Exploratory Visualization of Temporal Events in Epidemiological Research
Case Study of the Black Death

Peter Farai Madzudzo
March, 2007

Exploratory Visualization of Temporal Events in Epidemiological Research
Case Study of the Black Death

by

Peter Farai Madzudzo

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Geoinformatics

Thesis Assessment Board

Chairman: Dr. R.A. de By
External examiner : Dr Ir. R.M. Hootsmans
Supervisor: Prof. Dr. M.J. Kraak
Second Supervisor: Ms Ir. P.W.M. Augustijn

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Abstract

In epidemiological research incorporating the concept of time is very important as it can result in an improved perspective on the progression of epidemic events in relation to space and time. Health scientists would like to understand spatiotemporal relationships and associated patterns of such events to be able to take necessary measures to solve the problems. The visualization of the distribution and propagation of the epidemics can be very supportive in this process. However, previous attempts to visualize temporal relationships in health studies had been hampered by several limitations. Studying of epidemiological and related geoscience literature revealed that the often erratic and incomplete data collections, the unavailability of suitable data representations, as well as the problem of not being able to clearly disseminate the results had a negative influence on research activities.

This thesis suggests an alternative visualization method for highlighting spatiotemporal relationships in epidemiological research by putting the space-time-cube in the centre of a multiple coordinate view environment. In this approach, the traditional space-time-cube, which includes space (x-y plane) and time (the z-axis), is extended with the third spatial dimension to represent the terrain. The other views include the parallel coordinate plot, the scatter plot and a histogram. Data from the Black Death case study recently made available was used for showing the usability of the suggested approach for health studies. This Black Death data has interesting characteristics since at that time, the 14^{th} century, there were no proper recording of events resulting in data which has some characteristics of uncertainty and incompleteness. The approach taken in this thesis tried to incorporate these deficiencies into temporal visualizations which are based on statistical and other computational methods. To allow more insight into the data, the alternative visualization approach provides for user interaction with the data and the views, offering brushing functionalities in the multi-linked data frames. The studied approach has been compared with existing traditional visual representations. The project also covered the possibility of incorporating real time data entry to allow health scientists to monitor today's epidemics as well.

Keywords: Visualization, Epidemiological Research, Space Time Cube, Multiple Coordinate View.

Acknowledgements

The successful completion of this thesis project has come from a very long way, and along the way many hands, heads and hearts made generous contributions.

First of all, I would like to thank my first supervisor, Prof. Dr. M.J. Kraak for his continued support and patience all along the way. For the months I have worked on this thesis project, he was always around to assist. He also initiated and supported my attendance and presentation of scientific papers at international GIS conferences. Thank you very much.

I would like to thank my second supervisor Ms Ir. P.W.M Augustijn. She was very supportive of the project and facilitated some meetings with public health specialists with whom I discussed my work. She made valuable contributions and suggestions to most of the problems that I encountered. It was a pleasure working under her guidance.

I would also like to thank Ir. Bas Ratsios for his technical assistance and patience during the course of the thesis. Without him this thesis project would have been very difficult to complete. His contribution is greatly appreciated.

I would like to thank the ITC staff for the support, in various forms, which made the stay here more comfortable and turning the long 18 months look like a very short period. I would also like to thank NUFFIC for the funding, without which this programme could have proved impossible for me.

I would like to thank my classmates for all the support that they generously offered. The thesis period was a difficult period but they always had time for others. My special thanks go to Yashar Fallah Vazirabad, Sanka Perera and Mohamed Thamby Mohamed Rafeek who were always around and urging me to keep going.

I would like to thank all the members of my family, Vimbayi, Chipo and Farai for the support and understanding during the long period of my absence. Special thanks go to my brother, Elias, for periodically popping in to check on my progress.

dedicated to all those who never got the chance to go to school

Table of contents

1. Introduction 1
1.1. Motivation and Problem Statement 1
1.2. Research Identification 2
1.3. Research Objectives 2
1.4. Research Questions 2
1.5. Innovation Aimed At 2
1.6. Research Methodology 3
1.7. Thesis Structure 4
1.7.1. Chapter 1. Introduction. 4
1.7.2. Chapter 2. Multivariate Spatiotemporal Visualization 4
1.7.3. Chapter 3. Epidemiological Research 4
1.7.4. Chapter 4. Black Death Case Study 4
1.7.5. Chapter 5. Visualization Approach to Support Epidemiologists 4
1.7.6. Chapter 6. Prototype Development 4
1.7.7. Chapter 7. Prototype Evaluation. 5
1.7.8. Chapter 8. Conclusions and Recommendations 5
2. Spatiotemporal Visualization 7
2.1. Introduction 7
2.2. Spatiotemporal Data 7
2.3. Spatiotemporal Representation 8
2.4. Temporal Representations 8
2.4.1. Narrative Static Single Graphics 8
2.4.2. Small Multiples 9
2.4.3. Temporal Animation 10
2.4.4. Brushing 10
2.5. Alternative Temperal Visual Representations 11
2.5.1. Space Time Cube 11
2.6. Parallel Coordinate Plots 15
2.7. The Graph 16
2.8. Scatter Plot 17
3. Epidemiological Research 19
3.1. Introduction 19
3.2. Areas of Interest in Disease Mapping 19
3.3. Visualization in Epidemiological Research 20
3.3.1. Cholera Deaths in London 20
3.3.2. Spatio-Temporal GIS Analysis for Environmental Health 21
3.3.3. Prostate Cancer Mortality in the United States 22
3.3.4. Spatio-Temporal Analysis Toolkit (ESTAT) 23
3.3.5. Spatiotemporal Modelling and Mapping Plague Epidemic in India 24
4. Case Study Data – Black Death 27
4.1. Introduction 27
4.2. The Black Death 27

4.3. Sources of Black Death Data ... 28
4.4. Characteristics in Black Death Data ... 29
4.5. Topographic and Social effects on Black Death Propagation and Severity ... 31
5. Visualization Approach to Support Epidemiologists ... 33
5.1. Introduction ... 33
5.2. Answering Questions through Visualization ... 33
5.3. Space-Time Cube ... 34
5.4. The Parallel Coordinate Plot ... 34
5.5. Histogram ... 35
5.6. Scatter Plot ... 36
5.7. Multiple Views, Linking and Brushing ... 36
5.8. Summary ... 37
6. Prototype Development ... 39
6.1. Introduction ... 39
6.2. Case Study Data ... 39
6.2.1. Database Design ... 39
6.3 Prototype Development ... 40
6.3.1. The Space Time Cube ... 40
6.3.2. The Parallel Coordinate Plot ... 41
6.3.3. The Graph ... 42
6.3.4. Scatter Plot ... 42
6.3.5. Linking and Brushing ... 42
6.3.6. Digital Terrain Model ... 43
6.4. The Prototype Interface ... 43
7. Prototype Evaluation ... 44
7.1. Introduction ... 44
7.2. Evaluation Methods ... 44
7.3. Evaluation Method Selected for this Project ... 46
7.4. Evaluation Procedure ... 46
7.5. Evaluation Team ... 46
7.6. Results from the Evaluation Process ... 47
7.7. Summary ... 48
8. Conclusion and Recommendations ... 50
8.1. Research Questions ... 50
8.1.1. Previous Work on Visualization of Temporal Events ... 50
8.1.2. Alternative Visual Representation ... 51
8.1.3. How Effective are these Representations ... 51
8.1.4. How can the solution deal with real-time monitoring ... 52
8.2. Recommendations ... 52
9. Bibliography ... 54
10. Appendix ... 58

List of figures

Figure 1: Methodology for the thesis project....3
Figure 2 The basic components of the Triad framework 8
Figure 3 Minard 1812. Napoleon's March to Moscow. 9
Figure 4 Multivariate multiple time slices showing temporal pollutant levels in Los Angels .10
Figure 5 Aquarium 11
Figure 6 Space Time Path 12
Figure 7 Space Time Prism/Potential Path Area 12
Figure 8Co-location 13
Figure 9Co-location 13
Figure 10 Napoleon's march on Moscow 14
Figure 11 Parallel Coordinate Plot....15
Figure 12 Simple Bar Graph 16
Figure 13 Line Graph 17
Figure 14 Scatter plot correlations 18
Figure 15 Environmental Factors Affecting Patterns Appearance of Diseases 19
Figure 16 Dr John Snow. Cholera Deaths in central London, September 1854. 21
Figure 17 Geospatial lifelines of three individuals over nine decades 22
Figure 18 Screen shot of STIS software, the three linked views. 23
Figure 19 ESTAT Geovisualization Toolkit 24
Figure 20 Space-time mortality rate maps during 1902–1903....25
Figure 21 Black Death summary for Switzerland 30
Figure 22 Trajectory of an epidemic....34
Figure 23 Parallel Coordinate Plot of the Black Death Dataset....35
Figure 24 Black Death Histogram....35
Figure 25 Scatter Plot of the Black Death....36
Figure 26 Screen shot of the STC showing a space-time path. 41
Figure 27 Parallel Coordinate Plot (PCP) showing the Black Death variables. 41
Figure 28 Linking and brushing....42
Figure 29 The prototype showing the site map and the space-time cube. 43

1. Introduction

1.1. Motivation and Problem Statement

The world we live in today is faced with so many challenges, including health related problems. Advances in technology and medicine have done a great deal to attempt to bring health to all people, however, newer diseases spring up now and then. Attending to these epidemiological problems is often slowed down by, among other things, our lack of understanding of the characteristics of the epidemics and at times by the incorrect data that is received concerning these problems. There are so many varied reasons why the data received is at times very difficult to deal with. Technology, politics, economics, social, cultural, and other variables in our societies have a lot to play in the shaping of our modern world and can have an impact on information dissemination. Pressures on some of these variables tend to have a ripple effect on the epidemiology data quality. In some areas correct data cannot be collected due to lack of proper technology, the wrong data is collected or data is purposely altered so as to reduce the impacts of such revelations on the other variables. The end result is course data which is characterized by uncertainty. Statistical methods have to be used to help solve this problem.

An epidemic that had characteristics similar to the statement above is the Black Death. The Black Death was an epidemic that went through Europe, in the 14th Century and whose origin is not known. Record keeping, in the 14th Century was not systematic and it varied from one region to the other. Some regions had very little information recorded as compared to other regions which had well documented records. It is also interesting to note that the census, as we know it today, did not exist until the 15th century, so as a result the population figures are mare estimates derived from knowledge basis from various sources. Various statistical methods have been employed to come up with the data for the areas affected by Black Death (Christakos et al, 2005).

With all the new technology that we have now, including improved data collection methods, data processing methods and other information technology we enjoy, so why use data which is about 640 years old? We have had Aids, Bird Flue, Mad-cow, SARS, and other problematic epidemiological cases, so what is interesting about such old data which is incomplete in some aspects and also lacks some concrete demographic data that is essential for explorative analysis? Most of the data sources used to calculate the deaths and the population figures are subjective, for example, use of parish records is based on the assumption that most people were religious and had records kept by the clergymen or that the clergymen recorded all mortalities irrespective of their spiritual orientation. All these questions and comments about the Black Death data of 1347-51 AD is what makes it more interesting to demonstrate how exploration visualization can be applied to represent such temporal events in epidemiological research.

1.2. Research Identification

The main research objective is to design an exploratory visualization tool that can be used by health experts to visualize spatiotemporal epidemiological data. A prototype is going to be developed which will be tested using the case study data – Black Death.

There is need to design an exploratory visualization tool that can handle multivariate data from which analysis can be made. Exploratory visualization is unique and is different from the mere presentation graphics, in that in the later, ideas are presented to the user as opposed to the former which aims at presenting the raw data as true as possible and then assist the user in extracting interesting characteristics of the data.

1.3. Research Objectives

The following are the research objectives that this study will cover:

- To identify tools and variables that are best suited for exploratory visualization which can assist in attending to the problem of uncertainty and vagueness in spatiotemporal data.
- To test the visualization tool with the case study data and evaluate the usefulness of the tool in epidemiological research. To study the characteristics of the epidemiological data, Black Death of 1347-51, so as to see how the problem of coarseness, incompleteness and uncertainty can be solved by making use of statistics and visualization.
- To assess the results of the visualization and see how its effectiveness can be used to solve some epidemiological problems using real time mapping.

1.4. Research Questions

The following research questions are going to be addressed by this study.

- What work has already been done in this area of space-time visualization of epidemics and how much has been achieved?
- Which alternative visual representations are suitable to deal with ‘epidemic’ space-time data and can also handle their sometimes uncertain and vague character?
- How effective are these representations?
- How can the solution be extrapolated to deal with real-time monitoring of spreading diseases?

1.5. Innovation Aimed At

This project aims at designing an effective visualization tool to support epidemiological research. The tool will enhance the visualization of spatiotemporal data by allowing the temporal aspect to be treated as a dimension rather than as an attribute. The project will also look at the possibility of retaining the three spatial dimensions (x,y,z) whilst displaying the temporal aspect as one of the dimensions.

1.6. Research Methodology

The first stage in this research project is to look at visualization and then narrow down to exploratory visualization. The next stage will look at how exploratory visualization can be used to assist in epidemiological research. The project will look at the main problems that are faced by epidemiologists in addressing health problems and how visualization can assist in solving some of these problems. A prototype is going to be designed, using Open Source software, to illustrate how visualization can assist in public health research. To test the usefulness of the prototype, data from the Black Death case study is going to be used. The possibility of using this method with mobile technology is going to be looked at towards the end of the project. The diagram below shows how the project is going to progress.

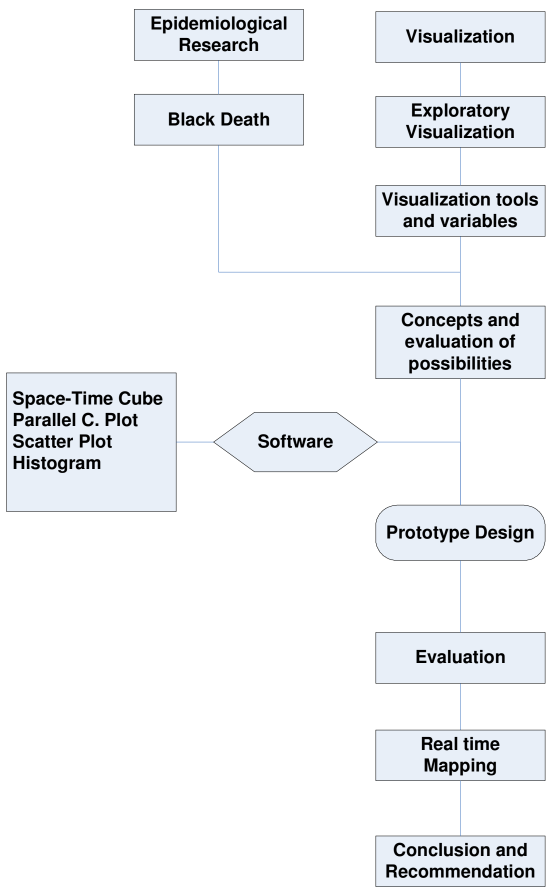


Figure 1: Methodology for the thesis project.

1.7. Thesis Structure

1.7.1. Chapter 1. Introduction.

The first chapter is going to introduce the thesis. Then it will look at exploratory visualization and its contribution to understanding the importance of visualization in the field of Epidemiological research. This chapter will cover at the brief overview of the processes that are going to be involved in this research project.

1.7.2. Chapter 2. Multivariate Spatiotemporal Visualization

This chapter is going to cover literature review in visualization. It is going to highlight the concept of space-time visualization and explain why it is so important in GIS. Various visualization tools and techniques are going to be looked at highlighting their strengths and weaknesses in representing space and time in GIS.

1.7.3. Chapter 3. Epidemiological Research

In this chapter, some previous projects that have been done dealing with the topic of space-time visualization in public health and epidemiological research are reviewed. It highlights how visualization, and in particular spatiotemporal visualization, has been applied to solve health related problems.

1.7.4. Chapter 4. Black Death Case Study

The chapter summarise the details Black Death epidemic that swept through Europe during the 14^{th} Century. The nature of the data, especially its problems, and the sources of the data are going to be identified. The source of the information in this chapter is the case study report by Christakos and others.

1.7.5. Chapter 5. Visualization Approach to Support Epidemiologists

This chapter is going to look at the conceptual design concept of visualization approach to support epidemiologists.

1.7.6. Chapter 6. Prototype Development

This chapter will combine the knowledge gained to highlight ideas and possible solutions that can be used to deal with the problems faced with space-time visualization in epidemiological research. Suggestions of how to deal with space-time visualization will be noted and evaluated to select the ones which will best address such issues in health studies.

