An Evaluation to Different Types of Data Visualization

A study submitted in partial fulfilment of the requirements for the degree of MSC Information Management at THE UNIVERSITY OF SHEFFIELD

by

BIN SHU

September 2016
Acknowledgment

It is pleasure to thank those who contribute to this study. First and foremost, I would like to express my deepest gratitude to my supervisor, Paula Goodale, for her constant support and guidance throughout the dissertation journey. This paper would not have reached the current stage without her patient instructions.

I am also greatly indebted to all respondents of questionnaire who provide valuable data for this study. Moreover, I want to thank my friends and families who have helped me keep great confidence to overcome difficulties of this study by their continuous encouragement and loving considerations.
Abstract

Background
Data can be treated as a new resource for the public and becomes increasingly important in the world. It may appear and spread through the Internet or other medium. Data has a high increasing speed. Many people are considering how to use the needed data.

Aims
The study aims to analyse individuals’ attitudes towards these essential features of data visualization. It also collects their responses for different types of data visualization, compare and analyse the performance of each type of data visualization in expression meanings to people.

Methods
An online questionnaire was developed, based on instruments used in previous studies, and was piloted with the students in Sheffield University. There are 60 responses for the questionnaire.

Results
The importance of the features in data visualization was verified. The legend and data source are the most important aspects in the data presenting and data quality respectively. There are also acquired the evaluations for different types of data visualization. Some improvements are suggested to overcome the drawback of the data visualization.

Conclusions
It is concluded that the features may influence the performance of different types of data visualization. The infographic may have the advantages to combine each kind of data visualization to tell a whole story for the readers.
## Contents

Acknowledgement ............................................................................................................. 1
Abstract .................................................................................................................................. 2
1. Introduction ......................................................................................................................... 1
   1.1 Context of the Research ................................................................................................. 1
   1.2 Research Question ........................................................................................................ 3
   1.3 Research Aims and Objectives .................................................................................... 3
      1.3.1 Research Aims ....................................................................................................... 3
      1.3.2 Objectives ............................................................................................................ 3
   1.4 Structure of Dissertation .............................................................................................. 3
2. Literature Review ............................................................................................................... 5
   2.1 Introduction .................................................................................................................... 5
   2.2 Development of Data Visualization ............................................................................ 5
   2.3 Essential Features of Data Visualization .................................................................... 5
   2.4 Classification of Data Visualization .......................................................................... 6
   2.5 Standards in Data Visualization Design .................................................................... 8
      2.5.1 Unsolved Information Visualization Problems .................................................... 9
      2.5.2 Measures to do in Design .................................................................................. 9
      2.5.3 Measures not to do in Design ............................................................................. 10
3. Methodology ....................................................................................................................... 11
   3.1 Introduction ................................................................................................................... 11
   3.2 Research Approach ....................................................................................................... 11
   3.3 Data Collection Method .............................................................................................. 12
      3.3.1 Online Survey Questionnaire ............................................................................. 12
      3.3.2 Question Design ................................................................................................ 12
      3.3.3 Participants ......................................................................................................... 13
   3.4 Ethical Aspects .............................................................................................................. 13
   3.5 Limitations of Methodology ......................................................................................... 13
4. Results and Findings ......................................................................................................... 15
   4.1 Introduction ................................................................................................................... 15
   4.2 Participants Information .............................................................................................. 15
   4.3 Basic Attitude towards Data Visualization ................................................................ 17
      4.3.1 Presenting Data .................................................................................................... 18
      4.3.2 Data Quality ....................................................................................................... 19
4.4 Results for Different Types of Data Visualization .......................................................... 21
 4.4.1 Result for Bar Chart .................................................................................................... 21
 4.4.2 Result for Line Chart ................................................................................................ 24
 4.4.3 Result for Pie Chart .................................................................................................. 27
 4.4.4 Result for Bubble Chart ............................................................................................ 29
 4.4.5 Result for Map ............................................................................................................ 32

4.5 Chi-Squared for Language ............................................................................................. 35

5. Discussion ................................................................................................................................. 37
  5.1 Introductions ..................................................................................................................... 37
  5.2 Importance of Features in Data Visualization ................................................................. 37
    5.2.1 The Features in Presenting Data ............................................................................... 37
    5.2.2 The Features in Data Quality .................................................................................. 38
  5.3 Evaluation of Different Data Visualization ....................................................................... 38
  5.4 Infographic and Data Visualization ................................................................................. 40

6. Conclusion ................................................................................................................................. 42
  6.1 Summary of the Research ............................................................................................... 42
  6.2 Limitation and Recommendation ..................................................................................... 43

Reference ..................................................................................................................................... 44
1. Introduction

1.1 Context of the Research

Nowadays, data can be treated as a new resource for the public and becomes increasingly important in the world. It may appear and spread through the Internet or other medium. Data has a high increasing speed. Many people are considering how to use the needed data. Borgman (2007) suggests the definition of the data that data can be used to show some incidents which are occurred before. In order to acquire the data, some essential measures are needed include the observation, test and record. For most people, the data cannot be used because it is abstract and discrete (Rosenberg, 2013). It may need a process to deal with the data so that the original data can acquire some special meanings. Five indexes can be created to identify different kinds of data (Kitchin, 2014). When considering the form of data, the data can be divided into two types: the quantitative data and the quality data. The quantitative data mainly refers to the physical value of the item while the quality data may concern more about the non-physical characters of data. There are four different data measurement scales to classify the types of data (Tate, 1999), including the nominal, ordinal, interval and ratio. The nominal data is always used to make a label for the variable and in has no numerical value. The ordinal explains the rank to array the data in an order such as from low to high. It may useful to express the degree of the feelings. Interval shows an extent between the continuously items. Once the order of the variable is suggested, the interval may give a description to explain the exact difference of these variables. Ratio includes the characters of the three types above. Meanwhile, it can provide an origin to expand the data range. Considering the structure of data, three types can be defined, which have the structured data, the semi-structured data and the unstructured data. In general, the structured data may be more useful because this type of data has a better relationship and it may easier to be managed. The data can also be divided into captured, exhaust, transient and derived data (Manyika, 2011). This classification is based on the different sources when collect the data. Data can be created by different producers, so it may have different types as the primary data, secondary data and tertiary data. Users may generate the indexical data, attributed data and metadata according to different requirements.

Data has different classification and can be divided into different types. However, people may have difficulty to use the data directly. Frické (2009) suggests a knowledge pyramid model to show the difference between data, information, knowledge and wisdom. In this DIKW model, Data is considered as the
abstracted elements and located in the bottom. Users may need to gives some description to the data and transfer the data into information. Compared with data, the information seems to be some meaningful items which have a wide link to other elements. In the next process, the information is required to have an organising measurement. This can make the information has a potential value and turned to be knowledge. When knowledge is under analysing and applying, it may create a concept of wisdom.

In order to make the transfer process more successfully, some tools or methods are needed. Data visualization can be treated as a useful method to achieve this goal. Few (2009) gives a definition to the data visualization that data visualization is a method to express the abstract data with some vivid graphs. Generally, the data visualization focus on two aims: one is the sense-making which concentrate on the data analysis and the other one is communication which help the readers understand the data better. When the data visualization is used to the data analysis, people may concern more about the results or the finding of the data. By showing the visualized data to others, people may feel easier to realize the relationships of the related data. In addition, when using the data visualization to enhance the communication, it may focus on the expression of the data. In order to make people understand the meanings, some pictures or images may be appeared. Data visualization is helpful to promote the visual perception abilities of people. By expanding the cognitions, people may feel easier to accept the presentation of data (Kirk, 2012).

Azzam (2013) suggests three criteria to define the data visualization. Firstly, the data visualization selects the qualitative data or quantitative data as the basic data source. Secondly, the selected data would be shown by some graphs or pictures. Finally, there will have an access to allow the users to analyse the data or make some communications. For the first criteria, the quantitative data and quality data refers to the data which people are likely to use in description for some items. In the second criteria, it would give some standards for the transfer of data to data visualization. The data need to be recorded carefully. When dealing with the collected data, designers may need to focus on the integrity and the consistency of the data to ensure the expression is correct. The last criteria aim to test the result of data visualization. The researcher should have a deep understanding for the data visualization and can point some main elements in the graphs.

MacEachren (1995) suggests some questions to the data visualization focus on cartography. These questions may include the identification about the meanings in data visualization, the method which can
used to collect the main items from graphs and the creation of specific situation from the meanings. These questions may be used to improve the performance of communication aspects in data visualization. Readers can also improve their ability to understand the data visualization by realizing the question above.

1.2 Research Question

While there are many research analyse the use of data visualization, few of them concern the performance of the specific types such as the bar chart and line chart. It may appear some question as follows:

What are the responses of people when they encounter the charts?

