Increasing the Accuracy of Urban Population Analysis With Dasymetric Mapping

Jeremy Mennis
Temple University

Abstract

Many types of urban policy analyses, particularly those relating to exposure to hazards or accessibility to resources, rely on accurate and precise spatial population data, although such data are not always available. Dasymetric mapping is a technique for disaggregating population data from one set of source spatial units to a finer resolution set of target spatial units through the use of an ancillary dataset, typically land use, zoning, or similar nominal datasets related to population distribution. Dasymetric mapping operates by employing weights that capture both the relative areas of the target spatial units and the relative population densities of the different nominal ancillary classes, and it is typically implemented in Geographic Information System, or GIS, software. An example application demonstrates the efficacy of the dasymetric approach by comparing census tract-level and dasymetric data in an assessment of the population living in proximity to hazardous air pollutant releases in Philadelphia, Pennsylvania, using block-level data as a validation dataset.

Introduction

Many types of urban policy analyses rely on accurate and precise spatial population data. Of particular note are analyses of exposure and accessibility, where one must assess the population in proximity to, or overlapping with, some geographic feature. Examples of such analyses include the estimation of population exposed to natural and technological hazards, such as flooding or air pollution. Other relevant research applications concern access to amenities and resources, such as recreation facilities, health centers, nutritious food, or employment opportunities.

Although the U.S. Census Bureau provides high-resolution demographic data for the United States, certain variables may be available only over coarser spatial units, such as census tracts.

Other population-related datasets, such as disease incidence data, may be limited to distribution at a coarse spatial resolution for purposes of privacy protection. In many developing nations, population data at a fine resolution are not available at all, because many countries do not have the resources to invest in census infrastructure. In addition, in all these cases, population data are likely to be available aggregated to spatial units that are derived by convenience of enumeration or are a reflection of administrative or political jurisdiction boundaries and, consequently, are unlikely to capture the nature of the actual population distribution. Thus, the development of small-area estimates for urban population data remains a challenge in both developed and developing nations.

Dasymetric mapping is a technique for estimating population in small areas in situations where one has access to population data aggregated only at a relatively coarser scale (Mennis, 2009). It uses ancillary data, an additional dataset related to the distribution of population but distinct from it, to disaggregate population data from one set of spatial units to another set of smaller spatial units. The formal principles of dasymetric mapping were initially developed for a Russian mapping project in the early 20th century (cf. Petrov, 2012) and were introduced to English-speaking audiences in a series of articles appearing in the 1920s and 1930s, most notably in an article by Wright (1936). The dasymetric mapping technique, however, was little known outside cartographic circles until the widespread availability of Geographic Information System (GIS) software and digital data products that could serve as ancillary data, such as those derived from remotely sensed imagery, spurred the growth of dasymetric mapping algorithms and applications beginning in the 1990s through the present.

Dasymetric mapping more recently has been employed for a wide variety of applications that benefit from high spatial resolution population data, including environmental justice (Mennis, 2002), public health (Maantay, Maroko, and Porter-Morgan, 2008), crime (Poulsen and Kennedy, 2004), and historical population estimation (Gregory and Ell, 2005). It has also been used to create national-level, high-resolution population datasets (Bhaduri et al., 2007).

The purpose of the present article is to describe dasymetric mapping, its theoretical basis, and its implementation using GIS software. As an illustration of dasymetric mapping and its application to urban analysis, an example is presented for Philadelphia, Pennsylvania, where tract-level population data are disaggregated to sub-tract-level spatial units. These data are then used for an analysis of population residing in close proximity to facilities releasing hazardous pollutants to the atmosphere. The tract and the dasymetric data are then compared with an analogous analysis using census block-level data for accuracy assessment.

The Dasymetric Mapping Technique

Dasymetric mapping can be considered a form of areal interpolation, the transformation of data from one set of spatial units to another set of spatial units; for example, the assignment of population originally encoded in U.S. counties to a set of watershed boundaries. The original set of spatial units is referred to as *source zones* and the set of destination spatial units is referred to as *target zones*. The simplest approach to areal interpolation is areal weighting, which assumes a homogeneous distribution of the data within the source zones. Thus, data are apportioned to the target zones based on the proportional area that each source zone contributes to each target zone.

