Final Project:
An Income and Education Study of Washington D.C.

Barbara J. McKay Archibald

Geography 586

Dr. Justine Blanford

12 December, 2009
As the seat of the United States Federal Government, Washington D.C. is a critical city within the U.S. Washington D.C. is home to a diverse and dynamic population. It has a resident population including civil servants and government contractors as well as an affluent transient population consisting of politicians, lobbyists, and others. Unfortunately, Washington D.C. is also home to an extensive poor population (Phillips & Beasley, 2005). Given this diversity, there are vast disparities in income. According to sampled data from the 2000 Census 12.6% of families earn less than $10,000 annually while 7% of families earn greater than $200,000 (U.S. Census Bureau, 2000a). (See Table 1.) Of the 115,963 families sampled 19,365 or 16.7% of them live below the poverty line (U.S. Census Bureau, 2000c). This is far greater than the national average: 9.2% (U.S. Census Bureau, 2000b). Disparity in income may translate to inequality in access to resources such as education, medical care, and transportation as well as exposure to social strife which may reflect in crime rates. With these consequences in mind, this study will scratch the surface of family economics in Washington D.C. in an attempt to supply a foundation for addressing issues related to poverty. This will examine the population diversity as it relates to income and education. Income determines access and power of families while education may be a factor that influences income.

The scope of the study is restricted to Washington D.C. rather than the entire metro area. The metro area may be useful since the U.S. Government, contractors, and lobbyists operate throughout the region, and their employees often live outside of the city. But, keeping the scope of this research at the scale of one government will keep some unknown variables out of the study. Working with the metro area introduces variation caused by other municipalities, as well as other states – Maryland and Virginia. In addition, the data set will be more manageable in the timeframe of the project.

Several decisions must be made in order to examine the income and education of Washington D.C. residents. How will this study look at these variables? Will the study work with individual data or aggregate data? What constitutes the economic unit of study? How will education data be aggregated? The data available and a desire for detailed data answer these questions. The U.S. Census Bureau is the primary source of demographic data in the U.S. Privacy for the individual is critical to the government’s ability to collect personal information; therefore individual level data are not available. Aggregate data are available. The data are in units as small as the block group. Given this research seeks to find the distribution of wealth and education in the city, fine detail is desirable. The census block group is used
Census data have several options for income: household, family and individual. This study uses the family economic unit since a social aggregate may better describe a

