TOWARDS A UNIFORM SPATIAL SCALE
(AN APPROACH TO SOLVING THE REOCCURRING MODIFIABLE AREAL UNIT PROBLEM)

USING POPULATION DENSITY AS A VARIABLE FOR MONTREAL AND TORONTO

TRANSPORTATION AND LAND DEVELOPMENT : URBP 619 : WINTER 2005 : SCHOOL OF URBAN PLANNING : MC GILL UNIVERSITY

A S H I S H   G H A T E
Mc Gill I.D. : 260154443

Contents

April 12, 2005

All maps, tables and graphs have been made by the author, using data supplied by Prof. Murtaza Haider , Assistant Professor, School of Urban Planning , McGill University.

Ashish Ghate (B. Arch)

Masters of Urban Planning, Candidate '04 -'06.
School of Urban Planning, McGill University.
Montreal, Canada.

ashish.ghate@mail.mcgill.ca

Abstract

Aggregated demographic datasets are associated with analytical problems due to the arbitrary nature of areal unit partitioning. Perhaps the most prominent of these issues is the modifiable areal unit problem (MAUP), defined as a situation in which modifying the boundaries and/or scale of data aggregation significantly affects the results of spatial data analysis (Openshaw 1983). The view expressed here is that to ignore these problems in spatial analysis and modeling would imply that location and spatial scale is not relevant for spatial analysis. This paper brings out the discrepancies surrounding the use of spatial data at different scales and describes a set of spatial analysis methodologies to permit scale-sensitive and location specific analyses of population data. As a demonstration/comparison, 300, 600 meter-resolution interpolated raster grids and inverse distance weighted population surfaces are generated from Canadian Census Tract and Enumeration Area data for Montreal and Toronto regions. The paper demonstrates that there is a need for more than one approach to address the issue of spatial analysis problems. The examples describe a methodology for generating a surface-based representation of population that minimizes these problems.

Key Words:

Spatial Scale, MAUP, Spatial Analyst, Population density surface – Montreal Toronto.

Introduction

{In the 1950s and 1960s, gross urban population density across Canada and America fell quickly and a new phrase**, "urban sprawl",** was coined to describe the phenomenon}.[1] The regions which defined the central regions of these cities greatly expanded and led to a range of development from satellite towns to industrial districts and employment hubs. This was but natural considering the demand for housing, increasing urban/suburban population, availability of cheaper land on the outskirts of the city, and infrastructure - predominantly transport. {However, in the 1970s, 1980s, and 1990s, new development in the form of clustered housing, in-fill, redevelopment, and conversions helped raise densities in parts of some urban regions}. [2]

While density is almost universally regarded as one of the essential components of sprawl, there is little agreement about the appropriate specification of its measurement. There are a number of important considerations in determining how the relationship between density and sprawl should be evaluated. These include:-

- The best variable to use for representing density,
- The density level at which a city might be regarded as sprawling,
- The scale at which density should be measured, and the extent of space over which density should be characterized.

A number of variables have been used to represent density, most commonly density of housing units, population, and/or employment. {While each of these variables has the capacity to capture the density characteristics of sprawl in a given city, it is unclear as to which variables work best.}[3] The scale at which density is studied is also an important consideration. Depending on the scale of observation—the metropolitan area, a district within a city, a neighborhood—measurements of urban density look quite different. The geography over which densities should be measured is also contentious.

Despite the level of attention that is given to sprawl, their remains relatively little understanding of its determinants and its constitution. In recent years researchers have traded conceptual explorations of the sprawl phenomenon: its causes, characteristics, costs, and potential controls. However, most of these analytical studies are done only at one spatial scale, either at Census Tracts or Enumeration Areas. Most of these studies overlook the fact that when scales of observation or analysis change, i.e., when the unit size, shape, spacing or extent is altered, statistical results change dramatically. This is a phenomenon which geographers commonly refer as the *Modifiable Areal Unit Problem (MAUP).* The kinds of results that may change include estimates of the population mean and variance. There is definitely a need to identify the influence of observational scale at the same time side by side and the statistical results.

The Modifiable Area Unit Problem

{The MAUP is classic problem associated with the design and display of boundaries. The MAUP is 'a form of fallacy associated with the aggregation of data into *areal units* for geographical analysis. This aggregated data is

[1, 2] **Urban sprawl in Canada and America: just how dissimilar?**
John R. Miron (Professor of Geography and Planning. University of Toronto at Scarborough)
http://www.citieslab.utsc.utoronto.ca/Papers/UrbanSprawl.pdf

[3] **Measuring Sprawl, Paper 27, Paul M. Torrens, Marina Alberti,** November 2000
Center for Advance Spatial Analysis , University College London
http://www.casa.ucl.ac.uk/working_papers.htm

then treated as individuals in analysis'. Many spatial datasets are collected on a fine resolution (i.e. a large number of small spatial units) but, for the sake of privacy and/or size concerns, are released only after being spatially aggregated to a coarser resolution (i.e. a smaller number of larger spatial units). The chief example of this process is census data which are collected from every household, but released only at the Enumeration Area or Census tract level of spatial resolution. When values are averaged over the process of aggregation, variability in the dataset is lost and values of statistics computed at different resolutions will be different; this change is called the ***scale effect.*** The results of aggregate analysis are not only related to the degree of aggregation but also strongly dependent on the choice of reporting zones. Areal units, such as administrative boundaries or even image pixels, are usually determined arbitrarily and are modifiable in that they can be spatially aggregated in an infinite number of ways. One also gets different values of statistics depending on how the spatial aggregation occurs; this variability is called the ***zoning effect***. In Openshaw's study of the MAUP (1984), he states:

'*The process of defining or creating areal units would be quite acceptable if it were performed using a fixed set of rules, or so that there was some explicit geographically meaningful basis for them. However there are no rules for areal aggregation, no standards, and no international conventions to guide the spatial aggregation process. Quite simply, the areal units used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating'*. }[4]

This was in 1985, and to date the situation has not improved much. {Openshaw (1996) points out that in order to compare zonal objects it must be certain, from a geographical point of view, that they are indeed comparable objects, otherwise there is a realm of comparing chalk with cheese. For example, correlating the percentage of elderly voters with Republican voters in Iowa counties, Openshaw and Taylor (1979) produced almost any result from perfect positive to perfect negative correlation by manipulating the reporting zone boundaries.}[5]

The purpose of studying the MAUP is to try to estimate the true values of the statistics at the original level of spatial resolution. This problem is integral to the display of demographic data as the information displayed is a product of the size, shape and scale of the administrative boundaries used in the data collection. In the past, because boundaries were assumed to be fixed, researchers had to use whatever boundaries were available. Consequently, the user had little, if any, control over the MAUP.

Modeling geographic information across scales in a data-rich environment brings together two of the most difficult aspects of scale in a geographic sense:

{• Aggregation and scale effects.

• Local analysis of spatial data and determining the sampling scale.

Scale effects are prevalent in all data sets and subsequently in all spatial analysis. After a thorough review of the landscape ecology and geographical literature, Jelinski and Wu (1996) concluded that there was no suitable encompassing theory for indicating how sensitive results are to the scale of the analysis and to variations in the way in which data are represented. Openshaw and Clarke (1996), Fotheringham (1998), Fischer et al. (1996) and many others have also voiced concern over this lack of theory behind data representation and a lack of theory behind GIS in general. Also,

[4] **The modifiable Area Unit Problem: Empirical analysis by statistical simulation. Harold David Reynolds.** University of Toronto, 1998, PhD. Thesis.

The modifiable Area Unit Problem. Openshaw and Taylor 1981.

[5] **Analyzing data across geographic scales in Honduras :detecting levels of organization within systems.** Andrew Nelson, School of Geography, UNiversitu of Leeds, Leeds.

it is often unclear whether the results of census data or Enumeration Area analysis indicate some reality about the individuals living in that region or are strictly a function of the particular areal unit used in the analysis (Openshaw 1984). It is of paramount importance to realize that when analyzing spatial data, it may be incorrect to assume that the results obtained from a study region apply equally to all individuals within that region (Fotheringham, 1997). Adopting techniques that can identify local rather than global spatial relationships, can help in avoiding this type of ecological fallacy, as can improved access to data. When spatial data form part of an analysis, it would seem prudent that the results should be in the form of a map or image as well as just tabulated results and general statistics.}[6]

A related problem concerns the display of demographic data. Maps of population density by administrative areal unit give the impression that population is distributed homogeneously throughout each areal unit, even when portions of the region are, in actuality, uninhabited. <u>One potential solution to these problems is a surface-based demographic data representation, in which data are modeled as a continuous field that is not dependent on an irregular partitioning into arbitrary areal units</u>. {Areal unit versus surface models of population can be understood via the object versus field representations of geographic reality considered in geographic information science (Goodchild1992). The object view treats population as a set of individual geographic entities to which population attributes may be attached. The field view, on the other hand, treats population as a continuously varying surface whose value (i.e., population density) may be measured at any given location. Of course, population is in reality composed of individual people; both object and field representations of population are thus abstractions of that reality. In geographic information systems (GIS), the object view is typically represented using points, lines, and polygons in the vector data model, and the field view is typically represented by an exhaustive tessellation of square grid cells in the raster data model.}[7]

Surface-based population representation offers certain advantages over areal unit representation. A surface-based representation allows for population data aggregation to nearly any desired areal unit and hence is not subject to the MAUP and other areal unit-derived problems (Bracken 1993). {In addition, because surface representations can present a graphic unit of display (a grid cell) that is uniform in size across a region, surfaces of population may offer a more accurate cartographic representation of population distribution than do conventional maps (Langford and Unwin 1994)}[8]. While most demographic data are not available in a surface-based format, raster GIS provides an environment in which to develop reliable and useful surface based representations of population and population character from aggregated census/enumeration data.

Scale effects complicate any straightforward understanding of spatial data and there is a need to explore and quantify their nature. Adopting an approach that furthers understanding about scale effects should enable greater focus on the scales that relate to the process under study. The potential for misleading or inappropriate analysis and error is magnified by the availability of more data, and users who are unaware of the problem. Aggregation of data must be a controlled and well-defined process. There is a grave danger of obtaining biased, misleading, or poor results when data for possibly inappropriate areal units are studied.

{Recent texts relating to spatial issues, spatial data integration and multi-disciplinary GIS applications such as Ecological Scale (Peterson and Parker, 1998), People and Pixels (Liverman et al., 1998) and GIS solutions in Natural

[6] **Analyzing data across geographic scales in Honduras :detecting levels of organization within systems.** Andrew Nelson, School of Geography, UNiversitu of Leeds, Leeds.

[7,8] **Generating Surface Models of Population Using Dasymetric Mapping.** Jeremy Menis, University of Colorado, The Professional Geographer 55(1) 2003 pg31-42, © 2003 by Association of American Geographers.

