Reclassification of Sustainable Neighborhoods: An Opportunity Indicator Analysis in Baltimore Metropolitan Area

Chao Liu [1], Eli Knaap[2], and Gerrit Knaap[3]

[1]National Center for Smart Growth, University of Maryland, cliu8@umd.edu
[2]National Center for Smart Growth, University of Maryland, eknaap@umd.edu
[3] National Center for Smart Growth, University of Maryland, gknaap@umd.edu

Keywords: factor analysis, cluster analysis, sustainable neighborhoods

ABSTRACT

The "Sustainable neighborhoods" has become widely proposed objective of urban planners, scholars, and local government agencies. However, after decades of discussion, there is still no consensus on the definition of sustainable neighborhoods (Sawicki and Flynn, 1996; Dluhy and Swartz 2006; Song and Knaap,2007; Galster 2010). To gain new information on this issue, this paper develops a quantitative method for classifying neighborhood types. It starts by measuring a set of more than 100 neighborhood sustainable indicators. The initial set of indicators includes education, housing, neighborhood quality and social capital, neighborhood environment and health, employment and transportation. Data are gathered from various sources, including the National Center for Smart Growth (NCSG) data inventory, U.S. Census, Bureau of Economic Analysis (BEA), Environmental Protection Agency (EPA), many government agencies and private vendors. GIS mapping is used to visualize and identify variations in neighborhood attributes at the most detailed level (e.g census tracts). Factor analysis is then used to reduce the number of indicators to a small set of dimensions that capture essential differences in neighborhood types in terms of social, economic, and environmental dimensions. These factors loadings are used as inputs to a cluster analysis to identify unique neighborhood types. Finally, different types of neighborhoods are visualized using a GIS tool for further evaluation.

The proposed quantitative analysis will help illustrate variations in neighborhood types and their spatial patterns in the Baltimore metropolitan region. This framework offers new insights on what is a sustainable neighborhood.

1 INTRODUCTION

"Sustainable communities" has become a common term in the discussion among planners, scholars, policy decision makers, and the general public in recent years. The World Commission on Environment and Development (1987) proposed the most consensual definition of sustainable development to date: "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." As specified in the Brundtland report (WCED, 1987) the concept of sustainable development comprises three aspects: economic, social and environmental. However, the definition of sustainable communities and how the sustainable communities should be operationalized has not reached a consensus, especially at the neighborhood scale. In the history of neighborhood indicator use, several lessons were emphasized. First, neighborhood classification is imperative to facilitate the policy discussion. Well - structured indicators can improve the planning process in terms of understanding the attributes and issues of the neighborhoods. Second, geographic details play a specific role since policy is administrated through local geographic units and because neighborhoods themselves affect the quality of people's lives (Sawicki, 1996). Third, each indicator should have its unique characteristics and should be unbundled. Fourth, some of the indictors should be used as tools for policy evaluation.

On June 16, 2009, the U.S. Department of Housing and Urban Development (HUD), U.S. Department of Transportation (DOT), and the U.S. Environmental Protection Agency (EPA) formed the Partnership for Sustainable Communities to help communities nationwide improve access to affordable housing, increase transportation options, and lower transportation cost while protecting the environment. The partnership agencies incorporate six principles of livability into federal funding programs, policies, and future legislative proposals. The six livability principles include: provide more transportation choices, promote equitable and affordable housing, enhance economic competiveness, support existing communities, coordinate and leverage federal policies and investment, value communities and neighborhoods. For Baltimore region in particular, one more principle is added: protect the Chesapeake Bay. Under the seven principles, the Baltimore Metropolitan Council (BMC) is proposing a Regional Plan for Sustainable Development (RPSD) that is aiming to support sustainable regional planning efforts that integrate housing, land-use, economic and workforce development, transportation, and infrastructure developments in a collective manner to create more job and economic opportunities for Baltimore region. The Opportunity Collaborative is the consortium charged with developing the RPSD for the Baltimore region. The consortium is comprised of 26 members including local governments, state agencies, universities and nonprofit organizations. Over the process of developing the RPSD, opportunity mapping is a tool to analyze a variety of opportunities at the neighborhood level that exist throughout the region. Understanding the disparities of the access to different opportunities is a crucial step to developing the RPSD and further helps the Opportunity Collaborative to connect housing, transportation and workforce development in the region. Meanwhile, we also obtained a rich dataset through the opportunity mapping process to further analyze neighborhood typology.

This paper develops a quantitative method of classifying neighborhoods. We developed a rich data set containing a total of 128 indicators measuring multiple dimensions including education, crime and neighborhood quality, health, employment and transportation, and population and labor force through using Geographic Information System (GIS). Then, exploratory factor analysis (EFA) is utilized to generalize the dimensions of the neighborhood attributes based on raw data. Cluster analysis is the next step to group the neighborhoods based on EFA results. Clusters are finally visualized using GIS tools to illustrate disparities of opportunity within the region.

