EXPLORATORY SPATIAL DATA ANALYSIS OF MISSING
& FOUND PERSONS IN VANCOUVER

by

Nikki Thompson
B.A., Mount Royal College, 2001

THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

In the School
of
Criminology

© Nikki Thompson 2003

SIMON FRASER UNIVERSITY

July 2003

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author.
APPROVAL

Name: Nikki Thompson

Degree: M.A.

Title of Thesis: Exploratory Spatial Data Analysis of Missing & Found Persons in Vancouver.

Examining Committee:

Chair: Brian Buritch, Ph.D.

Patricia D. Brantingham, Ph.D.
Senior Supervisor

Paul J. Brantingham, J.D.
Member

Neil Boyd, LLM.
Member

Darryl Pisa, Ed.D.
External Examiner
Professor
University College of the Fraser Valley (UCFV)

Date Approved: July 18, 2003
PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Exploratory Spatial Data Analysis of Missing & Found Persons in Vancouver.

Author

(Signature)

Nikki Thompson

(Name)

August 14 2003

(Date)
ABSTRACT

This thesis explored methods that can be used to improve the spatial analysis of crime. The emphasis was on exploratory spatial data analysis (ESDA) in the study of missing and found persons calls for service in Vancouver in 1996. Traditionally, spatial crime analysis has been limited to the cartographic display of clustering or hot spots of criminal events. Research in spatial clustering typically uses limited statistical analysis. This thesis addressed the value of spatial statistics in understanding crime patterns by using global and local methods of analysis on non-standardized and standardized missing and found persons data. Local analysis indicated that clustering exists within both the missing and found persons calls for service. The found persons point data indicate stronger clustering and less dispersion; whereas the missing persons point data illustrate a more dispersed pattern highlighting multiple clusters existing within Vancouver. Findings on a global level indicate that the non-standardized data demonstrate spatial autocorrelation and association within both the missing and found persons datasets; whereas analysis conducted on the standardized data indicate no spatial autocorrelation within the missing persons data. Spatial autocorrelation was present when analysis was conducted on found persons data, and high concentrations (clusters) were present specifically in the downtown east side of Vancouver.
DEDICATION

I dedicate this work to my son, Damyn Thompson, who is the inspiration driving my success in life. Without his love and support I would not be where I am today. I also dedicate this work to the most encouraging and supportive family, Sandy, Terry and Ryan Thompson. With your wonderful guidance, love and support I am indebted to you. Thank you for always being there for me.
ACKNOWLEDGEMENTS

With my deepest gratitude I would like to thank both Dr. Patricia L. Brantingham and Dr. Paul J. Brantingham. Your overwhelming support and dedication to your students is greatly appreciated. You have inspired me more than words can say. I thank you for everything.

I would also like to thank Neil Boyd and Darryl Plecas. You were both very encouraging and supportive. Your knowledge and support with this work were a valuable contribution to my successful defence.

Finally, to Bryan Kinney, Aili Malm and my ICURS Lab Family. Your constant support and encouragement was very appreciated. Our lab stuck together through the good and the bad and I could not have done this without all of you. Our ICURS family has been the most wonderful support system and I am blessed to have been part of it. Thank you.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPROVAL</td>
<td>II</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>III</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>IV</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>V</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>VI</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>IX</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>XI</td>
</tr>
<tr>
<td>LIST OF STATISTICAL EQUATIONS</td>
<td>XII</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS AND ACRONYMS</td>
<td>XIII</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER ONE: SPATIAL ANALYSIS IN ENVIRONMENTAL CRIMINOLOGY</td>
<td>5</td>
</tr>
<tr>
<td>Hot Spot Analysis</td>
<td>10</td>
</tr>
<tr>
<td>Crime Generators and Attractors</td>
<td>10</td>
</tr>
<tr>
<td>CHAPTER TWO: MISSING PERSONS RESEARCH</td>
<td>12</td>
</tr>
<tr>
<td>Involuntary</td>
<td>13</td>
</tr>
<tr>
<td>Voluntary</td>
<td>13</td>
</tr>
<tr>
<td>INVOLUNTARY MISSING PERSONS</td>
<td>14</td>
</tr>
<tr>
<td>Parental Abduction</td>
<td>14</td>
</tr>
<tr>
<td>Profile of Victims</td>
<td>15</td>
</tr>
<tr>
<td>Profile of Abductors</td>
<td>16</td>
</tr>
<tr>
<td>Non-Parental Abduction</td>
<td>16</td>
</tr>
<tr>
<td>Profile of Victims</td>
<td>18</td>
</tr>
<tr>
<td>Profile of Abductors</td>
<td>19</td>
</tr>
<tr>
<td>Circumstantial and Situational Factors</td>
<td>21</td>
</tr>
<tr>
<td>Altered State of Consciousness / Dementia</td>
<td>22</td>
</tr>
<tr>
<td>Profile of Wandering Individuals</td>
<td>23</td>
</tr>
<tr>
<td>VOLUNTARY MISSING PERSONS</td>
<td>24</td>
</tr>
<tr>
<td>Runaways</td>
<td>24</td>
</tr>
<tr>
<td>Profile of Runaways</td>
<td>26</td>
</tr>
<tr>
<td>Throwaways</td>
<td>27</td>
</tr>
<tr>
<td>FUTURE CONSIDERATIONS FOR RESEARCH</td>
<td>28</td>
</tr>
<tr>
<td>CHAPTER THREE: SPATIAL ANALYSIS</td>
<td>32</td>
</tr>
<tr>
<td>LOCAL/POINT DATA ANALYSIS</td>
<td>35</td>
</tr>
<tr>
<td>Complete Spatial Randomness (CSR)</td>
<td>35</td>
</tr>
<tr>
<td>Spatial Intensity</td>
<td>37</td>
</tr>
<tr>
<td>Nearest Neighbor Analysis Index (NNA)</td>
<td>38</td>
</tr>
<tr>
<td>Standard Deviation Ellipses (SDE)</td>
<td>39</td>
</tr>
<tr>
<td>GLOBAL/POLYGON DATA ANALYSIS</td>
<td>41</td>
</tr>
<tr>
<td>Spatial Clustering</td>
<td>41</td>
</tr>
<tr>
<td>Defining Nearest Neighbors</td>
<td>45</td>
</tr>
</tbody>
</table>
Dissolving Census Tracts .............................................................. 130
Spatial Join of CT or EA and Addresses ...................................... 130
Summarizing Data by CT or EA .................................................. 130
Summarizing Data by Address .................................................. 131
Joining the CT Tables .............................................................. 131
APPENDIX F: INTENSITY SURFACE PLOTS ......................... 133
Missing Persons ....................................................................... 133
Found Persons ......................................................................... 134
LIST OF FIGURES

Figure 1 Classification of Measures for Spatial Analysis ................................................................. 34
Figure 2 Point Data Observations. Source: (Adapted from Griffith, 1987) .......................... 43
Figure 3 Connections of Point Data Observations. Source: (Adapted from Griffith, 1987) .... 43
Figure 4 Adding Perpendicular Bisectors. Source: (Adapted from Griffith, 1987) ............ 43
Figure 5 Creation of Thiessen Polygon. Source: (Adapted from Griffith, 1987) .................. 44
Figure 6 Observation Plot .................................................................................................................. 50
Figure 7 Map of the City of Vancouver 1996 .............................................................................. 66
Figure 8 Missing Persons Point Data Vancouver 1996 ............................................................. 68
Figure 9 Missing Persons Polygon Data Vancouver 1996 ....................................................... 69
Figure 10 Found Persons Point Data Vancouver 1996 ............................................................. 70
Figure 11 Found Persons Polygon Data Vancouver 1996 ....................................................... 71
Figure 12 Missing Persons Point Data by Average Income in Vancouver 1996 ................. 72
Figure 13 Found Persons Point Data by Average Income in Vancouver 1996 ................. 73
Figure 14 Missing Persons Point Data by Rate of Females Ages 10-19 Years ..................... 74
Figure 15 Missing & Found Persons Bivariate Scatterplot by X, Y Locations ..................... 76
Figure 16 GHAT for Missing Persons Point Data ..................................................................... 78
Figure 17 GHAT for Found Persons Point Data ....................................................................... 79
Figure 18 FHAT for Missing Persons Point Data ..................................................................... 80
Figure 19 FHAT for Found Persons Point Data ....................................................................... 81
Figure 20 Spatial Intensity Plot of Missing Persons by X,Y Locations Using Binning ........ 82
Figure 21 Spatial Intensity Plot of Found Persons by X,Y Locations Using Binning .......... 84
Figure 22 Nearest Neighbor Analysis Index for Missing Persons ........................................... 85
Figure 23 Nearest Neighbor Analysis Index for Found Persons ................................................. 86
Figure 24 Standard Deviational Ellipses for Missing and Found Persons ................................. 88
Figure 25 Standard Deviational Ellipse for Found Persons Point Data .................................... 89
Figure 26 Standard Deviational Ellipse for Missing Persons Point Data .................................... 90
Figure 27 Local Indicators of Spatial Association for Missing Persons by ZI on Non-Standardized
Data Using Geary's C ............................................................................................................. 96
Figure 28 Local Indicators of Spatial Association for Found Persons by ZI on Non-Standardized
Data Using Geary's C ............................................................................................................. 96
Figure 29 Local Indicators of Spatial Association for Missing Persons by ZI on Standardized
Data Using Moran's I ............................................................................................................. 100
Figure 30 Point Observations .................................................................................................... 111
Figure 31 Surface Plot of Missing Persons .................................................................................. 133
Figure 32 Surface Plot of Found Persons ...................................................................................... 134
LIST OF TABLES

Table 1 Contiguity Matrix. Source: (Adapted from Griffith, 1987) ........................................... 44
Table 2 Geary's C Interpretation Table. Source (Griffith & Amrhein, 1991, 137) .................. 47
Table 3 Moran's I Interpretation Table. Source (Griffith & Amrhein, 1991, 137-138) .......... 48
Table 4 Missing Persons Variable Set A .................................................................................... 57
Table 5 Found Persons Variable Set B ....................................................................................... 57
Table 6 Canadian Census Data by Census Tract ....................................................................... 58
Table 7 Frequencies of Calls For Service By Type .................................................................... 65
Table 8 Spatial Autocorrelation: Non-Standardized Moran's I Results ................................. 94
Table 9 Spatial Autocorrelation: Non-Standardized Geary's C Results ................................. 95
Table 10 Spatial Autocorrelation: Standardized Moran's I Results ........................................ 98
Table 11 Spatial Autocorrelation: Standardized Geary's C Results ......................................... 99
Table 12 Spatial Autocorrelation Comparison of Standardized and Non-Standardized Data ... 100
LIST OF STATISTICAL EQUATIONS

Equation 1 GHAT Test for Spatial Randomness. Source: (Kaluzny et al., 1998) ......................... 36

Equation 2 FHAT Test for Spatial Randomness. Source: (Kaluzny et al., 1998) ......................... 36

Equation 3 Nearest Neighbor Index. Source: (Levine, 2002) ....................................................... 38

Equation 4 Geary’s C Statistic. Source: (Fortin, Dale & Hoef, 2002, 5) ................................. 46

Equation 5 Moran’s I Statistic. Source: (Fortin, Dale & Hoef, 2002, 5) ................................. 48

Equation 6 Moran’s I Standardized. Source: (Rogerson, 2001, 167) ............................................ 49

Equation 7 Getis G Statistic. Source: (Chakrovorty & Pelfrey Jr., 2000, 73) ............................... 51

Equation 8 Local Gi Statistic. Source: (Chakrovorty & Pelfrey Jr., 2000, 73) ............................... 51

Equation 9 Local Moran’s I Statistic. Source: (Rogerson, 2001) .................................................. 52

Equation 10 Regression Equation .................................................................................................. 53

Equation 11 Spatial Autoregression. Source: (Fotheringham, Brunsdon & Charlton, 2000, 168) ................................................................................................................................. 54
# LIST OF ABBREVIATIONS AND ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD</td>
<td>Computer Assisted Dispatch</td>
</tr>
<tr>
<td>CSR</td>
<td>Complete Spatial Randomness</td>
</tr>
<tr>
<td>DBF</td>
<td>Digital Boundary File</td>
</tr>
<tr>
<td>ESDA</td>
<td>Exploratory Spatial Data Analysis</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information Systems</td>
</tr>
<tr>
<td>LISA</td>
<td>Local Indicators of Spatial Association</td>
</tr>
<tr>
<td>MAUB</td>
<td>Modifiable Areal Unit Boundary</td>
</tr>
<tr>
<td>NNA</td>
<td>Nearest Neighbor Analysis</td>
</tr>
<tr>
<td>SAC</td>
<td>Spatial Autocorrelation</td>
</tr>
<tr>
<td>SDE</td>
<td>Standard Deviational Ellipses</td>
</tr>
</tbody>
</table>
INTRODUCTION

As crime mapping is becoming a more accepted practice within policing, it is also being criticized for its lack of analytic use within police agencies. It has been found that police agencies currently use crime mapping techniques; however, they typically use only visual descriptive cartographical analysis (Manning, 2001). While crime mapping is an excellent tool, it is necessary to use the tool correctly to properly explain a phenomenon.

Crime mapping is a process of extracting location information (i.e., X and Y address locations of a crime or criminal event) from a database such as CAD (computer assisted dispatch). After standardizing the addresses, the data is then geocoded and imported into a geographical information system (GIS) tool such as ArcView for mapping. The geocoded data contain attributes (i.e., time, date), which are information variables associated with the location information. Once geocoded, the data are then illustrated in the form of a shape file. A shape file can be represented by points (e.g., X and Y locations) or aggregated and represented as a polygon (e.g., aggregate counts in a census tract). These shape files are then overlaid on a basemap (e.g., the city of Vancouver). Once the data are in the shape file form they are ready for descriptive and statistical analysis. Crime mapping is a process of interpreting the spatial and/or temporal dynamics of crime.

The argument that police agencies lack analytic power is a concern within spatial analysis. There are various levels of analysis that can be conducted once the data are standardized and imported into a mapping program. The first level of analysis is purely a visual description; for example, mapping the events of robberies within a city. This
allows police officers to see where clustering of the events occur. It does not allow for understanding of the clusters, patterns, dispersions or neighbor associations. Most importantly, simple mapping fails to explain why the crimes occur where they do. Highlighting that visualization is effective only for the first level of analysis, Manning (2001) discusses a concern that maps can be easily distorted to create an inaccurate picture of the events. Depending on the projection of the map and the size that you choose as a study area, events can be perceptually distorted. Manning’s argument that crime mapping is not being applied properly in police agencies may not be correct when the department implements it only at a visual level. Manning argues that additional statistical analysis of patterns, dispersion and neighbor associations creates a better representation for understanding crime.

The most important aspects of crime mapping are theory and associated research of the event being studied. Theory should be the foundation of the explanation for crime concentrations in certain areas before mapping is conducted. With theory and research, crime analysis of patterns, clusters and dispersion can not only be visualized, but also explained. Models in spatial analysis of autoregression draw on variables (i.e., age, income, shopping malls, etc.) to explore the explanations as to why crime occurs in certain areas. This analysis is vital to crime mapping for police agencies in that the patterns, dispersions and explanatory factors enhance our knowledge of crime.

Based on Manning’s criticism of crime mapping, this thesis was conducted detailing the exploratory stages of analysis in crime mapping and outlining essential steps in the future of crime analysis. To demonstrate the value of mapping in crime analysis, a phenomenon was chosen that had never been spatially analyzed. Thus, missing and found persons calls for service data in Vancouver 1996 were selected not
only because of deficiencies in research regarding typologies, but specifically, for the lack of prior spatial analysis.

The objective of this study was to demonstrate how an increased reliance on spatial analysis is necessary as an effective tool for understanding crime. Specifically, this study applied exploratory spatial data analysis (ESDA) methods of patterning missing (N=2627) and found persons (N=625) calls for service data in Vancouver, 1996. This study employed two methods of exploratory spatial data analysis, local and global, to demonstrate the value of spatial data analysis.

Local methods of analysis include tests for complete spatial randomness (CSR), spatial intensity, nearest neighbor index analysis, and standard deviational ellipses. Local analysis was conducted on point (X, Y) missing and found persons location data, whereas global analysis was conducted with spatial neighbors, spatial autocorrelation (Moran’s I and Geary’s C), as well as local indicators of spatial association on polygon data. Polygon data is an aggregate count of missing and found persons locations over a boundary area (i.e., census tract).

The initial stage of this study contained a theoretical trace of the foundations of environmental criminology. Theories such as Routine Activities, Rational Choice and Pattern Theory are discussed to provide a foundation for conducting research on the spatial factors of missing and found persons. The second chapter contains a review of the current literature on missing persons in Canada and the United States. Given the lack of research on spatial typologies of missing persons, a review of existing typologies provides a foundation of factors that influence the involuntary and voluntary aspects of why people go missing.
The third chapter is a comprehensive review of the global and local methods for conducting spatial analysis. These methods are examined to provide an overview of each method and its use, as well as methods of calculation for each test. The fourth chapter provides a detailed description of the methods and procedures conducted in this exploratory spatial data analysis, including steps conducted in the process of the mapping of data. The fifth chapter contains the results of the analysis including bivariate, cartographical, and statistical representations of local and global analysis. The sixth chapter is a discussion of the pertinent literature and theory in support of the results, as well as a discussion on future research considerations. Finally, issues pertaining to the implementation of spatial analysis in criminology are addressed to provide recommendations of where crime mapping should be directed in the future.
CHAPTER ONE: 
SPATIAL ANALYSIS IN ENVIRONMENTAL CRIMINOLOGY

The primary focus of environmental criminology is to better understand crimes by considering them as events that are the result of how individuals with varying levels of motivation interact with their surroundings (Brantingham & Brantingham, 1997). Environmental criminology researchers "actively explore crimes as events with broad ranging diversity, but which are understandable when explored by jointly considering potential offenders and their proximal and distal surroundings" (Brantingham & Brantingham, 1997, 31). The following section outlines the role environmental criminology plays when examining the phenomenon of crime, but more specifically its role in the analysis of missing persons.

The focus of environmental criminology is directed towards "understanding the criminal event and how it is related to the individual motivation, to victims and targets, and to the legal, social, psychological and physical backcloth against which crime occurs" (Brantingham & Brantingham, 1997, 31). Environmental criminology is an integrated approach to the study of crime and criminality. The orienting strategy in environmental criminology is not reliant on a singular theoretical foundation. Rather, it has a common theme within several theoretical approaches, all of which have an orientation towards understanding patterns of crimes within a complex environment (Brantingham & Brantingham, 1997, 31).

Environmental criminology research covers three areas: (1) understanding the "where" and "when" of different types of crime; (2) understanding how settings facilitate crime; and (3) exploring how activity patterns and individual decisions shape the location
and timing of crime (Brantingham & Brantingham, 1997). This thesis focuses on the first area of environmental criminology described above where patterns of missing and found persons concentrations were explored. The location of concentrations will be analyzed spatially and compared visually to social factors. A wide range of social factors must be considered when viewing patterns of missing persons. For example, spatial patterning of missing persons may be related to specific locations such as a hospital or nursing homes.

