SpAM

SpAM (Spatial Analysis and Methods) presents short articles on the use of spatial statistical techniques for housing or urban development research. Through this department of Cityscape*, the Office of Policy Development and Research introduces readers to the use of emerging spatial data analysis methods or techniques for measuring geographic relationships in research data. Researchers increasingly use these new techniques to enhance their understanding of urban patterns but often do not have access to short demonstration articles for applied guidance. If you have an idea for an article of no more than 3,000 words presenting an applied spatial data analysis method or technique, please send a one-paragraph abstract to* rwilson@umbc.edu *for review.*

Measuring Distance to Resources

Ron Wilson
University of Maryland, Baltimore County

Alexander Din
Dewberry

Abstract

Mapping counts or rates of residents by areal geographies is useful for visualizing distributions across regions. However, this approach limits the understanding of resource proximity to visual approximations. Taking advantage of exact location information in a geographic information system (GIS), direct proximity statistics can be created by geoprocessing residence locations to population centers. In this article, we demonstrate how to geoprocess location information to create a table of the distances between resident locations and the nearest population centers to gain a more precise understanding of how far people live, as groups, from their closest resource centers.

Relationship Between Voucher Residences and Population Centers

Proximity analysis is an analytical approach used to capture the distance between neighboring locations. Measuring the proximity of a residential location to the closest population center (PC) is a common approach toward understanding general resource accessibility. Similarly, the proximity of PCs to residence locations is useful for understanding distance costs from a central location to nearby residences. Because residence locations and the center of a PC are represented as discrete spatial locations in a GIS, distance measurement techniques can succinctly capture proximity and present the results as a set of statistics. A proximity analysis also allows for the visualization of the geographic distribution of locations and their distances around a PC that can be used in conjunction with distance analysis tables that offer more comprehensive pictures of accessibility or outreach challenges.

We demonstrate the proximity analysis approach by geoprocessing two sets of locations to create a distance statistics table to gauge how proximate Housing Choice Voucher (HCV) program participant locations are to their nearest PC. Assigning residential locations to the closest, or proximate, PC allows for deriving a set of summary statistics on the distance between individuals and the likely places they frequent for resources and services. The assumption rests on a complex interaction between self-selection and economic realities that occurs in residential selection, wherein people often cluster together residentially due on similarities in social, demographic, and economic characteristics (Tittle and Rotolo, 2010). HCV recipients, as an example, typically cluster where housing is affordable.

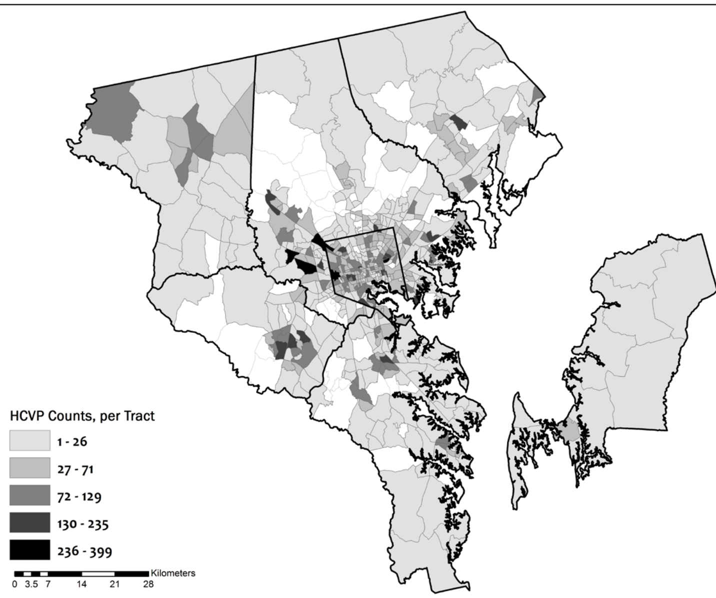
Data

The data we used in this example are the counts of 2016 HCV program participants by census tract in the Baltimore-Columbia-Towson, MD Core Based Statistical Area (CBSA)—Baltimore CBSA hereafter. We downloaded the data from the U.S. Department Housing and Urban Development (HUD) enterprise GIS storefront. Exhibit 1 shows the geographic distribution of program participants across the CBSA.

