Geospatial Analysis of Environmental Health

Geotechnologies and the Environment

Volume 4

Series Editors:

Jay D. Gatrell, *School of Graduate Studies and Department of Geography, Geology, and Anthropology, Indiana State University, Terre Haute, IN, USA*
Ryan R. Jensen, *Department of Geography, Brigham Young University, Provo, UT, USA*

The “Geotechnologies and the Environment” series is intended to provide specialists in the geotechnologies and academics who utilize these technologies, with an opportunity to share novel approaches, present interesting (sometimes counter-intuitive) case studies, and most importantly to situate GIS, remote sensing, GPS, the internet, new technologies, and methodological advances in a real world context. In doing so, the books in the series will be inherently applied and reflect the rich variety of research performed by geographers and allied professionals.

Beyond the applied nature of many of the papers and individual contributions, the series interrogates the dynamic relationship between nature and society. For this reason, many contributors focus on human-environment interactions. The series are not limited to an interpretation of the environment as nature per se. Rather, the series “places” people and social forces in context and thus explore the many socio-spatial environments humans construct for themselves as they settle the landscape. Consequently, contributions will use geotechnologies to examine both urban and rural landscapes.

For further volumes:
http://www.springer.com/series/8088

Juliana A. Maantay · Sara McLafferty
Editors

Geospatial Analysis of Environmental Health

Springer

Editors
Prof. Juliana A. Maantay
Earth, Environmental, and Geospatial
Sciences Department
City University of New York
Lehman College
Bronx, NY 10468
USA
Juliana.maantay@lehman.cuny.edu

Prof. Sara McLafferty
Department of Geography
University of Illinois
Urbana, IL 61801
USA
smclaff@uiuc.edu

Cover artwork by Kristen Grady, Urban GISc Lab, Lehman College, City University of New York, adapted from Figure 2.19, showing a continuous surface created by a Geographically Weighted Regression (GWR) analysis.

ISBN 978-94-007-0328-5 e-ISBN 978-94-007-0329-2
DOI 10.1007/978-94-007-0329-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011921119

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Recently, two disciplines – geospatial analysis and environmental health began to interact with each other. Each of them emerged from a decade of growth in their knowledge base, numbers of engaged professionals, and validated methodologies. Out of these separate growths, a new interdisciplinary area of study, geospatial analysis and environmental health, has emerged and this book captures this moment. It is an area of practice as well as an area of knowledge, because many studies in this area lead to policy changes and actions directly associated with improved health outcomes for affected populations. Studies in this area directly contribute to accelerating the health transition in those areas and populations where society is disposed to go from knowledge to action when environmental causes of ill-health are identified

"Environment" in "environmental health" is broadly defined From physical elements such as water, air, toxic materials; to the built environment with its access to elements that promote well-being or that promote poor diets and little physical exercise; to societal decisions that promote health or more negatively increase health risks. All these elements have spatial patterns and populations differ in their interactions with them. Advances in methods of measuring, storing, and accessing such data have brought the concept of geospatial data into prominence and an increasing recognition that exposures to these elements by individuals or groups can be measured using geospatial methods. Along with developments in measuring human exposure to environmental factors comes the search for relationships, if any, between measures of environmental exposure and health outcomes. After relationships are established, questions arise about placing responsibilities for reducing the risks of exposures, particularly those that are experienced by vulnerable population groups.

This useful and timely book provides examples of the principles described above as well as descriptions of recently developed methods such as participatory mapping, spatial regression and distance decay methods that are often used in studies of the environment and health. This book brings together the kind of studies that are more often widely distributed among specialist journals that are difficul to fin and access. It will be valuable especially for students and the general public whose widespread interest in population health is coupled with their commitment to developing a more healthy environment for the future.

Iowa City, IA

Gerard Rushton

Contents

Part I General Considerations in Geospatial Analysis of Environmental Health

Part II Impacts on Environmental Health (Topical Case Studies)

Part III Geospatial Methods in Investigating Environmental Health

Chapter 2
Using Geovisualization and Geospatial Analysis to Explore Respiratory Disease and Environmental Health Justice in New York City

Andrew Maroko, Juliana A. Maantay, and Kristen Grady

Abstract The goal of this chapter is to illustrate how complex issues in environmental health justice analysis can benefit from geovisualization and exploration within a Geographic Information Science (GISc) framework. Individual health outcome variables, such as hospitalizations due to respiratory disease, can be very difficult to interpret without a geographic context; and interactions amongst variables such as disease, socio-demographic characteristics, or environmental exposures, further complicate an accurate interpretation of the data. Data exploration and visualization through mapping and spatial analysis often provides a more robust understanding of the data, as well as improved clarity in viewing the phenomena under study, which will lead to better design of further analyses and additional hypothesis generation, in an iterative fashion. In the first part of this chapter, we use a hypothetical data set to illustrate some of the data exploration, geovisualization, statistical methods, and geospatial analyses. In the second part of the chapter, we use a worked example of respiratory disease and socio-demographic variables in New York City to assess potential environmental justice impacts, in order to further demonstrate the importance of geovisualization and geospatial analysis in achieving a better understanding of environmental health issues.

Keywords Geovisualization · Geostatistics · New York City · Asthma · Environmental Justice · Environmental Health

2.1 Introduction

The goal of this chapter is to illustrate how complex issues in environmental health justice analysis can benefit from geovisualization and exploration within a Geographic Information Science (GISc) framework. Individual health outcome variables, such as hospitalizations due to respiratory disease, can be very difficult

---

A. Maroko (✉)
MPH Program, Department of Health Sciences, Lehman College, City University of New York, Bronx, NY, USA; Urban GISc Laboratory, Lehman College, Bronx, NY, USA
e-mail: andrew.maroko@lehman.cuny.edu

J.A. Maantay, S. McLafferty (eds.), *Geospatial Analysis of Environmental Health*,
Geotechnologies and the Environment 4, DOI 10.1007/978-94-007-0329-2_2,

to interpret without a geographic context; and interactions amongst variables such as disease, socio-demographic characteristics, or environmental exposures, further complicate an accurate interpretation of the data. Data exploration and visualization through mapping and spatial analysis often provides a more robust understanding of the data, as well as improved clarity in viewing the phenomena under study, which will lead to better design of further analyses and additional hypothesis generation, in an iterative fashion.

Geovisualization is defined as "the use of computer technology for exploring data in visual form.... and the use of computer graphics for acquiring a deeper understanding of data" (Visvalingam, 1994). Visualization can also be thought of as "the interplay between technology and the human mind" (Davies and Medyckyj – Scott, 1994). The impetus behind such a geovisualization process is to "see the unseen" in these increasingly large and complex datasets, where, without computational exploratory mapping, it is unlikely that we would be able to ferret out many of these "unseen" relationships (Orford, 2005).

Maps are, and always have been, rich sources of data. GISc increases the richness of the data, and the functions of GISc make it possible to look at the data in many different ways and from various viewpoints. We can manipulate the data, examine its statistics, plot graphs of it, classify and reclassify it with different schemes of class breaks and classification methods, and look at multiple views of the data at the same time. This kind of geovisualization has become easier and more productive since the advent of accessible forms of GISc software and their computerized cartographic capabilities (Kraak and Orneling, 1996; MacEachern and Kraak, 1997).

In the first part of this chapter, we use a hypothetical data set to illustrate some of the data exploration, geovisualization, statistical methods, and geospatial analyses. In the second part of the chapter, we use a worked example of respiratory disease and socio-demographic variables in New York City to assess potential environmental justice impacts, in order to further demonstrate the importance of geovisualization and geospatial analysis in achieving a better understanding of environmental health issues.

2.2 Environmental Health Justice

Environmental Justice (EJ), as a research framework, is the attempt to document and address the disproportionate environmental and health burdens borne by the poor, people of color, and other vulnerable populations. In a broader context, EJ theory encompasses everything that is unsustainable about the world we have created, including rampant population growth, industrialization, pollution, consumption patterns, energy use, food production, and resource depletion. "The EJ movement has sought to redefine environmentalism as much more integrated with the social needs of human populations, and, in contrast with the more eco-centric environmental movement, its fundamental goals include challenging the capitalist growth economy, as well" (Pellow and Brulle, 2005, 3).

Environmental Justice, both as an advocacy movement and as a field of research, came into being over 20 years ago, and ever since that time, Geographic Information

Science has been used to examine the spatial realities of environmental injustice (Boer et al., 1997; Bowen et al., 1995; Burke, 1993; Chakraborty and Armstrong, 1997; Chakraborty et al., 1999; Maantay et al., 1997; Maantay, 2002; Morello-Frosch et al., 2001; Neumann et al., 1998; Perlin et al., 1995; Pollock and Vittas, 1995; Sheppard et al., 1999).

GISc methods have been used in environmental justice research primarily to analyze the spatial relationships between sources of pollution burdens and the socio-demographic characteristics of potentially affected populations, and for the most part, researchers have found strong associations between race, class, and environmental burdens. More recently, health outcomes and exposure measures have also been included in order to draw more definite connections between pollution, poor people, communities of color, other vulnerable populations, and adverse health outcomes. GISc technology is particularly well-suited for EJ research because it allows for the integration of multiple data sources (e.g., location of polluting facilities, population characteristics, and disease rates), representation of geographic data in map form, and the application of various spatial analytic techniques (e.g., buffering) for proximity analysis (Zandbergen and Chakraborty, 2006).

2.3 Data Exploration Example Using Hypothetical Data Set

GISc can be invaluable in data exploration. Looking at data spatially allows a much more holistic and complete view of the phenomena or processes under study. Spatial analysis requires data with a geographic identifier any data that has a locational component (e.g., a street address, latitude/longitude, zipcode, census tract) can be mapped and analyzed with GISc.

When starting the process of data analysis, it is traditional to run some basic descriptive statistics (e.g. mean, median, mode, standard deviation, etc.) on the pertinent datasets. Although these numbers can provide some extremely useful summaries of the data, they do not provide a spatial understanding – that is, they do not show how values vary from place to place. For instance, if you are presented with a dataset that contains thousands of samples (e.g., census tracts), each of which has information regarding the population and a disease of interest, an a-spatial analysis may not provide you with all the information that is required for a complete interpretation, vis-à-vis, how disease varies across geographic space.

