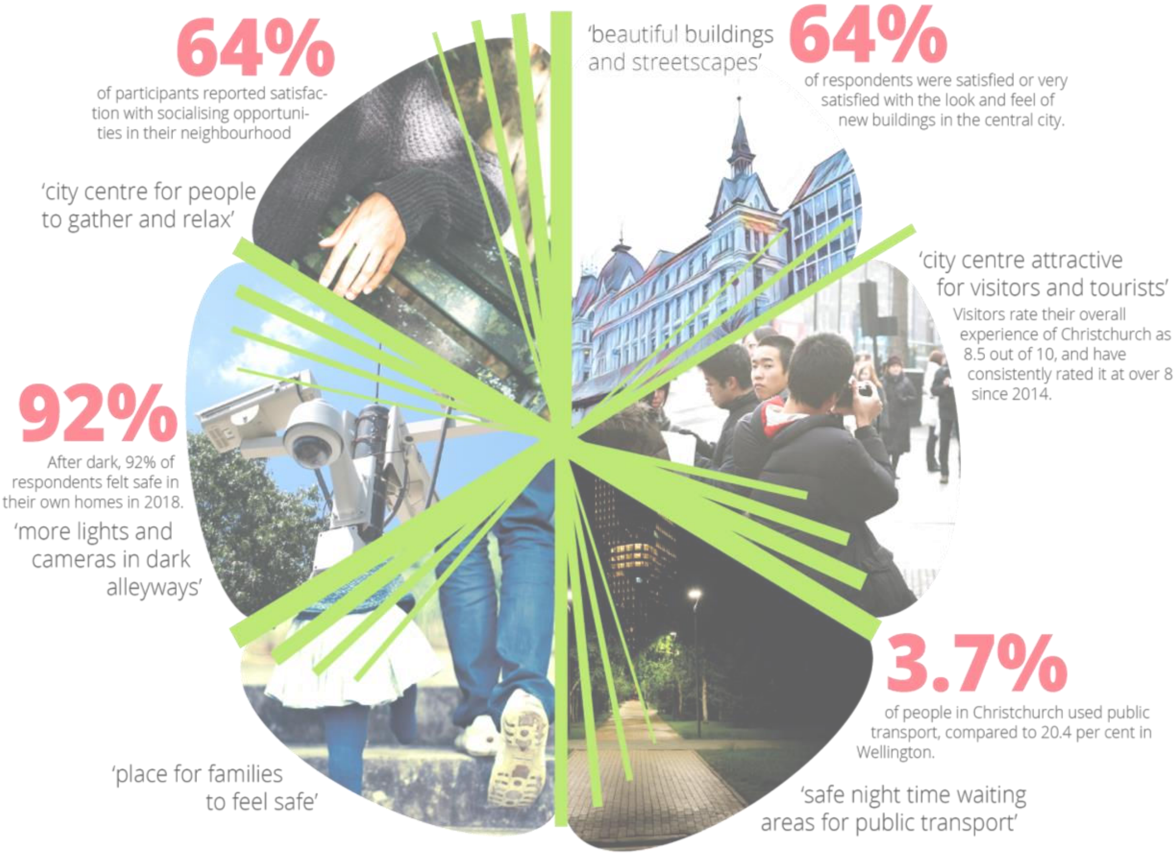
BUILDING BETTER HOMES, TOWNS AND CITIES | Ko Ngā wā Kainga hei whakamāhorahora | National Science Challenges

Data Collection, Data Analytics, Data Visualisations and Data Storytelling



Mark Dyer, Rachel Dyer, Tomás García Ferrari, Min-Hsien Weng, Jordan Wilson, Rebekah Wilkins and Shaoqun Wu.

Building Better Homes, Towns and Cities:
Urban Narrative June 2019

Building Better Homes, Towns and Cities Working Paper

Title	Data Collection, Data Analytics, Data Visualisations and Data Storytelling
Author(s)	Mark Dyer Rachel Dyer Tomás García Ferrari Min-Hsien Weng Jordan Wilson Rebekah Wilkins Shaoqun Wu

Acknowledgements

This research was funded by Building Better Homes, Towns and Cities National Science Challenge.

The paper acknowledges Urban Narrative is partnered with Napier City Council, Christchurch City Council and OPUS. However, none of these organisations are responsible for the information in this paper. Every effort has been made to ensure the soundness and accuracy of the opinions and information expressed in this report. While we consider statements in the report are correct, no liability is accepted for any incorrect statement or information.

Recommended citation

Dyer, M., Dyer, R., García Ferrari, T., Weng, M-H., Wilson, J., Wilkins, R., & Wu, S. (2018) Data Collection, Data Analytics, Data Visualisations and Data Storytelling. Waikato, NZ: National Science Challenge 11: Building Better Homes, Towns and Cities.

Building Better Homes, Towns and Cities National Science Challenge

Private Bag 50 908
Porirua 5240
New Zealand

www.buildingbetter.nz
buildingbetter@branz.co.nz

Table of Contents

1. Introduction

Cities are more than infrastructure, they are communities of individuals and families with different backgrounds, needs and aspirations. To make a city thrive the needs of all these inhabitants must be met. But, meeting the needs of communities can only happen when citizens are engaged in the shaping of their cities and engaged in the design process from the very start.

The United Nations Sustainability Goal 11 highlights the challenges faced by planners in managing urban population growth, emphasising 'growing numbers of slum dwellers, worsening air quality and insufficient basic urban services and infrastructure' as a result of urbanisation (United Nations, 2018). These statistics indicate a failure to sufficiently address the needs of those who live in cities.

Urban practitioners need to listen to 'what' people need, but also understand the values that underpin these needs in in order to understand 'why'. Using the collective knowledge and experience of communities by making values the basis of a design process will inherently lead to more innovative and effective design solutions, rather than predetermined design outcomes.

This collaborative approach to urban design is not an easy task. It requires an entirely new set of attitudes, systems and tools. Key to bridging the chasm between top-down and bottom-up planning processes is community capacity building through the promotion of data literacy and the co-design of digital tools to enable a more collaborative approach to urban design.

With the above in mind, the Urban Narrative Group have worked in partnership with Napier City Council and Christchurch City Council with two separate neighbourhoods that involve Maori and non-Maori communities to develop a suite of prototype digital tools and protocols for data storytelling to inform value based urban design.

This report outlines the design, development and deployment of these tools which are designed to facilitate evidence-based decision-making at multiple scales. These tools are capable of interrogating and disseminating large qualitative data sets (i.e. citizen stories) using syntactic data analysis to reveal citizen values and infrastructure needs.

The tools have been developed comprising data collection, analysis and visualisation techniques for data storytelling. The chapters within this report outline the development of all these techniques and showcases their application using data from a public consultation undertaken by Christchurch City Council, post-earthquake, 2011.

A basic step-by-step guide to the toolkit of these techniques from data collection through to data storytelling is shown Fig. 1.1. The term data storytelling is used to describe the evolving narrative of a place based of citizen stories, or 'lived experience' of a place, in terms of their values and demands on infrastructure.

The research recognises the city to be a complex system operating at multiple level. The toolkit offers a means to understand the urban system through the relationships between infrastructure types and the values that weave them together, allowing practitioners to respond in a more agile way to the continually changing needs of people, improving the livelihood and liveability of cities.

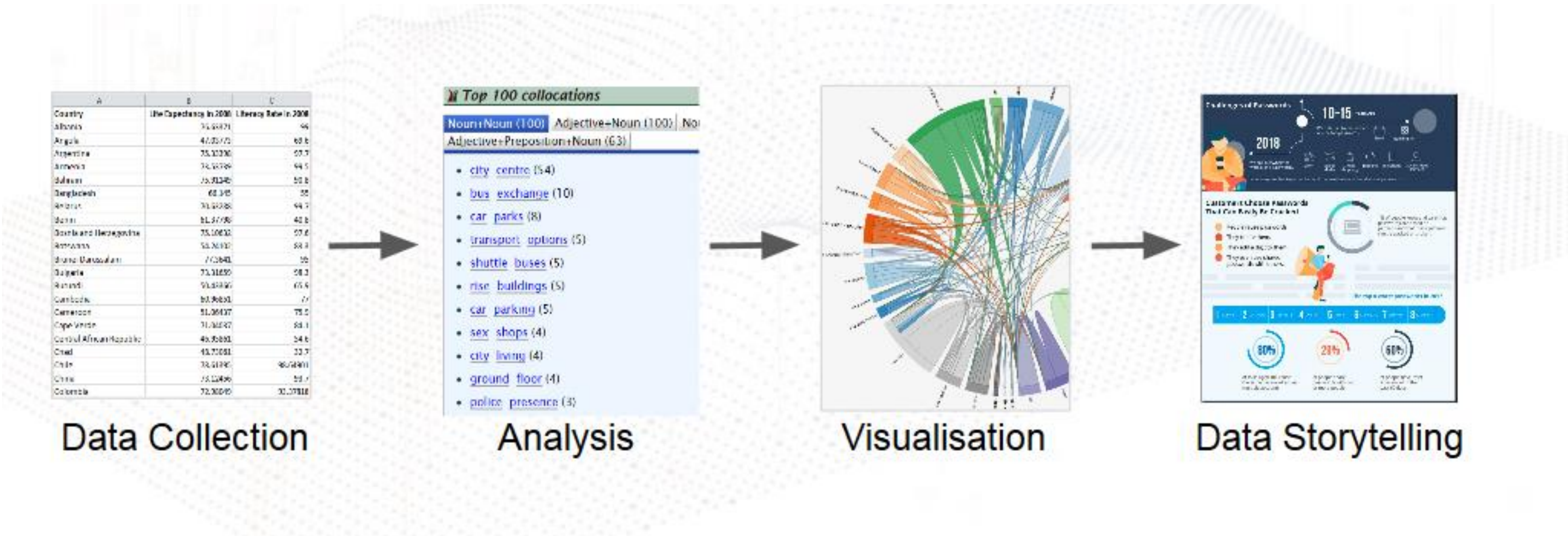


Figure 1.1 Flow Diagram Illustrating the Individual Aspects of the E-Toolkit

2. Data Collection

2.1 Introduction

This chapter details the data collection app, also referred to as webCrowd or the app. The app has been developed with the intention of capturing large quantities of data relating to people's lives in an urban setting. WebCrowd aims to allow people to share their stories in an open way without influencing their views and opinions by using a simple medium and easy-to-use interface. This method enables people to share what they want when they want, in the realm of ‘living in an urban environment’. WebCrowd must be easy to use and accessible to as many people in the partnered communities as possible so as to fully represent all demographics and views.

The aim of the app is to create an experience that is rewarding and simple. An experience can be rewarding when one is able to see what others have to say and know what is going on in the people's lives around them. This mostly relies on the people in these communities actively engaging with the app. To achieve a rewarding experience, simplicity of the app is crucial. A simple experience can be achieved by the ease of use of the app.

2.2 Development Process

WebCrowd has been the product of many design iterations and features that are distilled down to create an accessible and intuitive interface. At first, the design of the interface was very simple with a single person contributing to the design of the app. The focus at this stage was what to display and what is interactable on the interface. This gave an overview of what the app would be and how users would interact with it. Through our weekly team meetings, we arrived at two main objectives. One is to streamline the process of informing the user of the purpose of the app, and the other was to direct people to share their stories using the interactive map.

The page initially presented to the user was a subject of great importance as it gives the user the first impression of the app, which meant the page had to be easy for the user to understand. At the beginning, the first page was designed so that users would share their stories there, allowing them to instantly start using the app without any disconnection or disruption. However, as the project moved forward, it became apparent that the information and introduction of the app needed to be separated from the main feature of story-sharing. This separation brought a flow to the app with the landing page being specifically made for a brief overview of the app and the partnerships that have been made.

The page with the interactive map (share page) was also important as it was the key page in the app which houses all of the functionality. On this page, users are not only able to share their own personal stories but also view others' stories that have been already added to the map. In order to achieve these features, it took many iterations of the app. Each time, a part of the previous version that worked remained and was combined with a new layout and new features until these features were filtered down and the layout was finalised. Through this process, we created step-by-step instructions to give the users clear directions on how to use the elements on the page so that they can maximise their experience and are motivated to return to the page and see what others have shared.

Data visualisations are very important parts of the project as they allow the users to see and understand what is happening with their stories. The data visualisations were originally on the landing page of the app, but as it became clear that they were important not only for the researchers but also for the users, it was decided that the page dedicated to the data visualisations would be made. The development of this page was relatively straightforward compared to the other pages, but it provided more structure and flow to the app. The main idea for this particular page was to separate the various visualisations from one another so that they could each be viewed independently of one-another.

Finally, the about page simply gives a more detailed overview of the project as well as some accompanying images related to the project and the work of the team. This page, however, does not provide information such as the individual team members since the Urban Narrative web page provides all of that information and is linked to from webCrowd.

2.3 Development Tools

Throughout the design process, many changes were made in the code base. As new features were introduced, some features were removed, which left a lot of code obsolete with new and optimised code taking its place in the function. The code base was rather disorganised as it was in a state of constant change with many features being added and removed. Therefore, once the design had been finalised, the remaining code was cleaned up to optimise the interactions.

During the development of the app, it was decided that certain development tools were to be used. The app itself would be made using a framework called Angular, which would be run using a server called Tomcat from Apache. We would also use the relational database MariaDB and the reverse proxy Nginx. Later, the visualisation side of the app was separated from the Angular app, as it would require more access to the database to create the visualisations from the data. Figure 1 shows a chart showing the map of architecture used and the flow of a request through it.

2.3.1 Angular

Angular was chosen as the framework for the front end since the development time for adding features and deployment time are low, allowing a fast coding experience and addition of advanced features with relative ease. With large amounts of data being recorded inside the app, the components must be able to handle many different occurrences, and we can be certain that Angular is capable of this with the code provided by Angular modules. The Angular command line interface (Angular CLI) makes it easy to perform specific tasks with commands. Consequently, it makes compiling and hosting the app very streamlined, maximising time spent creating the features.

2.3.2 Tomcat

Tomcat is a server for hosting web-applications and services. Although it has been in use for a long time, it is very reliable and being improved and updated regularly, which were some of the reasons it was chosen to be used as the server. Tomcat allows for fast deployment in conjunction with Angular and is easy to configure in order to achieve the goals set in place. In addition, Tomcat works in parallel with our reverse proxy Nginx. This reverse proxy handles all incoming traffic (Requests) and distributes them to the correct places. Together with Tomcat, Nginx is able to handle upwards of 10,000 connections without slowing down. This means that a large number of people are able to access the app simultaneously without noticing a change in the load time or the service becoming unavailable.

2.3.3 Javascript

The visualisations were created using Javascript and the D3 library. They were made their own service within the Tomcat, not only because Angular and D3 library do not integrate well but also because the logic of creating the visualisations can be kept separate from the main app. This separation allows the computation of the data visualisations to be done on the server and not on the user's computer. The Javascript program creates the visualisations from the data for the users to see, as opposed to the user's computer requesting the data from the Nginx and then using the returned data to create the visualisations.

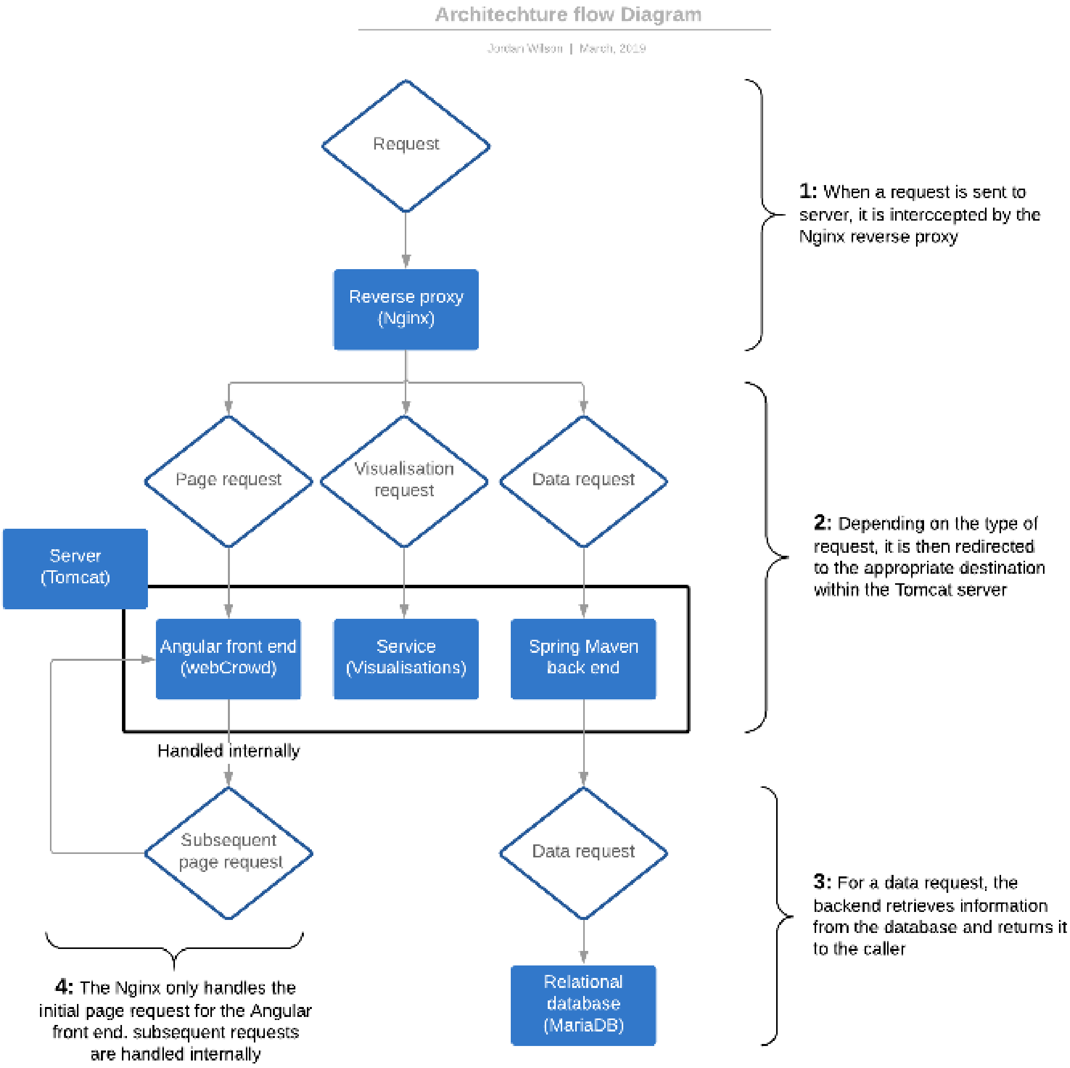


Figure 2.1. Architecture flowchart showing the interactions when a request is sent to the server.

2.4 WebCrowd

WebCrowd consists of four pages - a home or main page, a share page, a visualisation page and an about page. These pages serve the different functions of the website and provide the ease of use of the app as the users may go different routes through the app depending on the purpose of the visit. The separation of the components prevents the app from being overwhelming to its users, as each page has its distinct purpose, making it easy for users to understand what each page is for.

2.4.1 Pages

The home or main page for Addington is shown in Figure 2. The main purpose of this page is to briefly introduce the participating communities and to illustrate their city with an image. While this page does not contain much information, it is important that those who are visiting the page know who is involved and whom those communities are made up of.

