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 ronald.e.wilson@hud.gov for review.*

Changing Geographic Units and the Analytical Consequences: An Example of Simpson's Paradox

Ron Wilson
U.S. Department of Housing and Urban Development
University of Maryland, Baltimore County

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research or the U.S. Department of Housing and Urban Development.

Foreclosures and Crime

The rapidly degrading housing market of the mid-2000s caused local governments to be concerned about the multitude of problems foreclosures could wreak on their jurisdictions (Wilson and Paulsen, 2008). One concern was the escalation of crime and disorder in neighborhoods with concentrated foreclosures. Several researchers who examined the relationship between foreclosure and crime had conflicting results (Arnio and Baumer, 2012; Arnio, Baumer, and Wolff, 2012; Baumer, Wolff, and Arnio, 2012; Cui, 2010; Ellen, Lacoe, and Sharygin, 2011; Goodstein and Lee, 2010; Immergluck and Smith, 2006; Jones and Pridemore, 2012; Katz, Wallace, and Hedberg, 2011; Kirk and Hyra, 2012; Stucky, Ottensmann, and Payton, 2012; Wallace, Hedberg, and Katz, 2012). The assortment of geographic units used in these studies is extensive, consisting of property locations, block faces, census block groups, census tracts, customized local geographies, grid cells, cities, counties, and metropolitan statistical areas. The variety of factors, constructs, and variables the researchers used in these studies certainly contributed to their conflicting results, but the range of geographies likely played a role in the outcome differences, because the underlying data were aggregated to different geographic scales.

Conflicting results are common in social science research from the use of different geographic units of analysis (Coulton et al., 2001; Hipp, 2007; Macintyre, Ellaway, and Cummins, 2002; Rengert and Lockwood, 2009; Taylor, 2012). None of the cited studies, though, included tests of the foreclosure and crime relationship with multiple geographic units to gauge the effect on results. I illustrate in this article how changing geographic units can produce converse results with an example of foreclosure and crime modeling drawn from Wilson and Behlendorf (2013). I also conduct a spatial analysis to identify which geographic unit is best for modeling foreclosures and crime in the Wilson and Behlendorf (2013) example, using several spatial analysis techniques.

Modeling Foreclosures and Neighborhood Crime in Charlotte and Mecklenburg County

Wilson and Behlendorf (2013) analyzed the relationship between foreclosures and neighborhood crime in the city of Charlotte and Mecklenburg County, North Carolina, with four crime constructs[1] for the years 2006 and 2007. Point-level crime[2] and single-family foreclosure[3] locations were aggregated to census block group and tract geographies to compare results. Several demographic, economic, and environmental variables were included to form a set of explanatory factors (concentrated disadvantage, neighborhood quality, residential stability, and immigration concentration) known to be associated with neighborhood crime. The spatial proximity of other foreclosures and the temporal occurrence of crime were also included as controls to account for concurrent events in nearby places and time that may have an influence on the outcome. A negative binomial regression count model was used:

$$\lambda_{crimetype} = \ln(n_{pop-at-risk}) + \beta_{foreclosure} x_{foreclosure} + \sum_{k=0}^{k} \beta_k x_{ik}[CONCENTRATED\ DISADVANTAGE]$$
$$+ \sum_{k=0}^{k} \beta_k x_{ik}[NEIGHBORHOOD\ QUALITY] + \sum_{k=0}^{k} \beta_k x_{ik}[RESIDENTIAL\ INSTABILITY]$$
$$+ \sum_{k=0}^{k} \beta_k x_{ik}[IMMIGRATION\ CONCENTRATION] + \sum_{k=0}^{k} \beta_k x_{ik}[CONTROLS] \qquad (1)$$

The central finding from Wilson and Behlendorf (2013) was that the rate of foreclosures had a positive and significant association with crime increases in 2006 and 2007, but results differed between geographic units. The full output for the two geographies is shown in exhibits 1 (tracts) and 2 (block groups), but I focus on the residential instability factor and the spatial lag variable for the remainder of this analysis.

[1] Crimes of (1) violence, (2) property, (3) residential burglary, and (4) minor property damage.

[2] Crime data were supplied by the Charlotte-Mecklenburg Police Department; followed the Uniform Crime Report, or UCR, classifications; and were geocoded to the specific street of occurrence.

[3] The Department of Geography and Earth Sciences at the University of North Carolina at Charlotte provided parcel data. I identified foreclosed properties where the title transfer date indicated bank repossession ending in involuntary vacancy.

Exhibit 1

Negative Binomial Regression: Crime on Single-Family Foreclosures, Census Tracts, 2006 and 2007 (1 of 2)

	2006				2007			
	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes
Foreclosure rate								
Previous year's foreclosure rate per 100 housing units	0.148	0.161	0.294***	0.006	0.145	0.246**	0.351***	0.118
	(0.116)	(0.104)	(0.108)	(0.095)	(0.116)	(0.099)	(0.107)	(0.098)
Interaction between neighborhoods with high disadvantage (+1 sd) and the previous year's foreclosure rate	– 0.258	– 0.322**	– 0.321*	– 0.207	– 0.218	– 0.274**	– 0.244*	– 0.190
	(0.183)	(0.155)	(0.172)	(0.153)	(0.148)	(0.118)	(0.130)	(0.120)
Social disorganization								
Concentrated disadvantage	0.293**	0.400***	0.285**	0.410***	0.313***	0.479***	0.244**	0.289***
	(0.136)	(0.121)	(0.130)	(0.115)	(0.120)	(0.104)	(0.110)	(0.106)
Residential stability	– 0.137	– 0.306***	– 0.153*	– 0.113	– 0.082	– 0.273***	– 0.080	– 0.074
	(0.091)	(0.084)	(0.080)	(0.073)	(0.095)	(0.079)	(0.077)	(0.078)
Percent Hispanic	– 0.983	– 0.005	– 1.971*	– 1.168	– 0.084	0.944	– 1.325	0.297
	(1.206)	(1.110)	(1.114)	(1.025)	(1.228)	(1.144)	(1.057)	(1.070)
Percent foreign born	7.659	– 0.298	– 0.544	7.770	9.594	1.506	13.654	8.336
	(9.866)	(8.959)	(9.323)	(8.098)	(10.190)	(8.602)	(9.383)	(8.656)
Neighborhood construction and quality								
Population density	– 0.000	– 0.000	0.000	– 0.000	– 0.000	– 0.000	0.000	– 0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Median age of neighborhood	– 0.022***	– 0.014***	– 0.016***	– 0.012***	– 0.021***	– 0.013***	– 0.018***	– 0.013***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Percent of neighborhood built after 1999	1.977***	1.936***	1.418***	1.724***	1.660***	1.772***	1.500***	1.131
	(0.507)	(0.475)	(0.484)	(0.415)	(0.518)	(0.473)	(0.454)	(0.458)
Percent of units coded as good or better construction	0.543*	– 0.256	– 0.115	0.397	0.845**	– 0.000	0.175	0.761**
	(0.329)	(0.314)	(0.323)	(0.285)	(0.343)	(0.317)	(0.326)	(0.300)

