goodman, D., Morrison, M., & Eich, B. (2007). *Javascript® bible*: John Wiley & Sons, Inc.
- Gutwin, C., & Greenberg, S. (1998). *Design for individuals, design for groups: tradeoffs between power and workspace awareness*. Paper presented at the Proceedings of the 1998 ACM conference on Computer supported cooperative work.
- Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information visualization*, 7(1), 49-62.
- Heer, J., & Perer, A. (2011). *Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks*. Paper presented at the Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on.
- Helen Petrie, L. B. (2010). Collaborative Heuristic Evaluation: improving the effectiveness of heuristic evaluation. *UPA2010*.
- Henry, N. (2007). Visually exploring large social networks *Human-Computer Interaction—INTERACT 2007* (pp. 604-610): Springer.
- Henry, N., Fekete, J.-D., & McGuffin, M. J. (2007). NodeTriX: a hybrid visualization of social networks. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1302-1309.
- Herman, I., Melançon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 6(1), 24-43.
- <http://nvd3.org/>. Last accessed: 20/06/2014
- <http://www.naturearthdata.com/>. Natural Earth Data Last accessed: 20/06/2014
- <http://www.visualizing.org/full-screen/54850>. (2013). Last accessed: 20/06/2014
- Hu, Z., Snitkin, E. S., & DeLisi, C. (2008). VisANT: an integrative framework for networks in systems biology. *Briefings in bioinformatics*, 9(4), 317-325.
- Huffaker, B., Nemeth, E., & Claffy, K. (1999). *Otter: A general-purpose network visualization tool*. Paper presented at the International Networking Conference (INET).
- Ian1971. (2013). Excel Data Reader - Read Excel files in .NET.
- Ingebrigtsen, E. (2013). *SignalR: Real-time Application Development*: Packt Publishing.
- Isenberg, P., & Carpendale, S. (2007). Interactive tree comparison for co-located collaborative information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1232-1239.
- Isenberg, P., & Fisher, D. (2009). *Collaborative Brushing and Linking for Co - located Visual Analytics of Document Collections*. Paper presented at the Computer Graphics Forum.
- Isenberg, P., Tang, A., & Carpendale, S. (2008). *An exploratory study of visual information analysis*. Paper presented at the Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.
- Kang, H., Plaisant, C., Lee, B., & Bederson, B. B. (2007). NetLens: iterative exploration of content-actor network data. *Information Visualization*, 6(1), 18-31.
- Kapler, T., & Wright, W. (2005). Geotime information visualization. *Information Visualization*, 4(2), 136-146.
- Keel, P. E. (2006). *Collaborative visual analytics: Inferring from the spatial organization and collaborative use of information*. Paper presented at the Visual Analytics Science And Technology, 2006 IEEE Symposium On.
- Law, E. L.-C., & Hvannberg, E. T. (2004). *Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation*. Paper presented at the Proceedings of the third Nordic conference on Human-computer interaction.
- Lee, B., Czerwinski, M., Robertson, G., & Bederson, B. B. (2005). *Understanding research trends in conferences using PaperLens*. Paper presented at the CHI'05 extended abstracts on Human factors in computing systems.

