韓國地形空間情報學會論文集 第1卷 第2號 1993年 12月 pp. 223-235.

시카고 부분지역의 사회경제적 특성에 대한 지형공간정보체계의 이용

Socio-economic Features of One Slice of Chicagoland Using A Geo-Spatial Information System

오 종 우*
OH Jong-Woo

要 旨

시카고 광역도시지역 중에서 사회경제적 특징이 모식적으로 포함되는 지역을 선정하여, 어떠한 영향으로 그 특성들이 형성분포되는가를 파악하기 위하여 본 연구가 시작되었다. 연구대상지역에 대한 지형공간정보체계(GSIS)의 접근으로서 14개의 변수와 44개의 관측값을 선정하여 3종류의 통계적 방법으로, 지역성에 대한 사회경제적 양상을 비교하였으며, 이를 3가지의 도식화 방법으로 GSIS에 의한 결과를 도출하였다. 군집분석 I (Grouping analysis)에서의 결과를 볼 때 위성도시에 해당되는 ELMHRST, ELMWOOD PARK, MELROSE PARK, NORTH LAKE에서 연수입치가 높게 나타나서 상기 분석치를 인정할 수 있었다. 인자분석(Factor analysis)에서는 사회-민족적, 경제적, 남미유입인(HISPANICS), 흑인, 인생기대치, 차량수 등이 주요 관점이었는데, 사회민족적 인자가 가장 중요한 상관관계 계수를 나타내고 있어 대도시의 지역특성에 의한 결과로 작용된 것을 알 수 있다. 지공간 입체도화방식에 의한 사회경제적 특색들이 3가지 도식으로 표현되었다. 그 결과에 의하면 경제적 수준이 높은 지역으로서 서부의 DuPage County의 위성도시들과 시카고 동부의 Lake Michigan 인접지역이 가장 높게 나타나 있어, 지역분화에 따른 사회경제적 표현이 상이하게 나타났다. 따라서 흑인과 남미유입인의 거주지역과 백인의 거주지역은 사회경제적 기준에 의해서 구분되고 있는 것으로 나타났다. 그 예로서 상한연수입이 $30,311인데 비하여 하한 연수입이 $4,364에 불과하여 큰 격차를 나타낸다.

ABSTRACT

This study associates with socio-economic status in a slice section area of Chicago metropolitan to get spatial patterns of urban windows. GSIS(Geo-spatial Information System) has been monitored with several statistic methods, and geo-spatial map presentations. From the grouping analysis, the result displays that most suburban town have high income values, such as Elmhurst, Melrose Park, North Lake(Income ranges between $25,000 ~ 30,000: 1980 Sensus data). The factors produced form both anlayses of SAS and BMDP are socio-ethnic, economic, hispanic, black, life expectancy, and multiple car ownership. In the study area the socio-ethnic factor is striking, and is composed of nine out of the fourteen varialbles. Geo-spatial 3-D mapping represents a socio-economic configuration of the study area. The high income value areas are Elmhurst and North Lake, and a spot between Belmont Ave. and the Lake Shore. Economic configuration is a vital importance of socio-economic activities in the urban areas. In the study area a minimum average income level is about $4,364 and maximun is $30,311.

* 경희대학교 강사

I. INTRODUCTION

This paper deals with socio-economic activities of one slice area(a window area) of Chicagoland (Fig. 1). The study area is one of the complex urban slices in the Chicago metropolitan area, because many different ethenic groups inhabits, such as Hispanics, Blacks, and Whites. In Chicago metropolitan area, different ethenics have different social, cultural, and economic characteristics. The result may exhibit these differences from the Lake site to suburban areas. The result will be a part of a socio-economic model of inhabitants in Chicagoland.

The window area provides a few different interesting aspects: comparison between areal characteristics, such as income values, vehicle ownerships, and blacks socio-economic status. These research points were selected because two different Counties were involved, at least three ethic groups(white, blacks, hispanics) were distributed, and the most typical socio-economic mixture of the metropolitan spatial examples in United States.

The study area in the Chicago metropolitan area is sixteen miles long from Lake Michigan in the east to York Road on the west border. The study area is a strip of the Chicago urban landscape. It is two miles wide from North Avenue on the south boundary and Belmont Avenue on the north boundary. The east DuPage County. The study area is one of the complex urban slices of the Chicago metropolitan area. The study area consists of Cook County and DuPage County, two highways(Interstate 294 and 94/90), many rails, and two rivers(the Des Plaines River, the Chicago River)(Fig.1). There is no data near the river strips since there is no habitation.

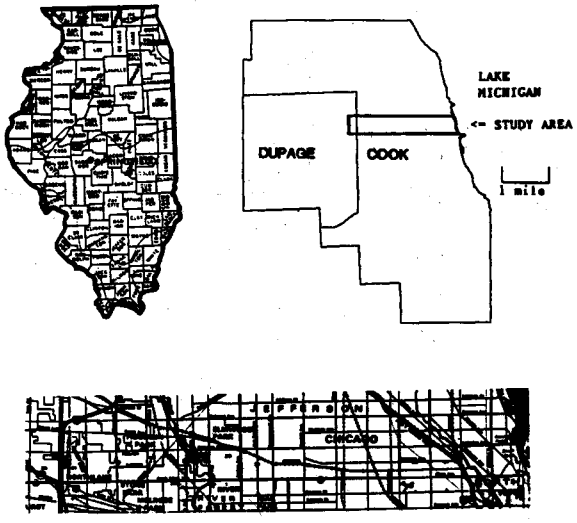

Figure 1. Study area (Left : Illinois map, Right : County boundary map, Lower : road map)

2. METHODS

Using GSIS(Geo-Spatial Information System) to approach socio-economic activities in urban areas can be one of the useful methods to produce systematic quantitative results[1]. In this study a small selected slice area(windows) (Fig.1) was chosen in order to generate a database (1980's Census data have been monitored, and for urban studies using GIS packages The University of Illinois main computer system was used).