Exhibit 1

Negative Binomial Regression: Crime on Single-Family Foreclosures, Census Tracts, 2006 and 2007 (2 of 2)

	2006				2007			
	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes
Other control variables								
Adults-to-children ratio	0.002	– 0.001	– 0.004	0.002	0.002	0.001	– 0.003	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Percent of population that are males ages 14–24	1.402	**3.558*****	0.823	– 0.284	1.387	1.153	– 0.350	1.857
	(1.638)	(1.563)	(1.600)	(1.414)	(1.711)	(1.538)	(1.554)	(1.510)
Percent currently divorced	4.718*	**8.982*****	2.848	2.979	**5.191****	**8.701*****	2.602	**5.198****
	(2.441)	(2.330)	(2.270)	(2.046)	(2.503)	(2.207)	(2.179)	(2.161)
Spatial lag	0.001	0.004	0.003	0.005*	0.000	0.004	– 0.000	0.004
	(0.001)	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)
Two-year crime lag	**0.002*****	**0.010*****	**0.008*****	**0.013**	**0.002*****	**0.009*****	**0.009*****	**0.011*****
	(0.000)	(0.002)	(0.002)	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)
_cons	37.753***	19.961**	25.664***	18.372**	35.925***	19.125**	29.729***	18.264**
	(9.354)	(8.983)	(9.286)	(7.950)	(9.611)	(8.897)	(8.651)	(8.501)
/lnalpha	– 0.773***	– 1.129***	– 0.998***	– 1.226***	– 0.719***	– 1.182***	– 1.099***	– 1.067***
	(0.123)	(0.156)	(0.145)	(0.146)	(0.124)	(0.155)	(0.147)	(0.143)
Number of observations	143	143	143	143	143	143	143	143
Adjusted R^2	0.070	0.140	0.083	0.097	0.069	0.148	0.086	0.088
Log-likelihood	– 917.96	– 594.67	– 662.69	– 665.03	– 926.53	– 584.05	– 657.57	– 686.14
aic	1,869.926	1,223.334	1,359.375	1,364.054	1,887.062	1,202.095	1,349.143	1,406.276
bic	1,920.295	1,273.702	1,409.744	1,414.422	1,937.430	1,252.463	1,399.511	1,456.645

aic = Akaike Information Criterion. bic = Bayesian Information Criterion. sd = standard deviation.

** $p < .10$. ** $p < .05$. *** $p < .01$.*

Note: Bold values are the most statistically significant coefficients.

Exhibit 2

Negative Binomial Regression: Crime on Single-Family Foreclosures, Block Groups, 2006 and 2007 (1 of 2)

	2006				2007			
	Property Crimes	**Violent Crimes**	**Residential Burglary**	**Minor Property Crimes**	**Property Crimes**	**Violent Crimes**	**Residential Burglary**	**Minor Property Crimes**
Foreclosure rate								
Previous year's foreclosure rate per 100 housing units	0.073 (0.057)	**0.153***** (0.051)	**0.157***** (0.053)	0.049 (0.049)	**0.156***** (0.055)	**0.255***** (0.045)	**0.259***** (0.050)	**0.165***** (0.047)
Interaction between neighborhoods with high disadvantage (+1 sd) and the previous year's foreclosure rate	– 0.183* (0.103)	**– 0.192***** (0.069)	**– 0.160**** (0.075)	– 0.103 (0.070)	– 0.115 (0.076)	**– 0.183***** (0.058)	– 0.128* (0.068)	**– 0.133**** (0.063)
Social disorganization								
Concentrated disadvantage	**0.198***** (0.069)	**0.379***** (0.065)	**0.202***** (0.065)	**0.287***** (0.060)	**0.179***** (0.069)	**0.443***** (0.062)	**0.145**** (0.066)	**0.269***** (0.061)
Residential stability	**0.203***** (0.057)	**0.315***** (0.052)	**0.117**** (0.048)	**0.162***** (0.048)	**0.208***** (0.059)	**0.314***** (0.049)	**0.114**** (0.051)	**0.226***** (0.049)
Immigration concentration	0.005 (0.053)	0.056 (0.048)	– 0.093* (0.049)	0.000 (0.047)	0.047 (0.054)	**0.140***** (0.045)	– 0.070 (0.050)	0.064 (0.047)
Neighborhood construction and quality								
Population density	– 0.000* (0.000)	**– 0.000**** (0.000)	0.000* (0.000)	– 0.000* (0.000)	**– 0.000***** (0.000)	**– 0.000**** (0.000)	0.000 (0.000)	**– 0.000***** (0.000)
Median age of neighborhood	**– 0.026***** (0.003)	**– 0.020***** (0.003)	**– 0.017***** (0.003)	**– 0.014***** (0.003)	**– 0.027***** (0.003)	**– 0.021***** (0.003)	**– 0.017***** (0.003)	**– 0.017***** (0.003)
Percent of neighborhood built after 1999	**2.265***** (0.326)	**1.681***** (0.298)	**1.816***** (0.303)	**1.685***** (0.279)	**2.151***** (0.334)	**1.617***** (0.290)	**1.906***** (0.310)	**1.342***** (0.285)
Percent of units coded as good or better construction	– 0.114 (0.178)	**– 0.926***** (0.182)	**– 0.452**** (0.176)	**– 0.413**** (0.166)	0.035 (0.186)	**– 0.605***** (0.179)	– 0.263 (0.177)	– 0.132 (0.168)