<table>
<thead>
<tr>
<th>Subject</th>
<th>Families: Numbers</th>
<th></th>
<th>Families: Percent</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Married-couple families</td>
<td>Female householder, no husband present</td>
<td>Total</td>
<td>Married-couple families</td>
</tr>
<tr>
<td></td>
<td>115,963</td>
<td>58,050</td>
<td>47,558</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Less than $10,000</td>
<td>14,559</td>
<td>2,039</td>
<td>11,129</td>
<td>12.6</td>
<td>3.5</td>
</tr>
<tr>
<td>$10,000 to $14,999</td>
<td>5,829</td>
<td>1,393</td>
<td>3,901</td>
<td>5.0</td>
<td>2.4</td>
</tr>
<tr>
<td>$15,000 to $19,999</td>
<td>6,162</td>
<td>1,770</td>
<td>3,632</td>
<td>5.3</td>
<td>3.0</td>
</tr>
<tr>
<td>$20,000 to $24,999</td>
<td>6,608</td>
<td>1,820</td>
<td>3,881</td>
<td>5.7</td>
<td>3.1</td>
</tr>
<tr>
<td>$25,000 to $29,999</td>
<td>6,257</td>
<td>1,931</td>
<td>3,503</td>
<td>5.4</td>
<td>3.3</td>
</tr>
<tr>
<td>$30,000 to $34,999</td>
<td>6,264</td>
<td>2,341</td>
<td>3,141</td>
<td>5.4</td>
<td>4.0</td>
</tr>
<tr>
<td>$35,000 to $39,999</td>
<td>5,834</td>
<td>2,292</td>
<td>2,869</td>
<td>5.0</td>
<td>3.9</td>
</tr>
<tr>
<td>$40,000 to $44,999</td>
<td>5,310</td>
<td>2,438</td>
<td>2,299</td>
<td>4.6</td>
<td>4.2</td>
</tr>
<tr>
<td>$45,000 to $49,999</td>
<td>4,298</td>
<td>2,022</td>
<td>1,789</td>
<td>3.7</td>
<td>3.5</td>
</tr>
<tr>
<td>$50,000 to $59,999</td>
<td>8,206</td>
<td>4,505</td>
<td>3,034</td>
<td>7.1</td>
<td>7.8</td>
</tr>
<tr>
<td>$60,000 to $74,999</td>
<td>9,584</td>
<td>5,633</td>
<td>3,016</td>
<td>8.3</td>
<td>9.7</td>
</tr>
<tr>
<td>$75,000 to $99,999</td>
<td>11,792</td>
<td>7,959</td>
<td>2,897</td>
<td>10.2</td>
<td>13.7</td>
</tr>
<tr>
<td>$100,000 to $124,999</td>
<td>7,508</td>
<td>5,773</td>
<td>1,250</td>
<td>6.5</td>
<td>9.9</td>
</tr>
<tr>
<td>$125,000 to $149,999</td>
<td>4,468</td>
<td>3,856</td>
<td>455</td>
<td>3.9</td>
<td>6.6</td>
</tr>
<tr>
<td>$150,000 to $199,999</td>
<td>5,209</td>
<td>4,799</td>
<td>309</td>
<td>4.5</td>
<td>8.3</td>
</tr>
<tr>
<td>$200,000 or</td>
<td>8,075</td>
<td>7,479</td>
<td>453</td>
<td>7.0</td>
<td>12.9</td>
</tr>
</tbody>
</table>
Median income (dollars) | 46,283 | 77,015 | 26,499
Mean income (dollars) | 78,192 | 117,734 | 37,038


Table 1. Sample income data for families in Washington D.C. in dollars and percent (Modified from U. S. Census Bureau 2000a).

<table>
<thead>
<tr>
<th>Poverty Status of Families</th>
<th>Number</th>
<th>Percent below poverty level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All income levels</td>
<td>Below poverty level</td>
</tr>
<tr>
<td>Washington D.C.</td>
<td>115,963</td>
<td>19,365</td>
</tr>
<tr>
<td>United States</td>
<td>72,261,780</td>
<td>6,620,945</td>
</tr>
</tbody>
</table>

Source: U.S. Census Bureau, Census 2000 Summary File 3, Matrices P90, P91, P92, P93, PCT59, PCT60, and PCT61.

Table 2. Sample poverty data for families in Washington D.C. and the United States in numbers of families and percent (Modified from U. S. Census Bureau 2000b, c).

person’s economic standing. The U. S. Census Bureau has education data available broken down by age and sex and aggregated with several other variables. This study compiles the sexes into one class of education and uses the education breakdown original to the Census Bureau data.
Now that the data are identified, the goals of the study beyond the general notion of income and education must be identified. The study’s focus is to understand the distribution of Washington D.C.’s population based on income and look at its distribution based on education achieved. With this information the study looks to see if the distributions of income and education co-occur. To do this the data sets are analyzed for clustering behavior. If clusters are identified, overlay analysis is conducted with the cluster layers to identify if they co-occur and to what extent they co-occur.