Resource Management (Morain, 1999) make no reference to the MAUP, suggesting that either the problem is unknown, considered unimportant, or that it has been deliberately ignored. Yet as Openshaw and Clarke (1996) adroitly put it 'It is unacceptable to assume that the MAUP does not exist'.}[9]

Much progress and understanding can be gained if the analyst is able to not only demonstrate the effects of scale and the MAUP for themselves, but can also control and define the aggregation process thus creating suitable areal units for the data (Haining, 1990; Fotheringham and Wong, 1991; Openshaw and Rao, 1995; Martin, 1998). Another option is to define new units of measurement. In this research, methods for delineating new areal boundaries have been investigated using a combination of existing GIS methodologies, new spatial analysis techniques. The paper within its limitations is a focused review of spatial analysis and has a two fold objective. The first half of the paper encompasses general spatial analytical issues and identifies their relevance to problems in urban analysis. It identifies interesting traditional spatial analytical issues of, **how spatial scale biases spatial statistics.** This stage of the analysis is both exploratory and empirical. Emphasis is placed on visualization of the data at a range of scales, to detect the variation in variables and relationships with respect to scale. Population density[10] was chosen as the variable to represent Urban Sprawl at the two different scales viz. Census Tract and Enumeration area for the spatial structures of the two largest Canadian metropolitan areas: {**Toronto**, (population 4.4 million in 1996, Canada's principal business and financial service centre), and **Montreal** (population 3.4 million, regional centre for high-order services and Canada's principal manufacturing centre.)}[11]

The second half of the paper suggests improved scale-sensitive process orientated representations and models. Here the paper attempts to address the spatial scale problem by generating surface-based representations to permit scale-sensitive and location specific analyses of population data. Interpolated raster grids and inverse distance weighted population surfaces are generated from Canadian Census Tract and Enumeration Area data for Montreal and Toronto regions. Finally the paper concludes that there is a need for more than one approach to address the issue of spatial analysis problems, based on the examples described for generating a surface-based representation of population that minimizes these problems.

[9] **Generating Surface Models of Population Using Dasymetric Mapping.** Jeremy Menis , University of Colorado, The Professional Geographer 55(1) 2003 pg31-42, © 2003 by Association of American Geographers

[10] Population density was by using Population from Census Canada 2001 for census tracts and enumeration areas of Montreal and Toronto.

[11] **A tale of four cities: intra-metropolitan employment, distribution in Toronto, Montréal, Vancouver, and Ottawa hull, 1981 - 1996.** Richard Shearmur and William J Coffey

Methodologies

The data

The key to this study was the ready availability of comparable data for these two cities under study. The data required for this study was:-

- "e00." Vector maps for CMA areas and EA areas of Montreal and Toronto. From Census 2001'Geographic Files
- Census Tract Data about Population (100% sample) for the year 2001 From CHASS website: - http://www.chass.utoronto.ca
- Enumeration Area Data about Population (100% sample) for the year-2001 and Centroid Point maps, (Montreal / Toronto, EA/CT) for raster generation was provided by Prof. Murtaza Haider , Assistant Professor, School of Urban Planning , McGill University.

The software environment used to analyze this data was Microsoft excel and ArcMap.

Methodologies

For the first half of the paper, which concentrates on the spatial scale factor described in brief in the preceding pages, two distinct approaches were used to examine the data generated from analyzing the census Tracts (henceforth referred to as CT) and Enumeration Areas (henceforth referred to as EA):

- Comparative analysis using Histograms of the "population density" , Area, Population etc from CT's and EA's of these two cities, and
- A vector analysis based on concentric rings in Arc Map.

It is clear that only statistics, tables, graphs and traditional exploratory analysis are not enough to describe the complex, scale-dependent relationships that exist in geographic data. From that viewpoint, it becomes imperative that any analysis pertaining to be spatially explicit must also have results that are mappable. For this reason, statistics, table's graphs etc and primarily maps represent the majority of the examples in the paper. Neither approach, on its own, provides a complete picture, but the combination of both approaches, although does not necessarily capture the full complexity of spatial patterns, provides a certain depth of vision.

The population density for each CT and EA was calculated in each census period as :-
(Population of CT / Area of CT) and (Population of EA / Area of EA).

Excel tables containing data gathered for CT and EA both for Montréal and Toronto were then combined with the respective shape files, in ArcMap to do a vector based analysis of the problem.

1. Population Density was projected for each census period on the shape files of respective cities.
2. Concentric rings with the CBD's as the center and each expanding with a 5 kms interval were drawn on the GIS map for the census periods.
3. CT's and EA's lying within these concentric rings namely, 0 to 5 kms from the CBD, 5 – 10 kms , 10 – 15 kms and so on up to 35 kms, were selected using the selection tools in ArcMap and exported in dbf format in excel to calculate the average population density.

These steps were performed for both cities to obtain the data in the following simple format:-

MONTREAL - AVERAGE POPULATION DENSITY - 2001				
		CT - LEVEL		EA LEVEL
0-5 KMS		11097		24505
5 -10 KMS		7708		12390
10-15 KMS		4015		10399
15-20 KMS		2436		10179
20-25 KMS		1933		3230
25-30 KMS		1484		2945
30-35 KMS		921		4736
35 +		709		4798

TORONTO - AVERAGE POPULATION DENSITY - 2001				
		CT - LEVEL		EA LEVEL
0-5 KMS		11221		157473
5 -10 KMS		6583		151279
10-15 KMS		5663		172348
15-20 KMS		4261		111004
20-25 KMS		3276		138395
25-30 KMS		2523		96382
30-35 KMS		2229		77050
35 +		1448		40409

A comparative study and a cursory analysis of these data for Toronto and Montreal, by line graphs and histograms in the following pages along with the GIS maps brings out interesting differences, as one can already see between the statistics generated by analyzing the two cities at different scales. findings and results in the following pages.

An analysis of whether the centroid of the EA's lies within the CT's to which they belong was also done for <u>some random polygons</u> for both the cities. The Population counts of the EA's which fall within a particular CT were added and compared with the count of the same CT. The total population was found to be similar (within negligible errors) to that of the CT.

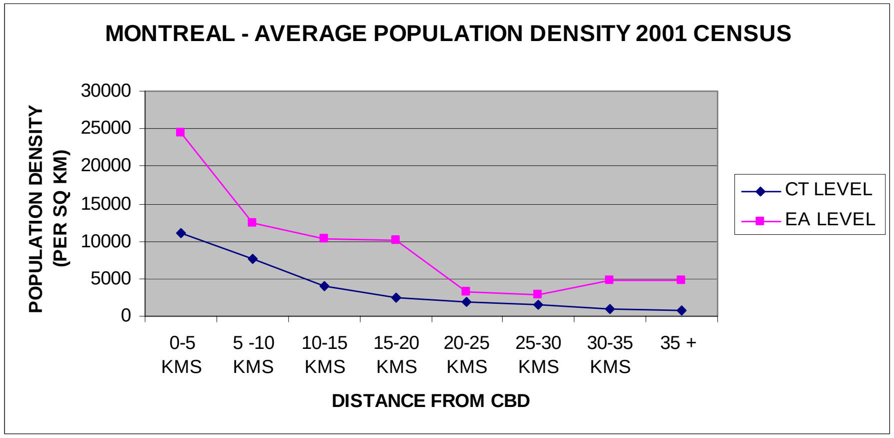
Observations and Findings

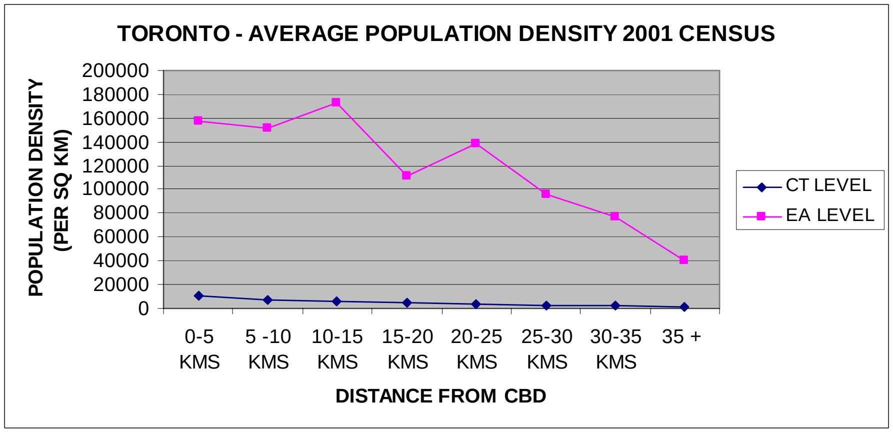
One of the most common approaches to quantifying density is using a density gradient. {The idea of measuring how density of urban activity declines along a gradient with growing distance from a designated center has been around for some time. Density gradients are potentially useful indices of sprawl for several reasons: they permit comparisons over time and between cities, they incorporate crucial elements of urban land use, and they overcome some traditional constraints in the measurement of urban densities. If a population density gradient falls over a specified period, for example, we may say that the urban area has sprawled—in relative terms—over that time. Likewise, the gradient measure allows us to make comparisons between cities and to gauge the relative degree of sprawl between them. A city with a small population (or perhaps employment or household) density gradient can be said to be more sprawling in its relative density than a city with a comparatively larger gradient. } [12]

Simple density gradient graphs using the average population densities both for Montreal and Toronto (Graphs – Next page) brings out the difference between the statistics generated at the two different spatial scales. The Population densities calculated at the EA level are almost 1.5 to 3 times higher in the case of Montreal while for Toronto they are 8 to 25 times higher.

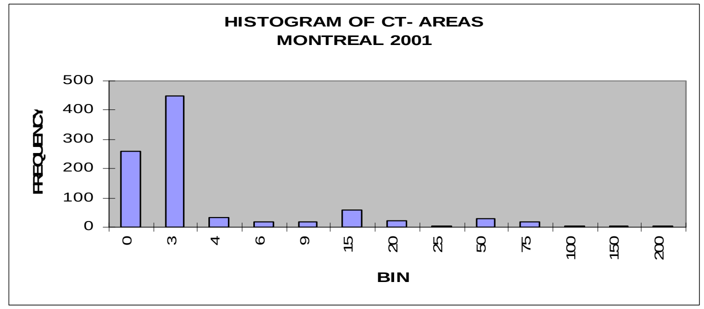
Censuses and other data collections base their statistics on reporting zones or like CT's or EA's which tend to vary in size and shape. Histograms of the Land Area of CT's and EA's were plotted to better understand this problem, for both the cities. (Histograms - next page).

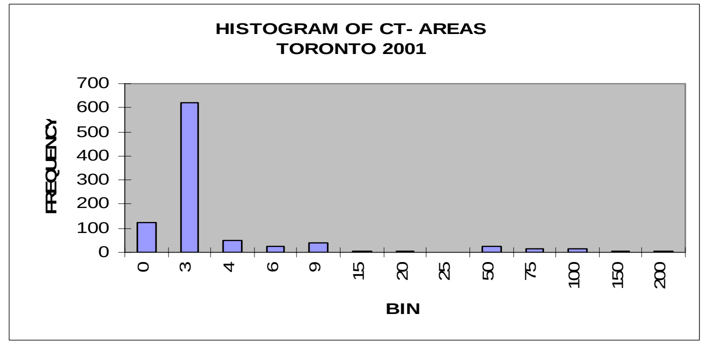
12 **Measuring Sprawl, Paper 27, Paul M. Torrens, Marina Alberti,** November 2000
Center for Advance Spatial Analysis , University College London
http://www.casa.ucl.ac.uk/working_papers.htm