2 PREVIOUS WORK ON CLASSIFICATION OF SUSTAINABLE NEIGHBORHOOD

2.1 State of the practice

In recent years, there are several HUD grantees that developed the sustainable community indictors. For example, the Puget Sound Regional Council and Kirwan Institute jointly developed "Growing Transit Communities," funded through an SCRPG grant. The steering committee worked with a variety of stakeholders and advocates throughout the region to select a set of opportunity indicators representing five key elements of neighborhood opportunity: education, economic health, housing and neighborhood quality, mobility and transportation, and health and environment. The opportunity mapping tool has been a catalyst for community discussion and has led to many findings and policy implications, including: (1) the total population, disabled population, and the foreign-born population are more or less evenly distributed across the opportunity spectrum. (2) Subsidized housing can be a strategy to help disadvantaged population to access healthy food or high-performing schools. (3) About half of the people living in poverty are located in the areas of low or very low opportunity. (4) About one third of current and proposed light rail stations are in areas of low and very low opportunity.

The Central Texas Opportunity initiative was established by the Community Partnership for the Homeless and involved a steering committee representing a consortium of organizations in the Central Texas region. The committee worked with the Kirwan Institute to identify and gather data for indicators of opportunity in the region. Their categories of indicators include: education, economic, mobility and transportation, health and environment, and neighborhood quality. The results show that areas of high and low opportunities are not evenly distributed throughout the region. Specifically, higher opportunity areas in the region are primarily concentrated west of I-35, which is also the divider in education conditions. In terms of housing and neighborhood quality, public health and environment, they also found that neighborhoods west of I-35 performed better. Similar to the results in the Central Puget Sound Region, the findings in Austin demonstrated that subsidized housing sites are less likely to exist in high-opportunity areas.

The Denver SCRPG-funded initiative is focused on ensuring that the region's significant investment in new rail and bus service will provide greater access to opportunity and a higher quality of the life for all the residents, especially for the disadvantaged populations who benefit the most from transit service. The five categories of opportunity maps are: population and demographic characteristics, housing, job and economic development, education, health. The opportunity mapping results show that the region has a significant opportunity to increase transportation options through transit expansion. Many low-income and other economically disadvantaged populations, however, cannot currently take advantage of affordable transit choices. In addition, even though many of the region's affordable housing units are located near the current or proposed transit stations, the demand for housing near transit is expected to grow fast in the coming decades.

However, none of the above attempts to identify the types of the neighborhood or a clear analysis framework.

2.2 State of the art

There are several earlier studies that have studied neighborhood inequality or neighborhood distress that have used a wide range of social and economic variables as a tool for assessing neighborhood disparity throughout neighborhoods and communities (Hill et al. 1998; Sawicki & Flynn, 1996; Mikelbank, 2004; Song & Knaap, 2007; Vicino et al. 2007; Hanlon, 2010; Jennings, 2012).

Hill et al. (1998) collected data for 508 central cities and used hierarchical cluster analysis and discriminant analysis to reveal five types of distressed central cities and eight distinct types of healthy central cities. In this study, cluster analysis was utilized to group observations that are similar based on multiple variables and discriminant analysis was used to identify the driving forces that distinguish the neighborhoods. Mikelbank (2004) extended this approach to 3,567 non-central city incorporated places by using the same methodology. The results revealed two types of healthy suburbs and two types of middle-America suburbs. The two studies were both national in scope and provided an extensive breadth of study area. Regional and neighborhood level differences, however, were neglected in the analysis.

Mikelbank (2006) then further analyzed suburban places within two similar census regions: the East North Central and Middle Atlantic regions. He only focused on each suburb's share of metropolitan – level population change through the use of location quotients. The results suggested that the comparison across different groups is more relevant in the region.

Vicino et al. (2007) investigated 12 Consolidated Metropolitan Statistical Areas of Washington-Baltimore, Philadelphia, New York, and Boston by using principal components analysis to identify the differences among the study areas. Then they used cluster analysis based on the previously derived components. They distinguished the following types of suburbs: Middle America, Affluent, Places of Poverty, Immigrant Gateways, and Black Middle Class. Their results highlighted that the growing disparities not only between cities and suburbs as traditionally thought, but also among suburbs themselves.

Hanlon (2010) used a similar approach in identifying five distinct types of inner-ring suburbs from among the more than 1,700 inner-ring suburbs. The dimensions identified in the research are race, class, ethnicity, and socio-economic status. The types are: Vulnerable, Ethnic, Lower Income, and Mixed, Old, and Middle Class.