Exhibit 2

Negative Binomial Regression: Crime on Single-Family Foreclosures, Block Groups, 2006 and 2007 (2 of 2)

	2006				2007			
	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes	Property Crimes	Violent Crimes	Residential Burglary	Minor Property Crimes
Other control variables								
Adults-to-children ratio	0.001	0.000	– 0.003	0.001	0.003	0.002	– 0.001	0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Percent of population that are males ages 14–24	– 0.533	0.968	– 0.480	– 1.112	– 0.876	– 1.815*	**– 2.231****	– 1.559
	(1.285)	(1.091)	(1.122)	(1.072)	(1.295)	(1.015)	(1.096)	(1.023)
Percent currently divorced	**4.345*****	**5.647*****	1.567	**3.386*****	**3.754*****	**4.848*****	0.623	**3.059*****
	(1.153)	(1.085)	(1.119)	(0.995)	(1.176)	(1.024)	(1.087)	(1.017)
Spatial lag	0.001*	**0.015*****	**0.009****	**0.011*****	0.001	**0.015*****	0.006*	**0.008****
	(0.001)	(0.004)	(0.004)	(0.003)	(0.001)	(0.003)	(0.004)	(0.003)
Two-year crime lag	**0.004*****	**0.020*****	**0.016*****	**0.018*****	**0.004*****	**0.018*****	**0.016*****	**0.018*****
	(0.000)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
_cons	46.582***	34.037***	27.077***	22.780***	48.807***	35.697***	27.829***	27.981***
	(6.329)	(5.806)	(6.087)	(5.419)	(6.276)	(5.489)	(5.931)	(5.475)
/lnalpha	– 0.518***	– 1.022***	– 0.837***	– 0.959***	– 0.475***	– 1.155***	– 0.827***	– 0.892***
	(0.077)	(0.108)	(0.095)	(0.096)	(0.077)	(0.114)	(0.096)	(0.095)
Number of observations	372	372	372	372	372	372	372	372
Adjusted R^2	0.067	0.163	0.080	0.094	0.066	0.174	0.081	0.093
Log-likelihood	– 2,046.04	– 1,196.89	– 1,403.63	– 1,404.95	– 2,062.29	– 1,158.82	– 1,397.64	– 1,435.57
aic	4,126.079	2,425.785	2,839.258	2,841.893	4,156.588	2,349.649	2,827.285	2,903.130
bic	4,192.700	2,488.487	2,901.960	2,904.595	4,219.290	2,412.352	2,889.987	2,965.833

aic = Akaike Information Criterion. bic = Bayesian Information Criterion. sd = standard deviation.

** $p < .10$. ** $p < .05$. *** $p < .01$.*

Note: Bold values are the most statistically significant coefficients.

The residential stability coefficients changed dramatically between tracts and block groups. Residential stability represents the level of social connections between neighborhood residents. Stable neighborhoods have a constancy of residents who remain in their homes over long periods of time and they know, trust, like, and communicate with their neighbors. Residential stability degrades when residents leave and new ones move into a neighborhood—that is, turnover—and social bonds are broken. Crime can increase if residential turnover is frequent, because social connections are strained and neighbors do not trust or know each other (Garcia, Taylor, and Lawton, 2007; Shaw and McKay, 1942). The residential stability factor was constructed as a scale centered on 0 and includes the percentage of (1) residents who are 5 years of age and older who lived in the same house 5 years earlier, (2) owner-occupied homes, and (3) single-family and multifamily housing units. The scale was reverse coded to represent instability with positive numbers and stability with negative numbers.

Exhibit 1 shows that residential instability is significantly associated with all crime constructs in both 2006 and 2007 for block groups, but only for violence with tracts (exhibit 2). Not only did the signs change from negative to positive between geographies, but also the coefficients remained statistically significant with large effects. With tracts, the interpretation is that crime decreased as residential instability increased, but, with block groups, the converse was true in that crime increased in less residentially stable neighborhoods; the latter scenario is theoretically expected. This sign switching between geographies is indicative of local spatial patterns being lost with the use of larger geographic units—the significance change of the spatial lag coefficient between the two geographies highlights this point.

The spatial lag variable represents measures of similarity and dissimilarity with nearby geographic units for a foreclosure contagion effect. The significance level of the spatial lag variable means a spatial effect is present in the relationship and should be modeled. Ignoring the spatial effect can bias parameter estimates and significance levels (Anselin et al., 2000), because existence of a spatial effect is an artifact of the measured relationships. The spatial lag coefficient is significant for block groups for most crime constructs across both years, but it is not significant for tracts. Tract results suggest no spatial contagion effect exists for foreclosures in relation to crime. This finding indicates the inability of census tracts to capture an existing spatial relationship between foreclosures and crime.

Conflicting Results and Simpson's Paradox

Coefficient sign reversals, especially when they remain statistically significant, can indicate model misspecification. Sign reversals can also occur when different geographic units are used, however, because the change alters data distribution patterns. Known as *Simpson's Paradox*, the repartitioning of the underlying data from smaller to larger geographic units can cancel out or reverse patterns in smaller units. The paradox is a consequence of the modifiable areal unit problem (MAUP)[4] in which statistical results are affected by modifications to the geographic unit's boundary size and/or shape. Aggregated data are uniquely partitioned by their geography and, when geographic units are

[4] For an indepth technical discussion of MAUP, see Openshaw (1994).

changed, the new boundary sizes and shapes are repartitioned. Exhibit 3 depicts how Simpson's Paradox occurs between census block groups and tracts for residential stability and 2006 violent crime counts for Charlotte and Mecklenburg County.

