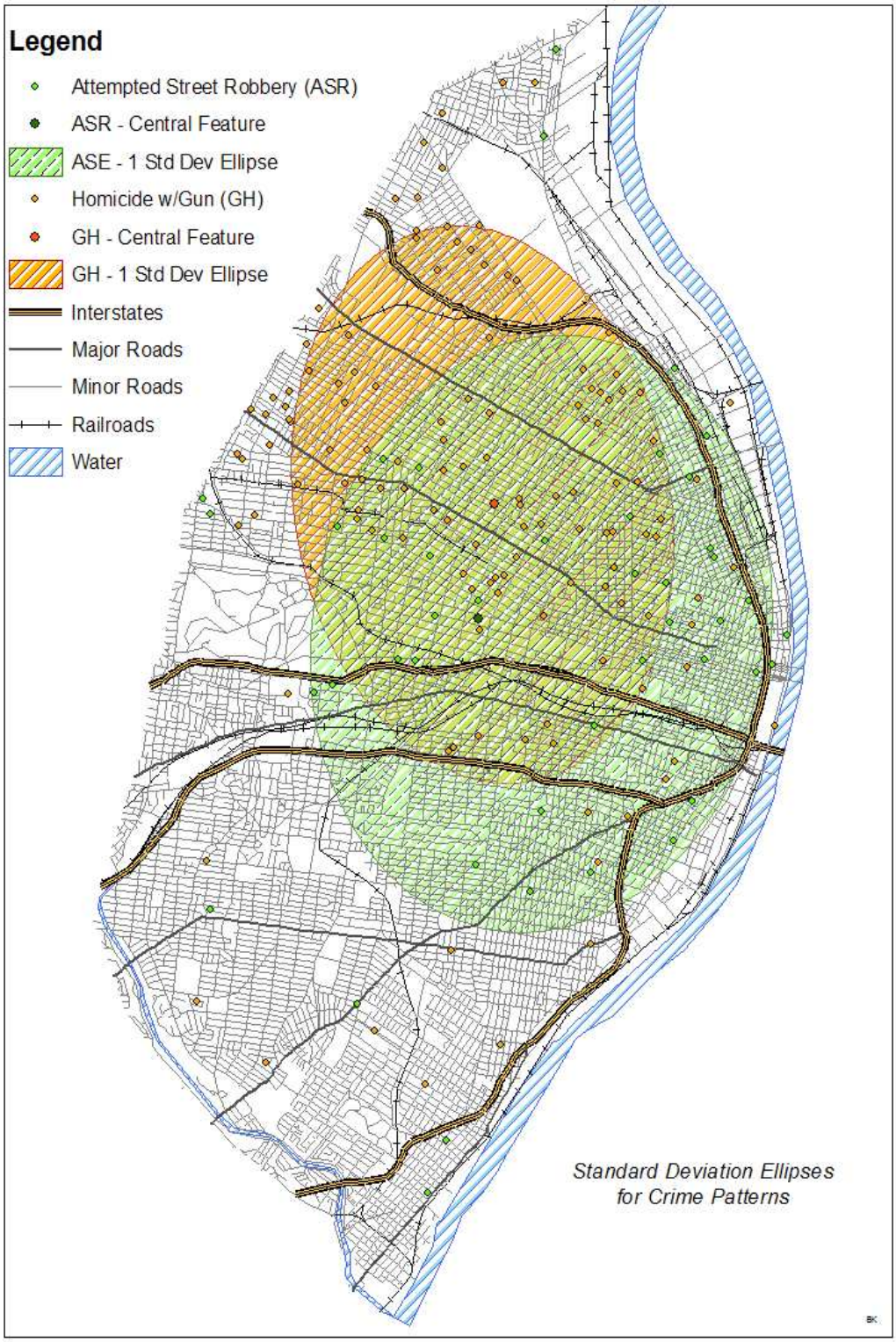
1. Standard Deviational Ellipses for the two crime patterns.