- Liu, Z., Navathe, S. B., & Stasko, J. T. (2013). Ploceus: Modeling, visualizing, and analyzing tabular data as networks. *Information visualization*.
- Livny, M., Ramakrishnan, R., Beyer, K., Chen, G., Donjerkovic, D., Lawande, S., . . . Wenger, K. (1997). *DEVise: integrated querying and visual exploration of large datasets*. Paper presented at the ACM SIGMOD Record.
- Löwgren, J. (2007). Pliability as an experiential quality: Exploring the aesthetics of interaction design. *Artifact*, 1(2), 85-95.
- Mackey, A. (2010). Windows Communication Foundation *Introducing .NET 4.0* (pp. 159-173): Springer.
- Medlock, M. C., Wixon, D., Terrano, M., Romero, R., & Fulton, B. (2002). Using the RITE method to improve products: A definition and a case study. *Usability Professionals Association*.
- Nielsen, J. (1994). *Usability inspection methods*. Paper presented at the Conference companion on Human factors in computing systems.
- Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Partners, N. NVD3 re-usable charts for d3.js.
- Rao, R., & Card, S. K. (1994). *The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Rosling, H. (2009). Gapminder. *GapMinder Foundation* <http://www.gapminder.org>.
- Roth, S. F., Lucas, P., Senn, J. A., Gomberg, C. C., Burks, M. B., Stroffolino, P. J., . . . Dunmire, C. (1996). *Visage: a user interface environment for exploring information*. Paper presented at the Information Visualization'96, Proceedings IEEE Symposium on.
- Sander, N., Abel, G. J., Bauer, R., & Schmidt, J. (2014). Visualising migration flow data with circular plots: Vienna Institute of Demography Working Papers.
- Scott, S. D., Grant, K. D., & Mandryk, R. L. (2003). *System guidelines for co-located, collaborative work on a tabletop display*. Paper presented at the ECSCW 2003.
- Shneiderman, B. (1998). Treemaps for space-constrained visualization of hierarchies.
- Shneiderman, B., & Aris, A. (2006). Network visualization by semantic substrates. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5), 733-740.
- Silva, S. F., & Catarci, T. (2000). *Visualization of linear time-oriented data: a survey*. Paper presented at the Web Information Systems Engineering, 2000. Proceedings of the First International Conference on.
- Spence, R., & Press, A. (2000). Information visualization.
- Stasko, J., Görg, C., & Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2), 118-132.
- Stolte, C., Tang, D., & Hanrahan, P. (2002). Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1), 52-65.
- Stolte, C., Tang, D., & Hanrahan, P. (2008). Polaris: a system for query, analysis, and visualization of multidimensional databases. *Communications of the ACM*, 51(11), 75-84.
- Stonebraker, M., Chen, J., Nathan, N., Paxson, C., & Wu, J. (1993). *Tioga: Providing data management support for scientific visualization applications*. Paper presented at the VLDB.
- Tang, A., Tory, M., Po, B., Neumann, P., & Carpendale, S. (2006). *Collaborative coupling over tabletop displays*. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in computing systems.
- Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: The research and development agenda for visual analytics*: IEEE Computer Society Press.
- Thudt, A., Hinrichs, U., & Carpendale, S. (2012). *The bohemian bookshelf: supporting serendipitous book discoveries through information visualization*. Paper presented at the Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems.

- Tobiasz, M., Isenberg, P., & Carpendale, S. (2009). Lark: Coordinating co-located collaboration with information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6), 1065-1072.
- Tollis, I., Eades, P., Di Battista, G., & Tollis, L. (1998). *Graph drawing: algorithms for the visualization of graphs* (Vol. 1): Prentice Hall New York.
- Tory, M., & Moller, T. (2005). Evaluating visualizations: do expert reviews work? *Computer Graphics and Applications, IEEE*, 25(5), 8-11.
- Tufte, E. R. (1985). The visual display of quantitative information. *Applied Optics*, 24, 145.
- van Ham, F., Schulz, H.-J., & Dimicco, J. M. (2009). Honeycomb: Visual analysis of large scale social networks *Human-Computer Interaction—INTERACT 2009* (pp. 429-442): Springer.
- van Velsen, L., van der Geest, T., & Klaassen, R. (2007). *Testing the usability of a personalized system: comparing the use of interviews, questionnaires and thinking-aloud*. Paper presented at the Professional Communication Conference, 2007. IPCC 2007. IEEE International.
- Van Velsen, L., Van Der Geest, T., Klaassen, R., & Steehouder, M. F. (2008). User-centered evaluation of adaptive and adaptable systems: a literature review. *Knowledge Engineering Review*, 23(3), 261.
- Viégas, F. B., & Donath, J. (2004). *Social network visualization: Can we go beyond the graph*. Paper presented at the Workshop on Social Networks, CSCW.
- Vredenburg, K., Mao, J.-Y., Smith, P. W., & Carey, T. (2002). *A survey of user-centered design practice*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Wang, X. (2006). A practical way to teach web programming in computer science. *Journal of Computing Sciences in Colleges*, 22(1), 211-220.
- Ware, C. (2000). *Information visualization* (Vol. 2): Morgan Kaufmann San Francisco.
- Wattenberg, M. (2006). *Visual exploration of multivariate graphs*. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in computing systems.
- Weibelzahl, S., Jedlitschka, A., & Ayari, B. (2006). *Eliciting requirements for an adaptive decision support system through structured user interviews*. Paper presented at the AH'06 Workshop Proceedings.
- Wesche, G., Wind, J., Gobel, M., Rosenblum, L., Durbin, J., Doyle, R., . . . Fischer, M. (1997). Applications of the responsive workbench. *Computer Graphics and Applications, IEEE*, 17(4), 10-15.