Exhibit 3

Residential Stability and Violent Crime Scatter Plots for Charlotte and Mecklenburg County, (a) Census Block Groups, (b) Census Tracts

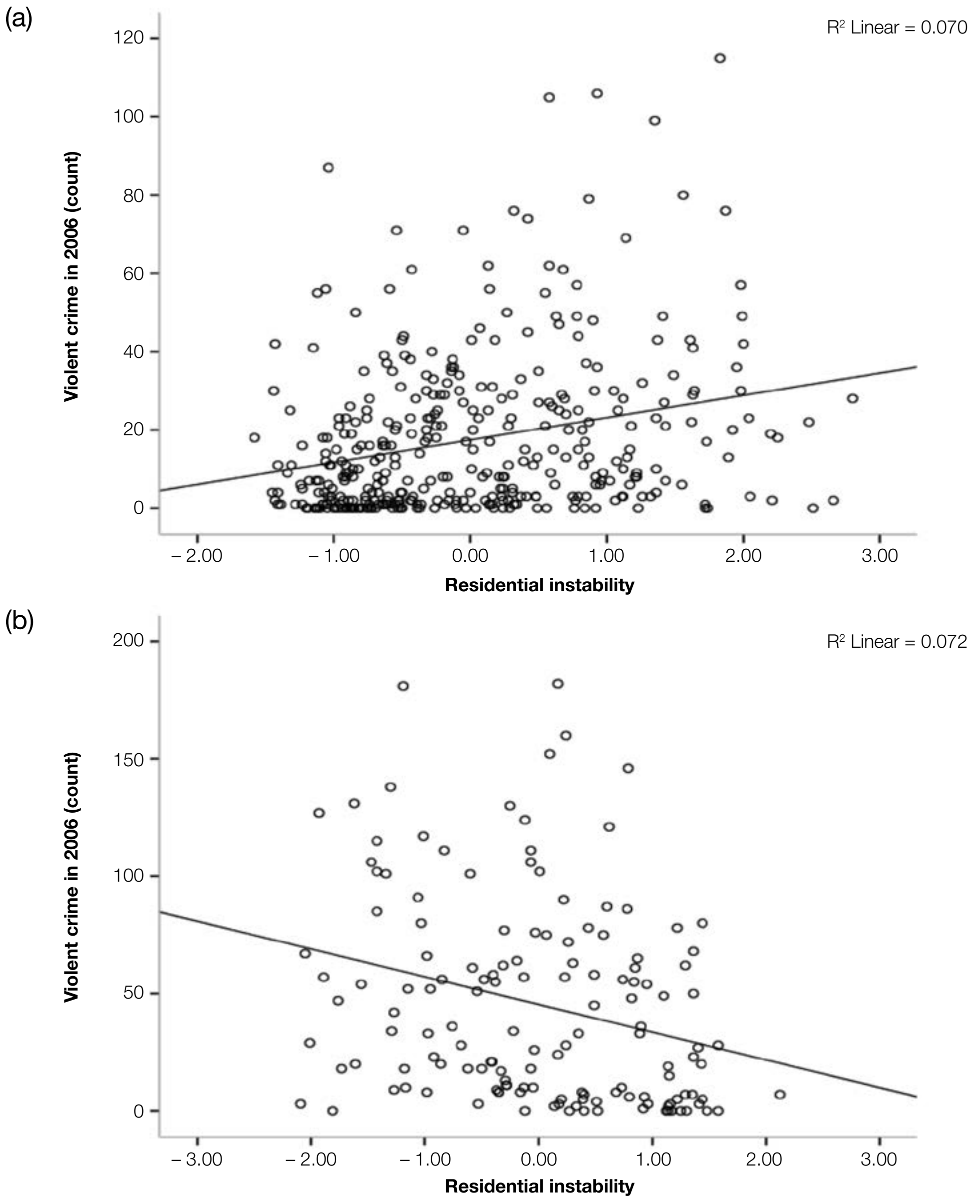


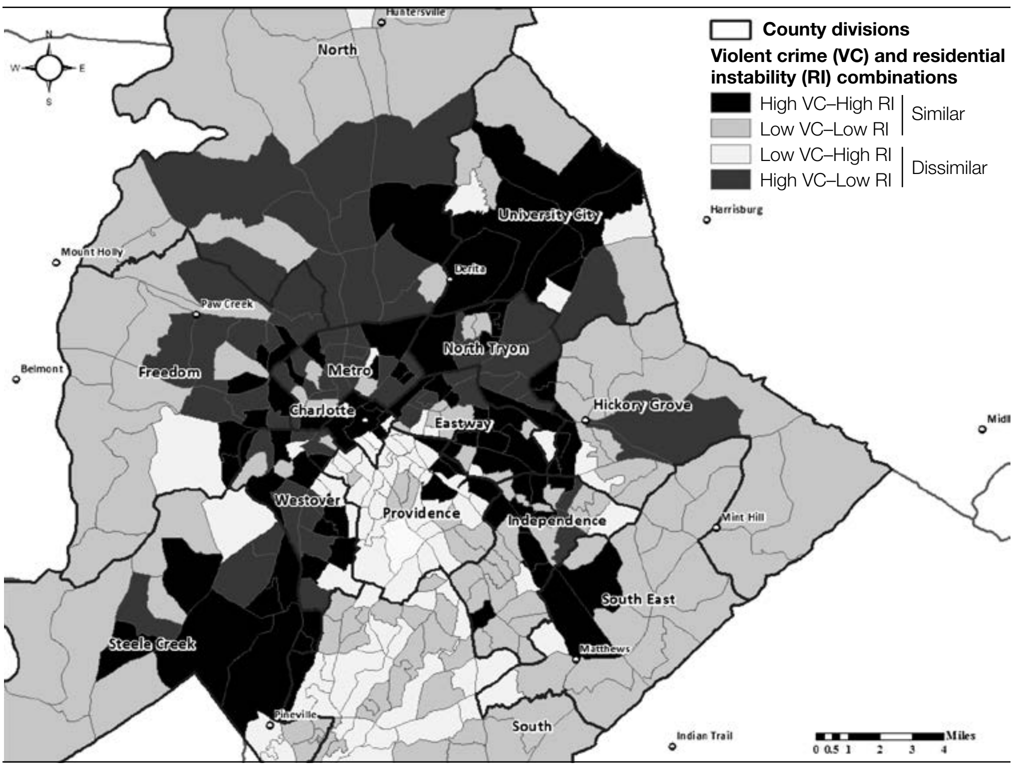
Exhibit 3 shows Simpson's Paradox with the trend lines in the two scatter plots being the inverse of each other with the data clouds practically being mirror images of each other. Exhibit 3a shows the block group data pattern and is interpreted as—when residential instability increases, crime also increases. The opposite pattern occurs for tracts (exhibit 3b) and is interpreted as—when crime decreases, residential instability increases.