A particular case of areal interpolation occurs when the target zones are formed by the geometric intersection of the source zones with another—ancillary—polygon data layer, so that the target zones spatially nest perfectly within the source zones, and each source zone can be disaggregated into one or more target zones. Areal weighting in this case implies that given a target zone f nested within a source zone g, such that $f \in g$, then the population of the source zone can be distributed to its constituent target zones based on the area ratio (AR) of each target zone, where $AR_f = A_f / A_g$, and where A is area. The target zone population can then be estimated as $\hat{y}_f = y_g AR_f$, where $\hat{y}_f$ is the estimated count of the target zone and y_g is the population of the host source zone (Goodchild and Lam, 1980).

Dasymetric mapping can be viewed as an extension of areal weighting in which the ancillary dataset overlaid with the source layer is typically an area-class map, which exhaustively tessellates a region into nominal classes that are related to the distribution of the variable being mapped. Thus, dasymetric mapping incorporates not only the relative proportion of the contributing area of each target zone but also its ancillary class to redistribute data from the source zone to its constituent target zones. As such, dasymetric mapping employs not only the area ratio but also a density ratio among the ancillary classes to make target-level estimates. If we formally consider an ancillary class c associated with target zone f, the density ratio (DR) can be defined as $DR_c = (\hat{D}_c)/(\sum \hat{D}_c)$, where $\hat{D}_c$ is the estimated density of the ancillary class c. The total fraction (TF) integrates the area ratio and density fraction into a single term, where

$$TF_{fc} = \frac{AR_f \, DR_c}{\sum_{f \in g}(AR_f DR_c)} \,. \tag{1}$$

The target zone population can then be estimated as

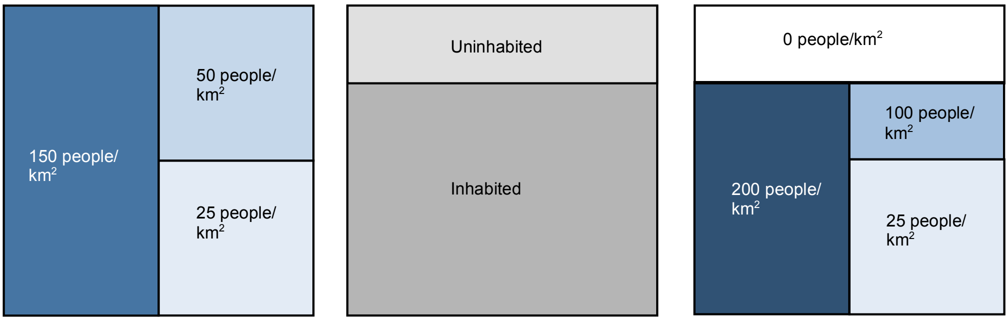
$$\hat{y}_f = y_g (TF_{fc}). \tag{2}$$

Note that the sum of the population of each source zone is maintained in the dasymetric output (Tobler, 1979), because the area ratio and density ratio both sum to 1 for each source zone. The value of $\hat{D}_c$ can be set by the analyst through his or her own expert knowledge (Eicher and Brewer, 2001) or it can be estimated by sampling the variable values of source layer zones that are spatially coincident with different ancillary data classes (Mennis, 2003). Values of DR_c can also be set directly by the analyst without setting the values of $\hat{D}_c$ in cases in which one is knowledgeable about only the relative densities among the classes.