Because HUD does not provide program participant locations, we simulated the locations in a two-stage process to create residential locations based on where people actually live. First, we proportionally divided the counts of program participants according to the proportion of residents within each block group contained in each census tract. Second, we created a set of randomly distributed locations within each block group to simulate the locations based on known residential patterns. This two-stage process allows for a reasonable approximation of where HCV program participants live and reduces the risk of placing them in areas where populations do not reside (for example, forested portions, lakes, parks areas, and industrial sites).

Exhibit 1

Thematic Map of Housing Choice Voucher Program Participant Counts in the Census Tracts



HCVP = Housing Choice Voucher participant.

Proximity Analysis

A proximity analysis of this type rests on the geographic theory of spatial interaction, in which individuals interact with places by traveling to nearby locations to obtain amenities, services, or resources. With HCV program participants, the interaction results from traveling to the closest PC where daily needs resources (for example, jobs, groceries, goods, and doctors) are concentrated. Interaction for outreach involves centralizing resources to deliver in a PC and venturing into surrounding neighborhoods to reach the intended constituents. In each of these cases, the “least cost principle” is applied to the geographic theory of spatial interaction to assume that people—particularly those with limited abilities to travel—will travel to the closest place for any needs.

Method

We use ArcGIS software to develop this method, but any GIS software program can be used to replicate our method. The GIS software programs available contain the tools required to prepare the data, calculate proximity distances, and visualize the results. The method demonstrated in this article follows a series of six stages to create a table of summary distance statistics between HCV program participant residences and their closest PCs, and also output for visualizing the distances. We provide a script in the appendix to automate the following method.

Stage 1: Identifying Initial Population Centers

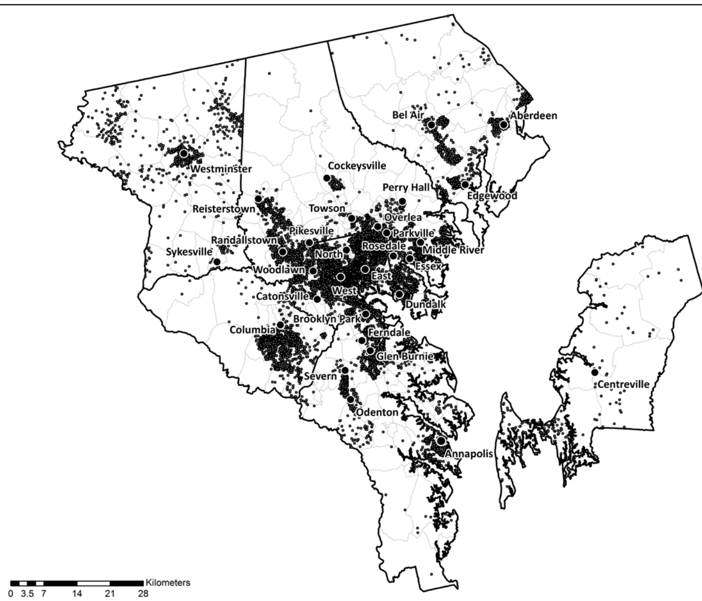
In the first stage, we selected a set of locations (point geometries) to represent the PCs that the HCV participant residence locations will be assigned. Our method gives analysts the quantitative and qualitative flexibility to identify PCs that serve as resource centers. The initial PCs can be selected objectively based on a population threshold, for example, that will ensure that PCs likely have a concentration of amenities, services, or resources. Alternatively, the initial selection of PCs can be determined subjectively by choosing sites that have importance or fill coverage gaps between places.

Any given coverage area, such as the Baltimore CBSA, likely has too many PCs for residential assignment. The initial centers should be places with a certain number of residents but recognizable by name to residents. Selecting too many PCs leads to an oversaturation with too many centers too close to each other, creating a large summary table with too few distances to each PC. Too large a table makes the statistics unreliable and unreasonable to communicate the likely centers to which residences travel for resources. Selecting too few PCs would aggregate residential locations to PCs that are unlikely to be places to which residences travel for resources. Both situations provide unrealistic pictures of people and the PCs with which they interact. The selection of PCs, therefore, must be well thought out.

