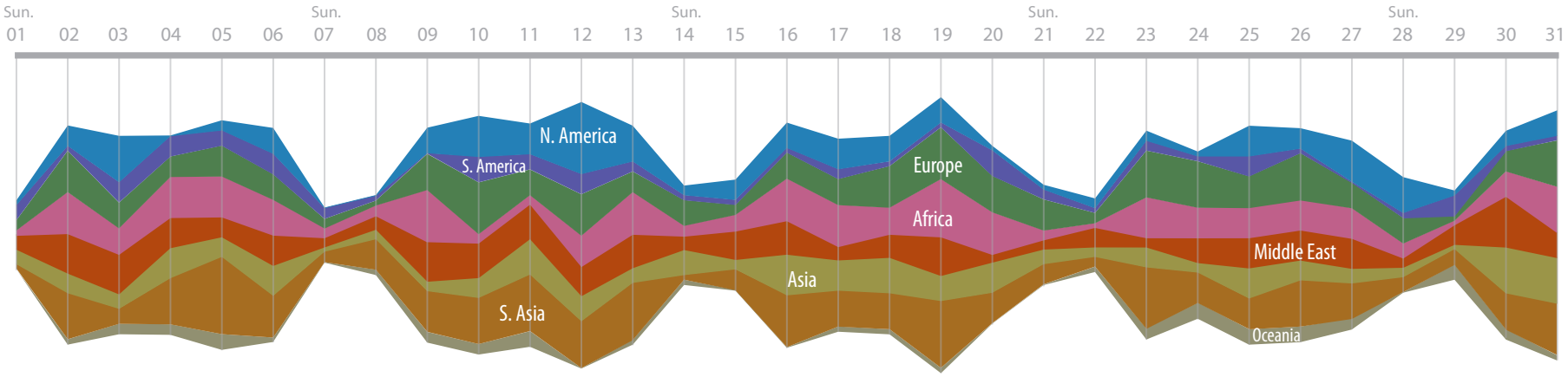
Mapping the recent past: Visualization of online news archives



Author: Li Xin

MA Thesis, 2007

Media Lab, University of Art & Design, Helsinki (TAIK)

Acknowledgements:

It would like to thank the following individuals:

Lily Diaz, my supervisor, for inspiration, encouragement and patience.

Heidi Tikka, Rasmus Vuori and fellow students in Final Thesis Seminar, for great discussions. Especially thanks to Marcus S Malka, my opponent, for important comments and suggestions.

Markku Reunanen and Samu Mielonen, teachers in Dynamic Visualization Design Seminar, for introducing me to the subject of information visualization.

Pipsa Asiala, for supporting the thesis process.

Heli Rantavuo, for her advice on academic writing.

Brenda Castro, for thesis proofreading and encouragement.

My dearest friends Ida Blekeli, Bjorn Saunes and Viki Olgod.

Eirik Fatland.

Table of Contents

Abstract

This thesis is about studying online news archives through design and visual representation. For the design work, I have chosen to focus on a case study: the archives of BBC online news. Two experiments were carried out: a quantitative data visualization of news coverage by country, and a set of visualizations of the textual contents of the news articles.

This thesis is an exploration of the subject of information visualization from a graphic designer's point of view. The field of information visualization has roots in multi-disciplinary research and practice. This work explores how graphic design can contribute to the field of information visualization, as well as to digital media in general. The work also examines some of the opportunities and limitations of digital media when it comes to facilitating graphic design and visual communication.

This thesis is relevant for research in information visualization, graphic design in digital media and online news archives.

1 Introduction

The work of this thesis is an exploration of the field of information visualization from a graphic designer's approach. Within the time limits of a master thesis work, I have chosen to focus on a case study- visualizing the archive of BBC online news. This thesis is the documentation of my design. It explains how I approach the field of information visualization as a graphic designer, and demonstrates how I use visualization as a tool to explore the subject of online news archives. It also shows how I communicate my understanding of the subject through the form of visual language.

Why should one be interested in information visualization? The increasing possibilities brought by digital communication technologies also come with increasing complexity. With large amounts of information easily available at the fingertips of individuals, the computer shifts from being a data processing machine to being an information display platform. Reading is what we spend most of the time doing in front of personal computers. Traditionally, configuring the reading experience has been the responsibility of the graphic designer. However, when dealing with massive amounts of information, the reading experience is not only about readability, legibility and aesthetics. It also concerns revealing patterns and complex relationships between data, in order to provide the reader with overview and context, and to aid the reader navigating and exploring the ever-growing information sphere. The ongoing field of information visualization is essential to meet this need.

Information visualization is a multi-disciplinary field. As Tufte (2001) has shown, this field grows out from graphic design, especially information graphics and statistics, and is inspired by cartography. Furthermore, the field is strongly related to graphical user interface design, information architecture, and human computer interaction (HCI/CHI).

This thesis is about visual representation and design in the field of information visualization, about the use of visualization as an external tool for exploration and communication, and about the classical techniques of design in printed media as they might be applied to digital media. This thesis is also about how we deal with large amounts of information when designing visual interfaces for them and about the possibilities and limitations that digital media offers for conveying information in a visual form.

Although every topic has its own specific set of problems, the principles of graphic design are applicable across a wide range of contexts. In this thesis, I have chosen to focus on the topic of visualizing an online news archive to explore the role of graphic design in information visualization.

1.1 Motivation

The work is the result of two personal motivations: a passion for graphic design, and an interest in studying online news as a digital cultural product.

About graphic design

For the last few centuries, graphic design has mainly been dealing with print media, such as paper and fabric. Traditional graphic design can be seen as the art of conveying messages on a two-dimensional surface. The age of the personal computer has fundamentally changed the nature of graphic design. Graphic designers must deal not just with changing tools and processes, but also with changing design problems and changing distribution of the work. The affordances and constraints of a computer system are mainly perceived through human vision. This is related to the way we interact with the machine, the position of the user, the surroundings and the screen itself. Designing visual elements falls into the domain of graphic design. However, digital media also offers opportunities for visual expression such as dynamic visual representations and animations which are not familiar to the traditional field of graphic design.

In the field of information visualization, much attention has been paid to data mining theory, real-time graphical rendering techniques, software algorithm design and novel interaction methods. Ironically, however, attention to the quality of graphic design is notably absent in the generally accepted definitions of "information visualization", as well as in the research and practice of the field. Currently, the graphic design of digital media is mainly seen as a form of artistic expression, or reduced to a merely decorative function. This is a misunderstanding of the nature of graphic design, which is not primarily about style but about clear communication in a language that includes such things as dimension, shape, color, composition and typography.

About online News

Through the Internet we are exposed to information and stories about places we have never been to, events we have not personally experienced. In a way the function of online news is not only to report on current events, but also to build up some common memories of the "recent past", the past that is too recent to be recorded in history books.

Most of us did not personally experience the Iraq war of 2003, SARS in east Asia, or the bombings in the London metro of 2005. However, these are familiar landmarks in our memory when we think back on the past five years. One of the reasons is that the same stories have been told by different people with different perspectives, and we have experienced these stories over and over again with different versions through various media, such as news articles, photo documentation, video footage on YouTube, numerous blogs, or movies made partially based on true stories. New events occur rapidly, and we are busy with experiencing. However, there is lack of appropriate tools to reflect on what we have been exposed to: What are the consequences of these mediated experiences? How do these experiences contrast from our personal experience in the physical environment? Most importantly, how do these mediated experiences alter our perception of the world? These questions are too complex to be fully explored in the frame of a masters thesis. However, to consider these questions is an important motivation for the work of this thesis.

Reading news as current events and reading news archives are different types of readings with different purposes. Reading news as current events focuses on getting an overview of the most recent events reported by various news media. In this case, the reader is mainly interested in a short time span. Old news is irrelevant for this type of reading and considered as information noise. However, news archives treat news as dynamic cultural products. Old news reports are important because the reader can learn from them how news changes over time, notice the permanent topics that our society considers "news worthy", and try to comprehend how the news media portrays the world. These questions should not ony be studied by media analysts. Tools to aid exploration of news archives would help readers and the general public to better contextualize and make sense of what they read.

1.2 Case study: BBC online news archive

Within the frame of this thesis, a limited dataset was collected for testing and prototyping purposes. The data was collected through the RSS feeds provided by BBC news' website in the duration of six month. A database containing all these news articles was constructed. The website (news.bbc.co.uk) is published by the news division of the BBC, the UK's publicly owned broadcaster. It is one of the oldest and most popular news websites worldwide, averaging around 15 million visitors per month (Wikipedia contributors).

The choice of dataset might seems arbitrary. I am aware that another set of news article collection may tell entirely different narrative. However, the focus of this thesis is not on the results of visualization or the narratives presented, but on exploring how visualization can be a means to explore news articles and reveal patterns, relationships and narratives which are not obvious by reading single articles. Although the dataset is limited, it is sufficiently large and varied that it can be used to realistically test design concepts for the visualization of news archives in general.

This case study comprises two visualization experiments with different approaches: quantitative data visualization of news coverage by country and visualization on textual contents of news articles. In these experiments, the design aims to provide visual external tools for reflecting, thinking, and comparing.

Experiment 1: visualizing news coverage

News Coverage Map, the results of the first experiment, aims to illustrate BBC news coverage by country, and visually test hypotheses about why some countries and regions might be given more coverage than others. The result is presented in the form of a screen-based, interactive prototype. The design of *News Coverage Map* is presented in Chapter 3.

Experiment 2: visualizing news contents

The second experiment deals with the visualization of news contents. The project explores how news portrays different places and topics to its reader, the relationships between these, and how news topics change over time. This experiment contains three distinctive visual design demonstrations, approaching these challenges from different angles. The results are presented in the form of sketches and visual mock-ups. The design process of this experiment is discussed in chapter 4.

Although the topic of the thesis is visualization of online news archives, the general visual design principles discussed here can be applied not only to online news archive visualization, but also to other types of data visualization. This will be discussed in detail in chapter 3 and chapter 4 in relation to each experiment.

2 Visual external aid

> *The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities. How have we increased memory, thought, and reasoning? By the invention of external aids: It is things that make us smart.*
>
> *Norman, Donald A. (1993, 43)*

Written texts and drawings are artifacts that function as significant external aids to human cognition. They serve two main purposes; The first of these is to exchange information, to aid communication amongst many. Without such external artifacts of communication we would have nowhere to store knowledge except in our own minds; no way to share knowledge except through word of mouth, which is highly limited both in distance and range. Furthermore, as memory is unstable, unreliable and fragile, much knowledge would not be inherited by future generations, but lost forever.

The second purpose of such external aids is to aid thinking, reasoning, exploring and performing a task at hand. These external artifacts act like information holders - extending a person's working memory. Examples of such practices are ubiquitous, from shopping lists made in everyday life to diagrams drawn for scientific research. Even with a relatively simple task, such as multiplying a pair of two digit numbers, the average person performs about five times faster using paper and pencil as an external aid than when using the mind alone. This clearly demonstrates the power of external aids (Card et al., 1-2).

Maps, graphs, and other types of visualization are also visual external aids. These visual external aids have advantages over written texts when we wish to gain a quick overview of information, deal with massive amounts of data with complex relationships, or work with data that

has a spatial property in nature. Visual representation takes advantage of the human ability to interpret visual and spatial data, thus reducing the cognitive effort required for processing large quantities of complex information. Information may be encoded in visual language such as color, shape, spatial proximity, or a combination of these properties. The images produced in this way differ from traditional portrait or landscape paintings, because these images depict relationships between objects, rather than the objects themselves. They provide an immediate overview by revealing patterns and relationships in the information that are not obvious in textual or other forms. The information in these images is perceptible and comparable, thus enabling deeper exploration of the subject that the image represents.

2.1 Graphic design and information graphics

Traditional graphic design can be seen as the art of conveying messages on a two-dimensional surface. Information graphics (also known as "data graphics") is one branch of this broad field. Unlike design for advertising and marketing purposes, in which the design is meant to persuade the viewer into a course of action, the essence of information graphics is to present information objectively, allowing the readers to ask and answer questions, make comparisons, and draw conclusions (Victor). Information graphics are widely used - we can find examples in everyday life, such as bus route schedules, stock graphics, and shopping catalogs.

The history of information graphics is strongly related to that of statistics and cartography. Many inventions of visual forms for abstract data visualization, such as the time-series graph, bar chart and polar graph

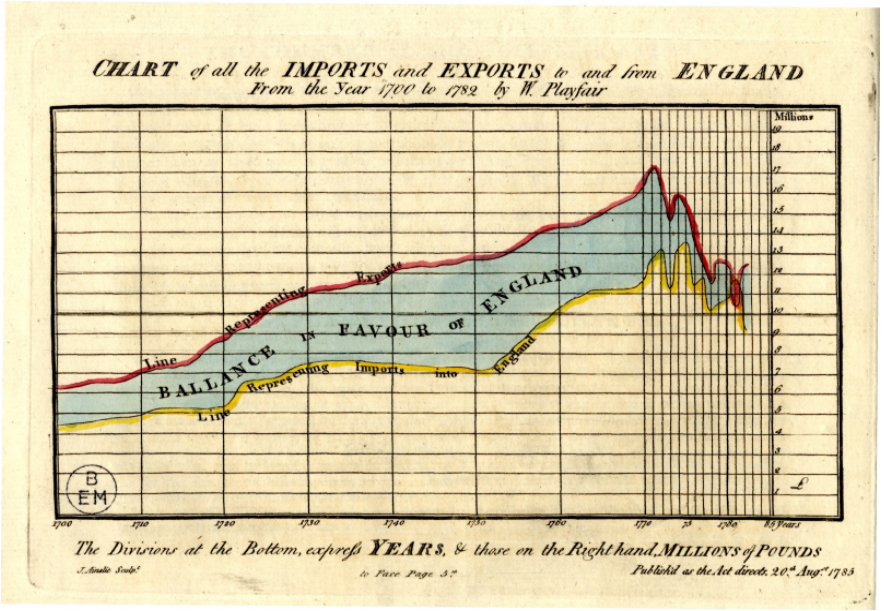


Figure 2.1 "chart of all the imports and exports to and from England from year 1700 to 1782" by William Playfair (Tufte, 2001,32)

were not based on the desire to show artistic expression, but were intended as visual aids to clarify data relationships, explore data in depths and communicate findings to others.

One great inventor in the field of information graphics is William Playfair, the Scottish economist. He is believed to be the first person to use time-series to display economic data and the inventor of a visual form, the bar chart (Tufte, 2001, 33-34). The graph (Seee Figure 2.1), entitled "chart of all the imports and exports to and from England from year 1700 to 1782", clearly plotted the duration and amount of two data variables: the import rate and the export rate. By mapping out those two variables, the balance

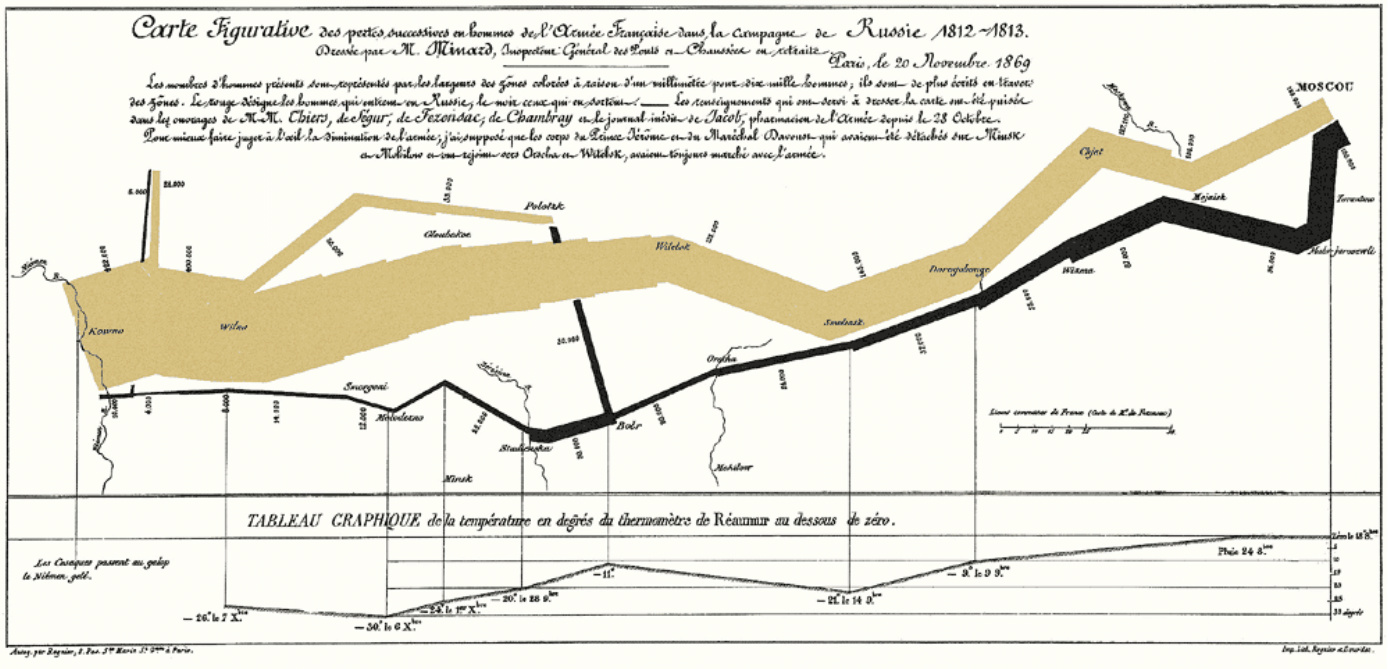


Figure 2.2 The graph of Napoleon's march, by Charles Joseph Minard. (Tufte, 2001, 41)

of import and export was revealed in the shape of the data itself. The grid in the background supports data comparison of different time durations: the x-axis represents time and the y-axis represents quantity. With the aid of a grid, the eyes can easily spot out data quantity in a certain time period.

A well-planned information graphics enables display of multiple data variables in a single coherent image. One of the most famous examples is probably the graph of Napoleon's march (See Figure 2.2), made in 1869 by Charles Joseph Minard, the French engineer. On this map, four variables were plotted on a single two-dimensional static image: the size of the French army, its location, the direction of its movement, and the temperature on different dates. The thick yellow line represents the invading army towards Russia, and the black line below represents the returning army. The thickness of the line indicates the size of the army at different locations. The graph clearly indicates the great loss of Napoleon's army, 422,000 men, of whom only 10,000 survived after the invasion. Furthermore, a few causes of the great loss are revealed in the graph: the low temperature and the crossing of the frozen waters of Berezina river. Minard's graph is an excellent combination of

statistical data and topological data (Mijksenaar, 29), as well as a brilliant information graphic that adopts time-series display while adding a spatial dimension (Tufte, 2001, 40).

The practice of graphic design, especially information graphics, has proven that visualization can be an effective way of communication, as shown in Playfair's chart. Furthermore, a well designed visualization not only displays information, but also conveys narrative, as shown in Minard's graph. Visual design practices from the branch of information graphics is one of the important roots for current research and practice in the domain of information visualization.