Summary of Questions:

- Is income distributed evenly across space?
- Does income cluster and where?
- Is education distributed evenly across space?
- Do people cluster by education levels?
- Does income distribution correspond with education?
Several approaches for determining distribution are available. Mean nearest neighbor distribution calculates the average of the distances between an event and its closest neighbors using the Pythagorean Theorem (O’Sullivan and Unwin, 2003). The nearest neighbor ratio’s relationship to 1 determines clustering or dispersion. Z-score tests significance of the null hypothesis, $H_0$. P-value is a probability of falsely rejecting the null hypothesis (ESRI, 2009a,b). Census and sample quadrat counts measure densities of values by either establishing a grid over the study area or creating a random sample of the study area using a quadrat (sample unit) of fixed size. The Variance Mean Ratio’s relationship to 1 indicates clustering or dispersion. $X^2$ indicates the acceptance or rejection of the $H_0$. Finally, Kernel Density Analysis creates a density surface using the density of events in a defined circle using a carefully selected radius. A density map is constructed using this result to show density patterns of events in a study area (O’Sullivan and Unwin, 2003).

These methods could be used to evaluate the presence or absence of clustering of income or education level and to map the values of income and education across space. But, there are a few problems with these tools. First, these tools are meant to be used on point data. Point data can be constructed from the census block groups. But, census block group size varies, and distances between points vary. Distance is the critical measure common to mean nearest neighbor distribution, quadrat count analysis and Kernel density analysis. Therefore, the varying sizes would interfere with the evaluation of income and education distribution across space. Second, although data is determined clustered or not, there is no definition of the clusters.

The approach to determine distribution in this project is a measure of spatial autocorrelation. Autocorrelation is a concept where a spatial data set “is likely to have characteristic distances or lengths at which it is correlated with itself” (O’Sullivan and Unwin 2003, p.108). Otherwise, data sets within the characteristic distances will be similar.” Determining spatial autocorrelation will identify grouped income and education data.

Moran’s $I$ is the spatial autocorrelation applied in this project. It translates non-spatial correlation to spatial correlation (O’Sullivan, 2009). It is designed to work with area and linear data (ESRI 2009c). Its general formula considers areal adjacency and the weight can consider adjacency of units rather than a variation on distance. And, Moran’s $I$ works with non-binary data. It calculates $I$ as follows:

$$I = \frac{n}{\text{variance}} \left[ \frac{\text{Weight(covariance)}}{\text{number of joins}} \right]$$
The covariance evaluates the data. The weight modifies the data. The number of joins and n/Variance normalize in terms of the number of zones, the number of adjacencies, and the range of y. A positive I is a positive autocorrelation. A negative is negative or inverse correlation (O’Sullivan and Unwin, 2003).

Local Indicators of Spatial Association (LISA) are applied in this project. LISA indicate where the clusters occur (O’Sullivan, 2009). Clusters are determined by evaluating the Moran’s I for each area unit against its probability value. If the p-value is appropriate, the areal unit experiences autocorrelation and is displayed on the map. The clusters are revealed (Anselin, 1993, ESRI 2009d).

Method

Now that the major decisions regarding the processing of Washington D.C.’s economic and educational data have been discussed, the data are ready to be processed. The raw data are: BlockGroupPly.zip, P77. Median Family Income in 1999 (dollars), and P37. Sex by Educational Attainment for the Population 25+ Years. BlockGroupPly.zip is the shape file. It is projected to NAD_1983_StatePlane_Maryland_FIPS_1900 and is ready for use. The remaining data are text files and require processing. The data files for Median Family Income in 1999 (dollars) and P37. Sex by Educational Attainment for the Population 25+ Years are imported into Excel. This triggers the Text Import Wizard. In the wizard select Delimited as the file type. Then select Tab, Comma and Other as the Delimiters. Indicate “|” as Other. Finally keep Column Data as General, and click finish. This produces an Excel file which only requires small modifications, primarily field expansion, to be readable and in a desirable format.

At this point P77. Median Family Income in 1999 (dollars) table is imported into ArcMap. It is joined to the shape file using the shape file’s 1500USFIPS field and the table’s GEO_ID field. Some fields are removed and the file shape file is exported and added to the map.

