Using Geographic Information Systems to Define and Map Commuting Patterns as Inputs to Agent-Based Models

David P. Chrest and William D. Wheaton

June 2009

NOTE: Online readers can click on most maps to view a larger interactive version in a separate window.

RTI Press

About the Authors

David P. Chrest, BS, is a geographic information systems analyst at RTI International.

William D. Wheaton, MA, is manager of RTI's Geographic Information Systems program.

Using Geographic Information Systems to Define and Map Commuting Patterns as Inputs to Agent-Based Models

David P. Chrest and William D. Wheaton

Abstract

By understanding the movement patterns of people, mathematical modelers can develop models that can better analyze and predict the spread of infectious diseases. People can come into close contact in their workplaces. This report describes methods to develop georeferenced commuting patterns that can be used to characterize the work-related movement of US populations and help agent-based modelers predict workplace contacts that result in disease transmission. We used a census data product called "Census Spatial Tabulation: Census Track of Work by Census Tract of Residence (STP64)" as the data source to develop commuting pattern data for agent-based synthesized populations databases and to develop map products to visualize commuting patterns in the United States. The three primary maps we developed show inbound, outbound, and net change levels of inbound versus outbound commuters by census tract for the year 2000. Net change counts of commuters are visualized as elevations. The results can be used to quantify and assign commuting patterns of synthesized populations among different census tracts.

Contents

RTI INTERNATIONAL

Introduction

Models of Infectious Disease

The National Institute of General Medical Sciences of the National Institutes of Health created the Models of Infectious Disease Agent Study (MIDAS) network in 2004. The purpose of this network, which is a collaborative project of scientists, is to develop computational models of the interactions between infectious agents and their hosts, disease spread, prediction systems, and response strategies. MIDAS models are useful not only to other scientists, but also to policy makers because the models provide insights into infectious disease dynamics; they can be applied to examine the implications of policy options and suggest response options.

The models developed for MIDAS include stochastic, spatial, agent-based models in which the interactions of people (referred to as "agents") are specified in terms of both time and space. For some infectious diseases, close contact of individuals enhances the spread of the pathogen. Therefore, understanding where people are in close contact and the patterns of movement between places of close contact enhances the ability of the models to simulate the spread of disease in the population. People come into close contact with others in many different situations, such as in houses, schools, workplaces, dormitories, and mass transit modes (e.g., buses, airplanes, and subways). The patterns of movement between places of close contact affect the patterns of spread of infectious agents.

Geographic Information Systems

The results of geographic information system (GIS) analysis are often displayed on a map, which enables users to identify immediately the spatial data characteristics over an area. A map brings forth a visual impact of the analysis and gives it spatial realism. While reviewing a well-designed map, the user can immediately see what and where the entity represented by the data will, is, or has occurred.

An advantage to GIS mapping is that multiple data sets can be simultaneously displayed, which enriches understanding. For example, to research and investigate the cause of a historic plague outbreak in a city, population concentrations, areas of social status, hydrography, areas of river contamination, sewer lines, known rat-infested areas, and disease-infested locations over time can be overlaid on a map for evaluation and analyses. After all of these data are mapped, users may then be able to see clearly that, for example, a plague originated from a rat-infested area in a poor neighborhood near a riverbank or contaminated sewer. Using robust GIS software, this and many other types of spatial analyses can be performed. The spread of disease from its origin over time can even be animated on a map.

GIS scientific analysis of the data and the map resulting from the analysis combine to communicate data to the viewer. The most extensive data maps, such as a cancer atlas (Pickle et al., 2003) or a count of galaxies (Knapp, 1997), reveal millions of pieces of information on a single page. No other method for the display of statistical information is so powerful.

Data Visualization

Cartography, the study and practice of making geographical maps, models reality and disseminates spatial information through a map (Tufte, 2001). A cartographer takes real-world data and information and conceptualizes a map through design and symbols. Maps do more than show location: they can be insightful tools and can lead to new ways of understanding through geographic patterns of phenomena (MacEachren, 1995).

In the mid-1980s, Abraham Verghese, MD, was surprised when patients with HIV, which was known to be more of an urban than a rural problem, began to dominate his rural practice in eastern Tennessee. Dr. Verghese first took a local map and located where the HIV patients in his service area lived. Then, he located where his patients had lived between 1979 and 1985 on a national map. The maps revealed that most of his patients grew up in eastern Tennessee, and then many of them moved to large cities, such as New York City, Miami, and San Francisco. Then the patients migrated back to their hometowns, where they eventually died (Krygier and Wood, 2005). Similar to the insights that were gained for the spread of HIV, the commuting maps described in this report can lead to important insights about spatial patterns of disease spread from workplace to home social contacts.

Three Exemplary Maps

For the MIDAS project, we created three US commuting maps that show the movement of people during daily commutes to and from their places of residence and work so we could visualize a common human activity that occurred each weekday. These maps illustrate commuting levels and reveal where people traveled to work, where people came from when they commuted to work, and the net change in levels of inbound versus outbound commuters (ESRI, 2007b). In fall 2005, we used US Census Bureau data and ArcGIS software from Environmental Systems Research Institute (ESRI) International to design the following three poster-sized US commuting maps (Maps 1–3). The viewer can easily inspect them and gain an understanding of both national and local patterns of commuters. Thus, these maps are an important visualization tool for model developers.

1. Commuting in 2000: Where People Are Going (Map 1)

2. Commuting in 2000: Where People Are Coming From (Map 2)

3. Commuting in 2000: Places Gaining People vs. Places Losing People (Map 3).

We presented these maps at the Map Gallery exhibit during the 26th Annual ESRI International User Conference (August 7–11, 2006) in San Diego, California.

The principal purposes of this report are to describe the data product Census 2000 Spatial Tabulation Product 64: Census Tract of Work by Census Tract of Residence: 2000, otherwise known as STP64, and document and how we used it as a source to create the synthesized population database (for a detailed description, see Wheaton et al., 2009). The STP64 data were used with business data sources and a synthetic population database (Wheaton et al., 2009) to assign workers who reside in one tract to a place of work in another tract. Essentially, the place of residence and place of work data from STP64 define the correct flow of commuters and therefore enhance the development of realistic contact patterns that models use to predict disease spread.

An additional purpose for the STP64 data is to visualize commuting patterns and to develop maps that illustrate these data. Thus, a second purpose of this report is to explain the methods for developing GIS maps and, in particular, the approach for assigning commuting patterns to populations for use in agent-based models.

For this methods report, the PDF versions of certain maps have up to five additional layers. Online readers of this report can click on most figures to view a larger version in a separate window on screen.

Click on a map to open a separate, interactive version of the map that has multiple layers. Use the Adobe Reader Layers feature to turn the layers on or off. (In Adobe Reader version 8 or 9, the icon looks like two diamonds. Adobe Reader 7 instead has a "Layers" tab on the far left of the screen.) Click on this icon (or tab) presents a list of the map's layers. Click on an eye symbol to activate/deactivate that layer on the map.

