Cartographic Guidelines for Geographically Masking the Locations of Confidential Point Data

Michael Leitner
Assistant Professor
Department of Geography and Anthropology
Louisiana State University
mleitne@lsu.edu

Andrew Curtis
Assistant Professor
Department of Geography and Anthropology
Louisiana State University
acurti1@lsu.edu

This research proposes cartographic guidelines for presenting confidential point data on maps. Such guidelines do not currently exist, but are important for governmental agencies that disseminate personal data to the public because these agencies have to balance between the citizens' right to know, and preserving a citizen's right to privacy.

In an experiment, participants compared an original point pattern of confidential crime locations with the same point pattern being geographically masked. Ten different masking methods were tested. The objective was to identify appropriate geographic masking methods that preserve both the confidentiality of individual locations, and the essential visual characteristics of the original point pattern. The empirical testing reported here is a novel approach for identifying various map design principles that would be useful for representing confidential point data on a map.

The results of this research show that only two of the ten masking methods that were tested yield satisfactory solutions. The two masking methods include aggregating point locations at either (1) the midpoint of the street segment or (2) at the closest street intersection. The cartographic guidelines developed from this research suggest a combination of both masking strategies. Future research should focus on the refinement and further testing of these, and other alternative masking methods.

Keywords: Geographic masking, cartographic design, privacy

I. Introduction

Different governmental agencies have long stored information in restricted access databases. The advent of on-line data entry and analysis, and subsequent distribution of data to the public, has created a need for a more rigid set of visualization rules that preserve individual confidentiality. For example, when crime data are disseminated to the public in the form of crime maps via the Internet, law enforcement agencies have to balance between citizens' rights to know the dangers they face in their neighborhoods, while at the same time preserving the confidentiality rights of the victim. Similarly in health data, it is important to know which "risks" pregnant mothers face in particular neighborhoods, while preserving the actual birth outcomes of women living in those neighborhoods.

In this paper we discuss different geographic masking strategies of crime location records that protect the confidentiality of individuals, and at the same time preserve essential visual characteristics of the true, original spatial distribution of those records. Five different global and five different local geographic masking methods (for a total of ten) were tested and compared. The decision to distinguish between global and local methods reflects a recent trend in spatial analysis to develop local analysis tools as extensions of already existing global measurements (Getis and Ord, 1996; Fotheringham, Brunsdon and Charlton, 2002; Anselin, 2003). Armstrong, Rushton and Zimmerman (1999) introduced the term 'geographic masking' into the literature, suggesting similar geographic masking methods that are applied in this research. Whereas their research discusses the influence geographically masked data have on the results of geographically based analyses, the research reported here identifies acceptable design solutions for presenting confidential point data on a map. Acceptable design solutions define any geographic masking method that would preserve as many visual spatial characteristics as possible, while reducing the likelihood of individual identification to an acceptable level. A review of the literature suggests that the methods reported here are the first to utilize empirical perceptual studies to assess methods for presenting confidential point data on maps. This work also continues a long-standing tradition of empirical research in map design as a paradigm for eliciting and formalizing cartographic design knowledge (Leitner, 1997; Leitner and Buttenfield, 2000; Aerts, Clarke and Keuper, 2003).

The subject matter of this research relates to an area of geography and related disciplines that has been receiving a fair amount of attention lately, especially as it pertains to the mapping of disease and crime information (Armstrong, Rushton and Zimmerman, 1999; Wartell and McEwen, 2001; Monmonier, 2002; Leitner and Curtis, 2003). The intention of the research reported here is to develop appropriate guidelines for mapping the location of individual-level data that is considered to be confidential. Such guidelines do not currently exist, but their development becomes increasingly important as more and more governmental agencies disseminate their data on maps via the Internet. To put it differently, this research is concerned with masking locations of individual-level data, rather than the attribute information that could be associated with such locations. Masking strategies for attribute data have already been widely discussed in the literature for some time (Duncan and Pearson, 1991; Cox, 1994; 1996).

"The intention of the research reported here is to develop appropriate guidelines for mapping the location of individual-level data that is considered to be confidential."

Results of this research are of interest to any agency needing to display confidential data at neighborhood levels. As data become more widely available on the WEB, and as the public increasingly realizes the power of local level rather than global level aggregate maps, the need for accurate and easy geographic masking of confidential data will be of utmost importance to all governmental agencies. According to a website updated by the Mapping and Analysis for Public Safety (MAPS) program (sponsored by NIJ), about 50 local law enforcement agencies in the US now provide online data/maps (see http://www.ojp.usdoj.gov/nij/maps/weblinks.html). Access to this type of data will certainly continue to increase.

II. Methodology

A. Data

All incident locations displayed in the test maps were the residences of homicide victims in the city of Baton Rouge between 1991 and 1997. There were a total of 301 homicide victims, of which 285 were successfully address-matched (95% success rate) to the Baton Rouge street network. The street network and the census tract boundaries were from the Census 2000 TIGER/Line Data from the U.S. Bureau of the Census. All data were projected to the Universal Transverse Mercator coordinate system, North American Datum 83 (NAD83), Zone 15 North.

B. Preparation of test maps

Only a subset of all incident locations was selected for inclusion in the test maps. This subset of incident locations was displayed in yellow on top of a blue background. In one-third of the test maps, no additional information was shown. In the second one-third, census tract boundaries were added to the test maps in black. The remaining one-third of the test maps included the complete street network in dark blue.

Incident locations were geographically masked at either the global or local level. Global geographic masking means that every single incident location was spatially displaced by the same exact amount (Figure 1). For three of the five local geographic masking methods, a regular grid was first superimposed over the selected study area. Each grid cell was a square measuring 500 meters per side. Incident locations falling into the same grid cell were then spatially displaced by the same exact amount, but this displacement vector changed randomly between grid cells (Figure 2). The regular grid was used in the preparation of the test maps but not shown on the maps that were included in the experiment.

C. Global geographic masking methods

Figure 1 compares the original incident locations (Figure 1-A) with the same incident locations after being geographically masked by five different global masking methods (Figure 1-B through 1-F). For privacy concerns only the blue background was used. Three incident locations were labeled to show the effect that each global masking method had on the incident locations.

"Incident locations were geographically masked at either the global or local level."

In Figure 1-B the original incident locations were flipped about the horizontal central axis of the map, and subsequently placed on top of the closest street segment. In Figure 1-C all incident locations were flipped about the vertical central axis of the map and then moved on top of the closest street segment. In Figure 1-D incident locations were flipped about both the horizontal and the vertical central axes of the map and then placed on top of its closest street segment. In Figure 1-E all incident locations were rotated around the map center either by 60º to the right or by 120º to the left (Figure 1-F) and subsequently placed on top of the closest street segment.

