**Handout for Presentation: Issues in Spatial Analysis**

January 14, 2003, 10.50-12.20


By Dr. Carl Amrhein[1], Dr. Rick Glazier[2], MA. Peter Gozdyra[3]

[1] Professor, Department of Geography and Planning, and Dean, Faculty of Arts and Science, University of Toronto
tel 416 978-3383, fax 416 978-3887
email dean@artsci.utoronto.ca

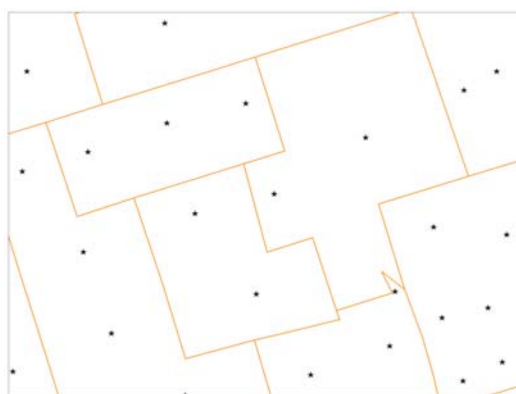
[2] Associate Professor, Department of Family and Community Medicine, Department of Public Health Sciences, University of Toronto, Inner City Health Research Unit (St. Michael's Hospital), Adjunct Scientist ICES
tel 416 864-6060 x. 2574, fax 416 864-5485
email richard.glazier@utoronto.ca

[3] Research Coordinator, Inner City Health Research Unit (St. Michael's Hospital)
tel 416 864-6060 x. 2749, fax 416 864-5485
email gozdyrap@smh.toronto.on.ca


## 1) Selected spatial elements of data (Carl Amrhein)

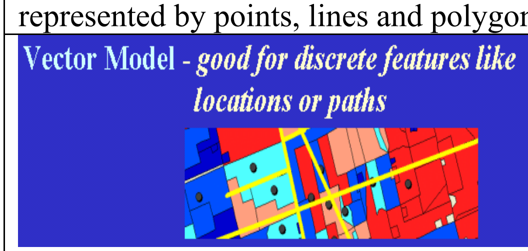
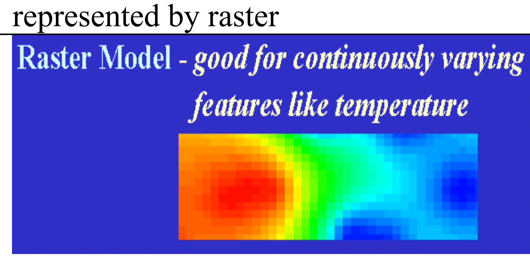
a) Point-assigned information vs. areal information

*Figure 1 – point and area geographical objects*

| *Point geographical objects (e.g.: block face) vs. areal objects (e.g.: Stats Can 2001 dissemination areas)* |
|---|
|  |

b) Continuous vs. discrete spatial phenomena

*Figure 2 – Vector and raster GIS data models*

| Discrete phenomena can be well represented by points, lines and polygons | Continuous phenomena can be well represented by raster |
|---|---|
|   |   |

c) The time element in data

d) Spatial objects of the study area: 'block face (BF)', 'postal codes (PC's)', 'enumeration areas (EA's)', 'census tracts (CT's)', 'forward sortation areas (FSA's)'

*Figure 3 - Statistics Canada 2001 geographic divisions (source: Statistics Canada)*