Map 1. Outbound commuters

Map 2. Inbound commuters
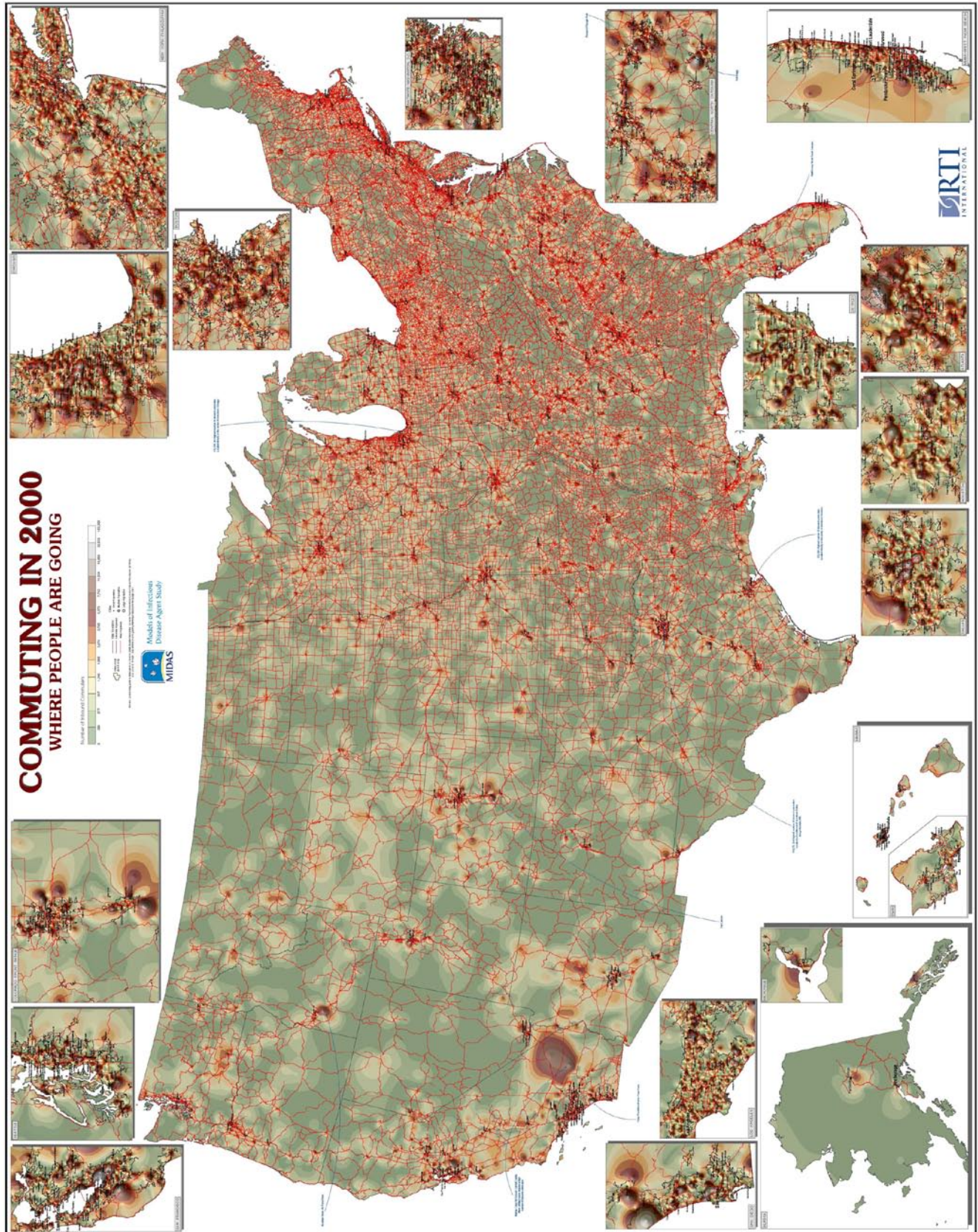
Map 3. Net change of commuters
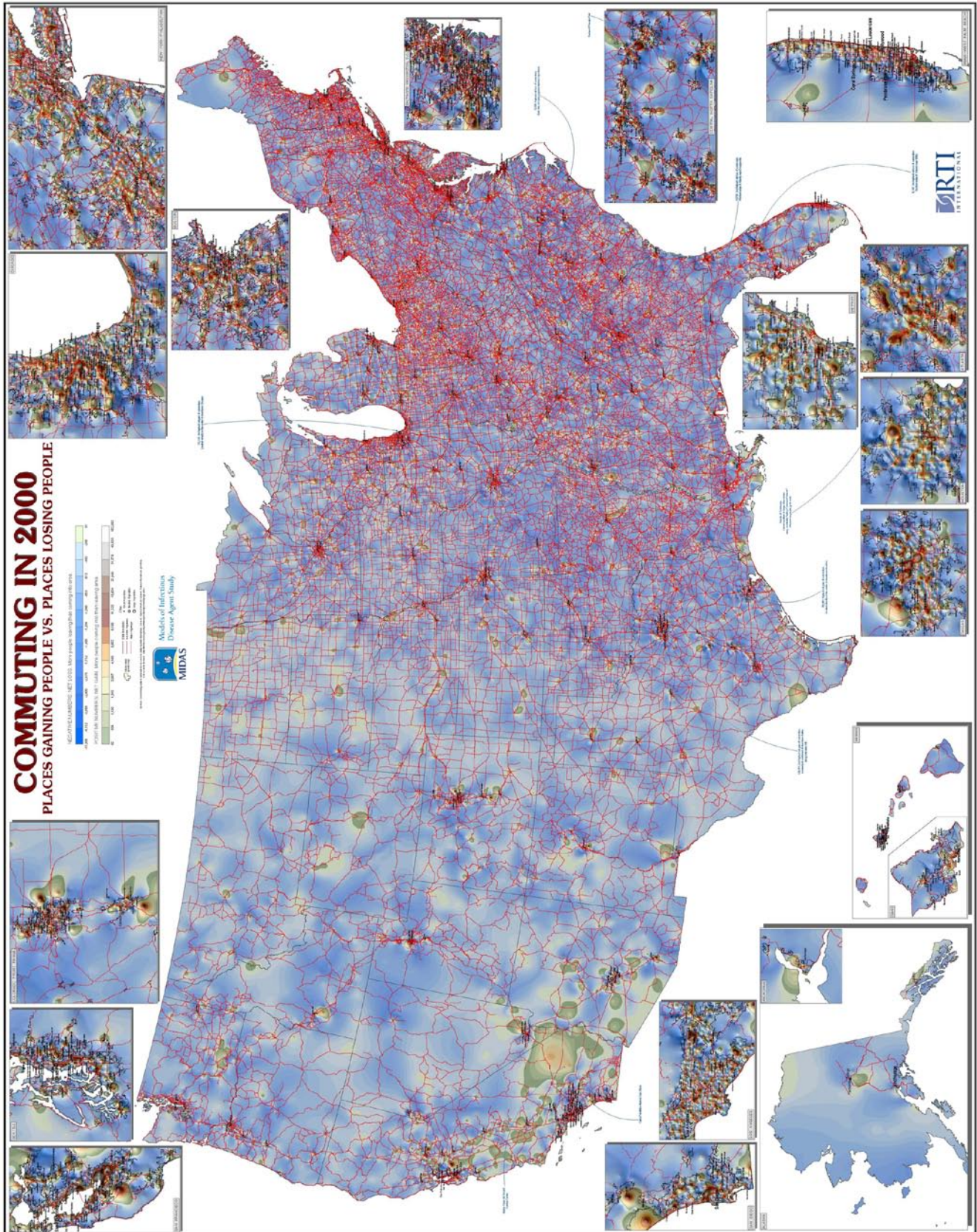
Methods

The Data Source

In 2004, the US Census Bureau released STP64, which includes the number of workers in each census tract of work by tract of residence (US Census, 2004). Table 1 depicts how an STP64 data table is organized (using hypothetical census tracts).

Table 1. Organization of an example STP64 data table

Record Number	Tract of Residence	Tract of Work	Number of Workers
1	A	A	123
2	A	B	351
3	A	C	58
4	B	A	209
5	B	B	312
6	B	C	245
7	C	A	290
8	C	B	176
9	C	C	99

Note: This is a hypothetical example that includes three census tracts denoted A, B, and C.

For each tract of residence and tract of work combination with a flow of commuters, a record in the table indicates the census tract of residence, the census tract of work, and the number of workers. However, this table does not include information on which intermediate census tracts were transited during the commute.

The STP64 table is used to assign synthesized agents to appropriate workplaces. Assigning synthesized agent workers to places of work creates potential contact relationships between people who live in different census tracts but who work in the same census tract.

For mapping purposes, we had to process the STP64 table to produce a table that combines for each census tract of residence a summary of all the workers who work in other census tracts. The results of the table frequency and summation process that we ran on the STP64 data yielded variables and data like those shown in Table 2. The format for the data file, therefore, is a flat file with these variables. A similar frequency summation run on the census tract of work

field yielded a table that shows the inflowing workers to each census tract of work (Table 3).

Further processing of these tables produces a single table with a record for each census tract that contains counts of gross outflowing workers, counts of gross inflowing workers, and the net change inflow or outflow of workers. A net gain indicates that, during the course of a day, more people commute into the census tract than leave to work outside that census tract.

The US Census Bureau defines workers as anyone 16 years of age or older who was employed at work, full time or part time, during the Census 2000 reference week (the last week in March 2000). Question 22 on the Census 2000 Long Form asks respondents where they worked the prior week. If they worked at more than one location, they should write in the address where they worked most the prior week; this address was used to determine the census tract of the respondent's workplace. Consequently, a worker with multiple worksites is linked to only one worksite.

Table 2. Workers who depart each census tract of residence for work elsewhere

Tract of Residence	Outflowing Workers
A	409 (sum of workers for records 2 and 3 in Table 1)
B	454 (sum of workers for records 4 and 6 in Table 1)
C	466 (sum of workers for records 7 and 8 in Table 1)

Table 3. Workers who arrive at each census tract of work

Tract of Work	Inflowing Workers
A	499 (sum of workers for records 4 and 7 in Table 1)
B	527 (sum of workers for records 2 and 8 in Table 1)
C	303 (sum of workers for records 3 and 6 in Table 1)