D. Local geographic masking methods

Figure 2 displays the original incident locations (Figure 2-A) together with the same incident locations after being geographically masked by five different local masking methods (Figure 2-B through 2-F). To protect the privacy of the individual, only the blue background was used in all maps. Four sample locations were selected and labeled so their locations could be traced between the original and the five locally masked maps. Note that Figure 2 shows only a small portion of each test map that was used in the experiment.

The first local masking method aggregated incident locations at the midpoint of their street segments (Figure 2-B). A street segment was defined as the portion of a street between two adjacent street intersections. In Figure 2-C incident locations were aggregated to their closest street intersection, which was defined as the intersection of three or more streets. The next three local masking methods were based on a regular grid. In the first instance, incident locations were flipped either about the vertical, horizontal, or both central axes of each grid cell and then moved on top of

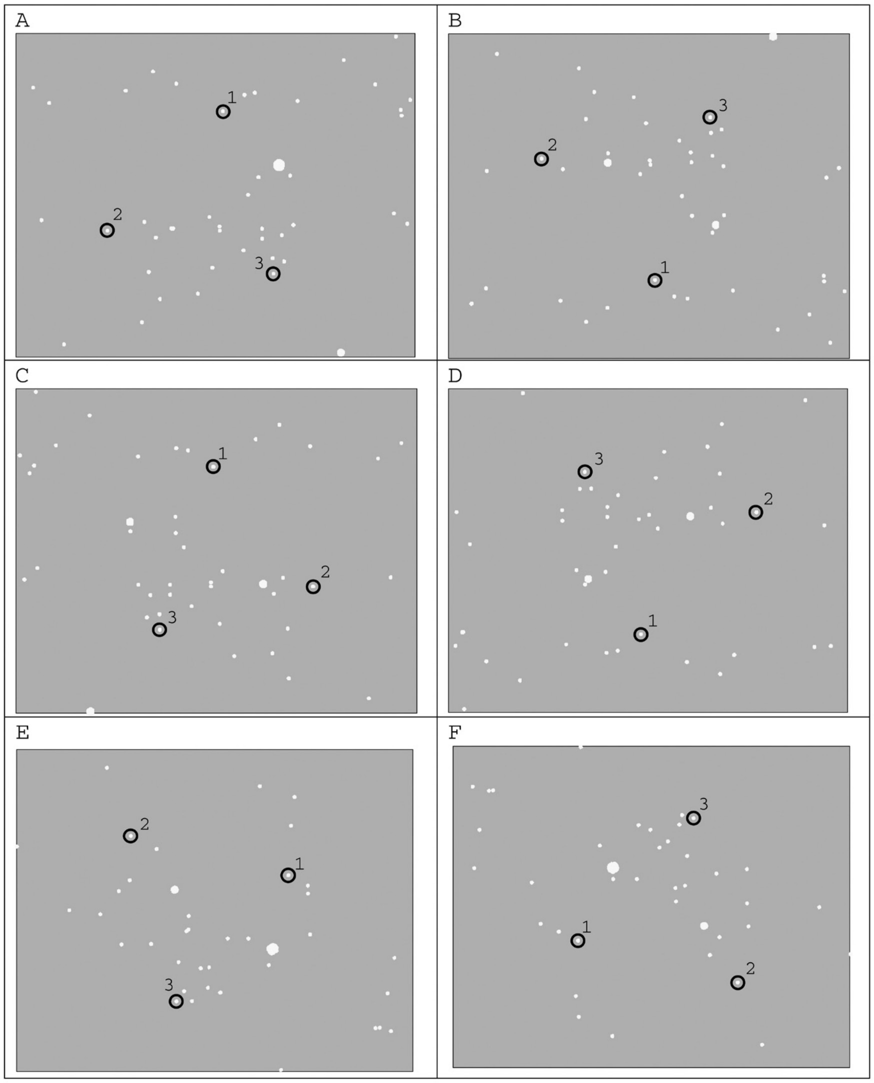


Figure 1. Global geographic masking methods used in this experiment.
A) Original incident locations, B) flipping about horizontal central axis of the map, C) flipping about vertical central axis of the map, D) flipping about both central axes of the map, E) rotating around the map center by 60º to the right, F) rotating around the map center by 120º to the left.
(See page 83 for a reduced color version)

its closest street segment. The type of flipping changed randomly between the cells of the regular grid (Figure 2-E). For example, the incident location labeled as #1 was flipped vertically; the other three incident locations (#2, #3 and #4) were each flipped horizontally. In Figure 2-D incident locations were first rotated by some random degrees around the center of each grid cell and then placed on top of the closest street segment. The incident location #1 was rotated by 120º to the left; #2 by 60º to the right; #3 by 240º to the right; and #4 by 120º to the right. In Figure 2-F incident locations were first translated by some random distances and then moved on top of their closest street segment. The incident location labeled as #1 was translated 150 meters in x- and 350 meters in y-direction from the lower left corner of its grid cell; #2 was translated 300 meters in x and 400 meters in y; #3, 450 meters in x and 50 meters in y; and #4, 300 meters in x and 200 meters in y.

The size of the grid cells, the angles of rotation and the translation distances were chosen arbitrarily when masking point patterns. Clearly, different choices might have yielded different results. For example, smaller