Whether there is a more suitable condition for these different types of data visualization?

Is there a format to combine these different types of data visualization to express the data?

1.3 Research Aims and Objectives

1.3.1 Research Aims

When design the data visualization, some people may concern more about the tittle or the data sources while the others may concern the chart type. The dissertation intends to analyse individuals’ attitudes towards these essential features of data visualization. It also collects their responses for different types of data visualization, compare and analyse the performance of each type of data visualization in expression meanings to people.

1.3.2 Objectives

- Summary an initial classification of data visualization and select some basic types of data visualization for the survey.
- Verity people’s attitudes towards essential features of data visualization.
- Evaluate each type of data visualization and analyse the performance of these data visualizations.
- Analyse the relationship between the infographic and data visualization
- Analyse some improvements to help the infographic telling story to readers

1.4 Structure of Dissertation

The dissertation has six chapters. It is beginning with the introduction and the other five chapters are as
follows:

Chapter 2: Review some related literatures about data visualization, including the development of data visualization and some essential features in data visualization, followed by some opinions about classification of data visualization. There is also has some analysis about the standards when designing the data visualization.

Chapter 3: Describe the methodology in the research. It gives an explanation about how to select the methodology and how to design the questionnaire in the survey. There is an explanation about the method to collect the responses. It also lists an ethical aspect to ensure the information security of participants.

Chapter 4: Present the results of the survey, including the attitude of essential features and some performance of different types of data visualization. The chapter adds some findings based on the survey.

Chapter 5: Discuss the findings from Chapter 4 and compare the five selected types of data visualization. Verify the participants’ attitude towards some essential features in data visualization and analyse the importance of each feature. Give an exploration to explain the performance of different data visualization. Describe the relationship between the data visualization and infographics and take a suggestion to integrate the separate data visualization as a whole to tell the story by using infographic.

Chapter 6: Summary the discussion of the research and give an explanation about the research aims and some objectives. In addition, it also suggests some recommendation for the future research.
2. Literature Review

2.1 Introduction

This chapter reviews some relate literatures about the data visualization. It has four parts. Firstly, it describes the development of data visualization, followed by an explanation about some essential features which people may need to consider when they design the data visualization. The next part is some opinions about the classification to the data visualization. Finally, there is a summary about the standards about the design of the data visualization.

2.2 Development of Data Visualization

Data visualization has a long history in its development. Few (2007) describes a timeline about the changes of data visualization. The data visualization has a widely development in the last 30 years. Since the computer can interact with the researchers to analyse the data, more data visualization are created. Nowadays, data visualization is always used in the business intelligence. Healy (2014) explains that the data visualization becomes the centre of the social scientific work. He confirms the plausibility of the data visualization and treat it as an effectively way to use data visualization to integrate the data sources. Few also show that the data visualization is a better way to help people communicate with others by using the graphs or diagrams. However, the data visualization has an incorrect trend that it concerns more about the display of different data and ignores the meanings of the information. Another field of the data visualization is the use of geo-spatial visualization. It based on the popularity of google earth and some other map services. The data visualization can combine the data with the real world to express some relevant information to the users. It is also useful when building a powerful sales web for business. In the future, the data visualization will appear more in displaying the relationship between items rather than showing the values of variables (Few, 2007).

2.3 Essential Features of Data Visualization

There are 10 features which relevant to the design and understanding of data visualization. They are format, chart type, dataset, data source, axis, scale, legend, variables, outliers and input area. Format has two different models, the interactive visualization and the static visualization. O'Donnell & David (2000) identify the select of format may affect the presentation of the data visualization. Dilla & Steinbart (2005) give an opinion that the suitable format to express data can help readers focus on more related information about
the data visualization. Chart type can be treated as a necessary part which the designers should consider before making the data visualization (Spath, 2007). Generally, the chart type includes some basic kinds such as the bar chart, line chart, or pie chart. Meanwhile, it can also combine some different types to make the data visualization more vivid. The dataset is used to store the data which will be used in the data visualization. The data in dataset always exist in the status of database or the spreadsheet. The lines and the rows would help the designers select a suitable variable or scale. Data source is usually shown in the data visualization. Through the data source, readers can realize where the data of the visualization from, it is helpful to improve the reliability of the chart (Ledesma, 2016).

Axis and scale are appears together, the axis can provide the readers a place to read the value of the variables and the scale is always show the specific values by the marks. Legend is like an explanation which help the reader identify each items of the data visualization. It may appear in the side of the data visualization (Dykes, 2010). When using the data visualization to express the information, some items may change the names in order to make the data visualization more consistent. The variables may be used to replace the names. If the dataset has some uncommon data which is difficult to appear in the data visualization, there will add an outliers to show that data (Gunisetti, 2011). Some data visualization may have an input area which allows the readers to search for an interesting part in the data visualization. In order to catch the main points of data visualization, five items are necessary. They are the key, data, data source, scale and title. All of the five items are listed in the data visualization and readers can understand the overall information about the data visualization.

2.4 Classification of Data Visualization

People may create the data visualization for different communication purposes. In this consideration, data visualization can be divided into five types: comparing categories, assessing hierarchies and part-to-whole relationships, showing changes over time, plotting connections and relationships, mapping geo-special data (Kirk, 2012). Comparing categories are always used to compare some related items and show the difference between these items. The bar chart may be a convenience method to describe this kind of situation. It has some common types such as the column chart, floating bar and pixelated bar chart. Except the bar chart, dot plot and histogram are also popular in comparing related items. Assessing hierarchies and part-to-whole relationships are used to describe the percentages or relationships in one or more items. The pie chart may be common in this type, including the formal pie chart and some special format such as the
square pie, tree map and cycle packing diagram. In addition, bubble hierarchy and tree hierarchy are important as useful parts of the assessing hierarchies and part-to-whole relationships. If the designers consider describing the changes of the items value with the time passing, they may be suggested to use the line chart area chart or horizon chart. For other types of data visualization, the mapping geo-spatial data would provide an intuitive way to help people recognise the distribution of some items. This kind of data visualization includes the choropleth map, dot plot map, bubble plot map and isarithmic map. There is also exit a kind of network connection map to describe not only the distributions but also the relationships of for elements.

Segel & Heer (2010) describe the data visualization in five basic categories: Time series or temporal data, Statistical distributions, Geo-spatial data, Hierarchies and Networks. Time series or temporal data is used to show the changes for item value with the time passing. It may draw the graphs with a huge number of data in different time. By comparing the data, reader may conclude the trend of the item. Some stacked graphs may help readers analyze the aggregate situation of data. People can read an aggregate pattern for different value and identify each trend of the elements.

If the data visualization intends to express the distribution of items, it may need to create a statistical visualization. This kind of visualization requires the designer to have a data analysis towards the statistical data. Stem-and-Leaf Plots is always used to show the common distribution of statistical data. It uses the numerical variables to describe each level of the data. Q-Q Plots seems to focus on two types of most important data. It is used to show the relevance between these two probability distributions. The central diagonal may be used when the two selected data would have linearly relation with each other. Parallel coordinates may use different kinds of methods to show the multivariate data. Unlike the other data visualization mentioned, Parallel coordinates can express many different variables in one time.

Maps are always considered as a meaningful method to help people visualize the data. At the beginning, the map is used to show the geographical situation. In addition, it can offer people the rich information base on the image. Flow map is based on a real geographical map and use some flow lines to express the process. Readers can realize more details through analyzing the flow line in the map. Choropleth map can be created to reflect the status in one or more areas (Crampton, 2004). To make the status more detailed, the map may also need to list some statistical data. Furthermore, all of the statistical data may be
transfered into normalized values with a standard to create a density map. Graduated symbol map can be treated as an alternative type of the choropleth map. It uses the labels to show the density of some related situations. This type may reduce the garble problems in geographical maps. The designers may also add extra information on the label to make the label spread more dimensions in one part. Cartogram are designed to encode a variable in one area, it always neglect the geographical situations. In this map, the data can be shown directly by simple symbol.

The hierarchies can be used to show the data in different levels. This type of data visualization can express the results more clearly by defining several levels of the data. In the situation of measuring a huge number of data, the hierarchies may have the advantages. Furthermore, the node-link diagrams may use the interchangeably with hierarchy to achieve the visualization. Usually, the node-link diagrams are related to the tree-layout algorithms with some algorithms to identify the different levels of data. Adjacency Diagrams is designed as a cycle which put the highest level of data in the centre and the lower levels of data will be arranged around the higher one. Card (2009) gives a description about the enclosure diagrams. The tree map can be transferred into the rectangles. This data visualization may show the node clearly and the readers can recognise the size of the node easily.