Census tracts are statistical subdivisions of a county that are used for data organization purposes. Census tracts usually contain between 2,500 and 8,000 persons. The Census Bureau attempts to define tracts so that they include persons with similar demographic characteristics, economic status, and living conditions. The geographic size of census tracts varies widely, depending on an area's population density. For example, a census tract in Manhattan, New York, can be as small as 0.03 square miles, or just more than 18 acres in size (Figure 1), whereas a census tract in sparsely populated central Nevada is 8,689 square miles, which is much larger than most counties and larger than the individual states of Rhode Island, Delaware, Connecticut, and New Jersey (Figure 2). Because of this variability in tract size, in cities of substantial population, STP64 data reveal geographically detailed commuting patterns over short distances. Yet in sparsely populated, rural areas, especially in the western part of the country, where distances between cities can be vast, there is much less commuting from one tract to another, as shown in (the detailed maps discussed previously) Maps 1–3.
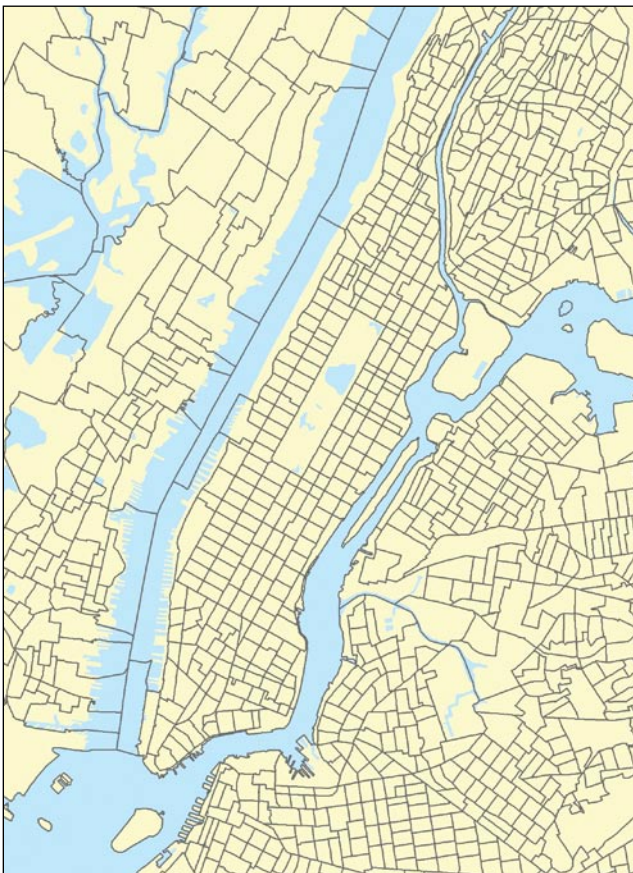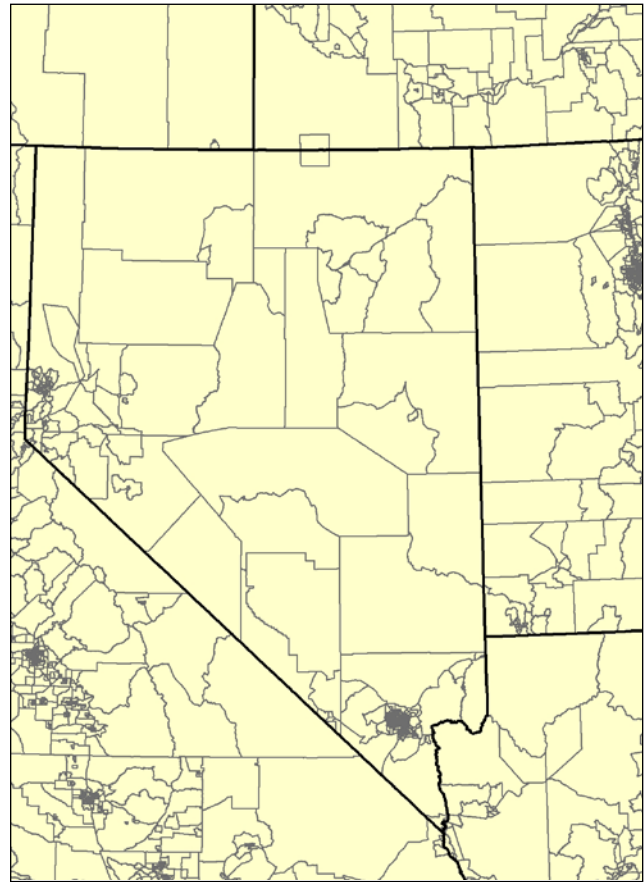
Figure 1. Manhattan, N.Y., census tracts

Figure 2. Nevada census tracts

Developing the Maps

To map the STP64 data effectively so that they are comprehensible to users, map designs cannot be either overly abstract or overly detailed. Users must be able to quickly perceive the flow patterns of commuters across the country.

We designed the maps for this project to enable users to visualize as elevations the counts of inbound, outbound, and net gain or loss of commuters. The maps use colors and shading similar to those of common terrain maps; however, instead of showing a range of heights, the colors represent a range of commuters.

The maps were created using points that represent the center of each census tract and associate the STP64 commuter data with the points. The areas between these points were then interpolated to create elevations that represent not actual terrain but the numbers of commuters. For example, Figure 3 shows inbound commuters arriving at their Los Angeles,
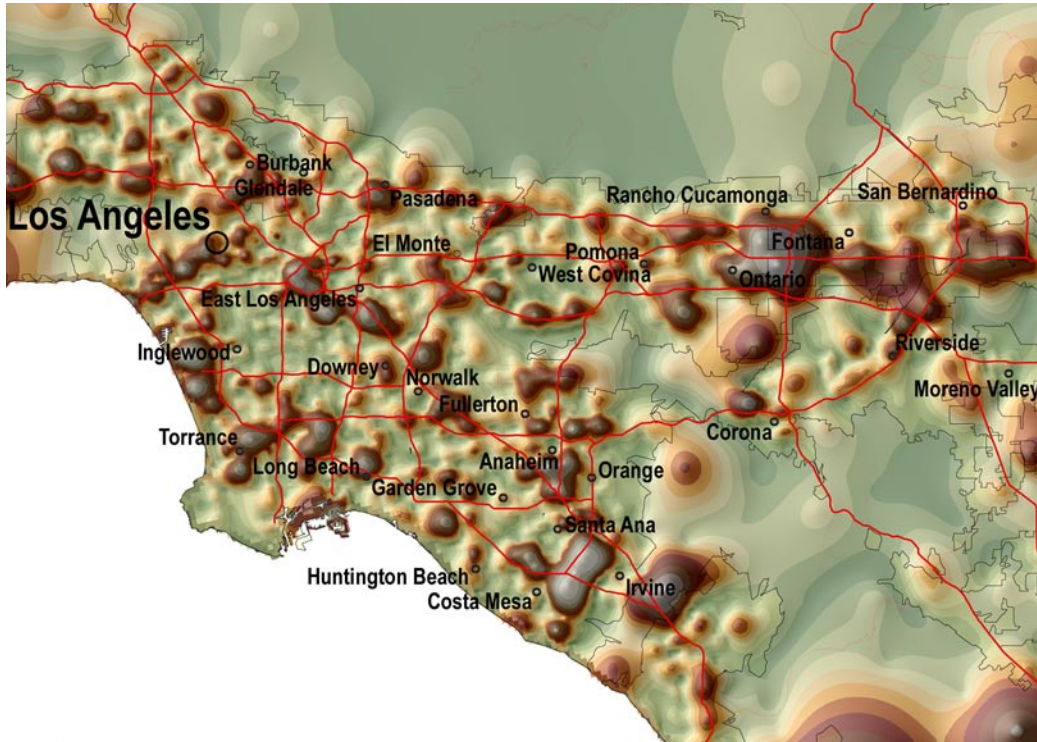
California, area census tract of work. A hillshade layer was then created for effect. Generated from the STP64 point values, a hillshade is a raster GIS data layer that is designed to simulate the sun's rays over the varied terrain. Shadows create a three-dimensional effect that provides a sense of visual relief to a map. We added main highways, cities, and states to complete the maps. We enlarged labels and city symbols for better visibility in all figures.

The following describes the main steps that we used to create these map effects. As stated earlier, we used ESRI's ArcGIS Desktop GIS software for this work.

1. Defining centroids. We first created census tract centroid points. That is, we created a geographic point data layer (map) of all 50 states and the District of Columbia with a point located in the center of each tract. We then joined the STP64 inbound, outbound, and net loss or gain of commuters data for each tract to all of the tract centroid points. Every census tract has a unique identifier code that is present in the tract centroid points, as well as in these STP64 data. Therefore, STP64 data were accurately matched with

the centroid points using these unique identifier codes. The result of the matching is a GIS point data layer that represents all census tracts and contains the commuting attributes. Figure 4 shows tract centroid points in census tract polygons for the Raleigh, Durham, and Chapel Hill area in North Carolina.

Figure 4. Census tract centroid points for the Raleigh, Durham, and Chapel Hill area in North Carolina.
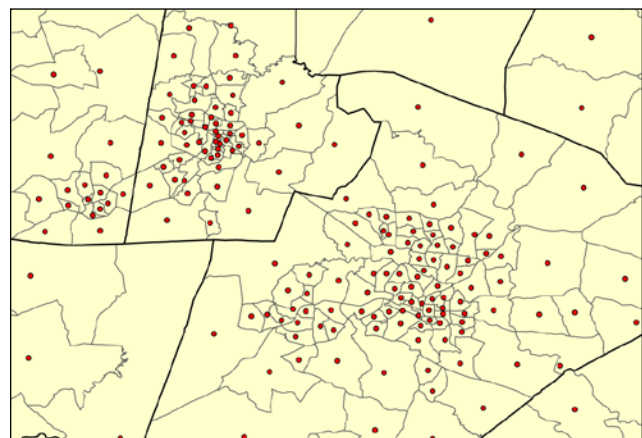
2. Creating raster data layers. We treated the commuting values for each tract point as if they were elevations. Using ArcGIS Desktop's Topo to Raster Tool, we interpolated the values between all known neighboring points, creating a surface layer. In this case, the census tract centroid points with STP64 data (i.e., the known surface values) are unevenly spaced throughout because the census tracts are smaller in areas of higher population density. The consequence for this raster data layer is that the valleys and peaks are tighter and close together in cities where the tracts are smaller; the slopes are longer and more gradual in sparsely populated areas.

This type of pattern is evident in Figure 3 (previous page). High peaks, hills, and valleys abound in the dense, suburban area of Los Angeles, California. They then drop off to a smooth plain north of the city, where the population immediately decreases along the San Gabriel Mountains.

Any census tract or block group of block data values assigned to centroid points can be similarly interpolated. For this project, we used worker numbers, but any point data with numerical values, such as crimes, air temperatures, or mortality rates, can be represented as a surface layer.