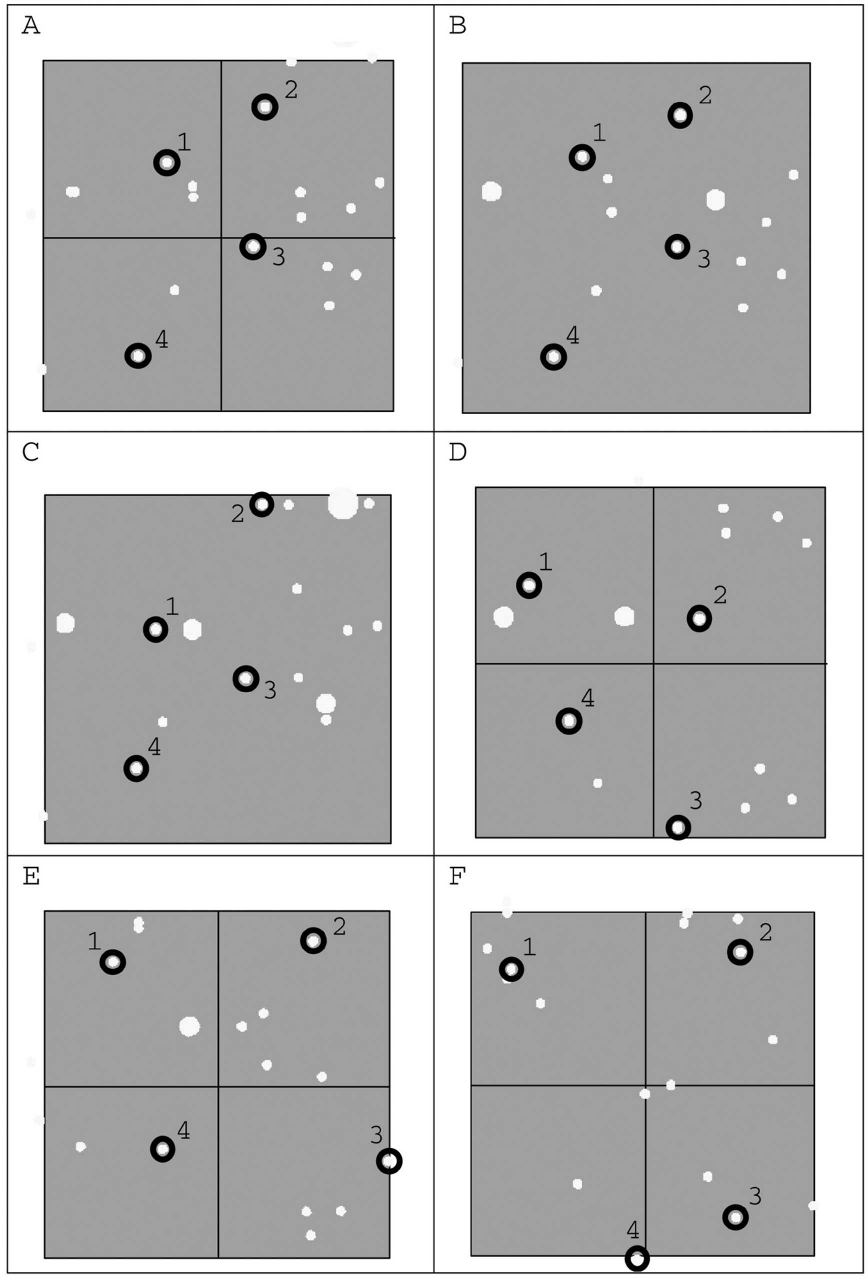


Figure 2. Local geographic masking methods used in the experiment (only a portion of the entire test map is shown).
A) Original incident locations, B) spatial aggregation at the midpoint of the street segment, C) spatial aggregation at the closest street intersection, D) flipping randomly either about the vertical, horizontal or both central axes of each grid cell, E) rotating by some random degree around the center of each grid cell, F) translating by some random distance.
(See page 83 for a reduced color version)

rotation angles and shorter translation distances in Figures 2-E and 2-F would have resulted in shorter point movements. Consequently, these two masked point patterns would have been more similar to those patterns exhibited in Figures 2-B and 2-C. In addition, smaller cell sizes would have shortened point movements and vice versa. Investigating the impact differently masked point patterns have on people's visual perception is new — the selection of all masking methods used in this study, including the translation distances, rotation angles and the size of the grid cells, should be understood as an exercise in exploratory analysis. It is hoped that results of this research will create a base line for the development of guidelines for geographic masking.

III. Procedure

Two experiments were conducted, one testing 30 map pairs, the other testing 34 map pairs. One map of each pair always showed all incidents in their correct location. In the second map, all incident locations were geographically masked, either locally or globally. All five local and all five global geographic masking methods (see Section II) were tested in each experiment. The incidents in each map pair were displayed in yellow on top of a blue background (Figure 4). The experiment with 30 map pairs included each of the ten masked point patterns for each of the three different backgrounds (i.e., no background, census tract boundaries, and street network information). For these 30 map pairs, participants were asked to identify hot spot areas in the masked point pattern. The experiment with 34 map pairs included the same 30 map pairs tested in the first experiment—again hot spot areas were marked in the masked point pattern. However, for the additional four map pairs participants were asked to identify hot spots in the original or unmasked point pattern, rather than in the masked point pattern. For all map pairs in both experiments, participants did not know which point pattern was masked and which one was unmasked. The order of presentation of the map pairs in each experiment was random.

". . . all incident locations were geographically masked, either locally or globally."

Each slide in this presentation shows two point pattern maps that are displayed next to each other.

Your first task is to compare the two point patterns and decide how similar or different they are.

Please make this decision by choosing a whole number between **1** and **7.** The number **1** means that the two point patterns are **VERY SIMILAR** and the number **7** means that they are **VERY DIFFERENT.**

Your second task is to identify areas within one of the two point pattern maps that show **a high concentration of points**. Mark those areas with a pen or pencil directly on the hard copy maps provided.

There is no time limit to complete this experiment.

Thank you for your participation!

Figure 3. First slide of the experiment, including the instructions.

"For all map pairs in both experiments, participants did not know which point pattern was masked and which one was unmasked."

Each experiment was put together as a power point presentation. In addition, all participants received a color printout (hard copy) of the entire presentation. The first slide of the presentation included the instructions (see Figure 3); all other slides included the map pairs; one map pair per each slide (Figure 4). Instructions were briefly repeated at the bottom of each slide showing a map pair. Overall, 82 participants completed the experiment. 44 of those participants were students from computer cartography, geographic information systems, and methods of spatial analysis classes at Louisiana State University in Baton Rouge, LA. The remaining

38 participants were students from two introductory spatial analysis and one cartography and visualization classes from the School of Geoinformation, University of Applied Science located in Villach, Austria.