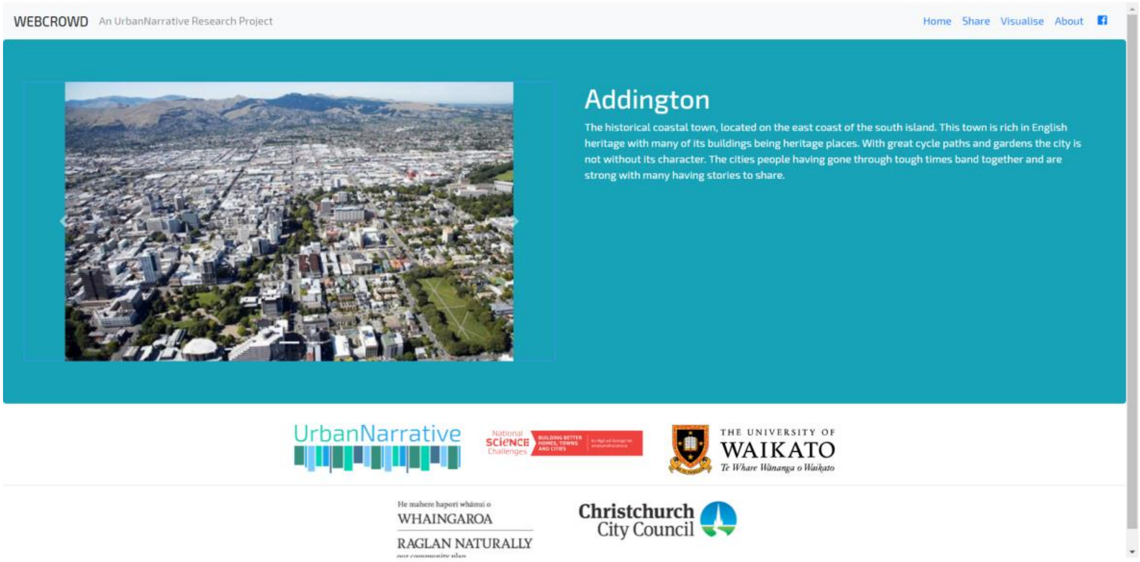


Figure 2.2. Main page of webCrowd

The share page, the user is able to share their story or view past stories that other users have shared, is the main part of the app and where most of the logic is performed. This page plays a significant role in determining whether or not users would return to the site. The interaction on this page must be simple and fast whilst providing adequate information about other stories. This in return should entice the user to return to the page and ultimately share their own stories more often.

One of the aims of this app has been to offer its users a simple experience when using the app. This was implemented by simplifying the instructions. The user is presented with a 3-step worded guide of how to share their story. First of all, they are provided with categories of the story and must choose one from them. Then, they place a marker, which represents the place of interest, on the map. They subsequently fill out their story in the text field and optionally upload a photograph to illustrate their story. Once completed, their story will be visible on the map as a marker that shows the story connected to it when clicked. This page also provides a feed of the most recent stories, sorted by the time the story was created. This feature aims to keep users interested in the page and to motivate them to return to the page to see the latest shared stories. Figure 3, 4 and 5 show the example of the share page and its interactions.

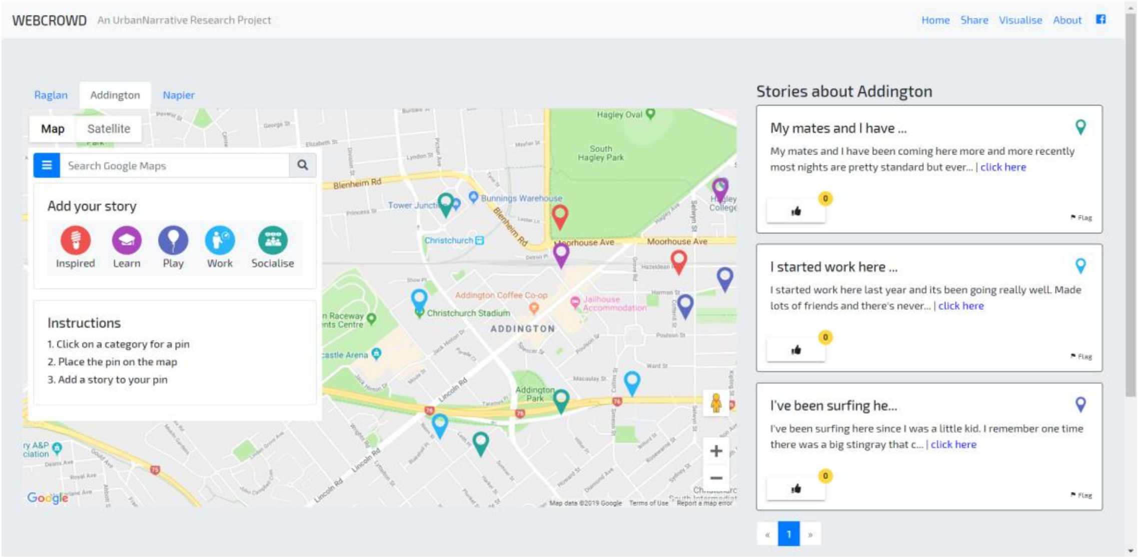


Figure 2.3. Share page of webCrowd

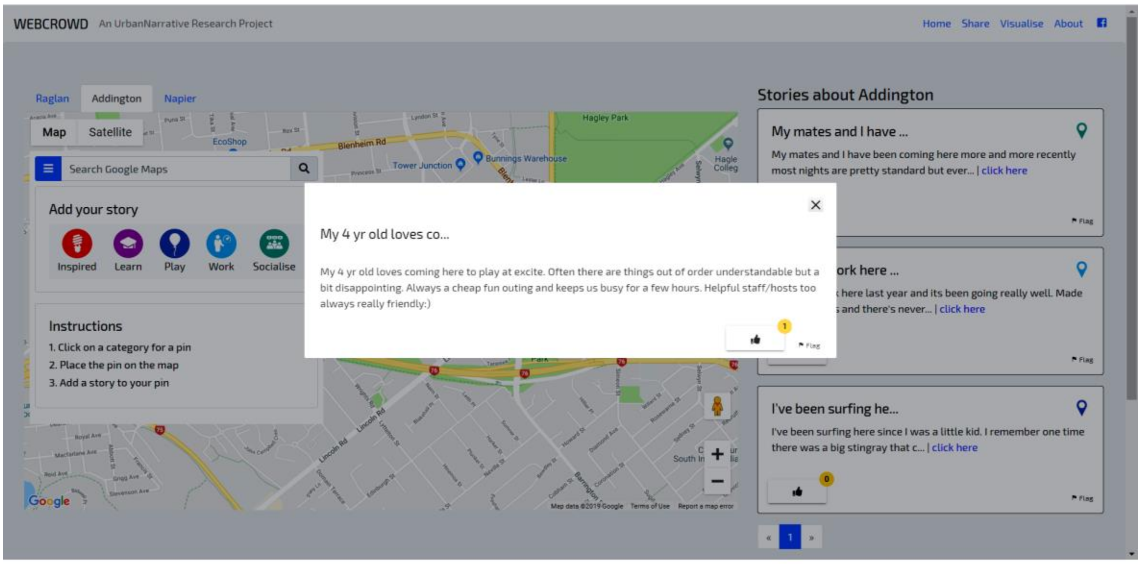


Figure 2.4. Marker of previously added story shows the story when clicked

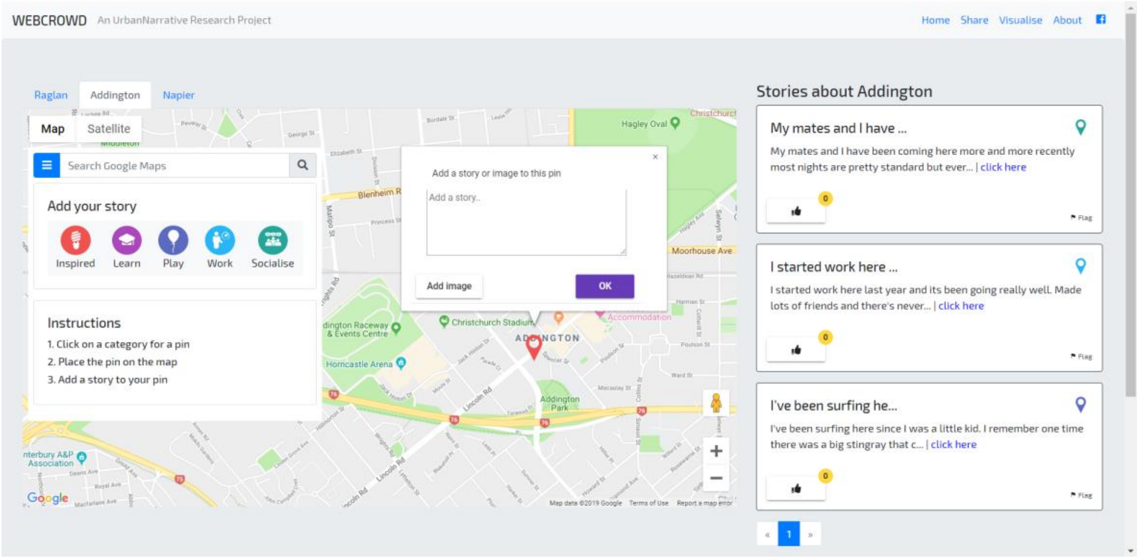


Figure 2.5. Example of the interface when adding a story to the map

The about page is devoted to informing the user not only about the webCrowd but also about the project in general. It provides an overview of what the project entails as well as images from past events held by the Urban Narrative team. The users of the app may visit this page to acquire necessary information such as the detail of the project or the process of the collected data. An image of the about page can be seen in Figure 6.

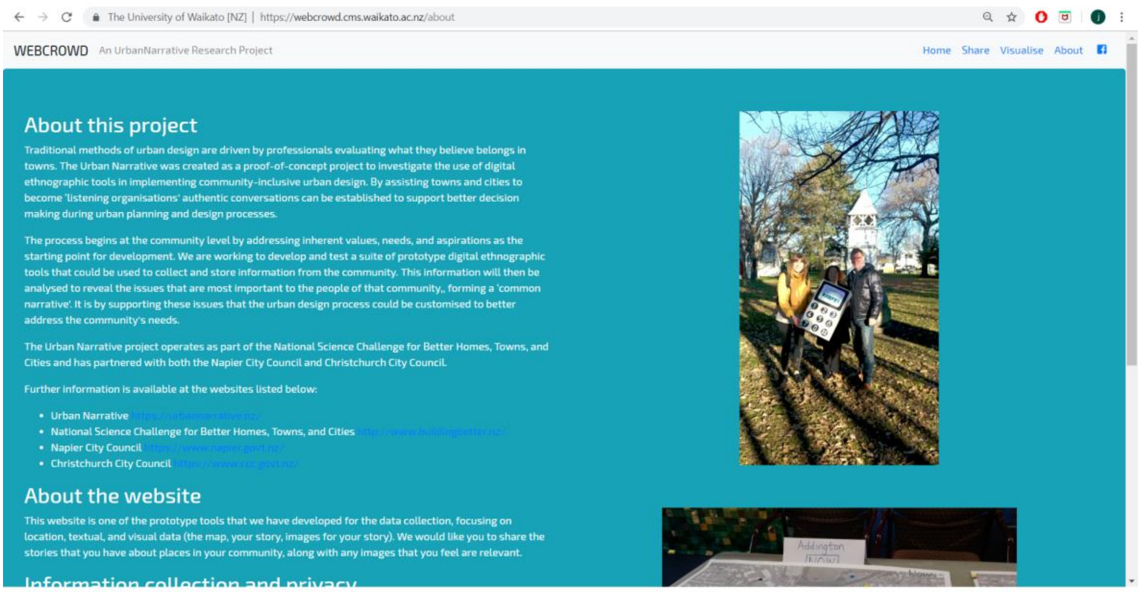


Figure 2.6. Image of the about page of webCrowd

2.4.2 Generalisation

The current app is a generalised version and can be customised to each participating community. Various changes can be made to accommodate differences in communities to bring out a personalised and unique sentiment. This localisation can be achieved in a cultural way by representing the diversity of the participating community or in a societal way by representing demographics or current events in the community. Customising the app to each community would help people to feel more connected to the app and to engage more with the service it provides. It is therefore important to represent each community appropriately and accurately.

The generalisation of the app is achieved by removing elements related to specific cultures or communities so that the all participating communities can be shown on the single app without being related to a certain demographic. This generalisation is essential where some communities have a large presence of one demographic while other communities may not. Using a generalised version of the app would avoid representing one community accurately but misrepresenting another as the app has been simplified to a neutral colour scheme without cultural icons present. On the other hand, the multiple versions of the app could be created, with each one being tied to a single community. The localised versions of the app would be unique to each community but with the same layout to keep coherency.

2.5 Data collection

The data that we collect is primarily text-based with some metadata including time and location. The data by itself can be very hard to categorise and subsequently analyse. To improve our categorisation and final results, we require the data to first be categorised by the user upon creation. Then using these categories as a starting point, we can sort the data before analysis.

2.5.1 Values and Activity

The idea of categories and values has been a quintessential part of the Urban Narrative and has been developed over the workshops involving the communities in Napier's Maraenui and Christchurch Addington. The Maraenui workshop focused more on the values and needs of the Maori community, comparing the different value sets and design principles of Maori with that of Maslow's hierarchy of needs. At the workshop, participants took part in various group tasks to ascertain the key values from the two sets in relation to Maraenui. The key values that the participants decided upon were:

- Whanaungatanga (Relationship/kinship)
- Wairuatanga (Spirituality)
- Safety
- Kotahitanga (Unity)
- Orangatanga (Health and wellbeing)
- Mauritanga (Life force)
- Self Esteem

- Manaakitanga (Hospitality)
- Matauranga (Knowledge)
- Kaitiakitanga (Guardianship)
- Rangatiratanga (Chieftainship)
- Sense of belonging

Following a preliminary review of the data, we were able to translate each value into actions and guiding principles for urban planners (Dyer, M., Hinze, A., & Dyer, R., 2018).

During the Addington workshop, there was an importance put upon time. (I.e. where participants go now and where they went in the past). During the workshop they were asked to place dots on two maps, one representing present day and one the past. These dots indicate what they would do at certain locations and how the passage of time has affected their path (M. Dyer et al., 2018). The activities that the dots represented were:

- To work
- To learn and upskill
- To be inspired
- To relax
- To meet family and friends
- To remember the past

From these two workshops, five categories were chosen to appear on the app with the intent of not limiting what people can share. The five chosen were Inspired, Learn, Play, Work and Socialise. While these categories represent the Addington workshop in full, the values taken from the Maraenui workshop identified the values that we are looking for from the analysis of the data.

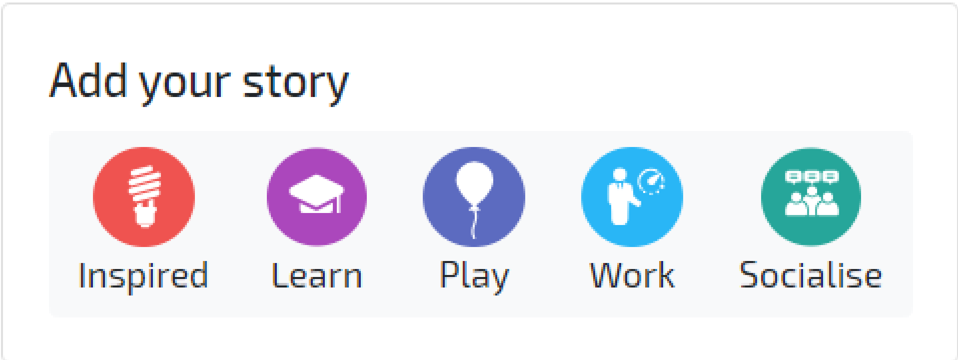


Figure 2.7. Categories of webCrowd

2.6 Technical Summary

The main computation of the app takes place on the share page. This page, as discussed, houses all the functionality for sharing and viewing stories. When a user navigates to the page, they are presented with an interactive map that they can use to share a story. To share a story, the user must first choose a category from the ones presented by clicking on it. By selecting a category, a marker unique to each category is selected. When the user clicks on a location on the map, this marker is placed on that location, allowing them to add a story to the marker. Figure 5 shows the example of the interface of this process. A user has selected the 'Inspire' category and placed a marker on the map. Once a user has clicked on the map, an empty text field appears for them to enter their story about that location. Their entry is then added to the data structure which consists of various pieces of data collected from the map and system. The main pieces of data that we, as researchers, are interested in are the date the story was created and the location clicked on the map. The date is in the form of 'day, month, year, time' and recorded at the moment the user clicked on the map and began writing their story. The location is provided by Google maps and consists of the street address, zip code, city, country, the latitude and longitude. The latitude and longitude are most important among location data because they are used to place the marker on the map in the correct location. With the use of bounds of the latitude and longitude, it is also possible to easily see how many stories were added in a certain area. Once the user has finished writing their story and pressed OK, the story is sent over HTTPS to the server where it is saved in the database. This ensures that the users' data is safe and handled properly. A flow chart of what happens when a user adds a marker on the map and when a user shares their story are detailed on figures 8 and 9.

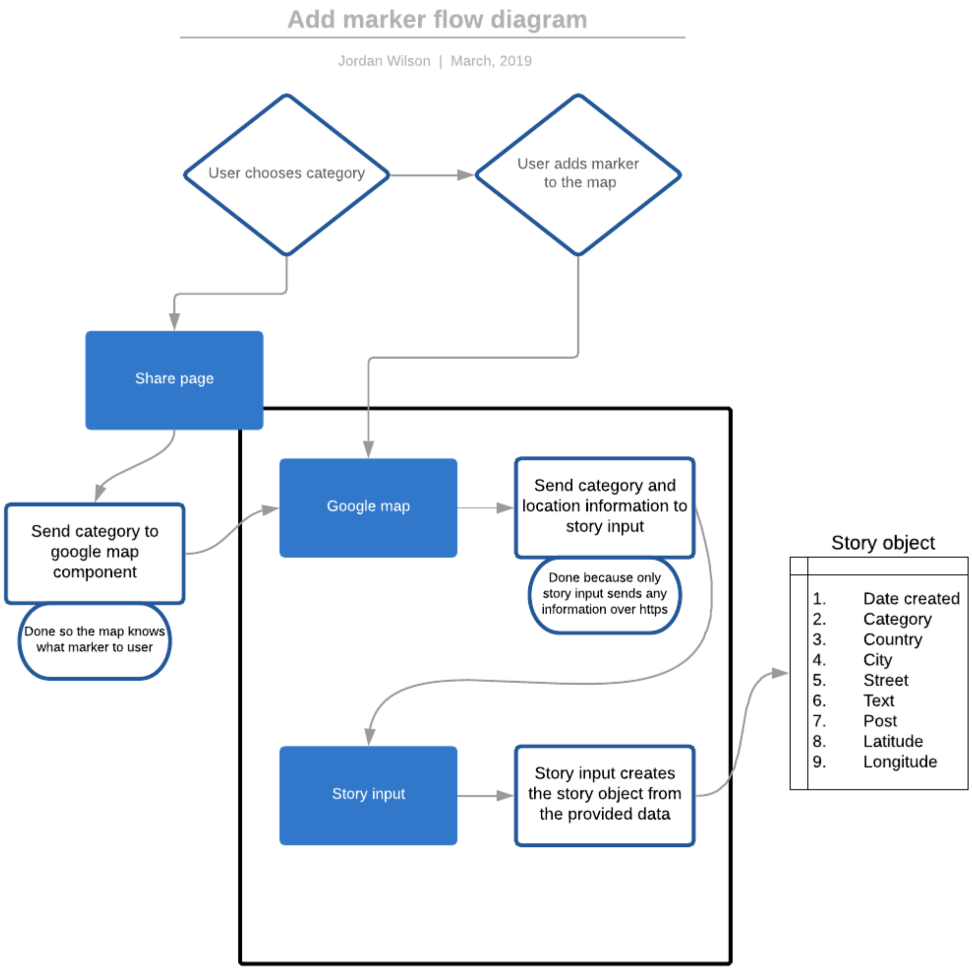


Figure 2.8. The interactions after a user chooses a category and adds a marker to the map.