By far the most common ancillary dataset used in dasymetric mapping of population is land use or land cover data, often derived from classified remotely sensed imagery. The most basic dasymetric mapping implementation involves the use of such ancillary data to simply distinguish between inhabited and uninhabited land area; for instance, by distinguishing between developed regions and those occupied by water or barren land. In this case, all population in a source zone bisected by uninhabited and inhabited land would be allocated to the land classified as inhabited, leaving the remaining portion of the source zone with zero population. Exhibit 1 illustrates this principle of dasymetric mapping using a schematic diagram. A set of source zones with observed population densities is shown on the left, with an ancillary land cover data layer used in the dasymetric

Exhibit 1

Illustration of the Binary Method of Dasymetric Mapping: Source Zones (left), Ancillary Layer (middle), and Target Zones (right)



km² = square kilometers.

mapping shown in the middle. The resulting dasymetric map is shown on the right, where all the population of a source zone with both inhabited and uninhabited regions is apportioned to the inhabited region. Thus, the population density of the inhabited regions of a source zone increases and is conversely held to zero in the uninhabited regions.

This general principle can be easily extended to the use of other ancillary datasets besides land use/cover, including road density data (Reibel and Bufalino, 2005), zoning- and parcel-level data (Maantay, Maroko, and Herrmann, 2007), address point data (Zanderbergen, 2011), unclassified remotely sensed spectral values (Holt, Lo, and Hodler, 2004), and scanned raster reference maps (Langford, 2007). In addition, as equation 1 indicates, the technique can accommodate not only two ancillary classes with DR_c values of zero (uninhabited) and one (inhabited) but also multiple classes with values ranging continuously between zero and one, so that population can be allocated in a more complex manner than by simply distinguishing between inhabited and uninhabited regions.

Demonstration

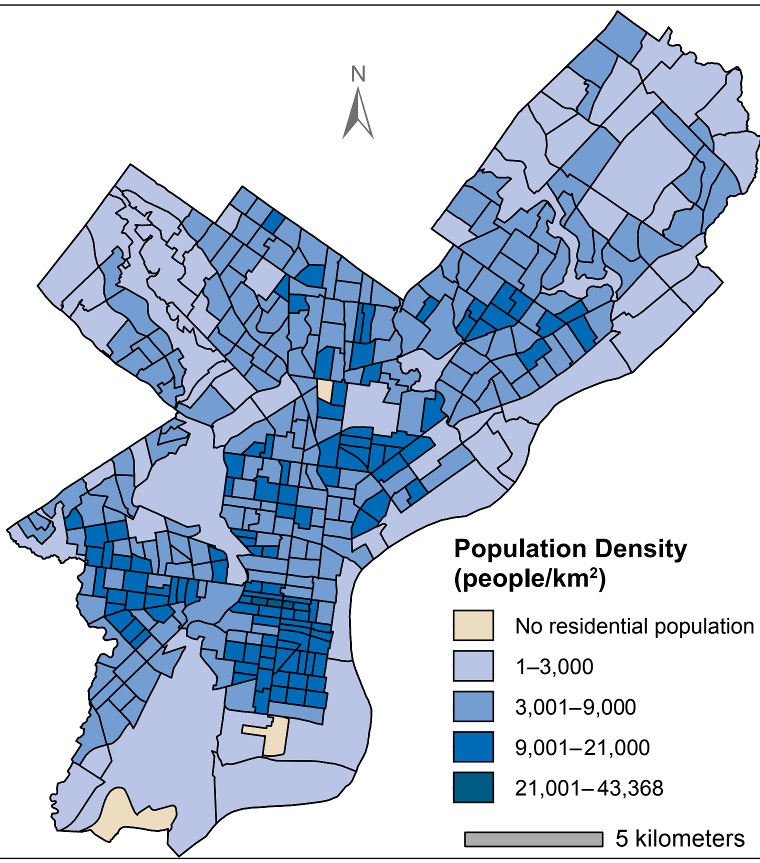
As a demonstration of the dasymetric mapping for small-area estimation of urban population, dasymetric mapping is used to estimate population data for the city of Philadelphia, mapped to the census tract level at sub-tract spatial units. Tract-level data and dasymetric mapping-level data are then compared in an analysis of the population at risk of air pollution to illustrate how dasymetric mapping can be used for urban analysis. Population data at the census block level are used as a validation dataset to compare the analytical results of the tract- and dasymetric-based measurements of population exposure. Note that this analysis is intended strictly for purposes of illustration of the dasymetric approach and is not intended to address issues of environmental exposure to air pollutants in Philadelphia, which would require a more thorough analysis.