Three specific theories within environmental criminology will be used in this thesis: Routine Activities Theory, Rational Choice Theory, and Pattern Theory. These theories vary in scale. Routine Activities explores aggregate crime activities. Rational Choice Theory looks at individual decisions. Pattern theory moves between the aggregate or macro scale and the individual or micro scale.

Routine Activities Theory created by Felson and Cohen (1979) does not solely focus on the characteristics of the offender, but more on the characteristics of the event. Routine Activity Theory contains three propositions to understanding criminal events. For a crime to occur there must be a motivated offender, a suitable target, and an absence of capable guardians (Felson & Cohen, 1980; Felson, 1987). The first element dealing with the motivated offender entails provocation, boredom or temptation. It is the notion that crimes are more likely to occur when there is a confluence of these characteristics. The second element necessary for the offence is a suitable target (Felson & Cohen, 1980; Felson, 1987). Targets are created in many ways. The potential target must have an attractiveness or be ‘suitable’ to the offender in committing the crime. For example, theft of televisions increased as the size and weight of the television decreased, but lost attractiveness when televisions saturated the market (Felson & Cohen, 1980). Absence of guardians is the third element that facilitates the
commission of an offence (Felson & Cohen, 1980; Felson, 1987). A lack of parents, neighbors, friends or authorities is believed to promote the offence when the offender has a suitable target and motivation for committing the offence. If one of these three elements is missing, then the crime is less likely to occur.

As stated before, Felson and Cohen's perspective of Routine Activity Theory is macro level theory. Some theorists who write on environmental criminology take a micro level approach. In particular, Clarke and Cornish (1985) developed a micro level theory called Rational Choice Theory. They look at the decision-making process of the offender, rather than focusing on aggregate or summary characteristics of crime. Rational Choice Theory was developed in theoretical support of situational crime prevention measures (Cornish, 1993). The main component of this theory is the offender's decision-making process. Two assumptions are vital to understanding offender's choices. The first assumption is that crime is purposeful and is for the benefit of the offender (Felson & Clarke, 1998). The second assumption is that offenders are goal-oriented. To commit a crime, the offender believes that the benefit from committing the crime outweighs the consequences or risk of being caught. This should not be viewed as an abstract rationality. The decision-making process occurs without the full picture in mind (Felson & Clarke, 1998). This limitation is usually called bounded rationality.

This theory differs from Routine Activities in that it is a micro approach based on the individual decision-making process. Routine Activity Theory is from a macro perspective analyzing how broader patterns of individuals' daily routines and motivation, intended targets and with a lack of guardians encourage criminal events. Routine Activity Theory is more like demographics or human ecology. Routine Activities and Rational Choice theories both contribute to the understanding of crime from a macro and
micro perspective. Neither of these theories, however, addresses spatial distribution of events or activities. Pattern theory adds this component.

Pattern theory is seen as a multidisciplinary approach to understanding crime and criminality. It is built on the geometry of crime while incorporating parts of Rational Choice and Routine Activities (Brantingham & Brantingham, 1993). As the name implies, Pattern Theory explores both spatial and temporal patterns that exist for crime. It is assumed that crimes do not occur randomly or uniformly across space and time (Brantingham & Brantingham, 1993).

Pattern theory looks at crimes as dynamic events that frequently reoccur to create common patterns in space and time. This likelihood of a criminal event occurring is dependent on the "backcloth, the site, the situation, and individuals readiness, routine activity patterns, and the distribution of targets" (Brantingham & Brantingham, 1993, 266). Viewing criminal events and behaviour from these conditions allows the analysis of individual and aggregate patterns over diverse urban backcloths (Brantingham & Brantingham, 1993). Pattern Theorists propose that crime is an event process that must be examined by understanding the activity backcloth, readiness/willingness of individuals to commit an offence, the distribution of potential targets, and the structural backcloth of urban form (Brantingham & Brantingham, 1993).

The activity backcloth are patterns that exist based on repeated human actions and interactions. Everyday patterns exist whether you drive to work, go to school or walk to the mall. Typically, the path you take to get to these destinations is the same, rarely deviating from the familiar, thereby creating recurring routines in daily life. Through these patterns of regular and repeated behaviour, people tend to cluster towards predictable locations. These locations, typically called 'nodes', are activity
centres for which daily life routines lead people to cluster. An offender is more likely to commit a crime when the opportunity arises within a familiar routine (Brantingham & Brantingham, 1993). In the case of missing and found persons, the activity backcloth is the rhythm or pulse of a city that helps identify locations from where a person goes missing, or is found.

Once routines exist, when an individual may be triggered into committing the crime, the actual crime is more likely to originate from within that persons routine activity space (Brantingham & Brantingham, 1993). Unlike many theories in criminology, it is believed there is not one single goal behind all types of crime (Brantingham & Brantingham, 1993). Readiness to commit a crime will vary by crime type. For missing persons, an example of a trigger could be a fight with a parent or schoolteacher when a teenager is involved. A specific fight could trigger a person to run away from his or her routine. From a Pattern Theory perspective the teenager who runs away is likely to be attracted to other known areas.

The structural backcloth is important in understanding routine movement patterns. It identifies major travel routes and activity nodes. For example, when looking at the pattern of commercial burglary, crimes will be clustered in more industrial/commercial zones. Some commercial areas will be more attractive to burglars just as some commercial areas are more attractive to shoppers. It is important to look at the structural backcloth to see how it varies depending on the city structure and type of crime. For a teenager who is triggered into running away, the structural backcloth may help identify places that attract the runaway.

In the case of Vancouver, missing persons are likely to be attracted to high activity areas like downtown Vancouver, but not far from the bus route used in the past
by the individual. Using pattern theory in analyzing missing and found persons, patterns may be uncovered relating to the homes, school, work and normal daily routines.

**Hot Spot Analysis**

The routine behaviours of people in daily life create concentrations of activity in the spatio-temporal environment. These concentrations made by daily routines create possible hot spot formations of criminal activities (Brantingham & Brantingham, 1997). Hot spots can become recognizable to an observer through visual representation on a map. A theoretical approach based on prior knowledge of routine activities helps in the search for the explanation of the clustering (Brantingham & Brantingham, 1997).

Hot spots are found for all types of activities. The objective of this study is to explore whether hot spots exist for missing and found persons. It is noted by Brantingham & Brantingham (1997) "hot spots are not accidental and are linked to the backcloth, movement and mobility, crime generators and crime attractors" (13). It is important to study the environment surrounding the clusters of missing persons to look for location attractors and generators.

**Crime Generators and Attractors**

Crime generators are considered “particular areas to which large numbers of people are attracted for reasons unrelated to any particular level of criminal motivation they might have or to any particular crime they might end up committing” (Brantingham & Brantingham, 1995, 7). This idea of crime generators applies to the analysis of missing persons in that generators may include places such as malls, hospitals, schools, halfway houses, and group homes. These areas are not necessarily related to missing persons. They are however environments that attract large numbers of people, some of whom
may have the potential to go missing. Crime attractors would be more pertinent to involuntary missing persons, which includes kidnapping, abduction, altered state of consciousness or to locations that attract youth who are, in turn, likely to attract people who will lead them to leaving home and entering a lifestyle involving the sex trade and/or drugs.

Crime attractors are considered "places, areas, neighborhoods, districts which create well-known criminal opportunities to which strongly motivated, intending criminal offenders are attracted because of the known opportunities for particular types of crime" (Brantingham & Brantingham, 1995, 8). In the context of this study, crime attractors include kidnappings concentrated around schools, shopping malls or areas that encourage youth to leave home. Crime generators and attractors are extremely important in exploring hot spot formation of missing persons and locations where missing persons are found. To fully examine the environmental or spatial characteristics of missing and found persons, it is important to examine past research in this specific area.
CHAPTER TWO:
MISSING PERSONS RESEARCH

The reason for choosing missing persons data as the example for spatial analysis is lack of research in this area. For example, missing persons have never been spatially analyzed and specific categories of missing persons have never been fully developed in Canada. Current American researchers have informed us of the need to examine spatial factors pertaining to missing persons and their environment.

Generalized missing persons data will be used to conduct the analysis, so it is important to understand the definition of missing persons pertaining to the data collected. An extensive literature review uncovered many definitions of 'missing'; however, there is no universal agreement among scholars or police. The Police Reference Manual for Cases of Child Abduction and Runaway Youth (1993), discusses how there is only one aspect of current missing persons definitions that has common agreement, that the person is not present in their usual residence (Weiler et al).

Further definitions range in perspective: the psychology of individual behaviour, the philosophy of 'missing', and the sociology of individual routines. After examination of the literature, the definition of missing persons used in this study is based on research conducted by Swanton and Wilson (1989). Their definition of missing persons, which is applicable to both children and adults, states that a missing person is:

[O]ne who, not being the subject of lawful commitment order, is absent from his/her normal haunts [social centre] in breach of the reasonable expectations and/or responsibilities of another by reason of abduction, altered state of consciousness or voluntary decision, and whose location is either not known or, if known, who is illegally detained (Swanton & Wilson, 1989, 2)
This definition is appropriate because it encapsulates aspects of ‘missing’ in a range that can apply to a wide variety of cases. This thesis analyzes data from all missing persons calls for service in Vancouver in 1996. It is important that the definition is not limiting, but rather a generalized definition of missing persons. With this definition, it is important to understand theoretical components from the field of environmental criminology, in order to explore the spatial relations of missing persons.

The review of the literature pertaining to the types and characteristics of missing persons will follow the format below dealing with involuntary and voluntary forms of missing persons.

**Involuntary**

1. Forcible abduction:
   - Parental, e.g. Non-custodial abduction;
   - Non-Parental, e.g. Kidnapping;
2. Non-Forcible abduction, e.g. Inducement, persuasion, enticement.
3. Altered state of consciousness:
   - Amnesia;
   - Confusion, drug induced or otherwise;
   - Senility, Alzheimer’s disease;

**Voluntary**

1. Removal of self from unsatisfactory physical/social/economic circumstances:
   - Running away from home by juveniles; (Derived from Swanton et al, 1988, 46)
   - [Throwaways – Parental abandonment]
Involuntary Missing Persons

Parental Abduction

A wide range of terminology is used to explain parental abductions. The most common definition in the Canadian literature is from the Solicitor General of Canada’s Police Reference Manual for Cases of Child Abduction and Runaway Youth:

A parental abduction occurs when a child under fourteen years old is taken by a parent or guardian without permission or legal authority from the parent or guardian with lawful custody of the child. The sections of the Criminal Code that may apply to parental abductions include: Abduction in Contravention of a Custody Order (section 282); and Abduction Where No Custody Order (section 283) (Weiler et al, 1993, 5).

This definition accurately depicts parental abduction based on analysis from Canadian law, and will be used in this section for explanatory purposes.

Parental abduction is typically the most researched aspect of missing persons. The Victims of Violence group found that in Canada children are in far more danger of being abducted by someone they know as opposed to a stranger (2001). Furthermore, the Missing Children Society of Canada estimates that the majority of abductions transcend national boundaries, and parental abductions account for the highest amount of total abductions (1999). International abductions are the most difficult to trace because, when a parent abducts a child and crosses national boundaries, it creates the need for jurisdictional cooperation between countries to retrieve the child.

Parental abduction most commonly occurs "after divorce or separation of the parents when the break-up had not been amicable and there has been disagreement
about custody and access rights" (Weiler et al, 1993, 49). Whether custodial or non-custodial, the parent abducts the child in order to deprive the other parent of any rights to custody. Although in Canada rates of parental abduction are not considered significant when compared to rates in the United States, the statistics are still very concerning.

In 1991, 346 cases of parental abductions in Canada involved a custody order (section 282), where the child was abducted in a situation where a custody order for one parent was assigned in court (Weiler et al, 1993). There were 252 cases of abductions not involving a custody order, which means no order has previously been assigned to either parent (Weiler et al, 1993). In Canada, parental abductions occur at a rate of 2 abductions per 100,000 children per year (Weiler et al, 1993). This is double the rate of non-parental abductions in Canada according to the RCMP Missing Children's Registry 1991 Annual Report (Weiler et al, 1993). These are the only statistics provided regarding Canada's current parental abduction rates, and only limited information was found on victims of parental abduction.

Profile of Victims

According to the Solicitor General of Canada's Report, there is very limited profile information with respect to the victims of parental abductions (Weiler et al, 1993). This report found gender differences among victims of parental abductions were not statistically significant (Weiler et al, 1993). Girls and boys were abducted at approximately an equal rate, which is consistent with American Statistics.

In the United States, the majority of parental abductions were children ranging from two through eleven years (Weiler et al, 1993), but there is no information in Canada
regarding the age range of abducted children. Currently, this is the only statistic provided on victims’ profiles in Canada. Research is similarly deficient in Canada in terms of abductors’ profiles.

Profile of Abductors

There have been no Canadian studies conducted dealing with the characteristics of parental abductors, but American research has found that 75% of the abductors were male (Weiler et al, 1993). After reviewing the aforementioned profile of parental abductions, the research is deficient in the areas of statistical rates, victims, abductors, circumstances surrounding abduction and environmental factors. Further critique will be discussed in the recommendations section.

Non-Parental Abduction

Weiler et al supply the most appropriate definition of non-parental abduction:

A non-parental abduction occurs when a child under sixteen years old is taken by a non-parental abductor. A non-parental abductor is legally defined as someone other than the parent, guardian, or other person having, or entitled to, lawful care or charge of the child. The sections of the Criminal Code that may apply to non-parental abduction include: Kidnapping (section 279(1)); Abduction of Victim under Sixteen (section 280); and abduction of Victim under Fourteen (281) (1993, 5).

This definition is considered to be most accurate description of non-parental abduction based on Canadian law and will be relied on for the exploratory nature of this study.

The Adam Walsh case in Florida, involving the abduction and murder of John Walsh’s son in 1981, falls within the category of non-parental abduction also denoted as ‘stranger abduction’ or ‘stranger danger’ (Johnson, 1988). Non-parental abduction is
very uncommon in Canada (Weiler et al, 1993), yet this is not a reason to ignore it; it is more important to learn how to deal with stranger abduction, for when it happens, knowledge is required by police agencies for a rapid response to find the child.

Research in the United States illustrates two common types of non-parental abduction. The first type "involve[s] the abduction of females in their pre- to mid-teens, usually for sexual purposes" (Weiler et al, 1993, 23). The Second type is "baby snatching," in which infants are abducted either for illegal adoptions or by childless women or couples (Weiler et al, 1993). In the United States, a common legal division differentiates between 'baby snatching' and 'teenage abduction'; however, no research has been conducted in Canada to determine if these differentiations are consistent with the United States. The infrequency of this type of abduction in Canada is reflected in the lack of statistics on the subject.

In 1991, there were 77 non-parental abductions in Canada involving the abduction of persons from the age of fourteen to sixteen (section 280) (Weiler et al, 1993). Furthermore, an additional 418 cases involved the abduction of persons under fourteen (section 283) (Weiler et al, 1993). In Canada, non-parental abductions occur at a rate of 1 victim per 100,000 children each year. This is approximately half of the parental abduction rate in Canada (Weiler et al, 1993).

The only other information available in Canada is from the RCMP Missing Children's Registry, which states, "that relatively few of the non-parental abductions are true 'stranger' abductions (where the abductor is unknown by the victim and parent)" (Weiler et al, 1993, 24). In Canada, there is no research analysing 'true stranger' and 'friend stranger' abductions. Deficiencies in research are also found in the profiling of victims.
Profile of Victims

Given the deficiencies in research within Canada, this section addresses research conducted in the United States to emphasize the importance of future research in Canada. Weiler et al suggest, "In the United States, half of the victims abducted by a non-family member are twelve years old or older" (1993, 24); furthermore, the victims of non-parental abductions are typically female, occurring at a ratio of 3:1 (1993).

Analysis conducted by James Tedisco and Michele Paludi (1996) of profiled victims of non-parental abductions, found that there were several psychological and socio-economic predictors that potentially make children more vulnerable to being abducted by a stranger. Their predicting factors are as follows:

Those children that are more vulnerable to stranger abductions are the quiet, thoughtful ones, children who appear to have special and intense needs for adult affection and approval. Other vulnerable children include those who are loners—withdrawn, with poor social skills with children their own age. Children who look unclean or unkempt, thus in need of attention, are also vulnerable to abductors. However, we should point out, that any child who has discipline problems at school or at home may be vulnerable. Also vulnerable are those involved in situations such as separation, or divorce, or illness (1996, 49).

Despite the importance of identifying these factors, theoretically, this personality profile could be applied to all children in society. Notwithstanding the advances purported by this personality profile, the overall profile provided here leads us no closer to prediction and understanding. Although the victim component of this model is criticized, Tedisco and Paludi's explanation of the abductor strengthens the typology of stranger abductions.
Profile of Abductors

Supporting research in Canada in the area of the non-parental abductor profiles is extremely deficient. Weiler et al discuss information regarding abductors in Canada demonstrating that "most of the abductors are relatives, friends, or acquaintances" (1993). In contrast, the majority of abductors in the United States are complete strangers (Weiler et al, 1993, 25); therefore, for the purposes of exploratory research, it is important to analyze American literature from Tedisco and Paludi (1996) to illuminate important areas for future research with respect to profiles of abductors in Canada.

In a study of missing children, Tedisco and Paludi (1996) found several areas to focus in regard to profiling the abductor. First, they found that there were four categories of non-parental abductors: pedophiles, profiteers, serial killers and childless psychotics. In regard to the first category:

Pedophiles constitute the single largest number of child abductors. According to the Diagnostic and statistical Manual of Mental Disorders of the American Psychiatric Association, paedophilia is 'recurrent, intense, sexual urges and sexually arousing fantasies, of at least six month's duration, involving sexual activity with a prepubescent child'. The age of the child is generally 13 or younger (Tedisco & Paludi, 1996, 45).

The second category of profiteers includes criminal exploiters. Exploiters are defined as abductors that sell children to pornographers or adoption rings (Tedisco & Paludi, 1993). Once sold to pornographers or adoption rings, the child is exploited in business profiteering.

The third category of abductors includes serial killers. This group includes "[c]hild abductors [that] may move from state to state kidnapping and murdering their victims" (Tedisco & Paludi, 1996, 47). The researchers conclude that the actions of
serial killers in the abduction of children are typically methodical and ritualised. The most common themes surrounding the abduction are power, control, and dominance (1996).

The fourth category of abductors is *childless psychotics*. This category includes abductors such as 'baby snatchers', where the abductor was not physically able to have children or has lost a child. If the individual cannot fill the void of losing a child, they may seek to abduct another individual's child in hopes of compensating for their perceived personal loss (Tedisco & Paludi, 1996, 47).

The categories provided by Tedisco and Paludi reflect abduction trends in the United States. The only concerns with these categories are that they are not mutually exclusive and mutually exhaustive. Abductors may overlap categories or not fall into one category at all; for example, a neighbour, realizing that a child is being beaten constantly next door, and that social services will not take the child away, decides to abduct the child in hopes of providing a better future for the child is considered to have committed non-parental abduction. This scenario is not reflective of the aforementioned categories.