In addition to the efforts that focused on race, class, socio-demographic attributes, Song and Knaap (2007) also used similar quantitative methods for classifying neighborhood types based on the measurements of physical neighborhood form. Through using 21 urban form attributes including density, design, diversity, accessibility and natural environment measurements, this study derived eight dimensions: street design, density, commercial use, transit, home size, public and industry use, nature environment, and multi-family use. The paper was only intended to study the physical urban form typology, which did not consider any social and economic characteristics.

3 DATA

3.1 Study Area

Our study area is the Baltimore Metropolitan Area, with a population over 2.5 million people, it is one of the twenty largest metropolitan regions in the United States. This region consists of Baltimore City, and Ann Arundel, Baltimore, Carroll, Harford and Howard Counties (Figure 1).

Figure 1. Study Area---- Baltimore Metropolitan Area

3.2 Neighborhood definition

Before elaborating the details on the sustainable indicator measurements, it is important to note that an appropriate unit of analysis is essential to the analysis. In past studies, census tracts, traffic analysis zones (TAZs), zip codes or other pre-defined neighborhood boundaries have been used to compute measurements of neighborhood types. Census tracts, however, are a widely used geographic unit in most of the studies and most of data are available at the census tract level. We also want to keep all the data in a consistent format for further analysis.

3.3 Neighborhood indicator selection

As part of the structure of the Opportunity Collaborative, a Nexus Committee was been formed to oversee the development of RPSD, the consortium's opportunity mapping project, and indicator category development. At this point, the Nexus Committee has identified six categories of indicators, for which the NCSG has identified numerous potential indicators: education, housing, neighborhood health and environment, neighborhood quality and social capital, workforce development, and transportation and mobility. Based on input from the Nexus Committee regarding the issues and concerns in the region, experience of other regions, as well as the literature justifying the connection between indicators and opportunity, we identified a total of 128 indicators for the six categories. We then refined the data to 48 indicators for the factor analysis. Descriptive statistics of all the variables are provided in Table 1. A detailed inventory of the indicators, measurement approaches, and data source are discussed here below.

Education data were originally obtained from the Maryland State Department of Education (MSDE) Maryland Report Card dataset, which includes school locations, student enrollment, students' performance on different subjects by grade levels, and teacher's qualification. All the information is available at three different levels (i.e. elementary, middle and high school). In addition, we also gathered the information on advanced placement and SAT scores by subject of all the high schools in the region. All the education indicators were merged to the school boundaries that were obtained from the school board of each county (with the exception of high schools in Baltimore City[1]). Then the school boundary data with the attributes are further aggregated to census tract for further analysis. Education indicators include:

- AdvProES: Percent of elementary school students obtained proficient and advanced level of all the subjects;
- 3ReadingES: Percent of elementary school students obtained proficient and advanced level of 3rd grade reading;
- 3MathES: Percent of elementary school students obtained proficient and advanced level of 3rd grade math;
- 5ReadingES: Percent of elementary school students obtained proficient and advanced level of 5th grade reading;
- 5MathES: Percent of elementary school students obtained proficient and advanced level of 5th grade math;
- TeacherES: Percent of high qualified[2] elementary school teachers;
- AdvProMS: Percent of middle school students obtained proficient and advanced level of all the subjects;
- 8MathMS: Percent of middle school students obtained proficient and advanced level of 8th grade math;
- TeacherMS: Percent of high qualified middle school teachers;
- AdvProHS: Percent of high school students obtained proficient and advanced level of all the subjects;
- TeacherHS: Percent of high qualified high school teachers;
- AP35HS: Percent of students obtained advanced placement with scores 3-5;
- CRHS: SAT scores of critical reading;
- MHS: SAT scores of math;
- WHS: SAT scores of writing;

[1] We did not merge the high school performance data to the Baltimore City. In 2005, Baltimore City Public School System (BCPSS) has initiated a citywide system of choice and neighborhood school enrollment that are no longer assigned by school districts.

[2] According to MSDE definition, highly qualified teacher referred to: requires verification of 3 years of full-time professional school-related experience, 6 semester hours of acceptable credit; and a master's degree, or a minimum of 36 semester hours of post baccalaureate course work which must include at least 21 hours of graduate credit.

- ProxPrivateCareer: Euclidian distance to the closest private career centers.

Neighborhood health and environment data were collected from multiple sources. Respiratory risk, neurological risk, and cancer risk data were obtained from National Air Toxics Assessments (NATA) of the U.S. Environmental Protection Agency (EPA 2005). Infant mortality rate, teen birth rates, late or no prenatal care rates, and low birth weights rates were gathered from the Department of Health and Mental Hygiene (DHM 2011). Hospital data including number of beds and location were also collected from DHMH. Grocery store location data were obtained from Quarterly Census Employment and Wages (QCEW) of the Department of Labor and Licensing (DLLR 2007). Ambulance service locations were gathered from QCEW and MdProperty View data of Maryland Department of Planning (MDP 2010). Park data were collected from the Maryland Department of Natural Resources (DNR 2012).