To illustrate this point, a hypothetical study area was created containing a 30 × 30 grid of cells ($n = 900$), each of which could be considered the geographic unit of analysis (resolution) in a GISc study. Each cell is 1,000 by 1,000 ft, and contains a value for population data as well as disease data. A random population between 500 and 600 was given to each unit. Disease data, however, was not randomly distributed. Instead, "disease centers" were defined around two locations. Geographic units that were greater than 5,000 ft from either of the source points were assigned a random disease rate between 1 and 2 cases per 100 persons. As the distance from a cell to either of the centers decreases, the rates increase. In other words, the rates in the geographic units proximal to the disease centers were calculated as higher than distant geographic units.

If this data were to be looked at tabularly (Table 2.1) or with a graph (Fig. 2.1), relatively few inferences could be made regarding the nature of the phenomenon, how it is distributed in space, and any relationship between the areas and the rates. However, by mapping the data, spatial patterns can emerge, and provide more explanatory power in terms of the geographic context. The relative clarity of the spatial patterns depends on many things, such as the choice of data classification method (e.g. natural breaks, quantile, etc.) and the type of thematic map (e.g. dot density for number of cases, choropleth for rates, etc.), as discussed below.

Data Classification Process: When mapping quantitative information, it is usual to classify the numerical data into ranges of numbers. This is done for convenience and for ease of reading and interpreting the information on the map, since it is often impossible or impractical to represent every unique value in the data with a different unique symbol on the map.

Table 2.1 Tabular view of the first eleven records in the hypothetical dataset

Cell id	Population	Number of cases	Disease rate (×100)
1	571	10	1.7513
2	553	8	1.4467
3	558	9	1.6129
4	529	7	1.3233
5	530	7	1.3208
6	577	10	1.7331
7	501	5	0.9980
8	578	33	5.7093
9	576	10	1.7361
10	581	11	1.8933
11	571	10	1.7513
...	...	...	...

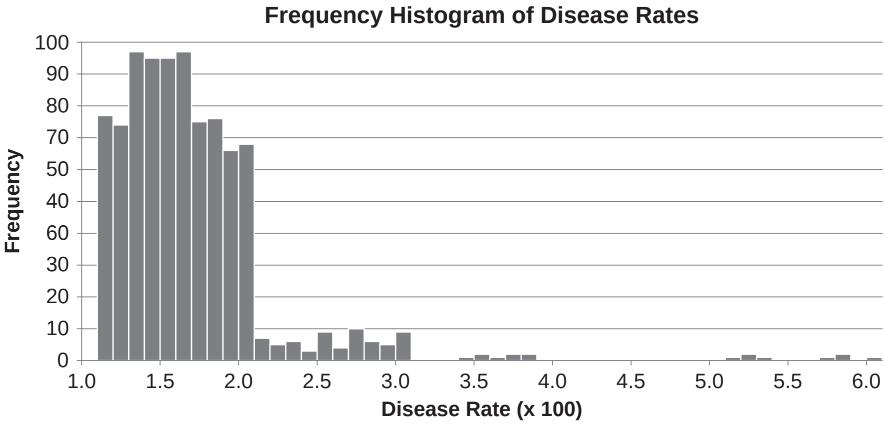

Fig. 2.1 Graph (histogram) of the hypothetical dataset

There are a number of data classification methods that are commonly used, and which one is selected will depend to a large degree on the dataset being mapped – its content, configuration and its "shape." This entails, for instance, whether or not the data has a normal distribution, and other particularities of the data set, which can be ascertained in general terms by running some basic descriptive statistics. The data classification method ultimately chosen will affect the emphasis of the mapped data, and some cartographers maintain that the "best" way to present quantitative data is by classifying it in several different ways, allowing the analyst to piece them together and form a composite picture of different aspects of the data.

The data classification methods briefly described here are the ones most generally available with standard GIS software packages – equal interval, quantile, natural breaks, and standard deviation – and these tend to cover most circumstances. There are other useful methods, such as arithmetic, geometric, harmonic, and nested mean, which can be used in a GIS software program by manually calculating and setting the class breaks.

Equal Interval: in some ways this is the easiest classification method to understand and use, and maps created with equal interval class breaks have an inherent logic and intuitive feel about them. With this method, the data range is broken into classes, each containing the same interval. For instance, if we were interested in mapping percent minority population by census tract for a city, and our data range was 1–100 (1 being the minimum value and 100 being the maximum value) and we wanted five classes, our classes would be 1–20; 21–40; 41–60; 61–80; 81–100. This method works well for a dataset that has a normal distribution of values. The drawback to Equal Interval comes if your dataset contains values that primarily fall in just one or two of the classes. The resulting map would show only one or two classes leaving other classes unrepresented. As such, little useful information about the spatial distribution of the variable would be visible.

Quantile: this method produces a map in which every class has an equal number of areal units or observations. Let us imagine that we have the same dataset as above with percent minority population (data range: 1–100) which are aggregated into 200 census tracts (i.e. $n = 200$). If we want five classes again, then each class will have 40 census tracts in it, and that will determine the class breaks. If the data values are arrayed in order of magnitude, the class breaks would be drawn so that each class includes a set of 40 census tracts, consecutively based on the data values. Although this method produces a map that can be more visually interesting than that produced by equal interval (because by definition each class will have some units in it) it can also be misleading to the map viewer, since the map will show each data class with equal weight (the same number of areal units per class) even if the values are not that different between classes. This method tends to work best when data is aggregated by areal units that are roughly the same size. Additionally, outliers may be de-emphasized in the quantile method, due to the grouping of values by ordinal ranking.

Natural Breaks: very often, this is the default classification method in GIS software. The assignment of class breaks in this method is very dependent on the dataset you are working with. It uses an algorithm to create classes which are as homogeneous as possible internally, while maximizing differences amongst the classes. It does this by organizing an array of the data, and then finding the "natural" break points or discontinuities in the array, thus combining the values that are similar into classes. Very often, this is the most realistic or "true" view of the data, since each class is internally consistent. The drawback is that map viewers may have a more difficult time interpreting maps made with natural breaks, since the class ranges in the map legend may appear random and arbitrary, and some classes may be overly inclusive, and other classes may contain very few data values.

Standard Deviation: The standard deviation method of classification groups the data values into classes based on the mean and the standard deviation of the data set. Each class represents one (or one half, one third, etc.) standard deviation above or below the mean (the arithmetic average) of the dataset. The standard deviation classification can result in classes containing class break values outside the actual range of the data, due to the way standard deviations are constructed. The standard deviation classification method is especially useful when performing longitudinal studies (comparing different time periods) or for comparisons amongst datasets that vary widely in their mean, median, or other measurement of central tendency. In standard deviation classification the mean and the standard deviation are the basis of the class boundary formation, and so the mean and standard deviation are more comparable across datasets with varying ranges.

The process of data classification is itself a type of data exploration. It allows the analysts to familiarize themselves with the data, and see if and how the map changes based on the classification method used. This can reveal information about the data and affect the ease of data interpretation. Maps made with the exact same data but using different classification methods will, more often than not, look markedly different. In addition to the classification method selected, factors such as how many classes are created and how many areal units are included in the data set, will also have an impact on the map's appearance and its interpretation. Below is a graphic representation of how different classification methods would group the data from our 900 sample hypothetical dataset of disease rates differently in terms of class breaks (Fig. 2.2) as well as number of samples per class (frequency, Fig. 2.3).

Types of Thematic Maps: The types of maps usually used for data exploration and geovisualization are, broadly speaking, thematic maps, and they can be either quantitative or qualitative in nature. Thematic maps express a "theme" of information, as opposed to a reference map, which is used for way-finding and identifying actual locations. Although it is common for a thematic map to only represent one "theme" or variable, effective maps can also be made to show multiple variables, often by employing several types of thematic maps in one. The following types of thematic maps are used most frequently for data exploration and visualization:

- Dot density
- Choropleth
- Proportional symbol
- Isoline
- Continuous surface (interpolation of point data)
- Cartogram

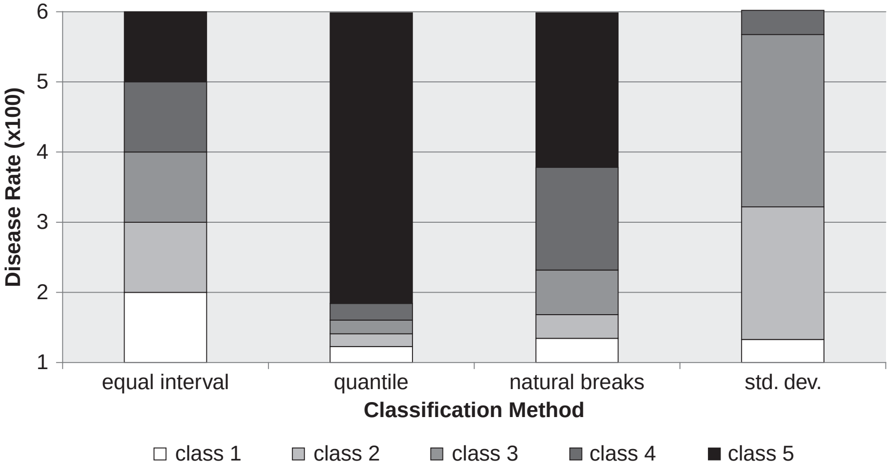

Fig. 2.2 Class breaks of disease rates using four different classificatio methods (hypothetical data)

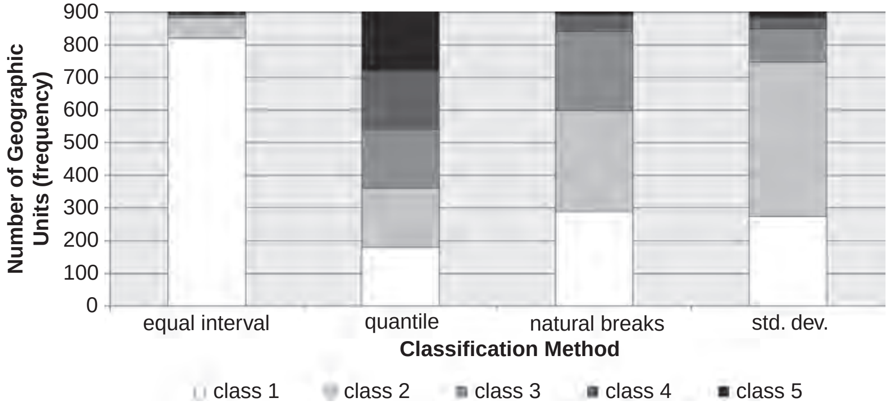

Fig. 2.3 Frequency of geographic units per class using four classificatio methods (hypothetical data)

Dot Density Maps: A dot density map is used to plot the absolute numbers of things – people, cases, or events – as aggregated by a geographic area. Usually, each dot represents a certain number of the things being mapped, e.g., one dot equals 10 cases of tuberculosis. The dots are not intended to correspond to the actual locations of these things, but rather are a random distribution of the points within each area of aggregation. Dot density maps are very useful for obtaining a "snapshot" of the distribution of the variable within the larger geography, and are often used for data exploration with the dots "on top" of a choropleth map showing rates or percentages of another variable. This is an easy way to investigate the potential spatial correspondence of two or more variables, or different aspects of the same variable. Dot maps have been in use for hundreds of years, for instance, John Snow's dot map of cholera deaths in Soho in nineteenth century London (Fig. 2.4), which is a modified use of a dot map, where one dot equals one event at its actual location.