2.2 Information visualization

The ubiquitous digital communication technologies have created an enormous space for transfer, store, and sharing of information. For example, the nature of web includes the aspect of archiving (personal blog archives, public news archives, academic journal archives and the like). Many of these archives can be easily accessed across time and space through the network connection. However, the current available technologies have not completely proved to be a more efficient mechanism of knowledge communication. In the public library, for example, the task of finding relevant information would be assisted by professional librarians. However, the easy access to massive information through the digital communication platform, such as the web, may imply that individuals are left alone mostly to face the frustration of filtering and processing information with great energy and time waste.

When we are in an unfamiliar city, we look for external aids, such as maps, to guide us through city: to get to know where we are, or to find out about interesting places and the way to get there. This applies to physical space as well as to information space since maps depict not only the physical world, but also reveal conceptual relationships. The term to describe abstract data mapping in digital media is called information visualization. A general accepted definition (Card et al) as follows:

> *"Information visualization: use of interactive visual representation of abstract, nonphysically based data to amplify cognition."*

Tufte presented one of the most important aspects about graphic design that applies also to information visualization, noting that "graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency" (Tufte, 2001, 51). By contrast, the lack of attention towards graphical excellence is notably absent in the definition of information visualization, as well as in general practices in this field. Graphic design is generally considered to be of lower importance, because of the misunderstanding of graphic design as making visual components more attractive rather than representing the message of data in precision. As Ben Fry argues that minor problems in visual representation of a smaller dataset are vastly magnified in a larger one (Fry).

As stated, this work is about graphic design in the field of information visualization. The work explores how graphic design may contribute to information visualization when dealing with large and complex datasets. The work also experiments with the use of different information organization principles, dynamic visual display and navigation techniques which are specific to digital media. This exploration is essentially about creating visual external aids for digital communication.

3. Visualizing news coverage

In this chapter, I present my first design experiment of visualizing BBC online news archive: *News Coverage Map* - a screen-based information visualization which illustrates BBC news coverage by country and over time. I intended to use this visualization as an example to discuss how traditional graphic design knowledge can be applied to the field of information visualization for the purpose of communication. The discussion is based on the dialog between my design choice and my inspiration from the traditional graphic design, the practices information graphics and the theory of visual perception. Furthermore, through the design work, I explored some opportunities and limitations of facilitating graphic design in digital media.

The term "news coverage" is generally understood as the amount of news reports about a person, an event or a country in a given time period. In this thesis, I refer "news coverage" strictly to the amount of news reports about a country.

In a given time period, such as a day or a week, we could assume there are more newsworthy stories in one country than the other, such as war/conflict, natural catastrophes and new inventions or scientific discoveries. However, some countries/regions seem to receive more news coverage consistently. This observation made me ask the question: Are some countries/regions more important to the news than others? What makes them more important? It is reasonable for a news organization to choose this type of focus for its reporting. But it is important to consider this question because there are no straightforward ways for a news reader to identify the news coverage over longer periods of time. In most of the news websites, such as BBC, the focus of news coverage is hidden when we browse the news, since such sites devote their front page to current events, giving attention to the places where those events occur.

The goal of the *News Coverage Map* is to make this focus visible to its readers by providing a tool to show news coverage of BBC news website over a longer time-span. Visualization is a useful tool for this purpose,

because it can be used to provide an immediate overview and reveal patterns of the changes, which human eyes and brain are extremely sensitive to (Ware, 70).

To clarity the purpose of the *News Coverage Map*, the following questions are addressed in this visualization:

1. How does BBC online news divide their attention, the news coverage, between countries of the world?
2. Does the news coverage change over time?
3. Whether there is any correlation between the news coverage and geopolitical variables such as territory, population, GDP, and a country's relationship to the United Kingdom?

3.1 Data

A data collection was made comprising six months BBC online news, using the database mentioned in section 1.2. The choice of which data variables should be presented in *News Coverage Map* is determined by the questions the visualization attempts to answer. In order to describe the news coverage, I have chosen to emphasize three data variables: the names of individual countries, the number of news stories mentioning the country, and the span of time measured. Based on an assumption that articles displayed on the front page of the website are more important than other articles in the archive, I also collected information about whether a given news article appeared on the front page of the BBC news website.

In order to explore the possible correlations with the differences in news coverage, a few data variables specific to each country were collected: population per country, geographical size, continent, GDP and whether the country is a member of Commonwealth of Nations[1]. These data are mainly gathered from the CIA World Factbook.

Notice that the data variables are numerous in the data set. It is difficult to present multi-variable data in a single static visual representation. A more detailed discussion on how to translate data variables to visual variables is also discussed in chapter 3.3.

An interactive prototype of *News Coverage Map* was built and tested with two variations:

1. News coverage in all articles monthly (See Attached CD-ROM)
2. News coverage of front-page articles monthly (See Attached CD-ROM)

The prototype was implemented in Macromedia Director[2]. The results of the prototype are discussed in detail in section 3.4.

1 My original approach was to gather data of the countries which is former British colony. Commonwealth nations are the most similar data available. All members of commonwealth nations are former British colony.

2 A mulimedia authoring tool, a product of Adobe software company

3.2 Inspiration

Global Attention Profiles (GAP)

The concept of *News Coverage Map* was partly inspired by *Global Attention Profiles* (GAP). GAP takes a quantitative approach to the study of news coverage by country over multiple English language news websites (Zuckerman). GAP collects country-by-country story counts over a period of time in a news website and correlates them to a wide range of country datasets. This data is visualized on a map display (See Figure 3.1). However, the visualization suffers from a few inappropriate visual design decisions. First of all, the lack of text labeling of a country's name on the map makes it difficult to read the meaning of every geometric shape. For instance, we may not be able to point to the right shape that corresponds to the country Tunisia, because human memory is constrained and imprecise. Secondly, the inconsistence of color coding results in a visual representation that is hard to read.

To some degree, *News Coverage Map* can be seen as a continuation of GAP, since *News Coverage Map* focuses on the same issue: the study of news coverage in web media. However, I have chosen to study the subject of news coverage with a different approach: first of all, *News Coverage Map* was developed by focusing on a limited dataset from a single news website. By visualizing and studying one news website, I assumed that the method and design solutions can be applied to other news websites. Secondly, the aim of *News Coverage Map* is to provide a visual external aid for everyday news readers to discover the differences of news coverage among countries as well as to explore possible correlations between such differences.

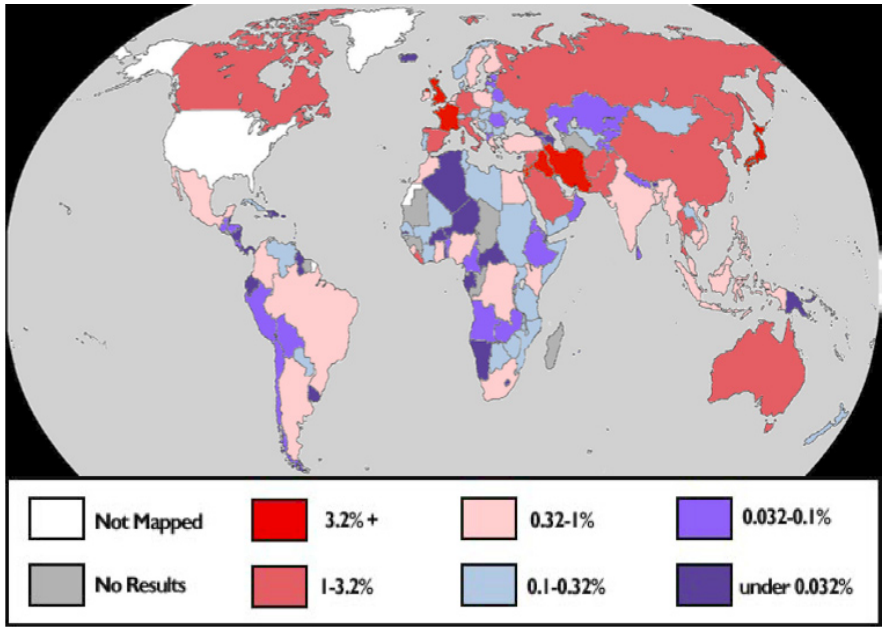


Figure 3.1 An example of the map created by GAP project. The map represents the news coverage of CNN News in the period of 1996 to 2003. (Zuckerman)

Cartograms

The visual design of *News Coverage Map* draws inspiration from various cartograms. A cartogram is a type of thematic map, where the appearance of countries, cities or regions is distorted according to a data variable which uses country, city or region as its measurement unit. A cartogram usually preserves the geographical spatial proximity among these units. The visual representation can be perceived by size differences of each individual units, as well as by the distortions of geographical space (Bertin, 120).

The images created in this way are striking. The distortion of the geographical map contrasts to our familiar visual representation of the world, and this contrast is the soul of a cartogram. It results in an image that surprises and draws the readers' interest. The readers are immediately struck by the gigantic proportion of South America's native plant species in the top map in Figure 3.2, as well as the enormous proportion of Asia's population in the bottom map.

Although the overall pattern of the data is clearly revealed in cartograms, distortion makes it difficult to read detailed information, such as the identification of the individual countries in Figure 3.2.

The *News Coverage Map* is essentially an interactive cartogram. The use of the principle of cartogram in *News Coverage Map* is discussed in section 3.3.2, and how *News Coverage Map* addresses the drawback of typical cartogram is discussed in section 3.3.3.

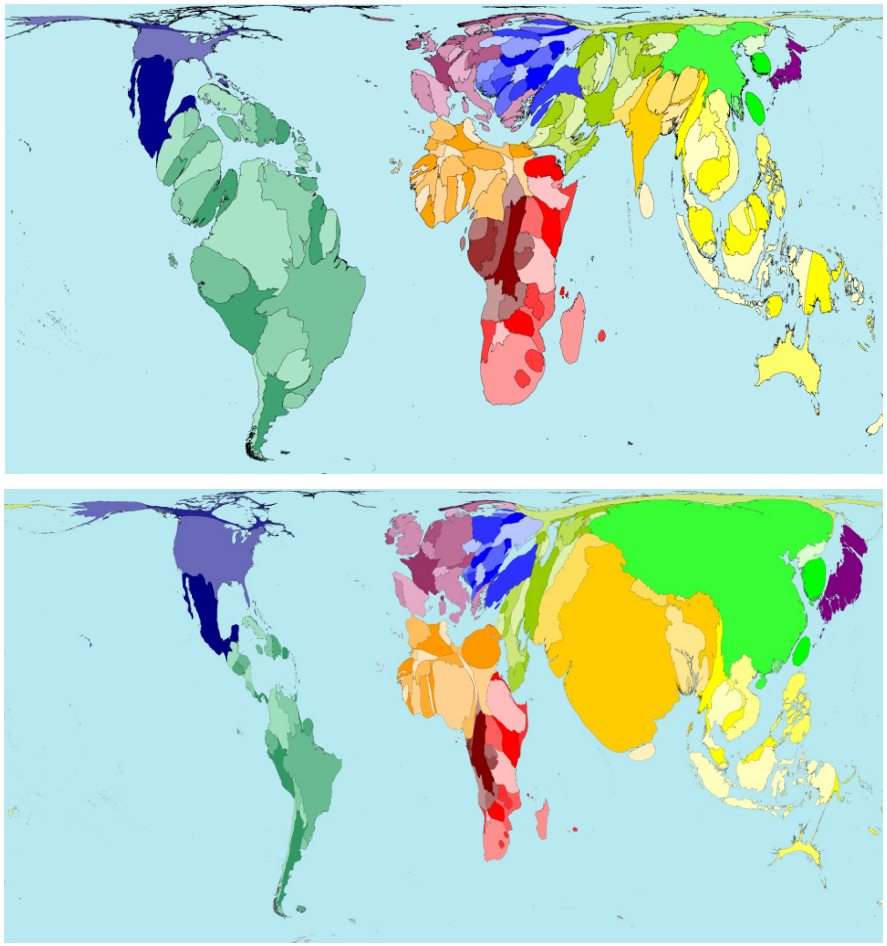


Figure 3.2 Two examples of cartograms (Worldmapper)
Map of native plant species (top): the territory size is proportional to the number of native plant species. Map of total world population (bottom): the territory size is proportional to the world's population living there.

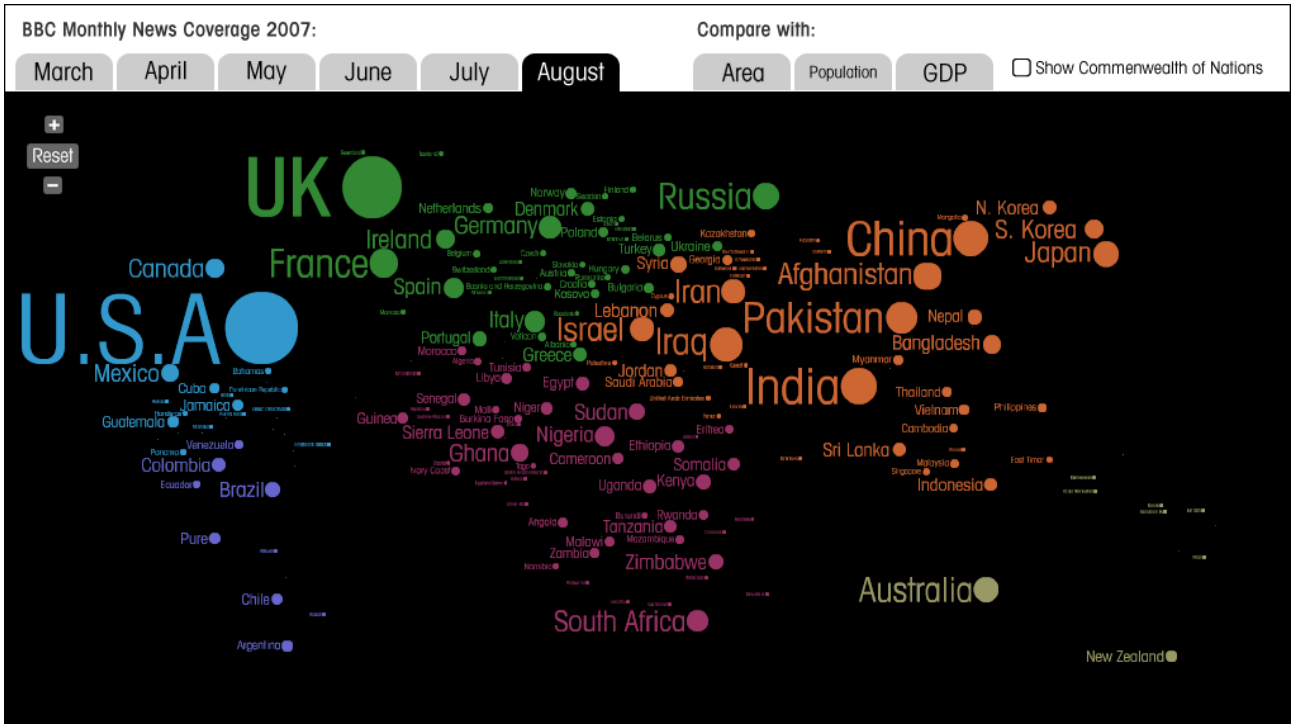


Figure 3.3 Screenshot of *News Coverage Map*. In this view, the size of the country displayed in the map is proportional to the amount of news stories reported about a country in BBC online news from the month of August, 2007

3.3 The design of *News Coverage Map*

The main screen of *News Coverage Map* displays a world map. However, instead of outlining the geographical shape of each country, the map uses the text of a country's name as its basic visual element. A navigation aid in the form of folder tabs that is located on the top of the map allows the reader to switch between different views, thereby updating the map with different data variables. The size of a country is correlated with data variables that have proportional quantities, such as news coverage, population, geographical size, and GDP. The color of each country is correlated with data variables that can be easily grouped into categories, in this case, continent, and whether the country is a member of Commonwealth of Nations. Two design decisions follow from opportunities specific to screen-based visualization: the use of animated transitions, emphasizing the change from one variable to another, and the mechanism of highlighting visual elements - allowing the user to focus on a smaller set of data.

In the following sections, I will discuss each design decision in depth, which includes: the choice of map display; the design strategies of mapping data variables to visual variables; the design of the basic visual elements to construct the map; the use of animated transition and highlight mechanism; the considerations regarding to the navigation and user interface.

3.3.1 The map display form

Time usually is the most prominent attribute in a news archive. But in the case of the *News Coverage Map*, the aim is to discover the differences of the news coverage between countries. Map display form was chosen, because most of the data variables in the data collection use country as measurement units (such as how many news stories about "Iraq" can be found in a given month's news articles).

Researching on the topic of maps and mapping has been part of my design process from an early stage. Maps are interesting for working in the field of information visualization, because a map can be used in a way that facilitates spatial understanding of places, objects, concepts and events. Maps as artifices are interesting to artists, historians and anthropologists, since these visual representations reflect human cultural activities, perception of the surroundings, as well as people's knowledge of the world.

The definitions of map are many and broad. From the point of view of display form, the word "map" refers to a visual representation that has a geo-spatial analogy. Thematic maps are typical examples of the visualizations that uses map display form.

A classical example of a thematic map is Dr. John Snow's incidence of cholera epidemic map from 1854 (See Figure 3.4). The dataset used in this map, the deaths by cholera, were originally organized as a time series. Instead of using a timeline display, John Snow plotted the location of deaths caused by cholera in central London during the month of September 1854 on a geographical map. By looking at the spatial pattern of the death locations on the map, he was able to discover that the origin of the spreading of disease was the water pump that was located in the area where most deaths occurred (Tufte, 1997, 30). This example demonstrated the power of maps to display spatial information, such as location.