To show Simpson's Paradox geographically, I converted the violent crime and residential instability values into z-scores to categorize their relationship as *similar*[5] and *dissimilar*.[6] Exhibits 4 (block groups) and 5 (tracts) exemplify Simpson's Paradox more visually by displaying the similar and dissimilar categories. Stark geographic pattern changes occur for the violent crime and residential instability relationship between the geographic units.

The exhibits show several areas across the county that change from similar levels of violent crime and residential instability to dissimilar levels when switching from block groups to tracts. For example, the Providence and Independence Divisions (patrol divisions of the Charlotte-Mecklenburg Police Department jurisdiction) south of city center, show near complete reversal patterns. The

Exhibit 4

Violent Crime and Residential Instability Combinations at Block Group Geography

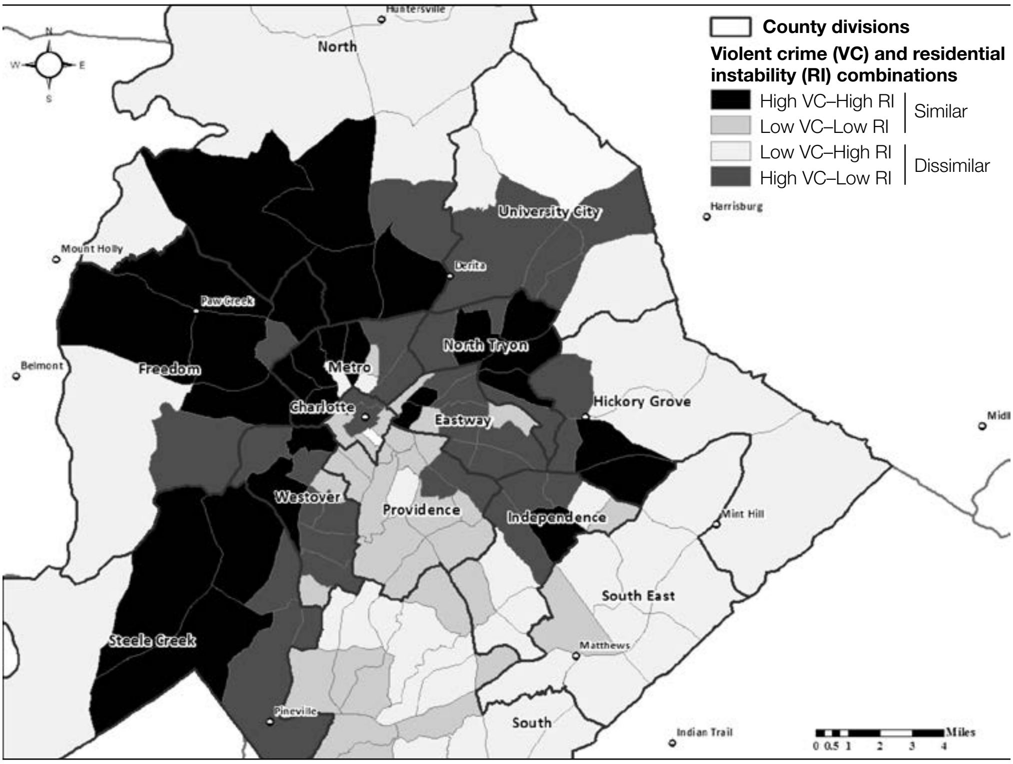


[5] Violent crime and residential instability values both have positive or negative z-scores.

[6] Violent crime and residential instability values have conflicting positive and negative z-scores.

Exhibit 5

Violent Crime and Residential Instability Combinations at Tract Geography



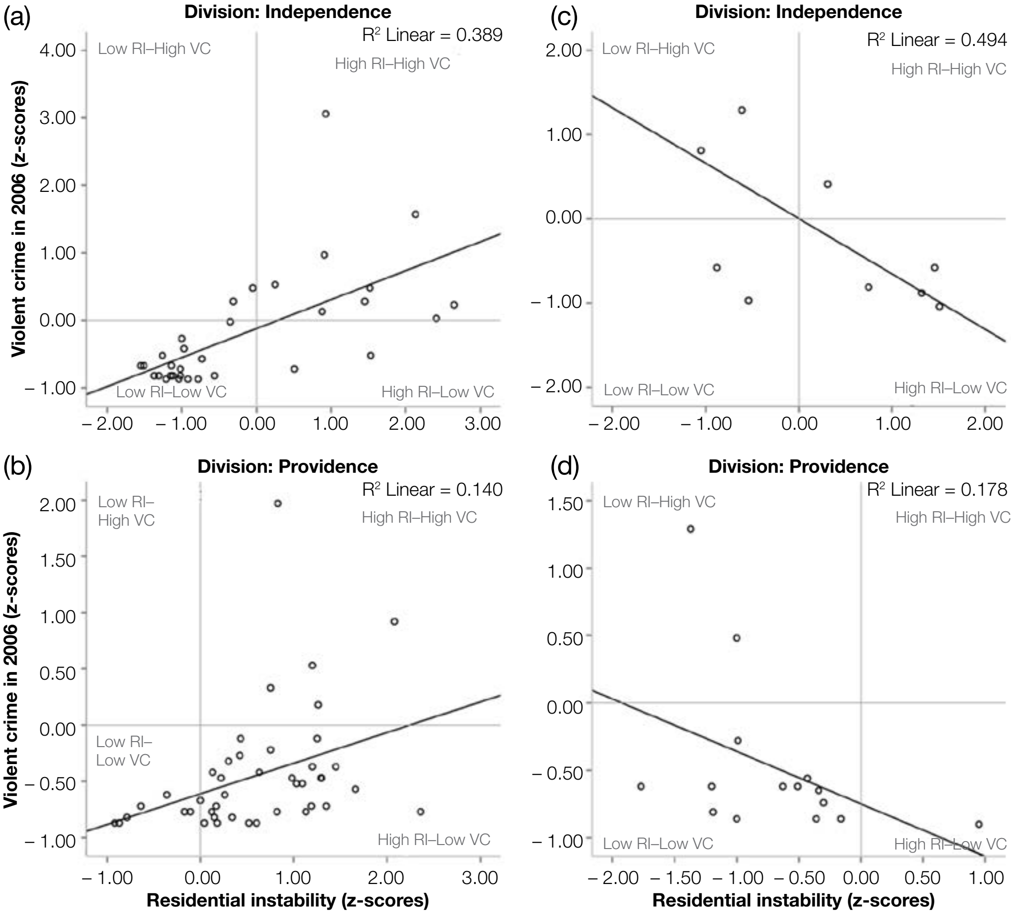
Providence Division primarily contained low levels of violent crime with high residential instability at the block group level, indicating dissimilarity between the two variables. With census tracts, the violent crime and residential instability relationship changed from similarity to dissimilarity. In the Independence Division, the opposite was true, in that a similarity existed between high and low levels of violent crime and residential instability within block groups, but changed to a dissimilar relationship within tracts.