The data was collected in arbitrarily in quarter mile sections for a total of 129 sections. The distribution of the townships will be Lakeview, Jefferson, Leyden, and Addison.[2] The study area from west to east near the shore of Lake Michigan has a dense population and is increasing prominently.

Grouping analysis, cluster analysis, and factor analysis were tested for general to specific

socio-economic characteristics, and provide limitations and further suggestions for GSIS approaches. Finally, map presentations may demonstrate incom values using trend surface map(TSM) and Symvu surface map(SSM), and vehicle ownership using symbolic plan map(SMP). Geospatial distribution of urban characteristic using GSIS technics provides perfect function of this kinds of complex urban systems.

3. RESULTS

3.1 GROUPING ANALYSIS

The spatial grouping analysis provides a geographic distribution in a group which combines observations and variables. The step graph of grouping is produces by the distance sqaured matrix between each observation. The gouping tree shows combined groupings individually.

The study area is a combination of western Cook County and eastgern DuPage County. These two areas are part of the western suburbs of Chicago(Fig.1).

In order to get the grouping trees, all 14 variables and 44 observations have been used. The variables are house holds, median house size, median family income, total population, ages less than six, median age, blacks, hispanics, number of college graduates, no vehicles, one vehicles, and two vehicles per house. The second grouping tree represents the education status of hispanics with income values. The three selected variables are median family, population, and number of college graduations.

3.1.1 THE GROUPING TREE; ALL VARIBLES

The early groupings(Fig.2) show seven initial sectors; the second group has five early groups(sub A group has 1,2, and 3, sub B group has 4 and 5). The third one has two early groups. The sub-group 4 and 7 indicate high positive correlation wiht a majority of the early groupings. The first two isolates are lacated in the Elmhurst area, and the other three isolates are located in between Northlake and Melrose Park. They exhibiyt high income values, because they are located in suburb areas near to west of Chicago.

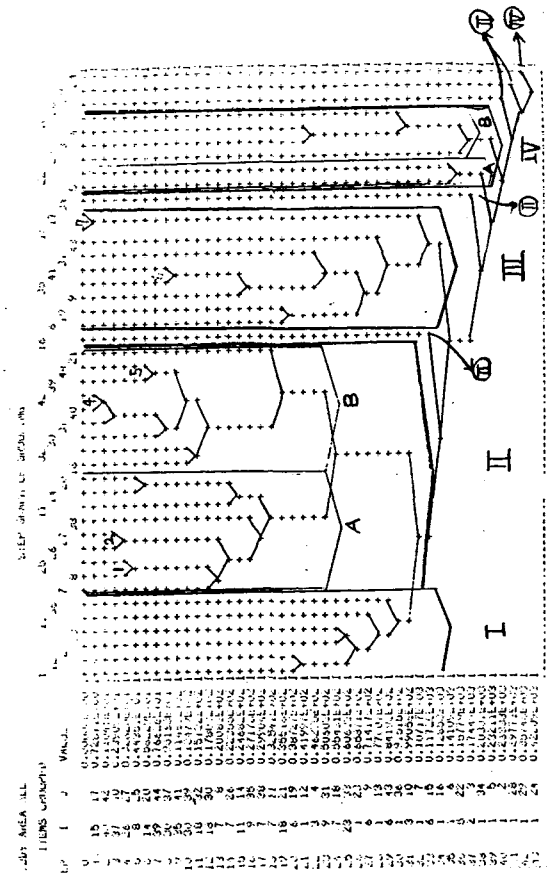

Figure 2. Grouping tree using all variables : numbers indicate ordering number and groups of close correlated observations.

3.1.2 SELECTED HISPANICS GROUPING TREE: THREE VARIABLES

Hispanes are one of the largest ethnic groups in Chicagoland.

There is a significant positive correlation between income and college graduations. The selected grouping analysis may be used for determining the differences of correlation between hispanics' education status and income(Fig.3).

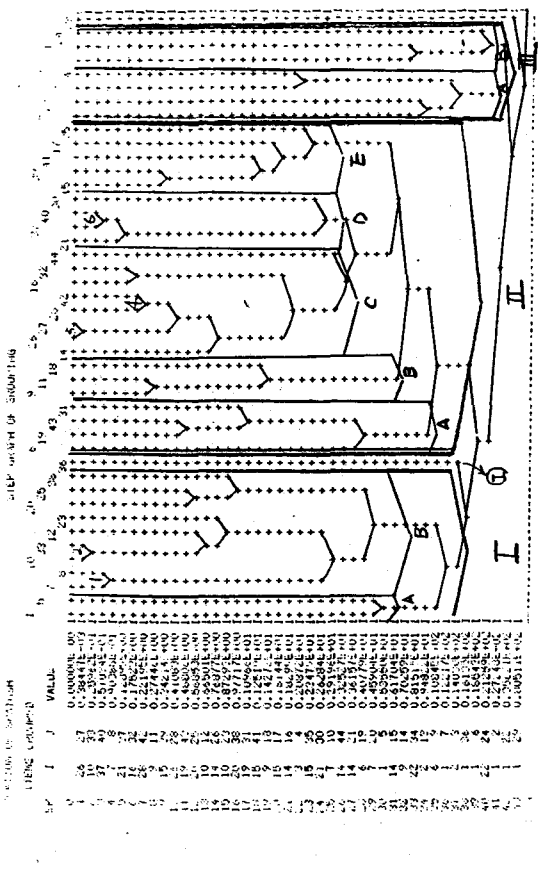

Figure 3. Grouping tree using 3 variables for Hispanics : numbers are orderings, alphabets are subgroups, roman numerics are major groups.