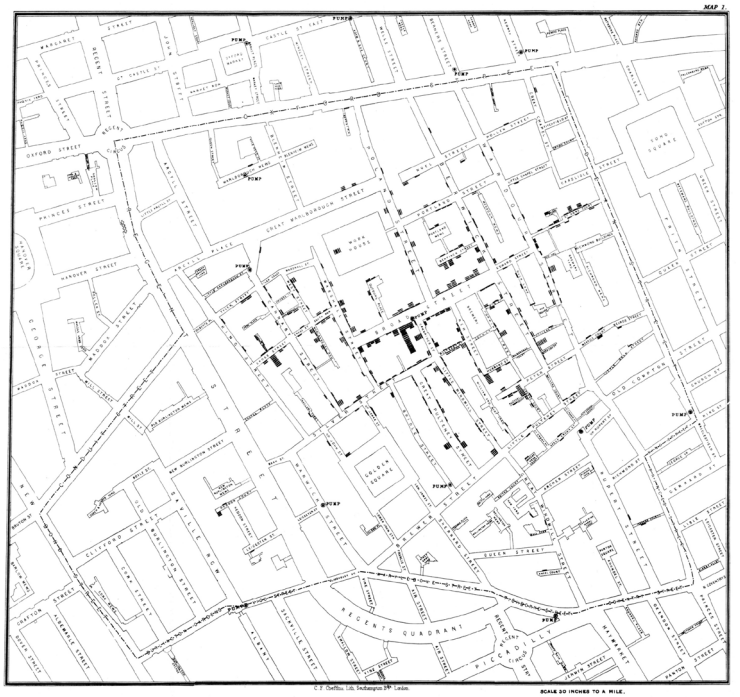


Figure 3.4 Map of cholera epidemic in London, 1854, By John Snow (Tufte, 1997, 31)

It is not a coincidence that many visual display forms, such as map display, have been inherited from stone and paper to digital media. The similarity of these media for visual display is that we have not escaped dealing with visual representations in two-dimensional surfaces, no matter whether the surface is a flat stone or a computer screen. The media evolve as new technologies become available to society. As we discussed in chapter 2, the invention of visual external aids holds a common goal which is to amplify human cognition. The availability of new types of data results both from current computation technologies and new types of human activities. This may requires the designers to rethink designing external aids to facilitate exploring, understanding, and learning.
In this case, the design of the *News Coverage Map* deals with online

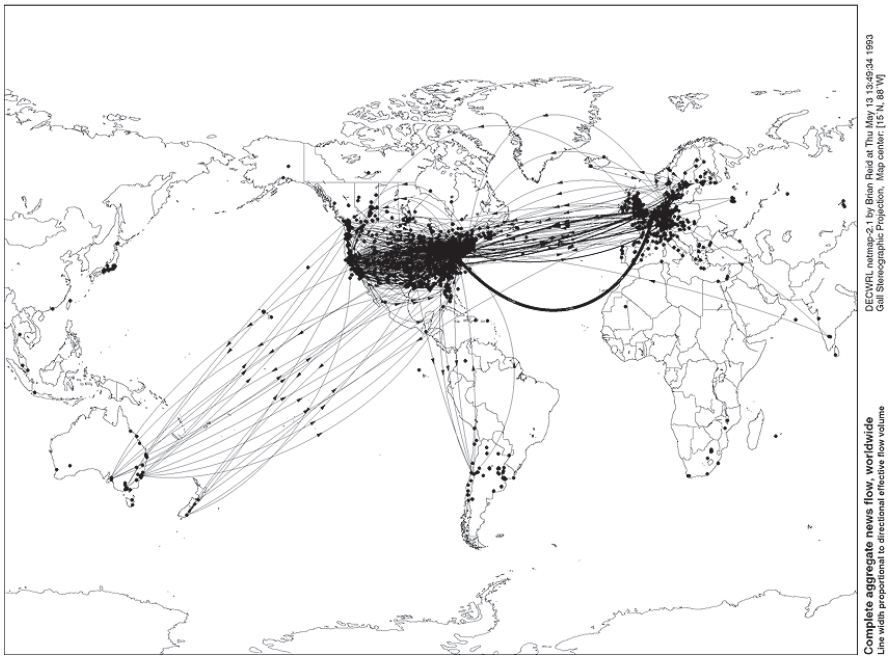


Figure 3.5 The map of *Usenet data flow*, by Brian Reid, May 1993 (Dodge & Kitchin, 53)

news. Can map facilitate our understanding of objects, concepts or events which belong to virtual environment, such as online news?

The following example demonstrates the possibilities of using map to facilitate understanding of a virtual community. The map *Usenet Data Flow* (See Figure 3.5), produced by Brian Reid, shows data traffic on the Usenet network in May, 1993. It was created as a part of a decade-long research project on the number of hosts and the amount of data exchange traffic in the Usenet network. Although virtual environments, such as those created through websites, do not have clear geographical borders, there are now many instruments and methods to measure the Internet and information sphere in relation to real-world geography. Where are Internet hosts located? Where are the user communities? In the case of *News Coverage Map*, we could measure where are the news stories come from. The art of cartography have broadened its borders from physical world into cyberspace (Dodge). Modern cartography deals with human activities in both physical world and virtual environments, for virtual environments gradually become an extension space for everyday social interaction.

In this map (See Figure 3.5), simple black dots represent the Usenet sites and smoothly curving lines represent the data flows between sites. The line thickness is proportional to volume of flow and small arrows indicate the direction of flow. This map may not be presented in the ideal way to facilitate reading and interpreting very detailed information because of heavily graphical overlapping in North America and Europe region on the map. But by having spatial analogy, the map clearly illustrates the essential fact about Usenet sites and reveals the demography of the community: there was very little connection to Usenet sites outside of North America and Europe.

Map display form facilitates a better understanding of data with spatial attributes in two aspects: the primary one is that spatial proximity is a powerful perceptual cue and organizing principle in design (Ware, 189-190). This is because human perception naturally groups visual elements which are located close to each other. The order of location reveals additional patterns, which may not be obvious or visible in the dataset, as in the example of John Snow's map and Reid's *Usenet Data Flow*. The second aspect is that map display is ubiquitous. For example, the order of the country's spatial analogy on the map is part of most people's knowledge from early experiences of map reading. Presenting spatial analogy of country or city increases the readability of the visualization and make it easier to make comparisons, as in the example of *Usenet Data Flow*.

As every visual display form has disadvantages, map display also has its own limitations. With map display form, the two dimensions (x-axis and y-axis) are occupied by one data variable, the geographic entity (Bertin, 24). In contrast, with a scatter diagram, the two dimensions are able to carry two different data variables (See Figure 3.6). Scatter diagram were one of considerations in deciding display form in *News Coverage Map*. However, the choice of visual display is related to the types of data and to the purpose of visualization. In the case of *News Coverage Map*, geo-spatial analogy provides a more intuitive comparison on the difference between countries than displaying data in scatter diagram with an additional data variable.

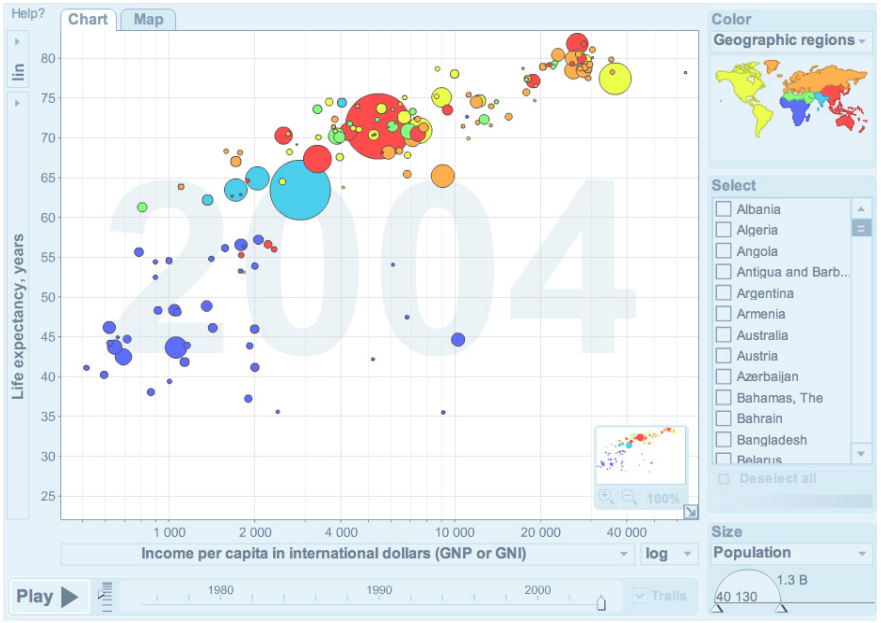


Figure 3.6 Screenshot of *GapMinder* application, an example of a scatter diagram (Gapminder)

3.3.2 The properties of visual elements

On a two-dimensional surface, a given data variable can be represented visually through size, value, texture, color, orientation, shape and spatial position. These variables are visual variables. In the context of discussing the design of *News Coverage Map*, I mainly focus on the use of size and color.

The data variables with proportional quantities in *News Coverage Map* are visually represented by size (See Figure 3.7). Quantitative perception can only be perceived through size, and no other visual variable has the same capacity (Bertin, 68-71). Note that the size of the country on the news coverage view does not necessary represent the amount of articles about these countries in the news archives in a given time, it represents the proportion the a country's news coverage in the archive compare to other countries.

By resizing the size of visual objects which representing countries according to not only the size of geographical area, but to other types of data variables, such as news coverage by country, population density and GDP, *News Coverage Map* is essentially a type of cartogram. One could evaluate the quantities of these data variables by size difference, as well as by distortion of geographical space. Because the order of geographical relation is preserved as in most of cartogram, the overall pattern is clearly revealed. For example, one could easily spot out on the map that South America received very little attention in BBC news (See Figure 3.7).

In *News Coverage Map*, categories are represented through color coding. The attribute of color is not intrinsic to the object and has very little capacity to describe the object's shape and spatial relationship. However,

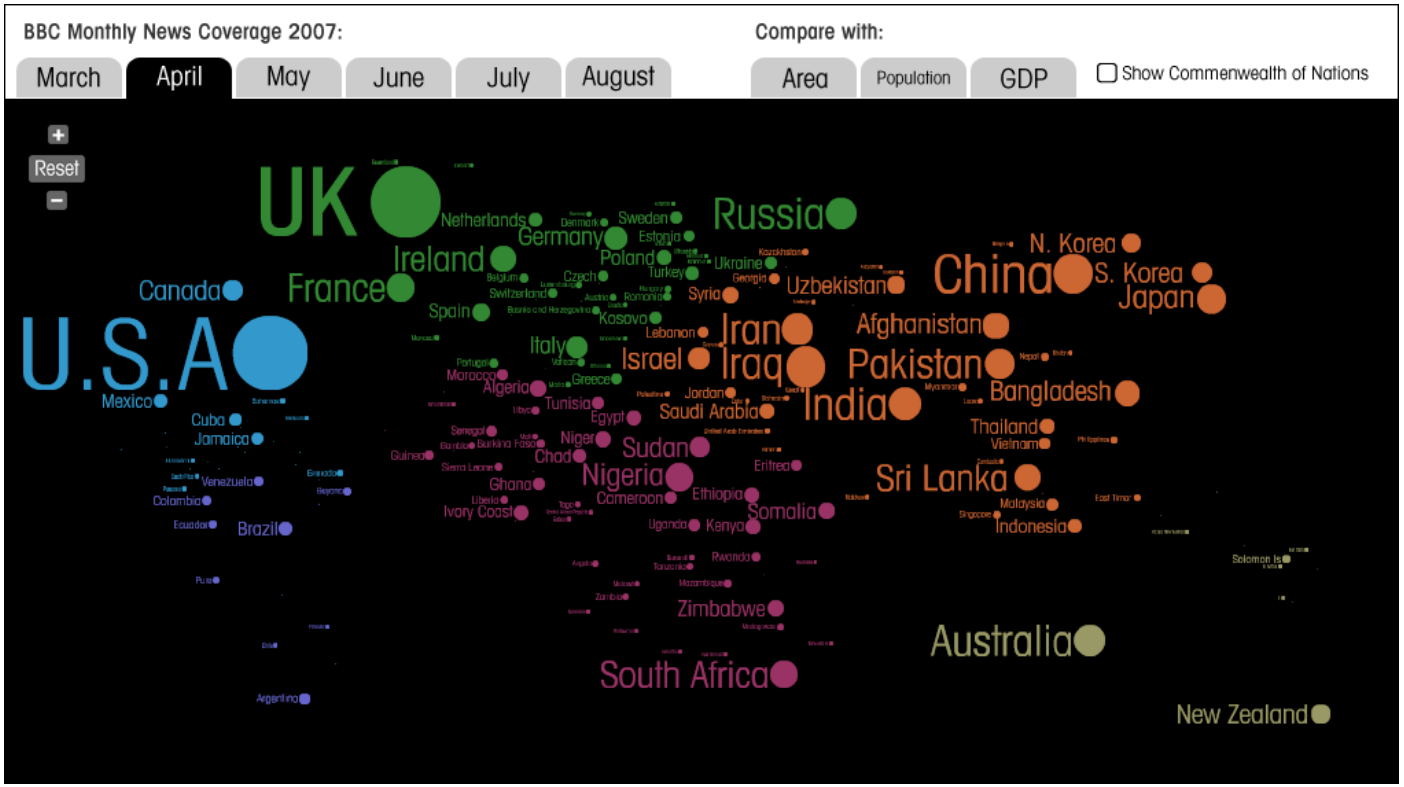


Figure 3.7 Screen shot of *News Coverage Map*. In this view, the size of a country displayed in the map is proportional to the amount of news stories reported about a country in BBC online news during the month of April, 2007

when classifying visual objects into separate categories, giving the objects distinctive colors is often the most effective solution (Ware, 116), as we can see in Figure 3.8 and Figure 3.9.

Each visual variable has different properties and capacities for portraying given types of information (Bertin, 42-43). The choice of visual variables is dependent on the types of data variables. Visual variables as pure decorations do not contribute to readers' understanding of the representation. In contrast, they result distraction, misleading or confusion. In visualization, a meaningful visual variation has to correlate with a data variation, to present the data variable in a perceivable visual form. This design principle applies universally to all types of visualizations that focus on communicating the message of data.

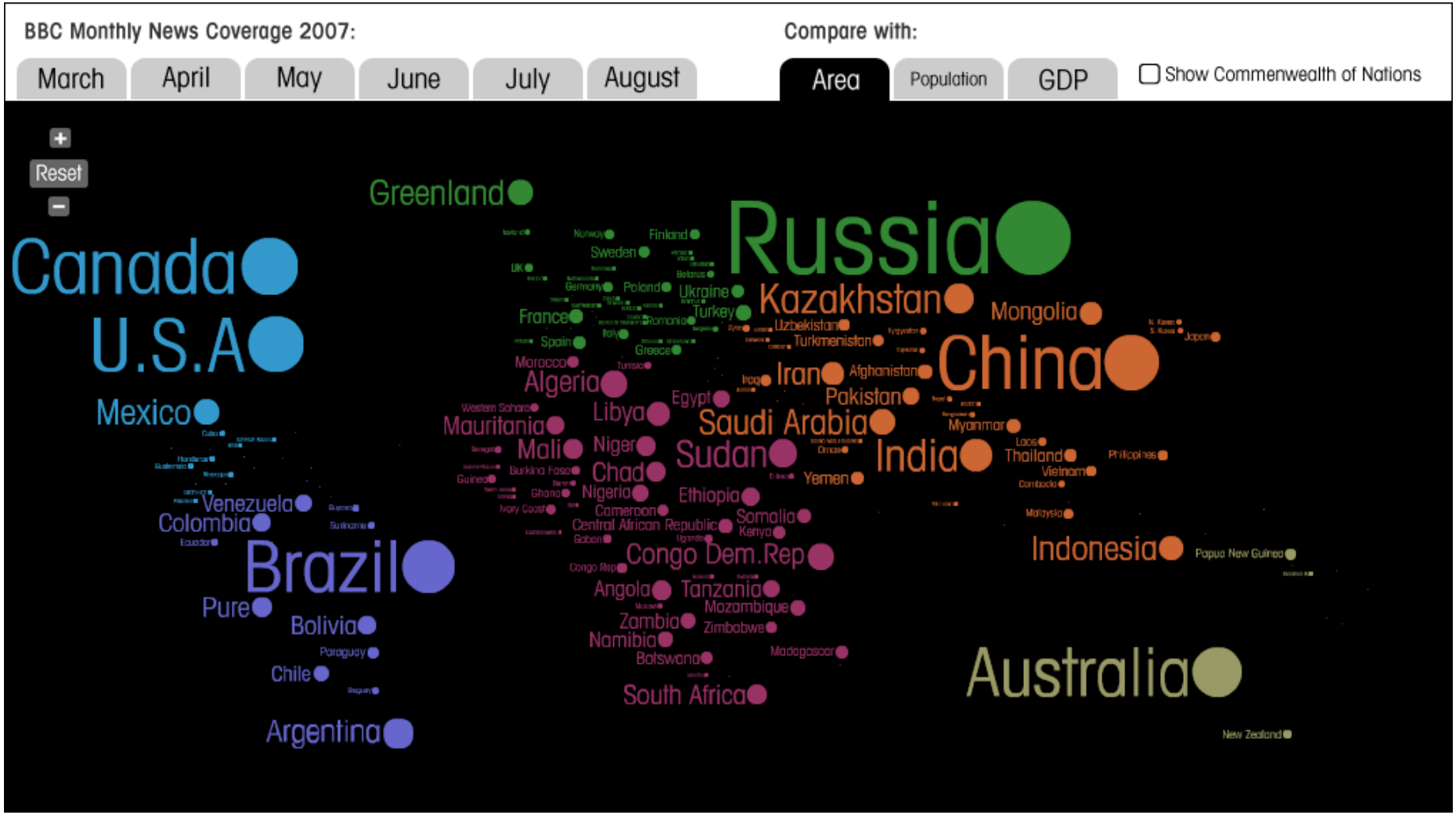


Figure 3.8 Screenshot of *News Coverage Map*. In this view, the size of a country displayed in the map is proportional to the size of geographical territory

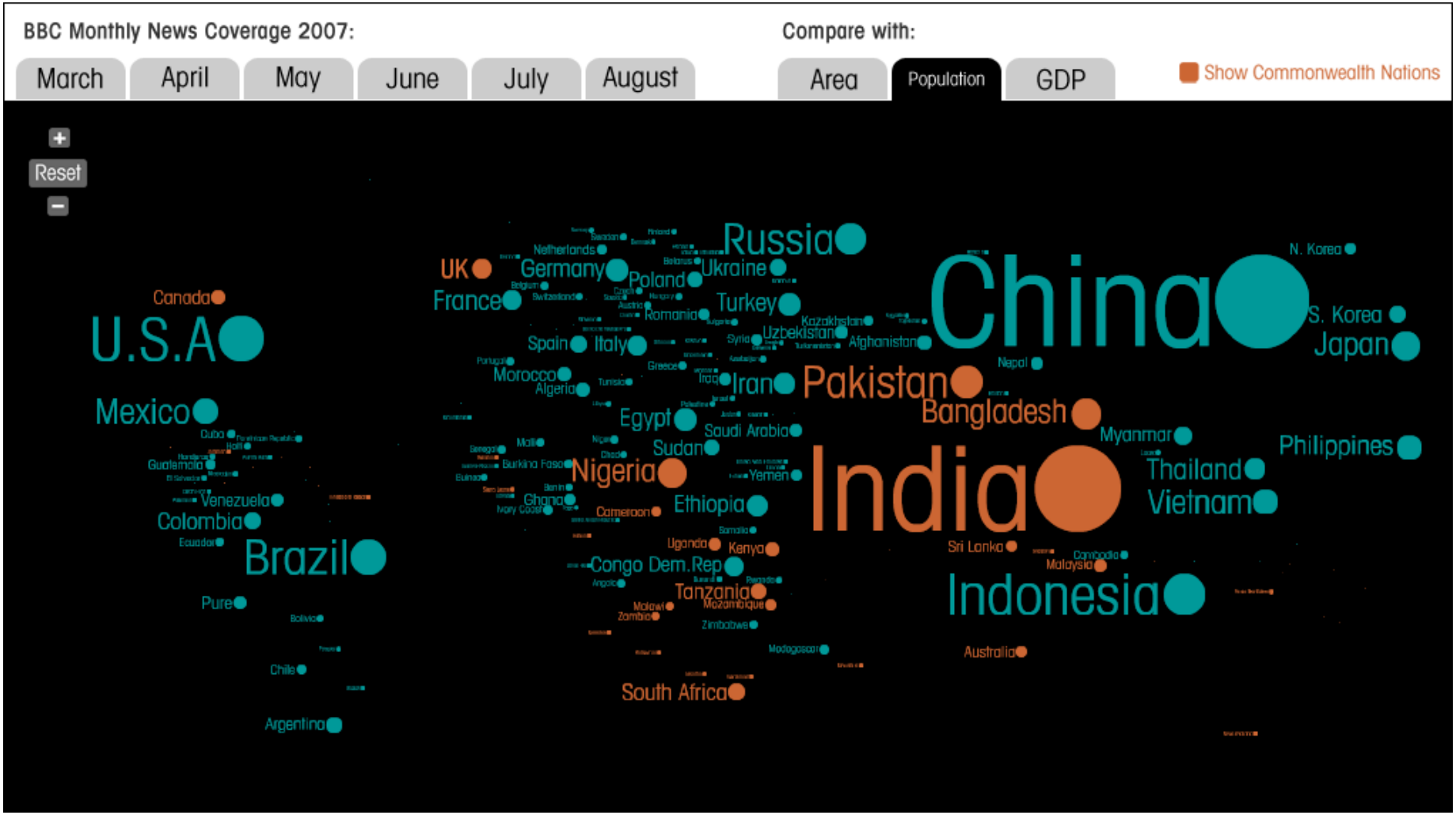


Figure 3.9 Screenshot of *News Coverage Map*. This screen shot illustrates the use of color coding for displaying categories. The countries with orange color are members of the Commonwealth of Nations.
The size of a country displayed in the map is the same as in Figure 3.7. It is proportional to the amount of news stories reported about a country in BBC online news during the month of April, 2007.

