UNIVERSITY OF ALBERTA

SPATIAL DATABASE INTEGRATION, A GIS APPROACH- A CASE STUDY FROM COLOMBO, SRI LANKA

BY

ANKUMBURE DEWAYALAGE GUNADASA

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
of DOCTOR OF PHILOSOPHY

DEPARTMENT OF EARTH AND ATMOSPHERIC SCIENCES

EDMONTON, ALBERTA

SPRING, 1999

0-612-39534-0

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: ANKUMBURE DEWAYALAGE GUNADASA

TITLE OF THESIS: SPATIAL DATABASE INTEGRATION, A GIS APPROACH - A CASE STUDY FROM COLOMBO, SRI LANKA

DEGREE: DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED: 1999

Ankumbure D. Gunadasa
Department of Geography
University of Ruhuna
Matara
Sri Lanka

December 7, 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Spatial Database Integration, A GIS Approach - A Case Study from Colombo, Sri Lanka** submitted by Ankumbure Dewayalage Gunadasa in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

M. John Hodgson

Nigel Waters

E.L. Jackson

Mark Rosenberg

Harvey Krahn

Date: Dec 4 1998

DEDICATION

To our two children, my wife, and deceased parents

ABSTRACT

Social area analysis is an analytical technique developed in the 1940s and 1950s for analyzing citywide differentiation (Yeh et al. 1995) and is identified principally with the work of Shevky and Bell (White, M., 1987). With the development of the statistical technique of factor analysis, social area analysis was replaced by "factorial ecology," whereby, rather than create indexes of social variation, the several variables could be subjected to factor analysis procedures. These multivariate analytical methods provided urban social geographers with a tool to measure social dimensionality and associated spatial patterning of cities. A popular multivariate technique of this type has been factorial analysis.

Factorial studies have been conducted on both developed and developing world cities with various limitations. One major limitation is that many of the available studies are limited to only a single source of information even though the importance of integrating information from different sources for the study of social structure of cities has been highlighted by many researchers (Murdie, 1968; Davies and Murdie, 1991; Morris and Pyle, 1972; Lo, 1986; Kasarda and Berry, 1977). Furthermore, several of the available studies are based on data obtained from only one type of spatial unit, namely census tracts (or polygons). Moreover, many available studies, especially in the developing world, were not established with an appropriate set of variables. For example, while the poor have been known to play a dominant role in shaping the social fabric of the cities of the developing world, their basic characteristics have not been adequately defined or measured. Census data do not provide detailed information on the poor and, therefore, alternatives sources of data or methods of study should be considered.

The present study attempts to fill the gap of our knowledge by demonstrating how Geographic Information Systems (GIS) tools can be used to integrate information provided not only by different sizes of one spatial unit, which has been a common problem that most researchers have highlighted, but also by various sizes of different spatial units, i.e., polygons, lines and points.

Colombo, the capital of Sri Lanka, was chosen as the case study as it possesses all the basic characteristics common to cities of the developing world. The researcher was confronted with numerous problems other than the critical problem mentioned above. Major problems included disorganized data, maps with numerous errors, information not in digital form, and census information that failed to analyze key concerns.

In this thesis the researcher applied simple raster GIS tools to address the critical issues related to the data and maps. Data given by different spatial units and sizes were converted to a compatible format by developing innovative methods. Finally, the researcher explored the social structure of Colombo, Sri Lanka, using data integrated from different sources. Quantitative methods such as factor and cluster analyses, database tools, and GIS techniques were used to reveal the city's social structure and the resulting spatial patterns. These methods can provide geographers, sociologists, and planners alike with tools to produce precise measurements of the differential character of areas within a city.

AKNOWLEDGEMENTS

This thesis is a product of the constant encouragement, patience and understanding that the researcher has always received from his supervisor, Dr. John Hodgson. This thesis would not be possible without his invaluable participation; therefore, I trust that it will serve as a testament to Dr. Hodgson's constant support.

Further, I am very grateful to Dr. E.L. Jackson and Dr. Nigel Waters, my committee members, for providing assistance in the statistical analysis component of the present work and with the organization of the thesis itself. Their comments helped the researcher perform the statistical analyses in meaningful ways and present them in an appropriate manner.

I would also like to express my thanks to the Canadian Commonwealth Organization for providing financial support that enabled me to study in Canada. I thank both Dr. Kanthi Vitharana of the Department of Geography, University of Colombo, Sri Lanka for coordinating and supervising secondary data collection and the two research assistants who helped in the data collection. I am especially thankful to the University of Ruhuna, Sri Lanka for approving my leave to complete my studies.

I wish to extend my sincere appreciation to Helen Lambert, Dr. J. Mensah and Dr. P. Kandanarachchi for proof reading the thesis. Finally, I thank my wife Priyanthi and our two children, for their patience, understanding and support during this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

Introduction

Cities have always been of interest to geographers and from the earliest times, studies of regional geographies have addressed their impact. But in spite of attempts to emphasize the advantages of location, during the early days, the geographical study of towns was essentially descriptive. Concerns with regular patterning of urban areas and the development of social structural studies have a long history. The work of social commentators such as Engels (1845) or Mayhew (1862) can be considered as representing the beginning of interest in the subject in the modern period. Since the introduction of urban ecological models (Burgess, 1925; Hoyt, 1939: Harris and Ullman, 1945), academic applications of these models in North America and other developed world cities have become quite extensive and they have brought a new light to the understanding of the urban mosaic.

The work of Shevky and Bell (1949, 1955) on social variation of cities can be highlighted as a breakthrough in urban social studies. They first initiated social area analysis in their study of Los Angeles and San Francisco with the use of three basic constructs: social rank, urbanization and segregation. These three constructs summarized the important social differences between census tracts and patterns of residential differentiation. Despite criticisms of the lack of strong theoretical justification for using the three constructs to differentiate social areas, this approach laid the foundation for the analysis of urban social spatial structure.

With the development of digital computers in the 1960s, which could handle large quantities of census data, the factorial ecological approach, using principal component analysis and factor analysis on a large number of variables, was commonly used to examine the components and spatial patterns of residential differentiation. A number of studies have been carried out in North America and other parts of the developed and developing world using this approach. The early studies in North America (Murdie, 1969; Davies and Barrow, 1973; and Rees, 1979) showed that the three components identified, namely socioeconomic status, family status, and ethnic status have exhibited significant regularities in a number of cities. These regularities were also identified in the cities of Australia (Jones, 1969), New Zealand (Johnston, 1973), and other Western cities. Social space in Western cities is dominated by a socioeconomic status component, with a second component characterized by family status/life-cycle characteristics, and a third component relating to segregation along ethnic lines. Socio-economic status shows a dominantly sectoral pattern with family status distributed spatially by concentric zones. The final dimension, ethnic status, finds certain ethnic groups relegated to separate sectors within the city. Interestingly, British cities were found to deviate from the American model due to modifications brought about by public housing sectors (Herbert, 1968; Evans, 1973).

The focus of the early factorial studies of cities was to examine Shevky and Bell's three major axes. Therefore, these studies were based on a limited set of variables that would generate biased structures. Application of multivariate techniques widened the scope and the understanding of social variation of cities. Following the birth of factorial studies, a series of studies was generated for the mathematical refinement of research techniques (Cattel, 1968; Harman, 1976; Harman and Jones, 1966; Harris and Kaiser, 1964; Levine, 1977; and Rao, 1955). In later works, additional axes other than Shevky and Bell's three were postulated (Davies, 1983; Lo, 1986; White, 1987; Hamm et al. 1988; Davies and Murdie, 1991; Yeh et al., 1995). Furthermore, the changing dynamic of the social geography of cities has been captured by recent studies with additional dimensions.

The objectives of the initial factorial studies of the developing world (Abu-Lughod, 1966; Brand, 1972; McElrath, 1968; Morris and Pyle, 1972) were similar to early factorial studies that were conducted in the Western countries. Most of these studies of emerging nations were unable to parallel Shevky and Bell's three dimensions as they were found in the developed world. Later studies conducted in the developing world cities have revealed the city structure in detail with new componenets (Berry and Rees, 1969; Abu-Lughod, 1980; Lo, 1986; and Yeh et al., 1995) not based on previous assumptions.

The latest trend in social area analysis is the focus in investigating consistency of social dimensions among the disparate cities and the changing, evolutionary patterns of the social structures of a city (Davies and Murdie, 1991; Hamm, et al., 1988; Lo, 1986; Le and Beaudry, 1988; Murdie, 1987; and White, 1987) rather than the exploration of the social dimensionality of a single city.

Rationale for the Study

The underlying multivariate studies show that they have captured the social dimensionality of cities in both the developed and the developing world. Despite these technical advances, one problem related to the data capturing still remains to be solved. This issue is known as "generating integrated spatial databases" by which data given at different scales are brought into a common format to reveal the social complexity of cities.

Integrated spatial databases provide the researcher with vital information that cannot be obtained from a single data source. When adequate and appropriate information is not employed in factorial studies, the reality of the city is not being revealed as fully as would be desired. Valuable information coming from other data sources should be integrated in order to capture many possible sources of social variation in a city. The most popular information source that researchers have used to reveal the urban social structure of cities has been census records where in information was collected by a standard set of zones (polygons). Information coming from sources other than the census such as police, health, education, municipalities, and other institutions is also needed for appropriate studies. The need and the importance of integrating information on social structural studies have been echoed by many researchers in the field, for example, Davies and Murdie, 1991; Murdie, 1969; Morris and Pyle, 1972; Lo, 1986; Kasarda and Berry, 1977.

Furthermore, temporal patterns of the social structure of cities have been conducted under a number of limitations due in part to changes in the census tract boundaries. It seems these studies halted at the end of the 1980s because valuable data were being collected by different zoning systems or were subject to changing boundaries of census zones, and appropriate methods to deal with these problems did not exist. As Murdie (1969) highlights in his study, changing boundaries of census tracts was a serious problem for his study. In his words:

> Changes in the boundaries of tracts between censuses are a particularly severe problem for comparative analyses. Subdivisions

> of existing tracts which are made necessary by population increases often result in boundary changes. Murdie, 1969, p.65

Davies and Murdie observed that

> ...ideally, more indicators would have been desirable, but census sources could not provide appropriate additional variables, and the requirement for a complete inventory of information for all CMAs precluded the possibility of using additional indicators. Davies and Murdie, 1991, p.60

Although this issue was raised in the early 1970s, previous researchers were unable to integrate information coming from other sources since the information was not given by compatible spatial units and the techniques available could not solve this problem. The assumption is that researchers have diverted their attention to other areas of social area analysis, i.e., mathematical refinement of techniques and comparative studies. This has created a gap in the knowledge base, as the available techniques do not allow researchers to perform an in-depth study of a city, with integration of information. The present study attempts to fill this knowledge gap by generating integrated spatial databases and performing multivariate analyses on them.

Thesis Objectives

Two main specific objectives of the study are:

1. To demonstrate the potential of GIS applications in a spatial database integration to overcome the limitations of previous multivariate analytical studies of the social structure of cities.
2. To explore the social structure of Colombo, capital of Sri Lanka, based on an integrated spatial database.

To accomplish these goals, the study applies GIS tools and statistical techniques on the data and maps of Colombo. Chapter Two provides background for the present study. It first discusses the major factorial studies available for both developing and developed world cities. Major findings of both groups are described with emphasis on the limitations of their studies. Secondly, Chapter Two gives a brief description of Colombo, the study area. The current characteristics of the city are covered under selected themes in order to make the reader familiar with the study area and to serve as an introduction to the expected sources of social variation in Colombo. The major areas covered are: present land use pattern, urban growth of the city, ethnic composition, housing structure, educational structure, and public health facilities. Finally, data collection and storage systems used in Sri Lanka will be described to highlight the nature of the systems used and to show the need for organized information systems to gain a better understanding of the city.

Chapter Three describes how the researcher prepared map sheets and how they were converted to appropriate digital format, which was an integral part of the research. Chapter Four details how the researcher used GIS tools to integrate data from different sources of information. The objective of this chapter is to show how one can apply simple GIS tools in Raster GIS to solve complicated problems related to integrating spatial databases. Generation of an acceptable city boundary for Colombo, the creation of data layers on selected variables for the study of urban social structures, and the integration of all the data layers into a common format will be described in detail.

Chapters Five and Six are the main statistical analysis components of the study. In Chapter Five, factor analysis is performed on the data set that was generated applying GIS techniques to extract major sources of variations. Different social groups are identified in the following chapter. The dimensions identified in Chapter Five are used as synthesized variables to extract distinct social groups. The final chapter presents the main conclusions drawn from the study, as well as its limitations, and proposes some directions for future research.

CHAPTER TWO

BACKGROUND TO THE STUDY

Introduction

A city is not a randomly generated structure but is recognized as a product of various processes that result in distinct patterns. Researchers have attempted to explain and describe the multivariate reality of urban society using different techniques and approaches. The purposes of this chapter are to review briefly the development of urban social structural studies in the developed and the less developed world, and to give a general overview of the case study area, Colombo, Sri Lanka. Finally, data collection and storage systems used by government and non-governmental organizations in Sri Lanka are described briefly in order to make the reader familiar with data sets applied in the present study.

The multi-dimensional nature of cities makes it difficult, if not impossible, to conceive of a completely unified and systematic analytical approach to the study of urban social character. Despite a plethora of literature on the character of urban places, most is written by specialists working in a wide variety of different disciplines such as urban geography, urban history, urban sociology, urban economics and urban planning. The information generated by different disciplines is difficult to integrate (Davies, 1983). The knowledge has been filtered through quite different sets of disciplinary perspectives and a variety of temporal and spatial perspectives. The stress that these disciplines place upon any given part of the urban character to the exclusion of other parts means it is difficult to obtain a balanced and comprehensive view. In addition, the range of possible sources of differentiation within cities is so immense and relates to such different phenomena, that it is difficult to see how they can be integrated into one set of generalizations.

The concern with the regular patterning of urban areas and the development of social structural studies have a long history. Engels (1845) or Mayhew (1862) can be considered as representing the beginning of interest in the subject in the modern period. But those studies were primarily designed to draw attention to the problems of urban poverty; they were hardly systematic inquiries in the sense of developing concepts or generalizations appropriate to the field. The development of a systematic interest in the field of urban social differentiation can be associated with the work of Booth (1903) in late nineteenth century London, and Park, Burgess, Wirth and others of the Chicago School of Urban Ecology after World War I. Booth produced a classic series of portraits of working class life and labor in late nineteenth century London, using several census variables to produce an aggregate score of what he called the 'social condition' for each area in London.

Social Area Analysis

The modern phase of systematic analysis of urban social structure can be said to have begun in 1955, when two influential studies were published, namely, Shevky and Bell's 'Social Area Analysis', an extension of previous work carried out by the senior author in Los Angeles in 1949 (Shevky and William, 1949) and Tryon's (1955) 'The Identification of Social Area by Cluster Analysis'. Although they dealt with the problems of defining social areas in cities in different ways, they were alike in seeking a quantitative measurement of the social character of urban areas using census based-variables. These two studies made a great impact by focusing attention upon one part of the overall field of urban ecology: that of understanding the recurring patterns of urban social differentiation rather than behavioral or land use concerns of urban ecology in general.

Shevky and Bell's greatest contribution was that their work linked the study of social variations in urban areas with a theory of social change. From their study, they identified three axes of variation in a city, originally called social rank, urbanization and segregation, and measured each city area on these axes of variation. In so doing, Shevky and Bell (1955) provided a clear demonstration of the need to separate the study of social dimensionality -- sources of axes of differentiation in cities -- from the spatial patterns of these dimensions as expressed in the relative scores of each area on the axes.

The limitations of social area analysis are well known (Timms, 1971). The utility of this particular theory of social change can and has been questioned. A city is a product of numerous forces but Shevky and Bell used very few variables to describe the multi-dimensional nature of a city. Alternative deductions of the constructs and the indicators may also be made. Davies (1983) argues that the ratio measures used to define the axes represented relatively little advance on the ideas of Booth since there was no guarantee that the individual indicators that were added together to form the index of social variation co-varied. Although the particular procedures can be criticized, the Social Area approach deserves an important place in the history of urban social area differentiation.

McElarth (1968) revised the original model presented by Shevky and Bell (1955) to produce a four axis structure in which migration was added to the list of the original constructs. Timms (1971) used these four constructs to produce a family of social area types based on the degree of separation of these axes.

Tryon (1955) took a different approach. His concern was primarily with the classification of areas. In his study of San Francisco, he used a set of thirty-three variables to define the social variation among 243 census tracts. A similarity index was calculated between each pair of areas on the basis of their similarities over the set of variables, and cluster analysis was used to assemble these areas into a limited number of groups. Although the clustering procedures that Tryon applied are not advanced, his work placed the study of urban social differentiation firmly within the developing literature on multivariate analysis. However, comparatively few workers followed his path in the 1960s and early 1970s. The precise clustering techniques used in those studies are not described in detail.

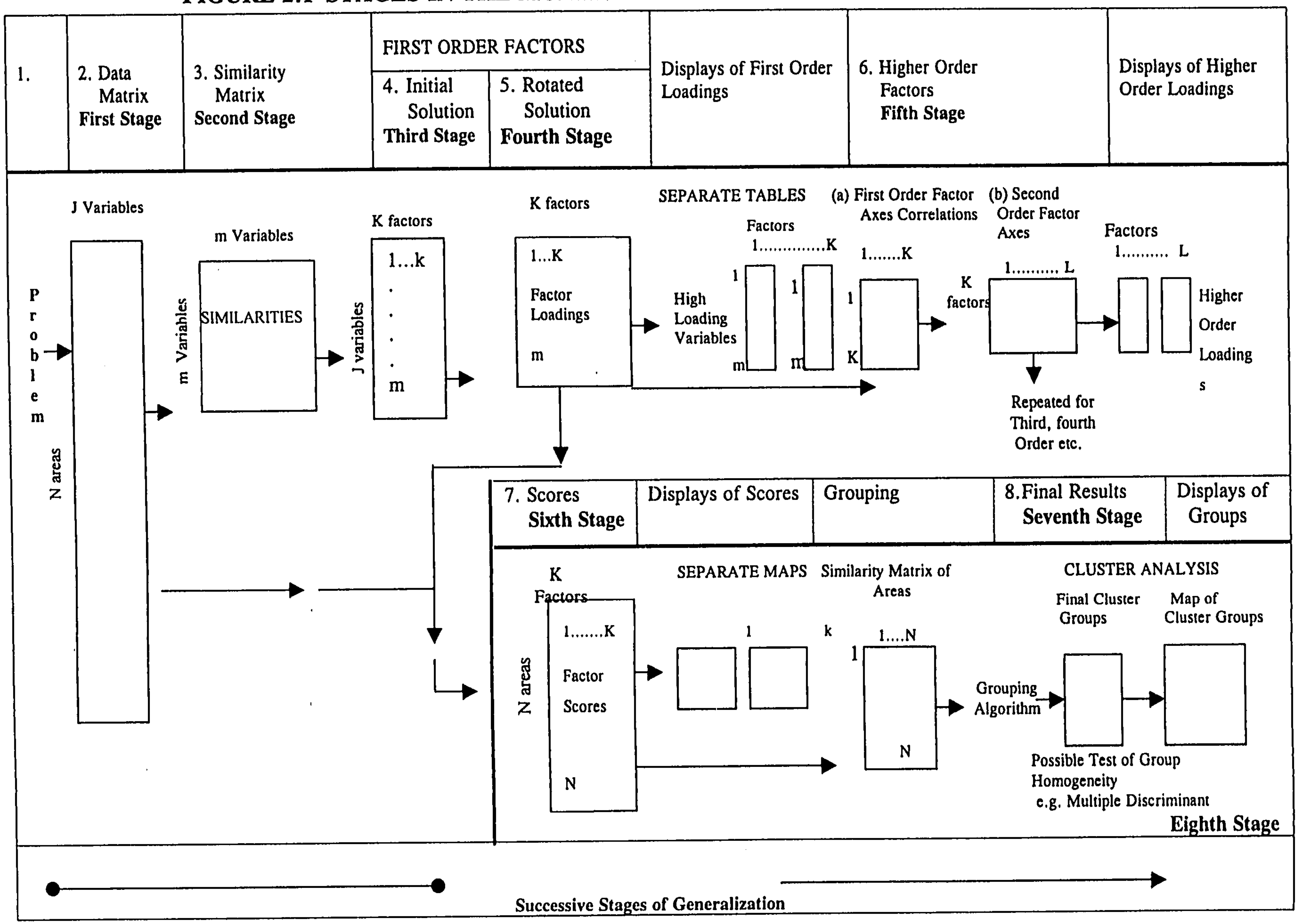
Multivariate Approach

In the late 1960s, a new methodological approach to the study of urban social differentiation was defined. The new approach was called 'Factorial Ecology' because it applied factor analysis to urban ecology. Figure 2.1 highlights major stages of this factorial approach (also known as multivariate methodology).

Multivariate Studies in Developed World Cities

Much of the early factorial work was really based on the Social Area method, since many researchers attempted to establish the Shevky-Bell model of urban social differentiation, with its three constructs. Bell (1963), using the same data set for San Francisco as in his earlier Social Area Analysis with Shevky (1955), maintained that the factorial results showed that the three axis model of Economic Status, Family Status, and Ethnic Status was appropriate, in the sense that the three axes were discrete factors of

FIGURE 2.1 STAGES IN THE MULTIVARIATE ANALYSIS



(Adapted from Davies, 1983)

variation. In later work, additional axes other than Shevky and Bell's three axes have been found. This may be partly because most of the early tests on Shevky and Bell's models used very small data sets with a small number of variables. Nevertheless the use of higher order factor methods on large data sets demonstrated that some of these more specialized axes collapse into the traditional Shevky-Bell axes at higher levels of generalization (Davies and Lewis, 1973).

A summary of the major findings of multivariate studies in developed world cities during the last three decades (1960 – 1990) is shown in the Table 2.1. In most studies, three major patterns of socio-economic differentiation were identified and where a full range of census data was employed or where a smaller set of data was carefully selected from a large set of variables, these reflect the social area indexes of economic status, family status, and ethnic status. The family status pattern appears to be more stable than the others. Composition of the remaining factors varies with the type of city being analyzed and the characteristics selected for study. The results of a factor analysis, of course, depend entirely on the nature of the variables which enter the analysis.

Very few multivariate studies have specifically investigated the spatial differentiation of urban social structure. From the limited number of studies that are available, there is strong evidence to support the concentric distribution of family status and the sectoral distribution of economic status.

The application of quantitative methods to the social sciences and the development of larger and faster computers with the associated growth of sophisticated statistical software has led to an increase in the use of factor analysis to the study of urban social structures. These developments have broken the previous barriers to the use of such sophisticated multivariate procedures, namely, the problem of recreating all the necessary statistical routines, and the need to spend enormous amounts of time calculating the results by hand or with the aid of simple calculators. Murdie's study (1969) on "Factorial Ecology of Metropolitan Toronto – 1951 – 1961" is a major advance in the use of factor analysis (see Table 2.1).

Table 2.1 Factorial/Multivariate Studies in the Developed World

Bell San Francisco and Los Angeles 1963 6 Variables 3 Factors	Schmid and Tagashira Seattle 1964 42 Variables 5 Factors	Sweetser Boston 1965 20 Variables 5 Factors	Sweetser Helsinki and Boston 1965 42 Variables 3 Factors	Berry and Tennant North-eastern Illinois 1965 50 Variables 7 Factors	Carey Manhattan New York City 1966 33 Variables 5 Factors	Pederson Copenhagen Sweden 1967 14 Variables 3 Factors	Murdie Toronto 1969 46 Variables 9 factors	Hunter & Latif Winnipeg 1972 15 Variables 3 Factors
1.Economic Status 2.Family Status 3.Ethnicity	1.Economic Status 2.Family Status 3.Ethnic Status 4.Maleness 5.Population Stability	1.Economic Status 2.Feminine Carcerism 3.Residentialism 4.Established Familism 5.Postgeniture	1.Economic Status 2.Progeniture * 3.Urbanism	1.Economic Status 2.Family Status 3.Ethnic Status 4.New Suburbs 5.Distance 6.Density 7.Housing Vacancies	1.General Manhattan Population Residence 2.Puerto Rican Sub-Population 3.Middle Income Negro 4.Low Density Transient Residence 5.West side Rooming House	1.Family Status 2.Socio-economic Status 3.Population Growth and Mobility	1.Economic Status 2.Family Status 3.Ethnicity 4.Italian 5.Jewish 6.Recent Growth 7.Service employment Clerical 8.Household Characteris--tics 9.Household & Employment Characteristics	1.Economic Status 2.Family Status 3.Ethnicity

* Progeniture is asociated with fertility and new housing while urbanism is closely related to women in labor force.

Figure 2.1 Contd.

Figure 2.1 Contd.

Johnston Christchurch (New Zealand) 1973 14 Variables 3 Factors	Davies Calgary 1975 60 Variables 6 Factors	Newton and Johnston Christchurch (New Zealand) 1976 14 Variables 3 Factors	Herbert Cardiff 1977 24 Variables 8 Factors	Davies Cardiff 1983 26 Variables 11 Factors	White 21 American SMSAS 1987 14 Variables 3-4 factors	Bourdais and Beaudry 1988 Montreal 59 Variables 6 Factors	Currie and Forde 1988 Winnipeg 56 Variable 10 Factors	Davies and Murdie 24 Metropolitan Areas in Canada 1991 35 variables 9 Factors
1.Socio-economic Status 2.Life-style / Age Structure 3.Housing Type	1.Family Status and Age 2.Socio-economic 3.Migrant Status 4.Economic Participation (non-affluent) 5. Ethnicity (Canadian Immigrant) 6.Family Size East European	1.Socio-economic Status 2.Life-style / Age Structure 3.Housing Type	1.Decayed/Old/ Age/ Foreign 2.Middle City/ Better Residence 3.Mixed/Private Council 4.New Status Suburbs 5.Rental 6.Shared 7.Established Council Estates- Low Ethnicity 8.Highest Status-Old	1.Socio-economic Status 2.Substan-dardness (Non Affluence) 3.Urban Fringe 4.Life Cycle & Family 5.Young Adult	Prominent Factors 1.Socioecono-mic 2.Life Cycle 3.Ethnicity 4.Middle Class Suburbs 5.Black	1. Family 2.Socio-economic 3.Ethnic 4.Socio-economic 5.Family 6.Ethnic	1.High S.E.S* 2.Middle S.E.S* 3.Ethnic-Religious 4.Low Income Elderly 6.Older Housing 7.New Immigrants 8.Native 9.German 10. French 11. Jewish	1.Economic Status 2.Impoverish-ment 3.Ethnicity 4.Early and Late Family 5.Family/Age 6.Pre-Family 7.Non-Family 8.Housing 9.Migrant Status

* Socioeconomic Status

Davies and Lewis's study of Leicester (1973) confirms the validity of Murdie's (1969) case study of Toronto. Although they found eight first order social dimensions collapsed into the traditional Shevky-Bell axes at higher levels of generalization.

Davies' study (1975) "A multivariate description of Calgary's community areas," is another milestone in factorial analysis studies. Comparatively few factorial ecologists had gone beyond the description of the factor loadings and factor score matrices. Davies (1975) applied a cluster analysis to the factor scores and produced a more parsimonious description than the original data set. Davies identified six major factors accounting for 67.9 per cent of the initial variance (Table 2.1).

The study confirmed the familiar Family Status dimension recognized in many factorial studies. The distribution of scores in this dimension showed a concentric pattern as described by Burgess (1925). Spatially, the pattern of scores on the Socio-Economic dimension shows a completely different pattern to the family status. This axis splits Calgary into an east and west zone. The Migrant Status factor that emerged from the study is a vital part of the 'complexity of society' as postulated by Shevky and Bell (1955). Comparatively few factorial studies of urban character have identified this dimension. Migration status is linked to both the growing edge of the city and to the areas of transformation around the commercial center (Davies 1975). Ethnicity status displays a more clustered pattern. Davies' results of high order factor solution supports McElarth's (1968) four axis model of the bases of differentiation of urban areas rather than the more widely used Shevky-Bell constructs.

Johnston's (1973) factorial study of Christchurch and other New Zealand urban areas has shown that the inductive 'theory' developed to account for residential patterns in North American cities is relevant to New Zealand. The only difference relates to ethnic groups. In North America, especially in the USA, blacks usually form a separate residential dimension in factorial ecologies, but in New Zealand, Maoris tend to be associated with the more general socioeconomic status dimension.

Bourne and Murdie's (1972) study entitled "Multivariate Analysis of Metropolitan Toronto" attempts to fill a gap between studies on spatial dimensions of social attributes and physical attributes. In their study, they attempt to assess the aggregate ecological relationships between the social and physical attributes of the city and to use those results in outlining a typology of urban neighborhood types. The complete analysis is built on their previous studies (1969 and 1970) of Metropolitan Toronto, one dealing with social characteristics and the other land use.

Sixty-nine variables were selected to represent the social matrix compared to only 29 measures of land use (physical space). First, two independent factors for both data sets were extracted for 62 zones of approximately equal population (Table 2.2). These factors provided realistic separate descriptions of the complex structure of social space and of land use. As shown in the Table 2.1, Bourne and Murdie (1972) found seven social space factors and ten physical space factors and then merged them into a composite analysis of urban spatial and ecological structure.

The study clearly showed that areas of relative economic and social deprivation are associated closely with the core area and to a lesser extent with zones of railroad and industrial and institutional use. Also, areas of high economic status were associated with suburbanization, apartments, and private schools or locations within the middle and outer rings of the metropolitan area which tended to be environmentally attractive. In the next step, factor scores from the socio-economic analysis were combined with those from the land use analysis and subjected to a second factor analysis. The new analysis derived eight factors which revealed stability in the new dimensions derived.

Davies and Barrow (1973) attempted to find first and second order dimensions in a comparative ecological framework in a case study of Winnipeg, Regina, and Edmonton. This was the first comparative ecological study of three cities. The selected cities were incorporated within the framework of a single analysis rather than being analyzed separately and then linked together by verbal description. The results confirm the established generalization of the social structure of western cities.

Davies's (1983) multivariate analysis of Cardiff and its region study goes beyond the general frame of factorial studies. It tests the changing pattern of social structure at different geographical scales and then applies cluster analysis to extract similar social areas at the selected scales. Not only are the changing patterns of social structure at selected areal scales tested but the changing patterns of social structures under different factorial methods are tested as well.

Three axes, namely, Socio-Economic Status, Life Cycle, and Young Adult, were very similar in all scales. Moreover, some rather typical results were obtained between the two levels of aggregation, for the ward level axes were more general and accounted for more variance than the detailed enumeration district results. At the intra-urban scale, the results of the study show the inadequacy of Shevky and Bell's theory (1955), McElarth's revisions (1968), and Timms' ideas (1971) since far more axes of differentiation were identified than were proposed in these models (Table 2.1). Davies (1983) proposed that any new theory of social differentiation at the intra-urban scale should take account of the Late Family, Young Adult and Rural Fringe axes and probably a sub-standardness dimension, in addition to the traditional Social Status, Age or Family, Ethnicity and Migrant axes. These results demonstrated that these Welsh towns have rather different dimensional structures from those of many other cities in the Western world, and therefore, the researcher suggests that these results would justify a modification of the established theory.

Relatively few factorial ecologies have explored the consistency of the social dimensionality of urban areas for more than a few cities. Davies and Murdie (1991) explored the consistency of the social dimensionality in 24 Canadian metropolitan areas. The study identifies a persistent similarity in six of the seven to nine dimensions found in separate analyses of three city size categories. The combined study of all the centers shows that 85% of the variability can be summarized by nine dimensions (Table 2.1). The evidence for several different family-related axes illustrates the increasing complexity of the social dimensionality of modern cities based on family differentiation.

The authors acknowledge that their study is only a partial view of the social complexity of intra-urban space because census sources could not provide appropriate additional variables. In addition to the limitations posed by a single source of information, some of the census tracts are excluded from the study due to incompatibility and lack of proper information. For example, fifty-one tracts had to be excluded because of low population counts and/or incomplete data in one or more of the variables.

There are some common limitations and problems related to developed world factorial studies. Some of these limitations will be discussed in the following chapter under variable selection. One common problem is that the basic unit of analysis is limited to census tracts or, in some instances, wards. Information available at other areal scales should be explored and tested against the results of the census information. Not only is the basic unit of analysis limited to the above areal units in all but three instances (Johnston, 1973; Davies, 1983; Davies and Murdie, 1991), but also the information used is usually from a single source. The value of using integrated information in social structural studies has been highlighted in many studies, i.e., Davies; 1983, Davies and Murdie, 1991; Johnston, 1973; Murdie, 1969. As Davies (1983) pointed out, non-census data and indicators may be more directly related to 'quality of life' and thus they should be integrated in order to obtain better results, especially for planning policies. As described above, Murdie (1969) used coarse methods to compare 1951 census data with 1961 census data. For example, 1961 census tract boundaries were changed and some of the census tracts of the 1961 census were excluded in order to obtain two compatible data sets. Changing patterns of socio-spatial structure over a given period cannot be compared accurately using the present approaches. Murdie (1969) discussed up this problem in his study and stated that changes in the boundaries of tracts between censuses are a particularly serious problem for comparative analysis.

Furthermore, on many occasions, information on physical space is not stored by census tracts. As Bourne and Davies (1972) showed in their study, information on the physical space of a city is a critical factor to the understanding of the social ecology of a city. Available methods do not integrate information satisfactorily when it is given at different scales.

Multivariate Structural Studies in the Developing World

Table 2.2 summarizes major studies related to the developing world. The most often referred to study among the first factorial studies in the developing world cities is Abu- Lughod's (1969) analysis of the social geography of Cairo. Her major purpose was to find factors similar to the Social Area Dimensions. To study the structure of Cairo, principal axes factor analysis was used to extract orthogonal factors from replicated 13 X

Table 2.2 Factorial / Multivariate Studies in the Developing World

Abu-Lughod Cairo (Egypt) 1966 13 Variables 3 Factors	McElarth Accra (Ghana) 1968 14 variables 4 Factors	Lo Hong Kong 1971		Brand Accra (Ghana) 1972 13 Variables 4 Component	Fred and Gerald Rio De Janeiro 1972 22 Variables 3 Factors	Berry and Rees Calcutta 1977 37 Variables 10 Factors
1.Life Style 2.Male Dominance 3.Social Disorganization	1.Economic Status 2.Family Status 3.Ethinicity 4.Migration Status	1961 28 Variables 6 Factors	1971 28 variables 5 Factors	1.Middle class Communities 2.Urban Villagers 3.Distance – Density Decay 4.Employment	1.Socio-economic 2.In-migration 3.Stage in Life Style	1.A Land use and Familism Gradient 2.Higher Status Muslim Concentration 3.Axiality in Literacy 4.Substancial Residential Areas 5.Traditional Commercial Areas 6.Pheripheral Ring 7.Rural-Higher Family Size 8.Entertainament District 9.Medical Center 10.Factory Zone
		1.High-income Expatriate Workers 2.Low-income Workers 3.Old Age Workers 4.Youthful Immigrants 5.Ethnic Contrast 6.Sex Contrast	1.High-income Non-Cantonese 2.Low-income Blue Collar Workers 3.High-income Expatriate Workers 4.Public Housing Residents 5.Sex Contrast			

TABLE 2.2 CONTD.

Table 2.2... contd.

Abu-Lughod Morocco 1980 27 Variables 6 Factors	Lo, 1986		Anthony, Huaying and Xuequiang Guangzhou (China) 1995 67 Variables 5 Components
	Hong Kong 95 Variables 10 Factors	Hong Kong and Territories 95 variables 8 Factors	
1.Socio-economic Status 2.housing Quality 3.Male Dominance 4.Family 1 5.Family 2 6.Migration	1.Elderly Population 2.Blue-collar Workers 3.Low-rent Public Housing Residents 4.Private Housing 5.Nucliar Families 6.Large Household Size 7.High-income Professional 8.High Educated Group 9.Farmers and Fishermen 10.Armed forces	1.Socio-economic Status 2.Recent Chinese Immigrants 3.Elderly Vs Youth 4.Self-employed Farmers or Fishermen 5.Housing Condition 6.White-collar and Service Workers 7.South Asian Group 8.Young Children	1.Population Density 2.Education 3.Employment 4.Housing Quality 5.Household Composition

13 matrices of variables reflecting the socio-economic status, familial and ethnic characteristics of the population in 206 subdivisions of Cairo in 1947 and 1960. In each data year, half of the total matrices' variance was accounted for by a single "life-style" vector on which were loaded not only variables of social rank but of family life as well. The results of the study can be questioned for several reasons: First, the metropolitan boundaries of the city changed over the selected census years and adjustments to assure congruency of census tract boundaries were not explained clearly. Second, the number of census tracts was not increased proportionately to the population change from 1947 to 1960. Cairo had a population of over two million in 1947 (Abu-Lughod, 1966) and this increased to three and a half million by 1960 but despite this, the number of census tracts was not changed. Finally, very few variables were used to measure the sources of variation of urban social structure, and none of the factors derived has higher loadings on more than two variables, with the exception of the first, second, and third factors.

Brand (1972) attempted to describe the social dimension of a postcolonial city, Accra, Ghana. The selected city is a product of cultural contact, innovation, and diffusion. These factors can be referred to as foci of the processes of modernization that are revolutionizing traditional societies. The study was based on thirteen variables using 1960 census data as the array of socioeconomic, demographic, and ethnic data reported at the statistical area scale of aggregation was not available for enumeration areas. The four factors extracted from the factor analysis show the social structure of Accra to consist of Bourgeois Migrant Communities, Urban Villagers (the Subculture of Traditionalism), Distance Decay Parameters and Hard-Core Unemployment. Among all the dimensions, migration status and its spatial manifestations were the most powerful within Accra. Migrant destination nodes were themselves highly differentiated, and there was also differentiation between elite migrant tracts and most central city indigenous tracts. This skewed modernization surface of Accra is typical of most postcolonial cities in transition. The study displays clearly that the pre-industrial and western-industrial generalized models of urban social morphology do not correspond very closely to the situation in Accra. Brand's (1972) attempt to describe the social structure of Accra is based on a crude and small choice of variables. Some of the dimensions extracted from the study had only one or two higher loadings. The dimensions were named and described in detail although they did not have adequate number of variables with higher factor loadings.

Inspired by Murdie's (1969) comparative factorial study of Toronto, Lo (1975) did a comparative factorial study on the city-colony of Hong Kong using 1961 and 1971census data and applying various factorial methods. Interpretations revealed two robust factors of 'high-income expatriate workers' and 'low-income blue- and white-collar workers' in 1961. By 1971, a more diversified socioeconomic stratification of the Chinese and high-income expatriate workers had emerged. There appeared to be a decline in importance of the expatriate factor and the rise of an economically more varied Chinese social class, which led to the implication that there had been a greater equality in the distribution of income despite the remaining contrast between the rich and the not so rich. The study also revealed that the family status of the population is not well expressed in the factor analysis. These conclusions on the social dimensions and spatial patterns of Hong Kong are based on highly aggregated data, only 27 spatial units; therefore, complex patterns of social structure are not evident.

Morris and Pyle's (1972) study on Rio de Janeiro demonstrates the nature of developing world data sets. Three data sets were available for the metropolitan area, but none covered the entire metropolitan area. Furthermore, they were not compatible with each other. The information for one of the states of metropolitan Rio de Janeiro was given at a completely different scale of measurement than the rest, therefore, the information given for that state could not be used with other states. Disparate data sets available were not integrated due to the nature of their boundaries and independent factorial analyses were conducted for each data set. The results of the four analyses show that Rio displays distinctive patterns of social geography.

Among other things, Morris and Pyle (1972) found that the metropolitan area of Rio de Janeiro displays a homogeneous pattern in the life cycle, with the exception of the old central area. The slums and shanties form a pattern different from the basic one of the city in that they are scattered "on top of" the city itself, imposed but not integrated into any other patterns of the city's social space. Morris and Pyle (1972) suggest that these ingredients clearly indicate a traditional urban ecology under the impact of modernization. These characteristics are similar to the features of pre-industrial cities as outlined by Sjoberg (1960). Rio de Janeiro is a common model of Latin American cities which are a product mainly of Portuguese and Spanish rule. One of the major features of these colonial cities is that the elite lives close to the core (Sjoberg, 1960) in marked contrast to the cities of North America. This study would be very meaningful had the researchers attempted to describe the social geography of the metropolitan area based on integrated data. That way the researchers could have studied the entire area and described the social dimensions and their distributions more appropriately.

The study on Calcutta (Berry and Rees, 1969) is an example from South Asia. The social geographies of South Asian cities have strong similarities except for the caste system, which is stronger in India than in neighboring countries. This study attempted to initiate systematic cross-cultural ecological analysis by means of a structured factorial ecology of Calcutta. Principal axis factor analysis with normal varimax rotation of all factors with eigenvalues exceeding unity was used to analyze an 80-ward X 37-variable data matrix for Calcutta.

The social structure of Calcutta can be synthesized in ten dimensions. The study reveals that Calcutta is characterized by a broadly concentric pattern of familism. The percentage of married women over fifteen years of age is high in the peripheral wards. Moreover, wards around the periphery of the city have the poorest residents of the city. The study also revealed that the land use pattern of Calcutta is a mixture of pre-industrial and industrial ecologies. Based on these mixtures, they concluded that Calcutta is in a transitional development stage. Compared to the city size, spatial units used (wards) in the analysis are extremely large and homogeneity within these units cannot be expected. Furthermore, the researchers tabulated their data manually since it was not properly formatted. This illustrates one of the shortcomings of developing world data sets. Although the data used were given for large areal units, the study; however, clearly reveals that social structure of South Asian cities is remarkably different from that of North American cities.

Results of the social structure of Rabat-Sale', capital of Morocco, done by Abu-Lughod (1980) showed how Muslim-dominant metropolitan social structures are different from other cities. The researcher attempts to compare the social structures of Rabat-Sale' with three basic dimensions of Shevky and Bell's social area analysis. The findings do not echo North American social structures. Ethnicity could not be separated from socio-economic status, which suggests strongly that interpretation of the geographical pattern of Rabat-Sale' is essentially as a caste-apartheid model. Also she finds that literacy rates of male and female were "style of life" measures as much as they were indicators of economic power.

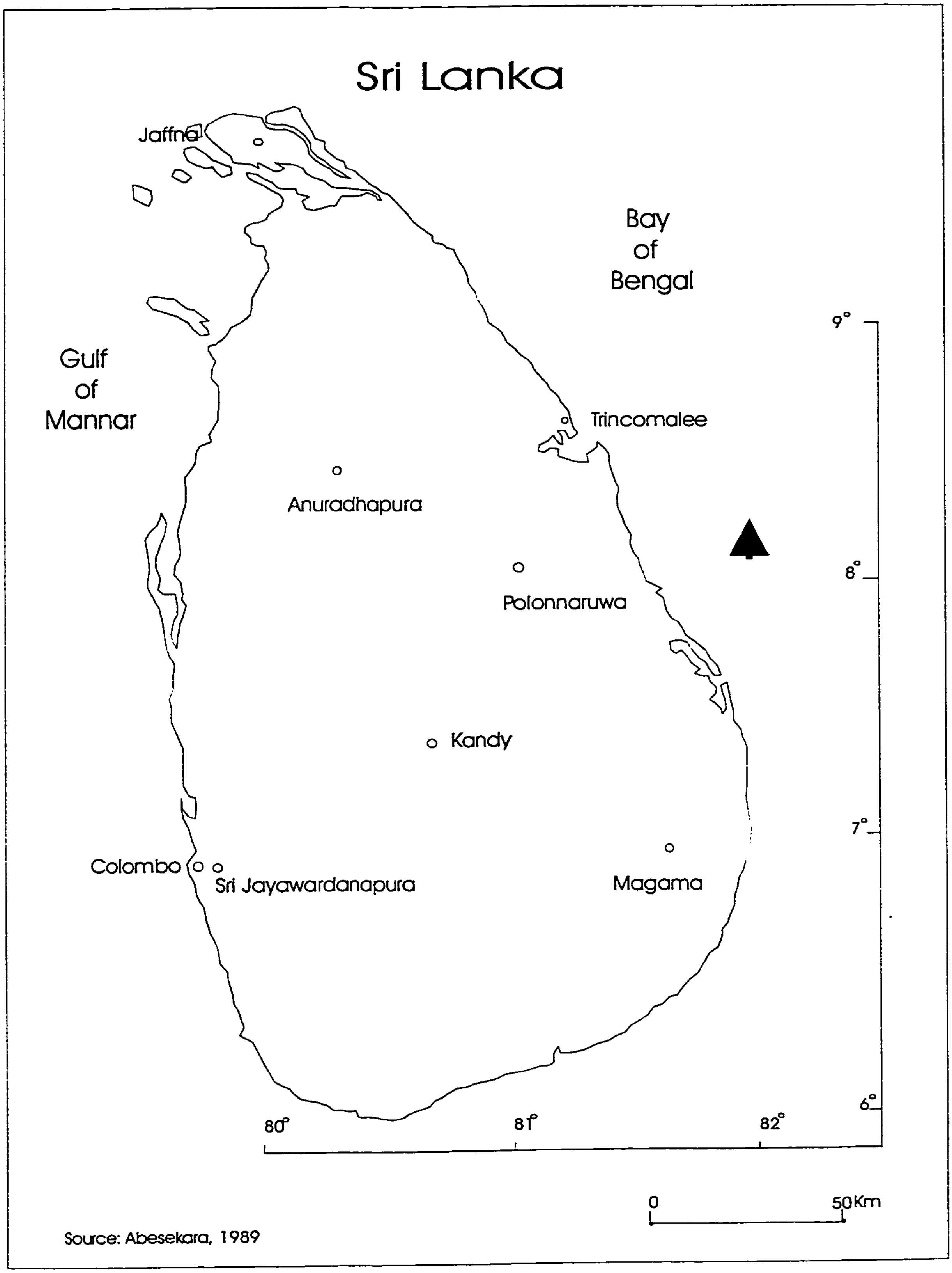
Using principal component analysis and cluster analysis, Yeh et al.(1995) demonstrated the urban social spatial structure of Guangzhou, China, a city in socialist China. The nature of the data used in this study is different from most of the studies available which are based on census data. However, the data for the Guangzhou study comes mainly from the Guangzhou resident trips survey conducted in 1984. The information was allocated to the urban blocks defined by streets. They found that the main components of social space in Guangzhou are population density, educational level, employment, housing quality, and household composition. The cluster analysis revealed five types of social areas. The spatial pattern of social areas of the city has a concentric elliptical shape. The urban social spatial structure of the city reflects the history of urban development, the housing allocation system and socialist urban land use planning. The dominant factors determining the residential location and the social spatial structure of Chinese cities were occupation and location of employment.

THE CASE STUDY AREA, COLOMBO, SRI LANKA

Introduction

The next section of this chapter will briefly describe the case study area, its data collection, and the storing and maintenance systems practiced. The Democratic Socialist Republic of Sri Lanka is an island located in the Indian Ocean (Figure 2.2), north of the equator. This island is separated from India by less than 35km. Sri Lanka lies between 79 ^{0}30" and 82 ^{0} 00" east longitude and 5 ^{0} 45" and 10 ^{0} 00" north latitude (Abeysekara, 1989). Colombo, the capital of Sri Lanka, is located in the western part of the country. The geography and politics of the region have played a

Figure 2.2. Location of Sri Lanka



major role in the formation of the present city, exerting both a positive and negative influence.

Climate

The country may be divided into two major climatic regions: the wet zone being about one third of the land area, and the dry zone. A widely used boundary differentiating these two zones is the 1900mm isohyet (Figure 2.3). Most of the country's hills are to be found in the center of the wet zone (Abeysekara, 1989).