9 APPENDIX

9.1 GLOSSARY

Term	Definition
AJAX	Asynchronous JavaScript and XML
API	Application programming interface
Brushing and linking	“Brushing” means that different part of the visualisation should be chained with each other, so that select on part can reflect to other parts. “Linking” represents that changes of a variable are reflected in all places it shown
Co-located	Multiple users work in a shared location/workspace
D3	Data driven document
DOM	Document object model
HTML5	The fifth generation of Hypertext Mark-up Language
Loosely coupled	opposite with tightly couple, where individual can work independently before having to interact with others
JavaScript	an interpreted popular web programming language
JQuery	A cross-browser JavaScript library designed to simplify scripting of HTML
Information Visualisation	The study of the interactive visual representations of abstract data to reinforce human cognition.
Network Visualisation	An effective means to understand the patterns of interaction between entities, to discover entities with interesting roles, and to identify inherent groups or clusters of entities (Liu et al., 2013)
RITE	Rapid Iterative Testing and Evaluation
SVG	Scalable vector graphics
SignalR	ASP.NET SignalR is a new library for ASP.NET developers that makes it incredibly simple to add real-time web functionality to your applications.
Scalability	In this literature it represents using filtering and aggregation strategy to handle massive network data
Tabletop	a table-sized device with multi-touch screen
WCF	Windows Communication Foundation
UCD	User-centred design
UCE	User-centred evaluation

Table 9-1 Terminology

9.2 TRANSCRIPTS OF HEURISTIC EVALUATION

10:00 am-11:50 am Friday, 23 May 2014

Attendees:

Sunny Avdihodzic - ANZ

Clifford, Dale - ANZ

Chris Garden - ANZ

Chunlei - Monash

Tim-Monash

James-Monash

Speaker & Transcription	Notes
<p>A: Enterprise Architect for cooperate centre segment so I am responsible for HR, Finance, GSS & Risk and one of the major components of the cooperate centre is one of horizontal sections for information. <u>So we are looking at major investments into information management capabilities. On the information management side, as well as BI and visual analytics.</u></p> <p>A: At the moment we have not very sophisticated tools for some of the reporting. We are moving towards modern tools. <u>The intent was how to look at some novel visualisations and link it together to the collaboration capabilities to provide rich user experience and gain some insight into the data. It's really important</u>The strategy and planning division is set by Patrick Maes which is chief technology officer and general manager for Strategy and Planning to report directly to the management board member Alistair Curry. We are effectively developing the reference architecture of all the maps and have a strategic discussion on the investments we make around the different capabilities. One of the things we also do rather just defining the reference architecture and solutions from the business problem we want to solved. We do a lot of work on validation. This is probably a subset of that as <u>we are engaged with the university as a sort of vendor so we test and verify certain concepts. That helps us inform the future investment in that area and helps us validate the reference architecture activities as well.</u> So it is really great that we engage you guys and you provide us with some insights to us into what is</p>	

possible and we really value that.

A: Unfortunately we would have a broader audience. I think there will be probably follow up sessions following today's meeting we will invite a broader audience to present the outputs you have created. The way we use this is we typically present this to the various business and technologies stakeholders. We capture the learnings, we try to incorporate into our future thinking. It is not just standing still in particular in the dynamic financial services industry you need to look at the different way. The data visualisations is one of the probably top three priorities with all of your visualisations, we are really keen to understand how you can get some better insights on the large volume data we have internally as well as externally. So with that in mind, I am going to hand it over to you guys.