When the data visualization aims to express the relationship between the items, the network may be a suitable method. This kind of data visualization has three types which are force-directed layouts, arc diagrams and matrix views. The force-directed layouts intend to create a physical system to show the relationship. In this system, the items which have a strong relationship with others will be arranged in the centre and the lines connected with others are designed in bond. Meanwhile, some items which have weak connection with others may be putted at the edge. This kind of network can help readers identify and understand the main relationships from amount of data. Arc diagrams are designed on a one-dimensional line and all of the elements are added in the line. Some relevant items are connected with each other via the arc. Some items with strong relationship will have more arcs in the line so that this part will appears a crowded situation of the arcs. Matrix views use the colours or saturations to express the relationship between the nodes. This method may provide a deeper structure for the network.

2.5 Standards in Data Visualization Design

Fry (2008) describes seven main stages for designing the data visualization. Firstly, the designers need to
identify the useful data and collect the needed data from some data sources. There is no limitation for the type of the collected data. Then the designers may parse the collected and classify the data into different categories. For example, the data can be sorted in different structures or orders. It may need a filter to select the data which will be used in the data visualization and the selected data would be stored. Some suitable patterns may be chosen by reorganization. In the process of data presenting, the designers need to compare some basic visual models and choose a more suitable one to express the data. In order to improve the performance of the data visualization, designers may refine the result. Finally, it may add some methods to measure the data according the interaction.

2.5.1 Unsolved Information Visualization Problems

Chen (2005) defines some unsolved information visualization problems and these problems are divided into different levels. When focus on the users, it may need to ensure the usability of the data visualization. People always see an original data or the visualized graphs but they know little about the process of the visualization. It may require the develop tools select an open source or high free system (Chen, 2006). Another problem is the understanding the basic perceptual-cognitive task. Some people may have a deep effect by the information retrieval and misunderstand the target of data visualization. The education or the training can be a challenge for the users to understand the meanings of data visualization. It may need the researchers to share the principles or technologies of data visualization. In the technical level, the data visualization may face the problems include the quality measures, scalability and the aesthetics.

2.5.2 Measures to do in Design

Clayton (2014) suggests a squint test in the design of data visualization. The test is focus on the integrity of the data and test the data visualization can cover all of the information. In some condition, the visualization may need to be printed out to check the visibility. Dando (2014) gives the opinion that designers need to keep an objective attitude to express the data so that the readers may not affect by the designers’ views. In addition, the designers need to consider the organization of the data. It can help the readers understand the relationship between the variables. Dando also suggest controlling the frequency of the colours. Even though the colour can make the data visualization seems more vivid, too many colour may disturb the readers reading the graphs.
2.5.3 Measures not to do in Design

Tufte (2001) stresses the graphical integrity when design or create the data visualization. The designers are abandoned to use any fake information which the dataset is not included. All of the data in the data visualization should have a real source. While the technology has a huge development to create some complex pictures, Few (2012) suggests reducing the 3D effect for the data visualization. This opinion has the explanation that the effect of some new technology may make the readers feels difficult to understand the information. Wong (2010) limits some combination of colours which makes the graph unreadable. For example, the red may become more dazzle when it is putted near the green.
3. Methodology

3.1 Introduction
This chapter describes the main methodology used in the study. There are four sections which are the research approach, data collection method, ethical aspects and limitations of the methodology.

3.2 Research Approach
The dissertation selects the quantitative approach achieve the evaluation for different types of data visualization. Quantitative research is generally used to acquire some related explanations to some opinions or some phenomenon from people or special groups. It also can be used to collect some predictions from participants (Glesne & Peshkin, 1992). In order to reach a generalizable result, the research needs to create a careful sampling strategy and experimental design. During the research, the researcher may observe and measure the process as an objective role.

Bryman (2015) explains the quantitative research has the main scopes. Firstly, the quantitative research can be treated as a deductive approach. It may be used in testing some relationship between theories and researches. Secondly, in epistemological orientation, the research can combine the natural scientific model and positivism such as the practices and norms.

Generally, the quantitative research has eleven main processes: Theory, Hypothesis, research design, devise measure of concepts, select research sites, select research subjects, collect data, process data, analyse data, findings, conclusions. The dissertation aims to collect the responses of participants towards the different types of data visualization so that it will summarize participants’ felling when they facing these kinds of data visualization instead of set a hypothesis at the beginning.

Thomas (2003) describes three types of quantitative method which are surveys, correlation analyses and experiments. Survey methods focus on several related variables. It will produce a summary about the results towards the collected data. Correlation analyses aims to find the relationship between two variables. It is usually being designed as exploring the changes when one variable affect to another, sometimes, the extent of the changes may also be considered. The experiments need to create an object according to the definition, then, it may test the outcome of object to find the influence by treatment. As a result, the
experiment method would give an explanation about the effect.

3.3 Data Collection Method

3.3.1 Online Survey Questionnaire

During the evaluation of different types of data visualization, it may have more attention on the responses of participants on data presenting, data quality. The survey method may have more convenient to achieve the requirement. It has two methods: the structured interview and the online self-administered questionnaire. Comparing these two methods, the structured interview method has more disadvantages which may not applicable to the research. Firstly, the structured interview requires the researchers to communicate with the participants face to face or via phone and record the whole interview which may increase the time cost. Secondly, the structure interview needs to record all of the information during the interview. However, the research may only use a few part of the data, it may produce some mistakes in the data translation. Thirdly, the collected data of structure interview is based on the communication between interviewer and interviewee. The interviewee may be affected by the interviewer and gives the choice without objective feelings. This condition may influence the analysis results. (Groves, 2009)

3.3.2 Question Design

The questions of the research are designed into two sections. The first part has some basic questions which include participants’ general information such as gender and study level. It also lists the multiple-choice grids about data presenting and data quality so that participants can select the importance of some main features in data visualization. As there are many elements in data visualization, the questionnaire will select four features to test the importance of data presenting and three features to test the importance of data quality. Four presenting data features include the chart type, scale, legend and colour. Three data quality features include the title, data source and variables. All of these questions in this part aims to help participants has a brief understanding of the research. In addition, the results of first part may reveal a general expression of participants’ feeling toward data visualization. In the second part are the questions about different basic data visualization such as the bar chart, line chart, pie chart, bubble chart and map. In this part, participants are invited to answer same question for each type of data visualization. The questions include the frequent of participants’ encounter / create of data visualization, the participants’ attitude towards the data types such as their feelings to data types in the chart or the style of the chart.
The collected data would be used in the evaluation of these types of data visualization. In the part two, there also has a textbox which allow participants to enter there extra comments about that type of data visualization. If participants have some suggestions for the type of data visualization, they can enter their word into the box.

3.3.3 Participants

The research intends to find some students of Sheffield University to take participant in the data collection. This design is based on two considerations. Firstly, the students may have more opportunities to encounter different kinds of the data visualization, especially some basic types such as bar chart and line chart. They may have an initial understanding of these data visualization so that they would give a reasonable response of the questionnaire. Secondly, the volunteer system of Sheffield University may be helpful to spread the questionnaire to students ensure the research can acquire enough responses in time. In addition, there is no limitation for different kinds of students such as undergraduate or postgraduate, science or non-science. The online questionnaire aims to collect nearly 50 responses.

3.4 Ethical Aspects

The research also considers the ethical aspect. When participants take part in the research, they may need to input some personal information such as the ages and the income. Meanwhile, the IP address may be recorded when they doing an online survey. Diener (1978) gives some describes of ethical issues such as the stress, physical harm, interference of development. Holmes (2004) suggests some methods to protect participants’ data include escaping record some addresses or names of participants, code the data files.

In this research, the ethical aspect can be defined as ‘low risk’. All of the questions are designed as insensitive to participants. Participants may be suggested to complete some basic questions such as select their gender and study level, but they are not required to enter their name or address. In addition, when browsing a type of data visualization, their feelings toward different types of data visualization have no harm to themselves. The questionnaire is created in google form, and the IP address of participants will not recorded to researchers.

3.5 Limitations of Methodology
There are some limitations of this methodology. Firstly, because of the time limitation, the sample size of the questionnaire is designed as 50 responses which may not be large for an overall analysis. Secondly, the research takes a self-administered questionnaire to collect, when participants have some confusion about the question, there is no further prompts. This may occur a misunderstanding and affect the participants’ judgment. Finally, once the participants have some extra responses, they may feel hard to express in the questionnaire, this may miss some salient results.
4. Results and Findings

4.1 Introduction

The research receives 60 responses from the students in Sheffield University. This chapter describes the results of the questionnaire which collects the responses of participants towards different types of data visualization. Firstly, there is an overall statistics about participants’ basic information, followed with a description of participants’ attitudes about presenting data and data quality in data visualization. For different types of data visualization, there are independence explanations to show participants’ evaluation. Finally, there are some correlation descriptions to explore how participants’ attitudes may influence the evaluation of different types of data visualization.