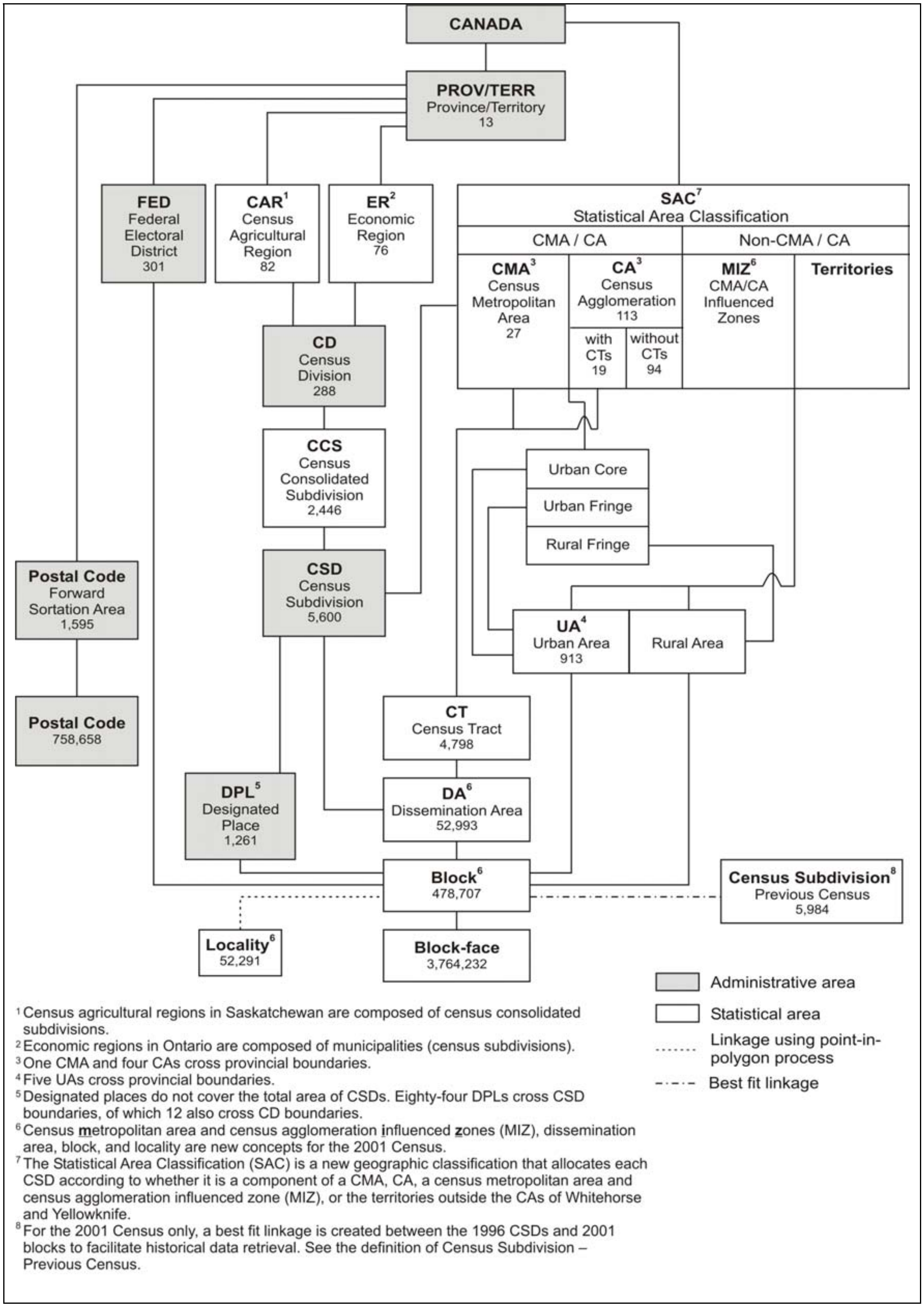


*Figure 4 – Three levels of statistical spatial units*

| Enumeration areas | Census tracts | Forward sortation areas |
|---|---|---|
|  |  |  |

**2) Selected spatial topics: spatial autocorrelation; aggregation of spatial data for (health) analysis (Carl Amrhein)**

a) Spatial autocorrelation of data.

(See Appendix 1)

b) Why data need to be aggregated spatially?

i) Matching available data sets to the smallest common spatial denominator

ii) Making spatial units compatible with units used in other studies

iii) Removing small cells problem

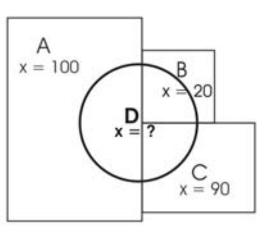
iv) Creating meaningful areal units for analysis

c) Basic methods of spatial aggregations

i) Aggregation of several smaller areas to one larger zone

ii) Spatial-weighting during aggregation in case of boundary misalignments at different levels of aggregation

*Figure 5 – Area-weighting during data transfers between polygon layers*

| Area-weighted data transfer from areas A, B and C to area D |
|---|
|  |



$X_D = 1/5X_A + 1/2X_B + 1/3X_C$

$X_D = 20 + 10 + 30$

$X_D = 60$

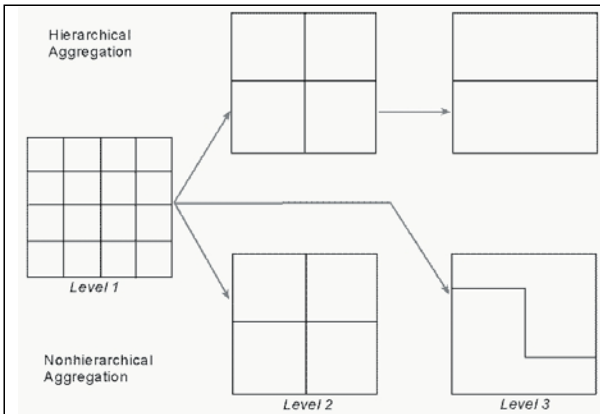
iii) Point-to-polygon data aggregation (e.g.: BF to EA)