1. The user chooses a category.
   1.1. The Share page selects the category from the available list and passes this information to the Google map component.
2. The user clicks the map to add the marker they have just selected.
   2.1. The map, using the coordinates from the clicked location, turns them into a street address and passes this information to the story input component.
   2.2. The story input component then creates the Story object, which is where all the information is kept. This is done in the story input component because only it does any transfer of data to the server.

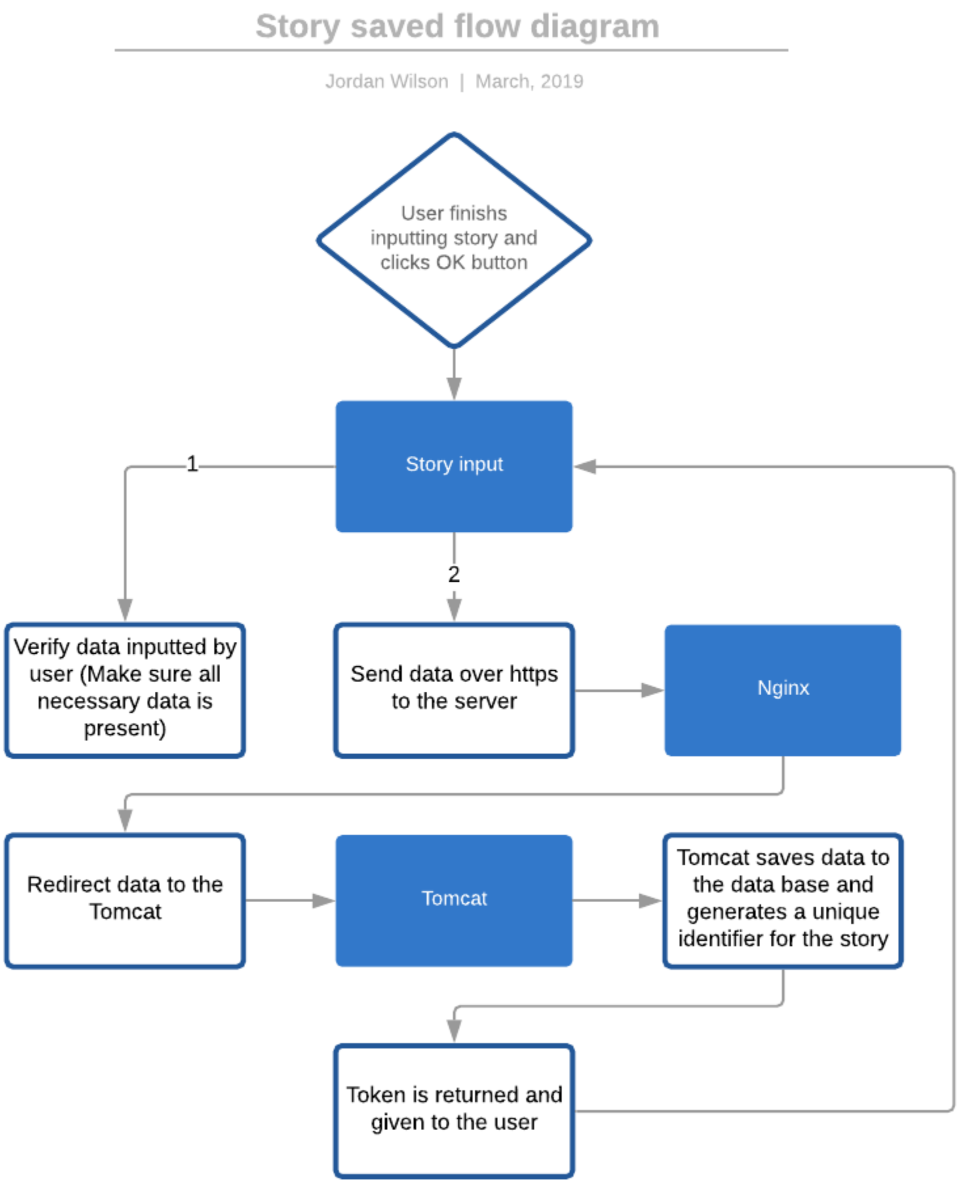


Figure 2.9. The interaction when a user has saved their story

1. The user finishes writing and saves the story.
    1.1. The story input component verifies that all required data is present.
    1.2. The data is then sent over a HTTPS connection to ensure that the data is safe.
    1.3. The Nginx reverse proxy then intercepts the request and redirects it to the backend running on the Tomcat server.
    1.4. The Tomcat server then saves the story object to the database and generates a unique identifier token.
    1.5. This unique identifier token which is used when a user wants to delete their story is then returned back to the webCrowd app.

2.7 Summary

WebCrowd consists of four pages with each page specifically designed to fulfil a purpose so that when a user navigates to a page they are visiting that particular page to fulfil said purpose. These pages are the result of multiple iterations of the app, taking elements that work from previous versions and finally generalising the end product to create a neutral platform. The main purpose of the app is to collect large scale data from communities using a map that allows users to add their data in the form of a story tied to a location. The data they enter is part of the category system designed by the researchers that has been a central topic of development. Finally, an overview of the technical aspects of webCrowd shows how the app uses different components to generate the data that we use for analysis.

3. Syntax Text Data Analytics

3.1 Introduction

This section introduces the design of data analytics and visualization system, and its application on urban design data. Our project uses Christchurch City Centre Common themes from Public Ideas (Christchurch City Council, 2011) as the data source. This document contains many community-based stories and captures people's practical needs toward urban infrastructures of city centre, where each story can be associated with one or multiple infrastructures.

However, due to the fact these stories are not well structured, it is quite difficult to manually categorize all the stories, sort out their connections with other stories and represent a top-down view of these stories' relations. This typically requires a tool to help urban designers to effectively gain the insights from these stories and efficiently convey ideas of urban plan to users, and thus leads to our main research question.

Can the public data be mapped into urban infrastructures and their complex relations be represented visually?

This question can be transformed to two parts: firstly identifying the category of the story, and secondly converting all the categorised stories into interactive charts, such as a Chord diagram. Our project aims to build a system to address this issue by performing a two-phased operation: text categorization using data analytics technique and data display using data visualisation.

The data analytics technique ("Defining Text Analytics | The Intelligent Enterprise Blog," 2009), or more specifically text mining, can derive useful information from plain or unstructured texts through the statistics patterns, such as usage frequency, to discover the new information and find the "**values**". The data visualisation technique (Friendly, n.d.) converts the numerical data into a graph to "**communicate information**" clearly and efficiently.

In this project, our data analytics is more like statistical text categorization, one of typical text mining tasks, to assign each story to one or more urban infrastructure categories according to the words used in the story. Our data visualisation represents these categorical data as several statistical and interactive graph using the web design. Our project combines these two techniques, but also includes a prototype module as the intermediate layer to sort out the categories of each story from raw data sources, and transform the data into the format required for visualization.

This chapter focuses on syntax text data analytics. It firstly gives an overview of our data analytics and visualisation architecture, and then discusses how our data analytics extracts the knowledge from plain texts to structured format and categorises the stories into six infrastructures using three components: Flax collocation library, glossary terms and story categorisation module.

3.2 Data Analytics and Visualisation Architecture

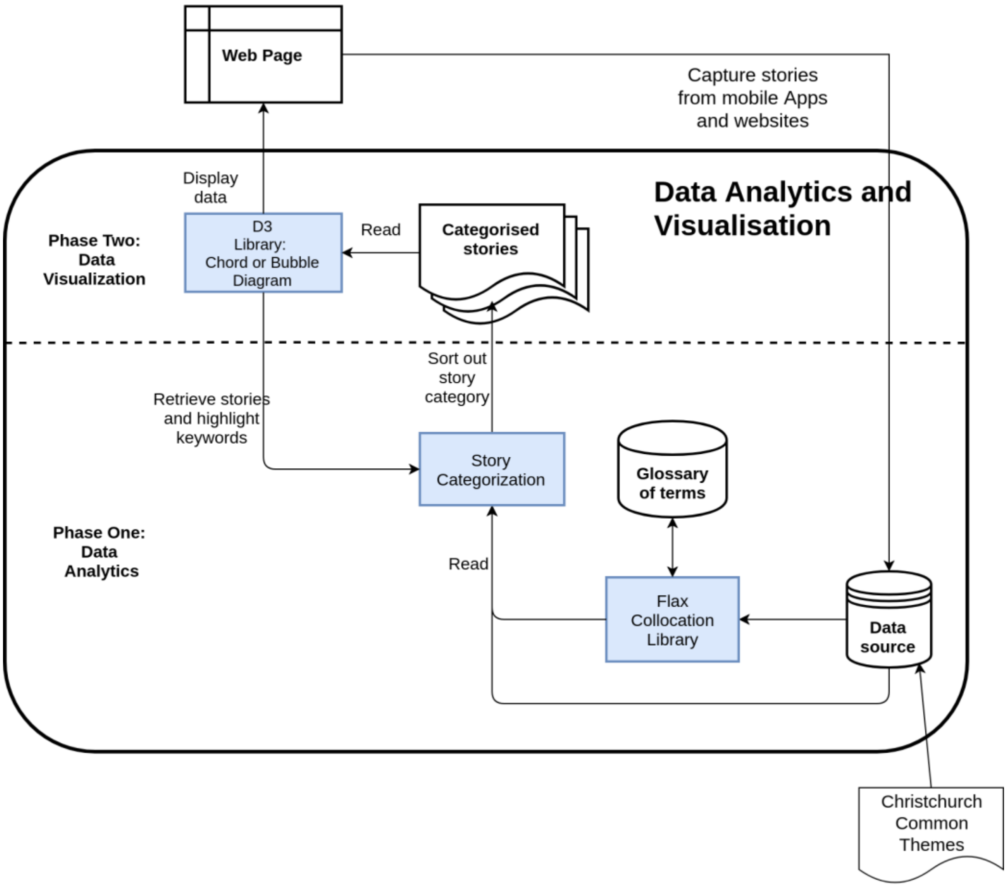


Figure 3.1 Data Analytics and Visualisation Architecture

Our system is composed of data analytics and visualisation interface. Our data analytics incorporates Flax (Flexible Language Acquisition), story categorization module and urban infrastructures (Grey, Dyer, & Gleeson, 2017) as back-end. Our visualisation interface is built on top of Javascript D3 (Data-Driven Documents) library as front-end to display the data.

The data analytics backend firstly uses Flax to pre-processes the data sources and collect the collocation words, and then the story categorization module is kicked in to classify the categories of stories in term of collocation word usage and produce the categorical data as output. Lastly, the visualisation interface reads in the output data from our backend, and encodes these data to several interactive diagrams. The procedure (see Figure 3.1) is the control flow graph of our system.

1. In phase one (data analytics), we use Flax library to read in the raw data source (Christchurch Common themes From Public Ideas) and extract the top 100 frequently used collocation terms, which each term is a combination of words appearing together in the text, such as light tram, and produces keyword terms for categorization.

   Then the story categorization module is kicked in to classify the stories based on the word usage. First, urban design experts are involved in the process and take the Flax collocation terms and link each term to one infrastructure category. Second, the module takes the glossary of categorical terms as input, and goes through each story and checks if any term of the glossary is used, and if so, associates the story with the category which that term is labelled to. After finalising the story categorisation, it outputs all the categorical stories to one single Excel file (CSV format) for D3 visualization, and lastly performs some statistical calculation, such as the proportion of stories in each category.

2. In phase two (data visualisation), we use D3 library along with story categorization module to create the chord diagram, word frequency charts and word bubble charts. As these charts are designed to be dynamic and interactive in the web browsers, we also define several user interaction behaviours to view the data in different aspects and provide better access to the data.

The above is a shorter description of our system. In the following sections, we will provide more details about Flax collocation library and story categorization module. D3 visualisation will be described in the next chapter.

3.3 Flax Collocation Generated Terms

Collocations (Witten, Wu, Li, & Whisler, n.d.) are two or more consecutive words that appear more frequently than random, for example, we would use "*shuttle bus*" in the sentence, rather than *bus shuttle*. In other word, the collocations are the high-frequent word combinations that people use in the sentences, and it also means that these word collocations can be the trending topics that people are interested in and talk often in the stories. We can make use of this property to analyse the data and find the keyword term for story categorization.

However, acquiring the collocations from a document is difficult (Nesselhauf, 2003) because there are a great number of collocation words in a language. Moreover, identifying these collocations in a free-formatted document is not straightforward as expected as it requires domain knowledge from language experts and has formal definitions and criterion as well as grammar rules. Therefore, it presents a great challenge, and we need a tool to help us automate this process and extract collocations from our unstructured data.

3.3.1 Flax Collocation Learning System

Flax collocation learning system (Wu & Witten, 2015) (FlaxCLS) can automatically use a set of syntactic patterns to extract the collocations from a document and establish a list of collocation words for retrieval. Flax collocation extraction heuristic procedure is as follows. It firstly parses a document and assigns part-of-speech tags to all the words using Apache OpenNLP (*Apache OpenNLP*, n.d.) (Open Natural Language Processing) library. Secondly, it matches these tagged words with a set of predefined patterns and identifies the collocation words. Lastly, it collects all the collocations, groups them by the matched patterns and sorts them by the frequency usage. FlaxCLS has been successfully used to search for the collocations from 3 trillion-word Wikipedia documents.

Part-of-speech tagging technique used in FlaxCLS is one of essential features for our project, because it can identify the collocation word as noun, verb, adjective or adverbs, etc, and label it with the corresponding word type. FlaxCLS part-of-speech tagging adopts Penn Treebank tagset at word level and assigns one POS tag to each word. For example, the collocation "*green spaces*" is tagged as "green/JJ spaces/NNs". The symbol 'JJ' represents the word 'green' as adjective, and 'NNs' indicates the word 'spaces' a plural common noun. These tagging information is helpful for our backend to recognize and highlight keywords in the stories, which will be discussed in next section.

In our project, we use Flax library to extract the top 100 most-appearing collocation words from our data source, and output the glossary of keyword terms, as the following xml. The glossary of keyword terms is composed of infrastructure terms and attribute words.

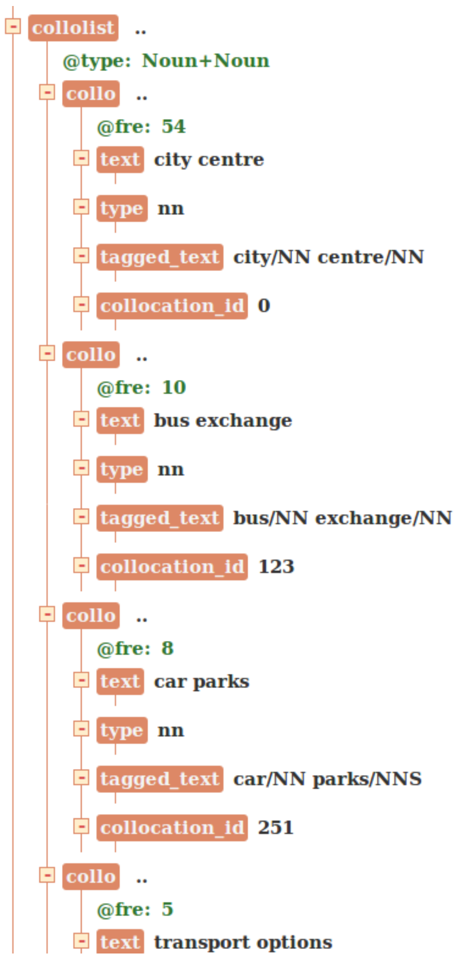


Figure 3.2(a) Infrastructure Terms

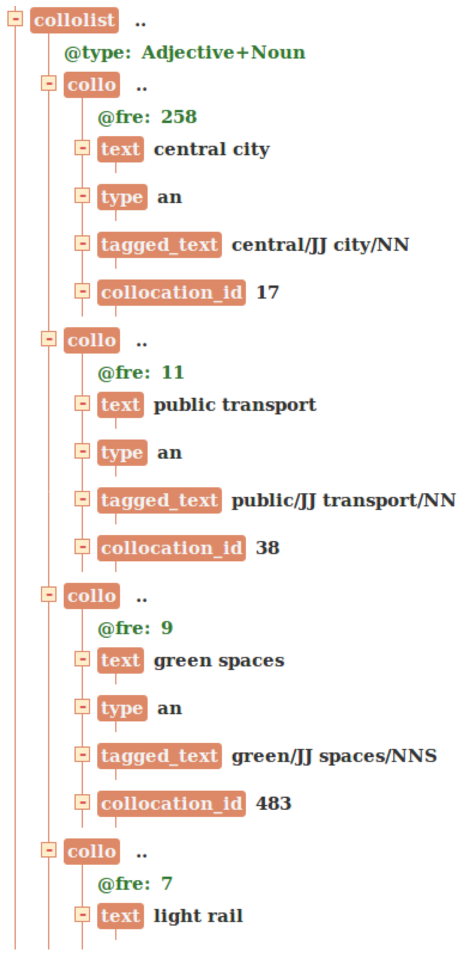


Figure 3.2(b) Attribute Words

3.3.2 Hard + Soft City Infrastructure Terms (Grey et al., 2017)

The infrastructure terms come from the Flax collocations of syntactic 'Noun + Noun' pattern. As shown in Figure 3.2(a), each infrastructure terms is encoded as a structured xml to show the words ('text' element), usage frequency ('fre' attribute value) and part-of-speech tag ('tagged_text' element). In this example, the term 'car parks' appears 8 times in all story contexts, and the word 'car' is tagged as a singular noun and ‘parks’ as a plural noun.

Every infrastructure term can be mapped into one of “hard” or “soft” infrastructure categories, and the categories for the hard infrastructure include “Utilities”, “Public Space” and “Building”, and for the soft infrastructure including “Institutional”, “Community” and “Personal”. For example, “bus exchange” and “car parks” are associated with “Public Space” infrastructure and “transport options” is assigned to “Utilities”.

3.3.3 Desired Attributes of City Infrastructure

The attribute words are the bridging terms used to establish the connection of infrastructure categories. These words come from the Flax collocation of the form of ‘Adjective + Noun’, ‘Adjective + to + Verb’, ‘Adjective+Preposition+Noun’. Similar to infrastructure terms, attribute words are also encoded as a xml document where the adjective is denoted as “JJ”. Consider the attribute words “light rail” in Figure 3.2(b). The word “rail” belongs to the Utilities category, and we take the adjective “green” as a link to the Utilities category.

3.4 Glossary Terms and Story Categorization

Once we collect infrastructure terms and attribute words, we can start story categorization module. Story categorization module has three parts: building the glossary of terms, labelling the stories and performing the statistics. The procedure is as follows. The module firstly classifies all the stories according to the usage of glossary terms and labels each story with one or more relevant infrastructure categories. Upon the completion of story categorization, it counts the number of stories in each of category and the number of stories between every two categories, to produce the relation matrix, which the matrix is used to represent the amount of stories associated with two categories. Lastly, this module goes through each story and searches for the key collocates from our glossary of terms and highlight them for display.