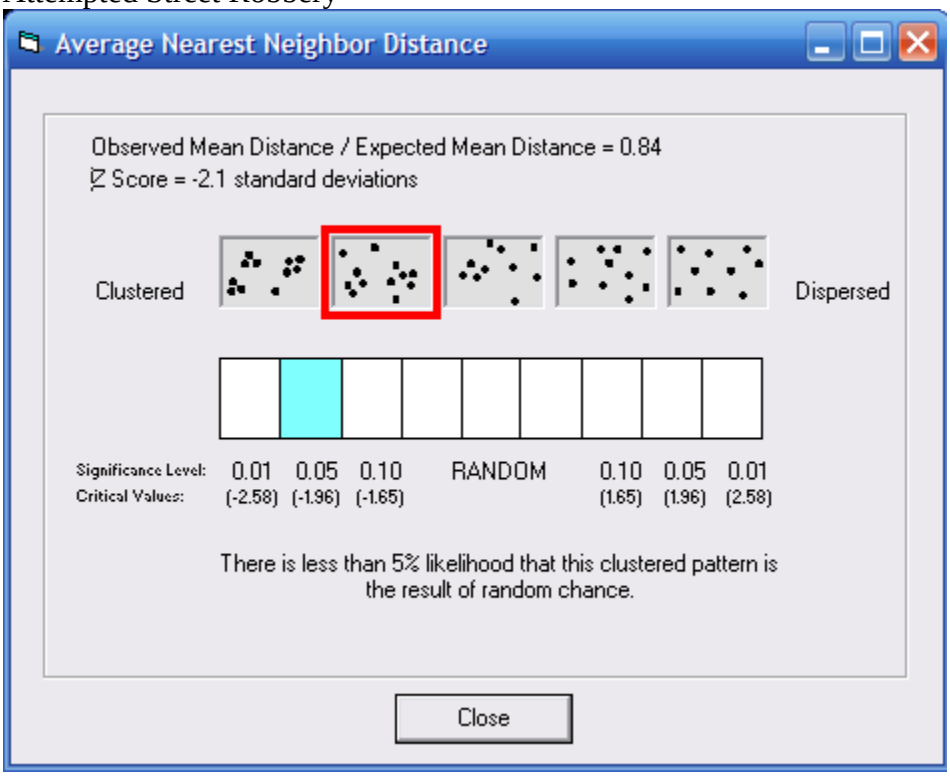
I expected a much more E-W summary ellipse for the Gun Homicides (GH) - see the density map. The study area is probably what skews the ellipse; the smaller number of points at the south end of the map is

just enough to tilt the ellipse in that direction. The large number of points just outside the GH ellipse is most likely related to edge effect. It seems hard (impossible?) to have part of the summary ellipse outside of the study area, meaning there is no way for these points to be in the summary ellipse. This limits its effectiveness as a tool.

The angle of the Attempted Street Robbery (ASR) ellipse initially led me to believe that data is more random in nature since the angle of the ellipse looks the same as the general angle of the shape of the study area. The size of ellipse also implies more ASR data is 'normal' than the GH. The central feature for the ASR is quite offset from the center of the ellipse, showing the data is weighted spatially toward the river.