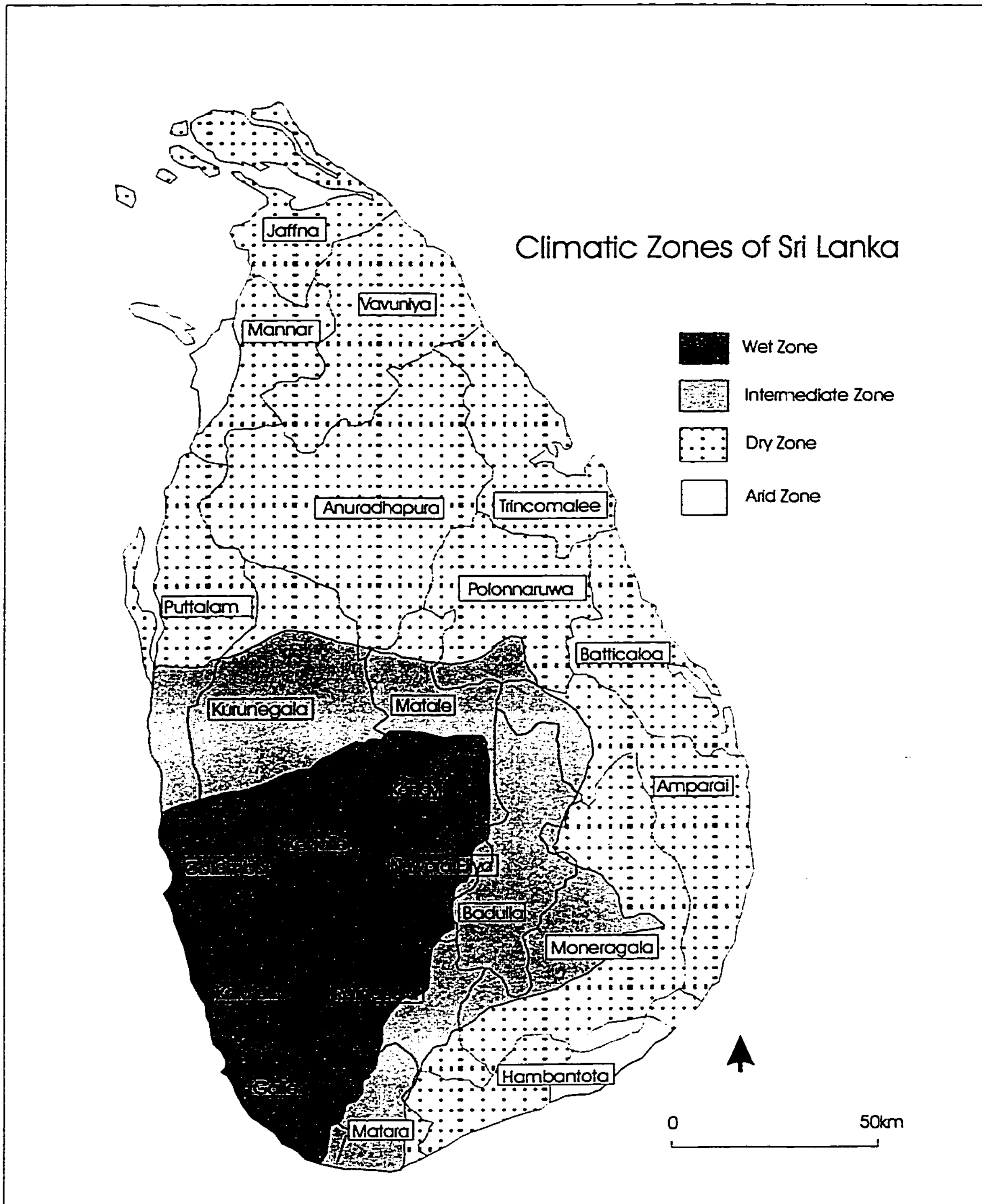
The Colombo district is located on the west boundary of the wet zone facing the sea (Figure 2.3). At present, the wet zone is the most favorable region for human settlement in terms of physical characteristics and services available. The region attracts rural people resulting in a very large urban population. In 1901, 29.6% of the population of Colombo district resided in urban centers. By 1991, this figure increased up to 58.8 %. Presently, two main internal migration streams are visible. The dominant one is to the Colombo district which contains the capital. The other migration stream is to the dry zone where successive governments of post-independent Sri Lanka have encouraged peasant agriculture.

The Colonial Imprints on the Townscape

Sri Lanka was under colonial rule, Portuguese, Dutch and British, for more than four centuries (from the 16th century to the middle of the 20th century) before finally achieving independence from the British in 1948. Naturally, colonialism has strongly influenced the present condition of Colombo and the country as whole. The introduction of different ethnic groups and religions, the establishment of a new capital (Colombo), plantation agriculture, a new administration system, a modern harbor and British education and justice systems made dramatic changes in the country, especially in the capital.

The introduction of plantation agriculture, and the neglect of traditional agricultural practices on which dry zone settlements mainly depended, greatly affected internal and external migration patterns (Arachchige-Don, 1994). With the plantation agriculture, a significant change in the pattern of immigration occurred. Labor for road and other construction works, and for plantation agriculture was brought in from south India. Since the British rulers did not encourage or maintain the traditional agricultural system,

Figure 2.3 Climatic Zones and Administrative Districts of Sri Lanka



Source: ESCAP (1975:17)

eventually, famine, poverty, destruction and starvation plagued the dry zone of Sri Lanka causing people to migrate into the wet zone. The majority of these people moved into the capital looking for labor oriented occupations. Most could not find proper work to survive and settled in areas which were marginal to other economic activities, building temporary houses with semi-permanent materials (shanty) or moving to the over-crowded buildings (slums) closer to the city center. Poverty became a new element of the plurality in this multi-racial society. Figure 2.4 shows the spatial distribution of slums and shanties of Colombo in 1990.

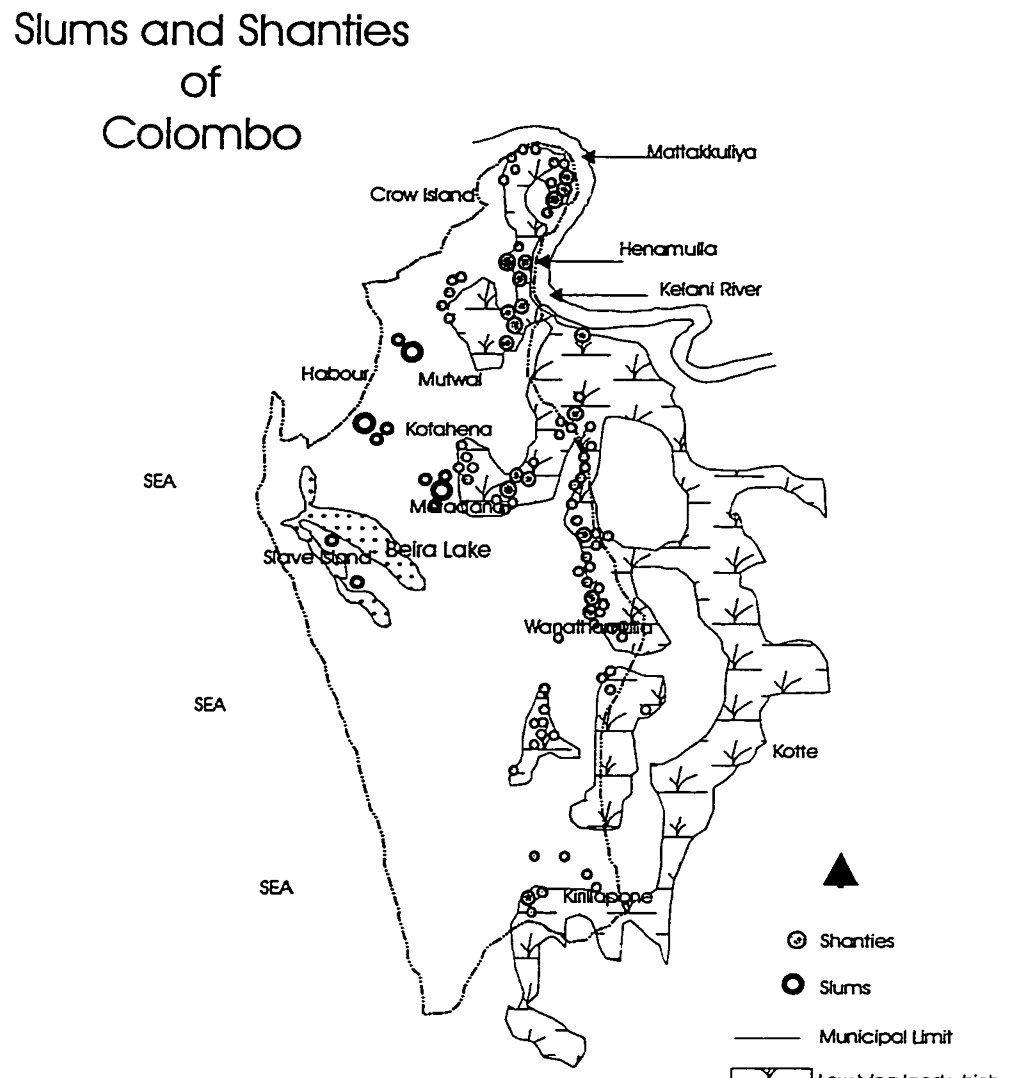
At the other end of the changing social spectrum was the appearance of a new class of elite Sri Lankans. This group used the wealth they had accumulated through capitalist enterprises to acquire high social status. In addition to their investment in plantation agriculture, they controlled graphite mining and the profitable liquor industries. They used their wealth to educate their children in European universities. This wealthy elite have the political power, or at least have access to it. Consequently, they are a dominant factor in controlling the city growth. Their land and properties are safe under the existing municipal and government regulations. Therefore, illegal settlements in the city are usually restricted to certain parts of the city.

The Growth of Colombo

The Colombo Municipal Council was established under Ordinance 17 of 1865 by the legislative council of Ceylon (People's Bank of Sri Lanka, 1994). As Figure 2.5 shows, the geographic boundaries of the Colombo municipality have not remained constant over the decades. Figure 2.5 indicates that the city area under Colombo has grown from 24.48 sq. km (9.5 sq. miles) in 1871 to 38.95 sq. km (15.1 sq. mile) in 1991 (National Housing Development Authority of Sri Lanka, 1991). The growth of Colombo continues to be dynamic and exerts a wide and important influence beyond its physical city limits that extends over the greater metropolitan area (the Urban Colombo District- Figure 2.6).

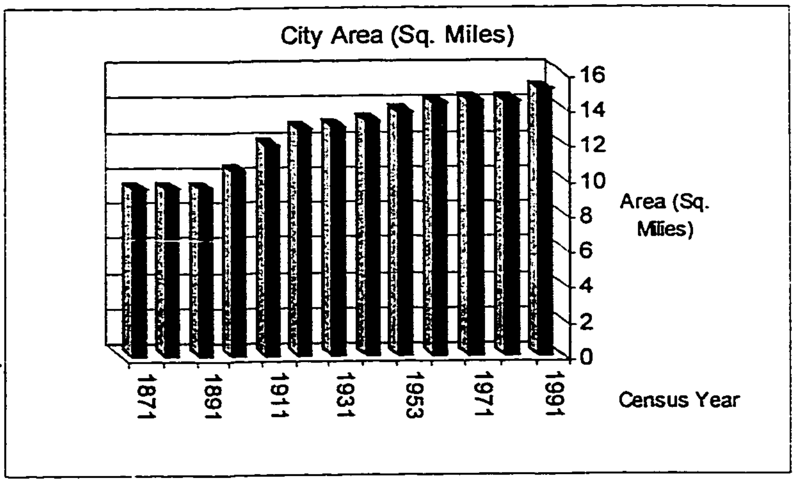
Three major suburban satellites namely, Dehiwela - Mt. Lavinia, Kotte, and Moratuwa, are included under the Urban Colombo District. Figure 2.6 shows the spatial distribution of the urban Colombo District (the greater metropolitan area) and

Figure 2.4 The Spatial Distribution of Slums and Shanties of Colombo



Source: Colombo Municipal Council, 1990

Figure 2.5 Growth of the Colombo Municipal Area



Census and Statistics Department, 1995: Statistical Report, Sri Lanka.

the area under the Colombo municipality. These suburban satellites are contiguous areas of the present city although they are not considered as areas within the city boundaries. The researcher understands these suburbs play a vital role in terms of absorbing the immigration to the city and influencing the growth of the city; however, it is important to note that the present study is based only on Colombo, due to the reasons discussed below.

The major reason for limiting the study to only the area under the Colombo municipal limit was that appropriate data were not available for the suburbs. Specifically, unlike the caseof the Colombo municipality, detailed information on slums and shanties was not available. Furthermore, published census statistics do not provide data on the subdivisions of the three suburbs. Moreover, data collecting agencies such as the Census and Statistics Department of Sri Lanka, Urban Housing Development Authority, and the Colombo Municipality (see the section under data collecting and storing systems of this chapter) collect and store data only for the area under the Colombo Municipal Council. The suburban cities, Dehiwela, Moratuwa, and Kotte, have their own city administrative systems.

As Figures 2.7, 2.8 and 2.9 demonstrate, the city defined by Colombo Municipality still plays a major role in terms of Colombo Metropolitan growth.

Figure 2.6 Urban Colombo

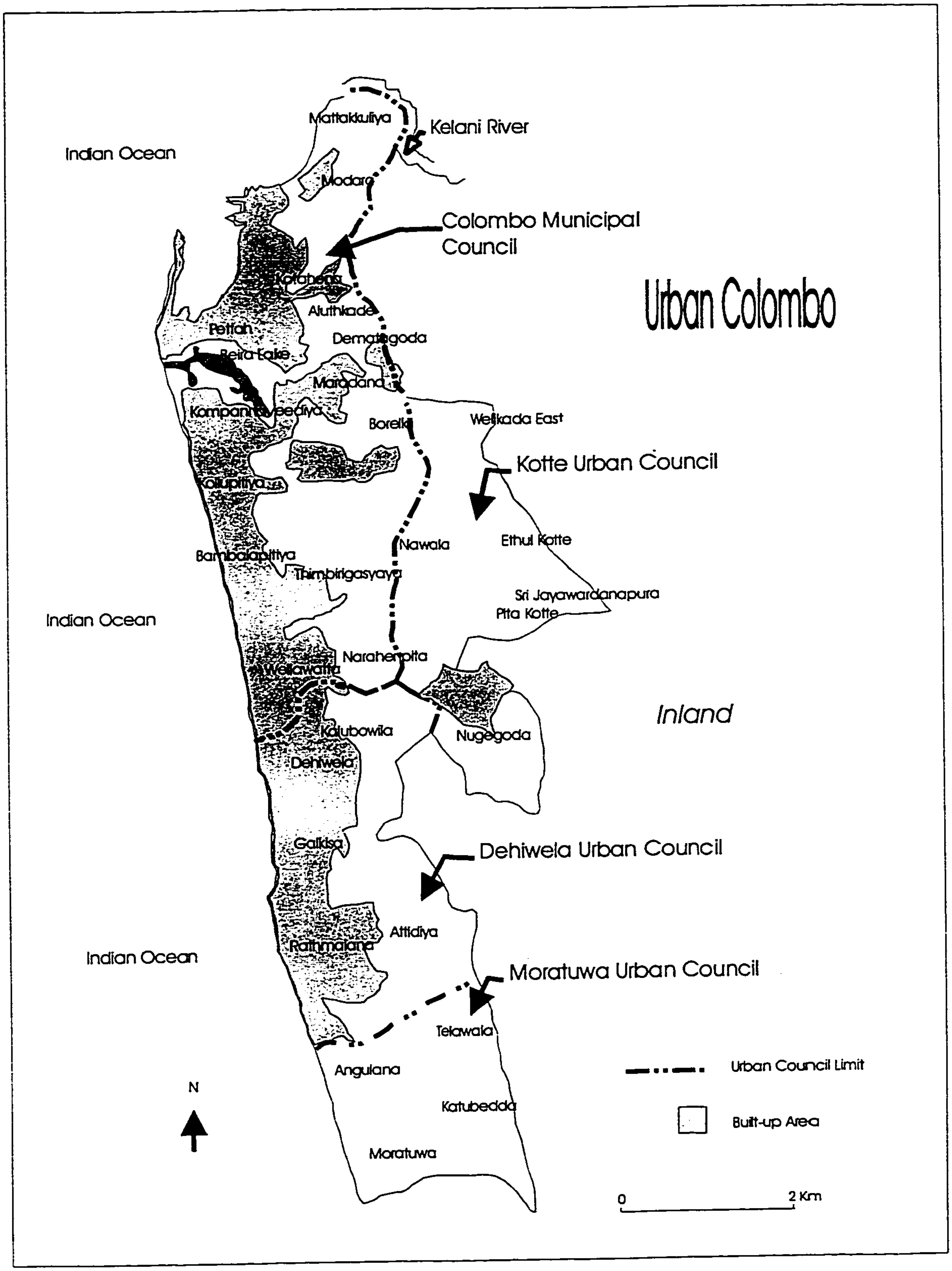
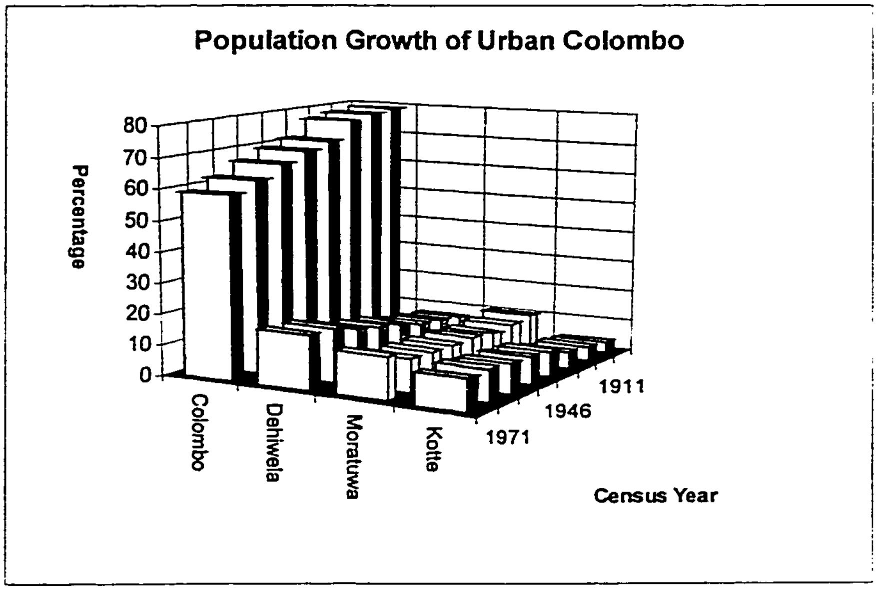
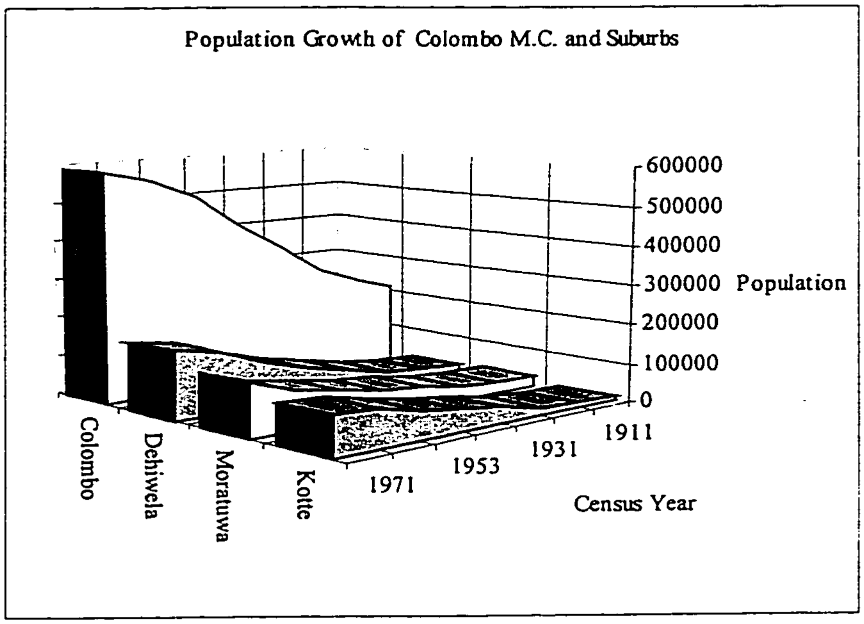


Figure 2.7 Population Growth of Urban Colombo



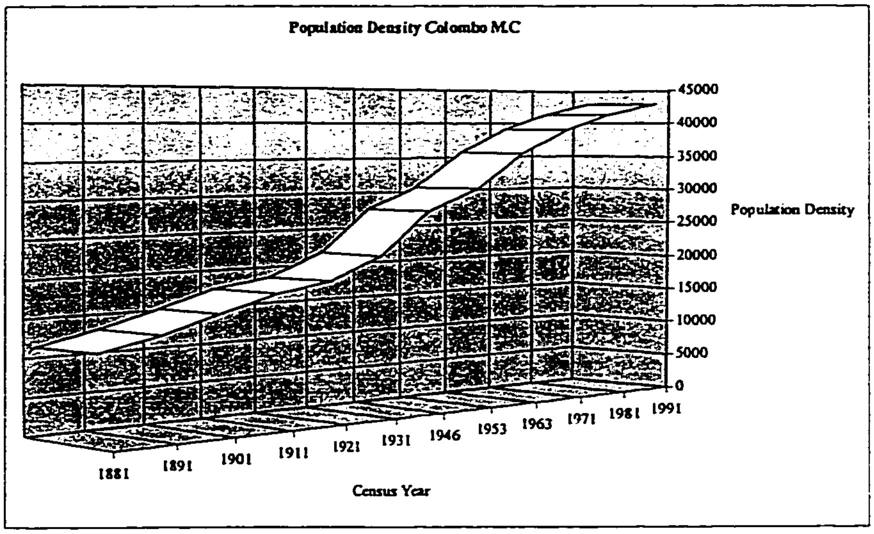
Source: Census and Statistics Department, 1991

Figure 2.8 Population Growth of Colombo M.C. and Suburbs



Source: Census and Statistics Department, 1991

Figure 2.9 Population Density of Colombo



Source: Census and Statistics Department, 1991

Although its contribution to the total population of Urban Colombo is decreasing gradually, the city defined by the municipal boundaries contributes almost 60% to the total population of Urban Colombo, which is fairly high (Figure 2.7). The growth of the city of Colombo can be measured through an analysis of its demographic changes as well. As Figure 2.9 suggests, the population density of Colombo is increasing constantly. The density figures given are for the entire area under the city limits; however, the density figures should be converted into area under residential land use, to obtain a real image of the city. To illustrate, there were 5.1268 Sq. miles under residential land use in 1992, and an estimated population of 685,376 for the city during the same year. Thus, the population density under residential land use in 1992 was 133,683 per sq. mile, which is more than three times (3.286) higher than the density figures given for 1982 for the entire city (including non- residential land use).

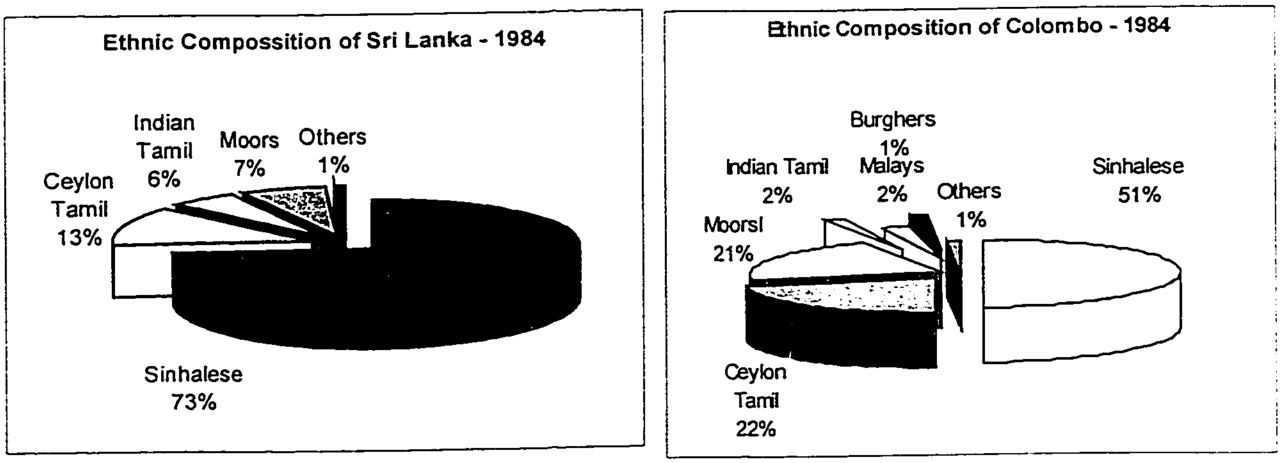
Furthermore, the proportion of built-up area within the Colombo city limits is remarkably larger than in its suburbs (Figure 2.6). City growth is restricted by the sea on the west and by the Kelani river on the north. Therefore, the main growth of the city is directed towards the south and southeast. In 1978, the administrative capital of Sri Lanka was shifted to a Colombo suburb at Sri Jayawardanapura, Kotte (Figure 2.6). Except for a few administrative institutions, which were shifted to the new capital, Colombo still retains most of the administrative institutions and it remains the financial and commercial center for Sri Lanka. Considering the facts discussed above, it is clear that the city of Colombo plays a more important role than do its suburbs.

Ethnic Composition:

Over the centuries, the city's population has been a mixture of three main ethnic groups, namely, Sinhalese, Tamils, and Moors. In the census records, the Tamil population has been grouped into two main categories: Indian Tamils and Ceylon Tamils. Ceylon Tamils have been in the country for centuries, but Indian Tamils were brought as laborers to the country during the British rule. Ceylon Tamils are much better off compared to Indian Tamils. Indian Tamils tend to be poorly educated; therefore, they do not have access to higher paying jobs. The Ethnic composition of Colombo is somewhat different from the general pattern in Sri Lanka (Figure 2.10).

While Sinhalese are the dominant ethnic group, their position in Colombo (51%) is less than the island average (73%). Moors contribute a significant proportion to the city population (21%). Other minority ethnic groups contribute fairly large percentages to the Colombo total population relative to the country's situation, making the city truly a multicultural society.

Figure 2.10 Ethnic Composition of Colombo and Sri Lanka



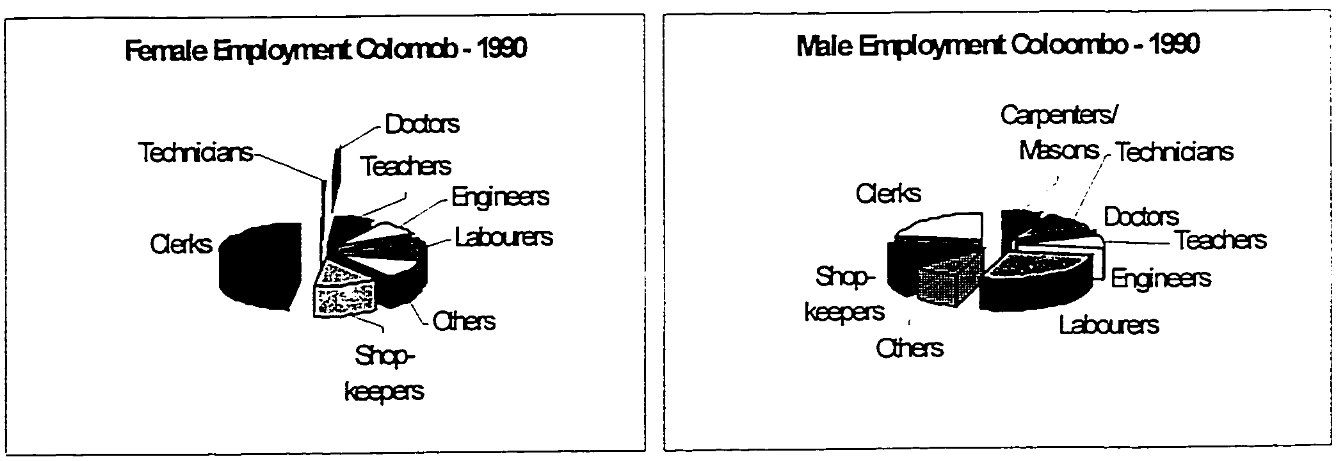
Sources: Department of Census and Statistics (1985) and Colombo M.C., 1988

Occupational Structure:

The socio-spatial structure of a city is directly correlated to the occupation structure. The present occupational structure of most developing world cities is a result of a colonial era and the activities of their present governments. During the colonial era, export-oriented products such as tea, rubber, cocoa, coconut, and spices were processed and packaged at Colombo. Besides these small scale processing industries, Colombo has been playing a pivotal role as the major administrative and service center for centuries. Major governmental and international financial and service institutions, offices, retail trade, personal (consumer) service, and professional services are all located in Colombo.

Figure 2.11 illustrates the employment structure of Colombo. Jobs can be divided into two sectors: formal and informal (Drakakis-Smith, 1992). Formal employment may be classified as the 'labor aristocracy' of relatively well-paid, privileged workers, such as, doctors, engineers, teachers, clerks and executive professionals. The informal sector is dominated by low skilled, manual labor intensive, poorly paid occupations such as laborers, carpenters, masons, security guards, housemaids, and street venders. It is clear that of educated women, most are employed in the 'formal' sector of the workforce, while employed males predominate in the 'informal' sector. A large percentage of poor and uneducated men are employed in the 'informal' sector. In contrast, the formal sector is

Figure 2.11 Colombo Employment Structure



Source: Resource profile of Colombo A.G.A. Division, Colombo, 1990, Ministry of Policy Planning & Implementation.

dominated by females for the following reason. Giving a dowry in cash or properties to the groom by the bride family was once common in all major ethnic groups, Sinhalese, Tamils and Moors. This dowry system has now been replaced by education. Parents do their best to provide their daughters with a good education in order to qualify them for a more advantageous marriage. This may have been reflected in the formal sector of employment with a higher percentage of women in occupations such as doctors, engineers and clerks.

Education Facilities:

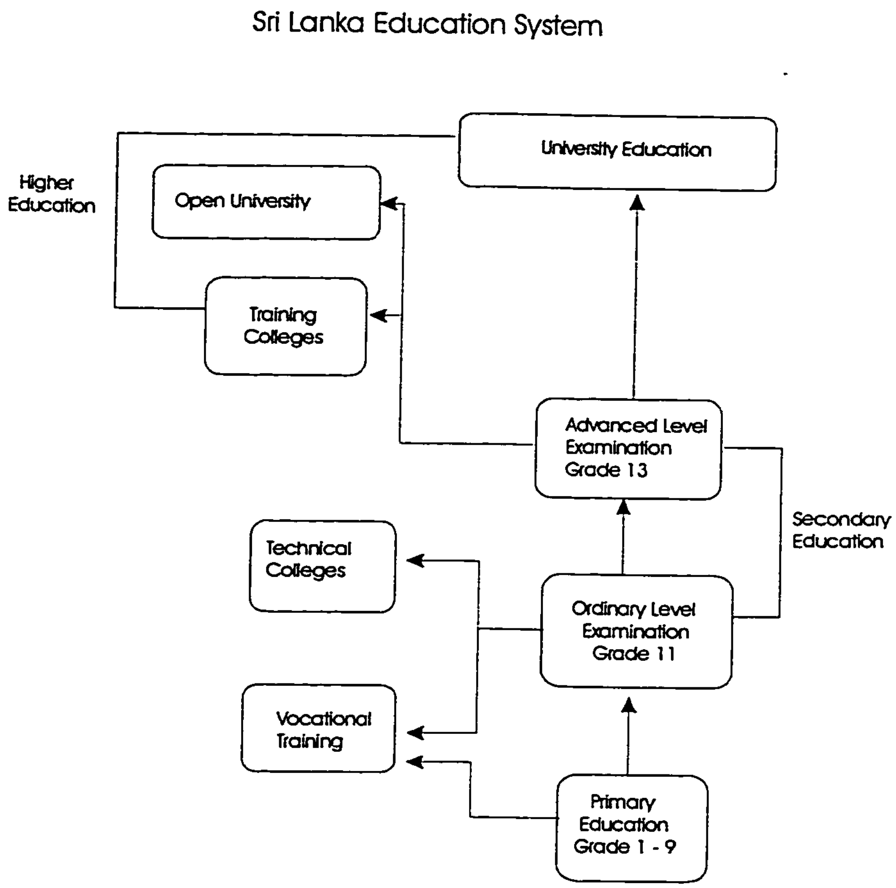
In order to understand the socio-economic profile of Colombo, one should pay attention to the structure of the education system including the distribution of facilities, teachers and their training, and resources of schools. Since the early 1940s, education has been viewed by successive governments in Sri Lanka as a strategy for reducing social injustice and inequalities, narrowing the gap between the 'have' and the 'have nots', and enhancing development (Hemachandra, 1991).

As a consequence of this political commitment, educational reforms and measures were introduced over the last five decades, making education compulsory up to the age of 14 years, and expanding education through a network of primary and secondary schools spread throughout the island. The introduction of free education from kindergarten through university, the provision of a common curriculum including science and English up to grade 11 (General Certificate of Education - Ordinary Level), the establishment of a wide range of scholarships at the university, and the giving of free mid-day meals, books, and school uniforms, all represent positive changes at every level of the education system.

Because of the various educational programs implemented over the past five decades, Sri Lanka's primary school enrollment rate rose from 57.6 percent in 1946 to 83.7 percent in 1991. A survey conducted on a sample of 2337 slum and shanty families in the city in 1984 by the Non-formal Education Branch of the Ministry of Education revealed that 22.1 percent of primary school aged children did not have any schooling (Ministry of Education, 1986). The educational level among slum and shanty dwellers is even lower than the urban average (Arachchige-Don, 1994). The main factors affecting low education levels among the children living in the slums and shanty dwellings are poverty, ignorance, and separation of children from parents when they later take up low paid employment especially in the Middle East countries (Ministry of Policy Planning and Implementation, 1990).

Figure 2.12 shows the Sri Lankan Government Education System and the critical stages of the system. Among all the students sitting for the Advanced Level

Figure 2.12 Sri Lanka Education System



Examination, approximately five percent are selected to enter the university annually. Although the country has ten large universities all over the island, they cannot meet the country's demand for higher education. Therefore, a large percentage of qualified students do not have access to the free university education system. Their next choice is to enter the Open University operated by the government. To register in the Open University, students pay a nominal fee.

Only those teachers with Advanced Level Qualification can enter teachers' training colleges to study for the Teachers' Training College Certificate. Teachers with a university degree in any field are considered the most qualified and are paid higher than teachers with only a Teachers' Training College Certificate.

Because Colombo is easily accessible from any part of the island, there are a number of higher educational institutions located within the city limits in addition to the very popular government and public schools. These higher education institutions include the University of Colombo, Law College, Ayurvedic University, Institute of Aesthetic Studies under the University of Kelaniya, Postgraduate Institution of Medicine, and Open University. When the ratio of teachers to students in Colombo's schools is considered, it becomes evident that there are some schools without enough qualified teachers. Most of the popular schools are staffed by qualified teachers in many areas. However, some schools do not have qualified teachers for some disciplines at all, for example, in such specialized areas as mathematics, science, music, and sports.

Public Health Facilities:

In Coiombo there were 23 public health centers run by the government and 28 private health centers in 1990 (Ministry of Policy Planning and Implementation, 1990). The ratio of people to doctors was 900: 1. But, as is often the case, these statistics do not reveal the true health situation in the city. The government provides free healthcare facilities to the public but, most of these have inadequate facilities compared to private health centers which benefit the wealthy with state of the art equipment and services. Outpatients at public health centers receive an admission number from the public hospital for specialized services such as eye, ear, throat, and heart. The numbers are allocated on a first come first served basis. It is very common to see long line ups for these services very early in the morning (around 1.00 a.m) at the hospitals. Location and the distance to the health centers may play a role in terms of access to the public health centers.

The situation at these government centers is further deteriorating since the government launched severe cutbacks to social services programs in the 1980s. Although there are no exact statistics to show this, the proportion of government spending per capita on health care has fallen in the last 10 years, mainly due to the increase in military expenditures.

In order to ascertain the urban social structure of the study area, detailed information on the city is needed. The following section describes the major information sources, data collecting and storage systems, and problems related to them.

DATA COLLECTING AND STORAGE SYSTEMS:

To explore the data collection and storage systems being practiced in the study area, the initial interviews held with the relevant authorities were very useful and augmented the researcher's practical experience collecting data for his master's thesis (Gunadasa, 1988). There are two types of institutions involved in this process: governmental and non-governmental. The Housing Development Authority, Urban Development Authority, Colombo Municipality, Census and Statistics Department, Survey Department of Sri Lanka, Grama Sevaka Officers, and Ministry of Planning and Implementation are the major governmental institutions from which the researcher collected data. Besides these government institutions, there are numerous non-government institutions involved in data collecting and storing. The Marga Institute and Center for Housing Planning and Building are the key non-government institutions that the researcher accessed.

The Census and Statistics Department, which handles the census data collection and storage, is the largest data collecting institution in Sri Lanka. Every ten years, it conducts an island wide census covering the total population. Besides the major censuses, it conducts sample censuses for various other purposes. This institution is the major data source as it covers the total population and conducts regular censuses.

First, the Census and Statistics Department aggregates the collected data by census enumeration tracts and stores them at this level. At this point in the process, the individual identity of the data disappears. The next step is to further aggregate the data into wards, the highest level of aggregation at which data are maintained. The stored data can be retrieved by census enumeration tracts or wards. Although data can be retrieved by census enumeration tracts, they cannot be linked to the spatial units from which they were collected because a census enumeration map has not yet been constructed for Colombo. The only available map for the collected data is the Colombo ward map. In addition to the ten-year census survey, the Census and Statistics Department conducts sample surveys for

Colombo during irregular time intervals. These sample surveys are used to estimate the population trends between the ten-year census periods.

The Colombo Municipal Council which collects and stores data about city residents is another data source. The Colombo Municipal Council provides and maintains certain municipal activities such as collecting garbage, assessing property taxes on housing, approving and issuing building permits, land use zoning, road maintenance, and electrical supply. The Municipal Council keeps records by wards about each city dweller. This is the same ward system that the Census and Statistics Department utilizes to maintain and store its data. The geographical location of each citizen is recorded by address but actually locating someone on the city map is difficult, as the addresses are not given in a standard format. Sometimes the address is referenced at a particular road name, a housing scheme, or by a road intersection. As an added complication, many of these addresses cannot be located because the available maps are not detailed enough for the purpose.

As the data are not computerized and do not have clear and proper geographical coordinates, they cannot easily be used in a GIS easily. In addition to regularly updating the information on city dwellers, the Colombo Municipal Council (CMC) and National Housing Development Authority (NHDA) conduct detailed surveys on selected themes. For example, in the past, the CMC and NHDA have conducted the following surveys with respect to the shanties and slums in Colombo.

The first census of the shanties within the city of Colombo was conducted by CMC in 1960. Neither the data of this survey nor the report based on the survey is available. In 1973, another census of the shanties was conducted by CMC to gather information on the provision of basic amenities to shanty dwellers. This survey was restricted to the shanties on state and municipal lands. Data were collected for specific slums and shanty areas and collated by those specific areas. The names of the slums and shanty areas are given, but maps of those areas are not available. This data, therefore, cannot be combined with spatial units. It was observed that there were 7,593 shanties on the selected lands. The number of shanty families was 10,000 while the total shanty population numbered 60,000.

Again in 1978, the National Housing Development Authority undertook a census of the shanties and slums that covered the entire Colombo. The purpose of this survey was to identify the slums and shanty areas to be declared as special project areas and to propose suitable building guidelines for such areas. The report does not provide detailed information by any spatial unit. The major objective of the next survey conducted in 1988 by the Colombo Municipal Council was to collect detailed information on shanties and slums to enable the government to both design and upgrade slums and shanties and to ensure the successful implementation of the proposed programs. Under the United National Party government (1977- 1994), various housing programs were implemented to improve the living standards of the shanties and slum dwellers. The UHSP (Urban Housing and Sub Program), Major Settlement Schemes Housing Sub Program (MSSHSP.) and MHP (Million Houses Program) are examples.

In the 1990 survey, data were collected and maintained by road segments referenced to each slum and shanty unit. The nature of the data is different from the common forms. The data are georeferenced by street segments; therefore, the location of these road segments could be located on the Colombo A-Z Street Atlas map. In order to combine this type of information with other information given by administrative units, special techniques needed to be developed.

Another set of available data for the study area is given at the Grama Sevaka Area/Divisional level collected by Grama Sevaka officers. These data are less aggregated compared to ward-level data. One Grama Sevaka officer is assigned to each of 55 administrative areas/divisions of Colombo. The officer is responsible for collecting and maintaining records on people living in his/her area. These records show basic information on each household such as the number of persons, their age and education, household amenities, death records, and major features such as schools, religious places, and land under different land use categories. The Grama Sevaka officer updates all the information every year and sends it to the Ministry of Housing and Construction. Because these data are maintained at GSD level, they are spatially referenced.

In conclusion, most of the urban social structural studies available for the developed and developing world cities are conducted under various constraints. The researchers were forced to conduct their study based on a single source of information since the available techniques did not support integrating data from various sources. Furthermore, some analysis was done based on a very limited number of variables. In addition, the selected variables did not represent the socioeconomic environment in developing world cities well. This study attempts to fill this gap by applying GIS techniques to integrate data from various sources and thereby reveal the socioeconomic environment in a more appropriate way. Therefore, the selected case study area is a good model for the objectives of this research as adequate information on the social geography of Colombo was lacking. The following chapter will describe how the researcher gathered and processed information in order to bring it into GIS environment.

CHAPTER THREE

DATA PREPARATION

Introduction:

Data preparation for the study area was a great challenge in terms of problems encountered and developing new methods to overcome them. As noted above, many of the previous works in this area are based on either a single data source or more than one where the spatial units used in the analysis are completely compatible. In many cases, some of the census tracts are excluded or merged with others under various conditions such as data unavailability, inadequacy, or incompatibility. Some of the procedures taken to achieve the stated objectives are questionable, and the reasons given are unsatisfactory. Because some of the data were discarded, the analyses done based on the remainder do not represent reality. Therefore, many researchers have commented on the importance of integrating information from different sources.

For the present study, the researcher could not obtain adequate information from a single source on the socioeconomic background of Colombo; therefore, data from different sources were explored and then integrated. This chapter describes how this information was laid out to work in a GIS environment for statistical procedures. Almost all of the collected information is secondary, except for several interviews with directors of selected organizations. The following sections will show how the researcher processed spatial and attribute data to convert them into an-appropriate format for the study.

Data collection:

Secondary data for the study were collected during the months of April, May, and June of 1994. The researcher was not physically involved in the secondary data collection due to both monetary constraints and security concerns. It was undesirable to travel to Sri Lanka during that period due to political unrest in the country at that time. Therefore, it was decided to hire two local research assistants to collect the required information.

Dr. Kanthi Vitharana, Department of Geography, University of Colombo, Sri Lanka undertook to help in coordinating and guiding the research assistants. Both research assistants were final year students in the geography honors degree program, Department of Geography, University of Colombo, Sri Lanka. These students were selected because they had extensive experience in field research. The researcher was closely involved in the entire research project, coordinating research assistants and contacting Dr. Kanthi Vitharana through the internet, by telephone, and by mail. Daily situations were reported to the researcher via e-mail.

Informal Interviews:

Before the secondary data collection, research assistants were guided to conduct interviews with the following authorities: the Directors of the Housing Development Authority, the Urban Development Authority, the Center for Housing Planning and Building, the Census and Statistics Department, and the Chairperson of the Colombo Municipal Council. These interviews provided valuable information on the authorities' research projects, data collection methods, data storage systems, data analysis and their views on the socio-spatial structure of Colombo, its history, development, and future trends. Dr. Kanthi Vitharana attended two of these meetings. The research assistants were able to gather many materials including very rare and important research reports, data, maps, and journals from the directors. The interviews with the directors of the Housing Development and the Urban Development Authorities enabled the researcher to understand the government housing development programs and their targets. Further, the directors explained their views on the location of services for the poor, as well as general city land use planning and its impact on the poor. The institution's facilities such as library, workspace and photocopiers were made accessible to the research assistants. These interviews helped the researcher understand the availability and nature of the information and the type of data that should be collected to provide a spatial database for the city of Colombo.

Problems of Data Collection: Validity, Reliability, and Accuracy

The following discussion reviews the possible sources of errors related to the collected data. The level of error in a GIS needs to be managed so that data errors will not invalidate the information that the system is used to provide (Aronoff, 1991).

The Census and Statistics Department of Sri Lanka maintains records of individuals only at the ward level which is highly aggregated. This is in contrast to most developed countries where census results are reported at a number of resolutions and using a variety of media (Fisher, 1991, Fulton and Ingold, 1989). Due to national security concerns, the data stored on the computer systems are restricted only to authorized personnel. The research assistants needed to receive special security permission from the relevant authorities even to access the Census and Statistics Department resources. The researcher had access only to data published by this department at ward level. The Census and Statistics Department publishes data under a limited number of themes such as demography, age, and religion at the ward level. The other information available in the reports is a summary for the city.

Reported censuses are, however, subject to numerous problems of accuracy and reliability. The foremost of these is that individuals are enumerated in Sri Lanka by their 'usual place of residence' which may be different from their legal residence. Undercounting is a perennial problem due to the lack of cooperation from some residents with limited education and from illegal residents of the city. As well, some of the slums and shanty units are not registered in the government-housing registry; therefore, people living in those units are not counted. Furthermore, in Sri Lanka, it is usual in a census to count people by their nighttime location. Therefore, only a poor representation of daytime and working place population distributions may be recorded.

Even if one accesses the data stored by the department at enumeration level, these data cannot be integrated with the other data originating from different sources, because a census enumeration or a census tract map is not constructed even in an analogue form. A census enumeration or a census tract map may be developed based on the sketches given to the census data collection officers when they collect information from the field. The possibility of developing such a map was discussed with the director of the Census and Statistics Department. As the officer pointed out, it would not be easy to find those sketches as they are not well organized or well kept. The researcher's experience constructing a census tract map of a small city in Sri Lanka for his Master's thesis led to the conclusion that it would take months to produce an analogue map of census enumeration areas for Colombo.

The department did not conduct its regular island-wide census in 1991 due to ethnic and communal clashes within the country. For the present study, therefore, 1990 mid-period estimated data were selected. Demographic information is usually very time sensitive and can change significantly over the course of a year (Aronoff, 1991). Hence, one should be careful about any decisions based on this type of data.

The most disaggregated data available are from the street level. This data set is from the Report on the Survey of Squatter Settlements in the City of Colombo (Colombo Municipal Council, 1990). There are 203 street segments with data. Only a few variables were selected due to the nature of the data available from this source. Most of the report is comprised of tabular information on Colombo by electorates (there are five electorates) giving a general picture of the city. The report would have been more valuable for the present research if all the information were presented by streets. As it was, any information given by street segment was extracted.