3.4.1 Glossary of Terms

This term mapping requires the domain knowledge of urban design, and relies on the professionals to classify each term into the infrastructure category. Once Flax library extracts all the collocate words of ‘Noun + Noun’ and ‘Adjective + Noun’ patterns, we list these terms and invite well-trained urban designers to go through each term, read the story that the term refers to, and assign the category of each term based on their expertise and judgement.

Category	Word	Frequency
Building	rise buildings	5
Building	sex shops	4
Building	city living	4
Building	ground floor	4
Building	ground level	2
Public Space	central city	258
Public Space	city centre	54
Public Space	car parks	8
Public Space	peoples district	2
Utilities	public transport	11
Utilities	bus exchange	10
Utilities	transport options	5
Utilities	shuttle buses	5
Utilities	car parking	5
Utilities	speed limits	3

Figure 3.3 A screenshot of glossary of hard infrastructure terms

Figure 3.3 shows parts of hard infrastructure terms sorted by the usage, and these terms are assigned with one category and will be used to identify the category each story belongs to. For example, “shuttle bus” is the characteristic term of “Utilities”, and any story that mentions “shuttle bus” will be categorised to “Utilities”.

Category	Word	Frequency
Community	police presence	3
Community	security wardens	2
Community	sex businesses	2
Community	night workers home	2
Community	street vendors	2
Community	support cbd business	2
Institutional	government departments	3
Institutional	Increased police presence	2
Institutional	visible policing	2
Institutional	have police	2
Personal	young people	10
Personal	older people	9

Figure 3.4 Snippet of glossary of soft Infrastructure Terms

We find this approach can well recognize the hard infrastructure, but also notice that it needs some improvement for identifying soft infrastructure because these soft terms may have a number of variations. As shown in Figure 3.4, “young people” belongs to “Personal” soft category so the “Personal” category should comprise all its synonyms, including “teenagers”, “adolescent”, “youth”, “young person”, “teen”, etc. However, finding all the synonyms for every soft term by-hand is time-consuming and error-proning, and this requires further tools to automate the process and reduce the error rates.

3.4.2 Story Labelling

The story labelling phase chooses and assigns the infrastructure categories to all the stories, based on their usage of glossary terms. As each story may be linked to more than one category, labelling these stories becomes a typical question of multi-label text classification. Due to the time constraints and technical difficulties, we decided to transform this multi-label classification to multiple single-label classification questions, and solve each question using text categorization technique. The single-label story classification questions include

1. Does this story use any glossary of term from *Utilities* infrastructure?
2. Does this story use any glossary of term from *Public Space* infrastructure?
3. Does this story use any glossary of term from *Building* infrastructure?
4. Does this story use any glossary of term from *Institutional* infrastructure?
5. Does this story use any glossary of term from *Community* infrastructure?
6. Does this story use any glossary of term from *Personal* infrastructure?

We use text categorization technique along with the glossary of terms from FlaxCLS system to perform these single-label tasks and check if a story is associated with any of the infrastructure categories.

Category	Story	Keywords
|Public Space|Institutional	Mix of views on university – some say bring it back into the city centre others say leave it at Ilam.	|Public Space:-city centre-city centre||Institutional:-university|
|Public Space|Building|Institutional|Community	"Some suggest bringing components back into city centre (e.g. Music school, artistic displays, law school, study space, accommodation) ""We really have to be positive about bringing the whole university back into town."	|Public Space:-city centre-city centre||Building:-study space||Institutional:-university-music school-artistic displays-law school||Community:-town|

Figure 3.5 Snippet of categorized stories CSV file

Figure 3.3 shows a list of our categorized stories. Each row contains one story, and its associated categories ('Category' column) and the keywords used for category identification ('Keywords' column). Our project categorises the stories based on the keyword usage, and transforms the multi-label to single-label classification. By doing so, the complexity of our project can be greatly reduced and our prototype can be built up within given time limit. However, we admit that our story categorisation may not be as precise as other approaches but these results can be baseline for evaluating other categorisation techniques in the future work. For example, MEKA (Read, Reutemann, Pfahringer, & Holmes, 2016) (Weka based multi-label learning toolkit) can build up the machine learning model and train the model to predict and annotate one or more categories for each story automatically.

3.4.3 Statistical Matrix Calculation

D3 Library contains several open access software solutions for data visualisation. For example, D3 chord diagrams (Bostock, n.d.) are one of solutions that illustrates the relationships between different parameters or categories of data.

In our test case, we have 6 different infrastructure categories, and each category is associated with a number of different stories, and each of these stories may have links to other categories. To visualise the interrelation among these categories, we need to create a chord diagram, and represent the linking stories as a set of chords where each chord connects from one source category with other target category. As such, our D3 chord diagram requires a 6x6 matrix to denote the directed flow amongst a complete graph of 6 categories. Once we completed the story categorization, we go through the output CSV file and categories, and count how many stories are linked between two categories and produce the relation matrix.

The relation $matrix$ is denoted as an array of 6 items, and each $matrix[i]$ is also an array of 6 items, and $matrix[i][j]$represents the number of stories from ith category to jth category.

```
[
  [0,5,1,1,3,2], // Utilities
  [5,0,2,3,5,2], // Public Spaces
  [1,2,0,1,2,1], // Building
  [1,3,1,0,1,1], // Institutional
  [3,5,2,1,0,2], // Community
  [2,2,1,1,2,0]  // Personal
]
```

The matrix row and column order are the same: Utilities, Public Spaces, Building, Institutional, Community and Personal. For example, the 1st row $matrix[0]$ represents all the links from Utilities category to 6 categories. $matrix[0][1]$is the link from Utilities to Public Spaces, and $matrix[0][5]$is the link from Utilities to Personal. $matrix[0][0]$ is zero because no link from/to Utilities itself. And this matrix is bi-directional, i.e. $matrix[i][j] = matrix[j][i]$.

3.5 Summary

Our syntax text data analytics combines the computerised tool and human domain knowledge to identify key infrastructure terms and perform story categorisation. It firstly uses the FLAX natural language processing technique to extract key terms from the unstructured and plain-text stories. Secondly, urban design professionals are involved to assign each key term to either the hard or soft city infrastructure category to build up the glossary of terms. Lastly, our analytics takes the glossary of terms and data stories as input, maps each story to one or more categories with respect to its word usage, and outputs the required data for producing the data visualisation (discussed in the next chapter). Our preliminary results show that, for a small-sized data source, our data analytics can provide sufficient information to generate useful and exploratory data visualisation.

4. Data Visualisation and Infographics

4.1 Introduction

Presenting datasets visually utilises the human brain's natural ability to detect patterns and process images quickly. The data visualisation helps explain the complicated scientific data much easily, and represent abstract topics with tangible figures (e.g. maps, star charts, diagrams) as a mean to increase the engagement of the public. With the constant and limitless availability of information in the modern era, many content creators began to focus on aesthetic appeal as a means to make their content stand out, relying on the application of graphic design principles. This movement towards aesthetic appeal and away from hard data presentation formed the area of information graphics ('infographics').

This chapter first gives the definitions of infographics and data visualisation, finds their common and distinct traits, and recognizes the convergence of these two terms. Then we present our data visualisation software to encode stories from common themes data source to visual forms, and display the interrelationship of these story categories with chord diagram.

4.2 Definitions and Spectrum of Data Visualisation and Infographic

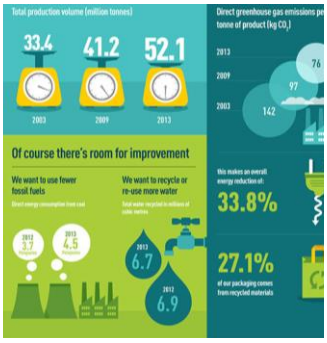
Presumably due to the relatively new emergence of the field of information visualisation, no standardised terminology has been established. As such, the distinction and relationship between the terms 'data visualisation' and 'information graphics' ('infographics') is the subject of dispute. Despite differing opinions on the relationship between the terms 'data visualisation' and 'infographic', the criteria that define each term are shared across all theories.

For the purposes of clarity and explanation of existing theories, term 'data visualisation' will be used in this document to denote traditional chart/graph-like representations of complete datasets. The term 'infographic' will be used to denote data narratives that contain visual representations of data that accompany textual information and are arranged in a narrative fashion with an introduction, elaboration, and conclusion.

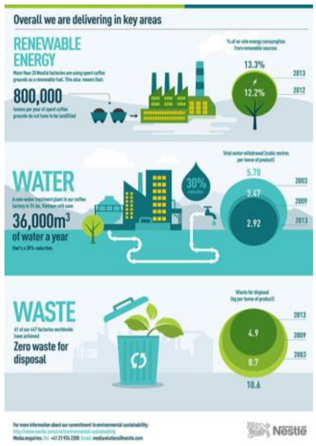
4.2.1 Definition of 'Infographic'

The term 'infographic' is a word blend from the phrase 'information graphic', referring to the style's predominant use of graphic elements (typography, imagery, iconography, etc.). Infographics illustrate a specific topic by providing enough information that the reader is able to understand the subject, the intention, and the perspective.

(a)



(b)



(c)

Figure 4.1 Infographics Examples

Infographics are created with a topic, a purpose, and an intended audience making each one unique. To do so effectively requires that each be created manually and emphasis must be placed on the formation of a coherent 'data story' so that communication of the intended perspective is successful. The intended audience must also be considered during the design and development process so that the final product can be tailored to interest the desired demographic. Such customisation and structural focus makes this development process longer and more demanding than the creation of simple charts and graphs.

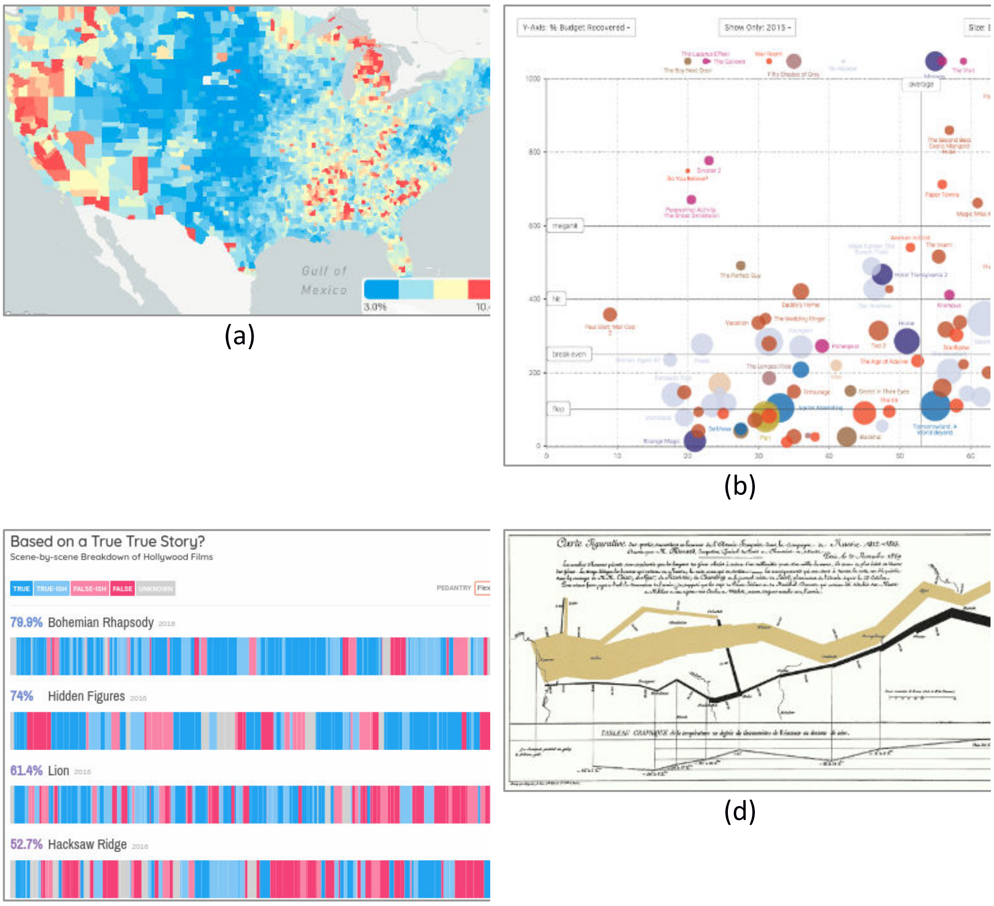
Figure 4.1 shows three example graphs using infographics technique. We notice that

1. Each infographic graph ("Infographic," 2019) has more textual content than other data visualisation styles, so that they are easy-to-read and convey their stories clearly. By posting these graphs on Facebook or Tweeter, we can spread the information to as many people as possible whilst avoiding distorting the stories that these graphics are intended to say.
2. Each infographic is designed to express a specific opinion or message - the "storyline". In our example, each graph talks only one topic. Graph (a) shows the environmental sustainability policy of a company, and Graph (c) indicate that the company has zero-waste policy and uses the renewable energy to reduce the carbon footprint.
3. Infographics should only include data that is relevant to the storyline. Graph (b) includes the statistical data to shows the effectiveness of the company's sustainability, such as the reduction of greenhouse gas emission.

4.2.2 Definition of 'Data Visualisation'

Data visualisation is the process of presenting complex data visually, rather than using textual methods. Andy Kirk, author of the book Data Visualisation: A handbook for data driven design, adds that "the representation and presentation of data to facilitate understanding" (Kirk, 2016) describes both a process and a product. The process he describes involves using computer programs to transform raw datasets into visual representations which are then presented in visual contexts (i.e. the product).

The transformation from data to product is usually performed by applying a general algorithm to a dataset. Using general algorithms to transform the data makes the process more adept for large datasets and also allows the model to be reused with other datasets. The datasets can then be displayed in a single view so that global patterns and areas of interest can be identified quickly. Investigation of internal patterns is also possible by extracting specific factors of the data and displaying it on its own. By combining multiple visualisations on a 'visualisation dashboard', a detailed view of the dataset is available. Adding interactive features to the visualisation enables viewers to explore the data and formulate unique perspectives and enquiries.



(a) (b) (c) (d)

Figure 4.2 Data Visualisation Examples

We notice there are some common characteristics in data visualisation.

1. The visualisation chart utilises the human brain's aptitude for processing visual information quicker than text. For example, Figure 4.2 (a) encodes a large number of geographical and numerical data into a map and distinguishes different levels by using colours. This helps to support the further data analysis and find the patterns.
2. The visualisation chart enables large datasets to be displayed in one view, and can be used interactively to explore the data. As shown in Figure 4.2 (b) and (d), users can easily identify the outliers in the graph and mouse over those data points to show the detailed statistical numbers, and this information helps find out the cause of the errors.
3. Each visualisation ("Data visualization," 2019) focuses on one aspect of the data, and encourages to compare different pieces of data. Figure 4.3 (c) illustrates the authenticity of a movie, i.e. how close it is related to the true

story, and lists the truth percentage of each movie for comparison. However, when comparing the data with multiple factors, it is agreed that we need other type of ‘data visualisations’.

4.2.3 Information Visualisation Spectrum

The first theory posits that the terms ‘data visualisation’ and ‘infographic’ can be used interchangeably. Instances of this stance do not include direct statements of the belief that the terms are synonymous; they simply alternate the two words throughout publications at will.

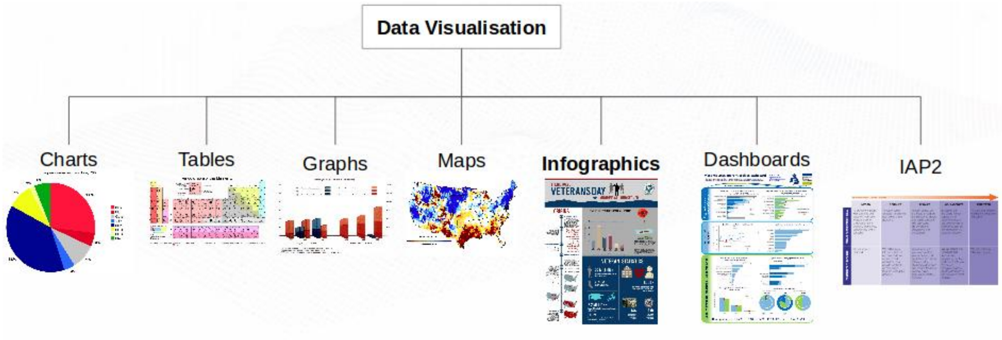
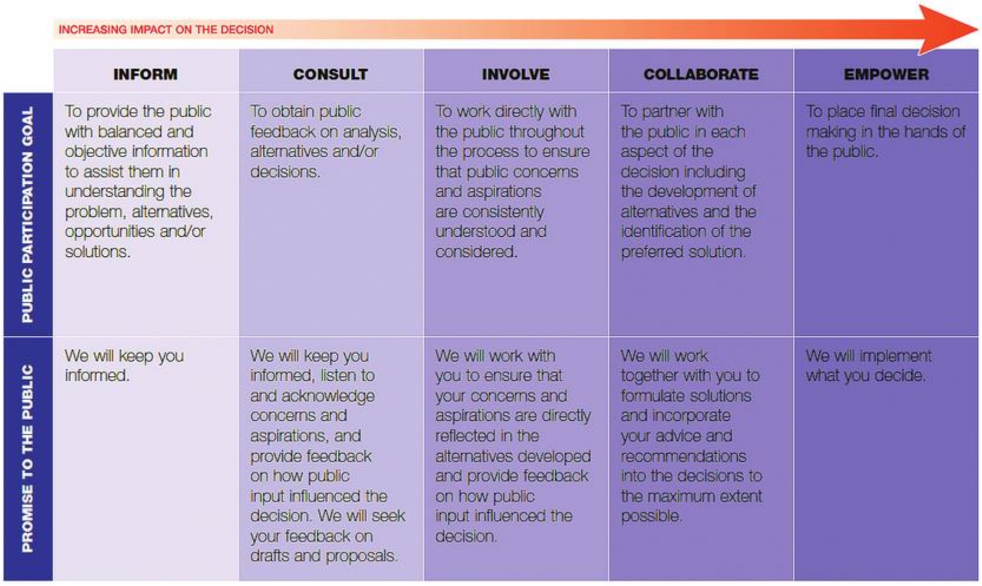


Figure 4.3. A spectrum of data visualisation and infographics Stakeholder participation: IAP2 public participation spectrum (“Stakeholder participation,” 2017)

It is mostly agreed that the definition of ‘data visualisation’ is the transformation of data into visual form. But it is still a dispute that infographics are a style of ‘data visualisation’ that is evolved to address a wider audience. It is often argued (Krum, 2013) that infographics often contain ‘data visualisations’ while ‘data visualisations’ cannot contain infographics. However, some visualisation research (Ryan, 2016) suggests that ‘data visualisation’ is an umbrella term that encompasses the ‘infographic’ style, and includes a variation of graph forms. The information spectrum in Figure 4.3 illustrates three important imperatives of data visualization. Through charts, tables, graphs and dashboard, the visualisation enables users to explore big data, analyse the data and manage the goals. The visualisation engages users with data interaction and improve user experience, e.g. the use of map help explores a large number of geographic data. And the visualisation with infographic style includes explicit contexts on the graph to communicate the big and abstract data effectively and correctly.