To enhance the explanatory power of the non-parental abduction model, Tedisco and Paludi (1996) list the most frequent lures in the abductor profile. The list of common lures includes:

**Asking for directions:** This is a frequently used lure for stranger abductors, who trick children into believing they are friends and need their help.

**Asking for help to locate a pet:** This lure works successfully most times because it is a request that children find difficult to refuse. This request (often the abductor shows the child a leash) is frequently
accompanied by a detour in to a wooded or other secluded area.

**Telling the child that a parent has been in an accident and is hurt**

**Knocking at a house door/ringing the doorbell to gain entry into the house.**

**Offering to give children a ride home to their parents**  
(Tedisco and Paludi, 1996)

Researchers (Greeman-Longo and Wall, 1986; Malinosky-Rummell and Hansen, 1993; Quina and Carlson, 1989; Russell, 1975 as quoted by Tedisco and Paludi, 1996) have also discovered the following regarding the psychological profile of stranger abductors:

[Child Abductors:]

- Express little or no concern, trust, or empathy for others, especially their victims;
- Cannot express anger in ways that are not violent, and express their anger toward children because the latter are less likely to confront their powers;
- Typically view sexual victimization as an element of the masculine role in society;
- May have been sexually abused themselves;
- Believe that the abduction does not have serious consequences for their victims;
- Are likely to be repeat offenders since the underlying psychological problems are not resolved by the abduction (Tedisco and Paludi, 1996, 53).

A great deal of research examines profiles of abductors in the United States. The information above indicates areas in need of elaboration in Canadian research. A final aspect of non-parental abduction is the circumstances surrounding the abduction.

**Circumstantial and Situational Factors**

Understanding situational and circumstantial factors is important in understanding non-parental abduction. If specific locations or settings are found to be
conducive to abduction, understanding these patterns will be critical in contributing to the
creation of a typology. Statistics found by Weiler et al show:

- Over two thirds of the cases involved the sexual assault of girls, who were mostly taken from the street.
- Force was involved in 85% of the cases.
- Most non-family abductions last less than one day, with approximately 20% lasting less than one hour.
- In up to 21% of the cases the children were physically injured.

These statistics are crucial in demonstrating that knowledge of situational circumstances may eventually uncover patterns by which to predict stranger abductions; however, further research is necessary.

**Altered State of Consciousness / Dementia**

One of the forms of involuntary missing persons occurring more often is exit-seeking behaviour caused by dementia or altered states of consciousness. It is commonly found that “every year, in long-term care facilities across the nation, dementia residents find their way out of the facility, get lost, and later are found injured or, worse, found dead” (Lucero, 2002, 277). Wandering is a term that defines a wide range of behaviours that relate to dementia-diagnosed elderly persons. For the purpose of this thesis, wandering is considered dementia-related behaviour used to seek an exit from a facility because of an altered state of consciousness or confusion (exit-seeking). Wandering has been found to occur in 17.4% of new patient referrals, and approximately 50% in cases of severe dementia (Hope et al., 2001).
Profile of Wandering Individuals

It has been found that "exit-seeking behaviour is characteristic of middle-stage dementia residents" (Lucero, 2002, 277). The common characteristics of an exit-seeking wanderer include:

Higher functioning abilities, have severe short-term memory loss, poor reasoning and judgement, severe spatial disorientation...and lack even the most basic safety awareness....they retain good social skills and relatively good communication ability, leading people who are unaware of their dementia to perceive them as 'normal' when first meeting them or interacting with them for a short time (Lucero, 2002, 277).

Two classifications of wanderers (elopers and runaways) have been developed to differentiate between types of wandering behaviour. Elopers "are seemingly unconcerned about the fact that they are in a long-term care facility and, as such, have easy-going, calm demeanors" (Lucero, 2002, 278). They typically consider themselves visitors to the facility, rather than actual residents (Lucero, 2002). In contrast, runaways have more insight into their current situation and have anger, anxiety or confusion about being in the facility (Lucero, 2002); they typically feel they are being held against their will, and their desire to leave is fuelled by their concern to return to their family members (Lucero, 2002).

In a study conducted by Koester and Stookbury (1995), it was found that dementia, particularly Alzheimer's and wandering behaviour was severely under-researched. Their study compared the dementia patients with non-dementia patients possessing normal cognitive abilities (Koester & Stookbury, 1995). Wanderers were distinctly different from a lost elder, displaying constant disorientation, better social skills, inability to know they were lost and were typically more active (Koester & Stookbury,
Wanderers were typically found within one mile of the point of last seen, and that among those patients found within the first 24 hours, no deaths occurred (Koester & Stookbury, 1995). The mortality rate of wanderers was 19%, and death occurred most frequently from hypothermia, dehydration or drowning (Koester & Stookbury, 1995). The wanderer’s probability of mortality increased rapidly if the missing person was not found within the first day.

The majority of patients were last seen at their nursing home or residence, and typically travelled along roads (Koester & Stookbury, 1995). It is interesting to note that this study was conducted in Virginia which consists of a swampy region with rolling hills and heavily forested mountains (Koester & Stookbury, 1995). It is important to mention that even though the residents were found one mile from the point of last seen, this distance may increase in a less densely vegetative environment. Wanderers differ from runaways based on the involuntary component of their decision. Wanderers are typically in an altered state of consciousness, whereas a typical runaway makes a cognitive choice to leave the point of last seen.

**Voluntary Missing Persons**

**Runaways**

Runaway children account for the most common statistic with respect to missing persons, and these cases are typically resolved very quickly (Weiler et al, 1996). According to Weiler et al, the definition of a runaway child is:

Runaways are [considered] children who voluntarily flee from their home. Children who have run from problems associated with their home, school, or personal life; Children who have failed to return to their place of residence, in violation of specific rules; or Children who have left their place of residence in an attempt to establish independence (1993, 79).
While there are multiple definitions explaining a variety of runaways' behaviours, the above definition is based on the Solicitor General of Canada’s Report (1993), and will be used in the exploration of this study.

The majority of research shows that running away is not a one-time event. Youths begin with an infrequent running away period, demonstrating typical patterns that evolve into chronic or frequent running away (Weiler et al, 1993). Jones (1988) created a typology of youth runaways demonstrating that explanations of runaways were typically problems relating to ‘family dynamics’ (Jones, as quoted by Payne, 1995).

Jones also argues there is reason to distinguish between youths who leave on their own volition, and youths who are “pushed out of their home by parental attitudes, either of indifference or active hostility” (Jones, as quoted by Payne, 1995, 339). A distinction is clear between a person leaving on their own (runaway) and a person leaving without a choice (throwaway).

Finally, an important aspect to consider is the theory surrounding why youths run away. Naomi Golan (1978) developed Crisis Theory, which was intended for treatment use in crisis situations. A summary of the theory states:

In Crisis Theory, ‘hazardous’ events cause the individual to be in a ‘vulnerable state’ in which their capacity to deal with problems is inhibited and they are worried about their life. This leads to additional stress and tension. A ‘final straw’, the precipitating factor which may in fact be a quite minor problem, sets off a ‘state of active crisis’ in which their typical ways of solving problems do not work and they have to cast around, ever more separately, for ways to resolve problems (Payne, 339).
Crisis Theory is considered a plausible explanation to date, explaining a significant portion of runaway youth behaviour. Researchers have not concluded what degree of behaviour this theory explains; however, it is widely used throughout the literature on runaways.

Research in Canada indicates that “the overall number of runaway cases reported to the RCMP Missing Children’s Registry in 1991 was 43,786, down slightly from 1990” (Weiler et al, 1993, 81). In the 1991 Annual Report, the rate in Canada was approximately 13 runaways per 100,000 children per year. This is drastically different from the United States, where it was reported that there were approximately 205 runaways per 100,000 children (Weiler et al, 1993).

A final statistical conclusion regarding the differences in family environments put forth by De Rocher (1987) studied the family environments of ‘normal’ youth and ‘runaway’ youth. De Rocher found that an increase in family cohesiveness coupled with an increase in positive interaction with children led to decreased provoking of running away (De Rocher as quoted by Payne, 1995).

**Profile of Runaways**

Canadian research shows that over 90% of the runaways studied were between ten and eighteen years old, with thirteen to fifteen year olds making up 60% of the statistic (Weiler et al, 1993). Research in Canada and the United States also describe the majority of runaways as females around the age of fifteen (Weiler et al, 1993).

Many factors are considered in the profile of a runaway; however, the most common incidents of runaways consistently demonstrate problems in school and educational history. Displays of ‘poor behaviour’ in social, and school settings are also a
common trend (Weiler, 1993). This is the only research in Canada depicting the profiled characteristics of a runaway.

**Throwaways**

Throwaways were added to the category of missing children because this category is currently becoming a widely accepted sub-group of runaways. Runaways commonly display a characteristic of fleeing a problem, situation, or family setting. This definition neglects youths who are *thrown away* from something without a choice.

The current literature is still in its infancy regarding throwaways. One definition suggests a partial explanation of this phenomenon. This definition is from the Police Reference Manual for Cases of Child Abduction and Runaway Youth. In this manual, Weiler et al (1993) define a throwaway as a child "forced to leave their home" including:

- Children who are told to leave their place of residence by a parent or another adult living there;
- Children who are abandoned by their parents;
- Children who have run but have not been missed by, searched for, or reported to the police by their legal caretakers; and
- Children who are denied re-entry to their place of residence after running away (79).

As previously mentioned, the term throwaway is relatively new. There are no studies conducted with respect to statistical information differentiating between runaways and throwaways in Canada; however, research in the United States demonstrates several important trends of throwaways. Weiler et al found that a majority of throwaways are between the ages of sixteen and seventeen years old (1993). They also concluded that the majority of throwaways stem from single parent living situations; furthermore, most throwaway youths do not venture farther than a ten mile radius from their home (1993).
Future Considerations for Research

A common criticism of the literature on missing children is the extreme lack of research in the areas of parental abduction, non-parental abduction, runaways, and throwaways. This section will highlight areas that are deficient, with special attention to the categories of parental abduction, non-parental abduction, wanderers, runaways and throwaways, in order to determine recommendations for future research.

In regard to parental abductions, statistical overviews of the rates of abductions are deficient. In order to fully understand this phenomenon, a general exploratory study needs to be conducted in Canada highlighting areas such as:

1. Characteristics of parental abductors including:
   - Personality characteristics
   - Environmental characteristics
   - Circumstantial patterns
   - Differences between male and female abductors
   - Characteristic differences between international abductors and Non-International abductors

2. Profile of the victims including:
   - Critical age ranges of victims
   - Gender differences

Research needs to be focused on these areas in order to explain and predict parental abduction. Canada's abduction rates in relation to the United States are low; however, this does not diminish the need for research on parental abductions.

As noted, there is a substantial amount of research available dealing with profiles of victims and profiles of abductors associated with non-parental abductions. Unfortunately, this research is specific to an American context. Canadian statistics depict
differences in non-parental abduction between Canada and the United States, but not within Canada alone. Areas in need of significant research in Canada include:

1. Percentage of non-parental abductions including:
   - Percent of true ‘stranger’ versus ‘non-stranger’ abduction
   - Percentage of teenage abduction compared to ‘baby snatching’

2. Profile of non-parental abduction victims:
   - Personality characteristics
   - Situational characteristics

3. Patterns of non-parental abduction in Canada, including:
   - Spatial patterning of the abductor
   - Temporal patterning

Non-parental abduction is well researched in the United States. A concern is that Canadian literature depicts the majority of non-parental abductions in Canada are by friends or relatives, whereas in the United States it is found that the majority of abductions are by ‘true strangers’; therefore, analyzing American literature in this paper demonstrates how research in Canada should focus on utilizing similar profiling techniques. One cannot however assume that patterns in the United States are comparable with Canada’s.

In relation to the involuntary category of altered state of consciousness exit-seekers, there is a striking lack of research pertaining to environmental aspects of wandering. A few studies have been done pertaining to the profiles of wanderers which include characteristics of the wandering behaviour, and typical environmental aspects (Koester & Stooksbury, 1995; Lucero, 2002; Hope et al, 2001). Further research must be directed towards determining the environmental factors that may play a role in wandering behaviour; for example, what the typical mode of transportation is (i.e., walking, buses), or where wanderers typically head, towards a point last known to them, towards a downtown centre or do they just walk aimlessly? Researchers should also
focus on basic issues surrounding the patterns of dementia related wandering behaviour. With research pertaining to these questions above, and spatial representation of wanderer's point of last seen, and point of being found, this may further our understanding of wanderers' exit-seeking behaviour patterns.

The next category discussed was runaways. Deficiencies are emphasized in terms of the profile of runaway children. Areas seen throughout the literature include types and differences among runaways. A common aspect of runaways is the pattern of behaviour with respect to frequency and consistency of running. Suggestions for research in this area include:

1. Patterns of running:
   - Frequency of the event of being 'missing';
   - Temporal patterning;
   - Spatial patterns in terms of the behaviour while 'missing';

2. Personality of a runaway including:
   - Social characteristics (including patterns of behaviour in school, work or home);
   - Personality characteristics;

3. Defining clearer boundaries between a runaway and a throwaway.

Although it is clear that research is sufficient in some areas, the concern is that research has not looked at the interaction between people and subsequent patterns. In order to explore the behaviour of a runaway, patterns and personality may need to be studied as a whole, rather than individually.

In terms of research recommendations for throwaways, the list is rather general and similar to that of runaways. The reason for this is that the main suggestion for future research in this area is that exploratory analysis needs to be conducted in order to understand the phenomenon of throwaways. The above recommendations were
highlighted based on concern over the literature in Canada and the United States. This section has depicted the current research in Canada, and highlighted the current abduction trends in Canada, providing an overview of required research, in order to fully understand typologies of missing children.

Reliance on American literature can illuminate the deficient areas of Canadian research; however, rates and trends of missing children are not directly comparable to Canadian research. In order to understand patterns, profiles and situations, research is required to create specific typologies of missing persons in Canada. These typologies will lead to further understanding and prediction of voluntary and involuntary acts of missing persons. Unfortunately the data used for this study did not contain victim or offender information; therefore, an exploratory spatial data analysis was conducted in order to identify whether further research should be conducted regarding spatial characteristics of missing and found persons.
Statistical measures are used to explain or predict a phenomenon under study. To break this term down, a measure is a method used to obtain a result and a statistic is a number describing the characteristics of a sample (Miller & Whitehead, 1996). Traditionally, statistical measures used in analysing crime data have relied on criminological methods from the social sciences. The concern with using only criminological methods is that crime is not static in space and time. Within the social sciences, statistical analysis has recognized the importance of temporal factors; yet, it has failed to emphasize the concept of space in criminological measures.

A trend in the past 20 years has been to conduct crime analysis utilizing statistics from the field of geography. The reason for such a shift in methods is from the addition of spatial factors within the statistical analysis. The objective of this study was to explore and explain common geographical methods used in analysing the phenomenon of crime. In order to achieve this objective, this section provides a detailed overview of the typical methods that may be utilized in conducting exploratory spatial data analysis (ESDA).

Exploratory spatial data analysis (ESDA) is an inductive approach to data analysis used to provide insight about spatial patterns without necessarily having a formulated hypothesis about the outcome (Anselin & Getis, as cited in Chakravorty & Pelfrey Jr, 2000). The first step in this data-driven process is a confirmatory approach. This approach was utilized in this study by first conducting a visual inspection of the data by looking at a bivariate distribution of the missing and found persons point data. The second stage tested for complete spatial randomness (CSR). This is conducted by
utilizing CSR techniques, described later in this section, that test whether clustering is present in comparison with a random pattern. The final stage in the exploratory spatial data analysis process is statistical testing of the data. This was conducted using global techniques such as spatial autocorrelation.

The following is a description of several methods for conducting the exploratory data analysis including techniques and methods for conducting analysis. Not all of these techniques are appropriate for all levels of data; therefore, not all of the techniques below are used in this study, but they are still necessary to provide an overview of exploratory spatial data analysis.

The objective of this section is to introduce various methods of cluster analysis in their application to spatial data analysis. Besag and Newell (1991) categorized cluster analysis into three categories: global statistics, local statistics and focused statistics (as quoted in Rogerson, 2001). As Rogerson (2001) describes, global statistics test for a deviation from randomness. Their intention is to test the null hypothesis, and to test for an overall pattern in the data (Rogerson, 2001). If no significant deviation from randomness can be found using global statistics, then the second category of local statistics may be used to uncover hotspots (Rogerson, 2001). Local tests evaluate the clusters around specific locations (Rogerson, 2001), such as a crime rate surrounding a shopping mall.

Finally, the third category of focused statistics is utilized for uncovering the size and location of clusters (Rogerson, 2001). Examples of tests for this method include Diggle's method and the score test of Lawson and Waller. Global and local methods test for clustering without a location component pre-determined in the hypotheses; therefore, this paper focuses on the discussion surrounding only these two approaches. Below is a
diagram of the statistics that will be discussed, categorized according to their local/point and global/polygon measures.

**Classification of Measures for Spatial Analysis**

<table>
<thead>
<tr>
<th>Global / Polygon Measures</th>
<th>Local / Point Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Autocorrelation</strong></td>
<td><strong>Complete Spatial Randomness Test (CRS)</strong></td>
</tr>
<tr>
<td>- Moran’s I Statistic</td>
<td>- GHAT</td>
</tr>
<tr>
<td>- Geary’s C Statistic</td>
<td>- FHAT</td>
</tr>
<tr>
<td><strong>G Statistic</strong></td>
<td><strong>Spatial Intensity</strong></td>
</tr>
<tr>
<td><strong>Spatial Autoregression</strong></td>
<td><strong>Nearest Neighbor Analysis (NNA)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Standard Deviation Ellipses (SDE’s)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Local Indicators of Spatial Association (Lisa)</strong></td>
</tr>
<tr>
<td></td>
<td>- Moran’s I</td>
</tr>
<tr>
<td></td>
<td>- Geary’s C</td>
</tr>
</tbody>
</table>

Figure 1 Classification of Measures for Spatial Analysis

The first stage in exploratory data analysis is visual inspection of the data on a map or planar surface. Visualization makes it possible to see patterns with the human eye that aggregate spatial statistics do not identify. For example, statistical analysis of the point data may conclude that the points demonstrate complete spatial randomness; however, this should not mean the end to the analysis, because other patterns may exist.
locally within the data. Visualization allows one to look at the 'big picture' and the 'small picture' at the same time.

**Local/Point Data Analysis**

**Complete Spatial Randomness (CSR)**

The main assumption of complete spatial randomness is that the events must conform to a *homogenous Poisson process* across the study area (Bailey & Gatrell, 1995). Two conditions that should be considered when conducting CSR are point independence and uniformity. Unfortunately, in criminology, these conditions are often violated; for example, the condition of independence implies that “the selection of a location for a point in no way influences the selection of locations for any other point” (Boots & Getis, 1988, 13). This condition is sometimes violated because criminologists frequently work with exhaustive samples and certain locations do influence crime; for example, a person's primary activity node such as a home or work, may influence the choice of where to commit an offence.