3. Define data breaks. The surface raster was divided into 13 categories using the natural-breaks method of data classification (ESRI, 2007a). Data are partitioned into classes based on natural groups in the data distribution using an algorithm that groups data by minimizing the variance within groups, or categories, and then maximizing the variance between clusters. This classification scheme determines the best arrangement of values into classes by iteratively comparing the sums of the squared difference between observed values within each class and class means (TerraSeer, 2008). The natural-breaks method is popular for mapping quantitative data because it attempts to find the most suitable class ranges by testing them against the distribution of the entire dataset so that the resulting class ranges reflect the structure of the distribution (Jenks, 1967).

4. Assign color ramps for outbound and inbound commuting. The commuting raster surface layers for the inbound and outbound commuting maps are symbolized using the same colors. To imitate recognizable terrain maps, we created green-to-brown-to-white color ramps. The color ramps closely mimic the ramps used in relief maps made by Raven Maps. The colors in the maps are designed to give an idea of commuting levels in conjunction with the shaded terrain effect rather than depict discrete values.

The color ramps begin with the lowest commuting values as dark green, move to lighter green, then to light beige, followed by darker browns that turn gray. They end with white to represent the peaks of the highest values. We spent considerable time choosing correct colors so that users will see a smooth transition in values. Figure 5 displays the inbound commuter categories and color ramp.

5. Assign color ramps for net change commuting. We created the map showing net gain or loss of commuters using different color schemes. Although the inbound and outbound commuting maps show only positive data values, the STP64 values were both negative and positive the net change commuting map. Census tracts with positive values mean that, during the course of the day, more people are coming into the tract to work than are leaving. Census tracts with negative values mean that more people are leaving the tract to go to work than coming into the same tract to work.

The negative values are displayed using blue color ramps; this color is commonly used to designate below sea level on bathymetric maps. Dark blue is used for the lowest numbers (areas with the greatest movement out), and then the color lightens over 12 categories until positive numbers are reached (very light green). Figure 6 shows the color ramp for the negative STP64 values.

To illustrate this combination (i.e., positive and negative values), which denotes net gains, Figure 7 shows the final result in the map for central North Carolina.
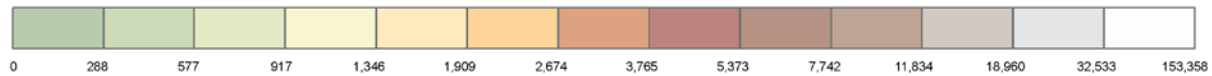
Figure 5. Inbound commuters color ramp

| 0 | 288 | 577 | 917 | 1,346 | 1,909 | 2,674 | 3,765 | 5,373 | 7,742 | 11,834 | 18,960 | 32,533 | 153,358 |

Figure 6. Outbound commuters color ramp

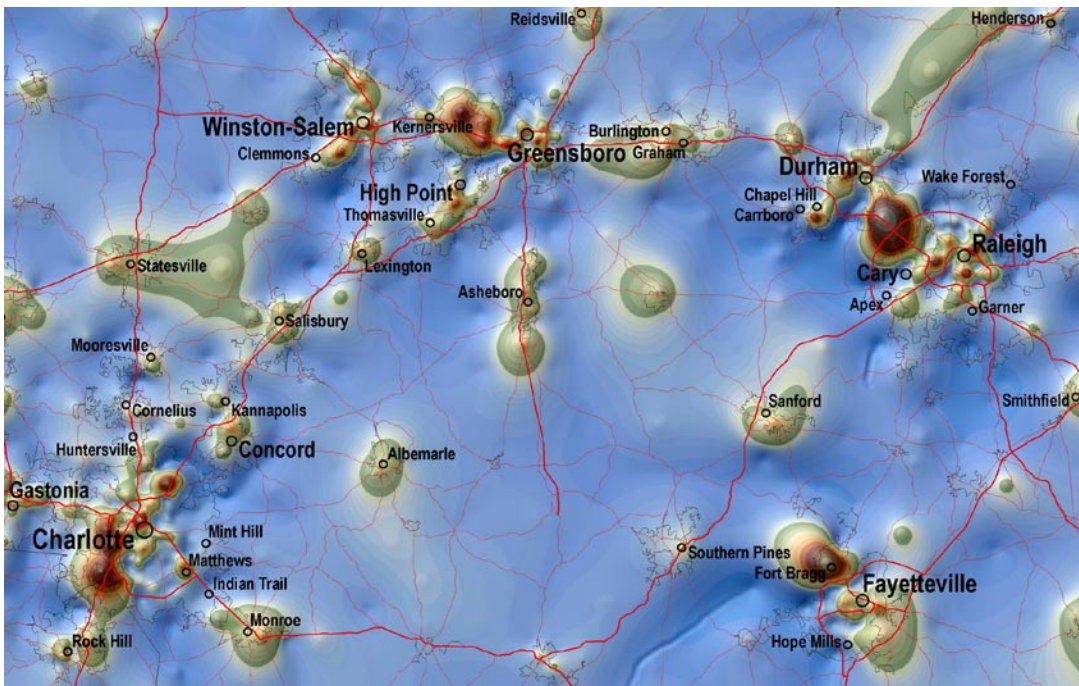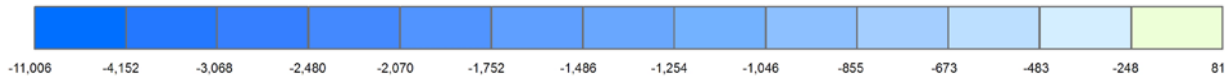| -11,006 | -4,152 | -3,068 | -2,480 | -2,070 | -1,752 | -1,486 | -1,254 | -1,046 | -855 | -673 | -483 | -248 | 81 |

Figure 7. Central North Carolina commuter map showing net change in inbound and outbound commuters

Click on the map to view the layers in an interactive PDF.

6. Generate a three-dimensional effect. To bring depth to the terrain and enhance hills, mountains, and valleys as physical features, we created a hillshade raster data layer for each map. The Hillshade Tool in ArcGIS Desktop computes hillshade values for a raster surface by considering the illumination angle and shadows. When designing the map layout, we placed the hillshade raster layer directly beneath the STP64 commuting raster surface. We assigned a transparency level of 50 percent to the commuting raster layer so that users can see the hillshade underneath while the map retains the colors of the commuting data. This combination gives the maps an enriched cartographic effect, as observed in Figures 3 and 7. Figure 8 shows the hillshade layer below the data shown in Figure 3. (Click here to see the maps together.)

Figure 8. Hillshade layer that shows the Los Angeles, Calif., area

Click on the map to view the layers in an interactive PDF.
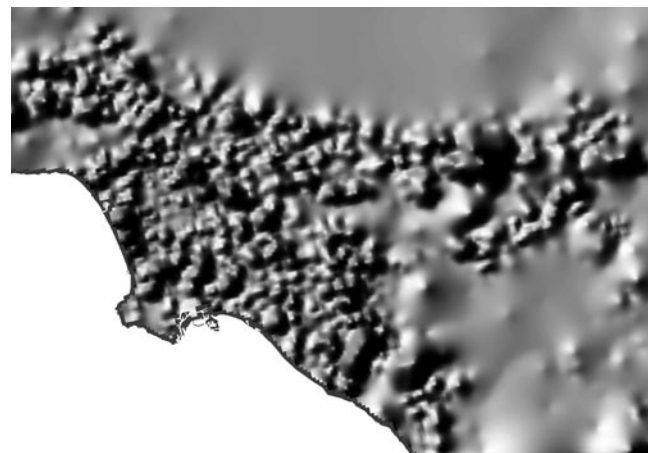
7. Handle boundaries. When produced, raster layers default to a rectangular extent. Consequently, to achieve the more realistic geographical boundaries, we used a detailed US outline vector data layer to clip both the hillshade and commuting rasters. The simple Clip Tool is used for this purpose, to erase all raster values beyond the US boundary. We did this for the states of Alaska and Hawaii.

8. Make finishing touches. We then added the final data layers. For example, we added a thin boundary outline to smooth the raster cells that are visible at the edge of Alaska and Hawaii, as well as at the edge of the coterminous United States. Urban areas, major highways, interstate highways, labeled cities, and state boundaries provide points of reference. We took great care to symbolize these layers so that they did not overpower the main message of the map. Commuting data had to be clearly visible, whereas the lines that delineated the roads, urban areas, and state boundaries were thin but present.

Results

We used the methods described in the previous section to generate commuting maps for the entire United States. We recognized to present readable data, that these maps would need many insets because of the complex commuting pattern over small areas in some major metropolitan areas. The detailed patterns were difficult to discern in certain areas if the map displayed only the coterminous states. Even when the map was displayed as a large-size poster, these detailed metropolitan areas were hard to decipher. In these areas, commuting patterns were very localized; therefore, a close-up view of the country's major urban areas was necessary to see the fine detail.

Table 4 lists the 19 inset areas that we added to the maps. The selection of areas for insets was dictated by the maximum number that we could accommodate on the map. The choice of some major urban areas (e.g., New York City, Los Angeles, Chicago, and most of the megalopolis from Boston to Washington, DC) was obvious because of their size and social importance.

Table 4. Areas for which detailed insets were included in the three poster-size maps (Maps 1–3)

States	Substate or Cross-State Regions	Metropolitan Areas and Cities
Alaska	San Francisco-Oakland-San Jose area, California	Anchorage, Alaska
Hawaii	Denver-Colorado Springs area, Colorado	Los Angeles and suburbs, California
	Oahu island, Hawaii	San Diego and suburbs, California
	New York City, New York-northern New Jersey-Philadelphia, Pennsylvania, area	Miami-West Palm Beach area, Florida
	Seattle-Tacoma area, Washington	Atlanta and suburbs, Georgia
	Washington, DC-Baltimore, Maryland, area	Chicago and suburbs, Illinois
		Boston and suburbs, Massachusetts
		Detroit and suburbs, Michigan
		Central North Carolina, from Charlotte to Raleigh and Fayetteville
		Dallas-Fort Worth and suburbs, Texas
		Houston and suburbs, Texas

Note: Areas are listed in alphabetical order by state.