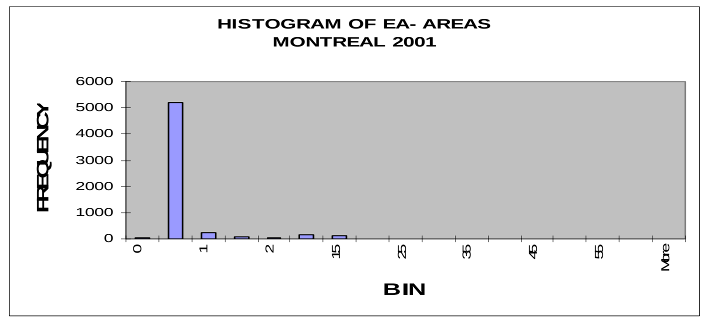


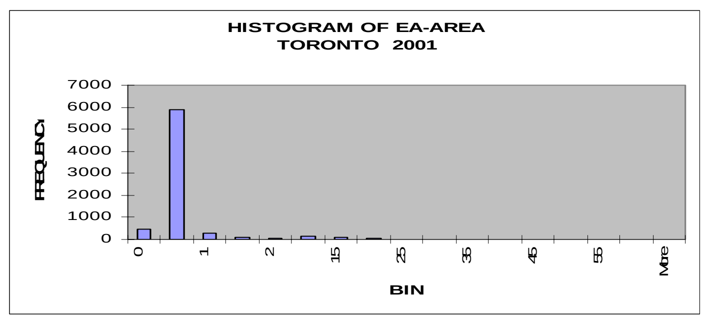


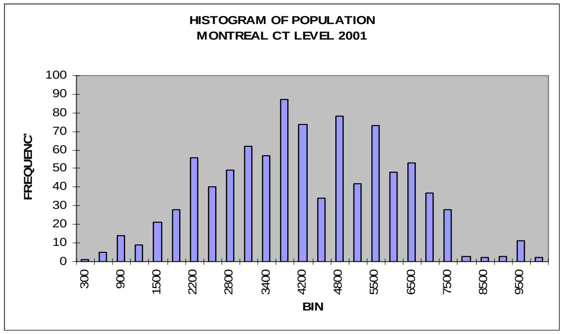
As one can see form these Histograms, most of the CT's have an area between 1 – 3 Sq Kms both for Montreal and Toronto, while The EA's have a land area of less than 1 sq Km, some of them being as small as the foot print of a building which could be the geographic area canvassed by one census representative. (ref. Notes for definitions of CT's and EA's.) These small areas of EA's go a long way in shooting up the average Population densities of the cities under analysis and thereby leading to inconsistent and misleading statistical and analysis.

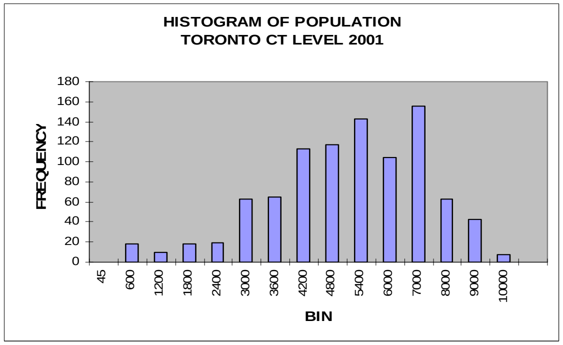


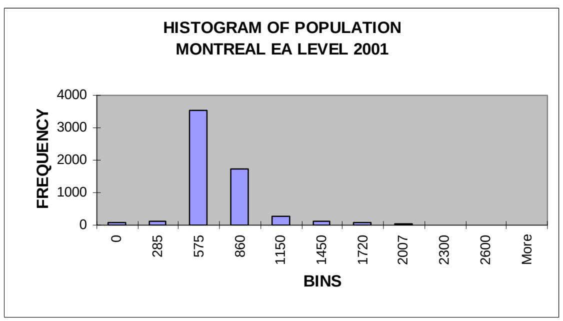


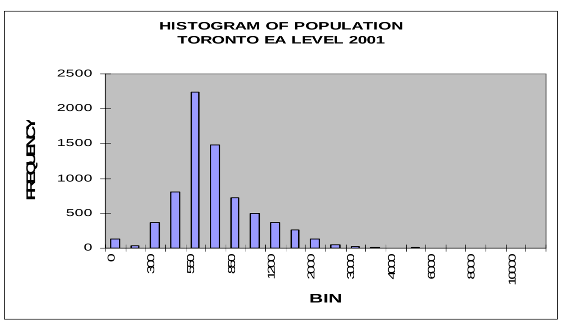


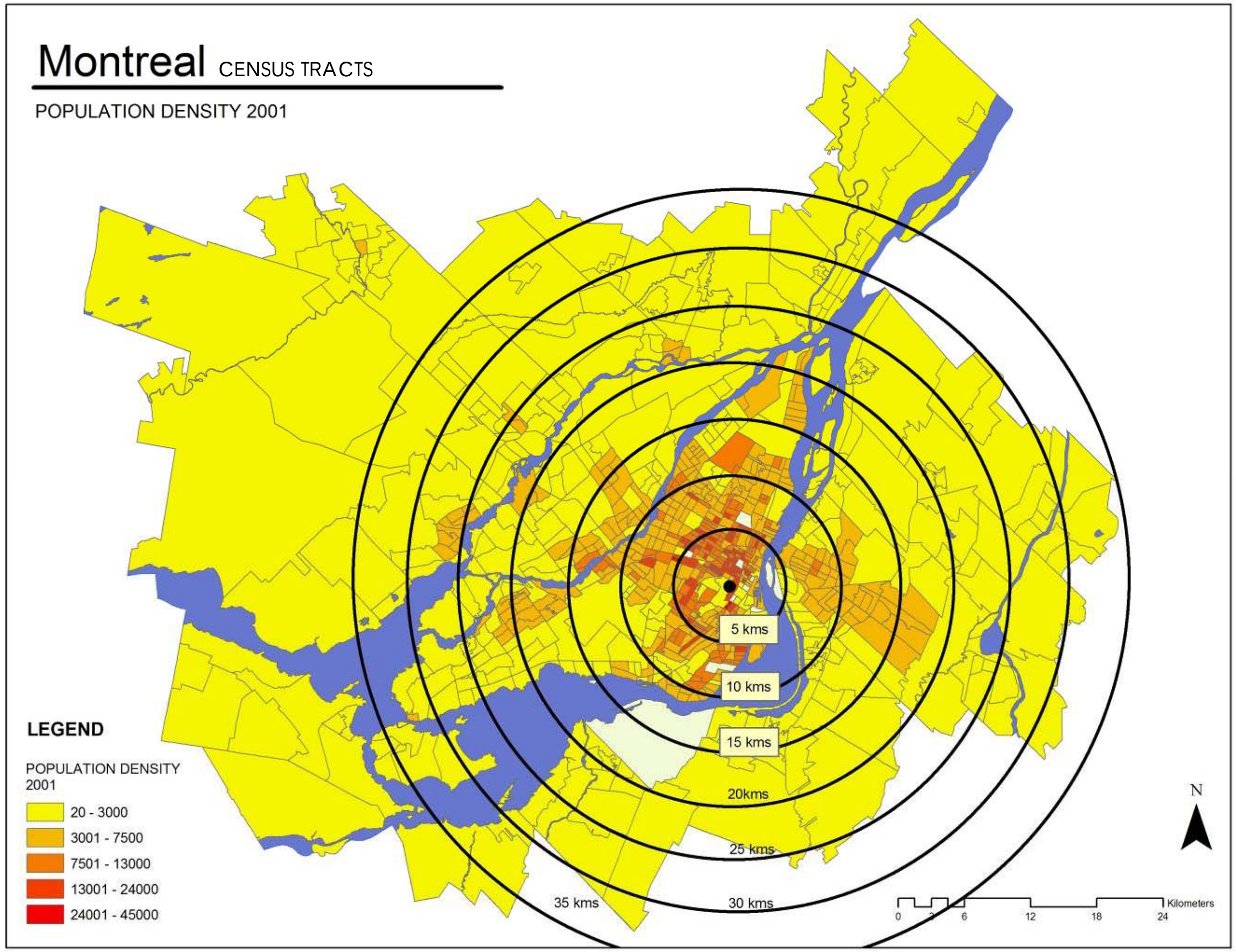






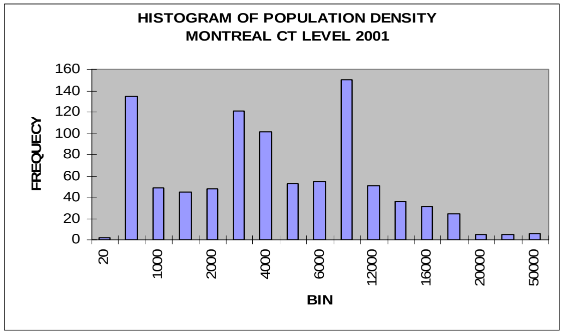


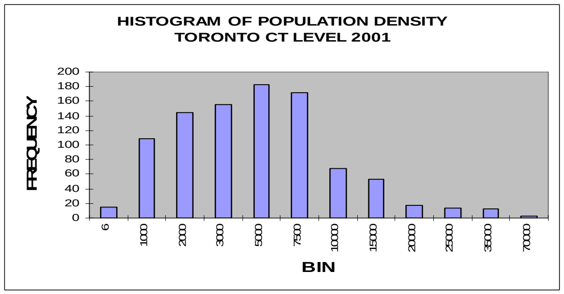


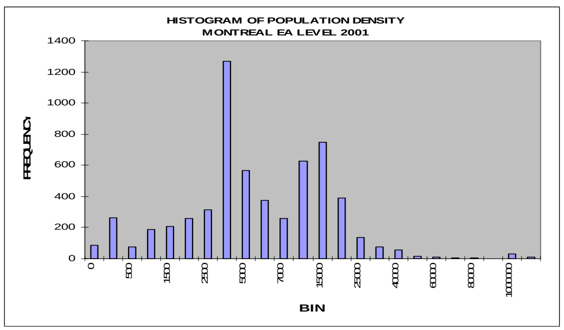


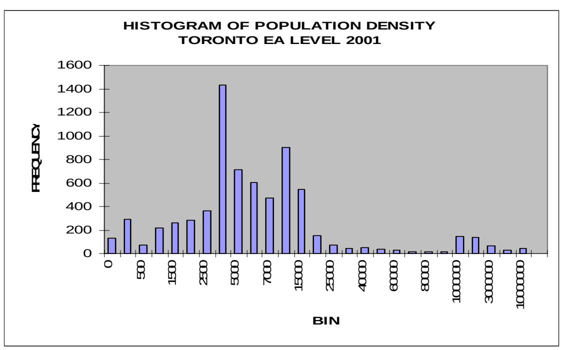
A second set of Histograms for the Population Counts within these spatial units were plotted for these two cities. Most of the CT's for Montreal and Toronto had a population count within the Range of 2000 – 7500 people. (Ref. definition of CT in Notes)