A: I spoke to Adam yesterday and I think we provide you a large set of transactions of the accounts payable in January last year?

Chunlei: Yes, I think is fiscal year 2013.

And I think you are doing this for your master thesis. And I am Monash alumni as well.

[Exchange name cards]

[Introduction]

Chris - lead manager VI analytic space. He is in charge of VI practice.

Dale, Technology Strategy & Enterprise Architect

Adam - EA

Tim - Graph visualisation

Chunlei

James - RA - BI - Funding set up lab building practical information visualisation - robust versions.

[Video introduction 7 minutes]

Can you just go back to the technologies has been used in the prototype?

[Discussion on the technologies] (Web services/ C#/ JS/CSS/HTML)

<p>T: See what we can do with html5. We are excited about the idea of using restful interfaces to build particularly collaborative interface. Tablet in private workspace and shared work space.</p> <p>They tend to work together, they lock themselves into the conference room, work on their laptops and periodically post their results on the walls.</p> <p>We want to reproduce and automate that public/private workspaces.</p>	
<p>A: One of the things with collaborative analytics is that typically users interact with visualisations, they would like to share with other users as well. They typically will have some interaction within the application where they put their comments in.</p> <p>-> <u>Preserving a particular view. Storing a list of URL that you can get back from that view.</u></p> <p>-> A more common thing is wanting <u>to annotate the data.</u> And <u>send to more senior members concerning that data.</u> e.g., we have an analyst lock himself in the room, gets whole lot of understanding and what the first they want to do is <u>annotate that data.</u> So the <u>next person or the more senior members can look at that data, can see analysis and the annotation the data is done to understand why the numbers are the way they are.</u></p> <p>-> They want to know why and what are we doing to stop it. <u>That's the annotation showing that: hey we have already done this, but what they really want to know is why, these are the steps taken to make sure it's not the case again in the future.</u></p>	<p><u>Preserving a particular view = 1</u> where user can go back to that view and see the comments</p> <p><u>Annotate data = 1</u></p> <p><u>Comparison analysis = 1</u> (Spend VS Revenue)</p>
<p>Tim: So, where would you like the annotations appear? Directly on the display?</p> <p>Dale: -> <u>Being able to see spend data and revenue assists analysis. Having large spend numbers is not the whole story, you need to see the revenue. We need to put these numbers in context that's what the analyst want to do, and they will be able to annotate,</u> and put the numbers right there.</p> <p>18% actually not bad compare with average.</p>	<p><u>Improving colour scheme so the hierarchy can be easily identified = 1</u></p> <p><u>Add a level indicator to tree map = 1</u></p> <p><u>Annotation =2</u></p> <p><u>Data in context =1</u></p>

-> The colour isn't immediately obvious to show levels. Suggest to add the board for each rectangle. The number in its self is not relevant.

A: -> if you click to the next level, it actually doesn't say which level of the org are located.

I suggest to add a level indicator on the group level -> you know where you are.

-> Highlight different levels of where you are.

Discussion on hard to see the colour and levels. Tim suggests adjusting the colors and adding a strong border as ways of distinguishing levels.

Enhancing clarity of navigation - at a glance verifying level and location in the hierarchy. A. suggests a sidebar showing current position in the hierarchy. So the breadcrumbs are in there, it's just a little hard to pick up at a glance.

What are other business units at the same level of that hierarchy and then do a comparison across. What you're doing in drilling down is really good, drilling down to the numbers, but it needs to have the context of being able to compare across, to do peer analysis.

[Discussion on the usage]

A: You could click tree map and update the map functionally but the laptop doesn't support that performance.

What are the different colour are representing?

Each circle represents a spend area and you can mouse over to see the town name, post code and its related spend amount in that particular area.