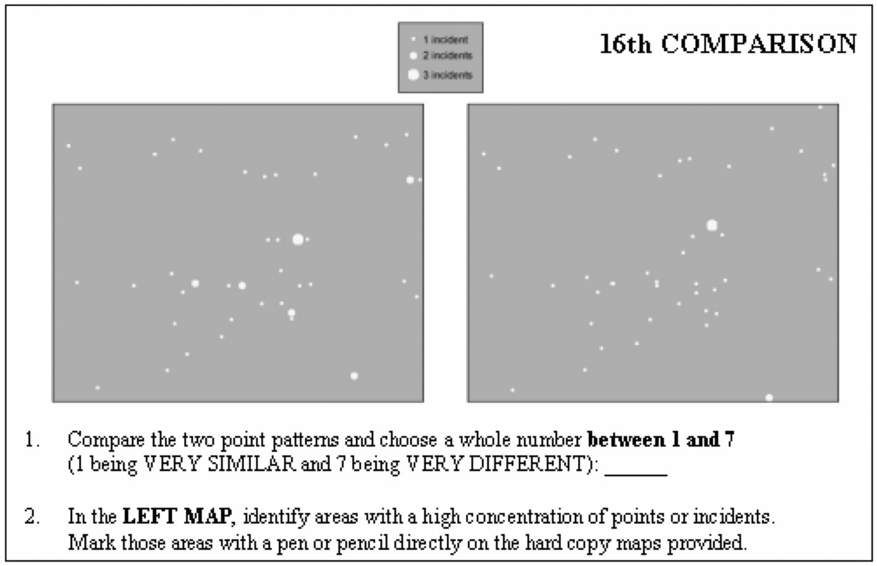


Figure 4. Example of a pair of test maps included in the experiment. All incident locations in the left map are locally masked by spatially aggregating them to their closest street intersection. All incidents in the right map are shown with their true, unmasked location. (See page 83 for color version)

Experiments were conducted during class time, lasting between 13 and 26 minutes. For each map pair, participants were asked to complete two tasks. The first task was to compare the two point pattern maps and decide if the two patterns were similar or different. Participants were asked to rank the two point patterns as "very similar" (rank value of 1) to "very different" (rank value of 7). This number was recorded on the hardcopy maps. The second task involved identifying areas within one of the two point pattern maps (in most instances this was the map for which the incident locations were geographically masked, but the participant did not know that) that showed a high concentration of incidents. Test participants were asked to mark those areas, if they thought they existed, with a pen or pencil directly onto the hardcopy maps.

"One objective of this study is to investigate how much each geographic masking technique changes the original pattern of incident locations."

IV. Results

One objective of this study is to investigate how much each geographic masking technique changes the original pattern of incident locations. This objective is addressed in two ways, first, by comparing the change in the overall point pattern and secondly, by comparing the change in the number and location of hot spots. Test participant's responses to how similar the (overall) original and the masked pattern were are summarized in two frequency tables. Table 1 includes the results for the five global, Table 2 for the five local masking techniques.

The results in Table 1 and Table 2 show that the selection of a masking technique is very important. The differences between the masking techniques can be seen in the differences between the relative frequency distributions. Additionally, the mean rank for each masking technique was calculated. The closer the mean rank is to one, the more similar the original and the masked point pattern were perceived. The closer the mean

Similarity between original and masked point pattern	Flipping about vertical central axis of the map	Rotating around the map center by 60º to the right	Flipping about horizontal central axis of the map	Flipping about both central axes of the map	Rotating around the map center by 120º to the left
1 (very similar)	3.7	1.1	2.4	2.8	2.8
2	19.5	4.9	7.3	6.5	2.8
3	22.4	13.4	10.6	8.5	4.1
4	23.2	18.7	13.5	11.8	10.2
5	17.1	23.3	18.0	24.4	14.6
6	7.7	22.6	27.3	25.2	28.0
7 (very different)	6.5	15.9	20.8	20.7	37.4
Mean Rank	**3.80**	**4.90**	**5.02**	**5.07**	**5.65**
Total number of map comparisons	246	283	245	246	246

Table 1. Comparison between the original and five different global geographically masked point patterns. All entries in the Table are percentages of map comparisons falling into one of seven categories ranging from 1=very similar to 7=very different. The percentages are calculated from all map comparisons irrespective of background information (i.e., no background, census tract and street network information).

rank is to seven, the more different the two point patterns were perceived. A mean rank lower than four indicates that the two point patterns were perceived to be more similar than different from each other. Accordingly, mean ranks above four indicate that the two point patterns were perceived to be more different than similar. In both Tables, the total number of map comparisons for each masking technique varies between 245 and 283. This can be explained by participants missing the answer to this question and/or by the difference in the total number of test maps used between the different groups of test participants.

Similarity between original and masked point pattern	Spatial aggregation at midpoint of street segment	Spatial aggregation at street intersection	Rotating by some random degree around the center of each grid cell	Flipping randomly either about the vertical, horizontal or both central axes of each grid cell	Translating by some random distance
1 (very similar)	43.3	15.9	4.5	2.0	0.8
2	42.0	45.5	15.9	12.2	4.1
3	9.4	24.0	31.3	17.9	14.2
4	3.3	10.6	20.3	17.9	17.9
5	0.8	2.0	17.5	24.4	22.0
6	0.8	2.0	8.5	18.7	21.5
7 (very different)	0.4	0.0	2.0	6.9	19.5
Mean Rank	**1.80**	**2.43**	**3.64**	**4.34**	**4.99**
Total number of map comparisons	245	246	246	246	246

Table 2. Comparison between the original and five different local geographically masked point patterns. All entries in the Table are percentages of map comparisons falling into one of seven categories ranging from 1=very similar to 7=very different. The percentages are calculated from all map comparisons irrespective of background information (i.e., no background, census tract and street network information).

V. Analysis

Overall, the mean rank varies greatly from a low of 1.80 to a high of 5.65. The lowest mean rank (1.80) is calculated for the masking technique that aggregates point locations at the midpoint of its street segment, followed by the masking technique that aggregates point locations at their closest street intersection (mean rank = 2.43). These two local masking techniques are clearly below a mean rank of four, which means that they were perceived rather similar (than different) to the original point pattern. Two additional masking methods yield a mean rank less than four, including the local masking method 'rotating by some random degree around the center of each grid cell' (mean rank = 3.64) and the global masking method 'flipping about the vertical central axis of the map' (mean rank = 3.80) (Tables 1 and 2).

Altogether, three local and one global masking method are the only ones below a mean rank of four. This means that they are the only masking methods of all ten methods tested for which the original and the masked point pattern were perceived to be more similar than different from each other. This further means that they are the only four masking methods identified in this research that comply with one important objective of an appropriate masking technique for the location of confidential data, namely to preserve as much as possible the visual characteristics between the original and the masked point pattern. Accordingly, the remaining four global and the other two local masking methods are inappropriate and should not be used.

". . . three local and one global masking method are the only ones below a mean rank of four. . . .they are the only masking methods of all ten methods tested for which the original and the masked point pattern were perceived to be more similar than different from each other."