After these insets were in place, we selected the country's other major urban areas and corridors from all the areas that might be displayed. Some examples of areas that could have used a closer view include Phoenix, Arizona; Cleveland, Ohio; Minneapolis-St. Paul, Minnesota; Kansas City and St. Louis, Missouri; and Tampa-St. Petersburg, Florida; there was simply not enough room to accommodate these areas. Other cities, such as Houston and Dallas-Fort Worth, Texas, and Atlanta, Georgia, were considered to be larger urban areas and were, consequently, displayed. A somewhat even spread across the country was also a factor in choosing the insets; therefore, we elected to show the Denver, Colorado, and Seattle, Washington, areas.

We had to organize, size, and place these insets carefully. To retain the coterminous United States as the largest feature of the maps in sufficient detail, while also displaying the insets in sufficient detail, we had to make compromises. The three-map set—i.e.,

Maps 1–3—maintained consistency for display and practicality purposes. Furthermore, we elected to make the insets equal in size; the size of insets is often used to indicate priority or importance (such that larger insets are higher priority or more important than smaller insets). Figures 9–11 are the inbound commuter insets for the New York–Philadelphia, Chicago, and Atlanta areas, respectively.
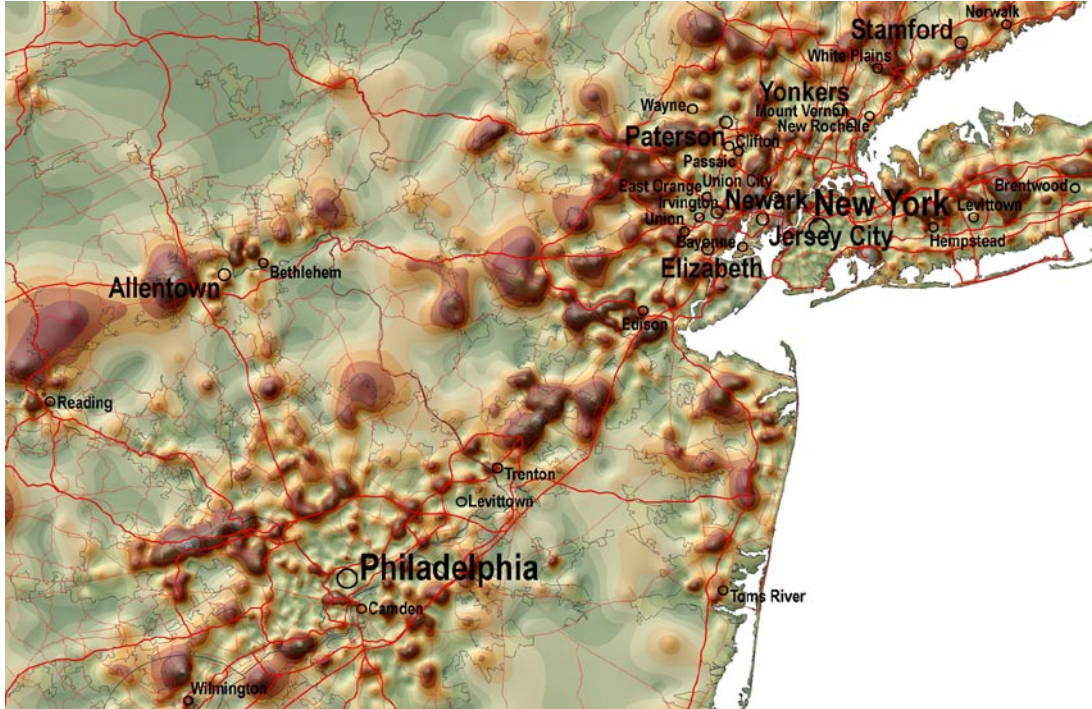

Figure 9. Inbound commuting map that shows the New York City, N.Y.–Philadelphia, Pa., area

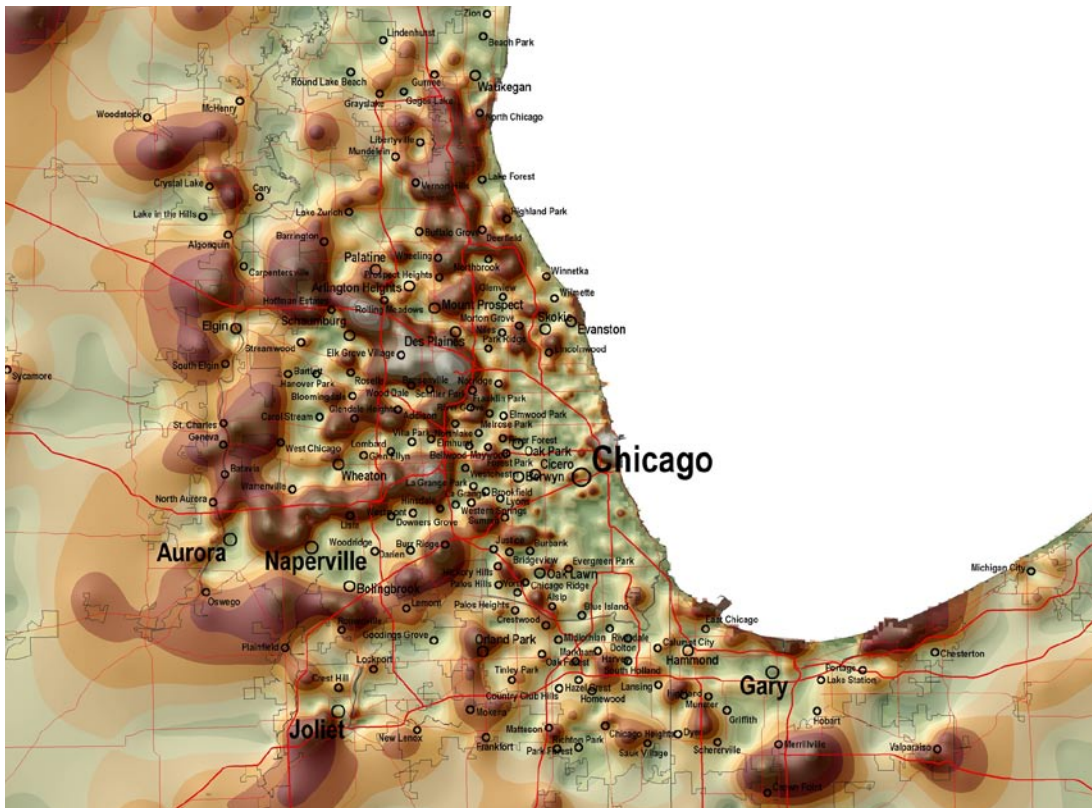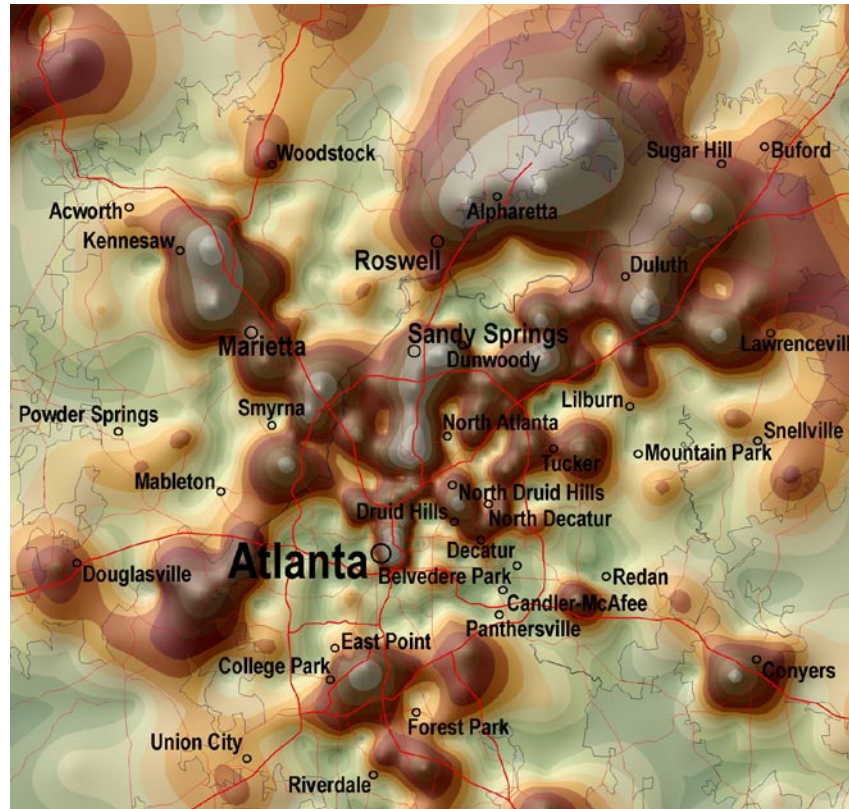*Click on the map to view the layers in an interactive PDF.*

Figure 10. Inbound commuting map that shows the Chicago, Ill., area

Click on the map to view the layers in an interactive PDF.