The second condition that is often violated in criminology is uniformity. This condition implies that all locations must have an equal chance of being selected within the study area (Boots & Getis, 1988). This is not always the case with crime. If the crime being analyzed was residential break and enter, then industrial locations within the study area do not have an equal chance of being selected; therefore, researchers must be cautious of their results when violating these statistical assumptions.

The two tests for complete spatial randomness include the nearest neighbor distance method $\hat{G}$ and the origin to point method $\hat{F}$. The nearest neighbor method specifically looks at interactions between points on a small scale (Kaluzny et al., 1998).
This method “defines a nearest neighbor distance $d_i$ as the distance from the $i$th point to the nearest other point in $A$, the bounded region of interest” (Kaluzny et al., 1998, 152).

The GHAT is expressed as:

$$\hat{G}(y) = n \cdot \sum_{d_i \leq y} 1$$

*Where $n$ is the number of points in $A$

**Equation 1** GHAT Test for Spatial Randomness. Source: (Kaluzny et al., 1998).

A pattern is considered clustered if “the points are significantly more grouped in the study area than they are in CSR” (Boots & Getis, 1988, 15); therefore, in the interpretation of the GHAT, clustering exists when there is an excess of short distance neighbors (Kaluzny et al., 1998). Whereas, if regularity exists within the point distribution, then an excess of long distance neighbors would be expected (Kaluzny et al., 1998).

This interpretation is the opposite for the interpretation of an FHAT. The FHAT is an origin-to-point nearest neighbor statistic. It is “defined by overlaying a $k \times k$ grid on the region $A$, then comparing the distances from the $m$ resulting origins to their nearest neighbor” (Kaluzny et al., 1998, 155) and is defined as:

$$\hat{F}(x) = m \cdot \sum_{e_i \leq x} 1$$

*Where $e_i$ is the distance from the $i$th origin to the closest of the $n$ points in the data.

**Equation 2** FHAT Test for Spatial Randomness. Source: (Kaluzny et al., 1998).

The interpretation of the FHAT is the opposite of the GHAT. For example, if there is an excess of long distances, then clustering is evident, whereas regularity in the distribution
is found through an excess of short distances (Kaluzny et al., 1998). Once the methods for complete spatial randomness have been conducted, the next stage of exploratory spatial data analysis would be to conduct spatial intensity to determine the intensity of clustering that exists.

**Spatial Intensity**

Several methods of spatial intensity exist. For the purpose of this thesis, only first-order intensity will be discussed. The properties of the first-order process of a point distribution describe the intensity of the point through space based on the number of points per the unit area (Kaluzny et al., 1998). Several types of smoothing techniques currently exist, but the two methods that will be discussed are binning using a loess smoothing and kernel estimation. The “binning method uses a two-dimensional histogram to form rectangular bins” (Kaluzny et al., 1998, 159). This method displays the intensity of the point data throughout the entire study area allocating intensity estimates varying throughout the surface area.

Kernel smoothing is a slightly different technique in which spatial smoothing of different bandwidths around the data points is used to determine density of point data. The result of spatial smoothing shows crime density as a “spatially continuous variable, with ‘peaks’ representing areas of high crime (hot spots) and ‘valleys’ representing areas of low crime” (McLafferty, Williamson, & McGuire, 2000, 79). The output created depicts the density at each location in the study area similar to the binning method. Kernel smoothing accounts not only for the amount of points in a given area, but provides further consideration to the spatial arrangement of the points; for example, if two grid points have the same amount of points in them, their density will be different if the spatial
pattern of the two sets differ (McLafferty, Williamson, & McGuire, 2000). Thus, closer points are weighted as higher density for kernel smoothing.

The overall benefit to Kernel smoothing is that it can also analyze a temporal component of hot spots. This method allows for the overlay of hot spots in order to analyze change in size, shape and location over a period of time (McLafferty, Williamson, & McGuire, 2000). Hot circles and ellipses have been used to identify areas where the highest density of crime occurs (McLafferty, Williamson, & McGuire, 2000). It has been noted however, that a problem with using hot circle analysis is that it does not account for irregularly shaped areas (McLafferty, Williamson, & McGuire, 2000).

Another technique in the analysis of point data is the nearest neighbor index, which calculates distances between each point to determine how far clustering is present based on distance between points.

**Nearest Neighbor Analysis Index (NNA)**

Comparison of nearest points is calculated by analyzing on the distances that would be expected on the basis of chance (Levine, 2002). The nearest neighbor index equation, as outlined below is calculated based on d(NN) which is the calculation of the nearest neighbor distance between points, and the d(ran), which is the calculation of the mean random distance. The index is then a ratio based on these two calculations expressed as (Levine, 2002):

\[
NNI = \frac{d(NN)}{d(ran)}
\]

*Equation 3 Nearest Neighbor Index. Source: (Levine, 2002).*
The “index compares the average distance from the closest neighbor to each point with a distance that would be expected on the basis of chance” (Levine, 2002, 173). The interpretation of the NNI is that the closer it is to 1.0, the less clustering exists, whereas less than 1.0 indicates clustering is present. Finally, after conducting nearest neighbor indexes to determine distances by neighbor order, it is important to create standard deviation ellipses to determine the dispersion of the point data.

**Standard Deviation Ellipses (SDE)**

Robert Langworthy and Eric Jeffris (2000) describe standard deviation ellipses as visual depictions of observations around major and minor axes in two-dimensional space. The major axis is considered a line of best fit, whereas the minor axis is perpendicular to the major axis which runs through the mean centre (Langworthy & Jeffris, 2000). The variation is considered the standard deviation around the axes (Langworthy & Jeffris, 2000). The coefficient of circularity \((CC)^3\) is produced from this measurement of variation from the major to minor axes and this coefficient represents the amount of linearity in the distribution (Langworthy & Jeffris, 2000).

The area of the standard deviation ellipse (SDE) is calculated by \(\pi \times S_x \times S_y\) (Langworthy & Jeffris, 2000). The requirements necessary to compute a SDE is a mean latitude (X), a mean longitude (Y), the angle of rotation, a standard deviation for the transformed Y, a standard deviation for the transformed X, calculation of the coefficient of circularity \((CC)^3\), and the calculated area of the ellipse (Langworthy & Jeffris, 2000). With these results, statistical analysis can then be performed on the ellipse. Analysis is necessary in order to test for significance based on the mean centres, diffuse distributions and shape of the distribution (Langworthy & Jeffris, 2000).
The primary reason for testing the mean centres of the ellipse is to compare differences between two groups; for example, with the distribution of missing and found persons point data a t-test can be performed in order to compare these groups. With the t-test analysis it can be determined whether there is any significant difference between the latitude and longitude of group 1 in comparison to group 2.

The second measurement that may be performed is to explore the diffusion within the distribution by testing for equality of variance (Langworthy & Jeffris, 2000). In order to determine any significant difference between the groups, examination of the transformed standard deviation for each group can be compared (Langworthy & Jeffris, 2000).

Shape of the distribution can be determined by analysing the output of the coefficient of circularity. The coefficient determines the amount to which the ellipse is linear or circular. To determine circularity, a CC that is approaching 1 indicates circularity in the point distribution, whereas a CC that is approaching 0 indicates linearity in the distribution of the point (Langworthy & Jeffris, 2000).

The practicality of standard deviation ellipses is to determine not only hot spots, but also to identify changes associated with hot spots by comparison of ellipses. The changes can represent temporal changes, such as a shift of linearity or circularity over time. An example of this would be the changes that would occur from vehicle theft in the daytime compared to vehicle theft during the evening hours. Standard deviation ellipses also determine changes in the area. For example, vehicle theft may occur more frequently in residential areas during the evening hours; however, the ellipse may show change during the daytime hours, illustrating higher theft in commercial areas.
As mentioned before, there are several variations to these techniques that can be utilized in exploratory data analysis, but they are beyond the scope of this paper. One method of analysis, local indicators of spatial association, is considered a local measure, yet, in this study it is utilized with polygon level data and therefore will be discussed in the global measures section of this thesis. Local point data analysis is only the first stage of exploratory analysis that is incorporated into this study. It is important to review the methods and techniques of global polygon methods in order to not only understand what type of clustering may exist, but to statistically analyze what level of association is present in the data.

**Global/Polygon Data Analysis**

**Spatial Clustering**

Deductive approaches in the study of crime and criminality typically utilize correlative statistics to test for a relationship between two variables. Correlation determines whether the relationship between two variables was produced based on low levels of chance. In criminology $p \leq .05$ and $p \leq .01$ are frequently used. Correlation cannot infer anything other than the probability of a relationship existing between the variables.

In order to examine characteristics between the variables, a measure such as Pearson's $r$ can be used. Pearson's $r$ is a measure of a linear relationship between two variables (Miller & Whitehead, 1996). The measure determines whether the relationship is significant, and with what strength and direction the variables are associated.

When comparing correlation and autocorrelation, the key difference lies in the concept of space. Correlation utilizes non-spatial data, whereas autocorrelation accounts for the relationship based on spatial ordering [arrangement] of variables.
Therefore, spatial autocorrelation is defined as a relationship among values of a variable that is attributed to spatial ordering (Griffith, 1987).

To visually illustrate the spatial ordering, authors such as Daniel Griffith explain the relationship through the depiction of a matrix of observations. Griffith’s example is derived from point data that “can be described as a configuration for the n observed values of some variable X, and may be depicted by an n-by-n table or matrix of values…” (Griffith, 1987, 9). This matrix model created by Griffith is utilized below to demonstrate how spatial autocorrelation is calculated.

The first approach to demonstrating autocorrelation is to create a binary matrix. This matrix is a table in which each entry represents the spatial contiguity of the variables. Griffith explains that,

Zero entries in this matrix would indicate that the corresponding row and column observations are not side-by-side, or in juxtaposition with one another, in the associated ordering. Unity entries in this matrix would indicate that the corresponding row and column areal units are juxtaposed (1987, 9).

A zero entered into the matrix depicts no spatial juxtaposition (adjacent) relationship between the pair of observations on a planar surface; however, a one (or unity entry) indicates that the two observations are juxtaposed on a planar map indicating a contiguous relationship. The ordering of observations that is produced in the matrix describes how areal units are spatially arranged on a planar map (Griffith, 1987).

In order to visually illustrate this concept of spatial autocorrelation, Griffith uses a series of figures to depict the theoretical construction of autocorrelation. All figures presented below are a personal interpretation of Griffith’s original figures (Griffith, 1987).
The initial figure presented is of the distribution of points (observations) on a planar map. Consider these points to be car thefts around a mall.

After the observations are arranged on the map, the next step is to connect each dot together with a line.

Next, Griffith (1987) adds a perpendicular bisector on the lines between each point.
The final step is the creation of a thiessen polygon by connecting the bisectors. The reason for creating the polygons is to depict how the theoretical boundaries are delineated. This boundary illustrates the juxtaposition of each point with its neighbor.

![Figure 5 Creation of Thiessen Polygon. Source: (Adapted from Griffith, 1987)](image)

As seen through the illustrations above, the original points (observations) are transformed into theoretical polygons. These polygons decipher the borders in order to determine which observation is juxtaposed with another observation. Remember that a single observation can never be adjacent to itself; therefore, the entry on the matrix would be zero in the case of an observation made on the same variable. The matrix for this example would look as follows:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 Contiguity Matrix. Source: (Adapted from Griffith, 1987)

Each variable on the planar surface is illustrated in the matrix as a one or a zero depending on whether the observations share a common border with the other points.
This process illustrates Griffith’s theoretical approach to understanding the foundation of autocorrelative statistics. Therefore, spatial autocorrelation is best described “as the relationship among values of some variable that is attributable to the manner in which the corresponding areal units are ordered or arranged on a planar surface” (Griffith, 1987, 11). Autocorrelation is useful in the identification of clusters; however there are multiple statistics that can be applied to identify and analyze such clusters. The first stage in cluster analysis is defining the neighbors within a study region.

**Defining Nearest Neighbors**

The initial stage of data analysis is to define the neighbors where the information exists for sub-regions or a polygon division of the study area. The reason for this is that polygon level analysis is based on the aggregated counts which are cartographically displayed in a boundary form. This study utilized the census tracts from Census Canada 1996 as the boundary and therefore, will discuss boundaries based on census tract data.

A “spatial process is modelled by predicting the outcomes of each region based partially on its dependence on nearby or neighbouring regions” (Kaluzny et al., 1998, 111). There are several methods for defining neighbors. They can be measured by a specified distance from one region to another, or by sharing a common border. For this study, nearest neighbors are defined by their juxtaposition with another region, and therefore first-order neighbor method was utilized. First-order method consists of creating a contiguity matrix based on adjacency of borders. Once the nearest neighbor contiguity matrix has been created, polygon analysis can be performed using the neighbor matrix as a neighbor object in the structuring of spatial correlations.
Spatial Autocorrelation

Geary’s C Statistic

Several methods can be used for determining presence and direction of spatial autocorrelation including the Geary’s C and Moran’s I statistics. Both tests measure for the presence of autocorrelation using interval/ratio scales (Griffith & Amrhein, 1991). The purpose of these statistical methods is to determine whether an overall clustering exists. They do not determine hot spots; rather they determine the significance of a cluster of points. Geary’s C is represented by the following equation:

$$c(d) = \frac{\sum \sum w_{ij}(d)(x_i - x_j)2/2W(d)}{\sum (x_i - \bar{x})^2/(n-1)}$$

Where:

- $w_{ij}$ = elements of a weight matrix for which a value of 1 indicates a pair of two samples, $x_i$ and $x_j$, are in the distance class $d$ and a value of 0 indicates all of the other cases.
- $W(d)$ = the sum of $w_{ij}(d)$

Equation 4 Geary’s C Statistic. Source: (Fortin, Dale & Hoef, 2002, s)

The Geary’s C statistic tests for clusters of similar values in order to provide results of autocorrelation or randomness found in the analysis of the ordering of observations. Both the Geary’s C and Moran’s I are very similar mathematically, but a slight variation in the statistics produces a different interpretation of the data. The results of conducting this measure include identification of similar or dissimilar clustering. Below is a summary detailing the interpretation of the results for Geary’s C.
In comparison, simple correlative tests such as Pearson's $r$, determine whether two variables show a significant relationship based on probability. Global methods are similar; however, they determine a relationship based on a spatial dependence of variables. These methods determine significance based on spatial dynamics such as the arrangement of observations on a map, rather than a non-dimensional analysis. Moran's $I$ is also utilized for determining spatial autocorrelation but has a slightly different interpretation.

**Moran's $I$ Statistic**

Moran's $I$ is more common as compared to other methods based on its ability to measure the degree of spatial autocorrelation in areal data (Rogerson, 2001). Moran's $I$ analyzes areal data while adding a component of spatial weighting. Weighting can be expressed in several different forms "such as simple contiguity (having a common border), distance contiguity (having centroids within the critical distance band), or in a function of inverse distance or squared inverse distance (Anselin 1992, as quoted by Chakrovorty & Pelfrey Jr., 2000, 69).

Spatial weighting is an important aspect as it adds a measurement of distance, which is key to spatial analysis. An example of this is weighting based on observations having commonality of a distance of 0.25 mile from each centroid being observed.
Provided with this distinction, Moran's I would be calculated by taking into account the spatial weighting assigned. Moran's I is therefore represented by the following equation:

\[ I(d) = \frac{\sum \sum w_{ij}(d)(x_i - \bar{x})(x_j - \bar{x})/W(d)}{\sum (x_i - \bar{x})^2/n} \]

Where:

- \( w_{ij} \) = elements of a weight matrix for which a value of 1 indicates a pair of two samples, \( x_i \) and \( x_j \), are in the distance class \( d \) and a value of 0 indicates all of the other cases.
- \( W(d) \) = the sum of \( w_{ij}(d) \)

Equation 5 Moran's I Statistic. Source: (Fortin, Dale & Hoef, 2002, 5)

In this study, observations would be aggregated to census tracts. Moran's I accounts for the number of observations in each census tract and accounts for the weighting or spatial proximity between the tracts (i.e., distance from centroids).

When reading the output of Moran's I, it is determined that values nearing +1 indicate strong spatial clustering of high values, and high spatial clustering of low values (Rogerson, 2001; Chakrovorty & Pelfrey Jr., 2000). Values that approach 0 indicate an absence of spatial clustering (Rogerson, 2001).

<table>
<thead>
<tr>
<th>Moran's I</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Similar X values clustering = Positive Spatial Autocorrelation Assumed</td>
</tr>
<tr>
<td>-1</td>
<td>Dissimilar X values clustering = Negative Spatial Autocorrelation Assumed</td>
</tr>
<tr>
<td>0</td>
<td>Random; No Spatial Autocorrelation Assumed</td>
</tr>
</tbody>
</table>

Table 3 Moran's I Interpretation Table. Source (Griffith & Amrhein, 1991, 137-138)
A more common method of analysing the Moran's I is to standardize the score in order to make the analysis and interpretation comparable. Therefore, transformation of the variable:

\[ z = \frac{(x - \bar{x})}{s} \]

Results in the expression:

\[ I = \frac{n \sum \sum w_{ij} z_i z_j}{(n-1) \sum \sum w_{ij}} \]

Where:
- \( n \) = the number of regions
- \( w_{ij} \) = a measure of spatial proximity between regions \( i \) and \( j \)

Equation 6 Moran's I Standardized. Source: (Rogerson, 2001, 167)

In regard to the analysis of this standardized equation, Rogerson states that "pairs of regions where both regions exhibit above average scores (or below average scores) will contribute positive terms to the numerator" (2001, 165) equating to positive spatial autocorrelation. For example, depicted below is the Moran's Scatterplot, which is a plot of spatially lagged values of each observation graphed against the standardized observation (Chakrovorty & Pelfrey Jr., 2000). Each quadrant provides a point for each observation.

The results describe each tract mean compared to its neighbor's mean for the variable being measured. If a score measured a tract with a crime rate below the mean, as well as a neighbor's rate also below the mean, they would be plotted in the lower-left quadrant (Chakrovorty & Pelfrey Jr., 2000). If a rate were found above the mean for a
tract, as well as above the mean for its neighbor, than the point would be plotted in the higher-right quadrant.

\[
\begin{array}{c|c}
\text{Low High} & \bullet & \text{High High} \\
\hline
\text{Low Low} & \bullet & \text{High Low}
\end{array}
\]

Figure 6 Observation Plot

If no spatial autocorrelation were indicated from the calculations, then these points would form a cloud (grouping) around their origin (Chakrovorty & Pelfrey Jr., 2000).

These local measures are very useful for identification of clusters from a generalized and global approach; for example, with missing persons data, the census data could be analyzed using these methods to determine the overall clusters of missing persons in neighborhoods with high crime rates. If you want to determine the clusters around a particular location that has a high crime rate, then local measure are necessary. The following are two local measures, local Gi statistic and local Moran’s I.