If there were a choice between the three local and the one global geographic masking method with a mean rank below four, which one should be chosen? The obvious choice would be the masking method with the lowest mean rank. In order to find out, if 1.80 is a statistically significant lower mean rank than any one of the other three mean ranks (2.43, 3.64, or 3.80), a Mann-Whitney U[1] test was applied. Analysis shows that a mean rank of 1.80 is indeed significantly lower than a mean rank of 2.43 (z-test statistic = -7.530, p-value < 0.001). From this result it follows that a mean rank of 1.80 is also significantly lower than a mean rank of 3.64 and 3.80. Additional analyses show that a mean rank of 2.43 is lower than a mean rank of 3.64 (z-test statistic = -9.992, p-value < 0.001), and that there is no difference between a mean rank of 3.64 and 3.80 (z-test statistic = -0.906, p-value = 0.365). Consequently, when there is a choice between the four different masking methods with a mean rank below four, then the first choice would be to aggregate points at the midpoint of its street segment. The second choice would be to aggregate points at their closest street intersection and the third choice would be to either rotating incident locations by some random degrees around the center of each grid cell or flipping all incident locations about its vertical central axis of the map.

The order of mean ranks measuring the similarity between an original and different masked point patterns is purely based on perceptual (subjective) interpretations of test participants. An interesting question to ask is if the order would remain the same when the similarity between two point patterns is measured with a statistical (objective) method. To answer this question the straight-line distances from each incident location in the original point pattern were measured to its nearest neighbor in each of the four masked point patterns with a mean rank lower than four (see Tables 1 and 2) and mean distances calculated. A shorter mean distance between an original and its masked point pattern would be interpreted as a higher similarity, whereas a longer mean distance would indicate a lower similarity. For this research the shortest mean nearest neighbor distance of 38.04 meters was calculated between the original point pattern and the masked point pattern that aggregates points at the midpoint of its street segment. The second mean nearest neighbor distance of 56.40 meters was calculated between the original point pattern and the masked point pattern that aggregates points at their closest street intersection. The third shortest mean nearest neighbor distance of 131.15 meters was calculated between the original point pattern and the masked point pattern that rotates incident locations by some random degrees around the center of each grid cell. The longest mean nearest neighbor distance of 158.00 meters was derived between the original and the masked point pattern that flips incident locations about its vertical central axis of the map. These results show that the shortest mean distance was calculated between the same original and masked point patterns that also yielded the lowest mean rank. Similarly, the second shortest mean distance was calculated between the same two point patterns that also yielded the second lowest mean rank. The same is true for the third and fourth shortest mean distances. In general, the results show that no differences were found whether the similarity (between one original and four different masked point patterns) was measured by a perceptual (subjective) or a statistical (objective) technique. This agreement is important, as it supports the robustness of the experimental design and perceptual framework used to collect data for this research.

Clearly, the mean ranks for the three methods that mask their incident locations locally within each grid cell (see Figure 2) depend on the chosen grid cell size. It is expected that the smaller the cell size, the lower the mean rank would be. Therefore, it is quite possible that with a small enough cell size, all three local masking methods would yield a mean rank lower than four. Additional experiments should therefore be conducted to explore the relationship between cell size and perceived similarity between an original and a geographically masked point pattern. The goal would be to possibly identify a small enough cell size resulting in a mean rank lower than four.

"Additional experiments should therefore be conducted to explore the relationship between cell size and perceived similarity between an original and a geographically masked point pattern."

VI. Analyzing the influence of different base map information

The previous section identified four masking methods that were appropriate for the display of confidential point locations. This section analyzes if different base map information might influence participants' perception of the similarity between the unmasked and any of these four masked point patterns. Results in Table 3 reveal that the mean rank changes somewhat across the three base maps for each of the four masking methods. As a reminder, the closer the mean rank is to one, the more similar the masked and unmasked point pattern is. A value above four indicates that the masked point pattern is more different (than similar) to the original, unmasked point pattern. All mean ranks are below four, with the exception of the masking method 'flipping about its vertical central axis of the map' with the street network as the base map. Three of the four masking methods yield the lowest mean ranks when census tracts are used as the base map, whereas, two of the four masking methods show the highest mean ranks when the base map consists of the street network. Due to the small sample size, one has to be cautious to not draw general conclusions from these results. Again, additional research is warranted.

The results further show that for each masking method, the variability in the mean ranks across the three base maps seems to increase with higher mean ranks. Are these variations statistically significant for any of the four masking methods?

	Spatial aggregation at midpoint of street segment			Spatial aggregation at street intersection		
Similar / Different	**No base map**	**Census tract boundaries**	**Street network**	**No base map**	**Census tract boundaries**	**Street network**
1 (very similar)	36.6	45.1	48.1	18.3	20.7	8.5
2	41.5	45.1	39.5	43.9	39.0	53.7
3	12.2	7.3	8.6	23.2	29.3	19.5
4	4.9	2.4	2.5	11.0	9.8	11.0
5	2.4	0.0	0.0	1.2	1.2	3.7
6	1.2	0.0	1.2	2.4	0.0	3.7
7 (very different)	1.2	0.0	0.0	0.0	0.0	0.0
Mean Rank	**2.04**	**1.67**	**1.70**	**2.40**	**2.32**	**2.59**
Total number of map comparisons	82	82	81	82	82	82

	Rotating by some random degree around the center of each grid cell			Flipping about vertical central axis of the map		
Similar / Different	**No base map**	**Census tract boundaries**	**Street network**	**No base map**	**Census tract boundaries**	**Street network**
1 (very similar)	9.8	2.4	1.2	2.4	3.7	4.9
2	29.3	7.3	11.0	14.6	34.1	9.8
3	28.0	30.5	35.4	26.8	12.2	28.0
4	14.6	28.0	18.3	23.2	24.4	22.0
5	12.2	19.5	20.7	20.7	14.6	15.9
6	4.9	11.0	9.8	4.9	9.8	8.5
7 (very different)	1.2	1.2	3.7	7.3	1.2	11.0
Mean Rank	**3.10**	**3.93**	**3.90**	**3.89**	**3.46**	**4.04**
Total number of map comparisons	82	82	82	82	82	82

Table 3. Comparison between the original and the masked point pattern for four different geographic masking methods. All entries are percentages of test participants falling into one of seven categories ranging from 1=very similar to 7=very different.

