Chapter 1

Introduction

In this chapter we will describe the background and rationale, objective, expected advantages and literature review of this study.

1.1 Background and rationale

Nowadays, many of research use the geographic information system (GIS) to analyze spatial data. GIS makes their information more interesting and easy to understand. Many people give the meaning of Geographic Information System (GIS), including:

A Geographic Information System (GIS) is a tool for displaying and analyzing spatial data. It uses a relational a database management system for managing spatial data and attribute data (Demers, 1999).

Geographic Information System (GIS) is a computer system used as a tool for manipulation, collection, storage, analysis, and displaying data. It manages all complex data in a format that can be utilized as needed. It uses geography to link with statistical data or attribute data (Phoipikul, 2007).

Geographic information systems are a special class of information systems that keep track not only of events, activities and things, but also of where these events, activities and things happen or exist (Longley et al, 2001).

A Geographic Information System (GIS) is a computer system for capturing, storing, querying, analyzing, and displaying geographic data (Chang, 2002).

In summary, Geographic Information System (GIS) is a tool for manipulation, storage, analysis, and displaying spatial data and attribute data.

The main function of GIS is the analysis of data that is spatial data and attribute data. It is useful for a wide range of purposes. GIS can provide many formats of spatial data and attribute data analysis, including those for managing new groups, creating new frameworks, linking analysis, for a transformation coordinates system, map connection, calculating area, calculating perimeter and calculating distance, using color shade on maps to represent the density of population, showing bar graphs on maps, etc.

Figure 1.1 shows the combining of regions. At first there are regions A, B, C and D. When we combine A with D to make E, we arrive at the last result of regions B, C and E.

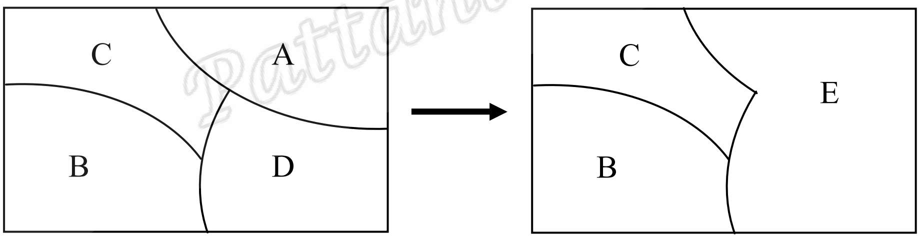

Figure 1.1: The example shows the combining of regions

Figure 1.2 shows an example of selection of regions. From regions called A, B, C, D, E and F, the user selects three regions to new data sets that are B, E and F.

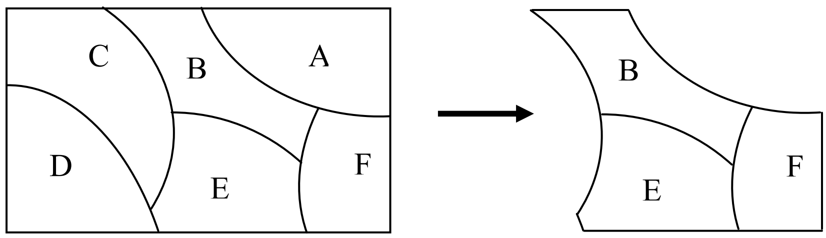

Figure 1.2: The example shows selection of regions

GIS can be applied in various fields such as business, engineering, public utility systems, public health, epidemiology, biology, agriculture and telecommunications. Some examples include health care use of the GIS to decide where to locate new clinics and hospitals and public health use to show patterns and trends of disease outbreaks. Some delivery companies use GIS to decide the routes and schedules of their vehicles, often on a daily basis. Transportation authorities use GIS to select routes for new highways. Fishery companiescan use it to decide how to allocate funds for building sea defenses. Travelers can use it to find their way through airports, give and receive driving directions and select hotels in unfamiliar cities. Farmers can employ new information technology to make better decisions about the amounts of fertilizer and pesticide to apply to their fields. Military can use it to make decisions about which areas to avoid, which are safe, how and where to deploy their troops, equipment, and expertise. In addition, GIS can be used to display floods, hurricanes, traffic and pollution.