This shape file “BGPLY_E06” is imported into GeoDa (Anselin, 1993). A weights file is created using the Tools>Weights>Create. BGPLY_E06 is the input file. BGPLY_E06_WT is the output file. Geo_ID2 is the ID variable. And the Contiguity Weight is Queen Contiguity (all adjacent polygons). Click Create to create the .gal file. Then run Univariate LISA analysis. (Space>Univarite Lisa). Select PO77001 (income field) in the Variable Settings box; click OK. Select the newly created .gal file in the Select Weights box; click OK. Select Significance Map, Cluster Map and Moran Scatter Plot from the “What window to open?” Box; click OK. A map showing the p-value for each block group and the LISA map showing the
clusters of block groups are produced. The LISA map indicates low incomes and high incomes which cluster and low-high and high-low income relationships which show negative autocorrelation. These results use a sample of 99 and a significance value of 0.05. Then the results at the 0.05 significance level are saved by right clicking the cluster map and selecting “Save Results”. Finally the analysis is made permanent by saving the shape file (Table>Save to Shape File As...).

Figure 3. Image of the Creating Weights menu from GeoDa.

The shape file is then imported back into ArcMap. A map is created indicating block groups with high income clustering, low income clustering, low-high income negative autocorrelation, high-low income negative autocorrelation and insignificant values. This is the map that is used in the education – income map overlay.

P37. Sex by Educational Attainment for the Population 25+ Years Excel file requires extensive processing before it is imported into ArcMap. The file has 32 education classes. The classes are no schooling, nursery school to 4th grade, 5th to 8th grade, 9th grade, 10th grade, 11th grade, 12th grade, high school
graduate, some college less than 1 year, some college greater than 1 year, associate’s degree, bachelor’s degree, master’s degree, professional school degree and doctorate degree for male and female populations. The male and female classes are combined into 16 classes. Then education data are graphed for each of the 433 block groups, and the median and modes are determined. The mode and the median are then converted to an ordinal numeric scale compatible with Moran’s I. Finally the table is reduced to critical fields.

<table>
<thead>
<tr>
<th>No.</th>
<th>Class Description</th>
<th>No.</th>
<th>Class Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No Schooling</td>
<td>9</td>
<td>High school graduate</td>
</tr>
<tr>
<td>2</td>
<td>Nursery to 4th grade</td>
<td>10</td>
<td>Some college, less than 1 year</td>
</tr>
<tr>
<td>3</td>
<td>5th to 6th grade</td>
<td>11</td>
<td>Some college, greater than 1 year</td>
</tr>
<tr>
<td>4</td>
<td>7th to 8th grade</td>
<td>12</td>
<td>Associate’s degree</td>
</tr>
<tr>
<td>5</td>
<td>9th grade</td>
<td>13</td>
<td>Bachelor’s degree</td>
</tr>
<tr>
<td>6</td>
<td>10th grade</td>
<td>14</td>
<td>Master’s Degree</td>
</tr>
<tr>
<td>7</td>
<td>11th grade</td>
<td>15</td>
<td>Professional Degree</td>
</tr>
<tr>
<td>8</td>
<td>12th grade</td>
<td>16</td>
<td>Doctorate</td>
</tr>
</tbody>
</table>

Table 3. Numeric and descriptive classes for the median and modal education data.

P37. Sex by Educational Attainment for the Population 25+ Years table is imported into ArcMap. It is joined to the shape file using the shape file’s 1500USFIPS field and the table’s Geography Identifier field. The shape file is exported and added to the map.

This shape file “BGPLYAbrevVE” is imported into GeoDa (Anselin, 1993). A weights file is created using the Tools>Weights>Create. The input file, output file, and ID variable are designated. The Contiguity Weight is Queen Contiguity. Click Create to create the .gal file. Then run Univariate LISA analysis (Space>Univariate LISA). Select “NumericMed” in the Variable Settings box; click OK. Select the .gal file in the Select Weights box; click OK. Select Significance Map, Cluster Map and Moran Scatter Plot from the “What window to open?” Box; click OK. A map showing the p-value for each block group and the LISA map showing the clusters of block groups are produced. The LISA map indicates high autocorrelation and low autocorrelation of education and low-high and high-low education relationships which show negative autocorrelation. These results use a sample of 99 and a significance value of 0.05. The 0.05 significance level results are saved by right clicking the cluster map and selecting
“Save Results”. Finally the analysis is made permanent by saving the shape file (Table > Save to Shape File As...).