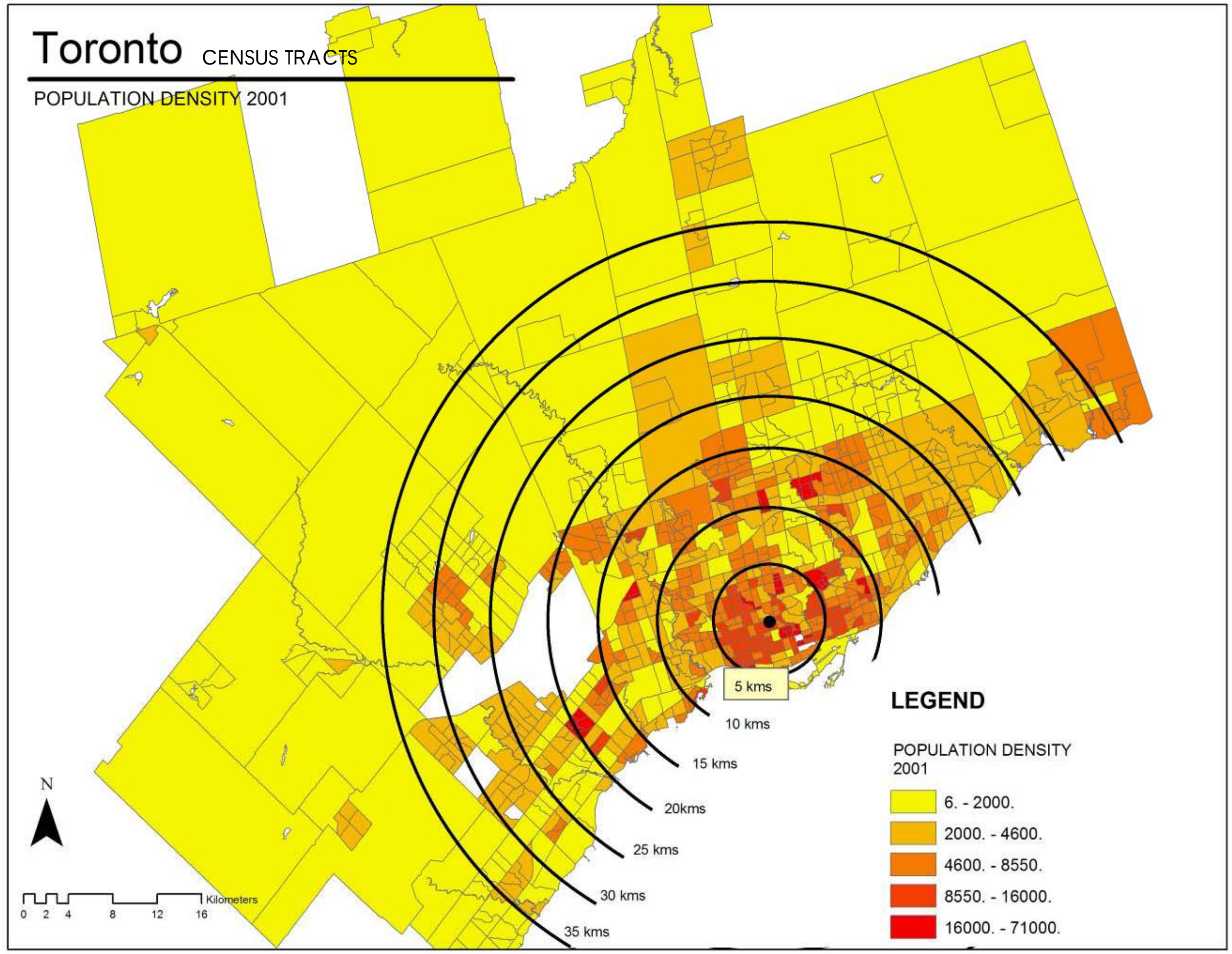
However most of the Population counts for these two cities were within the range of 550 – 900 people per EA. This clearly explains the ballooned population densities and the incoherent spatial statistics. The histograms of the population densities on the next page will further highlight this problem of spatial scale.





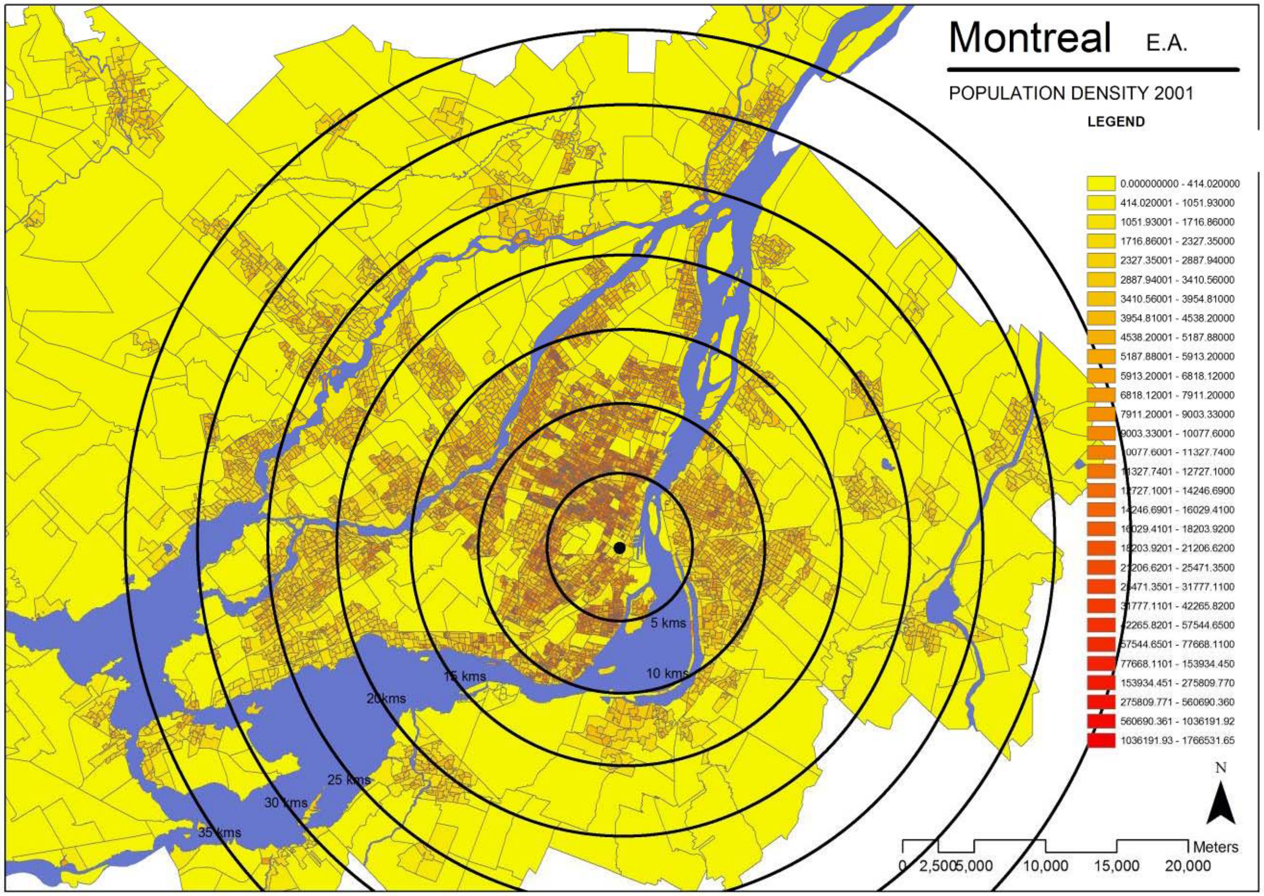






While these histograms may show that most of the observations of the population density calculated at the two different spatial scales lie with in the same range (1000 – 12000 persons per Sq kms) for Montreal and (1000-15000 persons per Sq Kms for Toronto), a few observations of very high densities in the range of 1000,000-10,000,000 persons per sq kms) are observed in the Histograms of Montreal and Toronto at EA level. Such results of densities profiles are Impossible and have an impact the mean and variance of the Data Set.

Increases in personal mobility, manufacturing and service decentralization, and the fission of contemporary lifestyles all provide powerful shifts in the locus of urban development. Urban density profiles, focused upon historic central business district (CBD) areas, nevertheless remain centrally important to urban analysis for a number of reasons. {First, cities historically grow outwards from their central seed sites, and built form freezes the social processes that led to the production of the built environment. Second, the quest for more sustainable urban futures implies a refocusing of functions back towards city centers. And third, an urban density profile provides a transect across the maximum spanning distance of most urban settlements which is likely to be representative of all stages of urban growth.}[13] Here, however, the priority was to bring out the difference in the statistics generated for the same variable over two different spatial scales, for Montreal and Toronto.



It is obvious from the results that the regions of Montreal and Toronto provide an example in which EA/ CT area and population density vary significantly from urban core to the to suburbs of the cities . EA/ CT areas groups in urban areas are relatively small and of homogeneous population density, while in suburbs they are typically much larger, and have a much more heterogeneous population distribution. Areal Interpolation methods (described in the next few pages) I feel should therefore be used to generate a surface model that provides a more accurate representation of population within suburban block groups, as well as in urban block groups that contain parks, cemeteries, and other features that may control the within block group distribution of population.

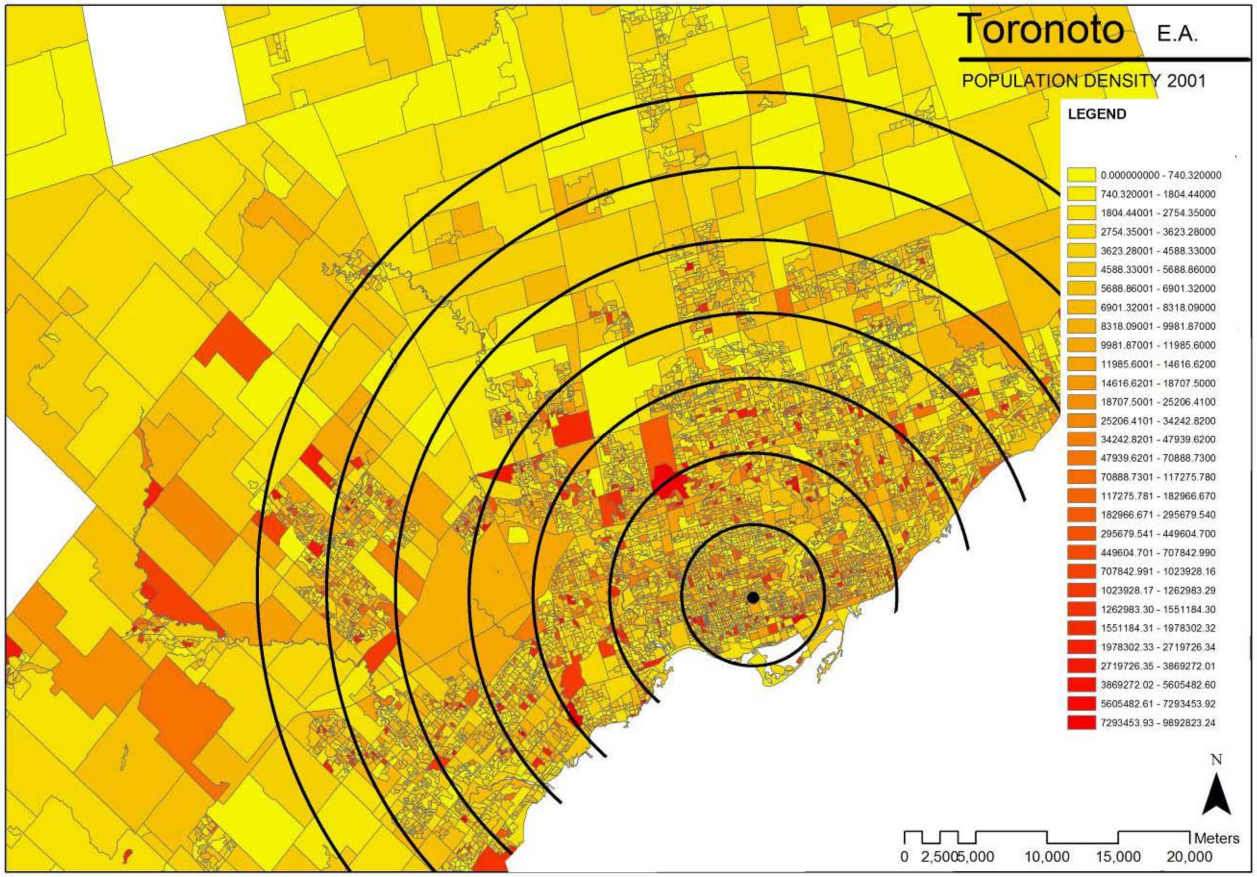
---

[13] **On the measurement and Generalization of urban form.** Paul Longley (School of Geographical Sciences, University of Bristol, , UK. and Victor Mesev School of Environmental Studies, University of Ulster.