There are many types of software that support the analysis and display of spatial data. The user should choose the program that is reasonable for organization of the task and for the budget. Most importantly, the user must have knowledge and capability to use that program. Some examples of GIS programs include: ArcGIS, ArcInfo and

ArcView (developed by Environmental Systems Research Institute [ESRI] Inc), ArcCAD (developed by Autodesk Inc), IDRISI (developed by Clark Labs), ILWIS (developed by International Institute for Aerospace Survey and Earth Sciences, the Netherlands) and MapInfo (developed by MapInfo Corporation), etc. Many of these are commercially available (Chang, 2002).

R is free software and an open-source project, in the sense that we can download it from the Internet, make as many copies as we want, and give them away. R is a functional language and software environment for statistical computing and graphics.

The initial version of R was developed by Robert Gentleman and Ross Ihaka. They are from the University of Auckland, New Zealand. The current design has been influenced largely by S, and more recently by S-Plus. Now, the development and distribution are carried out by R Development Core Team (Paradis, 2002). R is available under the terms of the GNU General public license as published by the Free Software Foundation, in source code format. R also can run on multi-platform such as Windows, Mac OS X, and Linux (Verzani, 2008).

R provides a suite of software facilities for reading and handling data, complex computations, analysis, exploration of statistical models (linear and non-linear modeling), classification, statistical tests, displaying of results, and can generate high quality graphics as a result of its computations.

R is not just a statistical environment but it is also a true computer programming language and it is also an interpreted programming environment. It allows the user to create his or her own packages and the use of R has appealed to academic researchers in various fields of applied statistics. There are a lot of R users, including those using it for medical and public health applications, environmental statistics, econometrics, and bioinformatics, among others. Because of this, R has a large number of packages that extend its functionality, creating extra statistical techniques. R has uncommon features in statistical software, like the possibilities to interface with external programming languages like C or FORTRAN, and to create complete software applications even with a Graphical User Interface or database access. The URL of R is *http://www.r-project.org*.

So, R is an alternative for teaching and learning, especially for students who want to analyze the data of their research, when most of the software that supportstatistical analysis are copyrighted programs, commercially available but usually fairly expensive. On the other hand, R is a command line program and the user needs to know about the syntaxes or the functions in R.

After the program R is set up and opened up, the result will be the RGui (R Graphic User Interface) as shown in figure 1.3. This is divided into two parts; the first is a menu bar that we will call A. There are menus of “File”, “Edit”, “View”, “Misc”, “Packages”, “Windows” and “Help”. The second part is R Console that we will call B. The R Console is an area for the user to write the commands. It has a “>” symbol. The user can write the command after this symbol, to call the function.