Exhibit 6 shows the violent crime and residential instability scatter plots for the Independence and Providence Divisions. The scatter plots show the same overall trends of violence and residential instability as exhibit 3 shows for the county, but they help explain the pattern changes between exhibits 4 and 5. For example, 30 of the 43 (69 percent) block groups within the Providence Division have high residential instability with low violent crime, but 12 of 15 tracts (80 percent) now have low violence and low residential instability categories. The similarity and dissimilarity categories were altered as the x and y axes in the scatter plots shifted significantly to contain different observations. Exhibits 4 and 5 also show that category changes significantly alter the trends across the county.

Simpson's Paradox prompts a dilemma in deciding which geographic unit to use for further analysis. Theoretical or expected results could guide the selection of geography, but they may not solve

Exhibit 6

Violent Crime and Residential Instability Scatter Plots for the Independence and Providence Divisions of Charlotte and Mecklenburg County, (a and b) Census Block Groups, (c and d) Census Tracts



RI = residential instability. VC = violent crime.

Note: The upper right quadrant is a similarity between high VC and high RI. The lower left quadrant is a similarity between low VC and low RI. The upper left quadrant is a dissimilarity between low VC and high RI. The lower right quadrant is a dissimilarity between high VC and low RI.

the paradox. In the next section, I demonstrate how to identify which geographic unit is more appropriate for measuring the relationship of foreclosures and crime in Wilson and Behlendorf (2013).

Examining Local Data To Identify the Spatial Extent of Foreclosures

An important aspect in Wilson and Behlendorf (2013) was the inclusion of the spatial lag (autocorrelation) measure of foreclosures to test for a contagion effect. The spatial lag coefficient in exhibits 1 and 2 was significant for block groups, but not for tracts. If a spatial contagion effect exits amongst foreclosures, then it is important to identify which geographic unit would best capture the effect, because crimes related to those properties would occur at a similar scale. I used several spatial analysis techniques to measure foreclosure concentration and to determine which geographic unit would be better to model with crime-related factors.

I first conducted a nearest neighbor analysis on foreclosed parcels from 2003 to 2008[7] to obtain the average distance between the properties. The nearest neighbor index was 0.3835 (z = -136.92), which indicates a strong clustering pattern. The average distance between foreclosed properties is 264.3 feet, with a standard distance of 374.8 feet. These two results indicate that foreclosures were very close to each other and often on the same or adjacent streets.

Next, to provide a measure of the contagion effect, I conducted a near repeat analysis[8] on foreclosures to identify how far and fast foreclosures were spreading (see exhibit 7). The near repeat

Exhibit 7

Observed Over Mean Expected Values for Near Repeat Foreclosures

	0 to 30 Days	31 to 60 Days	61 to 90 Days	91 to 120 Days	121 to 150 Days	151 to 180 Days	More Than 180 Days
Observed over mean expected frequencies table							
Same location	**13.69**	0.88	0.96	0.83	0.77	0.74	0.56
1 to 660 feet	**1.54**	**1.22**	**1.29**	**1.15**	**1.13**	**1.16**	0.95
661 to 1,320 feet	**1.31**	**1.12**	**1.12**	**1.08**	**1.07**	**1.08**	0.97
1,321 to 1,980 feet	**1.04**	**1.05**	**1.02**	**1.05**	**1.04**	1.01	0.99
1,981 to 2,640 feet	0.98	1.00	0.98	**1.04**	0.98	1.00	1.00
2,641 to 3,300 feet	**1.02**	**1.03**	1.00	**1.06**	1.01	**1.05**	0.99
3,301 to 3,960 feet	1.00	**1.04**	1.02	1.01	**1.10**	**1.05**	0.99
3,961 to 4,620 feet	1.01	**1.03**	1.02	**1.03**	**1.02**	**1.05**	0.99
4,621 to 5,280 feet	1.02	**1.02**	1.01	**1.04**	1.02	1.02	1.00
More than 5,280 feet	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**
Statistical significance table							
Same location	**0.05**	0.95	0.85	1.00	1.00	1.00	1.00
1 to 660 feet	**0.05**	**0.05**	**0.05**	**0.05**	**0.05**	**0.05**	1.00
661 to 1,320 feet	**0.05**	**0.05**	**0.05**	**0.05**	**0.05**	**0.05**	1.00
1,321 to 1,980 feet	**0.05**	**0.05**	**0.05**	**0.05**	**0.05**	0.25	1.00
1,981 to 2,640 feet	0.85	0.60	0.85	**0.05**	0.90	0.55	0.30
2,641 to 3,300 feet	**0.05**	**0.05**	0.45	**0.05**	0.30	**0.05**	1.00
3,301 to 3,960 feet	0.45	**0.05**	0.10	0.45	**0.05**	**0.05**	1.00
3,961 to 4,620 feet	0.20	**0.05**	0.15	**0.05**	**0.05**	**0.05**	1.00
4,621 to 5,280 feet	0.25	**0.05**	0.20	**0.05**	0.15	0.20	1.00
More than 5,280 feet	1.00	1.00	1.00	1.00	1.00	1.00	**0.05**

Note: Bold values are statistically significant at the p ≤ .01 level.