Solutions

The problem of creating a population surface from areal unit data is essentially one of **areal interpolation,** the transformation of geographic data from one set of boundaries to another. {Areal interpolation is typically used to compare two or more spatial datasets that are stored in incompatible areal units, such as Enumeration area and census tracts. In a sense, the generation of a raster population surface is a special case of areal interpolation, because the desired (target) areal unit (a raster grid cell) is intended to approximate a continuous surface; hence, it is necessarily much smaller than the size of the original areal unit of data aggregation.

The most straightforward technique for transformation to raster format is **areal weighting,** whereby each grid cell (size specified by the analyst) is assigned a population value based on its percentage area of the host areal unit. In this method maintains the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units—i.e. ‘‘[P]eople are not destroyed or manufactured during the redistribution process’’ (Langford and Unwin, 1994, 24).



A more sophisticated approach to developing surface representations of demographic data for census, enumeration areas is that of **Inverse distance weighted interpolation**.[14] Population counts are assigned to a set of summary points generated from the centroids of the original areal units. A moving window operation over an ‘‘empty’’ raster grid then assigns to the window kernel a value according to the population values of those centroids contained within the window, with closer centroids having more ‘‘weight’’ than those centroids farther away. The relative density of centroids around the kernel determines the size of the window. This method assumes that population density decreases away from the centroid according to some distance decay function and allows for some areas of the raster surface to contain zero population.}[15]

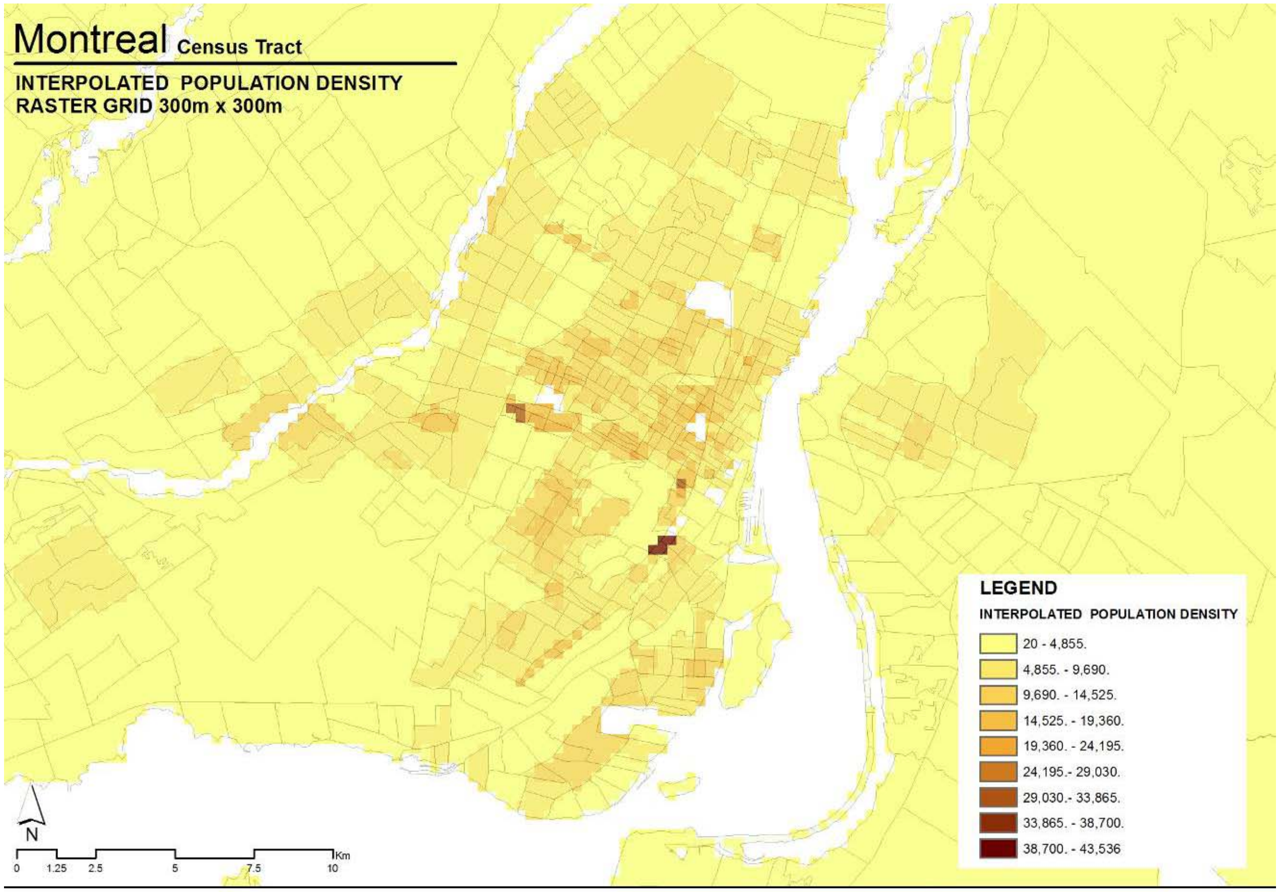
---

[14] Refer Notes

[15] **ArcGIS, Spatial Analyst Tutorial , ESRI Pg 132-141 and Spatial Data Analysis, Theory and Practice,** Robert Haining, University of Cambridge , 2004.

Three raster grids were then created by converting the vector CT / EA coverage to a raster grid for each city, for the areal weighting technique.
While a 300 sq m grid was used for studying the Census tracts two grids 300 sq m and 600 sq m were applied to the EA coverages.

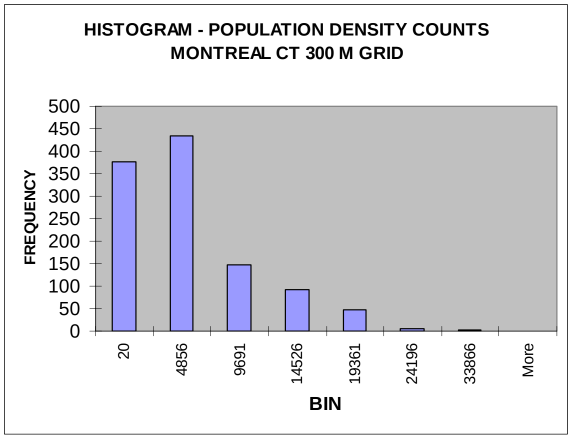
The adjoining map shows the raster population surface for Montreal, CT. While in the urban core areas it does not appear to differ significantly from the vector CT map, in the urban areas where there are parks or cemeteries, this map is significantly more detailed. This is especially evident in the south west and northern part of the city where the vector representation indicates a relatively homogeneous population density, whereas the raster surface has concentrated that population in certain sub-block group regions. In suburban areas, block groups are typically much larger than in urban areas, and these block groups may contain urbanized areas within which a majority of that block group's population resides. These Raster maps redistributes population within such block groups. This redistribution of population is particularly evident in the central parts of the city, where population in the raster surface is clearly concentrated in a number of sub-block group-sized high urbanization areas. This raster approach however does not allow for a lot of



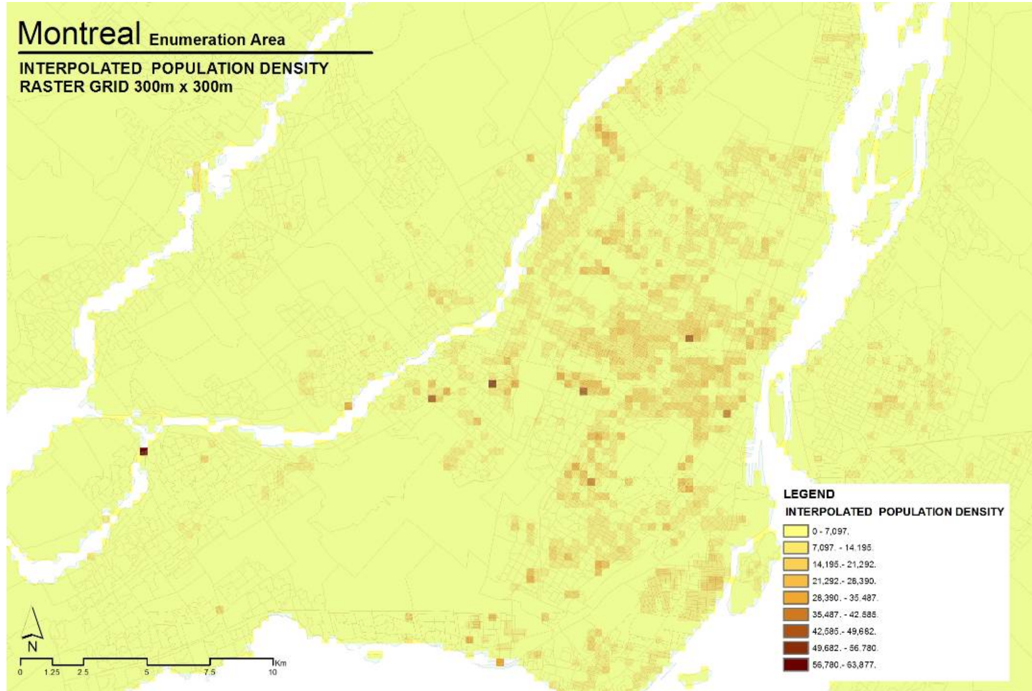
quantitative analysis, as it gives us only the values of each cell and its count. This approach also does not let us do a multiple buffer analysis to get **means** of the population density.

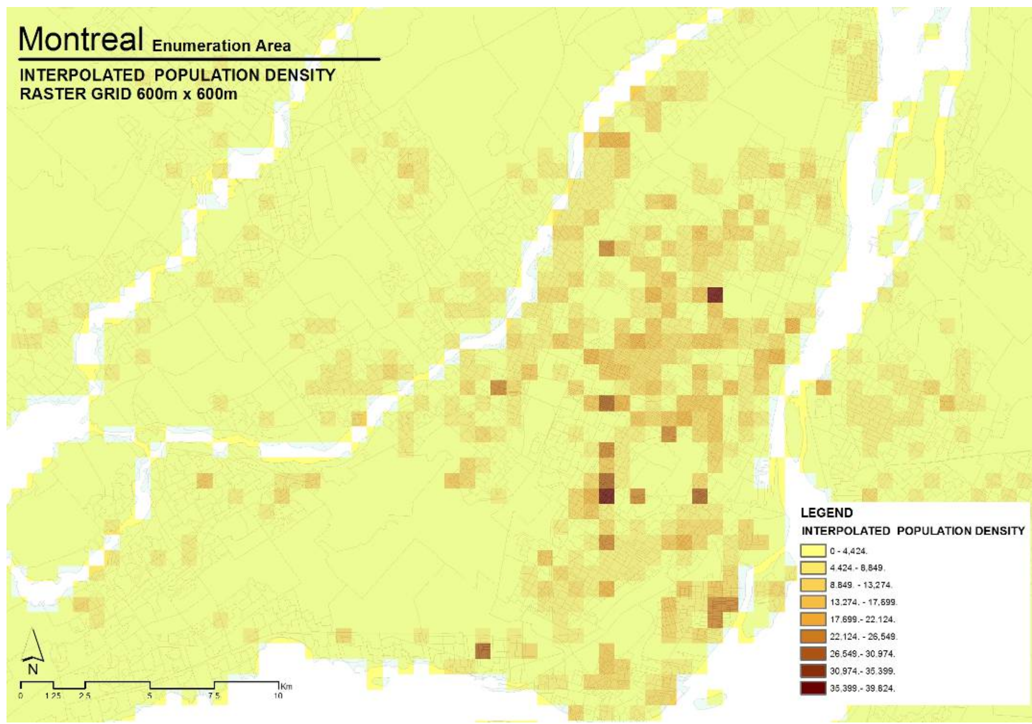
MONTREAL CT 300M X 300M		
POPULATION COHORTS	LAND AREA SQ M	% AREA
20-4885	3817.08	95.40
4856 -9690	138.33	3.46
9691-14525	33.93	0.85
14526-19360	9.99	0.25
19361-24195	0.9	0.02
24196 -33865	0.36	0.01
33866-43700	0.45	0.01
	4001.04	

The % of land area under each population cohort was calculated. The chart shows that 95 % of land area in Montreal CMA region has a population density between 20 – 5000 persons /sq km. Very few concentrated areas in the city 0.01 % boast of higher population density counts. The histogram of the raster grids also correlates to this observation.