*Figure 6 – Point-to-polygon data transfers*

Aggregation of point data to areas


iv) Hierarchical vs. non-hierarchical aggregations

*Figure 7 - Hierarchical vs. non-hierarchical spatial units aggregation*




d) Consequences of spatial data aggregation

i) Loss of homogeneity within larger units

ii) Loss of positional accuracy of count/frequency data summed up in areal units (See Appendix 2)

iii) Changed statistical characteristics of data (MAUP: scale and zoning effects; several examples of the impacts of spatial aggregations on means, averages, correlation coefficients etc)

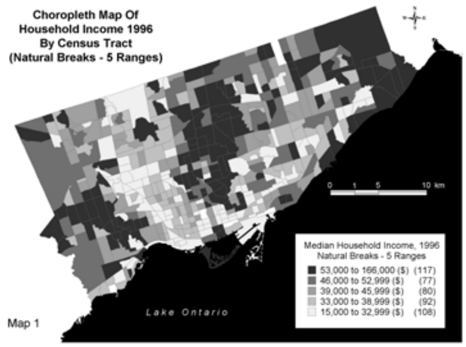
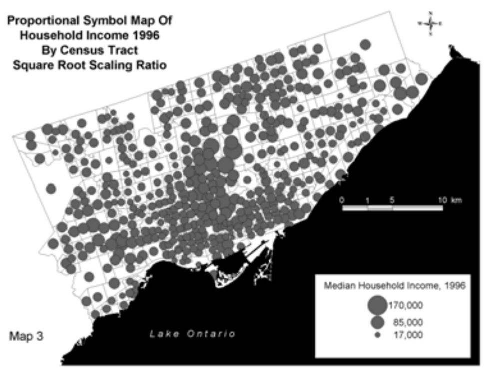
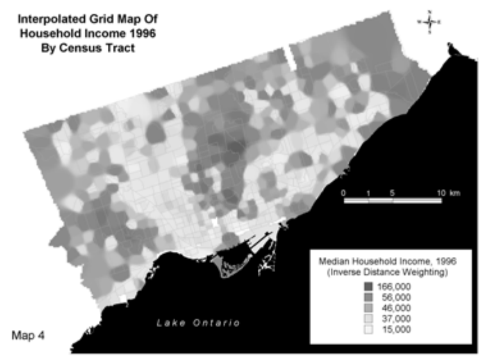
iv) Guidelines for choosing the most appropriate spatial units

3) **Selected issues in data mapping (Peter Gozdyra)**

a) Thematic maps

i) Types of thematic maps and variable types best shown by these maps

*Figure 8 – Various types of thematic maps*

<table>
<tr><td>Choropleth (shaded) map</td><td>Dot density map</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Proportional symbol map</td><td>Interpolated grid map</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

ii) Strengths and weaknesses of various map types

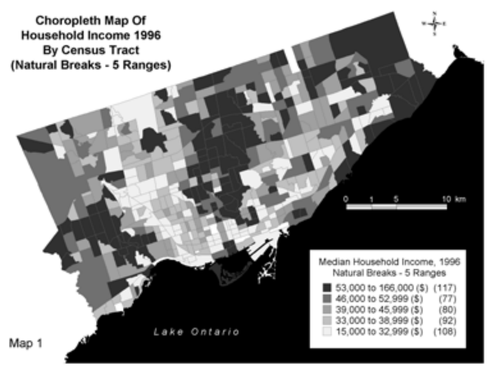
*Table 1 – Pros and cons of using various types of thematic maps*

