

Practically Meaningful (but Statistically Insignificant) Method for Smoothing Census Data for Inclusion into a Risk Terrain Model

Keywords: Census data, smooth, disaggregate, socio-demographic risk factor

Aim: To propose a method for disaggregating Census data and smoothing the harsh boundaries between polygon features.

Excerpt:

- "This means that the smoothing technique demonstrated here removed the harsh boundaries between polygons and smoothed the data in such a way that made it practically more meaningful, but not significantly different from the original data."

By: Joel M. Caplan

Assistant Professor, Rutgers School of Criminal Justice

Background

When census or other data variables are aggregated to polygons such as tracts, blockgroups, or blocks, analysts often feel the need to aggregate all other data to these polygons out of necessity--since disaggregating such data is considered impossible. Aggregating point-level data to polygons raises several issues, including the modifiable areal unit problem (MAUP) and edge effects.

While the value of a Census polygon is representative of the space within it, it does not account for the variation of values among micro-level places within the polygon, nor does it account for the (likely) more realistic gradual changes that occur across polygons throughout the terrain.

Raster tools can be used to model the "smoothed" or gradual variation of values within and across polygons. Presented here is an example of how this can be done. The data source is U.S. Census year 2000 vacancy rates per block (calculated as: number of vacant properties / total number of dwellings). The color scheme is such that darker colors represent higher values.

Smoothing Procedure

Figure 1 is a thematic map that shows vacancy rates per Census block. Within each block polygon is the centroid point. For reasons I will explain momentarily, the average distance between the block centroids was calculated using the "Nearest Neighbor Analysis Tool" in ArcToolbox.

Then, this feature map of Census blocks was converted to a raster grid layer using the "Convert > Features to Raster" function in the ArcGIS Spatial Analyst extension. "Vacancy rate" was the reference value and the cell size was set to 100ft. The resulting raster map is shown in Figure 2.

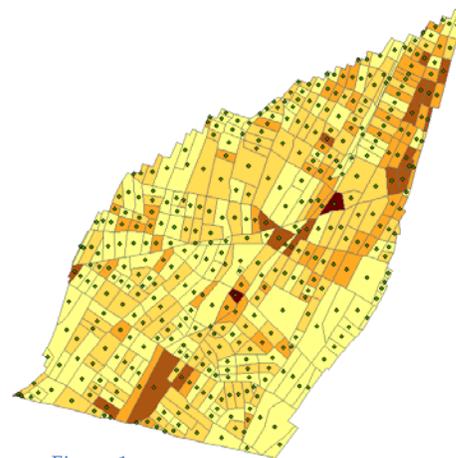


Figure 1

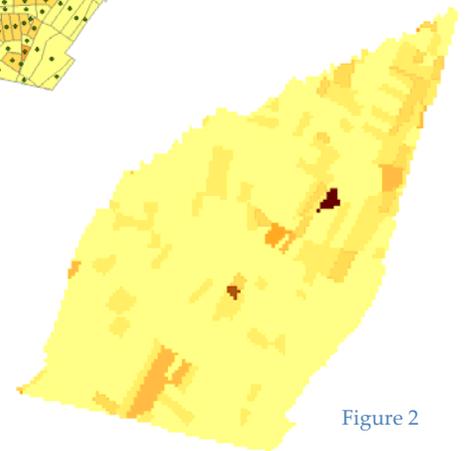
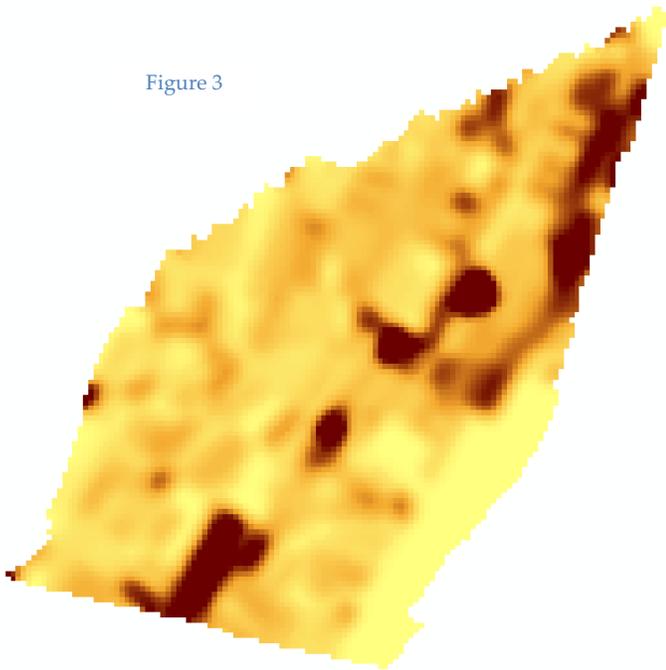


