Segregation and Subprime Lending Data: GIS Instructional Memo

Jackelyn Hwang, Michael Hankinson, and K. Steven Brown

July 10, 2014

The purpose of this memo is to guide researchers through the steps necessary to compose our Targeting Index dependent variable and the Clustering Index explanatory variable that pertain to the manuscript, "Racial and Spatial Targeting: Segregation and Subprime Lending within and across Metropolitan Areas." This memo follows steps applicable to ArcMap 10.

- The instructions below follow for .csv datasets created for each individual CBSA. A bold researcher may try to run all 100 CBSAs at once within ArcMap, but let it be done at his/her own risk.

- For a CBSA, create a .csv file containing information for all tracts in each CBSA and create a new column of overall subprime lending rate by dividing the "subprime_count" by "loanscount." Create a similar new column for the subprime lending rate of loans home purchase by dividing "subprimepurch_count" by "loanspurch_count."

- Create a series of new columns to be used as indicators for each tract's percentage of blacks, Hispanics, and blacks-Hispanics combined. Code these indicators as 1 for a minority population greater than 50 percent, as 0 for a minority population less than 40 percent, and a range from 0 to 1 for a minority population between 40 and 50 percent. To create the indicator for blacks, we first need new columns displaying the percentage of each group within the total population. For blacks, divide the total

number of blacks in the census tract (black_nh) by the tracts total population (total pop). Create new columns for Hispanics and black-Hispanic using "totalpop_hispan" and "bhcount," respectively. With these new percentages, we can build our indicators. If the percent black data is located in column AF, we can use the following code:

$$= IF(AF2 >= .5, 1, IF(AF2 <= .4, 0, (AF2 - .4)/.1)).$$

In two new columns, create a similar indicator for both Hispanics and black-Hispanic.

- Save the newly updated spreadsheet as an Excel 97-2003 Workbook, as these files function better within ArcMap 10.

- Open a Blank Map within ArcMap and use the "Add Data" feature to import the newly updated CBSA file. Similarly, import the 2000 U.S. Census tract boundaries shapefile.

- To isolate the specific CBSA, join the .xls document to the census shapefile. First, right click the icon for the tracts shapefile within the Table of Contents. Next, select "Join and Relates," then "Join." Within the "Join Data" menu, under "Choose the field in this layer that the join will be based on:" select "STFID," as this field is Numeric. Under "Choose table," select the CBSA's .xls file. Finally, under "Choose the field in the table to base the join on:," select "geoid." To eliminate extraneous census tract, mark the "Keep only matching records" icon. Click "OK."

- To confirm the join, export the tracts shapefile by right clicking the shapefile icon, selecting "Data," then "Export Data." Save the output under a new name and be sure to save it as a shapefile. ArcMap will offer to add the exported data to the map as a layer. Select "Yes" and right click the new icon, followed by "Zoom To Layer."

- First, let's find the Moran's I for the clustering of each minority racial group. Using the ArcToolbox, select "Spatial Statistics Tools," then "Analyzing Patterns." Finally, open "Spatial Autocorrelation (Morans I)." For "Input Feature Feature Class," use your newly exported CBSA shapefile. For "Input Field," select one of your racial indicators (black, Hispanic, or black-Hispanic). For "Conceptualization of Spatial Relationships," select "Polygon_Contiguity (First Order)." For "Standardization," be sure to select "Row." Click "OK." ArcMap should produce a report in the bottom right corner of the screen. Click on the report and it should offer an Index (your Moran's I) along with a ZScore and PValue. Record the Index as your Clustering Index for the CBSA. Also, note if the PValue is greater than 0, indicating a problem of statistical significance.

- To identify the clustering of black/Hispanic/black-Hispanic neighborhoods, first open the ArcToolbox, following the toolbox offerings from "Spatial Statistics Tools," to "Mapping Clusters," and finally to "Cluster and Outlier Analysis (Anselin Local Moran's I)."

- Within the cluster analysis menu, define your "Input Feature Class" as your recently exported CBSA shapefile. Let's start with mapping the clusters of tracts with high black populations. For "Input Field," select the label for your newly created indicator for black tracts. Under "Output Feature Class," define the new shapefile's name (black-cluster). For "Conceptualization of Spatial Relationships," select "Polygon_Contiguity (First Order)." For "Standardization," be sure to select "Row." ArcMap should add the clustering shapefile as a new layer. Repeat this step for both Hispanics and black-Hispanics, titling your new shapefile's accordingly. You should finish with three shapefiles of racial clustering.

- For the subprime lending rates, repeat the previous step, but for "Input Field," select your newly created column of overall subprime lending rate. Repeat this step once

more for subprime lending rates of homes purchased.

- Join both of your new subprime cluster shapefiles to your first exported file of all the CBSA data. Use the same process as your first join, but select "FID" for both the layer field and the table field. Be sure to "Keep all records." Remember, join both the cluster map of overall subprime loans and of subprime loans for homes purchased.

- To create our Targeting Index, we need separate shapefiles of those tracts belonging to minority clusters. Let's again start with only the clustering of blacks. First, right click the "blackcluster" shapefile and "Open Attribute Table." Scroll right and right click the heading for "COType IDW," selecting "Sort Descending." Using the Shift key, select both the first listing of "HH" and the last listing of "HH." Now that the data is selected, right click the "blackcluster" shapefile again, and export the data as a new shapefile. Repeat this step for both Hispanics and black-Hispanics, creating shapefiles of extracted tracts for each.

- Repeat the previous step, but use the base CBSA shapefile to which we joined the cluster maps for both the total subprime rate and the subprime rate for homes purchased. Having joined the three files, our subprime data also contains population statistics, vital for creating the Targeting Index.

- At this point, we should have five shapefiles of extracted cluster tracts ("HH" tracts only), one for blacks, one for Hispanics, one for black-Hispanic, the others for subprime lending rates and subprime lending rates of homes purchased. To find where they overlap, click the "Geoprocessing" dropdown menu and select "Intersect." Under "Input Features," select one of the racially-based shapefiles and one of the subprime lending shapefiles. Check to save your file under a memorable name. Click "OK." Your new shapefile of intersection should appear in the Table of Contents.

- To find the Targeting Index, open MS Excel, or a similar data processing application. In Excel, open the appropriately compatible version of the intersect shapefile and use the summation function to find the total population of the intersection. Repeat the same steps for the exported shapefile of only clustered subprime tracts (either all or just purchased, depending on which one is used in the Intersect) to find the total population of the subprime lending targeted areas.

- To calculate the Targeting Index, divide the population of the intersection by the population of the entire subprime lending cluster (again, either all subprime loans or just purchased-based loans).

- Now, repeat the Intersect procedure for all combinations of racial groups (black, Hispanic, and black-Hispanic) and subprime loans (all and purchase-based).

- For future analysis, join all three of the shapefiles for racial clustering to the base CBSA shapefile, keeping all records. In Excel, create new columns granting a dummy variable to whether the each tract is part of a larger cluster (one column for black, one for Hispanic, and one for black-Hispanic). For instance, if column Z contains the COType, the new column's formula would read: $= IF(Z = "HH", 1, 0)$. Copy and paste the values of these new columns so that you may delete the now vestigial cluster-join data. Along with the original census data, you should now have new columns regarding subprime lending rate, percent minority, minority indicators, and dummy variables regarding placement as part of a larger cluster.

With that, congratulations on the replication and we look forward to your comments, criticisms, and suggestions!