-> Suggestions a: it has a quite large number of cities. But it seems to be difficult to see because so many circle overlap each other. Maybe when you highlight this actually allows zoom and actually provide bigger views

Change an icon for breadcrumbs = 1

Comparison analysis = 2 (Spend VS Revenue)

Data in context =2

Circles are too dense on the map = 1

Google Map API = 1

Replace map with Google map API =2

so that you can actually zoom in the circles. You know how we can zoom in the Google maps, people can immediately see what spend is in this suburb or whatever, spent about this amount.

A: When you get the data from spreadsheet, it just load it into the memory, how long does it take?

C: you mean the first time loading? The dataset is about 100mb so it would taking half minutes to do that.

A: so it is relatively quick to load the new dataset.

C: yes, as long as it finish loading, the queries will be very quick to be performed.

A: I like the networking graph but for me it is a bit too busy.

C: yes, that is something we really want to simplify for next phase. Because for now it need visualise all the related vendors and sometimes it would be a lot.

A: so what is actually left and right represent?

C: on both sides, left and right are all vendor companies. For example

A: so the spend category in the middle is linking to the vendors, right?

C: yes.

A: so when I look at this, I immediately link the left item to the right item and I look for the linkages because I was thinking particular vendor related to other vendor. We paid money to them and they paid to somebody else.

Chris: I actually mistakenly think that was business unit on the left and vendors on the right. And it is usually have this spend of different business. That's is actually on my mind.

A: that might be actually a better way.

T: That's interesting.

A: so if you have a business on the left and you actually have a hierarchy on the relationship. You click on that and it will show you all the vendors in this particular business unit. It is actually link up to your previous screen and follows the paths.

T: Also, Chunlei, you can filter this view by dragging specific business unit

C: -> the other one is a lot of animation we seen from other tools, they

Confused by network

layout = 1

(With too many data and nodes)

Confused by network

layout = 2(at first

sight users tend to think left node have some relationship with the right nodes, and linkages are representing the relationship)

<p>animate to the point of where you got results, then they stop animation. No more than 3 seconds. The research shows that if the animation keeps going it is actually a distraction to the point people are not actually digesting information you are presenting because it is still moving. E.g., the bubble is still moving when I trying to watch the number. So animation is cool, it is a cool factor. Bring it in with a bit animation and stop, so that people can focus on the numbers.</p> <p>C: for current mechanism, it is 5 seconds for animation, then it just stops.</p> <p>A: So when you click on that bubble, what is that saying to me?</p> <p>C: this is software maintenance category, and in total have displayed spending. It is all related to vendor companies.</p> <p>A: so why is that bigger than this?</p> <p>C: that is because this company probably relationship with other spend category.</p> <p>T: so that is total spend on that vendor.</p> <p>Clifford: the problem is when you put it in context, that's 1.5 mill, the other one is 105 thousand in the side. It doesn't match the particular context.</p> <p>C: yes, right now we can't distinguish in particular context.</p> <p>A: yes, we should highlight spend related category is in the context of the overall spend of this company. When you look at this, why this is bigger and this is smaller? Because immediately the size of the bubble to the amount of dollars.</p> <p>A: but I think once you understand how it works It can be very powerful. It is fantastic.</p> <p>A: can I try it on the machine? What is the cost of this big screen?</p> <p>T: it take us about four thousand for us but they are about to get a lot cheaper.</p> <p>C: let's see the speed of the initial loading. It will take about half minutes.</p> <p>A: so every time you actually start the application it needs to load the data from the memory?</p> <p>T: The idea is we will have a server there all the time.</p>	<p><u>Annotation =3</u></p> <p><u>Animation overuse=</u> <u>1</u></p> <p><u>Data in context =3</u></p>
--	--