**Figure 11.
Inbound
commuting map
that shows the
Atlanta, Ga., area**

*Click on the
map to view
the layers in an
interactive PDF.*

Discussion

MIDAS Models and Commuting Maps

Mapping the US Census Bureau STP64 commuting data in a physical-terrain-like design produces discernable patterns in the inbound, outbound, and net gain or loss commuting maps. The MIDAS models focus on the contact between individuals. The distance traveled to make the contact is not as important as the fact that the contact was made. For instance, two equivalently infected individuals who come into contact with 10 people during a day have the same probabilities of transmission, regardless of the distance they have to travel to make those 10 contacts. However, what is relevant to the dynamics of the spread of infectious disease is the understanding of the different patterns of contact in various areas of the country. The patterns of contact affect the patterns of possible transmission of infection and, therefore, the methods of public health intervention.

These commuting data do have some limitations. The data reflect the Census from 2000, and no further detailed STP64 data will be available until 2010 Census; nonetheless, the maps give a good indication of the general commuting patterns among census tracts even as 2010 approaches. Another drawback is that the Census Long Form that was used to derive the STP64 files allows respondents to record only one work address where they worked most in the past week; the locations of the other places of employment are also not known. Therefore, complete commuting data are not captured for individuals who work more than one job. In addition, individuals are not asked which shift they work; therefore, determining how workers come in contact during different shifts is not possible.

Finally, the commuting data are aggregated by census tract so that these data can be used for mapping. For agent-based modeling purposes, this requires that an allocation process be used to determine which individuals from one tract travel to another tract. Owing to the variability in census tract size, the commuting data details vary geographically; however, the variability in census tract size is not a critical feature for these models. Wheaton et al. (2009) describe more completely how the STP64 data are used to develop commuting assignments in agent-based synthesized agent databases.

Examples of Information Gained from MIDAS Models and Maps

Inbound Map (Map 2). The surface of the inbound map, which shows where people are commuting to work, is a variety of peaks that rise anywhere there is a highly populated town or city and, hence, work opportunities. Figure 12, which focuses on Tennessee and parts of surrounding states, is a good example of this inbound commuting pattern. Notice the high peak on the Tennessee-Kentucky border, just northwest of Clarksville, Tennessee. This high peak is Fort Campbell, which is a major employer in the area. Fort Knox also shows up as a high peak, just south of Louisville, Kentucky. Likewise, other military bases (e.g., Fort Bragg near Fayetteville, North Carolina; Camp Pendleton north of San Diego, California; and Fort Carson near Colorado Springs, Colorado) appear prominently on the map as high peaks.

The census tract that has the highest number of inbound commuters (153,358) is located in the center of downtown Houston, Texas. The second highest number of inbound commuters is just northwest of downtown Dallas, Texas (130,272), and the third

highest figure is in the center of downtown Chicago, Illinois (116,503).

Further inspection of the US inbound map reveals other interesting aspects of commuting. For example, the Walt Disney World Resort complex appears on the map as an obvious high peak of inbound commuters just southwest of Orlando, Florida. In addition, Research Triangle Park, a high peak that rises between Raleigh and Durham, North Carolina, contains a very large number of employees, all of whom commute to this area.

The surface pattern of the STP64 data also reveals that, in most large cities, many commuters who live in the outer suburbs do not commute into the inner city to work; instead, they commute to jobs in outlying suburbs. This is most apparent in the Atlanta, Georgia, inset (Figure 11), where areas in the central city attract few commuters and instead show deep, widespread valleys. Yet large peaks rise high in the growing northern suburbs of Marietta, Alpharetta, and Sandy Springs, Georgia. *Not apparent in the map are the number of people who live and work in the same census tract in the inner city.* This
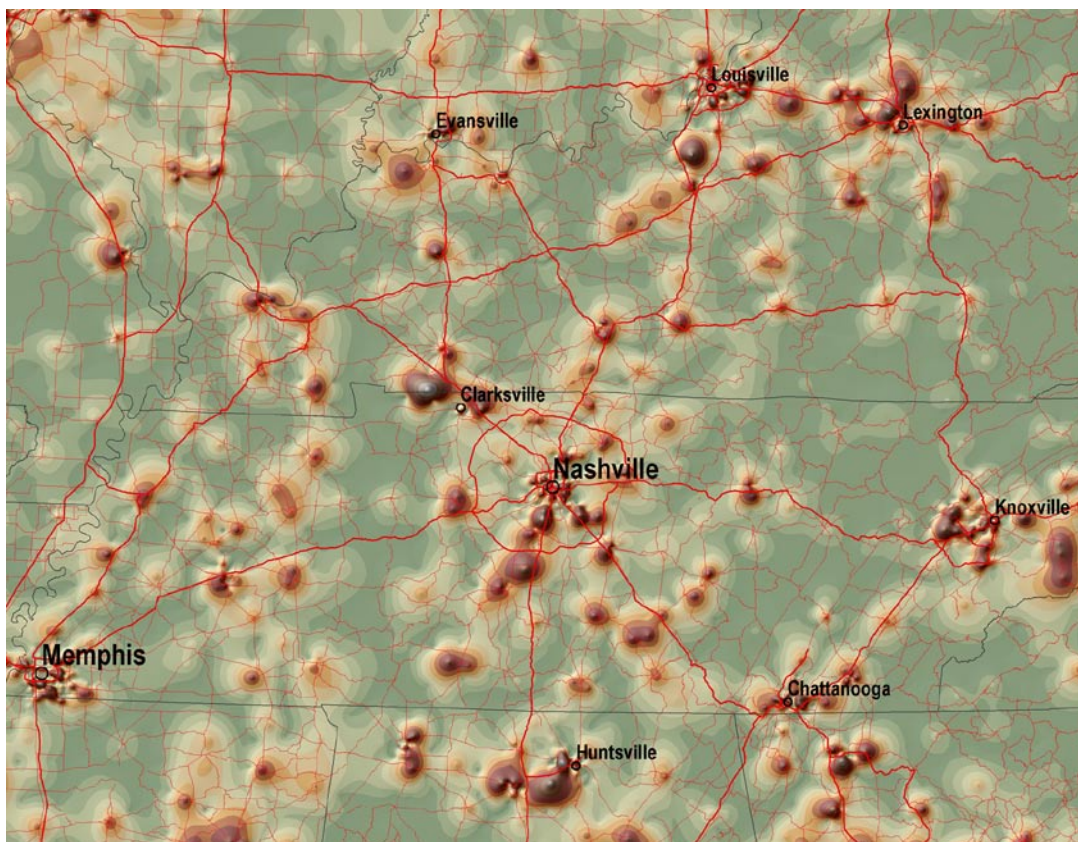

Figure 12. Inbound commuting map that shows Tennessee and surrounding states

Click on the map to view the layers in an interactive PDF.

pattern of flocking to the suburbs for job growth is also visible in other large cities, such as Detroit, Michigan; Washington, DC; and Philadelphia, Pennsylvania. In Hawaii, the Honolulu coast and the military bases near Pearl Harbor draw a high number of commuters.

Outbound Map (Map 1). The outbound commuting map and insets, which show where people are coming from when they leave work and return to their census tract of residence, reveal a much different surface from that of the inbound map. Immediately apparent is the fact that vast areas of the country have a high number of commuters who leave their residential census tract to go to work elsewhere. Symbolizing the upper elevations of the surface (i.e., the highest number of outbound commuters), dark brown and light gray areas spread wide with far-reaching fingers.

The outbound map displayed in Figure 13 shows roughly the same area of Tennessee and adjoining states as Figure 12. Areas with high numbers of outbound commuters expand from major cities and other areas. This confirms that many people over wide areas travel at least outside their census tract of residence to go to work. For example, the large

peak of outbound commuters southeast of Nashville, Tennessee, sits over the City of Murfreesboro, Tennessee. The inbound map (Figure 12) shows that Murfreesboro does not have a large number of people who commute into the city; therefore, one might conclude that a large number of people may be commuting to Nashville. Figure 13 clearly shows a great deal of daily movement of people out of their census tract of residence to get to work.

Interestingly, the census tracts with the three highest numbers of outbound commuters all are located in military bases. Camp Pendleton, north of San Diego, California, has the highest number of outbound commuters (23,417), followed by Fort Bragg near Fayetteville, North Carolina (19,191), and the Naval Station in Norfolk, Virginia (16,959). These numbers may suggest that spouses of personnel on these bases and other nearby residents leave these bases to return to different residential census tracts.