Neighborhood health and environment indicators include:

- CancerRisk: estimated probability per million people at census tract level of developing cancer over a lifetime by combining the information from modeled exposure to air toxics;
- RespRisk: refers to “hazard index” at census tract level where a value of “1” is considered safe and any number above 1 that can potentially result in health effects;
- NeuRisk: refers to “hazard index” at census tract level where a value of “1” is considered safe and any number above 1 that can potentially result in health effects;
- AmbSerArea: derived from network analyst tool to estimate 10 – minute service area using Maryland State Travel Modeling (MSTM[3]) network. Percent of census tract that is covered by service areas was calculated;
- HospitalProximity: kernel density of hospital at census tract level weighted by the number of beds using a 30-mile search radius;
- AccePark: kernel density of park at census tract level weighted by number and size of the parks using a half mile search radius;
- GroProx: Access to grocery stores: kernel density of grocery stores at the census tract level using a half-mile search radius;
- InfantMortality: Number of infant deaths per 1,000 live births at census tract level. Infants are defined as children under one year of age;
- TeenBirth: Percent of all births that are to mothers 15 to 19 years old at census tract level;
- NoPrena: Number of births to women receiving late (from third trimester) or no prenatal care at the census tract level;
- LBW: Percent of all births that are babies of low birth weight (less than 2,500 grams or 5.5 pounds) at census tract level.

Neighborhood quality and social capital data were primarily collected from the U.S. Census American Community Survey (2007-2011) 5-year estimations. Religious and social

[3] MSTM is a multi-layer model working at a regional, statewide, and urban level. The model is driven by the economic and land use assumptions and includes both person and freight travel. The passenger and truck trips from both the BMC regional (long-distance) and Statewide (short-distance) model components provide traffic flows allocated to a time period (AM peak, PM peak or off-peak) that are input to a single Multiclass Assignment.

organizations data were extracted from QCEW and MdProperty View data. A crime risk index was obtained from Applied Geographic Solutions (AGS 2011). All the data in this category are available at the census tract level.

Neighborhood quality and social capital indicators include:

- Religious_Social: kernel density of social and religious organizations[4] using a half – mile radius and further aggregated to census tract;
- Public_Inst: kernel density of social and religious organizations using a half – mile radius and further aggregated to census tract;
- Population density: number of people per square mile at census tract level;
- Median income: median income at census travel level;
- % with HS diploma: Percent of population with high school diploma or higher at census tract level;
- % with Bachelors: Percent of population with bachelor degree or higher at census tract level;
- Population2544: Population ages of 25 – 44 years at census tract level;
- Racial diversity: Diversity index = $\frac{1-\sum_{k=1}^{n} P_k^2}{1-\frac{1}{k}}$

  Where P_k is the percent of population in race group K in a census tract, and n is the total number of race groups in a census tract.
- Crime Index Total: Total crime risk index at census tract level;
- Crime Index Property: Property crime risk index at census tract level;
- Crime Index Personal: Personal crime risk index at census tract level;
- Crime Index Rape: Rape crime risk index at census tract level;
- Crime Index Murder: Murder crime risk index at census tract level;
- Crime Index Robbery: Robbery crime risk index at census tract level.

Workforce and transportation data were derived from the following sources: employment data were obtained from the Longitudinal Employer and Housing Dynamics (LEHD 2010) of the U.S. Census. Workers by skill level data were gathered from the U.S. Census Decennial data . Skill level estimates are based solely on educational attainment levels of all the people 25 years and older. Low-skill workers include people have no more than a high school education (high school diploma or less). Middle-skill workers include people have associate or post-secondary degrees. High-skill workers include people who have a bachelors degree or higher. We used the same criteria for job skill classification. To calculate the job accessibility index, we used a Visual Basic script and travel skim data of AM peak hour conditions from the MSTM travel model. Accessibility scores at the State Modeling Zone (SMZ) were then aggregated to census tract level.

Workforce and transportation indicators include:

[4] Religious and social organizations include: social advocacy organizations; Grant-making and giving services; social and volunteer clubs; non-profit trade associations; labor unions and similar organizations; political organizations; veteran's organizations. Public institution include: public library, museums, book stores, etc.