Figure 2

The "Neighborhood Statistics" function in the Spatial Analyst Extension was used to "smooth" the raster grid layer (depicted in Figure 2) using the mean "statistic type." This assigned each cell in the grid layer a new value according to the average value calculated from all nearby cells up to a certain distance. This "Neighborhood" of nearby cells is one of the parameters of the Neighborhood Statistics operation, and for this example, it was set as a "radius" of 292 feet. This radius was selected because it was the average nearest neighbor distance between each Census block's centroid point.

The Neighborhood Statistics function was performed on the original raster grid layer (i.e. Figure 2) and was then repeated on the resulting output layer, respectively, for a total of five times. Not much changed visually to the map after the 3rd iteration, but the repetition of five averaging operations ensured a completely smoothed map layer. The resulting raster map layer after five iterations is shown below as Figure 3 (This map was clipped to the city outline feature to trim the edges--which grew outward slightly beyond the study area with each iteration).

Figure 3

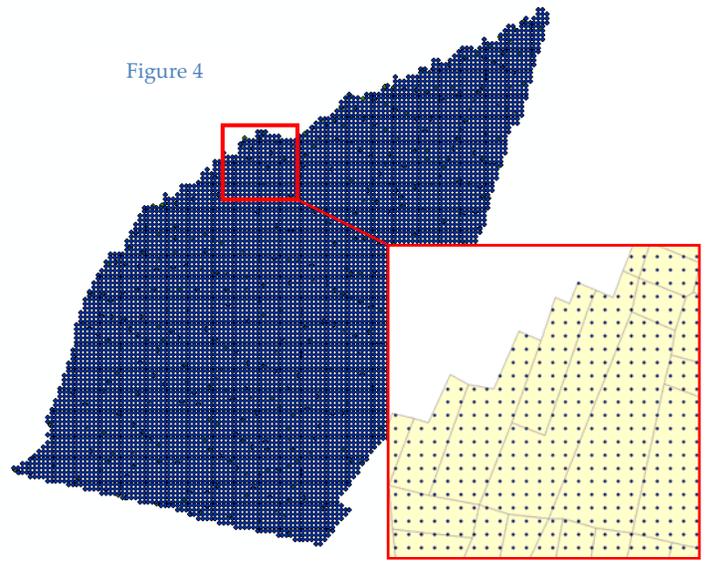


Validating the Smoothing Procedure

Results of this raster grid "smoothing" procedure were compared to the original thematic map. To do this, the "smoothed" raster map was converted to point features (using "Convert > Raster to Features" in the Spatial Analyst Extension). This resulted in a map similar to Figure 4, with each point having the attribute value of the smoothed

layer's respective cell value. These points were "Spatially Joined" to each Census block polygon of the original shapefile (i.e., Figure 1) so that the "mean" value of all points that fall within each polygon was calculated. Now, the original thematic map of Census blocks has two attribute values for each block feature: 1) the vacancy rate obtained from the original census data, and 2) the average value of all smoothed points (points are actually the centroids of each 100ftx100ft raster cell; that's how the conversion from raster to point features works...the points are placed at the center of each cell).

Figure 4



The attribute table of the "Joined" Census block thematic map was exported as a DFB table and opened in SPSS Statistical Software. A Paired T-Test was performed to compare the vacancy rate values before and after the smoothing procedure.

There were 409 census blocks in the study area. The original mean vacancy rate was .084; the smoothed mean vacancy rate per block was .086. Results suggest that these means were highly correlated (Correlation=.84, $p < 0.01$). And, as shown in the table below, the mean vacancy rates for census blocks were not significantly different ($p = 0.51$). This means that the smoothing technique demonstrated here removed the harsh boundaries between polygons and smoothed the data in such a way that made it practically more meaningful, but not significantly different from the original data.

Table: T-Test Result of Feature_VacancyRate -- RasterCell_AvgValue

Paired Differences					t	df	2-tailed Sig.
			95% Conf. Int.				
Mean	Std. Dev.	Std. Error	Lower	Upper			
-.002	.060	.003	-.008	.004	-.647	408	.518