3.3.3 Minimize visual ambiguity

The design of basic visual elements in *News Coverage Map*, tries to minimize the problem of visual ambiguity. The decision on using text to represent countries/areas instead of the geographical shape is due to the fact that the name of each country/area is easier to recognize than the geographical shape of the country or any other abstract visual elements (most of readers might be familiar with the picture-gram of the country "France", but few might recognize the country "Tunisia"). The text of country's name conveys information. At the same time it functions as label for visual elements. In addition, the geographical shape does not contribute to clarify data variation in this particular visualization. On the contrary, it might introduce visual ambiguity. For example, in the case of displaying geographical size, it is hard to spot the countries that are similar in geographical size but drastically different in geographical shape.

Using a country's name instead of the image of geographical shape solves the problem of the difficulty of labeling numerous visual elements on a single screen. However, it introduces a crucial problem for conveying data in precision when size is used as a visual variable. Because the length of the country's name varies, countries with a name in great length would appear larger visually than the actual proportional quantities. To adjust this visual distortion of data variables, a dot element has been added behind each text element. The size of each dot follows the data variable as the size of the texts does. With the aid of this visual element, readers can compare the data which has been depict with more precision. As mentioned in section 3.2, a typical cartogram suffers the problem of identifying individual countries because of the distortion of the geographical space. *News Coverage Map* as cartogram does not have the same problem, precisely because of the design of the basic visual element emphasizes the visibility of the label.

Labeling visual elements is a primary way of avoiding visual ambiguity. In abstract visualization, visual elements are context-dependent rather than visually descriptive. For example, a simple diagram cannot convey any meaning without proper labeling to put each visual element in contexts.

Furthermore, interpretation of visual elements is sensitive to people's past experiences and context-dependent. Miscommunication is a problem that not only occurs in design, but also in everyday life (See Figure 3.10). Labels can function in a way similar to language in everyday life: to clarify meaning. (Tufte, 2001, 55-56).

Fig 3.10 Drawing by CEM (Tufte, 2001, 56)

The importance of labeling visual elements is often ignored in screen-based visualizations. In some cases, the labeling system is only described in "help" or "project info" documents; in others, due to screen size limitation and numerous visual elements presented in a single screen, the labels are only visible when readers use mouse hover over an visual element. In the latter case, one effective adjustment is labeling the important elements in the visual representation. A good example of this can be seen in the project *Name Voyage*. As shown in Figure 3.11, the image on the top shows the trends of boys name from 1880 to 2006. A few visual objects were labeled. These labels give a glance of what are the most popular names for boys in a time period and how the trends change. The image on the bottom shows how a label is displayed when a mouse over action is applied to a visual object which the label is initially invisible.

Although each design task has a set of its own particular problems, there is no universal solution for avoiding visual ambiguity. The solution of labeling visual elements in *News Coverage Map* or *NameVoyage* may not be suitable for other visualizations. When designing screen-based visualization, labels should not be treated as secondary information or additional explanation. The combination of image and words to display information in coherence with clarity needs more attention and further exploration in screen-based visualization.

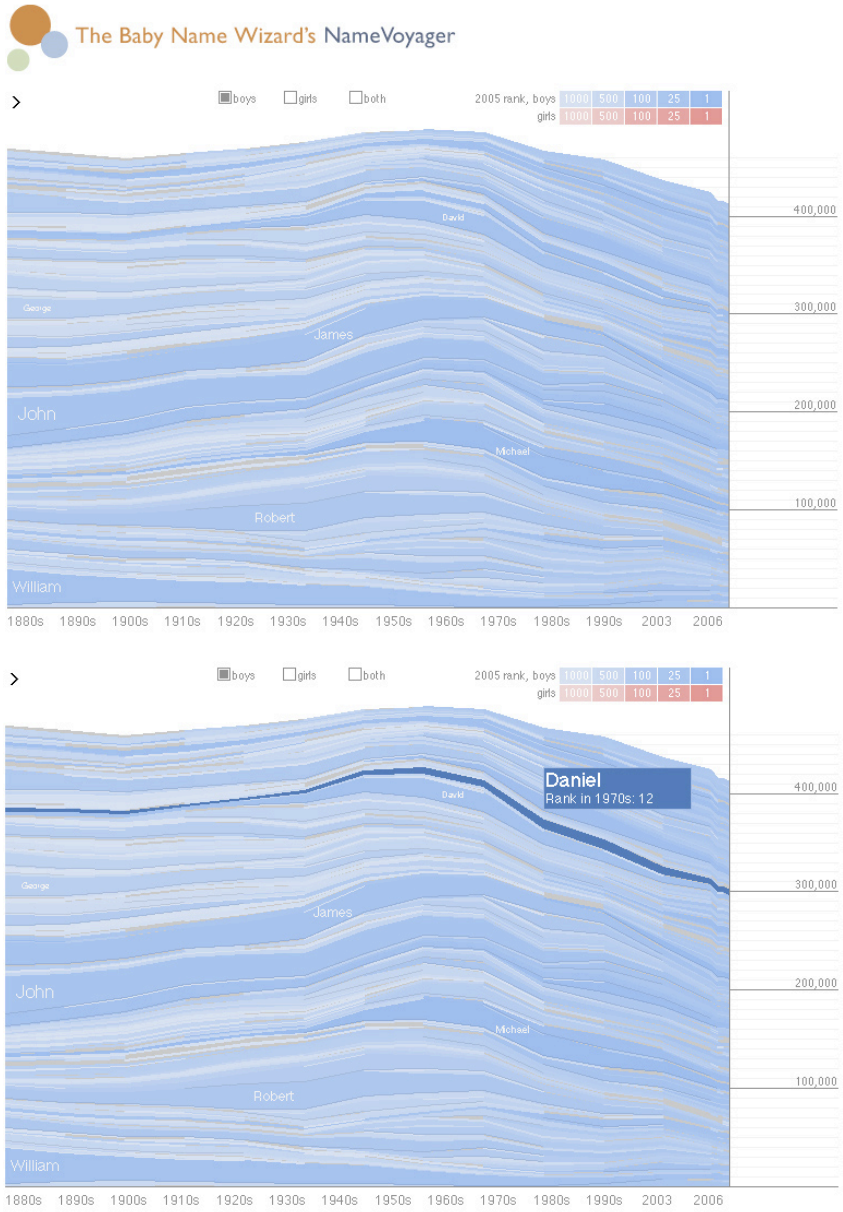


Figure 3.11 Screenshots of *NameVoyage* (The Baby Name Wizard)

3.3.4 The use of animated transition

A shift from one view to another in a visual display, I call it "transition". An important feature in *News Coverage Map* is the animated transition between different views that displays data variables and data values. In this case, Animated transition include two components: motion, and resizing a visual element according to different data value. When interacting with *News Coverage Map*, the readers can clearly see that some countries are expanding by increasing their size, and pushing away the neighbored countries. Some other countries are shrinking and disappearing from the map (See Figure 3.12). The use of animated transition here does not represent a data variable as such, it works as an adjective: it enables readers to perceive correlation between data values and it emphasizes changes.

Motion is interesting in information visualization because we are sensitive to motion cue. The real world environment is not static; things are changing around us all the time and human eyes develop high sensitivity to these changes. One good example which demonstrates the sensitivity of human perception towards motion is the experiment "Dalmatian Dog" (Bach). The experiment shows a static image and an animation. In the static image, we only perceive black and white dots arranged randomly. However, when these dots move in a certain way a walking dalmatian dog can be perceived.

Although the perception of dynamic patterns is not as well understood as the perception of static patterns, studies have shown that animated transition has a great potential to facilitate visual communication in information visualization (Shanmugasundaram et.; Ware, 219-226). However, the use of animation in abstract data visualization needs to applied with care. Animated visual element can easily be distractive precisely because we are very sensitive to it.

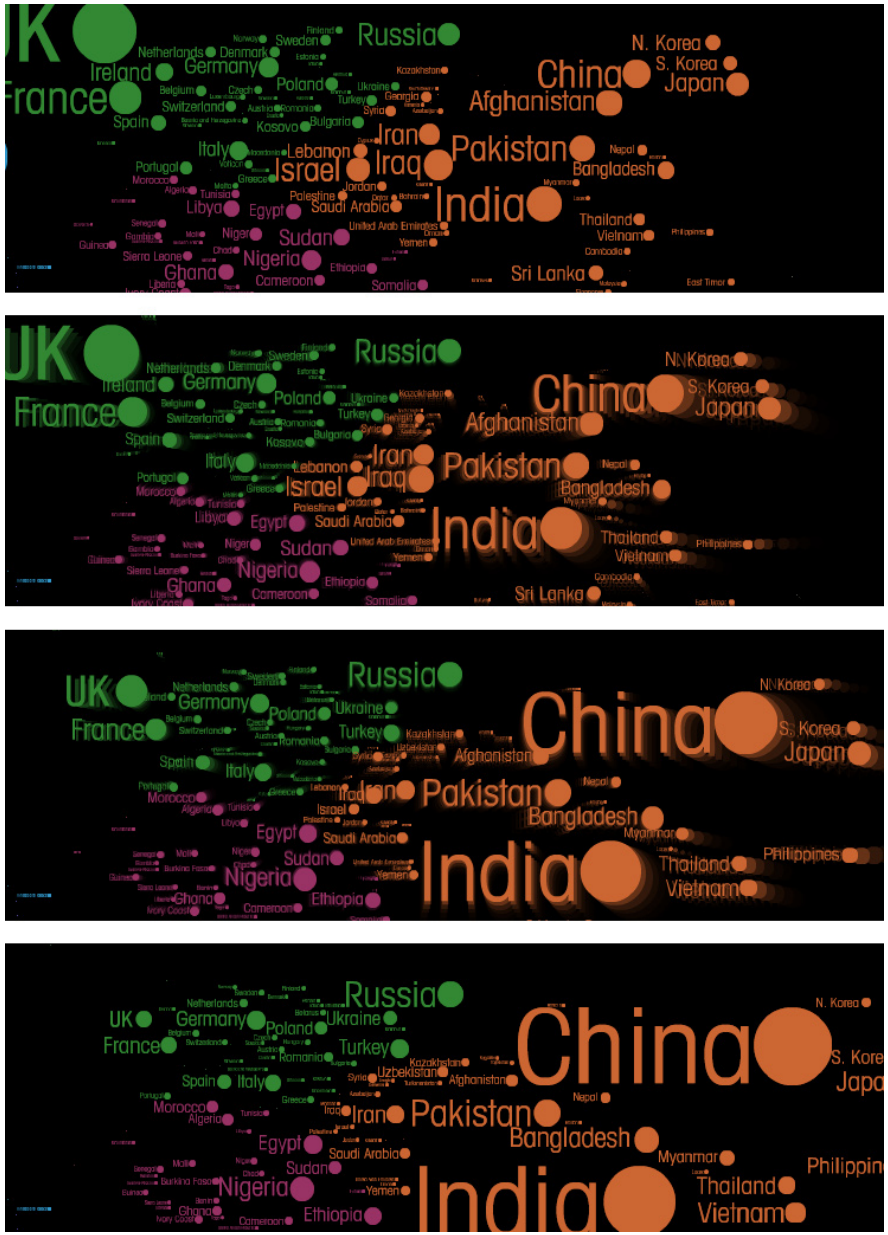


Figure 3.12 Screen shot of animated transition in *News Coverage Map*. The images show the transition from a view of the month of July (top), to a view showing population sizes (bottom).

A potential problem regarding the use of animated transition in *News Coverage Map* is that later frames overwrite previous frames: when the animated transition reaches the chosen data variable, the value of previous data may already be forgotten. The current design solution addresses this particular problem by allowing the readers control which view is displayed, so they can always go back and see the previous one. A more optimal solution could be to have a function that allows readers to select two views and the application would automatically switch in between both of them.

One attempt of addressing this problem during the design process was to display several maps in parallel. This visual technique is what Tufte called "small multiples". As he further explained in the text, "small multiples" is "a series of graphics, showing the same combination of variables, indexed by changes in another variable" (Tuft, 2001, 170). However, the technique of "small multiples" has limited use on screen-based visualization because of the limited visible space of the screen. "Small multiples" requires space for displaying each frame with readable size. Another crucial problem made me realize that "small multiples" may not be suitable for *News Coverage Map,* because resizing visual objects also results on the change of positioning of the objects on display. Without motion correlation, it is hard to spot the changes at a glance.

The comparison of the animated transition and "small multiples" addresses an interesting question: the advantages and limitations of different media for graphic design. Digital media enables displaying changes in time and in contrast, print media enables displaying changes in space. The use of animated transition in digital media emphasizes the change between views and the use of multiple views while print media provides each step of change in the same time. This suggests that some graphic design challenges are medium specific. It also indicates different design strategies are needed if *News Coverage Map* was designed for displaying in print form.

3.3.5 Navigation and user interface

As Norman puts it (2002, 189-190), there are three distinctive aspects of mental models regarding the design and the use of design products or systems:

1. the design model, the designer's conceptualization of the system
2. the user model, the user's interpretation of the system
3. the system image, the system's appearance, operation, the way it responds and the additional instructions of the system.

In information visualization, the purpose of the navigation is to clearly state the content of the system, what is the possible path to get from point A to point B and where are the readers located currently in the system. The display form and the basic visual elements contribute to build understanding of the system image as well as expectations of a particular visualization. Navigation is important for bridging these three mental models because navigation reflects the designer's mental classification of the data set, states what the system contains, and suggests user how to operate the system.

In *News Coverage Map*, the navigation is divided into two types: navigation for manipulating the dataset in display and navigation for viewing the map. The first type of navigation allows readers to select a viewpoint from multiple data variables and the second type allows readers to view the map display in details.

Navigation for manipulating the dataset in display:

In *News Coverage Map*, the navigation for switching the display among different data variables is located on the top of the map display. In addition, the navigation was grouped into two categories: A time serials tabs in the top left of the map, controlling the display of the data variables which describes how news coverage changes over time. Another set of tabs located in the top right, controls the display of data variables which indicates the possible correlation of the news coverage phenomenon.

Navigation for viewing map:

The navigation for viewing the map is displayed on the top left corner of the map. This follows the convention of current map display interfaces on the web, such as Google map and Yahoo map. Since the screen size and resolution is limited, zooming and panning are used as aids for a detailed map exploration. The possibility of switching among different datasets while the map is zoomed in provides a better comparison. The reset button takes the reader back to the overview instantly when the map is zoomed in to a large scale.

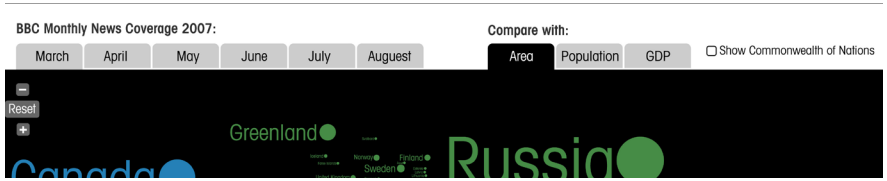


Figure 3.13 The navigation of *News Coverage Map*.

Highlight Mechanism

Highlight mechanism is another an effective navigation aid for data comparison. The readers can focus on reading the smaller set of data without missing the overview.

In the current prototype of *News Coverage Map*, readers can select to highlight a certain country by mouse clicking a visual element, and switching between different views to observe how does this particular country's data value changes in different views. A further implementation of highlighting mechanism would be allowing user to select multiple countries, to make comparison of different data values in these particular countries.

Figure 3.14 Two screenshots of *News Coverage Map* with the highlight mechanism applied.

3.4 Discussion

3.4.1 Results of two prototypes

The design of *News Coverage Map* started by presenting three questions that are related to the issue of news coverage by country and whether there is correlations between the news coverage and other country's dataset. The following is the discussion of how the visualization addresses these questions.

Results of news coverage in all articles monthly

The result of this prototype clearly illustrates the overview of the focus of BBC online news: a broad news coverage in most parts of the world in continental level, but very little coverage on news report about South America in general. In relation to monthly news coverage comparison, one might think that news coverage would change dramatically from month to month according current events. However, the prototype demonstrates a contradictory story to this assumption: the overview of news coverage from month to month are surprisingly similar, especially the size of each continent stays almost the same.

Although the overall pattern of news coverage are relatively similar from one month to another, there are interesting changes among countries within the same continent. The size of the country varies from month to month. These changes capture our eyes, because the change is displayed in motion. Furthermore the map also demonstrates that some countries always receive heavy news coverage, such as such as USA, UK, Iraq UK, US, Australia, and countries in Middle East. Other countries need to have something unusual occur in order to receive media attention, such as most of countries in Africa and South America.

Result of news coverage of front page articles monthly

The news archive is collected from the web, it may be interesting to look at front page news coverage since the front page articles of a news site is often what catches readers eyes. The results of this prototype show clearly how most of the time, African continent appears smaller than in the news coverage of all article comparisons. In contrast Asian continent appear larger. The map also reveals more dramatic changes by country comparison.

Compare news coverage with country's dataset

Although the visualization well describes the phenomenon of BBC news coverage, it falls short in visualizing the correlation factors of BBC news. It is interesting to compare dramatic difference between the view displaying news coverage and the views displaying the geographical size of the country, population per country, and GDP. These differences triggers curiosity and raise questions. But it does not describe patterns of correlations between news coverage and these other country's datasets. This indicates the complexity of the different news coverage by country may not correlate with a single factor but multiple combinations.