- HskillAUTO: Total number of high - skill jobs that can be accessible within 30 minutes auto commute;
- HskillTRANSIT: Total number of high - skill jobs that can be accessible within 30 minutes transit commute;
- MskillAUTO: Total number of middle - skill jobs that can be accessible within 30 minutes auto commute;
- MskillTRANSIT: Total number of middle - skill jobs that can be accessible within 30 minutes transit commute;
- LskillAUTO: Total number of low - skill jobs that can be accessible within 30 minutes auto commute;
- LskillTRANSIT: Total number of low - skill jobs that can be accessible within 30 minutes transit commute;
- HighSkillPct: Percent of high-skill workers at census tract level;
- MidSkillPct: Percent of middle-skill workers at census tract level;
- LowSkillPct: Percent of low-skill workers at census tract level;

Table 1. Summery Statistics for all variables (N=663)

Variables	Minimum	Maximum	Mean	Std. Dev.
AdvProES	0.00	98.86	78.43	16.93
3ReadingES	0.00	100.34	79.53	17.14
3MathES	0.00	100.00	81.91	17.01
5ReadingES	0.00	100.00	84.11	17.07
5MathES	0.00	100.00	76.96	18.50
TeacherES	0.00	90.76	56.14	15.83
AdvProMS	0.01	185.82	74.60	31.24
8MathMS	0.01	180.00	62.72	31.88
TeacherMS	0.01	177.80	59.32	27.89
AdvProHS	0.00	100.00	74.54	19.49
TeacherHS	0.00	85.09	55.06	15.85
AP35HS	0.00	88.52	37.86	26.61
CRHS	0.00	578.74	423.00	103.30
MHS	0.00	609.15	424.79	109.98
WHS	0.00	581.44	393.66	116.75
ProxPrivateCareer	0.00	26.00	3.70	4.45
Religious_Social	38.50	15287.00	1397.86	2218.10
Public_Inst	0.00	16875.00	454.83	954.87
Population 2544	0.00	74.40	34.76	11.80
Racial Diversity	0.00	0.69	0.32	0.18
% with HS Diploma	0.45	1.00	0.86	0.11
% with Bachelors	0.00	1.00	0.33	0.21
Median Income	0.00	196250.00	68100.00	32339.95
Population Density	0.00	97.32	9.09	10.09
Crime Index Total	7.00	639.00	163.50	142.75
Crime Index Property	3.00	470.00	142.68	109.48

Crime Index Personal	9.00	1016.00	232.93	238.16
Crime Index Murder	1.00	1813.00	276.95	381.87
Crime Index Rape	3.00	449.00	89.78	80.55
Crime Index Robbery	4.00	1718.00	305.04	388.77
CancerRisk	0.00	0.00	0.00	0.00
NeruRisk	0.03	0.44	0.09	0.04
RespRisk	1.22	7.85	3.27	1.27
InfantMortality	0.00	0.04	0.01	0.01
TeenBirth	0.00	0.50	0.07	0.07
NoPrena	0.00	0.27	0.07	0.05
LBW	0.00	0.40	0.09	0.06
AmbSerArea	0.00	52.37	19.10	13.89
HospitalProximity	21.00	752.00	549.57	218.25
AccePark	0.00	153.30	30.68	24.25
GroProx	0	1734.6	211.21	224.51
HskillAUTO	1097.10	208290.00	108002.40	64271.92
HskillTRANSIT	5.96	43204.00	8559.16	11970.39
MskillAUTO	763.36	133740.00	69973.79	41935.65
MskillTRANSIT	3.77	24775.00	4907.58	6837.00
LskillAUTO	3949.70	719320.00	373826.10	214320.16
LskillTRANSIT	27.41	112290.00	23902.28	31143.33
LowSkillPct	0.00	1.00	0.37	0.19
MidSkillPct	0.00	0.67	0.27	0.08
HighSkillPct	0.00	0.93	0.36	0.22

4 METHOD

4.1 Factor analysis

We began by applying Exploratory Factor Analysis (EFA) to identify a parsimonious pattern of the correlated indictors. EFA is a statistical approach used to determine correlation among variables. By grouping the variables based on strong correlations, EFA can reduce the redundant variables and generate more a cleaner structure of the data for the further analysis of neighborhood typology.

Based on 48 variables that measure different categories, EFA extracts eight dimensions (components). The EFA results are shown in Table 2. Principle Component Analysis (PCA) was utilized for extraction and Varimax was used for rotation of the factors. Several iterations were conducted for the analysis including incorporating different variables and using different extraction and rotation approaches. Only the final results were reported in Table 2. The extracted factors can explain about 74.19% of the total variance among the 48 indicators.