1.7.7. Chapter 7. Prototype Evaluation.

The chapter will cover how the prototype will be tested for its usability in exploratory space-time visualization. It also covers the various methods of evaluation. Comments and results from the testing will be used to improve on the prototype design and give suggestions for future work.

1.7.8. Chapter 8. Conclusions and Recommendations

This chapter will wind up the research topic by summing up the findings, commenting on the usefulness of the research in helping improve research in epidemiological studies. It will also look at the possibility of applying real time monitoring to the prototype. Recommendations of future work will also be made.

2. Spatiotemporal Visualization

2.1. Introduction

The primary analysis tool is the human imaginative mind, which has scientifically been agreed that it operates mainly with images (Andrienko & Andrienko 2006). In general terms, visualization is a process of making something visible or making it perceptible to the mind or imagination.

According to the Random House Unabridged Dictionary, to visualize is; to recall or form mental images or pictures, to make visual or visible, to form a mental image of or to make perceptible to the mind or imagination.

The term perception should not be confused with what is termed seeing. Seeing is simply the capturing of an object by the eye whereas perception means that the brain responds to the object and processes it to come up with a specific meaning. The detection of the symbol and the understanding of its meaning is what is referred to as visual perception. The visual concepts are obtained either directly from the environment or are formulated in the brain. Mental image is the term that is used to refer to the concept that is extracted from the human memory. Together with domain expertise, vision is assumed to be able to turn large, complex, heterogeneous data volumes into information and to integrate it with other information (Blok, 2004; after Card et al, 1999).

Visualization provides three functions, recording, communicating and processing information. Each function works well depending on the type of visualization that is at hand. The recording and communicating function is most suitable in presentation graphics, where an idea or a theme is being presented for visualization. The communicating and processing function is ideal for graphics where some analysis need to be performed on the data. The function of processing information is required where the data is being presented in an exploratory visualization environment. This thesis is going to concentrate on the exploratory visualization environment.

2.2. Spatiotemporal Data

The aspect of spatiotemporal data that is interesting to the user is the understanding of an object’s position in terms of geographic space and time. In visualizing spatiotemporal data the user is interested in the properties of each of the elements; location or space, time and objects, and how they are interrelated. According to Peuquet (1994), in her Triad framework she stated that the user of spatiotemporal data is interested in addressing three questions ‘what’, ‘where’ and ‘when’ as illustrated in the diagram below and which are composed of the following formations.

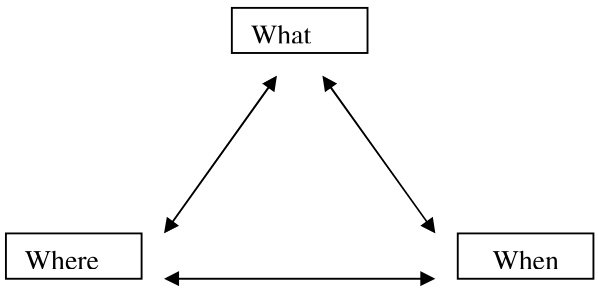


Figure 2 The basic components of the Triad framework (Peuquet, 1994)

a) when + where → what
This describe the objects or set of objects (what) that are present at a given location or set of locations (where) at a given time or set of times (when)

b) when + what → were
This describe the location or set of locations (where) occupied by a given object or set of objects (what) at a given time or set of times (when).

c) where + what → when
This describe the times or set of times (when) that a given object or set of objects (what) occupied a given location or set of locations (where). (Peuquet, 1994)

2.3. Spatiotemporal Representation

There have been a number of graphic representations of space time relationships. Each tool or technique of visualizing space time has its advantages and disadvantages and can only be effectively applied in some situations and not in all. The following examples are some of the most common methods that have been used to represent space time relations.

Space time visualization is a graphical presentation in which time is included as a dimension rather than an attribute. It departs from the traditional option of visualizing three dimensional data where three dimensions (x,y,z) represent the spatial geographical location attributes and the height value by representing time in the height dimension. Digital elevation models constructed using this visualization method show differences in time rather than differences in height. There are various methods that are used to represent the temporal characteristics of data. The following sections take a look at some of the methods and how they hand the time aspect.

2.4. Temporal Representations

2.4.1. Narrative Static Single Graphics

In a narrative static graphic, all the information is included in one frame. Normally this is a two dimensional graphic. The temporal aspect and other data will be represented in other forms, usually text and symbols, and not as dimensions. The graphic by Charles Joseph Minard depicting Napoleon's march, in 1812, to Moscow is accepted as one of best space time static graphics ever presented (Tufte, 1982). In his graphic, Minard demonstrated how multivariate complexity can be simplified by graphic presentation. The graphic shows the size of the army, the direction of movement, loses incurred along the march, temperature during the retreat, spatial location of the army at each point, etc. Figure 3 below shows the original graphic by Minard.

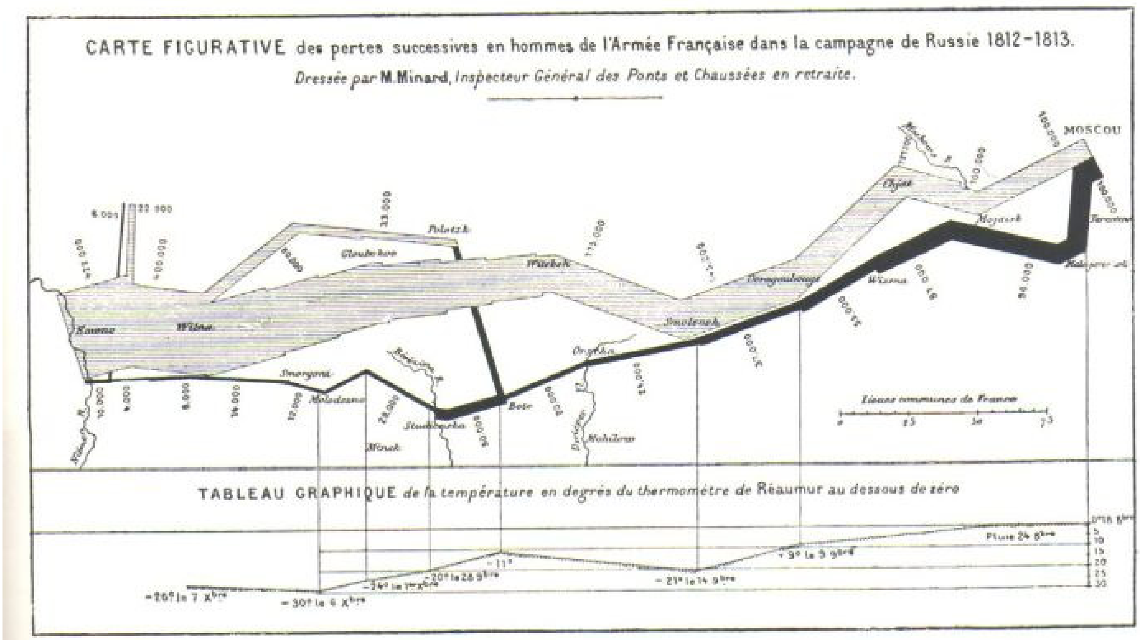


Figure 3 Minard 1812. Napoleon's March to Moscow. (Source: Tufte, 1983)

2.4.2. Small Multiples

In a small multiples graph, the temporal information is recorded in the form of time slices which are events at the same place recorded over different temporal intervals. Spatial dynamics have to the mentally extracted by making comparisons between the time slices. The intervals can be uniform or can be event determined. The uniform intervals are normally used to depict growth in a static spatial phenomena, for example; expansion of a city of growth in population over different years. The event based recording is for occurrence of such phenomenon like earthquakes, floors, droughts, which might be missed if the interval method is used. The occurrence date might fail to coincide with the interval date, hence the need to record it as a separate event.

Figure 4 below shows daily emissions of nitrogen oxides, carbon monoxide and reactive hydrocarbons from both industries and traffic recorded in Los Angelis in 1979. The daily recordings were grouped into 4 time slices each spanning over three hours. Vertically, the slices show the different pollutants and horizontally they show the individual pollutant levels at each spatial location and at different temporal space.

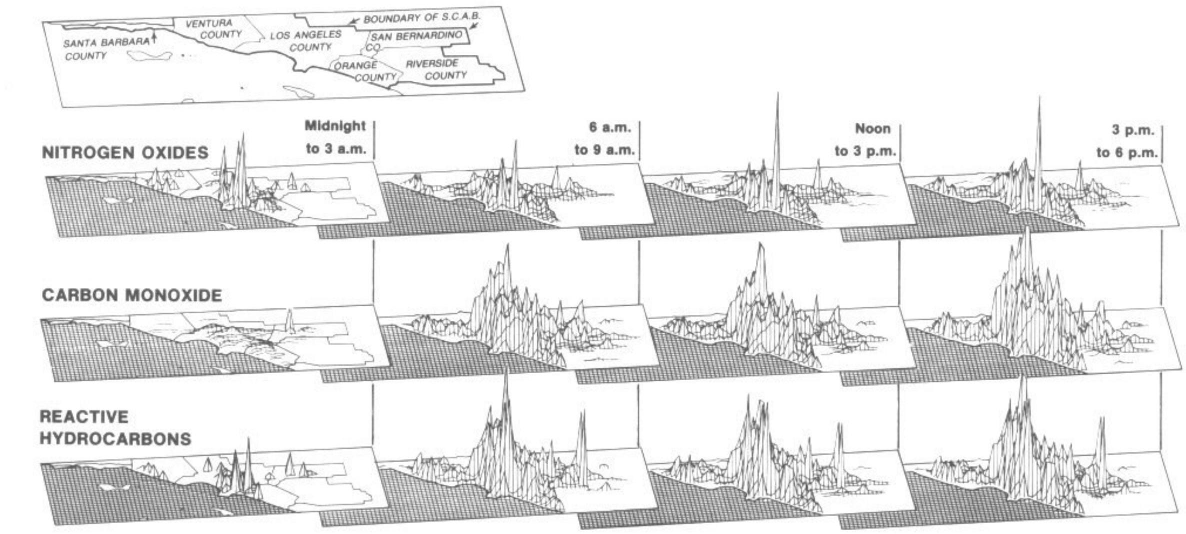


Figure 4 Multivariate multiple time slices showing temporal pollutant levels in Los Angels (Source: Tufte, 1983)

Multiple time slices have the advantage of displaying all the data at once and therefore providing a platform for easier comparisons between the data. Comparisons or patterns forming between distant frames can easily be visualized without further manipulation. Problems with this form of data visualization will arrive when this method is applied to data recorded over vast number of time slices, for example multi-dimensional data recorded weekly over one year resulting in 52 time slices. The human mind is limited to the number of images it can comprehend at a given time. The number of images may vary from 6 to 8 depending on the level of complexity in the composition of each image and also varies from one person to another. The other problem is the display space. Trying to fit the images on the screen will result in them becoming too small for meaningful visual comparisons and also enlarging them would result in some of the images going out of sight and therefore losing the advantages of having the images displayed at the same instance.

2.4.3. Temporal Animation

Temporal animation is a technique in which multiple time slices are played to produce a movie-like effect. The animation can be interactive, allowing the user to have control on the progression of the animation. Such controls like speed control, fast forward, reverse, selecting time slices and viewing time slices individually. The technique requires a media player. It has the added advantage of creating a visual perception which is identical to the real situation. The main disadvantage of the technique is that it can consume a lot of computer memory.

2.4.4. Brushing

Brushing is a technique which can be applied to linked view. The effect is that if one object is selected in one view, all the other corresponding objects in the other views are selected as well. This is ideal for visual exploration.

2.5. Alternative Temperal Visual Representations

2.5.1. Space Time Cube

Around the early seventies, Hägerstrand(1970) devised an alternative way of incorporating time in visualization. He designed it so that the geographic spatial dimensions (x,y) formed the base of the graphic and the temporal dimension was represented in the height (z) direction. His emphasis was on showing life histories of people and how people interact in space and time. His model is often seen as the start of the time-geography studies (Kraak, 2003).

The space-time cube evolved from Hägerstrand's aquarium. It dates back to the 1970's when Hägerstrand developed the concept of time geography. Time geography studies humans in space-time with an associated set of visual tools for looking at geographic reality at the individual level (Moore et al, 2003). It deals with complex space-time phenomena by reducing space-time patterns to the individual level, observing paths of individuals through space and time and their interactions. The main components of the time geography are lifelines, stations, bundles, domains and prisms, as illustrated in the figure 5.

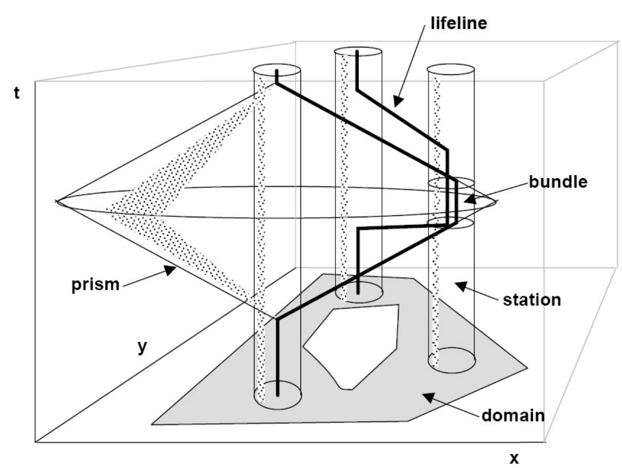


Figure 5 Aquarium (Source: Moore et al, 2003)

Lifelines are the paths that result as an object moves in space and in time. A vertical line indicates that there was movement in the temporal direction and non in the spatial direction, meaning the object was stationary. Speed of the object is determined by the angle of the trajectory from one station to the other. A steep gradient is an indication that the object took much longer to move from one station to the other and a gentle gradient is an indication that the object took less time. Lifeline, also referred to as the space-time path, answers the 'what' question by showing the identity of the object, the 'where' by referring to the domain, and the 'when' by showing the temporal position along the 't' axis.

A bundle refers to the point where individual objects from different stations coincide both in time and in space. Movement in space-time can be constrained by the boundaries of some domain, a physical manifestation of authority constraints. A prism represents the total area of space reachable by an individual in the time available (ibid).

The space-time cube (STC) has retained and refined some of the basic building blocks of the time geography aquarium. The concept of the STC was introduced during the time when graphics were limited to manual methods which made it very difficult to apply. Technological advancement currently prevailing provide much better functions for visualizing temporal data using this method.

The STC presents data in a three dimensional orthogonal viewpoint. The horizontal axes is used to record the geographical position of the object. The vertical axis is used to provide an ordered and synchronized sequential order of events, as well as relative position of specific events that occurred throughout the lifespan of each individual (Abrahart and Bradshaw, 2004). In the STC, the space-time path shows the object's trajectory through space and time. This is similar to the lifeline in the original space-time cube of the seventies. The space-time prism denotes the possible locations that an object could travel within under certain set of constraints. When the space-time prism is projected on a two dimensional surface it produces the potential path area.

The six illustrations below show the space-time path, the space-time prism and the potential path area, and the spatiotemporal relationships that can take place in a space-time cube.

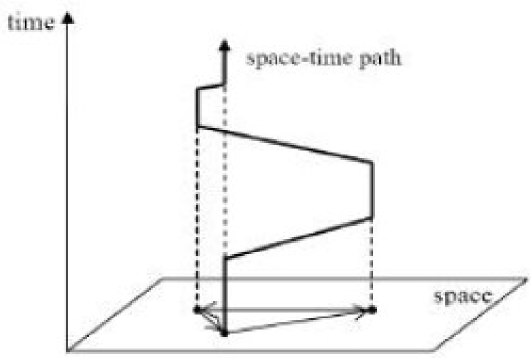


Figure 6 Space Time Path   (Source: Abrahart and Bradshaw, 2005 after Yu, 2004)

The space-time path represents the trajectory of the object in both space and time, as illustrated in figure 6. The dotted lines show the object's position in relation to space and time. With this function the object can be traced in both time and space.