The Resource Profile of Colombo A.G.A Division (Assistant Government Agent's Division), Colombo District (Ministry of Planning and Implementation, 1990) is the most valuable published data source in terms of the amount of information provided on the selected theme, and the completeness, accuracy and timeliness of the data. In 1990, the Ministry of Policy Planning and Implementation undertook the first ever attempt to compile information on both the needs and the constraints to development in all Assistant Government Agent Divisions on the island. The report on Colombo provides information on the Colombo Assistant Government Agent at the Grama Sevaka level. Not all the data contained in the report are referenced to Grama Sevaka Divisions (GSD); therefore, some valuable information could not be integrated to the spatial database easily.

The data set is complete in terms of coverage as it gives the information on selected themes for all the GSDs covering the city. Accuracy is expected to be high because the data were collected by Grama Sevaka officers living in the local area. The officer is familiar with the area, data set and the people living there and he or she updates the information annually. Therefore, this can be viewed as a reliable source of the latest information even though it contains omissions such as missing data for some Grama Sevaka Divisions. Also, while the rows are referenced to Grama Sevaka Divisions (records), the order of rows was changed from data set to data set.

Data Sources

The available studies on urban social structure rely on a single data source. Census data given by census tract, enumeration or ward level is the primary information source for almost all studies conducted in both developed and developing countries.

Not only are the existing studies limited to a single data source, many studies used only a limited number of variables to explore the urban social space. For example: Morris and Pyle, (1972) used only 22 variables while Abu-Lughod(1966) used 13 variables, Abu-Lughod (1980) used 27 variables, and Brand (1972) used 13 variables. Urban social structure should be described and explained in a multidimensional perspective using many variables in order to obtain an image closer to reality.

A large percentage of city dwellers in the developing world live in slums and shanties. Their role in shaping the city is extremely important (United Nations Center for Human Settlements, 1996; Michael, et. al 1996). Most of the available studies on urban social structure have not considered this issue adequately. The census questionnaire is designed to collect only general information from the city dwellers. Therefore, it does not provide the desired information on the slums and shanty dwellers.

The spatial data integrated for the present study were collected from the following sources:

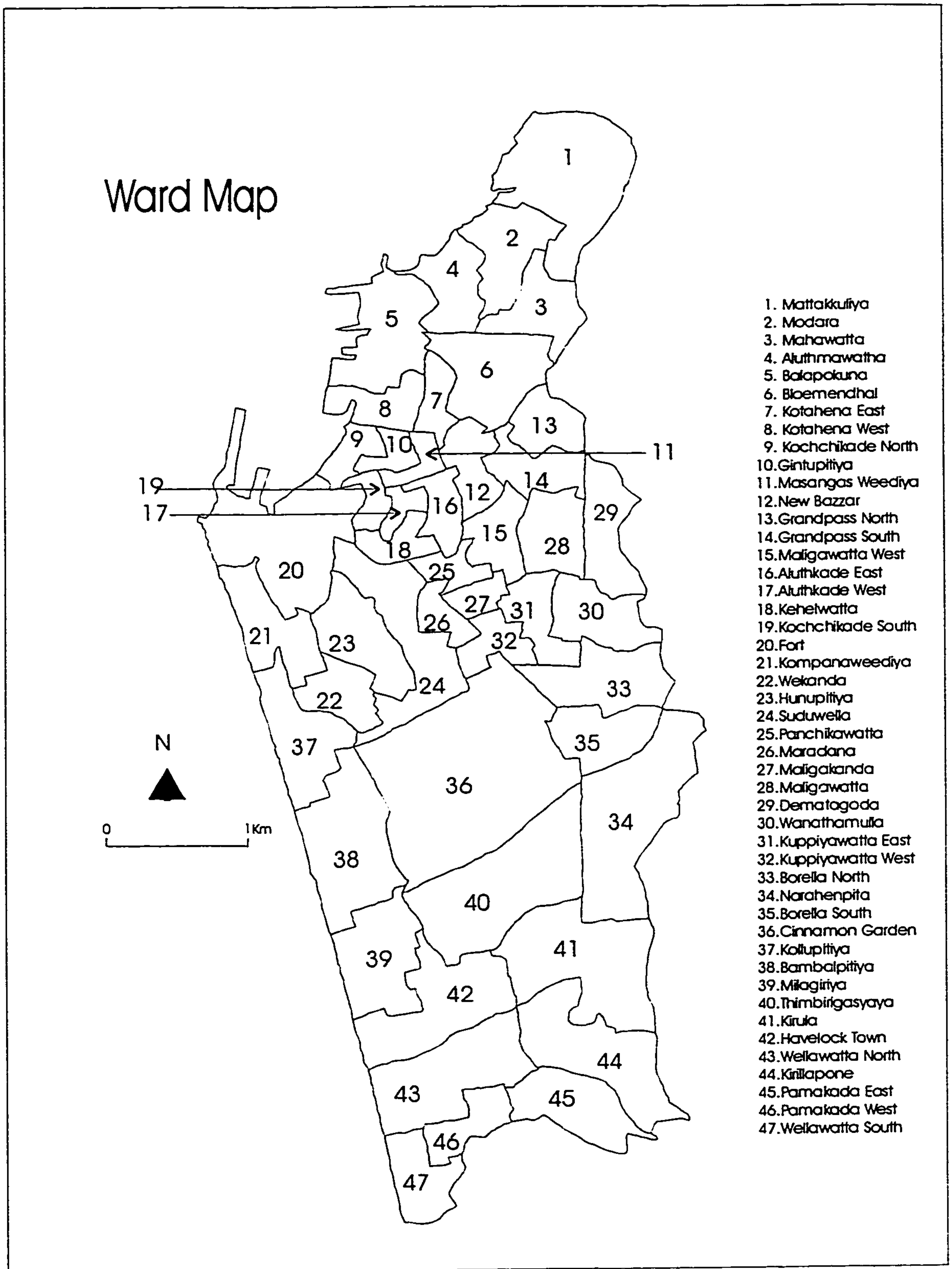
I. Ward level data
II. Grama Sevaka level data (village officer's administrative unit)
III. Street level data
IV. Colombo land use, and A to Z Colombo Street maps

The following sections describe these data sources in detail.

Ward level data:

Colombo has 47 administrative wards (Figure 3.1). There are two main data sets given at ward level that have been employed in the present study. The first is data collected in 1990 by the Census and Statistics Department of Sri Lanka (Census and Statistics Department, 1990). The second is from the Report on the Survey of Squatter Settlements in the city of Colombo (Colombo Municipal Council, 1990). Appendix 1 shows a list of variables from the first data set collected by the Census and Statistics

Figure 3.1 Colombo Ward Map



Source: Colombo Municipal Council: 1990

Department for 1990. From the list, only the following variables are selected for the spatial database, considering the reasons given below.

I. *Percentage of Sinhalese*

II. *Percentage of Tamils*

III. *Percentage of Moors*

IV. *Percentage of Population under Five years of Age*

The first three variables represent the ethnic composition of Colombo. These three ethnic groups represent 95.2 % of the total population; the remaining ethnic groups constitute a very small fraction and do not play a visible role in terms of their geographical distribution. The ethnic component is important to consider when examining the urban social structure of a city. The available literature shows that most North American and European cities have one or two ethnic dimensions in the urban social structure (Murdie, 1969; Davies, 1975) but none of the available studies for developing countries, with the exception of Berry and Kasarda's (1977) study on Calcutta show a clear ethnic dimension.

From the above variable list of variables, the first variable, *percentage of Sinhalese*, represents the majority ethnic group. The Sinhalese represent about 75% of the total population of Sri Lanka. The second variable, the *percentage of Tamils*, is comprised of both Ceylon Tamil and Indian Tamil which together form 19% of the Sri Lankan population. These ethnic groups are similar in many aspects, except that Ceylon Tamils have a longer history in Sri Lanka. The third variable refers to Moors. Based on the researcher's knowledge of ethnicity in Sri Lanka, minorities, especially, Moors, tend to isolate themselves from the Sinhalese. Moreover, although Tamils and Moors speak the same language, their cultural background is different. Unlike Tamil women, most of the Moor women are isolated from other ethnic groups, therefore the percentage of each ethnic group is very important to determine how the ethnic groups are spatially distributed. The three dominant ethnic groups discussed above were selected to investigate how they are associated with poverty and how they fit within the urban spatial structure.

Religion and ethnicity are superimposed in Sri Lanka. For example, almost all Moors are Muslims while most Tamils are Hindus except for a very small percentage that is Christian. The majority of Sinhalese in Colombo are Buddhists (67%), the remainder belongs to Christianity and to other small religious groups. Variables related to religion, therefore, were excluded to avoid repetition.

The fourth variable, *Population under five years of age*, was selected to determine its correlation with poverty. This has been a widely used variable in factorial ecological studies and most studies conducted in the developing world have employed it (Brand, 1972; Abu-Lughod, 1969 and 1980; Lo, 1986; Morris and Pylr, 1972; Yeh, et. al 1995). For the study area, there could be a positive correlation between this variable and slums and shanties, which represent the poorest people in the city since poor families tend to have a fairly large number of children compared to wealthy families (Hemachandra, 1991).

There are two levels of data encountered in the second source of information (The Report on the Survey of Squatter Settlements in the City of Colombo by Colombo Municipal Council, 1990). One set of data is given by street segments of which there are 203. The second set of data was given by wards. This data set has been generated by aggregating the street level data into ward level. The following is the list of variables that are given by wards:

I. *Shanty population*

II. *Age structure of shanty population*

III. *The percentage of shanties reached*

IV. *The percentage of shanties to be reached*

It is evident that the information in the report is mainly related to slums and shanties in the city of Colombo (see Chapter 2). Using the variables given above, seven variables were constructed and utilized in the study:

I. *Shanty population density per hectare*

II. *Percentage of population under 6 years*

III. *Percentage of population between 6 - 17 years*

IV. *Percentage of population 18 - 55 years*

V. *Percentage of population over 56 years*

VI. *Percentage of shanties reached*

VII. *Percentage of shanties to be reached*

The first variable, *shanty population density per hectare* (for an area under residential land use), is a very useful indicator of the geographical distribution of shanty population and its relationship to the other socioeconomic variables in developing world cities. None of the urban social structure studies conducted on developed or developing world cities was based on appropriate density measures. For example, general population density as a variable was included in many studies (Lo, 1975 and 1986; Yeh et. al 1995), but the population density figures estimated for the entire area do not represent the actual spatial distribution of population. To see the spatial distribution of density values, and its relationship with other socio-economic conditions, population and housing density values should be estimated on areas where people live excluding non-residential land use. In the previous studies, population density of a given census tract, enumeration area, or census district was estimated by dividing the total population of the selected area by the total area. This is a very crude estimation of how people and housing are concentrated over space, because the area under water bodies, parks, roads, playgrounds, commercial activities, industries, and other areas where people do not live were not excluded from the total area. The present study proposes the estimation of a "physiological density' (De Blij and Muller, 1994) where only the areas under residential land use are considered.

The next four variables, II - V, exhibit the demographic composition of the shanty population. The *percentage of population under 6 years* is a good indicator for determining whether or not the shanty dwellers have higher percentage of young dependents. *Percentage of population over 56 years* tells us about the elderly dependents. In Sri Lanka, the government retirement age is 55. The *percentage of population between 6 – 17 years* age represents school age children since at six years of age students enter the primary education system and graduate at the age of 17 from grade

12 to qualify for university education. Next, the *percentage of population 18 - 55 years* represents the work force. Previous researchers did not follow a common age structure to define dependents, labor force, and school age children (Abu-Lughod, 1969; Brand, 1972; Gerald, 1972; Berry and Karsada, 1977; Lo, 1975 and 1986). The sixth and seventh variables summarize the progress of the shanty upgrading programs launched by government and non-governmental organizations during the 1978 - 1988 development year.

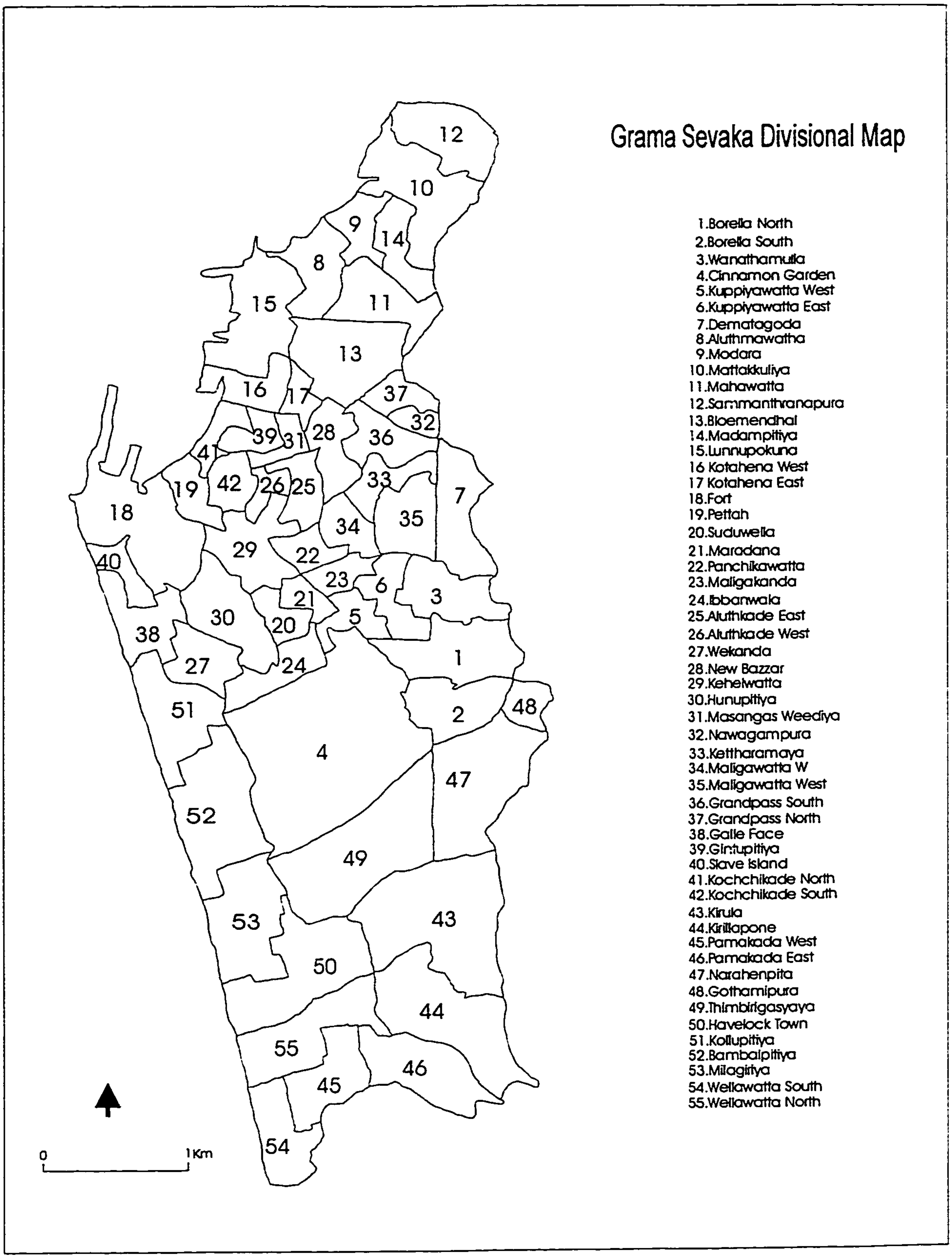
Under the shanty development programs, the slum and shanty dwellers received new houses or upgraded their units during the above period. The report classifies those shanties converted to permanent buildings with long lasting materials such as bricks, stones, and asbestos as "shanties reached." The other shanties, which were not transformed to permanent buildings, are designated as "shanties to be reached." These last two variables will indicate if there is any significant improvement in terms of raising the shanty dwellers' socio-economic situation.

The Grama Sevaka Divisional data

The Grama Sevaka Divisions are the smallest administrative units in Sri Lanka. Size of the unit varies over space. There are 57 Grama Sevaka Divisions (GSD) in Colombo (Figure 3.2). Each GSD is assigned to one village officer who is in charge of collecting, updating, and maintaining records of people as well as properties in his or her administrative area. In some cases, records are by streets and in others the officer aggregates them as total of GSD. Other agencies also collect data and store them by GSDs. These include the Urban Development Authority, Colombo Municipal Council and Ministry of Policy Planning and Implementation. The variables chosen from GSD records were modified and are shown in Table 3.1.

The first three variables indicate the demographic composition of the city. Unfortunately, the data set does not provide detailed information on age structure, restricting the ability to estimate the dependent population. It is expected that these variables would have assisted in understanding the demographic composition and its spatial distribution in the city.

Figure 3.2 Grama Sevaka Divisional Map



Source: Colombo Municipal Council: 1990

Table 3.1 Grama Sevaka Level Variables

- a) *Percentage of population less than 5 years of age*
- b) *Percentage of population 5 - 18 years of age*
- c) *Percentage of population 19 - 29 years of age*
- d) *Female/male employment ratio*
- e) *Percentage of male employment of the total work force*
- f) *Percentage of female seasonal employment of the total work force*
- g) *Percentage of male seasonal employment of the total work force*
- h) *Female/male unemployment ratio*
- i) *Percentage of male unemployment of the total work force*
- j) *Percentage of female foreign employment of the total work force*
- k) *Percentage of male foreign employment of total work force*
- l) *Percentage of employment in high income group (engineers, doctors, teachers, clerks, technicians)*
- m) *Percentage of employment in low income categories (carpenters, laborers, trade persons, and other minor income groups)*
- n) *Number of unemployed people with a degree Qualification per 10,000 unemployed people*
- o) *Number of unemployed people with Advanced Level Qualification (grade 13) per 1,000 unemployed people (male and female)*
- p) *Number of unemployed people with Ordinary Level (grade 11) Qualification per 1,000 unemployed people*
- q) *Number of unemployed people with 8^{th} grade standard education per1,000 unemployed people*
- r) *Number of unemployed people with technical education per 1,000 unemployed people*
- s) *Number of schools per 10,000 students*
- t) *Number of students per teacher*
- u) *Average school building square feet per student*
- v) *Number of student desks per 10,000 students*
- w) *Number of student chairs per 10,000 students*
- x) *Trained Mathematics and Science teachers per 10,000 students*
- y) *Trained Arts Teachers per 10,000 students*
- z) *Trained Commerce Teachers per 10,000 students*
- aa) *Science and Mathematics graduate teachers per 10,000 students*
- bb) *Arts and Music graduate teachers per 10,000 students*
- cc) *Commerce and other graduate teachers per 10,000 students*
- dd) *Untrained English teachers per 10,000 students*
- ee) *Untrained Dance, Music, Sports and other teachers per 10,000 students*

The variables (d) to (k) show the 'nature of employment' by 'sex' in the city. The detailed employment structure of a city is one of the best sets of variables to unveil the urban social structure of a city. Information on this theme has been used in all the studies available for both developed and developing world cities. The percentages of unemployed and seasonally employed men and women who tend to have low incomes are taken into account by this data set. The researcher has some doubts that this might not give the realimage of the nature of employment of the study area. On the one hand, many women, although they are counted as employed, earn very little money working as housemaids, operating small businesses, or working for a private factory. The same applies to men who have very low incomes. To investigate the relationship between poverty and employment characteristics, the variables above were selected.

The next two variables, *percentage of employment in high* and *low-income* categories are related to the type of occupation. Based on the available employment classification, two major employment categories are created: *percentage of people employed in higher income occupations* and *lower income occupations*. Doctors, engineers, teachers, clerks, technicians, earn a higher income than laborers, masons, carpenters, and small business owners. Although employment categories are very crude, it provides a general picture of the spatial employment structure of Colombo. None of the data sources had information on income, so this would be a good surrogate variable for income.

None of the studies available on developing world cities, except a study by Yeh et al. on Guangzhou in China (1995), are based on such a detailed employment structure. Seasonal employment, which has not been counted in any of the previous studies, is a major component of the employment of developing world cities. As many as 30%, or over 100 million people, in developing world urban areas may be considered as heavily underemployed or seasonally employed (Michael et al., 1996). As a large proportion of the city dwellers in the developing world cities live in squatter settlements, the studies on these cities should be focused on capturing the contribution of these squatters to the urban social structure by selecting the proper variables in the urban social context of the selected city.

The variables from (n) to (r) portray the unemployment structure by the educational level achieved. This is a new set of variables that was not employed in earlier studies in both developing and developed world cities. Information like this is a critical measurement for the understanding of the nature and spatial distribution of unemployment. Most poor people do not have higher education and as such have limited access to permanent and higher paid occupations. Therefore, economically disadvantaged areas may have a high percentage of unemployed people with low levels of education. Thus, variables were modified to suit the information available. For example, very few unemployed people have a university degree; therefore, the number of such people was calculated per 100,000. In order to keep uniformity among the variables, the same operation was applied to all the categories of unemployment with the given educational standings.

The subsequent five variables were selected in order to investigate the spatial distribution of government allocation of educational resources and their relationship with the low-income areas. Literacy rate was used in previous factorial ecological studies to measure the spatial differentiation of literacy. The nature of the selected variables are unique to this study since they will help the researcher determine whether or not the poor are disadvantaged in terms of access to adequate levels of educational resources. To illustrate, some schools do not have enough desks or chairs for students or are; otherwise, they are overcrowded. Thus, the *average number school buildings per square feet per student* may be a good indicator to capture the spatial distribution of poverty and its relation to educational facilities. Similarly, the *number of students per teacher* may be a good expository variable in examining whether the resources are allocated equitably throughout the city. Original data values have been converted to "per10, 000 students" to maintain uniformity among the variables.

The variables (y) to (ee) show the educational qualifications of the teachers. In Sri Lanka, there is a shortage of qualified graduate science and mathematics teachers. The trained teachers, those who have only a diploma, are not permitted to teach grade 12 and 13 classes, but they have better qualifications than do untrained teachers. Untrained teachers, those without any teaching experience, degree or diploma in their field, do have only a grade 12 certificate (G.C.E A/L) and are allowed to teach from grades 1 to 9 (primary level). In some cases, trained science specialists teach grade 12-13 classes where graduate teachers are not available. From personal experience living in Colombo and other areas of Sri Lanka, the researcher believes that teachers, no matter what their qualifications, prefer to teach in popular schools which are located close to good residential neighborhoods. Teaching in these schools brings many advantages not the least of which are safety, pride and prestige. Because of this, schools located in prestigious neighborhoods are staffed by teachers with high academic qualifications. These variables are unique to this study. By including these variables, the researcher expects to determine how selected resource allocation affects the poor. All variables have been converted to units per 10,000.

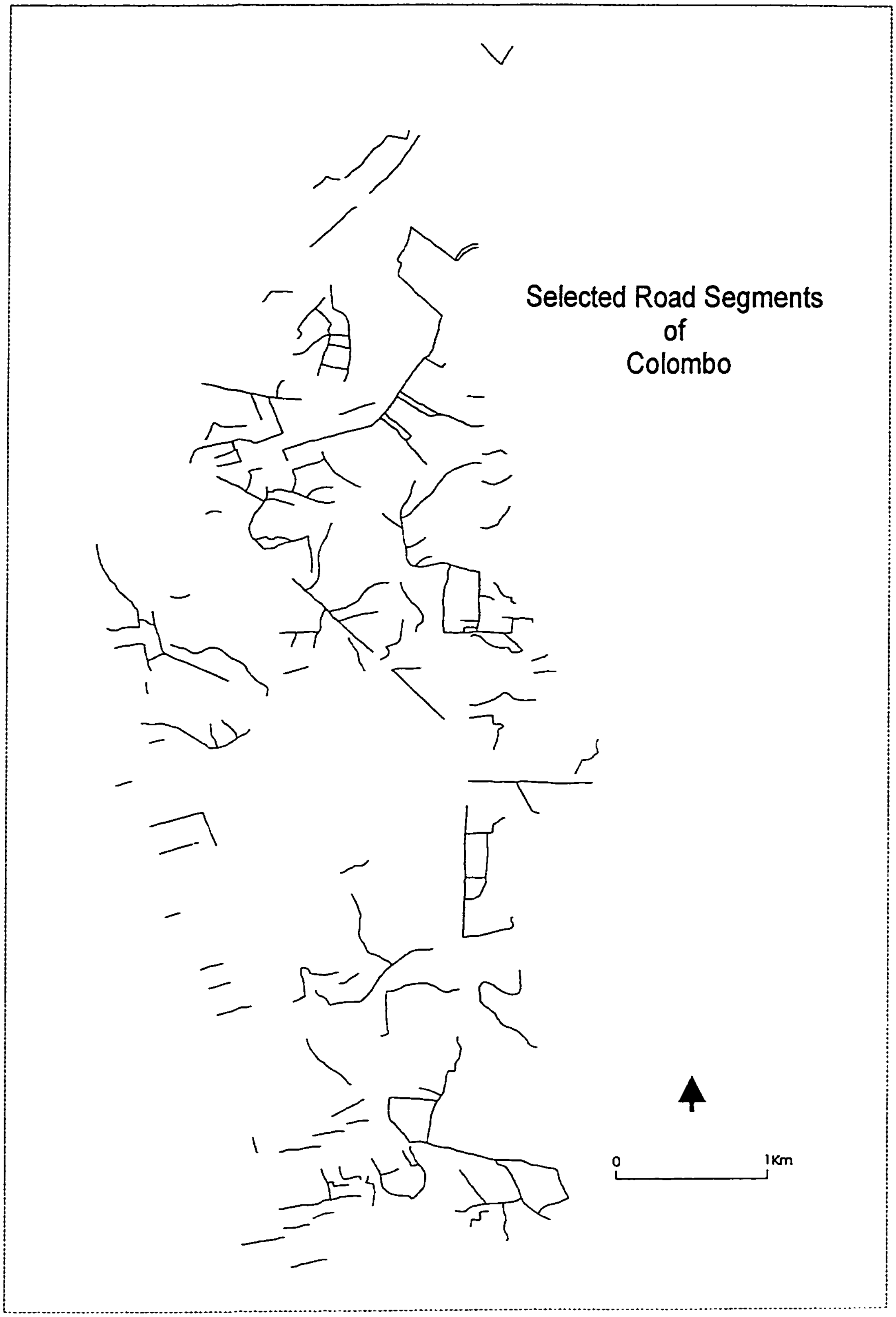
Street Level Data:

This is the third information source (Figure 3.3). The characteristics shown in the Table 3.2 are the only ones on slums and shanties given at street level. All the variables provided in the report are used to identify the socio-spatial structure of Colombo. This is the first time in the study of urban social structural applications that detailed information on slum and shanty dwellers and their dwellings at street level has been merged with information based on the general population and then analyzed. In 1972, Morris and Pyle's study of Rio de Janeiro included some information on shanty dwellers, but the

Table 3.2 Variables Given at the Street Level

I.	*Average number of years of residency: less than ten years*
II.	*Average number of years of residency: 10 - 20 years*
III.	*Average number of years of residency: more than 20 Years*
IV.	*Percentage of good housing units*
V.	*Percentage of renovated buildings*
VI.	*Percentage of slums and shanties*

Figure 3.3 Map of Selected Street Segments of Colombo



study excluded some parts of the metropolitan area. Furthermore, Morris and Pyle and Gerald (1972) used only 22 variables and did not collect adequate information on the shanty dwellers.

The detailed information recorded on length of residency was grouped into three categories: less than 10 years, 10 - 20 years, and over 20 years. As the available sources do not provide direct information on migration patterns of the city dwellers, these three variables can be used to reveal the migration patterns of the slums and shanty dwellers. Areas where a higher percentage of shanty and slum dwellers have fairly short length of residency (in this study *less than 10 years of residence*) can be identified as newly growing shanty areas. Thus, the length of residence of shanty residents is a good indicator of the spatial distribution of shanty development.

The remaining three variables are related to the housing quality. The characteristics of housing quality and the spatial distribution of the housing types are a good indicator of housing stock in a given area. *Good housing units* have basic amenities such as water, toilets, and enough space for the people living in the unit. In addition, the structure and materials used to construct the units are permanent and the units are relatively new. *Renovated buildings* are quite old and their appearance does not look pleasant. In most cases, these units have basic amenities such as water, electricity, and sewage facilities (Ministry of Housing and Construction, 1991). These buildings, owned by the residents, are located closer to the city center but are not overcrowded as are the slums.

Data from the Maps:

In addition to the information mentioned above, the researcher acquired more information from the following maps which he used to calculate measurements related to distance:

i. Location of Public Health Centers (Figure 3.4)

ii. Location of Public High Schools with science and mathematics facilities for Advanced Level (Grade 12) (Figure 3.5).

iii. Colombo Land Use map - 1992 (Developed by the Colombo Urban Development Authority, Sri Lanka (Figure 3.6)

Figure 3.4 Location of Public Health Centers of Colombo

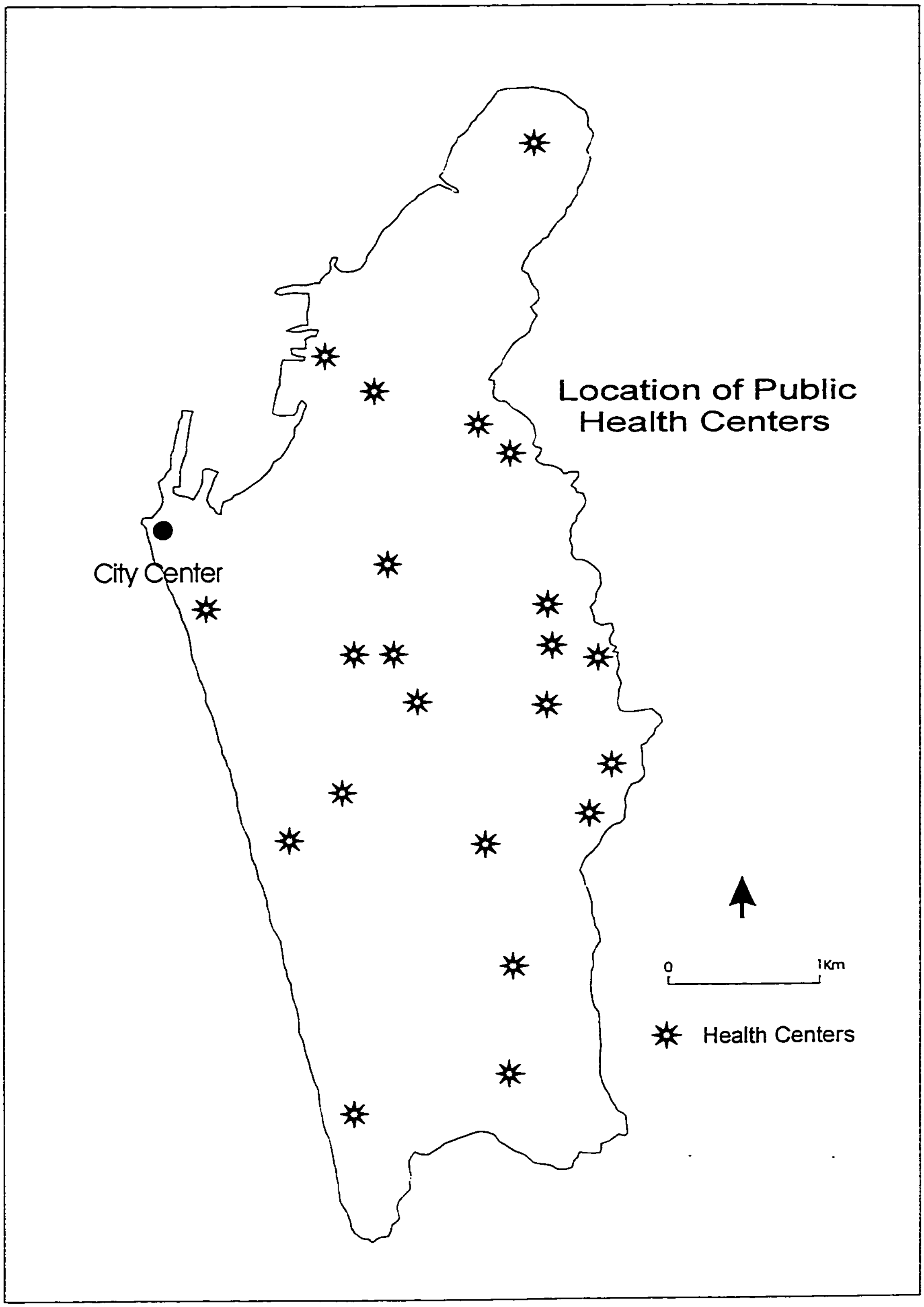


Figure 3.5 Location of Secondary Schools of Colombo

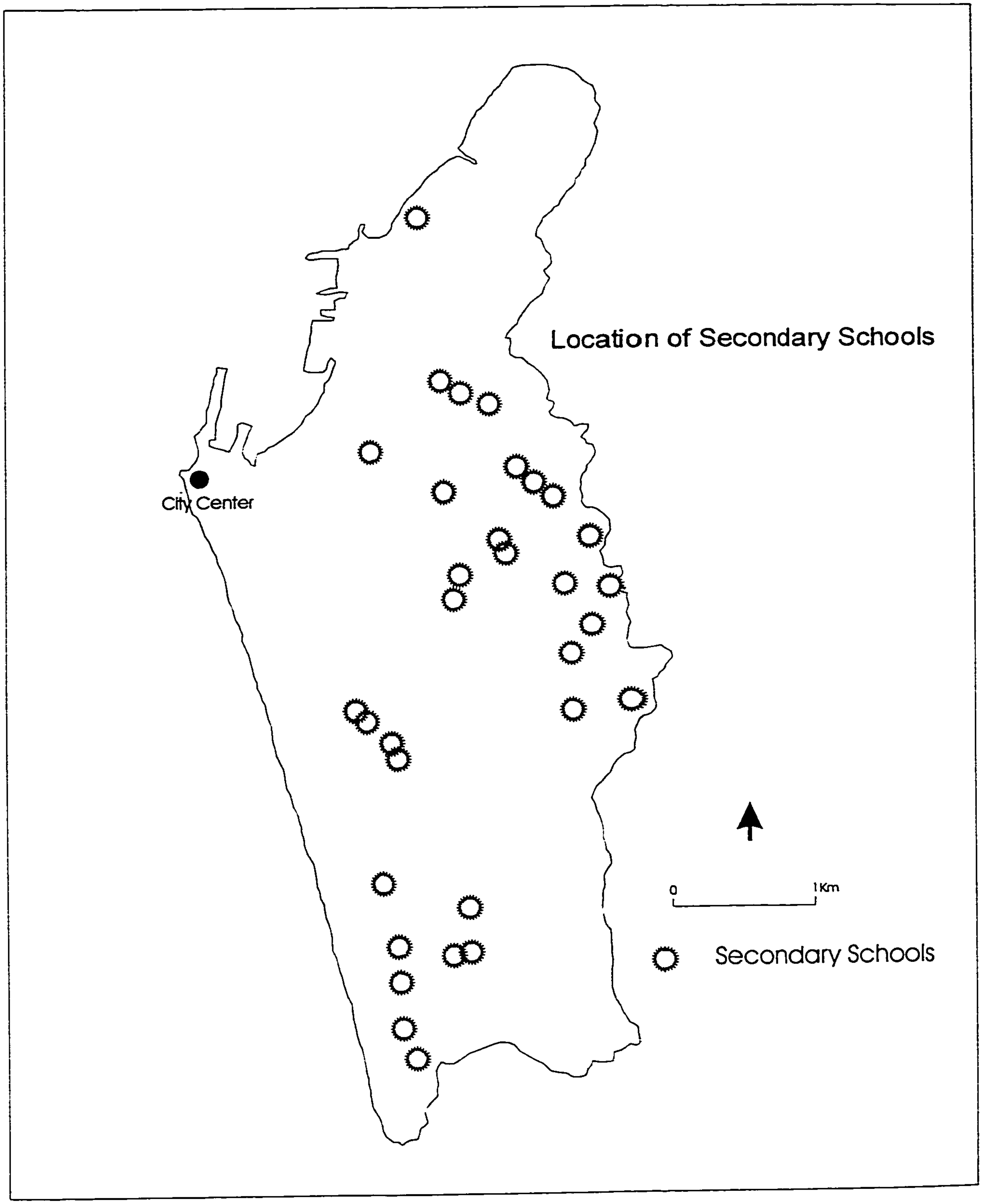
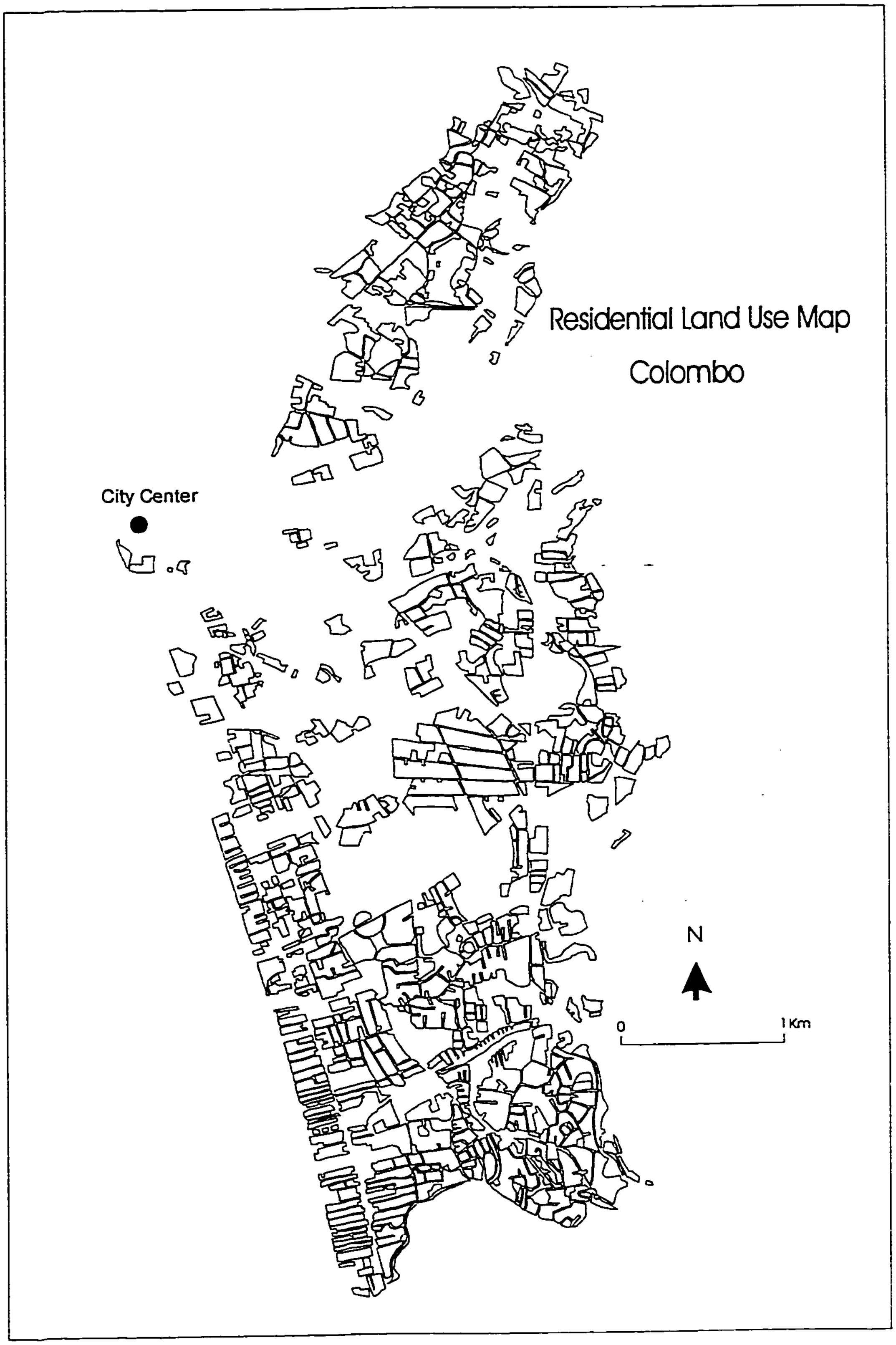


Figure 3.6 Residential Land Use Map - Colombo



Geographic dimensions such as distances to the nearest service centers were selected as variables to determine how they correlate with the spatial distribution of the low-income areas. The extracted factors will be explored further with selected distance measurements in order to determine any relationship between them. The remaining variables from the three information sources cited above are considered to be expository variables as they are all employed to describe the urban social structure of the selected city.

The A to Z Colombo Atlas, developed by the Survey Department of Sri Lanka, was important as it was the only map with UTM coordinates. This map was employed to locate all the facilities selected, and to extract the street network of Colombo. The original map scale was 1: 12,500; however, it was reduced for the digitizing process to 1: 18,750 since a map with a scale of 1:12,500 was extremely large. The addresses of all the names of public health centers and high schools with science and mathematics facilities for Advanced Level (Grade 13) were obtained from the Resources Profile of Colombo A.G.A. Division, Colombo District 1990. The geographic coordinates of public health services and high schools were also derived from the A to Z Colombo Atlas map.

The land use map of Colombo developed by the Urban Development Authority of Colombo (1992), Sri Lanka, was used to create a Residential Land use map (Figure 3.6) and to identify the Central Business District (CBD). The data from different sources were assigned only to areas of the category "residential land use". By assigning density figures related to population only to those areas under residential land use, one can derive a better picture of the spatial distribution of the theme considered than by allocating those values to the entire area. Therefore, the land use map was very helpful. Nine major land use categories and 73 subdivisions were identified. Land use categories "detached and semi detached household units", "flats and apartments", "row houses", "residential hotels", "transient lodging and guest houses", "hostels and group quarters", and "slums and shanties" are classified as being under residential land use by the Urban Development authority of Colombo.

To identify the CBD of Colombo, land use categories under "retail trade", "banking institutions", "personal (consumer) service", and "professional service (legal, insurance, and travel)" were considered.

Preparing the Map Sheets and Digitizing:

None of the maps used in the research was originally in digital form. To be used in a GIS, map data had to be digitized. This step is the largest task of the research project in terms of time consumption and problem resolution. Preliminary preparations and the actual process of geocoding the maps were started in December 1995. Once a map is drawn using a geographical coordinate system, the absolute position features can be mapped. Moreover, the information originating from another map, which has a geographical coordinate system, can be integrated. A GIS requires that a common coordinate system be used for all the data sets that will be used together (Aronoff, 1991).

Digitizing software named TOSCA was used to convert map information into digital format (Clark University, 1997). TOSCA supports most of the popular digitizer models and includes the ability to create and edit georeferenced data that can be directly integrated with the GIS package (this will be discussed later) that the researcher has selected for the study.

The following maps were prepared for geocoding, a process that took about three months.

- Colombo Ward Map (Figure 3.1)
- Colombo Grama Sevaka Division Map (Figure 3.2)
- Colombo Selected Street Segment Map (Figure 3.3)
- Public Health Centers (Figure 3.4)
- Public Schools with Science and Mathematics for Grade 13 (Figure 3.5)
- Colombo Residential Land use Map (Figure 3.6)
- Colombo City Center

Colombo Ward Map:

The ward map (Figure 3.1) that was in the Report on the Survey of Squatter Settlements in the City of Colombo (Colombo Municipal Council, 1990) was used for geocoding. This was the only ward map that the researcher was able to acquire. Since the map had been produced by a government organization, it was assumed to be fairly accurate, as cartographers normally follow standard maps. Unfortunately, the map had neither a scale nor any geographic coordinates indicated.

Each of the 47 wards in Colombo was treated as a polygon. The original map was in very good shape and was not shrunk or stretched. Therefore the original paper map was chosen for digitizing. The researcher decided to use a user-specified XY coordinate system for the map to be digitized. For this purpose, the researcher used an arbitrary XY coordinate system for the map as described below. A rectangular frame was drawn on the map on a scale, the horizontal and vertical axes of the rectangular frame were considered as X and Y axes respectively. The minimum and maximum values of the X-axis were set to 0" and 10" respectively. For Y-axis, these values were 0" and 12". All the boundaries of the 47 wards were digitized as separate lines giving each line a user identifier (a code). The errors involved in the lines were corrected and the coordinates of the lines were used to build the topology. Typically, a GIS maintains the arc as the basic unit, storing it with the information on left and right polygons. A polygon is comprised of a closed chain of lines that represents the boundaries of the area, in this case a ward. Polygon topology shows the arcs that make up the boundaries of each polygon. Based on this information, forty-seven polygons were constructed, each representing a ward.

Grama Sevaka Division Map

Although three Grama Sevaka Divisional maps were found in the literature, it was decided to use the Grama Sevaka Divisional Map (GSD Map – Figure 3.2) that was in the Report of Resource Profile of Colombo A.G.A division. The other two maps were sketches of the GSD map and some parts of the city were not shown clearly. The map used in the report was published in 1990, the latest map available. Colombo comprises 57 Grama Sevaka Divisions. The selected map does not have either a scale or a coordinate system. A user defined coordinate system was imposed for geocoding purposes. The same procedure applied to the ward map was used here with the same parameters for X and Y-axes. All the boundaries of the 57 GSDs were digitized as separate lines giving each line a user identifier (a code). The errors involved in the lines and topology were corrected and 57 polygons were constructed using TOSCA.

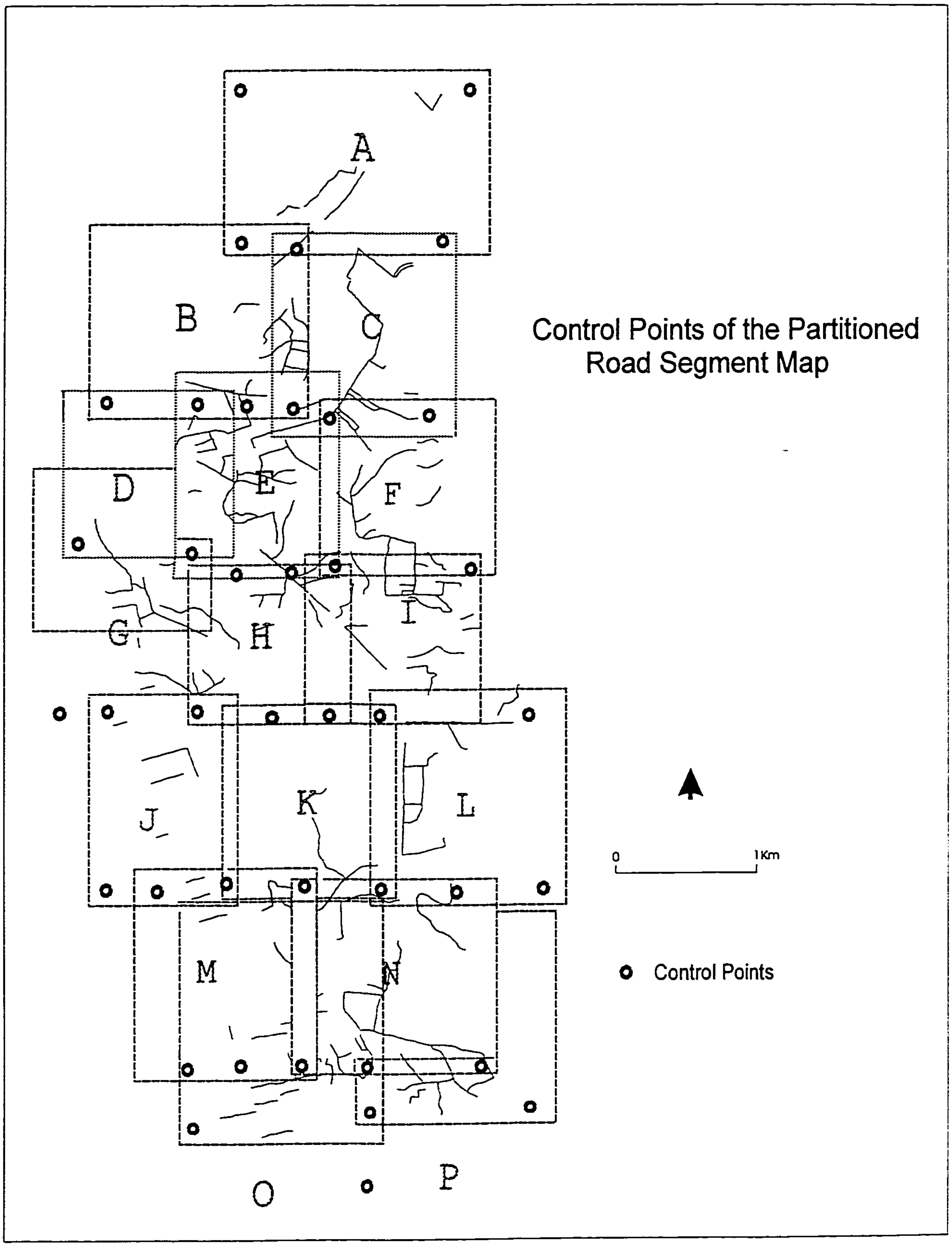
Street Segment Map:

Preparing the map manuscript for automation was done in several steps. The Report on the Survey of Squatter Settlements in the City of Colombo (Colombo Municipal, 1990) presents data on 203 road segments. A map showing the street segments (Figure 3.3) was not included in the report but the names of the street segments were. Therefore, the researcher had to construct a map to show the location of each street segment. The A-Z map is a small atlas. All the pages of the atlas were photocopied and a base map of Colombo was constructed at 1: 12,500 scale. All the names of road segments were obtained from the above report. The 203 road segments were then located on the A-Z Atlas Map and traced to a MYLAR sheet to avoid the roughness at the edges that were used to merge with other portions of the map. The A-Z atlas comes with UTM (Universal Transverse Mercator) coordinates.