This capacity of data visualisations and infographics to communicate large, abstract, data effectively to citizens and decision-makers make them critical tools in enabling more meaningful public participation processes as set out in Fig. 4.4. The ability to share knowledge across multiple scales is an important step towards achieving a more collaborative approach to urban design and planning, bridging that gap between top-down and bottom-up planning processes. Also, the ability to engage people in data interaction is an important step towards improving data literacy among citizens and building community capacity. As set out in Chapter 5, improving the quality, frequency and availability of data on the environment is a key factor in facilitating a more dynamic decision making framework. But, this is not possible without the use of the data visualization techniques outlined above..



INCREASING IMPACT ON THE DECISION

	INFORM	CONSULT	INVOLVE	COLLABORATE	EMPOWER
PUBLIC PARTICIPATION GOAL	To provide the public with balanced and objective information to assist them in understanding the problem, alternatives, opportunities and/or solutions.	To obtain public feedback on analysis, alternatives and/or decisions.	To work directly with the public throughout the process to ensure that public concerns and aspirations are consistently understood and considered.	To partner with the public in each aspect of the decision including the development of alternatives and the identification of the preferred solution.	To place final decision making in the hands of the public.
PROMISE TO THE PUBLIC	We will keep you informed.	We will keep you informed, listen to and acknowledge concerns and aspirations, and provide feedback on how public input influenced the decision. We will seek your feedback on drafts and proposals.	We will work with you to ensure that your concerns and aspirations are directly reflected in the alternatives developed and provide feedback on how public input influenced the decision.	We will work together with you to formulate solutions and incorporate your advice and recommendations into the decisions to the maximum extent possible.	We will implement what you decide.

Figure. 4.4 Stakeholder participation: IAP2 public participation spectrum (“Stakeholder participation,” 2017)

4.3 General Characteristics and Comparison of Data Visualisation and Infographics

People with less understanding of visual information tend to have a mix-up use of visualisation and infographics terms as both of terms are very closely similar on the definitions and refer to the tools for representing data and conveying information. Some visualisation researchers (Iliinsky & Steele, 2011) adds that terms ‘data visualisation’ and ‘infographic’ simply refer to the tone of the object. For example, “Some people use infographic to refer to representations of information perceived as casual, funny, or frivolous, and visualisation to refer to designs perceived to be more serious, rigorous, or academic.” This quote displays the belief that the difference between the two terms is merely the tone that the image expresses and also the intended audience.

The terms ‘data visualisation’ and ‘infographics’ are mutually exclusive but share some traits. This relationship between the two terms shows that they are separate entities that often share the characteristics. The blog (Hagley, n.d.) tried to compare and contrast these two terms from the perspective of a graphic designers. They concluded that the common traits of data visualisation and infographics are that both can convert data into visual representation and provide informative contents.

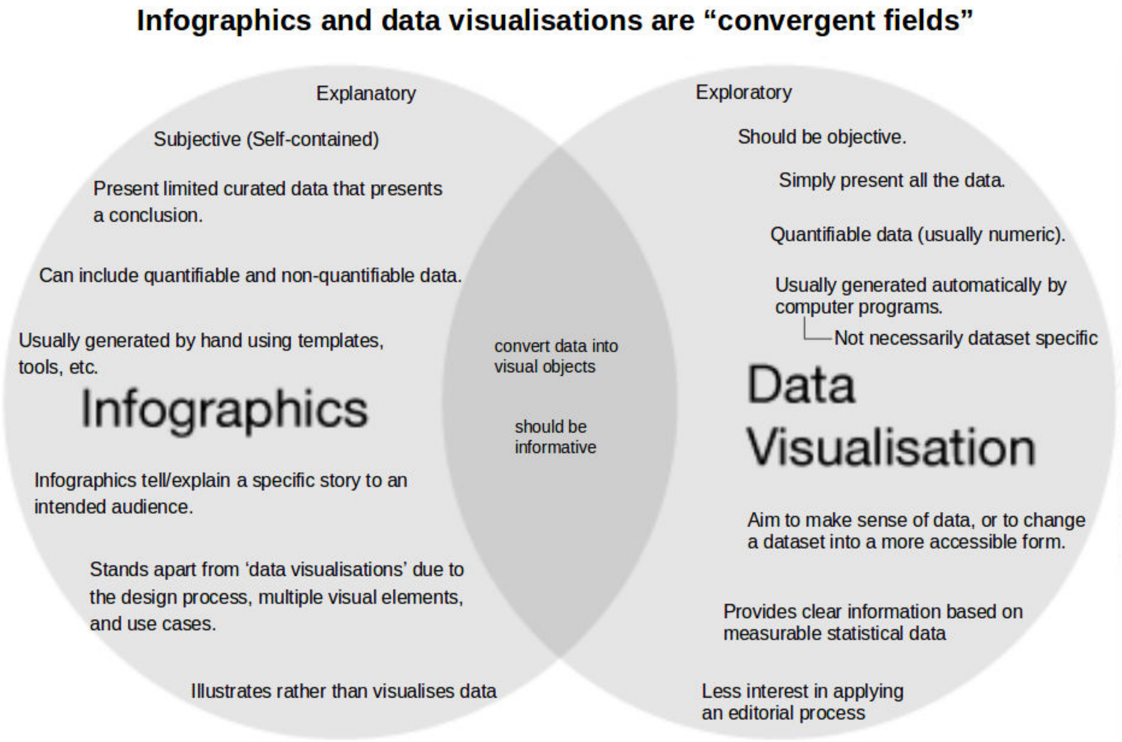


Figure 4.4 Convergence of Data Visualisation and Infographics

Figure 4.4 shows the distinction between infographic and data visualisation. The infographics uses the curated data set, and the graph is usually generated by hand so it is often considered subjective. On the other side, the data visualisation is considered objective as it tends to present the entire data set and encode the numeric data to the graph using the algorithms and programs automatically without human intervention. The infographics also presents the information with some texts to explain the contexts itself whilst the data visualization provides users with a better access to the raw data for exploratory analysis.

These two terms target on different kinds of audience. Infographics may include quantification and non-quantifiable data to tell a specific story to the intended people whilst the data visualisation provide information based on measurable and statistical data for general use. The major difference is that the infographics brings in the selected data and requires further graphic design to facilitate the illustration whereas the data visualisation represents the data with enabled user interaction and less editing efforts for better exploration. Lastly, the graphic designer (Hagley, n.d.) described data visualisation and infographics as “convergent fields”, in that they both seek to make sense of data but with differing methods, styles, and intentions. And they share the same goal: transition data to the information.

4.4 D3 Data Visualisation Software

Our project includes public data from city council and real-time stories from mobile apps. These data are plain texts and there are many of them, so it is not only time-consuming but also difficult to explain these stories and make them understandable to people, and most importantly, find the valuable insights. We need a helper tool to provide a better data access for users to explore these stories, find the trends and make the data-driven decisions.

Visualisation (Friendly, n.d.) is one of essential techniques used to display massive quantities of data by using graphical representations, e.g. maps, charts and diagrams and graphs. As people tend to comprehend the graphs faster than texts and quickly identify the trends from a chart or diagram, the use of data visualisation can make the story telling of large-scaled textual data much easy and efficient, and assist us to highlight the key points and extract useful information.

In this project, we include two kinds of data sources: top-down stories (public data) and bottom-up and real-time stories from mobile apps. Because these two sources provide different aspects of views and each has a large amount of data, exploring these stories requires the visualisation tool to be *flexible*, *easy-to-adopt* and *scalable*.

4.4.1 Web Based Visualisation Tools

There are a number of web-based and off-of-shelf visualization tools. We select three well-known tools: D3.js, Google Chart and vis.js for comparison, and choose D3 as our main development tool as it meets our project requirements. D3 is a free and open-source Javascript source library and provides a number of visualisation and interaction techniques to create the diagrams from quantitative or categorical data, and it enables users to have the access to interact with the data in the web browser, such as Firefox (preferred) or Chrome.

1. *Easy-to-adopt* dynamic model eases the learning curve of using tools. D3.js uses a similar dynamic document manipulation as JQuery library to meet interaction needs for graphical display. For example, we can define the mouse-over event on a D3 SVG graph to retrieve more data. This is the same event type as JQuery Document Object Model (DOM), so anyone who is familiar with the web application development can easily pick up the ideas and quickly write the D3 application. Google Chart APIs also allows similar event customisations. vis.js allows users to define the function triggered by interaction such as clicking or double clicking, but its use is not as intuitive as DOM and requires further learning efforts.
2. Massive data processing capability improves the *scalability* of visualisation tools. These three tools make a statement that they can process large amounts of data, and have many successful use cases. D3.js is used by 10 or more companies to explore a variety of application data, e.g. cloud data on Github. Google Chart APIs has a certain level of scalability as it is developed by a giant company. vis.js has several showcases, but has not explicitly shown the capability.
3. *Flexible* support accommodates the freedom of design and development, and adopts the changes constantly and rapidly during the design process. Google Chart APIs has a gallery of ready-to-use charts to help designers to convert their data to a diagram with minimal coding efforts, and it also provides the customisation to create a new chart type to fit into the data. D3.js and vis.js both provide Javascript APIs to create a chart and change its styles of charts, so the programming skill is required to use these two libraries. However, D3.js includes a set of ready-made online examples (D3 Gallery) to explore D3 charts and change the example chart by modifying the parameters only without writing the entire program. This could help

designers quickly build up the prototype chart using D3 library without extra coding.

In summary, Google Chart APIs and D3.js are a good candidate for our project because they are easy to learn for the web developers. They both can process large quantities of data, and provide a number of example charts for rapid prototyping. By comparing the chart types of these two tools, we find that D3.js has greater number of chart types relevant to **textual data visualisation**, e.g. chord charts and collapsible charts, than Google Chart APIs, and thus D3.js becomes our primary tool-kit. In this project, we use D3 library to create simple frequency chart, word bubble chart and relationship chord chart, etc., and aim to build up a dashboard as a part of future work to view all available visualisation graphs, and find the values of the trend.

4.4.2 D3 Charts

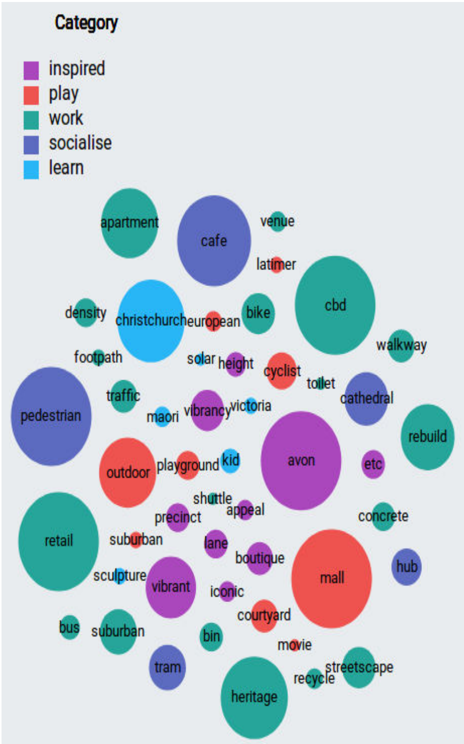
Identifying the key-phrases and terms in the texts for categorization is one of important tasks in natural language processing. For example, we can use sentiment analysis (Jurafsky, Chahuneau, Routledge, & Smith, 2014) to process the online restaurant reviews and frame the positive/negative opinions to learn customer preferences, so that we can provide insightful customer feedback to improve the restaurant's rating. These analysis results are usually plain texts and numbers, and require further encoding to graphical representations for people to interpret and understand.

A wide range of tools can be used to highlight these keywords and interact with data, e.g. we can use Microsoft Excel to create word bubbles or word clouds to display the frequency of words. But Excel does not provide the interactivity as web page, and has its limitations, e.g. displaying data across different categories (we need to create all the graphs). So in this project, we use D3.js, rather than Excel, as we use categorical data and need to retrieve and compare the stories between two categories.

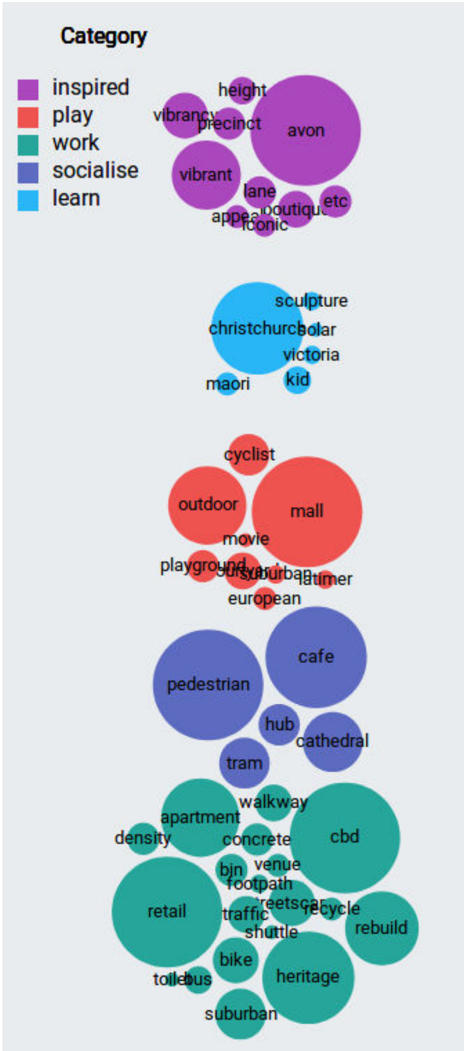
Scattertext (Kessler, 2017) is one of visualisation tools that enables the interaction among categories. It can display top ranked word lists in two dimensional graph and compare the word frequency in two categories. We could incorporate its ideas and create our own text charts as a part of future work. In this section, we will give more details of our visualisation graphs using the D3.js tool.

Word Bubble Chart

We use D3.js to create a bubble chart to display the high frequent word list from Flax collection. Each bubble represents one specific word and its size is determined by the word frequency in the collection and its bubble color is associated with the category of the word. By using word bubbles, we can find the most trending words in each category of our data collection.



(a) No group



(b) Group by category

Figure 4.5 Word bubble charts of high-frequent words

We also provide three kinds of grouping: all, category and experimental similarity. The group-by-all order does not specify the order of words but brings the bubbles to the center of the graph. As shown in Figure 4.1, “pedestrian”, “avon”, “cafe” are the most frequent words based on their bubble sizes. In Figure 4.2, we group word bubbles with their categories, and split the layout into five parts. And we can obtain the trending words in each category, e.g. “mall” and “outdoor” are the highly associated words in “play” category. These frequency information can indicate the popularity of a topic. High frequent words can highlight the trending topics in a category, whereas low-usage words indicate that people mention that topic less, e.g. the “outdoor” has higher frequency than “movie” in play category.

We include an experimental grouping to order the words by their similarity as a future work, because this feature requires a further definition of word similarity, e.g. word embedding or edit distance of words. Furthermore, a variation of bubble charts on two dimensional graph (x axis is the word category and y axis is the similarity), like Chart Studio (“Make a Bubble Chart Online with Chart Studio and Excel,” n.d.), can be used to show the similar/in-similar words in each category.

Chord Relationship Diagram for Christchurch City Centre: Common Themes

Chord diagram (Holten, 2006) is a common graph used to display the complicated inter-relationships between hierarchical data, e.g. the stories shared by two categories in our case. It has a radial layout and uses a line segment on the circle to represent the stories of each category, and includes one edge from one segment to another to show the relation between two categories. In this project, we use D3.js to produce the chord chart and also use MarkJS ("mark.js – JavaScript keyword highlight," n.d.) to highlight the stories with hard/soft infrastructure terms and attribute words.

The visualisation page, which holds all of the data visualisations, was conceived rather late in the development cycle of the app. Although the idea of having visualisations on the app was present from the early stage of the app development, it was not originally separated out into its own page and was incorporated into the main page. The change was made, however, as it became apparent that visualisations were a very important part of the process and were in fact more in-depth than originally thought. Therefore, they were made a page of their own, which consists of tabs that divide the different visualisations by type, with each visualisation showing a different aspect of the data. These visualisations could be expanded upon and exchanged over time to keep users interested.

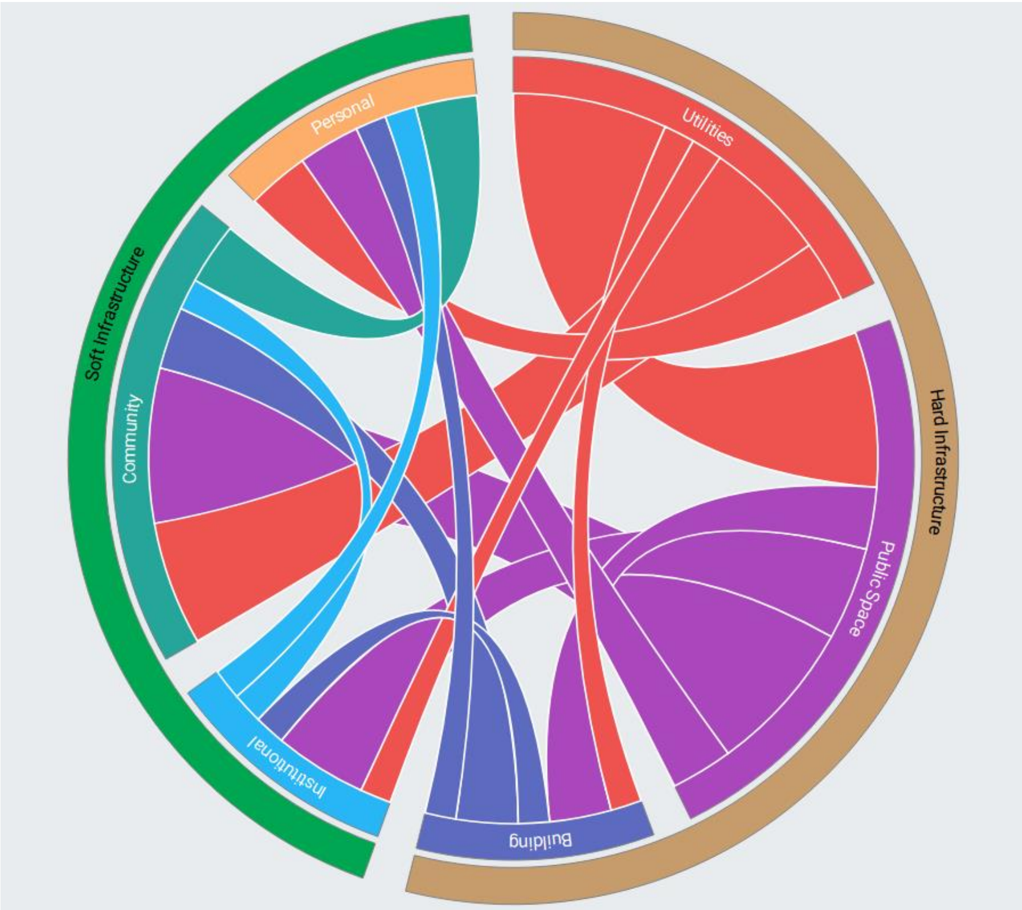


Figure 4.6 Screenshot of chord diagram

We developed a D3 front-end to utilise D3 library and perform two main tasks: generate the chord chart, and enable user interaction. D3 front-end reads in all the outputs of story categorization module along with the glossary of terms from Flax to produce the chord diagram and display the chart on the left side and the associated stories on the right. Figure 9 shows an overview of chord diagram that comprises 10 data stories. These stories are arranged in the categories, which each is represented by one line segment on the circles, and the inter-relations of data stories are drawn as the pairing edges between two categories.