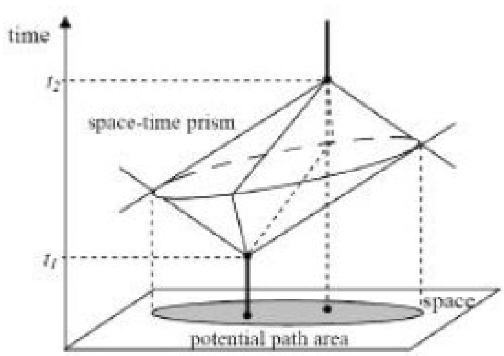
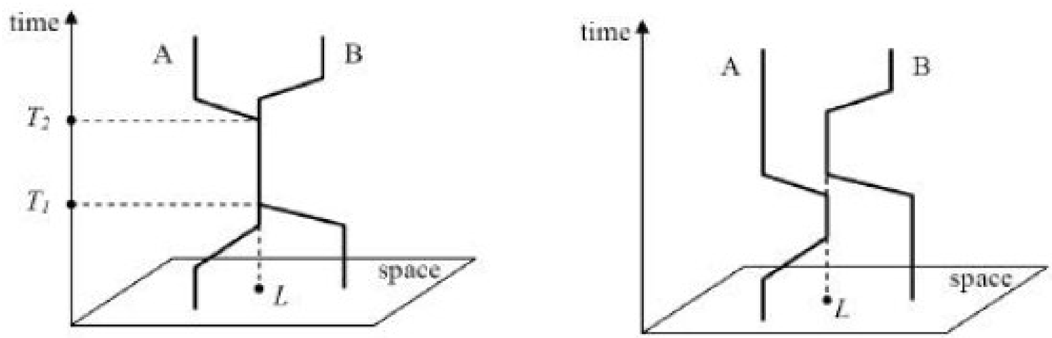


Figure 7 Space Time Prism/Potential Path Area (Source: Abrahart and Bradshaw, 2005 after Yu, 2004)

The space-time prism shows the possible positions that the object can be at in relation to time. The extent of the prism is determined by the constraints that the object will be subjected to. The potential path is the spatial coverage of the possible positions that the object can possibly be at between t1 and t2.

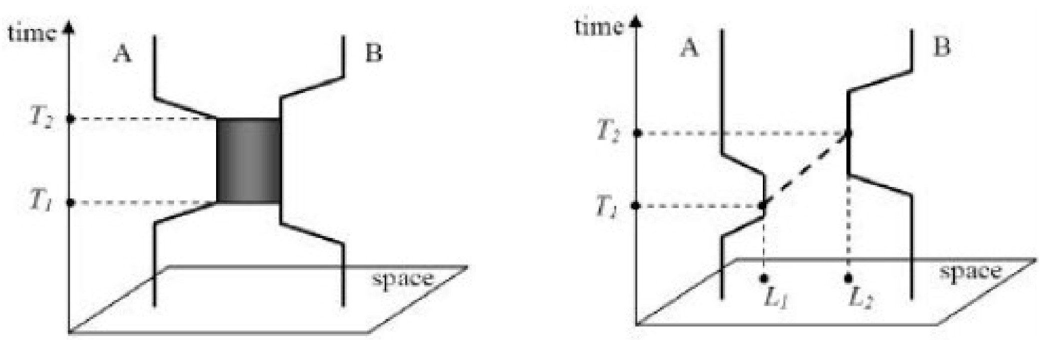


Co-location in space and time

Co-location in space

Figure 8 Co-location (Source: Abrahart and Bradshaw, 2005 after Yu, 2004)

The two diagrams in figure 8 show how objects can co-locate in space and time, and in space. When two objects are at the same place at the same time, they are said to be co-located in space and in time. It can happen that the two objects can be located at the same place but in different times, which is co-locating in space. In the left figure below, two objects can be at two different stations at the same time. They are referred to as having co-located in time. It also can happen that there is no co-location neither in time no in space as illustrated by the right figure. (Abrahart and Bradshaw, 2005)



Co-location in time

No co-location in space or time

Figure 9 Co-location (Source: Abrahart and Bradshaw, 2005 after Yu, 2004)

In a space-time cube the symbols are placed vertically in relation to the time of their occurrence. To distinguish between the attributes of the data, variation of size and colour can be applied to the graphic. Events that occur at the same location and at different times, will be shown as a chain of vertically aligned symbols. The orientation of the symbols can reveal a particular relationship within the data. The spatial plane at the bottom of the cube can be moved up or down along the time axis. This allows comparison of data at different temporal scales. Rotating the cube will allow visualization of the data from different angles. The figure 10 below shows Napoleon's march on Moscow. This is an example of how the space-time has been used to show the trajectory of an event in space and in time.

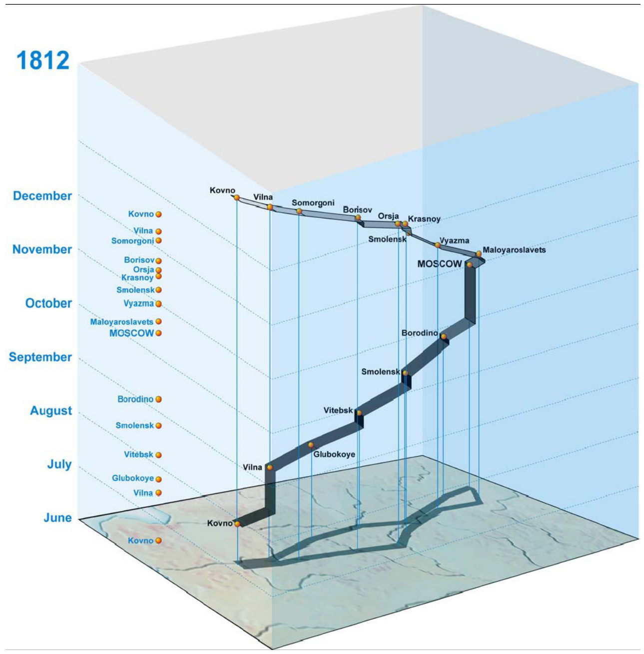


Figure 10 Napoleon's march on Moscow (Source: Kraak and Ormeling, 2003)

Referring back to our questions, the STC is very efficient in responding to the 'where' and 'when' questions. The 'what' question is not so obvious in the STC, depending on the question and the type of phenomena being analysed. If prior knowledge of the phenomena exists the questions will be made easier but if no prior knowledge exists then there is need to have more information to identify the object. To improve the response of the STC to the 'what' question, there is need to link the STC to other data sources that can provide more information about the identity 'what' of the object. Such representations like the parallel coordinate plot, the histogram and scatter plot. This can be done by dynamic linking the STC to these other representations. The STC has an additional functionality to move the space plane along the temporal dimension. This improves visualizing the relationship between object and place at each temporal scale.

The space-time cube offers an effective visualization environment for spatiotemporal data. It improves the way of visualizing data both in space and in time. The cube can display both continuous and discrete temporal data. Different temporal categories can be visualised using this method. The cube eliminates the problem of over-plotting, that is, when events take place at the same place and at different times. Clusters in data, both temporal and spatial, can be effectively displayed. The space-time path improves the understanding of how an object moves in space and in time.

The disadvantage with this representation method is that distances cannot be easily and correctly perceived. Some solutions to the limitations is to incorporate interactive manipulations, that is, changing the viewing perspective, temporal focusing, and dynamic linking with a map display which will allow simultaneous highlighting of corresponding symbols. The other limitation is when a huge amount of dataset is to be visualized. If many objects are displayed in the cube, it becomes difficult to visualize some of these data. This problem can be minimised by use of temporal zooming.

2.6. Parallel Coordinate Plots

The parallel coordinate plot (PCP) was originally proposed and implemented by Inselberg (1985). Its main advantage over other types of graphic representation tools is its ability to display multi-dimensional data in one representation. The principle is simple: it is that of taking all the axes of the multidimensional space and arranging them in order, but parallel to each other (Spence, 2001). Each variable is plotted along its group axis in proportional to its value. When all variables are plotted and viewed alongside each other, the lines joining the values in different axis may show patterns revealing some form of correlation between the variables. Figure 11 below illustrates the principle of the PCP. The figure shows the relationship between three variables, VI, Longitude and Latitude. The VI values are plotted along the VI axis according to their value placement along the axis continuum of 2.90 to 27.10. Each VI value has a corresponding Longitude and Latitude value. The values of each variable are then joined to form one continuous line. The relationship between the variables is not always a one to one relationship. There are times when there is either a one to many relationship or a many to one relationship.

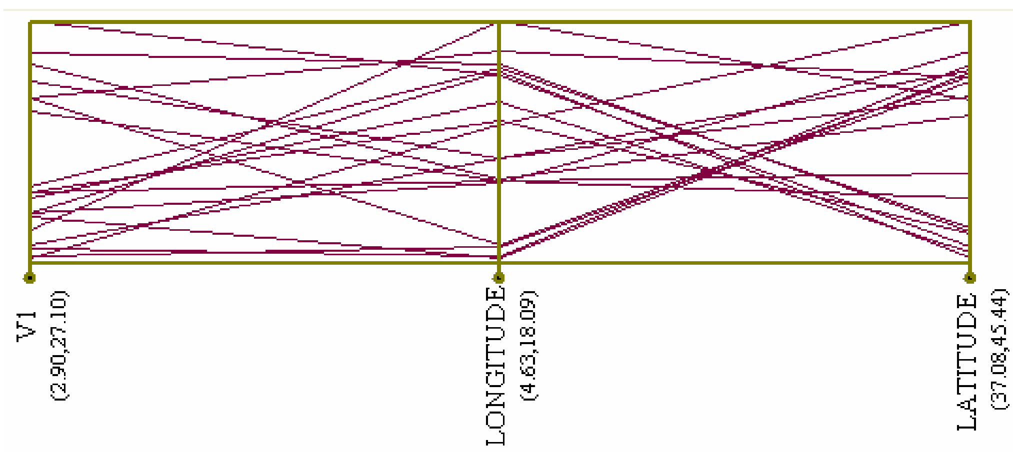


Figure 11 Parallel Coordinate Plot

The PCP has been transformed into various application specific parallel coordinate plots. There are various ways to enhance visualizing PCP data. Different colours can be used to indicate different groups of data. The main advantage of the PCP is that it allows visual pattern recognition of multidimensional data. Any number of dimensions may be used but as more axis and data are introduced, the patterns formed may be difficult to visually detect. All variables are treated equally. Some modification to address this problem was to introduce interactive selective viewing of the axis. Some axis can be switched off when not needed for the analysis. In some environments, the PCP offers the option to rearrange the order of the parallel axis. The order of the axis can have an influence on the pattern that can be revealed by the PCP. Understanding how the combination of variables can influence the resulting pattern is very important as failure to do so might result in wrong conclusions being drawn.

2.7. The Graph

Graphs come in various forms. Some of the types of graph that are normally used in multidimensional data are; the bar graph, the line graph, the pie, the area graph and the XY graph. The principle behind the different graphs is basically the same so here only the bar graph and the line graph are going to be discussed.

A bar graph is used to show relationships between groups. The two items being compared do not need to affect each other. The bar graphs can be applied in various forms; single, multiple or cumulative graphs that can be used to represent quantitative data of one or more variables.

The bar graph is simple to construct. These graphs can be applied to other graphics, like maps, or appear separately. As separate graphics they have the disadvantage of not being associative and not linked to any spatial data. When they are shown as graphics in another graphic they have the disadvantage of obstructing other data within the graphic. Figure 12 below is an example of a single bar graph showing the relationship between bean growth and type fertilizer used. The advantage of using the bar graph is its capability to show distribution or frequency over time. The problem with it is that it does not show the relationships with other variables.

:

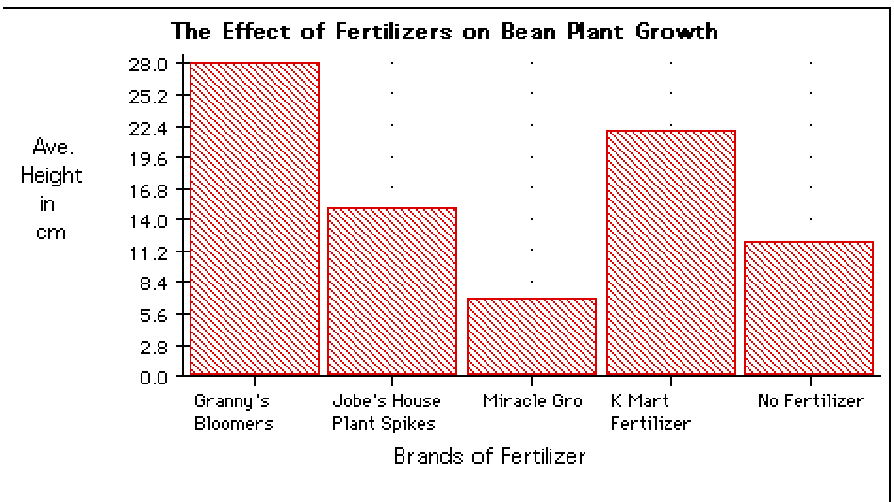


Figure 12 Simple Bar Graph (Marszalek, 2004)

In a line graph, instead of having bars, a line joins the maximum data points. The advantages of a line graph over the bar graph is that more variables can be plotted within the graph without creating too much noise. The line graph resembles the PCP but the main difference is that in a line graph you can only plot variables of the same magnitude or belonging to the same themes. Figure 13 below shows the pulse and the time recordings of two people exercising.

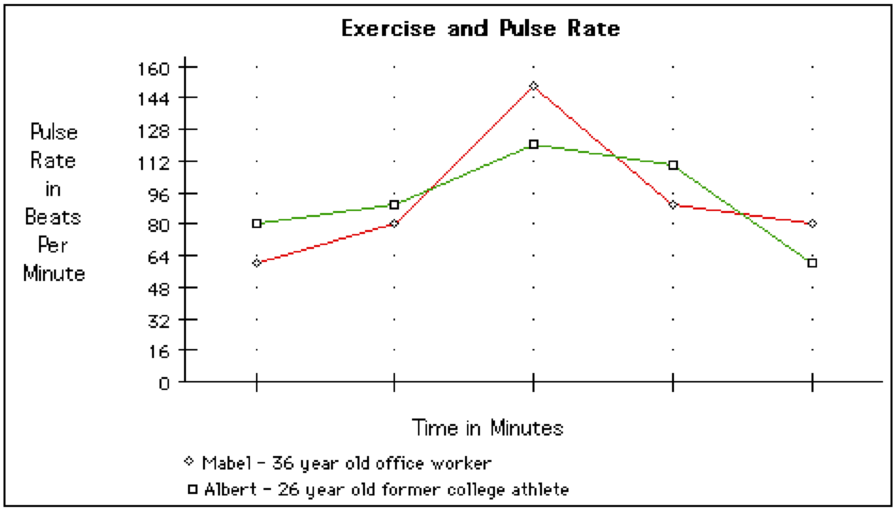


Figure 13 Line Graph (Marszalek, 2004)

2.8. Scatter Plot

Scatter plots are similar to line graphs in that they use horizontal and vertical axes to plot data points. However, they have a very specific purpose. Scatter plots show how much one variable is affected by another. The relationship between two variables is called their correlation.

Scatter plots usually can hand very large data sets. The closer the data points come when plotted to making a straight line, the higher the correlation between the two variables, or the stronger the relationship. An example of correlation is between studying and grades obtained. If the students are spending more time studying and they are receiving high grades, then the relationship between studying and passing is said to be highly correlated. In this case it will be referred to as a positive correlation. An example of negative correlation is when a group of smokers quit smoking and their weight increase.

If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a positive correlation. If the line goes from a high-value on the y-axis down to a high-value on the x-axis, the variables have a negative correlation. The figure 14 below illustrate the correlations between variables.

In spatiotemporal data, the scatter plot can be very useful in highlighting the effect of changes in one variable on the other variables.