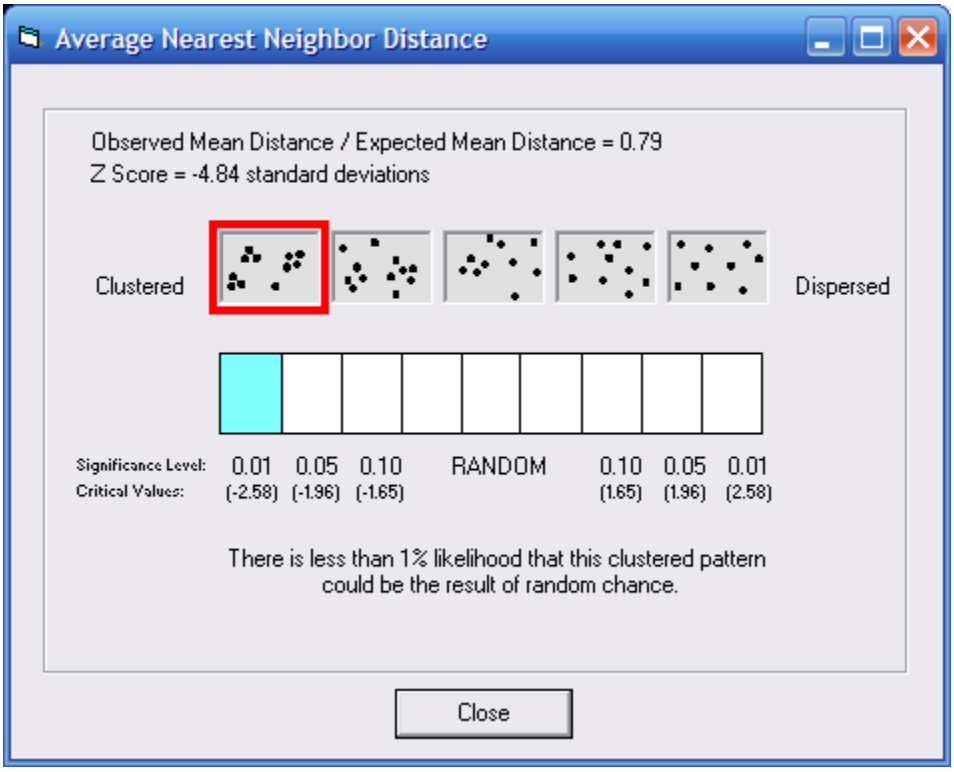
2. Nearest neighbor distance statistics

Attempted Street Robbery



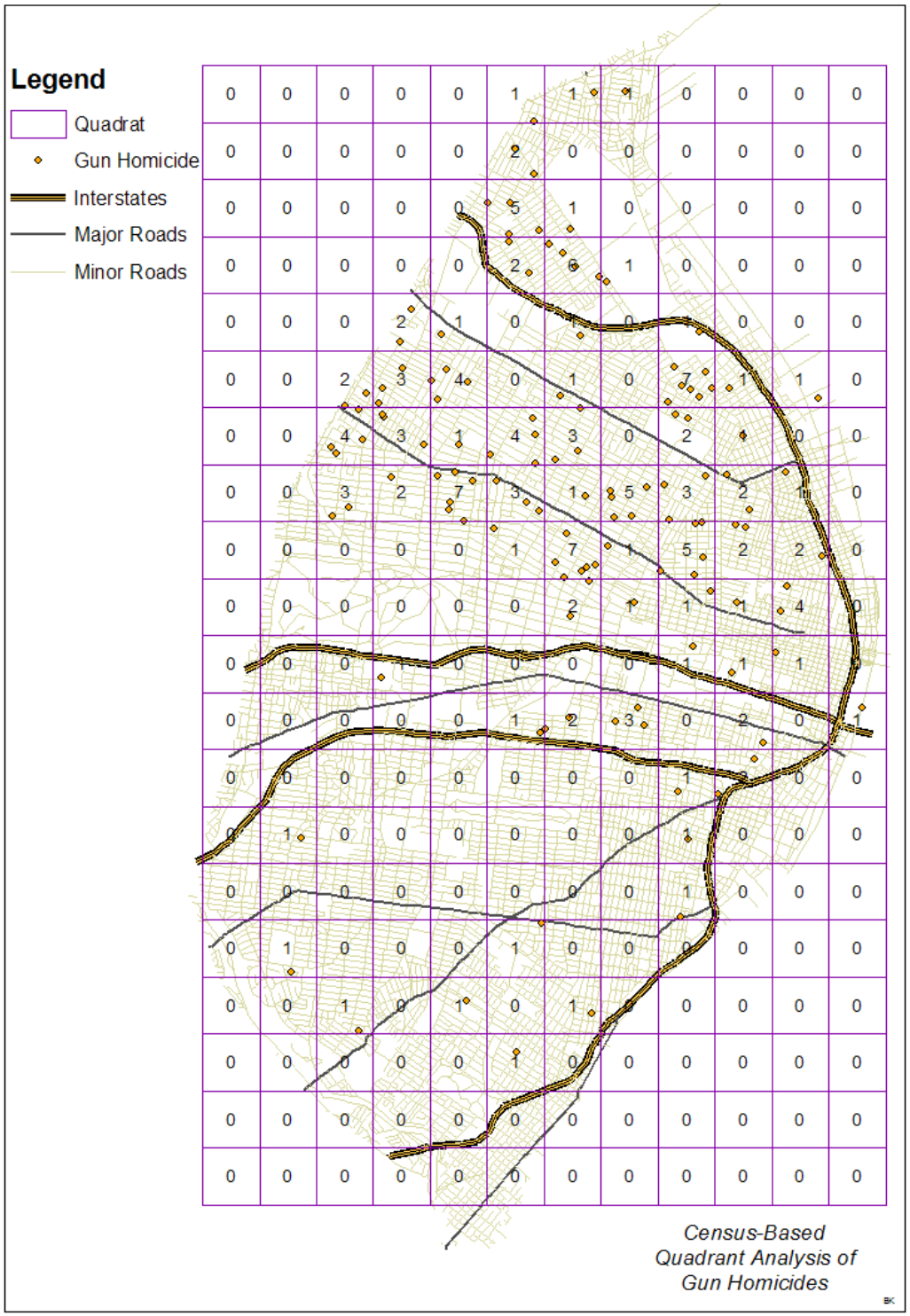
I was a little surprised by this data as I didn't expect the clustering level to be so high. I agree with the text that this method isn't very useful. It does begin to answer the '*how clustered?*' question but it is difficult to apply it in a concrete way to what we see on the map.