The groupng treeis divided into three groups with sub-groups and two isolated observations. The first and third groupings each have two sub-groups. The second grouping consists of five sub-groups. Isolates show between the first and second, and the end of the third grouping. These two isolates both belong to the first grouping according to the test of the distance sqared matrix. The isolate between the first and second grou0ping is located in Franklin Park, and the other one is in Melrose park. Obviously, the third grouping looks clearly devided into each sub-group, wheras the second grouping looks extremely complex. Even though the second grouping has complicated groups, there are isolates, and also the final tree matches almost perfectly.

In the second grouping, the early groupings and distribution show more clusting than the other two groups. The income values in the Ellmhurst area are comparatively high, however, the highest income area in the study areas is between Melrose Park and Elmwood Park(Income ranges between $25,000 to 30,000 annually).

Even though the variables income and college graduation rate are strongly correlated, in this study the distribution of the college attendance and income values will not match. There must be either no correlation between the income value and the college attendance rate or a bad choice of data, or neither.

3.2 CLUSTER ANALYSIS

The single method has been used for the cluster analysis. The sub single groupings have been drawn by hand useing data provided(Fig.4). The five groupings and one isolate have been produced from the 14 variables applied for the cluster analysis. According to ther cluster trees, the fourth, fifth, and sixth groupings are perfectly matched but the first, second, and third groupings are not

correlately matched.

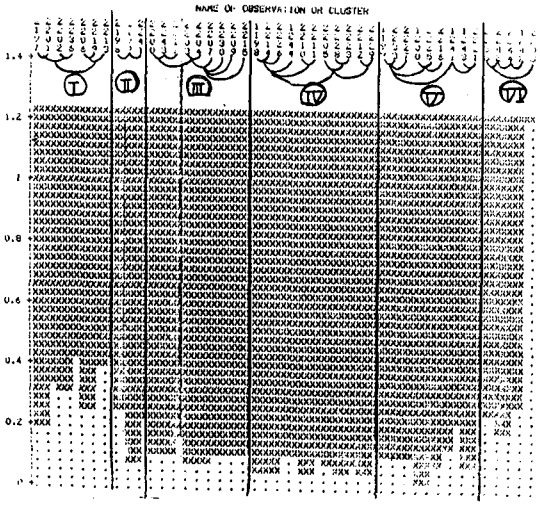

Figure 4. Clustering tree using SAS displays 6 groups.

A comparison result between the cluster analysis(Fig.4) and the grouping analysis(Fig2,3) demonstrates some similarities even though their shapes differ from each other. There must be a special correlation because both trees are used by all 14 variables.

Although there are several deep valleys in each unit, they are finally joined to each other. Therefore the cluster analysis has to be verified in a step by step approach, because the random groupings yield huge ranges of the standard deviations.

Cluster analysis exhibits a simiar result of above. Suburban town have high income values, such as Elmhurst, Elmwood park, Melrose park, and North Lake.

3.3 FACTOR ANALYSIS

Factor analysis is a method of spatial analysis. "Factor analysis assumes that ther varialbles are linear combinations of some underlying(hypothetical of unobservable) factors[3]. As Johnston[4](1980) mentioned that the purpose of factor analysis is th identify groups of variables with shared common variance, factor analysis will explain the correlation between variables and observations in a plot, data matrix, and statistical data. One of the clearest examples of quantitative analysis using factor analysis by Bell(1955) on hypothesis-testing[5].

There are two factor analysis methods that have been used, SAS factor analysis and BMDP analysis.

3.3.1 DATA

The data ofr the study area was chosen from 129 observations and fourteen variables(Table 1). There are many missing values because of the paths of the highways, rivers, and rails. Fourteen out of 129 observations have zero values. These useless values have been deleted in order to get a more accurate output. The BMDP program produced six plots of unrotated and rotated factor loadings and three plots of factor scores. Furthermore, the SAS factor program produced 36 plots which cover almost all of the plottings of the factors.

Table 1. A list of the fourteen variable

No.	Variable Names	V.
X 1	Total Number of Households	(HHLD)
X 2	Median Household of Size	(MHSZ)
X 3	Median Number of Workers	(MWK)
X 4	Median Family Income	(MFI)
X 5	Total Number of Persons	(TPOP)
X 6	Total Number of Persons Less than Six Years Old Median Age	(LSIX)
X 7	Median Age	(MAGE)
X 8	Total Number of Black	(BL)
X 9	Total Number of Hispanic	(SPAN)
X10	Total Number of College Students	(COL)
X11	Total Number of Households with Zero Vehicle	(VO)
X12	Total Number of Households with one Vehicle	(V1)
X13	Total Number of Households with two Vehicle	(V2)
X14	Total Number of Households with three Vehicle	(V3)

3.3.2. ANALYSIS AND RESULT

The data will be anayzed with two different factor computer programs. One program is produced by SAS and the other is by BMDP.

The SAS factor analysis consists of eigenvalues for ther correlation matrix, factor pattern, orthogonal transformation matrix, rotated factor pattern, and pots. The BMDP factor analysis is composed of statistics for each variable, correlation matrix, squared multiple correlations, communallities, unrotated and rotated factor loadings, estimated factor scores, distances(Chi-square), and plots.