<p>A: when you drag and drop from this view into this, does it automatically update all of the vendors?</p> <p>C: Yes. All of vendors,</p> <p>A: so immediately we I look at this I would say you have too many vendors. The visualisation is difficult.</p> <p>A: can I select just one vendor?</p> <p>C: yes.</p> <p>A: -> <u>in the map it will display the history of queries but at here it doesn't have. [Inconsistency]</u></p> <p>A: so let me have a play because I am interested in it.</p> <p>[Touch and try]</p> <p>C: the response of the screen is a bit flat.</p> <p>A: so when I click on technology, this doesn't update.</p> <p>[Discussion about data]</p> <p>A: it is interesting, because do this you can immediately identify the data quality issues because something doesn't make sense.</p> <p>A: This is fantastic, I love it.</p> <p>A: What do we need to set up to run this application?</p> <p>C: you need a web server, then you can run the client on any device. Right now we are using temporary host on another server.</p> <p>[Discussion on data]</p> <p>A: it is fantastic insight.</p> <p>A: it is actually quite interesting because one of the things I am thinking is it is great to have visualisation, but financial people just want rows with numbers.</p> <p>[Using touch to zoom]</p> <p>T: the idea is that people can collaborating on such tabletop together.</p> <p>A: [compliment]....</p> <p>[Showing the line chart]</p> <p>A: explain to me what this view is about?</p>	<p><u>The inconsistency of the queries history</u></p> <p>=1</p>
--	--

C: this is detailed spend information by filtering a specific company.

A: I just have some new finding, you see we spend most of the money, just before the financial year. What's this, is it representing the amount?

[Discussion on the data]

[Photo]

Clifford: What is the difference between two views there?

C: the view below is like a miniature of above view. It like a zoomed view and this represent a full year and this represent a full month.

A: this is great. We should present this to our GM. There is nothing on the market offer view like that.

C: the touch experience is actually a lot better on iPad.

A: This is your network links.

T: This is not at all the most sophisticated network view. This is one to grasp the quick concept.

A: I think what will be really useful if we can have a Google map and zooming and just bubble and information.

Clifford: yes, you can actually get straight into the Google apps API, and the google can give you entirely world based on what postcode you got in the list.

Chris: so how many man days do you think to develop it?

C: It would be about 100 days?

Clifford: I am interested in where you come up with criteria. You could add heat maps and try to identify the areas effectively on different usages of bubble charts. Is that something's you done with Cathy or is that something you done based on research saying this is the best way to display analysis data? Is it based on technology you wanted to try? How did you come out with the selection of these visualisations? What sort of selection process you going through? You've chosen heat-map, line

Identify patterns = 1

Confirming

Google map API =2

<p>chart... there are a lot of other visualisation methods out there, what guided your choice of the ones you have used here.</p> <p>T: It was a bit ad hoc. Just through the discussions and looking at different techniques. Probably a big part of Chunlei's contribution, as much as the architecture and concept of how you combine different visualisations collectively. The actual choice of the visuals used is almost arbitrary rather than a scientific process.</p> <p>Clifford: One of the big things in business intelligence is there are a lot of tools on offer, what vendors have is basically a grab bag, they have a dozen of different way to display it, everyone just wants to design a solution. <u>What we want to do is take a step back and consider the business process, how is this actually going to be consumed, what is the business process that you are going to walk through when you are consuming this solution.</u> From there that's where we can try to tailor the presentation of the data and the business information to support the processes. We need to understand the business process the visualisations are intended to support and then understanding those processes and tying that back to visualisation styles. <u>Accountants still just want their spreadsheets, they will still want to export to Excel.</u></p> <p>T: [discussion of what James will be working on] I guess, as a student research project we can just jump straight to the fun part of this.</p> <p>C: what is the actual thesis about?</p> <p>T: the general theme of the thesis is involved with exploring collaborative analytics and the software architecture for supporting that and the touch table and the different devices, allowing a shared workspace.</p> <p>C: How are we defining collaboration?</p> <p>T: <u>Allowing individual analysis and coming back to the table with insights.</u> It's very broad strokes, we'd like to hear your ideas about how that could work.</p> <p>C: Is it about communication? Would the addition of video of others analysis be useful? Tracking of comments? There are lots of different elements of collaborations. <u>And then there's the level of reporting, we do a lot of regulatory reporting, where can I annotate? Where can I put in my</u></p>	<p>Export data to spread sheet =1</p> <p><u>Collaborative scenarios = 1</u></p>
---	---