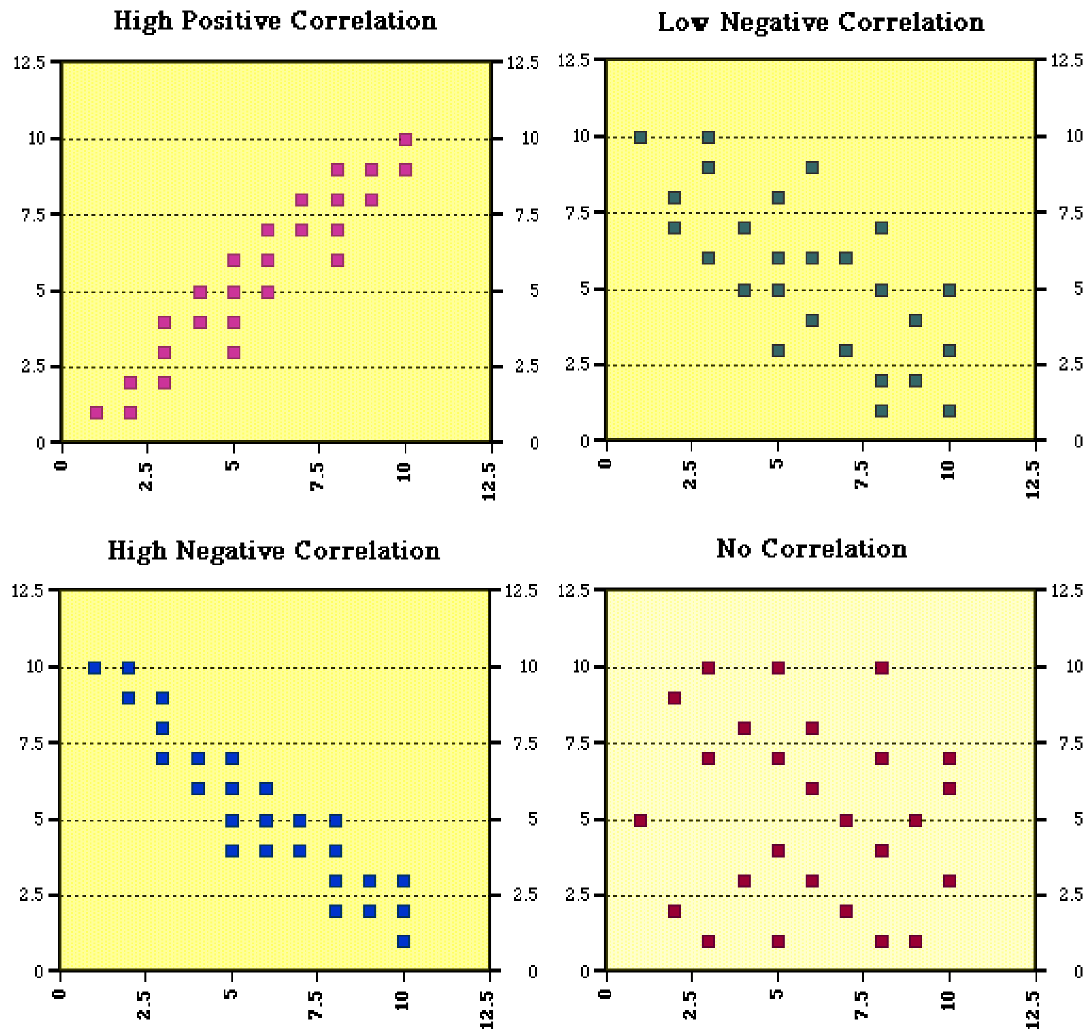


Figure 14 Scatter plot correlations (Source: http://www.mste.uiuc.edu/courses/ci330ms/youtsey/scatterinfo.html)

3. Epidemiological Research

3.1. Introduction

Understanding spatiotemporal characteristics is important in epidemiological research as it assists in identifying trends in space and time, detect problem areas of epidemics, improve assessment of interventions, identifying population at risk, and lead to generation of hypotheses. Aggregated spatio-temporal disease data present unique statistical challenges, in particular one must account for: spatial and temporal dependence; the inherent instability of rare events; errors in numerators and denominators; and problems due to aggregation, which if not addressed may lead to "ecological bias" (Wakefield, 2006).

3.2. Areas of Interest in Disease Mapping

The following are the areas that the researcher in epidemiology is interested in. The objective of exploratory visualization is to highlight these characteristics within the data being presented and to show these within the graphics. Figure 15 by Maheswaran and Craglia (2004) highlights the important factors of the natural, social technically altered environments that should be considered in epidemiological research.

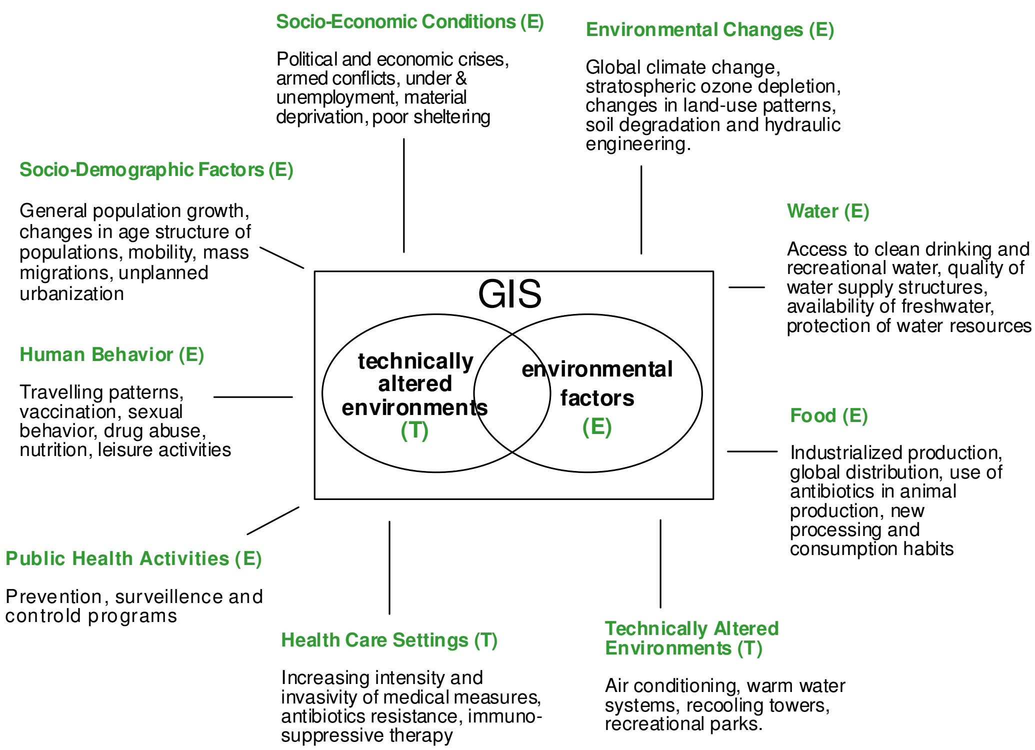


Figure 15 Environmental Factors Affecting the Space Time Patterns Appearance of Diseases
(after Maheswaran & Craglia, 2004)

- Spatial characteristics. Information about the place and its geographic attributes is important in the research. Some spatial characteristics like topography, location, transportation, etc. will have an impact on the way the disease or the epidemic behaves within that locality.

- Demographic characteristics. Understanding the demographic and social characteristics of the affected regions may provide an insight of some of the propellants of the epidemic. Characteristics such as population density and standard of living may appear as regulators of the epidemic impact.

- Epidemic propagation characteristics. This shows how the epidemic is moving around the area. It shows the area covered, the propagation velocity, the seasonal variations, spatial variations. It also shows for how long the epidemic stays in one place and whether is revisits the same place or re-infects the same people over and over again.

- The origin and transmission characteristics. The place of origin of the problem is important in the research. The other important characteristic is how the disease is transmitted from person to person. It is important for the researcher to know if it is a communicable disease, in this case the host, the agent and the environment will be important factors to check for control.

Each of the above factors' spatiotemporal characteristic is analysed to identify its 'hot spot' or 'cold spot'. By using some of the spatiotemporal visualization methods the relationships between the factors can be ascertained. Some of the visualization methods are going to be discussed later in this thesis.

3.3. Visualization in Epidemiological Research

Visualization in epidemiological research is not new. Some previous attempts to use visualization in this field were discouraged by the incapability of the visual tools to handle multivariate data.

3.3.1. Cholera Deaths in London

One of the famous visualization projects was carried out in the 19th century by a British researcher, Dr John Snow, on a cholera outbreak in London. By plotting the spatial deaths locations against the spatial locations of the pumps supplying the water to the residential areas, he managed to identify death occurrences clustering around the pump located on one of the streets. The London research demonstrated the strength of visualization in exploratory work. John Snow has been accredited with the introduction of the map-supported spatiotemporal analysis into inductive infectious disease cause research (Maheswara & Craglia, 2004). In figure 16, the map by John Snow is presented.

Figure 16 Dr John Snow. Cholera Deaths in central London, September 1854.(Source: Tufte, 1983)

3.3.2. Spatio-Temporal GIS Analysis for Environmental Health

A study by Mark et. al(1999) and others focused on extraction of health related information from geospatial lifelines. Geospatial lifelines show individual's locations in geographic space at regular or irregular temporal intervals. The objective of the study was to trace locations of individuals back through time so as to identify spatial clusters of affected people and to determine environmental exposures. The main issues dealt with were; generating hypothesis of spatial and temporal relationships between diseases and exposures, visualization techniques for assessing exposure data and disease incidents or mortality data for hypothesis generating and development of statistical methods to account for uncertainty resulting from factors such as measurement errors and missing data. The data collected consisted of series of discrete space-time samples describing an individual's location in geographic space. The research was based on Forer's space-time aquarium, which is also based on Hagerstrand (1967; 1970) space-time cube (refer to Figure 5).

In figure 17, the timeline shows the locational history or an accounting of an individual's spatial location over a period of time, indicating origins and destinations. In this study the term 'geospatial lifeline was used to refer to the space-time path. The basic element of the space-time path is an observation consisting of three items; an identity, a spatial location and the time. They used the following example to illustrate how visualization can be employed to investigate a problem area, B, where the time lines of three people affected by the type of infection are coinciding. All the three people have at some point in time visited area B. Therefore, more investigations need to be carried out to determine the characteristics of area B.

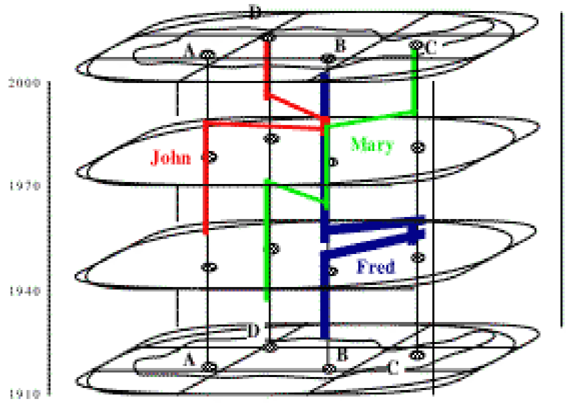


Figure 17 Geospatial lifelines of three individuals over nine decades (source: Mark & others, 1999).

3.3.3. Prostate Cancer Mortality in the United States

This project researched on the occurrence of prostate cancer mortality in the United States, from 1970 to 1994, focussing on space time analysis of spatial patterns of cancer incidents (Goovaerts, 2005). The objective of the project was to look at the disparities between the black and white community's prostate cancer mortality over space and time. In order to address this limitation, he used the Terraseer Space-Time Information Systems (STIS) technology.

This is a system which incorporates temporal information into the spatial data structures and it can handle epidemiological queries that are not possible with spatial GIS systems. In STIS time is an integral part of the data and the software interface. Thus, all views of the data can be animated, from maps to histograms to tables, giving the user a natural, intuitive understanding of data changes through space and time. STIS links chosen data views together, and helps make statistical inference about patterns in the data as well. It allows interaction with and analysis of space-time data (http://www.terraseer.com/products/stis.html). The result of this project was the Atlas of Cancer Mortality in the U.S., covering the period 1950-94 and it provides interactive maps, text, tables and figures showing geographical patterns and time trends of cancer death rates for the whole period. However, there is a 30 day trial version that is available for downloading. Figure 18 is a screen shot from the STIS software illustrating the disparities between the races.

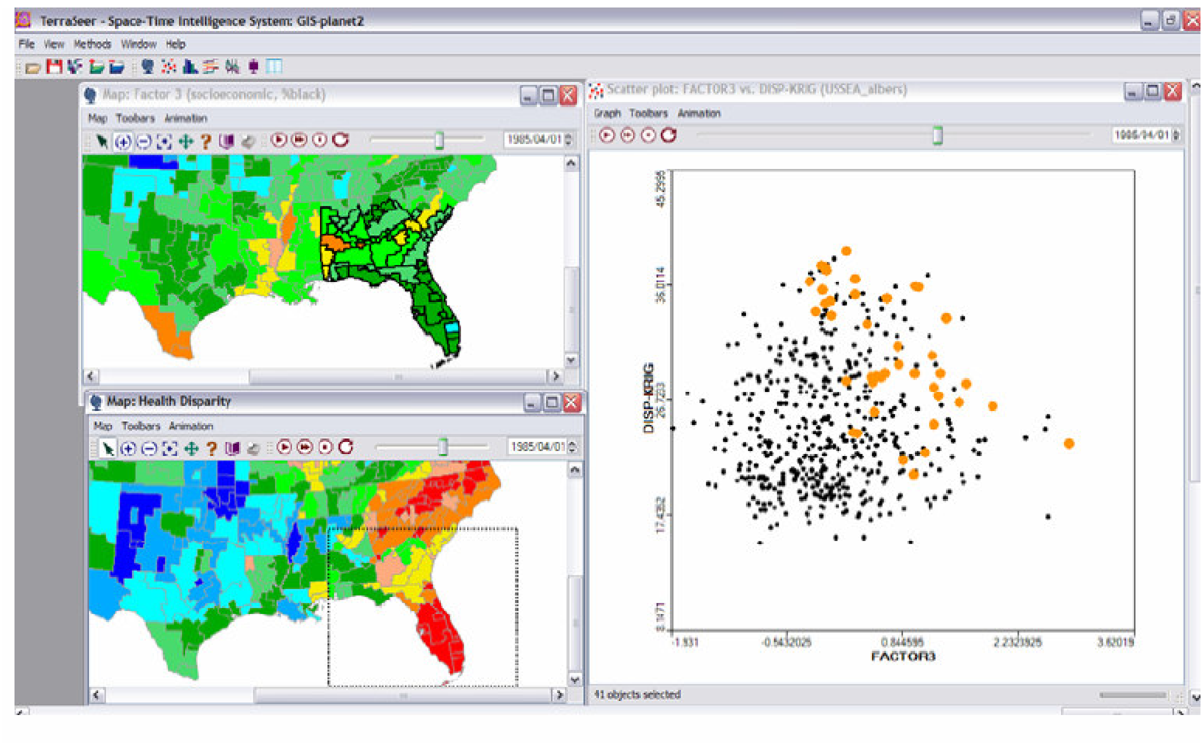


Figure 18 Screen shot of STIS software, showing the "brushing" function in the three linked views. (Source: Goovaerts, 2005)

3.3.4. Spatio-Temporal Analysis Toolkit (ESTAT)

This thesis project by Robinson, evaluated and explored a toolkit developed with the aim assisting in the exploratory visualization of multivariate spatiotemporal cancer data. This toolkit, Exploratory Spatio-Temporal Analysis Toolkit (ESTAT) houses four frames, each of which show the following data display; a scatterplot, a bivariate mapping tool, parallel coordinate plot (PCP), and a time series graph. These four frames are linked and the toolkit has a brushing functionality. The user has a wide variety of interactive functions to access. For assessment of the usability of the tool, Robinson used four assessment methods; card sorting, verbal protocol analysis (VPA), focus groups, ethnographic case studies.

There have been several versions of the ESTAT toolkit in response to the user assessment processes. Robinsons reported that epidemiologist had indicated a quest to initially attempt a confirmatory analysis before or during exploratory analysis (Robinson, 2005). So they have redesigned the toolkit to cater for this demand for confirmation. The following figure 19 is an illustration from the study.