Gun Homicide



The z-score given in the GH data shows a very high value that is off the charts! At least at an extreme end of the spectrum this data does tell us - yes this is really clustered data.

3. Quadrat Analysis



Gun Homicide(all 12x20 quadrats)

Quadrat Count Statistics

Method	Census
#Points (n)	141
#Quadrats (x)	240
μ	0.5875

Export

k	x	$k-\mu$	$(k-\mu)^2$	$x(k-\mu)^2$
0	173	-0.5875	0.3452	59.7120
1	36	0.4125	0.1702	6.1256
2	13	1.4125	1.9952	25.9370
3	7	2.4125	5.8202	40.7411
4	4	3.4125	11.6452	46.5806
5	3	4.4125	19.4702	58.4105
6	1	5.4125	29.2952	29.2952
7	3	6.4125	41.1202	123.3605

Variance (s^2)	1.6257		
VMR	2.7671	χ^2	664.11

P = 0

Gun Homicide (road intersecting quadrats)

Quadrat Count Statistics

Method	Census
#Points (n)	141
#Quadrats (x)	178
μ	0.7921

Export

k	x	$k-\mu$	$(k-\mu)^2$	$x(k-\mu)^2$
0	111	-0.7921	0.6275	69.6500
1	36	0.2079	0.0432	1.5555
2	13	1.2079	1.4589	18.9662
3	7	2.2079	4.8747	34.1227
4	4	3.2079	10.2904	41.1616
5	3	4.2079	17.7061	53.1184
6	1	5.2079	27.1219	27.1219
7	3	6.2079	38.5376	115.6128

Variance (s^2)	2.0298		
VMR	2.5625	χ^2	456.12

P = 0

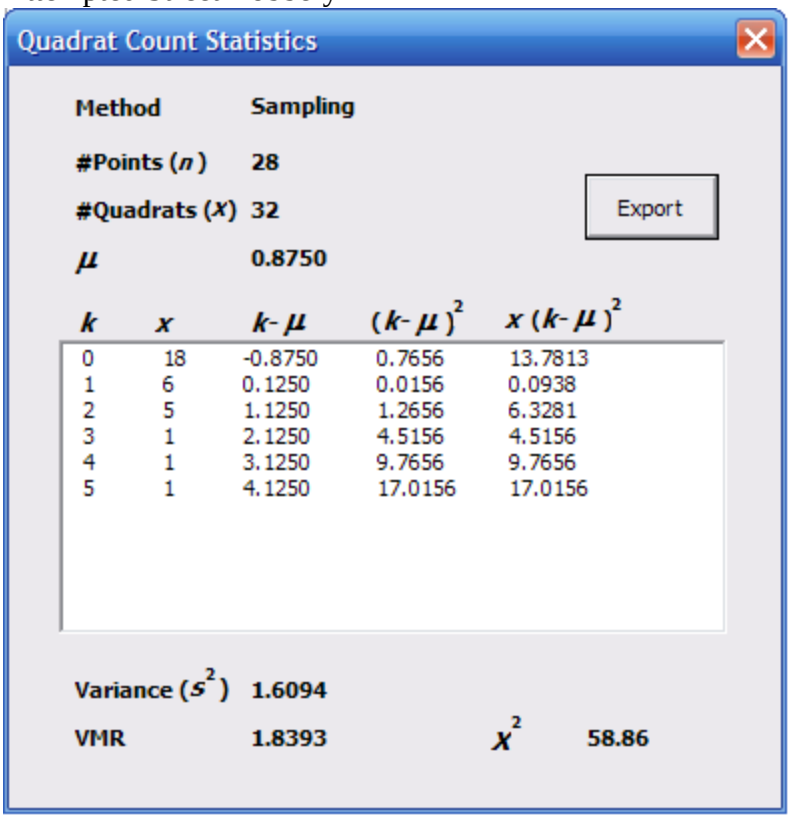
Census analysis was chosen for the Gun Homicides. This was done in order to include the entire study area. From the summary ellipse, it appeared the data points in the southern part of the city have an effect on the overall data. A quadrat size of 1 square km was chosen. This seemed a generally reasonable size and made it easy to use the quadrat analysis tool.