3.4.2 Further research

A further development of *News Coverage Map* can be taken in two directions:

1. Explore News coverage in different news websites

News Coverage Map shows the strength of displaying the phenomenon of a certain news website's global coverage. The same visualization can also fulfill the needs to compare the presence of countries in different news websites in a given time span, and to explore how news coverage differ from one news website to another.

2. Develop *News Coverage Map* as a general tool for displaying quantitative data variables which use country, city or region as measurement units.

With the countries shrinks and grows from different months in *News Coverage Map*, one might ask question such as "What happens in Cameroon in May?" or what topics makes a country's size grows significantly on the map in a given month? These question will partly addressed by the following chapter: Visualizing news content.

4 Visualizing news contents

News Coverage Map, discussed in chapter 3, illustrates the BBC news coverage by country over time. However, it tells little about the actual contents of the news. In the following experiment, the goal is to look beyond "where does BBC online news devote their attention" and to capture the topics of news as well as to explore how certain places and topics have been portrayed in the news.

Most tools for handling news archives have focused on the task of finding specific news articles. For example, news websites usually provide a "search" function that allows readers to look for articles by keyword retrieval. Less attention has been paid to the overall patterns of news contents that accumulate from daily news reports over time. The patterns of news contents presented by news media over time are themselves significant to the readers. Providing an overview of news contents may help with the following:

1. to contextualize news contents to regular readers
2. to form an impression of the news source for the readers who are not familiar with it.
3. to find news articles that readers only have vague memories of.
4. to reveal information which the reader missed out on when it was current, but which he/she would be interested in exploring.

Three visual design demonstrations are presented in this chapter. Although each is designed with its own approach to the subject of visualizing online news archives, these design demonstrations continue the theme of working with textual contents from the BBC news websites. The three visual demonstrations are:

1. *Newsness Map* whose objective is to explore the common and distinctive newsworthy topics in different continents.
2. *Keywords Chart*, through which I explore the common and distinctive news topics, as well as the temporal pattern of news topics in different places.
3. *News Proximity Map* in which I seek to analyze topic associations within a corpus of news articles.

These visual design demonstrations deal with information that exists in the form of text within a corpus of documents. They are document visualizations (Spence, 175; Card et al., 409).Visualizations such as these explore hidden patterns that are hard to detect when reading single documents. Unlike the representation of quantitative information, such as in *News Coverage Map* (See Chapter 3), which has been well studied since long before the age of digital media (Tufte, 2001; Bertin), document visualization is a young field that grows out from the need of handling the rapidly increasing amount of electronic documents, such as digital libraries, electronic journals, online conversations, personal emails and the like, available in digital media (Card et al., 409-410).

As the patterns of news contents are hidden between lines of texts, the process of designing visualizations with different approaches is a main part of my voyage to understand this particular data.

4.1 Data gathering and text analysis method

The visual design demonstrations presented in this chapter are in some ways more complex than the *News Coverage Map*. In order to discuss these demonstrations, we need to take a closer look at the issue of data gathering.

Data, information and knowledge are not separate entities, but different steps of a single process. This process describes how the human mentality relates to the world through different stages. Data is relatively raw material. It is collected through observations or measurements, and presented in numbers, statistics, words, voice, visual recordings, and other forms. Data is not the events or objects themselves, but records, descriptions or memories of events or objects (Bateson, xxiv). Data is transformed to informatio through a process of contextualizing the data collections by grouping, classifying and comparing. "Information" consists of descriptions or interpretations of relationships between data. It comes from the way data is presented and organized. Extracting useful information from data by statistical study and text processing is referred to as "data mining" or "data analysis". This is the first step of visualizing information.

In this section I present the method that was used to prepare the data used in creating the visual design demonstrations discussed in this chapter. The method extracts existing metadata and performs text analysis.

The first stage involved extracting metadata from each news article on the BBC news website such as time of publication, topic section and headline. Metadata is the description of data. By itself metadata describes little of the actual content of news. However, this metadata is useful if one thinks of it as an identifier for each of the articles within a corpus.

During the second stage, further text analysis was done with software made for this project. Text analysis is an ongoing research topic in the field of computer science and computer natural language processing. Text analysis software is built upon the foundations of data mining, which uses statistical analysis to pull information from large amounts of data. "Text analysis" deals specifically with written sources, such as email content, news articles, electronic journals and the like. The patterns are extracted from natural language text rather than from structured databases. To further understand the use of text mining in document visualization, one must be aware that most text mining software does not automatically extract meaning from data. There is always need for human interpretation, because the nature of computation is numeric and does not handle natural language as humans do (Hearst). Furthermore, although

text mining is one of the fundamental steps for document visualization, a relatively well- selected data set resulting from text mining does not automatically convert to a meaningful visual representation.

To describe my method of text analysis, The phrase “document collection” refers to a collection of news articles with certain search parameters. The parameter can be time, topical section or a particular keyword. A group of document collections, I call “a corpus”. A corpus is used here to describe all the document collections which are displayed in one visualization.

In my process, the text analysis used involves the following steps:

1. Retrieve a document collection from the corpora. For example, time duration, section, or keywords are parameters that can be used in a search to retrieve a particular document collection.
2. Go through a document collection and extract all the words used in the collection. Words that lose much of their meaning when taken out of context, such as “and”, “if”, “the”, “according”, are ignored.
3. Measure and score word occurrence frequency, and display the frequency alongside each word.
4. Display how many articles are correlated with each of the words.
5. Identify approximately which country the news article is about through word matching: if the country “China” is mentioned in the article, it assumed that the content of the article is about China.

The software developed here is not a mature product and has a number of limitations that result in ambiguities in the data collection. First of all, the algorithm does not recognize synonyms (such as “oil” and “fuel”) and words with the same stems (such as “Iraq” and “Iraqi”). Secondly, the algorithm treats all articles equally in a collection. For instance, there is no emphasis on front-page articles. Thirdly, if the content-parsing deals with the article collection about “oil”, it treats all the articles which contain the word “oil” equally, with no emphasis on the articles that mentioned word more often than others. Finally, the software is not able to determine accurately where the news story is from. The articles mentioning “China” do not always have a story related to China. Nevertheless, the software serves a purpose for my design process: to work with a data collection that will be similar to the results from a sophisticated text analysis algorithm.

It is hard, if not impossible, to design one visualization that captures all dimensions and aspects of a large corpus, such as half a year of news articles. A selection of data has to be made by either reducing the level of detail in order to create a summary, or by identifying a document collection with defined search parameters to work with parts of the data. However, the selection is difficult in this case because the patterns of news contents are hidden between lines of text. The way I approach this difficulty is by starting with some questions which I am personally interested in. According to these questions, I decided on selections of data that might answer those questions. In some cases, one visual design demonstration led to further questions which inspired the design of the next demonstration. I will discuss these three visual design demonstrations in the following sections.

4.2 Visual design demonstration 1: *Newsness Map*

By "newsness", I mean the characteristics of news contents, the topics that are constantly reported by a news organization. In this demonstration, the *Newsness Map*, I am interested in exploring BBC news with two main questions:

1. What kind of topics does BBC news report?
2. What are the common and distinctive topics associated with different continents?

To identify "newsness", one can not rely on daily news content. In this visual design demonstration, the parameter of time - in units of months - is used to retrieve a document collection containing all the news articles related to a continent. All the news articles are treated equally and that means the newer articles are not more prominent than the older ones. The visual demonstration presented in this section uses the data of all news articles about each continent in July, 2007 as sample data.

The following pages contain the visual design of *Newsness Map*. The design is explained through a series of visual mock-ups, still images of a hypothetical dynamic application. These mock-ups illustrate some of the key steps a reader might take when navigating through the visualization.

4.2.1 The Map display

As shown in Figure 4.1, the *Newsness Map* has two main visual elements: a typographic world map representing a corpus of news contents, showing as clusters of keywords arranged by continents on the map, and a timeline below the map, indicating the scope of the dataset.

As mentioned in chapter 3.3.1, the map display form is conventional, and most readers are familiar with the order of continents in a world map. *Newsness Map* makes use of this particular advantage to provide a visual categorization system of the contents without needing extra textual labels. This is different from most of the conventional labeling system on the web, such as the one on the Flickr website, where labels are organized alphabetically.

The map display in *Newsness Map* also provides the location context of each keyword through the spatial proximity of words. The word "nuclear" placed in East Asia on the map describes more than the word "nuclear" displayed in an alphabetically organized list.

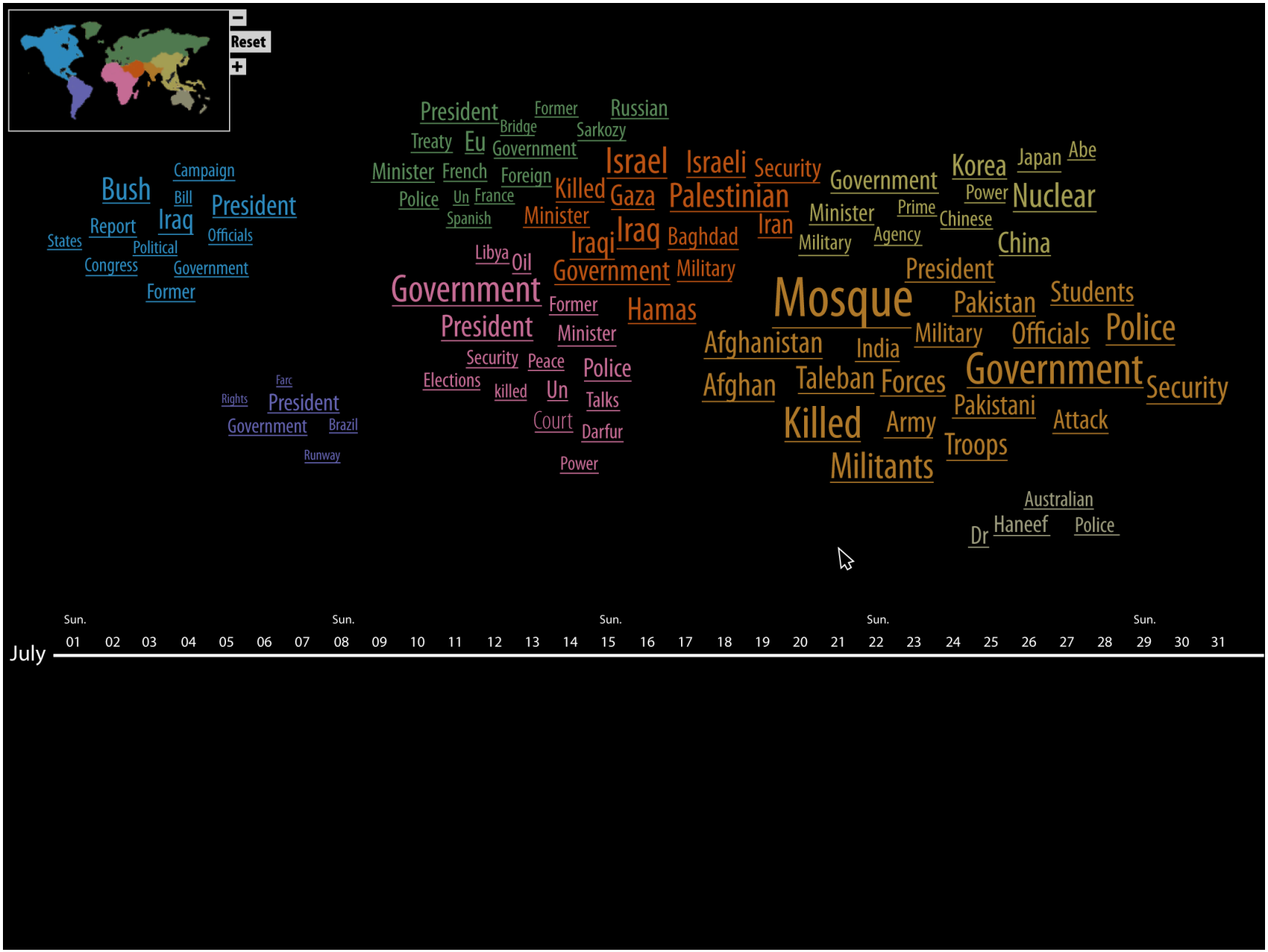


Figure 4.1 Visual mock-up of *Newsness Map*. The image shows the visual structure of the visualization: a typographical world map combined with a timeline

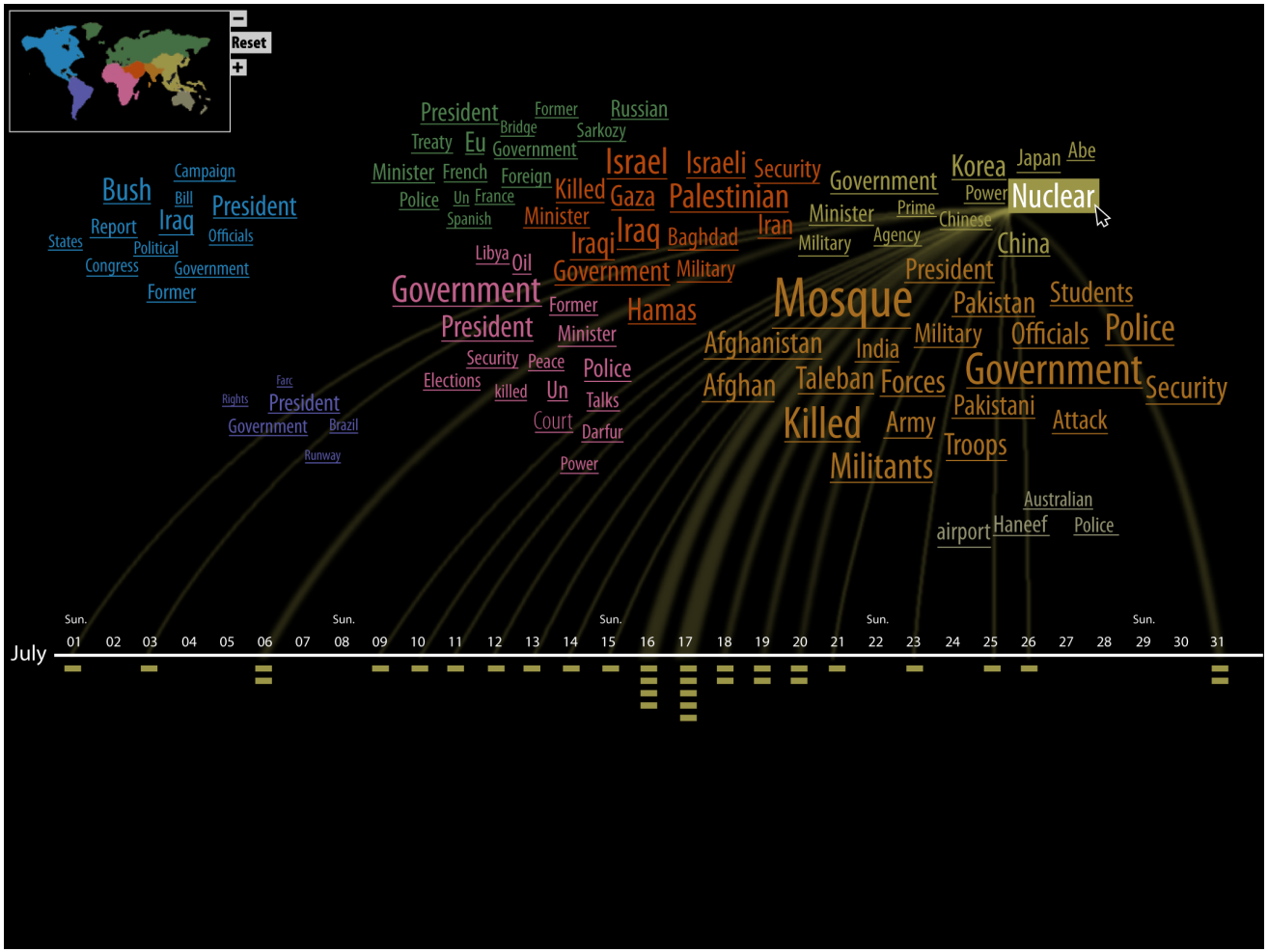


Figure 4.2 Visual mock-up of *Newsness Map*. The image shows the appearance of the visualization when a mouse over action is performed over a keyword.

4.2.2 The Timeline

The timeline format serves two functions in this visualization. First, it indicates the scope of data set. But more importantly, it can display the temporal changes of the occurrence of each keyword within the corpus. The keywords on the map are much like hypertext. One can simply click a word to explore in depth. When a word is chosen, the timeline below the map displays a graphical representation of the amount of articles containing the word in each day. Each article on the timeline display is represented by a small rectangle. These rectangles lead readers on to the contents of the article.

This graphical representation describes whether a keyword is specific to a short time span or if it is a general news topic reported on throughout the month (See Figure 4.2, Figure 4.3).

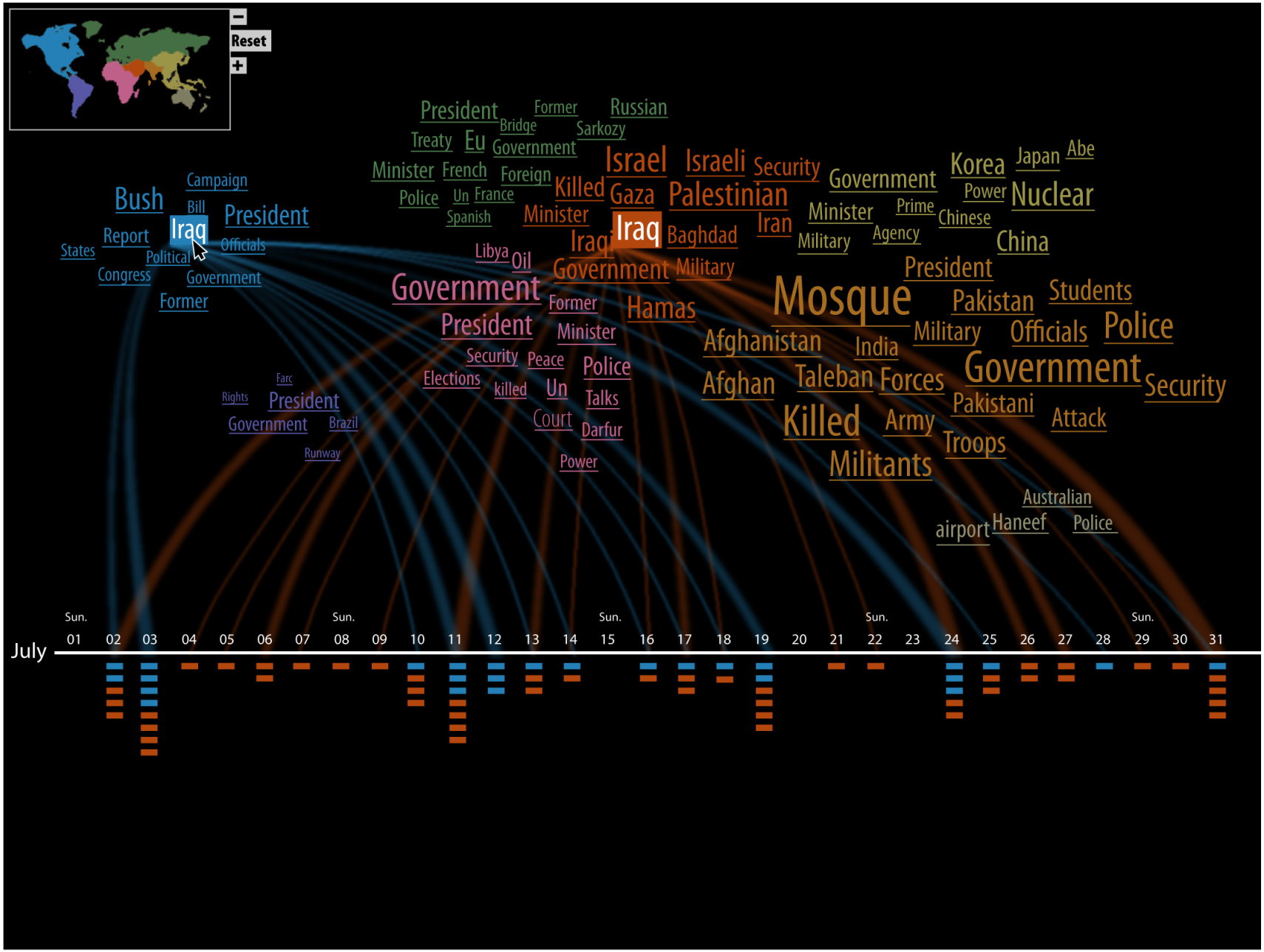


Figure 4.3 visual mock-up of *Newsness Map*. The image shows the appearance of the visualization when a mouse over action is performed over a keyword which is shared by more than one continent.