In their simplest form, one dot equals one case. Figure 2.5 depicts how altering the dot value changes our perception and possible interpretation of the data, and reveals why it is important to select the dot value wisely. As with many cartographic decisions, there is no necessarily "right" choice, but through a process of

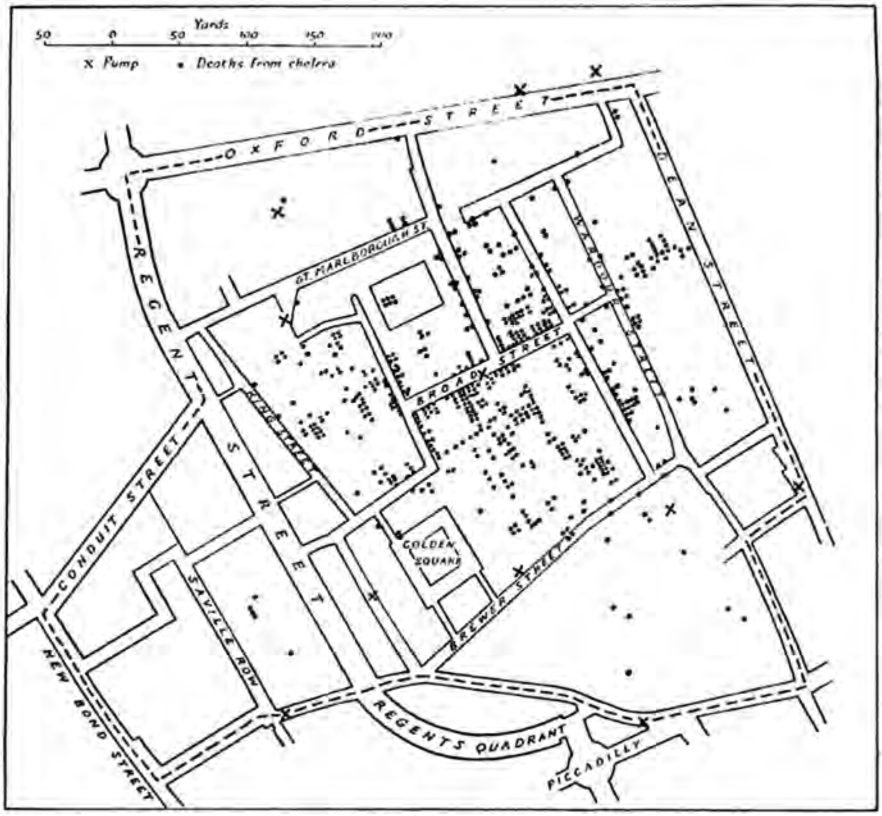

Fig. 2.4 By plotting deaths from cholera on to a map and revealing the geographical relationship between deaths and the location of the Broad Street pump, Dr. Snow had showed how maps could provide a unique insight into the patterns, processes, and relationships of spatial phenomena. The relationship between polluted drinking water and cholera was not self-evident and had to be graphically displayed before the connection could be made. (Orford, 2005:190–191). This seminal thematic dot map by Dr. John Snow of deaths from Cholera in 19th century London, as reproduced in E.W. Gilbert's 1958 article, is often thought to be the inspiration for the discovery of the waterborne nature of the disease's transmission (although this commonly-held idea of the map's role in this discovery is disputed as apocryphal by several authors, notably Tom Koch in *Cartographies of Disease*, and Orford in Visualization and Cartography). Map Source: "Pioneer Maps of Health and Disease in England," *Geographical Journal*, 124 (1958), 172–183

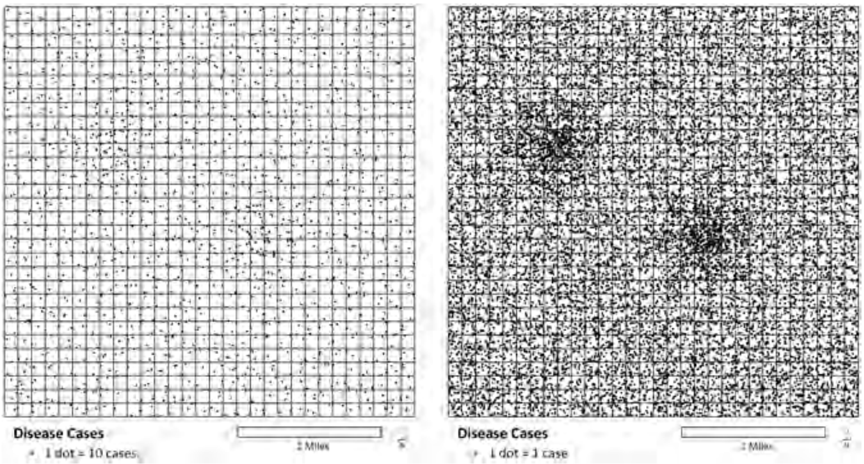

Fig. 2.5 Dot density maps of hypothetical data showing how dot value selection can affect ease of visual interpretation (*left*) shows 1 *dot* = 50 cases (*right*) shows 1 *dot* = 5 cases

"intelligent" trial and error, one can find the optimal way to present the data graphically in order to show the greatest detail and meaning in the data, and to distinguish any spatial pattern.

Choropleth maps: Choropleth maps, also called graduated color maps, are used to show rates, percentages, or ratios of a phenomenon, as aggregated by some geographical unit, such as a census tract, zip code, county, or state. In a choropleth map, each unit receives a color, pattern, or tone that designates the value range of the variable. These colors or tones are graduated in hue or intensity to denote the relative magnitude of the variable. Intuitively, the progression of shades makes sense: as the rate or percentage increases, the color deepens. It is easy to compare two choropleth maps showing different variables in order to ascertain geographic distributions and any spatial correspondence amongst the mapped variables, provided the maps have the same extent and unit of analysis. Using the natural breaks classification method on the hypothetical dataset, a choropleth map was created that renders the areas with unusually high disease rates easily identifiable (Fig. 2.6).

Proportional symbol: A proportional symbol map, also called a graduated symbol map, shows relative or absolute amounts of the variable by using a symbol placed at the corresponding point on the map, or at the centroid of each unit of aggregation. Data can be utilized in the form of absolute numbers or counts, or in the form of percentages, ratios, or proportions. Proportional symbol maps can also depict ordinal, or ranked, data, such as small, medium and large, or high, medium, and low density. These maps are useful to obtain an overview of the variable, and are particularly informative when used in combination with other types of thematic maps. Using proportional symbols on the hypothetical data, areas of high disease rates are again easily visible (Fig. 2.7).

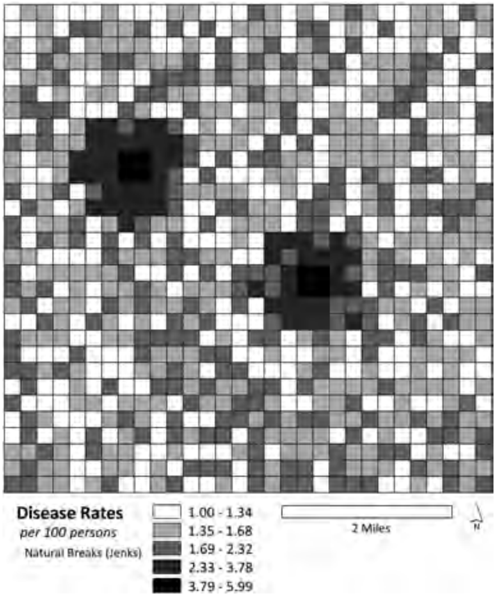

Fig. 2.6 Choropleth map of hypothetical data using "natural breaks" classificatio

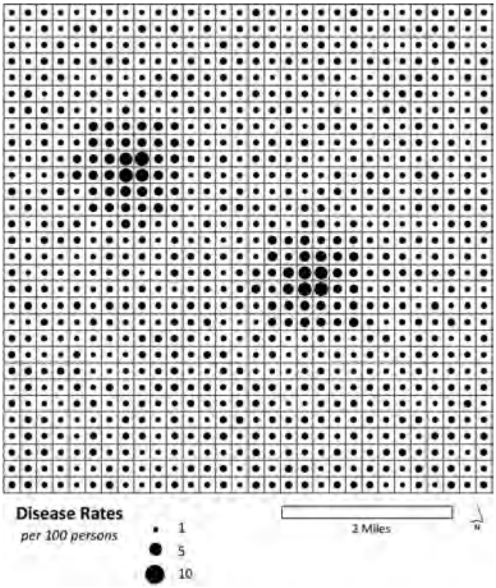

Fig. 2.7 Proportional symbol map of the hypothetical data

Continuous Surface: When data is presented as a continuous surface, spatial trends can often be intuitively and easily seen. A surface can be created by interpolating values between sampled points, essentially estimating values between the known samples, thereby "smoothing" the raw data. Although continuous surfaces are most often used to represent environmental data such as temperature, rainfall, air quality, and elevation, they can be used for nearly any type of dataset. However, caution must be used when interpolating sparse data (i.e. few sample points) or data that is inherently delineated by discrete boundaries (e.g. legal jurisdictions). There are many ways to achieve this interpolation, including inverse distance weighting, spline, and Kriging. Although an in depth discussion of each technique is beyond the scope of this chapter, the nature of the data being analyzed and the desired product will often dictate which methodology is most appropriate. In our hypothetical dataset, ordinary Kriging was used to estimate the values of points between the centroids of the original polygons. The output is in raster format (gridded) and can be viewed as either a true continuous surface where there are no classified divisions in the data (i.e. "stretched") or as a classified surface utilizing the same classification schemes discussed earlier (Fig. 2.8). Just as with other techniques, classification choices, as well as additional choices which must be made regarding interpolation, can profoundly affect the appearance and ease of interpretation of the product.