When viewing the US outbound map's insets, one can immediately see that the high number of people living in outer city fringes commute elsewhere. Figure 14 is the Atlanta, Georgia, inset from the outbound map. Deep valleys exist in the downtown

Figure 13. Outbound commuting map that shows Tennessee and surrounding states
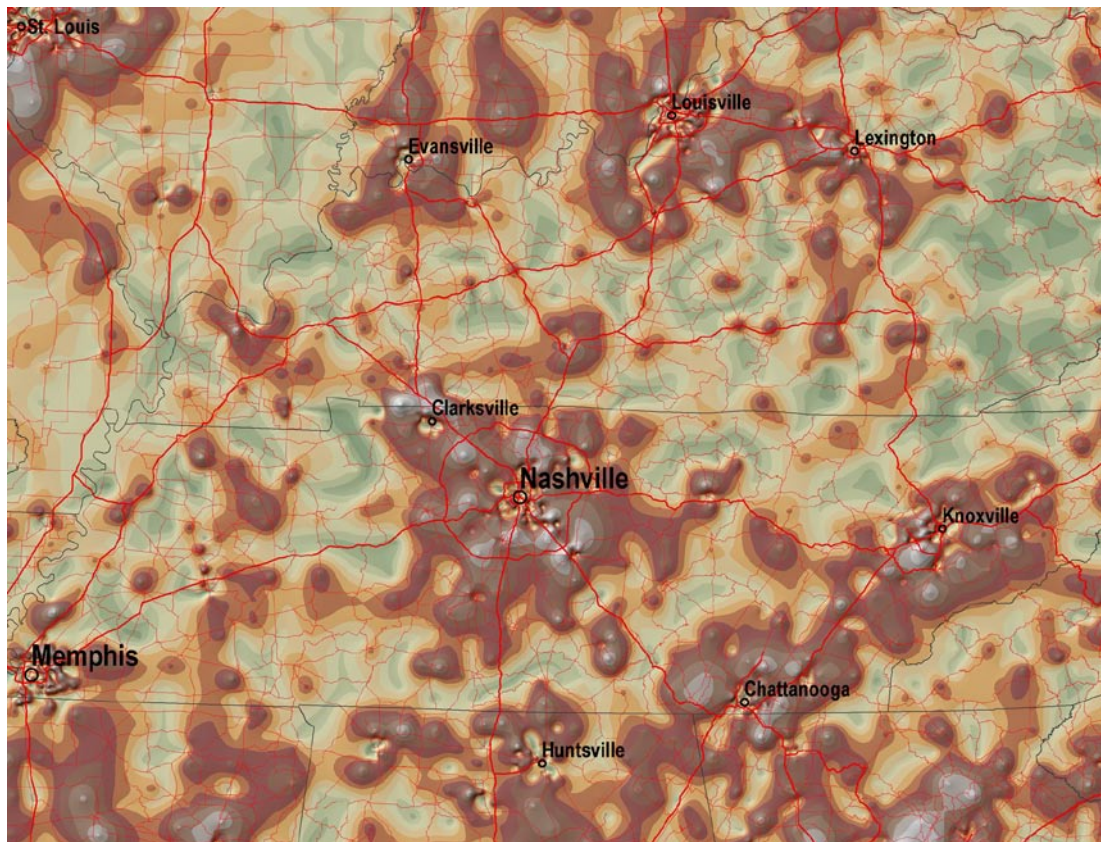
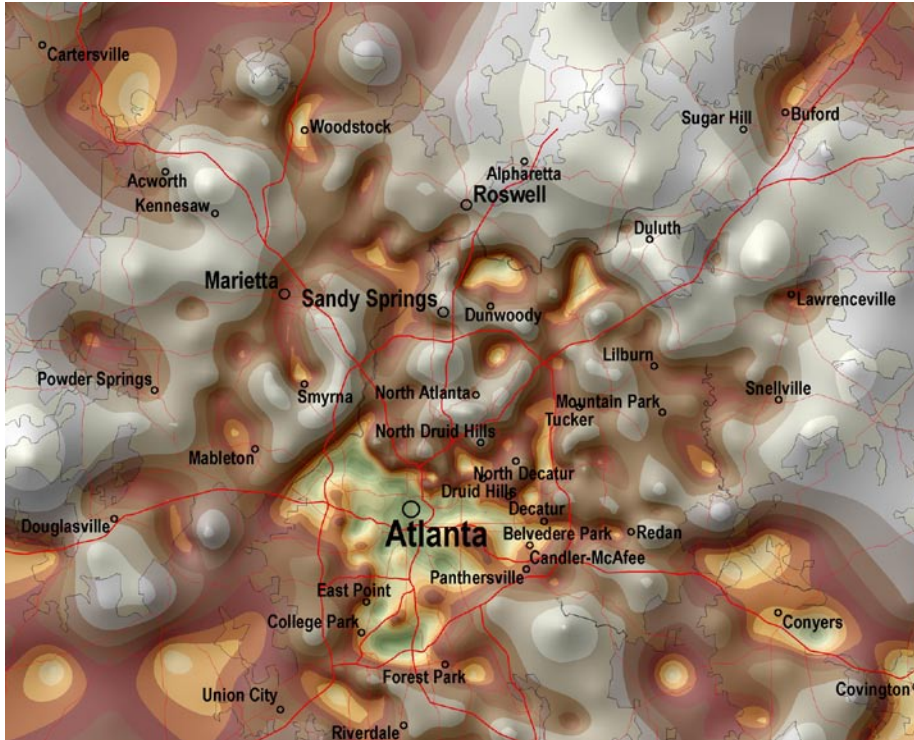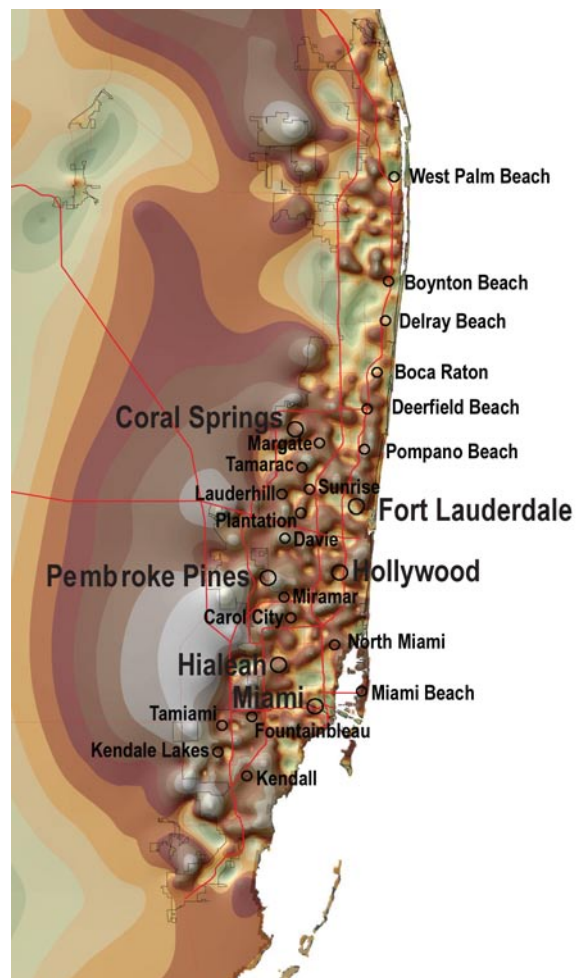*Click on the map to view the layers in an interactive PDF.*

area, but massively high peaks and plateaus of outbound commuters who live in the suburbs surround the city.

This type of terrain is also observed in other large cities such as Houston, Texas; Detroit, Michigan; Indianapolis, Indiana; and Chicago, Illinois. Great numbers of people live in suburbs, but they must travel a distance away from home every day to get to work. In Miami, Florida, bedroom communities are built on the urban area's western edge near the Everglades (Figure 15).

Figure 15. Outbound commuting map that shows the Miami, Fla., area

Net Change Map (Map 3). On the net change commuting map, the areas that gain more people because commuters travel there for work stand out prominently; these areas are symbolized by green-to-brown-to-white color ramps. These areas are surrounded by extensive blue areas from which more people leave for work than travel to for work. Figure 16 shows an area similar to that in Figures 12 and 13. Notice that the areas that gain people are the same locations as the areas of peaks in the inbound map, although they are smaller and more compact as a result of presenting these data as a net gain or loss of commuters. (Click here to see the maps together.)

Figure 17 is the Atlanta inset from the net gain or loss map. This map clearly shows the "islands of commerce," which is a common pattern in large cities where deep, residential "seas" are surrounded by "islands" where more people go to work.

The highest net gain of commuters (152,062) is located directly in the center of downtown Houston. The second highest net gain of commuters is located just northwest of downtown Dallas (129,551), followed by the center of downtown Chicago (112,118). These three places are precisely the same as the three top inbound areas.

When comparing Figure 11 with Figure 17 for Atlanta inbound and net change commuters, respectively, the pattern of hills on the inbound map closely matches the pattern of islands on the net change map. (Click here to see the maps together.) Areas with high numbers of inbound commuters also have high net gain numbers. These areas appear to be sizable employment centers that also reflect a high net gain of commuters.

With respect to net loss of commuters, an area on the east side of Camp Lejeune Marine Corps Base (near Jacksonville, North Carolina) has the highest net loss (11,006). An area west of Atlanta ranks second (10,760), followed by an area near Bithlo, Florida, on the eastern edge of Orlando (10,147).

Policy Applications

We developed these maps for investigators and policy makers to understand population movements between social settings where infectious disease may spread. They have many other potential uses, however. For example, analysis of these maps can assist decision makers in identifying segments of main roadways and interstate highways that experience heavy commuting traffic. This information can direct them to focus on areas that may need road repair or widening sooner than others.

Corridors likely to warrant more attention from urban planners for mass transit solutions, such as light rail systems, can be identified. Figure 10 shows corridors that experience high numbers of inbound commuters along the interstates in the western and northwestern suburbs of Chicago, which is a result of urban sprawl and affects infrastructure. Heavy commuting corridors surrounding Atlanta are visible in Figure 11 along the major highways that branch into the sprawling suburbs from downtown Atlanta, especially into the northern suburbs of Marietta, Alpharetta, and Duluth.