Data and Implementation

Total population derived from the 2010 Census at the tract and block levels were acquired from the Census Bureau American Factfinder website. These data include 384 tracts (18,872 blocks) with a total population of 1,526,006 (exhibit 2). Ancillary data related to population distribution in Philadelphia are necessary to facilitate the dasymetric mapping. For this purpose, zoning data for Philadelphia were acquired from the City of Philadelphia. These polygon data encode allowable land uses and building restrictions, coded as the following zoning classes: high-density residential, low-density residential, commercial/residential mixed use, commercial nonresidential, industrial nonresidential, parks and other related nonresidential land uses, and nonresidential transportation infrastructure. These classes were aggregated to reflect low-density residential, high-density residential (including commercial/residential mixed use), and nonresidential areas (exhibit 3).

Exhibit 2

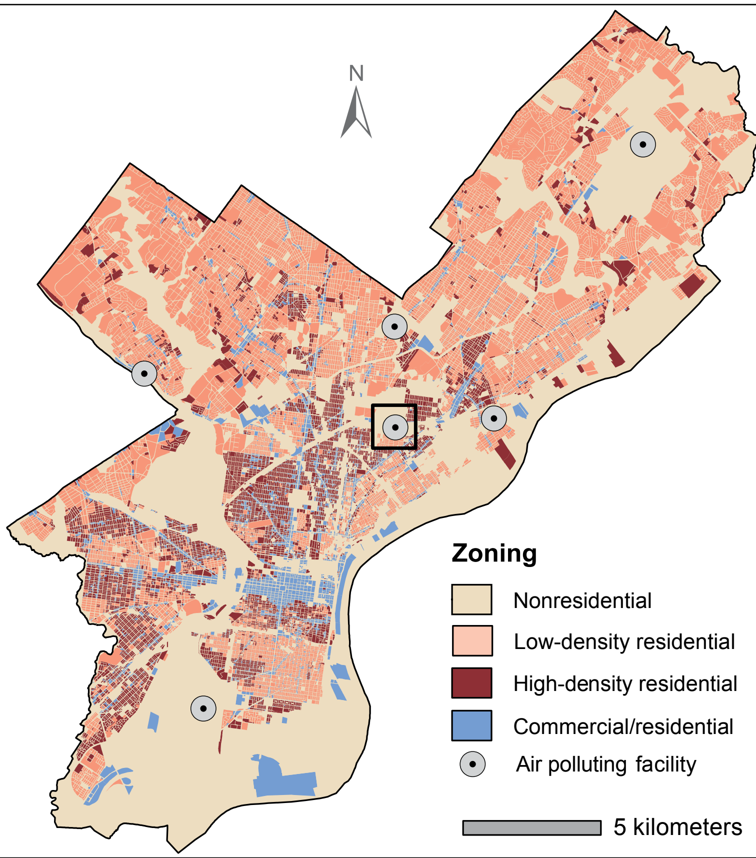
Map of Population Density by Tract in Philadelphia



km^2 = square kilometers.
Source: U.S. Census Bureau, with calculations by the author

Exhibit 3

Map of Zoning Classes Used As Ancillary Data in the Dasymetric Mapping: Nonresidential, High-Density Residential, and Low-Density Residential



Note: The locations of air polluting facilities are also included.
Source: City of Philadelphia, with calculations by the author