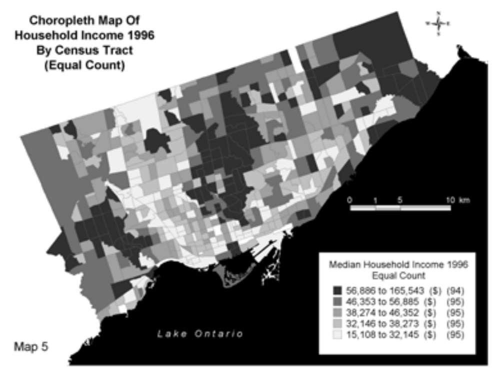
| **Choropleth (shaded) map** | **Dot density map** |
|---|---|
| Variables types:<br>rates, ratios, density per areal unit<br>Shows well:<br>value ranges for specific areas, overall pattern<br>Shortcomings:<br>small areas are visually overpowered by large areas | Variables types:<br>counts/frequencies<br>Shows well:<br>spatial distribution, clusters<br>Shortcomings:<br>deceptive impression of accurate locations of dots in randomly distributed dot maps |

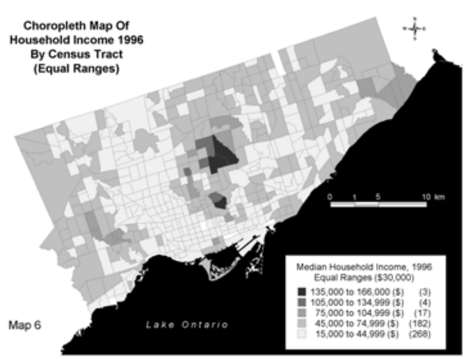
| **Proportional symbol map** | **Interpolated grid map** |
|---|---|
| <u>Variables types:</u><br>rates, ratios, counts/frequencies<br><u>Shows well:</u><br>magnitude of phenomenon, comparison of areas of different sizes<br><u>Shortcomings:</u><br>overlaps of symbols on each other and on other map layers | <u>Variables types:</u><br>rates, ratios, density per areal unit<br><u>Shows well:</u><br>gradual change of continuous-type phenomenon over space<br><u>Shortcomings:</u><br>values interpolated between known data points, centroids etc., pattern depends on a chosen algorithm |

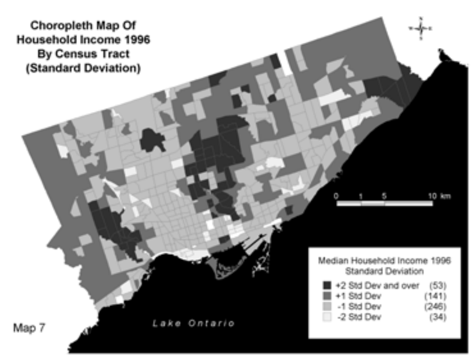
iii) Defining limits and the number of ranges on choropleth maps (map patterns)

*Figure 9 – Different ways of data classification on choropleth maps*

<table>
<tr><td>Natural breaks</td><td>Equal count (equal number of areas in each category)</td></tr>
<tr><td>


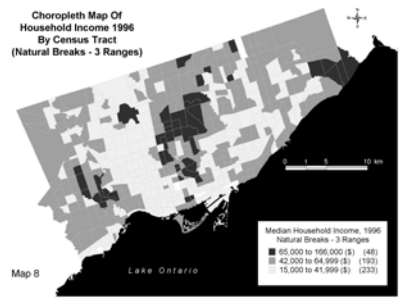
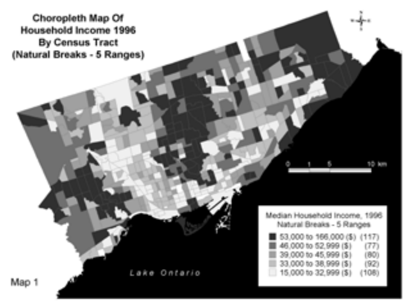
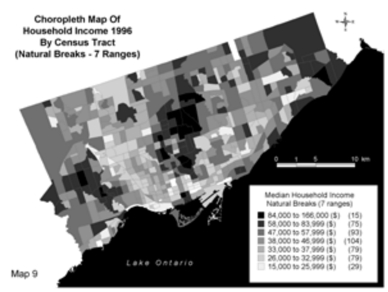
</td><td>


</td></tr>
<tr><td>Equal range (equal data ranges in each category)</td><td>Standard division</td></tr>
<tr><td>


</td><td>


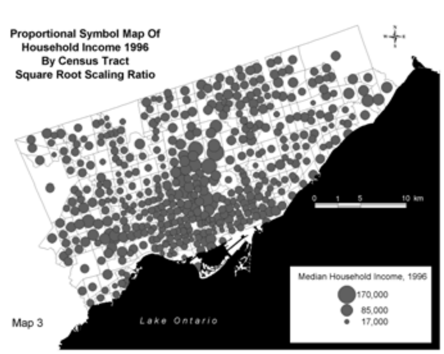
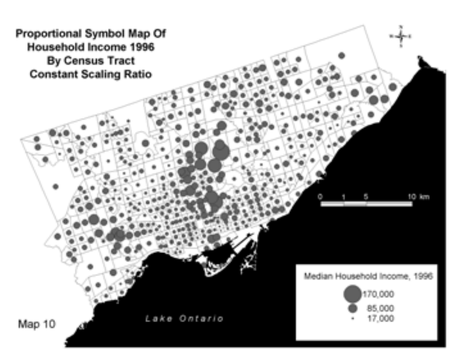
</td></tr>
</table>