<p><u>work flow process? Where can I have conversations and store the conversations? And how can I see what this person is saying at the same time? And how can it capture all of that stuff? It's got to work for our process.</u></p> <p>A: [discussion of ANZ internal accounting processes]</p> <p>You know what, I think this is all great but when you actually go and report to senior stakeholders you need to provide the artwork of your analysis.</p> <p>[Discussion on data]</p> <p>A: speed is ok.</p> <p><u>C: this is not actually full dataset. I have delete all transaction under \$20 in order to better performance on this laptop.</u></p> <p>A: When you actually go back and present to senior stack holders you need to show the outcomes of your analysis and have let's say 5 or 6 points. We have peaks in our costs in March. <u>Have you done any analysis to identify our patterns here? (For example in other years is this comparative?)</u> So is there any analysis to show that there have always been peaks in March. The data might show that there are drops in December January due to people on holidays. So you can see that people come back (from holidaying) and spending goes up. It's been shown that 30% of spend is unclassified and we're talking about 1.8 billion dollars, that's a lot.</p> <p>Tim: what would that actually do...?</p> <p><u>A: Well it could be many things it could be that people are hiding costs and deliberating not classifying it so that it won't be assigned to that business group</u></p> <p>A: when was this? Was this 2013?</p> <p>Tim: yes</p> <p><u>A: on the credit card payroll system there was a flaw in the system, so say</u></p>	<p><u>Performance =1</u></p> <p><u>Annotation =4</u></p> <p><u>Pattern identification = 2</u></p> <p><u>New findings</u></p> <p><u>Pattern identification =3</u></p>
--	--

<p><u>that if you didn't classify within a certain period, you couldn't reconcile the books and people would just dump these off into a generic account and so it never got traced.</u></p>	
<p><u>A: But you could figure it out if you looked at the vendors. It should obviously be part of risk management.</u></p>	<p><u>Annotation = 5</u></p>
<p><u>C: It would be useful to put annotations to tell the story, you want to tell insides of the business, you can't go inside and then replicate the analysis every time.</u> If you just had the view saved at this particular point and just basically highlight this and say spanning on instant 59 it's not in our vendor network</p>	<p><u>Preserve a link = 2</u> <u>Share the link = 1</u></p>
<p>A: I think it's a powerful tool to gain insight</p>	
<p>Chunlei: the idea is it that we can actually <u>preserve the link and then you can send it via email then the other people can click on the link then it would connect to the server</u></p>	<p><u>Annotation = 6</u></p>
<p>A: we were just saying <u>if you had a commentary box it would be very useful</u></p>	
<p>Tim: a lot of the time when we're showing presentations like that, as much as we're showing the concept, we're showing a real life example of how they will use it. And I think that's what Sunny's saying</p>	<p><u>Annotation = 7</u></p>
<p>It's... Okay so as you can see, here is our heat map as you can see here <u>there are Australian retail is here da ding double click and you can see this guy da ding and there you can see this this... And there we've driven down to this data and there's a problem here and we can investigate that further on our network chart and once we've seen this we can look at this. Ok so here's a real life scenario of where we've identified an issue and you can save that as a link and share that up. It defines the story and at the same time you're demonstrating the features of your tool as well.</u></p>	

Sunny: The other one is that obviously people in particular finance people will look at the percentages and identify immediately the issues ... you know you have 600 million here and zero percent, so that's not zero percent of that. What does that mean? What does that zero percent represent?

Chunlei: that's zero percent of total

A: but it still shouldn't be zero because 300 million dollar sub 4 billion is what..?

Chunlei: probably the amount is less than three decimal points

300 billion out of 4 million... no that wouldn't be right

[Bugs on percentage display on 6th layer in tree map]

Tim: I think we've found a bug

A: a lot of visualisations are useful because they help you identify the data issues. Because people will assume you have a clear/clean dataset and you'll never have issues and you'll create those reports and you spend a lot of time transforming the data into a form that can be used. So something like this shows you that there are some significant data quality issues but we can't manufacture and create the classifications because the classifications don't exist in the source data. And sometimes by focusing on fixing this you can actually reduce our spend. Because you improve the transparency around your...