The SAS factor analysis produced a factor loading matrix in which all fourteen varialbes combined with four out of fourteen factors(Table 2). The importance of the factor groups is the eigen values which is the total of the squared loadings, indicates the total variance explained by the factor.

Table 2. Factor loading SAS

	UNROTATED				ROTATED			
No.	FAC1	FAC2	FAC3	FAC4	FAC1	FAC2	FAC3	FAC4
X 1	0.947	-	-	-	0.951	-	-	-
X 2	-	-	0.605	-	-	-	0.401	-
X 3	-	-	0.748	-	-	-	0.933	-
X 4	-	0.565	-	-	-	-	0.847	-
X 5	0.962	-	-	-	0.790	-	-	-
X 6	0.643	-	-	-	-	0.837	-	-
X 7	-	0.677	-	-	-	-	-	0.933
X 8	0.659	-	-	-	0.474	-	-	-
X 9	0.568	-	-	-	-	0.869	-	-
X10	0.808	-	-	-	0.857	-	-	-
X11	0.903	-	-	-	0.954	-	-	-
X12	0.941	-	-	-	0.916	-	-	-
X13	0.607	-	-	-	0.473	-	-	-
X14	-	-	-	0.535	-	-	0.123	-
EIGEN	6.442	2.385	1.888	1.142	4.779	2.090	1.963	1.126

The highest value of the unrotated factor loading SAS which belongs th the factor 1 is the total population varialble. The rotated factor loading's highest value, however, is different from the unrotated one. Also, the lowest value of the unratated and rotated factor loading is not matched(Table 2). When the factor loading matrix is rotated, the different factors are produced. The eigen values are all above 1.0.

There are some different varialbes shown. The number of the factor 1 in the unrotated factor loading shows nine varialbles. However, the number of the factor loading in the ratated factor consists of seven varialbes. Consequently, the ranking of factor 3 in the rotated factors is comparatively higher(Table 3).

Table 3. Factor loading SAS

UNROTATED				ROTATED			
FAC1	FAC2	FAC3	FAC4	FAC1	FAC2	FAC3	FAC4
HHLD	MFI	MHSZ	V3	HHLD	LSIX	MHSZ	MAGE
TPOP	MAGE	MWK	-	TPOP	SPAN	MWK	-
LSIX	-	-	-	BL	-	MFI	-
BL	-	-	-	COL	-	V3	-
SPAN	-	-	-	VO	-	-	-
COL	-	-	-	V1	-	-	-
VO	-	-	-	V2	-	-	-
V1	-	-	-	-	-	-	-
V2	-	-	-	-	-	-	-

Table 4. Factor loading BMDP

	UNROTATED				ROTATED			
No.	FAC1	FAC2	FAC3	FAC4	FAC1	FAC2	FAC3	FAC4
X 1	0.945	-	-	-	0.953	-	-	-
X 2	-	-	0.733	-	-	0.607	-	-
X 3	-	-	0.705	-	-	-	0.967	-
X 4	-	0.668	-	-	-	-	0.870	-
X 5	0.963	-	-	-	0.779	-	-	-
X 6	0.680	-	-	-	-	0.793	-	-
X 7	-	0.644	-	-	-	-	-	0.358
X 8	0.669	-	-	-	0.750	-	-	-
X 9	0.580	-	-	-	-	0.845	-	-
X10	0.807	-	-	-	0.954	-	-	-
X11	0.903	-	-	-	0.915	-	-	-
X12	0.939	-	-	-	0.946	-	-	-
X13	0.609	-	-	-	-	-	-	-
X14	-	-	-	0.509	-	-	-	0.692
EIGEN	6.391	2.441	1.933	1.053	4.779	2.506	2.205	1.335

The BMDP factor analysis is also composed of a factor loading matrix in which all fourteen variables are combined into four factors. The eigenvalues are all above 1.0. When the factor loading matrix is rotated, the different factors are produced(Table 4). The variables within each factor should be interrelated and produce the different spatial pattern on the study area.

The number of variables in factor 1 in the unrotated factor loading consists of nine variables wheras the number in factor 1 in the rotated factor loading is composed of only six variables. the varables of other three factors are more evenly distributed(Table 5).

Table 5. Factor loading BMDP

UNROTATED				ROTATED			
FAC1	FAC2	FAC3	FAC4	FAC1	FAC2	FAC3	FAC4
HHLD	MFI	MHSZ	V3	HHLD	MHSZ	MWK	MAGE
TPOP	MAGE	MWK	-	TPOP	LSIX	MFI	V2
LSIX	-	-	-	BL	SP	-	V3
BL	-	-	-	COL	-	-	-
SPAN	-	-	-	VO	-	-	-
COL	-	-	-	V1	-	-	-
VO	-	-	-	-	-	-	-
V1	-	-	-	-	-	-	-
V2	-	-	-	-	-	-	-

In summary, factor analysis has been used with two programs; SAS and BMDP in order to analyze a spatial distribution from fourteen variables. The variables of the highest value and lowest value for the both unrotated SAS and BMDP factor loadings are the same. However, the variables of the highest value and lowest value for the both rotated SAS and BMDP factor loadings are not the same. On the other hand, the number of group factors and the name of the variables for the both SAS unrotated loading of SAS and BMDP are exactly the same. The number of group factors and the name of the variables for neither rotated loading of SAS and BMDP are more evenly shared than the unrotated factor loadings. Also, the factors for the BMDP variables are more supplementary to act than that of the SAS factors.

3.4 MAP PRESENTATIONS

Statistical data can explain various ways of expresssions.