As another example: Figure 7 shows an arch of cities from Charlotte to Raleigh in North Carolina. Between Durham and Raleigh appears a prominent dark brown peak with a white summit. This color ramp arises from the high number of commuters to Research Triangle Park, a 7,000-acre site that straddles several census tracts. The Park was originally designed to be centrally located among several urban areas. Today, more than 50,000 people commute to the Park for work. Having identified this location with a high number of commuters, officials might want to examine transportation options (as indeed is the case already for planners in the surrounding counties). They will, upon reviewing the map in detail, note that the Park is served by few major highways or thoroughfares, relative to the number of daily inbound commuters. This fact supports the need to examine transportation options to alleviate traffic congestion and its attendant burden on those commuters. Such maps can help planners, urban designers, legislators, and others identify areas of heaviest use and areas that might warrant various transportation options.
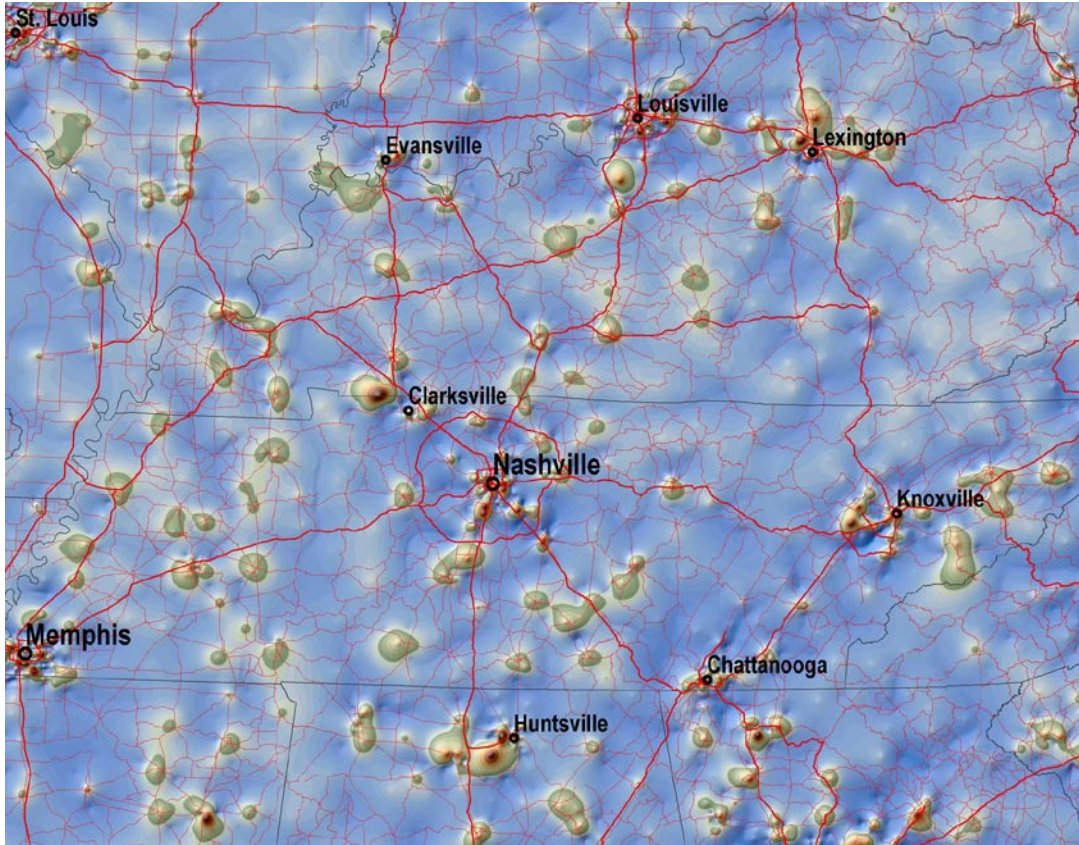
Figure 16. Inbound versus outbound commuting map that shows Tennessee and surrounding states

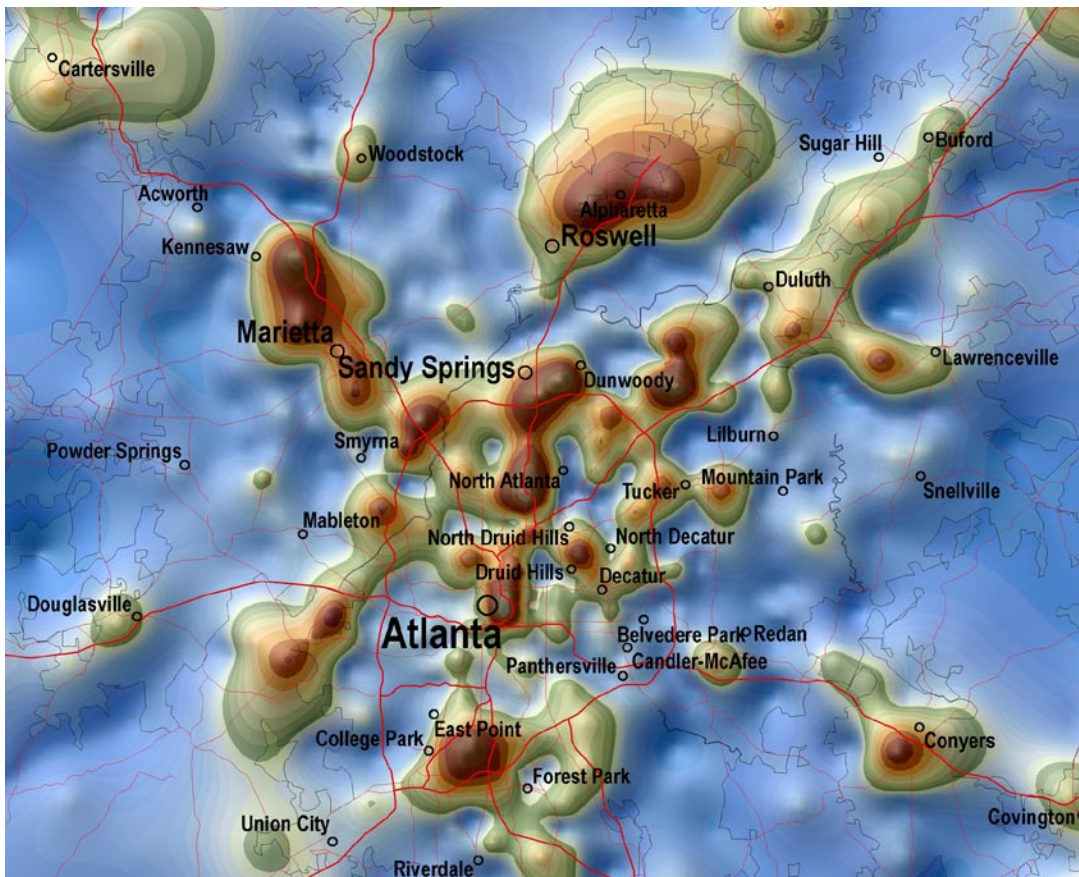*Click on the map to view the layers in an interactive PDF.*

Figure 17. Inbound versus outbound commuting map that shows the Atlanta, Ga., area

Click on the map to view the layers in an interactive PDF.

Summary

We used commuting pattern data from the US Census Bureau as a source of information for two purposes on the MIDAS project. First, we used the data to create workplace assignments whereby individual persons in a synthesized population database were assigned to places of work based on commuting patterns. We then used these data to develop maps that serve as visual information tools for use by urban planners, transportation officials, and other policy makers. Not only do these maps complement other analytical tools and computer models, but they also help public health officials understand the dynamics of the population so that they can effectively respond to naturally occurring events or a bioterrorism attack and manage infectious disease outbreaks.

GIS software is a powerful tool that enriches projects with spatially related data and analysis. An understanding of people's movement patterns can help researchers and government officials better predict the spread of infectious disease and prepare for potential outbreaks.

References

ESRI (Environmental Systems Research Institute International). (2007a). ArcGIS 9.2 desktop help. Retrieved March 6, 2008 from http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?topicname=natural_breaks_(jenks)

ESRI (Environmental Systems Research Institute International). (2007b). *ESRI map book.* Vol. 22. Redlands, CA: ESRI Press.

Jenks, G. F. (1967). The data model concept in statistical mapping. *International yearbook of cartography.* Guntersloh, Germany: Bertelsmann. 7: 186–190.

Knapp, G. R. (1997). Mining the heavens: The Sloan Digital Sky Survey. *Sky and Telescope* 94(2), 40–48. August.

Krygier, J., & Wood, D. (2005). *Making maps.* New York: The Guilford Press.

MacEachren, A. M. (1995). *How maps work.* New York: The Guilford Press.

Pickle, L. W., Feuer, E. J., & Edwards, B. K. (2003). *US predicted cancer incidence, 1999: Complete maps by county and state from spatial projection models.* NCI Cancer Surveillance Monograph Series, Number 5. Bethesda, MD: National Cancer Institute. National Institutes of Health Publication No. 03-5435.

TerraSeer space-time information system [Desktop support]. (2008). About Jenks' natural breaks. Retrieved March 11, 2008 from www.terraseer.com/help/stis/default.htm#interface/map/classify/About_natural_breaks.htm

Tufte, E. R. (2001). *The visual display of quantitative information.* 2nd ed. Cheshire, CT: Graphics Press.

US Census. (2004). *Census 2000 special tabulation product: Census tract of work by census tract of residence (STP 64).* Retrieved March 6, 2008 from http://www.census.gov/mp/www/spectab/stp64-webpage.html

Wheaton, W. D., Cajka, J. C., Chasteen, B. M., Wagener, D., Cooley, P. C., Ganapathi, L., Roberts D., & Allpress, J. L. (2009). Synthesized population databases: A US geospatial database for agent-based models. Research Triangle Park, NC: RTI Press.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory and chemistry services.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International. More information about RTI Press can be found at www.rti.org/rtipress.