**Local Indicators of Spatial Association**

**Local Gi Statistic**

If no significant deviation from randomness is found with Moran’s I, then local statistics can be used to possibly identify hot spots. Local Moran’s statistic and Getis Gi statistic were chosen for discussion because they evaluate clusters based on particular locations using areal data (Rogerson, 2001). Where global statistics measure patterns on a general level, local statistics supersede global in that they focus on distinguishing clusters of values between each parcel or census tract rather than general overall
patterns (Chakrovorty & Pelfrey Jr., 2000). The G statistic which is actually a global measure, is represented by the following equation:

\[ G(d) = \sum_{ij} w_{ij}(d) x_i x_j / \sum_{j} x_j \]

Where:
- \( x_i \) = Is the value observed at location \( i \),
- \( w_{ij}(d) \) = Is derived from an non-standardized contiguity matrix derived, from \( d \)
- \( d = \) Critical distance (which is the distance within which every \( j \) is considered a neighbor in the case of the numerator)


The Getis G statistic as depicted above is considered a global statistic. The addition that differentiates the \( G_i \) statistic from its global statistic is the addition of local measures. With this addition of localizing, the measure is then:

\[ G_i = \sum_{j} w_{ij}(d) x_j / \sum_j x_j \]

Where \( G_i \) is a derived measure for every parcel \( i \) where:
- \( x_i \) = Is the value observed at location \( i \),
- \( w_{ij}(d) \) = Is derived from an non-standardized contiguity matrix derived, from \( d \)
- \( d = \) Critical distance (which is the distance within which every \( j \) is considered a neighbor in the case of the numerator) differing by the number of observations in the denominator.

Equation 8 Local \( G_i \) Statistic. Source: (Chakrovorty & Pelfrey Jr., 2000, 73).

This \( G_i \) statistic is superior as a local statistic because it derives a measure for every parcel analyzed, rather than all parcels measured together; for example, a measure would be determined for each census tract crime rate and missing persons clusters, rather than the global approach of measuring all tracts combined.
Local Moran Statistic

The same is true for the local Moran statistic regarding the addition of a measurement within each parcel in the areal data. The local Moran statistic is delineated by the following equation:

\[ I_i = n (y_i - \bar{y}) \sum_{j \neq i} w_{ij} (y_j - \bar{y}) \]

Where:
- \( n \) = the number of regions
- \( w_{ij} \) = is a measure of spatial proximity between regions \( i \) and \( j \) (the sum of the local Moran's is equal to the global Moran's)

Equation 9 Local Moran's I Statistic. Source: (Rogerson, 2001)

This local Moran statistic accounts for the specific parcel that is being measured; therefore, "the sum of local Moran's is equal to, up to a constant of proportionality, the global Moran; [i.e., \( \sum I_i = I \)]" (Rogerson, 2001, 173). In other words, each census tract (as described in the Moran's I statistics section) has its own calculable local Moran statistic. All local Moran's combined would equal the global Moran. Finally, if spatial autocorrelation or association were present within the data, autoregressive models would be needed to further explore predictive models.

Spatial Autoregressive Models

The final area in conducting statistics on spatial data that is important to creating a solid model is spatial autoregressive statistics. In order to understand spatial autoregressive models, basic regression will be examined.
Regression is a very common statistic used in the social sciences based on its predictive ability. The basic function of regression is that it determines the amount of change in the dependent variable (Y), which occurs from change in the independent variable (X). The regression equation is described as follows:

\[ Y = \alpha + \beta X + \varepsilon \]

Where:
- \( Y \) = the dependent variable
- \( X \) = the independent variable
- \( \alpha \) = the y-intercept
- \( \beta \) = the slope
- \( \varepsilon \) = residual

Equation 10 Regression Equation

The equation illustrates how regression is used for prediction. Given the equation above, the independent variable (X) is used as the predictor variable, and the dependent variable (Y) is the variable that is being predicted given a change in (X).

The problem with regression is that it is not conducive to spatial factors. It is a model used to predict on a non-dimensional field. In order to more accurately predict the (Y) variable spatially, a geographic component is added to the equation in order to account for two-dimensional surface analysis. Many regression models exist in spatial statistics. A simple explanation will be provided for this study in order to explore basic spatial autoregressive models.

When attempting to analyze a geographical component of regression, it is important to address spatial dependency (Fotheringham, Brunsdon & Charlton, 2000); for example, consider dividing the city into small enumeration areas. It is typically found that variables in an enumeration area will have similar results to enumeration areas that
are adjacent; therefore, an extra explanatory variable is needed in a spatial autoregressive model in order to account for spatial dependency (Fotheringham, Brunsdon & Charlton, 2000). This is achieved by calculating a mean of the rates for the juxtaposed zones and transforming the dependent variable into a vector of adjacent means. Once this adjacent mean variable is calculated, it is added to the regression equation which produces the following:

\[ Y = X\beta + \rho Wy + \epsilon \]

Where:
\[ \rho = \text{regression coefficient for the adjacent-mean variable} \]
\[ \beta = \text{vector of regression coefficients} \]
\[ \epsilon = \text{vector of random errors} \]

Equation 11 Spatial Autoregression. Source: (Fotheringham, Brunsdon & Charlton, 2000, 168)

The final stage to the regression is to apply a maximum likelihood estimation method. The reason for this is to be able to estimate the likelihood of variable (Y) occurring, given the values of \( \rho \) and \( \beta \) (Fotheringham, Brunsdon & Charlton, 2000, 168). Through this use of spatial autoregressive statistics, stronger prediction can be made on the dependent variables.

The concept of crime is not static in time and space. The global and local methods in this study demonstrate geographical measures that can be used in the study of crime. These methods achieve the goal of measuring changes in crime spatially and temporally; for example, social science measures can account for temporal variables, but they are usually unable to consider the aspect of crime weighting from the centroid of each observation. In addition, they do not account for the juxtaposition of each observation, and therefore are non-dimension linear measures.
The analysis of missing persons was, for example, exploratory spatial data analysis because it contains point data, along with areal census data. The measures discussed can be used with any crime data provided the data requirements of point or areal data are met. Geographical measures are not new. This study demonstrates how they are a vital component in crime analysis when attempting to measure a relationship using environmental variables.
CHAPTER FOUR: METHODS

When conducting spatial analysis, several areas need to be addressed regarding the type of data necessary for conducting statistical tests. The objective of this study was to explore patterns of missing and found persons in Vancouver through use of spatial analysis.

Sample

The sample obtained was secondary data from the Vancouver Police Department. This was an exhaustive sample for the time period studied (1996). It contains all cases of reported missing persons (N=2627) and reported found persons (N=625) from January 1st, 1996 to December 31st, 1996 (See Table 1 Below). The study is probability based and the results are generalizable to Vancouver in 1996.

Point Data

The two data sets obtained from the Vancouver Police Department detailed in the table below, contain point data for reports of missing and found persons in 1996. Point data is where each observation (i.e., missing person), is represented on a map by a single point (Fotheringham, Brunsdon & Charlton, 2000). The reason for choosing the 1996 data was based on availability of accurate census data. In Canada, the census is collected on the first and sixth year of each decade. Therefore, to accurately measure the phenomenon of missing persons, the researcher chose to analyze the missing persons data most compatible with the census tract and variables for that time period. The dataset consists of all calls for service, and missing and found reports from the Vancouver Police Department in 1996. Calls for service are all reports made directly to
the police. The tables below describe each variable set in which only address information was used.

---

### Vancouver Police Departments Missing Persons Data:
**Calls for Service in 1996**

**Variable Set # A**
- Reported address of missing person
- Date person was reported missing
- Time person was reported missing

Table 4 Missing Persons Variable Set A

---

### Vancouver Police Departments Found Persons Data:
**Calls for Service in 1996**

**Variable Set # B**
- Reported address of found person
- Date person was reported found
- Time person was reported found

Table 5 Found Persons Variable Set B

Analysis was conducted using a geographical mapping program called ArcView. Both variable sets (A and B) contain the reported addresses of the missing and found calls for service. The addresses are extracted from the police database and geocoded into point data for analysis. Geocoding data into ArcView is a process of creating X,Y coordinates (non-dimensional points) for each address in order to create a visual depiction of the address on a basemap. A basemap is a combination of street and land networks for any given area (i.e., Vancouver).

### Areal Data

The table below presents a list of census variables that were collected for the exploratory spatial data analysis. The census variables are considered areal data and
were used for descriptive visual analysis. Areal data is point data that has been aggregated and represented in a two-dimensional polygon. The rate of unemployment would be considered areal data if it were representative rate for a specific census tract. Instead of each house representing its unemployment rate on the map, it is represented within an aggregated rate in its tract (polygon).

The list below displays each variable chosen from the census database. Appendix B provides information on each variable including the population sample size for each variable, the definition of each variable, and the categories included in the census survey. Census data summarizes data over a spatial field called a polygon. In this study, the polygon is a boundary called a census tract.

<table>
<thead>
<tr>
<th>Canadian Census Data by Census Tract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable Set # 3</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Gender Demographics</td>
</tr>
<tr>
<td>Age Demographics</td>
</tr>
<tr>
<td>Education Demographics</td>
</tr>
<tr>
<td>Income Demographics</td>
</tr>
<tr>
<td>Unemployment Demographics</td>
</tr>
</tbody>
</table>

Table 6 Canadian Census Data by Census Tract

Access to the census data was achieved through license agreement with the Research Data Library at Simon Fraser University. Census tract data were downloaded from the library in a Beyond 20/20 Browser format. Census tracts were chosen based on consideration of the modifiable areal unit boundary problems (MAUB), which are further discussed in chapter 6.

The census data were transferred to a database format at which point it was assigned a number that corresponds to its Digital Boundary File (DBF). A Digital
Boundary File is a polygon theme overlaid on the basemap. This file delineates the boundaries of the census tracts for Vancouver (Research Data Library, 1996) allowing for census boundaries to be cartographically displayed. Once this DBF is created, the corresponding numbers of each census tract (i.e., 933 0007.00 – Vancouver) are then associated with the correct demographics for that tract.

Finally, once each attribute table is created, the census demographics are converted into a shape file. These shape files can be added as a theme to the basemap, allowing for a layering of data; for example, using a basemap, a shape file of census tract boundaries is added, and then the theme of income is added. The reason for doing this is to cartographically display the varying levels of income across Vancouver by each census tract defined. Once all shape files are created, analysis can be conducted on the data (See Appendix C for detailed data steps on Missing Persons, Appendix D for found persons and Appendix E for census data steps).

**Statistical Analysis Procedures**

Statistical analysis for this study was conducted using S-Plus, CrimeStat II and the ArcView spatial statistics module. After producing the geocoded data and the attribute tables for all variables, the next step was to display the data cartographically. A choropleth map creates a visual depiction of summarized or grouped counts (observations) (Clarke, as quoted by Sutton, 1999); therefore, census tracts and boundaries can be displayed in the form of a choropleth map. A point map can be created by plotting the coordinates for missing persons and found persons. Both maps create a visual display, and provide the foundation for conducting spatial analysis. The following section details the procedures that were conducted in the exploratory spatial data analysis of missing and found persons in Vancouver 1996.
Point Data:

The first stage of the analysis was to visually analyze the point data in a bivariate form. The missing and found persons point data were graphically displayed by x and y coordinates on a scatterplot. This bivariate representation produced through S-Plus demonstrated the importance of further analysis of clusters based on the apparent visual clustering that existed in the downtown eastside (Refer to the results section for further interpretation).

The second stage of analysis was to test for complete spatial randomness. As defined in the previous chapter, a test for CSR is important in point pattern analysis because it determines whether the point arrangement is based on random chance or spatial clustering. Spatial randomness contains two assumptions: that the point data intensity does not vary; and, that there are no interactions between the points (MathSoft, 2000).

The FHAT and GHAT tests were then conducted on these coordinates producing a plot of the spatial randomness by distance respectively from neighbor or origin respectively. The test for CSR produced results that illustrated possible spatial clustering beyond chance; therefore, the next stage of analysis was to conduct a spatial intensity plot. This was conducted in order to illustrate the smoothing estimate of the point pattern displaying spatial intensity of possible clusters. Again, the data were set location 1 as X and location 2 as Y for each of the two datasets. The method chosen for an intensity test was binning and the smoothing parameter selected was 0.25. Binning was chosen as it “uses a two-dimensional histogram to form rectangular bins” (MathSoft, 2000, 52). The analysis was conducted with contour plots, filled contour plots and surface plots including the points in order to present the best visual representation of the.
data. Point pattern intensity was more dispersed for the missing persons data set. A higher level of intensity was evident in the found persons data.

Nearest neighbor indexes were calculated to determine how far each point is from its neighbor. This index helps determine whether there are larger distances within the data, or whether spatial clustering exists at a smaller distance from its neighbor. Nearest neighbor analysis was conducted using CrimeStat II. Each dataset was defined in CrimeStat II and the number of nearest neighbors computed was 50. The surface area was defined as 114kms, and there were no border correction methods chosen. Border corrections are methods of dealing with boundary problems, which will be discussed further in the final chapter of this study.

After the nearest neighbor index was created, the data was imported into S-Plus in order to create a two-dimensional plot of the index by the order of the neighbors. In interpreting the nearest neighbor plot, values closer to 1.0 indicate lack of clustering. A general rule is that values below approximately 0.7 indicate spatial clustering. As will be discussed in the results chapter of this paper, clustering was evident for both datasets, but given that the clustering of missing persons data points was more dispersed, a final measure, standard deviation ellipse analysis was conducted.

Standard deviation ellipses are the most appropriate method for determining the dispersion of the data points. This technique allows for the comparison of missing and found person points by looking at the circularity of the ellipse in order to determine dispersion in the patterns of the data. Analysis was conducted using CrimeStat II software. In conducting the analysis, both datasets were both analyzed by calculating mean centres, standard deviational ellipses and median centres. Once these were calculated, the data for each were saved as shape files and imported into ArcView. The
shape files were then overlaid on the point data for each dataset, allowing for visual and statistical analysis of dispersion within the missing and found persons datasets. After ellipses were analyzed, it was necessary to see if the patterns that were evident in the point data had any autocorrelative results when analyzed as an aggregate polygon form.

**Polygon Data:**

The first step in analysing polygon shape data was to define spatial neighbors. This process was necessary, as discussed in chapter three, in order to create a matrix that defines each polygon’s neighbors. The neighbor calculation utilized boundaries based on the census tracts from Vancouver census data for 1996. The neighbors can be defined by several methods. The method that was most appropriate for this study was defining first order neighbors based on adjacency. This procedure defined neighbors based on the juxtaposition of all census tracts bordering on a target census tract. This procedure was conducted through the spatial analyst utility in ArcView. The first order neighbors were created and then used to define neighbor boundaries when conducting spatial autocorrelation tests and identify local indicators of spatial association.

The second stage of analysis of polygon data was to calculate spatial autocorrelation. It was in the interest of exploratory spatial data analysis to determine whether the point patterns, once aggregated to polygon form, provide any information about whether the clustering that was present was independent of its neighbors, or whether the interactions were dependent on their neighbor. If the data are spatially autocorrelated, then it is important to conduct spatial modelling by use of autoregressive methods.
Spatial autocorrelation was conducted on both the missing person and found person datasets. This procedure was conducted on aggregated point data in counts, as well as in rates. It is the belief of the author that both methods were necessary in the exploratory analysis of this phenomenon. The first reason is that typically data were standardized in a rate form in order to create standardized tracts based on population for analysis. In the case of spatial data analysis, it is important to note that standardization by population may not always be the most appropriate method for point data analysis; therefore, both procedures were conducted, and issues relating to both of these arguments will be further discussed in chapter 6 of this study.

Spatial autocorrelation testing was conducted using the nearest neighbor defined for each missing and found persons based on Vancouver census tracts. Given the exploratory nature of this study, it was important to compare results from different statistical calculations; therefore, Moran’s I and Geary’s C statistics were both chosen as methods for analysis. The sampling type chosen was nonfree and the number of permutations run for each set were 1000. The Moran’s I and Geary’s C results using both standardized and non-standardized data indicated differing conclusions about whether autocorrelation was present within the data, so local indicators of spatial association (LISA) were calculated in order to assess whether association was present within the data or not.

LISA testing (local indicators of spatial association) is used when spatial autocorrelation is not statistically significant, yet apparent visual clustering seems to exist within certain neighborhood boundaries. LISA was conducted using Local Moran’s and Geary’s C statistics on both the missing and found datasets, using 1000 permutations. The results, as will be discussed in the following chapter, indicate that
association does exist and modelling of the data may be necessary for future exploration of the data.
CHAPTER FIVE: RESULTS

The results are discussed in three parts: the first section deals with the general descriptive analysis of the missing and found persons calls for service data, along with an overview of the census data for Vancouver in 1996; the second section of the results deals with the analysis of missing and found persons point data; and the third section is a discussion of the results for polygon data analysis of missing and found calls for service.

### Total Missing and Found Persons Calls For Service By Type

<table>
<thead>
<tr>
<th>Call For Service Type</th>
<th>Total Call Count 1996</th>
<th>Summed Counts by Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Persons</td>
<td>N = 2627</td>
<td>N = 1207</td>
</tr>
<tr>
<td>Found Persons</td>
<td>N = 625</td>
<td>N = 527</td>
</tr>
<tr>
<td>Total Calls Missing &amp;</td>
<td>N = 3252</td>
<td>N = 1734</td>
</tr>
<tr>
<td>Found Persons 1996</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7 Frequencies of Calls For Service By Type

As depicted above, the total calls for service count for missing persons (N=2627) was a great deal larger than the amount of calls for found persons (N=625). The discrepancy between counts does pose some methodological concerns and therefore will be highlighted in the discussion section of this study. Below is a cartographical display of the City of Vancouver (Figure 7), missing persons point data (Figure 8), missing persons polygon data (Figure 9), found persons point data (Figure 10), and found persons polygon data (Figure 11) across Vancouver in 1996.

**Descriptive Analysis**

For the purpose of this thesis, the map below depicts the city of Vancouver, the major roadways, the downtown core and suburbs for visualization purposes.
Figure 7: Map of the City of Vancouver 1996
From a visual interpretation of the point data in Figure 8, it is apparent that missing persons have high counts at repeated addresses. The overall picture illustrates that slight clustering may exist; however, it looks as though it is not concentrated in one specific area within Vancouver, but it is distributed throughout the residential areas in the city.
Vancouver: There is a distinct pattern in the visualization of the data. East of Hastings Street, even though there are reports of found persons throughout, depicts a different pattern in the form of clusters of found persons exist on the downtown. The found persons point data show a different picture (Figure 10). Visual analysis

Figure 9: Missing Persons Polygon Data Vancouver 1996

Aggregate Count of Missing Persons by Census Tract

Figure: From the polygon representation below (Figure 9) it is even more apparent that the data
The next area of visualization is purely descriptive analysis using missing and found persons in relation to a few socio-economic variables. Statistical analysis was not conducted in this comparison of the rates of missing and found persons with census variables because it was beyond the scope of this paper. The census data were cartographically displayed and overlaid with missing persons and found persons point data purely for visual descriptive purpose. The results were suggestive and call out for future analysis.