One of the most effective explanations of data is graphics. This graphic skill can provide many different types of mapping as if the graphic displays were the actual formations. 3-D angles from the statistical data give us easy way to understand the data. Specific geographic areas also can be described using SYMAP and SYMVU mapping procedures. These two analyses consist of income density in a selected Chicago area (Lakeshore & North Ave., Lakeshore & Belmont Ave., North Ave. & York Rd. and Belmont Ave. & York Rd.).

3.4.1 TREND SURFACE MAPS(TSM)

Trend surface maps describe different levels of income status in different shades. TSM was used by geographers from the mid 1960's(Tobler, 1964; Chorley and Haggett, 1965) for atmospheric problems and landform surface processes. Three types of trend surface maps are provided. The first and second order of trend surface maps seems similar. Thus the first, third and sixth degree orders have been chosen in order to get more of a dramatic change in intervals.

Fig. 5 is a first order trend surface of income values. The coefficient of determination on R squared value for this first order trend surfaces shows approximately 35% of variation explained (Table 6). Obviously, the sest suburban area has high income values, but east Chicago has lower

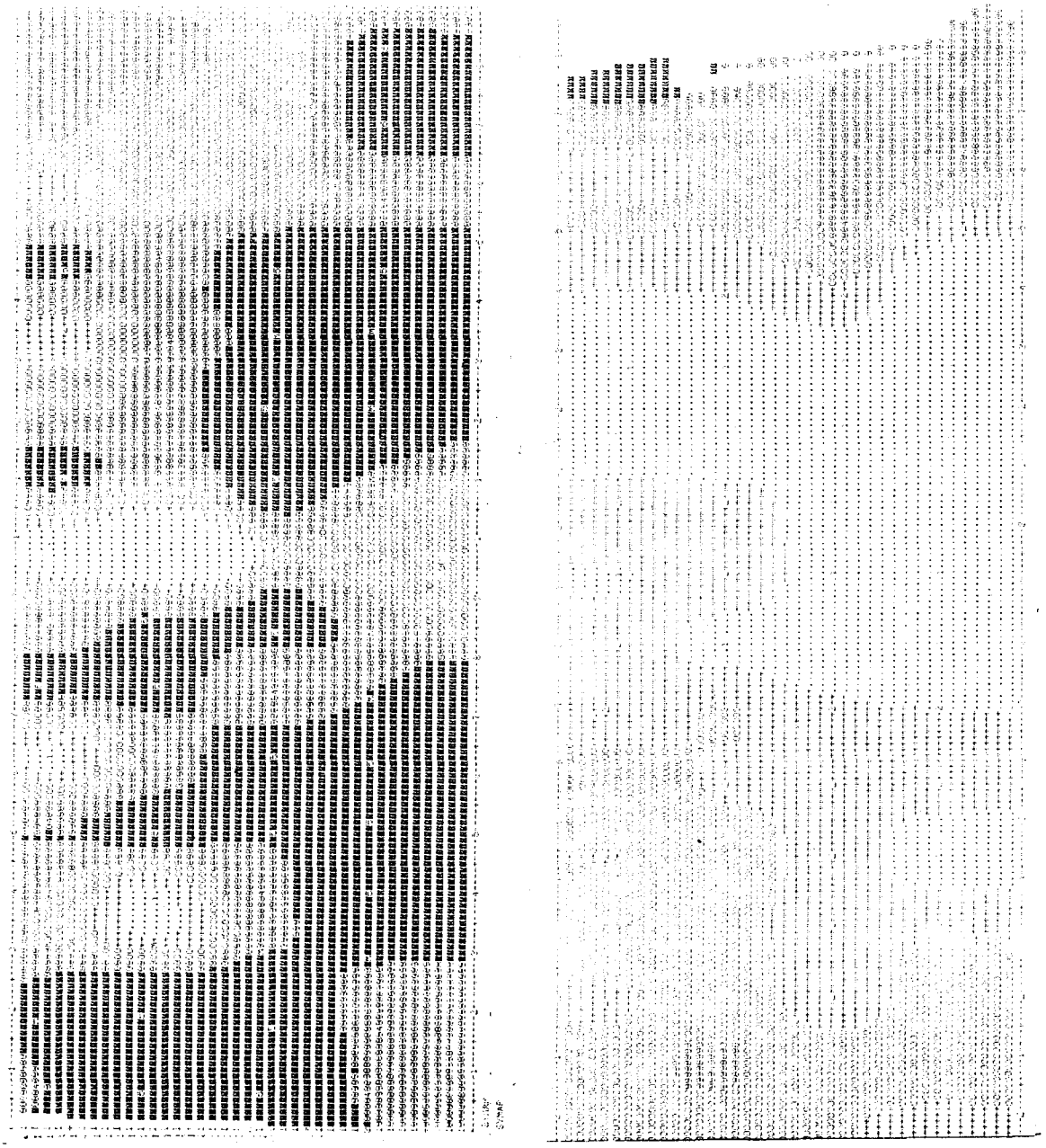

Figure 5. Trend surface map : first order for income values.