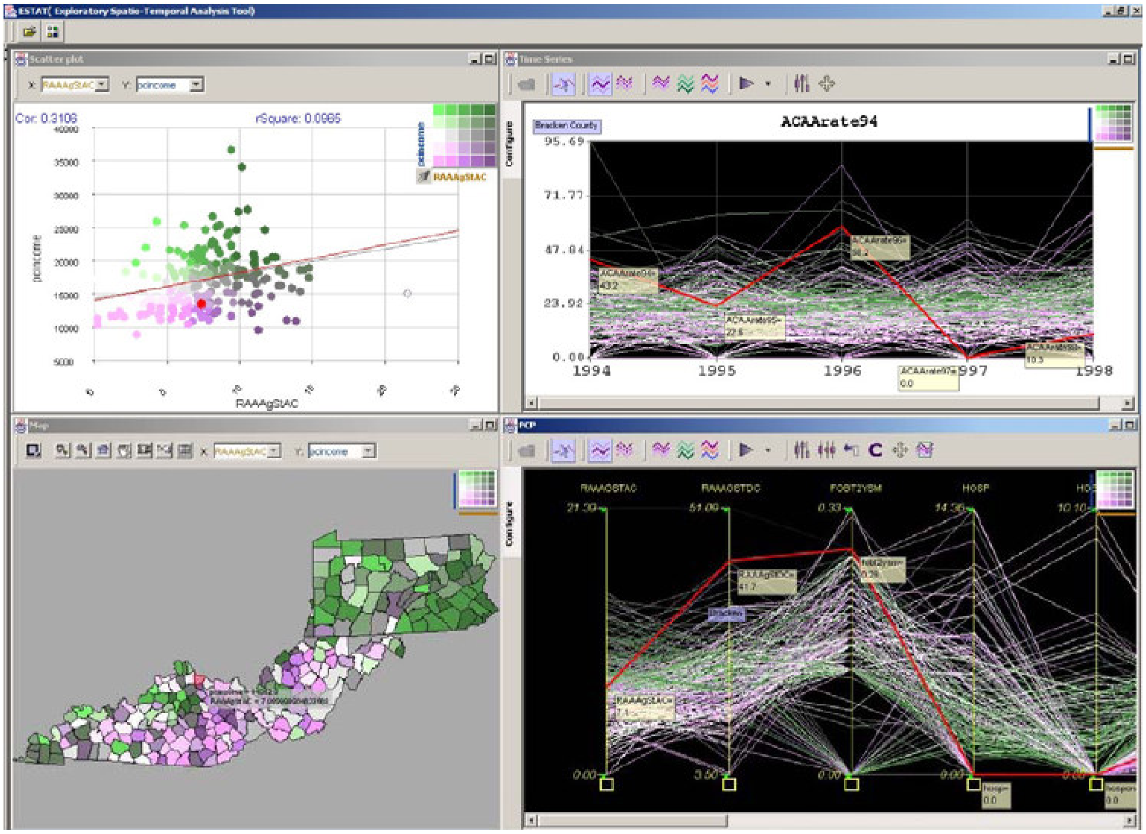


Figure 19 ESTAT Geovisualization Toolkit (source: Robinson, 2005)

The top left quadrant is the bivariate scatterplot. At bottom left is a bivariate map tool. The top right shows a time series graph, and at bottom right is the parallel coordinate plot. This capture shows colon cancer incidence rates and associated socioeconomic variables

3.3.5. Spatiotemporal Modelling and Mapping of the Bubonic Plague Epidemic in India

This research project studied the spatiotemporal evolution of bubonic plague in India during 1896-1906 (Christakos et.al, 2005) The plague was very similar to the Black Death but some of its characteristics were somehow different from the earlier Black Death. Unlike the Black Death, the Indian plague showed spatiotemporal characteristics of disappearing and reappearing in infected areas. More on the Black Death is going to be covered in Chapter 4. Monthly mortality rates for the period September 1896-October 1906 were collected or generated using the proposed space-time method, leading to a complete set of 122 mortality maps in space-time. The maps allow displaying and studying in detail the geographical distribution, temporal evolution and seasonal effects of the deadly epidemic (ibid). Time slices of both mortality and space-time maps were created for each month. The results of the study also revealed a link between the railway system and the epidemic distribution. The space-time maps revealed that the spread speed of the epidemic was higher along trade routes than in the native areas. The plague study showed that it had a strong re-emergence pattern, invaded new areas and slowed down during warmer months. The statistical method of Spatiotemporal Random Fields (S/TRF) was used to evaluate the uncertainty sources concerning the space-time distribution of the epidemic. Animation of the time slices was used to highlight the space-time propagation of the epidemic. The following maps in figure 20 are part of the set of the spatiotemporal maps for the Indian epidemic.

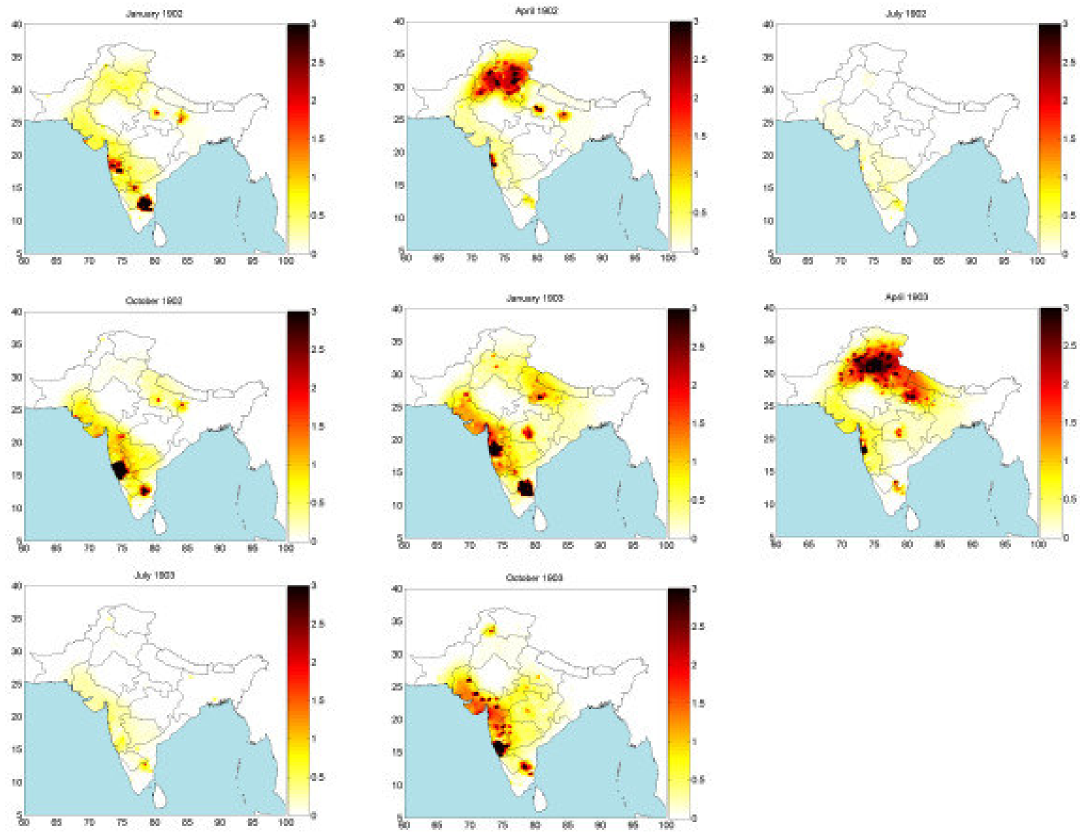


Figure 20 Space-time mortality rate maps during 1902–1903.

4. Case Study Data – Black Death

4.1. Introduction

The material in this chapter has been summarised from the case study, 'Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death' (Christakos et. al, 2005). The Black Death was an epidemic which is understood to have come to Europe in October 1347 and was in the region until around January, 1351. Europe, in the mid 14^{th} century, was occupied by fewer people than the present and they did not move much from place to place. Superstition played an important role in local communities, poor hygiene conditions prevailed, and there was no significant medical advances made during this period. There was a great famine in the region from 1315 to 1317. Also, the famous Hundred Years War had begun in 1337.

The epidemic started in Sicily, Italy, in October 1347. It moved to Austria and Switzerland in 1348. From Austria it moved to Germany and the Czech Republic. In Scandinavia it started in April. Marseille, France was the origin of the epidemic in France in December 1347. In June 1348 another front originated in France. The two fronts in France started moving in the opposite direction and they met in November 1348. It then almost disappeared in winter and resurfaced in May 1349 in Belgium and in June 1349 in Switzerland. At the same it then entered Germany. It appeared for the first time in the Iberia Peninsula in April 1348. Another front then came later in May of the same year from France. A third front started in the north-western part of Spain. The three fronts met in central Spain around October 1348. The epidemic entered into Germany from France, Austria, and Switzerland. The forth front to enter Germany came from Luxemburg in early 1350. Ireland got its first bout in August 1348. (Christakos & others, 2005).

4.2. The Black Death

The symptoms stated below were found in most people infected by the disease.

- Formation of buboes in different parts of the body up to the size of a chicken egg, mostly in the groins, armpits and neck. Buboes were painful inflamed lymphatic nodes that sometimes were opened to alleviate the suffering and increase the likelihood of survival.
- Severe inflammation of the throat and lungs which was evidenced by coughing of blood. This symptom is understood to have led to eminent death within 2 days of showing symptoms.
- Darkish (Black) blotches produced by bleeding under the skin and skin lesions in the form of eruptions covering large areas of the body. These symptoms resulted in the coining of the term 'Black Death' because of how they left the skin looking like and it could result in death within hours of the symptoms showing.

There are various other names by which the epidemic has been cited in literature books. In some English literature is has been named as *magna pestilencia* and *magna mortalitas* which respectively mean Great Pestilence and Great Mortality. Some other symptoms of the disease were prostration, high fever, and unbearable stench from the body. Different forms of the disease were reported from different geographic areas (Christakos et. al, 2005).

4.3. Sources of Black Death Data

It should also be noted that the systematic census of modern days was not in existence in the 14^{th} century. Printing was also not yet developed. Illiteracy levels were very high. The educated were the clergy and the nobility. The source of information was rumours and superstition was strong amongst people.

- Ecclesiastical records. These records were prepared by the church and they indicated the names of the new priests and the names of those they were replacing but however did not indicate the reason for the replacement.
- Parish records. Records from the church turned out to be very useful and some of them give a good picture of the impact of the epidemic on the population. The only missing item in all of these records is the total population of each area.
- Testaments. The wills made at the notaries and hospitals were another source of information. This source of information had problems of being only associated with the rich and the literate. These were the groups that had such kind of exposure. However, the fact that the increase in these could be ascertained was a helpful source of information to map the impact of the black death.
- Tax records. The Domesday Book and the Poll Tax of 1377 provided information about landholding and the ages of those above 14, respectively. The Poll Tax records are the closest records to a census record. Records indicate the total number of those above the age of 14 that could be liable for taxing so as to fund the war between England and France.
- Court rolls. These records indicated the proceedings at the court in terms of property transfers.
- Chronicles. These are contemporary accounts of the epidemic produced by some writers some years after the event. The problem with these as sources of information , is that, most of the compilers of these were not eye witnesses of the epidemic and a number of them have shown conflicting information. So these have not been regarded as reliable sources of information.
- Donations to the church. Church records indicated an increase in donations during the period. This is believed to have been due to the superstitious belief that the epidemic was a punishment from God and donating generously was believed to be a way that the Christians were buying forgiveness.
- Financial transactions. The drop in financial transactions was associated with the raging epidemic.
- Other sources. Death of famous people, personal or open letters narrating developments or sending warnings of the epidemic. Hospital records, occurrence of new cemeteries, tombstones and an abnormal increase in adoptions are the other sources of the impact of the epidemic on the locals (Christakos & others, 2005).

4.4. Characteristics in Black Death Data

Population figures had to be inferred as the modern day census did not take place until the 15th century. The population figures were obtained by a direct reference to the number of residents, tax records and other sources not initially meant for the enumeration of people. Population figures were mentioned in the chronicles but these figures were just simple educated guesses (Christakos & others, 2005).

Other sources of demographic information that were used include; consumption patterns of a settlement, for example, wheat consumed per period. Giovanni Villani, a merchant and chronicler used daily wheat consumption to estimate the population of Florence. The number of baptisms and weddings were other sources that were used for demographic figure estimates.

Most records of the Black Death are not complete and have so many gaps in them. The extract below, figure 21, is an example of the data obtained for the Black Death. This record extract is for Switzerland.

Place	Pre-plague population	Start	End	Mortality
Aargau		Sept. 8, 1349[Biraben 1975:77]		
Basel		May, 1349[Fössel 1987:9]		
Bellinzona		October-November 1348 [Biraben 1975:75]		
Bern		February, 1349[Fössel 1987:9]		
Chillon		December, 1348[Fössel 1987:9]		
Dissentis		December 1348[Sticker 1908:56]		
Engelberg		Sept. 8, 1349[Fössel 1987:9]	January 6, 1350[Fössel 1987:9]	
*Entremont**	4,500[Benedictow 2004:328]			41%[Benedictow 2004:328]
Geneva		August 10, 1348[Andenmatten and Morerod 1987:26]	October 11, 1348[Andenmatten and Morerod 1987:26]	
Lausanne		Nov. 10, 1348[Andenmatten and Morerod 1987:24-25]	August 18, 1349[Andenmatten and Morerod 1987:24-25]	Wills, starting in October 1348 [0 3 6 2 7 5 8 9 5 4 2 0] [Pasche 1998:126-129]
Lucerne		March, 1349[Fössel 1987:9]		
Monthey*	1,200 in 1329[Benedictow 2004:329-330]			43%[Benedictow 2004:329-330]
Nyon		Sept. 20, 1348[Andenmatten and Morerod 1987:26]		
Pfäfers		May, 1349[Biraben 1975:77]	November 1349[Hoeniger 1882:17]	
Rutwil		March, 1349[Biraben 1975:77]		
Saint Gallen		April, 1349[Fössel 1987:9]		

Figure 21 Black Death summary for Switzerland (source: Christakos & others, 2005)

The following three levels of uncertainty were considered, when calculating the mortality values for each place:

- Minimal uncertainty: Where there is a reasonable amount of information about all the assumptions and values that are part of the calculations resulting in a monthly mortality value. A standard deviation of 0.05 x mean was assumed.
- Medium uncertainty: Where at least one parameter is not known which may require taking a value from a neighbouring place or using a global value for the specified region. A standard deviation of 0.1 x mean was assumed.
- Maximum uncertainty: A minimum amount of information is available about the locality of interest, typically the date of beginning of the epidemic and the population size. A standard deviation of 0.2 x mean is assumed.

4.5. Topographic and Social effects on Black Death Propagation and Severity

Studies have been carried out to determine the effect of the geography on the propagation and the severity of the epidemic. Many writers argued that the epidemic hit harder at the ports and large cities along trade routes. While ports and trade centers were in many cases the focal point or the beginning of the epidemic in a region, there is no evidence that receiving the epidemic first made it more virulent (Christakos & others, 2005). Others also argued that mountainous areas had lower mortality than the valley plains. A study by Santamaria (1969) concluded that mortality was proportionally higher in the mountainous areas than along the coast. In some extreme cases for example, Pyrenees and the Alps, were the epidemic never reached had no settlements during the time.

The complexity of the Black Death is that it has no evidence of having affected those with a lower standard of living as compared to those with a higher. Results from the studies carried out indicated no co-relation between poverty and severity of the disease. Some chroniclers reported that in some places it affected the affluent and powerful more than it did to the lower classes (Christakos & others, 2005).

5. Visualization Approach to Support Epidemiologists

5.1. Introduction

This chapter is going to introduce the tools and techniques that can support epidemiologists in the understanding of disease propagation. These representations will then be used to design a prototype to test their effectiveness. Chapter 2 covered the various tools and techniques which are used to visualize multivariate spatiotemporal data. Multivariate data is complex and cannot be efficiently visualized by a single tool or technique, hence, the need to combine various tools and techniques. In selecting the most appropriate tools and techniques, the chapter is going to refer back to the triad framework of Peuquet (1994). Each will be assessed on how well it can answer the 'what', 'where' and 'when'. These basic questions are going to be extended to the type of questions that researchers in public health will be interested in.

5.2. Answering Questions through Visualization

Epidemiologists are interested in understanding the characteristics of epidemics and other public health problems. They would like to know how the epidemics propagate in space and in time, the characteristics of the propagation and other factors that might be affecting the propagation of the epidemics. Visualization can enhance their knowledge about the problems but finding a proper visual tool is the challenge. Some visual tools and techniques do not provide adequate information that can assist in research. This chapter is going to highlight how far the selected tools and techniques are going to address this problem.

In selecting the tools and techniques for visualizing spatiotemporal epidemiological data, typical questions that would be considered of interest to researchers in epidemiology:

- Where is the origin of the epidemic
- What is its spatial/temporal propagation speed
- When did it start
- When did it stop
- Where did it go to
- Where did it come from
- What are its characteristics

The questions that the visual tool should address are related in one respect or the other and cannot be answered using a single tool. Different tools and methods will be required to address these questions, hence, the need for a prototype. The prototype will be made up of the space-time cube, the parallel coordinate plot, the histogram and a scatter plot.

5.3. Space-Time Cube

In this research project, the STC shows the propagation of disease in space and in time. The space time path shows how the epidemic is moving. The path also models the speed at which the epidemic is spreading. In the figure 22 below, assuming that the space-time path represent the trajectory of a disease. There are two stations that the epidemic propagated to before it came back to the initial station. The trajectory shows that the epidemic stayed for shorter periods in each consecutive station visited.