*Figure 10 – Number of ranges on choropleth maps (map pattern)*

<table>
<tr><td>3 ranges</td><td>
</td></tr>
<tr><td>5 ranges</td><td>
</td></tr>
<tr><td>7 ranges</td><td>
</td></tr>
</table>

iv) Ways of scaling symbols on proportional symbol maps

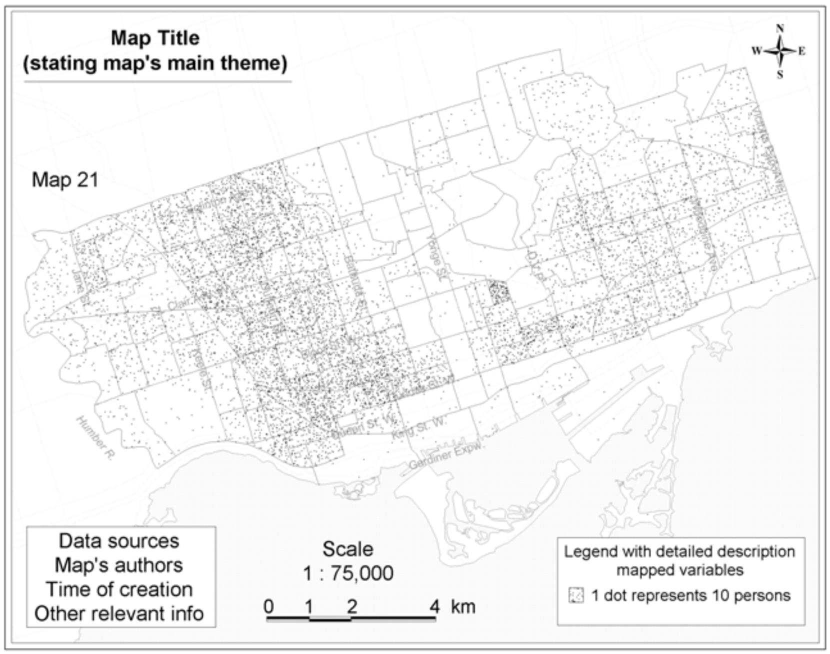
*Figure 11 – Scaling schemes of proportional symbols*

<table>
<tr><td>Square root scaling scheme – size of symbols proportional to magnitude of phenomena</td><td>Constant scaling scheme – height of symbols proportional to magnitude of phenomena</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