Isoline: Isoline maps, or contour maps, are made from a continuous variable, by connecting places of equal value of the variable. All points along any given line are presumed to have the same value for the particular variable being mapped. As isolines are essentially an alternate way to represent a continuous surface, discrete occurrences of a phenomenon (e.g. land use) should not be mapped using this method. Although topography is the most common variable used with isolines,

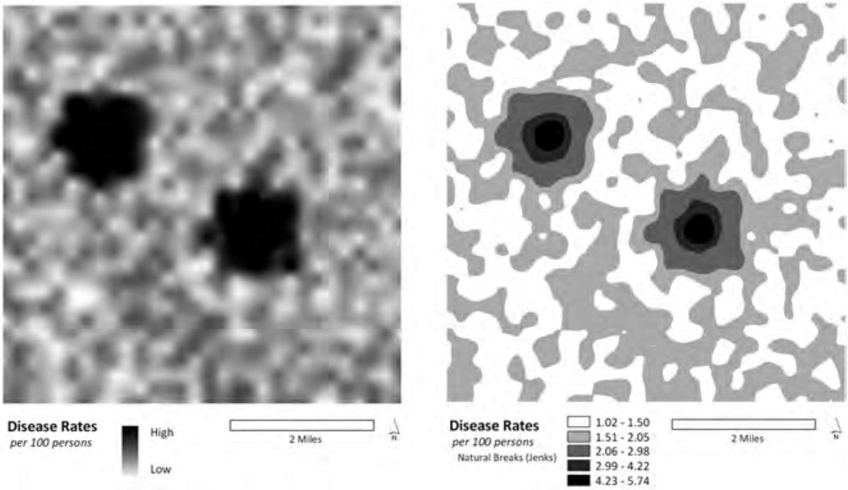

Fig. 2.8 Continuous surface of hypothetical dataset (*left*) shows a "stretched" surface and (*right*) shows a classified continuous surface

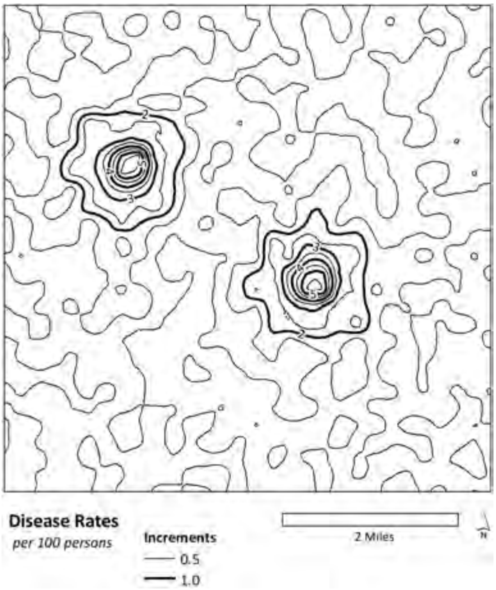

Fig. 2.9 Isolines of the hypothetical dataset derived from the continuous surface

any continuous data, such as most environmental variables as well as some human variables (e.g. median housing costs, population density, and disease rates) can be mapped in this fashion. For the hypothetical dataset, isolines were created based on the interpolated surface described above (Fig. 2.9).

Cartograms: Cartograms are not very commonly used in public health mapping, but deserve a wider exposure, as they can be quite effective in helping to visualize data. Cartograms have been used to good effect, for instance, in Danny Dorling's World Mapper cartogram series (Dorling, accessed 2010), which includes maps that informatively present complex and varied data in a "snapshot" way, and are especially useful for displaying data at a global extent. They have also been used in environmental health literature to illustrate the spatiality of environmental public health research (Sui and Holt, 2008; Houle et al., 2009). Cartograms can help in visually comparing regions and variables across regions at a glance, but are less useful for determining actual quantities. Therefore, they are primarily used as working or exploratory maps to indicate potential areas for more detailed study, and as presentation maps to inform and communicate to the public about different issues.

A cartogram, also called a "density-equalizing map," "equal-area map," "isodemographic map," and "value-by-area map," shows land areas sized to reflect the magnitude of the variable being mapped. Normally, the geographic units shown on maps reflect their real geographic size. Not cartograms – cartograms ignore true geographic size. In other quantitative thematic maps, data is mapped by symbolizing the variable's quantity and placing the symbol in or on the geographic units. In the cartogram, the size of the geographic unit itself is intended to communicate the variable's quantity. For instance, in a cartogram of world population by country, the geographic size of the countries would be drawn in proportion to their population size, not to their geographic size. China, India, the United States, and Indonesia, the world's four most populous countries, would therefore be drawn with the largest geographic extents. The size of the country itself on the map thereby represents the variable of population size. The cartogram does not lose data by classification or generalization, as do choropleth maps, for instance. However, the map user may find it difficult to understand the cartogram, depending on the map user's existing knowledge of and familiarity with the geography being portrayed. Many mapmakers choose to include a small inset map with the cartogram. The inset map can be used to remind the map user of the real relationship and physical sizes of the geographic units being mapped, so the information embedded in the altered sizes of the geographic units in the cartogram can be more easily interpreted. The geographic shapes on many cartograms can be highly generalized, often with box-like forms that make no attempt to conform to true shape, resulting in a map that looks more like an organizational chart (Maantay and Ziegler, 2006). Cartograms are often used in combination with other thematic mapping methods, such as choropleth or proportional symbol maps, which add richness and nuance to the data presentation.

The cartograms contained in this chapter are density-equalizing cartograms, in which areal units are drawn proportionally to the salient population characteristics. The benefit of this type of cartogram is to overcome the inevitability in non-cartogramatic maps for a variable to "show high incidence in cities and low incidence in rural areas, solely because more people live in cities," (Gastner and Newman, 2004, 7499). The maps below (Fig. 2.10) show the juxtaposition of the choropleth and cartogram of the hypothetical data disease rates. Notice how the two "disease centers" are not only brought out by the shading, but are also enlarged cartogramatically for a dramatic visual effect.

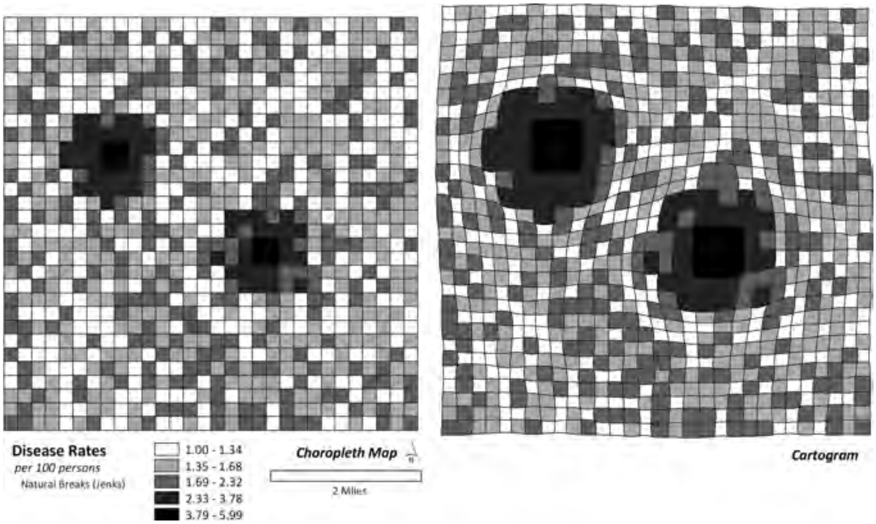

Fig. 2.10 Choropleth (*left*) and cartogram (*right*) of the hypothetical dataset

Geostatistics is a branch of statistics which focuses on the analysis of spatial data. As was demonstrated with the simple hypothetical example above, the locational dimension of a dataset is tremendously important in understanding environmental and health concerns. What follows is not meant to be an exhaustive discussion of geostatistics or geospatial analysis by any means, but rather a simple demonstration of how this type of quantitative exploration can help the researcher to better understand the nature of what is under investigation. This is done by working through only a few techniques: Moran's I and Local Indicators of Spatial Autocorrelation (LISA) are introduced below; and Geographically Weighted Regression is introduced in the next section. Many other measures and analyses could have been used (e.g. Getis-Ord G or Gi*, spatial autoregressive models, etc.), but that is beyond the scope of this introductory chapter. For more information see Chapters 17 and 20 (Chakraborty; Hu et al).

Spatial autocorrelation is an important concept in spatial analysis. If the value in one geographic unit is correlated with the values in neighboring units, a variable can be considered spatially autocorrelated (Cliff and Ord, 1973). Moran's I is a global measure for autocorrelation which ranges from –1 to 1. When values approach 0, there is no spatial autocorrelation; as the Moran's I approaches 1 or –1, there is positive (clustering) or negative (dispersion) autocorrelation, respectively. A standardized Z score can be used to assess significance (with the null hypothesis representing a random spatial distribution). In order to calculate a Moran's I, a spatial weights matrix must first be defined. This matrix defines the spatial relationship among the samples (e.g. census tracts). There are many options regarding the definition of the spatial weights matrix, the most common of which include polygon contiguity, simple distance threshold, distance decay, and k-nearest neighbors. For more information regarding spatial autocorrelation, see Chapter 17 by Chakraborty, in this book.

The Moran's I for our dataset suggests a statistically significant clustering of disease rates ($I = 0.62$ when first order contiguity is used, $Z\ Score = 26.4$). This clustering can be further explored using local indicators of spatial autocorrelation (LISA). A LISA can be used to quantify spatial autocorrelation locally by calculating a Moran's I and an associated significance level for each spatial unit. The sum of all of the LISAs will be proportional to the global measure of spatial autocorrelation (Anselin, 1995). When our dataset is looked at in this fashion, once again using first order contiguity, the areas with local clusters of high disease rates can be clearly seen ($p<0.01$) (Fig. 2.11). It is important to note that what is being seen in this map are the local clusters of similar or dissimilar rates (neighboring geographic units in this case), rather than the rates themselves.