So what I would suggest I'm not sure I'll talk to Allan we definitely need to get Patrick across this and probably Sarah as well because it would be very useful.

Tim: so we've still got half an hour do you want to do the tasks just to see the

Bugs in view =1

Identify the patterns

Chunlei: actually, can you guys play around a bit more with a few of the tasks and give feedback

Advanced search =1

Tim: we have no idea what these tasks really interest you or not it's just to systematically explore the tool and get some more feedback

A: It would be really useful to have a search for the different types you know say like categories you can select categories and do a search and it immediately creates that view for you...

Tim: so you've got a search there just on business units... Is that alright?

Chunlei: that's just a search on business units

A: so does that update all automatically?

Chunlei: this one just updates map and the other one updates the network

A: how do we clean these one's up?

Chunlei: you need to refresh everything

You need to have the ability to remove those (search histories)

A: So refresh and start again

Can you change the time period of the tree map view?

Sometimes the sum amount of the category is not equal to the sum of its vendors' spend this is another bug

So postage costs being 220 thousand I'd expect to see all the vendors that add up that 220 thousand

So when I look at the sum of all these in 220

Chunlei: I will add a feature to make the temporal filter work on both map

and tree map

A: So it would be good if you had this so it follows the graph and shows the numbers

A: Can you make it like the other ones but just see Australia though?

Tim: This system has really been ready and this is the first time we've gathered feedback of actual content. The whole reason for this, is so users can interpret data... The whole reason we're doing visualisation is we're trying to encourage a better way to interpret and understand data that's more intuitive so the rationalisation for why you're using certain visualisations in certain ways is and what's the goal you're trying to get to is just as important and it's really fundamental to the underlying process of intuitiveness as much as the visualisation is itself.

(The reason for visualisation technology is to give users a better more intuitive understanding of different datasets. So that people have a comparative, more holistic view of a certain dataset)

C: So this is the design that you're actually developing?

Chris: So we've got an enterprise business intelligence platform which is oracle. We've purchased oracle business intelligence enterprise edition and that's implemented here. So this is the style guide for that platform so that's one of the platforms that I own and anyone that sells. So I allow business units and technology people to build ports in that space and to send that out to a wider community and just as you have your operation component there we have set reports that everyone uses and we also have analysts that can come in and actually take that same data and we can analysis it further and they can push those result sets up to a shared space where other analysts can have access to it as well and they can push that out as well. So very similar to what you're talking about the collaboration here

But what we model is a set of styles so that when you have a user come

Data in context =4

in, regardless of whether or not they're looking at HR data or Australian retail data the look and the feel of the reports and the records are the exact same. So it doesn't matter the context of the data they already have a level of intuitiveness and can consume the data and then it's just a matter of subject area, context of what they're trying to get across. That's why these style guides are really important and for me something like more researching spaces and going beyond that and really going into the whole how's and whys of why you use visualisations and these tools, what makes a good case to use visualisations or not use visualisations. Being able to document that and clarify that in the form of a guide of dos and don'ts

Tim: There were principles of standard rule of thumb of designing visualisations providing that top down view of the data so that you can see the whole universe. And then we shifted in 2010, we started getting user feedback and we found that people were getting really excited about the that first but they didn't finish up using on a day to day basis so we did all this study that looked at why, and what we found was that very much like people would see a complicated network view for the first time they would they'd get scared there's too much happening and that was it a huge turn off so we re-engineered to be centre around their standard work space which for developers is ... so that instead of the universe visualisation to a detailed slice of assistance that add to the visualisation directly from the code and build up a much more focused diagram where everything from the diagram was there to query so that was much more useful and in the end a couple of weeks later it really ended up being the most used parts of the Tools

C: I get that because they've already got the context from the visualisation and now there just looking at the numbers and figures -- understand the bigger picture and being able to but the data they're looking at into context and thus making it easier to understand.

Data in context = 5s

Table 9-2 Transcripts of heuristic evaluation