The first census polygon description that was created was for average income. Typically, lower income areas have played a large role in explaining patterns of crime; however, typologies of missing persons are lacking in explanatory spatial and social models. Therefore, it was important to create a cartographical representation of areas...
where people typically go missing layered on the average income from the neighborhood in which they are missing. As seen below, visual analysis indicates that the majority of missing persons are last seen in the lower to moderate income areas.

Found persons are also highly concentrated in moderate to low income areas as seen in Figure 13 below. These maps indicate low numbers of missing and found persons in high-income areas. The issue is whether income is actually an explanatory variable or whether another phenomenon may exist. For example, the large clusters that contain up to 110 missing persons in a year may possibly be institutions such as group homes or hospitals. Found persons visually cluster around the downtown area.
however, this may not be explained by the low income levels of the parts of the city, but rather explained by routine activities and normal mobility toward the city centre.

Based on the fact that there are not a large number of care homes, group homes, etc. in higher income areas, the visualization may be misleading. Real estate market and planning decisions may result in the grouping of institutions within lower to moderate income areas. This should be studied in future research on missing persons.

The next visual description is the rate of female missing persons aged 10-19. The reason for conducting this research was based on research from the Solicitor General of Canada's Police Reference Manual for Cases of Child Abduction and
Runaway Youth (1993). In this report Weiler et al. found that 90% of the runaways are females between the ages of ten and eighteen years old (1993). Based on American research that repeatedly indicated the same results as Weiler et al., it was necessary to run a cartographical representation of the missing persons counts overlaid on polygon shape file of females from ages ten to nineteen.

![Figure 14 Missing Persons Point Data by Rate of Females Ages 10-19 Years](image)

It is evident from this map that further statistical analysis is necessary. There is one cluster that has extremely high counts of missing persons as well as a high rate of females aged ten to nineteen. The conclusion is that missing persons seems somewhat
dispersed and further analysis would be necessary to determine if age and gender are characteristics for missing persons spatial dispersion.

As mentioned previously, the data were divided into point and polygon data analysis. Analysis conducted on the point data included testing for complete spatial randomness (CSR), spatial intensity, nearest neighbor analysis, and standard deviation ellipses. The methods used on polygon data include spatial autocorrelation and local indicators of spatial association. The following section will outline the results that were found from the analysis of the point data.

**Point Data Analysis**

The first stage of point data analysis includes visualization of the bivariate point distributions. Figure 15 below depicts the point locations of the missing persons calls for service along with the found persons call for service location data. These addresses are displayed on a scatterplot according to their X and Y coordinates. The first visual depiction of the point data highlights an immediate pattern demonstrating a shift between the two types of data. For example, it is apparent that the missing persons data is spread throughout the suburbs and downtown of Vancouver.
Figure 15 Missing & Found Persons Bivariate Scatterplot by X, Y Locations
Looking at the found persons locations, it is apparent that there is a higher frequency of people reported missing in the suburbs, and the point concentration of found persons was located on the downtown eastside of Vancouver. This visualization of the data raises the awareness that patterns may be evident and therefore further analysis must be conducted to determine whether this patterning in the visual analysis is based on randomness or whether clusters truly exist.

**Spatial Randomness**

A method to determine whether these points are random or clustered relies on a test for complete spatial randomness (CSR). The two methods of testing for CSR that were utilized were the GHAT and the FHAT statistics (for description of these methods refer to Chapter 3 Spatial Analysis). The GHAT was the first statistic ran, which determined randomness based on point-to-point methods of determining clustering from its nearest neighbor.

In the analysis of GHAT results, clustering is present when an excess of short distances between points occurs. Dispersed data or regularity in data occurs when there is an excess of long distances between points in the study area (Kaluzny et al., 1998).
Figure 16 above depicts clustering in the missing persons data. The excess of short distances between the points in the study area lead one to conclude that there is a possibility of clustering that does exist within the missing persons point locations. Therefore, it is apparent that the data were not necessarily random, but that there is a pattern in the point-to-point data. This does not infer strength; it only tests for randomness of points.

Similarly, Figure 17 below demonstrates the same results for found persons point data. There is a very high excess of short distances between points in the data. This suggests that the data were not random, but that there is clustering in the point data of found persons locations.
The second form of testing for complete spatial randomness within the data is the FHAT. This statistic differs from the GHAT in that it determines whether clustering exists based on an origin-to-point statistic. The statistic is read slightly differently from the GHAT. The FHAT interpretation demonstrates clustering when there is an excess of large distances between points and the grid origin; whereas regularity or dispersed data are represented by shorter distances between the point and grid origin.

The results of the FHAT below indicate comparable results to the GHAT. Regarding Figure 18, there is an apparent excess of larger distances from the grid origin to the points of the data, thus implying clustering of the missing persons data was present.
The FHAT results for found persons point data is comparable to the GHAT. Larger distances are more obvious in the FHAT results of found persons points. Figure 19 below demonstrates a spatial clustering found from point-to-grid origin.
After reviewing the results for complete spatial randomness (CSR), it was determined that further testing of the data would be beneficial. Results evidently depict spatial clustering and are not based solely on randomness. Therefore, the next stage in point analysis is to determine the intensity of the clustering.

**Spatial Intensity**

Spatial intensity entails the surface output of clustering by intensity; therefore, the highest cluster of values will be the centre point of the intensity grid by X and Y coordinates. The further rings from the centre delineate the least amount of intensity in the clusters. According to the results of the intensity plot, missing persons point data were found to have the highest intensity dispersing towards neighborhoods surrounding the downtown leading to suburban areas.
Figure 20: Spatial Intensity Plot of Missing Persons by X,Y Locations Using Binning
The intensity plot for the found persons data (Figure 21), produced different results. Rather than the two intensity rings being dispersed as seen in the missing persons data, the intensity plot for found persons showed smaller dispersion in the rings. Only one ring existed which was concentrated primarily on the downtown eastside of Vancouver (See Appendix F for the surface plots of the intensity maps).
Figure 21 Spatial Intensity Plot of Found Persons by X,Y Locations Using Binning
The results of these intensity plots for missing and found persons point data highlight an interesting pattern. From a visualization of the data intensity it is apparent that people are more often being reported missing from suburban Vancouver and the found reports are concentrated towards the downtown eastside of Vancouver. The next step in exploratory analysis was to determine how far the clustering exists from its closest neighbor by using a form of nearest neighbor analysis.

**Nearest Neighbor Index Analysis**

Nearest neighbor indexes were created to demonstrate the distances between neighbors order of clustering. As depicted below in Figure 22, the missing persons clustering is sparse clustering between neighbors. Interpretation of this plot indicates that a slight clustering of the data points is evident within the first five neighbors; however, after the first five neighbors it becomes more distributed throughout the study area.

![Nearest Neighbor Analysis - Missing Persons](image)

*Figure 22 Nearest Neighbor Analysis Index for Missing Persons*
Found persons analysis of nearest neighbor index is slightly stronger than that of the missing persons index. It demonstrates a tighter clustering within the first two neighbors, but then it is more drastically distributed after the first five neighbors.

![Nearest Neighbor Analysis - Found Persons](image)

Figure 23 Nearest Neighbor Analysis Index for Found Persons

Interpretation of the nearest neighbor indices indicates that both datasets contain clustering in the data. The missing persons data depicts a larger area of neighbors being clustered. The found persons data indicates that one particular area may be highly concentrated with found persons. The next stage in exploring these distances and clusters was conducted through standard deviational ellipses in order to understand the pattern and dispersion of the points.

**Standard Deviational Ellipses**

Standard deviational ellipses are conducted to determine the dispersion of the data across the study area. For example, the more circular the ellipse, the more concentrated the distribution will be. The size and the shape of the ellipse will determine
what dispersion of points influence data analysis. Illustrated in Figure 24 is a map of the combined standard deviational ellipses conducted for both missing and found persons. This cartographical representation illustrates the blue ellipse as the found persons with the blue dot as the mean centre of the data points; the red indicates the missing persons ellipse and mean centre. Each ellipse was conducted individually and then overlaid on the map.
Figure 24 Standard Deviational Ellipses for Missing and Found Persons
Interpretation of the ellipses together indicates that the smaller, more circular shape of the found persons ellipse illustrates a smaller dispersion within the data. There are not many clusters that draw the ellipse away from its mean centre; however, looking at the missing persons data points, it is obvious that the data are more dispersed and more clusters are drawing the ellipse away from its circular shape to a more elongated linear ellipse.

Interpreting these individually is also important in order to understand why the ellipse is elongated or circular. The map below is of the found persons standard deviational ellipse.

Figure 25 Standard Deviational Ellipse for Found Persons Point Data
It is apparent through visualization of the SDE that two patterns within the data are contributing to the stretching of the circularity of the ellipse. The first cluster is the downtown eastside, which is represented by the points at the northwest corner of the ellipse. This pattern is drawing the ellipse towards a more elongated shape because of the second pattern, which is at the centre of the ellipse running in an east/west direction. This second pattern is a clustering of points along a primary road (Broadway) within Vancouver. These two patterns, along with the dispersion of the points spread through the remainder of Vancouver, create an elongated ellipse. The missing persons data, as displayed below, indicate a more dispersed pattern as compared to the found persons ellipse, thus resulting in a more elongated shape.
The SDE for missing persons illustrates three patterns that are contributing to the elongating of the ellipse. The first cluster is the downtown eastside / Hastings Street which is found at the north and northwest portions of the ellipse. This clustering stretches the ellipse out from the mean centre based on the strength of the other two clusters. The second cluster is again the same as found persons; it is located along a primary road (Broadway), where the mean centre is located running in an east/west direction. The final cluster stretches the ellipse to a more linear shape and is located at the southern portion of the ellipse. This is a location in which several group homes, hospitals and care facilities are assumed to possibly have some influence in the clustering of the data. These three patterns create more of a linear ellipse rather than a circular. This implies that clustering may exist, but it is more linear and dispersed, whereas the found data indicates more circular and clustered patterning.

The next stage of the results introduces measures of analysis for polygon level data. Creating polygons was a process of aggregating each of the found and missing persons data points into polygon boundaries. Once these data were aggregated, it was then analyzed by polygon methods to determine statistical relationships within the data based on the patterns that have been identified above.

**Polygon Data Analysis**

Analysis of polygon data is conducted at two levels of analysis. The first level is global which is an analysis of the data on a large scale measuring patterns throughout the study area. The second level is local analysis which allows for patterns to be found within smaller areas of study boundaries.
The first stage of analysis entailed the defining of spatial neighbors and the spatial autocorrelation of the data. If no spatial autocorrelation was present then local measures, such as the LISA statistic could be used to determine whether association of the data exists at a smaller level. If spatial autocorrelation was present in the data, then regression modelling should be conducted in order to create an explanatory model. Therefore, polygon analysis was conducted by first defining the spatial neighbors of the data and then conducting a test for autocorrelation.

**Defining neighbors:**

The neighbors were defined using the ArcView spatial neighbors method for calculating neighbor weights. The first-order method for calculating neighbor weights based on polygon boundaries was performed on the data. This method was chosen because the missing and found persons data was being analyzed on an aggregate level utilizing census tracts (polygons) as the boundaries for analysis.

The first-order neighbor analysis creates a matrix depicting polygons that are adjacent to its neighboring polygon (sharing a common border). The result of defining the neighbors is that the matrix created specified 92 original polygons (census tracts) sharing a total of 534 common neighbor borders. After these neighbors were defined, they were used in the computation of the spatial autocorrelation of missing and found persons.

**Spatial Autocorrelation**

These results were produced using Moran's I and Geary's C statistics for determining the presence of spatial autocorrelation. An interesting issue in determining autocorrelation was whether or not to standardize the data when conducting analysis.
Typically researchers have utilized both methods in the past, and therefore, as mentioned in the methods chapter of this study, both Moran’s I and Geary’s C autocorrelative tests were conducted on both non-standardized and standardized data. Results will be discussed and issues surrounding the use of each statistic and level of data will be discussed in chapter 6.

**Moran’s I Results: Non-Standardized Polygon Analysis**

After the neighbors were defined, exploratory spatial data analysis was conducted to test for spatial autocorrelation within both the missing and found persons datasets. In order for clusters to exist within the data using Moran’s I, the correlation value must be greater than 0, and the permutation p-values must be statistically significant ($p<.05$). The summary chart below displays the presence of spatial autocorrelation in the missing persons standardized dataset at, $I = 0.1034$ which was significant at a permutation p-value $0.044$. It is apparent that through the analysis of Moran’s I statistic, spatial autocorrelation is present in the data, meaning that clustering of similar values at the polygon level of analysis is evident. Therefore, we are able to reject the null hypothesis of no spatial autocorrelation in the missing persons non-standardized data.

The spatial autocorrelation for the found persons calls for service data were also conducted using the Moran’s I statistic. There was a significant presence of spatial autocorrelation in the found persons data as well, $I = 0.3507$ which is highly significant at a permutation p-value $0.00$. Therefore, the null hypothesis of no spatial autocorrelation is rejected, and the presence of spatial clustering of values in the found persons calls for service data is accepted. However, these results were not similar when testing the same data using the Geary’s C Statistic.
Spatial Autocorrelation: Non-Standardized – Moran’s I Results

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Variance</th>
<th>Std.Error</th>
<th>Normal p-value (2-sided)</th>
<th>Permutation p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Persons</td>
<td>0.1034</td>
<td>0.003457</td>
<td>0.0588</td>
<td>0.05179</td>
<td>0.044</td>
</tr>
<tr>
<td>Found Persons</td>
<td>0.3507</td>
<td>0.003457</td>
<td>0.0588</td>
<td>7.701e-10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8 Spatial Autocorrelation: Non-Standardized Moran’s I Results

Geary’s C Results: Non-Standardized Polygon Analysis

As discussed in chapter 3, the two tests for spatial autocorrelation are the Moran’s I and Geary’s C statistic. They both test for the presence of spatial autocorrelation, but they are slightly different in their mathematical computations. Therefore, both were conducted to see if any discrepancies in the computations existed. In order to determine whether the results of the Geary’s C statistic are significant, the results are interpreted as: if $0 < C < 1$ then clustering exists, if $C = 1$ or greater no clustering exists.

The results below show that there is only a very slight presence of spatial autocorrelation in the missing persons data because the value is approaching 1, $C = 0.9023$, however the result is not significant permutation p-value = 0.175. Therefore, because of the slight presence of clustering, and the lack of significant findings, we fail to reject the null hypothesis of no spatial autocorrelation in the missing persons data.

The Geary’s C statistic was then performed using the same procedures and produced similar results for the found persons calls for service data. The table below demonstrated that again there is a very slight level of clustering existing, $C = 0.8519$, however the results are not significant, $p = 0.218$. Therefore, we fail to reject the null hypothesis of no spatial autocorrelation present in the found persons data.
Table 9 Spatial Autocorrelation: Non-Standardized Geary’s C Results

The key point regarding the results discussed is that two similar tests, the Moran’s I and the Geary’s C, produced completely different results using the exact same datasets. The next step in the analysis would be to create spatial regression modelling when spatial autocorrelation is present in the data. However, because of the differing results in the autocorrelative tests, it is important to conduct a local test on the Geary’s C data in order to determine whether there is a spatial association. Local indicators of spatial association (LISA), were performed on the Geary’s C statistics to identify if spatial clustering occurred locally rather than only globally as seen in the results above.

Local Indicators of Spatial Association - Non-Standardized

Seeing the results of the Geary’s C test indicated there was no spatial autocorrelation present in the data, local measures of association were used to determine whether the patterns are associated. Local measure of association (LISA) was then calculated on the missing and found persons data using a Geary’s C statistic to determine whether spatial association was present at a local level. After computation for both data sets were conducted, the Geary’s C statistic was mapped for both data sets. Below in Figure 26 and Figure 28 are the cartographical representations of the Geary’s C z-statistics for missing and found persons.
Figure 27 Local Indicators of Spatial Association for Missing Persons by Zi on Non-Standardized Data Using Geary's C

Figure 28 Local Indicators of Spatial Association for Found Persons by Zi on Non-Standardized Data Using Geary's C
According to the Geary’s C statistics run on LISA, it is very apparent that some form of spatial association does exist. The missing persons map displays association clustered in the suburban areas. The found persons map visually depicts the Geary’s C association occurring in the downtown eastside of Vancouver. Therefore, though Geary’s C displayed no spatial autocorrelation on a global level, it did indicate that a slight association exists at a local level of analysis.

The results of spatial autocorrelation and spatial association globally and locally indicate that there are patterns within the data that suggest spatial association exits. The final stage to the analysis is then to fit a regression model using the spatial autocorrelation for the Moran’s I statistic in order to create spatial modelling of the data. However, as mentioned previously, as part of the exploratory process of this study, it is important to compare the results of the standardized polygon data to the non-standardized. Therefore, the same statistical procedures were conducted on the datasets using standardized rates of missing persons using population as the denominator.

**Moran’s I Results: Standardized Polygon Analysis**

The analysis and interpretation for the standardized polygon data were identical to the non-standardized data. The only difference was in the results of the statistical tests. The missing persons point data were standardized, based on a 1,000-population denominator. For the results of the missing persons standardized analysis of spatial autocorrelation using Moran’s I, the results were not statistically significant and no spatial autocorrelation was found at $I = 0.06734$, permutation p-value = 0.085.
This was not the case for the found persons dataset. The found persons data were also standardized at a 1,000-population denominator. The results for Moran’s I was statistically significant and clustering was present within the data at $I = 0.3731$ and a permutation p-value = 0.00.

**Spatial Autocorrelation: Standardized – Moran’s I Results**

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Variance</th>
<th>Std.Error</th>
<th>Normal p-value (2-sided)</th>
<th>Permutation p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Persons</td>
<td>0.06734</td>
<td>0.003457</td>
<td>0.0588</td>
<td>0.1828</td>
<td>0.085</td>
</tr>
<tr>
<td>Found Persons</td>
<td>0.3731</td>
<td>0.003457</td>
<td>0.0588</td>
<td>6.436e-11</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 10 Spatial Autocorrelation: Standardized Moran’s I Results

**Geary’s C Results: Standardized Polygon Analysis**

The Geary’s C results for the standardized data for both missing and found persons were consistent with the results of the Moran’s I analysis. The procedures for computation were identical to the non-standardized data. The only difference again is that a 1,000-population standardization was used as the denominator.

The results of the spatial autocorrelation using Geary’s C for the missing persons standardized data produced a relationship that was not statistically significant indicating no spatial clustering was present within the data. The results were reported at an $C = 0.9599$, and permutation p-value of 0.314.