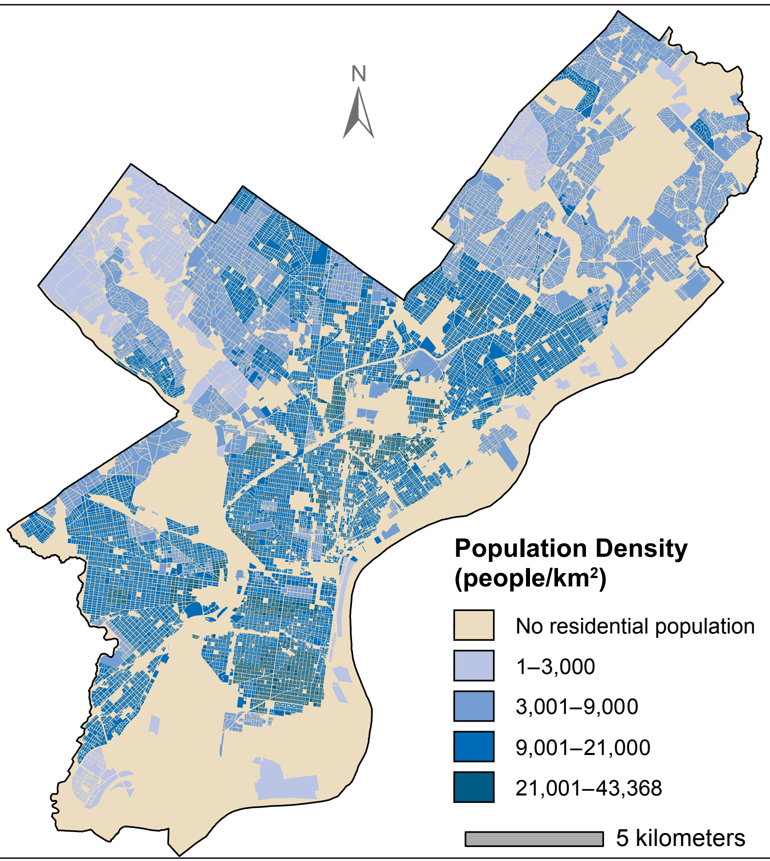
GIS software was used to process the data, perform the dasymetric estimation, and implement equations 1 and 2. The tract layer and the ancillary zoning layer were intersected to produce a target layer composed of 30,271 polygons. Each term in equation 1 was then calculated and stored in a field in the target layer attribute table. The area ratio (AR_f) was calculated as the ratio of the target zone area to its host source zone (tract) area. The density ratio for the spatial ancillary zoning class data (DR_c) was set to values of 0.30 for low-density residential areas, 0.35 for both the high-density residential areas and the commercial and residential mixed-use areas (typically downtown apartment buildings with stores located on the first floor), and 0.0 for the other nonresidential zoning classes. Although these values are acknowledged as being somewhat arbitrary, they reflect the exclusion of population from nonresidential areas of the city and also the greater concentration of population in high- versus low-density residential zoning classes.

The value for $AR_f DR_c$ was calculated and encoded in another field in the target layer attribute table. Using a summarize function in the GIS, the sum of all target layer polygons' $AR_f DR_c$ value was calculated for each individual tract, and these data were joined onto the target layer attribute table. The TF_{fc} value—that is, the ratio of each target polygon's $AR_f DR_c$ to the sum of all $AR_f DR_c$ in that target polygon's host tract—was then calculated and stored in another field in the target layer's attribute table. Finally, the population value for each target polygon was calculated according to equation 2 by multiplying the host tract population by the TF_{fc} for each target polygon.

The resulting dasymetric map of population density is shown in exhibit 4. One can clearly see that the precision of population distribution is far higher in the dasymetric map as compared with the tract-level map. The tract-level map has only three small tracts that have no residential population, whereas the zoning map clearly shows large areas of the city that are nonresidential. The dasymetric map prohibited population from the areas zoned nonresidential and allocated the remaining population to the residentially zoned areas. Areas zoned high-density residential and commercial and residential were allocated population at a greater proportion as compared with low-density residential zoned land, after accounting for differences in the area of the target polygons.

Exhibit 4

Dasymetric Map of Population Density in Philadelphia



km^2 = square kilometers.
Source: Calculations by the author

Calculating the Population Located Near Air Polluting Facilities

As a way of illustrating the utility of the dasymetric mapping, a relatively simple analysis is conducted of the population located near facilities releasing pollutants to the air. Data on the locations of facilities releasing or disposing of more than 10,000 pounds of hazardous air pollutants on site in 2010 were acquired from the U.S. Environmental Protection Agency's Toxic Release Inventory Program. Six facilities were mapped (exhibit 3) and population counts were tallied within a series of distances from the facilities, using the tract, block, and dasymetric data population maps. First, the total populations of the six tracts, blocks, and dasymetric polygons that contained the six facilities were calculated. Then, those tracts, blocks, and dasymetric polygons within 0.25 kilometer of a facility were selected, those within 0.5 kilometer of a facility were selected, and so on up to 1.0 kilometer from a facility.