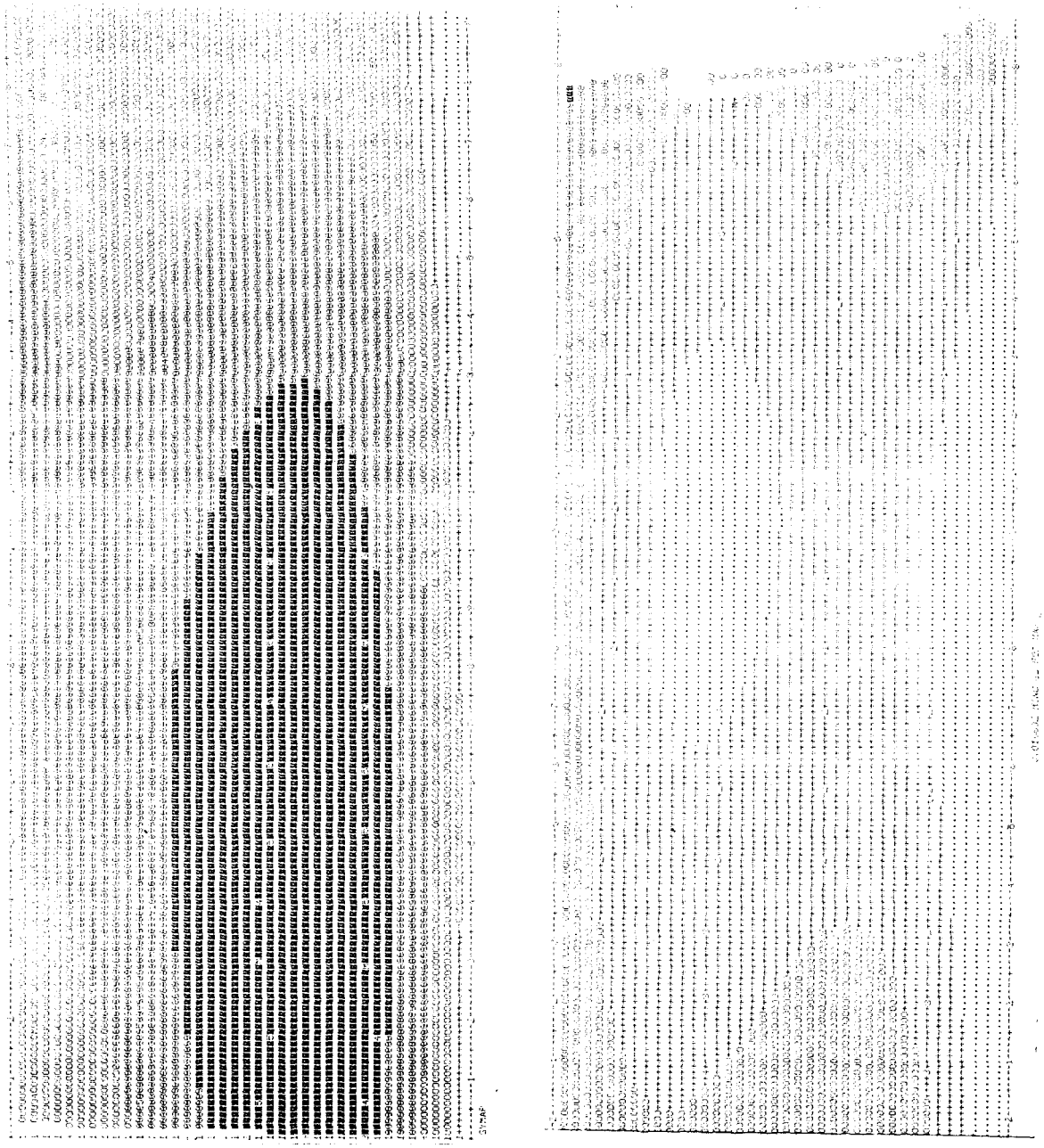

Figure 6. Trend surface map : third order for income values.