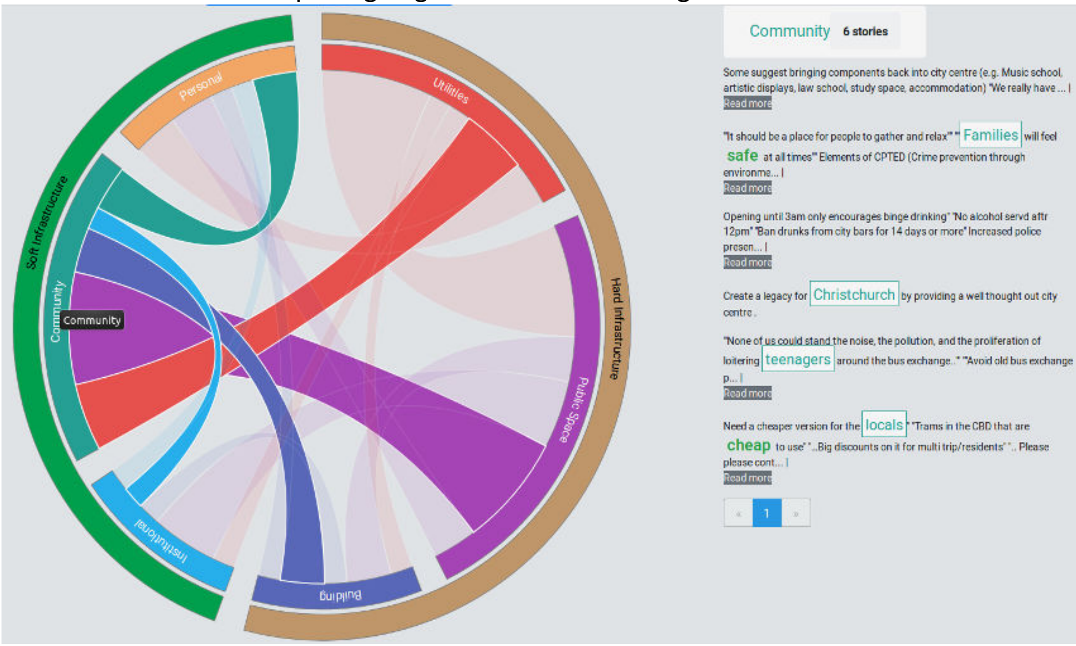


Figure 4.7 Screenshot of the data stories from "Community" Category

D3 front-end also enables user iteration events in the chord chart to retrieve the stories of a specific category, or the stories between two categories and then highlight the keywords in each story. We will use the example in Figure 4.7 to illustrate the user interaction of D3 front-end. When user moves the mouse over the soft infrastructure "Community" category, it will trigger D3 front-end to obtain all the stories linked to community category, and then go through each of these retrieved stories and highlight the Utilities infrastructure terms and attribute words using regular expression technique and MarkJS Javascript library. As the number of stories may be too large to fit into a single page, we also include the pagination on the end of the story list, and show a short version for each story (within 150 characters) and use the "read more" button to display the full content of a long story.

When the user move the mouse over the edge, the front-end will change the story list and make the associated stories bold. When the user decide to investigate the interrelationship and double click on the edge, the front-end will filter out all the irrelevant edges and keep the selected edge and also freezes the story list. The locked status can be restored by clicking on the 'lock' icon on top of story list.

4.5 Summary

Data visualisation transforms the entire numeric data into interactive graphs for exploratory study whilst infographics curates quantitative and qualitative data to graphs along with some texts for self-explanatory purpose. These two techniques however become convergent and mixed at recent researches. We have developed a D3 web-based visualisation software to display the data relationships using a chord diagram and highlight basic elements of data stories. Our chord diagram can

1. Illustrate the relation of hard + soft city infrastructures with desired attributes.

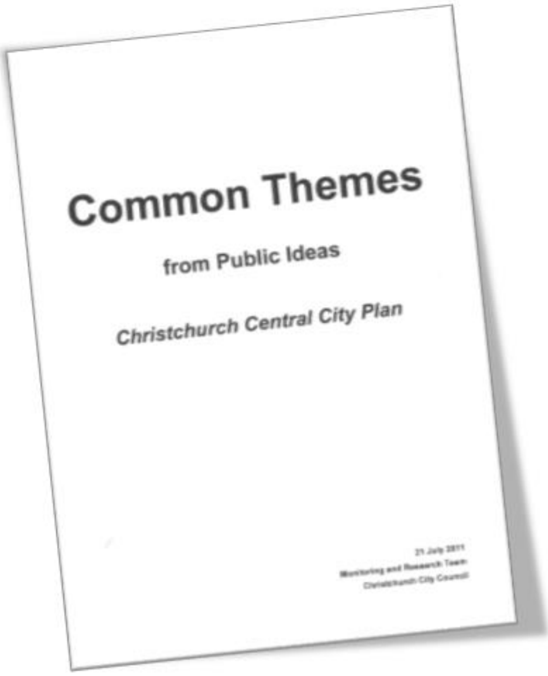
2. Highlight a specific aspect of hard infrastructures (Utilities, Public Spaces and Building) that different stakeholder (community, institutional and personal) group finds important.

The next challenge is to represent the persona and for storytelling. We may render the curated data stories of pairing soft and hard infrastructures into persona and then construct the good and coherent scenarios. These persona-scenarios (Madsen & Nielsen, 2010) can help the storyteller to create reasonable and new urban design ideas, and to explain these ideas to the intended audience by using the data visualisation with infographic style.

5. Data Storytelling for Christchurch City Centre: Common Themes

5.1 Introduction

The data collection, analysis and visualisation techniques outlined in the preceding chapters have been applied to the case study of Christchurch using data from a public consultation undertaken by Christchurch City Council following the earthquake in February 2011. The public consultation comprises a total of 105, 991 submissions from a variety of sources including workshops, surveys, questionnaires, graffiti, post-it-notes, videos, voicemail messages, radio, letters and emails. Although it has not been possible to obtain a transcript of all the submissions, the Council published a report in July 2011 entitled 'Common Themes' (Christchurch City Council, 2011) which contains over a 1,000 direct quotes from members of the public. This data has been extracted from the Common Themes documents for the purposes of this research.


Common Themes

from Public Ideas

Christchurch Central City Plan

5.2 Categorisation of Interrogative Infrastructure Model

The ability to understand the relationship between city infrastructure(s) at different spatial scales is critical in developing a decision-making framework that is evidence based and capable of responding to the continuous changing needs of citizens. Similarly, if cities are to be considered systems (Fuller & Moore, 2017; Newman & Jennings, 2012; White, Engelen, & Uljee, 2015) urban practitioners must be capable of responding to changes in the urban environment albeit human or natural. The quality, frequency and availability of data about the environment is therefore critical in facilitating a more dynamic decision making framework. But, this is only possible if the various methods of data collection, analysis and visualization techniques outlined above are applied in practice. If applied, these techniques have the capacity to share knowledge about urban systems at multiple spatial scales which will enable a more collaborative approach to urban planning and design, breaking down the chasm between top-down and bottom-up planning process (Campbell, 2011) - see Figure. 5.1.

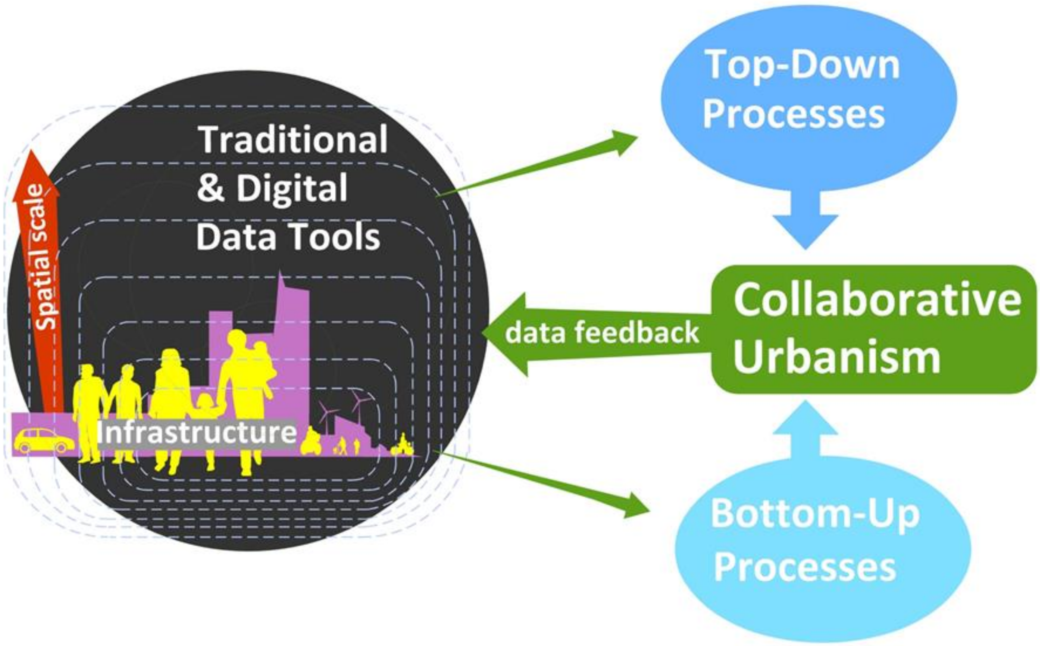


Figure. 5.1 Big and small data informing top-down and bottom up processes (Mark Dyer, Gleeson, & Grey, 2017).

In the contact of this research, infrastructure is defined as 'the basic physical and organizational structures and facilities needed for the operation of a society' (Oxford Dictionary, 2019). Whereas hard infrastructures refer to physical components, soft infrastructures often relate to the organisational or relational aspects of society (Casey, 2005; Landry, 2012; Tonkiss, 2014). The categorisation of infrastructure types is taken from the 'Framework for Interrogative Infrastructures' developed by Dyer, Gleeson and Grey (2017). The framework outlines three hard infrastructures and three soft infrastructures based on a review of existing literature. Figure 5.2 outlines a summary of each infrastructure category: utilities, urban space, building, institutional, community and personal.

HARD INFRASTRUCTURE	
UTILITIES	Physical services such as transportation, water and waste systems, ICT etc. Utilities connect and operate equally across all urban scales including national and international.
URBAN SPACE	Bounded space in the form of streets, urban plazas, local squares, playgrounds, parks etc. Typically identifiable at the neighbourhood scale or district scale.
BUILDING	Physical structures including a single or group of buildings often forming part of an urban block such as dwellings, education buildings, healthcare buildings

SOFT INFRASTRUCTURE	
INSTITUTIONAL	Formal public and private systems which provide services within the city such as local government, as well as services for healthcare, education, sport, art and culture or official community support organisations.
COMMUNITY	Informal networks or community groups that occur within neighbourhoods and districts that often rely on bridging and linking social capital. These can be online or place specific delineated by political, parish or physical boundaries.
PERSONAL	Individual support systems at a family or friendship level often bonding social capital where membership of a group is critical to a sense of belonging. It also includes education attainment and other support systems that occur at an individual level.

Figure 5.2 Definitions of hard and soft infrastructures taken from the Framework for Interrogatives Infrastructure (Mark Dyer, Gleeson, et al., 2017).

Using the stories extracted from the 'Common Themes' document a glossary of terms for each infrastructure type has been manually compiled by urban practitioners within the research team based on the definitions outlined in the 'Framework for Interrogative Infrastructures'.. This initial glossary of terms was expanded to create a more comprehensive infrastructure glossary in FLAX. As set out in the preceding chapters, FLAX uses a combination of natural learning processes and expert knowledge to identify hard and soft infrastructure, as well as the human attributes or values that link infrastructures together.

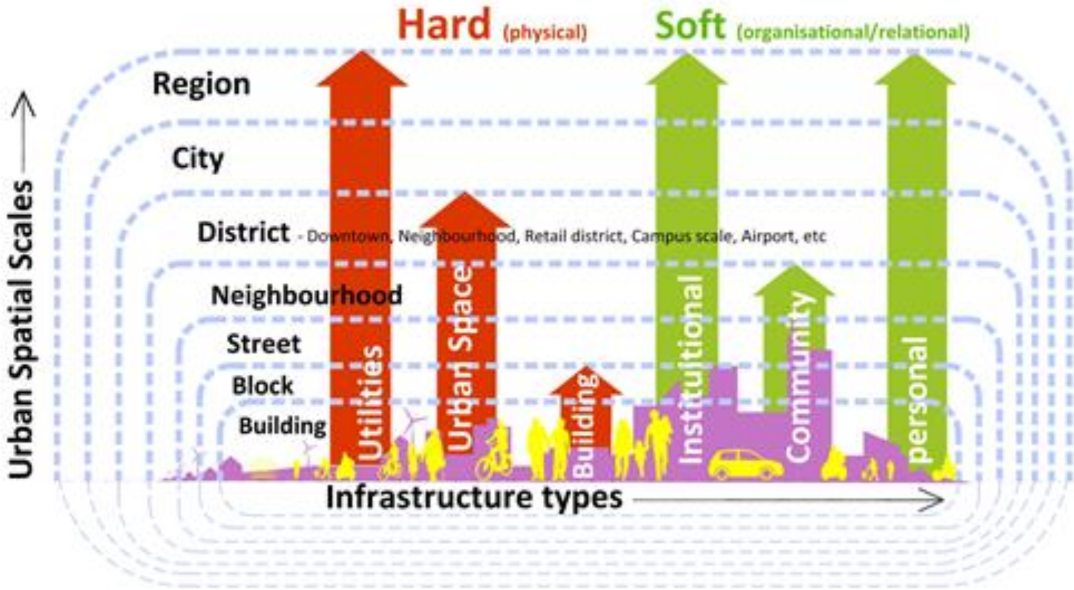


Figure. 5.3 Relationship between city Infrastructure types and spatial scale (Mark Dyer, Gleeson, et al., 2017).

Data collection, analysis and visualization methods must also be cognisant of urban scale. Figure 5.3 illustrates the degree of interaction between types of infrastructure depending on the spatial scale. (Grey, Dyer, & Gleeson, 2017, p36). While the soft infrastructure such as community and personal are more informal, so difficult to map to a particular spatial scale, the importance of understanding the relationship between soft and hard infrastructures at different geographical scales is critical in understanding the liveability and livelihood of cities. This has been reaffirmed in the results from the community workshops undertaken in Napier and Christchurch, which highlighted the importance of informal community networks at a neighbourhood scale (Dyer et al., 2018).

5.3 Data Analysis Using FLAX: Stories from Christchurch

Using 1,000 stories taken from the Common Theme's documents, FLAX has extracted 300 stories that relate specifically to the city centre. Figure 5.4 shows a list of the top 10 most frequent (noun + noun) collocations taken from these 300 stories using Flax. The list identifies 'city centre' to be the most frequent (noun + noun) collocation, mentioned in 54 stories. This is not unsurprising as this is the *core* topic of conversation. But, if the collocation 'city centre' is excluded from the list it is possible to examine the priorities of citizens in relation to the city centre itself.

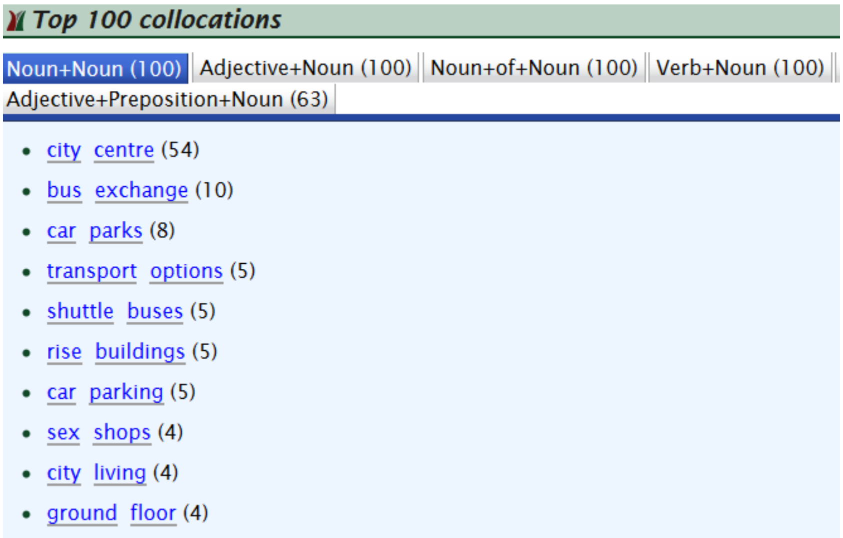


Figure 5.4 Top 10 Most Frequent (Noun+ Noun) Collocations using stories in Common Themes for 'City Centre'

The collocation 'bus exchange' is the most frequently occurring collocation (noun+nouin). However, across the list, collocations that relate to transport (i.e. bus exchange, car parks, transport options, shuttle buses and car parking) make up a substantial number of stories, 36 in total. Interestingly, the collocations about transport relate primarily to two topics, cars (16 stories) and buses (15 stories). In both instances it is the destination/arrival points which are the main topic of discussion i.e. the exchange and parking areas. While the collocation 'transport options' requires further contextualisation, this still supports the finding that 'transport' is a top trending topic amongst citizens in relation to the city centre of Christchurch.

Notwithstanding the top three most frequent collocations, the frequency of the remaining collocations on the list in Figure 5.4 is considerably less, containing between five or four stories each. These include 'rise buildings' which relates to 5 stories about building heights, both high and low rise. Buildings use, in terms of appropriateness ('sex shops') and physical location ('ground floor') within the city centre are also topics of discussion. The collocation 'city living' primarily relates to discussions around increasing activity within the city centre at different times of day and night through the re-introduction of residential uses.

Figure 5.5 Top 10 Most Frequent (Adjective Preposition + Noun) Collocations using stories in Common Themes for 'City Centre'

FLAX can also be used as a means to identify the underlying values of citizens in Christchurch. Using the same data shown from Common Themes document, the most frequent collocations for the word arrangement 'adjective + preposition + noun' have been used to represent citizen values. As shown in Figure 5.5 a core value of citizens is safety ('safety around the central city' and 'safe at all times'). The collocation 'present at all times' also relates to safety in regard to the presence of security wardens in the city centre.

Another theme revealed from the list is the importance of visitors to the city centre ('great for tourists' and 'attractive for visitors'). The collocation 'to its enhanced function' also relates to the retention of tourism through quality public realm. Connectivity between the city centre and suburbs are also important themes in terms of 'interconnecting with the suburbs', as does the topic of inclusivity in regard of the collocation 'accessible to everybody'.

Although 300 hundred stories is a relatively small sample size compared to the total number of stories received during the course of the entire public consultation, analysis of the frequency of collocations demonstrates how natural language processing and expert knowledge can be used to quickly reveal the top trending topics, or priorities, among citizens in relation to the city centre of Christchurch. While the individual topics identified in FLAX need further contextualisation at times, this snapshot can inform the foundation of an urban design brief or set of guiding principles for a community plan.

5.4 Data Visualisation for Christchurch City Centre

5.4.1 Chord Diagram to Explore the Relationship of City Infrastructures

Based on the six types of infrastructure outlined in Figure 5.2, a chord chart is capable of illustrating nine different types of infrastructure relations as depicted in Figure 5.7. From the 300 stories identified by FLAX as relevant to the city centre, 10 have been displayed in the Chord Chart shown in Figure 5.6. The visualization displays the data in a digestible format that allows the user to quickly identify core citizen values and subsequent demands on infrastructure. Immediately, the chord diagram shows Christchurch City Centre to comprise a complex network of infrastructures, supporting the prevailing rhetoric of the city as a system.