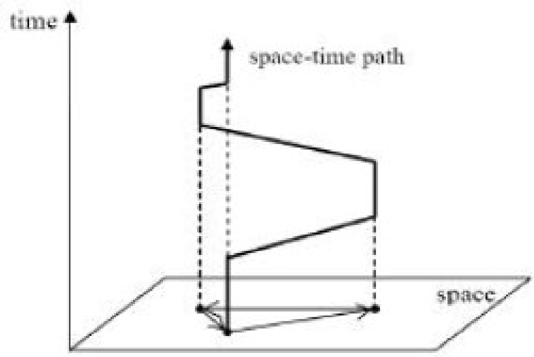


Figure 22 Trajectory of an epidemic

Trajectories also reveal co-location relationships. Co-location relationships identify the epidemic 'hot spots'. Co-location in space and time indicates that more than one trajectory affected the same place. Co-location in space shows that the epidemic has a re-infection pattern as the same place is affected twice. There is an option to visualize trajectories with attributes. The segments of the trajectory can have varied widths or colours.

The space plane slides along the time axis. This allows matching of the stations with the space in time. Two layers of space are going to be designed for the prototype cube. One layer will be the 2D space and the other layer will be a 3D digital elevation model. The 3D space will have vertical exaggeration functions. The cube will allow optional viewing of either or neither space. The 3D layer will provide information bout the impact of terrain on Black Death propagations.

The stations can be visualized with attributes of the station. In this study, the stations will show the mortality distribution of the epidemic at each station. Different colours can be used to vary the stations.

5.4. The Parallel Coordinate Plot

The parallel coordinate plot (PCP) was originally proposed and implemented by Inselberg (1985). The basic principle behind it is that it allows visualization of multidimensional data in one view.

The PCP show the relationship between the various attributes in the dataset. The dataset contains about twelve attributes. Some of the attributes are not useful for analysing for relationships. The most important ones are Place (ID), Country, Longitude, Latitude, Date, Month, Year, V1(mortality), V2 (mortality standard deviation), Starting date, Ending date, Code (data source), Distance (from the sea).

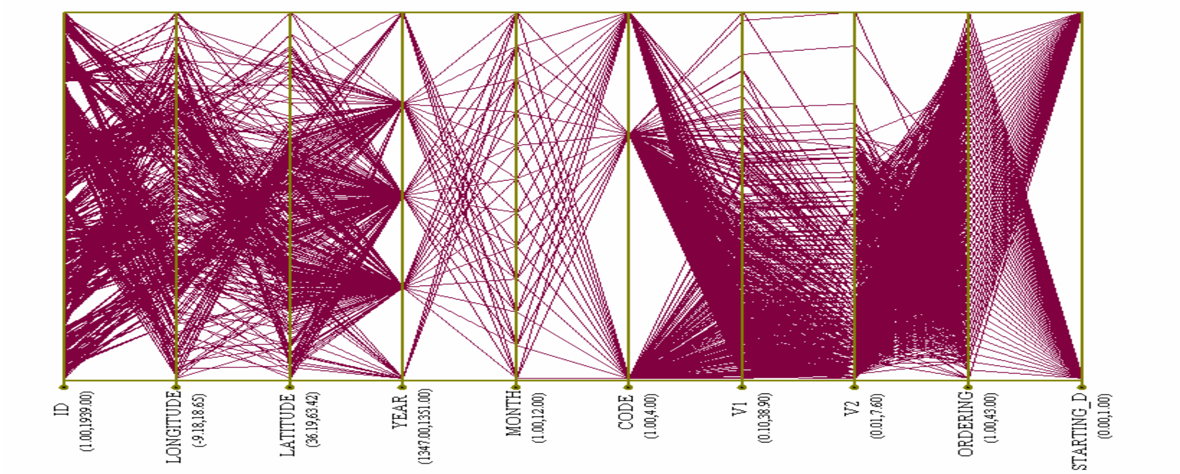


Figure 23 Parallel Coordinate Plot of the Black Death Dataset

Any of the above variables in Figure 23 can be analyzed with the PCP. There is an option to select the variable combinations. The order of the variables can be rearranged two. Order of variables in PCP can influence the creation of patterns that do not really exist. The PCP shows the relationship between the variables.

5.5. Histogram

The histogram, figure 24, in this prototype shows the mortality values. The prototype has an option of which histogram to visualize. It can be shown as a global histogram for the entire data set or can be shown for a selected place in the map view or a station in the STC.

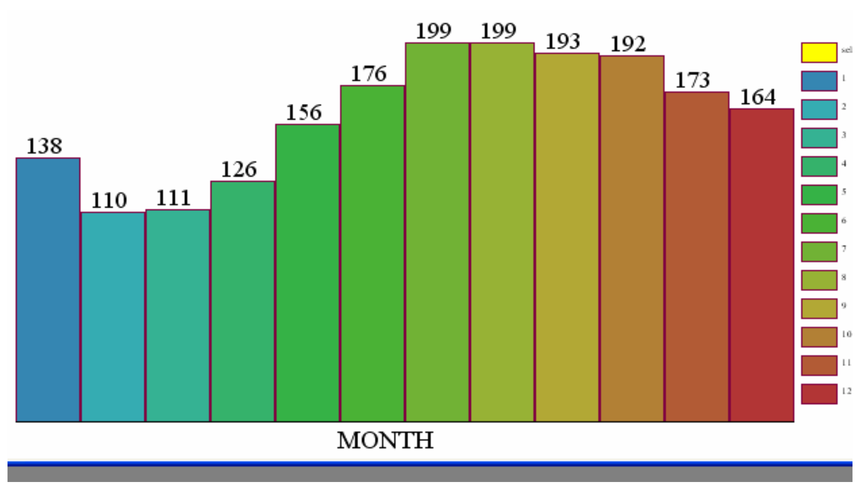


Figure 24 Black Death Histogram

5.6. Scatter Plot

The scatter plot shows the correlation between two variables. Any two variables can be plotted against each other to establish if there is any relationship. If the variables show a positive correlation then it means the positively affect each other, that is, if one changes positively then it means the other will follow the same pattern. Understanding correlations is important in disease control. A correlated variable of the epidemic could be a stimuli of the epidemic and with this knowledge controlling is much easier. The example below, figure 25, is a scatter plot of mortality plotted against standard deviation.

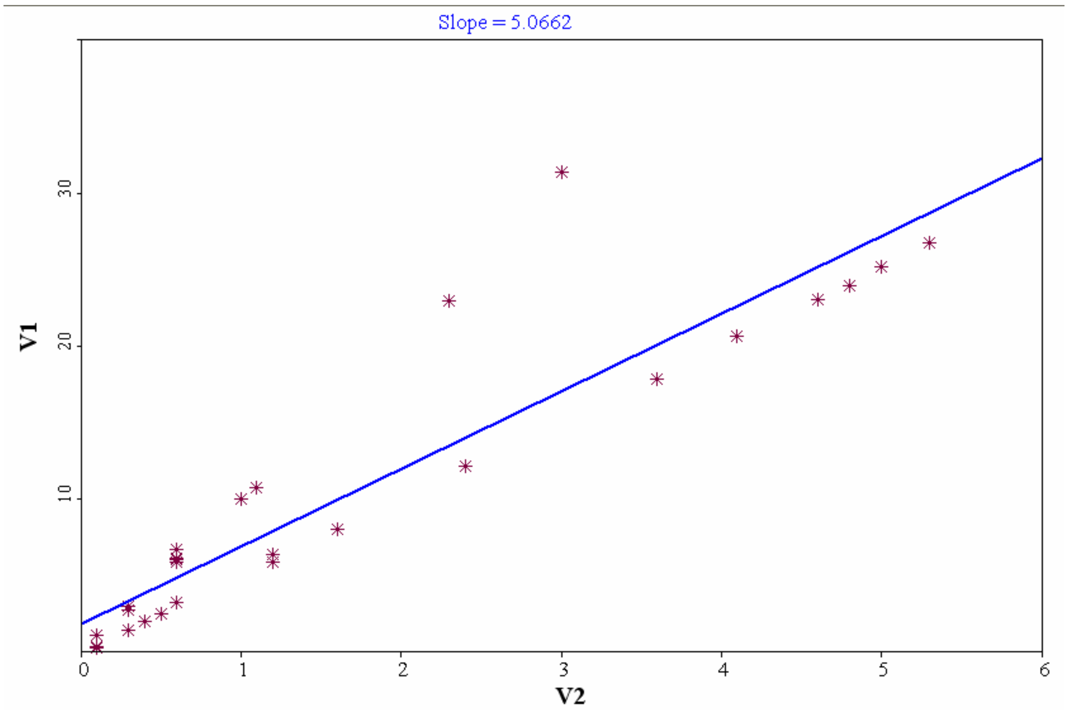


Figure 25 Scatter Plot of the Black Death

5.7. Multiple Views, Linking and Brushing

The prototype have four views displayed at a time. The views are linked. The brushing function is also part of the prototype. Brushing allows user to highlight corresponding objects in all views. Queries are made by using information icon. Views which are not in use can be turned off. The four views are; space-time cube, parallel coordinate plot, scatter plot and histogram.

The multiple views, linked and with a brushing functionality will answer the three basic questions, 'what', 'when' and , 'where' all in a single view.

5.8. Summary

When working with multidimensional spatiotemporal data visualization, it is important to understand what type of tool or tools can provide answers for the identification questions (what), the location questions (where), and the temporal questions (when). Due to the interrelated characteristics of these three basic questions, the resultant solution is an interactive multiple linked view environment.

6. Prototype Development

6.1. Introduction

Of all the tools that were introduced in the previous chapter, the tool that has not been fully implemented in space-time visualization is the Space Time Cube. Previous attempts to use this technique had been made difficult because of the lower levels of technological advancement during the past eras.

6.2. Case Study Data

Data from the Black Death epidemic is going to be used to test the usability of the prototype. The data was collected for a period covering 40 months, which is the estimated time that the epidemic was believed to have gone through Western Europe. The data was compiled by Christakos and others for the publication titled, "Interdisciplinary Public Health Reaasoning and Epidemic Modelling: The Case of Black Death".

6.2.1. Database Design

The dataset that was supplied by the authors of the case study book had already been processed to deal with the incompleteness and uncertainty. As a result, there was not much preprocessing required to be performed on the dataset. The generated data resulted in 8 relational items; Place, Country, Longitude, Latitide, Year, Code, V1 (mortality) and V2 (variance).

The Place item contains the names of the Medieval age cities as they where known in the 14^{th} century. The spatial location of the city is given in geographic coordinates. The name of the country is also given as a separate item. The country names are the current country names as they are known today. The time is split between the year and the month. The code is an indication of the source of the data. They range from code 1 to 6, with 6 indication absense of mortality. The different codes have different levels of uncertianty associated with them. The item V1 refers to the monthly mean estimated mortality calculated for each point. The V2 item is the variance for each point in space and in time.

The date came as two separate columns of Month and Year. These two columns were combined into a new column called Dates. However, the two columns of Month and Year are still available as additional items that can be incorporated into the analysis. This was done so as to allow analysis of data which might require just one of the variables. Another addition to the dataset that was made is the distance of each station from the shoreline.

6.3. Prototype Development

The main objective of this project is to develop a prototype that will show how space-time visualization can be achieved by using multiple linked views in exploratory epidemiological research. Multivariable data characteristics cannot be efficiently presented by a single tool. To be able to provide more insight into the data there is need to combine a variety of visualization tools. The prototype that is going to be designed in this project will comprise of the space-time cube, the parallel coordinate plot, the bar graphs and the scatter plot. The prototype is designed by creating plug-ins to the UDIG programme. UDIG is an Open Source software and this makes it more appropriate and easier to use.

The prototype that is going to be designed in this chapter is going to consist of four linked views. The views will have interactive functions which will provide the user with such choices of which view to turn off and which to turn on. The design will also ensure that the user can interact with the data in any of the views. This function is important in enhancing exploratory work as it provides the user with more control of how the data can be displayed.

6.3.1. The Space Time Cube

The Space Time Cube (STC), that is going to be used in this prototype, was developed as a plug-in software for UDIG. The STC can be used with a number of data formats. It works with both raster and vector layers and both forms of data can be visualized at the same time. The data to be visualized in an STC, in 3d space, has to have a data/time attribute in it. The strength of the STC that makes it ideal for this project is its ability to show the phenomena trajectories. In epidemic research it is important to understand how the disease is moving in space and in time.

This STC offers various options to deal with raster layers. This function will allow the importation of images from different sources for viewing in the prototype. It is designed to incorporate a digital terrain model in addition to the map layer at the base of the cube. This allows the user to visualize the influence of terrain in the propagation of the epidemic. The prototype will have the option of using both images at the same time or selecting one of the two depending on the analysis to be carried out. This functionality allows the user to view both the terrain as an extra height dimension and the time as Z dimensions, in one view. This enables data with four dimensions, X,Y,Z and T, to be visualized without having to switch views or switch off some of the dimensions. The STC will also have a function which allows control of vertical exaggeration of the digital elevation model. Layers can be switched off and on at the user’s discretion. It also provides an option of projecting referral lines to either x, y or z dimension. The figure below, figure 26, show a UDIG session with the map view and the STC with guide lines projected to the xy plane.

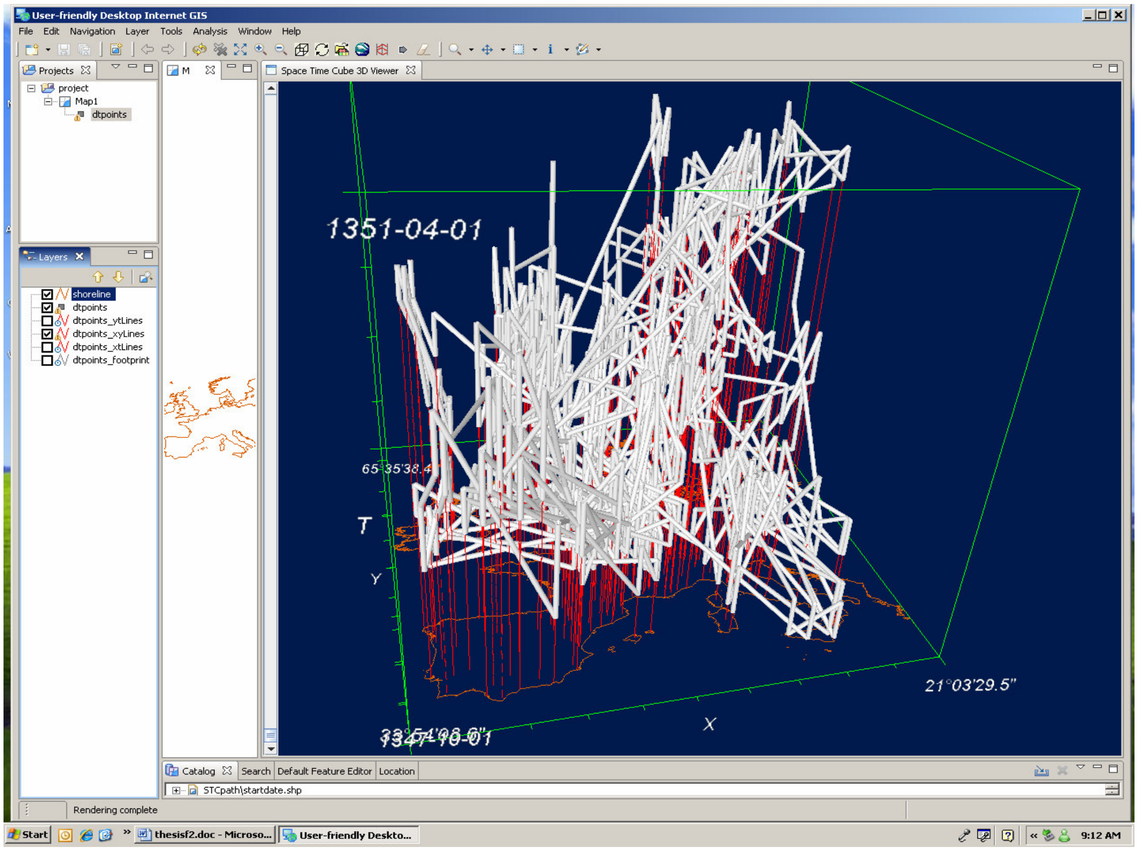


Figure 26 Screen shot of the STC showing a space-time path.

6.3.2. The Parallel Coordinate Plot

The PCP incorporated into this prototype, will allow the user to select the variables to visualize. This function will be useful in dealing with the problem of overloading. Overloading is a problem where the amounts of observations have become so huge that no meaningful observation can be achieved by the user as the patterns become distorted.