b) Map’s elements

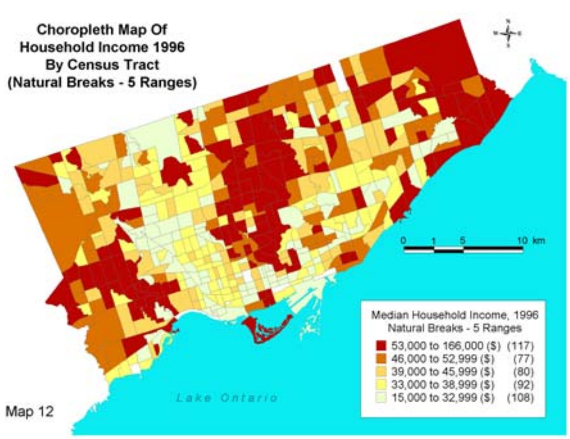
i) Basic map elements

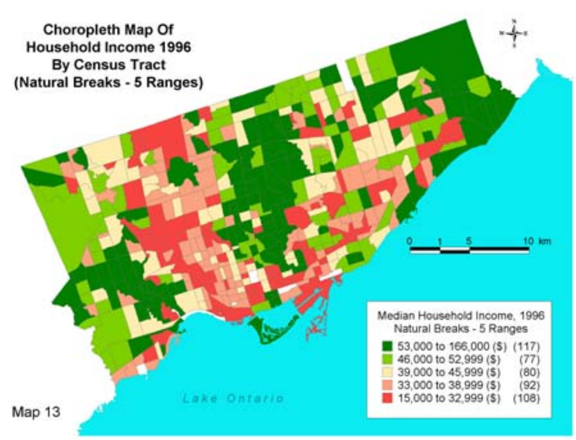
*Figure 12 – Basic map elements*




ii) Map colours

*Figure 13 – Use of colours on ‘Average Household Income’ choropleth map*

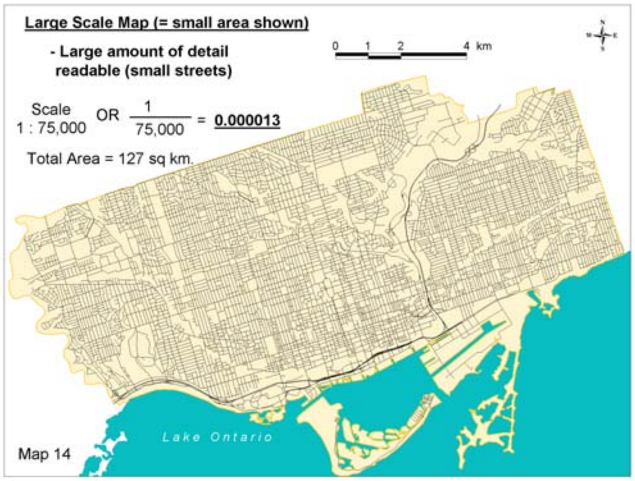
<table>
<tr><td>Poor colour choice - red colour indicates intuitive ‘negative’ type phenomena, while the increasing level of average income is a ‘positive’ event</td><td>Good colour choice – two-colour scheme reflects well ‘negative’ aspect of low levels of average income (red) and ‘positive’ aspect of higher levels of income (green)</td></tr>
<tr><td>


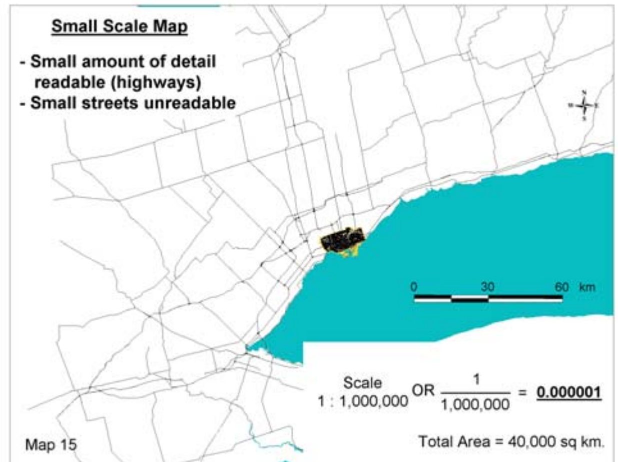
</td><td>


</td></tr>
</table>

iii) Map scale

*Figure 14 – Relationship between map's readable detail levels and map's scale.*

<table>
<tr><td>Large scale map 1 : 75,000 – detailed street coverage easily readable</td><td>Small scale map 1 : 1,000,000 – detailed street coverage unreadable</td></tr>
<tr><td>