Table 2. Factor Analysis of Sustainable Neighborhood Dimension

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
	Neighborhood Quality	Neighborhood Health	Education - Elementary School	Education - High School	Labor Force	Employment Accessibility	Education - Middle School	Population & Diversity
Crime Index Personal	0.88							
Crime Index Total	0.88							
Crime Index Robbery	0.82							
Crime Index Rape	0.77							
Crime Index Property	0.76							
Crime Index Murder	0.74							
Religious_Social	0.70							
Public_Inst	0.58							
RespRisk		0.93						
CancerRisk		0.91						
AmbSerArea		0.89						
HospitalProximity		0.78						
AccePark		0.66						
GroProx		0.61						
TeenBirth		0.59						
NeruRisk		0.59						
InfantMortality		0.53						
LBW								
NoPrena								
5ReadingES			0.95					

AdvProES			0.94					
3MathES			0.94					
3ReadingES			0.93					
5MathES			0.89					
TeacherES			0.75					
CRHS				0.96				
MHS				0.95				
AdvProHS				0.93				
TeacherHS				0.91				
WHS				0.89				
AP35WHS				0.69				
% with Bachelors					0.95			
HighSkillPct					0.94			
LowSkillPct					-0.85			
Median Income					0.69			
% with HS Diploma					0.66			
MidSkillPct					-0.52			
HskillAUTO	0.42					0.85		
LskillAUTO	0.44					0.85		
MskillAUTO	0.44					0.84		
HskillTRANSIT						0.76		
LskillTRANSIT						0.75		
MskillTRANSIT						0.75		
ProxPrivateCareer						-0.64		
AdvProMS							0.86	
TeacherMS							0.82	

8MathMS							0.81	
Population 2544								0.85
Racial Diversity								0.71
% of Variance	17.34	11.47	11.37	11.31	8.56	7.33	5.34	3.25

There are eight factors we extracted: neighborhood quality, high school education, neighborhood health, elementary school education, employment, labor force, middle school education, and population diversity. Among the eight factors, the neighborhood quality factor accounts for the highest percentage of the total variance while the population diversity factor has the lowest percentage. The neighborhood quality reflects neighborhoods with higher crime risk, higher population density, easier access to religious and social organizations, public institutions, and private career centers. Among all the indicators of this factor, crime risk indices contribute the most to the factor loadings.

It is interesting to note that we initially had 15 indicators of education. EFA analysis groups them into three factors by school levels instead of just one factor. High school education has the second highest explanatory power of all the eight factors, which explains 11.31% of the total variance in the data. The elementary school education factor and middle school education factor account for 11.37% and 5.34%, respectively. The high school education factor reflects the high school quality in neighborhood that has higher SAT scores, better student performance on all subjects, and a higher percent of teachers with advanced qualifications. The elementary school education factor indicates the level of elementary school quality in the neighborhood: more students can achieve better scores in all subjects, higher math and reading scores of 3rd grade and 5th grade students, and a higher percent of teachers with advanced qualifications. The middle school education factor reflects similar cases with slightly lower loadings on the factor.
The third factor relates to neighborhood health: higher cancer and non-cancer risk, higher infant mortality rate, higher teen birth rates, more access to hospitals, and less access to park and recreation sites.

Factor 5 reflects employment opportunity in the neighborhoods: higher auto and transit accessibility of all skill levels of employment. Factor 6 relates to labor force: a higher percent of high skill labor force, and a higher percent of population with high school diploma or greater, a higher percent of population that has a bachelor degree and higher, and a higher median income.

The last factor reflects population and diversity and indicates a higher percent of population ages 25-44 years and a higher racial diversity in the neighborhoods.

4.2 Cluster analysis

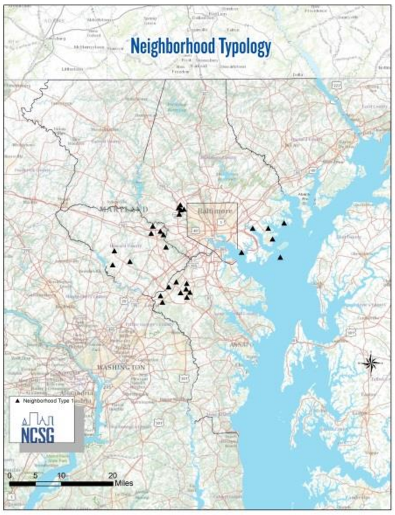
Gaining a better knowledge of the sustainable neighborhood typology in the Baltimore metropolitan region is the primary goal of this research. To identify groups of neighborhoods that are similar to each other but different from neighborhoods in other groups based on the eight dimensions that are derived from the above analysis, we used cluster analysis. There are a variety of methods of cluster analysis; we used the K-Means clustering method. The desired number of clusters needs to be determined in advance. The algorithm is called K-means, where K is the number of clusters. The initial step in K-means clustering is finding the K centers. This is done iteratively. The process starts with an initial set of centers then modifies them until the change between two iterations is deemed sufficiently small. After the initial cluster centers have been selected, each case is assigned to the closest cluster, based on its distance from the cluster centers. After all the cases have been assigned to clusters, the cluster centers are recomputed based on all the cases in the cluster. This process is repeated until no cluster center changes appreciably or the maximum number of iterations is reached.

A series of iterations by specifying the number of clusters was conducted. Based on the interpretability and cluster statistics, the five-cluster scenario is found to be the most meaningful and reasonable solution. The means of five-cluster is presented in Table 3 and spatial distribution of the clusters is illustrated in Figure 2.