The copied map with road segments was reduced to 1: 18,750 scale to decrease the number of portions (tiles) to be digitized. The entire map was partitioned into 16 portions making some overlaps because each map portion should have some common areas from all the adjacent map portions in order to construct a map that merges all the digitized map portions accurately. Figure 3.7 shows the partitioned road segments map and the control points used for digitizing. Each map portion should have at least four control points for digitizing (Clark University, 1997). For some map portions, seven to nine control points were selected. The digitizing program (TOSCA) gives the error component of control points called the Root Mean Square (RMS).

The RMS is calculated by determining the positional error of the test points, squaring the individual deviations and taking the square root of these values (Clark University, 1997). The allowable amount of RMS error depends mainly on the map scale. The control points selected on the map are considered as true values. When these points are transformed into digital format by a digitizer, the exact locations of the control points are changed due to factors such as digitizing device errors, technician handling errors, and map surface errors. Based on the control points, TOSCA can calculate the error involved

Figure 3.7 Control Points Used for Digitizing the Partitioned Road Segment Map



in the translation of map coordinates to digitizer device coordinates. It was able to achieve an acceptable level of error for the map portions after increasing the number of control points in some instances. Based on a rectangular grid, with minimum and maximum values of 'X' and 'Y' axes of each map portions set as previously described, the street segment map was digitized in 16 portions (tiles), checked for errors and merged.

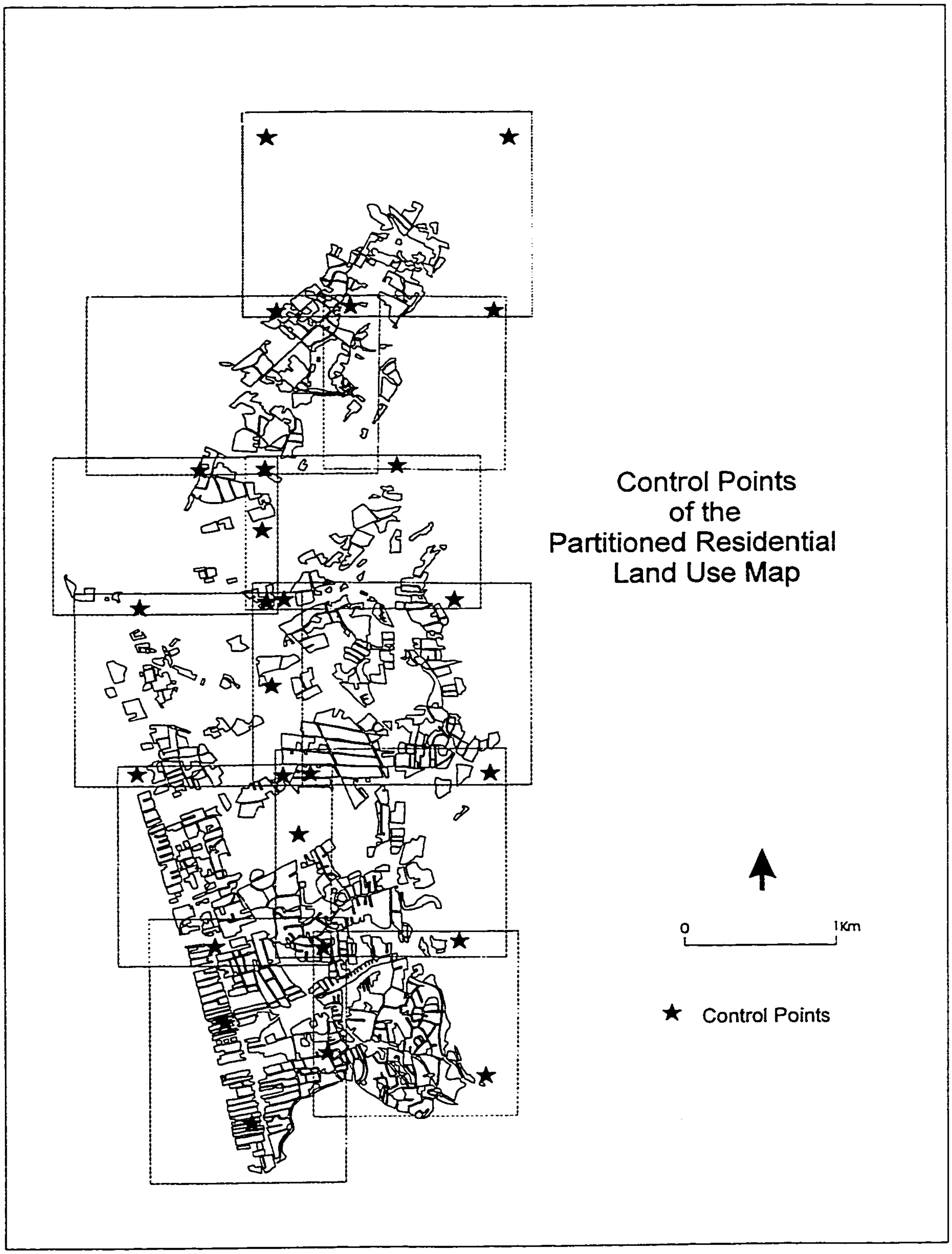
The Residential Land Use Map:

Three land use maps of Colombo were available. It was decided to use the land use map developed by the Urban Development Authority of Colombo (UDA), Sri Lanka in 1992 because the other two maps, both produced in 1978, do not have detailed information on land use categories and furthermore, the scales are extremely large. The UDA land use map utilized in the current study is the latest, and the most reliable map as it was constructed based on air photographs taken in 1990 at 1:11,250 scale. This is the largest and most complicated map to be automated. It was decided to use the original land use map at the same scale because if the map scale were decreased, some small polygons, which represent residential land use, would not be captured due to their extremely small size.

The complete map was automated in 11 portions (Figure 3.8). Each portion of the original map was digitized using user defined X, Y coordinate system. The minimum and maximum coordinates of the X-axis were placed as 0" and 12" respectively for each portion and for the Y axis, the minimum and maximum were 0" and 16" respectively. Based on the map UTM coordinates, the residential land use map was digitized in 11 portions (tiles) giving each residential polygon a value starting from "1". There were 596 separate residential land use polygons. The digitized portions were not perfectly matched at the edges. As a result, some residential land use polygons were not confirmed between sheets.

This problem was solved by finding common polygon boundaries for the adjacent sheets and then running an "automate feature matching procedure" over the boundary. This procedure was very time consuming, as the user had to examine common boundaries of both map sheets very carefully to match the corresponding polygon on the adjacent

Figure 3.8 Control Points Used for Digitizing the Partitioned Residential Land Use Map



map sheet. Once the corresponding polygons were identified on the adjacent map sheet, the researcher needed to combine each corresponding point at the boundary manually. Fortunately, it was possible to merge all portions. It took about four weeks to digitize the residential land use map as there were hundreds of small polygons representing areas under residential land use. Some portions of the map had to be redigitized due to the high error involved in the process.

Government Public Health Centers, Pubic High Schools with Science and Mathematics for Advanced Level classes (Grade 12- 13) and City Center:

This is the last set of information to be extracted from the maps. The report on the Resource Profile of Colombo A.G.A. Division contains addresses for all public health centers and public high schools with different facilities within the Colombo Municipal Council. According to the above report, there were 23 public health centers and 33 schools with the considered facilities in Colombo in 1990. This information was used to locate and trace the public health centers and schools from the A-Z Colombo Atlas Map. The city center was located and traced using the same map. Locations of schools, health centers, and the city center were traced on the map used to trace road segments, and all were digitized as points.

In conclusion, to study the social fabric of Colombo, all the possible sources of social variation were considered. Data and maps related to the above theme were collected and transformed into digital form. It is important to note that all available data were in different sets of zones and the digital map boundaries of Colombo were different. Ideally, for study, all of the data should be allocated only to areas under residential land use. Finally, the data from different sources had to be integrated in order to conduct multivariate analysis to reveal the social structure of Colombo. These critical issues were addressed by applying GIS (Geographic Information Systems) techniques. The following chapter will describe how the researcher applied GIS techniques to address these critical issues.

CHAPTER FOUR

SPATIAL DATABASE INTEGRATION

Introduction

In most previous studies, exploring the socioeconomic structure of cities has been restricted to information based on a single set of spatial zones. There are only two studies available which utilized information from two incompatible zoning systems whereby information from one zoning system was interpolated to another zoning system manually. In most cases, the studies were limited to a single source, ignoring information from other sources. Furthermore, some areas of the case study areas were excluded from the studies as they were not compatible with other zoning systems. This naturally results in the reality of the city not being revealed as fully as would be desired. The issue mentioned above has been in the field of urban factorial ecological analysis for the last three decades, but none of the available studies has thus far addressed the issue satisfactorily. The main purpose of this chapter is to demonstrate how GIS techniques can be used to solve problems related to disparate data sets.

Developing an integrated spatial database for the study area was a great challenge. The available data for the study area were given in three different formats: data collected by line segments, data referenced to points, and data collected by polygons. It was necessary to integrate these different data sets in order to build an appropriate database for the study. The tools available in the GIS were explored, and the researcher was able to reconcile the different data sets to allow the processing of the available data.

Selection of an Appropriate Data Model and Spatial Database Integration

Once all spatial information on wards, GSDs, road segments, schools with the facilities considered, and the city center were in digital format, the researcher had to confront several critical issues related to spatial database development and the need to select an appropriate data model to represent the available spatial data. The main problems were due to the fact that:

(a) The spatial data available were for different spatial units (points, lines, and areas). Moreover, several different zonal systems were used. For example, the wards and GSDs are totally different systems.

(b) The digital maps had different coordinate systems. Only the street segment map and selected services (school, health), and city center were in the UTM coordinate system.

(c) Beside the spatial units being incompatible, the ward and GSD maps showed a different city boundary.

The digital data generated had to be further processed in order to use them in the GIS environment. The first task was to choose an appropriate data model to portray the spatial data collected. There are two fundamental approaches to the representation of the spatial component of geographic information: the ***vector*** model (objects or conditions of the real world are represented by the points and lines that define their boundaries) and the ***raster*** model (location of geographic objects or conditions is defined by the row and column position of the cells they occupy). Information on different attributes is stored as "layers, images, or coverages" in the raster model. Both models have strengths and weaknesses for describing conditions in the real world (Burrough and McDonnell, 1998; Worboys, 1994; Keith, 1997 and Aronoff, 1991).

Selecting an appropriate data model for an analysis mainly depends on the purpose of the research. Considering their capabilities, both data models are appropriate for the present research. The vector data model is popular among social scientists for issues such as those that this research is examining. The maps generated by the vector approach are visually very attractive, but the present study is attempting to explore the capabilities of a raster data model with a low cost raster GIS package for finding solutions for the issues raised above.

Although there are not many factorial ecology studies based on integrated spatial data available, there is a good deal of literature on studies based on points, lines, and polygons. Methods have been developed to estimate data from one spatial unit to another through spatial data interpolation. The values derived in this way are not necessarily the true values; they are mathematical "best guesses" based on the known values (Aronoff, 1991). For example, point data can be converted to polygon based data through interpolation. Since the researcher's ultimate goal is to integrate spatial data after they were interpolated to a selected spatial unit, the available interpolation techniques were explored. But, the researcher was aware that such conversion through interpolation was not always appropriate.

A popular method of interpolation in geographical studies is the use of "Thiessen polygons" (Voronoi diagrams) to convert point data into polygons. Thiessen polygons are constructed around a set of points in such a way that the polygon boundaries are equidistant from the neighboring points (Okabe and Sugihara, 1992; Gold, 1991). The final result of this procedure is the creation of a tessellation of contiguous, space-exhaustive polygons (Boots and Getis, 1988) in which each location within a polygon is closer to its contained point than to any other point.

Geographical interest in Thiessen polygons originates with the climatologist Thiessen (1911), who assigned proximity polygons to observation sites in order to improve the estimation of precipitation averages over large areas. His method was worked out in detail by Horton (1917) who proposed the name "Thiessen polygons," crediting Thiessen with the idea. Thiessen polygons play four roles in geographical research: as models for spatial processes, as non-parametric techniques in point pattern analysis, as an organizing structure for displaying spatial data, and for calculating individual probabilities in point patterns (Franz, A., 1991). Thiessen's pioneering work has been continued by many researchers in various fields such as urban planning, marketing, ecological contact modeling, cartography, facility location and spatial interpolation (Snyder, 1962; Gambini, 1966; Boots, 1979; Mollison, 1977; McLain, 1976; Arnold and Milne, 1984; and Okabe et al., 1988).

In the present study, the major issue is not dealing with point data but with data given by line segments. Although the selected facility locations, which are given by points, are considered in the analysis to calculate distance to the nearest facility, space allocation is not performed based on their distribution.

The applications of Voronoi polygons for allocating street segment data were explored. Geographical studies on line data allocation to polygons and integration of them with other spatial data do not exist in the available literature. In computational geometry, generalized Voronoi diagrams for lines have been studied since the late 1970s by Drysdale and Lee, 1978; Drysdale, 1979; Imai and Murota, 1985; Kokubo ,1985; Fortune,1986; and Yap,1987 among others. These studies show that Voronoi polygons can be constructed based on different combinations of spatial objects (points, lines, and polygons).

The basic concept of constructing Voronoi diagrams for lines is similar to Voronoi diagrams based on points (ordinary Voronoi diagrams); therefore, a line Voronoi diagram is a generalization of the ordinary Voronoi diagram (Okabe and Sugihara, 1992) which divides the space among the spatial objects based on equidistance from each spatial object. The boundaries (edges) between the selected spatial objects are straight lines or parabolic arcs. Okabe and Sugihara (1992) argue that although curved lines may not be exactly represented by those lines, we can approximate a curved line by a chain of small straight line segments. Following the procedures described above for constructing Voronoi polygons for lines, street segment data can be converted to polygons.

One of the researcher's goals is to allocate polygon data (GSDs, wards and polygons based on street segments) only to the areas under residential land use and then to integrate them for multivariate analysis to identify the urban social structure of Colombo. Polygons in residential land use are very small (Figure 3.6). In order to allocate data from each of these polygons to residential land use polygons, an **OVERLAY** operation (described above) should be performed.

IDRISI for Windows, Version I, was selected as an appropriate GIS package for the present research. The following section titled "Appropriate GIS Software and Hardware" discusses this software in detail. IDRISI is a low-cost raster GIS software and the tools available in the program were explored to find the answers to the basic problems raised under the section "Selection of an Appropriate Data model and Spatial Database Integration." The major problems were transforming all the digital maps into a common coordinate system, building an appropriate city boundary for Colombo based on the different city boundaries, and allocating data from incompatible spatial units to a compatible spatial unit.

The selected raster GIS software has a built-in, powerful tool named *RESAMPLE* to manipulate spatial data in different coordinate systems. *RESAMPLE* registers the data in one grid system to a different grid system covering the same area. The process uses polynomial equations to establish a rubber sheet transformation, as if one of the grids were placed on rubber and warped to make it correspond to the other. The actual process is one in which a new grid is constructed and a set of polynomial equations is developed to describe the spatial mapping of data from the old grid into the new one. The new grid is then filled with data values by resampling the old grid and estimating, if necessary, the new value. Therefore, transforming all the digital maps into a common coordinate system is not very difficult.

The available maps have slightly different city boundaries. Arithmetic raster overlay operations can be applied to build an appropriate city boundary for Colombo. The part of the city area covered by all three images can be transformed as Boolean images and the *OVERLAY* function can be performed to identify the common city area covered by all images.

The researcher needed to develop new methods to convert street segment data into polygons similar to Voronoi line polygons using tools available in the raster GIS. It was found that the *DISTANCE* function in IDRISI combined with the *ALLOCATE* tool would give a solution for the problem of handling street segment data. The *DISTANCE* function calculates the true Euclidean distance (straight distance between two points) of each cell to the nearest of a set of target cells as specified. This function can be applied to calculate the distance from all the street segments.

The *ALLOCATE* function allocates the space (each cell) to the nearest of a set of designated features based on the distance. Finally, each cell in the space is allocated the identifier of the closest of the original features. For the present study, the *ALLOCATE* function can be performed to allocate space to the nearest street segment based on the distance calculated using *DISTANCE* function. The final image would have a set of raster polygons corresponding to each road segment as the space is allocated among all the road segments by *ALLOCATE* function. Then, the data referenced to each road segment can be simply assigned to the relevant raster polygon.

The next step would be assigning data given for street segments to corresponding raster polygons. The researcher proposes that socio-economic data collected should be assigned only to the areas that are under residential land use. This can be achieved by overlaying the rasterised Boolean image of residential land use upon the image converting street segments into polygons. Information is collected from the residents living on the both sides of the street segment. As the *ALLOCATE* procedure described above generates polygons, assigning areas on both sides of the street segments, it can be assumed that residential areas are allocated to the nearest road segment (residential areas are running along the streets - see Figure 3.6). The results of this procedure can be investigated by overlaying street segments and residential land use polygons over the polygons generated by applying *DISTANCE* function.

Dealing with GSD and ward data is not significantly more difficult as they are referenced to polygons. These polygons can be simply rasterised and data can be assigned to corresponding polygons. The data referenced to wards and GSD's can be assigned only to areas under residential land use by overlaying the rasterised wards and GSD images with the Boolean image of residential land use. Multivariate analysis to identify the socio-economic dimensions of Colombo can be performed on these data since they are all in one format.

Appropriate GIS Software and Hardware

Developing a spatial database model under minimum resources was a great challenge to the researcher as he had to consider many aspects of the problem. Software, hardware, and technology should be appropriate to the research as well as to the developing world (Daniel, 1987; Laksman, 1991; Yeh, 1991; Prithvish, 1987; Sundeep and Geoff, 1996; and Perera and Tateishi, 1995). There are many GIS packages in the commercial market place with various capabilities at different costs. An appropriate system for the present study was chosen based on the cost of the particular GIS package and on its capability of handling the collected data to achieve the research goals.

Capability of handling the collected data to achieve the research goals

As described in the previous sections of this chapter, the use of GIS was needed in order to construct a spatial database which is assembled from data in various formats. The following are the major tasks requiring GIS operations:

- Convert vector data generated through digitizing into raster format
- Convert different map coordinates into a standard coordinate system
- Polygonize street segments
- Build a city boundary
- Assign and classify image values
- Manipulate attribute data
- Provide an image import and export capability

GIS software packages that can handle raster data were explored to find solutions for the problems mentioned above and it was found that IDRISI for Windows Version I could provide the solutions for at a low cost. The first problem the researcher faced was converting digitized spatial data into raster format. The tool titled "The Raster/Vector Conversion" provides conversion between the two formats, vector and raster. *POINTRAS* sub menu under the tool above converts point data to raster format by assigning a new value to all cells in which a vector point is located. This new value may be the identifier of the vector point (Boolean one or zero) that simply indicates the presence of one or more points, or an identifier equal to the number of points falling in the pixel. For example, for the present study, the locations of all the public health centers would be "1"and the rest of the area as being "0". *LINERAS* converts digitized street segment data, for example 203 street segments, to raster format. All pixels intersected by a vector line are assigned the identifier of the vector line. *POLYRAS* converts polygon vector data into raster format by assigning all pixels covered by a vector polygon to the identifier of the polygon.

The coordinate systems used for digitizing the map information were different. In order to bring different spatial data into an integrated spatial database system, there should be a common coordinate system. The *RESAMPLE* option in the selected software is used to georegister an image or vector file to a reference system or to another file. In the present study, street segments, selected services, and the city center are based on UTM coordinate systems. GSDs, wards, and residential land use maps are in a user-defined coordinate system. The *RESAMPLE* option takes information -- for example, about the positions of a set of control points in the street segment file and the same positions for the wards in a user-defined system, and converts the ward file to a UTM coordinate system.

Polygonization of street segments is needed to transform street segment data into polygonized data, so that that information can be integrated with one another. It was realized that this problem can be solved in two phases calculating distances (using the *DISTANCE* tool in IDRISI) to the nearest road segment and allocating (applying *ALLOCATE*) the raster cells to the nearest road segments as described in a previous section.

In the selected software, the *OVERLAY* function allows the user to construct a common boundary based on different city boundaries derived from GSD and ward boundaries. The boundary of the common area created by overlaying all maps can be considered a solution to the Colombo city boundary problem. The last three tasks could be achieved by the basic GIS functions and by tools such as DATABASE WORKSHOP which can handle data manipulation tasks.

Besides the capabilities discussed above, the price of the GIS package was considered as well. As the study is addressed to developing world social scientists, planners, and government institutions, this feature was very important. The government institutions cannot afford high priced GIS software such as ArcInfo, TransCAD, TRMapper and ArcView. IDRISI for Windows Version I is a geographic information and image processing software system developed by the Graduate School of Geography at Clark University, Worcester, USA. It is designed to provide professional-level geographic research tools on a low-cost non-profit basis.

SPSS (Statistical Package for Social Sciences) for Windows 7.0 Graduate Pack Version was used for statistical analysis such as factor, cluster and multiple regression analysis. For creating charts and tables, Microsoft Excel 97 was used, and for image editing, drawing and vector file editing, Corel Photo-Paint and Corel Draw Version 7 were very helpful. SPSS is a comprehensive and flexible statistical analysis and data management system. It can take data from almost any type of file and use them to generate tabulated reports, charts and plots of distribution and trends, descriptive statistics, and complex statistical analyses. Corel Draw and Corel Photo-Paint programs were excellent in terms of "tuning up" the GIS image and vector outputs.

Georegistration

One of the critical problems related to the digital data was that the spatial information on selected spatial units was in different coordinate systems. Only the street segments were accurately registered based on UTM coordinates. In order to analyze data in a GIS environment, all the spatial information had to be in a standard coordinate system. It was decided to register all the spatial information to the street segment information making minimal displacements. The registration was achieved through the use of a systematic *rubber sheet* transformation process that gently warps an image (through the use of polynomial equations) to correspond to the known positions of a set of widely dispersed control points. This capability is provided in IDRISI through the module *RESAMPLE.*

The entire procedure of spatial database development is demonstrated in a cartographic model in Figure 4.1. A cartographic model is a graphic representation of the data and analytical procedures used in a study. Its purpose is to help the user organize and structure the procedures that will be performed in a study and to identify the data needed to complete them. It also serves as a source of documentation and reference for the analysis (Clark University, 1997). As shown in Figure 4.1, the final product of the present study is an integrated spatial database. The symbology conventions proposed by the IDRISI manual (Figure 4.2) were adopted for the research. Image files are represented by rectangles, vector files by triangles, values files by ovals, and tabular data by a page with a corner turned down. Individual fields in a database are identified with the field name underlined. File names are all written in uppercase, inside the symbol. Operations are indicated by writing the module names above or by arrows connecting input and output data files. Module names are written in uppercase and italicized. When an operation requires the input of two files, those two files are joined with lines, then a single arrow

Figure 4.1 Cont.

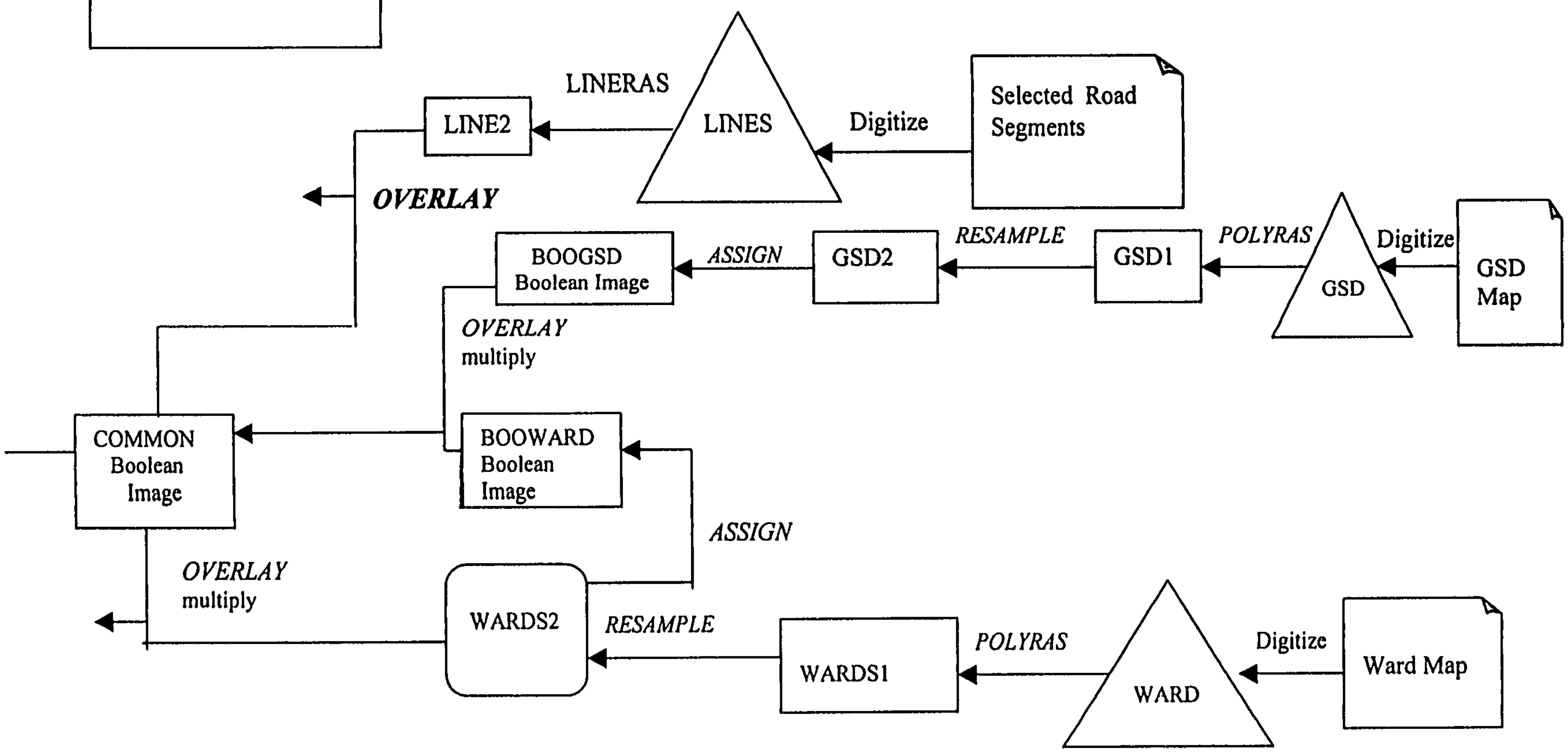


FIGURE 4.1 – CARTOGRAPHIC MODEL OF SPATIAL DATABASE DEVELOPMENT

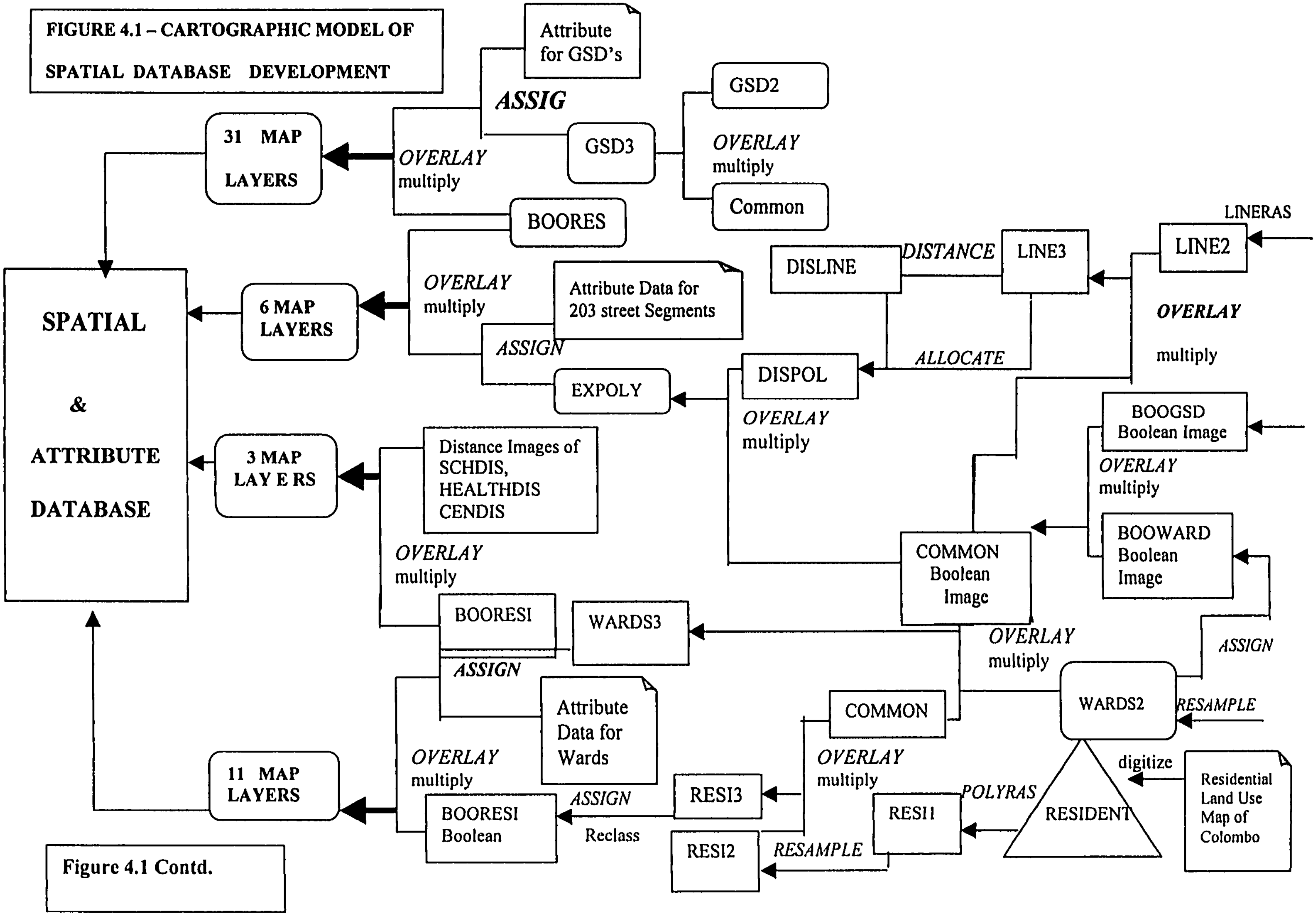


Figure 4.1 Contd.

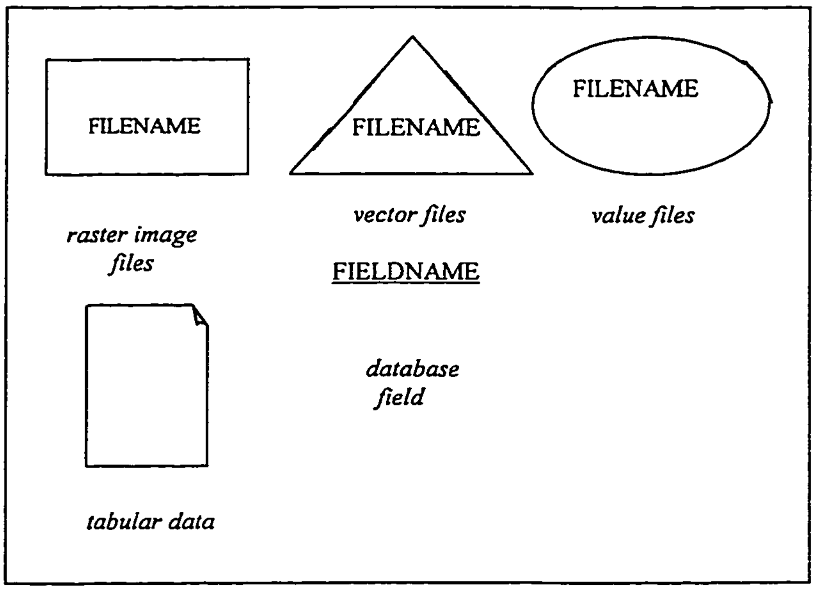
points to the results. In those cases, the module name is written between the input files. Important parameters may be noted in the cartographic model near their module names in lowercase italics.

Georegistration was done in two phases as follows:

a) Converting Digitized Files to Raster Image Files

The resolution of the raster cells was determined from the spatial distribution of the city. The city can be masked by a 6 km x 3 km rectangle creating 30m x 30m pixels. This pixel size is a reasonable size for the present study, as the data are not given for a smaller than this format.

Figure 4.2 Symbols used in the Cartographic Model



Digitized administrative and residential land use polygons were converted to image files, using *POLYRAS* function while *LINERAS* function was applied to generate an image file for street segments. The last step was to convert the rest of the point files into raster format. They were converted and named as HEALTH1 (Location of Public Health Centers), SCHOOL1 (Location of Public Schools with Selected Facilities) and CITY1 (City Center).

b) Resampling

The second phase of Georegistration involves ***Resampling***, which is a procedure for spatially georeferencing an image to its known positions on the ground. The procedure used to accomplish this correction is also termed "**registration**" (Aronoff, 1991, Burrough and McDonnell, 1998). Map layers for the city must all be registered so that the same location in each overlay has the same map coordinates. From the two approaches available for registering map layers, the registration by relative location technique was chosen considering the nature of the maps available. In this procedure, one map layer, termed the "**slave**", is registered to a second map layer, termed the "**master**" (Environmental Systems Research Institute, Inc., 1995). The relative position of the "**slave**" data layer is adjusted by choosing features that can be easily and precisely identified on the two data layers to be registered.

For the research, the street segment map layer is treated as the "**master**" layer and all the other map layers are considered to be "**slave**" layers. The street map layer is the most accurate and the most suitable "**master**" layer as it contains road intersections and comes with UTM coordinates. The selected road intersections, as test points for map registration from the master map layer, could be easily identified in the other map layers. Both administrative maps (GSD and ward) mainly follow the road network for their boundaries. The public health centers, the schools with selected facilities, and the city center images/ map layers are already georeferenced using the same UTM coordinate system as the street segment map layer. The last slave map layer/image is the Colombo residential land use map. It was not difficult to find similar reference points from this slave and master layer because residential land use distribution is based on the street network.

Initially, it was decided to utilize four test points from each map layer to assess the variability of the measurements against their true values. This was one of the most cumbersome processes of the research. There needed to be a balance between choosing the number of test points to be selected and selecting the locations of these test points in order to achieve a small value for the root mean square error (RMS). It is obvious that locations of sample points selected from slave layers are different from the street segment image. The locations of test points of the slave layer need to be compared with their locations on the street segment map. The RMS error is extremely important for the *RESAMPLING* process as the lower the error term, the better the transformation into the new coordinate system.

The allowable amount of RMS error depends mainly on the slave map scale. The errors given by IDRISI were compared with the revised version of the US National Map Accuracy Standard proposed by a committee of the American Society of Photogrammetry and Remote Sensing (Merchant, 1987) which specifies acceptable root mean square error terms for horizontal locations for various maps.

The residential land use map gave the maximum RMS error; therefore, 16 control points were selected for that map. Table 4.1 shows the information on the error statistics obtained and the allowable RMS error at a given map scale. Since neither GSD nor ward maps indicated scale, appropriate map scales were estimated by comparing the control points selected with their location on the street segment map.

For both administrative maps, three additional test points were employed to gain a small RMS value. All the test points were obtained from a good distribution of control

Table 4.1 Error Statistics

Map	Scale	Allowable RMS Error	Calculated RMS Error	Number of Control Points Selected
Ward	1 : 30,000	5.0	3.4567	7
GSDs	1 : 29,000	5.0	2.7889	7
Residential Land Use	1 : 11,250	2.5	2.3345	16

points. The initial RMS value for the residential land use image was as high as 3.48. Therefore, 16 test points were used to reduce the error for that image. Image *RESAMPLING* procedures were applied to obtain estimated UTM coordinates for all images after obtaining appropriate error terms.

Building a Common City Boundary

The constructed images featured varying city boundaries. Due to mapping, digitizing, and transformation errors, the administrative maps did not display a precise common boundary for the city. The next task was to build an appropriate city boundary for the study area. This was accomplished in the following manner.

Both administrative images, wards and GSDs, were transformed into Boolean images (Areas within the city boundary were coded as "1" and the rest of the area as "0") exercising the ***ASSIGN*** tool in IDRISI. This simple tool can be applied to solve a critical problem in dealing with the incompatible data sets. The GSD2 and WARD2 images were transformed into Boolean images and labeled as BOOGSD and BOOWARD respectively (Figure4.7). Simply applying the *OVERLAY* operation with option "*multiply*", BOOGSD and BOOWARD images were overlain on each other. When both images are Boolean, various logic operations (Boolean algebra) can be undertaken. "*Multiply*" operation is simply a multiplication between the selected Boolean images. Only the area common to both images can be extracted, where the condition "1" is met in both images, by multiplying the two images. In addition to the area common to both images, a "sliver"/ "splinter" running along the city limit was shown on the overlain image (Figure 4.3). This "sliver" is visible only on some parts of the city limit. The area of the sliver and the area common to both images were calculated and compared with the total area under the city limit given in the Report on the Survey of Squatter Settlements in the City of Colombo (Colombo Municipal Council, 1990).

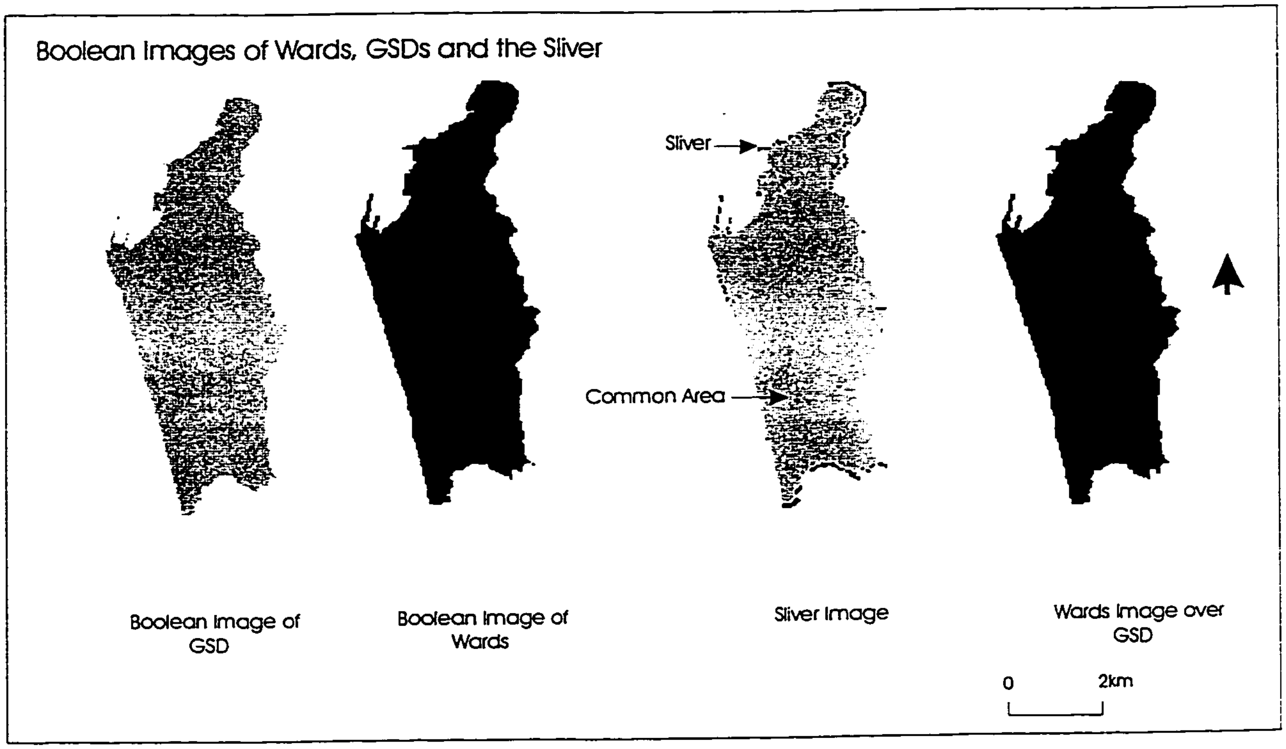
According to the report, the city covers at present an area of 3731.28 hectares. The overlapping area generated only 3661.51 hectares and it was decided to treat the common area as city area for the following reasons. The total area under the common region is 98.1 % of the total area given in the report. In effect, the study loses only 1.86 % of the given total area. The reason for this 'loss' may be attributed to the fact that:

- The figures presented in the report may not be precise.
- The areas under the dockyard and the construction site related to the Colombo harbor had been excluded in both administrative maps. Therefore, discrepancy may have arisen between the figures.

Trimming the Boundaries of all the Images to Fit into the New City Boundary

To integrate spatial databases and to perform GIS operations, all the images should be under a common boundary. It was decided to bring all of the images under the boundary of the **COMMON** image. This was done by overlaying all images with the COMMON image. This procedure created one single boundary for all the images. The

Figure 4.3 Boolean Images of Wards, GSDs and the Sliver



area under the original ward and GSD boundary were compared with the area under the new city boundary to investigate whether there were large gaps. Figure 4.3 shows two different combinations of overlays, GSD over wards and wards over GSD. The GSD boundary is slightly smaller than the ward boundary from north and south of the city boundary. Other than this, both boundaries correspond very well. The north boundary of the city is the Kelani river and the south boundary is a water channel. It is not uncommon to see slight changes to the city boundary when it involves a river or a channel. These small differences are not very significant.

Dealing with Street Segments:

The rasterized street image had to be further processed, to integrate the spatial database given at street level with other spatial databases. In order to achieve this goal, the following procedures were taken to build polygons based on the rasterised street segments:

a) Pixels of the street segment image were assigned to each road segment based on the Euclidean distance, that is, the straight distance from the closest street segment. The DISLINE image (Figure 4.4) was created applying the ***DISTANCE*** function. The image displays the distance from the nearest street segment.

b) The next step was to assign each pixel in the image to the nearest street segment to generate polygons representing the nearest street segment. The ***ALLOCATE*** tool was employed to assign each cell to the nearest street segment. Each cell of the image was associated with one of the street segments of the original image from which the distance was calculated and the new image was named as DISPOLY (Figure 4.5).

c) The last step was to limit pixels only to areas under the new city limit. This was done simply by overlaying the DISLINE image on the COMMON image with a sub-operation *multiply*.

The final product was named EXPOLY (Figure 4.6). These are the experimental polygons in raster format constructed on the street segments and to which the data from street segments are assigned. The accuracy of building polygons based on the street segments and allocating residential polygons to the nearest road segment were explored by overlaying street and residential polygons over the EXPOLY polygons. It was evident that residential polygons were assigned to the nearest road segments and road segments are located within the polygons in the EXPOLY image.

Aggregation always gives some error. Therefore, this estimation procedure is an approximation to the reality. Some of the residential polygons, for which the data are referenced, may have not been assigned to the precise road segment.