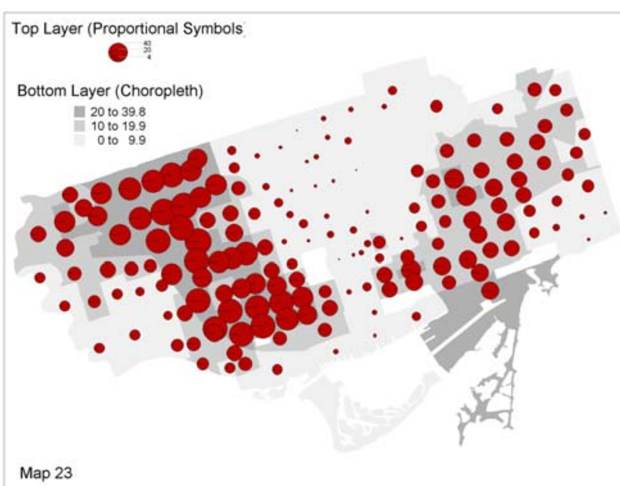
</td><td>


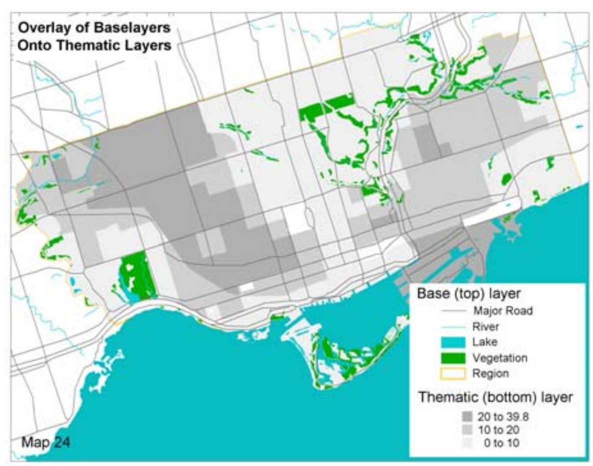
</td></tr>
</table>

c) Overlaying of variables

i) Contents and stacking of layers

*Figure 15 – overlaying of layers on the map*

<table>
<tr><td>Overlay of two thematic layers (proportional symbol layer is placed on the top of choropleth layer)</td><td>Overlay of several baselayers (roads, rivers, green spaces) onto one thematic layer (choropleth)</td></tr>
<tr><td>


</td><td>


</td></tr>
</table>

d) Basic model for utilization of GIS and maps in health care

**Defining questions -> Identification and Selection of Variables -> Mapping -> Map Interpretation -> Decision Making**

**<u>Appendix 1</u> - Spatial autocorrelation** (source: P. Gozdyra)

When examining spatial data the values tend to either cluster together showing groupings of regions with similar characteristics or they demonstrate less ordered patterns within the area of study. In the latter case regions show weaker data patterns or even a patchwork indicating that regions with relatively different values tend to neighbour with each other. The two most common ways of measuring the clustering of similar values in two-dimensional space are the Moran Coefficient [MC] and Geary Ratio [GR] defined as:

$$MC = \left(\frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}\right)\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \qquad (1)$$

$$GR = \left(\frac{n-1}{2\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}\right)\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \qquad (2)$$

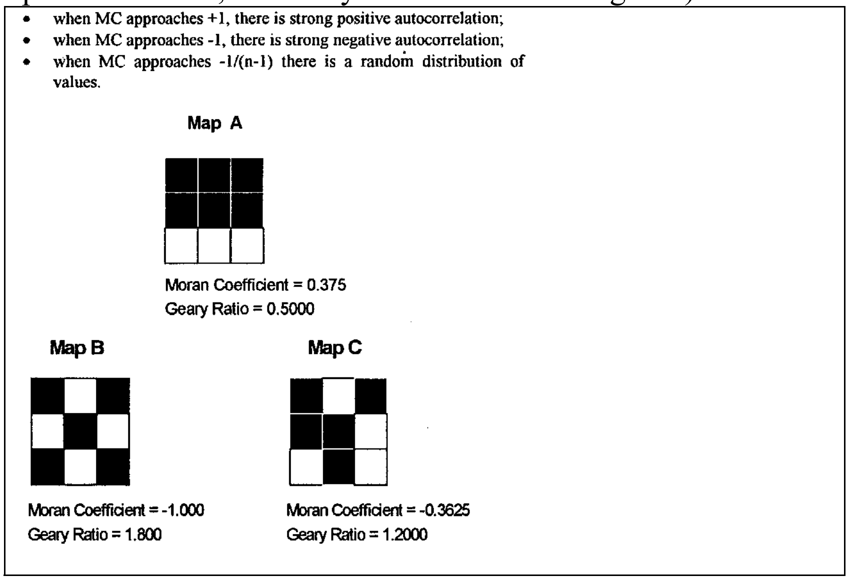
where: $n$ is the number of regions, $x_i$ and $x_j$, are the attribute values of the neighbouring spatial units, $c_{ij}$ is the weighting component from the connectivity matrix for neighbouring units $x_i$ and $x_j$, $\bar{x}$ is the mean attribute value of a whole region

As shown, formulas (1) and (2) utilize a weighting scheme, which in both cases is usually a binary connectivity matrix based on the adjacency of regions. The interpretation of the values of MC and GR is shown in Table 1. Later sections of this paper describe some relationships between spatial autocorrelation and MAUP.