To this end, a Kruskal-Wallis H test[2] was applied to explore if test participants' responses varied significantly due to different base map information. This test was carried out for the four masking methods with a mean rank below four. Results revealed that three of the four masking methods did not show statistically significant differences across the three different base maps. When incidence locations are aggregated at the midpoint of its street segment then the Kruskal-Wallis H test calculated a chi-square statistic of 4.501, with a p-value of 0.105 and two degrees of freedom. When incidence locations are aggregated at the closest street intersection, then the results yielded a chi-square statistic of 1.361, with a p-value of 0.506 and two degrees of freedom, respectively. The corresponding test statistics and probabilities for 'rotating by some random degree around the center of each grid cell' and 'flipping about its vertical central axis of the map' are a chi-square statistic of 20.228 with a p-value of 0.000 and a chi-square statistic of 5.743 with a p-value of 0.057, respectively.

Based on a significance level of 0.01, these results mean that for three of the four masking methods analyzed, it would not matter if a base map includes political-administrative boundaries, or the street network, or no base map at all. The only significant chi-square statistic was calculated for the local masking method that rotates incident locations by some random degree around the center of each grid cell.

VII. Analyzing hot spots for different masking methods and different base maps

Areas within each point pattern that show a high concentration of incident locations are usually of great interest and worthwhile to be studied further. Any good and appropriate masking method needs to preserve the number, locations, sizes and shapes of these 'hot spot' areas. For this reason, participants were asked to identify high concentrations of incident locations in all geographically masked maps and in the original, unmasked maps. Hot spot areas for those four masked point patterns that were perceived to be more similar than different from the original unmasked point pattern are displayed in Figure 5. These hot spots are compared with hot spots from the unmasked point pattern for each of the three different base maps. Base map information was not included in Figure 5, so that hot spots would be clearly visible.

"Areas within each point pattern that show a high concentration of incident locations are usually of great interest and worthwhile to be studied further."

In general, the results in Figure 5 clearly show that the size, shape, location and number of hot spots change little *across* the different types of base maps. However, differences can be observed *between* some of the masking methods. Specifically, there seems to be much agreement in the size, shape, location and number of hot spots between the original, unmasked point pattern, and the two masked point patterns that aggregate their incident locations either at the midpoint of its street segment or at their closest street intersection. This is an important result and it makes these two local masking methods very useful and appropriate tools to visualize the location of confidential data. A further observation is that of a hierarchical arrangement of differently sized hot spots. There are two small (local) hot spots to the right of the center of each map. These two are combined into one medium-sized hot spot. Then there is a larger (regional) hot spot apparent across the lower portion of each map that encloses the two small and the one medium-sized hot spot.

However, differences in the size, shape, location and number of hot spots are visible when comparing the original point patterns with the other two masked point patterns. For example, when the point pattern is masked by rotating point locations by some random degrees around the center of each grid cell then a distinct, local hot spot appears in the upper half of the map, just right of the center. This particular hot spot is not visible in the original and in any of the other masked point patterns. This points to a particular danger of using an inappropriate masking method, namely the likelihood of an incident cluster to appear in a neighborhood of a city, where such a cluster does not exist in reality. If this information is distributed to the public, consequences may include false perceptions regarding the nature of the distribution and/or unfair informal redlining methods employed by some insurance and banking companies. For

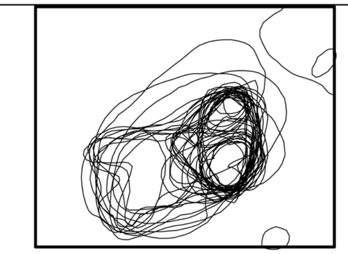
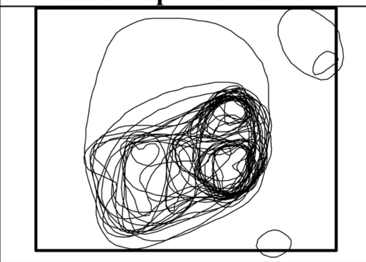
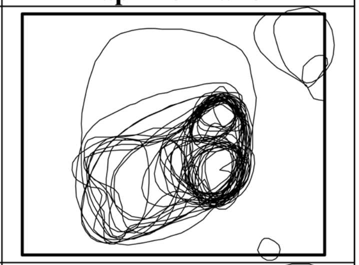
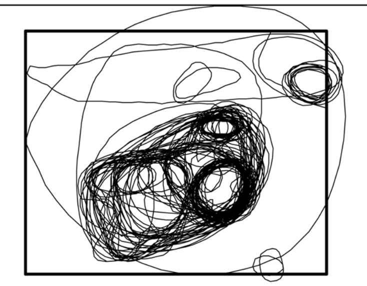
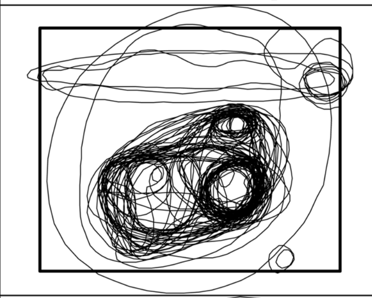
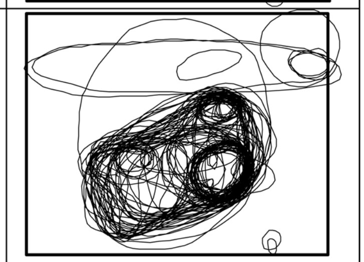
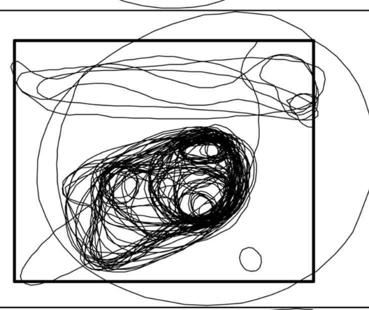
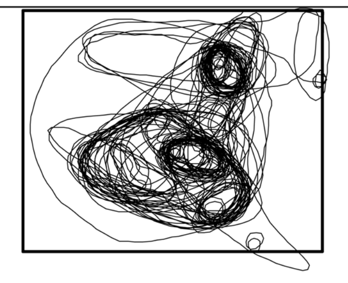
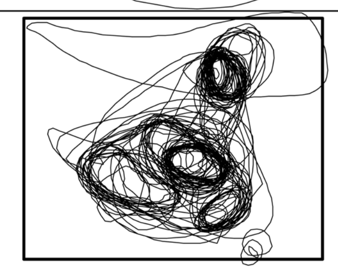
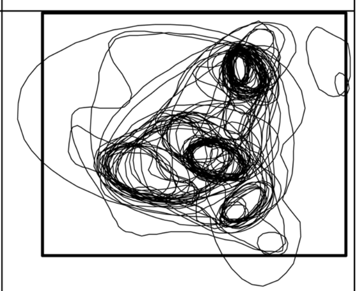
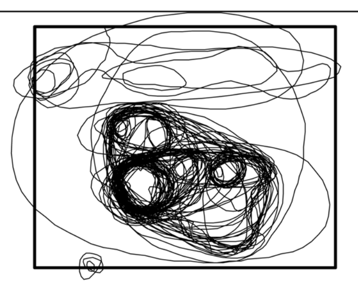
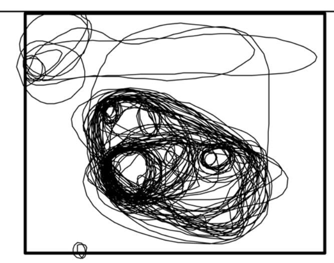
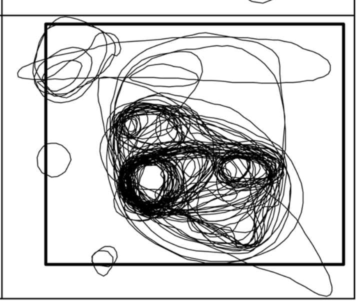
	No base map information	Census tract boundaries as base map information	Street network as base map information
Original, geographically unmasked point pattern			
Geographically masked by aggregating point locations at the midpoint of its street segment			
Geographically masked by aggregating point locations at their closest street intersection			
Geographically masked by rotating point locations by some random degree around the center of each grid cell			
Geographically masked by flipping point locations about the vertical central axis of the map			