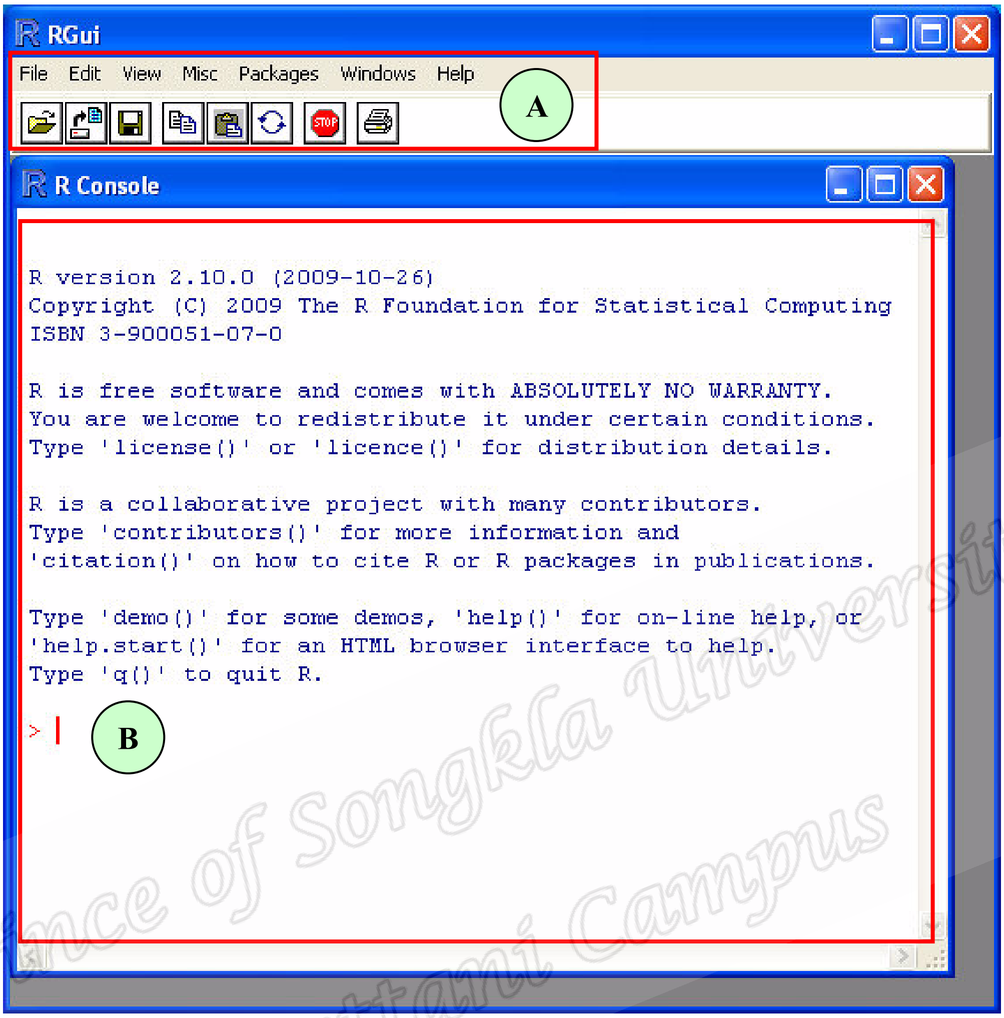

Figure 1.3: Show RGui

In R, there are many packages to display maps and analyze spatial data but most of these packages use ESRI shapefiles to create maps and analyze spatial data. ESRI shapefiles have their own specific format, and users cannot modify them. In the analysis of spatial data, users may want to manage boundaries of regions to present their information, for example combining regions with the same geographic characteristic, such as mountains, lakes, rural areas and urban area, or population characteristics, such as religion and language.

In this study, we aim to create and manage spatial data by using our own functions written in R that read in text files or other files, which can read with functions in R Base.

1.2 Objective

To create functions for handling spatial data of simple regions and complex regions.

1.3 Expected advantages

1. It can facilitate the analysis of spatial data, in program R.
2. It can provide an alternative for teachers, students and researchers.

1.4 Literature review

In R, there are many packages to display maps and analyze spatial data that include:

AIGIS package, created by Bryant and Westerling (2009). The full name is Areal Interpolation for GIS data. There are 23 functions in this package. These functions are oriented toward convenient interpolation of specific US census data for California. *AIGIS* was originally built for specific research on estimating losses associated with wildfires in California.

Automap package, created by Hiemstra (2009) to perform an automatic interpolation by automatically estimating the variogram. There are 7 functions in this package.

GRASS package, developed by Bivand (2010). There are 17 functions, used for interface between GRASS 5.0 geographical information system and R, based on starting R from within the GRASS environment, using values of environment variables set in the GISRC file.

Maps and *mapdata* packages were developed by Brownrigg (2010). The Original S code was developed by Richard A. Becker and Allan R. Wilks. The *maps* package is used for display of maps in projection code. There are 25 functions. And the *mapdata* is a supplement to *maps* package, providing for use of larger databases. This package has 5 functions.

Mapproj package was developed by Brownrigg and Minka (2009). There are 2 funtions in this package. It is used for converting latitude and longitude into projected coordinates. *Themapproj* package also supplements the *maps* package.

Maptools package was created by Lewin-Koh and Bivand (2010). It is a tool for reading and handling spatial objects, in particular ESRI shapefiles. There are 40 functions are contained in this package.

RArcInfo package was created by Gómez-Rubio (2009). There are 19 functions in this package, used for importing geographical information in Arc/Info V 7.x format and E00 files into R variables.

Shapefiles package was created by Stabler (2009) to read and write ESRI shapefiles. There are 15 functions. The ESRI shapefiles are composed of three files. The first file is *.shp* file, which contains the geography of each shape. The second file is *.shx* file which is an index file which contains record offsets. The third file is *.dbf* file, which contains feature attributes with one record per feature.

In summary, there are many types of software that support the analysis and display of spatial data. Many of them are commercially available but R is free software and an open-source project. In R there are many packages to analyze spatial data but most of these packages use ESRI shapefiles to create maps and analyze spatial data, and these data have their own specific format. In this study we use program R to create functions for handling spatial data that read in text files or other files, which can be read with functions in R Base.

For next chapter, we will describe the methodology of this study including format of input file, theory related and algorithm of functions.