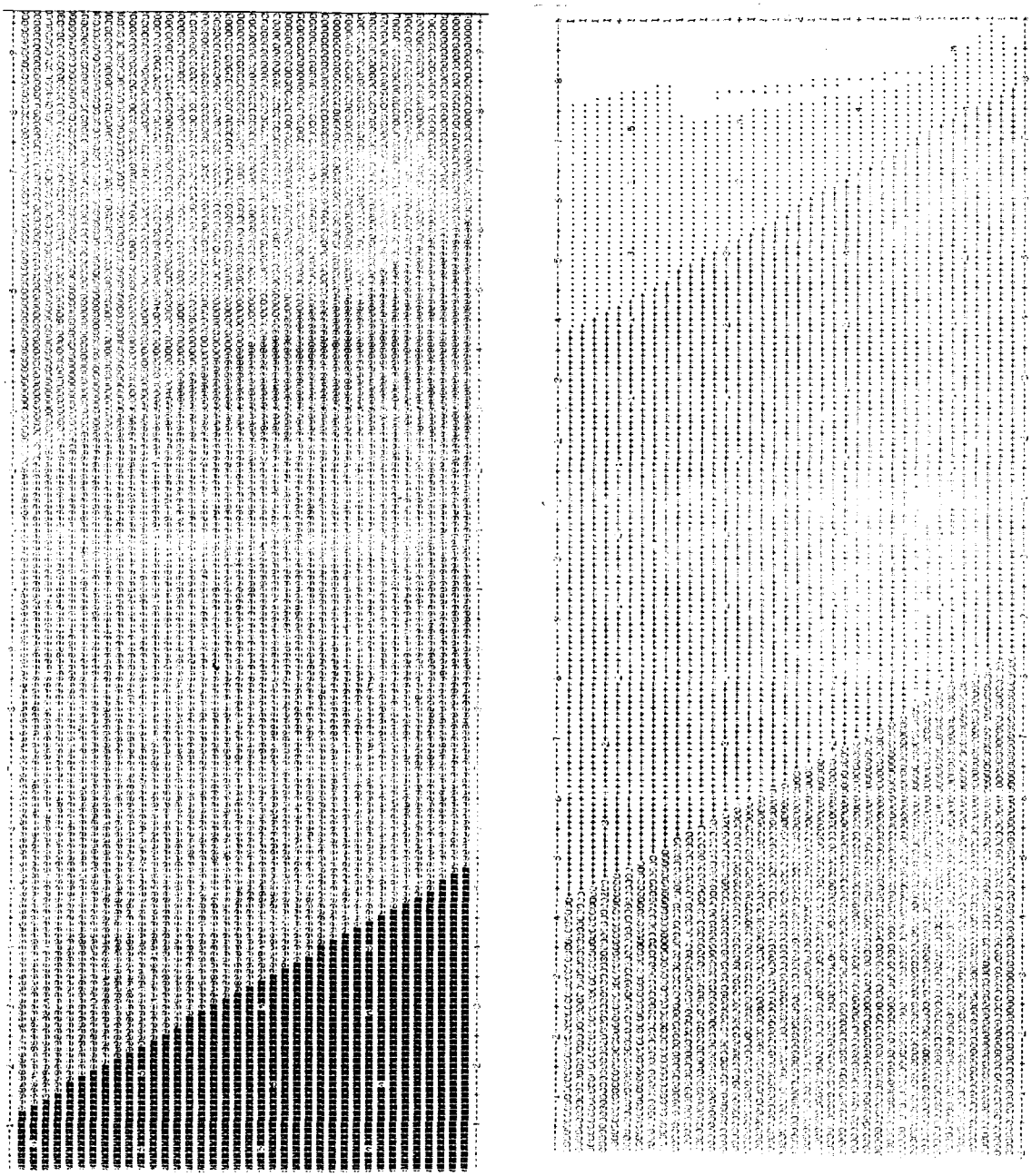

Figure 7. Trend surface map : sixth order for income values.

income values except the lakeshore area along Lake Michigan.

The factor analysis evaluates each factors strength for explanation against the other factors. The factors produced from both analysis are socio-ethnic, economic, hispanic, black, life expectancy, and multiple car ownership. For the stydy area the socio-ethnic factors are striking and are composed of nine out of the fourteen varialbles.

Table 6. R squared and R values from trend surface maps

DEGREES	R SQUARED	R VALUES
1st order	0.34969133	0.59134704
2nd order	0.47529531	0.68941665
3rd order	0.51474971	0.71745431

Fig. 6 is third order trend surface map. This map has a circular shape of income values. R squared of the third ordering map is approximately 48% of variation explained. Fig. 3 shows a high income value in the lakeshore, Elmhurst, North Lake, Stome Park, Melrose Park, Franklin Park, River Grove, and Elmwood Park. Besides, the western suburbs lakeshore areas are also predominant. The lowest income value areas, however, are shown in Pulaski, Kimball & Diversey, and the east half of North Ave. The eastern part of the study area near the lakeshore area shows the fourth or fifth out of tenth income divisions. There are many low income hispanics that are densely distributed. As a matter of fact, the income value along the lakeshore area shows as high a figure as the western part of the suburban areas.

Fig. 7 is a sixth degree ordering map. R squared of the sixth ordering map shows 52% of variation explained. The result of the income value shows almost the same as the third degree order map(Table 6).

3.4.2 SYMVU SURFACE MAPS

Fig. 8 and 9 are both income values in the study area. The SYMVU surface maps are viewed from ther southwest and southeast tipped 45 degrees and displays income values. A minimum average income valued in the study area is about $4,364 and maximum is $30,311.

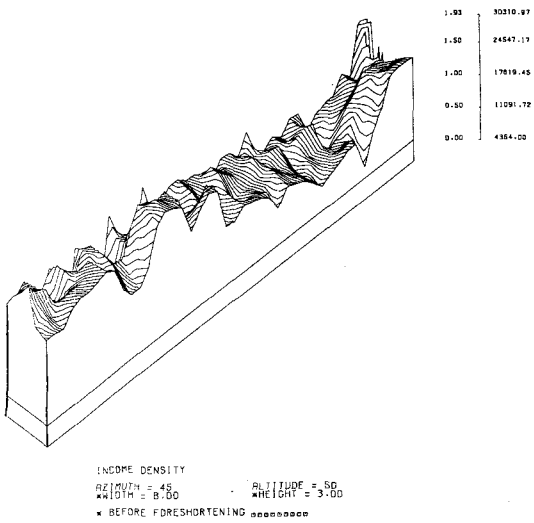

Figure 8. SYMVU 3-D surface map for income values in different azimuth, altitude, height, and width

As we expeted the high value of income featured in the Northwest part of the suburbs and Northeast part of lakeshore study areas show higher summits. The high income value areas are Elmhurst and North Lake, and a spot between Belmont Ave. and the Lakeshore. Also, two evenly high income value areas show up on Belmont and Austin &Central Avenue.

The trend surface map explained 47% in 3rd order. This map is not only enough to use a

regional income analysis without looking at residual maps, but also can be accepted as the best fit of ther regional trend for ther study area. On the other hand, in the SYMVU map procedure the altitude, 50 degrees, of ther map is not enough to show all surfaces of the study area because of a high height, 3.0". Fig. 6, however, explains better than Fig. 5 because its height was only 1.0" so that using the value of the altitude has to be considered.