[7] I used date ranges beyond our focus years to ensure capture of the long-term distribution patterns and reduce the likelihood of any anomalous cluster patterns that might have occurred at the peak of the housing crisis.

[8] The concept of near repeats extends to housing research because several documented problems—such as voucher relocations, property-code violations, price shocks, or foreclosures—have been shown to have patterns of spatial and temporal proximity. The theoretical underpinnings of near repeat research in criminology are geographic in nature, rooted in the first law of geography (Miller, 2004) that everything is related, but closer things are more related because they share common characteristics. I used the near repeat calculator, which can be found at http://www.temple.edu/cj/misc/nr/.

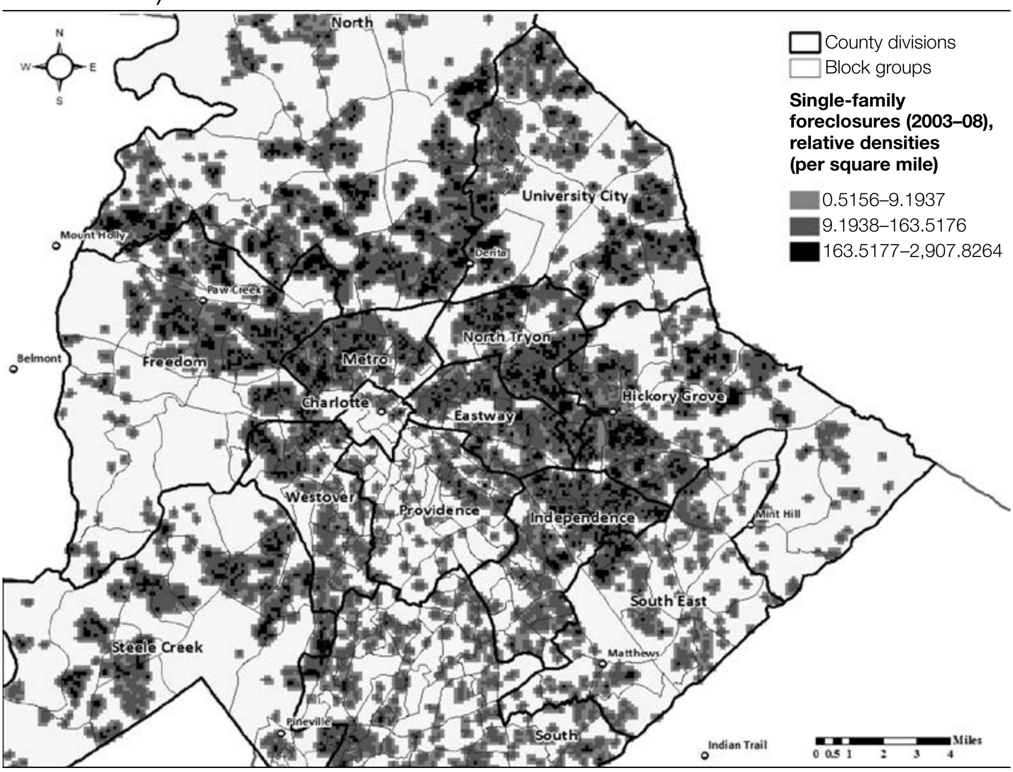
concept posits that events geographically concentrate by spreading from one location to another in a systematic manner. The near repeat analysis shows that properties within short distance intervals—up to 1,320 feet (0.25 miles)—are likely to go into foreclosure within 90 days. These results support the nearest neighbor analysis results in that foreclosures are occurring at the street block level.

Finally, I conducted a kernel density estimation (KDE) analysis to visualize the foreclosure cluster patterns across Charlotte and Mecklenburg County.[9] I overlaid the block group boundaries with the KDE output surface to examine how well foreclosure clusters aligned with the geographic units. Exhibit 8 shows foreclosed properties are highly concentrated within and across the block groups and that even units as small as block groups can still be too large and mask or dissect true local patterns. Nevertheless, the block groups generally capture the spatial extents of foreclosure concentration better than tracts.

The spatial analysis results show foreclosures to be highly concentrated at the micro scale and a contagion effect is more accurately measured with the block group geography.

Exhibit 8

Single-Family Foreclosure Relative Densities, 2003 to 2008 (geometrical interval classification)



[9] I used parameters from a distance analysis that revealed that 8 miles is the threshold at which clustering of foreclosures dissipated. I used a negative exponential function to model the distribution, because construction patterns often have houses that are tightly grouped within small neighborhoods. I used the geometric interval classification scheme to thematically map the density patterns to reveal the core areas of the clusters.

Analytical and Policy Considerations

The examples and results in this article demonstrate the impact Simpson's Paradox has on analysis. My analysis revealed that using a geographic unit larger than a block group in Wilson and Behlendorf (2013) would have compromised the analysis. My findings substantiate the concerns of several authors from the cited research who acknowledged that their results could change using smaller geographic units. Baumer, Wolff, and Arnio (2012), for example, thought that results from using large geographic units were speculative about local conditions and suggested more detailed analyses be conducted within cities. In another example, Kirk and Hyra (2012) recognized that increased crime from foreclosures might exhibit stronger relationships in select neighborhoods because of localized effects. Simpson's Paradox can be mitigated through a number of methods to meet these concerns, such as data normalization, transformation (Wilson, 2011), or optimization (Mu and Wang, 2008), as long as the data are reliable (Sperling, 2012). When these methods cannot be employed, however, identifying the geography that captures an existing spatial effect is the best approach.