Semantic zoom

The design of *Newsness Map* is meant to be explored in detail, using the technique of "semantic zoom". Semantic zoom is a technique for balancing the context and detail of graphic representations on screen (Modjeska). In contrast to ordinary zoom functions, semantic zoom goes beyond magnification of graphical representation, and modifies the level of detail to be displayed. As illustrated in Figure 4.4, one can explore a certain area by zooming in on the map. When a zoom action is performed, it not only enlarges the keywords on the map, but also reveals more keywords.

The use of semantic zoom solves a disadvantage of screen-based visualization: lack of screen display space. More importantly, it also provides a possibility to organize information according to different levels of detail. The display can provide an overview of information without needing to show an overwhelming amount of detail at first glance. It also can take input from reader's preference and display the detailed information according to their interests.

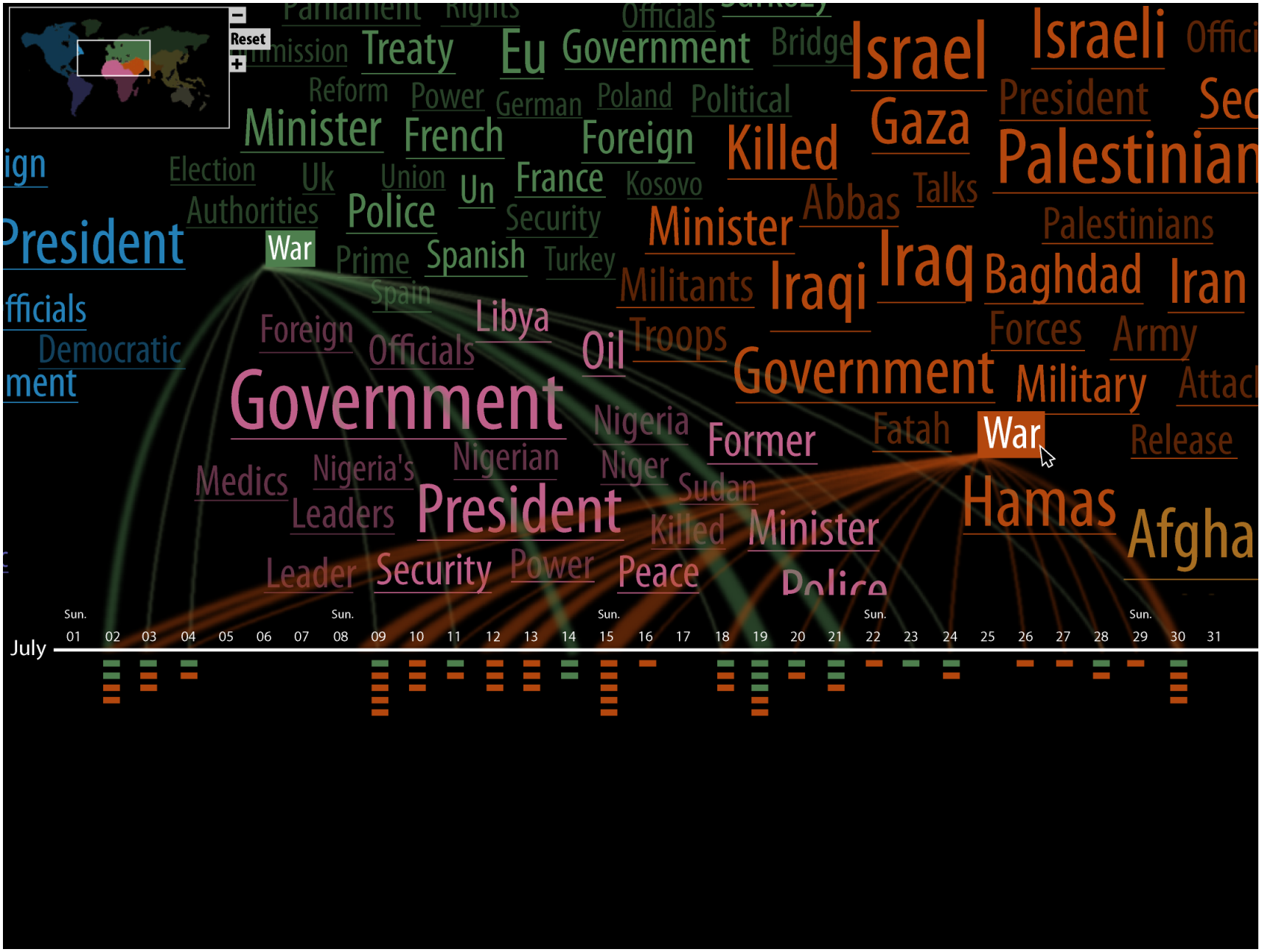


Figure 4.4 Visual mock-up of *Newsness Map*. The image shows the visual appearance of the visualization when the map is zoomed in to a certain area.

4.2.3 The visual appearance of keywords

In *Newsness Map*, Keywords are shown in different sizes depending on their frequency of occurrence in a document collection. This is based on the assumption that the more frequently used words are more representative of the contents, therefore their appearance should be more prominent in the representation. The number of keywords displayed in the map reflects the size of each document collection. The combination of keyword clouds and map display provide a quantitative comparison of the amount of news in different continents (the size of tag clouds), as well as qualitative information of news (the meaning of the words). As I discussed in chapter 3.3.3, because the length of the words varies, the use of the size of the words to represent a quantitative data variable introduces visual ambiguity. However, the objective of this visualization is not to compare the importance of each individual keyword in a document collection, but rather to provide an overview of the topics in a particular corpus of news articles.

Each keyword on the map is underlined. Text with underline is an arbitrary visual code, and the meaning depends on cultural and community conventions. Ware (10-18) describes arbitrary visual codes as those aspects of visual representation that has no perceptual basis and must be learned. Furthermore, he points out that the understanding of arbitrary visual codes is persistent and embedded in cultures. This might explain the following example: Underlining text on paper is usually understood as emphasizing certain words or sentences. On the web, however, the text would be interpreted as a hyperlink. This visual appearance of hypertext is a convention, which is applied by most web browsers and is a familiar visual code for most Internet users. If a designer chooses to emphasize certain texts by underlining them in a web page, the visual representation could be mis-interpreted as hypertext. Likewise, if a hyperlink is not underlined in a web page, it is far less recognizable as hypertext. Since *Newsness Map* is meant to be displayed on a screen, the use of underlined text indicates that these keywords are similar to hypertext on a web page: they lead readers to additional information about words on the map, and they encourage further explorations.

4.2.4 Discussion

The map display form and keyword highlight mechanism facilitate comparison of common and distinctive keywords in different continents. For example, the word “nuclear” is unique to East Asia, as shown in Figure 4.2, while the word “Iraq” is shared by North America and Middle East, as shown in Figure 4.3. Some words are shared by almost all continents, such as “government”, “president” and “security”. These words may not be very descriptive of events as such, however they depict well how BBC news differs from entertainment-oriented or technology-oriented news.

One interesting discovery made through the timeline display is that, with few exceptions, most of the words displayed on the initial map (before zooming) are ongoing topics being reported by BBC, rather than single incidences. They relate to articles that appear in different dates throughout the month, although with different temporal patterns (See Figure 4.2, Figure 4.3). This indicates that the method used in *Newsness Map* may not capture well the breaking news stories, the unusual events within the news. However, it serves well as a summary of the “Newsness”, the types of news that BBC reports.

The results of this visualization may not provide readers who are familiar with the BBC news webiste any additional unknown patterns. It is more useful to the readers who are not familiar with the site. The visualization

may help the readers to form an impression of the overall news contents, and may inform their decision of whether this news content provider is relevant or interesting to them.

A crucial drawback of *Newsness Map* is the visual ambiguity: the abstraction of a continental level containing keywords may be misleading. In many cases, a keyword is related to multiple countries, which are not necessary located next to each other geographically. For example, if a keyword occurs both in news about Norway and Italy, it would be difficult to determine where on the map the word should be placed.

4.3 Visual design demonstration 2: *Keywords Chart*

The problem of visual ambiguity that was brought on by the use of a map display in the *Newsness Map* leads to the following experiment: *Keywords Chart*. Instead of using a map display, *Keywords Chart* experiments with a different visual display form: words were piled into stacks according to their categories. This visual appearance, at certain degrees, resembles a bar chart display. However, by presenting words in a stack without borders, the visual representation is clearly distinguished from a traditional bar chart. This visual distinction is essential for precisely communicating the order of information within this visualization: in *Keywords Chart*, the primary objective is to provide an overview of common and distinctive topics in different continents, rather than a quantitative comparison of the amount of articles. Since bar charts are normally used for quantitative comparison, an appearance that is too similar to a conventional bar chart would be misleading.

As we can see in Figure 4.5, *Keywords Chart* provides an at-a-glance overview of the similarities and differences between news topics covered during a given month, sorted by continent. In contrast with a normal list of words, the different heights of the stacks, and the size of the words in each stack, provide heterogeneous visual cues. This makes it easier to re-find a specific stack of words. Furthermore, these two visual variables provide a layer of subtle information: the height of the stack reveals the amount of news articles in each document collection; while the size of the text reveals the relative importance of each keyword.

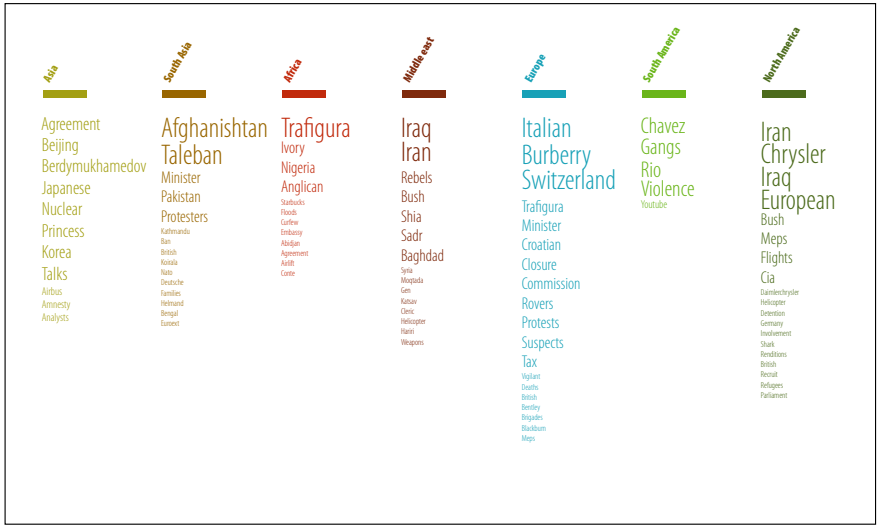


Figure 4.5 An early sketch of *Keyword chart*

4.3.1 An inspiration: *Themail*

The project *Themail* (Viegas) was an important inspiration for the *Keywords Chart* concept. *Themail* is a visualization project which depicts the temporal changes in a person's email conversation. The visualization uses a similar visual display form as *Keywords chart*: it arranges keywords, extracted from email conversations, into stacks (See Figure 4.6). One distinctive aspect of *Themail* is that the stacks of words are arranged in a time series. This opened up my understanding of the potential of this particular visual form: its flexibility. The visual display form is not tied to one kind of information organization principle, but rather, it enables different types of organization principle to be displayed.

Figure 4.6 Screenshot of *Themail*, the visualization depicts a person's email conversation with a friend during the last 18 months.

4.3.2 Beyond the flatland: *3D Keywords Chart*

The flexibility of the Keywords Chart visual form in displaying variety of data sets drove me to further explore the possibility of displaying multiple charts at the same time. For example, we might want to get an overview over BBC news reporting for given continents over time, while simultaneously having an overview of the common and distinctive topics of different continents. However, with a 2D representation, the limitation of the visible screen space made it difficult to display more than one chart without loosing overview.

A third dimension was introduced for the visualization in Figure 4.7. This visualization is a composite of two charts: One represents the time axis, displaying keywords extracted from news contents sorted by month; the other represents a geographical axis, displaying keywords sorted by continent. The 3D representation inherited most of the properties from the previous 2D representation, such as the overview of common and distinctive keywords among stacks, and the heterogeneous visual cues . In addition, it made it possible to display a greater density of data.

Navigating the *3D keywords chart*

In *3D keywords chart*, the point of intersection determines the content of both axes. By changing the point of intersection, the reader will select different data to display. To change the intersection, the reader can drag the red point along one axis. Directly dragging the point of intersection is a natural way of navigating this particular visualization. However, the drawback of this technique is that the "dragable" function of the point can not be easily perceived. One must know it in order to use it.

One of the two axes is always semi-transparent. Transparency was used to clearly separate the two axes, making it easier to focus on the information in the highlighted plane, while having effective preview of the other. The reader can choose to highlight either axis according to their interest.

Figure 4.7 interface mock-up of the *3D Keywords Chart*. In this visualization, we see news topics relating to the Middle East (highlighted axis) sorted by month, and news topics from the month of June sorted by continent (transparent axis).

Non-perspective projection

Although the 3D Keywords Chart uses a form of 3D representation, the visualization excludes the perspective cue. As Ware points out there is little evidence that a perspective image tells more than a non-perspective image in terms of data density on screen. He further explains this by showing the following example (See Figure 4.8).

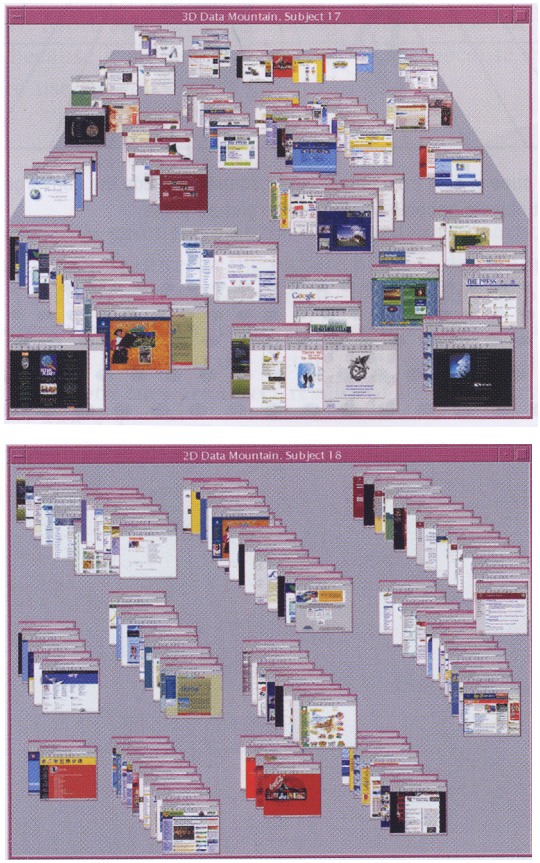


Figure 4.8 An example of a 3D representation with (top) and without (bottom) perspective (Robertson et al, 1998, cited in Ware, 262-263).

One advantage of non-perspective projection is that it not only indicates the foreground and background relationship of visual elements, but also preserves the details of foreground and background on a uniform plane. As shown in Figure 4.7, the *3D Keywords Chart* excludes the perspective cue for the purpose of displaying an equal amount of detail of each word stack on a single axis.

4.3.3 Discussion:

The information presented in this chapter is abstract. It does not form a familiar image that associates it with a phenomenon we might see in a real environment. For example, the use of the third dimension in *3D Keywords Chart* does not aim to achieve a sense of presence or engagement, but rather to effectively communicate abstract information with optimal use of screen space.

There are two problems related to the design of the *3D keywords chart*. First of all, there is a limited amount of words that can be listed in a stack due to the constraints of the visible screen space and font size of the words.

Secondly, when the user changes the point of intersection, the new intersection point has to be moved to the center of the screen, in order to optimally utilize the screen space. This means that the position of the stacks has to change in order to accommodate the changed intersection point. For example, if we wish to choose a different set of data by moving the intersection point to “June” and “South Asia” instead of “Middle East” as in figure 4.7, this means that the stacks representing “North America” and “South America” have to be moved to the left, in order to keep the intersect point in the center of the screen. The change of the position of each stack is undesirable in this visualization, as it requires that readers have to depend upon the label of each stack, and can not rely on getting familiar with the spatial organization of the visualization.

4.4 Visual design demonstration 3: *News Proximity Map*

As revealed from *Newsness Map* and *Keywords Chart*, keywords extracted from collections of news articles are multidimensional in nature: they are often related to numerous articles in different collections. The two previous visualizations sorted keywords according to pre-defined categories (months and continents), and the overall visual representation was determined by these categories. Unlike *Newsness Map* and *Keyword Chart*, *News Proximity Map* breaks down the hierarchical structure, and visually explores the semantic relationships, the relationships of meaning, between keywords. By showing the semantic relationships of these keywords in a visual form, *News Proximity Map* is designed to allow readers to freely jump from topic to topic, rather than being constrained to strict categories. By providing an overview of these relationships between articles, one might discover patterns and certain points of view in BBC news.

4.4.1 The use of Hypertext:

As mentioned in section 4.2 and 4.3, in the context of this work, the keywords extracted from news articles are treated much like hypertext. However, very few of the potential uses of hypertext were explored in the previous two visual design demonstrations. In *News Proximity Map*, the use of hypertext is extended.