These types of visualization and simple spatial analysis methods can lead to interesting findings and raise interesting new questions. For instance, in our hypothetical data, what may be causing these clusters of elevated disease rates? Is there a pollution source or other environmentally burdensome land use? Are there populations who are particularly vulnerable due to demographic, genetic, or social characteristics? These types of questions often can only be answered with a good knowledge of the study area's physical and social environments.

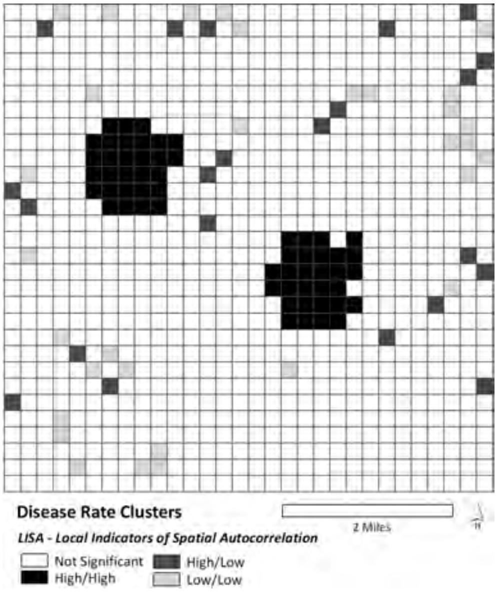

Fig. 2.11 Local Indicators of Spatial Autocorrelation (LISA) clusters of disease rates in the hypothetical dataset. High/high suggests local clusters of high disease rates (high values surrounded by neighbors of similarly high values). Low/low indicates local clusters of low disease rates. High/low and low/high suggest local statistical outliers (high values surrounded by low values or low values surrounded by high values, respectively)

When the number of variables or the complexity of the relationships increase, it can be very useful to explore the phenomena through more complex statistics rather than simply cartographically visualizing the individual variables. There are many ways to look at this data statistically, and each dataset may lend itself to one technique or another. A common approach when trying to quantify relationships among variables is regression analysis. Similar to the previously discussed exploratory techniques, regressions can be approached a-spatially or spatially. Ordinary least squares regression (OLS) results in summary statistics for the entire study area (i.e. global), however if the relationship(s) being examined are geographic in nature and have a truly spatial component, it can be beneficial to perform the regressions geographically (e.g. spatial regressions) or even locally (e.g. geographically weighted regression). The remainder of this chapter will go through a worked real-world example demonstrating the utility of data visualization and geostatistics.

2.4 Respiratory Disease and Environmental Health Justice in New York City

Respiratory disease rates can be a major concern in urban environments. One of the causes of high rates of respiratory disease may be poor air quality due to close proximity of residential areas to areas with high vehicular traffic, industrial land uses, high population densities, and other environmentally burdensome land uses (Aylin et al., 2001; Edwards et al., 1994; Maantay et al 2008; Smargiassi et al., 2009). Respiratory disease is an environmental justice concern that combines issues of socio and economic vulnerability with unequal environmental exposure. Some research has suggested that not only are lower-income populations and communities of color more likely to live in close proximity to environmentally burdensome facilities and thus be more exposed to pollution, but that the health effects of exposure to these burdens are further modified by socio-economic status, and "due to material deprivation and psychosocial stress [these populations] may be more susceptible to the health effects of air pollution," (O'Neill et al., 2003, 1861). Therefore, vulnerable populations, such as those with limited income or educational attainment, may suffer more from the same exposures when compared to other groups. What follows serves as an example of how geovisualization and exploratory spatial analysis can be used to examine respiratory disease vis-à-vis socio-demographic variables in New York City in order to assess potential environmental justice impacts.

Data: The data used in these analyses are all publically available information aggregated to the census tract, and includes socio-demographic information from the 2000 US. census and health information (hospital admissions) from New York Statewide Planning and Research Cooperative System (SPARCS) via Infoshare.org. The socio-demographic data consist of adults (> 25) who do not have a high school diploma, non-Hispanic black persons, Hispanic persons, and persons below the federal poverty level. Other variables were explored as well; however, due to concerns such as excessive colinearity, they were not included in these analyses. All of these variables were examined as rates (e.g. percentage of population that is below poverty thresholds, per census tract). As can be seen by examining the maps (choropleth and cartograms), the socio-demographic variables are not evenly distributed across NYC (Figs. 2.12 and 2.13).

Health outcomes were represented by the number of individuals who were hospitalized for respiratory illness, geocoded to the census tract of their residence. The SPARCS data was queried using ICD-9 codes for acute respiratory infections (ICD-9: 460–466), chronic obstructive pulmonary diseases and allied conditions (ICD-9: 490–496), and pneumonoconioses and other lung diseases from external agents (ICD-9: 500–508). Five years of data were combined (1998–2002, inclusive) and averaged in order to stabilize the rates. To further stabilize the rates, census tracts with fewer 250 persons or fewer than 20 people hospitalized for respiratory disease over the 5 years were excluded from the analysis. It is important to note that "persons hospitalized" is not necessarily equivalent to incidence or prevalence. Even though our previous research studies have used hospitalization records for health outcomes, the data has limitations and may be biased due the way in which NYC residents utilize hospitals and how existing disease is managed due to differences in health insurance coverage, physical access to primary care treatment and prevention, education regarding maintenance of health conditions, and other social and economic issues (Maantay 2007; Maantay et al., 2007, 2009a, b). The NYC respiratory hospitalization data was then age-adjusted using New York State as the standard population. As can be seen in Fig. 2.14 (left), there are high rates of respiratory hospitalizations in certain areas of NYC, which are indicated by the dramatically ballooned sizes of the affected census tracts. Some of these areas correspond to areas of high poverty and high minority population as is shown

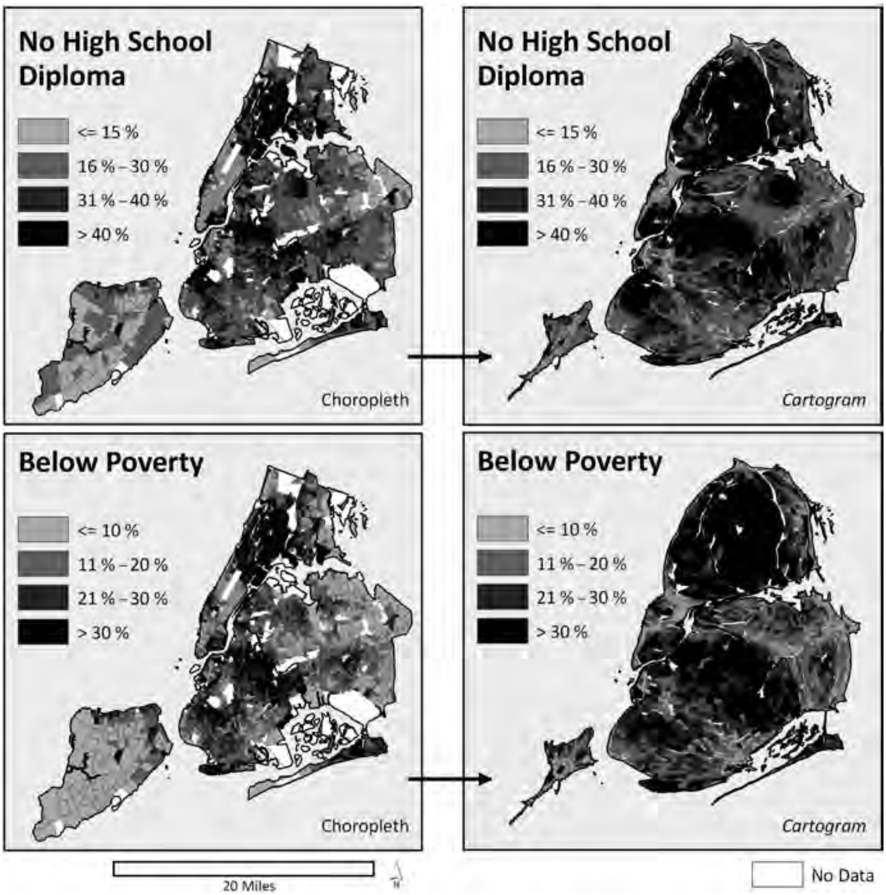

Fig. 2.12 Percentage of adults without a high school diploma and percent below poverty by census tract in NYC using choropleth mapping and cartograms

in Figs. 2.11 and 2.12. Once again, the utilization of cartograms can allow for an effective visualization of the data (Fig. 2.14, right).

Analysis: When the number of variables or the complexity of the relationships increase, it can be very useful to explore data statistically to augment the visualization of the variables individually. With a study such as this one, multiple ordinary least squares regression (OLS) is a natural choice. Age-adjusted rates for persons hospitalized for respiratory disease were used as the dependent variable, and percent Hispanic, percent non-Hispanic black, percent of adults without a high school education, and percent of people living below poverty were used as the independent variables. Note that there is no exposure estimate in the model, simply the health outcome versus the socio-demographic variables. Once again, rates were stabilized by excluding census tracts with fewer than 250 people or fewer than 20 people hospitalized between 1998 and 2002, inclusive ($n = 1{,}880$ tracts). The hospitalization

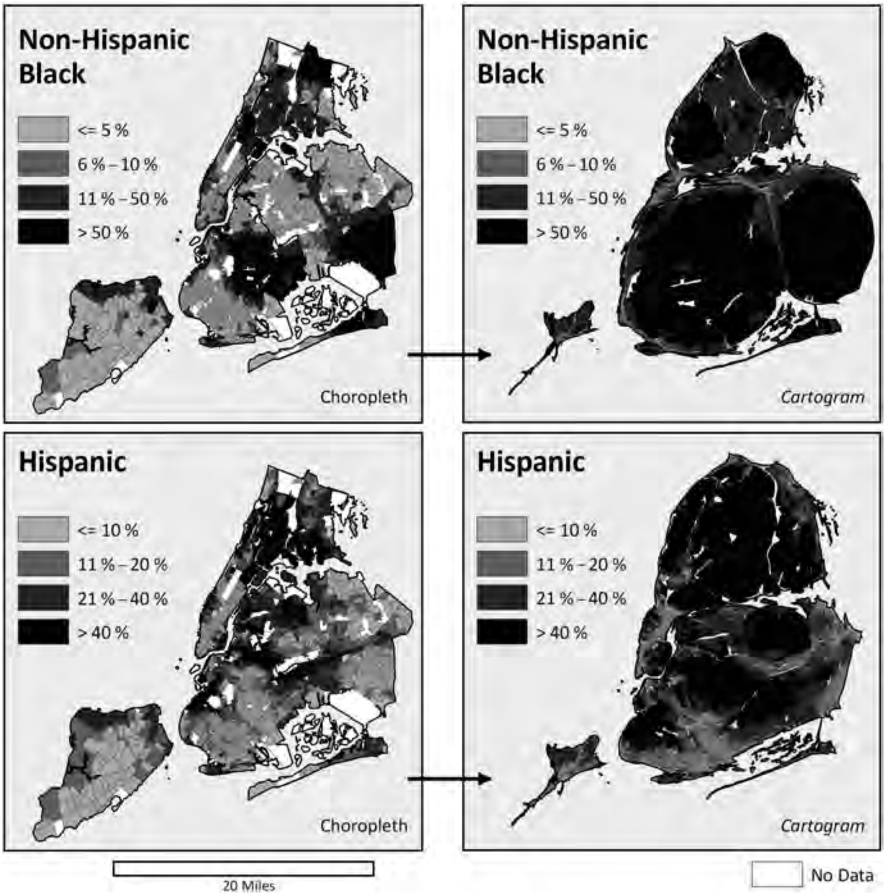

Fig. 2.13 Percent non-hispanic black and percent Hispanic by census tract in NYC using choropleth mapping and cartograms

data was log10 transformed in order to approximate a normal distribution in the residuals.