Figure 4.4 DISLINE Image
(Linear Distance from Street Segments to Pixels)

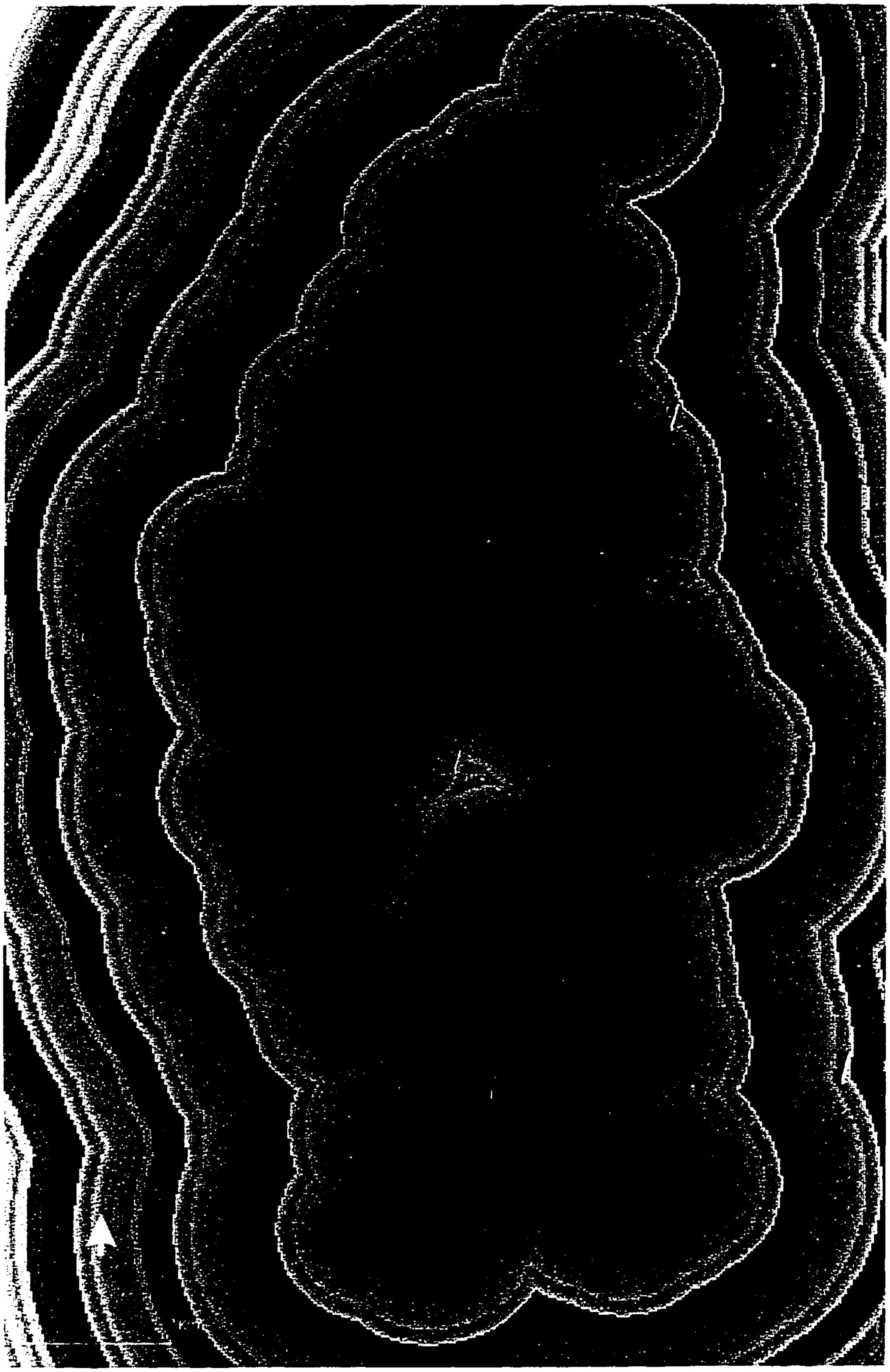


Figure 4.5 DISPOLY Image (Allocating Pixels to the Nearest Street Segment)

Figure 4.6 EXPOLY Image

Experimental Polygons in Raster Format Constructed on the Street Segments

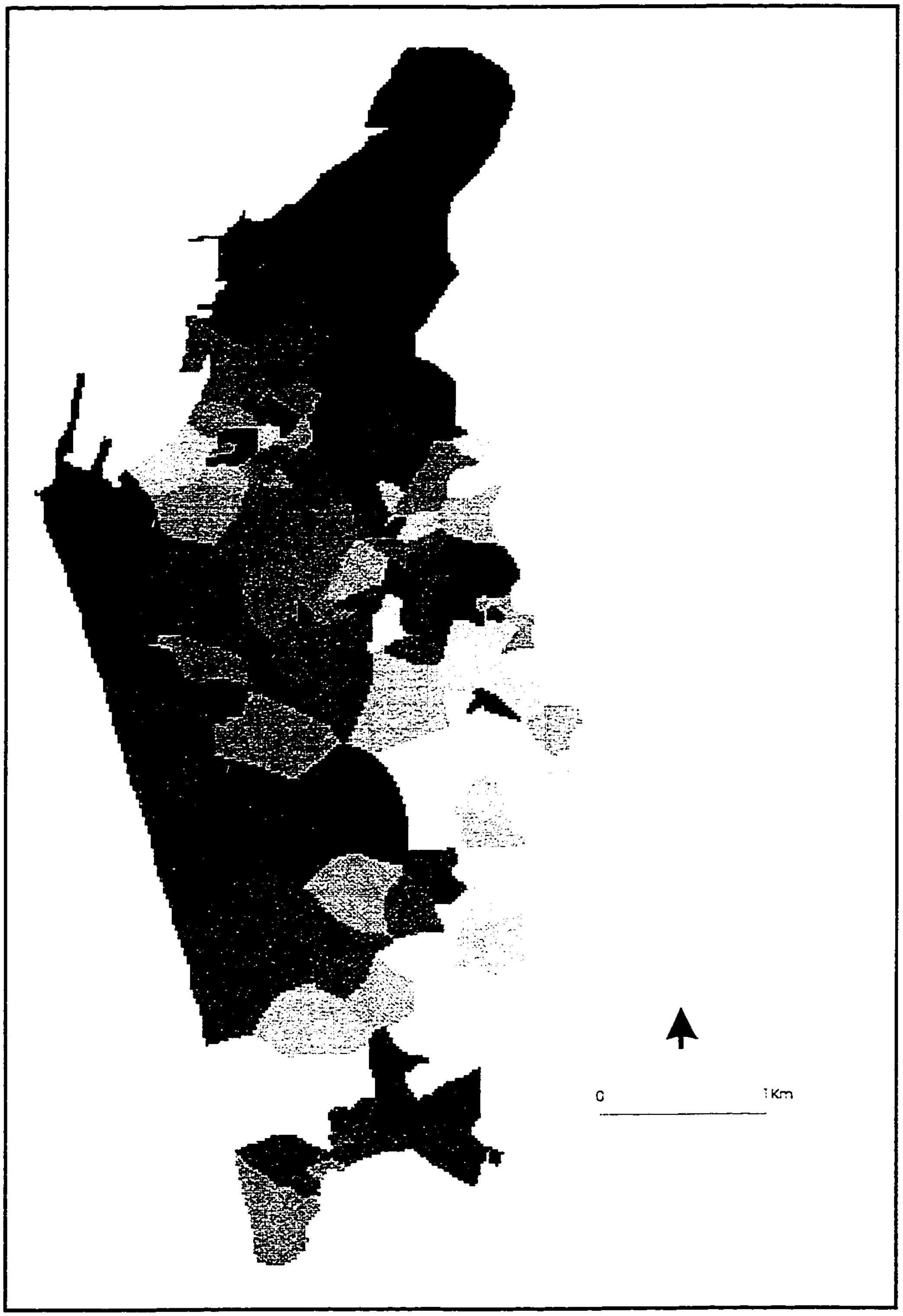


Figure 4.7 SCHDIS Image
(Linear Distance to the Nearest School with Selected Facilities)

Figure 4.8 HEALTHDIS Image
(Linear Distance to the Nearest Public Health Center)

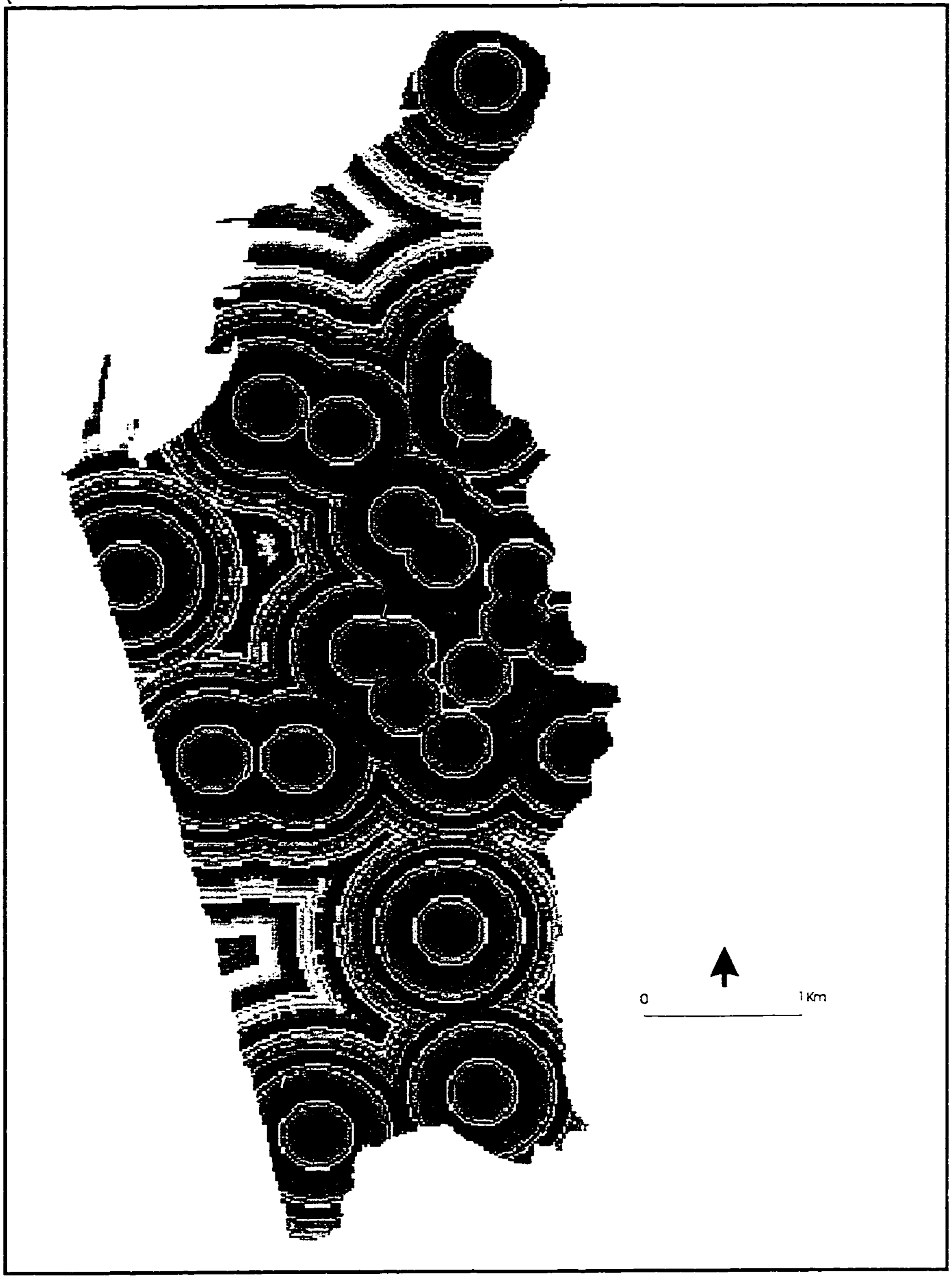


Figure 4.9 CENDIS

Generating Distance Measurement:

The distance measurements described above were calculated by applying the following procedures. The rasterised image of the locations of public schools with Science and Mathematics for Grade 13 (Figure 3.5) was used to calculate the Euclidean distances to the nearest such high school. The distance to the nearest school for each pixel was extracted using the **DISTANCE** function and the new image named ***SCHDIS*** (Figure 4.7). Utilizing the same procedure on rasterised images of locations of public health facilities and the location of city center, images ***HEALTHDIS*** (Figure 4.8), and ***CENDIS*** (Figure 4.9) were established, generating the required distances.

It was decided to build three separate attribute database files for wards, GSDs, and street segment data and to generate a map layer (theme) for each variable selected from various sources as described above. The database management system, together with the selected GIS package, was used to capture attribute data. This process involves a simple tabular data file storing system which relates the standard attributes of each geographic feature in the map layer to a corresponding record in the feature attribute table.

Capturing and Assigning Attribute Data to Images:

The feature attribute data capturing process consists of the following major steps:

- Creating a new data file to hold the attributes
- Adding the attribute values to the data file
- Linking the data file to the feature attribute table for the map layer

After the digitized maps are imported to IDRISI, the next step is to create a new data file to hold attribute information. For complicated data manipulations such as sampling and integrating different attribute databases, the Microsoft ACCESS 97 database system was utilized. Except for these complicated data manipulations that occurred on a few occasions, the Database Management System (DBMS) in IDRISI was used to store and manage information.

The first column of the database is reserved for feature ID which represents a code number for the spatial unit. Three databases were constructed to hold information for wards (eleven variables), GSD (thirty-one variables) and street segments (six variables).

The original values were transformed to percentages, ratios, densities, and so on using the database tools. The final step was to assign these attribute values to the images of wards, GSD's and street segments. To move data from the database to an image layer, one can use the *ASSIGN* option of the Database Workshop Link menu which creates new images by linking the geography of features defined in an image file with attributes defined in an attribute values file. Each attribute from the three attribute databases was selected and assigned to a separate map layer. Fifty-one map layers were produced based on the variables described in Chapter 2 (see Appendix 2). These map layers needed to be further processed in order to allocate information only to the areas falling under residential land use. This was done in several steps.

First, the residential land use map was transformed into a Boolean image giving a value of "1" for all the residential land use polygons and "0" for the rest (Figure 4.10).

Figure 4.10 **Overlay Operation with Multiply Option**

0	*23*	*23*	*23*	*23*
0	*0*	*23*	*23*	*23*
0	*0*	*0*	*23*	*23*
11	*11*	*0*	*0*	*23*
11	*11*	*0*	*0*	*0*

Boolean Image of Residential Land Use

0	*23*	*23*	*23*	*23*
0	*0*	*23*	*23*	*23*
0	*0*	*0*	*23*	*23*
11	*11*	*0*	*0*	*23*
11	*11*	*0*	*0*	*0*

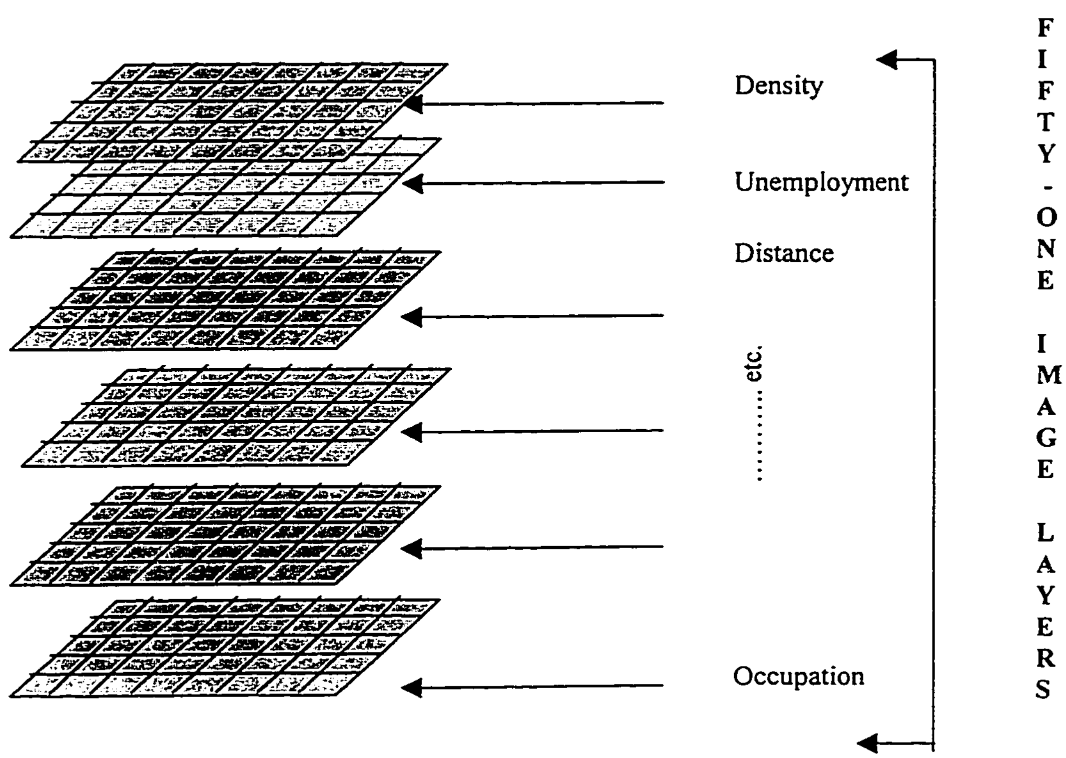
Percentage of Unemployment Distribution

0	*0*	*0*	*23*	*23*
0	*0*	*0*	*23*	*23*
0	*0*	*0*	*0*	*0*
11	*0*	*0*	*0*	*23*
0	*11*	*0*	*0*	*0*

New Image

The second step was to *OVERLAY* each map layer (fifty-one map layers) described above with the Boolean image of residential land use with the *multiply* option. The selected map layer is "Percentage of Unemployment Distribution". The data from this layer were extracted to a new layer indicating the "Percentage of Unemployment Distribution" only for the residential areas. In order to allocate the data from the selected layer to the new layer, the Boolean image is overlain with the residential land use layer with the *multiply* option. Multiplication has taken place by multiplying values of pixels of an image with the corresponding pixel values of the other image. The resulting new layer is labeled " New Image". The same overlaying procedure was performed for all the map layers in order to allocate values only to pixels that represent residential land use. The final product of this procedure was a number of new map layers with positive values assigned to pixels under residential land use. As shown in Figure 4.11, the procedure described

Figure 4.11 Compatible Data Layers



above could bring the data from different sources into a compatible format on which further statistical analysis could be performed. Each pixel of an image has a corresponding value from different layers. These map layers can be converted and exported to any statistical or database program for more complex statistical calculations to reveal the urban social structure of Colombo.

The size of each image was 600 X 300 pixels. Each pixel represents a 30m X 30m area. Therefore, there were 180,000 pixels in each image layer, out of which 43,281 fell within the city boundary. Of these, there were 14,754 pixels under residential land use.

The image layers were further processed in order to reveal the urban social structure of Colombo, extracting data from the layers. The researcher assumes the layers created do not reflect the exact area they are indicating. It follows that since the socio-economic data were collected from the people, this data should only be allocated to the residential areas. This goal was achieved in the following manner.

First, the rasterised residential land use map was converted to a Boolean image following the same procedures that was used to create Boolean images of wards, and GSDs described in detail above. Secondly, all the map layers were overlain with a Boolean image of residential land use to assign values to the pixels under residential land use. This procedure produced 51 map layers with positive values assigned only to pixels under residential land use, thereby generating a compatible data set for further analysis. Although there were 180,000 observation units (pixels) in one image, only 14,754 observation units, which fell under residential land use, were used for the statistical analysis.

The researcher proposes that not all of the pixels under the residential land use should be used to reveal the social structure of Colombo. If all were included in the analysis, the results would be greatly affected by high spatial autocorrelation, since all the residential land use pixels, under a given zone (i.e. Ward, GSD), each would carry the same values on each variable; and thus, neighboring pixels would display similar values. The reader should be aware that the original data set may have had some spatial autocorrelation to a certain degree since some data were referenced to zones.

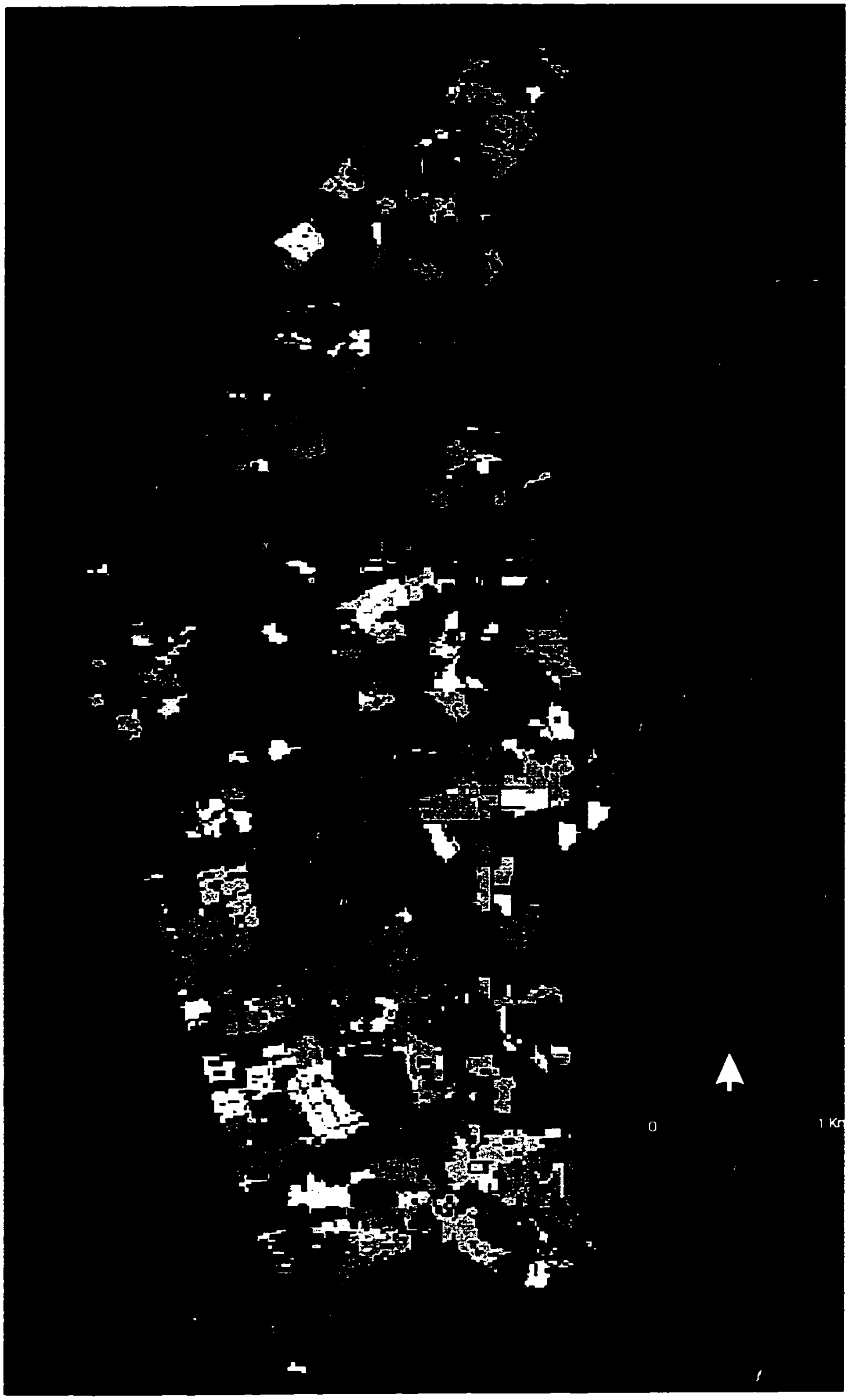
For the present study, therefore, it was decided to select one pixel from each group of residential land use pixels, in order to minimize autocorrelation. This was achieved, generating **"UNIQUE ZONES,"** by applying Raster GIS operations on all the data layers overlaying each other and taking one pixel from each to represent a unique spatial combination. Since only one pixel is selected from each zone, it is expected that autocorrelation is greatly reduced in the statistical analysis. The generation of **"UNIQUE ZONES"** and the preparation of data set are described as follows.

First, the information from all the map layers (fifty-one) was brought into a single map layer overlaying, each one with ADDITION arithmetic operation in raster GIS. The researcher assumes each map layer is unique in terms of the values under the corresponding pixel since they were calculated to the sixth decimal place.

The second step was to generate a new set of zones / homogeneous group of pixels based on the new map layer that was generated by combining values from all the map layers. Generating homogeneous groups of pixels was done with the operation called *GROUP* in IDRISI. *GROUP* determines contiguous groupings of identically valued integer cells in an image. A total of 1195 unique zones/groups were generated by this procedure and the final image was named as **GROUP** (Figure 4.12).

The third step was to select one pixel to represent each zone / group. Because only one value was selected to represent each group, each group had equal representation in the analysis. This procedure is similar to applying statistical techniques on area based data such as census tracts. Microsoft Access Database was used for this stage of data preparation. The values of the GROUP image and fifty-one image layers were exported to Microsoft Access Database as two separate database files for further calculations. The two database files exported to Access database were linked with the identifier of each pixel and joined into one database file which had 52 fields (including zone number) and 180,000 cases. The "Query" function in the database was applied to extract only one record containing a group's measurement on all the variables. This query analysis created a new database with 1195 observations (one for each group) and 52 fields. This file was exported to SPSS for statistical analysis to reveal the social structure of Colombo.

Figure 4.12 Group Image
Generating Unique Zones Based on the Fifty-One Map Layers



In conclusion, by applying GIS techniques and database tools, the researcher was able to develop a data set to study the complexity of the social structure of Colombo. The following chapter will exhibit the findings of the multivariate analysis performed on the integrated spatial database of Colombo to reveal the urban social structure of the city.

CHAPTER FIVE

Urban Social Structure of Colombo

Introduction

As described in the previous chapter, a large data set with 1195 records was developed from various sources of information by applying GIS techniques and database tools. The purpose of this chapter is to describe the urban social structure of Colombo by applying factor analysis on the data set described in Chapter Four. Factor analysis reduces a set of variables into a smaller set. This technique is based on the assumption that some underlying dimensions/factors/structures are responsible for the covariation among the observed variables, and that the factors can, in turn, be used to explain complex phenomena (Bailey and Gatrell, 1995). The major sources of social variation and their distribution pattern in Colombo will be discussed in detail. The factor analysis results were exported back to IDRISI and maps were developed to display the spatial patterns.

Factorial Ecology of Colombo: Dimensions of Variations

Data

The data preparation query created a database with 1195 records and fifty-one fields. Each record represents one of 1195 groups. There are no duplicates for any group, so that, the new data set which was generated following the procedures described in Chapter Four is similar to the original data set that the researcher initially collected.

Analysis Procedure

Factor analysis methods can be used in either hypothesis testing or in an exploratory context by using either the common factor or component models. Most researchers have used these techniques to explore the sources of the variation in the original data set. Given the objective of the present research, factor analysis was performed in an exploratory context. The principal component factoring technique was applied to data followed by Varimax rotations using the 'R mode' procedure in which the similarities between the columns or variables of the matrix are calculated using the SPSS package (SPSS Inc., 1996).

These technical procedures were chosen from a variety of alternative methods reported in previous work (Yeh et al., 1995; Davies, 1984; Davies and Murdie, 1991; Lo, 1986). The Varimax rotation procedure is a technique designed to increase the simplicity of the factor structure and to enhance the interpretability of the factors themselves (Girden, 1996 and O'Brien, 1992). This is a common procedure that many researchers have followed in previous studies.

Davies (1983) says that factor methods need not be used in a strictly inferential context; they are sufficiently comprehensive and stable enough to be used for synthesizing data sets without reference to the inferential tests. Indeed, Rummel (1971) and Mather (1976) have referred to these methods as the "calculus" of social science and believed that researchers should employ less rigorous techniques in order to better understand the sources of social variation.

The Choice of Indicator Variables

It was decided to extract a simplified version of the social structure of Colombo based on a key diagnostic set of variables. All but the distance variables were run through an exploratory factor analysis to extract the best set of variables (variable filtering procedure). This procedure is highly recommended by many researchers. For example, Davies's (1983) study on Cardiff and its region used 'expected' factors and areas as guides for the choice of variables and areas.

Applying filtering eliminates some of the original information. But, this procedure was needed in order to eliminate redundant information emanating from different data sources and to achieve a meaningful set of social dimensions. It is important to have a simplified, meaningful version of social dimensions based on a small number of key or diagnostic indicators rather than a complicated set of factors, based on a large set of attributes, which cannot be described meaningfully (Mather, 1993).

First of all, distance variables, namely, *Distance to the City Center*, *Distance to the Nearest Public Health Center*, and *Distance to the Nearest Public School with Science and Mathematics for Advanced Level*, were reserved to test their relationship with the factors that are to be extracted. Very few past studies have included distance variables in factor analysis; a few notable exceptions are Brand, 1972 and Bourne and Murdie 1972.

Variables used in the analysis provide a representative cross-section of information under the major headings of population, age distribution, ethnic origin, dwelling characteristics, employment structure, education, length of residence, and facilities. Variable selection, the nature of variables and their sources were discussed in detail in Chapter Three. All the variables used in the analysis are shown in the Appendix 2.

Redundant (duplicate) information contained in data can enter through the inclusion of multiple variables (or attributes) in an analysis, whereby the attributes themselves are repetitious, at least to some degree. Such redundant information causes multi-collinearity to be present in the correlation matrix (the multivariate problem is concerned with avoiding double counting of any repeated attribute information). Tabachnick and Fidell (1996) show the importance of screening data and offer a checklist as a guide to evaluation of data prior to analysis.

The overall strength and direction of relationships among the selected variables in the 1195 selected pixels have been determined by means of correlation analysis (Pearson's Product Moment Correlation Coefficients – Appendix 3). The variables with correlation coefficients higher than 0.9 are considered to be redundant (Griffith and Amrhein, 1997, Goddard and Kirby, 1976).

The variables with high correlation coefficients were examined carefully before they were excluded from the final analysis. Exploratory factor analysis was used as a tool to determine which variable should be retained from a pair of highly correlated variables. The program was run repeatedly excluding one variable at a time and the changing pattern of factor loadings was examined. The variable (from a pair of variables with high correlation) which contributed higher factor loadings and made factors more meaningful was retained. Six variables, *Number of student desks per 10,000 students (DESKPER), Percentage of employment in high income group (engineers, doctors, teachers, clerks, technicians) (EMPH), Number of unemployed people with 8th grade standard education per1, 000 unemployed people (male and female) (EIGHTUNM), Shanty population density per hectare (SSPOPDN), Percentage of population under 6 years (LESSIX) and Percentage of population less than 5 years of age (LESSPIVE)* were eliminated due to redundancy (see the Appendix 3 for more details).

Beside the variables mentioned above, another group of variables was dropped from the analysis for a variety of reasons including the fact that they did not load on expected axes making loadings meaningful - some variables had high loadings on more than one factor. And some had very low factor loadings (loadings lower than 0.50 are considered to be low, Davies, 1983). White (1987) points out that identifying key indicators for underlying constructs based on exploratory factor analysis is one useful by-product of factor analysis. The variable filtering / data 'grooming' process was performed in two stages. The results of these stages and the explanation for eliminating each variable at given exploratory stages are shown in Appendices 4 and 5.

Analysis

The final factor analysis was performed on a set of 24 variables (Appendix 6), filtered through the data 'grooming' procedures. Determining the "appropriate number of dimensions / axes / factors" is one of the most troublesome problems in factor analysis. A wide range of very different procedures, ranging from "rule of thumb" guidelines to statistical tests for the number of common factors, has been developed to deal with this problem. The cut-off point for determining the number of factors was selected using the criterion that all components contributing Eigenvalues less than 1.0 should be excluded because they would have too large a random-error variance. The Eigenvalue 1.0 is recognized as a nontrivial, minimum amount of variance (Griffith and Amrhein, 1997; Girden, 1996 and Goddard and Kirby, 1976). There were six factors conforming to the selected criterion.

The "Scree diagram" (Appendix 6), a two-dimensional plot of the Eigenvalues (variance accounted for by each factor) versus factors, was used to confirm the validity of the 'rule of thumb' described above. In a "Scree diagram," the X-axis represents each factor extracted, while Y shows the Eigenvalue of each factor. The curve shows a distinct elbow in the slope of the distribution of Eigenvalues at factor 6 (six-factor solution). Therefore, the appropriate number of factors / dimensions for the present study was determined as six. The six-factor solution accounted for 75.1 percent (Table 5.1) of the initial variance.

Furthermore, the overall *communality*, i.e. the proportion of a variable accounted for by the common factors in a factor analysis of all the variables used in the analysis was very high. Seventeen of 24 variables had a *communality* over 73.5 percent, an indication that the factor solution is strong. In addition, the factors extracted are qualitatively more meaningful in terms of the variables loaded to each factor.

Cronbach's alpha coefficient of reliability was used to determine the relative homogeneity of variables for the entire scale and within each factor. Test scores range from zero to unity; a score of 0.40 or greater is generally treated as acceptable level of reliability. The alpha score for the entire scale was 0.84 and the scores for Factors 1 through 6 were 0.82, 0.78, 0.64, 0.62, 056, and 0.52, respectively. Therefore, the entire scale and the six factors demonstrate a strong degree of internal consistency.

Table 5.1 shows Eigenvalues, the standard variance associated with a particular factor and the cumulative variance accounted for by cumulative dimensions. It tells us

Table 5.1 Eigenvalues and the Variance Accounted for by the Six Dimensions

Factor	Eigenvalue	Variance accounted by each Factor (%)	Cumulative Variance Accounted for (%)
I	7.0433	29.3	29.3
II	3.7972	15.8	45.2
III	2.3109	9.6	54.8
IV	2.1466	8.9	63.7
V	1.6773	7.0	70.7
VI	1.0545	4.4	75.1

that the mathematical solution is powerful since each factor had a high Eigenvalue. Even the sixth factor accounted for 4.4 % of the variance, indicating the strength of the solution. However, statistics do not tell us whether a solution or individual factors are particularly useful or interpretable. Do they constitute a logical set? Do the variables separate sufficiently into different factors? Can we identify and name the "hidden" dimensions or vectors that underlie each set of variables? For this we must look to the differential projections of the variables on each factor.

Table 5.2 shows the differential projections of the variables upon each of the six orthogonal factors. A simple descriptive title has been given to each factor to summarize the pattern of loadings. Factor loading represents the weight attached to each factor, in other words, the amount of variability in a variable that is associated with a particular factor.

The "loadings" of each variable indicate the association of the variable to the factor / vector / dimension. The results show that practically all of the variables used in the analysis load on to one of the factors with loadings higher than 0.5. This implies that the variable set used in this study could not be further reduced in number without serious effects upon the quality of the results. The data screening procedure left 18 new variables and six variables that are employed in the traditional factorial analyses.

Social Dimensions and Distribution Patterns

FACTOR 1: Quality of Unemployed Labor Force

The academic or technical background of unemployed people turned out to be the most important dimension, explaining 29.3 percent of the total variance. The factor loadings of all of the variables loaded to this dimension are higher than 0.93; therefore, the factor is very strong in terms of the explanatory power of the variables loaded to the factor. Moreover, variables loaded to the factor show a clear consistency (Cronbach's alpha: 0.82) since all the variables show the academic or technical background of unemployed people. The **Quality of Unemployed Labor Force** shows the educational and technical level achieved by the unused labor force. Previous studies in developing world cities do not identify a specific structure like the one above because most of the studies available are not based on detailed information of the unemployed. Unfortunately, some census information sources do not have detailed indicators on the specific characteristics of unemployment. Studies that were conducted by Davies and Murdie, 1991, Yeh et al., 1995, and Hamm et al., 1988 used male unemployment as an indicator and loaded it under characteristics of employment or under a general socio-economic dimension.

Table 5.2 Loadings of Variables on the Six Factor Solution

Quality of Unemployed Labor Force	**FACTOR 1**	**Poverty**	**FACTOR 2**	**Ethnicity**	**FACTOR 3**
Short Title of Variable	Loadings	Short Title of Variable	Loadings	Short Title of Variable	Loadings
Degree Qualification	0.96139	Shanty Population	0.78405	Percentage of Sinhalese	-0.81868
Ordinary Level (Grade 10)	0.93507	Blue collar employment	0.71278	Renovated buildings	0.78765
Advanced Level (Grade 12)	0.93155	Population under five years	0.67612	Unemployment ratio	0.59919
Technical Qualification	0.93056	Good houses	-0.55557	Moor population	0.57043
		Male unemployment	0. 50888		
Cronbach's Alpha	0.82	Cronbach's Alpha	0.78	Cronbach's Alpha	0.64

Educational Facilities	**FACTOR 4**	**Nature of Employment**	**FACTOR 5**	**Foreign Employment**	**FACTOR 6**
Short Title of Variable	Loadings	Male employment (seasonal)	0.86124	Male foreign employment	0.87571
Schools per 10,000 students	0.79844	Male employment (permanent)	-0.77371	Female foreign employment	0.69872
English teachers	0.72551	Female employment (seasonal)	0.73638		
Chairs per 10,000 students	0.72136				
Building space per student	0.69309				
Number of students per teacher	0.64532				
Specialized teachers	0.52425				
Cronbach's Alpha	0.62	Cronbach's Alpha	0.56	Cronbach's Alpha	0.52

Factor scores were estimated using score coefficients (a conventional method) to identify the spatial distribution of the dimensions. The factor scores of each observation on **Quality of Unemployed Labor Force** were mapped (Figure 5.1). High positive factor scores indicate a high level of unemployed labor possessing advanced levels of academic or technical training while the high negative factor scores show the areas with low levels of unemployed people with high academic or technical qualifications. Three main concentrations of high levels of unemployed people with advanced the qualifications considered are apparent. They are: Cinnamon Garden, Borella North, and Maligakanda region (Northeastern sector of the city); Mahawatta, Modara and Aluthmawatha region (East section of the city); and Wellawatta south region (Southern part of the city). The lowest rate of unemployment manifests around Thimbirigasyaya, (South section of the city), which is mainly an affluent area.

FACTOR 2: Poverty

This is the second most important social structure extracted from the data set, explaining 15.8 % of the total variance. **FACTOR 2** shows a specialized socio-economic profile, labeled **Poverty.** The variables that loaded on to the factor show strong consistency (Cronbach's alpha: 0.78), and the loadings are high, indicating the importance of their contribution to the factor. All the variables loaded to this factor are clear indicators of poverty. It has been shown in previous studies that poor families in Colombo, Sri Lanka, have a fairly high young dependency ratio (Arachchige-Don, 1994). Therefore, this analysis is consistent with the results of previous studies since the variable *Percentage of population under five years of age (POPFIV)* was loaded with a high weight (0.6761).

Furthermore, the high positive loading of *Percentage of slums and shanties (SHANPC)* is to be expected given the general relationship between poverty and housing structure. This relationship is further confirmed with the high negative loading (-0.73254) between *Percentage of good housing units (GHOUSE)* and the **Poverty** dimension. It is obvious that this variable should be lower in the areas where slums and shanties are dominant since the affluent people would prefer to move away from these areas.

Figure 5.2 shows the spatial distribution of factor scores on **Poverty** dimension. The class interval of the map equals one standard deviation of the factor scores. The

Figure 5.1 Distribution of Factor Scores on the Unemployment Dimension

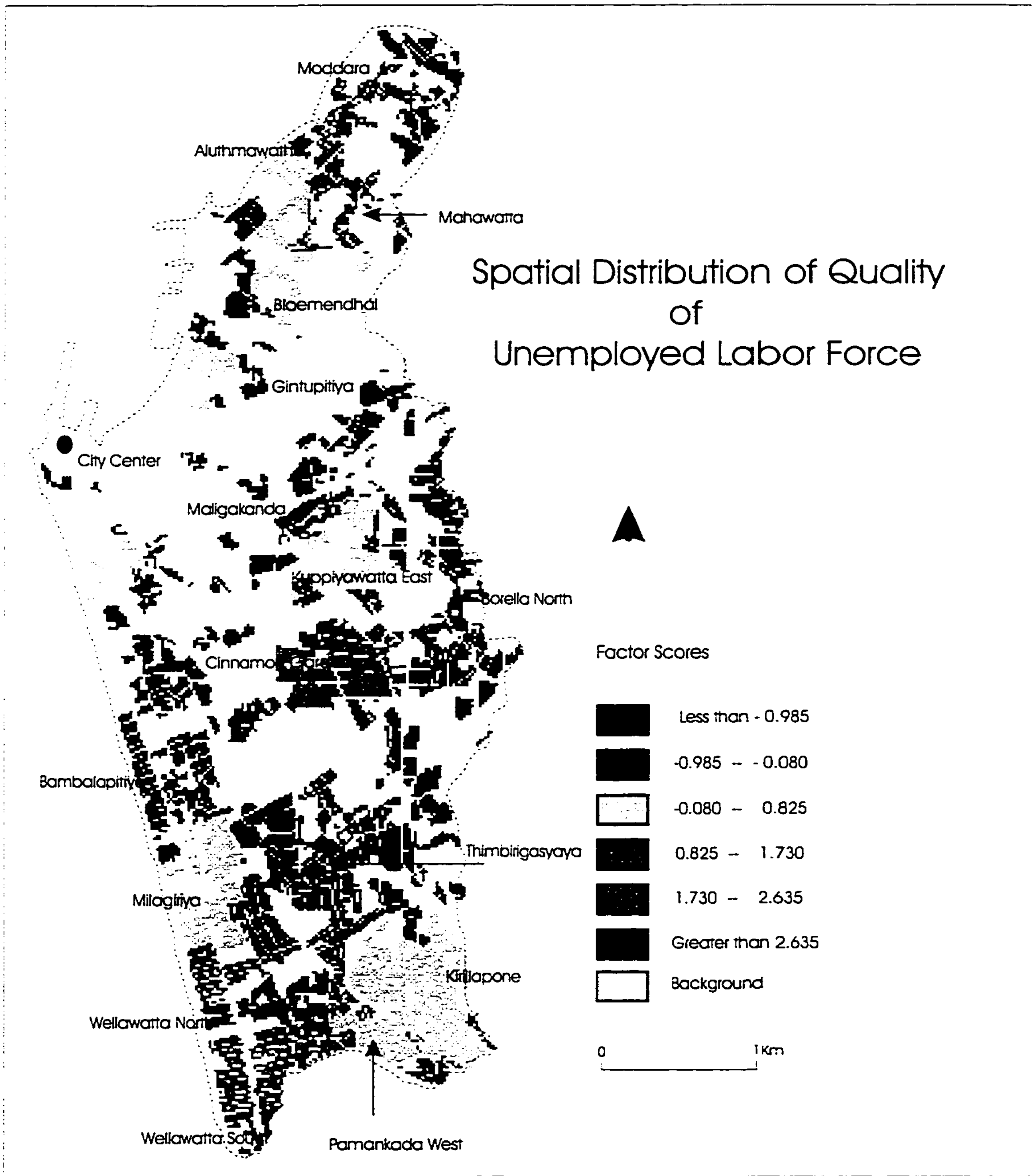
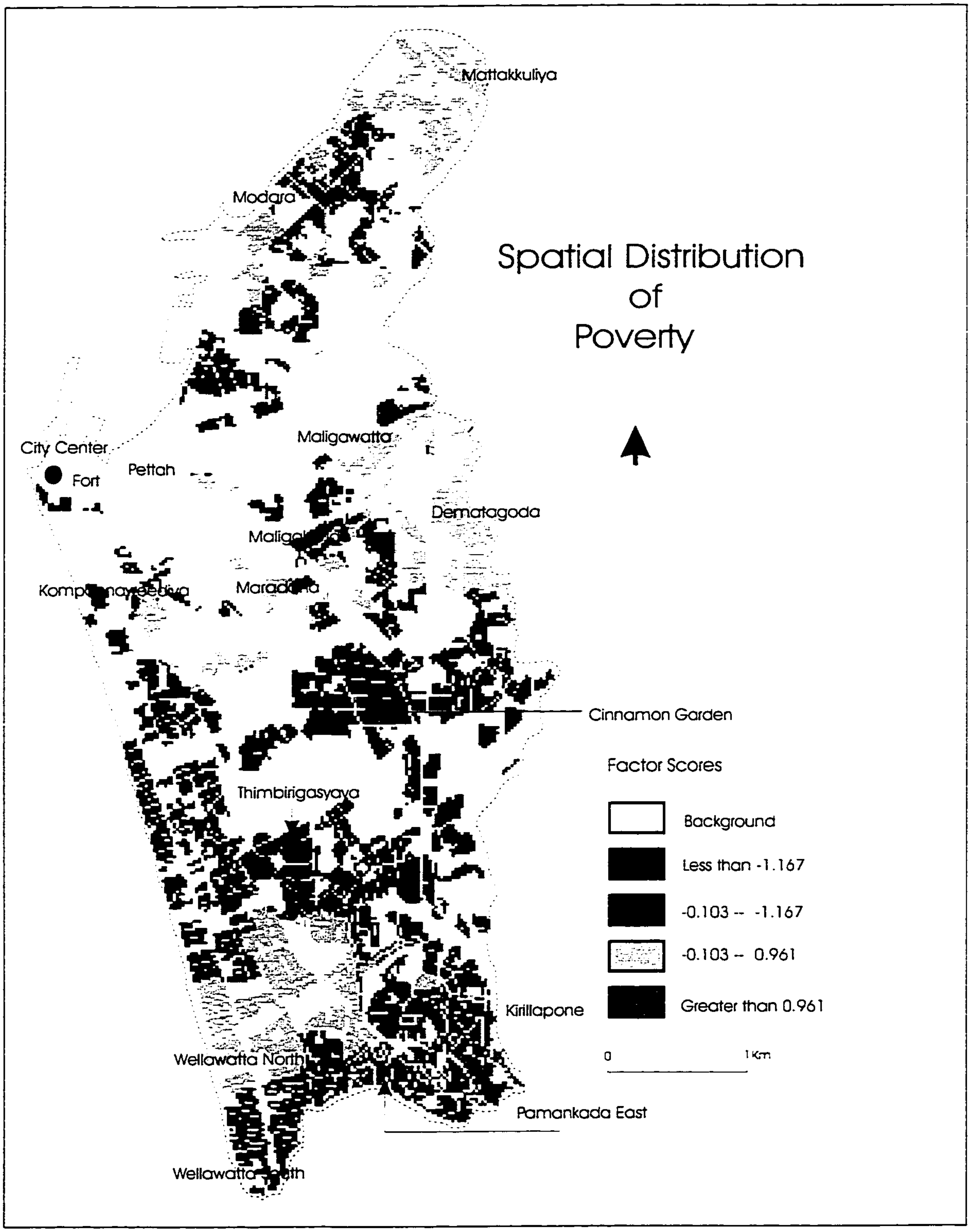


Figure 5.2 Distribution of Factor Scores on the Poverty Dimension



spatial distribution of factor scores confirms that the majority of the residential land in the city is occupied by the affluent and the poor are limited to a relatively small area. Poverty tends to be concentrated close to the periphery of the commercial areas (e.g. Compannaweediya, Maradana, Kochchikade and Masangasweediya) and at the city periphery (e.g.: Kirillapone, Pamankada East, Mattakkuliya, and Modara). Within the city, the sector most heavily inhabited by the poor runs along the coast north of the city center. Affluent areas are also located at the city periphery, running like an inconsistent belt around the city core. It is evident that both the poor and the affluent share the city periphery and the areas within the city limits. In spite of this, only the affluent enjoy the best land and the poor are restricted to the marginal land found in small clusters around the city limits and within close proximity of the city business area. Most of these marginal land sites are swamps, low lying areas that are susceptible to floods, or areas that are seriously overcrowded with buildings in a dilapidated condition (see Chapter Two).

This dimension confirms the findings of some of the previous studies of developed and developing world cities. But past factorial studies did not identify a dominant **Poverty** dimension. Although most factorial ecologies have identified these types of characteristics within the general socio-economic dimension, their intensity did not produce a separate poverty dimension. For example, Lo, 1986; Davies, 1983; White, 1987; and Hamm et al., 1988. Davies and Murdie's (1991) study on 24 metropolitan areas in Canada featured a dimension named "Impoverishment" which is quite different from previous findings. Unlike studies conducted in 1970s and 1980s on North American and other Western cities, it appears that studies carried out in the 1990s on these cities demonstrate the changing nature of the social geography of Western cities. As discussed in the previous chapter, most of the studies available for developing world cities are not based on adequate information about the poor. Hence, the characteristics of poverty have not been a major dimension of the available studies. When the Spatial distribution of **Poverty** is compared with the previous map, **Quality of Unemployed Labor Force,** it is evident that those areas with a significant level of unemployed labor possessing a high level of academic or technical qualifications are located in the parts of the town where the affluent are dominant.

FACTOR 3: Ethnicity

The **Ethnicity** dimension explains 9.6 percent of total variance. This factor provides insight into the ethnic diversity of Colombo, showing ethnic contrast between the Sinhalese and Moors. Beside the loadings of ethnic variables, two other variables related to Moors were loaded to this factor. They were: *Percentage of renovated buildings (UPGRA)* and *Female/ male unemployment ratio (female unemployment / male unemployment) (UNEMRAT)*. This is consistent with the observation in Chapter Three on female employment of Moors. Most Moor females stay at home taking care of family instead of working outside the house. Therefore, *Female/ male unemployment ratio* among the Moors is high.

The spatial distribution of **Ethnicity** is displayed in the Figure 5.3. Areas with high positive scores indicate where the Moors are dominant, while high negative factor scores show where the Sinhalese are. The spatial distribution pattern of Moors is very clear. They are concentrated around the city center, where most of the buildings are renovated/upgraded. That may be why the *Percentage of renovated buildings* was loaded on this factor with high positive loading. Most of the Moors are business oriented and live close to the city center. Areas such as Compannaweediya, Kochchikade, Kotahena, Maligawatta are dominated by Moors. The majority of Sinhalese live close to the city periphery (e.g.: Borella South, Kirillapone, Cinnamon Garden and Thimbirigasyaya).

FACTOR 4: Educational Resources

The dimension on **Educational Resources** explains 8.9 percent of the total variance. This is a clearly definable factor because of the distinct consistency of the variables loaded to this factor (Cronbach's alpha: 0.62). All the variables loading on this factor are associated with educational resources, and their contributions to the factor are strong.