Figure 5. Comparing hot spots between the original and four different geographically masked point patterns for three different base maps. The base map information is not included in any of the maps.

the global masking method, hot spots, when compared to the original, unmasked point pattern, are clearly flipped about the vertical central axis of the map. Again, hot spots appear in parts of the city, where they do not exist in reality. For this reason, the two masking methods 'rotating by some random degree around the center of each grid cell' and 'flipping about the vertical central axis of the map' are inappropriate and should not be used.

Figure 5 provides a visual comparison of hot spots between the different test maps. A tabular comparison of the same information is shown in Table 4, in which the number of hot spots (none, one, two, more than two) for the masked and unmasked point patterns for all three base maps is listed. Relative percentages are shown in square parenthesis. For the category 'more than two hot spots', the number of hot spots (3, 4, 5, etc.) is shown in round parenthesis. For this category, the relative percentages are calculated only for the total number of maps with more than two hot spots.

	Number of hot spots	No base map information	Census tract boundaries as base map information	Street network as base map information
Original, geographically unmasked point pattern	**No hot spot**	1 [2.7%]	0 [0.0%]	2 [5.4%]
	One hot spot	25 [67.6%]	26 [70.3%]	21 [56.8%]
	Two hot spots	10 [27.0%]	7 [18.9%]	12 [32.4%]
	More than two hot spots	1 (3) [2.7%]	3 (3), 1 (4) [10.8%]	2 (3) [5.4%]
	TOTAL MAPS	*37*	*37*	*37*
Geographically masked by aggregating point locations at the midpoint of its street segment	**No hot spot**	3 [3.7%]	1 [1.2%]	1 [1.2%]
	One hot spot	49 [59.8%]	55 [67.1%]	60 [73.2%]
	Two hot spots	19 [23.2%]	19 [23.2%]	11 [13.4%]
	More than two hot spots	5 (3), 5 (4), 1 (5) [13.4%]	3 (3), 4 (4) [8.5%]	8 (3), 2 (4) [12.2%]
	TOTAL MAPS	*82*	*82*	*81*
Geographically masked by aggregating point locations at their closest street intersection	**No hot spot**	3 [3.7%]	4 [4.9%]	10 [12.2%]
	One hot spot	50 [61.0%]	57 [69.5%]	50 [61.0%]
	Two hot spots	21 [25.6%]	14 [17.1%]	14 [17.1%]
	More than two hot spots	5 (3), 2 (4), 1 (5) [9.8%]	7 (3) [8.5%]	7 (3), 1 (4) [9.8%]
	TOTAL MAPS	*82*	*82*	*82*
Geographically masked by rotating point locations by some random degree around the center of each grid cell	**No hot spot**	9 [11.0%]	9 [11.0%]	1 [11.0%]
	One hot spot	30 [36.6%]	36 [43.9%]	42 [51.2%]
	Two hot spots	28 [34.1%]	17 [20.7%]	17 [20.7%]
	More than two hot spots	11 (3), 2 (4), 2 (6) [18.3%]	3 (16), 4 (4) [24.4%]	7 (3), 7 (4) [17.1%]
	TOTAL MAPS	*82*	*82*	*82*
Geographically masked by flipping point locations about vertical central axis of the map	**No hot spot**	2 [2.4%]	3 [3.7%]	10 [12.2%]
	One hot spot	51 [62.1%]	55 [67.1%]	44 [53.7%]
	Two hot spots	16 [19.5%]	15 [18.3%]	17 [20.7%]
	More than two hot spots	9 (3), 2 (4), 2 (5) [19.5%]	5 (3), 2 (4), 1 (5), 1 (6) [11.0%]	9 (3), 1 (4), 5 (1) [13.4%]
	TOTAL MAPS	*82*	*82*	*82*

Table 4. Comparing the number of hot spots between the original and four geographically masked point patterns for three different base maps.

In most instances, the majority of all test participants identified exactly one hot spot in each map, independent of the base map information and whether the point pattern was masked or not. Two hot spots were identified by approximately one-quarter of all test participants. Relatively few participants identified no hot spots, or identified more than two hot spots. The highest number of hot spots identified was six (Table 4).

VIII. Preserving the privacy of confidential incidence locations

The following final analysis investigates to what extent one might be able to identify an individual incident location after this location has been geographically masked. Recall that each incident location identifies the street address where a homicide victim resided. Consequently, this section discusses to what extent the privacy of the confidential location of an individual residence would be preserved if that location has been geographically masked.

To answer this question, a sample of fourteen different street addresses were randomly selected from a total of 48 addresses used in this research. All addresses are located in a residential neighborhood in Baton Rouge with mostly single-family houses. Churches, schools, office buildings, etc. are interspersed between the residences. Each of the fourteen residences associated with the street address from the sample was visited, and the number of residences on either side of the street segment that included the residence of interest was counted. Accordingly, all residences that were located on either side of all street segments of the street intersection to which the residence of interest was closest to were also counted. The results in Table 5 show that the number of residences along either side of a street segment ranges from a minimum of two to a maximum count of 29 residences. If the number of residences along the street segments with a common intersection are counted then the minimum number is 15, the maximum is 58. Both minimum and maximum numbers include the street address where a homicide victim resided.