4.2 Participants Information

There are some basic information graphs about the participants in the research. In order to make the percentages of each element more clearly, it use a pie chart to express the results.

![Pie chart showing gender distribution](image)

**Figure 1: Participant Gender Group**
In the 60 responses, 43.3% (26) are male students while 56.7% (34) are female students. Most of the participants are postgraduate students which occupy a percentage of 76.7%; the other 23.3% are undergraduate students. There are 70% science students with a number of 42 and the major of non-science has 18 students.
Figure 4: Participant Native Language Situation

The questionnaire also collects the native language situation of the participants. In the results, there are 27 participants whose native language is English. Because the questions in the questionnaire are designed in English, the English speaking participants may understand the meaning of the questions better. In the other 55% participants, 23 participants’ native language is Chinese, 3 participants’ native language is Hindi, 3 Malay speakers. Indonesian, Spanish, Russian, and Croatian have 1 participant respectively. The relationship of the language and understanding of each type of data visualization will be tested later.

4.3 Basic Attitude towards Data Visualization

The overall attitude includes two parts which are the attitude towards presenting data and the attitude towards data quality in data visualization. For both of the parts, there is a reliability analysis to test the reliability of the results. As the description of Field (2013), the higher reliability of the questionnaire would be helpful to reduce the measure error of the research and bring the confidence to the researcher. In SPSS, the reliability is tested with a Cronbach’s alpha value. Bland & Altman (1997) show that if the value is between 0.7 and 0.8, it may need to revise few questions, if the value is higher than 0.8, the value is an acceptable value in the research. For this questionnaire, there are seven items which test participants’ attitude about the importance of different features of data visualization.

<table>
<thead>
<tr>
<th>Reliability Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s Alpha</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 1: Reliability Analysis

In the form of reliability analysis, the Cronbach’s alpha of the seven items is 0.833, which is higher than 0.8. It can achieve a condition that the collected data can be treated as an acceptable result.

4.3.1 Presenting Data

The questionnaire firstly collects the participants’ attitudes towards presenting data in data visualization from four features. When combine the four features together, the situation of each feature is listed in follows.

![Figure 5: Importance Score Situation of Presenting Data](image)

This graph shows the detailed score situation of each features in presenting data. For the importance of chart type, 1 participant (1.7%) marks 1 score which means the participant considers the chart type is unimportant for the data visualization to presenting data. 3 participants (5%) mark the chart type's importance with 2 score and 8 participants (13.3%) choose 3 score. There are 48 participants (80%) choose the score higher than 3, including the 21 participants (35%) select 4 score and 27 (45%) select 5 score. The trend of scale is similar with chart type. From score 1 to score 5, the participants are 1 (1.7%), 5 (8.3%), 12 (20%), 20 (33.3%) and 22 (36.7%) respectively. Most of participants treat the legend as an important
feature in presenting data, with a number of 30 which occupies a percentage of 50%. Colour seems is less important than others, many participants mark it with the score of 3 or 4. Only 13 participants (21.7%) treat the colour’s importance as 5 score (very important).

### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChartType.imp</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>4.17</td>
<td>.960</td>
</tr>
<tr>
<td>Scale.imp</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>3.95</td>
<td>1.032</td>
</tr>
<tr>
<td>Legend.imp</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>4.22</td>
<td>.958</td>
</tr>
<tr>
<td>Colour.imp</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>3.55</td>
<td>1.048</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Mean analysis of Presenting Data

Legend has the highest mean (4.22) of the four features, followed by the chart types with 4.17. Colour gets the lowest mean score. However, the questionnaire defines the score from 1 to 5 to show the importance of these four features. The mean which is higher than 3 refers to an important roles for participants. The result shows that participants in this survey consider all of these four features are necessary for data visualization to present data.

### 4.3.2 Data Quality

The questionnaire collects the participants' attitude towards data quality in data visualization from three features. In the results, the situation of each feature is listed in follows.
More participants are likely to select the data source as a most important part in data quality. Only 3 participants treat the data source as a little unimportant. No one select unimportant option. 25 participants (41.7%) believe the data source is very important in data quality of data visualization. When participants evaluate the importance of title, they generally select the less important option, for example, 16 participants (26.7%) select the score 3 for title, which is much higher than the other two features. Variables has same situation in the scores which are below 4, but in the higher score, more participants would prefer score 4.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title.imp</td>
<td>60</td>
<td>1</td>
<td>5</td>
<td>3.82</td>
<td>1.017</td>
</tr>
<tr>
<td>Source.imp</td>
<td>60</td>
<td>2</td>
<td>5</td>
<td>4.13</td>
<td>.892</td>
</tr>
<tr>
<td>Variables.imp</td>
<td>60</td>
<td>2</td>
<td>5</td>
<td>4.08</td>
<td>.869</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Mean analysis of Data Quality

Data source has the highest mean (4.13) of the three features, followed by the variables with 4.08. These
two features are similar to express hat both of them have similar roles in data quality of data visualization. Title gets the lowest mean score which is differ with which mentioned in the literature review. The result shows that participants in this survey consider all of these four features are necessary for data visualization to present data.

4.4 Results for Different Types of Data Visualization

The questionnaire collects responses for each kind of data visualization. There will be some descriptions for each type respectively.

4.4.1 Result for Bar Chart

There are two questions which collect the frequency situations of participants encounter and creation of the bar chart in their daily life. In order to make the results more recognizable, these two responses are listed in a graph.

![Figure 7: Encounter and Creation of Bar Chart](image)

All of the participants have encountered the bar chart before. Most of them encounters encounter this type of data visualization about one month a time, 16 participants (26.7%) show that they encounter bar chart weekly. Only 3 participants (5%) keep a daily encounter. When it comes to creation aspect, there are 4 participants (6.7%) who never create a bar chart. Unlike the status of encounter, 31 participants create
this chart yearly with a percentage of 51.7%.

The evaluation of bar chart is also divided into two parts. The first part describes the participants’ responses of presenting and understanding towards the bar chart. The second will show the chart style and data type.

<table>
<thead>
<tr>
<th>Bar.presenting</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Useless</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Useless</td>
<td>2</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Fair</td>
<td>10</td>
<td>16.7</td>
<td>20.0</td>
</tr>
<tr>
<td>Useful</td>
<td>35</td>
<td>58.3</td>
<td>78.3</td>
</tr>
<tr>
<td>Very Useful</td>
<td>13</td>
<td>21.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Presenting Data of Bar Chart

<table>
<thead>
<tr>
<th>Bar.understand</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Difficult</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Difficult</td>
<td>3</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Fair</td>
<td>12</td>
<td>20.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Easy</td>
<td>32</td>
<td>53.3</td>
<td>78.3</td>
</tr>
<tr>
<td>Very Easy</td>
<td>13</td>
<td>21.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Understanding of Bar Chart
Participants’ response in chart presenting is similar with their understanding situation. 35 (58.3%) participants consider the bar chart is useful to present the information to them, and as a result, 32 (53.3%) participants express that it is easy to help them understand this chart. Nobody thinks the bar chart is very useless in presenting information and nobody understands the bar chart very difficultly. For 13 participants (21.7%), the bar chart is very easy for them to understand.

![Figure 8: Style of Bar Chart](image1)

![Figure 9: Data Type of Bar Chart](image2)

The style of the bar chart seems to be normal for participants, 34 participants (56.7%) select fair when they
evaluate the chart style. Only 1 participant (1.7%) treats this kind of data visualization as a very interesting chart style. When considers the data type, more participants think it is simple or very simple. No one feels the data of this chart is complex.

In addition, there are some participants express their extra comments. One has mentioned that the bar chart is difficult to show a trend of some phenomenon. Someone also shows that the colour in the example is not suitable which strain the participants’ eyes.

4.4.2 Result for Line Chart

![Line Chart Frequency and Percentage](image)

Even though most participants have seen the line chart in their daily life (only 1 participant has not encountered line chart before), there are still many participants (23.3%) who have never created the type of chart. 44 participants (73.3%) encounter the line chart once or more times in a month. 23 participants (38.3) show that they create the line chart yearly.
Table 6: Presenting Data of Line Chart

<table>
<thead>
<tr>
<th>Line Understanding</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Useless</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Useless</td>
<td>2</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Fair</td>
<td>17</td>
<td>28.3</td>
<td>31.7</td>
</tr>
<tr>
<td>Useful</td>
<td>33</td>
<td>55.0</td>
<td>86.7</td>
</tr>
<tr>
<td>Very Useful</td>
<td>8</td>
<td>13.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Understanding of Line Chart

Like the performance of bar chart, few participants think the line chart is difficult to understand. 33 (55%) of the participants consider this type of data visualization is useful to present data while 25 (41.7%) feels easy to understand the line chart. There are 8 participants (13.3%) treat the line chart as a very useful tool to present data and the number of participants who can understand the line chart is 8 similarly.
34 participants (56.7%) think the data type of line chart is fair. For the style of the line chart, they seem to prefer an interesting option. No one consider that the data type is very simple while 2 treat this type as very complex. The participants who think the style of line chart is boring only occupy a percentage of 15%.