The spatial distribution of factor scores on **Educational Resources** is demonstrated in Figure 5.4. Areas with high negative factor scores do not have adequate educational resources compared to areas with a high positive factor scores. When the spatial distribution patterns of Educational Resources and Poverty (Figure 5.2) are compared, it does not confirm the hypothesis that areas with high poverty concentration are disadvantaged in terms of inadequate educational resources since affluent areas do not

Figure 5.3 Distribution of Factor Scores on the Ethnicity

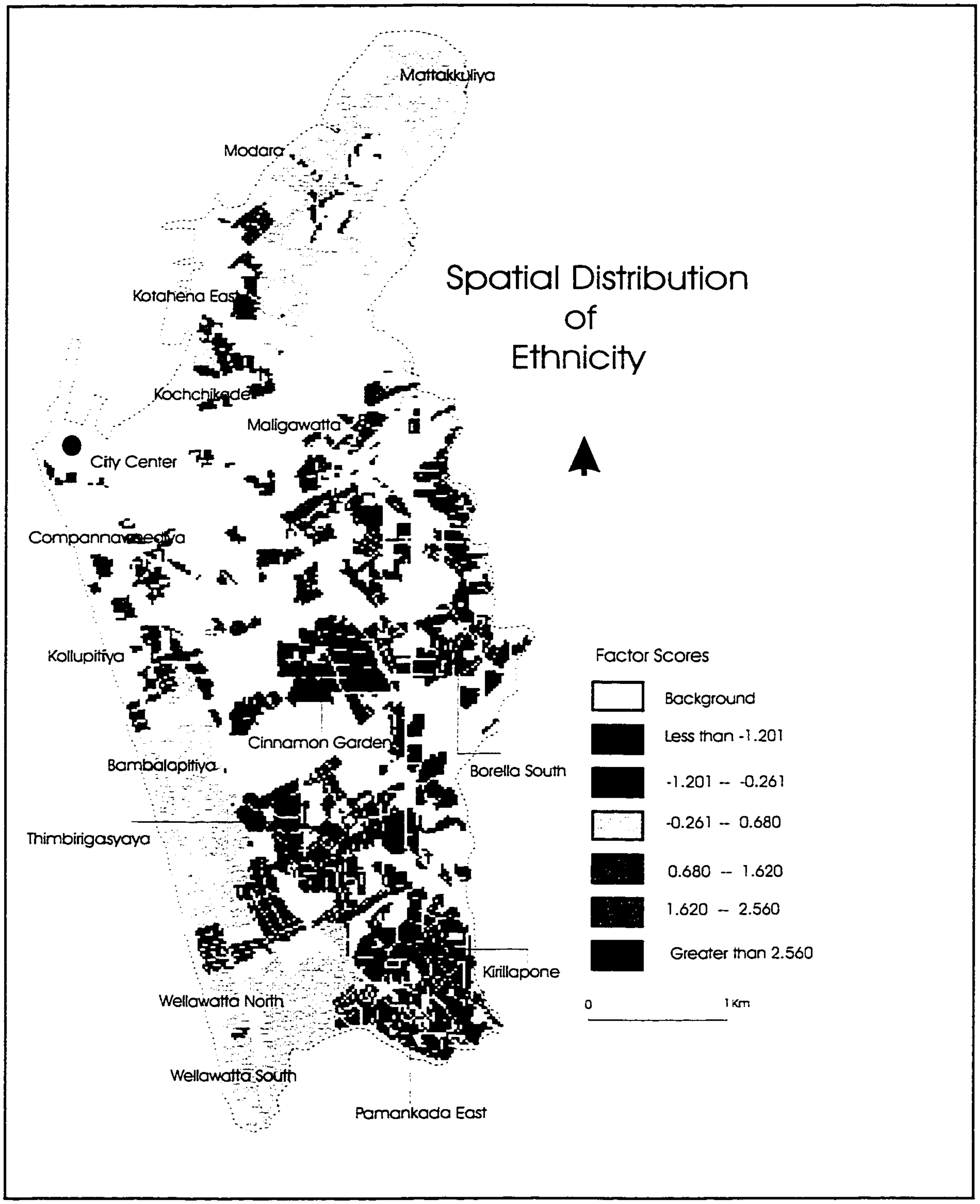
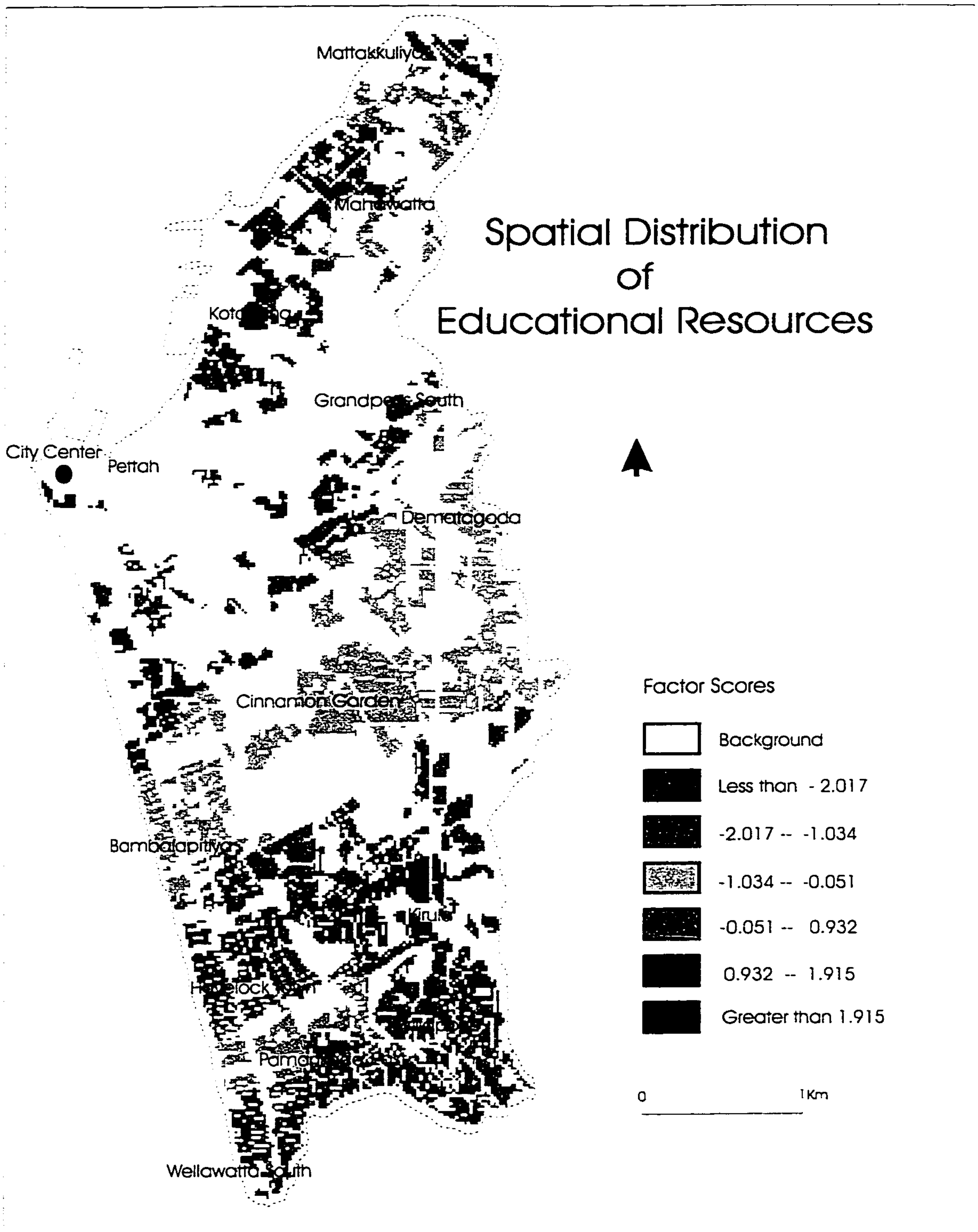


Figure 5.4 Distribution of Factor Scores on the Educational Resources Dimension



enjoy a higher level of **Educational Resources**, either. For example, although areas such as Cinnamon Garden, Borella South and Bambalapitiya are wealthy but they are below the city average in **Educational Resources**. As cited previously, most of the rich are interested in the private educational system to the detriment of public system.

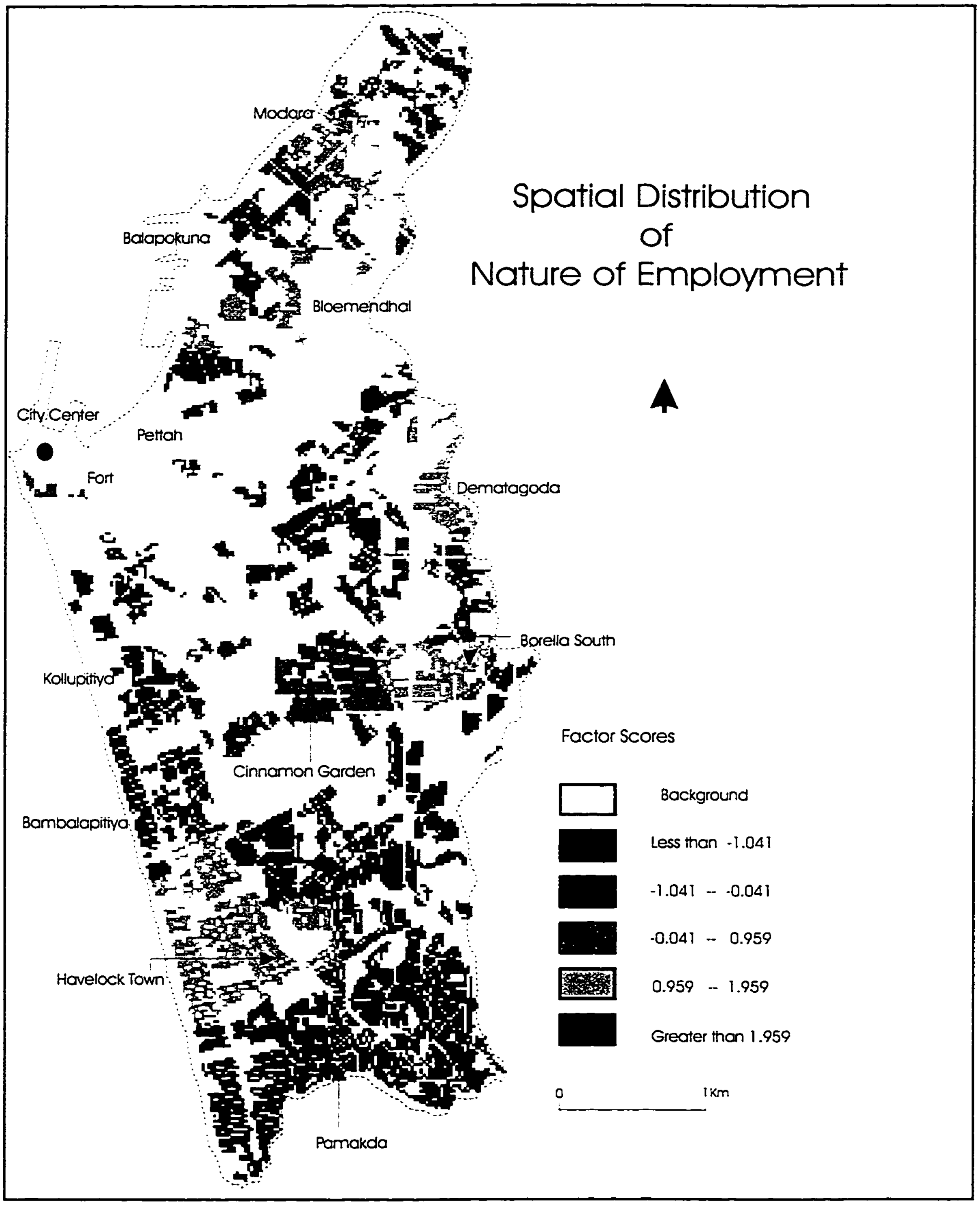
FACTOR 5: Nature of Employment

Although the explanatory power of this factor is not as great as the previous dimensions, this is a clearly definable factor in terms of the variables loaded to it. Three variables loading heavily on to this factor are: *Percentage of male employment of the total work force (MALEPERM), Percentage of female seasonal employment of the total work force (FSEASO)* and *Percentage of male seasonal employment of the total work force (MSEASON).* Not only the employment, but whether or not it is seasonal or permanent, has been captured by this factor.

The high positive factor scores of this dimension are related to high levels of female and male seasonal employment and low levels of permanent male employment. Since the variable, *Percentage of permanent male employment of the total work force*, is loaded negatively, a higher percentage of people do not have permanent work in the areas with high negative factor scores.

The spatial distribution of the factor scores tells us the pattern of the **Nature of Employment** of the city (Figure 5.5). Three main city regions with high levels of permanent employment are: North-Eastern sector (Balapokuna, Modara and Bloemendhal); central part (Borella South, Demtagoda and Cinnamon Garden); and Southern coastal zone (Milagiriya and Havelock Town). The highest seasonal employment areas are confirmed around the Balapokuna region, which displays the highest incidence of poverty concentration (see Figure 5.2). The interesting feature of the distribution is that the high rate of seasonal employment and low level of male employment are not restricted to the poor. The common view of the employment of poor is that they do not have permanent occupations, therefore, the majority of poor people are engaged in seasonal works (Arachchige- Don, 1994).

Figure 5.5 Distribution of Factor Scores on the Nature of Employment Dimension



FACTOR 6: Foreign Employment

This is the least important factor since it explaines only 4.4 percent of total variation. Although not as strong as the previous dimensions in terms of explanatory power, the factor does provides an example of a balanced factor scale.

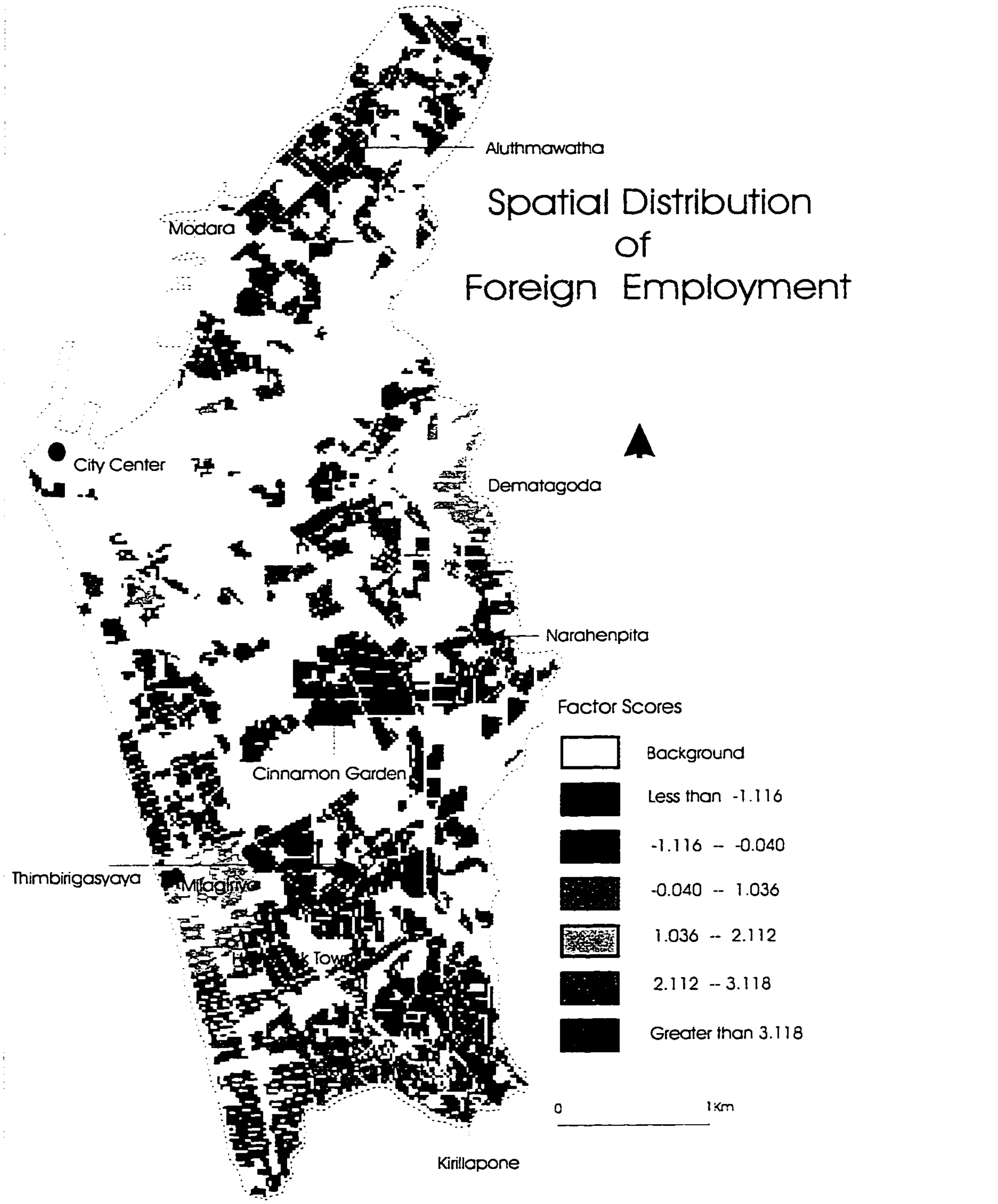
Until the end of the 1970s, foreign employment was not open to low income people because the Sri Lankan government imposed immigration restrictions for overseas employment. This situation was completely changed after the United National Front government came to power in 1978. Employment opportunities became available for Sri Lankans especially in the

Middle East, Singapore, Indonesia, and other neighboring countries. The major employment categories that attracted Sri Lankan labor were housemaids, drivers, garment and construction workers. These occupations do not need a high level of training; therefore, low income people began to migrate to these countries at a tremendous rate. Presently, one of Sri Lanka's major sources of foreign exchange comes from its Middle East employees. This latest employment trend has changed the social fabric of the capital and, indeed, of the entire country. The **Foreign Employment** dimension captures the latest trends in the employment structure of the city. Rapidly changing patterns of employment structure are occurring not only in Sri Lanka, but also in other neighboring countries such as Bangladesh, India, Pakistan, Maldives, and Thailand, since these countries are also major sources of cheap labor to Middle East countries.

Figure 5.6 shows the spatial distribution of factor scores on the **Foreign Employment** dimension. High positive factor scores show the areas with high levels of foreign employment. The dominant clusters are located at the South West corner of the city namely, Pamankada West, Havelock Town, and the Milagiriya areas. Three more distinct areas are located at Narahenpita, Dematagoda, and Modara. A notable pattern of the spatial distribution of foreign employment is that the major clusters are not spotted in the areas where the elite are dominant (see Figure 5.2).

The study reveals that the major dimensions derived from Colombo differ from previous studies on developing and developed world cities in many ways. Some researchers have stated that the limitations of the data sources have prevented the identification of an appropriate set of dimensions in their studies. As noted in the Chapter

Figure 5.6 Distribution of Factor Scores on the Foreign Employment Dimension



Two, a higher percentage of city dwellers in the developing world cities live in slums and shanties as compared to the developed world and thus play a dominant role in shaping the city. Therefore, this study employed an appropriate set of variables to capture the best set of dimensions to describe the urban social structure of developing world cities. The researcher proposes that the dimensions extracted, **Quality of Unemployed Labor Force, Poverty, Ethnicity, Educational Resources, Nature of Employment**, and **Foreign Employment**, represent the social fabric of the city satisfactorily. All the dimensions are distinct from the other studies and are capable of explaining the latest socio-economic trends of the city.

Association of Distance Variables with the Factors

The next step in the study was to explore the association of these social structures with the distance measurements selected. This was explored with multiple correlation and regression analyses. It was decided to extract six standard multiple regression (simultaneous) models foreach dependent variable (DV) to represent six social dimensions. This strategy calls for entry of all independent variables (IVs) into the regression equation at once. The guidance suggested by Tabachnick and Fidell (1996) was followed in the selection of an appropriate multiple regression model. As they point out, to simply assess relationships among variables and answer the basic question of multiple correlations the method of choice would be standard multiple regression. Reasons for using other methods might be theoretical or for development of hypotheses.

Before running multiple regression analysis, the nature of the data was explored in order to obtain appropriate results. The fit of the data with assumptions underlying parametric multivariate statistics should be checked before they are used in multivariate analyses (Girden, 1996; Griffith and Amrhein, 1997; O'Brien, 1992; Shaw and Wheeler, 1994 and Tabachnick and Fidell, 1996). The major issues considered, related to the data, were: missing data, outliers and normality.

Factor analysis and distance operations (see the previous chapter) generated factor scores and distance measurements respectively for each pixel under residential land use. Therefore, there were no missing values.

The researcher considered the search for outliers as a screening procedure for locating cases to purify the data set. Many researchers recommend using a standardized score of – or + 0.300 as a cut for identifying outlying cases (Tabachnick and Fidell, 1996). This analysis was based on that criterion. The distance measurements were converted to standardized format before identifying outliers. The outliers of six dependent variables (DVs) and independent variables (IVs) were inspected with tables of univariate descriptive statistics. The **Poverty** dimension had 38 cases while the **Foreign Employment** dimension had 10 cases with extreme values (outliers), defined by the above criterion. Extreme cases (of which there were 48) were excluded before the final analysis in order to avoid deleterious effects on regression solutions as on other multivariate solutions. This practice is recommended by many researchers (Tabachnick and Fidell, 1996, Girden, 1996; Griffith, and Amrhein, 1997; O'Brien, 1992; and Shaw and Wheeler, 1994).

The shape of the distribution of data points for each variable (DVs and IVs) was investigated after the outliers were excluded. This test of normality of variables was conducted through the SPSS (SPSS Inc., 1996) package by using a descriptive program in which measures of skewness were produced for distribution of variables. The test confirmed that none of the variables, tested, was skewed at $p = 0.01$ significance level. Therefore, the standardized original data set was used in the analysis, without any transformations.

Table 5.3 shows the standard multiple regression models obtained for each social dimension (dependent variables - DVs) while Table 5.4 shows the ANOVA results of standardized multiple regression coefficients. Three distance measurements, *Distance to the city center*, *Distance to the nearest public health center* and *Distance to the nearest public school with mathematics and science curriculums for Advanced Level classes (Grade 12)* were used as explanatory or independent variables (IV) to explain variation in each factor/dimension. Findings of the multiple regression analysis can be summarized as follows.

Analysis of the relationships between the social dimensions and the distance variables revealed that three distance variables, (IVs / predictor variables) play a significant role explaining the variance of the factors. These findings are acceptable as

Table 5.3 -Test of Multiple Correlation Coefficients (R)

Y_1 = Quality of Unemployed Labor Force; Y_2 = Poverty; Y_3 = Ethnicity; Y_4 = Educational Resources;
Y_5 = Nature of Employment; Y_6 = Foreign Employment
X_1 = SCHODIST (Distance to the nearest Public School)
X_2 = HELTHDIS (Distance to the nearest Public Health Center)
X_3 = CENDIST (Distance to the City Center)
R = Multiple Regression Coefficient

Section I

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1076.05	3	358.6682	386.97	0.00
Residual	13655.35	14733	0.926854		
Total	14731.35	14736			
A Dependent Variable: **Quality of Unemployed Labor Force (Factor 1)**					
B Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:
$Y_1 = -0.394 - 0.033X_1 + 0.085X_2 + 0.264X_3$; **R = 0.27 R^2 = 7.3%**

Section II

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	239.44	3	79.81	81.13	0.00
Residual	14494.30	14733	0.99		
Total	14733.70	14736			
A Dependent Variable: **Poverty (Factor 2)**					
B Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:
$Y_2 = -0.39E{-}04 - 0.003X_1 + 0.285X_2 + 0.095X_3$; **R = 0.127 R^2 = 1.6%**

Section III

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	2397.67	3	799.23	953.85	0.00
Residual	12344.70	14733	0.84		
Total	14742.40	14736			
A Dependent Variable: **Ethnicity (Factor 3)**					
B Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:
$Y_3 = 0.595 - 0.329X_1 + 0.229X_2 - 0.185X_3$; **R = 0.403 R^2 = 16.2 %**

Section IV

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1038.23	3	346.08	371.96	0.00
Residual	13707.70	14733	0.93		
Total	14745.90	14736			
Dependent Variable: **Educational Facilities (Factor 4)**					
Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:
$Y_4 = -0.705 + 0.188X_1 + 0.189X_2 - 0.113X_3$; **R = 0.265 R^2 = 7.0%**

Table 5.3 Contd.

Section V

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	3075.33	3	1025.11	1294.77	0.00
Residual	11664.60	14733	0.79		
Total	14739.90	14736			
Dependent Variable: **Nature of Employment** (Factor 5)					
Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:

$Y_5 = 0.917 - 0.320X_1 + 0.172X_2 - 0.196X_3$; $R = 0.457$ $R^2 = 20.9\ \%$

Section V I

ANOVA					
Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4619.36	3	1539.787	2246.11	0.00
Residual	10100.00	14733	0.686		
Total	14719.40	14736			
Dependent Variable: **Foreign Employment (Factor 6)**					
Independent Variables: (Constant), SCHODIST, HELTHDIS, CENDIST					

Regression Equation:

$Y_6 = -1.086 + 0.444X_1 + 0.225X_2 - 0.429X_3$; $R = 0.560$ $R^2 = 31.4\%$

Table 5.4

Test of Multiple Regression Coefficients

Section I

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-0.394	0.02788		-14.14	2.87E-24
	CENDIST	-1.9380E-05	4.91E-06	-0.032	-3.94	7.99E-05
	HELTHDIS	0.0001	1.79E-05	0.085	10.76	2.87E-24
	SCHODIST	0.0004	1.47E-05	0.263	31.66	2.87E-24

Dependent Variable: Factor Scores of **Quality of Unemployed Labor Force**

Section II

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	3.9E-04	0.029		0.014	0.989
	CENDIST	-2.1E-06	0.000	-0.003	-0.408	0.684
	HELTHDIS	1.9E-04	0.000	0.085	10.340	0.000
	SCHODIST	-1.7-04	0.000	-0.095	-11.086	0.000

Dependent Variable: Factor Scores of **Poverty**

Section III

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
	(Constant)	0.595	0.027		22.46	0.000
	CENDIST	-1.9E-04	0.000	-0.329	-41.63	0.000
	HELTHDIS	5.2E-04	0.000	0.229	30.39	0.000
	SCHODIST	-8.2E06	0.000	-0.005	-0.59	0.558

Dependent Variable: Factor Scores of **Ethnicity**

Section IV

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-0.705	0.028		-25.24	0.000
	CENDIST	1.1E-04	0.000	0.188	22.60	0.000
	HELTHDIS	4.3E-04	0.000	0.189	23.83	0.000
	SCHODIST	-2.0E-04	0.000	-0.113	-13.51	0.000

Dependent Variable: Factor Scores of **Educational Facilities**

Section V

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
	(Constant)	0.92	0.026		35.606	0.000
	CENDIST	-1.9E-04	0.000	-0.320	-41.640	0.000
	HELTHDIS	3.9E-04	0.000	0.172	23.437	0.000
	SCHODIST	-3.50E-04	0.000	-0.196	-25.491	0.000

Dependent Variable: Factor Scores of **Nature of Employment**

Section VI

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
	(Constant)	-1.086	0.024		-45.302	0.000
	CENDIST	2.6E-04	0.000	0.444	62.051	0.000
	HELTHDIS	5.1E-04	0.000	0.225	32.973	0.000
	SCHODIST	-7.6E-04	0.000	-0.429	-59.925	0.000

Dependent Variable: Factor scores of **Foreign Employment** the IVs are not intercorrelated. All the explained variances between the two groups of variables (DVs and IVs) are significant at $p = 0.001$. But before accepting these outcomes, the validity of the regression models should be investigated since some of the regression coefficients may not be significant.

The ANOVA results of standardized regression coefficients were explored to investigate which models can be applied to predict the relationship between each factor and each distance measurement. The outcomes reveal that all the regression coefficients except for the **Poverty** and **Ethnicity** dependent variables are significant at $p = 0.0001$. In the multiple linear regression model of **Poverty**, the coefficient for *Distance to the city center* and 'intercept / constant' are not significant ($p = 0.989$ and $p = 0.684$ respectively). Therefore, *Distance to the city center* can not be employed to predict Poverty. As seen in Figure 5.2, poverty areas are located closer to the city center and also to the city periphery. This pattern is quite different from that of western cities. Since the coefficients are in standardized format and the IVs are not intercorrelated, the importance of each distance measurement can be described. Regression coefficients tell us the amount of change in the dependent variable for a one-unit change in the independent variable. Only selected important findings follow.

Distance to the nearest public school cannot be utilized to predict the spatial distribution of **Ethnicity**. The regression coefficient between the **Ethnicity** dimension (Moors dominate the **Ethnicity** dimension) and *Distance from the city center* shows that the selected minority group is living closer to the Central Business District (CBD). Most of the buildings closer to CBD are renovated. As shown in the Appendix 3, the correlation coefficient between the *Distance from the city center* and *Percentage renovated buildings* (- 0.609) is significant at .05, indicating that renovated/upgraded buildings are concentrated within a short distance of the CBD. Furthermore, as noted earlier in this chapter, *Upgraded /Renovated Buildings* and *Percentage of Moors* loaded strongly to the **Ethnicity** dimension. All of these findings indicate that most of the Moors are concentrated around the CBD and they live in renovated buildings.

There is a positive relationship between **Foreign Employment** and *Distance to the nearest public school* and to *Distance to the nearest public health center*. Both the

Nature of Employment and the **Quality of Unemployed Labor Force** dimensions delineate a negative relationship with the *Distance to the city center*.

In conclusion, the foregoing analysis shows that social variation of Colombo can be simplified into six components. The urban social structures derived adequately represent the current social variation of Colombo. The spatial distribution pattern of six dimensions extracted is unique to the study area. The GIS techniques applied could generate a compatible data set from different data sources and the new data set may be strong enough to produce a simplified version of social structures. Although the variable filtering process was very cumbersome, it helped the researcher to obtain a meaningful set of dimensions to describe the social geography of Colombo. Finally, multiple regression analysis revealed that the selected distance measurements are powerful predictors to explain the variance of social structures. The following chapter will go a step further to describe the social areas of Colombo based on cluster analysis techniques.

CHAPTER SIX
SOCIAL GROUPS IN COLOMBO

Introduction

As described in the previous chapter, the researcher was able to disentangle the different sources of social differentiation and their associated patterns within the complex social mosaic found in the study area. The purpose of this chapter is to go one step further and synthesize the results of the individual distribution maps in order to produce a final classification of the sub areas of the city. Unlike factor analysis, cluster analysis helps the researcher detect and investigate the ***combination*** of different sources of social variations as a means of understanding the characteristics of different social groups in Colombo and their spatial distribution patterns. Factor analysis synthesizes a large number of variables or attributes into a small number of variables or attributes based on the similarities among them. In contrast, cluster analysis groups the observations of a study into a limited number of groups or clusters, considering the similarities of the observations.

Comparatively few factorial ecologists have gone beyond the description of factor loadings and scores. Some exceptions being, Davies, 1975 and 1983; Herbert, 1977; Lo, 1975; and Yeh et al., 1995 who have applied cluster analysis to explore the spatial distribution pattern of the social groups in a city. The standard procedure is to apply a cluster analysis solution to the similarity matrix, which was calculated between each pair of areas on the basis of the factor scores. This researcher decided to follow the same path in order to compare his results with others. It was expected that an even more parsimonious description of the original data set could be produced by applying cluster analysis to the factor scores.

The factor scores generated were further analyzed to reveal the social groups of Colombo. The cluster analysis was performed using SPSS K-means Cluster analysis. This procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics using an algorithm that can handle a large number of cases. This procedure, however, requires that the user specify the number of clusters (SPSS Inc., 1996; Griffith and Amrhein, 1997; Goddard and Kirby, 1976; and Girden, 1996).

The procedure was initially carried out for two through five clusters on the six factors described in the Chapter Five. Selection of the optimum number of clusters was achieved by applying three criteria. First, the minimum distance among the factors had to be large enough to separate each cluster distinctively. The program calculates the Euclidean distance between final cluster centers. Secondly, none of the clusters could be extremely large or small in terms of the number of observations contained in the cluster. The researcher formulated the third criterion, that is, each cluster group generated by the program should make qualitative sense. After the clusters were generated based on all the factors, the program calculated Cluster Center Values (CCV's) on each factor. A CCV represents a mean value for a given factor which shows its contribution to a given cluster (Griffith and Amrhein, 1997 and Girden, 1996). If the CCV's are distinctly different on each cluster, the clusters generated are easy to interpret. There is no use having too many clusters if they do not make sense in terms of the combination of items contained. Therefore, subjective elements should be included in order to decide the appropriate number of clusters.

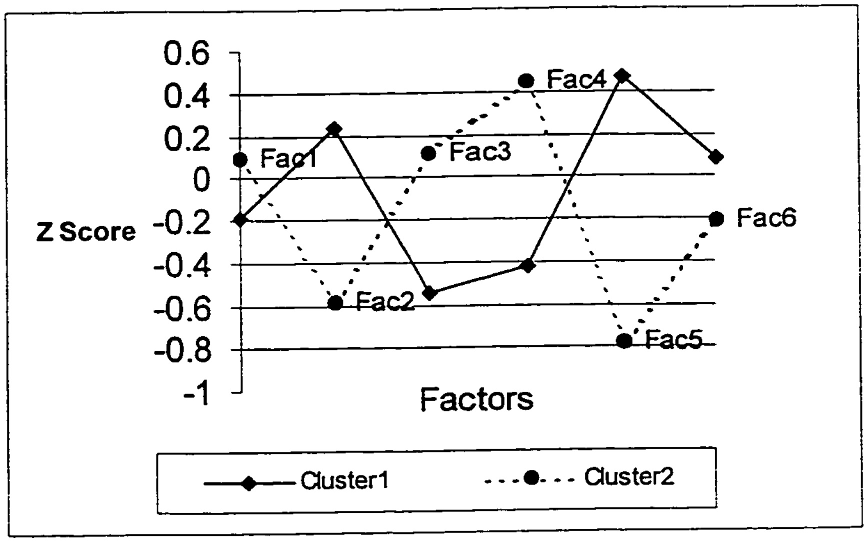
Two-Cluster Solution

The initial cluster solution of two clusters shows the most general pattern of social areas in Colombo. Table 6.1 and Figure 6.1 show the results of the Two-Cluster solution. The cluster analysis produced two homogeneous clusters based on the CCV of each factor employed in the analysis. The CCV on a selected factor should be described in the same way that the particular factor was described.

Table 6.1 Two-Cluster Solution

	Quality of Unemployed Labor Force	Poverty	Ethnicity	Educational Resources	Nature of Employment	Foreign Employment
Cluster1	-0.1927	0.2334	-0.5449	-0.4237	0.4686	0.0833
Cluster2	0.0841	-0.5937	0.1012	0.4385	-0.7851	-0.2203

Figure 6.1 Two Clusters Based on Six Dimensions



Factor 1: **Quality of Unemployed Labor Force**

Factor 2: **Poverty**

Factor 3: **Ethnicity**

Factor 4: **Educational Resources**

Factor 5: **Nature of Employment**

Factor 6: **Foreign Employment**

For example, the CCV of the second cluster for **Nature of Employment** is -0.7851 which indicates that this cluster has a strong negative relationship with this factor. The major characteristics of this factor are high levels of male and female seasonal employment and low levels of permanent employment.

Figure 6.1 shows the changing pattern of CCV of the Two-Cluster solution presented as standard scores. Each line in the graph represents a cluster and a factor score gives the boundary for each cluster. The two lines are distinct from each other, creating a meaningful set of clusters. Except for **Quality of Unemployed Labor Force** and **Foreign Employment,** all the other factors provided distinct CCV's between two clusters. These two groups generated the maximum variance between two clusters and the minimum variance within the clusters.

The results of cluster analysis (cluster number and the referenced observation number/group number) were manipulated with Microsoft Access Database to generate maps on clusters. The initial data file with 180,000 cases (see Chapter Four) and the new file with 1195 cases were linked by the group number. The "Update Query" function was applied to update the original data file with the new attribute, cluster value. Only the attribute field containing cluster values for 180,000 observations was exported to IDRISI and images were produced to show the spatial distribution of clusters. The clusters are mapped in Figure 6.2.

The social geography of Colombo has been separated into two major groups which are strongly distinct from each other with respect to **Poverty**, **Ethnicity**, **Educational Resources**, and **Nature of Employment**. Sinhalese, high levels of poverty and seasonal employment and low levels of educational resources dominate the first cluster. This cluster represents mainly the poor areas of the city. When Figure 6.2 is compared with Figure 5.2 (which shows the spatial distribution of poverty), one can see that all the areas with high levels of poverty fall under Cluster One. The second cluster represents the areas dominated by the elite of the city. Since the CCV for **Ethnicity** is not strong (0.101), the researcher suggests that a mixture of Sinhalese and other minorities represent the cluster. The major distinguishing characteristics of this cluster are low levels of poverty, higher levels of educational resources, and permanent employment.

Three-Cluster Solution

To determine the best number of clusters, the cluster program was run again for three clusters and the results were plotted on a graph (Figure 6.3). The relative level of contribution of each factor was determined by the magnitude of the CCVs. The three clusters identified should minimize the variation within the cluster while they maximize the variation between clusters or groups. As the lines on the graph show, each line is distinct from each other, creating a meaningful set of clusters. The first and second clusters should contrast with each other since the lines that represent those clusters display opposite characteristics. The gap between the two lines is high except at the fourth factor, **Educational Resources**. This indicates that these clusters have some similarities in terms of the educational resources considered in the factor analysis.

The four and five cluster solutions were explored to see whether they brought greater insight to the social groups in the city. As far as the first and second criteria are concerned, both four and five-cluster solutions were appropriate since the minimum distance between the clusters produced in both stages was adequate enough to well separate the clusters (5.5945 and 4.4313 respectively). However, the latter solution was rejected based on the third criterion mentioned above.

Figure 6.2 The Distribution of Cluster Membership (After Two-Cluster Solution)

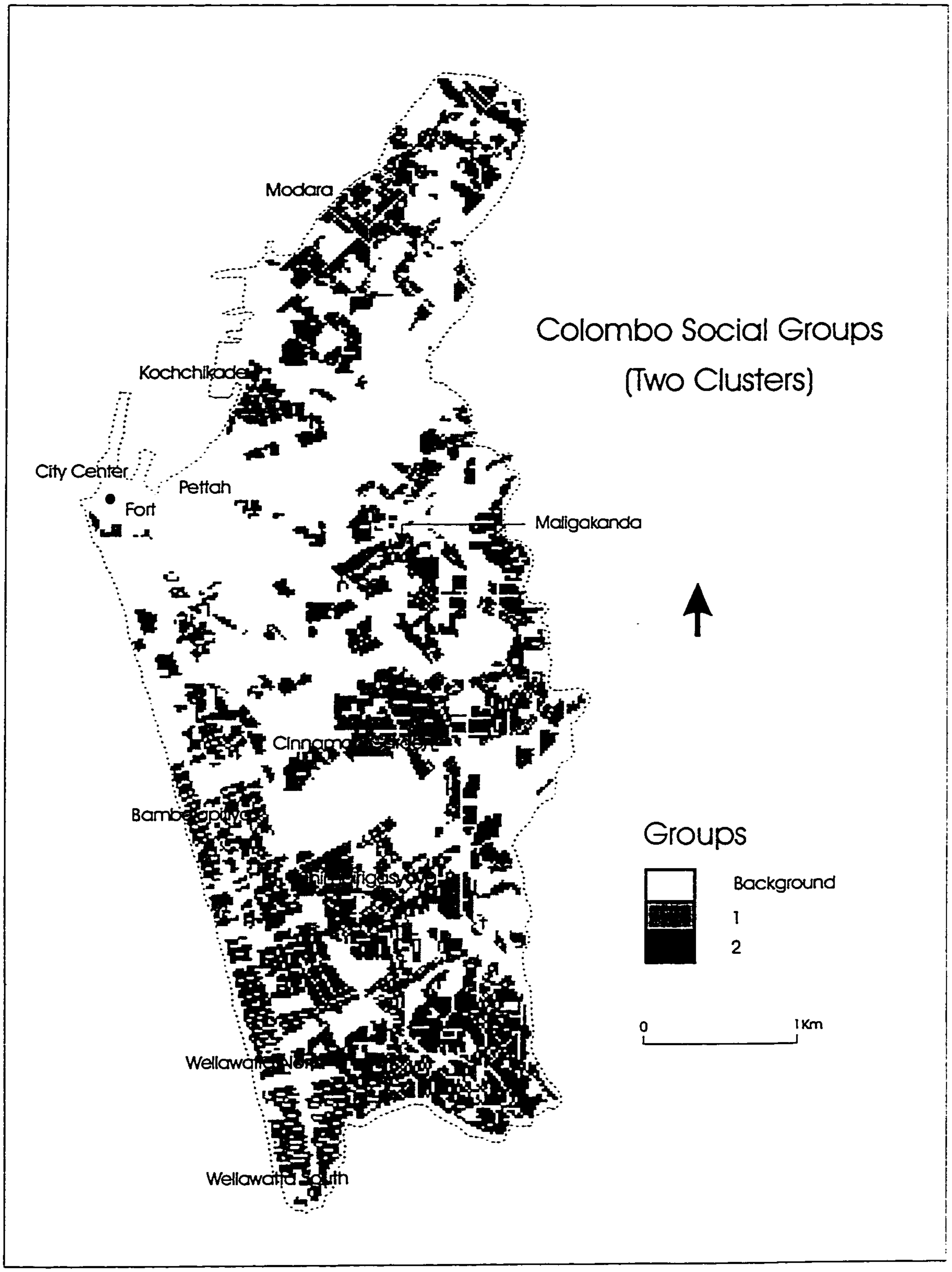
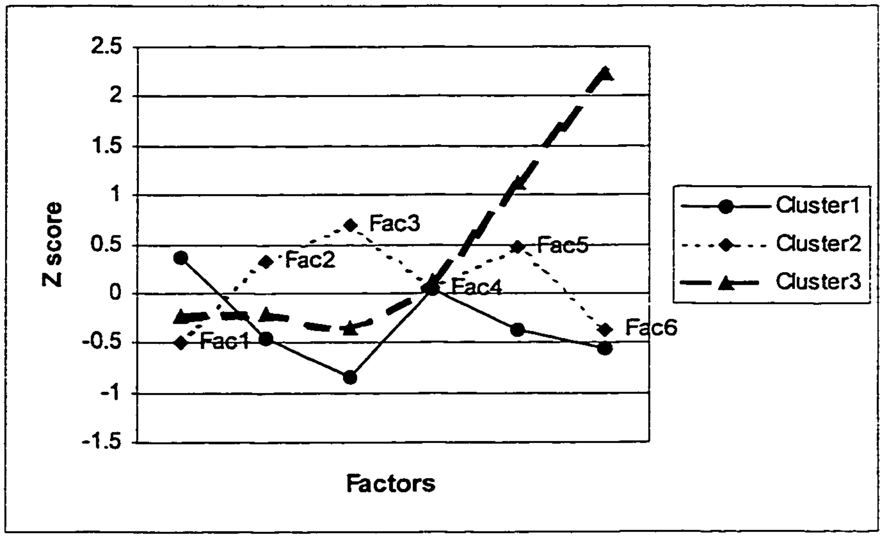


Figure 6.3 The Three-Cluster Solution



Factor 1: **Quality of Unemployed Labor**

Factor 2: **Poverty**

Factor 3: **Ethnicity**

Factor 4: **Educational Resources**

Factor 5: **Nature of Employment**

Factor 6: **Foreign Employment**

The Four-Cluster Solution

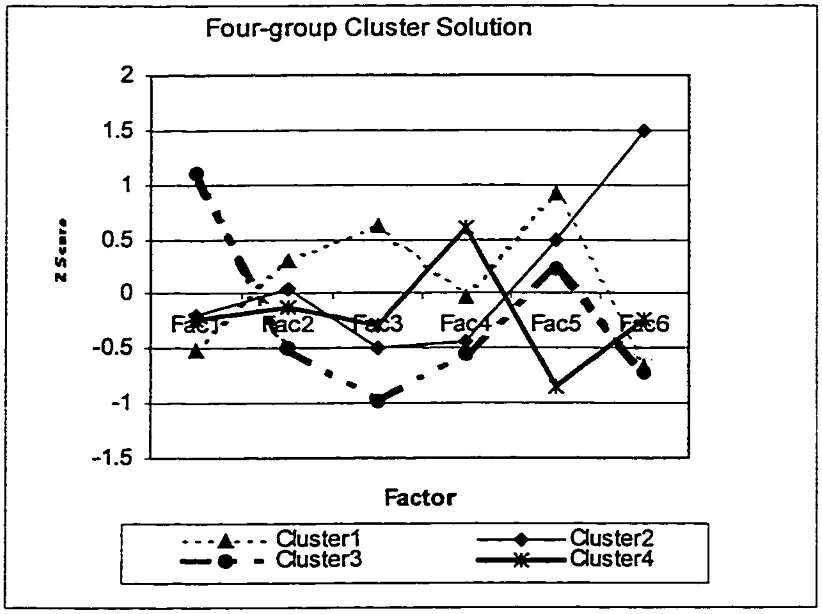
The results of four-cluster solution are shown in Figure 6.4 and their spatial distribution patterns are mapped in Figure 6.5. The four-cluster solution produced a clearly definable and meaningful set of clusters. Using the results of the given figures, it is possible to identify the distinguishing characteristics of each cluster.

Cluster 1: ***The Poorest People of the City***

This is the most distinct cluster produced by the four-factor solution, clearly defined by almost all the factors that were employed to reveal the social groups of the city. This cluster is meaningful in terms of the social characteristics it displays. **Quality of Unemployed Labor Force, Poverty, Ethnicity, Foreign Employment and Type of Employment** make this cluster clearly distinct from others.

The city's lowest level of Unemployed People with High Academic or Technical Qualifications (**FACTOR 1**) is found under this category. As quoted from different sources previously, poor people are likely to have very low academic background. Therefore, this situation is expected. Since the CCV for **Poverty** is higher than the city average (0.3037), it indicates that this social group has the highest level of poverty in the city. Furthermore, the high positive CCV of **Ethnicity** indicates that areas under this cluster are highly dominated by minorities.