". . . to what extent one might be able to identify an individual incident location after this location has been geographically masked."

The question that needs to be addressed now is: What is the minimum number of all residences, including the residence of interest, so that the privacy of an individual residing at this address is not compromised? For this research the confidentiality rules used by the U.S. Census Bureau of how to protect the privacy of individuals is followed. Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's or business' data can be identified. The U.S. Census Bureau uses different disclosure limitation procedures to protect the confidentiality of data, including suppression, data swapping and protection of micro data files (see http://factfinder.census.gov). Among the methods used by the Census Bureau, suppression is directly applicable within the context of this research. The U.S. Census Bureau defines suppression as "a method of disclosure limitation used to protect individuals' confidentiality by not showing (suppressing) the cell values in tables of aggregate data for cases where only a few individuals or businesses are represented." For example, cell values for up to three individuals or businesses were suppressed, but a cell value of seven was not. Accordingly, the U.S. Census Bureau must use a cut-off value for data suppression that lies somewhere between four and six.

If the same guidelines from the U.S. Census Bureau were applied to this research then four of the fourteen residences (#3, #6, #8 and #9 from Table 5) for the masking method 'aggregating incident locations at the midpoint of its street segment' would have been suppressed. In other words, such geographically masked locations should not be displayed in a map,

Confidential location of individual residences, numbered consecutively	Number of residences on either side of the street segment containing the confidential location	Number of residences on either side of all street segments with a common interesection. One of these street segments includes the confidential location
1	22	50
2	29	52
3	4	20
4	12	29
5	9	17
6	2	25
7	28	54
8	6	18
9	2	24
10	9	15
11	25	26
12	27	58
13	24	34
14	29	34
TOTAL	228	456
MEAN	16.3	32.6
MINIMUM	2	15
MAXIMUM	29	58

Table 5. Comparing the number of residential buildings along one street segment or along all street segments with a common intersection. The location of one of the residential buildings is confident.

because the privacy of individuals living at these residences would not be guaranteed. In cases like this, one solution would be to display each of these four residence locations at their closest intersection. By doing so the confidentiality of the individuals living at these residences would be protected.

IX. Summary

This research suggests that an appropriate masking method is one that combines the two strategies of aggregating confidential incident locations at either 'the midpoint of their street segment' or 'at their closest street intersection'. More specifically, the first strategy should be applied when a total of seven or more incident locations ('background population') that also includes the confidential location(s), can be counted on either side of the street segment. When that number falls below seven, confidential incident locations should be aggregated to their closest street intersection. Seven is the cut-off value for data suppression used by the U.S. Census Bureau to protect the confidentiality of its published data.

Based on the above discussion, the following general guidelines for geographically masking any location that contains confidential data are proposed. These guidelines lay the groundwork for the development of mapping and/or GIS modules that enable confidential data to be appropriately masked before released to the public.

"This research suggests that an appropriate masking method is one that combines the two strategies of aggregating confidential incident locations at either 'the midpoint of their street segment' or 'at their closest street intersection'."

1. Aggregate each confidential location (the address of a residence) to the midpoint of its street segment.

2. Count the number of all residences (addresses) on either side of the street segment that also included the confidential location. This count defines the 'background population' and includes the confidential location.

3. For street segments where this count is less than seven, move the confidential location from the midpoint of its street segment to its closest intersection. If seven or more residences are counted on either side of the street segment, do not move the confidential location.

NOTES

[1]The Mann-Whitney U test is the nonparametric equivalent to the t test and tests whether two independent samples are from the same population. It requires an ordinal level of measurement. U is the number of times a value in the first group precedes a value in the second group, when values are sorted in ascending order (SPSS Inc., 2001).

[2]The Kruskal-Wallis H test is the nonparametric equivalent to the one-way ANOVA. It tests whether several independent samples are from the same population and requires an ordinal level of measurement (SPSS Inc., 2001).

ACKNOWLEDGEMENTS

The authors would like to thank the Baton Rouge Police Department for providing the crime data used in this research. The authors would also like to thank all students from the computer cartography, methods of spatial analysis and geographic information system classes from Louisiana State University in Baton Rouge, LA who participated in this experiment. Additionally, we would like to thank all students from two different introductory spatial analysis and one cartography and visualization classes from the School of Geoinformation, University of Applied Science located in Villach, Austria. Finally, we would like to thank the reviewers of the original manuscript for their valuable and insightful comments and Monika Arthold for her support with this research.

REFERENCES

Aerts, J.C.J.H., Clarke, K.C. and Keuper, A.D., 2003. Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30(3):249-261.

Anselin, L., 2003. *GeoDa™ 0.9 User's Guide*, Spatial Analysis Laboratory and Center for Spatially Integrated Social Science (CSISS), Department of Agriculture and Consumer Economics, University of Illinois, Urbana-Champagne.

Armstrong, M.P., Rushton, G. and Zimmerman, D. L., 1999. Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18(5):497-525.

Cox, L.H., 1996. Protecting confidentiality in small population health and environmental statistics. *Statistics in Medicine*, 15:1895-1905.

Cox, L.H., 1994. Matrix masking methods for disclosure limitation in microdata. *Survey Methodology*, 20(2):165-169.

Duncan, G.T. and Pearson, R. W., 1991. Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Statistical Science*, 6(3):219-239.

Fotheringham, A.S., Brunsdon, C. and Charlton, M., 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, West Sussex, England: John Wiley and Sons.

Getis, A. and Ord, J.K., 1996. Local spatial statistics: An overview, In Longley, P. and Batty, M. (Eds.) *Spatial Analysis: Modeling in a GIS Environment*, 269-285. Cambridge, England: Information International.

Leitner, M., 1997. *The Impact of Attribute Certainty Displays on Spatial Decision Support*, unpublished dissertation, Department of Geography, State University of New York at Buffalo.

Leitner M. and Buttenfield, B.P., 2000. Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science*, 27(1):3-14.

Leitner, M. and Curtis, A., 2003. Cartographic guidelines for the visualization of confidential data using geographic masking. *21st International Cartographic Conference*, Durban, South Africa, 101-110.

Monmonier, M., 2002. *Spying with Maps: Surveillance Technologies and the Future of Privacy*. Chicago, IL: University of Chicago Press.

©SPSS Inc., 2001. *SPSS for Windows*, Release 11.0.1. Chicago, IL: SPSS Inc.

Wartell, J. and McEwen, J. T., 2001. *Privacy in the Information Age: A Guide for Sharing Crime Maps and Spatial Data*, U.S. Department of Justice, Office of Justice Programs.