Ultimately, the model had an R^2 value of 0.50, suggesting that approximately 50% of the variance in the hospitalization rates can be explained by the socio-demographic measures. All of the SES variables showed a significant ($p<0.01$) positive association with respiratory hospitalization rates except for the education variable which was not significant (Table 2.2).

It can be informative to map out some of the OLS results, particularly the residuals, in order to get a more complete geographic understanding of the results. As can be seen in the map, there appears to be clustering of tracts with similar values. For example, the Whitestone/College Point neighborhoods in Queens show a group of tracts where the OLS underestimated the hospitalization rate (area of interest in Fig. 2.15). The Moran's I statistic ($I = 0.23$ with first order contiguity, $Z\ Score = 16.93$) statistically confirms spatial autocorrelation. This suggests that

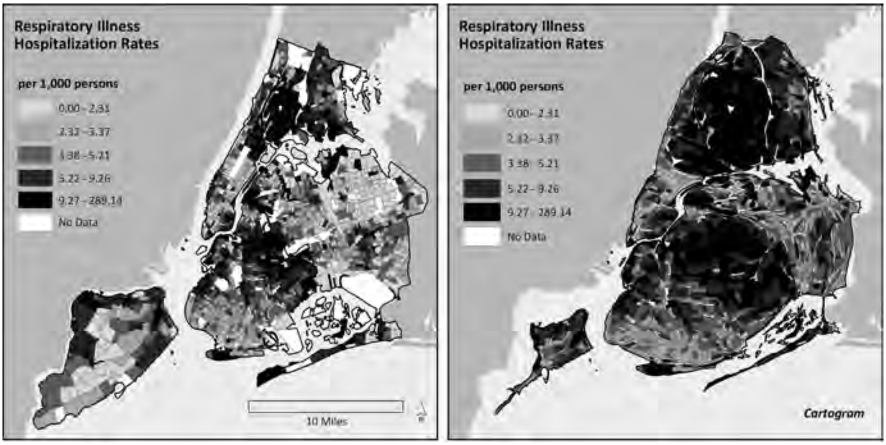

Fig. 2.14 Age-adjusted respiratory hospitalization rates in NYC visualized using a choropleth map (*left*) and cartogram (*right*)

Table 2.2 OLS parameter estimates

Parameter	Estimate	Std err	*t* value
Intercept[a]	0.29848	0.0128328	23.26
Pct without HS diploma	−0.00020	0.0007151	–0.29
Pct below poverty[a]	0.00919	0.0006683	13.75
Pct non-Hispanic black[a]	0.00276	0.0001960	14.07
Pct Hispanic[a]	0.00533	0.0003606	14.78

[a] $p<0.01$.

the residuals of the OLS were not randomly distributed which is a violation of OLS assumptions.

It may be illuminating, then, to examine the local nature of relationships between respiratory health and the socioeconomic variables. Geographically weighted regression (GWR) is a technique developed by Fotheringham et al. (1998) which quantifies locally varying relationships among data, rather than computing a global relationship as OLS does (similar to the distinction between Moran's I and a LISA). Instead of calculating global parameter estimates based on one regression, GWR performs many local regressions, each of which is influenced by the surrounding data resulting in a set of summary statistics for each regression point. In this way, GWR shows local variations in the regression relationships and is able to account for potential spatial non-stationarity, where the relationships among the independent and dependent variables vary over space. Locally varying relationships may suggest a number of things, including possible model misspecification sampling variation, or simply a relationship that intrinsically varies over space (Fotheringham et al., 2002). In this study, GWR was used as an exploratory tool to discover

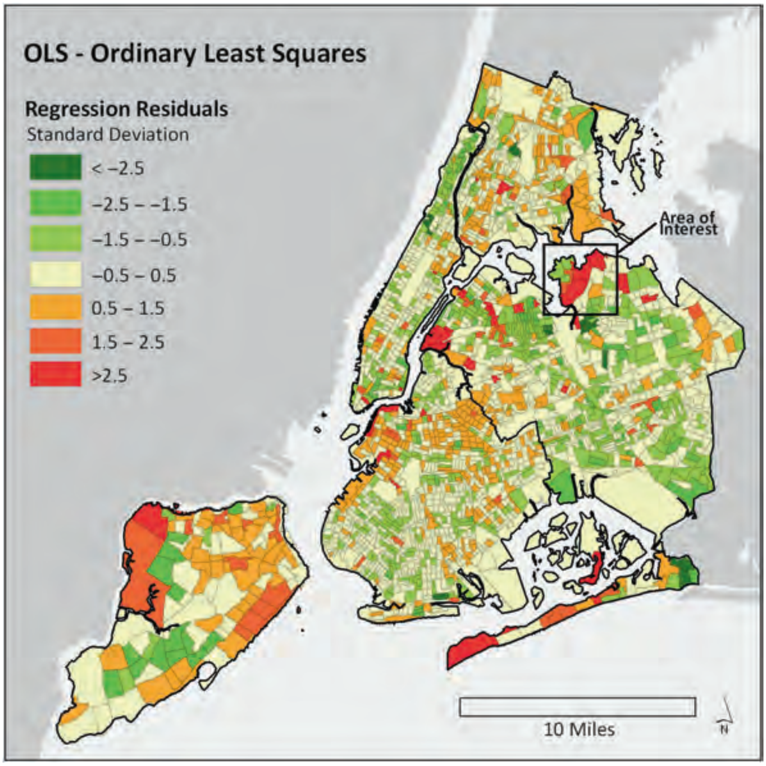

Fig. 2.15 OLS regression model residuals

how the relationships may vary and hypothesize as to why they may be spatially heterogeneous.

The GWR model was specified in the same way as the OLS, with respiratory rate as the dependent variable, and race/ethnicity, education, and poverty as the independent variables. GWR requires that the analyst specify a kernel bandwidth (distance radius) for the model identifying the size/radius of the area in which the local regression model is estimated. The two most common choices are a fixed (Gaussian) or adaptive (bi-square) kernel. The fixed kernel uses a constant radius, whereas the adaptive kernel involves varying the radius across the study area. The adaptive kernel works in a similar way to k-nearest neighbors, selecting a certain number of samples per local regression. As such, the adaptive kernel is able to “grow” when the samples are sparse, and “shrink” when there is a high density of sample points. Both fixed and adaptive bandwidths weight distant samples less heavy than proximal ones.

In our case study, a fixed bandwidth (rather than adaptive kernel) was assigned using an iterative process within the GWR3 software designed to minimize the Akaike Information Criterion (AIC) – a diagnostic statistic which describes the performance of the model. The result was a bandwidth of just under 1.5 miles.

Table 2.3 Comparison of OLS and GWR model diagnostics

Diagnostic	OLS	GWR
Residual sum of squares	104.3	72.2
Standard error	0.236	0.202
Akaike information criterion	−88.8	−570.8
R^2	0.501	0.655
$R_a{}^2$	0.500	0.635

Table 2.4 Five number summaries for GWR parameters. A Monte Carlo test of the local parameter estimates reveals significant spatial variability for all of the variables ($p<0.01$)

Parameter	Minimum	1st quartile	Median	3rd quartile	Maximum
Intercept	–0.2300	0.2366	0.2802	0.3661	1.3016
Pct without HS diploma	–0.0671	−0.0014	0.0007	0.0042	0.0767
Pct below poverty	–0.0423	0.0050	0.0081	0.0105	0.0620
Pct non-Hispanic Black	–0.0861	0.0023	0.0030	0.0037	0.0655
Pct Hispanic	–0.0424	0.0041	0.0057	0.0075	0.0304

Diagnostics of the GWR results suggest that it performed better than the global estimate as can be seen by the decrease in AIC, increase in R^2, and changes in other diagnostics such as residual sum of squares and standard error (Table 2.3).

Examining parameter estimates can be a bit more difficult in a GWR than an OLS. Global OLS regression results in one set of summary statistics for each parameter; however GWR has a set for each regression point, which in this case is each census tract centroid. Although it is not uncommon to report the results as a five number summary (Table 2.4), it is much more revealing to examine the geographically-varying parameters cartographically. This can include mapping values for the local R^2, error, parameter estimates, Cook's distance, or *t*-values.