Two separate result of the analysis are shown: one with all 240 quadrats show on the map and the other with only the quadrats containing roads. While removing the quadrats located completely outside the study area does technically make the data statistically more random (with a lower VMR): it is of no consequence, either way the data is dominated by the clusters. In both cases the p-value is zero (to at least twenty decimal places), meaning there is no chance the data is the result of a randomly distributed point pattern.

The census method does have one other benefit. It begins to show what the results of density maps will be with seven quadrats with five or more points.

Legend
Attempted Street Robbery (ASR)
Interstates
Major Roads
Minor Roads
Railroads
Sample-Based
Quadrant Analysis of
Attempted Street Robberies

Attempted Street Robbery



Quadrat Count Statistics

Method	Sampling
#Points (n)	28
#Quadrats (x)	32
μ	0.8750

Export

k	x	$k-\mu$	$(k-\mu)^2$	$x(k-\mu)^2$
0	18	-0.8750	0.7656	13.7813
1	6	0.1250	0.0156	0.0938
2	5	1.1250	1.2656	6.3281
3	1	2.1250	4.5156	4.5156
4	1	3.1250	9.7656	9.7656
5	1	4.1250	17.0156	17.0156

Variance (s^2)	1.6094		
VMR	1.8393	χ^2	58.86

P=0.0018

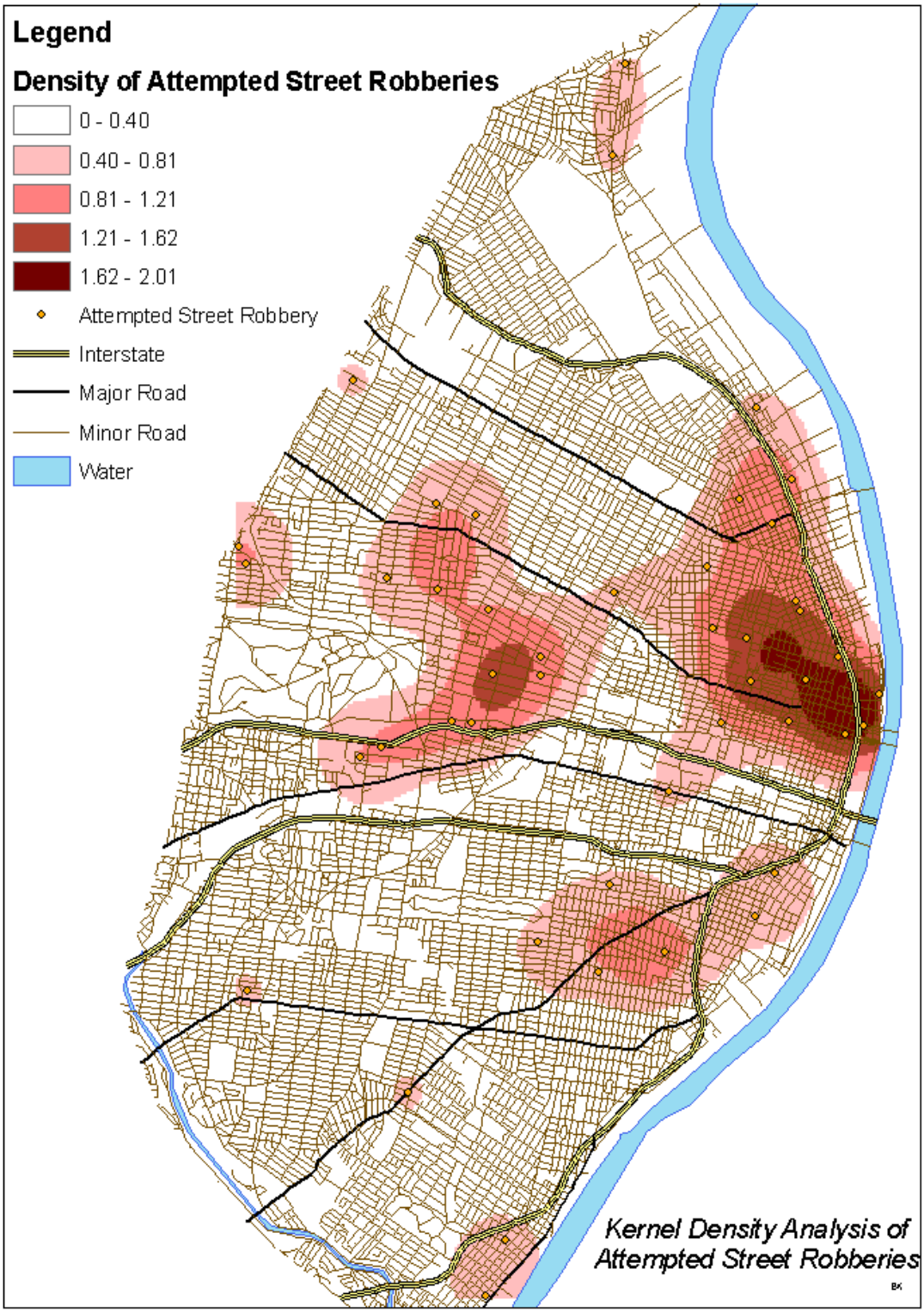
For the Attempted Street Robberies, a sample-base analysis was chosen. The goal here was two-fold: first, to create a sample area with as little as possible space with no data and secondly, to concentrate on the points in the middle section of the map to see if they were randomly spaced within the sample area.