Figure 6.4 The Four-Cluster Solution



Factor 1: **Quality of Unemployed Labor**

Factor 2: **Poverty**

Factor 3: **Ethnicity**

Factor 4: **Educational Resources**

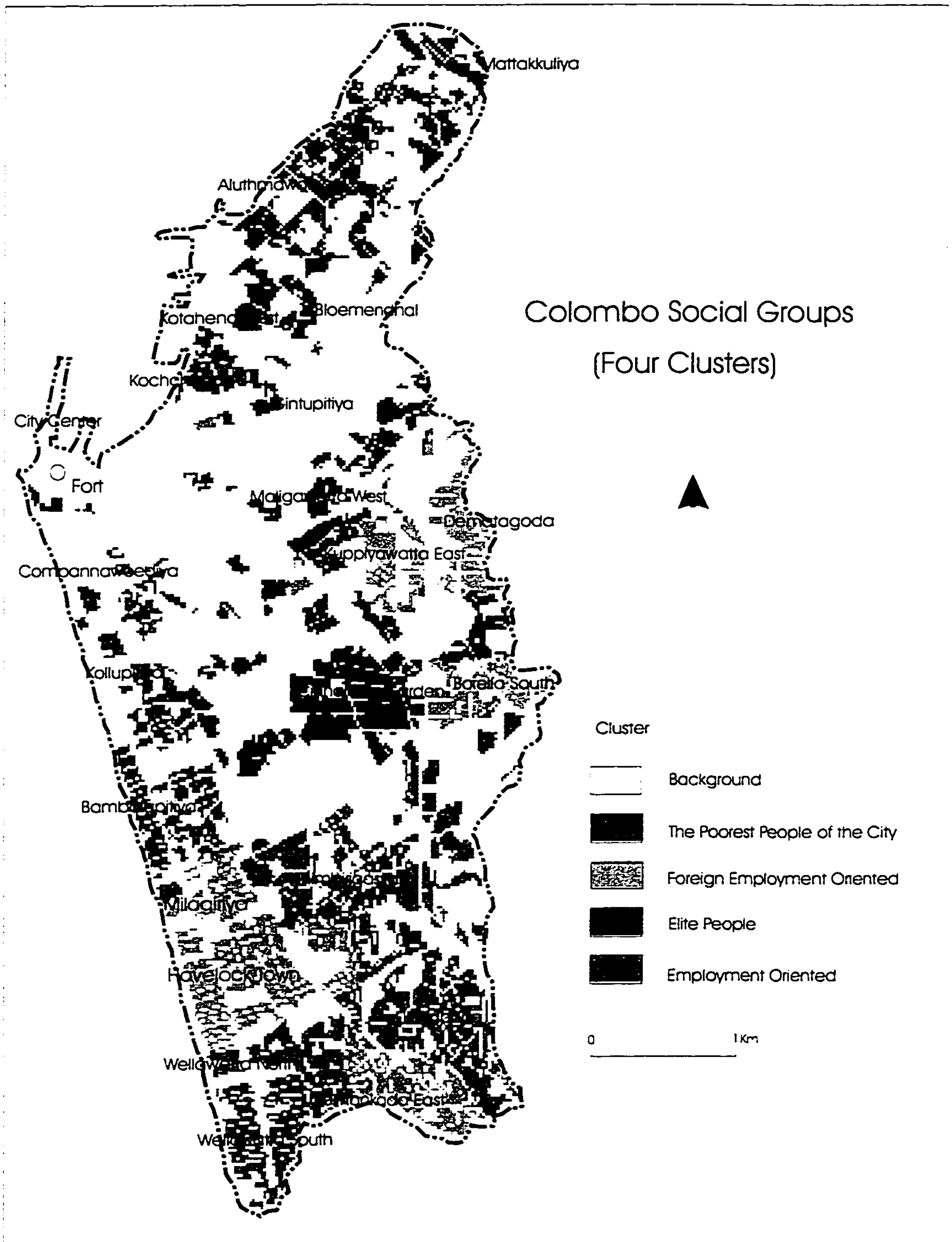
Factor 5: **Nature of Employment**

Factor 6: **Foreign Employment**

The values on **Type of Employment** indicates that the majority of both males and females in this category do not have permanent work, but rather, seasonal occupations. The CCVs on **Educational Resources** indicates its contribution to this cluster. It has been shown that this social class enjoys fair access to educational resources. Indeed their level of access to education is higher than that available to the wealthiest social group. Finally, it can be noted that males and females of this cluster employed in foreign occupations of this cluster are lower than the city average.

The researcher concludes that this cluster represents the poorest social group of the city. The spatial distribution of the cluster is very distinct since all the areas under this group are located around the city center and in the northeastern part of the city (Figure 6.5). The major clusters are located at Bloemendhal, Mattakkuliya and Gintupitiya. These areas are not ideally suited for human settlements since they are located on marginal land such as swamps or other low lying areas that are prone to flooding. This reinforces the observation that shanty and slum dwellers are the poorest social group of the city in terms of the underlying sources of variation considered.

Figure 6.5 Distribution of Cluster Membership (After Four-Cluster Solution)



Cluster 2: ***Foreign Employment Oriented***

This is one of the emerging social groups in the city. When Figures 6.5 and 5.1 (Spatial Distribution of **Quality of Unemployed Labor Force**) are compared, it is evident that all areas with a high level of **Foreign Employment** do not coincide with those areas in which a high level of unemployed people with high academic or technical qualification are found. Although these people are not very well off, their level of poverty is not as severe as the previous social group, in fact, it is just above the city average. However, this group shows some signs of low level socio-economic characteristics in terms of **Nature of Employment**. The CCVs on this factor indicate that this group has a high percentage of male and female employed in seasonal work, above the city average. Unlike the previous social group, Sinhalese dominate this group.

Cluster 3: ***Elite People (Exclusive High-Class Residential Area)***

This group represents the wealthiest people in the city. The CCVs on the factors applied to identify the clusters reveal the apparent characteristics of this group. As shown in Figure 6.4, this group is distributed at the opposite end of the first social group, ***The Poorest People of the City***. The **Poverty** factor is defined by loadings of five variables. Since this social group has opposite characteristics from the first cluster, the people under this group should have higher *Percentages of good houses*, higher *Percentages of employment in high income categories*, and lower *Percentages of population under five years of age*. The most interesting characteristic is that the majority of the population consists of Sinhalese. The CCV for **Ethnicity** is extremely high (-0.854) compared with CCVs of other clusters. This indicates that the contribution of the minority group to this cluster is very low.

Another pronounced feature of this cluster is that the lowest level of **Foreign Employment** is associated with this social group. As observed in the previous chapter, the **Foreign Employment** factor is related to poor people engaged in unskilled labor. This is further confirmed by the CCV on **Quality of Unemployed Labor Force**. Unlike all the other social groups in the city, this social group consists of the highest level of **Quality of Unemployed Labor Force**.

The most interesting finding was that this social group has the lowest level of access to **Educational Resources.** This does not mean that the elite do not have better access to high levels of **Educational Resources** because there are many causative factors other than the variables employed to measure the level of **Educational Resources**. For example, distance to a popular school is not a barrier to the elite since they have the financial resources and the political power to get to the school. Moreover, it is a well known fact that the elite are not interested in the public school system but rather in private schools such as 'international schools' where the medium of instruction is English. Only the elite can afford these institutions which are seen as gateways to overseas universities. Thus, having a low level of accessibility to **Educational Resources** does not necessarily mean that the high-income group does not have access to a high level of **Educational Resources**. Based on the dominant characteristics of the cluster, the researcher proposes that this cluster represents the most affluent social group of the city.

Cluster 4: ***Employment Oriented***

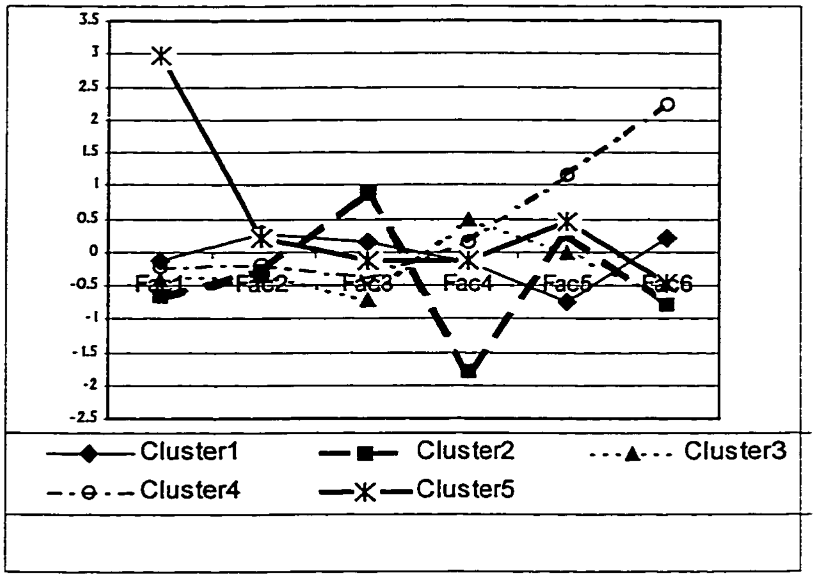
Based on the CCVs, the fourth social group can be distinguished on three major factors: **Type of Employment, Educational Resources,** and **Foreign Employment**. This social group is better off than ***The Poorest People of the City***, and ***Foreign Employment Oriented***, and therefore can be labeled ***Middle Class People with Permanent Employment***. It is evident that social groups dominated by Sinhalese are better off than minorities. The highest levels of **Permanent Employment** and **Educational Resources** are recorded among the people in this category. Since the level of **Permanent Employment** is dominant in this group, the percentage of people employed in other categories would be expected to be low. This situation is confirmed as the lowest level of **Foreign Employment** is related to this social group. As in the previous two social groups, the Sinhalese dominates this cluster. The most distinct characteristic of this group is that they enjoy the best access to **Educational Resources**.

The Best Number of Clusters

The results of the five-cluster solution were investigated to see whether they differentiated clusters adequately enough and whether the clusters are meaningful (Figure

6.6). The five-cluster solution is acceptable only under the first two criteria since the minimum distance between clusters was satisfactory (4.4313) and the allocation of cases

Figure 6.6 The Five-Cluster Solution



Factor 1: Quality of Unemployed Labor

Factor 2: Poverty

Factor 3: Ethnicity

Factor 4: Educational Resources

Factor 5: Nature of Employment

Factor 6: Foreign Employment

among the clusters was satisfactory. But the biggest problem that the researcher confronted was that the clusters derived cannot be described meaningfully.

The researcher suggests that the solution does not differentiate clusters adequately enough. In addition to, some of the clusters being uni-dimensional, the clusters do not separate distinctively. Therefore, clusters cannot be described meaningfully based on the CCVs obtained. For example, the first and fifth clusters are characterized by one factor **Type of Employment** and the **Quality of Unemployed Labor Force** respectively. The purpose of cluster analysis is to extract social groups based on the factors utilized in the analysis. If the social groups are defined by one factor, the results of such a procedure are similar to the areas identified by factor analysis, which is not the goal of this analysis. Furthermore, most of the CCVs on factors are scattered around the city average (Figure 6.6). This indicates that many of the clusters are not very distinct from the city averages making variance between the clusters low.

Therefore, the researcher proposes that the five-cluster solution is beyond the objectives of the study. Given that the five cluster solution for six factors failed to differentiate adequately among the clusters generated, the researcher suggests that the four-cluster solution generates the best pattern.

Spatial Distribution of Four-Clusters and Explanation for the Patterns

As we saw in the previous sections, two, three and five cluster solutions were rejected, considering the criteria employed. We should explore the distribution pattern of the four-cluster solution and examine why the patterns occurred more closely. The four-cluster solution separates clusters adequately enough showing their differences distinctively. The variance among the clusters and homogeneity within each cluster is high. The spatial distribution of the first social group, ***poorest people of the city***, shows a distinct pattern (Figure 6.5). Two dominant areas with the poorest people can be highlighted. The first area is closer to the city center and the other is at the city periphery, albeit limited only to the northeastern sector of the city. Furthermore, this social group occupies a smaller percentage of residential land use proportionate to their population size than any other group in the city.

This spatial distribution pattern occurs because all the areas where the poor are located are areas that are marginal to the other land use categories, and are not of interest to the other social groups. These areas do not provide a healthy living environment for humans. One set of areas is located within the vicinity of the city center, where the poor occupy slum dwellings. The slums are encountered in decaying urban areas, which include older properties, abandoned mansions of an earlier elite, and the tenements built by the commercial companies for their employees. These areas are seriously overcrowded and the buildings are in a dilapidated condition. Other than these slums, a number of shanty units located close to the city center are near Lake Beira which is badly polluted with disposal from near by shanties and factories (Figure 2.4). This group of poor people settled closer to the city center as they lacked the means to commute longer distance to the CBD. They provide cheap labor as housemaids, factory workers, temporary laborers while others survive working as street venders, prostitutes, and other minor low skilled occupations.

The lifestyle of the second group of poor people in the peripheral areas of the city is different. They are settled mainly in the low lands of the city (Figure 2.4) which are prone to frequent floods by the Kelani river or other water channels. As the areas are flooded very often, the soil is very fertile. The poor here take advantage of the rich soil by growing vegetables, mainly greens, for the Colombo market.

The second social group, ***Foreign Employment Oriented,*** is a result of recent changes to the social fabric of the city. It is characterized by foreign employment and is better off financially than the poorest social group of the city. When Figure 6.5 and Figure 6.2 are compared, it is evident that the ***Foreign Employment Oriented*** clearly a sub-group of the first social group, the poor, identified in the two-cluster solution. This social group may well represent social upgrade from the lowest to next lowest level. Four prominent clusters visible are: Pamankada East, Havelock Town, Borella South, and Kuppiyawatta East. Most of the people in this category are living at the city periphery.

Due to the limitations of the data set and other information, the researcher is not able to provide a good explanation for the distribution pattern. Therefore, detailed information from the areas under this social group should be collected in order to give a better explanation for the reasons behind the distribution pattern. After the United National Front government came to the power in 1978, not only did they relax the immigration restrictions that had been imposed on overseas employment, but also launched training and information programs for interested people. A few foreign employment bureaus were opened in the city to fulfill these initiatives. The reason for concentrating this social group in specific areas of the city might be that because such as the ones previously mentioned, were located in those areas of the city that would be accessible to that social group that would take advantage of the opportunity to secure overseas employment.

The exclusive high-class people, dominated by Sinhalese ***(Elite People),*** live close to the city center, enjoying the best city services. The main concentration centers are located at Cinnamon Garden and Modara, of which the former is the dominant cluster. They are buffered from other social groups by non-residential land use, namely parks and playgrounds. In this way, the elite areas consume an enormous amount of the city's land while housing only a small proportion of its residents.

Cinnamon Garden is the most sought after area by the elite due to its colonial legacy. This area, once occupied by colonial administrators has the best land plots in the city. Only a few indigenous people were able to locate in these areas when the country was under colonial rule but, after Sri Lanka achieved her independence in 1948, the indigenous elite took full control of this prestigious residential area.

The last social group, ***Employment Oriented,*** is a sub-set of the affluent social group recognized in the two-cluster solution. People who belong to this group live on the south side of the city. The residential areas which fall into this category are not mixed with the neighborhoods of ***The Poorest People of the City***, Cluster One, of the city except a small cluster under this category located at the north-east sector of the city.

Spatial distribution of this cluster is mainly determined by the land use pattern, the neighborhood characteristics, and educational resources. People under this social group tend to be isolated from the poor neighborhoods. The residential areas under this social group are not associated with swamps or low lying areas that are susceptible to floods (Figure 2.4). The characteristics and reasons for the distribution pattern of this social group confirm the researcher's findings from previous work (Gunadasa, 1988) on the town of Matara, Sri Lanka. This previous study revealed that many people who have permanent work are migrating to the city, where popular schools are located, hoping to provide an excellent education for their children. Their primary reason for relocating to a new residence is proximity to a popular school since the distance from the residence to a popular school plays a critical role obtaining admission to that school.

The number of social groups, their characteristics, and spatial distribution patterns are controlled by various factors. The major factors are the socio-economic variables used in the analysis, socio-economic background and geography of the study area, government and non-governmental policies imposed on the distribution pattern of social groups, and the technical procedures applied to identify social groups. The researcher states that the present study does not reveal the only possible or in some cases adequate explanation for the spatial distribution of the social groups identified in the study. The problem areas should be selected and further, more detailed study should be conducted for a more thorough assessment.

The results of cluster analysis cannot be closely compared with the few available studies (Davies, 1975 and 1983; Herbert, 1977; Lo, 1975; and Yeh et al., 1995) for two main reasons. First, there are only two other studies available on developing world cities. One (Yeh et al., 1995) is on an ideal socialist city, Guangzhou, China. In that study, the major criteria applied to identify the social areas are population density, level of education, and agricultural land use. By contrast, the socio-economic environment of Colombo is based on a capitalist frame; therefore, the characteristics of the social groups identified and their spatial arrangement are different from the Guangzhou study. For example, some of the social groups found in Guangzhou are *"Worker Areas"*, *"Cadre Areas"* and *"Scattered Agricultural Areas"* and they are characterized by the criteria mentioned above.

The other study on a third world city (Lo, 1975) is on Hong Kong which was a British colony when the field research took place. The majority of the people are Chinese; therefore, there is not much diversity in terms of ethnicity, but Lo was able to identify five social groups based on the other socio-economic characteristics. The five social groups found and their distribution patterns are unique to Hong Kong. For example, *High-income expatriates, Old established blue and white-collar workers' belt,* and *Rural area* are three of the social groups found in the study and their characteristics are not similar to the social groups found in this study of Colombo.

The studies from developed world cities (Davies, 1975 and 1983; Herbert, 1977) reveal that their social groups, characteristics, and distribution patterns are mainly a result of industrialization. Sri Lanka has not reached that level; therefore, social groups and their characteristics are different, making the comparison between the two groups complicated.

In conclusion, the results of the social dimensions and the patterns of social areas of Colombo differ from those of Western cities and developing world cities. The nature of the social dimensions and the social areas defined by the social dimensions / structures mainly depend on the nature and the type of variables included in a study. As a result of this, the social groups defined by social structures are unique to Colombo. But the researcher proposes that the urban social groups, their distribution patterns, and explanations for particular patterns produced based on the four-cluster solution represent a realistic image of a third world city as the data set adequately represents the role of the poor. The researcher's primary goal was to perform factor and cluster analysis to investigate the social structure of Colombo based on an integrated spatial database. This chapter completes the discussion of the major findings of the cluster analysis. The next chapter concludes the study by summarizing the entire work.

CHAPTER SEVEN
SUMMARY AND CONCLUSIONS

Introduction

This thesis has attempted to further our knowledge by filling a gap in the factorial analysis approach and by demonstrating how urban geographers can apply simple raster GIS techniques to bring incompatible data sets into a compatible format for their studies. This approach allows a researcher to access a large set of variables which can then be manipulated to extract the best set of variables for defining the different sources of social variation in a city. As we saw in the preceding chapters, urban geographers have attempted to explain and describe the social geography of cities by applying different techniques. Many of the studies have been based on a single source of information, mainly census tracts. The demerits of a single source of information have been highlighted by many researchers in this field, but the problem has not been solved successfully. The first section of this concluding chapter provides a synthesis and discussion of the major findings of the research. In this section, an overview of the spatial database integration is given and the findings of factor and cluster analysis are discussed. The next section details the limitations of the present study and offers suggestions for future research.

Synthesis and Discussion

The key component of the research was to develop an integrated spatial database for the study of the social geography of Colombo, Sri Lanka. The four different data sets selected were incompatible with each other. The problem that the researcher confronted was more complicated than the ones that other previous researchers faced. Whereas the other researchers were only challenged with different sizes of one spatial unit, polygons, the present researcher had to deal with different types of spatial units as well.

One of the researcher's purposes was to explore the suitability of the selected inexpensive raster GIS software for integrating incompatible spatial databases. The researcher was able to find appropriate solutions to various problems, deal with digitized spatial data, and integrate them into a GIS environment. The city boundary was formulated from two maps, Wards and GSDs, by applying GIS tools and adequately representing the area under the present city. The area under the estimated city boundary was compared with the total area under the city limit given in the Report on the Survey of Squatter Settlements in the City of Colombo (Colombo Municipal Council, 1990). It was found that 98.1 % of the total area given in the report was covered by the new city limit as designed by the researcher.

The technique developed to deal with data referenced to street segments is appropriate. This is the first time in factorial studies that the data referenced to street segments have been utilized. The researcher suggests that polygons generated based on the road segments for allocating data from that source are meaningful. The appropriateness of this procedure was explored by **OVERLAYING** street segments over new polygons. The new polygons are similar to the Voronoi line polygons. Therefore, they adequately represent the data source.

The socio-economic data, which were gathered from the four sources, were only allocated to the areas under residential land use. Therefore, they appropriately represent the social variation of the city. Allocating data like this, to areas under residential land use only, was not done in previous studies, but it has been highly recommended by many researchers (Martin, 1991; Carruthers, 1985; and Bracken and Martin, 1989).

The amount of spatial autocorrelation in the final output should be minimal because of the observation selection procedure. This was achieved by generating **"UNIQUE ZONES,"** by applying Raster GIS operations on all the data layers overlaying each other, and by taking one pixel from each to represent a unique spatial combination. Since only one pixel is picked up from each zone, it is expected that autocorrelation is greatly reduced in the statistical analysis. Considering the simplicity and the outcomes of the analyses, the raster GIS approach should be considered as good as vector GIS approach for urban studies. The current study proposes that factorial studies in combination with GIS give simple solutions for complicated problems that have not been addressed adequately.

Exploring the social geography of Colombo, based on the integrated spatial database, was the second objective of the study. The six factors extracted from the factor analysis accounted for 75.1 percent of the initial variance of the data set. Seventeen of 24 variables had communality over 73 percent, indicating the factor solution is strong. Furthermore, the factors extracted are qualitatively meaningful in terms of the variables loaded to each factor. In addition, Cronbach's alpha coefficients for all the factors indicated that all the factors demonstrate a strong degree of internal consistency.

The factors extracted demonstrate the recent trends of the social geography of Colombo. Due to the nature of the variables selected, most of the social dimensions are different from previous studies in developing world cities. Since the researcher was able to integrate different data sets, specific information on certain social groups could be included in the analysis to emphasize their role in shaping the city. For example, detailed information on low income people, foreign employment, and educational resources are not covered by general censuses, but this study is full of such information. This procedure helped the researcher to reveal the latest and appropriate sources of social variation.

Quality of Unemployed Labor Force, i.e., strength of the unused labor, was found to be the most important factor. High levels of unemployed labor possessing high levels of academic or technical qualifications were found in the areas where the rich are highly concentrated. This outcome was expected due to the fact that the poor do not have a good educational background. The spatial distribution of **Poverty** map and the results of multiple regression analysis reveal that poverty areas do not have a strong association with the *Distance to the city center*, *Distance to the nearest public health center* and *Distance to the nearest public school* distance from the city center. The spatial distributions of **Ethnicity** and **Poverty** reveal that Sinhalese, the majority ethnic group, are dominant in those areas where the rich are highly concentrated. This situation was unexpected. Moors are concentrated around the city center. The **Educational Resources** factor was helpful to understand the spatial variation of the level of educational resources. Surprisingly, this factor does not confirm that the poor are disadvantaged in terms of educational resources. The **Nature of Foreign Employment** factor demonstrates the recent changes in the social geography of Colombo.

Multiple regression analysis disclosed the importance of distance variables explaining the variation of the factors extracted. The distance variables play a significant role explaining the variation of all the factors. But the contribution of these distance measurements on **Poverty** is very low. It indicates that Poverty cannot be explained successfully by the selected independent variables. Therefore, other alternatives should be explored to more adequately explain this phenomenon. Furthermore, the test of regression coefficients proves that one regression coefficient and the intercept of the multiple regression model generated for the poverty dimension are not significant to accept the model. Excluding **Poverty** and **Ethnicity** dimensions, the remainder of the extracted multiple regression models are valid with high level of significance (p = 0.0001).

Cluster analysis helped the researcher to explore the ***combination*** of different sources of social variation and to understand the characteristics of different social groups in Colombo and their spatial distribution patterns. The four-cluster solution was the best solution in terms of the criteria applied. The four-clusters produced were: ***The Poorest People of the City, Foreign Employment Oriented, Elite people (Exclusive Residential Area),*** and ***Employment Oriented.***

Poor people are distinct in terms of **Poverty, Nature of Employment**, and **Ethnicity**. One interesting finding was that the poor people have better access to public education than does the wealthiest social group. This indicates that the poorest people are not disadvantaged in terms of access to the educational resources that are used in the analysis. The poor are dominant at the north-eastern part of the city and around the CBD of the city. This social group with highest incidence of foreign employment is a clear new dimension found from this study, reflecting the current social trends in the city. The social status of this new social group is higher than the previous social group.

The elite of the city are mainly Sinhalese. They are the largest ethnic group of the city with, unexpectedly, the highest level of skilled unemployment. Surprisingly, the lowest level of public educational resources is recorded in the areas where the wealthiest people reside. This is perhaps due to a overwhelming preference for private schools. The final social group, ***Employment Oriented***, has distinct characteristics in terms of employment and educational resources since the city's highest level of permanent employment and educational resources are found in this group. This is clearly a sub-group of the elite and is also dominated by Sinhalese.

The Poorest People of the City and ***Elite People*** social groups are evident in many studies in the developing and developed world cities, but the spatial distribution of these two groups and the characteristics of the ***Elite People*** group are unique to Colombo (see Chapter Five). ***Foreign Employment Oriented*** displays the recent changes in the social fabric of the city and this group is unique to the present study. As indicated in the previous chapter, characteristics of this social group indicate that this group may well be a social upgrade from ***The Poorest People of the City.***

Limitations of this Study and Suggestions for Future research

The current study does not cover Urban Colombo, which includes the suburban satellites, Kotte, Dehiwela Mt-Lavinia, and Moratuwa urban municipalities. The researcher understands that the peripheral urban municipalities should be included in order to generate a more complete picture of the city since peripheral urban centers strongly influence the growth of the city. The area under the Colombo Municipal limit was considered as the study area, since there were no appropriate data available for the suburbs. The Colombo Municipality still plays a major role in terms of Colombo Metropolitan growth. As highlighted in Chapter Two, the city defined by municipal limits contributes almost 60% to the total population of Urban Colombo, and the population density of Colombo is increasing constantly.

One of the biggest problems that the current study faced was that none of the maps were in digital form. Furthermore, some were not based on any geographic coordinate system and the city boundary did not coincide on some maps. Manual digitizing, converting data into computer readable format, and bringing them into the GIS environment were major challenges of the research. Since the Census Department of Sri Lanka does not have a census tract map, even in hard copy format, the researcher was unable to utilize fine data given at the census tract level. Therefore, the current study strongly proposes that the Department of Census and Statistics of Sri Lanka develop a census tract map of Colombo and that the census data should be stored, at least at census tract level. Furthermore, the government should take immediate steps to computerize the information collected by different agencies where it is possible. It is encouraging to note that at this time, The Survey Department of Sri Lanka has embraced GIS, and is presently converting a series of Colombo maps into digital format. Future studies will benefit from this procedure.

This study is lacking very fine data such as census enumeration or census tract level data. Two of the data sets used in the study, Wards and GSD, are highly aggregated.

The only detailed data set was the one given by street segments. The changing pattern of social dimensions and social groups should be studied using disaggregated level data. Notwithstanding these limitations the current study has been able to offer significant insight into the spatial structure of the city of Colombo.

Factor dimensions and social groups identified for Colombo mainly depend on the variables selected and the techniques applied for the study; therefore, these results may not reflect the realistic image of the city. New sources of social variation, such as land use and value, zoning, criminal activity and environmental qualities should be considered. This research may be the first of a series of studies for the exploration of the social geography of Colombo.

More research is also needed in uncovering more precisely the relationship between the extracted factors and distance measurements. As was apparent in Chapter Five, the association between some factors and selected distance measurements are not strong. For example, only 1.6 percent of **Poverty** is explained by the selected distance measures. Other distance measurements should be explored and included in the multiple regression analysis to see their contribution. As the poor areas are associated with low lying and marginal land use categories, distance to the Kelani river or Beira Lake (see Figure 2.4) or distance to the sea, industrial establishments or the nearest railway station should be investigated.

Integrating spatial databases by applying Raster GIS techniques opens a new avenue for future research in urban social geography. Planners, researchers, and policy agencies need more appropriate information to counteract the burgeoning urban problems of developing countries. Unfortunately, most of the developing world does not have well-planned database systems to provide crucial information. Therefore, there is a need for integrated spatial databases in the developing world. Once the data from various sources are allocated to pixels (data are allocated to 30m x 30m pixels in the present study), the latest information such as that gained from high-resolution satellite imagery can be integrated to produce large data sets. Lo has (1977, 1980, 1986, 1989, and 1995), demonstrated how socioeconomic data can be extracted from high-resolution satellite images and aerial photographs. The present research can be extended by combining information from satellite images to track the changing pattern of the social geography of Colombo.

The findings of this study are applicable not only to Colombo but to any city with reliable but isolated information sources which can be integrated to obtain more precise results. This study also reveals that when GIS is employed as a tool in our research studies, simple analytical techniques currently available can define patterns and solve very complicated problems with relative ease.

BIBLIOGRAPHY

Abeysekara, D., 1989: Urbanization and the Growth of Small Towns in Sri Lanka, *Papers of the East-West Population Institute*, Hawaii, 67, 4 - 32.

Abu-Lughod, J., 1966: Testing the Theory of Social Area Analysis: The Ecology of Cairo, Egypt, *American Sociological Review*, 34, 198 - 212.

Abu-Lughod, J., 1980: Rabat: Urban Apartheid in Morocco. Princeton, NJ: Princeton University Press.

Arachchige-Don, N., 1994: Patterns of Community Structure in Colombo, Sri Lanka: An Investigation of Contemporary Urban Life in South Asia, University Press of America, New York.

ARC/INFO, Environmental Systems Research Institutes, Redlands, CA.

Arnold, D.B. and Milne, W.J., 1984: The Use of Voronoi Tessellation in Processing Soil Survey Results, *IEEE*, 4, 22-30.

Aronoff, S., 1991: Geographic Information Systems: a Management Perspective, WDL Publications, Ottawa, Canada.

Bailey, T. and Gatrell, A., 1995: Interactive Data Analysis, Harlow, U.K., Longman Scientific Publishing.

Berry, B.J.L. and Karsada, J. D., 1977: Contemporary Urban Ecology, New York: Macmillan.

Berry, B.J.L. and Rees, P.H., 1969: The Factorial Ecology of Calcutta, *American Journal of Sociology*, 21, 537-543.

Boots, B.N. and Getis, A., 1988: Point Pattern Analysis, in Thrali, G., ed. Scientific Geography Series, Vol.8, Sage Publication, Inc., Beverly Hills.

Booth, C., 1903: Life and Labor of the People in London, Presidential Address, *Journal of the Royal Statistical Society*, 55(4), 557-91.

Boots, B.N., 1979: Weighting Thiessen Polygons, *Economic Geography*, 248-259.

Bourne, L.S. and Murdie, R.A., 1972: Interrelationships of Social and Physical Space in the City: A Multivariate Analysis of Toronto, *Canadian Geographer*, 3, 211 - 29

Bracken, I., and Martin, D., 1989: The Generation of Population Distribution from Census Centroid Data, *Environment and Planning A*, 21, 537 - 543.

Brand, R. R., 1972: The Spatial Organization of Residential Areas in Accra, Ghana, with Particular Reference to Aspects of Modernization. *Economic Geography*, 48, 284-98.

Burgess, E.W., 1925: "The Growth of the City', Park, R. E. et al. (1925), The City (The University of Chicago Press; Reprinted 1965), 37-44.

Burrough, P.A., and McDonnell, R.A., 1998: Principles of Geographic Information Systems, Oxford University Press, New York

Carter, H., 1995: The Study of Urban Geography, London, Arnold, 36 - 40.

Cattell, R.B. 1968: 'Higher Order Factor Structure', in Babks, C., and Broadhurst, P.L., (eds.), Studies in Psychology, University of London Press, 223-66.

Census and Statistics Department, 1990: Census Report for Colombo M.C., Sri Lanka

Census and Statistics Department, 1995: Statistical Report, Sri Lanka.

Center for Housing and Planning and Building, 1983: Slum and Shanty Upgrading in Sri Lanka, Center for Housing and Planning and Building, 23-34.

Clark University, 1997: IDRISI for Windows Tutorial Exercises Version 2.0, Clark University.

Colombo Municipal Council, 1990: Report on the Survey of Squatter Settlements in the City of Colombo, Colombo Municipal Council.

Daniel, Z.S., 1987: GIS and Urban Studies: Positivism, Post Positivism, and Beyond, *Urban Geography*, 15, 258 - 278.

Davies, W.K.D., 1975: A Multivariate Description of Calgary's Community Areas, *Western Geographical Series*, 2.

Davies, W.K.D., 1978 a: Charles Booth and the Measurement of Urban Social Structure, *Canadian Geographer*, 22, 273-97.

Davies, W.K.D., 1978 b: Alternative Factorial Solutions and urban Social Structure: A Data Analysis Exploration of Calgary in 1971, *Canadian Geographer*, 22 (4), 273 - 297

Davies, W.K.D., 1983: Urban Social Structure. A Multivariate-structural Analysis of Cardiff and Its Region, Cardiff: University of Wales Press.

Davies, W.K.D., 1984: Factorial Ecology, Aldershot: Gower

Davies, W.K.D., and Barrow, G.J., 1973: 'The Urban Dimensions of Leicester, England', in B. Clark and B. Gleave (eds.), 'Social Patterns in Cities', Institute of British Geographers Special Publication, 5, 71-86.

Davies, W.K.D. and Lewis, G.J., 1973: A Comparative Factorial Ecology of Three Canadian Prairie Cities, *Canadian Geographer*, 17 (4), 327-57.

Davies, W.K.D, and Murdie, R.A., 1991: Consistency and Differential Impact in Urban Social Dimensionality: Intra-urban Variations in the 24 Metropolitan Areas of Canada, *Urban Geography,* 12 (1), 55-79.

De Blij, Harm J. and Muller, Peter O, 1994: Geography : Realms, Regions, and Concepts, Wiley & Sons, New York.

Drakakis-Smith, D., 1992: The Third World City, Routledge, New York

Drysdale, R.L., 1979: Generalized Voronoi Diagrams and Geometric Searching Ph.D. Thesis, Department of Computer Science, Stanford University.

Drysdale, R.L., and Lee, D. T., 1978: Generalized Voronoi Diagram in the Plane, *Proceedings of the 16th annual Allerton Conference on Communications, Control and Computing*, 833-842.

Dykes, J. A., 1994: " Visualizing Spatial Association in Area-value data" in Worboys, (ed.), in Innovations in GIS (Selected Papers from the *First National Conference on GIS Research* UK), Taylor & Francis, 149 – 159.

Engels, F. (1845): The Conditions of the Working Class in England, Dover Publication, London.

Environmental Systems Research Institute, Inc., 1995: Understanding GIS, The ARC/INFO Method, ESRI, USA.

Fernando, M., 1993: This Piece of Planet Earth – Sri Lanka- and the Character of Its People, Institute of Integral Education, Piliyandala, Sri Lanka, 104 - 120

Fisher, P. F., 1991: 'Spatial Data Sources and Data Problems." in Maguire,D.J., et al.,(eds.), Geographical Information Systems, New York, Longman Scientific & Technical, 175- 89.

Flowerdew, R., 1991: "Spatial Data Integration" in Maguire, D.J et al., (eds.), Geographical Information Systems, New York, Longman Scientific & Technical, 375-87.

Fortune, S, 1986: A Sweepline Algorithm for Voronoi Diagrams, Algorithmica, 2, 153-174.

Franz, A., 1991: Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure, *ACM Computing Surveys*, 23 (3), 345 - 405.

Fulton, P.N. and Ingold, J.H., 1989: Highlights of Geographic Coverage and Content for the 1990 Census Data Products Program. In: Frazier, J.W., Epstein, B.J., Schoolmaster, F.A., (eds.) *Papers and Proceedings of Applied Geography Conferences*, 12, 215-17.

Gambini, R., 1966: A Computer Program for Calculating Lines of Equilibrium between Multiple Centers of Attraction, Center for Regional Studies, University of Kansas, Lawrence, Kan.

Girden, E.R., 1996: Evaluating Research Articles from Start to Finish, Sage Publications, London

Giulio, M., 1987: Raster versus Vector Data Encoding and Handling: A Commentary, *Photogrammetric Engineering and Remote Sensing*, 53 (10), 1397-1398

Goddard, J., and Kirby, A., 1976: An Introduction to Factor Analysis, CATMOG, No. 7, Geo Abstracts Ltd., University of East Anglia, Norwich.

Gold, C.M., 1991: Problems with Handling Spatial Data - The Voronoi Approach, *CISM Journal*, ACSGC, Spring 1991, 65 - 80.

Griffith, D.A and Amrhein, C.G, 1997: Multivariate Statistical Analysis for Geographers, Prentice-Hall, New Jersey.

Gunadasa, A.D., 1988: An Analysis of Urban Residential Patterns (Pertaining to the Town of Matara), Unpublished M.A. Thesis, University of Colombo, Sri Lanka.

Hamm, B., Currie, R., and Forde, D., 1988: A Dynamic Typology of Urban Neighborhoods, *Canadian Review of Sociology and Anthropology*, 25, 439-455.

Harman, H.H., 1976: Modern Factor Analysis, University of Chicago Press, Revised Edition.

Harman, H.H. and Jones, W.H, (1966): 'Factor Analysis by Minimizing Residuals,' *Psychometrica,* vol. 31: 351-68.

Harris, C.D, and Ullman, E.L. 1945: The Nature of Cities, *Annals of American Academy of Political Science*, 2242, 7-17.

Harris, C.W. and Kaiser, H.F, (1964): 'Oblique Factor analytic Solutions by Orthogonal Transformations,' *Psychometrica,* vol.29: 347-62

Hemachandra, (1988): Low Income Housing in Colombo City, *Economic Review*, June, Colombo.

Hemachandra, 1991: Primary Education among Children in Slum and Shanty Settlements, *Economic Review*, 1, 24 - 32.

Herbert, D.T., 1968: Principle Component Analysis and British Studies of Urban Social Structure, *Professional Geographer*, 20, 280 - 283.

Herbert, D.T., 1977: 'An Areal and Ecological Analysis of Delinquency Residence: Cardiff 1966 and 1971', Tijdscgrift v. E.S. *Geografie*, 68, 83-99

Horton, R.E., 1917: Rational Study of Rainfall Data Makes Possible Better Estimates of Water Yield, Eng. News-Record, 211-13

Hoyt,H., 1939: The Structure and Neighborhoods in American Cities, Washington, Federal Housing Administration.

Hunter, A. A. and Latif, A.H., 1973: "Stability and the Change in the Ecological Structure of Winnipeg', *Canadian Review of Sociology and Anthropology*, 6, 167-78.

Imai, H., Iri., and Murota, K, 1985: Voronoi Diagram in the Laguerre Geometry and its Applications, *SIAM Journal of Computing*, 14 (1), 93-105.

Johnston, R.J., 1973: "Residential Differentiation in Major New Zealand Urban Areas: A Comparative Factorial Ecology", *Institute of British Geographers*, Special Publication, 5, 143 - 168.

Johnston, R.J., 1976: Residential Characteristics and Residential Area Homogeneity: Further Thoughts on Extensions to the Factorial Method, *Environment and Planning A*, 8, 543-52.

Jones, F.L., 1969: Dimension of Urban Social Structure, Canberra, Australian National University Press.

Kasarda., J.D. and Berry., B.J.L., 1977: Contemporary Urban Ecology, Macmillan Publishing Co., Inc. New York, 108 - 178.

Keith, C.C., 1997: Getting Started with Geographic Information System, Prentice Hall, New Jersey.

Kokubo, I, 1985: Computational Methods of Generalized Voronoi diagrams, Master's Thesis, Department of Mathematical Engineering and Information Physics, University of Tokyo.

Laksman, S.Y., 1991: Is GIS Appropriate Technology, *International Journal of Geographic Information Systems*, 5 (1), 41-58.

Levine, M.S., (1977): Canonical Analysis and Factor Comparison, London: Sage, University Papers

Le Bourdais, C. and Beaudry, M., 1988: The Changing Social Structure of Montreal: 1971-81, *The Canadian Geographer*, Vol. 32, 98-113.

Lo, C.P., 1975: Changes in the Ecological Structure of Hong Kong 1961 - 1971: A Comparative Analysis, *Environment and Planning A*, 7, 941 - 963.

Lo, C.P., Pannel, C.W. and Welch, R., 1977: Land use Changes and City Planning in Shenyang and Canton, *Geographical Review*, 67, 268 - 283.

Lo, C.P., 1980: Shaping Socialist Chinese Cities - A Model of Form and Land Use. C.K. Leung and Norton Ginsberg, eds. China: Urbanization and National Development. Department of Geography, University of Chicago, Research Paper No. 196, 130 - 155.

Lo, C.P., 1986: The Evolution of the Ecological Structure of Hong Kong: Implications for Planning and Future Development, *Urban Geography*, 7, 311 - 335.

Lo, C.P., 1995: Automated Population and Dwelling Unit Estimation from High-Resolution Satellite Images: A GIS Approach, International Journal of Remote Sensing, Vol. 16, 1, 17-34.

Maguire, D.J., Goodchild, M.F., Rhind, D.W., 1991: Geographical Information Systems: Principals and Applications, Longman Scientific & Technical, New York

Mayhew, H., 1862: London Labor and London Poor, 1-4, Dover Publications, London.

Martin, D., 1991: Geographic Information Systems and Their Socioeconomic Applications, London, 148 - 160.

Mather, P.M., 1976: Computational Methods of Multivariate Analysis in Physical Geography, London, Wiley and Sons.

Mather, P.M., 1993: Geographical Information Handling: Research and Applications, New York, Wiley and Sons.

McElrath, D.C., 1968: 'Societal Scale and Societal Differentiation, Accra, Ghana', in S. Greer, et al. (eds.), The New Urbanization, St. Martin's Press, New York, 33-52.

McLain, D.H., 1976: Two Dimensional Interpolation from Random Data, *Computer Journal*, 19, 178-81.

Merchant, D.C., 1987: Spatial Accuracy Specifications for Large Scale Topographic Maps, *Photographic Engineering and Remote Sensing* 53, (7): 958-61.

Michael, A. C., Blair, A. R., Joseph, S.T., Allison, M. G., 1996 (eds.): Urban Future: Global Pressures and Local Forces, The Woodrow Wilson Center Press, Washington.

Ministry of Education, 1986: Primary Education among Slum and Shanty Dwellers, Non-formal Education Branch of the Ministry of Education, Ministry of Education, 25 - 45.

Ministry of Housing and Construction, 1991: Shelter in Sri Lanka, Ministry of Housing and Construction, 3- 39

Ministry of Policy Planning and Implementation, 1990: Resource Requirements and Economic Activities, Colombo Divisional Secretary's Division, Colombo.

Mollison, D., 1977: Spatial Contact Model for Ecological and Epidemic Spread, *Stat. Soc. B,* 39, 283-326.

Morris, F. B., and Pyle, G.F., 1972: The Social Environment of Rio de Janeiro in 1960, *Economic Geography*, 48, 286 -99

Murdie, R.A., 1969: Factorial Ecology of Metropolitan Toronto, 1951- 61, Department of Geography, University of Chicago, Research Paper No.116.

Murdie, R.A., 1980: Factor Scores: A Neglected Element of Factorial Studies, *Urban Geography*, 14, 295 - 316.

Murdie, R., 1987: Social Geography of Metropolitan Toronto : 1951-81, Paper Presented to the Annual Meeting of the Canadian Regional Science Association, Hamilton, Ontario.

National Housing Development Authority, 1987: Case Study on Land Tenure Regulations in Wanathamulla Squatter Upgrading Projects, Sri Lanka.

National Housing Development Authority of Sri Lanka, 1991: Development of Urban Low Income Settlements in the City of Colombo, National Housing Development Authority, 1 - 7.

Newton, P.W and Johnston, R.J, 1976: Residential Characteristics and Residential Area Homogeneity: Further Thoughts on Extensions to the Factorial Ecology Method, *Environment and Planning A*, 8, 543-52.

O'Brien, A., 1993: Introducing Quantitative Geography: Measurement, Methods and Generalized Linear Models, Routledge, London.

Okabe, A., Boots. B., and Sugihara. D., 1992: Spatial Tessellations, Concepts and Applications of Voroni Diagrams, John Wiley & Sons, New York.

Okabe, A., Yoshikawa, T., Fujii, A., and Oikawa, K., 1988: The Statistical Analysis of a Distribution of Active Points in Relation to Surface-like Elements, *Environmental and Planning A*, 20, 609-620.

Pannell, C.W., 1980: The Internal Structure of Chinese Cities: C.K. Leung and Norton Ginsberg, eds., China Urbanization and National Development. Department of Geography, University of Chicago, Research Paper 196, 187 - 214.

Park, R.E, 1925: The City as a Social Laboratory in T.V. Smith and L. White (eds.), Chicago: An Experiment in Social Science Research, University of Chicago Press, 46-63.

Peoples' Bank of Sri Lanka, 1994: Metropolitan Development, *Economic Review*, April, 1994.

Perera, L.K. and Tateishi, R, 1995: Do Remote Sensing and GIS Have a Practical Applicability in Developing Countries? (Including Some Sri Lankan Experience), *International Journal of Remote Sensing*, 16 (1), 35-51.

Prithvish, N., 1987: A Proposed Base for a Geographical Information System for India, International Journal of Geographical Information System, 2, 181-187

Rao, R.C, (1955): 'Estimation and Test of Significance in Factor Analysis,' *Psychometrika*, Vol. 20: 93-111

Rees, P.H., 1979: Residential Patterns in American Cities: 1960, Department of Geography, University of Chicago, Research Paper No. 189.

Rhind, D.W., Green, N.P.A., Mounsey, H.M., Wiggins, J.C., 1984: The Integration of Geographical Data. Proceedings of Austra Carto, Perth, Australian Cartographic Association, Perth, 81-94.

Rummel, R. J., 1971: Applied Factor Analysis, Northwestern University Press.

Schmid, C.F. and Tagashira, K., 1964: Ecological and Demographic Indices: A Methodological Analysis, *Demography*, 1, 194-211.

Shaw, G., and Wheeler, D., 1994: Statistical Techniques in Geographical Analysis, Halsted Press

Shevkey, E., and William, M., 1949: The Social Areas of Los Angeles, University of California Press.

Shevkey, E., and Bell, W., 1955: Social Area Analysis, Stanford, Stanford University Press.

Sjoberg, G., 1960: The Pre-industrial City, Past and Present, The Free Press, U.S.A

Snyder (1962): Urban Places in Uruguay and the Concept of a Hierarchy. Festschrift: C.F. Jones, *Northwestern University Studies in Geography*, 6, 29-46.

SPSS Inc., 1996: SPSS Advanced Statistics 7.0, Chicago, 145 - 158

Sundeep and Geoff., 1996: Implementation of GIS in India: Organizational Issues and Implications, *International Journal of GIS*, 4, 385-404

Survey Department of Sri Lanka, (1991): A to Z Colombo Atlas and Street Index, Colombo.

Sweetser, F.L., 1965b: 'Factorial Structure as Ecological Structure in Helsinki and Boston', *Acta Sociologica*, 8, 205-25

Tabachnick, B.G, and Fidell, L.S, 1996: Using Multivariate Statistics, Harper & Row Publishers, New York.

Thiessen, A.H., 1911: Precipitation Averages for Large Areas, *Monthly Weather Review* 39, 1082 - 1084.

Timms, D.W.G., 1971: The Urban Mosaic: Towards a Theory of Residential Differentiation, Cambridge: Cambridge University Press.

Tryon, R.G., 1955: The Identification of Social Area by Cluster Analysis, University of California Press, Berkeley.

Urban Development Authority of Colombo, 1992: The Land Use Map of Colombo, Urban Development Authority of Colombo, Sri Lanka

United Nations Center for Human Settlements (HABITAT), 1996: An Urbanizing World: Global Report on Human Settlements, 1996, Oxford University Press.

Webber, R.J., 1977: The National Classification of Residential Neighborhood, Planning Research Application Group, Paper No.23, Center for Environmental Studies, London.

White, M., 1987: American Neighborhood and Residential Differentiation, New York, Sage Foundation

Worboys, M.F., 1994: Innovations in GIS: Selected Papers from the First National Conference on GIS Research UK, Taylor and Francis Ltd., Bristol.

Yap, C.K, 1987: An O (n log n) Algorithm for the Voronoi Diagram of a Set of Simple Curve Segments, *Discrete and Computational Geometry*, 2, 365-393.