A participant shows the line chart is an “ideal for time-varying series” in the comments. Another participant complains there are too many clutter and gridlines in the example line chart.
4.4.3 Result for Pie Chart

Most of participants seem to encounter and create the pie chart monthly. Few participants show that they create the chart weekly or daily. However, 27 participants (45%) encounter the pie chart weekly which is more frequently than the creation. There are 3 participants have never encountered the pie chart.

<table>
<thead>
<tr>
<th>Pie.presenting</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Useless</td>
<td>1</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Useless</td>
<td>3</td>
<td>5.0</td>
<td>6.7</td>
</tr>
<tr>
<td>Fair</td>
<td>12</td>
<td>20.0</td>
<td>26.7</td>
</tr>
<tr>
<td>Useful</td>
<td>37</td>
<td>61.6</td>
<td>88.3</td>
</tr>
<tr>
<td>Very Useful</td>
<td>7</td>
<td>11.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Presenting Data of Pie Chart
### Table 9: Understanding of Pie Chart

<table>
<thead>
<tr>
<th>Attitude</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Difficult</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Difficult</td>
<td>3</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Fair</td>
<td>15</td>
<td>25.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Easy</td>
<td>25</td>
<td>41.7</td>
<td>71.7</td>
</tr>
<tr>
<td>Very Easy</td>
<td>17</td>
<td>28.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

The attitude of participants towards pie chart on presenting is positive. 37 participants (61.7%) think pie chart is useful to present information and 7 participants (11.7%) treat it as very useful. Only 4 participants select the option of “Useless” or “Very Useless”. In this condition, 17 participants (28.3%) can understand the chart very easily. No one has the very difficult feelings in understanding.

![Pie.style](image)

**Figure 14: Style of Pie Chart**
Figure 15: Data Type of Pie Chart

8 participants (13.3%) think the style of pie chart is boring and 23 participants (38.3%) hold an opinion of fair. This status may be influenced by their attitude to the data type in this pie chart. 42 participants (70%) feel the data type of this chart is simple or very simple. It may reduce their interest on that kind of data visualization.

The participants suggest that the pie chart may perform better on presenting some percentages but not suitable for quantity comparison. The reader may face some difficult to read this type of chart because they cannot compare the size of each slices. A participant also points that the number record in the example pie chart should be changed into a total statistics.

4.4.4 Result for Bubble Chart
The bubble chart appears rarely in people’s daily life. There are 22 participants (36.7%) who have never encountered the bubble chart. 23 participants (38.3%) show that they encounter the bubble chart yearly. Only 1 participant reads this kind of chart every day. Almost 80% participants do not have any experience of creating bubble chart. The number of participant who creates the bubble chart daily is only one. The participants who keep a weekly encounter or creation have a number below 5 repetitively.

<table>
<thead>
<tr>
<th>Bubble.presenting</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Useless</td>
<td>1</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Useless</td>
<td>22</td>
<td>36.7</td>
<td>38.4</td>
</tr>
<tr>
<td>Fair</td>
<td>22</td>
<td>36.7</td>
<td>75.1</td>
</tr>
<tr>
<td>Useful</td>
<td>13</td>
<td>21.7</td>
<td>96.8</td>
</tr>
<tr>
<td>Very Useful</td>
<td>2</td>
<td>3.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Presenting Data of Bubble Chart
Table 11: Understanding of Bubble Chart

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Difficult</td>
<td>9</td>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Difficult</td>
<td>27</td>
<td>45.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Fair</td>
<td>17</td>
<td>28.3</td>
<td>88.3</td>
</tr>
<tr>
<td>Easy</td>
<td>5</td>
<td>8.3</td>
<td>96.6</td>
</tr>
<tr>
<td>Very Easy</td>
<td>2</td>
<td>3.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Unlike the other three types of data visualization above, the bubble chart seems difficult for participants to understand. Only 7 participants (11.7%) explain that they can understand the bubble chart easily. Instead, the number of participants who feel difficult or very difficult to understand the bubble chart is 36. Participants’ attitude of presenting may be affected by the difficult understanding. 22 participants (36.7%) believe that it is useless to choose this type for presenting information. Another 22 participants hold a fair view.

![Figure 17: Style of Bubble Chart](image-url)
28 participants (46.7%) feel the data type of bubble chart is complex. Nobody treat it as a very simple feature. However, participants are likely to accept this style of bubble chart. 25 participants (41.7%) give their choice that the style is interesting, which appears a different phenomenon compared with the performance in data type.

In the comments, many participants express their confused about the bubble chart. They complain that they do not understand the meaning of each bubble in the graph. The different size of bubble may need more explanation. Someone have no sense about the intersection between the bubbles.

4.4.5 Result for Map
Using map to visualize the data has two different performances in encounter and creation. Participants have a high frequency to encounter the map. 7 participants (11.7%) keep a daily encounter. There are 21 participants (35%) face this type of chart monthly. In addition, only 1 participant has not encountered the map before. However, most of the participants do not have the experience of creating map. 11 participants (18.3%) create the map monthly, which is nearly of the encounter.

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Useless</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Useless</td>
<td>2</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Fair</td>
<td>16</td>
<td>26.7</td>
<td>30.0</td>
</tr>
<tr>
<td>Useful</td>
<td>27</td>
<td>45.0</td>
<td>75.0</td>
</tr>
<tr>
<td>Very Useful</td>
<td>15</td>
<td>25.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Presenting Data of Map
### Table 13: Understanding of Map

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Difficult</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Difficult</td>
<td>7</td>
<td>11.7</td>
<td>11.7</td>
</tr>
<tr>
<td>Fair</td>
<td>11</td>
<td>18.3</td>
<td>20.0</td>
</tr>
<tr>
<td>Easy</td>
<td>29</td>
<td>48.3</td>
<td>68.3</td>
</tr>
<tr>
<td>Very Easy</td>
<td>13</td>
<td>21.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Though the frequency of encounter and creation in map is quite differ, the performances of understanding and presenting are similar. Participants think this kind of data visualization is easy to realize and it also useful to express the information. 15 participants (25%) even mark this graph is very useful presenting tools and 13 participants (21.7%) feels very easy to understand.
Figure 21: Data Type of Map

36 participants (60%) are interested in the style of map and 9 participants (15%) think it is very interesting. Most participants hold a fair opinion for the data type. The numbers of very complex and very simple feelings to the data type are 1 and 2 respectively.

Participants have a positive attitude to this type of data visualization. Someone suggests that the map may be interactive so that reader can zoom some areas to find more information.

4.5 Chi-Squared for Language

The questionnaire is designed in English, so it may cause some misunderstanding to those participants whose native language is not English. In the Chi-square part, it will test if different native language speakers may have different performance in the charts understanding.

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Value</th>
<th>Df</th>
<th>Asymp.Sig. (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar Chart Pearson Chi-Square</td>
<td>23.160(^a)</td>
<td>21</td>
<td>.335</td>
</tr>
<tr>
<td>Line Chart Pearson Chi-Square</td>
<td>31.263(^a)</td>
<td>21</td>
<td>.070</td>
</tr>
</tbody>
</table>
Agresti & Kateri (2011) describe if the Asymp.Sig is higher than 0.05, it means there is no relations between the two variables. In the Chi-square test, the first three types of data visualization have the Asymp.Sig results which are 0.335, 0.07 and 0.707 respectively. It can get the explanation that the different languages have no influence on the understanding of bar chart, line chart and pie chart. For the last two which are lower than 0.05, their minimum expected counts are quite little (0.03 and 0.12), so the bubble chart and map do not have the relation either. According to the analysis, it can get the results that the language of data visualization will not affect the participants’ understanding in the research. The evaluation result is equal even though it is use single language to collect the responses.

