Early Work in Social Area Analysis

Establishing the Idea of Typologies of Urban Neighborhoods

Common sense and everyday observation tell us that the residential sections of urbanized areas such as Cincinnati are divided into several diverse communities, ranging from slums to high income sections. It is also no secret to community leaders and planners that the social characteristics and needs of these various communities vary greatly, and that policies and programs need to be designed accordingly. But, because urban areas are too complex to allow public officials to rely completely on common sense and personal observations, planners and other students of the city constantly seek empirical tools that will provide a more reliable understanding of the changing character of large urban areas.

One such planning tool is Social Areas Analysis. It is a method of classifying and describing different communities which has been in use since Shevky and Williams(1) applied it to Los Angeles in 1949. Its originators called social areas analysis "...a method of analysis of population data ... to describe the uniformities and broad regularities observed in the characteristics of urban population."(2)

As various economists, geographers, sociologists, and other social scientists have established, there are various kinds of orderly patterns underlying the apparent unsystematic nature, growth, and changes of urban neighborhoods.(3) Social area analysis takes data from the decennial census and they are used to classify each residential census tract in the city, according to a typology which makes possible comparative studies among cities.

Census data are used to construct indicators of the economic, family, and ethnic characteristics of each neighborhood. An analysis of each tract according to its indicators is an empirically tested(4) instrument for determining the small social units of the large urban area. "Boiling down" the long list of possible variables available from the census to their three indicators is described by Shevky(5):

When the social characteristics of urban populations are studied statistically, it is observed that they follow certain broad regularities, and that the variations in the social characteristics are graded and measurable. When different attributes of a population are isolated or measured, they are found to vary in relation to other attributes of the same population in an orderly manner.

Social areas analysis as developed by Shevky and Bell was more appropriate for describing Los Angeles in 1949 than Cincinnati in 2010. Their approach has been described here mainly as an introduction to this type of methodology. A variation of this methodology developed by the Census Bureau is the actual methodology used in the present report.

The New Haven Census Use Study

In 1967 a dress rehearsal of the 1970 census was conducted in New Haven, Connecticut. Census data were combined with other information sources to develop a health information system, which in turn was used to construct social indicators at the census tract and block group level.

Components of the information system were:

a) Census data - 100 percent and 25 percent samples

b) Family Health Survey

c) Vital Records

d) Hospital obstetrical records

The purposes of the New Haven work were (1) to demonstrate how small area analysis of related health and socioeconomic characteristics might identify "high risk" populations; (2) to establish a system whereby related data can be readily retrieved and analyzed using computer technology; and (3) to produce information which would point out health issues, social

problems and needs upon which planners can act and to clearly display those data in a manner which would be convincing to budget directors and consumers.

To organize the large mass of data and to compress the social indexes into a smaller number of indicators (composite variables) one needed to arrive at a measure of socio-economic status (SES). SES was thought of as broader than also, the traditional use of the construct, and approximates an indicator of quality of social life. The large mass of data were then entered into correlation and factor analysis. Of the total number of indicators, those which are most related to each other are selected out and combined into constructs.

The one construct which seemed the most discernible was SES. From correlational analysis and factor analysis, as well as from a theoretical point of view, it was decided that SES is really a combination of five variables – income, occupational status, educational status, family organization, and housing. Health variables tended to display two kinds of clustering which made them either inefficient or too discrete for use in delineating social areas. Many health variables have a high correlation with SES, while others were not associated with SES or each other.

An SES delineation made up of a composite, rather than measured along one dimension such as family income or occupational status, is much more useful for planning purposes. The

problem with using one-dimensional definitions is that the emphasis is usually placed on either the economic or social, rather than the interaction of both. An SES delineation based solely on family income would emphasize the economic while ignoring the social qualities such as family organization and educational status. It would classify as low SES highly educated professionals who have just begun their careers. Family organization is another facet of SES. Families typified by the absence of a male breadwinner considerably reduce the potential for acquiring greater income, better housing, and higher status occupations. We assumed that the methodology of the New Haven study was valid and applied it to Cincinnati. One limitation was the non-availability of health and social data from the human service agencies.(6)

Applying the New Haven Method for Cincinnati

On the basis of the New Haven study and similar studies in Mecklenburg and Forsythe counties in North Carolina, a correlation matrix of 20 variables was developed using Cincinnati census tract data from the American Community Survey 2005-2009 (ACS) (population characteristics and housing characteristics). The 20 variables are presented in Table 2b. The Correlation Matrix (Table 1b) shows the degree of relationship between the five variables which are defined in Table 1a.

Table 1b is a matrix in which the rows correspond to the columns. Row 1 and Column 1

TABLE 1A	
DEFINITION OF SES INDEX AND ITS INDICATORS	
SES Index	The Socio-Economic Status Index is a composite scale developed from the comparative ranking scores of five indicators derived from data from the 2005-2009 American Community Survey (ACS)[a]
Family Income Indicator	Median family income
Education Indicator	Percent of population 25 years of age or older with less education than a high school diploma
Occupation Indicator	Percent of workers in unskilled and semi-skilled occupations
Family Structure Indicator	Percent of children (under the age of 18) living in married-couple, family households
Crowding Indicator	Percent of housing units with more than one person per room
[a] Previous editions and their data are based on data from the decennial census.	

are median family income which are perfectly correlated as shown by the value 1.000. The value -0.592 means that the median family income and education have a negative correlation of 0.592. Remember that the education index is the percentage of the adult population with less than a high school population. So, as income goes up, the education indicator goes down. The value -0.674 means that income and occupation (percentage of blue collar and service workers) are negatively correlated, and so on. The factor that is most highly correlated in Cincinnati with socio-economic status is edu-

cation (0.821). Occupation is second at -0.807.

This represents an identical pattern with that discovered in the first edition of this report based on the 1970 census. One of the highest correlations in the 2005-2009 data is between family structure and occupation (0.674). The correlation between family income and family structure is almost equally high (0.662).

TABLE 1B

CORRELATION MATRIX FOR SES VARIABLES, 2005-2009

	Family Income Indicator	Education Indicator	Occupation Indicator	Crowding Indicator	Family Structure Indicator	SES Index
Family Income Indicator	1.000	-0.592	-0.674	-0.260	0.662	0.794
Education Indicator		1.000	0.654	0.330	-0.517	-0.821
Occupation Indicator			1.000	0.346	-0.444	-0.807
Crowding Indicator				1.000	-0.144	-0.471
Family Structure Indicator					1.000	0.781