The hypertext concept was first proposed by Vannevar Bush in 1945 (Bush), while the term "hypertext" was coined by Theodore Nelson in 1965. Horn (Horn, 6) defines hypertext as:

> *"the ability to link any place in text stored in a computer with any other place in the same or different texts, that permits rapid access through buttons and other tools across non-linear pathways."*

In addition, Horn describes six possible user modes for a hypertext system: Browsing, training, briefing, learning and analysis, helping, and referencing (Horn, 34). *News Proximity Map* attempts to provide a tool for the browsing mode. The goal of "browsing" is to quickly skim through a large amount of information to find areas of interest. Supporting the browsing activity is important, because information-seeking is an imprecise process: when readers approach a large information system (for example, thousands of news articles in an archive), they often have only a vague and fuzzy understanding of what they are looking for and how they can achieve their goals.

Current news archive interfaces on the web, although they are hypertext applications, rarely explicitly support browsing activity. Keyword retrieval and topical categorization are two common ways of browsing news on the web. With keyword retrieval, such as that provided by the Google search engine, requires readers to know what they are searching for. It supports the user modes of referencing, when readers are searching for specific information known to exist within the system, rather than browsing. Topical categorization, such as the sections of the BBC news website, does not reflect the complexity of a large collection of documents, nor the non-linear nature of hypertext and the nature of human thought. Vannevar Bush, in "As we may think":

> *"Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing... The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. "(Bush, 44)*

Much like explorers of an unfamiliar city, we need maps. Visualization can be used to aid the exploration of unfamiliar territory in the information space, supporting the browsing activity.

4.4.2 Problems with the Node-Link diagrams

In *News proximity Map*, keywords extracted from news are usually connected with multiple articles from multiple categories. A common design strategy for visualizing these types of relationships is the Node-Link diagram (also known as network diagram). However, the most common form of the Node-Link diagram has a number of problems which are not applicable to this particular dataset. First of all, although the distance and links between nodes convey information in the diagrams, the horizontal and vertical position of nodes in relation to the screen appears arbitrary and homogeneous. As shown in Figure 4.9, the diagrams A and B convey the same information, even though the position of nodes on the screen is entirely different. Secondly, the arbitrary position of nodes makes it hard to re-find a specific node. This difficulty increases with the increasing number of nodes in a diagram. Thirdly, large scale Node-Link diagrams are often hard to understand for readers, since the relationships between nodes is mainly expressed by lines connecting nodes. The overlapping lines within the visualization result in noise and undesired visual patterns which do not convey relevant information.

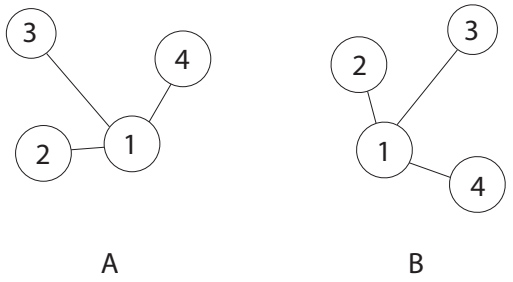


Figure 4.9 An example of Node-Link diagram

4.4.3 An Inspiration: *TextArc*

TextArc (Paley) is a visualization which provides an overview of a collection of textual data, such as a novel. The visualization also reveals the semantic relationships between words within the texts, thereby providing its users with a different reading experience. The most interesting concept in TextArc is the use of spatial organization for visual elements on the screen: the entire text was rendered as an ellipse with a very small font size. The keywords extracted from the text are repeated and placed within the ellipse. The placement of keywords is based on the occurrence of the word within the text. In this way, if a word appears evenly throughout the text, it will be placed in the center area of the ellipse, while if a word appears unevenly, it will be placed close to the chapter where the word occurs most often.

4.4.4 The design of *News Proximity Map*

In *News Proximity Map*, I try to study the news data with a finer granularity: the following visualization uses weekly news as duration parameter, rather than monthly, as in the two previous visual demonstrations. The corpus that is being visualized comprises of all the news articles about each continent during a week. The articles are visually represented by their titles. A large amount of keywords are extracted from the corpus. In the visual mock-up that is presented here (See Figure 4.10), about 300 keywords are used.

Figure 4.10 Visual mock-up of *News Proximity Map*.

Borrowing from *TextArc*, in *News Proximity Map* the article titles and keywords are arranged in the form of an ellipse. The ellipse is mainly intended to utilize the screen display space optimally. Additionally, the ellipse shape suggests a clock metaphor. This leads to the design decision of arranging the news titles around the ellipse clockwise, sorted by time. Between daily news articles, a small black gap was used to separate each day's news article. The titles were drawn in a font size which although too small to read, provides chronological relationships between articles. In addition, it provides information about the amount of news reported daily.

Much like the architectural foundation of a building, the ellipse constructed by news article titles is the foundation for this visualization. With this foundation, the positions of keywords in relation to the screen frame are not random. Keyword positions are determined by the position of articles that contain the word. In this way, if a word appears in the news every day, with an approximately equal amount of articles, it will be located more or less in the center of the circle; Likewise, if a word occurs in news during a specific day, it will be located at the edge of the circle, close to the articles which contain the word.

Navigating *News proximity Map*

In *News Proximity Map*, there are two kinds of distinctive nodes: the article titles and the keywords. The links between nodes are represented visually by lines. However, unlike the conventional Node-Link diagram, the lines are organized in many invisible layers in this visualization. They appear when we wish to see them, when the mouse pointer hovers over a certain node (See Figure 4.11), but are otherwise hidden. In this way, the visualization avoids the readability problems of displaying numerous overlapping links.

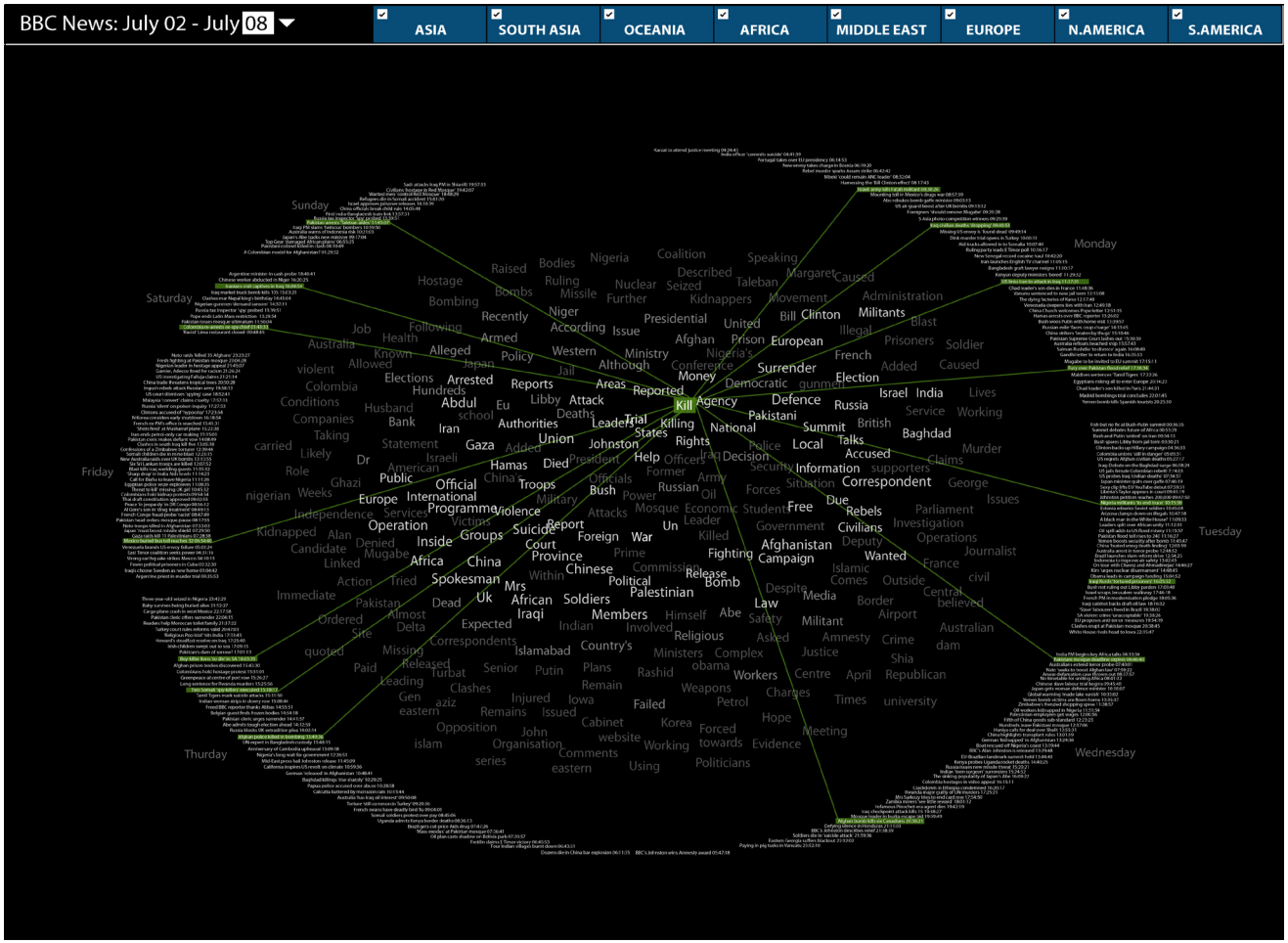


Figure 4.11 Visual mock-up of *News Proximity Map*.

When the mouse pointer hovers over a keyword, the visualization shows the related articles by highlighting them. In addition, the visible links between the keyword and its related articles, emphasizing the connection between these nodes (See Figure 4.11). Since the news article titles are sorted by time along the ellipse, the pattern revealed by the links gives a similar result as in *Newsness Map* which I discussed in section 4.2: it provide a clear overview of the temporal changes of a word occurrence within the corpus that has been visualized.

As shown in Figure 4.12, the article title as node behaves in a similar manner as keywords. When a mouse over action is performed over an article title, the visualization displays the links between the article and its related keywords.

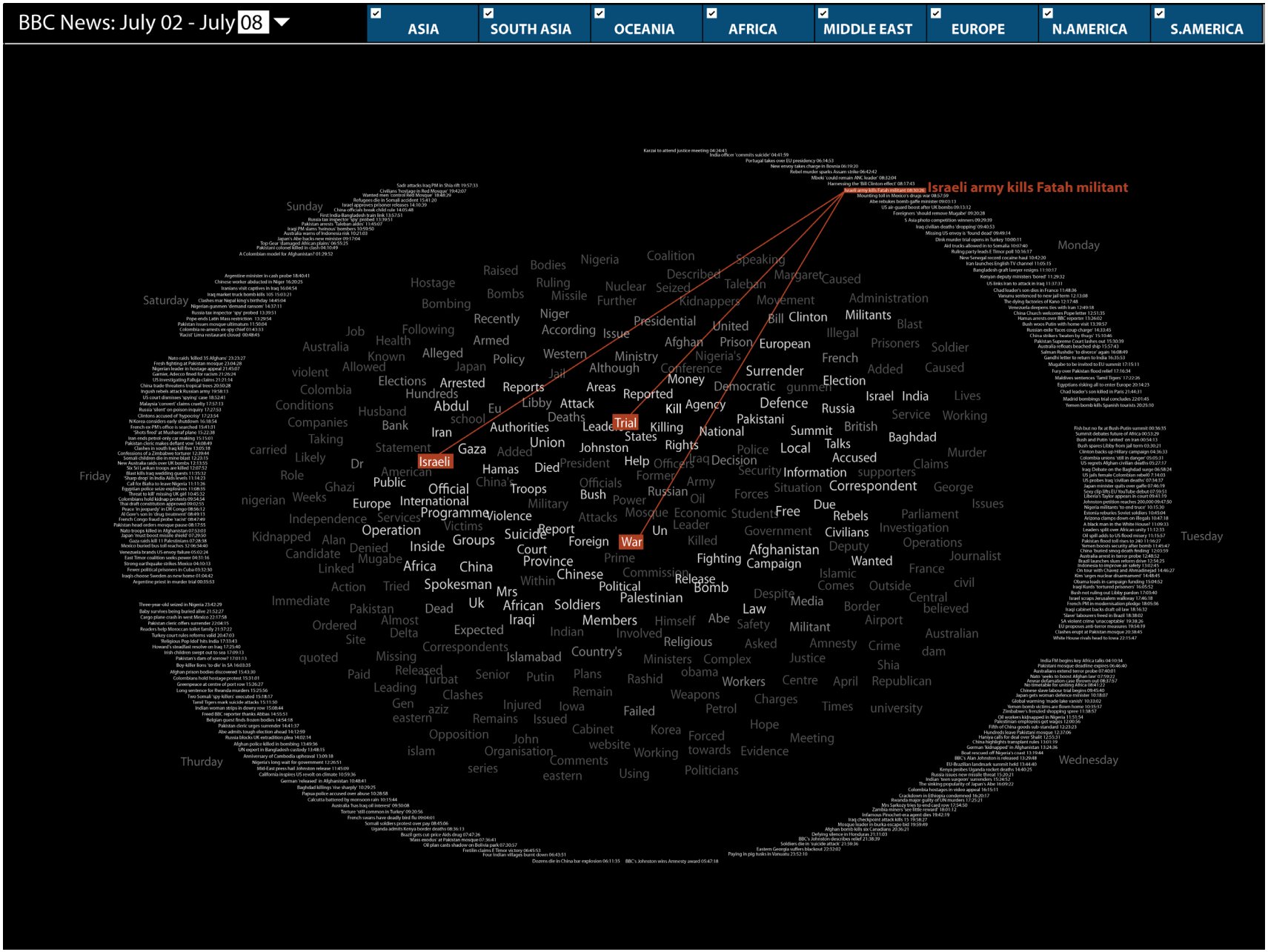


Figure 4.12 Visual mock-up of *News Proximity Map*.

When a mouse click action is performed over a node, its links will remain visible until the next mouse click action is performed over another node. For example, we could click the word “kill” in the visualization. The links between the word and its related articles remain visible on the screen while the mouse pointer is moved away from the word. Now we could use a mouse over action to explore other nodes. For example, as shown in Figure 4.13, we could follow the articles that are related to the word “kill” through the visible lines, to explore further the other words that relate to a certain article. We could also use a mouse over action over another keyword to compare the patterns revealed (See Figure 4.14).

In this way, we could explore the corpus that has been visualized by following the links, jumping from words to articles, and further to other words. This visualization encourages and supports further explorations of the corpus, through both our visual perception and linguistic knowledge.

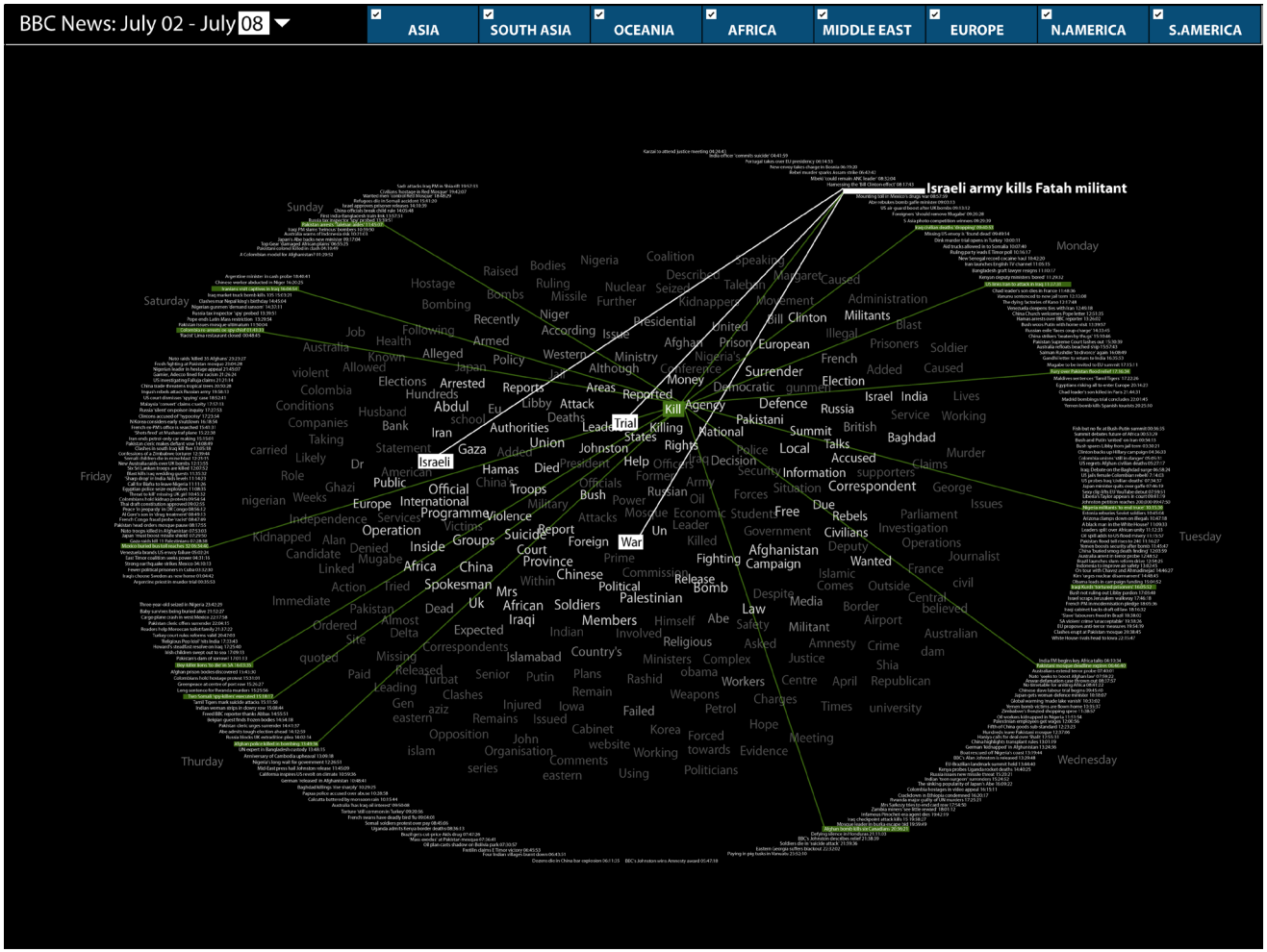


Figure 4.13 Visual mock-up of *News Proximity Map*.