A circle was chosen to for the quadrat shape to limit longer road distances which actually extend the distance beyond the size of the sample radius. A 1 km radius was used and 32 samples were taken. Based on the size of quadrat about 64 samples would cover the sample area assuming no overlap, so this means 32 quadrats covers area (including overlap) equal to half the sample area.

Overall the sample method seems to have worked here. It identifies a cluster on the east side but shows a more random spread elsewhere on in the sample area. If the two quadrats with x=4 and x=5 are reduced to x=1 the chi-squared value would only be 24.75 which gives a p-value of 0.77.

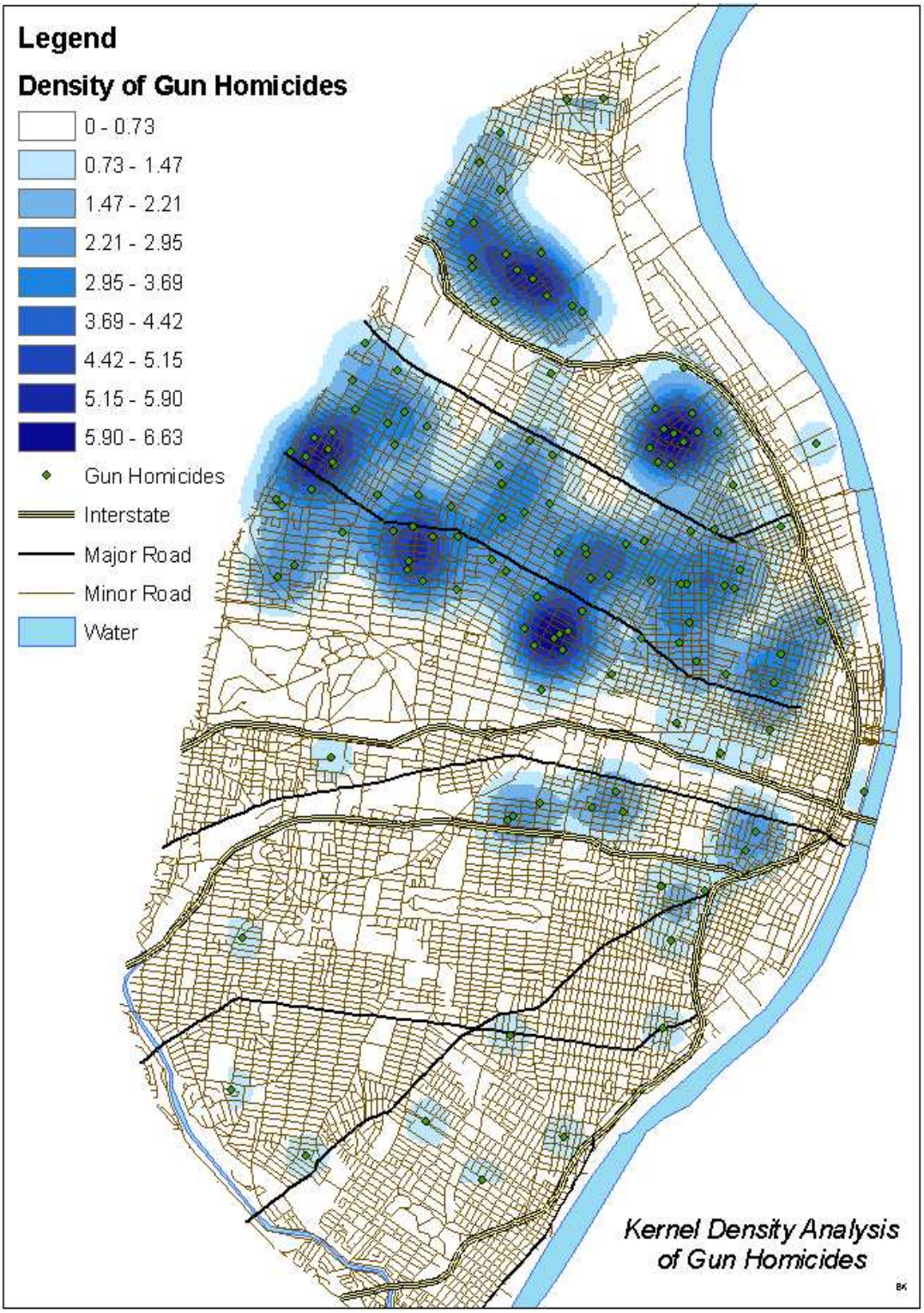
This situation indicates more sample quadrat scenarios should be run (Monte Carlo style) to calculate what the average chi-squared value. This will determine if the cluster is recognized consistently or not. Using ten samples, a range for chi-squared from 26 to 64 was returned with a mean value of 49.6. Two interesting points, the chi-squared value of 26 missed the cluster at the east end almost entirely (two samples in that area had just one point and none had two or more). Conversely, a chi-squared of 63 or higher was calculated in three of the ten samples. In these samples, a second cluster in the middle of the sample area is also indicated; furthering the idea this data is not the result of a random process.

4. Density maps



Bandwidth = 1500m Cell Size = 50m

The bandwidth used here was chosen through trial and error. Larger bandwidth created too much smoothing and smaller bandwidths did not provide enough values to the upper ranges of an equal interval classification. The values used here do confirm the conclusions reach in the quadrat analysis - the existence of two clusters: a strong one on the east side and a secondary cluster in the center. There is some feeling of make the results fit the data here, which makes the analysis questionable