Our first selection reduced the 72 PCs in the Baltimore CBSA to 30 by removing nearly all centers with fewer than 10,000 residents, making our initial analysis primarily quantitatively selective. We subjectively included additional PCs to cover large gaps between centers (see exhibit 2).

Exhibit 2

Initial Population Centers With Randomized Housing Choice Voucher Program Participant Locations



For example, we kept Cockeysville (northern Baltimore County) and Centreville (Eastern Shore in Queen Anne's County) to reduce gaps in coverage, even though their populations are less than 10,000. Additionally, because the city of Baltimore is a single PC in the geometric center of the city jurisdiction boundary, we added three PCs (North, East, and West Baltimore) within the administrative sections of the city. The initial selection of 30 PCs provided a comprehensive coverage across the Baltimore CBSA in relation to residential distributions of HCV program participants.

Stage 2: Assigning HCV Recipient Locations to the Nearest Population Centers

Before conducting a proximity analysis, we updated the HCV data to contain X and Y coordinates, to create and draw proximity links between HCV recipient locations and the nearest PCs. These proximity links are known as *spider diagrams*, which aid in the visualization of the geographic distribution of locations around a central location.

We identified the nearest PC to each participant location based on the direct (Euclidean) distance using the Near technique in the ArcToolbox. This technique identifies the closest PC to each participant location and records the following three geographic characteristics of the nearest PC to each participant location in the resident location attribute table.

- *Unique identifier* of the nearest PC location.
- *Distance* from the HCV program participant location to the nearest PC.
- *X and Y coordinates* of the nearest PC location.

These three characteristics allow for a summarization of the linkage distances into a set of summary statistics of residences and the closest PCs. They also provide the data that creates the spider diagrams for visualization.

Stage 3: Calculating the Proximity Link Between Locations

In this stage, we calculated the distances between HCV recipient locations and PCs and created proximity links for the spider diagram. Using the X and Y coordinates of the closest PC recorded for each program participant location, we produced a proximity layer of lines between the residences and PCs showing those linkages.

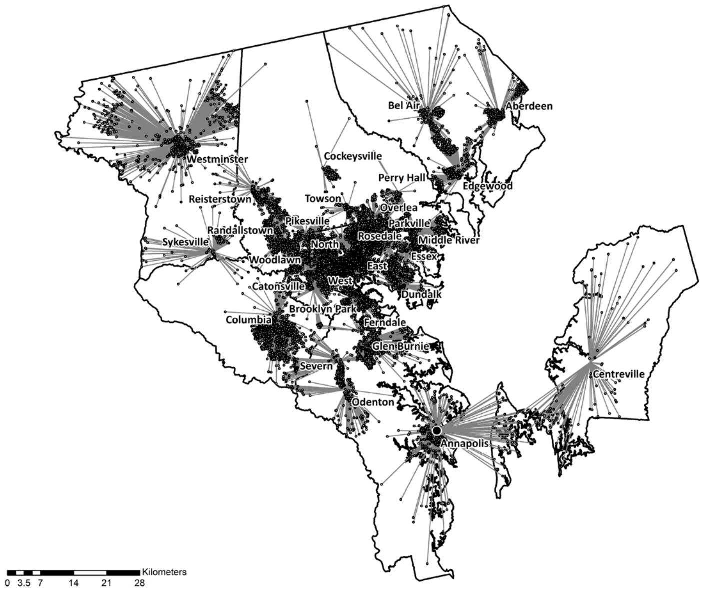
We first had to generate an additional set of X and Y coordinates for the participant locations to create the proximity links. Creating a layer of proximity links requires start and end coordinates that allow for the lines to be drawn between locations. With the Add Geometry Attributes tool in the ArcToolbox, we added the centroid coordinates for the HCV resident locations. Doing so gave the program participant layer the coordinate pairs needed to create proximity link lines between the residences and the population centers.