When the *t*-values of the GWR results are mapped by interpolating the values between regression points (creating a continuous surface), the spatial variability in the relationships becomes visible. For instance, the relationship between percent non-Hispanic black and respiratory hospitalization rate are relatively stable while adjusting for the other variables, with only one area showing a negative association (area of interest in Fig. 2.16). The relationships with percent Hispanic and percent below poverty behave similarly to one another. They are somewhat stable, although both show significant negative relationships near the Whitestone/College Point neighborhoods of Queens (areas of interest in Figs. 2.17 and 2.18). Mapping the *t*-values for the percent of adults without a high school diploma, a variable that had no significance in the OLS, reveals that although the majority of NYC does not show a significant relationship, distinct areas of negative and positive associations can be identified (Fig. 2.19). The area with the positive relationship is, once again, around the Whitestone/College Point neighborhoods of Queens. The educational

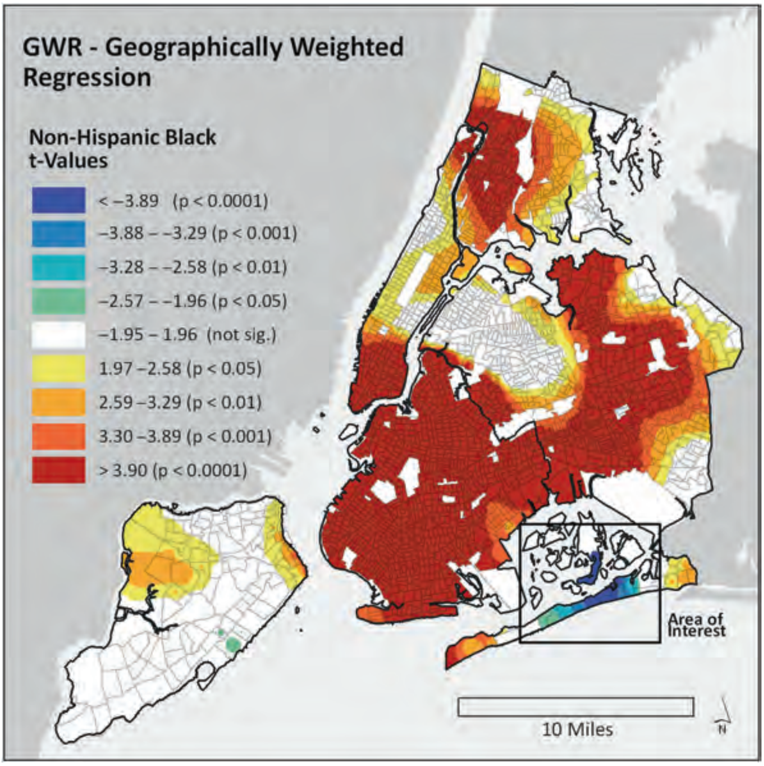

Fig. 2.16 GWR output – percent non-Hispanic black *t*-values

attainment variable must be interpreted with caution however, since the relationships are relatively weak and inconsistent. Also, excessive colinearity with other variables (e.g. poverty) may be skewing the results.

The potential spatial non-stationarity revealed by the GWR can identify areas of interest for further investigation. For instance, it may be useful to conduct a more qualitative assessment of the Whitestone/College Point neighborhoods, which suffer from high hospitalization rates and have high OLS residuals, in order to hypothesize about possible causes of the varying relationships of respiratory hospitalizations with percent Hispanic, poverty, and educational attainment. It is possible that these discrepancies are due to particular environmental conditions (built, physical or social) occurring in the area, which is information not available in the data sets used for the geovisualization. Such information could form the basis for additional questions and hypotheses which could be examined using more in-depth qualitative and quantitative methods.

Results Summary: As can be seen in Tables 2.2, 2.3, and 2.4, both the OLS and GWR suggest that there are measurable associations between hospitalizations for respiratory disease and selected socio-demographic variables. More specifically, as the percent of the population that is below the poverty level, non-Hispanic

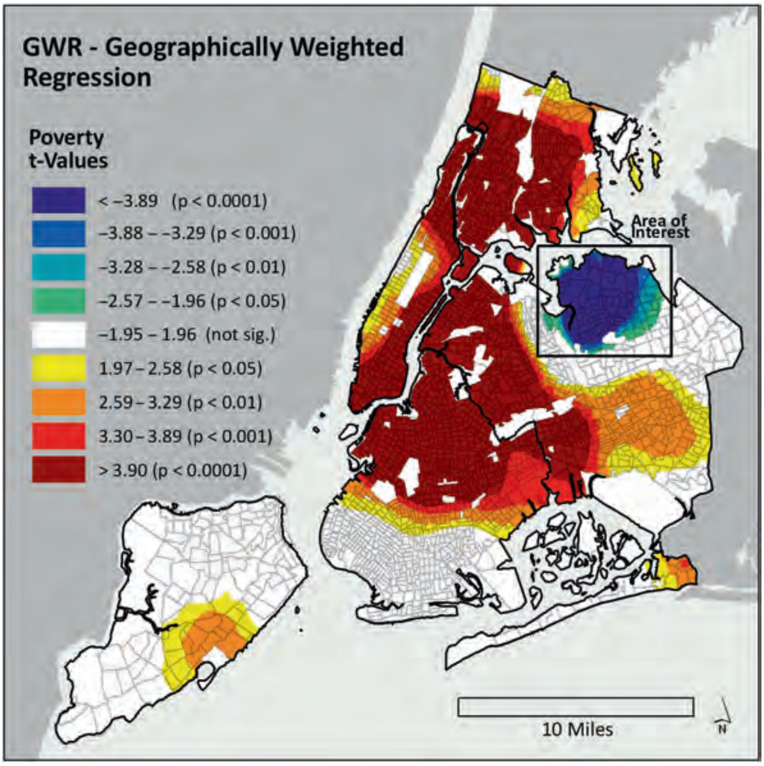

Fig. 2.17 GWR output – poverty rate *t*-values

black, or Hispanic increases, so do the hospitalization rates for respiratory disease. Approximately 50% of the variance of the hospitalization data is predicted by race/ethnicity and poverty status in the OLS model (AIC: –88.9) – approximately 64% is explained in the GWR (AIC: –570.8). Even though the OLS results in a high R^2, according to the AIC the local analysis (GWR) outperforms the global analysis (OLS). This appears to be mainly due to the non-stationary nature of the relationships around the Whitestone/College Point neighborhoods of NYC.

2.5 Conclusions

Geovisualization is an "intelligent" trial and error iterative process, involving various types of thematic mapping, data classification schemes, and geospatial analysis and geostatistics. It can be extremely useful in developing further hypotheses for testing, as well as guiding future analyses. It can also, in its own right, be helpful in answering questions about complex geospatial problems.

However, interpretation of the geovisualized results will have a higher degree of reliability if the analyst has a holistic and intimate knowledge of not only the data

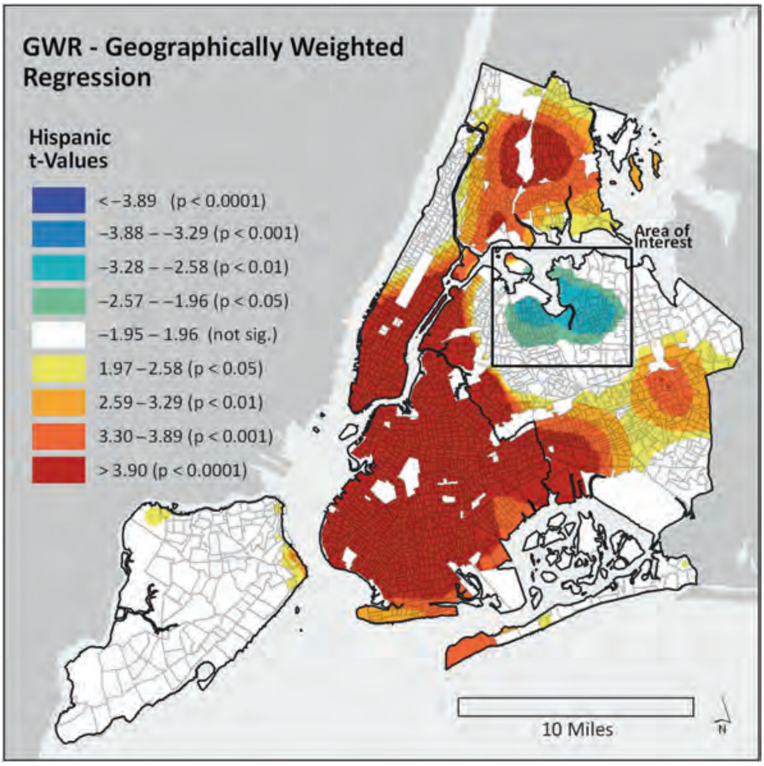

Fig. 2.18 GWR output – percent Hispanic/Latino *t*-values

itself, but also the geography of the study area and familiarity with the built, natural, and social environments. Otherwise, the explanatory power of geovisualization will be limited, and explications of any anomalous situations will be highly speculative, if they are possible at all.

As shown in the example of respiratory disease and environmental health justice in New York City, geovisualization techniques identified areas of concentrations of individual variables and potential spatial co-incidences amongst them, as well as geostatistical trends and anomalies. For instance, although most of NYC exhibits a positive relationship between census tracts with high proportion of non-Hispanic Black residents, Hispanic residents, and those below poverty with respiratory disease hospitalization rates, there are areas which suggest a spatially varying relationship. This is clearly seen when the results of the GWR are mapped, but may have otherwise not been detected.

The geospatial analysis gave us some answers, but also presented new questions, which led us to plans for a more detailed qualitative and quantitative analyses to ferret out some of the spatially inconsistent associations which only became observable with the geovisualization of the data and statistical results.