MONTREAL EA 300M X 300M		
POPULATION COHORTS	**LAND AREA SQ M**	**% AREA**
0-7097	3866.10	96.59
7097 - 14195	102.63	2.56
14195 - 21292	24.57	0.61
21292 - 28390	6.56	0.16
28390 – 42585	1.27	0.03
42585 - 49682	0.54	0.013
49682 - 63877	0.9	0.02
	4002.57	

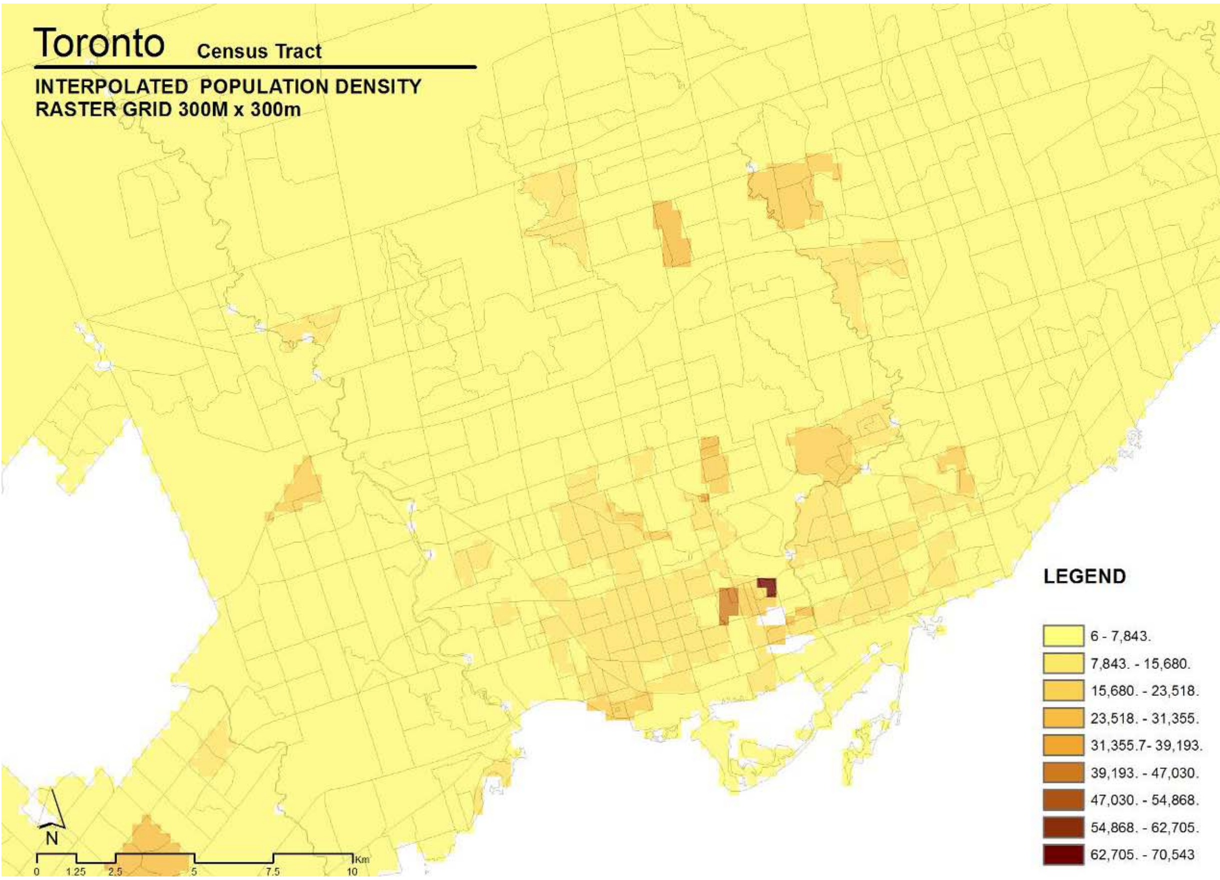


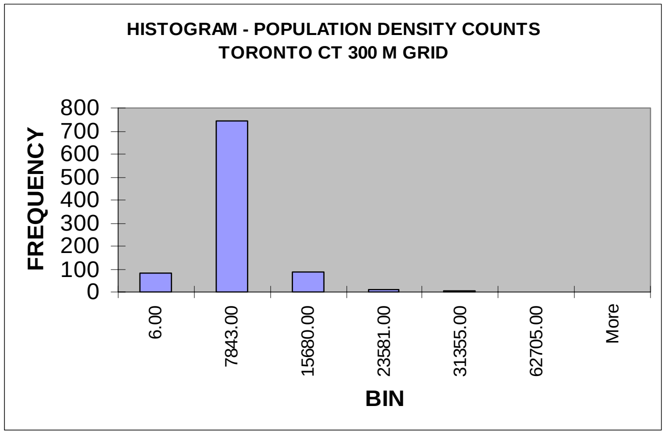


MONTREAL EA 600M X 600M		
POPULATION COHORTS	LAND AREA SQ M	% AREA
0 - 4424	3824.61	95.55
4424 - 8849	134.33	3.35
8849 – 13274	32.63	0.81
13274 - 17700	8.75	0.21
17700 - 22125	0.9	0.02
22125- 31000	0.63	0.015
31000- 39825	0.72	0.017
	4002.57	

Similarly, land area calculations were done for both 300 and 600 sq m grids generated using EA block for Montreal give us comparable results. What is of interest to note here is that while a 300 sq m grid gives us higher population density, a 600 m grid gives us results almost similar to that generated by the 300 m grid of CT. The possible reason for this phenomenon is that a 600 m grid is synonymous to the aggregation of EA's to make a CT.

TORONTO CT 300M X 300M		
POPULATION COHORTS	LAND AREA SQ M	% AREA
6--7843	4850.28	98.38
7843 -- 15680	55.17	1.11
15680 --23518	13.95	0.28
23581 --31355	9.90	0.20
31355 --54868	0.72	0.07
62705 –70543	0.27	< 0.01
	4930.29	