Using the XY to Line technique in the ArcToolbox, we created a new geographic layer of lines between participant locations and PCs and loaded the links into ArcGIS to visualize (see exhibit 3). The links allow for a visual depiction of the proximity assignment to ensure that coverage is adequate.

The PCs can now be visually inspected in conjunction with the distance summary statistics table to begin evaluating whether the results are reasonable to meet the objectives of the analysis.

Exhibit 3

Initial Population Center Selection Links Between Housing Choice Voucher Program Participants and Nearest Population Centers



Stage 4: Summarizing the HCV Counts by Population Centers

In this stage, we summarized the distances between the HCV recipient locations and PCs into a table of proximity distance statistics. We used these statistics to evaluate the proximity assignment results and determine if PCs should be removed from, or added to the analysis.

We summarized the distances associated by the unique identifier for each PC, recorded from the near analysis of each resident location in stage 2. We then used the Summary Statistics tool in the ArcToolbox to create several statistical point estimates for each PC and specified Count, Sum, Mean, Standard Deviation, Minimum, and Maximum to be calculated using the NEAR_FID unique identifier as the Case field. The Case field enables ArcGIS to summarize distances by each individual PC and create a table of summary statistics.

Stage 5: Assessing Proximity Results by Population Center

At the final stage, we imported the summary distances into an Excel spreadsheet to create a distance statistics table (exhibit 4) corresponding with the thematic map in exhibit 3 to visualize the frequencies of each PC. This visualization allows for an examination of the distance summary statistics. In this stage, we generated a set of links that show which HCV recipient locations are proximately associated with the PCs and inspected the attribute table of the PCs layer to ensure that we achieved a reasonable set of statistics.

Exhibit 4

Distance Statistics (in Kilometers) of Initial Housing Choice Voucher Program Participation Assignment to Nearest Population Centers

Pop Center	Count	Sum	Mean	Std. Dev.	Minium	Maximum	Coeff. Var.
East Baltimore	4,002	8,452	2.1	0.9	0.1	4.7	0.43
North Baltimore	1,437	4,099	2.9	0.7	0.7	4.5	0.24
West Baltimore	3,514	7,590	2.2	0.9	0.0	5.0	0.44
Aberdeen	322	1,376	4.3	3.5	0.1	16.6	0.81
Annapolis	407	2,024	5.0	5.0	0.1	28.2	1.01
Bel Air	366	1,378	3.8	3.8	0.1	25.8	1.02
Brooklyn Park	845	2,177	2.6	1.1	0.2	5.6	0.44
Catonsville	458	1,432	3.1	1.4	0.2	8.2	0.44
Centreville	92	1,275	13.9	5.9	0.8	28.3	0.43
Cockeysville	96	232	2.4	2.0	0.0	14.6	0.81
Columbia	1,329	6,654	5.0	2.2	0.0	14.6	0.43
Dundalk	780	1,618	2.1	1.2	0.1	8.7	0.58
Edgewood	438	1,457	3.3	2.0	0.0	8.0	0.60
Essex	430	1,008	2.3	0.8	0.3	6.7	0.33
Ferndale	154	352	2.3	1.0	0.2	8.0	0.44
Glen Burnie	755	2,707	3.6	2.0	0.2	14.0	0.55
Middle River	624	1,409	2.3	1.4	0.2	7.3	0.61
Odenton	205	839	4.1	3.3	0.2	12.5	0.80
Overlea	412	981	2.4	1.0	0.2	4.6	0.41
Parkville	773	2,019	2.6	0.7	0.1	4.8	0.25
Perry Hall	168	409	2.4	1.6	0.2	9.8	0.64
Pikesville	1,090	2,804	2.6	0.9	0.4	6.5	0.36
Randallstown	746	1,732	2.3	1.6	0.2	7.5	0.68
Reisterstown	360	1,306	3.6	2.1	0.2	9.4	0.58
Rosedale	753	2,051	2.7	0.8	0.1	4.5	0.31
Severn	227	915	4.0	3.4	0.2	12.9	0.85
Sykesville	85	460	5.4	4.7	0.0	15.5	0.87
Towson	227	388	1.7	1.3	0.1	5.8	0.77
Westminster	708	5,194	7.3	6.3	0.1	25.2	0.86
Woodlawn	1,278	3,324	2.6	1.0	0.1	6.4	0.38