The shape file is then imported back into ArcMap. A map is created indicating block groups with high education clustering, low education clustering, low-high education negative autocorrelation, high-low education negative autocorrelation and insignificant values using median data. This is the map that is used in the education – income map overlay.

At last, Univariate LISA analysis is run on the “NumericMod” variable. The .gal file created for the median data is reused with the modal data LISA. The same significance map, cluster map and Moran’s I scatter plot are created. The LISA map shows the high and low education modal autocorrelation and the Low-High and High-Low education negative autocorrelation. A 99 count sample is used at the 0.05 significance level. The results are saved and made permanent by saving the shape file.

The shape file is imported into ArcMap. A map is created indicating block groups with high education clustering, low education clustering, low-high education negative autocorrelation, high-low education negative autocorrelation and insignificant values using modal data. This map is used in the modal education – income map overlay.

Now that an income layer, a median education layer, and a modal layer exists they undergo overlay analysis. (Note, since the income layer and education layers work with the same spatial units a join would have worked equally.) A union is performed (Analysis Tools > Overlay > Union). This preserves spatial data and merges the attribute data. Then using the Select By Attributes tool, CL_NumericMed =1 is selected and refined by selecting from the current selection CL_PO77001= 1. This identifies all high median education block groups which also have high median income. Again, using the Select By Attribute tool, CL_NumericMed =2 and CL_PO77001= 2 are selected. These are the low median education, low median income block groups. This procedure is then repeated with the income and modal education layers to compare the high median income and high modal education and low median income and low modal education block groups.

Results

Figure 4 shows the Moran’s I Scatter Plot and the LISA Cluster Map from Washington D.C.’s block group median Income data. The Moran’s I statistic of 0.7211 is positive and distant from 0. This indicates strong clustering. The Moran’s I Scatter Plot graphs the standardized relationship between a variable
on the x-axis and its spatial lag on the y-axis (Anselin, 1993). The points highlighted in yellow coordinate with the clustered block groups in the LISA Cluster Map. These points appear to be distant from the origin which represents random distribution on both axis (0,0). The LISA Cluster Map shows only the significantly clustered data. Clustered high income block groups are in red. Clustered low income block groups are in dark blue. Pink represents block groups with high values that have neighbors with low values. Light blue represents block groups with low values that have neighbors with high values. These lighter colors contribute to global negative autocorrelation or dispersion of the Moran’s I.
Figure 5 displays the LISA Cluster Map with its LISA Significance Map from Washington D.C.’s block group median income data. The LISA Cluster Map is the same as that in Figure 4. It shows the significantly clustered data. The LISA Significance Map maps the point at which a block group is significant. White Block groups are not significant. Light green block groups are significant at the 0.05 level. Medium green block groups are significant at the 0.01 level. This data set does not have block groups with a significance level at the 0.001 level. The significance map corresponds to the block groups included in the cluster. In this case all green block groups are highlighted in the high and low positively autocorrelated block groups since the LISA cluster map is at a p-value of 0.05.
Figure 6. Cluster results from LISA analysis of income data.

Figure 6 more clearly displays the income clusters calculated in figures 4 and 5. Eighty-three red block groups represent the significant block groups from Quadrant I (+,+), and these are the high incomes which fall between $65,833 and $200,001. One hundred fifteen dark blue block groups represent significant block groups in Quadrant III (-,-) of the Moran’s I Scatter Plot. These are low incomes which fall between $7,122 and $57,031. Light blue and pink block groups represent those block groups that contribute to global negative autocorrelation, and grey block groups have p-values higher than 0.05, therefore are insignificant regarding positive or negative autocorrelation.