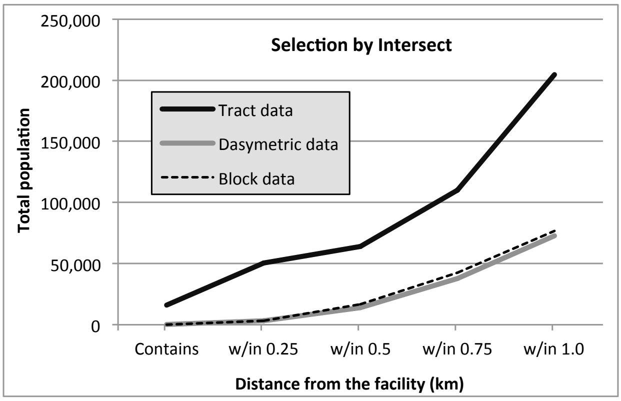
Two prominent GIS-based methods of selecting polygons within a certain distance of a set of point features were employed (Mennis, 2003). The first method, referred to as the *intersect* method, selects all those polygons that overlap the distance buffer. So, for example, a tract for which any portion of the tract falls within the specified distance of a facility would be selected as being within that distance of that facility. The second method, called the *centroid* method, selects only those polygons whose geometric center falls within the buffer distance.

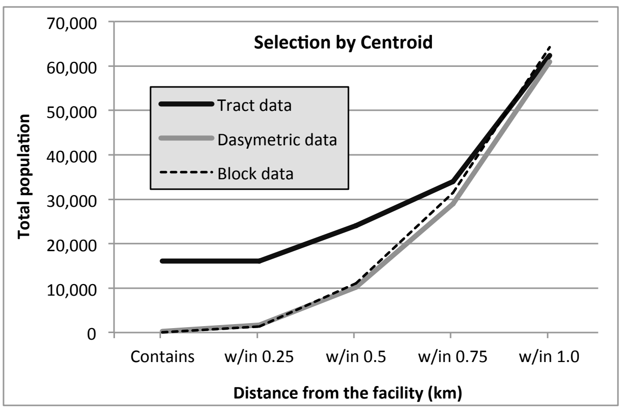
Exhibit 5 shows two graphs indicating the differences among the tract-level, block-level, and dasymetric-level population calculations, with total population shown on the y-axis for each measured distance from the each facility. The graph at the top shows the results for the intersect method of polygon selection and the bottom graph shows the results using the centroid method. For both methods of selection, the tract-based calculations clearly tend to overestimate the total population nearby as compared with the dasymetric-based calculations. For the intersect method of selection, the difference between the tract data and the dasymetric data increases with increasing distance. At a distance of 1 kilometer, the tract-level population estimate is nearly three times the dasymetric-level estimate. This pattern occurs because the number of tracts under consideration increases substantially as the distance from a facility increases—because the tract data are a much coarser resolution than the dasymetric data, the area considered within any given distance of a facility using the intersect method is much larger using the tract data as compared with the dasymetric data.

For the centroid method, the maximum difference between the tract and dasymetric datasets is observed at distances nearest to the facilities. At a distance of 0.25 kilometer, the tract-level population estimate is nearly 10 times the dasymetric-level estimate. As the distance increases, the estimation of the total populations for the different datasets tends to converge at a distance of approximately 1 kilometer. The reason for this can be observed in exhibit 6, which shows a closeup view of the facility in the bolded box in exhibit 3, where the bolded circle and cross-hatch pattern show the area within 0.5 kilometer of the facility. The tract data on the left indicate that the facility lies nearly at the intersection of three separate tracts; thus, the calculation of the population at risk, using data derived from the host tract, is likely to be inherently inaccurate, because air pollutants would likely spread across tract boundaries. The dasymetric data on the right shows a far greater spatial variation in population distribution, where it is clear that most of the area within 0.5 kilometer of the facility is nonresidential. Indeed, the entire population within the 0.5-kilometer distance is concentrated in the southern portion of the buffer.