We added the coefficient of variation (CV) to the table to compare the spread of HCV program participant locations around the average from the PCs. The CV is particularly important in spatial analysis when comparing groups of locations or distances around a central point, because less space exists in which the locations closer to the center can be distributed. That is, locations toward the center are likely to be closer together and have naturally smaller distances, because less space exists for the locations to be distributed–, making their variation from the mean appear less varied. The converse is true for locations farthest aware from the center; that is, they appear more dispersed and have larger distances. Thus, using the standard deviation distances alone to compare the summarized distances around each PC location is untenable, because the differences in space restrictions will make PCs with HCV participants closer to the PC to appear less dispersed than those farther away. Therefore, to determine if HCV residences are more dispersed around the different PCs, the distances around the average distance for each group are contextualized by dividing the standard deviation of the distances, the mean distance to create a CV of distances to make comparisons.

The results of our first analysis show that using 30 PCs created a large distance analysis table, including a number of PCs with few HCV program participants. Centreville, Cockeysville, Ferndale, Odenton, Perry Hall, Severn, Sykesville, and Towson are centers with a relatively few HCV program participants in proximity. Reducing the size of the table is not a simple matter of removing these small number of residences. First, removing each location will impact the counts of nearby PCs, because the program participants associated with one of the removed centers will get assigned to the next nearest PC. Second, removing a center because it does not meet a certain threshold may lead an unreasonable distance to be considered proximate. For example, removing Centreville would show many more HCV program participants on the Eastern Shore traveling to Annapolis for services, which is an unreasonable and unlikely assumption. As such, we subjectively determined whether to keep or discard a PC for the next analysis.

Stage 6: Adjustment Recalculations To Refine the Results

To refine the analysis, we removed five of the eight PCs with fewer than 300 associated HCV resident locations—Cockeysville, Ferndale, Odenton, Perry Hall, and Sykesville. We eliminated two additional PCs, Essex and Rosedale, because they are in close proximity to each other and other centers. We removed Sykesville in the west and Cockeysville in the north; although doing so left gaps in the coverage area, the HCV program participant locations were close enough to be assigned to a nearby PC on a regional scale.

Most participants near Cockeysville were reassigned to Towson, thus increasing Towson's numbers. Towson is reasonable reassignment choice because amenities, services, and other resources are more clustered in that center than in Cockeysville. We also removed Parkville and Overlea because of their proximity to Towson, which has greater name recognition and is more meaningful to residents. We replaced North Baltimore with Northeast Baltimore to provide coverage in the city's northeast. Few to no HCV program participant locations actually lived close to the North Baltimore location, even though it was the most proximate PC for about 1,400 participants. Further, given the economic status of North Baltimore, resources for HCV participants are not in that section of the city. These two reasons made North Baltimore less than an ideal choice to represent a central location for participants to travel to for resources.

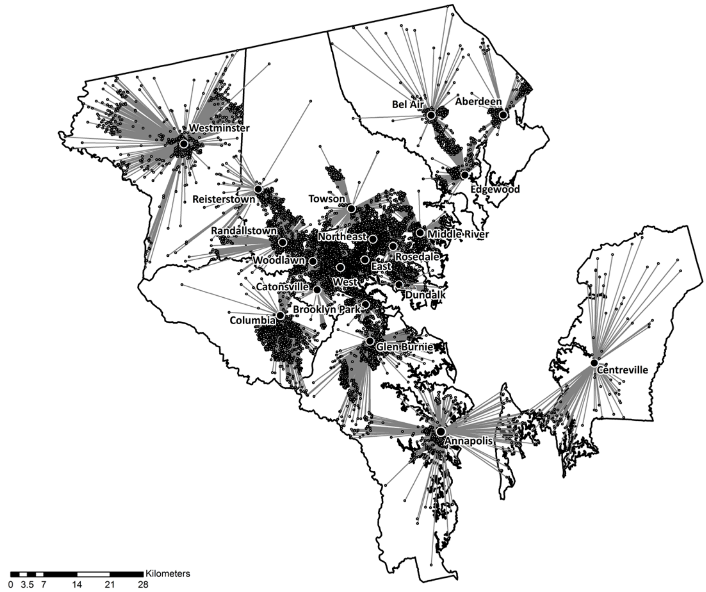
We removed Ferndale and Odenton, because they are in close proximity to Glen Burnie and Odenton to Annapolis, respectively. HCV participant locations around Parkville and Perry Hall, slightly northeast of the city of Baltimore, were reassigned to several PCs in the vicinity, which also increased the count of HCV program participants around Towson (see exhibit 5).