<table>
<thead>
<tr>
<th>Data Visualization</th>
<th>Chi-Square Value</th>
<th>Degrees of Freedom</th>
<th>Asymp.Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pie Chart Pearson Chi-Square</td>
<td>17.073&lt;sup&gt;a&lt;/sup&gt;</td>
<td>21</td>
<td>.707</td>
</tr>
<tr>
<td>Bubble Chart Pearson Chi-Square</td>
<td>61.133&lt;sup&gt;a&lt;/sup&gt;</td>
<td>28</td>
<td>.000</td>
</tr>
<tr>
<td>Map Pearson Chi-Square</td>
<td>33.919&lt;sup&gt;a&lt;/sup&gt;</td>
<td>21</td>
<td>.037</td>
</tr>
</tbody>
</table>

Table 14: Chi-Square for Five Types of Data Visualization Understanding
5. Discussion

5.1 Introductions

This chapter will verify the participants’ attitude towards some essential features in data visualization and analyse the importance of each feature. Based on the responses of participants, different types of data visualization will be evaluated and compared. According to the evaluation and comparison, there will have an exploration to explain this situation. Focusing on the drawbacks of these types of data visualization, it gives a suggestion about the design. In addition, it describes the necessity of some combined data visualization in presenting information. Finally, it describes the relationship between the data visualization and infographics, followed by a suggestion to integrate the separate data visualization as a whole to tell the story by using infographic.

5.2 Importance of Features in Data Visualization

As the literature review mentioned before, some essential features is helpful for not only the readers who encounter the data visualization but also the designers who create the data visualization. The features are still being divided into two groups, the presenting data and the data quality. The importance of each feature in data visualization can be verified in groups.

5.2.1 The Features in Presenting Data

When considering the features in data presentation, the chat type, format, scale, legend, axis and colour can be all included. In the research, it selects four of them which are chart type, scale, legend and colour. As a result, all of the four features appear the important roles to participants with a score which is higher than 3. Jeffrey, Michael & Vadim (2010) suggest that the organization of chart types should base on the primary methods of data expression. They also list some typical chart types to cover users’ requirements. In the research, the importance of chart type acquires a score of 4.17 which means the participants think this feature is important for presenting data. Another feature whose score is higher than 4 is the legend. The legend seems to be a further explanation for the data visualization. While some data visualization uses different shapes or sizes to express the data, an extra legend would help reader identify the meanings of these values. When design the legend, it usually rely on the designers’ subjective attitude (Dykes, Wood & Slingsby, 2010). However, a current legend would help the readers realize the main points of the data visualization and this condition is verified in the survey. The importance of scale gets 3.95 which is nearly
the important level. As the scale shows the range of values, it may seem less important than the other two above. Meanwhile, some data visualization may have no scales. This condition is more likely to result the lower score in the research. The last part of the four features is the colour with a score of 3.55. This situation may show participants would care less about the colour than other features in data visualization. It can give a suggestion to the designers that they do not need to consider more about the colour of the data visualization. Instead, they may concern on the other elements in visualization such as the selection of chart types.

5.2.2 The Features in Data Quality

Data quality is treated as another essential part to help people understand the data visualization. For the three features in the survey, both of data source and variable get the score higher than 4. The importance of title is satisfied in 3.82. When encounter data visualization, people may concern more about the data source at beginning, it can help them identify the credibility of this kind of data visualization (Soukup & Davidson, 2002). It means a reliable data source in the data visualization would increase the acceptance of the readers. Variable is usually used to replace the name of the items in data visualization. As participants mark the importance of the variables with 4.08, it may suggest the designers to consider a suitable chart to express the relationship between the variables. The title of the data visualization seems have less concern in this survey. According to the responses, participants complain that some abbreviations make them confused about the information in the data visualization and they may misunderstand the meaning in the chart. This survey shows an improvement in the data visualization that the designers may not use the Abbreviations in the title or give a further explanation for the data visualization.

5.3 Evaluation of Different Data Visualization

The survey aims to evaluate different types of data visualization. From the results of online survey, it can acquire different responses of these types of data visualization. By comparing their performance, it will discuss the similar or difference between each kind of data visualization and explore the reasons what occur the situation.

The encounter frequencies of bar chart and line chart have similar situations. Both of the two charts has a widely encounter situation in people and many people have the experience to create these charts. Dando
(2014) describes the bar chart is usually used to express different size or importance of the variables by equal weight bars. If the variable is important or has a large value, the height of the bar would be higher. Sometimes, bar chart can also show a comparison percentage for different opinions in one question. This condition mainly appear in the questions which have two opposite opinions, the bar is divided into two parts with different colour. Readers can realize the percentage of each opinion and identify the high percentage one easily. The line chart is used to show the trend of one item which is related the time. The chart does not need to start at origin place which is different with bar chart (Wong, 2013). The style of bar chart and line chart are described as fair. It means that these two charts may difficult to attract the readers to concern. To solve the problem, designers may add some elements to improve the attitudes of readers. In order to express the change of a single quantity or a set of quantities over time, the bar chart can be combined with the line chart, by using the line to show a trend and the bar to show the values, readers can have a specific understanding (Hardin, 2012). In addition, both of the bar chart and line chart are easy to be understand by people, this condition ensure the combination will not increase the realize difficulty.

The bubble chart may have some difficulties to be understanding by readers. This status also appears in the frequency of encounter and creation. In the survey, few people show they have created the bubble chart before. While this chart use different sizes of bubbles to show the different values of variables, people may feel confused because there is no specific scale. (Fortino, Alenius & Greco, 2015) In the survey of the verify process, participants give a high score for the importance of scale. It may be able to explain the phenomenon that the unsuitable scale causes a confused on the understanding. It suggests that the designers may need to give a detailed scale to help the readers realize the meanings of each bubble in the data visualization. In addition, the style of bubble chart is interesting to people, it shows a fact that people is likely to encounter this kind of data visualization once the data is more clearly.

The pie chart is easy to understand for the readers. In the research, most participants show the data type in the pie chart is very simple. Lown (2012) suggests that the pie chart need to be designed more sparingly because readers may feel difficult to distinguish the difference between each part if the sizes are similar. Wong (2010) also list some other disadvantages about the pie chart that it always occupy a huge space and difficult to be labelled. As many participants think the data type of the pie chart is complex, there may need an explanation for some key data.
The map seems to be more interesting for readers. Usually, the map is based on the real geography map in the world (Sarma, 2013). In order to express the different situations of each area, the designers can set the colour from the light to deep to show the changes from low level to high level. The readers can understand the meanings better if they have an overview about the geography map. However, if the map is related to an incorrect place or the readers are unfamiliar to the map, the map may perform worse in presenting data. Meanwhile, the readers may only realize the values but know little about their further information such as the relationship between each area. As a result, the map still needs to add a description of the pictures to help readers acquire the needed information. When design the map, designers are also need to check the elements of this format. In addition, the data type of the map seems a little simple for readers, and it gives a suggestion that the map can include the bar chart or the bubble chart (Helouvry, 2015) to make the values more obviously. In this condition, the bars or bubbles of different heights or sizes would express the different values.

5.4 Infographic and Data Visualization

Helouvry (2015) defines the infographic is a common data visualization. According to the definition, infographic is described as using the textual and highlight key pictures to tell a story to the readers. However, Krum (2013) holds an opposite opinion that the infographic is another communication tools that combine the data visualization together to tell the complete story. In the research, all of the types of data visualization are designed to show one kind of data. Meanwhile, the data visualization prefers to be shown in one space. Both of these situations are different to the attitude of infographic. It is more realizable to define the data visualization can be considered as part of the infographic. Compared with list the data, people are more likely to accept an infographic which is using a story to communicate with readers.

While using the infographic to tell the story to people, there may have some risks. That may make the infographic to be a poor design. The poor design may appear as follows: visualize the data in an incorrect method, fail to express the essential information, make the readers feels confused with the graphs, express in an informal style. To escape these risks, some improvements can be used. Firstly, the huge numbers of data visualization may make the misunderstanding to readers that they do not confirm which one is the essential graph. So the infographic may control the frequency of data visualization. Secondly, while the infographic express the main information for the readers, it may loss some relevant information. This requires the designers have an overall view about the data set and select suitable ones to achieve the goal.
Finally, the infographic may need to add some explanation for the data visualization so that the style should be designed as suitable for the subjective.
6. Conclusion

6.1 Summary of the Research

When people analyse the data visualization, they usually concern more about the design tools or some complex data visualization format. The research mainly focuses on the specific types of data visualization such as the bar chart and line chart. In order to make the research more acceptable, some features of data visualization are also verified. These features are divided into two aspects: the data presenting includes the chart type, scale, legend and colour; the data quality includes title, data source and variable. An online questionnaire is designed to collect the responses from participants. It has two parts, one is the attitude survey to investigate the importance of each features in participants’ opinions; the other is to collect the evaluation of each type of data visualization.

As a result, all of the four features in the data presenting appears their importance in data visualization. When ranking these features in score, the order is legend, chart type, scale and colour. It can considered that the legend seems to be the most important element in data presenting, it also suggest the designers to add the legend for their data visualization to help readers realize the meanings of each items in the graphs. The other three features in data quality are all important and the data source has the highest score. It means the readers would concern more about the sources so that the designers may need to indicate the source of the data to improve the realizable of the data visualization.