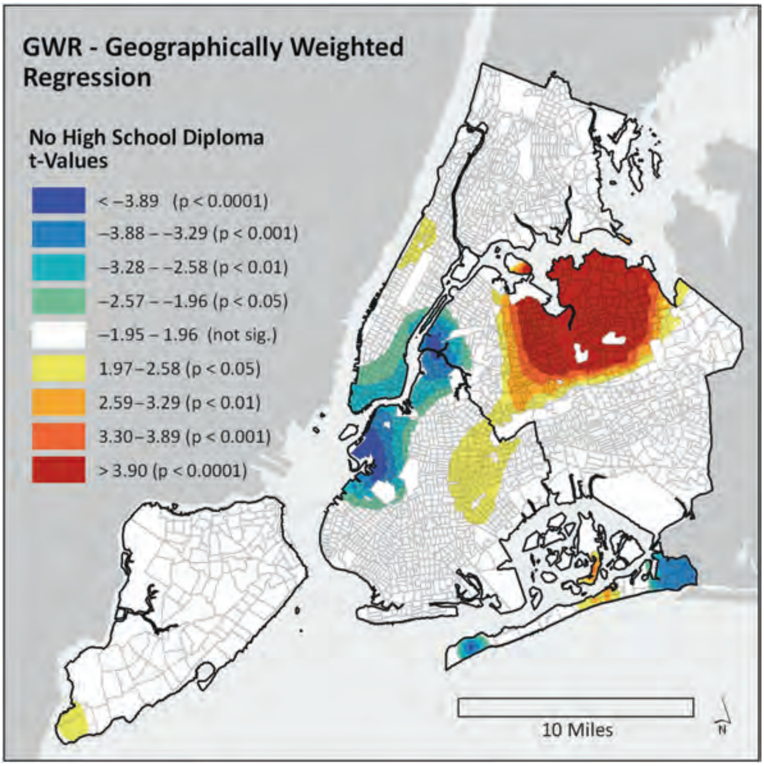

Fig. 2.19 GWR output – rate of adults with no high school diploma *t*-values

Acknowledgments This chapter resulted from on-going research, partially supported by the National Oceanic and Atmospheric Administration's Cooperative Remote Sensing Science and Technology Center (NOAA-CREST) under NOAA grant number NA17AE162. The National Institute of Environmental Health Sciences of the National Institutes of Health also provided critical support for this research under grant number 2 R25 ES01185-05. The National Center for Minority Health and Health Disparities, National Institutes of Health also helped fund this research as part of the Bronx CREED (the Bronx Center to Reduce and Eliminate Racial and Ethnic Health Disparities) under grant P60-MD0005-03. The statements contained within this paper are not the opinions of the funding agencies or the US. government, but reflect the authors' opinions. Thanks are also due to the member organizations of the South Bronx Environmental Justice Partnership, who understood the relevance of this research to environmental health justice and gave their unstinting encouragement and assistance in the effort.

Author Biographies

Andrew Maroko is an Assistant Professor in the Department of Health Sciences at Lehman College, City University of New York (CUNY). He received his Ph.D. in geography from The Graduate Center of CUNY, a GISc Graduate Certificate from Lehman College, CUNY, and his B.A. in biology from Rutgers University. His research interests include the spatial analysis of environmentally-linked diseases and environmental health justice, focusing on the associations between public health and socio-demographics in an explicitly geographical context. Much of his work examines methodological issues related to the quantification of the associations described above (e.g. geo-statistics, exposure modeling, and dasymetric disaggregation). He is a member of the Urban GISc Lab at Lehman College and teaches in the MPH program and in both the Health Sciences and Geography departments, and is an active collaborator in the South Bronx Environmental Justice Partnership, as well as both the Bronx CREED (the Bronx Center to Reduce and Eliminate Ethnic and Racial Health Disparities), and the CUNY Institute for Health Equity at Lehman College.

Juliana A. Maantay is Professor of Urban Environmental Geography, the Director of the GISc Program, and Director of the Urban GISc Lab at Lehman College, City University of New York. She also is a faculty member in Lehman's MPH program, the Earth and Environmental Sciences Ph.D. program at the CUNY Graduate Center, and the doctoral program (DrPH) in Public Health, as well as a research scientist with NOAA-CREST, the National Oceanic and Atmospheric Administration Center of Remote Sensing Science and Technology. Prior to her academic career, Dr. Maantay worked as an urban and environmental planner and policy analyst with governmental agencies, non-profit organizations, and private sector consulting firms and has been active in environmental health justice research and advocacy for more than 15 years.

Kristen Grady (Ph.D. Student, Earth and Environmental Sciences, City University of New York (CUNY) Graduate Center; B.A. 2008, Geography, Hunter College) is an Adjunct GISc Lab Technician in the Environmental, Geographic and Geological Sciences Department at CUNY Lehman College, and a Research Assistant at several CUNY institutes and centers, including the Urban GISc Lab at CUNY Lehman College, the Center for Urban Research at the CUNY Graduate Center, and the CUNY Institute for Sustainable Cities at CUNY Hunter College. Kristen is also a USDA Fellow in the Department of Environmental, Geographic and Geological Sciences at Lehman College. Recent GIS research projects include measuring access to parks in New York City, analyzing food deserts and access to healthy food in the Bronx in a study of urban agriculture and community gardening, mapping future flood zones due to climate change-induced sea level rise for New York City, and analyzing New York City's historical land use changes. Kristen's research interests include GIS Technologies, Urban Mental Health, Geographic Perception, and Cartography.

References

Anselin L (1995) Local indicators of spatial association – LISA. Geogr Anal 27:93–115

Aylin P, Bottle A, Wakefield J, Jarup L, Elliott P (2001) Proximity to coke works and hospital admissions for respiratory and cardiovascular disease in England and Wales. Thorax 56(3): 228–233

Boer J, Pastor M Jr et al (1997) Is there environmental racism? the demographics of hazardous waste in Los Angeles County. Soc Sci Q 78(4):793–810

Bowen W, Salling M et al (1995) Toward environmental justice: Spatial equity in Ohio and Cleveland. Ann Assoc Am Geogr 85:641–663

Burke L (1993) Race and environmental equity: a geographic analysis in Los Angeles. GEO Inf Syst October: 44–50

Chakraborty J, Armstrong M (1997) Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment. Cartogr Geogr Inf Syst 24(3):145–157

Chakraborty J, Schweitzer L et al (1999) Using GIS to assess the environmental justice consequences of transportation system changes. Transact GIS 3(3):239–258

Cliff AD, Ord J (1973) Spatial autocorrelation. Pioneer, London

Davies C, Medyckyj-Scott D (1994) Introduction: the importance of human factors. In: Hearnshaw HM, Unwin DJ (eds) Visualization in geographical information systems. Wiley, New York, NY, pp 189–192

Dorling D, Newman M, Allsopp G, Barford A. http://www.worldmapper.org/. Worldmapper Project Retrieved 1/6/2010

Edwards J, Walters S, Griffith R (1994) Hospital admissions for asthma in preschool children: relationship to major roads in Birmingham, United Kingdom. Arch Environ Health 49(4): 223–227

Fotheringham SA, Brunsdon C et al (2002) Geographically weighted Regression: the analysis of spatially varying relationships. Wiley, Chichester, UK

Fotheringham SA, Brunsdon C, Charlton ME (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plan A 30: 1905–1927

Gastner M, Newman M (2004) Diffusion-based method for producing density equalizing maps. Proc Natl Acad Sci 101(20):7499–7504

Houle B, Holt J et al (2009) Use of density-equalizing cartograms to visualize trends and disparities in state-specific prevalence of obesity: 1996–2006. Am J Public Health 99(2):308–312

Koch T (2005) Cartographies of disease: maps, mapping, and medicine. ESRI Press, Redlands, CA

Kraak MJ, Orneling FJ (1996) Cartography: visualization of spatial data. Pearson Education Ltd, Harlow

Maantay J (2002) Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. Environ Health Perspect 110(2):161–171

Maantay J (2007) Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. Health and place. Linking Environ Justice Popul Health Geogr Inform Sci 13:32–56, special issue

Maantay J, Maroko A, Alicea C, Strelnick A (2009a) Geographic information systems, environmental justice, and health disparities: the need for an interdisciplinary approach to study asthma and air pollution in the Bronx, New York. In: Freudenberg N, Saegert S Klitzman S (eds) Urban health and society: interdisciplinary approaches to research and practice. Jossey Bass, pp 93–126

Maantay JA, Maroko AR, Porter-Morgan H (2008) A new method for population mapping and understanding the spatial dynamics of disease in urban areas. Urban Geogr 29(7):724–738

Maantay J, Maroko A, Herrmann C (2007) Mapping population distribution in the urban environment: the cadastral-based expert dasymetric system (CEDS). Cartogr Geogr Inf Sci 34(2):77–102 special issue: Cartography 2007: Reflections Status, and Prediction

Maantay J, Timander L, Graziosi G, Meyer L (1997) The Bronx Toxic Release Inventory Report: US. Environmental Protection Agency Region 2 and Hostos Community College/Center for a Sustainable Urban Environment [CSUE]

Maantay J, Tu J, Maroko A (2009b) Loose-coupling an air dispersion model and a geographic information system (GIS) for studying air pollution and asthma in the Bronx, New York City. Int J Environ Health Res 19(1):59–79

Maantay JA, Ziegler J (2006) GIS for the urban environment. Environmental Systems Research Institute (ESRI), Redlands, CA

MacEachren AM, Kraak MJ (1997) Exploratory cartographic visualization: advancing the agenda. Comput GEO sci 23:64–81

Morello-Frosch R, Pastor M et al (2001) Environmental justice and southern California's "riskscape" – the distribution of air toxics exposures and health risks among diverse communities. Urban Aff Rev 36(4):551–578

Neumann CM, Forman DL et al (1998) Hazard screening of chemical releases and environmental equity analysis of populations proximate to toxic release inventory facilities in Oregon. Environ Health Perspect 106(4):217–226

O'Neill M, Jerrett M, Kawachi I, Levy J, Cohen A, Gouveia N et al (2003) Health, wealth, and air pollution: advancing theory and methods. Environ Health Perspect 111(16):1861–1870

Orford S (2005) Cartography and Visualization. In: Castree N, Rogers A, Sherman D (eds) Questioning geography: fundamental debates. Blackwell Publishing, Oxford

Pellow D and Brulle R (eds) (2005) Power, justice, and the environment: a critical appraisal of the environmental justice movement. MIT Press, Cambridge

Perlin SA, Setzer RW et al (1995) of industrial air emissions by income and race in the United States: an approach using the toxic release inventory. Environ Sci Technol 29(1):69–80

Pollock PH, Vittas ME (1995) Who bears the burden of environmental pollution? race, ethnicity, and environmental equity in Florida. Soc Sci Q 76(2):294–309

Sheppard E, Leitner H et al (1999) GIS-based measures of environmental equity: exploring their sensitivity and significance J Expo Anal Environ Epidemiol 9:18–28

Smargiassi A, Kosatsky T, Hicks J, Plante C, Armstrong B, Villeneuve P (2009) Risk of asthmatic episodes in children exposed to sulfur dioxide stack emissions from a refiner point source in Montreal, Canada. Environ Health Perspect 117(4):653–659

Sui D, Holt J (2008) Visualizing and analysing public-health data using value-by-area cartograms: toward a new synthetic framework. Cartographica 43(1):3–20

Visvalingam M (1994) Visualization in GIS, Cartography and ViSC. In: Hearnshaw HM Unwin DJ (eds) Visualization in geographical information systems. Wiley, New York, NY, pp 18–25

Zandbergen PA, Chakraborty J (2006) Improving environmental exposure analysis using cumulative distribution functions and individual geocoding. Int J Health Geogr 5:23