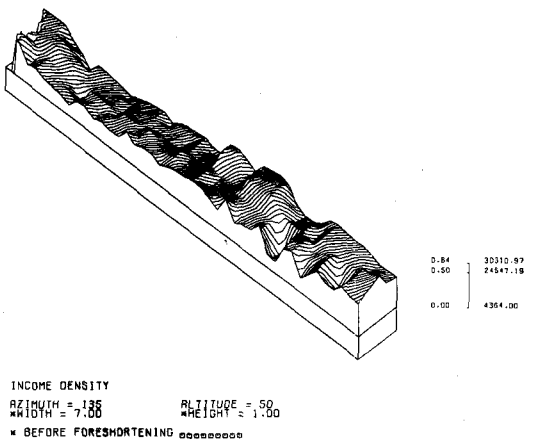

Figure 9. SYMVU 3-D surface map for income values in different azimuth, altitude, height, and width

3.4.3 SYMBOLIC MAP

Private ownership of vehicles is one of the most important describing transpotation in the United States in addition to public trasportation. In the study area, the east half of Cook County in Chicago shows a high percentage of no vehicles per house units, whereas the western half of Cook County and DuPage County indicate high vehicle ownership rates(Fig. 10).

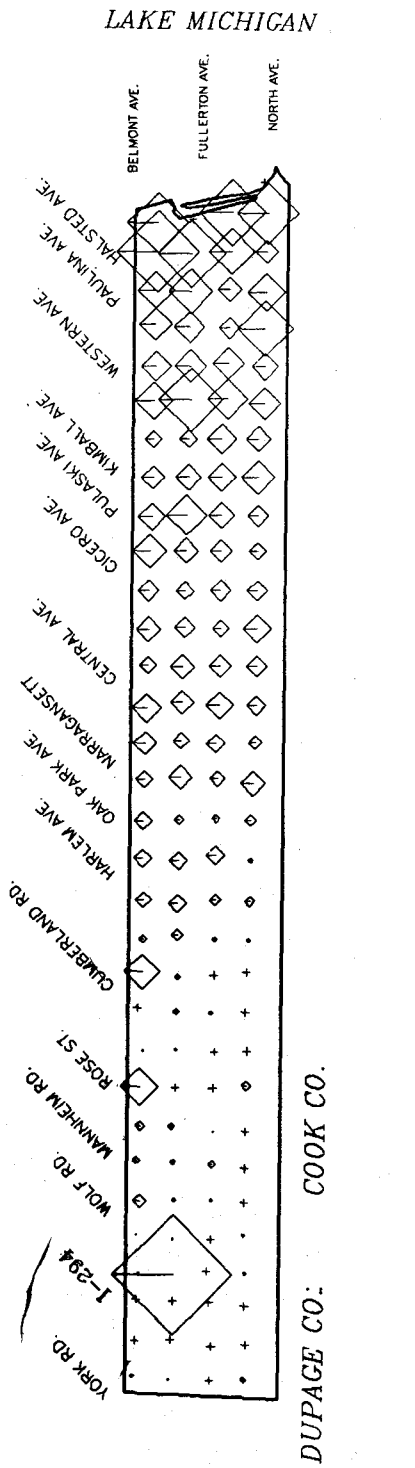

Figure 10. Symbolic map for vehicle ownership : size indicates high income holders.

A Tendency of higher rates of no vehicles in Chicago might be either people hold lower income or they use public transit. On the other hand, a tendency of low rate of no vehicle area in the west suburban areas might be either people hold high income or they do not have a chace to use public transit.

One exceptional spot which lays in the border between Cook County and DuPage County shows the highest percent of no vehicles per house units. The reason for this is that this area has the highest population density in the western suburbs. There is a significant correlation between high population density areas and high population density areas and high rates of no vehicles holders in this study area.

3. CONCLUSIONS

This study demonstrates a socio-economic status in a slice section area of Chicago metropolitan to get spatial patterns of urban windows. GSIS(Geo-spatial Information System) has been monitored with several statistic methods, and geo-spatial map presentations.

Economic configuration is a vital importance of socio-economic activities in the urban area. From the grouping analysis, the result displays that most suburban town have high income values, such as Elmjurst, Melrose Park, North Lake(Income ranges between $25,000~$30,000 : 1980 Sensus data). Cluster analysis exhibits similar result of above. Suburban areas contain high income values, such as Elmhurst, Elmwood Park, Melrose Park, and North Lake. The factors produced from both anayses of SAS and BMDP are socio-ethnic, economic, hispanic, black, life expectancy, and multiple car ownership. In the study area the socio-ethnic factor is striking, and is composed of nine out of nine out of the fourteen variables.

Geo-spatial 3-D mapping represents a socio-economic configuration of the study area. The high income value areas are Elmhurst and North Lake, and a spot between Belmont Ave, and the Lake Shore. In the study area including Cook County(Chicago city area) and DuPage County (Chicago suburbs) a minimum average income level is about $4,364 and maximum is $30,311.

ACKNOWLEDGEMENT

The author thanks to Dr. Siim soot for his direction and suggestions on this GIS project, and for his permission using the Main Computer system in the University of Illinois-Chicago.

참고문헌

1. Maquire, D.J., Computers in Geography, Longman Scientific & Technical, England. 1989, pp. 172.
2. Kim, J. and Mueller, C., Factor Analysis-statistical Methods and Practical Issues, Sage University paper 14, England, 1978, pp. 8.
3. Johnston, R.J., Multivariate statistical analysis in Geography, John Willey & Sons, Inc., New York, 1980, pp. 162.
4. Goddard, J. and Kirby, A., An Introduction to factor Analysis, Concepts and Techniques in Modern Geography, Catmog 7, GEO Abstacts LTD., England, 1976, pp. 34.
5. Unwin, D., An Introduction to Trend Surface Analysis, Concepts and Techniques in Modern Geography, Catmog 5, GEO Abstracts LTD., England, 1975, pp. 5.