Bar chart and line chart have a widely used in people. This situation is also appears in the creation aspect. The data type of these two charts are simple, however, most people think the style of the tow chart are fair. It gives the suggestion that designer may need to make the charts more vivid. The pie chart performs well but occupies more space than other charts. In order to solve this problem, the pie chart can include some other information to enrich the meanings. Bubble chart seems difficult to people, few people have encountered this type of data visualization and few have created it. It requires the designers add more explanations to help readers understand the chart. Map expresses the data successfully but the data type looks simple. An improvement may combine the map with bar chart or bubble chart to expand the data types.

Infographic can be treated as a new format which includes the data visualization and any other expression
models. While using many types of data visualization to tell the story, infographic may need to control the numbers of charts but cover all of the requirements to express the useful information.

### 6.2 Limitation and Recommendation

There are three limitation of the research. As it mentioned in the methodology chapter, the research decides to collect 50 responses and only collect a numbers of 60. While the research aims to collect the response of public, it may needs to have a huge of the participants to express their feelings. This number of collections is not able to reflect the general reactions to the data visualization. In addition, in order to acquire the response in a short time, the participants are defined as the students in Sheffield University. Generally, the students group is just one part of the public and it cannot show the possibility of other groups. It makes the research limited to evaluate the students’ responses and may ignore the situations in other groups such as the workers in different occupations. The last limitation is the selection of features. The research only focuses on some main elements in data visualization and verifies the importance of the selected features. It may ignore some potential factors which are also important to data visualization. In the analysis of the results, the lack of features may make some situation difficult to be explained.

For further similar research, some recommendations are listed to improve reasonability of the survey. Firstly, the further may expand the sample size to ensure the collected responses have a high quality. Meanwhile, considering the difficult in research public’s attitudes, the further research may focus on one group such as the students. More features can be added in the verify process that researchers can have an overall realization about the elements which would influence the performance of data visualization. The further research may set some hypothesis to give an explanation about the relationship between the features and the performance of each type of data visualization.

**Word Count: 11023**
Reference


Hardin, M., Hom, D., Perez, R., & Williams, L. (2012). Which chart or graph is right for you?. *Tell Impactful Stories with Data*. Tableau Software.


Appendix

Appendix A: Ethics application

Application 010491

Section A: Applicant details

Created:
Thu 30 June 2016 at 16:09

First name:
Bin

Last name:
Shu

Email:
bshu1@sheffield.ac.uk

Programme name:
Information Management

Module name:
Dissertation
Last updated:
12/07/2016

Department:
Information School

Date application started:
Thu 30 June 2016 at 16:09

Applying as:
Undergraduate / Postgraduate taught

Research project title:
An Evaluation to Different Types of Data Visualization

Section B: Basic information

1. Supervisor(s)

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paula Goodale</td>
<td><a href="mailto:p.goodale@sheffield.ac.uk">p.goodale@sheffield.ac.uk</a></td>
</tr>
</tbody>
</table>
2: Proposed project duration

Proposed start date:
Thu 30 June 2016

Proposed end date:
Wed 31 August 2016

3: URMS number (where applicable)

URMS number
- not entered -

4: Suitability

Takes place outside UK?
No

Involves NHS?
No

Healthcare research?
No

ESRC funded?
No

Involves adults who lack the capacity to consent?
No

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial?
No

Social care research?
No

5: Vulnerabilities

Involves potentially vulnerable participants?
No

Involves potentially highly sensitive topics?
No

Section C: Summary of research
1. Aims & Objectives

The dissertation aims to collect people’s responses toward different types of data visualization, compare and analyse the advantages of each type of data visualization in expression meanings to people. Finally, the dissertation will give some suggestions for some data visualization types according to the evaluation results.

Objectives: to identify each type of data visualization, analyze the advantages and disadvantages of different kinds of data visualization, make a statistic of people’s responses toward each type of data visualization, raise an evaluation and give some suggestions to improve the performance of the data visualization.

2. Methodology

It will use a questionnaire online to measure the response of people to data visualization through professional survey website ‘questionpro’. It may requires people express their subjective feelings, therefore, the questions provides a range which is marked from 1 to 5 to show whether the people may accept the type of data visualization or understand the meanings of visualization. Then use the SPSS to analyse the collected data. According to the results it will give more comparisons and suggestions.

3. Personal Safety

Raises personal safety issues? No
Personal safety management
- not entered -

Section D: About the participants

1. Potential Participants

Potential participants will be the students enrolled at the University of Sheffield. There is no limitation about students’ subject or degree. All of students can suit the requirements of the questionnaires.

2. Recruiting Potential Participants

The potential participants will be sent an email via the student volunteers list by my supervisor

2.1 Advertising methods

Will the study be advertised using the volunteer lists for staff or students maintained by CICS? Yes
Even though the survey can be advertised by twitter or Facebook, it may needs a huge numbers of followers which is not suitable for me.
3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

The Information Sheet/Consent Form (attached) will be the first and second pages of the survey and will describe what will be asked of participants. An email will be supplied for further information if a participant has questions that are not answered on the Information Sheet.

4. Payment

Will financial/in kind payments be offered to participants? No

- not entered -

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

The risks of participating are the same as those experienced in everyday life.

How will this be managed to ensure appropriate protection and well-being of the participants?

During the design of the survey, it will be checked the questions to ensure the suitability of each question.

The responses collection will use the 'questionpro' which is more professional in each processes of the survey.

After the survey, all related information will be deleted.

Section E: About the data

1. Data Confidentiality Measures

Each survey response will be associated with an arbitrary number; all responses will be anonymous. Any IP addresses associated with the dataset will be deleted as soon as the file is downloaded from the web.

2. Data Storage

The data will be exported from Survey Questionpro and stored on the Information School’s research data drive which can be accessed by only by me, my supervisor, and the School’s Examinations Officer and ICT staff operating the facility. This data will be deleted 3 months after the dissertation has been completed. I will also store a password protected back up copy on my personal laptop.

Section F: Supporting documentation

Information & Consent
Participant information sheets relevant to project?
Yes

**Participant Information Sheets**
- Ethics_Info_Consent_template.docx
  (Document 025068)

Consent forms relevant to project?
No

**Additional Documentation**
None

**External Documentation**
-not entered-

Official notes
-not entered-

**Section G: Declaration**

Signed by:
Bin Shu
Date signed:
Thu 30 June 2016 at 16:15
Appendix B: Information sheet

The University of Sheffield.
Information School

An Evaluation to Different Types of Data Visualization

Researchers
Bin Shu on 07421227389 / email bshu1@sheffield.ac.uk.

Purpose of the research
The research is going to identify each types of data visualization, analyze the advantages and disadvantages of different kinds of data visualization, make a statistic of people's responses toward each type of data visualization, raise an evaluation and give some suggestion to improve the performance of the data visualization.

Who will be participating?
Potential participants will be the students enrolled at the University of Sheffield.

What will you be asked to do?
We will ask you to complete a brief demographics questionnaire so that we have a profile of our participant group. Then we will conduct a 15 minute questionnaires about the response to different types of data visualization.

What are the potential risks of participating?
The risks of participating are the same as those experienced in everyday life.

What data will we collect?
We are collecting the responses of participants when they facing different kinds of data visualization and recording their subjective marks about their feelings.

What will we do with the data?
We will be analyzing the data for inclusion in my master's dissertation. After that point, the data will be destroyed.

Will my participation be confidential?
Each survey response will be associated with an arbitrary number; all responses will be anonymous. Any IP addresses associated with the dataset will be deleted as soon as the file is downloaded from the web.

What will happen to the results of the research project?
The results of this study will be included in my master's dissertation which will be publicly available. Please contact the School in six months.

I confirm that I have read and understand the description of the research project, and that I have had an opportunity to ask questions about the project.

I understand that my participation is voluntary and that I am free to withdraw at any time without any negative consequences.
I understand that I may decline to answer any particular question or questions, or to do any of the activities. If I stop participating at all time, all of my data will be purged.

I understand that my responses will be kept strictly confidential, that my name or identity will not be linked to any research materials, and that I will not be identified or identifiable in any report or reports that result from the research.

I give permission for the research team members to have access to my anonymised responses.

I give permission for the research team to re-use my data for future research as specified above.

I agree to take part in the research project as described above.

Participant Name (Please print)  Participant Signature

Researcher Name (Please print)  Researcher Signature

Date

Note: If you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, please contact Dr Jo Bates, Research Ethics Coordinator, Information School, The University of Sheffield (ischool_ethics@sheffield.ac.uk), or to the University Registrar and Secretary.
Appendix C: Letter of Approval from Information School Research Ethics Panel

The University Of Sheffield.