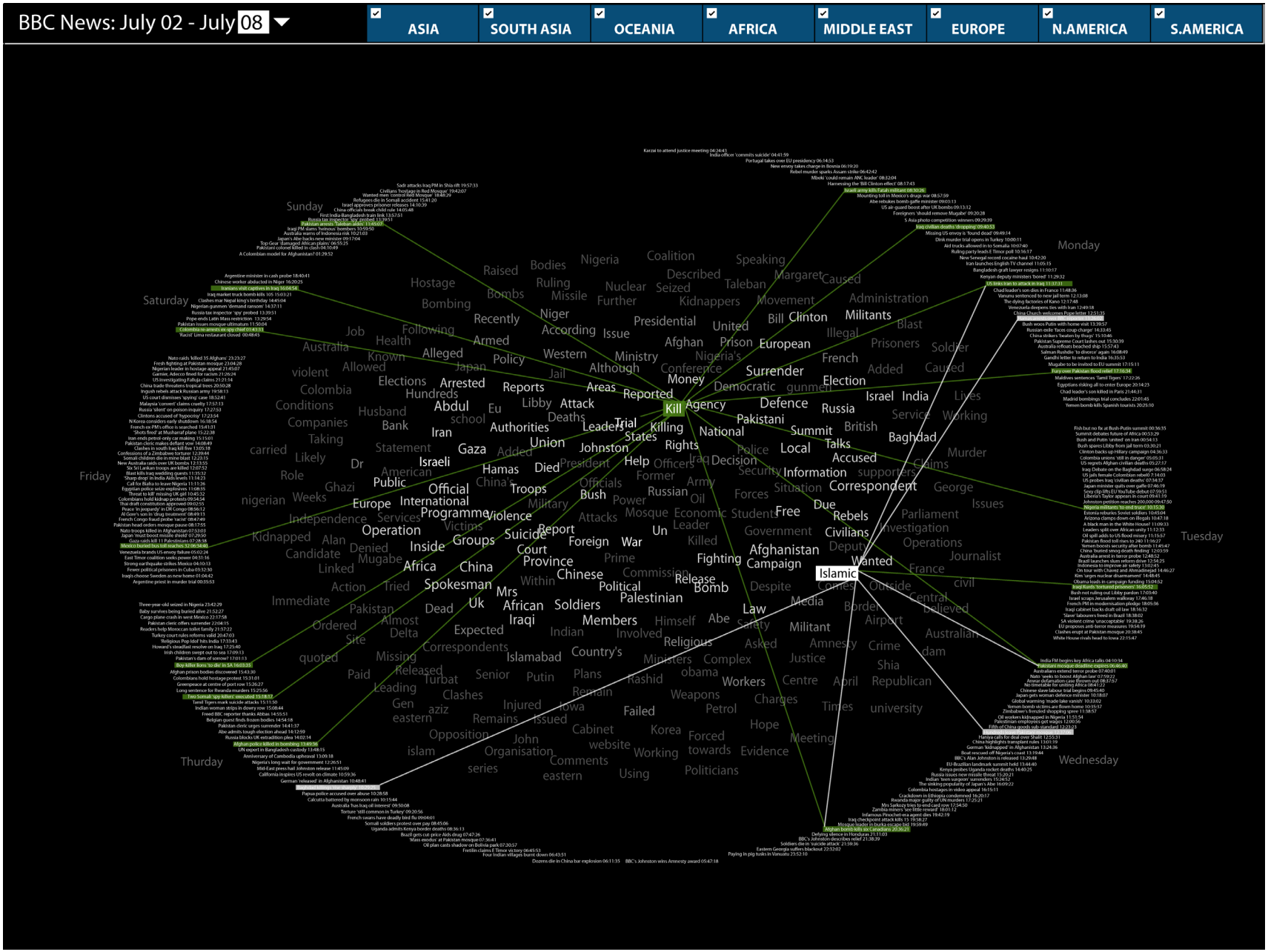


Figure 4.14 Visual mock-up of *News Proximity Map*.

Detail-on-demand

The detail-on-demand principle allows readers to select a visual element or group of elements and get details when they are needed (Shneiderman, 1996). By organizing multiple links in numerous invisible layers and displaying layers according to readers preferences (through mouse over or mouse click actions), this visualization utilizes the detail-on-demand principle for effectively organizing information that has multidimensional relationships.

To apply detail-on-demand in design, it requires that the designer decides on an appropriate level of detail and overview of the dataset in order to support exploring and further understanding. More importantly, as mentioned in section 3.3.5, the graphical representation of navigation is a bridge between the designers mental model, the system image and the user's mental model. The graphical representation of the relationship between overview and the detail view is important, as it describes the information organization principle to its reader.

4.4.5 Discussion

The design of *News Proximity Map* resembles a Node-Link diagram to some degree. However, it avoids some of the problems of a typical Node-Link diagram. By constructing a visual form that puts two types of nodes (the keywords and the article titles) in a semantic relationship, it avoids the arbitrary positioning of nodes in relation to the screen. Furthermore, by applying the detail-on-demand principle, it organizes links into multiple layers, avoiding the problem of overlapping links.

News Proximity Map is currently a visual mock-up. Without the implementation of an algorithm that could precisely calculate each word's position in the design, some of the problems/challenges may not be predictable. The process of designing *News Proximity Map* reveals the difficulty of working individually with information visualization, since several different specializations are involved in the field.

Nevertheless, by creating visual mock-ups of *News Proximity Map,* a number of unsolved problems within the concept are already revealed. First of all, although the ellipse shape is an effective form to utilize the visible screen space, there is a limited amount of articles that can be displayed at the same time. Secondly, the use of the ellipse shape to display a linear timeline is problematic in the *News Proximity Map*: it breaks the consistency of spatial proximity. As we can see in the visual mock-ups, the news articles from July 2nd (Monday) is located next to the news articles from July 8th (Sunday), while a natural mapping would suggest they be displayed the furthest from each other. Furthermore, displaying keywords in relation to articles sorted by time does not provide sufficient context - e.g. "Nuclear" related to 2nd July communicates less than "nuclear" related to Asia.

A further exploration would be to experiment with a different organization principle for displaying article titles. Two sketches are included in the appendix (See page 65), exploring the idea of using categories, rather than time, as the information organization principle.

4.5 Conclusion

During the design process, many design challenges were strongly related to information organization and the choice of visual display form. Each visual display form has its own strengths and weaknesses. The strength of a visual display form will be amplified if it is chosen according to the type of data. As I mention in the beginning of this chapter, the design process

serves as a way of understanding the data. It is hard to predict the right visual form for a given set of data, when we do not know the data itself that well.

The purpose of the three visual demonstrations is to provide tools which can be used to gain an preliminary understanding of some of the unique issues involved in the analysis of news contents. To fully examine the strengths and weaknesses of these visual design demonstrations, a further implementation of the demonstrations is needed. Nevertheless, the visual design demonstrations may suggest different approaches for building online news archive applications.

Although the work deals with textual contents of news, some of the design demonstrations may also be applicable to other types of textual content, such as emails, online journals and personal blogs.

5 Reflection

The thesis demonstrates that attention to graphic design in visualization can aid both visual exploration and effectively communicate discoveries and insights regarding a given subject, such as a news archive.

Although I have been designing for the screen display, many traditional graphic design principles can be applied to information visualization. Through the design work, I also gained a better understanding of how graphic designers have to take into consideration the properties of different media. Digital media has limitations as well as opportunities for visualization. The design techniques discussed in the thesis, such as animated transition, highlight, detail-on-demand and semantic zoom, are applied to digital media precisely to compensate for the limitations of screen display resolution and visible screen space. However, these design techniques also introduce different information organization principles and sometimes more effective solutions for visualizing large, complex and ever growing amounts of information.

In this thesis, I generally focus on the discussion of design and visual representation. Little has been discussed of my understanding and interpretation of the BBC news contents, as it has developed through my work. As visualization is not a form of persuasion, but rather exploration and communication. I try to leave an open space for readers to gain their own understandings through my design.

The field of information visualization is a multi-disciplinary field. Although I focus on one aspect of this field, graphic design, the process cannot be separated from the fields of statistics, information architecture and computer algorithms design. Input from experts in these fields would have been beneficial to my work, and might help gain more valuable results from news data.

As discussed in chapter two, the practice of mapping data in visual form has a rich history. What is new is that the availability of information is dramatically increasing, and we may need to design/invent new ways of making sense of and obtaining knowledge from information.

Reference:

Books:

Bertin, Jacques (1983)
Semiology of Graphics, The University of Wisconsin Press, USA.

Bateson, Gregory (2000)
Steps to an ecology of mind, The University of Chicago Press, Chicago.

Card, Stuart K., Mackinlay, Jock D. & Shneiderman, Ben (ed.) (1999)
Information Visualization: Using Vision To Think, Morgan Kaufmann Publisher, San Francisco.

Dodge, Martin and Kitchin, Rob (2001)
Atlas of Cyberspace, Pearson Education Ltd, UK.

Horn, Robert E. (1989)
Mapping Hypertext, The Lexington Institution, Massachusetts.

Mijksenaar, Paul (1997)
Visual Function: An Introduction to Information Design, Princeton Architectural Press.

Norman, Donald A. (1993)
Things That Makes Us Smart, Addison-Wesley Publishing Company, USA.

Norman, Donald A. (2002)
The Design of Everyday Things, 2002 edition, Basic Books, New York.

Spence, Robert (2001)
Information Visualization, Pearson Education Limited, England.

Tufte, Edward Rolf (1997)
Visual Explanations: Image and Quantities, Evidence and Narrative, Graphics Press LLC, USA.

Tufte, Edward Rolf (2001)
The Visual Display of Quantitative Information, Second edition, Graphic Press LLC, USA.

Ware, Colin (2004)
Information Visualization: Perception for Design, Morgan Kaufmann Publishers, San Francisco.

Articles:

Bush, Vannevar (1945)
"As We May Think", *The New Media Reader*, The MIT Press, Massachusetts, pp. 35-47.

Dodge, Martin (2001)
"Cybercartography", *Environment and planning B: Planning and Design*, Volume 28, pp 1-2.
http://www.cybergeography.org/martin/martin.html (Retrieved on June 12, 2007)

Fry, Ben (2004),
"Computational information design", doctor thesis, Massachusetts Institute of Technology.
http://acg.media.mit.edu/people/fry (Retrieved on June 12, 2007)

Hearst, Marti (1999)
"Untangling Text Data Mining", in *Proceedings of ACL: the 37th Annual Meeting of the Association for Computational Linguistics.* University of Maryland, June 20-26, 1999, pp. 1-2.

Hornbaek, Kasper, Bederson, Benjamin B. & Plaisant, Catherine (2002)
"Navigation Patterns and Usability of Zoomable User Interfaces with and without an Overview", *ACM Transactions on Computer-Human Interaction*, Vol. 9, No. 4, ACM Press, New york.

Modjeska, David (1997)
"Navigation in Electronic Worlds: Research Review for Depth Oral Exam", Technical Report. Computer Systems Research Group, University of Toronto.
http://www.dgp.utoronto.ca/people/modjeska/Pubs/lit_rvw.pdf (Retrieved on October 19, 2007)

Paley, W. Bradford (2002)
"TextArc: Showing Word Frequency and Distribution in Text", Poster presented at IEEE Symposium on Information Visualization (InfoVis 02).
http://www.textarc.org/appearances/InfoVis02/InfoVis02_TextArc.pdf (Retrieved on May 10, 2007)

Shanmugasundaram, M., Irani, P., and Gutwin, C. (2007)
"Can smooth view transitions facilitate perceptual constancy in node-link diagrams?", *Proceedings of Graphics interface 2007*, Montreal, Canada, May 28 - 30, 2007. GI '07, vol. 234, pp. 71-78. ACM Press New york. (Retrieved on June 12,2007)

Shneiderman, Ben (1996)
"The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", *Proceedings of the IEEE Symposium on Visual Languages*, Sept 3-6, 1996, pp. 336-343.

Viegas, Fernanda B., Golder, Scott & Donath, Judith (2006)
"Visualizing email content: portraying relationships from conversational histories", in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, ACM, New York, pp. 979–988.

Online publications and resources:

Bach, Michael: Optical Illusions & Visual Phenomena
"Hidden Figures – Dalmatian Dog"
http://www.michaelbach.de/ot/cog_dalmatian/index.html (Retrieved on October 12, 2007)

BBC NEWS
http://news.bbc.co.uk (Retrieved on November 18, 2007)

Gapminder
http://www.gapminder.org (Retrieved on November 18, 2007)

CIA, the world factbook
https://www.cia.gov/library/publications/the-world-factbook/index.html (Retrieved on September 1, 2007)

Flickr: share your photo, watch the world
http://flickr.com/ (Retrieved on November 18, 2007)

The Baby Name Wizard: Name Voyage
http://www.babynamewizard.com/namevoyager (Retrieved on July 12, 2007)

TextArc: an alternative way to view a text
http://www.textarc.org (Retrieved on November 18, 2007)

Victor, Bret
"Magic Ink: Information Software and the Graphical Interface".
http://worrydream.com/MagicInk (Retrieved on June 10, 2007)

Wikipedia contributors
"BBC News", *Wikipedia, The Free Encyclopedia.* ,
http://en.wikipedia.org/w/index.php?title=BBC_News&oldid=171757009 (Retrieved on September 17, 2007)

WorldMapper
http://www.worldmapper.org (Retrieved on October 15, 2007)

Zuckerman, Ethan
"Global Attention Profiles - A working paper: First steps towards a quantitative approach to the study of media attention", The Berkman Center for Internet & Society Research Publication Series.
http://cyber.law.harvard.edu/home/2003-06 (Retrieved on June 12, 2007)

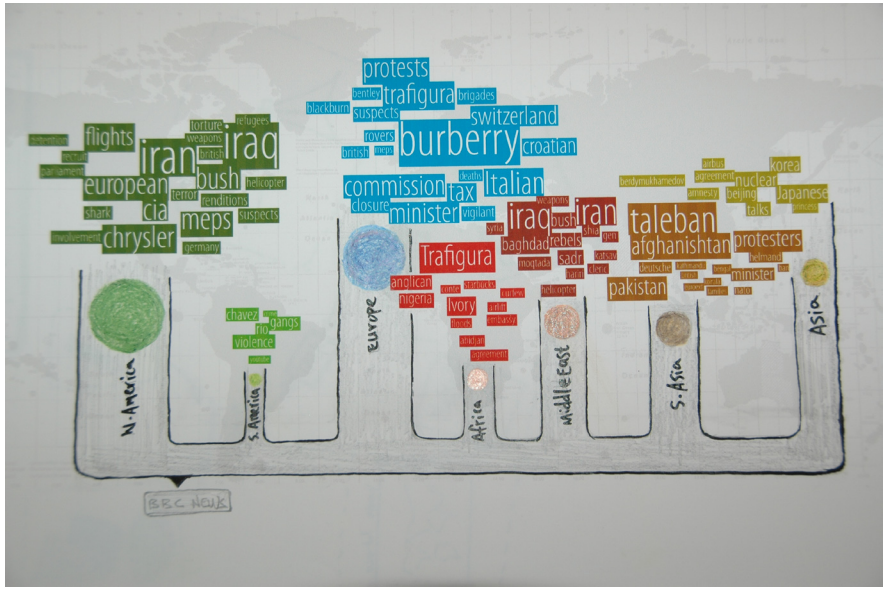
Contents of the attached CD-ROM

Mapping the recent past

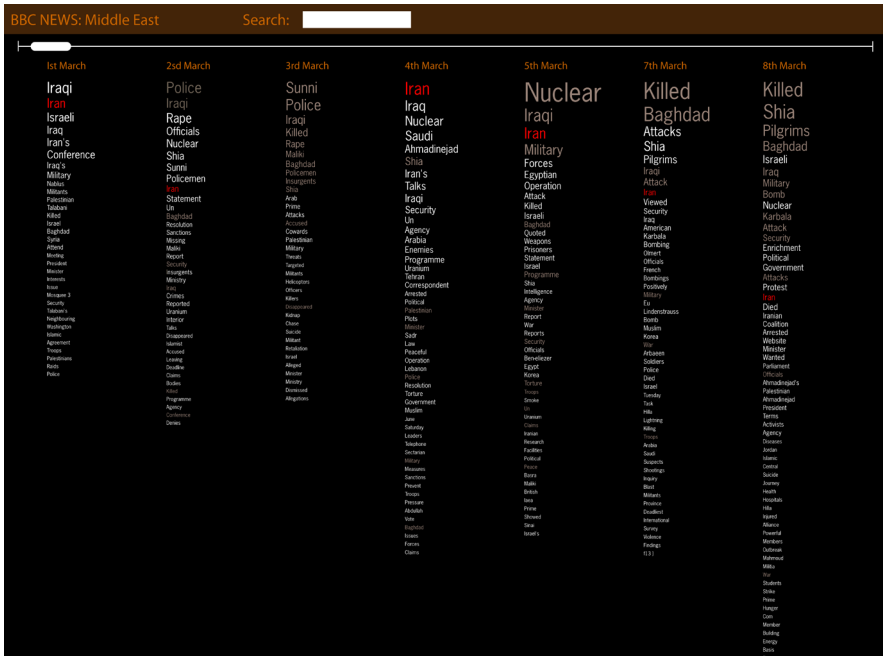
- Chapter 3 prototype
  - exe_file [prototype of *News Coverage Map*, PC version]
    - ncm.exe
    - ncm_front news.exe
  - dcr_file [prototype of *News Coverage Map*, Broswer version, requires Shockwave plugin]
    - ncm.exe
    - ncm_front news.exe
- Chapter 4 image
  - Newsness map [visual mockup, 4 still images]
  - Keywords chart [visual mockup, 1 still images]
  - News proximity map [visual mockup, 5 still images]

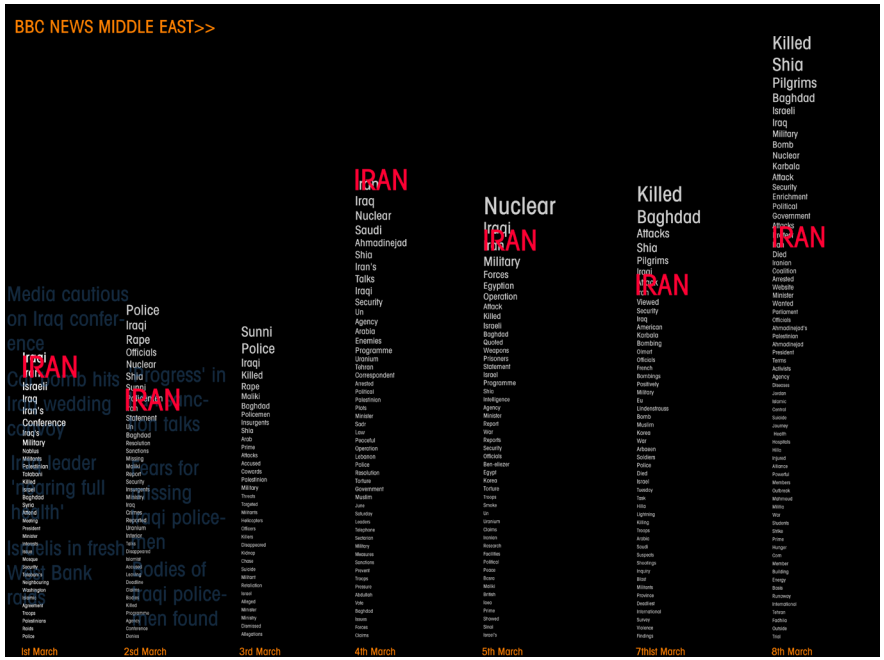
Appendix

Some early sketches of *Newsness Map*



Sketches of *Keyword Chart*





Sketches of *News Proximity Map*