An improvement that has been made to the PCP in this prototype is controlling the order of the variables being considered (ref. Figure 24). Selecting the order of appearance of the variables is important to avoid the problem of wrong patterns being visualised. The graphic might show some exaggerated relationships in some places where the relationship is not very strong if too many variables are displayed at the same time. This PCP will allow the use to select the variables to analyse and the order of the variable axis. The data from the PCP will indicate the relationships that exist between the variables.

Figure 27 Parallel Coordinate Plot (PCP) showing the Black Death variables.

6.3.3. The Graph

The graph in this prototype will display a histogram when nothing is selected or queried. When a query is executed about a particular object, the graph will show a line graph. The line graph will only show the distribution of total mortality against time at that particular point. The time can be Date, Month or Year. The idea to show the line graph instead of a bar graph was to allow display of information from more one point. This will allow comparison of data from a number of points.

6.3.4. Scatter Plot

The scatter plot will show the correlation between the selected two variables. The prototype will allow the user to select the variable for plotting. Correlation information is useful in finding the influence of one variable on the other. For example, if the spread of the epidemic shows a positive correlation between the mortality and distance from the shoreline then it will give the researcher a starting point of looking into how these two variables could be related. Data from the scatter plot will reduce time wasting by concentrating on variables which indicate some relationships.

6.3.5. Linking and Brushing

Linking views helps in exploratory visualization task as the linked views will act as one component. The views become coordinated allowing tasks performed in one view to be reflected in the other. The term brushing refers to any dynamic query tools where query results are presented by marking rather than by filtering (Andrienko and Andrienko, 2005). The diagram below, figure 28, shows the linking and brushing for the Black Death data. The yellow coloured portions of the graphic show the brushed items.

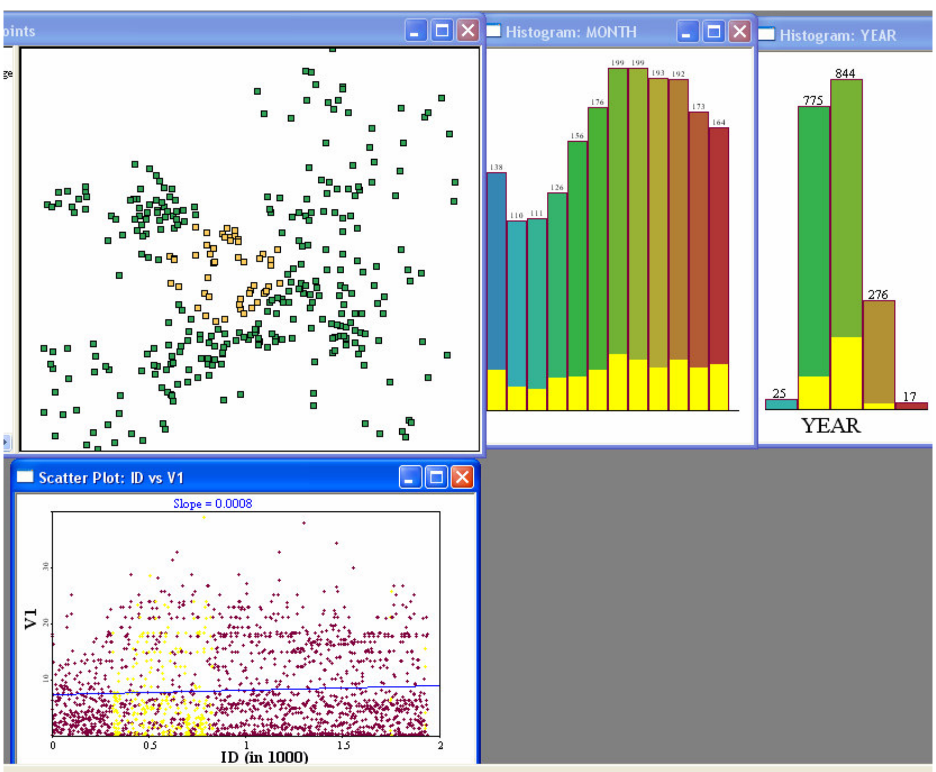


Figure 28 Linking and brushing.

6.3.6. Digital Terrain Model

One of the objectives of the prototype was to introduce the height dimension in the space-time cube so as to give the user more information to work with in exploratory work. In order to achieve this, a digital elevation model (DEM) of the area is required. Construction of a DTM requires elevation data, which can be obtained by various earth observation methods. For this research the GTOPO/SRTM30 data was used as the source for the DTM. GTOPO30 is a global digital elevation model with a horizontal grid spacing of 30 arc seconds, which approximates to one kilometre. The GTOPO DEMs are available in the tiles.

6.4. The Prototype Interface

All in all the prototype will be able to visually display data as a cumulative graph or a point line graph, a parallel coordinate plot, a site map, a space-time cube and a scatter plot. Figure 29 shows the prototype interface displaying the histogram, the parallel coordinate plot, the space time cube and the map.

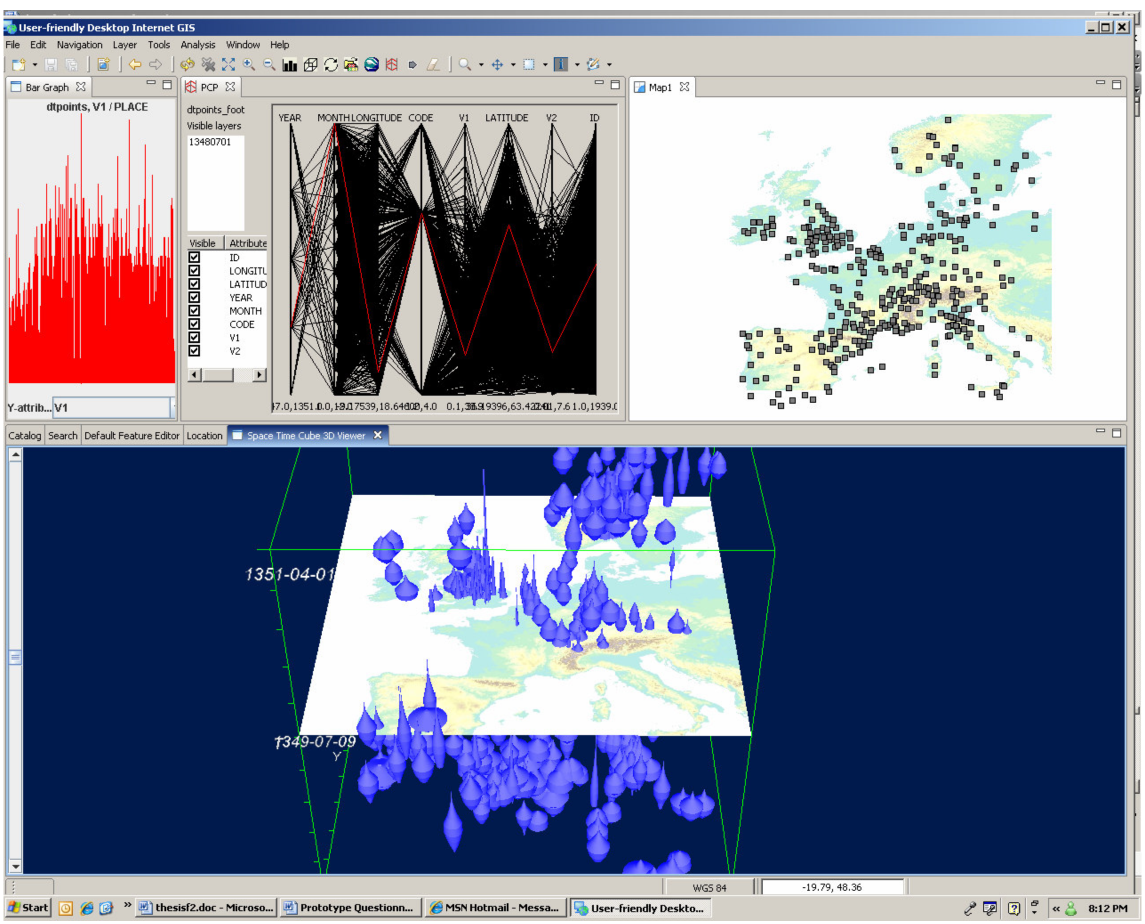


Figure 29 The prototype showing the bar graph, the parallel coordinate plot, the site map and the space-time cube.

7. Prototype Evaluation

7.1. Introduction

One of the main objectives of this research project was to develop a tool that will assist in public health research by provision of an effective exploratory visualization tool. To evaluate if such a tool will be of much use to the intended users, a usability test was conducted. Evaluation is a procedure for checking the suitability of a design and how well it achieves the aims and objectives of the design. There is need to check if more knowledge is gained from use of the prototype when compared to the other methods that are in use or have been used before. One of the advantages of getting the usability evaluation is to get feedback on the performance of the prototype so as to go back to the design stage and look at what can be improved to achieve its objectives.

Evaluating exploratory visualization is more difficult than evaluating other forms of visualization. The idea behind exploratory visualization is to provide visualization tools that allow researchers to come up with characteristic information about the data that would call for further research. As a result of the nature of exploratory work, the assessment of the efficiency of a tool is made difficult as there is no preset target that should be achieved. A pattern may or may not be present in phenomena and failing to identify one does not necessarily mean inefficiency of the tool or of the researcher. In some instances there will be nothing to reveal. The other problem that is faced in usability evaluation of exploratory work is that, at times it takes some time before a researcher finally comes up with some notable observations. Some time is also spent learning how the tool is operated at the expense of time being spent on how efficient the tool is in achieving its objective. There is no set time limit that one researcher has to be exposed to data so as to become familiar with it in order to comprehend the relationships within the data. Knowledge about the topic being evaluated can also determine the evaluator's comments on the prototype.

There are several evaluation methods that can be used in the usability assessment of the prototype. The following section is going to briefly point out some of the advantages and disadvantages of each method.

7.2. Evaluation Methods

a) Think aloud user testing

This is a performance testing procedure where the user's interaction with the product is videotaped and later analyzed, as a way of improving the product. The testing is ideal for one participant at a time. The test can either be scenario or task oriented. The users can be new to the discipline or domain experts. The user is given instructions to think aloud as they work through the task. In this test method there is very little interaction between the user and the experimenter (Axup, 1998).

The main advantage of this method of evaluation testing is that it gives a very close approximation to the actual feeling of an individual's usage of the product. The other advantage is that it requires fewer users, which might lead to less experimental costs.

The main problem with this method is that this kind of scenario is unnatural to the user and it rarely works well. The other problem is associated with language. The user must be fluent in the test language otherwise there is need to translate as the user thinks aloud.

An extension of the simple think aloud method is called the constructive interaction (Axup, 1998).
In this case instead of one person being observed, two people are being observed solving the same problem. In this setup, tasks are solved collaboratively. The main advantage of this method is that it gives more insight into competitive problem solving methods and opinions.

b) Focus Groups

This form of evaluation is normally done during the product conceptualization phase. It involves identification of the potential users of a product or service and having them get involved in the design of a product by asking them their opinions about the potential product (Axup, 1998). In this method, a group of users is asked for their opinions about the product. The methods might involve making use of scenarios or demos. The main advantage of this method is a large quantity of possible design choices and new features to the product. This evaluation method also acts as an assurance that the users will be interested in using the product as they feel to be part of the total product. The disadvantages of this method is that it works well with a customer specific product, does not test users actual interaction with the product, and also that it records what users want, not necessarily what they would actually use.

Another form of focus groups is the 'expert review'. This method makes use of people having sufficient experience or an advanced degree in a related discipline. There are no guidelines nor scenarios or tasks here since the process is dealing with experts. The advantage of this method is that it uses experts so there is a possibility of covering the entire problem space. The main disadvantage of this method is that it may be difficult to standardize design changes. The other problem associated with this method is inherent bias (ibid).

c) Guidelines

Guidelines are procedures which are supposed to be followed in designing a particular product or service. A guideline assessment or evaluation examines whether a product is in line with the stated guidelines. This method of evaluation is more applicable to solutions to known problems (ibid). It also has the advantage of being a cheaper way of assessing the usability of a product. Guidelines have the disadvantage of requiring a long time to construct and some may show bias towards some standards. Guidelines lack innovation as they do not allow experimenting outside the stated guides.

d) Questionnaire

The questionnaire method consists of a series of questions that should be answered by the user for the purpose of gathering information (Deigham, 2005). These have the advantages of being faster and cheaper to implement. There are various forms of questionnaires that are in use, which fall into two major categories; structured and unstructured. Structured ones guide the user into selecting responses from pre-formulated a set of answers. There is not much room for the user to state some other things which are not provided in the questionnaire. With an unstructured method, it gives room to respond to the question the way that the user feels is the appropriate response. Structure questionnaires are ideal where the information to be collected is huge and when the process needs to be computerised. Unstructured questionnaires are ideal for cases where few users are being considered for evaluation and where it is important to understand the different expressions of each individual user.

7.3. Evaluation Method Selected for this Project

The method that was selected for evaluating this prototype is the unstructured questionnaire. This method was selected because of it is a quicker, does not require special environment setup and that it allows expression of user's views. The time that was available to evaluate the prototype was very short. Initially the evaluation method that had been planned for this project was to have epidemiological researchers evaluate the prototype using the think aloud method. However, because of the time constraint on this project, the unstructured questionnaire was chosen as the best alternative. As mentioned before, this method is easy to implement and more manageable for small user groups. The following sections are going to give a brief overview of the evaluation and the nature of the evaluators and their comments about the usability of the prototype.

7.4. Evaluation Procedure

The evaluation was planned in such a way that the prototype would be introduced first by explaining what it is all about, how it functions and what the objectives of it being designed where. Then the data set to be used for the evaluation was explained. After the verbal introduction, the prototype was then run and the features of the prototype where explained to the evaluators. Since the prototype runs as plug-in of another fully functional program, the host programme was also briefly introduced to the evaluators.

The animated version of the same dataset was also played to demonstrate the alternative visualization method. Then, a small demonstration was carried out to illustrate the capabilities of the prototype, using another data set of an earlier time series project.

The main phase involved working with the actual case study data. The data set was loaded and the evaluators were instructed to spend some time doing some analysis operations. After about checking the data with the prototype, the evaluators were then asked to answer some questions contained in the questionnaire. These questions were all open ended questions, allowing them to express how they felt about the usefulness of the prototype in exploratory work and its functions.

7.5. Evaluation Team

The evaluation team comprised of five members. Two of the members are Phd. students and the other three are MSc. students at ITC. All the evaluators have good GIS background and some knowledge about public health. All have good experience working with computers and working with maps. The detailed information about their backgrounds and experience can be found in the appendix of this thesis.

7.6. Results from the Evaluation Process

In this section, the results from the five evaluators are going to be presented. Some general and specific questions were posed to the evaluators. They were also invited to include their own comments and recommendations. The following summarised responses were given by the evaluators in response to the questions and task assigned to them. The full questionnaire is in the appendix.

- **Identification**
  The first and second questions asked the users whether they could identify and explain changes. The responses to this question were that generally all of them could identify a number of space-time changes. A few problems were encountered, mostly to do with the development, which made it difficult for some of the benefits of the prototype to be explored. These were noted and would be rectified.

- **Change Detection**
  The question about the space-time path received mixed comments. Some had no problems in using this function but some appreciated the concept but stated that the prototype was having problems in making this function comfortable to use. They suggested that the space-time path should have a direction component in it. It was difficult to continue from where there is a co-location in space and in time.

- **Elevation Model**
  The question of the benefits of including a digital terrain model to the prototype also received mixed comments. It was not appreciated much mostly because of the scale. They commented that it would be more applicable if the prototype would have a zooming function which would provide more detailed information at different scales. Only one commented that it was useful.

- **Brushing**
  This function was accepted by all as useful in exploratory work. Some problems were encountered with brushing in the linked views. The prototype's brushing show flashing objects which later go off after flashing for a few moments. This made it difficult for the users to effective make use of this function.

- **Multiple Linked Graphics**
  The multiple linked graphics approach was accepted by all that it helps in improving their understanding of the dataset. Some suggestions were made to improve the linkage by including more.

7.7. Summary

The conclusion reached from the evaluation process is that the prototype can be a useful tool in exploratory work. The multiple linked graphics method helps to improve the understanding of the dataset and the relationships between the data variables. However, it was also noted that the prototype still requires some further development so that it becomes fully functional. The comments from the evaluator also highlighted the need for having the prototype tested by the intended users.

8. Conclusion and Recommendations

This research focused on the use of visualization in multivariate exploratory data analysis. The Black Death epidemic, of the 14^{th} century was chosen as the case study to illustrate how the process of exploratory visualization can be achieved. The main aim of the research project was to come up with a tool that will enable researchers dealing with time series data gain more insight of the data. The prototype was designed for this project after considering the questions that need to be answered in exploratory work, as an individual tool will not be able to address them all.