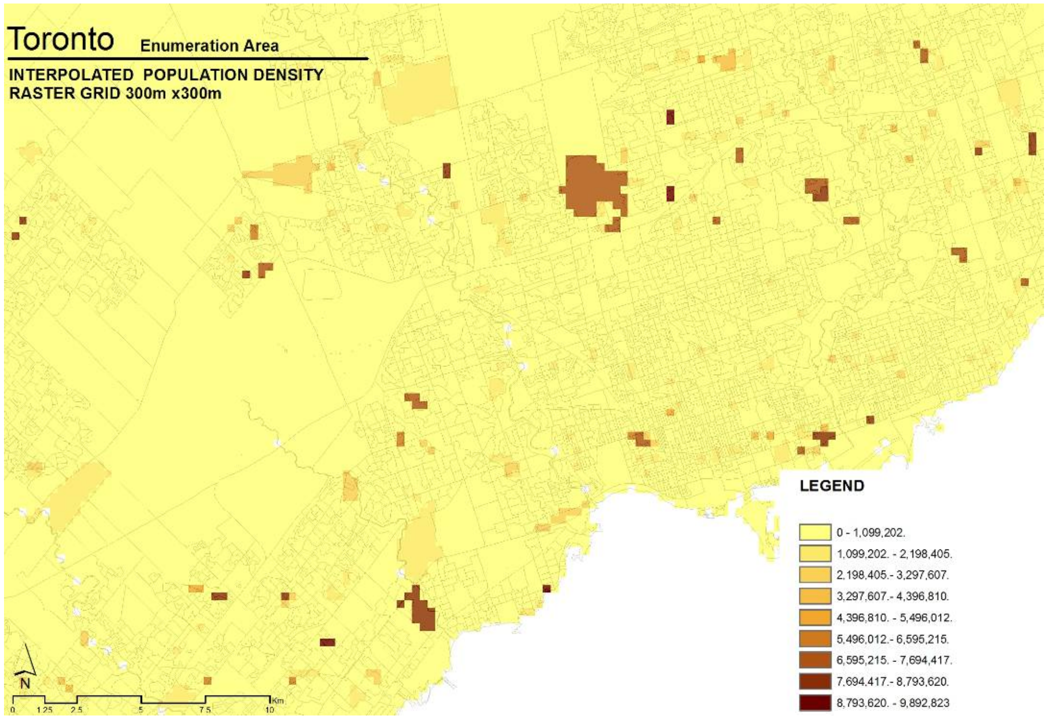


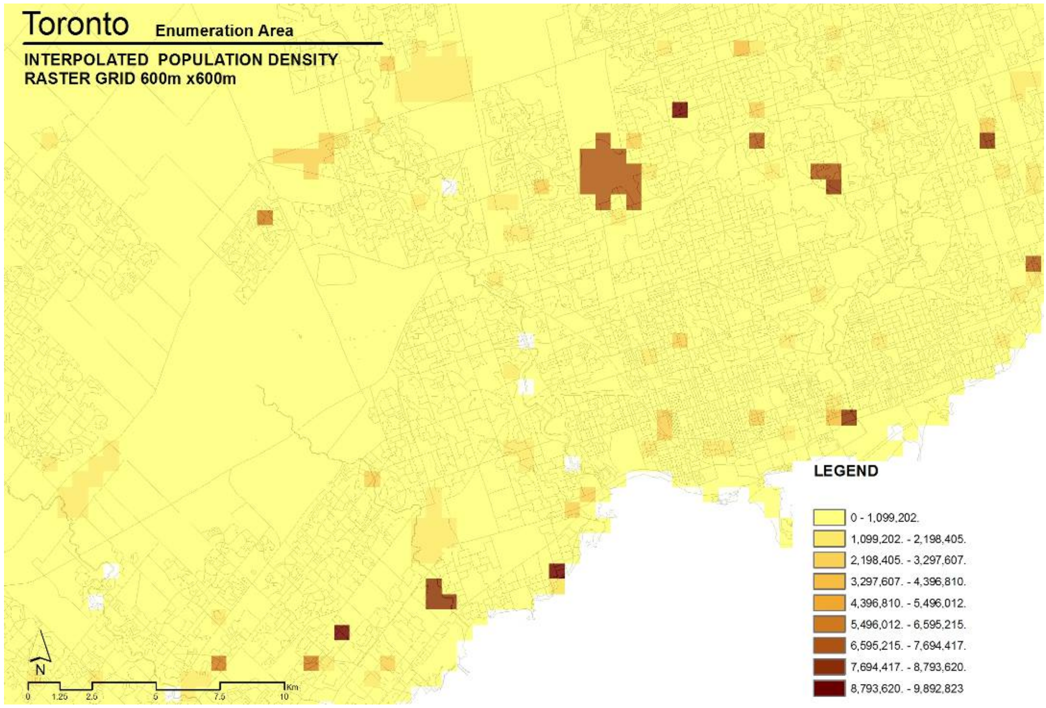


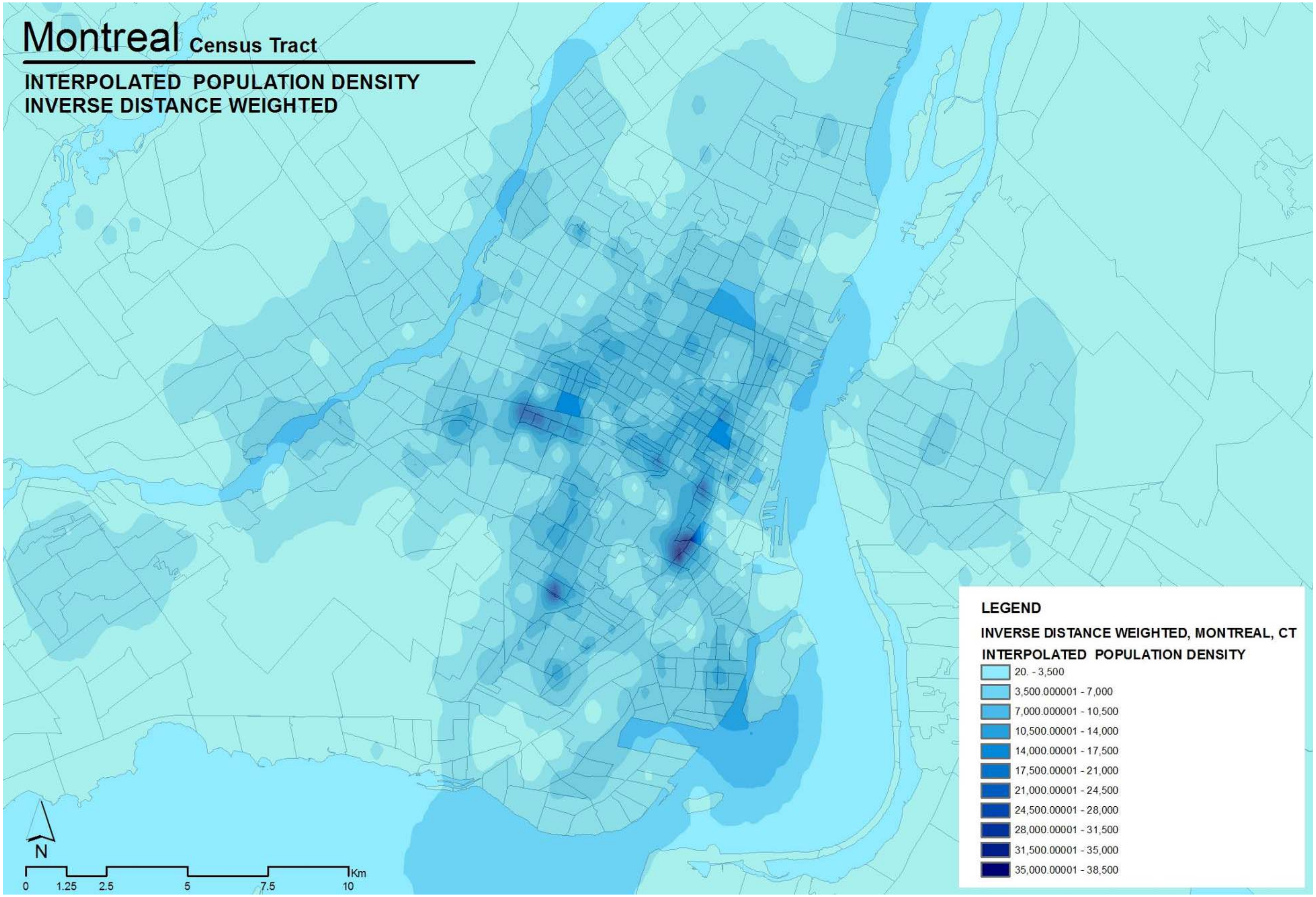
Similarly a 300 m grid for Toronto was drawn for CT block groups. The map again brings out significant details which are missed in a vector representation. Histograms and Land Area chart shows that the average range of population is higher than that of Montreal and at places as high as 70,000 persons per sq m. However these places cover less that 0.01% land area of the CMA region of Montreal. 98% of the land area of Toronto has a population density range of about 6 – 7850 persons per sq km.

Similar maps when drawn for Toronto, EA blocks do not give any meaningful results. Again, like for vector EA maps extremely small enumeration areas, like high rise apartment buildings are responsible for these results.

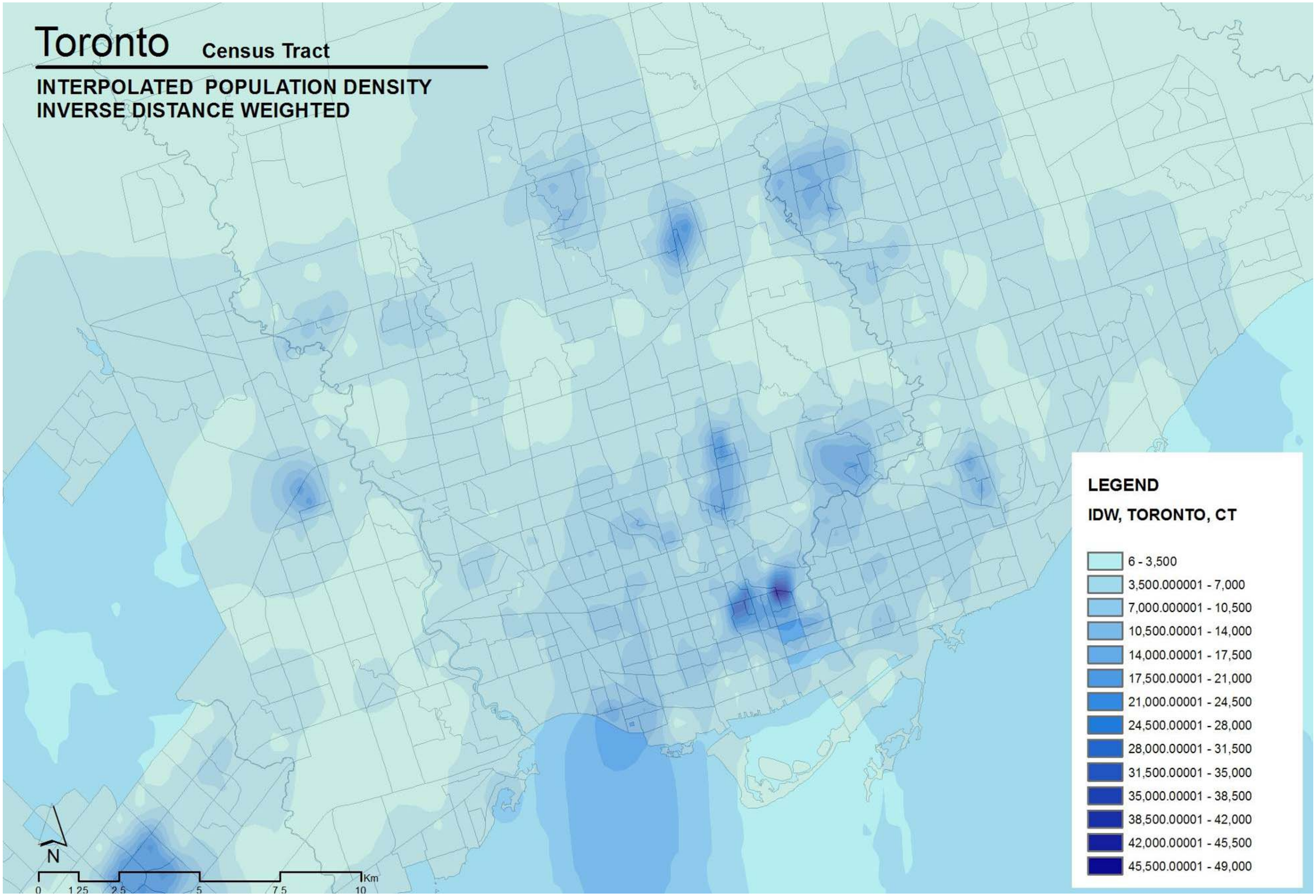
<u>Such raster maps help us to do a micro analysis of a city as they capture significant details missed which are overlooked in a typical vector representation. Also we can clearly see that there is no distinct rule of thumb that can be followed for calculating population density and perhaps ever other socio demographic statistic for cities.</u>



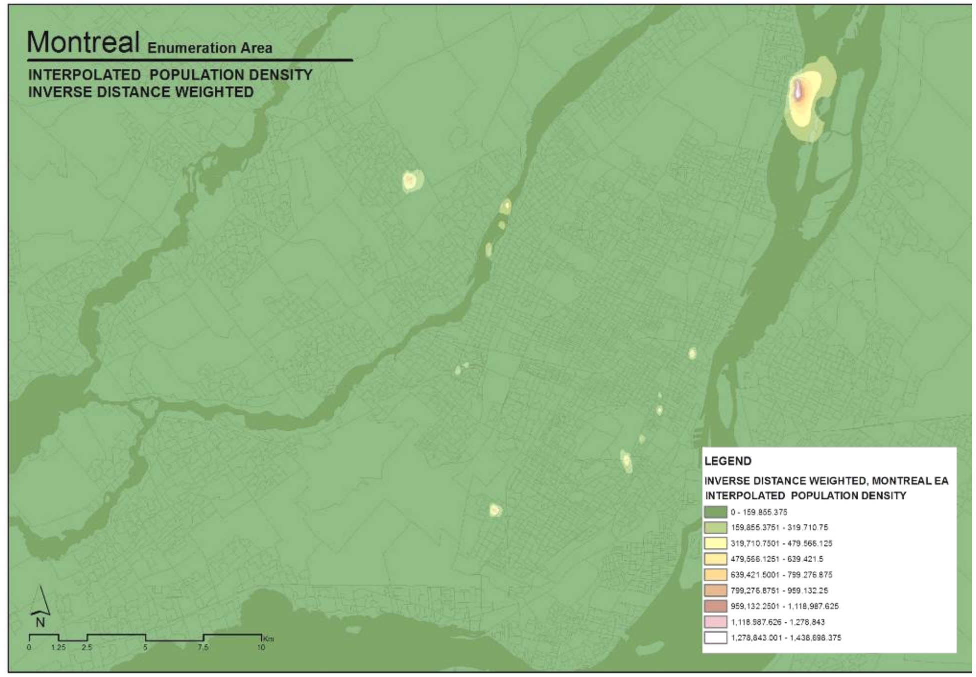




Montreal Census Tract
INTERPOLATED POPULATION DENSITY
INVERSE DISTANCE WEIGHTED
LEGEND
INVERSE DISTANCE WEIGHTED, MONTREAL, CT
INTERPOLATED POPULATION DENSITY
20. - 3,500
3,500.000001 - 7,000
7,000.000001 - 10,500
10,500.00001 - 14,000
14,000.00001 - 17,500
17,500.00001 - 21,000
21,000.00001 - 24,500
24,500.00001 - 28,000
28,000.00001 - 31,500
31,500.00001 - 35,000
35,000.00001 - 38,500
N
0
1.25
2.5
5
7.5
10
Km


Toronto Census Tract
INTERPOLATED POPULATION DENSITY
INVERSE DISTANCE WEIGHTED
LEGEND
IDW, TORONTO, CT
6 - 3,500
3,500.000001 - 7,000
7,000.000001 - 10,500
10,500.00001 - 14,000
14,000.00001 - 17,500
17,500.00001 - 21,000
21,000.00001 - 24,500
24,500.00001 - 28,000
28,000.00001 - 31,500
31,500.00001 - 35,000
35,000.00001 - 38,500
38,500.00001 - 42,000
42,000.00001 - 45,500
45,500.00001 - 49,000
N
0
1.25
2.5
5
7.5
10
Km

The second approach used to map population density as a surface was the Inverse *distance weighted interpolation* method. This approach captures minute details and groups similar pixels to generate a surface. Population density count for each pixel can be calculated by simply clicking over the desired location. These surface maps correspond, both visually and in terms of the data to the maps generated by the areal weighting technique. However these maps are visually more communicative as they generate a smoother surface and project a more realistic image of the phenomenon under study. <u>The barrier feature (which limits the surface over a geographical region failed to work and as a result data is projected even on the geographical features such as lakes and rivers.</u>



This approach did not however give meaningful results when applied both to Montréal and Toronto EA level block groups. Again the reason for this phenomenon as we have seen the small zones of very high population densities of EA's which give extremely misleading points to the surface calculator in ArcGIS, and skew the inverse distance weighing algorithm.