Exhibit 5

Graphs of the Total Population Within Certain Distances of the Air Polluting Facilities Using the Tract, Block, and Dasymetric Population Data

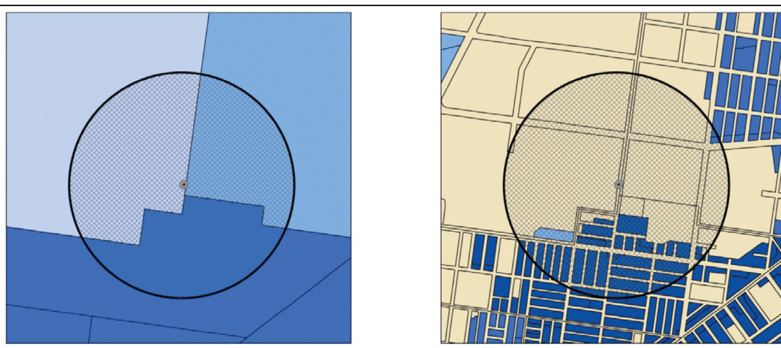




km = kilometer. w/in = within.

Exhibit 6

A Visual Comparison of the Tract (left) and Dasymetric (right) Population Data Within 0.5 Kilometer of the Air Polluting Facility Shown in the Box in Exhibit 3



Importantly, for both the intersect and centroid selection methods, the graph lines for the block-level data closely mirror those of the dasymetric data. Indeed, it is very clear visually that the tract-level data for both methods substantially overestimate the population within a given distance of a facility, whereas the dasymetric data provide a far more accurate depiction.

Conclusion

This article outlines the general principles of dasymetric mapping and offers a demonstration of its efficacy in urban population analyses. The use of coarse resolution population data relative to the scale of analysis, or the use of population data aggregated to spatial units unrelated to the actual distribution of population, can result in inaccurate assessments of urban population exposure and accessibility. Dasymetric mapping offers substantial potential for improving estimates of population exposure and accessibility through the estimation of population at a much finer scale, through the integration of often publicly available ancillary data. In addition, the basic principles of dasymetric mapping are relatively easy to implement in many commercial and open-source GIS software packages.

More sophisticated approaches to dasymetric mapping that rely on regression, kriging, and iterative algorithms have also been developed (for example, Leyk, Nagle, and Buttenfield, 2013; Liu, Kyriakidis, and Goodchild, 2008). Research suggests, however, that, although the accuracy of dasymetric mapping is dependent on the nature of the algorithm and ancillary data source, even relatively simple efforts to incorporate ancillary data into dasymetric population estimation typically result in significant improvements in population estimations over areal weighting (Langford, 2013; Zanderbergen and Ignizio, 2010). Thus, urban analysts with even basic knowledge in GIS should be able to effectively implement and benefit from dasymetric mapping.[1]

Acknowledgments

The author thanks the editors for their helpful comments in improving this article.

Author

Jeremy Mennis is a professor in the Department of Geography and Urban Studies at Temple University.

[1] The dasymetric mapping technique described here was implemented in two publicly available scripts for ArcGIS (Environmental Systems Research Institute, Inc.). The first was implemented by Rachel Sleeter at the U.S. Geological Survey (Sleeter and Gould, 2007) and is available at http://geography.wr.usgs.gov/science/dasymetric/. The second was implemented by Torrin Hultgren (Mennis and Hultgren, 2006) and is available at http://enviroatlas.epa.gov/enviroatlas/Tools/Dasymetrics.html. Responsibility for the use and application of the dasymetric mapping scripts and their products lies with the user.

References

Bhaduri, Budhendra, Edward Bright, Phillip Coleman, and Marie L. Urban. 2007. "LandScan USA: A High-Resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics," *GeoJournal* 69 (1–2): 103–117.