Similar to the result of the Moran’s I statistics, Geary’s C provided statistically significant results for the standardized found persons data. The results indicate that there is definite spatial autocorrelation within the data at an $C = 0.5686$, and a permutation p-value of 0.00.
**Spatial Autocorrelation: Standardized – Geary’s C Results**

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>Variance</th>
<th>Std.Error</th>
<th>Normal p-value (2-sided)</th>
<th>Permutation p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Persons</td>
<td>0.9599</td>
<td>0.005672</td>
<td>0.07531</td>
<td>0.5944</td>
<td>0.314</td>
</tr>
<tr>
<td>Found Persons</td>
<td>0.58686</td>
<td>0.005672</td>
<td>0.07531</td>
<td>1.014e-8</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 11 Spatial Autocorrelation: Standardized Geary’s C Results

The interesting note regarding these tests was that when the data were not standardized, the Moran’s I results indicated spatial autocorrelation for both missing and found persons; whereas in the standardized analysis, neither of the statistical tests indicated any presence of autocorrelation for missing persons. With this lack of spatial autocorrelation in the Geary's test, local indicators of spatial analysis was conducted to determine if there were any spatial association. In keeping consistent with the analysis of the non-standardized data, LISA was conducted using local Moran’s I to determine if association was present within the missing persons data.

**Local Indicators of Spatial Association - Standardized**

When spatial autocorrelation is not found within the data a further test, local indicators of spatial association, can be utilized to determine if clustering exists at a local level of analysis. Therefore, LISA was conducted on the missing standardized dataset in order to determine whether association may be present locally, even though clustering did not exist globally. Below (Figure 29) is the map of the standardized z-scores from the local Moran’s I test for association.
From visual interpretation of the map above, it is evident that once the missing persons data were standardized, there is slight but very minimal association that exists. Therefore, within the analysis of the standardized missing persons data, we fail to reject the null hypothesis of spatial autocorrelation and association. Below is a summary chart of the results of both standardized and non-standardized spatial autocorrelative tests for missing and found persons.

**Table 12 Spatial Autocorrelation Comparison of Standardized and Non-Standardized Data**

<table>
<thead>
<tr>
<th></th>
<th>Moran's I Significant</th>
<th>Geary's C Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Non-Standardized</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Found Non-Standardized</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Missing Standardized</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Found Standardized</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>
The final stage of data analysis would be to create a spatial autoregressive model for the data that was autocorrelated to create a stronger explanatory model, however this is beyond the scope of this study. As mentioned in the previous chapter, future research should explore regression models with variables such as the number of hospitals, group homes, care facilities, shopping malls etc., in order to create a stronger model of explaining spatial factors of missing and found persons.
CHAPTER SIX: DISCUSSION

This particular study was unique in that it analyzed a phenomenon that has never been studied spatially. With such a lack of research in this area, this study was conducted in an exploratory manner in order to highlight whether spatial research should be continued in this area. This chapter initially outlines the results of this study supported by environmental criminological theory and current missing person's research. The second section identifies the strengths and weaknesses applicable to the exploratory spatial data analysis that was conducted on the missing and found persons calls for service data. Third, issues pertaining to the implementation of GIS in crime analysis are covered highlighting the positive and negative aspects of crime mapping. And finally, recommendations for future research are discussed to enhance future missing and found persons typologies in Canada.

Current typologies of missing persons in Canada are extremely deficient. Unfortunately, very little research has been conducted on this phenomenon in Canada and it has never been spatially analyzed on an international level. Therefore, the objective of this study was to conduct an exploratory spatial data analysis of the missing and found persons calls for service data in order to demonstrate the value of not only analyzing data using criminological statistics, but also the incorporating of other methods of analysis such as spatial measures.

The results of this study were not black and white. Current spatial methods of analysis, similar to other disciplines of analysis, are not perfect for analyzing crime in space; however, the analysis conducted does provide insight into the phenomenon of
missing and found persons indicating that further research is necessary to reliably inform current typologies of missing persons.

Through a test for complete spatial randomness of the point data, results indicate that both missing and found persons data exhibit clustering that is not based on randomness. This is consistent with the premise of environmental criminology in that crime is not seen as random, but rather demonstrates that individuals' routine activities, and environmental and social backcloth contribute to creating a clustering of events. If a person travels every day to and from work by the Sky Train to downtown Vancouver, then later in life enters a nursing home and falls ill to Alzheimer's or some form of dementia, the path travelled would be that of the routine best known. Therefore, knowing individuals' routines, along with known hot spots of found persons, could be used to enhance the search for the missing person. The results of the test for spatial randomness were not surprising as they follow the theoretical foundation that routine activities and environmental factors typically dictate where clusters may occur.

Spatial intensity was then conducted to determine intensity estimates across the study area for missing and found persons. It was found that the estimate for missing persons was most intense at the downtown / East Hastings Street area as well as an intensity hot spot found in a southern suburban area of Vancouver. The remainder of estimates were dispersed throughout the residential areas. This is somewhat consistent with literature in that there have been few environmental factors discussed in the global missing persons typologies. Environmental factors do contribute to individual typologies. The concern with this is that each type of missing person is unique, and studying the combined data of missing persons allows for analysis globally but not individually. This may indicate why clusters were distributed throughout the city of Vancouver. For example, the “point of last seen” for a runaway will typically be their residence or group
home. Therefore, clusters are higher where repeat calls are made in residential areas. However, Alzheimer’s patients typically leave the care facility or hospital, and walk along major roadways creating a different “point of last seen” scenario (Koester & Stookbury, 1995). Therefore the runaway and dementia cases will be independent of each other, causing dispersion in the intensity estimates of missing persons as a combined typology.

The found persons intensity was not nearly as dispersed as the missing persons point data. The found persons contained intensity estimates that did not widely disperse throughout the city of Vancouver. Rather, the estimates demonstrated a tight cluster occurring specifically in the downtown and East Hastings Street area. Again, this is supported by routine activities in that people tend to be attracted towards the downtown core of a city. Thus, found persons had stronger clustered intensity estimates, whereas the missing persons points created a more dispersed estimate.

Nearest neighbor analysis indexes were then created to measure the order based on distance from point-to-point. This order was plotted by the neighbors index and order to demonstrate how far from each neighbor distance clusters existed. It was found that clustering existed within the first five neighbors and gradually increased for missing persons, indicating a larger number of clusters existed within the study area; however, found persons had tight clustering occurring within the first two neighbors, and minimal clustering throughout the remainder of Vancouver.

Finally, standard deviational ellipses (SDE) were conducted in order to determine whether circularity or linearity occurred within the dispersion of the data. These ellipses indicated that the found persons point data demonstrate a more circular ellipse than the missing persons. It was also found that two clusters of found persons were elongating the ellipse to a less circular shape. The missing persons ellipse indicated that three
patterns existed, and that these three patterns created a more elongated, linear ellipse indicating more dispersion within the point data.

This is consistent with criminality of place theories in that it is believed that place has an effect on the location of an event occurring (Eck & Wiesburd, 1995). In this study, missing persons was a combination of all types of missing, and therefore, clusters would be dependent on the surroundings. For example, some of the clusters that occurred in the data were of group home, hospital and shopping mall locations. Unfortunately, it was beyond the scope of this study to elaborate on exact details pertaining to these locations. Nonetheless, it is necessary to highlight that the point data from missing persons initially illustrated some type of dependence on the type of missing persons and location. Therefore, when a person is reported missing, it would be expected that the data may be slightly more dispersed and location dependent. Further research into the connection between place and type of missing persons would be a vital contribution to future research on typologies.

Polygon analysis was then conducted to statistically determine whether autocorrelation was present within the data. Standardized and non-standardized data were used for both datasets. These results are extremely interesting; when the data were manipulated into a rate rather than the aggregate of counts, the data presented opposing results. For the Moran’s I test on the non-standardized count data, spatial autocorrelation was found to be statistically significant for both missing and found data. However, Geary’s C indicated that there was no spatial autocorrelation present for either dataset when conducted with the same non-standardized data. Once the data were standardized, Moran’s I and Geary’s C results both specified that there was no spatial autocorrelation present in the missing persons data, yet spatial autocorrelation was present in the found persons data. Past research demonstrates analysis was conducted
using both methods; however, it has not been found that either method is incorrect. The only consideration may be that standardization based on population may be a more accurate form of analysis.

It is of the author's opinion that in order to discuss the end results, the data must be looked at cumulatively. The point data indicated that spatial clustering is apparent in both datasets. The clustering was more evident in the found persons calls for service. It was also found that the missing persons data were more dispersed than the found persons data. Once the global analysis was conducted, differing results of autocorrelation lead us to look at local indicators of spatial association. This test demonstrated that association did, to some extent, exist within the missing persons, but not to any large degree.

The final conclusion regarding the exploratory spatial data analysis of the missing and found persons calls for service is that spatial association and clustering is somewhat evident in missing persons calls for service. Further steps in analysis would be necessary to separate each category and analyze them individually in order to see if the clusters are more apparent. In regard to the found persons calls for service, clusters and spatial autocorrelation were present. The final stage of research would be to conduct spatial autoregression to create an explanatory model of found persons calls for service. After conducting this exploratory spatial data analysis, it is important to highlight the strengths and weaknesses of the methods and procedures utilized in this study.

**Future Research Considerations**

There were several strengths and weaknesses that need to be addressed pertaining to future research considerations regarding the exploratory spatial data
analysis of missing persons. The first area is in regard to the violation of data assumptions for statistical analysis. Unfortunately, geographic and social methods of analysis are not created with the idea of crime in the calculations. Therefore, a weakness in using spatial methods is that two of the data assumptions of testing for complete spatial randomness tend to be violated when analyzing crime data; however, not all crime data violate these assumptions, but they are important to consider when conducting research.

The first assumption that can be violated is the concept of uniformity. When analyzing spatial randomness, it is assumed that every area within the boundaries being studied has an equal chance of receiving a point. This is an assumption that is not always accurate when analyzing criminological data; for example, if theft of cars were being analyzed, then typically cars would be stolen from roadways or parking lots. City park areas however, (i.e., Stanley Park) would not have an equal chance of receiving a point; therefore, criminologists must be aware that this assumption can be violated when analyzing complete spatial randomness.

In this study, the data were fairly consistent with uniformity in that a person can be reported missing or found from practically anywhere within the study area; therefore, this condition was not violated, in regard to the analysis of missing and found persons. The second condition that is typically violated in criminology is of independence. This implies that the points should be independent of each other and are not interacting. A case in which this condition could be violated is when analyzing data of criminal event location and offenders activity nodes. The offender's activity nodes such as housing, work, school etc. can influence where they choose to do the crime. Thus, the condition in this case would be violated, because the points may influence each other.
In the analysis of the missing and found persons data, it is assumed that a reported missing person from location ‘A’ most likely does not influence the report of missing persons from location ‘B’. The only case that the data points may lose independence is when repeat calls occur at the same location. In this case, the data may violate the condition of independence. Though the statistics are not entirely perfected in this area for analyzing crime data, they continue to enhance the level of analysis that currently exists in criminology.

The primary strength of this research is the ability to meet the objective set out in this study. The objective entailed an exploratory spatial data analysis of the missing and found persons calls for service data in order to highlight the value of utilizing spatial methods in criminological research. It was found that patterns and spatial association were apparent within the datasets not based on randomness, indicating that spatial methods enhance the current knowledge of missing and found persons spatial factors. It was also determined that spatial autoregressive models are necessary to determine an explanatory model of ‘why’ the calls for service are clustering at various locations. The strength of this study was that spatial patterns were clearly identified indicating further research must be directed towards uncovering the reasons for these patterns.

A weakness of this study is the lack of spatial typologies and simple typologies of missing and found persons that exist in Canada. Unfortunately, research is extremely lacking on missing persons. Typologies have been attempted in the United States; however, they fail to incorporate analysis of spatial factors in the typologies. Without proper spatial analysis to inform the typologies, analysis relies solely on theory. The theoretical support in this study was founded on environmental criminology, which illustrates the importance of studying missing and found persons spatially; however,
environmental factors must be studied more in depth pertaining to the typologies that exist in order to incorporate spatial factors in explanatory regression modelling.

Another weakness was the low autocorrelative results of the missing persons standardized data. A future consideration would be to conduct the spatial analysis using individual typologies of missing persons rather than the cumulative database as was used in this study. Unfortunately, because of data constraints this was not possible. Future research could focus on the individual typologies to highlight whether patterns exist within each type of missing persons categorizations, rather than globally across all types of missing persons.

A final possibility regarding future research in this area is that of conducting analysis using border correction methods. The modifiable area unit boundary and edge effect may have slightly influenced the results of the data analysis (which is discussed further in the next section); therefore, future research should attempt the spatial analysis of missing persons, allowing for border corrections to demonstrate clustering with point averaging for nearest neighbors that fall outside of the boundary. The next section of this thesis is a discussion of issues pertaining to the implementation of spatial analysis in criminology.

**Implementation Issues of Spatial Analysis**

The analysis of crime has, for decades, relied on spatial environmental factors. It has not been until more recently that analysis was conducted through the use of geographic technology. Geographical information systems have allowed criminologists to create descriptive and analytic cartographic representations of crime in space. The concern is that software and concepts which have been "borrowed" from geography and
applied to crime have resulted in the development of an imperfect crime analysis process. This section highlights issues with crime mapping including: modifiable area unit boundary problems, edge effects, descriptive versus analytic purposes as well as the practicality of crime mapping in police agencies. Crime mapping is consistently increasing police recognition of environmental factors that influence crime. Even with such benefits of mapping, the negatives associated with the process create some reluctance in the acceptance of mapping as a crime-fighting tool.

The first area to address is in regard to the modifiable area unit boundary problem (MAUB). Areal-based data is considered crime data that has been aggregated into geographical areas (McLafferty, Williamson, & McGuire, 2000). The reason for aggregation is that many police departments measure crimes based on precincts, districts or census tracts. A problem with areal measurement is that if crimes occur along area boundaries or intersection boundaries (McLafferty, Williamson, & McGuire, 2000), analysis may not be accurate or may not be accurately represented.

A second problem with the data is that "changes in the number and configuration of areas used can strongly affect the results" (McLafferty, Williamson, & McGuire, 2000, 78). This is called the modifiable areal unit boundary problem. It has been argued that current spatial studies may produce a different result depending on how the zone boundaries are delineated.

One of the problems associated with boundaries is point weighting. This is a concern when clusters may occur toward one edge of a boundary but not the other edge. For example, diagram "A" below depicts crime event locations in a boundary as point data. Once the data is aggregated, the analysis assumes a consistency across the boundary as seen in diagram "B". The problem is that it does not weight the polygon
according to the amount of points in each section of the boundary; rather it is seen as a constant across the entire boundary.

A – Point data before aggregation

B – Aggregate shape file of A

Figure 30 Point Observations

Note the majority of points are clustered to the lower right quadrant of the boundary in diagram "A". Once the data is aggregated to polygon form for analysis, the boundary points are represented by a constant over the entire boundary.

Edge effects are another concern that occurs when comparing neighbors, as neighbors can fall outside of the study boundary (Brantingham & Brantingham, 1995). If this occurs it can potentially bias the results of the nearest neighbor calculation. An example of this is in the analysis of missing persons point data where the nearest neighbor point may be in Burnaby, which is a city that shares a border with Vancouver. Also, clustering of points that are adjacent to water can also create an edge effect. If this occurs, the program picks a point closest to the neighbor being defined to create a neighbor distance seeing there is a lack of data on the other side of the boundary (i.e., no data in Burnaby). The concern is that this can exaggerate the nearest neighbor resulting in an overestimation of distance between neighbors (Levine, 2002).

Two types of edge corrections, rectangular and circular, can be utilized in accounting for the edge effect on nearest neighbors. Both methods calculate an average nearest neighbor distance and compare this to the theoretical average distance.
under random conditions (Levine, 2002). The concern is that neither method is a perfect measurement, but is an averaged corrected method.

Another concern with boundary problems is roadways. Roads tend to dictate where certain crimes are likely or not likely to happen; however, census data often contain boundaries in the middle of the roadways. For example, one house might have been broken into on the north side of the street, whereas another may have occurred the same day on the other side of the street. When comparing the neighbors by census tracts, these houses could be located in two different tracts, which have a potential of affecting the resulting data. This can occur with any form of boundary that currently exists, such as enumeration areas, census tracts, police beats, etc. Though several measures have been found to help in overcoming the modifiable areal unit boundary problem, it has not yet been fully resolved in statistical analysis of areal-based data.

The final area of discussion surrounding issues in the implementation of GIS in crime analysis focuses on the practicality of crime mapping in police departments. These issues are discussed pertaining to crime mapping by applicability of the software, police infrastructure, database management and communication.

In regard to the applicability of crime mapping in police departments, it has been argued that theoretically it appears as a positive contribution to policing, but is viewed as not being practical. Crime mapping is a process of extracting address information, geocoding it into a crime mapping program and then conducting analysis on the data through the use of point and polygon shape files. It is extremely beneficial as it provides the ability to conduct visual analysis on all crime data through the highlighting of crime hot spots. In addition, statistical analysis can be conducted to inform officers and policy makers as to reasons for crime concentrations at certain locations.
The concern raised by Ross Swope (2001) is that police agencies typically use crime mapping software only for descriptive levels of analysis. Descriptive analysis demonstrates visually where hot spots may occur. They do not determine why they occur; therefore, it would appear that police departments are likely already familiar with locations where certain crimes are concentrated. The issue that is most relevant is not specifically where crimes are concentrated but the reason as to why they are concentrated in that particular location. Swope found that the component that was missing in the analytic side of crime mapping was the theoretical support for analysis.

Without theoretical support, crime analysis is considered just cartography. Theoretical support is vital to the creation of a good explanatory model. In order to create a practical application of mapping for police agencies three areas are essential for analysis: theoretical foundation supporting social and environmental factors of the crime; descriptive crime mapping for visualization of hot spots and patterns; and finally, statistical analysis such as spatial regression modelling to create an understanding of why these trends or patterns are occurring.

It is of no surprise that researchers such as Manning (2001) argue that this technology has very little effectiveness in policing when the tools are used as a descriptor rather than an analytic aid supported by theory. Future implementation of crime mapping must incorporate theory and model building in addition to cartography if police effectiveness is to improve.

The second issue regarding the implementation of GIS is directly associated with police department infrastructure. A study conducted by Manning (2001) found that in police departments the technical support staff are typically overworked and are inadequately trained in the maintenance of such complex electronic infrastructure. He
also found that the lack of evaluation and supervision in the different levels of policing lead officers to have little interest in mapping (Manning, 2001). This is unfortunate, as the only way crime mapping will ever be an effective tool is if it is properly maintained and accepted within the police agencies.

This leads to the third issue of database management. Without the proper resources it is not possible to keep crime mapping consistent. Without consistency in the maps, data and file management analysis will be virtually impossible. Manning found that the lack of a centralized database created a crime mapping department with no link to any other departments such as prevention or community policing (2001). He found that the various departments are physically and functionally isolated creating a lack of information transfer between departments (Manning, 2001).

