CHAPTER 1

Introduction

1.1 Background and rationale

There is a book called “Thinking, Fast and Slow” written by Daniel Kahneman discussing about human thinking systems. There are two systems, fast thinking and slow thinking. Whereas fast thinking is the way that uses a little effort of thinking or no effort, such as we can compute the result of 1+1 instantly, called System 1, Slow thinking is when human finds the problem harder, say, 15*17. In this situation we cannot compute it easily (other than exceptional people), however, we can compute it but time or pencil and paper may be used. He called the “Slow thinking” or System 2. Thus, System 2 takes more effort to solve the problem. Consider LINE A and LINE B in Figure 1.1 developed by Franz Müller-Lyer. Our eyes tell us that the LINE B is longer than LINE A. That is the function of System 1. Even if you already measured that the two lines are long equally, but your eyes still tell you the same result that the LINE B is longer.

We link this idea to data representation in statistical graphics, such as a bubble plot, a bar chart, or a pie chart. Our eyes can easily distinguish differences while looking at those such graphics, say, which size is bigger or which line is longer than the others. Because the System 1 is used to assess results. However, even human eye distinguished the different but if we analyze the data with statistical methods the result may difference.

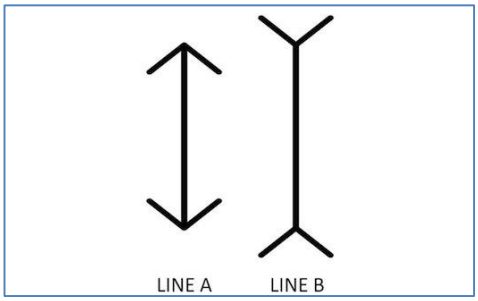

Figure 1.1 Franz Müller-Lyer's image

(Taken from: http://www.popsci.com/files/imagecache/article_image_large/articles/ALTER ART 21_adjusted1.jpg)

Also there is a sentence, "A picture is worth a thousand words", that describes the benefit of a picture or image in explaining a lot of information including hidden information by just one picture. Typically, explaining information with a word is difficult for people with less communication skills and it is time consuming. Oppositely, only words may mislead the viewer. Some people have the ability to explain a word beautifully but most of them do not. Therefore, showing a picture, graph, diagram or histogram can reduce an explanation in words. In addition, the picture may give hidden information which can be useful, such as the pattern of data. This thesis tried to find a good way to display information that makes it easy for people to read and understand.

Therefore, the idea mentioned above and the sentence "a picture is worth a thousand words" give us the idea to study and develop a system that uses the benefit of System 1 for data visualization.

Computer science and Statistics have become important factors for human being, mainly for education and business organizations. The computer makes problems easier to solve than in the past, particularly statistical problems. In the past using statistics was

difficult because the calculation was time consuming, but nowadays, statistics can be computed very fast by using statistics software, and even produce graphs, images and histograms very efficient. Statistical graphs is very important for displaying information as well as statistical methods, which are used for analyzing data. Typically, statistical methods can be easily understood by statisticians, but it is difficult for people without statistical knowledge. But statistical graphics can be understood by people with less statistics knowledge. However, the limitation of the statistical graphic is lack of ability to interact. For example it cannot zoom, tilt, rotate or even make the bird's eye view. Therefore this thesis tried to develop these issues as we mentioned.

Currently, the World Wide Web (WWW) technology is common; results from statistics should be easily publishable over the Internet (say, an interactive display system for statistics over the Internet). There is widespread software available, both commercial and freeing download-able, but most of them cannot produce a dynamic graphic, they can only export to a static image (say, jpeg, gif or bmp) but they are not interactive images. An example is when we use a bubble plot with a lot of circles in a small area so points are very close together or overlapping, making them messy. Thus, the challenge is how to distinguish those points which look like just one point. The only way to distinguish is to zoom closer and observe the area, then scale (x and y axis) with points expanding rising from the others, and then we can simply distinguish them.

Another example is how to display statistics results instead of using a table of data. Bar charts, pie charts, bubble charts, are all available, even using different colours as well. But these charts do not show geographic variation. So we must combine statistics with a geographical information system (GIS), which will provide clearer information (by

putting a bar chart on the real area in the map, say). This will tell the viewer the exact location rather than in our imagination.

The statistician themselves sometimes find it is very difficult to compare two similar statistical results (for example, incidence rates of Muslim and non-Muslim residents in three southernmost of Thailand, including 4 districts of Songkhla province). The primitive way is to print out graphs and then lay down them close together and try to compare by eye; this is not efficient and is time consuming, also wasteful of resources, particularly paper. Therefore, we try to invent a system that can compare results with a few mouse clicks. Particularly, it should work over the Internet. This would save the cost.

There are two ways of displaying information over the WWW: static and dynamic. Static is a primitive way to publish information over a web site which employs Hyper Text Markup Language (HTML), similar to reading a book over computer networks. The content in the book will be the same anytime we read it. Dynamic is the new technique to present information over the web. Script languages (JavaScript is the well-known script computer language) which can select different information to display on the same form, make information change depending on the situation. For example, in a news paper's web page, the headline changes every day, even on the same page because they will be new news. Sometimes, the static is better than the dynamic way in case of implementation. However, in some situations, like making a page for students as a university, it is a not good way to make a page individually for each student because it will take a lot of working time. Thus, it is much better to make just one page and then upload a student's individual data to the page from a database.