Figure 7 shows the Moran’s I Scatter Plot and the LISA Cluster Map from Washington D.C.’s block group median education data. The Moran’s I statistic of 0.5830 is positive and distant from 0. This indicates strong clustering. The Moran’s I Scatter Plot graphs the relationship between each block group and its neighboring block groups. The points appear at intervals because the values plotted are ordinal. Yellow
highlighted points correspond to the clustered block groups in the LISA Cluster Map. These points are distant from the origin which represents random distribution. The LISA Cluster Map shows the significantly clustered data. Clustered high median education block groups are in red. Clustered low median education block groups are in dark blue. Pink represents block groups with high values with low valued neighbors. Light blue represents block groups with low values with high valued neighbors. Lighter colors contribute to global negative autocorrelation.

Figure 7. Cluster and Moran’s I results from LISA analysis of median education data.

Figure 8 displays the LISA Cluster Map with its LISA Significance Map from Washington D.C.’s block group median education data. The LISA Cluster Map is the same as that in Figure 7 which shows the significantly clustered income data. The LISA Significance Map demonstrates which block groups are significant and at what level. White Block groups are not significant. Light green block groups are significant at the 0.05 level. Medium green block groups are significant at the 0.01 level. There is no block group with a 0.001 significance level in this data. The significance map indicates which block group in the Cluster map should be mapped as clustered. All green block groups are highlighted in the high and low positively autocorrelated block groups since the LISA cluster map is at a p-value of 0.05.
Figure 8. Cluster and significance results from LISA analysis of median education data.

Figure 9 displays the median education clusters calculated in figures 7 and 8. One hundred and two red block groups represent the significant block groups from Quadrant I of the Moran’s I Scatter Plot. These represent high median education levels which include master’s degrees, bachelor’s degrees, one associates degree and three some college greater than one year. Sixty-six dark blue block groups represent significant block groups in Quadrant III of the Moran’s I Scatter Plot. These are low median education levels which include high school graduates and below. Light blue and pink block groups represent the block groups that contribute to global negative autocorrelation. Grey block groups are insignificant regarding positive or negative autocorrelation.
Figure 9. Cluster results from LISA analysis of median data.

Figure 10 shows the Moran’s I Scatter Plot and the LISA Cluster Map from Washington D.C.’s block group modal education data. The Moran’s I statistic of 0.5034 indicates strong clustering. The Moran’s I Scatter Plot graphs the relationship between the block groups and their neighbors. The points are spaced because the values are ordinal. Yellow highlighted points are the clustered block groups in the LISA Cluster Map. These points are distant from the origin. The LISA Cluster Map shows the significantly clustered data which is dictated by the probabilities plotted on the LISA Significance Map (not shown here). Clustered high modal education block groups are in red. Clustered low modal education block groups are in dark blue. Pink represents block groups with high values with low valued neighbors. Light blue represents block groups with low values with high valued neighbors. Lighter colors contribute to global negative autocorrelation.
Figure 10. Cluster and Moran’s I results from LISA analysis of modal education data.