Downloaded: 21/07/2016
Approved: 12/07/2016

Bin Shu
Registration number: 150213361
Information School
Programme: Information Management

Dear Bin

PROJECT TITLE: An Evaluation to Different Types of Data Visualization
APPLICATION: Reference Number 010491

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 12/07/2016 the above-named project was approved on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 010491 (dated 30/06/2016).
- Participant information sheet 1020833 version 1 (30/06/2016).

The following optional amendments were suggested:
Update the consent form for web use. Consider targeting specific types of students from the volunteer list.

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Matt Jones
Ethics Administrator
Information School
Appendix D: Questionnaire

Basic Information
1. What is your gender?
   ○ Female
   ○ Male

2. What is your study level?
   ○ Undergraduate
   ○ Postgraduate

3. What is your major?
   ○ Science
   ○ Non-science

4. What is your native language?
   ○ English
   ○ Other:__________

Here is an example of data visualization.
How important are each of these features in data quality? from 1 (unimportant) to 5 (very important)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How important are each of these features in presenting data? from 1 (unimportant) to 5 (very important)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chart Type (eg. pie/bar/line chart etc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axes/Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legend/Key</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. How often do you encounter this kind of data visualization?
- Never
- Yearly
- Monthly
- Weekly
- Daily

8. How easy is it to understand this kind of data visualization?
- Very Difficult
- Difficult
- Fair
- Easy
- Very Easy

9. How often do you create this kind of data visualization?
- Never
- Yearly
- Monthly
- Weekly
- Daily

10. How useful is this type of data visualization for presenting data?
- Very Useless
- Useless
- Fair
- Useful
- Very Useful

11. What do you think about this chart style?
- Very Boring
- Boring
- Fair
- Interesting
- Very Interesting

12. What do you think about this type of data?
- Very Simple
- Simple
- Fair
- Complex
- Very Complex

13. Do you have any further comments about this data visualization?
14. How often do you encounter this kind of data visualization?
   ○ Never  ○ Yearly  ○ Monthly
   ○ Weekly  ○ Daily
15. How easy is it to understand this kind of data visualization?
   ○ Very Difficult  ○ Difficult  ○ Fair
   ○ Easy  ○ Very Easy
16. How often do you create this kind of data visualization?
   ○ Never  ○ Yearly  ○ Monthly
   ○ Weekly  ○ Daily
17. How useful is this type of data visualization for presenting data?
   ○ Very Useless  ○ Useless  ○ Fair
   ○ Useful  ○ Very Useful
18. What do you think about this chart style?
   ○ Very Boring  ○ Boring  ○ Fair
   ○ Interesting  ○ Very Interesting
19. What do you think about this type of data?
   ○ Very Simple  ○ Simple  ○ Fair
   ○ Complex  ○ Very Complex
20. Do you have any further comments about this data visualization?

---

Line Chart

[Image of Line Chart: Real House Prices]
21. How often do you encounter this kind of data visualization?
○ Never ○ Yearly ○ Monthly
○ Weekly ○ Daily

22. How easy is it to understand this kind of data visualization?
○ Very Difficult ○ Difficult ○ Fair
○ Easy ○ Very Easy

23. How often do you create this kind of data visualization?
○ Never ○ Yearly ○ Monthly
○ Weekly ○ Daily

24. How useful is this type of data visualization for presenting data?
○ Very Useless ○ Useless ○ Fair
○ Useful ○ Very Useful

25. What do you think about this chart style?
○ Very Boring ○ Boring ○ Fair
○ Interesting ○ Very Interesting

26. What do you think about this type of data?
○ Very Simple ○ Simple ○ Fair
○ Complex ○ Very Complex

27. Do you have any further comments about this data visualization?
28. How often do you encounter this kind of data visualization?
○ Never ○ Yearly ○ Monthly
○ Weekly ○ Daily

29. How easy is it to understand this kind of data visualization?
○ Very Difficult ○ Difficult ○ Fair
○ Easy ○ Very Easy

30. How often do you create this kind of data visualization?
○ Never ○ Yearly ○ Monthly
○ Weekly ○ Daily

31. How useful is this type of data visualization for presenting data?
○ Very Useless ○ Useless ○ Fair
○ Useful ○ Very Useful

32. What do you think about this chart style?
○ Very Boring ○ Boring ○ Fair
○ Interesting ○ Very Interesting

33. What do you think about this type of data?
○ Very Simple ○ Simple ○ Fair
○ Complex ○ Very Complex

34. Do you have any further comments about this data visualization?
35. How often do you encounter this kind of data visualization?
   ○ Never ○ Yearly ○ Monthly
   ○ Weekly ○ Daily

36. How easy is it to understand this kind of data visualization?
   ○ Very Difficult ○ Difficult ○ Fair
   ○ Easy ○ Very Easy

37. How often do you create this kind of data visualization?
   ○ Never ○ Yearly ○ Monthly
   ○ Weekly ○ Daily

38. How useful is this type of data visualization for presenting data?
   ○ Very Useless ○ Useless ○ Fair
   ○ Useful ○ Very Useful

39. What do you think about this chart style?
   ○ Very Boring ○ Boring ○ Fair
   ○ Interesting ○ Very Interesting

40. What do you think about this type of data?
   ○ Very Simple ○ Simple ○ Fair
   ○ Complex ○ Very Complex

41. Do you have any further comments about this data visualization?

________________________________________________________________________
Access to Dissertation

A Dissertation submitted to the University may be held by the Department (or School) within which the Dissertation was undertaken and made available for borrowing or consultation in accordance with University Regulations.

Requests for the loan of dissertations may be received from libraries in the UK and overseas. The Department may also receive requests from other organisations, as well as individuals. The conservation of the original dissertation is better assured if the Department and/or Library can fulfill such requests by sending a copy. The Department may also make your dissertation available via its web pages.

In certain cases where confidentiality of information is concerned, if either the author or the supervisor so requests, the Department will withhold the dissertation from loan or consultation for the period specified below. Where no such restriction is in force, the Department may also deposit the Dissertation in the University of Sheffield Library.

To be completed by the Author – Select (a) or (b) by placing a tick in the appropriate box

If you are willing to give permission for the Information School to make your dissertation available in these ways, please complete the following:

- (a) Subject to the General Regulation on Intellectual Property, I, the author, agree to this dissertation being made immediately available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

- (b) Subject to the General Regulation on Intellectual Property, I, the author, request that this dissertation be withheld from loan, consultation or reproduction for a period of [   ] years from the date of its submission. Subsequent to this period, I agree to this dissertation being made available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

Name: BIN SHU
Department: Information School
Signed: BIN SHU
Date: 30th AUG 2016

To be completed by the Supervisor – Select (a) or (b) by placing a tick in the appropriate box

- (a) I, the supervisor, agree to this dissertation being made immediately available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project.

*Special restrictions

- (b) I, the supervisor, request that this dissertation be withheld from loan, consultation or reproduction for a period of [   ] years from the date of its submission. Subsequent to this period, I, agree to this dissertation being made available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project
Name
Department
Signed
Date

THIS SHEET MUST BE SUBMITTED WITH DISSERTATIONS BY DEPARTMENTAL REQUIREMENTS.
CONFIRMATION OF ADDRESS
The University student record system has the capability to record up to three different addresses for you at any one time: HOME address, TERM-TIME address and CORRESPONDENCE address.

Please note that the University and the School will use the details on your HOME address record when posting out information to you, e.g. Statement of Results, Notification of Degree Results, Degree Ceremony information etc. It is therefore YOUR RESPONSIBILITY to ensure this address information is up to date so that communications we post to you get to you.

However, if you know that you will be staying at a temporary/other address rather than your home address after completing the programme, and you wish correspondence to be sent to this temporary/other address it is YOUR RESPONSIBILITY to add the address details to the CORRESPONDENCE address on your student record. If you have provided a Correspondence address this will be used instead of the Home address to post out communications to you. (See www.shef.ac.uk/ssid/record/correspondence.html for more information) The Schools will not send out correspondence to your TERM address once you have completed your studies with us.

You can check and update all your address information by logging on to the SSID website at www.shef.ac.uk/ssid/record/pin.html. You will need your UCARD PIN number to access your student record.

I confirm that I have checked, and updated if necessary, my HOME address details on my University student record

I wish correspondence to go to a temporary/other address and not my Home address, and I have therefore provided CORRESPONDENCE address details on my University student record.

Alumni Information

I agree that the Information School may pass details regarding the result of my degree to the relevant Professional Body Organizations for the purposes of assisting these organisations in deciding my eligibility for the award of professional accreditation.

I would like my name to go on the School’s Alumni webpage.
I would like my name and email address to go on the School’s Alumni webpage
My email address is __________________________________________
(please do not use your Sheffield University email address)

I would like to be forwarded details of any job vacancies received by the School.
My email address is __________________________________________
(please do not use your Sheffield University email address)