It would be interesting to know again, the % of land cover under each population cohort generated in each of these maps. Unfortunately these maps though give us the pixel size do not give us its respective counts.

Conclusions

This research has presented a series of techniques and tools that allow spatial data sets to be constructed and de-constructed in a generalized, yet context sensitive manner. I feel, that that the outputs from such techniques should be explored and described through various user-defined levels, before making a generalization, thereby revealing the spatial patterns and processes that are arguably more useful than raw data or standard representations. Further, hypotheses and models can be developed based on the improved understanding that such mapping techniques provide.

Additionally, the opportunity to re-express the data at different levels — **levels appropriate to the analyst** — enabling conflicts to be rapidly highlighted and the effects of a decision at one level to be visualized at other levels of organization. This research demonstrates the use of the spatial analyst tool in ArcGIS to create raster surface representations of population density. Even though Population density as a variable for Urban Sprawl was studied for two examples over two spatial scales, these techniques can be applied to most common GIS data types, and operate very rapidly across several scales. The cross-scale analysis/results that are generated are a method of attaching a degree of confidence to spatial data by assessing their scale dependence and ability to faithfully represent spatial patterns.

The two techniques (viz. **areal weighting** and **Inverse distance weighted interpolation**) that are used to map population density as a surface phenomenon for Montreal and Toronto at the two different spatial scales may be improvements in the previous methods of surface Mapping. While Areal weighting (Raster Grids approach suggests an empirical sampling approach to assessing population density. {This approach does not provide a predictive model of this relationship, Langford and colleagues(1991) have shown that the derivation of such models is made difficult due to the spatial variation in the nature of land use, urbanization, or other land-cover-based classifications as they relate to population density. }[16]

However this approach does let us empirically study the land area under the population density assigned to a particular land cover class based solely on the subjective decision of the analyst. The obvious drawback to the approach described here is that it assumes that at least some of the original areal units of population are small enough to be contained entirely within the area of each EA /CT, so that the population density of the EA/CT may be sampled. This drawback however, may be countered by reducing the size of the grid.

Inverse Distance weighted technique while did not allow an empirical analysis, was very communicative visually for both cities under analysis. Even though spatial clusters were generated, information for every interpolated pixel for the data projected could be obtained by simply pointing the mouse over a desired region.

The two approaches used to generate surface were neither perfect nor all encompassing but they did address some of the important issues and problems that plague many spatial analysis applications and leave us at a point where we can:-

- {Quantify and visualize the spatial drift within a data set.
- Produce better representations and hence understanding of local phenomena where spatial variation is found to be significant.

---

[16], [17] **Generating Surface Models of Population Using Dasymetric Mapping.** Jeremy Menis, University of Colorado, The Professional Geographer 55(1) 2003 pg31-42, © 2003 by Association of American Geographers.

- Help to generate new hypotheses and experiments based on this understanding.
- Automatically define regions and system boundaries for further analysis.}[17]

The identification of appropriate scales for analysis and prediction is an interesting and challenging problem. Although the factors producing scale-dependent patterns may yet, not be clearly understood, accurate and reliable descriptions of scale dependent patterns and processes are required, to design data sampling procedures and test the accuracy and reliability of methods of prediction. There is clearly some way to go before scale effects can be fully understood and accommodated, but this research has aimed to be a 'next step' in that process. These techniques may be easily incorporated within spatial analysis using raster GIS packages. While I demonstrated these techniques here using ArcGIS, most GIS packages that support vector and raster data handling also support the calculations necessary for these techniques using simple drop-down menus or push-button controls. In other words, programming is not required to implement these techniques in ArcGIS and most other GIS packages. In addition, the techniques described here are generalizable to a variety of settings. While I used urbanization data (Montreal and Toronto) for the spatial analysis, one could easily use these techniques with other categorical ancillary data that have a demonstrable spatial relationship with the distribution of population.

{One way of assessing the importance of scale effects is to document the effects by reporting results at different levels of data manipulation. However, great care must be taken to ensure that these levels are context specific and not imposed on the data a priori. Such context specific reporting can be made easier by the increased use of techniques such as those presented here.}[18]

[18] **Generating Surface Models of Population Using Dasymetric Mapping.** Jeremy Menis, University of Colorado, The Professional Geographer 55(1) 2003 pg31-42, © 2003 by Association of American Geographers.

{Census Tracts (CT):-

A permanent, small neighborhood-like or rural community-like area established in large urban centers with the help of local specialists interested in urban and social science research. Census tracts are delineated jointly by a local committee and Statistics Canada. **The population must be between 2,500 and 8,000, with a preferred average of 4,000 persons (except for those CTs in central business districts, in other major commercial and industrial zones, or in peripheral rural or urban areas that may have either a lower or higher population).** Also, when first delineated or subsequently subdivided, CTs must be as socio-economically homogenous and compact in shape as possible, and follow permanent, easily recognizable physical features.

All CMAs and CAs in Canada containing a CSD having a population of 50,000 or more at the previous census are eligible for a census tract program. Census tracts cover all 25 CMAs and 14 of the 115 CAs.
There are 4,068 CTs in Canada.

Enumeration Area (EA):-

The geographic area canvassed by one census representative. The number of dwellings in an EA generally varies between a maximum of 375 in large urban areas to a minimum of 125 in rural areas. An EA always respects higher level geographic areas recognized by the census, and is the smallest geographic area for which census data are available. EAs are defined solely by Statistics Canada. There are 45,995 EAs in Canada. }[19]

Interpolation

{*Interpolation* predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data: elevation, rainfall, chemical concentrations, noise levels, and so on. The left-hand graphic above is a point dataset of known values. The right-hand graphic is a raster interpolated from these points. Unknown values are predicted with a mathematical formula that uses the values of nearby known points.

Fig: Point dataset of known values.

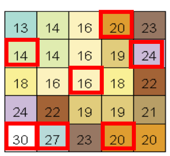


Fig: Raster Interpolated from the points. Cells Highlighted in Red Indicate the values of Input Point Dataset.

Visiting every location in a study area to measure the height, magnitude, or concentration of a phenomenon is usually difficult or expensive. Instead, strategically dispersed sample input point locations can be selected and a predicted value can be assigned to all other locations. Input points can be either randomly or regularly spaced points containing height, concentration, or magnitude measurements. The assumption that makes interpolation a viable option is that spatially distributed objects are spatially correlated; in other words, things that are close together tend to have similar characteristics. For instance, if it is raining on one side of the street, you can predict with a high level of confidence that it is also raining on the other side of the street. You would be less sure if it was raining across town and less confident still about the state of the weather in the next county. Using this analogy, it is easy to see that the values of points close to sampled points are more likely to be similar than those that are further apart. This is the basis of interpolation.}

Inverse Distance Weighted (IDW)

IDW estimates cell values by averaging the values of sample data points in the vicinity of each cell. The closer a point is to the center of the cell being estimated, the more influence, or weight; it has in the averaging process. This method assumes that the variable being mapped decreases in influence with distance from its sampled location. For example, when interpolating a surface of consumer purchasing power for a retail site analysis, the purchasing power of a more distant location will have less influence because people are more likely to shop closer to home.

With IDW you can control the significance of known points upon the interpolated values, based upon their distance from the output point. By defining a high power, more emphasis is placed on the nearest points, and the resulting surface will have more detail (be less smooth). Specifying a lower power will give more influence to the points that are further away, resulting in a smoother surface. A power of 2 is most commonly used, and is the default.

The characteristics of the interpolated surface can also be controlled by applying a search radius (fixed or variable), which limits the number of input points that can be used for calculating each interpolated cell. }[20]

---

[19] **A Comparison of Census Geographic Areas of Canada and the United States** by Carolyn Weiss Geography Division Statistics Canada,
Michael Ratcliffe and Nancy Torrieri Geography Division U.S. Bureau of the Census

[20] **ArcGIS, Spatial Analyst Tutorial , ESRI** Pg 132-141

Literature Cited

1. **A tale of four cities: intrametropolitan employment distribution in Toronto, Montréal, Vancouver, and Ottawa hull, 1981 - 1996.**
Richard Shearmur and William J Coffey
Received 28 November 2000; in revised form 8 October 2001

2. **Measuring Sprawl, Paper 27, Paul M. Torrens, Marina Alberti,** November 2000
Center for Advance Spatial Analysis , University College London
http://www.casa.ucl.ac.uk/working_papers.htm

3. **Urban sprawl in Canada and America: just how dissimilar?**
John R. Miron (Professor of Geography and Planning. University of Toronto at Scarborough)
http://www.citieslab.utsc.utoronto.ca/Papers/UrbanSprawl.pdf

4. **The modifiable Area Unit Problem: Empirical analysis by statistical simulation**.
Harold David Reynolds. University of Toronto, 1998, PhD. Thesis.
The modifiable Area Unit Problem. Openshaw and Taylor 1981.

5. **SPRAWL AND URBAN GROWTH,** Edward L. Glaeser (Harvard University and NBER) and Matthew E. Kahn. (Tufts University)
http://post.economics.harvard.edu/hier/2003papers/2003list.html

6. **When Census Geography Doesn't Work:** Using Ancillary Information to Improve the Spatial Interpolation of Demographic Data.
Paul R. Voss, David D. Long, Roger B. Hammer.
Center for Demography and Ecology, University of Wisconsin-Madison

7. **The Visualization of Area-based Spatial Data,** Stephen Wise, Robert Haining & Paola Signoretta.
Sheffield Centre for Geographic Information and Spatial Analysis, University of Sheffield.

8. **Population Density Surface:** A New Approach to an Old Problem
ZOLA K. MOON, FRANK L FARMER, School of Human Environmental Sciences
University of Arkansas

9. **On the measurement and generalisation of urban form** Paul Longley* and Victor Mesev**, *School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK. Corresponding author ** School of Environmental Studies, University of Ulster.

10. **A balanced view of scale in spatial statistical analysis.**
J. L. Dungan, J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jakomulska, M. Miriti and M. S. Rosenberg.

11. **Generating Surface Models of Population Using Dasymetric Mapping.** Jeremy Menis , University of Colorado, The Professional Geographer 55(1) 2003 pg31-42, © 2003 by Association of American Geographers.

12. **Analyzing data across geographic scales in Honduras :detecting levels of organization within systems.** Andrew Nelson, School of Geography, University of Leeds, Leeds.

13. **Spatial Data Analysis, Theory and Practice**, Robert Haining, University of Cambridge , 2003.

14. **ArcGIS, Spatial Analyst Tutorial, ESRI** Pg 132-141