Eicher, Cory L., and Cindy A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation," *Cartography and Geographic Information Science* 28: 125–138.

Goodchild, Michael F., and Nina S. Lam. 1980. "Areal Interpolation: A Variant of the Traditional Spatial Problem," *Geo-Processing* 1: 297–312.

Gregory, Ian N., and Paul S. Ell. 2005. "Breaking the Boundaries: Integrating 200 Years of the Census Using GIS," *Journal of the Royal Statistical Society, Series A* 168: 419–437.

Holt, James B., C.P. Lo, and Thomas W. Hodler. 2004. "Dasymetric Estimation of Population Density and Areal Interpolation of Census Data," *Cartography and Geographic Information Science* 31: 103–121.

Langford, Mitchel. 2013. "An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data," *Geographical Analysis* 45: 324–344.

———. 2007. "Rapid Facilitation of Dasymetric-Based Population Interpolation by Means of Raster Pixel Maps," *Computers, Environment and Urban Systems* 31: 19–32.

Leyk, Stefan, Nicholas N. Nagle, and Barbara P. Buttenfield. 2013. "Maximum Entropy Dasymetric Modeling for Demographic Small Area Estimation," *Geographic Analysis* 45: 285–306.

Liu, X.H., Phaedon C. Kyriakidis, and Michael F. Goodchild. 2008. "Population-Density Estimation Using Regression and Area-to-Point Residual Kriging," *International Journal of Geographical Information Science* 22 (4): 321–447.

Maantay, Juliana A., Andrew R. Maroko, and Christopher Herrmann. 2007. "Mapping Population Density in the Urban Environment: The Cadastral-Based Expert Dasymetric System (CEDS)," *Cartography and Geographic Information Science* 34: 77–102.

Maantay, Juliana A., Andrew R. Maroko, and Holly Porter-Morgan. 2008. "Research Note: A New Method for Mapping Population and Understanding the Spatial Dynamics of Disease in Urban Areas: Asthma in the Bronx, New York," *Urban Geography* 29: 724–738.

Mennis, Jeremy. 2009. "Dasymetric Mapping for Estimating Population in Small Areas," *Geography Compass* 3: 727–745.

———. 2003. "Generating Surface Models of Population Using Dasymetric Mapping," *The Professional Geographer* 55: 31–42.

———. 2002. "Using Geographic Information Systems To Create and Analyze Statistical Surfaces of Population and Risk for Environmental Justice Analysis," *Social Science Quarterly* 83 (1): 281–297.

Mennis, Jeremy, and Torrin Hultgren. 2006. "Intelligent Dasymetric Mapping and Its Application to Areal Interpolation," *Cartography and Geographic Information Science* 33: 179–194.

Petrov, Andrey. 2012. "One Hundred Years of Dasymetric Mapping: Back to the Origin," *The Cartographic Journal* 49: 256–264.

Poulsen, Erika, and Leslie W. Kennedy. 2004. "Using Dasymetric Mapping for Spatially Aggregated Crime Data," *Journal of Quantitative Criminology* 20 (3): 243–262.

Reibel, Michael, and Michael E. Bufalino. 2005. "Street Weighted Interpolation Techniques of Demographic Count Estimation in Incompatible Zone Systems," *Environment and Planning A* 37: 127–129.

Sleeter, Rachel, and Michael Gould. 2007. "Geographic Information System Software To Remodel Population Data Using Dasymetric Mapping Methods," *U.S. Geological Survey Techniques and Methods* 11–C2: 15.

Tobler, Waldo. 1979. "Smooth Pycnophylactic Interpolation for Geographical Regions," *Journal of the American Statistical Association* 74: 519–530.

Wright, John K. 1936. "A Method of Mapping Densities of Population," *Geographical Review* 26: 103–110.

Zanderbergen, Paul A. 2011. "Dasymetric Mapping Using High Resolution Address Point Datasets," *Transactions in GIS* 15(s1): 5–27.

Zanderbergen, Paul A., and Drew A. Ignizio. 2010. "Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates," *Cartography and Geographic Information Science* 37 (3): 199–213.