Although multiple combinations are conceivable, it possible to quickly break down the types of relationship into three categories 1) soft and hard infrastructures 2) soft and soft infrastructures and 3) hard and hard infrastructure. As shown in Figure 5.6, the chord between hard and soft infrastructures is largest with 19 stories. This implies people often spoke about a combination of hard and soft infrastructures in a single story which suggests the complementary nature of hard infrastructures and soft infrastructures, whereby one support or amplifies the value of another. Interestingly, the relationship between two hard infrastructures was more common (16 links) than those between two soft infrastructures (8 links).

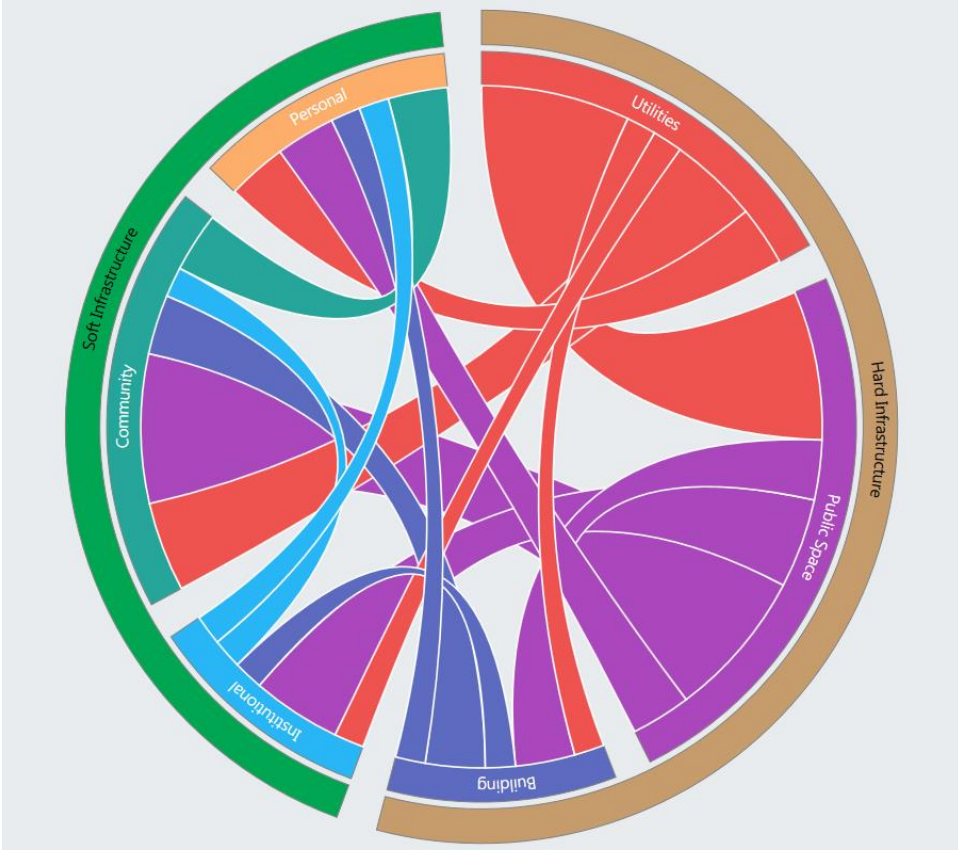


Figure. 5.6 Chord Diagram showing 10 Stories (noun+ noun) that include the collocation 'City Centre' with Common Themes

5.4.2 Individual Chords for Public Space

A preliminary reviews of the chord chart in Figure 5.6 shows residents spoke about hard infrastructure more than soft infrastructure; the latter included 12 stories and the former 20 stories. The hard infrastructure, 'public space' had the greatest number of stories as well as the highest number of connections with other infrastructure types. Figure 5.8 shows all the chords in relation to 'public space'. This result suggest that public space is a top trending topic amongst residents and should be a priority for urban practitioners in the future development of the city centre of Christchurch.

Given the value residents attribute to public space, the three infrastructure relations that relate to 'public space' as highlighted in grey in Figure. 5.7 shall be explored in more detail in this chapter. This is to demonstrate the capacity of FLAX, as analytical tool, to extract stories from individual chords in the diagram and interrogate these stories in more detail.

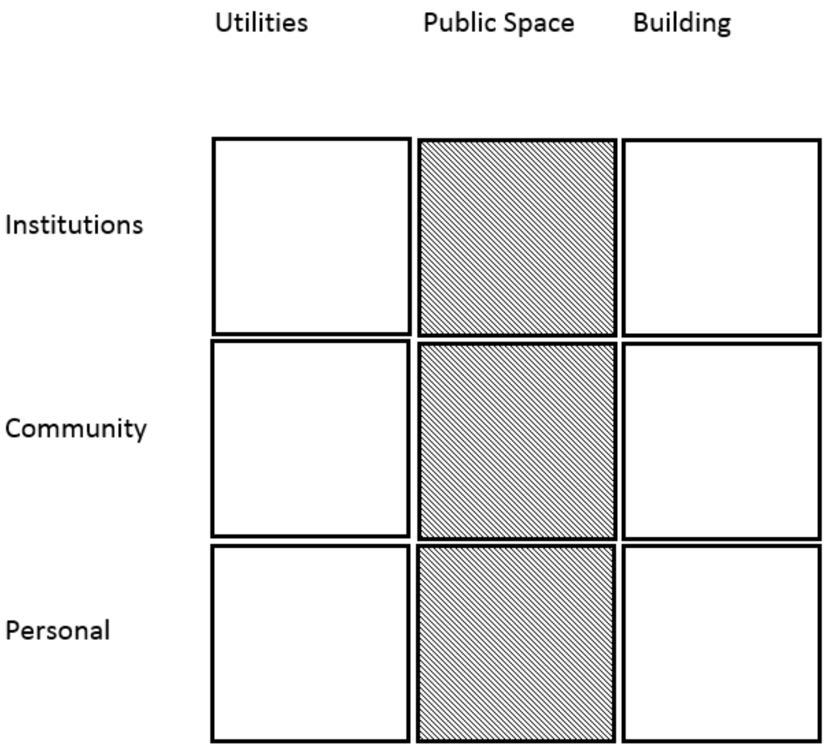


Figure. 5.7 Chord Diagram Combinations of Infrastructure Types

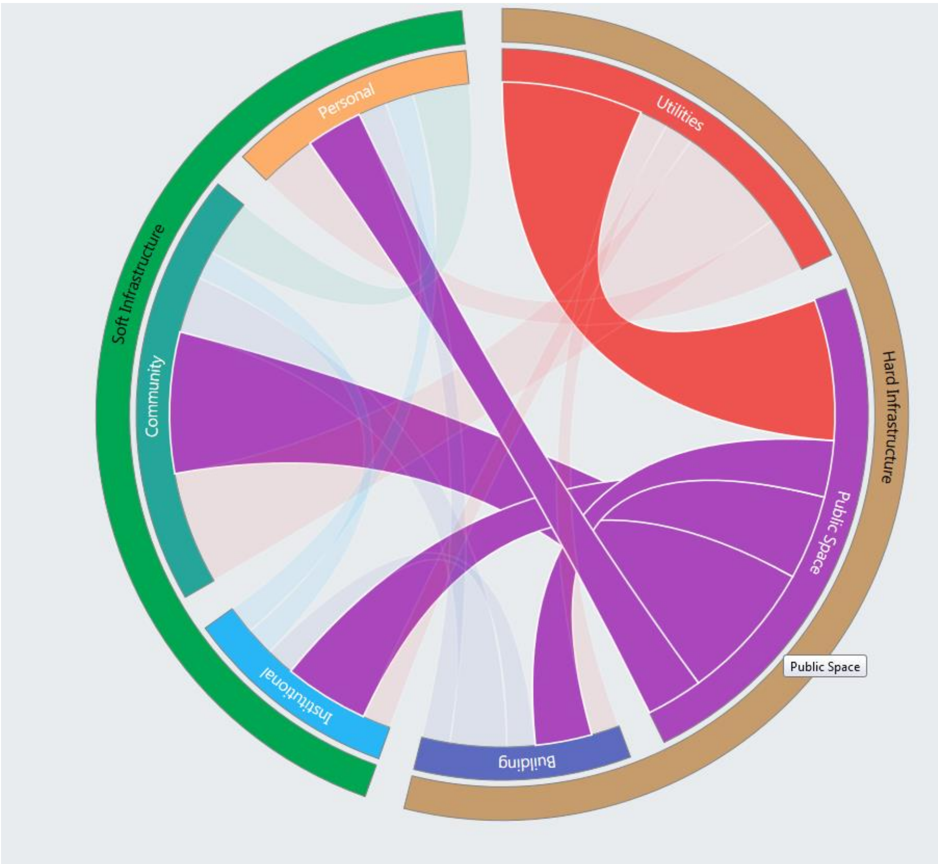


Figure. 5.8 Chord Diagram with All Public Space Chords Highlighted.

Chord: Public Space and Community

Figure. 5.9 shows all the stories that relate to 'public space' and 'community' based on the glossary of keywords used by FLAX. Five stories have been identified. Here, the keywords for 'public space' are in purple and the keywords for 'community' in blue. In addition, citizen attributes have been highlighted in green. The attributes reflect resident values so explain *why* residents value a particular infrastructure type or relationship. Based on the keywords highlighted it is possible to quickly capture the narrative in an individual story in terms of infrastructure priorities and related citizen values.

For example, the expanded story in Figure 5.9 shows the 'city centre' as a keyword identified under the category 'public space' and 'locals' as the keyword for 'community'. The attribute that links 'locals' and the 'city centre' is 'cheap'. From this snapshot it is possible to determine the need to connect locals with the city centre in an affordable manner. This example shows how interrogation of infrastructure relationships can be used to highlight a particular need of a citizen, but also the values which underpin this need.

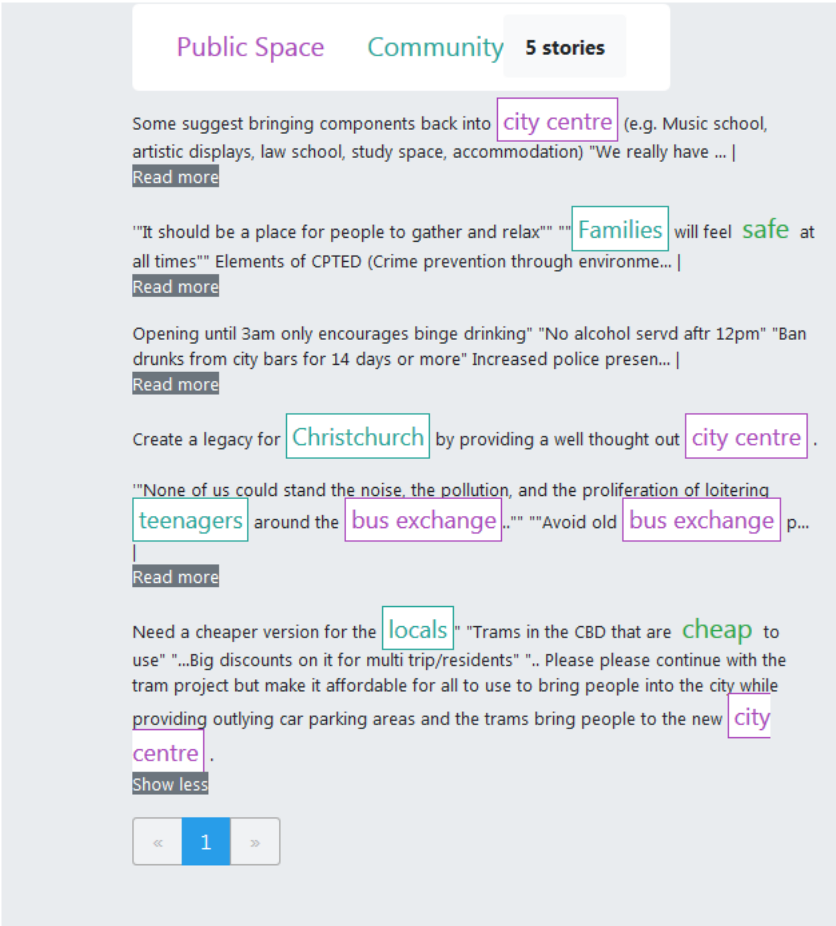


Figure. 5.9 Stories extracted from the Chord Diagram showing Public Space and Community Infrastructures.

Chord: Public Space and Institutional

Figure. 5.10 shows all the stories that relate to 'public space' and 'Institutions'. Three stories have been identified. The keyword for public space remains in purple, whereas the keywords for institutions is in blue. In this example, public space is being talked about in relation to the 'city centre' or the 'heart of the city centre'. The 'institution' in this story is the 'police', in particular in relation to the visibility and presence of the police within the city centre. The attributes that are highlighted in green are primarily negative in terms of dark, seedy, late, low etc. Although these words are negative, so perhaps do not reflect a value which is often positive, the adjectives appear to relate to a sense of safety. In short, the narrative is one of the role police have in securing the safety of public spaces in Christchurch city centre.

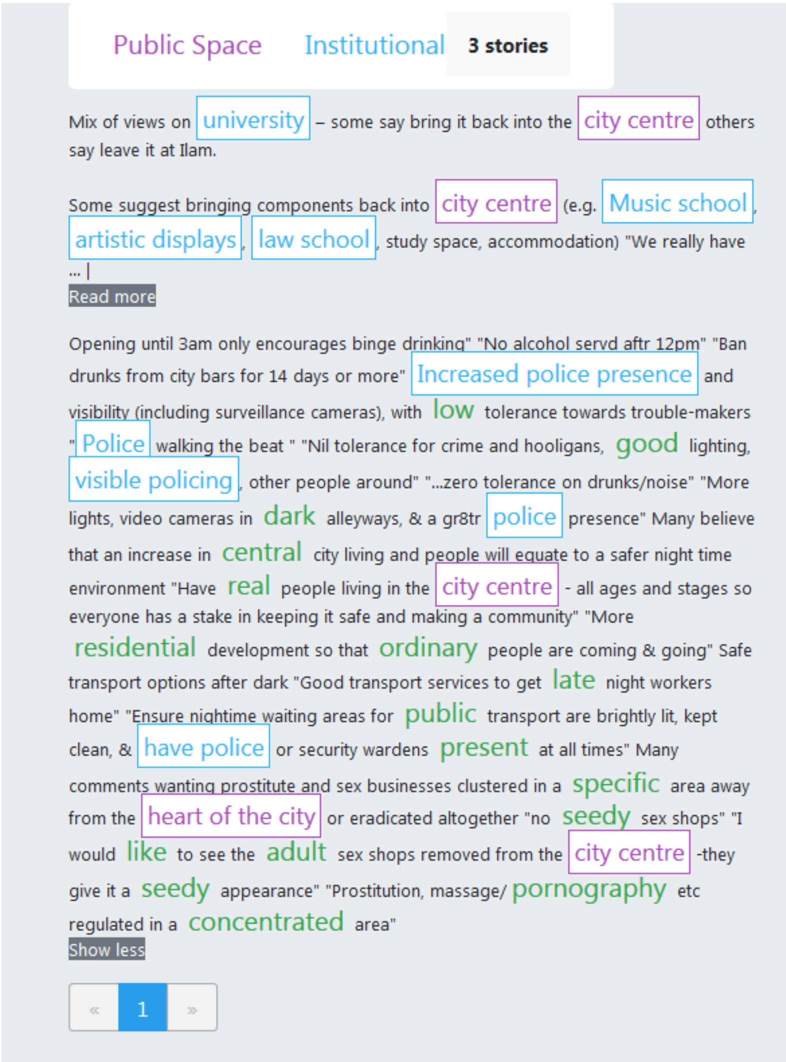


Figure. 5.10 Stories extracted from the Chord Diagram showing Public Space and Institutions Infrastructures.

<u>Chord: Public Space and Personal</u>

Figure. 5.11 shows all the stories that relate to the infrastructure categories 'public space' and 'personal'. Taking the first story in the list, the terms identified in relation to public space are 'city square', 'alleys' and 'corners'. These are very

specific physical attributes compared to the previous examples Figure 5.10. FLAX has also highlighted broader terms such as ‘environmental design’ and ‘urban design ‘within public space category which are not place specific. The keywords for the infrastructure ‘personal’ include ‘people’ and ‘residents’. This story begins to show a relationship between locals and the city centre, particularly in relation to routes and nodes (i.e. alleyways and corners) as a perceived environmental barriers. If we examine the keywords for the attributes ‘safe’, ‘attractive’ and ‘mixed’ are identified. This implies the narrative is one of safe streets, while also providing potential design solutions such as mixed use, aesthetics etc.

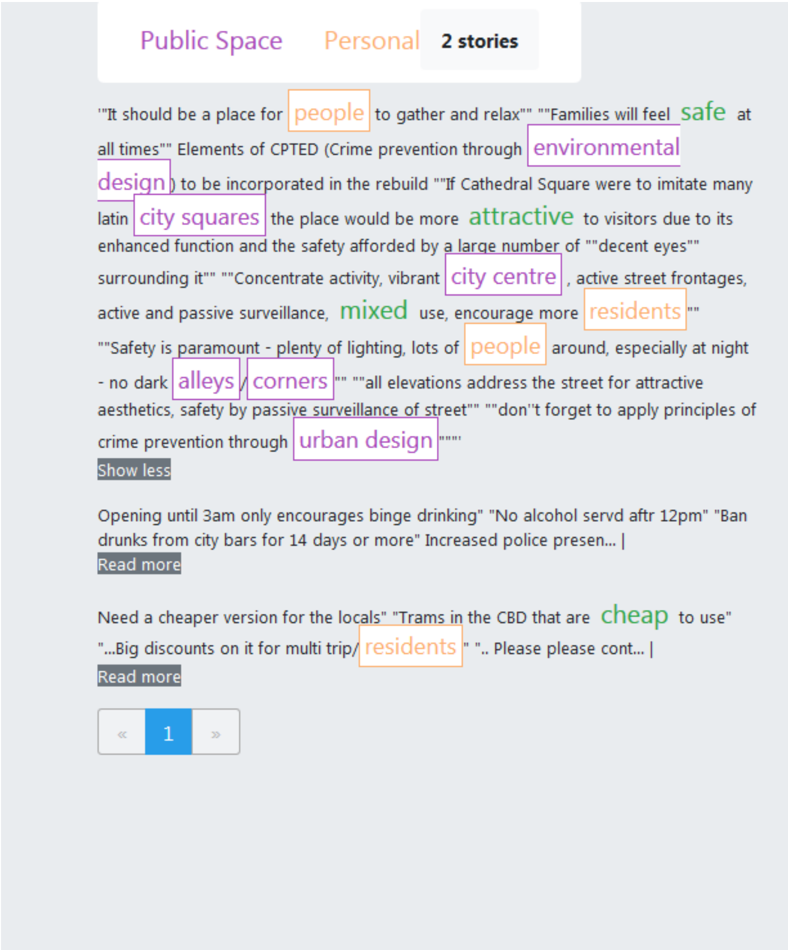


Figure. 5.11 Stories extracted from the Chord Diagram showing Public Space and Personal Infrastructures.

5.4.3 Summary of Citizen Priorities for Public Space in Christchurch City Centre

To summarise, the chord charts shown in Figure 5.9, 5.10 and 5.11 represent the relationship between public space in relation to institutions, community and personal. In terms of community and personal, the importance of *locals* or *residents* in the city centre is highlighted. However the keyword ‘people’ is often captured which is less informative, as such this is an area of the syntactic analysis software which needs further refinement.