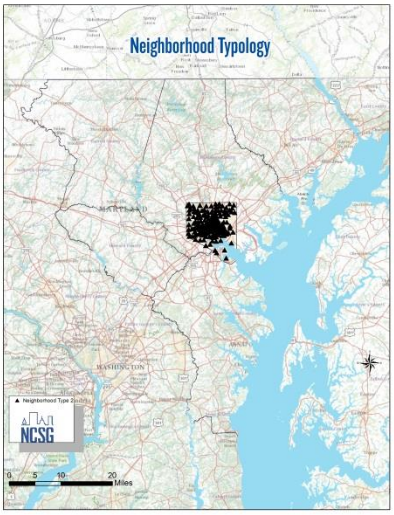
Table 3. Cluster Means in eight dimensions

	Cluster				
	1	2	3	4	5
Neighborhood Quality	-0.96	1.35	-0.19	-0.63	-0.95
Neighborhood Health	0.23	-0.40	1.15	-0.53	-0.03
Education -Elementary School	0.72	-0.16	0.19	0.28	-3.72
Education - High School	-3.44	-0.31	0.32	0.39	-0.14
Labor Force	0.41	-0.13	0.38	-0.28	0.61
Employment Accessibility	-0.09	0.29	0.02	-0.19	-0.19
Education - Middle School	0.38	-0.12	-0.26	0.21	0.11
Population	0.12	0.02	-0.25	0.19	-0.38
counts	28	181	175	253	26
Percentage of all	4%	27%	26%	38%	4%

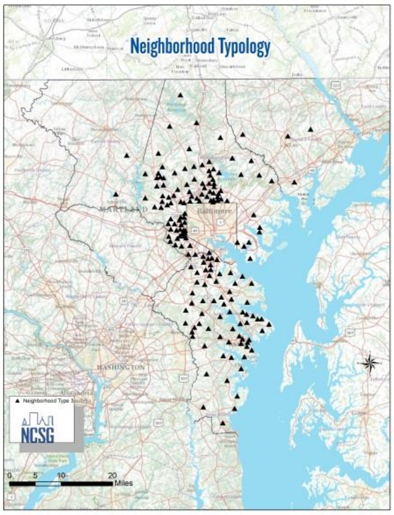
Figure 2. Neighborhood Classification



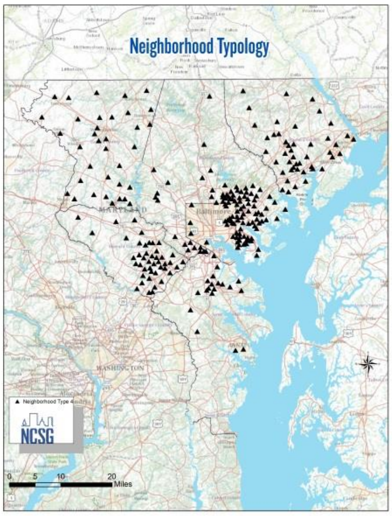
2a. Neighborhood Type 1



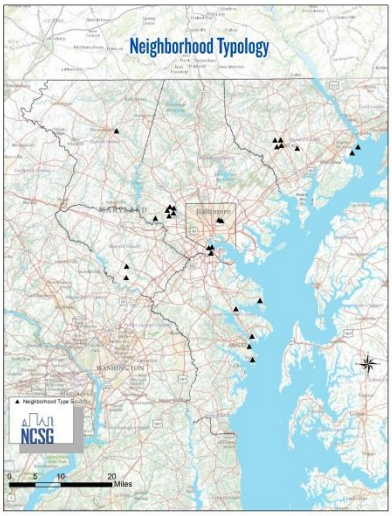
2b. Neighborhood Type 2



2c. Neighborhood Type 3



2d. Neighborhood Type 4



2e. Neighborhood Type 5

Figure 2a shows the spatial distribution of neighborhoods in the Baltimore region. Neighborhood Type 1 account for 4.2% of the total neighborhoods. Based on the attributes of Table 3, this neighborhood has features showing: better environment and health quality, better middle school quality, and higher median income, and higher percent of high-skill workers. However, job accessibility, high school quality and access to social organization are relative low.

As shown in Figure 2b, Neighborhood Type 2 are predominantly concentrated in Baltimore, and represents the second largest neighborhood type with 27% of the total neighborhoods. Table 3 reveals the cluster means of this type indicating that this cluster has high employment accessibility and high population diversity. The information also shows that these neighborhoods are characterized by high crime rate and higher access to religious and social organizations and public institutions.

Most of the Type 3 neighborhoods are located in Baltimore County and Ann Arundel County. This type of neighborhood has almost the opposite characteristics of Type 2 neighborhoods: less crime, better health quality, better education quality of all three levels, higher income and percent of high-skill workers and lower population diversity.