The removal of the aforementioned five centers left Centreville, Severn, and Towson remaining in the second analysis.

In the second analysis, all PCs had resident counts at least above a reasonable threshold to create estimates that represent large enough groups, with no two PCs too close to each other. We repeated stages 2 to 4 and reexamined the results, as in stage 5.

Exhibit 5

Final Population Center Selection Links Between Housing Choice Voucher Program Participants and Nearest Population Centers



Graphing the Final Proximity Analysis Results

Once we summarized the voucher holder distances, we assessed the resulting table to determine if we needed to adjust the number and inclusion of PCs. We imported the summary table into a Microsoft Excel spreadsheet with the distance statistics and calculated the corresponding CVs to evaluate the results (exhibit 6).

Exhibit 6 offers several statistics for reporting on approximately how far HCV program participants travel for services or other resources. These statistics provide the number of participants in proximity to each PC, how far on average they would have to travel to a particular PC, and how dispersed they are within the area. The minimum and maximum figures give the distance ranges of the closest and farthest HCV program participant residences from a PC. Finally, we calculated the CV so that the dispersion of the distances for each PC can be compared.

Exhibits 7 and 8 show graphs of the average distances and CVs, respectively, to visually gauge the similarities and differences for each PC.

The summary table displayed in exhibit 6 can also be joined to the PC layer to thematically map HCV program participant volumes of those proximate to that PC or any of the summarized distance statistics. The final map to accompany exhibit 6 contains only the PCs with corresponding HCV program participant counts (exhibit 9).

Exhibit 6

Distance Statistics (in Kilometers) of Final Housing Choice Voucher Program Participation Assignment to Nearest Population Centers

Pop Center	Count	Sum	Mean	Std. Dev.	Minium	Maximum	Coeff. Var.
East Baltimore	4,005	6,587	1.9	0.9	0.1	4.7	0.44
Northeast Baltimore	1,611	6,612	2.8	1.1	0.1	8.3	0.39
West Baltimore	3,514	9,861	2.5	1.3	0.0	6.5	0.51
Aberdeen	322	1,376	4.3	3.5	0.1	16.6	0.81
Annapolis	439	2,539	5.8	5.6	0.1	28.2	0.98
Bel Air	367	1,394	3.8	3.9	0.1	25.8	1.02
Brooklyn Park	926	2,464	2.7	1.1	0.2	5.7	0.42
Catonsville	461	1,456	3.2	1.4	0.2	8.3	0.45
Centreville	92	1,275	13.9	5.9	0.8	28.3	0.43
Columbia	1,400	7,527	5.4	2.7	0.0	21.8	0.51
Dundalk	805	1,732	2.2	1.3	0.1	8.7	0.59
Edgewood	439	1,466	3.3	2.0	0.0	8.9	0.60
Glen Burnie	1,157	6,149	5.3	3.6	0.2	19.4	0.67
Middle River	1,142	3,988	3.5	2.1	0.2	9.7	0.60
Randallstown	961	3,174	3.3	3.0	0.2	20.9	0.91
Reistertown	376	1,540	4.1	3.2	0.2	23.7	0.77
Rosedale	1,263	1,712	2.8	1.1	0.1	6.0	0.38
Towson	986	2,869	4.0	2.4	0.1	14.7	0.61
Westminster	722	5,516	7.6	6.6	0.1	27.3	0.87
Woodlawn	2,093	8,284	3.5	1.3	0.1	7.0	0.37

Exhibit 7

Average Distances of Housing Choice Voucher Program Participant Residences to Nearest Population Centers