In terms of institutions, the role of the police in ensuring a safe city centre at all times, albeit day and night, is a common narrative. Similarly, the terms 'city centre' frequently occurs under the category public space which is less informative so should be filtered out by FLAX. However, specific examples such as squares, alleys and corners show how FLAX is capable of selecting specific features of the built environment that are perceived as important features of public space by citizens.

Although the syntax used to identify keywords for attributes needs further refinement, a review of the attributes identified in the three examples quickly captures a narrative around safety, affordability and attractiveness. If expanded upon further, these three attributes/values could form the foundation of an urban design brief or set of guiding principles. Further research into the linguistics around infrastructure and citizen attributes/values is required to determine the most appropriate word arrangements required to identify the values of citizens.

5.5 Data Storytelling Leading to Personas and Infographics to Inform Evidence-Based Design Briefs.

5.5.1 Personas

The information illustrated in the chord diagram can be used to develop personas based the number and type of chords in an individual story, as well as the attributes and values identified within them. Personas are currently used as a tool to enable designers to understand a user's needs (Lupton, 2017). However, due to time and resource constraints, 'dummy' personas are often used rather than authentic user profiles which require hours of interviews, transcribing and analysis (Nielsen, 2012).

The ability to automatically generate personas based on the analysis of large qualitative data sets using FLAX and data visualisations would be a valuable tool for urban practitioners, enabling them to design and plan urban spaces based on the needs of users. These personas could reflect the evolving needs of citizens in line with the emerging data available, allowing designers to adapt spaces or buildings according to the changing needs of users. Thematic analysis of trends could allow for future casting in order to predict the needs of future users.

To demonstrate how the data analysis and visualization techniques outlined in this report could be used to generate user profiles, two personas have been created based on the stories extracted from the Common Themes document. The information has been manually organised, but the content is taken from information displayed in the Chord Chart. Further research is required to explore how this information could be automatically uploaded into personas.

Persona #1: The Night Shift Worker

The persona shown in Figure 5.12 is developed from a story in the Chord Chart between Utilities and Institutions. On the left hand side, the user priorities of hard and soft infrastructure are shown along with the attributes linked to these infrastructure priorities. The keywords have been identified by FLAX. Based on this information, it is possible to gain a snapshot of the user needs. In terms of utilities; lighting, cameras and public transport are key priorities, whereas the institutional priorities relate to policing and security wardens. While the attribute keywords identified require further refinement, the underlying value is one of safety which is shown on the right hand side table. In addition to 'safety and security', 'visual aesthetics' has been highlighted as core values for the user. These core values are drawn from an earlier review of value systems with environmental psychology and Maori health models (Dyer, M. et al., 2018). The quotes on the right hand side support the values, while informing the design attributes. This design attributes show how a persona can then be used to form guiding principles for an urban design brief.

Persona #2: The Parent

The persona in Figure. 5.13 relates to Utilities and Communities. This is a shorter story compared to the previous personas as such the number of keywords is less, but it is still possible to gain a snapshot of the user priorities. Here, lighting appears as a priority in terms of utilities, and the community focus is that of families. The attributes keywords are more tangible in this instance so can be easily translated into values. However, not all the attribute words captures the value in the stories, for example 'social interaction' is overlooked as the words 'gather' and 'relax' are adjectives so are not picked up by FLAX. Again, this highlights a need to further refine the natural language processing used by FLAX in order to pick up alternative word arrangements in relation to attributes.

Persona #1

The Night Shift Worker

Hard

Surveillance cameras, lighting, video cameras, transport options, transport services, public transport

Soft

police presence, trouble makers, hooligans, all ages, real people, night workers, security wardens

Attribute Keywords

low tolerance, dark, visible, residential, ordinary, present, seedy, concentrated, grieve

Values	Quotes	Design Attributes
safe and security	– 'safe transport options after dark' – 'good transport services to get late night workers home'	regular, safe and available public transport
	– 'ensure night-time waiting areas for public transport are brightly lit' kept clean, and have police or security wardens present at all times' – 'more lights in dark alleyways'	lighting
	– 'increased police presence and visibility' – 'police walking the beat' – 'zero tolerance on drunks and noise' – 'surveillance cameras' – 'video cameras in dark alleyways'	visible law enforcement security cameras
	– 'more residential development' – 'increase in central city living'	more residential development within the city
	– 'prostitute and sex businesses clustered in a specific area away from the heart of the city or eradicated al together' – 'adult sex shops removed from the city centre' – 'Prostitution, massage/pornography etc regulated in a concentrated areas'	sex economy / appropriate uses in the city centre
visual aesthetics	'beautiful buildings and shops'	

Persona #2

The Parent

Hard
lighting

Soft
families

Attribute Keywords
safe, attractive, mixed

Values	Quotes	Design Attributes
safety and security	– 'active and passive surveillance' – 'safety is paramount – plenty of lighting, lots of people around...'	passive surveillance
	– 'no dark alleyways and corners' – 'a large number of "decent eyes"'	lighting
	– 'crime prevention through environmental design'	environmental design
foster social interaction	– 'place for people to gather and relax' – 'encourage more residents'	
visual aesthetics	– 'imitate Latin city squares', 'attractive aesthetics' – 'all elevations address the street'	
vibrant and active city centre	– 'concentrate activity... active street frontages' – 'enhanced function'	

Figure. 5.13 Persona #2: the Parent

5.5.2 Urban Design Briefs

Infographics can also be developed using stories from the Common Themes document as a tool for data storytelling (see Chapter 4). The results from the Chord Chart in conjunction with the keywords identified by FLAX, have been used to develop an infographic that illustrates the data narrative for Christchurch City Centre which could form the basis of an urban design brief, see Figure 5.14.

The larger text in quotation marks are the core values of citizens, many of which are direct quotes that feature in the personas in Figure. 5.12 and 5.13. These core values have then been paired with municipal data (Christchurch City Council, 2019) to show how these tools can be used to bridge the gap between bottom up and town down processes within planning. In this instance the statistical data is past, but statistical projections could be used to aid future casting.

Although the infographic's shown in Figure 5.14 draw upon a relatively small data source, this shows the capacity of infographics to bring together bottom up and top down approaches in planning to create a more holistic narrative of a city. This has the potential to be a powerful tool to facilitate a more collaborative approach to decision-making that is evidence based and valued-based.

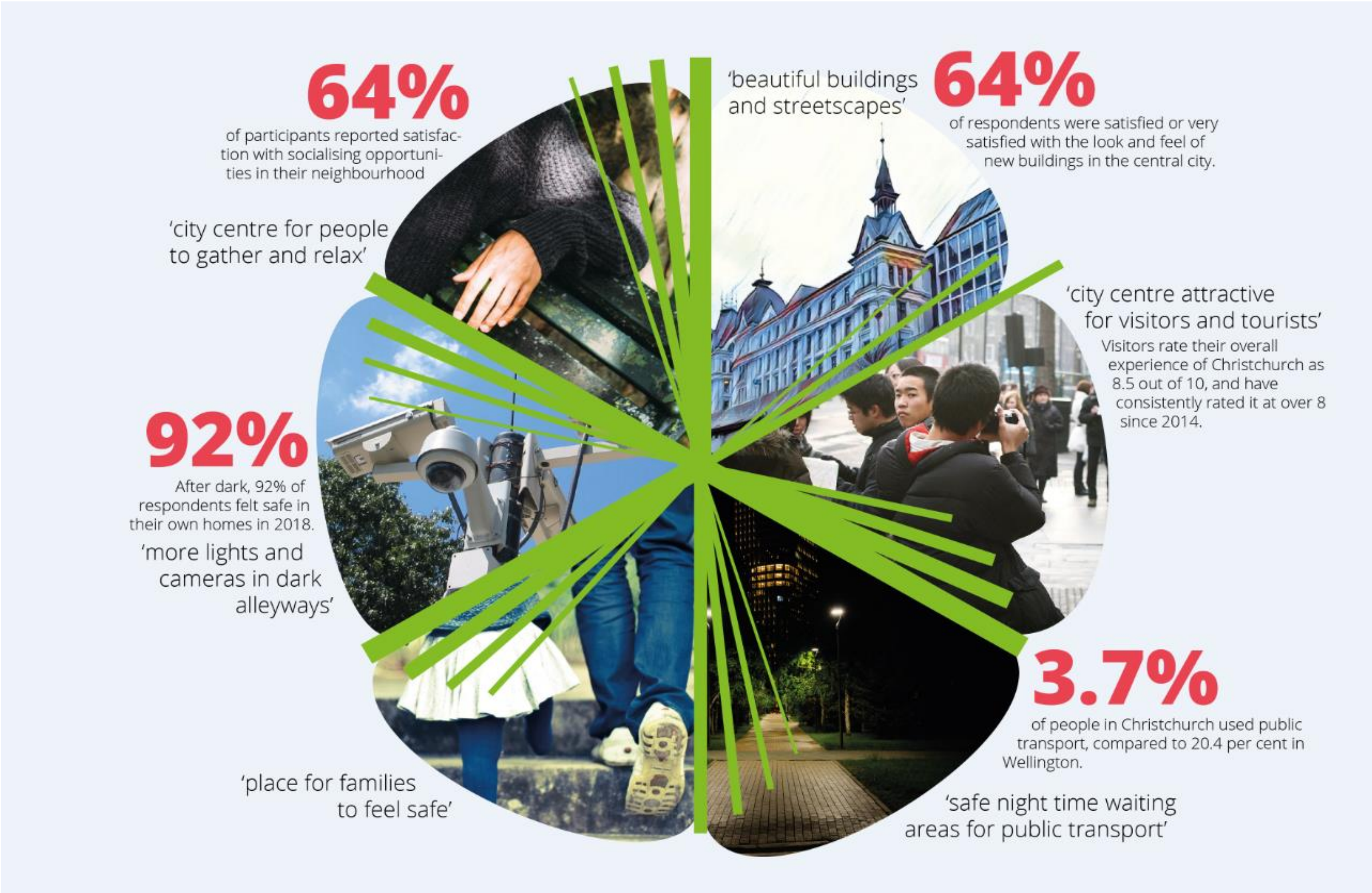


Figure. 5.14 Collage Based on the 10 Stories from the Common Themes Document using Municipal Data and User Priorities from Persona

5.6 Summary

The application of the data collection, analysis and visualization techniques to the case study of Christchurch shows how these tools can be used to rapidly identify the needs and priorities of citizens in terms of underlying values and subsequent infrastructure requirements. Having gained an awareness of citizen priorities, the ability to then analyse the interrelationships between infrastructures shows how the tools can be used by practitioners to examine the urban system in a meaningful way. The use of data visualisations and infographics also make it possible to communicate these results in an easy to digest format through data storytelling and personas, both of which can form the basis of an urban design brief or set of guiding principles. This case study shows the very real potential of these tool in enabling urban practitioners to respond to the evolving needs of citizens, identifying specific design attributes required to foster particular infrastructure relations, in order to improve the livelihood and liveability of city infrastructures.

6. Conclusion

The United Nations estimates that 68% of the world's population will be living in urban areas by 2050 (United Nations, 2018). The systems, tools and interfaces developed to shape, design and plan our towns and cities now must therefore be capable of responding to an increasingly complex urban system to ensure future inhabitants have access to 'safe and adequate housing, clean air and basic services and live in resilient and sustainable communities ' (United Nations, Sustainability Goal 11, 2018)

The suite of tools developed by the Urban Narrative group demonstrate how data gathering can be a valuable asset to effective city governance, facilitating evidence-based decision-making and innovative design solutions. The ability to harness the collective knowledge of citizens in order to understand not just 'what' people want, but also the 'why' they want it, has the potential to shift design thinking allowing for higher quality design that effectively meets the needs of citizens.

Using Christchurch as a case study, it is possible to see the powerful role data analytics and visualisations can have in facilitating our understanding of urban system through the relationships between infrastructures and the values that weave them together. Equally, the ability to distil large qualitative data into usable outputs such as personas and urban design briefs through the curation of data can effectively aid practitioners in the decision-making process.

There are a number of lessons to be learnt from the initial suite of prototype tools. These limitations are areas of further research and are briefly outlined below.

1) The glossary of terms in FLAX was constructed manually by urban designers to map each key term to an associative infrastructure category based on the story contexts. This procedure has proven time intensive and error-prone, and thus could cause imprecision and inconsistency of story categorisation. Further research needs to be undertaken to develop a smart tool to assist this mapping procedure and improve the accuracy and quality of glossary of terms. A combination of syntactic and semantic analysis should be explored to better understand the sentiment and context behind words or phrases. Both these avenues of research would improve the overall accuracy of the tools while ensuring their capable of running larger qualitative data sets.

2) The personas and design brief's at present are built manually using urban design expertise which is resource intensive. This process also excludes users who are not urban practitioners form generating such outputs. Further research should be undertaken to develop a template to extract data to automatically generate personas and design brief. This will make the software accessible to non-experts, while educating a wider audience about the needs of citizens in relation to the urban systems.

3) The link between the data stories and existing statistical data, as shown in the urban design brief, is manually curated. The research acknowledges that data stories is one tool amongst many that can help to create a compelling and inclusive vision for our cities. As shown in the infographic of the design brief, both qualitative and quantitative formats should be used to open debates. Further research should explore methods of linking existing statistics with the data stories to provide a holistic vision of the city.

In summary, cities must gain a deeper understanding of data and its role in a collaborative urban systems**.** In line with this, cities must commit to establishing clear guidelines on data governance to ensure ethical collection, storage and use of data.

.

Bibliography

Apache OpenNLP. (n.d.). Retrieved from https://opennlp.apache.org/

https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

Bostock, M. (n.d.). D3.js - Data-Driven Documents. Retrieved May 31, 2019, from https://d3js.org/

Campbell, K. (2011). Smart urbanism: Making massive small change. *Journal of Urban Regeneration & Renewal*, *4*(4), 304–311.

Casey, S. (2005). *Establishing standards for social infrastructure*. UQ Boilerhouse, Community Engagement Centre.

Christchurch City Council. (2019). Facts, stats and figures. Retrieved April 11, 2019, from https://www.ccc.govt.nz/culture-and-community/christchurch/statistics-and-facts/facts-stats-and-figures/

Data visualization. (2019). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Data_visualization&oldid=886400533

Defining Text Analytics | The Intelligent Enterprise Blog. (2009, November 29). Retrieved March 25, 2019, from https://web.archive.org/web/20091129171151/http://intelligent-enterprise.informationweek.com/blog/archives/2007/02/defining_text_a.html

Dyer, M., Dyer, R., Hinze, A., Mackness, K., Wilkins, R., Wilson, J., & Wu, S. (2018). *St. Mary's Church, Addington, Community Workshop Project Report #2.*

Dyer, M., Hinze, A., & Dyer, R. (2018). *Building Research Capacity in Communities: Community Workshop Feedback Report #1.* Waikato.

Dyer, Mark, Corsini, F., & Certomà, C. (2017). Making urban design a public participatory goal: toward evidence-based urbanism. *Proceedings of the Institution of Civil Engineers - Urban Design and PlaBostock, M. (n.d.). D3.js - Data-Driven Documents. Retrieved May 31, 2019, from https://d3js.org/nning*, *170*(4), 173–186. https://doi.org/10.1680/jurdp.16.00038

Dyer, Mark, Gleeson, D., & Grey, T. (2017). Framework for collaborative urbanism. In *Citizen empowerment and innovation in the data-rich city* (pp. 19–30). Springer.

Friendly, M. (n.d.). *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. 79.

Fuller, M., & Moore, R. (2017). *The death and life of great American cities*. Macat Library.

Grey, T., Dyer, M., & Gleeson, D. (2017). Using Big and Small Urban Data for Collaborative Urbanism. In C. Certomà, M. Dyer, L. Pocatilu, & F. Rizzi (Eds.), *Citizen Empowerment and Innovation in the Data-Rich City* (pp. 31–54). https://doi.org/10.1007/978-3-319-47904-0_3

Hagley, J. (n.d.). What's the difference between an Infographic and a Data Visualisation? Retrieved April 8, 2019, from https://www.jackhagley.com/What-s-the-difference-between-an-Infographic-and-a-Data-Visualisation

Holten, D. (2006). Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 741–748. https://doi.org/10.1109/TVCG.2006.147

Iliinsky, N., & Steele, J. (2011). *Designing Data Visualizations: Representing Informational Relationships*. O'Reilly Media, Inc.

Infographic. (2019). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Infographic&oldid=888308079

Jurafsky, D., Chahuneau, V., Routledge, B. R., & Smith, N. A. (2014). Narrative framing of consumer sentiment in online restaurant reviews. *First Monday*, *19*(4). https://doi.org/10.5210/fm.v19i4.4944

Kessler, J. S. (2017). Scattertext: a Browser-Based Tool for Visualizing how Corpora Differ. *ArXiv:1703.00565 [Cs]*. Retrieved from http://arxiv.org/abs/1703.00565

Kirk, A. (2016). *Data Visualisation: A Handbook for Data Driven Design*. SAGE.

Krum, R. (2013). *Cool infographics: Effective communication with data visualization and design*. John Wiley & Sons.

Landry, C. (2012). *The art of city making*. Routledge.

Lupton, E. (2017). *Design is Storytelling*. Retrieved from https://books.google.co.nz/books?id=1wPlAQAACAAJ

Madsen, S., & Nielsen, L. (2010). Exploring Persona-Scenarios - Using Storytelling to Create Design Ideas. In D. Katre, R. Orngreen, P. Yammiyavar, & T. Clemmensen (Eds.), *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts* (pp. 57–66). Springer Berlin Heidelberg.

Make a Bubble Chart Online with Chart Studio and Excel. (n.d.). Retrieved April 2, 2019, from https://help.plot.ly/excel/bubble-chart/

mark.js – JavaScript keyword highlight. (n.d.). Retrieved March 28, 2019, from https://markjs.io/

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Newman, P., & Jennings, I. (2012). *Cities as sustainable ecosystems: principles and practices*. Island Press.

Nielsen, L. (2012). *Personas - User Focused Design*. Springer Publishing Company, Incorporated.

Read, J., Reutemann, P., Pfahringer, B., & Holmes, G. (2016). Meka: a multi-label/multi-target extension to weka. *The Journal of Machine Learning Research*, *17*(1), 667–671.

Ryan, L. (2016). Introduction. In L. Ryan (Ed.), *The Visual Imperative* (pp. xix–xxi). https://doi.org/10.1016/B978-0-12-803844-4.00020-0

Tonkiss, F. (2014). *Cities by design: the social life of urban form*. John Wiley & Sons. United Nations. (2018, May 16). 68% of the world population projected to live in urban areas by 2050, says UN | UN DESA | United Nations Department of Economic and Social Affairs. Retrieved May 31, 2019, from https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

United nations. (2018). Goal 11: Make cities and human settlements inclusive, safe, resilient and sustainable — SDG Indicators. Retrieved May 31, 2019, from https://unstats.un.org/sdgs/report/2018/goal-11/

White, R., Engelen, G., & Uljee, I. (2015). *Modeling cities and regions as complex systems: From theory to planning applications*. MIT Press.

Witten, I. H., Wu, S., Li, L., & Whisler, J. L. (n.d.). *A new approach to computer-assisted language learning*. 124.

Wu, S., & Witten, I. H. (2015, October 9). *Using Wikipedia for language learning*. 124–131. Retrieved from https://hdl.handle.net/10289/9686

LAST PAGE