Figure 11 displays the modal education clusters. One hundred two red block groups represent the significant block groups from Quadrant I of the Moran’s I Scatter Plot. These represent high modal education levels which include one professional degree, master’s degrees, bachelor’s degrees, and one some college greater than one year. Forty-five dark blue block groups represent significant block groups in Quadrant III of the Moran’s I Scatter Plot. These are low modal education levels which include nine bachelor’s degrees, one some college greater than one year and high school graduates. Light blue and pink block groups represent the block groups that contribute to global negative autocorrelation. Grey block groups are insignificant regarding autocorrelation.
Figure 11. Cluster results from LISA analysis of modal data.
Figure 12 indicates the results of overlaying the median income block group layer over the median education block group layer and selecting for high income and high education values and low income and low education values. Block groups that contain high income autocorrelation and high median education autocorrelation are represented in dark blue. From the one hundred two block groups that show high median education autocorrelation and the eighty-three block groups that show high income autocorrelation, eighty-two block groups are preserved. This means eighty-two block groups show high median education autocorrelation and high income autocorrelation. Block groups that contain low income autocorrelation and low median education autocorrelation are represented in red. From the sixty-six block groups that exhibit low median education autocorrelation and the one hundred fifteen block groups that exhibit low income autocorrelation, fifty-five block groups are preserved. This means fifty-five groups show low median education autocorrelation and low income autocorrelation.
Figure 13 displays the results of overlaying the median income block group layer over the modal education block group layer and selecting for high income and high modal education values and low income and low modal education values. Block groups that contain high income autocorrelation and high modal education autocorrelation are represented in green. From the one hundred two block groups that show high modal education autocorrelation and the eighty-three block groups that show high income autocorrelation, seventy-eight block groups are preserved. This means seventy-eight block groups show high modal education autocorrelation and high income autocorrelation. Block groups that contain low income autocorrelation and low modal education autocorrelation are represented in brown. From the forty-five block groups that exhibit low modal education autocorrelation and the one hundred fifteen block groups that exhibit low income autocorrelation, thirty-seven block groups are preserved. This means thirty-seven block groups show low modal education autocorrelation and low income autocorrelation.
The population of Washington D.C. clearly experiences autocorrelation in income, median education and modal education attained by persons over 25 years of age. All Moran’s I values were positive and distant from 0. LISA shows certain clusters for all groups. About 20 percent of block groups indicate positive autocorrelation of high incomes above $65,833. Above 25 percent of block groups indicate positive autocorrelation of low incomes. Almost 24 percent of block groups indicate positive autocorrelation of high median education and high modal education. While 15 percent and 10 percent of block groups indicate autocorrelation of low median education and low modal education.

When the layers undergo overlay analysis data continue to cluster. Nineteen percent of block groups identify with high income autocorrelation and high median education autocorrelation. Thirteen percent of block groups identify with low income autocorrelation and low median education autocorrelation. While eighteen percent of block groups identify with high income autocorrelation and high modal education autocorrelation, and 9 percent of block groups identify with low income and low modal education autocorrelation.

The trends in these analyses suggest several things. These analyses have differing success rates at capturing populations. Identifying clusters of block groups by income work well. Identifying clusters of block groups with high median and modal education attainment works extremely well. One can see nearly all block groups with above a bachelor’s degree median or mode are within the clusters. But, identifying clusters of block groups by low median and modal education works less well. The raw data indicates many block groups of education values equal to those in the low education clusters are scattered throughout the insignificant zone. After the overlay is applied, identifying clusters of block groups with high education and high income is successful, but identifying clusters of block groups with low education and low income is far less successful. Finally, when one compares the results from the analysis using median education and modal education it suggests that analysis using median education produces better (larger) clusters and is more likely to correspond to income clusters.

The implications of these results are mixed for the goals of this study. The methods used identify poor populations. The block groups loosely cluster around two centers in the city. Education clusters very well for highly educated block groups but less so for less educated block groups. This impacts the ability for low education clusters to correspond with low income populations. None the less, 15 percent of Washington D.C.’s block groups cluster by combined low income and low education. This is a positive outcome for beginning to identify variables related to low income. But more importantly, it may indicate other variables come into play with these populations.
Sources


ESRI 2009a ArcGIS Desktop Help: Average nearest neighbor.

ESRI 2009b ArcGIS Desktop Help: How average nearest neighbor distance (Spatial Statistics) works.

ESRI 2009c ArcGIS Desktop Help: Spatial Autocorrelation (Moran’s I) (Spatial Statistics).

ESRI 2009d ArcGIS Desktop Help: How Cluster and Outlier Analysis: Anselin Local Moran’s I (Spatial Statistics) works