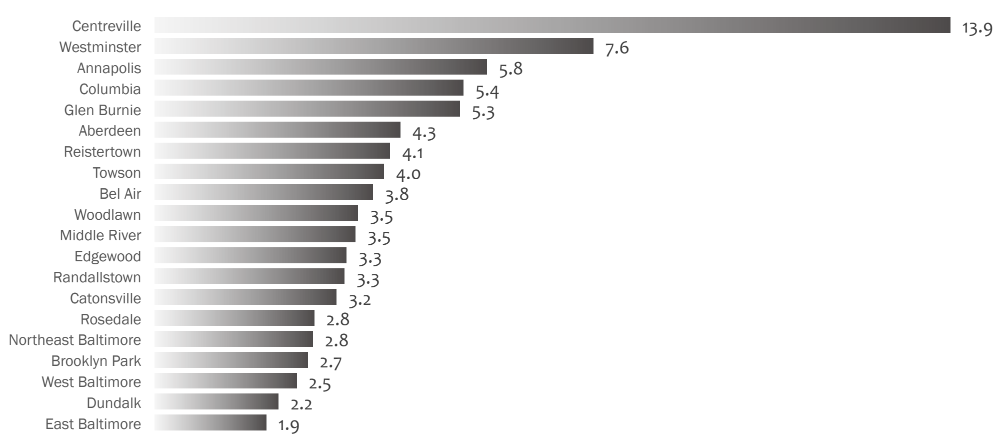


Exhibit 8

Distance Coefficients of Variation of Housing Choice Voucher Program Participant Residences to Nearest Population Centers

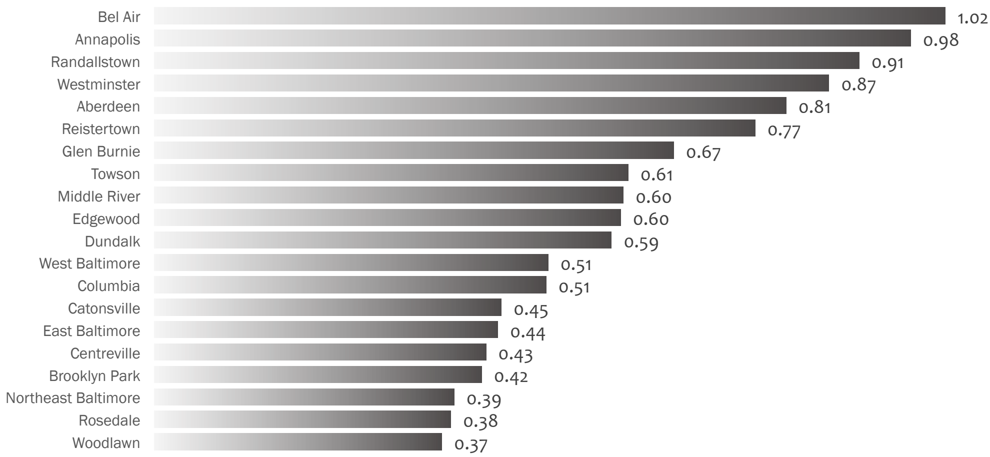
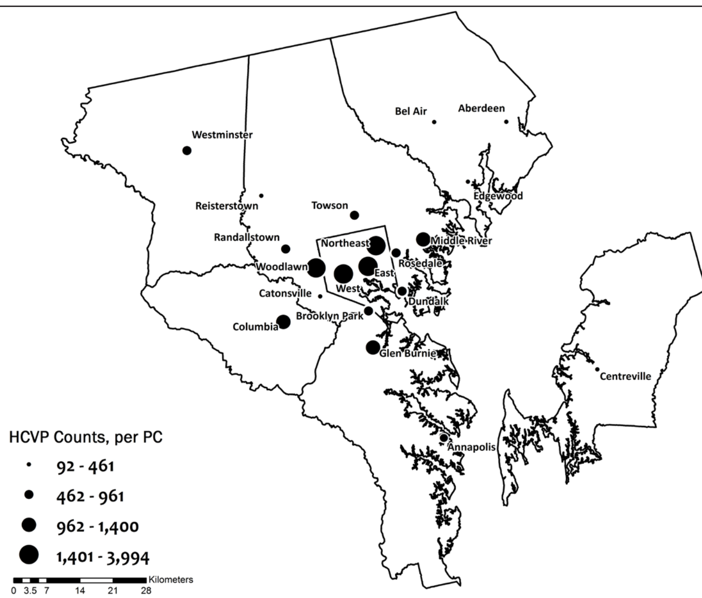


Exhibit 9

Final Population Center Selection With Counts of Housing Choice Voucher Program Participants



HCVP = Housing Choice Voucher participant. PC = population center.