The focus of the project was to introduce a new concept of visualizing spatiotemporal data that offers a better understanding of the temporal events. It aimed at providing the user with more interactive and dynamic tools. Interaction improves the understanding of the 'what', 'where' and 'when' questions about the subject of interest. In some cases it attempts to answer the 'how' part to some extent. Looking at previous attempts to tackle the issue of space-time visualization it was noted that the space-time cube (STC) which was introduced in the early 1970s had not been utilised because of the technical environment then which made its implementation difficult. After weighing the advantages of the STC in space-time visualization, the project then concentrated on how best it can be combined with other tools to come up with an effective exploratory visualization tool. As a result the prototype was developed with four dynamic linked views. Then it was later tested to check its usability. The results from the testing revealed that the space-time approach of the STC was an acceptable tool and if given enough attention, it can be a very useful tool in the study of spatiotemporal phenomena and in exploratory work.

8.1. Research Questions

The following research questions were drawn up at the beginning of the research project.

- What work has already been done in this area of space-time visualization of epidemics and how much has been achieved?
- Which alternative visual representations are suitable to deal with 'epidemic' space-time data and can also handle their sometimes uncertain and vague character?
- How effective are these representations?
- How can the solution be extrapolated to deal with real-time monitoring of spreading diseases?

8.1.1. Previous Work on Visualization of Temporal Events

A number of research projects have been undertaken to address the issue of visualization of temporal events. Many have tried different visualization tools and some combinations of a number of tools in attempting to improve space-time visualization in exploratory work.

A closer look at most of the projects that have been undertaken reveals that there have been attempts to address the problem of visualizing time as a dimension but including this dimension has not been easy. There are examples of work that has been done and have used the space-time cube, however, most of them have been used this tool to map the incidence of 'hotspots' or places of common occurrences that might need attention.

In most of the other projects, animation seems to be the preferred method of representing the temporal dimension of the data. The problem with this method of representing time is what is referred to as the 'short term memory' effect. The human brain is understood to be capable of only holding between five and nice images at a time, depending on the complexity of the images. So in order for the brain to effectively register images, the animation has to be played over and over again. The other option of multiple time slices has not been able to give reasonable results because of lack of association with the timeline.

8.1.2. Alternative Visual Representation

Concluding from these previous attempts to represent time visually, it became evident that the space-time cube can prove to be a good alternative method for visualizing spatiotemporal data in epidemiological research. With its visual strength of showing both the phenomena distribution in space and in time, and its capability to show space-time trajectory lines, the STC is an effective space-time visualization tool.

8.1.3. How Effective are these Representations

Effectiveness is defined as a measure of producing the intended or expected result. The aim of this research project was to produce a tool that can be used for visualizing spatiotemporal data so as to assist in exploratory work. The expectation from exploratory work is not measured by a standard. Exploratory work can produce nothing or can have as many findings depending on the subject that is being researched on.

However, the design of the prototype was meant to efficiently show the distribution of phenomena in both space and time. So the effectiveness of the prototype was measured by assessing how well it managed to represent the phenomena in both space and time, which will lead to the detection of patterns and other relationships in epidemiological research. To measure the effectiveness of the representations, the Black Death case study data was used to test the usability of the prototype. Effectiveness was based on the user comments on whether they were able to carry out some prescribed tasks.

The exercise was carried over a period of two hours and in normal exploratory work, this is not enough time for one to have managed to get a good understanding of the nature of the data. The other thing is that effectiveness might be stated as low being a reflection of the evaluator and not necessarily the nature of the prototype. However, since a total of five evaluators evaluated the prototype this problem of inefficiency of the evaluator has the chance of being neutralised.

8.1.4. How can the solution deal with real-time monitoring

Real-time monitoring allows the points to be captured in the field immediately displayed on the interface. This process will result in information about public health being transmitted simultaneously as it is capture. The monitoring system can be designed so that it can be transmitted in various forms so as to be suitable for visualizing on various platforms. Currently information can be visualized on cellphones, laptops, PDA's, televisions and personal computers.

Due to lack of time, this part was not part of the main design of the prototype. It will be put forward as future work.

8.2. Recommendations

Some modifications to the prototype will improve its effectiveness as an exploratory tool. In improving its design the following recommendations can also be considered.

- Design the application of real time monitoring of epidemics. The design should take into account the viewing platforms that are in use, like PDA's, mobile phones, iPods.

- From the evaluator's comments, the prototype can be designed further to include additional visualization tools, and improve on zooming to include informative zooming where more information is obtained as the user increases the zooming.

- The multiple linked view can be designed in such a way that it offers the option of animating the whole prototype view or manually control the viewing.

9. Bibliography

AvRuskin G.A, Jacquez G.M., Meliker J.R., Slotnick M.J, Kaufmann A.M., Nriagu J.O., (2004), Visualization and Exploratory Analysis of Epidemiologic Data Using a Novel Space Time Information System, International Journal of Health Geographics, 2004, 3:26 http://www.ij-helthgeographics.com/content/3/1/26

Andrienko N., Andrienko G. (2006). Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach. Springer, Berlin. Germany

Blok C.A., (2005). Dynamic Visualization Variables in Animation to Support Monitoring of Spatial Phenomena. Phd Thesis. ITC, Enschede, Netherlands

Christakos G., Olea R.A., Serre M.L., Yu H.L., Wang L.L. (2005). Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death. Springer, Berlin, Germany.

Clarke C.K., McLafferty S.L., Tempalski B.J., (1996), On Epidemiology and Geographic Information Systems: A Review and Discussion of Future Directions. Emerging Infectious Diseases Vol 2. No.2

Edsall R.M., (2003), The Parallel Coordinate Plot in Action: Design and Use for Geographic Visualization. Elsevier, Computational Statistics & Data Analysis 43, 605-619. www.elsevier.com/locate/csda

Goovaerts P., (2005), Analysis and Detection of Health Disparities Using Geostatistics and a Space-Time Informations System, The case of prostate cancer mortality in the United States 1970-1994. GIS Planet, Estoril, Portugal.

Hinman S.E, Blackburn J.K., Curtis A., (2006). Spatial and temporal structure of typhoid outbreaks in Washington, D.C., 1906-1909: evaluating local clustering with the G statistic. International Journal of Health Geographics 2006, 5:13

Koua E.L., Kraak M.J. (2004). Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics 2004, 3:12 http://www.ij-healthgeographics.com/content/3/1/12

Kraak M.J. (2005), Timelines, Temporal Resolution, Temporal Zoom and Time Geography, XXII International Cartographic Conference (ICC2005), Spain.

Kraak M.J., Ormeling F. (2003). Cartography, Visualization of Geospatial Data, 2nd Edition, Prentice Hall. London WIT 4LP

Kwan M., (2000), Interactive Geovisualization of Activity-Travel Patterns Using Three-Dimensional Geographic Information Systems: a Methodological Exploration with a Large Data Set, Transportation Research Part C 8, 185-203, Elsevier Science Ltd.

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. (1999). Geographic Information Systems. Principles and Technical Issues. 2nd Edition. John Wiley & Sons. New York.

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. (2005). Geographic Information Systems and Science, 2nd Edition. John Wiley & Sons. New York.

MacEachren A.M., Chen J., Guo D., (2006), Visual Inquiry Toolkit – An Integrated Approach for Exploring and Interpreting Space-Time, Multivariate Patterns. Proceeding, AutoCarto 2006, Vancouver, WA.

MacEachren A.M., Brewer C.A., (1995), Mapping Health Statistics: Representing Data Reliability, 17th International Cartographic Conference, Barcelona, 311-319

Maheswaran R., Massimo C., (2004), GIS in Public Health Practice, CRC Press, Florida 33431, USA.

Malunda A.A., (2005) Interactive dynamic visualization environment linking animated map and animated parallel coordinate plot for monitoring geospatial dynamics. MSc. Thesis. ITC, Enschede. Netherlands.

Mark D.M., Egenhofer M.J. (1999), Spatio-Temporal GIS Analysis for Environmental Health

Meliker J.R., Slotnick M.J. (2006), Individual Lifetime Exposure to Inorganic Arsenic using a Space-Time Information System. Int. Arch Occup Environ Health, Springer-Verlad 2006.

Miller J.H., (1991), Modelling Accessibility using Space-Time Prism Concepts within Geographical Information Systems. International Journal of Geographical Information Systems, 5, 287-301

Ogao P.J. (2000). Exploratory Visualization of Temporal Geospatial Data Using Animation, ITC, Enschede, Netherlands.

Ott. T, Swaczny F., (2001). Time-Integrative Geographic Information Systems, Management and Analysis of Spatio-Temporal Data. Springer. Berlin

Pent Z.R, Tsou M.H. (2003). Internet GIS, Distributed Geographic Information Services for the Internet and Wireless Networks. John Wiley & Sons Inc. New Jersey

Peuquet, D. J., 1994. It's about time: A conceptual framework for the representation of spatiotemporal dynamics in geographic information systems. Annals of the Association of American Geographers 84: 441-461.

Ratsios, B. (2006), Space Time Cube Plug-in for uDig – Users Guide, ITC, Enschede, Netherlands.

Robinson A.C., (2005), Geovisualization and epidemiology: a general design framework. GeoVISTA Center, Department of Geography, The Pennsylvania State University.

Spence R. (2001). Information Visualization. ACM Press. New York

Tufte R.E., (1982), The Visual Display of Quantitative Information, Graphics Press, Connecticut 06410, USA

van Elzakker C. P. J. M (2004) The use of maps in the exploration of geographic data. Phd Thesis. Utrecht / Enschede. ITC. Enschede. Netherlands

Xia L. (2005). New Methods of Visualization of Multivariable Spatiotemporal Data: PCP Time-Cube and Multivariable-Time-Cube. MSc Thesis. ITC. Enschede.

Yattaw N.J., (1999), Conceptualizing Space and Time: A Classification of Geographic Movement, Cartography and Geographic Information Science, Vol. 26

Yu H., Christakos G. (2006). Spatiotemporal modeling and mapping of the bubonic plague epidemic in India. International Journal of Health Geographics 2006, 5:12

Yu H., Kolovos A., Christakos G., Chen J. Warmerdam S, Dev B., (2005), Interactive Spatiotemporal Modelling of Health Systems: The SEKS-GUI Framework, Report #06-39

Wakefield, J.C., (2006) Spatio-Temporal Epidemiology: Methods And Applications, Cancer Control Research http://cancercontrol.cancer.gov/grants (accessed 10/2/2007)

Wilkinson L., (1999), The Grammar of Graphics, Springer, Verlag, New York

Zhang J., Goodchild M., (2002), Uncertainty in Geographic Information, Taylor and Francis, London EC4P 4EE, Uk

URLs

http://www.statsoft.com/textbook/glosi.html

http://www.terraseer.com/products/stis.html

10. Appendix

<u>Appendix 1</u>

<u>Usability Evaluation of the Prototype</u>

<u>BACKGROUND INFORMATION</u>

1. **Educational Background**
   - Certificate
   - Diploma
   - Masters / MSc
   - Phd
   - Other (please specify)

     ……………………………………………………………………………………

2. **Occupation or Field of Specialization**

   ……………………………………………………………………………………….

3. **Computer Experience (number of years)**

   ………………………………………

4. **Experience using GIS software (number of years)**

   …………………………………………………….

5. **Level of GIS knowledge**

   a) low b) moderate c) high d) very high

6. **Experience using UDIG**

   a) low b) moderate c) high d) very high

7. **Experience using ArcGIS**

   a) low b) moderate c) high d) very high

8. Previous work with Epidemiological Data (Public Health)

a) none at all b) little experience c) more experienced

9. Experience using paper maps

a) low b) moderate c) high d) very high

10. Experience using digital maps

a) low b) moderate c) high d) very high

11. Exploratory work experience

a) low b) moderate c) high d) very high

12. Experience with Time Series data

a) low b) moderate c) high d) very high

13. Familiarity with Black Death Data

a) not at all b) little c) quite familiar

PROTOTYPE EVALUATION DATA

- To run the prototype go to d:\stc\udig_release\eclipse\udig.exe
- Data for the evaluation is located in the folder d:\stc\thesis data\total figures
- Before loading the evaluation data, open a new map
- To load the files, navigate to the folder and then drag the files into the drawing area of the prototype.

14. Using this prototype, can you detect any space-time changes that are taking place within the data?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………..

15. Please can you describe the changes that you have noted as you move along the temporal dimension.

……………………………………………………………………………………….

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………

16. Do you notice any patterns, trends or other relationships within the data. If so please can you explain

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

………………………

17. The Space Time Cube shows the Location of the phenomena in time and in space. It also allows the user to control the space-time position. Comparing with other space-time representation methods, like animation, do you find this functionality adding more insight into your data analysis? Explain.

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

………………………………….

18. Did you find the idea of combining the Space-Time Cube, The Parallel Coordinate Plot, The Bar/Line Graph useful in improving your understanding of the data relationships. What can be improved?

....

....

....

....

....

....

19. What did you like about this prototype

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

………………………

20. What did you not like about this prototype

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

………………………

21. What suggestions would you like to give to improve this prototype

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Appendix 2

Monthly Mortality Data

Content:
All information used in Black Death Study
Columns:
Place/Country/Longitude/Latitude/Year/Code/V1/V2

Column Description:

Column Name	Column Description
Place	The name of city or location in the Medieval age
Country	The name of present country where the place locates
Longitude	The coordinate of the place
Latitude	The coordinate of the place
Code	Indicator for the information type (Christakos et al, 2005) 1: for W~N(V1,V2^2) and Z|W~U(0.75W,1.1W) (p.187) 3: for Z~N(V1,V2^2) (p.180) 4: for W~N(V1,V2^2) and Z|W~T(0.33W,W,3.3W) (p.188) 5 and 6: for Z~0.0, no mortality at all
V1	The parameter for soft information (mean mortality)
V2	Same as above (standard deviation of mortality)

Description of the Black Death data contents. (Source: Christakos et al, 2005)

Appendix 3

Prototype Interface Icons

The icons and what they do in the prototype.

 ***Open the STC***

This icon will let the user open the STC. When this command is issued, uDIG copies the data layers from the currently displayed map into the STC.

 ***Reset the STC***

This will reset the STC to its original starting position. If the user finds out that the operations that have been performed on the STC are now too many and cannot be undone to revert back to the original position, pressing this button will send the STC back to the initial settings.

 ***Load Layers in the STC***

This icon will let the currently selected layers be loaded into the STC. The icon can be used to load new layers or to overwrite the layers that are currently displayed in the STC by calling new ones. On its own, the STC is not capable to automatically load layers. Layers being displayed in an STC when the session starts are simply those layers that were last loaded in the previous session. It should also be noted that the STC can only open layers from one uDIG map at a time.

 ***Open in Google Earth***

Pressing this icon will result in all what is visible in the STC being loaded into Google Earth. However, it should be noted that Google Earth must be installed and running properly on the host machine. The command will create a KML file, the file type that Google Earth works with (at the time of writing, this was not fully functional. The function does not work with other file types).

 ***Open the PCP***

This icon will open the parallel coordinate plot. The parallel coordinate plot loads the attributes of the data displayed in the current map. The user has the option to select which attribute items to display and which to switch off.

 ***Open Bar/ Line Graph***

With this icon, the data presented in the STC will be plotted as a bar graph in a separate window. When the user selects a point from the reference map, the bar graph is then plotted as a line graph showing the distribution of that particular attribute across time.

Appendix 4

Evaluator's Details

Occupation or Specialization	**Candidate One**	**Candidate Two**	**Candidate Three**	**Candidate Four**	**Candidate Five**
Educational background	MSc	Phd	Phd	MSc	MSc
	Civil Engineer	Earth Observation	Geoinformation Processing	Geoinformatics	Geoinformatics
Computer Experience	4 years	8 years	15 years	+10 years	12
Experience Using GIS Software	2 years	7 years	12 years	7	8
Level of GIS knowledge	high	high	high	high	high
Experience using uDIG	low	low	moderate	low	moderate
ArcGIS knowledge	moderate	moderate	moderate	high	high
Experience using paper maps	high	moderate	very high	high	very high
Experience using digital maps	high	moderate	very high	high	high
Understanding of exploratory work	moderate	moderate	high	moderate	high
Background knowledge on time series data	high	moderate	moderate	low	moderate
Familiarity with Black Death	little	not at all	not at all	not at all	little
Previous work on health data	none	none	none	none	little