Type 4 neighborhoods make up the largest neighborhood type with almost 40% of all the neighborhoods in the Baltimore region. Like Type 3 neighborhoods, Type 4 neighborhoods also have high education quality at all levels. They are also characterized by lower crime rates, lower access to religious and social organizations, lower access to public institutions, lower environment and health quality, and lower employment accessibility.

Type 5 neighborhoods are located sporadically across the Baltimore region. These types of neighborhoods represent about only 4% of the total neighborhoods. According to values in Table 3, these neighborhoods are characterized by higher median income, higher percent of high-skilled workers, and better middle school quality.

5 CONCLUSIONS

In this paper we developed a quantitative analysis framework to classify neighborhoods in the Baltimore Metropolitan area. Exploratory Factor Analysis (EFA) was utilized to derive eight dimensions based on 48 predefined indicators. These factors are: neighborhood quality, neighborhood health, education quality of elementary school, education quality of high school, labor force, employment accessibility, education quality of middle school, and population and diversity. Cluster analysis was then applied to identify five types of neighborhoods based on their similarity and dissimilarity according to the eight factors that were derived from the previous EFA. Finally, cluster types were displayed using ArcGIS to reveal the spatial distribution of different types of neighborhoods. This analysis approach provides a useful framework for analyzing neighborhood landscapes across multiple dimensions. The results suggest that the predominant type 4 neighborhoods are characterized by better education quality, larger share of population aged 25-44 years and higher racial diversity are located in the suburbs of the region. The second largest are the central city neighborhoods with higher employment accessibility, higher access to social organizations and public institutions, and higher crime rates. The third largest group is identified as the suburbs of Baltimore, which are characterized by better health quality, better education quality at all three levels, higher income and percent of high-skill workers and lower population diversity.

These results indicate that it is promising to classify neighborhoods using multiple categories. The analysis framework developed in this study is a useful and statistically based tool to assess the neighborhoods and can further be used as inputs for policy evaluation process.
It is important, however, to point out several limitations of the approach:

First, we only include education, health, neighborhood quality, and employment in the analysis. At the moment, we are still working on getting additional indicators that are lacking from the current analysis. Second, while EFA was applied in this analysis, there are several factors that have cross loadings on other factors. Solving this issue requires the Confirmative Factor Analysis (CFA) approach. We also find that factors covary with each other. Therefore, Structural Equation Modeling (SEM) should be conducted to investigate the interactions among factors. Third, it is also important to note that this classification scheme was developed for the Baltimore Metropolitan region only. Future research should be carried out for other regions to get a general pattern of neighborhood classification picture. Finally, and perhaps most importantly, although the analysis provides important statistical measures of differences in neighborhood types, they reveal less about what neighborhoods should be considered as “sustainable.” This we will also explore in future work.

REFERENCES

Brundtland, Gro Harlem. "World commission on environment and development." Our common future (1987).

Dluhy, Milan and Nicholas Swartz. (2006). Connecting Knowledge and Policy: the Promise of Community Indicators in the United States. Social Indicators Research. 79. DOI10.1007/s1205-005-3486-2.

Galster, George. (2009). Neighborhood Social Mix: Theory, Evidence, and Implications for Policy and Planning. Paper presented at the 2009 International Workshop at Technion University: "Planning for / with People" Presented at Haifa, Israel June, 2009.

Jennings, J. (2012). Measuring neighborhood distress: a tool for place-based urban revitalization strategies. Community Development, 43(4), 464-475.

Hanlon, Bernadette, John R Short, and Thomas J Vicino. Cities and suburbs: New metropolitan realities in the US. Taylor & Francis, 2010.

Hill, Edward W, John F Brennan, and Harold L Wolman. "What is a central city in the United States? Applying a statistical technique for developing taxonomies." Urban Studies 35.11 (1998): 1935-1969.

Reibel, Michael, and Moira Regelson. "Quantifying neighborhood racial and ethnic transition clusters in multiethnic cities." Urban Geography 28.4 (2007): 361-376.

Mikelbank, Brian A. "A typology of US suburban places." Housing Policy Debate 15.4 (2004): 935-964.

Sawicki, David and Patrice Flynn. (1996). Neighborhood Indicators: A Review of the Literature and an Assessment of Conceptual and Methodological Issue. Journal of the American Planning Association. 62(2).

Song, Yan and Gerrit –Jan Knaap. (2007). Quantitative Classification of Neighborhoods: The Neighborhoods of New Single-family Homes in the Portland Metropolitan Area. Journal of Urban Design. 12(1).

Vicino, Thomas J, Bernadette Hanlon, and John Rennie Short. "Megalopolis 50 years on: the transformation of a city region." International Journal of Urban and Regional Research 31.2 (2007): 344-367.