Decentralization is a key, not only in crime mapping, but also as it applies to the entire discipline of criminology. Database decentralization is problematic as it causes a cessation of information sharing between vital departments. This leads into the final area to address the communication of information. Crime mapping is becoming more vital to policing; however, to be effective, police officers must be able to access the information through a) a properly managed centralized database, b) a system that is accessible to other departments within the police agency, and c) a method of accuracy is necessary in communicating this information. Complete accessibility to this information for all those in need either on foot or through on-board patrol car systems are mandatory. Without meeting these requirements, crime mapping simply becomes a tool that would be used by only a limited number of investigators, therefore failing to provide street officers with extremely beneficial information.
These are only a few issues with the current state of crime mapping implementation in criminology. Despite these negatives, it is apparent from this study that crime mapping, if conducted properly, based on theory, descriptive and statistical analysis can and does provide a substantial amount of information that would not be obvious to researchers without the use of such tools. The issue is not whether to implement crime mapping tools in policing agencies, but rather what is the best approach to implementing them, in order to achieve the most beneficial results in policing.

**Concluding Remarks**

The objective of this study was to conduct an exploratory spatial data analysis of missing persons in Vancouver 1996. This was conducted in order to demonstrate the value of spatial statistics as an effective tool in the analysis of criminological data. Specifically, global and local measures of analysis were conducted on missing and found persons point data in Vancouver. The results illustrate missing persons as containing slight clustering within the data, with dispersed clustering throughout the suburbs of Vancouver; whereas, the found persons dataset illustrated strong clustering within the downtown eastside of Vancouver. These results indicate that spatial factors do have an influence on the patterning of missing and found persons in Vancouver, highlighting the importance of exploring the spatial components of missing and found persons further.

The criticism discussed pertaining to the need to enhance the analytic power of crime mapping in police agencies is supported in the results of this study. Cartographically, the data visually depicts clusters and possible patterns; however, statistical analysis uncovered clustering that existed globally and locally within the data.
It also demonstrated that reliance on one form of data might not be accurate. Depending on methods, boundary issues and standardization of data, the results can be slightly altered. This highlights the value of statistical exploration when conducting spatial analysis of events.

It was determined that spatial analysis is important to explore patterning in missing and found persons data. Future research must be conducted in order to create spatial regression modelling in order to account for the explanatory modelling of missing and found persons calls for service in Vancouver. Exploratory spatial data analysis is only the first step in the crime mapping process. If crime mapping is to become more widely accepted within police agencies, analysis must not only begin with exploratory measures, but also enhance the methods with explanatory models in order to fully understand criminal events.
REFERENCES


MathSoft, Inc. (2000). *S+ spatial stats version 1.5 supplement*. Data Analysis Products Division, MathSoft, Seattle, WA.


Ms. Nikki Thompson  
Graduate Student  
School of Criminology  
Simon Fraser University  

Dear Ms. Thompson:  

Re: Exploratory Spatial Data Analysis of Missing and Found Persons in Vancouver  

The above-titled ethics application has been granted approval by the Simon Fraser Research Ethics Board, in accordance with Policy R 20.01, "Ethics Review of Research Involving Human Subjects".  

Sincerely,  

Dr. Hal Weinberg, Director  
Office of Research Ethics
Appendix B: Canada Census 1996 Descriptions

Census Variables

% Sampled 1996 – Explanation of Variable

(Research Data Library: Census Text, 1996)

Population:

100 % Census sample

Population includes:

The Population Universe of the 1996 Census includes the following groups:

- Canadian citizens (by birth or by naturalization) and landed immigrants with a usual place of residence in Canada;

- Canadian citizens (by birth or by naturalization) and landed immigrants who are abroad, either on a military base or attached to a diplomatic mission;

- Canadian citizens (by birth or by naturalization) and landed immigrants at sea or in port aboard merchant vessels under Canadian registry;

- Persons in Canada claiming refugee status and members of their families living with them;

- Persons in Canada who hold student authorizations (student visas or student permits) and members of their families living with them;

- Persons in Canada who hold employment authorizations (or work permits) and members of their family living with them;

- Persons in Canada who hold Minister's permits (including extensions) and members of their family living with them.

For census purposes, the last four groups in this list are referred to as "non-permanent residents". For further information, refer to the variable Immigration: Non-permanent Resident.
The Population Universe of the 1996 Census does not include foreign residents because they were not enumerated in 1996. Foreign residents are persons who belong to the following groups:

- Government representatives of another country attached to the embassy, high commission or other diplomatic body of that country in Canada, and members of their families living with them;
- Members of the Armed Forces of another country who are stationed in Canada, and members of their families living with them;
- Residents of another country visiting Canada temporarily (for example, a foreign visitor on vacation or on business with or without a visitor’s permit). (Research Data Library, 1996, 2-23)

Gender:

- 100% Census Sample
- Gender: Refers to the gender of the respondent
- Survey Categories Include: Male; Female

(Authority Data Library, 1996, 2-23)

Age:

- 100% Census Sample
- Age: Refers to the age at last birthday (as of the census reference date, May 14, 1996). This variable is derived from date of birth.
- Survey: Ages 0-121

(Authority Data Library, 1996, 2-23)

Education:

- 20% Census Sample
- Education: Refers to the highest degree, certificate or diploma obtained.
- Survey Categories Include: No degree, certificate or diploma; Secondary (high) school graduation certificate or equivalent; Trades certificate or diploma; Other non-university certificate or diploma; University certificate or diploma below bachelor level; Bachelor’s degree(s); University certificate or diploma above bachelor level; Degree in medicine, dentistry, veterinary medicine or optometry; aster’s degree; Earned doctorate (Research Data Library, 1996, 2-23)

Income:

- 20% Census Sample
- Income: Refers to the total money income received from the following sources during calendar year 1995 by persons 15 years of age and over:
  - wages and salaries (total);
  - net farm income;
  - net non-farm income from unincorporated business and/or professional practice;
- federal Child Tax benefits;
- Old Age Security pension and Guaranteed Income Supplement;
- benefits from Canada or Quebec Pension Plan;
- benefits from Unemployment Insurance;

- other income from government sources;

- dividends, interest on bonds, deposits and savings certificates, and other investment income;
- retirement pensions, superannuation and annuities, including those from RRSPs and RRIFs;
- other money income.

Survey Response: Positive or negative dollar value or nil.

(Research Data Library, 1996, 2-23)

Unemployment Rate:

20% Census Sample

Unemployment Rate: Refers to the unemployed labour force expressed as a percentage of the total labour force in the week (Sunday to Saturday) prior to Census Day. Data are available for persons 15 years of age and over, excluding institutional residents. (Research Data Library, 1996, 2-23)

Rate Calculated: \[ \frac{\text{Unemployed labour force}}{\text{Total labour force}} \times 100 \]
Appendix C: Missing Persons Data Steps

N=2627 (Original 2677)

1. Cut the themes of Vancouver by municipal boundaries shape file. In this process you can cut the road network.
2. Extract data from CAD files. The Master data file contains the GEO # described below, along with all information pertaining to the case (i.e., date, time, code etc.)
3. Clean data by attaching a code called GEO to each case.
4. Extract only addresses and GEO # for the location tables .dbf
5. This new original file was then named Combined 96-01 renamed to Standardized 96-01-12 Test dbf and excel.
6. Add DBF table 'standardize 96-01-12 Test 01.dbf to the ArcView tables.
7. Open your DMTl attribute table and start the Geocoding Process with Standardizing the original DMTl Road network so that the address table you are using is compatible with the DMTl addresses (steps below).

Standardizing Addresses

- Click on scripts (this button is located in the dialogue box on the left)
- Click on new
- Load Text File (click on file folder at top to do this), the text file that you want to load is stanshape.ave which I can give you a copy of or you can get it from the esri.com website.
- Click on the Compile
- Open your Road file – Click on the Street tab so that the tab is highlighted
- Click on Running Man and save file – rename so you know it is your standardized address file
- A box will pop up asking for an address standardizer – click on unique ID
- A choose address box will pop up asking to choose the address field – this should be street, click ok
- Now, on your new table sort by street name so it is in order, Double check to make sure the values turned out the same

Calculating values:

- Now you need to calculate values for this table the first thing you need to do is to Edit the table. You can do this by clicking on the attributes table and then click on Table then click on Start Editing
- Then you need to add a new field, click on Edit and then add field
- Name the new field newstreet, click on string value and put 60 characters
- Select the newstreet field and use the Field Calculator (click on the picture of the calculator) to make the values in newstreet exactly the same as those in streetname.
• Sort the table based on the **streetname** field so that all the numbered streets are at the top of the table.

• **Manually select** all the numbered streets (by **holding down** the shift key and selecting with the mouse).

• Again use the **Field Calculator**, but this time calculate the **newstreet** field to = the values in you joined street name column (the one that contains the 1st, 33rd, etc. I think it's usually called **street_nam**)

• **Save your edits** and stop editing the table. Remove the selection

**Mapping Data**

8. Highlight the Vancouver roads to make it active. Click on Theme then Preferences to change the geocoding preferences.

9. Change the predir to suffdir to correct the E/W direction problem.

10. Change the suffdir to none.

11. Street Type should equal Street Type.

12. Click on View, then geocode address.

13. Reference Theme is the attribute table of the road network that was clipped and standardized.

14. The address style used was US Streets.

15. Address Table was the Standardize 96-01-12 Test 01.

16. Your Address Field is then the new address field that was standardized (i.e., NewStreet). This was picked from the scroll down menu.

17. The first Batch Match produced 2589 records out of 2677 matched. Meaning 88 records did not match.

   *(Shape File) Missing Match 1 N=2589
   *(DBF File) Standardize Match #1*

18. Interactive Re-Match was then used to determine why the records did not match.

19. Interactive Re-Match was used for address that were within the same 100 block. Addresses were changed to the next highest address in the interactive rematch providing it was within the same 100 blocks. Even and Odd addresses were kept consistent unless it did not exist in the database. In this case, the next highest addresses in the same block were chosen.

20. The addresses that were not found through the interactive re-match, were then compared to the actual DMTI road network by opening up the Vancouver road file (that I previously cut and made a theme). Once opened, the remaining addresses were compared manually to the file to see if there was a road network that existed. In this case, 50 addresses were deemed NOT geocodable, and therefore, were cut and pasted into an excel document for future reference if needed. However, they were deemed insufficient and discarded from the sample size.

   *(Excel File) Not Mappable Standardized N=50*

21. The other 38 addresses were changed as mentioned in step #19. This file was then renamed:

   *(Shape File) Missing Match #2 N=38*
22. These two databases were then combined to create a total so that they could be mapped together. This table includes a total N=2627:

**Standardize 96-01-12 Test**

This is the end of the Geocoding Process for Missing Persons, the next step is to repeat the steps for the found persons data spatial analysis.
Appendix D: Found Persons Data Steps

Found Persons N= 625 (Original N=637):

1. Cut the themes of Vancouver by municipal boundaries shape file, in this process you can cut the road network.
2. Extract data from CAD files. The Master data file contains the GEO # described below, along with all information pertaining to the case (i.e., date, time, code etc.)
3. Clean data by attaching a code called GEO to each case.
4. Extract only addresses and GEO # for the location tables .dbf
5. Change all # Ave to #-Ave in order to delete the 'th' after the # (find/replace every # from 1 to 10)
6. Change all intersections to include ' W 4-Ave & Cambie St'
7. Deleted any 'ND' and 'RD' and 'TH' and 'ST' (for 1st)
8. Change under geocoding predir to suffdir to correct the E/W direction problem
9. Change street type to suff type

Geocoding Data Process:

1. Batch Match #1 after cleaning the data ended up with a result of N=514 Matched Cases and N=123 Unmatched
   
   **FP Match 1 (N=514)**

2. The second stage to was to Sum FP Match 1 to create a count in which the data could depict the amount of calls at each location.
   
   Sum FP Match 1- Process discussed in next section (census data)

3. Cleaned data again making sure that spelling of the 123 unmatched addresses was corrected, interactive rematch was used for address that were within the same 100 block. Addresses were changed to the next highest address in the interactive rematch providing it was within the same 100 blocks. After matching again, their were N = 40 more that properly matched, N = 83 No Match.

   **FP Match 2 - N= 40**

4. The remaining addresses are then stored for future reference and the number of found persons cases is N=625.
Appendix E: Census Data Steps

1. Add Census Tract (CT) or Enumeration Area (EA) boundaries as a theme and cut the theme by the Vancouver Municipal Boundaries Shape.
2. To cut them you will need to use the cut function, through the geoprocessing wizard. You can do this by clicking on File, then extensions, and then scroll down to find geoprocessing wizard. Then click on View, then Geoprocessing Wizard, then clip one them based on another. Your Input theme is your CT’s or EA’s, and your polygon them is your municipal boundary shape file.
3. To clean your census data before adding it to ArcView, open census tables in excel.
4. Sort the records by their CMACT name i.e., Vancouver would have 102 CMAT CT’s zones. The concern with this is that there is only 92 actual CT’s and 10 are duplicates (because of mapping issues) therefore, you must dissolve the 10 duplicates (through a dissolve feature discussed later) in order to match actual CT data by their corresponding CMAT tract.
5. All census variables for Vancouver were copied from the CT file. The list of variables include:

<table>
<thead>
<tr>
<th>Aboriginal</th>
<th>Immigration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Income 1</td>
</tr>
<tr>
<td>Dwelling</td>
<td>Income 2</td>
</tr>
<tr>
<td>Dwelling 2</td>
<td>Labour 1</td>
</tr>
<tr>
<td>Education</td>
<td>Labour 2</td>
</tr>
<tr>
<td>Ethnicity 1</td>
<td>Land 1</td>
</tr>
<tr>
<td>Ethnicity 2</td>
<td>Land 2</td>
</tr>
<tr>
<td>Family 1</td>
<td>Mobility</td>
</tr>
<tr>
<td>Family 2</td>
<td>Plac</td>
</tr>
<tr>
<td>Housing</td>
<td></td>
</tr>
</tbody>
</table>

6. Once variables are downloaded and saved in excel as dbf tables, they must be cleaned according to the CMAT. Make sure there are the appropriate number of tracts per data and no duplicates exist.
7. To make it easier in ArcView, I then combined all the census variables that were of interest to my study into one dbf table and added it to the ArcView tables. Rather than having to add every census table created.
8. Next step is to enter the tables into ArcView but first, the census tract file (in ArcView) must have the 10 duplicate tracts dissolved.
Dissolving Census Tracts

Why? The reason is that multiple CT's and EA's exist (the same number is allocated for two geographical locations, but are still considered the same CT or EA), you must blend them to equal on CT or EA, otherwise your data will not join (i.e., Vancouver has 102 CT's but once dissolved only 92 actual CT's exist).

1. To dissolve census tracts, first make sure that you have loaded the extension geoprocessing wizard. You can do this by clicking on File, then extensions, and then scroll down to find geoprocessing wizard.
2. Then click on view, then geoprocessing wizard, then dissolve themes based on an attribute.
3. Click next, then the in the box theme to dissolve click on your census tract theme.
4. In the next box attribute to dissolve click on CMACT (this is your number that is associated with each census tract).
5. Save this in an output file. The theme is then entered into your table (in my case CT Vancouver) containing the dissolved CT tracts. In my case, there is now 92 tracts instead of 102.

Renamed CT tract table is:

Dissolved Tracts n=92.shp

Spatial Join of CT or EA and Addresses

6. The next step is to add CT or EA numbers to your geocoded addresses. To do this, click on view, geoprocessing wizard.
7. Click on Assign data by location (spatial Join). Then the theme to assign data TO is you geocoded address shape file. In my case, the Missing Persons n=2627 shape file.
8. The theme to specify that you want the data assigned FROM is the dissolve CT file that you created (i.e., in my case, dissolved tracts n=92). Click finish. It will look as though nothing has occurred, but open your table of your geocoded addresses and you will notice that each address has now been assigned a CT or EA number.
9. The next step is summarizing your point data so that you can combine your point data into a polygon theme. As well as summarize it by location. The next step will deal with summarizing your data by CT or EA so that it can be created into polygon data.

Summarizing Data by CT or EA

10. The next step in summarizing is to open your shape file that has the original geocoded addresses in it (in my case, the Missing persons data with n=2627 locations mapped). Once this table is open, click on the column label that states the CT or EA code (For CT it is CMACT, and EA is PRFEDEA). Only the top label i.e., CMACT should be highlighted.
11. Next, click on the sum button (Σ). In the box that pops up, the field should remain as 'SHAPE' and the summarized by should be 'MERGE'. Click on add to add this to the empty panel. **BEFORE** clicking ok, make sure to save your file as something you will remember (i.e., MP_summed_CT).

12. This will then add a new shape to your view. You can then open the table and will notice that there are no addresses rather your data is now summarized by the CT or EA number. After this you will want to join this new shape file to your dissolved CT or EA shape file. If you wish to do this, then skip to the Joining the Tables section step #16. Or, you can first summarize your addresses by location. To do this, follow then next steps below.

**Summarizing Data by Address**

13. The next step in summarizing is to open your shape file that has the original geocoded addresses again (in my case, the Missing persons data with n=2627 locations mapped). Once this table is open, click on the column label that states the address field (this is the column that you used to map the original addresses from) Only the top label i.e., Address should be highlighted.

14. Next, click on the summarize button (Σ). In the box that pops up, the field should remain as 'SHAPE' and the summarize by should be 'MERGE'. Click on add to add this to the empty panel. **BEFORE** clicking ok, make sure to save your file as something you will remember (i.e., MP_summed_Address).

15. This will then add a new shape to your view. After this, you then want to join your address table and your CT table (from the prior steps) to your CT dissolved tracts table.

**Joining the CT Tables**

16. The next step would be to join the CT tract dissolved table with the census data table.

17. **HINT:** to join tables, save and close all tables, then re-open the CT dissolved and shape file you created above for the CT or EA summed data (i.e., MP_summed_CT) then join. You can only do a spatial join on an **ORIGINAL** table **TWICE**. This is important, if you are trying to join a table that already has been joined more than once, it will not work. If this occurs, you can convert (under theme option) your shape file into a shape file (this is making a duplicate of your table, only the data is then considered a part of the shape file rather than just joined to your table. Then you can use the new shape file to attach to your dbf census data. For the most appropriate method, follow the next few steps.

18. Create a new shape file of the dissolved tracts. Make sure that your dissolved tracts theme is active. Click on Theme, then click convert to shape file, then name this so that you will know that this new shape file is going to be of your dissolved CT or EA's combined with your point (address) data (i.e., MP by CT). Click ok, and it will give you a new theme, i.e., MP by CT and this is now the table that we are going to use to join.

19. Open the table from this new theme i.e., MP by CT. Also, open your table that you created of the spatially joined CT or EA #'s with your geocoded data (i.e., MP_summed_CT). Click and highlight the label CMACT (for CT data) or PRFDEEA (for EA data) in the MP_summed_CT table. Then click on the label CMACT or PRFDEEA in the other table i.e., MP by CT (this is the new shape file
you just converted). Once you have the same label highlighted in both tables, click on the button above that does a spatial join (hint: it looks like two pages with an arrow going across it and it is white and blue).

20. Finally, you should notice that now you have assigned your point data to a polygon shape file (MP by CT). At this point, you can now display your aggregated point data visually by your CT or EA polygons.
APPENDIX F: Intensity Surface Plots

Missing Persons

Surface Plot of Missing Persons

Figure 31 Surface Plot of Missing Persons
Figure 32 Surface Plot of Found Persons