*Table 2: Value ranges for Moran Coefficient and Geary Ratio*

| Statistic\Value | -1.0 | 0.0 | 1.0 | 2.0 |
|---|---|---|---|---|
| Moran Coefficient | Strong negative autocorrelation | Random distribution of values | Strong positive autocorrelation | |
| Geary Ratio | | Strong positive autocorrelation | Random distribution of values | Strong negative autocorrelation |

*Figure 16* - Sample visual patterns of spatial autocorrelations (source: Practical Handbook of Spatial Statistics, Edited by: Sandra Lach Arlinghaus)

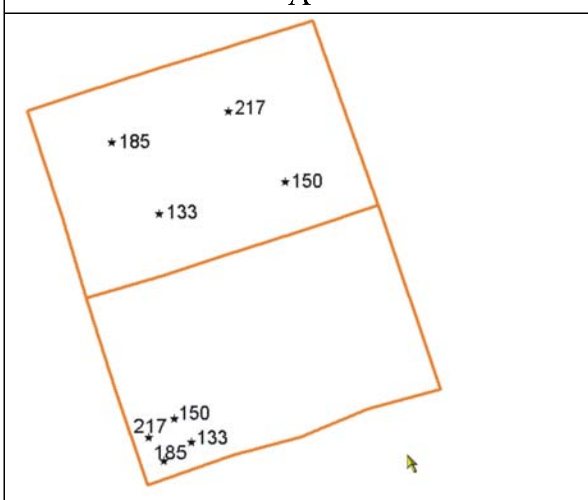
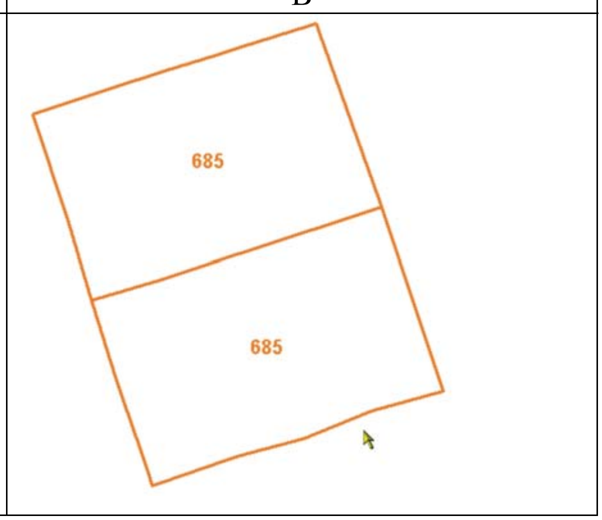



**<u>Appendix 2</u> - Loss of spatial accuracy due to aggregation of point data to areas** (source: P. Gozdyra)

**Use of point-define data vs. polygon-summed data.**

In the table below (column A) the points in the higher-located polygon are spread out more evenly than in the lower-located polygon. Upon aggregation of point data to polygons (column B) only one number represents the whole data in each polygon. The higher-located polygon may be thought of as well representing the distribution of the original point data, while the lower-located polygon conceals the true distribution of highly concentrated point data (loss of positional accuracy of data)

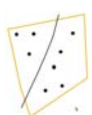
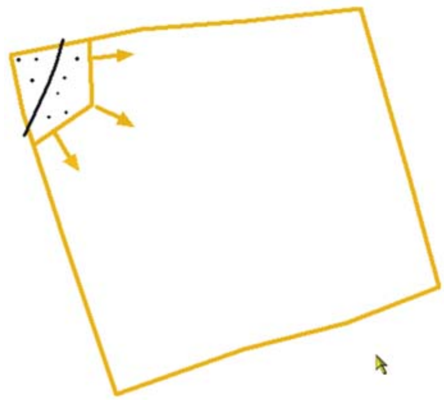
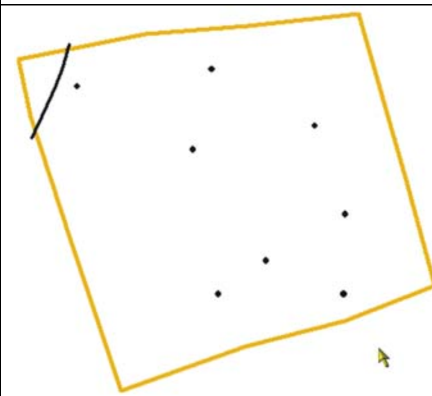
*Figure 17 – Appropriateness of areas to represent point data*

| A | B |
|---|---|
|   |   |

**Representation of count data within areas by random points (dot density map).**

In the table below data points are in reality located close to the line (e.g. population living close to a street). When a polygon-summed data is represented by random points (e.g.: one dot representing 100 persons) a use of a smaller polygon (column A) makes the points fall closer to the true physical location of the population. An aggregation of population to larger polygon (column B) causes the random points to be located further away from the true physical location of data (column C).

*Figure 18 – Loss of spatial proximity of point data to its sources during their aggregation to larger areal units*

| A | B | C |
|---|---|---|
|  |  |  |