The policy consequences of Simpson's Paradox are equally as important. Urban policy often targets places, and as such, the spatial extent of those policies should match the geographic coverage area of the problem to be effective in mitigation. Using the wrong geographic unit could lead to policies that do not fully address the problem. In Wilson and Behlendorf (2013) the use of census tracts would have led to a conclusion of no spatial contagion between foreclosures and that any crime associated with those properties also did not spread into adjacent neighborhoods. The tract model results, then, might have prompted the formulation of ineffective, or less than optimal, policy in containing the spread of foreclosures and any associated crime.

Acknowledgments

The author thanks Brent Mast and Jon Sperling from the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development for providing valuable comments toward improving this article. The author also thanks Brandon Behlendorf from the University of Maryland for his work on the original foreclosure and crime working paper.

Author

Ron Wilson is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development and an adjunct faculty member of the Geographic Information Systems program at the University of Maryland, Baltimore County.

References

Anselin, Luc, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita. 2000. "Spatial Analyses of Crime." In *Criminal Justice 2000*, vol. 4, *Measurement and Analysis of Crime and Justice*, edited by David Duffee. Washington, DC: National Institute of Justice: 213–262.

Arnio, Ashley N., and Eric P. Baumer. 2012. "Demography, Foreclosure, and Crime: Assessing Spatial Heterogeneity in Contemporary Models of Neighborhood Crime Rates," *Demographic Research* 26 (18): 449–488.

Arnio, Ashley N., Eric P. Baumer, and Kevin T. Wolff. 2012. "The Contemporary Foreclosure Crisis and U.S. Crime Rates," *Social Science Research* 41 (6): 1598–1614.

Baumer, Eric P., Kevin T. Wolff, and Ashley N. Arnio. 2012. "A Multicity Neighborhood Analysis of Foreclosure and Crime," *Social Science Quarterly* 93 (3): 577–601.

Coulton, Claudia J., Jill Korbin, Tsui Chan, and Marilyn Su. 2001. "Mapping Residents' Perceptions of Neighborhood Boundaries: A Methodological Note," *American Journal of Community Psychology* 29 (2): 371–383.

Cui, Lin. 2010. Foreclosure, Vacancy and Crime. Working paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1773706 (accessed May 1, 2013).

Ellen, I.G., J. Lacoe, and C.A. Sharygin. 2011. Do Foreclosures Cause Crime? Unpublished manuscript. New York University.

Garcia, Ruth M., Ralph B. Taylor, and Brian A. Lawton. 2007. "The Impacts of Violent Crime and Neighborhood Structure on Trusting Your Neighbors," *Justice Quarterly* 24 (4): 679–704.

Goodstein, Ryan M., and Yan Y. Lee. 2010. Do Foreclosures Increase Crime? Working paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1670842 (accessed May 1, 2013).

Hipp, John R. 2007. "Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder As a Case in Point," *American Sociological Review* 72 (5): 659–680.

Immergluck, Daniel, and Geoff Smith. 2006. "The Impact of Single-Family Mortgage Foreclosures on Neighborhood Crime," *Housing Studies* 21: 851–866.

Jones, Roderick W., and William A. Pridemore. 2012. "The Foreclosure Crisis and Crime: Is Housing-Mortgage Stress Associated With Violent and Property Crime in U.S. Metropolitan Areas?" *Social Science Quarterly* 93 (3): 671–691.

Katz, Charles M., Danielle Wallace, and E.C. Hedberg. 2011. "A Longitudinal Assessment of the Impact of Foreclosure on Neighborhood Crime," *Journal of Research in Crime and Delinquency*. DOI: 10.1177/0022427811431155.

Kirk, David S., and Derek S. Hyra. 2012. "Home Foreclosures and Community Crime: Causal or Spurious Association," *Social Science Quarterly* 93 (3): 648–670.

Macintyre, Sally, Anne Ellaway, and Steven Cummins. 2002. "Place Effects on Health: How Can We Conceptualise, Operationalise and Measure Them?" *Social Science & Medicine* 55: 125–139.

Miller, Henry. 2004. "Tobler's First Law and Spatial Analysis," *Annals of the Association of American Geographers* 94: 284–289.

Mu, Lan, and Fahui Wang. 2008. "A Scale-Space Clustering Method: Mitigating the Effect of Scale in the Analysis of Zone-Based Data," *Annals of the Association of American Geographers* 98 (1): 85–101.

Openshaw, Stan. 1994. *The Modifiable Areal Unit Problem*. London, United Kingdom: Geo Books. Also available at http://qmrg.org.uk/files/2008/11/38-maup-openshaw.pdf.

Rengert, George F., and Brian Lockwood. 2009. "Geographical Units of Analysis and the Analysis of Crime." In *Putting Crime in Its Place: Units of Analysis in Geographic Criminology*, edited by David Weisburd, Wim Bernasco, and Gerben J.N. Bruinsma. New York: Springer-Verlag: 109–122.

Shaw, Clifford, and Henry D. McKay. 1942. *Juvenile Delinquency in Urban Areas*. Chicago: University of Chicago Press.

Sperling, Jon. 2012. "The Tyranny of Census Geography: Small-Area Data and Neighborhood Statistics," *Cityscape* 14 (2): 219–224.

Stucky, Thomas D., John R. Ottensmann, and Seth B. Payton. 2012. "The Effect of Foreclosures on Crime in Indianapolis," *Social Science Quarterly* 93 (3): 602–624.

Taylor, Ralph B. 2012. "Defining Neighborhood in Space and Time," *Cityscape* 14 (2): 225–230.

Wallace, Danielle, E.C. Hedberg, and Charles M. Katz. 2012. "The Impact of Foreclosures on Neighborhood Disorder Before and During the Housing Crisis: Testing the Spiral of Decay," *Social Science Quarterly* 93 (3): 625–647.

Wilson, Ronald E. 2011. "Visualizing Racial Segregation Differently: Exploring the Changing Patterns From the Effect of Underlying Geographic Distributions," *Cityscape* 13 (2): 163–174.

Wilson, Ronald E., and Brandon P. Behlendorf. 2013. The Neighborhood Context of Foreclosures and Crime. Working paper. Washington, DC: U.S. Department of Housing and Urban Development.

Wilson, Ronald E., and Derek J. Paulsen. 2008. "Foreclosures and Crime: A Geographic Perspective," *Geography & Public Safety* 1 (3): 1–2.