but it is hard to see what other conclusions can be reached in a similarly visual supported way. Another interesting result is the high density areas actually have few point in them are exist mostly because there are a number of nearby point. However, this seems appropriate due to spatial autocorrelation more armed robberies are likely to take place in the high density areas.



Bandwidth = 1000 Cell size = 50

The Gun Homicide density map shows five very strong clusters that were hinted at in the quadrat analysis. The bandwidth used is the same as the size of the quadrat used earlier. The results provide a clear indication of the spatial characteristics of this type of crime.

The density analysis shows the data points in the southern half of the city bear no spatial relationship to each other. These events are more or less random spatially and most likely the reasons are quite different as well.

5. Conclusions about the study area

The study area is not effectively set by the dataset. The Modifiable Areal Unit Problem (MAUP) is relevant here. It would be very helpful to know what is happening to the west of the city especially for the north half of the map. Does the high rate of crime continue to the west? How far does it continue? If that area includes more data points, the analysis become richer and the density and clustering become more complex. If there is truly a cessation of crime to the west (like to the south) then the clustering characteristics of the data become even more powerful. Coupled with demographic data this data would be far more telling than the dataset given just by the city itself.

A discussion of the demographics is inescapable here. The situation in North St Louis in 1982 existed for a number of reasons. But the methods by which data is collected and subject to the MAUP and the ecological fallacy limit the power of the spatial statistical tools to find the relationships which can be addressed to help reduce the problem.