There are two studies in this thesis. The first study focuses on statistical and graphical methods based on non-Muslim populations in the three southern most provinces of Thailand including four districts of Songkhla province, using data provided by the Deep South Coordination Center. This study used a bubble chart to present the injuries death pattern in different areas, having 37 districts. At the end we grouped some contiguous districts together into 15 super-districts because the non-Muslim populations are too different. We also used a bubble chart to determine differences of injury rates between non-Muslims in terms of sex, with males at greater risk than females. Finally we fitted models with Poisson, negative binomial and log-normal distributions. We found the log-normal fitted better than the others.

The second study concentrates on trying to display information on the real location in three-dimensions by using Google Earth and R programs. The data collected by DSCC from 2004 to 2010 for non-Muslims was selected to do analysis because Muslims were affected less than non-Muslims (mainly Buddhists). We used confidence interval plots for data and then separated these into three groups, above the mean (most risk), near the mean and below the mean (less risk). Furthermore, we plotted two-dimensional rectangle with different colours on the surface of map. Also, at the end, we used Keyhole Markup Language (KML) to provide the ability to interact with Google Earth. However, we did not try to create KML files directly, but used R programs to generate these files. That means the KML file is changed dynamically every time we re-run the R codes.

1.2 Literature review

The literature review in this section will discuss the southern Thailand situation and who have studied statistical graphics. We discuss thematic maps and three-dimension interactive systems.

Southern Thailand terrorism

In the past Thailand, which used to be call Siam, expanded a Colony to the south which covered some provinces of Malaysia, Pattani, Yala and Narathiwath which were mostly Muslim. But Thailand is Buddhist country, so they forced the Muslims to assimilate. Muslims were not happy with that situation, and they wanted to become a separatist state. The well-known separatist organization is the Pattani United Liberation Organization (PULO). See (2009) claimed that the separatists of southern Thailand used religions to drive their objective to separate the southern area of Thailand to be an independent state. However, the situation after that, deteriorated in 2004, when the current terrorism sparked by an incident in Tak Bai when Muslim protestors died from suffocation after being arrested by Thai authorities. Many innocent people, particularly, school teachers, soldiers, police, and other residents have been killed or injured. The separatist used gunshot to kill targeted people and bombs for creating a risky situation. As a result, there are now a lot of widows and orphans in this area. The Thai government tried remedies to help those affected by the situation. Finally, the Deep South Coordination Center (DSCC) was established to help the Thai government collect accurate data on the incidence of the terrorism injuries.

There have been many studies about the southern Thailand situation, the following are some examples.

Komolmalai et al (2012) studied incidence rates of Thai Muslim in three southern-most provinces and four eastern districts of Songkhla. They found the annual risk of Muslim comparing to non-Muslim to be 48/121 per 100,000 residents. That means non-Muslims have almost 3 times the risk of Muslims. Interestingly, Muslims were injured by gunshot but non-Muslim mostly were victims of bombing.

Pimpa (2008) studied about motivation of teachers in five provinces of southern Thailand. He found that terrorism had an effect an economic and educational situation in the area. A lot of Thai officials moved out of the area, especially teachers. This caused more problem in the area. Then his suggestion was to encourage teachers in the five provinces to live in the area and do their work, with increased salary.

Graphical

"A picture is worth a thousand words", (Wikipidea 2013) means just one still image can explain a lot of information. Note that the original proverb comes from the Chinese philosopher Confucius who is credited with saying, "A picture is worth 10,000 words". For example, in statistics one graph or plot can give more information than just a pure numeric value. Also, readers will be able to absorb information quicker by looking at a picture than by reading.

The first known graph was developed in the 10^{th} century (Tufte 1983). Nicolas of Cusa plotted a graph of distance versus speed in the 15^{th} century. An important initiator of graphical methods was Playfair (1786, 1801). He developed line, bar and pie charts.

Whenever an image is created, it encodes information in it (Cleveland, 1985) through shape, position, colours, and symbols. The decoding system is in the viewer's eyes. If a viewer look at the graph and gets encoded information from the graph, graph is successful.

Tufte (1997) said that to present a huge amount of information in a small area is difficult without including a graphic or image. Additionally, how an image is represented can mean whether they will die or live. For example, the Challenger space shuttle exploded because an O-rings problem due to O-ring temperature chart showed unclear information, the O-ring that were used could not survive very cold conditions.

There is a paper that tried to compare learning efficiency with animated and static images Vogel-Walcutt et al (2010), selected two groups, fire support teams taught procedural skills and conceptual knowledge using static pictures with the first group, and using animated pictures with the second group. Results found, for the first test, the static pictures were better. However, for the second test, they found the result of the animated pictures were better. Therefore, in long term, the human brain can remember animated pictures better than static pictures.

1.3 Objectives

Objectives of this thesis are to find a new method to represent data in term of visualization which helps the reader better to understand the meaning of data. The data is the incidence rates in the three southern-most provinces and 4 districts of Songkhla. These methods are developed in following steps for the example chosen.

(1) Fit a statistical model to incidence rates of injuries for non-Muslims in,

(2) Show the incidence rates of injuries for non-Muslims.

(3) Make a dynamic thematic map to show the terrorism incidence rates.

(4) Make a three-dimensional interactive system for showing incidence rates.