Yeh, A.G.O., 1991: The Development and Applications of Geographic Information Systems for Regional Planning in the Developing Countries, *International Journal of Geographic Information Systems*, 5 (1), 5-27.

Yeh, A.G.O, Xu, X., and Hu, H., 1995: The Social Space of Guangzhou City, China, *Urban Geography*, 16 (7), 595-621.

Appendix 1

Variables Available at the Ward Level

Section I

I. *Population under five years of age*

II. *Number of Sinhalese*

III. *Number of Sri Lankan Tamils*

IV. *Number of Indian Tamils*

V. *Number of Moors*

VI. *Number of Burghers*

VII. *Number of Malay*

VIII. *Number of Others*

IX. *Total Population*

X. *Number of Buddhists* *(Religion)*

XI. *Number of Hindus* (,,)

XII. *Number of Muslims* (,,)

XIII. *Number of Christians* (,,)

Appendix 2
All the Variables from Four Different Information Sources

Ward Level Data:

I. *Percentage of Sinhalese (SINPER)*

II. *Percentage of Tamils (TAMPC)*

III. *Percentage of Ceylon Moors (MOOPC)*

IV. *Percentage of population under five years of age (POPFIV)*

V. *Shanty population density per hectare (SSPOPDN)*

VI. *Percentage of population under 6 years (LESSIX)*

VII. *Percentage of population between 6 - 17 years (SXSEVEN)*

VIII. *Percentage of population 18 - 55 years (FIFTYFIVE)*

IX. *Percentage of population over 56 years (OVRFFTYX)*

X. *Percentage of shanties reached (REACHPER)*

XI. *Percentage of shanties to be reached (TOREACH)*

Grama Sevaka Level Data:

XII. *Percentage of population less than 5 years of age (LESSPIVE)*

XIII. *Percentage of population 5 - 18 years of age (EIGTEEN)*

XIV. *Percentage of population 19 - 29 years of age (TWEYNINE)*

XV. *Female/male employment ratio (female employment / male employment) (EMRATIO)*

XVI. *Percentage of male employment of the total work force (MALEPERM)*

XVII. *Percentage of female seasonal employment of the total work force (FSEASO)*

XVIII. *Percentage of male seasonal employment of the total work force (MSEASON)*

XIX. *Female/ male unemployment ratio (female unemployment / male unemployment) (UNEMRAT)*

XX. *Percentage of male unemployment of the total work force (MALUN)*

XXI. *Percentage of female foreign employment of the total work force (FFORINEM)*

XXII. *Percentage of male foreign employment (MFORIN)*

XXIII. *Percentage of employment in high income group (engineers, doctors, teachers, clerks, technicians) (EMPH)*

XXIV. *Percentage of employment in low income categories (carpenters, laborers, trade persons, and other minor income groups) (EML)*

XXV. *Number of unemployed people with a degree Qualification per 10,000 unemployed people (male and female) (UNEMDEGR)*

XXVI. *Number of unemployed people with Advanced Level Qualification (grade 13) per 1,000 unemployed people (male and female) (ALUNEMP)*

XXVII. *Number of unemployed people with Ordinary level (grade 11) Qualification per 1,000* *unemployed people (male and female) (OLUNEMP)*

XXVIII. *Number of unemployed people with 8th grade standard education per 1,000 unemployed people (male and female) (EIGHTUNM)*

XXIX. *Number of unemployed people with technical education per 1,000 unemployed people (male and female) (UNEMPTEC)*

XXX. *Number of schools per 10,000 students (SCHOOLS)*

XXXI. *Number of students per teacher (STDTEA)*

XXXII. *Average school building square feet per student (BLDGSTUD)*

XXXIII. *Number of student desks per 10,000 students (DESKPER)*

XXXIV. *Number of student chairs per 10,000 students (CHAIRS)*

XXXV. *Trained Mathematics and Science teachers per 10,000 students (TMSTEACH)*

XXXVI. *Trained Arts Teachers per 10,000 students ((TARTSPER)*

XXXVII. *Trained Commerce Teachers per 10,000 students (TCOMTEC)*

XXXVIII. *Science and Mathematics graduate teachers per 10,000 students (GRADSCE)*

XXXIX. *Arts and Music graduate teachers per 10,000 students (GRADARTS)*

XL. *Commerce and other graduate teachers per 10,000 students (GRADCOM)*

XLI. *Untrained English teachers per 10,000 students (UNENGLIS)*

XLII. *Untrained Dance, Music, other, and Sports teachers per 10,000 students (UNDANCE)*

Street Level Data

XLIII. *Average number of years of residency less than 10 years (RESITEN)*

XLIV. *Average number of years of residency 10 - 20 years (RESTNTWE)*

XLV. *Average number of years more than 20 years (RESTUWEN)*

XLVI. *Percentage of good housing units (GHOUSE))*

XLVII. *Percentage of renovated buildings (UPGRA)*

XLVIII *Percentage of slums and shanties (SHANPC)*

Distance Measurements:

XLIX. *Distance to the City Center (CENDIS)*

L. *Distance to the Nearest Public School with Science for Advanced Level (SCHDIS)*

LI. *Distance to the nearest Government Health Center(HELTHDIS)*

APPENDIX 3

CORRELATION MATRIX

	ALUNEMP	BLDGSTUD	CHAIRS	DESKPER	EIGHTUNM	EIGTEEN	EML
ALUNEMP	1.00000						
BLDGSTUD	-.13981	1.00000					
CHAIRS	.01615	.29773	1.00000				
DESKPER	.02479	.32252	.90671	1.00000			
EIGHTUNM	.04427	.15923	.09538	.01472	1.00000		
EIGTEEN	-.03786	.14701	-.30238	-.31973	-.16191	1.00000	
EML	.03354	-.34726	-.31373	-.34657	-.15666	.40246	1.00000
EMPH	-.02536	.23556	.24553	.29457	.05566	-.38546	- .98761
EMRATIO	-.23112	-.08154	.13312	.16208	-.09700	-.10230	.18644
FFORINEM	-.10017	.05386	-.15268	-.09142	-.21080	.44766	.09771
FIFTYFIV	.25972	-.21364	-.11882	-.24713	-.04590	.12338	.12172
FSEASO	-.06894	-.30460	-.10374	-.09585	-.24322	.22678	.47676
GHOUSE	-.20910	.38611	.32629	.32564	.19456	-.31119	-.74474
GRADARTS	-.15733	.34104	.13467	.32805	.19344	-.07319	-.25746
GRADCOM	-.11507	.48416	.41692	.37372	.15291	-.36960	-.52946
GRADSCE	-.01892	.41141	.33803	.23983	.29207	-.34252	-.31233
LESSPIVE	-.08991	-.10690	.23686	.27294	.38358	-.44966	.80862
LESSSIX	-.04375	.06133	-.01440	.04281	-.02871	.05819	.88478
MALEPERM	-.07478	.39397	.13271	.02787	.27302	-.12282	-.47114
MALUN	.31200	-.40071	-.07722	-.03181	-.15832	-.16760	.43671
MFORIN	-.05023	.03795	-.27811	-.21494	-.38224	.25035	.24293
MOOPC	.21941	-.30659	-.36379	-.40736	-.06845	.46269	.74432
MSEASON	.05335	-.20185	-.14319	-.11397	.00882	.06642	.30639
OLUNEMP	.85664	-.26564	.02870	.02756	.10587	.01330	.20327
OVRFFTYX	.00048	-.16355	-.10872	-.30081	-.02444	.03902	-.01194
POPFIV	.21221	-.20659	-.12359	-.38736	-.06523	.36269	.64432
REACHPER	-.14060	.00971	.41702	.21026	.13649	-.10121	-.14308
RESITEN	.05470	.31682	-.16171	-.23615	.04591	.32445	.03593
RESTNTWE	-.16050	-.34881	.20218	.28844	-.06497	-.60377	-.08016
RESTUWEN	.22870	.11415	-.14780	-.15170	.07316	.27332	.00380
SCHOOLS	.16598	.41632	.46765	.34854	-.04232	.07284	.01586
SHANPC	.25637	-.40055	-.16912	-.22820	-.12109	.12888	.71358
SINPER	-.04303	.06894	.35169	.27248	.26936	-.32855	-.39270
SSPOPDN	-.01152	-.11284	-.09239	-.09973	-.05990	.09185	.08605
STDTEA	.25426	.09815	.41145	.30305	-.01170	.15434	.35566
SXSEVEN	-.16758	.15482	.16296	.27460	.06193	-.20321	-.19700
TAMPC	.05729	.05972	-.14364	-.13860	-.27926	.11298	.31152
TARTSPER	.06848	.15790	.17794	.18351	-.03840	.12313	-.17682
TCOMTECH	.10120	.46445	.03969	.13402	.02592	.01935	-.37892
TMSTEACH	-.21341	-.07281	.29270	.43909	.02438	-.68163	-.38846
TOREACH	.22975	-.17266	-.22729	-.24805	-.20966	.27719	.43040
TWEYNINE	-.09146	.26183	.18960	-.00331	-.05820	.22664	.02621
UNDANCE	-.01141	.36769	.46590	.35554	.07986	.03137	-.16533
UNEMDEGR	.84565	-.19831	.06090	.07600	-.90178	-.07794	.14685
UNEMPTEC	.79005	-.21855	.04234	.04503	.67890	-.04287	.24705
UNEMRAT	-.02781	-.32338	-.16385	-.07732	-.06689	.25144	.36119
UNENGLIS	-.22802	.74995	.39506	.35748	.30629	.11266	-.31792
UPGRA	.08152	-.23452	-.38379	-.31851	-.20212	.39938	.51002

	EMPH	EMRATIO	FFORINEM	FIFTYFIV	FSEASO	GHOUSE	GRADARTS
EMPH	1.00000						
EMRATIO	-.18644	1.00000					
FFORINEM	-.09771	-.11464	1.00000				
FIFTYFIV	-.12172	-.29453	.01616	1.00000			
FSEASO	-.47676	.09004	-.00963	.24667	1.00000		
GHOUSE	.74474	-.15525	-.04037	-.12476	-.51005	1.00000	
GRADARTS	.25746	.13122	.29386	-.37561	-.00187	.23530	1.00000

GRADCOM	.52946	-.12563	.03002	-.04226	-.54275	.62088	.09822
GRADSCE	.31233	-.18024	-.25752	-.00156	-.47882	.41957	-.23028
LESSPIVE	.20862	.11966	-.12748	-.20933	-.23587	.25304	.24371
LESSSIX	-.08478	-.02583	.01639	-.46138	-.19189	-.03007	.04434
MALEPERM	.47114	.05451	-.17273	-.22055	-.56025	.50243	-.04727
MALUN	-.43671	.25636	.01339	.08476	.25897	-.59958	-.00850
MFORIN	-.24293	-.25229	.51946	.29633	.09579	-.19913	.04457
MOOPC	-.74432	.05225	.08634	.03958	.16699	-.69600	-.30603
MSEASON	-.30639	-.02153	.03474	.30004	.55304	-.36677	.11971
OLUNEMP	-.20327	-.00524	-.13602	.17255	.20472	-.43536	-.08840
OVRFFTYX	.01194	-.13159	-.20186	.54282	-.00955	-.03380	-.39325
POPFIV	-.68452	.03255	.04554	.01121	.23399	-.58800	-.28603
REACHPER	.14308	-.11308	-.13895	-.09225	-.26367	.26409	-.23945
RESITEN	-.03593	-.32602	.25097	.02688	-.20509	.08833	-.12400
RESTNTWE	.08016	.09194	-.12187	-.14932	-.10029	.16590	.06324
RESTUWEN	-.00380	-.00504	-.00922	.10943	-.14887	.02624	.08275
SCHOOLS	-.01586	.06601	-.18819	-.03119	.02594	.02492	-.06362
SHANPC	-.71358	.08907	.06973	.17957	.42749	-.84395	-.19849
SINPER	.39270	-.15409	-.03389	-.01538	-.13390	.40972	.00427
SSPOPDN	-.08605	.03759	-.00600	.02821	.08555	-.10548	-.04073
STDTEA	-.35566	.00402	.05587	.17049	.20148	-.36016	-.00245
SXSEVEN	.19700	.29903	.03564	-.48316	.00633	.16415	.38509
TAMPC	-.31152	.21732	.03207	.17234	.10680	-.20259	.02612
TARTSPER	.17682	-.04740	.22011	-.13669	.04596	.02157	.31375
TCOMTECH	.37892	.07467	.16180	-.07719	-.26500	.29851	.52641
TMSTEACH	.38846	.09428	-.03294	-.12921	-.09061	.34959	.35542
TOREACH	-.43040	.14197	.04801	.17079	.14196	-.36882	-.20892
TWEYNINE	-.02621	-.26214	-.02762	.00868	-.10707	-.01885	-.30260
UNDANCE	.16533	-.21233	.06909	-.05218	-.26237	.30902	.01649
UNEMDEGR	-.14685	-.13411	-.13604	.10784	.05190	-.32879	-.05924
UNEMPTEC	-.24705	.11177	-.12445	.08584	.07422	-.42491	-.00750
UNEMRAT	-.36119	.41760	-.10042	-.11096	.36857	-.51410	-.01348
UNENGLIS	.31792	.03846	-.06442	-.21255	-.13174	.43164	.40639
UPGRA	-.51002	.17006	-.00707	.02263	.41683	-.81109	-.19164

	GRADCOM	GRADSCE	LESSPIVE	LESSSIX	MALEPERM	MALUN	MFORIN
GRADCOM	1.00000						
GRADSCE	.70732	1.00000					
LESSPIVE	.10611	-.06504	1.00000				
LESSSIX	-.08089	-.11581	.90522	1.00000			
MALEPERM	.49150	.38472	.13483	.08683	1.00000		
MALUN	-.40337	-.33532	.16461	.11158	-.53477	1.00000	
MFORIN	-.10411	-.12934	-.17081	-.01739	-.47205	.26049	1.00000
MOOPC	-.46831	-.25214	-.18854	.20443	-.15884	.38764	.20353
MSEASON	-.39303	-.34015	.04988	-.09626	-.70093	.42183	.39888
OLUNEMP	-.34607	-.17698	-.10923	-.14004	-.23786	.50419	-.11863
OVRFFTYX	.11392	.26849	-.31805	-.35358	.10813	-.04585	-.05014
POPFIV	-.35431	-.15214	.90228	.91267	-.18554	.28764	.14453
REACHPER	.15204	.28005	.29011	.28137	.03387	-.10136	-.13938
RESITEN	.00498	.15606	-.25181	-.02409	.30498	-.28412	.20938
RESTNTWE	.24037	.08038	.34170	-.05393	-.08738	.07707	-.02736
RESTUWEN	-.08355	-.07390	.09186	.22481	.06295	.01026	.05230
SCHOOLS	.07313	.25189	-.23927	-.02737	.21549	-.25702	-.26900
SHANPC	-.50918	-.34285	.80531	.89418	-.43503	.73000	.23946
SINPER	.34236	.17687	.41538	.09048	.19416	-.07331	-.25801
SSPOPDN	-.15149	-.12302	-.11801	-.09513	-.11065	.06439	-.01793
STDTEA	-.05101	.09632	-.23137	-.15371	-.32437	.34959	.24529
SXSEVEN	.09725	.03066	-.27751	-.50843	.04429	-.15394	-.18793
TAMPC	-.14273	-.04654	-.22854	-.07126	-.05259	.13884	.23290
TARTSPER	.05263	-.14360	-.10919	-.04680	.06524	-.16971	-.01666
TCOMTECH	.40387	.02438	.03633	.01920	.22574	-.09087	.09512
TMSTEACH	.37502	.15220	.37835	-.16596	-.10536	.10205	.07029
TOREACH	-.22210	-.24358	-.17660	.06648	-.17198	.09542	.11740
TWEYNINE	.10170	.30047	-.24028	.18411	.16871	-.19571	.01327
UNDANCE	.40085	.34178	.08425	.12309	.14531	-.31732	.07936

UNEMDEGR	-.25001	-.07345	.12674	.06177	-.19893	.54347	-.11448
UNEMPTEC	-.30803	-.14153	.08720	-.03447	-.19494	.59289	-.11686
UNEMRAT	-.50320	-.43702	-.08079	-.00865	-.24926	.18947	-.16612
UNENGLIS	.35868	.30320	.05005	.04361	.26485	-.39814	-.19257
UPGRA	-.52157	-.35429	-.38078	-.05271	-.39677	.24114	.08282

	MOOPC	MSEASON	OLUNEMP	OVRFFTYX	POPFIV	REACHPER	RESITEN
MOOPC	1.00000						
MSEASON	.10630	1.00000					
OLUNEMP	.32699	.21253	1.00000				
OVRFFTYX	.13920	-.10699	-.02500	1.00000			
POPFIV	.74572	.10630	.32699	.13920	1.00000		
REACHPER	-.16296	.01275	-.21153	.04387	-.16296	1.00000	
RESITEN	.13081	-.26648	-.05540	.06701	.13081	.02157	1.00000
RESTNTWE	-.21346	-.04599	-.19849	-.11487	-.21346	.07664	-.26490
RESTUWEN	.22354	.09770	.15109	.13016	.22354	-.03184	-.18984
SCHOOLS	-.01524	-.21132	.09638	.01634	-.01524	.24102	.20670
SHANPC	.65112	.34456	.42485	.03195	.65112	-.09103	-.05134
SINPER	-.50396	-.00129	-.05630	-.15364	-.50396	.42020	-.00850
SSPOPDN	.08818	.04535	.06298	.02356	.68818	-.00797	-.03804
STDTEA	.34094	.24548	.32526	.08831	.34094	.11273	-.04595
SXSEVEN	-.31858	-.09737	-.00729	-.40396	-.31858	-.21983	-.03730
TAMPC	.21129	.00684	-.02136	.14582	.21129	-.21361	-.02415
TARTSPER	.00169	-.03727	.13148	-.25532	.00169	-.08114	-.03668
TCOMTECH	-.25068	-.06983	-.04809	-.08286	-.25068	-.18370	-.05926
TMSTEACH	-.55107	.05414	-.22579	-.13966	-.55107	-.08125	-.32759
TOREACH	.34792	.13530	.22283	-.06502	.34792	-.13128	-.13602
TWEYNINE	.14762	-.21854	-.17318	.21163	.14762	.37974	.21075
UNDANCE	-.21750	-.17291	-.16560	-.11822	-.21750	.45865	.29764
UNEMDEGR	.29035	.20320	.92580	-.09802	.29035	-.08989	-.00514
UNEMPTEC	.35129	.21261	.92914	-.12705	.35129	-.16260	-.07199
UNEMRAT	.23197	.12404	.19339	-.10576	.23197	-.20655	-.22218
UNENGLIS	-.35982	-.04567	-.22390	-.07369	-.35982	.23095	.15515
UPGRA	.49701	.25846	.28915	.02844	.49701	-.35845	-.09832

	RESTNTWE	RESTUWEN	SCHOOLS	SHANPC	SINPER	SSPOPDN	STDTEA
RESTNTWE	1.00000						
RESTUWEN	-.50039	1.00000					
SCHOOLS	-.20787	-.03433	1.00000				
SHANPC	-.06045	-.04258	.01602	1.00000			
SINPER	.15328	-.29716	-.03752	-.17423	1.00000		
SSPOPDN	-.05507	-.02563	-.01829	.91232	-.09105	1.00000	
STDTEA	-.10504	.10443	.39885	.46793	-.26298	.00859	1.00000
SXSEVEN	.24384	-.36789	.02879	-.25619	-.00132	.05350	-.02374
TAMPC	-.16024	.43434	.25123	.24377	-.61100	-.03278	.31019
TARTSPER	-.23592	.12175	.22209	-.08950	.00628	.05458	.08376
TCOMTECH	-.21148	.36818	-.10108	-.29022	-.09041	-.05948	-.02501
TMSTEACH	.72229	-.33999	-.35938	-.26866	.21283	-.10888	-.13663
TOREACH	-.14814	.07248	.16032	.30533	-.22124	.05054	.13538
TWEYNINE	-.35595	.12121	.33473	.02609	-.08075	-.02556	.25641
UNDANCE	-.03702	-.02174	.33294	-.18759	.31628	-.09404	.16274
UNEMDEGR	-.14651	.16013	-.01421	.39794	.10194	.01110	.25912
UNEMPTEC	-.15904	.16997	.01202	.47003	-.01380	.03753	.33586
UNEMRAT	-.02583	-.00555	-.10097	.25821	-.36997	.10427	.06437
UNENGLIS	-.31147	.21948	.41467	-.44496	.17579	-.07626	.03880
UPGRA	-.22078	-.00006	-.06269	.37094	-.52093	.11754	.11078

	SXSEVEN	TAMPC	TARTSPER	TCOMTECH	TMSTEACH	TOREACH	TWEYNINE
SXSEVEN	1.00000						
TAMPC	-.11647	1.00000					
TARTSPER	.21237	-.09836	1.00000				
TCOMTECH	.07430	.22993	.31212	1.00000			
TMSTEACH	.36253	-.15810	-.23190	.07620	1.00000		

TOREACH	-.19606	.14411	.06673	-.07720	-.37864	1.00000	
TWEYNINE	-.28091	.12981	.23620	-.00277	-.47737	-.08901	1.00000
UNDANCE	-.03878	-.08494	.22284	.19321	-.06473	-.04099	.37619
UNEMDEGR	-.12414	-.13256	.01838	-.02753	-.14996	.06968	-.15940
UNEMPTEC	-.00934	-.00984	.05578	.01730	-.16933	.19892	-.21350
UNEMRAT	.11288	.13359	-.00073	-.10192	-.13083	.23352	-.15056
UNENGLIS	.13988	.02422	.17613	.40598	-.04757	-.27765	.16614
UPGRA	-.00625	.08241	.05873	-.20183	-.31311	.30272	.00370

	UNDANCE	UNEMDEGR	UNEMPTEC	UNEMRAT	UNENGLIS	UPGRA
UNDANCE	1.00000					
UNEMDEGR	-.08834	1.00000				
UNEMPTEC	-.14472	.93435	1.00000			
UNEMRAT	-.35636	.05243	.22732	1.00000		
UNENGLIS	.48282	-.16852	-.18869	-.26943	1.00000	
UPGRA	-.33129	.13438	.22221	.60821	-.26412	1.00000

The following pair of variables had very high correlation coefficients, showing redundancy:

Pair of Variables	Correlation Coefficient
1. *Number of student desks per 10,000 students (DESKPER) Vs Number of student chairs per 10,000 students (CHAIRS))*	0.90671
2. *Percentage of employment in high income group (engineers, doctors, teachers, clerks, technicians) (EMPH) Vs Percentage of employment in low income categories carpenters, laborers, trade persons, and other minor income groups) (EML)*	-0.98761
3. *Number of unemployed people with 8th grade standard education per1,000 unemployed people (male and female) (EIGHTUNM) Vs Number of unemployed people with a degree Qualification per 10,000 unemployed people (male and female) (UNEMDEGR)*	-0.90178
4. *Shanty population density per hectare (SSPOPDN) Vs Percentage of slums and shanties (SHANPC)*	0.91232
5. *Percentage of population under 6 years (LESSIX) Vs Percentage of population under five years of age (POPFIV)*	0.97267
6. *Percentage of population less than 5 years of age (LESSPIVE) Vs Percentage of population under five years of age (POPFIV)*	0.90228

Appendix 4

EXPLORATORY FACTOR ANALYSIS - STAGE 1

Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
		*				
ALUNEMP	.90409	*	1	8.76308	20.9	20.9
BLDGSTUD	.78677	*	2	4.20896	10.0	30.9
CHAIRS	.81481	*	3	3.10317	7.4	38.3
EIGTEEN	.81305	*	4	2.99899	7.1	45.4
EML	.84478	*	5	2.62854	6.3	51.7
EMRATIO	.72079	*	6	2.24405	5.3	57.0
FFORINEM	.64774	*	7	2.15646	5.1	62.2
FIFTYFIV	.79382	*	8	1.75987	4.2	66.3
FSEASO	.75048	*	9	1.36521	3.3	69.6
GHOUSE	.82866	*	10	1.22835	2.9	72.5
GRADARTS	.72791	*	11	1.20520	2.9	75.4
GRADCOM	.80827	*	12	1.06942	2.5	77.9
GRADSCE	.75824	*				
MALEPERM	.80422	*				
MALUN	.85782	*				
MFORIN	.78091	*				
MOOPC	.85004	*				
MSEASON	.70782	*				
OLUNEMP	.93880	*				
OVRFFTYX	.73467	*				
POPFIV	.85004	*				
REACHPER	.69398	*				
RESITEN	.76692	*				
RESTNTWE	.67031	*				
RESTUWEN	.76068	*				
SCHOOLS	.84600	*				
SHANPC	.66627	*				
SINPER	.80171	*				
STDTEA	.76668	*				
SXSEVEN	.66566	*				
TAMPC	.76597	*				
TARTSPER	.76576	*				
TCOMTECH	.74398	*				
TMSTEACH	.86098	*				
TOREACH	.66367	*				
TWEYNINE	.73795	*				
UNDANCE	.64723	*				
UNEMDEGR	.96374	*				
UNEMPTEC	.94783	*				
UNEMRAT	.69058	*				
UNENGLIS	.82550	*				
UPGRA	.75619	*				

Rotated Factor Matrix:

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
MOOPC	.84934	.18795	.04734	-.04936	.00747
POPFIV	.84934	.18795	.04734	-.04936	.00747
EML	.65127	.05395	.42281	-.00223	.02904
SHANPC	.45515	.32936	.34298	.07399	.11192
UNEMDEGR	.05487	.96999	.10235	.00320	.00410
UNEMPTEC	.12018	.93937	.06836	.03968	-.08797
OLUNEMP	.17299	.91236	.12026	.03104	-.07327
ALUNEMP	.05505	.90678	-.10815	.03415	.10516
MALUN	.19244	.57150	.37834	-.10145	.03161
MALEPERM	-.12476	-.13750	-.80883	.03352	-.01943

MSEASON	-.09843	.11495	.79732	-.06295	.09803
FSEASO	.16151	-.06410	.79049	.02026	-.02256
GRADCOM	-.46247	-.12377	-.55766	.34066	.25435
GRADSCE	-.26948	-.03223	-.47954	.37012	.23484
SCHOOLS	.11256	-.02095	-.07949	.77716	-.20957
CHAIRS	-.20343	.13894	-.09608	.75473	-.04088
STDTEA	.30008	.27145	.15589	.69332	.09881
UNENGLIS	-.49925	-.17835	.01614	.61981	-.11174
BLDGSTUD	-.43769	-.16906	-.22083	.58954	-.00374
UNDANCE	-.17025	-.03216	-.24707	.51523	.06029
FIFTYFIV	-.07964	.08013	.23031	.00574	.78196
SXSEVEN	-.05472	-.02377	.01234	.07514	-.74842
OVRFFTYX	.03183	-.13379	-.16125	.05702	.74456
GRADARTS	-.36014	.05440	.14466	.26061	-.48551
UNEMRAT	.19843	.00120	.06176	-.14767	-.10157
UPGRA	.33037	.02957	.19600	-.19180	-.05072
GHOUSE	-.49269	-.22019	-.33645	.07922	-.03278
EIGTEEN	.33271	-.24247	.15439	.03295	.12119
MFORIN	.13780	-.11572	.23199	-.05755	.13471
FFORINEM	.06896	-.13005	.00051	-.06608	-.21411
TCOMTECH	-.22580	.12064	-.30641	.23849	.08783
TAMPC	.18921	-.05807	-.02382	.16809	.07535
EMRATIO	.06201	-.08964	-.00262	.03232	-.28722
TARTSPER	.10806	.02470	-.12175	.30677	-.45916
RESTNTWE	.10214	-.14961	-.11646	-.00430	-.09772
TMSTEACH	-.54881	-.02749	.03606	-.04019	-.04522
RESTUWEN	.02779	.09408	-.11156	.00494	.17063
REACHPER	-.03666	-.19464	-.09426	.16811	.00275
SINPER	-.50538	.08935	.03402	-.00174	.08816
TOREACH	.29222	.11952	.02369	.00796	.05339
TWEYNINE	.23290	-.24722	-.22381	.29230	.02321
RESITEN	.09794	-.03654	-.22001	.01399	-.02532

	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
MOOPC	.19904	.09699	.06115	.05902	.17261
POPFIV	.19904	.09699	.06115	.05902	.17261
EML	.17020	-.03848	.37850	-.02239	.01294
SHANPC	.24100	.01687	.36608	-.11488	-.03714
UNEMDEGR	-.00044	-.06255	-.02969	.04378	-.03539
UNEMPTEC	.10835	-.05250	.12190	.03380	.01763
OLUNEMP	.06195	-.12245	-.09288	.07523	.10060
ALUNEMP	-.11747	.00212	-.08644	.08926	.08096
MALUN	.11517	.12221	.48652	-.14706	-.03831
MALEPERM	-.06431	-.24646	.05563	.12917	-.02782
MSEASON	-.01913	.11978	.10558	.08351	-.03203
FSEASO	.16809	-.18067	-.00150	.08421	.13508
GRADCOM	-.21706	.09332	.01055	-.15200	-.08514
GRADSCE	-.36694	-.23329	.02208	-.10667	-.02796
SCHOOLS	-.21316	-.31586	-.03123	.16229	.02912
CHAIRS	.00272	-.04759	-.05180	-.22566	-.28315
STDTEA	-.03868	.17318	.13833	-.11464	.04709
UNENGLIS	-.07992	-.04608	.10497	.31151	-.05392
BLDGSTUD	-.08563	.10337	.10991	.17395	.07578
UNDANCE	-.20037	.19335	-.07187	.01033	-.42979
FIFTYFIV	-.10259	-.02280	-.09698	.03096	.06182
SXSEVEN	.01649	-.15653	-.07845	-.16824	.18780
OVRFFTYX	-.02023	-.23078	-.06975	.04352	.17329

GRADARTS	.06477	.43328	-.01305	.01503	.22036
UNEMRAT	.75515	-.09301	.10684	.01160	.11009
UPGRA	.66997	-.05246	-.08004	.00069	.33210
GHOUSE	-.58465	.02333	-.17258	.06947	-.19393
EIGTEEN	.38332	.33165	-.19605	.36682	-.05007
MFORIN	-.26919	.74819	.01745	-.08326	.15459
FFORINEM	.01749	.73636	-.09798	.06961	-.09130
TCOMTECH	.11502	.46848	.25827	.17833	.22851
TAMPC	-.16684	.06775	.68300	.18281	.35221
EMRATIO	.36429	-.17070	.64625	-.03938	-.00397
TARTSPER	.02519	.17297	-.47408	.16229	.15683
RESTNTWE	-.01350	-.05161	-.00954	-.77343	-.01737
TMSTEACH	-.08393	.19010	.13516	-.66669	.08408
RESTUWEN	-.15635	.15567	.26650	.55807	.11920
REACHPER	-.21032	-.04908	-.05329	.03166	-.73314
SINPER	-.17225	-.11189	-.11870	-.05422	-.67673
TOREACH	.01222	.03317	.00804	.03933	.06394
TWEYNINE	-.08580	-.01937	-.10729	.20997	-.16373
RESITEN	-.16556	.10075	-.14156	.13380	-.01882

	Factor 11	Factor 12
MOOPC	.04590	-.02120
POPFIV	.04590	-.02120
EML	.23833	-.08318
SHANPC	-.08965	-.00156
UNEMDEGR	.00599	-.03676
UNEMPTEC	.07340	.02804
OLUNEMP	.08997	.06593
ALUNEMP	.13732	.01512
MALUN	-.17742	.14240
MALEPERM	-.03176	-.16613
MSEASON	.03673	-.02040
FSEASO	.08955	.01539
GRADCOM	-.01199	-.03002
GRADSCE	-.17272	-.17590
SCHOOLS	.07287	.00116
CHAIRS	-.09258	.17024
STDTEA	-.10674	.10396
UNENGLIS	-.06361	-.15492
BLDGSTUD	.01422	-.32245
UNDANCE	-.06300	-.12477
FIFTYFIV	.29574	.06189
SXSEVEN	.04056	.00408
OVRFFTYX	-.18408	.08826
GRADARTS	-.09049	.14702
UNEMRAT	.04821	.10184
UPGRA	.00841	.00256
GHOUSE	.04797	-.00045
EIGTEEN	.36069	-.20353
MFORIN	-.05896	-.07990
FFORINEM	.10221	-.01041
TCOMTECH	.00392	.07867
TAMPC	-.04280	.18457
EMRATIO	.15024	.14747

TARTSPER	-.09983	.34280
RESTNTWE	.09824	.05720
TMSTEACH	-.16326	.12008
RESTUWEN	-.03837	.51250
REACHPER	-.16898	.03262
SINPER	.02676	-.10720
TOREACH	.72530	.16586
TWEYNINE	-.62934	-.01756
RESITEN	-.21100	-.76575

The following variables will be eliminated at the next (Appendix 5) exploratory stage, considering the reasons given.

Variable Name	**Reason**
GRADCOM *(Commerce and other graduate teachers per 10,000 students)*	Not loaded to expected factor and High loadings on more than one factor
GRADSCE *(Science and Mathematics graduate teachers Per 10,000 students)*	Not loaded to expected factor and High loadings on more than one factor
GRADARTS *(Arts and Music graduate teachers per 10,000 students)*	Not loaded to expected factor
EIGTEEN *(Percentage of population 5 - 18 years of age)*	Low loadings
TCOMTECH *(Trained Commerce teachers per 10,000 Students)*	Not loaded to expected factor and Low loadings
TARTSPER *(Trained Arts Teachers per 10,000 students)*	Not loaded to expected factor
TMSTEACH *(Trained Mathematics and Science teachers per 10,000 students)*	Not loaded to expected factor
TWEYNINE *(Percentage of population 19 – 29 years of Age)*	Loadings do not make sense
REISTEN *(Average number of years of residency less than 10 years)*	Generates a factor with one variable

Appendix 5

EXPLORATORY FACTOR ANALYSIS STAGE 11

Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
		*				
ALUNEMP	.91020	*	1	7.45675	22.6	22.6
BLDGSTUD	.66531	*	2	3.88396	11.8	34.4
CHAIRS	.73878	*	3	2.64760	8.0	42.4
EML	.78784	*	4	2.47501	7.5	49.9
EMRATIO	.74433	*	5	2.22027	6.7	56.6
FFORINEM	.56116	*	6	1.77552	5.4	62.0
FIFTYFIV	.81711	*	7	1.59717	4.8	66.8
FSEASO	.72634	*	8	1.29957	3.9	70.8
GHOUSE	.78098	*	9	1.10284	3.3	74.1
MALEPERM	.75783	*	10	1.04046	3.2	77.3
MALUN	.86472	*				
MFORIN	.81115	*				
MOOPC	.92212	*				
MSEASON	.77782	*				
OLUNEMP	.94107	*				
OVRFFTYX	.79900	*				
POPFIV	.92212	*				
REACHPER	.67402	*				
RESTNTWE	.66542	*				
RESTUWEN	.70142	*				
SCHOOLS	.77651	*				
SHANPC	.67399	*				
SINPER	.77992	*				
STDTEA	.80520	*				
SXSEVEN	.68190	*				
TAMPC	.78763	*				
TOREACH	.83521	*				
UNDANCE	.63349	*				
UNEMDEGR	.96495	*				
UNEMPTEC	.94426	*				
UNEMRAT	.51494	*				
UNENGLIS	.79987	*				
UPGRA	.73259	*				

F A C T O R   A N A L Y S I S

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6
UNEMDEGR	0.96313	0.09520	-0.00203	0.09752	-0.00818	0.03987
UNEMPTEC	0.93417	0.16068	0.01586	0.07729	-0.08935	-0.04129
OLUNEMP	0.92587	0.16175	0.02470	0.10775	-0.052223	-0.15350
ALUNEMP	0.91265	0.16667	0.04224	-0.10139	0.11107	-0.05209
MALUN	0.56191	0.25950	-0.14505	0.35228	0.04067	0.04816
MOOPC	0.16254	0.92725	-0.08623	-0.00802	0.00242	-0.12714
POPFIVE	0.06254	0.92335	-0.06210	-0.01803	0.00255	-0.12584
EML	0.04473	0.68668	-0.04474	0.40348	0.01890	-0.02013
GHOUSE	-0.17337	-0.65088	0.16730	-0.39171	0.06338	0.28904
SHANPC	0.21715	0.53420	-0.00006	0.33997	0.12731	0.05782
UPGRA	-0.01201	0.49954	-0.25724	0.28192	-0.21132	-0.49666
SCHOOLS	0.00615	0.01382	0.82196	-0.11351	-0.06990	0.01279
CHAIRS	0.16460	-0.17100	0.72694	-0.06688	-0.02676	0.24959
STDTEA	0.28737	0.34185	0.68440	0.11924	0.13348	-0.00419

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6
UNENGLISH	-0.21175	-0.36273	0.64354	0.07124	-0.11897	0.17363
BLDGSTUD	-0.19751	-0.33889	0.62755	-0.13848	-0.08773	0.02142
MSEASON	0.09449	-0.00959	-0.10118	0.84273	0.04041	0.08599
FSEASO	-0.05403	-0.19116	0.00426	0.78023	0.00496	-0.21564
MALEPERM	-0.15806	-0.15563	0.06932	-0.77237	-0.01637	0.07407
SXSEVEN	0.01359	-0.34340	0.15119	-0.45048	-0.01339	-0.29514
FIFTYFIV	0.10618	-0.17363	-0.03481	0.25471	0.81074	-0.08535
RESTUWEN	0.09260	0.09132	0.02203	-0.15752	0.58394	-0.00507
SINPER	0.09404	-0.52570	-0.03745	0.10905	0.02878	0.64552
UNDANCE	-0.04743	-0.19235	0.49132	-0.16643	0.00595	0.52537
REACHPER	-0.19311	-0.06575	0.14432	-0.07828	-0.03299	0.51028
UNEMRAT	-0.02051	0.32922	-0.23195	0.13924	-0.19937	-0.35472
TAMPC	-0.04341	0.10140	0.20977	-0.08245	0.19317	-0.24077
EMRATIO	-0.39096	0.11019	-0.02127	0.06310	-0.30518	-0.10040
MFORIN	-0.10766	0.06121	-0.04347	0.19216	0.11768	-0.51065
FFORINEM	-0.10612	0.01148	-0.09510	0.00105	-0.25902	0.00343
RESTNTWE	-0.12037	0.04622	-0.03490	-0.14462	-0.03257	0.04689
OVRFFTYX	-0.14399	0.06017	0.03673	-0.17878	0.21063	-0.11170
TOREACH	0.13837	0.20248	-0.05252	0.08952	0.04055	-0.06926

VARIABLE	FACTOR 7	FACTOR 8	FACTOR 9	FACTOR 10
UNEMDEGR	-0.06880	-0.07834	0.07025	-0.03740
UNEMPTEC	0.13694	-0.09352	0.44156	0.02036
OLUNEMP	-0.03686	-0.10976	0.02823	0.07034
ALUNEMP	-0.10171	0.01691	0.08633	0.17832
MALUN	0.50875	0.12922	-0.02506	-0.23740
MOOPC	-0.00617	0.07592	0.05193	0.06114
POPFIVE	-0.00412	0.07555	0.05666	0.02555
EML	0.31805	-0.01400	-0.04462	0.21324
GHOUSE	-0.18658	0.14001	0.06099	-0.01271
SHANPC	0.38663	-0.02420	-0.11715	-0.03753
UPGRA	-0.07806	-0.19265	0.01452	0.05036
SCHOOLS	0.04661	-0.20324	0.01825	0.19729
CHAIRS	0.06193	-0.08181	-0.20879	-0.17998
STDTEA	0.15569	0.22815	-0.07354	-0.15363
UNENGLISH	-0.01888	-0.15571	0.35179	-0.10765
BLDGSTUD	-0.08050	0.04648	0.27691	-0.07052
MSEASON	0.00934	0.11654	0.15914	0.01873
FSEASO	0.03584	-0.14906	-0.02456	0.08646
MALEPERM	-0.00814	-0.29760	0.11319	0.00762

SXSEVEN	-0.00890	-0.13981	-0.21648	-0.05641
FIFTYFIV	-0.02106	0.02928	-0.02314	0.20780
RESTUWEN	0.23706	0.12917	0.44219	0.04021
SINPER	-0.15348	-0.16152	-0.07405	-0.09180
UNDANCE	-0.12608	0.12055	0.00174	0.13658
REACHPER	-0.09634	-0.07149	-0.02822	-0.07527
UNEMRAT	0.25233	-0.31236	-0.07798	-0.00656
TAMPC	0.72555	0.18719	0.22908	0.11102
EMRATIO	0.54237	-0.50101	-0.02989	0.02020
MFORIN	-0.02014	0.85959	0.03514	-0.00087
FFORINEM	-0.01225	0.68640	0.04436	0.01896
RESTNTWE	0.04689	0.02047	-0.78377	0.03691
OVRFFTYX	-0.11170	-0.22184	0.55564	-0.21732
TOREACH	-0.06926	0.01621	-0.01455	0.86873

The following variables will be eliminated at the final factor analysis, for the reasons given below.

Variable	Reason
Percentage of population between 6 - 17 years (SXSEVEN)	High loadings on more than one factor
Percentage of Tamils (TAMPC)	Did not load to expected factor
Female/male employment ratio (female employment / male employment) (EMRATIO)	High loadings on more than one factor
Percentage of population over 56 years (OVRFFTYX)	Loading does not make sense
Percentage of shanties reached (REACHPER)	Did not load to expected factor
Average number of years of residency 10 - 20 years (RESTNTWE)	High loadings on more than one factors
Percentage of population 18 - 55 years (FIFTYFIVE)	Loading does not make sense
Percentage of shanties to be reached (TOREACH)	Generates a single factor
Average number of years more than 20 years (RESTUWEN)	Loading does not make sense

Appendix 6

Variables Used in the Final Factor Analysis

I. *Percentage of Sinhalese (SINPER)*

II. *Percentage of Ceylon Moors (MOOPC)*

III. *Percentage of population under five years of age (POPFIV)*

IV. *Percentage of male employment of the total work force (MALEPERM)*

V. *Percentage of female seasonal employment of the total work force (FSEASO)*

VI. *Percentage of male seasonal employment of the total work force (MSEASON)*

VII. *Female male unemployment ratio (female unemployment / male unemployment) (UNEMRAT)*

VIII. *Percentage of male unemployment of the total work force (MALUN)*

IX. *Percentage of female foreign employment of the total work force (FFORINEM)*

X. *Percentage of male foreign employment (MFORIN)*

XI. *Percentage of employment in low income categories (carpenters, laborers, trade persons, and other minor I Income groups) (EML)*

XII. *Number of unemployed people with a degree Qualification per 10,000 unemployed people (male and female) (UNEMDEGR)*

XIII. *Number of unemployed people with Advanced Level Qualification (grade 13) per 1,000 unemployed people (male and female) (ALUNEMP)*

XIV. *Number of unemployed people with Ordinary level (grade 11) Qualification per 1,000 unemployed people (male and female)(OLUNEMP)*

XV. *Number of unemployed people with technical education per 1,000 unemployed people (male and female) (UNEMPTEC)*

XVI. *Number of schools per 10,000 students (SCHOOLS)*

XVII. *Number of students per teacher (STDTEA)*

XVIII. *Average school building square feet per student (BLDGSTUD)*

XIX. *Number of student chairs per 10,000 students (CHAIRS)*

XX. *Untrained Dance, Music, other, and Sports teachers per 10,000 students (UNDANCE)*

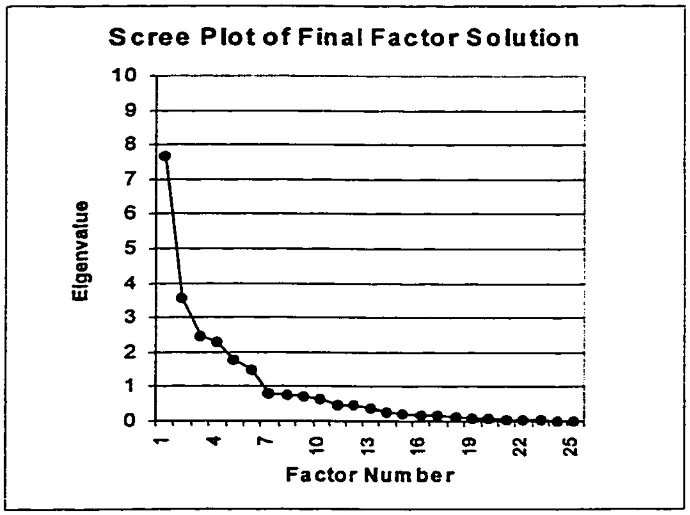
XXI. *Percentage of good housing units (GHOUSE))*

XXII. *Percentage of renovated buildings (UPGRA)*

XXIII. *Percentage of slums and shanties (SHANPC)*

XXIV. *Untrained English teachers per 10,000 students (UNENGLIS)*

Final Factor Analysis



Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
		*				
ALUNEMP	.87712	*	1	7.04326	29.3	29.3
BLDGSTUD	.68668	*	2	3.79722	15.8	45.2
CHAIRS	.67010	*	3	2.31089	9.6	54.8
EML	.73510	*	4	2.14655	8.9	63.7
FFORINEM	.67114	*	5	1.67727	7.0	70.7
FSEASO	.69399	*	6	1.05445	4.4	75.1
GHOUSE	.78759	*				
MALEPERM	.74971	*				
MFORIN	.82443	*				
MOOPC	.84751	*				
MSEASON	.77739	*				
OLUNEMP	.93724	*				
POPFIV	.84751	*				
SCHOOLS	.71362	*				
SHANPC	.75023	*				
SINPER	.75661	*				
STDTEA	.74591	*				
UNDANCE	.66892	*				
UNEMDEGR	.95734	*				
UNEMPTEC	.93271	*				
UNEMRAT	.65215	*				
UPGRA	.76960	*				
MALUN	.68724	*				
UNENGLIS	.76978	*				

Rotated Factor Matrix:

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6
UNEMDEGR	0.96139	0.11852				
OLUNEMP	0.93507	0.14111	0.14717			-0.10705
ALUNEMP	0.93155					
UNEMPTEC	0.93056	0.20739				-0.10544

SHANPC	0.20384	0.78405			0.22121	-0.10951
EML		0.71278	0.14405		0.12506	
POPFIVE	0.18096	0.67612	0.25043		-0.10367	0.17123
GHOUSE	-0.15512	-0.57557	-0.24350	0.17842	-0.23007	0.13700
MALUN	0.20534	0.56888		-0.13131	0.21495	
SINPER		-0.17313	-0.81868			-0.22213
UPGRA		0.12043	0.78765	-0.24984	0.24181	-0.10770
UNEMRAT		0.14411	0.59919	-0.20400	0.13545	-0.25397
MOOPC	0.19096	0.20271	0.57043		-0.10123	
SCHOOLS				0.79844	-0.15240	-0.21749
UNENGLIS	-0.15631	-0.14066	-0.14756	0.72551	0.13085	-0.10701
CHAIRS	0.12211		-0.23188	0.72136		-0.13220
BLDGSTUD	-0.15081	-0.20117		0.69309		
STDTEA	0.25956	0.15758		0.64532		0.20634
UNDANCE		-0.11674	-0.21793	0.52425	-0.17341	0.11734
MSEAN	0.11011				0.86124	0.13542
MALEPERM	-0.14601	-0.19812			-0.77371	-0.26870
FSEASO		0.25219	0.21308		0.73638	-0.18486
MFORIN					0.22132	0.87571
FFORINEM	-0.11638					0.69872