Summary

The proximity method is an approach that enables analysts, researchers, and others to transform location data into a set of a descriptive statistics for stakeholders, planners, and other officials to make informed decisions based on how proximate residents are to resources or how far outreach must be to reach specific populations. Information compiled as a set of statistics allows for that information to be better understood than by the visual display of numerous data points or color-shaded areas that show concentrations of locations. The use of GIS enables analysts to implement this technique quickly, which gets information to users faster than before. If too many locations are assigned to a PC, many of those locations are likely not really proximate, which undermines the objective of providing local-level counts so that outreach or services can be centered in places that serve the maximum population and minimize travel distance. The advantage of the method presented in this article is that the analyst has control over data selection to make the results as generalized as needed but more accurate that other methods.

Appendix: Python Script of Proximity

```
import arcpy
## set the workspace environment
arcpy.env.workspace      = "C:\my\workspace\and\geodatabase.gdb"
arcpy.env.overwriteOutput = True
## set variables:
## these variables you may need to change as the analysis changes
points     = “randomized_voucher_locations” ## HCV program locations
popCenters = “population_centers”           ## population centers
nearTable  = “Near_Table”                   ## the near distances table
## these variables always stay the same, no change required
## the below is the summarized table of the near table and the accompanying statistics
myFields = [
                         [ “NEAR_DIST”, ”COUNT”],
                         [ “NEAR_DIST”, ”MIN”  ],
                         [ “NEAR_DIST”, ”MAX”  ],
                         [ “NEAR_DIST”, ”MEAN” ],
                         [ “NEAR_DIST”, ”STD”  ]
        ]
## sumTable is the summary statistics of the near distances table
sumTable = "%s_SUMMARY" %(nearTable)
## generate near table
arcpy.GenerateNearTable_analysis(points,popCenters,nearTable,””,”LOCATION”,””,”CLOSEST”)
## summarize the near table
arcpy.Statistics_analysis(nearTable,sumTable,myFields,”NEAR_FID”)
## add coefficient of variation and name fields
arcpy.AddField_management(sumTable,”NEAR_DIST_CV”,”DOUBLE”)
arcpy.AddField_management(sumTable,”NAME”,”TEXT”)
## join in the population centers to get their names, it makes reading the table easier
arcpy.AddJoin_management(sumTable,”NEAR_FID”,popCenters,”OBJECTID”)
## calculate coefficient of variation and bring in the name of the population center
arcpy.CalculateField_management(sumTable,”NEAR_DIST_CV”, “!STD_NEAR_DIST! / !MEAN_NEAR_DIST!”,
”PYTHON_9.3”)
expression   = “!%s.NAME!” %(popCenters)
sumTableName = “%s.NAME”   %(sumTable)
arcpy.CalculateField_management(sumTable,sumTableName,expression,”PYTHON”)
## remove the join
arcpy.RemoveJoin_management(sumTable,popCenters)
```

Acknowledgments

The authors thank Katherine Condon of the U.S. Citizenship and Immigration Services for motivating us to develop and publish this method and for providing valuable comments toward improving the manuscript. The authors also thank several others who motivated us to publish this work.

Authors

Ron Wilson is an adjunct faculty member of the Geographic Information Systems Program at the University of Maryland, Baltimore County.

Alexander Din is a geographic information system analyst at Dewberry.

References

Tittle, Charles R., and Thomas Rotolo. 2010. “Socio-Demographic Homogenizing Trends Within Fixed-Boundary Spatial Areas in the United States,” *Social Science Research* 39 (2): 324–340.